INVESTIGATING THE GENETIC AETIOLOGY OF THREE FACIAL DYSOSTOSES IN SOUTH AFRICA

Patracia Nevondwe

A dissertation submitted to the Faculty of Health Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Master of Science in Medicine

Johannesburg, 2019

DECLARATION

I, Patracia Livhuhani Nevondwe, declare that this dissertation is my own work. It is being submitted for the degree of Master of Science in Medicine in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

……. …….

Patracia Livhuhani Nevondwe Date

DEDICATION

In loving memory of my mother

Emily Malixu Nevondwe

1969 - 2016

PRESENTATIONS ARISING FROM THIS STUDY

Patracia Nevondwe, Robyn Kerr, Candice Feben, Heather Seymour , Maria Mudau, Zané Lombard, Careni Spencer, Amanda Krause, Nadia Carstens***. "Investigating the genetic aetiology of three facial dysostoses in South Africa"*** Poster presentation at the Faculty of Health Sciences Research day and postgraduate Expo (University of the Witwatersrand, 06 August 2018)

Patracia Nevondwe, Robyn Kerr, Candice Feben, Heather Seymour , Maria Mudau, Zané Lombard, Careni Spencer, Amanda Krause, Nadia Carstens. ***"Investigating the genetic aetiology of three facial dysostoses in South Africa"*** Oral presentation at the University Of Limpopo Faculty Of Health Sciences Research Conference (University of Limpopo, 12-14 September 2018)

Patracia Nevondwe, Robyn Kerr, Candice Feben, Heather Seymour, Maria Mudau, Zané Lombard, Careni Spencer, Amanda Krause, Nadia Carstens. ***"Investigating the genetic aetiology of three facial dysostoses in South Africa"*** Oral presentation at the 2018 Joint Congress of the South African Bioinformatics Society and South African Genetics society (Free state, 16-18 October 2018, Golden gate).

ABSTRACT

Treacher Collins (TCS), Nager (NS) and Miller syndromes (MS) are genetic developmental disorders that show overlapping clinical features. No molecular testing for these facial dysostoses (FDs) disorders is available in South Africa (SA). This is due to a number of challenges, most notably that we have no published data on the genetics of these disorders in the South African populations. Clinicians therefore depend on family history and clinical features to make a diagnosis, a task complicated by the variable expression and reduced penetrance seen in these conditions. Diagnosis of these disorders is further complicated by features overlapping with condition such as Broncho-oto-renal (BOR) syndrome, Mandibulofacial dysostosis with microcephaly (MFDM) and CHARGE syndrome.

Fifteen South African patients with TCS and NS including their differential diagnoses of BOR- and CHARGE syndrome were recruited from the participating clinics of the Division of Human Genetics, Wits and NHLS. Of the 15 patients recruited, ten were Africans, four were Caucasians and one was of Indian ancestry. The majority (7/15) of the patients had a provisional clinical diagnosis of TCS followed by CHARGE syndrome (5/15). Two patients had a suspected diagnosis of TCS, BOR- or CHE syndrome and a single patient was clinically diagnosed with NS.

Genomic DNA was extracted from whole blood and targeted next-generation sequencing-based (NGS) mutation screening was performed to analyse twelve genes known to cause or interact with causative genes of the disorders under study. Sequencing was performed on an Illumina MiSeq platform. Putative pathogenic variants were identified through a tiered filtering approach. Firstly, variants with a minor allele frequency of more than 0.05 in gnomAD exomes and genomes datasets were excluded. Subsequently, bioinformatics prediction tools and the mode of inheritance were then used to prioritise variants. Lastly, the American College of Medical Genetics (ACMG) guidelines for variant interpretation were used to classify and identify putative disease-causing variants.

Seven putative disease-causing variants were identified in seven of 15 unrelated patients. Putative disease-causing variants were identified within three genes: *CHD7, POLR1D* and *TCOF1* genes. These consisted of five deletions (*CHD7* c.1931delA, *CHD7* c.3309_3310delCA, *TCOF1* c.4369_4373delAAGAA, *TCOF1* c.3708delC and *POLR1D* c.261delA) and two single nucleotide variants (*CHD7* c.232C>T and *CHD7* c.643C>T). Of the seven putative disease-causing variants identified, three (*TCOF1* c.4369_4373delAAGAA, *TCOF1* c.3708delC and *POLR1D* c.261delA) were identified in patients with TCS and four (*CHD7* c.232C>T, *CHD7* c.1931delA, *CHD7* c.3309_3310delCA and *CHD7* c.643C>T) were identified in patients with CHARGE syndrome. Six of these variants (*CHD7* c.232C>T, *CHD7* c.1931delA, *CHD7* c.3309_3310delCA, *CHD7* c.643C>T *TCOF1* c.3708delC and *POLR1D* c.261delA) are not reported in public mutation databases and one (*TCOF1* c.4369_4373delAAGAA) is a common recurring TCS pathogenic mutation. The overall diagnostic yield of this study was 47%. In addition, a variant of unknown significance (VUS) (*TCOF1* c.3183G>C) predicted to affect splicing was also identified in one patient with TCS.

The present study is, to the best of our knowledge, the first study to perform a molecular analysis on TCS, NS, MS, BOR and CHARGE syndrome in South African patients. This study has produced a baseline mutation profile of TCS and CHARGE syndrome in the South African population and has demonstrated that targeted NGS with multigene panel testing is an acceptable diagnostic method which could be successfully implemented for the molecular diagnosis of TCS and CHARGE syndrome in the South African population. The successful implementation of an NGS-based diagnostic approach for TCS and CHARGE syndrome in SA will enable their molecular diagnosis and this will be important for confirmation of a clinical diagnosis. Furthermore, understanding the genetic basis of these conditions in the South African population will thus not only have practical implication for the patients and their families but also address the paucity of data on genetic conditions in SA.

ACKNOWLEDGEMENTS

I would like to thank the following people:

- The participants who have been involved in this study since the beginning, without whom this work would have not been possible I hope that the results of this study will be useful to you
- My supervisors, Dr Careni Spencer, Dr Nadia Carstens and Dr Robyn Kerr. Thank you for giving me an opportunity to work on this interesting project and thank you again for all your support, guidance, encouragement, patience and wisdom over the past two years. I couldn't have asked for better supervisors
- The postgraduate students and staff in the Division of Human Genetics at the NHLS for their help, support and on-going encouragement
- Staff from Diagnostech, Michelle Bronze and Dean Harris, for their emotional and technical support during lab work of this study
- FRASC working group, Maria Mudau, Dr Candice Feben, Prof Zane Lombard, Heather Seymour and Prof Amanda Krause. Thank you for being the most amazing group one can ever work with and thank you for your expertise and experience projected to this project
- My family, thank you for your support, love and encouragement. I know it was not easy for you
- Funders:
    - NHLS Research Trust
    - Faculty Research Committee
    - National Research Foundation, South Africa
- Lastly, I would like to thank my friends for their unwavering support

TABLE OF CONTENTS

LIST OF FIGURES

Referencing

The reference list in this dissertation was generated using used Harvard referencing style incorporated by the Reworks reference managing tool.

LIST OF TABLES

ABBREVIATIONS

%	Percentage
≤	Less than or equal to
≥	Greater than or equal to
µg	Microgram
µl	Microliter
A	Adenine
ACMG	American College of Medical Genetics
AD	Autosomal Dominant
AFDs	Acrofacial dysostoses
AR	Autosomal Recessive
BMP	Bone morphology protein
BOR	Broncho-oto-renal syndrome
BR	Broad range
C	Cytosine
CADD	Combined annotation-dependent depletion
CHD7	Chromodomain-helicase-DNA-binding protein 7
DANN	Deleterious Annotation of genetic variants using Neutral Networks
DHODH	Dihydroorotate Dehydrogenase
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EFTUD2	Elongation factor Tu GTP binding domain containing 2
FATHMM	Functional Analysis through Hidden Markov Models
FDs	Facial Dysostoses
G	Guanine
GERP++	Genomic Evolutionary Rate Profiling
gnomAD	Genome Aggregation Database
HGMD	Human gene mutation database
HS	High sensitivity
HSF	Human splice finder
IGV	Integrative genome viewer
Indel	Insertion and deletion variant

Kb	Kilo base
MAF	Minor allele frequency
MFDs	Mandibulofacial dysostoses
mRNA	Messenger Ribonucleic acid
MS	Miller syndrome
NCCs	Neural Crest Cells
NGS	Next Generation Sequencing
NHLS	National Health Laboratory Services
nM	Nanomolar
NMD	Nonsense mediated decay
NS	Nager syndrome
OMIM	Online Mendelian Inheritance in Man
POADS	Post axial acrofacial dysostoses
POLYPHEN-2	Polymorphism Phenotyping v2
POLR1C	RNA Polymerase I subunit C
POLR1D	RNA Polymerase I subunit D
RNA	Ribonucleic acid
rRNA	Ribosomal Ribonucleic acid
SA	South Africa
SAV	Sequencing alignment viewer
SF3B4	Splicing Factor 3B unit 4
SIFT	Sorting Intolerant from Tolerant
SMAD4	Mothers against decapentaplegic homolog 4
SMAD5	Mothers against decapentaplegic homolog 5
SMAD6	Mothers against decapentaplegic homolog 6
SMAD7	Mothers against decapentaplegic homolog 7
T	Thymine
TCOF1	Treacle Ribosome Biogenesis factor 1
TCS	Treacher Collins Syndrome
tRNA	Transfer Ribonucleic acid tRNA
VEST	Variant Effect Scoring Tool
Wits	The University of the Witwatersrand

1 INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Rare diseases are those only affecting a small portion of the population and the majority of them have a genetic basis (Field *et al*., 2010). The frequency that defines a disease as rare differs amongst publications; however, in general, diseases are considered rare if they affect fewer than five in 10 000 people (Richter *et al*., 2015). To date, more than 6000 rare diseases are recorded in the Orphanet database (Weinreich *et al*., 2008) and the facial dysostoses (FDs) form part of this group. The FDs are a group of clinically and genetically heterogeneous developmental disorders, caused by the abnormal development of the first and the second branchial arches and their derivatives during early embryonic life (Noden and Trainor, 2005).

During the early stages of embryonic development, flattened neural epithelium forms the neural plate, a structure which ultimately forms the neural tube. The neural tube is a hollow structure which gives rise to the brain and the spinal cord. The dorsal region of the neural tube is covered by the ectoderm cell layer which ultimately gives rise to the neural crest cells (NCCs) (Rubenstein *et al*., 1998). The NCCs are migratory multipotent cells that give rise to a range of diverse derivatives. Depending on their destination, NCCs are classified into four categories and these include the cranial, trunk, vagal and the cardiac NCCs (Bronner-Fraser, 1989). Cranial neural crest cells migrate from the dorsal region of the neural tube to form and populate the first three of five human branchial arches and their derivatives. The first branchial arch gives rise to the middle ear (incus and malleus), maxillary, zygomatic, mandibular and temporal bones. The second branchial arch gives rise to the lesser horn of hyoid, styloid process and the middle ear ossicle, while the third branchial arch give rise to the greater horn of hyoid (Grabb, 1965). The derivatives of the first two branchial arches form the blueprint of the human craniofacial skeleton and impairment in their development leads to the manifestation of FDs (Wieczorek, 2013).

The FDs are broadly classified into two categories: the mandibulofacial dysostoses (MFDs) and the acrofacial dysostoses (AFDs). The MFDs are characterized by craniofacial malformations including down-slanted palpebral fissures, eye lid colobomas, zygomatic hypoplasia, micrognathia and microtia. The AFDs are characterised by craniofacial malformations similar to

those seen in MFDs with the addition of limb defects. Depending on the limb defects involved, AFDs are further sub-classified into the pre-axial, post-axial and other groups. The pre-axial AFDs exhibit thumb anomalies, while post-axial AFDs exhibit anomalies of the fifth and/ fourth ray of both the upper and lower limbs. The third group comprises of those not fitting into the first two groups (Wieczorek, 2013).While there are at least eight different MFDs and eighteen different AFDs reported in literature (Wieczorek, 2013), the present study focused on three most common and well understood FDs, namely, one MFD (Treacher Collins syndrome (TCS)) and two AFDs (Miller syndrome (MS) and Nager syndromes (NS)).

Treacher Collins syndrome is the most prevalent facial dysostosis estimated to occur in 1 in 10 000 – 50 000 live births (Wieczorek, 2013). Nager syndrome has a prevalence rate of 3 in 1 000 000 and MS is even rarer, occurring in 1 per 1 000 000 individuals in Caucasian populations (Fazen *et al*., 1967; Halonen *et al*, 2006).

There is no curative treatment for FDs. There are, however, management strategies to improve the patient's quality of life and ease the burden on their families (Trainor and Andrews, 2013). Despite this, a molecular confirmation of diagnosis can still have a positive effect on a patient's life as research shows that individuals who undergo genomic sequencing for rare disorders report interest in receiving results for reasons beyond clinical utility. These include well-informed reproductive decision-making, enhancing self-knowledge as well as identifying useful support and advocacy groups (ACMG board of directors, 2015; Kohler *et al*., 2017). Globally, a molecular diagnosis of TCS, NS and MS is obtained using different techniques. These include single gene analyses, chromosomal microarray analysis, multiplex ligation-dependent probe amplification (MLPA) and next generation sequencing technologies (Katsanis and Jabs, 2004).
Next-generation sequencing (NGS) technologies, first launched in 2005, are high throughput sequencing technologies which have revolutionised the study of genomics and molecular biology (Besser *et al*., 2017). The ability of these technologies to enable rapid DNA sequencing makes them faster and more cost-effective compared to the traditional Sanger sequencing method (Behjati and Tarpey, 2013). Since their emergence, NGS technologies have been primarily used

in the research setting up until 2011, when they started being rapidly incorporated into routine clinical laboratory testing, mainly in the USA and Western Europe. NGS technologies are now used for improving diagnosis, directing management strategies and enabling early thoughts on the development of therapeutic measures (Mardis, 2011). However, the pace of NGS technologies implementation in developing and under-resourced countries, such as SA, is not equally met (Bahasi and Stambrook, 2014). This is due to the high costs of the laboratory tests, as well as a requirement for technical and bioinformatics expertise associated with this technique.

Currently, no molecular genetic testing for TCS, NS or MS is available in SA. This is due to several challenges, most notably the fact that we have no published data on the genetic aetiology of these disorders in South African populations. As a result, clinicians depend solely on patient family history and clinical phenotypes to make a diagnosis. This can be challenging when phenotypes are overlapping, leading to misdiagnoses. Furthermore, because of the variable expressivity displayed by these disorders, very mildly affected individuals may go undiagnosed. The present study sought to use a targeted NGS approach to generate a mutation profile for TCS, NS and MS in South African patients. The availability of genetic data on these disorders in South African populations would inform development and implementation of appropriate cost-effective NGS-based diagnostic approaches for TCS, MS and NS in South Africa. Literature supporting and informing this study is discussed in the next section.

1.2 Treacher Collins syndrome

Treacher Collins syndrome (OMIM 154500) is an autosomal dominant (AD) mandibulofacial dysostosis estimated to occur in 1 in 10 000 to 50 000 live births. TCS was first described by Thomson, Toynbee and Berry independently in the 1840s (Berry, 1889; Thomson, 1847; Toynbee, 1847). However, it was only in 1900 that Edward Treacher Collins, an ophthalmologist, described the critical components of the syndrome (Treacher, 1900). Following his work, Franceschetti and Klein comprehensively reviewed the syndrome and because of its clinical presentation, they coined the descriptive term 'mandibulofacial dysostosis' (Franceschetti and Klein, 1949).

1.2.1 Clinical features of TCS

Treacher Collins syndrome exhibits variable expressivity and, in some cases, reduced penetrance (Wieczorek, 2013). Treacher Collins syndrome has a wide spectrum of craniofacial features which are usually bilateral and symmetrical in nature. These include down slanting palpebral fissures, underdevelopment of the zygomatic complex, malar hypoplasia and micrognathia, complete or partial absence of the lower eyelashes, lower eyelid coloboma, as well as microtia and hair displacement (Posnick and Ruiz, 2000; Trainor *et al*., 2013).

Other less common clinical features observed in TCS include cleft palate with or without cleft lip, choanae atresia or stenosis. Several complications have been described secondary to the craniofacial features such as upper airway obstruction, feeding difficulties, conductive hearing loss and visual abnormalities (Wieczorek, 2013). According to Katsasis and Jabs (2004), TCS genetic testing should be considered when a patient exhibits at least two major or three minor clinical features. The major and minor clinical features of TCS are listed in Table 1.1 below.

Table 1.1: Major and minor clinical features of TCS (Adapted from Katsanis and Jabs, 2004)

Major	Minor

• Hypoplasia of the zygomatic bones and mandible resulting in:
  ○ midface hypoplasia
  ○ micrognathia
  ○ retrognathia
• External ear abnormalities including:
  ○ absent
  ○ small
  ○ malformed ears
  ○ rotated ears
• Lower eyelid abnormalities including:
  ○ coloboma
  ○ sparse eyelashes
• autosomal dominant inheritance family history
	• External ear abnormalities including:
	  ○ atresia
	  ○ stenosis
	  ○ conductive hearing loss
	• Cleft lip and/or palate
	• Pre-auricular hair displacement
	• Uni or bilateral choanae stenosis or atresia
	• Delayed motor or speech development
	• Ophthalmologic defects including:
	  ○ vision loss
	  ○ amblyopia
	  ○ refractive errors
	  ○ anisometropia
	  ○ strabismus

1.2.2 The genetics of TCS

In 1996, the Treacher Collins Syndrome Collaborative Group identified mutations in the *TCOF1* gene as a cause of TCS (Treacher Collins Syndrome Collaboration Group, 1996). Mutations in two additional genes: *POLR1C* and *POLR1D* were later found to account for a proportion of *TCOF1* negative TCS cases (Dauwerse *et al*., 2011). The *TCOF1* gene is the most common mutated gene in TCS, with variations in this gene accounting for more than 90% of TCS cases (Trainor *et al.*, 2009; Weiner *et al*., 2012). Mutations in the *POLR1C* and *POLR1D* genes are together associated with less than 8% of TCS cases (Dauwerse *et al*., 2011). Nonetheless, there is still a subset of people clinically diagnosed with TCS in whom no causative mutation has been identified (Horiuchi *et al*., 2005; Masotti *et al*., 2009; Splendore *et al*., 2000; Teber *et al*., 2004). It is estimated that approximately 60% of individuals genetically diagnosed with TCS have *de novo* mutations and 40% will have a positive family history (Splendore *et al*., 2002). The former can create an additional complication in providing genetic counselling, as although penetrance of TCS genes mutations is reported to be high, few cases of reduced or incomplete penetrance have been reported (Dauwerse *et al*., 2011; Dixon *et al*., 2004; Katsanis *et al*., 2003; Marres *et al*.,

2002) and there is significant individual phenotypic variability ranging from very mild to severely affected (Martelli-Junior *et al*., 2009). Mildly affected individuals might only be diagnosed retrospectively after the birth of a severely affected family member. In contrast, severe cases can lead to perinatal death (Jones *et al*., 1999). While most cases of TCS are inherited in an autosomal dominant fashion, 1% will be autosomal recessive due to mutations in either *POLR1C* or *POLR1D* genes (Dauwerse *et al*., 2011; Schaefer *et al*., 2014).

According to the Human Gene Mutation Database (HGMD®) (Krawczak and Cooper, 1998), approximately 335 different mutations have been associated with TCS, and the majority of them are small frameshift deletions of up to 40 nucleotides. Literature reports that the majority of TCS mutations are private (family or individual specific) with the exception of a commonly reported 5-bp deletion (c.4366_4370delGAAAA) mutation in exon 24 of the *TCOF1* gene, found to occur in approximately 17% of affected families reported in literature (Edwards *et al*., 1997; Su *et al*., 2006; Trainor *et al*., 2009). Although TCS exhibits a variable phenotype, there is no distinct association between a specific mutation and the resulting phenotype (Edwards *et al*., 1997). It is important to note that most of the genetic studies on rare disorders are carried out in Caucasian populations and genetic studies in individuals of African ancestry are rare. To the best of our knowledge, no study analysing the genetics of TCS and those disorders which can be considered under its differential diagnoses, has been carried out on African populations. As a result, no mutation data is available on these populations.

1.2.2.1 The *TCOF1* gene

The *TCOF1* gene is localised to the long arm of human chromosome 5q32-q33.1 and consists of 26 exons (Figure 1.1). Previous studies have identified several mutational hotspot regions within the *TCOF1* gene and over 50% of *TCOF1* mutations are reported to occur in exon 10, 15, 16, 23, 24 or 25 (Splendore *et al*., 2002). Figure 1.1 below shows a schematic representation of the *TCOF1* gene.

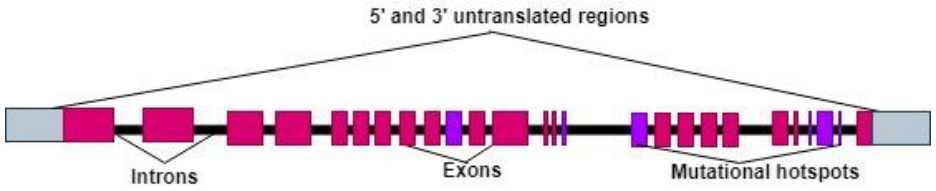

Figure 1.1. A schematic representation of the *TCOF1* gene
The *TCOF1* gene comprises 26 exons (Transcript: TCOF1-001 ENST00000323668.7, Ensemble, release 75). Exons 10, 15, 16, 23, 24 and 25 are hotspot loci for mutagenesis, with over 50% of the *TCOF1* mutations occurring in these regions. Mutational hotspots are in purple.

The *TCOF1* gene encodes a 1489 amino acid protein called treacle (Treacher Collins Collaboration Group, 1996). Treacle protein is a low complexity, serine/alanine-rich, nucleolar phosphoprotein involved in pre-ribosomal processing and ribosomal biogenesis (Teber *et al*., 2004).

1.2.2.2 The *POLR1C* and *POLR1D* genes

The *POLR1C* and *POLR1D* genes are localised to the short arm of chromosome 6p21.1 and the long arm of chromosome 13q12.213, respectively. Figure 1.2 below shows a schematic representation of the *POLR1C* and *POLR1D* genes.

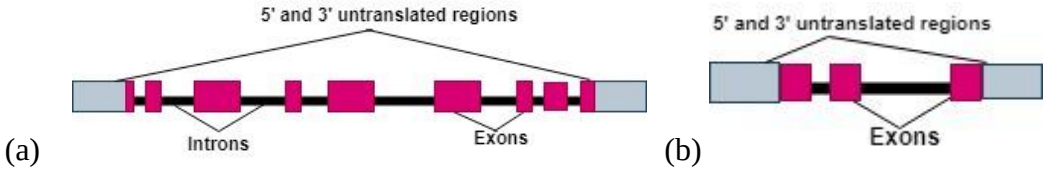

Figure 1.2: Schematic representation of *POLR1C* and *POLR1D* genes
(a) A representation of the *POLR1C* gene consisting of nine exons (Transcript: POLR1D-004 ENST00000399697, Ensemble, release 75) (b) A representation of the *POLR1D* gene consisting of three exons (Transcript: POLR1D-004 ENST00000399697, Ensemble, release 75). No mutational hotspots have been recorded in these two genes.

The *POLR1C* and *POLR1D* genes encode for the AC40 and AC19 proteins, respectively. Both the AC40 and AC19 proteins are located in the α-subunits of the RNA polymerase I and RNA polymerase III enzymes involved in rRNA transcription (Laferté *et al*., 2006). The RNA polymerase I enzyme transcribes the rRNA necessary for the formation of the structural and catalytic components of the ribosomes including the 5S, 5.8S and the 28S rRNAs. RNA polymerase III transcribes the 5S rRNA, tRNA and some non-coding RNAs, important for stabilising the structure of the ribosome (Lafontaine and Tollervey, 2001; Szymanski *et al*., 2002). Thus, the majority of ribosomal biogenesis transcription is carried out by RNA polymerase I and RNA polymerase III enzymes (Paula and White, 2000), and disruptions in ribosome biogenesis can result in early embryonic development disorders (Yelick and Trainor, 2015). The discovery of mutations in the *TCOF1*, *POLR1C* and *POLR1D* as a cause of TCS led to the establishment of functional studies elucidating the pathogenesis of TCS.

1.2.3 The pathogenesis of TCS

Mouse- model studies have revealed that TSC is caused by premature reduction of neural crest cell precursors through a mechanism involving augmented cell death and reduced cell division (Dixon *et al*., 2006; Jones *et al*., 2008). The *TCOF1*, *POLR1C* and *POLR1D* gene products are involved in rRNA transcription and pre-processing of the rRNA transcript (Dauwerse *et al*., 2011; Valdez *et al*., 2004; Yao *et al*., 1997). Mutations in these genes disturb ribosomal RNA transcription, which in turn, compromise the production of mature ribosomes in the pre-migratory NCCs (neuroepithelial cells) and NCCs (Dixon *et al*., 2006). Reduced ribosomal biogenesis is believed to activate the p53 protein (a tumour suppressor and a cell cycle regulator). An activated p53 leads to neuroepithelial apoptosis and the arrest of the G1 phase of NCCs cell cycle in the early stages of embryogenesis. The former can lead to diminished NCCs moving to the cranial region to form the craniofacial skeleton. Arrest of the G1 cycle reduces the proliferation capacity of the migrating NCCs and neuroepithelial cells leading to an inadequate number of NCCs populating both the first and second branchial arches and their derivatives. This results in the manifestation of characteristic craniofacial features of TCS. Although the pathogenesis of TCS has been elucidated, it is still not yet understood why TCS present with a variable phenotype. Future studies elucidating this variable expression will be required. Interestingly, knowledge on

the pathogenesis of TCS has prompted researchers to investigate and comment on possible therapeutic measures of the disorder.

1.2.4 Proposed therapeutic measures of TCS

The pathogenesis of TCS is mainly due to excessive cell death in pre-migratory NCCs, hence the principle behind the proposed therapeutic measures is to suppress apoptosis and ensure a sufficient population of migratory NCCs required for facial skeletal formation. Mouse model studies have revealed that p53 inhibition in the early stage of development rescues the mouse-TCS phenotype (Jones *et al*., 2008). As a result, the investigators have proposed exploring human p53 inhibition in the early stage of development as a potential therapeutic measure to rescue the human TCS phenotype. However, p53 is also important as a tumor suppressor and its subsequent association with cancer and tumorigenesis would likely mean that alteration of its function can be associated with other problems. Trainor *et al*. (2009) proposed that stem cells might play a role in improving surgical outcomes. As a result, it is proposed that stem cells could be introduced in TCS causing tissues, such as cartilage and bone *in utero*.

1.3 Miller syndrome

Miller syndrome (MS) (OMIM 263750) is an autosomal recessive (AR) post-axial acrofacial dysostosis. MS was first described in 1969 by Genée and his report assumed the condition to be an extreme form of TCS (Genée–Weidemann, 1969), but it was later identified as a separate condition (Weidemann, 1973). Miller *et al.* (1979) reported several more cases with similar post-axial clinical defects describing them as 'post-axial acrofacial dysostosis' (Miller *et al*., 1979). The disorder is currently referred to as Miller syndrome, Genée–Weidemann syndrome, Wildervanck-Smit syndrome or POADS. This is a rare condition, estimated to affect 1 in 1 000 000 individuals but despite this, it is the most common and best understood post-axial acrofacial dysostosis (Wieczorek, 2013). As with TCS, MS presents with a wide spectrum of variable clinical features.

1.3.1 Clinical features of MS

Miller syndrome affects the development of the craniofacial skeleton and post-axial limbs. The craniofacial anomalies of MS overlap with those of TCS, but in addition post-axial limb defects occur (Miller *et al*., 1979; Ng *et al*., 2010). Limb anomalies include hypoplastic or complete absence of the fifth ray affecting the upper and lower limbs, with or without the underdevelopment of the ulna and the radius (Wieczorek, 2013). Abnormalities of other digits are also reported and may include absent fourth digits, various degrees of syndactyly and/or clenched hypoplastic thumbs (Donnai *et al*., 1987). Table 1.2 below shows the most common clinical features of MS. The list was compiled from Fang *et al*., 2012 and Rainger *et al*., 2012.

Table 1.2: Common clinical features of MS

Region	Features	Frequency
Craniofacial	Micrognathia	100%
	Malar hypoplasia	100%
	Cleft lip and/or palate	80%
	Low-set, dysmorphic pinnae	75%
	Cupped and low-set ears	75%
	Coloboma of the eyelids	43%
	Accessory nipples	40%
Limbs	Absence of the fifth (and/or fourth) ray of limbs	80%
	Hypoplasia of the Ulna bone	60%

1.3.2 The genetics of MS

Mutations in the *DHODH* gene were first identified as a cause of MS in 2010 (Ng *et al*., 2010). Mutations in the *DHODH* are accountable for approximately 69% - 75% cases of human MS (Wieczorek, 2013). According to the HGMD, approximately 16 different *DHODH* mutations have been associated with MS and the majority of these are compound heterozygous missense mutations.

1.3.2.1 The *DHODH* gene

The *DHODH* gene is located on the long arm of chromosome 16q22.2 and contains nine exons. It encodes a mitochondrial dihydroorotate dehydrogenase enzyme involved in mammalian *de novo* pyrimidine synthesis (Evans and Guy, 2004; Löffler *et al*., 2005). Figure 1.3 below shows a schematic representation of the *DHODH* gene.

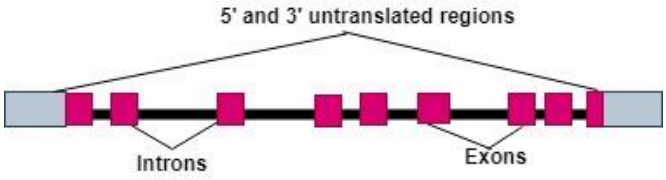

Figure 1.3: A schematic representation of the *DHODH* gene
The *DHODH* gene contains nine exons (Transcript: DHODH-201 ENST00000219240.4, Ensemble, release 75).

1.3.3 The pathogenesis of MS

The mitochondrial dihydroorotate dehydrogenase protein encoded by the *DHODH* gene catalyses the rate limiting step of mammalin *de novo* pyrimidine synthesis (Evans and Guy, 2004; Löffler *et al*., 2005). Mammalian *de novo* pyrimidine synthesis is essential and increased in proliferating cells to meet an increased demand for nucleic acid precursors and other cellular components during early development (Duley *et al*., 2016; Loffler *et al*., 1998; Rawls *et al*., 2000). Mutations in the *DHODH* are believed to impair the function of the dihydroorotate dehydrogenase protein, which in turn, may compromise *de novo* production of pyrimidines in proliferating cells, such as the NCCs (Fang *et al*., 2012). The molecular mechanism in which the impaired *de novo* biosynthetic pathway results in the characteristic phenotype of MS is not yet clear (Fang *et al*., 2012; Duley *et al*, 2016). Duley *et al*, 2016 suggest that reduced functionality of the *de novo* pyrimidine biosynthetic pathway during early embryogenesis might not be the only mechanism involved in the manifestation of MS phenotype. They hypothesise that other possible mechanisms, such as impaired cell signalling, perturbation of pyrimidine nucleotides synthesising RNA or mitochondrial dysfunction leading to localised tissue energy deficits during embryological development, could also underlie the manifestation of MS phenotype (Duley *et al*., 2016). Future studies investigating the relationship between *de novo* pyrimidine biosynthetic pathway and the mechanisms suggested by Dulley and colleagues will be required to understand the pathogenesis of MS. Although the molecular mechanism in which impaired *de novo* biosynthetic pathway results in the characteristic phenotype of MS is not yet clear, a few studies investigating possible theurapeutic measures of MS have been conducted (Duley *et al*., 2016; Fukushima *et al*., 2009).

1.3.4 Proposed theraupetic measures

Mouse studies have demonstrated that a maternal high-uridine diet during pregnancy reverses an induced mouse-like MS phenotype (Fukushima *et al*., 2009). As a result, it is proposed that maternal consumption of high-uridine foods such as tomatoes, mushrooms, broccoli, pork and beef liver during pregnancy could be explored to try and rescue the MS phenotype in humans (Duley *et al*., 2016). However, this could work best in individuals with a family history of MS or in cases in which a pre-implantation and/or early prenatal diagnosis is established.

1.4 Nager syndrome

Nager syndrome (NS) (OMIM 154400) is a rare pre-axial acrofacial dysostosis disorder. NS was first described by Slingenberg in 1908, and later recognized as a pre-axial acrofacial dysostosis by Nager and Reynier in 1948 (Slingenberg, 1908; Nager and Reynier, 1948). The inheritance pattern of NS remains unclear as both AD and AR patterns of inheritance have been observed, suggesting genetic heterogeneity (Kavadia, 1993; Lowry, 1997; McDonald and Goski, 1993). Nager syndrome is estimated to 3 in 1 000 000 individuals in Finland (Halonen *et al*., 2006)

1.4.1 Clinical features of NS

The clinical features on NS consist of craniofacial and limb abnormalities (Nager and Reynier, 1948) The craniofacial features described in NS are similar to those of TCS and MS, but in addition pre-axial limb defects occur (Wieczorek, 2013). Nager syndrome limb abnormalities involve the radial elements of the upper limbs and may present as hypoplasia or total absence of the thumb, triphalangeal thumbs, radial hypoplasia or aplasia, and radioulnar-synostosis (Le *et al*., 1989). The presence of pre-axial upper-limb defects as opposed to post-axial upper-limb differentiates NS from MS (Trainor and Andrews, 2013). Table 1.3 below summarises the most common clinical features of NS. The list was compiled from Fang *et al*., 2012 and Rainger *et al*., 2012.

Table 1.3: A list of common clinical features of NS

Region	Clinical features	Frequency
Craniofacial	Retrognathia and/or micrognathia	96%
	Downward slanted palpebral fissures	95%
	Conductive hearing loss	95%
	Dysplastic ears	90%
	Malar hypoplasia	83%
	Cleft palate deformities	80%
	Absent medial lower lid eyelashes	43%
Upper limbs	Hypoplasia or complete absence of the thumb	59%
	Radio-ulnar synostosis	50%
	Radial bone hypoplasia	43%
Others	Delayed development	50%

1.4.2 The Genetics of NS

Mutations in the splicing factor3b, subunit4 (*SF3B4*) were first identified as a cause of NS in 2012 (Bernier *et al*., 2012). This was later confirmed by Czeschik and Petit, respectively (Czeschik *et al*., 2013; Petit *et al*., 2014). Further support of *SF3B4*-association with NS comes from Waggoner *et al*, who described a child presenting with a NS phenotype with a deletion of the chromosome 1q12-q2 region, which contains the *SF3B4* gene (Waggoner, 1999). Mutations in the *SF3B4* gene are associated with approximately 67% of NS (Bernier *et al*., 2012; Petit *et al*., 2014). According to the HGMD, over 30 different *SF3B4* mutations have been identified in unrelated patients with NS and the majority of these mutations are point mutations. Large deletions have only been reported in two NS (Lund *et al*., 2016 and Waggoner *et al*., 1999). While most people with NS represent a *de novo* change, few familial cases have been reported (Cassina *et al*., 2016; Petit *et al*., 2014). In these instances, AD (Aylsworth *et al*., 1991; Hall, 1989) and AR (Chemke *et al*., 1988; Kennedy and Teebi, 2004) patterns of inheritance have been described, suggesting genetic heterogeneity.

1.4.2.1 The *SF3B4* gene

The *SF3B4* gene is localised to the long arm of chromosome 1q21.2. The *SF3B4* gene consists of six exons. The *SF3B4* gene encodes a protein called spliceosome-associated protein 49 (SAP49). The SAP49 protein is involved in spliceosomal assembly and is further linked to the bone morphology pathway (Bernier *et al*., 2012 and Watanabe *et al*., 2007). Figure 1.4 below shows a schematic representation of the *SF3B4* gene.

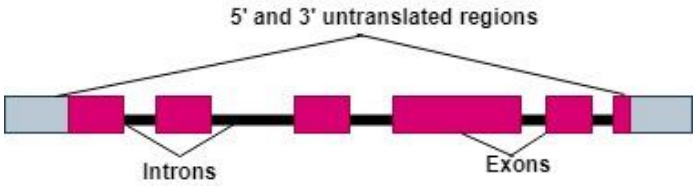

Figure 1.4: A schematic representation of the *SF3B4* gene
The *SF3B4* gene consists of six exons (Transcript: SF3B4-001 ENST00000271628.8, Ensemble, release 75).

1.4.3 Pathogenesis of NS

The exact mechanism in which mutations in the *SF3B4* gene results in the manifestation of NS phenotype is not currently understood (Petit *et al*., 2014). Bernier *et al*. (2012) hypothesised that the SAP49 protein could be involved in the splicing of genes involved in limb and craniofacial development. One obstacle hindering investigations into the cellular pathogenesis of NS was a lack of an animal model which could be used in animal studies. Fortunately, Devotta and colleagues recently identified established Xenopus laevis as a suitable animal to use in investigating the pathogenesis of human NS and research elucidating the pathogenesis of this disorder is expected to be undertaken in the future (Devotta *et al*, 2016). No therapeutic measures of NS have been proposed yet and this is likely a result of its unknown pathogenesis.

1.5 Differential diagnosis

Treacher Collins syndrome, MS and NS share overlapping clinical features with various disorders. The conditions considered as differential diagnoses in this study include Broncho-oto-renal syndrome (BOR), CHARGE syndrome and Mandibulofacial dysostosis with microcephaly (MFDM). CHARGE is an acronym for Coloboma, Heart defects, Atresia of choanal, Retardation of growth and development, and Ear abnormalities (Pagon *et al*., 1981). CHARGE syndrome is an AD condition estimated to affect 1 in 8 500-10 000 new-borns (Klingenberg and Andersen, 2008). Most cases are sporadic due to *de novo* mutations in the *CHD7* gene (Zentner *et al*., 2010). Individuals with CHARGE syndrome exhibit a broad phenotypic spectrum including choanal atresia, hypoplastic semi-circularcanals, coloboma of the eye, cranial nerve and external ear anomalies (Janssen *et al*., 2012). The overlapping clinical features shared with the FDs include palatal abnormalities, choanal atresia, ear anomalies and hearing loss

Broncho-oto-renal syndrome is an AD disorder estimated to affect 1 in 40 000 people (Fraser *et al*., 1980). The disorder is caused by mutations within the *EYA, SIX1* and *SIX5* genes, with mutations in the *EYA1* identified n 40% of the patients. Broncho-oto-renal syndrome is characterised by anomalies of the second branchial arch, preauricular pits and/or tags, hearing loss, and renal abnormalities (Chen *et al*., 1995). While renal anomalies are less commonly seen

in patients with TCS, MS or NS, pre-auricular tags, palate and second branchial arch anomalies and hearing loss are frequently reported in individuals with TCS, MS or NS. Mandibulofacial dysostosis with microcephaly is an AD congenital disorder (Wieczorek, 2013). The incidence of MFDM is unknown. The disorder is characterised by craniofacial anomalies, anomalies of the ear and hearing, intellectual disability, microcephaly and short stature (Guion-Almeida *et al*., 2009). This syndrome is caused by mutations within the *EFTUD2* gene (Luquetti *et al*., 2013). There is he clear phenotypic overlap between these disorders, it is therefore believed that testing for BOR, CHARGE syndrome and MFDM in this study is justified, particularly given the benefit of a potential conclusive diagnosis for the management of the patient and genetic counselling of the family.

1.6 Management of TCS, MS and NS

Management of TCS, MS and NS necessitates a multidisciplinary approach which brings together expertise from craniofacial surgeons, plastic surgeons, ear nose and throat specialists, audiologists, dieticians, genetic counsellors/clinicians and occupational therapists. The exact timeline of these procedures is complex and depends on the severity of the phenotype (Thompson *et al*., 2009). Management of these FDs requires multiple hospital visits and surgeries. This places a heavy financial burden on the patient's family members and the healthcare system. However, despite multiple surgeries that these patients normally undergo, they are rarely fully corrective (Trainor *et al*., 2009). Patients and their family members should be offered genetic counselling to provide information on the genetics as well as the clinical picture, its management and prognosis (Marsella *et al*., 2011).

1.7 Mutation detection strategies

In previous years, mutation screening of TCS, MS and NS mutation was performed using direct sequencing of the candidate genes in combination with methods analysing deletions and duplications. These include chromosomal microarray, quantitative PCR and multiplex ligation-dependent probe amplification analysis (reviewed in Katsanis and Jabs, 2004; Trainor *et al*., 2013). However, these methods have their own limitations, necessitating a new and improved

approach. With advances in technology, we are now able to screen for TCS, MS and NS mutations using NGS.

1.7.1 Next generation sequencing

Twelve years after the discovery of the double helix structure of DNA (Maxam and Gilbert, 1977), the first-generation sequencing methods emerged, and the Sanger sequencing method dominated (Sanger *et al*., 1977). While the Sanger sequencing method has led to major accomplishments including sequencing of the human genome (Lander *et al*., 2001 and Venter *et al*., 2001), the method's limitations presented a need for new and improved technologies to sequence large numbers of human and other genomes. Next generation sequencing is a term used to describe deep, high throughput sequencing technologies, which first emerged in the early 2000s with the primary aim of improving the existing sequencing method (Mardis, 2008). Since their emergence, NGS technologies, combined with Bioinformatics tools, have transformed the scope of genomics and molecular biology by enabling rapid DNA and RNA sequencing, testing multiple genes at once in a relatively short time and at relatively low costs, and thereby yielding more molecular diagnoses (Dark, 2013). While NGS was primarily used in the research environment, it has since been rapidly incorporated into routine clinical diagnostics in international clinical laboratories since 2011 (Dunne *et al*., 2012; Mardis, 2011). However, the pace of NGS technologies implementation in developing and under-resourced countries, such as SA, is not equally met (Bahassi and Stambrook, 2014) and this might be due to the high costs associated with the technology and a lack of knowledge transfer.

In summary, FDs are a group of congenital developmental disorders caused by the abnormal embryonic development of the first and second branchial arches and their derivatives. Accordingly, FDs are divided into two categories: MFDs and AFDs. While there are at least 30 FDs reported in literature (Wieczorek, 2013), the present study focused on the three most common and best understood namely, TCS, MS and NS. TCS is caused by mutations in the *TCOF1, POLR1C* and *POLR1D* genes. Mutations in the *DHODH* gene are known to cause MS and changes in the *SF3B4* gene cause NS. Table 1.4 below summarises information on the three disorders under study.

Table 1.4: Brief review on the three facial dysostoses, TCS, MS and NS

FDs	Disorder	Prevalence	Mode of inheritance	Genes involved	% of mutation accounted for by gene	Mutation profiles	% of mutation accounted for by each profile[1]	Mutation hotspots	Reference
MFDs	TCS	1 in 10 000 - 50 000	AD/AR	*TCOF1* *POLR1C* *POLR1D*	>90% <4% <4%	Point mutations Large indels/duplication	~86% ~14%	*TCOF1* exon 10,15, 16, 23, 24 and 25	Dauwerse *et al*., 2011
AFDs	MS	3 in 1 000 000	AR	*DHODH*	69-75%	Point mutations Large indels/duplication	~70% ~30%	-	Ng *et al.*, 2010
	NS	1 in 1 000 000	AD/AR	*SF3B4*	57-63%	Point mutations Large indels/duplication	~97% ~3%	-	Bernier *et al*., 2012

[1] Human Gene Mutation Database (HGMD®)

Abbreviations: -, none; AD, Autosomal dominant; AFDs, Acrofacial dysostoses; AR, Autosomal recessive; *DHODH*, Dihydroorotate dehydrogenase; FDs, Facial dysostoses; MFDs, Mandibulofacial dysostoses; MS, Miller syndrome; NS, Nager syndrome; *POLR1C*, RNA polymerase I and III subunit C; *POLR1D*, RNA polymerase I and III subunit D; *SF3B4*, Splicing factor 3b subunit 4; TCS, Treacher Collins syndrome; *TCOF1*, Treacle ribosome biogenesis factor 1

1.8 The purpose of the study

1.8.1 Rationale

TCS, NS and MS share overlapping clinical features and display variable expressivity and, in a few cases, reduced penetrance (Dauwerse *et al*., 2011; Wieczorek, 2013). The genetic aetiology of these disorders in South African patients is currently unknown, as there is no published literature on this topic. No molecular diagnostic testing for these conditions is available in South Africa and thus clinicians depend on family history and clinical features to make a diagnosis. The present study sought to generate a mutation profile of TCS, NS and MS in South African patients. The findings of this study can be used to comment on an appropriate genetic testing strategy for the FDs in South Africa. The successful implementation of an NGS-based diagnostic approach for TCS, MS and NS in South Africa will enable molecular diagnosis important for confirmation of a clinical diagnosis. Early genetic confirmation of these conditions will enable an accurate recurrence risk for parents including the potential for prenatal or preimplantation genetic testing. Understanding the genetic basis of these conditions in the South African population will thus not only have practical implication for the patients and their families, but also address the paucity of data on genetic conditions in South Africa.

1.8.2 Aim

The aim of this study was to use targeted NGS to generate a mutation profile for TCS, NS and MS in South African patients.

1.8.3 Objectives

The objectives of this study were:

- To recruit 16 participants diagnosed or presenting with TCS, NS or MS clinical features and their immediate family members, where possible.
- To identify appropriate target genes to include on an NGS panel.
- To extract DNA and prepare sequencing libraries using the Agilent SureSelect ^{QXT} target enrichment system.

- To sequence captured libraries and analyse data.
- To filter and interpret variants using the American College of Medical Genetics Guidelines (ACMG) for variant interpretation (Richards *et al*., 2015).
- To validate any putative pathogenic mutations using Sanger sequencing.

2 MATERIALS AND METHODS

This chapter seeks to give a summary of the research methods employed to achieve the aim and objectives outlined in the previous section

2.1 Study participants

2.1.1 Patient recruitment

Participants were identified at the participating genetic clinics of the Division of Human Genetics, National Health Laboratory Services (NHLS) and The University of the Witwatersrand (Wits). Most participants were referrals from genetic clinics at Charlotte Maxeke Academic Hospital, Rahima Moosa Hospital and the Chris Hani Baragwanath Hospital (all in the greater Johannesburg area) and a few cases were referred from doctors in private practice. Medical records of the identified patients were screened by a qualified Medical Geneticist using the Facial Dysostoses Tick Sheet (See Appendix B1, page 117, developed in-house by Medical Geneticists, Dr Careni Spencer and Dr Candice Feben, to identify suitable participants who were then invited to participate in the study. The study included patients clinically diagnosed and/or suspected to have TCS, NS and MS and undiagnosed individuals with clinical phenotypes forming part of the broader phenotypic spectrum of the three FDs. In addition, persons with a clinical diagnosis and/or suspected Mandibulofacial dysostoses with microcephaly (MFDM) CHARGE and Broncho-oto-renal (BOR) syndromes were also included as this can be considered in the differential diagnosis of the three FDs of interest. Patients with no clinical features of the three FDs under study or their differential diagnoses were excluded. Patients were not selected on the basis of ethnic background. The parent/guardian of any patient recently seen in the genetics clinics, who met the clinical criteria, was approached.

2.1.2 Sample collection

Five to ten millilitres of whole blood were collected from each participant in Ethylenediaminetetraacetic acid (EDTA) tubes. Blood samples were collected by a Medical Geneticist or a qualified phlebotomist.

2.1.3 Informed Consent and Voluntary Participation

The study aim, objectives, risks and benefits were explained to the parents and/or legal guardians of the participants by a qualified Medical Geneticist and/or a Genetic Counsellor from the Division of Human Genetics, Wits and NHLS. Parents and/or legal guardians were given time to decide regarding participation, including time for consultation with family members whenever needed. Written informed consent for genetic testing was obtained from all patients, and specific written informed consent for the publication of photos was also obtained (See Appendix B2, page 119). Patient recruitment was carried out retrospectively and prospectively. Retrospectively, patients who had attended the division's clinic were invited to take part in the study and prospectively, new patients attending our genetic clinic were told about the study. Parents or legal guardians of the deceased retrospective patients gave written informed consent for the use of their deceased children's DNA for genetic testing.

2.1.4 Anonymity and Confidentiality

All blood samples were de-identified and a unique bar code was assigned to each sample. The assigned code only retained relevant information and not the participant's name and/or other identifying details. To maintain confidentiality, participant's identifying and personal information were kept in a Microsoft Excel database, accessed through google drive and only accessible to individuals directly involved in this study.

2.1.5 Ethical clearance

The present study was a sub-study for a broader project titled 'Using Next-Generation Sequencing Technologies to Investigate the Genetic Aetiology of Developmental Disorders in South Africa'. The principal investigator of the project is Dr Nadia Carstens, who also supervised the present study. Ethical clearance for the broader project (M160830) and this sub-study (M170760) was obtained from the University of the Witwatersrand Human Research Ethics Committee (Medical). Ethical clearance certificates are shown in Appendix B3, page 126.

2.2 Methods

2.2.1 DNA extraction

Genomic DNA was extracted from whole blood samples using the modified version of the salting-out method (Miller *et al*., 1988). This method is routinely used for diagnostic and research applications in the Division and has proved to yield adequate DNA quality and quantities for NGS and other sequencing technologies. DNA was extracted as part of a routine diagnostic procedure by a qualified medical scientist from the Division of Human Genetics at the NHLS. Briefly, the red blood cells were lysed and washed with sucrose Triton-X buffer. Through a series of washes and centrifugation steps, the red blood cell debris was removed and the white blood cells pelleted. An overnight incubation at 37°C with proteinase K, sodium dodecyl sulphate and EDTA was used to break down proteins and the lipid bilayer of the white blood cells and the DNA was recovered with salt and ethanol. The DNA was then re-suspended in Tris buffer and kept at 4°C until ready to use.

2.2.2 DNA quantification

DNA was quantified using the NanoDrop®ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The 260/280 ratio of the sample absorbance was used to assess the purity of the DNA sample. A ratio of ~1.8 normally represents pure nucleic acids and a ratio lower than 1.8 indicates the presence of possible contaminants.

2.2.3 Agarose gel electrophoresis

The process of DNA extraction predisposes DNA to mechanical damage. As a result, it is important to assess the integrity of the DNA prior to further analysis or sequencing, as failure to do this, may negatively affect downstream DNA processing. DNA integrity was assessed using agarose gel electrophoresis. Briefly, the procedure involves loading DNA and the loading dye in parallel with a one kilo base (kb) ladder into precast wells on a one percent agarose gel. The gel was then immersed in a one times Tris/Borate/EDTA buffer and run at 120 voltages until separation is achieved. Fragments were then visualised and imaged using an ultraviolet

transilluminator (Omega Fluor™ Gel Documentation System, Vacutec, and Johannesburg, South Africa).

2.2.4 Targeted next-generation sequencing

2.2.4.1 Custom gene panel design

An extensive review of the literature was carried out in order to identify genes associated with the three disorders and those of their differential diagnoses. The present study was part of a broader project titled 'Using Next-Generation Sequencing Technologies to Investigate the Genetic Aetiology of Developmental Disorders in South Africa'. The project had three sub-studies (two MSc's and one PhD), each focusing on a different developmental disorder or group of disorders. The three studies used a single custom designed gene panel to sequence the genes associated with each of the sub-studies. The custom gene panel was designed by Maria Mudau (PhD student and a medical scientist) using the Agilent SureDesign software package (Agilent, 2017). The gene panel included coding regions, intron-exon boundaries, untranslated regions and ten flanking bases of 49 genes known to cause or interact with causative genes across the three sub-studies. Of the 49 genes included in the panel, regions of interest of 12 genes, covering 28.27kb of the human genome, were associated with the present study.

2.2.4.2 DNA quantification

The concentration of the starting material is crucial to the optimal performance of an NGS experiment. Therefore, prior to target library preparation, genomic DNA was quantified using both double-stranded DNA Broad Range (BR) and High Sensitivity (HS) assay kits for the Qubit Fluorometric quantification system following the manufacturer's protocol (Thermofisher Scientific, Johannesburg, South Africa). The HS kit accurately quantifies samples with concentration ranging from ten pg/µl - 100 ng/µl and the BR kit accurately quantifies samples with a concentration of 100 pg/µl - 1000 ng/µl (Thermofisher Scientific, Johannesburg, South Africa). All DNA samples were normalised to a concentration range of 18 ng/µl – 25 ng/µl and two µl of each sample were used to prepare targeted DNA libraries.

2.2.4.3 Target DNA library preparation

Targeted DNA libraries were prepared using the Agilent SureSelect ^{QXT} target enrichment kit. The workflow is summarised in Figure 2.1 below

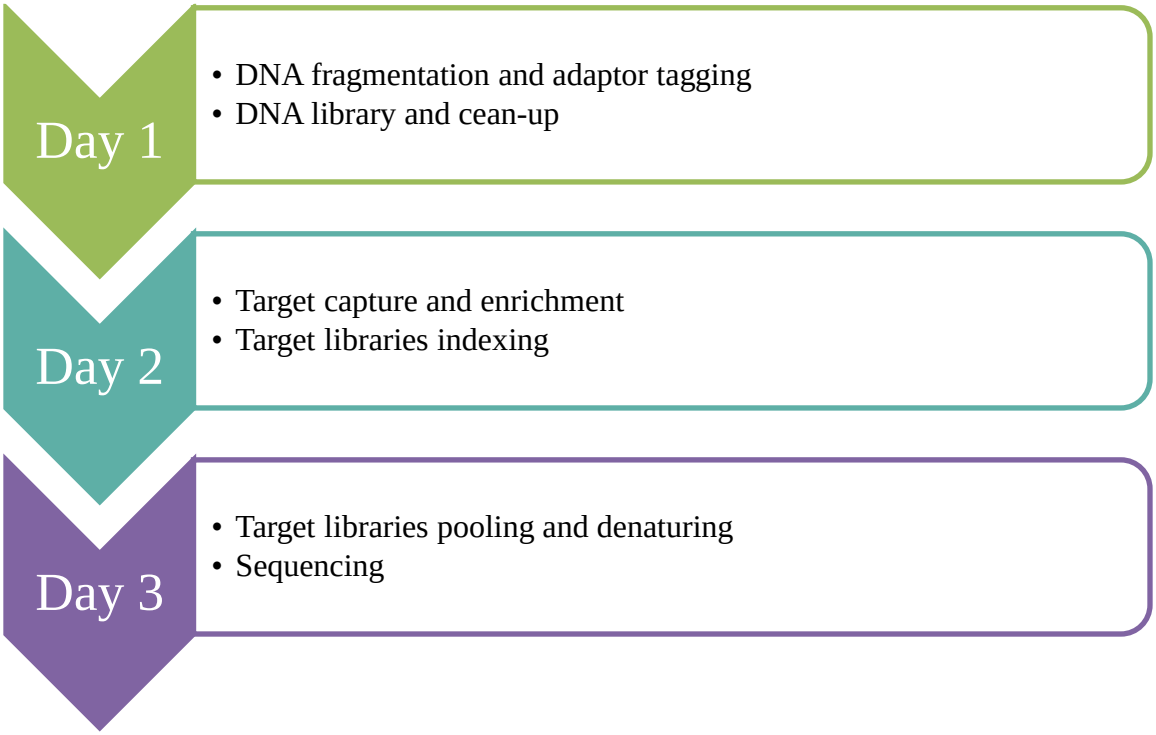

Figure 2.1: A summary of the workflow employed to prepare targeted DNA libraries
The workflow is adapted from the manufacturer's protocol (Agilent, 2016).

DNA libraries were prepared using an enzymatic shearing approach. Each DNA sample was enzymatically sheared using a transposon enzyme and fragments were adaptor tagged and amplified following the Agilent SureSelect QXT target enrichment protocol. Following fragmentation and tagmentation, DNA fragments were then purified and size selected using the SPRIselect® beads (Agencourt, Beckman Coulter, Johannesburg, South Africa). The quality and quantity of each DNA library were evaluated using the Bioanalyzer DNA 2100 HS kit (Agilent, Johannesburg, South Africa) and the Qubit BR kit (Thermofisher Scientific, Johannesburg, South Africa) following the manufacturer's instructions. DNA libraries were hybridised to custom designed capture probes through incubation on a plate shaker for 30 minutes at 1800 revolutions per minute. Targeted regions were then in-solution captured using MyOne Streptavidin T1

magnetic beads (Life Technologies, Johannesburg, South Africa) and non-target regions were washed away. Each enriched library was indexed using dual barcode primes through PCR amplification. Indexing allows sequencing of multiple samples in a single experiment. Indexed libraries were then purified using the SPRIselect® (Agencourt, Beckman Coulter, Johannesburg, South Africa) following the manufacturer's protocol. The purification step was crucial for removing excess PCR reagents that could interfere with downstream sequencing. Finally, indexed enriched target sequencing libraries were quantified using the Qubit HS kit (Thermofisher Scientific, Johannesburg, South Africa) and qualified using the Bioanalyzer DNA 1000 HS kit (Agilent Technologies, Johannesburg, South Africa).

2.2.4.4 Library pooling and denaturing

The addition of index primers during the library preparation process allows multiple samples to be sequenced together in a single experiment. To ensure equal coverage among samples, each DNA library quantified and size analysed using the Qubit HS kit (Thermofisher Scientific, Johannesburg, South Africa) and the concentration in ng/µl was converted to Nanomolar using the formula below (Figure 2.2). Each library in Nanomolar (nM) was then diluted to 4nM with nuclease-free water and five µl of each diluted library was pooled in a single tube.

$$\frac{\text{Concentration in ng/ul} \times 10^{6}}{\text{dsDNA Molecular weight (654.6 g/mol)} \times \text{insert size (base pairs)}} = \text{Molar concentration (nM)}$$

Figure 2.2: A calculation formula used to calculate molar concentration of each prepared targeted DNA library

The pooled library was denatured into single-stranded DNA using sodium hydroxide and diluted to eight Picomoles using the HT1 buffer (Illumina, Johannesburg, South Africa) and an internal

control for Illumina experiments, PhiX control version three (Illumina, Johannesburg, South Africa), was added at a concentration of five percent.

2.2.4.5 Sequencing

The pooled library combined with five percent PhiX control was sequenced on an Illumina MiSeq platform with 150bp paired-end sequencing using the MiSeq reagent kit version two Nano (300 cycles) and Micro sequencing kits (300 cycles) following the manufacturer's protocol (Illumina, Johannesburg, South Africa). Sequencing in this study was performed in three separate runs.

2.3 Data analysis

The workflow employed and tools used during data analysis are summarised in Figure 2.3 below.

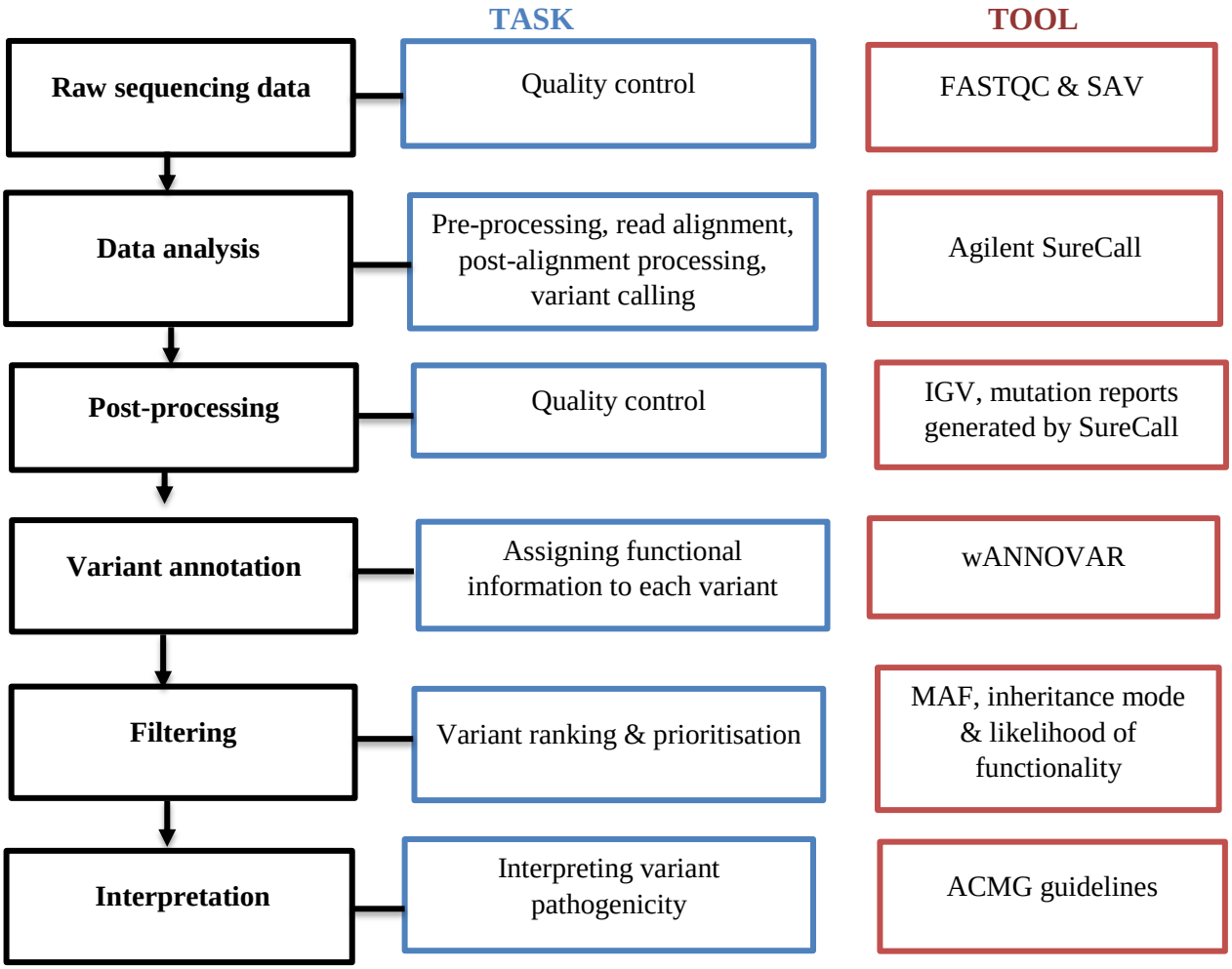

Figure 2.3: A summary of steps taken and tools used to analyse data
Abbreviations: SAV, sequencing analysis viewer; IGV, integrative genomic viewer; MAF, minor allele frequency; ACMG, American College of Medical Genetics.

2.3.1 Sequencing data quality control (QC)

The quality of the sequencing data was assessed using the Illumina Sequencing Analysis Viewer (SAV) and FASTQC files. The SAV was used to visualise important quality metrics of the overall experiment such as the cluster density, experiment output and the quality score (Q-score) distribution and the FASTQC files were used to assess the performance of the experiment as well as the quality metrics of each sample.

2.3.2 Read alignment and variant calling

Following sequencing quality control (QC) evaluation, low-quality reads and sequencing adaptors were removed using the Agilent SureCall software under default parameters. Reads were then aligned to the human genome reference (GRCh37/hg19) using the Burrows-Wheeler Aligner-MEM version 0.5.9 (Li and Durbin, 2009) under default parameters. Post alignment, a QC step which included removal of duplicate reads, indel local realignment and base quality score recalibration was performed using the SureCall software on default parameters.

2.3.3 Post-processing quality check

A post-processing data quality check was then performed to ensure that the target regions were covered to an adequate depth for all samples and to identify regions of interest that were poorly covered. Post-processing quality control check in this study focused on read alignment and mutation reports generated by the SureCall software.

2.3.4 Variant annotation

The latest version of wANNOVAR (Wang *et al*., 2010) was used to annotate variants identified in each sample.

2.3.5 Variant filtering and prioritisation

Following variant annotation, a tiered variant filtering approach was used to identify putative disease-causing variants. Firstly, variants with a minor allele frequency (MAF) of more than five percent in the Exomes- and Genomes All datasets of the genome Aggregation Database (gnomAD) were filtered out. Then, variants were prioritised based on their inheritance modes. Lastly, various bioinformatics prediction tools, which predict the effect of variants on protein sequence and/or structure and conservation, were then used to prioritise variants. A list of the bioinformatics prediction tools used in this study can be seen in Table 2.1 below. The pathogenicity of each variant was then evaluated using known information about the variant from the literature, wANNOVAR (Wang *et al*., 2010) and Varsome (Kopanos *et al*., 2018)

software. Based on this knowledge, the ACMG guidelines for sequence variant interpretation (Richards *et al*., 2015), were then applied to each variant. Based on these guidelines, variants were classified as either likely benign, benign, pathogenic, likely pathogenic, or variant of unknown significance in instances where there was inadequate or no formal information available.

Table 2.1: A list of bioinformatics prediction tools used in this study

Tool	**Category**	**Reference**
CADD	Multi-data integration	Kircher *et al*., 2014
DANN	Multi-data integration	Quang *et al.*, 2015
FATHMM	Missense	Shihab *et al*., 2014
GERP++	Conservation	Davydov *et al*., 2010
HSF	Splicing	Desmet *et al*., 2009
IGV	Variant visualiser	Robinson, 2011
Mutalyzer	Checks sequence variant nomenclature	Den Dunnen, 2016
Mutation Assessor	Multi-data integration	Reva *et al*., 2017
Mutation Taster	Multi-data integration	Schwarz *et al*., 2014
MutPred Splicing	Multi-data integration	Mort *et al*., 2014
Polyphen-2	Missense	Adzhubei *et al*., 2010
PROVEAN	Missense and Indel	Choi and Chan, 2015
SIFT	Missense	Sim *et al*., 2012
SIFT Indel	Indels	Hu *et al*., 2013
VEST3	Missense	Carter *et al*., 2013

Abbreviations: CADD, Combined annotation-dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; GERP++, Genomic Evolutionary Rate Profiling; HSF, Human Splice Finder; IGV, Integrative Genomic Viewer; PolyPhen-2, Polymorphism Phenotyping v2; SIFT, Sorting Intolerant from Tolerant

2.4 Variant validation

Primers for validation of putative-disease causing variants through Sanger sequencing were designed using the free web-based software Primer3 under default parameters. A list of primers designed is shown in Table 2.2.

Table 2.2: Primers designed for validation of putative disease-causing variants by Sanger validation

	Variant	**Forward primer (5- 3')**	**Reverse primer (5-3')**
1	*TCOF1* c.4369_4373delAAGAA	TTTCAGGCATCAGAACCAATGT	GTGTCCCCAGATTACTCTTCCTC
2	*POLR1D* c.261delA	CAACATTGACCCCAGCACTT	GCTTGGGTTTGGGTCTGTTT
3	*CHD7* c.232C>T	AATGATGAGCAACACCCCTG	TGCCATATAGCTGCCCATCT
4	*CHD7* c.1931delA	ACAGCAGCCACAACAAAAGA	TCTCCTTGGGTTCTTTCGGT
5	*TCOF1* c.3708delC	AATTCCCAGGCCTCAAAAGC	ATGTTGGGAGGACTGTTGCT
6	*CHD7* c.3309_3310delCA	GGGTCGAGTGATAAAGGGGT	TTGAGTCCCTCCAACAGCTT
7	*CHD7* c.643C>T	ATATGGCACGTGGGGATTTT	AGTTAAGGGTTTGCGGCCTA

3 RESULTS

This chapter seeks to summarise all the findings obtained in this study

3.1 Demographic data

A total of 15 patients and 19 family members were recruited for this study. The objective of the study was to recruit 16 participants; however, because of the rare nature of the disorders studied, we only managed to recruit 15 participants. Patients recruited were those clinically diagnosed and/or suspected to be affected by TCS, NS, BOR- or CHARGE syndrome. No patients with MS or MFDM syndrome were recruited. The cohort represented three South African demographic groups. A summary of the 15 patients recruited in this study is presented in Table 3.1 below.

Table 3.1: A summary of the 15 patients recruited in this study

Case	Age[1]	Gender	Ethnicity	Clinical/ suspected diagnosis	Number of family members recruited	Relationship to the patient
FRASC13	3	F	African	TCS	1	Mother
FRASC23	1	M	Indian	CHARGE syndrome	1	Mother
FRASC25	7	F	Caucasian	BOR/TCS?	3	Mother, father and a sibling
FRASC26	12	F	African	TCS	3	Mother and two siblings
FRASC27	3	F	African	TCS	1	Mother
FRASC28	7	M	Caucasian	TCS	3	Mother, father and a sibling
FRASC31	6	M	African	TCS	1	Mother
FRASC33	Deceased[2]	M	Caucasian	NS	-	-
FRASC34	3	F	African	CHARGE syndrome	1	Mother
FRASC35	1	F	Caucasian	CHARGE syndrome?	1	Mother
FRASC52	Deceased[3]	F	African	TCS	2	Mother and father
FRASC54	7	M	African	CHARGE syndrome	-	-
FRASC57	1	M	African	TCS	1	Mother
FRASC59	5	F	African	CHARGE syndrome	1	Mother
FRASC61	1	M	African	CHARGE syndrome	-	-

[1] At the time of recruitment
[2] The patient had passed on at the time of recruitment. The patient passed on when he was 6-months old. Banked DNA was used for genetic testing
[3] The patient had passed on at the time of recruitment. The patient passed on when she was 5 years old. Banked DNA was used for genetic testing
Abbreviations: -, not available; BOR, Broncho-oto-Renal syndrome; CHARGE, Coloboma , Heart defects, Atresia choanae, Growth retardation, Genital abnormalities and Ear abnormalities; F, female; M, male; NS, Nager syndrome; TCS, Treacher Collins Syndrome.

3.2 Genes included in a multigene panel

A total of 12 candidate genes were identified for inclusion on the multigene targeted NGS panel. The gene list included *TCOF1, POLR1C, POLR1D, SF3B4, EFTUD2, DHODH, EYA1, CHD7, SMAD4, SMAD5, SMAD6 and SMAD7*. This list includes known genes for the three FDs under study: the *TCOF1, POLR1C* and *POLR1D* genes are known causative genes of TCS (Treacher Collins Syndrome Collaboration Group, 1996, Dauwerse *et al*., 2011). The *SF3B4* gene is a causative gene of NS (Bernier *et al*., 2012) and mutations in the *DHODH* gene cause MS (Rainger *et al*., 2012). Genes known to cause the three differential diagnoses under study were also included: the *EFTUD2* is a causative gene of MFDM (Lines *et al*, 2012), changes in *EYA1* cause BOR (Pierides *et al*., 2002) and mutations in *CHD7* cause CHARGE syndrome (Johnson *et al*., 2006). The *SMAD4, SMAD5, SMAD6 and SMAD7* genes were included as it is the investigator's opinion that they play a role in the pathogenesis of NS.

The exact mechanism by which mutations in *SF3B4* lead to the manifestation of the NS phenotype is not yet understood (Devotta *et al*., 2016). The *SF3B4* gene encodes a protein which forms part of the spliceosome complex and is also suspected to play a role in bone morphogenic protein signalling (BMP) SMAD-dependent pathway (Bernier *et al*., 2012 and Watanabe *et al*., 2007).

In the BMP SMAD-dependent pathway, BMPs binds to type II serine/threonine kinase receptor to initiate the signal transduction cascade (Kawabata *et al*., 1998). The active type II receptor then transphosphorylates the type I receptor, and the type I receptor phosphorylates the regulatory smads proteins (*SMAD1/SMAD5/SMAD8*). Phosphorylated *SMAD1/SMAD5/SMAD8* associates with the co-Smad (*SMAD4*) and the complex moves to the nucleus where it associates with coactivators or corepressors (*SMAD6* and *SMAD7*) to regulate gene expression (Miyazo *et al*., 2010).

The canonical BMP pathway is dependent on *SMAD1, SMAD4, SMAD5, SMAD6, SMAD7* and *SMAD8* for normal BMP-inducing osteogenesis and impairment in this pathway may affect bone development (Hu *et al*., 1998). Mouse knockouts of various SMADs results in the manifestation of human NS-like limb phenotype: Yang and colleagues observed that Smad4 knockout in mice results in hearing loss and inner ear malformation (Yang *et al*., 2009). Smad5 mice knockouts

have multiple embryonic defects including the axial skeletal defects (Chang *et al*., 1999) and Estrada and colleagues reported that Smad6 and Smad7 knockouts result in defects of both the axial and appendicular skeleton (Estrada *et al*., 2013). These suggest that *SMAD4, SMAD5, SMAD6* and *SMAD7* could have an important role in normal bone formation. It is speculated that should the pathogenesis of NS be associated with the BMP signalling, then mutations in BMP pathway critical SMADs could also contribute to the development of NS.

3.3 DNA target library preparation QC

Prior to target library pooling, the size distribution of each library was determined using the Bioanalyzer DNA 1000 kit HS (Agilent, Johannesburg, South Africa). All target libraries passed this QC step and proceeded to sequencing. Figure 3.1-3.4 below shows bioanalyzer traces of the 15 samples.

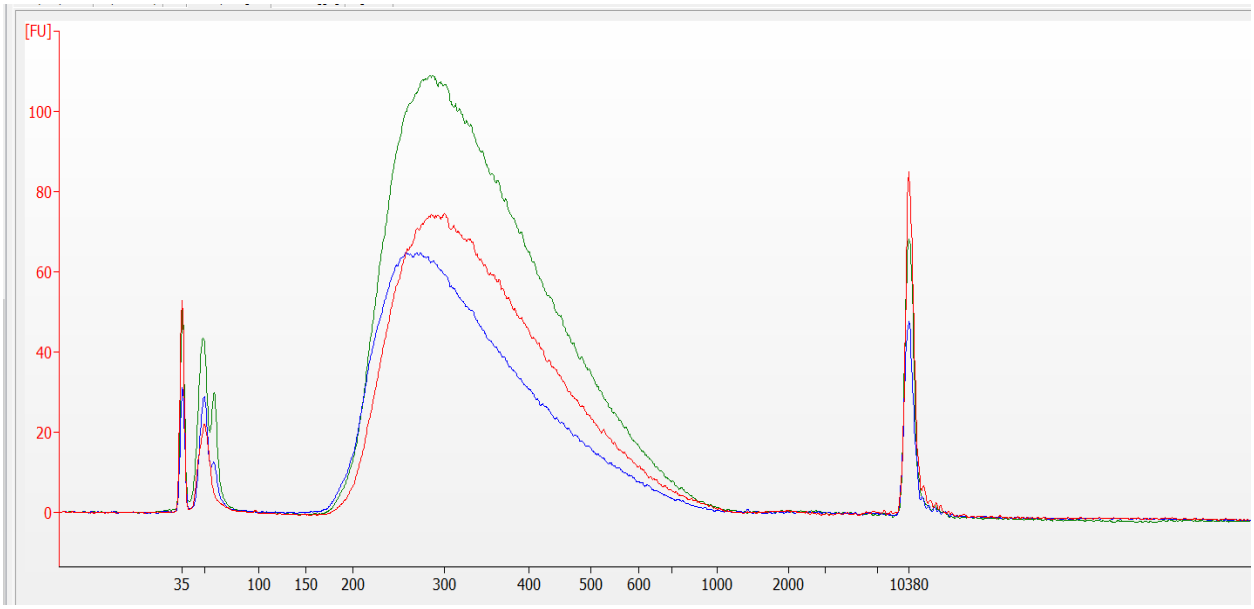

Figure 3.1: Overlaid bioanalyzer electropherograms showing the size distribution of FRASC13, FRASC27 and FRASC28.

Overlaid bioanalyzer electropherograms, showing the size distribution of FRASC13 (red), FRASC27 (green) and FRASC28 (blue). The x-axis displays the number of base pairs whereas the Y-axis plots the intensity of the sample. The first and fourth peaks represent the lower and upper marker, respectively. Sample peak heights for FRASC13 and FRASC27 are centred between 325bp and 450bp, as expected. FRASC28 is centred under 300bp and this could result in the introduction of sequencing artefact during sequencing

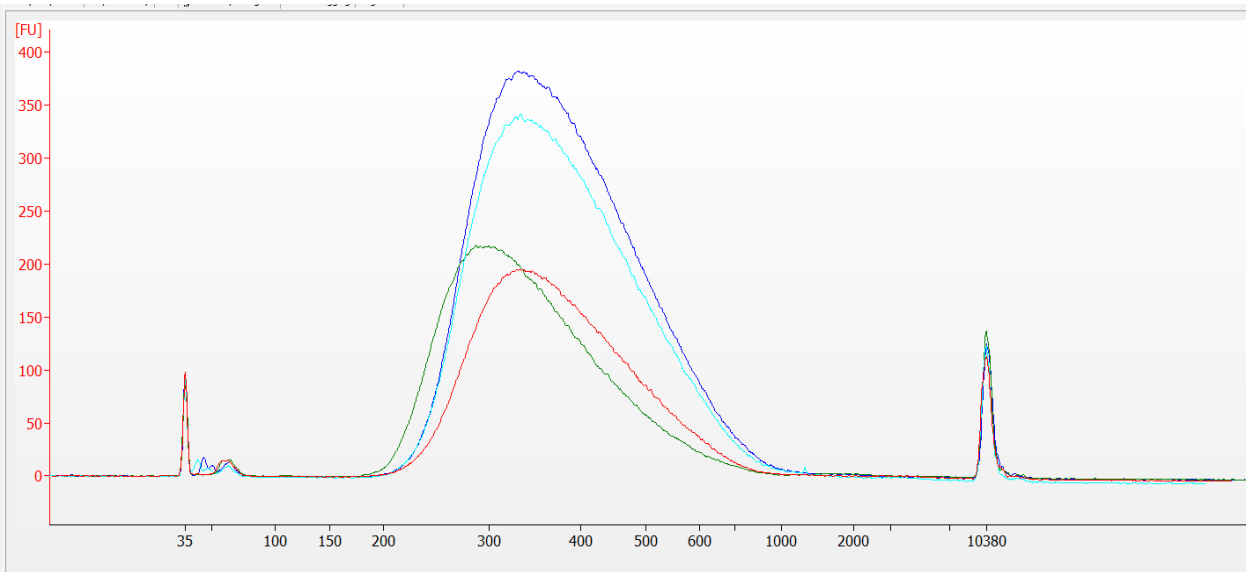

Figure 3.2: Overlaid bioanalyzer electropherograms showing the size distribution of FRASC34, FRASC61, FRASC25 and FRASC52.

Overlaid bioanalyzer electropherograms showing the size distribution of FRASC34 (red), FRASC61 (green), FRASC25 (cyan) and FRASC52 (blue). The x-axis displays the number of base pairs whereas the Y-axis plots the intensity of the sample. The first and forth peaks represent the lower and upper marker, respectively. Sample peak heights of FRASC34, FRASC25 and FRASC52 are centred between 325bp and 450bp, as expected. FRASC61 is centred around under 300bp and this could result in the introduction of sequencing artefact during sequencing.

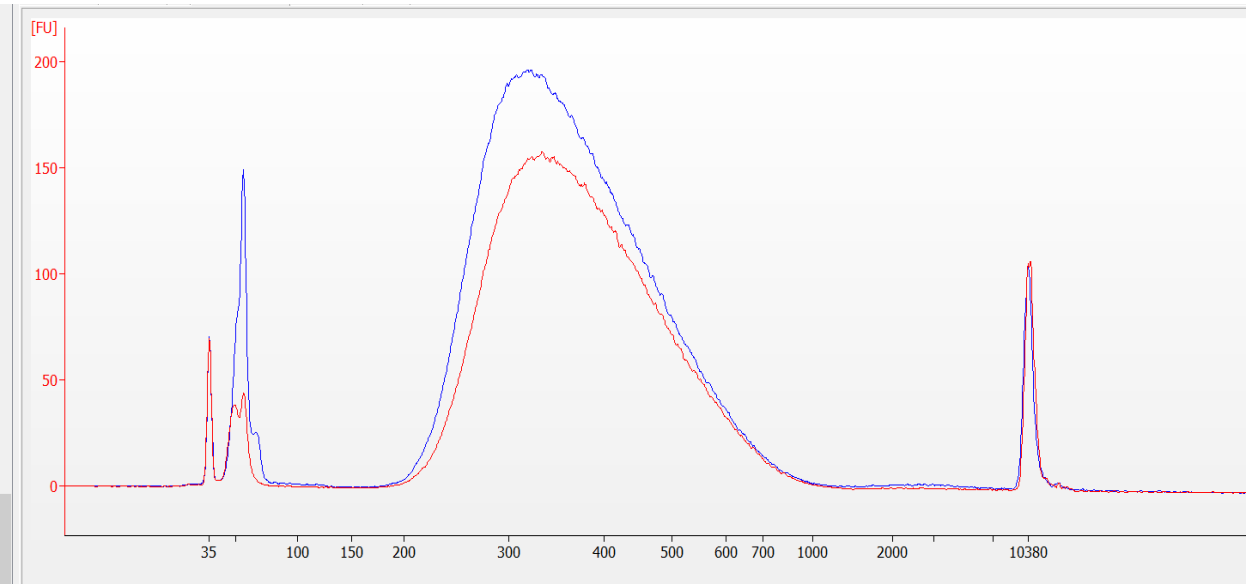

Figure 3.3: Overlaid bioanalyzer electropherograms showing the size distribution of FRASC23 and FRASC26

Overlaid bioanalyzer electropherograms showing the size distribution of FRASC23 (red) and FRASC26 (blue). The x-axis displays the number of base pairs whereas the Y-axis plots the intensity of the sample. The first and forth peaks represent the lower and upper marker, respectively. Sample peak heights are centred between 325bp and 450bp, as expected.

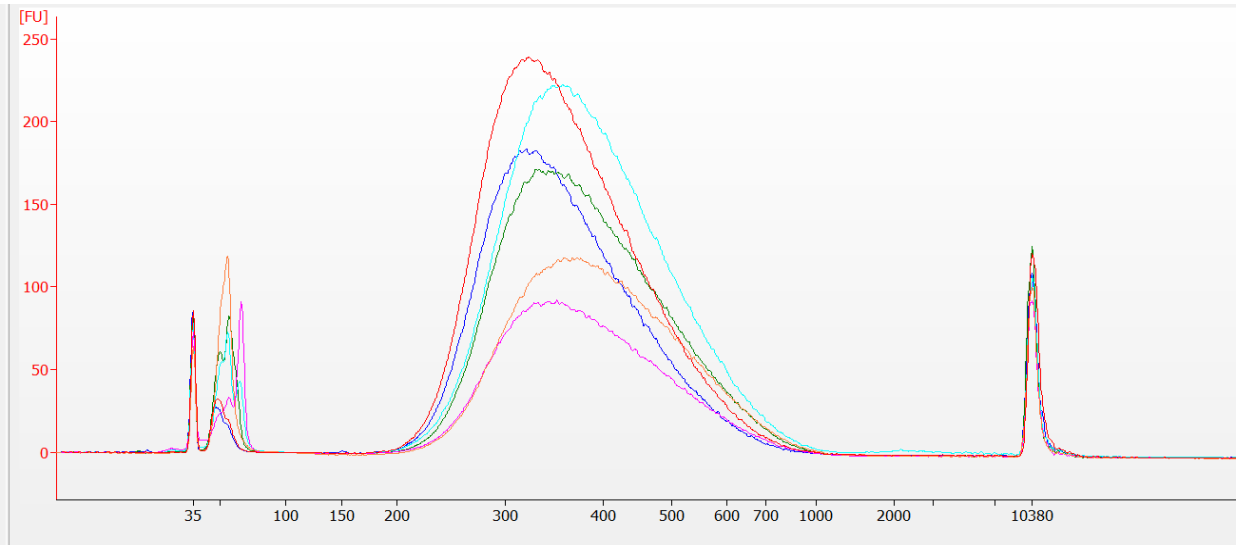

Figure 3.4: Overlaid bioanalyzer electropherograms showing the size distribution of FRASC31, FRASC33, FRASC57, FRASC35, FRASC54 and FRASC59.
Overlaid bioanalyzer electropherograms showing the size distribution of FRASC31 (red), FRASC33 (green), FRASC57 (blue), FRASC35 (cyan), FRASC54 (orange), and FRASC59 (pink).The x-axis displays the number of base pairs whereas the Y-axis plots the intensity of the sample. The first and forth peaks represent the lower and upper marker, respectively. Sample peak heights are centred between 325bp and 450bp, as expected.

3.4 Sequencing QC

Target DNA libraries that passed QC were sequenced on an Illumina MiSeq in three separate runs. During sequencing, an Illumina sequencing internal control, PhiX, was used as a technical control for the clustering reaction. PhiX library was spiked into the prepared DNA libraries and it was expected to yield a certain number of clusters. The presence of PhiX clusters is an indication that the run was successful. The PhiX was successful in all three runs, yielding an average data output of 0.02 gigabytes (GB).

With PhiX being successful in all three runs, the quality metrics of each run was then evaluated (Table 3.2). The QC metrics evaluated in this study included the number of DNA clusters generated per square millimetre for the run (cluster density), the number of clusters generated which passed an Illumina filter (cluster pass filter) and the number of bases with a Q-score of ≥ Q30. The first and third runs slightly under clustered at 628.0 k/mm^2 and 742.2 k/mm^2, respectively. The second run was under clustered at 404.6 k/mm^2. Despite the observed under

clustering, on average, 92.9 % of the data had a Q-score of ≥ Q30, indicating high-quality data generated. While under clustering results in the reduction of the output data, we had enough high-quality data to enable variant analysis. Further optimisation would be useful to optimise cluster density in future runs using this probe design.

Table 3.2: The quality metrics of the three experiments

Run	1	2	3
MiSeq v2 sequencing kit used	Nano	Micro	Micro
Amount of library loaded[a]	7	8	8
Optimal cluster density (k/mm^2)[b]	865-965	865-965	865-965
Cluster density(k/mm^2)[c]	628.0	404.6	742.2
Density Pf (k/mm^2)[d]	599.5	345.1	629.1
Number of reads generated[e] (millions)	253.6	592.0	688.5
% reads Q ≥ 30[f]	94.0	94.3	90.5

[a] The concentration of the DNA library loaded on a flow cell during sequencing
[b] The cluster density range in which optimal sequencing is predicted
[c] Cluster densities generated from our sequencing run. Shows the number of clusters per square millimetre for the run
[d] Cluster density generated from our sequencing run which passed Illumina chastity filter (usable clusters)
[e] The number of reads generated in each run in millions
[f] Percentage of bases that have a probability of incorrect base calling of 1 in 1000
Abbreviations: k\mm^2, Thousand per millimetre square; Pf, pass filter, pM=Pico molar.

Following run QC, the data quality metrics of each sample was evaluated (Table 3.3). The quality metrics evaluated in each sample included but were not limited to the number of raw reads generated in each sample and the sequencing depth. The number of raw reads generated across samples ranged from 110 000 to 661 000 reads. The average sequencing depth for target regions was 111.1× (range: 37× - 204×). In general, an average read depth of more than 20× is considered sufficient to call variants (Hancock-Hanser *et al*., 2013; Prasad *et al*., 2016). The data generated in all three runs had sufficient read depth to call variants with confidence. However, visualisation of target regions on the Integrative Genomics Viewer (IGV) revealed that some target regions in FRASC27 were not adequately covered (Figure 3.5) and analysis of mutation reports showed that a small portion of target regions in FRASC27, FRASC28 and FRASC54 had a Q-score of ≤ 20 and this may interfere with variant calling for these targeted regions.

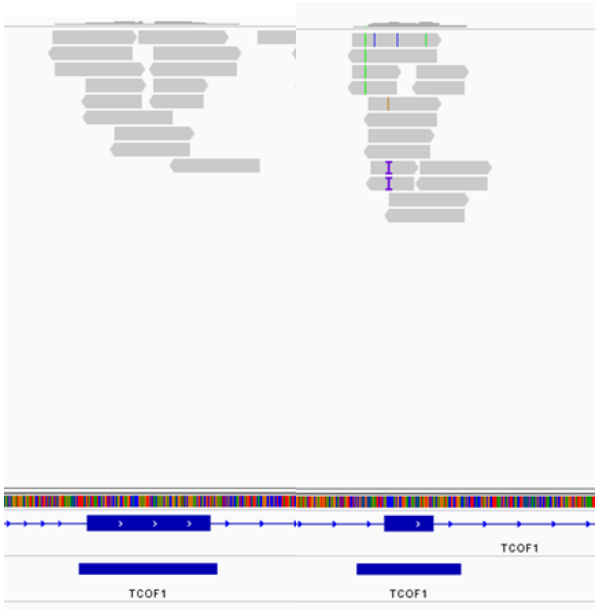

Figure 3.5: IGV screenshot showing exon eight and 21 of the *TCOF1* gene covered inadequately in FRASC27

Visualisation of targeted regions in FRASC27 revealed that some target regions, including exon eight and 21 of the *TCOF1* gene were not adequately covered and had less than 20 reads. This could compromise variant calling in this patient

Table 3.3: The quality metrics of each sample

Patient	Run #	Reads generated[1]	Average read depth[2]
FRASC13	1	159,223	56×
FRASC23	2	110, 873	37×
FRASC25	3	561,829	199×
FRASC26	2	324,689	119×
FRASC27	1	114,851	38×
FRASC28	1	110,873	37×
FRASC31	2	293,398	105×
FRASC33	2	316,383	102×
FRASC34	3	467,832	166×
FRASC35	2	304,680	104×
FRASC52	3	561,339	204×
FRASC54	2	353,053	127×
FRASC57	2	324,517	94×
FRASC59	2	288,813	88×
FRASC61	3	611,209	190×

[1] The number of raw reads generated by the sequencer for each sample
[2] The number of times each base has been read in a sample on average

3.5 Clinical and Molecular analysis results

In this study, 12 candidate genes, their exons plus 10bp flanking regions of intron-exon boundaries and their UTRs were sequenced in 15 patients using a targeted NGS approach. In this section, a summary of each patient's clinical presentation and their identified candidate variants are presented. A list of all variants identified in each patient can be seen in Appendix C1, page 128.

3.5.1 FRASC13

The patient is a three-year old female born to non-consanguineous parents. She presented with clinical features suggestive of TCS based on malar flattening, mandible hypoplasia, microtia, and lower-lid coloboma. The patient does not have a family history of TCS. Figure 3.6 below shows the clinical presentation of FRASC13.

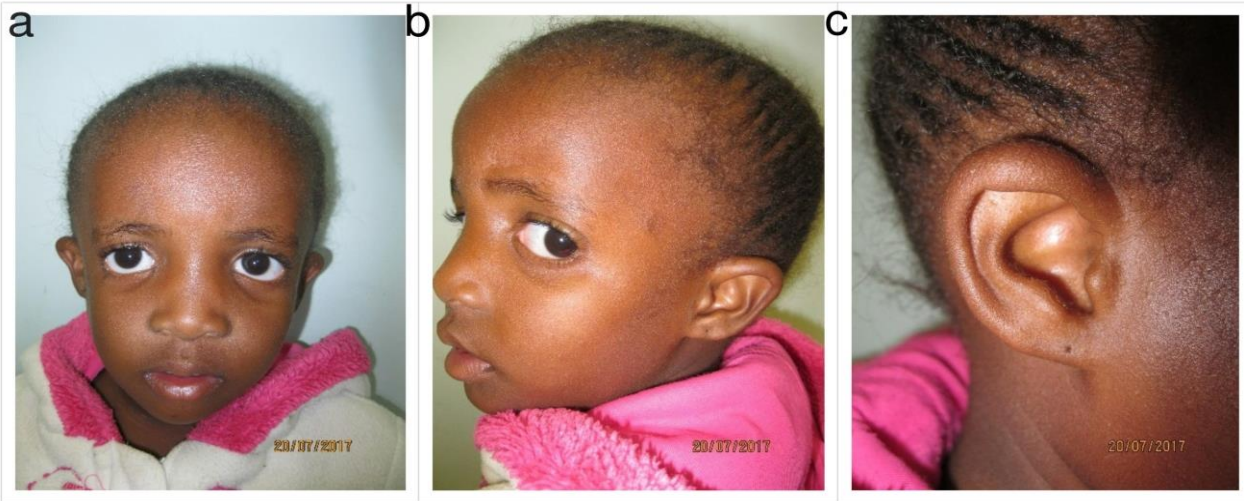

Figure 3.6: FRASC13 presenting with clinical features suggestive of TCS
Photographs of FRASC13 presenting with clinical features of TCS. Note (**a-b**)malar flattening, mandible hypoplasia, microtia and (**c**) Ear pit.

Variant calling yielded 16 variants in this patient. Of the 16 variants, eight were missense, seven were synonymous and one was an intronic variant. Variants were filtered and prioritised using the MAF and the disease model: variants with a MAF of >0.05 in gnomAD Exomes and Genomes datasets and those not segregating with the disease model were excluded from downstream analysis. Following variant filtering and prioritisation, three candidate variants were identified (Table 3.4).

Table 3.4: Candidate variants identified in FRASC13

Gene	*TCOF1*	*TCOF1*	*POLR1C*
HGVS nomenclature[1]	c.1133C>T	c.3183G>C	c.8C>T
Protein change	p.Ala378Gly	p.Gln1061His	p.Ala3Val
Function	Exonic	Exonic	Exonic
MAF[2]	0.000	-	0.000
MAF[3]	0.010	-	0.000
CADD Scaled score[4]	22.500	25.600	18.420
DANN	0.996 (Pathogenic)	0.996 (Pathogenic)	0.997 (Pathogenic)
FATHMM	0.670 (Deleterious)	0.861 (Deleterious)	0.677 (Deleterious)
GERP++	2.634 (Non-conserved)	4.801 (Conserved)	4.490 (Conserved)
HSF	NSV	SV	NSV
Mutation Assessor	0.466 (Benign)	0.654 (Deleterious)	0.431 1(Benign)
Mutation Taster	0.290 (Neutral)	0.515 (Deleterious)	0.990 (Deleterious)
MutPred Splicing	NSV	SV	NSV
PolyPhen-2	0.293 (Tolerated)	1.000 (Deleterious)	0.172 (Benign)
PROVEAN	-2.146 (Neutral)	-2.981 (Deleterious)	-0.270 (Neutral)
SIFT	0.410 (Deleterious)	0.000 (Deleterious)	0.091 (Tolerated)
VEST3	0.056 (Benign)	0.460 (Benign)	0.590 (Deleterious)
ClinVar variation ID	352202	-	-
ClinVar classification	Likely benign	-	-
ACMG codes	BS1,BP4 & BP6	PP3 & PM2	BS1 & BP4
ACMG classification	Likely benign	VUS	Likely benign

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, not available; A, Adenine; Ala, alanine; C, Cytosine; CADD, Combined Annotation-Dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, Guanine; GERP++, Genomic Evolutionary Rate Profiling; Gln, Glutamine; His, Histidine; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor allele frequency; NSV, non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; *POLR1C*, RNA polymerase I and III subunit C; SIFT, Sorting Intolerant From Tolerant; SV, Splice variant; T, Thymine; *TCOF1*, Treacle ribosome biogenesis Factor 1; V, Valine; VEST3, Variant Effect Scoring Tool 3; VUS, Variant of unknown significance.

The heterozygous *TCOF1* c.1133C>T (p.Ala378Gly) variant identified in this patient is a missense variant reported in dbSNP (rs75181211). Computational analysis of this variant yielded five benign outcomes from Mutation Assessor, Mutation Taster, PolyPhen-2, PROVEAN and VEST3. The ACMG code BP4 therefore applies (Richards *et al*., 2015). Analysis of this variant by two splicing prediction tools predicts that the variant doesn't affect splicing. Two sources

classify this variant as benign and one classifies the variant as likely benign in ClinVar (Landrum *et al*., 2014). The ACMG code BP6 therefore applies (Richards *et al*., 2015). Missense variants could possibly cause disease through alteration of the protein function or disruption of splicing but the *TCOF1* c.1133C>T variant is neither predicted to disrupt splicing nor alter protein function. Furthermore, the variant has been observed in 585 and 310 apparently healthy African individuals in gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to cause disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes BS1, BP4 and BP6, the variant was likely benign (Richards *et al*., 2015).

The heterozygous *TCOF1* c.3183G>C (p.Gln1061His) is a missense variant. The variant is not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The variant lies in a conserved region and is one nucleotide away from the splice site. Computational analysis of this variant yields seven damaging predictions from CADD, DANN, FATHMM, Mutation Assessor, PolyPhen-2, PROVEAN and SIFT (Adzhubei *et al*., 2013; Choi and Chan, 2015; Sim et al., 2012; Kircher *et al*., 2014; Quang *et al*., 2014; Shihab *et al*., 2013; Reva *et al*., 2017) and two benign predictions from VEST3 and Mutation Taster (Carter *et al*., 2013; Schwarz *et al*., 2014). The ACMG code PP3 therefore applies (Richards *et al*., 2015). Furthermore, analysis of this variant on HSF and MutPred Splicing (Desmet *et al*., 2009; Mort *et al*., 2014) predicts that the variant affects the natural 5' splice site. Based on the ACMG codes PP3 and PM2, the variant was a variant of unknown significance (VUS).

The heterozygous *POLR1C* c.8C>T (p.Ala3Val) variant identified in this patient is a missense variant reported in dbSNP (rs138184356). Four computational tools yield a damaging prediction (FATHMM, DANN, VEST3 and Mutation Taster) and five yield a benign prediction (CADD, SIFT, PolyPhen-2, Mutation Assessor and PROVEAN). The ACMG code BP4 therefore applies (Richards *et al*., 2015). Analysis of this variant by two splicing prediction tools (MutPred Splicing and HSF) predict that the variant doesn't affect splicing. Furthermore, the variant has been observed in 92 and 55 apparently healthy African individuals in gnomAD Exomes and Genomes databases, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to cause

disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes BS1 and BP4 the variant was likely benign (Richards *et al*., 2015).

3.5.2 FRASC23

The patient is a one year-old male born to non-consanguineous parents. He is the only child to his parents and has four maternal and paternal half-siblings who are clinically unaffected. The patient presented with ocular coloboma, clinodactyly, cardiac anomalies, renal anomalies, chronic lung disease, undescended testes, mild intellectual disability and hearing loss. These features were suggestive of CHARGE syndrome. Photographs of this patient were not available for publication. Variant calling yielded 31 variants in this patient. Of the 31 variants called, 23 were intronic, three were missense, four were synonymous and one was a frameshift variant. These were filtered and prioritised using the MAF and disease model. Following variant filtering, one candidate variant was identified (Table 3.5).

Table 3.5: A candidate variant identified in FRASC23

Gene	*SMAD6*
HGVS nomenclature[1]	c.887C>T
Function	Exonic
Protein change	p.Ser296Phe
MAF[2]	0.000
MAF[3]	-
CADD Scaled score[4]	34.000
DANN	0.998 (Pathogenic)
FATHMM	0.956 (Deleterious)
GERP++	5.620 (Conserved)
HSF	NSV
Mutation Assessor	0.576 (Deleterious)
Mutation Taster	1.000 (Deleterious)
MutPred Splicing	NSV
PolyPhen-2	0.990 (Deleterious)
PROVEAN	-3.110 (Deleterious)
SIFT	0.000 (Deleterious)
VEST3	0.670 (Deleterious)
ClinVar variation ID	-
ClinVar classification	-
ACMG codes	BS1 &PP3
ACMG classification	VUS

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (*Kircher et al., 2014*).
Abbreviations: &, and; -, not Available C, Cytosine; CADD, Combined annotation-dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, non-splice variant; Phe, Phenylalanine; PolyPhen-2; Polymorphism Phenotyping v2; Ser, Serine; SIFT, Sorting Intolerant From Tolerant; *SMAD6*, Mothers Against Decapentaplegic Homolog 6; T, Thymine; VUS, Variant of Unknown Significance

The heterozygous *SMAD6* variant c.887C>T (p.Ser296Phe) identified in this patient is a missense variant reported in dbSNP (rs200374822). Eight computational tools yield a damaging or pathogenic prediction (SIFT, PolyPhen-2, FATHMM, CADD, PROVEAN, DANN, Mutation Assessor and Mutation Taster). The ACMG code PP3, therefore applies (Richards *et al*., 2015). The variant lies in a conserved region and it is not predicted to affect splicing. Although eight computational tools yield a damaging classification on this variant, the variant has been observed in 230 apparently healthy individuals in gnomAD Exomes dataset, suggesting that it is unlikely to cause disease (Karczewski *et al*., 2019). The ACMG code BS1 therefore applies (Richards *et*

al., 2015). Based on the ACMG code PP3 and BS1, the variant was a VUS (Richards *et al*., 2015).

3.5.3 FRASC25

The patient is an eight-year-old female born to non-consanguineous parents. The patient presented with features suggestive of TCS or BOR based on malar flattening mandibular hypoplasia, renal anomalies, hearing loss and external ear defects which include: microtia, ear pits and ear tags, external canal atresia. The patient does not have a family history of TCS or BOR. Photographs of this patient were not available for publication. Thirty-four (34) variants were called in this patient. Of the 34 variants called, 24 were intronic, two were missense, seven were synonymous and one was a frameshift deletion. Following variant filtering and prioritisation, two candidate variants were identified (Table 3.6).

Table 3.6: Candidate variants identified in FRASC25

Gene	*POLR1D*	*SMAD7*
HGVS nomenclature[1]	c.26+10G>A	c.1086C>T
Function	Intronic	Exonic
Protein change	-	p.Pro362=
MAF[2]	-	0.000
MAF[3]	-	0.000
CADD Scaled score[4]	-	-
DANN	0.969(Pathogenic)	0.790 (Benign)
FATHMM	-	-
GERP++	-3.540 (Non-conserved)	3.810 (Non-conserved)
HSF	NSV	NSV
Mutation Assessor	-	-
Mutation Taster	-	-
MutPred Splicing	NSV	NSV
PolyPhen-2	-	-
PROVEAN	-	-
SIFT	-	-
VEST3	-	-
ClinVar variation ID	-	-
ClinVar classification	-	-
ACMG codes	PM2	BS1 & BP7
ACMG classification	VUS	Likely benign

[1] den Dunnen, 2016
[2] Karczewski *et al.*, 2019
[3] Karczewski *et al*., 2019
[4] A Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher et al., 2014).
Abbreviations: &, and; -, not Available; A, Adenine; C, Cytosine; CADD, Combined annotation-dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, Guanine; GERP++, Genomic Evolutionary Rate Profiling; HGVS, human genome variation society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, non-splice Variant; *POLR1C*, RNA polymerase I and III subunit C; *POLR1D*, RNA Polymerase I And III Subunit D; PolyPhen-2, Polymorphism Phenotyping v2; Pro, Proline; SIFT, Sorting Intolerant From Tolerant; *SMAD7*, Mothers Against Decapentaplegic Homolog 7; T, Thymine; VUS, Variant of Unknown Significance.

The heterozygous *POLR1D* c.26+10G>A is an intronic variant not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The variant lies in a non-conserved region, 10 nucleotides away from the splice site and it is not predicted to affect splicing (Jian *et a*l., 2014; Mort *et al*., 2014). A single computational tool (DANN) yields a pathogenic classification. Based on the ACMG code PM2, the variant was a VUS (Richards *et al*., 2015).

The heterozygous *SMAD7* c.1086C>T (p.Pro362=) variant identified in this patient is a synonymous variant reported in dbSNP (rs149492644). The c.1086C>T variant is located 13 bases away from the splice site and is not predicted to impact splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). Synonymous variants could possibly cause disease through disruption of splicing or altering gene expression (Sauna *et al*., 2001), but the c.1086C>T variant is neither predicted to disrupt splicing nor alter gene expression. Furthermore, the variant has been observed in 117 and 86 apparently healthy African individuals in gnomAD Exomes and Genomes databases, respectively. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes BP7 and BS1, the variant was classified as likely benign (Richards *et al*., 2015).

3.5.4 FRASC26

The patient is a 13-year old female born to non-consanguineous parents. She presented with ptosis, lower-lid coloboma, malar flattening, mandibular hypoplasia, low-set ears, microtia, external ear canal and sensorineural hearing loss suggesting a clinical diagnosis of TCS. Her mother and two siblings also presented with features suggesting TCS. In addition, she has a family history of similarly affected females from her maternal side. Figure 3.7 below shows the pedigree of the patient.

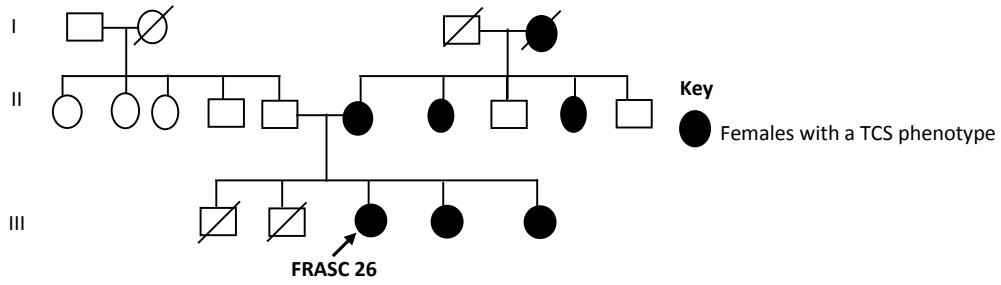

Figure 3.7: Pedigree of FRASC26
The patient is the third born to her non-consanguineous parents. Her mother and her two siblings also presented with clinical features suggestive of TCS. Furthermore, she has a family history of similarly affected females from her maternal side.

Variant calling yielded 33 variants in this patient. Of the 33 variants, three were missense, four were synonymous and 26 were intronic variants. One frameshift variant was identified through manual inspection of the BAM file. Following variant filtering and prioritisation, two candidate variants were identified (Table 3.7).

Table 3.7: Candidate variants identified in FRASC26

Gene	*EFTUD2*	*TCOF1*
HGVS nomenclature[1]	c.26+10C>T	c.4369_4373delAAGAA
Protein change	-	p.Lys1457GlufsTer12
Function	Intronic	Exon
MAF[2]	-	0.000
MAF[3]	0.042	-
CADD Scaled score[4]	-	-
DANN	0.446 (Benign)	-
FATHMM	-	-
GERP++	-7.560 (Non-conserved)	4.960 (Conserved)
HSF	NSV	-
Mutation Assessor	-	-
Mutation Taster	-	-
MutPred Splicing	NSV	-
PolyPhen-2	-	-
PROVEAN	-	-
SIFT	-	-
SIFT indel	-	Non-NMD
VEST3	-	-
ClinVar variation ID	-	19002
ClinVar classification	-	Pathogenic
ACMG codes	BA1 & BP4	PM1, PSV1 & PP5
ACMG classification	Benign	Pathogenic

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] A scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, Not available; A, Adenine; CADD, Combined annotation-dependent depletion; C, Cytosine; DANN, Deleterious annotation of genetic variants using neutral networks; *EFTUD2*, Elongation factor Tu GTP-Binding domain-containing protein 2; FATHMM, Functional analysis through hidden markov models, G, Guanine; GERP++, Genomic Evolutionary Rate Profiling; Glu, Glutamic acid HGVS, Human Genome Variation Society; HSF, Human Splice Finder; Lys, Lysine; MAF, Minor Allele Frequency; -, no prediction; NSV, non-splice variant; *POLR1D*, RNA Polymerase I And III Subunit D; PolyPhen-2, Polymorphism Phenotyping v2; *POLR1C*, RNA polymerase I and III subunit C; SIFT, Sorting Intolerant From Tolerant; T, Thymine; *TCOF1*, Treacle ribosome biogenesis factor 1;Ter, termination.

The heterozygous *EFTUD2* c.26+10C>T identified in this patient is an intronic variant reported in dbSNP (rs8080227). The variant lies in a non-conserved region and is an NSV (Desmet *et al*., 2009; Mort *et al*., 2014). One computational tool (DANN) yields a benign classification and no benign prediction is yielded. The ACMG code BP4 therefore applies (Richards *et al*., 2015). The variant has been observed in 1265 apparently health African individuals, suggesting that it is unlikely to cause disease. Moreover, the variant has an allele frequency of more than 0.05 in 1000 genomes dataset and gnomAD African dataset, with 8722 alleles observed in the population (Karczewski *et al*., 2019; The 1000 Genomes Project Consortium, 2015). Therefore, the ACMG code BA1 applies (Ghosh *et al*., 2018). Intronic variants located in non-regulator regions and not predicted to alter splicing are unlikely to cause disease (Shaez *et al*., 2014). Based on the ACMG codes BA1 and BP4, the variant was classified as benign (Richards *et al*., 2015)

The heterozygous *TCOF1* c.4369_4373delAAGAA deletion identified in this patient was identified through manual inspection of the BAM file. The deletion lies in one of the *TCOF1* reported mutation hot spots, exon 24 (Dauwerse *et al*., 2011). Therefore, the ACMG code PM1 applies (Richards *et al*., 2015). The *TCOF1* c.4369_4373delAAGAA deletion introduces a premature stop codon at amino acid 1469 of the treacle protein (see Figure 3.8) and its analysis by the SIFT Indel prediction tool (Hu and Ng, 2013), predicts that variant's pre-mRNA transcript is not a target of NMD, however, the deletion is located in a known functional domain, the ACMG code PSV1 therefore applies (Tayoun *et al*., 2018). Four reports classify this deletion as pathogenic in ClinVar (Landrum *et al*., 2014). The ACMG code PP5 therefore applies (Richards *et al*., 2015). Based on the ACMG codes PM1, PSV1 and PP5, the variant is pathogenic (Richards *et al*., 2015). Figure 3.9 below shows an image of the *TCOF1* c.4369_4373delAAGAA deletion as visualised on IGV.

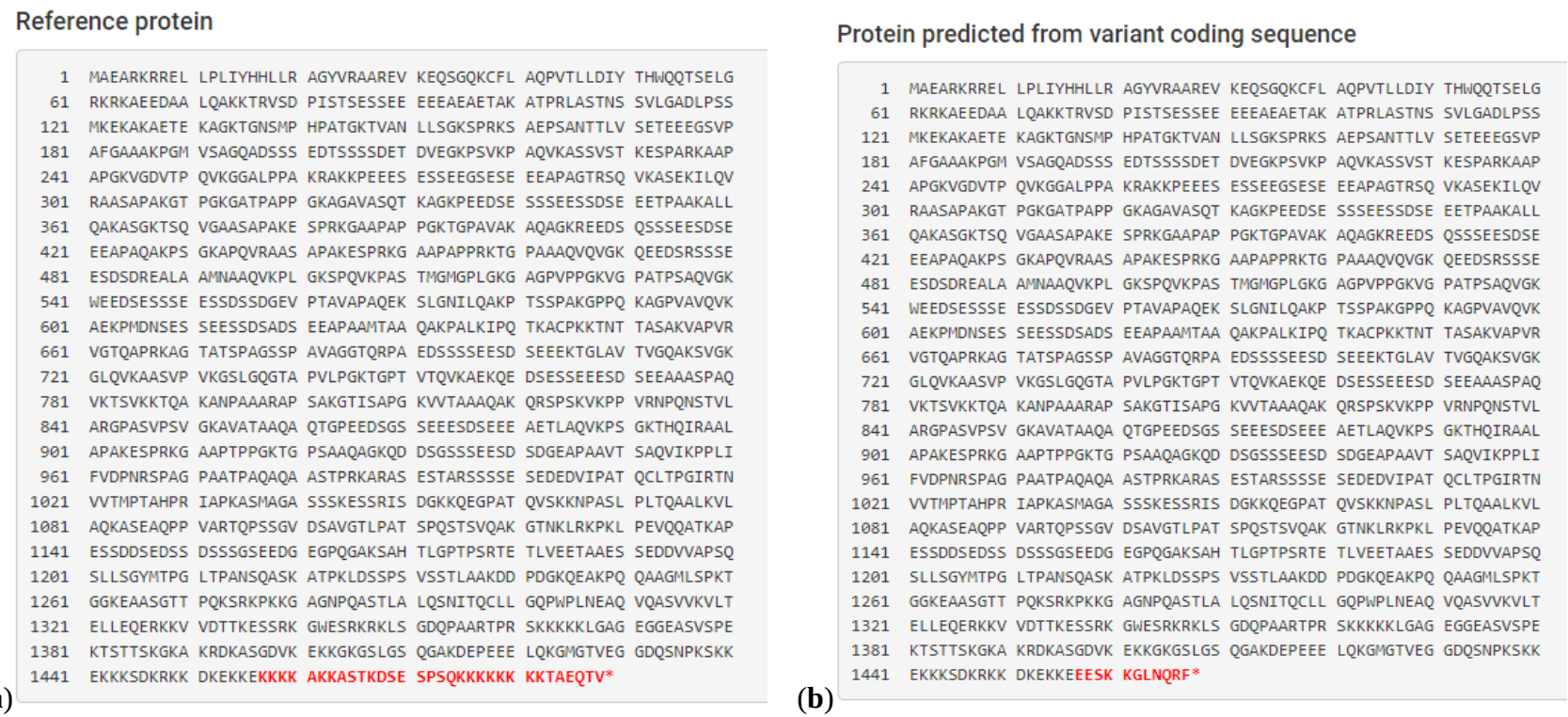

Figure 3.8: The *TCOF1* c.4916delAAAAG resultant predicted protein

The *TCOF1* c.4916delAAAAG deletion introduces a premature stop codon at amino acid 1469 of the treacle protein. Note (**a**) the TCOF1 reference protein and (**b**) the TCOF1 predicted truncated protein. The amino acid sequences were obtained from Mutalyzer.

Figure 3.9: A screenshot of IGV showing the *TCOF1* c.4916delAAAAG variant identified in FRASC26

The genetic variant was visualised by importing the BAM file into IGV

3.5.5 FRASC27

The patient is a three-year female born to non-consanguineous parents. She presented with malar flattening, grade 3 microtia, mandibular hypoplasia, ear pits and external canal atresia. These findings suggested a clinical diagnosis of TCS. Figure 3.10 below shows the clinical presentation of FRASC27.

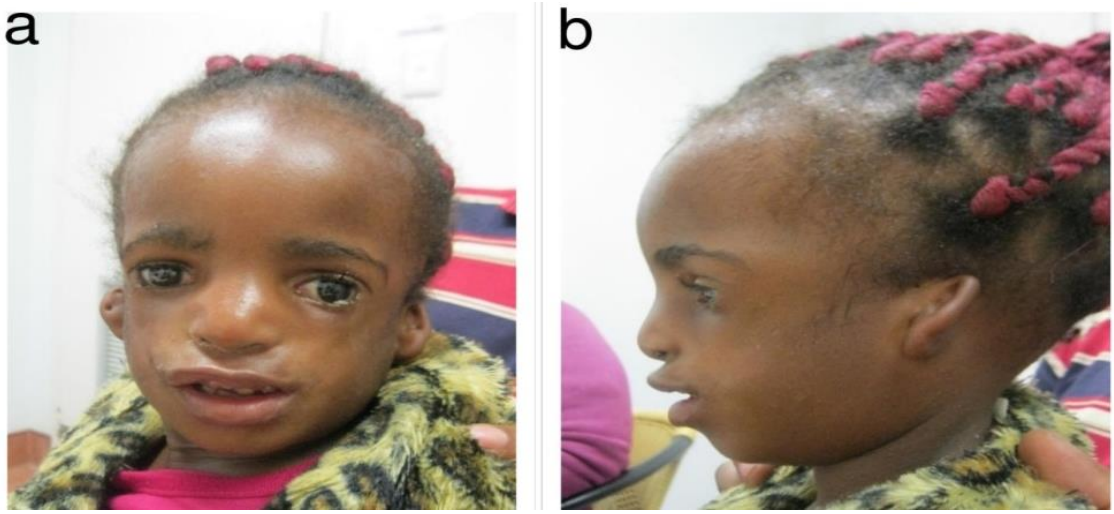

Figure 3.10: FRASC27 presenting with clinical features suggestive of TCS
Photographs showing frontal (**a**) and lateral (**b**) views of FRASC27 presenting with malar flattening, mandible hypoplasia, grade three microtia and external canal atresia.

Six variants were called in this patient. Of the six variants called, three were synonymous, two were missense and one variant was a frameshift. Following variant filtering and prioritisation, one candidate variant was identified (Table 3.8).

Table 3.8: A candidate variant identified in FRASC27

Gene	*POLR1C*
HGVS nomenclature[1]	c.510C>G
Protein change	p.Thr170=
Function	Exonic
MAF[2]	0.000
MAF[3]	0.000
CADD Scaled score[4]	-
DANN	0.749 (Benign)
FATHMM	-
GERP++	2.460 (Non-conserved)
HSF	NSV
Mutation Assessor	-
Mutation Taster	-
MutPred Splicing	NSV
PolyPhen-2	-
PROVEAN	-
SIFT	-
VEST3	-
ClinVar variation ID	-
ClinVar classification	-
ACMG codes	BS1
ACMG classification	Likely benign

1 den Dunnen, 2016
2 Karczewski *et al*., 2019
3 Karczewski *et al*., 2019
4 A Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, not Available; C, cytosine; CADD, Combined annotation-dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, guanine, GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, non-splice variant; *POLR1C*, RNA polymerase I and III subunit C; PolyPhen-2, Polymorphism Phenotyping v2; SIFT, Sorting Intolerant From Tolerant; Thr, Threonine; VUS, Variant of Unknown Significance.

The *POLR1C* c.510C>G (p.Thr170=) is a synonymous variant reported in dbSNP (rs140188270). The patient is a heterozygote for this variant. The variant is eight bases away from the splice site and is not predicted to affect splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). The variant lies in a non-conserved region (Jian *et al*., 2014). A single computational tool (DANN) yields a benign classification and no pathogenic prediction is yielded. The ACMG code BP4 therefore applies (Richards *et al*., 2015). Furthermore, the variant has been observed in 12 and 10 apparently healthy African individuals in the gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to

cause disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Synonymous variants could possibly cause disease through disruption of splicing or altering gene expression (Sauna *et al*., 2001), however, the *POLR1C* c.510C>G variant is not predicted to disrupts splicing nor alter gene expression. Based on the ACMG code BP4, BS1 and BP7, the variant was likely benign (Richards *et al*., 2015).

Sequencing generated 114 851 raw reads in this sample and on average, each base was read 38 times (38×). When visualised on IGV, 14 *TCOF1* exons (exon seven, eight, nine, ten, 11, 12, 14, 16, 17, 19, 21, 23, 24 and 25) were not covered adequately and a number of target regions in this sample also had a Q-score of less than 20 (93% of *DHODH* exon five, 100% of *SF3B4* exon four, 100% of *TCOF1* exon 21, 100% of *TCOF1* exon 22, 58% of *TCOF1* exon seven and 100% of *TCOF1* exon eight). Thus, variant calling could have been compromised in this patient.

3.5.6 FRASC28

The patient was a seven-year old male born to non-consanguineous parents. The patient presented with malar flattening, mandibular hypoplasia, low set ears, ear tags, undescended testes and hearing loss. These findings suggested a clinical diagnosis of TCS. Figure 3.11 below shows the clinical presentation of FRASC28.

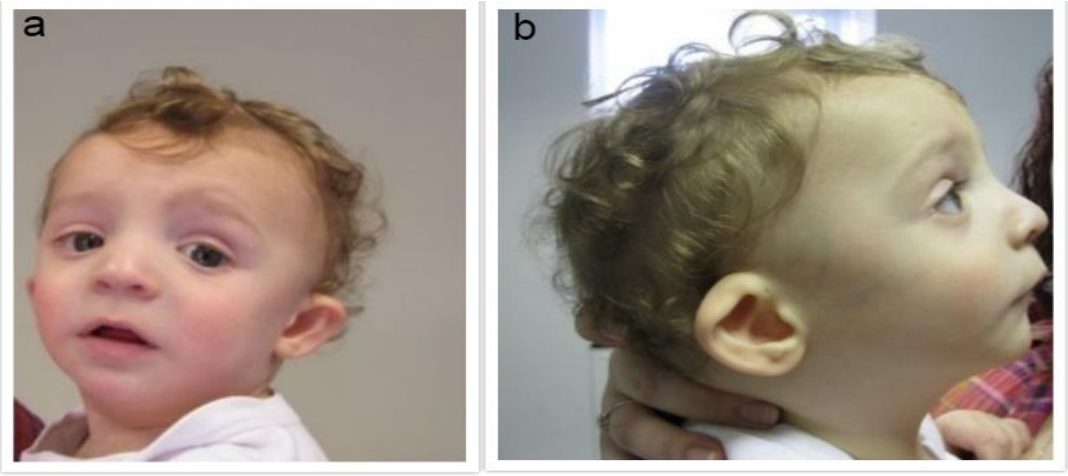

Figure 3.11: FRASC28 presenting with clinical features suggestive of TCS
Photographs showing frontal (**a**) and lateral (**b**) views of FRASC28 patient with malar flattening, mandible hypoplasia, low set ears and ear pits.

Variant calling in this patient yielded six variants. Of the six variants, two were missense, two were synonymous and two were frameshift deletions. Following variant filtering and prioritisation, three candidate variants were identified (Table 3.9).

Table 3.9: Candidate variants identified in FRASC28

Gene	*SF3B4*	*POLR1D*	*POLR1D*
HGVS[1]	c.735C>T	c.261delA	c.300C>T
Protein change	p.Pro245	p.Gly88ValfsTer13	p.Ser100
Function	Exonic	Exonic	Exonic
MAF[2]	0.003	-	0.021
MAF[3]	0.003	-	0.019
CADD Scaled score[4]	-	-	-
DANN	0.717 (Benign)	-	-
FATHMM	-	-	-
GERP++	-3.080 (non-conserved)	-11.700 (non-conserved)	-
HSF	NSV	-	NSV
Mutation Assessor	-	-	-
Mutation Taster	-	-	-
MutPred Splicing	NSV	-	NSV
PolyPhen-2	-	-	-
PROVEAN	-	-	-
SIFT	-	-	-
SIFT Indel	-	Non-NMD	-
VEST3	-	-	-
ClinVar variation ID	277035	-	-
ClinVar classification	Likely benign	-	-
ACMG codes	BS1, BP6 & BP7	PSV1 & PM2	BS1 & BP7
ACMG classification	Likely benign	Likely pathogenic	Likely benign

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, Not Available; A, Adenine; C, cytosine; CADD, Combined annotation-dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, guanine; GERP++, Genomic Evolutionary Rate Profiling; Gly, Glycine; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, non-splice variant; *POLR1D*, RNA Polymerase I And III Subunit D; PolyPhen-2, Polymorphism Phenotyping v2; *POLR1C*, RNA polymerase I and III subunit C; Pro, Proline; Ser, Serine; SIFT, Sorting Intolerant From Tolerant; T, Thymine; Ter, termination; Val, Valine; VUS, Variant of Unknown Significance.

The *SF3B4* c.735C>T (p.Pro245) is a synonymous variant reported in dbSNP (rs113949235). The patient is a heterozygote for this variant. The c.735C>T variant lies in an evolutionary non-conserved region and is not predicted to affect splicing. The ACMG code BP7 therefore applies

(Richards *et al*., 2015). A single source classifies this variant as likely benign in ClinVar (Landrum *et al*., 2014). Therefore, the ACMG code BP6 applies (Richards *et al*., 2015). The variant has been observed in 16 and 14 apparently healthy African individuals in the gnomAD Exomes and Genomes databases, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to cause disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Synonymous variants could possibly cause disease through disruption of splicing or altering gene expression (Sauna *et al*., 2001), however, the *SF3B4* c.735C>T variant is not predicted to disrupts splicing nor alter gene expression. Based on the ACMG codes BS1, BP6 and BP7, the variant was likely benign (Richards *et al*., 2015).

The heterozygous *POLR1D* c.300C>T (p.Ser100) variant identified in this patient is a synonymous variant reported in dbSNP (rs41291680). Analysis of this variant on MutPred Splicing predicts that the variant doesn't affect splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). The variant has been observed in 72 and 34 apparently healthy African individuals in gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to cause disease. Therefore, the ACMG code BS1 applies (Richards *et al*., 2015). Based on the ACMG codes BP7 and BS1, the variant was likely benign classification (Richards *et al*., 2015).

The *POLR1D* c.261delA (p.Gly88ValfsTer13) identified in this patient is not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The deletion introduces a premature stop codon at position 13 downstream (Figure 3.12). Analysis of this variant by SIFT Indel (Hu and Ng, 2012), predicts that the variant's pre-mRNA product is not a target for NMD. However, 19% of the protein is removed by this variant. The ACMG code PSV1 therefore applies (Tayoun *et al*., 2018). Based on the ACMG codes PSV1 and PM2, the variant was likely pathogenic (Richards *et al.*, 2015). Figure 3.13 shows an image of the c.261delA deletion identified in this patient visualised on IGV.

(**a**)

Reference protein

```
  1  MEEDQELERK ISGLKTSMAE GERKTALEMV QAAGTDRHCV TFVLHEEDHT LGNSLRYMIM
 61  KNPEVEFCGY TTTHPSESKI NLRIQTRGTL PAVEPFQRGL NELMNVCQHV LDKFEASIKD
121  YKDQKASRNE STF*
```

(**b**)

Protein predicted from variant coding sequence

```
  1  MEEDQELERK ISGLKTSMAE GERKTALEMV QAAGTDRHCV TFVLHEEDHT LGNSLRYMIM
 61  KNPEVEFCGY TTTHPSESKI NLRIQTRVPF QLLSHFREA*
```

Figure 3.12: The *POLR1D* c.261delA resultant predicted protein

The *POLR1D* c.261delA introduces a premature stop codon at amino acid 87 of the POLR1D protein. Note (**a**) The normal POLR1D protein reference sequence (b) The protein predicted from the c.261delA. The amino acid sequences were obtained from Mutalyzer.

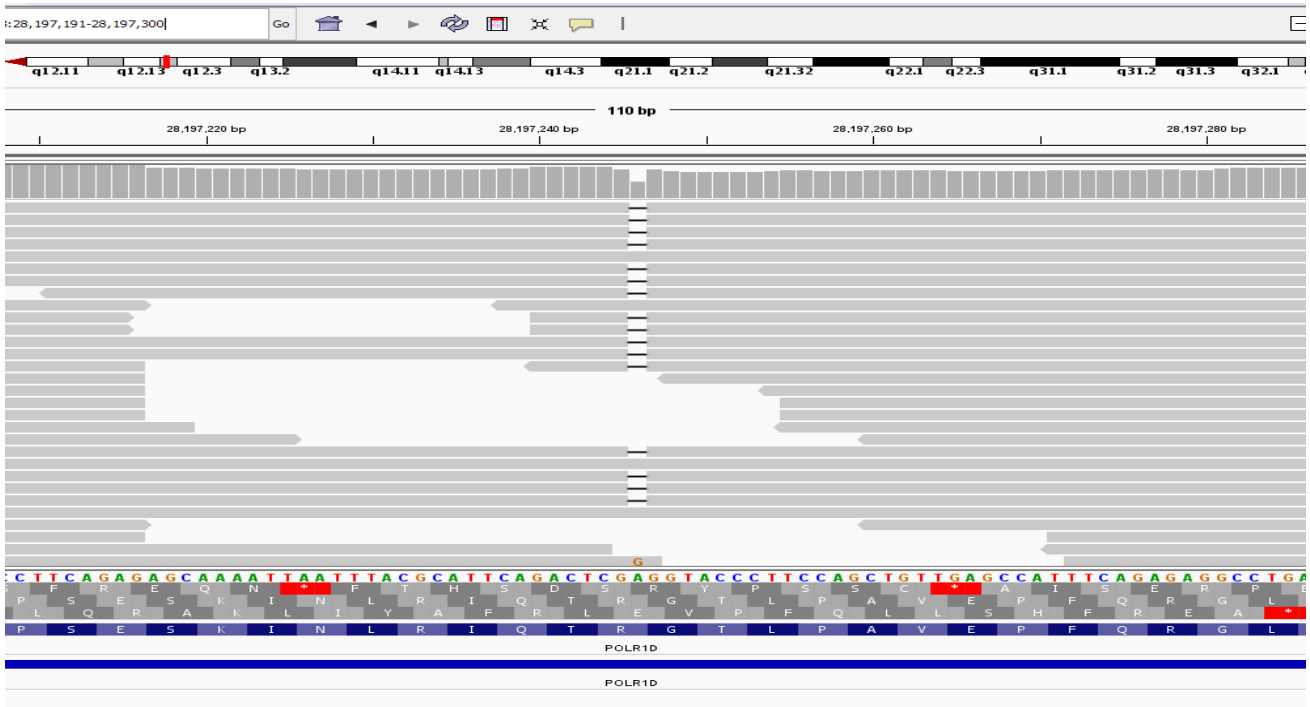

Figure 3.13:A screenshot of IGV showing the *POLR1D* c.261delA variant identified in FRASC28

The genetic variant was visualised by importing the BAM file into IGV.

3.5.7 FRASC31

The patient is a six-year old male. He is the only child to his non-consanguineous parents. The patient presented with malar flattening, mandibular hypoplasia, low set ears, microtia, ear tags, ear pits, external canal atresia and hearing loss. These findings suggested a clinical diagnosis of TCS. The patient doesn't have a family history of TCS. Figure 3.14 below shows the clinical presentation of FRASC31

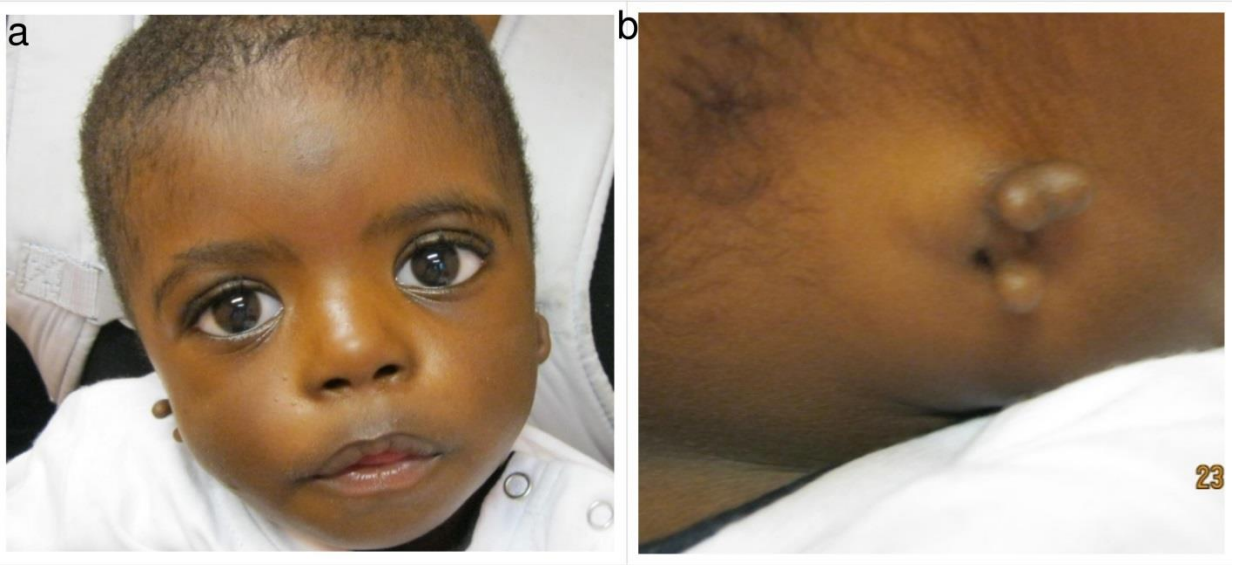

Figure 3.14: FRASC31 presenting with clinical features suggestive of TCS
Photographs showing FRASC31 presenting with clinical features suggestive of TCS. Note (**a**) mandible hypoplasia, malar flattening (**b**) grade 3 microtia.

Variant calling yielded 30 variants in this patient. Of the 30 variants, 22 were intronic, 4 were missense and 4 were synonymous variants. Variants were filtered and prioritised using the MAF and disease model. Following variant filtering and prioritisation, one candidate variant was identified (Table 3.10).

Table 3.10: A candidate variant identified in FRASC31

Gene	*CHD7*
HGVS nomenclature[1]	c.3697G>A
Protein change	p.Gly1233Ser
Function	Exonic
MAF[2]	0.000
MAF[3]	0.000
CADD Scaled score[4]	27.500
DANN	0.998 (Pathogenic)
FATHMM	0.912 (Damaging)
GERP++	5.470 (Conserved)
HSF	NSV
Mutation Assessor	0.08 (Neutral)
Mutation Taster	1.000 (Damaging)
MutPred Splicing	NSV
PolyPhen-2	0.750 (Damaging)
PROVEAN	0.06 (Neutral)
SIFT	0.100 (Tolerated)
VEST3	Neutral
ClinVar variation ID	191709
ClinVar classification	Benign/likely benign
ACMG codes	BS1 & BP6
ACMG classification	Likely benign

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; A, Adenine; CADD, Combined annotation-dependent depletion; *CHD7*, Chromodomain Helicase DNA Binding Protein 7; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, Guanine; GERP++, Genomic Evolutionary Rate Profiling; Gly, Glycine; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; Ser, Serine; SIFT, Sorting intolerant from tolerance.

The heterozygous *CHD7* c.3697G>A (p.Gly1233Ser) identified in this patient is a missense variant reported in dbSNP (rs190548814). Five computational tools yield a damaging (FATHMM, CADD, DANN, PolyPhen-2 and Mutation Taster) and four yield a benign verdict (PROVEAN, VEST3, Mutation Assessor and SIFT). The variant lies in a conserved region and is predicted to not affect splicing. Three separate reports classify the variant as benign and one as likely benign in ClinVar (Landerum *et al*., 2014). The ACMG code BP6 therefore applies (Richards *et a*l., 2015). Furthermore, the variant has been observed in 136 and 64 apparently

healthy African individuals in the gnomAD Exomes and Genomes databases, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to cause disease. The ACMG code BS1 therefore applies (Richards *et a*l., 2015). Based on the ACMG codes BS1 and BP6, the variant was likely benign (Richards *et a*l., 2015).

3.5.8 FRASC33

The patient is a second child of non-consanguineous parents. The patient passed on when he was 6 months old due to unknown causes. At the time of consultation, the patient presented with micrognathia, down-slanting palpebral fissures, preauricular skin tags, cleft palate and hypoplasia of the right thumb. These findings suggested a clinical diagnosis of NS. The patient doesn't have a family history of NS and his images were not available for publication. Twenty-one (21) variants were called in this patient. Of the 21 variants called, 16 were intronic, two were missense, two were synonymous and one was a frameshift deletion. Following variant filtering and prioritisation, one candidate variant was identified (Table 3.11).

Table 3.11: A candidate variant identified in FRASC33

Gene	*CHD7*
HGVS nomenclature[1]	c.307T>A
Protein change	p.Ser103Thr
Function	Exonic
MAF[2]	0.012
MAF[3]	0.012
CADD Scaled score[4]	19.780
DANN	0.985 (damaging)
FATHMM	0.827 (deleterious)
GERP++	5.360 (conserved)
HSF	NSV
Mutation Assessor	0.610 (Deleterious)
Mutation Taster	0.997 (Deleterious)
MutPred Splicing	NSV
PolyPhen-2	0.378 (Benign)
PROVEAN	-1.07 (neutral)
SIFT	0.010 (Deleterious)
VEST3	Benign
ClinVar variation ID	95782
ClinVar classification	-
ACMG codes	BS1 & BP6
ACMG classification	Likely benign

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014). **Abbreviations**: &, and; -, Not Available; A, adenine; CADD, Combined annotation-dependent depletion; *CHD7*, Chromo domain Helicase DNA Binding Protein 7; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; Ser, Serine; SIFT, Sorting Intolerant From Tolerant; T, thymine; Thr, Threonine.

The *CHD7* c.307T>A (p.Ser103Thr) variant is a missense variant reported in dbSNP (rs41272435). The patient is a heterozygote for this variant. Five computational tools yield a damaging (SIFT, FATHMM, DANN, Mutation Assessor, Mutation Taster) and three yield a benign verdict (CADD, PROVEAN, VEST3 and PolyPhen-2). The variant is 481 nucleotides from the splice site and its analysis on MutPred Splicing tool (Mort *et al*., 2014), predicts that it doesn't impact splicing. The *CHD7* c.307T>A variant is classified as benign in other CHARGE syndrome studies (Bergman *et al*., 2012; Bartels *et al*., 2010; Kim *et al*., 2008; Wincent *et al*., 2008). Eight separate reports classify the variant as benign and two classify it as likely benign in

ClinVar (Landerum *et al*., 2014). Therefore, the ACMG code BP6 applies (Richards *et al*., 2015). The variant has been observed in 49 and 38 apparently healthy African individuals in gnomAD Exomes and Genomes databases, (Karczewski *et al*., 2019), suggesting that the variant is unlikely to cause disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes BS1 and BP6, the variant was likely benign (Richards *et a*l., 2015).

3.5.9 FRASC34

The patient is a three-year old female. The patient presented with developmental delay, bilateral ocular coloboma, low-set ears and hearing loss. Her features suggested a clinical diagnosis of CHARGE syndrome. The patient doesn't have a family history of CHARGE syndrome and her images were not available for publication. Twenty-eight (28) variants were called in this patient. Of the 28 variants called, 22 were intronic, five were missense, and one was a stop gain variant. Following variant filtering, two candidate variants were identified (Table 3.12).

Table 3.12: Candidate variants identified in FRASC34

Gene	*TCOF1*	*CHD7*
HGVS[1]	c.2699T>C	c.232C>T
Function	Exonic	Exonic
Protein change	p.Leu900Ser	p.Gln78Ter
MAF[2]	0.000	-
MAF[3]	0.000	-
CADD Scaled score[4]	8.890	35.000
DANN	0.924 (Benign)	(Pathogenic)
FATHMM	0.600 (Benign)	0.920 (Pathogenic)
GERP++	1.870 (Non-conserved)	Conserved
HSF	NSV	NSV
Mutation Assessor	0.002(Neutral)	-
Mutation Taster	0.090 (Neutral)	1.000 (deleterious)
MutPred Splicing	NSV	NSV
PolyPhen-2	0.020 (Tolerated)	-
PROVEAN	0.033 (Neutral)	-
SIFT	0.12 (Tolerated)	-
VEST3	0.030 (Benign)	-
ClinVar variation ID	-	-
ClinVar classification	-	-
ACMG codes	BS1 & BP4	PSV1, PP3 & PM2
ACMG classification	Likely benign	Pathogenic

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski et al., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher et al., 2014).
Abbreviations: &, and; -, not Available; C, Cytosine; CADD, Combined annotation-dependent depletion; *CHD7*, Chromodomain Helicase DNA Binding Protein 7; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; Leu, Leucine; MAF, Minor Allele Frequency; NSV, non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; Ser, Serine; SIFT, Sorting Intolerant From Tolerant; T, thymine; *TCOF1*, Treacle Ribosome Biogenesis Factor; Ter, termination

The heterozygous *TCOF1* c.2699T>C (p.Leu900Ser) variant identified in this patient is a missense variant reported in dbSNP (rs751264680). Missense variants could possibly cause disease through alteration of protein function or affecting splicing. Eight computational tools (DANN, SIFT, PolyPhen-2, FATHMM, PROVEAN, DANN, Mutation Assessor and Mutation Taster) yield a benign classification. The ACMG code BP4 therefore applies (Richards *et al*., 2015). According to GERP++ (Davydov *et al*., 2010) and MutPred Splicing (Mort *et al*., 2014),

the variant lies in a non-conserved region and is predicted to have no impact on splicing. Furthermore, the variant has been observed in five apparently healthy individuals of which three are Africans in gnomAD datasets (Karczewski *et al*., 2019), suggesting that it is unlikely to cause disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes BS1 and BP4, the variant was likely benign (Richards *et al*., 2015).

The heterozygous *CHD7* c.232C>T (p.Gln78Ter) variant identified in this patient is not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). Figure 3.10 below shows an image of *CHD7* c.232C>T variant as visualised on IGV (Figure 3.15). The substitution of cytosine with thymine in this position introduces a stop codon at amino acid 78 into the protein (Figure 3.16). The ACMG code PSV1 therefore applies (Richards *et al*., 2015). The variant lies in a conserved region across species and alterations in conserved regions are highly pathogenic (Cooper *et al*., 2005). Four computational tools (FATHMM, CADD, DANN and Mutation Taster) yield a pathogenic verdict. The ACMG code PP3 therefore applies (Richards *et al*., 2015). Based on the ACMG codes PVS1, PP3 and PM2 the variant was pathogenic (Richards *et al*., 2015).

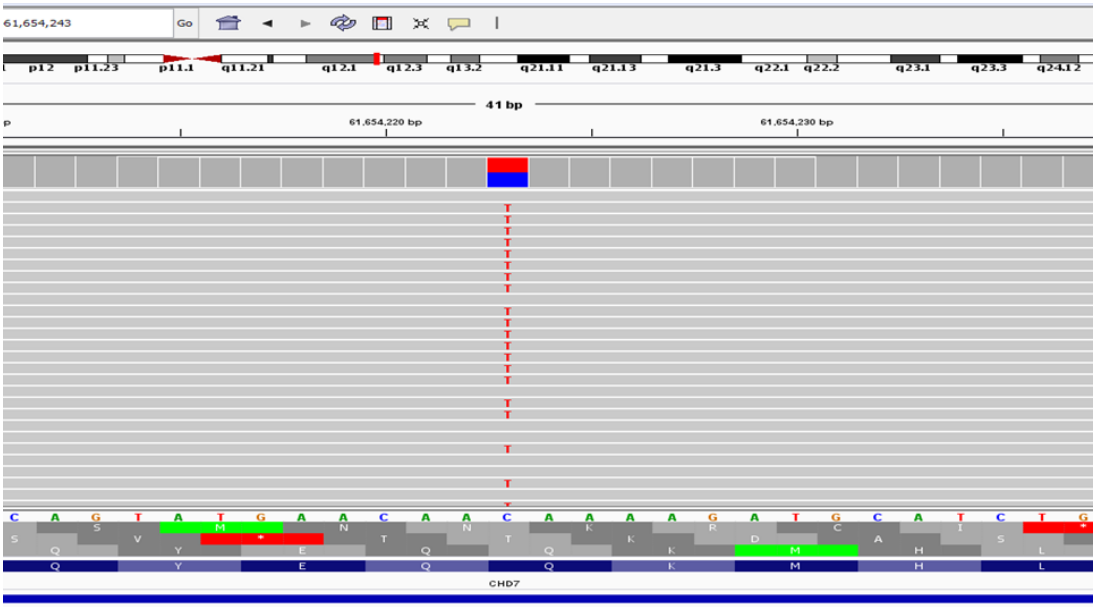

Figure 3.15: A screenshot of IGV showing the *CHD7* c.232C>T variant identified in FRASC34
The genetic variant was visualised by importing the BAM file into IGV.

(a)

Reference protein

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFSQGQEGL NQGNPFIATS GPGHLSHVPQ
 241  QSPSMAPSLR HSVQQFHHHP STALHGESVA HSPRFSPNPP QQGAVRPQTL NFSSRSQTVP
 301  SPTINNSGQY SRYPYSNLNQ GLVNNTGMNQ NLGLTNNTPM NQSVPRYPNA VGFPSNSGQG
 361  LMHQQPIHPS GSLNQMNTQT MHPSQPQGTY ASPPPMSPMK AMSNPAGTPP PQVRPGSAGI
 421  PMEVGSYPNM PHPQPSHQPP GAMGIGQRNM GPRNMQQSRP FIGMSSAPRE LTGHMRPNGC
 481  PGVGLGDPQA IQERLIPGQQ HPGQQPSFQQ LPTCPPLQPH PGLHHQSSPP HPHHQPWAQL
 541  HPSPQNTPQK VPVHQHSPSE PFLEKPVPDM TQVSGPNAQL VKSDDYLPSI EQQPQQKKKK
 601  KKNNHIVAED PSKGFGKDDF PGGVDNQELN RNSLDGSQEE KKKKKRSKAK KDPKEPKEPK
 661  EKKEPKEPKT PKAPKIPKEP KEKKAKTATP KPKSSKKSSN KKPDSEASAL KKKVNKGKTE
 721  GSENSDLDKT PPPSPPPEED EDPGVQKRRS SRQVKRKRYT EDLEFKISDE EADDADAAGR
 781  DSPSNTSQSE QQESVDAEGP VVEKIMSSRS VKKQKESGEE VEIEEFYVKY KNFSYLHCQW
 841  ASIEDLEKDK RIQQKIKRFK AKQGQNKFLS EIEDELFNPD YVEVDRIMDF ARSTDDRGEP
 901  VTHYLVKWCS LPYEDSTWER RQDIDQAKIE EFEKLMSREP ETERVERPPA DDWKKSESSR
 961  EYKNNNKLRE YQLEGVNWLL FNWYNMRNCI LADEMGLGKT IQSITFLYEI YLKGIHGPFL
1021  VIAPLSTIPN WEREFRTWTE LNVVVYHGSQ ASRRTIQLYE MYFKDPQGRV IKGSYKFHAI
1081  ITTFEMILTD CPELRNIPWR CVVIDEAHRL KNRNCKLLEG LKMMDLEHKV LLTGTPLQNT
1141  VEELFSLLHF LEPSRFPSET TFMQEFGDLK TEEQVQKLQA ILKPMMLRRL KEDVEKNLAP
1201  KEETIIEVEL TNIQKKYYRA ILEKNFTFLS KGGGQANVPN LLNTMMELRK CCNHPYLING
1261  AEEKILEEFK ETHNAESPDF QLQAMIQAAG KLVLIDKLLP KLKAGGHRVL IFSQMVRCLD
1321  ILEDYLIQRR YPYERIDGRV RGNLRQAAID RFSKPDSDRF VFLLCTRAGG LGINLTAADT
1381  CIIFDSDWNP QNDLQAQARC HRIGQSKSVK IYRLITRNSY EREMFDKASL KLGLDKAVLQ
1441  SMSGRENATN GVQQLSKKEI EDLLRKGAYG ALMDEEDEGS KFCEEDIDQI LLRRTHTITI
1501  ESEGKGSTFA KASFVASGNR TDISLDDPNF WQKWAKKAEL DIDALNGRNN LVIDTPRVRK
1561  QTRLYSAVKE DELMEFSDLE SDSEEKPCAK PRRPQDKSQG YARSECFRVE KNLLVYGWGR
1621  WTDILSHGRY KRQLTEQDVE TICRTILVYC LNHYKGDENI KSFIWDLITP TADGQTRALV
1681  NHSGLSAPVP RGRKGKKVKA QSTQPVVQDA DWLASCNPDA LFQEDSYKKH LKHHCNKVLL
1741  RVRMLYYLRQ EVIGDQADKI LEGADSSEAD VWIPEPFHAE VPADWWDKEA DKSLLIGVFK
1801  HGYEKYNSMR ADPALCFLER VGMPDAKAIA AEQRGTDMLA DGGDGGEFDR EDEDPEYKPT
```

```
1861  RTPFKDEIDE FANSPSEDKE ESMEIHATGK HSESNAELGQ LYWPNTSTLT TRLRRLITAY
1921  QRSYKRQQMR QEALMKTDRR RRRPREEVRA LEAEREAIIS EKRQKWTRRE EADFYRVVST
1981  FGVIFDPVKQ QFDWNQFRAF ARLDKKSDES LEKYFSCFVA MCRRVCRMPV KPDDEPPDLS
2041  SIIEPITEER ASRTLYRIEL LRKIREQVLH HPQLGERLKL CQPSLDLPEW WECGRHDRDL
2101  LVGAAKHGVS RTDYHILNDP ELSFLDAHKN FAQNRGAGNT SSLNPLAVGF VQTPPVISSA
2161  HIQDERVLEQ AEGKVEEPEN PAAKEKCEGK EEEEETDGSG KESKQECEAE ASSVKNELKG
2221  VEVGADTGSK SISEKGSEED EEEKLEDDDK SEESSQPEAG AVSRGKNFDE ESNASMSTAR
2281  DETRDGFYME DGDPSVAQLL HERTFAFSFW PKDRVMINRL DNICEAVLKG KWPVNRRQMF
2341  DFQGLIPGYT PTTVDSPLQK RSFAELSMVG QASISGSEDI TTSPQLSKED ALNLSVPRQR
2401  RRRRRKIEIE AERAAKRRNL MEMVAQLRES QVVSENGQEK VVDLSKASRE ATSSTSNFSS
2461  LSSKFILPNV STPVSDAFKT QMELLQAGLS RTPTRHLLNG SLVDGEPPMK RRRGRRKNVE
2521  GLDLLFMSHK RTSLSAEDAE VTKAFEEDIE TPPTRNIPSP GQLDPDTRIP VINLEDGTRL
2581  VGEDAPKNKD LVEWLKLHPT YTVDMPSYVP KNADVLFSSF QKPKQKRHRC RNPNKLDINT
2641  LTGEERVPVV NKRNGKKMGG AMAPPMKDLP RWLEENPEFA VAPDWTDIVK QSGFVPESMF
2701  DRLLTGPVVR GEGASRRGRR PKSEIARAAA AAAAVASTSG INPLLVNSLF AGMDLTSLQN
2761  LQNLQSLQLA GLMGFPPGLA TAATAGGDAK NPAAVLPLML PGMAGLPNVF GLGGLLNNPL
2821  SAATGNTTTA SSQGEPEDST SKGEEKGNEN EDENKDSEKS TDAVSAADSA NGSVGAATAP
2881  AGLPSNPLAF NPFLLSTMAP GLFYPSMFLP PGLGGLTLPG FPALAGLQNA VGSSEEKAAD
2941  KAEGGPFKDG ETLEGSDAEE SLDKTAESSL LEDEIAQGEE LDSLDGGDEI ENNENDE*
```

(b)

Protein predicted from variant coding sequence

```
 1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
61  QNQTKLTHFD HYNQYEQ*
```

Figure 3.16: The *CHD7* c.232C>T resultant predicted protein

The *CHD7* c.232C>T introduces a stop codon at amino acid 78 into the protein. Note (**a**) The normal CHD7 protein reference sequence (**b**) Protein predicted to result from the CHD7 c.232C>T variant. The amino acid sequences were obtained from Mutalyzer.

3.5.10 FRASC35

The patient is a six-year old female. She is the fourth child of her non-consanguineous parents. The patient presented with malar flattening, choanae atresia, hearing loss, sensorineural and mild intellectual disability. Diagnoses of CHARGE syndrome and TCS were considered in this patient. She doesn't have a family history of either CHARGE syndrome or TCS and her photographs were not available for publication. Twenty (20) variants were called in this patient. Of the 20 variants called, 15 were intronic, two were missense and two synonymous and one was a frameshift deletion. Following variant filtering and prioritisation, one candidate variant was identified (Table 3.13).

Table 3.13: Candidate variant identified in FRASC35

Gene	*EFTUD2*
HGVS nomenclature[1]	c.1608-55_1608-53delAAA
Protein change	-
Function	Intronic
MAF[2]	0.027
MAF[3]	0.000
CADD Scaled score[4]	-
DANN	-
FATHMM	-
GERP++	0.220 (Non-conserved)
HSF	NSV
Mutation Assessor	-
Mutation Taster	-
MutPred Splicing	-
Polyphen-2	-
PROVEAN	-
SIFT	-
VEST3	-
ClinVar variation ID	-
ClinVar classification	-
ACMG codes	BS1
ACMG classification	VUS

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014). **Abbreviations:** &, and; -, Not Available; A, Adenine; CADD, Combined annotation-dependent depletion; DANN, Deleterious Annotation of genetic variants using Neutral Networks; *EFTUD2*, Elongation factor Tu GTP binding domain containing 2; FATHMM, Functional Analysis through Hidden Markov Models; GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; NSV, Non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; SIFT, Sorting Intolerant From Tolerant; VUS; variant of unknown significance

The heterozygous *EFTUD2* c.1608-55_1608-53delAAA variant identified in this patient is an intronic variant reported in dbSNP (rs753601433). The deletion lies in a homopolymer region. Homopolymer regions are regions which include repetition of the same nucleotide and these repetitions could make the region susceptible to DNA replication errors. Furthermore, the variant has been observed in 1398 apparently healthy individuals across eight inter-continental populations in the gnomAD Exomes database (Karczewski *et al*., 2019; Fu *et al*., 2013), suggesting that it is unlikely to cause disease (Kopanos *et al*., 2018). The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG code BS1, the variant was a VUS (Richards *et a*l., 2015).

3.5.11 FRASC52

The patient is a deceased female. The patient passed on when she was five-years old and due to unknown causes. At the time of initial consultation, the patient presented with microtia, external canal atresia, mandibular hypoplasia and choanae atresia. These findings suggested a clinical diagnosis of TCS. The patient's father and sibling show features in keeping with TCS (Figure 3.17).

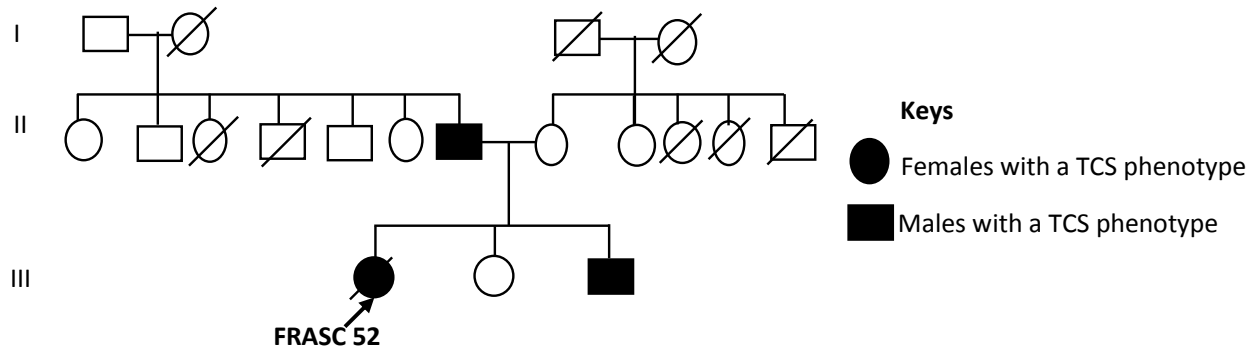

Figure 3.17: Pedigree of FRASC52
The patient is the third born to her non-consanguineous parents. Her father and brother also presented with clinical features suggestive of TCS.

Variants calling yielded 41 variants in this patient. Of the 41 variants called, 29 were intronic, three were missense, eight were synonymous and one was a frameshift deletion. Following

variant filtering, two candidate variants were identified (Table 3.14). Following variant filtering and prioritisation, one candidate variant was identified.

Table 3.14: Candidate variant identified in FRASC52

Gene	*POLR1D*
HGVS nomenclature[1]	c.26+10G>A
Protein change	-
Function	Intronic
MAF[2]	-
MAF[3]	-
CADD Scaled score[4]	-
DANN	0.968 (Pathogenic)
FATHMM	-
GERP++	-3.540 Non-conserved
HSF	NSV
Mutation Assessor	-
Mutation Taster	-
MutPred Splicing	NSV
PolyPhen-2	-
PROVEAN	-
SIFT	-
VEST3	-
ClinVar variation ID	-
ClinVar classification	-
ACMG codes	PM2
ACMG classification	VUS

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4]Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, not available; A, adenine; CADD, Combined annotation-dependent depletion; CHD7, Chromodomain Helicase DNA Binding Protein 7; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, guanine; GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; MAF, Minor Allele Frequency; -, no prediction; NSV, Non-splice variant; *POLR1D*, RNA polymerase I and III subunit D; PolyPhen-2, Polymorphism Phenotyping v2; SIFT, Sorting Intolerant From Tolerant; VUS, Variant of Unknown Significance.

The *POLR1D* c.26+10G>A is an intronic variant which is not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The patient is a heterozygote for this variant. The variant lies in a non-conserved region, 10 nucleotides away from the splice site and it is not predicted to affect splicing. Computational analysis of this variant yields a pathogenic prediction from DANN (Quang *et al*., 2014). The variant is unlikely to cause disease as it is neither predicted to affect

splicing nor situated in regulatory regions. Based on the ACMG code PM2, the variant was a VUS (Richards *et al*., 2015).

3.5.12 FRASC54

The patient is a seven-year old male. He is a second child to a non-consanguineous marriage. The patient presented with clinical features suggestive of CHARGE syndrome based on ocular coloboma, low set ears, clinodactyly, mild to moderate intellectual disability, myopia, strabismus, renal anomalies and undescended testes. The patient doesn't have a family history of CHARGE syndrome and his images were not available for publication. Variant calling yielded 36 variants in this patient. Of the 36 variants called, 24 were intronic, six synonymous, four were missense and two were frameshift variants. Following variant filtering, two candidate variants were identified (Table 3.15).

Table 3.15: Candidate variants identified in FRASC54

Gene	*CHD7*	*CHD7*
HGVS nomenclature[1]	c.2124T>C	c.1931delA
Protein change	p.Ser708=	p.Lys644ArgfsTer67
Function	Exonic	Exonic
MAF[2]	0.013	-
MAF[3]	0.017	-
CADD Scaled score[4]	-	-
DANN	-	-
FATHMM	-	-
GERP++	3.110 (Non-conserved)	5.520 (conserved)
HSF	NSV	-
Mutation Assessor	-	-
Mutation Taster	-	-
MutPred Splicing	NSV	-
PolyPhen-2	-	-
PROVEAN	-	-
SIFT	-	-
SIFT Indel	-	NMD
VEST3	-	-
ClinVar variation ID	95777	-
ClinVar classification	Likely benign	-
ACMG codes	BA1, BP6 & BP7	PSV1 & PM2
ACMG classification	Likely benign	Likely pathogenic

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, Not Available; A, Adenine; Arg, Arginine; C, Cytosine; CADD, Combined annotation-dependent depletion; CHD7, Chromodomain Helicase DNA Binding Protein 7; DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; GERP++, Genomic Evolutionary Rate Profiling; HGVS, Human Genome Variation Society; HSF, Human Splice Finder; Lys, Lysine; MAF, Minor Allele Frequency; NMD, nonsense-mediated decay; NSV, non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; Ser, Serine; SIFT, Sorting Intolerant From Tolerant; T, Thymine; Ter, termination; VUS, Variant of Unknown Significance.

The heterozygous *CHD7* c.2124T>C (p.Ser708=) variant identified in this patient is a synonymous variant reported in dbSNP (rs79302359). The variant lies in a non-conserved and its analysis on MutPred Splicing (Mort *et al*., 2014) predicts that the variant doesn't affect splicing. The variant has been observed in 843 and 485 apparently healthy African individuals in gnomAD Exomes and Genomes databases, respectively. In addition, the variant has an allele frequency of 0.06 in gnomAD African dataset and there are 15256 observed alleles in this position (Karczewski *et al*., 2019). The ACMG code BA1 therefore applies (Ghosh *et al*., 2018).

Six sources classify this variant as benign and two classify it as likely benign in ClinVar (Landerum *et al*., 2014), therefore, the ACMG code BP6 applies (Richards *et al.,* 2015). Synonymous variants could possibly cause disease through disruption of splicing or altering gene expression (Sauna *et al*., 2001), but the *CHD7* c.2124T>C is predicted to not affect splicing; therefore, the variant is unlikely to cause disease. Based on the ACMG codes BP6, BP7 and BA1, the variant was benign (Richards *et al*., 2015).

The *CHD7* c.1931delA (p.Lys644ArgfsTer67) variant is absent from gnomAD datasets despite good coverage (Fu *et al*., 2013: Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The variant lies in a conserved region across species. The deletion introduces a premature stop codon at position 67 downstream (Figure 3.18). The ACMG code PSV1 therefore applies (Richards *et al*., 2015). The deletion is located in a homopolymer region and its analysis on SIFT indel (Hu *et al*., 2013), reveals that its pre-mRNA transcript is a target of NMD (Hu and Ng, 2012). Based on the ACMG codes PSV1 and PM2, the variant was likely pathogenic (Richards *et al*., 2015). Figure 3.19 below shows an image of the *CHD7* c.1931delA variant as visualised on IGV.

(a)

Reference protein

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFSQGQEGL NQGNPFIATS GPGHLSHVPQ
 241  QSPSMAPSLR HSVQQFHHHP STALHGESVA HSPRFSPNPP QQGAVRPQTL NFSSRSQTVP
 301  SPTINNSGQY SRYPYSNLNQ GLVNNTGMNQ NLGLTNNTPM NQSVPRYPNA VGFPSNSGQG
 361  LMHQQPIHPS GSLNQMNTQT MHPSQPQGTY ASPPPMSPMK AMSNPAGTPP PQVRPGSAGI
 421  PMEVGSYPNM PHPQPSHQPP GAMGIGQRNM GPRNMQQSRP FIGMSSAPRE LTGHMRPNGC
 481  PGVGLGDPQA IQERLIPGQQ HPGQQPSFQQ LPTCPPLQPH PGLHHQSSPP HPHHQPWAQL
 541  HPSPQNTPQK VPVHQHSPSE PFLEKPVPDM TQVSGPNAQL VKSDDYLPSI EQQPQQKKKK
 601  KKNNHIVAED PSKGFGKDDF PGGVDNQELN RNSLDGSQEE KKKKKRSKAK KDPKEPKEPK
 661  EKKEPKEPKT PKAPKIPKEP KEKKAKTATP KPKSSKKSSN KKPDSEASAL KKKVNKGKTE
 721  GSENSDLDKT PPPSPPPEED EDPGVQKRRS SRQVKRKRYT EDLEFKISDE EADDADAAGR
 781  DSPSNTSQSE QQESVDAEGP VVEKIMSSRS VKKQKESGEE VEIEEFYVKY KNFSYLHCQW
 841  ASIEDLEKDK RIQQKIKRFK AKQGQNKFLS EIEDELFNPD YVEVDRIMDF ARSTDDRGEP
 901  VTHYLVKWCS LPYEDSTWER RQDIDQAKIE EFEKLMSREP ETERVERPPA DDWKKSESSR
 961  EYKNNNKLRE YQLEGVNWLL FNWYNMRNCI LADEMGLGKT IQSITFLYEI YLKGIHGPFL
1021  VIAPLSTIPN WEREFRTWTE LNVVVYHGSQ ASRRTIQLYE MYFKDPQGRV IKGSYKFHAI
1081  ITTFEMILTD CPELRNIPWR CVVIDEAHRL KNRNCKLLEG LKMMDLEHKV LLTGTPLQNT
1141  VEELFSLLHF LEPSRFPSET TFMQEFGDLK TEEQVQKLQA ILKPMMLRRL KEDVEKNLAP
1201  KEETIIEVEL TNIQKKYYRA ILEKNFTFLS KGGGQANVPN LLNTMMELRK CCNHPYLING
1261  AEEKILEEFK ETHNAESPDF QLQAMIQAAG KLVLIDKLLP KLKAGGHRVL IFSQMVRCLD
1321  ILEDYLIQRR YPYERIDGRV RGNLRQAAID RFSKPDSDRF VFLLCTRAGG LGINLTAADT
1381  CIIFDSDWNP QNDLQAQARC HRIGQSKSVK IYRLITRNSY EREMFDKASL KLGLDKAVLQ
1441  SMSGRENATN GVQQLSKKEI EDLLRKGAYG ALMDEEDEGS KFCEEDIDQI LLRRTHTITI
1501  ESEGKGSTFA KASFVASGNR TDISLDDPNF WQKWAKKAEL DIDALNGRNN LVIDTPRVRK
1561  QTRLYSAVKE DELMEFSDLE SDSEEKPCAK PRRPQDKSQG YARSECFRVE KNLLVYGWGR
1621  WTDILSHGRY KRQLTEQDVE TICRTILVYC LNHYKGDENI KSFIWDLITP TADGQTRALV
1681  NHSGLSAPVP RGRKGKKVKA QSTQPVVQDA DWLASCNPDA LFQEDSYKKH LKHHCNKVLL
1741  RVRMLYYLRQ EVIGDQADKI LEGADSSEAD VWIPEPFHAE VPADWWDKEA DKSLLIGVFK
1801  HGYEKYNSMR ADPALCFLER VGMPDAKAIA AEQRGTDMLA DGGDGGEFDR EDEDPEYKPT
1861  RTPFKDEIDE FANSPSEDKE ESMEIHATGK HSESNAELGQ LYWPNTSTLT TRLRRLITAY
1921  QRSYKRQQMR QEALMKTDRR RRRPREEVRA LEAEREAIIS EKRQKWTRRE EADFYRVVST
1981  FGVIFDPVKQ QFDWNQFRAF ARLDKKSDES LEKYFSCFVA MCRRVCRMPV KPDDEPPDLS
2041  SIIEPITEER ASRTLYRIEL LRKIREQVLH HPQLGERLKL CQPSLDLPEW WECGRHDRDL
2101  LVGAAKHGVS RTDYHILNDP ELSFLDAHKN FAQNRGAGNT SSLNPLAVGF VQTPPVISSA
2161  HIQDERVLEQ AEGKVEEPEN PAAKEKCEGK EEEEETDGSG KESKQECEAE ASSVKNELKG
2221  VEVGADTGSK SISEKGSEED EEEKLEDDDK SEESSQPEAG AVSRGKNFDE ESNASMSTAR
2281  DETRDGFYME DGDPSVAQLL HERTFAFSFW PKDRVMINRL DNICEAVLKG KWPVNRRQMF
2341  DFQGLIPGYT PTTVDSPLQK RSFAELSMVG QASISGSEDI TTSPQLSKED ALNLSVPRQR
2401  RRRRRKIEIE AERAAKRRNL MEMVAQLRES QVVSENGQEK VVDLSKASRE ATSSTSNFSS
2461  LSSKFILPNV STPVSDAFKT QMELLQAGLS RTPTRHLLNG SLVDGEPPMK RRRGRRKNVE
2521  GLDLLFMSHK RTSLSAEDAE VTKAFEEDIE TPPTRNIPSP GQLDPDTRIP VINLEDGTRL
2581  VGEDAPKNKD LVEWLKLHPT YTVDMPSYVP KNADVLFSSF QKPKQKRHRC RNPNKLDINT
2641  LTGEERVPVV NKRNGKKMGG AMAPPMKDLP RWLEENPEFA VAPDWTDIVK QSGFVPESMF
2701  DRLLTGPVVR GEGASRRGRR PKSEIARAAA AAAAVASTSG INPLLVNSLF AGMDLTSLQN
2761  LQNLQSLQLA GLMGFPPGLA TAATAGGDAK NPAAVLPLML PGMAGLPNVF GLGGLLNNPL
2821  SAATGNTTTA SSQGEPEDST SKGEEKGNEN EDENKDSEKS TDAVSAADSA NGSVGAATAP
2881  AGLPSNPLAF NPFLLSTMAP GLFYPSMFLP PGLGGLTLPG FPALAGLQNA VGSSEEKAAD
2941  KAEGGPFKDG ETLEGSDAEE SLDKTAESSL LEDEIAQGEE LDSLDGGDEI ENNENDE*
```

(b)

Protein predicted from variant coding sequence

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFSQGQEGL NQGNPFIATS GPGHLSHVPQ
 241  QSPSMAPSLR HSVQQFHHHP STALHGESVA HSPRFSPNPP QQGAVRPQTL NFSSRSQTVP
 301  SPTINNSGQY SRYPYSNLNQ GLVNNTGMNQ NLGLTNNTPM NQSVPRYPNA VGFPSNSGQG
 361  LMHQQPIHPS GSLNQMNTQT MHPSQPQGTY ASPPPMSPMK AMSNPAGTPP PQVRPGSAGI
 421  PMEVGSYPNM PHPQPSHQPP GAMGIGQRNM GPRNMQQSRP FIGMSSAPRE LTGHMRPNGC
 481  PGVGLGDPQA IQERLIPGQQ HPGQQPSFQQ LPTCPPLQPH PGLHHQSSPP HPHHQPWAQL
 541  HPSPQNTPQK VPVHQHSPSE PFLEKPVPDM TQVSGPNAQL VKSDDYLPSI EQQPQQKKKK
 601  KKNNHIVAED PSKGFGKDDF PGGVDNQELN RNSLDGSQEE KKKRKGQRQK KTRRNRKNPR
 661  RKKSPRNPRP RKPLRFPKSQ RKRKQKLPRQ NPNPAKSQVI RNLTQKQVL*
```

Figure 3.18: The *CHD7* c.1931delA resultant predicted protein

The *CHD7* c.3708delC introduces a premature stop codon at amino acid 644 into the CHD7 protein. Note (**a**) The normal CHD7 protein reference sequence (**b**) and protein predicted from the *CHD7* c.1931delA variant.

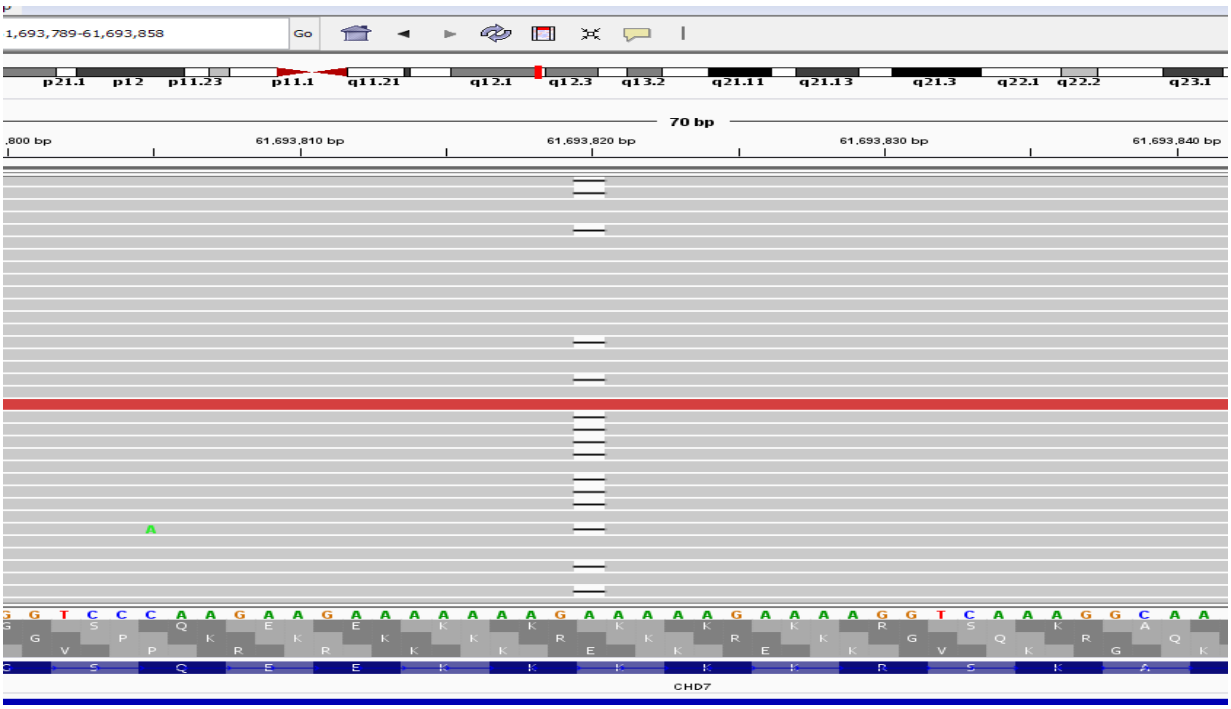

Figure 3.19: A screenshot of IGV showing the *CHD7* c.1931delA identified in FRASC54
The genetic variant was visualised by importing the BAM file into IGV.

3.5.13 FRASC57

The patient is a one-year old male. He is the second child of a non-consanguineous marriage and presented with clinical features suggestive of TCS based on malar flattening, mandibular hypoplasia, lower-lid, coloboma, external canal atresia, lateral cheek hair projection, microtia and gastro-oesophageal reflux. He doesn't have a family history of TCS and his photographs were not available for publication. Variant calling yielded 30 variants in this patient. Of the 30 variants called, 19 were intronic, two were missense, seven were synonymous and two were frameshift deletions. Following variant filtering, three candidate variants were identified (Table 3.16).

Table 3.16: Candidate variants identified in FRASC57

Gene	*TCOF1*	*POLR1D*	*TCOF1*
HGVS nomenclature[1]	c.4455G>A	c.26+18G>A	c.3708delC
Protein change	p.Glu1485=	-	p.Ala1237Pro
Function	Exonic	Intronic	Exonic
MAF[2]	0.002	0.000	-
MAF[3]	0.008	0.001	-
DANN	0.833	0.969 (Pathogenic)	-
FATHMM	-	-	-
CADD Scaled score[4]	-	-	-
GERP++	4.100 (Conserved)	1.710 (Non-conserved)	-
HSF	NSV	NSV	-
Mutation Assessor	-	-	-
Mutation Taster	-	-	-
MutPred Splicing	NSV	NSV	-
PolyPhen-2	-	-	-
PROVEAN	-	-	-
SIFT	-	-	-
SIFT Indel	-	-	NMD
VEST3	-	-	-
ClinVar variation ID	352237	-	-
ClinVar classification	Likely benign	-	-
ACMG codes	BS1 & BP6	BS1 & PP3	PSV1 & PM2
ACMG classification	Likely benign	VUS	Likely pathogenic

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, Not Available; A, adenine; Ala, Alanine; C, Cytosine CADD, Combined annotation-dependent depletion, DANN, Deleterious Annotation of genetic variants using Neutral Networks, FATHMM, Functional Analysis through Hidden Markov Models, G, Guanine, GERP++, Genomic Evolutionary Rate Profiling, HGVS, Human Genome Variation Society; Glu, Glutamine; HSF, Human Splice Finder , MAF, Minor Allele Frequency, NMD, nonsense-mediated decay; NSV, non-splice variant; *POLR1D*, RNA polymerase I and III subunit D, PolyPhen-2, Polymorphism Phenotyping v2, Pro, Proline; SIFT, Sorting Intolerant From Tolerant, *TCOF1*, Treacle ribosome biogenesis factor 1, VUS, Variant of Unknown Significance.

The heterozygous T*COF1* c.4455G>A (p.Glu1485=) variant identified in this patient is a synonymous variant reported in polymorphism dbSNP (rs116268092). The variant is 15 bases away from the splice site and is not predicted to impact on splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). Synonymous variants could possibly cause disease through disruption of splicing or altering gene expression (Sauna *et al*., 2001), but the *TCOF1* c.4455G>A variant neither disrupts splicing nor is it predicted to alter gene expression. The ACMG code BP7 is usually applied. However, in this case, the variant lies in a conserved region, therefore, BP7 cannot be applied. The variant has been observed in 449 and 229 apparently healthy African individuals in the gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019), suggesting that it is unlikely to cause disease. The ACMG code BS1 therefore applies (Richards *et al*., 2015). The variant is classified as likely benign in ClinVar (Landerum *et al*., 2014); therefore, the ACMG code BP6 applies (Richards *et al*., 2015). Based on the ACMG codes BP6 and BS1, the variant was classified as likely benign (Richards *et al*., 2015).

The *POLR1D* c.26+18G>A is an intronic variant reported in dbSNP (rs2232681). The variant lies in a non-conserved region, 18 nucleotides away from the splice site and isn't predicted to affect splicing. Although the variant is predicted to not affect splicing, the variant is not synonymous: therefore, the ACMG code BP7 cannot be applied (Richards *et al*., 2015). A single computational tool (DANN), yield a pathogenic classification and no benign prediction is yielded. The ACMG code PP3 therefore applies (Richards *et al*., 2015). Intronic variants could possibly cause disease through disruption of splicing but the *POLR1D* c.26+10G>A is predicted to neither affect splicing nor is situated in regulatory regions. In addition, the variant has been observed in 21 and 13 apparently healthy African individuals in gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019). The ACMG code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes PP3 and BS1, the variant was a VUS (Richards *et al*., 2015).

The heterozygous *TCOF1* c.3708delC (p.Ala1237Pro) variant identified in this patient is not observed in either of the gnomAD datasets despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The deletion of cytosine in this

position introduces a premature stop codon at position 70 downstream resulting in a truncated protein (see Figure 3.21). Therefore, the ACMG code PSV1 applies (Richards *et al*., 2015). Analysis of this variant by the SIFT indel predicts that the variant's transcript is a target of NMD. An Application of the ACMG codes PSV1 and PM2 yields a likely pathogenic classification (Richards *et al*., 2015). Figure 3.20 shows an image of the *TCOF1* c.3708delC variant as visualised on IGV.

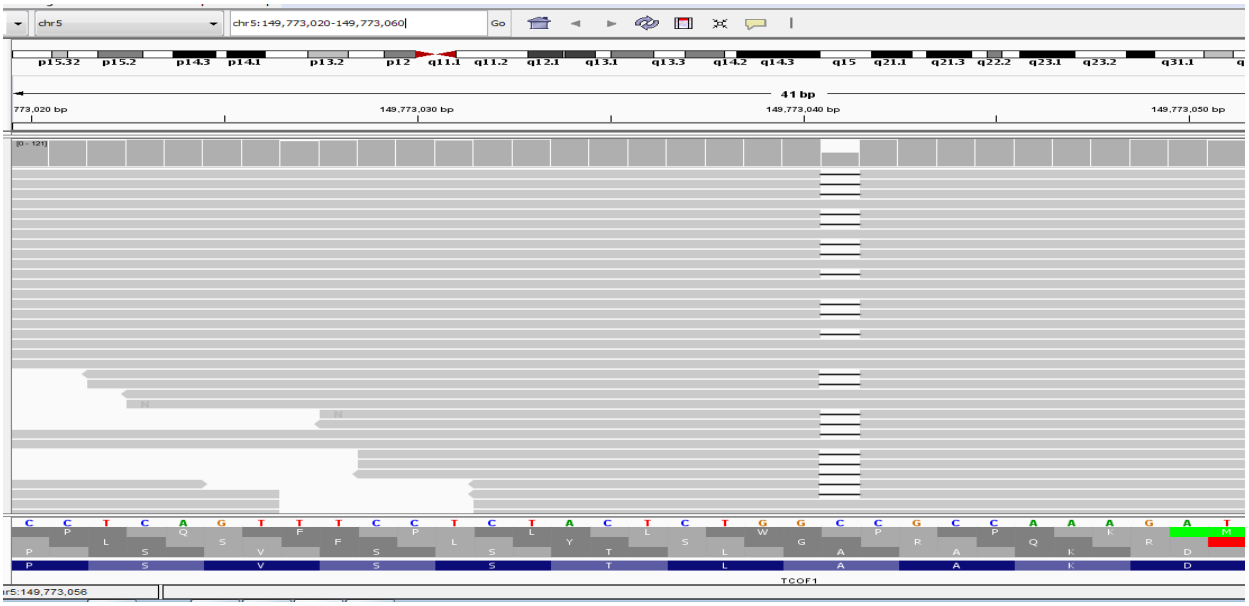

Figure 3.20:A screenshot of IGV showing the *TCOF1* c.3708delC variant identified in FRASC57
The genetic variant was visualised by importing the BAM file data into IGV.

(a)

Reference protein

```
   1  MAEARKRREL LPLIYHHLLR AGYVRAAREV KEQSGQKCFL AQPVTLLDIY THWQQTSELG
  61  RKRKAEEDAA LQAKKTRVSD PISTSESSEE EEEAEAETAK ATPRLASTNS SVLGADLPSS
 121  MKEKAKAETE KAGKTGNSMP HPATGKTVAN LLSGKSPRKS AEPSANTTLV SETEEEGSVP
 181  AFGAAAKPGM VSAGQADSSS EDTSSSSDET DVEGKPSVKP AQVKASSVST KESPARKAAP
 241  APGKVGDVTP QVKGGALPPA KRAKKPEEES ESSEEGSESE EEAPAGTRSQ VKASEKILQV
 301  RAASAPAKGT PGKGATPAPP GKAGAVASQT KAGKPEEDSE SSSEESSDSE EETPAAKALL
 361  QAKASGKTSQ VGAASAPAKE SPRKGAAPAP PGKTGPAVAK AQAGKREEDS QSSSEESDSE
 421  EEAPAQAKPS GKAPQVRAAS APAKESPRKG AAPAPPRKTG PAAAQVQVGK QEEDSRSSSE
 481  ESDSDREALA AMNAAQVKPL GKSPQVKPAS TMGMGPLGKG AGPVPPGKVG PATPSAQVGK
 541  WEEDSESSSE ESSDSSDGEV PTAVAPAQEK SLGNILQAKP TSSPAKGPPQ KAGPVAVQVK
 601  AEKPMDNSES SEESSDSADS EEAPAAMTAA QAKPALKIPQ TKACPKKTNT TASAKVAPVR
 661  VGTQAPRKAG TATSPAGSSP AVAGGTQRPA EDSSSSEESD SEEEKTGLAV TVGQAKSVGK
 721  GLQVKAASVP VKGSLGQGTA PVLPGKTGPT VTQVKAEKQE DSESSEEESD SEEAAASPAQ
 781  VKTSVKKTQA KANPAAARAP SAKGTISAPG KVVTAAAQAK QRSPSKVKPP VRNPQNSTVL
 841  ARGPASVPSV GKAVATAAQA QTGPEEDSGS SEEESDSEEE AETLAQVKPS GKTHQIRAAL
 901  APAKESPRKG AAPTPPGKTG PSAAQAGKQD DSGSSSEESD SDGEAPAAVT SAQVIKPPLI
 961  FVDPNRSPAG PAATPAQAQA ASTPRKARAS ESTARSSSSE SEDEDVIPAT QCLTPGIRTN
1021  VVTMPTAHPR IAPKASMAGA SSSKESSRIS DGKKQEGPAT QVSKKNPASL PLTQAALKVL
1081  AQKASEAQPP VARTQPSSGV DSAVGTLPAT SPQSTSVQAK GTNKLRKPKL PEVQQATKAP
1141  ESSDDSEDSS DSSSGSEEDG EGPQGAKSAH TLGPTPSRTE TLVEETAAES SEDDVVAPSQ
1201  SLLSGYMTPG LTPANSQASK ATPKLDSSPS VSSTLAAKDD PDGKQEAKPQ QAAGMLSPKT
1261  GGKEAASGTT PQKSRKPKKG AGNPQASTLA LQSNITQCLL GQPWPLNEAQ VQASVVKVLT
1321  ELLEQERKKV VDTTKESSRK GWESRKRKLS GDQPAARTPR SKKKKKLGAG EGGEASVSPE
1381  KTSTTSKGKA KRDKASGDVK EKKGKGSLGS QGAKDEPEEE LQKGMGTVEG GDQSNPKSKK
1441  EKKKSDKRKK DKEKKEKKKK AKKASTKDSE SPSQKKKKKK KKTAEQTV*
```

(b)

Protein predicted from variant coding sequence

```
   1  MAEARKRREL LPLIYHHLLR AGYVRAAREV KEQSGQKCFL AQPVTLLDIY THWQQTSELG
  61  RKRKAEEDAA LQAKKTRVSD PISTSESSEE EEEAEAETAK ATPRLASTNS SVLGADLPSS
 121  MKEKAKAETE KAGKTGNSMP HPATGKTVAN LLSGKSPRKS AEPSANTTLV SETEEEGSVP
 181  AFGAAAKPGM VSAGQADSSS EDTSSSSDET DVEGKPSVKP AQVKASSVST KESPARKAAP
 241  APGKVGDVTP QVKGGALPPA KRAKKPEEES ESSEEGSESE EEAPAGTRSQ VKASEKILQV
 301  RAASAPAKGT PGKGATPAPP GKAGAVASQT KAGKPEEDSE SSSEESSDSE EETPAAKALL
 361  QAKASGKTSQ VGAASAPAKE SPRKGAAPAP PGKTGPAVAK AQAGKREEDS QSSSEESDSE
 421  EEAPAQAKPS GKAPQVRAAS APAKESPRKG AAPAPPRKTG PAAAQVQVGK QEEDSRSSSE
 481  ESDSDREALA AMNAAQVKPL GKSPQVKPAS TMGMGPLGKG AGPVPPGKVG PATPSAQVGK
 541  WEEDSESSSE ESSDSSDGEV PTAVAPAQEK SLGNILQAKP TSSPAKGPPQ KAGPVAVQVK
 601  AEKPMDNSES SEESSDSADS EEAPAAMTAA QAKPALKIPQ TKACPKKTNT TASAKVAPVR
 661  VGTQAPRKAG TATSPAGSSP AVAGGTQRPA EDSSSSEESD SEEEKTGLAV TVGQAKSVGK
 721  GLQVKAASVP VKGSLGQGTA PVLPGKTGPT VTQVKAEKQE DSESSEEESD SEEAAASPAQ
 781  VKTSVKKTQA KANPAAARAP SAKGTISAPG KVVTAAAQAK QRSPSKVKPP VRNPQNSTVL
 841  ARGPASVPSV GKAVATAAQA QTGPEEDSGS SEEESDSEEE AETLAQVKPS GKTHQIRAAL
 901  APAKESPRKG AAPTPPGKTG PSAAQAGKQD DSGSSSEESD SDGEAPAAVT SAQVIKPPLI
 961  FVDPNRSPAG PAATPAQAQA ASTPRKARAS ESTARSSSSE SEDEDVIPAT QCLTPGIRTN
1021  VVTMPTAHPR IAPKASMAGA SSSKESSRIS DGKKQEGPAT QVSKKNPASL PLTQAALKVL
1081  AQKASEAQPP VARTQPSSGV DSAVGTLPAT SPQSTSVQAK GTNKLRKPKL PEVQQATKAP
1141  ESSDDSEDSS DSSSGSEEDG EGPQGAKSAH TLGPTPSRTE TLVEETAAES SEDDVVAPSQ
1201  SLLSGYMTPG LTPANSQASK ATPKLDSSPS VSSTLAPKMT QMASRRQSPN RQQACCPLKQ
1261  VEKRLLQAPH LRSPGSPRKG LGTPKPQPWR CKATSPSASW ANPGP*
```

Figure 3.21: The *TCOF1* c.3708delC resultant predicted protein
The *TCOF1* c.3708delC introduces a premature stop codon at amino acid 1236 into the Treacle protein. Note (**a**) normal Treacle protein reference sequence (**b**) truncated predicted protein from the c.3708delC variant.

3.5.14 FRASC59

The patient is a five-year old female. She is the only child to her non-consanguineous parents. The patient presented with clinical features suggestive of CHARGE syndrome based on malar flattening, prominent low set ears, ocular coloboma, kyphosis, strabismus and mild conductive hearing loss. In addition, she has a severe intellectual disability, cardiac anomalies and bilateral small kidneys. There is no family history of CHARGE syndrome and her photographs were not available for publication. Variant calling yielded 30 variants in this patient. Of the 30 variants called, 17 were intronic, four were missense, seven were synonymous and two were frameshift deletions. Following variant filtering, three candidate variants were identified (Table 3.17).

Table 3.17: Candidate variants identified in FRASC59

Gene	*CHD7*	*CHD7*	*CHD7*
HGVS[1]	c.2829G>A	c.8322C>G	c.3309_3310delCA
Protein change	p.Glu943=	p.Gly2774=	p.Ala1237ProfsTer70
Function	Exonic	Exonic	Exonic
MAF[2]	0.000	-	0.000
MAF[3]	0.001	-	0.000
DANN	0.829 (Benign)	0.766 (Benign)	-
FATHMM	-	-	-
CADD Scaled score[4]	-	-	-
GERP++	3.450 (non-conserved)	3.510 (Non-conserved)	5.280 (conserved)
HSF	NSV	NSV	-
Mutation Assessor	-	-	-
Mutation Taster	-	-	-
MutPred Splicing	NSV	NSV	-
PolyPhen-2	-	-	-
PROVEAN	-	-	-
SIFT	-	-	-
SIFT Indel	-	-	NMD
VEST3	-	-	-
ClinVar variation ID	NA	260916	-
ClinVar classification	-	Conflicting classification	-
ACMG codes	BS1, BP7 & PP3	BS1 & BP7	PSV1, PM1 & PM2
ACMG classification	Likely benign	Likely benign	Pathogenic

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher et al., 2014).
Abbreviations: &, and; -, not available; A, adenine; Ala, Alanine; C, cytosine; CADD, Combined annotation-dependent depletion; *CHD7*,Chromodomain helicase DNA binding protein 7 DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, Guanine; GERP++, Genomic Evolutionary Rate Profiling; Glu, Glutamine; Gly, Glycine; HGVS, Human Genome Variation Society; HSF, Human Splice Finder ; MAF, Minor Allele Frequency; NMD, nonsense-mediated decay; NSV; non-splice variant PolyPhen-2, Polymorphism Phenotyping v2; Pro, Proline; SIFT, Sorting Intolerant From Tolerant; Ter, termination; VUS, variant of unknown significance.

The heterozygous *CHD7* c.2829G>A (p.Glu943=) variant identified in this patient is a synonymous variant recorded in dbSNP (rs374877439). The variant lies in a non-conserved region and is predicted to not affect splicing (Jian *et a*l., 2014; Mort *et al*., 2014). The ACMG code BP7 therefore applies (Richards *et al*., 2015). A single computational tool (GERP++) yields a pathogenic verdict versus no benign predictions. Therefore, the ACMG PP3 applies (Richards *et al*., 2015). In addition, the variant has been observed in 23 apparently healthy African individuals, suggesting that it is unlikely to cause disease (Kopanos *et al*., 2018). The ACMG

code BS1 therefore applies (Richards *et al*., 2015). An application of the ACMG codes BS1, BP7 and PP3 classify the c.2829G>A variant as likely benign (Richards *et al*., 2015).

The heterozygous *CHD7* c.8322C>G (p.Gly2774=) variant identified in this patient is a synonymous variant recorded in dbSNP (rs376063472). The variant lies in a non-conserved and is predicted to not alter splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). A single report classifies this variant as likely benign and two reports classify it as VUS in ClinVar (Landerum *et al*., 2014), therefore, the ACMG code BP6 cannot be applied (Richards *et a*l., 2015). Furthermore, the variant has been observed in 18 and 2 apparently healthy African individuals in the gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019). Therefore, the ACMG BS1 applies (Richards *et al*., 2015). Based on the ACMG code BS1 and BP7, the variant was classified as likely benign (Richards *et al*., 2015).

The *CHD7* c.3309_3310delCA (p.Ala1237ProfsTer70) variant is not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The variant introduces a stop codon at amino acid 1103 into the CHD7 protein (Figure 3.22); therefore, the ACMG code PSV1 applies (Richards *et al*., 2015). The variant lies in a conserved ATP binding motif of the *CHD7* protein. The ACMG code PM1 therefore applies (Richards *et al*., 2015). Analysis of this variant by the SIFT Indel (Hu and NG, 2013), predicts that the variant's transcript a target of NMD. Based on the ACMG code PSV1, PM2 and PM1, the variant is classified as pathogenic (Richards *et al*., 2015). Figure 3.23 shows the *CHD7* c.3309_3310delCA variant as visualised on IGV.

(a)

Reference protein

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFSQGQEGL NQGNPFIATS GPGHLSHVPQ
 241  QSPSMAPSLR HSVQQFHHHP STALHGESVA HSPRFSPNPP QQGAVRPQTL NFSSRSQTVP
 301  SPTINNSGQY SRYPYSNLNQ GLVNNTGMNQ NLGLTNNTPM NQSVPRYPNA VGFPSNSGQG
 361  LMHQQPIHPS GSLNQMNTQT MHPSQPQGTY ASPPPMSPMK AMSNPAGTPP PQVRPGSAGI
 421  PMEVGSYPNM PHPQPSHQPP GAMGIGQRNM GPRNMQQSRP FIGMSSAPRE LTGHMRPNGC
 481  PGVGLGDPQA IQERLIPGQQ HPGQQPSFQQ LPTCPPLQPH PGLHHQSSPP HPHHQPWAQL
 541  HPSPQNTPQK VPVHQHSPSE PFLEKPVPDM TQVSGPNAQL VKSDDYLPSI EQQPQQKKKK
 601  KKNNHIVAED PSKGFGKDDF PGGVDNQELN RNSLDGSQEE KKKKKRSKAK KDPKEPKEPK
 661  EKKEPKEPKT PKAPKIPKEP KEKKAKTATP KPKSSKKSSN KKPDSEASAL KKKVNKGKTE
 721  GSENSDLDKT PPPSPPPEED EDPGVQKRRS SRQVKRKRYT EDLEFKISDE EADDADAAGR
 781  DSPSNTSQSE QQESVDAEGP VVEKIMSSRS VKKQKESGEE VEIEEFYVKY KNFSYLHCQW
 841  ASIEDLEKDK RIQQKIKRFK AKQGQNKFLS EIEDELFNPD YVEVDRIMDF ARSTDDRGEP
 901  VTHYLVKWCS LPYEDSTWER RQDIDQAKIE EFEKLMSREP ETERVERPPA DDWKKSESSR
 961  EYKNNNKLRE YQLEGVNWLL FNWYNMRNCI LADEMGLGKT IQSITFLYEI YLKGIHGPFL
1021  VIAPLSTIPN WEREFRTWTE LNVVVYHGSQ ASRRTIQLYE MYFKDPQGRV IKGSYKFHAI
1081  ITTFEMILTD CPELRNIPWR CVVIDEAHRL KNRNCKLLEG LKMMDLEHKV LLTGTPLQNT
1141  VEELFSLLHF LEPSRFPSET TFMQEFGDLK TEEQVQKLQA ILKPMMLRRL KEDVEKNLAP
1201  KEETIIEVEL TNIQKKYYRA ILEKNFTFLS KGGGQANVPN LLNTMMELRK CCNHPYLING
1261  AEEKILEEFK ETHNAESPDF QLQAMIQAAG KLVLIDKLLP KLKAGGHRVL IFSQMVRCLD
1321  ILEDYLIQRR YPYERIDGRV RGNLRQAAID RFSKPDSDRF VFLLCTRAGG LGINLTAADT
1381  CIIFDSDWNP QNDLQAQARC HRIGQSKSVK IYRLITRNSY EREMFDKASL KLGLDKAVLQ
1441  SMSGRENATN GVQQLSKKEI EDLLRKGAYG ALMDEEDEGS KFCEEDIDQI LLRRTHTITI
1501  ESEGKGSTFA KASFVASGNR TDISLDDPNF WQKWAKKAEL DIDALNGRNN LVIDTPRVRK
1561  QTRLYSAVKE DELMEFSDLE SDSEEKPCAK PRRPQDKSQG YARSECFRVE KNLLVYGWGR
1621  WTDILSHGRY KRQLTEQDVE TICRTILVYC LNHYKGDENI KSFIWDLITP TADGQTRALV
1681  NHSGLSAPVP RGRKGKKVKA QSTQPVVQDA DWLASCNPDA LFQEDSYKKH LKHHCNKVLL
1741  RVRMLYYLRQ EVIGDQADKI LEGADSSEAD VWIPEPFHAE VPADWWDKEA DKSLLIGVFK
1801  HGYEKYNSMR ADPALCFLER VGMPDAKAIA AEQRGTDMLA DGGDGGEFDR EDEDPEYKPT
1861  RTPFKDEIDE FANSPSEDKE ESMEIHATGK HSESNAELGQ LYWPNTSTLT TRLRRLITAY
1921  QRSYKRQQMR QEALMKTDRR RRRPREEVRA LEAEREAIIS EKRQKWTRRE EADFYRVVST
1981  FGVIFDPVKQ QFDWNQFRAF ARLDKKSDES LEKYFSCFVA MCRRVCRMPV KPDDEPPDLS
2041  SIIEPITEER ASRTLYRIEL LRKIREQVLH HPQLGERLKL CQPSLDLPEW WECGRHDRDL
2101  LVGAAKHGVS RTDYHILNDP ELSFLDAHKN FAQNRGAGNT SSLNPLAVGF VQTPPVISSA
2161  HIQDERVLEQ AEGKVEEPEN PAAKEKCEGK EEEEETDGSG KESKQECEAE ASSVKNELKG
2221  VEVGADTGSK SISEKGSEED EEEKLEDDDK SEESSQPEAG AVSRGKNFDE ESNASMSTAR
2281  DETRDGFYME DGDPSVAQLL HERTFAFSFW PKDRVMINRL DNICEAVLKG KWPVNRRQMF
2341  DFQGLIPGYT PTTVDSPLQK RSFAELSMVG QASISGSEDI TTSPQLSKED ALNLSVPRQR
2401  RRRRRKIEIE AERAAKRRNL MEMVAQLRES QVVSENGQEK VVDLSKASRE ATSSTSNFSS
2461  LSSKFILPNV STPVSDAFKT QMELLQAGLS RTPTRHLLNG SLVDGEPPMK RRRGRRKNVE
2521  GLDLLFMSHK RTSLSAEDAE VTKAFEEDIE TPPTRNIPSP GQLDPDTRIP VINLEDGTRL
2581  VGEDAPKNKD LVEWLKLHPT YTVDMPSYVP KNADVLFSSF QKPKQKRHRC RNPNKLDINT
2641  LTGEERVPVV NKRNGKKMGG AMAPPMKDLP RWLEENPEFA VAPDWTDIVK QSGFVPESMF
2701  DRLLTGPVVR GEGASRRGRR PKSEIARAAA AAAAVASTSG INPLLVNSLF AGMDLTSLQN
2761  LQNLQSLQLA GLMGFPPGLA TAATAGGDAK NPAAVLPLML PGMAGLPNVF GLGGLLNNPL
2821  SAATGNTTTA SSQGEPEDST SKGEEKGNEN EDENKDSEKS TDAVSAADSA NGSVGAATAP
2881  AGLPSNPLAF NPFLLSTMAP GLFYPSMFLP PGLGGLTLPG FPALAGLQNA VGSSEEKAAD
2941  KAEGGPFKDG ETLEGSDAEE SLDKTAESSL LEDEIAQGEE LDSLDGGDEI ENNENDE*
```

(b)

Protein predicted from variant coding sequence

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFSQGQEGL NQGNPFIATS GPGHLSHVPQ
 241  QSPSMAPSLR HSVQQFHHHP STALHGESVA HSPRFSPNPP QQGAVRPQTL NFSSRSQTVP
 301  SPTINNSGQY SRYPYSNLNQ GLVNNTGMNQ NLGLTNNTPM NQSVPRYPNA VGFPSNSGQG
 361  LMHQQPIHPS GSLNQMNTQT MHPSQPQGTY ASPPPMSPMK AMSNPAGTPP PQVRPGSAGI
 421  PMEVGSYPNM PHPQPSHQPP GAMGIGQRNM GPRNMQQSRP FIGMSSAPRE LTGHMRPNGC
 481  PGVGLGDPQA IQERLIPGQQ HPGQQPSFQQ LPTCPPLQPH PGLHHQSSPP HPHHQPWAQL
 541  HPSPQNTPQK VPVHQHSPSE PFLEKPVPDM TQVSGPNAQL VKSDDYLPSI EQQPQQKKKK
 601  KKNNHIVAED PSKGFGKDDF PGGVDNQELN RNSLDGSQEE KKKKKRSKAK KDPKEPKEPK
 661  EKKEPKEPKT PKAPKIPKEP KEKKAKTATP KPKSSKKSSN KKPDSEASAL KKKVNKGKTE
 721  GSENSDLDKT PPPSPPPEED EDPGVQKRRS SRQVKRKRYT EDLEFKISDE EADDADAAGR
 781  DSPSNTSQSE QQESVDAEGP VVEKIMSSRS VKKQKESGEE VEIEEFYVKY KNFSYLHCQW
 841  ASIEDLEKDK RIQQKIKRFK AKQGQNKFLS EIEDELFNPD YVEVDRIMDF ARSTDDRGEP
 901  VTHYLVKWCS LPYEDSTWER RQDIDQAKIE EFEKLMSREP ETERVERPPA DDWKKSESSR
 961  EYKNNNKLRE YQLEGVNWLL FNWYNMRNCI LADEMGLGKT IQSITFLYEI YLKGIHGPFL
1021  VIAPLSTIPN WEREFRTWTE LNVVVYHGSQ ASRRTIQLYE MYFKDPQGRV IKGSYKFHAI
1081  ITTFEMILTD CPELRNIPWR CVV*
```

Figure 3.22:The *CHD7* c.3309_3310delCA resultant predicted protein

The *CHD7* c.3309_3310delCA introduces a stop codon at amino acid 1103 into the CHD7 protein. Note (**a)** The normal CHD7 Protein reference sequence **(b**) and the protein sequence predicted from *CHD7* c.3309_3310delCA variant.

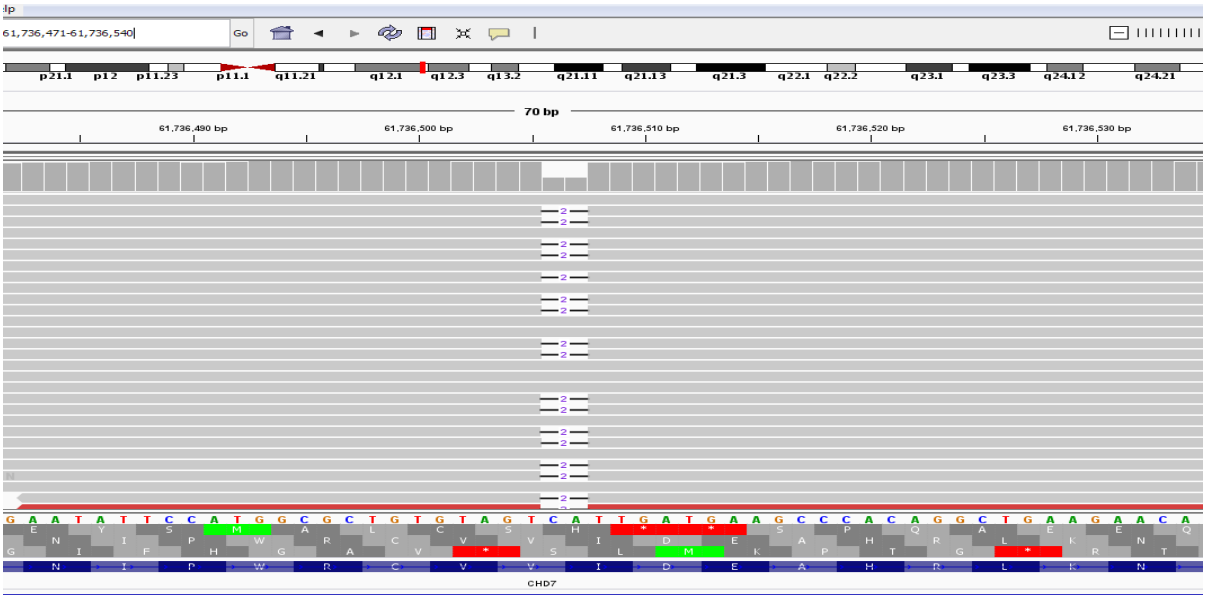

Figure 3.23: A screenshot of IGV showing the *CHD7* c.3309_3310delCA variant identified in FRASC59
The genetic variant was visualised by importing the BAM file data into IGV.

3.5.15 FRASC61

The patient is a one-year old male who presented with clinical features suggestive of CHARGE syndrome based on cleft palate, ocular coloboma, choanal atresia, cardiac anomalies, hypothyroidism, gastro-oesophageal reflux, small penis and severe intellectual disability. He has no family history of CHARGE syndrome and his images were not available for publication. Variant calling yielded 42 variants in this patient. Of the 42 variants called, 31 were intronic, six were synonymous, three were missense, one was a deletion and one was a nonsense variant. Following variant filtering, three candidate variants were identified (Table 3.18).

Table 3.18: Candidate variants identified in FRASC61

Gene	*CHD7*	*CHD7*	*CHD7*
HGVS[1]	c.7278G>A	c.5307C>T	c.643C>T
Protein change	p.Gln2426=	p.Ala1769=	p.Gln215Ter
Function	Exonic	Exonic	Exonic
MAF[2]	0.000	0.004	-
MAF[3]	0.001	0.018	-
DANN	0.732 (Benign)	0.454 (Benign)	0.998 (Pathogenic)
FATHMM	-	-	0.928 (Damaging)
CADD Scaled score[4]	-	-	-
SIFT	-	-	-
PolyPhen-2	-	-	-
GERP++	3.310 (Benign)	-8.690 (Non-conserved)	4.280 (Conserved)
Mutation Taster	-	-	0.810 (Deleterious)
Mutation Assessor	-	-	-
VEST3	-	-	-
PROVEAN	-	-	-
MutPred Splicing	NSV	NSV	NSV
HSF	NSV	NSV	NSV
ClinVar variation ID	95810	95794	-
ClinVar classification	Benign	Benign	-
ACMG codes	BS1, BP6 & BP7	BS1, BP6 & BP7	PSV1, PM1 & PM2
ACMG classification	likely benign	likely benign	Pathogenic

[1] den Dunnen, 2016
[2] Karczewski *et al*., 2019
[3] Karczewski *et al*., 2019
[4] Scaled score of 20-29= Mutation falls within 1% of the most deleterious mutations. Scaled score of 30-99= Mutation falls within 0.1% of the most deleterious mutations (Kircher *et al*., 2014).
Abbreviations: &, and; -, not available; A, adenine; Ala, Alanine; C, cytosine; CADD, Combined annotation-dependent depletion; *CHD7*, DANN, Deleterious Annotation of genetic variants using Neutral Networks; FATHMM, Functional Analysis through Hidden Markov Models; G, Guanine; GERP++, Genomic Evolutionary Rate Profiling; Gln, Glutamine; HGVS, Human Genome Variation Society; HSF, Human Splice Finder ; MAF, Minor Allele Frequency; NMD, nonsense mediated decay; NSV, non-splice variant; PolyPhen-2, Polymorphism Phenotyping v2; SIFT, Sorting Intolerant From Tolerant; T, Thymine.

The heterozygous *CHD7* c.7278G>A (p.Gln2426=) variant identified in this patient is a synonymous variant recorded in dbSNP (rs187311127). The variant lies in a non-conserved region and is not predicted to alter splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). Two sources classify this variant as benign and three as likely benign in ClinVar (Landrum *et al*., 2014). The ACMG code BP6 therefore applies (Richards *et al*., 2015). Furthermore, the variant has been observed in 62 and 40 apparently healthy African individuals in gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019). The ACMG

code BS1 therefore applies (Richards *et al*., 2015). Based on the ACMG codes BP7, BP6 and BS1, the variant was likely benign (Richards *et al*., 2015).

The heterozygous *CHD7* c.5307C>T (p.Ala1769=) variant identified in this patient is a synonymous variant recorded in dbSNP (rs16926499). The variant lies in a non-conserved and is predicted to not alter splicing. The ACMG code BP7 therefore applies (Richards *et al*., 2015). Six sources classify this variant as benign in and two as likely benign in ClinVar (Landrum *et al*., 2014). The ACMG code BP6 therefore applies (Richards *et al*., 2015). The variant has been observed in 879 and 554 apparently healthy African individuals in gnomAD Exomes and Genomes datasets, respectively (Karczewski *et al*., 2019). Furthermore, the variant has an allele frequency of 0.06 in gnomAD African dataset and there are 15286 observed alleles in this position (Karczewski *et al*., 2019). The ACMG code BA1 therefore applies (Ghosh *et al*., 2018). In keeping with the ACMG codes BS1, BP7 and BP6, the variant was likely benign (Richards *et al*., 2015).

The *CHD7* c.643C>T (p.Gln215Ter) variant identified in this patient is not observed in either of the gnomAD databases despite good coverage (Karczewski *et al*., 2019). The ACMG code PM2 therefore applies (Richards *et al*., 2015). The *CHD7* c.643C>T is a nonsense variant which introduces a premature stop codon at amino acid 215 into the CHD7 protein (Figure 3.24**)**. The ACMG code PSV1 therefore applies (Richards *et al*., 2015). The variant lies in a conserved region across species and alterations in this region are mostly pathogenic (Cooper *et al*., 2005). Four prediction tools yield a pathogenic classification (FATHMM, CADD, DANN and Mutation Taster). The ACMG code PP3 therefore applies (Richards *et al*., 2015). Based on the ACMG codes PVS1, PP3 and PM2, the variant is pathogenic (Richards *et al*., 2015). Figure 3.25 shows the c.643C>T variant as visualised on IGV.

(a)

Reference protein

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFSQGQEGL NQGNPFIATS GPGHLSHVPQ
 241  QSPSMAPSLR HSVQQFHHHP STALHGESVA HSPRFSPNPP QQGAVRPQTL NFSSRSQTVP
 301  SPTINNSGQY SRYPYSNLNQ GLVNNTGMNQ NLGLTNNTPM NQSVPRYPNA VGFPSNSGQG
 361  LMHQQPIHPS GSLNQMNTQT MHPSQPQGTY ASPPPMSPMK AMSNPAGTPP PQVRPGSAGI
 421  PMEVGSYPNM PHPQPSHQPP GAMGIGQRNM GPRNMQQSRP FIGMSSAPRE LTGHMRPNGC
 481  PGVGLGDPQA IQERLIPGQQ HPGQQPSFQQ LPTCPPLQPH PGLHHQSSPP HPHHQPWAQL
 541  HPSPQNTPQK VPVHQHSPSE PFLEKPVPDM TQVSGPNAQL VKSDDYLPSI EQQPQQKKKK
 601  KKNNHIVAED PSKGFGKDDF PGGVDNQELN RNSLDGSQEE KKKKKRSKAK KDPKEPKEPK
 661  EKKEPKEPKT PKAPKIPKEP KEKKAKTATP KPKSSKKSSN KKPDSEASAL KKKVNKGKTE
 721  GSENSDLDKT PPPSPPPEED EDPGVQKRRS SRQVKRKRYT EDLEFKISDE EADDADAAGR
 781  DSPSNTSQSE QQESVDAEGP VVEKIMSSRS VKKQKESGEE VEIEEFYVKY KNFSYLHCQW
 841  ASIEDLEKDK RIQQKIKRFK AKQGQNKFLS EIEDELFNPD YVEVDRIMDF ARSTDDRGEP
 901  VTHYLVKWCS LPYEDSTWER RQDIDQAKIE EFEKLMSREP ETERVERPPA DDWKKSESSR
 961  EYKNNNKLRE YQLEGVNWLL FNWYNMRNCI LADEMGLGKT IQSITFLYEI YLKGIHGPFL
1021  VIAPLSTIPN WEREFRTWTE LNVVVYHGSQ ASRRTIQLYE MYFKDPQGRV IKGSYKFHAI
1081  ITTFEMILTD CPELRNIPWR CVVIDEAHRL KNRNCKLLEG LKMMDLEHKV LLTGTPLQNT
1141  VEELFSLLHF LEPSRFPSET TFMQEFGDLK TEEQVQKLQA ILKPMMLRRL KEDVEKNLAP
1201  KEETIIEVEL TNIQKKYYRA ILEKNFTFLS KGGGQANVPN LLNTMMELRK CCNHPYLING
1261  AEEKILEEFK ETHNAESPDF QLQAMIQAAG KLVLIDKLLP KLKAGGHRVL IFSQMVRCLD
1321  ILEDYLIQRR YPYERIDGRV RGNLRQAAID RFSKPDSDRF VFLLCTRAGG LGINLTAADT
1381  CIIFDSDWNP QNDLQAQARC HRIGQSKSVK IYRLITRNSY EREMFDKASL KLGLDKAVLQ
1441  SMSGRENATN GVQQLSKKEI EDLLRKGAYG ALMDEEDEGS KFCEEDIDQI LLRRTHTITI
1501  ESEGKGSTFA KASFVASGNR TDISLDDPNF WQKWAKKAEL DIDALNGRNN LVIDTPRVRK
1561  QTRLYSAVKE DELMEFSDLE SDSEEKPCAK PRRPQDKSQG YARSECFRVE KNLLVYGWGR
1621  WTDILSHGRY KRQLTEQDVE TICRTILVYC LNHYKGDENI KSFIWDLITP TADGQTRALV
1681  NHSGLSAPVP RGRKGKKVKA QSTQPVVQDA DWLASCNPDA LFQEDSYKKH LKHHCNKVLL
1741  RVRMLYYLRQ EVIGDQADKI LEGADSSEAD VWIPEPFHAE VPADWWDKEA DKSLLIGVFK
1801  RTPFKDEIDE FANSPSEDKE ESMEIHATGK HSESNAELGQ LYWPNTSTLT TRLRRLITAY
1861  QRSYKRQQMR QEALMKTDRR RRRPREEVRA LEAEREAIIS EKRQKWTRRE EADFYRVVST
1921  FGVIFDPVKQ QFDWNQFRAF ARLDKKSDES LEKYFSCFVA MCRRVCRMPV KPDDEPPDLS
1981  SIIEPITEER ASRTLYRIEL LRKIREQVLH HPQLGERLKL CQPSLDLPEW WECGRHDRDL
2041  LVGAAKHGVS RTDYHILNDP ELSFLDAHKN FAQNRGAGNT SSLNPLAVGF VQTPPVISSA
2101  HIQDERVLEQ AEGKVEEPEN PAAKEKCEGK EEEEETDGSG KESKQECEAE ASSVKNELKG
2161  VEVGADTGSK SISEKGSEED EEEKLEDDDK SEESSQPEAG AVSRGKNFDE ESNASMSTAR
2221  DETRDGFYME DGDPSVAQLL HERTFAFSFW PKDRVMINRL DNICEAVLKG KWPVNRRQMF
2281  DFQGLIPGYT PTTVDSPLQK RSFAELSMVG QASISGSEDI TTSPQLSKED ALNLSVPRQR
2341  RRRRRKIEIE AERAAKRRNL MEMVAQLRES QVVSENGQEK VVDLSKASRE ATSSTSNFSS
2401  LSSKFILPNV STPVSDAFKT QMELLQAGLS RTPTRHLLNG SLVDGEPPMK RRRGRRKNVE
2461  GLDLLFMSHK RTSLSAEDAE VTKAFEEDIE TPPTRNIPSP GQLDPDTRIP VINLEDGTRL
2521  VGEDAPKNKD LVEWLKLHPT YTVDMPSYVP KNADVLFSSF QKPKQKRHRC RNPNKLDINT
2581  LTGEERVPVV NKRNGKKMGG AMAPPMKDLP RWLEENPEFA VAPDWTDIVK QSGFVPESMF
2641  DRLLTGPVVR GEGASRRGRR PKSEIARAAA AAAAVASTSG INPLLVNSLF AGMDLTSLQN
2701  LQNLQSLQLA GLMGFPPGLA TAATAGGDAK NPAAVLPLML PGMAGLPNVF GLGGLLNNPL
2761  SAATGNTTTA SSQGEPEDST SKGEEKGNEN EDENKDSEKS TDAVSAADSA NGSVGAATAP
2821  AGLPSNPLAF NPFLLSTMAP GLFYPSMFLP PGLGGLTLPG FPALAGLQNA VGSSEEKAAD
2881  KAEGGPFKDG ETLEGSDAEE SLDKTAESSL LEDEIAQGEE LDSLDGGDEI ENNENDE*
```

(b)

Protein predicted from variant coding sequence

```
   1  MADPGMMSLF GEDGNIFSEG LEGLGECGYP ENPVNPMGQQ MPIDQGFASL QPSLHHPSTN
  61  QNQTKLTHFD HYNQYEQQKM HLMDQPNRMM SNTPGNGLAS PHSQYHTPPV PQVPHGGSGG
 121  GQMGVYPGMQ NERHGQSFVD SSSMWGPRAV QVPDQIRAPY QQQQPQPQPP QPAPSGPPAQ
 181  GHPQHMQQMG SYMARGDFSM QQHGQPQQRM SQFS*
```

Figure 3.24: The *CHD7* c.643C>T resultant predicted protein

The *CHD7* c.643C>T introduces a premature stop codon at amino acid 215 into the CHD7 protein. Note (**a**) The normal CHD7 protein reference sequence (**b**) and the protein sequence predicted from *CHD7* c.643C>T variant (b).

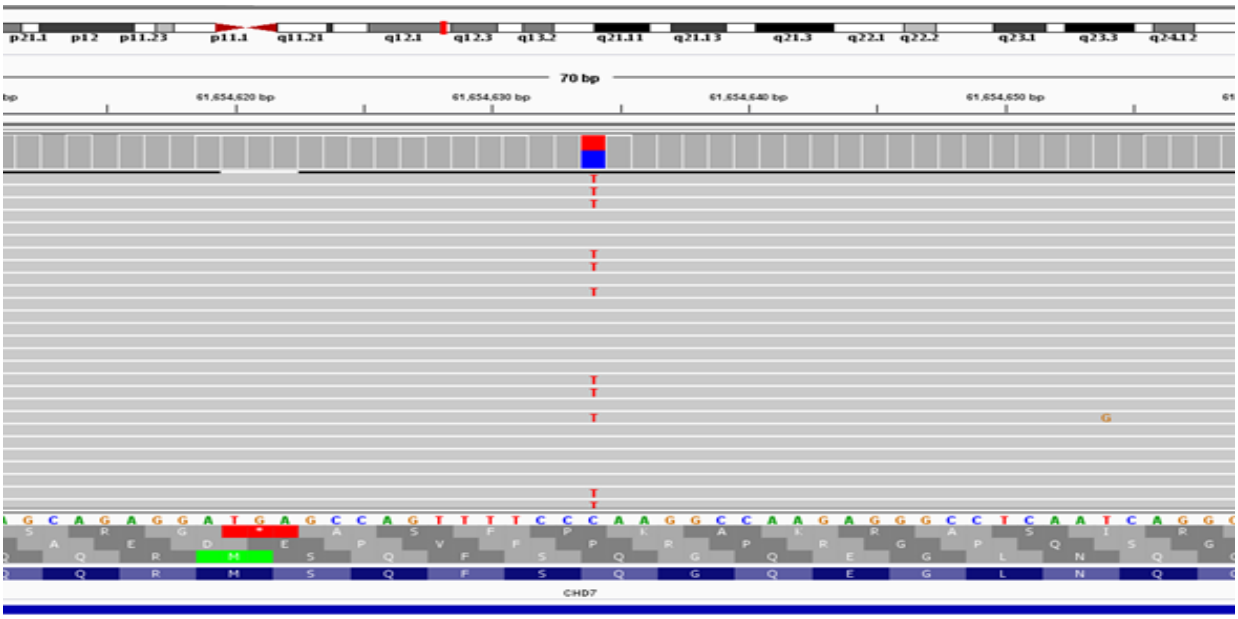

Figure 3.25: A screenshot of IGV showing the *CHD7* c.643C>T variant identified in FRASC61
The genetic variant was visualised by importing the BAM file into IGV.

3.6 Overall diagnostic yield

In this study, out of the 15 patients tested, seven disease-causing mutations were identified in seven unrelated patients giving an overall diagnostic yield of 47%. Of the seven identified disease-causing variants, three were associated with TCS and four were identified in clinically diagnosed CHARGE syndrome patients. In addition, one VUS, predicted to affect splicing was identified in one TCS patient. A list of putative disease-causing variants identified in this study is summarised in Table 3.19 below and Figures 3.26-3.28 show the distribution of these variants within their specific genes.

Table 3.19: A summary of putative disease-causing variants identified in this study

Patient	Ethnicity	Clinical diagnosis	Gene	HGVS[1]	Zygosity	Nature of mutation	Coding impact	ACMG classification
FRASC26	African	TCS	*TCOF1*	c.4369_4373delAAGAA	Het	Five bp deletion	Frameshift	Pathogenic
FRASC28	Caucasian	TCS	*POLR1D*	c.261delA	Homo	One bp deletion	Frameshift	Likely pathogenic
FRASC34	African	CHARGE	*CHD7*	c.232C>T	Het	Substitution variant	Nonsense	Pathogenic
FRASC54	African	CHARGE	*CHD7*	c.1931delA	Het	One bp deletion	Frameshift	Likely pathogenic
FRASC57	African	TCS	*TCOF1*	c.3708delC	Het	One bp deletion	Frameshift	Likely pathogenic
FRASC59	African	CHARGE	*CHD7*	c.3309_3310delCA	Het	Two bp deletion	Nonsense	Pathogenic
FRASC61	African	CHARGE	*CHD7*	c.643C>T	Het	Substitution variant	Nonsense	Pathogenic

[1] den Dunnen, 2016

Abbreviations: -, Not available; A, Adenine ; Ala , Alanine ; bp, base pairs; C, Cytosine; CHARGE, Coloboma; Heart defects; Atresia choanae ; growth retardation; Genital abnormalities; and Ear abnormalities ; *CHD7*,Chromodomain helicase DNA binding protein 7; G, Guanine; Het, heterozygous; Homo, Homozygous; *POLR1D*, RNA polymerase I and III subunit D; TCS, Treacher Collins Syndrome; *TCOF1*,treacle ribosome biogenesis factor

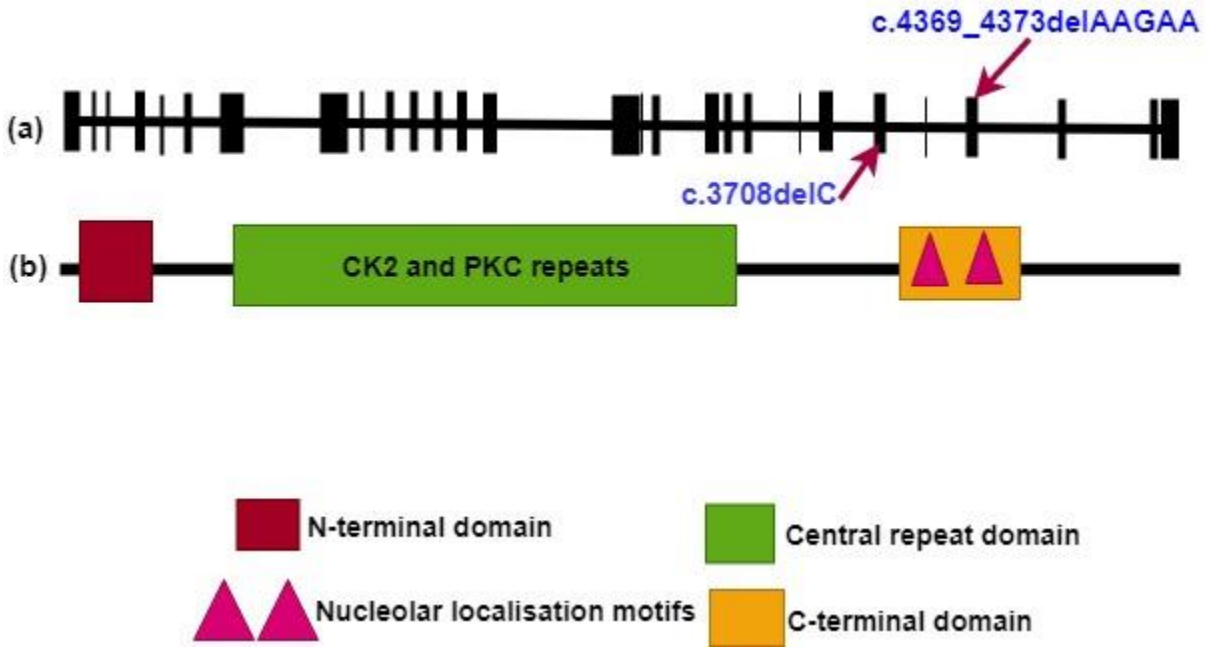

Figure 3.26: Distribution of putative disease-causing variants identified within the *TCOF1* gene

(**a**) A schematic representation of the *TCOF1* gene showing the location of putative disease-causing variants identified in this study *TCOF1* gene (**b**) and the treacle protein structure (adapted from Winokur and Shiang, 1998).

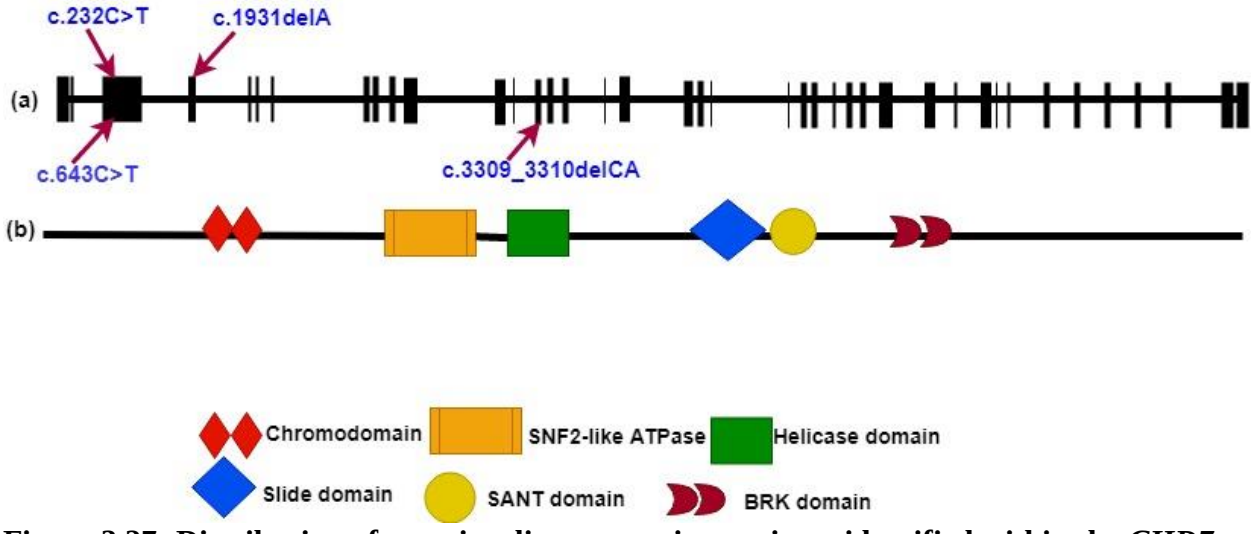

Figure 3.27: Distribution of putative disease-causing variants identified within the *CHD7* gene

(a) A schematic representation of the *CHD7* gene showing the location of putative disease-causing variants identified in this study *CHD7* gene (**b**) and the CHD7 protein structure (Adapted from Bouazoune and Kingston, 2012).

Figure 3.28: Figure 3.28: A schematic representation of putative disease-causing variant identified in *POLR1D*

(**a**) A schematic presentation of the *POLR1D* gene showing the location of putative disease-causing variant identified in this study (**b**) a schematic presentation of the POLR1D protein (Yao *et al*., 1997).

3.7 Sanger sequencing

Following primer design, PCR was optimised and the PCR products were separated using Agarose gel electrophoresis. Figure 3.29 below shows a collection of gel images as visualised and imaged using an Ultraviolet transilluminator (Omega Fluor™ Gel Documentation System, Vacutec, Johannesburg, South Africa). However, owing to time constraints, Sanger validation and segregation analysis could not be performed.

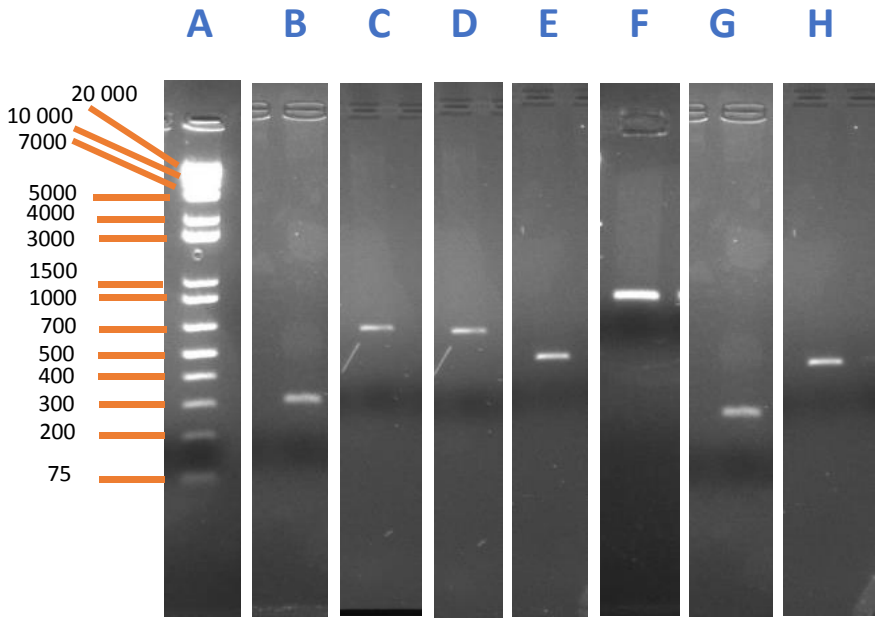

Figure 3.29: A collection of gel images showing PCR optimisation using the seven sets of primes designed for Sanger validation

Gel images showing optimisation of the seven sets of primers designed to validate identified putative disease–causing variants. Please note (**A**) Thermofisher 1kbp plus DNA ladder (**B**) primers targeting the *TCOF1* c.4369_4373delAAGAA (**C**) *POLR1D* c.261delA (**D**) *CHD7* c.232C>T (**E**) *CHD7* c.1931delA (**F**) *TCOF1* c.3708delC (**G**) *CHD7* c.3309_3310delCA and (**H**) *CHD7* c.643C>T variants.

4 DISCUSSION

TCS, NS and MS share overlapping clinical features and display variable expressivity. The genetic aetiology of these facial dysostoses disorders in South African patients is currently unknown, as there is no literature published on this topic. As a result, no genetic testing is available in the country and clinicians depend on the assessment of clinical features and family history to make a diagnosis; this can be problematic due to phenotypes overlapping and the variable expressivity displayed by these disorders and could lead to mis- or under diagnosis. Clinical recognition of these disorders is further complicated by disorders which can be considered under their differential diagnoses like Broncho-oto-renal (BOR) syndrome and CHARGE syndrome. Internationally, a clinical diagnosis of TCS, NS or MS is confirmed at a molecular level using different techniques such as single gene analysis, chromosomal microarray analysis, multiplex ligation-dependent probe amplification (MLPA) and NGS panels (Katsanis and Jabs, 2004). In this study targeted NGS-based mutation screening was performed on 15 unrelated patients with a clinical diagnosis of TCS, NS, BOR syndrome or CHARGE syndrome. This study aimed to generate a mutation profile of these disorders in the South African population. The availability of this data could inform the development of an appropriate and cost-effective NGS-based diagnostic test of these disorders in SA.

4.1 Demographic data

A total of 15 patients and 19 family members were recruited for this study. The objective of the study was to recruit 16 participants; however, because of the rare nature of the disorders studied, we only managed to recruit 15 participants. Of these 15 patients, two patients were deceased. The average age of the remaining 13 patients was 4.4 years (range: 1-12 years). Most individuals were of African ancestry (10/15), followed by Caucasians (4/15) and there was a single Indian patient. The majority (7/15) of the patients had a provisional clinical diagnosis of TCS followed by CHARGE syndrome (5/15). No MS patients were recruited for this study most

4.2 Clinical and Molecular analysis

In the present study, seven putative disease-causing variants were identified in seven patients. Of the seven variants, three were identified in patients with TCS and four were identified in patients with CHARGE syndrome, confirming the initial clinical diagnosis in all seven cases. The overall diagnostic yield of the present study was 47%. We discuss the clinical phenotype and diagnostic yield of TCS and CHARGE syndrome. In addition, we speculate on the mechanism of disease for each identified disease-causing variant.

4.2.1 TCS phenotype

Treacher Collins syndrome is an AD congenital disorder characterised by the presence of variable craniofacial, ophthalmic and ear anomalies. Several studies have defined the major and minor diagnostic criteria of TCS (Dauwerse *et al*., 2011; Teber *et al*., 2004; Trainor *et al*., 2009; Vincent *et al*., 2016) and are listed in Table 1.1. According to Katsanis and Jabs (2004), genetic testing for TCS should be considered when a patient exhibits at least two major or three minor clinical features. All persons clinically diagnosed with TCS in the present study met these criteria. The most commonly seen clinical features were malar flattening (100%), mandible hypoplasia (100%) and microtia (86%), of which 50% were a grade three microtia. External ear canals and hearing loss were present in 57% of the patients and minor clinical features such as lateral cheek hair displacement, ear tags, ear pits and ptosis were only present in 14% of the patients.

4.2.2 Diagnostic yield of TCS

Three putative disease-causing variants were identified in three of seven clinically diagnosed TCS patients. Our TCS diagnostic yield was therefore 43%. In other studies, the pick-up rate of TCS is estimated to range from 63% - 93% (Katsanis and Jabs, 2004). This skewed pick-up rate could be attributed to the cohort's small sample size (n=7) in contrast to larger sample size of n=58, n=46, n=146 reported in previous studies (Edwards *et al*., 1997; Teber *et al*., 2004; Vincent *et al*., 2016). Nevertheless, the pick-up rate is still informative as to the utility of the platform used. According to the HGMD, the most prevalent TCS mutations within the three

known TCS causative genes are point mutations, accounting for approximately 86% of TCS cases. Therefore, our findings further suggest that a targeted NGS mutation testing could be an acceptable primary screening method for TCS point mutations in the South African population.

4.2.3 Putative disease-causing variants identified in patients with TCS

Mutations in the *TCOF1*, *POLR1C* and *POLR1D* genes are a known cause of TCS (Dauwerse *et al*., 2011: Treacher Collins Syndrome Collaborative Group, 1996). The *TCOF1* gene encodes a nuclear phosphoprotein called treacle. Treacle is localised to both the cytoplasm and nucleolus during cell division. During embryogenesis, treacle is expressed in the neuroepithelial cells that give rise to the neural crest cells, the latter differentiating into cartilage, bones, and connective tissue of the head and the face (Dixon *et al*., 1997; Marsh *et al*., 1998; Sakai *et al*., 2012). The *POLR1D* and *POLR1C* genes encode for subunits shared by RNA Pol I and RNA Pol III involved in ribosomal RNA transcription during ribosomal biogenesis (Lafontaine and Tollervey, 2001).

The *TCOF1* gene is the most commonly mutated gene in patients with TCS. Globally, more than 90% of patients with a molecular diagnosis have mutations in this gene. The same trend can be seen in this cohort as the majority of putative disease-causing variants identified occurred within the *TCOF1* gene. Literature reports that mutations within the *POLR1C* and *POLR1D* genes are associated with 1.2% and 6% of TCS cases, respectively (Dauwerse *et al*., 2011; Vincent *et al*., 2016). In this study, a single variant (c.261delA) was identified in *POLR1D,* yielding a pick-up rate of 33% and no mutations were identified within the *POLR1C* gene. Similarly, all putative disease-causing TCS variants in this study were frameshift deletions of up to five base pairs. This agrees with what has been reported in other studies which found that the majority of TCS mutations are indels of up to 40 nucleotides, resulting in frameshift mutations (Splendore *et al*., 2000).

Literature shows that the majority of TCS mutations are private (Dauwerse *et al*., 2011; Trainor and Andrews, 2013), with few common mutations being described. In our cohort, private mutations were identified in two patients (FRASC28 and FRASC57) and the third (FRASC26) harbored a well-documented 5-bp deletion (Bowman *et al*., 2012; Dixon *et al*., 2004; Edwards *et*

al., 1997; Ellis *et al*., 2002; Horiuchi *et al*., 2005 and Splendore *et al*., 2000), which is located in one of the TCS mutation hot spots, exon 24 of the *TCOF1* gene. It is hypothesised that 60% of TCS mutations occur *de novo* (Jones *et al*., 1975; Splendore *et al*., 2002); however, segregation analysis was not performed in our cohort, owing to time constraints.

Each putative disease-causing variant identified in patients with TCS has convincing evidence to suggest pathogenicity. This evidence was gathered and assessed by the ACMG guidelines for sequence variant interpretation. Literature shows that haploinsufficiency is the proposed mechanism of disease of TCS (Trainor and Andrews, 2013). We, therefore, speculate on the mechanism of disease of all identified putative disease-causing variant based on the possibility of their mRNA transcript to be targeted by NMD as predicted by SIFT indel (Hu *et al*., 2013) and their respective location and amino acid residues they affect within their respective protein sequences as visualised on Mutalyzer.

4.2.3.1 The *TCOF1* c.3708delC putative disease-causing variant

The likely pathogenic *TCOF1* c.3708delC variant identified in FRASC57 (Table 3.16) is a single nucleotide frameshift deletion that affects amino acid 1236 of the treacle protein (Figure 3.21) (Winokur and Shiang, 1998). This position is not located within a known functional domain of treacle (Figure 3.26). Analysis of this variant by the SIFT Indel prediction tool predicts that the c.3708delC mRNA transcript is a target of degradation by NMD. Although mostly degraded, some transcripts escape this mechanism (Neu-Yilik *et al*., 2011). Visualisation of the treacle protein sequence on Mutalyzer shows that the c.3708delC variant would introduce a premature termination codon at amino acid 70 of 1489, should its mutated transcript escape NMD. Thus, the resultant protein would lack the C-terminal functional domain located from amino acids 1307 to amino acid 1439 (Figure 3.26). Functional studies have revealed that a treacle protein lacking the C-terminal domain is not expressed in cells as it lacks the ability to enter into the nucleolus, the site of ribosomal biosynthesis and pre-ribosomal assembly (Isaac *et al*., 2000). FRASC57 is heterozygous for this variant and thus haploinsufficiency of the treacle protein is the probable mechanism for disease in this patient.

4.2.3.2 The *TCOF1* c.4369_4373delAAGAA putative disease-causing variant

The pathogenic c.4369_4373delAAGAA identified in FRASC26 (Table 3.7) is a commonly reported 5bp frameshift deletion. This variant has been observed in various population groups including Caucasians (Conte *et al*., 2011; Edward *et al*., 1997; Splendore *et al*., 2000; Splendore *et al*., 2002), Brazilians (Conte *et al*., 2011), Japanese (Horiuchi *et al*., 2005) and Chinese (Chen *et al*., 2018). This deletion is located in a known TCS mutational hotspot within the C-terminal of the treacle protein (Mash *et al*., 1998), which constitutes two nucleolar localization signal motifs which are critical for the identification of the nucleolus during treacle translocation from the cytoplasm into the nucleolus (Figure 3.26) (Marsh *et al*., 1998; Winokur and Shiang, 1998). Because of its location, the c.4369_4373delAAGAA mutation likely disrupts the second nucleolar localization signal motif. Disruption in either of the localisation motifs leads to reduced localisation signal and as a consequent, reduced gene expression (Isaac *et al*. 2000). This abnormal protein could therefore possibly have reduced nucleolar localisation ability.

Interestingly, the c.4369_4373delAAGAA's mRNA transcript is not targeted for degradation by NMD. While the NMD mechanism generally degrades transcripts which harbour premature termination codons, there are exceptions to this mechanism. One such exception is that NMD is not activated in transcripts which harbour premature termination codon within the NMD insensitive regions. These regions include the 50 nucleotides before the last exon and the whole last exon of the gene (Kurosaki and Maquat, 2016; Lykke-Andersen and Janssen, 2015). Visualization of the treacle protein sequence on Mutalyzer shows that c.4369_4373delAAGAA variant introduces a premature termination codon 21 amino acids away from the native stop codon (Figure 3.8). Thus, the c.4369_4373delAAGAA variant occurs within an NMD insensitive region. FRASC26 is heterozygous for this variant. Therefore, haploinsufficiency of the treacle protein is the possible mechanism for disease in this patient.

4.2.3.3 The *POLR1D* c.261delA putative disease-causing variant

The likely pathogenic *POLR1D* c.261delA identified in FRASC28 (Table 3.9) is a single nucleotide frameshift mutation located in the third and last exon of the *POLR1D* gene (Figure 3.28). This exon houses the RNA polymerase dimerization domain crucial for normal

functioning of the RNA pol I and pol III enzymes during rRNA and/or tRNA transcription (Valdez *et al*., 2004; Yao *et al*., 1999). Analysis of this variant by the SIFT Indel prediction tool predicts that the c.261delA mRNA transcript is not a target of NMD. Visualization of the POLR1D protein sequence on Mutalyzer shows c.261delA variant will introduce a premature termination codon at amino acid 13 downstream (Figure 3.12). Impaired RNA polymerase dimerization could reduce the function of Pol I and/or Pol III, which could consequently result in a deficiency of rRNA and/or tRNA in this patient (Szymanski *et al*., 2002). FRASC28 is homozygous for this variant. Therefore, reduced expression of the *POLR1D* gene and subsequent haploinsufficiency of the protein could possibly be a mechanism of disease. Further RNA-level investigations and segregation studies in FRASC28's family will be required to elucidate the precise deleterious effect. Homozygous *POLR1D* mutations are not common. Literature only documents two cases of *POLR1D* homozygous mutations identified in two separate consanguineous families by Schaefer in 2014 (Schaefer *et al*., 2014).

4.2.3.4 The *TCOF1* c.3183G>C putative disease-causing variant

The *TCOF1* c.3183G>C variant identified in FRASC13 (Table 3.4) is a missense variant located at the distal end of exon 19 of the *TCOF1* gene. Missense variants cause disease through alterations of the protein function and/or affecting splicing. Functional studies have revealed that variants located at the end of exons tend to inhibit splicing factors recognising the exon/intron boundaries and usually result in exon skipping (Horiuchi *et al*., 2005; Talerico and Berget, 1990). It is important to note that the guanine nucleotide at position 3183 is highly conserved across species, suggesting that alterations in this nucleotide may not be tolerated (Davydov *et al*., 2010). Two potential disease-causing substitutions of this guanine have been identified in two separate TCS studies: firstly by Horiuchi *et al.* (2005) and then by Bownman *et a*l. (2012). Horiuchi *et al*. (2005) identified a substitution of this guanine by thymine (*TCOF1* c.3183G>T) in one TCS patient. Interestingly, analysis of their proband's RNA revealed that the variant resulted in skipping of exon 18 and 19. Thus in their case, the variant was confirmed to cause disease through aberrant splicing (Horiuchi *et al*., 2005). Bownman *et al*. (2012) identified a substitution of this guanine by adenine (*TCOF1* c.3183G>A) in one patient with TCS. The investigators classified their *TCOF1* c.3183G>A variant as potentially pathogenic owing to

suggested effects on splicing predicted by in silico prediction tools. Analysis of the *TCOF1* c.3183G>C variant identified in the present study using Human Splicing Finder (Desmet *et al*., 2009) and MutPred Splicing (Mort *et al*., 2014) prediction tools predict that the G>C substitution could also disrupt the natural 5' splice site of intron 19. Therefore, this *TCOF1* variant identified in FRASC13 could possibly affect splicing as well. However, following the ACMG guidelines, the variant was classified as a VUS (Richards *et al*., 2015). It is recommended that RNA and cDNA sequencing should be carried out in order to elucidate the contribution of this variant to the development of disease in this patient. The mechanism of disease in this patient is possibly aberrant splicing.

4.2.4 CHARGE syndrome phenotype

CHARGE syndrome is an AD congenital developmental disorder characterised by a wide spectrum of clinical features. The most common clinical features of CHARGE syndrome include coloboma, heart defects, atresia choanae, growth retardation, genital abnormalities and ear abnormalities. In 1998, Blake *et al.* (1998) defined the major and minor diagnostic criteria of CHARGE syndrome, which was later updated in 2005 by Verloes (Verloes, 2005). The major criteria being coloboma, choanal atresia and hypoplastic semicircular canals and the minor criteria being malformation of the middle and/or external ear, mental retardation, anomalies of the heart and oesophagus, hypothalamo-hypophyseal and rhombencephalic dysfunction (Verloes, 2005). Verloes's criteria classify CHARGE syndrome cases into three categories: typical, partial or incomplete and atypical CHARGE syndrome diagnosis. The presence of all three major criteria, or two major and two minor criteria, warrant a typical CHARGE syndrome diagnosis. The presence of two major and one minor criteria is suggested to warrant a partial or incomplete CHARGE syndrome diagnosis and the presence of two or one major and thee minor criteria, warrant an atypical CHARGE syndrome diagnosis (Verloes, 2005). Based on Verloes's diagnostic criteria, four of the five patients with CHARGE syndrome (FRASC23, FRASC54, FRASC59, and FRASC61) presented with an atypical CHARGE syndrome and a single patient (FRASC61) met the diagnostic criteria for a typical CHARGE syndrome diagnosis.

4.2.5 Diagnostic yield of CHARGE syndrome

Four putative disease-causing variants were identified in four of five clinically diagnosed CHARGE syndrome patients, yielding a pick-up rate of 80%. Previous studies have reported a CHARGE syndrome pick up rate varying between 65%-70% in cohorts studying both patients with typical and suspected CHARGE syndrome (Aramaki *et al*., 2006: Jongmans *et al*., 2006: Zentner *et al*., 2010). The sample size in this study was obviously small, so interpretation of these figures should be undertaken with caution. Our diagnostic yield was similar to the 80% (4/5) reported by Janssen *et al* in a Caucasian cohort using whole exome sequencing (Janssen *et al*., 2012). This skewed pick up rate could be attributed to the small sample size; however, it is still informative as to the platform used as well as the gene sequenced. Our findings suggest that the *CHD7* gene remains a relevant gene to test for CHARGE syndrome in the South African populations. According to the HGMD and CHD7 database (Janssen *et al*., 2012), the most prevalent CHARGE syndrome mutations are point mutations, accounting for approximately 89% of CHARGE syndrome mutations, therefore, our findings further suggest that a targeted NGS mutation screening is an acceptable screening method for CHARGE syndrome and that *CHD7* screening would be an appropriate diagnostic strategy for CHARGE syndrome in SA, as it is globally.

4.2.6 Putative disease-causing variants identified in patients with CHARGE syndrome

The *CHD7* gene encodes a chromodomain helicase DNA binding protein 7 (CHD7) which regulates gene expression during embryonic development through nucleosomal remodeling (Bajpai *et al*., 2010; Engelen *et al*. 2011). The *CHD7* gene is the only mutated gene amongst patients with CHARGE syndrome (Engelen *et al*., 2011; Lalani *et al*., 2006; Vissers *et al*., 2004). The same trend can be seen in this study. Similarly, the majority of the *CHD7* putative disease-causing variants identified in this study are nonsense variants. All putative disease-causing variants in this cohort were private, which is in accordance with what has been observed in previous studies in other populations (Bartels *et al*., 2010; Vissers *et al*., 2004). The majority of *CHD7* mutations have a *de novo* origin (Vissers *et al*., 2004). However, owing to time limitations, segregation analysis was not performed in this study.

We speculate on the mechanism of disease of all identified putative disease-causing variant based on the possibility of their mRNA transcript to be targeted by NMD as predicted by SIFT indel (Hu *et al*., 2013) and their respective location and amino acid residues they affect within their respective protein sequences as visualised on Mutalyzer.

4.2.6.1 The *CHD7* c.232C>T and *CHD7* c.643C>T putative disease-causing variants

The pathogenic *CHD7* c.232C>T and c.643C>T variants identified in FRASC34 (Table 3.12) and FRASC61 (Table 3.18), respectively, are both nonsense variants that affect amino acids 78 and 215 of the CHD7 protein, respectively (Figure 3.16 and Figure 3.24). These positions are not located within a known functional domain of the CHD7 protein (Bouazoune and Kingston, 2012) (Figure 3.27). Analysis of these variants by the SIFT Indel prediction tool predicts that both c.232C>T and c.232C>T mRNA transcripts are a target of degradation by NMD. Visualisation of the CHD7 protein sequence on Mutalyzer shows that the c.232C>T and c.643C>T variants could therefore introduce premature termination codons at amino acid 78 and 215, respectively should their mutated transcripts escape NMD. Thus, the resultant proteins could lack all functional domains of the CHD7 protein (See Figure 3.27). The CHD7 protein lacking functional domains completely abolishes the *CHD7* nucleosomal remodeling activities in animal models (Bouazoune and Kingston, 2012). Both FRASC34 and FRASC61 are heterozygous for their specific variants; suggesting they could only possess a single functional *CHD7* allele. Thus, haploinsufficiency of the CHD7 protein during embryonic development is the possible mechanism for disease in these patients.

4.2.6.2 The *CHD7* c.3309_3310delCA putative disease-causing variant

The pathogenic c.3309_3310delCA variant identified in FRASC59 (Table 3.17) is a two bp deletion resulting in a nonsense mutation that affects amino acid 1104 of the CHD7 protein (Figure 3.22). This position is located within the helicase domain of the *CHD7* protein (Bouazoune and Kingston, 2012). This helicase domain of the *CHD7* protein houses the ATP-binding motif critical for ATP hydrolysis (Bouazoune and Kingston, 2012; Thompson *et al*., 2009). Analysis of the c.3309_3310delCA variant by the SIFT Indel prediction tool predicts that

the c.3309_3310delCA mRNA transcript is a target of degradation by NMD. Visualisation of the CHD7 protein sequence on Mutalyzer shows that the c.3309_3310delCA variant could therefore introduce a premature termination codon at amino acid residue 711 of 2998 should its mutated transcript escape NMD. The resultant protein could have an impaired ATP-binding domain and in addition, lack five of its seven functional domains of the CHD7 protein (See Figure 3.27). Thus, the resultant protein could lose its ability to catalyse nucleosome remodeling (Bouazoune and Kingston, 2012). FRASC59 is heterozygous for this variant; therefore, the patient could possibly only have a single functional *CHD7* allele. Thus, again, haploinsufficiency of the CHD7 protein during embryonic development is the possible mechanism for disease in this patient.

4.2.6.3 The *CHD7* c.1931delA putative variant

The likely pathogenic *CHD7* c.1931delA variant identified in FRASC54 (Table 3.15) variant is a single base deletion resulting in a frameshift that affects amino acid 644 of the CHD7 protein (Figure 3.18). This position is not located within a known functional domain of the CHD7 protein (Bouazoune and Kingston, 2012) (Figure 3.27). Analysis of this variant by the SIFT Indel prediction tool predicts that the c.1931delA mRNA transcript is a target of degradation by NMD. Visualisation of the CHD7 protein sequence on Mutalyzer shows that the c.1931delA variant could therefore introduce a premature termination codon at amino acid residue 711 of 2998 should its mutated transcript escape NMD. Thus, the resultant protein could lack all functional domains of the CHD7 protein (See Figure 3.27). Functional studies have revealed that CHD7 protein lacking functional domains completely abolishes *CHD7* nucleosomal remodeling activities in animal models (Bouazoune and Kingston, 2012). FRASC54 is heterozygous for this variant; therefore, the patient probably has only a single functional *CHD7* allele. Haploinsufficiency of the CHD7 protein is the possible mechanism for disease in this patient.

4.3 Pathogenic mutation-negative patients

In this study, no putative disease-causing variants were identified in eight of the 15 patients (FRASC13, FRASC23, FRASC25, FRASC27, FRASC31, FRASC33, FRASC35 and FRASC52). It is known that 4% and 5%-10% of clinically diagnosed TCS and CHARGE syndrome cases respectively, do not have a molecular diagnosis, irrespective of technique/s used for mutation screening (Janssen *et al*., 2012; Vincent *et al*., 2016). There are various reasons which could explain the lack of mutation pick-up in certain patients. These include the technique used in this study; the most common bioinformatics algorithm and/or tools employed for the analysis of NGS mutations are less sensitive for detecting large mutations (Spencer *et al*., 2013); therefore, larger deletions, duplications or inversions would not be detected using the tools employed in our data analysis. Our custom multigene panel only targeted certain regions of the genes (exons, intron-exon boundaries, UTR's and 10 flanking base pairs); therefore, these patients may have mutations in regions that were not targeted like the introns or the control regions.

Furthermore, the lack of putative disease-causing variants in these patients could suggest that other genes, not investigated here, may be responsible for the manifestation of these phenotypes in these patients. Due to the variability in clinical phenotype and the remarkable overlap with other genetic syndromes, it is also possible that the clinical diagnosis may be incorrect. In another person with TCS (FRASC27), visualisation of their sequenced reads revealed sequencing gaps in the *TCOF1* gene. The gaps were present in positions covering known functional domains of the treacle protein and certain targeted regions in this sample were of low sequencing quality. This raises concern that variant calling could have been compromised by the quality of data generated in this patient. Future studies resequencing this patient, with a higher sequencing coverage and possible sequencing regions with gaps using Sanger sequencing will be required to further investigate a putative disease-causing in this patient.

4.4 Limitations of the study

The limitations of this study are acknowledged. The cohort size was small due to the rare nature of the disorders under study. Therefore, the findings may not necessarily be inferable to the entire South African population. The target gene panel and bioinformatics workflow employed in this study could pick up most of common types of mutations described in patients with TCS, NS and CHARGE syndrome; however, it would not be able to detect deep intronic and larger mutations. Furthermore, owing to time constraints, family studies and validation of putative disease-causing variants using Sanger sequencing could not be performed.

While validation of putative disease-causing variants could not be performed, there is currently no definite guidance on validation of NGS detected variants in a clinical environment. According to the College of American Pathologists guidelines, it is up to the laboratory performing sequencing to determine whether validation testing is appropriate (Aziz *et al.*, 2015), and the ACMG clinical laboratory standards for NGS recommends that all NGS detected variants should be validated using a secondary test (Rehm *et al*., 2013).

Several studies assessing the necessity to validate NGS detected variants using Sanger sequencing have been undertaken. One such was carried out by Mu *et al*. (2016). In this study, the investigators aimed to assess the need to validate NGS detected variants using Sanger sequencing in 7843 variants. The study reported a validation rate of 98.7% (n=7746). The remaining 1.3% (n=99) variants were located in complex genomic regions. The authors concluded that validation of NGS targeted variants is required for variants located in difficult to sequence regions.

In contrast, other studies have shown that Sanger validation of NGS results is not required, provided a minimum coverage was achieved for the NGS: In 2013, Sikkema-Raddatz *et al* obtained a 100% (n=168) validation of variants identified through targeted NGS. In this study, the authors concluded that an at least 30x NGS coverage is equivalent to the specificity and sensitivity of Sanger sequencing and variants identified through targeted NGS with a minimum of 30× coverage do not necessarily need to be validated (Sikkema-Raddatz *et al*., 2013). Baudhuin *et al*. (2015) also reported a 100% (n=919) validation of variants identified through

targeted NGS. The authors concluded that validation of SNVs with a Q score of ≥20 and a high coverage of ≥100× could be unnecessarily redundant but recommended sequencing of indels for correct genomic location defining (Baudhuin *et al*., 2015). A high proportion of variants validated through Sanger sequencing is also reported in data generated through whole exome sequencing: Strom *et al*. (2014), reported a 99.090% (n=109 of 110) validation of variants identified through whole exome sequencing. In a larger study, Beck *et al*. (2016), reported a 99.965% (n=5658 of 5660) Sanger sequencing validation rate. The findings and recommendations of these studies suggest that the genetics community has not yet reached a consensus on this matter.

4.5 Feedback to the patients

All patients and/or families recruited for this study showed interest in receiving feedback on the study's findings. Studies have shown that individuals who undergo genomic sequencing for rare disorders report interest in receiving results for reasons beyond clinical use such as informed reproductive decision-making, enhancing self-knowledge as well as identifying useful support- and advocacy groups (ACMG board of directors, 2015; Kohler *et al*., 2017). The findings of this study will be validated under diagnostic conditions and reported back to the families in an appropriate clinical and counselling setting.

4.6 Future Work and Recommendations

We acknowledge that certain aspects of this study will require prospective work. This will include validation of identified putative disease-causing variants and possibly, increasing the sample size. While all identified putative disease-causing variants were subjected to careful bioinformatics analyses, the possibility of false-positive results cannot be excluded. Putative disease-causing variants identified in this study were identified based on computational predictions and available knowledge in the literature; therefore, functional studies to elucidate the precise deleterious effect of each variant will be required. Additionally, segregation analysis to establish if variants segregate with disease within a family will be required. In contrast to previous studies in other populations, our study cohort was small, thus, future studies analysing a larger cohort will be required in order to confidently infer these results to the study population.

According to the HGMD and *CHD7* database, a subset of TCS and CHARGE syndrome mutations are large indels and/ or duplications, thus, it is possible that the eight patients without an identified point putative disease-causing variants could have one of these. Therefore, future work analysing large indels and/or duplications, using appropriate techniques such as MLPA, would be required in these patients. Furthermore, it is possible that other genes/genomic regions, not included in our gene panel, could be responsible for the manifestation of facial dysostosis phenotypes. In such cases, whole exome- and genome sequencing should be considered.

4.7 Conclusion

The present study is, to the best of our knowledge, the first study to perform a molecular analysis on TCS, NS, MS, MFDM, BOR- and CHARGE syndrome in South African patients. In our analysis of 15 patients, seven putative disease-causing variants were identified in seven patients clinically diagnosed with TCS and CHARGE syndrome. The overall pick-up rate for this study was therefore 47%. In addition, a VUS predicted to affect splicing was also identified in a single TCS patient. These findings demonstrate that targeted NGS with multigene panel testing is an acceptable diagnostic method which could be successfully implemented for the molecular diagnosis of TCS and CHARGE syndrome in the South African population. The successful implementation of an NGS-based diagnostic approach for TCS and CHARGE syndrome in SA will enable their molecular diagnosis which in turn can inform clinical management and accurate genetic counselling. Although NGS-based approaches are rapidly being incorporated into routine diagnostics in international laboratories, the pace of NGS technologies implementation in developing and under-resourced countries, such as SA, is not equally met. This is due to the high costs, as well as the need for technical and bioinformatics expertise that are associated with this technique. While TCS and CHARGE syndrome, like most inherited diseases, are incurable, studies have described various benefits of receiving a definite diagnosis of incurable disorders (ACMG board of directors, 2015; Kohler *et al.*, 2017). The present study has produced a baseline mutation profile of TCS and CHARGE syndrome in the South African population.

REFERENCES

Journal articles

ACMG board of directors 2015, 'Clinical utility of genetic and genomic services: a position statement of the American College of Medical Genetics and Genomics' *Genet Med*, 17, 6, 506-507.

Aylsworth, AS, Lin, AE & Friedman PA 1991, 'Nager acrofacial dysostosis: male-to-male transmission in 2 families' *Am J Med Genet*, 41, 1, 83-88.

Aziz, N, Zhao, Q & Bry, L 2015, 'College of American Pathologists laboratory standards for next-generation sequencing clinical tests', *Arch Pathol Lab Med*, 139, 4, 481-493.

Bahassi, M & Stambrook, PJ 2014, 'Next-generation sequencing technologies: breaking the sound barrier of human genetics' *Mutagenesis*, 29, 5, 303-310.

Bajpai, R, Chen, DA & Rada-iglesias, A 2010, 'CHD7 cooperates with PBAF to control multipotent neural crest formation' *Nature*, 18, 463, 958-962.

Bartels, CF, Scacheri, C & White, L 2010, 'Mutations in the CHD7 gene: the experience of a commercial laboratory' *Genet Test Mol Biomarkers*, 14, 6, 881-891.

Baudhuin, LM, Lagerstedt, SA & Klee, EW 2015, 'Confirming variants in next-generation sequencing panel testing by Sanger sequencing' *J Mol Diagn*, 17, 4, 456-461.

Beck, TF, Mullikin JC & Biesecker LG 2016, 'Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants' *Clin Chem*, 62, 4, 647-54.

Behjati, S & Patrick, S 2013, 'What is next generation sequencing?', Arch *Dis Child Educ Pract Ed*, 98, 6, 236-238.

Bernier, FP & Caluseriu, O 2012, 'Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome', A*m J Hum Genet*, 90, 5, 925-933.

Berry, G 1889, 'Note on congenital defect (coloboma?) of the lower lid', *R Lond Ophthalmic Rep*, 12, 3, 255-57.

Besser, J, Carleton, HA & Gerner-smidt, P 2017, 'Next-generation sequencing technologies and their application to the study and control of bacterial infections', *Clin Microbiol Infect*, 24, 4, 4335-341.

Blake, KD, Davenport, SL & Hall, BD 1998, 'CHARGE association: An update and review for the primary pediatrician', *Clin Pediatr*, 37, 3, 159-173.

Bouazoune, K & Kingston RE 2012, 'Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders', *Proc Natl Acad Sci*, 109, 47, 19238-19243.

Bowman, M, Oldridge, M & Archer, C 2012, 'Gross deletions in TCOF1 are a cause of Treacher-Collins-Franceschetti syndrome', *Eur J Hum Genet*, 20, 7, 769-777.

Bronner-Fraser, M & Fraser, S 1989, 'Developmental potential of avian trunk neural crest cells in situ', *Neuron*, 3, 6, 755-766.

Carter, H, Douville, C & Stenson, PD 2013, 'Identifying Mendelian disease genes with the variant effect scoring tool', *BMC Genomics*, 14, 3, S3.

Cassina, M, Cerqua, C & Rossi, S 2016, 'A synonymous splicing mutation in the SF3B4 gene segregates in a family with highly variable Nager syndrome', *Eur J Hum Genet*, 25, 3, 371-375.

Chang, H, Huylebroeck, D & Verschueren, K 1999, 'Smad5 knockout mice die at mid-gestation due to multiple embryonic and extraembryonic defects' *Dev Biol*, 382, 8, 375-84.

Chemke, J, Mogilner, BM & Ben-itzhak, I 1988, 'Autosomal recessive inheritance of Nager acrofacial dysostosis' *J Med Genet*, 25, 4, 230-232.

Chen, A, francis, M, Ni , G & Smang, HG 1995, 'Phenotypic manifestations of branchiootorenal syndrome', *Am J Med Genet*, 58, 4, 365-370.

Chen, Y, Guo, L & Lic, L 2018, 'Mutation screening of Chinese Treacher Collins syndrome patients identified novel TCOF1 mutations', *Mol Genet Genomics*, 293, 2, 569-577.

Choi, Y & Chan, AP 2015, 'PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels', *Bioinformatics*, 31, 16, 2745-2747.

Conte, C, D'apice, MR & Rinaldi, F 2011, 'Novel mutations of TCOF1 gene in European patients with Treacher Collins syndrome', *BMC Med Genet*, 12, 2, 125.

Cooper, DN & Krawczak, M 1998, 'The human gene mutation database', *Nucleic Acids Res*, 26, 3, 285–287.

Cooper, GM, stone, EA & Asimenos, G 2005, 'Distribution and intensity of constraint in mammalian genomic sequence', *Genome Res*, 15, 7, 901–913.

Czeschik, JC, Voigt, C, Alanay, Y & Go, D 2013, 'Clinical and mutation data in 12 patients with the clinical diagnosis of Nager syndrome', *Hum Genet*, 132, 8, 885-898.

Dark, M J 2013, 'Whole-genome sequencing in bacteriology: state of the art', *Infect Drug Resist*, 6, 2, 115–123

Dauwerse, JG, Dixon, J & Seland, S 2011, 'Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome', *Nat Genet*, 43, 1, 20-22.

Davydov, EV, Goode, DL, Sirota, M & Huu, E 2010, 'Identifying a high fraction of the human genome to be under selective constraint using GERP++', P*LoS Comput Biol*, 6, e1001025.

Den dunnen, JT 2016, 'Sequence Variant Descriptions: HGVS Nomenclature and Mutalyzer', *Curr Protoc Hum Genet*, 1, 1, 7.

Desmet, FO, Hamroun, D, Lalande, M, Furu, MH 2009', 'Human Splicing Finder: an online bioinformatics tool to predict splicing signals', *Nucleic Acids Res*, 37, 9, e67.

Devotta, A, Juraver-geslin, H & Gonzalez, JA 2016, 'Sf3b4-depleted Xenopus embryos: A model to study the pathogenesis of craniofacial defects in Nager syndrome', *Dev Biol*, 415, 2, 371-382.

Dixon, J, Edwards, SJ & Anderson, I 1997, 'Identification of the complete coding sequence and genomic organization of the Treacher Collins syndrome gene', *Genome Res*, 7, 3, 223–34.

Dixon, J, Ellis, I & Bottani, A 2004, 'Identification of mutations in TCOF1: use of molecular analysis in the pre- and postnatal diagnosis of Treacher Collins syndrome', *Am J Med Genet*, 127, 3, 244-248.

Dixon, J, Jones, NC, Sandel, LL & Goodwill, E 2006, 'Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities', *Proc Natl Acad Sci*, 103, 36, 13403-13408.

Dixon, J, Trainor, P & Dixon, MJ 2007, 'Treacher Collins syndrome', *Orthod Craniofac Res*, 10, 5, 88-95.

Donnai, D, Hughes, HE, Winter, RM 1987, 'Postaxial acrofacial dysostosis (Miller) syndrome', *J Med Genet*, 24, 422-425.

Duley, JA, Henman, MG, Carpenter, KH 2016, 'Elevated plasma dihydroorotate in Miller syndrome: Biochemical, diagnostic and clinical implications, and treatment with uridine', *Mol Genet Metab*, 119, 83-90.

Dunne, JR, Westblade LF & Ford, B 2012, 'Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory', *Eur J Clin Microbiol Infect Dis*, 31, 8, 1719-1726.

Edwards, SJ, Gladwin, AJ & Dixon, MJ 1997, 'The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon', *Am J Hum Genet*, 60, 3, 515-524.

Ellis, PE, Dawson, M & Dixon, MJ 2002, 'Mutation testing in Treacher Collins Syndrome', *J Orthod*, 29, 4, 293-297.

Engelen, E, Akinci, U & Bryne JC 2011, 'Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes', *Nat Genet*, 43, 6, 607-611.

Estrada, KD, Retting, KN, Chin, AM & Cho, D 2013, 'Smad6 is essential to limit BMP signaling during cartilage development', *J Bone Miner Res*, 26, 10, 2498-2510.

Evans, DR & Guy, HI 2004, 'Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway', *J Biol Chem*, 279, 32, 33035-33038.

Fang, J, Uchiumi, T, yagi, M, Hong, E 2012, 'Protein instability and functional defects caused by mutations of dihydro-orotate dehydrogenase in Miller syndrome patients', *Biosci Rep*, 32, 6, 631-639.

Fazen, LE, Elmore, J & Nadler, HL 1967, 'Mandibulo-facial dysostosis (Treacher Collins)', *Am J Dis Child*, 113, 4, 405-406.

Franceschetti, A. & Klein, D 1949, 'The mandibulofacial dysostosis; a new hereditary syndrome', *Acta Ophthalmol*, 27, 2, 143-224.

Fraser, FC, Sproule, J & Halal, F 1980, 'Frequency of the branchio-oto-renal (BOR) syndrome in children with profound hearing loss', *Am J Med Genet*, 7, 3, 341-349

Fukushima, R, Kanamori, S & Hirashiba, M 2007, 'Teratogenicity study of the dihydroorotate-dehydrogenase inhibitor and protein tyrosine kinase inhibitor Leflunomide in mice', *Reprod Toxicol*, 24, 3-4 310-316.

Fukushima, R, Kanamori, S, Hirashiba, M 2009, 'Inhibiting the teratogenicity of the immunosuppressant leflunomide in mice by supplementation of exogenous uridine', *Toxicol Sci*, 108, 2, 419-426.

Genée, E 1969, 'An extensive form of mandibulo-facial dysostosis', *J Genet Hum*, 17, 1, 45-52.

Grabb, WC 1965, 'The first and the second branchial arches', *Plast Reconstr Surg*, 36, 5, 485-508.

Guion-almeida, ML, Zechi-ceide, RM & Vendramini S 2006, 'A new syndrome with growth and mental retardation, mandibulofacial dysostosis, microcephaly, and cleft palate', *Clin Dysmorphol*, 15, 3, 171–174.

Hall, BD 1989, 'Nager acrofacial dysostosis: autosomal dominant inheritance in mild to moderately affected mother and lethally affected phocomelic son', *Am J Med Genet*, 33, 3, 394-397.

Halonen, K, Hukki, J, Arte, S & Hurmerinta, K 2006, 'Craniofacial structures and dental development in three patients with Nager syndrome', *J Craniofac Surg* 17, 6, 1180-1187.

Hancock-hanser, BL, Frey, A, Leslie, MP & Ng, R 2013, 'Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics', *Mol Ecol Resour*, 13, 2, 254-268.

Horiuchi, K, Ariga, T, & Fujioka, H 2005, 'Mutational analysis of the TCOF1 gene in 11 Japanese patients with Treacher Collins Syndrome and mechanism of mutagenesis', *Am J Med Genet*, 134, 4, 363-367.

Hu, J & Ng PC 2013, 'SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins', *PLoS One*, e77940.

Isaac, C, Marsh, KL & Paznekas, WA 2000, 'Characterization of the nucleolar gene product, treacle, in Treacher Collins syndrome', *Mol Biol Cell*, 11, 9, 3061-3071.

Janssen, N, Bergman, JE, Swertz, MA & Maggy, Y 2012, 'Mutation update on the CHD7 gene involved in CHARGE syndrome', *Hum Mutat*, 33, 8, 1149-1160.

Johnson, D, Morrison, N, Grant, L & Gu, D 2006, 'Confirmation of CHD7 as a cause of CHARGE association identified by mapping a balanced chromosome translocation in affected monozygotic twins', *J Med Genet*, 43, 3, 280–284.

Jones, KL, Smith, DW, Harvey, MA & Smith, W 1975, 'Older paternal age and fresh gene mutation: data on additional disorders', *J Pediatr*, 86, 1, 84-88.

Jones, NC, Farlie, PG & Minichiello, J 1999, 'Detection of an appropriate kinase activity in branchial arches I and II that coincides with peak expression of the Treacher Collins syndrome gene product, treacle', *Hum Mol Genet*, 8, 12, 2239-2245.

Jones, NC, Lynn, ML, Gaudenz, K & Smith, DW 2008, 'Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function', *Nat Med*, 14, 2, 125-133

Karczewski, KJ, Francioli, LC, Tiao, G & Jones, E 2019, 'Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes' *BioRxiv* 53121, https://www.biorxiv.org/content/10.1101/531210v2

Kavadia, S, Kaklamanos, E & Antoniadis, K 1993, 'Nager syndrome (preaxial acrofacial dysostosis): a case report', *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 97, 6, 732-738.

Kawabata, M, Inoue, H, Hanyu, A & Lorry, H 1998, 'Smad proteins exist as monomers in vivo and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors', *EMBO J*, 17, 14, 4056-4065.

Kennedy, SJ & Teebi, AS 2004, 'Newly recognized autosomal recessive acrofacial dysostosis syndrome resembling Nager syndrome', *Am J Med Genet*, 15, 1, 73-76.

Kircher, M, Witten, DM, Jain, P & Jui, R 2014, 'A general framework for estimating the relative pathogenicity of human genetic variants', *Nat Genet*, 46, 1, 310-315.

Klingenberg, C & Andersen WH 2008, '[The CHARGE syndrome]', *Tidsskr Nor Laegeforen*, 128, 1, 1401-1405.

Kohler, JN, Turbitt, E & Bieseker, B 2017, 'Personal utility in genomic testing: a systematic literature review', *Eur J Hum Genet*, 25, 6, 662-668.

Kopanos, C, Tsiolkas, V & Kouris, A 2018, 'VarSome: The Human Genomic Variant Search Engine', *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bty897

Kurosaki, T & Maquat, LE 2016, 'Nonsense-mediated mRNA decay in humans at a glance', *J Cell Sci*, 129, 3, 461-467.

Laferté, A, Favry, E, Sentenac, A, Jim, A 2006, 'The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components', *Genes Dev*, 15, 20, 2030–2040.

Lafontaine, DL & Tollervey D 2001, 'The function and synthesis of ribosomes', *Nat Rev Mol Cell Biol*, 2, 7, 514-520.

Lander, ES, Linton, LM & Birren, B 2001, 'Initial sequencing and analysis of the human genome', *Nature*, 409, 6822, 860-921.

Lines, MA, Huang, L, Schwartzentruber, J & Josh, G 2012, 'Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly', *Am J Hum Genet*, 90, 2, 369-77.

Löffler, M, Fairbanks, LD, Zameitat, E & Smith, DW 2005, 'Pyrimidine pathways in health and disease', *Trends Mol Med*, 11, 9, 430-437.

Lowry, RB 1977, 'The Nager syndrome (acrofacial dysostosis): Evidence for autosomal dominant inheritance', *Birth Defects Orig Artic Se*r, 13, 9, 195-220.

Lund, C, Vestergaard, EM & Christensen, R 2016, 'Prenatal diagnosis of Nager syndrome in a 12-week-old fetus with a whole gene deletion of SF3B4 by chromosomal microarray', *Eur J Med Genet*, 59, 1, 48-51.

Luquetti, DV, Hing, AV, Rieder, MJ & Shiang, H 2013, 'Mandibulofacial dysostosis with microcephaly caused by EFTUD2 mutations: expanding the phenotype', *Am J Med Genet*, 161, 1, 108–113.

Lykke-Andersen, S & Jensen, TH 2015, 'Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes', *Nat Rev Mol Cell Biol*, 16, 11, 665-677.

Mardis, ER 2011, 'A decade's perspective on DNA sequencing technology', *Nature*, 470, 7333, 198-203.

Mardis, ER 2008, 'Next-generation DNA sequencing methods', A*nnu Rev Genomics Hum Genet*, 9, 1, 387-402.

Marres, HA 2002, 'Hearing loss in the Treacher-Collins syndrome', *Adv Otorhinolaryngol*, 61, 3, 209-215.

Marsella, P, Scorpecci, A & Pacifico, C 2011, 'Bone-anchored hearing aid (Baha) in patients with Treacher Collins syndrome: tips and pitfalls', *Int J Pediatr Otorhinolaryngol*, 75, 10, 1308-1312.

Marsh, KL, Dixon, J & Dixon, MJ 1998, 'Mutations in the Treacher Collins syndrome gene lead to mislocalization of the nucleolar protein treacle', *Hum Mol Genet*, 7, 1795-1800.

Martelli-junior, H, Coletta, RD, Miranda, RT & Mack, G 2009, 'Orofacial features of Treacher Collins syndrome', *Medicina Oral Patologia Oral y Cirugia Bucal*, 14, 5, 344-348.

Masotti, C, Ornelas, CC & Splendore-gordonos, A 2009, 'Reduced transcription of TCOF1 in adult cells of Treacher Collins syndrome patients', *BMC Med Genet*, 14, 2, 136.

Maxam, AM & Gilbert, W 1977, 'A new method for sequencing DNA', *Proc Natl Acad Sci U.S.A*, 74, 2, 560-564.

Mcdonald, MT & Gorski, JK 1993, 'Nager acrofacial dysostosis', *J Med Genet*, 30, 3, 779-782.

Miller, M, Fineman, R & Smith, DW 1979, 'Postaxial acrofacial dysostosis syndrome', *J Med Genet*, 95, 4, 970-975.

Miller, SA, Dykes, DD & Polesky, HF 1988, 'A simple salting out procedure for extracting DNA from human nucleated cells', *Nucleic Acids Res*, 16, 3, 1215.

Morgan, TH 1910, 'Sex Limited Inheritance in Drosophila', *Science*, 32, 812, 120-122.

Mort, M, Sterne-weiler, T & LI, B 2014, 'MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing', *Genome Biol*, 15, R19.

Mu, W, Lu, HM & Chen, J 2016, 'Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing', *J Mol Diagn*, 18, 6, 923-932.

Nager, F & Reynier, D 1948, 'Das Gehorogan bei den angeborenen kopfmissnildungen', *Pract Otorhinolaryngol*, 10, 1, 1-128.

Neu-yilik, G, Amthor, B & Gehring, NH 2011, 'Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon', *RNA*, 17, 5, 843-54.

Ng, SB, Buckingham, KL & Lee, C 2010, 'Exome sequencing identifies the cause of a mendelian disorder', *Nat Genet*, 42, 1, 30-35.

Noden, DM & Trainor, PA 2005, 'Relations and interactions between cranial mesoderm and neural crest populations, *J Anat*, 207, 5, 575-60.

Pagon, RA, Graham, JM & Zonana, J 1981, 'Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association', *J Pediatr*, 9, 2, 223-237.

Petit, F, Escande, F, Jourdain, A & Mash, G 2014, 'Nager syndrome: confirmation of SF3B4 haploinsufficiency as the major cause', *Clin Genet*, 86, 3, 246-251.

PIerides, AM, Athanasiou, Y, Demetriou, K & Yang, Z 2002, 'A family with the branchio-oto-renal syndrome: clinical and genetic correlations', *Nephrol Dial Transplant*, 17, 6,1014-1018.

Posnick, JC & Ruiz RL 2000, 'Treacher Collins syndrome: current evaluation, treatment, and future directions', *Cleft Palate Craniofacial J*, 37, 5, 434.

Quang, D, Chen, Y & Xie, X 2014, 'DANN: a deep learning approach for annotating the pathogenicity of genetic variants', *Bioinformatics*, 31, 5, 761-763.

Rainger, J, Bengani, H & Campbell, L 2012, 'Miller syndrome represent a clinically and biochemically distinct subgroup of postaxial acrofacial dysostosis associated with partial deficiency of DHODH', *Hum Mol Genet*, 21, 18, 3969-3983.

Rawls, J, Knecht, W & Diekert, K 2000, 'Requirements for the mitochondrial import and localization of dihydroorotate dehydrogenase', *Eur. J. Biochem*, 267, 7, 2079-2087.

Rehm, HL, Bale, SJ & Bayrak-toydemir, P 2013, 'ACMG clinical laboratory standards for next-generation sequencing', *Genet Med*, 15, 9, 733-47.

Reva, B, Antipin, Y & Sander, C 2017, 'Determinants of protein function revealed by combinatorial entropy optimization', *Genome Biol*, 8, 11, R232.

Richards, S, Aziz, N & Bale, S 2015, 'ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology',, *Genet Med*, 17, 5, 405-24.

Richter, T, Nestler-parr, S, Babel, AR & Mang, T 2015, 'Rare disease terminology and definitions-a systematic global review: report of the ISPOR rare disease special interest group', *Value in Health*, 18, 6, 906-914.

Robinson, JT, Thorvaldsdóttir, H, Winckler, W & Wang, J 2011, 'Integrative genomics viewer', *Nat Biotechnol*, 29, 1, 24–26.

Rubenstein, JL, Shimamura, K & Martine, S 1998, 'Regionalization of the prosencephalic neural plate', *Annu Rev Neurosci* , 21, 1, 445–77.

Sakai, D, Dixon, J, Dixon, MJ & Su, H 2012, 'Mammalian neurogenesis requires Treacle-Plk1 for precise control of spindle orientation, mitotic progression, and maintenance of neural progenitor cells', *PLoS Genet*, 8, e1002566.

Sanger, F, Nicklen, S & Coulson, AR. 1977, 'DNA sequencing with chain-terminating inhibitors', *Proc Natl Acad Sci U S A*, 74, 12, 5463-5467.

Schaefer, E, Collet, C, Genevieve, D & Mu, M 2014, 'Autosomal recessive POLR1D mutation with decrease of TCOF1 mRNA is responsible for Treacher Collins syndrome', *Genet Med*, 16, 9, 720-724.

Schwarz, JM, Cooper, DN & Schuelke, M 2014, 'MutationTaster2: mutation prediction for the deep-sequencing age', *Nat Methods*, 11, 4, 361-362.

Shihab, HA, Gough, J & Cooper, DN 2013, 'Predicting the functional consequences of cancer-associated amino acid substitutions', *Bioinformatics*, 29, 12, 1504-1510.

Sikkema-raddatz, B, Johansson, LF & de Boer, EN 2013, 'Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics', *Hum Mutat*, 34, 7, 1035-1042.

Sim, NL, Kumar, P & Hu, J 2012, 'SIFT web server: predicting effects of amino acid substitutions on proteins', *Nucleic Acids Res*, 40, 4, W452-W457.

Slingenberg, B 1908, 'Missbildungen von extremitaten. Virchows', *Arch Pathol Anat*, 193, 2, 1–91

Splendore, A, Silva, EO & Alonso, LG 2000, 'High mutation detection rate in TCOF1 among Treacher Collins syndrome patients reveals clustering of mutations and 16 novel pathogenic changes', *Hum Mutat*, 16, 4, 315-322.

Splendore, A, Jabs, EW & Passos-bueno, MR 2002, 'Screening of TCOF1 in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle', *J Med Genet*, 39, 7, 493–495.

Strom, SP, Lee, H, & Das, K 2014, 'Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory', *Genet Med*, 16, 7, 510-515.

Szymanski, M. Barciszewska, MZ & Erdmann, VA 2002, 'Ribosomal RNA Database', *Nucleic Acids Res*, 30, 32, 176-178.

Talerico, M & Berget, SM 1990, 'Effect of 5' splice site mutations on splicing of the preceding intron', *Mol Cell Biol*, 10, 12, 6299-6305.

Tayoun, A, Pesaran, NT & Distefano, MT 2018, 'Sequence Variant Interpretation Working Group (ClinGen SVI) (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion', *Human Mut*, 39, 1, 1517-152.

Teber, OA, Gillessen-kaesbach, G & Fischer, S 2004, 'Genotyping in 46 patients with tentative diagnosis of Treacher Collins syndrome revealed unexpected phenotypic variation', *Eur J Hum Genet*, 12, 11, 879-90.

Thompson, JT, Anderson, PJ & David, DJ 2009, 'Treacher Collins syndrome: protocol management from birth to maturity', *J Craniofac Surg*, 20, 6, 2028-2035.

Thomson, A 1847, 'Notice of Several Cases of Malformation of the External Ear, Together with Experiments on the State of Hearing in Such Persons', *Mon J Med Sci*, 10, 2, 729-738.

Toynbee, J 1847, 'Description of a Congenital Malformation in the Ears of a Child', *Mon J Med Sci*, 1, 10, 738-740.

Trainor, PA & Andrews, BT 2013, 'Facial dysostoses: Etiology, pathogenesis and management', *Am J Med Genet*, 163, 4, 283-94.

Trainor, PA, Dixon, J & Dixon, MJ 2009, 'Treacher Collins syndrome: etiology, pathogenesis and prevention', *Eur J Hum Genet*, 17, 3, 275-283.

Treacher, CE 1900, 'Case with symmetrical congenital notches in the outer part of each lowerlid and defective development of the malar bones', T*rans Amer Opht Society UK*, 20, 3, 90.

Treacher Collins syndrome collaborative group 1996, 'Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome', *Nat Genet*, 12, 3, 130-136.

Valdez, B C, Henning, D, So, RB & Dixon, J 2004, 'The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor', *Proc Natl Acad Sci*, 101, 29, 10709-10714.

Venter, JC, Adams, MD, Myer, EW & Su, I 2001, 'International Human Genome Sequencing Consortium', *Nature*, 409, 77, 860-921.

Verloes, A 2005, 'Updated diagnostic criteria for CHARGE syndrome: A proposal', *Am J Med Genet*, 133, 3, 306-308.

Vincent, M, Geneviève, D & Ostertag, A 2016, 'Treacher Collins syndrome: a clinical and molecular study based on a large series of patients', *Genet Med*, 18, 1, 49-56

Vissers, LE, van Ravenswaaij, CM, Admiraal, R Vries, BB & 2004, 'Mutations in a new member of the chromodomain gene family cause CHARGE syndrome', *Nat Genet*, 36, 9, 955-957.

Waggoner, DJ, Ciske, DJ, Dowton, SB & Watson, MS 1999, 'Deletion of 1q in a patient with acrofacial dysostosis', *Am J Med Genet*, 82, 4, 301-304.

Wang, K, Li, M & Hakonarson, H 2010, 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Res*, 38, 16, e164.

Watanabe, H, Shionyu, M, Kimura, T & Kimata, K 2007, 'Splicing factor 3b subunit 4 binds BMPR-IA and inhibits osteochondral cell differentiation', *J Biol Chem*, 282, 28, 20728-20738.

Weidemann, HR 1973, 'Malformation-retardation syndrome with bilateral absence of the 5th rays in both hands and feets, cleft palate, malformed ears and eyelids, radioulnar synostosis', *Klin Padiatr*, 185, 3, 181-186.

Weiner, AM, Scampoli, NL & Calcaterra, NB 2012, 'Fishing the molecular bases of Treacher Collins syndrome', *PLoS One*, 7, 1, 25.

Wieczorek, D 2013, 'Human facial dysostoses', *Clin Genet*, 83, 499-510.

Wright, CF, Fitzgerald, TW & Jones, WD 2015, 'Molecular diagnosis', *Lancet*, 385, 9975, 1305-1314.

Winokur, ST & Shiang R 1998, 'The Treacher Collins syndrome (TCOF1) gene product, treacle, is targeted to the nucleolus by signals in its C-terminus', *Hum Mol Genet*, 7, 12, 1947-1952.

Yang, SM, Hou, ZH, Yang, G & Zhang, JS 2009, 'Chondrocyte-specific Smad4 gene conditional knockout results in hearing loss and inner ear malformation in mice', *Dev Dyn*, 238, 8, 1897-1908.

Yao, Y, Yamamoto, K, Nishi, Y & Muramatsu, M 1996, 'Mouse RNA polymerase I 16-kDa subunit able to associate with 40-kDa subunit is a homolog of yeast AC19 subunit of RNA polymerases I and III', *J Biol Chem*, 271, 51, 32881-32885.

Yelick, PC & Trainor, PA 2015, 'Ribosomopathies: Global process, tissue specific defects', *Rare Dis*, 3, 1, e1025185.

Zentner, GE, Layman, W S, Martin, DM & Scapheri, 2010, 'Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome', *Am J Med Genet*, 152, 3, 674–686.

Books

Adzhubei, I, Jordan, DM & Sunyaev, SR 2013, 'Predicting functional effect of human missense mutations using PolyPhen-2', *Curr Protoc Hum Genet*, Chapter 7, Unit 7.20.

Field, MJ, Boat, TF & Editors 2010, 'Rare Diseases and Orphan Products: Accelerating Research and Development', available from: https://www.ncbi.nlm.nih.gov/books/NBK56184/, accessed 07 November 208

Katsanis, SH & Jabs, EW 2004, 'Treacher Collins Syndrome', available from: https://www.ncbi.nlm.nih.gov/books/NBK1532/, accessed 14 February 2019

Lalani,SR, Hefner, MA & Belmont, JW 2006, 'CHARGE syndrome', available from http://www.ncbi.nlm.nih.gov/books/NBK1117/, accessed 19 January 2019

Electronic sources

Agilent 2016. SureSelect QXT Target Enrichment for Illumina Multiplexed Sequencing. Retrieved from: https://www.agilent.com/cs/library/usermanuals/public/G9681-90000.pdf , accessed 07 September 2017

CHD7 mutation database, https://molgenis51.gcc.rug.nl/, accessed 09 March 2019

Ensemble, http://grch37.ensembl.org/index.html , accessed 22 February 2019

Human Gene Mutation Database, http://www.hgmd.cf.ac.uk/ac/index.php, accessed 09 March 2019

IGV, https://software.broadinstitute.org/software/igv/, accessed 16 September 2018

Mutalyzer, https://mutalyzer.nl/, accessed 12 January 2019

Orphanet database, https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=EN, accessed 06 January 2019

Primer3 (v 0.4.0), http://frodo.wi.mit.edu/, accessed 28 November 2018

Refworks, https://www.refworks.com/refworks2/

APPENDICES

Appendix A: Plagiarism declaration

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Patricia Nevondwe (Student number: 1619460) am a student registered for the degree of MSc in Medicine in Human Genetics in the academic year 2019.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: [signature] Date: 20/03/2019

1

Appendix B1: Facial dysostosis tick sheet

Name: _____ **Male** ☐ **Female** ☐

DOB: _____ **Caucasian** ☐ **Black** ☐

Hospital: _____ **Indian** ☐ **Mixed ancestry** ☐

Hospital Number: _____ **Participant number:** _____

Attending Clinician: _____ **Possible diagnosis:** _____

GROWTH

Height:	<3rd centile	☐	3rd – 97th centile	☐	>97th centile	☐	SD ____
Weight:	<3rd centile	☐	3rd – 97th centile	☐	>97th centile	☐	SD ____
OFC:	<3rd centile	☐	3rd – 97th centile	☐	>97th centile	☐	SD ____

FACIAL FEATURES

General:	Asymmetry 0000324	☐					
Eyes/ periorbital region:	DSPF 0000494	☐	Ptosis 0000508	☐	Lower lid coloboma 0000652	☐	Ectropion 0000656 ☐
	Ocular colobomas 0000589	☐					
Mouth:	Cleft palate 0000175	☐			Mandibular hypoplasia 0000330	☐	
Nose/midface:	Malar flattening 0100846	☐			Choanal atresia 0000416	☐	
Ears:	Low set 0000369	☐	Microtia 0008551	☐	Ear pits 0030025	☐	Ear tags 0030021 ☐
	External canal atresia 0000402	☐					
Hair:	Lateral cheek hair 0009554	☐					

Other: _____

CARDIAC ANOMALIES

ASD ☐ 0010445 | VSD ☐ 0001629 | PDA ☐ 0001643 | AS ☐ 0001650 | PS ☐ 0001642

Other:_____

SKELETAL ANOMALIES

Upper limbs:	Thumb hypoplasia ☐ 0005699	Thumb aplasia ☐ 0009777	Radial hypoplasia ☐ 0002984
	Radio-ulnar synostosis ☐ 0002174	Limb reduction defect ☐	Syndactyly ☐ 0001159
	Absent 5th digit ☐	Ulnar hypoplasia ☐ 0005033	Olidodactyly ☐ 0001588
	Clinodactyly ☐ 0012165		
Spine:	Kyphosis ☐ 0002808	Scoliosis ☐ 0002650	Vertebral defects ☐ 0005640

GENITALIA

Undescended testes ☐ 0000028 | Small penis ☐ 0000054 | Hypospadias ☐ 0000047 | Hypoplastic labia minora ☐ 0000064

NEURODEVELOPMENT / CENTRAL NERVOUS SYSTEM

Normal ☐ | Mild ID ☐ 0001256 | Moderate ID ☐ 0002342 | Severe ID ☐ 0010864

SENSORY DEVELOPMENT

Vision:	Myopia ☐ 0000545	Strabismus ☐ 0000486	
Hearing:	Hearing loss ☐	Conductive ☐ 0000405	Sensorineural ☐ 0000407

GASTROINTESTINAL SYSTEM

Gastro-oesophageal reflux ☐ 0002020 | Malrotation ☐ 0002566 | Anal atresia ☐ 0002023

OTHER SIGNIFICANT ABNORMALITIES

Renal anomalies

Appendix B2: Consent forms used in this study

B2.1: CONSENT FORM for adults

"The purpose of this study, procedures to be followed, risks and benefits, have been explained to me. I have been allowed to ask the questions I have, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this consent form and agree to participate in this research study with the understanding that I may withdraw at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one."

_____ _____

Signature of Participant Date

_____ _____

Signature of Person Obtaining Consent Date

B2.2: CONSENT FORM for children

"The purpose of this study, procedures to be followed, risks and benefits, have been explained to me and my child. I have been allowed to ask the questions I have, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this consent form and agree to allow my child to participate in this research study with the understanding that I may withdraw my child at any time. We have discussed the study with my child to the best of our ability, who agrees to be in the study. We have explained to my child that he/she may approach the researchers to be re-consented when he/she turns 18 if he/she wishes and that he/she may withdraw from the study at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one."

_____	_____
Name of Child	Date
_____	_____
Signature of Parent or Legal Guardian	Date
_____	_____
Signature of Person Obtaining Consent	Date

B2.3: CONSENT FORM for ADULTS for DNA storage for possible future use

I, the undersigned, have been fully informed as to the procedure to be followed and have been given a description of the risks and benefits of having my DNA extracted and stored for possible further research.

In signing this form, I agree to have my DNA stored at the Division of Human Genetics in the Molecular Genetics Laboratory.

I understand that most of the DNA that is extracted from my blood will be stored in Johannesburg for an indefinite period of time and may be used for further genetic testing for genetic- or developmental disorder-related research only. This may include sending a portion of the sample overseas for further tests. Any further testing on these samples will be monitored by the Wits Ethics Committee.

I understand that if I have any questions at any time, they will be answered.

I understand that I can choose to have my DNA destroyed at any time.

I understand that confidentiality will be maintained at all times through the use of a code and numbering system.

A signed copy of this consent form will be made available to me if I want one.

The information around the blood and DNA samples taken from me is clear.

YES ☐ NO ☐

The purpose of this consent is for me to inform the study what they can or cannot do with these samples.

YES ☐ NO ☐

I acknowledge that all procedure/tests on the stored blood and DNA samples have been or will be approved by the Human Research Ethics Committee of the University of the Witwatersrand.

YES ☐ NO ☐

I am in agreement that my DNA may be stored and used for the purposes described above.

YES ☐ NO ☐

I am in agreement that the data generated from my DNA may be made available in a public domain without any identifiers.

YES ☐ NO ☐

I agree that a small bit of my DNA may be sent out of the country if the research cannot easily be done in South Africa.

YES ☐ NO ☐

I understand that every time a new study is done on my DNA, permission will be obtained from the ethics committee for the study to make sure that it is used only for the purposes stated above.

YES ☐ NO ☐

I understand that I may not benefit directly from the research done on my DNA and my family's DNA. I understand that the tests will reveal information about my biological ancestors, about my ability to respond to certain drugs and thatgenetic variants that may affect my risk for some diseases may be identified.

YES ☐ NO ☐

If this research does identify information that would be important or of benefit to my family, or me I would like to be contacted to receive the information.

YES ☐ NO ☐

I understand that my family and I may withdraw from the study at any time.

YES ☐ NO ☐

RESEARCHER:

Printed Name Signature Date

PARTICIPANT:

_____

Printed name Signature Date

WITNESS:

Printed Name Signature Dat

B2.4: CONSENT FORM for CHILDREN for DNA storage for possible future use

I, the undersigned, have been fully informed as to the procedure to be followed and have been given a description of the risks and benefits of having my child's blood sampled for the purpose of DNA extraction and storage for possible further research.

In signing this form, I agree to have my child's DNA stored at the Division of Human Genetics in the Molecular Genetics Laboratory.

I understand that most of the DNA that is extracted from my child's blood will be stored in Johannesburg for an indefinite period of time and may be used for further genetic testing for genetic- or developmental disorder-related research only. This may include sending a portion of the sample overseas for further tests. Any further testing on these samples will be monitored by the Wits Ethics Committee. I understand that that my child may approach the researchers to be re-consented when he/she turns 18 if he/she wants to and that he/she may withdraw from the study at any time.

I understand that if I have any questions at any time, they will be answered.

I understand that I can choose to have my child's DNA destroyed at any time.

I understand that confidentiality will be maintained at all times through the use of a code and numbering system.

I have discussed the study with my child, who agrees to the DNA storage.

A signed copy of this consent form will be made available to me if I want one.

The information around the blood and DNA samples taken from my child is clear.

YES ☐ NO ☐

The purpose of this consent form is for me to inform the study what they can or cannot do with these samples.

YES ☐ NO ☐

I acknowledge that all procedure/tests on the stored blood and DNA samples have been or will be approved by the Human Research Ethics Committee of the University of the Witwatersrand.

YES ☐ NO ☐

I am in agreement that my child's DNA may be stored and used for the purposes described above.

YES ☐ NO ☐

I am in agreement that the data generated from my child's DNA may be made available in a public domain without any identifiers.

YES ☐ NO ☐

I agree that a small bit of my child's DNA may be sent out of the country if the research cannot easily be done in South Africa.

YES ☐ NO ☐

I understand that every time a new study is done on my child's DNA, permission will be obtained from the ethics committee for the study to make sure that it is used only for the purposes stated above.

YES ☐ NO ☐

I understand that my child is unlikely to benefit directly from the research done on the DNA. I understand that the tests will reveal information about my child's biological ancestors, about my child's ability to respond to certain drugs and that genetic variant that may affect my child's risk for some diseases may be identified.

YES ☐ NO ☐

I understand that my child may withdraw from the study at any time.

YES ☐ NO ☐

CHILD:

Printed Name

Printed Name Signature Date

PARENT OR LEGAL GUARDIAN:

_____

Printed name Signature Date

WITNESS:

Printed Name Signature Date

B2.5: CONSENT for photos

Patient Consent for Medical Photography

Patient name: _____ Date: _____

☐ check here if minor or unable to provide consent

I consent for medical photographs to be made of me or my child (or person for whom I am legal guardian). I understand that the information may be used in my medical record, for purposes of medical teaching, or for publication in medical textbooks or journals as I have designated below. By consenting to these medical photographs I understand that I will not receive payment from any party. Refusal to consent to photographs will in no way affect the medical care I will receive. If I have any questions or wish to withdraw my consent in the future I may contact:

_____

By signing this form below I confirm that this consent form has been explained to me in terms which I understand.

1) I consent for these photographs to be used in medical publications, including medical journals, textbooks, and electronic publications. I understand that the image may be seen by members of the general public, in addition to scientists and medical researchers that regularly use these publications in their professional education. Although these photographs will be used without identifying information such as my name, I understand that it is possible that someone may recognize me. I also agree for my image to be shown for teaching purposes and to be used for my medical record.

_____ (Signature) _____ (Witness)

2) I agree for my image to be shown for teaching purposes **AND** to be used for my medical record but **NOT FOR** medical publication:

_____ (Signature) _____ (Witness)

3) I agree to use of my image for medical records **ONLY**:

_____ (Signature) _____ (Witness)

For patients between ages 7 and 18 years, a signature below indicates that the information in this consent form has been explained to me, and I assent to use of my images as outlined above:

_____ _____

(Signature of patient) (Witness)

Appendix B3: Ethics certificates

B3.1: Ethics certificate for the broader project

R14/49 Dr Nadia Carstens

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M160830

NAME: (Principal Investigator)	Dr Nadia Carstens
DEPARTMENT:	School of Pathology Division of Human Genetics National Health Laboratory Service
PROJECT TITLE:	Using Next-Generation Technologies to Investigate the Genetic Aetiology of Development Disorders
DATE CONSIDERED:	26/08/2016
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	14/10/2016

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in August and will therefore be due in the month of August each year.

Principal Investigator Signature

Date 18 October 2016

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

B3.2: Ethics certificate for the present study

R14/49 Miss Patracia Nevondwe et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M170760

NAME: (Principal Investigator)	Miss Patracia Nevondwe et al
DEPARTMENT:	School of Pathology Division of Human Genetics
PROJECT TITLE:	Investigating the Genetic Aetiology of Three Facial Dysostoses in South Africa
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M160830 Dr Nadia Carstens
SUPERVISOR:	Nadia Carstens
APPROVED BY:	Prof P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	19/07/2017

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting `the study was initially reviewed. In this case, the study was initially reviewed in July and will therefore be due in the month of July each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature _____ Date _____

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix C1: A list of all variants called in each patient

Table C1.1: A list 16 variants called in FRASC13

Chr	Star	End	Ref	Alt	location	Gene	dbSNP	ClinVar	gnomAD Exomes All (MAF)	gnomAD Genomes All (MAF)	Mode of inheritance
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754229	149754229	C	T	exonic	*TCOF1*	rs75181211	Likely benign	0.0027	0.0102	Het
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	Likely benign	0.1416	0.1838	Hom
chr5	149755340	149755340	G	T	exonic	*TCOF1*	rs7701163	Likely benign	0.0663	0.0951	Het
chr5	149755421	149755421	A	G	exonic	*TCOF1*	rs2071239	Likely benign	0.1428	0.1877	Hom
chr5	149755845	149755845	A	G	exonic	*TCOF1*	rs34796297	Likely benign	0.0051	0.0198	Het
chr5	149759096	149759096	T	C	exonic	*TCOF1*	rs7713638	Likely benign	0.1428	0.1886	Het
chr5	149759201	149759201	C	T	exonic	*TCOF1*	rs114689020	Likely benign	0.0068	0.021	het
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Hom
chr5	149769586	149769586	G	C	exonic	*TCOF1*	.	.	.	.	Het
chr6	43484855	43484855	C	T	exonic	*POLR1C*	rs138184356	.	0.0006	0.0018	Het
chr8	61765560	61765560	G	A	exonic	*CHD7*	rs2068096	Benign	0.0613	0.1243	Het
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign	0.3943	0.3567	Hom
chr8	72129009	72129009	G	A	exonic	*EYA1*	rs4738118	Benign	0.218	0.1508	Het
chr8	72267083	72267083	G	C	exonic	*EYA1*	rs1445404	Benign	0.0134	0.053	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other	0.5531	0.5464	Het

Table C1.2. A list 31 variants called in FRASC23

Chr	Start	End	Ref	Alt	location	Gene	dbSNP	ClinVar	gnomAD Exome_All (MAf)	gnomAD Genome_All (MAF)	Mode of inheritance
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149776232	149776232	C	T	exonic	*TCOF1*	rs15251	Benign	0.2598	0.219	Het
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	Hom
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Hom
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign	0.3943	0.3567	Hom
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Hom
chr8	72127563	72127563	G	A	intronic	*EYA1*	rs7846086	.	.	0.3158	Hom
chr8	72127764	72127764	C	A	intronic	*EYA1*	rs3735935	.	0.3774	0.3171	Hom
chr8	72129009	72129009	G	A	exonic	*EYA1*	rs4738118	Benign	0.218	0.1508	Hom
chr8	72211834	72211834	A	C	intronic	*EYA1*	rs3779747	.	0.4165	0.3979	Het
chr13	28239940	28239940	G	C	exonic	*POLR1D*	rs11029	.	0.3618	0.2973	Het
chr13	28239970	28239970	G	A	exonic	*POLR1D*	rs14105	.	0.4392	0.3872	Het
chr15	67008708	67008708	G	A	intronic	*SMAD6*	rs2278603	.	0.1198	0.1215	Het
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	Het
chr15	67008771	67008771	C	T	exonic	*SMAD6*	rs200374822	.	0.0009	.	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	Benign	0.5531	0.5464	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42931596	42931596	G	A	intronic	*EFTUD2*	rs369476889	.	0.0033	9.70E-05	Het
chr17	42937749	42937749	C	G	intronic	*EFTUD2*	rs9903106	.	0.1363	0.1377	Het
chr17	42937964	42937965	TT	-	intronic	*EFTUD2*	rs757112069	.	0.0733	0.0018	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42949808	42949808	G	A	intronic	*EFTUD2*	rs11654183	Other	0.2687	0.278	Het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	Other	0.5609	0.6292	Hom
chr18	46474746	46474746	C	T	intronic	*SMAD7*	rs3736242	.	0.2217	0.2112	Het
chr18	48584855	48584855	-	T	intronic	*SMAD4*	rs760402978	.	0.1703	0.037	Hom
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom
chr18	48586184	48586184	A	G	intronic	*SMAD4*	rs948589	Benign	.	0.0552	Het
chr18	48586344	48586344	C	T	intronic	*SMAD4*	rs948588	.	.	0.0576	Het

Table C1.3: A list of variants called in FRASC25

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	Hom
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	benign|Likely benign	0.1416	0.1838	Hom
chr5	149755340	149755340	G	T	exonic	*TCOF1*	rs7701163	benign |Likely benign	0.0663	0.0951	Hom
chr5	149755421	149755421	A	G	exonic	*TCOF1*	rs2071239	benign |Likely benign	0.1428	0.1877	Hom
chr5	149759096	149759096	T	C	exonic	*TCOF1*	rs7713638	benign |Likely benign	0.1428	0.1886	Hom
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Hom
chr5	149772237	149772237	G	A	intronic	*TCOF1*	rs78716239	.	0.0651	0.092	Hom
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	benign |Likely benign	0.0638	0.0878	Hom
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	Hom
chr6	43488536	43488536	A	-	intronic	*POLR1C*	rs111455262	.	0.0092	0.0358	Hom
chr8	27641609	27641609	G	A	intronic	*ESCO2*	rs1824449	Benign	0.9948	0.9795	Hom
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Benign	0.8091	0.7793	Hom
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Hom
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign	0.3943	0.3567	Het
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Het
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	Het
chr8	72127563	72127563	G	A	intronic	*EYA1*	rs7846086	.	.	0.3158	Het
chr8	72127764	72127764	C	A	intronic	*EYA1*	rs3735935	.	0.3774	0.3171	Het
chr8	72211834	72211834	A	C	intronic	*EYA1*	rs3779747	.	0.4165	0.3979	Hom
chr8	72246440	72246440	G	A	intronic	*EYA1*	rs113812295	.	0.0041	0.0161	Het
chr8	72267083	72267083	G	C	exonic	*EYA1*	rs1445404	Benign	0.0134	0.053	het
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom

Table C1.4: A list of 33 variants called in FRASC 26

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	Hom
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	benign|Likely benign	0.1416	0.1838	Hom
chr5	149755340	149755340	G	T	exonic	*TCOF1*	rs7701163	benign |Likely benign	0.0663	0.0951	Hom
chr5	149755421	149755421	A	G	exonic	*TCOF1*	rs2071239	benign |Likely benign	0.1428	0.1877	Hom
chr5	149759096	149759096	T	C	exonic	*TCOF1*	rs7713638	benign |Likely benign	0.1428	0.1886	Hom
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Hom
chr5	149772237	149772237	G	A	intronic	*TCOF1*	rs78716239	.	0.0651	0.092	Hom
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	benign |Likely benign	0.0638	0.0878	Hom
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	Hom
chr6	43488536	43488536	A	-	intronic	*POLR1C*	rs111455262	.	0.0092	0.0358	Hom
chr8	27641609	27641609	G	A	intronic	*ESCO2*	rs1824449	Benign	0.9948	0.9795	Hom
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Benign	0.8091	0.7793	Hom
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Hom
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign	0.3943	0.3567	Het
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Het
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	Het
chr8	72127563	72127563	G	A	intronic	*EYA1*	rs7846086	.	.	0.3158	Het
chr8	72127764	72127764	C	A	intronic	*EYA1*	rs3735935	.	0.3774	0.3171	Het
chr8	72211834	72211834	A	C	intronic	*EYA1*	rs3779747	.	0.4165	0.3979	Hom
chr8	72246440	72246440	G	A	intronic	*EYA1*	rs113812295	.	0.0041	0.0161	Het
chr8	72267083	72267083	G	C	exonic	*EYA1*	rs1445404	Benign	0.0134	0.053	het
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom

Table C1.5: A list of 6 variants called in FRASC27

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr6	43488020	43488020	C	G	exonic	*POLR1C*	rs140188270	.	6.53E-05	0.0003	Het
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign	0.3943	0.3567	Het
chr8	72267083	72267083	G	C	exonic	*EYA1*	rs1445404	Benign	0.0134	0.053	Het
chr13	28239970	28239970	G	A	exonic	*POLR1D*	rs14105	.	0.4392	0.3872	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Het

Table C1.6: A list of the 6 variants called in FRASC28

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr1	149897906	149897906	G	A	exonic	*SF3C1*	rs113949235	Likely benign	0.0033	0.0033	Het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149776232	149776232	C	T	exonic	*TCOF1*	rs15251	other|Benign	0.2598	0.219	Het
chr13	28197246	28197246	A	-	exonic	*POLR1D*					Hom
chr13	28240021	28240021	C	T	exonic	*POLR1D*	rs41291680	.	0.0205	0.0193	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Het

Table C1.7: A list of 30 variants called in FRASC31

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	other|Likely benign	0.1416	0.1838	Hom
chr5	149755340	149755340	G	T	exonic	*TCOF1*	rs7701163	other|Likely benign	0.0663	0.0951	Hom
chr5	149755421	149755421	A	G	exonic	*TCOF1*	rs2071239	other|Likely benign	0.1428	0.1877	Hom
chr5	149759096	149759096	T	C	exonic	*TCOF1*	rs7713638	other|Likely benign	0.1428	0.1886	Hom
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Hom
chr5	149772237	149772237	G	A	intronic	*TCOF1*	rs78716239	.	0.0651	0.092	Hom
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	other|Likely benign	0.0638	0.0878	Hom
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	Hom
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	Hom
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Het
chr8	61743055	61743055	G	A	exonic	*CHD7*	rs190548814	Benign	0.0006	0.0021	Het
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign|	0.3943	0.3567	Het
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Het
chr8	72211295	72211295	T	C	exonic	*EYA1*	rs1445398	Benign	0.1292	0.0945	Het
chr8	72211834	72211834	A	C	intronic	*EYA1*	rs3779747	.	0.4165	0.3979	Het
chr15	67008708	67008708	G	A	intronic	*SMAD6*	rs2278603	.	0.1198	0.1215	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42932244	42932244	C	T	intronic	*EFTUD2*	rs2289672	.	0.1151	0.1425	Het
chr17	42934415	42934415	T	C	intronic	*EFTUD2*	rs58089352	.	0.0749	0.0819	Het
chr17	42937964	42937966	TTT	-	intronic	*EFTUD2*	rs753601433	.	0.0274	0.0003	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42949992	42949992	C	T	intronic	*EFTUD2*	rs8080227	.	.	0.042	Het
chr17	42959146	42959146	G	A	intronic	*EFTUD2*	rs115665005	.	.	0.0022	Het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	Other	0.5609	0.6292	Hom
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom

Table C1.8: A list of 21 variants called in FRASC33

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr8	61654298	61654298	T	A	exonic	*CHD7*	rs41272435	Benign|other|Likely benign	0.0123	0.0123	het
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	het
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	het
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	het
chr13	28239940	28239940	G	C	exonic	*POLR1D*	rs11029	.	0.3618	0.2973	het
chr13	28239970	28239970	G	A	exonic	*POLR1D*	rs14105	.	0.4392	0.3872	het
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	het
chr15	67073549	67073549	C	T	exonic	*SMAD6*	rs12591946	Benign	0.0154	0.0102	het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929759	42929759	C	G	intronic	*EFTUD2*	rs117345300	Benign	0.0257	0.0221	het
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	het
chr17	42930859	42930859	G	A	intronic	*EFTUD2*	rs78620114	.	0.1099	0.0949	het
chr17	42932244	42932244	C	T	intronic	*EFTUD2*	rs2289672	.	0.1151	0.1425	het
chr17	42937964	42937966	TTT	-	intronic	*EFTUD2*	rs753601433	.	0.0274	0.0003	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	het
chr17	42953409	42953409	A	G	exonic	*EFTUD2*	rs2289674	Other	0.1698	0.2046	het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	Other	0.5609	0.6292	het
chr18	46474746	46474746	C	T	intronic	*SMAD7*	rs3736242	.	0.2217	0.2112	het
chr18	48584855	48584855	-	T	intronic	*SMAD4*	rs760402978	.	0.1703	0.037	Hom

Table C1.9: A list of 28 variants called in FRASC34

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	het
chr5	135510054	135510054	G	A	intronic	*SMAD5*	rs77131299	.	0.0032	0.014	het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149759135	149759135	T	C	exonic	*TCOF1*	rs751264680	.	2.03E-05	6.46E-05	het
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	het
chr6	43488354	43488354	C	T	intronic	*POLR1C*	rs113383614	Likely benign	0.0093	0.0325	het
chr8	61654223	61654223	C	T	exonic	*CHD7*	.	.	.	.	het
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732518	61732518	C	G	intronic	*CHD7*	rs79276682	Benign	0.0024	0.0106	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Hom
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	het
chr8	72211834	72211834	A	C	intronic	*EYA1*	rs3779747	.	0.4165	0.3979	het
chr8	72233922	72233922	C	T	intronic	*EYA1*	rs59927438	.	0.0043	0.0163	het
chr8	72267083	72267083	G	C	exonic	*EYA1*	rs1445404	Benign	0.0134	0.053	het
chr13	28239970	28239970	G	A	exonic	*POLR1D*	rs14105	.	0.4392	0.3872	het
chr15	67008708	67008708	G	A	intronic	*SMAD6*	rs2278603	.	0.1198	0.1215	het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42937963	42937963	-	T	intronic	*EFTUD2*	rs34404174	.	0.1085	0.1203	Hom
chr17	42937964	42937966	TTT	-	intronic	*EFTUD2*	rs753601433	.	0.0274	0.0003	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42949808	42949808	G	A	intronic	*EFTUD2*	rs11654183	Other	0.2687	0.278	het
chr17	42949992	42949992	C	T	intronic	*EFTUD2*	rs8080227	.	.	0.042	het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	Other	0.5609	0.6292	Hom
chr18	46474746	46474746	C	T	intronic	*SMAD7*	rs3736242	.	0.2217	0.2112	het
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom

Table C1.10: A list of 20 variants called in FRASC35

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149749197	149749197	C	T	intronic	*TCOF1*	rs77741284	.	0.0543	0.0515	het
chr5	149776232	149776232	C	T	exonic	*TCOF1*	rs15251	other|Benign	0.2598	0.219	het
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	other	0.8091	0.7793	Hom
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	other	0.8175	0.8044	Hom
chr13	28239940	28239940	G	C	exonic	*POLR1D*	rs11029	.	0.3618	0.2973	Hom
chr13	28239970	28239970	G	A	exonic	*POLR1D*	rs14105	.	0.4392	0.3872	Hom
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	Hom
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Hom
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	het
chr17	42937963	42937963	-	T	intronic	*EFTUD2*	rs34404174	.	0.1085	0.1203	Hom
chr17	42937964	42937966	TTT	-	intronic	*EFTUD2*	rs753601433	.	0.0274	0.0003	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	other	0.5609	0.6292	het
chr18	46474746	46474746	C	T	intronic	*SMAD7*	rs3736242	.	0.2217	0.2112	het
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom

Table C1.11: A list of 41 variants called in FRASC52

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	Het
chr5	135510054	135510054	G	A	intronic	*SMAD5*	rs77131299	.	0.0032	0.014	Het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149749197	149749197	C	T	intronic	*TCOF1*	rs77741284	.	0.0543	0.0515	Het
chr5	149754991	149754991	C	T	Exonic	*TCOF1*	rs2071238	Likely benign	0.1416	0.1838	Het
chr5	149755340	149755340	G	T	Exonic	*TCOF1*	rs7701163	Likely benign	0.0663	0.0951	Het
chr5	149755421	149755421	A	G	Exonic	*TCOF1*	rs2071239	Likely benign	0.1428	0.1877	Het
chr5	149759096	149759096	T	C	Exonic	*TCOF1*	rs7713638	Likely benign	0.1428	0.1886	Het
chr5	149763305	149763305	A	G	Exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Het
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	Likely benign	0.0638	0.0878	Het
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	Het
chr6	43488536	43488536	A	-	intronic	*POLR1C*	rs111455262	.	0.0092	0.0358	Hom
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	Het
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Het
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Het
chr8	61765273	61765273	C	T	intronic	*CHD7*	rs3763592	Benig	0.0664	0.1429	Het
chr8	61765560	61765560	G	A	Exonic	*CHD7*	rs2068096	Benign	0.0613	0.1243	Het
chr8	72111599	72111599	A	G	Exonic	*EYA1*	rs10103397	Benign	0.3943	0.3567	Het
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Het
chr8	72127563	72127563	G	A	intronic	*EYA1*	rs7846086	.	.	0.3158	Het
chr8	72127764	72127764	C	A	intronic	*EYA1*	rs3735935	.	0.3774	0.3171	Het
chr8	72211834	72211834	A	C	intronic	*EYA1*	rs3779747	.	0.4165	0.3979	Het
chr8	72233935	72233935	G	A	intronic	*EYA1*	rs113993220	.	0.0013	0.0053	Het
chr8	72267083	72267083	G	C	Exonic	*EYA1*	rs1445404	Benign	0.0134	0.053	Het
chr8	72268678	72268678	A	G	intronic	*EYA1*	rs201352094	.	0.0002	0.0009	Het
chr13	28196156	28196156	G	A	intronic	*POLR1D*					Het
chr13	28240021	28240021	C	T	Exonic	*POLR1D*	rs41291680	.	0.0205	0.0193	Het
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	Het
chr16	72042635	72042635	A	G	upstream	*DHODH*	rs34270657	Likely benign	0.0609	0.0587	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42934415	42934415	T	C	intronic	*EFTUD2*	rs58089352	.	0.0749	0.0819	Het
chr17	42937963	42937963	-	T	intronic	*EFTUD2*	rs34404174	.	0.1085	0.1203	Hom

chr17	42937964	42937967	TTTT	-	intronic	*EFTUD2*	.	.	0.0071	.	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42949992	42949992	C	T	intronic	*EFTUD2*	rs8080227	.	.	0.042	het
chr17	42953409	42953409	A	G	Exonic	*EFTUD2*	rs2289674	Other	0.1698	0.2046	Het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	Other	0.5609	0.6292	Hom
chr18	46474746	46474746	C	T	intronic	*SMAD7*	rs3736242	.	0.2217	0.2112	Het
chr18	48584855	48584855	-	T	intronic	*SMAD4*	rs760402978	.	0.1703	0.037	Hom

Table C1.12: A list of 36 variants called in FRASC54

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	Het
chr5	135510054	135510054	G	A	intronic	*SMAD5*	rs77131299	.	0.0032	0.014	Het
chr5	135513085	135513085	-	C	Exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	Exonic	*TCOF1*	rs2071238	other|Likely benign	0.1416	0.1838	Hom
chr5	149755340	149755340	G	T	Exonic	*TCOF1*	rs7701163	other|Likely benign	0.0663	0.0951	Hom
chr5	149755421	149755421	A	G	Exonic	*TCOF1*	rs2071239	other|Likely benign	0.1428	0.1877	Hom
chr5	149759096	149759096	T	C	Exonic	*TCOF1*	rs7713638	other|Likely benign	0.1428	0.1886	Hom
chr5	149763305	149763305	A	G	Exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Hom
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	other|Likely benign	0.0638	0.0878	Hom
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	Hom
chr6	43488536	43488536	A	-	intronic	*POLR1C*	rs111455262	.	0.0092	0.0358	Hom
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	Het
chr8	61693820	61693820	A	-	Exonic	*CHD7*					Hom
chr8	61707572	61707572	T	C	Exonic	*CHD7*	rs79302359	Benign|other|	0.0125	0.0173	Het
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732518	61732518	C	G	intronic	*CHD7*	rs79276682	Benign	0.0024	0.0106	Het
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Hom
chr8	61765273	61765273	C	T	intronic	*CHD7*	rs3763592	Benign|	0.0664	0.1429	Het
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	Hom
chr15	66995675	66995675	A	G	Exonic	*SMAD6*	rs374058045	.	6.65E-05	0.0013	Het
chr16	72042682	72042682	A	C	Exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42930859	42930859	G	A	intronic	*EFTUD2*	rs78620114	.	0.1099	0.0949	Het

chr17	42932244	42932244	C	T	intronic	*EFTUD2*	rs2289672	.	0.1151	0.1425	Het
chr17	42937964	42937966	TTT	-	intronic	*EFTUD2*	rs753601433	.	0.0274	0.0003	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42949992	42949992	C	T	intronic	*EFTUD2*	rs8080227	.	.	0.042	Het
chr17	42953409	42953409	A	G	Exonic	*EFTUD2*	rs2289674	Other	0.1698	0.2046	Het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	Other	0.5609	0.6292	Hom
chr17	42964145	42964145	C	T	intronic	*EFTUD2*	rs115624297	.	0.0074	0.0288	Het
chr18	46447937	46447937	G	A	Exonic	*SMAD7*	rs149492644	.	0.0006	0.0028	Het
chr18	48584856	48584856	T	-	intronic	*SMAD4*	rs540116851	.	0.1576	0.0274	Hom

Table C1.13: A list of 30 variants called in FRASC57

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	other|Likely benign	0.1416	0.1838	het
chr5	149755340	149755340	G	T	exonic	*TCOF1*	rs7701163	other|Likely benign	0.0663	0.0951	het
chr5	149755421	149755421	A	G	exonic	*TCOF1*	rs2071239	other|Likely benign	0.1428	0.1877	het
chr5	149759096	149759096	T	C	exonic	*TCOF1*	rs7713638	other|Likely benign	0.1428	0.1886	het
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	het
chr5	149772237	149772237	G	A	intronic	*TCOF1*	rs78716239	.	0.0651	0.092	het
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	other|Likely benign	0.0638	0.0878	het
chr5	149773041	149773041	C	-	exonic	*TCOF1*					Hom
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	het
chr5	149778599	149778599	G	A	exonic	*TCOF1*	rs116268092	Likely benign	0.002	0.0076	het
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	other	0.8091	0.7793	Hom
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	other	0.8175	0.8044	het
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign|Benign|Benign	0.3943	0.3567	Hom
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Hom
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	het
chr8	72127563	72127563	G	A	intronic	*EYA1*	rs7846086	.	.	0.3158	het
chr8	72127764	72127764	C	A	intronic	*EYA1*	rs3735935	.	0.3774	0.3171	het
chr8	72127947	72127947	A	T	exonic	*EYA1*	rs112593082	Likely benign	2.03E-05	6.46E-05	het
chr8	72229769	72229769	C	A	intronic	*EYA1*	rs113694988	Benign	0.0066	0.0269	het
chr13	28196164	28196164	G	A	intronic	*POLR1D*	rs2232681	.	0.0002	0.0005	het

chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42932244	42932244	C	T	intronic	*EFTUD2*	rs2289672	.	0.1151	0.1425	het
chr17	42937964	42937965	TT	-	intronic	*EFTUD2*	rs757112069	.	0.0733	0.0018	Hom
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42953409	42953409	A	G	exonic	*EFTUD2*	rs2289674	other	0.1698	0.2046	Hom
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	other	0.5609	0.6292	Hom

Table B5.14: A list of 30 variants called in FRASC59

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	other|Likely benign	0.1416	0.1838	Hom
chr5	149755340	149755340	G	T	exonic	*TCOF1*	rs7701163	other|Likely benign	0.0663	0.0951	Het
chr5	149755421	149755421	A	G	exonic	*TCOF1*	rs2071239	other|Likely benign	0.1428	0.1877	Hom
chr5	149755845	149755845	A	G	exonic	*TCOF1*	rs34796297	other|Likely benign	0.0051	0.0198	Het
chr5	149759096	149759096	T	C	exonic	*TCOF1*	rs7713638	other|Likely benign	0.1428	0.1886	Hom
chr5	149759201	149759201	C	T	exonic	*TCOF1*	rs114689020	other|Likely benign	0.0068	0.021	Het
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Hom
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	other|Likely benign	0.0638	0.0878	Het
chr5	149776113	149776113	G	A	exonic	*TCOF1*	rs114169102	Likely benign	0.0028	0.011	Het
chr5	149777778	149777778	C	T	intronic	*TCOF1*	rs11167501	.	.	0.085	Het
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	Het
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	Other	0.8175	0.8044	Het
chr8	61734480	61734480	G	A	exonic	*CHD7*	rs374877439	.	0.0001	0.0007	Het
chr8	61736506	61736507	CA	-	exonic	*CHD7*					Hom
chr8	61761780	61761780	A	G	intronic	*CHD7*	rs16926500	.	.	0.0169	Het
chr8	61777820	61777820	C	G	exonic	*CHD7*	rs376063472	Conflicting interpretations	7.81E-05	6.46E-05	Het
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	Het
chr13	28196204	28196204	G	T	intronic	*POLR1D*	rs2232682	.	.	0.0229	Het
chr13	28240021	28240021	C	T	exonic	*POLR1D*	rs41291680	.	0.0205	0.0193	
chr15	67008737	67008737	C	A	intronic	*SMAD6*	rs2278604	.	0.2274	0.229	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Het
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom
chr16	72058012	72058012	T	G	intronic	*DHODH*	rs113617640	.	0.012	0.0132	Het

chr17	42934415	42934415	T	C	intronic	*EFTUD2*	rs58089352	.	0.0749	0.0819	Het
chr17	42937964	42937965	TT	-	intronic	*EFTUD2*	rs757112069	.	0.0733	0.0018	Hom
chr18	46474746	46474746	C	T	intronic	*SMAD7*	rs3736242	.	0.2217	0.2112	Het
chr18	48584855	48584855	-	T	intronic	*SMAD4*	rs760402978	.	0.1703	0.037	Hom

Table B5.15: A list of 42 variants called in FRASC61

Chr	Start	End	Ref	Alt	Location	Gene	dbSNP	ClinVar	GnomAD Exome_All (MAF)	GnomAD Genome_All (MAF)	Mode of inheritance
chr1	149899578	149899578	G	A	intronic	*SF3B5*	rs113558439	.	2.03E-05	.	Het
chr5	135508381	135508381	A	G	intronic	*SMAD5*	rs4585442	.	0.3095	0.3601	Het
chr5	135513085	135513085	-	C	exonic	*SMAD5*	rs55765823	.	1	1	Hom
chr5	149754991	149754991	C	T	exonic	*TCOF1*	rs2071238	other|Likely benign	0.1416	0.1838	Het
chr5	149763305	149763305	A	G	exonic	*TCOF1*	rs7715100	.	0.1401	0.1839	Het
chr5	149772237	149772237	G	A	intronic	*TCOF1*	rs78716239	.	0.0651	0.092	Het
chr5	149772932	149772932	C	T	intronic	*TCOF1*	rs11743855	other|Likely benign	0.0638	0.0878	Het
chr6	43488536	43488536	A	-	intronic	*POLR1C*	rs111455262	.	0.0092	0.0358	Hom
chr8	61654634	61654634	C	T	exonic	*CHD7*	.	.	.	.	Het
chr8	61655690	61655690	G	A	intronic	*CHD7*	rs7836586	Other	0.8091	0.7793	Het
chr8	61707725	61707725	G	A	intronic	*CHD7*	rs4540437	Other	0.829	0.8435	Hom
chr8	61713126	61713126	-	TGGACT	intronic	*CHD7*	rs5891777	Likely benign	0.8008	0.7775	Hom
chr8	61732521	61732521	A	G	intronic	*CHD7*	rs6471902	other	0.8175	0.8044	Het
chr8	61754339	61754339	C	T	intronic	*CHD7*	rs115999896	.	0.0009	0.0044	het
chr8	61761616	61761616	C	T	exonic	*CHD7*	rs16926499	Benign||Likely benign	0.004	0.0182	Het
chr8	61761780	61761780	A	G	intronic	*CHD7*	rs16926500	.	.	0.0169	Het
chr8	61765273	61765273	C	T	intronic	*CHD7*	rs3763592	Benign|Benign|Benign	0.0664	0.1429	Het
chr8	61765560	61765560	G	A	exonic	*CHD7*	rs2068096	Benign|Benign|Benign	0.0613	0.1243	Het
chr8	61769117	61769117	G	A	exonic	*CHD7*	rs187311127	Benign|Likely benign|	0.0003	0.0013	Het
chr8	72111599	72111599	A	G	exonic	*EYA1*	rs10103397	Benign|Benign|Benign	0.3943	0.3567	Het
chr8	72111678	72111678	T	C	intronic	*EYA1*	rs10090382	.	0.3946	0.3559	Het
chr8	72127562	72127562	C	T	intronic	*EYA1*	rs79867447	.	.	0.1175	Het
chr8	72127563	72127563	G	A	intronic	*EYA1*	rs7846086	.	.	0.3158	Het
chr8	72127764	72127764	C	A	intronic	*EYA1*	rs3735935	.	0.3774	0.3171	Het
chr8	72129009	72129009	G	A	exonic	*EYA1*	rs4738118	Benign|	0.218	0.1508	Het
chr8	72181947	72181947	T	G	intronic	*EYA1*	rs7835813	.	0.0041	0.016	Het
chr8	72233922	72233922	C	T	intronic	*EYA1*	rs59927438	.	0.0043	0.0163	Het
chr8	72267083	72267083	G	C	exonic	*EYA1*	rs1445404	Benign|	0.0134	0.053	Het
chr15	66996035	66996035	G	A	exonic	*SMAD6*	rs867644316	.	6.76E-05	3.43E-05	Het
chr16	72042682	72042682	A	C	exonic	*DHODH*	rs3213422	other|Benign	0.5531	0.5464	Hom
chr16	72050885	72050885	G	C	intronic	*DHODH*	rs11075914	.	0.9999	1	Hom

chr17	42929970	42929970	T	C	intronic	*EFTUD2*	rs2289673	.	0.6332	0.6773	Hom
chr17	42934415	42934415	T	C	intronic	*EFTUD2*	rs58089352	.	0.0749	0.0819	Het
chr17	42937963	42937963	-	TT	intronic	*EFTUD2*	rs34404174	.	0.02	0.0017	Hom
chr17	42937964	42937966	TTT	-	intronic	*EFTUD2*	rs753601433	.	0.0274	0.0003	Hom
chr17	42940288	42940288	C	A	intronic	*EFTUD2*	rs58107893	.	0.0042	0.0064	Het
chr17	42945296	42945296	A	G	intronic	*EFTUD2*	rs2120276	.	0.6327	0.6763	Hom
chr17	42949808	42949808	G	A	intronic	*EFTUD2*	rs11654183	other	0.2687	0.278	Het
chr17	42949992	42949992	C	T	intronic	*EFTUD2*	rs8080227	.	.	0.042	Het
chr17	42961009	42961009	C	T	intronic	*EFTUD2*	rs2289677	other	0.5609	0.6292	Het
chr18	48584469	48584469	A	G	intronic	*SMAD4*	rs375856998	.	0.0003	0.0008	Het
chr18	48584855	48584855	-	TT	intronic	*SMAD4*	rs760402978	.	0.181	0.307	Hom