

https://dspace.org/

Development of a multi-disease targeted next generation sequencing panel to study genetic aetiology of rasopathies

Mudau, Maria Mabyalwa

2024

https://hdl.handle.net/10539/37582

Downloaded from a DSpace repository. https://dspace.org/

DEVELOPMENT OF A MULTI-DISEASE TARGETED NEXT GENERATION SEQUENCING PANEL TO STUDY GENETIC AETIOLOGY OF RASOPATHIES

Maria Mabyalwa Mudau

Student number: 1143805



Supervised by

Dr Nadia Carstens

Professor Amanda Krause

A thesis submitted to the School of Pathology, Division of Human Genetics, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I, Maria Mabyalwa Mudau, declare that this Thesis is my own unaided work. It is being submitted for the Degree of Doctor of Philosophy 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.

Mudau

(Signature of candidate)

6th day of February 2023 in Johannesburg

Dedication

In loving memory of my

Mother-in-law Mrs Phophi Kwinda (January 2021)

Baby Girl Angel Phophi Mawela (May 2022).

Presentations from this study

MudauMM, Krause A and Carstens N. Molecular and clinical profiles of patients with RASopathies: First Next generation sequencing report for South African patients. Oral presentation at the virtual 4th Pathology Research and Development Congress – PathReD 2021, 19-23 August 2021

Mudau MM, Krause A and Carstens N. The use of NGS custom targeted panel and Whole Exome Sequencing to study genetic aetiology of RASopathies. Poster presentation at the Virtual Meeting 2020, American Society of Human Genetics (ASHG), 27-30 October 2020

Mudau MM, Krause A and Carstens N. The use of NGS custom targeted panel and Whole Exome Sequencing to study genetic aetiology of RASopathies. Oral presentation at the Wits Faculty of Health Sciences, 2020 Virtual research day and postgraduate expo, 15 October 2020. *Best oral presentation awarded for this presentation.*

Mudau MM, Nevondwe P, Seymour H, Kerr R, Feben C, Lombard Z, Spencer C, Krause A and Carstens N. Development of a cost-effective multi-disease targeted NGS panel for facial dysostoses RASopathies and Cohesinopathies (FRASC). Oral Presentation. 18th Biennial Congress of the Southern African Society for Human Genetics (SASHG) - 3 -6 August 2019, Century City Conference Centre, Cape Town, South Africa

Mudau MM, Nevondwe P, Seymour H, Kerr R, Feben C, Lombard Z, Spencer C, Krause A and Carstens N. Development of a cost-effective multi-disease targeted NGS panel for facial dysostoses RASopathies and Cohesinopathies (FRASC). 3rd Pathology Research and Development Congress – PathReD 2019, 18- 21July 2019 in Destiny Hotel and Convention Centre, Johannesburg, South Africa. *Awarded best oral presentation in Human Genetics*.

Mudau MM, Nevondwe P, Seymour H, Kerr R, Feben C, Lombard Z, Spencer C, Krause A and Carstens N. Development of a multi-disease targeted next generation sequencing (NGS) panel to investigate the genetic aetiology of RASopathies in South Africa. Poster presentation. Wits Research day and Postgraduate Expo. Wits Research day and Postgraduate Expo, 11 September 2018, Wits Medical School, Johannesburg

Carstens N, **Mudau MM,** Nevondwe P, Seymour H, Kerr R, Feben C, Lombard Z, Spencer C, and Krause A. Developing focused genetic approaches for developmental delay in South Africa. Poster presentation.17th Congress of the South African Society of Human Genetics (SASHG), August 13-16 2017, Durban, South Africa

Abstract

With Next Generation Sequencing (NGS), technologies it is now possible to screen a large number of genes simultaneously through massively parallel sequencing, significantly reducing costs and time generally associated with mutation screening. After an informal survey of the diagnostic needs of the clinicians in the Division of Human Genetics, National Health Laboratory Service (NHLS), it was established that the aetiology of genetic disorders called facial dysostoses, RASopathies and Cohesinopathies (FRASC) could be understood better using NGS. These are developmental disorders that are phenotypically different and are commonly referred to our genetic clinic, currently there is limited genetic testing for these conditions in South Africa. A NGS panel targeting genes associated with the FRASC disorders was designed and optimised. Upon successful optimisation the panel was then utilised to sequence samples from patients presenting with features suggestive of RASopathies. An overall detection rate of 56.6% (34/60) was obtained for all RASopathy patients sequenced in the current study. Detection rate of 46.7% (7/15) for neurofibromatosis type 1 (NF1) patients and 60% (27/45) for patients with non-NF1 RASopathies was obtained. No clinically significant variants were identified in 26 of the 60 patients (43.3%), two of the 26 had a VUS in the *MAP2K1* gene. Seven patients of the panel negatives were successfully sequenced using whole exome sequencing (WES), one patient had a pathogenic variant and the rest had variants of uncertain significance identified. This is the first report profiling mutations and clinical features in patients with RASopathies as a whole in South Africa, compared to a study focusing on NS patients only. The detection rates obtained were comparable to other international studies using NGS, except for detection rate of *LZTR1* and *BRAF1* gene variants observed in Noonan syndrome patients. The (35/60) 58.3% (panel and WES) of patients with a disease causing variant identified in this study now have a molecular confirmation that they previously did not have. Our results show that a targeted panel could be an effective first-line diagnostic testing for RASopathies in SA. The development of the multi-disease targeted panel in the current study has contributed to the design of the 500 gene inherited disease NGS targeted panel (IDP) that is currently being utilised in our facility for diagnostic testing of various monogenic disorders.

Acknowledgments

Doing a part time Doctor of Philosophy (PhD), employed as a full time Senior Medical scientist, as well as being a wife and a mother of two boys was quite challenging. The past five years (2017-2021) have been a journey in all aspects of my life, and I am thankful to God for being fundamental to my life.

I would like to firstly acknowledge and thank the patients and their relatives who enrolled and participated in the current study, without whom this project would have been impossible. I believe the outcomes of this study will benefit the management of their condition and improve their health care.

I want to thank my supervisors Dr Nadia Carstens and Professor Amanda Krause for their invaluable support and guidance throughout my study which led to the successful completion of this project. They both recognised my passion and dedication and I will forever cherish the opportunity I was given to do this project. I would like to extend my gratitude to Ms Patracia Nevondwe and Ms Heather Seymour and the Facial dysostoses, RASopathies and Cohesinopathies (FRASC) working group for their advice and support. I would like to also thank my proof-readers, Ms Nkateko Mayevu and Ms Merlyn Glass.

The success of this project was made possible by the research funding received from The National Research Foundation –Thuthuka Grant (NRF-TTK), The National Health Laboratory Service Research Trust (NHLSRT) as well as the University of the Witwatersrand Faculty Research Council Grant (FRC). I would also like to thank the University of the Witwatersrand and National Health Laboratory Service staff bursaries for funding my tuition fees, I am truly grateful to have had my degree accepted.

Special gratitude to the unending support from my husband Mashudu, my sons Mpho and Pfano, and my sisters Rofhiwa and Lydia who have all been my pillar of strength throughout my PhD.

This thesis is dedicated to my parents Mrs Jane Ngwako Mudau and Mr Patrick Mudau

Table of Contents

Declaration…… i

Dedication……. ii

Presentations from this study……. iii

Abstract……. iv

Acknowledgements…… v

Table of contents…… vi

List of Figures…… xi

List of Tables……. xv

List of abbreviations……. xviii

Chapter 1: Introduction…….. 1

1. Introduction…… 2

1.1.Genomic research in Africa……. 3

1.1.1. Population Genetics…… 3

1.1.2. Molecular basis of genetic disorders…… 4

1.2.Next generation sequencing technology……. 5

1.2.1. Whole genome sequencing……. 7

1.2.2. Whole exome sequencing…… 9

1.2.3. Targeted gene panels…….. 10

1.2.4. Analysis of NGS data …….. 11

1.2.5. Interpretation of NGS data…… 16

1.3. Cost considerations and diagnostic implementation of NGS…….. 16

1.3.1. Diagnostic utility…… 17

1.3.2. Sensitivity, specificity and efficacy……. 18

1.3.3. Cost implications……. 20

1.4. RASopathies……..…….. 21

1.4.1. Genetic aetiology of RASopathies……..... 21

1.4.2. Disease mechanism of RASopathies…… 23

1.4.3. Noonan syndrome……. 25

1.4.4. Neurofibromatosis type 1……..... 27

1.4.5. Neurofibromatosis-Noonan syndrome……..... 29

1.4.6. Noonan syndrome with multiple lentigines…….. 30

1.4.7. Capillary malformation-arteriovenous malformation syndrome……. 30

1.4.8. Costello syndrome…… 31
1.4.9. Cardio-facio-cutaneous syndrome…… 31
1.4.10. Legius syndrome…… 32
1.4.11. Neurofibromatosis type 2…….... 32
1.4.12. Schwannomatosis…… 33
1.5. Problem Statement……..34
1.6. Aims…….. 35
1.7. Objectives…….. 35

Chapter 2: Subjects and Methods…….. 36

2.1.Patient recruitment…… 37
2.1.1. Ethics…… 37
2.1.2. Patients and relatives…… 37
2.2. DNA extraction……. 38
2.3. DNA quality control……. 39
2.4. Targeted gene panel ……. 39
2.4.1. Custom-design of the targeted gene panel…… 39
2.4.2. Library preparation and target enrichment……... 41
2.4.3. Sequencing on the Illumina MiSeq…….. 43
2.4.4. Quality control of NGS results…… 44
2.4.5. Targeted panel data analysis and interpretation…… 45
2.5. Costing analysis……. 46
2.6.Whole Exome Sequencing using Ion-Torrent ThermoFisher technology……. 47
2.6.1. Library preparation…….. 47
2.6.2. Template preparation and chip loading…… 46
2.6.3. Sequencing using the Ion S5™ GeneStudio……. 48
2.6.4. Whole Exome Sequencing data analysis and interpretation…… 48
a) Ensembl Variant effect predictor……. 50
b) Moon diploid…… 50
c) Developmental disorder G2P pipeline…… 50
d) Interpretation using ACMG-AMP guidelines……. 51
2.7.Sanger Sequencing validations…… 52
Chapter 3: Results…….. 53

3.1.Demographics of patients recruited…….…… 54
3.2. Design and optimization of the targeted gene panel…….……... 54
3.2.1. Costing Analysis…… 57
3.2.2. Utilising FRASC panel to sequence RASopathy samples…… 59
3.2.3. Quality control of NGS data…….. 59
a) *PTPN11*:c.923A>G variant…….. 60
b) *NF1:c.27G>A* variant…… …… 62
3.3. Genetic variants detected using the FRASC targeted gene panel……..……66
3.3.1. Variants identified in the *PTPN11* gene……. 66
3.3.2. Variants identified in the *BRAF* gene……. 70
3.3.3. Variants identified in the *LZTR1* gene…….75
3.3.4. Variants identified in the *SOS1* gene…….. 79
3.3.5. Variants identified in the *RAF1* gene……. 80
3.3.6. Variants identified in the *NRAS* and *MAP2K2* genes…… 81
3.3.7. Variants identified in the *NF1* gene…….. 82
3.3.8. Variant of uncertain significance in the *MAP2K1* gene…… 86
3.3.9. Detection rate per clinical diagnosis and gene variant ….…… 87
3.3.10. Segregation analysis…….. 89
3.4.Additional testing on patients who tested negative on the targeted panel…... 90
3.4.1. Whole exome sequencing…… 91
a) Patient FRASC24……..……. 92
b) Patient FRASC55……..….. 93
c) Patient FRASC56…… 95
d) Patient FRASC58…… 97
e) Patient FRASC80…… 99
f) Patient FRASC96…… 100
g) Patient FRASC97…… 103
Chapter 4: Discussion……..... 105
4.1. Summary of study rationale…… 106
4.2. Design and Optimization of the FRASC targeted panel…… 107
4.2.1. Cost-effective strategy of using the panel…….. 107
4.2.2. Utilising panel to sequence RASopathies…….. 109
4.2.3. Quality control of the NGS data……. 109
a) False positive variants…….. 109

b) Quality metrics……. 110
c) Variant allele frequency threshold…….. 110
d) Reducing false positive calls in future…… 111
4.3.Genetic profiles obtained……. 112
4.3.1. Detection rate achieved in the current study..…… 112
4.3.2. Pathogenic and likely pathogenic variants in *PTPN11* gene…… 113
4.3.3. Pathogenic and likely pathogenic variants in *BRAF* gene…… 115
4.3.4. Pathogenic and likely pathogenic variants in *LZTR1* gene……. 117
4.3.5. Pathogenic variants in *SOS1* gene…… 118
4.3.6. Pathogenic variants in *RAF1* gene…….. 119
4.3.7. *NRAS* and *MAP2K2* pathogenic variants……. 119
4.3.8. Variants detected in NF1 patients……..120
4.3.9. Variants of uncertain significance…… 121
4.3.10. Benign variation in current study……. 122
4.3.11. Detection rate per clinical diagnosis…… 123
a) Noonan syndrome…….. 123
b) Noonan syndrome with multiple lentigines…… 125
c) Cardio-facio-cutaneous syndrome……. 125
d) Schwannomatosis…… 126
e) Neurofibromatosis type 1…… 126
4.3.12. Segregation analysis of RASopathy-causing variants…….. 128
4.4. Whole exome sequencing……..... 129
4.4.1. Detection rate achieved by WES data…….. 129
4.4.2. Variants of uncertain significance in the WES data…… 131
4.5. Targeted gene panel vs WES……. 133
4.6.Diagnostic utility of findings in the current study…….. 134
4.7.Limitations of current study…… 135
4.8.Recommendations and future directions……. 135
4.9.Conclusions……..... 137
4.10. Appendices……. 138
Appendix 1.1: Ethics clearance obtained by Mudau MM (Current PhD) from the University of the Witwatersrand, Medical Human Research Ethics Committe……... 138

Appendix 1.2: Ethics clearance obtained by Dr Nadia Carstens (Current PhD, Supervisor) from the University of the Witwatersrand, Medical Human Research Ethics Committee…….. 139
Appendix 1.3: Ethics clearance obtained by Dr Nadia Carstens (Current PhD, Supervisor) from the University of Pretoria, Faculty of Health Sciences, Research Ethics Committee……...140
Appendix 2: Clinical RASopathy data sheet…… 141
Appendix 3: Full list of demographics of RASopathy patients recruited in the current study…….……..... 143
Appendix 4: FRASC panel design report from Agilent SureDesign software. 145
Appendix 5: Patient quality control report example with duplicate information from Agilent SureCall software……. 148
Appendix 6: Primers used for Sanger sequencing ……. 151
Appendix 7.1: Summary of data obtained on Ion-540™ chip 4, WES run 2…153
Appendix 7.2: Assessing WES run report example from the Torrent _Suite software…… 154
Appendix 8: ACMG criteria for classifying pathogenic variants……. 155
Appendix 9: Turnitin report…….. 156
4.11. List of websites…….…… 157
4.12. References……. 158

List of Figures

Figure 1.1: Comparison of the sequencing approaches using NGS technologies; summarises the whole genome sequencing, whole exome sequencing and targeted gene panel approaches and demonstrate key differences. Figure obtained from (Han and Lee, 2020**)**

Figure 1.2: Three stages of data analysis performed during next generation sequencing

Figure 1.3: RAS/MAPK pathway showing some of the different genes associated with the RASopathies in different colours and where they act in the pathway. Figure obtained from Rauen,(2013)

Figure 1.4: Phenotypic features of the main RASopathies and their clinical overlap. The figure is modified from Stevenson et al., (2016)

Figure 1.5: Diagram showing some clinical features of Noonan syndrome typically seen in a newborn baby. Obtained from Bhambhani and Muenke (2014)

Figure 1.6: Diagram of an adolescent with Noonan syndrome. Obtained from Bhambhani and Muenke (2014)

Figure 1.7: Common clinical features of NF1 patient, cutaneous lesions (arrows); multiple café au lait lesions (CAL) (A); freckling and two CAL (B); discrete neurofibromas (C); pigmented skin lesions (epidermal nevus) misdiagnosed as CAL and freckling in a non-NF1 patients (D), obtained from (Rodrigues et al., 2014)

Figure 1.8: Skin with lentigines in individuals with NSML at various ages. Obtained from Sarkozy, Digilio and Dallapiccola (2008)

Figure 2.1: Patient recruited, their clinical diagnosis and testing performed

Figure 2.2: Summary of the SureSelect^{QXT} library preparation and target enrichment workflow, obtained from (https://www.agilent.com/cs/library/usermanuals/public/G9681-90000.pdf)

Figure 2.3: Summary of analysis of data generated by the targeted gene panel

Figure 2.4: Summary filtering strategy used for variants generated by whole exome sequencing

Figure 2.5: Summary of the whole exome sequencing process, from wet-lab to analysis

Figure 3.1: Sequence analysis viewer quality graphs, obtained for FRASC run1 on the MiSeq instrument

Figure 3.2: IGV image and Sanger sequencing electropherogram for FRASC91, one of the false positive variants for identified during data analysis (not observed on Sanger sequencing and also IGV showing uneven allele fraction)

Figure 3.3: IGV image and Sanger sequencing electropherogram for FRASC66, showing a true positive variant (*PTPN11*: c.923A>G) identified and confirmed on Sanger sequencing.

Figure 3.4: IGV image and Sanger sequencing electropherogram for FRASC67, showing a *NF1*:c.27G>A true positive variant identified and confirmed on Sanger sequencing.

Figure 3.5: Showing FRASC83 sample where the *NF1*:c.27G>A variant was not confirmed on Sanger sequencing and when the another panel was utilised for sequencing.

Figure 3.6: Sanger sequencing electropherograms of FRASC2 (A) and FRASC18 (B) showing the *PTPN11*:c.922A>G variant observed in the probands and their mothers

Figure 3.7: Sanger sequencing electropherograms of FRASC20 showing the *PTPN11*:c.182A>G variant observed in both child and mother.

Figure 3.8: Sanger sequencing electropherogram of the NSML patient, showing the *PTPN11*: c.1381G>A (p.A461T) de novo variant and the parents (homozygous wild type)

Figure 3.9: Protein Tyrosine Phosphatase protein domains (SHP) containing the N and C terminals (N-SH2 and C-SH2) as well as the protein tyrosine phosphate (PTP) domain showing where the *PTPN11* variants identified in the current study cluster.

Figure 3.10: FRASC4 (A) and FRASC64 (B) with a clinical diagnosis of NS both carrying the same *BRAF*:c.770A>G (p.Q257R) pathogenic variant

Figure 3.11: Sanger sequencing electropherogram of the three patients A (FRASC4), B (FRASC64) and C (FRASC73), showing the *BRAF*:c.770A>G variant and the parents (having homozygous wild-type variant)

Figure 3.12: Sanger sequencing electropherogram of the two patients, showing the *BRAF*:c.722C>T (A) and *BRAF*:c.736G>C (B) variants and both variants not observed in the mothers of the probands.

Figure 3.13: *BRAF* variants obtained in the current study are mostly clustered in exon 6 of the conserved region 1 (CR1) of the Ras-GTP-binding self-regulatory domain.

Figure 3.14: FRASC69 (A) with a clinical diagnosis of NS carrying the *LZTR1*:c.848G>A (p.R283Q) likely pathogenic variant and FRASC15 (B) with a clinical diagnosis of NS carrying the previously described *LZTR1*: c.730T>C (p.S544P) likely pathogenic variant

Figure 3.15: Sanger sequencing electropherogram of the three patients, showing the *LZTR1*: c.848G>A, c.690C>G and c.730T>C variants in A (FRASC69), B (FRASC21) and C (FRASC15). Variant *LZTR1*: c.848G>A was observed in both the FRASC69 patient and mother (A).

Figure 3.16: Pedigree and Sanger sequencing electropherogram of FRASC62, showing the disease-causing variant segregating through the family with Schwannomatosis
Figure 3.17: *LZTR1* variants obtained in the current study are clustered exons 8 and 9 within Kelch 3 and 4 protein domains, the variants are monoallelic and associated with autosomal dominant Noonan syndrome (Motta et al., 2019: 1; Pagnamenta et al., 2019).
Figure 3.18: FRASC14 with a clinical diagnosis of NS carrying *SOS1*:c.806T>C (p.M269T) pathogenic variant.
Figure 3.19: Sanger sequencing electropherogram of the two patients, showing the c.806C>T, identified in exon 6 (A) and c.1649T>C identified in exon 10 (B) of the *SOS1* gene. Variant *SOS1*: c.806C>T was observed in both FRASC14 patient and her father (A)
Figure 3.20: Sanger sequencing electropherogram of FRASC87 and FRASC17, showing the c.770C>T,(A) and c.781C>A (B) both identified in exon 7 of the *RAF1* gene.
Figure 3.21: Sanger sequencing electropherogram of FRASC22 and FRASC5, showing the *MAP2K2*: c.169T>G (A) and the *NRAS*:c.173C>T (B) variants.
Figure 3.22 Sanger sequencing electropherogram confirming two frame-shifts variants (c.496_497 delGT and c.5204_5205delAG) identified in the *NF1* gene of FRASC44 (A) and FRASC53 (B) where the variant was not observed in the father.
Figure 3.23: Sanger sequencing electropherogram confirming a nonsense variant, c.3721C>T identified in the *NF1* gene of FRASC45 child. Both parents had the homozygous wild type variant
Figure 3.24: Sanger sequencing electropherogram confirming two variants identified in this patient. One nonsense pathogenic *NF1*:c.6772C>T (A) and one missense VUS c.215G>A identified in the *MAP2K1* gene. The *MAP2K1* VUS was also observed in the mother (B)
Figure 3.25: *NF1* variants obtained in the current study are mostly clustered outside the functional domains. The type of variants obtained are mostly nonsense and frameshifts causing premature termination**.**
Figure 3.26: Sanger sequencing confirmation was performed for the *MAP2K1*:c.215G>A variant and was found in all three patients (Figure 3.26). Variants identified in FRASC98 father and FRASC42 mother.
Figure 3.27: Further testing done in 24 Noonan syndrome and Neurofibromatosis type 1 patients in whom no variants were identified using the FRASC targeted gene panel
Figure 3.28: IGV image of FRASC 24 showing the region of *RHOBTB2*: c.617C>G heterozygous variant.

Figure 3.29: IGV image of FRASC55 showing *DNAH9:*c.7169A>C variant in homozygous state.
Figure 3.30: IGV image of the *RYR3*: c.10693C>G heterozygous variant at 51x read depth
Figure 3.31: IGV image and Sanger sequencing electropherogram of the *CDK13*: c.2149G>A variant at 114x read depth and variant was observed using Sanger sequencing.
Figure 3.32: IGV image of the *MYH3*:c.3131A>T and *MYH3*:c.3056A>C heterozygous variants at 90x and 38x read depth with ~60% and 40% allele fractions. These variants were observed on Ensembl VEP, Moon diploid as well as the DDD G2P pipeline. Both these variants have not been reported on ClinVar
Figure 3.33: IGV image and Sanger sequencing electropherogram of the *TRPS1*:c.3403C>G heterozygous variant at 60x read depth and variant was observed using Sanger sequencing in both child and father.

List of Tables

Table 1.1: Genes associated with different RASopathies and their relative mutation contribution in the reported literature.

Table 1.2: Minor and major clinical criteria for the diagnosis of Noonan syndrome (Carcavilla *et al.*, 2020)

Table 2.1: Genes included in the custom-designed NGS targeted panel, their associated clinical phenotypes and ClinGen association classification (Rehm et al., 2015)

T**able 2.2:** Genes with 100% coverage after probe tiling parameters were modified

Table 2.3: Summary of filtering parameters used to shortlist targeted gene panel variants

Table 3.1: Summary of demographics of recruited patients

Table 3.2: FRASC samples used in the first panel validation run showing Q-scores and read depth obtained

Table 3.3: Costs of tests done prior to the current study

Table 3.4: Showing estimations of prices if Sanger sequencing were to be used to analyse all the exons of the genes associated with RASopathies

Table 3.5: Quality metrics obtained using MiSeq reagent v2 kits

Table 3.6: Patients with skewed allele fraction at *PTPN11*: c.923A>G (p.N308S) variant site and FRASC66 (highlighted) showing true positive variant.

Table 3.7: Eight patients with *NF1:* c.27G>A (p.W9X) false positive variant, IGV coverage as well as the respective allele fraction and FRASC67 (highlighted) showing the true positive variant.

Table 3.8: Pathogenic variants identified in the *PTPN11* (NM_002834.3) gene

Table 3.9: Typical clinical features of Noonan syndrome observed in patients with *PTPN11* variants

Table 3.10: List of the pathogenic variants identified in the *BRAF* (NM_004333.4) gene

Table 3.11: Summary of clinical features of patients with *BRAF* pathogenic variants

Table 3.12: List of likely pathogenic and pathogenic variants identified in the *LZTR1* (NM_006767.4) gene

Table 3.13: Summary of clinical features of Noonan syndrome patients with *LZTR1* variants observed in this study

Table 3.14: Two variants identified in the *SOS1* (NM_005633.3) gene.

Table 3.15: Two variants identified in the *RAF1* (NM_002880.3) gene

Table 3.16: Variants identified in the *NRAS* and *MAP2K2* genes

Table 3.17: Pathogenic variants identified in the *NF1* (NM_001042492.2*)* gene in patients with clinical diagnosis of NF1
Table 3.18: Summary of clinical features of patients with a pathogenic *NF1* variant.
Table 3.19: Summary of the detection rate in the RASopathy genes where clinically significant variants were identified in our cohort of 60 patients.
Table 3.20: Comparison of our frequency of clinically significant variants in a gene per RASopathy to the frequencies in existing literature
Table 3.21: Variants that were observed in probands and their relatives
Table 3.22: Summary of the run statistics obtained during WES, with quality of the reads, depth coverage and number of variants obtained per sample.
Table 3.23: Shortlisted variants for FRASC24, with the associated disease, mode of inheritance, zygosity, coverage depth, ACMG classification as well as interpretation
Table 3.24: Shortlisted variants for FRASC55, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification
Table 3.25: Short-listed variants for FRASC56 with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification.
Table 3.26: Putative disease-causing variants for FRASC58, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification
Table 3.27: Putative disease-causing variants for FRASC80, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification
Table 3.28: Shortlisted variants for FRASC96, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification
Table 3.29: Putative disease-causing variants for FRASC97, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification

List of abbreviations

µl -	microliter
A.I. -	Artificial intelligence
ACMG-AMP -	The American College of Medical Genetics and Genomics - the Association for Molecular Pathology
AD -	Autosomal dominant
AGVP-	The African Genome Variation Project
AR -	Autosomal recessive
B -	African ancestry
BAM -	Binary Alignment Map
bp -	base pair
CADD-	Combined Annotation Dependent Depletion
CdLS -	Cornelia de Lange syndrome
CFC -	Cardiofaciocutaneous syndrome
CHBAH -	Chris Hani Baragwanath Academic Hospital
CM-AVM -	Capillary malformation-arteriovenous malformation syndrome
CMJAH -	Charlotte Maxeke Johannesburg Academic Hospital
CNV -	Copy Number Variation
CS -	Costello syndrome
DDD -	Developmental Disorder Genes
DD-G2P -	Development Disorder Genotype - Phenotype Database
dsDNA BR -	double-stranded DNA Broad Range
dsDNA HS -	double-stranded DNA High Sensitivity
DSPF -	Downslanted palpebral fissures
EDTA -	Ethylenediaminetetraacetic acid

EF -	Epicanthic fold
ExAC -	The Exome Aggregation Consortium
F -	Female
FATHMM -	Functional Analysis through Hidden Markov Models
FRASC -	Facial dysostoses, RASopathies and Cohesinopathies
Gbp -	Giga-base pairs
gDNA -	genomic Deoxyribonucleic Acid
GnomAD -	The Genome Aggregation Database
GRCh37 -	Genome Reference Consortium Human Build 37
GRCh38 -	Genome Reference Consortium Human Build 38
H3Africa -	The Human Heredity and Health in Africa
HET -	Heterozygous
HGDP -	Human Genome Diversity Project
HGMD -	The Human Gene Mutation Database
HGVS -	Human Genome Variation Society
HOCM -	Hypertrophic obstructive cardiomyopathy
HOMO -	Homozygous
HPO -	Human Phenotype Ontology
HPO -	The Human Phenotype Ontology
HREC -	Human Research Ethics Committee
ID –	Intellectual diability
IDP -	Inherited disease targeted gene panel
IGV -	Interactive genome viewer
Indels -	Insertion–deletion mutations

JHB -	Johannesburg
Kbp -	Kilo-base pair
LMIC -	Low and Low Middle Income Countries
LS -	Legius Syndrome
M -	Male
MAF -	Minor Allele Frequency
MLPA -	Multiplex Ligation-dependant Probe Amplification
NCBI -	National Center for Biotechnology Information
NEB -	New England Biolabs
NF1 -	Neurofibromatosis type 1
NF2 -	Neurofibromatosis type 2
NGS -	Next-Generation Sequencing
NHLS -	National Health Laboratory Service
NS -	Noonan syndrome
NSML -	Noonan Syndrome with Multiple Lentigines
POLYPHEN -	Polymorphism Phenotyping
PTA -	Pretoria
PTP -	Protein tyrosine phosphatase
Q30 -	Quality score of 30
QC -	Quality Control
RAS/MAPK -	Ras/mitogen activated protein kinase
RMH -	Rahima Moosa Mother and Child Hospital
SAHGP -	The Southern African Human Genome Project
SAV -	Sequencing Analysis Viewer

SBAH -	Steve Biko Academic Hospital
SHP -	Src homology domain phosphatase
SIFT -	Sorting Intolerant From Tolerant
SNV -	Single-nucleotide variant
TCS -	Treacher Collins syndrome
TE -	Tris-EDTA
UBAM -	Unaligned/Unmapped Binary Alignment Map
VAF -	Variant allele frequency
VCF -	Variant Call Format
VEP -	Ensembl Variant Effect Predictor
VUS -	Variant of uncertain/unknown significance
W -	Caucasian
WANNOVAR -	Wed-based ANNOtate VARiation
WES –	Whole exome sequencing
WGS -	Whole genome sequencing
PS -	Pulmonary stenosis
ASD -	atrial septal defect
PDA -	Patent ductus arteriosus
VSD -	ventricular septal defect

CHAPTER 1

Introduction

1. Introduction

It is critical to diagnose genetic or rare disorders timeously and accurately, as this will enable a better understanding of the prognosis, improve availability of directed management and provide accurate recurrence risk information. Currently, in South Africa and the African continent as a whole, genetic services in the State hospitals are very limited; there are few tests to confirm a molecular diagnosis for many monogenic disorders that are clinically diagnosed. This may lead to misdiagnosis of genetic disorders, thus affecting the way the patient is managed.

Next generation sequencing (NGS) technologies have allowed a breakthrough in medical sciences in recent years. The current study aimed to utilise NGS technology to address the challenges of diagnosis in a resource-constrained environment with the ultimate goal being to develop an affordable approach to identifying disease-causing variants associated with various monogenic disorders in South African patients. The Division of Human Genetics, National Health Laboratory Service and The University of Witwatersrand, Johannesburg, South Africa, is one of the largest referral centers for genetic testing in the country and delivers most of its services in the public healthcare system. Departmental resources need to be utilised cost-effectively so that patients can get access to the care that they need without unnecessary burdens on limited public health care funds, as is the case in most low- and middle-income countries.

At the start of this study, an informal survey was conducted amongst seven medical geneticists from the Division of Human Genetics to investigate which disorders required urgent research in order to generate information that could improve clinical services and assist in test development. The group indicated that RASopathies, cohesinopathies and facial dysostoses represented some of the most common referrals to genetic clinics, but most of these patients could unfortunately not be offered any confirmatory molecular testing.

The strategy employed in the current study was to design a NGS targeted panel that contains genes associated with multiple monogenic disorders, so that more than one genetic disorder can be tested and studied simultaneously; this is the key benefit of NGS technology. Genes for multiple disorders could be included in the single targeted gene panel and analysed selectively based on the phenotype. This would allow batching, thus maximising sequencing capacity of the panel and saving costs. At the inception of this current study, NGS technologies such as

targeted gene panels and whole exome sequencing were costly. Our approach allowed for an more affordable way to utilise this technology as we were able to generate mutation profiles of three groups of monogenic developmental disorders simultaneously using a single NGS panel. The genes included in the targeted panel were associated with RASopathies which is the focus of the current study, as well as cohesinopathies and facial dysostoses. Variant analysis of the latter two formed part of two separate studies (Nevondwe, 2019; Seymour, 2019).

In this introductory chapter, genomic research in Africa is described; followed by, the description of next generation sequencing (NGS) technologies (since this was the first study using this technology in our setting a deep understanding of the technicalities and analysis tools is required for correct data interpretation). The cost implications and diagnostic utility of NGS in the described and lastly the types of genetic disorders that were studied using NGS in the current project – RASopathies are then outlined, followed by the problem statement, aims and objectives of the study.

1.1.Genomic research in Africa

1.1.1. Population Genetics

Africa is a continent rich in genetic diversity due to the complex history of the population and variation in environmental factors. This contributes to wide genetic and phenotypic differences in people of African ancestry compared to other global populations (Bergström et al., 2020). It is crucial to undertake genomic studies from African populations in order to understand the uniqueness of African genomes and how they contribute to health and disease. Genetic diversity studies on African populations are essential in uncovering patterns of genetic variation that are influenced by natural selection and migrations from or into Africa (Maxmen, 2020). In early sequencing projects such as the 1000 Genomes Project (https://www.genome.gov/27528684/1000-genomes-project) there was a large amount of genetic diversity observed in African compared to non-African populations (Devuyst, 2015a). This led to various genomic projects that aimed to contribute genetic data from African genomes; these included initiatives such as the African Genome Variation Project (AGVP) (Gurdasani et al., 2015), The Human Genetic Diversity Project (HGDP) (Cavalli-Sforza, 2005; Bergström et al., 2020), The Human Heredity and Health in Africa (H3Africa Consortium) (H3Africa Consortium et al., 2014)_as well as the Southern African Human Genome Programme (SAHGP) (Choudhury et al., 2017a).

These projects are teaching us about the importance of defining normal variation in populations in order to understand genetic diversity and how it contributes to disease susceptibility. With all these projects, there is still limited sampling of African populations, sampling of more individuals will make the characterization of genetic diversity more efficient and precise. Detailed analysis of the African genome will assist in addressing factors such as human origins, migrations and gene flow from one population to the other. Furthermore factors such as natural selection, recombination and variants occurring at specific parts of the genome influence the levels of genetic variability between human populations. Understanding these regions in the genome will assist in identifying genetic variation that contributes to disease susceptibility and other phenotypic variations (Choudhury et al., 2020; Ramsay, 2022). Globally European and Asian ancestry populations dominate genetic research; this paucity in research in Africa makes it difficult to unpack the molecular basis of genetic disorders in African populations.

1.1.2. Molecular basis of genetic disorders

Most monogenic disorders including RASopathies in African patients are only diagnosed clinically as molecular testing is generally unavailable or very limited (Kromberg, Sizer and Christianson, 2013). Misdiagnosis of these patients may occur due to the lack of molecular confirmation, which presents an issue when offering clinical services and genetic counselling. This is caused by limited research regarding the genetic structure of African populations and the molecular basis of many genetic conditions (May et al., 2013; Krause, 2019a; Baynam et al., 2020). This paucity of genomic research has contributed to the low detection rate and limited genetic services on the African continent, further exacerbated by a lack of resources and relevant technologies (Kamp, Krause and Ramsay, 2021).

Recent research has shown that more diversity studies in genomics are needed if low and middle-income countries (LMICs) are to benefit from genomic medicine. Once the genetic information is available, its contribution to monogenic disorders will be better understood in these previously understudied population (Popejoy and Fullerton, 2016; Baynam et al., 2020). Other factors that impede monogenic disease research implementation in LMICs include economic and cultural challenges, government policies regarding genomic research, inadequate research infrastructure and a limited number of genomic experts (Kingsmore et al., 2012; Tekola-Ayele and Rotimi, 2015a; Kamp, Krause and Ramsay, 2021). It is important for African-focused genomic disease research and clinical care infrastructure to be developed, as this will contribute to the development of appropriate health systems thus improving the

diagnosis of genetic or rare disorders timeously and accurately. However, it is challenging to obtain funding for genomic research by African researchers, because genomic research competes for resources with communicable diseases (Musanabaganwa et al., 2020). This has been exacerbated recently by the COVID19 pandemic. There has been some success through grant funding and African researchers and scientists are beginning to introduce technologies such as NGS to generate African specific genetic data that will be useful in diagnosing patients and then designing treatments and therapies that are unique to patients of African ancestry (Maxmen, 2020).

1.2. Next Generation Sequencing technology

Massively parallel sequencing or deep sequencing are terms used to describe the next generation sequencing (NGS) technology. This technology is used to sequence large parts of the genome rapidly or target specific regions more deeply, depending on what users want to achieve. NGS technology, since its inception in the early 2000s (Reinartz et al., 2002). has facilitated breakthroughs in medical science, unpacking the molecular basis of disease and revolutionising molecular medicine (Hu et al., 2021). It is now possible to screen a large number of genes simultaneously through massively parallel sequencing, significantly reducing costs and time associated with mutation screening (Adams and Eng, 2018).

Prior to NGS, Sanger Sequencing was used as a first line screening method for most monogenic disorders. Sanger sequencing only targets certain areas of the gene, hot spot regions or one gene at a time. It is laborious and time consuming to target many genes at the same time using Sanger sequencing which is why the introduction of NGS has improved how we study monogenic disorders. NGS is superior due to is its ability to sequence the same fragments many times in a parallel reaction allowing for deeper sequencing. The sequencing depth of each fragment relies on the kind of NGS performed and the question the researcher is seeking to answer. Mostly the minimum sequencing depth of 30 reads (30x) per target region is considered sufficient in most NGS procedures used to screen genetic data for rare variants (Petrackova et al., 2019a).

There are various manufacturers of NGS technology platforms and consumables but the basic steps involved in the NGS procedure are similar across all platforms and are based on a sequencing by synthesis principle. The DNA is fragmented into shorter fragments because this sequencing by synthesis technology only analyses short reads of less than 500bp in a process

termed library preparation. The fragments are then attached to the adapters and unique barcodes are added during target enrichment or template preparation; this is done so that each sample can be easily identified during analysis. The sequencing by synthesis principle relies on the normal incorporation of labelled nucleotides occurring, while being detected or imaged during sequencing and without being blocked allowing the next nucleotide to be incorporated, thereby generating newly synthesized DNA strands (Buermans and den Dunnen, 2014; Slatko, Gardner and Ausubel, 2018). Once the sequencing is complete, the short reads are attached back together to enable analysis.

The two common NGS platforms available are Illumina (https://www.illumina.com/) and Ion-Torrent (https://www.thermofisher.com/). They both have built in software that enables most processes of NGS sequencing by synthesis technology to happen in the machine up until the data is produced and ready to be analysed (Lahens et al., 2017). The difference between these two platforms is that the Illumina uses fluorescent labelling to read bases in a nucleotide sequence and the Ion Torrent uses semiconductor technology, which is based on pH measurements to read the nucleotide sequence. Other differences are that Illumina generates reads of the same length and the Ion Torrent generates read sequences of variable length; furthermore Illumina can sequence from both ends of the fragment, paired-ends whereas the Ion-torrent cannot (Slatko, Gardner and Ausubel, 2018; Kanzi et al., 2020).

NGS technology has been shown to increase the detection rate for many developmental disorders for which genetic testing was not previously available (Xue et al., 2015). Next generation sequencing technology (summarised in Figure 1.1) may be used to study the entire genome of a human, called whole genome sequencing (WGS). The NGS technology can be utilised to sequence the coding exons of the human genome, in whole exome sequencing (WES) and also can make use of targeted panels in a more directed way, where only genes or exons of interest are sequenced and investigated (Han and Lee, 2020b).

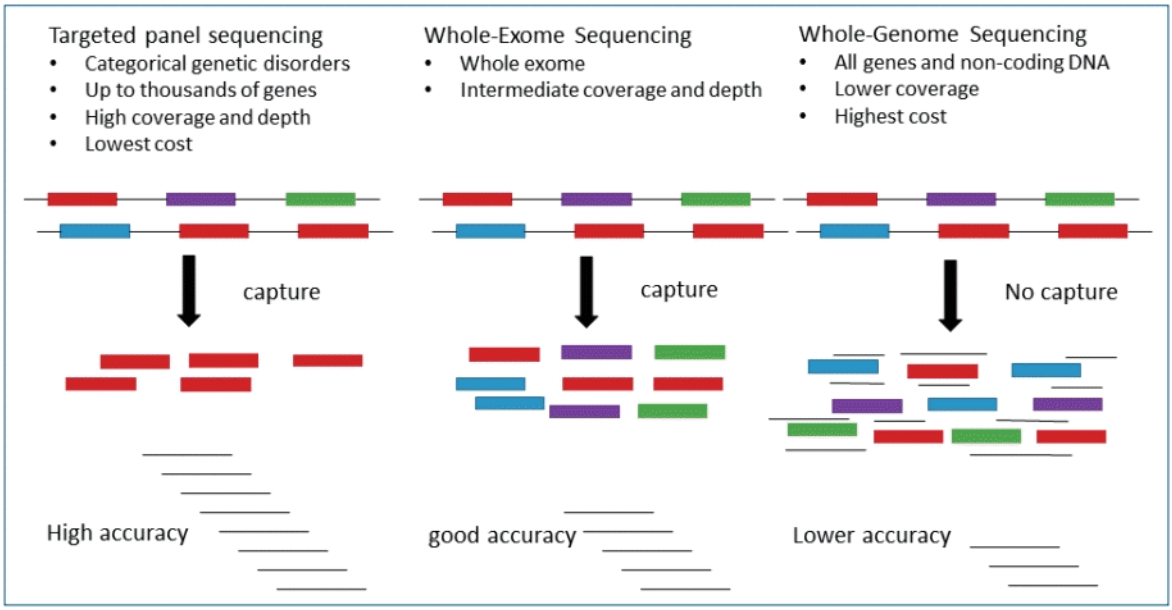


Figure 1.1: Comparison of the sequencing approaches using NGS technologies; summarises the whole genome sequencing, whole exome sequencing and targeted gene panel approaches and demonstrate key differences. Figure obtained from (Han and Lee, 2020**)**

The developed countries have been utilising this technology for more than ten years to generate genetic data for various monogenic disorders. African genetic data can be rapidly generated using this technology and this will enable us to understand our health and genetic conditions better (Maxmen, 2020). Combining NGS data with a detailed clinical report and family history is a powerful approach towards providing improved and appropriate healthcare and management of genetic disorders. For example, when NGS data is analysed, it is focused mainly on the clinical phenotype of the patient and identifying a genetic variant that is associated with the phenotype. There is a need to understand when and how to use NGS technology and data handling, because the likelihood of identifying novel variants of uncertain significance (VUS) is high in WGS and WES genetic data, especially in understudied populations such as those of African ancestry (Baynam et al., 2020).

1.2.1. Whole Genome Sequencing

This procedure is used to sequence the entire human genome at single nucleotide level resolution and covers both coding and noncoding regions. The clinical utility of WGS is improving as the technology improves (Schwarze et al., 2018; Palmer et al., 2021). This technology is capable of detecting a wide spectrum of common and rare genetic variants across the entire genome including single nucleotide variants (SNVs), insertion and deletion variants

(indels), copy number alterations, and translocations (Chrystoja and Diamandis, 2014; Lappalainen et al., 2019). These kind of data contribute to rare disease research and clinical applications.

The costs of WGS per sample is reducing drastically, with the estimated costs of ~$1000. This is because most laboratories in the developed countries are undertaking WGS for research and for diagnostics of rare genetic disorders with an uncertain cause(Austin-Tse et al., 2022). The disadvantage of WGS is complex analysis of generated data that is time consuming and it also costly as it requires highly qualified bioinformaticians to analyse and interpret the data (Wainschtein et al., 2022; Young, 2022).

The comprehensive nature of WGS leads to huge amounts of genetic data being generated. This amount of data needs good storage capacity and high internet speed facilities that would enable bioinformatics pipelines to be utilised efficiently during analysis. Also, it has been observed in various datasets, that the number of genomic variants generated from African individuals is 20% higher than that of Caucasians; this calls for the need for relevant bioinformatics pipelines that will not omit any disease causing variants when African genomes are analysed (Choudhury et al., 2017a).

In recent years, WGS has been shown to have improved detection rates in Mendelian disorders. A study by Posey (2019) identified a 22% improved detection rate in previously undiagnosed patients. Another study argues that WGS could be a first tier diagnostic test as they have observed a 41% detection rate with the discovery of several disease genes and novel variants in deep intronic regions and copy number variants (CNVs) that would not have been observed using WES and gene panels technologies (Lionel et al., 2018). WGS has been shown to add value in confirming some diagnoses where first line testing such as WES and targeted gene panels did not identify any causative variants. This technology showed an increased detection rate of where patients had prior WES sequencing, 37.5% of these had complex structural variations and 68% detection rate was obtained in patients that were initially analysed using a multi-gene panel, and variants were identified in genes that were not in the panel (Palmer et al., 2021). Although there is observed benefit of utilising WGS in a diagnostic setting in laboratories from developed countries, implementation in a low resource setting as a diagnostic tool may be challenging considering the skills needed for data analysis. It would be best for

African researchers to start introducing a technology as part of research and build capacity in skills needed to deal with WGS data analysis and interpretation.

1.2.2. Whole Exome Sequencing

WES is used to screen for clinically significant variants in the coding regions of the genome. Although exons cover only 1-2% of the genome, about 85% of known pathogenic mutations are located in these coding and functional regions of the genome (Rabbani *et al.,* 2014). The first proof of concept for WES was published in 2009, in this study, WES demonstrated the ability to target a causative gene, *MYH3* associated with Freeman-Sheldon syndrome (Ng et al., 2009). Early reports of utilising WES to make clinical diagnoses were reported in patients suspected of having Bartter syndrome, where homozygous variants were identified in the *SLC26A3* gene (Choi et al., 2009). Pathogenic variants were identified in patients with Miller syndrome in the *DHODH* gene (Ng et al., 2010).

After the initial success of WES in diagnosis, the technique has been utilised extensively across the world in medical laboratories to diagnose genetic disorders and find novel pathogenic mutations for known phenotypes (Seo et al., 2020; Alawi et al., 2021). WES is particularly useful in molecular medicine for the diagnosis of patients who present with features suggestive of a genetic disorder where an exact clinical phenotype is unknown, or nonspecific such as developmental delay (Wright et al., 2015; Mahler et al., 2019; Guo et al., 2021). It is also useful where the genetic etiology is complex with a large number of relevant genes, such as in hearing loss (Lewis et al., 2018; Zhou et al., 2020).

The detection rate of Mendelian disorders utilising WES, has been shown to increase in recent years. Mostly, the laboratories utilise WES when the traditional first line testing such as microarray and targeted sequencing do not confirm any diagnosis for the disorder. For this reason, most laboratories are exploring WES as the first line diagnostic test for genetic disorders where more genes may be involved. There are few studies in the literature where a detection rate of >50% was obtained using WES, such as in a cohort of patients with myopathy, where the study authors concluded that their phenotype driven exome analysis diagnostic approach improved the detection rate in this group of patients (Božović et al., 2021). In a carefully selected cohort of patients presenting with ciliopathy syndromes, a detection rate of 50% was obtained when utilising WES as a first line test (Alawi et al., 2021). In a cohort of patients presenting with cerebral palsy neurodevelopmental disorder, the detection rate ranged

from 11.2% in adult patients and 32.9% in pediatric patients, where these yields indicate that for some of the disorders there may be other factors, not necessarily genetic causing the disorder (Moreno-De-Luca et al., 2021). A study by Seo *et al* (2020) demonstrate the significance of having an analysis system that can assist with interpreting the WES data, a 42% detection rate was obtained in patients with various genetic disorders using their new variant interpretation system that is based on the human phenotype ontology (HPO) (Seo et al., 2020). Recently, a new American College of Medical Genetics (ACMG) and the Association for Molecular Pathology (AMP) practice guidelines (ACMG-AMP) recommendation was released on using WES/WGS as first or second line testing for congenital anomalies, developmental delay and intellectual disability, because of higher detection rates reported compared to single gene or targeted gene panel testing (Manickam et al., 2021).

1.2.3. Targeted next generation sequencing gene panels

Targeted NGS gene panels interrogate gene regions that are associated with the disorder or phenotypes of interest (Santani et al., 2017; Martin et al., 2019). They can be routinely used as a cost-effective and efficient tool to screen for pathogenic variants in Mendelian disorders and disorders with highly heterogeneous backgrounds, for example, cardiomyopathies, ataxia or Noonan syndrome (Ouellette et al., 2018; Wallace and Bean, 2020). It is common practice to use one panel for a group of phenotypically related disorders followed by targeted analysis.

Targeted NGS panels, as with all other NGS approaches, involve multiplexed sample preparation, followed by massively parallel sequencing and significant bioinformatics analysis (Kingsmore et al., 2012). Targeted NGS-based gene panel approaches remain cheaper to implement compared to whole genome or whole exome sequencing (WGS or WES) in diagnostic settings, and have consequently become a standard first tier test for many monogenic disorders in developed countries (Yohe *et al.*, 2015; Santani *et al.*, 2017). When utilised for addressing a basic and direct research question, NGS targeted panels have shown to be valuable due to their low cost and the higher sequencing depth compared to WGS and WES (Fujiki et al., 2018; Schwarze et al., 2018; Gulilat et al., 2019).

The ACMG-AMP practice guidelines recommend that to develop a useful and effective panel, genes included should be chosen based on evidence of a sufficient causative role they have on disease (Rehm et al., 2013; Martin et al., 2019). If a gene panel targets multiple disorders, then analysis must be based on the clinical phenotype. This is done to limit the number of VUSs

that may be identified in genes not associated with that disorder. It is important to ensure that genes included in the targeted panel are covered comprehensively with higher sequencing depth to allow for accurate and more reliable analysis (Strande et al., 2017a).

1.2.4. Analysis of next generation sequencing data

The data generated from NGS technologies is large and complex and requires specialist bioinformatics expertise to analyse (Rabbani et al., 2016; Seifi et al., 2017; Roy et al., 2018) and high throughput hardware infrastructure that is able to store large amounts of data is required (Kanzi et al., 2020; Pereira, Oliveira and Sousa, 2020). The typical NGS run must be inspected for overall quality to ensure success. Parameters such as the percentage of the target region sequenced, how many reads aligned to the reference sequence and average coverage and depth are checked in order to deem the NGS run successful and ready to analyse (Rehm et al., 2013; Strande et al., 2017a). NGS data analysis can be done in three phases; primary, secondary and tertiary (Figure 1.2).

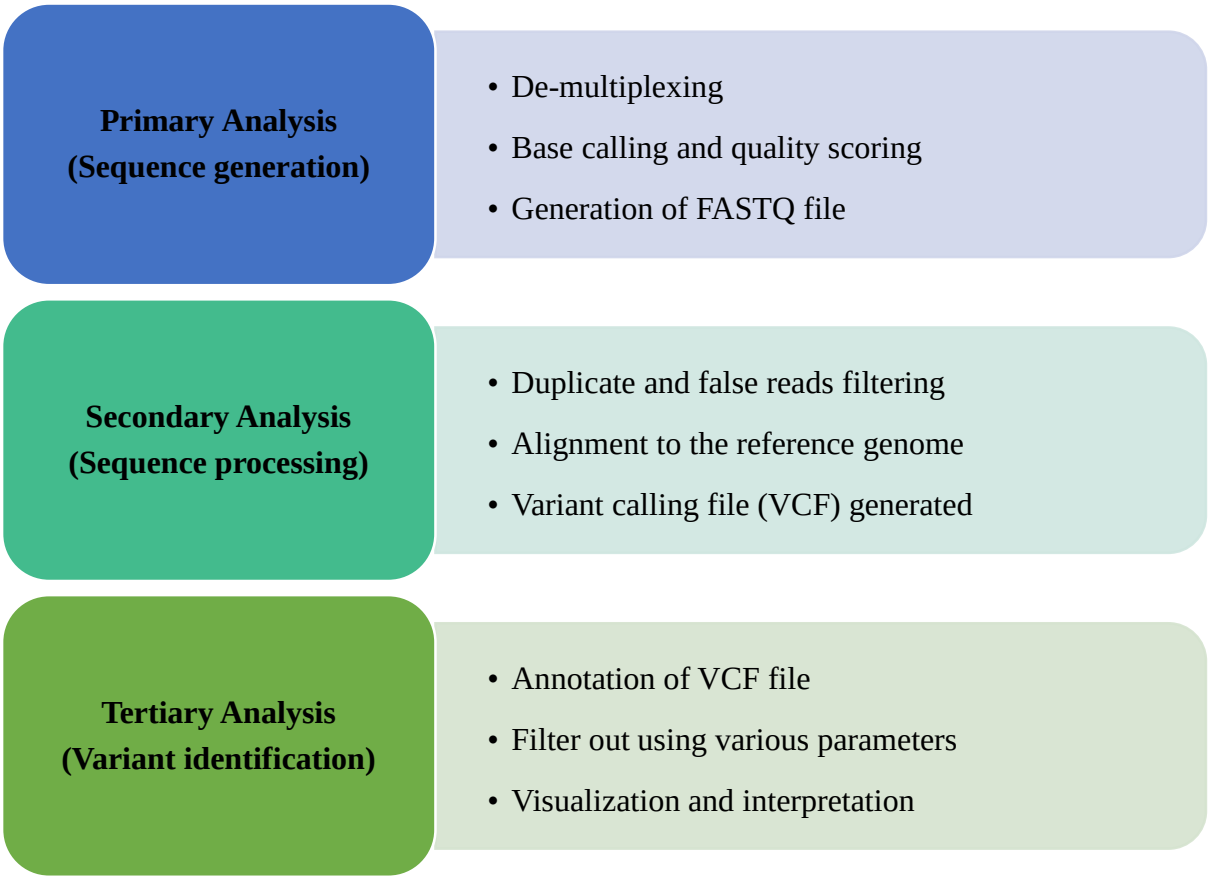


Figure 1.2: Three stages of next generation sequencing data analysis

Primary data analysis occurs in the NGS instrument, as built-in software. During this phase, the FASTQ files (Illumina instruments) or the unmapped binary alignment map file (UBAM) files (Ion Torrent) which consist of nucleotide base calls are generated. Various NGS platforms have the built-in de-multiplexing protocols that are required for the sequencing reads to be separated into individual files according to the barcodes used for each sample (Stawinski et al., 2016; Pereira, Oliveira and Sousa, 2020).

FASTQ/UBAM files contain all the raw reads, file names and quality values and they can function as the first quality control step (Pereira, Oliveira and Sousa, 2020), which guides the success of the entire NGS data analysis. Owing to the massive nature of the NGS technology, errors occur in both Ion Torrent and Illumina platforms. The errors are usually shown as quality scores of the base calls using the phred score. A phred score of 30, usually written as Q30, is the preferred minimum quality of the NGS data. A phred score of 30 means 1/1000 probability of an incorrect base or 99.9% accuracy (Liao, Satten and Hu, 2017). The phred score is also used as a measure to trim and filter out adapter sequences at the end of each read and this increases the quality of the acceptable sequence (Stawinski et al., 2016; Pereira, Oliveira and Sousa, 2020). When the sequences have enough quality reads, it can then be aligned or mapped against the reference genome.

Secondary data analysis starts with the alignment of reads to the reference human genome, Genome Reference Consortium Human Reference 37 (GRCh37/hg19) or (GRCh38/hg38). The human genome assembly, hg19 is the older version released in 2009 by the University of California and hg38 was released four years later (2014). Hg38 is currently used and most updated providing alternate sequences and additional updated sequences (Pan et al., 2019). The FASTA format sequence used by various NGS sequencers read alignment is different between reference genomes, hg19 and hg38, because the latter contains additional updated sequences, including those at centromeres, 261 alternate loci across 178 regions of the genome, filled in gaps from hg19 and also corrected misassembled regions. It is important to use the human reference assembly that is relevant to the study design and parameters, such as primer/probes for the targeted panel. Using an incorrect reference genome will lead to inaccurate data analysis.

Sequence alignment of the NGS data usually happens in the built-in bioinformatics pipelines of various sequencing platforms. NGS generates short sequencing reads that result in millions of DNA fragments being generated that need to be assembled to map to the reference genome.

This data needs good performance and a high throughput computational bioinformatics pipeline, because sometimes there are highly repetitive sequences that are not easy to align (Pereira, Oliveira and Sousa, 2020). Alignments can be viewed on the integrative genomics viewer (IGV), which is useful software to visualize aligned sequences and assess coverage of the regions targeted during sequencing (Robinson et al., 2017). When the alignment of the BAM file to the reference genome happens successfully, for example; if 30 reads have aligned to the reference sequence, which is 30x coverage depth, then the assembly will be ready for the variant call step.

Various analysis tools such as Agilent SureCall software (Agilent technologies, CA, USA) and the Ion-Reporter software (ThermoFisher Scientific, California, USA) are able to identify variants based on the number of high confidence base calls that deviate from the reference genome. Most of the tools use the BAM format input to generate a variant calling format (VCF) file as the output. The VCF file is a text file that is compiled with information about the variants detected during NGS run. Annotation follows, which add meta-data describing each variant (Tollefson et al., 2019).

Annotation of the VCF file happens in the final step of the NGS analysis called tertiary data analysis. Meta-data generated through annotation of VCF includes; minor allele frequency (MAF) in population databases like the 1000 Genomes project and the Genome Aggregation Database (gnomAD). The prediction scores of various *in-silico* protein function prediction tools and variant information from disease variant databases such as ClinVar are shown only after the annotation. Furthermore, information on the variant including the nucleotide changes affecting the gene, location of the variants (for example, upstream of a transcript in coding sequence or in non-coding RNA or in regulatory regions) is also generated. Information on chromosome location, reference allele, alternative allele and the consequence of the variants on the protein sequence, for example stop- gain, missense, stop-lost or frameshift is also shown when the VCF file is annotated (Tollefson et al., 2019).

Annotation tools use the VCF file to generate a report of variants identified and the biological function that it predicts. These tools include wANNOVAR software which allows scientists to submit the VCF file to the web server, select the desired annotation categories, and receive functional annotation back by emails (Chang and Wang, 2012; Yang and Wang, 2015). Another annotation tool which is commonly used and freely accessible is the Ensembl Variant Effect Predictor (VEP) which also uses the VCF file to determine the effect of the variants

(single nucleotide variants (SNVs) and small insertions and deletions) on genes, transcripts, and protein sequence, as well as regulatory regions for that particular sample (McLaren et al., 2016). Moon (Diploid) software was also employed in the current study, the input VCF file from targeted, WES or WGS together with (human phenotype ontology) HPO terms describing the patient's phenotype and known disorder manifestations as well as gender and age of onset which are critical during annotation by Moon software. Moon uses an artificial intelligence (A.I.) algorithm to analyse data and significantly reduces time taken to suggest variants associated with the clinical phenotype compared to other annotation software (Wenger et al., 2017; Clark et al., 2019)

Filtering information generated during variant annotation is aimed at reducing the number of variants needing further prioritization. Variants may be included or excluded based on presence in the disease variant databases, inheritance pattern, mutation type, presence and absence in the control populations, *de novo* occurrence, and gene expression pattern and *in-silico* prediction of protein function. The filtering process differs according to the disorder and requires a high level of genetics, molecular biology as well as bioinformatics knowledge (Rehm et al., 2013; Hume et al., 2019; Lalonde et al., 2020; Pereira, Oliveira and Sousa, 2020).

Below are some of the annotated characteristics that are used to prioritize or filter out variants

a) Minor allele frequency (MAF) in population databases

MAFs refers to the frequency of the minor allele of a specific variant, and usually applies to the genetic variants occurring at less than 5% frequency in the general population. The less common the variant is, the more likely it is to undergo further prioritization. Filtering out common variants in a population using MAF is often the first step in the analysis of NGS sequencing data (Linck and Battey, 2019). When the right parameters are used for filtering and the function of variants is understood, then the chances of finding putative disease causing variants is increased. Although filtering for most genetic variants uses <5% MAF, there are instances where specific cut-offs are used following extensive analysis of some disorders. For example, when analysing RASopathy gene variants a minor allele frequency (MAF) of 0.00025 (0.025%) is used to filter out variants for further analysis. This cut-off was determined by the RASopathy expert panel, the panel constituted to apply ACMG to the RASopathies (Gelb et al., 2018).

Population databases are useful in obtaining the frequencies of variants from large population studies. These include the 1000 Genomes Project, which is a database of variants found during

low and high coverage genomic and targeted sequencing from 26 populations (Devuyst, 2015b), and the Exome Aggregation Consortium (ExAC), which contains variants found during exome sequences of 60,706 unrelated individuals, sequenced as part of various disease-specific and population genetic studies (Lek et al., 2016a). The Genome Aggregation Database (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community. This database describes the aggregation of 125,748 exomes and 15,708 genomes from human sequencing studies into the Genome Aggregation Database (gnomAD) (Karczewski et al., 2019; Genome Aggregation Database Consortium et al., 2020). All these population data sets are essential in NGS variant analysis as they provide information to distinguish putative disease-causing variation from wild-type variation.

b) *In-silico* protein prediction software

During annotation, the *in-silico* tools that predict functional impact of the variant are included during tertiary analysis. The common ones include FATHMM (Shihab et al., 2013), SIFT (Kumar, Henikoff and Ng, 2009), CADD (Kircher et al., 2014), Polyphen-2 (Adzhubei et al., 2010; Adzhubei, Jordan and Sunyaev, 2013) and MutationTaster (Schwarz et al., 2010). These tools use amino acid properties, functional domains and structural attributes of the sequence variant to determine if it is deleterious or tolerated in the protein function (Mahmood et al., 2017).

c) Disease databases.

Additional disease/variant databases such as the ones listed below are used to get more information on the variant and the associated disorders. It is important for researchers to publish their data and submit variant information to these databases so that it would be easier to determine if the variant is novel or previously reported.

- ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/): This is a database of human genetic variants and interpretations of their significance to the disease phenotype (Landrum et al., 2018).
- Human Disease Mutation Database (HGMD) (https://www.hgmd.cf.ac.uk/ac/index.php): Collates all published variants associated with human disease and links it to the relevant publication (Stenson et al., 2003).

- GeneReviews® (https://www.ncbi.nlm.nih.gov/books/NBK1116/): Detailed description of diseases is available and can be used to verify the phenotype of patients and also obtain information on gene variants and how they can be tested (Adam et al., 1993).
  In projects where multiple disease genes are sequenced using one targeted panel, selective analysis of data based on the clinical phenotype of the patient becomes imperative, and the databases discussed here are very suitable.

1.2.5. Interpretation of next generation sequencing data

The standard way to interpret and classify sequence variants is by using the ACMG-AMP guidelines for Mendelian disease variants. The ACMG-AMP guidelines (Appendix 8) are a useful resource for data analysis and variant interpretation (Richards et al., 2015). These guidelines were developed to assist and guide researchers to have a standardised way to classify variants (Landrum et al., 2018; Marian, 2020).

The ACMG-AMP guidelines use a combination of the variant information, patient phenotype and family history to assist researchers come to conclusion in classifying the variants into benign, likely benign, VUS, likely pathogenic and pathogenic. Variants of uncertain significance means that the clinical significance of variants identified in a patient is not fully understood, or that the variant is novel, has never been seen before and there is not enough evidence to classify these variants as pathogenic or benign. As there are different tools for data analysis that are separate from the library preparation or wet-lab protocols, the ACMG-AMP guidelines recommend that when the laboratory customises the data analysis tools, an analytical tool validation must be put in place when the test is performed in a diagnostic setting (Rehm et al., 2013; Rehder et al., 2021).

In most cases, NGS data is analysed and interpreted in keeping with various clinical scenarios. It is important for a clinician to detail the clinical description so that interpretation is correctly aligned with the clinical phenotype. When a clinical phenotype is detailed, the prioritized variants are checked to see if they have been previously reported to be associated with the disease. The preferred way to interpret a variant is to find a single nucleotide variant (SNV) that was previously described in multiple patients with the same phenotype, but that is not always the case. As time goes on and as more genetic information is generated and the clinical phenotype is refined through family history or thorough assessments, VUS's may need to be reclassified (Rehm et al., 2013; Richards et al., 2015; Gelb et al., 2018).

1.3. Cost considerations and diagnostic implementation of next generation sequencing

Clinical utility, sensitivity, specificity and efficacy as well as cost-effectiveness need to be demonstrated before an NGS-based test can be brought onto a diagnostic platform. This is preceded by a significant amount of research and development, as well as validation of the test and analysis pipeline during the implementation phase (Jennings et al., 2017a; Hartman et al., 2019). A diagnostic service based on NGS testing also needs to be implemented into an appropriate clinical environment due to the need for proper clinical evaluation of patients, pre- and post-test genetic counselling, cascade screening and the need for complementary techniques such as Sanger sequencing, chromosomal microarray and others when necessary (Patch and Middleton, 2018a).

1.3.1. Diagnostic utility

Diagnostic utility of a test should be at the centre of why the technology is introduced. NGS technology has been shown to have the ability to identify the causative variant and help the clinician to confirm the diagnosis in patients with a particular disease. The diagnostic utility of NGS may vary depending on the selection or inclusion criteria of the patients for a particular test and study (Smith et al., 2019). The clinical evaluations done on the patients also do influence the diagnostic utility of a test. The technical aspecst of the NGS, for example if a test is performed using WGS, WES or targeted panels do affect the diagnostic utility of test. In most diagnostic laboratories, targeted gene panels are an easy and logical extension onto traditional methods of variant detection when compared to WES or WGS (Hayeems et al., 2020).

When using targeted gene panels the diagnostic yields are reported to be better compared to WGS/WES and this was demonstrated in a study by Scocchia and colleagues in patients with skeletal dysplasia (Scocchia et al., 2021). The ACMG guidelines recommend that laboratories initiate their NGS implementation in the diagnostic setting using disease-targeted gene panels and validate the protocol (Miller et al., 2017; Santani et al., 2017). When a test is limited to just regions and genes of interest, it is possible to have higher detection rates (due to targeted analysis based on the phenotype) which leads to increased diagnostic yields, and this is what targeted panels offer. When the technical quality and interpretation is well understood and validated, then laboratories can begin introducing WES/WGS into the diagnostic service (Rehm et al., 2013; Hume et al., 2019).

The analysis strategy of focusing only on single nucleotide variants (SNVs) in the coding regions of a particular gene, the entire gene including introns, exons with promoter regions or copy number variants also plays a major role in the diagnostic utility. This highly depends on the mechanism of disease in a gene or disorder being studied or tested, and the type of variants commonly associated with the disease (Kanzi et al., 2020). Analysing NGS data based on the clinical data such as phenotypic features and family history, is preferred and result in variants classified as pathogenic and likely pathogenic variants which are clinically significant and less VUSs. Variants of uncertain significance do not add value in the diagnostic or clinical utility as they neither confirm or refute the diagnosis and they are not easy to interpret and translate into clinical management (Federici and Soddu, 2020a; Mighton et al., 2021).

Another potential complication of NGS is the discovery of incidental findings, where an unexpected disease causing variant is identified, thus highlighting the need to have a proper clinical examination before any NGS procedure can be undertaken (Fujiki et al., 2018). When using NGS in a diagnostic setting, patients should have pre-test genetic counselling so that they have an idea of the procedure and what can be expected (Patch and Middleton, 2018b), because a test can be VUS or negative which adds anxiety to the patient and relatives. The genetic counselling process informs the patients and families affected by the genetic disorder of the chances of passing the disorders onto future generations and the available genetic interventions. When the diagnosis is confirmed, then clinicians or genetic counsellors communicate how the results will impact the patients and family; this demonstrate both diagnostic and clinical utility of a genetic test (Kromberg, Wessels and Krause, 2013).

1.3.2. Sensitivity, Specificity and efficacy

The ability of a test to accurately identify causative variant in patients with a disease and not in those without the disease is crucial to any new test implementation (Chin, da Silva and Hegde, 2013a). The possibility of errors that could occur during NGS runs make it difficult for most NGS platforms to correctly call low level pathogenic variants at less than 5-20% variant allele frequency (<5-20% VAF) (Mu et al., 2016a; Yan et al., 2021). These low-level variants are mosaic, mostly false positives, and could occur at any major step during NGS, they make it difficult to obtain accurate variant calls.

Some low level calls originate from sample or DNA processing artifacts. Mostly peripheral blood samples yield good quality DNA that is suitable in most NGS platforms whereas

formalin fixed samples usually yield poor quality DNA that leads to decreased amplification yields and induces artifacts that are non-specific and often occur as low-level sequence variants (Lucena-Aguilar et al., 2016; Robin et al., 2016). Poor quality DNA interferes with the accurate performance of NGS platforms as most of these include PCR amplification steps to amplify genes/ regions of interest. Successful amplifications rely on correct PCR cycles, good quality DNA, high fidelity DNA polymerase with the ability to reduce PCR artifacts; because PCR errors induced by the DNA polymerase and the process of sequencing by synthesis could contribute to sequence errors. Most PCR errors occur in AC and AT rich genomic regions, this result in low sequencing coverage affecting accurate variant detection (Brandariz-Fontes et al., 2015).

Illumina sequencing platforms are reported to provide high sequencing accuracy but are also prone to errors especially in GC-rich regions. The sequencing error rate on Illumina platforms is <0.8% and could affect low level variant detection accuracy (Stoler and Nekrutenko, 2021). The Ion Torrent platform causes most errors in homopolymer regions due to changes in pH that may affect semiconductor chips. Most Ion Torrent errors are indels (1%-2%) and substitution errors are reported at a lower rate (Singh, 2020a; Heo et al., 2021).

Data analysis is also another important step in eliminating false positive calls to achieve higher sensitivity, specificity and efficient NGS outcome. This is because data analysis includes filtering of variants based on set parameters on the analysis software depending on sequence quality and VAFs. Sequencing artifacts or false positive calls could originate if the set parameters are too strict or too lenient (Bauer et al., 2019). It is important to establish accurate filtering parameters based on sequencing platforms, types of variants and sequencing depth. Higher sequencing depth results in higher sensitivity, specificity and accuracy of the NGS procedure (Borges et al., 2020). The depth of coverage of the NGS reaction is one important parameter for quality control and confidence in variant calls; it refers to the number at which a target area was covered adequately and aligned to the specific genomic position during sequencing.

Depth of coverage value is usually the first one to be checked during quality assessment of most NGS runs. The depth varies depending on NGS procedure, targeted panel (30x - ≥200x), WES (30x-100x) and WGS (15x-50x). There are currently no established guidelines for adequate coverage of depth as this highly dependent on the aim of the experiment and NGS

platform used (Sims et al., 2014a; Petrackova et al., 2019a; Deng et al., 2020). There might be bias in the depth of coverage observed in NGS procedures, so it is good practice to assess the coverage throughout especially in the regions of interest. Some problematic regions such as GC-rich or homopolymers may have poor coverage leading to uneven distribution of depth (Ross et al., 2013; Nardi et al., 2021).
Sanger sequencing can be used to confirm NGS identified variants so that the sensitivity aspect of the NGS can be properly determined (Chin, da Silva and Hegde, 2013b; Mu et al., 2016a). However, there are studies that argue that Sanger sequencing is unnecessary to achieve high sensitivity and specificity, particularly in targeted gene panel sequencing where adequate depth of coverage is achieved (McCourt et al., 2013; Baudhuin et al., 2015a; De Cario et al., 2020).

For diagnostic implementation, validation of the initial phases of NGS findings should happen and must have high concordance to Sanger sequencing results. For some bigger validation cohorts, false positive calls have been observed to reduce sensitivity, mostly originating from variant calling processes. Assessing variant calling pipelines could further reduce false positive calls (Sivakumaran et al., 2013). Random Sanger-based confirmatory sequencing of a subset of NGS identified variants could be used to maintain high specificity and sensitivity in targeted panel based NGS test for diagnostics (Bauer et al., 2019).

1.3.3. Costs implications

When sequencing more than 20 target regions NGS is generally more cost-effective than Sanger sequencing. As already mentioned the ability for NGS to detect more genomic targets at the same time makes it more efficient and cost-effective than traditional ways of mutation analysis. Prior to NGS, available methods for mutation screening included micro-array, MLPA, hot-spot Sanger sequencing, karyotyping, common aneuploidy screening and others (Durmaz et al., 2015). With the introduction of NGS, all these tests could be reduced and health care resources could be used more efficiently and effectively, allowing for better management of the disorder. NGS is one technology that keeps up with or even surpasses Moore's law, which describes development trends in the computer hardware industry (Cavin, Lugli and Zhirnov, 2012). With recent reductions in NGS costs, it is appropriate to compare this trend to Moore's law. South Africa being resource constrained for genomic research, it is always preferable to introduce new technologies cost-effectively and maximize the resources (Kromberg, Sizer and Christianson, 2013; Tekola-Ayele and Rotimi, 2015b; Krause, 2019b).

There are a few factors to consider when determining the costs of NGS. These include reagents (DNA isolation, library preparation and sequencing reagents), consumables, NGS sequencing instruments, DNA quality assessment equipment, instrument maintenance, data analysis and storage. All these should be factored in when assessing costs per sample. Costs also depend on which NGS test is done; targeted panel, WES or WGS. As previously mentioned all these generate different amount of data, so the cost for analysis and storage would differ based on that. For example, Illumina MiSeq platform sequencing cost per gigabase (Gb) of DNA sequence is approximately US$500, instrument costs are estimated around US$128K (Gordon et al., 2020a; Pruneri et al., 2021). Minimum data output per patient for targeted gene panels are ~0.5-5 Gb depending on the number of genes the targeted panel contains and the sequencing kits used, for WES DNA sequence output is ~6-10 Gb and WGS data output is ~120-150 Gb (Wadapurkar and Vyas, 2018; Zhao et al., 2021).

In instances where there is no clear diagnosis of the disorder, it is reported that performing WES could be more cost-effective than targeted panel as it prevents additional unnecessary testing that are sometimes not genetic. The costs may be increased for data analysis as this analysis generate many variants to scrutinize compared to more targeted analysis using gene panel. Current costs of WES range from ~ US$2800 – US$4000 depending on the sequencing platform used and this is without taking account of data analysis and storage costs (Schwarze et al., 2018; Gordon et al., 2020b).

1.4. RASopathies

The following section will focus on the group of disorders studied in this project, RASopathies. Subsections includes genetic aetiology of RASopathies, their mechanism of disease as well as the description of each RASopathy syndrome from most common to rarest.

1.4.1. Genetic aetiology of RASopathies

RASopathies are caused by germline pathogenic variants found in genes regulating the Ras/mitogen activated protein kinase (Ras/MAPK) pathway (Rauen, 2013). RASopathies are clinically heterogeneous and associated with over 22 genes (Rauen, 2013; Jafry and Sidbury, 2020). Each RASopathy is a single-gene disorder, usually inherited in autosomal dominant manner with ~30% of the variants occurring *de novo* (Hebron, Hernandez and Yohe, 2022). The overall prevalence across all RASopathies is estimated to be one in 1000 new-borns (Jindal

et al., 2015; Hebron, Hernandez and Yohe, 2022). This makes them common and thus necessary to study their genetic aetiology.

There are approximately 22 known genes associated with the pathogenesis of RASopathies (Table 1.1), with some genes causing different conditions with overlapping features within the RASopathies, such as *MAP2K1* and *BRAF* mutations causing both Noonan syndrome and cardio-facio cutaneous syndrome and *PTPN11* mutations causing Noonan syndrome and Noonan syndrome with multiple lentigines. The data shown on Table 1.1 is mostly European (non-African) data; very little is known about mutation profiles for RASopathies in African populations. When this study was established, the only diagnostic testing for a RASopathy done in South Africa was for Costello syndrome, and only included screening for the common mutations in the *HRAS* gene (those mutations includes c.38G>A, c.37G>A, c.35G>C, c.35CG>AA and c.34G>A) (Fahmida Essop, personal communication, 2008).

Table 1.1: Genes associated with different RASopathies and their relative mutation contribution in the reported literature.

RASopathy	Gene	Chromosomal location	Contribution of variants	Reference
NS	*PTPN11* (OMIM 176876)	12q24.13	~40-50%	(Tartaglia et al., 2002; Xu et al., 2017; Athota et al., 2020)
	RAF1 (OMIM 164760)	3p25.2	3-17%	(Pandit et al., 2007; El Bouchikhi et al., 2016)
	SOS1 (OMIM 182530)	2p22.1	10-20%	(Roberts et al., 2007; Rauen, 2013; El Bouchikhi et al., 2016)
	SOS2 (OMIM 601247)	14q21.3	~4%	(Lissewski et al., 2021)
	LZTR1 (OMIM 600574)	22q11.1-q11.2	~8%	(Yamamoto et al., 2015)
	KRAS (OMIM 190070)	12p12.1	~5%	(Brasil et al., 2010)
	RIT1 (OMIM 609591)	1q22	5%	(Aoki et al., 2016)
	MAP2K1 (OMIM 176872)	15q22.31	<2%	(Nava et al., 2007a)
	BRAF (OMIM 164757)	7q34	~1.7-2.3%	(Sarkozy et al., 2009; Lee et al., 2011)
	NRAS (OMIM 164790)	1p13.2	(rare)~1%	(Cirstea et al., 2010; El Bouchikhi et al., 2016)
	RASA2 (OMIM 601589)	3q23	~1%	(Chen, Yin, et al., 2014)
	CBL (OMIM 165360)	11q23.3	~1%	(Rauen, 2013)
	SHOC2 (OMIM 602775)	10q25.2	~1%	(Chen, Yin, et al., 2014)
	A2ML1 (OMIM 610627)	12p13.31	rare	(Lee and Yoo, 2019a)
	RRAS (OMIM 165090)	19q13.33	rare	(Koh et al., 2019)

NSML	*PTPN11*	12q24.13	~90%	(Sarkozy, Digilio and Dallapiccola, 2008; Sarkozy et al., 2009; Lee et al., 2011)
CFC	*BRAF*	7q34	~75%	(Sarkozy et al., 2009; Pierpont et al., 2014)
	MAP2K1 and *MAP2K2* (OMIM 601263)	15q22.31 and 19p13.3	~25%	(Rodriguez-Viciana et al., 2006; Nava et al., 2007a)
CS	*HRAS* (OMIM 190020)	11p15.5	~80-90%	(Estep et al., 2006a; Bertola et al., 2017)
CM-AVM	*RASA1* (OMIM 139150)	5q14.3	~70%	(Eerola et al., 2003; Boon, Mulliken and Vikkula, 2005)
NF1	*NF1* (OMIM 613113)	17q11.2	>90%	(Rauen, 2013; Koczkowska et al., 2018)
NF2	*NF2* (OMIM 607379)	22q12.2	~90%	(Ahronowitz et al., 2007: 2; Tiwari and Singh, 2021: 2)
LS	*SPRED1* (OMIM 609291)	15q14	>99%	(Brems et al., 2012)

In patients of African ancestry, the genetic aetiology of these disorders is understudied. One study from South Africa focused on a few patients presenting with one RASopathy, namely; Noonan syndrome (Tekendo-Ngongang et al., 2019a). It has not always been feasible to study these disorders individually in a resource-constrained environment because of the time and cost involved in optimizing the sequencing and analysis of each gene. Using traditional Sanger sequencing, often only the hot-spot regions of the gene are targeted. The costs involved would have prevented the sequence and complete analysis of all the genes associated with the disorder. Next Generation Sequencing technologies have revolutionised the way we do research and develop diagnostic tests for these disorders.

1.4.2. Disease mechanism of RASopathies

Ras/MAPK (mitogen-activated protein kinase) pathway dysregulation is a common mechanism causing all RASopathies (Stevenson et al., 2016). This pathway (Figure 1.3) is critical in the regulation of cells involved in normal development (Aoki et al., 2016; Kratz and Zenker, 2018). The core components of the RAS/MAPK pathway are regulators of signalling which regulate intracellular signalling essential in the cell cycle, differentiation, growth and apoptosis, which is crucial throughout all developmental stages of life (Tidyman and Rauen, 2009). The growth factors are transduced from the extracellular environment into the cell, thereby activating signals in the genes involved in the different pathways that are critical during normal cell development (Tajan et al., 2018). The dysregulation of the RAS/MAPK dependent developmental process has multiple developmental consequences and that includes RASopathies (Tajan et al., 2018).

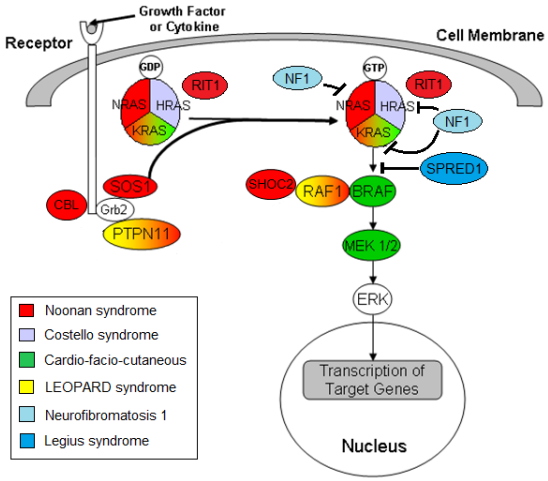


Figure 1.3: RAS/MAPK pathway showing some of the different genes associated with the RASopathies in different colours and where they act in the pathway. Figure obtained from Rauen,(2013)

Interruptions in the RAS/MAPK pathway leads to various RASopathies, including; neurofibromatosis type 1 (NF1), Noonan syndrome (NS), Noonan syndrome with multiple lentigines (NSML), capillary malformation-arteriovenous malformation syndrome (CMV-MS), Costello syndrome (CS), Cardio-facio-cutaneous syndrome (CFC) and Legius syndrome (LS). Each of these RASopathies exhibits a distinct phenotype, although there is considerable overlap in phenotypic features (Figure 1.4) (Rauen, 2013; Tajan et al., 2018).

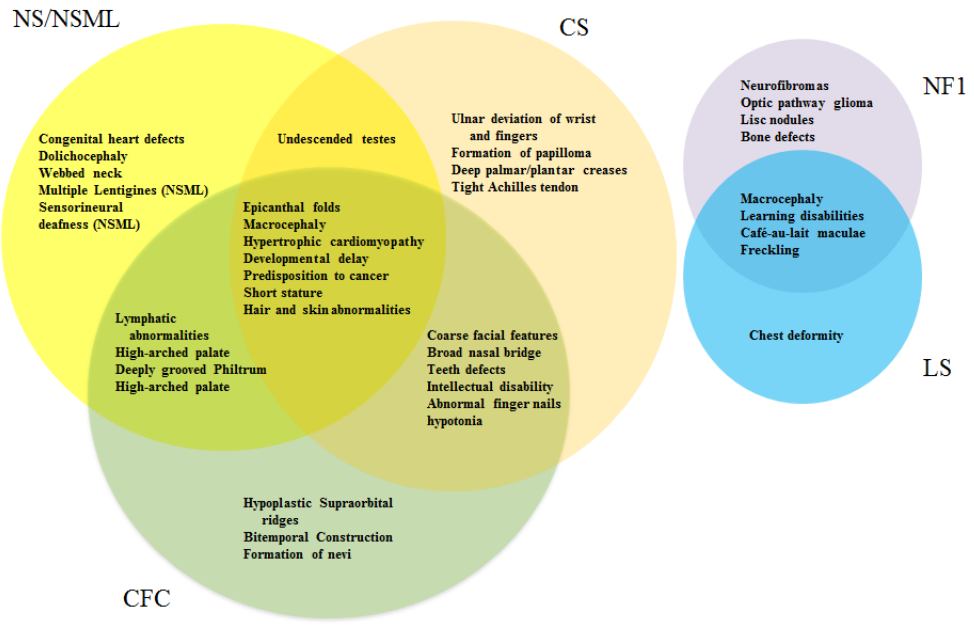


Figure 1.4: Phenotypic features of the main RASopathies and their clinical overlap. The figure is modified from Stevenson et al., (2016)

In the sections below each RASopathy will be further detailed. They are presented in order from most common to rarest based on their reported global prevalence with Noonan syndrome is the most common RASopathy followed by neurofibromatosis type 1. The rest of the RASopathies such as Noonan syndrome with multiple lentigines, capillary malformation-arteriovenous malformation syndrome, Costello syndrome, Legius syndrome, cardio-facio-cutaneous syndrome are all rare and for most of these the prevalence is not well characterised (Tidyman and Rauen, 2009; Rauen, 2013; Aoki et al., 2016; Hebron, Hernandez and Yohe, 2022).

1.4.3. Noonan syndrome (OMIM 163950)

Noonan syndrome (NS) is the most common autosomal dominant RASopathy with a prevalence rate of 1/1000 – 1/2500 live births (Rauen, 2013; Lee and Yoo, 2019a). NS can be either inherited or sporadic and is caused by gain of function mutations most commonly in the protein tyrosine phosphatase non-receptor type 11 (*PTPN11*) gene. Pathogenic variants associated with Noonan syndrome, also occur in other genes including the son-of-sevenless homolog 1 (*SOS1*) gene, neuroblastoma RAS viral oncogene homolog (*NRAS*) gene, Raf-1 proto-oncogene (*RAF1*) gene, proto-oncogene B-Raf (*BRAF)* gene, in cbl proto-oncogene (*CBL*) gene amongst others (Kang and Lee, 2019; Lee and Yoo, 2019a; Hebron, Hernandez and Yohe, 2022).

Mutations in the *PTPN11* gene, located on chromosome 12q24, account for 50% of NS cases. *PTPN11* encodes for the Src homology 2-containing phosphotyrosine phosphatase (SHP2) which activates multiple RAS/MAPK signalling pathways and is vital for the survival, growth and proliferation of tumour cells (Chong and Maiese, 2007).

About 60% of NS syndrome cases are caused by *de novo* mutations and approximately 40% are inherited (Bhambhani and Muenke, 2014). Almost all (~100%) pathogenic variants in genes associated with NS can be identified by sequence analysis (Allanson and Roberts, 1993; Bhambhani and Muenke, 2014).

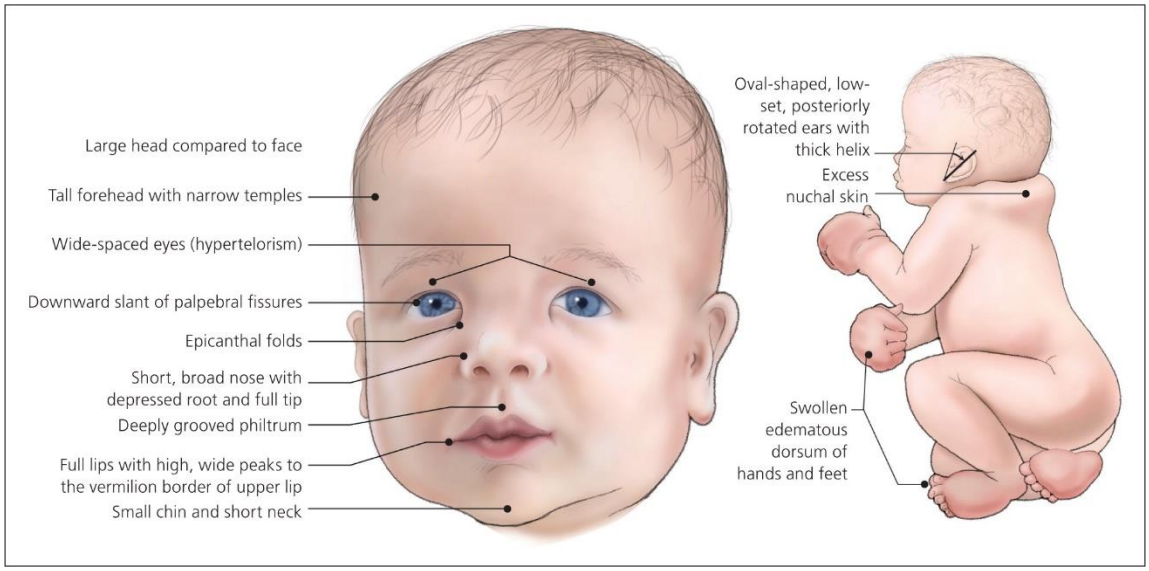


Figure 1.5: Diagram showing some clinical features of Noonan syndrome typically seen in a newborn baby. Obtained from Bhambhani and Muenke (2014)

Clinical diagnosis of patients with NS is based on typical facial abnormalities, including low-set posteriorly rotated ears, broad forehead, sparse eyebrows and a webbed neck (Rauen, 2013; Lee and Yoo, 2019a; Jafry and Sidbury, 2020). Other features include short stature, cardiac abnormalities and between 30-50% of NS patients have variable neurocognitive delay (Figures 1.5 and 1.6 show diagrams indicating the features of Noonan syndrome typically seen in a newborn baby and adolescent with NS). Features in a newborn baby are sometimes subtle and during adolescent years the features become prominent and more visible (Cesarini et al., 2009)

.

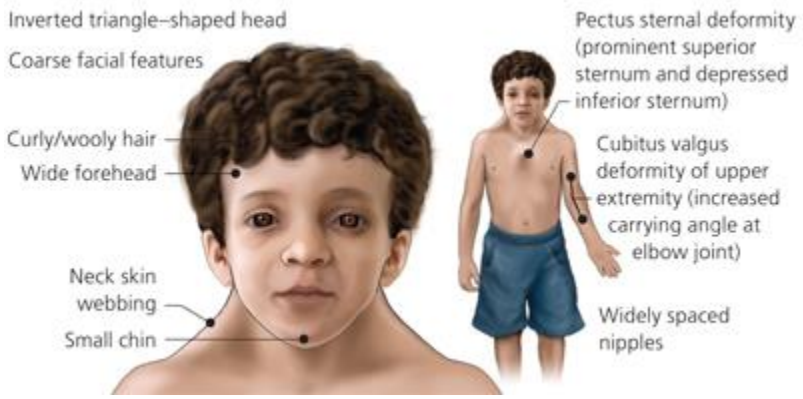


Figure 1.6: Diagram of an adolescent with Noonan syndrome. Obtained from Bhambhani and Muenke (2014)

The diagnosis of Noonan syndrome can be made using the criteria outlined in Table 1.2 adapted from (Carcavilla *et al.*, 2020)

Table 1.2: Minor and major clinical criteria for the diagnosis of Noonan syndrome (Carcavilla *et al.*, 2020)

Feature	A = Major criteria	B = Minor criteria	
1	Facial	Typical face dysmorphology (varies with age)	Suggestive face dysmorphology
2	Cardiac	Pulmonary valve stenosis, hypertrophic obstructive cardiomyopathy and/or ECG typical of NS	Other defect
3	Height*	< 3rd percentile	< Percentil p10
4	Chest wall	Pectus carinatum/excavatum	Broad thorax
5	Family history	First degree relative with definite NS	First degree relative with suggestive NS
6	Other	All of the following: Mental retardation, cryptorchidism and lymphatic dysplasia	One of the following: Mental retardation, cryptorchidism and lymphatic dysplasia

Noonan syndrome:
A (typical face dysmorphology) plus one other major criteria (2A-6A) or 2 minor criteria (2B-6B);
B (suggestive face dysmorphology) plus 2 major criteria (2A-6A) or 3 minor criteria (2B-6B).
* Height relative to applicable growth charts for age and sex.

1.4.4. Neurofibromatosis type 1 (OMIM 162200)

Neurofibromatosis type 1 is an autosomal dominant RASopathy and the first condition identified where the gene was identified to produce a protein implicated in the RAS-MAPK pathway. NF1 has complete penetrance but variable expression (Viskochil et al., 1990). NF1 is a relatively common disorder with a prevalence rate of 1/2500 - 1/3000 live births. Approximately 50% of the cases are familial and the other 50% are caused by *de novo* mutations (Anderson and Gutmann, 2015a).

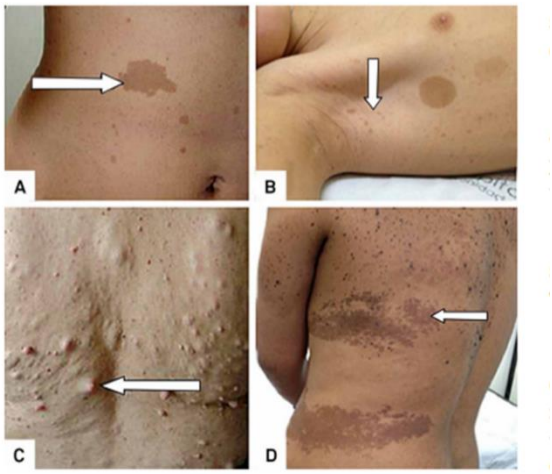


Figure 1.7: Common clinical features of NF1 patient, cutaneous lesions (arrows); multiple café au lait lesions (CAL) (A); freckling and two CAL (B); discrete neurofibromas (C); pigmented skin lesions (epidermal nevus) misdiagnosed as CAL and freckling in a non-NF1 patients (D), obtained from(Rodrigues et al., 2014)

Individuals with NF1 usually present with café-au-lait macules and cutaneous neurofibromas axillary and inguinal freckling and Lisch nodules of the iris. Other characteristics features includes bone malformations, neurocognitive impairments, benign and malignant tumours, structural cerebral abnormalities, and vascular anomalies (Figure 1.7) (Cimino and Gutmann, 2018; Mao et al., 2018; Kang and Lee, 2019; Jafry and Sidbury, 2020). The clinical diagnosis is made using the diagnostic criteria illustrated below in A and B (Legius et al., 2021) .

A: The diagnostic criteria for NF1 are met in an individual who does not have a parent diagnosed with NF1 if two or more of the following are present:

- Six or more café-au-lait macules over 5 mm in greatest diameter in prepubertal individuals and over 15 mm in greatest diameter in postpubertal individuals.
- Freckling in the axillary or inguinal region.
- Two or more neurofibromas of any type or one plexiform neurofibroma.
- Optic pathway glioma.
- Two or more iris Lisch nodules identified by slit lamp examination or two or more choroidal abnormalities (CAs)—defined as bright, patchy nodules imaged by optical coherence tomography (OCT)/near-infrared reflectance (NIR) imaging•
- A distinctive osseous lesion such as sphenoid dysplasia anterolateral bowing of the tibia, or pseudarthrosis of a long bone •
- A heterozygous pathogenic NF1 variant with a variant allele fraction of 50% in apparently normal tissue such as white blood cells

B: A child of a parent who meets the diagnostic criteria specified in A merits a diagnosis of NF1 if one or more of the criteria in A are present

Neurofibromatosis type1 is caused by loss of function mutations in the *NF1* tumor suppressor gene, which is located in the long arm of chromosome 17 (17q11.2), and codes for neurofibromin, a guanosine triphosphatase–activating protein that inhibits RAS activity.

Neurofibromin catalyses RAS GTPase activity, thereby negatively regulating the RAS/MAPK signalling pathway. The loss of function of the NF1 gene leads to the hyperactivation of the RAS/MAPK signalling (Jafry and Sidbury, 2020).

NF1 gene is one of the largest known genes with a genomic size of 282 kb, consisting of 57 constitutive exons and three alternatively spliced exons without any obvious hotspot regions. Over 2600 NF1 mutations have been reported in the Human Gene Mutation Database (HGMD) and ClinVar disease variant databases (Stenson et al., 2003; Landrum et al., 2018).

Pathogenic variants in the *NF1* gene result in abnormal cell growth, differentiation and proliferation; 60-90% of these variants can be identified by sequence analysis of the NF1 gene (Friedman, 1993; Mao et al., 2018). However ~22%-30% of the NF1 causing variants affect splicing and are not often detected by genomic DNA sequencing, thus for these kind of variants RNA analysis is appropriate (Evans et al., 2016).

Approximately 5%-10% of NF1-causing variants are copy number variations (CNVs) such as large deletions and whole gene deletions which are associated with a severe phenotype (Pasmant et al., 2010; Kehrer-Sawatzki and Cooper, 2012; Kehrer-Sawatzki, Mautner and Cooper, 2017). For these variants multiplex-ligation dependent probe amplification (MLPA) or microarray analysis may be utilised. There are four types of NF1 large deletion variants; type 1,2,3 and atypical, these are differentiated based on their size and breakpoints. Type 1 NF1 deletions encompass 1.4Mb and account for ~70-80% of all large NF1 deletions. Type 2 deletions covers 1.2Mb and accounts for ~10% of large NF1 deletions. Type 3 NF1 deletions are 1Mb in size and are very rare occurring in 1-4% of all patients with NF1 deletions. Atypical NF1 large deletions do not have recurrent breakpoints and are variable in size. About 8-10% of all large NF1 deletions are atypical (Kehrer-Sawatzki and Cooper, 2021; Legius et al., 2021; Kehrer-Sawatzki et al., 2022)

1.4.5. Neurofibromatosis-Noonan syndrome (OMIM 1601321)

Neurofibromatosis-Noonan syndrome (NFNS) is a rare phenotype that has both NF1 and NS clinical features. The RAS/MAPK pathway is responsible for the pathophysiology of both NF1 and NS phenotypes. The *NF1* gene regulates the RAS in the starting phase of the RAS/MAPK signalling along with the SHP2 which is encoded by the *PTPN11* gene for NS (Yapijakis, Pachis and Voumvourakis, 2017; Zhang et al., 2020). A study done by Reig and colleagues suggested that most patients with an NFNS phenotype have pathogenic variants in the *NF1* gene and not in both *NF1* and *PTPN11*, and these should be called NF1 syndrome with Noonan

like features (Reig et al., 2011). However, research done prior to 2011 identified some patients with this syndrome harbouring variants in both *NF1* and *PTPN11* genes (Stevenson et al., 2006; Thiel et al., 2009).

1.4.6. Noonan syndrome with multiple lentigines (OMIM 151100)

Noonan syndrome with multiple lentigines (NSML) was previously referred to as LEOPARD syndrome. LEOPARD is an acronym for the features of this disorder meaning; Lentigines, ECG conduction defects, Ocular hypertelorism, Pulmonary stenosis, Abnormalities of the genitalia, Retarded growth and Deafness. Genital abnormalities occur in 50% of individuals and sensorineural deafness is present in 20% of NSML cases (Sarkozy, Digilio and Dallapiccola, 2008). The multiple lentigines are often observed in the upper part of the body, primarily on the face. These continue to increase when the patient reaches puberty, Figure 1.8 (Jafry and Sidbury, 2020). This syndrome and NS have some overlapping clinical features including hypertrophic cardiomyopathy, short stature, pulmonic valve stenosis and hypertelorism. Most NSML (~90%) pathogenic variants are found in *PTPN11* (Gelb and Tartaglia, 1993; Tartaglia et al., 2002).



Figure 1.8: Skin with lentigines in individuals with NSML at various ages. Obtained from Sarkozy, Digilio and Dallapiccola (2008)

1.4.7. Capillary malformation-arteriovenous malformation syndrome (OMIM 608354)

Capillary malformation-arteriovenous malformation syndrome (CM-AVM) syndrome is caused by mutations in the *RASA1* gene. It is characterized by the presence of multiple small capillary malformations primarily localized on the face and limbs. Some affected (15% - 20%) individuals also have associated arteriovenous malformations (AVMs), fast-flow vascular anomalies that typically arise in the skin, muscle, bone, spine, and brain. Life-threatening

complications of these lesions can include bleeding, congestive heart failure, and/or neurologic consequences (Eerola et al., 2003; Larralde et al., 2014).

Approximately 70% of the CM-AVM cases are inherited and about 30% have a *de novo* pathogenic variant and 50% of these cases have variants in the *RASA1* gene, and of these, 92% can be identified using sequence analysis and only a small percentage (~8%) are reported to be deletions/duplications (Bayrak-Toydemir and Stevenson, 1993). Pathogenic variants in the *RASA1* gene lead to loss of function of the protein product (p120-*Ras*GAP), which is a regulator of the RAS-GTP signal transduction pathway. This then leads to the negative regulation of the RAS-MAPK signalling pathway. The penetrance for CM-AVM mutations range from 90 - 99% with a prevalence of 1/100,000 individuals (Eerola et al., 2003; Boon, Mulliken and Vikkula, 2005).

1.4.8. Costello syndrome (OMIM 218040)

Costello syndrome (CS) is an autosomal dominant disorder caused by heterozygous mutations in the *HRAS* gene on chromosome 11p15 (Estep et al., 2006b). It is a rare condition with a prevalence of 1/300,000 to 1/1.25 million (Abe et al., 2012; Giannoulatou et al., 2013). It is typically caused by a *de novo* variant in the *HRAS* gene, which is the only gene known in which pathogenic variants cause CS. More than 95% of pathogenic variants causing Costello syndrome affect the amino acid p.Gly12Ser or p.Gly13Asp (Bertola et al., 2017) and can be identified using sequence analysis.

Pathogenic variants in *HRAS*, which is a cell growth and replication regulator gene, place Costello patients at a greater risk of developing a malignancy including abdominal and pelvic rhabdomyosarcomas. The striking feature of Costello syndrome is cutis laxa (loose/sagging skin) which is observed on the neck, palms, fingers and soles in 99% of CS patients. They also have a risk of developing papillomas and postnatal failure to thrive (Gripp and Rauen, 1993). Costello syndrome has phenotypic overlap with cardiofaciocutaneous syndrome (CFC) and Noonan syndrome including macrocephaly, coarse facies, epicanthic folds and cardiac anomalies. (Hennekam, 2003; Gripp et al., 2019).

1.4.9. Cardio-facio-cutaneous syndrome (OMIM 115150)

Cardio-facio-cutaneous syndrome (CFC) is a rare RASopathy with an unknown prevalence; it has significant overlap in its clinical features with those in NS (Pierpont et al., 2014). These

include epicanthic folds, short stature, relative macrocephaly, cardiovascular abnormalities and hypertelorism with down-slanting palpebral fissures (Lee and Yoo, 2019a). However, there are distinct features that are observed only in CFC patients and include scaly skin, sparse and curly hair, absent or sparse eyebrows and eyelashes and keratosis pilaris. A clinical diagnosis of CFC is usually based on the ectodermal abnormalities (Pierpont et al., 2014). About 75% of CFC patients carry a mutation in the *BRAF* gene and other mutations are found in the *KRAS, MAP2K1* and *MAP2K2* genes (Rauen, 1993; Pierpont et al., 2014). Some of the complications associated with a CFC diagnosis are that during infancy, there are feeding difficulties and gastrointestinal problems, which lead to the need for nasogastric tube feeding (Rodriguez-Viciana et al., 2006; Pierpont et al., 2014). Genitourinary complications occur in 20-40% of CFC patients who usually need genital evaluation and renal sonograms. Seventy percent (70%) of patients have heart disease requiring surgical intervention; these include pulmonic valve stenosis, atrial septal defects, ventricular septal defects and hypertrophic cardiomyopathies (Rodriguez-Viciana et al., 2006; Lee and Yoo, 2019a; Jafry and Sidbury, 2020).

1.4.10. Legius syndrome (OMIM 611431)

Legius syndrome is an autosomal dominant disorder that has similar clinical features to NF1 but is less severe (Legius and Stevenson, 1993). Patients with Legius syndrome typically have multiple cafe-au-lait macules and variable dysmorphic features such as hypertelorism, lipomas, and mild learning disabilities. It is rare compared to NF1, occurring in 1/200,000 individuals. Legius syndrome is considered a RASopathy because it arises from mutations in the *SPRED1* gene, which encodes a negative regulator of the RAS-MAPK pathway (Brems et al., 2012). The *SPRED1* gene is located on chromosome 15q14 and screening for mutations in this gene is the effective diagnostic method for Legius syndrome (Legius and Stevenson, 1993; Benelli et al., 2015). Approximately 90% of the pathogenic variants can be identified by sequence analysis and 10% by deletion/duplication analysis of the *SPRED1* gene (Benelli et al., 2015).

1.4.11. Neurofibromatosis type 2 (OMIM 101000)

This disorder is clinically different from NF1 in that it is characterized by schwannomas, plexiform schwannomas, cortical wedge cataracts and retinal hamartomas. Adults with NF2 present with bilateral or unilateral vestibular schwannomas resulting in hearing loss and this is observed in 90-95% of the patients (Evans, 1993). Approximately 50% of NF2 adults also

present with multiple meningiomas. This disorder is inherited in an autosomal dominant manner and has an incidence of 1 in 25000-40000.

NF2 gene single nucleotide variants cause NF2 syndrome in 75% of individuals and deletions and duplications account for ~20% of affected individuals. Although patients may have a variant identified, variable expression is frequent and patients may have different sizes and number of tumors. However, there is 100% penetrance rate. Unlike NF1, large deletions and missense variants are associated with a milder NF2 phenotype; nonsense and frameshift variants are associated with severe disease (Tiwari and Singh, 2021).

1.4.12. Schwannomatosis (OMIM 162091/OMIM 615670)

Schwannomatosis is an autosomal dominant disorder that is phenotypically similar to NF2. In individuals presenting with multiple schwannomas, it is challenging for clinicians to distinguish between the two disorders (Dhamija et al., 1993; Alaidarous et al., 2019). Patients with these two conditions can present with two or more non-intradermal tumors suggestive of schwannomas, a unilateral vestubular schwannoma, an intracranial meningioma and have a family history of schwannomatosis. Patients have shown clinical variability and reduced penetrance. Schwannomatosis is associated with pathogenic variants in *SMARD1 and LZTR1* genes, Approximately 38% of the *LZTR1* variants account for the inherited *LZTR1* associated schwannomatosis and ~30% of the pathogenic variants are sporadic. In the *SMARCB1*-related schwannomatosis, 48% of cases are inherited and about 10% are sporadic. There are ~16% of inherited and ~60% of sporadic schwannomatosis cases that are not associated with the *LZTR1* and *SMARCB1* genes and their pathogenesis is unknown (Smith et al., 2015; Alaidarous et al., 2019).

1.5. Problem Statement

The lack of genetic data and appropriate, cost-effective local testing solutions on monogenic disorders has led to little molecular confirmation testing available for most monogenic disorders in the South African State healthcare system. NGS technology has enabled us to create genetic profiles that would assist us in the introduction of diagnostic testing for some of these disorders. Based on the review of the literature, it can be seen that RASopathies are common in global populations and little is known on their distribution in African populations. Understanding the genetics of these disorders in the African context would enable health professionals to treat and manage these group of disorders appropriately. Since these disorders are regulated by a single pathway, it makes sense to study their genetic profiles simultaneously. With the introduction of NGS in the medical sciences, studying these disorders has become possible, it is however important to have innovative and cost effective ways to adopt this state of the art technology in resource-constrained setting such as South Africa. A cost-effective RASopathy panel would enable us to generate mutation profiles in African RASopathy patients and the data would be useful for the development of an appropriate testing strategy. The additional benefit of NGS is that different disorders e.g. less common and more common conditions can be batched together on one assay to improve turnaround time and save costs.

1.6.Aims

Aim 1: To develop an efficient, cost-effective NGS-based targeted gene panel that can be used for the screening of clinically significant variants associated with RASopathies and other genetic disorders.

Aim 2: To determine the mutation profiles of RASopathies in South African patients and determine the most appropriate diagnostic detection approach

1.7. Objectives

Objectives for aim 1

a. To develop and design targeted gene panel to maximally utilise sequencing capacity
b. To optimise the performance of the custom-designed panel
c. Assess its cost effectiveness of implementing the NGS targeted panel testing

Objectives for aim 2

d. To recruit a cohort of patients with clinical features suggestive of a RASopathy.
e. To use the targeted gene panel to screen the cohort for disease-causing variants in known RASopathy genes.
f. To perform additional testing (WES/MLPA) on the patients where no pathogenic variants were identified using the custom-designed NGS targeted gene panel.

CHAPTER 2

Subjects and Methods

2.1.Patient recruitment

2.1.1. Ethics

This study forms part of an umbrella project for investigating developmental disorders using next generation sequencing (NGS) technologies, which was approved by the Human Research Ethics Committee (Medical) at the University of the Witwatersrand (Wits) with ethics clearance number: M160830 and the University of Pretoria (UP) ethics clearance number: 80/2018. Ethics clearance was obtained for the current project as a sub-study "Development of a multi-disease panel to study genetic aetiology of RASopathies" Ethics clearance number. M170163. The certificates are in Appendix 1.1-1.3. The ethics clearance was obtained for both Wits and UP as the recruitment and sampling was done by Medical Geneticists affiliated in the two Universities.

2.1.2. Patients and relatives

Sixty South African patients with clinical features suggestive of a RASopathy were recruited from genetic clinics around Johannesburg (JHB) and Pretoria (PTA), specifically at Chris Hani Baragwanath Academic Hospital (CHBAH), Charlotte Maxeke Johannesburg Academic Hospital (CMJAH), Rahima Moosa Hospital (RMH), National Health Laboratory Service, Braamfontein (NHLS) and Steve Biko Academic Hospital (SBAH) (Figure 2.1).

The patients were divided into Non-NF1 RASopathy patients as these have overlapping clinical features amongst each other, which is distinct to NF1 clinical phenotype. Fourty-one non-NF1 RASopathy patients were recruited this included, NS (n=37), NSML (n=1), CFC (n=3) (Figure 2.1). Fifteen NF1 patients that met NF1 clinical criteria were also recruited as well as four patients classified as other in Figure 2.1, which were two patients with clinical features overlapping between NS/CS and NS/NF1, and two patients diagnosed as having an unspecified RASopathy. In addition to the patients, 68 family members were recruited resulting in 11 trios (parents and proband) and 40 duos (2 with father and child, 38 with mother and child). The recruitments were performed by the Division of Human Genetics clinical team led by Professor Amanda Krause. A clinical RASopathy data sheet (Appendix 2) was designed using the main clinical features of Rasopathies and was utilised during patient recruitment as part of the clinical examination to standardise data collection. The importance of participating in the study was discussed with the participants who then signed the consent forms. This was followed by a clinical examination and blood drawing. Blood samples (in EDTA tubes) were taken from

patients and their relatives. Patients' photographs were taken with informed consent, the full list of recruited patients and their relatives in Appendix 3.

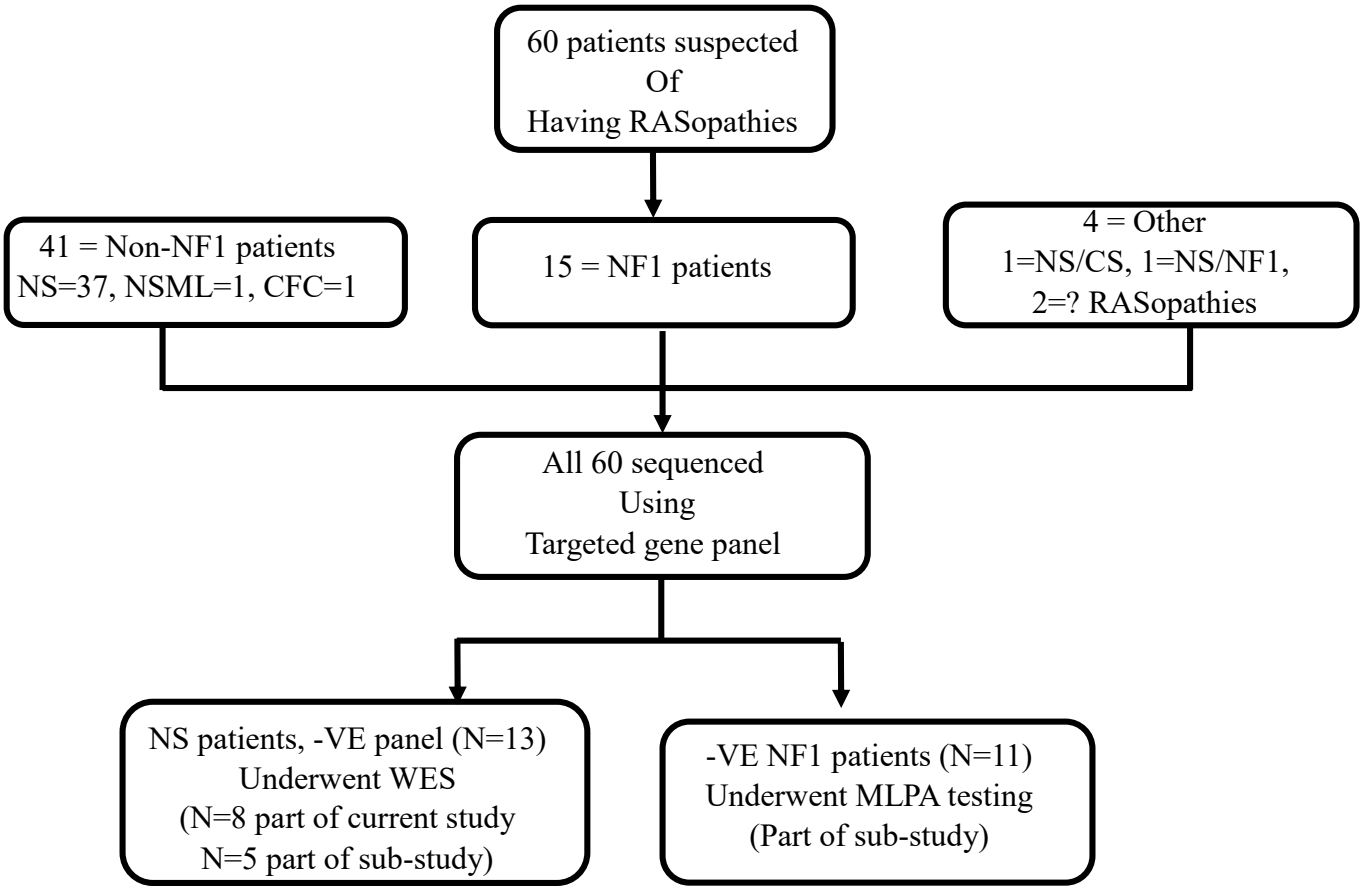


Figure 2.1: Patient recruited, their clinical diagnosis and testing performed

2.2. DNA Extraction

DNA was extracted from the blood samples using the salting out DNA extraction procedure, adapted from Miller, Dykes and Polesky (1988). The blood samples were frozen prior to extraction. In this procedure, cell membranes are broken to expose the DNA from the cell, the process is called cell lysis, detergents and surfactants as well as Proteinase-K enzyme are used in this process. The solution is then treated with a concentrated salt solution to make debris such as broken proteins, lipids and RNA to clump together. Centrifugation of the solution separates the clumped cellular debris from the DNA. DNA is then purified from detergents, proteins, salts and reagents used during cell lysis step using ethanol precipitation. Since DNA

is insoluble in alcohol, it will aggregate together, giving a pellet upon centrifugation. After isolation, the DNA was resuspended in Tris-EDTA (TE) buffer.

2.3. DNA quality control

A Nanodrop 2000 Spectrophotometer (Thermo-Fisher Scientific, Massachusetts, USA), was used for the initial quantification of the genomic DNA. The A260/A280 absorbance ratio of ~1.8 and A260/A280 of ~2.0 was used to assess the purity and quality of the DNA. The DNA samples were also assessed using agarose gel electrophoresis to check the molecular weight of the genomic DNA. DNA quality was assessed again during the library preparation procedure using the Invitrogen Qubit® v3.0 Fluorometer (Thermo-Fisher Scientific, Massachusetts, USA). The Qubit® dsDNA BR (Broad-Range) assay kit (assesses 100ng of double stranded DNA and higher) and Qubit® dsDNA HS (high sensitivity) assay kit (assesses less than 100ng of double stranded) kits depending on the concentration of the DNA sample as per the manufacturer's instructions.

2.4. Targeted gene panel (Agilent and Illumina technologies)

As indicated in Figure 2.1, all 60 patients were sequenced using the targeted panel. This section details how the multi-disease targeted panel was designed and utilised to enrich and prepare patient samples to be sequenced on the Illumina Miseq platform.

2.4.1. Custom-design of the targeted gene panel

Selection of the genes to be included in the custom-designed targeted panel was informed by an extensive literature review that included reviewing published information on genes associated with the disorders of interest. If there was limited information on a gene's role in disease, that is reported in just one family, it was not included on the panel. This was followed by reviewing the guidelines on the ClinGen gene-disease clinical validity classification framework (Strande et al., 2017). Most genes were classified as being definitively associated with the phenotype, but some genes with preliminary evidence indicating possible association with the clinical phenotypes to be studied were also added. These differential diagnosis genes were added as the molecular epidemiology for the target disorders had not yet been characterized in African patients and we wanted to have a broader group of genes for analysis.

The final list of 51 genes included in the panel (Table 2.1) consisted of genes known to be associated with RASopathies (including Noonan syndrome, Noonan syndrome with multiple lentigines, capillary malformation-arteriovenous malformation syndrome, Costello syndrome, cardio-facio-cutaneous syndrome, Legius syndrome and neurofibromatosis); cohesinopathies (Cornelia de Lange syndrome and related phenotypes); facial dysostoses (Treacher Collins syndrome, Nager syndrome and Miller syndrome) and CHARGE syndrome as well their differential diagnoses. The panel was referred to as FRASC targeted gene panel, derived from **F**acial dysostoses, **RAS**opathies and **C**ohesinopathies. The date of the final panel design was 08 August 2018.

The Agilent SureDesign software (version 4.0, released May 2017) (Agilent technologies, CA, USA) was used to design the targeted panel, with a probe region size of less than 499kbp. The probe size of 499kb is the smallest size of the panel (tier one, pricing) and was informed by the number of genes included in a panel and the size of the sequences flanking the coding regions of the genes, details of the panel can be seen in Appendix 4. For the current study, only coding regions of the genes in Table 2.1 were included, with an additional 10 bases from 3' end and 10 bases from 5' end of each target region and the coverage of bases overlapped by probes (extended +/- 100 bp) to represent likely capture

Table 2.1: Genes included in the custom-designed NGS targeted panel, their associated clinical phenotypes and ClinGen association classification (Strande et al., 2017)

Clinical Phenotype	Genes	ClinGen classification
RASopathies	*CBL, LZTR1, KRAS, NRAS, PTPN11, RAF1, SOS1, RIT1, SOS2, RASA1, BRAF, MAP2K1, MAP2K2, SPRED1, NF1, NF2, HRAS, SHOC2*	Definitive association
Treacher Collins	*TCOF1, POLR1C, POLR1D*	Definitive association
CHARGE Syndrome	*CHD7*	Definitive association
Cornelia de Lange	*NIPBL, HDAC8, RAD21, SMC1A, SMC3*	Definitive association
Additional preliminary/ differential diagnosis genes	Facial dysostoses*: SF3B4, DHODH, EFTUD2, SMAD4, SMAD5, SMAD6, SMAD7, EYA1, PDS5A, PDS5B, SCC4, STAG1, ESCO2.* Cornelia de Lange: *AFF4, ANKRD11, KMT2A, TAF1, TAF6, SETD5, SMARCB1, ARID1B.* RASopathies: *RRAS, RASA2, A2ML1*	Limited/Disputed association with the clinical phenotypes.

Definitive association: Genes fully contribute to the associated phenotype

Limited/disputed association: Genes that partially contribute to the associated phenotype

The probe tiling parameters were edited to ensure full coverage of the genes with high complexity and repetitive regions. Probe tiling is the ability of NGS probes to cover the target regions evenly or fully. In highly repetitive regions and GC-rich areas, the designed probes might not hybridise efficiently leading to reduced coverage during sequencing. Agilent SureDesign software uses stringent default probe tiling parameters and in our design, this was suitable for most genes, but a set of seven genes with inadequate coverage using the default parameters of panel design required further adjustments (Table 2.2). Probe tiling parameters were edited to be less stringent in an attempt to achieve full coverage as these seven genes had repetitive sequences.

T**able 2.2:** Genes with 100% coverage after probe tiling parameters were modified

Genes	**Ref Seq transcript**	**Region targeted: GRCh37 (h19) reference genome**	**Size (bp)**	**Coverage Default parameters**	**Coverage after modification**
ARID1B	NM_001371656.1	chr6:157099054-157529035	7150	97.5%	100%
HDAC8	NM_001166419.2	chrX:71549894-71792621	1472	98.7%	100%
PDS5A	NM_001100399.2	chr4:39827095-39978207	4687	98.9%	100%
PTPN11	NM_002834.3	chr12:112856906-112942578	2086	97.4%	100%
RASA1	NM_002890.3	chr5:86564259-86686710	3672	98.6%	100%
SMAD6	NM_005585.5	chr15:66995587-67073883	1571	96.8%	100%
STAG1	NM_005862.3	chr3:136057086-136349750	4437	99.3%	100%

The final panel designed contained 51 genes, the overall size of the panel was 245.950 kbp containing 3375 probes and was optimized using eight patients recruited to the FRASC study (3 Cornelia de Lange, 3 Treacher Collins and 2 RASopathy patient samples).

2.4.2. Library Preparation and Target Enrichment

DNA library preparation process was done using the SureSelect^{QXT} Target enrichment for Illumina paired-end multiplexed sequencing (Agilent Technologies, California, USA) following the manufacturer protocols. Libraries were prepared using 2µl of the ~23ng/µl of starting DNA material, following the manufacturer's protocol. DNA was fragmented using an enzymatic reaction; each fragment was then adaptor-tagged. This was followed by the amplification of the adaptor-tagged DNA library and purification using Agencourt AMPure

XP bead (Beckman Coulter® Life Sciences, California, USA) clean-up protocol. The cleaned DNA libraries were assessed using the Bioanalyzer DNA 1000 HS assay kit on the Bioanalyser 2100 (Agilent technologies, CA, USA) to check if the libraries were within the expected fragment sizes of 245 -345 bp.

The concentration of the libraries was assessed on a Qubit® Fluorometer using the Qubit® dsDNA HS assay kit where the concentration was expected to be ≥50ng/µl. These quality control steps are essential, as an amount of 750ng of the DNA libraries is required to proceed to the next step of hybridization of the designed FRASC panel probes capture to the purified patient libraries. Hybridized amplicons containing targeted regions were captured on Dynabeads™ MyOne™ Streptavidin T1 in solution (Thermo-Fisher Scientific, Massachusetts, USA). Untargeted regions were not hybridized and therefore washed off. Captured libraries with targeted regions were amplified using dual indexing or barcoding primers. This process adds unique tags to the libraries so that it will be easier to identify each sample during data analysis. This was followed by a final AMPure bead clean-up step. A summary of this process is shown in Figure 2.2.

The qubit® dsDNA HS assay kit was used to assess the DNA concentration obtained on the indexed libraries as a first check to assure that the indexed libraries had been successfully amplified. The Bioanalyzer DNA 1000 HS assay kit was then used to quantify and assess the size distribution of the amplified indexed DNA as per the manufacturer's protocol. The electropherograms generated by the Bioanalyzer were assessed to determine if they were within the recommended DNA fragment size of between 325 and 450bp.

The NEBNext® library quant kit for Illumina (New England Biolabs, Massachusetts, USA) was utilised for qPCR to achieve more precise quantitation of the target enriched samples before the samples could be pooled and sequenced on the Illumina platform. The use of the qPCR is more precise than using the Qubit® Fluorometer and Bioanalyser 2100 alone as it measures only the DNA fragments ligated to adaptors and is this gives an accurate concentrations of libraries before pooling. Accurate concentrations of libraries are crucial for the calculation of equimolar amounts for pooling in order to achieve even distribution of libraries. The quantitation of the libraries is critical for the efficient cluster generation in the Illumina flow cell. This kit contains primers that target the P5 and P7 Illumina adaptor sequences and a set of pre-diluted DNA standards to enable the quantitation of the diluted DNA

libraries between 150-1000bp. NEBNext library quant kit for Illumina workflow (NEB#E7630) was utilised following manufacturers guidelines for the quantitation of indexed libraries.

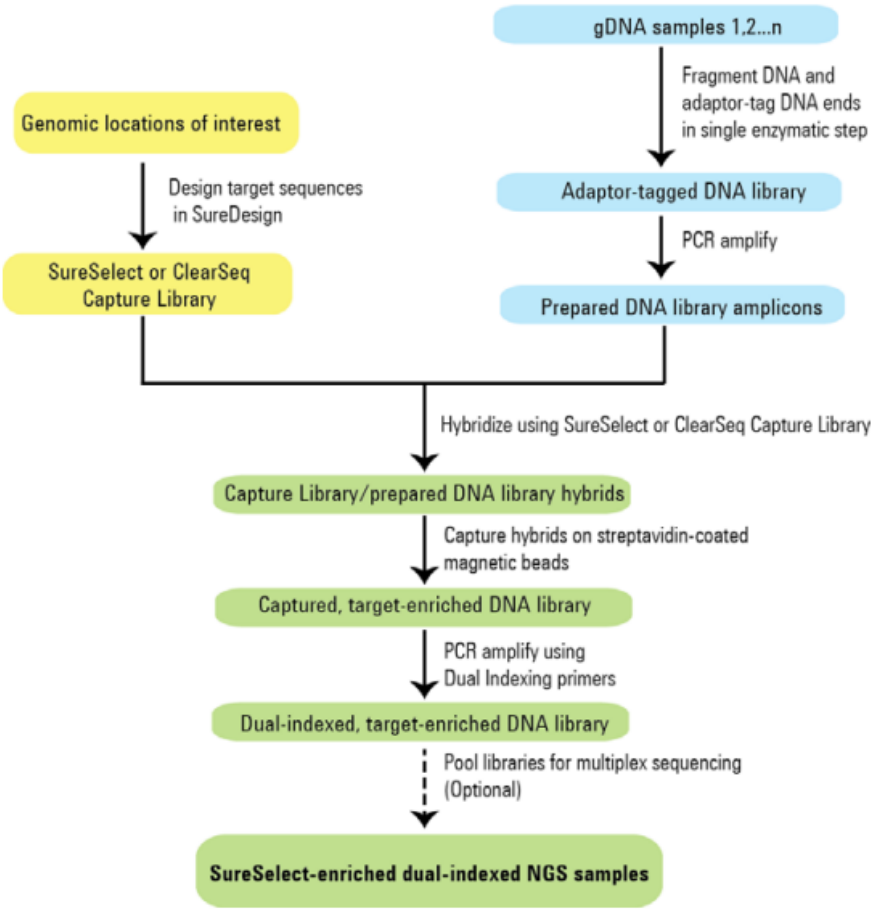


Figure 2.2: Summary of the SureSelect^{QXT} library preparation and target enrichment workflow, obtained from (https://www.agilent.com/cs/library/usermanuals/public/G9681-90000.pdf)

2.4.3. Sequencing of prepared libraries on the MiSeq NGS system

For sequencing on the Illumina MiSeq the libraries must be diluted into equimolar concentrations before they could be pooled for sequencing. The libraries were diluted in a serial dilution until they were at 4ng/µl, and then were pooled together into a 4ng/µl library. The pooled libraries were then denatured and diluted following the Illumina MiSeq system, denature and dilute libraries guide. For the MiSeq nano and micro sequencing kits, the final library concentration loaded on the cartridge was 8pM and for the standard MiSeq system the final library loaded was 12pM. The PhiX control library (Illumina, Inc., California, USA) was also denatured and diluted down to the final concentration of 12.5pM following the Illumina

MiSeq system, denature and dilute libraries guide. The PhiX control library was spiked into the library to be sequenced (570µl of the library and 30µl PhiX library). PhiX is an adaptor-ligated library used as a control for Illumina sequencing runs. The main function of the PhiX is to serve as a sequencing control, providing a quality control for cluster generation, sequencing, and alignment, and a calibration control for cross-talk matrix generation, phasing, and prephasing.

Each MiSeq sequencing kit cartridge comes with Illumina TruSeq primers that work for libraries prepared with Illumina kits. For this study, a SureSelect^{QXT} library preparation kit was used and therefore to ensure efficient sequencing of our libraries Agilent SureSelect QXT read primers 1 and read primer 2 as well as the index primer (3µl of each primer) were spiked into the Illumina TruSeq primers (597µl) wells (12, 13 and 14) on the cartridge. The sample (prepared libraries combined with the PhiX library) was then sequenced on the Illumina MiSeq Instrument (Illumina, Inc., California, USA) using the nano-, micro- or the standard MiSeq reagent v2 kits chemistry to obtain paired-end 2x150bp fragment using standard sequencing parameters and following the instrument quality control parameters. The MiSeq nano and micro kits were used to sequence a maximum of eight samples in low throughput sequencing reactions whereas the standard MiSeq reagent v2 kit was used to sequence 24 samples in high throughput sequencing reactions. These three different sized MiSeq sequencing kits were used in order to sequence various batches of libraries, from eight to 24 samples as when needed.

2.4.4. Quality control of NGS data

After a first round of analysis, the same variant was found in a number of patient samples and thus suspected as being a false positive, Sanger sequencing was used to confirm this. False positives were resolved by adjusting the VAF threshold from 10% to 30% and reanalysed (based on the manufacturer's instructions). The data was initially analysed on Agilent SureCall using the default VAF threshold (10%). Samples were re-analysed on the SureCall using two adjusted VAF thresholds of 20% and 30%. The VAF threshold was adjusted by changing the Minimum Allele Frequency from 0.1 to 0.2 and 0.1 to 0.3 (representing a VAF of 20% and 30% respectively) in the SNPPET SNP caller settings prior to analysing variants. The VAF threshold was adjusted a second time by changing the Minimum Allele Frequency to 0.3 (which represents a VAF of 30%) prior to analysing variants a third time. All targeted panel data in the study was then reanalysed using the VAF of 30% on the SureCall analysis software, as this was the recommended allele frequency threshold.

2.4.5. Targeted panel data analysis and interpretation

Read alignment and variant calling was done using the Agilent SureCall data analysis software (version 4.2.0.18) using default parameters. The MiSeq instrument has built in software that performs primary analysis including the demultiplexing of the indexed reads and generates FASTQ files. The FASTQ file served as an input file on the SureCall data analysis software. This workflow aligned the reads from FASTQ file to the human reference genome, hg19 (GRch build 37) and provided variant call files (VCFs) and Binary alignment files (BAM) output. Human genome hg19 was used to align the reads and not the newer hg38 as the panel SureDesign and SureCall software was based on hg19 at the time, the example of the quality control report of a sample analysed on the Agilent SureCall software can be seen in Appendix 5, showing duplicate information and other calling metrics.

The VCF file of each sample was annotated using wANNOVAR (Chang and Wang, 2012) and Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016a). Common variants were filtered out using a 0.025% minor allele frequency from population databases, The minor allele frequency cut-off was informed by the ACMG-AMP guidelines (Richards et al., 2015) revised for RASopathies by the ClinGen expert panel (Gelb et al., 2018). Other parameters used for filtering included, *in-silico* protein function predictors (variants with damaging or intolerated protein effect were prioritised) and mode of disease inheritance (Autosomal dominant for RASopathies). In addition, variant classification on ClinVar disease variant database and using the ACMG-AMP variant classification criteria (only VUS, likely pathogenic and pathogenic variants were prioritised) and the shortlisted variant were then interpreted against the clinical phenotype. The summary of the filtering strategy for this data is summarised in Table 2.3

Table 2.3: Summary of filtering parameters used to shortlist targeted gene panel variants

Filtering parameters	Shortlisted variants criteria
Minor allele frequency	0.025% (RASopathy ClinGen)
Protein function	Multiple in-silico predictions: Damaging/Intolerated
Inheritance pattern	Autosomal dominant for RASopathies
Variant type	Missense variants for Non-NF1 RASopathies
Clinical Significance	VUS, Likely pathogenic and pathogenic based on ACMG-AMP criteria
Prioritised variants	Interpret against clinical phenotype

All putative disease-causing variants were visualised using Integrative Genome Viewer (IGV) (Robinson et al., 2017) and evaluated for sequencing depth coverage and quality metrics. Putative disease-causing variants were then classified using the ACMG-AMP and the ClinGen RASopathy expert panel guidelines to identify putative disease causing variants. Varsome (Kopanos et al., 2019) was used to compare manual assignment of ACMG-AMP codes to the on-line automated ACMG codes application. This process is summarised in Figure 2.3 below.

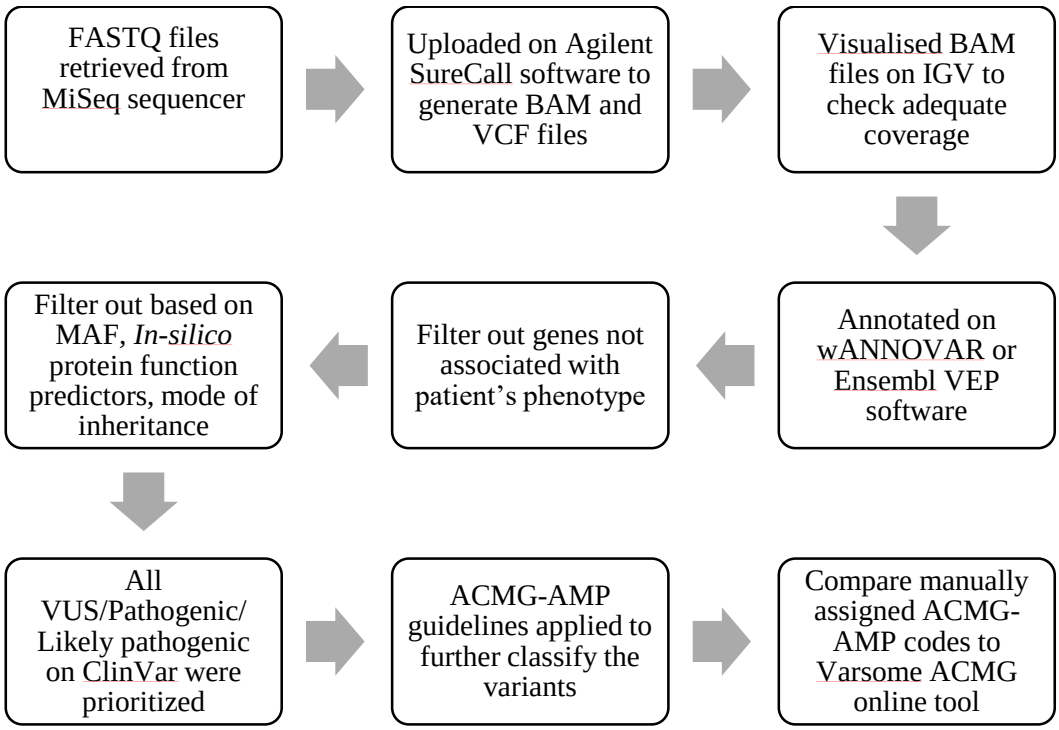


Figure 2.3: Summary of analysis of data generated by the targeted gene panel

2.5.Costing Analysis

Analysis of costs in the current study was done by assessing the costs of the negative testing done prior to this study in the patients where clinically significant variants were identified. This is to illustrate that resources were being inefficiently used on inappropriate testing. An estimation of how much testing would cost for these group of disorders if we were to sequence each gene using Sanger sequencing was also done. This was informed by the NHLS costing model as well as the test price list of the National Health Laboratory Service, South Africa.

The estimation of the current costs per sample on the Illumina platform were obtained from the https://emea.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-cost.html, Illumina website page using targeted panel of ~50-60 target regions/genes using MiSeq platform and the Nano reagent kit is ~ US$200 excluding data analysis and interpretation.

2.6. Whole Exome Sequencing (WES) using Ion Torrent ThermoFisher technology

WES was performed on a subset of the Non-NF1 RASopathy samples (n = 13) that had no mutations identified using the targeted gene panel mutation screen. Eight of the thirteen was sequenced as part of the current study and five samples formed part of a sub-study (MSc supervised by the current PhD candidate). The eight samples were selected as they had clinical features suggestive of NS and yet no variants were identified using the panel that contained common genes associated with NS. Further testing using WES was necessary on these to determine if there are other genes or rare variants that could not be detected using the targeted panel (Figure 2.1). The WES laboratory workflow includes library preparation, templating and chip loading followed by sequencing of the chip, which was done using the Ion Chef™ Instrument and the Ion S5™ sequencer (Ion Torrent by ThermoFisher Scientific, California, USA).

2.6.1. Library preparation

The library preparation for whole exome sequencing was done following the Ion AmpliSeq™ Exome RDY workflow. DNA concentration was measured using the Qubit™3.0 Fluorometer broad range kit. The total concentration of 50ng-100ng genomic DNA is required for each exome library target amplification reaction; 14µl of the 5x Ion Ampliseq™ HiFi mix was added to the 56ul DNA sample to a maximum volume of 70µl that was then added to the Ion AmpliSeq™ Exome RDY plate (1x8) (Ion Torrent by ThermoFisher Scientific, California, USA). The amplified libraries were then combined and digested using the FuPa reagent, switch solution and DNA ligase were used to ligate barcode adapters to the libraries. Barcodes used were from the Ion Xpress™ Barcode Adapter 1-16 Kit (ThermoFisher Scientific, California, USA). After the ligation, the libraries were purified using the Agencourt™AMPure™XP reagent and quantified using the Qubit™3.0 Fluorometer high sensitivity kit, the libraries were diluted down to ~22ng/mL (~100pM).

2.6.2. Template preparation and chip loading

Template preparation and chip loading was performed on the Ion Chef™ Instrument using the Ion 540™ Kit-chef and Ion 540™ Chip kits (ThermoFisher Scientific, California, USA) as per the manufacturer's instructions. Two patient libraries (12.5µl each) were pooled together to make 25µl per Ion 540™ chip. Four samples can be sequenced in one whole exome sequencing run using two Ion 540™ chips. For this particular study, two sequencing runs were performed to sequence eight patient samples.

2.6.3. Sequencing using the Ion S5™ GeneStudio system

The sequencer was initialised following manufacturer's instructions. This process involved placing all the buffers (wash buffer and cleaning solution), previously used Ion 540™ chip and sequencing cartridges needed for the sequencing of the chip. After the initialization run (45min) was complete, the previously used chip was removed and the new prepared chip was loaded on the Ion S5™sequencer (ThermoFisher Scientific, California, USA). Primary and secondary data analysis happened on the Torrent suite software (version 5.12.2) incorporated on the Ion S5 platform, created by ThermoFisher Scientific, California, USA.

2.6.4. WES data analysis and interpretation

When the WES run was complete, the data was collected on the Torrent-Suite data analysis software. Aligned BAM files and VCF files were produced as an output from the Torrent Suite software. The filtering strategy employed for the WES data is summarized in Figure 2.4. Additional annotation software was employed during analysis of WES data; as it was not clear what is expected, particularly because RASopathy features overlap with other syndromes. For targeted panel one annotation software is sufficient, as the output is smaller than that of the WES and easier to manipulate and target only the gene of interest.

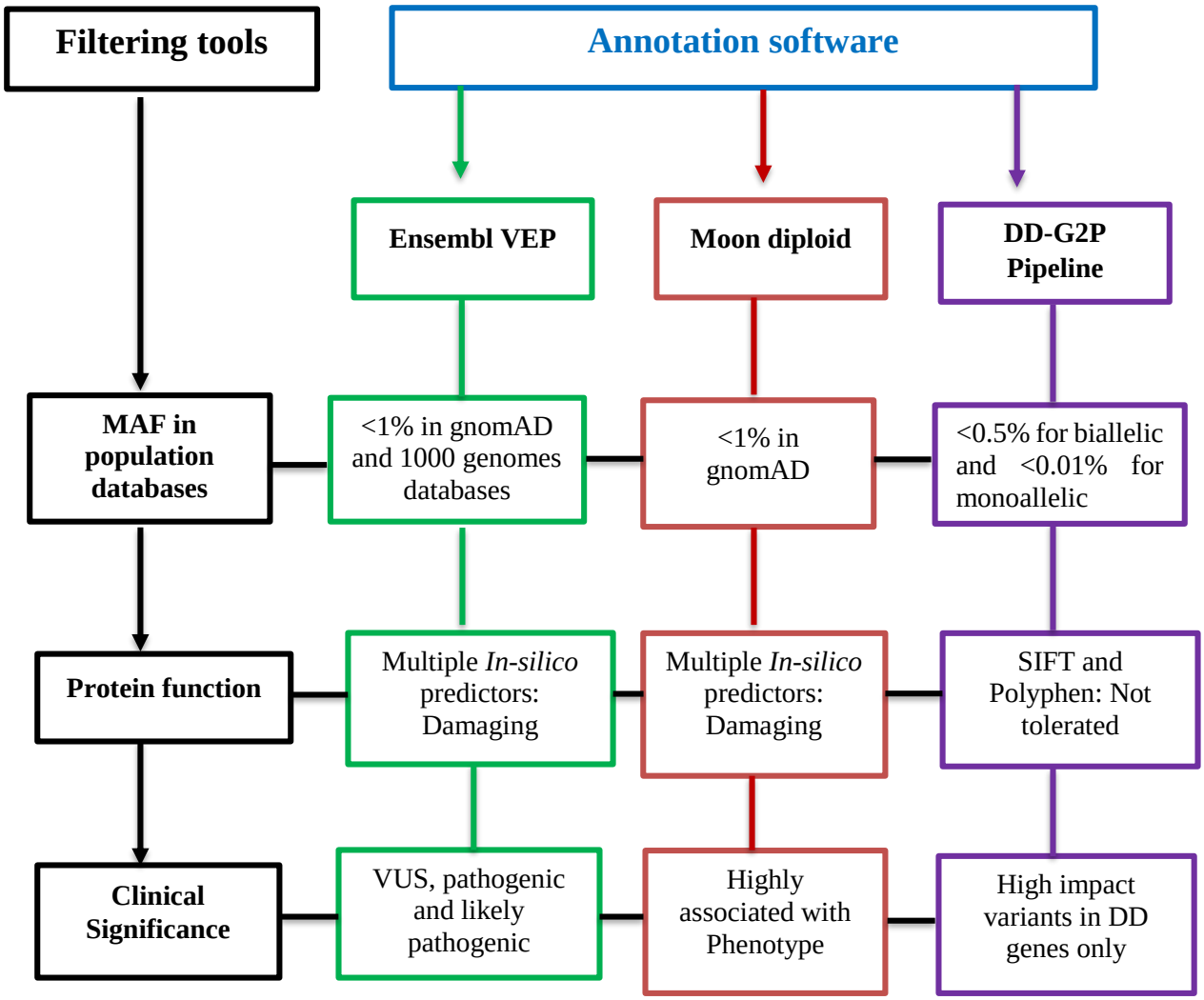


Figure 2.4: Summary filtering strategy used for variants generated by whole exome sequencing

Different analysis approaches were utilised for WES data to identify common variants of interest that may be observed in these three approaches; the Ensembl VEP (McLaren et al., 2016), the moon diploid (Lek et al., 2016) and the DD-G2P pipeline (Thormann et al., 2019). Figure 2.4 shows annotation software of three approaches employed for WES data analysis and the filtering parameters. Frequency of prioritised variants on Ensembl VEP; variants with MAF of less than 1% in gnomAD and 1000 genome datasets, moon diploid prioritised variants with less than 1% in gnomAD only and the DD-G2P pipeline prioritise variants with <0.5% for biallelic and <0.01% for monoallelic variants. Variants where multiple *in-silico* tools predicts damaging to the protein function were prioritised for ensemble VEP and moon diploid whereas for theDD-G2P the prioritised variants are those not tolerated by the protein using SIFT and

Polyphen protein function prediction tools. The significance of the variants is considered if they were VUS, likely pathogenic or pathogenic for Ensemble VEP through ClinVar disease variant database. For Moon diploid; the variants highly associated the phenotype are prioritised and the DD-G2P pipeline considered variants with high impact in DD genes only. Variants that were observed in all these three approaches were further prioritised and scrutinised.

a) Ensembl Variant Effect Predictor

The VCF files for the WES samples were first annotated on the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016a). Variants were filtered using standard filtering parameters such as allele frequencies in the population databases; this resulted in more than 200 variants observed per sample in genes that were not typically associated with RASopathies.

b) Moon by Diploid software

Moon by Diploid software, version 3.3.3 (http://www.diploid.com/moon) was utilised to assist with attempting to reduce candidate variants for analysis. A VCF file together with the patient's phenotypic features (HPO terms) were needed on the Moon analysis software. This is a more phenotype focused analysis of the WES data and clinical information of the patient is critical for this software to select the variants that are relevant to the HPO terms entered. A good clinical work up on the patients including family history and other parameters helps to reduce the variants and increases the chances of identifying the possible variant causing the disorder in less time using this method.

c) The Developmental disorder Gene2Phenotype (DD-G2P) VEP plugin

Following Moon software analysis and the on-line VEP analysis, variants identified were not in genes typically associated with the RASopathies but rather with random developmental disorders. This led us to exploring the G2P VEP plugin that was added using the command-line with the assistance of the DDD Africa team from the University of the Witwatersrand (Wits). The Variant Call Format (VCF) files were uploaded onto the Wits research computer cluster (ZA-Wits-Core cluster) where they underwent variant annotation using the command line version of Variant Effect Predictor (VEP). The VEP computes the consequence of a variant and can further annotate the variant with additional plugins. A Gene2Phenotype (G2P) VEP plugin was added to the script to filter for variants in likely disease-causing genes found in the G2P Developmental disorder (DD) panel database. The database contains 2429 genes known

to be associated with DD that have been constructed from the literature (Thormann et al., 2019). The plugin filters in variants that overlap with G2P genes, if listed as a severe consequence and if the allele frequencies in the reference populations are below a given threshold (bi-allelic gene: 0.005 and mono-allelic gene: 0.0001) will be filtered-in and analysed further https://www.ebi.ac.uk/gene2phenotype/.

d) Interpretation of the WES variants using ACMG-AMP guidelines

Prioritized variants were then classified as pathogenic, likely pathogenic or VUS using ACMG-AMP guidelines. The general ACMG guidelines (Richards et al., 2015), were used because the variants were not identified in genes that are commonly associated with RASopathies. Varsome, the human genome variant search engine (Kopanos *et al.*, 2019), was used to verify the ACMG classifications. Variants were also checked against ClinVar and HGMD disease variant databases to determine if they are existing or new. The whole process taken to perform WES in this study is summarised in Figure 2.5

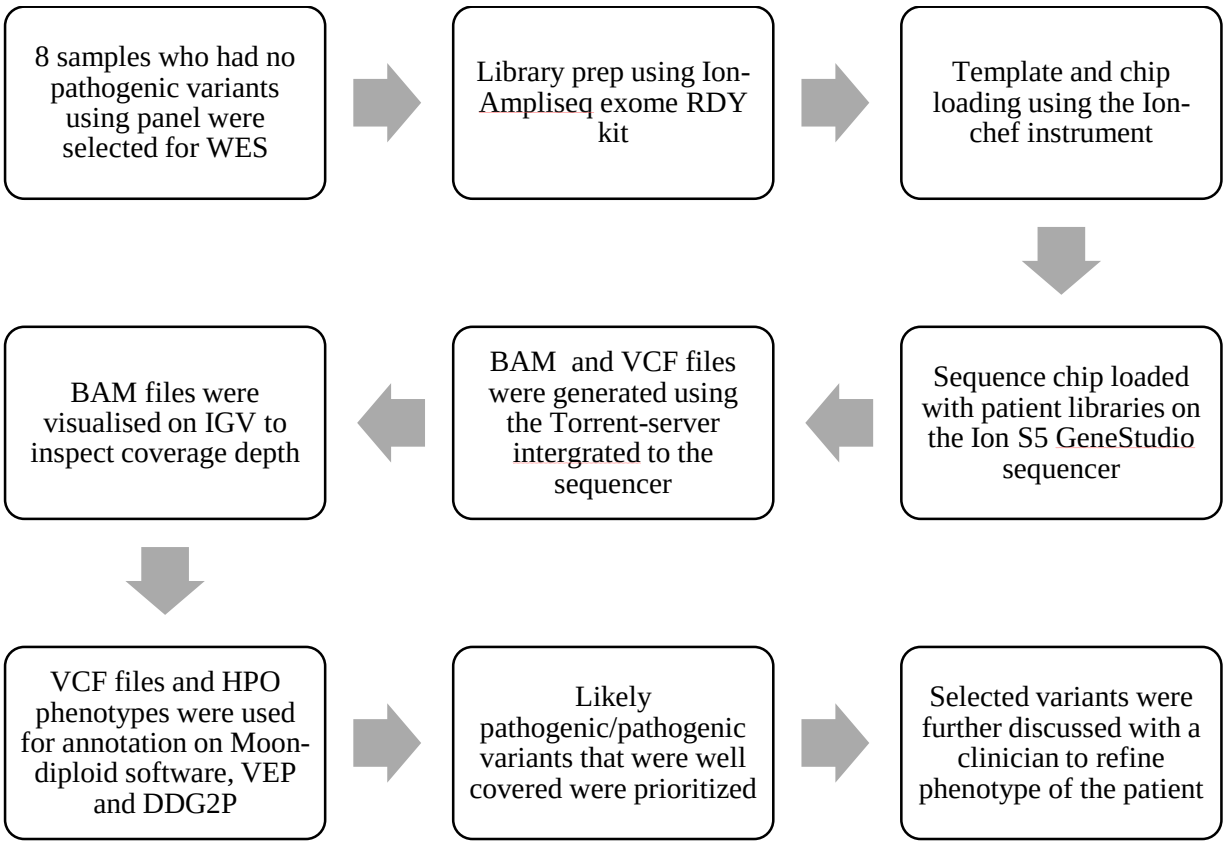


Figure 2.5: Summary of the whole exome sequencing process, from wet-lab to analysis

2.7. Sanger sequencing validations

Sanger sequencing was used to validate the presence of putative causal mutations identified by both the targeted panel and WES workflows. Where possible segregation analysis was performed by analysing whether pathogenic variants were present in family members. The primers for each clinically significant variant were designed using the primer3web version 4.1.0 software (Rozen and Skaletsky, 2000) targeting the area where the mutation was identified (Appendix 6). The DNA sample extracted from the patients' and their relatives' (where applicable) were amplified, followed by the purification of PCR products using Exo-SAP express kit (ThermoFisher Scientific, California, USA). Cycle sequencing of the purified products was performed using the BigDye v3.1 kit (ThermoFisher Scientific, California, USA). The resulting products were then purified using the Montage PCR_{96} plate and injection solution (Millipore®, Massachusetts, USA). Samples were then loaded on the 3500xl Genetic Analyzer to sequence using capillary sequencing technology (ThermoFisher Scientific, California, USA). Raw sequencing data was analysed using the CLC Bio software version 7.7.1 (Qiagen, Hilden, Germany).

CHAPTER 3

Results

3.1. Demographics of patients recruited

This section describes the demographics of the patients recruited in the current study as well as their clinical diagnoses. A total of 60 patients, 21 female and 39 male patients were recruited: Of the patients recruited, 45 had non-NF1 RASopathies and 15 had a clinical diagnosis of NF1. In addition to the patients, 69 family members were recruited and this resulted in 13 trios (parents and proband) and 39 duos (2 with father and child, 37 with mother and child). Eight adult patients were recruited as singletons and did not have any parents recruited. The demographics of these patients are summarised in Table 3.1, full list can be found in Appendix 3

Table 3.1: Summary of demographics of recruited patients

RASopathy patients	N=60
Females	21/60 (35%)
Males	39/60 (65%)
Non-NF1	45/60 (75%)
NF1	15/60 (25%)
Mean Age at recruitment (Non-NF1 RASopathies)	9y3m (5m-33yrs)
Mean Age at recruitment (NF1 patients)	14y4m (3-29yrs)
Trios	13/60 (21.5%)
Duos	39/60 (65%)
Singletons	8/60 (13.3%)
Relatives	69 (65 parents and 4 other)
African ancestry	56/60 (93.3%)
Caucasian	3/60 (5%)
Asian	1/60 (1.6%)

NF1: Patients presenting with NF1 clinical features, **Non-NF1**: Patients presenting with features suggestive of Noonan syndrome and other RASopathies except NF1, **Trios**: Patients and both parents recruited; **Duo**: Patient and one parent recruited, **Singletons**: Adult patients recruited without parents, **Other:** daughters, grandparent

3.2. Design and optimisation of the FRASC panel

In this section, the details of the design of the final custom-designed targeted panel are outlined and how it was first used to assess the performance of our library preparation and sequencing protocol. The probe region size of the custom NGS panel was 249.95kbp, containing 3375 total probes (Appendix 4) making the panel appropriately sized to work on most of the smaller benchtop NGS systems available to the researcher at the time of the study. The size of first tier priced panel is usually from 0-499kbp (e.g. Illumina) and ours is approximately half of that.

The performance of the panel was optimised using three patients clinically diagnosed with Cornelia de Lange syndrome (CdLS), three with Treacher Collins syndrome (TCS) and two

with suspected Noonan syndrome (NS) to assess how the protocol performs for a range of target genes in our setting and how to optimise in subsequent runs.

The sequencing run was performed using the MiSeq Reagent Kit v2 Nano, which is used for low output sequencing with an expected data output of ~0.3Gb (300Mb) and cluster density for this is 600-800K/mm2. The cluster density measures the efficiency of the sequenced library clusters passing the filters (PF) on the sequencing chip or flow cell, the first optimisation run in the current study had cluster PF of 95.93%. This eventually influences the run quality as underclustering may results in low sequence yield and poor coverage and overclustering may results in poor sequence that cannot be analysed. Good cluster density highly depend on the concentration of DNA libraries sequenced, its important during the DNA library pooling step to have accurate concentrations. The Illumina Sequence analysis viewer (SAV) QC metrics of the first optimisation run including a cluster density of 631k/mm2 and ~94% of bases with quality score of Q30 are shown in Figure 3.1.

The output achieved for the optimization run was 253.6Mb as can be seen in Figure 3.1. Although it is lower than the expected output, more that 94% of the target area had high quality reads.

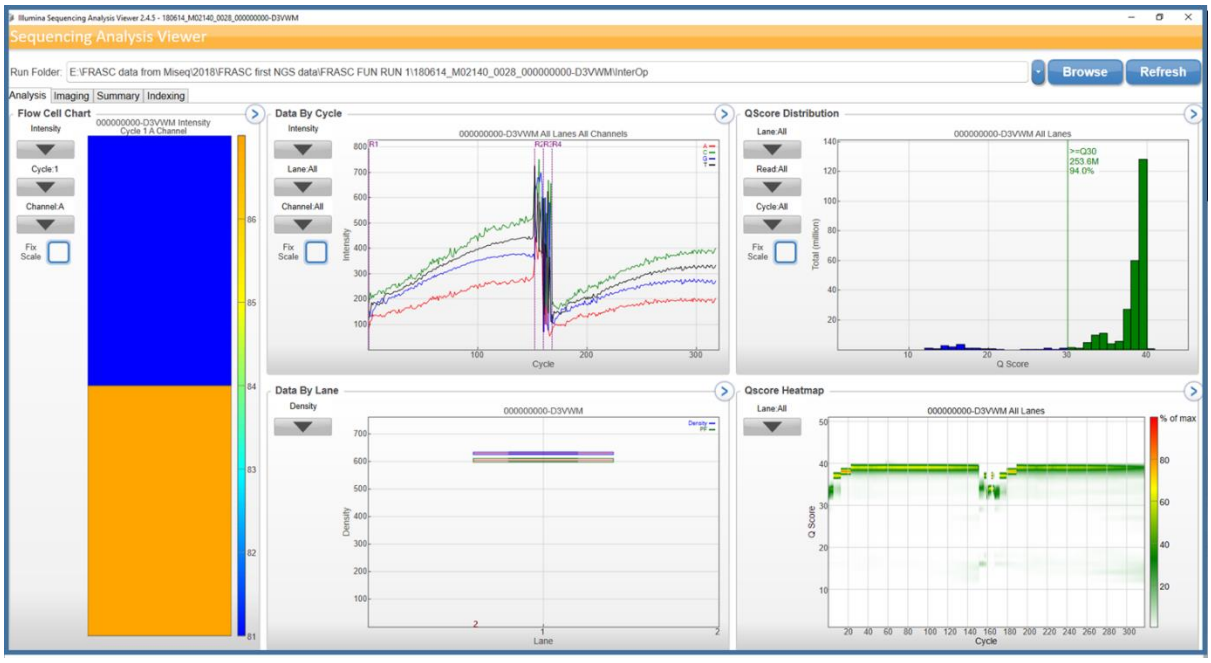


Figure 3.1: Sequence analysis viewer quality graphs, obtained for FRASC run1 on the MiSeq instrument

With the cluster density within expected range, this run was rendered successful and the coverage as well as the quality score of each sample included in this run was analysed further as shown in Table 3.2. All samples had good quality data with Q score of 39 indicating 99.9% accuracy in the base calling of the data. The output and the quality of data achieved was acceptable based on the size of the MiSeq reagent kit used (https://emea.illumina.com/systems/sequencing-platforms/miseq/specifications.html). The minimum read depth received for the optimisation run ~37x which is within the maximum acceptable read depth of 30x for the targeted gene panel (Table 3.2).

Table 3.2: FRASC samples used in the first panel validation run showing Q-scores and read depth obtained

Patient	Clinical diagnosis	Q-Score	Read depth
FRASC1	NS	39	50x
FRASC2	NS	39	51x
FRASC8	CdLS	38	52x
FRASC13	TCS	39	56x
FRASC27	TCS	39	38x
FRASC28	TCS	39	37x
FRASC32	CdLS	39	50x
FRASC41	CdLS	39	64x

NS=Noonan syndrome, CdLS=Cornelia de Lange syndrome and TCS=Treacher Collins syndrome

3.2.1. Costing analysis

The cost effectiveness of the strategy in the current study is shown in Table 3.3, which is showing testing that was done prior to the patients being enrolled in the current study. Of the 33 patients with pathogenic/likely pathogenic variant, 14 had prior testing; meaning 42% had inefficient testing prior to this study. At the time of the costing analysis the exchange rate between South African Rands (ZAR) and United States Dollar (USD) was approximately OneZAR = 14,758USD.

Table 3.3: Costs of tests done prior to the current study

Testing done prior to enrolment in current study	Patients who had prior testing	Price in ZAR of patients who had prior testing	Price in USD of patients who had prior testing
MLPA for microdeletion / duplications syndromes	3	12542,54	849,54
QF-PCR for aneuploidy testing	2	7332,00	496,61
Chromosomal analysis	2	5078,19	344,01
FISH testing for Di-George syndrome	2	5180,1	350,93
Mucopolyssacharides enzyme testing	2	329,8	20,23
Cystic fibrosis testing	2	3666,00	248,36
NF2 whole gene sequencing	1	9000,00	609,85
Total	14	43128,63	2901,53

Costing was done using NHLS costing model at ~14,758 USD vs 1 ZAR exchange rate

If we were to do Sanger sequencing in the eighteen genes with definite association to RASopathies, at ~R1242,62 (US$73,40) per region/exon, based on single regions Sanger sequencing utilizing both reverse and forward primers, the estimated prices are shown in Table 3.4.

Table 3.4: Showing estimations of prices if Sanger sequencing were to be used to analyse all the exons of the genes associated with RASopathies

Gene	Number of exons	Estimated Sanger sequencing price (ZAR)	Estimated Sanger sequencing price (USD)
BRAF	18	22367,16	1321,2
CBL	16	19881,92	1174,4
HRAS	6	7455,72	440,4
KRAS	6	7455,72	440,4
LZTR1	11	13668,82	807,4
MAP2K1	12	14911,44	880,8
MAP2K2	11	13668,82	807,4
NF1	61	75799,82	4477,4
NF2	17	21124,54	1247,8
NRAS	7	8698,34	513,8
PTPN11	15	18639,3	1101
RAF1	16	19881,92	1174,4
RASA1	25	31065,5	1835
RIT1	6	7455,72	440,4
SHOC2	26	32308,12	1908,4
SOS1	24	29822,88	1761,6
SOS2	13	16154,06	954,2
SPRED1	7	8698,34	513,8
Total	**297**	**369058,14**	**21799,8**

It can be seen in Table 3.4 that the cost of sequencing all the common genes associated with RASopathies would be high and is unlikely to be cost-effective. This is the reason why NGS is a breakthrough for this group of disorders. Based on the prices of reagents and consumable purchases for the current study using the targeted panel, the cost per sample was ~R14 186,20 (US$849, 68) and this is considerably cheaper, considering that all RASopathy gene exons and other additional genes are sequenced simultaneously in one reaction and that the price is little more than sequencing 1-2 of the smaller genes.

3.2.2. Utilising the FRASC panel to sequence RASopathy samples

Upon the successful validation of the panel, the rest of the RASopathy samples were sequenced using different MiSeq sequencing kits. MiSeq Reagent Kits v2 are ready to use sequencing cartridges, they have different data outputs and can be used for small batch of samples (from eight to 24 samples). Expected data outputs for the MiSeq Reagent nano, micro and standard sequening kits are 300Mb, 1200Mb and 5000Mb respectively.

As shown in Table 3.5, we were able to achieve adequate coverage (58x – 180x average read depth) of good quality (more than 90% of the reads had quality score of >Q30) while utilising MiSeq Reagent kits (Table 3.5). The number of variants obtained after annotation per sample (~170 variants) were similar regardless of the kit used for sequencing, this indicates that each sample had sufficient/ uniform amount of sequencing and there was no bias occurring during sequencing.

Table 3.5: Quality metrics obtained using MiSeq reagent v2 kits

Parameters	MiSeq Reagent Kit v2 Nano	MiSeq Reagent Kit v2 Micro	MiSeq Reagent Kit v2
Samples sequenced	6-8 per run	8-16 per run	16-24 per run
% >Q30	94.0	94.3	90.5
Cluster density	628	742	874
Average read depth	58x	101x	180x
Average number of variants identified per sample	170	169	174

3.2.3. Quality control of NGS data

This section explains processes taken to ensure that false positive variant calls were removed from data analysis. Variants with skewed variant allele fractions (VAF) and low read depth were filtered out and removed and not analysed together with true variants. Variants with skewed allele fractions were observed in *PTPN11* and *NF1* genes and further details of their analysis and troubleshooting are described in this section. Quality parameters such as read depth and VAF were analysed by manual inspection of the BAM files in IGV. The expected VAF for heterozygous loci should be near 50%, and for homozygous loci should be near 100%, and reference loci should be near 0%. Any deviations from these expected values should be assessed further to eliminate any form of false variant calls. Sanger sequencing was utilised to further confirm the true positives (variants that was confirmed using Sanger sequencing) from the false positives (variants that were detected during variant annotation but not observed using Sanger sequencing).

a) *PTPN11*: c.923A>G (p. N308S) variant

A *PTPN11*: c.923A>G variant was observed in 10 out of 45 Non-NF1 patients. Inspection of the BAM files at the variant site on IGV indicated that the allele fraction for this variant was skewed in 9 patients indicating false-positive calls (Table 3.6), even though the average read coverage is 195x. Sanger sequencing was performed in all samples with this variant and only one was found to be a true variant (FRASC66).

Table 3.6: Patients with skewed allele fraction at *PTPN11*: c.923A>G (p.N308S) variant site and FRASC66 (highlighted) showing true positive variant.

Patient	Dx	False positive variant	Read depth	Reference allele 'A'	Alternative allele 'G'
FRASC 3	NF1	*PTPN11*:c.923A>G	121x	99 (82%)	22 (18%)
FRASC38	NF1	*PTPN11*:c.923A>G	138x	119 (86%)	19 (14%)
FRASC42	NF1	*PTPN11*:c.923A>G	274x	212 (77%)	61 (22%)
FRASC43	NF1	*PTPN11*:c.923A>G	277x	226 (82%	50(18%)
FRASC44	NF1	*PTPN11*:c.923A>G	192x	166(86%)	26(14%)
FRASC62	NF1	*PTPN11*:c.923A>G	204x	132 (79%)	35(21%)
FRASC66	NS	*PTPN11*:c.923A>G	186x	103 (56%)	81(44%)
FRASC84	NS	*PTPN11*:c.923A>G	66x	54 (82%)	12 (18%)
FRASC91	NS	*PTPN11*:c.923A>G	221x	155 (70%)	66 (30%)
FRASC99	CFC	*PTPN11*:c.923A>G	258x	210 (81%)	47 (18%)

One example of the IGV plot showing the appearance of the false-positive variant is shown in Figure 3.2 below of FRASC91 which had 70% and 30% allele fraction and the variant was not observed on Sanger sequencing.

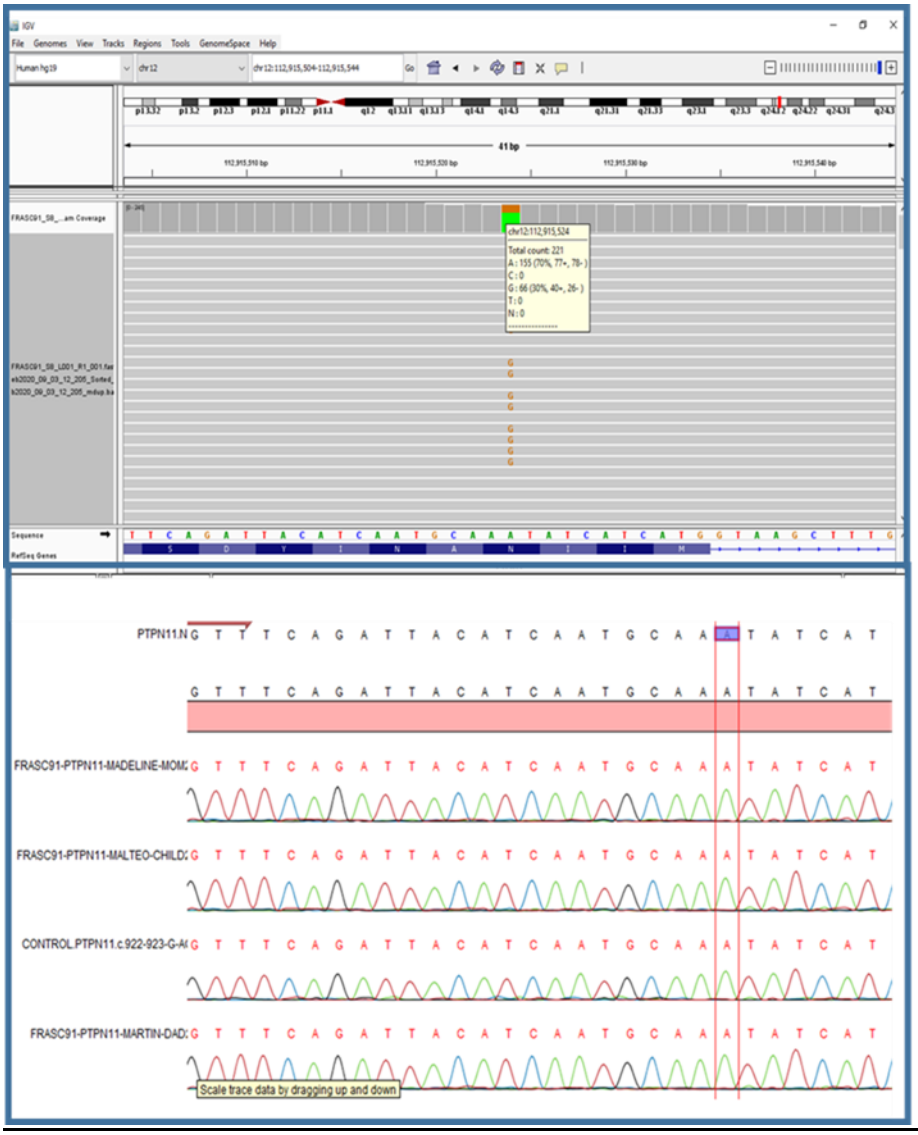


Figure 3.2: IGV image and Sanger sequencing electropherogram for FRASC91, one of the false positive variants for identified during data analysis (not observed on Sanger sequencing and also IGV showing uneven allele fraction)

The *PTPN11*: c.923A>G (p.N308S) variant was confirmed in FRASC66. This result is shown in (Figure 3.3). The allele fraction was 56% A and 44% G, which was not skewed from the expected.



Figure 3.3: IGV image and Sanger sequencing electropherogram for FRASC66, showing a true positive variant (*PTPN11*: c.923A>G) identified and confirmed on Sanger sequencing.

b) The *NF1:* c.27G>A (p.W9X) variant

This variant was observed in 9 patients (Table 3.7). Of the nine this variant was confirmed to be a true variant in one patient, FRASC67 (Figure 3.4).

Table 3.7: Eight patients with *NF1:* c.27G>A (p.W9X) false positive variant, IGV coverage as well as the respective allele fraction and FRASC67 (highlighted) showing the true positive variant.

Patient	Dx	False positive variant	Read depth	Reference allele 'G'	Alternative allele 'A'
FRASC65	NF1	*NF1*:c.27G>A	26x	8 (31%)	18(69%)
FRASC67	NF1	*NF1*:c.27G>A	48x	33 (69%)	15(31%)
FRASC68	NF1	*NF1*:c.27G>A	48x	27 (56%)	21 (44%)
FRASC70	NF1	*NF1*:c.27G>A	59x	32 (54%	27(46%)
FRASC82	NF1	*NF1*:c.27G>A	34x	21(62%)	13(38%)
FRASC83	NF1	*NF1*:c.27G>A	43x	18 (42%)	25(58%)
FRASC89	NF1	*NF1*:c.27G>A	114x	69 (61%)	45 (39%)
FRASC94	NS	*NF1*:c.27G>A	73x	45 (62%)	28(38%)
FRASC95	NF1	*NF1*:c.27G>A	59x	33(56%)	26(44%)

It can be seen in Table 3.7 that the average read depth at the variant site of all the nine samples is 56x which was much lower compared to the coverage of other exons in the *NF1* gene, observed on IGV during manual inspection (data not shown). This can be explained by the fact that this variant was found in exon 1 on the 27th nucleotide, which is at the very beginning of the gene. It is previously reported that sequencing at the beginning or end of the gene may have lower coverage. This is because it is closer to the primer binding site and the point where polymerase activity is just ending / starting as compared to the middle of the sequence where the activity is optimal (Hert, Fredlake and Barron, 2008; Ari and Arikan, 2016)

The allele fraction of the few reads obtained per sample was also skew compared to the expected heterozygous allele ratios. Poor coverage depth and uneven allele fraction could have been the cause for false positive calls. The *NF1*:c.27G>A variant was confirmed in one patient, FRASC67 using Sanger sequencing (Figure 3.4). This variant was also observed in the mother who was not formerly assessed prior to the enrolment to the current study, therefore it is unknown if she has NF1.

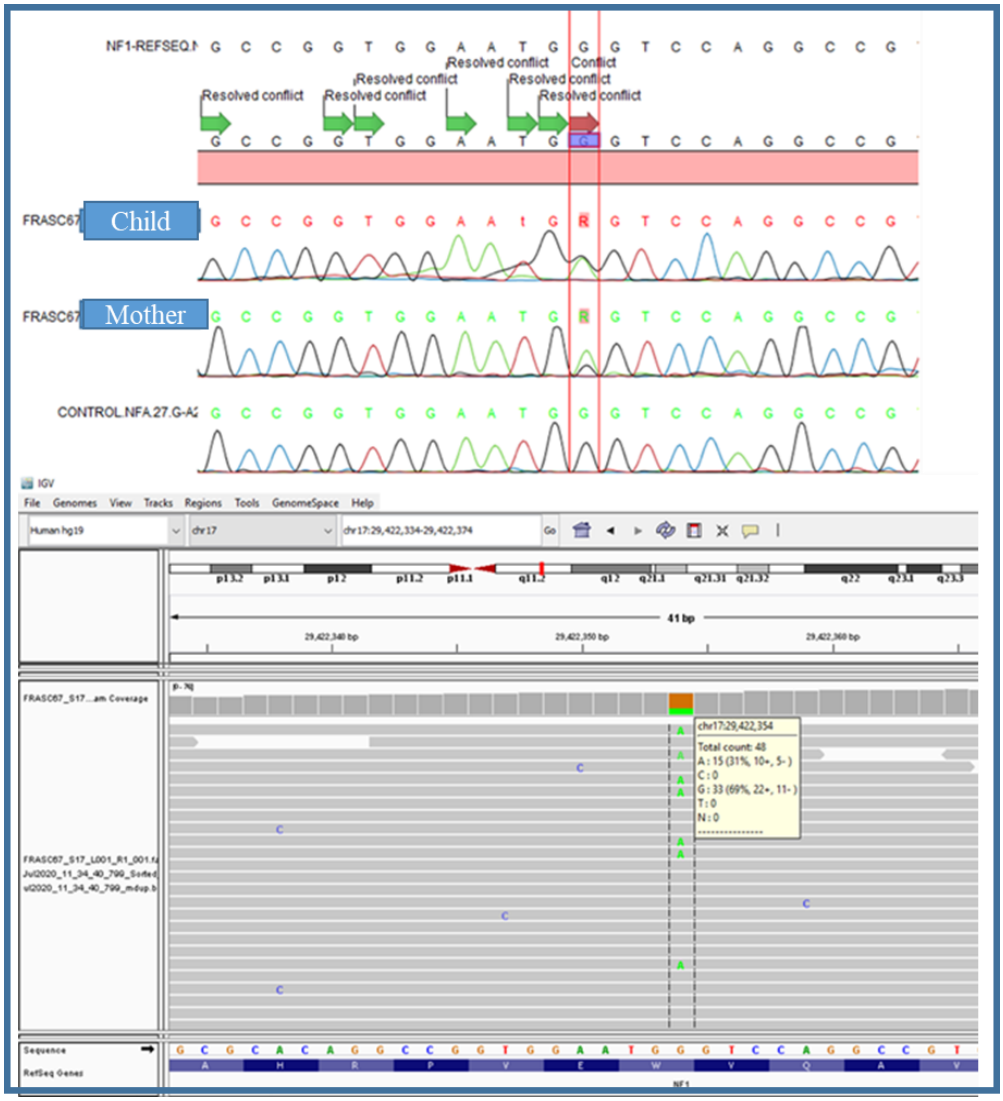


Figure 3.4: IGV image and Sanger sequencing electropherogram for FRASC67, showing a *NF1*:c.27G>A true positive variant identified and confirmed on Sanger sequencing.

In an attempt to verify if the *NF1*:c.27G>A was a false-positive call, FRASC83 sample, one of the samples with the false positive call was re-sequenced using a different chemistry (on the ThermoFisher, Ion-torrent S5 sequencing platform) utilising another panel with the *NF1* gene used in our facility to assess whether same result would l be observed. Figure 3.5 shows that this variant was not observed using both Sanger sequencing and the IDP panel. This additional QC step was introduced because the VAF skewing was not as informative as for the *PTPN11* variant.

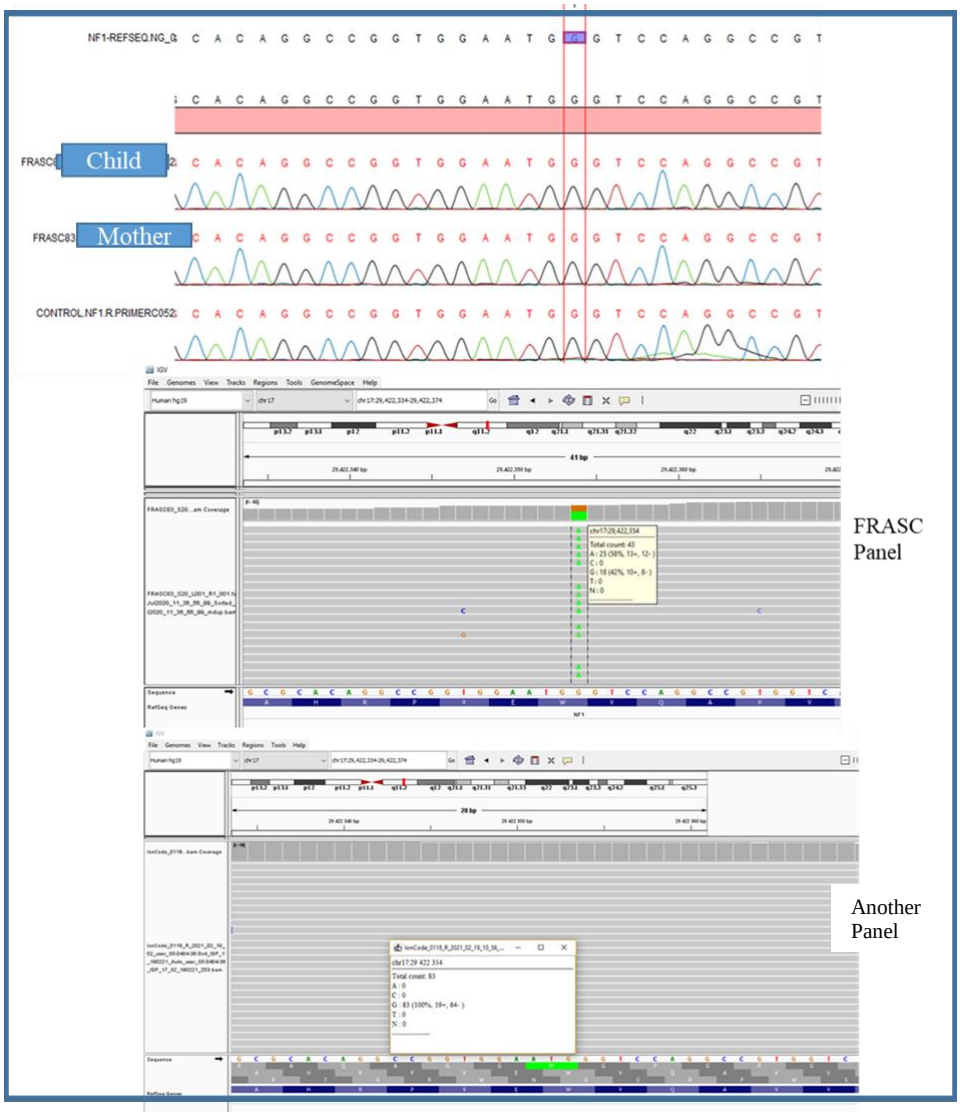


Figure 3.5: Showing FRASC83 sample where the *NF1*:c.27G>A variant was not confirmed on Sanger sequencing and when the another panel was utilised for sequencing.

False-positive calls observed for both the *PTPN11*:c.923A>G and *NF1*:c.27G>A in these patients were resolved by adjusting the variant call parameters utilised in the Agilent SureCall analysis software. The default parameters of this software were initially set to low stringency with a variant allele frequency (VAF) threshold at 10%. This was suspected to be contributor to the false-positives observed, as most analysis software has the default VAF threshold at 30% (Petrackova et al., 2019a; Song et al., 2021). So the VAF threshold on the SureCall software was adjusted to 30% and the data from these samples were reanalysed, Increasing the VAF threshold from 10% to 30% completely removed the *PTPN11*:c.923A>G variant false positive calls and the variant only remained in FRASC66, which was also identified using Sanger sequencing as a true variant.

However for the *NF1*:c.27G>A false positive data the variant remained even after the threshold of the VAF was adjusted to 30%. This indicates that when there is poor coverage of the variant, the analysis software may not be accurate and there is a higher chance of false positive calls.

Based on the findings from this section, all data was then analysed using a VAF threshold of 30% and all variants where confirmed using Sanger sequencing so that all false positive calls were not included in the dataset in the next section.

3.3. Genetic variants identified using the FRASC targeted gene panel

This section details all clinically significant variants (pathogenic, likely pathogenic and VUS) identified using the targeted panel. A total of 34 likely pathogenic/pathogenic variants were identified in eight RASopathy genes (*PTPN11, BRAF, LZTR1, MAP2K2, NRAS, NF1, RAF1* and *SOS1*), resulting in an overall variant detection rate of 56.6% (34/60). Sanger sequencing was performed to confirm all variants detailed in this section. Two patients had a VUS in the *MAP2K1* gene. Twenty-four patients (10 suspected of having NF1, 13 with clinical features suggestive of NS, and 1 with an unknown RASopathy) had no pathogenic variants identified.

3.3.1. Variants identified in the *PTPN11* gene

A total of ten pathogenic variants were identified in the *PTPN11* gene (Table 3.8). All these variants are heterozygous and have been previously reported in disease variant databases. ClinVar database accession numbers of these variants are also listed in the table. FRASC100 was diagnosed with NSML and the rest of the patients with *PTPN11* had a clinical diagnosis of NS.

Table 3.8: Pathogenic variants identified in the *PTPN11* (NM_002834.3) gene

Patients	Clinical disorder	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC1	NS	c.215C>G (p.A72G)	VCV000013325.13	Missense	Pathogenic
FRASC2	NS	c.922A>G (p.N308D)	VCV000013326.31	Missense	Pathogenic
FRASC10	NS	c.211T>C (p.F71L)	VCV000040499.3	Missense	Pathogenic
FRASC18	NS	c922A>G (p.N308D)	VCV000013326.31	Missense	Pathogenic
FRASC20	NS	c.182A>G (p.D61G)	VCV000013330.12	Missense	Pathogenic
FRASC36	NS	c.228G>T (p.E76D)	VCV000040502.6	Missense	Pathogenic
FRASC66	NS	c.923A>G (p.N308S)	VCV000013327.16	Missense	Pathogenic
FRASC81	NS	c.1510A>G (p.M504V)	VCV000040562.20	Missense	Pathogenic
FRASC86	NS	c.188A>G (p.Y63C)	VCV000013333.19	Missense	Pathogenic
FRASC100	NSML	c.1381G>A (p.A461T)	VCV000013342.10	Missense	Pathogenic

PTPN11 gene pathogenic variants were identified in (9/22) 41% of the patients with a clinical diagnosis of Noonan syndrome When segregation analysis was performed using Sanger sequencing, it was seen that the *PTPN11:* c.922A>G (p.N308D) variants for FRASC2 and FRASC18 were also observed in the mothers of both probands (figure 3.6 A and B). Similarly, the *PTPN11:* c.182A>G (p.D61G) variant was also observed in both child and mother for FRASC20 (Figure 3.7). Mothers for FRASC2 and FRASC20 were clinically diagnosed with NS and mother to FRASC18 was not formerly assessed.

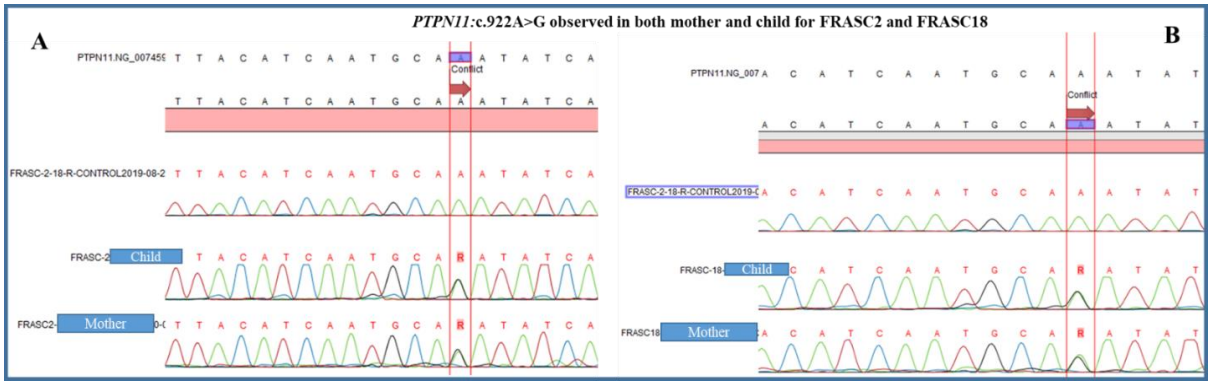


Figure 3.6: Sanger sequencing electropherograms of FRASC2 (A) and FRASC18 (B) showing the *PTPN11*:c.922A>G variant observed in the probands and their mothers

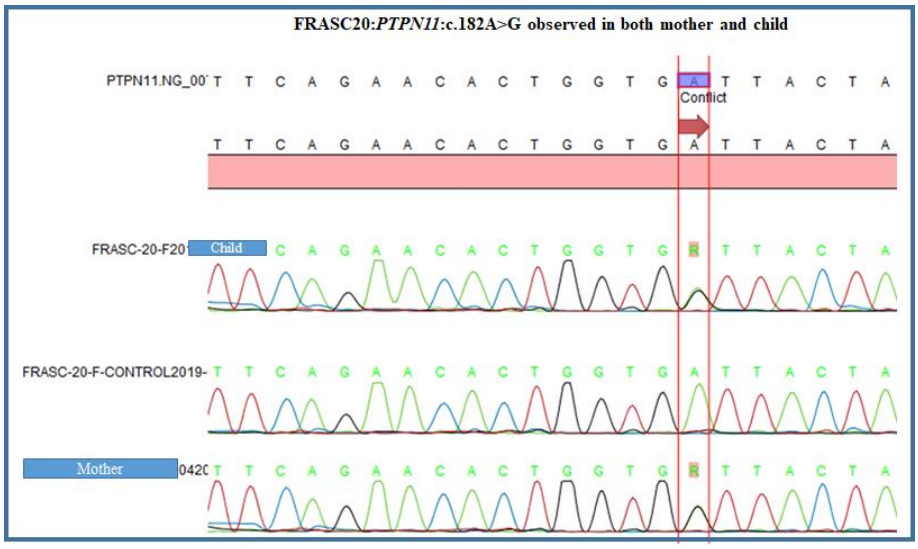


Figure 3.7: Sanger sequencing electropherograms of FRASC20 showing the *PTPN11*:c.182A>G variant observed in both child and mother.

The common clinical features of the patients with a pathogenic variant in the *PTPN11* gene are short stature, cardiac anomalies, webbed neck, low set ears, hypertelorism, epicanthic folds, intellectual disability and undescended testis in males. These are typical features associated with Noonan syndrome and are listed in Table 3.9 (Tartaglia et al., 2002: 11; Athota et al., 2020).

Table 3.9: Typical clinical features of Noonan syndrome observed in patients with PTPN11 variants

All patients listed in Table 3.9 had a clinical diagnosis of NS, except for FRASC100 with clinical diagnosis of NSML. X=present, PS= Pulmonary stenosis, ASD= atrial septal defect, PDA= Patent ductus arteriosus, VSD= ventricular septal defect, DSPF= downslanting palpebral fissures, EF= epicanthic fold, F=female, M=Male

Patient (FRASC No)	**1**	**2**	**10**	**18**	**20**	**36**	**66**	**81**	**86**	**100**
Short stature	X	X	X	X	X	X	X	X		X
Heart defects										
PS		X							X	X
ASD		X			X	X				
PDA										
VSD	X		X		X				X	
Face										
Low set ears	X	X	X		X	X	X	X	X	X
Posteriorly rotated ears		X			X		X	X		
DSPF			X		X	X	X	X	X	X
Ptosis		X			X		X			
EF		X	X	X		X	X	X	X	X
Hypertelorism		X	X	X		X		X		X
Webbed neck	X	X	X	X	X				X	X
ID										
Mild				X				X		X
Moderate		X	X		X		X			
Sex	F	M	F	M	F	F	M	F	M	M
Undescended testes		X		X			X		X	X

NSML patient

FRASC100 had a clinical diagnosis of NSML and had the pathogenic *de novo* variant *PTPN11*: c.1381G>A (p.A461T). Clinical features that distinguished this patient from other patients with *PTPN11* variants is that in addition to typical NS features, this patient had 10 CALs, multiple lentigines, hearing loss, pectus excavatum and myopic vision. Both parents were included for segregation analysis using Sanger sequencing and did not carry this variant (Figure 3.8)

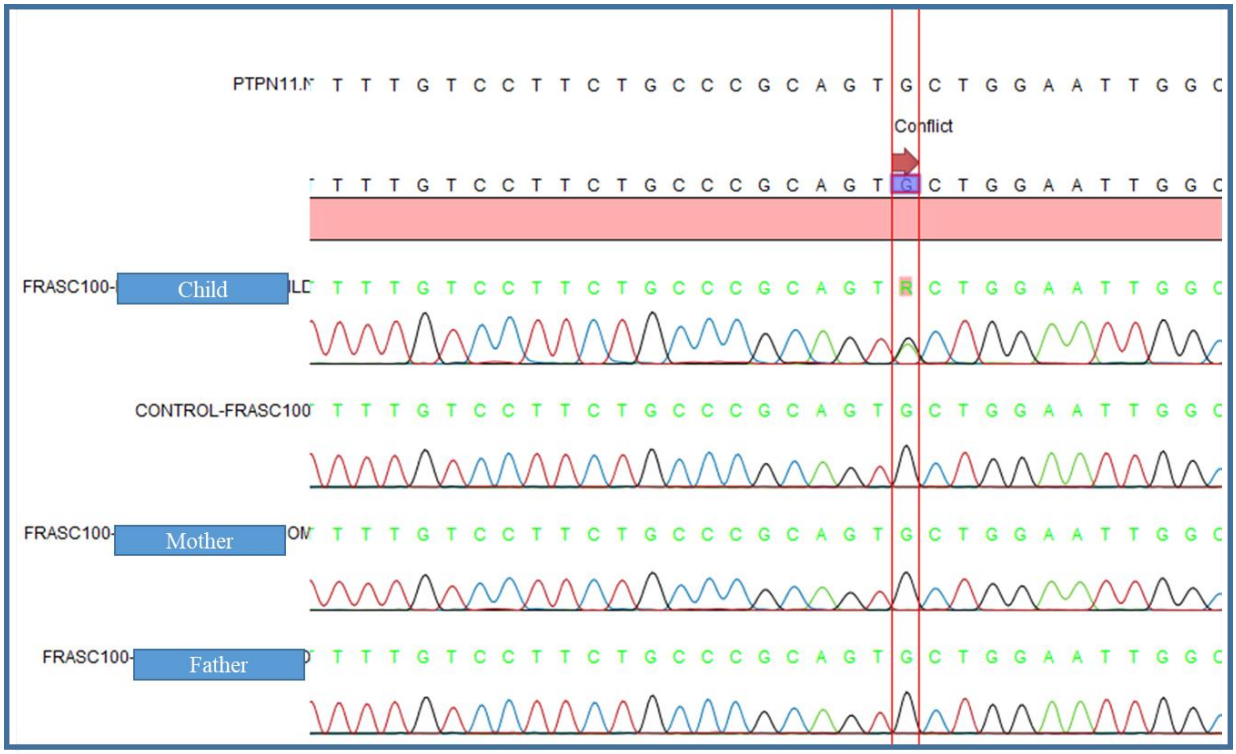


Figure 3.8: Sanger sequencing electropherogram of the NSML patient, showing the *PTPN11*: c.1381G>A (p.A461T) de novo variant and the parents (homozygous wild type)

All the variants identified in the *PTPN11* gene were mapped on the *PTPN11* protein domains to see where they cluster (Figure 3.9). Most pathogenic variants were missense mutations clustered in exons 3, 8, 12, and 13. Five variants (D61G, Y63C, F71L, A72G and E76D) were identified in exon 3 and occur within the N-terminal of the SH2 (Src Homology 2) domain, indicating that this is the hot-spot region as reported in the literature (Tartaglia et al., 2002; Athota et al., 2020). The two variants N308D and N308S were identified in exon 8 and occurs within the protein tyrosine phosphate (PTP) domain. The A461T variant observed in exon 12 of the PTP domain is associated with Leopard/NSML syndrome as identified in FRASC100 (ClinVar ID: VCV000013342.10).

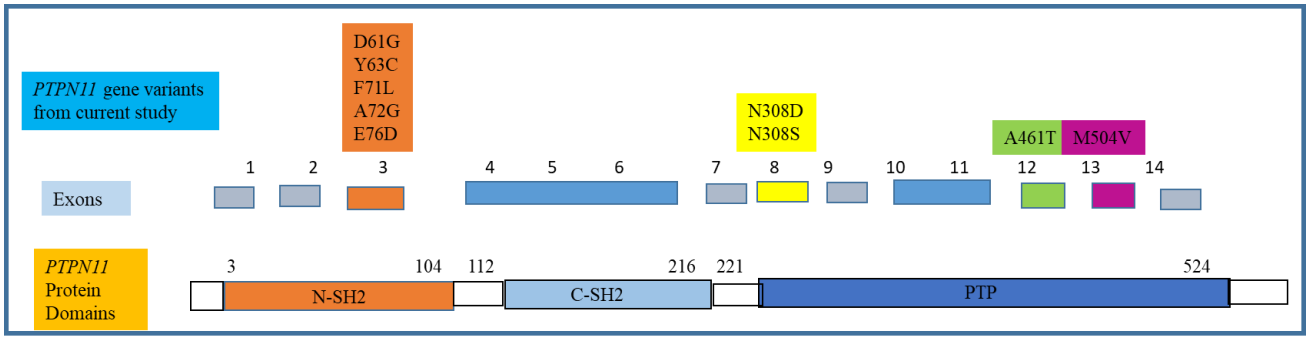


Figure 3.9: Protein Tyrosine Phosphatase protein domains (SHP) containing the N and C terminals (N-SH2 and C-SH2) as well as the protein tyrosine phosphate (PTP) domain showing where the *PTPN11* variants identified in the current study cluster.

3.3.2. Variants identified in the *BRAF* gene

Seven pathogenic variants in the *BRAF* gene were identified in the current study, five patient had a clinical diagnosis of NS while two had a clinical diagnosis of CFC (Table 3.10). All variants were heterozygous and previously reported in ClinVar and the accession numbers are listed in the Table 3.10.

Table 3.10: List of the pathogenic variants identified in the *BRAF* (NM_004333.4) gene

Patients	Clinical disorder	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC4	NS	c.770A>G (p.Q257R)	VCV000013973.17	Missense	Pathogenic
FRASC64	NS	c.770A>G (p.Q257R)	VCV000013973.17	Missense	Pathogenic
FRASC73	NS	c.770A>G (p.Q257R)	VCV000013973.17	Missense	Pathogenic
FRASC88	NS	c.736G>C (p.A246P)	VCV000013965.5	Missense	Pathogenic
FRASC39	NS	c.722C>T (p.T241M)	VCV000029805.6	Missense	Pathogenic
FRASC63	CFC	c.1502A>T (p.E501V)	VCV000040374.1	Missense	Pathogenic
FRASC99	CFC	c.1741A>G (p.N581D)	VCV000013979.4	Missense	Pathogenic

Three patients were found to have the same *BRAF:* c.770A>G (p.Q257R) pathogenic variant, see Figure 3.10 showing 2 of the 3 patients with this pathogenic variant. These two presented with clinical features such as moderate (patient 64) and severe (patient 4) ID, DSPF, low-set posteriorly rotated ears and epicanthic folds (Table 3.8).



A
FRASC4

B
FRASC64

Figure 3.10: FRASC4 (A) and FRASC64 (B) with a clinical diagnosis of NS both carrying the same *BRAF*:c.770A>G (p.Q257R) pathogenic variant

Clinical features of patients with the *BRAF* variants are summarised in Table 3.11, showing FRASC4, FRASC64, FRASC73, FRASC88 and FRASC39 had a clinical diagnosis of NS whereas FRASC63 and FRASC99 had a clinical diagnosis of CFC.

Table 3.11: Summary of clinical features of patients with *BRAF* pathogenic variants

Patient FRASC No.	**4**	**64**	**73**	**88**	**39**	**63**	**99**
Short Stature	X			X	X	X	
Skin							
Deep creases			X		X	X	X
Hyperkeratosis			X			X	X
Heart defects							
PS				X		X	
Mitral or tricuspid valve regurgitation			X				
HOCM					X		
Face							
Coarse features						X	X
Frontal bossing		X			X		
Low set ears	X	X	X	X		X	
Posteriorly rotated ears	X	X	X				
DSPF	X	X	X	X	X	X	X
Ptosis				X			X
EF	X	X	X	X	X	X	
Hypertelorism	X		X	X	X	X	X
Full lips			X		X	X	
Webbed neck			X				X
Broad chest		X	X				X
Pectus excavatum	X			X			
Hair							
Sparse		X	X	X	X	X	X
Nails							
Small						X	
Deep-set						X	X
Dystrophic						X	
ID							
Mild					X		
Moderate	-	X	X			X	
Severe	X						
Myopic vision							X
Strabismus		X				X	
Joint laxity		X	X		X		
***BRAF* variant**	c.770A>G	c.770A>G	770A>G	736G>C	722C>T	c.1502A>T	c.1741A>G

X=present, PS= Pulmonary stenosis, HOCM= Hypertrophic obstructive cardiomyopathy, DSPF= downslanting palpebral fissures, EF= epicanthic fold, F=female, M=Male

Sanger sequencing confirmation of the variants identified in patients FRASC4, FRASC64 and FRASC73 are shown in Figure 3.11 A and B. The mothers of both FRASC4 and FRASC64 did not have the variants and fathers were unavailable to test. The c.770A>G variant was also observed in FRASC 73, also included in Sanger sequencing confirmations in Figure 3.11 C.

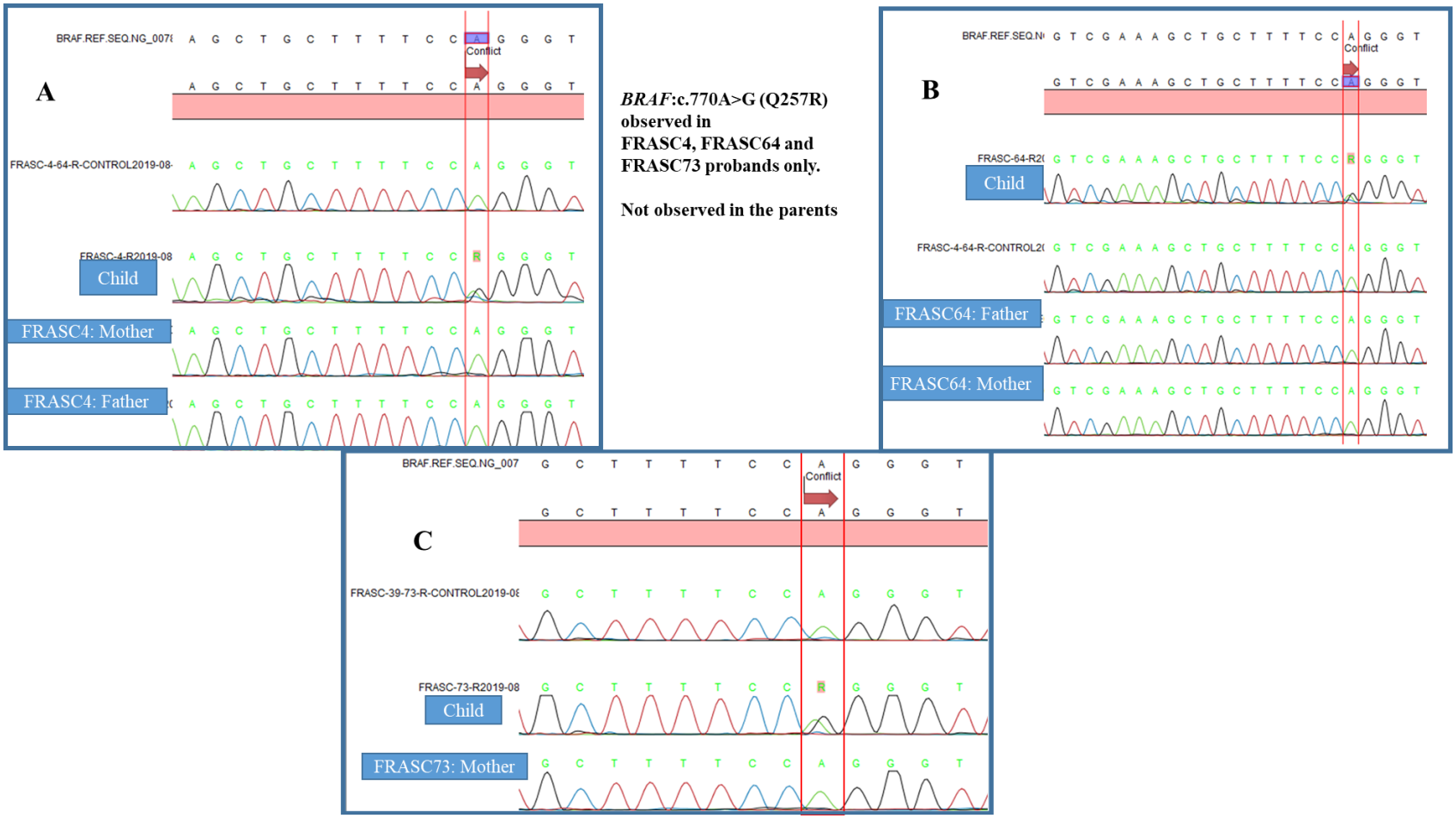


Figure 3.11: Sanger sequencing electropherogram of the three patients A (FRASC4), B (FRASC64) and C (FRASC73), showing the *BRAF*:c.770A>G variant and the parents (having homozygous wild-type variant.

Sanger validations of FRASC39 and FRASC88 are also indicated in Figure 3.12. The *BRAF*:c.722C>T (A) and *BRA*F:c.736G>C (B) variants were also not observed in mothers, fathers were not available to test and ascertain if variants are *de novo* or not.

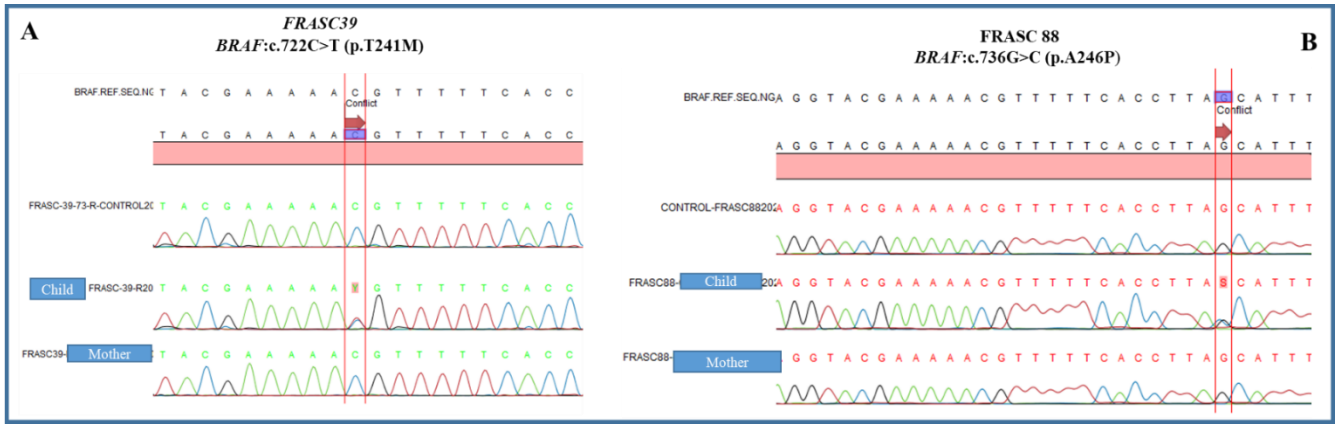


Figure 3.12: Sanger sequencing electropherogram of the two patients, showing the *BRAF*:c.722C>T (A) and *BRAF*:c.736G>C (B) variants and both variants not observed in the mothers of the probands.

All the variants (five) identified in the *BRAF* gene were mapped on the RAS binding and protein kinase protein domains including conserved regions (CR1-3). It can be seen in Figure 3.14 that three (Q257R, T241M and A246P) of the five are clustered in CR1 indicating a possible hot spot region. Two variants (E501V and N581D) were in exon 12 and 14, respectively of the conserved region 3 (CR3) of the protein kinase domain (Figure 3.13).

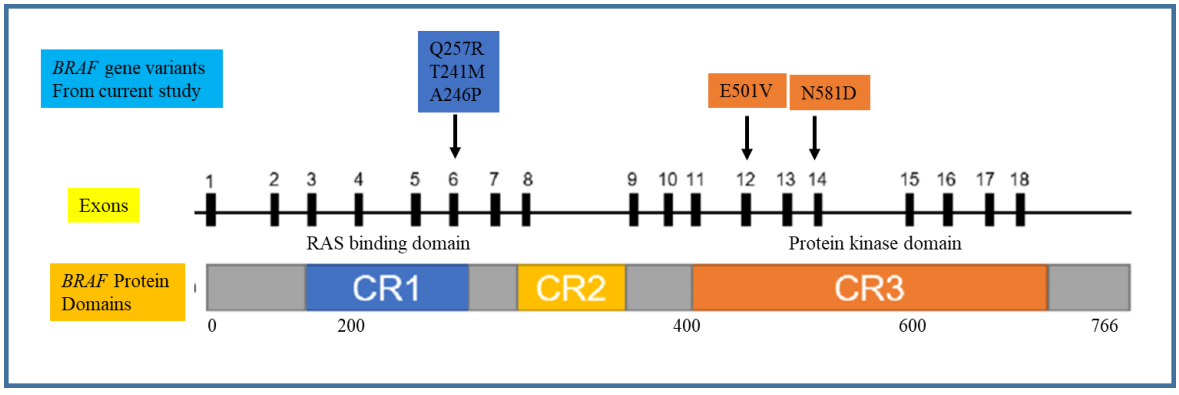


Figure 3.13: *BRAF* variants obtained in the current study are mostly clustered in exon 6 of the conserved region 1 (CR1) of the Ras-GTP-binding self-regulatory domain.

3.3.3. Variants identified in the *LZTR1* gene

Heterozygous *LZTR1* variants were observed in 3 patients with a clinical diagnosis of Noonan syndrome (FRASC15, FRASC21 and FRASC69) and 1 patient (FRASC62) presenting with clinical features of Schwannomatosis. The variants identified are listed in Table 3.12. The two variants identified in FRASC21 and FRASC62 have not been reported in clinical variant interpretation databases such as ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) and Human Genome Mutation Database (HGMD®) (http://www.hgmd.cf.ac.uk).

Table 3.12: List of likely pathogenic and pathogenic variants identified in the *LZTR1* (NM_006767.4) gene

Patients	Clinical disorder	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC15	NS	c.730T>C (p.S244P)	VCV000636792.1	Missense	Likely pathogenic
FRASC21	NS	c.690C>G (p.N230K)	Not reported	Missense	Likely pathogenic
FRASC69	NS	c.848G>A (p.R283Q)	VCV000561716.3	Missense	Likely pathogenic
FRASC62	Schwannomatosis	c.2269C>T (p.Q757X)	Not reported	Nonsense	Pathogenic

FRASC69 Figure 3.14 (A) was found to have the *LZTR1*:c.848G>A and FRASC15 in Figure 3.14 (B) was found to have a c.730T>C pathogenic variant in the *LZTR1* gene. Both FRASC69 and 15 with *LZTR1* pathogenic variants had typical Noonan syndrome clinical features such as mild ID, DSPF, epicanthic folds. FRASC 15 also presented with hypertrophic obstructive cardiomyopathy.



Figure 3.14: FRASC69 (A) with a clinical diagnosis of NS carrying the *LZTR1*:c.848G>A (p.R283Q) likely pathogenic variant and FRASC15 (B) with a clinical diagnosis of NS carrying the previously described *LZTR1*: c.730T>C (p.S544P) likely pathogenic variant

The *LZTR1*:c.848G>A variant observed in FRASC69, Figure 3.15 (A) has conflicting classifications on ClinVar of LP/VUS. Using ACMG –AMP guidelines (Richards et al., 2015; Gelb et al., 2018) PM2 (variant not found in gnomAD genomes/exomes), PP3 (multiple in-silico predictors, predicts pathogenic/deleterious effect) and PP4 (applied because the patient had a clinical diagnosis of Noonan syndrome) this variant was classified as likely pathogenic in the current study. This variant was also observed in the mother of the patient; however, no formal clinical assessment was conducted on the mother (Figure 3.15.A).

A *LZTR1*:c.690C>G novel variant was observed in FRASC21 (Figure 3.16 B).

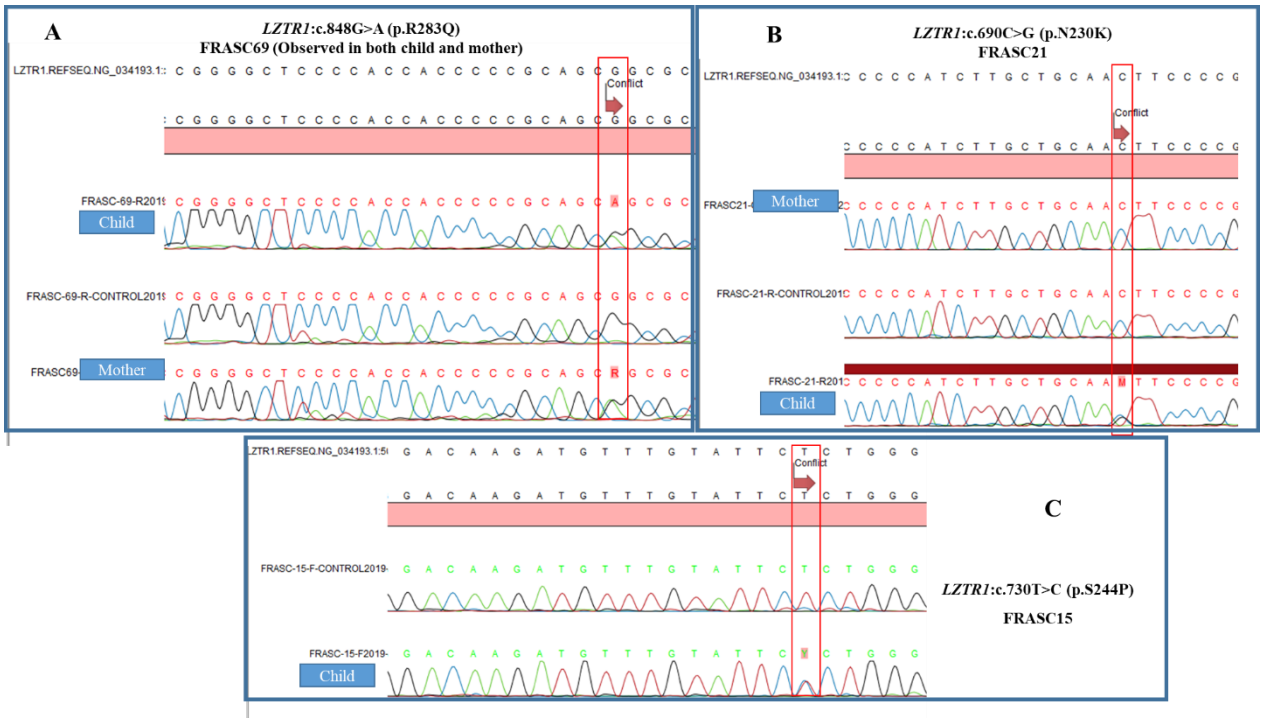


Figure 3.15: Sanger sequencing electropherogram of the three patients, showing the *LZTR1*: c.848G>A, c.690C>G and c.730T>C variants in A (FRASC69), B (FRASC21) and C (FRASC15). Variant *LZTR1*: c.848G>A was observed in both the FRASC69 patient and mother (A).

Clinical features of the patients with clinical features of Noonan syndrome that had a likely pathogenic variant in the *LZTR1* gene are summarised in Table 3.13.

Table 3.13: Summary of clinical features of Noonan syndrome patients with *LZTR1* variants observed in this study

Patient FRASC No.	**15**	**21**	**69**
Short Stature	X		
Skin			
Deep creases	X		X
Heart defects			
PDA		X	
HOCM	X		
Face			
Coarse features		X	
Low set ears		X	X
DSPF	X	X	X
Ptosis		X	X
EF	X		X
Hypertelorism		X	
Full lips		X	
Webbed neck		X	X
Broad/Shield chest		X	
Pectus excavatum		X	X
Hair			
Coarse		X	
Sparse			X
ID			
Mild	X	X	X
Short 5th Metacarpal			X
Variant identified	*LZTR1*:c.730T>C	*LZTR1*:c.690C>G	*LZTR1*:c.848G>A

FRASC62 had features suggestive of Schwannomatosis (Figure 3.16, A) and was shown to have a c.2269C>T (p.Q757X) variant in the *LZTR1* gene. As observed in the Figure 3.16 (B) the variant was identified in the two clinically affected daughters as well during segregation analysis using Sanger sequencing.

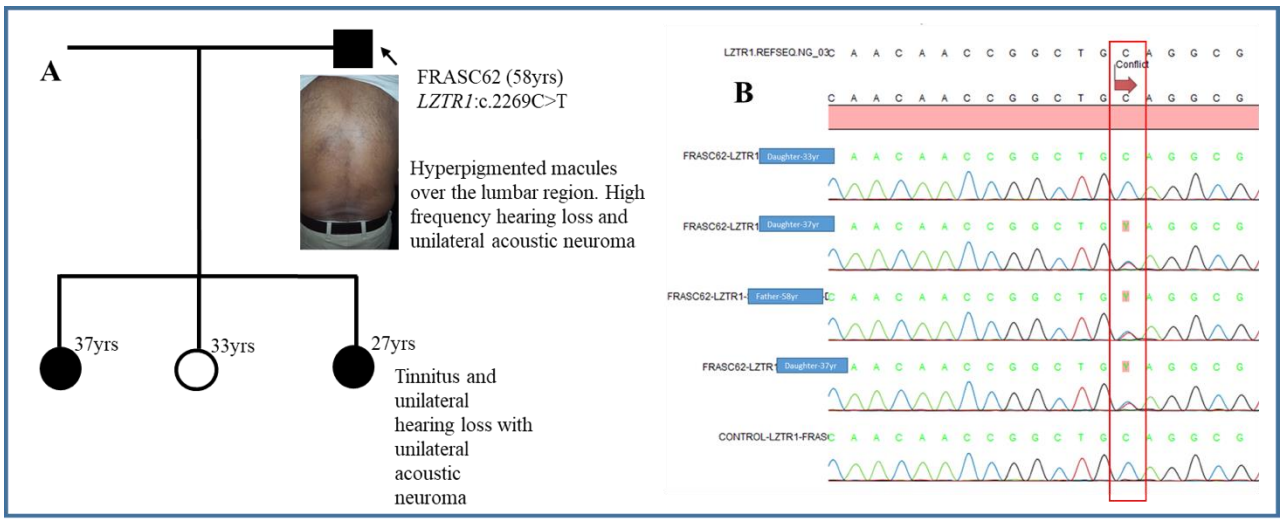


Figure 3.16: Pedigree and Sanger sequencing electropherogram of FRASC62, showing the disease-causing variant segregating through the family with Schwannomatosis

The *LZTR1* variants were mapped on the Kelch protein domains of the *LZTR1*, exon 8 and 9 of Kelch 3 and 4 had all three (N230K, S244P and R283Q) variants associated with NS and the the Q757X Schwannomatosis variant was found in in exon 19 was and was not within any of the functional domains (Figure 3.17).

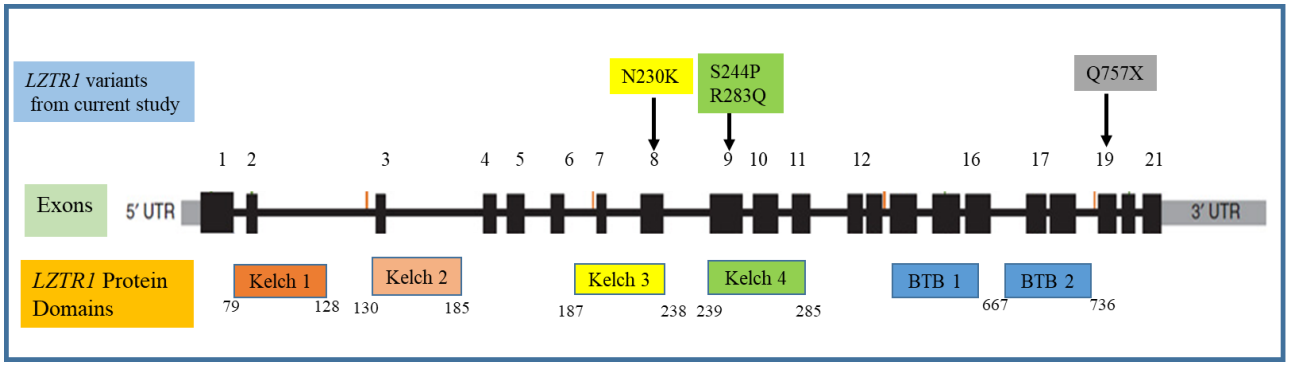


Figure 3.17: *LZTR1* variants obtained in the current study are clustered exons 8 and 9 within Kelch 3 and 4 protein domains, the variants are monoallelic and associated with autosomal dominant Noonan syndrome (Motta et al., 2019: 1; Pagnamenta et al., 2019).

3.3.4. Variants identified in *SOS1* gene

Heterozygous Noonan-causing variants in the *SOS1* gene were observed in 2 patients: FRASC14 and FRASC60. These two pathogenic variants are listed in Table 3.14, both variants are known variants reported in ClinVar to be associated with NS.

Table 3.14: Two variants identified in the *SOS1* (NM_005633.3) gene.

Patients	Clinical disorder	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC14	NS	c.806T>C (p.M269T)	VCV000040662.9	Missense	Pathogenic
FRASC60	NS	c.1649T>C (p.L550P)	VCV000040680.4	Missense	Pathogenic

FRASC14 in whom the *SOS1:* c.806T>C variant was identified is shown in Figure 3.18. This patient presented with typical Noonan syndrome features such as pulmonary stenosis, unusual faces, mild ID and strabismus



FRASC14

Figure 3.18: FRASC14 with a clinical diagnosis of NS carrying *SOS1*:c.806T>C (p.M269T) pathogenic variant.

Sanger sequencing was performed to confirm these variants and the electropherograms are indicated in Figure 3.19 (A). For FRASC60, this variant was confirmed in the child only (Figure 3.20 B).

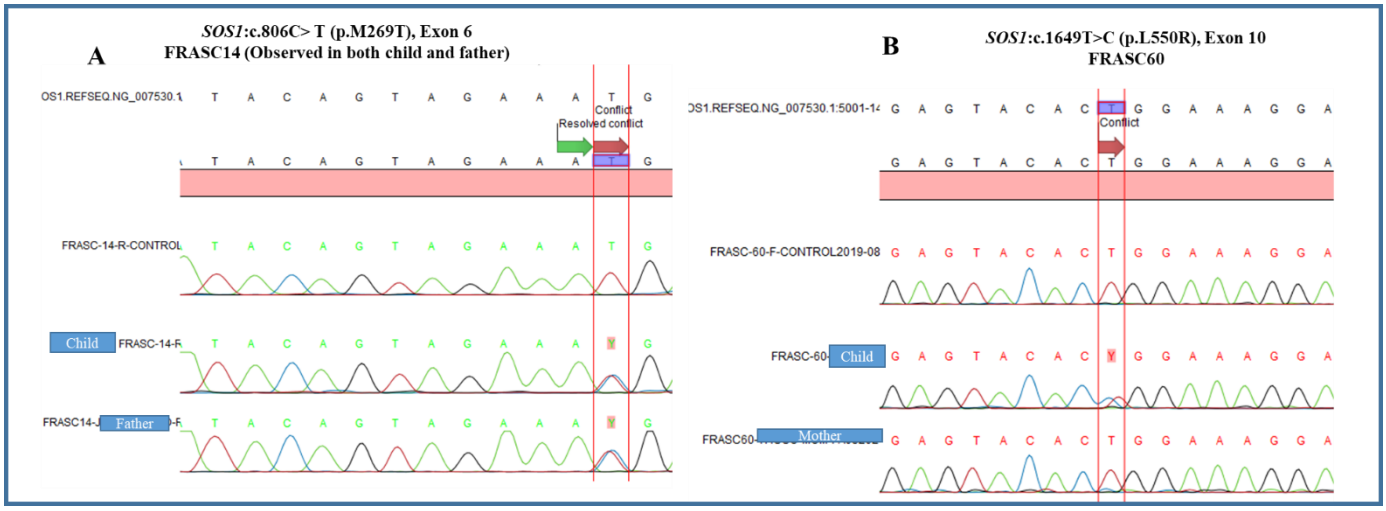


Figure 3.19: Sanger sequencing electropherogram of the two patients, showing the c.806C>T, identified in exon 6 (A) and c.1649T>C identified in exon 10 (B) of the *SOS1* gene. Variant *SOS1*: c.806C>T was observed in both FRASC14 patient and her father (A)

3.3.5. Variants identified in the *RAF1* gene

Two pathogenic variants were identified in the *RAF1* gene, both these variants have been previously reported in ClinVar (Table 3.15) and HGMD to be associated with Noonan syndrome.

Table 3.15: Two variants identified in the *RAF1* (NM_002880.3) gene

Patients	Clinical disorder	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC17	NS	c.781C>A (p.P261T)	VCV000040604.5	Missense	Pathogenic
FRASC87	NS	c.770C>T (p.S257L)	VCV000013957.18	Missense	Pathogenic

Sanger sequencing was performed to confirm the variants. These variants were observed in the probands only and were not observed in the mothers of the patients (Figure3.20).

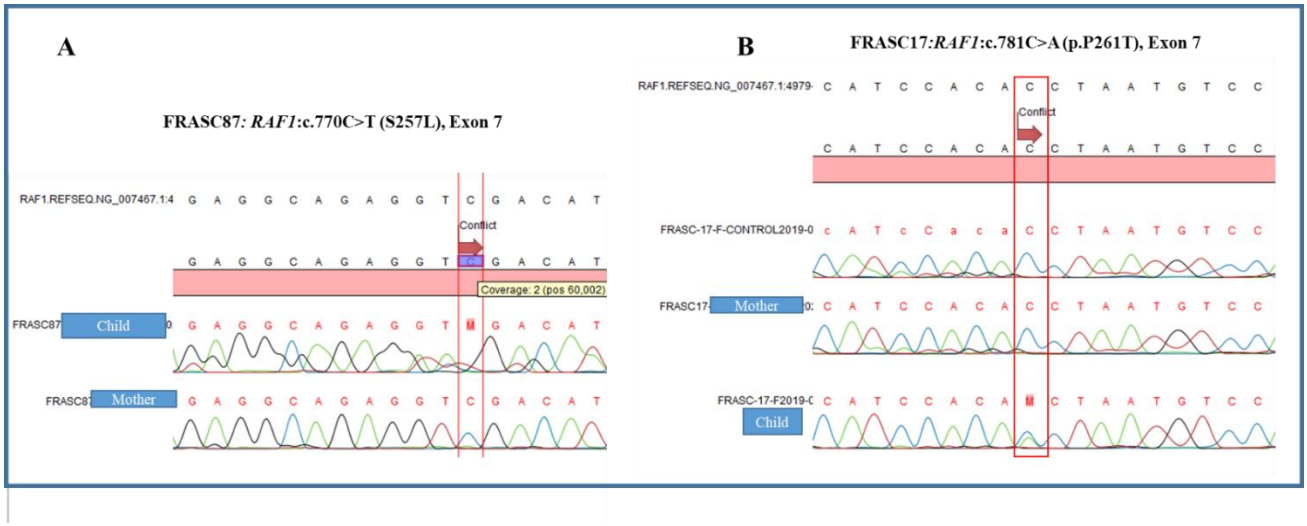


Figure 3.20: Sanger sequencing electropherogram of FRASC87 and FRASC17, showing the c.770C>T,(A) and c.781C>A (B) both identified in exon 7 of the *RAF1* gene.

3.3.6. Genes with single variants identified (*NRAS* and *MAP2K2*)

NRAS and *MAP2K2* genes each had one variant identified. The *NRAS* variant was identified in a patient with a clinical diagnosis of NS and the *MAP2K2* variant was identified in a patient presenting with clinical features of CFC. Both these variants have been previously reported in ClinVar to be associated with these syndromes (Table 3.16).

Table 3.16: Variants identified in the *NRAS* and *MAP2K2* genes

Patients	Clinical disorder	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC5	NS	*NRAS:*NM_002524.3:c.173C>T (p.T58I)	VCV001070042.1	Missense	Likely pathogenic
FRASC22	CFC	*MAP2K2:*NM_030662.3:c.169T>G (p.F57V)	VCV000008273.2	Missense	Pathogenic

The two variants were confirmed using Sanger sequencing and that is shown in Figure 3.21 A for the *MAP2K2*: c.169T>G and figure 3.21 B for the *NRAS*:c.173C>T variant

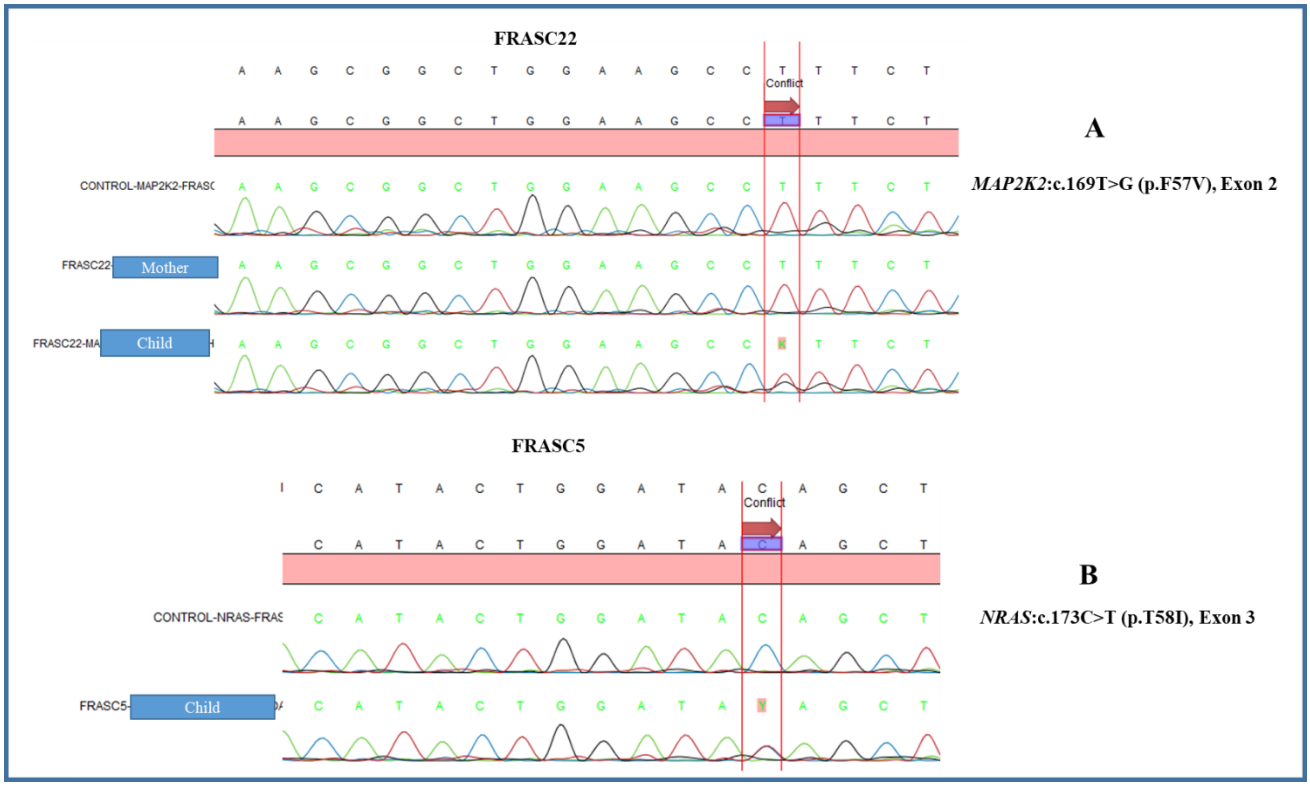


Figure 3.21: Sanger sequencing electropherogram of FRASC22 and FRASC5, showing the *MAP2K2*: c.169T>G (A) and the *NRAS*:c.173C>T (B) variants.

3.3.7. Variants identified in the *NF1* gene

Fifteen patients (25%) were recruited with clinical features suggestive of NF1 syndrome> Oof those NF1-causing mutations were identified in seven (46.7%). All variants were previously reported in ClinVar except for the *NF1:*NM_001042492.2:c.27G>A (p.W9X) variant, which was not reported in the ClinVar or HGMD databases. Mutation profiles of the seven patients are indicated in Table 3.17.

Table 3.17: Pathogenic variants identified in the *NF1* (NM_001042492.2*)* gene in patients with clinical diagnosis of NF1

Patients	Clinically significant variants identified	ClinVar ID	Type of mutation	Classification
FRASC38	c.1658A>G (p.H553R)	VCV000420076.5	Missense	Pathogenic
FRASC42	*NF1:* c.6772C>T (p.R2258X)	VCV000230389.15	Nonsense	Pathogenic and
FRASC44	c.496-497 del (p.V166Lfs*)	VCV000431562.4	Frameshift	Pathogenic
FRASC45	c.3721C>T (p.R1241X)	VCV000000361.11	Nonsense	Pathogenic
FRASC46	c.569T>G (p.L190X)	VCV000431974.2	Nonsense	Pathogenic
FRASC53	c.5204_5205 del (p.K1735Sfs*)	VCV001072614.1	Frameshift	Pathogenic
FRASC67	c.27G>A (p.W9X)	N/A	Nonsense	Pathogenic

Sanger sequencing was performed to confirm the variants identified in patients listed in Table 3.15 and this is shown in Figures 3.23 to 3.24.

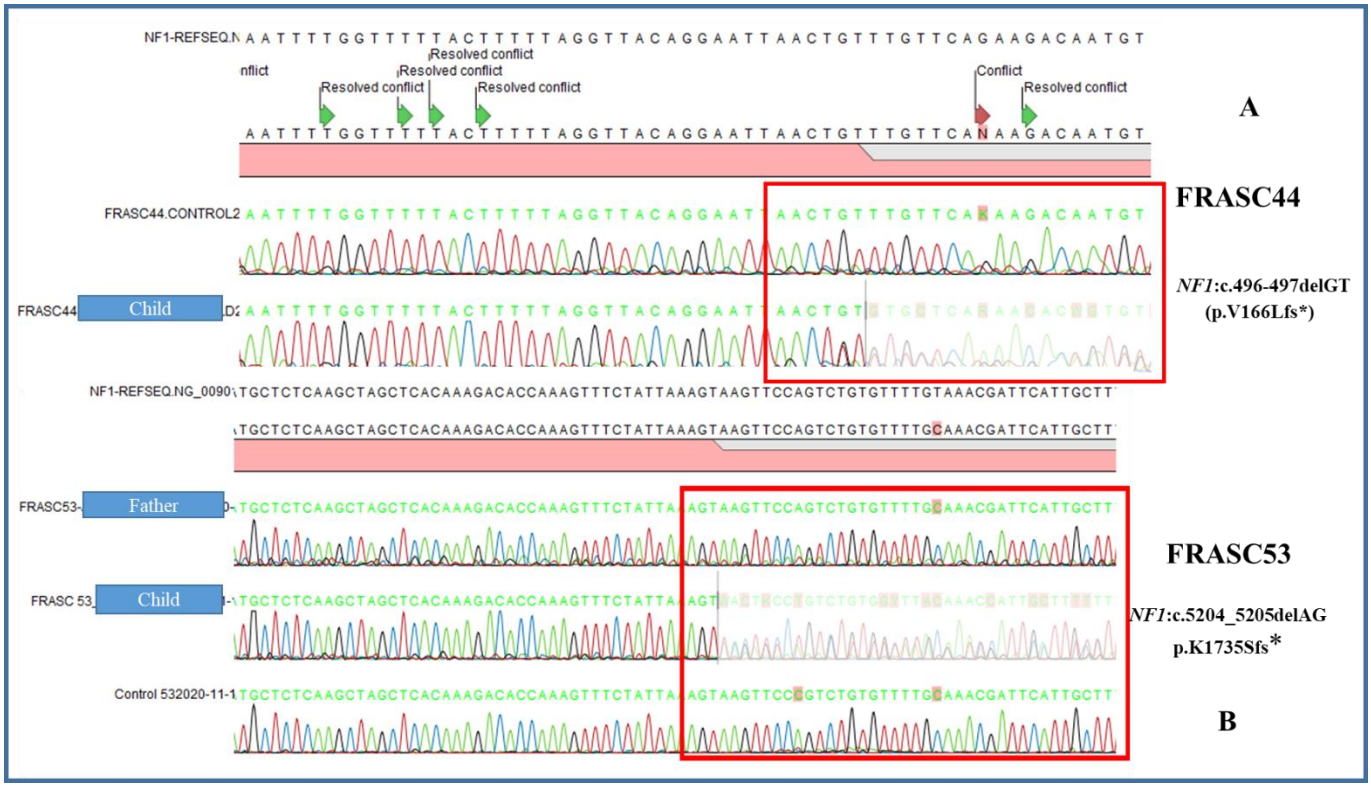


Figure 3.22 Sanger sequencing electropherogram confirming two frame-shifts variants (c.496_497 delGT and c.5204_5205delAG) identified in the *NF1* gene of FRASC44 (A) and FRASC53 (B) where the variant was not observed in the father.

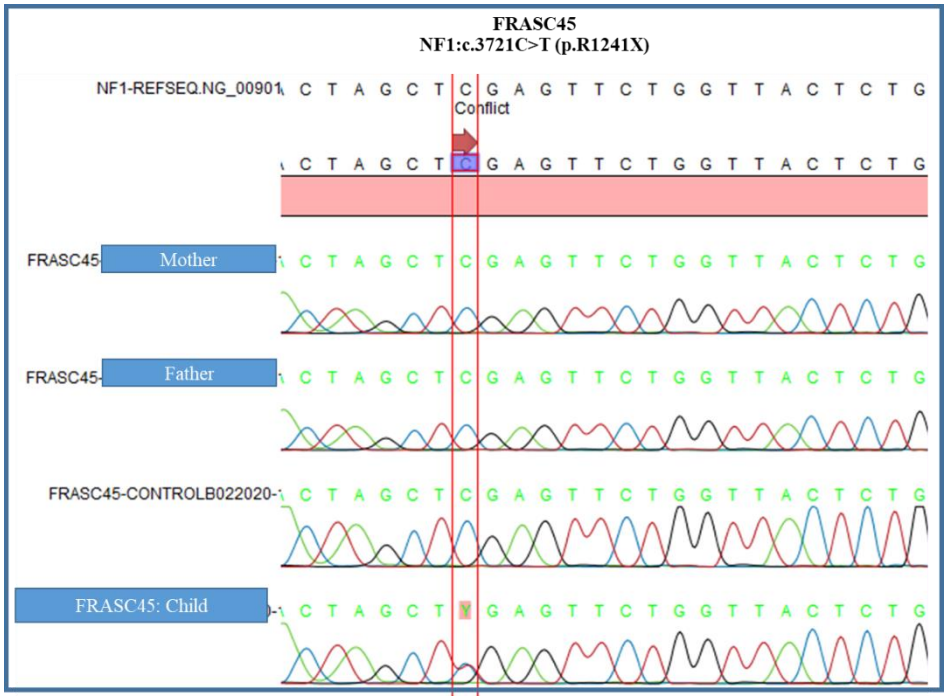


Figure 3.23: Sanger sequencing electropherogram confirming a nonsense variant, c.3721C>T identified in the *NF1* gene of FRASC45 child. Both parents had the homozygous wild type variant

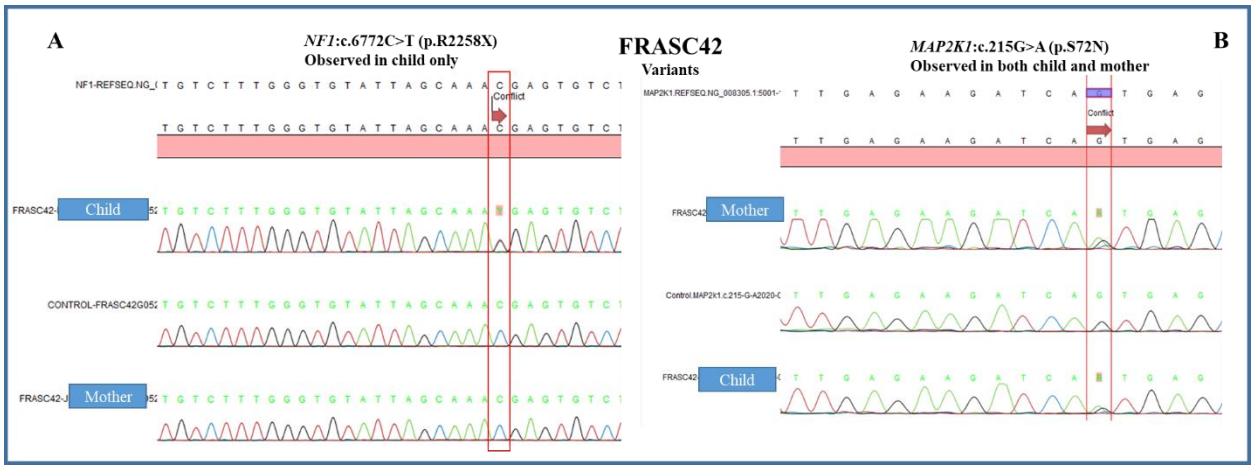


Figure 3.24: Sanger sequencing electropherogram confirming two variants identified in this patient. One nonsense pathogenic *NF1*:c.6772C>T (A) and one missense VUS c.215G>A identified in the *MAP2K1* gene. The *MAP2K1* VUS was also observed in the mother (B)

All these patients had a clinical diagnosis of NF1 presenting with more than six CALs of 1.5cm in size (Table 3.16). More clinical features in keeping with NF1 are summarised in Table 3.15.

With the pathogenic variant identified in all these patients (Table 3.18), the diagnosis of NF1 can be confirmed.

Table 3.18: Summary of clinical features of patients with a pathogenic *NF1* variant.

Patient FRASC No.	**38**	**42**	**44**	**45**	**46**	**53**	**67**
Age	3yr3m	12yr3m	6yr2m	8y9m	29yr9m	11yr8m	4yr9m
Café au lait (number : size)	10: No size	>6:>1.5cm	>10:>1.5cm	>6:>1.5cm	>6:1.5cm	>6:>0.5cm	>6: 5mmx5mm
Freckling							
Axillary	X	X	X	X	X	X	
Inguinal	X	X		X	X	X	
Hyperpigmentation	X						X
Skin							
Deep creases				X			X
Hyperkeratosis	X						
Neurofibromas		X		X	X	X	X
Plexiform neurofibromas		X		X	X	X	
Face							
Coarse features					X		
Frontal bossing	X						
Low set ears	X	X					
DSPF	X	X	X	X			
EF			X				
Hypertelorism			X				
Full lips	X						
Pectus Carinatum	X						
Short 5th metacarpal				X			
ID							
Normal			X	X	X	X	X
Mild	X	X					

All seven *NF1* variants were mapped on the neurofibromin domains to see how they cluster. Most *NF1* variants are nonsense and frameshifts and they do not fall within any functional

domain but the ability of these variants to cause premature termination of the protein formation and results in the shortened coding sequence in the gene have rendered them pathogenic. In Figure 3.25 it can be seen that only three variants (p.H553RM, p.K1241Sfs and p.K1735Sfs are within the neurofibromin domains.

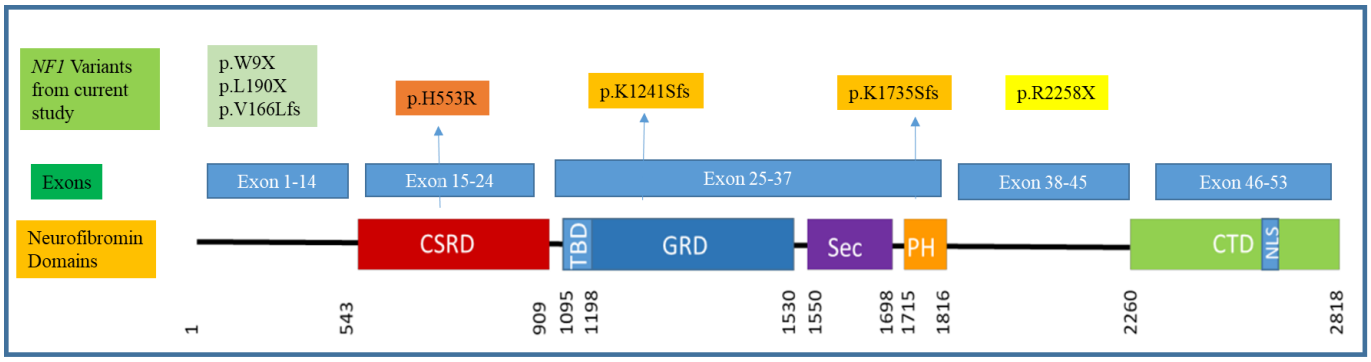


Figure 3.25: *NF1* variants obtained in the current study are mostly clustered outside the functional domains. The type of variants obtained are mostly nonsense and frameshifts causing premature termination**.**

3.3.8. Variant of uncertain significance, *MAP2K1:*c.215G>A (p.S72N)

One VUS in the *MAP2K1* gene was observed in 5% (3/60) of the patients included in this study. The *MAP2K1*:c.215G>A (p.S72N) VUS was observed in three patients in our study (FRASC 19, 42 and 98). Two of the patients had clinical features suggestive of Noonan syndrome and one patient (FRASC42) had another *NF1* pathogenic variant identified and for this patient ACMG rule BP5 (variant found in a case with an alternate molecular basis for disease) applies. The *MAP2K1:*c.215G>A (p.S72N) variant was not found in the ClinVar and HGMD® databases and was not *de novo* as two of the parents (mother to patient 42 and father to patient 98) were found to have this variant during Sanger sequencing analysis (Figure 3.26). The parents found with these variants were not clinically assessed. The significance of this variant is uncertain due to conflicting ACMG-AMP rules, the PM2 (variant have not been reported in population databases), PP2 (missense variants in the *MAP2K1* gene are not tolerated) and BP5 code explained earlier for FRASC42. The BP5 rule was applied with caution as cases with NF1-NS syndromes have been reported where patients presents with features suggestive of NF1 and NS syndromes and have been documented with variants in both genes (Reig et al., 2011; Yapijakis, Pachis and Voumvourakis, 2017).

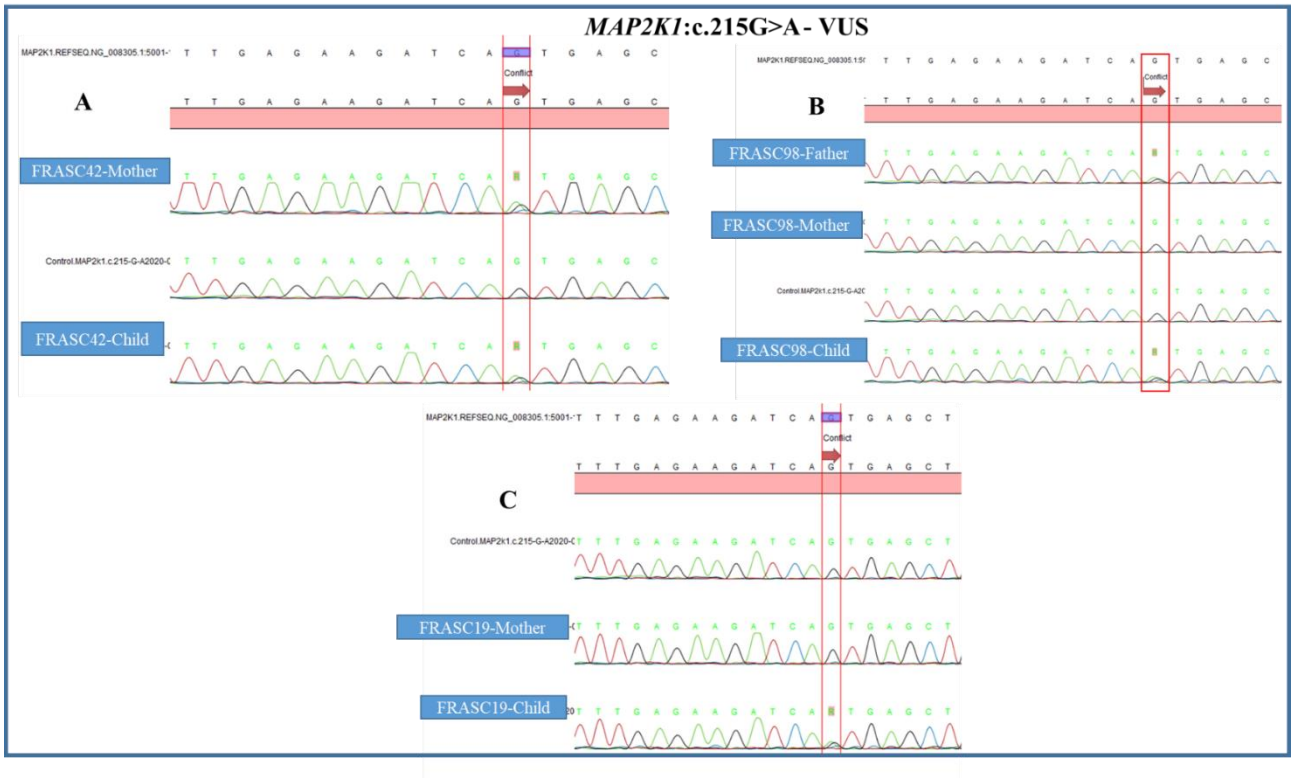


Figure 3.26: Sanger sequencing confirmation was performed for the *MAP2K1*:c.215G>A variant and was found in all three patients (Figure 3.26). Variants identified in FRASC98 father and FRASC42 mother.

3.3.9. Detection rate per clinical diagnosis and gene variant frequencies

In this section the detection rate per clinical diagnosis is summarised (Table 3.19) with the frequency of the variants identified per gene compared to the published data (Table 3.20).

Table 3.19: Summary of the detection rate in the RASopathy genes where clinically significant variants were identified in our cohort of 60 patients.

Clinical diagnosis	Patients recruited	P and LP variants	Number of patients with variants in a gene	Pick-up per diagnosis
Noonan	37	22	*9-PTPN11, 5-BRAF, 3-LZTR1, 2-RAF1, 2-SOS1,1-NRAS,*	59.5%
CFC	3	3	*1-MAP2K2, 2-BRAF*	100%
NSML	1	1	*1-PTPN11*	100%
NF1	15	7	*7-NF1*	46.7
Schwannomatosis	1	1	*1-LZTR1*	100%
Noonan/NF1	1	0		0%
?RASopathy	2	0		0%

P=pathogenic, LP=likely pathogenic

Table 3.20: Comparison of our frequency of clinically significant variants in a gene per RASopathy to the frequencies in existing literature

RASopathy	Gene	Our frequency	Literature	Ref
NS	*PTPN11*	41%	40-50%	(Tartaglia et al., 2002; Xu et al., 2017; Athota et al., 2020)
	BRAF	22.7%	1.7-2.3%	(Sarkozy et al., 2009; Lee et al., 2011)
	LZTR1	13.6%	8%	(Yamamoto et al., 2015)
	RAF1	9%	3-17%	(Pandit et al., 2007; El Bouchikhi et al., 2016)
	SOS1	9%	10-20%	(Roberts et al., 2007; Rauen, 2013; El Bouchikhi et al., 2016)
	NRAS	4.5%	(rare)~1%	(Cirstea et al., 2010; El Bouchikhi et al., 2016)
NSML	*PTPN11*	100%	90%	(Sarkozy, Digilio and Dallapiccola, 2008; Sarkozy et al., 2009; Lee et al., 2011)
CFC	*BRAF*	66.7%	75%	(Sarkozy et al., 2009; Pierpont et al., 2014)
	MAP2K2	33%	25%	(Rodriguez-Viciana et al., 2006; Nava et al., 2007a)
NF1	*NF1*	100%	90%	(Rauen, 2013; Koczkowska et al., 2018)
Schwannomatosis	*LZTR1*	100%	38%%	(Yamamoto et al., 2015)

Table 3.19 shows the detection rate in patients recruited per clinical disorder, for example, of the 15 patients recruited with clinical diagnosis of NF1, 7 (46.7%) had a pathogenic variant

identified. Table 3.20 shows the frequency of variants identified in a gene compared to the literature, for example, all variants in patients with clinical diagnosis of NF1 were identified in the *NF1* gene.

3.3.10. Segregation analysis in family members of variants identified on the targeted

This section describes the variants observed in family members during Sanger sequencing analysis following the identification of a variant in a proband. Clinical assessment of relatives was done in three patients FRASC2, FRASC20 and FRASC62. In other families where variants were observed in parents, no formal clinical diagnosis of the parents was performed Table 3.21.

Table 3.21: Variants that were observed in probands and their relatives

Patient number	Recruited	Clinical disorder	Variants identified and Classification	Sanger Confirmation	Relatives' clinical assessment
FRASC69	Duo	NS	*LZTR1*:c.848G>A Likely Pathogenic	Observed in mother and child	Mother not clinically assessed
FRASC98	Trio	NS	*MAP2K1*:c.215G>A VUS	Observed in father and child	Father not clinically assessed
FRASC42	Duo	NF1	*MAP2K1*:c.215G>A VUS	Observed in mother and child	Mother not clinically assessed
FRASC14	Duo	NS	*SOS1*:c.806T>C Pathogenic	Observed in father and child	Father not clinically assessed
FRASC62	Father and 3 daughters (see Figure 3.17)	NF1	*LZTR1*:c.2269C>T pathogenic	Observed in father and two daughters	Sister to the father previously tested +ve for the mutation in overseas lab
FRASC20	Duo	NS	*PTPN11*:c.182A>G pathogenic	Observed in mother and child	Mother clinically diagnosed with NS
FRASC2	Duo	NS	*PTPN11*:c.922A>G pathogenic	Observed in mother and child	Mother clinically diagnosed with NS
FRASC18	Duo	NS	*PTPN11*:c922A>G pathogenic	Observed in mother and child	Mother not clinically assessed
FRASC67	Duo	NF1	*NF1*:c.27G>A pathogenic	Observed in mother and child	Mother not clinically assessed

3.4. Additional testing on patients who tested negative on the FRASC targeted panel

There were 26 patients with no clinically significant variants identified using the FRASC targeted gene panel (Figure 3.27). The two patients FRASC19 and FRASC98 had the VUS and no further testing was done for these two. In panel-negative NS patients, WES was used as a secondary method for mutation screening, however for *NF1* patients MLPA (deletions/duplication screening) is the recommended second tier analysis due to the rate at which large *NF1* pathogenic deletions are found (Friedman, 1993). A subset of eight (mutation negative) samples with a clinical diagnosis of Noonan syndrome were included in the WES performed in the current study, the reason why they were chosen was because they had some features on Noonan syndrome as assessed by medical geneticist but no variant was identified in the common genes associated with Noonan syndrome. The number of the samples included (eight) was done based on the availability of WES reagents at the time.

The additional WES study and the NF MLPA was performed as part of a substudy MSc research project supervised by the candidate doing the current study as shown in Figure 3.27.

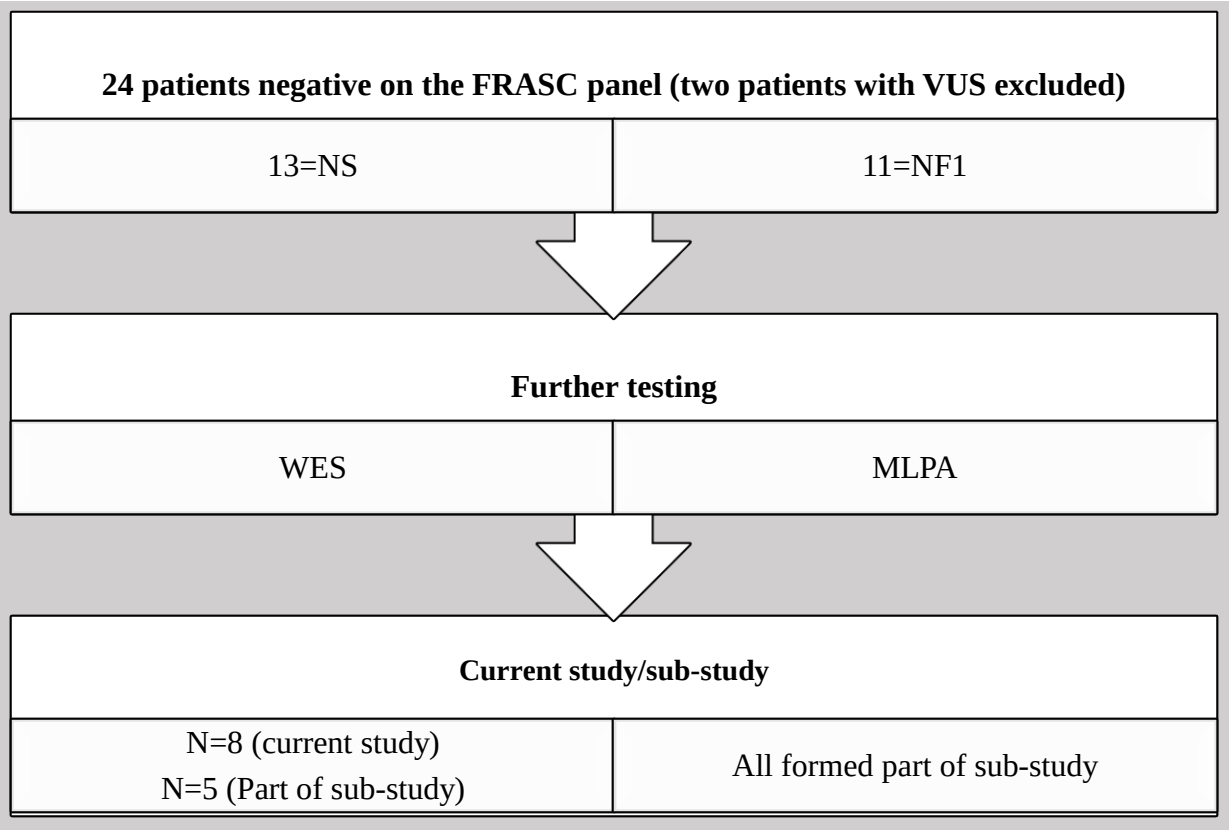


Figure 3.27: Further testing done in 24 Noonan syndrome and Neurofibromatosis type 1 patients in whom no variants were identified using the FRASC targeted gene panel

No clinically significant variants were identified in the five patients for WES and 11 NF1 patients for MLPA in the sub-study cohort. The results of the eight patients included in the WES for the current study are detailed in the following section.

3.4.1. Whole Exome Sequencing

In this section, WES findings of seven out of eight patients are outlined. One of the samples, FRASC71 failed QC and was therefore removed from further analysis (Highlighted in red, Table 3.22). The rest of the seven patients' WES shortlisted variants will be discussed in this section.

Table 3.22: Summary of the run statistics obtained during WES, with quality of the reads, depth coverage and number of variants obtained per sample.

	Run 1				**Run 2**			
Data	**Ion 540™ Chip 1**		**Ion 540™ Chip 2**		**Ion 540™ Chip 3**		**Ion 540™ Chip 4**	
Usable reads	52%		49%		47%		66%	
Total bases	7.8Gbp		7.64Gbp		9.35Gbp		16.9Gbp	
Key signal	74		65		71		88	
ISP loading	68%		58%		70%		88%	
Patient: FRASC No.	**24**	**55**	**56**	**58**	**71**	**80**	**96**	**97**
Mapped reads	~26mil	~20mil	~18mil	~21mil	~2mil	~45mil	~36mil	~49mil
≥Q20 base calls	~4bil	~3bil	~3bil	~4bil	~383mil	~7bil	~6bil	~8bil
Mean depth	78	55	53	64	7	140	114	153
Uniformity	92.11%	92.05%	93.34%	93.21%	86.63%	94.04%	93.71%	91.73%
Variants obtained	59617	44627	44543	49784	549	60606	60692	60579

It can be seen from the Table 3.22 that WES generated mapped reads between 18 and 49 million. The differences in the number of reads obtained per sample could be the bias in the sample concentrations, although equal concentrations are calculated before pooling the two samples together in a run. One sample may have had a higher concentration than the other, thus more reads. The minimum reads expected in WES runs per sample are between 25-65 million reads using the Ion Torrent 540 chip (thermofisher.com/genestudio), The average reads obtained in the current WES was 30 million which is within this range . The WES run report summary is shown in Appendix 7.1 indicating usable reads, total bases, key signal and ISP loading obtained on the Ion-540^{TM} chip 4 in run 2, an example of how to assess this is shown

in Appendix 7.2. Samples were sequenced adequately at more than 50 reads covering the targeted exome regions. More than 40 000 variants were identified per sample and this was in line with the estimated number of variants which is said to be more than 20,000 variants per patient (Ross, Dion and Rouleau, 2020).

Using ACMG criteria, short-listed VUSs favouring pathogenic (without any benign codes), likely pathogenic and pathogenic variants observed in the samples were short-listed and interpreted together with the patient's phenotype. Variants that could contribute to the phenotype in any way were further interrogated and their IGV images (indicating the sequence coverage or read-depth at each variant site) are shown. The following sub-sections, with each patients' unique code, will describe the variants that were shortlisted for in each patient.

a) Patient FRASC24

This patient is a six year nine month old male, presenting with deep palmar creases, eczema, hypertelorism, webbed neck, low set ears, low posterior hairline and shield chest. The patient had a clinical diagnosis of Noonan syndrome. Out of 59617 variants identified for this patient, prioritised variants were identified in three genes, *RHOBTB2, HYDIN* and *IGHMBP2* (Table 3.23).

Table 3.23: Shortlisted variants for FRASC24, with the associated disease, mode of inheritance, zygosity, coverage depth, ACMG classification as well as interpretation

Shortlisted variants	Zygosity	Depth	Disorder	Inheritance	ACMG	Interpretation
RHOBTB2:NM_015178.3: c.617C>G (p.Pro206Arg)	Het	204x C=50%:G=50%	Developmental delay and Epileptic Encephalopathy	Majority AD	VUS favouring pathogenic:PM2,PP2, PP3	limited contribution to phenotype (Adachi et al., 2012)
IGHMBP2:NM_002180.3 c.832C>G((p.His278Asp)	Homo	72x G=100%	Spinal muscular atrophy with respiratory distress type 1 (SMARD1)	AR	VUS favouring pathogenic PM2,,PP3	Disease not linked to phenotype
*HYDIN:*NM_001270974.2: c.6259C>T (p.Arg2087Cys)	Het	296x C=53%:T=46%	Primary ciliary dyskinesia	AR	VUS: PM2,PP3	Disease not linked to phenotype

Het: Variant observed in heterozygous state, Homo: Variant observed in homozygous state, AD: Autosomal recessive disorder, AR: Autosomal recessive disorder, ACMG classification: Appendix 8

All variants for this patient had uncertain significance and have not been reported on disease variant databases such as ClinVar and HGMD. Our interpretation of these variants were inconclusive because we did not have enough data to confidently discard or classify any of these variants as the most likely disease causing variant. However, variants in the *RHOBTB2* gene are reported to have limited contribution to RAS-MAPK phenotypes observed in the patient (Adachi et al., 2012). The variant c.617C>G identified in this gene is shown in (Figure 3.28). Future re-analysis of these variants including revisiting clinical phenotyping and considering alternative mutation screening strategies may benefit the patient.

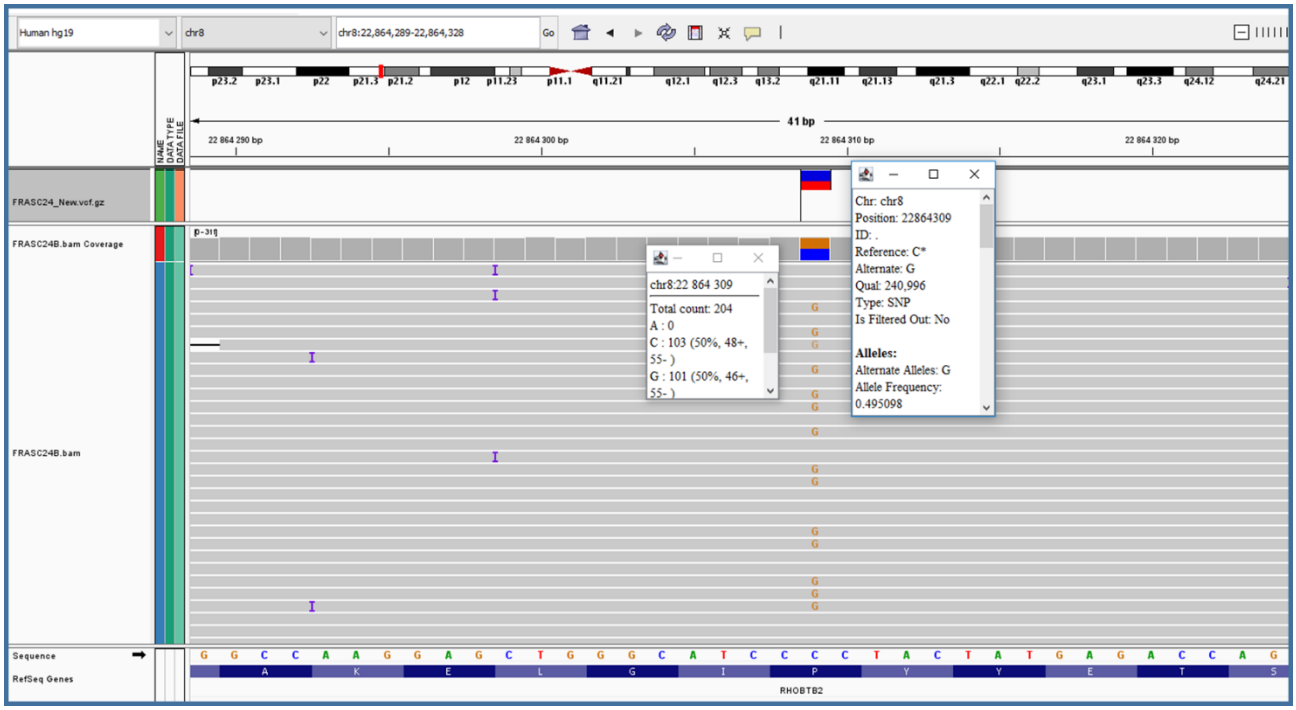


Figure 3.28: IGV image of FRASC 24 showing the region of *RHOBTB2*: c.617C>G heterozygous variant.

b) Patient FRASC55

This patient was a three months four days old male baby at recruitment presenting with bilateral choanal atresia, tricuspid regurgitation and dysmorphic features. A total of 44627 variants were identified for this patient. Of those the four variants in Table 3.24 were shortlisted for clinical review. This patient demised before this study was performed and informed consent was given by the mother and grandmother.

Table 3.24: Shortlisted variants for FRASC55, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification

Shortlisted variants	Zygosity	Depth	Disorder	Inheritance	ACMG	Interpretation
DNAH9:NM_001372.4: c.7169A>C (p.Glu2390Pro)	Homo	44x C=100%	Primary Ciliary dyskinesia	AR	Likely pathogenic: PVS1 and PM2	Partial contribution to phenotype
RTEL1:NM_032957:c.2487delG (p.Ser830HisTer98)	Homo	50x delC	Pulmonary Fibrosis and bone marrow failure	AR	Likely pathogenic: PVS1 and PM2	Disease not linked to phenotype
ACTL6B:NM_016188.4:c.1277C>T (p.Pro426Leu)	Het	90x C=71%:T=29%	Intellectual developmental disorder with severe speech and ambulation defects	AD	VUS: PM2, PP2,PP3	Disease not linked to phenotype
1DH1:NM_005896.2: c.964G>A(p.Gly322Arg)	Het	52x G=67%: T=33%	Metaphyseal chondromatosis with increased urinary excretion of D-2-hydroxyglutarate, IDH1-related	AD	VUS: PM2, PP2,PP3	Disease not linked to phenotype

Het: Variant observed in heterozygous state, Homo: Variant observed in homozygous state, AD: Autosomal recessive disorder, AR: Autosomal recessive disorder, ACMG classification: Appendix 8

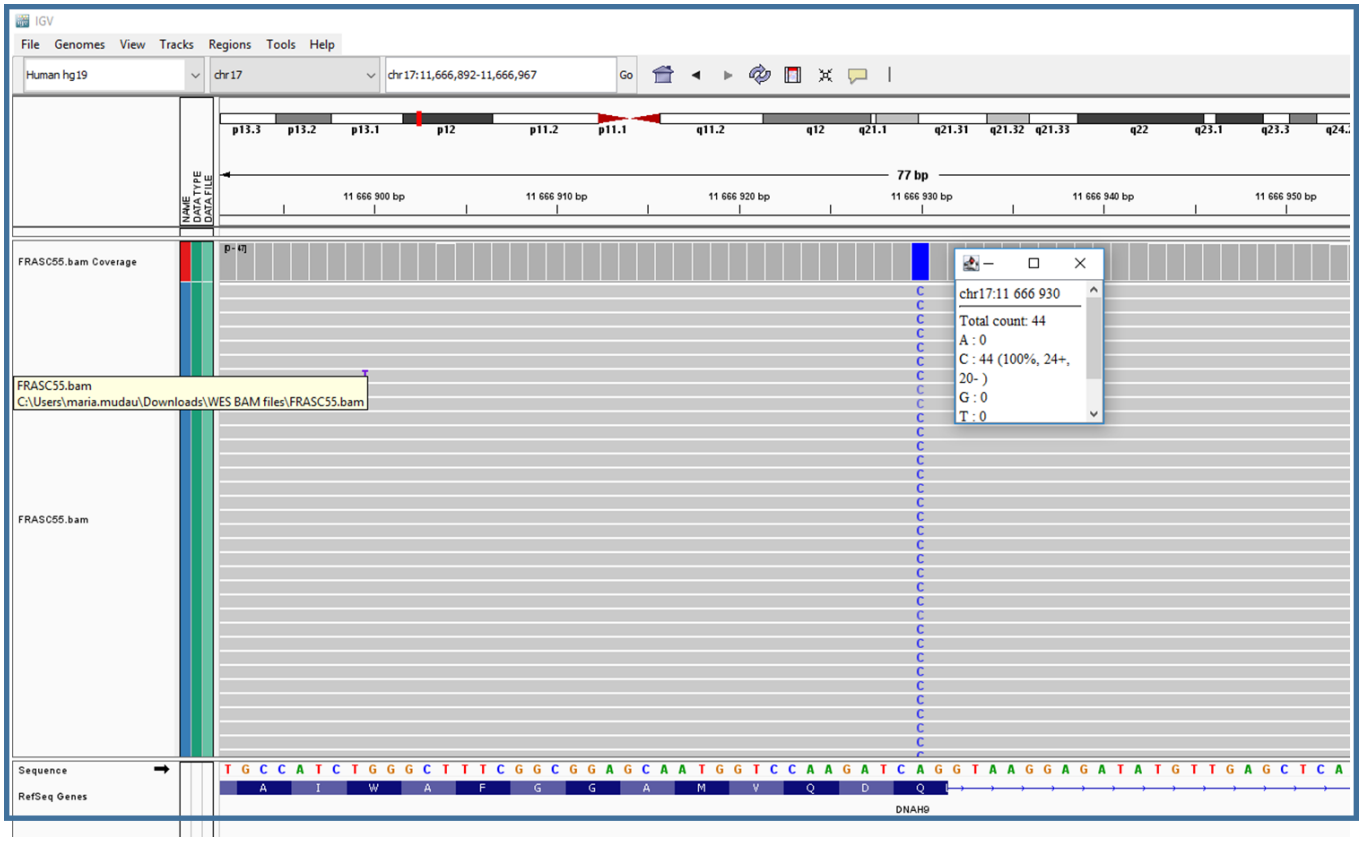


Figure 3.29: IGV image of FRASC55 showing *DNAH9:*c.7169A>C variant in homozygous state.

The variant shown in Figure 3.29 is at a splice site and has been classified as VUS by a single submitter on ClinVar (VCV001210273.1), who only identified one copy of this variant. No homozygotes are reported on gnomAD database, and therefore the ACMG rule, PM2 is applied here as it suggested that homozygous variants in this gene are rare in the population and could potentially contribute to the disease, Table 3.24. The *DNAH9*: c.7169A>C variant was identified in a homozygous state, at the splice-site, it may affect the RNA splicing process, leading to the loss of function of the *DNAH9* gene (ACMG rule, PVS1 applied). Additionally this gene is associated with Primary Ciliary dyskinesia and the clinical features of this disorder overlap with those observed in FRASC55 such as bilateral choanal atresia (Stenner and Rudack, 2014).

c) Patient FRASC56

This is a one-year-old male patient, presenting with clinical features suggestive of having Noonan syndrome. He has deep palmar creases, deep-set nails, frontal bossing, velvety skin,

ptosis, hypertelorism, kyphosis, mild intellectual delay, full lips, low set ears, posteriorly rotated ears; joint laxity, scoliosis, short stature, progressive microcephaly, sparse scalp hair, low-set posteriorly rotated ears. A total of 44543 SNPs were identified, after filtering. The variants in two genes in Table 3.25 below were shortlisted for clinical review.

Table 3.25: Short-listed variants for FRASC56 with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification.

Shortlisted variants	Zygosity	Depth	Disease	Inheritance	Classification	Interpretation
PLXNA1:NM_032242.3:c.613C>T (p.Arg205Trp)	het	75x C=47%;T=53%	Developmental encephalopathy, PLXNA1-related	AD	VUS: PM2,PP3	Limited contribution to patient's phenotype
*GIGYF1:*NM_001375761.1:c.1390G>A (p.Ala464Thr)	Het	94x G=61%;T=39%	*GIGYF1*-related developmental disorder	AD	VUS:PM2	Limited contribution to patient's phenotype

Het: Variant observed in heterozygous state, AD: Autosomal recessive disorder, ACMG classification: Appendix 8

It can be seen on Table 3.25 that only variants of uncertain significance were observed for this patient. Both variants are novel and have not previously been reported in disease variant databases such as ClinVar or HGMD. The *PLXNA1*c.613C>T variant has limited contribution to the phenotype as the major feature of PLXNA1-related Developmental encephalopathy is epilepsy (Park et al., 2017) and FRASC56 does not have epileptic features.

For the *GIGYF1* c.1390G>A variant, there is paucity of information in the literature, however there are few sequence and copy number variants in the *GIGYF1* gene reported in DECIPHER to be associated with intellectual disability, the disease mechanism of this gene is poorly characterised https://www.deciphergenomics.org/gene/GIGYF1/overview/clinical-info. The only feature linking this variant to the patient's phenotype is the intellectual delay observed which is insufficient and has limited contribution to the disease. Future re-analysis of these variants including revisiting the clinical phenotyping, segregation analysis and considering alternative mutation screening strategies may benefit the patient.

d) Patient FRASC58

The patient is an eight-year-old female, suspected of having a RASopathy. This patient presented with protruding ears, dystrophic nails, full lips, large mouth, hyperpigmentation of the skin, hyperkeratosis, camptodactyly of fingers, coarse facial features, small nails, short fifth metacarpalsand severe intellectual disability. 49784 variants were identified, after filtering, the variants in three genes in Table 3.26 below were shortlisted for clinical review.

Table 3.26: Putative disease-causing variants for FRASC58, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification

Shortlisted variants	Zygosity	Depth	Disease	Inheritance	Classification	Interpretation
*RYR3:*NM_001036.6*:*c.10693C>G (p.Leu3565Val)	Het	51x C=37%:G=63%	developmental and epileptic encephalopathy	AD	VUS PM2,PP3-	Limited association to phenotype.
*WFS1:*NM_006005.3*:*c.2407G>A (p.Val803Met)	Het	163x G=61%:A=49%	Wolfram Syndrome	AR	VUS:PM1,PM2,PP3	No association to patient's phenotype
*AGK:*NM_018238.4*:*c.445A>G(p.Ile149Val)	Het	77x A=40%:G=60%	Sengers syndrome	AR	VUS:PM2,PP3	No association to patient's phenotype

Het: Variant observed in heterozygous state, AD: Autosomal recessive disorder, AR: Autosomal recessive disorder, ACMG classification: Appendix 8

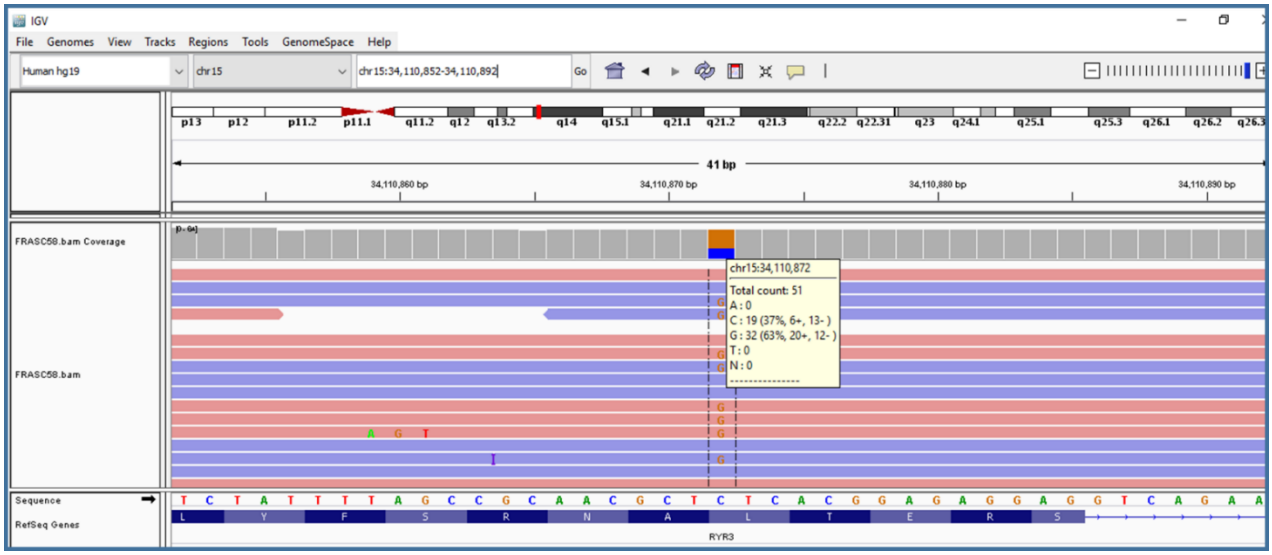


Figure 3.30: IGV image of the *RYR3*: c.10693C>G heterozygous variant at 51x read depth

The variant in Figure 3.30 has limited contribution to disease, currently very limited disease association in the literature linked to variants in *RYR3* gene (EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, and Epi4K Consortium, 2014). However, the severe intellectual ability and some features observed in this patient maybe associated with the *RYR3*-related developmental and epileptic encephalopathy. This variant has been reported as a VUS in ClinVar by a single submitter (VCV001002882.1), not observed on gnomAD database (PM2) and protein function predictors suggest intolerance (PP3). There is limited information in the literature on this variant and that makes the significance of this variant uncertain, further testing of the unaffected family members of this patient and refining the clinical phenotype may assist in clarifying the significance of this variant.

The *WFS1* c.2407G>A heterozygous variant was shortlisted as it had multiple predictors suggesting damaging effect (PP3), not found in gnomAD database (PM2) and also in a hot-spot region for Wolfram syndrome variants (PM1). This variant has been reported as a VUS on ClinVar by a single submitter (VCV000229643.2). This variant has no contribution to disease phenotype observed in FRASC58 as only one variant was identified and Wolfram syndrome is a recessive disorder. According to the literature 95% of pathogenic variants can be detected using sequence analysis and ~5% can be detected using deletion/duplication analysis (Tranebjærg, Barrett and Rendtorff, 1993). It could be suspected that the second variant if available was a CNV, which could not be detected by current assay. However, features associated with this syndrome are progressive optic atrophy, hearing loss and

progressive neurologic abnormalities which our patient does not have (Tranebjærg, Barrett and Rendtorff, 1993). Thus, it is unlikely that a second variant would be identified in this patient

The *AGK* c.445A>G heterozygous variant observed in this patient has been reported as a VUS on ClinVar by a single submitter (VCV001198422.2). However, it does not have any association with the patient's phenotype as *AGK* variants are associated with Sengers syndrome where patients have cardiomyopathy or skeletal myopathy (Guleray et al., 2020; Wang et al., 2021). Additionally, only one copy was observed for this patient and Sengers syndrome is autosomal recessive disorder, the second variant if available was not detected by current assay

For FRASC58 screening for clinically significant variants using a different method, including unaffected family members during testing and re looking at clinical phenotype and refine it better may assist in clarifying the significance of the variants identified.

e) Patient FRASC80

The patient is a three-year-old male patient presenting with clinical features suggestive of NS: velvety skin, deep-set nails, hypertrophic cardiomyopathy, shield chest, undescended testes, sparse hair, ptosis, hypertelorism, low set ears, and mild intellectual disability. A total of 60606 SNVs were obtained and analysed for this patient. The variants in genes *NOTCH3* and *CDK13* were shortlisted and interpreted together with the phenotype (Table 3.27).

Table 3.27: Putative disease-causing variants for FRASC80, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification

Shortlisted variants	Zygosity	Depth	Disorder	Inheritance	Classification	Interpretation
*NOTCH3:NM_000435.2:*c.3848delG (p.Gly1283fs)	Het	75x G= 60%: delG=40 %	Lateral meningocele syndrome	AD	Pathogenic : PVS1, PM2,PP3	Unknown contribution to phenotype
*CDK13:*NM_003718.4:c.2149G>A (p.Gly717Arg)	Het	114x G=56%:A =44%	Dysmorphic facial features and intellectual development disorder with or without congenital heart defects	AD	Pathogenic : PS1, PM1, PM2, PP2, PP3,PP5	Full contribution to phenotype

Het: Variant observed in heterozygous state, AD: Autosomal recessive disorder, ACMG classification: Appendix 8

The *NOTCH3*:c.3848delG variant was shortlisted as it is associated with Lateral meningocele syndrome which has some clinical features that overlap with those of the patient's (Ejaz, Carter and Gripp, 1993). However, this variant was identified in a homopolymer region (string of 'G's) therefore unreliable, as it could be a false-call due to a sequencing artefact.

The *CDK13*: c.2149G>A variant has been reported as pathogenic in ClinVar by multiple submitters (VCV0000375737.13). For this reason, this variant is most likely disease-causing variant in our patient as well. The dysmorphic facial features and intellectual development disorder with or without congenital heart defects disorder has many the clinical features overlapping with those in FRASC80. The mutation is thus likely to be causative of the phenotype of FRASC80 patient (Hamilton et al., 2018). The *CDK13*: c.2149G>A variant was further validated with Sanger sequencing (Figure 3.31) and was identified in the child, but not the mother.

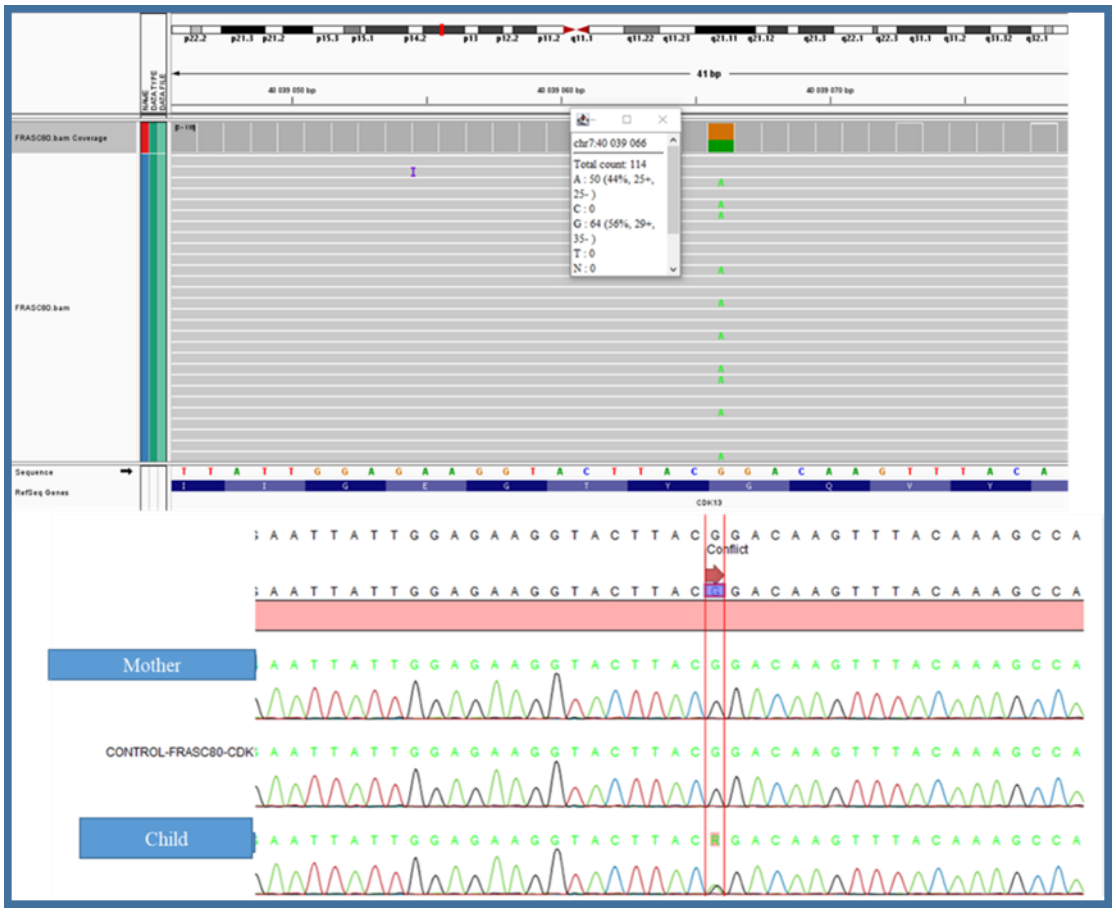


Figure 3.31: IGV image and Sanger sequencing electropherogram of the *CDK13*: c.2149G>A variant at 114x read depth and variant was observed using Sanger sequencing.

f) Patient FRASC96

This patient was a three months old female baby at recruitment with clinical features suggestive of Noonan syndrome: strabismus, aplastic/hypoplastic toenails, single palmar creases, wide intermammillary distance, posteriorly rotated ears, webbed neck, unilateral ptosis, downslanted palpebral fissures, small head circumference, height less than 3rd percentile. 60692 SNVs were obtained and analysed for this patient. Three variants in *MYH6, DNAH5 and MYH3* (Table3.28) were shortlisted as possible disease causing variants and were interpreted in relation to the patient's phenotype.

Table 3.28: Shortlisted variants for FRASC96, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification

Shortlisted variants	Zygosity	Depth	Disorder	Inheritance	Classification	Interpretation
*MYH6:*NM_002471.4:c.733T>A (p.Pheu245Ile)	Het	178x T=62%: A=37%	Primary familial dilated Cardiomyopathy	AD	VUS: PM2, PM5,PP3	No contribution to phenotype
*DNAH5:*NM_001369.3:c.13448C>T (p.Thr4483Met)	Het	167x C=53% :T=46%	Primary ciliary dyskinesia	AR	VUS PP3, PM2	No contribution to phenotype
*MYH3:*NM_002470.4:c.3131A>T:p.Lys1044Met	Het	90x A=42% :T=58%	Contractures, pterygia, and variable skeletal fusions syndrome 1B	AR	VUS favouring Likely Pathogenic: PM2, PP2, PP3	No contribution to disease
*MYH3:*NM_002470.4:c.3056A>C:p.Asn1019Thr	Het	38x A=40% :C=60 %	Contractures, pterygia, and variable skeletal fusions syndrome 1B	AR	VUS favouring Likely Pathogenic: PM2, PP2, PP3	No contribution to disease

Het: Variant observed in heterozygous state, AD: Autosomal recessive disorder, AR: Autosomal recessive disorder, ACMG classification: Appendix 8

All shortlisted variants observed (Table 3.28) in this patient are associated with diseases that do not overlap with clinical features observed in our patient, they are all not reported in ClinVar disease variant database and are unlikely to contribute to the pathogenesis of disease of FRASC96. Although the *MYH3* variants, c.3131A>T and c.3056A>C (Figure 3.32) have no contribution to disease phenotype, these variants were consistently highlighted up by all analysis methods used to annotate variants in the current study. These variants have not been reported in ClinVar or HGMD databases (suggesting novel), not found in gnomAD database (PM2), missense variants in this gene are not normally tolerated (PP2) and multiple protein predictors suggest pathogenicity (PP3). This would require further investigation.



Figure 3.32: IGV image of the *MYH3*:c.3131A>T and *MYH3*:c.3056A>C heterozygous variants at 90x and 38x read depth with ~60% and 40% allele fractions. These variants were observed on Ensembl VEP, Moon diploid as well as the DDD G2P pipeline. Both these variants have not been reported on ClinVar

FRASC96 did not have any formal cardiac assessment at the time of recruitment and the variants identified are partially linked to cardiac abnormalities. For this reason, currently variants identified for this patient are not associated with the diseases concerned. Testing of other family members and refining clinical features may assist in understanding the genetic aetiology of this patient's phenotype better. Particularly testing of the parents to determine if the two *MYH3* variants are inherited and in *cis* or *trans*, this would assist in clarifying the significance of these variants.

g) Patient FRASC97

This patient is a 2-year-old male patient with clinical features of Noonan syndrome. Those included, Ptosis, full lips, webbed neck, posteriorly rotated ears, undescended testes, coarse facial features, low-set-posteriorly rotated ears, sparse hair and downslanted palpebral fissures. A total of 60579 SNPs were obtained and analysed for this patient. Table 3.29 summarises putative disease causing variants for this patient.

Table 3.29: Putative disease-causing variants for FRASC97, with the associated disease, mode of inheritance, zygosity, coverage depth as well as ACMG classification

Shortlisted variants	Zygosity	Depth	Disorder	Inheritance	Classification	Interpretation
SETD1B:NM_001353345.2:c.3065_3067dupGCC:p.Arg1022dup	het	179x dupGCC	Intellectual disability, epilepsy and autism, *SETD1B*-related	AD	VUS: PP3, PM2	No contribution to disease
*TRPS1:*NM_014112.5:c.3403G>C:p.Pro1135Ala	het	60x G=52%:C=48%	Trichorhinophalangeal syndrome, type I	AD	VUS: PM2,PP2,PP3	Contributes to disease

Het: Variant observed in heterozygous state, AD: Autosomal recessive disorder, ACMG classification: Appendix 8

The *SETD1B*:c.3065_3067dupGCC variant has no contribution to the phenotype observed in our patient, the area is CG-rich and the variant is unreliable and could be associated with technical artefact usually observed in GC rich regions during sequencing reaction.

The *TRPS1*:c.3403C>G variant could contribute to the disease phenotype due to unusual facial features that overlaps in Trichorhinophalangeal syndrome I and Noonan syndrome patients.

This variant was further validated using Sanger sequencing (Figure 3.33). The father of FRASC97 patient was found to have the variant but he was not formally assessed during recruitment. This variant could be a novel familial variant as it is not observed in ClinVar or HGMD disease variant databases. The variant has not been observed in gnomAD (PM2), the gene does not tolerate missense variants (PP2) and multiple in-silico protein predictors predict pathogenic (PP3). Further clinical phenotyping of FRASC97 family members including the father, against the variant identified, the significance of this variant may be clarified.

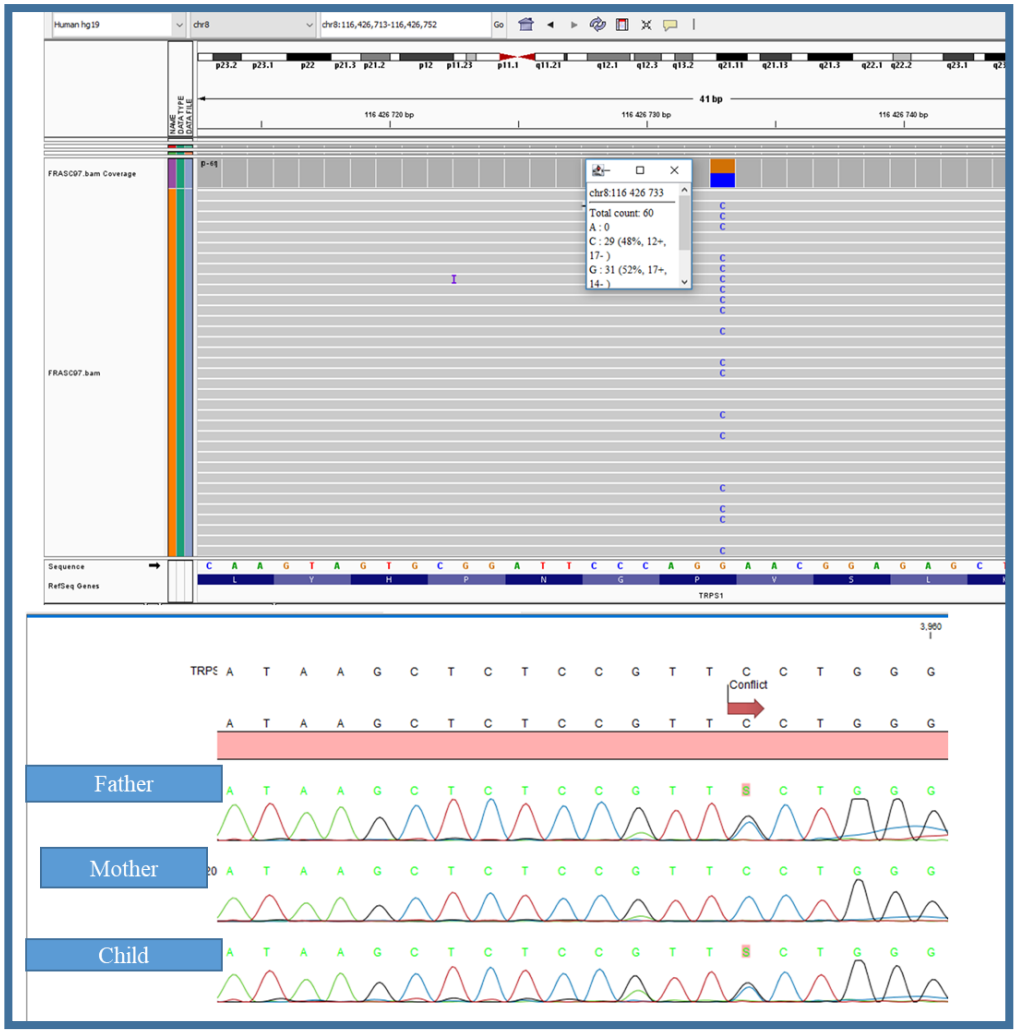


Figure 3.33: IGV image and Sanger sequencing electropherogram of the *TRPS1*:c.3403C>G heterozygous variant at 60x read depth and variant was observed using Sanger sequencing in both child and father.

CHAPTER 4

Discussion and Conclusion

4.1. Summary of the study rationale

It is critical to diagnose genetic disorders timeously and accurately, as this benefits the patient by enabling a better understanding of their prognosis and more specialized management and recurrence risk. Currently there is no molecular confirmation for many monogenic disorders that are diagnosed during clinical examination for the fortunate patients that can at least access clinical genetic service. In South Africa and the African continent as whole, genetic services in the state hospitals are very limited; this leads to misdiagnosis or non-diagnosis of genetic diseases, which further affects the way the disorders are managed. Next generation sequencing technologies have provided significant breakthroughs in the medical science; utilizing this technology enables scientists to study multiple genes simultaneously in one massively parallel reaction and allows for hypothesis free mutation screening. The approach taken in the current study aimed to utilize NGS technology to address the challenges of limited genetic services. The main goal was to develop an affordable approach to generate mutation profiles in a targeted group of monogenic developmental disorders. The strategy here was to design a NGS targeted panel that contains genes associated with multiple monogenic disorders, so that more than one genetic disorder can be studied simultaneously.

In this chapter, we discuss the performance of the FRASC NGS custom-designed targeted gene panel as well as the cost effective strategy of using the panel. This is followed by the contribution of known genes to the genetic profiles of RASopathies in this study and the role of WES when no pathogenic variants are identified using the custom panel.

4.2. Design and **Optimization run for the FRASC targeted gene panel**

The FRASC panel was optimized using eight patient samples; two (NS), three (TCS) and three (CdLS). The success of this run indicated that the panel's performance was promising and could be used with other MiSeq reagents kits. Although the panel was optimised on Treacher Collins syndrome, CHARGE syndrome, Cornelia de Lange syndrome as well as RASopathy syndromes, this approach of combining conditions to maximize sequencing capacity and analyzing many rare conditions simultaneously on one test, can be extended to other genetic disorders.

Evaluating and optimizing the performance of a targeted panel is recommended in the literature, particularly when the panel will be used in a diagnostic setting (Cleveland et al., 2018; Gulilat et al., 2019; Hume et al., 2019). The data obtained on the eight samples used for optimisation of our targeted panel indicated the effective performance of the panel as for all eight samples the data had base call quality of >Q30 and the depth of coverage was above the recommended minimum of 30x (Sims et al., 2014b; Petrackova et al., 2019b). A 100% sensitivity was achieved in the samples used for optimization,. The variants observed on the targeted panel were identified/confirmed when Sanger sequencing was used. The findings of our optimisation run was as expected as NGS has been shown to achieve high detection rate compared to traditional ways of mutation screening in various studies (Stasik et al., 2018; Singh, 2020b).

This panel was optimized on the Illumina MiSeq platform, but the approach is amenable to smaller bench-top NGS platforms such as the Illumina iSeq and Ion-torrent NGS platforms, which requires smaller capital investment, as well as other NGS chemistries. As many diagnostic laboratories move towards using NGS in low resource environments, cost-effectiveness should be a high priority due to the number of competing demands in the healthcare sector.

4.2.1. Cost-effective strategy of using the FRASC panel

Prior to the current study, testing for RASopathies was not performed as the genetically heterogeneous nature of the conditions and the limitations of Sanger sequencing made it too costly to test these group of disorders. In our positive patient cohort, 42% had testing that was largely inappropriate and possibly being done because was the only option available at the time. This amounted to the State paying ~ an average of US$291,53 (>R4300), in a low resource setting like ours, this money would be better spent on appropriate test rather than an

inappropriate but available test. Overuse of diagnostic testing in healthcare has been reported in various undiagnosed patients, which often results in more specialists visits, additional diagnostic procedures, and poor management that is often costly and unnecessary (Bouck et al., 2020; Radomski et al., 2020; Müskens et al., 2022).

The current study demonstrated that using an NGS targeted panel, each patient was sequenced at a cost of ~US$849,68 (R14186,20) and that covered all genes included in the targeted panel associated with the three groups of disorders including their differential diagnosis. Without NGS approximately US$21799,80 (>R300000) was estimated as the theoretical cost if Sanger sequencing was performed on all RASopathy genes together. At the Division of Human Genetics, molecular laboratory, if we were to perform Sanger sequencing in the hot-spot regions of the gene or common genes associated with a particular RASopathy. For example *PTPN11* gene sequencing would be performed for NS patients; this would cost ~R18639,00 for all 15 exons. For CFC patients, *BRAF* gene would sequenced at ~R22367,16 covering all exons and *NF1* gene would be screened for NF1 patients and that would cost~ R75799,82 for all 61 exons. This shows that even testing common or single genes per disorder would still be expensive compared to the NGS targeted panel.

A study done by Pruneri and collegues comparing single gene testing and NGS also found that NGS was cost-effective and their cost ranged from US$239,84 - 1498,05 when NGS was used compared to US$483 - 3838,96 for single gene testing (Pruneri et al., 2021). Targeted gene panels are still considerably cheaper than other NGS-based methods in LMIC. For this reason, there is an increased interest in using this approach for mutation variant screening of genetic diseases with relatively limited locus heterogeneity (Castellanos et al., 2017; Phillips et al., 2021).

The strategy employed in our study is cost-effective as it allows batching of samples from patients with different diseases in a single run, an approach that can be utilised with rare and less frequently ordered molecular diagnostic tests. It would reduce both sequencing costs as the sequencing capacity of the chip is maximised, while allowing for mixed batches to be run, thus reducing turn around time. This is feasible and cost-effective as the minimum tier pricing for the panel was used, this can be employed for the molecular diagnosis of a wide range of rare monogenic disorders in resource-constrained environment.

4.2.2. Utilizing the FRASC targeted gene panel to sequence RASopathy patients

As already indicated in the results chapter 3.2, the FRASC panel was optimised successfully and used to sequence the study cohort of 60 patients with features suggestive of RASopathies. The read depth of sequencing which indicates the number of times the targeted subset of the genome is sequenced was also above 30x which is reported to be sufficient for germline testing (Hume et al., 2019). According to (Jennings et al., 2017b; Gulilat et al., 2019) a minimum of 30x-50x when balanced forward and reverse reads are equally presented is usually an acceptable coverage for human genome mutation screening.

Our results show that all the kits employed in the current study yielded data with >90% base calls of >Q30 quality score and coverage obtained per regions targeted during sequencing was an average of 170x, which is higher than the minimum recommended read depth. This indicates the potential diagnostic utility of the FRASC targeted gene panel, for use in the subsequent NGS runs as and when needed. The diagnostic utility in this context refers to the FRASC targeted panel being developed as a diagnostic tool to perform mutation screening for a diverse range of genes, covering a number of unrelated disease groups, in a limited resource setting.

4.2.3. Quality control of the NGS data

It is imperative to have a way to identify and filter out false-positive variants from true positives in any data analysis platform (Manley, Ma and Levine, 2016; Ma, 2019). In the current project, the Agilent SureCall analysis software was utilized, this software uses built in algorithms for variant allele threshold and other parameters.

a) False-positive variants

When the RASopathy data were analysed there were two pathogenic variants that were each observed in more than one patient. The *PTPN11*:c.923A>G was observed in 10 patients where five had another pathogenic variant. It was evident that this was an unexpected finding because mode of inheritance for RASopathies is autosomal dominant meaning that only one pathogenic variant causes the phenotype. Upon manual inspection of the BAM files on IGV in all the samples that it was observed in, it was clear that there was some ‘A’ ref allele and ‘G’ alternative allele bias. Only one of the 10 *PTPN11*:c.923A>G variants was a true variant observed in FRASC66, as confirmed on Sanger sequencing and allele ratio. The second false positive variant observed was the *NF1*:c.27G>A, this novel pathogenic variant was observed

in nine patients and only one was found to be a true variant in FRASC67 after sequencing and adjustment of VAF. Sanger sequencing has been used to differentiate true positive and false positive variants in NGS data (Baudhuin et al., 2015b; Beck et al., 2016; Mu et al., 2016b). Detecting false positives during NGS data analysis is not uncommon; several publications have reported on detecting false positives and how fine-tuning by changing parameters settings for variant calling reduced false positive rates by 27% (Gampawar et al., 2019; Muyas et al., 2019).

b) Quality metrics of the false-positive variants

When comparing to the samples with *PTPN11* false positive variants, none of the patients with the *NF1*:c.27G>A variant had a different second pathogenic variant. The other observed difference was the read depth in these variants, the average read depth of *PTPN11*:c.923A>G false positive variant was 189.5x and 42.5x for the *NF1*:c.27G>A false positive variant. Ideally the analysis software should not call variants with allele bias and low coverage, as reported in various publications where false-positive variants were identified and investigated (Caspar et al., 2018; Ma, 2019; Pereira, Oliveira and Sousa, 2020).

c) Variant allele frequency threshold

When the VAF threshold was increased to 30%, the VCF files were reanalysed and the *PTPN11*:c.923A>G false positive variant could no longer be observed. However, the *NF1* false positive variant still remained when the VAF thresholds were adjusted, this could be explained by the lower average coverage depth of 42.5x compared to the average coverage of 189.5x of the *PTPN11*. The sample from FRASC83 was included for sequencing using another targeted panel (IDP on the ThermoFisher Ion torrent platform) for the purpose of investigating where the *NF1*:c.27G>A false-positive variant calls may have originated as the VAF skewing was not as clear as the *PTPN11* variant. This finding demonstrated that when using our custom-designed targeted panel, the coverage depth of the variant site should be above 60x reads when analysing using the Agilent SureCall software and the VAF threshold should be adjusted to 30% as with other analysis software (Strom, 2016).

d) Reducing false-positive calls in future

It is reported that the sequencing coverage level often determines whether variant identification can be made with a certainty at particular base positions (Roy et al., 2018; Hume et al., 2019). In our case we have certainly determined that reads below 50x may skew analysis. The putative

disease-causing variants identified using the custom panel in our current study were validated using Sanger sequencing. This was done in order to optimise and fine tune our protocol and to know when or when not to use Sanger sequencing as this was the first NGS study in our laboratory. This validation together with fine tuning of parameters used for NGS data quality control procedure has enabled us to determine when to perform Sanger validation as described in various guidelines for NGS implementation in a diagnostic setting (Aziz et al., 2015; Matthijs et al., 2016; Hume et al., 2019).

In our current setting, Sanger sequencing would be recommended where variant calling is uncertain, such as from homopolymer regions or poorly covered regions of the genome (below ~50x reads at the variant site). Furthermore, Sanger sequencing would be performed where there is skewed allele ratios of heterozygous variants (when the variant is not ~50% between the wild-type reads and the reads with the pathogenic variant) and for cascade testing in families. The decision on performing Sanger sequencing was based on our finding as well as the recommendations in various studies (Baudhuin et al., 2015; Beck et al., 2016; Mu et al., 2016).

From the current using our data and learning from the false-positive variants identified, we recommend that when using the Agilent SureCall analysis software, the reliable read-depth at variant site should be 50x. Below 50x read depth the possibility of obtaining false-positive calls increases. The outcome from these false-negative investigations has assisted us in determining the acceptable variant allele ratio and the coverage for future sequencing runs. It is recommended in the ACMG guidelines that laboratories will determine their own quality thresholds depending on the NGS platform used and the validation data (Rehm et al., 2013; Rehder et al., 2021). Going forward parameters such as read depth, sequence quality, allele fraction as well as manual inspection on IGV should be assessed to distinguish between false positives and true positives. Once we were able to distinguish false positives from true positives, we were able to accurately calculate mutation detection rates in our cohort.

4.3. Genetic profiles obtained in the current study

Out of 21 RASopathy genes included in the panel, no clinically significant variants were identified in 12 of the genes (*CBL, KRAS, RIT1, SOS2, RASA1, SPRED1, NF2, HRAS, SHOC2, A2ML1, RASA2* and *RRAS*). Pathogenic variants in these genes are rare in most RASopathy patient cohorts (Aoki et al., 2016; Jafry and Sidbury, 2020), In addition, none of the recruited patients in this study had a clinical diagnosis of Costello syndrome, Legius syndrome or CM-AVM. Thus, it is not surprising that variants were not identified in *HRAS, SPRED1* and *RASA1* genes respectively

Twenty-nine missense variants were identified in 8 genes (*PTPN11*, *BRAF*, *LZTR1*, *SOS1*, *RAF1*, *NRAS*, *MAP2K1*,*MAP2K2* and *NF1*). This is consistent with the literature, which reports that pathogenic missense variants in these genes account for most non-NF1 RASopathies (Rauen, 2013; Gelb et al., 2018).

The *NF1* gene was the only one where no missense variants were identified but instead, nonsense and frameshifts variants were identified. Our results are in keeping with the literature that most of the germline pathogenic variants identified in NF1 patients appear to cause severe truncation of the gene product (Fahsold et al., 2000; Mao et al., 2018).

4.3.1. Detection rate achieved in the current study

In the current study, an overall detection rate of 56.6% (pathogenic and likely pathogenic variants) was obtained in the 60 RASopathy patients sequenced when the FRASC targeted gene panel was used. Chen *et al.*, 2019 reported that approximately 70% of Chinese patients with a clinical diagnosis of RASopathy had a molecular confirmation (Chen, Li, et al., 2019)., Some studies such as, Bhoj *et al.*, 2017 had an overall detection rate of 19%, where only 24 of their 126 American RASopathy patients had a clinically significant variant identified. This highlights genetic heterogeneity and inconsistency with regards to the criteria used to select patients for RASopathy gene analysis. The authors related this low yield to the random ordering of RASopathy testing from various teaching hospitals and clinics even when patients did not meet the clinical criteria of RASopathies (Bhoj et al., 2017). In some LMICs countries laboratories that do diagnostic testing for NS still use traditional Sanger sequencing targeting *PTPN11* hot spot regions on exon 3 and exon 8. One such study was done on Egyptian patients and this study had the detection rate of 4.7% (Essawi et al., 2013). Detection rate of 29% was obtained for Indian patients with NS and this study used Sanger sequencing analysis of exon 2

to exon 16 of the *PTPN11* gene (Athota et al., 2020). A study done on Brazilian patients with NS utilising denaturing high-performance liquid chromatography (DHPLC) followed by sequencing of amplicons with an aberrant elution profile had the detection rate of 42% (Bertola et al., 2006).

In South Africa a 31.2% detection rate was obtained in a cohort of 16 NS patients utilizing an NGS targeted panel of 14 genes associated with NS (Tekendo-Ngongang et al., 2019b). The papers above indicate the lack of comprehensive testing for RASopathies as a whole in low resource setting. This could be because RASopathies have variable expression of phenotype and may affect selection criteria as well as population heterogeneity.

The cohort of 60 patient studied here were recruited if they had some suggestive clinical features of RASopathies. They were not limited to NS patients only. As the clinical profile in African patients is largely undescribed, (Kruszka et al., 2017; Tekendo-Ngongang et al., 2019b), braod, inclusive clinical criteria were used. Thus a number of patients did not have completely typical RASopathy features.

The 56.6% achieved in our study is fairly high compared to other studies. This study also served as a molecular confirmation to these patients that had no molecular testing prior to this study. What this means in the current context is that for RASopathy patients, using targeted a panel containing genes commonly known to be associated with these disorders is an effective way to offer diagnostic service. The current study illustrates example of an approach that LMICs can start using to introduce NGS technology into their diagnostic setting. In this approach, the costs become affordable, particularly considering high detection rates achieved compared to the traditional way of mutation screening using Sanger sequencing and hotspot analysis (Ari and Arikan, 2016; Matthijs et al., 2016; Hu et al., 2021; Scocchia et al., 2021).

4.3.2. Pathogenic and likely pathogenic variants in *PTPN11* gene

Heterozygous pathogenic variants in the *PTPN11* gene were identified in ten patients. Nine had a clinical diagnosis of NS and one patient (FRASC100) had a clinical diagnosis of NSML. Of the 22 NS patients with a pathogenic or likely pathogenic variant identified, 41% were found in the *PTPN11* gene. This is consisitent with global reports, where approximately 50% of NS patients globally are reported to have pathogenic variants in the *PTPN11* gene (Lee et al., 2011; Rauen, 2013; Bhoj et al., 2017; Lee and Yoo, 2019a). Some studies reported a higher detection rate in this gene, 58% (Lepri et al., 2014; El Bouchikhi et al., 2016). The single South African

study identified pathogenic variants in 5/16 (31%) NS patients, of the five, two (40%) had the variants in the *PTPN11* gene, although the sample size was small (Tekendo-Ngongang et al., 2019), the South African studies indicate that most African NS patients are likely to have *PTPN11* pathogenic variants as well.

All the variants identified in PTPN11 were previously reported as pathogenic in ClinVar and HGMD® disease variant databases. This confirms that *PTPN11* is a well-studied gene for patients presenting with NS and is an appropriate gene to test in African NS patients as the mutation profiles are not different and it accounts for the majority of mutation positive patients, with NS.

All NS patients with *PTPN11* pathogenic variant had clinical features that met the diagnostic criteria of NS; there were no specific differences in the features observed in these patients compared to the literature (Carcavilla et al., 2020). As seen in Table 3.7, 90% of the patients had short stature and low set ears, 80% had epicanthic folds, 70% had heart defects, DSPF, webbed neck and milder/moderate intellectual disability. All male patients with the *PTPN11* variant had undescended testes. All these clinical features are consistent with what is reported in the literature, with no clinical features observed completely consistently in all patients, indicating the clinical heterogeneity observed in NS patients (Rauen, 2013; Athota et al., 2020)

FRASC 2 and FRASC18 had the same pathogenic variant c.922A>G in the *PTPN11* gene. This variant has 47 submissions on ClinVar and was reviewed by the ClinGen RASopathy variant curation expert panel (VCV000013326) and is observed in NS patients from various populations globally, including Italy (Tartaglia et al., 2002), France (Edouard et al., 2010), Spain (Ezquieta et al., 2012), Singapore (Koh et al., 2019) and many more populations (Kruszka et al., 2017). Although it is observed in 2 of the nine NS patients with *PTPN11* variants, we do not have enough data to suggest that this is a common recurrent variant in African populations but the data are supportive. The mothers of these two patients were also diagnosed with NS and they were shown to have the c.922A>G variant. With a family history of NS and an identified pathogenic variant, these patients meet major criteria of NS diagnosis as described by (Carcavilla et al., 2020). Because NS is an autosomal dominant disorder, these mothers have 50% risk of passing this variant to their children and because the variants have been identified, prenatal testing can be offered.

FRASC100, presented with features of NSML and was found to have a heterozygous c.1381G>A variant in exon 12 of the *PTPN11* gene. This further confirmed the diagnosis of

NSML as these features have been previously reported to be associated with NSML (Digilio et al., 2002). More than 90% of the NSML pathogenic variants are found in the *PTPN11* gene with exon 12 reported as being the hot-spot region (Sarkozy et al., 2009), thus this result was as expected.

As seen in Figure 3.9, all likely pathogenic and pathogenic variants identified in the *PTPN11* gene in the current study are all within functional protein domains, with most of these variants (50%) in exon 3 within the Src homology 2 (SH2) domain. The SH2 domain is involved in the tyrosine kinase signalling for the regulation of gene expression through the nucleus and cytoplasm of the cell (Waksman, Kumaran and Lubman, 2004). In addition, 40% of the variants have clustered in the PTP domain in exons 8, 12 and 13. Exon 12 is reported as the hot-spot regions for *PTPN11* pathogenic variants associated with NSML (Athota et al., 2020), and FRASC100 variant was found in this hot-spot exon.

Looking at the importance of the protein domains these variants are found in, the protein tyrosine phosphatases (PTPs) are essential enzymes for the regulation and control of signal transduction pathways. They are involved in the reversal of growth factor signalling and have a critical function of maintaining phosphor-tyrosine levels in the cells. PTPs are essential for targeted treatment and therapy for NS and other RASopathies involving these enzymes, because inhibitors could be designed for these proteins or even methods to elevate the biological activity depending on the aim for therapy (Tabernero et al., 2008).

4.3.3. Pathogenic and likely pathogenic variants in *BRAF* gene

Seven (15.5%) pathogenic variants in 45 non-NF1 RASopathy cohort in the *BRAF* gene were identified in the current study; five patients had a clinical diagnosis of NS while two had a clinical diagnosis of CFC. The *BRAF* gene is reported to be involved in NS pathogenesis by enhancing the ERK/MAPK inactivation. Only 1.7%-2.3% of NS patients have *BRAF* pathogenic variants. The mutations cluster predominantly in conserved regions of CR1 and CR3 (Sarkozy et al., 2009; Lee et al., 2011). Although the detection rate of *BRAF*-positive NS patients were higher in this study than in most publications, this was lower than the one reported by (Xu et al., 2017) where 40% of *BRAF* pathogenic variants were observed in a cohort of 10 NS Chinese patients.

In the current study three of the five (FRASC4, FRASC64 and FRASC73) patients had the same *BRAF* pathogenic variant (*BRAF*: c.770A>G) and had a clinical diagnosis of NS. This

variant could be a global recurrent variant for patients presenting with NS or CFC RASopathies due to *BRAF* as it has been reviewed and classified as pathogenic by the ClinGen RASopathy expect panel (Gelb et al., 2018) and has 22 submissions on ClinVar (Accession:VCV000013973.15). Ten of these submissions associate this variant with CFC, and only two submissions associate this variant with NS and other submissions mention that an unspecified RASopathy is associated with this variant. All five patients with *BRAF* pathogenic variants in the current study had DSPF and epicanthic folds, 80% had hypertelorism and sparse hair. There was insufficient clinical information to exclude CFC completely in all NS cases; there is a need to evaluate these patients further.

All variants identified in the *BRAF* gene in the current study were mapped to the RAS binding and protein kinase protein domains including conserved regions (CR1 and CR3). It has been previously reported that pathogenic variants in the *BRAF* gene cluster predominantly in the conserved regions of CR1 and CR3 (Sarkozy et al., 2009; Lee et al., 2011). It can be seen in Figure 3.13 that three (Q257R, T241M and A246P) of the five are clustered in CR1 indicating a possible hot spot region (Aramini et al., 2015). Two variants (E501V and N581D) were in exon 12 and 14, respectively of the conserved region 3 (CR3) of the protein kinase domain (Figure 3.13). RAF kinases play a very important role in the mitogen-activated protein kinase (MAPK) cascade. All three CR regions are highly conserved among the RAF kinases. As already mentioned five of the variants identified in the *BRAF* gene are clustered in the CR1 region of the RAS binding domain (RBD). This domain plays critical step in the activation of BRAF protein serine/threonine kinase and stimulation of the mitogen-activated protein kinase-signalling pathway (Aramini et al., 2015). The two additional variants identified in the *BRAF* gene are mapped at the CR3 region in the protein kinase domain; this is a structurally conserved protein domain containing the catalytic function of protein kinases. They move a phosphate group onto proteins in a phosphorylation process (Knight et al., 2007). This indicates that the *BRAF* variants identified in the current study are in the critical functional domains for intracellular processes (Manning et al., 2002).

4.3.4. Pathogenic and likely pathogenic variants in *LZTR1* gene

Globally is it reported that a total of ~8% of variants in the *LZTR1* gene have been reported to be associated with NS (Bhambhani and Muenke, 2014; Yamamoto et al., 2015; Motta et al., 2019: 1) and 38% of variants in the *LZTR1* gene are associated with Schwannomatosis, a

distinct form of NF (Dhamija et al., 1993; Piotrowski et al., 2014). In the current study, four variants were identified in the *LZTR1* gene. Three patients with these variants had a clinical diagnosis of NS and one had a clinical diagnosis of Schwannomatosis. The detection rate of 13.6% of *LZTR1*-positive NS patients was observed in the current study, this was higher than that reported in Rauen (2013) and Yamamoto *et al.*, (2015) who reported a detection rate of 8% of *LZTR1*-positive NS patients.

The three patients (FRASC15, FRASC21 and FRASC69) with heterozygous likely pathogenic *LZTR1* variants had clinical features listed in Table 3.11. The clinical features observed in these patients are typical features associated with NS as described in the literature (Yamamoto et al., 2015; Carcavilla et al., 2020). .

All variants observed in the *LZTR1* gene were not confirmed in mothers of these probands and were found in heterozygous state this indicate that variants in the current study follow autosomal dominant (AD) mode of inheritance for NS. Publications such as Johnson *et al*., 2018 and Pagnamenta *et al* 2019 report cases of autosomal recessive NS caused by biallelic *LZTR1* variants (Johnston et al., 2018; Pagnamenta et al., 2019). The variants associated with the recessive NS are mostly splice-sites, nonsense and frameshifts and usually observed as compound heterozygotes in *trans* with a presumed hypomorphic variant (Johnston et al., 2018; Pagnamenta et al., 2019).

Contrary to the variants associated with the autosomal recessive NS which are said to be spread across the *LZTR1* gene, variants associated with the AD form of NS in the *LZTR1* gene are clustered in the Kelch 1 to Kelch 4 domains, particularly in Kelch 4 (Yamamoto et al., 2015; Motta et al., 2019). Kelch proteins mediate protein-protein interactions and pathogenic variants in these are predicted to disrupt protein function (Tidyman and Rauen, 2016).

This observation is similar to the current study, where two of the three variants were observed in the Kelch 4 domain and the third one in Kelch 3 (Figure 3.17). One of the two variants (p.S244P) observed in FRASC15 mapped to the Kelch 4 domain and was also observed in one patient by Pagnamenta et al. (2019) in a *de novo* state further indicating haploinsufficiency and the autosomal dominant mode of inheritance.

Dominant acting *LZTR1* pathogenic variants are said to enhance stimulus-dependent MAPK signalling, which may lead to the protein losing its function (Motta et al., 2019). In the context of our study cohort, the *LZTR1* variants are clustering in the region that has been reported for autosomal dominant *LZTR1* NS variants. The phenotype of the patients and the heterozygous

variant identified also support autosomal dominant NS. Additionally, no hypomorphic alleles were observed in our study cohort.

With regard to the testing strategy, no deletions or duplications *LZTR1* gene variants have been reported to be associated with NS, ~100% of pathogenic variants in the *LZTR1* gene can be identified using sequence analysis (Roberts et al., 2013). The detection rate achieved in our study cohort makes the *LZTR1* gene important to screen in African patients presenting with NS.

The LZTR1:c2269C>T nonsense heterozygous variant was detected in patient with Schwannomatosis, FRASC62, a 58-year-old with a family history of vertigo and high frequency hearing loss (Figure 3.16). Pathogenic variants in the *LZTR1* gene account for 38% of familial Schwannomatosis cases (Dhamija et al., 1993; Piotrowski et al., 2014). The *LZTR1*:c.2269C>T (p.Q757X) variant was found in exon 19 and not within a protein domain, however this is not relevant for variants like this that causes complete loss of protein function (Figure 3.17). This is a truncating variant that lead to shorter coding sequences of the gene and cause the nonsense mediated decay leading to the disorder.

4.3.5. Pathogenic variants in *SOS1* gene

SOS1 pathogenic variants are associated with 10-20% of NS cases (Lepri et al., 2011; Roberts et al., 2013; El Bouchikhi et al., 2016). In this current study (2/22) 9% of *SOS1* pathogenic variants were detected in the two patients presenting with NS. Our detection rate is comparable to the reported literature, although the variants were only identified in two patients it does indicate that *SOS1* is a relevant gene to test in our African NS patients. These two variants have been previously reported on ClinVar in various submissions to be associated with NS. NS observed with *SOS1* gene variants is associated with ectodermal abnormalities and pulmonary valve stenosis, as reported in genotype-phenotype correlation studies (Ko et al., 2008; Yamamoto et al., 2015). The two patients in the current study with the *SOS1* pathogenic variants had pulmonary valve stenosis and other clinical features suggestive of NS. One of the patients, FRASC14 with the *SOS1*:c.806T>C (p.M269T) pathogenic variant can be seen in Figure 3.18, presenting with typical NS features such as unusual faces, mild ID and strabismus. This is not uncommon because patients with NS have variable phenotypic expression of the disorder and parents usually become awaretaht there is NS in the family only after having a more severely affected child (Allanson and Roberts, 1993; van Trier et al., 2017).

4.3.6. Pathogenic variants in *RAF1* gene

Pathogenic variants in the *RAF1* gene are observed in 3-17% of NS cases (Pandit et al., 2007). Two heterozygous pathogenic variants in this gene were identified out of 22 Noonan-causing variants in the current study (9%). These variants, the c.770C>T (pS257L) and c.781C>A (p.P261T) are previously reported in the literature to be associated with NS (Aoki et al., 2016; Tidyman and Rauen, 2016). These two variants are reported to cause the increased kinase activity and MAPK1 activation compared to the wild-type during functional analysis (Pandit et al., 2007; Razzaque et al., 2007). Our patients had typical clinical features observed in NS patients (Lee and Yoo, 2019a; Athota et al., 2020; Carcavilla et al., 2020).

4.3.7. Variants identified in the *NRAS* and *MAP2K2* genes

Gene variants in the *NRAS* and *MAP2K2* genes were observed in one patient each. The *NRAS* variant was observed in a patient with clinical diagnosis of NS (FRASC5) and the *MAP2K2* variant was observed in a patient with clinical diagnosis of CFC (FRASC22). Pathogenic variants in the *NRAS* gene are rare causes for RASopathies and having been identified in one patient in our cohort in consistent with the literature that described variants in these genes as rare causes for RASopathies (Cirstea et al., 2010; El Bouchikhi et al., 2016). The *NRAS*:c.173C>T(p.T58I) pathogenic variant was found in one patient FRASC5, this male patient had undescended testis, and other NS clinical features similar to the ones reported in the literature (Altmüller et al., 2017), there was no particular features that sets this patient apart from other NS patients.

Over 25% of *MAP2K1* and *MAP2K2* germline mutations are associated with CFC syndrome (Rauen, 1993; Nava et al., 2007b; Sarkozy et al., 2009; Pierpont et al., 2014; Chen, Li, et al., 2019). In the current study one patient of the three patients recruited with clinical features suggestive of CFC had a CFC-causing variant in the *MAP2K2* gene(Rauen, 2013). Our CFC cohort was only 3 patients and they all had CFC pathogenic variant identified. Although only identified in one patient in our current study, the *MAP2K2* gene variants are prominent in the CFC patients and the detection rates in our African patients will be better understood in a larger cohort of patients. This patient had lentigines, DSPF, deep palmar creases, sparse hair which are typical clinical features associated with CFC syndrome (Pierpont et al., 2014) and there was no particular features that sets this patient apart from other CFC patients with *BRAF* variants observed in the current study.

4.3.8. Variants detected in NF1 patients

Seven of the 15 NF1 patients enrolled in this study had *NF1* mutations identified, this is a detection rate of about 46.7% for NF1 patients. Over 90% of NF1 pathogenic variants in NF1 patients occur in the *NF1* gene (Rauen, 2013; Koczkowska et al., 2018), similarly, in our current study, 100% of NF1-causing variants were found in the *NF1* gene and no pathogenic variants were observed in the *SPRED1* gene in our cohort. Figure 3.23 is showing the c.3721 C>T variant in the FRASC45 patient and the parents had homozygous wild type variant, and therefore suggesting *de novo* occurrence as it is reported in the literature that *de novo* variants occur on 50% of the NF1 patients (Friedman, 1993). In our study, we could only confirm one *de novo* variant and not on the rest as the segregation, analysis could not be performed in all variants identified due to lack of parent samples.

In the current study six truncating variants were identified (four nonsense and two frameshifts). Three are within the neurofibromin protein domains and the rest are outside the domains. It is reported that most of the NF1-causing variants are truncating variants that disrupt the nonsense mediated decay and result in premature termination leading to loss of function of the neurofibromin protein (Koczkowska et al., 2018; Chen, Xue, et al., 2019: 1; Castellanos et al., 2020).

Only one variant (observed in FRASC38) was a missense variant *NF1:* c.1658A>G (p.H553R) previously reported on ClinVar to be associated with NF1. All variants except the c.27G>A (p.W9X) have been previously reported on Clinvar to be associated with NF1. The c.27G>A variant observed in FRASC67 has never been reported before and it is not found in gnomAD population database, this variant was also observed in the mother of the patient who was not formally assessed clinically and it is unknown if she is affected. The detection rate of 46.7% suggest that in about half of our NF1 cohort, other types of NF1-causing variants may be the cause of the NF1 in these patients other than the SNVs and small indels detected by the custom targeted panel employed in our study.

It is reported that 22%-30% of the NF1 causing variants affect splicing and are often not detected by genomic DNA sequencing as this occur on the mRNA level. For these kind of variants RNA analysis must be employed as it may improve detection rate in the NF1 patients (Evans et al., 2016). About 5%-10% of NF1-causing variants are copy number variations (CNVs) such as deletions, duplications and whole gene deletions (Pasmant et al., 2010; Kehrer-

Sawatzki and Cooper, 2012; Kehrer-Sawatzki, Mautner and Cooper, 2017); for these variants MLPA or microarray analysis (for larger deletions including whole gene) may be employed.

Our detection rate was lower than the yield reported in various publications. More than 90% detection rate was achieved in 73/79 Italian patients that met the NF1 diagnostic criteria (Calì et al., 2017). A study done on Chinese patients with NF1 where testing was done using Sanger sequencing, cDNA analysis and MLPA analysis achieved a diagnostic rate of 89% (Zhang et al., 2015, 2020). This indicates that the multi-step testing improves the detection rate for NF1. When NGS targeted panel was utilised in French patients with NF1 syndrome, an 88% detection rate was achieved, in this study. The analysis pipeline included both SNVs and CNV pipeline (Pasmant et al., 2010, 2015).

Importantly, the testing strategy for NF1 that is reported to identify >95% of the variants is a multi-step testing of the DNA pathogenic variants, RNA splicing variants followed by CNV analysis when no SNVs pathogenic variants are identified (Pasmant et al., 2015; Evans et al., 2016) . This multi-step testing protocol is only recommended when testing is indicated and patients meet the NF1 diagnostic criteria. Particularly in instances where a child has a tumour and upon identifying an NF1 causing variant, the management is immediately affected or in a woman who plans to have children and would like a prenatal testing in the future. Incorporating SNV, CNV and RNA analysis testing strategy in our diagnostic setting will likely improve the detection rate compared to what we currently observed because we only used the targeted panel that detects DNA pathogenic single nucleotide variants only, this excluding CNVs

4.3.9. Variant of uncertain significance

One variant of uncertain significance *MAP2K1:*c.215G>A (p.S72N) was observed in three patients in the current study. One patient, FRASC42 has another pathogenic NF1 variant identified may shift this variant towards likely benign in the future. The ACMG rule BP5 that is applied when variant was found in a case with an alternate molecular basis for disease could possibly be applied in this case, however there are reports on patients with both NS and NF1 syndromes (Zhang et al., 2020) therefore we could not apply BP5 with certainty in this case. So this novel variant remains a VUS, it has not been reported in ClinVar and HGMD® disease variant databases and also it is absent in gnomAD population database.

Analysing genes known to be associated with the RASopathies may have contributed to low rate of VUSs that were obtained in our patient cohort. Because this is targeted analysis, it has

low VUS rate that comes as a benefit of utilizing targeted gene panels (Federici and Soddu, 2020b; Iancu et al., 2021) when compared to WES/WGS (Chrystoja and Diamandis, 2014; Trujillano et al., 2017). This is one of the advantages of utilizing targeted panels in a low resource setting like ours as we will not have lots of VUSs that will not translate into actionable outcome for the patients. *MAP2K1* gene variants accounts for ~2% of NS patients (Allanson and Roberts, 1993; Rauen, 2013) and about 10-25% of CFC patients (Pierpont et al., 2014).

4.3.10. Benign variation observed in the current study

Twenty-four of the 60 (40%) patients had benign variants identified and no confirmation of their diagnosis was achieved using the targeted gene panel. For Non-NF1 cohort of 45, 25 had the variant identified, two had the *MAP2K1* VUS and 18 had no clinically significant variants identified leading to (25/45) 55.5% detection rate for this cohort and (18/45) 40% benign variation. For the NF1 cohort of 15 patients, seven had a pathogenic variant identified (46.6%) and 8/15 (53.3%) had only benign variants identified in the *NF1* gene. The benign variants observed in our cohort of patients indicate that the patients may have other types of variants or variants in genes that could not be identified using the FRASC targeted gene panel. It is widely reported that most NGS targeted panels, including the FRASC panel, covers only the coding regions of the genes included and do not cover the non-coding, untranslated regions, promoter regions and other regions on the genome such as repeats regions (Lionel et al., 2018; Gulilat et al., 2019; Yu et al., 2019). Additionally the analysis pipelines of most targeted gene panel do not include CNV analysis, RNA analysis and structural analysis (Lek et al., 2016a; Lappalainen et al., 2019; Pereira, Oliveira and Sousa, 2020). These may be the reason of identifying no clinically significant variants in 40% of our patients.

4.3.11. Detection rate per clinical diagnosis

This section focuses on the detection rate per clinical diagnosis made in the group of patients recruited. We found variants in patients who were suspected to have NS, NSML, CFC, Schwannomatosis as well as NF1. These are discussed in more detail under each subheading

a) Noonan syndrome

In our cohort of NS patients with clinically significant variants (n=22), *PTPN11* (41%), *BRAF* (22.7%) and *LZTR1* (13.6%) were the genes with the highest observed number of pathogenic variants. This is different from the detection rates reported in the literature where *SOS1* and *RAF1* pathogenic variants are second and third respectively, after *PTPN11* in to NS patients (Ko et al., 2008; Rauen, 2013; Carcavilla et al., 2020).

The higher variant rate observed for *BRAF* and *LZTR1* genes in our cohort could be an indication of unique variation observed in the African genome compared to other global populations (May et al., 2013; Gurdasani et al., 2015; Choudhury et al., 2017). Noonan syndrome patients have high variant rates in *PTPN11*, *SOS1* and *RAF1* respectively in various studies reported for various populations such as in Korean patients (Lee et al., 2011; Lee and Yoo, 2019), Indian patients (Athota et al., 2020), American patients (Bhoj et al., 2017), Italian patients (Sarkozy et al., 2009) and Brazilian patients (Yamamoto et al., 2015). Pathogenic variants in *BRAF (1.7%-2.3%)* and *LZTR1* (~8%) genes are reported as rare in the literature (Sarkozy et al., 2009; Yamamoto et al., 2015) .

In our study both *SOS1* and *RAF1* accounted for 9% of Noonan-causing variants each, which is consistent with the literature for *RAF1* (Pandit et al., 2007) and a bit lower compared to the literature for *SOS1* (Lepri et al., 2011: 1; Roberts et al., 2013). The different detections rate in this study compared to the literature could be related to the limited genetic information from African populations in genetic variants population databases due to the understudy of the African genome and patients of African ancestry (Kruszka et al., 2017; Tekendo-Ngongang et al., 2019).

Pathogenic variants in the *BRAF* gene are mostly associated with CFC, however in the current study variants were observed in a substantial number of NS patients. This could suggest that the patients may need to be reassessed for features of CFC, however similar phenotypic overlap

has previously been observed amongst RASopathy genes (Rauen, 2013). No clinical features of CFC were noted specifically in these patients.

It is reported (Athota et al., 2020) that the four major features in 25% of NS patients are unusual facies (including hypertelorism, DSPF , ptosis and webbed neck), pulmonary valve stenosis (PVS), short stature and mental retardation (intellectual disability). 59.5% (22/37) of NS patients had a putative disease causing mutation identified. The common clinical features observed in the 22 samples with Noonan-causing variants, was that 72.2% had low set ears, (9/16) 56.3% of males had undescended testes, 65% had short stature and heart defects, 85% had DSPF and, hypertelorism was 63.3%. It was shown that PVS was the most frequent heart defect observed in 50% of our cohort followed by 41.6% for VSD. Intellectual disability was observed in 63.6% with patients having mild or moderate symptoms. No clinical feature was observed consistently in all patients, this indicates the clinical overlap or heterogeneity observed in NS patients (Rauen, 2013), it may also be that our sample size was small and there may have been ascertainment bias in our cohort ("Ascertainment bias", 2017; "Ascertainment bias", 2017)

Copy number variants in the NS genes play a minor role in the pathogenesis of NS (Allanson and Roberts, 1993). Three publications identified a duplication encompassing the *PTPN11* gene through array CGH or SNP array, this duplication was identified at the chromosome region 12q24.11q24.23. These were identified in three unrelated patients from USA (Shchelochkov et al., 2008; Graham et al., 2009) and China Chen *et al.*, 2014). All these three patients had clinical features on NS and the *PTPN11* gene duplication may suggest that increased dosage of SHP2 protein may have dysregulating effects on intracellular signaling of the RAS/MAPK pathway (Rehman *et al.*, 2018). A rare deletion of 183kb at 10q25.2 encompassing *SHOC2* gene was also identified in the Chinese patient (Chen *et al.*, 2014). In our study, CNVs for NS causing genes were not screened in patients without NS causing variant. Shchelochkov *et al*., (2008) suggest that 15-30% of the NS patients with no pathogenic variants identified through sequencing should be screened for CNVs. The authors recommend that analysis of copy number changes in the *PTPN11* should be included in the comprehensive testing for NS (Shchelochkov et al., 2008). This high CNV rate in *PTPN11* gene is only reported in this single paper and most NS papers report the low contribution of CNV variants in *PTN11* in patients with NS phenotype (Allanson and Roberts, 1993; Tartaglia *et al.*, 2002, Graham *et al.*, 2009, Bhambhani and Muenke, 2014). This suggest that more CNV studies involving NS genes are needed to clarify this discrepancy.

b) Noonan syndrome with multiple lentigines

The *PTPN11*:c.1381G>A variant was observed in the one patient that had the clinical diagnosis of NSML, FRASC100. Since more than 90% of the NSML pathogenic variants are found in the *PTPN11* gene (Sarkozy et al., 2009), this finding was expected. This patient had typical clinical features associated with NSML (Digilio et al., 2002). No CNVs have been reported to cause NSML, ~100% of the NSML pathogenic variants can be identified by sequence analysis (Gelb and Tartaglia, 1993).

c) Cardio-facio-cutaneous syndrome

Over 75% of *BRAF*, 25% *MAP2K1* and *MAP2K2* and <2% *KRAS* germline mutations are associated with CFC syndrome (Sarkozy et al., 2009). In the current study, three patients had a clinical diagnosis of CFC and a CFC-causing variant was identified in all three, indicating 100% detection rate. Two of the three patients (66.7%) were found in the *BRAF* gene; this detection rate is similar to what is reported in the literature (Pierpont et al., 2014) and one patient had a CFC-causing variant in the *MAP2K2* gene, the reported detection rate is 10-15 % (Rauen, 2013). No pathogenic or likely pathogenic variants were observed in *SHOC2* or *KRAS* genes, these are reported to be rare causes of CFC in the literature (Lee et al., 2011), additionally our CFC sample size was small. All three patients clinically diagnosed with CFC had lentigines, DSPF, deep palmar creases, sparse hair. These are typical clinical features associated with CFC syndrome (Pierpont et al., 2014). DNA testing strategy for CFC variants is approximately 100% utilizing sequencing analysis. There are few reports of CNVs associated with CFC syndrome reported for *BRAF* and *MAP2K2* genes(Yu and Graf, 2011; Nowaczyk et al., 2014). A single report of an American patient with a *BRAF* deletion associated with CFC-like phenotype was reported (Yu and Graf, 2011). Seven Canadian patients with clinical features overlapping with CFC features were reported to have *de novo* deletions of Chr19p13.3 including *MAP2K2* gene (Nowaczyk et al., 2014). For CFC testing, although >99% of variants are detected by sequencing analysis, ~1% may be CNVs such as deletions suggesting that, in addition to activating mutations, *MAP2K2* haploinsufficiency may lead to dysregulation of the RAS/MAPK pathway (Rauen, 1993; Nowaczyk et al., 2014).

d) Schwannomatosis

This patient was first thought to have NF2 when included in the current study, it was after recruitment when the patient's sister had testing overseas and pathogenic variant in the *LZTR1* gene associated with Schwannomatosis was identified in her. Similar phenotypic features such as Schwannomas and hearing difficulties between NF2 and Schwannomatosis have been previously reported in the literature (Evans, 1993; Seppälä et al., 1998; Tamura, 2021). The clinical features of this patient includes a left sided/unilateral acoustic neuroma, which supports clinical diagnosis of Schwannomatosis (Smith et al., 2015). The patient does not have any unusual skin growths but has two hyperpigmented macules over the lumbar region. The patient has three daughters that were tested for the *LZTR1*:c.2269C>T nonsense variant. Two of the daughters, a 27-year-old with tinnitus and unilateral hearing loss and and a 33-year-old daughter who did not show any obvious clinical features of Schwannomatosis both tested positive for the nonsense variant. There is variable expressivity in Schwannomatosis and this explains the two sisters with a variant identified but not presenting similarly (Dhamija et al., 1993).

e) Neurofibromatosis type 1

All seven patients with NF1-causing variant had CALs of >6:>15cm. 71% (5/7) of the patients who had NF1 mutation had neurofibromas, of those 80% (4/5) had plexiform neurofibromas, 85% (6/7) of the patients had axillary freckling with 42.8% (3/7) having both axillary and inguinal freckling. The clinical features are in keeping with the diagnostic criteria of clinical diagnosis of NF1 (Koczkowska *et al.*, 2018). It is reported in the literature that multiple café au lait spots occur in nearly all NF1 patients and freckling in 90% (Friedman, 1993; Anderson and Gutmann, 2015b; Cimino and Gutmann, 2018). The seven patients with NF1-causing variant had a mean age of 11 years, it is reported that in early childhood the plexiform neurofibromas grow rapidly and may cause disfigurement (Friedman, 1993), which may affect the quality of life in these children. Identifying the NF1-causing variants will immediately assist in the management of symptoms in these patients and contribute to the improvement of the quality of life. Neurofibromatosis type 1 has a highly variable expression and familial variation (Anderson and Gutmann, 2015b; Pasmant et al., 2015; Zhang et al., 2015). There are few genotype-phenotype correlations described in the literature but the clinical variability of NF1 arising from genetic or non-genetic factors make the correlations difficult in most cases (Anderson and Gutmann, 2015b; Koczkowska et al., 2018). The whole NF1 gene deletion is reported to be associated with severe clinical phenotypes such as cognitive defects, body or

facial dysmorphisms and early onset of cutaneous neurofibromas (Imbard et al., 2015), the 3bp inframe deletion of exon 17 (c.2970-2972 delAAT) in the NF1 gene associated with a milder phenotype without the neurofibromas (Upadhyaya et al., 2007).

In our context, a study by C.Smal (a sub-study under current study, supervised by current student) analysed 13 patients (seven were from current study NF1 cohort) with no SNVs observed using targeted panel utilising MLPA to screen for CNVs. Three of the 13 patients (23%) had deletions observed; of these two were whole gene deletions encompassing flanking genes. These two patients had features that suggest a severe NF1 phenotype, these were multiple subcutaneous neurofibromas, Iris lisch nodules, intellectual disability and speech delay, dysmorphic features, macrocephaly and skeletal abnormalities, these were observed over and above the typical NF1 features such as multiple café au lait macules, axillary freckling (MSc research, Unpublished data).

As I conclude the discussion on genetic profiles obtained by FRASC targeted panel as well as the detection rate per clinical diagnosis, most of the trends observed are similar to those reported in the literature. The detection rate obtained in our African cohort was similar to global yields. The only highlight to mention are the yields of *LZTR1* and *BRAF* genes variants that are reported as rare for NS patients had higher yields in our study. The FRASC multi-disease panel is appropriate cost-effective diagnostic tool and can be utilised alone for all RASopathies, however for NF1 where CNVs are reported to contribute 5-10% to the mutation load (Imbard et al., 2015; Kehrer-Sawatzki and Cooper, 2021), MLPA could be valuable secondary mutation screening tool. Microarray (Array CGH) cannot detect single exon deletions/duplications, intragenic CNVs or CNVs that are smaller than 200kb reliably. MLPA can correctly detect copy number changes as small as 50bp enabling identification of single exon deletions. Additionally MLPA is a cheaper alternative, costing approximately R2100 per analysis compared to microarray which costs R~8000. With regard to CNV analysis using WES, the CNV analysis pipeline using exome data is not yet well established in our setting, but would be a viable alternative worth pursuing. For these reasons, MLPA was employed as a secondary screening tool in the current study and not microarray or WES. Our laboratory could also benefit from optimizing RNA sequencing analysis and incorporate it in the NF1 testing strategy as the splicing variants also account for substantial number (~20%) associated with NF1 and are not often detected by genomic DNA sequencing, thus for these kind of variants RNA analysis may be appropriate (Pasmant et al., 2015; Evans et al., 2016). Overall detection rate

per clinical diagnosis in the current study were comparable to the literature, further indicating the potential of the panel to confirm or exclude the clinical diagnosis.

4.3.12. Segregation analysis of RASopathy-causing variants

It is reported in the literature that most inherited RASopathy-causing variants are inherited from a mother, as most of these variants except for NF1 variants have an adverse impact on male fertility (Jafry and Sidbury, 2020). In the current study, 39 parents were tested in during segregation analysis and 7 (17.9%) were found to carry putative disease causing variants, similar to those observed in the probands (children). Of those, five were mothers (71.4%) and two were fathers (28.5%), indicating that in the variants did not have adverse impact on their fertility. Four variants were found to be *de novo* (10%) after parents of FRASC45 (*NF1*:c.3721C>T), FRASC100 (*PTPN11*:c.1381G>A), FRASC4 and FRASC64 both carrying the *BRAF*:c.770A>G were sequenced using Sanger sequencing and were found to not carry the disease-causing variant identified in their children. The literature reports that ~50% of the variants in RASopathies are inherited *de novo* (Rauen, 2013; Tidyman and Rauen, 2016; Jafry and Sidbury, 2020), the low *de novo* rate observed in our study is related to the lack of availability of the second parent's sample due to unavailability of the parent during recruitments.

The mothers of FRASC2 and FRASC20 were also suspected of having NS during the recruitment clinics. Thus, when the Noonan-causing variant was identified in both of them, it was an expected outcome. In the case of FRASC62 father and two daughters were found to have the *LZTR1*:c.2269C>T variant, this indicate that this Schwannnomatosis-causing variant segregates within this family and it is the cause of their disorder. However, for the rest of the parents whom disease-causing variants were identified, no clinical assessment was done so it could not be ascertained if they were clinically presenting with the disease or not. It is reported that many affected adults are diagnosed only after the birth of an affected child in most RASopathies and for NS patients, mild expression is likely to be overlooked (Roberts et al., 2013; Jafry and Sidbury, 2020).

4.4. Whole exome sequencing

It is estimated that ~85% of clinically significant variants known to date can be identified by exome analysis (Chrystoja and Diamandis, 2014; Guo et al., 2021). Although the exome contains only 1% of the entire genome, it carries most clinically significant variants as it codes for all proteins critical for life (Ross, Dion and Rouleau, 2020). Whole exome sequencing is often utilised for its ability to identify clinically significant variants that could not be found using targeted gene panels (Xu et al., 2017). This technique has some limitations in detecting pathogenic variants in non-coding regions of the genome, some copy number variations as well as repeat expansions. In such cases WGS, microarray and other technologies may be utilised to further screen for variants in rare/undiagnosed diseases (Manickam et al., 2021).

In the current study WES was utilised in a subset of patients (n=7) where no clinically significant variants were identified using the FRASC targeted gene panel.

4.4.1. Detection rate achieved by whole exome sequencing

One of the seven samples analysed had a validated clinically significant variant (FRASC80), this lead to detection rate of 14%, however this is a small sample size to make significant conclusions. This indicate that WES as second line testing in a well phenotyped group of RASopathy patients would add little benefit once the targeted panel is implemented in a diagnostic setting. There are several studies that reported a better detection rate using NGS targeted gene panels as a primary testing approach over WES in several disorders (Ankala et al., 2015; Krenn et al., 2020; Platt et al., 2021). The study by Platt *et al* (2021) had a detection rate of 56% for NGS targeted panel and 45% was obtained for WES, both panel and WES had an overall yield of 58% in this study (Platt et al., 2021). Targeted panel had a detection rate of 46% compared to 18% of WES in Ankala *et al* (2015), the authors recommended that targeted panels should be used as a first tier testing based on detection rate, good coverage and cost effectiveness (Ankala et al., 2015). Another study reported a detection rate of 34.7% when targeted panel was used and 30.6% when WES was used to sequence patients with neuromuscular disorders (Krenn et al., 2020)

The variants identified using WES in the current study were not in any of the known RASopathy genes, indicating the phenotypic overlap of RASopathies with other disorders. The identification of one more pathogenic variant on WES improved detection rate from (34/60) 56.6% when targeted panel only was used to (35/60) 58.3% detection rate overall for targeted panel and WES utilized in the current study. This means that 58.3% of the patients enrolled in this current study have a clinical diagnosis and this contribute to clinical management they get

towards their disorder as well as informed genetic counselling sessions for patients and their families. The one patient (FRASC80) with a pathogenic variant in the *CDK13* gene is discussed below.

Patient FRASC80

This patient presented to our research clinic with clinical features suggestive of NS was found to have the *CDK13*-related dysmorphic feature and intellectual development disorder. The heterozygous *CDK13*:c.2149G>A missense pathogenic variant, associated with autosomal dominant dysmorphic facial features and intellectual development disorder with or without congenital heart defects (Hamilton et al., 2018: 13) was detected in this patient. This finding highlights other syndromes that may overlap with NS features; currently there is no literature that associate this disorder with NS.

There are less than 50 cases published in the literature that have *CDK13*-related dysmorphic feature and intellectual development disorder (Bostwick, 2019; Bostwick *et al.*, 2017). This syndrome was recently described with first publication in 2016 and *CDK13* variants were identified in patients enrolled in the DDD-UK study (Sifrim et al., 2016). Subsequent studies where the *CDK13* variants was identified were undertaken in 9 patients from USA (Bostwick et al., 2017) , 16 patients from the UK (Hamilton et al., 2018), seven patients from Japan (Uehara et al., 2018) and 15 patients from the Netherlands (van den Akker et al., 2018). Our variant is the same as the one identified in ten patients from these studies, the *CDK13*:c.2149G>A (p.Gly717Arg) has 10 submissions on ClinVar (VCV000375737.8) and so far has been reported as *de novo* (Bostwick et al., 2017). Most variants are identified in the exon 4 of the *CDK13* gene in the protein kinase domain and this serves as a potential hot spot region this disorder, the protein kinase domain is responsible for mRNA processing, mRNA splicing (https://www.uniprot.org/uniprot/Q14004). In our study the three year old male patient was found to have this variant, this is the first *CDK13* variant identified in an African patient, a study such as this help to understand the spectrum of disease in Africa which is poorly characterised. . The variant had full contribution to the patient's phenotype because most of the clinical features observed in this patient such as dysmorphic features, intellectual disability and heart defects. This finding highlight the usefulness of WES in this patient, without which the clinical management of the features was not going to be appropriate.

4.4.2. Variants of uncertain significance in whole exome sequencing data

Sequencing using WES is reported to generate more VUSs than targeted sequencing due to the large amount of data produced (Ari and Arikan, 2016), in the context of our setting this higher VUS rate is exacerbated by lack of African ancestry population genetic data. The lack of African specific genetic data in public databases complicates NGS-based testing and data analysis in non-European populations (Horak, Fröhling and Glimm, 2016; Pottinger Tess D. et al., 2020). This begs the need for researchers focusing on African genomes to collaborate and share their genetic data, so that the uncertainty can be resolved (Krause, 2019; Clift et al., 2020). In the current study, 6 out of seven (85.7%) patients where WES was performed (FRASC24, FRASC55, FRASC56, FRASC58, FRASC96 and FRASC97) had no clinically significant variants identified, the significance of the variants identified for these patients is unknown and they had limited contribution to patients' disorder when analysed against patient's phenotypic features noted down during recruitments.

To minimize the number of VUSs identified using WES, a study by Vears and colleagues recommend that parental samples be available so that the significance of these variants is understood during analysis (Vears et al., 2018). Various professional guideline bodies emphasize the importance of detailed clinical data is also reported as crucial aspect into understanding the significance of the VUS, the recommendations further state that no WES should be undertaken until detailed clinical data is provided. There is still no consensus with regard to reporting VUS from various guidelines organizations, ACMG does not support reporting of VUSs (Rehm et al., 2013), CCMG recommends analysis should be limited to disease-causing variants (likely pathogenic and pathogenic) only (Hume et al., 2019). Furthermore, ESHG does not address whether a variant should be reported (van El et al., 2013) and the EuroGentest recommends laboratories should report these at their discretion (Matthijs et al., 2016). Most diagnostic laboratories limit their reporting to those VUSs favouring pathogenic, these are those that may be upgraded to likely pathogenic by applying one more ACMG code and only if the VUS is related to the phenotype. In our context, all VUSs identified in the six patients were not directly related to the clinical diagnosis of our patients which is NS, there was limited information to correlate and we did not have access to histories and other investigations. Vears *et al* (2018) panelists have agreed that VUS where a gene is known to be relevant to the clinical question but lacking evidence of pathogenicity should be reported. The panel also discussed that single VUS in a recessive condition should be reported if it is well covered during sequencing and matches the phenotype and there might be the possibility of

second variant being identified using another method. This panel agreed with ACMG that VUSs in genes not related to the phenotype should not be reported. In our context these variants will be documented for future reanalysis using CNV pipelines which would clarify if low detection rate was due to the variants being deletions or duplications on not. Also having both parental samples could be useful during analysis to ascertain if a variant is *de novo*. The analysis pipelines employed in the current study such as moon diploid relies heavily of the detailed clinical information provided for the patient. The limited clinical information in our six patients could have contributed to the identification of VUSs. The other important aspect is loading these variants and their classifications to disease variant database such as ClinVar so that it is possible to reclassify as likely benign or likely pathogenic and eventually the number of VUSs will be reduced. In the context of African genomic data contributing to increase in VUSs, I believe having guideline body for African specific variant interpretation might contribute to the understanding and interpreting of VUSs in the African context.

Further analysis using a different testing method, utilising a different analysis pipeline, or revisiting shortlisted variants as well as refining clinical phenotypes may improve the detection rate observed in these patients. It is reported in the literature that reanalysis of whole exome data with new analysis approaches has shown to improve the detection rate of various genetic disorders (Bergant et al., 2018; Liu et al., 2021). Not all reanalysis results in the improved detection rate, reanalysing with new genomic data may also downgrade variants into likely benign as observed in the study where 82% of the hereditary cancer variants were downgraded to benign or likely benign (Esterling et al., 2020). In our context since most of these variants are not related to the clinical disorder they may be downgraded to likely benign, this will contribute to the reduction of VUSs.

4.5.Targeted gene panel vs Whole exome sequencing

Utilizing NGS techniques such as NGS panel or WES depends on the ability of each technique to have an accurate disease variant detection and to confirm or exclude the diagnosis. In LMICS the cost effectiveness of the technique is paramount. Targeted gene panels are often preferred in LMICS and have consequently become a standard first tier testing methodology for most monogenic disorders

phenotype of the patient. Sequencing using WES is much more expensive than the panel testing and the analysis is not as targeted, this is mostly suitable for disorders such as developmental delay and epilepsy phenotypes where the genetic cause is not directly known (Schwarze et al., 2018; Eisele and Kappelmann-Fenzl, 2021).

In our context, the targeted panel was utilised to sequence 60 RASopathy patients (current study), 14 Cornelia de Lange patients (Seymour, 2019) and 14 Facial dysostosis patients (Nevondwe, 2019). The diagnostic utility of this panel was demonstrated through these studies and the 56.6% detection rate achieved for the current study makes this evident. In our facility, targeted panel will be used as a first tier testing methodology as the affordability and utility was demonstrated through the current study.

The value of using WES as a secondary variant screening was also demonstrated in the current study, one of the seven patients had a pathogenic variant identified following negative results on a targeted panel. When clinically indicated through detailed phenotyping WES can be employed as a secondary screening tool. However, in our context, this can be done in a research setting only currently, as the cost of WES and other complexities such as bioinformatics expertise may not be feasible in a diagnostic setting.

4.6. Diagnostic utility of findings in the current study

Clinically significant variants were identified in 58.3% (utilizing both a targeted panel and WES) of the patients who had no molecular testing prior to this study, the identification of a causative variant immediately contributes to the clinical utility of that result to the patient. These variants were reported back to the patients, this lead to improved health care to the patient with regard to the directed management of the predicting symptoms, reproductive choices and valuable information to relatives and family members of the patient.

When the diagnosis of a RASopathy is confirmed, the patient needs constant follow-up through life for regular examination including; ophthalmologic and hearing evaluation, surveillance of malignancies or heart defects (Joseph et al., 2016). Although patients with RASopathies are associated with mild increase risk of cancer, surveillance for possible malignancies may be necessary particularly when suggestive clinical symptoms are present (Villani et al., 2017). Although ~50% of RASopathy-causing variants arise *de novo*, if a pathogenic variant is identified in the child, testing of parents is recommended to predict the risk of transmission to the child (Carcavilla et al., 2020).

Although the utility of the confirmed diagnosis benefits the patient more, for the health care provider, making a diagnosis, having an effective standard of care and recommending the effective course of treatment is the ultimate goal that will benefit all patients presenting with similar features. Identifying the importance of all the utilities involved in a molecular testing in support of patient health care is essential for informing the genomic knowledge needed in precision medicine (Farkas, 2016; Joseph et al., 2016).

4.7. Limitations of current study

The RASopathy clinical tick-sheet was designed for the purposes of recruiting patients in the current study. The limitations of this tick-sheets includes

- Limited clinical data related to the disorder that was suspected.
- Incomplete data leading to inability for us to distinguish whether the unticked features was absent or just not recorded
- No place for additional information or specific information

This limitation was observed during the assessment of the patients where clinically significant variants were not identified. The assessment of clinical details was important for the selection of patients into the WES run. The clinical data tick-sheets were not originally designed to do clinical phenotyping, but were designed to document positive clinical features of RASopathy, by ticking if a feature was present but did not have alternatives. The 'Not present vs not examined' options could not be distinguished, also there was no space to record additional features on the tick sheet. For future recruitments, the tick-sheets should include additional space for extra notes and clinicians should be trained on how to fill in the form in detail.

The 40% (24) of patients with no mutations identified using the custom designed targeted panel utilised in this study, indicate that there are other type of variants causing RASopathies or RASopathy-like phenotypes other than SNVs and small indels. The FRASC targeted panel only screened for variants in the coding regions of the genes included in the design; copy number variations, variants in the non-coding regions and coding regions of the genes that were not included in the panel could be the cause of the RASopathy-like phenotype observed in those patients.

The other limitation was not incorporating RNA sequencing, particularly for NF1 patients as it is reported in the literature that 20-30% of NF1 variants affect splicing, and can be identified using RNA sequencing, and could possibly be missed by DNA sequencing.

4.8.Recommendations and Future directions

Future work utilizing WES, WGS, MLPA and Microarray will contribute to the uncovering of novel gene variants associated with RASopathies or RASopathy-like phenotypes. Sequencing more exomes in a larger cohort will also add more value in understanding the overlapping developmental phenotypes to RASopathies.

In addition to RASopathies, somatic variants in the RAS/MAPK genes are reported to be associated with cancer. Studying the regulation of this pathway may benefit not only RASopathy patients but also cancer patients. Our findings may contribute to the National Cancer Institute initiative on Advancing RASopathy Therapies (ART) aimed at developing effective therapies and prevention strategies for the patients with RASopathies and tumors (Gross et al., 2020). Finding a causative variant in these patients may contribute to personalised therapy/treatment strategies as it explains the phenotype in a patient, and can guide specific disease management based on the established diagnosis. In the future it may guide pharmacologic interventions targeted to a specific developmental pathway. The patient's predicted responses and effectiveness of therapy could be gene and possible variant dependent.

The findings of this research add valuable information to the body of knowledge in RASopathies. Currently existing information in the literature contains limited genomic data from RASopathy patients of African ancestry. This current study served as a stepping-stone for subsequent NGS studies using a targeted gene panel and broader testing.

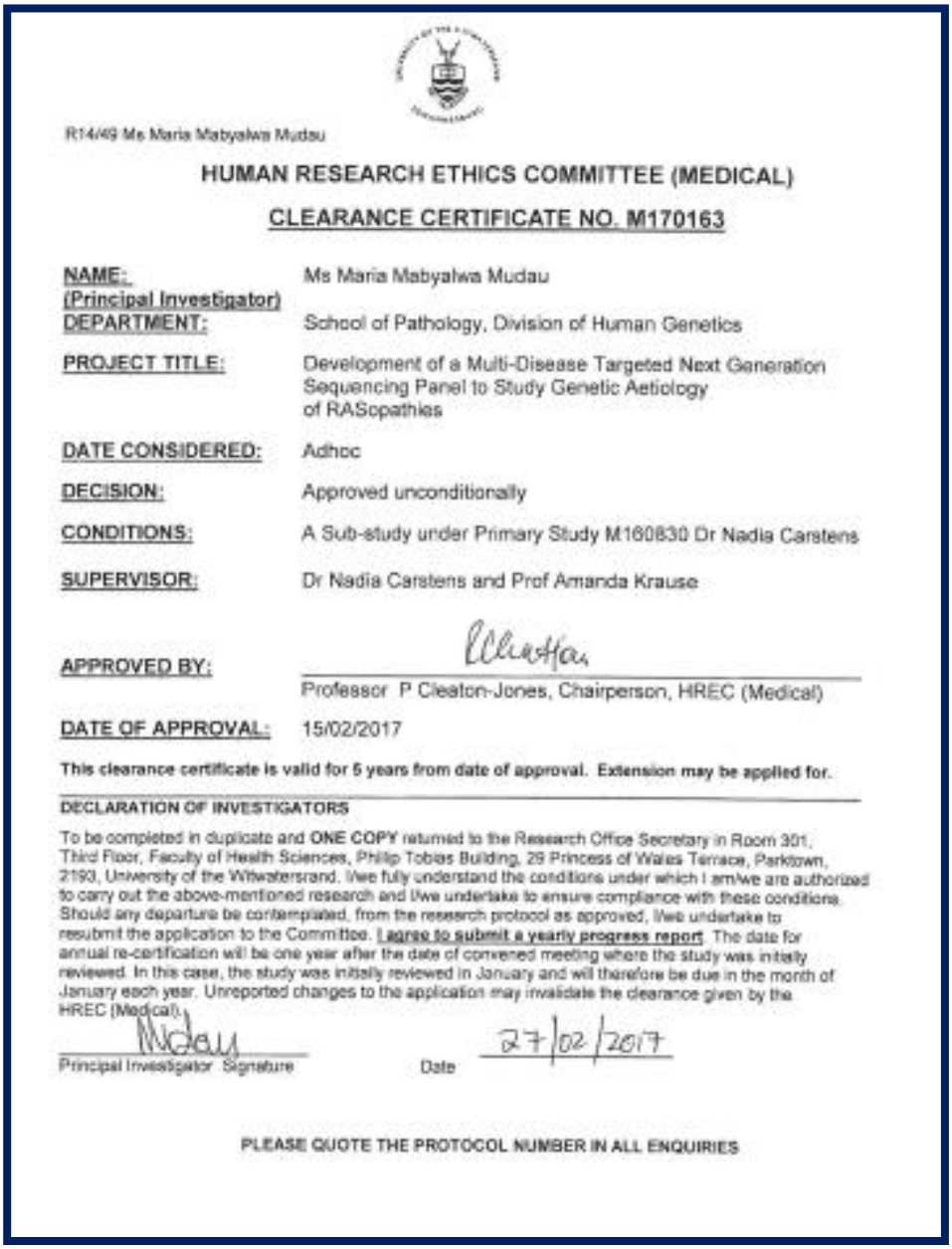
The current study was the first in our Division to custom design a local panel and initiate studies aimed at determining molecular profiles through NGS, leading to improved diagnostic testing.The NGS panel in the current study has the potential to improve not only the detection rate for RASopathies but for many other monogenic disorders through the inherited disease targeted panel (IDP) that was designed following the success of this current project. The Inherited disease panel is a targeted disease gene panel containing 500 genes associated with various monogenic disorders that was designed in the Division of Human Genetics, University of the Witwatersrand and the National Health Laboratory Service, Johannesburg. The FRASC panel employed in the current study was a proof of principle for the IDP panel. The future involves the implementation of the IDP panel and integrating it in the diagnostic setting, the initial validations of this panel have commenced.

4.9. Conclusion

Next generation sequencing has revolutionized how testing for monogenic disorders is undertaken globally and through the current study this technology has shown its potential to improve diagnosis in the South African State healthcare system. This study demonstrated that the approach of combining clinically distinct diseases on one sequencing panel, followed by selective analysis, is a feasible strategy and a cost-effective diagnostic tool for rare genetically heterogeneous diseases. The genetic aetiology of South African patients with RASopathies was studied in the current study; this is the first study looking at this group of disorders as a whole, rather than studying one RASopathy at a time. The molecular variant data obtained was compared to the existing RASopathy literature that is mostly not African. The molecular genetic profiles obtained on RASopathy patients were mostly similar to the published data except that for NS patients, the *BRAF* and *LZTR1* genes have been shown to be the second and third contributors to pathogenic variants in NS after *PTPN11*. This finding is different from the published European data and indicates the unique genetic profiles that South African patients may have. Utilizing WES in the current study has demonstrated that analysis of the data relies heavily on a detailed patient phenotype. All disease features observed during clinical examination are critical to reach a molecular diagnosis using WES data. Disease causing variants were identified in genes such as *CDK13* which are not considered RASopathy genes. This further indicates that there are many syndromes that have clinical features overlapping with RASopathy phenotypes. Only one more variant was detected with WES, this improved detection rate from 56.6% when targeted panel only was used to 58.3% detection rate overall for targeted panel and WES utilized in the current study. This means that 58.3% of the patients enrolled in this current study have a clinical diagnosis and this contribute to clinical management as disease specific guidelines can be followed as well as defining recurrence risk and disease risk to other family members. Developed countries are using WES as first-line testing for most monogenic disorders to reduce the number of diagnostic tests in their various laboratories (Yang et al., 2013; Posey et al., 2016; Arts et al., 2019; Waldrop et al., 2019). However, the recommendation of the current study is that; the targeted gene panels should be first tier test for patients with RASopathies. This is because of the detailed clinical phenotyping required before WES testing in addition to resources required to analyse the amount of data generated, bioinformatics complexities as well as costs associated with WES. Our strategy and recommendations would be useful in a laboratory with limited resources providing genetic diagnostic testing to the South African public health care system. Similar value would be expected in other limited resource countries.

4.10.Appendices

Appendix 1.1: Ethics clearance obtained by Mudau MM (Current PhD) from the University of the Witwatersrand, Medical Human Research Ethics Committee



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

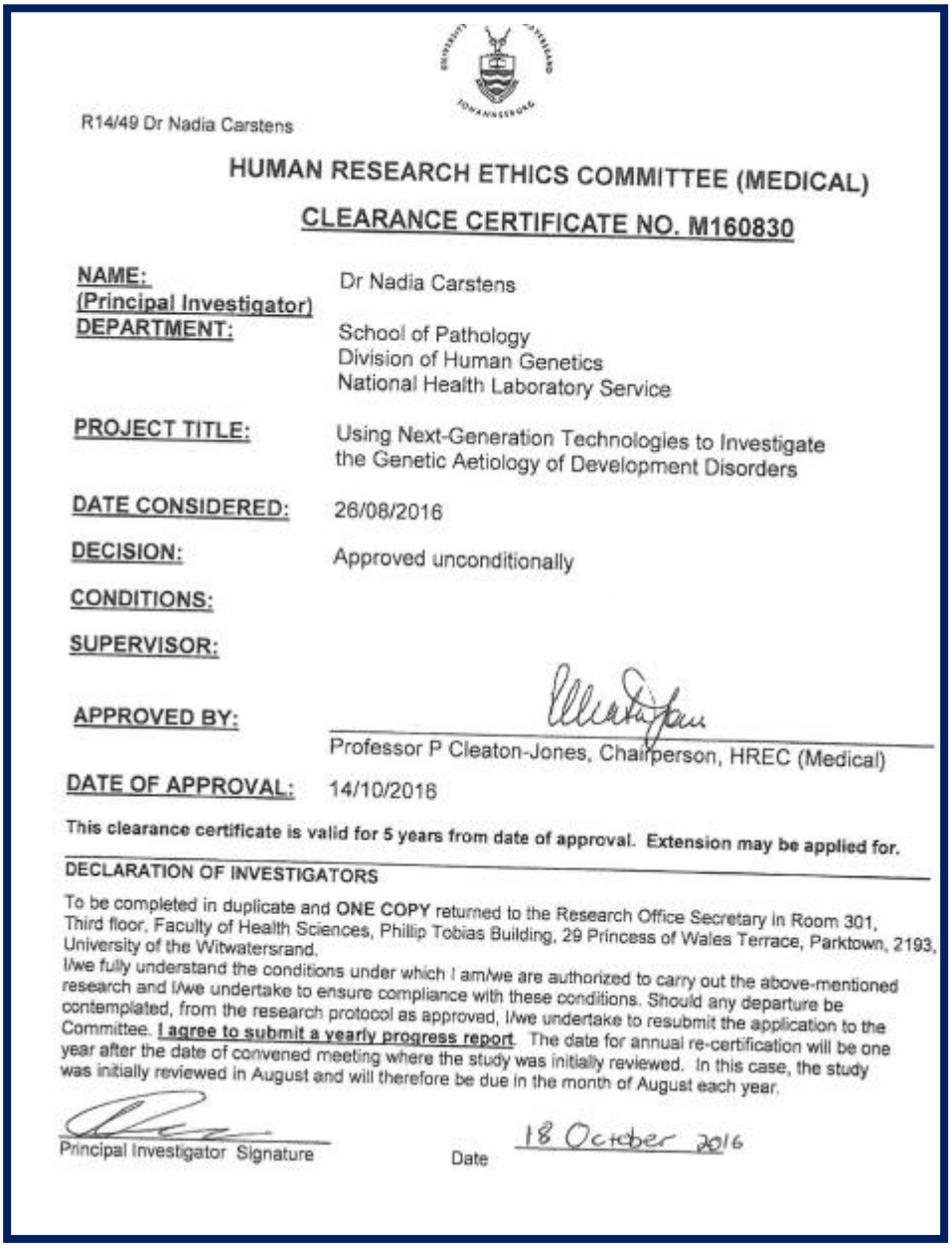
To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we 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 1.2: Ethics clearance obtained by Dr Nadia Carstens (Current PhD, Supervisor) from the University of the Witwatersrand, Medical Human Research Ethics Committee



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

Appendix 1.3: Ethics clearance obtained by Dr Nadia Carstens (Current PhD, Supervisor) from the University of Pretoria, Faculty of Health Sciences, Research Ethics Committee

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 2: Clinical RASopathy data 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	☐

Rasopathy Clinical Ticksheet

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

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: Full list of demographics of RASopathy patients recruited in the current study

FRASC No.	Sex	Age at recruitment	Ethnicity	Clinical diagnosis	Hospital	Family samples available	
FRASC1	F	33,4	B	NS	CMJAH	0	None
FRASC2	M	2,9	B	NS	CHBAH	1	Mother
FRASC4	M	5,2	B	NS	CMJAH	2	Mother and Father
FRASC5	M	4,9	B	NS	CHBAH	1	Mother
FRASC10	F	2,3	B	NS	CHBAH	1	Mother
FRASC14	F	7,7	W	NS	CHBAH	1	Father
FRASC15	M	4,5	B	NS	CHBAH	1	Mother
FRASC17	M	9,4	B	NS	CHBAH	1	Mother
FRASC18	M	7,3	B	NS	CHBAH	1	Mother
FRASC19	M	7,2	B	NS	CHBAH	1	Mother
FRASC20	F	3,9	B	NS	CHBAH	1	Mother*
FRASC21	M	8	B	NS	CHBAH	1	Mother
FRASC24	M	7,5	B	NS	CHBAH	2	Mother and Father
FRASC36	F	1,6	B	NS	CHBAH	1	Mother
FRASC37	F	4,8	B	NS	CHBAH	1	Mother
FRASC39	M	2	B	NS	CHBAH	1	Mother
FRASC55	M	2,4	B	NS	CHBAH	2	Mother and Granny
FRASC56	M	1,5	B	NS	CHBAH	1	Mother
FRASC60	M	4,6	B	NS	CHBAH	1	Mother
FRASC64	M	6,9	B	NS	SBAH	2	Mother and Father
FRASC66	M	4,6	B	NS	SBAH	0	None
FRASC69	M	4,4	B	NS	SBAH	1	Mother
FRASC71	M	3,2	B	NS	SBAH	1	Mother
FRASC73	F	8,7	B	NS	SBAH, PTA	1	Mother
FRASC80	M	3,2	B	NS	CMJAH	1	Mother
FRASC81	F	2,5	B	NS	CMJAH	2	Mother and Father
FRASC84	M	10,5	W	NS	SBAH, PTA	2	Mother and Father
FRASC87	M	11,4	B	NS	SBAH	1	Mother
FRASC88	M	0,5	B	NS	SBAH	1	Mother
FRASC90	F	12,8	B	NS	SBAH	1	Mother
FRASC91	M	5,5	B	NS	SBAH	2	Mother and Father
FRASC92	M	6,5	B	NS	SBAH	2	Mother and Father
FRASC94	F	0,6	B	NS	CMJAH	1	Mother
FRASC96	F	0,7	B	NS	RMH	2	Mother and Father
FRASC97	M	0,8	B	NS	CMJAH	2	Mother and Father
FRASC98	F	4,1	B	NS	RMH	2	Mother and Father
FRASC22	F	9,3	B	CFC	CHBAH	1	Mother
FRASC58	F	10,2	B	?RASopathy	NHLS	1	Mother
FRASC62	M	58	I	Schwannomatosis	NHLS	3	Daughters

FRASC63	F	13,8	B	CFC	RMH	1	Mother
FRASC70	M	6,6	B	NS/NF1	SBAH	1	Mother
FRASC85	M	15,7	W	?RASopathy	SBAH	1	Father
FRASC99	M	4,7	B	CFC	CHBAH	1	Mother
FRASC100	M	8,4	B	NSML	SBAH	2	Mother and Father
FRASC3	M	13,9	B	NF1	CHBAH	1	Mother
FRASC38	F	3,3	B	NF1	CHBAH	1	Mother
FRASC42	F	12,3	B	NF1	CHBAH	1	Mother
FRASC43	M	34	B	NF1	CMJAH	0	None
FRASC44	M	6,2	B	NF1	CMJAH	1	Mother
FRASC45	F	8,9	B	NF1	CMJAH	2	Mother and Father
FRASC46	M	29,9	B	NF1	CMJAH	0	None
FRASC53	M	11,8	B	NF1	RMH	1	Father
FRASC65	F	7,3	B	NF1	SBAH	1	Mother
FRASC67	M	4,9	B	NF1	SBAH	1	Mother
FRASC68	M	2,2	B	NF1	SBAH	1	Mother
FRASC82	M	27,2	B	NF1	CMJAH	0	None
FRASC83	M	18,4	B	NF1	CMJAH	1	Mother
FRASC89	F	6,2	B	NF1	SBAH	1	Mother
FRASC95	F	28,9	B	NF1	CHBAH	0	None

M=Male, F=Female, B=African ancestry, W=Caucasian, CHBAH=Chris Hani Baragwanath Hospital, CMJAH, Charlotte Maxeke Academic Hospital, RMH= Rahima Moosa Hospital, SBAH=Steve Biko Academic Hospital

Appendix 4: FRASC panel design report from Agilent SureDesign software

FRASC-Panel design summary

File Type: SureSelect DNA Design Report
Created By: SureSelect DNA Advanced Design Wizard
User: maria.mudau@nhls.ac.za
Workgroup: Wits Human Genetics
Folder: FRASC Final
Timestamp: 08-Aug-2018

Design Summary
Design Name: FRASC and NF2 final
Species: H. sapiens (UCSC hg19, GRCh37, February 2009)

Probegroup Summary
Number of Probegroups: **3**
Probegroup 1 : FRASC_problematic
Probegroup 2 : FRASC_ Main_Final
Probegroup 3 : NF2

Probe Summary
Total Probes: 3375
Total Probes Size: 245.950 kbp
Recommended Minimum Sequencing per Sample: **49.190 Mbp**

Target Parameters
Databases: RefSeq
Region: Coding Exons
Region Extension: 10 bases from 3' end and 10 bases from 5' end.
Allow Synonyms: No

Probe Tiling Parameters
Tiling density: 2x

Boosting: Balanced
Extension into repeats: 20
Strand: Sense

Target and Probe Details
Target ID: The identifier entered in the Targets list.
Interval: The genomic interval of the target.
Regions: The number of regions within this target.
Size: The total size (in base pairs) of the regions.
Database(s): The databases in which this target was found.

High Coverage: Number of regions with overlap >= 90%.

Low Coverage: Number of regions with overlap < 90%.

Target ID and Refseq transcript	Interval	Regions	Size	Coverage	High Coverage	Low Coverage
A2ML1 (NM_144670.6)	chr12:8975238-9027617	36	5088	100	36	0
*AFF4 (*NM_014423.4)	chr5:132216741-132272891	20	3892	100	20	0
ANKRD11 (NM_013275.6)	chr16:89334876-89383437	11	8212	100	11	0
ARID1B (NM_001374828.1)	chr6:157099054-157529035	20	7150	100	20	0
BRAF (NM_004333.6)	chr7:140434387-140624513	18	2661	100	18	0
CBL (NM_005188)	chr11:119077118-119170501	16	3041	100	16	0
CHD7 (NM_017780.4)	chr8:61653982-61778502	37	9734	100	37	0
DHODH (NM_001361.5)	chr16:72042654-72058108	9	1368	100	9	0
EFTUD2 (NM_004247.4)	chr17:42928632-42971899	27	3459	100	27	0
ESCO2 (NM_001017420.3)	chr8:27633022-27660965	10	2006	100	10	0
EYA1 (NM_000503.6)	chr8:72111565-72268731	17	2144	100	17	0
HDAC8 (NM_018486.3)	chrX:71549894-71792621	13	1472	100	13	0
HRAS (NM_005343.4)	chr11:532626-534332	5	752	100	5	0
KMT2A (NM_001197104.2)	chr11:118307218-118392897	36	12639	100	36	0
KRAS (NM_033360.4)	chr12:25362719-25398328	5	787	100	5	0
LZTR1 (NM_006767.4)	chr22:21336651-21351647	21	2943	100	21	0
MAP2K1 (NM_002755.4)	chr15:66679676-66782963	11	1402	100	11	0
MAP2K2 (NM_030662.4)	chr19:4090586-4123882	11	1423	100	11	0
MAU2 (NM_015329.4)	chr19:19431659-19466601	19	2222	100	19	0
NF1 (NM_001042492.3)	chr17:29422318-29701183	58	9741	100	58	0
NF2 (NM_000268.4)	chr22:29999978-30090801	17	2164	100	17	0
NIPBL (NM_133433.4)	chr5:36953789-37065004	46	9380	100	46	0
NRAS (NM_002524.5)	chr1:115251146-115258791	4	650	100	4	0
*PDS5A (*NM_001100399.2)	chr4:39827095-39978207	32	4687	100	32	0
PDS5B (NM_015032.4)	chr13:33222900-33349200	34	5024	100	34	0
POLR1C (NM_203290.4)	chr6:43484838-43497103	10	1348	100	10	0
*POLR1D (*NM_152705.3)	chr13:28196111-28240100	4	825	100	4	0

PTPN11 (NM_002834.5)	chr12:112856906-112942578	15	2086	100	15	0
RAD21 (NM_006265.3)	chr8:117859729-117878978	13	2156	100	13	0
RAF1 (NM_001354689.3)	chr3:12626003-12660230	16	2267	100	16	0
RASA1 (NM_002890.3)	chr5:86564259-86686710	26	3672	100	26	0
RASA2 (NM_001303246.3)	chr3:141205916-141331165	25	3065	100	25	0
RIT1 (NM_001256821.2)	chr1:155870169-155880686	6	831	100	6	0
RRAS (NM_006270.5)	chr19:50138823-50143365	6	777	100	6	0
SETD5 (NM_001349451.2)	chr3:9470613-9517785	22	4808	100	22	0
SF3B4 (NM_005850.5)	chr1:149895424-149899661	6	1395	100	6	0
*SHOC2 (*NM_007373.4)	chr10:112724107-112771586	8	1909	100	8	0
SMAD4 (NM_005359.6)	chr18:48573407-48604847	11	1879	100	11	0
SMAD5 (NM_005903.7)	chr5:135489440-135513178	6	1517	100	6	0
SMAD6 (NM_005585.5)	chr15:66995587-67073883	4	1571	100	4	0
SMAD7 (NM_005904.4)	chr18:46447732-46476804	4	1464	100	4	0
SMARCB1 (NM_003073.5)	chr22:24129347-24176377	9	1392	100	9	0
SMC1A (NM_006306.4)	chrX:53407014-53449559	26	4265	100	26	0
SMC3 (NM_005445.4)	chr10:112327565-112364070	29	4234	100	29	0
SOS1 (NM_005633.4)	chr2:39212955-39347573	23	4462	100	23	0
SOS2 (NM_006939.4)	chr14:50585052-50698011	23	4459	100	23	0
SPRED1 (NM_152594.3)	chr15:38545377-38643875	7	1475	100	7	0
STAG1 (NM_005862.3)	chr3:136057086-136349750	33	4437	100	33	0
TAF1 (NM_004606.5)	chrX:70586155-70683906	38	6448	100	38	0
TAF6 (NM_005641.4)	chr7:99704859-99717379	15	2445	100	15	0
TCOF1 (NM_001371623.1)	chr5:149737300-149778621	27	5028	100	27	0

Appendix 5: Patient quality control report example with duplicate information from Agilent SureCall software



SAMPLE ATTRIBUTES

Sample name: FRASC4_S5_L001_R1_001.fastq
Date: 15 Mar 2019
Genome Build: hg19
Design Name: FRASC and NF2 final
Design ID: 3161001
Analysis Method: Default SureSelect Method

QC METRICS

Metrics For Input File	Metrics Value
Number of reads in untrimmed fastq file	665,158
Number of reads in trimmed fastq file	665,156
Number of reads trimmed	140,148
Number of untrimmed reads in fastq file after trimming	525,008
Number of reads discarded	2
Number of Duplicate reads	118,949 (17.86%)

Overall metrics for all reads in BAM file	Metrics Value
Number of reads in bam file	518,555
Number of reads in bam file passing mapping quality filters *	511,418
Number of reads in covered regions	385,422 (75.36%)
Percentage of covered regions with zero coverage	0.00
Median QV bases used in variant calling	39
Average Read Length	147
Total high quality mapped bases	75,193,413

* SureCall uses reads that passed the mapping quality filter for the denominator in the calculations that follow.

For Research Use Only. Not for use in diagnostic procedures.

Agilent Technologies

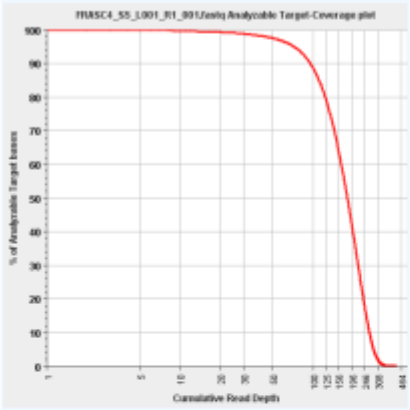
Overall metrics related to analyzable target regions	Metrics Value
Number of genomic bases in analyzable target region	245,950
Number of sequenced bases in analyzable target regions	44,699,563
Number of reads	385,422 (75.36%)
Average read depth in analyzable target regions	181
Median read depth in analyzable target regions	182

Percentage of analyzable target bases with	Metrics Value
at least 1 read	100.00
at least 5 reads	99.95
at least 10 reads	99.72
at least 20 reads	99.33
at least 50 reads	97.61
at least 100 reads	89.03
at least 200 reads	40.64
at least 500 reads	0.00
at least 1000 reads	0.00
at least 5000 reads	0.00
at least 10000 reads	0.00

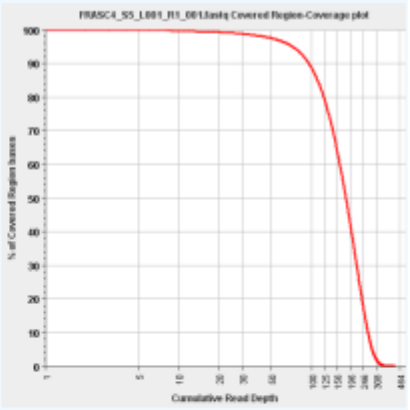
Percentage of analyzable target regions covered by	Metrics Value
at least 1 read	100.00
at least 5 reads	100.00
at least 10 reads	100.00
at least 20 reads	99.89
at least 50 reads	99.36
at least 100 reads	98.09
at least 200 reads	91.18
at least 500 reads	19.55
at least 1000 reads	2.55
at least 5000 reads	0.21

QC CHARTS

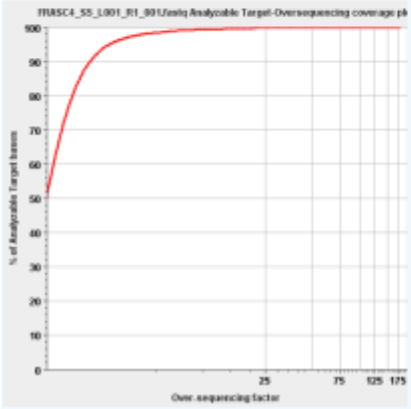
Analyzable Target-Coverage plot:



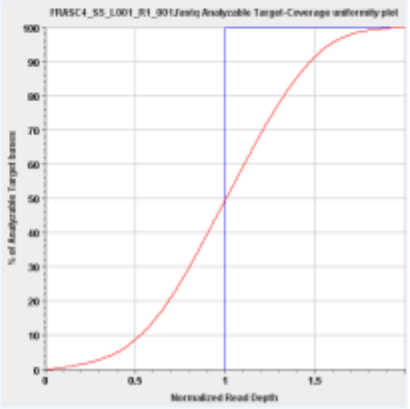
Covered Region-Coverage plot:



Analyzable Target-Oversequencing coverage plot:



Analyzable Target-Coverage uniformity plot:



EXONS TABLE

Gene	Transcript	Exon#	Percentage of positions < 20
ANKRD11	NM_001256182	NM_001256182.ex.14	3.36
PTPN11	NM_080601	NM_080601.ex.1	24.68
BRAF	NM_004333	NM_004333.ex.1	2.02
SF3B4	NM_005850	NM_005850.ex.3	3.51
PDS5B	NM_015032	NM_015032.ex.27	0.76
SOS1	NM_005633	NM_005633.ex.1	6.3
KMT2A	NM_005933	NM_005933.ex.1	18.94
NIPBL	NM_015384	NM_015384.ex.33	0.93
PDS5A	NM_001100400	NM_001100400.ex.4	100
ANKRD11	NM_001256182	NM_001256182.ex.10	6.98
NF1	NM_001042492	NM_001042492.ex.13	46.27
NF1	NM_001128147	NM_001128147.ex.15	9.08
NF1	NM_001042492	NM_001042492.ex.1	3.62
ARID1B	NM_020732	NM_020732.ex.1	5.97

Appendix 6: Primers used for Sanger sequencing

Primer	Sequence (5'-3' orientation)	Tm	Variants targeted
FRASC1/10- Forward	TGACCTGTATGGAGGGGAGA	58	*PTPN11:*c.215C>G (p.A72G) and *PTPN11:*c.211T>C (p.F71L)
FRASC1/10- Reverse	TCTGACACTCAGGGCACAAG	56.7	
FRASC2/18-Forward	GTGATCCCAATGAGCCTGTT	55.1	*PTPN11:*c.922A>G (p.N308D)
FRASC2/18-Reverse	GGGCTTTGAATTGTTGCACT	54	
FRASC14-Forward	GGCTTTTATGCAGACTTTTCG	53.4	*SOS1:*c.806T>C (p.M269T)
FRASC14-Reverse	TTTCCCCCAAACAGGATGTA	53.6	
FRASC17-Forward	CCCAAGGTGCCCTATTACCT	56.6	*RAF1:*c.781C>A (p.P261T)
FRASC17-Reverse	AGCCTAAGTGCCAATCATGG	54.9	
FRASC19-Forward	CCTTGCAGAAGAAGCTGGAG	56.8	*MAP2K1*:c.215G>A(p.S72N)
FRASC19-Reverse	AGAAGGCTTGTGGGAGACCT	57.9	
FRASC20-Forward	AAAATCCGACGTGGAAGATG	52.6	*PTPN11:*c.182A>G (p.D61G)
FRASC20-Reverse	ACTCAGCCAAAGTGGCAAAT	55.3	
FRASC39/73-Forward	CACATGTTTGACAGCGGAAA	54.8	*BRAF:*c.722C>T (p.T241M) and *BRAF:*c.770A>G (p.Q257R)
FRASC39/73-Reverse	AACCCCCGGTTTTCATTTTA	53.4	
FRASC60-Forward	CTTGGGCTGCATGTTCTCTT	55.3	*SOS1:*c.1649T>C (p.L550P)
FRASC60-Reverse	AATTGGATGGCAGCATTGAT	53	
FRASC81-Forward	CCTGGCTCTGCAGTTTCTCT	57.1	*PTPN11:*c.1510A>G (p.M504V)
FRASC81-Reverse	CCTGCGCTGTAGTGTTTCAA	55.6	
FRASC4/64-Forward	CACATGTTTGACAGCGGAAA	53.6	*BRAF:*c.770A>G (p.Q257R)
FRASC4/64-Reverse	AACCCCCGGTTTTCATTTTA	53.4	
FRASC21-Forward	CCAGGTTGAATGACATGTGG	53.5	*LZTR1:*c.690C>G (p.N230K)
FRASC21-Reverse	GTTATTTGGCTCCGCTTTG	52.3	
FRASC69-Forward	AGTCACCGTAAGGGATGCAG	56.5	*LZTR1:*c.848G>A (p.R283Q)
FRASC69-Reverse	CACCCCCAAACACATAGAGG	55.4	
FRASC15-Forward	TGAATCCTCATCTGGGGAAG	53.8	*LZTR1:* c.730T>C (p.S244P)
FRASC15-Reverse	GTTATTTTGGCTCCGCTTTG	52.6	
FRASC36-Forward	GGAGAAATTTGCCACTTTGG	52.3	*LZTR1:* c.730T>C (p.S244P)
FRASC36-Reverse	TCTGACACTCAGGGCACAAG	56.7	
FRASC63-BRAF-Forward	AGGTGATGTGGCAGTGAAAA	59.1	*BRAF:* c.1502A>T (p.E501V)
FRASC63-*BRAF*-Reverse	TTCCAGAGCCCAGGACTAAA	59.8	

FRASC5-*NRAS*-Forward	AGCAATTTGAGGGACAAACC	59	*NRAS:* c.173C>T (p.T58I)
FRASC5-*NRAS*-Reverse	TGGTAACCTCATTTCCCCATA	59.1	
FRASC42-*NF1*-Forward	GGCATGCATGAGAGATATTCC	59.5	*NF1:* c.6772C>T (p.R2258X)
FRASC42-*NF1*-Reverse	TTTCATTGACCTCAAATTTAAACG	59.4	
FRASC44/46-*NF1*-Forward	CCATCATGCCCAGCTAATTT	59.9	*NF1:* c.496-497 del (p.V166Lfs*) and *NF1:* c.569T>G (p.L190X)
FRASC44/46-*NF1*-Reverse	AGCACCCTTGCATTCTCTGT	59.8	
FRASC45-*NF1*-Forward	TCAGCCCTGCACATTCTGTA	60.4	*NF1*: c.3721C>T (p.R1241X)
FRASC45-*NF1*-Reverse	TTTTACCGGTTAATCCTTGGAA	59.7	
FRASC53-*NF1*-Forward	CTCCTGGGTCAGGGAGTACA	60.1	*NF1:* c.5204_5205 del (p.K1735Sfs*)
FRASC53-*NF1*-Reverse	AGATTACAGGTGCCCACCAC	59.8	
NF1-c.27G-A-Forward	CGTGGAAAGGATCCCACTT	59	*NF1:* c.5204_5205 del (p.K1735Sfs*)
NF1-c.27G-A-Reverse	CTTCCCTTCCTTTCCTCCAG	60.1	
FRASC62-*LZTR1*-Forward	GGAGGGGATTCATGGTTTTC	60.5	*LZTR1*: c.2269C>T (p.Q757X)
FRASC62-*LZTR1*-Reverse	CCTGACCTTGGTGAACTGGT	60.0	
FRASC22-*MAP2K2*-Forward	CCTGGATTCTGGTCATTGGT	59.7	*MAP2K2:* c.169T>G (p.F57V)
FRASC22-*MAP2K2*-Reverse	ACCTGCCTCCTGTTTCCAG	60.2	
FRASC100-*PTPN11*-Forward	TTCATTCTGTCCAACCAGCA	60.2	*PTPN11*: c.1381G>A (p.A461T)
FRASC100-*PTPN11*-Reverse	CCAGACTGTTTTCGTGAGCA	60	
FRASC99-*BRAF*-Foward	GGCTTGACTGGAGTGAAAGG	58.4	*BRAF:*c.1741A>G (p.N581D)
FRASC99-*BRAF*-Reverse	CTGTGGTATCCTGCTCTCCT	58.2	
FRASC38-*NF1*-Forward	ATTCAGGCTCTGCTGGTTCT	50	*NF1:* c.1658A>G (p.H553R)
FRASC38-*NF1*-Reverse	AGGAAGCAGACCACAAAATTCT	40.9	
FRASC80-*CDK13*-Forward	ACAATTGGGTGGCGAGTAGA	50	*CDK13:* c.2149G>A (p.Gly717Arg)
FRASC80-*CDK13*-Reverse	ACAGACCCACTTACATGTACTCA	43.5	
FRASC97-*TRPS1*-Forward	GACCTGCGAAACACCCAAAT	59	*TRPS1:* c.3403G>C (p.Pro1135Ala)
FRASC97-*TRPS1*-Reverse	GCCACAGTGCACACATTTTG	58.8	

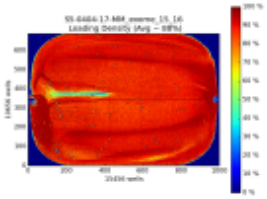
Appendix 7.1: Summary of data obtained on Ion-540™ chip 4, WES run 2

Run Summary

16.9 G	88
Total Bases	Key Signal

88%
ISP Loading

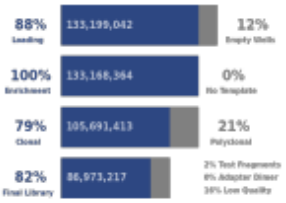
ISP Density



86,973,217
Total Reads

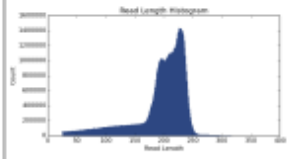
66%
Usable Reads

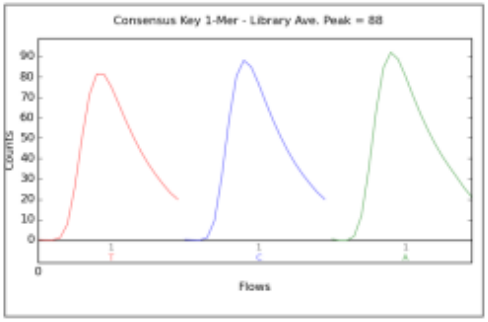
ISP Summary



194 bp	207 bp	226 bp
Mean	Median	Mode

Read Length





Addressable Wells	151,539,288	
With ISPs	133,199,042	87.9%
Live	133,168,364	100.0%
Test Fragment	1,824,955	01.4%
Library	131,343,409	98.6%

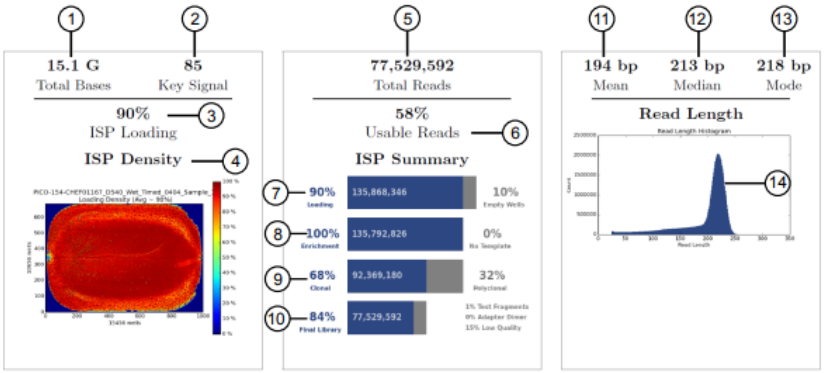
Library ISPs	131,343,409	
Filtered: Polyclonal	27,476,951	20.9%
Filtered: Low Quality	16,862,992	12.8%
Filtered: Adapter Dimer	30,249	00.0%
Final Library ISPs	**86,973,217**	**66.2%**

Barcode Name	Sample	Bases	$\geq Q20$	Reads	Mean Read Length	Read Length Histogram
No barcode	none	144,283,126	123,103,673	797,264	180 bp	
IonXpress_015	FRASC96	7,112,597,227	6,273,740,578	36,506,331	194 bp	
IonXpress_016	FRASC97	9,600,851,734	8,408,239,883	49,662,736	193 bp	

Appendix 7.2: Assessing WES run report example from the Torrent _Suite software

Run report metrics before alignment

To assess the sequencing run, you can review the quality metrics for the unaligned reads. Primary pipeline processing, basecalling, and signal processing generate these metrics. The following prealignment metrics are provided in the run report. Access the run report in the Torrent Suite™ Software using the **Data** tab.



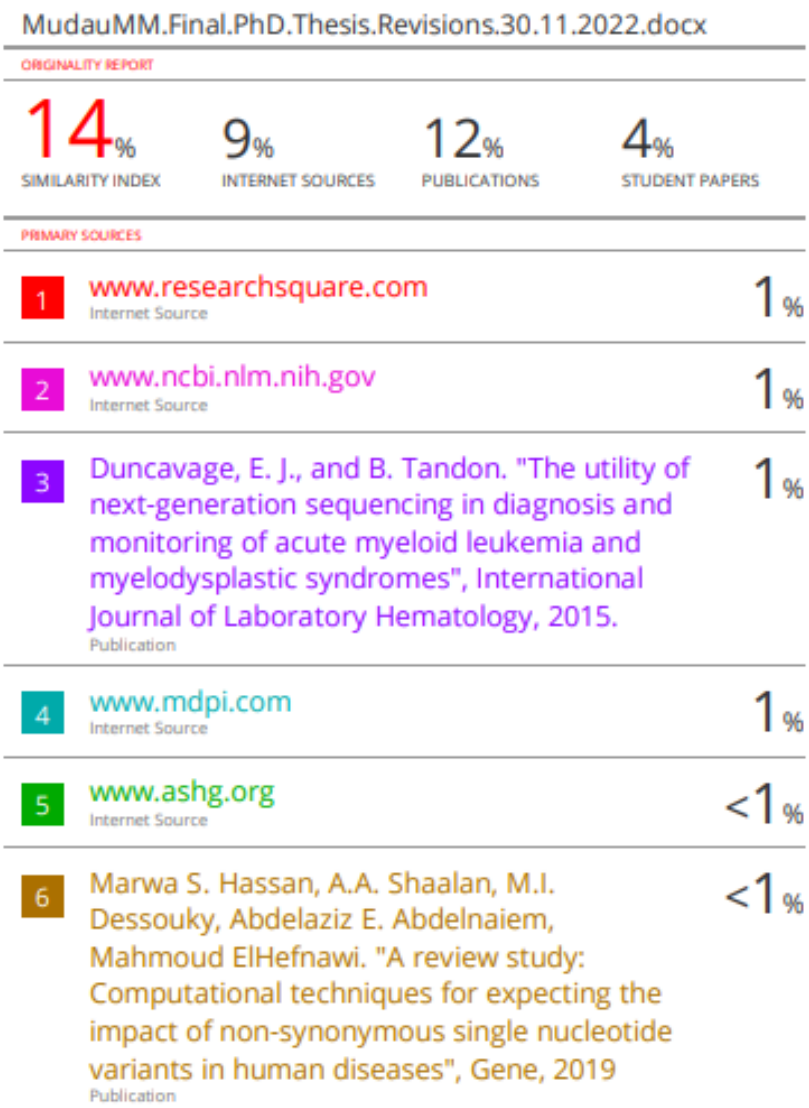
Number	Metric	Description
1	Total Bases	The number of filtered and trimmed base pairs reported in the output BAM file.
2	Key Signal	The average signal for all library ISPs with library key (TCAG).
3	ISP Loading	The percentage of chip wells that contain an Ion Sphere Particle (ISP). The percentage value considers only addressable wells.
4	ISP Loading Density	A visual representation of well loading distribution on the chip surface. Red color indicates areas of high loading and blue indicates areas of low loading.
5	Total Reads	The total number of filtered and trimmed reads independent of length reported in the output BAM file.
6	Usable Reads	The percentage of library ISPs that pass the polyclonal, low quality, and primer-dimer filters. This percentage is calculated by dividing final library ISPs by library ISPs.
7	ISP Summary-Loading	The percentage of chip wells that contain an ISP. The percentage value considers addressable wells.
8	ISP Summary-Enrichment	The predicted number of live ISPs that have a key signal identical to the library key signal or test fragment (TF) key signal. The Percent Enrichment value reported is the number of loaded wells with live ISPs that are Library ISPs or TF ISPs. This number is calculated by dividing wells with live ISPs by the number of wells loaded with ISPs.
9	ISP Summary-Clonality	The percentage of clonal ISPs (all library and TF ISPs that are clonal, not polyclonal). An ISP is clonal if all of its DNA fragments are cloned from a single original template. All the fragments on such an ISP are identical and they respond in unison as each nucleotide is flowed in turn across the chip. This percentage is calculated by dividing the number of ISPs with a single DNA template by the number of live wells.
10	ISP Summary-Final Library	The percentage of reads, which pass all filters, and which are recorded in the output BAM file. This value can be different from the Total Reads due to technicalities associated with read trimming beyond a minimal requirement that results in Total Reads being slightly less than Final Library.
11	Mean Read Length	The average length, in base pairs, of called reads.
12	Median Read Length	The median length, in base pairs, of called reads.
13	Mode Read Length	The mode length, in base pairs, of called reads.
14	Read Length Histogram	The read length histogram is a histogram of the trimmed lengths of all reads present in the output files.

Appendix 8: ACMG criteria for classifying pathogenic variants (Richards et al., 2015)

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

LOF, loss of function; OR, odds ratio.

Appendix 9: Turnitin report



MudauMM.Final.PhD.Thesis.Revisions.30.11.2022.docx

ORIGINALITY REPORT

14%	9%	12%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchsquare.com Internet Source	1%
2	www.ncbi.nlm.nih.gov Internet Source	1%
3	Duncavage, E. J., and B. Tandon. "The utility of next-generation sequencing in diagnosis and monitoring of acute myeloid leukemia and myelodysplastic syndromes", International Journal of Laboratory Hematology, 2015. Publication	1%
4	www.mdpi.com Internet Source	1%
5	www.ashg.org Internet Source	<1%
6	Marwa S. Hassan, A.A. Shaalan, M.I. Dessouky, Abdelaziz E. Abdelnaiem, Mahmoud ElHefnawi. "A review study: Computational techniques for expecting the impact of non-synonymous single nucleotide variants in human diseases", Gene, 2019 Publication	<1%

4.11. List of websites with date accessed

1000 genomes project: https://www.genome.gov/27528684/1000-genomes-project (accessed, February 2019)

Agilent SureCall: (accessed, January 2019) https://www.agilent.com/cs/library/software/public/Release%20Notes%20SureCall

Agilent SureDesign: https://earray.chem.agilent.com/suredesign/ (accessed, July 2018)

Agilent SureSelect: https://www.agilent.com/cs/library/usermanuals/public/G9681-90000.pdf (July 2018)

ClinGen: https://clinicalgenome.org/ (accessed, February 2018)

ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/ (accessed, February 2019)

Ensembl variant effect predictor: https://www.ensembl.org/info/docs/tools/vep/index.html (accessed, February 2019)

Exome Aggregation consortium (ExAC): https://exac.broadinstitute.org/ (accessed, February 2018)

Human disease mutation database (HGMD): http://www.hgmd.cf.ac.uk/ac/index.php (accessed, February 2019)

Illumina MiSeq: https://emea.illumina.com/systems/sequencing-platforms/miseq.html (accessed, November 2018)

Ion-Reporter software: https://ionreporter.thermofisher.com/ir/ (accessed, February 2021)

Ion-torrent ThermoFisher NGS platform: https://www.thermofisher.com/za/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-applications.html (accessed, November 2020)

MiSeq reagent kit used: https://emea.illumina.com/systems/sequencing-platforms/miseq/specifications.html (accessed, November 2018)

Moon (diploid) annotation software: https://www.diploid.com/moon (accessed, March 2021)

The Genome Aggregation Database (gnomAD): https://gnomad.broadinstitute.org/ (accessed, February 2020)

wANNOVAR software: https://wannovar.wglab.org/ (accessed, February 2019)

4.12.References

A. Mahler, E., Johannsen, J., Tsiakas, K., Kloth, K., Lüttgen, S., Mühlhausen, C., Alhaddad, B., B. Haack, T., et al. 2019. Exome Sequencing in Children. *Deutsches Ärzteblatt International*. 116(12):197–204. DOI: 10.3238/arztebl.2019.0197.

Abe, Y., Aoki, Y., Kuriyama, S., Kawame, H., Okamoto, N., Kurosawa, K., Ohashi, H., Mizuno, 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. DOI: 10.1002/ajmg.a.35292.

Adachi, M., Abe, Y., Aoki, Y. and Matsubara, Y. 2012. Epilepsy in RAS/MAPK syndrome: Two cases of cardio-facio-cutaneous syndrome with epileptic encephalopathy and a literature review. *Seizure*. 21(1):55–60. DOI: 10.1016/j.seizure.2011.07.013.

Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Mirzaa, G.M. and Amemiya, A. Eds. 1993. *GeneReviews®*. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1116/ [2022, February 03].

Adams, D.R. and Eng, C.M. 2018. Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. *New England Journal of Medicine*. 379(14):1353–1362. DOI: 10.1056/NEJMra1711801.

Adzhubei, I., Jordan, D.M. and Sunyaev, S.R. (in press). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*. 0 7:Unit7.20. DOI: 10.1002/0471142905.hg0720s76.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. 2010. A method and server for predicting damaging missense mutations. *Nature Methods*. 7(4):248–249. DOI: 10.1038/nmeth0410-248.

Ahronowitz, I., Xin, W., Kiely, R., Sims, K., MacCollin, M. and Nunes, F.P. 2007. Mutational spectrum of the NF2 gene: a meta-analysis of 12 years of research and diagnostic laboratory findings. *Human Mutation*. 28(1):1–12. DOI: 10.1002/humu.20393.

van den Akker, W.M.R., Brummelman, I., Martis, L.M., Timmermans, R.N., Pfundt, R., Kleefstra, T., Willemsen, M.H., Gerkes, E.H., et al. 2018. De novo variants in CDK13 associated with syndromic ID/DD: Molecular and clinical delineation of 15 individuals and a further review. *Clinical Genetics*. 93(5):1000–1007. DOI: 10.1111/cge.13225.

Alaidarous, A., Parfait, B., Ferkal, S., Cohen, J., Wolkenstein, P. and Mazereeuw-Hautier, J. 2019. Segmental schwannomatosis: characteristics in 12 patients. *Orphanet Journal of Rare Diseases*. 14(1):207. DOI: 10.1186/s13023-019-1176-4.

Alawi, I.A., Riyami, M.A., Barroso-Gil, M., Powell, L., Olinger, E., Salmi, I.A. and Sayer, J.A. 2021. The diagnostic yield of whole exome sequencing as a first approach in consanguineous Omani renal ciliopathy syndrome patients. 10:207. F1000Research. DOI: 10.12688/f1000research.40338.2.

Allanson, J.E. and Roberts, A.E. 1993. Noonan Syndrome. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, Eds. Seattle

(WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1124/ [2021, September 28].

Altmüller, F., Lissewski, C., Bertola, D., Flex, E., Stark, Z., Spranger, S., Baynam, G., Buscarilli, M., et al. 2017. Genotype and phenotype spectrum of NRAS germline variants. *European journal of human genetics: EJHG*. 25(7):823–831. DOI: 10.1038/ejhg.2017.65.

Anderson, J.L. and Gutmann, D.H. 2015a. Neurofibromatosis type 1. *Handbook of Clinical Neurology*. 132:75–86. DOI: 10.1016/B978-0-444-62702-5.00004-4.

Anderson, J.L. and Gutmann, D.H. 2015b. Chapter 4 - Neurofibromatosis type 1. In *Handbook of Clinical Neurology*. V. 132. M.P. Islam and E.S. Roach, Eds. (Neurocutaneous Syndromes). Elsevier. 75–86. DOI: 10.1016/B978-0-444-62702-5.00004-4.

Ankala, A., da Silva, C., Gualandi, F., Ferlini, A., Bean, L.J.H., Collins, C., Tanner, A.K. and Hegde, M.R. 2015. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Annals of Neurology*. 77(2):206–214. DOI: 10.1002/ana.24303.

Aoki, Y., Niihori, T., Inoue, S. and Matsubara, Y. 2016. Recent advances in RASopathies. *Journal of Human Genetics*. 61(1):33–39. DOI: 10.1038/jhg.2015.114.

Aramini, J.M., Vorobiev, S.M., Tuberty, L.M., Janjua, H., Campbell, E.T., Seetharaman, J., Su, M., Huang, Y.J., et al. 2015. The RAS-Binding Domain of Human BRAF Protein Serine/Threonine Kinase Exhibits Allosteric Conformational Changes upon Binding HRAS. *Structure (London, England : 1993)*. 23(8):1382–1393. DOI: 10.1016/j.str.2015.06.003.

Ari, Ş. and Arikan, M. 2016. Next-Generation Sequencing: Advantages, Disadvantages, and Future. In *Plant Omics: Trends and Applications*. K.R. Hakeem, H. Tombuloğlu, and G. Tombuloğlu, Eds. Cham: Springer International Publishing. 109–135. DOI: 10.1007/978-3-319-31703-8_5.

Arts, P., Simons, A., AlZahrani, M.S., Yilmaz, E., AlIdrissi, E., van Aerde, K.J., Alenezi, N., AlGhamdi, H.A., et al. 2019. Exome sequencing in routine diagnostics: a generic test for 254 patients with primary immunodeficiencies. *Genome Medicine*. 11(1):38. DOI: 10.1186/s13073-019-0649-3.

Ascertainment bias. 2017. Available: https://catalogofbias.org/biases/ascertainment-bias/ [2022, March 10].

Athota, J.P., Bhat, M., Nampoothiri, S., Gowrishankar, K., Narayanachar, S.G., Puttamallesh, V., Farooque, M.O. and Shetty, S. 2020. Molecular and clinical studies in 107 Noonan syndrome affected individuals with PTPN11 mutations. *BMC medical genetics*. 21(1):50. DOI: 10.1186/s12881-020-0986-5.

Austin-Tse, C.A., Jobanputra, V., Perry, D.L., Bick, D., Taft, R.J., Venner, E., Gibbs, R.A., Young, T., et al. 2022. Best practices for the interpretation and reporting of clinical whole genome sequencing. *npj Genomic Medicine*. 7(1):1–13. DOI: 10.1038/s41525-022-00295-z.

Aziz, N., Zhao, Q., Bry, L., Driscoll, D.K., Funke, B., Gibson, J.S., Grody, W.W., Hegde, M.R., et al. 2015. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. *Archives of Pathology and Laboratory Medicine*. 139(4):481–493. DOI: 10.5858/arpa.2014-0250-CP.

Baudhuin, L.M., Lagerstedt, S.A., Klee, E.W., Fadra, N., Oglesbee, D. and Ferber, M.J. 2015a. Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. *The Journal of molecular diagnostics: JMD*. 17(4):456–461. DOI: 10.1016/j.jmoldx.2015.03.004.

Baudhuin, L.M., Lagerstedt, S.A., Klee, E.W., Fadra, N., Oglesbee, D. and Ferber, M.J. 2015b. Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. *The Journal of molecular diagnostics: JMD*. 17(4):456–461. DOI: 10.1016/j.jmoldx.2015.03.004.

Bauer, P., Kandaswamy, K.K., Weiss, M.E.R., Paknia, O., Werber, M., Bertoli-Avella, A.M., Yüksel, Z., Bochinska, M., et al. 2019. Development of an evidence-based algorithm that optimizes sensitivity and specificity in ES-based diagnostics of a clinically heterogeneous patient population. *Genetics in Medicine*. 21(1):53–61. DOI: 10.1038/s41436-018-0016-6.

Baynam, G.S., Groft, S., van der Westhuizen, F.H., Gassman, S.D., du Plessis, K., Coles, E.P., Selebatso, E., Selebatso, M., et al. 2020. A call for global action for rare diseases in Africa. *Nature Genetics*. 52(1):21–26. DOI: 10.1038/s41588-019-0552-2.

Bayrak-Toydemir, P. and Stevenson, D.A. 1993. Capillary Malformation-Arteriovenous Malformation Syndrome. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K. Stephens, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK52764/ [2020, October 06].

Beck, T.F., Mullikin, J.C., NISC Comparative Sequencing Program and Biesecker, L.G. 2016. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clinical Chemistry*. 62(4):647–654. DOI: 10.1373/clinchem.2015.249623.

Benelli, E., Bruno, I., Belcaro, C., Ventura, A. and Berti, I. 2015. Legius syndrome: case report and review of literature. *Italian Journal of Pediatrics*. 41(1):8. DOI: 10.1186/s13052-015-0115-9.

Bergant, G., Maver, A., Lovrecic, L., Čuturilo, G., Hodzic, A. and Peterlin, B. 2018. Comprehensive use of extended exome analysis improves diagnostic yield in rare disease: a retrospective survey in 1,059 cases. *Genetics in Medicine*. 20(3):303–312. DOI: 10.1038/gim.2017.142.

Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science*. 367(6484):eaay5012. DOI: 10.1126/science.aay5012.

Bertola, D., Buscarilli, M., Stabley, D.L., Baker, L., Doyle, D., Bartholomew, D.W., Sol-Church, K. and Gripp, K.W. 2017. Phenotypic Spectrum of Costello Syndrome Individuals Harboring the Rare HRAS Mutation p.Gly13Asp. *American journal of medical genetics. Part A*. 173(5):1309–1318. DOI: 10.1002/ajmg.a.38178.

Bertola, D.R., Pereira, A.C., Albano, L.M.J., De Oliveira, P.S.L., Kim, C.A. and Krieger, J.E. 2006. PTPN11 gene analysis in 74 Brazilian patients with Noonan syndrome or Noonan-like phenotype. *Genetic Testing*. 10(3):186–191. DOI: 10.1089/gte.2006.10.186.

Bhambhani, V. and Muenke, M. 2014. Noonan Syndrome. *American Family Physician*. 89(1):37–43.

Bhoj, E.J., Yu, Z., Guan, Q., Ahrens-Nicklas, R., Cao, K., Luo, M., Tischler, T., Deardorff, M.A., et al. 2017. Phenotypic predictors and final diagnoses in patients referred for RASopathy testing by targeted next-generation sequencing. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. 19(6):715–718. DOI: 10.1038/gim.2016.169.

Boon, L.M., Mulliken, J.B. and Vikkula, M. 2005. RASA1: variable phenotype with capillary and arteriovenous malformations. *Current Opinion in Genetics and Development*. 15(3):265–269. DOI: 10.1016/j.gde.2005.03.004.

Borges, M.G., Rocha, C.S., Carvalho, B.S. and Lopes-Cendes, I. 2020. Methodological differences can affect sequencing depth with a possible impact on the accuracy of genetic diagnosis. *Genetics and Molecular Biology*. 43. DOI: 10.1590/1678-4685-GMB-2019-0270.

Bostwick, B. 1993. CDK13-Related Disorder. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K.W. Gripp, G.M. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK536784/ [2022, March 12].

Bostwick, B.L., McLean, S., Posey, J.E., Streff, H.E., Gripp, K.W., Blesson, A., Powell-Hamilton, N., Tusi, J., et al. 2017. Phenotypic and molecular characterisation of CDK13-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders. *Genome Medicine*. 9(1):73. DOI: 10.1186/s13073-017-0463-8.

Bouck, Z., Calzavara, A.J., Ivers, N.M., Kerr, E.A., Chu, C., Ferguson, J., Martin, D., Tepper, J., et al. 2020. Association of Low-Value Testing With Subsequent Health Care Use and Clinical Outcomes Among Low-risk Primary Care Outpatients Undergoing an Annual Health Examination. *JAMA Internal Medicine*. 180(7):973–983. DOI: 10.1001/jamainternmed.2020.1611.

Božović, I.B., Maver, A., Leonardis, L., Meznaric, M., Osredkar, D. and Peterlin, B. 2021. Diagnostic yield of exome sequencing in myopathies: Experience of a Slovenian tertiary centre. *PLOS ONE*. 16(6):e0252953. DOI: 10.1371/journal.pone.0252953.

Brandariz-Fontes, C., Camacho-Sanchez, M., Vilà, C., Vega-Pla, J.L., Rico, C. and Leonard, J.A. 2015. Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. *Scientific Reports*. 5(1):8056. DOI: 10.1038/srep08056.

Brasil, A.S., Pereira, A.C., Wanderley, L.T., Kim, C.A., Malaquias, A.C., Jorge, A.A.L., Krieger, J.E. and Bertola, D.R. 2010. PTPN11 and KRAS gene analysis in patients with Noonan and Noonan-like syndromes. *Genetic Testing and Molecular Biomarkers*. 14(3):425–432. DOI: 10.1089/gtmb.2009.0192.

Brems, H., Pasmant, E., Van Minkelen, R., Wimmer, K., Upadhyaya, M., Legius, E. and Messiaen, L. 2012. Review and update of SPRED1 mutations causing Legius syndrome. *Human Mutation*. 33(11):1538–1546. DOI: 10.1002/humu.22152.

Buermans, H.P.J. and den Dunnen, J.T. 2014. Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 1842(10):1932–1941. DOI: 10.1016/j.bbadis.2014.06.015.

Calì, F., Chiavetta, V., Ruggeri, G., Piccione, M., Selicorni, A., Palazzo, D., Bonsignore, M., Cereda, A., et al. 2017. Mutation spectrum of NF1 gene in Italian patients with

neurofibromatosis type 1 using Ion Torrent PGM™ platform. *European Journal of Medical Genetics*. 60(2):93–99. DOI: 10.1016/j.ejmg.2016.11.001.

Carcavilla, A., Suárez-Ortega, L., Sánchez, A.R., Gonzalez-Casado, I., Ramón-Krauel, M., Labarta, J.I., Gonzalez, S.Q., Galán, I.R., et al. (in press). Noonan syndrome: genetic and clinical update and treatment options. *Anales de Pediatría (English Edition)*. 93(1):61.e1-61.e14. DOI: 10.1016/j.anpede.2020.04.009.

Caspar, S. m., Dubacher, N., Kopps, A. m., Meienberg, J., Henggeler, C. and Matyas, G. 2018. Clinical sequencing: From raw data to diagnosis with lifetime value. *Clinical Genetics*. 93(3):508–519. DOI: 10.1111/cge.13190.

Castellanos, E., Gel, B., Rosas, I., Tornero, E., Santín, S., Pluvinet, R., Velasco, J., Sumoy, L., et al. 2017. A comprehensive custom panel design for routine hereditary cancer testing: preserving control, improving diagnostics and revealing a complex variation landscape. *Scientific Reports*. 7(1):39348. DOI: 10.1038/srep39348.

Castellanos, E., Rosas, I., Negro, A., Gel, B., Alibés, A., Baena, N., Pineda, M., Pi, G., 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. DOI: 10.1111/cge.13649.

Cavalli-Sforza, L.L. 2005. The Human Genome Diversity Project: past, present and future. *Nature Reviews Genetics*. 6(4):333–340. DOI: 10.1038/nrg1596.

Cavin, R.K., Lugli, P. and Zhirnov, V.V. 2012. Science and Engineering Beyond Moore's Law. *Proceedings of the IEEE*. 100(Special Centennial Issue):1720–1749. DOI: 10.1109/JPROC.2012.2190155.

Cesarini, L., Alfieri, P., Pantaleoni, F., Vasta, I., Cerutti, M., Petrangeli, V., Mariotti, P., Leoni, C., et al. 2009. Cognitive profile of disorders associated with dysregulation of the RAS/MAPK signaling cascade. *American Journal of Medical Genetics. Part A*. 149A(2):140–146. DOI: 10.1002/ajmg.a.32488.

Chang, X. and Wang, K. 2012. wANNOVAR: annotating genetic variants for personal genomes via the web. *Journal of Medical Genetics*. 49(7):433–436. DOI: 10.1136/jmedgenet-2012-100918.

Chen, H., Li, X., Liu, X., Wang, J., Zhang, Z., Wu, J., Huang, M., Guo, Y., et al. 2019. Clinical and mutation profile of pediatric patients with RASopathy-associated hypertrophic cardiomyopathy: results from a Chinese cohort. *Orphanet Journal of Rare Diseases*. 14(1):29. DOI: 10.1186/s13023-019-1010-z.

Chen, J.-L., Zhu, X., Zhao, T.-L., Wang, J., Yang, Y.-F. and Tan, Z.-P. 2014. Rare copy number variations containing genes involved in RASopathies: deletion of SHOC2 and duplication of PTPN11. *Molecular Cytogenetics*. 7:28. DOI: 10.1186/1755-8166-7-28.

Chen, L., Xue, F., Xu, J., He, J., Fu, W., Zhang, Z. and Kang, Q. 2019. Five novel NF1 gene pathogenic variants in 10 different Chinese families with neurofibromatosis type 1. *Molecular Genetics and Genomic Medicine*. 7(9):e904. DOI: 10.1002/mgg3.904.

Chen, P.-C., Yin, J., Yu, H.-W., Yuan, T., Fernandez, M., Yung, C.K., Trinh, Q.M., Peltekova, V.D., 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. DOI: 10.1073/pnas.1324128111.

Chin, E.L., da Silva, C. and Hegde, M. 2013a. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genetics*. 14:6. DOI: 10.1186/1471-2156-14-6.

Chin, E.L., da Silva, C. and Hegde, M. 2013b. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genetics*. 14:6. DOI: 10.1186/1471-2156-14-6.

Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloğlu, A., et al. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 106(45):19096–19101. DOI: 10.1073/pnas.0910672106.

Chong, Z.Z. and Maiese, K. 2007. The Src homology 2 domain tyrosine phosphatases SHP-1 and SHP-2: diversified control of cell growth, inflammation, and injury. *Histology and Histopathology*. 22(11):1251–1267. DOI: 10.14670/HH-22.1251.

Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., Chimusa, E.R., Christoffels, A., et al. 2017a. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nature Communications*. 8(1):2062. DOI: 10.1038/s41467-017-00663-9.

Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., Chimusa, E.R., Christoffels, A., et al. 2017b. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nature Communications*. 8(1):2062. DOI: 10.1038/s41467-017-00663-9.

Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., et al. 2020. High-depth African genomes inform human migration and health. *Nature*. 586(7831):741–748. DOI: 10.1038/s41586-020-2859-7.

Chrystoja, C.C. and Diamandis, E.P. 2014. Whole Genome Sequencing as a Diagnostic Test: Challenges and Opportunities. *Clinical Chemistry*. 60(5):724–733. DOI: 10.1373/clinchem.2013.209213.

Cimino, P.J. and Gutmann, D.H. 2018. Neurofibromatosis type 1. *Neurogenetics, Part II*. 799–811. DOI: 10.1016/B978-0-444-64076-5.00051-X.

Cirstea, I.C., Kutsche, K., Dvorsky, R., Gremer, L., Carta, C., Horn, D., Roberts, A.E., Lepri, F., et al. 2010. A restricted spectrum of NRAS mutations causes Noonan syndrome. *Nature Genetics*. 42(1):27–29. DOI: 10.1038/ng.497.

Clark, M.M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., Watkins, K., Ellsworth, K., Camp, B., et al. 2019. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Science Translational Medicine*. 11(489):eaat6177. DOI: 10.1126/scitranslmed.aat6177.

Cleveland, M.H., Zook, J.M., Salit, M. and Vallone, P.M. 2018. Determining Performance Metrics for Targeted Next Generation Sequencing Panels Using Reference Materials. *The Journal of molecular diagnostics : JMD*. 20(5):583–590. DOI: 10.1016/j.jmoldx.2018.04.005.

Clift, K., Macklin, S., Halverson, C., McCormick, J.B., Abu Dabrh, A.M. and Hines, S. 2020. Patients' views on variants of uncertain significance across indications. *Journal of Community Genetics*. 11(2):139–145. DOI: 10.1007/s12687-019-00434-7.

De Cario, R., Kura, A., Suraci, S., Magi, A., Volta, A., Marcucci, R., Gori, A.M., Pepe, G., et al. 2020. Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice? *Frontiers in Genetics*. 11:592588. DOI: 10.3389/fgene.2020.592588.

Deng, C., Daley, T., Calabrese, P., Ren, J. and Smith, A.D. 2020. Predicting the Number of Bases to Attain Sufficient Coverage in High-Throughput Sequencing Experiments. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*. 27(7):1130–1143. DOI: 10.1089/cmb.2019.0264.

Devuyst, O. 2015a. The 1000 Genomes Project: Welcome to a New World. *Peritoneal Dialysis International : Journal of the International Society for Peritoneal Dialysis*. 35(7):676–677. DOI: 10.3747/pdi.2015.00261.

Devuyst, O. 2015b. The 1000 Genomes Project: Welcome to a New World. *Peritoneal Dialysis International : Journal of the International Society for Peritoneal Dialysis*. 35(7):676–677. DOI: 10.3747/pdi.2015.00261.

Df, V., K, S. and P, B. 2017. Reporting practices for variants of uncertain significance from next generation sequencing technologies. *European journal of medical genetics*. 60(10). DOI: 10.1016/j.ejmg.2017.07.016.

Dhamija, R., Plotkin, S., Asthagiri, A., Messiaen, L. and Babovic-Vuksanovic, D. 1993. Schwannomatosis. In *GeneReviews*®. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK487394/ [2021, May 10].

Digilio, M.C., Conti, E., Sarkozy, A., Mingarelli, R., Dottorini, T., Marino, B., Pizzuti, A. and Dallapiccola, B. 2002. Grouping of multiple-lentigines/LEOPARD and Noonan syndromes on the PTPN11 gene. *American Journal of Human Genetics*. 71(2):389–394. DOI: 10.1086/341528.

Durmaz, A.A., Karaca, E., Demkow, U., Toruner, G., Schoumans, J. and Cogulu, O. 2015. Evolution of Genetic Techniques: Past, Present, and Beyond. *BioMed Research International*. 2015:e461524. DOI: 10.1155/2015/461524.

Edouard, T., Combier, J.-P., Nédélec, A., Bel-Vialar, S., Métrich, M., Conte-Auriol, F., Lyonnet, S., Parfait, B., et al. 2010. Functional effects of PTPN11 (SHP2) mutations causing LEOPARD syndrome on epidermal growth factor-induced phosphoinositide 3-kinase/AKT/glycogen synthase kinase 3beta signaling. *Molecular and Cellular Biology*. 30(10):2498–2507. DOI: 10.1128/MCB.00646-09.

Eerola, I., Boon, L.M., Mulliken, J.B., Burrows, P.E., Dompmartin, A., Watanabe, S., Vanwijck, R. and Vikkula, M. 2003. Capillary malformation-arteriovenous malformation, a

new clinical and genetic disorder caused by RASA1 mutations. *American Journal of Human Genetics*. 73(6):1240–1249. DOI: 10.1086/379793.

Eisele, M. and Kappelmann-Fenzl, M. 2021. NGS Technologies. In *Next Generation Sequencing and Data Analysis*. M. Kappelmann-Fenzl, Ed. (Learning Materials in Biosciences). Cham: Springer International Publishing. 47–58. DOI: 10.1007/978-3-030-62490-3_4.

Ejaz, R., Carter, M. and Gripp, K. 1993. Lateral Meningocele Syndrome. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K.W. Gripp, G.M. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK368476/ [2022, January 28].

El Bouchikhi, I., Belhassan, K., Moufid, F.Z., Iraqui Houssaini, M., Bouguenouch, L., Samri, I., Atmani, S. and Ouldim, K. 2016. Noonan syndrome-causing genes: Molecular update and an assessment of the mutation rate. *International Journal of Pediatrics and Adolescent Medicine*. 3(4):133–142. DOI: 10.1016/j.ijpam.2016.06.003.

van El, C.G., Cornel, M.C., Borry, P., Hastings, R.J., Fellmann, F., Hodgson, S.V., Howard, H.C., Cambon-Thomsen, A., et al. 2013. Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. *European journal of human genetics: EJHG*. 21(6):580–584. DOI: 10.1038/ejhg.2013.46.

Estep, A.L., Tidyman, W.E., Teitell, M.A., Cotter, P.D. and Rauen, K.A. 2006a. HRAS mutations in Costello syndrome: Detection of constitutional activating mutations in codon 12 and 13 and loss of wild-type allele in malignancy. *American Journal of Medical Genetics Part A*. 140A(1):8–16. DOI: 10.1002/ajmg.a.31078.

Estep, A.L., Tidyman, W.E., Teitell, M.A., Cotter, P.D. and Rauen, K.A. 2006b. HRAS mutations in Costello syndrome: Detection of constitutional activating mutations in codon 12 and 13 and loss of wild-type allele in malignancy. *American Journal of Medical Genetics Part A*. 140A(1):8–16. DOI: 10.1002/ajmg.a.31078.

Esterling, L., Wijayatunge, R., Brown, K., Morris, B., Hughes, E., Pruss, D., Manley, S., Bowles, K.R., et al. (in press). Impact of a Cancer Gene Variant Reclassification Program Over a 20-Year Period. *JCO precision oncology*. 4:PO.20.00020. DOI: 10.1200/PO.20.00020.

EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, and Epi4K Consortium. 2014. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *American Journal of Human Genetics*. 95(4):360–370. DOI: 10.1016/j.ajhg.2014.08.013.

Evans, D.G. 1993. Neurofibromatosis 2. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1201/ [2021, September 27].

Evans, D.G., Bowers, N., Burkitt-Wright, E., Miles, E., Garg, S., Scott-Kitching, V., Penman-Splitt, M., Dobbie, A., 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. DOI: 10.1016/j.ebiom.2016.04.005.

Ezquieta, B., Santomé, J.L., Carcavilla, A., Guillén-Navarro, E., Pérez-Aytés, A., Sánchez del Pozo, J., García-Miñaur, S., Castillo, E., et al. 2012. Alterations in RAS-MAPK Genes in 200 Spanish Patients With Noonan and Other Neuro-Cardio-Facio-Cutaneous Syndromes. Genotype and Cardiopathy. *Revista Española de Cardiología (English Edition)*. 65(5):447–455. DOI: 10.1016/j.rec.2011.12.017.

Fahsold, R., Hoffmeyer, S., Mischung, C., Gille, C., Ehlers, C., Kücükceylan, N., Abdel-Nour, M., Gewies, A., et al. 2000. Minor Lesion Mutational Spectrum of the Entire NF1 Gene Does Not Explain Its High Mutability but Points to a Functional Domain Upstream of the GAP-Related Domain. *American Journal of Human Genetics*. 66(3):790–818.

Farkas, D.H. 2016. Clinical Validity and Utility: Putting the Patient Front and Center. *The Journal of Molecular Diagnostics*. 18(5):635–637. DOI: 10.1016/j.jmoldx.2016.06.003.

Federici, G. and Soddu, S. 2020a. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *Journal of Experimental and Clinical Cancer Research*. 39(1):46. DOI: 10.1186/s13046-020-01554-6.

Federici, G. and Soddu, S. 2020b. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *Journal of Experimental and Clinical Cancer Research*. 39(1):46. DOI: 10.1186/s13046-020-01554-6.

Friedman, J.M. 1993. Neurofibromatosis 1. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1109/ [2021, September 28].

Fujiki, R., Ikeda, M., Yoshida, A., Akiko, M., Yao, Y., Nishimura, M., Matsushita, K., Ichikawa, T., et al. 2018. Assessing the Accuracy of Variant Detection in Cost-Effective Gene Panel Testing by Next-Generation Sequencing. *The Journal of Molecular Diagnostics*. 20(5):572–582. DOI: 10.1016/j.jmoldx.2018.04.004.

Gampawar, P., Saba, Y., Werner, U., Schmidt, R., Müller-Myhsok, B. and Schmidt, H. 2019. Evaluation of the Performance of AmpliSeq and SureSelect Exome Sequencing Libraries for Ion Proton. *Frontiers in Genetics*. 10. DOI: 10.3389/fgene.2019.00856.

Gelb, B.D. and Tartaglia, M. 1993. Noonan Syndrome with Multiple Lentigines. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1383/ [2021, September 28].

Gelb, B.D., Cavé, H., Dillon, M.W., Gripp, K.W., Lee, J.A., Mason-Suares, H., Rauen, K.A., Williams, B., et al. 2018. ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genetics in Medicine*. 20(11):1334–1345. DOI: 10.1038/gim.2018.3.

Genome Aggregation Database Consortium, Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 581(7809):434–443. DOI: 10.1038/s41586-020-2308-7.

Giannoulatou, E., McVean, G., Taylor, I.B., McGowan, S.J., Maher, G.J., Iqbal, Z., Pfeifer, S.P., Turner, I., et al. 2013. Contributions of intrinsic mutation rate and selfish selection to levels of de novo HRAS mutations in the paternal germline. *Proceedings of the National Academy of Sciences of the United States of America*. 110(50):20152–20157. DOI: 10.1073/pnas.1311381110.

Gordon, L.G., White, N.M., Elliott, T.M., Nones, K., Beckhouse, A.G., Rodriguez-Acevedo, A.J., Webb, P.M., Lee, X.J., et al. 2020a. Estimating the costs of genomic sequencing in cancer control. *BMC Health Services Research*. 20(1):492. DOI: 10.1186/s12913-020-05318-y.

Gordon, L.G., White, N.M., Elliott, T.M., Nones, K., Beckhouse, A.G., Rodriguez-Acevedo, A.J., Webb, P.M., Lee, X.J., et al. 2020b. Estimating the costs of genomic sequencing in cancer control. *BMC Health Services Research*. 20(1):492. DOI: 10.1186/s12913-020-05318-y.

Graham, J.M., Kramer, N., Bejjani, B.A., Thiel, C.T., Carta, C., Neri, G., Tartaglia, M. and Zenker, M. 2009. Genomic Duplication of PTPN11 is an Uncommon Cause of Noonan Syndrome. *American journal of medical genetics. Part A*. 149A(10):2122–2128. DOI: 10.1002/ajmg.a.32992.

Gripp, K.W. and Rauen, K.A. 1993. Costello Syndrome. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K. Stephens, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1507/ [2020, October 06].

Gripp, K.W., Morse, L.A., Axelrad, M., Chatfield, K.C., Chidekel, A., Dobyns, W., Doyle, D., Kerr, B., et al. 2019. Costello syndrome: Clinical phenotype, genotype, and management guidelines. *American Journal of Medical Genetics Part A*. 179(9):1725–1744. DOI: 10.1002/ajmg.a.61270.

Gross, A.M., Frone, M., Gripp, K.W., Gelb, B.D., Schoyer, L., Schill, L., Stronach, B., Biesecker, L.G., 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. DOI: 10.1002/ajmg.a.61485.

Guleray, N., Kosukcu, C., Taskiran, Z.E., Simsek Kiper, P.O., Utine, G.E., Gucer, S., Tokatli, A., Boduroglu, K., et al. 2020. Atypical Presentation of Sengers Syndrome: A Novel Mutation Revealed with Postmortem Genetic Testing. *Fetal and Pediatric Pathology*. 39(2):163–171. DOI: 10.1080/15513815.2019.1639089.

Gulilat, M., Lamb, T., Teft, W.A., Wang, J., Dron, J.S., Robinson, J.F., Tirona, R.G., Hegele, R.A., et al. 2019. Targeted next generation sequencing as a tool for precision medicine. *BMC Medical Genomics*. 12(1):81. DOI: 10.1186/s12920-019-0527-2.

Guo, Y.-X., Ma, H.-X., Zhang, Y.-X., Chen, Z.-H. and Zhai, Q.-X. 2021. Whole-Exome Sequencing for Identifying Genetic Causes of Intellectual Developmental Disorders . *International Journal of General Medicine*. 14:1275–1282. DOI: 10.2147/IJGM.S300775.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., et al. 2015. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 517(7534):327–332. DOI: 10.1038/nature13997.

H3Africa Consortium, Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V.M., Adebamowo, C., Adebiyi, E., Ademola, A.D., et al. 2014. Research capacity. Enabling the genomic revolution in Africa. *Science (New York, N.Y.)*. 344(6190):1346–1348. DOI: 10.1126/science.1251546.

Hamilton, M.J., Caswell, R.C., Canham, N., Cole, T., Firth, H.V., Foulds, N., Heimdal, K., Hobson, E., et al. 2018. Heterozygous mutations affecting the protein kinase domain of CDK13 cause a syndromic form of developmental delay and intellectual disability. *Journal of Medical Genetics*. 55(1):28–38. DOI: 10.1136/jmedgenet-2017-104620.

Han, J.Y. and Lee, I.G. 2020a. Genetic tests by next-generation sequencing in children with developmental delay and/or intellectual disability. *Clinical and Experimental Pediatrics*. 63(6):195–202. DOI: 10.3345/kjp.2019.00808.

Han, J.Y. and Lee, I.G. 2020b. Genetic tests by next-generation sequencing in children with developmental delay and/or intellectual disability. *Clinical and Experimental Pediatrics*. 63(6):195–202. DOI: 10.3345/kjp.2019.00808.

Hartman, P., Beckman, K., Silverstein, K., Yohe, S., Schomaker, M., Henzler, C., Onsongo, G., Lam, H.C., et al. 2019. Next generation sequencing for clinical diagnostics: Five year experience of an academic laboratory. *Molecular Genetics and Metabolism Reports*. 19:100464. DOI: 10.1016/j.ymgmr.2019.100464.

Hayeems, R.Z., Dimmock, D., Bick, D., Belmont, J.W., Green, R.C., Lanpher, B., Jobanputra, V., Mendoza, R., et al. 2020. Clinical utility of genomic sequencing: a measurement toolkit. *npj Genomic Medicine*. 5(1):1–11. DOI: 10.1038/s41525-020-00164-7.

Hebron, K.E., Hernandez, E.R. and Yohe, M.E. 2022. The RASopathies: from pathogenetics to therapeutics. *Disease Models and Mechanisms*. 15(2):dmm049107. DOI: 10.1242/dmm.049107.

Hennekam, R.C.M. 2003. Costello syndrome: An overview. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 117C(1):42–48. DOI: 10.1002/ajmg.c.10019.

Heo, Y., Manikandan, G., Ramachandran, A. and Chen, D. 2021. Comprehensive Evaluation of Error-Correction Methodologies for Genome Sequencing Data. In *Bioinformatics*. N. Helder I., Ed. Brisbane (AU): Exon Publications. Available: http://www.ncbi.nlm.nih.gov/books/NBK569557/ [2022, July 10].

Hert, D.G., Fredlake, C.P. and Barron, A.E. 2008. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis*. 29(23):4618–4626. DOI: 10.1002/elps.200800456.

Horak, P., Fröhling, S. and Glimm, H. 2016. Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls. *ESMO Open*. 1(5). DOI: 10.1136/esmoopen-2016-000094.

Hu, T., Chitnis, N., Monos, D. and Dinh, A. 2021. Next-generation sequencing technologies: An overview. *Human Immunology*. 82(11):801–811. DOI: 10.1016/j.humimm.2021.02.012.

Hume, S., Nelson, T.N., Speevak, M., McCready, E., Agatep, R., Feilotter, H., Parboosingh, J., Stavropoulos, D.J., et al. 2019. CCMG practice guideline: laboratory guidelines for next-

generation sequencing. *Journal of Medical Genetics*. 56(12):792–800. DOI: 10.1136/jmedgenet-2019-106152.

Iancu, I.-F., Avila-Fernandez, A., Arteche, A., Trujillo-Tiebas, M.J., Riveiro-Alvarez, R., Almoguera, B., Martin-Merida, I., Del Pozo-Valero, M., et al. 2021. Prioritizing variants of uncertain significance for reclassification using a rule-based algorithm in inherited retinal dystrophies. *npj Genomic Medicine*. 6(1):1–9. DOI: 10.1038/s41525-021-00182-z.

Imbard, A., Pasmant, E., Sabbagh, A., Luscan, A., Soares, M., Goussard, P., Blanché, H., Laurendeau, I., et al. 2015. NF1 single and multi-exons copy number variations in neurofibromatosis type 1. *Journal of Human Genetics*. 60(4):221–224. DOI: 10.1038/jhg.2015.6.

Jafry, M. and Sidbury, R. 2020. RASopathies. *Clinics in Dermatology*. (April, 1). DOI: 10.1016/j.clindermatol.2020.03.010.

Jennings, L.J., Arcila, M.E., Corless, C., Kamel-Reid, S., Lubin, I.M., Pfeifer, J., Temple-Smolkin, R.L., Voelkerding, K.V., et al. 2017a. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of molecular diagnostics: JMD*. 19(3):341–365. DOI: 10.1016/j.jmoldx.2017.01.011.

Jennings, L.J., Arcila, M.E., Corless, C., Kamel-Reid, S., Lubin, I.M., Pfeifer, J., Temple-Smolkin, R.L., Voelkerding, K.V., et al. 2017b. Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics*. 19(3):341–365. DOI: 10.1016/j.jmoldx.2017.01.011.

Jindal, G.A., Goyal, Y., Burdine, R.D., Rauen, K.A. and Shvartsman, S.Y. 2015. RASopathies: unraveling mechanisms with animal models. *Disease Models and Mechanisms*. 8(8):769–782. DOI: 10.1242/dmm.020339.

Johnston, J.J., van der Smagt, J.J., Rosenfeld, J.A., Pagnamenta, A.T., Alswaid, A., Baker, E.H., Blair, E., Borck, G., et al. 2018. Autosomal recessive Noonan syndrome associated with biallelic LZTR1 variants. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. 20(10):1175–1185. DOI: 10.1038/gim.2017.249.

Joseph, L., Cankovic, M., Caughron, S., Chandra, P., Emmadi, R., Hagenkord, J., Hallam, S., Jewell, K.E., 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. *The Journal of Molecular Diagnostics*. 18(5):605–619. DOI: 10.1016/j.jmoldx.2016.05.007.

Kamp, M., Krause, A. and Ramsay, M. 2021. Has translational genomics come of age in Africa? *Human Molecular Genetics*. (ddab180). DOI: 10.1093/hmg/ddab180.

Kang, M. and Lee, Y.-S. 2019. The impact of RASopathy-associated mutations on CNS development in mice and humans. *Molecular Brain*. 12(1):96. DOI: 10.1186/s13041-019-0517-5.

Kanzi, A.M., San, J.E., Chimukangara, B., Wilkinson, E., Fish, M., Ramsuran, V. and de Oliveira, T. 2020. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic

Inheritance. *Frontiers in Genetics*. 11. Available: https://www.frontiersin.org/article/10.3389/fgene.2020.544162 [2022, February 12].

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., et al. 2019. *The mutational constraint spectrum quantified from variation in 141,456 humans*. (Preprint). Genomics. DOI: 10.1101/531210.

Kehrer-Sawatzki, H. and Cooper, D.N. 2012. NF1 Microdeletions and Their Underlying Mutational Mechanisms. In *Neurofibromatosis Type 1: Molecular and Cellular Biology*. M. Upadhyaya and D.N. Cooper, Eds. Berlin, Heidelberg: Springer. 187–209. DOI: 10.1007/978-3-642-32864-0_14.

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. DOI: 10.1007/s00439-021-02363-3.

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. DOI: 10.1007/s00439-017-1766-y.

Kehrer-Sawatzki, H., Bäzner, U., Krämer, J., Lewerenz, J. and Pfeiffer, C. 2022. The NF1 microdeletion syndrome: early genetic diagnosis facilitates the management of a clinically defined disease. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft*. 20(3):273–277. DOI: 10.1111/ddg.14707.

Kingsmore, S.F., Lantos, J.D., Dinwiddie, D.L., Miller, N.A., Soden, S.E., Farrow, E.G. and Saunders, C.J. 2012. Next-generation community genetics for low- and middle-income countries. *Genome Medicine*. 4(3):25. DOI: 10.1186/gm324.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. 46(3):310–315. DOI: 10.1038/ng.2892.

Knight, J.D.R., Qian, B., Baker, D. and Kothary, R. 2007. Conservation, Variability and the Modeling of Active Protein Kinases. *PLoS ONE*. 2(10):e982. DOI: 10.1371/journal.pone.0000982.

Ko, J.M., Kim, J.-M., Kim, G.-H. and Yoo, H.-W. 2008. PTPN11 , SOS1 , KRAS , and RAF1 gene analysis, and genotype–phenotype correlation in Korean patients with Noonan syndrome. *Journal of Human Genetics*. 53(11):999–1006. DOI: 10.1007/s10038-008-0343-6.

Koczkowska, M., Chen, Y., Callens, T., Gomes, A., Sharp, A., Johnson, S., Hsiao, M.-C., Chen, Z., et al. 2018. Genotype-Phenotype Correlation in NF1: Evidence for a More Severe Phenotype Associated with Missense Mutations Affecting NF1 Codons 844–848. *American Journal of Human Genetics*. 102(1):69–87. DOI: 10.1016/j.ajhg.2017.12.001.

Koh, A., Tan, E., Brett, M.S., Lai, A.H.M., Jamuar, S.S., Ng, I. and Tan, E. 2019. The spectrum of genetic variants and phenotypic features of Southeast Asian patients with Noonan syndrome. *Molecular Genetics and Genomic Medicine*. 7(4):e00581. DOI: 10.1002/mgg3.581.

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C.E., Albarca Aguilera, M., Meyer, R. and Massouras, A. 2019. VarSome: the human genomic variant search engine. *Bioinformatics*. 35(11):1978–1980. DOI: 10.1093/bioinformatics/bty897.

Kratz, C.P. and Zenker, M. (in press). Inherited Disorders of the Ras-MAPK Pathway. *Blood*. 132(Supplement 1):SCI-41. DOI: 10.1182/blood-2018-99-109379.

Krause, A. 2019a. New genetic testing technologies: Advantages and limitations. *South African Medical Journal*. 109(4):207–209. DOI: 10.7196/SAMJ.2019.v109i4.13990.

Krause, A. 2019b. New genetic testing technologies: Advantages and limitations. *South African Medical Journal*. 109(4):207–209. DOI: 10.7196/SAMJ.2019.v109i4.13990.

Krenn, M., Tomschik, M., Rath, J., Cetin, H., Grisold, A., Zulehner, G., Milenkovic, I., Stogmann, E., et al. 2020. Genotype-guided diagnostic reassessment after exome sequencing in neuromuscular disorders: experiences with a two-step approach. *European Journal of Neurology*. 27(1):51–61. DOI: 10.1111/ene.14033.

Kromberg, J.G.R., Wessels, T.-M. and Krause, A. 2013. Roles of genetic counselors in South Africa. *Journal of Genetic Counseling*. 22(6):753–761. DOI: 10.1007/s10897-013-9606-2.

Kromberg, J.G.R., Sizer, E.B. and Christianson, A.L. 2013. Genetic services and testing in South Africa. *Journal of Community Genetics*. 4(3):413–423. DOI: 10.1007/s12687-012-0101-5.

Kruszka, P., Porras, A.R., Addissie, Y.A., Moresco, A., Medrano, S., Mok, G.T., Leung, G.K., Tekendo-Ngongang, C., et al. 2017. Noonan Syndrome in Diverse Populations. *American journal of medical genetics. Part A*. 173(9):2323–2334. DOI: 10.1002/ajmg.a.38362.

Kumar, P., Henikoff, S. and Ng, P.C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 4(7):1073–1081. DOI: 10.1038/nprot.2009.86.

Lahens, N.F., Ricciotti, E., Smirnova, O., Toorens, E., Kim, E.J., Baruzzo, G., Hayer, K.E., Ganguly, T., et al. 2017. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics*. 18(1):602. DOI: 10.1186/s12864-017-4011-0.

Lalonde, E., Rentas, S., Lin, F., Dulik, M.C., Skraban, C.M. and Spinner, N.B. 2020. Genomic Diagnosis for Pediatric Disorders: Revolution and Evolution. *Frontiers in Pediatrics*. 8:373. DOI: 10.3389/fped.2020.00373.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., et al. 2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*. 46(Database issue):D1062–D1067. DOI: 10.1093/nar/gkx1153.

Lappalainen, T., Scott, A.J., Brandt, M. and Hall, I.M. 2019. Genomic analysis in the age of human genome sequencing. *Cell*. 177(1):70–84. DOI: 10.1016/j.cell.2019.02.032.

Larralde, M., Abad, M.E., Luna, P.C. and Hoffner, M.V. 2014. Capillary malformation-arteriovenous malformation: a clinical review of 45 patients. *International Journal of Dermatology*. 53(4):458–461. DOI: 10.1111/ijd.12040.

Lee, B.H. and Yoo, H.-W. 2019a. Noonan syndrome and RASopathies: Clinical features, diagnosis and management. *Journal of Genetic Medicine*. 16(1):1–9. DOI: 10.5734/JGM.2019.16.1.1.

Lee, B.H. and Yoo, H.-W. 2019b. Noonan syndrome and RASopathies: Clinical features, diagnosis and management. *Journal of Genetic Medicine*. 16(1):1–9. DOI: 10.5734/JGM.2019.16.1.1.

Lee, B.H., Kim, J.-M., Jin, H.Y., Kim, G.-H., Choi, J.-H. and Yoo, H.-W. 2011. Spectrum of mutations in Noonan syndrome and their correlation with phenotypes. *The Journal of Pediatrics*. 159(6):1029–1035. DOI: 10.1016/j.jpeds.2011.05.024.

Legius, E. and Stevenson, D. 1993. Legius Syndrome. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K. Stephens, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK47312/ [2020, October 06].

Legius, E., Messiaen, L., Wolkenstein, P., Pancza, P., Avery, R.A., Berman, Y., Blakeley, J., Babovic-Vuksanovic, D., et al. 2021. Revised diagnostic criteria for neurofibromatosis type 1 and Legius syndrome: an international consensus recommendation. *Genetics in Medicine*. 23(8):1506–1513. DOI: 10.1038/s41436-021-01170-5.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., et al. 2016a. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536(7616):285–291. DOI: 10.1038/nature19057.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., et al. 2016b. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536(7616):285–291. DOI: 10.1038/nature19057.

Lepri, F., De Luca, A., Stella, L., Rossi, C., Baldassarre, G., Pantaleoni, F., Cordeddu, V., Williams, B.J., 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. DOI: 10.1002/humu.21492.

Lepri, F.R., Scavelli, R., Digilio, M.C., Gnazzo, M., Grotta, S., Dentici, M.L., Pisaneschi, E., Sirleto, P., et al. 2014. Diagnosis of Noonan syndrome and related disorders using target next generation sequencing. *BMC medical genetics*. 15:14. DOI: 10.1186/1471-2350-15-14.

Lewis, M.A., Nolan, L.S., Cadge, B.A., Matthews, L.J., Schulte, B.A., Dubno, J.R., Steel, K.P. and Dawson, S.J. 2018. Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. *BMC Medical Genomics*. 11(1):77. DOI: 10.1186/s12920-018-0395-1.

Liao, P., Satten, G.A. and Hu, Y.-J. 2017. PhredEM: A Phred-Score-Informed Genotype-Calling Approach for Next-Generation Sequencing Studies. *Genetic epidemiology*. 41(5):375–387. DOI: 10.1002/gepi.22048.

Linck, E. and Battey, C.J. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*. 19(3):639–647. DOI: 10.1111/1755-0998.12995.

Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M., Thiruvahindrapuram, B., Merico, D., et al. 2018. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genetics in Medicine*. 20(4):435–443. DOI: 10.1038/gim.2017.119.

Lissewski, C., Chune, V., Pantaleoni, F., De Luca, A., Capri, Y., Brinkmann, J., Lepri, F., Daniele, P., et al. 2021. Variants of SOS2 are a rare cause of Noonan syndrome with particular predisposition for lymphatic complications. *European journal of human genetics: EJHG*. 29(1):51–60. DOI: 10.1038/s41431-020-00708-6.

Liu, Y., Teng, Y., Li, Z., Cui, J., Liang, D. and Wu, L. 2021. Increase in diagnostic yield achieved for 174 whole-exome sequencing cases reanalyzed 1–2 years after initial analysis. *Clinica Chimica Acta*. 523:163–168. DOI: 10.1016/j.cca.2021.09.015.

Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Ávila, J.A., López-Guerrero, J.A. and Aguilar-Quesada, R. 2016. DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. *Biopreservation and Biobanking*. 14(4):264–270. DOI: 10.1089/bio.2015.0064.

Ma, X. 2019. Analysis of error profiles in deep next-generation sequencing data. 15.

Mahmood, K., Jung, C., Philip, G., Georgeson, P., Chung, J., Pope, B.J. and Park, D.J. 2017. Variant effect prediction tools assessed using independent, functional assay-based datasets: implications for discovery and diagnostics. *Human Genomics*. 11(1):10. DOI: 10.1186/s40246-017-0104-8.

Manickam, K., McClain, M.R., Demmer, L.A., Biswas, S., Kearney, H.M., Malinowski, J., Massingham, L.J., Miller, D., et al. 2021. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*. 23(11):2029–2037. DOI: 10.1038/s41436-021-01242-6.

Manley, L.J., Ma, D. and Levine, S.S. 2016. Monitoring Error Rates In Illumina Sequencing. *Journal of Biomolecular Techniques : JBT*. 27(4):125–128. DOI: 10.7171/jbt.16-2704-002.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. 2002. The protein kinase complement of the human genome. *Science (New York, N.Y.)*. 298(5600):1912–1934. DOI: 10.1126/science.1075762.

Mao, B., Chen, S., Chen, X., Yu, X., Zhai, X., Yang, T., Li, L., Wang, Z., et al. 2018. Clinical characteristics and spectrum of NF1 mutations in 12 unrelated Chinese families with neurofibromatosis type 1. *BMC Medical Genetics*. 19(1):101. DOI: 10.1186/s12881-018-0615-8.

Marian, A.J. 2020. Clinical Interpretation and Management of Genetic Variants. *JACC: Basic to Translational Science*. 5(10):1029–1042. DOI: 10.1016/j.jacbts.2020.05.013.

Martin, A.R., Williams, E., Foulger, R.E., Leigh, S., Daugherty, L.C., Niblock, O., Leong, I.U.S., Smith, K.R., et al. 2019. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics*. 51(11):1560–1565. DOI: 10.1038/s41588-019-0528-2.

Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., Race, V., Sistermans, E., et al. 2016. Guidelines for diagnostic next-generation sequencing. *European Journal of Human Genetics*. 24(1):2–5. DOI: 10.1038/ejhg.2015.226.

Maxmen, A. 2020. The next chapter for African genomics. *Nature*. 578(7795):350–354. DOI: 10.1038/d41586-020-00454-1.

May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., et al. 2013. Genetic diversity in black South Africans from Soweto. *BMC Genomics*. 14(1):644. DOI: 10.1186/1471-2164-14-644.

McCourt, C.M., McArt, D.G., Mills, K., Catherwood, M.A., Maxwell, P., Waugh, D.J., Hamilton, P., O'Sullivan, J.M., et al. 2013. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PloS One*. 8(7):e69604. DOI: 10.1371/journal.pone.0069604.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P. and Cunningham, F. 2016a. The Ensembl Variant Effect Predictor. *Genome Biology*. 17(1):122. DOI: 10.1186/s13059-016-0974-4.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P. and Cunningham, F. 2016b. The Ensembl Variant Effect Predictor. *Genome Biology*. 17(1):122. DOI: 10.1186/s13059-016-0974-4.

Mighton, C., Shickh, S., Uleryk, E., Pechlivanoglou, P. and Bombard, Y. 2021. Clinical and psychological outcomes of receiving a variant of uncertain significance from multigene panel testing or genomic sequencing: a systematic review and meta-analysis. *Genetics in Medicine*. 23(1):22–33. DOI: 10.1038/s41436-020-00957-2.

Miller, E.M., Patterson, N.E., Zechmeister, J.M., Bejerano-Sagie, M., Delio, M., Patel, K., Ravi, N., Quispe-Tintaya, W., et al. 2017. Development and validation of a targeted next generation DNA sequencing panel outperforming whole exome sequencing for the identification of clinically relevant genetic variants. *Oncotarget*. 8(60):102033–102045. DOI: 10.18632/oncotarget.22116.

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-De-Luca, A., Millan, F., Pesacreta, D.R., Elloumi, H.Z., Oetjens, M.T., Teigen, C., Wain, K.E., Scuffins, J., et al. 2021. Molecular Diagnostic Yield of Exome Sequencing in Patients With Cerebral Palsy. *JAMA*. 325(5):467–475. DOI: 10.1001/jama.2020.26148.

Motta, M., Fidan, M., Bellacchio, E., Pantaleoni, F., Schneider-Heieck, K., Coppola, S., Borck, G., Salviati, L., et al. 2019. Dominant Noonan syndrome-causing LZTR1 mutations specifically affect the Kelch domain substrate-recognition surface and enhance RAS-MAPK signaling. *Human Molecular Genetics*. 28(6):1007–1022. DOI: 10.1093/hmg/ddy412.

Mu, W., Lu, H.-M., Chen, J., Li, S. and Elliott, A.M. 2016a. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *The Journal of molecular diagnostics: JMD*. 18(6):923–932. DOI: 10.1016/j.jmoldx.2016.07.006.

Mu, W., Lu, H.-M., Chen, J., Li, S. and Elliott, A.M. 2016b. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *The Journal of molecular diagnostics: JMD*. 18(6):923–932. DOI: 10.1016/j.jmoldx.2016.07.006.

Musanabaganwa, C., Mihigo, B., Tumusime, R., Uwanyirigira, M., da Rocha, J., Hayat, M., Govender, M., Buto, P., et al. 2020. Building Skills and Resources for Genomics, Epigenetics, and Bioinformatics Research for Africa: Report of the Joint 11th Conference of the African Society of Human Genetics and 12th H3Africa Consortium, 2018. *The American Journal of Tropical Medicine and Hygiene*. 102(6):1417–1424. DOI: 10.4269/ajtmh.19-0837.

Müskens, J.L.J.M., Kool, R.B., Dulmen, S.A. van and Westert, G.P. 2022. Overuse of diagnostic testing in healthcare: a systematic review. *BMJ Quality and Safety*. 31(1):54–63. DOI: 10.1136/bmjqs-2020-012576.

Muyas, F., Bosio, M., Puig, A., Susak, H., Domènech, L., Escaramis, G., Zapata, L., Demidov, G., et al. 2019. Allele balance bias identifies systematic genotyping errors and false disease associations. *Human Mutation*. 40(1):115–126. DOI: 10.1002/humu.23674.

Nardi, V., Tsuchiya, K.D., Kim, A.S., Bean, L.J.H., Halley, J.G., Long, T.A., Szelinger, S., Vasalos, P., et al. 2021. Next-Generation Sequencing Somatic and Germline Assay Troubleshooting Guide Derived From Proficiency Testing Data. *Archives of Pathology and Laboratory Medicine*. 146(4):451–461. DOI: 10.5858/arpa.2020-0842-CP.

Nava, C., Hanna, N., Michot, C., Pereira, S., Pouvreau, N., Niihori, T., Aoki, Y., Matsubara, Y., et al. 2007a. Cardio-facio-cutaneous and Noonan syndromes due to mutations in the RAS/MAPK signalling pathway: genotype-phenotype relationships and overlap with Costello syndrome. *Journal of Medical Genetics*. 44(12):763–771. DOI: 10.1136/jmg.2007.050450.

Nava, C., Hanna, N., Michot, C., Pereira, S., Pouvreau, N., Niihori, T., Aoki, Y., Matsubara, Y., et al. 2007b. Cardio-facio-cutaneous and Noonan syndromes due to mutations in the RAS/MAPK signalling pathway: genotype-phenotype relationships and overlap with Costello syndrome. *Journal of Medical Genetics*. 44(12):763–771. DOI: 10.1136/jmg.2007.050450.

Nevondwe, P. 2019. Investigating the genetic aetology of three facial dysostoses in South Africa. Thesis. Available: http://wiredspace.wits.ac.za/handle/10539/28976 [2021, September 28].

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., et al. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*. 42(1):30–35. DOI: 10.1038/ng.499.

Nowaczyk, M.J.M., Thompson, B.A., Zeesman, S., Moog, U., Sanchez-Lara, P.A., Magoulas, P.L., Falk, R.E., Hoover-Fong, J.E., et al. 2014. Deletion of MAP2K2/MEK2: a novel mechanism for a RASopathy? *Clinical Genetics*. 85(2):138–146. DOI: 10.1111/cge.12116.

Ouellette, A.C., Mathew, J., Manickaraj, A.K., Manase, G., Zahavich, L., Wilson, J., George, K., Benson, L., et al. 2018. Clinical genetic testing in pediatric cardiomyopathy: Is bigger better? *Clinical Genetics*. 93(1):33–40. DOI: 10.1111/cge.13024.

Pagnamenta, A.T., Kaisaki, P.J., Bennett, F., Burkitt-Wright, E., Martin, H.C., Ferla, M.P., Taylor, J.M., Gompertz, L., et al. 2019. Delineation of dominant and recessive forms of

LZTR1-associated Noonan syndrome. *Clinical Genetics*. 95(6):693–703. DOI: 10.1111/cge.13533.

Palmer, E.E., Sachdev, R., Macintosh, R., Melo, U.S., Mundlos, S., Righetti, S., Kandula, T., Minoche, A.E., et al. 2021. Diagnostic Yield of Whole Genome Sequencing After Nondiagnostic Exome Sequencing or Gene Panel in Developmental and Epileptic Encephalopathies. *Neurology*. 96(13):e1770–e1782. DOI: 10.1212/WNL.0000000000011655.

Pan, B., Kusko, R., Xiao, W., Zheng, Y., Liu, Z., Xiao, C., Sakkiah, S., Guo, W., et al. 2019. Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinformatics*. 20(2):101. DOI: 10.1186/s12859-019-2620-0.

Pandit, B., Sarkozy, A., Pennacchio, L.A., Carta, C., Oishi, K., Martinelli, S., Pogna, E.A., Schackwitz, W., et al. 2007. Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. *Nature Genetics*. 39(8):1007–1012. DOI: 10.1038/ng2073.

Park, K., Seltzer, L.E., Tuttle, E., Mirzaa, G.M. and Paciorkowski, A.R. 2017. PLXNA1 developmental encephalopathy with syndromic features: A case report and review of the literature. *American Journal of Medical Genetics. Part A*. 173(7):1951–1954. DOI: 10.1002/ajmg.a.38236.

Pasmant, E., Sabbagh, A., Spurlock, G., Laurendeau, I., Grillo, E., Hamel, M.-J., Martin, L., Barbarot, S., et al. 2010. NF1 microdeletions in neurofibromatosis type 1: from genotype to phenotype. *Human Mutation*. 31(6):E1506–E1518. DOI: 10.1002/humu.21271.

Pasmant, E., Parfait, B., Luscan, A., Goussard, P., Briand-Suleau, A., Laurendeau, I., Fouveaut, C., Leroy, C., et al. 2015. Neurofibromatosis type 1 molecular diagnosis: what can NGS do for you when you have a large gene with loss of function mutations? *European Journal of Human Genetics*. 23(5):596–601. DOI: 10.1038/ejhg.2014.145.

Patch, C. and Middleton, A. 2018a. Genetic counselling in the era of genomic medicine. *British Medical Bulletin*. 126(1):27–36. DOI: 10.1093/bmb/ldy008.

Patch, C. and Middleton, A. 2018b. Genetic counselling in the era of genomic medicine. *British Medical Bulletin*. 126(1):27–36. DOI: 10.1093/bmb/ldy008.

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):132. DOI: 10.3390/jcm9010132.

Petrackova, A., Vasinek, M., Sedlarikova, L., Dyskova, T., Schneiderova, P., Novosad, T., Papajik, T. and Kriegova, E. 2019a. Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. *Frontiers in Oncology*. 9. Available: https://www.frontiersin.org/article/10.3389/fonc.2019.00851 [2022, March 23].

Petrackova, A., Vasinek, M., Sedlarikova, L., Dyskova, T., Schneiderova, P., Novosad, T., Papajik, T. and Kriegova, E. 2019b. Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. *Frontiers in Oncology*. 9. Available: https://www.frontiersin.org/article/10.3389/fonc.2019.00851 [2022, March 23].

Phillips, K.A., Douglas, M.P., Wordsworth, S., Buchanan, J. and Marshall, D.A. 2021. Availability and funding of clinical genomic sequencing globally. *BMJ Global Health*. 6(2):e004415. DOI: 10.1136/bmjgh-2020-004415.

Pierpont, M.E.M., Magoulas, P.L., Adi, S., Kavamura, M.I., Neri, G., Noonan, J., Pierpont, E.I., Reinker, K., et al. 2014. Cardio-Facio-Cutaneous Syndrome: Clinical Features, Diagnosis, and Management Guidelines. *Pediatrics*. 134(4):e1149–e1162. DOI: 10.1542/peds.2013-3189.

Piotrowski, A., Xie, J., Liu, Y.F., Poplawski, A.B., Gomes, A.R., Madanecki, P., Fu, C., Crowley, M.R., et al. 2014. Germline loss-of-function mutations in LZTR1 predispose to an inherited disorder of multiple schwannomas. *Nature Genetics*. 46(2):182–187. DOI: 10.1038/ng.2855.

Platt, C.D., Zaman, F., Bainter, W., Stafstrom, K., Almutairi, A., Reigle, M., Weeks, S., Geha, R.S., et al. 2021. Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. *Journal of Allergy and Clinical Immunology*. 147(2):723–726. DOI: 10.1016/j.jaci.2020.08.022.

Popejoy, A.B. and Fullerton, S.M. 2016. Genomics is failing on diversity. *Nature News*. 538(7624):161. DOI: 10.1038/538161a.

Posey, J.E. 2019. Genome sequencing and implications for rare disorders. *Orphanet Journal of Rare Diseases*. 14(1):153. DOI: 10.1186/s13023-019-1127-0.

Posey, J.E., Rosenfeld, J.A., James, R.A., Bainbridge, M., Niu, Z., Wang, X., Dhar, S., Wiszniewski, W., et al. 2016. Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genetics in Medicine*. 18(7):678–685. DOI: 10.1038/gim.2015.142.

Pottinger Tess D., Puckelwartz Megan J., Pesce Lorenzo L., Robinson Avery, Kearns Samuel, Pacheco Jennifer A., Rasmussen-Torvik Laura J., Smith Maureen E., et al. 2020. Pathogenic and Uncertain Genetic Variants Have Clinical Cardiac Correlates in Diverse Biobank Participants. *Journal of the American Heart Association*. 9(3):e013808. DOI: 10.1161/JAHA.119.013808.

Pruneri, G., De Braud, F., Sapino, A., Aglietta, M., Vecchione, A., Giusti, R., Marchiò, C., Scarpino, S., et al. 2021. Next-Generation Sequencing in Clinical Practice: Is It a Cost-Saving Alternative to a Single-Gene Testing Approach? *PharmacoEconomics - Open*. 5(2):285–298. DOI: 10.1007/s41669-020-00249-0.

Rabbani, B., Nakaoka, H., Akhondzadeh, S., Tekin, M. and Mahdieh, N. 2016. Next generation sequencing: implications in personalized medicine and pharmacogenomics. *Molecular BioSystems*. 12(6):1818–1830. DOI: 10.1039/C6MB00115G.

Radomski, T.R., Feldman, R., Huang, Y., Sileanu, F.E., Thorpe, C.T., Thorpe, J.M., Fine, M.J. and Gellad, W.F. 2020. Evaluation of Low-Value Diagnostic Testing for 4 Common Conditions in the Veterans Health Administration. *JAMA Network Open*. 3(9):e2016445. DOI: 10.1001/jamanetworkopen.2020.16445.

Ramsay, M. 2022. African genomic data sharing and the struggle for equitable benefit. *Patterns*. 3(1):100412. DOI: 10.1016/j.patter.2021.100412.

Rauen, K.A. 1993. Cardiofaciocutaneous Syndrome. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK1186/ [2021, September 28].

Rauen, K.A. 2013. The RASopathies. *Annual Review of Genomics and Human Genetics*. 14:355–369. DOI: 10.1146/annurev-genom-091212-153523.

Razzaque, M.A., Nishizawa, T., Komoike, Y., Yagi, H., Furutani, M., Amo, R., Kamisago, M., Momma, K., et al. 2007. Germline gain-of-function mutations in RAF1 cause Noonan syndrome. *Nature Genetics*. 39(8):1013–1017. DOI: 10.1038/ng2078.

Rehder, C., Bean, L.J.H., Bick, D., Chao, E., Chung, W., Das, S., O'Daniel, J., Rehm, H., 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: Official Journal of the American College of Medical Genetics*. 23(8):1399–1415. DOI: 10.1038/s41436-021-01139-4.

Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., et al. 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics*. 15(9):733–747. DOI: 10.1038/gim.2013.92.

Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., et al. 2015. ClinGen — The Clinical Genome Resource. *New England Journal of Medicine*. 372(23):2235–2242. DOI: 10.1056/NEJMsr1406261.

Reig, I., Boixeda, P., Fleta, B., Morenoc, C., Gámez, L. and Truchuelo, M. 2011. Neurofibromatosis-Noonan syndrome: Case report and clinicopathogenic review of the Neurofibromatosis-Noonan syndrome and RAS-MAPK pathway. *Dermatology Online Journal*. 17(4). DOI: 10.5070/D329s5592h.

Reinartz, J., Bruyns, E., Lin, J.-Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M. and Woychik, R. 2002. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics and Proteomics*. 1(1):95–104. DOI: 10.1093/bfgp/1.1.95.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., 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: Official Journal of the American College of Medical Genetics*. 17(5):405–424. DOI: 10.1038/gim.2015.30.

Roberts, A.E., Araki, T., Swanson, K.D., Montgomery, K.T., Schiripo, T.A., Joshi, V.A., Li, L., Yassin, Y., et al. 2007. Germline gain-of-function mutations in SOS1 cause Noonan syndrome. *Nature Genetics*. 39(1):70–74. DOI: 10.1038/ng1926.

Roberts, A.E., Allanson, J.E., Tartaglia, M. and Gelb, B.D. 2013. Noonan syndrome. *The Lancet*. 381(9863):333–342. DOI: 10.1016/S0140-6736(12)61023-X.

Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E. and Shay, J.W. 2016. Comparison of DNA Quantification Methods for Next Generation Sequencing. *Scientific Reports*. 6(1):24067. DOI: 10.1038/srep24067.

Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A. and Mesirov, J.P. 2017. Variant Review with the Integrative Genomics Viewer. *Cancer Research*. 77(21):e31–e34. DOI: 10.1158/0008-5472.CAN-17-0337.

Rodrigues, L.O.C., Batista, P.B., Goloni-Bertollo, E.M., de Souza-Costa, D., Eliam, L., Eliam, M., Cunha, K.S.G., Darrigo-Junior, L.G., et al. 2014. Neurofibromatoses: part 1 - diagnosis and differential diagnosis. *Arquivos De Neuro-Psiquiatria*. 72(3):241–250. DOI: 10.1590/0004-282x20130241.

Rodriguez-Viciana, P., Tetsu, O., Tidyman, W.E., Estep, A.L., Conger, B.A., Cruz, M.S., McCormick, F. and Rauen, K.A. 2006. Germline Mutations in Genes Within the MAPK Pathway Cause Cardio-facio-cutaneous Syndrome. *Science*. 311(5765):1287–1290. DOI: 10.1126/science.1124642.

Ross, J.P., Dion, P.A. and Rouleau, G.A. (in press). Exome sequencing in genetic disease: recent advances and considerations. *F1000Research*. 9:F1000 Faculty Rev-336. DOI: 10.12688/f1000research.19444.1.

Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C. and Jaffe, D.B. 2013. Characterizing and measuring bias in sequence data. *Genome Biology*. 14(5):R51. DOI: 10.1186/gb-2013-14-5-r51.

Roy, S., Coldren, C., Karunamurthy, A., Kip, N.S., Klee, E.W., Lincoln, S.E., Leon, A., Pullambhatla, M., et al. 2018. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *The Journal of Molecular Diagnostics*. 20(1):4–27. DOI: 10.1016/j.jmoldx.2017.11.003.

Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.)*. 132:365–386. DOI: 10.1385/1-59259-192-2:365.

Santani, A., Murrell, J., Funke, B., Yu, Z., Hegde, M., Mao, R., Ferreira-Gonzalez, A., Voelkerding, K.V., et al. 2017. Development and Validation of Targeted Next-Generation Sequencing Panels for Detection of Germline Variants in Inherited Diseases. *Archives of Pathology and Laboratory Medicine*. 141(6):787–797. DOI: 10.5858/arpa.2016-0517-RA.

Sarkozy, A., Digilio, M.C. and Dallapiccola, B. 2008. Leopard syndrome. *Orphanet Journal of Rare Diseases*. 3(1):13. DOI: 10.1186/1750-1172-3-13.

Sarkozy, A., Carta, C., Moretti, S., Zampino, G., Digilio, M.C., Pantaleoni, F., Scioletti, A.P., Esposito, G., et al. 2009. Germline BRAF mutations in Noonan, LEOPARD and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. *Human mutation*. 30(4):695–702. DOI: 10.1002/humu.20955.

Sb, N., Eh, T., Pd, R., Sd, F., Aw, B., C, L., T, S., M, W., et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 461(7261). DOI: 10.1038/nature08250.

Schwarz, J.M., Rödelsperger, C., Schuelke, M. and Seelow, D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*. 7(8):575–576. DOI: 10.1038/nmeth0810-575.

Schwarze, K., Buchanan, J., Taylor, J.C. and Wordsworth, S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in Medicine*. 20(10):1122–1130. DOI: 10.1038/gim.2017.247.

Scocchia, A., Kangas-Kontio, T., Irving, M., Hero, M., Saarinen, I., Pelttari, L., Gall, K., Valo, S., et al. 2021. Diagnostic utility of next-generation sequencing-based panel testing in 543 patients with suspected skeletal dysplasia. *Orphanet Journal of Rare Diseases*. 16(1):412. DOI: 10.1186/s13023-021-02025-7.

Seifi, M., Ghasemi, A., Raeisi, S. and Heidarzadeh, S. 2017. Application of Next-generation Sequencing in Clinical Molecular Diagnostics. *Brazilian Archives of Biology and Technology*. 60. DOI: 10.1590/1678-4324-2017160414.

Seo, G.H., Kim, T., Choi, I.H., Park, J., Lee, J., Kim, S., Won, D., Oh, A., et al. 2020. Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clinical Genetics*. 98(6):562–570. DOI: 10.1111/cge.13848.

Seymour, H. 2019. Mutation profiling in South African patients with cornelia de lange syndrome phenotype using targeted next generation sequencing. Thesis. Available: http://wiredspace.wits.ac.za/handle/10539/29922 [2021, September 28].

Shchelochkov, O.A., Patel, A., Weissenberger, G.M., Chinault, A.C., Wiszniewska, J., Fernandes, P.H., Eng, C., Kukolich, M.K., et al. 2008. Duplication of chromosome band 12q24.11q24.23 results in apparent Noonan syndrome. *American Journal of Medical Genetics Part A*. 146A(8):1042–1048. DOI: 10.1002/ajmg.a.32215.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M. and Gaunt, T.R. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*. 34(1):57–65. DOI: 10.1002/humu.22225.

Sifrim, A., Hitz, M.-P., Wilsdon, A., Breckpot, J., Turki, S.H.A., Thienpont, B., McRae, J., Fitzgerald, T.W., et al. 2016. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nature Genetics*. 48(9):1060–1065. DOI: 10.1038/ng.3627.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. and Ponting, C.P. 2014a. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*. 15(2):121–132. DOI: 10.1038/nrg3642.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. and Ponting, C.P. 2014b. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*. 15(2):121–132. DOI: 10.1038/nrg3642.

Singh, R.R. 2020a. Next-Generation Sequencing in High-Sensitive Detection of Mutations in Tumors: Challenges, Advances, and Applications. *The Journal of Molecular Diagnostics*. 22(8):994–1007. DOI: 10.1016/j.jmoldx.2020.04.213.

Singh, R.R. 2020b. Next-Generation Sequencing in High-Sensitive Detection of Mutations in Tumors: Challenges, Advances, and Applications. *The Journal of Molecular Diagnostics*. 22(8):994–1007. DOI: 10.1016/j.jmoldx.2020.04.213.

Sivakumaran, T.A., Husami, A., Kissell, D., Zhang, W., Keddache, M., Black, A.P., Tinkle, B.T., Greinwald, J.H., et al. 2013. Performance evaluation of the next-generation sequencing approach for molecular diagnosis of hereditary hearing loss. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*. 148(6):1007–1016. DOI: 10.1177/0194599813482294.

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. DOI: 10.1002/cpmb.59.

Smith, H.S., Swint, J.M., Lalani, S.R., Yamal, J.-M., de Oliveira Otto, M.C., Castellanos, S., Taylor, A., Lee, B.H., et al. 2019. Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. 21(1):3–16. DOI: 10.1038/s41436-018-0024-6.

Smith, M.J., Isidor, B., Beetz, C., Williams, S.G., Bhaskar, S.S., Richer, W., O'Sullivan, J., Anderson, B., et al. 2015. Mutations in LZTR1 add to the complex heterogeneity of schwannomatosis. *Neurology*. 84(2):141–147. DOI: 10.1212/WNL.0000000000001129.

Song, P., Chen, S.X., Yan, Y.H., Pinto, A., Cheng, L.Y., Dai, P., Patel, A.A. and Zhang, D.Y. 2021. Selective multiplexed enrichment for the detection and quantitation of low-fraction DNA variants via low-depth sequencing. *Nature Biomedical Engineering*. 5(7):690–701. DOI: 10.1038/s41551-021-00713-0.

Stasik, S., Schuster, C., Ortlepp, C., Platzbecker, U., Bornhäuser, M., Schetelig, J., Ehninger, G., Folprecht, G., et al. 2018. An optimized targeted Next-Generation Sequencing approach for sensitive detection of single nucleotide variants. *Biomolecular Detection and Quantification*. 15:6–12. DOI: 10.1016/j.bdq.2017.12.001.

Stawinski, P., Sachidanandam, R., Chojnicka, I. and Płoski, R. 2016. Chapter 2 - Basic Bioinformatic Analyses of NGS Data. In *Clinical Applications for Next-Generation Sequencing*. U. Demkow and R. Płoski, Eds. Boston: Academic Press. 19–37. DOI: 10.1016/B978-0-12-801739-5.00002-7.

Stenner, M. and Rudack, C. 2014. Diseases of the nose and paranasal sinuses in child. *GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery*. 13:Doc10. DOI: 10.3205/cto000113.

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S.T., Abeysinghe, S., Krawczak, M., et al. 2003. Human Gene Mutation Database (HGMD): 2003 update. *Human Mutation*. 21(6):577–581. DOI: 10.1002/humu.10212.

Stevenson, D., Viskochil, D., Rope, A. and Carey, J. 2006. Clinical and molecular aspects of an informative family with neurofibromatosis type 1 and Noonan phenotype. *Clinical Genetics*. 69(3):246–253. DOI: 10.1111/j.1399-0004.2006.00576.x.

Stevenson, D.A., Schill, L., Schoyer, L., Andresen, B.S., Bakker, A., Bayrak-Toydemir, P., Burkitt-Wright, E., Chatfield, K., et al. 2016. The Fourth International Symposium on Genetic

Disorders of the Ras/MAPK pathway. *American Journal of Medical Genetics. Part A*. 170(8):1959–1966. DOI: 10.1002/ajmg.a.37723.

Stoler, N. and Nekrutenko, A. 2021. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics and Bioinformatics*. 3(1):lqab019. DOI: 10.1093/nargab/lqab019.

Strande, N.T., Riggs, E.R., Buchanan, A.H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S.S., Goldstein, J., Ghosh, R., et al. 2017a. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *American Journal of Human Genetics*. 100(6):895–906. DOI: 10.1016/j.ajhg.2017.04.015.

Strande, N.T., Riggs, E.R., Buchanan, A.H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S.S., Goldstein, J., Ghosh, R., et al. 2017b. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *American Journal of Human Genetics*. 100(6):895–906. DOI: 10.1016/j.ajhg.2017.04.015.

Strom, S.P. 2016. Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biology and Medicine*. 13(1):3–11. DOI: 10.28092/j.issn.2095-3941.2016.0004.

Tabernero, L., Aricescu, A.R., Jones, E.Y. and Szedlacsek, S.E. 2008. Protein tyrosine phosphatases: structure–function relationships. *The FEBS Journal*. 275(5):867–882. DOI: 10.1111/j.1742-4658.2008.06251.x.

Tajan, M., Paccoud, R., Branka, S., Edouard, T. and Yart, A. 2018. The RASopathy Family: Consequences of Germline Activation of the RAS/MAPK Pathway. *Endocrine Reviews*. 39(5):676–700. DOI: 10.1210/er.2017-00232.

Tartaglia, M., Kalidas, K., Shaw, A., Song, X., Musat, D.L., van der Burgt, I., Brunner, H.G., Bertola, D.R., et al. 2002. PTPN11 Mutations in Noonan Syndrome: Molecular Spectrum, Genotype-Phenotype Correlation, and Phenotypic Heterogeneity. *The American Journal of Human Genetics*. 70(6):1555–1563. DOI: 10.1086/340847.

Tekendo-Ngongang, C., Agenbag, G., Bope, C.D., Esterhuizen, A.I. and Wonkam, A. 2019a. Noonan Syndrome in South Africa: Clinical and Molecular Profiles. *Frontiers in Genetics*. 10. DOI: 10.3389/fgene.2019.00333.

Tekendo-Ngongang, C., Agenbag, G., Bope, C.D., Esterhuizen, A.I. and Wonkam, A. 2019b. Noonan Syndrome in South Africa: Clinical and Molecular Profiles. *Frontiers in Genetics*. 10. Available: https://www.frontiersin.org/article/10.3389/fgene.2019.00333 [2022, March 09].

Tekola-Ayele, F. and Rotimi, C.N. 2015a. Translational Genomics in Low and Middle Income Countries: Opportunities and Challenges. *Public health genomics*. 18(4):242–247. DOI: 10.1159/000433518.

Tekola-Ayele, F. and Rotimi, C.N. 2015b. Translational Genomics in Low and Middle Income Countries: Opportunities and Challenges. *Public health genomics*. 18(4):242–247. DOI: 10.1159/000433518.

Thiel, C., Wilken, M., Zenker, M., Sticht, H., Fahsold, R., Gusek-Schneider, G.-C. and Rauch, A. 2009. Independent NF1 and PTPN11 mutations in a family with neurofibromatosis-Noonan

syndrome. *American Journal of Medical Genetics Part A*. 149A(6):1263–1267. DOI: 10.1002/ajmg.a.32837.

Thormann, A., Halachev, M., McLaren, W., Moore, D.J., Svinti, V., Campbell, A., Kerr, S.M., Tischkowitz, M., et al. 2019. Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nature communications*. 10(1):1–10.

Tidyman, W.E. and Rauen, K.A. 2009. The RASopathies: Developmental syndromes of Ras/MAPK pathway dysregulation. *Current opinion in genetics and development*. 19(3):230–236. DOI: 10.1016/j.gde.2009.04.001.

Tidyman, W.E. and Rauen, K.A. 2016. Pathogenetics of the RASopathies. *Human Molecular Genetics*. 25(R2):R123. DOI: 10.1093/hmg/ddw191.

Tiwari, R. and Singh, A. 2021. Neurofibromatosis Type 2. *StatPearls*. (August, 11). Available: https://www.statpearls.com/ArticleLibrary/viewarticle/25787 [2021, September 28].

Tollefson, G.A., Schuster, J., Gelin, F., Agudelo, A., Ragavendran, A., Restrepo, I., Stey, P., Padbury, J., et al. 2019. VIVA (VIsualization of VAriants): A VCF File Visualization Tool. *Scientific Reports*. 9(1):12648. DOI: 10.1038/s41598-019-49114-z.

Tranebjærg, L., Barrett, T. and Rendtorff, N.D. 1993. WFS1 Wolfram Syndrome Spectrum Disorder. In *GeneReviews®*. M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K.W. Gripp, G.M. Mirzaa, and A. Amemiya, Eds. Seattle (WA): University of Washington, Seattle. Available: http://www.ncbi.nlm.nih.gov/books/NBK4144/ [2022, February 19].

van Trier, D.C., Rinne, T., Noordam, K., Draaisma, J.M. and van der Burgt, I. 2017. Variable phenotypic expression in a large Noonan syndrome family segregating a novel SOS1 mutation. *American Journal of Medical Genetics. Part A*. 173(11):2968–2972. DOI: 10.1002/ajmg.a.38466.

Trujillano, D., Bertoli-Avella, A.M., Kumar Kandaswamy, K., Weiss, M.E., Köster, J., Marais, A., Paknia, O., Schröder, R., et al. 2017. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *European Journal of Human Genetics*. 25(2):176–182. DOI: 10.1038/ejhg.2016.146.

Uehara, T., Takenouchi, T., Kosaki, R., Kurosawa, K., Mizuno, S. and Kosaki, K. 2018. Redefining the phenotypic spectrum of de novo heterozygous CDK13 variants: Three patients without cardiac defects. *European Journal of Medical Genetics*. 61(5):243–247. DOI: 10.1016/j.ejmg.2017.12.004.

Upadhyaya, M., Huson, S.M., Davies, M., Thomas, N., Chuzhanova, N., Giovannini, S., Evans, D.G., Howard, E., et al. 2007. An Absence of Cutaneous Neurofibromas Associated with a 3-bp Inframe Deletion in Exon 17 of the NF1 Gene (c.2970-2972 delAAT): Evidence of a Clinically Significant NF1 Genotype-Phenotype Correlation. *The American Journal of Human Genetics*. 80(1):140–151. DOI: 10.1086/510781.

Vears, D.F., Sénécal, K., Clarke, A.J., Jackson, L., Laberge, A.M., Lovrecic, L., Piton, A., Van Gassen, K.L.I., et al. 2018. Points to consider for laboratories reporting results from diagnostic genomic sequencing. *European journal of human genetics: EJHG*. 26(1):36–43. DOI: 10.1038/s41431-017-0043-9.

Villani, A., Greer, M.-L.C., Kalish, J.M., Nakagawara, A., Nathanson, K.L., Pajtler, K.W., Pfister, S.M., Walsh, M.F., et al. 2017. Recommendations for Cancer Surveillance in Individuals with RASopathies and Other Rare Genetic Conditions with Increased Cancer Risk. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*. 23(12):e83–e90. DOI: 10.1158/1078-0432.CCR-17-0631.

Viskochil, D., Buchberg, A.M., Xu, G., Cawthon, R.M., Stevens, J., Wolff, R.K., Culver, M., Carey, J.C., et al. 1990. Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell*. 62(1):187–192. DOI: 10.1016/0092-8674(90)90252-a.

Wadapurkar, R.M. and Vyas, R. 2018. Computational analysis of next generation sequencing data and its applications in clinical oncology. *Informatics in Medicine Unlocked*. 11:75–82. DOI: 10.1016/j.imu.2018.05.003.

Wainschtein, P., Jain, D., Zheng, Z., Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D., et al. 2022. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nature Genetics*. 54(3):263–273. DOI: 10.1038/s41588-021-00997-7.

Waksman, G., Kumaran, S. and Lubman, O. 2004. SH2 domains: role, structure and implications for molecular medicine. *Expert Reviews in Molecular Medicine*. 6(3):1–18. DOI: 10.1017/S1462399404007331.

Waldrop, M.A., Pastore, M., Schrader, R., Sites, E., Bartholomew, D., Tsao, C.-Y. and Flanigan, K.M. 2019. Diagnostic Utility of Whole Exome Sequencing in the Neuromuscular Clinic. *Neuropediatrics*. 50(2):96–102. DOI: 10.1055/s-0039-1677734.

Wallace, S.E. and Bean, L.J. 2020. *Educational Materials — Genetic Testing: Current Approaches*. University of Washington, Seattle. Available: https://www.ncbi.nlm.nih.gov/books/NBK279899/ [2021, September 27].

Wang, B., Du, Z., Shan, G., Yan, C., Zhang, V.W. and Li, Z. 2021. Case Report: Two Chinese Infants of Sengers Syndrome Caused by Mutations in AGK Gene. *Frontiers in Pediatrics*. 9. Available: https://www.frontiersin.org/article/10.3389/fped.2021.639687 [2022, February 19].

Wenger, A.M., Guturu, H., Bernstein, J.A. and Bejerano, G. 2017. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. 19(2):209–214. DOI: 10.1038/gim.2016.88.

Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., 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. DOI: 10.1016/S0140-6736(14)61705-0.

Xu, S., Fan, Y., Sun, Y., Wang, L., Gu, X. and Yu, Y. 2017. Targeted/exome sequencing identified mutations in ten Chinese patients diagnosed with Noonan syndrome and related disorders. *BMC Medical Genomics*. 10(1):62. DOI: 10.1186/s12920-017-0298-6.

Xue, Y., Ankala, A., Wilcox, W.R. and Hegde, M.R. 2015. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-

gene, gene panel, or exome/genome sequencing. *Genetics in Medicine*. 17(6):444–451. DOI: 10.1038/gim.2014.122.

Yamamoto, G.L., Aguena, M., Gos, M., Hung, C., Pilch, J., Fahiminiya, S., Abramowicz, A., Cristian, I., et al. 2015. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. *Journal of Medical Genetics*. 52(6):413–421. DOI: 10.1136/jmedgenet-2015-103018.

Yan, Y.H., Chen, S.X., Cheng, L.Y., Rodriguez, A.Y., Tang, R., Cabrera, K. and Zhang, D.Y. 2021. Confirming putative variants at ≤ 5% allele frequency using allele enrichment and Sanger sequencing. *Scientific Reports*. 11(1):11640. DOI: 10.1038/s41598-021-91142-1.

Yang, H. and Wang, K. 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols*. 10(10):1556–1566. DOI: 10.1038/nprot.2015.105.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., et al. 2013. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England Journal of Medicine*. 369(16):1502–1511. DOI: 10.1056/NEJMoa1306555.

Yapijakis, C., Pachis, N. and Voumvourakis, C. 2017. Neurofibromatosis-Noonan Syndrome: A Possible Paradigm of the Combination of Genetic and Epigenetic Factors. *Advances in Experimental Medicine and Biology*. 987:151–159. DOI: 10.1007/978-3-319-57379-3_14.

Yohe, S., Hauge, A., Bunjer, K., Kemmer, T., Bower, M., Schomaker, M., Onsongo, G., Wilson, J., et al. 2015. Clinical Validation of Targeted Next-Generation Sequencing for Inherited Disorders. *Archives of Pathology and Laboratory Medicine*. 139(2):204–210. DOI: 10.5858/arpa.2013-0625-OA.

Young, A.I. 2022. Discovering missing heritability in whole-genome sequencing data. *Nature Genetics*. 54(3):224–226. DOI: 10.1038/s41588-022-01012-3.

Yu, S. and Graf, W.D. 2011. BRAF gene deletion broadens the clinical spectrum neuro-cardio-facial-cutaneous syndromes. *Journal of Child Neurology*. 26(12):1593–1596. DOI: 10.1177/0883073811413830.

Yu, A.C.-S., Yim, A.K.-Y., Chan, A.Y.-Y., Yuen, L.Y.P., Au, W.C., Cheng, T.H.T., Lin, X., Li, J.-W., et al. 2019. A Targeted Gene Panel That Covers Coding, Non-coding and Short Tandem Repeat Regions Improves the Diagnosis of Patients With Neurodegenerative Diseases. *Frontiers in Neuroscience*. 13. Available: https://www.frontiersin.org/article/10.3389/fnins.2019.01324 [2022, March 26].

Zhang, J., Tong, H., Fu, X., Zhang, Y., Liu, J., Cheng, R., Liang, J., Peng, J., et al. 2015. Molecular Characterization of NF1 and Neurofibromatosis Type 1 Genotype-Phenotype Correlations in a Chinese Population. *Scientific Reports*. 5(1):11291. DOI: 10.1038/srep11291.

Zhang, Z., Chen, X., Zhou, R., Yin, H. and Xu, J. 2020. Chinese patient with neurofibromatosis-Noonan syndrome caused by novel heterozygous NF1 exons 1–58 deletion: a case report. *BMC Pediatrics*. 20(1):190. DOI: 10.1186/s12887-020-02102-z.

Zhao, Y., Fang, L.T., Shen, T., Choudhari, S., Talsania, K., Chen, X., Shetty, J., Kriga, Y., et al. 2021. Whole genome and exome sequencing reference datasets from a multi-center and

cross-platform benchmark study. *Scientific Data*. 8(1):296. DOI: 10.1038/s41597-021-01077-5.

Zhou, Y., Tariq, M., He, S., Abdullah, U., Zhang, J. and Baig, S.M. 2020. Whole exome sequencing identified mutations causing hearing loss in five consanguineous Pakistani families. *BMC Medical Genetics*. 21(1):151. DOI: 10.1186/s12881-020-01087-x.