INVESTIGATING THE GENETIC AETIOLOGY OF RASOPATHIES IN A COHORT OF AFRICAN PATIENTS

Clarice Smal
Student number: 1930283

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

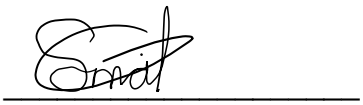
A Dissertation Submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine.

Supervisors:
Dr Nadia Carstens
Ms Maria Mudau

Johannesburg, 2022

Declaration

I, Clarice Smal, declare that this Dissertation is my own, unaided 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 any other University.



Clarice Smal

Date: 25th day of August 2022 in Johannesburg

Presentations arising from this research project

“Whole exome sequencing (WES): secondary testing for patients with RASopathies”, oral presentation at the National Health Laboratory Service PathReD Congress, 21 August 2021.

Abstract

The RASopathies are a group of genetic disorders with clinical overlap caused by germline pathogenic variants in genes of the RAS/MAPK pathway. Limited studies have been done in African populations and therefore the genetic aetiology of RASopathies remain poorly understood in individuals of African ancestry. Furthermore, there is currently no routine molecular diagnostic testing available for patients with a RASopathy in the State sector of South Africa. This study formed part of a larger investigation that used a custom designed targeted NGS gene panel to determine the genetic cause of RASopathies in South Africa and offer better diagnostic services to these patients. The primary study had a diagnostic yield of 56.6% which left a large proportion of the patient cohort without a genetic diagnosis. Additional molecular approaches using MLPA and WES were investigated in the current study to establish a molecular profile for the patients in whom no pathogenic disease-causing variants were identified and to determine whether MLPA and WES could improve the diagnostic yields for South African patients with a RASopathy. MLPA identified three deletions in three patients clinically diagnosed with Neurofibromatosis type 1 achieving a diagnostic yield of 25.0%. This included one multi-exon and two large deletions encompassing the whole *NF1* gene as well as flanking genes. WES identified two pathogenic and one likely pathogenic disease-causing variants in three patients achieving a diagnostic yield of 37.5%. Two of these variants were associated with rare developmental disorders unrelated to the RASopathies and changed the clinical diagnosis of the patients. The other variant was present in a known RASopathy gene not included on the targeted NGS panel. This study significantly improved the diagnostic rate for patients suspected of having a RASopathy and provides insight into improving genetic testing for patients in South Africa. Overall, this study demonstrated the value of reflex testing using both MLPA and WES in a South African setting.

Acknowledgements

I would like to thank my supervisors, Dr Nadia Carstens and Ms Maria Mudau, for giving me the opportunity to work on this project and for their support, guidance and patience over the past two years. Thank you for understanding and keeping me on track as a full-time student while also working full-time. I am grateful for the opportunity to have worked with you and have gained valuable knowledge. I would also like to acknowledge Prof Amanda Krause for her clinical expertise. Thank you to the funders of this project, the National Research Foundation (NRF) for the funding that was provided through the Thuthuka grants awarded to Dr Carstens and Ms Mudau, as well as the National Health Laboratory Service Research Trust (NHLSRT). A special thank you to Ian, my husband, for his unwavering love and support. Thank you for believing in me and motivating me when I needed it most. I would also like to thank my parents for their continued love, support and faith in me. Lastly, I would like to thank my fellow postgraduate students, as well as my friends for their encouragement and words of advice.

Table of Contents

List of Abbreviations

aCGH: Array comparative genomic hybridization

ACMG-AMP: American College of Medical Genetics and Genomics and the Association for Molecular Pathology

BAI: BAM index

BAM: Binary alignment map

CADD: Combined Annotation Dependent Depletion

CAS: Coffalyser Analysis Score

CFC: Cardio-facio-cutaneous syndrome

CM-AVM: Capillary malformation-arteriovenous malformation syndrome

CNV: Copy number variant

CS: Costello syndrome

DANN: Deleterious Annotation of genetic variants using Neural Networks

DD: DNA denaturation

DDD: Deciphering Developmental Disorders

ddNTP: dideoxynucleotide triphosphate

dNTP: deoxynucleotide triphosphate

dsDNA: double stranded deoxyribonucleic acid

ExAC: Exome Aggregation Consortium

FATHMM: Functional Analysis Through Hidden Markov Models

FMRS: Fragment MLPA Reaction Score

FR: Final Ratio

FRSS: Fragment Run Separation Score

FSLP: Final-Normalisation Signal Sloping Probes

G2P: Genotype2Phenotype

GAB1: GRB2 associated binding protein 1

GAP: GTPase activating protein

GDP: Guanosine diphosphate

GEF: Guanosine exchange factor

gnomAD: Genome Aggregation Database

GRB2: Growth factor receptor-bound protein 2

GRCh37: Genome Reference Consortium Human Build 37

GTP: Guanosine triphosphate

HGVS: Human Genome Variation Society

HI: Haploinsufficiency

HPO: Human Phenotype Ontology

HREC: Human Research Ethic Committee

IGV: Integrative Genomics Viewer

ISP: Ion Sphere Particle

LoF: Loss of function

LPO: Left probe oligonucleotide

LS: Legius syndrome

MAF: Minor allele frequency

MLPA: Multiplex ligation dependent probe amplification

mRNA: messenger RNA

NAHR: Non-allelic homologous recombination

NCBI: National Center for Biotechnology Information

NF1: Neurofibromatosis Type 1

NF2: Neurofibromatosis Type 2

NGS: Next generation sequencing

NHLS: National Health Laboratory Service

NS: Noonan syndrome

NS-LAH: Noonan syndrome with loose anagen hair

NSML: Noonan syndrome with multiple lentigines

o/e: Observed/expected

PCR: Polymerase chain reaction

pLI: Probability of LoF intolerance

PolyPhen-2: Polymorphism Phenotyping

PROVEAN: Protein Variation Effect Analyzer

PSLP: Preliminary Signal Sloping Probes

RAS/MAPK: RAS/mitogen-activated protein kinase

REVEL: Rare Exome Variant Ensembl

RPO: Right probe oligonucleotide

RPQ: Reference Probe Quality

RSQ: Reference Sample Quality

RTK: Receptor tyrosine kinase

SHP2: SH2 containing protein tyrosine phosphatase 2

SIFT: Sorting Intolerant From Tolerant

SNP: Single nucleotide polymorphism

SNV: Single nucleotide variant

SOS: Son of sevenless

TBE: Tris-Borate-EDTA

TE: Tris-EDTA

VCF: Variant call format

VEP: Variant Effect Predictor

VUS: Variant of uncertain significance

WES: Whole exome sequencing

WGS: Whole genome sequencing

WITS: University of the Witwatersrand

ZTTK: Zhu-Tokita-Takenouchi-Kim

List of Figures

List of Tables

1. INTRODUCTION

African populations are understudied which means that the genetic aetiology of inherited disorders in individuals of African ancestry remain largely unknown (Sirugo, Williams and Tishkoff, 2019; Choudhury et al., 2020). This as well as the unavailability of genetic testing in Africa consequently affects the diagnosis of patients with genetic disorders since it is predominantly based on their clinical presentation. A clinical diagnosis is not always accurate since some disorders have several overlapping features. It is therefore important to increase the current knowledge related to the genetic basis of inherited disorders in individuals of African ancestry. This will aid in directing genetic testing for these patients. An accurate diagnosis is crucial for the appropriate clinical management and care of patients (Watson, 2015). The RASopathies are one example of a group of disorders with overlapping clinical features (Rauen, 2013) and limited studies done on African populations (Ramanjam et al., 2006; Tekendo-Ngongang and Kruszka, 2020). No genetic testing was available for these patients in the State sector of South Africa at the time of commencement of this study. This study is a sub-study of a lager PhD study by Ms Maria Mudau which aimed to determine the genetic aetiology of patients with a RASopathy in South Africa through a targeted gene panel. However, a large proportion of patients remained without a genetic diagnosis following this approach. Additional molecular approaches have to be investigated to determine the genetic aetiology of these patients in South Africa.

1.1 RASopathies

The RASopathies are a group of genetic developmental conditions (Rauen, 2013) which comprise of several disorders, including Noonan syndrome (NS), Noonan syndrome with multiple lentigines (NSML), Noonan like syndrome with loose anagen hair (NS-LAH), Neurofibromatosis Type 1 (NF1), Neurofibromatosis Type 2 (NF2), Costello syndrome (CS), Cardio-facio-cutaneous syndrome (CFC), Capillary malformation–arteriovenous malformation (CM-AVM) syndrome and Legius syndrome (LS) (Tidyman and Rauen, 2016b). The RASopathies have a collective global incidence of approximately 1 in 1000 individuals (Rauen, 2013; Tajan et al., 2018; Riller and Rieux-Laucat, 2021). It affects all population groups with no known

difference in prevalence (Kruszka et al., 2017; Bergqvist et al., 2020). However, this data is derived from epidemiological studies that are largely focused on populations of European ancestry. This represents one of the largest and most common groups of genetic conditions, but each syndrome individually is rare (Tajan et al., 2018).

The disorders exhibit overlapping phenotypic features due to dysregulation of the common underlying RAS/mitogen-activated protein kinase (RAS/MAPK) pathway which plays a vital role in normal development (Rauen, 2013; Riller and Rieux-Laucat, 2021). Each RASopathy also presents with unique features attributed to the specific genes and manner in which this pathway is dysregulated (Tidyman and Rauen, 2016b). Features that are shared between the majority of RASopathies include cardiopathies, failure to thrive, neurocognitive impairments, craniofacial dysmorphology, cutaneous-, musculoskeletal- and ocular abnormalities and an increased risk for cancer (Rauen, 2013; Gross et al., 2020).

1.1.1 Noonan syndrome

Kobylinski reported on the first patient with NS in 1883 (Kobylinski, 1883), however it was not officially described until 1963 (Noonan and Ehmke, 1963). Noonan syndrome is the most common RASopathy with an estimated global prevalence of approximately 1 in 1000 to 1 in 2000 (Rauen, 2013; Riller and Rieux-Laucat, 2021). Although NS is clinically heterogenous, the major clinical characteristics usually associated with NS are distinctive craniofacial features, short stature, cardiac defects, developmental delay and learning disabilities, skeletal deformities and an increased risk for cancer (Rauen, 2013).

Typical craniofacial features reported in patients with NS include a broad forehead, hypertelorism, ptosis, down slanting palpebral fissures, low set and posteriorly rotated ears and a broad flat nose. A large proportion of affected individuals also have short, webbed necks. The most frequently reported cardiac defect is pulmonary stenosis followed by hypertrophic cardiomyopathy. Chest deformities such as inferior pectus excavatum and superior pectus carinatum are present in most affected individuals (Rauen, 2013; Lee and Yoo, 2019). Patients with NS have an increased risk of developing malignancies particularly childhood cancers and most often juvenile

myelomonocytic leukemia (Tajan et al., 2018; Riller and Rieux-Laucat, 2021). Patients can also present with coagulation defects, wide spaced nipples, low body weight and cryptorchidism in males (Lee and Yoo, 2019).

1.1.2 Noonan syndrome with multiple lentigines

Noonan syndrome with multiple lentigines was first reported by Zeisler and Becker in 1936 and later defined and initially termed LEOPARD syndrome in 1969 (Gorlin, Anderson and Blaw, 1969). The global prevalence of NSML is estimated to be 1 in 100 000 (Tajan et al., 2018). Patients typically present with similar craniofacial features as that seen in NS, as well as short stature and cardiac abnormalities. Cardiac defects most frequently reported are hypertrophic cardiomyopathy, however pulmonary stenosis is also observed. Affected individuals are distinguished from NS by the presence of multiple lentigines which are flat pigmented (black-brown) skin lesions mostly located on the face, neck and trunk. Other distinctive features commonly reported in patients with NSML include hearing loss and electrocardiogram abnormalities. Patients can also present with mild intellectual disability (Rauen, 2013; Tajan et al., 2018).

1.1.3 Noonan-like syndrome with loose anagen hair

Noonan-like syndrome with loose anagen hair was first described by Mazzanti et al. in 2003. The exact prevalence is unknown; however it is regarded as a rare disorder. The clinical features of NS-LAH are very similar to that of NS including dysmorphic craniofacial features and short stature. However, NS-LAH is predominantly distinguished from NS by the presence of sparse, thin, loose and slow-growing hair in the anagen phase. Other distinctive features include mild psychomotor delay with hyperactive behaviour and a higher frequency of cardiac defects, specifically mitral valve and septal defects, compared to NS (Tajan et al., 2018). Patients also present with hypernasal voices and hyperpigmented skin with eczema or ichthyosis (Huckstadt et al., 2021).

1.1.4 Cardio-facio-cutaneous syndrome

Cardio-facio-cutaneous syndrome was first described in 1986 by Reynolds et al. and Baraitser and Patton. The worldwide prevalence of CFC is unknown; however a study estimated the prevalence of CFC in Japan to be approximately 1 in 810 000 (Abe et al., 2012; Ella et al., 2014). Many of the clinical features observed in patients with CFC overlap with those seen in NS and CS. Affected individuals are typically characterized by dysmorphic craniofacial features similar to NS, cardiac defects, short stature, neurocognitive impairment, failure to thrive, ectodermal anomalies, ocular and musculoskeletal conditions (Rauen, 2013; Ella et al., 2014; Tajan et al., 2018).

Macrocephaly, a broad forehead, bitemporal narrowing, increased facial width and depth, a small chin, high arched palate, low set and posteriorly rotated ears, hypertelorism, epicanthic folds, ptosis and a short nose are some of the craniofacial features associated with CFC (Rauen, 2013; Ella et al., 2014). The most prevalent cardiac defect in patients with CFC is pulmonary stenosis, however septal defects and hypertrophic cardiomyopathy are also commonly seen. Almost all the patients have developmental delay and intellectual disability. Furthermore, seizures are also a common occurrence in the spectrum of neurocognitive impairments observed in affected individuals. An important distinguishing feature of CFC is the ectodermal manifestations. Patients typically present with sparse, brittle, and curly hair, as well as sparse eyebrows and eyelashes (Ella et al., 2014). Dystrophic nails and skin conditions such as hyperkeratosis, keratosis pilaris, ichthyosis and the progressive formation of benign nevi are also reported in most patients (Tajan et al., 2018). Unlike some of the other RASopathies, patients with CFC appear to not have an increased risk of cancer (Riller and Rieux-Laucat, 2021).

1.1.5 Costello syndrome

Costello syndrome was first described in 1977 and subsequently named after its founder (Costello, 1977). The worldwide prevalence of this syndrome is estimated to be 1 in 300 000 (Gripp, Morse, et al., 2019). Like most of the RASopathies, patients with CS have dysmorphic craniofacial features, cardiac anomalies, short stature, neurocognitive impairment, failure to thrive, ectodermal, ocular and musculoskeletal

anomalies, as well as an increased risk for cancer (Rauen, 2013; Tajan et al., 2018; Gripp, Morse, et al., 2019).

Craniofacial features associated with CS overlap to a degree with that of other RASopathies, however facial features are typically coarse in CS. Patients usually have macrocephaly, a prominent forehead, epicanthic folds, downslanting palpebral fissures, a short nose, low set posteriorly rotated ears and a large mouth with full lips. Hypertrophic cardiomyopathy is the most common cardiac defect reported in patients, although pulmonary stenosis and arrhythmias are also frequently observed (Rauen, 2013; Gripp, Morse, et al., 2019). Like CFC, most affected individuals have some degree of intellectual disability and developmental delay as part of the neurocognitive manifestations. Ectodermal findings include sparse, curly, and fine hair similar to CFC, however in contrast to CFC, patients with CS have thick and full eyebrows (Riller and Rieux-Laucat, 2021). Other distinctive dermatological features are soft, loose skin with redundant skin on hands and feet, as well as deep plantar and palmar creases (Rauen, 2013; Gripp, Morse, et al., 2019). Papillomas have a high prevalence in CS, but is not seen in other RASopthies. It is mainly present on the nose and can also be found on other areas of the face, ear lobes and perineal region (Rauen, 2013).

Individuals affected with CS have a 15% risk of developing malignancies by the age of 20 years. This is the highest risk for developing cancer among all the RASopathies. The most common type of malignancy is rhabdomyosarcoma in early childhood followed by neuroblastoma, as well as bladder carcinoma in adolescence and early adulthood (Tajan et al., 2018; Gripp, Morse, et al., 2019).

1.1.6 Capillary malformation arteriovenous malformation syndrome

Capillary malformation arteriovenous malformation syndrome was first described as a genetic syndrome by Eerola et al., 2003. This syndrome is clinically characterized by the presence of multiple small capillary malformations (CM) which present as flat cutaneous pink or red lesions. Capillary malformation arteriovenous malformation syndrome is also frequently accompanied by arteriovenous malformations and arteriovenous fistulas located in the skin, bone, muscle, spine and internal organs such as the heart and brain (Wooderchak-Donahue et al., 2018). Arteriovenous

malformations and arteriovenous fistulas can result in life threatening complications such as congestive heart failure, neurological consequences such as seizures, headaches, sensorimotor or other neurological deficits and hemorrhage. Although less commonly observed, some patients can also present with Parkes Weber syndrome which is characterized by hypertrophy of the bones and soft tissue of the limbs (Orme et al., 2013). It is still unknown if affected individuals have an increased risk of developing malignant tumours (Rauen, 2013). Capillary malformation arteriovenous malformation syndrome is estimated to have a prevalence of approximately 1 in 100 000 in Europe (Wooderchak-Donahue et al., 2018).

1.1.7 Neurofibromatosis Type 1

Neurofibromatosis Type 1 was originally described by Von Recklinghausen in 1882 and was the first identified RASopathy disorder. It is the second most common RASopathy with a global prevalence of 1 in 3000 (Tajan et al., 2018). The most frequent clinical features characterizing NF1 are the presence of multiple café au lait maculae, axillary and inguinal freckling, optic gliomas, iris Lisch nodules and multiple neurofibromas. Neurofibromas are mostly cutaneous, although plexiform neurofibromas are found in 50% of affected individuals and are mostly located internally (Rauen, 2013; Legius et al., 2021).

Approximately half of the patients with NF1 have mild learning disabilities. Cardiac anomalies such as vasculopathy, pulmonary stenosis, hypertrophic cardiomyopathy and hypertension are less frequently reported (Rauen, 2013). Some patients also present with distinctive skeletal manifestations such as long bone (especially tibia) or sphenoid dysplasia (Riller and Rieux-Laucat, 2021).

Patients with NF1 have an overall 5 to 15% increased risk of developing malignancies compared to the general population. Malignant peripheral nerve sheath tumours are the most common type of malignancy. Others include malignant optic pathway glioma, juvenile myelomonocytic leukemia, rhabdomyosarcoma and neuroblastomas. Affected individuals also have an increased risk of developing breast cancer, gastrointestinal cancer and lung cancers in adulthood (Landry et al., 2021; Riller and Rieux-Laucat, 2021).

1.1.8 Neurofibromatosis Type 2

Neurofibromatosis Type 2 was first described by Wishart in 1822. The global prevalence of NF2 is estimated to be 1 in 60 000 (Evans, 2009). The major feature present in almost all patients with NF2 is the development of bilateral vestibular schwannomas by the age of 30 years, however unilateral vestibular schwannomas can also occur. Schwannomatosis is a neurofibromatosis disorder characterized by the presence of multiple schwannomas similar to NF2. However, in contrast to NF2, patients with Schwannomatosis do not have any vestibular schwannomas. Neurofibromatosis Type 2 patients also commonly develop schwannomas of the other cranial, peripheral or spinal nerves, intracranial and intraspinal meningiomas and ependymomas (Tamura, 2021). Skin tumours such as plaque-like intracutaneous lesions, subcutaneous nodular tumours and intracutaneous tumours can also occur, as well as the occasional neurofibromas. However, these are few in comparison to NF1 (Evans, 2009). The size and location of the tumours determine the symptoms patients experience. Most patients present with hearing loss, tinnitus and balance dysfunction. Other manifestations include seizures, mononeuropathy affecting the facial nerve, polyneuropathy, paraesthesia, muscle weakness and visual disturbances or vision loss. Tumours are usually benign, however malignant transformation can occur (Evans, 2009; Tamura, 2021).

1.1.9 Legius syndrome

Legius syndrome was first identified in 2007 by Brems et al. and closely resembles the phenotype of NF1 although it is milder compared to NF1. The syndrome shares features with NF1, most predominantly café-au-lait maculae with or without axillary and inguinal freckling and mild neurocognitive impairment. However, patients do not have neurofibromas, central nervous system tumours or iris Lisch nodules commonly associated with NF1. Patients with LS also have macrocephaly, dysmorphic craniofacial features similar to NS, developmental delay, learning disabilities and attention deficit disorder. Some patients also present with short stature and lipomas (Tajan et al., 2018; Legius et al., 2021). The relative risk of developing cancer is still uncertain for patients with LS (Rauen, 2013). The worldwide prevalence of this disorder is estimated to be 1 in 46 000 to 1 in 75 000 (Legius et al., 2021).

1.2 Genetic aetiology of RASopathies

RASopathies are caused by germline pathogenic variants in one of the numerous genes directly involved or regulating the RAS/MAPK pathway. The genetic aetiology of RASopathies will be discussed in the following sections including the underlying pathway, genes and variants involved.

1.2.1 RAS/MAPK pathway

The RAS/MAPK pathway plays a critical role in normal development as it regulates cellular processes such as proliferation, differentiation, growth, apoptosis and senescence through a signaling cascade (Rauen, 2013; Riller and Rieux-Laucat, 2021). The pathway consists of numerous core pathway components and regulators of RAS signaling (Gross et al., 2020) (Figure 1.1). It is a complex pathway with multiple levels of regulation much of which remains unknown due to the large number of interactions between different molecules, as well as interactions with other signaling pathways (Tajan et al., 2018).

A central component of cell signaling in the RAS/MAPK pathway are the RAS proteins. The *RAS* genes are a family of genes including *KRAS, HRAS* and *NRAS* that code for the RAS proteins which are small guanosine triphosphatases (GTPases). RAS proteins switch between an active guanosine triphosphate (GTP) bound and an inactive guanosine diphosphate (GDP) bound state. In the active form, RAS can bind to a variety of effector proteins to transmit downstream signals (Riller and Rieux-Laucat, 2021).

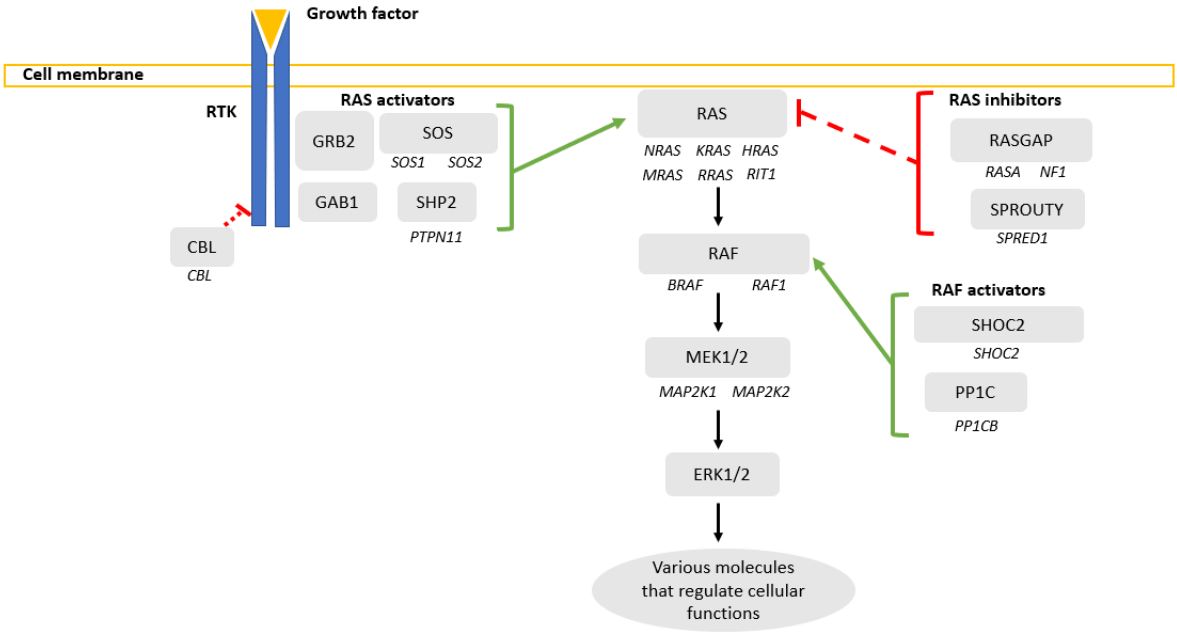


Figure 1.1. Schematic representation of the RAS/MAPK signaling pathway. Growth factors bind to RTKs on the cell membrane causing autophosphorylation. This recruits proteins such as SOS which leads to the activation of RAS proteins. Active RAS initiates a phosphorylation cascade starting with RAF which in turn activates MEK1/2 and subsequently ERK1/2. This ultimately regulates various molecules that control essential cellular functions. Various RAS activators and inhibitors, as well as RAF activators are involved in the regulation of the pathway. Grey blocks represent proteins and genes encoding the proteins are indicated below the blocks. Adapted from Tajan et al., 2018b.

RAS is activated through the binding of growth factors to receptor tyrosine kinases (RTKs), G-protein coupled receptors, cytokine receptors and extracellular matrix receptors. Binding of the growth factors to RTKs leads to RTK autophosphorylation on tyrosine residues. This recruits and facilitates the interaction of the receptors with cytoplasmic adapter proteins such as adapter protein growth factor receptor-bound protein 2 (GRB2) and GRB2 associated binding protein 1 (GAB1). GRB2 is complexed with guanosine exchange factors (GEFs) such as son of sevenless (SOS) and mediates the recruitment of SOS to the plasma membrane. SOS induces the exchange of GDP for GTP on RAS thereby switching RAS to an active state (Rauen, 2013; Riller and Rieux-Laucat, 2021). In addition to SOS, RAS activation is facilitated by SH2 containing protein tyrosine phosphatase 2 (SHP2) which dephosphorylates

several inhibitors of the RAS/MAPK pathway such as SPROUTY and RASGAP proteins (Tajan et al., 2018; Gross et al., 2020).

Inactivation of RAS requires GTPase activating proteins (GAPs) such as neurofibromin, RASA1, RASA2 or SYNGAP1 which promote GTP hydrolysis to switch back to a GDP-bound inactive state. The inactivation of RAS is further promoted through receptor ubiquitination by CBL, an E3 ubiquitin ligase (Gross et al., 2020), and sequestration of RAS activating proteins by SPROUTY proteins (Tajan et al., 2018).

Activated RAS initiates a phosphorylation cascade starting with the activation of RAF proteins. RAF proteins are the first MAPK kinases of the pathway and consist of a family of proteins including ARAF, BRAF and CRAF, also known as RAF1 (Riller and Rieux-Laucat, 2021). The full activation of RAF is facilitated through the dephosphorylation of an inhibitory residue on RAF by PP1C phosphatase, which is in turn activated by the SHOC2 scaffold protein (Tajan et al., 2018; Gross et al., 2020). RAF subsequently phosphorylates and activates the MAPK ERK kinases, MEK1 and/or MEK2. MEK1/MEK2 phosphorylates and activates the extracellular receptor kinases, ERK1 and/or ERK2. ERK1/2 are the ultimate effectors and regulate various downstream cytosolic and nuclear molecules. This includes transcription factors, membrane proteins and protein kinases which control essential cellular functions (Rauen, 2013; Riller and Rieux-Laucat, 2021).

1.2.2 Genetics of RASopathies

Pathogenic variants in one of the RAS/MAPK pathway genes lead to enhanced pathway signaling or sustained activation of the pathway. Most pathogenic variants lead to increased signaling of the pathway through a gain-of-function effect. This includes variants in genes that are involved in the activation of the pathway i.e., positive regulators. Alternatively, other variants in genes that negatively regulate or inhibit the pathway produce increased signaling through a loss-of-function effect (Tajan et al., 2018; Riller and Rieux-Laucat, 2021). Since the RAS/MAPK pathway has a vital role in development, the disruption of normal regulation and function of the pathway leads to disease (Tidyman and Rauen, 2016b; Tajan et al., 2018; Riller and Rieux-Laucat, 2021).

Seeing that all the RASopathies have a common underlying molecular mechanism of disease involving the same pathway, the clinical characteristics of the disorders overlap as discussed in section 1.1. However, there is also phenotypic variability between the different disorders depending on which genes are affected and the specific pathogenic variants that are present in these genes (Rauen, 2013). Variable expressivity has also been observed in individuals with the same pathogenic variant (Gelb et al., 2018; Grant et al., 2018).

To date, 23 genes have been established to be associated with the RASopathies (Table 1.1) (Tidyman and Rauen, 2016a; Grant et al., 2018; Tajan et al., 2018; Gripp, Schill, et al., 2019; Gross et al., 2020; Riller and Rieux-Laucat, 2021). Most of these genes (Table 1.1) have been studied extensively with supporting evidence for a causal role. In recent years, various new genes associated with RASopathies have been discovered due to improvements in technology. Compared to the well-established genes, there is not currently as much evidence for these genes to support a definitive causal role (Tidyman and Rauen, 2016a). The location of the proteins encoded by the RASopathy associated genes in the RAS/MAPK pathway plays a role in the severity of the disease. Genes that are part of the central pathway are associated with a more severe clinical presentation (Tajan et al., 2018). Several genes are associated with only one particular disorder, while other genes overlap between the different RASopathies where the phenotypic manifestations are more dependent on the presence of specific variants (Tidyman and Rauen, 2016b). Noonan syndrome is genetically the most heterogeneous of the disorders (Tajan et al., 2018) as reflected by the diversity of genes involved.

RASopathies are typically inherited in an autosomal dominant manner (Rauen, 2013; Tajan et al., 2018; Riller and Rieux-Laucat, 2021). However, a recessive pattern of inheritance has also been described for Noonan syndrome due to the presence of biallelic variants in the *LZTR1* gene (Johnston et al., 2018).

Table 1.1. Genes with pathogenic variants associated with different RASopathies.

Genes	RASopathy								
	NS	**NSML**	**NS-LAH**	**CS**	**CFC**	**NF1**	**NF2**	**LS**	**CM-AVM**
PTPN11	x	x							
SOS1	x								
RAF1	x	x							
KRAS	x				x				
NRAS	x	x		x	x				
MRAS	x								
SHOC2			x						
CBL	x								
BRAF	x	x			x				
SOS2	x								
RIT1	x								
RRAS	x								
RRAS2	x								
RASA2	x								
LZTR1	x								
PPP1CB			x						
HRAS				x					
MAP2K1	x	x			x				
MAP2K2	x				x				
NF1						x			
NF2							x		
SPRED1								x	
RASA1									x

NS- Noonan syndrome; NSML- Noonan syndrome with multiple lentigines; NS-LAH - Noonan-like syndrome with loose anagen hair; CS- Costello syndrome; CFC- Cardio-facio-cutaneous syndrome; NF1- Neurofibromatosis Type 1; NF2- Neurofibromatosis Type 2; LS- Legius syndrome; CM-AVM- Capillary malformation–arteriovenous malformation; *PTPN11*- protein tyrosine phosphatase, non-receptor type 11; *SOS1*- SOS Ras/Rac guanine nucleotide exchange factor 1; *RAF1*- Raf-1 proto-oncogene, serine/threonine kinase; *KRAS*- KRAS proto-oncogene, GTPase; *NRAS* - NRAS proto-oncogene, GTPase; *MRAS*- muscle RAS oncogene homolog; *SHOC2*- SHOC2 leucine rich repeat scaffold protein; *CBL*- Cbl proto-oncogene; *BRAF*- B-Raf proto-oncogene, serine/threonine kinase; *SOS2*- SOS Ras/Rho guanine nucleotide exchange factor 2; *RIT1*- Ras like without CAAX 1; *RRAS*- RAS related; *RRAS2*- RAS related 2; *RASA2*- RAS p21 protein activator 2; *LZTR1*- leucine zipper like transcription regulator 1; *PPP1CB*- protein phosphatase 1 catalytic subunit beta; *HRAS*- HRas proto-oncogene, GTPase; *MAP2K1*- mitogen-activated protein kinase kinase 1; *MAP2K2*- mitogen-activated protein kinase kinase 2; GTPase; *NF1*- neurofibromin 1; *NF2*- neurofibromin 2; *SPRED1*- sprouty related EVH1 domain containing 1; *RASA1*- RAS p21 protein activator 1

The majority of identified pathogenic variants are *de novo* single nucleotide variants (SNVs). Missense variants account for most of the SNVs and this is observed for the following genes: *PTPN11* (Athota et al., 2020), *RIT1* (Kouz et al., 2016; Yaoita et al., 2016), *HRAS* (Gripp, Morse, et al., 2019; Gripp et al., 2020), *SHOC2* (Cordeddu et al., 2009; J. L. Chen et al., 2014), *CBL* (Martinelli et al., 2015), *PPP1CB* (Gripp et al., 2016; Huckstadt et al., 2021), *RRAS* (Flex et al., 2014; Capri et al., 2019), *SOS2* (Lissewski et al., 2021), *MRAS* (Motta et al., 2020), *RAF1* (Kobayashi et al., 2010), *LZTR1* (Johnston et al., 2018; Pagnamenta et al., 2019), *NRAS* (Altmüller et al., 2017), *RASA2* (P. C. Chen et al., 2014), *SOS1* (Lepri et al., 2011), *BRAF* (Lee et al., 2021), *MAP2K1*, *MAP2K2* (Dentici et al., 2009; Čizmárová et al., 2016) and *KRAS* (Zenker et al., 2007).

In contrast to these genes, missense variants contribute to a lesser extent to the mutational spectrum of *NF1* (Kehrer-Sawatzki, Mautner and Cooper, 2017; Pinna et al., 2019), *NF2* (Halliday et al., 2017; Pasmant et al., 2018), *RASA1* (Revencu et al., 2013; Wooderchak-Donahue et al., 2018) and *SPRED1* (Brems et al., 2012). Pathogenic variants in these genes are mostly frameshift, nonsense and splicing variants. These types of variants are observed less frequently in the other genes. Frameshift variants have been reported for *PPP1CB* (Gripp et al., 2016; Huckstadt et al., 2021), as well as for the recessive form of *LZTR1*, while nonsense and splicing variants have also been detected for the recessive form of *LZTR1* and *CBL* (Martinelli et al., 2015).

Apart from SNVs, small in-frame duplication or deletion variants have also been reported for several genes including *CBL* (Martinelli et al., 2015), *SOS2* (Lissewski et al., 2015), *NF1* (Kehrer-Sawatzki and Cooper, 2021), *PTPN11* (Lee et al., 2005), *BRAF* (Suzuki-Muromoto et al., 2019), *RRAS2* (Capri et al., 2019), *RRAS* (Lorenz et al., 2013; Nagai et al., 2021) and *NRAS* (Altmüller et al., 2017). These variants do not occur commonly and in some instances only a few case reports have been published.

Pathogenic copy number variants (CNVs) involving single or multiple exons, whole genes and even surrounding genes also contribute to the mutational spectrum of RASopathies. Large pathogenic deletions account for 5-11% of variants in the *NF1* gene (Kehrer-Sawatzki and Cooper, 2021), 10-20% of variants in *NF2* (Halliday et al.,

2017; Pasmant et al., 2018), ~10% of *SPRED1* variants (Spencer et al., 2011a) and ~8% of *RASA1* variants (Revencu et al., 2013; Wooderchak-Donahue et al., 2018). Rare CNVs have also been observed in genes such as *PTPN11* (J. L. Chen et al., 2014), *SHOC2* (J. L. Chen et al., 2014), *RAF1* (Luo et al., 2012) and *MAP2K2* (Nowaczyk et al., 2014; Čizmárová et al., 2016). However, CNVs in these genes are not regarded as a major contributor to the mutation spectrum of RASopathies (Lissewski et al., 2015).

The genetic aetiology in 20-30% of patients with a clinical diagnosis of a RASopathy remains unknown (Lissewski et al., 2015). One of several factors could be responsible for this. It is possible that the patients could have causal variants in previously unidentified genes (Tidyman and Rauen, 2016a; Tajan et al., 2018). Alternatively, it could indicate that the current testing strategy is missing variants in known genes due to limitations of the technology. Finally, the patients could have a disorder that is unrelated to the RASopathies but has similar clinical features. We need to expand our current knowledge of the RAS/MAPK pathway and its regulation in order to clarify the genetic cause of RASopathies in the significant percentage of patients who remain undiagnosed, as well as to better understand the genetic aetiology of RASopathies. One way to achieve this is through the advancement and application of new technology, which has previously resulted in the discovery of new RASopathy associated genes (Tidyman and Rauen, 2016a).

Molecular genetic information for RASopathies is based primarily on populations of European ancestry. Limited studies have been carried out on other population groups, especially African populations (Ramanjam et al., 2006; Tekendo-Ngongang and Kruszka, 2020). Additionally, most published literature on African populations has a clinical focus (Odebode et al., 2005; Kruszka et al., 2017). A few studies have explored the genetics of NS to an extent (Essawi et al., 2013; Louati et al., 2014; El Bouchikhi et al., 2015, 2020; Ghedira et al., 2018; Tekendo-Ngongang and Kruszka, 2020). Tekendo-Ngongang et al., 2019 was the first group to report on the application of targeted next generation sequencing (NGS) to elucidate the possible molecular basis of Noonan syndrome in South Africa. This lack of data means the genetic aetiology for African populations remain poorly understood.

1.3 Diagnostic testing for RASopathies

1.3.1 Clinical diagnosis

Patients suspected to have a RASopathy are initially diagnosed based on their clinical presentation (Rauen, 2013). Each disorder has distinctive features which set it apart from the other RASopathies. Clinical diagnostic criteria have been published for most of the RASopathy disorders. There are well-established consensus diagnostic criteria for NF1, LS (Legius et al., 2021) and NF2 (Evans et al., 2018), for which revisions have been published. Diagnostic criteria for NS (Van der Burgt, 2007) and CM-AVM (Orme et al., 2013) have been proposed, but no consensus has been reached. Clinical guidelines exist for both CS (Gripp, Morse, et al., 2019) and CFC (Ella et al., 2014), however no clear set of criteria has been published.

The overlap of clinical features between the RASopathy disorders make it challenging to accurately diagnose a patient. In addition, a patient's phenotype can also vary based on age, as well as variable expression. This means a patient could either not have a clear diagnosis or be clinically misdiagnosed (Bhoj et al., 2017; Grant et al., 2018). Initial clinical misdiagnosis of patients is especially common in Africa since the diagnostic criteria are largely based on European populations and limited information about RASopathy phenotypes in African patients is available (Kruszka et al., 2017; Tekendo-Ngongang and Kruszka, 2020). The shortage of trained medical geneticists in Africa and lack of molecular testing further exacerbate the issue (Kruszka et al., 2017; Tekendo-Ngongang et al., 2019).

Therefore, it is crucial to follow up a clinical diagnosis with molecular genetic testing. This has the potential to identify the causal pathogenic variant and confirm the diagnosis of a specific RASopathy (Rauen, 2013; Grant et al., 2018). An accurate diagnosis is important in managing the health of the patient, improving the patient's outcome and quality of life, as well as offering appropriate genetic counselling (Lee and Yoo, 2019; Wise et al., 2019). Furthermore, an accurate and timely diagnosis reduces costs by minimizing inappropriate and unnecessary clinical interventions and disorganized, protracted diagnostic testing (Wise et al., 2019). Several potential

targeted therapies are also currently under development which will require patient specific variant information (Gripp, Schill, et al., 2019; Gross et al., 2020).

1.3.2 Molecular diagnostic testing

Targeted NGS gene panels are suggested as a first-tier molecular genetic testing strategy for patients with a RASopathy and are offered by several commercial laboratories in various countries (Bhoj et al., 2017; Gelb et al., 2018; Castellanos et al., 2020). NGS is a high-throughput sequencing technique where millions of DNA fragments are simultaneously sequenced. This allows for the sequencing of several genes at the same time using multi-gene panels to detect the causal pathogenic variant (Liu et al., 2019). This is especially useful for RASopathies where single gene testing is impractical due to phenotypic overlap and genetic heterogeneity. A gene panel comprising of all genes known the be associated with RASopathies is a more cost and time effective testing strategy (Bhoj et al., 2017; Castellanos et al., 2020).

Putative pathogenic variants must be classified and reviewed to determine if it is associated with a causal role in a RASopathy. It is therefore important to have a standardized curation of RASopathy associated variants and genes (Gelb et al., 2018; Grant et al., 2018). The ClinGen RASopathy Expert Panel is a disease-specific working group consisting of a team of various medical experts. They review identified genes to find sufficient supporting evidence of a causal role in RASopathies (Grant et al., 2018). The group also refined the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines to create standardized RASopathy specific criteria for variant classification (Gelb et al., 2018).

Additional genetic testing approaches are recommended for patients with negative results obtained using a targeted gene panel. Patients could have disease-causing variants in genes not included on the panel, either because the genes have insufficient evidence supporting a causal role in disease or the genes are not yet associated with RASopathies (Tidyman and Rauen, 2016a; Tajan et al., 2018). Alternatively, patients could have an unrelated disorder caused by pathogenic variants in different genes. It is thus useful to investigate additional genes, apart from known RASopathy genes and

this can be achieved through whole exome sequencing (WES) (Bhoj et al., 2017; Dillon et al., 2018; Liu et al., 2019). Furthermore, CNVs have been reported in a few genes associated with RASopathy disorders, particularly NF1, NF2, LS and CM-VM (Spencer et al., 2011; Brems et al., 2012; Halliday et al., 2017; Wooderchak-Donahue et al., 2018; Revencu et al., 2020; Kehrer-Sawatzki and Cooper, 2021) as discussed in section 1.2.2. These CNVs are usually not screened for using targeted NGS panels and thus alternative techniques such as multiplex ligation dependent probe amplification (MLPA) (Stuppia et al., 2012; Uludağ Alkaya et al., 2021) or array comparative genomic hybridization (aCGH) can be used (Tidyman and Rauen, 2016a; Wooderchak-Donahue et al., 2018) for this purpose. RNA analysis can also be used to identify variants such as deep intronic variants or splice variants that cannot be detected by DNA based testing strategies (Evans et al., 2016; Koster et al., 2021).

In South Africa there is currently no routine molecular diagnostic testing available for patients with a RASopathy in the State sector. The only known available test is carried out by the National Health Laboratory Service (NHLS) and is confined to six variants located in exon 2 of the *HRAS* gene for the diagnosis of Costello syndrome. In the private sector, molecular diagnostic testing for RASopathies is available through outsourcing of private overseas laboratories such as Invitae. These laboratories test by means of targeted NGS gene panels. For example, the Invitae RASopathies comprehensive panel comprises of 28 genes associated with RASopathies or related conditions. This testing is, however, expensive and not covered by medical aids which makes it infeasible for most patients. Furthermore, 83% of the population in South Africa is dependent on the public healthcare system (Ngobeni, Breitenbach and Aye, 2020) and thus testing in overseas laboratories is inaccessible to these patients. RASopathy testing is therefore not available to the majority of patients in South Africa since no molecular diagnostic testing is offered in the State sector and testing through the private sector is too expensive. This highlights the healthcare inequalities that exists in South Africa. As a result, the majority of South African patients with RASopathies are either not diagnosed or only diagnosed clinically if referred to a medical geneticist or appropriate specialist.

1.4 Reflex testing

In this section WES and MLPA will be discussed as reflex testing approaches for patients who have a negative result using a targeted RASopathy gene panel as a first-tier diagnostic test.

1.4.1 WES

WES is an application of NGS technology where all the coding regions of the genome or exons are sequenced. The exome comprises 1-2% of the genome and 85% of known pathogenic variants are located within these regions (Seaby, Pengelly and Ennis, 2016). Since WES is hypothesis-free and thus not limited to a few known disease-causing genes like targeted gene panels, it increases the likelihood of discovering pathogenic variants in genes not currently associated with the specific disease to establish novel gene-disease associations (Dillon et al., 2018). The use of WES is evident in a growing number of studies in recent years which identified newly associated RASopathy genes (Tidyman and Rauen, 2016a). This is also valuable in an African context where the genetic aetiology is still largely unknown and there might be involvement of genes not previously identified. Alternatively, when patients present with clinical features overlapping with those of other conditions, WES can be useful to identify variants in genes associated with a different condition that was not initially clinically suspected (Dillon et al., 2018). Another advantage of WES is that data can be reanalysed periodically when new gene-disease associations are discovered or new information emerges that supports gene-disease associations for which there is currently insufficient evidence (Fung et al., 2020). WES is thus the recommended second line testing method for a negative result on a targeted NGS panel and could potentially lead to an improved diagnostic yield.

Previous studies have demonstrated the effectiveness of WES, especially in patients who had a negative result on other genetic tests including targeted gene panels. The Deciphering Developmental Disorders (DDD) study used WES to achieve a diagnostic yield of 27% for patients with developmental disorders of which the majority had undergone at least one genetic test including clinical microarray, targeted genetic tests or both (Wright et al., 2015). The diagnostic yield of this study increased to 40% after

reanalysis of the original data as more information became available. This was largely attributed to the discovery of new genes and gene-disease associations, but also to improved filtering strategies and updated variant annotation (Wright et al., 2018). Another study using WES in patients presenting with a range of genetic disorders, mostly developmental disorders, had a diagnostic yield of 24%. The patients had either an unknown genetic disorder or a suspected disorder with no genetic diagnosis even after undergoing genetic testing including karyotyping, microarray or targeted gene panels (Zhu et al., 2015).

Even though WES has enormous potential in providing patients with a genetic diagnosis, it also poses a bioinformatic challenge due to the large volumes of data generated. The identification of disease-causing variants among the thousands of detected variants, as well as the interpretation thereof is complex. Additionally, incidental findings and the discovery of variants of uncertain significance (VUS) are also increased compared to targeted gene panel testing (Dillon et al., 2018).

CNVs can also be detected using WES data and several detection tools are available (Zhao et al., 2020). However, it is not used routinely in clinical settings as there are limitations and secondary methods are required to confirm findings (Pounraja et al., 2019). Limitations include high false positive rates, inconsistencies between different algorithms, variable depth of coverage of WES data and lack of available best practice guidelines (Zhao et al., 2020). It is recommended that all CNVs detected by WES in a diagnostic setting should be confirmed with a secondary method (Moreno-Cabrera et al., 2020) which adds additional costs that are currently not feasible in a resource constrained environment and is also not within the budget of my study. Therefore, instead of using WES data to detect CNVs, other methods that are able to detect CNVs such as MLPA and aCGH can be used.

1.4.2 MLPA

MLPA is a high throughput technique that can detect deletions and duplications of several regions in a single reaction. MLPA targets specific regions such as the exons of genes of interest and depending on the design of the MLPA assay this can establish if there is a partial or whole gene deletion/duplication or even a deletion/duplication of

multiple genes in a region (Stuppia et al., 2012). Array CGH is also a high throughput technique capable of detecting CNVs, however aCGH allows for more extensive coverage compared to MLPA by detecting CNVs throughout the genome with high resolution. Small deletions and duplications of approximately ~100 kb, as well as whole chromosome deletions/duplications can be identified by aCGH (Cheung and Bi, 2018).

Both MLPA (Spencer et al., 2011; Pasmant et al., 2018; Wu-Chou et al., 2018; Castellanos et al., 2020; Uludağ Alkaya et al., 2021) and aCGH (Revencu et al., 2013; Summerer et al., 2018; Wooderchak-Donahue et al., 2018) have been used successfully to detect CNVs in patients with a RASopathy including NF1 (Summerer et al., 2018; Wu-Chou et al., 2018; Castellanos et al., 2020; Uludağ Alkaya et al., 2021), NF2 (Pasmant et al., 2018), LS (Spencer et al., 2011a) and CM-AVM (Revencu et al., 2013; Wooderchak-Donahue et al., 2018). MLPA has a lower cost compared to aCGH which makes it more feasible in resource constrained environments such as the State sector of South Africa if shown to be successful for the genetic disorder in question. In addition, multiple commercial MLPA kits are available for specific diseases which removes the necessity of designing and validating an MLPA assay (Stuppia et al., 2012).

Commercial MLPA assays have particularly been used in RASopathies to detect deletions in patients clinically diagnosed with NF1. It is recommended to consider CNV analysis using a technique such as MLPA for patients with NF1 who are panel-negative (Friedman, 2019). The use of MLPA in combination with targeted NGS panels have shown to improve the diagnostic yield for patients with NF1, with an additional 2.5% (Castellanos et al., 2020), 10.5% (Uludağ Alkaya et al., 2021) and 15% (Wu-Chou et al., 2018) of patients in these cohorts receiving a genetic diagnosis following MLPA. Furthermore, since the molecular aetiology is largely unknown in individuals of African ancestry, CNVs could potentially have a higher contribution to disease in this population group compared to patients of European ancestry.

CNVs are not a common occurrence in RASopathy disorders, as discussed in section 1.2.2. However, DNA sequencing analysis can only detect 60–90% of disease-causing variants in NF1 (Friedman, 2019), whereas CNV analysis using alternative methods

can identify large deletions accounting for 5-11% of pathogenic variants in patients with NF1 (Kehrer-Sawatzki and Cooper, 2021). A focused CNV screen for selected disorders is a more viable option in a resource constrained environment and for the current study with limited funding.

1.5 RASopathy studies in South Africa

There are currently only two known studies that investigated the molecular aetiology of RASopathies in South Africa. The first study by Tekendo-Ngongang et al., 2019 investigated both the phenotypic and molecular profiles of South African patients with Noonan syndrome. Tekendo-Ngongang et al., 2019 performed NGS on 16 patients using a targeted panel consisting of 14 genes (*PTPN11, SOS1, RIT1, A2ML1, BRAF, CBL, HRAS, KRAS, MAP2K1, MAP2K2, NRAS, RAF1, SHOC2, and SPRED1*) known to be associated with Noonan syndrome. The study had a 31.2% detection rate and identified pathogenic variants in the *PTPN11, CBL* and *MAP2K1* genes.

The second study is an ongoing PhD study by Ms M. Mudau at the University of the Witwatersrand using a custom designed targeted NGS panel to investigate the genetic aetiology of all RASopathy disorders in South Africa. Ms Mudau's targeted gene panel comprised of 21 RASopathy associated genes (*CBL, LZTR1, KRAS, RRAS, NRAS, PTPN11, RAF1, SOS1, RIT1, SOS2, RASA2, RASA1, BRAF, MAP2K1, MAP2K2, SPRED1, NF1, NF2, HRAS, SHOC2, A2ML1*) and included the exons, as well as 10 base pairs on the exon/intron boundary of these genes. This targeted gene panel contained most of the common genes known to be associated with RASopathies at the time of commencement of the study. The study included 60 patients clinically diagnosed with one of the RASopathy disorders and had a diagnostic yield of 56.6%.

A large proportion of clinically diagnosed patients remained without a molecular diagnosis in both of these studies. This can partly be attributed to the limitations of targeted NGS panels. Targeted NGS panels are designed based on prior knowledge of genes (Liu et al., 2019) associated with RASopathies. It thus includes a restricted selection of known genes. Additionally, the data used to design these gene panels is primarily European based, because African populations are under-represented in the literature as discussed (Essawi et al., 2013; Louati et al., 2014; El Bouchikhi et al.,

2015, 2020; Ghedira et al., 2018; Tekendo-Ngongang and Kruszka, 2020). Patients could also potentially have a differential diagnosis of another rare genetic disorder caused by an unrelated gene not included on the panel. Targeted NGS panels are, however, more cost-effective, produce smaller volumes of data and limit the possibility of secondary findings and VUS compared to WES (Liu et al., 2019).

In view of Ms M. Mudau's study, which did not identify disease-causing variants in 44.4% of the patient cohort due to the limitations of targeted NGS panels, it is appropriate to investigate additional molecular approaches to elucidate the molecular basis of RASopathies in South Africa using the subset of her patient cohort that remained without a molecular confirmation/diagnosis.

1.6 Current study

1.6.1 Study rationale

The RASopathy disorders have several overlapping phenotypic features, therefore a molecular genetic diagnosis is essential to confirm the clinical diagnosis of the specific RASopathy. A correct and timely diagnosis is important for appropriate clinical management and genetic counselling of patients. It determines the prognosis of a patient which in turn affects clinical decision making, intervention and care of the patient. This could improve a patient's quality of life. An accurate diagnosis is also important to identify other at risk family members and aids in reproductive decision making. In South Africa there is currently no molecular diagnostic testing available for patients with a RASopathy in the State sector. Additionally, there is no clear understanding of the genetic aetiology in South African patients due to a lack of published data. This study formed part of a larger investigation aimed at improving our understanding of the genetic aetiology of RASopathies in South Africa in order to offer better diagnostic services to these patients. Secondary molecular techniques, such as WES and MLPA, were used to identify disease-causing variants in RASopathy cases where a targeted gene panel could not detect a pathogenic variant. This will potentially lead to the detection of pathogenic variants in genes other than the known RASopathy genes, the detection of pathogenic CNVs or a different diagnosis. It is therefore hypothesized that utilizing these secondary molecular techniques could lead to an

increase in the diagnostic yield of RASopathies in South Africa. The research question for this study is thus: Can utilizing molecular techniques such as WES and MLPA improve the diagnostic yield of RASopathies in patients where no causal variants were identified using a targeted gene panel?

1.6.2 Aim and Objectives

1.6.2.1 Aim

The aim of this study is to test whether an exome- or MLPA-based mutation screening approach will improve diagnostic yields for South African patients with a RASopathy.

1.6.2.2 Objectives

1. To identify a cohort of patients with a RASopathy with no pathogenic variants identified using a targeted NGS panel.
2. To perform WES on the subset of patients with a RASopathy with no causal variants identified on the targeted NGS panel to identify any putative disease-causing variants.
3. To perform MLPA on the subset of identified patients with NF1 to detect pathogenic CNVs.
4. To validate WES results using Sanger sequencing.

2. METHODOLOGY

2.1 Study samples

Thirteen study samples were primarily selected from a group of patients previously recruited for an ongoing RASopathy study by Ms M. Mudau. Patients were recruited at Genetics Clinics hosted by medical geneticists from the Division of Human Genetics, NHLS and the University of the Witwatersrand (Wits). In addition, patients were also recruited by Dr Engela Honey at Steve Biko Academic Hospital in collaboration with the University of Pretoria. Clinical assessment of the patients was performed by a medical geneticist using a clinical tick sheet (Appendix 2) consisting of features commonly associated with RASopathy disorders. Informed consent was obtained from the patients or legal guardians (as appropriate) following a discussion during which the study was explained. In addition, nine study samples were selected from patients with a clinical diagnosis of a RASopathy who were referred by medical geneticists from the Division of Human Genetics, NHLS and the University of the Witwatersrand for diagnostic testing using a targeted gene panel. Only patients who had a clinical diagnosis or strong suspicion of one of the RASopathy disorders and in which no disease-causing variants were identified using a targeted gene panel containing known RASopathy genes were included in this study. From this cohort of 22 patients, eight patients with a diagnosis of a RASopathy, excluding NF1, were selected for WES. Fourteen patients with a clinical diagnosis of NF1 were selected for MLPA. Patients were not excluded based on age, sex or ethnicity. Patients who had a previously identified pathogenic variant were excluded from this study.

2.2 Ethics clearance

This study has been approved by the University of the Witwatersrand, Human Research Ethics Committee (HREC) (Clearance certificate no: M200617, Appendix 1.1). Patient samples used in this study are a subset of patients recruited for a previous RASopathy study by Ms M. Mudau with required Wits HREC approval (Certificate no. M170163, Appendix 1.2). The before mentioned study was a sub-study of a primary study by Dr N. Carstens with Wits HREC approval (Certificate no. M160830, Appendix

1.3) and ethics approval by the University of Pretoria (Reference no. 80/2018, Appendix 1.4).

2.3 Quantification and quality assessment of genomic DNA

All samples were previously extracted using either the salting out method (Miller, Dykes and Polesky, 1988) or Qiagen Flexigene® DNA kit (Qiagen, Hilden, Germany) and stored at 4°C in Tris-EDTA (TE) buffer solution. DNA concentrations and quality were determined for all samples using the Implen NanoPhotometer® NP80 (Implen, Munich, Germany). The concentration was measured in ng/µl and the purity of the DNA was assessed based on the 260/280 and 260/230 absorbance ratios. A 260/280 absorbance ratio of ~1.8 and 260/230 absorbance ratio of ~2.0 were deemed acceptable with lower values indicative of the presence of contaminants (Green and Sambrook, 2018).

In addition, agarose gel electrophoresis using a 0.8% agarose gel in Tris-borate-EDTA (TBE) buffer was used to assess the molecular weight of the DNA. DNA migrated through the agarose gel proportional to size when an electric current was applied. Ethidium bromide, a fluorescent intercalating dye, was used to visualize DNA in the presence of ultraviolet light using the Omega Fluor™ Gel Documentation System (Vacutec, Johannesburg, South Africa). A 1 kb molecular weight marker was used to determine the size of the DNA. Only high molecular weight DNA was deemed acceptable and was represented by an intact band above the biggest fragment of the marker (Green and Sambrook, 2019).

A Qubit™ 4.0 fluorometer (ThermoFisher Scientific, US) was used to quantify double stranded DNA (dsDNA) for use in WES. The Qubit™ DNA quantification assays use dyes that bind selectively to DNA and only emit fluorescence when bound. Therefore, DNA concentrations are more accurate compared to conventional spectrophotometer quantifications which overestimate DNA concentrations due to contaminants absorbing at the same wavelength. Two Qubit™ DNA quantification assay kits were used: the Invitrogen™ Qubit™ dsDNA Broad-Range Assay Kit and the Invitrogen™ Qubit™ dsDNA High-Sensitivity Assay Kit. These kits detect double stranded DNA

with a broad range of DNA concentrations (~0.1-1000 ng/µl) and a low range of DNA concentrations (~0.01-100 ng/µl), respectively.

2.4 MLPA

MLPA was used to screen for CNVs in patients clinically diagnosed with NF1, but in whom no pathogenic variants were detected using a targeted gene panel. MLPA is a multiplex polymerase chain reaction (PCR) method used to detect CNVs. It makes use of multiple different probes targeting specific regions of interest in the same reaction (Schouten et al., 2002). The SALSA MLPA P081 NF1, P082 NF1 and P122 NF1 probemixes and the SALSA MLPA EK1 Reagent kit (FAM labelled) (MRC Holland, Amsterdam, The Netherlands) were used as per the manufacturer's instructions to detect single and multiple exon deletions or duplications in the *NF1* gene, as well as larger deletions encompassing the whole *NF1* gene and neighbouring genes.

The MLPA EK1 Reagent kit contains all the reagents that are required for the MLPA procedure including: MLPA buffer, Ligase-65, Ligase buffer A, Ligase buffer B, PCR primer mix and polymerase. The SALSA P081 and P082 probemixes combined contain at least one probe for each of the 58 exons of *NF1*. In addition, both probemixes also contain one upstream and downstream probe, one probe for intron 1 of the *NF1* gene and two probes for the *OMG* gene located in intron 36 of *NF1*. The SALSA P081 probemix contains 46 probes of which 35 are specific to *NF1* and 11 are reference probes located on chromosomal regions other than 17q11.2 (Appendix 3.1). The SALSA P082 probemix contains 44 probes including 35 *NF1* specific probes and nine reference probes (Appendix 3.2) (MRC Holland, 2021b). More extensive deletions of the region surrounding *NF1* is detected with the SALSA P122 probemix which contains 35 probes. This includes ten reference probes and 25 probes specific to *NF1* and genes located near *NF1* (Appendix 3.3) (MRC Holland, 2020).

Each probemix (P081, P082 and P122) was performed as a separate MLPA reaction. The MLPA procedure included: denaturation of DNA, hybridization of probes, ligation of probes, probe amplification and capillary electrophoresis (Figure 2.1). Each MLPA reaction required the use of three negative controls or reference samples known to

have no CNVs in the target regions, a positive control with a known deletion of *NF1* and a no template control.

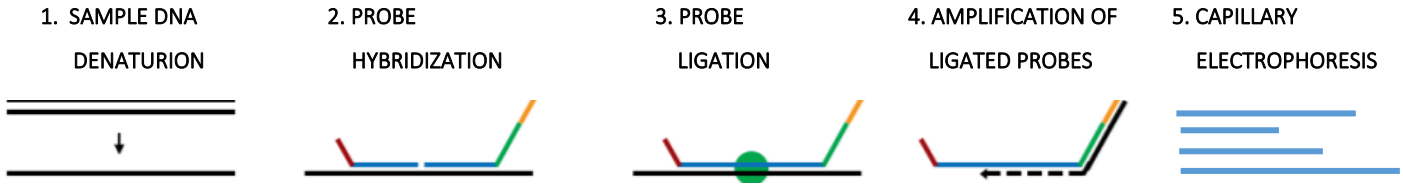


Figure 2.1. Diagram showing the MLPA procedure. DNA was denatured into single strands, followed by the hybridization of each probe to its target region. Probes were subsequently ligated and amplified using PCR. Lastly, amplified fragments were separated by capillary electrophoresis. Adapted from MRC Holland, 2019.

2.4.1 DNA denaturation

Each patient and control DNA sample were diluted in TE buffer to the required concentration of 40 ng/µl prior to starting the MLPA reaction. Double stranded DNA was denatured into single strands at 98°C using a thermocycler to allow for the hybridization of probes in the subsequent step.

2.4.2 Hybridization reaction

An MLPA probe set consists of two separate probe oligonucleotides, a left probe oligonucleotide (LPO) and a right probe oligonucleotide (RPO), which have sequences complementary to sites immediately adjacent to each other at the target region (Figure 2.2). During the hybridization reaction, multiple probes bind to their complementary sequences on the DNA, targeting different regions. A target must be present, i.e., not deleted for a probe to hybridize (Eid, 2017). Reference probe target regions are expected to be present in all samples. The hybridization reaction mix contained the appropriate probemix (P081, P082 or P122) and MLPA hybridization buffer. Hybridization occurred at 60°C for 18 hours.

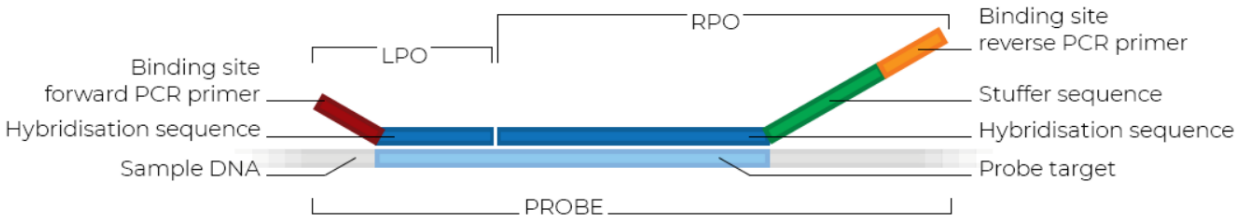


Figure 2.2. Illustration of an MLPA probe set and hybdridization to target DNA. The probe set consists of a LPO and RPO both of which hybridize to sequences immediately adjacent to the target site. Universal primer sequences flank each probe's target sequence. The RPO has an additional flanking stuffer sequence of defined length that differs for each probe. This is used to generate amplified fragments of different lengths in order to separate and distinguish probes from each other during capillary electrophoresis (MRC Holland, 2019).

2.4.3 Ligation reaction

The ligation reaction joined the two probe oligonucleotides that were hybridized to immediately adjacent sequences at the target site. Probe oligonucleotides could only be ligated if both were hybridized to their targets. Ligation occurred at 54°C using the Ligase-65 enzyme together with Ligase buffer A and Ligase buffer B as supplied in the MLPA EK1 Reagent kit.

2.4.4 Probe amplification

The ligated probes were amplified using the DNA polymerase and PCR primer mix supplied in the MLPA EK1 Reagent kit and standard PCR conditions (Table 2.1). A single pair of universal primers was used to amplify all the probes since all probes had the same primer binding sequence. One of the primers was fluorescently labelled with FAM which allowed visualization of the amplified products during fragment analysis. The resulting amplification products ranged in length due to probe design through the presence of a stuffer sequence flanking each probe which was different in length for each target site (Figure 2.2) (Eid, 2017).

Table 2.1. MLPA probe amplification thermocycler conditions.

Step	Temperature	Duration	Cycles
Denaturation	95°C	30 seconds	35
Annealing	60°C	30 seconds	
Elongation	72°C	60 seconds	
Final elongation	72°C	20 minutes	1
Hold	15°C	∞	1

2.4.5 Capillary electrophoresis

The amplified fragments were suspended in Hi-Di Formamide (ThermoFisher Scientific, US) and a GeneScan 500 LIZ size standard (ThermoFisher Scientific, US) before subjecting it to capillary electrophoresis using the Applied Biosystems 3500xL Genetic Analyzer (ThermoFisher Scientific, US). Hi-Di Formamide denatured the DNA fragments and kept it in a denatured state. The GeneScan 500 LIZ size standard contains DNA fragments of known sizes labelled with a fluorescent dye called LIZ and provides reference points for determining the length of the sample DNA fragments. Each target fragment in the MLPA reaction had a unique known size (ranging from 129 to 483 nucleotides) which allowed for the separation of the fragments based on their size through capillary electrophoresis. The sample was injected into a long, thin capillary containing a matrix polymer by applying an electric current. The electric current subsequently resulted in the migration of the fragments through the polymer in the capillary from the negative to the positive electrode where small fragments migrated faster than large fragments. Shortly before the fragment reached the positive electrode, the fragment moved past a laser beam which caused the dyes attached to the fragments to fluoresce. The fluorescence was detected and recorded by a CCD camera.

2.4.6 Data analysis

2.4.6.1 Quality control metrics

The data obtained by capillary electrophoresis was analysed using Coffalyser.Net™ software (MRC Holland, The Netherlands). Quality control metrics generated by the

Coffalyser.Net software assisted in the evaluation of results. The following quality indicators were assessed: DNA concentration, DNA denaturation (DD), Fragment Run Separation Score (FRSS), Fragment MLPA Reaction Score (FMRS) and the Coffalyser Analysis Score (CAS) which included the Preliminary Signal Sloping Probes (PSLP), Final-normalisation Signal Sloping Probes (FSLP), Reference Sample Quality (RSQ) and Reference Probe Quality (RPQ). FRSS measured the quality of capillary electrophoresis, and FMRS indicated the overall quality of the MLPA reaction for each sample. PSLP indicated the difference in signal sloping between each sample and the reference samples while FSLP showed if the correction for signal sloping could be applied successfully. RSQ and RPQ indicated whether the reference samples and reference probes provided reproducible results. The PSLP, FSLP, RSQ and RPQ scores were reflected in an overall score termed CAS. Good quality metrics were displayed as green bars and ticks, whereas metrics of lower or unacceptable quality were represented by yellow bars and red crosses.

2.4.6.2 CNV analysis

The analysis performed by Coffalyser.Net software consisted of two steps: fragment analysis and comparative analysis. Fragment analysis included independent sample processes of raw data such as the recognition and signal determination of peaks, size determination of peaks and linking peaks to the specific probes. Comparative analysis used the signal strength or fluorescence of probes for a set of three negative controls as a reference to which each patient sample was compared for data normalization and to determine final probe ratios. Final ratios (FR) are a measure of the amount of target sequence present in the patient DNA compared to the reference DNA and in essence provides information on copy numbers for each specific target probe. A normal result with two copies of the target region has a FR of ~1.0 (0.80<FR<1.20). This means the patient sample target region had the same number of copies as the reference sample. A FR of ~0.5 (0.40<FR<0.65) indicates that there was only one copy of the target region which is indicative of a heterozygous deletion. A homozygous deletion with no copies present is indicated by a FR of 0 and a heterozygous duplication with 3 copies of a specific target region is indicated by a ratio of ~1.5 (1.30<FR<1.65) (MRC Holland,

2021a). The data for each patient was visualized in a ratio chart using the Coffalyser.Net software.

2.5 WES

WES was used to screen for SNVs in patients clinically diagnosed with a RASopathy, but in whom no pathogenic variants were detected using a targeted gene panel. WES was conducted using the Ion Torrent system (ThermoFisher Scientific, US). The process included library and template preparation followed by automated sequencing (Figure 2.3).

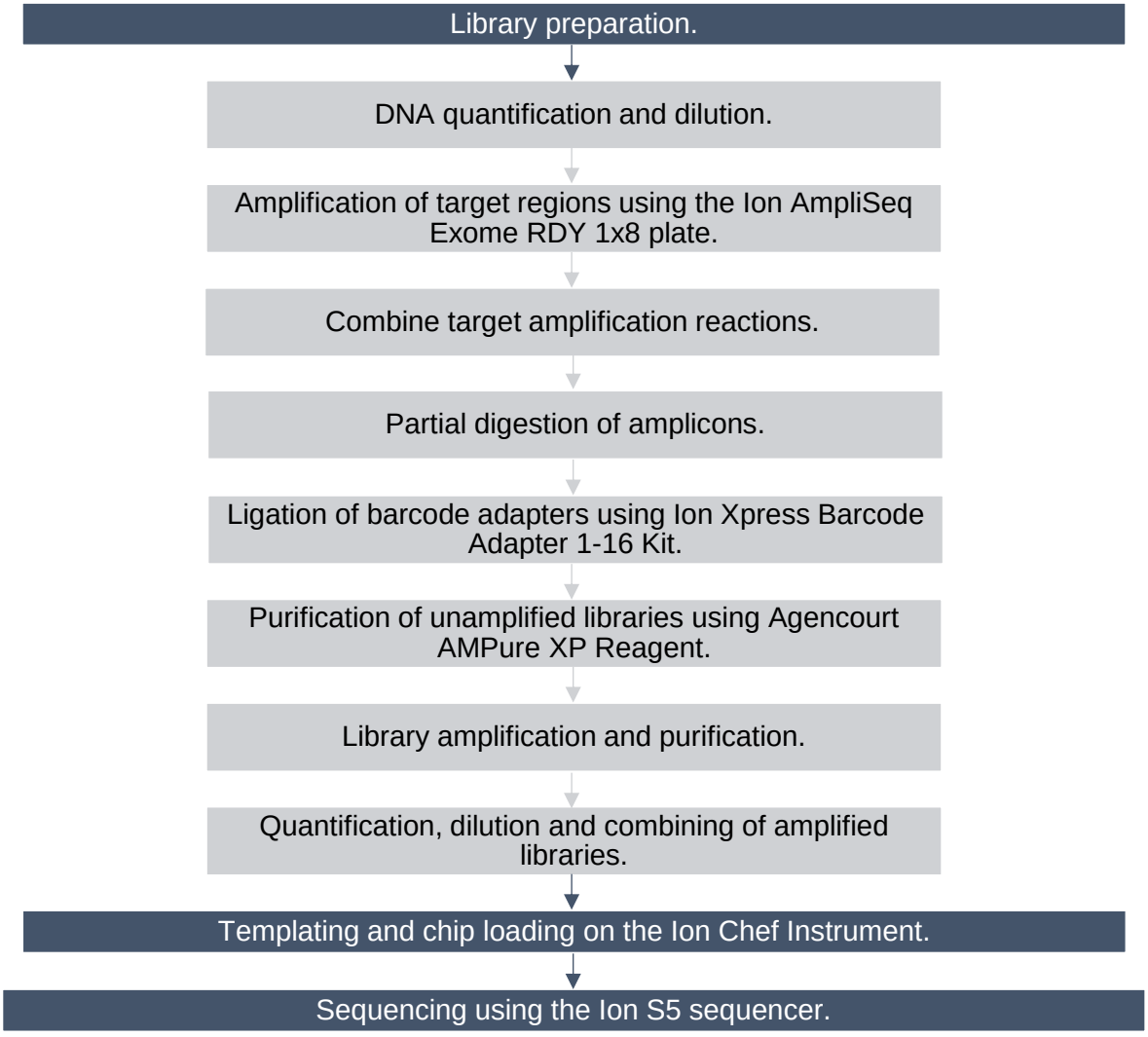


Figure 2.3. Diagram showing the process of WES using the Ion Torrent system. The main procedural steps included library and template preparation and sequencing.

2.5.1 Library preparation

The Ion AmpliSeq™ Exome RDY kit (ThermoFisher Scientific, US) was used for the preparation of exome libraries. The kit included reagents required for library preparation, as well as an Ion AmpliSeq™ Exome RDY 1x8 plate. Eight exome libraries can be prepared with this kit.

Genomic DNA samples for patients were diluted in nuclease free water to a concentration of 10 ng/µl prior to library preparation. A total of 100 ng of genomic DNA was used per sample in conjunction with 5X Ion AmpliSeq™ HiFi Mix for the amplification of target regions in an Ion AmpliSeq™ Exome RDY 1x8 plate containing 12 primer pools for each sample. Following target amplification, the reactions for each sample were combined and the amplicons partially digested using FuPa Reagent to prepare the ends of the amplicons for adapter ligation.

Barcode adapters were subsequently ligated to the amplicons using the Ion Xpress Barcode Adapter 1-16 Kit (ThermoFisher Scientific, US). The kit included both an Ion P1 Adapter and an Ion Xpress Barcode X adapter which was combined with Switch Solution and DNA Ligase. The Ion P1 Adapter attached to Ion Sphere Particles (ISPs) during template preparation and the Ion Xpress Barcode X adapter contained both a universal priming site for sequencing and a unique short sequence (barcode) to label each sample. The barcodes were used to identify sequence data belonging to a specific sample since libraries were pooled prior to templating. The unamplified libraries were purified after barcode ligation using Agencourt AMPure XP Reagent in a 1X bead-to-sample-volume ratio and 70% ethanol.

Library amplification was subsequently performed using 1X Library Amp Mix and 25X Library Amp Primers. Amplification of libraries is required to obtain enough material for quantification. Following amplification, the amplified libraries were purified in two rounds with Agencourt AMPure XP Reagent using a 0.5X bead-to-sample-volume ratio and 1.2X bead-to-original-sample-volume ratio. After the first round of purification, high molecular weight DNA was bound to the beads and discarded, while the amplicons and primers were retained in solution. A higher concentration of beads in the second round of purification resulted in the capture of smaller fragments i.e., the

amplicons, which were retained while the solution containing primers was discarded. The purified amplicons were eluted from the beads with TE.

The amplified libraries were quantified using the Invitrogen Qubit dsDNA High-Sensitivity Assay Kit and Qubit 4.0 fluorometer. The Agilent 4150 TapeStation System (Diagnostech, Johannesburg, South Africa) and Agilent High Sensitivity D5000 ScreenTape assay (Diagnostech, Johannesburg, South Africa) was also used for quantification and size determination of libraries.

2.5.2 Template preparation

Prior to template preparation, a planned run was created using the Ion Torrent Suite software (ThermoFisher Scientific, US). A planned run provides instructions to the Ion Chef instrument (ThermoFisher Scientific, US) for templating, as well as to the Ion S5 sequencer (ThermoFisher Scientific, US) for sequencing and post-sequencing data processing. Information such as the kit type, sample type, barcode set, number of flows, plugins such as coverage analysis and variant calling, as well as the assignment of patient identifiers to each barcode were included in the planned run.

Amplified libraries were normalized to a concentration of 22 ng/mL (~100 pM) based on the concentrations obtained from the Qubit 4.0 fluorometer. According to the manufacturer's instructions, two libraries can be pooled on one Ion 540™ Chip (ThermoFisher Scientific, US) to yield an average coverage depth of 100x. Therefore, two libraries were subsequently pooled by combining 15 µl of each library at equimolar concentrations of ~100 pM. A total of four pooled libraries were generated by combining eight individual barcoded libraries.

Template preparation and chip loading were automated processes performed by the Ion Chef Instrument using Ion 540 Chef Reagents (ThermoFisher Scientific, US), Ion 540 Chef Solutions (ThermoFisher Scientific, US), Ion 540 Chef supplies (ThermoFisher Scientific, US) and an Ion 540 Chip kit (ThermoFisher Scientific, US). These reagents, chips, as well as the pooled libraries were loaded onto the Ion Chef Instrument. During template preparation, each unique DNA amplicon was immobilized on an ISP through hybridization of the P1 adapter to complementary sequences on

the ISP. The amplicon was then clonally amplified using emulsion PCR to create an ISP containing multiple copies of the same amplicon. This was required to increase the signal from each amplicon to detectable levels (Slatko, Gardner and Ausubel, 2018).The ISPs were deposited into the microwells of an Ion 540™ Chip in a process known as chip loading. One pooled library was loaded per chip. The Ion Chef Instrument can only template and load two pooled libraries per run. Therefore, two separate runs were performed over the course of two days to produce a total of four loaded Ion 540™ Chips.

2.5.3 Sequencing

Sequencing of the loaded Ion 540™ Chip was done on the Ion S5 GeneStudio sequencer. The sequencer was initialized before sequencing to load and prepare all the reagents. An Ion S5™ Sequencing Reagents cartridge (ThermoFisher Scientific, US), Ion S5™ Wash Solution (ThermoFisher Scientific, US), Ion S5™ Cleaning Solution (ThermoFisher Scientific, US) and a previously sequenced chip was loaded onto the Ion S5™ sequencer for initialization. After completion of initialization, the old chip was removed, and one chip was sequenced at a time. The second chip was stored in a chip holder at 4°C during this time.

The Ion S5™ uses semiconductor sequencing technology. Chips are sequentially flooded by sequencing reagents containing a single species of nucleotides (dATP, dCTP, dGTP, dTTP) at a time. This is referred to as a flow. The incorporation of a complementary nucleotide into a growing DNA strand by a DNA polymerase leads to the release of a hydrogen ion (H^+). This results in a pH change of the solution which is detected by ion sensors incorporated on the chip and converted into a voltage signal. An algorithm subsequently converts this signal into a base call. If no nucleotide is incorporated, no H^+ is released, no pH change occurs, and no signal is generated. The chip was flooded with Ion S5™ Wash Solution in between nucleotide flows to remove unincorporated nucleotides from the previous flow (Slatko, Gardner and Ausubel, 2018; Pereira, Oliveira and Sousa, 2020). A total number of 550 sequencing flows were performed per chip for a read length of ~200 bp. After completion of sequencing of two chips, Ion S5™ Cleaning Solution was used to ensure that all the reagents from

the run were removed from the instrument. Sequencing of the four chips took two days to complete.

2.5.4 Data analysis

Following sequencing, the Ion Torrent Suite software performed data analysis for each run. This included generating reports containing quality control metrics and performance parameters for each run and patient, alignment of sequences to the human reference genome and variant calling. Through these processes, the software generated binary alignment map (BAM), BAM index (BAI) and variant call format (VCF) files required for variant annotation and interpretation.

2.5.4.1 Quality control metrics and performance parameters

The quality of the sequencing data was assessed by the metrics given in the sequencing run report generated by the Ion Torrent Suite™ software. The quality of the sequencing run was represented by metrics for both unaligned and aligned reads. The metrics were generated during primary pipeline processing, base calling and signal processing. Important quality metrics and their minimum acceptable threshold included throughput or total data yield (10-15 Gb), total number of reads (60-80 million), ISP loading density (>70%), polyclonality (<35%), low quality reads (<20%), usable reads which passed all quality filers (>55%) and sequencing read lengths (~200 bp) (De Abreu et al., 2016). The Coverage Analysis plugin generated statistics related to the sequence coverage for each patient. The minimum mean sequencing depth of coverage for WES is recommended to be between 75-100x. A minimum read depth of 15x with balanced reads is recommended for each target region to ensure sufficient confidence in the base call (Rehder et al., 2021).

2.5.4.2 Alignment and variant calling

The Ion Torrent Suite™ software aligned the sequencing reads to the human reference genome, Genome Reference Consortium Human Build 37 (GRCh37 or hg19). A BAM file which stored the aligned reads, as well as a BAI file which is required to view a BAM file were generated. In addition, the Variant Caller plugin generated a

VCF file which contained all the sequence variations between the human reference genome and the specific patient (Rehder et al., 2021).

2.6 Variant annotation and interpretation

Interpretation of variants included variant annotation, variant filtering and prioritization and classification of variants (Figure 2.4).

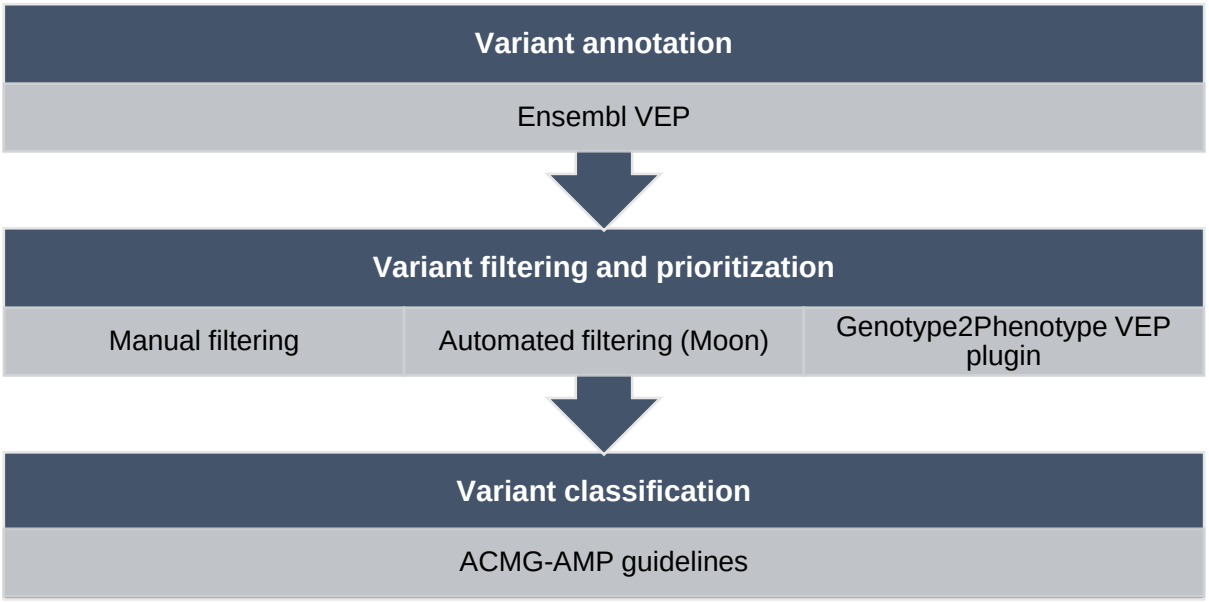


Figure 2.4. Diagram depicting the variant interpretation workflow. This included annotation of variants, three different strategies for filtering and prioritizing variants and classification of variants.

2.6.1 Variant annotation

Ensembl Variant Effect Predictor (VEP), an online tool, was used to annotate all patient variants using VCF files as input data (McLaren et al., 2016). Variant annotation is the process of assigning information to variants to aid in the interpretation thereof. Ensembl VEP uses a variety of databases such as dbNSFP (Liu et al., 2020) to retrieve relevant information for a variant. It provided information for different transcripts of a variant such as the genomic location, gene symbol, Human Genome Variation Society (HGVS) nomenclature (Den Dunnen et al., 2016), allele frequency data, etc. The type of variant, functional consequences of the variant and pathogenicity prediction scores from various tools were also given. Additionally, the clinical

significance as reported by ClinVar (Landrum et al., 2018) and identifiers such as PubMed IDs were provided for previously identified variants.

2.6.2 Variant filtering and prioritization

Three different filtering strategies were used to prioritize possible disease-causing variants for in-depth analysis. This included a manual filtering workflow, an automated filtering workflow using an online tool and the use of an additional plugin for Ensembl VEP.

2.6.2.1 Manual filtering workflow

The variant annotation generated by Ensembl VEP was investigated to reduce thousands of variants generated by WES to a few candidate variants. The filtering strategy that was used is shown in Figure 2.5. Variants were only interpreted in the context of the canonical transcript.

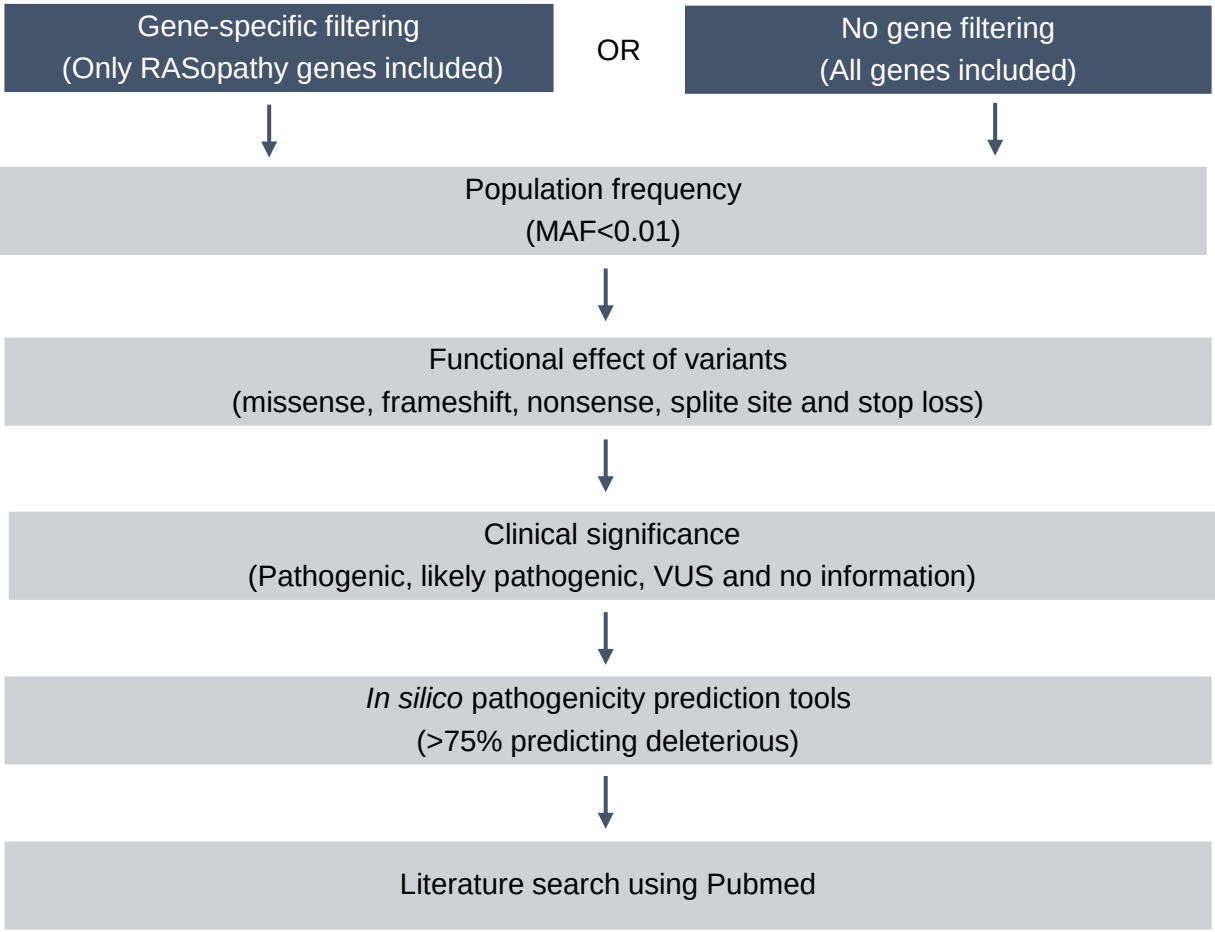


Figure 2.5. Diagram showing the manual filtering strategy that was used in this study. A gene-specific approach which only included known RASopathy genes was undertaken first. If no disease-causing variants were identified using this approach, a second filtering strategy was used which did not exclude any genes.

Two filtering approaches were undertaken in which variants were filtered in a step wise manner. The first approach applied a gene-specific filter in which only variants located in genes known to cause a RASopathy (Table 2.1) were prioritized and included in the next filtering step. The RASopathy gene list was compiled from literature and gene-disease curations (Rehm et al., 2015). This gene-specific filtering was applied since all patients were suspected of having a RASopathy disorder. The second filtering approach was only used for patients in whom no causal variant was identified with the first approach. As opposed to the first filtering approach, this strategy included variants in all the sequenced genes in the next filtering step. The filtering strategies were the same for both approaches after the initial step.

Following the initial filtering step, a population frequency filter was applied to the remaining variants in order to assess the prevalence of these variants across general healthy human populations. Variants were filtered based on their minor allele frequency (MAF) and only rare variants with a MAF of less than 1% (MAF<0.01) were prioritized and included in the next step. Allele frequency information was retrieved by Ensembl VEP from population databases such as the Genome Aggregation Database (gnomAD) (Gudmundsson et al., 2021), Exome Aggregation Consortium (ExAC) (Karczewski et al., 2017) and 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015). Since these population databases consist of mostly healthy individuals, it is assumed that variants not observed in these cohorts or observed at very low frequencies are more likely to cause disease phenotypes.

Rare variants remaining after the population frequency filtering were filtered based on the predicted functional effect of the variants. Missense, nonsense, stop-loss, splice and frameshift variants were prioritized (Dashti and Gamieldien, 2017; Pereira, Oliveira and Sousa, 2020). The next step included filtering of the remaining variants based on the clinical significance of these variants as reported in reputable public disease variant databases such as ClinVar (Landrum et al., 2018). ClinVar reports on the clinical significance of previously detected variants and their association with phenotypes while also providing supporting evidence for the data. Variants reported as benign or likely benign by ClinVar were excluded, while variants were retained if reported to be pathogenic, likely pathogenic or VUS. Variants that were not reported by ClinVar were also retained.

The remaining variants were subsequently filtered based on pathogenicity prediction scores by multiple *in silico* prediction tools. The tools assigned scores to variants based on different criteria to predict whether a variant is damaging. The tools that were used in this study include: Polymorphism Phenotyping (PolyPhen-2) (Adzhubei et al., 2010), Sorting Intolerant From Tolerant (SIFT) (Sim et al., 2012), MutationTaster (Schwarz et al., 2014), Protein Variation Effect Analyzer (PROVEAN) (Choi and Chan, 2015), Functional Analysis Through Hidden Markov Models (FATHMM) (Shihab et al., 2013), Mutation assessor (Reva, Antipin and Sander, 2011), Deleterious Annotation of genetic variants using Neural Networks (DANN) (Quang, Chen and Xie, 2015), Combined Annotation Dependent Depletion (CADD) (Rentzsch et al., 2019) and Rare

Exome Variant Ensembl Learner (REVEL) (Ioannidis et al., 2016). The scores assigned by all the tools were considered to determine the likely pathogenicity of a variant. Variants predicted to be damaging, intolerable or deleterious by >75% of tools were prioritized. Lastly, a literature search for shortlisted variants was performed using PubMed to establish a potential disease-causing role.

2.6.2.2 Automated filtering workflow

Moon (Diploid, Leuven, Belgium), a commercial genomic prioritization tool, was used for patients for whom no causal variants were identified using the manual filtering strategy. Moon aids in prioritizing potential causal variants for rare Mendelian genetic disorders using a phenotype-driven automated filter pipeline based on artificial intelligence algorithms and a disorder model. Input data included a VCF file and patient information such as sex, age of onset and clinical features provided as Human Phenotype Ontology (HPO) terms. Moon annotated the VCF file with information such as population frequency, known pathogenicity, inheritance, variant effect and phenotype overlap. Variants were then filtered, and a shortlist of relevant candidate variants was generated. Only genes with a known gene-disease association were considered (O'Brien et al., 2021). The variant shortlist was reviewed to identify possible casual variants. This entailed assessing the quality of each variant, as well as its pathogenicity and clinical significance for our patient.

2.6.2.3 Genotype2Phenotype Variant Effect Predictor plugin

The Genotype2Phenotype (G2P) Ensembl VEP plugin was used for the remaining patients for whom no causal variant was identified using the first two filtering strategies. This filtering workflow was performed with the assistance of a bioinformatics scientist at the Sydney Brenner Institute for Molecular Bioscience. A custom designed command line was used to select the most likely causal variants. The DDD study developed the G2P database with a curated list of all known genes causative of developmental disorders (Wright et al., 2015). These genes are categorized based on confidence of causing disease, variant consequence and allelic requirement (monoallelic or biallelic). In addition to the database, the DDD study developed a plugin that runs as an extension to Ensembl VEP analysis (Thormann et

al., 2019). The VEP-G2P plugin required a VCF file as input and used the default Ensembl VEP annotations, as well as knowledge from the G2P datasets to filter variants, identify and generate a list of potentially causal genotypes. Filtering such as the presence of a variant in a G2P gene, allelic requirement at a locus, variant consequence and allele frequency data were utilized (Thormann et al., 2019). The output was reviewed for data quality and clinical relevance of the variants.

2.6.3 Variant classification

Prioritized variants were visualized using the Integrative Genomics Viewer (IGV) and BAM and BAI files as input data (Robinson et al., 2017). Visual inspection of the data helped determine if a variant could be called with confidence and reduced the risk of false positives due to sequencing artefacts and low-quality data. Variants were assessed to determine whether the variant had adequate depth of coverage and whether there was strand bias.

The prioritized variants with high confidence were classified using the ACMG-AMP guidelines (Richards et al., 2015) and the ClinGen RASopathy expert panel specifications (Gelb et al., 2018) to determine the pathogenicity of the variants. The ACMG-AMP guidelines use a 5-tier classification system where variants are classified as pathogenic, likely pathogenic, VUS, likely benign or benign. The classification is based on applying a variety of criteria weighted as very strong, strong, moderate or supporting to a variant.

2.7 Sanger sequencing validation

Variants that were identified through WES and classified as either pathogenic or likely pathogenic according to the ACMG-AMP guidelines were validated using Sanger sequencing. In addition, a single probe deletion identified by MLPA was also validated using this method. Sanger sequencing is used in both clinical and research settings to confirm the presence of NGS variants (Beck, Mullikin and Biesecker, 2016).

2.7.1 Primer design

The web interface of Primer3 (v4.1.0), an online tool that evaluates and returns acceptable primer pairs for an entered template sequence, was used to design primers in this study (Table 2.2) (Untergasser et al., 2012). The template sequence included the variant and sequences up- and downstream thereof. Sequences were retrieved from the FASTA files of the RefSeq genomic reference sequences for the specified genes as obtained from the National Center for Biotechnology Information (NCBI) (Sayers et al., 2021). The specificity of primers was evaluated using an NCBI tool called Primer-BLAST. Primers should be specific to their intended target region and should not bind to any other regions in the human genome, as this will lead to aberrant results. SNPCheck (v3), an online tool, was used to identify any potential single nucleotide polymorphisms (SNPs) in the primer region using population frequency data from 1000 Genomes Project, dbSNP and the Exome Sequencing Project (Lai and Love, 2012). The presence of a SNP in these regions would prevent the primers from binding to their intended target region.

Table 2.2. Primers for Sanger sequencing validation of disease-causing variants.

Gene	**Variant**	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Product size (bp)**
RIT1	c.297T>G	ATAGCAGGTAATCGACATGCGT	AGAACCACAGGTGTATCGTCAG	262
SON	c.5753_5756del	TCAAGACGCAGGAGGAGAAG	CGGCTTGGGGAAATGCTAAA	266
CDK8	c.95A>T	CAGAGGCTGTGACAATGGACTATG	CGCAGGATTTCTCGTGGGTTT	332
NF1	Exon 6 deletion	TTCCTAGCAGACAACTATCGAG	TCTACCCAGTTCCAAAATGCC	480

2.7.2 Primer optimization and sequencing

Primers were ordered from Inqaba BioTech™ (Inqaba Biotechnical Industries, Pretoria, South Africa) and resuspended in TE buffer to a concentration of 100 ng/µl. The optimal annealing temperature of each primer pair was determined using gradient PCR and standard PCR reagents. The PCR thermocycler conditions and annealing temperature for each primer pair are shown in Table 2.3. Successful amplification of the target regions was assessed by running the PCR products on a 3% agarose gel

with visualization on the Omega Fluor™ Gel Documentation System. A 50 bp molecular weight marker was used to determine the size of the PCR products.

Table 2.3. PCR thermocycler conditions for Sanger validation.

Step	Temperature	Duration	Cycles
Initial denaturation	95°C	5 minutes	1
Denaturation	95°C	1 minute	30
Annealing*	55°C (*RIT1*) 59°C (*SON*) 53°C (*CDK8*) 55°C (*NF1*)	1 minute	
Elongation	72°C	1 minute	
Final elongation	72°C	10 minutes	1
Hold	15°C	∞	1

*Annealing temperature was dependent on the specific primer pair.

Following PCR, an enzymatic clean-up of the PCR products was done using ExoSAP-IT™ Express (ThermoFisher Scientific, US) to remove excess deoxynucleotide triphosphates (dNTPs) and primers. Cycle sequencing was subsequently performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific, US) and the thermocycler conditions in Table 2.4. Samples were sequenced in both the forward and reverse direction. Sanger sequencing is based on the use of fluorescently labelled dideoxynucleotide triphosphates (ddNTPs) in addition to dNTPs. The ddNTPs lack the 3'-OH group which is required to form a phosphodiester bond with the 5'-phospate group of an adjacent nucleotide during the elongation step of PCR. Extension is thus terminated at random positions as a result of the incorporation of a ddNTP, resulting in multiple fragments of varying lengths (Sanger, Nicklen and Coulson, 1977). Excess ddNTPs were removed from the sequencing reactions with the DyeEx 2.0 Spin Kit (Qiagen, Hilden, Germany) using gel filtration and spin columns. Sequencing reactions were subsequently dried using the Concentrator Plus vacuum centrifuge (Eppendorf, Hamburg, Germany) and resuspended in Hi-Di Formamide. Capillary electrophoresis was performed on the Applied Biosystems 3500xL Genetic Analyzer.

Table 2.4. Cycle sequencing thermocycler conditions.

Temperature	Duration	Cycles
96°C	30 seconds	1
50°C	15 seconds	25
60°C	4 minutes	
15°C	∞	1

CLC Bio Main Workbench software (Qiagen, Hilden, Germany) was used for the analysis of sequencing data. Low quality sequences at the ends of electropherograms were removed through trimming. Secondary peak calling was performed to identify positions where two peaks were present indicating a heterozygous variant. Samples were then aligned to their respective genomic reference sequences obtained from NCBI. Sequences were then investigated to identify conflicts, e.g., positions that differed from that of the reference sequence or positions where secondary peaks were called. Such positions were analysed to determine if it matched the variant identified in WES, thereby confirming the presence of the variant.

2.8 Feedback of findings

Variants identified in this study that were classified as pathogenic or likely pathogenic were given back to the referring clinician/genetic counsellor through a research report. The clinician/genetic counsellor communicated the results back to the patient/family if it was deemed appropriate by them. Variants of uncertain significance were not reported back to the referring clinician/genetic counsellor.

3. RESULTS

3.1 Patient cohort

A total of 22 patients with a clinical diagnosis of one of the RASopathies and in whom no disease-causing variants were identified using a targeted gene panel containing known RASopathy genes were selected for inclusion in this study (Table 3.1). The majority of patients (21) were of African ancestry and only one patient was of Caucasian ancestry. The age of the patients ranged from 2 to 37 years of which 45.5% were female and 54.5% were male. Thirteen patients had a definitive clinical diagnosis of NF1, whereas one patient's clinical features were a combination of that seen in NF1 and Noonan syndrome. Seven patients had clinical features consistent with a diagnosis of Noonan syndrome and one patient had typical RASopathy features, however it was non-specific, and the patient could not be allocated to a specific disorder in the RASopathy group. Based on their clinical diagnosis, fourteen patients were selected for inclusion in MLPA to screen for CNVs in the *NF1* gene and the remaining eight patients were selected for variant screening using WES.

Table 3.1. Characteristics of the patient cohort and variant screening method based on their suspected clinical diagnosis.

	PATIENT	SEX	AGE (YEARS)	ETHNICITY	CLINICAL DIAGNOSIS	FIRST LINE TESTING METHOD	REFLEX VARIANT SCREENING METHOD
1	IDP59	F	33	African	NF1	Diagnostic panel	MLPA
2	DD3072	M	12	African	NF1	Diagnostic panel	MLPA
3	FRASC3	M	17	African	NF1	Mudau's panel	MLPA
4	FRASC43	M	37	African	NF1	Mudau's panel	MLPA
5	FRASC65	F	11	African	NF1	Mudau's panel	MLPA
6	FRASC68	M	5	African	NF1	Mudau's panel	MLPA
7	FRASC82	M	30	African	NF1	Mudau's panel	MLPA
8	FRASC83	M	22	African	NF1	Mudau's panel	MLPA
9	FRASC89	F	9	African	NF1	Mudau's panel	MLPA
10	FRASC95	F	32	African	NF1	Mudau's panel	MLPA
11	IDP113	F	12	African	NF1	Diagnostic panel	MLPA
12	DD3546	F	5	African	NF1	None	MLPA
13	FRASC70	M	10	African	NF1/Noonan syndrome	Mudau's panel	MLPA
14	IDP89	M	7	African	NF1	Diagnostic panel	MLPA
15	FRASC85	M	19	Caucasian	Noonan syndrome	Mudau's panel	WES
16	FRASC90	F	16	African	Noonan syndrome	Mudau's panel	WES
17	DD3198	M	4	African	RASopathy	Diagnostic panel	WES
18	FRASC37	F	8	African	Noonan syndrome	Mudau's panel	WES
19	FRASC55	F	26	African	Noonan syndrome	Mudau's panel	WES
20	IDP17	M	4	African	Noonan syndrome	Diagnostic panel	WES
21	IDP27	F	2	African	Noonan syndrome	Diagnostic panel	WES
22	DD3425	M	3	African	Noonan syndrome	Diagnostic panel	WES

3.2 MLPA

3.2.1 MLPA data quality and troubleshooting

The quality of the MLPA data for all samples and controls included in the P081, P082 and P122 probemix reactions were assessed prior to analysis. This included both the fragment quality metrics and comparative quality metrics generated by the Coffalyser.Net™ software (Figure 3.1). The electropherograms were also evaluated to determine the cause when a sample failed quality control (Figure 3.2, Figure 3.3, Figure 3.4).

The majority of samples had acceptable scores for all quality metrics. All samples had an optimal score (>90%) for the FRSS quality metric which indicated acceptable capillary electrophoresis quality based on the size marker. In addition, DNA concentrations and denaturation were optimal for all samples, except one (FRASC3, P081 probemix). The X chromosome and Y chromosome internal control fragments were used to confirm the sex of the patients (Figure 3.1).

a) Fragment analysis

	sample name	sample type	bin smpl	FRSS	FMRS	probes	DNA	DD
1	P081R_Blank	no DNA				0/0		
2	P081R_DD3072_	sample				45/46		
3	P081R_DD3546_	sample				46/46		
4	P081R_FRASC3_	sample				28/46		
5	P081R_FRASC43_	sample				46/46		
6	P081R_FRASC65_	sample				46/46		
7	P081R_FRASC68_	sample				46/46		
8	P081R_FRASC70_	sample				46/46		
9	P081R_FRASC82_	sample				46/46		
10	P081R_FRASC83_	sample				46/46		
11	P081R_FRASC89_	sample				46/46		
12	P081R_FRASC95_	sample				46/46		
13	P081R_IDP59_	sample				46/46		
14	P081R_IDP89_	sample				46/46		
15	P081R_IDP113_	sample				46/46		
16	P081R_Negative_	reference				46/46		
17	P081R_Negative_	reference				46/46		
18	P081R_Negative_	reference				46/46		
19	P081R_Positive_	positive reference				46/46		

Comparative analysis

	sample name	sample type	FMRS	X	Y	gender	analy...	CAS	PSLP	FSLP	RSQ	RPQ
1	P081R_DD3072_	sample				male						
2	P081R_DD3546_	sample				female						
3	P081R_FRASC3_	sample				male						
4	P081R_FRASC43_	sample				male						
5	P081R_FRASC65_	sample				female						
6	P081R_FRASC68_	sample				male						
7	P081R_FRASC70_	sample				male						
8	P081R_FRASC82_	sample				male						
9	P081R_FRASC83_	sample				male						
10	P081R_FRASC89_	sample				female						
11	P081R_FRASC95_	sample				female						
12	P081R_IDP59_	sample				female						
13	P081R_IDP89_	sample				male						
14	P081R_IDP113_	sample				female						
15	P081R_Negative_	reference				female						
16	P081R_Negative_	reference				female						
17	P081R_Negative_	reference				female						
18	P081R_Positive_	positive reference				male						

b) Fragment analysis

	sample name	sample type	bin smpl	FRSS	FMRS	probes	DNA	DD
1	P082_Blank	no DNA				0/0		
2	P082_DD3072_	sample				44/44		
3	P082_DD3546_	sample				44/44		
4	P082_FRASC3_	sample				44/44		
5	P082_FRASC43_	sample				44/44		
6	P082_FRASC65_	sample				44/44		
7	P082_FRASC68_	sample				44/44		
8	P082_FRASC70_	sample				44/44		
9	P082_FRASC82_	sample				44/44		
10	P082_FRASC83_	sample				44/44		
11	P082_FRASC89_	sample				44/44		
12	P082_FRASC95_	sample				44/44		
13	P082_IDP59_	sample				44/44		
14	P082_IDP89_	sample				44/44		
15	P082_IDP113_	sample				44/44		
16	P082_Negative_	reference				44/44		
17	P082_Negative_	reference				44/44		
18	P082_Negative_	reference				44/44		
19	P082_Positive_	positive reference				44/44		

Comparative analysis

	sample name	sample type	FMRS	X	Y	gender	analy...	CAS	PSLP	FSLP	RSQ	RPQ
1	P082_DD3072_	sample				male						
2	P082_DD3546_	sample				female						
3	P082_FRASC3_	sample				male						
4	P082_FRASC43_	sample				male						
5	P082_FRASC65_	sample				female						
6	P082_FRASC68_	sample				male						
7	P082_FRASC70_	sample				male						
8	P082_FRASC82_	sample				male						
9	P082_FRASC83_	sample				male						
10	P082_FRASC89_	sample				female						
11	P082_FRASC95_	sample				female						
12	P082_IDP59_	sample				female						
13	P082_IDP89_	sample				male						
14	P082_IDP113_	sample				female						
15	P082_Negative_	reference				female						
16	P082_Negative_	reference				female						
17	P082_Negative_	reference				female						
18	P082_Positive_	positive reference				male						

c) Fragment analysis

	sample name	sample type	bin smpl	FRSS	FMRS	probes	DNA	DD
1	P122R_Blank	no DNA				3/0		
2	P122R_DD3072_	sample				35/35		
3	P122R_DD3546_	sample				35/35		
4	P122R_FRASC3_	sample				35/35		
5	P122R_FRASC43_	sample				35/35		
6	P122R_FRASC65_	sample				35/35		
7	P122R_FRASC68_	sample				35/35		
8	P122R_FRASC70_	sample				35/35		
9	P122R_FRASC82_	sample				35/35		
10	P122R_FRASC83_	sample				35/35		
11	P122R_FRASC89_	sample				35/35		
12	P122R_FRASC95_	sample				35/35		
13	P122R_IDP59_	sample				35/35		
14	P122R_IDP89_	sample				35/35		
15	P122R_IDP113_	sample				35/35		
16	P122R_Negative_	reference				35/35		
17	P122R_Negative_	reference				35/35		
18	P122R_Negative_	reference				35/35		
19	P122R_Positive_	positive reference				35/35		

Comparative analysis

	sample name	sample type	FMRS	X	Y	gender	analy...	CAS	PSLP	FSLP	RSQ	RPQ
1	P122R_DD3072_	sample				male						
2	P122R_DD3546_	sample				female						
3	P122R_FRASC3_	sample				male						
4	P122R_FRASC43_	sample				male						
5	P122R_FRASC65_	sample				female						
6	P122R_FRASC68_	sample				male						
7	P122R_FRASC70_	sample				male						
8	P122R_FRASC82_	sample				male						
9	P122R_FRASC83_	sample				male						
10	P122R_FRASC89_	sample				female						
11	P122R_FRASC95_	sample				female						
12	P122R_IDP59_	sample				female						
13	P122R_IDP89_	sample				male						
14	P122R_IDP113_	sample				female						
15	P122R_Negative_	reference				female						
16	P122R_Negative_	reference				female						
17	P122R_Negative_	reference				female						
18	P122R_Positive_	positive reference				male						

Figure 3.1. Images generated by Coffalyser.Net™ software showing the MLPA data quality metrics for fragment and comparative analysis for all samples and controls for the three probemixes: a) P081; b) P082; c) P122. The data was of good quality for most of the samples indicated by green bars and ticks. P081 data for FRASC89 was not optimal (FMRS, CAS and PSLP flagged), but was of acceptable quality for analysis. DD3072 failed quality control for P081 (FMRS, probes, CAS and RPQ failed) and capillary electrophoresis was repeated. FRASC82 failed quality control for P082 (FMRS and CAS flagged; PSLP failed) and the MLPA reaction had to be repeated. FRASC3 failed quality control for both P081 (probes and DD flagged; FMRS, CAS, PSLP, FSLP and RPQ failed) and P082 (CAS and RPQ flagged; PSLP failed) and both MLPA reactions had to be repeated.

Four samples did not pass quality control initially and their quality metrics will be described in this section. FRASC89 had acceptable, but not ideal quality scores for the P081 probemix reaction. The FMRS (55%), CAS (75%) and PSLP metrics were flagged. Based on the electropherogram pattern (Figure 3.2), showing a non-specific high primer peak and an overall low peak pattern for the MLPA probes, the quality concern was the result of reduced amplification efficiency resulting in a high percentage of residual primers. The data was of acceptable quality to proceed with analysis.

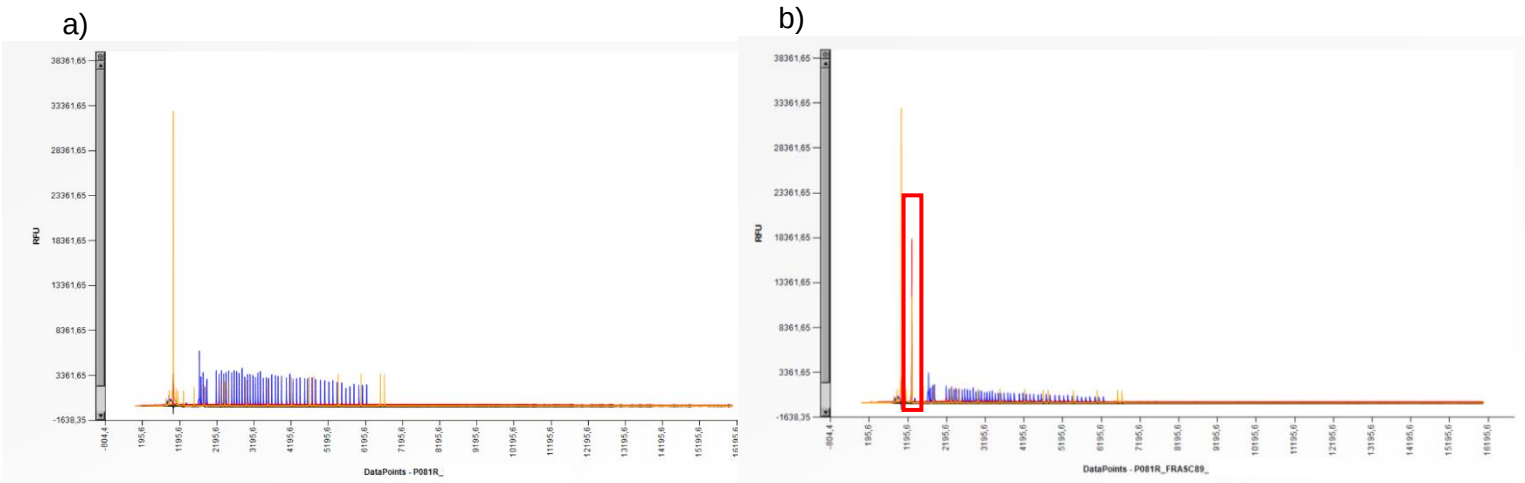


Figure 3.2. Raw data electropherograms generated by Coffalyser.Net™ software for the P081 probemix: a) A normal and acceptable quality sample. b) FRASC89 with the presence of a non-specific primer peak (indicated with a red block) and overall low peak pattern (low relative fluorescence units of the MLPA probes). This was indicative of a high percentage of residual primers resulting from a suboptimal PCR reaction. The data for FRASC89 was of acceptable quality for analysis.

Data for DD3072 P081 probemix reaction failed FMRS (10%), probe quantity, CAS (25%), and RPQ quality metrics. The low-quality data for this sample was caused by one of the reference probes deviating from the set bin, based on the quality scores and the electropherogram for this sample (Figure 3.3). As a result, the reference probe was not labelled resulting in a drop out of the probe and consequently affecting the number of probes detected. This was caused by migration problems during capillary electrophoresis and resolved by repeating the electrophoresis.

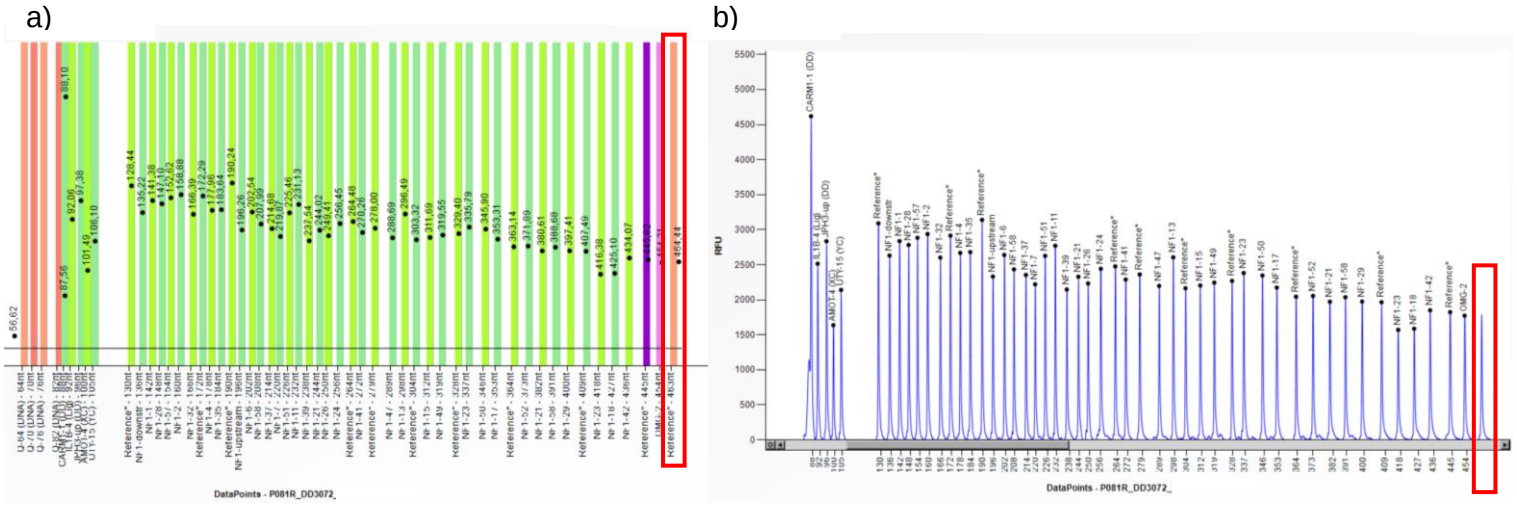


Figure 3.3. Images generated by Coffalyser.Net™ software for DD3072 P081 probemix: a) Binning profile showing one reference probe (indicated by a red block) falling outside of the bin. b) Genomic electropherogram showing one reference probe (indicated by a red block) not labelled due to the binning problem. The probe was consequently not detected, affecting the data quality. This was caused by migration problems during capillary electrophoresis.

FRASC82 failed quality control for the P082 probemix due to flagged FMRS (70%) and CAS (60%) quality metrics, as well as a failed PSLP quality check. It was caused by reduced amplification efficiency due to evaporation during hybridization or ligation. This resulted in a reduction in the signal intensity of the fragments proportional to their length referred to as sloping (Figure 3.4). The data was not reliable to analyse due to the extent of the sloping. The MLPA reaction was repeated which resolved the problem.

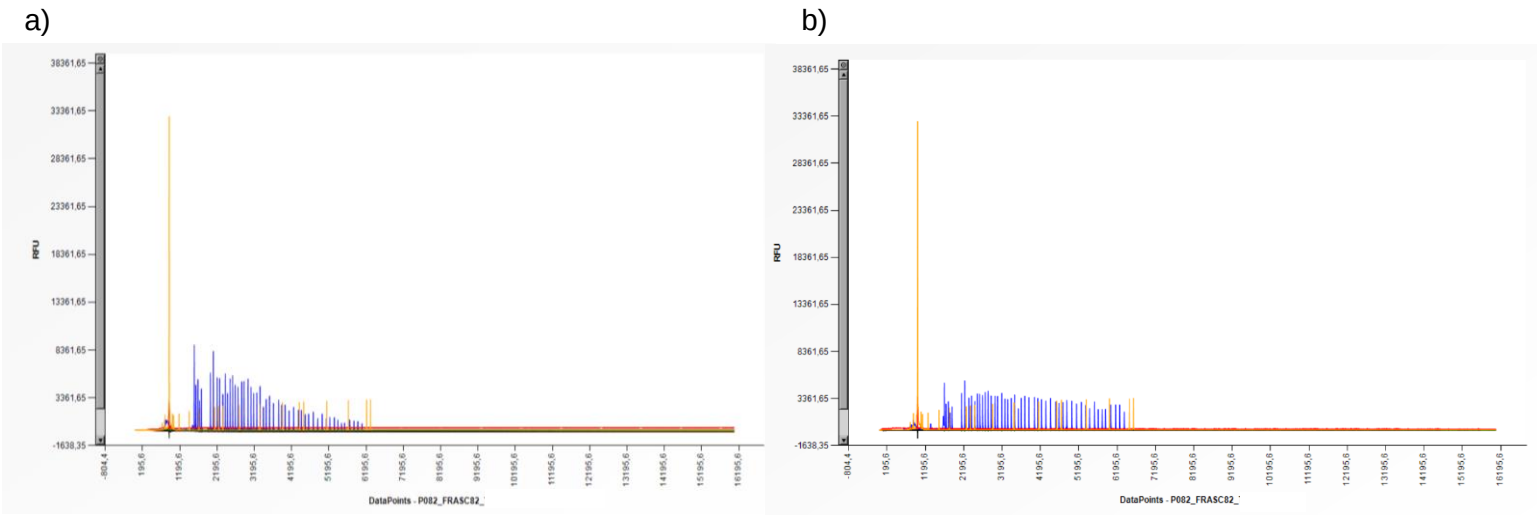


Figure 3.4. Raw data electropherograms generated by Coffalyser.Net™ software for FRASC82 P082 probemix. a) Initial MLPA reaction showing a sloping peak pattern of the MLPA probes indicative of evaporation during hybridization or ligation which resulted in reduced amplification efficiency. This data was not suitable for analysis. b) Repeated MLPA reaction showing no sloping of the MLPA probes, i.e., an even peak pattern. This data was of acceptable quality for analysis.

Data for FRASC3 was not of acceptable quality for analysis for both the P081 and P082 probemix reactions, however it was of sufficient quality for the P122 probemix reaction. DNA denaturation was incomplete for the P081 probemix reaction. In addition, the FMRS score (0%) indicated failure of amplification. As a result, the other quality metrics were also adversely affected including the CAS (25%), PSLP, FSLP and RPQ metrics. Similarly, the CAS (70%) and RPQ metrics were flagged and the PSLP quality check failed for the P082 probemix reaction. Based on these scores and the electropherograms (Figure 3.5) which indicated a high non-specific peak for P081 and sloping for both P081 and P082, the reaction failure was likely the result of contaminants in the sample DNA. MLPA reactions for both P081 and P082 probemixes were repeated, however quality control failed again and thus this sample was excluded from analysis.

a)

b)

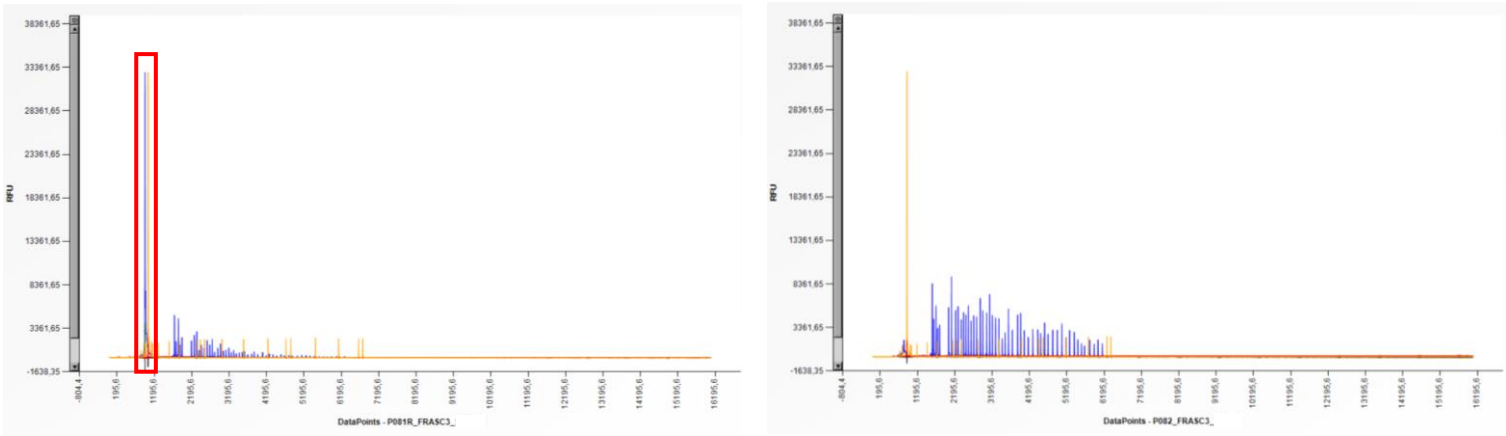


Figure 3.5. Raw data electropherogram generated by Coffalyser.Net™ software for FRASC3 a) P081 and b) P082 probemixes. a) Electropherogram showing a high non-specific peak (blue peak indicated by red block) and a sloping peak pattern. b) Electropherogram showing sloping of the MLPA probes indicating a reaction failure likely as a result of contaminants in the sample DNA. The data was not suitable for analysis.

3.2.2 MLPA data analysis

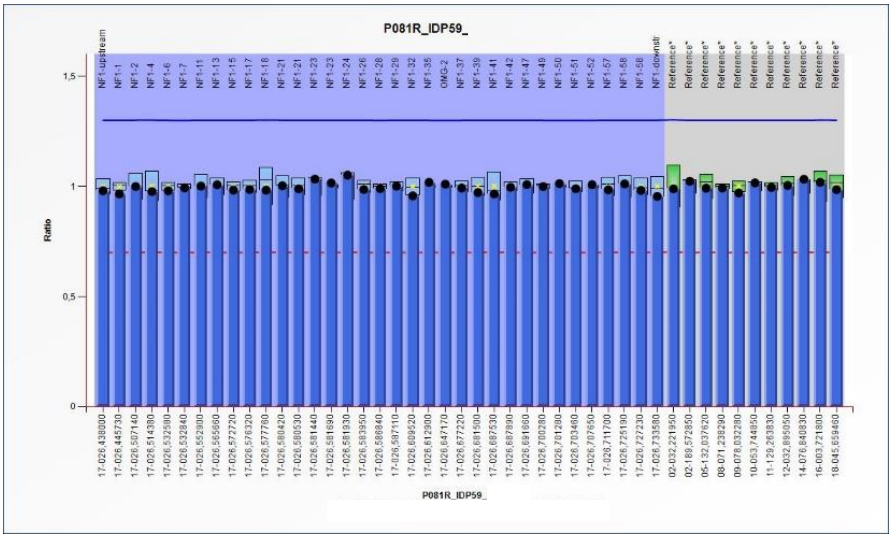
This section describes the results obtained for thirteen patients that passed the standard MLPA quality checks and were therefore included in MLPA analysis to detect CNVs of the *NF1* gene and genes flanking *NF1.*

3.2.2.1 Negative patients

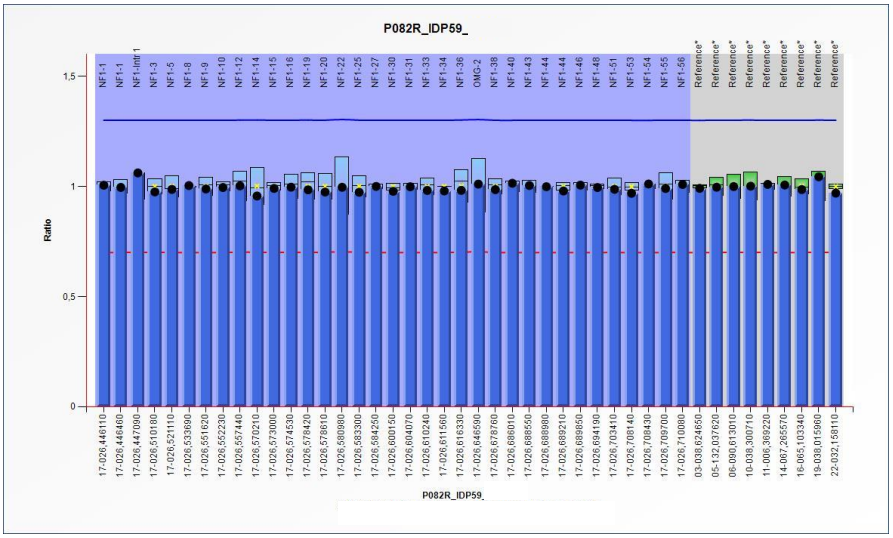
No CNVs were detected for nine of the patients using the P081, P082 and P122 probemixes. These patients were: IDP59, FRASC43, FRASC65, FRASC68, FRASC82, FRASC83, FRASC89, FRASC95 and FRASC70. Eight of these patients had a clinical diagnosis of NF1 and one patient's phenotype was consistent with both NF1 and Noonan syndrome (FRASC70).

Ratio charts generated by the Coffalyser.Net™ software for IDP59 for the three probemixes (P081, P082 and P122) are shown in Figure 3.6 as an example of a negative result. The other eight patients had similar ratio charts. The reference probes for the three probemixes had final ratios of approximately 1. This represented two copies of the target regions in the patient DNA compared to the reference sample DNA and was expected. All probes targeting the exons of the *NF1* gene and flanking genes also had final ratios of ~1. This indicated the presence of two copies of all 58 exons of *NF1,* as well as the genes flanking *NF1.* This means that we did not detect a single exon, multi-exon or whole *NF1* gene deletion or a deletion of the genes surrounding *NF1* for these patients using the P081, P082 or P122 MLPA probemixes. A genetic diagnosis for these patients could not be established.

a)



b)



c)

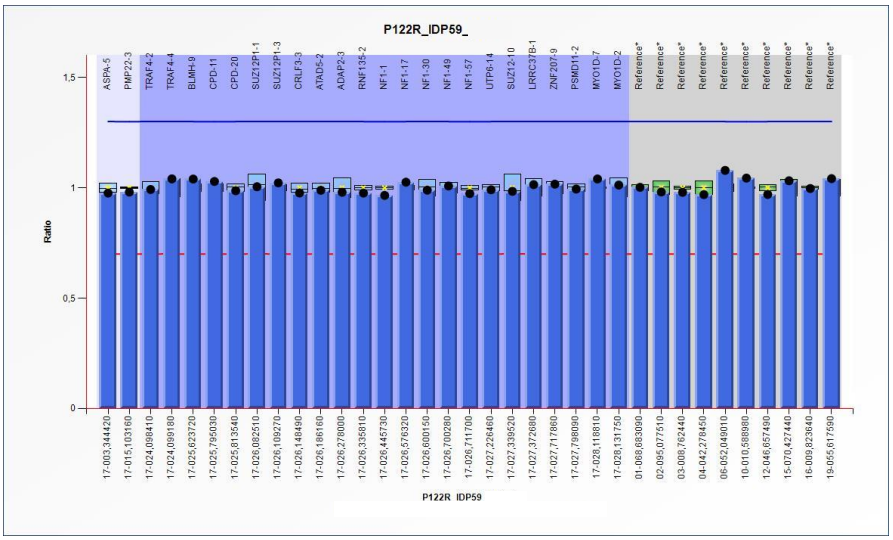


Figure 3.6. Negative result shown as ratio charts generated by Coffalyser.Net™ software for IDP59 for the a) P081; b) P082 and c) P122 probemix reactions. Each bar represents a single probe or target region. Genomic locations of each target region are displayed on the X-axis and the probe name (gene name and exon) is shown at the top of each bar. All probes had a ratio of ~1 (as seen on the Y-axis) which indicated the presence of two copies of the target regions i.e., 58 exons of *NF1,* as well as the genes flanking *NF1.* No single exon, multi-exon or whole *NF1* gene deletion or a deletion of the genes surrounding *NF1* were detected by the P081, P082 or P122 MLPA probemixes.

3.2.2.2 Positive patients

Deletions were identified for four of the patients using MLPA. This included two large deletions, one single exon deletion and one multi-exon deletion of *NF1* (Table 3.2).

Table 3.2. CNVs identified in *NF1* and flanking genes using MLPA P081, P082 and P122 probemixes.

Patient	**Type**	**Gene**
IDP113	Large deletion (Type 1)	*NF1* (exon 1 - 58)*; OMG; SUZ12P; CRLF3; ATAD5; ADAP2; RNF135; UTP6; SUZ12; LRRC37B*
IDP89	Large deletion (Type 3)	*NF1* (exon 1 - 58); *UTP6; LRRC37B*
DD3072	Multi-exon deletion	*NF1* (exon 19 and 20)
DD3546	Single exon deletion	*NF1* (exon 6)

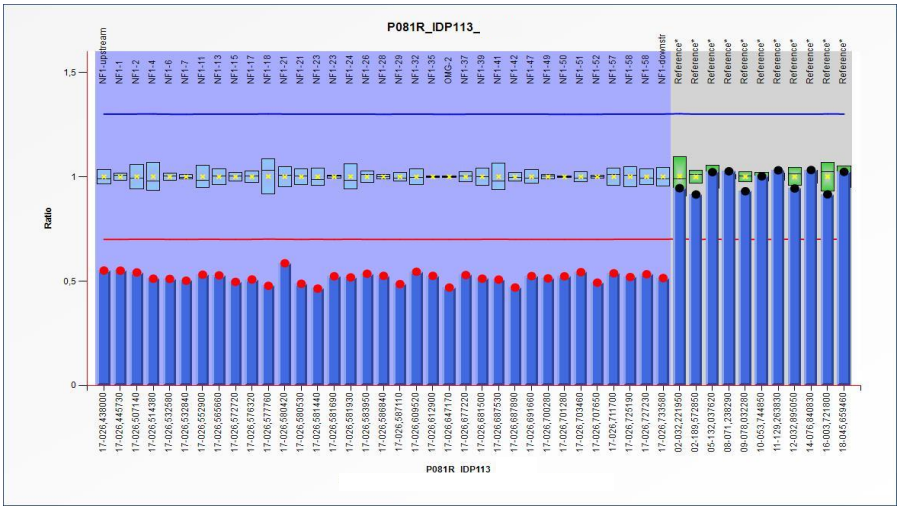
i. <u>IDP113</u>

IDP113 fulfilled the diagnostic criteria for NF1 due to the presence of multiple café au lait maculae (>6) of >1.5 cm in diameter, axillary and inguinal freckling, subcutaneous neurofibromas and iris Lisch nodules. Furthermore, she also had mild intellectual delay, dysmorphic features, macrocephaly and skeletal abnormalities.

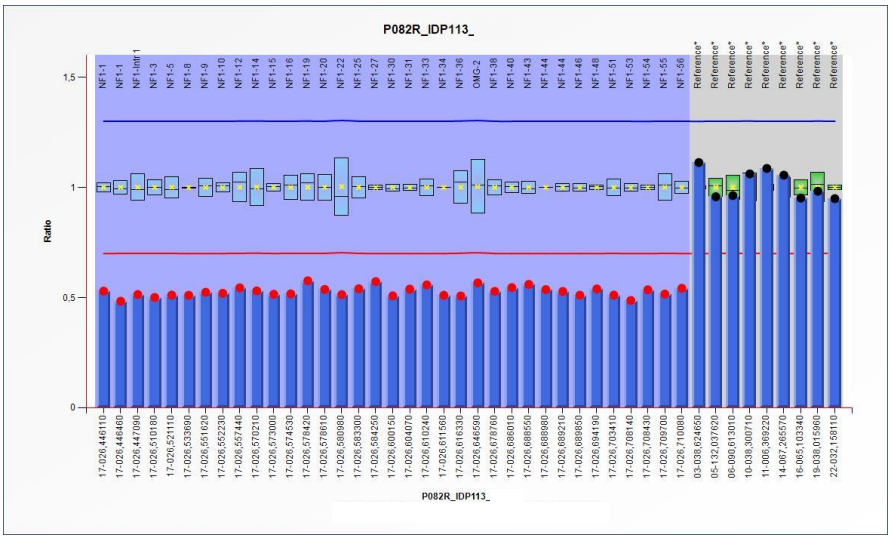
A large deletion encompassing the whole *NF1* gene and extending beyond *NF1* was identified in this patient (Figure 3.7). Ratio charts for P081 and P082 depict the heterozygous deletion of exons 1 to 58 of *NF1,* as well as a deletion of the *OMG* gene located in intron 36 of *NF1.* Probes corresponding to the following gene regions flanking *NF1* also indicated a heterozygous deletion as shown in the ratio chart for P122: *SUZ12P* exon 1 and 3 (363 kb and 336 kb upstream respectively); *CRLF3* exon 3 (297 kb upstream); *ATAD5* exon 2 (260 kb upstream); *ADAP2* exon 3 (168 kb upstream); *RNF135* exon 2 (110 kb upstream); *UTP6* exon 14 (515 kb downstream); *SUZ12* exon 10 (628 kb downstream) and *LRRC37B* exon 1 (661 kb downstream). The target regions described above had a final ratio of approximately 0.5 which is indicative of the presence of only one copy of these regions in the patient DNA, i.e., a heterozygous deletion. A heterozygous deletion of this extent is known as a type 1

deletion (Figure 3.8) and is the most commonly identified large deletion in patients with NF1. It is usually ~1.4 Mb in size (Kehrer-Sawatzki and Cooper, 2021), however the exact size cannot be determined by MLPA. This result confirms the clinical diagnosis of NF1 in this patient.

a)



b)



c)

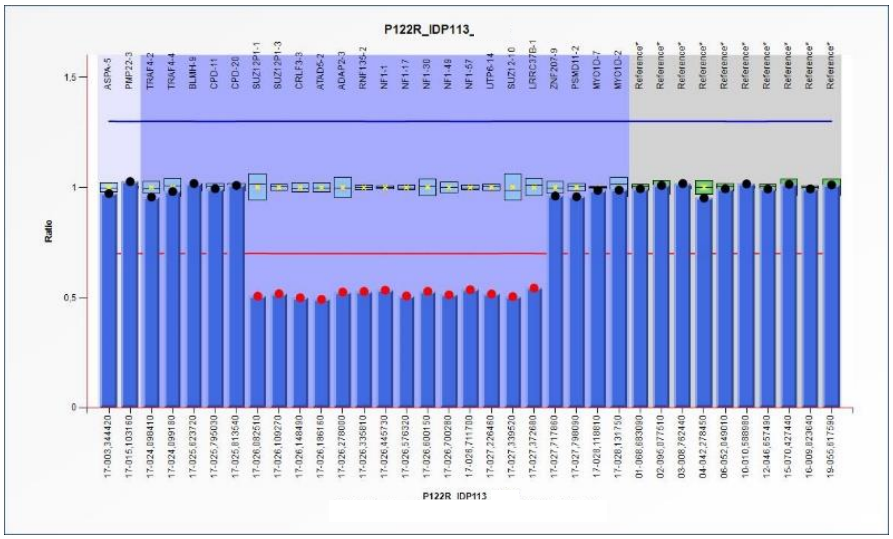


Figure 3.7. Type 1 deletion shown in ratio charts generated by Coffalyser.Net™ software for IDP113 for the a) P081; b) P082 and c) P122 probemix reactions. Target regions with a final ratio of ~0.5 indicated the presence of only one copy of the regions, i.e., a heterozygous deletion.

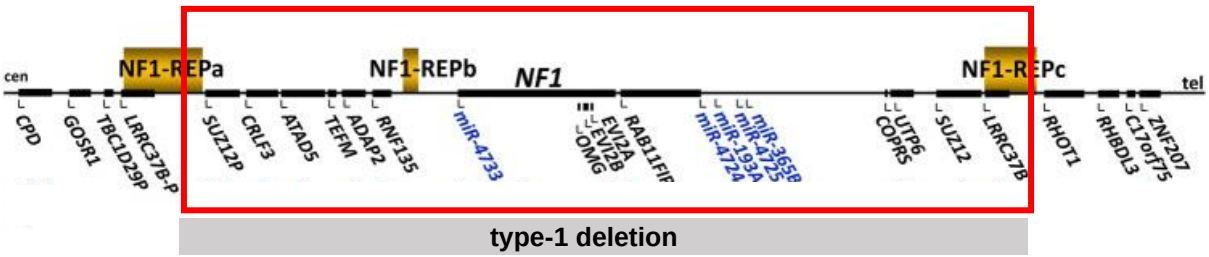


Figure 3.8. Image showing a typical large type 1 deletion (indicated by the red block) encompassing the *NF1* gene and flanking genes. This deletion is approximately 1.4 Mb in size. Adapted from Kehrer-Sawatzki and Cooper, 2021.

ii. IDP89

IDP89 fulfilled the diagnostic criteria for NF1 due to the presence of multiple neurofibromas and café au lait maculae (>6) of >1.5 cm in diameter and axillary freckling. In addition, he also had mild intellectual delay, speech delay and macrocephaly.

A large deletion of the whole *NF1* gene and extending beyond *NF1* was identified in this patient (Figure 3.9). Ratio charts for P081 and P082 show a heterozygous deletion of exons 1 to 58 of *NF1,* as well as a heterozygous deletion of the *OMG* gene located in intron 36 of *NF1.* Furthermore, a heterozygous deletion of the following flanking genes downstream of *NF1,* as shown in the ratio chart for P122, was also present in this patient: *UTP6* exon 14 (515 kb downstream); *SUZ12* exon 10 (628 kb downstream) and *LRRC37B* exon 1 (661 kb downstream). A final ratio of approximately 0.5 for these target regions indicated that there was only one copy of the respective regions present in the patient DNA. A heterozygous deletion of this extent is known as a type 3 deletion (Figure 3.10). These deletions are rare and are typically ~1 Mb in size (Kehrer-Sawatzki and Cooper, 2021). The exact size of the deletion in this patient cannot be determined by MLPA. This result confirms the clinical diagnosis of NF1 in this patient.

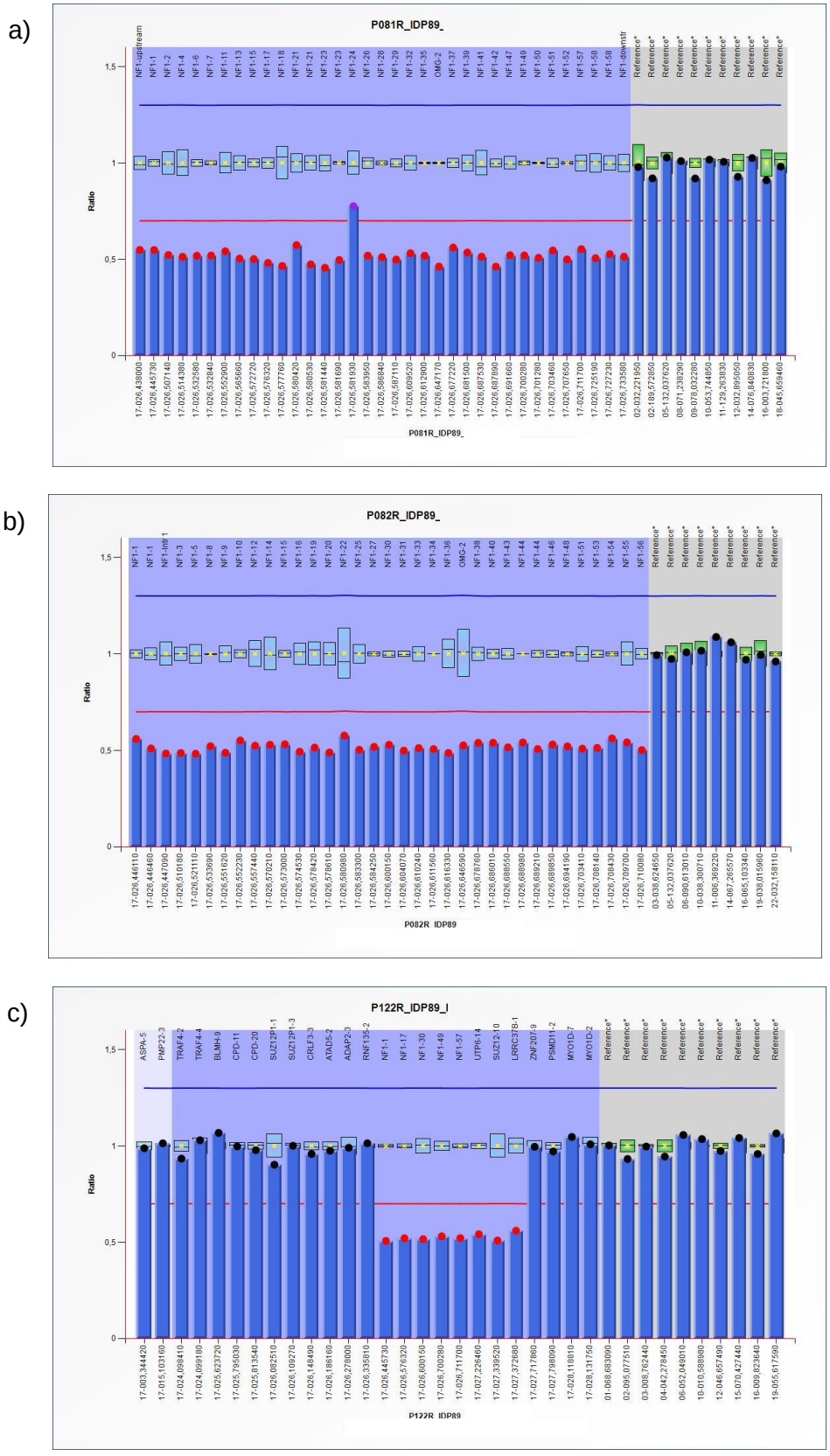


Figure 3.9. Type 3 deletion shown in ratio charts generated by Coffalyser.Net™ software for IDP89 for the a) P081; b) P082 and c) P122 probemix reactions. Target regions with a final ratio of ~0.5 indicated the presence of only one copy of the regions, i.e., a heterozygous deletion.

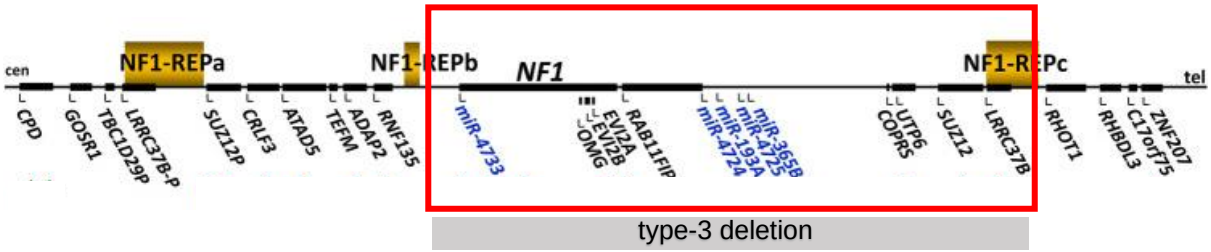


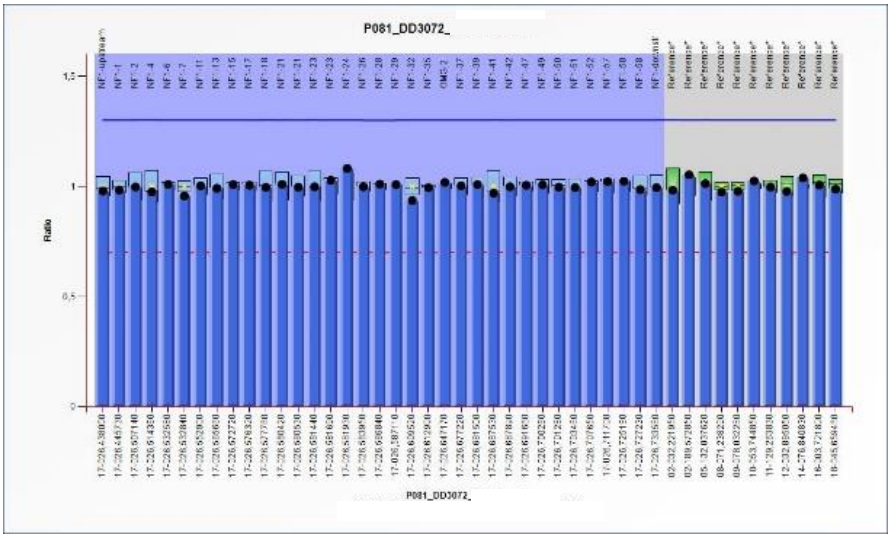
Figure 3.10. Image showing a typical type 3 deletion (indicated by the red block) encompassing the *NF1* gene and flanking genes. This deletion is approximately 1 Mb in size. Adapted from Kehrer-Sawatzki and Cooper, 2021.

iii. DD3072

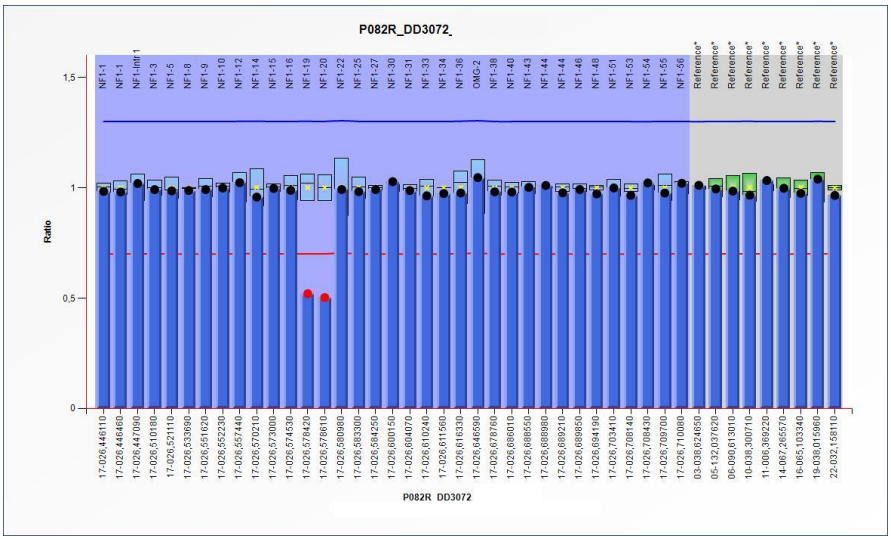
DD3072 met the clinical diagnostic criteria for NF1 due to the presence of multiple café au lait maculae, axillary and inguinal freckling, facial and occipital neurofibromas and iris Lisch nodules. Furthermore, he also presented with learning difficulties, macrocephaly and chronic headaches.

A heterozygous multi-exon deletion consisting of exon 19 and 20 of *NF1* was identified in the patient. This two exon deletion is shown in the ratio chart for P081 (Figure 3.11). A final ratio of approximately 0.5 for the probes targeting exon 19 and 20 indicated that there was only one copy of the respective exons present in the patient DNA. The target regions in the P081 and P122 ratio charts had a final ratio of ~1 which indicated the expected two copies for the other exons of *NF1,* as well as for the genes surrounding *NF1*. Deletions of multiple exons (more than one) have been reported in the literature to be associated with NF1 (Imbard et al., 2015) . The deletion of exon 19 and 20 of the *NF1* gene has not been reported previously. This result confirms the clinical diagnosis of NF1 in this patient.

a)



b)



c)

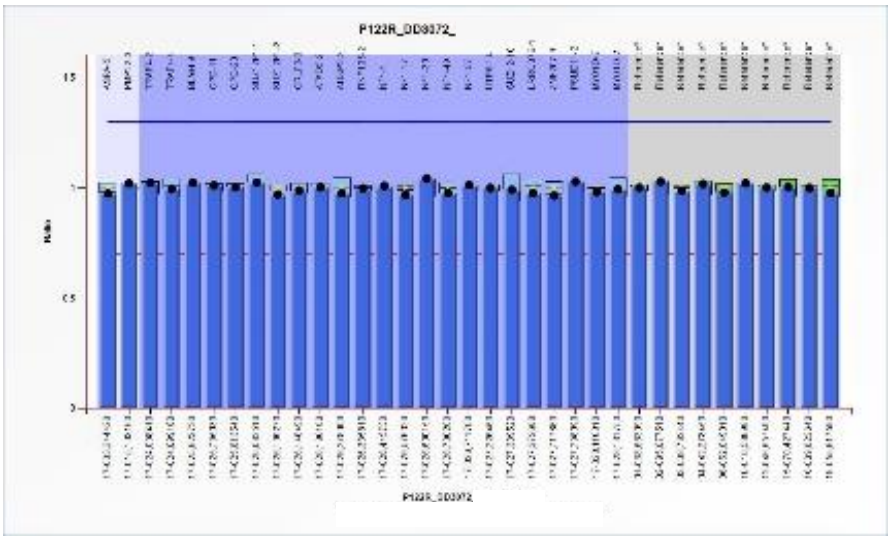


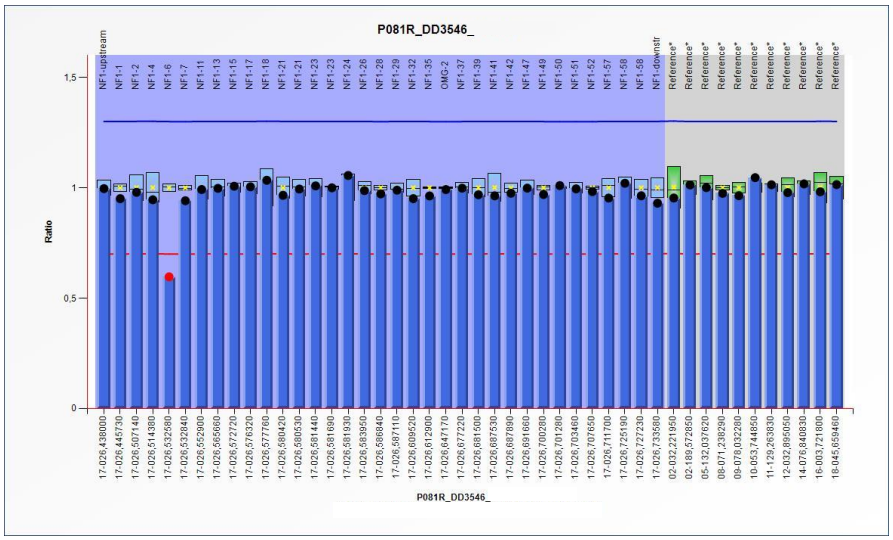
Figure 3.11. Two exon deletion in *NF1* shown in ratio charts generated by Coffalyser.Net™ software for DD3072 for the a) P081; b) P082 and c) P122 probemix reactions. Exon 19 and 20 had final ratios of ~0.5 which indicated the presence of only one copy of the exons as seen in b).

iv. DD3546

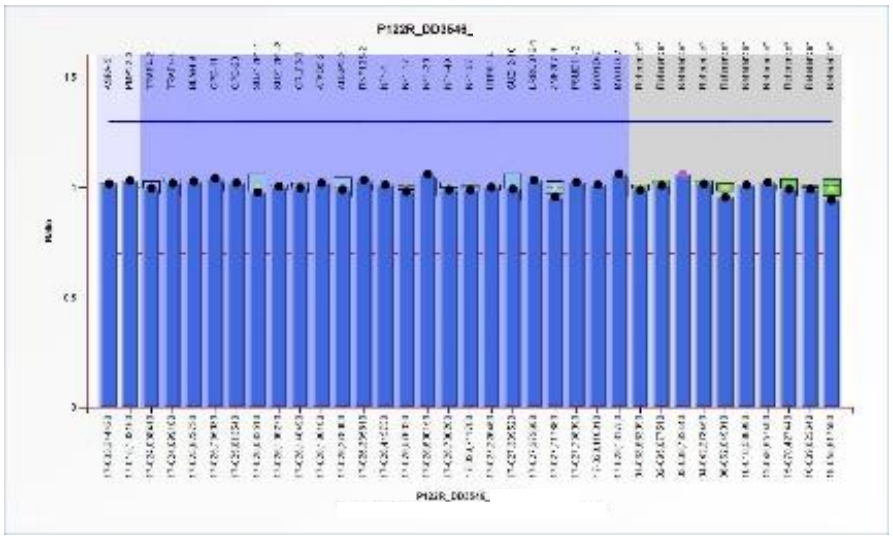
DD3546 met the clinical diagnostic criteria for NF1 due to the presence of multiple café au lait maculae (>6) of >5mm in size and iris Lisch nodules. Furthermore, the patient also presented with macrocephaly, short stature, joint hyperlaxity and asymmetry of the orbital area.

A heterozygous deletion of a single exon (exon 6) in the *NF1* gene was identified in the patient. This single exon deletion is shown in the ratio chart for P081 (Figure 3.12). A final ratio of approximately 0.5 for the probe targeting exon 6 indicated that there was only one copy of the respective exon present in the patient DNA. The target regions in the P081 and P122 ratio charts had a final ratio of ~1 which indicated the expected two copies for the other exons of *NF1,* as well as for the genes surrounding *NF1*. Single exon deletions have been reported in literature to be associated with NF1 (Imbard et al., 2015), however a heterozygous deletion of exon 6 has not been reported previously. Single probe deletions must be confirmed as it could be a false positive result caused by a SNP in the target sequence of the probe.

a)



b)



c)

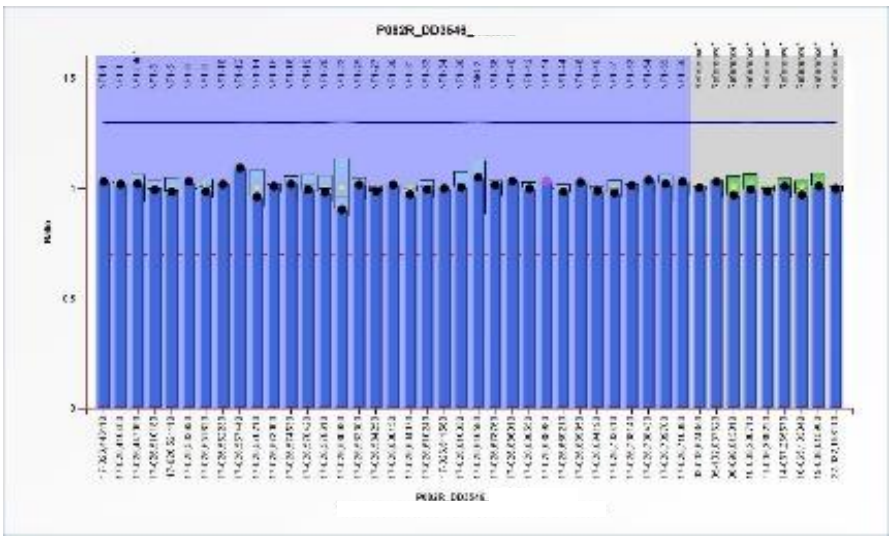


Figure 3.12. Single exon deletion in *NF1* shown in ratio charts generated by Coffalyser.Net™ software for DD3546 for the a) P081; b) P082 and c) P122 probemix reactions. Exon 6 had a final ratio of ~0.5 which indicated the presence of only one copy of the exon as seen in b).

Sanger sequencing was used to confirm the presence of the single probe deletion in DD3546. However, Sanger sequencing identified a heterozygous variant in the patient located in the 3' region of the target sequence of the LPO (Figure 3.13). This variant affected the hybridization of the probe to the target DNA, thereby resulting in reduced amplification of the probe for this region, which was interpreted incorrectly as a heterozygous deletion of exon 6. This result therefore excludes the presence of a heterozygous deletion of exon 6 in DD3546.

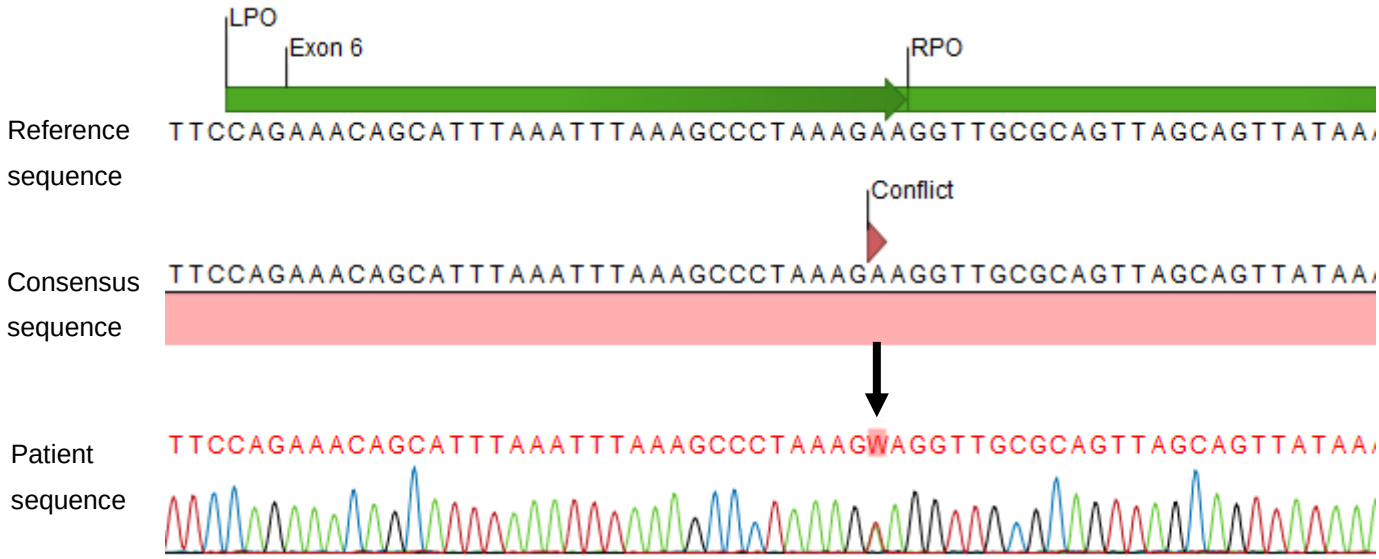


Figure 3.13. Sanger sequencing electropherogram for DD3546 showing the presence of a heterozygous variant (c.616A>T) (indicated with an arrow) located in the 3' region of the LPO targeting exon 6 in the P081 probemix.

The identified variant was a heterozygous nonsense variant, (NM_000267.3): c.616A>T (p.Lys206Ter), located in exon 6 of *NF1* (Table 3.3). Further investigation of this variant showed that it was a pathogenic variant as classified by the ACMG-AMP guidelines.

Table 3.3. Variant identified in DD3546 using Sanger sequencing.

Gene	Variant	Type	Associated disease	Inheritance	ACMG codes applied	ACMG classification
NF1	c.616A>T (p.Lys206Ter)	Heterozygous nonsense	Neurofibromatosis type 1	AD	PVS1, PM2, PP3	Pathogenic

The c.616A>T variant changes the Lysine amino acid at position 206 to a stop codon. This results in premature termination of translation and a truncated protein. Since this variant is located in exon 6 of 58 exons of the *NF1* gene, premature termination of translation will adversely affect the protein. It is expected that the protein will either be absent due to nonsense-mediated decay of the transcripts, or the protein will be significantly disrupted since only a small part will be translated. Mutalyzer 2.0.35 software was used to predict the effect of this nonsense variant on the NF1 protein (Figure 3.14).

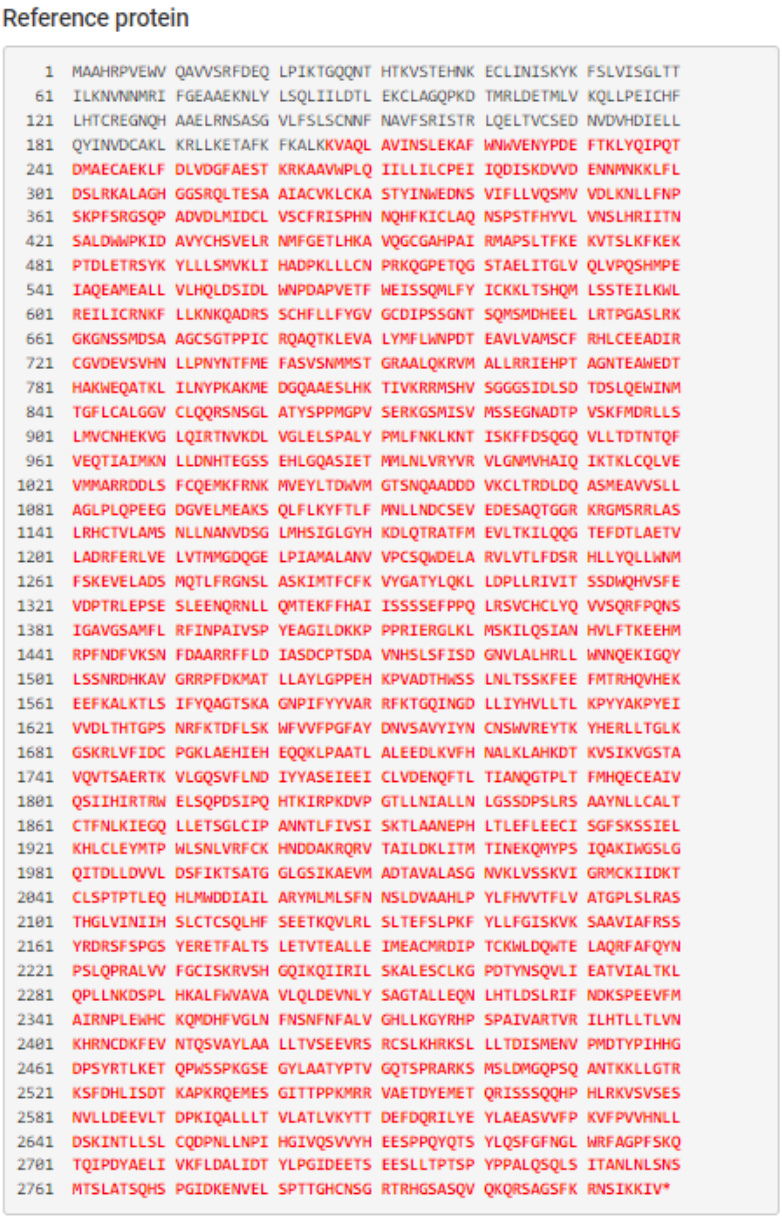


Figure 3.14. Image generated by Mutalyzer 2.0.35 software showing the effect of the c.616A>T (p.Lys206Ter) variant in *NF1*. The variant introduces a stop codon resulting in the premature termination of translation and a significantly truncated NF1 protein. This is indicated by the amino acids shown in red that are not translated.

Nonsense variants have been previously reported to cause NF1, mostly resulting in haploinsufficiency of *NF1* (Castellanos et al., 2020). *NF1* is intolerant to loss of function (LoF) variants as reflected in the probability of LoF intolerance (pLI) score (pLI = 0.9; o/e ratio = 0.22) and the ClinGen haploinsufficiency (HI) score of 3. Therefore, the ACMG code PVS1 could be assigned. The variant was not present in population databases such as gnomAD, ExAc and 1000 Genomes Project (ACMG code PM2).

Furthermore, this variant has not been reported in ClinVar or the literature and is therefore a novel variant. The pathogenicity of this variant was further supported by multiple *in silico* tools that predicted a deleterious effect of this variant. These tools included: DANN, MutationTaster, CADD and FATHMM (ACMG code PP3). Based on this evidence, the heterozygous c.616A>T (p.Lys206Ter) variant was classified as pathogenic and thus confirms the diagnosis of NF1 in DD3546.

3.2.3 Diagnostic yield

One patient failed quality control and was not included in the analysis. Of the remaining thirteen patients, MLPA identified heterozygous deletions of varying sizes in *NF1* for three patients thus confirming the clinical diagnosis of NF1. Furthermore, one patient was identified to have a heterozygous disease-causing SNV in *NF1*. No CNVs were identified for the other nine patients. The diagnostic yield of MLPA for the identification of CNVs in patients with a clinical diagnosis of NF1, but in whom no pathogenic variants were identified using a targeted NGS panel was 25.0% (3/12).

3.3 WES

3.3.1 Sequencing data quality

A subset of eight patients were included for WES after selection based on their clinical features suggestive of a RASopathy. Following library preparation, the Agilent 4150 TapeStation System and Agilent High Sensitivity D5000 ScreenTape assay was used to evaluate the quality of each sample prior to templating and sequencing (Figure 3.15). The fragment lengths ranged between 292-312 bp for the eight libraries and this was within the acceptable range of 200-350 bp. All libraries were thus of acceptable quality to proceed with templating and sequencing.

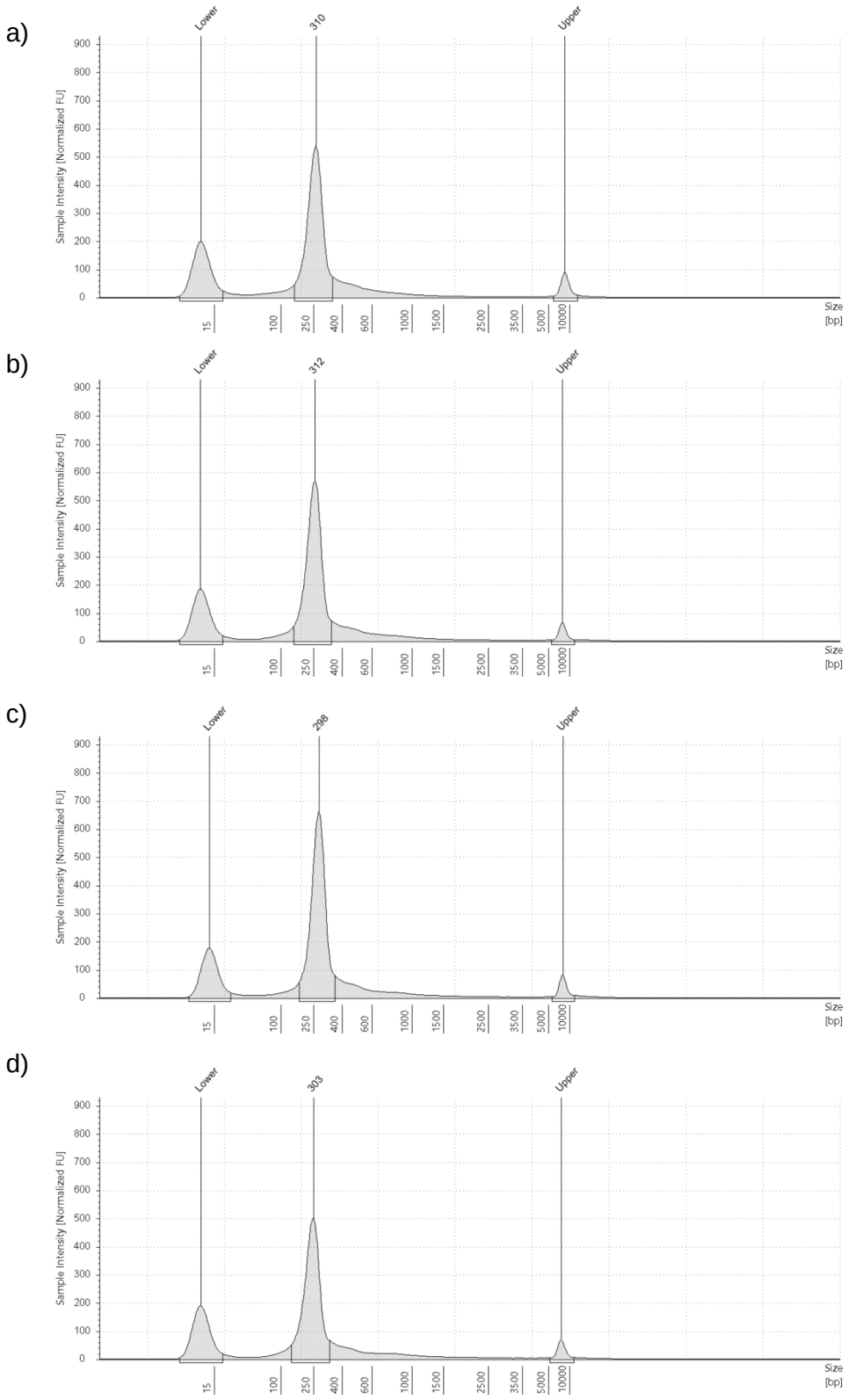
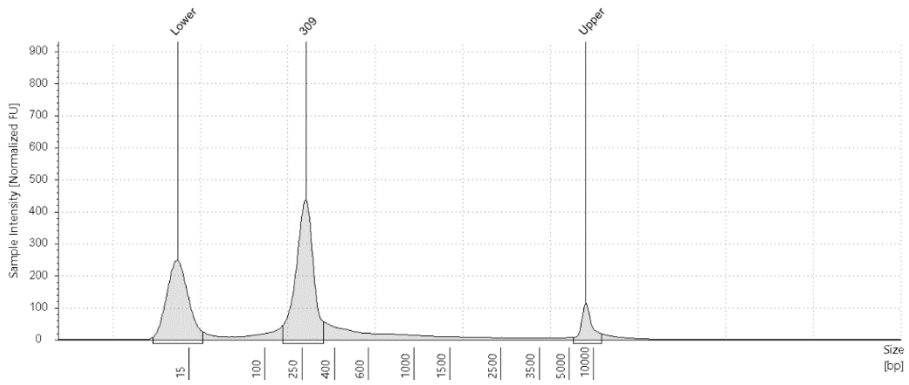
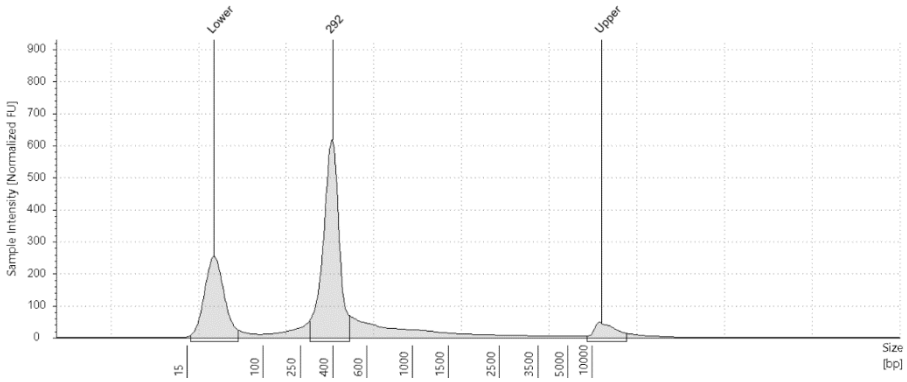


Figure 3.15. Electropherogram images for each sample following library preparation using the High Sensitivity D5000 ScreenTape assay. Each library had an acceptable fragment length for template preparation and sequencing: a) FRASC85; b) FRASC90; c) DD3198 and d) FRASC37.

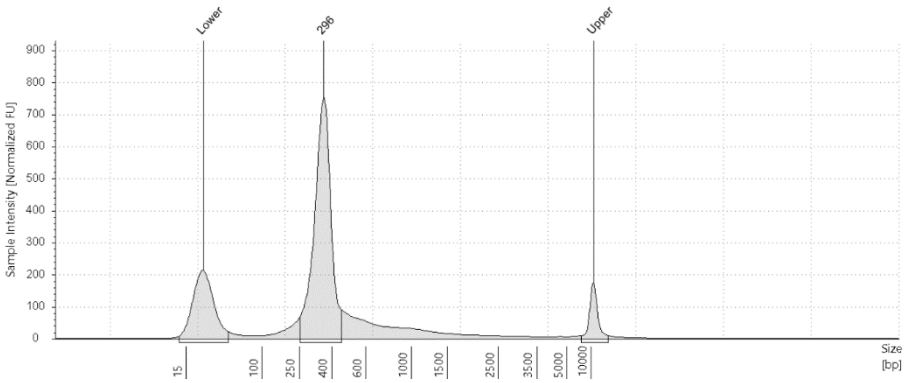
e)



f)



g)



h)

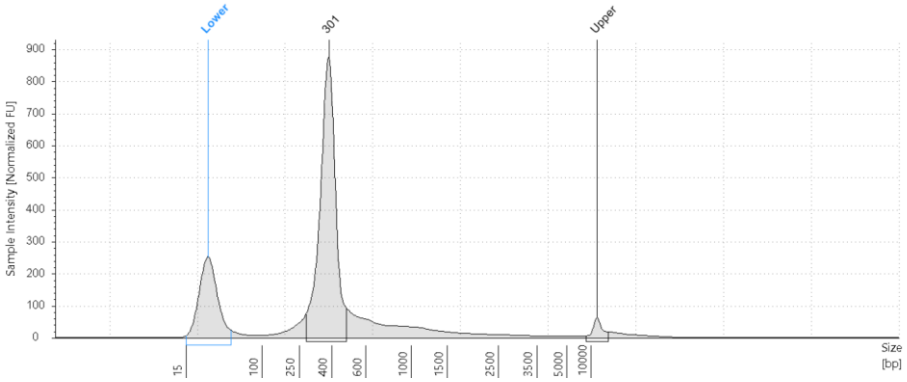


Figure 3.15 cont. Electropherogram images for each sample following library preparation using the High Sensitivity D5000 ScreenTape assay. Each library had an acceptable fragment length for template preparation and sequencing: e) FRASC55; f) IDP17; g) IDP27 and h) DD3425.

Four sequencing runs were performed where two patients were pooled together and sequenced on one Ion 540™ Chip (Table 3.4). The sequencing run reports generated by the Ion Torrent Suite™ software were used to assess the quality of the sequencing data (Figure 3.16). Both the unaligned and aligned quality metrics were evaluated. The run statistics for each of the four runs are summarized in Table 3.4. The total data yields and number of overall reads for all runs were higher than the expected range of 10-15 Gb and 60-80 million, respectively, for an Ion 540™ Chip as specified in the Ion Torrent Suite™ technical manual (Ion Torrent, 2019). Chip loading or ISP loading indicated the percentage of wells in the chip that contained an ISP. Chip loading was of good quality in all four runs, with values of >90%. This statistic was supported by a heat map which visually represented the loading on the chip with red indicating high loading density. The mean read lengths were ~200 bp for all runs as expected for WES using an Ion 540™ Chip.

Reads that passed all the quality filters including polyclonal, low quality and adapter dimer filters are referred to as usable reads. The percentage of usable reads ranged between 65-68% for the four sequencing runs. A polyclonal ISP consists of amplicons from two or more different templates and thus the reads on the ISP are not identical. This will adversely affect data analysis and thus polyclonal ISPs (~26%) were excluded from the usable reads. Low quality reads were observed to be between 9 and 12% and were also removed. No adapter dimers were detected for any of the runs. These metrics indicated overall good quality data.

Table 3.4. Summary of WES run statistics.

Run	Samples per run	Data yield (Gb)	Total number of reads	Chip loading (%)	Final library (%)	Clonal (%)	Polyclonal (%)	Low quality (%)	Adapter dimer (%)	Usable reads (%)
1	FRASC85 FRASC90	17.3	93,093,768	95	90	72	28	9	0	65
2	DD3198 FRASC37	18.2	97,532,286	95	92	74	26	7	0	68
3	FRASC55 IDP17	17.4	95,323,048	94	88	77	23	11	0	67
4	IDP27 DD3425	17.2	94,394,422	94	87	76	24	12	0	67

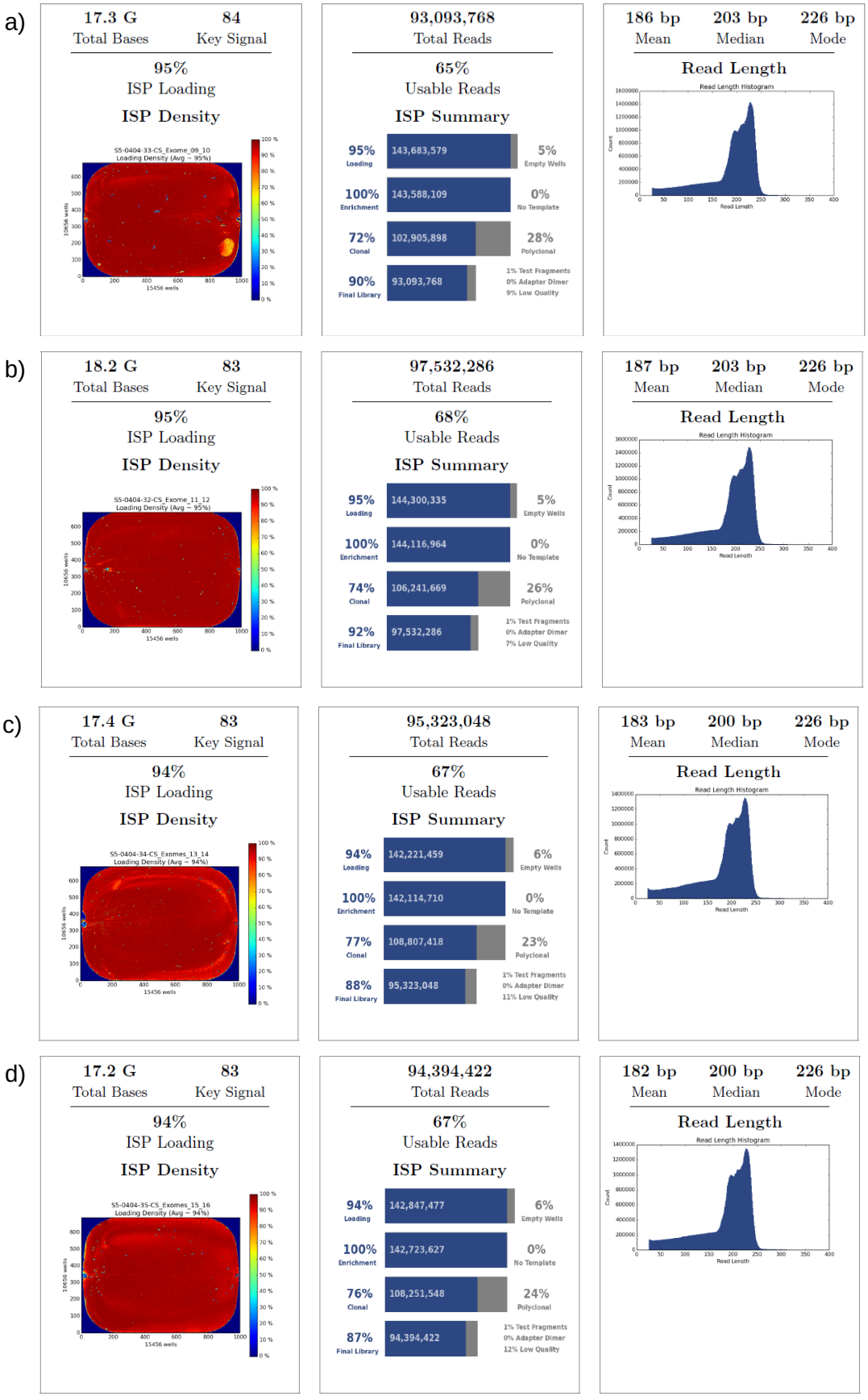


Figure 3.16. Images extracted from the run reports generated by the Ion Torrent Suite™ software indicating sequencing data quality metrics for all four sequencing runs. a) Run 1; b) Run 2; c) Run 3 and d) Run 4. The data obtained for all four runs were of good quality.

Individual sample statistics, as summarized in Table 3.5, were generated by the Coverage Analysis plugin and included in the run reports. The total number of reads that were mapped to the human reference genome varied between 40 and 53 million. The minimum mean sequencing depth of coverage for WES is recommended to be between 75-100x (Rehder et al., 2021). All eight patients that were sequenced had an overall mean depth of coverage between 123-160x with the majority being ~130x (Table 3.5). This is higher than the recommended depth of coverage. The uniformity of coverage was consistent across all samples (>92%) which indicated an equal distribution of reads and good coverage across the exomes, i.e., 92% of the bases had a read depth of at least 20% that of the patient's mean coverage depth.

Table 3.5. Sample specific statistics

Patient	Mapped reads	On Target (%)	Mean depth of coverage (x)	Uniformity of coverage (%)
FRASC85	43,505,923	94.3	131.0	94.0
FRASC90	47,254,867	94.7	141.5	93.6
DD3198	53,300,367	95.0	160.6	94.2
FRASC37	40,951,773	95.1	123.7	93.9
FRASC55	47,603,703	94.9	140.5	93.2
IDP17	45,371,285	94.5	133.3	94.2
IDP27	46,060,369	94.1	135.0	92.6
DD3425	46,123,025	94.4	135.4	92.3

3.3.2 Variant analysis

This section presents the disease-causing and potential disease-causing variants identified in the eight patients that underwent WES. This includes shortlisted variants for those patients in whom no pathogenic disease-causing variants were identified. The different filtering strategies that were utilized were discussed in section 2.6. The candidate variants are reported together with the evidence that was used to classify each variant according to the ACMG-AMP guidelines (Richards et al., 2015). This includes data obtained from population and disease databases, as well as the literature. The disease with which each variant is associated is also listed, as well as the inheritance pattern thereof. Coverage plots from IGV are presented for those

variants classified as pathogenic or likely pathogenic to determine the quality and confidence of these variants based on the coverage depth and allele fractions. Each patient's phenotype is also briefly described.

3.3.2.1 IDP17

IDP17 had a clinical diagnosis of Noonan syndrome based on the presence of hypertrophic obstructive cardiomyopathy, pulmonary stenosis and short stature. Furthermore, the patient also presented with hypotonia, failure to thrive, skin pigmentary abnormalities, dysmorphic features and urogenital abnormalities. There was no family history of Noonan syndrome.

The manual filtering workflow was used for variant analysis for this patient. Prior to filtering, there were 89029 variants identified in this patient. Variants were subsequently filtered in a stepwise manner starting with known RASopathy genes. Variants located in known RASopathy genes were retained, while variants in other genes were filtered out. A total of 94 variants remained after this step. Of the remaining variants, those with a MAF of more than 1% were filtered out. Only 25 variants passed this filter. Synonymous variants that were not predicted to impact splicing, intronic, up and downstream variants were excluded, which left one missense variant. The remaining variants were subsequently assessed based on their clinical significance as reported by ClinVar. Based on the clinical significance, the impact, as well as the *in silico* pathogenicity predictions, only one variant was prioritized for further investigation (Table 3.6).

Table 3.6. Shortlisted variant identified in IDP17.

Gene	Variant	Type	Associated disease	Inheritance	ACMG codes applied	ACMG classification
RIT1	c.297T>G (p.Phe99Leu)	Heterozygous missense	Noonan syndrome	AD	PS1, PM1, PM2, PP2, PP3, PP5	Pathogenic

The prioritized variant that was identified in IDP113 was a heterozygous missense variant in exon 5 of *RIT1* (NM_001256821.2): c.297T>G (p.Phe99Leu). *RIT1* is a known RASopathy associated gene and forms part of the RAS/MAPK pathway. This variant has been previously reported in the literature to be associated with Noonan syndrome (Yaoita et al., 2016) and has been classified as pathogenic by ClinVar (VCV000181522.15) (ACMG code PP5). In addition, two other missense variants resulting in the same amino acid change as our variant have both been reported as pathogenic by ClinVar (ACMG code PS1): c.297T>A (VCV000370035.6) and c.295T>C (VCV000183407.5). This is a gain of function variant affecting a highly conserved amino acid located in the switch II functional domain region (Figure 3.17) (Aoki et al., 2013). The majority of pathogenic variants in this gene are clustered within this domain and 22% of variants occur at the same codon as our variant (Kouz et al., 2016) (ACMG code PM1).

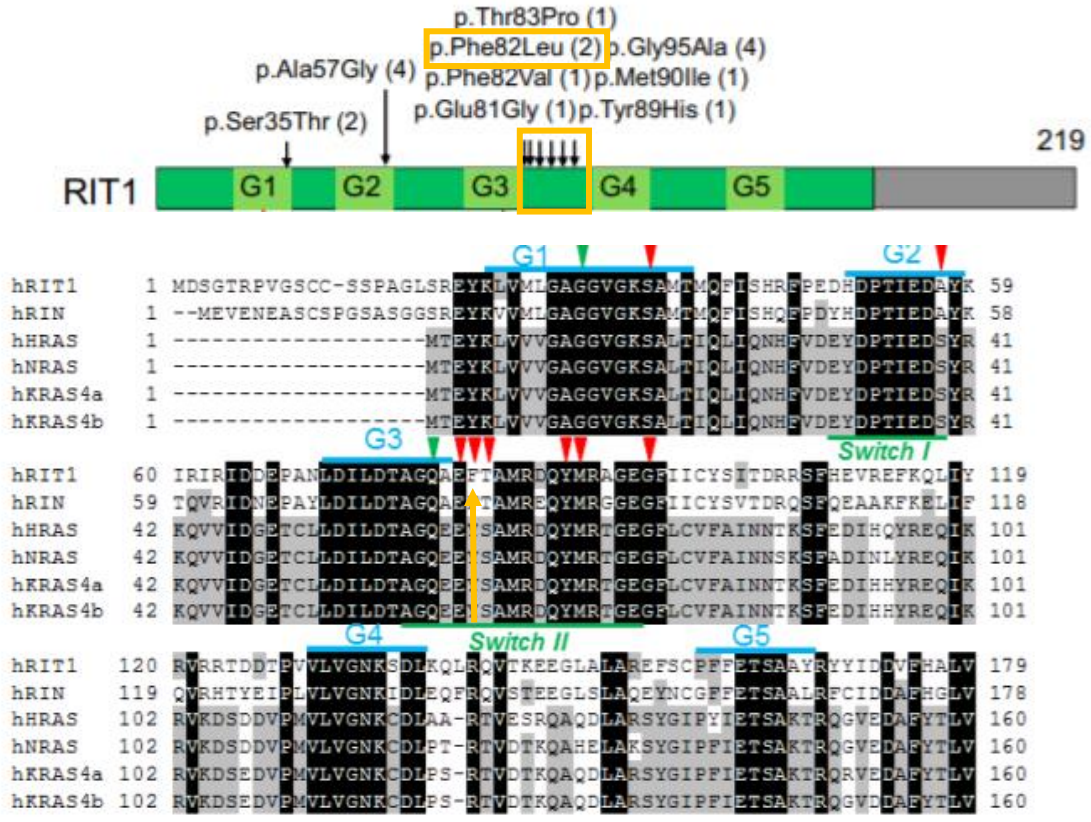


Figure 3.17. Diagram showing the location of the c.297T>G (p.Phe99Leu) variant in the switch II domain of the human *RIT1* gene (hRIT1) (yellow block and arrow). The variant is also known as p.Phe82Leu depending on the transcript. The switch II domain spans from amino acid residues 78–93 with the most mutations clustering around this region (red triangles). Adapted from Aoki et al., 2013.

RIT1 has a low rate of benign missense variation and therefore the PP2 ACMG code was assigned. *RIT1* has a missense z-score of 2.08 (o/e ratio = 0.52). Furthermore, the majority of variants causative of Noonan syndrome are missense variants (Tidyman and Rauen, 2016b). The identified variant was absent from population databases such as gnomAD, ExAc and 1000 Genomes Project (ACMG code PM2). Multiple *in silico* tools (REVEL, CADD, DANN, FATHMM, PolyPhen2, SIFT, MutationTaster, MutationAssesor and PROVEAN) predicted a deleterious effect of the variant which further supports the pathogenicity of this variant (ACMG code PP3).

The variant was visualized in IGV which revealed a depth of coverage of 167x for the variant (Figure 3.18). Furthermore, the variant was validated with Sanger sequencing as depicted in Figure 3.19. The *RIT1* c.297T>G missense variant has ample evidence to support the classification of the variant as a pathogenic disease-causing variant. This result provides molecular confirmation for the patient's clinical diagnosis of Noonan syndrome.

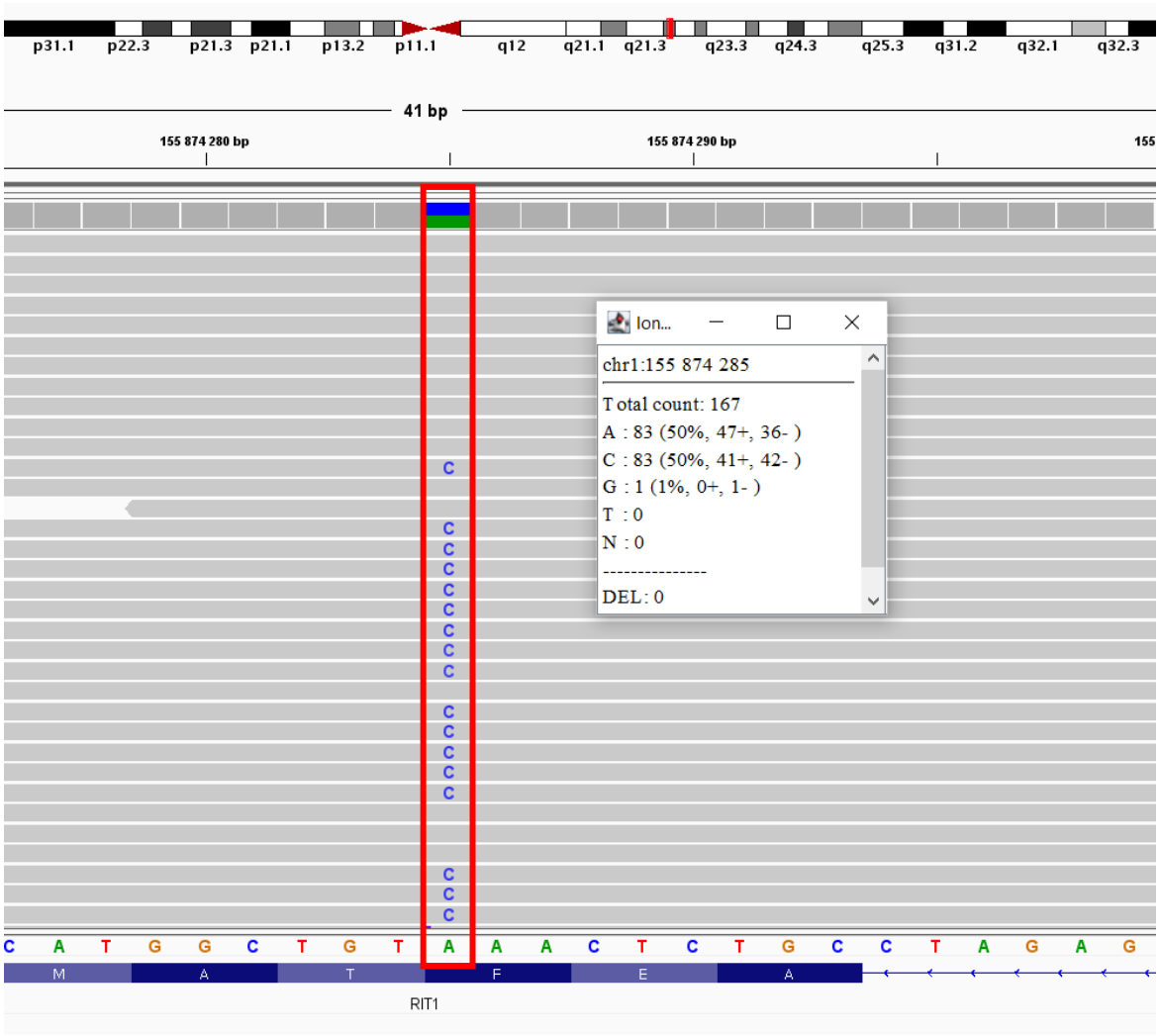


Figure 3.18. IGV image for IDP17 showing the c.297T>G variant in R*IT1* presented as a A>C change (indicated with a red block). The depth of coverage for the variant is 167x and there was no strand bias seen in the equal allele counts and ratios (0.5).

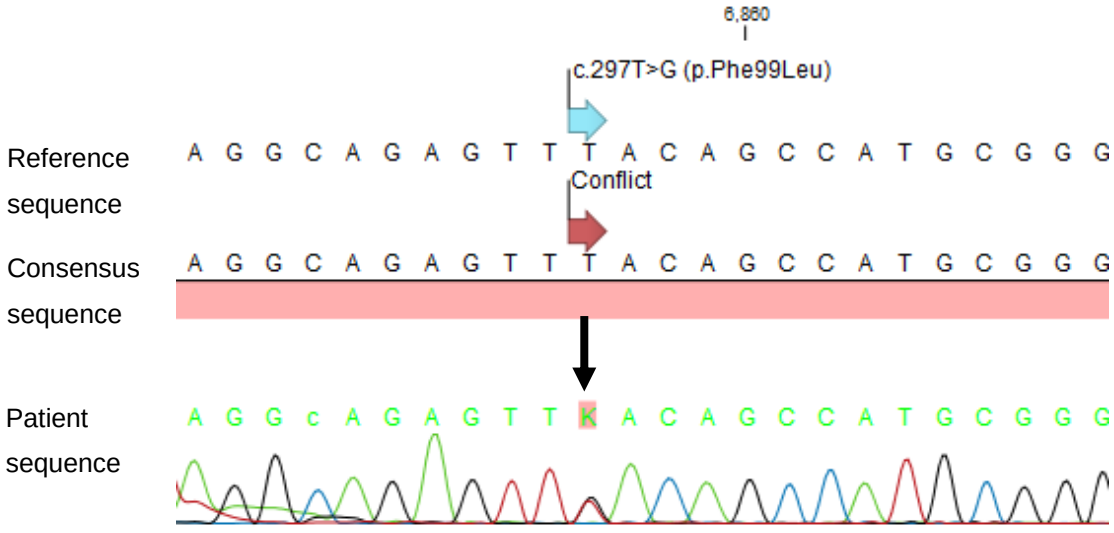


Figure 3.19. Sanger sequencing validation for *RIT1* c.297T>G. Sequence electropherogram showing the presence of the heterozygous variant c.297T>G variant (indicated by an arrow) in *RIT1* for IDP17.

3.3.2.2 DD3198

DD3198 was suspected to have a RASopathy and presented with the following clinical features: atrial septal defect and ventricular septal defect, macrocephaly, developmental delay, cryptorchidism, hypoplastic nails and dysmorphic features including a broad chest, short palpebral fissures, hooded eyelids and cup shaped ears.

A total of 90209 variants were identified in this patient through annotation using Ensembl VEP. The automated filtering workflow using Moon software returned a shortlist of 20 variants ranked according to likeliness of causing disease in our patient. Based on the data quality and other information returned in the Moon report, one variant was prioritized and investigated further (Table 3.7).

Table 3.7. Shortlisted variant identified in DD3198.

Gene	**Variant**	**Type**	**Associated disease**	**Inheritance**	**ACMG codes applied**	**ACMG classification**
SON	c.5753_5756del, (p.Val1918GlufsTer87)	Heterozygous frameshift	ZTTK syndrome	AD	PVS1, PM2, PP5	Pathogenic

A heterozygous variant in exon 3 of *SON* (NM_138927.4): c.5753_5756del (p.Val1918GlufsTer87) was identified in this patient. This variant is a four base pair deletion which causes a shift in the codon reading frame, i.e., frameshift, thereby resulting in a premature termination codon. Transcripts with this premature stop codon undergo degradation through nonsense-mediated decay resulting in reduced messenger RNA (mRNA) levels. This variant thus results in haploinsufficiency of the SON protein (J. H. Kim et al., 2016). Mutalyzer 2.0.35 software was used to predict the effect of this deletion on the SON protein with the amino acids indicated in red not being translated (Figure 3.20). The *SON* gene is extremely intolerant to LoF variants based on the pLI score (pLI =1; o/e ratio = 0.05) and the ClinGen HI score of 3. LoF is also a known mechanism of disease for ZTTK syndrome and thus the PVS1 ACMG code was assigned.

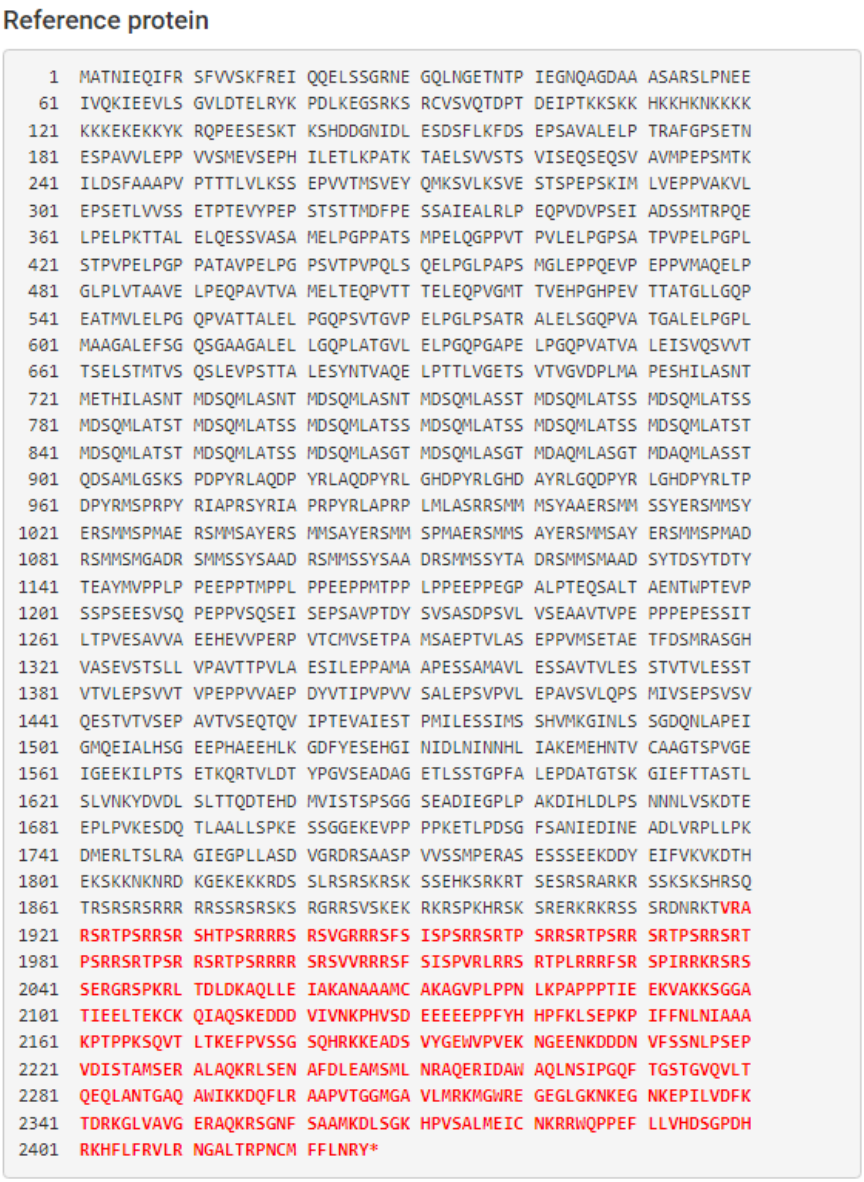


Figure 3.20. Image generated by Mutalyzer 2.0.35 software showing the effect of the c.5753_5756del frameshift variant in *SON*. The variant introduces a stop codon downstream of the deletion resulting in premature termination of translation and a truncated protein. The amino acids shown in red are not translated.

This deletion was previously reported in the literature to be associated with Zhu-Tokita-Takenouchi-Kim (ZTTK) syndrome and is a common recurring variant (Kushary et al., 2021). It has been classified as pathogenic in ClinVar (VCV000252929.18) (ACMG code PP5). ZTTK syndrome is a rare developmental disorder with a broad phenotypic spectrum with some features overlapping with that of the RASopathies. Pathogenic disease-causing variants in *SON* usually occur *de novo.* The variant was not present in population databases such as gnomAD, ExAc and 1000 Genomes Project (ACMG code PM2). The deletion was visualized in IGV (Figure 3.21) which revealed a depth

of coverage of 124x for the variant. The variant was validated using Sanger sequencing as depicted in Figure 3.22.

The c.5753_5756del variant present in *SON* has enough evidence to support the classification thereof as pathogenic. The disease-causing variant in the patient was identified, however this changes the clinical diagnosis of the patient from a RASopathy to ZTKK syndrome.

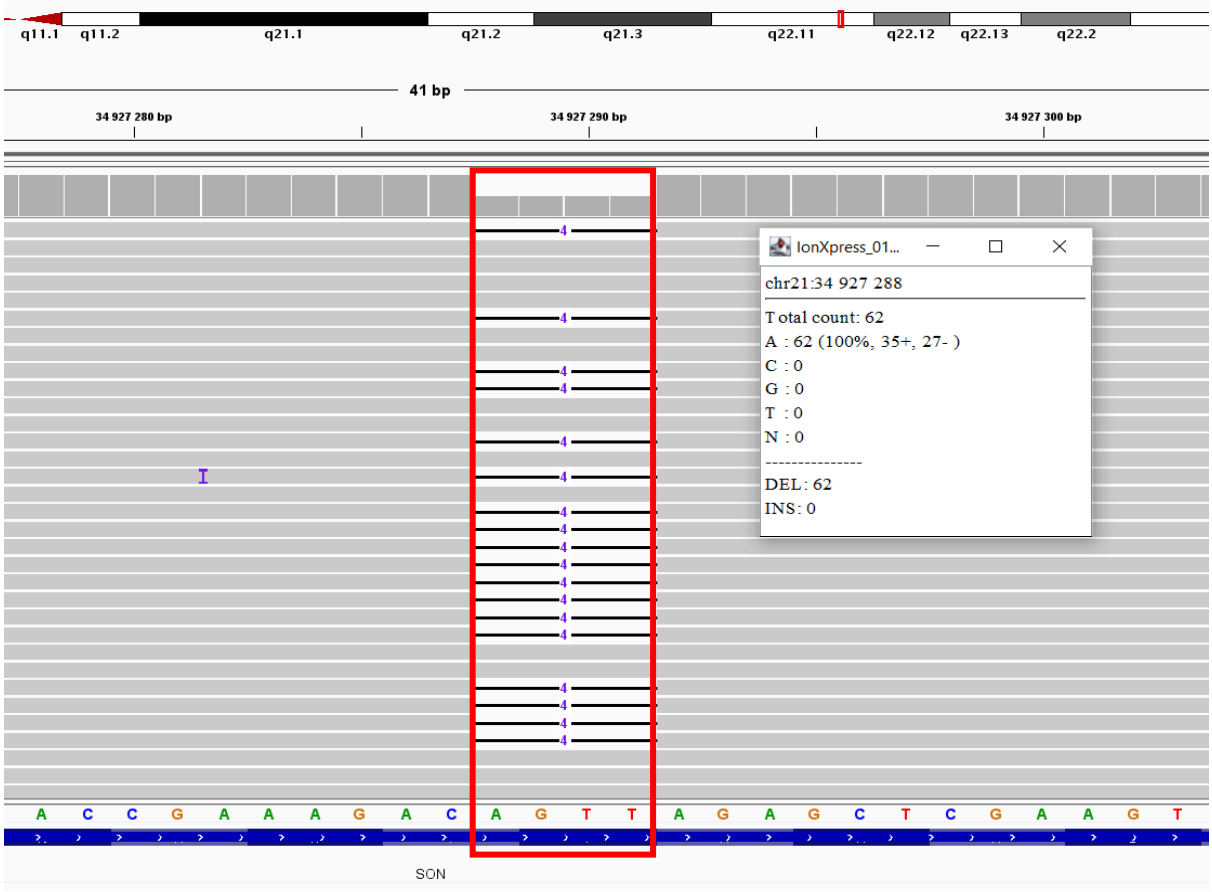


Figure 3.21. IGV image for DD3198 showing the c.5753_5756del variant in *SON* (indicated with a red block). The depth of coverage for the variant is 124x and there was no strand bias seen in the equal allele counts and ratios (0.5).

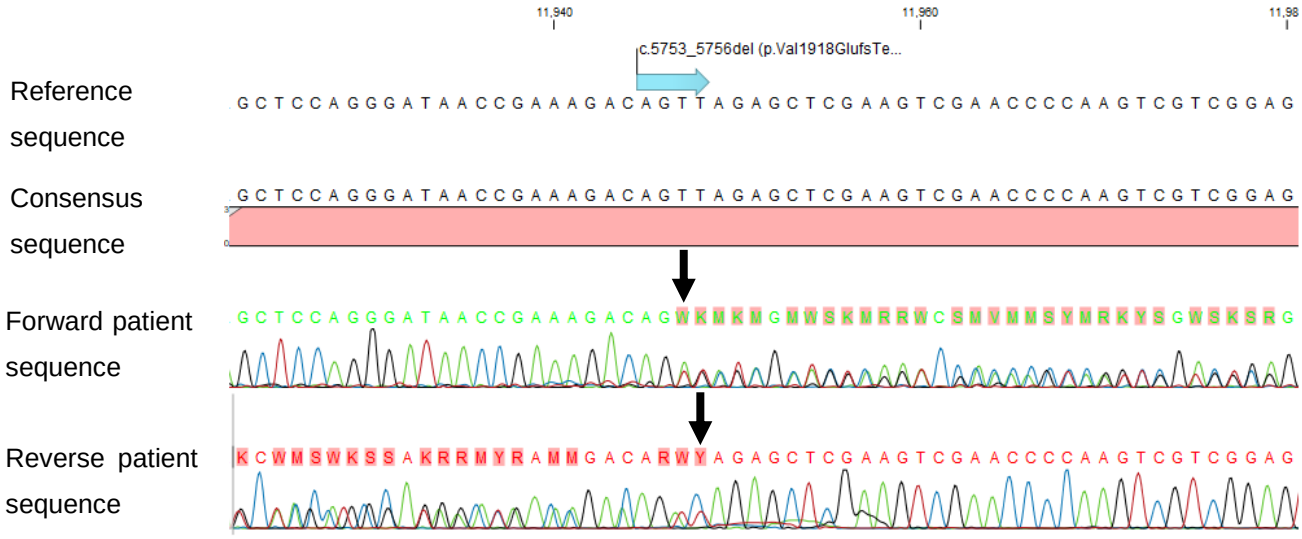


Figure 3.22. Sanger sequencing electropherogram showing the heterozygous *SON* c.5753_5756del variant for DD3198. The start position of the deletion is indicated by an arrow and differs depending on the direction of sequencing.

3.3.2.3 DD3425

DD3425 was clinically diagnosed with Noonan syndrome. The patient presented with short stature, macrocephaly, hypotonia, developmental delay and dysmorphic features including hypertelorism, epicanthic folds, a flat facial profile, low set posteriorly rotated ears and a short neck. Furthermore, he had prominent surface sulci and ventricles of the brain and minor seizures.

There were 86694 variants called for this patient prior to variant filtering. One candidate variant in *MTHFR* (NM_005957.5): c.760C>T (p.Pro254Ser) (Table 3.8) was identified by the manual filtering strategy. The c.760C>T variant has an ACMG classification of likely pathogenic, however, it is associated with an autosomal recessive disorder. A second pathogenic SNV in *MTHFR* could not be identified and thus this variant was viewed as an unlikely cause of disease in DD3425. CNVs could not be ruled out with this approach. The automated filtering strategy using Moon was utilized and a shortlist of 41 variants was returned. Based on the data quality and other information returned in the Moon report, one variant was prioritized for further investigation (Table 3.8).

Table 3.8. Shortlisted variants identified in DD3425.

Gene	Variant	Type	Associated disease	Inheritance	ACMG codes applied	ACMG classification
MTHFR	c.760C>T (p.Pro254Ser)	Heterozygous missense	Homocystinuria	AR	PM1, PM2, PP3, PP5	Likely pathogenic
CDK8	c.95A>T (p.Tyr32Phe)	Heterozygous missense	Developmental disorder	AD	PM1, PM2, PP2, PP3	Likely pathogenic

The identified prioritized variant in DD3425 was a heterozygous missense variant located in exon 1 of *CDK8* (NM_001260.3): c.95A>T (p.Tyr32Phe). This variant has not been reported in literature or in databases such as ClinVar and this suggests that this variant is novel. The variant was also not present in population databases such as gnomAD, ExAc and 1000 Genomes Project (ACMG code PM2).

Other variants in *CDK8* have been associated with a recently identified rare developmental disorder (Calpena et al., 2019). *CDK8* is intolerant of missense changes with a missense z-score of 3.68 (o/e = 0.35). There is a low rate of benign missense variation reported in this gene and all reported pathogenic variants associated with the disorder were missense variants (Calpena et al., 2019) (ACMG code PP2). The variant is present in the highly conserved Gly-rich loop (amino acid 27-35) of the protein kinase domain (amino acid 21–335) of CDK8 (Figure 3.23) where multiple other pathogenic variants cluster. The domain has a critical role in ATP binding and phosphoryl transfer reactions. ACMG code PM1 was thus assigned. Multiple *in silico* tools (CADD, DANN, PolyPhen2, SIFT, MutationTaster and PROVEAN) predicted the variant to be damaging which supports the pathogenicity thereof (ACMG code PP3).

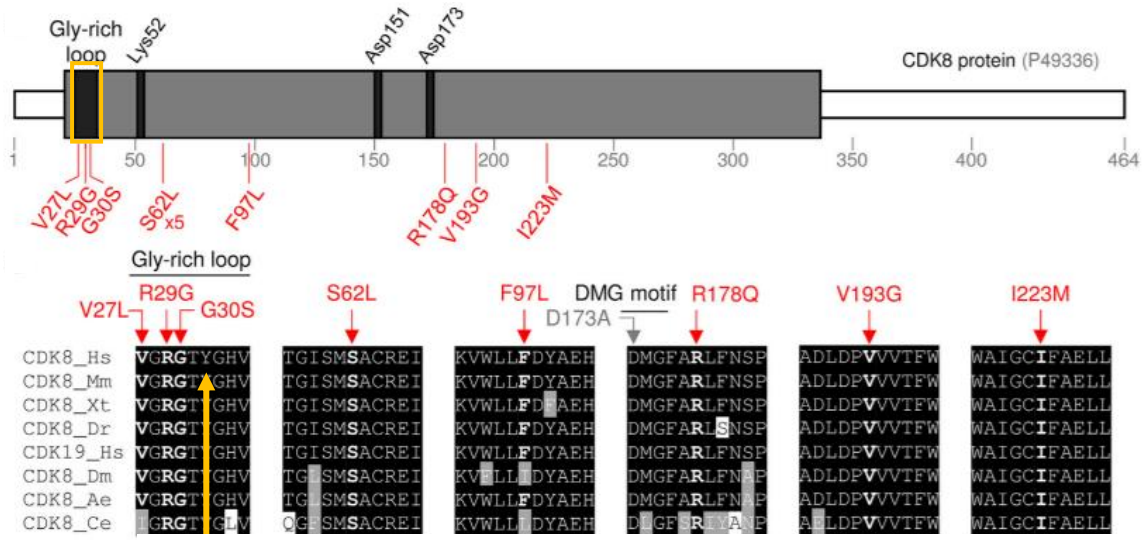


Figure 3.23. Image showing the position of the novel c.95A>T (p.Tyr32Phe) variant in the Gly-rich loop of the protein kinase domain (grey box) of human C*DK8* (CDK8_Hs) (indicated with a yellow box and arrow). Adapted from Calpena et al., 2019.

The variant had a read depth of 206x as seen in Figure 3.24 obtained from IGV. Furthermore, the variant was validated using Sanger sequencing as shown in Figure 3.25. The novel c.95A>T (p.Tyr32Phe) variant located in *CDK8* has sufficient evidence to support its classification as a likely pathogenic variant. Although this variant does not support the initial clinical diagnosis of the patient, it is consistent with the patient's clinical features including developmental delay, dysmorphic facial features, hypotonia and brain abnormalities (Calpena et al., 2019). The variant can thus be regarded as the putative disease-causing variant contributing to the phenotype of DD3425.

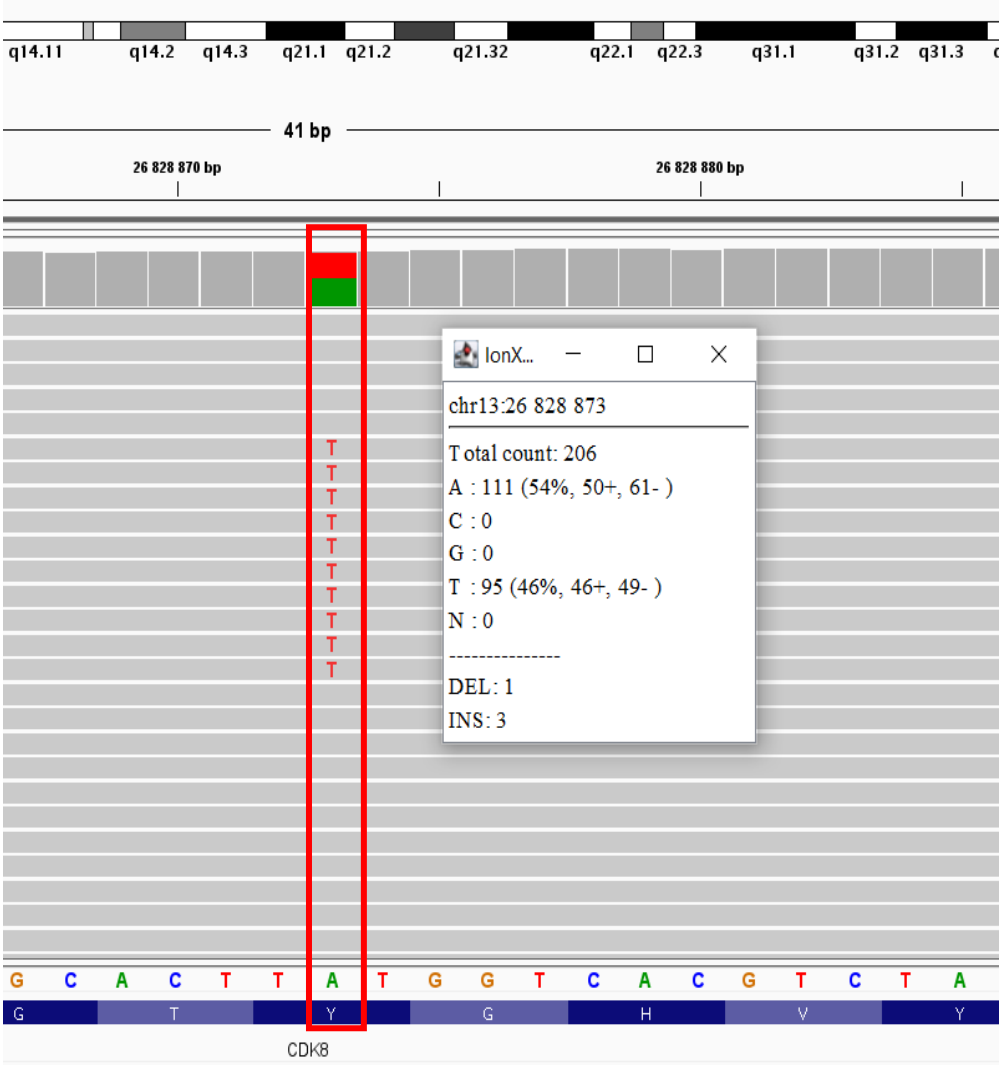


Figure 3.24. IGV image for DD3425 showing the c.95A>T variant in *CDK8* (indicated with a red block). The depth of coverage for the variant is 206x and the variants were present in nearly equal ratios (0.54 and 0.46).

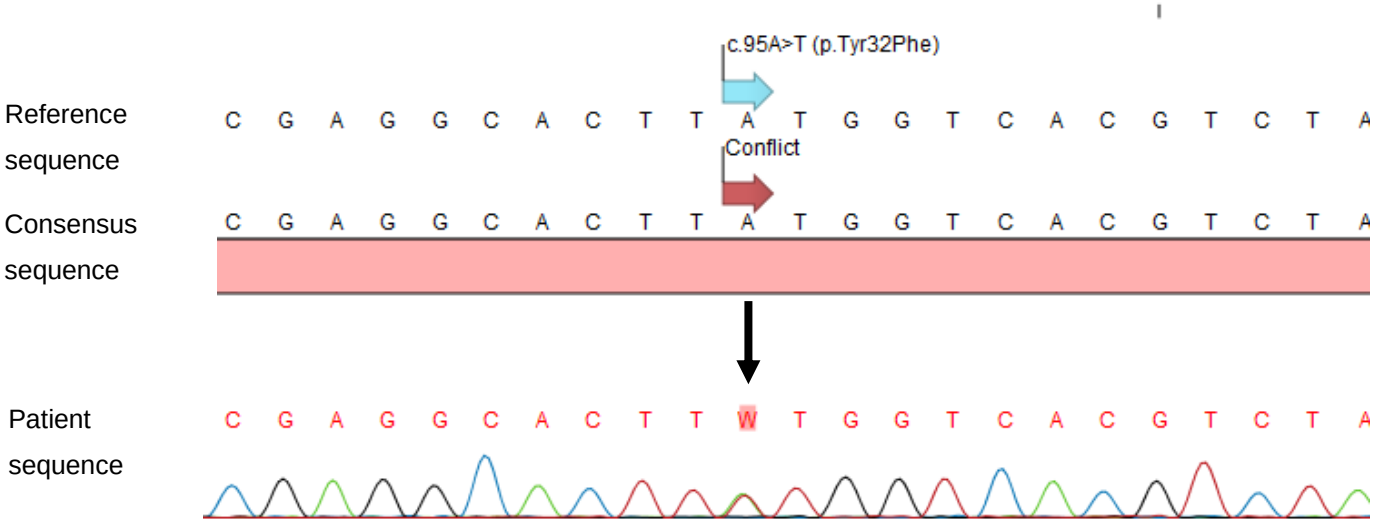


Figure 3.25. Sequencing electropherogram showing the presence of the heterozygous *CKD8* c.95A>T variant (indicated by an arrow) for DD3425.

3.3.2.4 FRASC85

FRAC85 presented with short stature, developmental delay, intellectual disability, a shawl scrotum and dysmorphic features including ptosis, a broad forehead, sparse lateral eyebrows and coarse facial features suggestive of Noonan syndrome. In addition, he also had immotile cilia, a hoarse voice, hepatomegaly and attention deficit hyperactivity disorder.

Ensembl VEP annotated 70954 variants in this patient. Four variants were prioritized for further investigation of which two (*CLCN4* c.1837A>G and *KIF1A* c.4131C>G) were identified by the automated filtering strategy utilizing Moon and two (*NAGA* c.479C>G and c.973G>A) were identified using the VEP-DDG2P plugin (Table 3.9).

Table 3.9. Shortlisted variants of uncertain significance identified in FRASC85.

Gene	**Variant**	**Type**	**Associated disease**	**Inheritance**	**ACMG codes applied**	**ACMG classification**
CLCN4	c.1837A>G (p.Thr613Ala)	Hemizygous missense	CLCN4-related neurodevelopmental disorder	X-linked recessive	PM2, PP2 PP3	VUS favouring likely pathogenic
KIF1A	c.4131C>G (p.Asn1377Lys)	Heterozygous missense	Syndromic intellectual disability	AD	PM2, PP3	VUS
NAGA	c.479C>G (p.Ser160Cys)	Heterozygous missense	Schindler disease	AR	PM2, PP3	VUS
NAGA	c.973G>A (p.Glu325Lys)	Heterozygous missense	Schindler disease	AR	PM2, PP3, PP5	VUS

The c.1837A>G (p.Thr613Ala) hemizygous missense variant located in exon 11 of *CLCN4* (NM_001830.4) was classified as a VUS favouring likely pathogenic. This variant was absent in population databases (ACMG code PM2) and has also not been reported previously in literature or in ClinVar. The *CLCN4* gene is intolerant to missense variants with a z-score of 4.52 (o/e = 0.32) and it has been reported that missense variants cause a more severe form of disease (He et al., 2021) (ACMG code

PP2). Furthermore, this variant position is highly conserved. Multiple *in silico* prediction tools predicted a damaging effect of this variant on the gene product (ACMG code PP3). The variant is located within the CBS1 functional domain where other pathogenic variants have been reported previously, however there is also benign variation within this domain and therefore PM1 could not be applied. The clinical features of the disorder have some overlap with our patient's phenotype including intellectual disability, developmental delay, behavioural problems such as hyperactivity and varying dysmorphic facial features. In addition, patients have also been reported to present with seizures, hypotonia and structural brain abnormalities (He et al., 2021). The variant could, however, not be classified as a disease-causing variant due to insufficient evidence.

The second shortlisted variant was a missense variant identified in exon 40 of *KIF1A* (NM_001244008.2): c.4131C>G (p.Asn1377Lys) and classified as a VUS (ACMG codes PM2 and PP3) owing to insufficient evidence. Furthermore, the phenotype was not consistent with that observed in our patient.

The two heterozygous missense variants in *NAGA* (NM_000262.3): c.479C>G (p. Ser160Cys) and c.973G>A (p.Glu325Lys), identified by the VEP-DDG2P plugin were also classified as VUS due to insufficient evidence. Pathogenic variants in this gene are associated with an autosomal recessive disorder known as Schindler disease. Schindler disease is a lysosomal storage disorder with extensive clinical heterogeneity of which some features overlap with that of our patient. Clinical features that have been reported include intellectual disability, seizures, facial coarseness, angiokeratoma, motor retardation, behavioural problems, cardiomyopathy and hepatomegaly (Keulemans et al., 1996; Bakker et al., 2001; Chabás, Duque and Gort, 2007). Both identified variants are present in population databases at a low frequency of 0.000744 and 0.002475, respectively, and since it is associated with an autosomal recessive condition, PM2 can be assigned. Both variants are predicted to be deleterious by multiple *in silico* pathogenicity tools (ACMG code PP3). Furthermore, both variants have been previously reported in the literature (Keulemans et al., 1996; Bakker et al., 2001) as well as in ClinVar. The c.479C>G variant has conflicting classifications in ClinVar (VCV000018165.20) with the majority assigning a VUS classification and therefore PP5 cannot be applied. The c.973G>A variant also has

conflicting interpretations (VCV000018162.22), however the majority of submitters agreed on a pathogenic classification (ACMG code PP5). There was not sufficient information available to confidently identify one of these variants as the disease-causing variant in this patient. Further in-depth clinical phenotyping of the patient and future reanalysis of the shortlisted variants are required for this patient.

3.3.2.5 FRASC90

Noonan syndrome was diagnosed in FRASC90 based on the patient's clinical features. The patient presented with moderate intellectual disability, cardiac anomalies, pectus excavatum, joint laxity, café au lait maculae and multiple lentigines. Furthermore, the patient had many dysmorphic features such as low set and posteriorly rotated ears, webbed neck, long face and hypertelorism.

There were 89193 variants identified in this patient by Ensembl VEP. The automated filtering strategy using Moon returned a shortlist of 22 variants of which only one candidate variant was prioritized (Table 3.10).

Table 3.10. Shortlisted variant of uncertain significance identified in FRASC90.

Gene	**Variant**	**Type**	**Associated disease**	**Inheritance**	**ACMG codes applied**	**ACMG classification**
SMARCA2	c.4227T>A (p.Asn1409Lys)	Heterozygous missense	Nicolaides-Baraitser syndrome	AD	PM2, BP4	VUS

The prioritized variant was a heterozygous missense variant in exon 29 of *SMARCA* (NM_003070.5): c.4227T>A (p.Asn1409Lys). Using the ACMG-AMP guidelines, this variant was classified as a VUS. The variant was absent in population databases (ACMG code PM2) and has not been previously reported in databases such as ClinVar. BP4 was assigned since the majority of *in silico* tools predicted the variant to be benign. The variant has been classified as a VUS due to conflicting evidence for

pathogenicity. It is unlikely that this is the disease-causing variant for FRASC90 based on the available information at this stage.

3.3.2.6 FRASC37

FRASC37 was clinically diagnosed with Noonan syndrome. The patient presents with patent ductus arteriosus, café au-lait maculae, joint laxity and dysmorphic features including hypertelorism, ptosis, clinodactyly and frontal bossing.

Ensembl VEP annotated a total of 85481 variants in FRASC37 prior to filtering. One variant identified using the VEP-G2P plugin was prioritized for further investigation (Table 3.11).

Table 3.11. Shortlisted variants identified in FRASC37.

Gene	Variant	Type	Associated disease	Inheritance	ACMG codes applied	ACMG classification
WDFY3	c.8147C>T (p.Ala2716Val)	Heterozygous missense	Primary Microcephaly/ Macrocephaly with Developmental Delay	AD	PM2, PP3	VUS

The c.8147C>T (p.Ala2716Val) variant identified in *WDFY3* (NM_014991.6) was classified as a VUS using the ACMG-AMP guidelines. This variant was absent in population databases (ACMG code 2) and has not been reported previously in disease databases such as ClinVar. Multiple *in silico* prediction tools predicted this variant to have a damaging effect on the gene product (ACMG code PP3). Variants in the *WDFY3* gene cause neurodevelopmental delay, micro- or macrocephaly and behavioural problems such as autism spectrum disorder (Le Duc et al., 2019). Our patient was not reported to present with those clinical features. In addition, *WDFY3* has been shown to have a crucial role in cardiac development (Luo et al., 2021) which could potentially overlap with the patent ductus arteriosus observed in our patient. The variant had insufficient evidence to classify it as a pathogenic disease-causing variant, however it seems unlikely that this variant is the disease-causing variant based on the

information currently available. Future reanalysis of the variant and refined phenotyping could benefit the patient.

3.3.2.7 FRASC55

FRASC55 presented with mild intellectual disability and dysmorphic features such as ptosis, low set ears, a webbed neck and camptodactyly. The patient had a clinical diagnosis of Noonan syndrome.

A total of 88059 variants were annotated by Ensembl VEP for this patient. Two variants were shortlisted for further investigation. One of these variants was identified and prioritized from the list of variants returned by Moon (*SLC25A24* c.1417G>A) and one additional variant (SCN2A c.5836A>G) was identified using the VEP-DDG2P plugin (Table 3.12).

Table 3.12. Shortlisted variants of uncertain significance identified in FRASC55.

Gene	**Variant**	**Type**	**Associated disease**	**Inheritance**	**ACMG codes applied**	**ACMG classification**
SLC25A24	c.1417G>A (p.Gly473Arg)	Heterozygous missense	Gorlin-Chaudhry-Moss syndrome	AD	PM2, PP3	VUS
SCN2A	c.5836A>G (p.Lys1946Glu)	Heterozygous missense	Complex neurodevelopmental disorder	AD	PM2, PP3	VUS

The heterozygous missense variant in exon 10 of *SLC25A24* (NM_013386.5): c.1417G>A (p.Gly473Arg) was classified as a VUS due to insufficient evidence. The variant was absent in population databases (ACMG code PM2) and multiple *in silico* tools predicted this variant to be damaging (ACMG code PP3). The only overlap in terms of clinical features for the associated Gorlin-Chaudhry-Moss syndrome are

dysmorphic facial features. Intellectual disability is not usually associated with this disorder (Rosti et al., 2013). This variant is unlikely to be the disease-causing variant in FRAC55.

The second candidate variant was also classified as a VUS due to insufficient evidence supporting pathogenicity. This variant in *SCN2A* (NM_001040142.2): c.5836A>G (p.Lys1946Glu) was absent in population databases (ACMG code PM2) and multiple *in silico* tools predicted the variant to be damaging (ACMG code PP3). Variants in *SCN2A* are associated with a complex neurodevelopmental disorder most notably associated with seizures, developmental delay and intellectual disability (Reynolds, King and Gorman, 2020). Intellectual disability overlaps with the phenotype of our patient.

None of the candidate variants had sufficient evidence to confidently classify it as the disease-causing variant. More extensive clinical phenotyping, functional assays for the variants and segregation analysis are required to improve the interpretation of these variants and their association with disease in our patient.

3.3.2.8 IDP27

The patient had clinical features suggestive of Noonan syndrome. She presented with cardiac defects (tricuspid and mitral valve regurgitation), macrocephaly, abnormal skin pigmentation and dysmorphic features including coarse facial features, micrognathia and posteriorly rotated ears.

Ensembl VEP annotated a total of 86614 variants in this patient. Two variants were shortlisted, one identified through the automated filtering workflow (*TP63* c.468C>A) and the second (*TET3* c.789G>C) through the VEP-DDG2P plugin (Table 3.13).

Table 3.13. Shortlisted variants of uncertain significance identified in IDP27.

Gene	Variant	Type	Associated disease	Inheritance	ACMG codes applied	ACMG classification
TP63	c.468C>A (p.Phe156Leu)	Heterozygous missense	*TP63*-related disorders	AD	PM2, PP3	VUS
TET3	c.789G>C (p.Gln263His)	Heterozygous missense	Beck-Fahrner syndrome	AD and AR	PM2	VUS

The c.468C>A (p.Phe156Leu) missense variant in *TP63* (NM_003722.5) was absent in population databases (ACMG coed PM2) and *in silico* tools predicted this variant to be damaging (ACMG code PP3). Based on this evidence, the variant was classified as a VUS. *TP63*-related disorders are a group of six syndromes with a broad spectrum of clinical features (Nanda et al., 2021). The variant identified in *TET3* (NM_001287491.2): c.789G>C (p.Gln263His) was also classified as a VUS due to insufficient available information. Only the PM2 ACMG code could be assigned due to the absence of the variant in population databases. *In silico* prediction tools had conflicting predictions with regards to the effect of this variant and thus PP3 could not be assigned. Both shortlisted variants had insufficient evidence to classify it as a disease-causing variant. It is unlikely that one of these variants is the disease-causing variant in our patient based on the currently available information. Future reanalysis of the variants and refined phenotyping could aid in identifying the putative disease-causing variant.

3.3.3 Diagnostic yield

Disease-causing variants were identified in three patients (Table 3.14). Two of these variants were classified as pathogenic using the ACMG-AMP guidelines, while the remaining variant was classified as likely pathogenic. The diagnostic yield for WES for patients in whom no pathogenic variants were identified using a targeted gene panel is thus 37.5% (3/8). The genetic diagnosis for one of these patients confirmed the suspected clinical diagnosis of Noonan syndrome. The other two variants that were identified were associated with disorders that do not form part of the RASopathies and

this changes the initial clinical diagnosis of these patients. Five patients remain without a molecular diagnosis.

Table 3.14. Summary of disease-causing variants identified using WES.

Patient	Suspected clinical diagnosis	Gene	Variant	Type	Associated disease	ACMG classification
IDP17	Noonan syndrome	*RIT1*	c.297T>G (p.Phe99Leu)	Heterozygous missense	Noonan syndrome	Pathogenic
DD3198	RASopathy	*SON*	c.5753_5756del, (p.Val1918GlufsTer87)	Heterozygous frameshift	ZTTK syndrome	Pathogenic
DD3425	Noonan syndrome	*CDK8*	c.95A>T (p.Tyr32Phe)	Heterozygous missense	Developmental disorder	Likely pathogenic

4. DISCUSSION

The RASopathies are a group of genetic developmental conditions (Rauen, 2013) and include the following disorders: NS, NSML, NS-LAH, NF1, NF2, CS, CFC, CM-AVM and LS (Tidyman and Rauen, 2016b). They are caused by germline pathogenic variants in one of several genes directly involved or regulating the RAS/MAPK pathway.

Overall, 20-30% of patients clinically diagnosed with a RASopathy remain without a molecular diagnosis (Lissewski et al., 2015). These patients either have causal variants in genes not yet discovered or were clinically misdiagnosed due to phenotypic overlap. Very few genetic studies have been carried out on African populations, especially those in sub-Saharan Africa (Essawi et al., 2013; Louati et al., 2014; El Bouchikhi et al., 2015, 2020; Ghedira et al., 2018; Tekendo-Ngongang and Kruszka, 2020). To the best of our knowledge, there are only two studies which have explored the genetic basis of RASopathies in South Africa to date. Both studies used targeted NGS panels containing only known RASopathy genes. The first study by Tekendo-Ngongang et al., 2019, had a diagnostic yield of 31.2% for Noonan syndrome and the second study by Ms Mudau as part of her PhD had an overall yield of 56.6% for a range of RASopathies (manuscript in preparation). This leaves a significant proportion of patients without a genetic diagnosis and thus further investigations are required to provide these patients with a molecular diagnosis and explore the genetic aetiology of patients with a RASopathy in South Africa.

The aim of this study was to determine whether the use of secondary testing methods could improve the diagnostic yields for South African patients with a RASopathy who remain without a molecular diagnosis following first line testing utilizing an NGS panel of known RASopathy genes. A subset of Ms Mudau's patient cohort and patients who formed part of an NGS validation in a diagnostic setting at the NHLS were investigated in this study. MLPA and WES were the methods of choice for reflex testing as outlined in chapter 2. This chapter will discuss the outcomes of this study, limitations, and future directions.

4.1 CNV detection in patients with NF1

NF1 is predominantly caused by pathogenic SNVs in the *NF1* gene. The variants include frameshift, nonsense, splicing and missense variants (Castellanos et al., 2020). The majority of these variants result in a loss of function of NF1 (Imbard et al., 2015; Büki et al., 2021) which is supported by ClinGen's curation of *NF1* as a haploinsufficient gene (pLI score = 0.9). Since most NF-causing variants are SNVs, the recommended first-tier molecular genetic test for NF1 is targeted sequencing of the *NF1* gene. The targeted NGS panel designed by Ms Maria Mudau as part of her PhD study as well as the IDP panel included *NF1,* however, no pathogenic variants could be identified in a large proportion of the patients with NF1 included in the Mudau PhD and IDP studies.

NF1 is also caused by CNVs, specifically deletions, in 5-11% of patients (Kehrer-Sawatzki and Cooper, 2021). These deletions range in size from a single exon, multiple exons (Imbard et al., 2015), the whole *NF1* gene and can also include genes flanking *NF1.* Four types of large *NF1* deletions, namely type 1, 2, 3 and atypical deletions, have been identified (Figure 4.1). These deletions encompass the entire *NF1* gene and flanking genes and are commonly termed NF1 microdeletions. The deletions differ in size and thus in the number of additional genes apart from *NF1* that are included within the deleted region (Kehrer-Sawatzki and Cooper, 2021). This leads to phenotypic variability and it is thus important to classify the type of deletion to establish an accurate genotype/phenotype correlation. Patients with large deletions typically have more severe clinical manifestations of NF1 compared to those with pathogenic SNVs (Kehrer-Sawatzki, Mautner and Cooper, 2017).

It was thus important to determine if those patients clinically diagnosed with NF1, but in whom no pathogenic variants were identified using targeted NGS panels had deletions of *NF1* and the flanking genes. MLPA was the chosen method in this study to investigate the presence of such deletions due to the availability of commercial kits that target *NF1* and the surrounding genes and the cost-effectiveness thereof compared to other available methods such as aCGH. This is especially important in a resource constraint environment such as the State sector of South Africa.

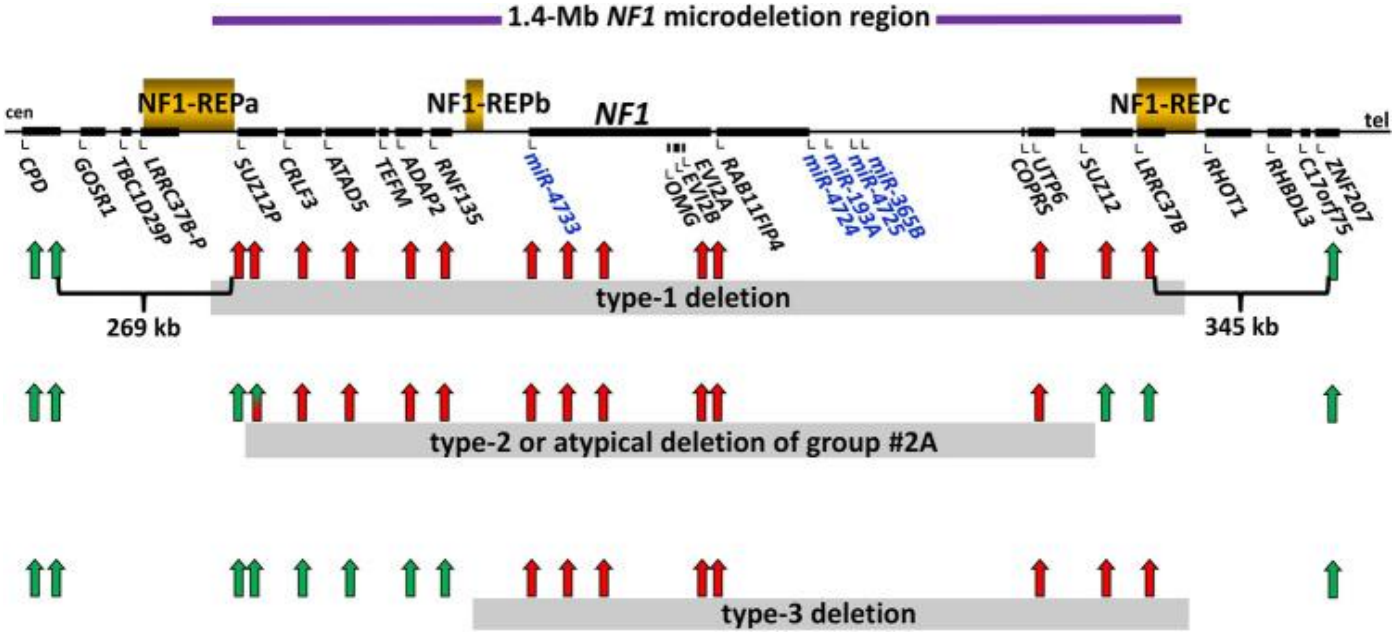


Figure 4.1. Different types of deletions identified in patients with NF1 (indicated by the gray blocks). The extent of the deletion and thus the number of deleted genes determine the type of deletion. The arrows indicate the location of the MLPA probes in the P122 probemix. The green arrows represent two copies of a particular probe/gene, i.e., not deleted and red arrows indicate the presence of only one copy of the probe/gene, i.e., a heterozygous deletion (Kehrer-Sawatzki and Cooper, 2021).

There are also multiple studies showing the effectiveness of MLPA as a secondary testing method for patients with NF1 to increase the diagnostic yield. Detection rates of deletions varied from 2.5% in Spain (Castellanos et al., 2020), 4% in France (Imbard et al., 2015), 10.5% in Turkey (Uludağ Alkaya et al., 2021), 11.5% in Italy (De Luca et al., 2007) and 15% in Taiwan (Wu-Chou et al., 2018). The only study conducted in Africa was in Egypt and had a deletion detection rate of 8% (Abdel-Aziz et al., 2021). To the best of our knowledge this is the first study in South Africa that used MLPA to detect deletions in patients clinically diagnosed with NF1.

Three different commercially available MLPA probemixes were used: P081, P082 and P122. Two of the probemixes, P081 and P082, collectively targeted each of the 58 exons of *NF1* and were used to detect single and multi-exon deletions, as well as whole *NF1* gene deletions. The third probemix, P122, predominantly targeted the genes flanking *NF1* and was used to determine the extent of the deletion. The MLPA P122 probemix is thus crucial to characterize the type of *NF1* deletion (Kehrer-Sawatzki and Cooper, 2021).

4.1.1 Identified deletions

Four different types of deletions in four patients were identified using MLPA. This included two large deletions: one type 1 and one type 3 deletion, as well as two smaller deletions: a single exon deletion (exon 6) and a multi-exon deletion (exon 19 and 20). The single exon deletion was determined to be a false positive result due to a pathogenic variant in the probe binding region. Therefore, only three of the identified deletions were true deletions responsible for the NF1 phenotype of the patients. Each of the deletions, as well as the identified variant will be discussed in the following sections.

4.1.1.1 *NF1* type 1 deletion

A heterozygous deletion of the whole *NF1* gene, as well as the following flanking genes, both upstream and downstream of *NF1*, were detected by MLPA in IDP113: *SUZ12P, CRLF3, ATAD5, ADAP2, RNF135, UTP6, SUZ12* and *LRRC37B*. This is classified as a type 1 deletion (Figure 4.1). Type 1 deletions are typically 1.4 Mb in size and include 14 protein coding and five microRNA genes (Kehrer-Sawatzki and Cooper, 2021). Even though every gene in the deleted region is not targeted by an MLPA probe, these genes can be assumed to be deleted since it is located within the boundaries of the MLPA targeted deleted genes and a contiguous deletion is expected.

A type 1 deletion is the most common NF1 microdeletion and accounts for 70-80% of large deletions (Kehrer-Sawatzki and Cooper, 2021). Type 1 deletions are mainly the result of interchromosomal non-allelic homologous recombination (NAHR) most commonly occurring during maternal meiosis. The breakpoints of these deletions are recurring and located in low copy repeat regions called NF1-REPa and NF1-REPc (Figure 4.1). The exact breakpoints cannot be determined by MLPA, however Summerer et al., 2018 investigated the breakpoints of type 1 deletions identified by MLPA using alternative methods such as aCGH and sequencing analysis of breakpoint regions to determine the efficiency of MLPA in the classification of these types of deletions. The study showed that MLPA can accurately classify type 1 deletions, since 99% of the type 1 deletions identified by MLPA were true type 1

deletions. It can thus be concluded that our patient's deletion was accurately classified as a type 1 deletion using MLPA and no additional investigation was pursued.

IDP113 had typical NF1 features including the presence of multiple café au lait maculae (>6), axillary and inguinal freckling, subcutaneous neurofibromas and iris Lisch nodules and thus met the diagnostic criteria for NF1. In addition to these features, the patient also presented with mild intellectual delay, dysmorphic features, macrocephaly and skeletal abnormalities. The latter, more severe clinical features are typically associated with type 1 deletions and are not commonly seen in patients with pathogenic SNVs in *NF1*. The full spectrum of features includes developmental delay, overgrowth, intellectual disability, dysmorphic facial features, macrocephaly, a large number of neurofibromas and cardiovascular anomalies (Büki et al., 2021). However, there is also clinical variability between patients with the same deletion (Kehrer-Sawatzki, Mautner and Cooper, 2017). Our patient's clinical presentation is thus consistent with that reported in the literature. Furthermore, this molecular result may assist in better defining the clinical presentation of the patient, as all the typical type 1 deletion features may not have been assessed accurately at the time of recruitment into the study.

The additional genes that are deleted in conjunction with *NF1* act as modifiers and are responsible for the more severe phenotype. Five of the protein-coding genes, including *NF1,* are intolerant to loss of function with pLI scores ranging between 0.97 and 1.0 and include: *ATAD5, OMG, RAB11FIP4* and *SUZ12*. These genes are therefore most likely to contribute to the phenotypic presentation of patients, however the contribution of the other deleted genes cannot be ruled out. *SUZ12* is a particularly important gene as haploinsufficiency thereof (pLI=1) leads to an increased risk (16-26%) of developing malignant nerve sheath tumours compared to typical patients with NF1 (8-13% risk). In addition, these malignancies also occur at an earlier age. This emphasizes the importance of identifying and accurately classifying large *NF1* deletions, since it is crucial for these high-risk patients to undergo regular screening (Kehrer-Sawatzki, Mautner and Cooper, 2017; Kehrer-Sawatzki and Cooper, 2021). Appropriate clinical screening, management and early intervention can now be received by IDP113 which could potentially lead to an improved quality of life for this patient.

4.1.1.2 *NF1* type 3 deletion

A type 3 NF1 deletion was identified in IDP89 using MLPA (Figure 4.1). This heterozygous deletion encompassed the entire *NF1* gene, as well as the following flanking genes downstream of *NF1* as identified by MLPA: *UTP6*; *SUZ12* and *LRRC37B* (Figure 4.1). Type 3 deletions are typically ~1Mb in size and encompass nine protein coding and five microRNA genes (Kehrer-Sawatzki and Cooper, 2021). These type of deletions are smaller than type 1 deletions and do not include any genes upstream of *NF1.*

Type 3 deletions are rare and constitute an estimated 1-4% of all NF1 microdeletions (Büki et al., 2021). Similar to type 1 deletions, the main causative mechanism of these deletions is NAHR, however the breakpoints are different (Figure 4.1). Type 1 and type 3 deletions share the NF1-REPc breakpoint, but the second breakpoint differs for type 3 deletions and is located in NF1-REPb (Kehrer-Sawatzki and Cooper, 2021). A previous study by Bengesser et al., 2010 used breakpoint spanning PCR to investigate the breakpoints of type 3 deletions in patients identified to have a loss of this region on MLPA. This study showed that MLPA is an efficient method to classify type 3 deletions. It is thus not necessary for us to confirm the breakpoints of the deletion in our patient using alternative methods as it is highly likely that this is a true type 3 deletion.

IDP89 had typical NF1 features including multiple neurofibromas, café au lait maculae and axillary freckling. He also presented with mild intellectual delay, speech delay and macrocephaly. Since type 3 deletions are rare, less is known about the associated phenotype compared to that of type 1 deletions. However, it is still expected that patients will present with a more severe phenotype compared to that of patients with pathogenic SNVs (Kehrer-Sawatzki and Cooper, 2021). Intellectual disability and dysmorphic features have been reported in patients with type 3 deletions (Bengesser et al., 2010). The severe phenotype can be attributed to the heterozygous loss of genes flanking *NF1.* Type 3 deletions also include *SUZ12* as in type 1 deletions. This means that patients with a type 3 deletion are also regarded as high-risk for developing malignancies at an early age and will therefore have to undergo regular surveillance to identify potential malignant tumours. This highlights the value of MLPA in providing

patients with an accurate genetic diagnosis, allowing for appropriate targeted clinical screening, management and early intervention.

4.1.1.3 *NF1* multi-exon deletion

A heterozygous multi-exon deletion encompassing two exons, exon 19 and 20, was identified in DD3072 using MLPA, specifically the P082 probemix. Various multi-exon deletions have been reported previously to be associated with NF1. These deletions varied in size from 2 to 48 exons (Wimmer et al., 2006; De Luca et al., 2007; Imbard et al., 2015; Wu-Chou et al., 2018; Castellanos et al., 2020; Kang et al., 2020). Most of the multi-exon deletions are unique to a single individual or family and have only been reported once (Wu-Chou et al., 2018; Kang et al., 2020). The deletion of exon 19 and 20 identified in our patient has not been reported in the literature or in disease databases such as Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER) (Firth et al., 2009) and therefore it is regarded as a novel deletion.

DD3072 presented with multiple café au lait maculae, axillary and inguinal freckling, facial and occipital neurofibromas and iris Lisch nodules typical of NF1. Furthermore, he also had learning difficulties, macrocephaly and chronic headaches. Since most of the multi-exon deletions are unique, a precise genotype/phenotype correlation has not been established (Wu-Chou et al., 2018; Kang et al., 2020). However, more severe phenotypes, particularly related to the presence of intellectual and learning disabilities, have been reported to be more common in patients with deletions and truncating variants compared to patients with pathogenic missense variants (Imbard et al., 2015; Kang et al., 2020). Functional studies should ideally be conducted to determine the effect of this deletion on the mRNA and NF1 protein structure and function.

4.1.1.4 *NF1* single exon deletion

A heterozygous single exon deletion of the *NF1* gene encompassing only exon 6 was identified in DD3546 using MLPA, specifically the P081 probemix. Similarly to multi-exon deletions, various different single exon deletions have been reported to be causative of NF1 (Wimmer et al., 2006; De Luca et al., 2007; Imbard et al., 2015; Kang

et al., 2020; Abdel-Aziz et al., 2021). However, single probe deletions must be confirmed by a secondary method since the presence of SNPs in the target sequence of the probe lead to false positive results.

Sanger sequencing was performed to confirm the single exon deletion using primers targeting regions flanking the probe target sequences. This method has been proven to be effective in previous studies (De Luca et al., 2007; Carrera et al., 2016). Sanger sequencing identified a heterozygous variant in the probe target sequence, thereby excluding the presence of a heterozygous deletion of exon 6. This variant affected the hybridization of the probe to the target DNA of the patient resulting in reduced amplification of the probe for this region and an incorrect interpretation thereof as a deletion (M. J. Kim et al., 2016).

The identified heterozygous c.616A>T, (p.Lys206Ter) variant in exon 6 of *NF1* was found to be a pathogenic disease-causing variant using the ACMG-AMP guidelines. This nonsense variant resulted in a premature termination codon expected to result in nonsense-mediated decay. Nonsense variants have been previously reported to cause NF1 due to haploinsufficiency and are the most prevalent type of SNVs (~30%) along with frameshift variants (Castellanos et al., 2020; Kang et al., 2020). This was a novel variant that has not been previously reported in the literature or population and disease databases.

DD3546 presented with typical NF1 features including multiple café au lait maculae, iris Lisch nodules, macrocephaly, short stature, joint hyperlaxity and asymmetry of the orbital area. An increased incidence of axillary freckling, large numbers of neurofibromas and iris Lisch nodules have been reported in patients with truncating and splicing variants compared to those with missense variants and inframe deletions (Kang et al., 2020). The heterozygous c.616A>T, (p.Lys206Ter) nonsense variant in DD3546 identified by Sanger sequencing thus confirms the diagnosis of NF1 in the patient. This patient underwent MLPA prior to first line testing using a targeted NGS panel due to time and sample constraints. NGS would have been able to detect this variant since the *NF1* gene is included on the IDP panel. However, this result highlights the importance of using secondary methods to exclude false positive results due to

the presence of either pathogenic variants or SNPs when single probe deletions are detected by MLPA.

4.1.2 Diagnostic yield of MLPA

The diagnostic yield of MLPA for the identification of CNVs in patients with a clinical diagnosis of NF1, but in whom no pathogenic variants were identified using a targeted NGS panel was 25.0% (3/12). MLPA thus improved the diagnostic yield of the primary study by providing a genetic diagnosis to three additional patients, thereby confirming their clinical diagnosis of NF1. Fourteen patients with a clinical diagnosis of NF1, but in whom no pathogenic variants were detected in the NF1 gene were selected for MLPA. Twelve patients were included in the diagnostic yield calculation and two patients were excluded. One of these patients failed quality control after performing the MLPA reaction and was therefore not included in the analysis. The other patient had a false positive result on MLPA confirmed to have been caused by the presence of a pathogenic SNV which would have been detected by a targeted NGS panel. The sample size in this study was small and a more accurate measure of the effectiveness of MLPA would be obtained with a greater sample size. The diagnostic yield of 25.0% obtained by this study indicates that MLPA is a valuable method for the identification of deletions in South African patients with NF1. MLPA was able to accurately identify and discern heterozygous *NF1* deletions of varying sizes (type 1, type 3 and a multi-exon deletion) enabling us to understand the patients' phenotypes better and direct clinical management and screening accordingly. It can thus be recommended to implement MLPA in a diagnostic setting as a reflex test for patients who remain without a genetic diagnosis following targeted NGS.

4.2 WES in RASopathy patients

In the second part of the study, WES was used to investigate patients with suspected RASopathy disorders other than NF1. Seven of these patients had a clinical diagnosis of Noonan syndrome and one patient had a non-specific RASopathy diagnosis. No pathogenic variants were identified for these patients in 21 known RASopathy genes included on Ms Mudau's targeted NGS panel (*CBL, LZTR1, KRAS, RRAS, NRAS, PTPN11, RAF1, SOS1, RIT1, SOS2, RASA2, RASA1, BRAF, MAP2K1, MAP2K2,*

SPRED1, NF1, NF2, HRAS, SHOC2 and *A2ML1)*; or the 13 RASopathy associated genes included on the IDP panel (*NF1, NF2, SPRED1, PTPN11, RAF1, SOS1, BRAF, KRAS, MAP2K1, MAP2K2, RASA1, LZTR1* and *NRAS)*.

RASopathies are predominantly caused by pathogenic SNVs of which the majority are missense variants (Rauen, 2013; Tidyman and Rauen, 2016b). The prevalence of CNVs in Noonan syndrome and RASopathy disorders other than NF1 and NF2 is rare. There are only a few case reports of CNVs recorded for *PTPN11*, *SHOC2* (J. L. Chen et al., 2014), *RAF1* (Luo et al., 2012) and *MAP2K2* (Nowaczyk et al., 2014; Čizmárová et al., 2016). CNVs are thus not considered to be a major contributor to the pathogenesis of these disorders (Lissewski et al., 2015). Therefore, a testing strategy to detect CNVs was not chosen for this group of patients as with the patients with NF1. Instead, WES was used to investigate potential disease-causing variants in genes not included on the targeted NGS panels.

WES is not limited to a few known disease-causing genes like targeted gene panels and therefore has the potential to establish novel gene-disease associations by identifying pathogenic variants in genes not previously recognized to be causative of RASopathies (Dillon et al., 2018). There have been a few studies in recent years in which newly associated RASopathy genes were identified using WES (Tidyman and Rauen, 2016a). This is especially useful in understudied populations, such as African populations, for whom the genetic aetiology of RASopathies is largely unknown.

Individuals of African ancestry harbour the most genetic variation of all population groups. In addition, there is also less mixing with non-African populations. Despite the genetic diversity of African populations, most genetic studies are centered on individuals of European ancestry and thus global genetic disease diversity is not representative (Wonkam, 2021). This limits our understanding of the genetic basis of disease, which in turn affects the translation of research into clinical practice. Our current knowledge is thus incomplete or incorrect in some instances for understudied population groups (Sirugo, Williams and Tishkoff, 2019; Wonkam, 2021).

This highlights the importance of including individuals of African ancestry in genetic studies. Due to the genetic diversity, studying African populations can lead to the

discovery of novel variants in known genes or novel genes contributing to disease that have not been revealed in previous European based studies (McClellan, Lehner and King, 2017; Wonkam, 2021). African genomes also provide a basis to determine a gene's tolerance to variants which aids in establishing gene-disease associations and interpreting the pathogenicity of variants. It also increases knowledge of the frequency of disease-causing variants in different population groups (McClellan, Lehner and King, 2017). This information is vital to improve our understanding of the genetic basis of disease, predict the course of disease and direct accurate diagnostic testing and appropriate clinical management, as well as establishing new therapies. These discoveries are not only applicable to African populations but can be extended to other population groups (Wonkam, 2021).

The clinical features of the RASopathy disorders not only overlap with those within the RASopathy group, but there is also clinical overlap between RASopathies and other disorders. It is therefore possible that a patient who is clinically diagnosed with a RASopathy could have a differential diagnosis. In such cases, WES can be useful to identify variants in genes unrelated to the RAS/MAPK pathway (Dillon et al., 2018). This has been demonstrated by Chen et al., 2014 who discovered a pathogenic variant in a gene associated with a rare developmental disorder distinct from RASopathies by using WES. This ultimately changed the patient's diagnosis.

WES was performed using the Ion Torrent system. Manual filtering was done for all patients following a sequential filtering process, however only one pathogenic disease-causing variant was identified using this strategy. The alternative analysis workflow that employed Moon, a commercial genomic prioritization tool using an automated filtering strategy and incorporating the phenotypes of patients, identified two additional disease-causing variants. The third filtering workflow, VEP-DDG2P, which incorporates all known genes causative of developmental disorders, did not yield any additional disease-causing variants. However, VEP-DDG2P was able to identify the two causal variants shortlisted by Moon which makes it a valuable tool to use for WES analysis. In comparison to Moon, VEP-DDG2P is more affordable, since it is free to use. However, VEP-DDG2P requires bioinformatics skills, whereas Moon's graphical interface makes it easier to use.

4.2.1 Identified pathogenic variants

Two pathogenic and one likely pathogenic disease-causing variant were identified in three patients using WES. Two of these variants were missense variants and one was a frameshift variant. Of these three variants, one was novel and two have been previously reported to be associated with a specific disease phenotype. Only one of the variants was in a known RASopathy gene and confirmed the patient's clinical diagnosis. The other two variants were in genes unrelated to the RAS/MAPK pathway or RASopathies and changed the diagnosis of the patients. Each of these variants will be discussed in the sections that follow.

4.2.1.1 *RIT1*: c.297T>G (p.Phe99Leu)

A heterozygous missense variant, c.297T>G (p.Phe99Leu), was identified in exon 5 of *RIT1* for patient IDP17 using WES and the manual filtering strategy. *RIT1* is a known RASopathy gene with pathogenic variants in the gene causing Noonan syndrome. A gene-disease association for *RIT1* and Noonan syndrome was first described by Aoki et al., 2013. RIT1 is part of the Ras protein subfamily. This family of GTPases cycle between an active GTP bound and an inactive GDP bound state. In the active state, Ras proteins interact with several effector proteins to transmit downstream signals and regulate the RAS/MAPK pathway (Van et al., 2020; Riller and Rieux-Laucat, 2021).

This c.297T>G (p.Phe99Leu) variant was classified as pathogenic using the ACMG-AMP guidelines. The variant has been reported previously to be associated with Noonan syndrome (Yaoita et al., 2016) and is a common recurrent variant in patients with a RASopathy (Van et al., 2020). A large majority of pathogenic variants (22%) affect codon 99 (alternatively codon 82 depending on the transcript). Two other missense variants (c.297T>A and c.295T>C) reported to be pathogenic and resulting in the same amino acid substitution as our variant support the pathogenicity of our variant. Furthermore, amino acid changes at codon 99 from phenylalanine to valine and serine, respectively, are also causative of Noonan syndrome (Kouz et al., 2016). The c.297T>G (p.Phe99Leu) variant affects a highly conserved amino acid located in the Switch II functional domain of RIT1 (Aoki et al., 2013; Yaoita et al., 2016; Van et al., 2020). This missense variant is expected to cause a gain of function effect and

lead to enhanced transactivation of ELK1, a transcription factor downstream of ERK in the RAS/MAPK pathway (Aoki et al., 2013; Kouz et al., 2016). The *RIT1* c.297T>G missense variant has ample evidence to support the classification of the variant as a pathogenic disease-causing variant.

IDP17 presented with hypertrophic obstructive cardiomyopathy, pulmonary stenosis, short stature, hypotonia, failure to thrive, skin pigmentary abnormalities, dysmorphic features and urogenital abnormalities suggestive of Noonan syndrome. The identified variant thus confirms the patient's diagnosis of Noonan syndrome. The incidence of hypertrophic cardiomyopathy and pulmonary stenosis is higher in patients with *RIT1* variants compared to Noonan syndrome patients with variants in other genes (Aoki et al., 2016; Yaoita et al., 2016). This leads to a higher mortality rate in patients with *RIT1* variants (Van et al., 2020).

Even though *RIT1* is a known RASopathy gene, this variant was not detected using the targeted NGS panel. *RIT1* was not included on the panel due to a limited number of genes that could be selected. At the time of the design of the panel, this gene was not considered to contribute significantly to RASopathies. Pathogenic variants in this gene are responsible for 5% of Noonan syndrome cases compared to the two major genes, *PTPN11* (50%) and *SOS1* (10-15%) (Aoki et al., 2016; Kouz et al., 2016). This highlights the importance of WES compared to NGS panels which could miss variants due to panel design and availability of information at the time of design.

4.2.1.1 *SON*: c.5753_5756del (p.Val1918GlufsTer87)

A heterozygous four base pair deletion, c.5753_5756del (p.Val1918GlufsTer87), was identified in exon 3 of the *SON* gene for DD3198. The *SON* gene does not form part of the RAS/MAPK pathway and is not associated with any of the RASopathy disorders. Pathogenic variants in the *SON* gene cause a rare autosomal dominant developmental disorder known as ZTTK syndrome (Kushary et al., 2021), first reported in 2015 by Zhu et al.. Only 60 cases have been reported in the literature to date, predominantly in European populations (Kushary et al., 2021). To the best of our knowledge, this is the first case of ZTTK syndrome identified in a patient of African ancestry.

This c.5753_5756del, (p.Val1918GlufsTer87) variant identified in our patient was classified as pathogenic using the ACMG-AMP guidelines. It is in the conserved Arginine/Serine-rich (RS) domain of *SON* and causes a frameshift resulting in a premature termination codon. This is predicted to lead to reduced mRNA levels due to nonsense-mediated decay and ultimately causes haploinsufficiency of the SON protein (J. H. Kim et al., 2016). Loss of function variants are a known cause of ZTTK syndrome with frameshift variants being the most prevalent (69%) (Kushary et al., 2021). The identified deletion has been previously reported and is the most common recurrent variant identified for ZTTK syndrome (Kushary et al., 2021). The SON protein is involved in constitutive as well as alternative RNA splicing. Therefore, haploinsufficiency of SON, as observed in patients with ZTTK syndrome, results in faulty RNA splicing of multiple genes that are vital for neural development. These genes show significantly reduced expression as a result of the accumulation of mis-spliced products. This affects multiple systems involved in growth and development (J. H. Kim et al., 2016; Tokita et al., 2016; Yang and Yang, 2020).

DD3198 presented with an atrial septal defect and ventricular septal defect, macrocephaly, developmental delay, cryptorchidism, hypoplastic nails and dysmorphic facial features and was suspected of having one of the RASopathy disorders. However, the identification of this pathogenic variant in *SON* changes the diagnosis of the patient. The phenotypic spectrum for ZTTK syndrome is broad with many of the features overlapping with that of the RASopathies. This is evident in a study by Kushary et al., 2021, where two of the patients in the cohort were also suspected of having a RASopathy. The majority of patients present with developmental delay and intellectual disability. Other clinical features commonly observed in patients include hypotonia, musculo-skeletal abnormalities, genitourinary abnormalities and dysmorphic facial features. Cardiac defects are more prevalent in patients that have the c.5753_5756del, (p.Val1918GlufsTer87) variant (Yang and Yang, 2020; Kushary et al., 2021).

A clinical misdiagnosis is not uncommon in patients with ZTTK syndrome. This can be attributed to the rarity and phenotypic variability of the disorder. Most patients reported in the literature were diagnosed by means of WES and had multiple genetic tests done prior to this (Kushary et al., 2021). Furthermore, disease-causing variants in *SON* are

usually *de novo* (Kushary et al., 2021). If this variant is confirmed to be absent in our patient's parents, it would have implications for the patient's parents and other family members. The recurrence risk for *de novo* variants is <1%, which would aid in the parents' reproductive decisions since the risk of having another child affected with ZTKK syndrome is low and prenatal testing is likely not necessary. Furthermore, the risk of other family members having an affected child is also low (Savatt and Myers, 2021).

This case highlights the importance and utility of WES in diagnosing patients with ZTTK syndrome, especially identifying African patients with rare or emerging phenotypes which could expand the phenotypic spectrum of these disorders and improve our understanding thereof.

4.2.1.2 *CDK8*: c.95A>T (p.Tyr32Phe)

A heterozygous missense variant located in exon 1 of *CDK8*: c.95A>T (p.Tyr32Phe) was identified in DD3425. The *CDK8* gene, similarly to *SON,* does not form part of the RAS/MAPK pathway and is not associated with any of the RASopathy disorders. Pathogenic variants in *CDK8* cause a recently identified, rare developmental disorder. To the best of our knowledge, only 14 cases have been reported in the literature (Calpena et al., 2019; Uehara et al., 2020) and this is the first case identified in an individual of African ancestry.

The c.95A>T (p.Tyr32Phe) variant has not been reported previously and is thus regarded as a novel variant. Using the ACMG-AMP guidelines this variant was classified as likely pathogenic. All reported pathogenic variants associated with this disorder are missense variants which supports the pathogenicity of our variant. These variants occur *de novo* (Calpena et al., 2019; Uehara et al., 2020), however inheritance has not been determined for our patient due to the unavailability of the patient's parents for testing. The variant is present in the highly conserved Gly-rich loop of the protein kinase domain of CDK8. Multiple other pathogenic variants cluster in this CDK8 domain. Furthermore, disease-causing variants are also found in other protein kinases. The protein kinase domain of CDK8 is crucial for ATP binding and phosphoryl transfer reactions which are vital for CDK8's function (Calpena et al.,

2019). CDK8 regulates transcription factors and numerous signaling pathways and therefore a change in its activity can have detrimental effects (Calpena et al., 2019). Studies have shown that missense variants in *CDK8* result in a functional loss of the kinase activity rendering it unable to phosphorylate its targets (Calpena et al., 2019; Uehara et al., 2020).

CDK8 forms part of a CDK/cyclin module consisting of CDK8, MED12 and MED13 or CDK19, MED12L and MED13L. Pathogenic variants in *MED12*, *MED13* and *MED12L* have been reported to cause developmental disorders with similar phenotypes to that of the *CDK8* associated disorder. Patients with the *CDK8* associated disorder typically present with intellectual disability, dysmorphic facial features, hypotonia and behavioral problems. Cardiac defects and brain abnormalities such as agenesis of the corpus callosum and seizures are also observed (Calpena et al., 2019). The clinical features of our patient, DD3425, are consistent with the phenotype of *CDK8* associated disorder. DD3425 presented with short stature, macrocephaly, hypotonia, developmental delay and dysmorphic facial features. Furthermore, he had prominent surface sulci and ventricles of the brain and minor seizures. He was, however, clinically diagnosed with Noonan syndrome and the identification of this variant changes his diagnosis. Since this *CDK8* associated disorder is extremely rare and not yet well described an accurate clinical diagnosis would have been unlikely. Similarly, to ZTTK syndrome all patients diagnosed with this disorder were identified through WES or whole genome sequencing (WGS). This highlights the value of WES to establish an accurate diagnosis for patients with rare developmental disorders. Furthermore, this study makes a valuable contribution to the phenotypic spectrum of an emerging developmental disorder by discussing this disorder's presentation in a patient of African ancestry.

4.2.2 Diagnostic yield

The diagnostic yield for WES in patients with a clinical diagnosis of RASopathies other than NF1, but in whom no pathogenic variants were identified using a targeted NGS panel was 37.5% (3/8). WES improved the diagnostic yield for these patients by providing a diagnosis to three additional patients. The genetic diagnosis for one of these patients confirmed the suspected clinical diagnosis of Noonan syndrome. The

other two variants that were identified were associated with developmental disorders that do not form part of the RASopathies and this changed the initial clinical diagnosis of these patients. The diagnostic yield of 37.5% is comparable to previously reported overall yields of 25-58% (Fung et al., 2020). Five patients remain without a molecular diagnosis and could benefit from future reanalysis and reevaluation of their clinical phenotypes since the significance of the shortlisted variants is currently uncertain.

4.3 Clinical utility of the study

The clinical utility of a testing strategy is the value or benefit that such a test or result provide to a patient and their family. The test or result should be able to direct patient care appropriately to improve the patient's clinical outcome with regards to providing a diagnosis, determining the progression of the disease, management and surveillance, treatment strategies, as well as reproductive decisions. An accurate diagnosis is a crucial step that will affect the other tiers of patient care (Joseph et al., 2016). The two reflex testing strategies that were used in this study both demonstrated clinical value and will be discussed in this section.

MLPA proved valuable by providing a genetic diagnosis to three patients clinically diagnosed with NF1, but in whom no disease-causing variants could be identified using a targeted NGS panel as a first line testing method. Numerous studies have demonstrated the efficiency of MLPA as a reflex test to identify CNVs (Imbard et al., 2015; Castellanos et al., 2020; Abdel-Aziz et al., 2021; Uludağ Alkaya et al., 2021). MLPA was able to accurately identify and discern two large deletions (Bengesser et al., 2010; Summerer et al., 2018) and one multi-exon deletion in our study cohort. The detection of large deletions in patients with NF1 is important since it is associated with a severe phenotype and an increased risk of developing malignancies at an early age (Büki et al., 2021). MLPA is thus able to direct the clinical management, surveillance and counselling of these patients. Early identification of malignancies will increase life expectancy. In addition, features that are not yet present or severe in the patient can be monitored and planned for accordingly to improve quality of life. Since only ~50% of causal variants are *de novo,* testing can also be offered to other at-risk family members and the recurrence risk can be established to guide reproductive choices (Ly and Blakeley, 2019).

WES was able to provide a genetic diagnosis to three patients when first line testing using a targeted NGS panel could not identify any causal variants. The patients were suspected of having a RASopathy based on their clinical presentation, however, two of the patients were diagnosed with developmental disorders unrelated to the RASopathies and RAS/MAPK pathway using WES. The phenotypes of these disorders (ZTTK syndrome and *CDK8* associated developmental disorder) overlapped with that of the RASopathies (Calpena et al., 2019; Kushary et al., 2021) which affected the ability to make an accurate clinical diagnosis. Furthermore, both disorders are rare, emerging disorders with limited information available in the literature. WES has been shown to be effective in providing a genetic diagnosis to patients with atypical phenotypes, complex phenotypes, disorders with overlapping features, rare or newly characterized disorders (Shickh et al., 2021). It is thus evident that if WES was not performed, the putative disease-causing variants would not have been identified in these patients leaving them with an incorrect clinical diagnosis which would have affected their clinical management and outcome. Clinical management of these patients can now be directed appropriately in line with their genetic diagnosis. Previous studies have shown how WES changed the management and care of patients after receiving an accurate genetic diagnosis (Wise et al., 2019). Parental testing can establish whether the variants are *de novo* which would inform the recurrence risk for the parents, as well as other family members. In addition, similarly to MLPA, the diagnosis would aid in counselling and reproductive decisions (Shickh et al., 2021).

In addition to providing patients with an accurate genetic diagnosis thereby directing their clinical care, WES and MLPA also contributed to expanding the phenotypic and molecular spectrum of NF1, Noonan syndrome, ZTTK syndrome and *CDK8* associated syndrome. This will lead to a better understanding of these disorders, especially in individuals of African ancestry. African populations are understudied despite having the most genetic variation of all population groups (Sirugo, Williams and Tishkoff, 2019; Choudhury et al., 2020). Literature and public databases are not representative of African genomes and our current knowledge of genetic variation in individuals of African ancestry and the contribution thereof to disease is insufficient (Choudhury et al., 2020). A substantial proportion of variants of uncertain significance are discovered through WES. These variants are difficult to interpret with the currently

limited available information, especially in understudied African populations. To the best of our knowledge, this is the first instance of ZTTK syndrome and *CDK8* syndrome reported in individuals of African ancestry. This study added information to the lack of genetic diversity in the literature and databases. This information will aid in identifying disease associated genes and interpreting the pathogenicity of variants in African populations. Furthermore, it will translate into appropriate clinical interventions and direct testing for individuals of African ancestry (Sirugo, Williams and Tishkoff, 2019; Choudhury et al., 2020; Shickh et al., 2021).

WES is expensive and bioinformatically challenging (Dillon et al., 2018) and even though this test proved to be valuable it will be a considerable challenge to implement it as a diagnostic test in the State sector of South Africa in the near future due to limited resources. On the contrary, MLPA is less costly and MLPA testing for other disorders is already performed routinely in a diagnostic setting in the State sector. Therefore, MLPA would likely be implemented as a reflex test for patients with NF1.

4.4 Limitations

A limitation of this study is that MLPA cannot define exact breakpoints of deletions. Studies have shown that MLPA can accurately detect type 1 and type 3 deletions as identified in this study. However, type 2 deletions and atypical deletions cannot be accurately classified using MLPA (Kehrer-Sawatzki and Cooper, 2021). Even though we did not identify any type 2 or atypical deletions, this should be mentioned as a potential limitation of the technique, especially when implemented in a diagnostic setting. Alternative techniques such as aCGH or breakpoint spanning PCR must be performed to confirm such deletions. Additionally, single probe deletions as identified in this study also have to be confirmed by using a secondary method since it could be a potential false positive result (M. J. Kim et al., 2016). The use of confirmatory techniques increases the cost and time of testing. This is also true for the implementation of reflex testing and is an important consideration for setting up RASopathy testing in low resource settings such as the State sector of South Africa.
Another limitation of the study is that CNVs were only investigated for patients with NF1 and not for patients with other RASopathy syndromes such as Noonan syndrome. Even though CNVs are rare in Noonan syndrome, there have been case reports of

such in *PTPN11*, *SHOC2* (J. L. Chen et al., 2014), *RAF1* (Luo et al., 2012) and *MAP2K2* (Nowaczyk et al., 2014; Čizmárová et al., 2016). CNVs can therefore not be ruled out and should be investigated further in future studies for patients who remain without a genetic diagnosis following WES.

The limited data available in the literature and in public databases for African populations, both in terms of phenotypes and molecular profiles, is also a limitation. Software such as Moon utilizes the phenotype of patients to identify causal variants (O'Brien et al., 2021). Gene-disease associations are based primarily on phenotypes of European populations, and this might result in causal variants being overlooked for African patients due to atypical phenotypes or clinical heterogeneity in these individuals. Additionally, insufficient data for African populations in public databases such as gnomAD affects the ability to accurately classify variants identified using WES. Variants are more likely to be classified as VUS due to a lack of evidence supporting the pathogenicity thereof. Variants could also be incorrectly classified as pathogenic due to the apparent rarity thereof (absent or negligible minor allele frequency) in African populations attributed to insufficient representation of these population groups (Sirugo, Williams and Tishkoff, 2019).

4.5 Future work

Even after undergoing testing using a targeted NGS panel and MLPA, nine patients presenting with features suggestive of NF1 remain without a molecular diagnosis. Further investigation to determine the genetic cause of NF1 in these patients should be undertaken. RNA analysis of NF1 has been shown to identify variants such as deep intronic variants or splicing variants that cannot be detected using DNA based testing such as NGS (Evans et al., 2016). These variants are usually present outside of the exon/intron boundary region sequenced in NGS based testing using DNA. Furthermore, RNA analysis can help determine the effect that a VUS has on splicing and thus its contribution to disease (Evans et al., 2016; Koster et al., 2021). Using RNA sequencing, Koster et al., 2021 identified pathogenic deep intronic variants in 25% of their cohort for which DNA based approaches did not yield a result. Utilizing RNA based methods such as RNA sequencing is thus recommended when DNA approaches fail to detect disease-causing variants in patients with NF1.

A pipeline to investigate CNVs using the WES data generated in this study should be explored for the five patients in whom no disease-causing SNVs were identified. There are several tools and software available for CNV calling using NGS data, each with its own advantages and disadvantages (Zhao et al., 2020). No standardized method has been established (Zanardo et al., 2020). Any CNVs identified using this approach should ideally be confirmed with a secondary method such as aCGH. Even though CNVs are rare and not regarded as a major cause for RASopathy disorders other than for NF1, a thorough investigation can rule out any potential deletions or duplications. Furthermore, as shown in this study, patients could also have a differential diagnosis not suspected clinically and thus potentially harbour CNVs in genes associated with a different developmental disorder.

Periodic reanalysis of WES data for patients who remain without a genetic diagnosis following this study has the potential to further improve the diagnostic yield. As more information becomes available, variants or genes that were not prioritized during the initial analysis could yield a genetic diagnosis by incorporating the new knowledge. Alternatively, prioritized variants that did not meet the criteria for a pathogenic classification could be reclassified as evidence emerges. The discovery of new gene-disease associations or evidence supporting current gene-disease associations will aid in reanalysis. Furthermore, sequencing of parental samples could also aid in identifying previously overlooked pathogenic variants (Fung et al., 2020). A reevaluation of the patient's phenotype can be valuable to direct further testing. A patient's clinical presentation can change with age and new diagnosis may arise as disorders are better described, diagnostic criteria are established, or phenotypes are expanded (Wise et al., 2019).

4.6 Conclusions

RASopathies have many overlapping features within the RASopathy spectrum, as well as with other disorders. Very little information about the genetic aetiology of RASopathies is available in South Africa. An accurate diagnosis is crucial for patient management and treatment of clinical manifestations. This study aimed to determine whether MLPA and WES could improve the diagnostic yield for patients with a RASopathy in South Africa after undergoing first line testing using targeted NGS

panels. This study provided a diagnosis for six additional patients. MLPA identified three deletions including one multi-exon deletion of *NF1* and two large deletions encompassing the whole *NF1* gene as well as flanking genes. WES identified two pathogenic and one likely pathogenic variant. Two of these variants were associated with developmental disorders other than one of the RASopathies and thus changed the diagnosis of the patients. This was the first cases of ZTTK syndrome and *CDK8* associated disorder identified in Africa and specifically South Africa. This study significantly improved the diagnostic rate and accuracy for patients suspected of having a RASopathy and provides insight into improving genetic testing for patients in South Africa. It highlighted the importance of WES in providing a molecular diagnosis for patients with rare disorders that would not be suspected clinically, especially when clinical features are less well defined or overlap with that of other disorders. It also resulted in the potential implementation of MLPA in a diagnostic setting as a reflex test for patients with NF1 who remain without a molecular diagnosis following first line testing using a targeted NGS panel. Overall, this study demonstrates the value of reflex testing using both MLPA and WES in a South African setting.

5. REFERENCES

Abdel-Aziz, N., El-Kamah, G., Khairat, R., et al. (2021). Mutational spectrum of NF1 gene in 24 unrelated Egyptian families with neurofibromatosis type 1. *Molecular Genetics and Genomic Medicine*. 9(12), e1361.

Abe, Y., Aoki, Y., Kuriyama, S., et al. (2012). Prevalence and clinical features of Costello syndrome and cardio-facio-cutaneous syndrome in Japan: Findings from a nationwide epidemiological survey. *American Journal of Medical Genetics Part A*. 158A(5), 1083-1094.

Adzhubei, I.A., Schmidt, S., Peshkin, L., et al. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*. 7(4), 248-249.

Altmüller, F., Lissewski, C., Bertola, D., et al. (2017). Genotype and phenotype spectrum of NRAS germline variants. *European Journal of Human Genetics*. 25(7), 823-831.

Aoki, Y., Niihori, T., Banjo, T., et al. (2013). Gain-of-function mutations in RIT1 cause Noonan syndrome, a RAS/MAPK pathway syndrome. *American Journal of Human Genetics*. 93(1), 173-180.

Aoki, Y., Niihori, T., Inoue, S.I. and Matsubara, Y. (2016). Recent advances in RASopathies. *Journal of Human Genetics*. 61(1), 33-39.

Athota, J.P., Bhat, M., Nampoothiri, S., et al. (2020). Molecular and clinical studies in 107 Noonan syndrome affected individuals with PTPN11 mutations. *BMC Medical Genetics.* 21(1).

Bakker, H.D., de Sonnaville, M.C., Vreken, P., et al. (2001). Human α-N-acetylgalactosaminidase (α-NAGA) deficiency: no association with neuroaxonal dystrophy? *European Journal of Human Genetics*. 9, 91-96.

Baraitser, M. and Patton, M.A. (1986). A Noonan-like short stature syndrome with sparse hair. *Journal of Medical Genetics*. 23(2), 161-164.

Beck, T.F., Mullikin, J.C. & Biesecker, L.G. (2016). Systematic evaluation of sanger validation of next-generation sequencing variants. *Clinical Chemistry*. 62(4), 647-654.

Bengesser, K., Cooper, D.N., Steinmann, K., et al. (2010). A novel third type of recurrent NF1 microdeletion mediated by nonallelic homologous recombination between LRRC37B-containing low-copy repeats in 17q11.2. *Human Mutation*. 31(6), 742-751.

Bergqvist, C., Servy, A., Valeyrie-Allanore, L., et al. (2020). Neurofibromatosis 1 French national guidelines based on an extensive literature review since 1966. *Orphanet Journal of Rare Diseases*. 15(1), 1-23.

Bhoj, E.J., Yu, Z., Guan, Q., et al. (2017). Phenotypic predictors and final diagnoses in patients referred for RASopathy testing by targeted next-generation sequencing. *Genetics in Medicine*. 19(6), 715-718.

Brems, H., Chmara, M., Sahbatou, M., et al. (2007). Germline loss-of-function mutations in SPRED1 cause a neurofibromatosis 1-like phenotype. *Nature Genetics*. 39(9), 1120-1126.

Brems, H., Pasmant, E., van Minkelen, R., et al. (2012). Review and update of SPRED1 mutations causing Legius syndrome. *Human Mutation*. 33(11), 1538-1546.

Büki, G., Zsigmond, A., Czakó, M., et al. (2021). Genotype-Phenotype Associations in Patients with Type-1, Type-2, and Atypical NF1 Microdeletions. *Frontiers in Genetics*. 12.

Calpena, E., Hervieu, A., Kaserer, T., et al. (2019). De Novo Missense Substitutions in the Gene Encoding CDK8, a Regulator of the Mediator Complex, Cause a Syndromic Developmental Disorder. *American Journal of Human Genetics*. 104(4), 709-720.

Capri, Y., Flex, E., Krumbach, O.H.F., et al. (2019). Activating Mutations of RRAS2 are a Rare Cause of Noonan Syndrome. *American Journal of Human Genetics*. 104(6), 1223-1232.

Carrera, P., Calzavara, S., Magistroni, R., et al. (2016). Deciphering Variability of PKD1 and PKD2 in an Italian Cohort of 643 Patients with Autosomal Dominant Polycystic Kidney Disease (ADPKD). *Scientific Reports*. 6.

Castellanos, E., Rosas, I., Negro, A., et al. (2020). Mutational spectrum by phenotype: panel-based NGS testing of patients with clinical suspicion of RASopathy and children with multiple café-au-lait macules. *Clinical Genetics*. 97(2), 264-275.

Chabás, A., Duque, J. and Gort, L. (2007). A new infantile case of α-N-acetylgalactosaminidase deficiency. Cardiomyopathy as a presenting symptom. *Journal of Inherited Metabolic Disease*. 30(1), 108.

Chen, J.L., Zhu, X., Zhao, T.L., et al. (2014). Rare copy number variations containing genes involved in RASopathies: Deletion of SHOC2 and duplication of PTPN11. *Molecular Cytogenetics*. 7(1).

Chen, P.C., Yin, J., Yu, H.W., et al. (2014). Next-generation sequencing identifies rare variants associated with Noonan syndrome. *Proceedings of the National Academy of Sciences of the United States of America*. 111(31), 11473-11478.

Cheung, S.W. and Bi, W. (2018). Novel applications of array comparative genomic hybridization in molecular diagnostics. *Expert Review of Molecular Diagnostics*. 18(6), 531-542.

Choi, Y. and Chan, A.P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 31(16), 2745-2747.

Choudhury, A., Aron, S., Botigué, L.R., et al. (2020). High-depth African genomes inform human migration and health. *Nature*. 586(7831), 741-748.

Čizmárová, M., Hlinková, K., Bertok, S., et al. (2016). New Mutations Associated with Rasopathies in a Central European Population and Genotype-Phenotype Correlations. *Annals of Human Genetics*. 80(1), 50-62.

Cordeddu, V., di Schiavi, E., Pennacchio, L.A., et al. (2009). Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nature Genetics*. 41(9), 1022-1026.

Costello, J.M. (1977). A new syndrome: mental subnormality and nasal papillomata. *Journal of Paediatrics and Child Health*. 13(2), 114-118.

Dashti, M.J.S. and Gamieldien, J. (2017). A practical guide to filtering and prioritizing genetic variants. *BioTechniques*. 62(1), 18-30.

De Abreu, F.B., Peterson, J.D., Amos, C.I., et al. (2016). Effective quality management practices in routine clinical next-generation sequencing. *Clinical Chemistry and Laboratory Medicine*. 54(5), 761-771.

De Luca, A., Bottillo, I., Dasdia, M.C., et al. (2007). Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. *Journal of Medical Genetics*. 44(12), 800-808.

Den Dunnen, J.T., Dalgleish, R., Maglott, D.R., et al. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*. 37(6), 564-569.

Dentici, M.L., Sarkozy, A., Pantaleoni, F., et al. (2009). Spectrum of MEK1 and MEK2 gene mutations in cardio-facio-cutaneous syndrome and genotype-phenotype correlations. *European Journal of Human Genetics*. 17(6), 733-740.

Dillon, O.J., Lunke, S., Stark, Z., et al. (2018). Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *European Journal of Human Genetics*. 26(5), 644-651.

Eerola, I., Boon, L.M., Mulliken, J.B., et al. (2003). Capillary Malformation-Arteriovenous Malformation, a New Clinical and Genetic Disorder Caused by RASA1 Mutations. *American Journal of Human Genetics*. 73, 1240-1249.

Eid, O.M. (2017). Molecular cytogenetic techniques for identification of copy-number variations. *Middle East Journal of Medical Genetics*. 6(1), 1-12.

El Bouchikhi, I., Samri, I., Iraqui Houssaini, M., et al. (2015). The first PTPN11 mutations in hotspot exons reported in Moroccan children with Noonan syndrome and comparison of mutation rate to previous studies. *Turkish Journal of Medical Sciences*. 45(2), 306-312.

El Bouchikhi, I., Bouguenouch, L., Moufid, F.Z., et al. (2020). Molecular and environmental characterization of Noonan syndrome in Morocco reveals a significant association with consanguinity and advanced parental age. *Egyptian Journal of Medical Human Genetics*. 21(1), 1-9.

Ella, M., Pierpont, M., Magoulas, P.L., et al. (2014). Cardio-Facio-Cutaneous Syndrome: Clinical Features, Diagnosis, and Management Guidelines. *Pediatrics*. 134(4), e1149-e1162.

Essawi, M.L., Ismail, M.F., Afifi, H.H., et al. (2013). Mutational analysis of the PTPN11 gene in Egyptian patients with Noonan syndrome. *Journal of the Formosan Medical Association*. 112(11), 707-712.

Evans, D.G.R. (2009). Neurofibromatosis type 2 (NF2): A clinical and molecular review. *Orphanet Journal of Rare Diseases*. 4(1).

Evans, D.G., Bowers, N., Burkitt-Wright, E., et al. (2016). Comprehensive RNA Analysis of the NF1 Gene in Classically Affected NF1 Affected Individuals Meeting NIH Criteria has High Sensitivity and Mutation Negative Testing is Reassuring in Isolated Cases with Pigmentary Features only. *EBioMedicine*. 7, 212-220.

Evans, G.D., King, A.T., Bowers, N.L., et al. (2018). Identifying the deficiencies of current diagnostic criteria for neurofibromatosis 2 using databases of 2777 individuals with molecular testing The English Specialist NF2 Research Group. *Genetics in Medicine*. 21(7), 1525-1533.

Firth, H. v., Richards, S.M., Bevan, A.P., et al. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics*. 84(4), 524-533.

Flex, E., Jaiswal, M., Pantaleoni, F., et al. (2014). Activating mutations in RRAS underlie a phenotype within the RASopathy spectrum and contribute to leukaemogenesis. *Human Molecular Genetics*. 23(16), 4315-4327.

Friedman, J.M. (2019). Neurofibromatosis 1. In Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallis, S.E., Bean, L.J.H., Stephens, K. and Amemiya, A. (eds). Genereviews®. Available online at: https://www.ncbi.nlm.nih.gov/books/NBK1109/

Fung, J.L.F., Yu, M.H.C., Huang, S., et al. (2020). A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *npj Genomic Medicine*. 5(1).

Gelb, B.D., Cavé, H., Dillon, M.W., et al. (2018). ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genetics in Medicine*. 20(11), 1334-1345.

Ghedira, N., Lagarde, A., Ben Ameur, K., et al. (2018). Clinical profile of comorbidity of rare diseases in a Tunisian patient: A case report associating incontinentia pigmenti and Noonan syndrome. *BMC Pediatrics*. 18(1), 1-7.

Gorlin, R.J., Anderson, R.C. and Blaw, M. (1969). Multiple Lentigines Syndrome. *American Journal of Diseases of Children*. 117(6), 652-662.

Grant, A.R., Cushman, B.J., Cavé, H., et al. (2018). Assessing the gene–disease association of 19 genes with the RASopathies using the ClinGen gene curation framework. *Human Mutation*. 39(11), 1485-1493.

Green, M.R. and Sambrook, J. (2018). Isolation and quantification of DNA. *Cold Spring Harbor Protocols*. 2018(6), 403-414.

Green, M.R. and Sambrook, J. (2019). Analysis of DNA by agarose gel electrophoresis. *Cold Spring Harbor Protocols*. 2019(1), 6-15.

Gremer, L., De Luca, A., Merbitz-Zahradnik, T., et al. (2010). Duplication of Glu37 in the switch I region of HRAS impairs effector/GAP binding and underlies Costello syndrome by promoting enhanced growth factor-dependent MAPK and AKT activation. *Human Molecular Genetics*. 19(5), 790-802.

Gripp, K.W., Aldinger, K.A., Bennett, J.T., et al. (2016). A novel RASopathy caused by recurrent de novo missense mutations in PPP1CB closely resembles Noonan syndrome with loose anagen hair. *American Journal of Medical Genetics Part A*. 170(9), 2237-2247.

Gripp, K.W., Schill, L., Schoyer, L., et al. (2019). The sixth international RASopathies symposium: Precision medicine—From promise to practice. *American Journal of Medical Genetics Part A*. 182(3), 1-10.

Gripp, K.W., Morse, L.A., Axelrad, M., et al. (2019). Costello syndrome: Clinical phenotype, genotype, and management guidelines. *American Journal of Medical Genetics Part A*. 179(9), 1725-1744.

Gripp, K.W., Baker, L., Robbins, K.M., et al. (2020). The novel duplication HRAS c.186_206dup p.(Glu62_Arg68dup): clinical and functional aspects. *European Journal of Human Genetics*. 28(11), 1548-1554.

Gross, A.M., Frone, M., Gripp, K.W., et al. (2020). Advancing RAS/RASopathy therapies: An NCI-sponsored intramural and extramural collaboration for the study of RASopathies. *American Journal of Medical Genetics Part A*. 182(4), 866-876.

Gudmundsson, S., Singer-Berk, M., Watts, N.A., et al. (2021). Variant interpretation using population databases: lessons from gnomAD. *Human Mutation*.

Halliday, D., Emmanouil, B., Pretorius, P., et al. (2017). Genetic severity score predicts clinical phenotype in NF2. *Journal of Medical Genetics*. 54(10), 657-664.

He, H., Guzman, R.E., Cao, D., et al. (2021). The molecular and phenotypic spectrum of CLCN4-related epilepsy. *Epilepsia*. 62(6), 1401-1415.

Huckstadt, V., Chinton, J., Gomez, A., et al. (2021). Noonan syndrome with loose anagen hair with variants in the PPP1CB gene: First familial case reported. *American Journal of Medical Genetics Part A*. 185(4), 1256-1260.

Imbard, A., Pasmant, E., Sabbagh, A., et al. (2015). NF1 single and multi-exons copy number variations in neurofibromatosis type 1. *Journal of Human Genetics*. 60(4), 221-224.

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics*. 99(4), 877-885.

Johnston, J.J., Van der Smagt, J.J., Rosenfeld, J.A., et al. (2018). Autosomal recessive Noonan syndrome associated with biallelic LZTR1 variants. *Genetics in Medicine*. 20(10), 1175-1185.

Joseph, L., Cankovic, M., Caughron, S., et al. (2016). The Spectrum of Clinical Utilities in Molecular Pathology Testing Procedures for Inherited Conditions and Cancer: A Report of the Association for Molecular Pathology. *Journal of Molecular Diagnostics*. 18(5), 605-619.

Kang, E., Kim, Y.M., Seo, G.H., et al. (2020). Phenotype categorization of neurofibromatosis type I and correlation to NF1 mutation types. *Journal of Human Genetics*. 65(2), 79-89.

Karczewski, K.J., Weisburd, B., Thomas, B., et al. (2017). The ExAC browser: Displaying reference data information from over 60 000 exomes. *Nucleic Acids Research*. 45(D1), D840-D845.

Kehrer-Sawatzki, H. and Cooper, D.N. (2021). Classification of NF1 microdeletions and its importance for establishing genotype/phenotype correlations in patients with NF1 microdeletions. *Human Genetics*. 140(12), 1635-1649.

Kehrer-Sawatzki, H., Mautner, V.F. and Cooper, D.N. (2017). Emerging genotype–phenotype relationships in patients with large NF1 deletions. *Human Genetics*. 136(4), 349-376.

Kim, J.H., Shinde, D.N.N., Reijnders, M.R.R.F., et al. (2016). De Novo Mutations in SON Disrupt RNA Splicing of Genes Essential for Brain Development and Metabolism, Causing an Intellectual-Disability Syndrome. *American Journal of Human Genetics*. 99(3), 711-719.

Kim, M.J., Cho, S.I., Chae, J.H., et al. (2016). Pitfalls of Multiple Ligation-Dependent Probe Amplifications in Detecting DMD Exon Deletions or Duplications. *Journal of Molecular Diagnostics*. 18(2), 253-259.

Kobayashi, T., Aoki, Y., Niihori, T., et al. (2010). Molecular and clinical analysis of RAF1 in Noonan syndrome and related disorders: Dephosphorylation of serine 259 as the essential mechanism for mutant activation. *Human Mutation*. 31(3), 284-294.

Kobylinski, O. (1883). Uber eine flughautahnliche Ausbreitung am Hales. *Archiv für Anthropologie*. 14, 343-348.

Koster, R., Brandão, R.D., Tserpelis, D., et al. (2021). Pathogenic neurofibromatosis type 1 (NF1) RNA splicing resolved by targeted RNAseq. *npj Genomic Medicine*. 6(1).

Kouz, K., Lissewski, C., Spranger, S., et al. (2016). Genotype and phenotype in patients with Noonan syndrome and a RIT1 mutation. *Genetics in Medicine*. 18(12), 1226-1234.

Kruszka, P., Porras, A.R., Addissie, Y.A., et al. (2017). Noonan syndrome in diverse populations. *American Journal of Medical Genetics Part A*. 173(9), 2323-2334.

Kushary, S.T., Revah-Politi, A., Barua, S., et al. (2021). ZTTK syndrome: Clinical and molecular findings of 15 cases and a review of the literature. American *Journal of Medical Genetics Part A*. 185(12), 3740-3753.

Lai, D. and Love, D.R. (2012). Automation of a primer design and evaluation pipeline for subsequent sequencing of the coding regions of all human Refseq genes. *Bioinformation*. 8(8), 365-368.

Landrum, M.J., Lee, J.M., Benson, M., et al. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*. 46(D1), D1062-D1067.

Landry, J.P., Schertz, K.L., Chiang, Y.J., et al. (2021). Comparison of Cancer Prevalence in Patients with Neurofibromatosis Type 1 at an Academic Cancer Center vs in the General Population from 1985 to 2020. *JAMA Network Open*. 4(3), e210945

Le Duc, D., Giulivi, C., Hiatt, S.M., et al. (2019). Pathogenic WDFY3 variants cause neurodevelopmental disorders and opposing effects on brain size. *Brain*. 142(9), 2617-2630.

Lee, B.H. and Yoo, H.-W. (2019). Noonan syndrome and RASopathies: Clinical features, diagnosis and management. *Journal of Genetic Medicine*. 16(1), 1-9.

Lee, W.H., Raas-Rotschild, A., Miteva, M.A., et al. (2005). Noonan syndrome type I with PTPN11 3 bp deletion: Structure-function implications. *Proteins: Structure, Function and Genetics*. 58(1), 7-13.

Lee, Y., Choi, Y., Seo, G.H., et al. (2021). Clinical and molecular spectra of BRAF-associated RASopathy. *Journal of Human Genetics*. 66(4), 389-399.

Legius, E., Messiaen, L., Wolkenstein, P., et al. (2021). Revised diagnostic criteria for neurofibromatosis type 1 and Legius syndrome: an international consensus recommendation. *Genetics in Medicine*. 23(8), 1506-1513.

Lepri, F., de Luca, A., Stella, L., et al. (2011). SOS1 mutations in Noonan syndrome: Molecular spectrum, structural insights on pathogenic effects, and genotype-phenotype correlations. *Human Mutation*. 32(7), 760-772.

Lissewski, C., Kant, S.G., Stark, Z., et al. (2015). Copy number variants including RAS pathway genes-How much RASopathy is in the phenotype? *American Journal of Medical Genetics Part A*. 167(11), 2685-2690.

Lissewski, C., Chune, V., Pantaleoni, F., et al. (2021). Variants of SOS2 are a rare cause of Noonan syndrome with particular predisposition for lymphatic complications. *European Journal of Human Genetics*. 29(1), 51-60.

Liu, X., Li, C., Mou, C., et al. (2020). dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Medicine*. 12(1), 103.

Liu, Z., Zhu, L., Roberts, R., et al. (2019). Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We? *Trends in Genetics*. 35(11), 852-867.

Lorenz, S., Lissewski, C., Simsek-Kiper, P.O., et al. (2013). Functional analysis of a duplication (p.E63_D69dup) in the switch II region of HRAS: New aspects of the molecular pathogenesis underlying Costello syndrome. *Human Molecular Genetics*. 22(8), 1643-1653.

Louati, R., Abdelmoula, N.B., Trabelsi, I., et al. (2014). Clinical and molecular findings of Tunisian patients with RASopathies. *Molecular Syndromology*. 5(5), 212-217.

Luo, C., Yang, Y.F., Yin, B.L., et al. (2012). Microduplication of 3p25.2 encompassing RAF1 associated with congenital heart disease suggestive of Noonan syndrome. *American Journal of Medical Genetics Part A*. 158A(8), 1918-1923.

Luo, S., Chen, L., Wei, W., et al. (2021). Prenatal Genetic Diagnosis in Three Fetuses with Left Heart Hypoplasia (LHH) From Three Unrelated Families. *Frontiers in Cardiovascular Medicine*. 8.

Ly, K.I. and Blakeley, J.O. (2019). The Diagnosis and Management of Neurofibromatosis Type 1. *Medical Clinics of North America*. 103(6), 1035-1054.

Keulemans, J.L., J Reuser, A.J., Kroos, M.A., et al. (1996). Human α-N-acetylgalactosaminidase (α-NAGA) deficiency: new mutations and the paradox between genotype and phenotype. *Journal of Medical Genetics*. 33, 458-464.

Martinelli, S., Stellacci, E., Pannone, L., et al. (2015). Molecular Diversity and Associated Phenotypic Spectrum of Germline CBL Mutations. *Human Mutation*. 36(8), 787-796.

Mazzanti, L., Cacciari, E., Cicognani, A., et al. (2003). Noonan-like syndrome with loose anagen hair: A new syndrome? *American Journal of Medical Genetics*. 118A(3), 279-286.

McClellan, J.M., Lehner, T. and King, M.C. (2017). Gene Discovery for Complex Traits: Lessons from Africa. *Cell*. 171(2), 261-264.

McLaren, W., Gil, L., Hunt, S.E., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*. 17(1).

Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*. 16(3), 1215.

Moreno-Cabrera, J.M., Del Valle, J., Castellanos, E., et al. (2020). Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *European Journal of Human Genetics*. 28(12), 1645-1655.

Motta, M., Sagi-Dain, L., Krumbach, O.H.F., et al. (2020). Activating MRAS mutations cause Noonan syndrome associated with hypertrophic cardiomyopathy. *Human Molecular Genetics*. 29(11), 1772-1783.

MRC Holland. (2019). MLPA ® General Protocol Instructions. MRC Holland. 1-12.

MRC Holland. (2020). Product Description SALSA ® MLPA ® Probemix P122-D2 NF1-area. MRC Holland. 1-8.

MRC Holland. (2021a). Coffalyser.Net Reference Manual Version 07. MRC Holland. 1-108.

MRC Holland. (2021b). Product Description SALSA ® MLPA ® Probemix P081-D1 NF1 mix 1 & P082-C2 NF1 mix 2. MRC Holland. 1-15.

Nagai, K., Niihori, T., Okamoto, N., et al. (2021). Duplications in the G3 domain or switch II region in HRAS identified in patients with Costello syndrome. *Human Mutation*. 43(1), 3-15.

Nanda, A., AILafi, A., Wolf, S., et al. (2021). TP63-related disorders: Two case reports and a brief review of the literature. *Dermatology Online Journal*. 27(11), 1-6.

Ngobeni, V., Breitenbach, M.C. and Aye, G.C. (2020). Technical efficiency of provincial public healthcare in South Africa. *Cost Effectiveness and Resource Allocation*. 18(1), 1-19.

Noonan, J.A. and Ehmke, D.A. (1963). Associated noncardiac malformations in children with congenital heart disease. *Journal of Pediatrics*. 63, 468-470.

Nowaczyk, M.J.M., Thompson, B.A., Zeesman, S., et al. (2014). Deletion of MAP2K2/MEK2: A novel mechanism for a RASopathy? *Clinical Genetics*. 85(2), 138-146.

O'Brien, T.D., Campbell, N.E., Potter, A.B., et al. (2021). Artificial intelligence (AI)-assisted exome reanalysis greatly aids in the identification of new positive cases and reduces analysis time in a clinical diagnostic laboratory. *Genetics in Medicine*. 24(1), 192-200.

Odebode, T.O., Afolayan, E.A.O., Adigun, I.A., et al. (2005). Clinicopathological study of neurofibromatosis type 1: An experience in Nigeria. *International Journal of Dermatology*. 44(2), 116-120.

Orme, C.M., Boyden, L.M., Choate, K.A., et al. (2013). Capillary Malformation - Arteriovenous Malformation Syndrome: Review of the Literature, Proposed Diagnostic Criteria, and Recommendations for Management. *Pediatric Dermatology*. 30(4), 409-415.

Pagnamenta, A.T., Kaisaki, P.J., Bennett, F., et al. (2019). Delineation of dominant and recessive forms of LZTR1-associated Noonan syndrome. *Clinical Genetics*. 95(6), 693-703.

Pasmant, E., Louvrier, C., Luscan, A., et al. (2018). Neurofibromatosis type 2 French cohort analysis using a comprehensive NF2 molecular diagnostic strategy. *Neurochirurgie*. 64(5), 335-341.

Pereira, R., Oliveira, J. and Sousa, M. (2020). Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. *Journal of Clinical Medicine*. 9(1).

Pinna, V., Daniele, P., Calcagni, G., et al. (2019). Prevalence, type, and molecular spectrum of NF1 mutations in patients with neurofibromatosis type 1 and congenital heart disease. *Genes*. 10(9).

Pounraja, V.K., Jayakar, G., Jensen, M., et al. (2019). A machine-learning approach for accurate detection of copy number variants from exome sequencing. *Genome Research*. 29(7), 1134-1143.

Quang, D., Chen, Y. and Xie, X. (2015). DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 31(5), 761-763.

Ramanjam, V., Adnams, C., Ndondo, A., et al. (2006). Clinical Phenotype of South African Children With Neurofibromatosis 1. *Journal of Child Neurology*. 21(1), 63-70

Rauen, K.A. (2013). The RASopathies. *Annual Review of Genomics and Human Genetics*. 14, 355-369.

Rehder, C., Bean, L.J.H., Bick, D., et al. (2021). Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*. 23(8), 1399-1415.

Rehm, H.L., Berg, J.S., Brooks, L.D., et al. (2015). ClinGen-The Clinical Genome Resource. *The New England Journal of Medicine*. 372(23), 2235-2242.

Rentzsch, P., Witten, D., Cooper, G.M., et al. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*. 47(D1), D886-D894.

Reva, B., Antipin, Y. and Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*. 39(17), e118.

Revencu, N., Boon, L.M., Mendola, A., et al. (2013). RASA1 Mutations and Associated Phenotypes in 68 Families with Capillary Malformation-Arteriovenous Malformation. *Human Mutation*. 34(12), 1632-1641.

Revencu, N., Fastre, E., Ravoet, M., et al. (2020). RASA1 mosaic mutations in patients with capillary malformation-arteriovenous malformation. *Journal of Medical Genetics*. 57(1), 48-52.

Reynolds, C., King, M.D. and Gorman, K.M. (2020). The phenotypic spectrum of SCN2A-related epilepsy. *European Journal of Paediactric Neurology.* 24, 117-122.

Reynolds, J.F., Neri, G., Hermann, J.P., et al. (1986). New multiple congenital anomalies/mental retardation syndrome with cardio-facio-cutaneous involvement-the CFC syndrome. *American Journal of Medical Genetics*. 25(3), 413-427.

Richards, S., Aziz, N., Bale, S., et al. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*. 17(5), 405-424.

Riller, Q. and Rieux-Laucat, F. (2021). RASopathies: From germline mutations to somatic and multigenic diseases. *Biomedical Journal*. 44(4), 422-432.

Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., et al. (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Research.* 77(21), e31-e34.

Rosti, R.O., Karaer, K., Karaman, B., et al. (2013). Gorlin-chaudhry-moss syndrome revisited: Expanding the phenotype. *American Journal of Medical Genetics Part A*. 161(7), 1737-1742.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 74(12), 5463-5467.

Savatt, J.M. and Myers, S.M. (2021). Genetic Testing in Neurodevelopmental Disorders. *Frontiers in Pediatrics*. 9.

Sayers, E.W., Beck, J., Bolton, E.E., et al. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. 49(D1), D10-D17.

Schouten, J.P., Mcelgunn, C.J., Waaijer, R., et al. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*. 30(12), e57.

Schwarz, J.M., Cooper, D.N., Schuelke, M., et al. (2014). Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*. 11(4), 361-362.

Seaby, E.G., Pengelly, R.J. and Ennis, S. (2016). Exome sequencing explained: A practical guide to its clinical application. *Briefings in Functional Genomics*. 15(5), 374-384.

Shickh, S., Mighton, C., Uleryk, E., et al. (2021). The clinical utility of exome and genome sequencing across clinical indications: a systematic review. *Human Genetics*. 140(10), 1403-1416.

Shihab, H.A., Gough, J., Cooper, D.N., et al. (2013). Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Human Mutation*. 34(1), 57-65.

Sim, N.L., Kumar, P., Hu, J., et al. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*. 40, W452-W457.

Sirugo, G., Williams, S.M. and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*. 177(1), 26-31.

Slatko, B.E., Gardner, A.F. and Ausubel, F.M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*. 122(1), e59.

Spencer, E., Davis, J., Mikhail, F., et al. (2011). Identification of SPRED1 deletions using RT-PCR, multiplex ligation-dependent probe amplification and quantitative PCR. *American Journal of Medical Genetics Part A*. 155(6), 1352-1359.

Stuppia, L., Antonucci, I., Palka, G., et al. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *International Journal of Molecular Sciences*. 13(3), 3245-3276.

Summerer, A., Mautner, V.F., Upadhyaya, M., et al. (2018). Extreme clustering of type-1 NF1 deletion breakpoints co-locating with G-quadruplex forming sequences. *Human Genetics*. 137(6-7), 511-520.

Suzuki-Muromoto, S., Miyabayashi, T., Nagai, K., et al. (2019). Leucine-485 deletion variant of BRAF may exhibit the severe end of the clinical spectrum of CFC syndrome. *Journal of Human Genetics*. 64(5), 499-504.

Tajan, M., Paccoud, R., Branka, S., et al. (2018). The RASopathy family: Consequences of germline activation of the RAS/MAPK pathway. *Endocrine Reviews*. 39(5), 676-700.

Tamura, R. (2021). Current understanding of neurofibromatosis type 1, 2, and schwannomatosis. *International Journal of Molecular Sciences*. 22(11), 5850.

Tekendo-Ngongang, C. and Kruszka, P. (2020). Noonan syndrome on the African Continent. *Birth Defects Research*. 112(10), 718-724.

Tekendo-Ngongang, C., Agenbag, G., Bope, C.D., et al. (2019). Noonan syndrome in South Africa: Clinical and molecular profiles. *Frontiers in Genetics*. 10.

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*. 526(7571), 68-74.

Thormann, A., Halachev, M., McLaren, W., et al. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nature Communications*. 10(1), 2373.

Tidyman, W.E. and Rauen, K.A. (2016a). Pathogenetics of the RASopathies. *Human Molecular Genetics*. 25(R2), R123-R132.

Tidyman, W.E. and Rauen, K.A. (2016b). Expansion of the RASopathies. *Current Genetic Medicine Reports*. 4(3), 57-64.

Tokita, M.J.J., Braxton, A.A.A., Shao, Y., et al. (2016). De Novo Truncating Variants in SON Cause Intellectual Disability, Congenital Malformations, and Failure to Thrive. *American Journal of Human Genetics*. 99(3), 720-727.

Uehara, T., Abe, K., Oginuma, M., et al. (2020). Pathogenesis of CDK8-associated disorder: two patients with novel CDK8 variants and in vitro and in vivo functional analyses of the variants. *Scientific Reports*. 10(1).

Uludağ Alkaya, D., Lissewski, C., Yeşil, G., et al. (2021). Expanding the clinical phenotype of RASopathies in 38 Turkish patients, including the rare LZTR1, RAF1, RIT1 variants, and large deletion in NF1. *American Journal of Medical Genetics Part A*. 185(12), 3623-3633.

Untergasser, A., Cutcutache, I., Koressaar, T., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*. 40(15), e115.

Van, R., Cuevas-Navarro, A., Castel, P. et al. (2020). The molecular functions of RIT1 and its contribution to human disease. *Biochemical Journal*. 477(15), 2755-2770.

Van der Burgt, I. (2007). Noonan syndrome. *Orphanet Journal of Rare Diseases*. 2(1).

Von Recklinghausen, F.D. (1882). Uber ide multiplen Fibrone der Haut und ihre beziehung zu den multiplen Neuromen. Berlin: Hirschwald. 3-18.

Watson, M. (2015). Clinical utility of genetic and genomic services: A position statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine*. 17(6), 505-507.

Wimmer, K., Yao, S., Claes, K., et al. (2006). Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Genes Chromosomes and Cancer*. 45(3), 265-276.

Wise, A.L., Manolio, T.A., Mensah, G.A., et al. (2019). Genomic medicine for undiagnosed diseases. *Lancet*. 394(10197), 533-540.

Wishart, J. (1822). Cases of tumours in the skull, dura mater and brain. *Edinburg Medical and Surgical Journal*. 18, 393-397.

Wonkam, A. (2021). Sequence three million genomes across Africa. *Nature*. 590, 209-211.

Wooderchak-Donahue, W.L., Johnson, P., McDonald, J., et al. (2018). Expanding the clinical and molecular findings in RASA1 capillary malformation-arteriovenous malformation. *European Journal of Human Genetics*. 26(10), 1521-1536.

Wright, C.F., Fitzgerald, T.W., Jones, W.D., et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of genome-wide research data. *The Lancet*. 385(9975), 1305-1314.

Wright, C.F., McRae, J.F., Clayton, S., et al. (2018). Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine*. 20(10), 1216-1223.

Wu-Chou, Y.H., Hung, T.C., Lin, Y.T., et al. (2018). Genetic diagnosis of neurofibromatosis type 1: Targeted next- generation sequencing with Multiple Ligation-Dependent Probe Amplification analysis. *Journal of Biomedical Science*. 25(1), 1-10.

Yang, L. and Yang, F. (2020). A de novo heterozygous variant in the SON gene is associated with Zhu-Tokita-Takenouchi-Kim syndrome. *Molecular Genetics and Genomic Medicine.* 8(11), e1496.

Yaoita, M., Niihori, T., Mizuno, S., et al. (2016). Spectrum of mutations and genotype–phenotype analysis in Noonan syndrome patients with RIT1 mutations. *Human Genetics*. 135(2), 209-222.

Zanardo, É.A., Monteiro, F.P., Chehimi, S.N., et al. (2020). Application of Whole-Exome Sequencing in Detecting Copy Number Variants in Patients with Developmental Delay and/or Multiple Congenital Malformations. *Journal of Molecular Diagnostics*. 22(8), 1041-1049.

Zeisler, E.P. and Becker, W.S. (1936). Generalised lentigo. *Archive for Dermatology and Syphilis*. 32, 109-25.

Zenker, M., Lehmann, K., Schulz, A.L., et al. (2007). Expansion of the genotypic and phenotypic spectrum in patients with KRAS germline mutations. *Journal of Medical Genetics*. 44(2), 131-135.

Zhao, L., Liu, H., Yuan, X., et al. (2020). Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinformatics*. 21(1), 1-10.

Zhu, X., Petrovski, S., Xie, P., et al. (2015). Whole-exome sequencing in undiagnosed genetic diseases: Interpreting 119 trios. *Genetics in Medicine*. 17(10), 774-781.

6. APPENDICES

Appendix 1. Ethics clearance certificates.

Appendix 1.1. Ethics clearance certificate obtained by Ms C. Smal .

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Ms C Smal

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M200617

NAME: (Principal Investigator)	Ms C Smal
DEPARTMENT:	School of Pathology Department of Human Genetics Medical School University
PROJECT TITLE:	Investigating the genetic aetiology of RASopathies in a cohort of African patients
DATE CONSIDERED:	2020/06/26
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr N Carstens and Ms M Mudau
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2020/08/04

This clearance certificate is valid for 5 years from the 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 on the 3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
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 submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **June** and will therefore reports and re-certification will be due early in the month of **June** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

20/08/2021
Date

Appendix 1.2. Ethic clearance certificate obtained by Ms M. Mudau.

R14/49 Ms Maria Mabyalwa Mudau

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M170163

NAME: (Principal Investigator)	Ms Maria Mabyalwa Mudau
DEPARTMENT:	School of Pathology, Division of Human Genetics
PROJECT TITLE:	Development of a Multi-Disease Targeted Next Generation Sequencing Panel to Study Genetic Aetiology of RASopathies
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	A Sub-study under Primary Study M160830 Dr Nadia Carstens
SUPERVISOR:	Dr Nadia Carstens and Prof Amanda Krause
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	15/02/2017

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 where the study was initially reviewed. In this case, the study was initially reviewed in January and will therefore be due in the month of January each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 27/02/2017

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix 1.3. Ethics clearance certificate obtained by Dr N. Carstens.

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

Appendix 1.4. Ethics clearance certificate obtained by Dr N. Carstens.

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee

20/06/2018

Approval Certificate
New Application

Ethics Reference No: 80/2018

Title: Using next-generation technologies to investigate the genetic aetiology of developmental delay

Dear Dr Nadia Carstens

The **New Application** as supported by documents specified in your cover letter dated 23/05/2018 for your research received on the 23/05/2018, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 20/06/2018.

Please note the following about your ethics approval:
- Ethics Approval is valid for 5 years
- Please remember to use your protocol number (**80/2018**) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, or monitor the conduct of your research.

Ethics approval is subject to the following:
- The ethics approval is conditional on the receipt of **6 monthly written Progress Reports**, and
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers; MBChB; MMed (Int); MPharMed,PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health).

☎ 012 356 3084 deepeka.behari@up.ac.za / fhsethics@up.ac.za http://www.up.ac.za/healthethics
✉ Private Bag X323, Arcadia, 0007 - Tswelopele Building, Level 4, Room 60 / 61, 31 Bophelo Road, Gezina, Pretoria

Appendix 1.5. Ethics approval for IDP patients.

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 I PO Box 1038, Johannesburg, 2000
[T] +27 11 489 9223 I [M] +27 78 080 8841 I [F] +27 11 489 9226 I [E] human.genetics@nhls.ac.za

Date: 19/03/2022

I wish to apply for permission to use data/samples from the Division of Human Genetics covered under the Ethics Clearance Certificate M180506:

Using anonymised residual diagnostic samples and –file data for audit, research and development:

Name of Staff member/Student:	Clarice Smal
Staff/ Student Number:	1930283
Degree/Non-degree purposes	Degree: MSc (Med) Human Genetics
E-mail address:	claricesmal@gmail.com / clarice.smal@nhls.ac.za
Contact Number:	0724042725
Name of Supervisor (if PI is a student):	Dr Nadia Carstens and Ms Maria Mudau
Staff/ Student Number:	A0040144/1143805
E-mail address:	maria.mudau@nhls.ac.za/mabyalwa.mudau@wits.ac.za
Contact Number:	0606970846

My study title is:

Investigating the genetic aetiology of RASopathies in a cohort of African patients

I require access to the following samples/patient files (Give numbers and description):

IDP59 ([redacted]), DD3072 ([redacted]), IDP113 ([redacted]), DD3546 ([redacted]), IDP89 ([redacted]), DD3198 ([redacted]), IDP17 ([redacted]), IDP27 ([redacted]), DD3425 ([redacted])

I will use samples/data covered under the following Section of the above mentioned Ethics clearance:

B,D,E. _____

Retrospective database of clinical patient records and residual diagnostic samples:

> **Category A**: Existing patient clinical records from 1970s to 1 September 2018, for which no consent has been obtained
>
> **Category B**: Existing residual diagnostic samples from mid 1980s for which no consent has been obtained

Prospective database of clinical patient records and residual diagnostic samples (after 1 September 2018)

> **Category C**: Clinical records of patients who presented to Human Genetics Clinics after 1 September 2018 from whom written consent is obtained.
>
> **Category D**: Residual Diagnostic samples from patients who presented to Human Genetics Clinics after 1 September 2018 from whom written consent is obtained.
>
> **Category E**: Residual Diagnostic samples from patients referred to the Division of Human Genetics from whom consent has not be obtained

I attach a summary of my main project objectives.

I am aware that if I may still need to apply to the Human Ethics Research Committee for approval for my research project

Signature: _____

Name: Clarice Smal_____

Supervisor signature: _____

Name: Maria Mudau

Approved:
Amanda Krause, MBBCh, PhD
Medical Geneticist/Associate Professor
Head of Division

Appendix 2. RASopathy clinical tick sheet.

Rasopathy Clinical Ticksheet

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 ____

PIGMENTARY ANOMALIES

CAL's 0000957 ☐ Number: _____ Average size: _____

Lentigines 0001003 ☐ Hyperpigmentation 0007440 ☐ Hypopigmentation 0007513 ☐

Freckling 0001480 ☐ Axillary 0000997 ☐ Inguinal 0030052 ☐

Other _____

ECTODERMAL FEATURES

Skin: Velvety 0000977 ☐ Deep palmar creases 0006191 ☐ Hyperkeratosis 0007543 ☐

Neurofibromas 0001067 ☐ Eczema 0000964 ☐

Hair: Sparse 0008070 ☐ Coarse 000208 ☐ Hypopigmented 0011358 ☐

Nails: Small 0001792 ☐ Deep-set 0001814 ☐ Dystrophic 0008404 ☐

Other _____

FACIAL FEATURES

General: Coarse 0000280 ☐ Frontal bossing 0002007 ☐

Eyes: DSPF 0000494 ☐ Ptosis 0000508 ☐ EF 0000286 ☐ Hypertelorism 0000316 ☐

Mouth: Full lips 0012471 ☐ Large 0000154 ☐

Neck: Webbed 0000465 ☐ Low posterior hairline 0002162 ☐

Ears: Low set 0000369 ☐ Posteriorly rotated 0000358 ☐

Other: _____

CARDIAC ANOMALIES

Hypertrophic cardiomyopathy 0001639 ☐ ASD 0010445 ☐ VSD 0001629 ☐ PDA 0001643 ☐ AS 0001650 ☐ PS 0001642 ☐

Other: _____

SKELETAL ANOMALIES

Chest: Shield chest 000914 ☐ Pectus excavatum 0000767 ☐ Pectus carinatum 0000768 ☐

Hands: Ulnar deviation 0009487 ☐ Short 5th metacarpal 0010047 ☐

Spine: Kyphosis 0002808 ☐ Scoliosis 0002650 ☐

Joints: Laxity 0001388 ☐ Camptodactyly 0100490 ☐

Other: _____

GENITALIA

Undescended testes 0000028 ☐ Small penis 0000054 ☐ Other: _____

NEURODEVELOPMENT

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

SENSORY DEVELOPMENT

Vision: Myopia 0000545 ☐ Strabismus 0000486 ☐ Other: _____

Hearing: Hearing loss ☐ Conductive 0000405 ☐ Sensorineural 0000407 ☐

OTHER IMPORTANT FEATURES/STRUCTURAL ANOMALIES

CNS:

Renal:

Appendix 3. MLPA probemixes

Appendix 3.1 Probes and target regions of MLPA probemix P081.

SALSA MLPA probe	Target region		
	Gene	Exon	Chromosome
Probe 18368-L23333	*NF1*	Upstream	17q
Probe 18363-L23328	*NF1*	Downstream	17q
Probe 02491-L29974	*NF1*	Exon 1	17q
Probe 02493-L01924	*NF1*	Exon 2	17q
Probe 02865-L02617	*NF1*	Exon 4	17q
Probe 02497-L03706	*NF1*	Exon 6	17q
Probe 18032-L22398	*NF1*	Exon 7	17q
Probe 13221-L26128	*NF1*	Exon 11	17q
Probe 02503-L22647	*NF1*	Exon 13	17q
Probe 04076-L22649	*NF1*	Exon 15	17q
Probe 02507-L22658	*NF1*	Exon 17	17q
Probe 12024-L26426	*NF1*	Exon 18	17q
Probe 21185-L29794	*NF1*	Exon 21	17q
Probe 21186-L29799	*NF1*	Exon 21	17q
Probe 18408-SP0653-L23405	*NF1*	Exon 23	17q
Probe 21000-L29222	*NF1*	Exon 23	17q
Probe 18033-SP0601-L29798	*NF1*	Exon 24	17q
Probe 03849-L18072	*NF1*	Exon 26	17q
Probe 18364-L23329	*NF1*	Exon 28	17q
Probe 04072-L03709	*NF1*	Exon 29	17q
Probe 02513-L01944	*NF1*	Exon 32	17q
Probe 18367-L23332	*NF1*	Exon 35	17q
Probe 02517-L26127	*NF1*	Exon 37	17q
Probe 02519-L01950	*NF1*	Exon 39	17q
Probe 02521-L22646	*NF1*	Exon 41	17q
Probe 03853-L29796	*NF1*	Exon 42	17q
Probe 04071-L01954	*NF1*	Exon 47	17q
Probe 02525-L22650	*NF1*	Exon 49	17q
Probe 02526-L01957	*NF1*	Exon 50	17q
Probe 19363-L25737	*NF1*	Exon 51	17q
Probe 02528-L01959	*NF1*	Exon 52	17q
Probe 05220-L03309	*NF1*	Exon 57	17q
Probe 02530-L01961	*NF1*	Exon 58	17q
Probe 19361-L26126	*NF1*	Exon 58	17q
Probe 04075-L03310	*OMG*	Intron 36 of NF1	17q
Reference probe 00797-L00463			5q
Reference probe 09940-L29795			8q
Reference probe 09836-L10246			11q
Reference probe 09265-L10877			10q
Reference probe 12437-L13438			14q
Reference probe 16436-L18889			18q
Reference probe 05388-L04785			12p
Reference probe 05953-L05397			2p
Reference probe 08725-L08736			9q
Reference probe 05026-L29797			2q
Reference probe 09908-L10321			16p

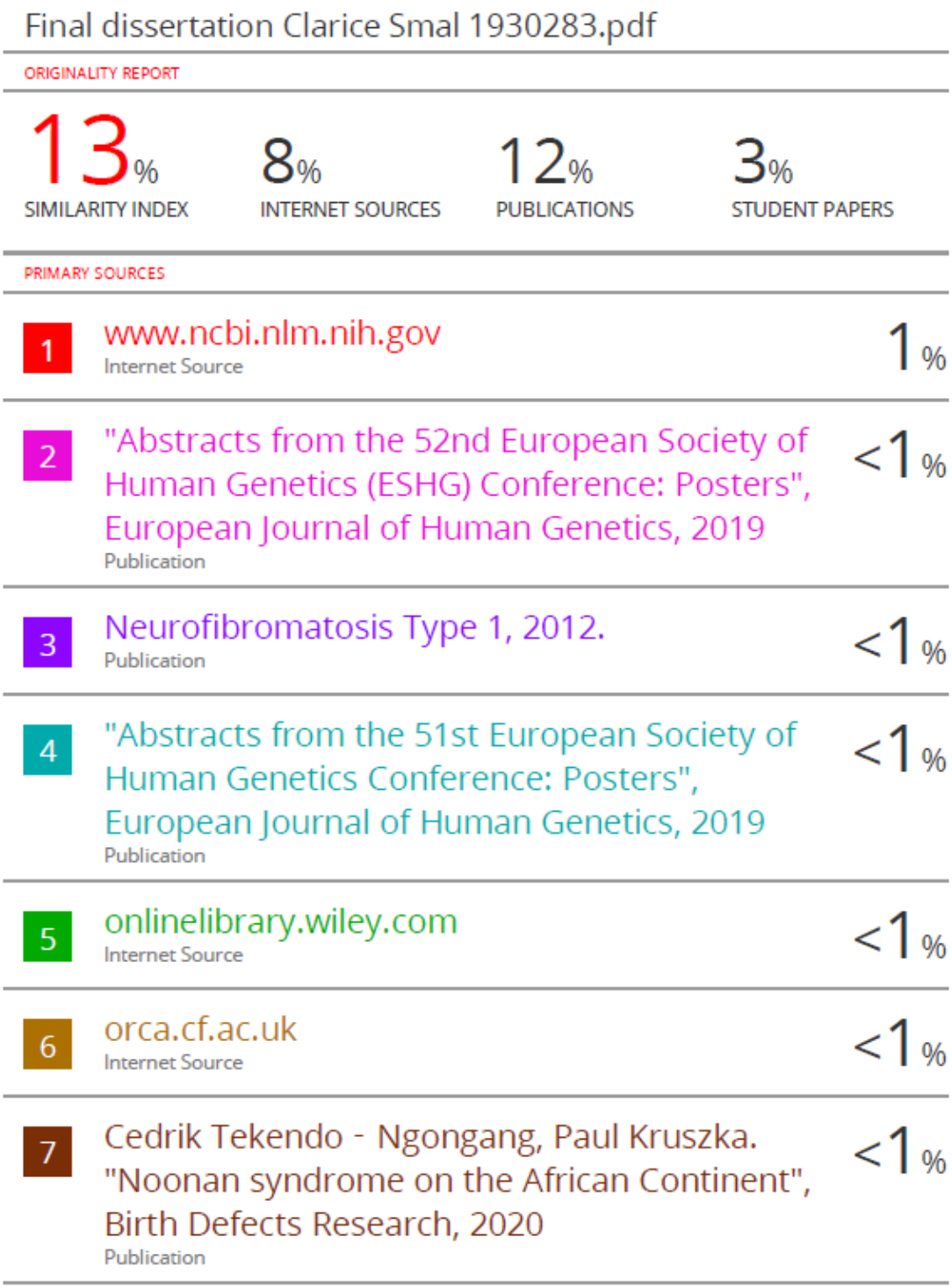
Appendix 3.2. Probes and target regions of MLPA probemix P082.

SALSA MLPA probe	Target region		
	Gene	Exon	Chromosome
Probe 18382-L19008	*NF1*	Exon 1	17q
Probe 19362-L26201	*NF1*	Exon 1	17q
Probe 02494-L01925	*NF1*	Exon 3	17q
Probe 18173-L22738	*NF1*	Exon 5	17q
Probe 02498-L22716	*NF1*	Exon 8	17q
Probe 18037-SP0602-L22403	*NF1*	Exon 9	17q
Probe 02500-L26202	*NF1*	Exon 10	17q
Probe 03778-L26198	*NF1*	Exon 12	17q
Probe 02504-L26817	*NF1*	Exon 14	17q
Probe 19364-SP0809-L25738	*NF1*	Exon 15	17q
Probe 18369-SP0646-L23334	*NF1*	Exon 16	17q
Probe 12025-L23157	*NF1*	Exon 19	17q
Probe 18370-L23335	*NF1*	Exon 20	17q
Probe 18036-L22765	*NF1*	Exon 22	17q
Probe 18174-SP0619-L22739	*NF1*	Exon 25	17q
Probe 18170-L26175	*NF1*	Exon 27	17q
Probe 02512-L01943	*NF1*	Exon 30	17q
Probe 18038-L26174	*NF1*	Exon 31	17q
Probe 18365-L23330	*NF1*	Exon 33	17q
Probe 02514-L01945	*NF1*	Exon 34	17q
Probe 18374-L26502	*NF1*	Exon 36	17q
Probe 02518-L01949	*NF1*	Exon 38	17q
Probe 02520-L26200	*NF1*	Exon 40	17q
Probe 03854-L23156	*NF1*	Exon 43	17q
Probe 12021-L26199	*NF1*	Exon 44	17q
Probe 03856-L03307	*NF1*	Exon 45	17q
Probe 02522-L01953	*NF1*	Exon 46	17q
Probe 02524-L22720	*NF1*	Exon 48	17q
Probe 13217-L22725	*NF1*	Exon 51	17q
Probe 12018-L12866	*NF1*	Exon 53	17q
Probe 18034-L22721	*NF1*	Exon 54	17q
Probe 02529-L01960	*NF1*	Exon 55	17q
Probe 18035-L22401	*NF1*	Exon 56	17q
Probe 12019-L12867	*NF1*	Intron 1	17q
Probe 04069-L03311	*OMG*	Intron 36 of NF1	17q
Reference probe 07808-L23525			3p
Reference probe 06708-L26176			10p
Reference probe 06676-L06254			11p
Reference probe 00797-L00463			5q
Reference probe 15957-L26197			6q
Reference probe 08893-L23475			14q
Reference probe 11571-L12318			16q
Reference probe 17862-L22121			19q
Reference probe 12427-L1342			22q

Appendix 3.3. Probes and target regions of MLPA probemix P122.

SALSA MLPA probe	Target region			
	Gene	*Exon*	Chromosome	Location relative to *NF1*
Probe 01325-L07456	*ASPA*	Exon 5	17p13.3	23101 kb upstream
Probe 01463-L00928	*PMP22*	Exon 3	17p12	11343 kb upstream
Probe 09176-L19109	*TRAF4*	Exon 2	17q11.2	2347 kb upstream
Probe 08620-L08632	*TRAF4*	Exon 4	17q11.2	2347 kb upstream
Probe 09627-L09912	*BLMH*	Exon 9	17q11.2	822 kb upstream
Probe 09628-L21977	*CPD*	Exon 11	17q11.2	651 kb upstream
Probe 09629-L09914	*CPD*	Exon 20	17q11.2	632 kb upstream
Probe 09629-L09914	*SUZ12P1*	Exon 1	17q11.2	363 kb upstream
Probe 11801-L12592	*SUZ12P1*	Exon 3	17q11.2	336 kb upstream
Probe 03780-L03289	*CRLF3*	Exon 3	17q11.2	297 kb upstream
Probe 03781-L03290	*ATAD5*	Exon 2	17q11.2	260 kb upstream
Probe 03782-L03291	*ADAP2*	Exon 3	17q11.2	168 kb upstream
Probe 03783-L03292	*RNF135*	Exon 2	17q11.2	110 kb upstream
Probe 02491-L01922	*NF1*	Exon 1	17q11.2	
Probe 02507-L01938	*NF1*	Exon 17	17q11.2	
Probe 02512-L01943	*NF1*	Exon 30	17q11.2	
Probe 02525-L01956	*NF1*	Exon 49	17q11.2	
Probe 05220-L03309	*NF1*	Exon 57	17q11.2	
Probe 03785-L03294	*UTP6*	Exon 14	17q11.2	515 kb downstream
Probe 03786-L03295	*SUZ12*	Exon 10	17q11.2	628 kb downstream
Probe 03787-L03296	*LRRC37B*	Exon 1	17q11.2	661 kb downstream
Probe 09637-L09949	*ZNF207*	Exon 9	17q11.2	1006 kb downstream
Probe 09632-L09917	*PSMD11*	Exon 2	17q11.2	1086 kb downstream
Probe 09631-L09916	*MYO1D*	Exon 7	17q11.2	1407 kb downstream
Probe 09630-L09915	*MYO1D*	Exon 2	17q11.2	1420 kb downstream
Reference probe 20555-L14777			1p31	
Reference probe 05714-L05152			2q11	
Reference probe 17349-L21793			3p25	
Reference probe 19616-L26684			4p13	
Reference probe 20960-L29094			6p12	
Reference probe 01232-L00780			10p14	
Reference probe 07404-L27958			12q13	
Reference probe 06719-L06305			15q24	
Reference probe 17436-L21192			16p13	
Reference probe 18498-L23723			19q13	

Appendix 4. Plagiarism form



Final dissertation Clarice Smal 1930283.pdf

ORIGINALITY REPORT

13%	8%	12%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	1%
2	"Abstracts from the 52nd European Society of Human Genetics (ESHG) Conference: Posters", European Journal of Human Genetics, 2019 Publication	<1%
3	Neurofibromatosis Type 1, 2012. Publication	<1%
4	"Abstracts from the 51st European Society of Human Genetics Conference: Posters", European Journal of Human Genetics, 2019 Publication	<1%
5	onlinelibrary.wiley.com Internet Source	<1%
6	orca.cf.ac.uk Internet Source	<1%
7	Cedrik Tekendo - Ngongang, Paul Kruszka. "Noonan syndrome on the African Continent", Birth Defects Research, 2020 Publication	<1%