

https://dspace.org/

The Utility of Clinical Exome Sequencing as a First-Tier Diagnostic Tool in Critically Ill Infants in South Africa

Campell, Lisa

2024
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/47277
Campell, Lisa. (2024). The Utility of Clinical Exome Sequencing as a First-Tier Diagnostic Tool in Critically Ill Infants in South Africa [PhD thesis, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/47277

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

THE UTILITY OF CLINICAL EXOME SEQUENCING AS A FIRST-TIER DIAGNOSTIC TOOL IN CRITICALLY ILL INFANTS IN SOUTH AFRICA

Lisa Campbell

Student number: 808632



A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor in Philosophy.

Supervisors: Dr Nadia Carstens (PhD) and Professor Amanda Krause (MBBCh, PhD)

Johannesburg, 2024

DECLARATION

I, Lisa Campbell, 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.



Lisa Campbell (Candidate)

14th day of August 2024 in Johannesburg, South Africa

PRESENTATIONS AND PUBLICATIONS

Publications and presentations arising from this study:

Presentations:

- Campbell, L., Krause, A. and Carstens, N. The utility of clinical exome sequencing as a first-tier diagnostic tool in critically ill infants in South Africa. Poster presentation. Wits Faculty of Health Sciences Research Day 2022. University of the Witwatersrand, Johannesburg, South Africa. 15th September 2022.
- Campbell, L., Krause, A. and Carstens, N. The utility of clinical exome sequencing as a first-tier diagnostic tool in critically ill infants in South Africa. Poster presentation. African Society of Human Genetics Young Investigator Forum. University of Cape Town, Cape Town, South Africa. 21st February 2023.
- Campbell, L., Krause, A. and Carstens, N. The utility of clinical exome sequencing as a first-tier diagnostic tool in critically ill infants in South Africa. Oral presentation.14th International Congress of Human Genetics. Cape Town International Convention Centre, Cape Town, South Africa. 22nd-26th February 2023.
- Campbell, L., Krause, A. and Carstens, N. The utility of clinical exome sequencing in a South African infant cohort. Virtual poster presentation. Genomics of Rare Diseases 2023. 24th-26th April 2023.
- Campbell, L., Krause, A. and Carstens, N. The utility of clinical exome sequencing in a South African infant cohort. Virtual poster presentation. European Society of Human Genetics (ESHG) European Human Genetics Conference 2023. 10th-13th June 2023.

Publications:

- Campbell, L., Fredericks, J., Mathivha, K., Moshesh, P., Coovadia, A., Chirwa, P., Dillon, B., Ghoor, A., Lawrence, D., Nair, L., Mabaso, N., Mokwele, D., Novellie, M., Krause, A. and Carstens, N. (2023). The implementation and utility of clinical exome sequencing in a South African infant cohort. *Front Genet,* 14(1), 1277948. doi:10.3389/fgene.2023.1277948

ABSTRACT

Genetic disorders are significant contributors to infant mortality, morbidity, hospitalisation, and the need for intensive care globally. The identification and diagnosis of genetic disorders in ill infants is challenging due to their indistinct, and often atypical disease presentations. Diagnosis is traditionally driven by a differential clinical diagnosis; however, broad genotype-first approaches are now the recommended strategy in ill infants. The use of NGS-based testing, including gene panels, whole exome sequencing and whole genome sequencing, has successfully been utilised to diagnose genetic disorders in ill infants and has begun to be implemented in global settings; however, representation of low- and middle-income countries is lacking. Local infrastructure, capacity and expertise significantly affect successful implementation in these contexts. We therefore aimed to investigate the utility and implementation of singleton virtual exome sequencing panels in a cohort of 32 ill infants in the South African State healthcare system, providing the first investigations into the use of NGS in a neonatal intensive care unit setting in Africa. Three virtual panels were used to analyse exome sequencing data: a curated panel of genes implicated in neonatal-onset conditions and enriched for variants in South African populations; a ClinVar panel; and the Developmental Disorders Genotype-to-Phenotype (DDG2P) panel. A diagnostic yield of 22% was achieved across the three virtual panels, providing a definitive molecular diagnosis in seven ill infants. These infants experienced changes in their clinical management as a result of this diagnosis, including the initiation of palliative care, familial screening and prenatal testing for future pregnancies. The adaptation of recommended implementation strategies was necessary in the South African context to address shortages in resources, infrastructure and bioinformatics capacity through the use of gene panels instead of whole exome sequencing and the use of locally available technologies; shortages in the trained genetics workforce through the reliance on primary care clinicians for referrals; and through the singleton sequencing approach to address the unavailability of parents for trio-sequencing and the additional cost factors. This pilot study demonstrated the utility of NGS for the diagnosis and management of ill infants in the South African State healthcare system and explored the challenges in implementing NGS technologies in resource-limited settings. The implementation strategy explored through this research provides a baseline from which advanced NGS diagnostic

strategies can be developed and from which scaled up investigations of utility in the South African State setting can commence.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following parties:

Firstly, to my supervisors, Dr Nadia Carstens and Professor Amanda Krause for your knowledge, support, guidance, and encouragement throughout this project. I am exceedingly grateful for the opportunity to have ventured outside of my comfort zone and pursued this research journey. My growth under your guidance has been immeasurable.

To my colleagues and friends at the Division of the Human Genetics for their support and assistance throughout my time at the division. Special thanks to the Clinical and Counselling Unit for their assistance with patient recruitment.

To the funders of this project, the National Research Foundation (NRF), which funded the larger umbrella study which encompassed this project through a Thuthuka Grant, and to the University of the Witwatersrand and NRF who supported me financially through this PhD through a Block Grant.

To the hospital staff at both recruitment sites for their willingness to be involved in this research and to all the families who participated in this study and made this research possible.

Lastly, to my family and friends, especially my mom and dad, Melba and Colin, my sister, Carmen, and, my partner, Blaine. Thank you for tolerating me, being patient with me and loving me through this long process. I love you all immensely. And you all deserve honorary PhDs.

TABLE OF CONTENTS

DECLARATION ii
PRESENTATIONS AND PUBLICATIONS iii
ABSTRACT iv
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS vii
NOMENCLATURE xii
LIST OF FIGURES xiii
LIST OF TABLES xv
1. INTRODUCTION 1
1.1. The Prevalence of Genetic Disorders in Infants 2
1.2. The Challenge of Determining a Genetic Diagnosis in an Infant 4
1.3. The Diagnosis of Genetic Disorders 6
1.3.1. The Phenotype-First Approach 6
1.3.2. The Genotype-First Approach 6
1.3.3. Phenotypes Suggestive of Genetic Disorders 7
1.3.4. Genetic Testing and Newborn Screening 8
1.4. Next-Generation Sequencing (NGS) 10
1.4.1. NGS Gene Panels 12
1.4.1.1. Targeted Gene Panels 13
1.4.1.2. Virtual Gene Panels 14
1.4.2. Whole Exome Sequencing 15
1.4.3. Whole Genome Sequencing 16
1.5. The Use of NGS in Infant Diagnostics 17
1.5.1. The Use of Gene Panels for Infant Diagnostics 18
1.5.2. The Use of WES in Infant Diagnostics 18
1.5.3. The Use of WGS in Infant Diagnostics 19

1.5.4. High Yield Phenotypes in NGS Diagnostics in Infants....20
1.5.5. Rapid NGS Testing in Infants....21
1.5.6. Changes in Clinical Management Arising from NGS Testing....21
1.5.7. The Cost Effectiveness of NGS Testing in Infants....22
1.6. Genetic Services and Access to NGS Testing in Africa....23
1.7. Barriers and Challenges to the Implementation of NGS in Africa....25
1.7.1. Resources, Infrastructure and Funding....26
1.7.2. Genetics Expertise and Formal Training....28
1.7.3. Data Interpretation from Individuals of African Ancestry....29
1.7.4. Language, Family Structure and Cultural Systems....30
1.8. The Study Rationale....32
1.9. Aim and Objectives....33
2. METHODOLOGY....34
2.1. Implementation Strategy....34
2.2. Ethics Clearance....35
2.3. Patient Cohort....35
2.4. Genomic DNA Extraction....37
2.5. Genomic DNA Quantification and Quality Assessment....37
2.5.1. Nanodrop Quantification and Quality Assessment....37
2.5.2. DNA Quality Assessment by Gel Electrophoresis....37
2.5.3. Qubit Quantification....38
2.6. Exome Sequencing....38
2.6.1. DNA Dilution and Library Preparation....39
2.6.2. Library Quantification, Dilution and Pooling....40
2.6.3. Library Templating and Sequencing....40
2.6.4. Data Processing....41
2.7. Variant Analysis....42
2.7.1. Small Nucleotide Variant Analysis....44

2.7.1.1. The Curated NICU/African Gene Panel Filtering Strategy 44
2.7.1.2. The ClinVar Filtering Strategy 46
2.7.1.3. The DDG2P Gene Panel 46
2.7.1.4. Manual SNV Prioritisation 46
2.7.1.5. SNV Classification 48
2.7.2. CNV Screening 48
2.8. Patient Feedback 49
2.9. Statistical Analysis 50
3. RESULTS 51
3.1. Curation of the NICU/African and DDG2P Virtual Panels 51
3.2. Cohort Demographics and Characteristics 53
3.3. Exome Sequencing Quality Metrics 55
3.4. SNV Analysis 58
3.4.1. Participants with Positive Variant Findings 59
3.4.1.1. Participant NE001 61
3.4.1.2. Participant NE021 62
3.4.1.3. Participant NE022 65
3.4.1.4. Participant NE026 66
3.4.1.5. Participant NE029 69
3.4.1.6. Participant NE033 72
3.4.1.7. Participant NE034 74
3.4.2. Participants Without a Causative Variant Identified 77
3.5. CNV Analysis 85
3.5.1. Participant NE006 88
3.5.2. Participant NE017 89
4. DISCUSSION 91
4.1. The Recruited African Infant Cohort 93
4.2. The Overall Diagnostic Yield and Utility 95

4.2.1. High Yield Phenotype Groups 95
4.2.2. Clinical Utility from NGS Testing 96
4.2.2.1. Time-to-a-Result 96
4.2.2.2. Changes in Clinical Management 97
4.2.2.3. Measures of Clinical Utility 98
4.2.2.4. Observed Personal and Clinical Utility in This Cohort 99
4.2.2.5. Clinical Utility in Infants Without a Confirmed Molecular Diagnosis 100
4.3. A Comparison of the Three Filtering Strategies Utilised 100
4.3.1. The NICU/African Panel 101
4.3.2. The ClinVar Filtering Strategy 102
4.3.3. The DDG2P Panel 104
4.4. Outcomes of the CNV Screen 105
4.5. Implementation Challenges in the Establishment of NGS Testing 106
4.5.1. Virtual Gene Panels vs WES and WGS 107
4.5.2. Singleton vs Trio Sequencing 109
4.5.3. Referrals from Neonatologists and Paediatricians 110
4.5.4. African Genetic Data Interpretation 111
4.6. Study Limitations, Recommendations and Future Directions 112
4.7. Concluding Remarks 114
REFERENCES 115
APPENDICES 140
Appendix A. Ethical Clearance 140
Appendix B. Hospital Approval Letters 142
Appendix C. Phenotype Collection Rubric 145
Appendix D. Table of genes included in the NICU/African virtual gene panel 148
Appendix E. Table of additional genes included in the DDG2P virtual gene panel 190
Appendix F. Age of infant cohort at enrolment 195
Appendix G. Run and sample statistics from the 17 exome sequencing runs 196

Appendix H. Statistical Analysis 198
Appendix I. Clinical Description of Participants Without A Causative Variant Identified 199
Appendix J. Published Paper - Campbell et al. (2023) 206
Appendix J: Plagiarism Declaration 219
Appendix K: Turnitin Report 220

NOMENCLATURE

ACMG	American College of Medical Genetics and Genomics
BAI	Binary Alignment Index
BAM	Binary Alignment Map
CF	Cystic Fibrosis
CMA	Chromosomal Microarray
CNV	Copy Number Variant
DDG2P	Developmental Delay Genotype-to-Phenotype
GC	Guanine-Cytosine
HPO	Harmonised Phenotype Ontology
ICU	Intensive Care Unit
IGV	Integrative Genomics Viewer
ISP	Ion Sphere Particle
LMIC	Low- and Middle-Income Country
NGS	Next Generation Sequencing
NICU	Neonatal Intensive Care Unit
NMCH	Nelson Mandela Children’s Hospital
PCR	Polymerase Chain Reaction
PICU	Paediatric Intensive Care Unit
RMMCH	Rahima Moosa Mother and Child Hospital
SDG	Sustainable Development Goal
SNV	Small Nucleotide Variant
TAT	Turn-around-time
UN	United Nations
VCF	Variant Call File
VEP	Variant Effect Predictor

LIST OF FIGURES

Chapter 1

Figure 1.1. The evolution of sequencing technologies. 10
Figure 1.2. African countries face significant challenges in the implementation of genomics medicine into clinical practice.....26

Chapter 2

Figure 2.1. Summarised workflow for exome library preparation using the Ion Ampliseq™ Exome RDY Library preparation kit. 39
Figure 2.2. Summary of the clinical exome analysis workflow utilised for variant analysis, filtering, prioritization, and interpretation. 43

Chapter 3

Figure 3.1. Gene disease associations and penetrance of the 1127 genes included in the NICU/African virtual panel. 52
Figure 3.2. IGV image depicting the pathogenic *STAC3*:c.851G>C (p.Trp284Ser) variant identified in Participant NE001.....62
Figure 3.3. Pedigree of the family of Participant NE021.....63
Figure 3.4. IGV image depicting the pathogenic *MTM1*:c.664C>T (p.Arg222Ter) variant identified in Participant NE021.....64
Figure 3.5. The predicted effect of the nonsense *MTM1*:c.664C>T (p.Arg222Ter) variant on the *MTM1* protein.....64
Figure 3.6. IGV image depicting the pathogenic *COL2A1*:c.3589G>A (p.Gly1197Ser) variant identified in Participant NE022.....66
Figure 3.7. IGV image depicting the pathogenic *SHOC2*:c.4A>G (p.Ser2Gly) variant identified in Participant NE026.. 69
Figure 3.8. The predicted effect of the nonsense *OCRL:*c.1621C>T (p.Arg514Ter) variant on the *OCRL* protein.....71
Figure 3.9. IGV image depicting the pathogenic *OCRL:*c1621C>T (p.Arg514Ter) variant identified in Participant NE029.....72

Figure 3.10. Sanger sequencing electropherograms confirming the carrier status of the mother of Participant NE029.. 72
Figure 3.11. IGV image depicting the pathogenic *NPHS1:*c.1379G>A (p.Arg460Gln) variant identified in Participant NE033.....74
Figure 3.12. IGV image depicting the pathogenic *TRPV4*:c.806G>A (p.Arg269His) variant identified in Participant NE034.....77
Figure 3.13. High confidence CNVs called in the 32 participants through the Germline Ampliseq™ Exome single sample analysis workflow using the Ampliseq™ exome CNV baseline. 86
Figure 3.14. Genomic location of the approximately 5.8 kb deletion detected in participant NE006.....88
Figure 3.15. Genomic location of the approximately 4.9 Mb deletion identified in participant NE017.....90

LIST OF TABLES

Chapter 2:

Table 2.1. Annotation databases utilised by the IonReporter software. 42
Table 2.2. Filtering parameters utilised during manual filtering and prioritization.....48

Chapter 3:

Table 3.1. Demographics and testing indications of the patient cohort referred for NGS testing. 54
Table 3.2. Summarised average run statistics for the 17 WES runs on the Ion GeneStudio S5 sequencer. 55
Table 3.3. Summarised average coverage statistics for the 32 WES samples. 56
Table 3.4. Variants called per participant using the Ion Reporter™ Germline Ampliseq™ exome single sample analysis workflow. 57
Table 3.5. Diagnostic yields achieved through the three filtering strategies. Demographics, clinical characteristics and diagnostic yields are listed. 59
Table 3.6. Summary table of reportable positive molecular findings in the infant cohort 60
Table 3.7. Shortlisted variants identified in Participant NE026 prioritized through the ClinVar filtering strategy. 69
Table 3.8. Shortlisted variants identified in Participant NE022 prioritized through the NICU/African virtual panel and the ClinVar filtering strategies. 75
Table 3.9. Demographics, clinical indications and the number of variants retained and prioritised across the three filtering strategies amongst participants who did not receive a molecular diagnosis. 79
Table 3.10. Shortlisted variants identified in the 25 participants who did not receive a confirmed molecular diagnosis from SNV filtering and analysis 81
Table 3.11. Potential pathogenic high confidence CNVs identified in Participants NE006 and NE017. 87

Appendices:

Table D.1. List of 1127 genes included in the NICU/African virtual gene panel 148

Table E.1. List of 1704 genes included in the DDG2P virtual gene panel that are not in the NICU/African virtual gene panel. 190

Table F.1. Age of participants in months at enrolment 195

Table G.1. Run statistics from the 17 exome sequencing runs on the Ion GeneStudio S5 sequencer. 196
Table G.2. Coverage statistics for the 32 clinical exome sequencing samples. 197

Table H.1. Fisher's Exact Test examining the influence of cohort characteristics (covariates) on positive diagnostic yield. 198
Table H.2. Logistic regression examining the relationship between positive diagnosis and age at enrolment in months. 198

1. INTRODUCTION

Single-gene, or Mendelian, conditions are considered to be a leading cause of infant and early childhood mortality and hospitalisations globally (Kingsmore et al., 2020). These conditions are estimated to affect 1% of newborns and to be responsible for 21% of infant deaths in developed nations (Wojcik et al., 2019, Hagen et al., 2022). Given the estimated prevalence of these conditions, genetic disorders have become a significant health priority, particularly in infants and children. Understanding the contribution of genetic diseases to infant morbidity and the spectrum of conditions affecting these vulnerable populations is vital in developing strategies to reduce mortality and morbidity in infant populations.

Infants pose a significant challenge in genetic disease diagnostics as they often present with non-specific phenotypes and progress rapidly (Wilkinson et al., 2016, Wright et al., 2018a). Next-generation sequencing (NGS) has been successfully utilised in many international settings to successfully diagnose ill infants, which in turn has resulted in improved clinical management of infants and their families, improved patient outcomes and reduced mortality and morbidity (Clark et al., 2018). NGS targeted gene panels provide a strategy for genetic testing in infants that minimises the cost and analysis time, a key consideration for the implementation of genetic testing in low-resource settings.

Despite global efforts to unmask the burden of genetic diseases in various infant populations, there is a paucity of data from low- and middle- income countries (LMICs), in which the genetic and rare disease burden is still largely unexplored. This has exacerbated healthcare disparities in LMICs, making the access to personalised healthcare and management difficult. A number of barriers to the implementation of NGS in LMICs have been identified, including resource and personnel scarcity and the complexity of data interpretation due to the absence of representative reference data (Kamp et al., 2021). To reduce the healthcare gap experienced by infant populations in LMICs affected by Mendelian conditions, implementation studies which explore the disease burden, aetiology, diagnostic utility and implementation challenges of NGS in these settings are needed.

The following chapter explores the global prevalence of genetic conditions amongst infant populations, the challenges observed in diagnosing these conditions and the

technologies utilised to achieve a diagnosis. The current state of genetic diagnostics in South Africa and in Africa are discussed, as well as the barriers and challenges to implementing NGS and genomic medicine in LMICs. This provides evidence for the necessity of implementation studies in diverse populations, particularly those of African ancestry, which are inadequately represented in global databases and literature and in which growing healthcare inequalities exist.

1.1. The Prevalence of Genetic Disorders in Infants

Genetic disorders are significant contributors to infant hospitalisations, morbidity and mortality globally, with approximately 21% of infant deaths in the developed world estimated to be the result of a genetic condition (Kingsmore et al., 2020). These include both chromosomal abnormalities and monogenic conditions. The real contribution of genetic conditions to infant mortality however may be significantly higher, as mortality statistics are inferred from immediate and underlying causes of death from death certificates, which may not appropriately capture the involvement of underlying genetic conditions. A study at a large paediatric medical centre in the United States found that 10% of infants who demised within a year of life had a diagnosed genetic condition which was not reported as a factor in their death. This was owing to the misclassification of their deaths with no attribution to congenital abnormalities, deformations and chromosomal abnormalities (Wojcik et al., 2021a). The investigators reported an association between misclassification of cause of death and the diagnosis of a monogenic condition. An earlier study by the same group, confirmed a molecular diagnosis in 22% of infants who died within a year of life and in 14% of participants in the infant cohort who did not present with a major congenital anomaly (Wojcik et al., 2019). Another recent US study reported a genetic disease prevalence of 15% amongst all infants at a paediatric hospital who died within a year of life, with an alarming 64% of the genetic diagnoses not disclosed on the death certificate (Owen et al., 2023). These studies highlight the possible underrepresentation of monogenic disorders in infant mortality statistics, in which their contribution is likely higher than the reported 14-22%.

Genetic and monogenic conditions contribute not only to the hospitalisation of infants and children, but also to the need for intensive care and specialist treatment (Wojcik et al., 2019). The prevalence of genetic conditions among infants and children has been reported to be between 16% and 22% in intensive care units (ICUs) in the United States (Ceyhan-Birsoy et al., 2019, Wojcik et al., 2019). The need for intensive care

treatment and overall longer hospital stays for infants who present with clinical features suggestive of a genetic condition, even if undiagnosed, are supported by findings by Schroeder et al. (2021). These authors observed higher neonatal ICU (NICU) and paediatric ICU (PICU) utilisation by infants who presented with features of genetic diseases, regardless of the specificity of these clinical features. Infants with broad clinical features suggestive of a genetic condition, those with commonly accepted suggestive features (multiple congenital anomalies, developmental delay and seizure), and those with less common features suggestive of a genetic condition who remained undiagnosed after first-line genetic testing, had 16 to 50 fold higher intensive care utilisation (Schroeder et al., 2021). Many of the leading causes of hospitalisation and death in NICU and PICU settings have associations with genetic disease, including congenital malformations, low birth weight, prematurity, sudden infant death syndrome, newborn sepsis, circulatory disorders, neonatal haemorrhage and metabolic abnormalities (Wojcik et al., 2019, Kingsmore et al., 2020). This highlights the significant role genetic conditions play in infant and paediatric morbidity and illness.

As diagnostic technologies continue to advance and become more readily available, the identification of genetic conditions within NICUs and PICUs is expected to increase and provide a better understanding of the global prevalence of these conditions. Over a four year period, in a study by Hagen et al. (2022), a 3% increase in the number of diagnosed infants in the NICU was reported with increasing access to broader, more advanced testing technologies, despite almost no change in the number of genetic tests requested during the same period (Hagen et al., 2022). This demonstrated the need for the availability and accessibility to these advancing technologies to accurately document the prevalence of genetic conditions among infants admitted to ICUs.

An understanding of the causes of infant and child death and hospitalisations is necessary to inform healthcare policies, surveillance, and intervention, to improve patient outcomes and to reduce mortality in infant populations. Considering the estimated prevalence of genetic diseases in infants and young children, strategies and frameworks to diagnose and manage these conditions are necessary to meet the Sustainable Development Goal (SDG) to reduce preventable deaths in neonates and children under five by 2030 set out by the United Nations (UN) (Collins et al., 2018). The SDGs aim to promote well-being and good overall health for individuals of all ages. A significant milestone in this endeavour is to reduce neonatal mortality to less than 12 in every 1000 births and to reduce under-five mortality to less than 25 in every 1000

births. Significant improvements in neonatal and under-five mortality have been observed globally, with the average number of neonatal and early-childhood deaths reduced by 10% between 2015 and 2021, however, sub-Saharan Africa remains a concern as the region with the highest mortality rate in both categories (United Nations, 2023). The prevalence of genetic conditions among infants, and their contribution to infant mortality, has not been thoroughly investigated in sub-Saharan Africa. As genetic conditions typically manifest early in childhood, and often in the neonatal period, understanding the prevalence of these conditions is critical in developing healthcare strategies and public policies to help achieve the SDG mortality targets by 2030, particularly in sub-Saharan Africa.

1.2. The Challenge of Determining a Genetic Diagnosis in an Infant

A key challenge in the management of genetic conditions is accurate and early identification. The diagnosis of genetic conditions is particularly challenging in infants as they often present with atypical or non-specific phenotypes, making targeted testing for a particular gene or condition difficult. For example, infantile epileptic encephalopathy, a rare condition which presents as recurrent, intractable seizures during the first few months of life, has been associated with variants in more than 50 genes, making single-gene testing in infants who present with these seizures impractical (Gursoy and Ercal, 2016). Many case reports of atypical disease manifestations have illustrated the difficulties in diagnosing infants with clinical presentations which deviate from the classical, known phenotypes associated with genetic conditions. A case report by Sanford et al. (2020), describes an infant who presented with vomiting, diarrhoea, fatigue, malignant hypertension and severe hyponatremia and in whom a clinical diagnosis of acute kidney injury with haemolytic uremic syndrome and acute glomerulonephritis was suspected. Genomic sequencing revealed a mutation in the *WT1* gene, confirming a diagnosis of Nephrotic syndrome type 4 (NPHS4). The patient did not present with oedema and hypoalbuminemia, the characteristic features of NPHS4, and the treatment and management of this patient was significantly altered as a result of this unexpected diagnosis (Sanford et al., 2020). A recent case study by Tajan et al. (2024) reported on an infant with severe anaemia suggestive of myelodysplastic syndrome. Targeted genetic testing through a leukaemia myelodysplastic syndrome and bone marrow failure panel however was negative. Comprehensive genomic sequencing revealed a mitochondrial DNA deletion conferring a diagnosis of Pearson syndrome in the patient, despite no signs of

pancreatic dysfunction (Tajan et al., 2024). Atypical presentations may skew and prolong the diagnostic process, which may have significant health consequences for the affected infant, particularly when limited testing is available.

Phenotypic presentations in infants often have overlapping features from multiple disorders, making a definitive clinical diagnosis challenging to determine and necessitating the need for genetic confirmation. Wojcik et al. (2019) described an infant with a suspected diagnosis of CHARGE syndrome in whom a diagnosis of Kabuki syndrome was determined from genomic sequencing. Significant phenotypic overlap has been reported between CHARGE syndrome and Kabuki syndrome, particularly during infancy when the distinguishing facial features of Kabuki syndrome are not yet apparent (Butcher et al., 2017). These overlapping features could have led to an incorrect clinical diagnosis and the request of an inappropriate targeted genetic test to confirm the diagnosis.

Many common phenotypic markers of genetic disease, including developmental and intellectual disability and delay, cannot be measured during the neonatal period, and phenotypes may not be fully apparent in infants. This may delay diagnosis by conventional testing methods, which can significantly affect medical management and overall patient outcomes. Specialists with experience in identifying the early, non-specific symptoms of a genetic condition and who can request the relevant investigations are needed (Wilkinson et al., 2016). Disease progression in infants is often very rapid, resulting in discharge or demise of infants before a cause can be identified (Saunders et al., 2012). The presence of comorbidities, including prematurity, birth trauma, sepsis and infection, which may obscure disease presentation, are another significant obstacle in the recognition of a genetic condition in ill infants (Clark et al., 2018).

An early definitive diagnosis of a genetic condition can improve an infant’s outcomes and allow for more personalised management of patients and their families. Early diagnosis can shorten a potential diagnostic odyssey, facilitate the initiation and redirection of treatment in a more targeted approach, allow for increased surveillance, initiate comfort and palliative care when necessary, minimise costly repeat, invasive and unnecessary testing, and instigate counselling for reproductive planning and testing for family members (Rabbani et al., 2012, Splinter et al., 2018, Wright et al., 2018a, Lunke et al., 2020). Infants with a definitive diagnosis may also benefit from more accurate risk assessments and prognoses, referrals to medical specialists,

support from support and advocacy organisations, government welfare, as well as access to clinical trials for emerging therapies. These changes in management create clinical and diagnostic utility and may facilitate early intervention, improved outcomes for infants, and overall decreased morbidity and mortality.

1.3. The Diagnosis of Genetic Disorders

1.3.1. The Phenotype-First Approach

The diagnosis of genetic disorders is traditionally guided by a differential clinical diagnosis from a medical geneticist (Petrikin et al., 2015). This phenotype-first approach has been shown to be successful in infants, producing high diagnostic yields of more than 50% in some cohorts (Gubbels et al., 2020). This approach allows for targeted testing and targeted analysis, which can reduce costs and accelerate the time to a result, a significant advantage in the diagnosis of ill infants where rapid diagnosis and personalised intervention can be lifesaving. Phenotype-driven testing may include chromosome analysis, single-gene sequencing and disease-specific gene panels (Miller et al., 2010, Petrikin et al., 2015). The phenotype-first approach however, is reliant on strong phenotype-genotype associations, limited phenotypic heterogeneity (clinical manifestation of the disease does not vary significantly from patient to patient) and limited genotypic heterogeneity (the resulting manifestation is the result of a small number of known variants within a known restricted subset of genes) (Xue et al., 2015). The phenotype-first approach is also dependent on clinical expertise to recognise the suggestive features of a suspected genetic condition, a considerable challenge in infants who may present with mild and non-classical disease manifestations and in settings where genetics expertise is scarce (Wenger et al., 2021). As a result, many phenotype-guided tests have inconclusive findings and many patients remain undiagnosed (Chong et al., 2015).

1.3.2. The Genotype-First Approach

A genotype-first approach, which examines genomic variation regardless of phenotype, has grown in popularity as a first-line approach for genetic testing in infants due to its more comprehensive nature (Kingsmore et al., 2019). This approach is not limited by known phenotype-genotype associations or restricted to a small subset of condition specific genes. The genotype-first approach is particularly useful in instances of clinical heterogeneity, where disease presentation may be atypical or have varying

severity, as is often observed in infants, and for conditions which are not fully penetrant (Wenger et al., 2021). An infant cohort study which evaluated the utility of exome sequencing in infants with hospital admissions within the first 100 days of life, reported that almost 40% of infants who received a diagnosis from whole exome sequencing (WES) presented with atypical features of their given condition and would have been inappropriately tested despite evaluation by qualified medical geneticists (Meng et al., 2017). A genotype-driven approach is also ideal for genotype-phenotype discovery as it allows for new phenotype associations to be determined, for current phenotypic associations to be expanded and for a better understanding of the mechanisms of disease (Wilczewski et al., 2023). Mercer et al. (2017) described a family with familial Ebstein anomaly and mild skeletal phenotypes, in whom a novel variant in the *FLNA* gene was identified through genotype-driven NGS. The identification of this variant expanded the known phenotype association of this gene with congenital heart defects, as familial Ebstein anomaly had not been previously reported to have an association with this gene and had not been frequently reported to segregate within families (Mercer et al., 2017).

1.3.3. Phenotypes Suggestive of Genetic Disorders

For both genotype- and phenotype-first approaches, a suspicion of a genetic condition is needed to initiate a diagnostic genetic investigation. A number of phenotypic features are frequently observed in infants and are suggestive of an underlying genetic condition, including congenital abnormalities, developmental delay and dysmorphisms (Suzuki et al., 2022). A study by Hagen et al. (2022) investigating the prevalence of genetic disorders in NICUs proposes a broader range of phenotypic symptoms, extending beyond visible congenital anomalies and dysmorphisms, which may warrant the initiation of a genetic investigation in ill Infants. These include abnormal growth, abnormal tone, skeletal abnormalities, organ dysfunction, seizures, encephalopathy, metabolic abnormalities, developmental delay, and unexplained respiratory, dermatological and haematological findings (Hagen et al., 2022). A family history of a genetic condition, as well as consanguinity, are also considered risk factors for genetic disease and may instigate a diagnostic investigation.

The introduction of more comprehensive genotype-driven approaches, including exome- and genome-wide sequencing and microarrays, has improved the diagnostic rate of many genetic conditions. These comprehensive, genotype-first approaches allow for more genes and genomic regions to be simultaneously examined and

investigated. This decreases the time to a diagnosis by eliminating the need for multiple single-gene tests, minimises testing costs, and allows for earlier personalised intervention (Hayeems et al., 2017, Stark et al., 2017, Tan et al., 2017, Dimmock et al., 2021). These benefits are not restricted to a particular category of conditions, as improved diagnostic yields and outcomes have been reported for those with congenital anomalies, developmental and intellectual disability, inborn errors of metabolism, hearing loss, inherited ophthalmic diseases, epilepsy, neurogenetic conditions, skeletal dysplasia, cardiovascular diseases, renal diseases and those with primary immunodeficiencies (Lee, 2023).

An evidence-based systematic review by the American College of Medical Genetics and Genomics (ACMG) determined the benefit of genome and exome sequencing to far outweigh the potential risk and harm for those with congenital anomalies, intellectual disability, and developmental delay as it provides key information which affects and directs patient and family management. The accompanying meta-analysis showed an 8% increase in short-term changes in clinical management, 10% increase in long-term changes in clinical management, 9% change in reproductive outcomes, 4% change in family focused outcomes and an overall 17% higher diagnostic yield for families who underwent genome-wide sequencing compared to standard genetic testing. This has resulted in genome and exome sequencing being recommended by the ACMG as a first- and second-tier test for infants with congenital anomalies under the age of one and in individuals under the age of eighteen with intellectual disabilities and developmental delay (Manickam et al., 2021). Similar recommendations have also been made for the use of exome-sequencing as a first-tier test for patients who present with neurodevelopmental disorders, including autism spectrum disorder (Srivastava et al., 2019). Considering the high prevalence of congenital anomalies and neurogenetic phenotypes amongst hospitalised infants and children, these recommendations could increase the rate of detection of genetic conditions amongst ill infants and provide valuable guidance for medical management.

1.3.4. Genetic Testing and Newborn Screening

Comprehensive genetic testing also offers the opportunity for expanded newborn screening, to diagnose more genetic conditions earlier in life and potentially initiate treatment prior to the onset of severe symptoms (Woerner et al., 2021). Newborn screening has traditionally been performed on dried blood spots using tandem mass spectrometry and is reliant on the detection of metabolic markers (Chace et al., 2002).

This inherently limits the number and type of conditions these screens can detect. The development of protocols to extract DNA from dried blood spots allowed for targeted genetic screening at a first- and second-tier level and the potential expansion of newborn screening to include known genetic conditions, including cystic fibrosis (CF), severe combined immunodeficiency and spinal muscular atrophy (Kwan et al., 2014, Mak et al., 2016, Kucera et al., 2021).

The number of conditions included in newborn screening panels varies from country-to-country, and even within countries, with a larger panel of 35 conditions standardised for use in the United States and a smaller panel of only nine conditions currently in use in the United Kingdom (UK Newborn Screening Programme Centre, 2012, Kellar-Guenther et al., 2020). Comprehensive genetic testing as a method of screening and diagnosis for conditions with known genetic cause in healthy and ill infants has been investigated in a number of studies worldwide, exploring the potential benefits and risks of gene panels, exome and genome sequencing (Ceyhan-Birsoy et al., 2019, Pereira et al., 2019, Adhikari et al., 2020, Roman et al., 2020, Wojcik et al., 2021b, Veldman et al., 2022). The BabySeq project, a randomised trial of genomic sequencing in healthy and ill infants, identified a risk of a childhood onset condition in 9.4% of infants admitted to the NICU, evidencing the diagnostic benefit of NGS for newborn screening (Ceyhan-Birsoy et al., 2019). A recent study by Chen et al. (2023) examining the use of gene panels for comprehensive newborn screening identified a risk for a severe condition in 813 infants in a cohort of 29 601 newborns from the general population. This study provided further evidence of the utility of comprehensive screening through NGS, with 7% of the positive screens not detected by biochemical newborn screening methods (Chen et al., 2023b).

Initiatives to pilot comprehensive newborn screening through genomic sequencing into existing healthcare systems are underway in the United States, United Kingdom and Australia through the BeginNGS programme, Newborn Genomes Programme and BabyScreen+ study (Kingsmore et al., 2022, Pichini et al., 2022, Lunke, 2024). These projects aim to explore the implementation of comprehensive genomic screening alongside traditional newborn screening. Despite the benefits of early diagnosis in ill infants demonstrated in research studies to date, a comprehensive risk-benefit analysis is still needed to evaluate the potential benefit and harm in infants who are clinically well.

1.4. Next-Generation Sequencing (NGS)

NGS is a massively parallel sequencing technology that allows the user to investigate large portions of the genome through the sequencing of millions of DNA fragments simultaneously. This comprehensive technology, termed second-generation sequencing technology, superseded first-generation Sanger sequencing (Figure 1.1.). Sanger sequencing allowed for the sequence of a single DNA fragment to be determined at a time through the termination of an elongating nucleotide chain (Sanger et al., 1977). High-throughput NGS-sequencing utilises either a ligation or synthesis approach to determine the nucleotides of the DNA sequence (Satam et al., 2023). Two of the most commonly utilised platforms are both synthesis-approach platforms: the Illumina fluorescent platform and Ion Torrent Semiconductor platform (Smith et al., 2019).

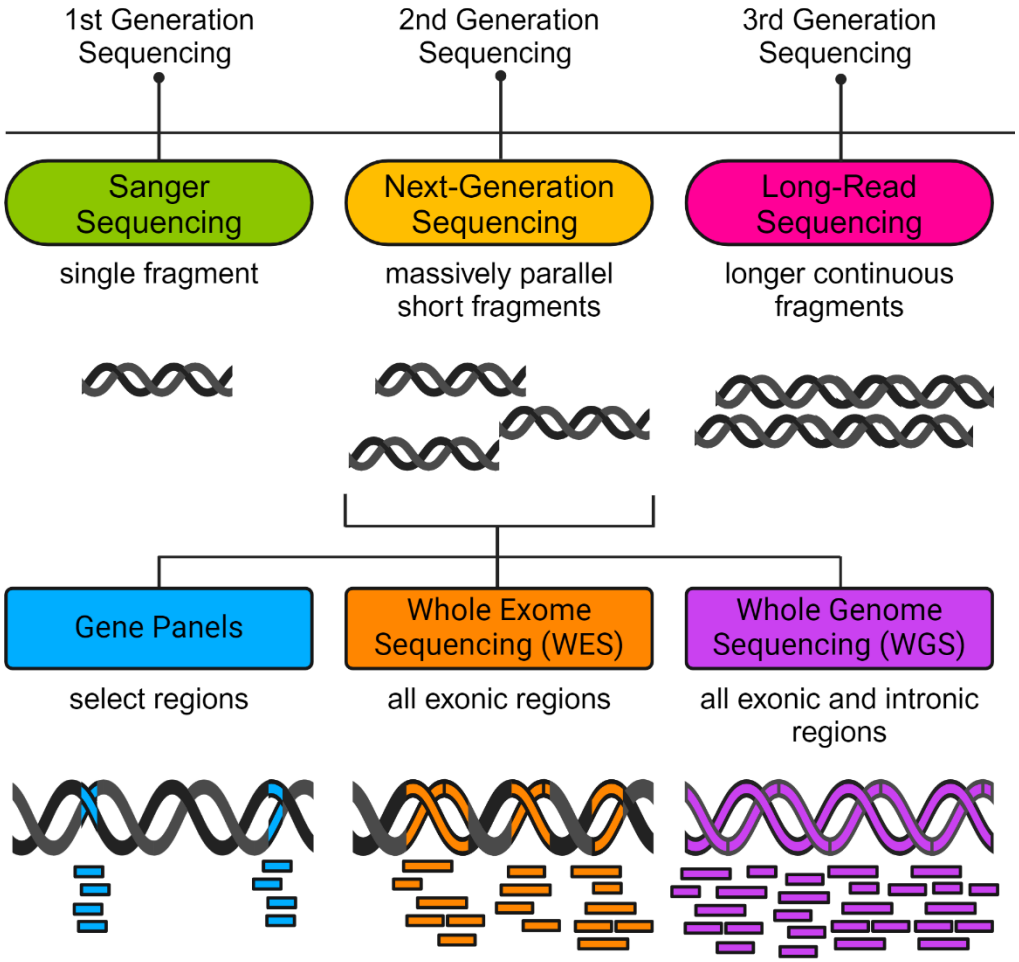


Figure 1.1. The evolution of sequencing technologies. Sanger sequencing, the first-generation sequencing technology, was superseded by second-generation next-generation sequencing (NGS). NGS encompasses three approaches: gene panels, whole exome sequencing (WES) and whole genome sequencing (WGS). Third generation long-read sequencing has in recent years succeeded next-generation sequencing, allowing for longer continuous fragments to be sequenced.

During this sequencing process, the genome is fragmented into 200 – 400 bp reads, ligated to adapters and barcodes and clonally amplified. The incorporation of a base into the synthesising fragment produces a measurable reaction, allowing the incorporated base to be called (Rothberg et al., 2011). The parallel nature of NGS offers a significant advantage over Sanger sequencing, as the sequencing of millions of reads simultaneously is faster, more affordable and more simplistic in its laboratory workflow compared to multiple Sanger sequencing reactions. NGS also allows for the detection of both small nucleotide variants (SNVs) and copy number variants (CNVs) simultaneously, another key advantage to this technology (Zhao et al., 2013).

Sequencing technologies have advanced to include a third-generation approach – long-read sequencing. Long-read technologies allow for the continuous sequencing of larger fragments of DNA, up to several megabases in length. This method of sequencing overcomes some of the shortfalls of short-read sequencing allowing for better base calling and alignment and for better detection of structural variation (Roberts et al., 2013, Goto et al., 2020). Long-read sequencing is more accurate at assembling regions of structural variation, repetitive sequences, regions with a high guanine-cytosine (GC) content, and regions with multiple homologous elements (Treangen and Salzberg, 2011). The phasing of variants is also easier to interpret from long-read data (Delaneau et al., 2013). Long-read sequencing has provided a platform for a more thorough investigation and understanding of the human genome, and through initiatives such as the Telomere-to-Telomere Consortium, provided access to complex regions of the genome not possible to interrogate using second-generation technologies (Nurk et al., 2022). This has significant implications for variant detection and for our understanding of human disease.

A recent study by Viering et al. (2022) demonstrated the clinical utility of long-read sequencing in identify causal variants undetected by short-read sequencing in 45 patients with recessive Gitelman syndrome. The participants had a single pathogenic variant associated with the recessive disease identified through short-read approaches, however, their respective second variants were only resolved through long-read sequencing. The identification of these second variants and the subsequent diagnosis of the participants allowed the researchers to examine the phenotypes and age of onset of disease, enhancing our knowledge of the biochemical manifestations of Gitelman syndrome (Viering et al., 2023).

Despite its advantages over short-read sequencing, long-read sequencing has its challenges. Third generation technologies are dependent on high quality high molecular weight DNA which requires optimised DNA extraction and storage protocols (Oehler et al., 2023). Long read sequencing has a higher error-rate than second-generation sequencing and the improved variant detection is not consistent across all variant types, particularly for small indels (Mahmoud et al., 2024). Despite decreasing costs, long-read technologies are still not affordable and accessible to many LMICs, as many are still challenged by the infrastructure and strategies required for the implementation of short-read technologies (Akoniyon et al., 2022).

NGS technologies are typically characterised by the region sequenced into 3 broad categories: Gene panels, whole exome sequencing (WES), and whole genome sequencing (WGS). Each of these approaches has its own benefits, limitations, and resources which determine its best application.

1.4.1. NGS Gene Panels

NGS gene panels are the least comprehensive NGS testing strategy but minimise the cost and time required for NGS data analysis. Gene panels assess variants in a limited subset of selected genes, typically associated with a specific phenotype or group of conditions. As fewer genes and regions are analysed, incidental findings are minimised through this method compared to comprehensive strategies like WES and WGS (Dillon et al., 2018). The interpretation of variants is often simpler compared to WES and WGS, as only genes of known clinical relevance are included (Boycott et al., 2015, Brunelli et al., 2019). Panels can be performed on more affordable benchtop instruments which are more accessible in resource-limited clinical settings (De Leeneer et al., 2015).

NGS panels can vary significantly in size and design, with the number of genes and the choice of genes included in the panel determined by each user. This variation in size and design can affect their yield, which is highly influenced by the strength of the disease association of the included genes and the specificity of the associated phenotypes. Panel selection is often driven by a clinical hypothesis and benefits from clinical expertise and accurate, comprehensive phenotyping (Tammimies et al., 2015, Lionel et al., 2018). Atypical clinical presentations can skew the selection of genes included in a particular panel or skew the choice of panel requested by a referring clinician. This affects the diagnostic yield and success of a panel and is a key

consideration when choosing an appropriate test for infant populations, in which atypical and indistinct disease presentations are common (Alazami et al., 2015, Saudi Mendeliome, 2015). Two approaches to sequencing panels can be utilised, a targeted wet-lab approach or a virtual bioinformatics approach.

1.4.1.1. Targeted Gene Panels

Targeted NGS panels capture only the targeted genes of interest during sequencing, and not the entire exome or genome. This results in a smaller number of genes being sequenced at a time, allowing for greater sequencing depth per gene and per exon (Lim et al., 2015, Saudi Mendeliome, 2015). These panels can be pre-designed and offered commercially or specifically designed for the users' need. The design and optimisation of individualised panels adds additional time and effort to a workflow, a consideration for successful implementation of these testing strategies. A significant drawback to the use of these panels is that they quickly become outdated and require frequent redesign with the discovery of new gene-disease associations and new variants and genes of interest. This may delay the workflow in routine use and accumulate additional costs (Robertson et al., 2023).

A study of all the gene panels in PanelApp Australia, an open database of gene panels in Australia, found that over a two-year period 121 new panels were added to the database and almost 64 000 changes to the 271 panels on the database were made (Robertson et al., 2023). The continuous discovery of new gene-disease evidence necessitates the revision and update of targeted panels to remain relevant and beneficial.

Targeted panels can vary significantly in size, assessing a few genes associated with a very specific phenotype or be large and expansive, such a Mendeliome panel or clinical exome panel, which assesses all genes of known clinical relevance with a Mendelian disease association (Saudi Mendeliome, 2015, Daoud et al., 2016, Brunelli et al., 2019). Targeted panels of various sizes have been utilised to investigate genetic diseases in infants and children in a number of studies, showing the success of this technique in infant and paediatric diagnostics (Saudi Mendeliome, 2015, Daoud et al., 2016, Dillon et al., 2018, Lionel et al., 2018, Brunelli et al., 2019, Maron et al., 2023).

The Saudi Mendeliome Group utilised a series of 13 curated, broad, symptom-based gene panels, covering a total of 3070 genes with confirmed disease associations to investigate the genetic cause of disease in approximately 2300 individuals of mostly

Arab ethnicity. These capture-based targeted panels varied in size from 69 genes in the renal panel to 758 genes in the neurology panel and produced an overall diagnostic yield of 43% with an analytical sensitivity of 79% (Saudi Mendeliome, 2015). The broad panel design and the redundancy between panels in this study were observed to be an effective strategy to identify causative variants in individuals with atypical disease presentations. The 13 broad panels achieved a molecular yield comparable to that of more comprehensive WES and reduced the interpretation cost significantly in this setting. A comprehensive 4503 gene targeted gene panel by Brunelli et al. (2019) produced a comparable diagnostic yield of 50% in a cohort of 20 ill newborns. These infants benefitted from directed changes in medical management due to the diagnosis, including the initiation of new medication, early tracheostomy and gastrostomy tube placement, follow ups with specialists, and screening and monitoring for the onset of cancers, ureteral obstructions, and Hirschsprung's disease (Brunelli et al., 2019). This panel illustrated the benefit of a comprehensive panel design in being inclusive of the maximum number of disease-associated genes while maintaining a more rapid analytical turn-around-time (TAT) compared to whole exome and genome approaches.

1.4.1.2. Virtual Gene Panels

Virtual gene panels utilise data from WES and WGS to restrict analysis to variants in a subset of genes. This approach, unlike targeted gene panels, does not increase the depth of coverage of genes and exons. This method, however, allows for frequent reanalysis through the expansion and redesign of the gene panel content without additional sequencing and wet-lab costs. This approach can be utilised as a conservative first-tier approach, allowing for simplified screening and interpretation of a set of known disease-causing genes, which can then be expanded to larger panels or comprehensive analysis of whole exomes and genomes if a diagnosis is not yielded (Wang et al., 2018, Quaio et al., 2021, Molina-Ramirez et al., 2022).This conservative approach to first assess genes with the most clinical relevance has been utilised in a number of larger studies, including the Deciphering Developmental Disorders (DDD) Study, which utilised a continually updated virtual panel of genes curated for their role in developmental disorders, termed DDG2P (Wright et al., 2015, Thormann et al., 2019).

The first published report of the DDD study utilised a panel of 1750 genes (which had 2312 known gene-disease associations) and produced a diagnostic yield of 27% in a cohort of 1133 children with developmental phenotypes (Wright et al., 2015). The latest

published report from the study reported a yield of 41% in a larger cohort of 13 449 individuals analysed using updated DDG2P panels, most of which were analysed using a panel of 1840 genes (Wright et al., 2023). Expansion of the original DDG2P panel has considered curated literature and genes observed to be enriched in the study cohort over 10 years. The use of a virtual panel has allowed for periodic updates to be made to the original DDG2P panel without an investment in the optimisation of targeted assays and allowed for subsequent reanalysis of the exome data without additional wet-lab costs. Reanalysis of the initial 1133 participants of the study in 2018 resulted in an increased diagnostic yield of 13% owing to updates in the analysis workflow and the addition of 286 new genes to the DDG2P panel (Wright et al., 2018b).

A number of curated virtual panels have been published in the literature and are available through curated online sources, such as PanelApp England and PanelApp Australia (Martin et al., 2019, Stark et al., 2021). Various commercial laboratories, including Invitae, also make their gene panel lists publicly available, which may provide a starting point for the design of an individualised virtual or targeted panel. Guidelines for panel design and maintenance have recently been developed by the American College of Medical Genetics and Genomics (ACMG) to assist with this process (Bean et al., 2020). Guidelines on establishing gene-disease validity by the Clinician Genome Resource (ClinGen) may also assist in the selection of genes during panel design (Strande et al., 2017). The choice and design of gene panels remains individualised in most settings, making a comparison of individual panels a challenge.

1.4.2. Whole Exome Sequencing

WES examines the protein-coding regions of the genome and their flanking regions, which constitute approximately 2% of the entire genome. The exome is highly conserved and approximately 85% of disease-associated variants are found in this region (Majewski et al., 2011). The examination of only 2% of the genome minimises the TAT and complexity of analysis, the associated cost, and the number of incidental findings and findings of uncertain significance compared to WGS (Wright et al., 2018a). As fewer regions are targeted during sequencing, WES can achieve a greater sequencing depth compared to WGS, while remaining broad enough for the detection of variants associated with non-specific disease aetiologies and conditions in which multiple genes may be implicated. WES has also been shown to be a powerful tool for gene discovery (Ng et al., 2010, Bamshad et al., 2019). It is estimated that 85% of all

Mendelian gene discovery is a result of NGS, particularly WES, demonstrating the utility of this technology in resolving gene-disease associations (Bamshad et al., 2019).

WES can be used to detect SNVs, indels and CNVs, making it a comprehensive tool for diagnostics. SNVs have been the primary focus of WES, as the majority of molecular diagnoses reported from WES, ranging from 65 to 90%, occur as SNVs and indels (Freed et al., 2020, D'Gama et al., 2022, Suzuki et al., 2022). Many pipelines and tools for detecting CNVs have been developed in recent years to address the gaps seen in SNV-only analysis (Zhao et al., 2013). The detection of CNVs and structural variants from WES data is complex, as exons are not uniformly covered, breakpoints are often found outside of the exons and associated flanking regions, and GC-rich regions make quantification of CNVs challenging, resulting in the over- or under-representation of copy regions and false copy number interpretation (Zhao et al., 2013). CNV calling from exome data also falls short for the detection of small single exon CNVs and intergenic CNVs (Royer-Bertrand et al., 2021). As a result of these limitations, confirmatory testing through chromosomal microarray (CMA) is often performed.

The addition of CNV analysis to WES analysis pipelines has been shown to increase their diagnostic yield and provide additional diagnoses for patients. An increase in diagnostic yield of 5% was reported in an infant cohort who underwent genome wide testing by Suzuki et al. (2022). A more substantial increase in the diagnostic yield of approximately 13% was reported by Freed et al. (2020) with the addition of CNV analysis to the analysis strategy in a cohort of ill children undergoing rapid exome sequencing. These studies illustrate the utility of a more comprehensive analysis strategy for WES through the addition of CNV analysis, making WES a promising tool for the diagnosis of genetic conditions (Pfundt et al., 2017, Royer-Bertrand et al., 2021).

1.4.3. Whole Genome Sequencing

WGS examines the entire genome, both coding and non-coding regions. This comprehensive strategy can be used to identify multiple types of variation, including SNVs, CNVs and structural variants, making it a promising technique for clinical diagnostics (Adams and Eng, 2018). WGS offers advantages over other types of NGS testing by its more comprehensive nature, as well as by allowing for more uniform coverage of exonic regions and GC-rich regions, improved detection of CNVs, and the ability to interrogate the mitochondrial genome (Belkadi et al., 2015, Lelieveld et al.,

2015, Alfares et al., 2018). A comparison of four popular WES platforms and WGS found WGS to produce more reproducible overall coverage, outperform all four WES platforms in regions of high GC content and have reads with better mappability compared to WES, a key contributor to overall sequence coverage and variant calling (Barbitoff et al., 2020). The superiority of WGS over WES in the coverage of protein coding regions has been demonstrated in a study by Meienberg et al. (2016). The authors reported 100% coverage of RefSeq coding regions compared to 98% in WES, mainly due to better coverage of GC-rich first exons by WGS despite overall depth of coverage being halved (Meienberg et al., 2016).

WGS also offers a high probability of gene discovery and the identification of new gene-disease associations owing to its more comprehensive nature. These benefits have been shown to increase the diagnostic and clinical utility of WGS compared to WES, gene panels and CMA (Soden et al., 2014, Stavropoulos et al., 2016, Lionel et al., 2018). Despite these advantages, WGS is not widely adopted in many parts of the world for clinical diagnosis. WGS generates a multitude of data which can be costly and time-consuming to analyse and store, complex to analyse, requires skilled bioinformatics expertise and substantial initial capital investment to establish and to run. As a result, less comprehensive sequencing techniques, such as WES and targeted gene panels have been more widely adopted.

1.5. The Use of NGS in Infant Diagnostics

NGS, in the form of genome, exome and panel sequencing, has been utilised in a multitude of studies globally to investigate the genetic cause of disease in infant populations, particularly those under intensive care in NICUs (Saunders et al., 2012, Willig et al., 2015, Daoud et al., 2016, Stark et al., 2016, Meng et al., 2017, van Diemen et al., 2017, Farnaes et al., 2018, Petrikin et al., 2018, Powis et al., 2018, Clark et al., 2019, Elliott et al., 2019, French et al., 2019, Kingsmore et al., 2019, de Castro et al., 2020, Dimmock et al., 2020, Freed et al., 2020, Gubbels et al., 2020, Lunke et al., 2020, Smith et al., 2020, Wang et al., 2020, Dimmock et al., 2021, Imafidon et al., 2021, Krantz et al., 2021, Maron et al., 2021, D'Gama et al., 2022, Denommé-Pichon et al., 2022, Halabi et al., 2022, Maron et al., 2023). These studies have achieved diagnostic yields ranging between 21 and 70% in infant populations, providing definitive molecular diagnoses to thousands of ill infants. The representation of individuals of African ancestry in these studies however is minimal, providing little insight into the utility and implementation in African settings.

1.5.1. The Use of Gene Panels for Infant Diagnostics

One of the largest neonatal gene panel studies to date, the GEMINI study, compared the diagnostic utility of a curated 1722-gene targeted gene panel to rapid WGS in a cohort of 400 ill infants with broad indications for testing (Maron et al., 2023). This panel provided a diagnosis to 109 participants, a yield of 27%, in a TAT of approximately 100 hours. The panel was noted to identify causal variants in 9 participants who were not diagnosed by rapid WGS due to poor gene coverage, computational filtering, and technical interpretation. The utility of NGS diagnostics was evidenced in this study through the 76 participants (19% of the total cohort) who experienced changes in medical management as a result of testing. This included those with a confirmed molecular diagnosis and some participants without. WGS produced a higher diagnostic yield in this cohort compared to targeted gene panels (49% vs 27%), but at the expense of TAT, which was on average 40 hours longer for WGS compared to the targeted gene panel (Maron et al., 2023). This difference in time to a result can have significant effects on medical management options, particularly in critically ill infants.

A similarly sized targeted panel, the NeoSeq panel, produced a diagnostic yield of 39% in a cohort of 33 ill newborns, assessing 1870 genes with known associations to neonatal-onset conditions (de Castro et al., 2020). The success and improved diagnostic performance of this panel, comparable to yields reported from comprehensive WES and WGS, was attributed to the phenotype-guided analysis, which prioritised a subset of the 1870 genes during analysis based on harmonised phenotype ontology (HPO) terms. The enrichment of this study cohort with participants with neurological features, including encephalopathies, seizures and hypotonia, is another considerable factor in the success of this panel, as those with neurological phenotypes have been previously reported to have high diagnostic yields for NGS testing in neonatal cohorts (Suzuki et al., 2022).

1.5.2. The Use of WES in Infant Diagnostics

The success of WES in diagnosing genetic conditions in ill infants and the subsequent utility in clinical decision making has been demonstrated in many research studies. A study of 80 participants in Australia identified a genetic cause behind the clinical presentation of 46 ill infants under the age of two through proband-only WES, a yield of more than 50% (Stark et al., 2016). A third of the diagnosed participants in this study

had changes in their medical management as a result of the diagnosis and more than half of these participants would have gone undiagnosed by standard care.

Advancements in sequencing and analysis protocols have allowed for faster TATs for WES in ill infants, maximising the potential benefits of time-sensitive changes in medical management. Rapid-WES was performed in a cohort of 33 ill infants under a year old admitted to ICUs in a study by Wang et al. (2020) using a rapid workflow to produce test results in 24 hours. This rapid workflow determined a genetic diagnosis in 70% of participants, most of whom presented with metabolic phenotypes. The clinical management of 44% of diagnosed participants was affected by this diagnosis, substantiating the utility of a confirmed molecular diagnosis for directed patient care (Wang et al., 2020).

1.5.3. The Use of WGS in Infant Diagnostics

The use of comprehensive WGS for the diagnosis of genetic conditions in ill infants has been explored extensively over the last 10 years, in research and clinical settings. One of the first cohort studies to investigate the utility of NGS sequencing in ill infants with a suspicion of a genetic condition was by Saunders et al. in 2012, initiating a decade of research into the utility and implementation of these technologies in NICUs and PICUs. The authors identified likely causative variants in four ill infants in a 50-hour WGS assay, resulting in a diagnostic yield of 75% amongst prospective participants (Saunders et al., 2012). An expansion of this study in 2014 achieved a diagnostic rate of 70% in a cohort of 16 ill infants admitted to the NICU (Soden et al., 2014).

Large cohort studies such as the NSIGHT2 study and NICUseq study have in recent years further evidenced the utility of WGS in NICUs and PICUs. The NSIGHT2 randomised controlled trial compared the analytical and diagnostic performance of WGS and WES in 213 critically ill infants in the NICU. This study observed identical diagnostic yields in both WES and WGS (21%), as well as identical TATs (Kingsmore et al., 2019, Dimmock et al., 2021). WGS however had better analytical performance with overall better nucleotide coverage compared to WES, resulting in 121 fold more total variants called, 258 fold more indels called, and 85 fold more rare variants called. The superior coverage and variant calling from WGS in this study emphasises an advantage of WGS in the clinical setting where the detection of rare variants can add

to the growing knowledge of a rare condition, as well as have direct management implications for the affected infant and family.

The NICUSeq randomised time-delayed trial, investigating the utility and timing of WGS in ill infants, reported a diagnostic yield of 31% in a large cohort of 354 ill infants under 120 days old, one of the largest WGS cohort studies in a NICU setting to date (Krantz et al., 2021). The study supported the use of WGS early in the diagnostic trajectory, as infants who received WGS within the first 15 days of enrolment experienced twice the number of medical management changes than those who only received WGS after 60 days. These changes included surgery, invasive procedures and the initiation of specific treatment regimens, and resulted in more directed management and patient care.

Implementation studies of WGS in real world hospital settings have also shown to be successful in diagnosing and directing the management of ill infants with genetic conditions. The FASTGENOMICS pilot study in France is one such study. The pilot project reported a diagnostic yield of 57% in a cohort of 37 newborns in need of urgent genetic testing and explored implementation challenges in bringing WGS into routine diagnostic care. These challenges included the meeting of multidisciplinary teams to discuss patient selection and variant interpretation and transportation delays from multiple hospital sites to a central laboratory (Denommé-Pichon et al., 2022). Despite the proposed diagnostic yield and utility of WGS, many more implementation studies are needed to evaluate the unique challenges in each individual hospital setting.

1.5.4. High Yield Phenotypes in NGS Diagnostics in Infants

The success of a chosen NGS testing method is influenced by patient and phenotype selection. Phenotype-driven approaches to patient selection have shown to produce higher diagnostic yields for NGS tests, however, these are not reflective of the full spectrum of genetic diseases ill infants may present with. These approaches may exclude affected infants with atypical disease presentations and require a level of genetic expertise amongst referring clinicians to identify infants best suited for a particular test (Gubbels et al., 2020, D'Gama et al., 2022). This has promoted the use of more comprehensive genotype-driven testing strategies, including large Mendeliome gene panels, WES and WGS, and the use of more inclusive, broad participant inclusion criteria. Infants with multiple congenital anomalies and neurological phenotypes are most frequently referred for NGS-based genetic testing

and these phenotype groups also exhibit the highest diagnostic rate for genome, exome and panel sequencing (Saunders et al., 2012, Willig et al., 2015, Farnaes et al., 2018, Elliott et al., 2019, D'Gama et al., 2022). Its success as a diagnostic tool has resulted in WGS and WES being the recommended first-tier test for children and infants who present with congenital anomalies, developmental delay, intellectual disability and neurodevelopmental phenotypes (Srivastava et al., 2019, Manickam et al., 2021).

1.5.5. Rapid NGS Testing in Infants

Improvements in data analysis and interpretation, as well as optimisation and automation of many laboratory and analytical steps, have allowed research and diagnostic facilities to deliver comprehensive NGS test results to ill infants in rapid TATs, some in a little as 13.5 hours (Owen et al., 2022). The rapid return of results allows for critical early intervention and in some cases may prolong life (Stark et al., 2019). These rapid workflows however are accompanied by additional costs and are reliant on well-established systems and workflows, which may be beyond the capacity of the genomics workforce and infrastructure in LMIC settings.

1.5.6. Changes in Clinical Management Arising from NGS Testing

Regardless of time-to-result, diagnoses received from genome, exome and panel testing has resulted in significant changes in the clinical management of ill infants, with up to 50% of infants tested having reported changes in their medical management (Willig et al., 2015, Stark et al., 2016, Meng et al., 2017, Farnaes et al., 2018, Petrikin et al., 2018, Elliott et al., 2019, de Castro et al., 2020, Dimmock et al., 2020, Freed et al., 2020, Gubbels et al., 2020, Lunke et al., 2020, Smith et al., 2020, Wang et al., 2020, Dimmock et al., 2021, Krantz et al., 2021, Maron et al., 2021, D'Gama et al., 2022, Denommé-Pichon et al., 2022, Maron et al., 2023).

In a study by Willig et al. (2015), 57% of infants undergoing trio-WGS received a positive molecular diagnosis. Changes in clinical management as a result of NGS-testing were initiated in 85% of diagnosed participants, including medication changes, the use of genetic counselling services and in 30% of diagnosed infants, palliative care was initiated (Willig et al., 2015). Similar changes in management were recorded in diagnosed participants in a cohort of 50 ill infants who underwent rapid-WES in a 2020 study (Gubbels et al., 2020). Amongst the 24 ill infants diagnosed in the NICU and PICU, 83% were referred for further screening and assessment by specialists, 17%

were started on new medication regimes, 38% had comfort care initiated, 13% had supportive procedures such as tracheostomy performed and 33% had familial testing to determine carrier status in family members (Gubbels et al., 2020).

A definitive diagnosis may not only provide an infant with more tailored clinical management and more accurate risk assessments and prognoses, but also access to clinical trials for emerging therapies, support from advocacy and support organisations and assistance from government welfare to relieve the financial burdens of specialised healthcare. In an implementation study by Lumaka et al. (2023), two infants who received a diagnosis through 40 hour rapid-WGS were enrolled in clinical trials as a result of their definitive diagnosis. One of the infants displayed unsatisfactory responses to routine treatment and the enrolment in the clinical trial, due to a conclusive diagnosis, provided a more successful treatment avenue (Lumaka et al., 2023).

Clinical decision making can also be affected in patients who do not receive a definitive diagnosis, through the cancellation of unnecessary medical procedures, cessation of inappropriate testing and the termination of inappropriate treatment regimens. In the study by Gubbels et al. (2020) four infants without a positive molecular diagnosis had further testing cancelled as a result of the negative WGS test. In the WES study by Lunke et al. (2020) 11% of participants who received negative NGS testing results had their management impacted by this result, including two participants who had invasive lung biopsies cancelled and two participants who had unnecessary medication regimens terminated. These cases provide evidence for the significant clinical and personal utility of NGS in the diagnosis of ill infants, regardless of the result received.

1.5.7. The Cost Effectiveness of NGS Testing in Infants

NGS has shown to be a more cost effective diagnostic approach in ill infants compared to traditional diagnostic tests, allowing for a comprehensive first-tier approach that minimises the subsequent testing costs which often accompany standard-of-care testing (Stark et al., 2017, Tan et al., 2017, Incerti et al., 2022). A model-cost analysis in 2022 found the cost of WGS to be equivalent to standard-of-care, however, the comprehensiveness of the genomic approach, the higher diagnostic yield and the higher probability of a diagnosis, resulted in WGS being the more cost-effective strategy in infants because of a lower cost per diagnosis (Incerti et al., 2022).

A cost-analysis involving 40 ill infants in the Australian healthcare system found standard-of-care testing to cost on average AUS$ 4733, however its substantially lower diagnostic yield of 18% resulted in a high cost per diagnosis of approximately AUS$ 27000. In comparison, WES cost AUS$ 3154 and had a cost per diagnosis of AUS$ 5447 due to its high diagnostic yield of more than 60% (Stark et al., 2017). This study found cost-saving to be highest when comprehensive WES was implemented as a first-tier strategy compared to utilisation as a replacement for some standard testing or once standard testing was exhausted, promoting the use of NGS testing early in the diagnostic trajectory (Stark et al., 2017). Identical conclusions were drawn from a study by Tan et al. (2017), who also reported a lower cost for WES compared to stand-of-care of testing, particularly when utilised as a first-tier test (Tan et al., 2017).

A recent cost-model determined WGS to be the most cost-effective strategy for the diagnosis of genetic conditions in paediatric patients compared to first-and second-tier WES and stand-of-care testing, further promoting the use of more comprehensive testing strategies amongst children and infants early in the diagnostic process (Nurchis et al., 2024). The cost-effectiveness of NGS compared to standard-of-care testing is an important consideration in low- and middle-income settings, where resources, both personal and from the State, are severely constrained.

1.6. Genetic Services and Access to NGS Testing in Africa

Despite the benefit of NGS technologies observed globally, the uptake of genome, exome and panel sequencing in many LMICs and African countries is low (Tekola-Ayele and Rotimi, 2015). Most of the published literature surrounding the utility and implementation of NGS testing in infants has focused primarily on populations of European ancestry and in high and upper-middle income countries, with a noticeable absence of data from continental Africans. The scarce uptake of these technologies is not surprising considering the infancy of genomic medicine and genetic services in LMICs and across the African continent, in which significant healthcare disparities still exist. The availability of medical genetics services differs significantly across the African continent. Genetic testing and clinical genetics services are offered in some form in Benin, the Democratic Republic of the Congo, Mali, Rwanda, Senegal, Cameroon, Sudan, Kenya and South Africa, however many testing modalities, including NGS, are offered primarily through research and not clinical service (Tekola-Ayele and Rotimi, 2015, Ndiaye Diallo et al., 2017, Kamp et al., 2021, Lumaka et al.,

2022). This heightens the access disparity for individuals who do not live in urban areas and who are not in close proximity to research institutions (Kamp et al., 2021).

In Ethiopia, a single clinical genetics laboratory with limited services was established in Addis Ababa in 2014. The establishment of the service in the major city makes access for those in rural regions a significant challenge and limits utilisation of genetic services (Quinonez and Terefework, 2021). The laboratory is a private institution and does not offer NGS testing, only multiple ligation probe assays (MLPAs) and real-time polymerase chain reaction (PCR) tests, limiting the available services to the public and making these services inaccessible to many who cannot afford it. Limited genetic services are also available in Rwanda, where karyotyping is the only available test carried out on-site, and where testing for monogenic conditions is only performed through international research collaborations (Uwineza and Mutesa, 2015). Genetic services have been established in Nigeria since the 1970s, offering cytogenetic testing and limited single-gene testing, particularly for sickle cell disease (Adeyemo et al., 2018). In Kenya, the outsourcing of clinical testing to international laboratories is common practice, increasing the range of genetic tests available to the population. Although PCR, Sanger sequencing, FISH, karyotyping and southern blot assays are available in-house in select clinical service laboratories, approximately 50% of all genetic testing is outsourced internationally. This has increased access to additional genetic tests, including methylation assessments, however, the cost and TATs associated with outsourcing is a barrier to the use of these services. (Zhong et al., 2021b). Outsourcing these tests also minimises efforts to grow local capacity, infrastructure and expertise. Updates on the establishment of genetic services in these countries is not well reported, therefore it is assumed minimal increases in services have been introduced in recent years.

Access to primary health care and diagnostic services in many African countries is challenging, exacerbating the difficulties in accessing more specialised services, such as medical genetics, if they are established. South Africa is noted to have one of the most advanced and established clinical genetics services across the continent, however, services in this setting are still very limited (Quinonez and Terefework, 2021, Thom and Haw, 2021). Only three of the nine South African provinces have operational service units in the State healthcare system, which is utilised by more than 80% of the South African population consisting of almost 60 million individuals (Kromberg et al., 2013, Thom and Haw, 2021, Stats SA, 2022).

Alternative health priorities, including the HIV-TB epidemic, the emerging prevalence of non-communicable diseases, poor maternal and child health, injury and violence, (identified as the quadruple burden of disease) and in recent years the COVID-19 pandemic, are prioritised in the South African healthcare system, resulting in many suffering from genetic conditions not accessing the limited available diagnostic services and clinical care (Hofman and Madhi, 2020, Van Der Merwe et al., 2022). Within the State genetic service in the Gauteng province, WES and WGS are not offered under routine diagnostic services, however an NGS panel of 500 genes has recently been introduced and is available at the request of a medical geneticist for a few select cases (Krause, 2023, personal communication).

Most ill infants do not have access to comprehensive genetic testing and the true burden of genetic diseases in the South African and in African populations is unknown. As the developed world shifts toward more rapid testing and precision medicine for genetic conditions, Africa requires significant efforts to bring the continent to the same level of clinical care for individuals affected by genetic conditions and prevent further widening of the healthcare gaps between LMICs and high-income, developed countries (Wonkam et al., 2011, Ramos et al., 2012, Munung et al., 2022). The implementation of NGS testing for ill infants in the South African setting should be informed by studies in the developed world and by international guidelines and recommendations but will require significant optimisation for the local context, which considers the unique challenges experienced in LMICs and across the African continent.

1.7. Barriers and Challenges to the Implementation of NGS in Africa

There are many challenges which need to be investigated and addressed for NGS testing to be implemented in an African State healthcare setting for the benefit of ill infants and children. Studies investigating the utility of genome, exome and panel testing need to consider these challenges for these technologies to be successfully integrated into local healthcare systems and to make genomic medicine achievable and beneficial in LMICs (Tekola-Ayele and Rotimi, 2015). Challenges to the incorporation of genomic medicine into standard healthcare across the African continent vary from-country-to-country due to social, economic, political and population diversity, making a one-size-fits-all approach to implementation implausible and necessitating exploratory studies of utility and implementation in each local setting

(Tiffin, 2014). Some of the key challenges, highlighted in Figure 1.2, will be discussed in the following subsections.

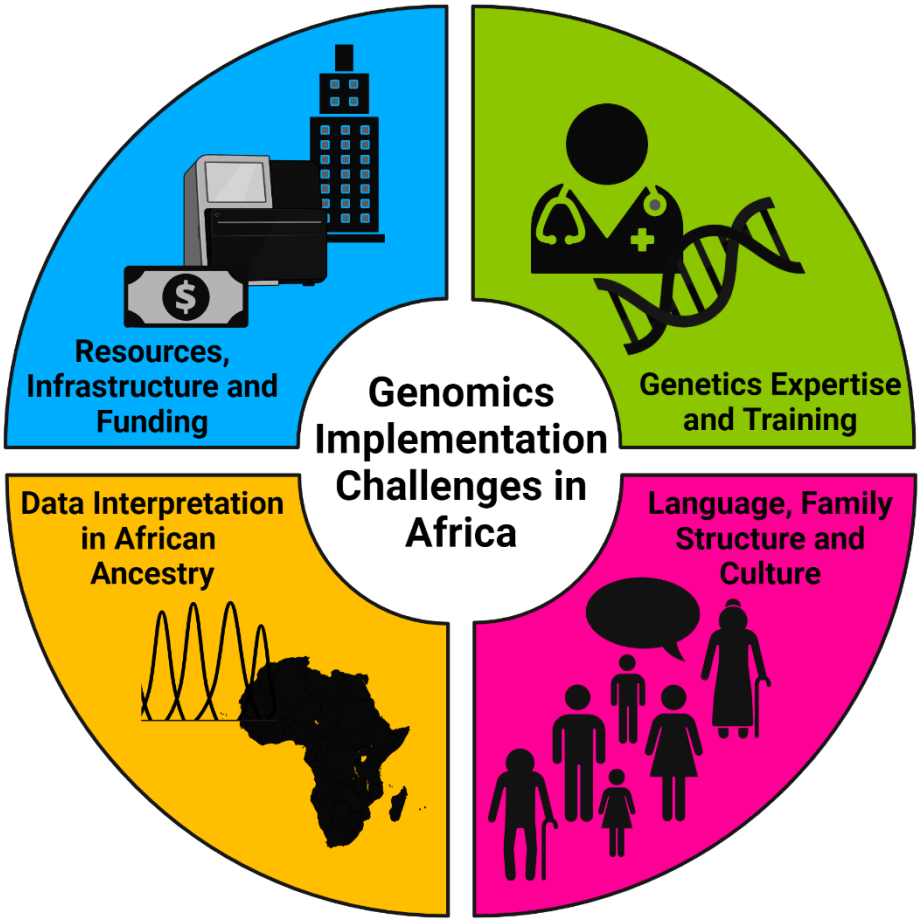


Figure 1.2. African countries face significant challenges in the implementation of genomics medicine into clinical practice. These challenges include a lack of resources, infrastructure and funding; limited genetics expertise and training; difficulties in the interpretation of genetic data from individuals of African ancestry; and language barriers, unconventional family structures and cultural beliefs.

1.7.1. Resources, Infrastructure and Funding

The lack of funding, resources, infrastructure and technology are one of the largest barriers to the implementation of NGS testing in the African context (Adebamowo et al., 2018). Shortages in field-specific experts, inadequate and poorly equipped infrastructure and inconsistent power supply threaten the ease of implementation and the development of genomic research and genetic services. A 2022 survey examined the readiness and challenges experienced by African researchers regarding the implementation of genomic medicine and services across 12 African countries. This study found that 81% of participants felt they lacked the necessary infrastructure and equipment to implement genomic services in their institutions and only 43% of

participants reported having access to genomic infrastructure locally, creating a reliance on international collaborators and service providers (Jongeneel et al., 2022).

The practical implications of the skills and infrastructure shortage and inconsistent power supply are evidenced in the implementation of prenatal genetic testing in Cameroon (Wonkam et al., 2011). During a two-year pilot of prenatal diagnostic testing, unpredictable electricity supply resulted in multiple occurrences of repeat sampling and testing, incurring an additional cost and inconvenience to participants. The absence of the necessary microscope and a trained technician to operate and analyse chorionic villus samples resulted in the complete absence of this type of sampling, delaying genetic testing to later during pregnancy and limiting termination of pregnancy choices (Wonkam et al., 2011).

Investment from external funders and local governments are needed to build the essential infrastructure and to purchase NGS technologies. Significant investment in clinical laboratories, sequencing and analysis facilities, computational facilities for bioinformatic analysis and training, data storage facilities and biobanks are needed to create operational and sustainable genomic services (Jongeneel et al., 2022). Despite positive views on the integration of genomic medicine into the African and South African healthcare system, the shortage of funding to manifest this is widely acknowledged by stakeholders (Van Der Merwe et al., 2022). Investments in infrastructure have the potential to not only benefit human and rare disease genetics, but to also be leveraged for the diagnosis of infectious and non-communicable diseases, other key healthcare concerns in Africa (Kamp et al., 2021, Pereira et al., 2021).

Economic evaluations are needed to show the benefit of these technologies in the local context to draw support and investment for resources from local policy makers, governments, and funders. The cost-savings and public health benefits of NGS have been demonstrated in studies from Australia, the United States, Canada, the United Kingdom, Netherlands, Norway, China, Thailand, Israel, Iran and Finland, but the absence of studies in Africa and in LMICs is a significant barrier when lobbying for resource investment and local uptake of these tools (Nurchis et al., 2022, Rezapour et al., 2023).

1.7.2. Genetics Expertise and Formal Training

The limited genetics workforce In South Africa, and the African continent at large, creates another barrier to implementation in the African context. Formal training of clinical geneticists, genetic counsellors and human genetic laboratory staff is scarce across the continent, and the availability of employment and formalised workforce structures to absorb internationally trained professionals is scarce (Kromberg et al., 2013, Kamp et al., 2021, Owolabi et al., 2023). This creates a reliance on primary healthcare providers, including nurses, to provide care for patients with genetic conditions, despite their limited formal training and genetic literacy (Malherbe et al., 2017). Appropriate training and knowledge of genetics and genomics is needed to consent patients and to translate and relay test findings (Ogunrin et al., 2019).

The 2022 survey by Jongeneel et al. reported insufficient education and training surrounding genetics and genomic practices for healthcare providers as a significant concern and barrier for implementation, as only a third of the represented African institutions offered formal genomic medicine training. Approximately 40% of participants expressed a basic understanding of genetics and its use in clinical practice but felt ill equipped to implement research and genetic services and had the need for further clinician-specific training (Jongeneel et al., 2022).

A study in South Africa assessing genetic knowledge and attitudes toward the use of genetics and genomic medicine for predictive testing in 45 healthcare providers and 79 medical students found participants to have adequate general genetic knowledge and positive attitudes toward the use and implementation of genetic testing (Naidoo and Reddy, 2022). Of the assessed active healthcare workers, 78% scored in the excellent knowledge category, however, genetic knowledge did not translate into ease of practice, as only 31% of practitioners in State and 18% of practitioners in private actually utilised genetic testing. Similar findings were observed in a study in Nigeria, in which 104 healthcare providers, mostly general practitioners, were questioned on their genetic knowledge and their use of genetics in clinical practice (Pughikumo and Pugikumo, 2022). Only half of the participants had adequate general knowledge of genetics and genomic medicine. Approximately 71% of the participants had made genetic diagnoses throughout their career, however, frequency of diagnoses were low (less than 10 throughout their career), indicating difficulties in identifying patients with genetic conditions and minimal referrals for confirmatory testing (Pughikumo and Pugikumo, 2022). In a 2021 study in South Africa only 50% of allied health

professionals (physiotherapists, occupational therapists and speech and language therapists) knew about the existing genetic services in the State healthcare system and only 52% of those who knew of the services had referred patients within the last year (Thom and Haw, 2021). Of the participants, 91% expressed a desire to know more about the most prevalent genetic conditions in South Africa, indicating a desire among allied healthcare providers to be knowledgeable regarding genetics and genetic services.

Initiatives to provide more widespread training on genomic medicine have begun in some African countries, such as the African Genomic Medicine Training (AGMT) Initiative and the training activities of the H3Africa Consortium, however, the skills and expertise to implement genomic medicine at levels on par with the developed world may take years to come to fruition (Mulder et al., 2018, Nembaware et al., 2019). An understanding of the availability, TATs and limitations of genetic tests is needed to successfully implement and utilise these technologies in the local healthcare system to best benefit affected patients.

1.7.3. Data Interpretation from Individuals of African Ancestry

The interpretation of genomic data from individuals of African ancestry remains a notable challenge for widespread implementation of NGS testing. The absence of data on genetic disease prevalence in African populations and on disease aetiology and phenotypes in African individuals makes the identification and treatment of patients with genetic conditions difficult (Adeyemo et al., 2018, Duncan et al., 2019, Baynam et al., 2020, Kamp et al., 2021). It is estimated that approximately 50 million Africans are affected by a rare disease. This estimate indicates a substantial number of individuals who would require diagnostic services and medical care (Lumaka et al., 2022). The true genetic disease burden, the distribution of genetic conditions amongst African populations, and the factors which modulate disease severity are under-investigated and need to be explored to develop and deliver appropriate and relevant testing in the African setting (Adeyemo et al., 2018, Baynam et al., 2020, Lumaka et al., 2022). The limited African data in publicly available databases and repositories, such as GnomAD, and the high level of natural genetic variation amongst African individuals, makes the interpretation of genomic variants difficult and time-consuming (Baynam et al., 2020, Choudhury et al., 2020, Chen et al., 2024).

Patients of African ancestry have more variants shortlisted during variant interpretation due to the absence of adequate reference sequences and are more likely to have variants misinterpreted and misclassified as pathogenic (Choudhury et al., 2020, Bowling et al., 2022). In a study of 367 ill infants who underwent WGS, individuals of African ancestry had 1.4 fold more variants retained for manual curation compared to those of European ancestry, increasing the time needed for analysis and the overall TAT of the diagnostic test (Bowling et al., 2022).

Population frequency data remains one of the strongest considerations during variant filtering and interpretation, providing evidence of rarity and novelty. A lack of representation of individuals of African ancestry therefore has significant consequences for variant identification amongst these populations (Lumaka et al., 2022). The absence of phenotypic data from individuals of African ancestry is also a considerable challenge for disease identification, even with the use of artificial intelligence (AI) and machine learning tools. This was observed in the case of a boy of Congolese descent diagnosed with Xia-Gibbs syndrome. The use of Face2Gene, an AI tool which classifies the distinctive facial features of individuals with neurodevelopmental and congenital disorders and models likely diagnoses, was unable to identify the appropriate potential diagnosis in the Congolese participant. This was the opposite phenomenon observed in patients of European decent with the same diagnosis, in which Xia-Gibbs was the first or second most likely diagnosis suggested by the AI tool (Mubungu et al., 2021). A 2021 study investigating the success of a facial analysis tool in detecting Down syndrome in individuals of Congolese decent reported better accuracy when the tool was trained with individuals of African ancestry compared to training with the general global population, as well as a 67% reduction in false positives (Porras et al., 2021). The paucity of African data in genomics databases, both genetic and phenotypic, deemed “the African genomics data gap”, is therefore a key contributor to health disparities experienced by continental Africans and African diaspora.

1.7.4. Language, Family Structure and Cultural Systems

Literacy and language, non-conventional family structure and cultural beliefs are other significant barriers in the uptake and implementation of NGS testing in Africa. Low genomic literacy amongst healthcare providers and the public, low general literacy amongst the population and the lack of standardised genomic terminology in local languages makes consent, the feedback of testing results and overall implementation

of NGS testing a considerable challenge (Solomon et al., 2012, Adebamowo et al., 2018). The translation of genetic terminology into local languages is complex and extends beyond finding equivalent linguistic terms. While scientific and genetic terminology is incorporated into English language overtime, cross-cultural translation into other languages is not and the meaning and context of terminology remains alien to other linguistic backgrounds (Kamaara et al., 2020). The use of visual aids and metaphoric language is recommended to overcome some of the language barriers, however, there is a risk of over simplification, misrepresentation and the distortion of key information – an ethical concern for consent (Goddard, 2004, Adebamowo et al., 2018, Black and Sykes, 2022).

Social nuances and non-traditional family structures makes the sequencing of parents (trio-sequencing) implausible in many cases in Africa. This increases the difficulty of variant interpretation and makes the determination of inheritance challenging. Trio-sequencing and the associated determination of inheritance, has been shown to provide incremental benefit to diagnostic yields and TATs compared to singleton “proband-only” sequencing, however the cost implications and practicalities require consideration in low-resource settings (Kingsmore et al., 2019, Tan et al., 2019).

Cultural and religious beliefs around disease causality, as well as fear of social stigma, may make individuals less likely to utilise genetic services in the African setting, prolonging diagnostic odyssey and preventing specialised treatment (Zhong et al., 2021a, Oladayo et al., 2024). A survey of 49 participants in Kenya and the Gambia found that genetic research and the confirmation of a genetic diagnosis could enhance stigmatization of already stigmatized groups, affirm antagonism toward specific groups of individuals and may contribute to apprehension to pursue testing (de Vries et al., 2012). Similar sentiments were shared in a study in Ethiopia examining the experiences of 46 individuals with podoconiosis, a genetically linked lymphoedema. Researchers found that fear of further stigmatization made individuals unlikely to seek confirmation of the genetic origin of their disease due to significant existing stigmatization from the community, including unwillingness to marry affected individuals, the avoidance of physical contact with affected persons and the shunning and social exclusion of affected individuals and their families (Tekola et al., 2009).

A study investigating the stigmatisation of a family in rural Cameroon affected by Fragile X syndrome describes the stigmatization of affected family members by the community, as well as by unaffected family members (Kengne Kamga et al., 2021).

Affected members of the family were subject to stereotypical derogatory names, were marginalised by the community, had to hide their neurological status to attract suitors for marriage and were subject to discrimination and poor treatment within their own family by nonaffected members.

Implementation strategies for NGS testing need to be sensitive to these cultural and social nuances for widespread uptake of these technologies and to maintain positive attitudes toward advancing healthcare practices. Scoping studies in varying contexts are needed to provide locally appropriate implementation strategies for NGS testing and to illuminate the unique challenges faced in each local setting.

1.8. The Study Rationale

The translation and implementation of genome, exome and panel sequencing into clinical practice has become the focus of many clinical niches. Real world studies which investigate the utility and implementation of genomics in clinics and hospitals from varied economic and social backgrounds are needed to evidence the benefit of these technologies on a global scale, including in LMICs. Investigations into the utility of NGS testing in NICUs have primarily focused on European, American and Chinese populations, with no data provided from African countries. These studies have evidenced the use and value of exome and panel sequencing in both acutely and critically ill infants in the NICU, with and without suspicion of a genetic condition. The utility and implementation of these testing technologies however remain unexplored in Africa, where the burden of genetic conditions in NICUs is still unknown.

Current literature indicates prematurity, hypoxia and infection as the leading causes of admission and death in South African NICUs, followed by congenital anomalies (Hoque et al., 2011, Moundzika-Kibamba and Nakwa, 2018, Nieuwoudt et al., 2022, Gabriels and Le Roux, 2023). Genetic contributions to preterm birth and hypoxic-ischemic encephalopathy have been identified, however, the genetic contribution to these phenotypes in NICU admission and infant death in the South African setting is unknown (Pasanen et al., 2023, Woodward et al., 2023).

Considering the absence of evidence for the diagnostic utility and implementation of NGS and virtual gene panels in South African NICUs, a scoping research study entitled “Evaluating the utility of clinical sequencing to improve diagnostic services for critically ill infants in South Africa” was initiated and lead by Dr Nadia Carstens. This larger study aimed to investigate the implementation of NGS into diagnostic service for the benefit

of ill infants in South Africa and to provide evidence of the feasibility of NGS testing in the local setting with the limited available resources and infrastructure. This PhD study evaluated the first 32 participants recruited to this larger project and established the collaborations, laboratory protocols and first analytical pipelines implemented to achieve this objective and explored the utility and unique challenges arising from the implementation of these technologies in a LMIC setting using locally available infrastructure and expertise.

1.9. Aim and Objectives

This study aimed to investigate the diagnostic utility of singleton exome-sequencing panels in ill infants with a suspected genetic condition in the South African State healthcare system.

This was achieved through the following specific objectives:

1. To recruit a cohort of 32 ill infants with a suspected genetic condition from two State hospitals in South Africa
2. To design and curate a virtual gene panel consisting of genes implicated in neonatal- and early-childhood-onset conditions, as well as genes with founder and high prevalence variants in African populations
3. To perform whole exome sequencing in-house on the Ion Torrent™ platform and to analyse the data using a series of three virtual panels
4. To assess the diagnostic utility of the three analytical approaches

2. METHODOLOGY

This chapter will describe the implementation strategy for the study, the recruited patient cohort, and the technical methodology of the NGS experiments performed. The curation of the gene panels used for NGS data analysis will also be described.

2.1. Implementation Strategy

This scoping study aimed to investigate the implementation and utility of clinical exome sequencing in the NICU environment in the State healthcare system of South Africa. This study was implemented in the NICU and associated paediatric wards at two State hospitals in Johannesburg, South Africa: Rahima Moosa Mother and Child Hospital (RMMCH), in Coronationville, and Nelson Mandela Children's Hospital (NMCH), in Parktown.

RMMCH is a regional hospital which serves State patients and is one of the largest birthing centres in Johannesburg. This hospital also serves as a diversion centre for other maternal units in regional hospitals in the Johannesburg district, often resulting in overcrowding and admissions beyond the capacity of the neonatal units. Dilapidated infrastructure and limited resources are documented ongoing issues at the facility. The contribution of these factors to neonatal infection, health and overall outcomes has not been established.

NMCH is a quaternary hospital serving both State and private patients, however State patients without financial means and alternative access to genetic services were the target population of this study. This hospital is a referral centre for the larger Gauteng province and primarily admits patients for surgical and intensive care services, skewing the admission cohort to those with very severe illness. NMCH is a specialist centre, therefore patients admitted to this facility experience easier access to specialist care (Pettifor, 2017).

Genetic services are offered at both recruiting hospitals through call-out consultations and at genetic clinics several times a week across both sites. Genetic services in the Gauteng province and in South Africa at large are limited and understaffed, resulting in long waiting times at clinics for consultations and follow ups (Kromberg et al., 2013, Wiener et al., 2023). There is a reliance on referrals from primary healthcare practitioners for patients in need of these services to be identified and for these

services to be accessed, as seen in many LMICs. This study relied on primary healthcare providers in the NICU and paediatric setting, with no formal genetics training, to screen admissions for infants who presented with features suggestive of a genetic condition and who would benefit from access to panel and exome-level testing, without formal consultation from a medical geneticist. Medical geneticists at genetics clinics and call-out consultations were utilised as a secondary referral method. This strategy was believed to best reflect the common practices in accessing genetic services in the under-resourced South African setting and in settings in which genetic services are severely restricted.

2.2. Ethics Clearance

Ethical clearance to conduct this study was obtained from the Human Research Ethics Committee (Medical) at the University of the Witwatersrand under certificate M201199 (Appendix A). Permission to recruit participants and conduct the study at the two respective hospital sites, Rahima Moosa Mother and Child Hospital and Nelson Mandela Children's Hospital, was obtained from their respective chief executive officers, supported by heads of department in neonatology and paediatrics (Appendix B).

2.3. Patient Cohort

A cohort of 32 ill infants with a suspected genetic condition was recruited for this study from two hospitals, RMMCH and NMCH, in Johannesburg, South Africa between December 2021 and November 2022. Participants were identified by neonatologists and paediatricians, or by medical geneticists during call-out consultations (hospital consultations requested outside of the genetics clinic) and referrals. Participants were recruited from NICUs, PICUs, paediatric wards and outpatient clinics. The parents or guardians of the identified participants were consented for inclusion in the research study following a discussion about the study and genetic testing by a principal investigator or genetic counsellor. Participants' families were recruited as duos (proband and one parent) and as trios (proband and both parents) depending on the availability of the parents at enrolment, however genetic testing was performed on probands only. Although trio sequencing is known to produced higher diagnostic yields and turn-around times, the availability of parents at enrolment in the local setting is a considerable limitation to the testing strategy. The confidentiality of participants and

their families was maintained throughout the study through the assignment of anonymised study numbers.

The study cohort consisted of infants under the age of two years at the time of recruitment, who presented with various features suggestive of a possible genetic condition. Participants were considered for inclusion if they had a history of serious illness or hospitalisation during the first two years of life or were referred to a genetic clinic within the first two years of life with concerns of a genetic condition. Admission to a NICU was not a hard inclusion criterion in this study. Due to the limited capacity of NICUs in the South African State healthcare system, many infants who are critically ill will not be admitted to these wards, therefore the inclusion of other paediatric wards and outpatient clinics provided a more representative range of ill infants within the State healthcare system.

Recruiting clinicians were provided broad inclusion and exclusion criteria to assist in the identification of infants with appropriate suggestive features of a genetic condition. An infant was considered eligible for inclusion in this study if they presented with any of the following: a concerning neurological status, dysmorphic features, multiple congenital anomalies, biochemical and metabolic abnormalities of an unknown cause, or if they had abnormal growth parameters. Infants were not eligible for inclusion in the study if they had diagnosed aneuploidies (aneuploidy screening is accessible to all patients in the State healthcare system and frequently requested within the NICU, PICU, paediatric wards and genetics clinics), had been significantly exposed to teratogens during gestation and labour, experienced birth asphyxia and trauma, or had a diagnosed infection that could be the primary cause of their illness. Prematurity was not considered a hard exclusion criterion, making infants whose condition could not be primarily explained by prematurity eligible for inclusion.

Phenotypic data was collected for each participant from the referring clinician using a phenotype collection rubric provided to clinicians as a short format RedCap form (Appendix C) (Harris et al., 2009, Harris et al., 2019). Clinicians were requested to provide any clinical and family history information they believed to be relevant to a potential genetic diagnosis.

A 2 to 5 mL EDTA-blood sample was collected from all probands following consent of their legal guardians. This was facilitated by an appropriate healthcare professional.

Blood samples were drawn, labelled with the appropriate anonymised study number, and stored at -20°C for sample batching and extraction.

2.4. Genomic DNA Extraction

Genomic DNA was extracted from the stored participant blood samples using a modified salting out method (Miller et al., 1988). Blood samples were batched for extraction and library preparation to minimise laboratory cost and time. Briefly, blood samples were thawed at room temperature and red blood cells lysed through consecutive washes with Triton-X-Sucrose lysis buffer (10 mM Tris, 5 mM Magnesium chloride, 0.32 M Sucrose, 0.01% Triton-X). White blood cells were lysed with 8% Proteinase K and 3% sodium dodecyl sulphate. Protein and cell debris were precipitated with sodium chloride and DNA precipitated with 100% ethanol. DNA was washed in 70% ethanol and dried at room temperature, before reconstitution and storage in tris-EDTA (TE) buffer at 4°C.

2.5. Genomic DNA Quantification and Quality Assessment

Genomic DNA was quantified photometrically and fluorometrically using the Nanodrop assessment method and the Qubit method respectively. DNA quality and integrity was assessed on the Nanodrop and by electrophoresis to ensure DNA was not fragmented prior to NGS library preparation. This ensured DNA was of a high quality and would perform optimally during the fragmentation steps of the library preparation, creating libraries of uniform size and coverage.

2.5.1. Nanodrop Quantification and Quality Assessment

DNA was quantified and its quality assessed photometrically on the NanoDrop™ 2000 (Thermo Fisher Scientific, USA). Nanodrop quantification was performed post-DNA extraction to estimate the concentration of DNA. Examination of the A260/A280 and A260/A230 ratios indicated contamination with protein and salts. The estimated concentration of DNA obtained from this assay was used to dilute genomic DNA to approximately 100 ng/ul for library preparation after further quality assessment by gel electrophoresis was performed.

2.5.2. DNA Quality Assessment by Gel Electrophoresis

Extracted genomic DNA was further assessed for quality through gel electrophoresis. A microlitre of genomic DNA was run on a 1% agarose gel at 5 V/cm and visualised

with ethidium bromide. The absence of smearing indicated intact, unfragmented DNA that could be carried forward for library preparation.

2.5.3. Qubit Quantification

Diluted genomic DNA was quantified fluorometrically for library preparation using the Qubit™ dsDNA High Sensitivity kit and the Qubit™ 4.0 fluorometer (Thermo Fisher Scientific, USA). This quantification method provided a more accurate measurement of double-stranded DNA concentration compared to the Nanodrop method (Section 2.5.1), which can be skewed by contaminants which absorb light at similar wavelengths to DNA. A microlitre of diluted 100ng/ul DNA was added to 199ul Qubit™ working solution, thoroughly mixed by vortexing and incubated at room temperature for 2 minutes to allow the fluorescent dye to bind to the DNA. Fluorescently labelled DNA was then measured against two standards of known concentration to determine the accurate concentration of DNA in the tested DNA sample.

2.6. Exome Sequencing

Library preparation for exome sequencing was performed using the Ion Ampliseq™ Exome RDY Library preparation kit (Thermo Fisher Scientific, USA), as per the manufacturer's instructions. The workflow for exome library preparation and library sequencing is summarised in Figure 2.1. Prepared libraries were diluted manually and templated onto sequencing chips using the automated Ion Chef instrument and Ion 540 Kit (Thermo Fisher Scientific, USA). Templated chips were sequenced on the Ion GeneStudio S5 sequencer (Thermo Fisher Scientific, USA).

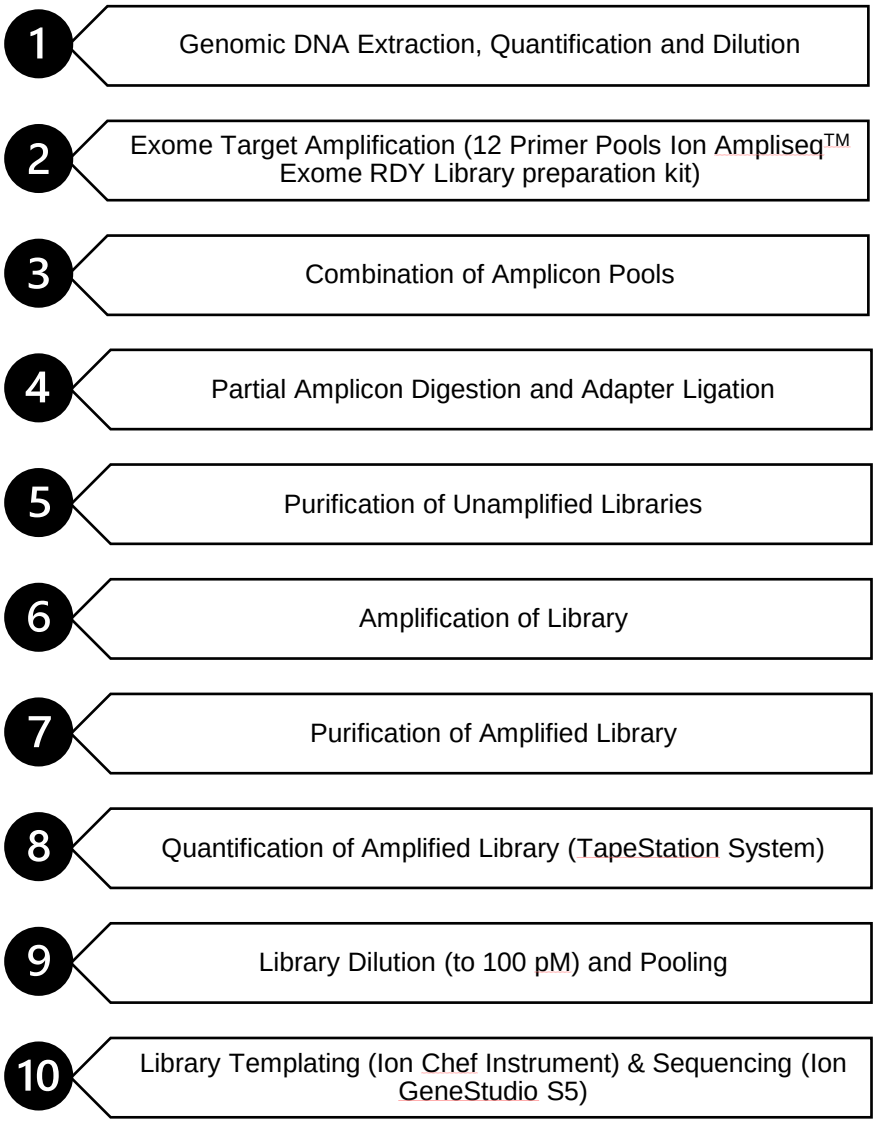


Figure 2.1. Summarised workflow for exome library preparation using the Ion Ampliseq™ Exome RDY Library preparation kit. Adapted from the Ion Ampliseq™ Exome RDY Library Preparation User Guide (Thermo Fisher Scientific, available at https://www.thermofisher.com/za/en/home/technical-resources/technical-reference-library/next-generation-sequencing-support-center/ngs-library-preparation-support.html).

2.6.1. DNA Dilution and Library Preparation

Genomic DNA was diluted for library preparation to approximately 10ng/ul in a series of dilutions. Diluted genomic DNA was quantified using the Qubit™ method (Section 2.5.3) to ensure accurate quantification of double stranded DNA. Approximately 50 ng of diluted DNA was utilised for exome target amplification on the Ion Amliseq™ Exome RDY plate (Thermo Fisher Scientific, USA). Exome target regions were amplified by PCR with twelve Ampliseq™ primer pools targeting approximately 97% of the coding consensus sequence. Exome amplicon pools were combined and partially digested with FuPa reagent in preparation for adapter binding. Ion Express barcode adapters

from the IonExpress Barcode Adapter 1-16 Kit (Thermo Fisher Scientific, USA) were ligated to digested amplicons, as well as a P1 sequencing adapter, through another PCR reaction. Amplicon libraries were purified from the reaction mixture with Agencourt AMPure XP reagent (Thermo Fisher Scientific, USA). The amplicon libraries were then further amplified in a PCR reaction with 1X Library Amp Mix and 1X Library Amp Primers (Thermo Fisher Scientific, USA) to obtain sufficient material for accurate quantification. Amplified libraries were then isolated through two rounds of purification with Agencourt AMPure XP Reagent (Thermo Fisher Scientific, USA). Final amplified libraries were reconstituted in low TE buffer, provided by the Ion Ampliseq™ Exome RDY Library preparation kit (Thermo Fisher Scientific, USA), and were quantified and stored at -20°C.

2.6.2. Library Quantification, Dilution and Pooling

DNA libraries were quantified for sequencing on the Agilent 4150 TapeStation System (Agilent Technologies, Germany) using the High Sensitivity D5000 ScreenTape Assay. This assay provided accurate quantification of libraries and sized library fragments to ensure they were an appropriate length and concentration for sequencing on the Ion S5 platform. Amplicon lengths of approximately 200bp are optimal for use on the Ion 540 chip. Library concentrations of between 300 and 1500 ng/mL are expected from library preparation using the Ion Ampliseq™ Exome RDY Library preparation kit. The High Sensitivity D5000 ScreenTape Assay can size fragments between 100 and 5000 base pairs in length, at a quantitative range of 10 to 1000 ng/mL with a sensitivity of approximately 5 ng/mL. Quantified, appropriately sized exome libraries were diluted to approximately 22 ng/mL (~100 pM), the suggested input for the Ion 540 chip, with nuclease-free water and pooled together in pairs for sequencing. Pooling libraries reduced the cost of sequencing, while still achieving an average sequencing depth of 100X on each Ion 540 sequencing Chip, as suggested by the manufacturer.

2.6.3. Library Templating and Sequencing

Automated templating of the Ion 540 sequencing chips was performed by the Ion Chef instrument using the IonChef Ion 540 Chef Reagents, Ion S5 Chef Supplies and Ion S5 Chef Solutions (Thermo Fisher, USA). The templating process immobilises amplicon libraries on Ion Sphere Participles (ISPs) and amplifies them through an emulsion PCR reaction. These clonally amplified ISPs are loaded by the automated system into the wells of the 540 chip for sequencing. Two chips were templated at a

time – one chip was then immediately sequenced and the other stored at 4°C until a sequencer became available.

Sequencing of the templated Ion 540 chips was performed on the Ion GeneStudio S5 sequencing platform (Thermo Fisher, USA). A single chip was sequenced at a time on an initialised sequencing machine. Initialisation of the S5 sequencer was performed concurrently with chip templating, to ensure a speedy workflow and to avoid lengthy storage times for templated chips. The necessary reagents for sequencing, Ion S5 Sequencing Reagents cartridge, Ion S5 Wash Solution and Ion S5 Cleaning Solution (ThermoFisher Scientific, US), were supplied to the machine during initialisation and used for the sequencing of two consecutive chips. The Ion Genestudio S5 sequencer utilised semiconductor sequencing technology to sequence immobilised amplicons in approximately 500 sequencing flows. Each flow flooded the Ion 540 chip with a series of four nucleotides, allowing for incorporation of the nucleotide into an extending DNA strand and the release of hydrogen ions. These hydrogen ions create a change in pH registered by sensors, generating a detectable base call signal.

2.6.4. Data Processing

Torrent Suite software (version 5.18.1), linked to the Ion S5 sequencer, was used for initial data processing. This software collected sequencing and base calling data, performed sequence alignment to the reference genome, build GRCh37/hg19 (the operating build on the sequencing platform at the time of sequencing and data processing). Sequence alignment also provided quality metric data for each sequencing run, accessible in a run report. Quality control data included metrics before and after read alignment. Pre-alignment metrics recorded the total bases reported, total number of reads, number of usable reads, ISP loading density, ISP loading and enrichment, polyclonality of ISPs, and read length data. Post-alignment metrics recorded the number of aligned bases, total number of aligned and unaligned reads, mean aligned read length, the longest alignment and the mean depth of coverage.

Read alignment generated binary alignment map (BAM) and binary alignment index (BAI) files. These files were then manually uploaded to Ion Reporter software for data analysis and could be utilised by Integrative Genomics Viewer (IGV) visualisation software during variant analysis and interpretation (Thorvaldsdottir et al., 2013).

2.7. Variant Analysis

Ion Reporter software (version 5.18 and 5.20) was utilised for variant calling and initial variant annotation. Uploaded BAM and BAI files were analysed using the Germline Ampliseq Exome single sample analysis workflow, a pre-existing workflow on the software. This workflow called SNVs and provided annotation from a series of online databases, listed in Table 2.1. CNVs were also called using this workflow, calling against the AmpliSeq exome CNV baseline. This baseline dataset is based on European data and no African baseline exists. CNV calling using this baseline was used as an initial screen for CNVs. Variant calling generated a Variant Call Format (VCF) file which could then be further annotated using Ensemble Variant Effect Predictor (VEP) and filtered using a series of three filtering approaches, which will be discussed in the following sections (McLaren et al., 2016).The analysis workflow utilised for SNVs and CNVs is summarised in Figure 2.2.

Table 2.1. Annotation databases utilised by the IonReporter software.

Annotation Tool/Database	Tool/Database Description	Reference
5000Exomes (v20161108)	Population frequency information from the 5000 exomes project	(Tennessen et al., 2012)
Canonical RefSeq Transcripts, Transcript Scores and Gene Model	Canonical transcript used to determine the coding region of the affected gene	(O'Leary et al., 2016)
ClinVar (20220709)	Variant impact from the NCBI ClinVar database	(Landrum et al., 2018)
dbSNP (v155)	Single nucleotide polymorphism (SNP) database of SNPs curated by USCS	(Sherry et al., 2001)
Database of Genomic Variants (DGV) (20200225)	Curated database of structural variation	(MacDonald et al., 2014)
Disease Research Area (DRA) (20170914)	Functional annotation by IonReporter™ software	-
Exome Aggregation Consortium (ExAC) (v1)	Catalogue of variant population frequencies	(Lek et al., 2016)
Gene Ontology (20210830)	Standard ontology of gene and its products, including their functional role and localisation	(Ashburner et al., 2000)
OMIM (Online Mendelian Inheritance in Man)	Database of genes and their associated genetic diseases	(McKusick, 1998)
Pfam	Protein domain families in the encoded protein	(Mistry et al., 2021)
PhyloP Scores	Conservation score of the affected protein across a range of organisms	(Ramani et al., 2019)

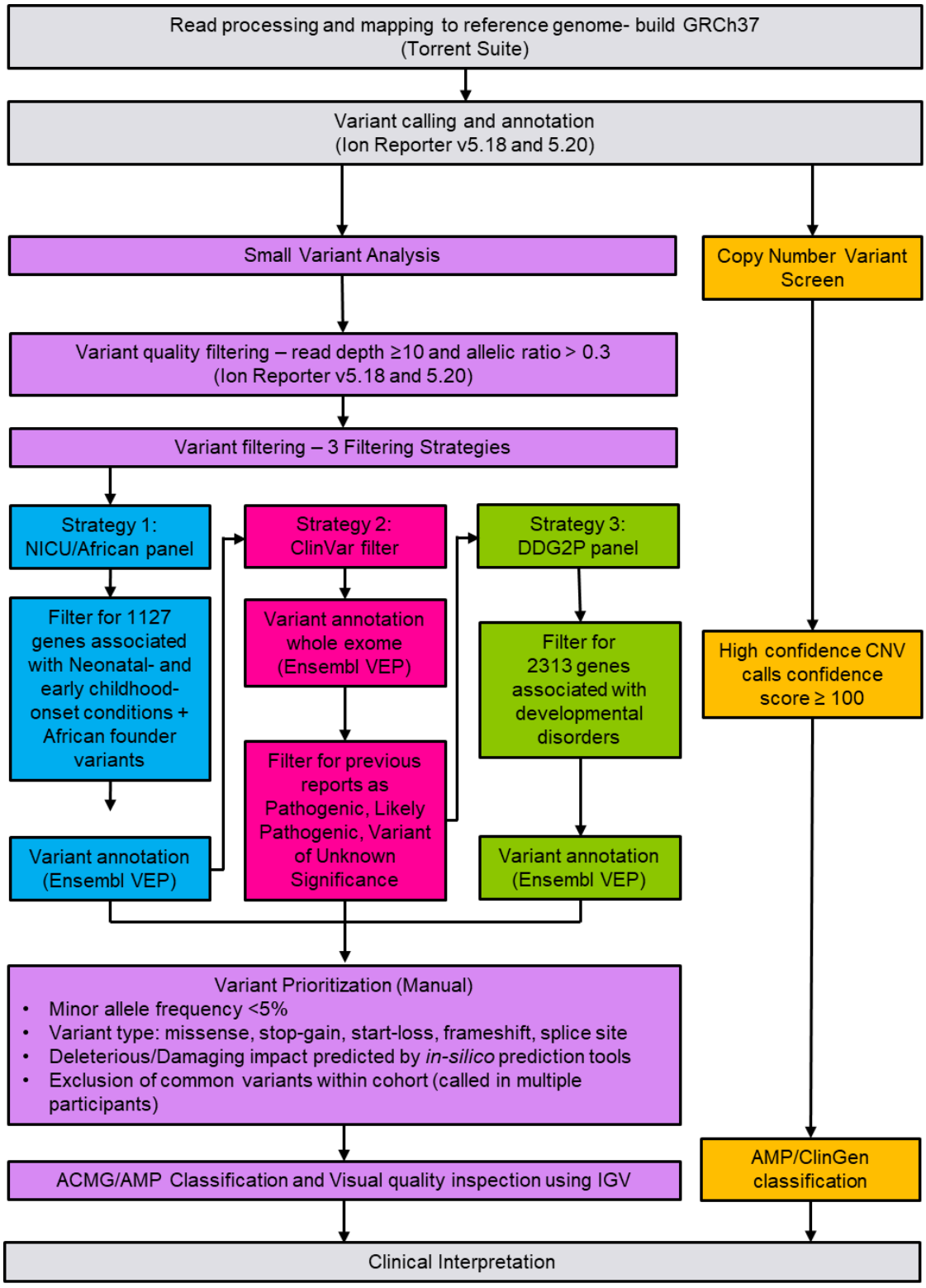


Figure 2.2. Summary of the clinical exome analysis workflow utilised for variant analysis, filtering, prioritization, and interpretation. Small nucleotide variants were filtered and prioritised using three filtering strategies: an in-house virtual gene panel of genes implicated in neonatal- and childhood-onset conditions and with known common and founder variants in African populations (NICU/African panel); a ClinVar filter to prioritise variants with previous pathogenic, likely pathogenic, uncertain significance and variants with conflicting interpretations (ClinVar filter); and a virtual gene panel of genes included in the Deciphering Developmental Disorders Developmental Disorder Genotype-to-Phenotype panel (DDG2P). Copy number variants were filtered based on calling confidence scores, with scores greater than or equal to 100 retained and classified.

2.7.1. Small Nucleotide Variant Analysis

A virtual gene panel approach was utilised in this study for SNV analysis. This approach allowed for a subset of genes to be assessed, limiting the time and complexity of the analysis. This approach has been utilised in a number of NICU studies allowing for time-sensitive clinically actionable findings to be determined. More comprehensive analysis strategies may be employed at a later time-point, at no additional laboratory cost, where necessary. Three sequential strategies for variant filtering and annotation were employed in this study. If no causative variant was identified using the first filtering approach, the second filtering approach was employed. If no causative variant was identified using the second filtering approach, the third filtering strategy was employed.

2.7.1.1. The Curated NICU/African Gene Panel Filtering Strategy

The first filtering approach filtered for variants in genes within a virtual curated gene panel. This gene panel consisting of genes curated for their involvement in neonatal- and early childhood-onset conditions, as well genes with African founder variants and common variants in African populations. The 1127 genes in this panel were selected from literature published in South Africa and by two international research groups with a focus on the implementation of genomic sequencing in newborn populations: The North Carolina Newborn Exome Sequencing for Universal Screening (NC NEXUS) project and The BabySeq Project (Ceyhan-Birsoy et al., 2017, Krause et al., 2018, Milko et al., 2019).

The BabySeq Project was a randomised trial investigating the utility of genomic sequencing in newborn care as an expanded form of newborn screening (Ceyhan-Birsoy et al., 2017). The researchers curated 1390 genes, assessing their 1514 gene-disease relationships, age of onset, penetrance, and inheritance, and classified them into three categories: highly penetrant genes with strong evidence for a disease relationship which presented in childhood, genes associated with conditions for which medical intervention is available during childhood, and those that did not meet either of these criteria. Genes falling into the first two categories, a total of 870 genes, were selected for inclusion in our curated gene panel as these were likely to be associated with conditions which presented in infancy and early childhood. The medical actionability of these genes was not assessed and considered during classifications, with the age of onset and strength of gene-disease associations prioritised.

The NC NEXUS project was another randomised controlled trial investigating the utility and implementation of exome sequencing into newborn screening practices (Milko et al., 2019). This study developed a semi-quantitative framework to assess the clinical actionability of genes based on age related factors to determine which genes are associated with neonatal- and paediatric-onset conditions, and evaluated 669 genes based on the severity of the associated disease, likelihood of disease manifestation (penetrance), efficacy of available interventions, the acceptability of available interventions considering risk of harm and evidence of gene-disease relationships. The semiquantitative framework classified genes into four categories: those associated with highly-actionable paediatric-onset conditions, those associated with paediatric-onset conditions with low actionability, those associated with highly-actionable adult-onset conditions, and those associated with adult-onset conditions with low actionability. Genes falling into the paediatric-onset categories were selected for inclusion in our curated gene panel, a total of 632 genes. Both high and low actionability genes were included to not limit the non-medical benefits of a potential clinical diagnosis.

A review by Krause et al. (2018) highlighted a number of genes with known common and founder variants associated with monogenic traits in South Africa. Ten genes were selected from this review for inclusion in our curated panel due to their relevance in African populations, regardless of the age of onset of the associated condition. Seven of these genes were already curated through the BabySeq and NC NEXUS projects. The remaining four genes, *FKBP10*, *HTT*, *JPH3* and *TYRP1* were included in our curated panel. During the inception of this study, the prevalence of *STAC3*-disorder amongst patients of African ancestry with myopathic clinical presentations was reported and investigated at the National Health Laboratory Service (Schoonen et al., 2019, Mhlongu et al., 2022). Due to its significance and prevalence in patients of African ancestry, *STAC3* was also included in our curated gene panel.

This resulted in 1127 genes being selected for inclusion in our curated neonatal- and paediatric-onset disease and African common and founder variant gene panel, referred to as the NICU/African panel. The full panel of genes, the associated phenotypes, inheritance patterns, evidence of disease associations, and respective disease penetrance is listed in Appendix D, Table D.1.

For our first filtering strategy, VCF files were filtered on Ion Reporter software by gene name using the curated NICU/African gene panel, before being downloaded and annotated using VEP. This filter retained novel variants.

2.7.1.2. The ClinVar Filtering Strategy

The second filtering approach was a ClinVar filter applied to the entire exome to retain variants previously reported on ClinVar as pathogenic, likely pathogenic, variants of uncertain significance (VUS) and variants with conflicting reports in these categories (Landrum et al., 2018) . This filter excluded variants with previous benign, likely benign or no previous classifications on ClinVar. VCF files from exome variant calling were not filtered on Ion Reporter but directly annotated with the latest ClinVar annotations using VEP, and then filtered manually by these ClinVar annotations in Microsoft Excel.

2.7.1.3. The DDG2P Gene Panel

The third filtering approach was a virtual gene panel consisting of genes identified in the Deciphering Developmental Disorders' Developmental Disorder Genotype-to-Phenotype panel (DDG2P) (static freeze from April 2023, accessed from the Genotype2Phenotype website at https://www.ebi.ac.uk/gene2phenotype/downloads) (Wright et al., 2015). This list of 2313 genes have well characterised genotype-phenotype relationships and encompass a broad range of disorders affecting growth and development (Thormann et al., 2019). Variants were filtered by this gene panel on Ion Reporter software, before the VCFs were downloaded and annotated using VEP. This filtering strategy retained novel variants.

All three filtering approaches prioritized a list of variants which were then further manually filtered and prioritised.

2.7.1.4. Manual SNV Prioritisation

Fully annotated variants from VEP were manually filtered to remove common variants, variant types predicted to have minimal functional impact on protein structure and function, any remaining variants with benign or likely benign classifications on ClinVar, and variants not predicted to have deleterious effects on protein function from *in silico* prediction tools. The filtering parameters and their respective cut offs used for prioritisation are summarised in Table 2.2.

Minor allele population frequencies below 5%, obtained from GnomAD, were considered rare and therefore prioritised. Variants predicted to have high and moderate

impact on protein function, including missense, stop-gain, start-loss, stop-loss, frameshift, splice-site, in-frame deletions and insertions, and transcript ablation variants were all retained and prioritised. Five *in-silico* prediction tools were utilised to predict potential deleterious effects of the variants on protein function. Three of these tools, BayesDel (without allele frequencies), MetaRNN and REVEL, are meta-prediction tools which incorporate prediction scores from multiple in-silico prediction tools (Ioannidis et al., 2016, Feng, 2017, Rentzsch et al., 2019, Li et al., 2022). These tools have been shown to outperform individual *in silico* predictors and provide a more accurate pathogenicity prediction (Tian et al., 2019). BayesDel and REVEL do not incorporate allele frequencies into their predictions, minimising the double-counting of evidence. The exclusion of minor allele frequencies for *in silico* predictions is proposed as a superior approach when analysing data from individuals of African ancestry, as these tools are trained on publicly available genomic data, of which there is an inadequate representation of African populations, which may skew predictions (Bope et al., 2019). SpliceAI was utilised to predict the deleterious effects of splice-site variants.

Prioritised variants were visually inspected using the IGV tool (Thorvaldsdottir et al., 2013). This allowed for the detection of sequencing artifacts and low-quality reads and variants calls. Variants with suspicious quality metrics were disregarded. Visual inspection ensured that candidate variants were well covered with no strand biases and increased the confidence of the variant calling.

Table 2.2. Filtering parameters utilised during manual filtering and prioritization.

Filtering Parameter	Cut Off	Reference
Minor allele frequency (MAF) in GnomAD	< 5%	(Chen et al., 2024)
Functional impact predicted by VEP	High and moderate impact - missense, stop-gain, start- and stop-loss, frameshift, splice site, transcript ablation and amplification, in-frame deletions and insertions, and protein altering variants	(McLaren et al., 2016)
ClinVar classification and interpretations	Pathogenic, likely pathogenic, VUS, conflicting interpretations, no previous reports on ClinVar (excluded previously reported as benign or likely benign)	(Landrum et al., 2018)
In-silico prediction on CADD, REVEL, BayesDel, MetaRNN and SpliceAI	CADD score > 15, REVEL score > 0.5, BayesDel > 0.1, MetaRNN score > 0.5, SpliceAI score > 0.5	(Ioannidis et al., 2016, Feng, 2017, Jaganathan et al., 2019, Rentzsch et al., 2019, Li et al., 2022)

2.7.1.5. SNV Classification

Prioritised variants were classified according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines for the interpretation of sequence variants (Richards et al., 2015). These guidelines classify variants into five categories based on supportive evidence in the literature and from annotation tools: pathogenic, likely pathogenic, VUS, likely benign and benign.

Candidate pathogenic and likely pathogenic variants were then reviewed in conjunction with the participant's phenotypic features to determine if the identified pathogenic or likely pathogenic variant was causative in that particular participant. This review process was performed in consultation with a medical geneticist and medical scientist.

2.7.2. CNV Screening

Quality metrics from CNV calling using the AmpliSeq™ exome CNV baseline was used to screen for potentially pathogenic CNVs in the participants. This baseline is based on European data and could therefore not confidently be used as a reference for the

calling of CNVs in this African cohort. This analysis strategy was therefore utilised as a screen, as no African baseline, from the manufacturer or in-house, was available during the course of this study. CNVs were filtered to exclude variants with a call confidence score of less than 100. This cut off was determined from the comparison of CNV calls to those detected and confirmed on CMA in five of the participants. Variants with a call confidence score of 100 could be seen on microarray and confidently called as true positives. Common variants, variants called in multiple participants in our cohort, were also removed. These were deemed unlikely causes of disease in a cohort of our size and assumed to be alignment artifacts, as disease-causing variants with severe health implications are likely to experience a selection pressure.

The remaining prioritised variants were classified for pathogenicity according to the Technical Standards for the Interpretation and Reporting of Constitutional Copy-Number Variants from the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen) (Riggs et al., 2020). These guidelines consider evidence of gene content, the inclusion of known haploinsufficient and triplosensitive genes, the number of genes in the duplicated or deleted region, data from published literature and databases, and inheritance patterns to classify variants into the five variant categories: pathogenic, likely pathogenic, VUS, likely benign and benign. An online tool to assist in the interpretation of CNVs and the assignment of points for supporting evidence is recommended by Riggs et al. (2020) and was utilised in this study. Similarly to candidate pathogenic and likely pathogenic SNVs, CNVs with pathogenic or likely pathogenic classifications were assessed in conjunction phenotypic data with a multidisciplinary team to determine if the variant was disease causing in a particular participant.

2.8. Patient Feedback

Pathogenic and likely pathogenic variants determined to be causative of the participants clinical phenotype, after review by a medical geneticist and medical scientist, were reported back to the referring clinician in a research report documenting the finding and its implications. These results were then fed back to the infants and their families by an appropriately qualified clinician, medical geneticist, or genetic counsellor and the appropriate medical intervention initiated. Incidental findings were not reported in this study.

2.9. Statistical Analysis

Fisher's exact test was used to statistically determine the relationship between a positive diagnosis and various attributes, including phenotypic features, sex and referral directly from a neonatologist or paediatrician. These analyses were performed in R (version 4.3.2) (R Core Team, 2024). The relationship between positive diagnosis and age, a continuous variable, was determined by logistic regression in R (version 4.3.2).

3. RESULTS

The results of participant recruitment, NGS technical reports and the results of NGS gene panel data analysis are presented in this chapter.

Positive SNV findings from this study were published in *Frontiers in Genetics* in November 2023 (Campbell et al., 2023)(Appendix J) and expanded on in the following chapter.

3.1. Curation of the NICU/African and DDG2P Virtual Panels

The NICU/African virtual panel contained 1127 genes pooled from two publicly available curated gene lists with five additional genes with known African specific variants. The genes in this virtual panel were associated with a range of genetic conditions affecting several physiological systems, including neurological, metabolic, and complex multisystem disorders. The range of associated phenotypes and their relative contribution to the panel can be seen in Figure 3.1. The most frequent gene-phenotype associations were with metabolic, multisystem and neurological phenotypes. Four-hundred and seventy-seven of the genes were curated by both international research groups for their association with neonatal- and early childhood-onset conditions. Of the 1127 genes, 738 genes had definitive evidence for their gene-disease association and 987 genes were considered to be highly penetrant for an associated phenotype (Ceyhan-Birsoy et al., 2017, Milko et al., 2019) (Figure 3.1, Appendix D - Table D.1). Genes with limited and disputed disease associations were included as some of the curated genes had high degrees of disease severity, early onset and available interventions which could provide utility to ill infants despite the limited associations.

The DDG2P virtual gene panel, consisting of 2313 genes from the April 2023 freeze, shared 609 of these genes with the curated NICU/African virtual panel. The use of the DDG2P filtering strategy therefore allowed for the prioritisation of variants in an additional 1704 genes. These additional genes are listed in Appendix E, Table E.1.

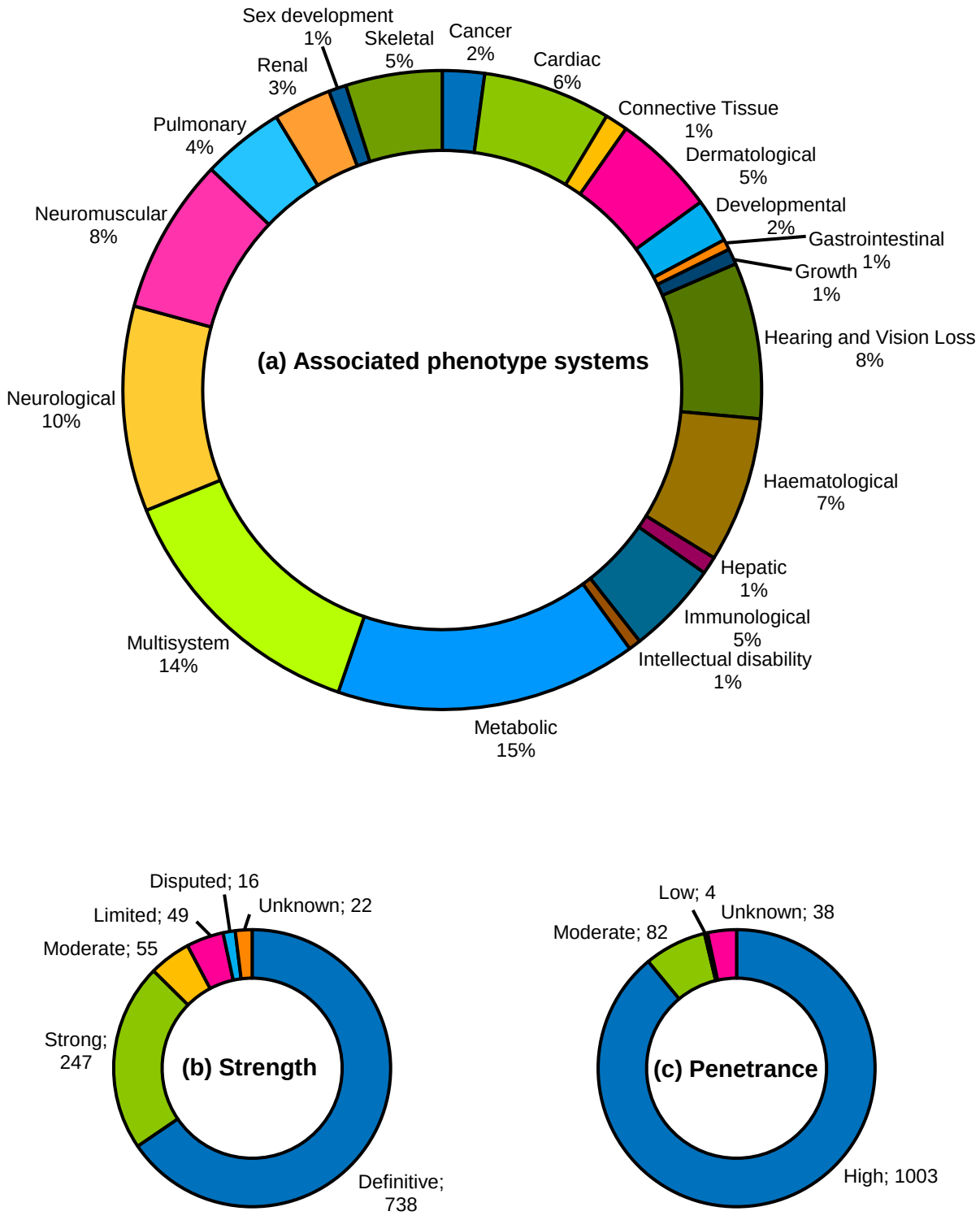


Figure 3.1. Gene disease associations and penetrance of the 1127 genes included in the NICU/African virtual panel. (a) Distribution of the phenotype and disease associations by affected physiological system given as percentages (b) Strength of the gene-disease association and the number of genes associated with each strength level (Ceyhan-Birsoy et al., 2017, Milko et al., 2019) (c) Penetrance of the associated disease phenotype and the number of genes with each penetrance level.

3.2. Cohort Demographics and Characteristics

Our participant cohort consisted of 32 ill infants under two years of age with a suspicion of a genetic condition. These infants had minimal genetic testing performed prior to recruitment, with only six of the participants having previous genetic investigations (excluding tests for aneuploidies involving chromosomes 13, 18, 21, X and Y): one disease-specific investigation and five chromosomal investigations (Table 3.1). One participant had received a supportive result for a single-gene CF test, which tested for 50 common variants in the *CFTR* gene in the South African population. This participant was found to carry a single heterozygous variant in the *CFTR* gene (c.3718-2477C>T) with no other variants in the gene identified. As CF is a recessive condition, these findings were supportive but not conclusive of a CF diagnosis. Five participants had CMA testing. All five received a negative result with no significant chromosomal aberration detected and reported. These six previous investigations did not explain the clinical phenotype of the six participants, therefore more comprehensive testing through NGS gene panels was recommended in these participants.

An equal proportion of male and female participants were recruited, with slightly more participants recruited from RMMCH (53%) than NMCH (47%). The majority of the participants were referred for NGS by a neonatologist or paediatrician without formal assessment by a clinical geneticist (23/32, 72%). Most participants were of African ancestry (26/32, 81%) and had non-consanguineous parents (30/32, 94%). Participants were referred for genetic testing for a variety of clinical indications, most frequently multiple congenital anomalies (9/32, 28%) and the presence of cardiac defects (6/32, 19%). Most participants did not have both parents present for consultation at the time of enrolment, (22/32, 69%). The average age at enrolment was 6.96 months old (SD = 7.21 month) (Appendix F, Table F.1). The demographics of the participant cohort are summarised in Table 3.1.

No participants in our cohort had any abnormalities detected prenatally on ultrasound or had any prenatal genetic investigations performed. This is not unexpected in the South African State healthcare system in which fetal anomaly scans are rarely performed within standard obstetric care and access to specialist fetal medicine clinics is limited.

Table 3.1. Demographics and testing indications of the patient cohort referred for NGS testing.

Characteristic		Number of Individuals	Percentage of Cohort (%)
Total Number of Participants		32	-
Previous genetic investigations	Chromosomal Microarray	5	15.6
	Single-Gene Test	1	3.1
Hospital Recruitment Site	RMMCH	17	53.1
	NMCH	15	46.9
Sex	Female	16	50.0
	Male	16	50.0
Referring Clinician	Neonatologist/ Paediatrician	23	71.9
	Geneticist	9	28.1
Ancestry	Black	26	81.3
	Asian and Middle Eastern	4	12.5
	Mixed	1	3.1
	White	1	3.1
Consanguinity	Non-Consanguineous parents	30	93.8
	Consanguineous parents	2	6.3
Indication for Genetic Testing	Multiple congenital anomalies	9	28.1
	Cardiac	6	18.8
	Dysmorphic	3	9.4
	Neuromuscular	3	9.4
	Developmental	2	6.3
	Hepatic	2	6.3
	Renal	2	6.3
	Skeletal	2	6.3
	Abnormal growth	1	3.1
	Gastrointestinal	1	3.1
	Neurological	1	3.1
Availability of Parents at Enrolment for Consent	Single parent available	22	68.8
	Both parents available	10	31.2

3.3. Exome Sequencing Quality Metrics

All 32 participants had libraries for sequencing prepared using the Ion Ampliseq Exome RDY kit and sized using the Agilent High Sensitivity D5000 ScreenTape assay. The library preparation assay produced libraries with an average amplicon length of 330 bp measured by the TapeStation D5000 High Sensitivity assay. This is larger than the ideal amplicon length of ~200 bp recommended for sequencing on the Ion Torrent platform. Libraries were sequenced in 17 sequencing runs on the Ion Genestudio S5 sequencer. The sequencing runs had an average data yield of 16.9 GB, 86 million reads, 90.9% Ion sphere particle (ISP) loading and 28.1% polyclonality per sequencing chip. Across all 17 runs, an average of 63.1% of reads were usable, with a mean read length of 195.6 bp. The average run statistics across all 17 runs can be seen in Table 3.2. Run statistics for each individual sequencing run are found in Appendix G, Table G.1. The average number of reads and data yields were higher than the expected 60-80 million reads and 10-15 Gb data for an Ion 540 chip. Polyclonality was also higher than the optimal 10-25%, however enough monoclonal ISPs were present to generate a high yield of usable sequencing reads.

Table 3.2. Summarised average run statistics for the 17 WES runs on the Ion GeneStudio S5 sequencer.

Run statistic	Average (Interquartile range)
Total data yield (GB)	16.9 (2.29)
Ion Sphere Particle (ISP) loading (%)	90.9 (5.25)
Total number of reads	86319414.6 (12270936)
Final library (%)	87.6 (5.75)
Polyclonality (%)	28.1 (2.5)
Low quality reads (%)	12 (5.5)
Usable reads (%)	63.1 (6)
Mean read length (bp)	195.6 (2.25)

The average coverage statistics across the sequenced samples are summarised in Table 3.3. Individual sample coverage statistics can be seen in Appendix G, Table G.2. An average of 41 706 681 reads mapped to the reference genome, build GRCh37/hg19, with a mean depth of 134X. An average of 95% of reads were on target and mapped to the exome target regions. Mapping was observed with an average

uniformity score of 90%, indicating that on average 90% of the exome was covered to at least 20% of the mean depth for each sample.

Table 3.3. Summarised average coverage statistics for the 32 WES samples.

Coverage statistic	Average (Interquartile range)
Number of mapped reads	41 706 681 (13 587 118.25)
Reads on target (%)	95 (1.20)
Mean depth (X)	134 (39.35)
Uniformity of coverage (%)	90 (3.16)

An average of 64 212 variants per participant sample were called by the Ion Reporter software using the Germline Ampliseq™ Exome single sample analysis workflow. These variants were mostly small variants, with an average of 59 007 SNVs called per participant sample. An average of 826 CNVs were called per participant sample with an average median absolute pairwise difference (MAPD) score of 0.44. A breakdown of the number of SNVs, CNVs, multiple nucleotide variants (MNVs) in which adjacent nucleotides vary from the reference, and indels called per patient sample can be seen in Table 3.4.

Table 3.4. Variants called per participant using the Ion Reporter™ Germline Ampliseq™ exome single sample analysis workflow.

Study ID	Total Variants	SNVs	Indels	MNVs	CNVs	MAPD score
NE001	66438	61995	3672	546	577	0.476
NE002	67068	61807	3576	561	1555	0.519
NE004	64321	59842	3200	509	1156	0.528
NE005	69085	63109	3492	506	2407	0.444
NE006	68585	61521	3531	574	3371	0.436
NE007	65452	61255	3367	538	620	0.456
NE009	65255	60821	3313	509	936	0.523
NE010	67151	62564	3691	513	722	0.391
NE011	66398	61295	3413	517	1515	0.401
NE012	56111	50762	3036	479	2153	0.348
NE013	66032	61172	3404	521	1304	0.359
NE014	65649	61794	3327	511	344	0.460
NE015	55295	62291	3561	549	240	0.394
NE016	66332	61268	3582	554	1259	0.476
NE017	65351	61615	3245	500	331	0.445
NE018	60725	56748	3251	488	534	0.428
NE019	65753	62046	3258	519	251	0.439
NE020	66590	62732	3330	633	360	0.389
NE021	54491	50667	3017	464	635	0.44
NE022	67089	62894	3665	555	356	0.396
NE023	67321	63445	3426	604	318	0.394
NE025	54148	50837	2925	571	277	0.373
NE026	65274	21301	3138	584	742	0.564
NE027	62867	59513	2689	552	567	0.575
NE028	65745	60982	3464	580	1051	0.496
NE029	65931	61772	3511	504	467	0.421
NE030	66508	62526	3376	577	462	0.441
NE031	64411	60971	2983	533	360	0.472
NE032	67894	63908	1310	616	460	0.427
NE033	66146	62182	3251	645	548	0.405
NE034	65520	61900	3248	621	257	0.406
NE035	53841	50674	2801	548	293	0.409
Average	64212	59007	3252	546	826	0.4
Interquartile range	2140	1633	312.3	64.3	724.3	0.01

3.4. SNV Analysis

SNV analysis using a series of 3 sequential filtering strategies yielded a reportable result in seven participants, a diagnostic yield of 22%. The demographics, clinical characteristics of these positive cases, and applied filtering strategies are summarised in Table 3.5. All 32 participants in the cohort had the NICU/African gene panel filter employed. Five participants received a confirmed molecular diagnosis using this strategy (Study ID NE001, NE021, NE022, NE029 and NE033). The identified causative variants in these participants were all previously reported on ClinVar.

Twenty-seven participants, in whom a causative variant was not identified using the first filtering strategy, had the ClinVar filter employed. Two additional participants, Study ID NE026 and NE034, received a molecular diagnosis using this filtering strategy.

Twenty-five participants had the DDG2P gene panel employed, however, no additional molecular diagnoses were achieved with this strategy. Six of the seven genes in which causative variants were identified appeared on the DDG2P panel and would have been detected if this filtering strategy was employed first. The DDG2P curated list of genes is continually updated to include the latest gene-disease evidence. The DDG2P panel used in this study was from a freeze of the panel for April 2023, which contained 2313 genes. The latest version of the panel, as of March 2024, contains 2377 genes and includes the *STAC3* gene, in which the remaining variant not initially detected using this filtering strategy occurs (latest freeze accessed through the Genotype2Phenotype website at https://www.ebi.ac.uk/gene2phenotype/downloads). Had this updated filtering strategy been employed first, all seven variants would have been identified.

A higher diagnostic rate was achieved in participants who were referred by a medical geneticist compared to those referred directly by a neonatologist or paediatrician, however this difference was not statistically significant (calculated by Fisher's exact test, Appendix H, Table H.1).

Participants who presented with skeletal, renal and developmental phenotypes had a higher diagnostic rate, however, these phenotypic categories had very few participants. Those with skeletal features were statistically more likely receive a positive diagnosis (p-value = 0.0423), however, this assertion is biased by a very small sample size of two participants (Appendix H, Table H.1). The increased diagnostic rate for infants who presented with renal and developmental phenotypes was not statistically significant.

Table 3.5. Diagnostic yields achieved through the three filtering strategies. Demographics, clinical characteristics and diagnostic yields are listed.

Characteristic		Number of Individuals	Number of Positive Diagnoses	Percentage of Positive Diagnoses
Total Number of Participants		32	7	21.9
Filtering Strategy Utilised for Diagnosis	NICU/African	32	5	15.6
	ClinVar	27	2	7.4
	DDG2P	25	-	-
Hospital Recruitment Site	RMMCH	17	6	35.3
	NMCH	15	1	6.7
Sex	Female	16	3	18.8
	Male	16	4	25.0
Referring Clinician	Neonatologist/Paediatrician	23	4	17.4
	Geneticist	9	3	33.3
Ancestry	African	26	5	19.2
	Other[$]	4	-	25.0
	Mixed	1	1	100.0
	European	1	-	-
Consanguinity	Non-Consanguineous parents	30	7	23.3
	Consanguineous parents	2	-	-
Indication for Genetic Testing	Multiple congenital anomalies	9	1	11.1
	Cardiac	6	-	-
	Dysmorphic	3	-	-
	Neuromuscular	3	2	66.7
	Developmental	2	1	11.1
	Hepatic	2	-	-
	Renal	2	1	50.0
	Skeletal	2	2	100
	Abnormal growth	1	-	-
	Gastrointestinal	1	-	-
	Neurological	1	-	-

[$] Other ancestry included individuals of Middle Eastern and Asian ancestry.

3.4.1. Participants with Positive Variant Findings

A causative pathogenic or likely pathogenic variant was identified in seven participants and reported back to the referring clinician and the participant family. These findings are summarised in Table 3.6.

Table 3.6. Summary table of reportable positive molecular findings in the infant cohort. These findings were reported back to the referring clinician and participant families. All seven variants were previously reported on ClinVar.

Study ID	Indication for Testing	Variant [$]	Variant Type	Filtering Strategy	ACMG/AMP Classification and Applicable Codes	ClinVar Accession ID and Interpretation	Variant Inheritance	Diagnosed Condition
NE001	Neuromuscular	*STAC3*:c.851G>C (p.Trp284Ser)	Missense	NICU/African	Pathogenic PM3_strong, PS3, PP3, PP5_moderate	VCV000088744 Pathogenic (2-star)	Autosomal recessive	Bailey-Bloch (*STAC3*) myopathy
NE021	Neuromuscular	*MTM1*:c.664C>T (p.Arg222Ter) [$]	Stop gain	NICU/African	Pathogenic PVS1, PP5_moderate, PM2_supporting	VCV000158994 Pathogenic (2-star)	X-linked recessive	X-linked myotubular myopathy.
NE022	Skeletal	*COL2A1*:c.3589G>A (p.Gly1197Ser) [$]	Missense	NICU/African	Likely Pathogenic PM1, PM5, PP5_moderate, PM2_supporting, PP2, PP3,	VCV000017361 Pathogenic/Likely Pathogenic (2-star)	Autosomal dominant	*COL2A1* skeletal dysplasia
NE026	Developmental	*SHOC2*:c.4A>G (p.Ser2Gly) [$]	Missense	ClinVar	Pathogenic PS2_very strong, PS4, PS3_moderate, PP5_strong, PM2_supporting, PP2	VCV000006821 Pathogenic (3-star)	Autosomal dominant	Noonan-like syndrome with loose anagen hair
NE029	Multiple congenital anomalies	*OCRL*:c.1621C>T (p.Arg541Ter) [$]	Stop gain	NICU/African	Likely Pathogenic PVS1_strong, PP5_moderate, PM2_supporting	VCV000521093 Pathogenic (2-star)	X-linked recessive	Lowe syndrome
NE033	Renal	*NPHS1*:c.1379G>A (p.Arg460Gln) [$]	Missense	NICU/African	Likely Pathogenic PM3_strong, PM1, PP5_moderate, PM2_supporting	VCV000056438 Pathogenic/Likely Pathogenic (2-star)	Autosomal recessive	Congenital nephrotic syndrome
NE034	Skeletal	*TRPV4*:c.806G>A (p.Arg269His) [$]	Missense	ClinVar	Pathogenic PS4, PS3, PM5, PM1, PP5_moderate	VCV000005000 Pathogenic (2-star)	Autosomal dominant	Spondylo-metaphyseal dysplasia

[$] Variants in genes which are on the DDG2P virtual panel (freeze for April 2023)

3.4.1.1. Participant NE001

Participant NE001 was a male participant who presented at birth with features suggestive of a severe myopathy and who was referred for exome sequencing by his neonatologist at two days of age. This participant presented with low-set posteriorly rotated ears, a tented upper lip vermillion, a high arched palate, myopathic facies, a wide intermammillary distance, a single transverse palmar crease, cryptorchidism, genu recurvatum, rocker bottom foot, reduced deep tendon reflexes, generalised hypotonia, and respiratory and swallowing difficulties.

The first filtering strategy, the NICU/African panel, prioritised 79 variants for manual curation. Two variants of interest were shortlisted from these curated variants. The first of these variants was in the *GFAP* gene, associated with Alexander disease, an autosomal dominant white matter disorder characterised by megalencephaly, seizures and failure to thrive (Springer et al., 2000). The characteristic features of Alexander disease did not fit the phenotype of Participant NE001, therefore this variant was excluded as disease-causing in this participant.

The second variant of interest was a homozygous variant in the *STAC3* gene. This variant was well covered during sequencing (123X) with no strand bias (Figure 3.2). The identified variant, *STAC3*:c.851G>C (p.Trp284Ser), has been reported more than 10 times on ClinVar, carrying a 2-star pathogenic classification (VCV000088744) (PP5_moderate) and has been reported in a homozygous state multiple times in the literature in patients with *STAC3* myopathy, also called Bailey-Bloch myopathy (PM3_strong) (Horstick et al., 2013, Telegrafi et al., 2017, Waldrop et al., 2017, Zaharieva et al., 2018, Schoonen et al., 2019, Ravenscroft et al., 2021, Saleh et al., 2021). This variant is predicted to have a deleterious effect on protein function by *in silico* meta-prediction tools REVEL, BayesDel and MetaRNN (PP3). These deleterious effects on protein function have been observed in zebrafish models, who display deficient excitation-contraction coupling in muscle tissue as a result of this variant (PS3) (Horstick et al., 2013, Linsley et al., 2017). This variant was classified as pathogenic according to ACMG/AMP guidelines.

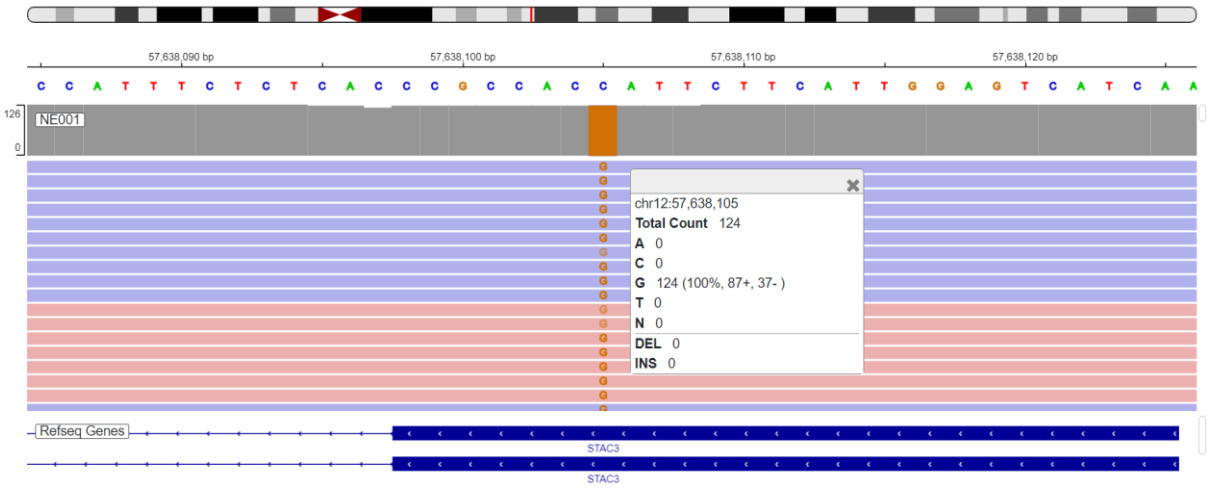


Figure 3.2. IGV image depicting the pathogenic *STAC3*:c.851G>C (p.Trp284Ser) variant identified in Participant NE001. Transcription of the *STAC3* gene occurs in the reverse direction, indicated by the <<< arrows in the gene box, therefore, the G>C change is represented visually as C>G. This variant was observed in a homozygous state and with 124X coverage. Reads are coloured by strand, purple representing reverse (-) strands and pink forward (+) strands.

This variant was determined to be causative in this participant, who presented with several characteristic features of *STAC3* myopathy. This diagnosis was reported back to the referring clinician and the family (Table 3.6). Due to limited resources in the State healthcare system and the poor prognosis of the condition, NE001 was not ventilated and demised within a week of life. The parents of Participant NE001 were offered prenatal testing for any future pregnancies. This could determine if the fetus has the same severe homozygous variant, at which there is a 25% risk.

3.4.1.2. Participant NE021

Participant NE021 was a male participant referred for exome sequencing by his neonatologist on day two of life due to his severe neuromuscular phenotype. The participant presented with widely open fontanelles, a low anterior hairline, low set posteriorly rotated ears, a high arched palate, retromicrognathia, myopathic facies, a narrow chest, hypertrichosis of the face and upper back, fifth finger clinodactyly, hypoplastic scrotum, cryptorchidism, hypoplastic toenails, generalised hypotonia and upper and lower limb weakness, areflexia, low power and breathing difficulties. Participant NE021 was delivered at 35 weeks gestation due to fetal distress and was born with very low muscle tone.

This participant had a family history of severe myopathy and had had a male sibling with a similar clinical presentation who demised within three days of life (pedigree can

be seen in Figure 3.3). The sibling had undergone genetic testing for spinal muscular atrophy and received a negative result for variants in the *SMN1* gene. As Participant NE021 was the second child with severe myopathy born to this couple, it was suggested that there may be a possible hereditary genetic cause to the disease presentation in this family.

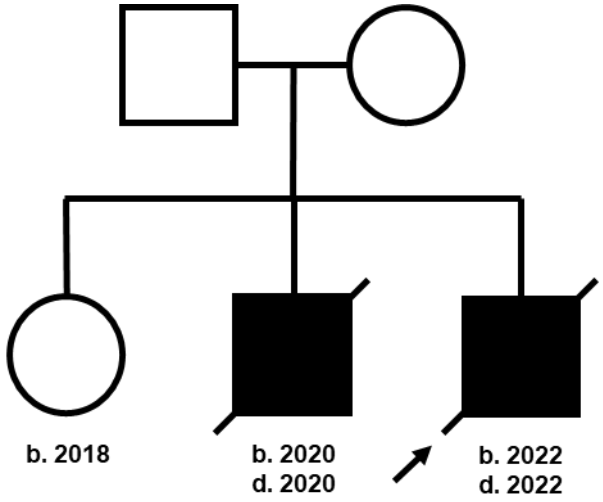


Figure 3.3. Pedigree of the family of Participant NE021. This family had two affected male children, born in 2020 and 2022. Both children demised within a week of life. This couple have an older female child who is healthy.

A single variant was shortlisted as a variant of interest from the 59 variants prioritised for manual curation through the NICU/African filtering strategy. This variant, *MTM1*:c.664C>T (p.Arg222Ter), was a nonsense variant in the *MTM1* gene on the X-chromosome. The pathogenic variant was observed in a hemizygous state at 47X coverage in Participant NE021 (Figure 3.4).

The *MTM1* gene encodes the myotubularin phosphatase domain of the MTM1 protein, a tyrosine phosphatase enzyme (Laporte et al., 1997). Variants in *MTM1* have been associated with X-linked myotubular myopathy, characterised by neonatal hypotonia, neonatal respiratory failure, muscle weakness, decreased muscle bulk, cryptorchidism and facial weakness (Laporte et al., 1997, Tanner et al., 1999, Biancalana et al., 2003, Longo et al., 2016, Amburgey et al., 2017). The identified variant, *MTM1*:c.664C>T (p.Arg222Ter), introduces a premature stop codon in the eighth exon of the *MTM1* gene (Figure 3.5), in which loss-of-function is a known mechanism of disease (PVS1_very strong). This variant is absent in GnomAD exomes and genomes (PM2_supporting) and carries a 2-star pathogenic classification on ClinVar (VCV000158994) (PP5_moderate). This variant has been reported in the literature in

six other patients with severe X-linked myotubular myopathy (Laporte et al., 1997, Biancalana et al., 2003, Tsai et al., 2005, Gurgel-Giannetti et al., 2012). Considering this evidence, this variant was classified as pathogenic.

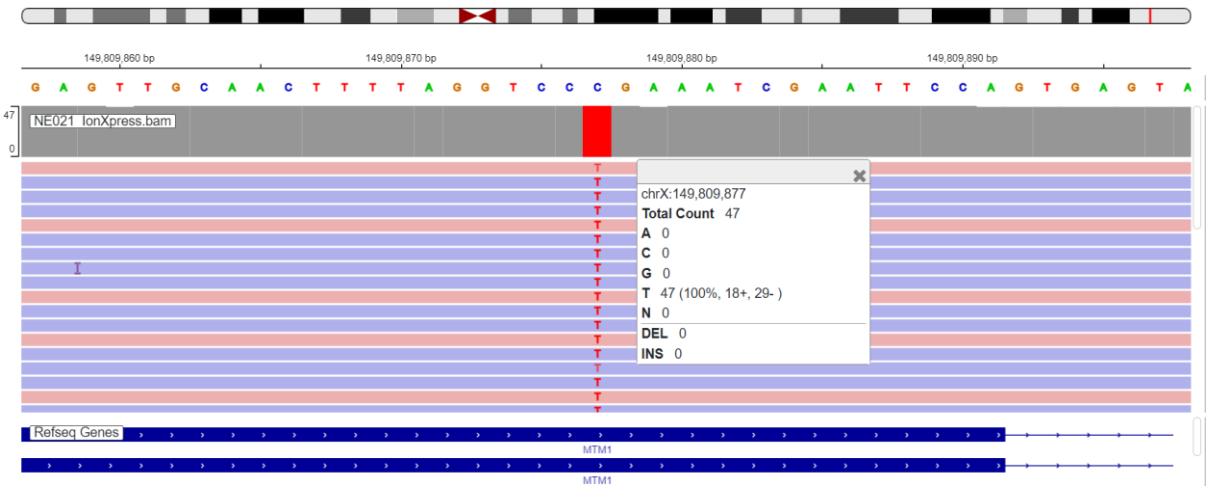


Figure 3.4. IGV image depicting the pathogenic *MTM1*:c.664C>T (p.Arg222Ter) variant identified in Participant NE021. This variant was observed in a hemizygous state and with 47X coverage. Reads are coloured by strand, purple representing reverse (-) and pink forward (+) strands.

(a) AFFECTED PROTEIN REFERENCE SEQUENCE

MASASTSKYNSHSLENESIKRTSRDGVNRDLTEAVPRLPGETLITDKEVIYICPFNGPIKGRVYITNYRLYLRSLETDSSLILD
VPLGVISRIEKMGGATSRGENSYGLDITCKDMRNLRFALKQEGHSRRDMFEILTRYAFPLAHSLPLFAFLNEEKFNVDGWTVYN
PVEEYRRQGLPNHHWRITFINKCYELCDTYPALLVVPYRASDDDLRRVATFRSRNRIPVLSWIHPENKTVIVRCSQPLVGMSGK
RNKDDEKYLDVIRETNKQISKLTIYDARPSVNAVANKATGGGYESDDAYHNAELFFLDIHNIHVMRESLKKVKDIVYPNVEESH
WLSSLESTHWLEHIKLVLTGAIQVADKVSSGKSSVLVHCSDGWDRTAQLTSLAMLMLDSFYRSIEGFEILVQKEWISFGHKFAS
RIGHGDKNHTDADRSPIFLQFIDCVWQMSKQFPTAFEFNEQFLIIILDHLYSCRFGTFLFNCESARERQKVTERTVSLWSLINS
NKEKFKNPFYTKEINRVLYPVASMRHLELWVNYYIRWNPRIKQQQPNPVEQRYMELLALRDEYIKRLEELQLANSAKLSDPPTS
PSSPSQMMPHVQTHF*

(b) AFFECTED PROTEIN PREDICTED SEQUENCE

MASASTSKYNSHSLENESIKRTSRDGVNRDLTEAVPRLPGETLITDKEVIYICPFNGPIKGRVYITNYRLYLRSLETDSSLILD
VPLGVISRIEKMGGATSRGENSYGLDITCKDMRNLRFALKQEGHSRRDMFEILTRYAFPLAHSLPLFAFLNEEKFNVDGWTVYN
PVEEYRRQGLPNHHWRITFINKCYELCDTYPALLVVPYRASDDDLRRVATFRS*

Figure 3.5. The predicted effect of the nonsense *MTM1*:c.664C>T (p.Arg222Ter) variant on the *MTM1* protein. Image generated using Mutalyzer v3.02.5. (a)This variant introduces a premature stop codon (*) at codon 222 in the 8th exon of MTM1, resulting in a truncated protein in which the red amino acids residues are absent. (b) The resulting truncated protein sequence.

Participant NE021's clinical features were consistent with a diagnosis of myotubular myopathy; therefore, this variant was determined to be disease-causing in this participant and reported back to the clinician and family (Table 3.6). Due to the severity

of this participants condition, the participant demised within a week of life, shortly after sample collection.

This finding of a X-linked condition had significant implications for the family, who indicated that they did want more children after the loss of their two male children. This family was offered prenatal testing for any future pregnancies. This could determine if the fetus harbours the same identified pathogenic variant.

3.4.1.3. Participant NE022

Participant NE022 was a female participant referred by a medical geneticist at 13 months of age with a suspicion of a skeletal dysplasia. This participant had a history of admission to the NICU for failure to thrive. She presented with a prominent forehead, epicanthic folds, infraorbital creases, grey sclera, a broad nasal bridge, malar hypoplasia, macrocephaly, a short neck, short sternal prominence, pectus carinatum, a short and narrow chest, flared ribs, Harrison's sulcus, a protuberant abdomen with a soft liver and spleen, exaggerated lumbar lordosis, restricted elbow extension, restricted hip abduction, mild rhizomelic shortening, and truncal hypotonia.

A single variant of interest was shortlisted from the 82 variants prioritised for manual curation through the NICU/African panel and classified according to ACMG/AMP guidelines. The variant shortlisted was a heterozygous missense variant in the *COL2A1* gene, *COL2A1*:c.3589G>A (p.Gly1197Ser).

COL2A1 encodes the pro-alpha1(II) chain of type II collagen and variants in this gene are associated with autosomal dominant type II collagenopathies, including several skeletal disorders and ocular disorders (Barat-Houari et al., 2016b). *COL2A1* is a predicted missense constraint gene and carries a Z-score of 5.37 on GnomAD (PP2). The identified variant affects a glycine residue, an essential component of the stabilising Xaa-Yaa-glycine triple repeat which stabilises collagen structure (PM1) (Beck et al., 2000, Bodian et al., 2008). Multiple *in silico* meta-prediction tools, including REVEL, MetaRNN, and BayesDel, predict this variant to be damaging to COL2A1 protein structure (PP3).

The variant is absent from GnomAD (PM2_supporting) but has been reported in the literature in multiple patients with Spondyloepimetaphyseal dysplasia (Chan and Cole, 1991, Barat-Houari et al., 2016a). The variant carries a 2-star pathogenic/likely pathogenic classification on ClinVar (VCV000017361), with more than ten pathogenic/likely pathogenic submissions (PP5_moderate). Alternative amino acid

substitutions at the same position, p.Gly1197Ala and p.Gly1197Arg, carry 2-star pathogenic classifications on ClinVar (VCV000988359 and VCV001326882), highlighting the potential importance of this glycine residue in COL2A1 structure (PM5). Considering this evidence, the variant was classified as likely pathogenic according to ACMG/AMP guidelines.

The *COL2A1*:c.3589G>A (p.Gly1197Ser) variant was observed in a heterozygous state and had coverage of 85X (Figure 3.6). The skeletal features of Participant NE022 are in keeping with a collagenopathy, therefore this variant was reported as disease-causing in this participant (Table 3.6). Due to the range of phenotypes associated with variation in the *COL2A1* gene, further clinical assessments are needed to assist in refining this diagnosis to a specific disorder. This participant will require further monitoring for eye abnormalities, cervical instability and decreased respiratory function which are common clinical manifestations of *COL2A1* disorders.

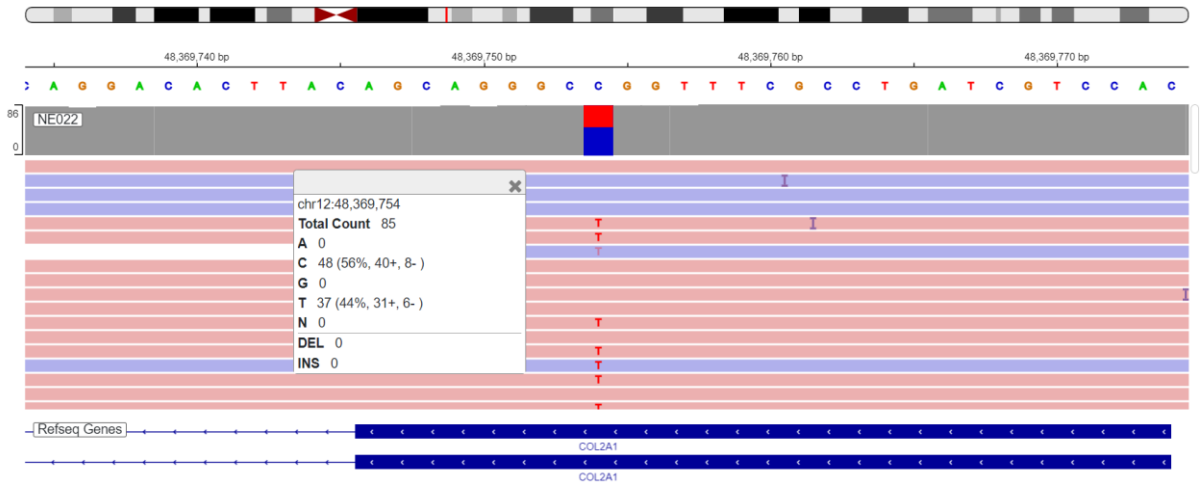


Figure 3.6. IGV image depicting the pathogenic *COL2A1*:c.3589G>A (p.Gly1197Ser) variant identified in Participant NE022. Transcription of the *COL2A1* gene occurs in reverse direction, indicated by the <<< arrows in the gene box, therefore, the G>A change is represented visually as C>T. This variant was observed in a heterozygous state and with 85X coverage. Reads are coloured by strand, purple representing reverse (-) and pink forward (+) strands.

3.4.1.4. Participant NE026

Participant NE026 was a male patient referred by his paediatrician at twenty-four months of age due to his developmental delay. This participant presented with scaphocephaly, alopecia, frontal bossing, sparse eyebrows, a flat nasal bridge, malar flattening, retromicrognathia, fragile teeth, mild abdominal distension, scrotal

dysplasia, dry skin, a history of neonatal seizures, mild hypotonia, hypocalcaemia, failure to thrive, and a history of feeding difficulties.

A single variant of interest in the *DCHR7* gene was shortlisted for curation through the NICU/African filtering strategy. Variants in this gene, encoding a sterol reductase enzyme involved in sterol biosynthesis, have been associated with autosomal recessive Smith-Lemi-Opitz syndrome, a congenital malformation and mental retardation disorder which typically manifests as pre- and post-natal growth restriction, intellectual disability, and microcephaly (Nowaczyk and Irons, 2012). This variant was excluded as it was observed in a heterozygous state and the associated phenotype was not fitting with the clinical presentation of Participant NE026.

As a causative variant was not identified using the first filtering strategy, a second strategy, the ClinVar filter, was utilised. Three variants of interest were shortlisted from these curated variants, including the *DHCR7* variant shortlisted by the NICU/African filter.

The first variant shortlisted from the ClinVar filter, *PADI3*:c.1669C>T (p.Arg557Trp), was a heterozygous missense variant in the *PADI3* gene. This gene encodes a peptidylarginine deiminase enzyme, detected primarily in the epidermis and in hair follicles (Nachat et al., 2005). Variants in this gene are associated with autosomal recessive central centrifugal cicatricial alopecia (CCCA), a common form of alopecia in individuals of African ancestry (Malki et al., 2019). The identified variant is present in GnomAD at a carrier frequency of less than 0.6% in individuals of African ancestry (PM2) and has been reported in three individuals of African ancestry with CCCA in the literature (Malki et al., 2019). This variant is predicted to be tolerated by in-silico meta-prediction tools MetaRNN and REVEL (BP4), however, a conflicting *in vitro* study in human keratinocyte cells showed altered intracellular localization of the PADI3 protein and decreased enzymatic activity due to this variant (PS3_moderate). This variant has a single submission on ClinVar as pathogenic (VCV000599196).

This variant was observed in Participant NE026 in a heterozygous state and a second pathogenic variant in the *PADI3* gene was not identified. Participant NE026 was therefore considered a carrier of the *PADI3* variant, an incidental finding that was not returned to the participant (Table 3.7). Although participant NE026 presented with alopecia, not all of his phenotypic features could be explained by a diagnosis of CCCA.

The second variant shortlisted through the ClinVar filter in Participant NE026 was a heterozygous missense variant in the *SHOC2* gene, *SHOC2*:c.4A>G (p.Ser2Gly). *SHOC2* encodes a scaffold protein implicated in the activation of the RAF/MAPK1 signalling pathway (Young et al., 2018). Variants in the *SHOC2* gene have been associated with autosomal dominant Noonan-like syndrome with loose anagen hair (NLSLAH) (Cordeddu et al., 2009). This condition shares clinical features with Noonan syndrome including short stature, characteristic facial features, growth hormone deficiency, cognitive impairment, and cardiac defects, but is distinguished by its hair associated anomalies (Komatsuzaki et al., 2010). The variant identified in Participant NE026, *SHOC2*:c.4A>G (p.Ser2Gly), is a well reported NLSLAH variant, with multiple NLSLAH patients harbouring the same variant reported in the literature (PS4) (Cordeddu et al., 2009, Hoban et al., 2012, Gripp et al., 2013, Gargano et al., 2014, Choi et al., 2015). This variant has also been reported in multiple *de novo* RASopathy cases (PS2_very strong) and has a minor allele frequency of less than 1% on GnomAD (PM2_supporting) (Cordeddu et al., 2009, Ekvall et al., 2011, Hoban et al., 2012, Gripp et al., 2013). This variant has been evaluated by the ClinGen Rasopathy Expert panel and given a 3-star pathogenic classification on ClinVar (VCV000006821) (PP5_strong). Functional studies in nematodes evidence the deleterious effect of this variant on the SHOC2 protein, which has a high rate of pathogenic missense variants, causing altered localisation of the SHOC2 protein and impaired MAPK activation (PP2, PS3_moderate) (Cordeddu et al., 2009, Gelb et al., 2018). Considering this evidence, this variant was classified as pathogenic according to ACMG/AMP guidelines.

Participant NE026 presented with many features characteristic of NLSLAH. Sparse, easily pluckable hair is a characteristic feature of the condition which distinguishes it from other RASopathies. This variant was observed in a heterozygous state, fitting the autosomal dominant mechanism of disease, and had coverage of 34X during sequencing (Figure 3.7). This variant was therefore determined to be causative in this participant and reported back to the clinician and family with a diagnosis of NLSLAH (Table 3.6 and Table 3.7). This participant will require specialist neurodevelopmental and cardiac assessments to monitor psychomotor delay and cardiac anomalies associated with NLSLAH.

Table 3.7. Shortlisted variants identified in Participant NE026 prioritized through the ClinVar filtering strategy. The variant, utilised filtering strategy, zygosity, coverage, variant type, associated phenotype, applicable ACMG/AMP codes, and the interpretation of the variant is provided. AR = autosomal recessive, AD = autosomal dominant

Gene (OMIM number)	Variant	Zygosity and Variant Coverage	Variant Type	Associated Phenotype	ClinVar Classification and Accession ID	Applicable ACMG/AMP Codes and Interpretation
PADI3 (606755)	c.1669C>T (p.Arg557Trp)	Heterozygous 78X	Missense	AR Central centrifugal cicatricial alopecia	Pathogenic (1-star) VCV-000599196	PS3_moderate, PM2_supporting, PP5, BP4 Heterozygous carrier (Incidental finding), Not reported
SHOC2 (602775)	c.4A>G (p.Ser2Gly)	Heterozygous 34X	Missense	AD Noonan-like syndrome with loose anagen hair	Pathogenic (3-star) VCV-000006821	PS2_very strong, PS4, PP5_strong PS3_moderate, PM2_suppprting, PP2 Pathogenic, Reported

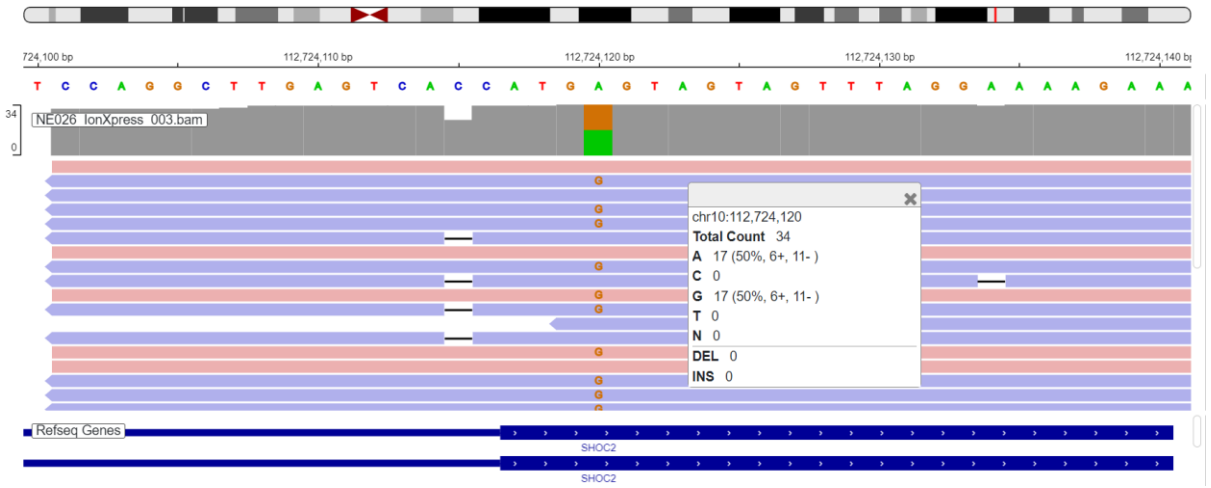


Figure 3.7. IGV image depicting the pathogenic *SHOC2*:c.4A>G (p.Ser2Gly) variant identified in Participant NE026. This variant was observed in a heterozygous state and with 34X coverage. Reads are coloured by strand, purple representing reverse (-) and pink forward (+) strands.

3.4.1.5. Participant NE029

Participant NE029 was a male patient referred by a medical geneticist at 1 year old as he presented with multiple congenital anomalies. This participant presented with open anterior fontanelle, cataracts, blocked tear ducts, bilateral eye opacity, bilateral

nystagmus, retrognathia, a head lag, redundant skin on the shoulder joints, brachydactyly, hyperlaxity of the finger joints and tapered fingers, osteomalacia and rickets (observed on x-ray), short stature, global hypotonia with reduced reflexes and power, proteinuria, and stagnating weight. The participant had no history of NICU admission; however, the patient was seen at multiple paediatric clinics, including genetics, during his first year of life for specialist consultations to address his neurodevelopmental, ophthalmological, orthopaedic, and renal concerns.

A single variant of interest was shortlisted in this participant. This variant, *OCRL:*c.1621C>T (p.Arg514Ter), was a nonsense variant in the phosphatidylinositol 4,5-bisphosphate-5-phosphatase encoding gene, *OCRL*. Phosphatidylinositol 4,5-bisphosphate-5-phosphatase, found in lysosomal compartments, plays a role in intracellular trafficking (Mehta et al., 2014). Variants in the *OCRL* gene are associated with Lowe oculocerebrorenal syndrome, an X-linked multisystem condition involving abnormalities of the eyes, kidneys, and central nervous system (Bokenkamp and Ludwig, 2016). The disease typically manifests with cataracts and vision impairment, renal tubular dysfunction and albuminuria, and hypotonia (Loi, 2006).

Deletions, frameshift variants and nonsense variants are most frequently observed as disease-causing in *OCRL* (Hichri et al., 2011). The shortlisted variant introduced a premature stop codon in the sixteenth of 24 of exons in *OCRL* (Figure 3.8), resulting in a truncated protein without two essential catalytic domains, the phosphatase domain and Rho GTPase-activating protein domain. Loss-of function variants in *OCRL* are a known mechanism of disease for Lowe syndrome, supported by a GnomAD pLOF score of 1 and a haploinsufficiency curation by Clingen (PVS1_strong). This variant is absent from GnomAD (PM2_supporting) and has been reported twice in a hemizygous state in the literature in patients with features suggestive of Lowe syndrome (Yang et al., 2014, Recker et al., 2015). This variant carries a 2-star pathogenic classification on ClinVar (VCV000521093) and has 5 non-conflicting pathogenic submissions (PP5_moderate). Considering this evidence, this variant was classified as likely pathogenic according to ACMG/AMP guidelines.

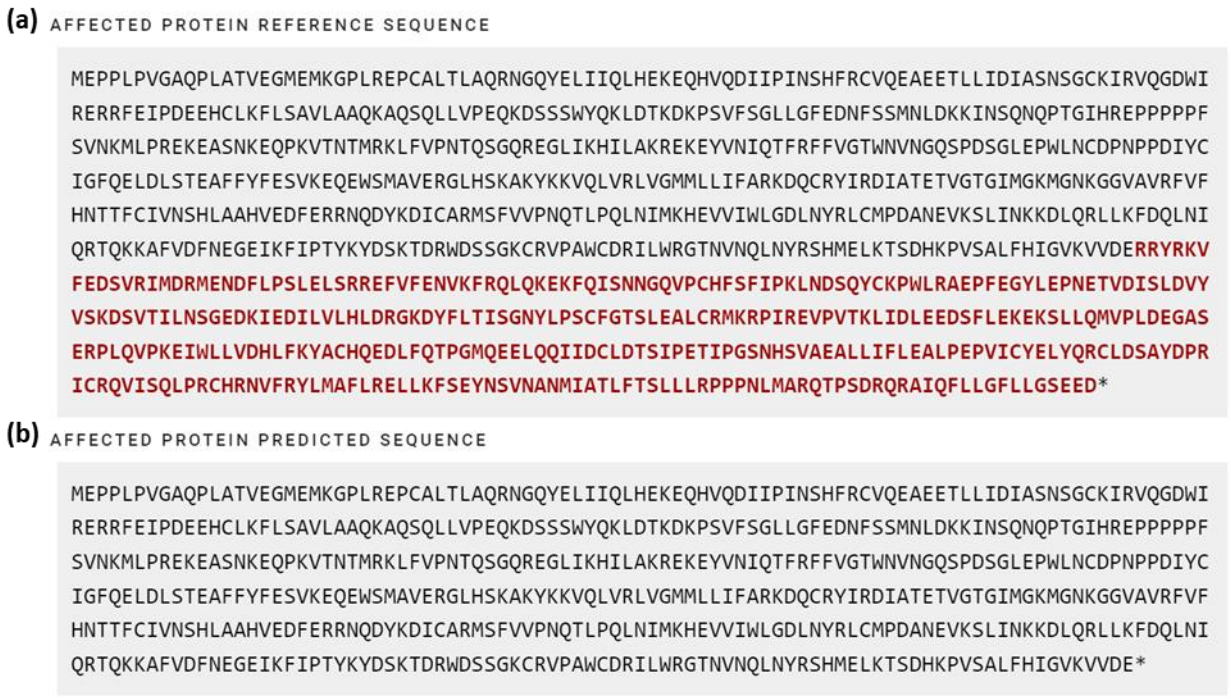


Figure 3.8. The predicted effect of the nonsense *OCRL:*c.1621C>T (p.Arg514Ter) variant on the *OCRL* protein. Image generated on using Mutalyzer v3.02.5. (a)This variant introduces a premature stop codon at codon 514 in the 16th exon of the *OCRL* gene, resulting in a truncated protein in which the red amino acids residues are absent. (b) The resulting truncated protein sequence in which the two essential catalytic domains, the phosphatase domain at amino acids 564-678 and the Rho GTPase-activating protein domain at amino acids 721-901, are absent.

The likely pathogenic *OCRL* variant was observed in a hemizygous state in Participant NE029 and was observed at a coverage of 16X (Figure 3.9). The coverage of this variant was low, therefore this variant was validated by Sanger sequencing (Figure 3.10). Participant NE029 presented with clinical features fitting with a diagnosis of Lowe syndrome, including the characteristic visual impairment due to cataracts, general hypotonia, and proteinuria, indicative of kidney dysfunction. This variant was therefore determined to be causative in this participant and reported back to the clinician and family (Table 3.6). This participant will be referred for ophthalmological treatment for cataracts and glaucoma, as well as electrolyte monitoring for renal tubular acidosis.

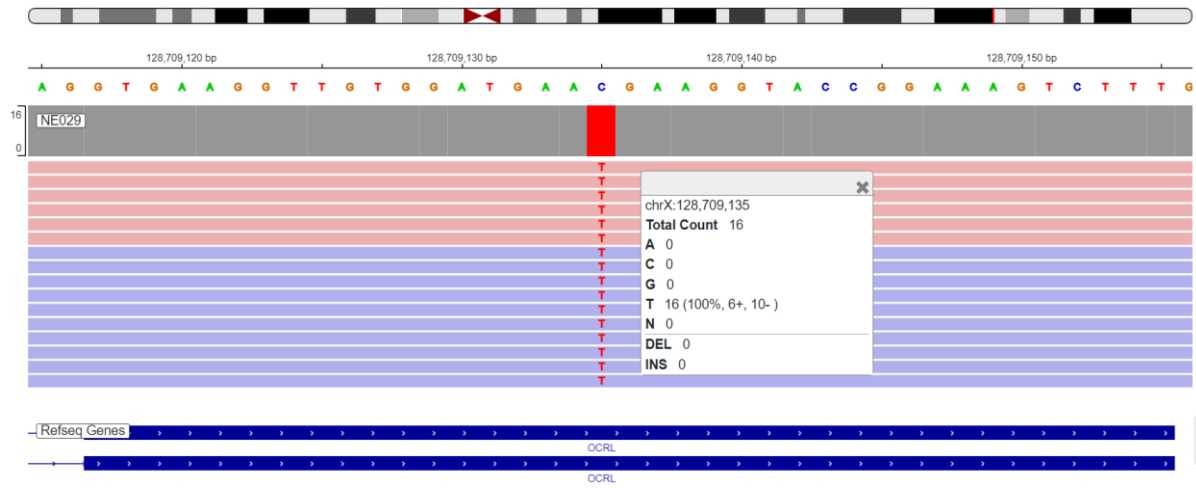


Figure 3.9. IGV image depicting the pathogenic *OCRL*:c1621C>T (p.Arg514Ter) variant identified in Participant NE029. This variant was observed in a hemizygous state and with 16X coverage. Reads are coloured by strand, purple representing reverse (-) and pink forward (+) strands.

Participant NE029 was recruited as a duo (mother and son). The mother of participant NE029 indicated a desire to have more children during consultations with a medical geneticist and therefore underwent carrier testing by Sanger sequencing to confirm if she harboured the same variant as her son. Sanger sequencing confirmed that the mother was not a carrier for the likely pathogenic *OCRL* variant identified in participant NE029 (Figure 3.10). This variant is thus assumed to be *de novo* in Participant NE029.

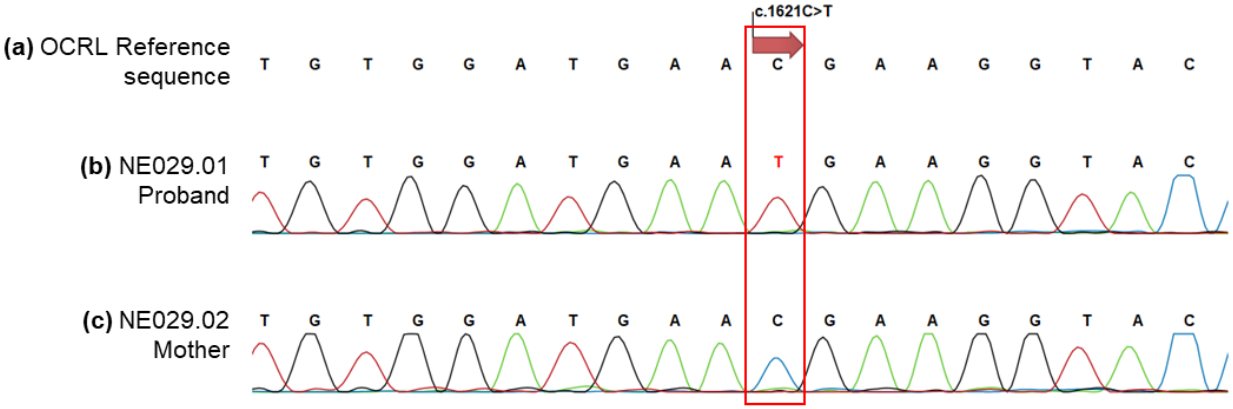


Figure 3.10. Sanger sequencing electropherograms confirming the carrier status of the mother of Participant NE029. The electropherogram confirmed that the mother of Participant NE029 does not carry the pathogenic OCRL variant (*OCRL*:c.1621C>T) identified in the proband. (a) reference sequence (from NCBI), (b) the proband sequence and (c) mother's sequence are shown.

3.4.1.6. Participant NE033

Participant NE033 was a female patient referred by her medical geneticist for exome sequencing at three months of age due to her severe nephrotic disease phenotype. This patient presented with congenital nephrotic syndrome, left ventricular hypertrophy,

right ventricle effusion, peripheral pulmonary stenosis, anasarca, proteinuria, electrolyte abnormalities and hypothyroidism.

A single variant of interest was shortlisted from the manually curated variants prioritised by the NICU/African panel. This shortlisted variant was in the *NPHS1* gene, encoding the nephrin protein involved in kidney glomerular filtration (Wartiovaara et al., 2004). Variants in *NPHS1* are associated with congenital nephrotic syndrome (Finnish type) (Kestila et al., 1998). The identified variant, *NPHS1*:c.1379G>A (p.Arg460Gln), has been reported in a compound heterozygous and homozygous state in multiple patients with nephrotic syndrome in the literature (PM3_strong) (Beltcheva et al., 2001, Koziell et al., 2002, Sako et al., 2005, Heeringa et al., 2008, Lee et al., 2009, Machuca et al., 2010, Schoeb et al., 2010, Mbarek et al., 2011, Mann et al., 2019). This variant affects the immunoglobulin-like 5 domain of the NPHS1 protein, in which no benign missense or frameshift variants are reported on Uniprot (PM1). This variant carries a 2-star pathogenic classification on ClinVar (VCV000056438), with three non-conflicting classifications as pathogenic (PP5_moderate) and is observed at a very low frequency in population databases (PM2_supporting). Considering this evidence, this variant was classified as likely pathogenic according to ACMG/AMP guidelines.

The likely pathogenic variant was observed in a homozygous state and with 87X coverage (Figure 3.11). Participant NE033 presented with features in keeping with a diagnosis of congenital nephrotic syndrome, therefore this variant was determined to be causative in this participant and reported back to the clinician and family (Table 3.6). Participant NE033's disease presentation was very severe; however, this participant was unable to access renal replacement therapy in the South African State healthcare system. Considering the poor prognosis of this participant and paucity of available treatments and medical interventions, the participant's family were counselled on comfort care. The parents of Participant NE033 were offered prenatal testing for any future pregnancies, owing to the recessive nature of the condition.

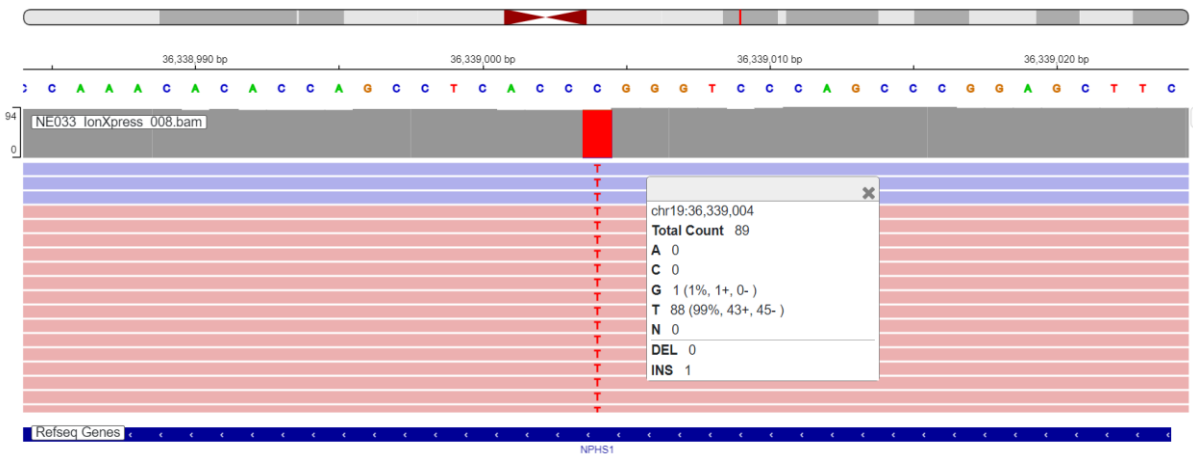


Figure 3.11. IGV image depicting the pathogenic *NPHS1*:c.1379G>A (p.Arg460Gln) variant identified in Participant NE033. Transcription of the *NPHS1* gene occurs in reverse direction, indicated by the <<< arrows in the gene box, therefore, the G>A change is represented visually as C>T. This variant was observed in a homozygous state and with 89X coverage. Reads are coloured by strand, purple representing reverse (-) and pink forward (+) strands.

3.4.1.7. Participant NE034

Participant NE034 was a female patient referred for testing by her neonatologist at four days old due to her skeletal abnormalities. This participant presented with a prominent occiput, restricted elbow extension, a proximally inserted thumb, bilateral tapered fingers, sacral dimple, fixed bilateral hip and knee extension, bilateral talipes equinovarus, metaphyseal widening of the right femur, a hypoplastic left tibia, reduced plantar creases and hypoplastic nails on both feet. The participant was also noted to have a midshaft fracture at the right femur. Participant NE034 was the second child to this couple with lower limb abnormalities, indicative of a possible genetic cause for the clinical presentation in this family.

Two variants of interest were shortlisted using the first filtering strategy, the NICU/African panel. The first shortlisted variant, *PAH*:c.890G>A (p.Arg297His), was a missense variant in the *PAH* gene, encoding the phenylalanine hydroxylase enzyme, a key enzyme in phenylalanine catabolism. Variants in *PAH* are associated with autosomal recessive phenylalanine intolerance and phenylketonuria, an inborn error of metabolism (Zurfluh et al., 2008). The shortlisted variant identified in Participant NE034 carries a 3-star pathogenic classification on ClinVar (VCV000092750) after review by the Phenylketonuria Variant Curation Expert Panel (PP5_strong). This variant was observed in a heterozygous state. Participant NE034 was therefore determined to be a carrier for the pathogenic *PAH* variant, a non-reported incidental finding (Table 3.8).

The second shortlisted variant, *ARSA*: c.434G>C (p.Arg145Pro), is missense variant in the arylsulfatase A encoding gene, *ARSA*, which encodes a lysosomal hydrolase enzyme implicated in the catabolism of sulfatides, a glycophospholipid component of myelin (Cesani et al., 2016). The variant was excluded as causative as it was observed in a heterozygous state with no additional variants in the *ARSA* gene identified and was not in keeping the phenotypic presentation of the participant (Table 3.8).

Table 3.8. Shortlisted variants identified in Participant NE022 prioritized through the NICU/African virtual panel and the ClinVar filtering strategies. The variant, zygosity, coverage, variant type, associated phenotype, applicable ACMG/AMP codes and the interpretation of the variant is provided. AR = autosomal recessive, AD = autosomal dominant

Gene (OMIM number)	Variant	Zygosity and Variant Coverage	Variant Type	Associated Phenotype	ClinVar Classification and Accession ID	Applicable ACMG/AMP Codes and Interpretation
PAH (612349)	c.890G>A (p.Arg297His)	Heterozygous 87X	Missense	AR phenylketon-uria and hyperphenyl alaninemia	Pathogenic (3-star) VCV-000092750	PP5_strong, PS3, PS1, PP3, PM2_supporting Heterozygous carrier (Incidental finding) Not reported
ARSA (607574)	c.434G>C (p.Arg145Pro)	Heterozygous 196X	Missense	AR meta-chromatic leuko-dystrophy	-	Heterozygous, Not reported
TRPV4 (605427)	c.806G>A (p.Arg269His)	Heterozygous 73X	Missense	AD spondylo-metaphyseal dysplasia	Pathogenic (2-star) VCV-000005000	PS3_strong, PM5_moderate, PM1_moderate, PP5_moderate, PS4_strong Pathogenic, Reported

As a causative variant was not identified from the first filtering strategy, the ClinVar filter, was employed. Two variants of interest were shortlisted from 43 prioritised variants, one being the *PAH* variant previously identified.

The second shortlisted variant was a missense variant in the *TRPV4* gene, *TRPV4*:c.806G>A (p.Arg269His). *TRPV4* encodes a calcium ion channel widely

expressed in the kidneys, lungs, brain, endothelial cells, and sensory ganglia (Kwon et al., 2023). Variants in *TRPV4* have been associated with a broad range of autosomal dominant skeletal dysplasias and neuromuscular disorders (Krakow et al., 2009, Auer-Grumbach et al., 2010, Deng et al., 2010, Landoure et al., 2010, Nishimura et al., 2010, Cho et al., 2012). The identified variant, *TRPV4*:c.806G>A (p.Arg269His) has been reported in multiple patients with skeletal disorders in the literature in a heterozygous state (PS4) (Auer-Grumbach et al., 2010, Deng et al., 2010, Landoure et al., 2010, Zimon et al., 2010, Echaniz-Laguna et al., 2014, Louis et al., 2014, Biasini et al., 2016, Fleming and Quan, 2016, Jedrzejowska et al., 2019).

This variant, in the fourth exon of *TRPV4*, occurs in the ARD domain, an ankyrin repeat domain, previously implicated in the normalisation of TRPV4 activity (PM1) (Landoure et al., 2010, Takahashi et al., 2014). The variant is predicted to be deleterious by REVEL and MetaRNN *in-silico* predictors, supported by *in vitro* functional studies which show altered calcium channel activity and decreased inositol binding and sensitivity as a result of this variant (PS3) (Deng et al., 2010, Landoure et al., 2010, Fecto et al., 2011, Klein et al., 2011, Takahashi et al., 2014). Alternative variants affecting the same arginine amino acid residue, p.Arg269Ser, p.Arg269Leu and p.Arg269Cys, have been reported as pathogenic in the literature, on ClinVar (VCV000005002) and on Uniprot, indicating the possible importance of this residue in channel activity (PM5) (Berciano et al., 2011, Evangelista et al., 2015, Vill et al., 2015). The *TRPV4*:c.806G>A (p.Arg269His) variant has been reported on ClinVar previously and carries a 2-star pathogenic classification (PP5_moderate). Considering this evidence, this variant was classified as pathogenic.

The pathogenic *TRPV4* variant was observed in a heterozygous state in Participant NE034, with coverage of 73X (Figure 3.12). Participant NE034 presented with skeletal features suggestive of a *TRPV4*-disorder, therefore the identified variant was determined to be causative in this participant and a diagnosis of spondylometaphyseal dysplasia reported back to the clinician and family (Table 3.6 and Table 3.8). Participant NE034 did not present with the neuropathic features commonly associated with *TRPV4*-disorders, however these have been reported to manifest after the neonatal period (Fleming and Quan, 2016). This participant will need to be monitored for the development of these features and require physical therapy to maintain function of the lower limbs.

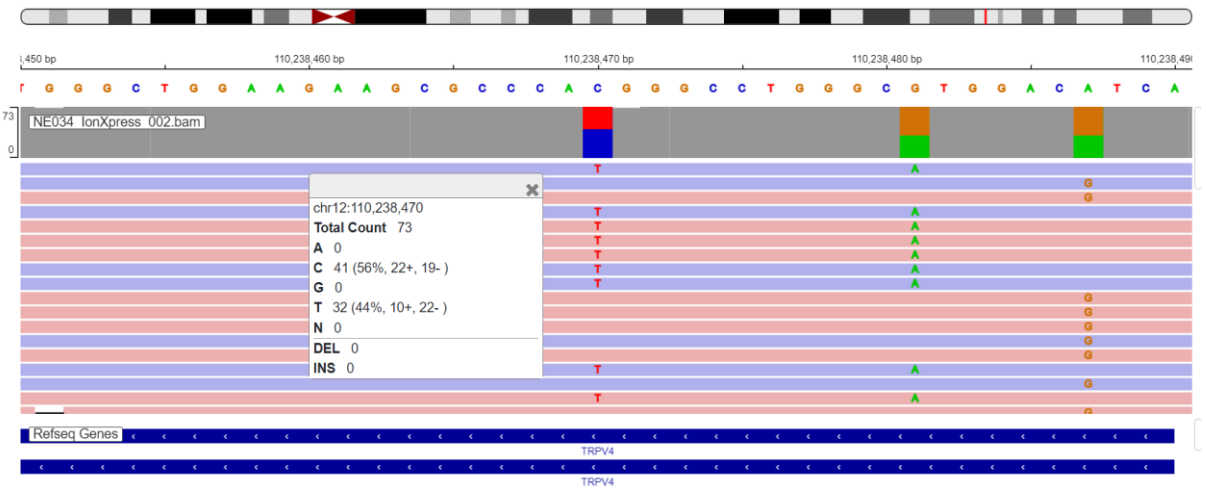


Figure 3.12. IGV image depicting the pathogenic *TRPV4*:c.806G>A (p.Arg269His) variant identified in Participant NE034. Transcription of the *TRPV4* gene occurs in the reverse direction, indicated by the <<< arrows in the gene box, therefore, the G>A change is represented visually as C>T. his variant was observed in a heterozygous state and with 73X coverage. Reads are coloured by strand, purple representing reverse (-) and pink forward (+) strands.

3.4.2. Participants Without a Causative Variant Identified

Twenty-five participants did not receive a molecular diagnosis after all three filtering strategies. The participant indications for testing, suspected diagnoses, the number of variants retained with each filter and the number of variants prioritised through manual filtering for manual curation are summarised in Table 3.9. The detailed clinical presentations of these participants are summarised in Appendix I.

The shortlisted but excluded and/or not reported variants, are summarised in Table 3.10. Six participants had no variants of interest identified from the variants shortlisted using the three filtering strategies. Twenty-four of the shortlisted variants occurred in genes associated with homozygous conditions but were observed in a heterozygous state, with a second pathogenic or likely pathogenic variant not identified. High confidence CNVs encompassing these genes were also not identified and could not account for a second causative variant. These variants were therefore excluded as causative.

Six variants were observed in an allelic ratio consistent with the zygosity of the associated condition, but the participants' phenotype was not fitting of the associated condition. These variants were also excluded as they could not be concluded as disease-causing in the associated participants.

Pathogenic variants in the *HBB* and *LDLR* genes were observed in Participants NE006 and NE035. The shortlisted heterozygous *HBB* variant, *HBB*:c.20A>T (p.Glu7Val) was prioritised through all three filtering strategies and is well documented as the most common variant associated with sickle cell disease. Homozygous genotypes present with sickle cell anaemia and heterozygotes with sickle cell trait (Williams and Weatherall, 2012). This variant, classified as pathogenic, was observed in this participant in a heterozygous state, conferring a diagnosis of sickle cell trait. As this diagnosis did not explain the presenting dysmorphic phenotype of Participant NE006, this finding was ruled an incidental finding and not reported back to the participant. The mother of participant NE006, who is of Ghanaian ancestry, was a known carrier of sickle cell trait.

The LDLR variant identified in Participant NE035, *LDLR*:c.1291G>A (p.Ala431Thr), was a missense variant in the low density lipoprotein receptor (LDLR) encoding gene, which plays a role in cholesterol homeostasis. The identified variant carries a 3-star pathogenic classification on ClinVar and has been reviewed by the Clingen Familial Hypercholesterolemia Variant Curation Expert Panel (PP5_strong). As symptoms of hypercholesterolemia may develop later in life, this variant was concluded to be an incidental finding and also not reported back to the participant's family.

Table 3.9. Demographics, clinical indications and the number of variants retained and prioritised across the three filtering strategies amongst participants who did not receive a molecular diagnosis. G = Geneticist, N/P = Neonatologist/Paediatrician, M = male, F = female

Study ID	Referring Clinician	Sex	Ancestry	Consang-uineous Parents	Primary Indication for Testing	Suspected Clinical Diagnosis	Variants Retained			Prioritised variants			
							NICU/ African Panel	ClinVar Panel	DDG2P Panel	Total	NICU/ African Panel	ClinVar Panel	DDG2P Panel
NE002	G	M	African	No	Cardiac	-	5824	76	9961	183	81	49	112
NE004	G	M	African	No	Developmental	RASopathy	5098	133	9029	214	62	44	145
NE005	G	F	African	No	Cardiac	-	5679	108	10041	264	81	66	171
NE006	N/P	F	African	No	Dysmorphic	Turner syndrome	5739	108	1041	253	106	42	151
NE007	N/P	M	African	No	Multiple congenital anomalies	Trisomy 18	5577	109	9944	224	74	50	148
NE009	N/P	M	African	No	Neuromuscular	Congenital myopathy	5461	99	9733	200	56	40	131
NE010	N/P	M	African	No	Cardiac	-	6013	129	10042	269	104	75	159
NE011	N/P	M	African	No	Dysmorphic	-	5764	109	10209	226	52	54	156
NE012	N/P	F	European	No	Dysmorphic	-	4596	70	8438	158	36	31	117
NE013	N/P	F	African	No	Abnormal growth	-	5678	130	9925	229	42	77	145
NE014	N/P	F	African	No	Multiple congenital anomalies	Trisomy 13	5439	108	9990	237	87	64	158
NE015	G	F	African	No	Multiple congenital anomalies	OEIS complex	5745	85	10286	210	62	48	140
NE016	N/P	M	African	No	Neurological	-	5659	98	10255	238	82	54	159
NE017	N/P	M	African	No	Renal	-	5596	117	9801	253	94	66	147

NE018	N/P	F	Other	No	Cardiac	-	5288	92	9242	225	79	54	140
NE019	N/P	F	African	No	Multiple congenital anomalies	-	5592	126	9922	240	77	72	149
NE020	N/P	M	African	No	Hepatic	-	5760	85	10352	274	74	67	177
NE023	N/P	F	African	No	Multiple congenital anomalies	-	5760	85	10352	269	101	76	166
NE025	N/P	F	Other	Yes	Gastrointestinal	-	4779	65	8341	189	64	21	104
NE027	N/P	F	African	No	Hepatic	Inborn error of metabolism	1980	43	124	183	61	30	124
NE028	G	M	African	No	Multiple congenital anomalies	-	5635	65	9888	200	73	44	137
NE030	N/P	M	African	No	Multiple congenital anomalies	Trisomy 18	5758	74	10215	257	73	53	197
NE031	N/P	F	African	No	Cardiac	-	6107	97	9453	211	85	64	130
NE032	N/P	F	African	No	Cardiac	-	6025	146	10496	284	99	80	166
NE035	G	M	Other	Yes	Multiple congenital anomalies	-	4533	54	8223	154	58	23	97

Table 3.10. Shortlisted variants identified in the 25 participants who did not receive a confirmed molecular diagnosis from SNV filtering and analysis. The participant study ID, variant name, zygosity, coverage, variant type, associated phenotype, applicable ACMG/AMP codes and the interpretation of the variant is provided.

Study ID	Gene (OMIM number)	Variant	Zygosity and Variant Coverage	Variant Type	Associated Phenotype	Interpretation [$]
NE002	-	-	-	-	-	No variants of interest identified
NE004	-	-	-	-	-	No variants of interest identified
NE005	*HEXB (606873)*	c.1597C>T (p.Arg533Cys)	Het, 84X	Missense	AR Infantile Sandhoff disease	Heterozygous, Did not fit phenotype
NE006	*HBB (141900)*	c.20A>T (p.Glu7Val)	Het, 10X	Missense	AD beta-thalassemia and AR sickle cell trait	PP5_moderate, PS4, PS3 Incidental finding, Not Reported
	ACADS (606885)	c.529T>C (p.Trp177Arg)	Het, 206X	Missense	AR butryl-CoA-dehydrogenase deficiency	Heterozygous, Did not fit phenotype
NE007	-	-	-	-	-	No variants of interest identified
NE009	*GRM6 (604096)*	1336C>T (p.Arg446Ter)	Het, 203X	Stop-gain	AR congenital stationary night blindness	Heterozygous, Did not fit phenotype
NE010	*SH3TC2 (608206)*	c.58G>T (p.Glu20Ter)	Het, 191X	Stop gain	AR Charcot-Marie-Tooth disease type 4C	Heterozygous, Did not fit phenotype
	NPHP1 (607100)	c.939+1G>T	Het, 109X	Splice donor	AR Nephronophthisis and	Heterozygous, Did not fit phenotype
	THOC6 (615403)	c.483+1G>A	Het, 408X	Splice donor	AR Beaulieu Boycott Inness syndrome	Heterozygous, Did not fit phenotype
	ATP6AP1 (300197)	c.938A>G (p.Tyr313Cys)	Hem, 13X	Missense	XLR immunodeficiency 47	Did not fit phenotype

NE011	*LRPPRC (607544)*	c.805dup (p.Thr269AsnfsTer11)	Het, 146X	Frameshift	AR mitochondrial complex IV deficiency	Heterozygous, Did not fit phenotype
	MCPH1 (607117)	c.233+2T>C	Het, 70X	Splice donor	AR primary microcephaly 1 (head circumference 3 to 11 SD below mean)	Heterozygous, Not reported - Remained variant of interest
NE012	*WFS1 (614296)*	c.2053C>T (p.Arg685Cys)	Het, 97X	Missense	AR Wolfram syndrome 1 and AD Wolfram-like syndrome	Did not fit phenotype
	BTD (606019)	641A>G (p.Asn214Ser)	Het, 55X	Missense	AR biotinidase deficiency	Heterozygous, Did not fit phenotype
NE013	-	-	-	-	-	No variants of interest identified
NE014	*LFNG (602576)*	c.219+G>A	Het, 184X	Splice donor	AR spondylocostal dysostosis 3	Heterozygous, Did not fit phenotype
NE015	*NPHS2 (604766)*	c.779T>A (p.Val260Glu)	Het, 66X	Missense	AR steroid resistant nephrotic syndrome	Heterozygous, Did not fit phenotype
NE016	*NCAPD2 (615638)*	c.-23-1G>A	Het, 147X	Splice acceptor	AR microcephaly with short stature	Heterozygous, Did not fit phenotype
	RAD50 (604040)	c.2185_2186dup (p.Met729IlefsTer8)	Het, 156X	Frameshift	AR Nijmegen breakage syndrome like disorder	Heterozygous, Did not fit phenotype
	RNASET2 (612944)	c.492+2T>G	Het, 24 X	Splice donor	AR leukoencephalopathy without megalocephaly	Heterozygous, Did not fit phenotype
NE017	*WRN (604611)*	c.2704del (p.Tyr902IlefsTer72)	Het, 54X	Frameshift	AR progeroid syndrome	Heterozygous, Did not fit phenotype
	GALNS (612222)	c.680T>A (p.Phe227Tyr)	Het, 139X	Missense	AR mucopolysaccharidosis	Heterozygous, Did not fit phenotype
NE018	*BAZ2B (605683)*	c.2770+2T>C	Het, 44X	Splice donor	AD neurodevelopmental disorder	Did not fit phenotype

NE019	*VPS13B (607817)*	c.2381del (p.Pro794GlnfsTer33)	Het, 68X	Frameshift	AR Cohen syndrome	Heterozygous, Did not fit phenotype
NE020	-	-	-	-	-	No variants of interest identified
NE023	*CD36 (173510)*	c.748+2T>A	Het, 59X	Splice donor	AR CD36 deficiency	Heterozygous, Did not fit phenotype
NE025	*KMT2D (602113)*	c.5468-1G>A	Hom, 255X	Splice acceptor	AD Kabuki syndrome 1	Did not fit disease mechanism
NE027	-	-	-	-	-	No variants of interest identified
NE028	*FTCD (606806)*	c.1607T>A (p.Leu536Ter)	Het, 118X	Stop gain	AR formiminoglutamic aciduria	Heterozygous, Did not fit phenotype
NE030	*MYCN (164840)*	c.1270A>G (p.Lys424Glu)	Het, 238X	Missense	AD Feingold syndrome	Did not fit phenotype
NE031	*SCNM1 (620107)*	c.508del (p.Ala170ProfsTer13)	Het, 140X	Frameshift	AR orofaciodigital syndrome	Heterozygous, Did not fit phenotype
NE032	*TBX19 (604614)*	c.387del (p.Asn130Thrfs Ter7)	Het, 79X	Frameshift	AR adenocorticotropic hormone deficiency	Heterozygous, Did not fit phenotype
	DGUOK (601465)	c.268C>T (p.Gln90Ter)	Het, 42X	Stop gain	AR mitochondrial depletion syndrome	Heterozygous, Did not fit phenotype
	ASL (609310)	c.281G>A (p.Arg94His)	Het, 116X	Missense	AR argininosuccinic aciduria	Heterozygous, Did not fit phenotype
	ACAD11 (614288)	c.37G>T (p.Glu13Ter)	Het, 30X	Stop gain	AR developmental and epileptic encephalopathy	Heterozygous, Did not fit phenotype
NE035	*LDLR (606945)*	c.1291G>A (p.Ala431Thr)	Het, 240X	Missense	AR/AD familial hypercholesterolemia	PS3, PS4, PP3, PP5_strong, PP4, PM2_supporting

						Incidental finding, Not reported
	CTRC (601405)	c.85C>T (p.Arg29Ter)	Het, 122X	Stop-gain	AD susceptibility to pancreatitis	Did not fit phenotype
	ADAMTS9 (605421)	c.3580-8_3580-2del	Hom, 122X	Spice acceptor	AR nephronophthisis related ciliopathy	Phenotype association unclear

[$] Heterozygous variants did not have a second variant in the gene identified and were interpreted as not disease causing. These were therefore not reported to the family

Het = heterozygous, Hom = homozygous, Hem = hemizygous, AR = autosomal recessive, AD = autosomal dominant

3.5. CNV Analysis

A CNV screen was performed using the Ampliseq exome CNV baseline on Ion Reporter. A total of 151 high confidence CNVs (call confidence score of at least 100) were called by the analysis pipeline across the 32 participants, with an average call confidence score of 182.5. Eight participants had no high confidence CNVs called (Figure 3.13). All high confidence CNVs were interpreted according to technical standards for the interpretation and reporting of constitutional copy-number variants from the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Twenty-eight potentially pathogenic CNVs were carried forward for manual curation. Of these 28 CNVs, 26 were found to be called in multiple participants, suggestive of a sequencing or calling artifact, or of common population variants, and were excluded as a likely cause of disease. Two remaining CNVs, in participants NE006 and NE017 (Table 3.11) were further examined for disease causality.

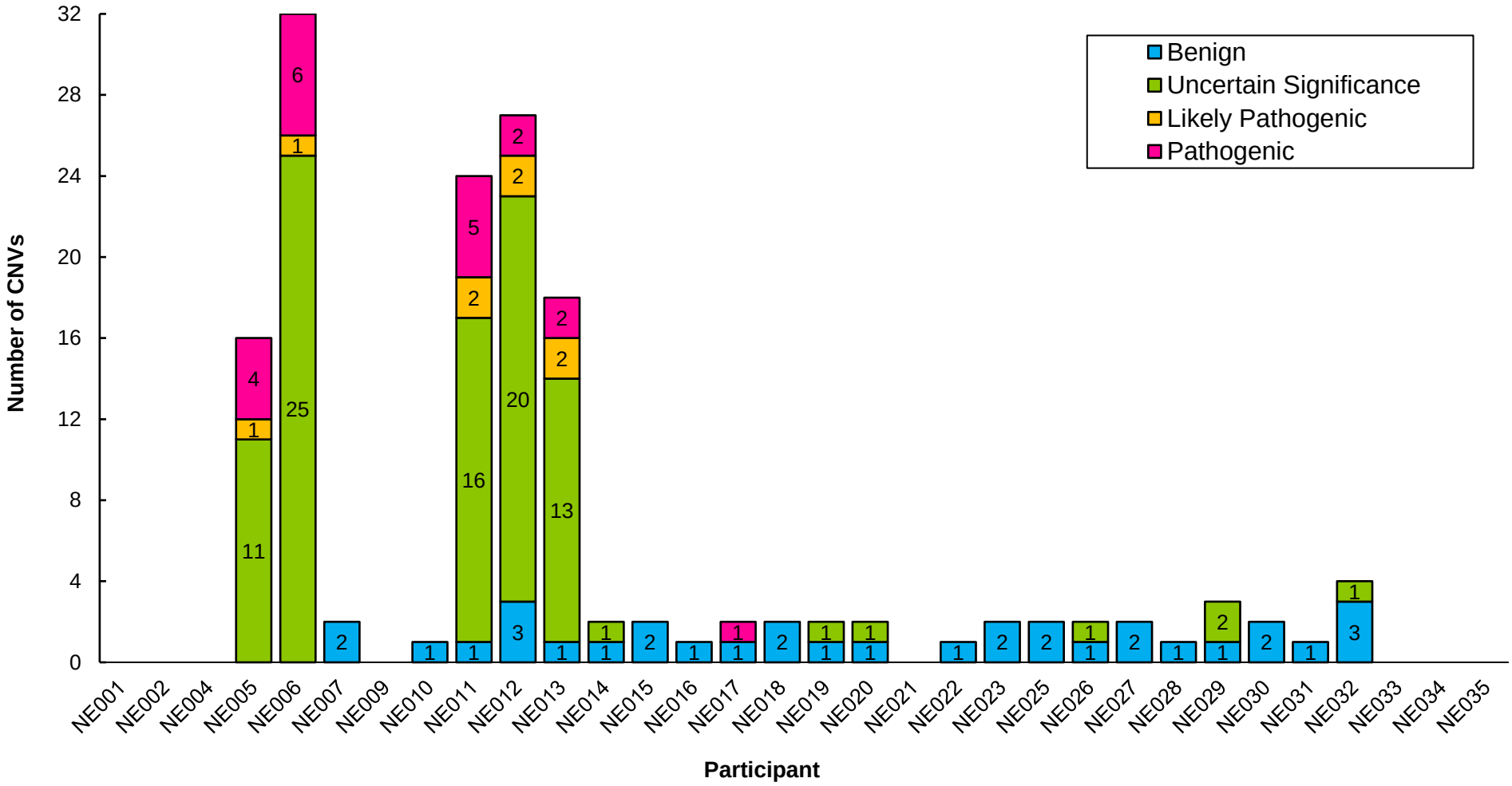


Figure 3.13. High confidence CNVs called in the 32 participants through the Germline Ampliseq™ Exome single sample analysis workflow using the Ampliseq™ exome CNV baseline. High confidence was defined by a confidence score of 100 or more.Variants were classified according to the technical standards and guidelines outlined by the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen).

Table 3.11. Potential pathogenic high confidence CNVs identified in Participants NE006 and NE017. The participant, phenotypic features, variant, variant type, estimated variant size, call confidence score, number of encompassed protein coding genes and the encompassed gene symbols are provided. Variants curated by the Clingen Dosage Sensitivity Working Group are indicated with asterisks (*).

Study ID	Phenotypic Features	Variant HGVS	Variant Type and Size (bp)	Call Confidence Score	Number Genes Affected	Gene Symbols
NE006	Hypodensities in left occipital lobe, low hairline, low set ears, submucosal cleft palate, small chin, microcephaly, short neck, horseshow kidney, contractures, global brisk reflexes, history of feeding and swallowing difficulties	19q13.12 (36164206-36222832) x1	Deletion 58626	106.85	3	*KMT2B**, *UPK1A*, *ZBTB32*
NE017	Utero-pelvic obstruction and bilateral hydronephrosis	17p13.1p12 (10233685-15142948) x1	Deletion 4 909 263	953.03	23	*ADPRM*, *ARHGAP44*, *CDRT15*, *COX10*, *DNAH9*, *ELAC2*, *HS3ST3A1*, *HS3ST3B1*, *MAP2K4*, *MYH1*, *MYH13*, *MYH2*, *MYH3*, *MYH4*, *MYH8*, *MYOCD*, *PIRT*, *PMP22**, *SCO1*, *SHISA6*, *TMEM220*, *TMEM238L*, *ZNF18*

3.5.1. Participant NE006

A 58 kb deletion was identified in participant NE006, who presented with dysmorphic features (summarised in Table 3.11 and in Section 3.4.2.4). This deletion encompassed 3 protein coding genes, including *KMT2B* (Table 3.11 and Figure 3.14). The *KMT2B* gene has been curated by the ClinGen Dosage Sensitivity Working Group as haploinsufficient and is associated with generalised childhood-onset dystonia (Zech et al., 2016).

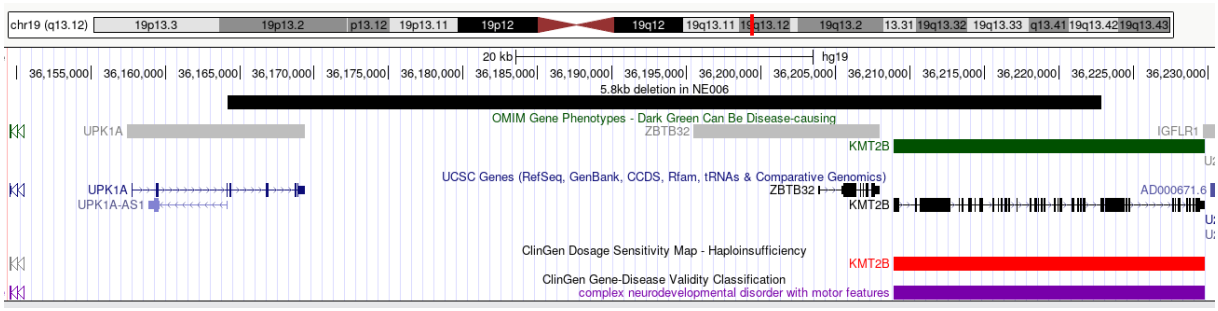


Figure 3.14. Genomic location of the approximately 5.8 kb deletion detected in participant NE006. Viewed in the UCSC Genome Browser, accessed 30 January 2024 (Kent et al., 2002). This deletion (black) encompasses three protein coding genes, seen in the OMIM (green) and UCSC (blue) tracks. *KMT2B*, of which 26 exons is deleted, has been curated for haploinsufficiency by ClinGen (red track) and has a curated association with a complex neurodevelopmental disorder with motor features (purple track).

This deletion removed the first 26 of 37 exons of the *KMT2B* gene, removing four key protein domains: the SET-binding domain, PHD domain, PHD-like domain and the FYRN domain (Zech et al., 2019, Cif et al., 2020). Lower limb dystonia has been reported as the most common clinical manifestation of *KMT2B*-dystonia, along with microcephaly, an elongated face and bulbous nasal tip (Gorman et al., 2018, Kawarai et al., 2018, Cif et al., 2020). Multiple microdeletions encompassing the *KMT2B* gene have been reported in the literature, with almost all cases reported to present with lower limb dystonia and dysarthria, and multiple reports of swallowing difficulties (Meyer et al., 2017, Zhao et al., 2018, Carecchio et al., 2019, Cif et al., 2020). Hypodensities of the globus pallidi have also been reported in multiple patients through magnetic resonance imaging (MRI), however, these hypodensities seem to be age dependent and diminish with age (Meyer et al., 2017).

Participant NE006 presented with arthrogryposis, elbow contractures, feeding difficulties, and occipital lobe hypodensities, which may be explained by the deletion in the *KMT2B* gene. No reports of kidney malformations, present in this participant, have

been reported in the literature in patients diagnosed with *KMT2B*-related dystonia and this clinical feature remains unexplained in Participant NE006. Notably, this participant did not present with the characteristic elongated face and bulbous nose associated with *KMT2B*-dystonia. Further clinical assessment is needed to determine if the participant has manifested any further clinical features indicative of progressive *KMT2B*-dystonia. Further genetic testing is needed to confirm the presence of this CNV and to further clarify the breakpoints.

3.5.2. Participant NE017

A 4.9 Mb deletion was identified in Participant NE017 (Table 3.11). This deletion encompassed 23 protein coding genes (Figure 3.15). A single gene in this region, *PMP22*, has been curated by the ClinGen Dosage Sensitivity Working Group as haploinsufficient. This deletion removed the last two of four exons of the *PMP22* gene. PMP22 has been found to play a dosage sensitive role in Charcot Marie Tooth disease type 1A (CMT1A), as well as Hereditary Neuropathy with liability to Pressure Palsies (HNPP), with duplications of the gene associated with CMT1A and deletions of the gene associated with HNPP (Lupski et al., 1991, Chance et al., 1993).

HNPP, also known as tomaculous neuropathy, is a mononeuropathy which results in the thickening of peripheral nerve myelin into tomaculae structures. This mononeuropathy typically affects the wrists, elbows, knees and shoulders, and usually has an onset in adolescence with a high penetrance (Chance, 2006). Participant NE017 did not present with any neuropathic clinical features typical of HNPP but with hydronephrosis and bilateral pelvic-utero obstruction (summarised in Table 3.13 and Section 3.4.2.14).

Kidney malformations have been reported twice in the literature in association with deletions of the 17p12 region. Two children, both aged 13, received a diagnosis of HNPP and presented with posterior urethral valves and renal hypodysplasia respectively (Figure 3.15) (Caruana et al., 2015, Verbitsky et al., 2015). Both children, harbouring a 1.3 Mb deletion of the 17p12 region and received a diagnosis of HNPP, however the neuropathic features of these patients were not reported. No reports of hydronephrosis or pelvic-utero obstruction have been reported in patients with a similar deletion.

A CMA was performed on this participant to validate the presence of this large deletion and to better estimate its breakpoints. Microarray confirmed the presence of a 4.8 Mb

deletion, 17p13.1p12 (10237819-15043412)x1, however the breakpoints did not include the *PMP22* gene. Therefore, this CNV could not be determined as the cause of this participant's presenting phenotype and was concluded to be an incidental finding.

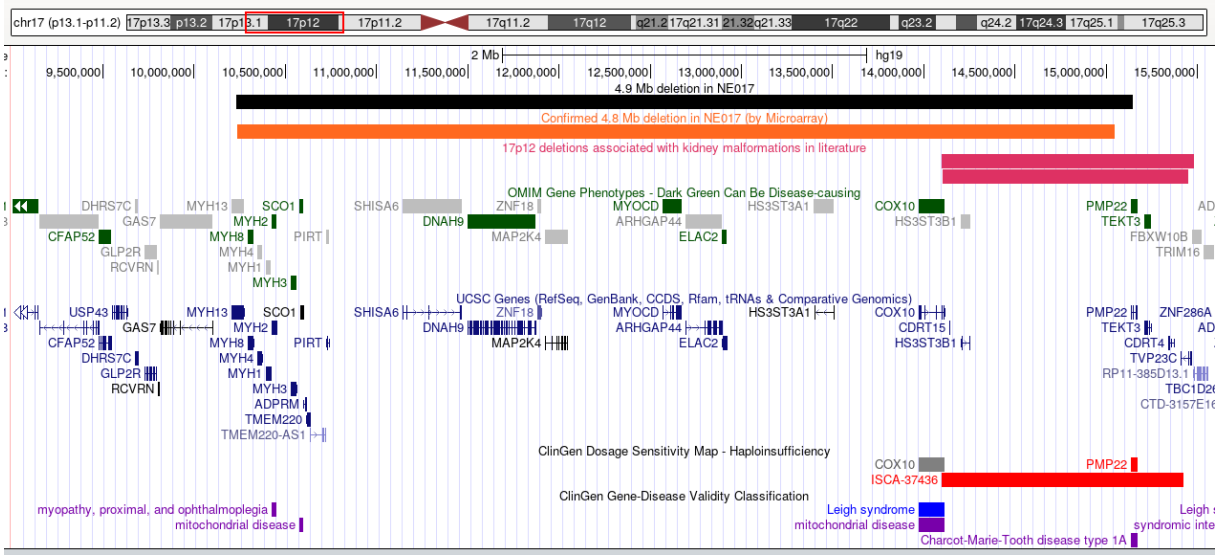


Figure 3.15. Genomic location of the approximately 4.9 Mb deletion identified in participant NE017. Viewed in the UCSC Genome Browser, accessed 30 January 2024 (Kent et al., 2002). This 4.9 Mb (black) deletion encompasses 23 protein coding genes, seen in the OMIM (green) and UCSC (blue) tracks. PMP22 is the only gene in this region that has been curated for haploinsufficiency by ClinGen (red track) and has a curated definitive association with Charcot Marie Tooth Disease type 1A purple track). This deletion was orthogonally confirmed by microarray in this participant (orange track). Two other deletions with kidney and urethral malformation associations which overlap with this region have been reported in the literature (Caruana et al., 2015, Verbitsky et al., 2015).

Neither CNV identified in this cohort could definitively be determined to be causative in the associated participant. This CNV screen therefore provided no further increase in the diagnostic yield of our analytical pipeline in this infant cohort.

4. DISCUSSION

The utility of NGS technologies for the diagnosis of critically ill infants has been explored in numerous studies globally, particularly for infants under intensive care in NICUs and PICUs. Between 16 and 22% of all infants admitted to NICUs and PICUs harbour a genetic condition, which impacts their length of hospital stay, need for intensive care and their overall health outcomes (Ceyhan-Birsoy et al., 2019, Wojcik et al., 2019, Kingsmore et al., 2020). Some of the leading causes of hospitalisation and death in infant populations have underlying genetic aetiologies, including congenital anomalies, low birthweight, prematurity, sepsis, and metabolic abnormalities (Wojcik et al., 2019, Kingsmore et al., 2020). These also form the leading causes of infant hospitalisation and death in South Africa (Hoque et al., 2011, Moundzika-Kibamba and Nakwa, 2018, Nieuwoudt et al., 2022, Gabriels and Le Roux, 2023). The contribution of genetic conditions to infant morbidity and mortality is therefore estimated to be significant globally and creates a need for effective diagnostic tools to identify genetic diseases in infants and to inform and direct their clinical management.

NGS technologies have been shown to be effective in diagnosing genetic conditions and in impacting the clinical management of infants and their families. Large cohort studies and randomised controlled trials have proven gene panels, WES and WGS to be effective in diagnosing genetic disorders in up to 70% of infants with features suggestive of a genetic condition and impacting their clinical management in up to 50% of infants tested, regardless of a positive or negative diagnosis (Willig et al., 2015, Kingsmore et al., 2019, Lunke et al., 2020, Krantz et al., 2021, Maron et al., 2023) These studies have largely included participants of European and Asian ancestry in high- and upper-middle-income countries, with little to no representation of individuals from LMICs and individuals of African ancestry. The utility of these tools in diagnosing individuals of African ancestry, as well as the implementation of these tools in LMICs, therefore warrants further investigation, to develop a representative global view of the benefit of NGS in infant medical management.

The uptake and implementation of NGS diagnostic tools is low in LMICs and across the African continent, therefore evidence of the utility in local African contexts is absent (Tekola-Ayele and Rotimi, 2015). The lack of evidence makes support for implementation and utilisation of these technologies from governments, policy makers

and the general public, as well as resources for implementation and maintenance of genetic services, a challenge to obtain (Kamp et al., 2021, Jongeneel et al., 2022). It is necessary to explore the utility of these tools to substantiate their necessity in low-resource settings, where access to basic services, including basic health services, is limited. The implementation of these tools is necessary to bridge the existing healthcare disparities and to achieve equitable access to genetic services.

Genetic services for ill infants and children are scarce and inconsistent across Africa. Services are limited in some countries, including Ethiopia, Rwanda and Mali, where testing is restricted to MLPA, real-time PCR and cytogenetics for the detection of chromosomal abnormalities (Landoure et al., 2010, Uwineza and Mutesa, 2015, Quinonez and Terefework, 2021). Broader, more established genetic services are offered in others, including Kenya, Nigeria and South Africa, in which the most established genetic service on the continent exists (Kromberg et al., 2013, Adeyemo et al., 2018, Gedleh et al., 2018, Zhong et al., 2021b). Established genetic services are also developing in North Africa, in Morocco, Egypt and Tunisia, in which consanguinity remains high and a considerable risk for inheritance of genetic conditions (Belhassan et al., 2016, Elloumi-Zghal and Chaabouni Bouhamed, 2018, El-Attar et al., 2022) The political, social, cultural, and economic landscapes differ significantly between countries in Africa; therefore, a blanket implementation approach is not appropriate for the diverse African setting. Studies which investigate the challenges and benefits of NGS technologies on infant healthcare and outcomes are needed in each local setting, to validate the utility claims frequently reported in high-income settings and to address the various challenges of implementation in these different contexts.

Equitable access to genetic services has the potential to not only directly benefit participants in African healthcare systems, but also the broader genetics community through access to more diverse reference data and more expansive phenotypic data. The paucity of African data, both genetic and phenotypic, increases the difficulty of variant interpretation and can result in the misclassification and misinterpretation of sequencing data in all populations (Baynam et al., 2020). The inclusion of more individuals of African ancestry in NGS testing may further assist in reducing the diagnostic gap between individuals of African ancestry and other ethnicities, in which those of African ancestry are half as likely to receive a genetic diagnosis from NGS testing and are faced with longer TATs for NGS-based tests (Choudhury et al., 2020, Bowling et al., 2022, Wright et al., 2023).

As the developed world accelerates its efforts to diagnose genetic conditions in ill infants in rapid and ultra rapid TATs, significant efforts are needed in African countries, and broadly in LMICs, to establish systems to provide genetic services and to prevent the exacerbation of the existing healthcare disparities (Kamp et al., 2021). The translation of NGS technologies into existing healthcare systems to benefit local populations requires careful consideration and should be led by local evidence of utility and successful integration and implementation. This study was conducted to provide the first insights into the implementation of exome sequencing panels in a NICU environment in an under-resourced African setting and to assess its utility under the unique challenges faced in this environment, particularly the limited bioinformatics capacity, limited access to trained medical geneticists, and the necessity for singleton sequencing over trio-sequencing due to the unique family structures and financial implications.

4.1. The Recruited African Infant Cohort

We recruited 32 participants from two hospitals in Johannesburg, South Africa, most of whom had histories of NICU and PICU admissions. All of the participants were less than two years old at enrolment and had a suspicion of a genetic condition from a neonatologist, paediatrician, or medical geneticist. Most of the participants were referred for testing as they presented with multiple congenital anomalies, cardiac abnormalities, neuromuscular phenotypes and/or dysmorphic features. This was concordant with what has been reported globally in similar studies, in which the most frequent indications for genetic testing amongst infants in NICUs are multiple congenital anomalies and neurological phenotypes (Willig et al., 2015, Stark et al., 2016, Petrikin et al., 2018, D'Gama et al., 2022). These two phenotype categories are large and encompass a broad range of specific clinical features, therefore, a direct comparison of specific features observed in our cohort and those reported in the literature is challenging. These two phenotype categories however are identified as broad recognizable features indicative of a potential genetic aetiology and have collectively been reported to account for more than 70% of testing referrals in several infant cohorts (Daoud et al., 2016, Powis et al., 2018, Elliott et al., 2019, de Castro et al., 2020, Gubbels et al., 2020, Krantz et al., 2021, D'Gama et al., 2022).

The overall prevalence of these clinical features amongst patients admitted to South African NICUs and PICUs is under-investigated, as well the prevalence of underlying genetic causes. Congenital anomalies are not reported as a common cause for NICU

admission in South Africa, where prematurity, hypoxia and neonatal infection are most frequently reported as drivers for the need for intensive care for infants (Hoque et al., 2011). Despite their low reported prevalence in South African NICUs, congenital anomalies have been identified as the fourth most common cause of death amongst NICU infants, highlighting their clinical severity and criticality (Hoque et al., 2011, Moundzika-Kibamba and Nakwa, 2018, Nieuwoudt et al., 2022, Gabriels and Le Roux, 2023).

A study in Zambia, another Sub-Saharan African country, found a strong association between the presence of congenital anomalies in infants admitted to the NICU and neonatal death, in which neonates who presented with congenital anomalies had a 1.6 fold higher risk of neonatal death (Tembo et al., 2024). This emphasises the need to address congenital anomalies to achieve the necessary reductions in neonatal deaths to reach the SDGs for neonatal and under-five mortality. Sub-Saharan Africa remains a considerable threat to the realisation of the SDGs as neonatal and under-five mortality are highest in this region (United Nations, 2023). Models of the declining mortality rates in 31 Sub-Saharan African countries indicate that countries with slow rates of mortality reduction will continue this trend and not reach the levels necessary to achieve the SDGs by 2030. These models predict that only five of the 31 Sub-Saharan countries investigated, Kenya, Rwanda, Senegal, Tanzania and Uganda, will reach the under-five mortality goal and only two countries, Rwanda and Tanzania, will reduce their neonatal mortality rates to reach the fewer than 12 deaths per 1000 births target (Mejia-Guevara et al., 2019).

The contribution of underlying genetic causes to severe congenital abnormalities is largely unknown in the South African context and requires investigation. A study at a regional hospital in the KwaZulu-Natal province reported a prevalence of 15.57 in every 1000 births for congenital anomalies and one in every 1000 births for multiple congenital anomalies (Saib et al., 2021). South Africa has no existing registries to monitor the prevalence and aetiology of genetic conditions and the overall contribution of genetic conditions to morbidity and mortality in South African NICUs remains unknown. Infants in this study were referred by clinicians in the NICU for testing making it challenging to determine the true prevalence of genetic conditions in South African NICUs as not all admissions were assessed or tested for potential genetic conditions. The prevalence of infants who did not present with characteristic features suggestive of a genetic condition, those without severe features, those not in a critical condition

under intensive care and those who demised prior to testing were therefore also undocumented.

4.2. The Overall Diagnostic Yield and Utility

This study achieved a diagnostic yield of 22% through the analysis of three consecutive gene panel filtering strategies, providing a definitive diagnosis for seven participants. This yield fell on the lower end of the spectrum achieved by international studies in similar NICU cohorts, in which yields of between 21 and 70% in infant populations have been reported, differing between studies based on cohort specificity, employed NGS method, cohort size and the analytical tools and pipelines utilised (Stark et al., 2016, Dimmock et al., 2020, Gubbels et al., 2020, Wang et al., 2020, Maron et al., 2023)d

The highest diagnostic yields amongst these studies were reported in participants who received trio-WES and trio-WGS and in those who presented with neuromuscular and neurological phenotypes (Elliott et al., 2019, Gubbels et al., 2020, Wang et al., 2020, Halabi et al., 2022). Only 31% of our study cohort had both parents available at the time of enrolment, making the employment of trio-sequencing unfeasible in this study and the contribution of trio-analysis to diagnostic yield undeterminable. Parents were unavailable for consent at enrolment due to variety of circumstances, including the inability to get time off work, residence in a different city or country, absence in the proband's life, and an unwillingness to participate in the research study.

4.2.1. High Yield Phenotype Groups

In our cohort we observed the highest diagnostic yields in participants who presented with skeletal anomalies and neuromuscular features. This was comparable to the highest yield phenotype groups reported in the literature (Stark et al., 2016, Farnaes et al., 2018, Elliott et al., 2019, Gubbels et al., 2020). A single study reported a diagnostic yield of 58% in infants who received singleton sequencing, the highest reported yield for singleton sequencing amongst a broad infant cohort (Stark et al., 2016). Extensive evaluation by the investigating team prior to enrolment, previous microarray testing to rule out structural variants and a phenotype-driven candidate gene list were employed in this study and may have favoured a greater diagnostic yield compared to our study in which screening by a trained medical geneticist was not a hard inclusion criterion. Suspected skeletal dysplasias in this high yield study were reported to have a diagnostic rate of more than 65%, concordant but lower than our

yield of 100% in this phenotype group (Stark et al., 2016). Our cohort however only had two participants who fell within this phenotype category.

Neuromuscular phenotypes, particularly hypotonia, have also been previously reported to have high diagnostic yields in infant cohorts. A trio-WES study in Canada reported a 100% diagnostic yield amongst the seven participants in their cohort who presented with neurological and neuromuscular phenotypes (Elliott et al., 2019). A WGS study in the United States reported a 100% diagnostic yield in infants who presented with musculoskeletal features, but similar to our cohort, had few participants who fell into the phenotype category (only one participant presented with isolated generalised hypotonia) (Farnaes et al., 2018). In another study investigating the use of rapid-WES in critically ill infants, an 80% diagnostic yield was achieved in participants who presented with neuromuscular phenotypes and hypotonia, again highlighting the success of this phenotype group diagnostically in NGS testing methods (Gubbels et al., 2020). These studies highlight the increased prevalence of underlying genetic causes in certain phenotype groups and promote the use of the NGS testing in infants who present with these phenotypic features, and as well as the inclusion of genes associated with these phenotypes in gene panels.

4.2.2. Clinical Utility from NGS Testing

Clinical utility, often inferred from changes in the clinical management of the proband and their family and from test TAT, was not directly measured in our study. Diagnostic utility was primarily utilised as proxy for overall utility.

4.2.2.1. Time-to-a-Result

Laboratory TAT for our clinical exome panel was approximately 5 days, followed by a variable number of days for data analysis which was dependent on the number of variants requiring manual curation. Rapid and ultra-rapid sequencing tests are increasingly being incorporated into genetic services for ill infants in the NICU globally, with TATs of less than 24 hours now achievable (Wang et al., 2020, Kingsmore and Cole, 2022). These rapid TATs allow for more directed management and for clinical intervention to occur sooner. The impact of this early result on clinical management is exemplified in the WGS randomised controlled trial by the NICUSeq study group, in which infants who received earlier diagnostic testing had double the number of changes to their medical care compared to those who received WGS 45 days later (Krantz et al., 2021).

TATs could not be directly measured in our study due to the batching of samples for DNA extraction, library preparation and sequencing, which was performed to minimise laboratory costs. For the rapid TATs reported in high-income settings to be reached in our South African setting considerable efforts would be required to address the plethora of implementation challenges faced and the lack of established baseline NGS sequencing services. A routine NGS-based genetic service, which includes consistent referrals of appropriate patients for established, routine and expansive NGS-testing, would provide a baseline from which testing could be optimised to reach rapid TATs. These services rely on adequate numbers of competent clinicians with appropriate genetics training to recognise and refer appropriate infants, established logistics workflows, established routine diagnostic service which can deliver NGS test results that meet international standards for sensitivity and specificity, and substantial reference data from individuals of African ancestry allowing for rapid and more accurate variant interpretation in individuals of this ancestral group. The lack of adequate genetics training amongst primary care clinicians, the limited availability of trained geneticists, the limited availability of genetic services and the lack of continental African reference data therefore form significant implementation barriers for the establishment of rapid testing services for infants in South Africa.

4.2.2.2. Changes in Clinical Management

Changes in clinical management as a result of a positive molecular diagnosis were not directly followed up in the seven participants who received a diagnosis in this study and their impact on perceived clinical utility not accessed. The clinical utility of a confirmed molecular diagnosis from NGS-testing to probands and their families has been reported in numerous studies in the literature (Petrikin et al., 2018, Freed et al., 2020, Lunke et al., 2020, Dimmock et al., 2021) This utility includes the adaption of treatment regimens, the cancellation or initiation of medical procedures, additional screening, specialist consults, and inclusion in medical trials for new therapies.

In a WES study by Meng et al. (2017), 52% of diagnosed infants experienced changes in their medical management as a result of a molecular diagnosis, including comprehensive genetic counselling for the families of 88% of diagnosed infants, the initiation of comfort or palliative care in 19% of infants, referral to and assessment by specialists in 25% of infants, changes in medication or diet in 7% of infants and the undertaking of major medical procedures, including organ transplants, in 5% of infants (Meng et al., 2017). In a 2023 study, a confirmed molecular diagnosis in 2 ill infants

redirected their medication regimens and allowed for participation in two clinical drug trials, providing utility by broadening the available treatment approaches (Lumaka et al., 2023).

The utility of a confirmed diagnosis extends beyond the proband to parents and families, who can then receive more accurate risk predictions and prenatal testing for future pregnancies, affecting the future medical management of the entire family unit, not just the proband. In a 2017 study investigating the utility of WGS in infants admitted to the NICU in the Netherlands, 2 couples who received a molecular diagnosis for their ill infant reconsidered their reproductive choices due to a confirmed diagnosis and the availability of prenatal testing. These couples as a result of a diagnosis in a child and the available directed medical management of future pregnancies, were provided reassurance that allowed them to consider having more children (van Diemen et al., 2017).

4.2.2.3. Measures of Clinical Utility

Measures of clinical utility of NGS in infant populations vary in the literature, however a recent review of 21 studies investigating genomic medicine for critically ill infants determined five common measures of utility frequently reported by clinicians and investigators across studies (Callahan et al., 2022). The first measure was changes in treatment, which was reported to occur in 14% of infants across the 21 studies. The second measure was redirection to palliative care, estimated to occur in 14% of all diagnosed infants. The third measure was additional screening and referrals to specialists which occurred in 15% of all diagnosed infants. The fourth measure was prognostic information, providing clarity on the future prognoses of the diagnosed infants. This is predicted to have utility in 16% of diagnosed infants. The fifth commonly reported measure of utility was reproductive information, providing recurrence risks to parents for future pregnancies. Reproductive information is estimated to be provided to 27% of families who undergo genomic testing. This review also highlighted the limited reporting of other measures of clinical utility in the literature, including personal utility from the perspectives of probands and families and the utility of negative results (Callahan et al., 2022).

The personal utility of NGS testing is subjective and considers many non-health related outcomes that may benefit the proband and the family. This personal utility can be divided into four domains, as defined by Kohler et al. (2017): affective outcomes,

cognitive outcomes, behavioural outcomes and social outcomes. Affective outcomes consider the utility of a diagnosis to the emotional state of the proband and family, including enhancing acceptance and coping with health risks, mental preparation for predicted future health outcomes, particularly negative outcomes, as well as feelings of responsibility and relief. Cognitive outcomes consider the utility of the information gained through a test result, including knowledge of the result itself, knowledge of the associated condition and knowledge of health status of the proband and the family. Behavioural outcomes consider the utility and practical utilisation of the result, including future planning and reproductive decision making. Social outcomes consider the social support received as a result of the diagnosis. This element of utility considers altruism, social support and access to supportive resources resulting from a diagnosis (Kohler et al., 2017). The personal utility of a diagnosis, and the absence of one, should be considered when evaluating the utility of implementing NGS technologies for ill infants. Personal utility is expansive and may impact many areas of the participants' lives. The subjective personal benefit of a diagnostic result requires investigation in the African context, in which social and cultural beliefs may be significantly impacted.

4.2.2.4. Observed Personal and Clinical Utility in This Cohort

Although extensive investigations into clinical utility were not conducted in this study, some utility and benefit were observed in the seven families who received a confirmed molecular diagnosis. A single participant, NE033, had comfort care initiated as a result of a molecular diagnosis, one of the five commonly reported measures of utility for genomic sequencing in infants. Carrier screening was initiated as a result of the diagnosis of Participant NE029, providing reproductive information for the mother of the participant. The identified X-linked recessive variant in this participant was assumed to be *de novo* as the mother did not carry it, providing reassurance of a low recurrence risk for future pregnancies. Prenatal testing for future pregnancies was offered to the two families of diagnosed infants who demised throughout the course of this study, providing some measure of clinical utility and personal utility to these families. Further changes in management can be monitored in diagnosed families going forward as all families will be followed up by a medical geneticist.

4.2.2.5. Clinical Utility in Infants Without a Confirmed Molecular Diagnosis

The clinical utility in participants who did not receive a diagnosis was not assessed in this study. Utility in infants who receive negative diagnostic results has been reported in the literature, however this is not as frequently explored as utility in infants with a positive result (Kohler et al., 2017). Changes in clinical management as a result of a negative diagnosis in an ill infant include the cessation of further testing and invasive procedures, the termination of ineffective medications and treatments, changes in diet, specialist referrals for follow up care and the initiation of life support (Dimmock et al., 2020, Freed et al., 2020, Gubbels et al., 2020, Lunke et al., 2020, Dimmock et al., 2021, Krantz et al., 2021). These benefits have mostly been reported and measured by clinicians.

Clinician's view on utility in participants who do not receive a result were echoed by families in a study investigating parental views on the return of WES and WGS results (Cakici et al., 2020). In this study 97% of parents considered NGS testing to be useful and provide clinical and personal utility despite only 23% of participants receiving a positive diagnosis for their affected infant. Parents found the additional knowledge from both positive and negative test results to be empowering, allowing them to make informed decisions regarding the future care of their children and future reproductive decisions. These sentiments regarding clinical and personal utility have been expressed in other studies with participants from diverse socioeconomic backgrounds (Halley et al., 2022, Lemke et al., 2023, Smith et al., 2023).

4.3. A Comparison of the Three Filtering Strategies Utilised

We employed a gene panel approach to analyse our exome data, analysing a total of 2831 genes, as well as any variant with previous pathogenic, likely pathogenic or VUS reports on ClinVar. The comprehensive gene panel approach provides an achievable genotype-driven analysis strategy that minimises the cost and time needed for the analysis of exome sequencing data. The analysis of exome data remains a challenge without deep-phenotype data to direct analysis and shortlist variants of interests. Studies show that deep-phenotyping and an increased number of phenotype terms correlate with a greater diagnostic yield in ill infants (Clark et al., 2019). Accurate phenotyping of ill infants is difficult due to atypical, non-specific and indistinct clinical features and requires a level of clinical and genetic expertise to retrieve an appropriate number of relevant phenotypic terms (Wright et al., 2018a, Clark et al., 2019).

Considering the limited genetics expertise and the reliance on untrained primary clinicians in the South African setting, the genotype-first panel approach, which is less reliant on extensive phenotyping, provides a practical alternative to phenotype-driven WES analysis. The virtual panel approach allows for frequent updating of the analysis panels without additional laboratory costs and optimization. This allows for the incorporation of new data from ClinVar, which is continually updated with new variant classifications from global users of the database, and the incorporation of newer iterations of the DDG2P panel, which is periodically updated with information from the literature and observed clinical data.

4.3.1. The NICU/African Panel

Five participants received a diagnosis from the use of our curated NICU/African panel, a yield of more than 15%. Diagnoses of neuromuscular conditions were most common with this panel, in which 8% of the genes have associations with neuromuscular phenotypes. This gene panel had a high content of genes associated with metabolic and multisystemic conditions, however, our cohort consisted of few to no participants who fell into these phenotype categories, which may have considerable effects on the yield and success of this filtering strategy.

The curated genes in the NICU/African panel were primarily selected from two curated lists from newborn screening projects under the NSIGHT consortium, the BabySeq Project and NC NEXUS Project (Ceyhan-Birsoy et al., 2017, Milko et al., 2019). Both newborn focused studies curated genes based on actionability and the onset of symptoms during the neonatal period. The BabySeq panel of 954 genes, associated with neonatal-onset conditions, when utilised in a cohort of 159 well and ill infants identified disease-causing variants in 9.4% of the total infant cohort, producing a diagnostic yield of 16% amongst ill infants, a similar yield to what was observed in our study (Ceyhan-Birsoy et al., 2019). A small proportion of this study cohort, 20%, was comprised of infants in the NICU. The inclusion of more ill infants with a suspicion of genetic diagnosis is therefore expected to increase the overall yield of this panel. Only 1.3% of this cohort consisted of individuals of African American ancestry. The additional 6% of seemingly well patients in whom a pathogenic or likely pathogenic variant was identified in the study had no suspicion of a genetic condition or had symptoms which did not appear in the neonatal period. These seemingly well infants evidence the complexity of clinical presentations in infants with genetic conditions and support the adoption of these technologies not just in critically ill infants with easily identifiable

syndromic features, but for all infants during newborn screening, allowing for early detection and intervention.

The NC NEXUS panel of 466 genes with high actionability in neonates, was also tested in a mixed cohort of well and ill infants, producing a diagnostic yield of 20% in the total cohort, a similar yield to our overall yield in this scoping study (Roman et al., 2020). This panel identified a causative variant in 44% of ill infants in the tested cohort, however, only two phenotype groups, metabolic conditions and hearing loss were included in this study. The specificity of the inclusion criteria for participants would have significantly skewed the diagnostic yield of this panel amongst the ill infant arm of the cohort. As our infant cohort did not contain any participants whose primary indication for testing fell into these phenotype groups, a significant difference between the diagnostic yields is to be expected. The lack of participants in these groups reflects the challenge of recognising these conditions in the South African NICU environment, and how more phenotypically recognisable manifestations, such as neuromuscular or skeletal conditions, may be easier for primary care clinicians to recognise and accurately phenotype.

Interestingly, only a single gene in which a causative variant was identified in our cohort, *COL2A1*, was curated by both newborn screening study groups and included in both the BabySeq and NC NEXUS gene lists. Three of the five genes from the NICU/African panel in which pathogenic variants were identified were curated for their associations with highly actionable, neonatal-onset conditions by the BabySeq study group only: *MTM1*, *OCRL*, and *NPHS1*. This suggests that the larger and broader gene list curated by the BabySeq project, may provide a better starting point for panel design in our setting and may better reflect phenotypes commonly observed in infants in our South African NICUs, despite the higher yield of 44% amongst ill infants observed in the NC NEXUS study.

4.3.2. The ClinVar Filtering Strategy

All seven variants identified as disease-causing in our cohort were previously reported and prioritised using the ClinVar filtering strategy. Two of these variants were in genes not in the NICU/African panel, *SHOC2* and *TRPV4*, and would have been missed had the first filtering strategy been the only strategy employed. The use of the ClinVar filter therefore increased the diagnostic yield in this cohort by 6%. The ClinVar filtering strategy shortlisted the fewest number of variants for manual review between the three

filtering strategies, demonstrating the use of this publicly available variant repository and its value in prioritising variants of interest in a short time frame.

This strategy is limited by the number and variety of variants uploaded into the database, excluding novel variants, and by the accuracy of their interpretation by reporters. The use of higher rated pathogenicity classifications, such as two- and three-star rated classifications, are suggested to be more plausible variants of interest, owing to their review and consensus classification by multiple individuals (Wright et al., 2021). Variants carrying lower minor population allele frequencies and those submitted to repositories by multiple clinical laboratories with appropriate clinical information, are also considered more likely to be correctly classified (Shah et al., 2018, Sharo et al., 2023). All seven variants identified in this study as disease causing carried a minimum two-star rated pathogenic or likely pathogenic classification of ClinVar, a key factor in their timely prioritization, as well as the confidence in the applied pathogenic classification.

Misclassification of variants is a considerable concern when using public variant repositories. Although efforts have been made to standardise variant interpretation, with most laboratories and research groups having adopted the guidelines of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP), classifications ultimately are determined at the discretion of the researcher and are open to interpretation (Richards et al., 2015, Niehaus et al., 2019). These interpretations are heavily influenced by publicly available data, particularly population frequency data. A lack of representative data from populations in LMICs and from continental Africans, can significantly skew the classification of variants, resulting in absent classifications, misclassification and misinterpretation. This not only affects immediate medical care for the affected individual, but also future novel variants classifications as classification codes are influenced by previous pathogenic classifications of genomic regions (Sharo et al., 2023).

Less than 1% of variant submissions to ClinVar are from African organisations, with the limited submitted data representing less than 10 African countries (Mulder et al., 2021). The impact of this lack of representation was seen in a study by Choudhury et al. (2020), in which 54 variants classified as pathogenic or likely pathogenic on ClinVar were discovered to have minor allele population frequencies greater than 5% in an African population, 13 of which had minor allele frequencies of less than 5% reported

in all population groups in GnomAD at the time of submission. Furthermore, almost every individual included in the 314-person high-coverage cohort, carried at least one variant classified as pathogenic on ClinVar, suggestive of high frequencies of misclassification and misinterpretation (Choudhury et al., 2020). ClinVar classifications may provide a good tool for initial variant prioritization, as seen in our study cohort, however other prioritization techniques are necessary when considering variant data from African individuals and when pursuing the discovery of novel variants. The paucity of data from individuals of African ancestry may lead to false conclusions of rarity and novelty, resulting in false pathogenicity classifications.

The inclusion of individuals of African ancestry and the sharing of data from individuals of African ancestry are vital in expanding our understanding of genetic diseases and in our interpretation of genomic data from all individuals, those of African ancestry and those not of African ancestry (Lumaka et al., 2022). The inclusion of data from diverse, representative African populations can decrease TAT for sequencing data analysis for individuals of African ancestry by reducing the number of variants requiring manual review and interpretation. This is a key area of focus in producing rapid-diagnostic tests and in settings where results can impact life-saving medical interventions. The inclusion of more African data can also resolve current conflicting interpretations and misclassification for global populations as observed by the H3Africa Consortium. The consortium was able to resolve 41 conflicting classifications in ClinVar with updated minor allele frequency data from African populations, reclassifying variants previously thought to be pathogenic as benign, owing to population frequencies greater than 5% in African populations (Lumaka et al., 2022). The inclusion of more African data and more diverse African data is needed to create more representative reference databases and to improve variant interpretation across all populations.

4.3.3. The DDG2P Panel

The third filtering strategy we employed was a curated panel of genes with known associations with developmental disorders, the DDG2P panel (April 2023 freeze, accessed thought the Genotype2Phenotype website at https://www.ebi.ac.uk/gene2phenotype/downloads) (Wright et al., 2015). No additional participants received a diagnosis from the use of this third filtering strategy, however six of the seven pathogenic variants identified in this study were in genes found in this gene panel: *MTM1*, *COL2A1*, *OCRL*, *NPHS1*, *SHOC2* and *TRPV4*. This panel, consisting of 2313 genes, if employed first, would have resulted in a diagnostic yield

of 19%, a considerable yield for an initial filtering strategy. The gene not included in this version of the DDG2P panel, *STAC3*, has subsequently been added to later versions of this panel (version 3.12, added in October 2023, accessed from PanelApp at https://panelapp.genomicsengland.co.uk). Employment of an updated version of this panel would therefore identify all seven causative variants identified in this study.

The DDG2P panel when utilised in a largely European cohort of more than 13 000 individuals with a suspicion of a developmental disorder, 4067 of which were infants admitted to the NICU, produced a diagnostic yield of 41% (Wright et al., 2023). This cohort however had a low representation of individuals of African ancestry, with only 1.8% of participants having inferred African ancestry. This panel in our predominantly African ancestry cohort still performed well and produced the highest individual yield of the three filtering strategies utilised in this study. This suggests that this panel would offer the best starting approach of the three virtual gene panels for variant filtering and prioritization in a cohort of ill infants, regardless of the ethnicity.

Combining the DDG2P panel with the five additional African-specific genes in the NICU/African panel (described in Section 2.7.1.1) would provide the highest diagnostic yield for our cohort and may be the best starting filtering approach in the future for infant cohorts of this nature in the South African NICU setting.

4.4. Outcomes of the CNV Screen

A CNV screen was employed to identify potential CNVs in our infant cohort. This screen identified two potentials CNVs in two participants who had no molecular findings from the SNV analysis. These CNVs required confirmatory testing to provide confidence to the variant calls, which would allow for these potentially pathogenic variants to be formally classified. A microarray was performed to validate the identified 4.9 Mb deletion in Participant NE017; however, a phenotypic association could not be determined and this CNV was concluded to be an incidental finding.

Microarrays were performed in five of the participants prior to exome panels, all yielding negative results. CNV screening was performed on all participants, including those who had received negative results on CMA. WGS and WES can provide better coverage of some genomic regions compared to microarray, yielding additional diagnoses and utility. This was seen in a study by Vissers et al. (2017) in which two participants with negative microarray results had pathogenic CNVs identified by WES due to better coverage of the affected chromosomal region. CMAs have been largely

superseded by exome and genome sequencing as the recommended first-line test for infants who present with congenital anomalies, intellectual disability and developmental delay (Srivastava et al., 2019, Manickam et al., 2021). These genome wide tests allow for the detection of both SNVs and CNVs compared to CNV-only analysis achieved with microarray, producing higher diagnostic yields for a single test (Willig et al., 2015, Meng et al., 2017, Vissers et al., 2017, Farnaes et al., 2018, Lionel et al., 2018, Petrikin et al., 2018). A 2018 meta-analysis found those who underwent WGS and WES to be 8 fold more likely to receive a diagnosis than those who received CMA, validating the use of a comprehensive NGS approach in the screening and diagnosis of ill infants (Clark et al., 2018).

The detection of CNVs from exome data is challenging due to non-uniform coverage, the positioning of breakpoints and frequent false copy number interpretations at GC-rich regions (Zhao et al., 2013). Despite these challenges, CNVs detected from exome data are predicted to increase the diagnostic yield from SNV analysis in infants by 5-6% (Suzuki et al., 2022, Wright et al., 2023). Advancements in calling and mapping tools are improving the performance, sensitivity and specificity of tools utilised for the detection of CNVs from WES data (Zhao et al., 2020). The multitude of tools available for the interrogation of CNVs from WES data allows the user to determine the best combinations of tools to meet the desired calling method and sensitivity and specificity based on their specific dataset (Louw et al., 2023). The use of different calling and mapping tools may therefore produce different CNVs of interest in our cohort and potentially provide additional diagnoses.

4.5. Implementation Challenges in the Establishment of NGS Testing

This scoping study provided insight into some of the challenges faced in LMICs regarding the implementation of NGS-based testing. Uptake of NGS technologies into standard diagnostic care in LMICs is slow, particularly in Sub-Saharan Africa, and may exacerbate the existing healthcare disparities experienced in these countries (Baine-Savanhu et al., 2023). Careful consideration of the unique challenges faced in each setting is necessary for these technologies to be successfully incorporated into healthcare structures, be accessible to all, and for health equity to be achieved (Kingsmore and Cole, 2022).

Many LMICs, including South Africa, are challenged by a lack of resources and infrastructure, shortages in genetics expertise amongst healthcare workers, poor

genomic literacy, lack of investment by funders and governments, limited data regarding country-specific genetic disease epidemiology, and fears of inaccessibility and biased benefit (Jongeneel et al., 2022). Our implementation strategy considered the local South African healthcare context and the population it served and was adapted from recommended international guidelines to accommodate some of these challenges. This study was executed in an academic hospital setting through referrals, therefore infants in rural settings with no access to specialist assessments by neonatologists, paediatricians and medical geneticists were excluded.

In this study we relied on primary care clinicians, who did not have extensive formal genetics training, for the referral of participants. This was done to address the shortages of medical geneticists and genetic counsellors in the South African State healthcare system. A singleton-sequencing strategy was implemented to address the social nuances which affect family structure and often result in the absence of one or both parents and their lack of availability to attend hospitals visits. The study design opted for the use of the available S5 sequencing platform, limiting investment in the procurement of additional technologies or the logistical costs to have NGS testing performed elsewhere. The use of virtual gene panels minimised the computational infrastructure and bioinformatics resources needed for data analysis and interpretation, addressing the shortage of available genomics infrastructure and bioinformatics capacity. Despite these adaptations several challenges to widespread implementation were observed and will be discussed in the following sections.

4.5.1. Virtual Gene Panels vs WES and WGS

We utilised a virtual gene panel analysis approach in this study, a less comprehensive analysis technique compared to WES and WGS. Despite their higher probabilities of pathogenic and causal variant identification, more comprehensive analysis strategies are more time and labour intensive, an important consideration in a low-resource setting and in NICU environments, where faster TATs for actionable results are necessary (Stark and Ellard, 2022). The analysis of a restricted subset of genes, as performed in our virtual gene panels, simplified the analysis workflow and reduced the dependence on computational infrastructure and bioinformatics capacity, both of which are limited in the South African healthcare context. Our gene panel strategies provided a successful first analysis approach for the participants in this cohort, achieving a notable diagnostic yield of more than 20%. More comprehensive analysis strategies,

such as an expanded panel and whole-exome analysis, could next be employed in participants who did not receive a positive molecular finding (Splinter et al., 2018).

The use of virtual panels negates additional laboratory costs when expanding analysis, removing the need for resequencing to incorporate more genes and regions of interest. This cost-saving approach provides additional benefit in low-resource settings. Our virtual gene panels were broad and encompassed genes associated with a large variety of phenotypes. This design strategy considered the often atypical and non-specific disease presentations of ill infants, which may skew the selection of genes in narrower panels or the selection of an inappropriate phenotype-specific panel (Alazami et al., 2015, Saudi Mendeliome, 2015).

Periodic and ad-hoc reanalysis may benefit the undiagnosed participants in this study, which constituted 78% of the total infant cohort. Reanalysis at a later time point can accommodate updates and expansions of the utilised gene panels to include new genes of interest, new variant information, including new pathogenicity reports, new *in silico* predictions, updated population frequencies, new phenotypic data from participants, and improvements and updates to bioinformatic tools (Ewans et al., 2018, Schobers et al., 2022). Reanalysis has been shown to increase the diagnostic yield of NGS in numerous studies, with an additional 10-32% of participants receiving a diagnosis after reanalysis (Wenger et al., 2017, Ewans et al., 2018, Wright et al., 2018b, Schmitz-Abe et al., 2019, Schobers et al., 2022).

Reanalysis may be particularly beneficial in our cohort, as phenotypes may become more defined and distinct in our infants as they age, allowing for better phenotyping and more accurate phenotype-genotype correlations. Reanalysis may also allow for variants of questionable quality, filtered out during the initial analysis due to cost and time restraints, to be reassessed and more closely scrutinised. This again may increase the diagnostic yield. Expansion and update of our curated panel to include additional genes with newly discovered neonatal and African relevance should be considered to maximise yields from reanalysis and incorporate the most up-to-date gene annotations. Continual updates to the DDG2P panel, as well as updates to ClinVar with the submission of new variant interpretations should also be considered for reanalysis.

Current recommendations suggest reanalysis a minimum of 18-24 months after the initial sequencing and analysis to maximise the probability of new relevant clinical and

variant data, however, it has been suggested that reanalysis be pursued sooner if new phenotypic features develop (Costain et al., 2018, Tan et al., 2020). Considering the rapid progression of infant disease presentations, more frequent reanalysis would therefore be recommended in ill infant populations. The resources required for this is a concern in the South African context in which resource scarcity makes frequent reanalysis of negative cases challenging and even impractical.

4.5.2. Singleton vs Trio Sequencing

We employed a singleton sequencing approach in our cohort despite the noted benefits and increased likelihood of diagnosis observed in previous studies with trio-sequencing (Clark et al., 2018, Wright et al., 2023). Trio-sequencing allows for the detection of *de novo* variants, phasing of compound heterozygous variants and an overall faster analysis time (Kingsmore et al., 2019, Wright et al., 2023). These benefits need to be weighed against the additional costs of two exomes per participant, especially in a low-resource setting, such as the South African State healthcare system, which is required to serve a large population size with limited financial resources.

A key consideration for the use of trio-sequencing in our study, in addition to the financial implications of performing thrice the number of exomes, was the availability of both parents at enrolment and sample collection. Only 31% of our cohort had both parents consented at enrolment due to a multitude of social and other factors. Some infants were admitted at hospitals far from their residence, some across provincial borders, and parents had no access to transportation for hospital visits and medical consultations. Other parents were unable to get time off work for hospital visits and could therefore not attend the enrolment session. Some participants in this study were immigrants who were not in South Africa as complete family units, making it impossible to have both parents available for recruitment. Many of our participants were born into single-parent families and would therefore be ineligible for inclusion in this study if the availability of both parents was a hard inclusion criterion.

Several diagnoses were achieved in our cohort with singleton sequencing, confirming the utility of this sequencing approach, as observed in other infant cohort studies, despite potentially more complex data analysis and longer TATs (Stark et al., 2016, Meng et al., 2017, Farnaes et al., 2018, Powis et al., 2018, French et al., 2019, Kingsmore et al., 2019, Dimmock et al., 2020, Smith et al., 2020). The implementation of NGS testing in South Africa, and in Africa, needs to consider and embrace the social

challenges and social dynamics of the local population, to prevent the exclusion of infants who may benefit from access to these genetic services.

Two significant disadvantages to the singleton sequencing approach should be noted. The first is the limited accuracy of recurrence risk calculations in the absence of genetic data from both parents as the inheritance of pathogenic variants is needed for these calculations. The second is the influence of inheritance information on variant classification and the number of VUSs. The inheritance of a variant is considered during variant classification, with *de novo* and assumed *de novo* variants providing moderate and strong evidence for pathogenicity (Richards et al., 2015). Segregation analysis within families may provide evidence of co-segregation and allude to novel variants and novel disease associations for variants which are completely penetrant. Inheritance information may therefore upgrade a variant from a VUS to pathogenic or likely pathogenic. The absence of inheritance data may be a considerable factor in the diagnostic yield of NGS tests in populations of African ancestry, in which high yields of VUSs are a common occurrence (Chen et al., 2023a).

4.5.3. Referrals from Neonatologists and Paediatricians

A significant proportion of our cohort (72%) were referred to the study for clinical exome sequencing directly by a neonatologist or paediatrician, without a formal consultation and assessment by a medical geneticist. A predictive genetic diagnosis, accompanied by appropriate phenotyping, can direct data analysis and interpretation, increasing the diagnostic yield and reducing the time-to-diagnosis (Trujillano et al., 2017, Gubbels et al., 2020). Frequent consultation from trained geneticists remains a considerable challenge in the South African State healthcare system due to the scarcity of medical geneticists and clinicians with formal genetics training, who serve a population of approximately 60 million individuals (Kromberg et al., 2013, Stats SA, 2022). The reliance on primary care clinicians, in this study, neonatologists and paediatricians in the NICU, is therefore necessary for appropriate patients to be identified and referred for NGS testing. The upskilling of these primary care clinicians is an important consideration for successful implementation of NGS testing into genetic services in low-resource settings. This allows for the limited formally trained genetics workforce to be efficiently and most appropriately utilised and for NGS testing to have a broader reach through other clinical disciplines (Dragojlovic et al., 2020, Chou et al., 2021).

Improved training of the NICU and PICU teams at our recruiting hospitals may help these clinicians to identify more appropriate infants who would most benefit from access to panel and exome sequencing, which may further increase the diagnostic yield of these tools in our setting and limit unnecessary testing and costs. The feedback of positive and negative results to referring clinicians may highlight key phenotypes and features to be aware of when referring patients for testing and indicate the depth of phenotyping necessary to facilitate variant interpretation. More formalised genetics training is being explored in Africa through initiatives such as the African Genomic Medicine Training (AGMT) Initiative and efforts by the H3Africa consortium (Mulder et al., 2018, Nembaware et al., 2019). Access to these formalised training initiatives may provide clinicians with tools and resources to confidently identify appropriate patients for referral and adequately assess their clinical features.

The challenge of inadequate genetics-trained personnel in establishing and implementing genetic services includes the shortage of both medical geneticists and genetic counsellors. Genetic counsellors play key roles in the feedback of results to participants and in assisting families to navigate the medical, psychological and familial implications of a genetic diagnosis (Resta et al., 2006). The absence of these services may influence the personal utility of the diagnosis for participants and families.

4.5.4. African Genetic Data Interpretation

The lack of representation of individuals of African ancestry in public databases complicates variant interpretation and increases the risk of variant misclassification, a challenge in the implementation of NGS testing in an African setting. Access to genetic testing across Africa is extremely limited and achieved in rare instances through research institutions (Kamp et al., 2021, Baine-Savanhu et al., 2023). As a result, African populations are significantly underrepresented in public databases, such as GnomAD and ClinVar, which makes the interpretation of rare variants and genomic data from individuals of African ancestry complex (Mulder et al., 2021). Individuals of African ancestry were half as likely to receive a diagnosis from NGS testing, compared to individuals of European ancestry in the DDD study (Wright et al., 2023). Data analysis in individuals of African ancestry is also more time-consuming as less variants are removed when filtering by minor population allele frequencies, due to the absence of documented common variation in African populations, who are more genetically diverse (Choudhury et al., 2020, Bowling et al., 2022).

These data shortages extend to include a lack of epidemiological data on genetic conditions affecting populations in Africa, stemming from the absence of African disease registries (Kamp et al., 2021). It is necessary to understand the type of conditions affecting African populations and their aetiologies to design and implement genetic tests that are appropriate for the African setting and to improve genetic services and patient management for patients of African ancestry. Through this study we contributed valuable information on the molecular profile of genetic conditions present in our South African population and contributed to the growing knowledge of the genetic disease epidemiology.

The inclusion of genes with known African-specific founder variants and variants of higher frequency in African populations were a notable element of our NICU/African panel design, allowing for the detection of variants specific to our unique population. An example of this is the inclusion of *STAC3* in the NICU/African panel, which was absent from both the BabySeq and NC NEXUS study lists and from DDG2P at the time of retrieval (freeze from April 2023). The variant observed in Participant NE001, associated with *STAC3* myopathy, has been well reported in individuals of African ancestry and has the highest prevalence in GnomAD in individuals of African and African American ancestry (Zaharieva et al., 2018, Gromand et al., 2022). The detection of a pathogenic variant in this gene, which is reported to have a prevalence of 20% in *SMN1* negative myopathy patients of African ancestry, supports its inclusion in routine diagnostic service in South Africa (Mhlongu et al., 2022). The introduction of *STAC3* single-gene testing into routine diagnostic services in South Africa during the course of this study highlights the importance of developing and implementing locally appropriate tests, which consider the variants and conditions most frequently observed in our South African population.

4.6. Study Limitations, Recommendations and Future Directions

A significant proportion of participants in this study cohort (78%) did not receive a molecular diagnosis through the use of the three filtering strategies employed. This is concordant with what has been reported in ill infant cohorts in the literature. The analysis workflow provided a starting analysis strategy, from which further genetic investigations may proceed. More comprehensive analyses, including the analysis of complete exome or an expanded panel, may address gaps in our analysis and provide additional diagnoses in this cohort.

Both the ClinVar and DDG2P panels are frequently updated to include new information from the literature and global clinical settings. A significant number of submissions to both of these publicly available datasets have been made throughout the course of this study. Reanalysis of our study cohort at a later time point with these updates may reveal new potential variants of interest, provide new insights into VUSs and therefore potentially increase our diagnostic yield. The DDG2P analysis strategy was observed to produce the highest yield amongst our African infant cohort. The DDG2P filter, supplemented with genes enriched for individuals of African ancestry, is therefore recommended as the starting analysis approach for future panel-based implementation strategies. The use of the updated ClinVar repository may assist in shortlisting variants of interest during variant prioritization using this strategy.

Comprehensive CNV analysis using optimised tools for our dataset may also reveal additional variants of interest, potentially providing additional diagnoses in our cohort. Reassessment and re-examination of our infant cohort may also provide further phenotypic information, as clinical features are likely to have progressed or become more distinct with age. These additional phenotypic terms may assist in variant interpretation and prioritization.

Broad inclusion criteria were used in this study to identify patients with a suspicion of a genetic condition. Considering the lack of formal genetics training by the primary referring clinicians, a limitation in this study, further refinement of the inclusion criteria may assist in the identification of appropriate patients for WES and gene panel analysis and ensure the limited resources for testing are best utilised. Upskilling of the recruiting clinicians for comprehensive phenotyping may assist in patient screening for inclusion and may assist in variant interpretation in the future. The utilisation of medical geneticists and genetic counsellors in participant screening after referral would provide significant benefit when available.

TATs and clinical utility for medical decision making were not formally assessed in this study, a limitation when inferring overall utility of NGS implementation in the South African context. Utility of a diagnosis was inferred for the seven participants who received a positive diagnosis; however, formal utility evaluation is needed for both positive and negative molecular findings. Follow ups with the primary referring clinicians to determine clinical utility with regards to change in clinical management, as well as with families to determine the extent of the personal and clinical utility of a result is needed and should be pursued in the future.

This pilot study provided key insights into the implementation of WES virtual panels for ill infants in South Africa through its utilisation in 32 ill infants. Upscaling of this project to recruit more infants will provide further insights into implementation challenges, benefits and potential harms, and allow for formal assessments of utility and economic evaluations to be performed.

4.7. Concluding Remarks

This explorative study investigated the implementation and utility of NGS testing in ill infants under intensive care in the South African State healthcare setting, providing some of the first insights into NGS-based diagnostics in the NICU in Africa. This study illustrated the utility of clinical exome gene panels through the diagnosis of seven families, who without access to this testing would likely have gone undiagnosed, producing a diagnostic yield of 22%. These diagnoses provided utility in initiating palliative care, familial screening, and prenatal testing for future pregnancies. The utilisation of singleton sequencing, primary care clinicians for referrals and virtual curated gene panels for data analysis considered the implementation challenges faced in translating these technologies into service in LMICs, where resources, infrastructure, and capacity are low. Despite these challenges, continued efforts are needed to establish genetic services, to prevent the exacerbation of the existing healthcare gaps in Africa. This study provides a point of departure for the development of appropriate exome sequencing strategies to benefit ill infants in the South African State healthcare system and other low-resource settings and to address the access disparities to genomic medicine in LMICs.

REFERENCES

Adams, D. R. & Eng, C. M. (2018). Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. *N Engl J Med,* 379(14)**,** 1353-1362.

Adebamowo, S. N., Francis, V., Tambo, E., et al. (2018). Implementation of genomics research in Africa: challenges and recommendations. *Glob Health Action*, 11(1), 1419033. doi:10.1080/16549716.2017.1419033.

Adeyemo, A. A., Amodu, O. K., Ekure, E. E., et al. (2018). Medical genetics and genomic medicine in Nigeria. *Mol Genet Genomic Med,* 6(3)**,** 314-321.

Adhikari, A. N., Gallagher, R. C., Wang, Y., et al. (2020). The role of exome sequencing in newborn screening for inborn errors of metabolism. *Nat Med,* 26(9)**,** 1392-1397.

Akoniyon, O. P., Adewumi, T. S., Maharaj, L., et al. (2022). Whole Genome Sequencing Contributions and Challenges in Disease Reduction Focused on Malaria. *Biology (Basel)*, 11(4), 587. doi:10.3390/biology11040587.

Alazami, A. M., Patel, N., Shamseldin, H. E., et al. (2015). Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. *Cell Rep,* 10(2)**,** 148-161.

Alfares, A., Aloraini, T., Subaie, L. A., et al. (2018). Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet Med,* 20(11)**,** 1328-1333.

Amburgey, K., Tsuchiya, E., De Chastonay, S., et al. (2017). A natural history study of X-linked myotubular myopathy. *Neurology,* 89(13)**,** 1355-1364.

Ashburner, M., Ball, C. A., Blake, J. A., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25(1)**,** 25-29.

Auer-Grumbach, M., Olschewski, A., Papic, L., et al. (2010). Alterations in the ankyrin domain of TRPV4 cause congenital distal SMA, scapuloperoneal SMA and HMSN2C. *Nat Genet,* 42(2)**,** 160-164.

Baine-Savanhu, F., Macaulay, S., Louw, N., et al. (2023). Identifying the genetic causes of developmental disorders and intellectual disability in Africa: a systematic literature review. *Front Genet*, 14(1), 1137922. doi:10.3389/fgene.2023.1137922.

Bamshad, M. J., Nickerson, D. A. & Chong, J. X. (2019). Mendelian Gene Discovery: Fast and Furious with No End in Sight. *Am J Hum Genet,* 105(3)**,** 448-455.

Barat-Houari, M., Dumont, B., Fabre, A., et al. (2016a). The expanding spectrum of COL2A1 gene variants IN 136 patients with a skeletal dysplasia phenotype. *Eur J Hum Genet,* 24(7)**,** 992-1000.

Barat-Houari, M., Sarrabay, G., Gatinois, V., et al. (2016b). Mutation Update for COL2A1 Gene Variants Associated with Type II Collagenopathies. *Hum Mutat,* 37(1)**,** 7-15.

Barbitoff, Y. A., Polev, D. E., Glotov, A. S., et al. (2020). Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Sci Rep*, 10(1), 2057. doi:10.1038/s41598-020-59026-y.

Baynam, G. S., Groft, S., Van Der Westhuizen, F. H., et al. (2020). A call for global action for rare diseases in Africa. *Nat Genet,* 52(1)**,** 21-26.

Bean, L. J. H., Funke, B., Carlston, C. M., et al. (2020). Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med,* 22(3)**,** 453-461.

Beck, K., Chan, V. C., Shenoy, N., et al. (2000). Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc Natl Acad Sci U S A,* 97(8)**,** 4273-4278.

Belhassan, K., Ouldim, K. & Sefiani, A. A. (2016). Genetics and genomic medicine in Morocco: the present hope can make the future bright. *Mol Genet Genomic Med,* 4(6)**,** 588-598.

Belkadi, A., Bolze, A., Itan, Y., et al. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A,* 112(17)**,** 5473-5478.

Beltcheva, O., Martin, P., Lenkkeri, U., et al. (2001). Mutation spectrum in the nephrin gene (NPHS1) in congenital nephrotic syndrome. *Hum Mutat,* 17(5)**,** 368-373.

Berciano, J., Baets, J., Gallardo, E., et al. (2011). Reduced penetrance in hereditary motor neuropathy caused by TRPV4 Arg269Cys mutation. *J Neurol,* 258(8)**,** 1413-1421.

Biancalana, V., Caron, O., Gallati, S., et al. (2003). Characterisation of mutations in 77 patients with X-linked myotubular myopathy, including a family with a very mild phenotype. *Hum Genet,* 112(2)**,** 135-142.

Biasini, F., Portaro, S., Mazzeo, A., et al. (2016). TRPV4 related scapuloperoneal spinal muscular atrophy: Report of an Italian family and review of the literature. *Neuromuscul Disord,* 26(4-5)**,** 312-315.

Black, G. F. & Sykes, P. (2022). Steps Toward Engagement Integrity: Learning From Participatory Visual Methods in Marginalized South African Communities. *Front Public Health*, 10(1), 794905. doi:10.3389/fpubh.2022.794905.

Bodian, D. L., Madhan, B., Brodsky, B., et al. (2008). Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry,* 47(19)**,** 5424-5432.

Bohring, A. (2002). OEIS complex, VATER, and the ongoing difficulties in terminology and delineation. *Am J Med Genet,* 107(1)**,** 72-76.

Bokenkamp, A. & Ludwig, M. (2016). The oculocerebrorenal syndrome of Lowe: an update. *Pediatr Nephrol,* 31(12)**,** 2201-2212.

Bope, C. D., Chimusa, E. R., Nembaware, V., et al. (2019). Dissecting in silico Mutation Prediction of Variants in African Genomes: Challenges and Perspectives. *Front Genet*, 10(1), 601. doi:10.3389/fgene.2019.00601.

Bowling, K. M., Thompson, M. L., Finnila, C. R., et al. (2022). Genome sequencing as a first-line diagnostic test for hospitalized infants. *Genet Med,* 24(4)**,** 851-861.

Boycott, K., Hartley, T., Adam, S., et al. (2015). The clinical application of genome-wide sequencing for monogenic diseases in Canada: Position Statement of the Canadian College of Medical Geneticists. *J Med Genet,* 52(7)**,** 431-437.

Brunelli, L., Jenkins, S. M., Gudgeon, J. M., et al. (2019). Targeted gene panel sequencing for the rapid diagnosis of acutely ill infants. *Mol Genet Genomic Med*, 7(7), e00796. doi:10.1002/mgg3.796.

Butcher, D. T., Cytrynbaum, C., Turinsky, A. L., et al. (2017). CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. *Am J Hum Genet,* 100(5)**,** 773-788.

Cakici, J. A., Dimmock, D. P., Caylor, S. A., et al. (2020). A Prospective Study of Parental Perceptions of Rapid Whole-Genome and -Exome Sequencing among Seriously Ill Infants. *Am J Hum Genet,* 107(5)**,** 953-962.

Callahan, K. P., Mueller, R., Flibotte, J., et al. (2022). Measures of Utility Among Studies of Genomic Medicine for Critically Ill Infants: A Systematic Review. *JAMA Netw Open*, 5(8), e2225980. doi:10.1001/jamanetworkopen.2022.25980.

Campbell, L., Fredericks, J., Mathivha, K., et al. (2023). The implementation and utility of clinical exome sequencing in a South African infant cohort. *Front Genet*, 14(1), 1277948. doi:10.3389/fgene.2023.1277948.

Carecchio, M., Invernizzi, F., Gonzalez-Latapi, P., et al. (2019). Frequency and phenotypic spectrum of KMT2B dystonia in childhood: A single-center cohort study. *Mov Disord,* 34(10)**,** 1516-1527.

Caruana, G., Wong, M. N., Walker, A., et al. (2015). Copy-number variation associated with congenital anomalies of the kidney and urinary tract. *Pediatr Nephrol,* 30(3)**,** 487-95.

Cesani, M., Lorioli, L., Grossi, S., et al. (2016). Mutation Update of ARSA and PSAP Genes Causing Metachromatic Leukodystrophy. *Hum Mutat,* 37(1)**,** 16-27.

Ceyhan-Birsoy, O., Machini, K., Lebo, M. S., et al. (2017). A curated gene list for reporting results of newborn genomic sequencing. *Genet Med,* 19(7)**,** 809-818.

Ceyhan-Birsoy, O., Murry, J. B., Machini, K., et al. (2019). Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project. *Am J Hum Genet,* 104(1)**,** 76-93.

Chace, D. H., Kalas, T. A. & Naylor, E. W. (2002). The application of tandem mass spectrometry to neonatal screening for inherited disorders of intermediary metabolism. *Annu Rev Genomics Hum Genet,* 3(2002)**,** 17-45.

Chan, D. & Cole, W. G. (1991). Low basal transcription of genes for tissue-specific collagens by fibroblasts and lymphoblastoid cells. Application to the characterization of a glycine 997 to serine substitution in alpha 1(II) collagen chains of a patient with spondyloepiphyseal dysplasia. *J Biol Chem,* 266(19)**,** 12487-12494.

Chance, P. F. (2006). Inherited focal, episodic neuropathies: hereditary neuropathy with liability to pressure palsies and hereditary neuralgic amyotrophy. *Neuromolecular Med,* 8(1-2)**,** 159-174.

Chance, P. F., Alderson, M. K., Leppig, K. A., et al. (1993). DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell,* 72(1)**,** 143-151.

Chen, E., Facio, F. M., Aradhya, K. W., et al. (2023a). Rates and Classification of Variants of Uncertain Significance in Hereditary Disease Genetic Testing. *JAMA Netw Open*, 6(10), e2339571. doi:10.1001/jamanetworkopen.2023.39571.

Chen, S., Francioli, L. C., Goodrich, J. K., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. *Nature,* 625(7993)**,** 92-100.

Chen, T., Fan, C., Huang, Y., et al. (2023b). Genomic Sequencing as a First-Tier Screening Test and Outcomes of Newborn Screening. *JAMA Netw Open*, 6(9), e2331162. doi:10.1001/jamanetworkopen.2023.31162.

Cho, T. J., Matsumoto, K., Fano, V., et al. (2012). TRPV4-pathy manifesting both skeletal dysplasia and peripheral neuropathy: a report of three patients. *Am J Med Genet A,* 158A(4)**,** 795-802.

Choi, J. H., Oh, M. Y., Yum, M. S., et al. (2015). Moyamoya syndrome in a patient with Noonan-like syndrome with loose anagen hair. *Pediatr Neurol,* 52(3)**,** 352-355.

Chong, J. X., Buckingham, K. J., Jhangiani, S. N., et al. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet,* 97(2)**,** 199-215.

Chou, A. F., Duncan, A. R., Hallford, G., et al. (2021). Barriers and strategies to integrate medical genetics and primary care in underserved populations: a scoping review. *J Community Genet,* 12(3)**,** 291-309.

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

Cif, L., Demailly, D., Lin, J. P., et al. (2020). KMT2B-related disorders: expansion of the phenotypic spectrum and long-term efficacy of deep brain stimulation. *Brain,* 143(11)**,** 3242-3261.

Clark, M. M., Hildreth, A., Batalov, S., et al. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci Transl Med*, 11(489), eaat6177. doi:10.1126/scitranslmed.aat6177.

Clark, M. M., Stark, Z., Farnaes, L., et al. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med*, 3(2018), 16. doi:10.1038/s41525-018-0053-8.

Collins, T., Mikkelsen, B., Adams, J., et al. (2018). Addressing NCDs: A unifying agenda for sustainable development. *Glob Public Health,* 13(9)**,** 1152-1157.

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

Costain, G., Jobling, R., Walker, S., et al. (2018). Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet,* 26(5)**,** 740-744.

D'gama, A. M., Del Rosario, M. C., Bresnahan, M. A., et al. (2022). Integrating rapid exome sequencing into NICU clinical care after a pilot research study. *NPJ Genomic Medicine*, 7(2022), 51. doi:10.1038/s41525-022-00326-9.

Daoud, H., Luco, S. M., Li, R., et al. (2016). Next-generation sequencing for diagnosis of rare diseases in the neonatal intensive care unit. *CMAJ,* 188(11)**,** e254-E260.

De Castro, M. J., González-Vioque, E., Barbosa-Gouveia, S., et al. (2020). Rapid phenotype-driven gene sequencing with the NeoSeq panel: a diagnostic tool for critically ill newborns with suspected genetic disease. *J Clin Med*, 9(8), 2362. doi:10.3390/jcm9082362.

De Leeneer, K., Hellemans, J., Steyaert, W., et al. (2015). Flexible, scalable, and efficient targeted resequencing on a benchtop sequencer for variant detection in clinical practice. *Hum Mutat,* 36(3)**,** 379-387.

De Vries, J., Jallow, M., Williams, T. N., et al. (2012). Investigating the potential for ethnic group harm in collaborative genomics research in Africa: is ethnic stigmatisation likely? *Soc Sci Med,* 75(8)**,** 1400-1407.

Delaneau, O., Howie, B., Cox, A. J., et al. (2013). Haplotype estimation using sequencing reads. *Am J Hum Genet,* 93(4)**,** 687-696.

Deng, H. X., Klein, C. J., Yan, J., et al. (2010). Scapuloperoneal spinal muscular atrophy and CMT2C are allelic disorders caused by alterations in TRPV4. *Nat Genet,* 42(2)**,** 165-169.

Denommé-Pichon, A.-S., Vitobello, A., Olaso, R., et al. (2022). Accelerated genome sequencing with controlled costs for infants in intensive care units: a feasibility study in a French hospital network. *Eur J Hum Genet,* 30(5)**,** 567-576.

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

Dimmock, D., Caylor, S., Waldman, B., et al. (2021). Project Baby Bear: Rapid precision care incorporating rWGS in 5 California children's hospitals demonstrates improved clinical outcomes and reduced costs of care. *Am J Hum Genet,* 108(7)**,** 1231-1238.

Dimmock, D. P., Clark, M. M., Gaughran, M., et al. (2020). An RCT of rapid genomic sequencing among seriously ill infants results in high clinical utility, changes in management, and low perceived harm. *Am J Hum Genet,* 107(5)**,** 942-952.

Dragojlovic, N., Borle, K., Kopac, N., et al. (2020). The composition and capacity of the clinical genetics workforce in high-income countries: a scoping review. *Genet Med,* 22(9)**,** 1437-1449.

Duncan, L., Shen, H., Gelaye, B., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*, 10(2019), 3328. doi:10.1038/s41467-019-11112-0.

Echaniz-Laguna, A., Dubourg, O., Carlier, P., et al. (2014). Phenotypic spectrum and incidence of TRPV4 mutations in patients with inherited axonal neuropathy. *Neurology,* 82(21)**,** 1919-1926.

Ekvall, S., Hagenas, L., Allanson, J., et al. (2011). Co-occurring SHOC2 and PTPN11 mutations in a patient with severe/complex Noonan syndrome-like phenotype. *Am J Med Genet A,* 155A(6)**,** 1217-1224.

El-Attar, E. A., Helmy Elkaffas, R. M., Aglan, S. A., et al. (2022). Genomics in Egypt: Current Status and Future Aspects. *Front Genet,* 13797465.

El-Hattab, A. W., Skorupski, J. C., Hsieh, M. H., et al. (2010). OEIS complex associated with chromosome 1p36 deletion: a case report and review. *Am J Med Genet A,* 152A(2)**,** 504-511.

Elliott, A. M., Du Souich, C., Lehman, A., et al. (2019). RAPIDOMICS: rapid genome-wide sequencing in a neonatal intensive care unit—successes and challenges. *Eur J Pediatr,* 178(8)**,** 1207-1218.

Elloumi-Zghal, H. & Chaabouni Bouhamed, H. (2018). Genetics and genomic medicine in Tunisia. *Mol Genet Genomic Med,* 6(2)**,** 134-159.

Evangelista, T., Bansagi, B., Pyle, A., et al. (2015). Phenotypic variability of TRPV4 related neuropathies. *Neuromuscul Disord,* 25(6)**,** 516-521.

Ewans, L. J., Schofield, D., Shrestha, R., et al. (2018). Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet Med,* 20(12)**,** 1564-1574.

Farnaes, L., Hildreth, A., Sweeney, N. M., et al. (2018). Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *NPJ Genom Med*, 3(2018), 10. doi:10.1038/s41525-018-0049-4.

Fecto, F., Shi, Y., Huda, R., et al. (2011). Mutant TRPV4-mediated toxicity is linked to increased constitutive function in axonal neuropathies. *J Biol Chem,* 286(19)**,** 17281-17291.

Feng, B. J. (2017). PERCH: A Unified Framework for Disease Gene Prioritization. *Hum Mutat,* 38(3)**,** 243-251.

Fleming, J. & Quan, D. (2016). A case of congenital spinal muscular atrophy with pain due to a mutation in TRPV4. *Neuromuscul Disord,* 26(12)**,** 841-843.

Freed, A. S., Candadai, S. V. C., Sikes, M. C., et al. (2020). The impact of rapid exome sequencing on medical management of critically ill children. *J Pediatr,* 226(2020)**,** 202-212.

French, C. E., Delon, I., Dolling, H., et al. (2019). Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med,* 45(5)**,** 627-636.

Gabriels, C. & Le Roux, D. (2023). In-hospital neonatal mortality in a level-2 hospital in Cape Town, South Africa. *SAJCH,* 17(3)**,** 129-134.

Gargano, G., Guidotti, I., Balestri, E., et al. (2014). Hydrops fetalis in a preterm newborn heterozygous for the c.4A>G SHOC2 mutation. *Am J Med Genet A,* 164A(4)**,** 1015-1020.

Gedleh, A., Lee, S., Hill, J. A., et al. (2018). "Where Does it Come from?" Experiences Among Survivors and Parents of Children with Retinoblastoma in Kenya. *J Genet Couns,* 27(3)**,** 574-588.

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

Goddard, C. (2004). The ethnopragmatics and semantics of 'active metaphors'. *Journal of Pragmatics,* 36(7)**,** 1211-1230.

Gorman, K. M., Meyer, E. & Kurian, M. A. (2018). Review of the phenotype of early-onset generalised progressive dystonia due to mutations in KMT2B. *Eur J Paediatr Neurol,* 22(2)**,** 245-256.

Goto, Y., Akahori, R., Yanagi, I., et al. (2020). Solid-state nanopores towards single-molecule DNA sequencing. *J Hum Genet,* 65(1)**,** 69-77.

Gripp, K. W., Zand, D. J., Demmer, L., et al. (2013). Expanding the SHOC2 mutation associated phenotype of Noonan syndrome with loose anagen hair: structural brain anomalies and myelofibrosis. *Am J Med Genet A,* 161A(10)**,** 2420-2430.

Gromand, M., Gueguen, P., Perville, A., et al. (2022). STAC3 related congenital myopathy: A case series of seven Comorian patients. *Eur J Med Genet*, 65(10), 104598. doi:10.1016/j.ejmg.2022.104598.

Gubbels, C. S., Vannoy, G. E., Madden, J. A., et al. (2020). Prospective, phenotype-driven selection of critically ill neonates for rapid exome sequencing is associated with high diagnostic yield. *Genet Med,* 22(4)**,** 736-744.

Gurgel-Giannetti, J., Zanoteli, E., De Castro Concentino, E. L., et al. (2012). Necklace fibers as histopathological marker in a patient with severe form of X-linked myotubular myopathy. *Neuromuscul Disord,* 22(6)**,** 541-545.

Gursoy, S. & Ercal, D. (2016). Diagnostic Approach to Genetic Causes of Early-Onset Epileptic Encephalopathy. *J Child Neurol,* 31(4)**,** 523-532.

Hagen, L., Khattar, D., Whitehead, K., et al. (2022). Detection and impact of genetic disease in a level IV neonatal intensive care unit. *J Perinatol,* 42(5)**,** 580-588.

Halabi, N., Ramaswamy, S., El Naofal, M., et al. (2022). Rapid whole genome sequencing of critically ill pediatric patients from genetically underrepresented populations. *Genome Med*, 14(1), 56. doi:10.1186/s13073-022-01061-7.

Halley, M. C., Young, J. L., Fernandez, L., et al. (2022). Perceived utility and disutility of genomic sequencing for pediatric patients: Perspectives from parents with diverse sociodemographic characteristics. *Am J Med Genet A,* 188(4)**,** 1088-1101.

Harris, P. A., Taylor, R., Minor, B. L., et al. (2019). The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*, 95(1), 103208. doi:10.1016/j.jbi.2019.103208.

Harris, P. A., Taylor, R., Thielke, R., et al. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform,* 42(2)**,** 377-381.

Hayeems, R. Z., Bhawra, J., Tsiplova, K., et al. (2017). Care and cost consequences of pediatric whole genome sequencing compared to chromosome microarray. *Eur J Hum Genet,* 25(12)**,** 1303-1312.

Heeringa, S. F., Vlangos, C. N., Chernin, G., et al. (2008). Thirteen novel NPHS1 mutations in a large cohort of children with congenital nephrotic syndrome. *Nephrol Dial Transplant,* 23(11)**,** 3527-3533.

Hichri, H., Rendu, J., Monnier, N., et al. (2011). From Lowe syndrome to Dent disease: correlations between mutations of the OCRL1 gene and clinical and biochemical phenotypes. *Hum Mutat,* 32(4)**,** 379-388.

Hoban, R., Roberts, A. E., Demmer, L., et al. (2012). Noonan syndrome due to a SHOC2 mutation presenting with fetal distress and fatal hypertrophic cardiomyopathy in a premature infant. *Am J Med Genet A,* 158A(6)**,** 1411-1413.

Hofman, K. & Madhi, S. (2020). The unanticipated costs of COVID-19 to South Africa's quadruple disease burden. *S Afr Med J,* 110(8)**,** 698-699.

Hoque, M., Haaq, S. & Islam, R. (2011). Causes of neonatal admissions and deaths at a rural hospital in KwaZulu-Natal, South Africa. *SAJEI,* 26(1)**,** 26-29.

Horstick, E. J., Linsley, J. W., Dowling, J. J., et al. (2013). Stac3 is a component of the excitation-contraction coupling machinery and mutated in Native American myopathy. *Nat Commun*, 4(2013), 1952. doi:10.1038/ncomms2952.

Imafidon, M. E., Sikkema-Raddatz, B., Abbott, K. M., et al. (2021). Strategies in rapid genetic diagnostics of critically Ill children: experiences from a Dutch university hospital. *Front Pediatr*, 9(2021), 600556. doi:10.3389/fped.2021.600556.

Incerti, D., Xu, X. M., Chou, J. W., et al. (2022). Cost-effectiveness of genome sequencing for diagnosing patients with undiagnosed rare genetic diseases. *Genet Med,* 24(1)**,** 109-118.

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

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., Mcrae, J. F., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell,* 176(3)**,** 535-548.

Jedrzejowska, M., Debek, E., Kowalczyk, B., et al. (2019). The remarkable phenotypic variability of the p.Arg269HiS variant in the TRPV4 gene. *Muscle Nerve,* 59(1)**,** 129-133.

Jongeneel, C. V., Kotze, M. J., Bhaw-Luximon, A., et al. (2022). A View on Genomic Medicine Activities in Africa: Implications for Policy. *Front Genet*, 13(2022), 769919. doi:10.3389/fgene.2022.769919.

Kamaara, E., Kong, C. & Campbell, M. (2020). Prioritising African perspectives in psychiatric genomics research: Issues of translation and informed consent. *Dev World Bioeth,* 20(3)**,** 139-149.

Kamp, M., Krause, A. & Ramsay, M. (2021). Has translational genomics come of age in Africa? *Hum Mol Genet,* 30(20)**,** 164-173.

Kawarai, T., Miyamoto, R., Nakagawa, E., et al. (2018). Phenotype variability and allelic heterogeneity in KMT2B-Associated disease. *Parkinsonism Relat Disord,* 52(2018)**,** 55-61.

Kellar-Guenther, Y., Mckasson, S., Hale, K., et al. (2020). Implementing Statewide Newborn Screening for New Disorders: U.S. Program Experiences. *Int J Neonatal Screen*, 6(2), 35. doi:10.3390/ijns6020035.

Kengne Kamga, K., Munung, N. S., Nguefack, S., et al. (2021). Negotiating political power and stigma around fragile X Syndrome in a rural village in Cameroon: A tale of a royal family and a community. *Mol Genet Genomic Med*, 9(3), e1615. doi:10.1002/mgg3.1615.

Kent, W. J., Sugnet, C. W., Furey, T. S., et al. (2002). The human genome browser at UCSC. *Genome Res,* 12(6)**,** 996-1006.

Kestila, M., Lenkkeri, U., Mannikko, M., et al. (1998). Positionally cloned gene for a novel glomerular protein--nephrin--is mutated in congenital nephrotic syndrome. *Mol Cell,* 1(4)**,** 575-82.

Kingsmore, S. F., Cakici, J. A., Clark, M. M., et al. (2019). A Randomized, Controlled Trial of the Analytic and Diagnostic Performance of Singleton and Trio, Rapid Genome and Exome Sequencing in Ill Infants. *Am J Hum Genet,* 105(4)**,** 719-733.

Kingsmore, S. F. & Cole, F. S. (2022). The Role of Genome Sequencing in Neonatal Intensive Care Units. *Annu Rev Genomics Hum Genet,* 23(2022)**,** 427-448.

Kingsmore, S. F., Henderson, A., Owen, M. J., et al. (2020). Measurement of genetic diseases as a cause of mortality in infants receiving whole genome sequencing. *NPJ Genom Med*, 5(1), 49. doi:10.1038/s41525-020-00155-8.

Kingsmore, S. F., Smith, L. D., Kunard, C. M., et al. (2022). A genome sequencing system for universal newborn screening, diagnosis, and precision medicine for severe genetic diseases. *Am J Hum Genet,* 109(9)**,** 1605-1619.

Klein, C. J., Shi, Y., Fecto, F., et al. (2011). TRPV4 mutations and cytotoxic hypercalcemia in axonal Charcot-Marie-Tooth neuropathies. *Neurology,* 76(10)**,** 887-894.

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

Komatsuzaki, S., Aoki, Y., Niihori, T., et al. (2010). Mutation analysis of the SHOC2 gene in Noonan-like syndrome and in hematologic malignancies. *J Hum Genet,* 55(12)**,** 801-809.

Koziell, A., Grech, V., Hussain, S., et al. (2002). Genotype/phenotype correlations of NPHS1 and NPHS2 mutations in nephrotic syndrome advocate a functional inter-relationship in glomerular filtration. *Hum Mol Genet,* 11(4)**,** 379-388.

Krakow, D., Vriens, J., Camacho, N., et al. (2009). Mutations in the gene encoding the calcium-permeable ion channel TRPV4 produce spondylometaphyseal dysplasia, Kozlowski type and metatropic dysplasia. *Am J Hum Genet,* 84(3)**,** 307-15.

Krantz, I. D., Medne, L., Weatherly, J. M., et al. (2021). Effect of whole-genome sequencing on the clinical management of acutely ill infants with suspected genetic disease: a randomized clinical trial. *JAMA Pediatr,* 175(12)**,** 1218-1226.

Krause, A., Seymour, H. & Ramsay, M. (2018). Common and Founder Mutations for Monogenic Traits in Sub-Saharan African Populations. *Annu Rev Genomics Hum Genet,* 19(1)**,** 149-175.

Kromberg, J. G., Sizer, E. B. & Christianson, A. L. (2013). Genetic services and testing in South Africa. *J Community Genet,* 4(3)**,** 413-423.

Kucera, K. S., Taylor, J. L., Robles, V. R., et al. (2021). A Voluntary Statewide Newborn Screening Pilot for Spinal Muscular Atrophy: Results from Early Check. *Int J Neonatal Screen*, 7(1), 20. doi:10.3390/ijns7010020.

Kwan, A., Abraham, R. S., Currier, R., et al. (2014). Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA,* 312(7)**,** 729-738.

Kwon, D. H., Zhang, F., Mccray, B. A., et al. (2023). TRPV4-Rho GTPase complex structures reveal mechanisms of gating and disease. *Nat Commun*, 14(1), 3732. doi:10.1038/s41467-023-39345-0.

Landoure, G., Zdebik, A. A., Martinez, T. L., et al. (2010). Mutations in TRPV4 cause Charcot-Marie-Tooth disease type 2C. *Nat Genet,* 42(2)**,** 170-174.

Landrum, M. J., Lee, J. M., Benson, M., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res,* 46(1)**,** 1062-1067.

Laporte, J., Guiraud-Chaumeil, C., Vincent, M. C., et al. (1997). Mutations in the MTM1 gene implicated in X-linked myotubular myopathy. ENMC International Consortium on Myotubular Myopathy. European Neuro-Muscular Center. *Hum Mol Genet,* 6(9)**,** 1505-11.

Lee, B. H., Ahn, Y. H., Choi, H. J., et al. (2009). Two Korean infants with genetically confirmed congenital nephrotic syndrome of Finnish type. *J Korean Med Sci,* 24(Suppl 1)**,** 210-214.

Lee, N. C. (2023). The incorporation of next-generation sequencing into pediatric care. *Pediatr Neonatol,* 64(Suppl 1)**,** 30-34.

Lek, M., Karczewski, K. J., Minikel, E. V., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536(7616)**,** 285-291.

Lelieveld, S. H., Spielmann, M., Mundlos, S., et al. (2015). Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum Mutat,* 36(8)**,** 815-22.

Lemke, A. A., Thompson, M. L., Gimpel, E. C., et al. (2023). Parents' Perspectives on the Utility of Genomic Sequencing in the Neonatal Intensive Care Unit. *J Pers Med*, 13(7), 1026. doi:10.3390/jpm13071026.

Li, C., Zhi, D., Wang, K., et al. (2022). MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. *Genome Med*, 14(1), 115. doi:10.1186/s13073-022-01120-z.

Lim, E. C., Brett, M., Lai, A. H., et al. (2015). Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients. *Hum Genomics*, 9(2015), 33. doi:10.1186/s40246-015-0055-x.

Linsley, J. W., Hsu, I. U., Groom, L., et al. (2017). Congenital myopathy results from misregulation of a muscle Ca2+ channel by mutant Stac3. *Proc Natl Acad Sci U S A,* 114(2)**,** 228-236.

Lionel, A. C., Costain, G., Monfared, N., 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. *Genet Med,* 20(4)**,** 435-443.

Loi, M. (2006). Lowe syndrome. *Orphanet J Rare Dis*, 1(1), 16. doi:10.1186/1750-1172-1-16.

Longo, G., Russo, S., Novelli, G., et al. (2016). Mutation spectrum of the MTM1 gene in XLMTM patients: 10 years of experience in prenatal and postnatal diagnosis. *Clin Genet,* 89(1)**,** 93-98.

Louis, E. D., Hernandez, N., Ottman, R., et al. (2014). Essential Tremor in a Charcot-Marie-Tooth Type 2C Kindred Does Not Segregate with the TRPV4 R269H Mutation. *Case Rep Neurol,* 6(1)**,** 1-6.

Louw, N., Carstens, N., Lombard, Z., et al. (2023). Incorporating CNV analysis improves the yield of exome sequencing for rare monogenic disorders-an important consideration for resource-constrained settings. *Front Genet*, 14(1), 1277784. doi:10.3389/fgene.2023.1277784.

Lumaka, A., Carstens, N., Devriendt, K., et al. (2022). Increasing African genomic data generation and sharing to resolve rare and undiagnosed diseases in Africa: a call-to-action by the H3Africa rare diseases working group. *Orphanet J Rare Dis*, 17(1), 230. doi:10.1186/s13023-022-02391-w.

Lumaka, A., Fasquelle, C., Debray, F. G., et al. (2023). Rapid Whole Genome Sequencing Diagnoses and Guides Treatment in Critically Ill Children in Belgium in Less than 40 Hours. *Int J Mol Sci*, 24(4), 4003. doi:10.3390/ijms24044003.

Lunke, S. (2024). The BabyScreen+ Study: a genomic newborn screening pilot. *Pathology*, 56(Suppl 1), 20. doi:10.1016/j.pathol.2023.12.080.

Lunke, S., Eggers, S., Wilson, M., et al. (2020). Feasibility of Ultra-Rapid Exome Sequencing in Critically Ill Infants and Children With Suspected Monogenic Conditions in the Australian Public Health Care System. *JAMA,* 323(24)**,** 2503-2511.

Lupski, J. R., De Oca-Luna, R. M., Slaugenhaupt, S., et al. (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell,* 66(2)**,** 219-232.

Macdonald, J. R., Ziman, R., Yuen, R. K., et al. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res,* 42(1)**,** 986-992.

Machuca, E., Benoit, G., Nevo, F., et al. (2010). Genotype-phenotype correlations in non-Finnish congenital nephrotic syndrome. *J Am Soc Nephrol,* 21(7)**,** 1209-1217.

Mahmoud, M., Huang, Y., Garimella, K., et al. (2024). Utility of long-read sequencing for All of Us. *Nat Commun*, 15(1), 837. doi:10.1038/s41467-024-44804-3.

Majewski, J., Schwartzentruber, J., Lalonde, E., et al. (2011). What can exome sequencing do for you? *J Med Genet,* 48(9)**,** 580-589.

Mak, D. Y., Sykes, J., Stephenson, A. L., et al. (2016). The benefits of newborn screening for cystic fibrosis: The Canadian experience. *J Cyst Fibros,* 15(3)**,** 302-308.

Malherbe, H. L., Christianson, A. L., Woods, D., et al. (2017). The case for the genetic nurse in South Africa. *J Community Genet,* 8(2)**,** 65-73.

Malki, L., Sarig, O., Romano, M. T., et al. (2019). Variant PADI3 in Central Centrifugal Cicatricial Alopecia. *N Engl J Med,* 380(9)**,** 833-841.

Manickam, K., Mcclain, M. R., Demmer, L. A., 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). *Genet Med,* 23(11)**,** 2029-2037.

Mann, N., Braun, D. A., Amann, K., et al. (2019). Whole-Exome Sequencing Enables a Precision Medicine Approach for Kidney Transplant Recipients. *J Am Soc Nephrol,* 30(2)**,** 201-215.

Maron, J. L., Kingsmore, S., Gelb, B. D., et al. (2023). Rapid Whole-Genomic Sequencing and a Targeted Neonatal Gene Panel in Infants With a Suspected Genetic Disorder. *JAMA,* 330(2)**,** 161-169.

Maron, J. L., Kingsmore, S. F., Wigby, K., et al. (2021). Novel variant findings and challenges associated with the clinical integration of genomic testing: an interim report of the genomic medicine for ill neonates and infants (GEMINI) study. 175(5), 205906. doi:10.1001/jamapediatrics.2020.5906.

Martin, A. R., Williams, E., Foulger, R. E., et al. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet,* 51(11)**,** 1560-1565.

Mbarek, I. B., Abroug, S., Omezzine, A., et al. (2011). Novel mutations in steroid-resistant nephrotic syndrome diagnosed in Tunisian children. *Pediatr Nephrol,* 26(2)**,** 241-249.

Mckusick, V. A. 1998. *Mendelian inheritance in man: a catalog of human genes and genetic disorders*, John Hopkins University Press.

Mclaren, W., Gil, L., Hunt, S. E., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biol*, 17(1), 122. doi:10.1186/s13059-016-0974-4.

Mehta, Z. B., Pietka, G. & Lowe, M. (2014). The cellular and physiological functions of the Lowe syndrome protein OCRL1. *Traffic,* 15(5)**,** 471-487.

Meienberg, J., Bruggmann, R., Oexle, K., et al. (2016). Clinical sequencing: is WGS the better WES? *Hum Genet,* 135(3)**,** 359-362.

Mejia-Guevara, I., Zuo, W., Bendavid, E., et al. (2019). Age distribution, trends, and forecasts of under-5 mortality in 31 sub-Saharan African countries: A modeling study. *PLoS Med*, 16(3), e1002757. doi:10.1371/journal.pmed.1002757.

Meng, L., Pammi, M., Saronwala, A., et al. (2017). Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. *JAMA Pediatr*, 171(12), e173438. doi:10.1001/jamapediatrics.2017.3438.

Mercer, C. L., Andreoletti, G., Carroll, A., et al. (2017). Familial Ebstein Anomaly: Whole Exome Sequencing Identifies Novel Phenotype Associated With FLNA. *Circ Cardiovasc Genet*, 10(6), e001683. doi:10.1161/CIRCGENETICS.116.001683.

Meyer, E., Carss, K. J., Rankin, J., et al. (2017). Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. *Nat Genet,* 49(2)**,** 223-237.

Mhlongu, F., Bhengu, L., Dillon, B., et al. (2022). Published. STAC3 disorder: Frequent genetic cause of congenital hypotonia in African patients. University of the Witwatersrand Faculty of Health Sciences Research Day, 24 September 2022, Johannesburg, South Africa.

Milko, L. V., O'daniel, J. M., Decristo, D. M., et al. (2019). An Age-Based Framework for Evaluating Genome-Scale Sequencing Results in Newborn Screening. *J Pediatr,* 209(1)**,** 68-76.

Miller, D. T., Adam, M. P., Aradhya, S., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet,* 86(5)**,** 749-764.

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

Mistry, J., Chuguransky, S., Williams, L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res,* 49(1)**,** 412-419.

Molina-Ramirez, L. P., Kyle, C., Ellingford, J. M., et al. (2022). Personalised virtual gene panels reduce interpretation workload and maintain diagnostic rates of proband-only clinical exome sequencing for rare disorders. *J Med Genet,* 59(4)**,** 393-398.

Moundzika-Kibamba, J.-C. & Nakwa, F. (2018). Neonatal mortality at Leratong hospital. *SAJCH,* 12(1)**,** 24-28.

Mubungu, G., Makay, P., Boujemla, B., et al. (2021). Clinical presentation and evolution of Xia-Gibbs syndrome due to p.Gly375ArgfsTer3 variant in a patient from DR Congo (Central Africa). *Am J Med Genet A,* 185(3)**,** 990-994.

Mulder, N., Abimiku, A., Adebamowo, S. N., et al. (2018). H3Africa: current perspectives. *Pharmgenomics Pers Med,* 2018(1)**,** 59-66.

Mulder, N., Zass, L., Hamdi, Y., et al. (2021). African Global Representation in Biomedical Sciences. *Annu Rev Biomed Data Sci,* 4(1)**,** 57-81.

Munung, N. S., De Vries, J. & Pratt, B. (2022). Towards equitable genomics governance in Africa: Guiding principles from theories of global health governance and the African moral theory of Ubuntu. *Bioethics,* 36(4)**,** 411-422.

Nachat, R., Mechin, M. C., Takahara, H., et al. (2005). Peptidylarginine deiminase isoforms 1-3 are expressed in the epidermis and involved in the deimination of K1 and filaggrin. *J Invest Dermatol,* 124(2)**,** 384-393.

Naidoo, L. & Reddy, P. (2022). Stakeholder knowledge and attitudes toward the use of predictive genetic testing in South Africa. *J Community Genet,* 13(6)**,** 567-577.

Ndiaye Diallo, R., Gadji, M., Hennig, B. J., et al. (2017). Strengthening human genetics research in Africa: report of the 9th meeting of the African Society of Human Genetics in Dakar in May 2016. *Glob Health Epidemiol Genom*, 2(1), e10. doi:10.1017/gheg.2017.3.

Nembaware, V., African Genomic Medicine Training, I. & Mulder, N. (2019). The African Genomic Medicine Training Initiative (AGMT): Showcasing a Community and Framework Driven Genomic Medicine Training for Nurses in Africa. *Front Genet*, 10(1), 1209. doi:10.3389/fgene.2019.01209.

Ng, S. B., Buckingham, K. J., Lee, C., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet,* 42(1)**,** 30-35.

Niehaus, A., Azzariti, D. R., Harrison, S. M., et al. (2019). A survey assessing adoption of the ACMG-AMP guidelines for interpreting sequence variants and identification of areas for continued improvement. *Genet Med,* 21(8)**,** 1699-1701.

Nieuwoudt, L., Mackay, C. A. & Mda, S. (2022). Causes of and Modifiable Factors Contributing to Neonatal Deaths at Dora Nginza Hospital in the Eastern Cape, South Africa. *Glob Pediatr Health*, 9(1), 2333794X221139413. doi:10.1177/2333794X221139413.

Nishimura, G., Dai, J., Lausch, E., et al. (2010). Spondylo-epiphyseal dysplasia, Maroteaux type (pseudo-Morquio syndrome type 2), and parastremmatic dysplasia are caused by TRPV4 mutations. *Am J Med Genet A,* 152A(6)**,** 1443-1449.

Nowaczyk, M. J. & Irons, M. B. (2012). Smith-Lemli-Opitz syndrome: phenotype, natural history, and epidemiology. *Am J Med Genet C Semin Med Genet,* 160C(4)**,** 250-262.

Nurchis, M. C., Radio, F. C., Salmasi, L., et al. (2024). Cost-Effectiveness of Whole-Genome vs Whole-Exome Sequencing Among Children With Suspected Genetic Disorders. *JAMA Netw Open*, 7(1), e2353514. doi:10.1001/jamanetworkopen.2023.53514.

Nurchis, M. C., Riccardi, M. T., Radio, F. C., et al. (2022). Incremental net benefit of whole genome sequencing for newborns and children with suspected genetic disorders: Systematic review and meta-analysis of cost-effectiveness evidence. *Health Policy,* 126(4)**,** 337-345.

Nurk, S., Koren, S., Rhie, A., et al. (2022). The complete sequence of a human genome. *Science,* 376(6588)**,** 44-53.

O'leary, N. A., Wright, M. W., Brister, J. R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res,* 44(1)**,** 733-745.

Oehler, J. B., Wright, H., Stark, Z., et al. (2023). The application of long-read sequencing in clinical settings. *Hum Genomics*, 17(1), 73. doi:10.1186/s40246-023-00522-3.

Ogunrin, O., Taiwo, F. & Frith, L. (2019). Genomic Literacy and Awareness of Ethical Guidance for Genomic Research in Sub-Saharan Africa: How Prepared Are Biomedical Researchers? *J Empir Res Hum Res Ethics,* 14(1)**,** 78-87.

Oladayo, A. M., Odukoya, O., Sule, V., et al. (2024). Perceptions and beliefs of community gatekeepers about genomic risk information in African cleft research. *BMC Public Health*, 24(1), 507. doi:10.1186/s12889-024-17987-z.

Owen, M. J., Lefebvre, S., Hansen, C., et al. (2022). An automated 13.5 hour system for scalable diagnosis and acute management guidance for genetic diseases. *Nat Commun*, 13(1), 4057. doi:10.1038/s41467-022-31446-6.

Owen, M. J., Wright, M. S., Batalov, S., et al. (2023). Reclassification of the Etiology of Infant Mortality With Whole-Genome Sequencing. *JAMA Netw Open*, 6(2), e2254069. doi:10.1001/jamanetworkopen.2022.54069.

Owolabi, P., Adam, Y. & Adebiyi, E. (2023). Personalizing medicine in Africa: current state, progress and challenges. *Front Genet*, 14(1), 1233338. doi:10.3389/fgene.2023.1233338.

Pasanen, A., Karjalainen, M. K., Finngen, et al. (2023). Meta-analysis of genome-wide association studies of gestational duration and spontaneous preterm birth identifies new maternal risk loci. *PLoS Genet*, 19(10), e1010982. doi:10.1371/journal.pgen.1010982.

Pereira, L., Mutesa, L., Tindana, P., et al. (2021). African genetic diversity and adaptation inform a precision medicine agenda. *Nat Rev Genet,* 22(5)**,** 284-306.

Pereira, S., Robinson, J. O., Gutierrez, A. M., et al. (2019). Perceived Benefits, Risks, and Utility of Newborn Genomic Sequencing in the BabySeq Project. *Pediatrics,* 143(Suppl 1)**,** 6-13.

Petrikin, J. E., Cakici, J. A., Clark, M. M., et al. (2018). The NSIGHT1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *NPJ Genomic Medicine*, 3(1), 6. doi:10.1038/s41525-018-0045-8.

Petrikin, J. E., Willig, L. K., Smith, L. D., et al. (2015). Rapid whole genome sequencing and precision neonatology. *Semin Perinatol,* 39(8)**,** 623-631.

Pettifor, J. M. (2017). The new Nelson Mandela Children's Hospital -a white elephant or an essential development for paediatric care in Johannesburg? *SAJCH,* 11(1)**,** 4-4.

Pfundt, R., Del Rosario, M., Vissers, L., et al. (2017). Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. *Genet Med,* 19(6)**,** 667-675.

Pichini, A., Ahmed, A., Patch, C., et al. (2022). Developing a National Newborn Genomes Program: An Approach Driven by Ethics, Engagement and Co-design. *Front Genet*, 13(1), 866168. doi:10.3389/fgene.2022.866168.

Porras, A. R., Bramble, M. S., Mosema Be Amoti, K., et al. (2021). Facial analysis technology for the detection of Down syndrome in the Democratic Republic of the Congo. *Eur J Med Genet*, 64(9), 104267. doi:10.1016/j.ejmg.2021.104267.

Powis, Z., Hagman, K. D. F., Speare, V., et al. (2018). Exome sequencing in neonates: diagnostic rates, characteristics, and time to diagnosis. *Genet Med,* 20(11)**,** 1468-1471.

Pughikumo, D. T. & Pugikumo, C. O. (2022). Physician knowledge, attitude and practice of medical genomics/genetically related diseases in Bayelsa State, Nigeria. *International Journal of Research Publication and Reviews,* 3(4)**,** 589-595.

Quaio, C., Obando, M. J. R., Perazzio, S. F., et al. (2021). Exome sequencing and targeted gene panels: a simulated comparison of diagnostic yield using data from 158 patients with rare diseases. *Genet Mol Biol*, 44(4), 20210061. doi:10.1590/1678-4685-GMB-2021-0061.

Quinonez, S. C. & Terefework, Z. (2021). The introduction of clinical genetic testing in Ethiopia: Experiences and lessons learned. *Am J Med Genet A,* 185(10)**,** 2995-3004.

R Core Team. 2024. *R: A language and environment for statistical computing* [Online]. Available: https://www.R-project.org/ [Accessed 01 February 2024].

Rabbani, B., Mahdieh, N., Hosomichi, K., et al. (2012). Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet,* 57(10)**,** 621-632.

Ramani, R., Krumholz, K., Huang, Y. F., et al. (2019). PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. *Bioinformatics,* 35(13)**,** 2320-2322.

Ramos, E., Callier, S. L. & Rotimi, C. N. (2012). Why personalized medicine will fail if we stay the course. *Per Med,* 9(8)**,** 839-847.

Ravenscroft, G., Clayton, J. S., Faiz, F., et al. (2021). Neurogenetic fetal akinesia and arthrogryposis: genetics, expanding genotype-phenotypes and functional genomics. *J Med Genet,* 58(9)**,** 609-618.

Recker, F., Zaniew, M., Bockenhauer, D., et al. (2015). Characterization of 28 novel patients expands the mutational and phenotypic spectrum of Lowe syndrome. *Pediatr Nephrol,* 30(6)**,** 931-943.

Rentzsch, P., Witten, D., Cooper, G. M., et al. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res,* 47(1)**,** 886-894.

Resta, R., Biesecker, B. B., Bennett, R. L., et al. (2006). A new definition of genetic counseling: National Society of Genetic Counselors' task force report. *Journal of genetic counseling,* 15(1)**,** 77-83.

Reutter, H., Keppler-Noreuil, K., C, E. K., et al. (2016). Genetics of Bladder-Exstrophy-Epispadias Complex (BEEC): Systematic Elucidation of Mendelian and Multifactorial Phenotypes. *Curr Genomics,* 17(1)**,** 4-13.

Rezapour, A., Souresrafil, A., Barzegar, M., et al. (2023). Economic evaluation of next-generation sequencing techniques in diagnosis of genetic disorders: A systematic review. *Clin Genet,* 103(5)**,** 513-528.

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

Riggs, E. R., Andersen, E. F., Cherry, A. M., et al. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med,* 22(2)**,** 245-257.

Roberts, R. J., Carneiro, M. O. & Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biol*, 14(7), 405. doi:10.1186/gb-2013-14-6-405.

Robertson, A. J., Tran, K., Patel, C., et al. (2023). Evolution of virtual gene panels over time and implications for genomic data re-analysis. *GIM Open*, 1(1), 100820. doi:10.1016/j.gimo.2023.100820.

Roman, T. S., Crowley, S. B., Roche, M. I., et al. (2020). Genomic Sequencing for Newborn Screening: Results of the NC NEXUS Project. *Am J Hum Genet,* 107(4)**,** 596-611.

Rothberg, J. M., Hinz, W., Rearick, T. M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature,* 475(7356)**,** 348-352.

Royer-Bertrand, B., Cisarova, K., Niel-Butschi, F., et al. (2021). CNV Detection from Exome Sequencing Data in Routine Diagnostics of Rare Genetic Disorders: Opportunities and Limitations. *Genes (Basel)*, 12(9), doi:10.3390/genes12091427.

Saib, M. Z., Dhada, B. L., Aldous, C., et al. (2021). Observed birth prevalence of congenital anomalies among live births at a regional facility in KwaZulu Natal Province, South Africa. *PLoS One*, 16(8), e0255456. doi:10.1371/journal.pone.0255456.

Sako, M., Nakanishi, K., Obana, M., et al. (2005). Analysis of NPHS1, NPHS2, ACTN4, and WT1 in Japanese patients with congenital nephrotic syndrome. *Kidney Int,* 67(4)**,** 1248-1255.

Saleh, S., Beyyumi, E., Al Kaabi, A., et al. (2021). Spectrum of neuro-genetic disorders in the United Arab Emirates national population. *Clin Genet,* 100(5)**,** 573-600.

Sanford, E., Wong, T., Ellsworth, K. A., et al. (2020). Clinical utility of ultra-rapid whole-genome sequencing in an infant with atypical presentation of WT1-associated nephrotic syndrome type 4. *Cold Spring Harb Mol Case Stud*, 6(4), a005470. doi:10.1101/mcs.a005470.

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

Satam, H., Joshi, K., Mangrolia, U., et al. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)*, 12(7), 997. doi:10.3390/biology12070997.

Saudi Mendeliome, G. (2015). Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol*, 16(1), 134. doi:10.1186/s13059-015-0693-2.

Saunders, C. J., Miller, N. A., Soden, S. E., et al. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med*, 4(154), 154ra135. doi:10.1126/scitranslmed.3004041.

Schmitz-Abe, K., Li, Q., Rosen, S. M., et al. (2019). Unique bioinformatic approach and comprehensive reanalysis improve diagnostic yield of clinical exomes. *Eur J Hum Genet,* 27(9)**,** 1398-1405.

Schobers, G., Schieving, J. H., Yntema, H. G., et al. (2022). Reanalysis of exome negative patients with rare disease: a pragmatic workflow for diagnostic applications. *Genome Med*, 14(1), 66. doi:10.1186/s13073-022-01069-z.

Schoeb, D. S., Chernin, G., Heeringa, S. F., et al. (2010). Nineteen novel NPHS1 mutations in a worldwide cohort of patients with congenital nephrotic syndrome (CNS). *Nephrol Dial Transplant,* 25(9)**,** 2970-2976.

Schoonen, M., Smuts, I., Louw, R., et al. (2019). Panel-Based Nuclear and Mitochondrial Next-Generation Sequencing Outcomes of an Ethnically Diverse Pediatric Patient Cohort with Mitochondrial Disease. *J Mol Diagn,* 21(3)**,** 503-513.

Schroeder, B. E., Gonzaludo, N., Everson, K., et al. (2021). The diagnostic trajectory of infants and children with clinical features of genetic disease. *NPJ Genom Med*, 6(1), 98. doi:10.1038/s41525-021-00260-2.

Shah, N., Hou, Y. C., Yu, H. C., et al. (2018). Identification of Misclassified ClinVar Variants via Disease Population Prevalence. *Am J Hum Genet,* 102(4)**,** 609-619.

Sharo, A. G., Zou, Y., Adhikari, A. N., et al. (2023). ClinVar and HGMD genomic variant classification accuracy has improved over time, as measured by implied disease burden. *Genome Med*, 15(1), 51. doi:10.1186/s13073-023-01199-y.

Sherry, S. T., Ward, M. H., Kholodov, M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res,* 29(1)**,** 308-311.

Smith, H. S., Ferket, B. S., Gelb, B. D., et al. (2023). Parent-Reported Clinical Utility of Pediatric Genomic Sequencing. *Pediatrics*, 152(2), e2022060318. doi:10.1542/peds.2022-060318.

Smith, H. S., Swint, J. M., Lalani, S. R., et al. (2020). Exome sequencing compared with standard genetic tests for critically ill infants with suspected genetic conditions. 22(8)**,** 1303-1310.

Smith, H. S., Swint, J. M., Lalani, S. R., et al. (2019). Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature. *Genet Med,* 21(1)**,** 3-16.

Soden, S. E., Saunders, C. J., Willig, L. K., et al. (2014). Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med*, 6(265), 265ra168. doi:10.1126/scitranslmed.3010076.

Solomon, G., Greenberg, J., Futter, M., et al. (2012). Understanding of genetic inheritance among Xhosa-speaking caretakers of children with hemophilia. *J Genet Couns,* 21(5)**,** 726-740.

Splinter, K., Adams, D. R., Bacino, C. A., et al. (2018). Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *N Engl J Med,* 379(22)**,** 2131-2139.

Springer, S., Erlewein, R., Naegele, T., et al. (2000). Alexander disease--classification revisited and isolation of a neonatal form. *Neuropediatrics,* 31(2)**,** 86-92.

Srivastava, S., Love-Nichols, J. A., Dies, K. A., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med,* 21(11)**,** 2413-2421.

Stark, Z. & Ellard, S. (2022). Rapid genomic testing for critically ill children: time to become standard of care? *Eur J Hum Genet,* 30(2)**,** 142-149.

Stark, Z., Foulger, R. E., Williams, E., et al. (2021). Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am J Hum Genet,* 108(9)**,** 1551-1557.

Stark, Z., Schofield, D., Alam, K., et al. (2017). Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. *Genet Med,* 19(8)**,** 867-874.

Stark, Z., Schofield, D., Martyn, M., et al. (2019). Does genomic sequencing early in the diagnostic trajectory make a difference? A follow-up study of clinical outcomes and cost-effectiveness. *Genet Med,* 21(1)**,** 173-180.

Stark, Z., Tan, T. Y., Chong, B., et al. (2016). A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med,* 18(11)**,** 1090-1096.

Stats Sa. (2022). General Household Survey 2021. STATISTICS SOUTH AFRICA. Pretoria, South Africa.

Stavropoulos, D. J., Merico, D., Jobling, R., et al. (2016). Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. *NPJ Genom Med*, 1(1), 15012. doi:10.1038/npjgenmed.2015.12.

Strande, N. T., Riggs, E. R., Buchanan, A. H., et al. (2017). Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet,* 100(6)**,** 895-906.

Suzuki, H., Nozaki, M., Yoshihashi, H., et al. (2022). Genome Analysis in Sick Neonates and Infants: High-yield Phenotypes and Contribution of Small Copy Number Variations. *J Pediatr,* 244(1)**,** 38-48.

Tajan, A., Riebel, A., Zavala, M. J., et al. (2024). Atypical presentation of Pearson syndrome in an infant with suspected myelodysplastic syndrome. *Pediatr Nephrol,* 39(2)**,** 447-450.

Takahashi, N., Hamada-Nakahara, S., Itoh, Y., et al. (2014). TRPV4 channel activity is modulated by direct interaction of the ankyrin domain to PI(4,5)P(2). *Nat Commun*, 5(1), 4994. doi:10.1038/ncomms5994.

Tammimies, K., Marshall, C. R., Walker, S., et al. (2015). Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *JAMA,* 314(9)**,** 895-903.

Tan, N. B., Stapleton, R., Stark, Z., et al. (2020). Evaluating systematic reanalysis of clinical genomic data in rare disease from single center experience and literature review. *Mol Genet Genomic Med*, 8(11), 1508. doi:10.1002/mgg3.1508.

Tan, T. Y., Dillon, O. J., Stark, Z., et al. (2017). Diagnostic Impact and Cost-effectiveness of Whole-Exome Sequencing for Ambulant Children With Suspected Monogenic Conditions. *JAMA Pediatr,* 171(9)**,** 855-862.

Tan, T. Y., Lunke, S., Chong, B., et al. (2019). A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *Eur J Hum Genet,* 27(12)**,** 1791-1799.

Tanner, S. M., Schneider, V., Thomas, N. S., et al. (1999). Characterization of 34 novel and six known MTM1 gene mutations in 47 unrelated X-linked myotubular myopathy patients. *Neuromuscul Disord,* 9(1)**,** 41-49.

Tekola-Ayele, F. & Rotimi, C. N. (2015). Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. *Public Health Genomics,* 18(4)**,** 242-247.

Tekola, F., Bull, S., Farsides, B., et al. (2009). Impact of social stigma on the process of obtaining informed consent for genetic research on podoconiosis: a qualitative study. *BMC Med Ethics*, 10(1), 13. doi:10.1186/1472-6939-10-13.

Telegrafi, A., Webb, B. D., Robbins, S. M., et al. (2017). Identification of STAC3 variants in non-Native American families with overlapping features of Carey-Fineman-Ziter syndrome and Moebius syndrome. *Am J Med Genet A,* 173(10)**,** 2763-2771.

Tembo, D., Abobo, F. D. N., Kaonga, P., et al. (2024). Risk factors associated with neonatal mortality among neonates admitted to neonatal intensive care unit of the University Teaching Hospital in Lusaka. *Sci Rep*, 14(1), 5231. doi:10.1038/s41598-024-56020-6.

Tennessen, J. A., Bigham, A. W., O'connor, T. D., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* 337(6090)**,** 64-69.

Thom, J. & Haw, T. (2021). Awareness of genetic counseling services among allied healthcare professionals in South Africa. *J Genet Couns,* 30(6)**,** 1649-1657.

Thormann, A., Halachev, M., Mclaren, W., et al. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat Commun*, 10(1), 2373. doi:10.1038/s41467-019-10016-3.

Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14(2)**,** 178-192.

Tian, Y., Pesaran, T., Chamberlin, A., et al. (2019). REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Sci Rep*, 9(1), 12752. doi:10.1038/s41598-019-49224-8.

Tiffin, N. (2014). Unique considerations for advancing genomic medicine in African populations. *Per Med,* 11(2)**,** 187-196.

Treangen, T. J. & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet,* 13(1)**,** 36-46.

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., et al. (2017). Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet,* 25(2)**,** 176-182.

Tsai, T. C., Horinouchi, H., Noguchi, S., et al. (2005). Characterization of MTM1 mutations in 31 Japanese families with myotubular myopathy, including a patient carrying 240 kb deletion in Xq28 without male hypogenitalism. *Neuromuscul Disord,* 15(3)**,** 245-252.

Uk Newborn Screening Programme Centre (2012). Health professional Handbook: a guide to newborn blood spot screening for healthcare professionals. UK Newborn Screening Programme Centre London.

United Nations (2023). The Sustainable Development Goals Report 2023: Special Edition Towards a Rescue Plan for People and The Planet. New York, NY, USA.

Uwineza, A. & Mutesa, L. (2015). Medical genetics and genomic medicine in Rwanda. *Mol Genet Genomic Med,* 3(6)**,** 486-489.

Van Der Merwe, N., Ramesar, R. & De Vries, J. (2022). Whole Exome Sequencing in South Africa: Stakeholder Views on Return of Individual Research Results and Incidental Findings. *Front Genet*, 13(1), 864822. doi:10.3389/fgene.2022.864822.

Van Diemen, C. C., Kerstjens-Frederikse, W. S., Bergman, K. A., et al. (2017). Rapid Targeted Genomics in Critically Ill Newborns. *Pediatrics*, 140(4), doi:10.1542/peds.2016-2854.

Veldman, A., Kiewiet, M. B. G., Heiner-Fokkema, M. R., et al. (2022). Towards Next-Generation Sequencing (NGS)-Based Newborn Screening: A Technical Study to Prepare for the Challenges Ahead. *Int J Neonatal Screen*, 8(1), 17. doi:10.3390/ijns8010017.

Verbitsky, M., Sanna-Cherchi, S., Fasel, D. A., et al. (2015). Genomic imbalances in pediatric patients with chronic kidney disease. *J Clin Invest,* 125(5)**,** 2171-2178.

Viering, D., Hureaux, M., Neveling, K., et al. (2023). Long-Read Sequencing Identifies Novel Pathogenic Intronic Variants in Gitelman Syndrome. *J Am Soc Nephrol,* 34(2)**,** 333-345.

Vill, K., Kuhn, M., Glaser, D., et al. (2015). Long-Term Observations in an Affected Family with Neurogenic Scapuloperoneal Syndrome Caused by Mutation R269C in the TRPV4 Gene. *Neuropediatrics,* 46(4)**,** 282-286.

Vissers, L., Van Nimwegen, K. J. M., Schieving, J. H., et al. (2017). A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet Med,* 19(9)**,** 1055-1063.

Waldrop, M. A., Boue, D. R., Sites, E., et al. (2017). Clinicopathologic Conference: A Newborn With Hypotonia, Cleft Palate, Micrognathia, and Bilateral Club Feet. *Pediatr Neurol,* 74(1)**,** 11-14.

Wang, H., Qian, Y., Lu, Y., et al. (2020). Clinical utility of 24-h rapid trio-exome sequencing for critically ill infants. *NPJ Genom Med*, 5(1), 20. doi:10.1038/s41525-020-0129-0.

Wang, X., Shen, X., Fang, F., et al. (2018). Phenotype-Driven Virtual Panel Is an Effective Method to Analyze WES Data of Neurological Disease. *Front Pharmacol*, 9(1), 1529. doi:10.3389/fphar.2018.01529.

Wartiovaara, J., Ofverstedt, L. G., Khoshnoodi, J., et al. (2004). Nephrin strands contribute to a porous slit diaphragm scaffold as revealed by electron tomography. *J Clin Invest,* 114(10)**,** 1475-1483.

Wenger, A. M., Guturu, H., Bernstein, J. A., et al. (2017). Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med,* 19(2)**,** 209-214.

Wenger, B. M., Patel, N., Lui, M., et al. (2021). A genotype-first approach to exploring Mendelian cardiovascular traits with clear external manifestations. *Genet Med,* 23(1)**,** 94-102.

Wiener, E. K., Buchanan, J., Krause, A., et al. (2023). Retrospective file review shows limited genetic services fails most patients - an argument for the implementation of

exome sequencing as a first-tier test in resource-constraint settings. *Orphanet J Rare Dis*, 18(1), 81. doi:10.1186/s13023-023-02642-4.

Wilczewski, C. M., Obasohan, J., Paschall, J. E., et al. (2023). Genotype first: Clinical genomics research through a reverse phenotyping approach. *Am J Hum Genet,* 110(1)**,** 3-12.

Wilkinson, D. J., Barnett, C., Savulescu, J., et al. (2016). Genomic intensive care: should we perform genome testing in critically ill newborns? *Arch Dis Child Fetal Neonatal Ed,* 101(2)**,** 94-98.

Williams, T. N. & Weatherall, D. J. (2012). World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harb Perspect Med*, 2(9), a011692. doi:10.1101/cshperspect.a011692.

Willig, L. K., Petrikin, J. E., Smith, L. D., et al. (2015). Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir Med,* 3(5)**,** 377-387.

Woerner, A. C., Gallagher, R. C., Vockley, J., et al. (2021). The Use of Whole Genome and Exome Sequencing for Newborn Screening: Challenges and Opportunities for Population Health. *Front Pediatr*, 9(1), 663752. doi:10.3389/fped.2021.663752.

Wojcik, M. H., Schwartz, T. S., Thiele, K. E., et al. (2019). Infant mortality: the contribution of genetic disorders. *J Perinatol,* 39(12)**,** 1611-1619.

Wojcik, M. H., Stadelmaier, R., Heinke, D., et al. (2021a). The Unrecognized Mortality Burden of Genetic Disorders in Infancy. *Am J Public Health,* 111(2)**,** 156-162.

Wojcik, M. H., Zhang, T., Ceyhan-Birsoy, O., et al. (2021b). Discordant results between conventional newborn screening and genomic sequencing in the BabySeq Project. *Genet Med,* 23(7)**,** 1372-1375.

Wonkam, A., Tekendo, C. N., Sama, D. J., et al. (2011). Initiation of a medical genetics service in sub-Saharan Africa: experience of prenatal diagnosis in Cameroon. *Eur J Med Genet,* 54(4)**,** 399-404.

Woodward, K. E., Murthy, P., Mineyko, A., et al. (2023). Identifying Genetic Susceptibility in Neonates With Hypoxic-Ischemic Encephalopathy: A Retrospective Case Series. *J Child Neurol,* 38(1-2)**,** 16-24.

Wright, C. F., Campbell, P., Eberhardt, R. Y., et al. (2023). Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. *N Engl J Med,* 388(17)**,** 1559-1571.

Wright, C. F., Eberhardt, R. Y., Constantinou, P., et al. (2021). Evaluating variants classified as pathogenic in ClinVar in the DDD Study. *Genet Med,* 23(3)**,** 571-575.

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

Wright, C. F., Fitzpatrick, D. R. & Firth, H. V. (2018a). Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet,* 19(5)**,** 253–268.

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

Xue, Y., Ankala, A., Wilcox, W. R., et al. (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. *Genet Med,* 17(6)**,** 444-451.

Yang, Y., Muzny, D. M., Xia, F., et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA,* 312(18)**,** 1870-1879.

Young, L. C., Hartig, N., Boned Del Rio, I., et al. (2018). SHOC2-MRAS-PP1 complex positively regulates RAF activity and contributes to Noonan syndrome pathogenesis. *Proc Natl Acad Sci U S A,* 115(45)**,** 10576-10585.

Zaharieva, I. T., Sarkozy, A., Munot, P., et al. (2018). STAC3 variants cause a congenital myopathy with distinctive dysmorphic features and malignant hyperthermia susceptibility. *Hum Mutat,* 39(12)**,** 1980-1994.

Zech, M., Boesch, S., Maier, E. M., et al. (2016). Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. *Am J Hum Genet,* 99(6)**,** 1377-1387.

Zech, M., Lam, D. D. & Winkelmann, J. (2019). Update on KMT2B-Related Dystonia. *Curr Neurol Neurosci Rep*, 19(11), 92. doi:10.1007/s11910-019-1007-y.

Zhao, K., Van Der Spoel, A., Castiglioni, C., et al. (2018). 19q13.12 microdeletion syndrome fibroblasts display abnormal storage of cholesterol and sphingolipids in the endo-lysosomal system. *Biochim Biophys Acta Mol Basis Dis,* 1864(6 Pt A)**,** 2108-2118.

Zhao, L., Liu, H., Yuan, X., et al. (2020). Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinformatics*, 21(1), 97. doi:10.1186/s12859-020-3421-1.

Zhao, M., Wang, Q., Wang, Q., et al. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics*, 14(Suppl 11), S1. doi:10.1186/1471-2105-14-S11-S1.

Zhong, A., Darren, B., Loiseau, B., et al. (2021a). Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: a systematic review. *Genet Med,* 23(12)**,** 2270-2280.

Zhong, A., Xia, K., Hadjis, Z., et al. (2021b). Opportunities and barriers for genetic service delivery in Kenya from a health personnel perspective. *J Community Genet,* 12(4)**,** 525-538.

Zimon, M., Baets, J., Auer-Grumbach, M., et al. (2010). Dominant mutations in the cation channel gene transient receptor potential vanilloid 4 cause an unusual spectrum of neuropathies. *Brain,* 133(6)**,** 1798-1809.

Zurfluh, M. R., Zschocke, J., Lindner, M., et al. (2008). Molecular genetics of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. *Hum Mutat,* 29(1), 167-175.

APPENDICES

Appendix A. Ethical Clearance

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R49 Dr N Carstens and Ms L Campbell

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M201199

NAME: (Principal Investigator)	Dr N Carstens and Ms L Campbell
DEPARTMENT:	School of Pathology Department of Human Genetics National Health Laboratory Service
PROJECT TITLE:	*Evaluating the utility of clinical sequencing to improve diagnostic services for critically ill infants in SA*
DATE CONSIDERED:	2020/11/27
DECISION:	Approved unconditionally
CONDITIONS:	New investigator added 2021/07/28
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Not applicable
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2021/05/31

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **November** and therefore reports and re-certification will be due in the month of **November** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator — Date: 01/06/2021

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

2022/06/17

Dr N Carstens and Ms L Campbell
School of Pathology
Department of Human Genetics
National Health Laboratory Service

Sent by e-mail to: Nadia.Carstens@wits.ac.za

Dear Dr Carstens

Re: **Protocol Ref No:** M201199
Protocol Title: *Evaluating the utility of clinical sequencing to improve diagnostic services for critically ill infants in SA*
Principal Investigators: Dr N Carstens and Ms L Campbell

Thank you for your letter of 2022/06/06.

I confirm that we have noted and approve of the addition of the Nelson Mandela Children's Hospital as an additional study site.

Thank you for keeping us informed.

Yours Sincerely

.....
Mr I Burns
For the Human Research Ethics Committee (Medical)

.....
Dr CB Penny, Chairperson, Human Research Ethics Committee (Medical)

Appendix B. Hospital Approval Letters

(a) Rahima Moosa Mother and Child Hospital

GAUTENG PROVINCE
HEALTH
REPUBLIC OF SOUTH AFRICA

Rahima Moosa Mother and Child Hospital
Enquiries: Adjunct Professor Ashraf Coovadia
Tel: 011 470 9284/9100
Email: Karen.Marshall@wits.ac.za

TITLE OF RESEARCH PROJECT:
"EVALUATING THE UTILITY OF CLINICAL SEQUENCING TO IMPROVE DIAGNOSTIC SERVICES FOR CRITICALLY ILL INFANTS IN SA"

NAME OF RESEARCHER:
Dr Nadia Carstens
Division of Human Genetics, School of Pathology,
National Health Laboratory Service
University of the Witwatersrand

NHRD REF NO: GP_202109_062

Dear Dr Carstens,

Permission is granted for you to conduct the research as indicated in the title above.

The terms under which this permission is granted is contained in the Researcher Declaration form that you have signed. Failure to comply with these conditions will result in the withdrawal of such permission.

It is crucial for you to inform the Research Coordinator, Karen Marshall of the actual start and end dates of your study. This could be done by e-mail.

Should the study commence more than 12 months after receipt of this approval letter you will have to go through the process of applying again.

You are strongly advised to keep a signed copy of the declaration form to ensure that the terms of this agreement are always complied with.

Yours sincerely,

DR NP MKABAYI
Chief Executive Officer
Rahima Moosa Mother and Child Hospital
2021.11.09

(b) Nelson Mandela Children's Hospital

Nelson Mandela Children's Hospital
A family dedicated to care

Date: 25 May 2022

Title of Research Project	:	**EVALUATING THE UTILITY OF CLINICAL SEQUENCING TO IMPROVE DIAGNOSTIC SERVICES FOR CRITICALLY ILL INFANTS IN SA**
Department of Programme	:	**Department of NICU and PICU**
Principal Investigator	:	**Dr Nadia Carstens**
Site Investigator	:	**Dr Khakhu Mathivha and Dr Porai Moshesh**

Dear Dr Nadia Carstens,

You have been given permission to conduct research at Nelson Mandela Children's Hospital. Please note that the following conditions should be met:

1. Ethical Approval should be sort from a Human Ethics Research Committee of any recognised institution. Once such approval is granted, it must be submitted to the Hospital Research and Ethics Committee for record-keeping.
2. There will be no cost to the hospital associated with conducting this research.
3. Once the research is completed, the Hospital will be notified, and a report of the results provided.
4. The hospital will be acknowledged in any publications and presentations that will be given in relation to this research.

Documents submitted and required:

Title of document	Required documents
Approval from Ethics Committee	
Ethics Clearance Certificate	
Research Protocol	
Report of the results	

All the best with your research

Recommended/Not Recommended

Prof. Hopewell Ntsinjana
Academic Committee Chairperson
Date:

Supported/Not Supported

Dr Andrew Grieve
Deputy Chairperson
Date: 25/5/2022

NPC Registration No: 2011/009134/08
Web address: http://www.nelsonmandelachildrenshospital.org
P.O. Box 797, Highlands North, 2037. 6 Jubilee Road, Parktown 2193. Tel: (+27 10) 133-0600

Directors
Phuthuma Nhleko (Chairperson) • Moses Kgosana • Richard Lebethe • Jacko Maree • Sibongile Mkhabela
Sam Mokgokong • Kgomotso Moroka • Moss Ngoasheng • Victor Sekese • Joe Seoloane • Sithembiso Velaphi • Martin Veller

Nelson Mandela Children's Hospital
A family dedicated to care

Title of Research Project : **EVALUATING THE UTILITY OF CLINICAL SEQUENCING TO IMPROVE DIAGNOSTIC SERVICES FOR CRITICALLY ILL INFANTS IN SA**

Supported/Not Supported

Dr Pinky Chirwa
Clinical Director
Date:

Approved/Not Approved

Dr Nonkululeko Boikhutso
Chief Executive Officer
Date: 26/04/2022

NPC Registration No: 2011/009134/08
Web address: http://www.nelsonmandelachildrenshospital.org
P.O. Box 797, Highlands North, 2037. 6 Jubilee Road, Parktown 2193. Tel: (+27 10) 133-0600

Directors
Phuthuma Nhleko (Chairperson) • Moses Kgosana • Richard Lebethe • Jacko Maree • Sibongile Mkhabela
Sam Mokgokong • Kgomotso Moroka • Moss Ngoasheng • Victor Sekese • Joe Seloane • Sithembiso Velaphi • Martin Veller

Appendix C. Phenotype Collection Rubric

Phenotype Form NICU Exomes Project Wits					
Please complete the phenotype form below.					
Recruitment Information:	Patient Name and/or Study Number				
	Recruiting Hospital				
	Parent Name(s) and Contact No.(s)				
	Referring Doctor Name and Contact No				
Birth Parameters	Birth Weight (g)				
	Birth weight centile				
	Height (cm)				
	Height centile				
	Head Circumference (cm)				
	Head circumference centile				
	APGARs				
	APGAR centile				
	Sex				
	Specify other sex				
Neurological Assessment	Has the infant experienced seizures?	YES		NO	
	Seizure details				
	Does the infant have neuropathy or nerve dysfunction?	YES		NO	
	Neuropathy or nerve dysfunction details				
	Is the infant encephalopathic?	YES		NO	
	Encephalopathy details				
	Does the infant have any brain malformations?	YES		NO	
	Brain malformation details				
Head and Facial Assessment	Does the infant have any dysmorphic features affecting the following:				
	Head and neck shape	YES		NO	
	Head and neck dysmorphism details				
	Eyes	YES		NO	
	Eye dysmorphism details				
	Ears	YES		NO	
	Ear dysmorphism details				

	Nose	YES		NO	
	Nose dysmorphism details				
	Mouth	YES		NO	
	Mouth details				
	Is the infant micro- or macrocephalic?	YES		NO	
	Micro-/Macrocephaly details				
	Is the infant hearing impaired?	YES		NO	
	Hearing impairment details				
	Is the infant visually impaired?	YES		NO	
	Visual impairment details				
Cardiovascular and Respiratory Assessment:	Does the infant have any cardiac defects or malformations?	YES		NO	
	Cardiac malformation details				
	Is the infant cardiomyopathic?	YES		NO	
	Cardiomyopathy details				
	Does the infant have any respiratory abnormalities or difficulties?	YES		NO	
	Respiratory difficulties details				
	Metabolic Assessment:				
	Does the infant have any inborn errors of metabolism?	YES		NO	
	Inborn errors of metabolism details				
	Does the display features of an autoimmune disorder?	YES		NO	
	Autoimmune disorder details				
Abdominal and Genitourinary Assessment:	Does the infant have any defects in the following organ systems:				
	Gastrointestinal	YES		NO	
	Gastrointestinal defect details				
	Liver	YES		NO	
	Liver defect details				
	Kidney	YES		NO	
	Kidney defect details				
	Genitourinary	YES		NO	
	Genitourinary defect details				
	Does the infant have any other abnormalities along the trunk?	YES		NO	
	Trunk abnormality details				
Dermatological Assessment:	Does the infant have any dermatological features?	YES		NO	

	Dermatological features details				
Skeletal and Muscle Assessment:	Does the infant have any skeletal abnormalities?	YES		NO	
	Skeletal abnormality details				
	Does the infant have any joint abnormalities?	YES		NO	
	Joint abnormality details				
	Does the infant have abnormal muscle tone?	YES		NO	
	Muscle tone details				
	Does the infant have any abnormalities of hands, feet or limbs?	YES		NO	
	Hands, feet and limb abnormality details				
General and Behavioural Assessment:	Is the infant failing to thrive?	YES		NO	
	Failing to thrive details				
	Does the infant have an abnormal weight or growth parameters?	YES		NO	
	Weight and growth details				
	Does the infant fatigue easily?	YES		NO	
	Fatigue details				
	Does the infant have feeding difficulties?	YES		NO	
	Feeding difficulties details				
Miscellaneous features and History:	Does the infant have any other relevant phenotypic features or history? Please provide details in the box below				
Thank you for completing					

Appendix D. Table of genes included in the NICU/African virtual gene panel

Table D.1. List of 1127 genes included in the NICU/African virtual gene panel. The gene symbol, the curated associated phenotypes, strength of the association, typical inheritance pattern, penetrance of the associated phenotype, and phenotype category are listed, as well as the genes evaluation status and categorization in the Babyseq and NC NEXUS gene lists.

Gene	Curated disease	Evidence for gene-disease association	Typical inheritance	Penetrance	BabySeq Category	BabySeq	NC NEXUS Category	NC NEXUS	Phenotype Category
AAAS	Achalasia-addisonianism-alacrimia syndrome	Definitive	AR	High	A	Yes		No	Neuromuscular
AARS	Charcot-Marie-Tooth disease	Strong	AD	High	A	Yes		No	Neuromuscular
ABCA12	Ichthyosis, congenital, autosomal recessive	Definitive	AR	High	A	Yes		No	Dermatological
ABCA3	Surfactant metabolism dysfunction, pulmonary, 3	Definitive	AR	High	A	Yes	2	Yes	Pulmonary
ABCA4	Stargardt disease	Definitive	AR	High	A	Yes		No	Hearing and vision loss
ABCB11	Cholestasis, progressive familial intrahepatic 2	Definitive	AR	High	A	Yes		No	Hepatic
ABCB4	Cholestasis, progressive familial intrahepatic 3	Definitive	AR	High	A	Yes		No	Hepatic
ABCC6	Pseudoxanthoma elasticum	Definitive	AR	High	A	Yes		No	Dermatological
ABCC8	Hyperinsulinemic hypoglycemia, familial	Definitive, Strong	AR, AD	High, Moderate	A	Yes	1	Yes	Metabolic
ABCC9	Cardiomyopathy, dilated, (CANTU syndrome) Hypertrichotic osteochondrodysplasia	Moderate, Strong	AD	High, Moderate	B, A	Yes	1	Yes	Multisystem
ABCD1	Adrenoleukodystrophy	Definitive	XLR	High	A	Yes	1	Yes	Neurological
ABCG5	Sitosterolemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ACAD8	Isobutyryl-CoA dehydrogenase deficiency	Definitive	AR	High	A	Yes		No	Metabolic
ACAD9	ACAD9 deficiency	Strong	AR	High	A	Yes		No	Multisystem
ACADM	Medium chain acyl CoA dehydrogenase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ACADS	Short-Chain Acyl-CoA Dehydrogenase (SCAD) Deficiency - Infantile	Definitive	AR	Moderate	C	Yes	2	Yes	Metabolic
ACADSB	2-Methylbutyrylglycinuria	Definitive	AR	Low	C	Yes	2	Yes	Metabolic
ACADVL	VLCAD deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ACAT1	Alpha-methylacetoacetic aciduria	Definitive	AR	High	A	Yes	1	Yes	Metabolic

ACE	Renal tubular dysgenesis	Strong	AR	High	A	Yes		No	Renal
ACOX1	Peroxisomal acyl-CoA oxidase deficiency	Strong	AR	High	A	Yes		No	Neurological
ACSF3	Combined malonic and methylmalonic aciduria	Strong	AR	High	A	Yes		No	Metabolic
ACTA1	Nemaline myopathy	Definitive	AD	High	A	Yes		No	Neuromuscular
ACTA2	Aortic aneurysm, familial thoracic	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
ACTB	Baraitser-Winter syndrome	Strong	AD	High	A	Yes		No	Developmental
ACTC1	Cardiomyopathy, dilated, Cardiomyopathy, familial hypertrophic	Moderate, Strong	AD	Moderate	B	Yes	1	Yes	Cardiac
ACTG1	Deafness, autosomal dominant , Baraitser-Winter syndrome 2	Strong	AD	High	A	Yes	1, 2	Yes	Hearing and vision loss
ACTG2	Megacystis-microcolon-intestinal hypoperistalsis syndrome	Strong	AD	High	A	Yes		No	Neuromuscular
ACTN1	Macrothrombocytopenia, Platelet bleeding disorder type 15	Defintive	AD	High	A	Yes	2	Yes	Haematological
ACTN2	Cardiomyopathy, familial hypertrophic	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
ACTN4	Glomerulosclerosis, focal segmental, 1	Definitive	AD	High	A	Yes		No	Renal
ACVR1	Fibrodysplasia ossificans progressiva	Definitive	AD	High	A	Yes		No	Neuromuscular
ACVRL1	Telangiectasia, hereditary hemorrhagic, type 2	Definitive	AD	High	A	Yes	1	Yes	Haematological
ADA	Severe combined immunodeficiency due to ADA deficiency	Definitive	AR	High	A	Yes	1	Yes	Immunological
ADAMTS13	Thrombotic thrombocytopenic purpura, familial	Definitive	AR	High	A	Yes		No	Haematological
ADAMTSL2	Geleophysic dysplasia 1	Strong	AR	High	A	Yes		No	Connective tissue
ADCY1	Deafness, autosomal recessive 44	Limited	AR	High		No	2	Yes	Hearing and vision loss
ADGRV1	Usher Syndrome, type 2C	Definitive	AR	High		No	1	Yes	Hearing and vision Loss
ADK	Hypermethioninemia due to adenosine kinase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
AGA	Aspartylglucosaminuria	Definitive	AR	High	A	Yes	2	Yes	Neurological
AGL	Glycogen storage disease IIIa	Definitive	AR	High	A	Yes	1	Yes	Metabolic
AGRN	Myasthenia, limb-girdle, familial	Strong	AR	High	A	Yes		No	Neuromuscular
AGXT	Hyperoxaluria, primary, type 1	Definitive	AR	High	A	Yes		No	Metabolic
AHCY	Hypermethioninemia with deficiency of S-Adenosylhomocysteine Hydrolase	Moderate	AR	High		No	2	Yes	Metabolic
AHI1	Joubert syndrome-3	Definitive	AR	High	A	Yes		No	Multisystem
AIFM1	Cowchock syndrome	Strong	XLR	High	A	Yes	2	Yes	Neurological

AIP	Pituitary adenoma	Definitive	AD	Moderate	B	Yes	1	Yes	Cancer
AIRE	Autoimmune polyendocrinopathy syndrome , type I, with or without reversible metaphyseal dysplasia	Definitive	AR	High	A	Yes		No	Immunological
AK2	Reticular Dysgenesis	Definitive	AR	High		No	1	Yes	Immunological
AKAP9	Long QT Syndrome 11	Disputed	AD	Unknown	C	Yes	2	Yes	Cardiac
AKR1D1	Bile acid synthesis defect, congenital, 2	Strong	AR	High	A	Yes		No	Hepatic
ALAS2	Anemia, sideroblastic, X-linked	Definitive	XLR	High	A	Yes		No	Haematological
ALB	Analbuminemia	Strong	AR, AD	High	A	Yes	2, 1	Yes	Metabolic
ALDH18A1	Cutis laxa, autosomal recessive, type IIIA	Strong	AR	High	A	Yes		No	Dermatological
ALDH3A2	Sjogren-Larsson syndrome	Definitive	AR	High	A	Yes	2	Yes	Multisystem
ALDH5A1	Succinic semialdehyde dehydrogenase deficiency	Definitive	AR	High	A	Yes		No	Metabolic
ALDH7A1	Pyridoxine-dependent epilepsy	Definitive	AR	High		No	1	Yes	Metabolic
ALDOB	Fructose intolerance	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ALG1	Congenital disorder of glycosylation, type Ik	Strong	AR	High	A	Yes		No	Metabolic
ALG12	Congenital disorder of glycosylation, type Ig	Strong	AR	High	A	Yes		No	Metabolic
ALG3	Congenital disorder of glycosylation, type Id	Strong	AR	High	A	Yes		No	Metabolic
ALG6	Congenital disorder of glycosylation, type Ic	Strong	AR	High	A	Yes		No	Metabolic
ALG8	Congenital disorder of glycosylation, type Ih	Strong	AR	High	A	Yes		No	Metabolic
ALMS1	Alstrom syndrome	Definitive	AR	High	A	Yes	1	Yes	Multisystem
ALOX12B	Ichthyosis, congenital, autosomal recessive	Strong	AR	High	A	Yes		No	Dermatological
ALOXE3	Ichthyosis, congenital, autosomal recessive	Strong	AR	High	A	Yes		No	Dermatological
ALPL	Hypophosphatasia	Definitive	AR	High	A	Yes		No	Skeletal
ALS2	Amyotrophic lateral sclerosis	Definitive	AR	High	A	Yes		No	Neuromuscular
ALX4	Parietal foramina 2	Strong	AD	High	A	Yes		No	Skeletal
AMELX	Amelogenesis imperfecta	Definitive	XLD	High	A	Yes		No	Skeletal
AMN	Megaloblastic anemia-1, Norwegian type	Strong	AR	High	A	Yes		No	Haematological
AMT	Hyperglycinaemia, non-ketotic	Strong	AR	High	A	Yes		No	Metabolic
ANK1	Spherocytosis	Definitive	AD	High	A	Yes		No	Haematological
ANK2	Long QT syndrome	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
ANKH	Craniometaphyseal dysplasia	Definitive	AD	High	A	Yes		No	Skeletal

ANKRD1	Cardiomyopathy, dilated	Moderate	AD	Moderate	B	Yes	2	Yes	Cardiac
ANKRD26	Thrombocytopenia 2	Strong	AD	High	A	Yes		No	Haematological
ANO10	Spinocerebellar ataxia, autosomal recessive 10	Strong	AR	High	A	Yes		No	Neurological
ANO5	Muscular dystrophy, limb-girdle, type 2L	Strong, Moderate	AR, AD	High	A	Yes		No	Neuromuscular
ANO6	Scott syndrome	Moderate	AR	High		No	2	Yes	Haematological
ANTXR2	Hyaline fibromatosis syndrome	Definitive	AR	High	A	Yes		No	Dermatological
AP3B1	Hermansky-Pudlak syndrome 2	Strong	AR	High	A	Yes	2	Yes	Multisystem
APC	Adenomatous polyposis coli and attenuated	Definitive	AD	High	A	Yes	1	Yes	Cancer
APOB	Apolipoprotein B deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
APOL1	Focal segmental glomerulosclerosis	Definitive	AR	Low		No	2	Yes	Renal
APTX	Ataxia, early-onset, with oculomotor apraxia and hypoalbuminemia	Definitive	AR	High	A	Yes	2	Yes	Neuromuscular
AR	Androgen insensitivity	Definitive	XLR	High	A	Yes		No	Sex development
ARFGEF2	Periventricular heterotopia with microcephaly	Strong	AR	High	A	Yes		No	Neurological
ARG1	Arginase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ARHGEF6	Mental Retardation, X-linked 46	Disputed	X-linked	High		No	2	Yes	Intellectual disability
ARID1B	Coffin-Siris syndrome	Strong	AD	High	A	Yes		No	Multisystem
ARMC4	Primary ciliary dyskinesia	Strong	AR	High	A	Yes	1	Yes	Pulmonary
ARSA	Metachromatic leukodystrophy	Definitive	AR	High	A	Yes	2	Yes	Neurological
ARSB	Mucopolysaccharidosis type VI (Maroteaux-Lamy)	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ARX	Lissencephaly, X-linked 2	Definitive	XLR	High	A	Yes		No	Neurological
ASAH1	Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy	Limited	AR	High		No	2	Yes	Neuromuscular
ASCL1	Central hypoventilation syndrome / Haddad syndrome	Unknown	AD	Unknown	C	Yes	2	Yes	Neurological
ASL	Argininosuccinic aciduria	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ASPA	Canavan disease	Definitive	AR	High	A	Yes		No	Neurological
ASS1	Citrullinemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ATM	Ataxia-telangiectasia	Definitive	AR	High	A	Yes		No	Neurological
ATP1A2	Hemiplegic migraine	Definitive	AD	High	A	Yes		No	Neurological
ATP2A1	Brody myopathy	Strong	AR	High	A	Yes		No	Neuromuscular

ATP6V0A2	Cutis laxa, autosomal recessive, type IIA	Strong	AR	High	A	Yes		No	Dermatological
ATP6V1B1	Renal tubular acidosis	Definitive	AR	High	A	Yes	1	Yes	Renal
ATP7A	Menkes syndrome, Occipital horn syndrome	Definitive	XLR	High	A	Yes	1	Yes	Multisystem
ATP7B	Wilson disease	Definitive	AR	High	A	Yes	1	Yes	Hepatic
ATP8B1	Cholestasis, progressive familial intrahepatic 1	Definitive	AR	High	A	Yes		No	Hepatic
ATRX	Alpha-thalassemia/mental retardation syndrome	Definitive	XLD	High	A	Yes		No	Intellectual disability
AUH	3-methylglutaconic aciduria, type I	Strong	AR	High	A	Yes		No	Metabolic
AVPR2	Diabetes insipidus, nephrogenic	Definitive	XLR	High	A	Yes		No	Renal
B3GALTL	Peters-Plus syndrome	Strong	AR	High	A	Yes		No	Multisystem
BAAT	Bile acid amidation defect	Strong	AR	High	A	Yes		No	Hepatic
BAG3	Cardiomyopathy, dilated	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
BBS1	Bardet-Biedl syndrome	Definitive	AR	High	A	Yes		No	Multisystem
BBS10	Bardet-Biedl syndrome	Definitive	AR	High	A	Yes		No	Multisystem
BBS12	Bardet-Biedl syndrome	Definitive	AR	High	A	Yes		No	Multisystem
BBS2	Bardet-Biedl syndrome	Definitive	AR	High	A	Yes		No	Multisystem
BBS4	Bardet-Biedl syndrome	Definitive	AR	High	A	Yes		No	Multisystem
BBS5	Bardet-Biedl syndrome	Strong	AR	High	A	Yes		No	Multisystem
BBS7	Bardet-Biedl syndrome	Strong	AR	High	A	Yes		No	Multisystem
BBS9	Bardet-Biedl syndrome	Strong	AR	High	A	Yes		No	Multisystem
BCHE	Increased sensitivity to choline ester anesthesia	Strong	AR	Unknown		No	2	Yes	Metabolic
BCKDHA	Maple syrup urine disease	Definitive	AR	High	A	Yes	1	Yes	Metabolic
BCKDHB	Maple syrup urine disease	Definitive	AR	High	A	Yes	1	Yes	Metabolic
BCS1L	Bjornstad syndrome	Definitive, Limited	AR	High	A	Yes	2	Yes	Multisystem
BDNF	Central hypoventilation syndrome	Unknown	AD	Unknown	C	Yes	2	Yes	Neurological
BICD2	Congenital spinal muscular atrophy	Strong	AD	High	A	Yes		No	Neurological
BIN1	Myopathy, centronuclear, autosomal recessive	Definitive	AR	High	A	Yes		No	Neuromuscular
BLM	Bloom syndrome	Definitive	AR	High	A	Yes	2	Yes	Dermatological
BLOC1S3	Hermansky-Pudlak Syndrome 8	Moderate	AR	High	C	Yes	2	Yes	Multisystem
BLOC1S6	Hermansky-Pudlak syndrome, 9	Defintitive	AR	High	C	Yes	2	Yes	Multisystem

BMPR1A	Juvenile polyposis syndrome	Definitive	AD	High	A	Yes	1	Yes	Gastrointestinal
BMPR2	Pulmonary hypertension, familial primary	Definitive	AD	Moderate	B	Yes	1	Yes	Pulmonary
BRAF	Cardiofaciocutaneous syndrome	Definitive	AD	High	A	Yes		No	Multisystem
BRCA2	Fanconi anemia, complementation group D1	Definitive	AR	High	A	Yes	1	Yes	Multisystem
BRIP1	Fanconi Anemia complementation group J, Familial ovaraian cancer	Definitive	AR, AD	High		No	1	Yes	Multisystem
BSCL2	Berardinelli-Seip lipodystrophy	Definitive	AR	High	A	Yes		No	Multisystem
BSND	Bartter syndrome with sensorineural deafness	Strong	AR	High	A	Yes	1	Yes	Renal
BTD	Biotinidase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
BTK	Agammaglobulinemia, X-linked 1	Definitive	XLR	High	A	Yes		No	Immunological
CA2	Osteopetrosis, autosomal recessive 3, with renal tubular acidosis	Definitive	AR	High	A	Yes		No	Skeletal
CACNA1A	Episodic ataxia, type 2	Definitive	AD	High	A	Yes		No	Neurological
CACNA1C	Brugada syndrome	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
CACNA1D	Sinoatrial Node Dysfunction and Deafness	Moderate	AR	High	C	Yes	2	Yes	Multisystem
CACNA1F	Night blindness, congenital stationary (complete), 1A, X-linked	Definitive	XLR	High	A	Yes		No	Hearing and vision loss
CACNA1S	Malignant hyperthermia susceptibility 5	Moderate	AD	Low	C	Yes	1	Yes	Skeletal
CACNB2	Brugada Syndrome 4	Moderate	AD	High	C	Yes	1	Yes	Cardiac
CALM1	Arrhythmia (Long QT Syndrome 14 / Ventricular tachycardia, catecholaminergic polymorphic, 4)	Definitive, Moderate	AD	High		No	1	Yes	Cardiac
CALM2	Long QT Syndrome 15	Definitive	AD	Moderate		No	1	Yes	Cardiac
CALR3	Cardiomyopathy, Familial Hypertrophic, 19	Limited	AD	Unknown		No	2	Yes	Cardiac
CAPN3	Muscular dystrophy, limb-girdle, type 2A	Definitive	AR	High	A	Yes		No	Neuromuscular
CARD11	Immunodeficiency 11	Definitive	AR	High		No	1	Yes	Immunological
CASK	Mental retardation and microcephaly with pontine and cerebellar hypoplasia	Definitive	XLD	High	A	Yes		No	Intellectual disability
CASQ2	Ventricular tachycardia, catecholaminergic polymorphic	Definitive	AR, AD	High	A	Yes		No	Cardiac
CASR	Hyperparathyroidism, neonatal severe, AD, w/ or w/o Bartter syndrome, hypercalcemia	Definitive	AR, AD	High		No	1	Yes	Metabolic
CATSPER2	Sensorineural Deafness and Male Infertility	Limited	AR	High		No	2	Yes	Hearing and vision loss
CAV1	Pulmonary hypertension, primary, 3	Definitive	AD	High		No	2	Yes	Pulmonary
CAV3	Caveolinopathy	Definitive	AD	High	A	Yes	2	Yes	Neuromuscular

CAVIN1	Lipodystrophy, congenital generalized, type 4	Strong	AR	High	A	Yes		No	Metabolic
CBL	Noonan syndrome-like disorder with or without juvenile meylomonocytic leukemia	Definitive	AD	High	A	Yes		No	Multisystem
CBLIF	Intrinsic factor deficiency	Moderate	AR	High		No	1	Yes	Immunological
CBS	Homocystinuria, B6-responsive and nonresponsive types	Definitive	AR	High	A	Yes	1	Yes	Metabolic
CC2D2A	Cilliopathy	Definitive	AR	High	A	Yes		No	Pulmonary
CCDC103	Ciliary dyskinesia, primary, 17	Definitive	AR	High	C	Yes	1	Yes	Pulmonary
CCDC114	Ciliary dyskinesia, primary, 20	Definitive	AR	High		No	1	Yes	Pulmonary
CCDC151	Ciliary dyskinesia, primary, 30	Definitive	AR	High		No	1	Yes	Pulmonary
CCDC39	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
CCDC40	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
CCDC50	Deafness, autosomal dominant 44	Limited	AD	High	C	Yes	2	Yes	Hearing and vision loss
CCDC65	Ciliary dyskinesia, primary, 27	Limited	AR	High		No	1	Yes	Pulmonary
CCNO	Ciliary dyskinesia, primary, 29	Definitive	AR	High		No	1	Yes	Pulmonary
CCNQ	Syndactyly - telecanthus - anogenital and renal malformations	Strong	XLD	High	A	Yes		No	Multisystem
CD247	Immunodeficiency due to defect in CD3-zeta	Definitive	AR	Unknown		No	2	Yes	Immunological
CD36	Platelet Glycoprotein IV Deficiency	Moderate	AR	Unknown	C	Yes	2	Yes	Haematological
CD3D	Immunodeficiency 19	Definitive	AR	High		No	1	Yes	Immunological
CD3E	Immunodeficiency 18 (Severe Combined Immunodeficiency)	Definitive	AR	High		No	1	Yes	Immunological
CD3G	Immunodeficiency 17, CD3 gamma deficient	Definitive	AR	Unknown		No	2	Yes	Immunological
CD40LG	Immunodeficiency, X-linked, with hyper-IgM	Definitive	XLR	High	A	Yes	1	Yes	Immunological
CDAN1	Anemia, congenital dyserythropoietic, type I	Definitive	AR	High	A	Yes		No	Haematological
CDC73	Hyperparathyroidism, familial primary	Definitive	AD	High		No	1	Yes	Metabolic
CDH1	Gastric cancer	Definitive	AD	Moderate	B	Yes	3	Yes	Cancer
CDH23	Deafness, autosomal recessive , Usher syndrome, type 1D	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
CDKL5	Epileptic encephalopathy, early infantile, 2	Definitive	XLD	High	A	Yes		No	Neurological
CDKN1C	Beckwith-Wiedemann syndrome	Definitive	AD	High	A	Yes		No	Growth
CDKN2A	Melanoma	Definitive	AD	Moderate	B	Yes	1	Yes	Cancer
CDSN	Hypotrichosis	Strong	AD	High	A	Yes		No	Dermatological

CEACAM16	Non-syndromic hearing loss, autosomal recessive or dominant	Strong, Moderate	AR, AD	High	C	Yes	2	Yes	Hearing and vision loss
CEL	MODY, type VIII (MODY8)	Moderate	AD	High		No	2	Yes	Metabolic
CENPF	Stromme Syndrome	Limited	AR	Moderate		No	2	Yes	Multisystem
CEP152	Seckel syndrome	Definitive	AR	High	A	Yes		No	Growth
CEP290	Joubert syndrome	Definitive	AR	High	A	Yes		No	Multisystem
CFAP298	Ciliary dyskinesia, primary, 26	Limited	AR	High		No	1	Yes	Pulmonary
CFC1	Congenital heart defects	Strong	AD	High	A	Yes		No	Cardiac
CFL2	Nemaline myopathy	Definitve	AR	High	A	Yes		No	Neuromuscular
CFP	Properdin deficiency, X-linked	Strong	XLR	High	A	Yes		No	Immunological
CFTR	Cystic fibrosis	Definitive	AR	High	A	Yes	1	Yes	Multisystem
CHAT	Congenital myasthenic syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
CHD2	Developmental delay, intellectual disability, epilepsy	Strong	AD	High	A	Yes		No	Developmental
CHD7	CHARGE syndrome	Definitive	AD	High	A	Yes		No	Multisystem
CHKB	Muscular dystrophy, congenital, megaconial type	Strong	AR	High	A	Yes		No	Neuromuscular
CHM	Choroideremia	Definitive	XLD	High	A	Yes		No	Hearing and vision loss
CHRNA1	Congenital myasthenic syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
CHRND	Congenital myasthenic syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
CHRNE	Congenital myasthenic syndrome	Definitive	AR	High	A	Yes		No	Neuromuscular
CHRNG	Pterygium syndrome	Strong	AR	High	A	Yes		No	Developmental
CIB2	Nonsyndromic hearing loss	Definitive	AR	High		No	1	Yes	Hearing and vision loss
CISD2	Wolfram syndrome 2	Strong	AR	High	C	Yes	1	Yes	Multisystem
CLCN5	Dent disease	Definitive	XLR	High	A	Yes		No	Renal
CLCNKB	Bartter syndrome, type 3	Strong	AR	Moderate		No	1	Yes	Renal
CLDN14	Hearing loss, non-syndromic, autosomal recessive	Strong	AR	High	A	Yes	1	Yes	Hearing and vision loss
CLDN19	Hypomagnesemia 5, renal, with ocular involvement	Strong	AR	High	A	Yes		No	Renal
CLN3	Ceroid lipofuscinosis, neuronal, 3	Definitive	AR	High	A	Yes	2	Yes	Neurological
CLN5	Ceroid lipofuscinosis, neuronal, 5	Definitive	AR	High	A	Yes	2	Yes	Neurological
CLN6	Ceroid lipofuscinosis, neuronal, 6	Definitive	AR	High	A	Yes	2	Yes	Neurological

CLN8	Ceroid lipofuscinosis, neuronal, 8	Strong	AR	High	A	Yes	2	Yes	Neurological
CLPP	Deafness, autosomal recessive 81 (aka Perrault syndrome 3)	Definitive	AR	High	C	Yes	1	Yes	Hearing and vision loss
CLRN1	Usher syndrome, type 3A	Strong	AR	High	A	Yes	1	Yes	Hearing and vision loss
CNGB3	Achromatopsia-3	Definitive	AR	High	A	Yes		No	Hearing and vision loss
COCH	Deafness, non-syndromic, autosomal dominant	Strong	AD	High	A	Yes	3	Yes	Hearing and vision loss
COL11A1	Stickler syndrome and Fibrochondrogenesis 1	Definitive	AD	High	A	Yes	1, 2	Yes	Multisystem
COL11A2	Otospondylomegaepiphyseal dysplasia	Definitive	AD	High	A	Yes	1, 2	Yes	Multisystem
COL17A1	Epidermolysis bullosa, junctional, non-Herlitz type	Definitive	AR	High	A	Yes		No	Dermatological
COL1A1	Osteogenesis imperfecta, type I	Definitive	AD	High	A	Yes	1	Yes	Skeletal
COL1A2	Osteogenesis imperfecta, type I to IV, Ehlers-Danlos syndrome	Definitive	AD, AR	High	A	Yes	1, 2	Yes	Skeletal, Connective tissue
COL2A1	Stickler syndrome, Achondrogenesis type 2, Platyspondylic dysplasia, spondylometaphyseal dysplasia, Kniest dysplasia, Stickler syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem disorder
COL3A1	Ehlers-Danlos syndrome, type IV	Definitive	AD	High	A	Yes	1	Yes	Connective tissue
COL4A3	Alport syndrome	Definitive	AD, AR	High	A	Yes	1	Yes	Multisystem
COL4A4	Alport syndrome	Definitive	AR	High	A	Yes	1	Yes	Multisystem
COL4A5	Alport syndrome	Definitive	XLD	High	A	Yes	1	Yes	Multisystem
COL5A1	Ehlers-Danlos syndrome, type I	Definitive	AD	High	A	Yes		No	Connective tissue
COL5A2	Ehlers-Danlos syndrome	Definitive	AD	High	A	Yes		No	Connective tissue
COL6A1	Ullrich congenital muscular dystrophy	Definitive	AR	High	A	Yes		No	Neuromuscular
COL6A2	Ullrich congenital muscular dystrophy	Definitive	AR	High	A	Yes		No	Neuromuscular
COL6A3	Ullrich congenital muscular dystrophy	Definitive	AR	High	A	Yes		No	Neuromuscular
COL7A1	Epidermolysis bullosa dystrophica	Definitive	AD	High	A	Yes		No	Dermatological
COL9A1	Stickler syndrome, type IV	Limited	AR	Moderate		No	1	Yes	Multisystem
COLQ	Congenital myasthenic syndrome	Definitive	AR	High	A	Yes		No	Neuromuscular
COQ6	Coenzyme Q10 deficiency, primary, 6	Limited	AR	Moderate	C	Yes	1	Yes	Multisystem

COQ8A	Coenzyme Q10 deficiency, primary, 4	Unknown	AR	High		No	2	Yes	Multisystem
CORO1A	Immunodeficiency 8	Definitive	AR	High		No	1	Yes	Immunological
CP	Aceruloplasminaemia	Definitive	AR	High	B	Yes	3	Yes	Metabolic
CPS1	Carbamoylphosphate synthetase I deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
CPT1A	Carnitine palmitoyltransferase I deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
CPT2	Carnitine palmitoyltransferase 2 deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
CREBBP	Rubinstein-Taybi syndrome	Definitive	AD	High	A	Yes		No	Multisystem
CRLF1	Crisponi syndrome	Definitive	AR	High	A	Yes		No	Multisystem
CRPPA	Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies), type A, 7	Definitive	AR	High	A	Yes		No	Neuromuscular
CRTAP	Osteogenesis imperfecta, type VII	Strong	AR	High	A	Yes		No	Skeletal
CRYAB	Cardiomyopathy, dilated, Myofibrillar myopathy	Moderate, Strong	AD	Moderate, High	B, A	Yes	2	Yes	Cardiac
CRYL1	Nonsyndromic hearing loss	Limited	AR	Moderate		No	1	Yes	Hearing and vision loss
CSF2RA	Pulmonary alveolar proteinosis	Strong	XLR	High	A	Yes	1	Yes	Pulmonary
CSF2RB	Surfactant metabolic dysfunction, pulmonary, 5 / Familial pulmonary alveolar proteinosis	Unknown	AR	High	C	Yes	2	Yes	Pulmonary
CSRP3	Cardiomyopathy, familial hypertrophic, 12	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
CSTB	Epilepsy, progressive myoclonic 1A	Defintive	AR	High	A	Yes		No	Neurological
CTC1	Coats plus syndrome	Strong	AR	High	A	Yes		No	Multisystem
CTNNA3	Arrhythmogenic right ventricular dysplasia 13	Limited	AD	Unknown		No	2	Yes	Cardiac
CTNS	Nephropathic cystinosis	Definitive	AR	High	A	Yes	1	Yes	Renal
CTSA	Galactosialidosis	Definitive	AR	Moderate		No	2	Yes	Multisystem
CTSD	Ceroid lipofuscinosis, neuronal, 10	Definitive	AR	High	A	Yes	2	Yes	Neurological
CTSK	Pycnodysostosis	Definitive	AR	High	A	Yes	2	Yes	Skeletal
CUBN	Megaloblastic anemia-1, Finnish type	Strong	AR	High	A	Yes		No	Haematological
CUL7	3-M syndrome	Definitive	AR	High	A	Yes		No	Skeletal
CYBA	Chronic granulomatous disease	Definitive	AR	High	A	Yes		No	Immunological
CYBB	Chronic granulomatous disease	Definitive	XLR	High	A	Yes		No	Immunological
CYP11A1	Adrenal insufficiency, congenital, with 46XY sex reversal, partial or complete	Strong	AR	High	A	Yes		No	Sex development

CYP11B1	Adrenal hyperplasia, congenital, due to 11-beta-hydroxylase deficiency	Definitive	AR	High	A	Yes	1	Yes	Sex development
CYP21A2	Adrenal hyperplasia, congenital, due to 21-hydroxylase deficiency	Definitive	AR	High	A	Yes	1	Yes	Sex development
CYP27A1	Cerebrotendinous xanthomatosis	Definitive	AR	High	A	Yes		No	Metabolic
CYP27B1	Vitamin D-dependent rickets, type I	Definitive	AR	High	A	Yes	1	Yes	Skeletal
CYP4F22	Ichthyosis, congenital, autosomal recessive	Strong	AR	High	A	Yes		No	Dermatological
D2HGDH	D-2-hydroxyglutaric aciduria	Strong	AR	High	A	Yes		No	Neurological
DBT	Maple syrup urine disease	Definitive	AR	High	A	Yes	1	Yes	Metabolic
DCLRE1C	Severe combined immunodeficiency, Athabascan type	Definitive	AR	High	A	Yes	1	Yes	Immunological
DCX	Lissencephaly, X-linked	Definitive	XLR	High	A	Yes		No	Neurological
DDB2	Xeroderma pigmentosum	Strong	AR	High	A	Yes		No	Dermatological
DDC	Aromatic L-amino acid decarboxylase deficiency	Strong	AR	High	A	Yes		No	Neurological
DES	Cardiomyopathy, dilated, Arrythmogenic right ventricular cardiomyopathy	Defintive, Moderate	AD	Moderate, High	A, B	Yes	1	Yes	Cardiac
DGUOK	Mitochondrial DNA depletion syndrome	Definitive	AR	High	A	Yes		No	Immunological
DHCR7	Smith-Lemli-Opitz syndrome	Definitive	AR	High	A	Yes	2	Yes	Developmental
DIABLO	Deafness, autosomal dominant 64	Limited	AD	Unknown	C	Yes	2	Yes	Hearing and vision loss
DIAPH1	Deafness, autosomal dominant 1	Definitive	AD	High	C	Yes	2	Yes	Hearing and vision loss
DKC1	Dyskeratosis congenita	Definitive	XLR	Moderate	B	Yes	2	Yes	Multisystem
DLD	Maple syrup urine disease, type III	Definitive	AR	High	A	Yes	2	Yes	Metabolic
DLL3	Spondylocostal dysostosis, autosomal recessive, 1	Strong	AR	High	A	Yes		No	Skeletal
DMD	Duchenne muscular dystrophy, Becker muscular dystrophy, Cardiomyopathy, dilated	Definitive	XLR, XLD	High, Moderate	A, B	Yes	2	Yes	Neuromuscular
DMP1	Hypophosphatemic rickets, AR	Strong	AR	High	A	Yes		No	Skeletal
DMPK	Myotonic dystrophy 1	Definitive	AD	High	A	Yes		No	Neuromuscular
DNAAF1	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
DNAAF11	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
DNAAF2	Ciliary dyskinesia, primary, 10	Strong	AR	High	C	Yes	1	Yes	Pulmonary
DNAAF3	Ciliary Dyskinesia, Primary, 2	Definitive	AR	High	C	Yes	1	Yes	Pulmonary
DNAAF4	Ciliary dyskinesia, primary, 25	Limiited	AR	High		No	1	Yes	Pulmonary

DNAAF5	Ciliary dyskinesia, primary, 18	Limited	AR	High		No	1	Yes	Pulmonary
DNAH1	Spermatogenic failure, Ciliary dyskinesia, primary, 37	Definitive, Limited	AR	Unknown		No	2	Yes	Pulmonary
DNAH11	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
DNAH5	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
DNAH8	Spermatogenic failure	Strong	AR	Unknown		No	2	Yes	Pulmonary
DNAI1	Primary ciliary dyskinesia	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
DNAI2	Ciliary dyskinesia, primary, 9, with or without situs inversus	Definitive	AR	High	C	Yes	1	Yes	Pulmonary
DNAJB13	Ciliary dyskinesia, primary, 34	Limited	AR	Moderate		No	1	Yes	Pulmonary
DNAJB6	Muscular dystrophy, limb girdle	Definitive	AD	High	A	Yes		No	Neuromuscular
DNAL1	Ciliary dyskinesia, primary, 16	Limited	AR	Moderate	C	Yes	1	Yes	Pulmonary
DNMT3B	Immunodeficiency-centromeric instability-facial anomalies syndrome 1	Strong	AR	High	A	Yes	1	Yes	Immunological
DOCK8	Hyper-IgE syndrome	Definitive	AR	High	A	Yes	1	Yes	Immunological
DOK7	Congenital myasthenic syndrome	Definitive	AR	High	A	Yes		No	Neuromuscular
DPAGT1	Congenital disorder of glycosylation, type Ij	Strong	AR	High	A	Yes		No	Metabolic
DRC1	Ciliary dyskinesia, primary, 21	Limited	AR	High		No	1	Yes	Pulmonary
DSC2	Arrhythmogenic right ventricular cardiomyopathy	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
DSG2	Arrhythmogenic right ventricular cardiomyopathy	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
DSP	Arrhythmogenic cardiomyopathy	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
DSPP	Deafness, autosomal dominant 36, with dentinogenesis	Definitive	AD	Moderate		No	2	Yes	Hearing and vision loss
DUOX2	Thyroid dyshormonogenesis	Definitive	AR	High	A	Yes	1	Yes	Metabolic
DUOXA2	Thyroid dyshormonogenesis 5	Moderate	AR	Moderate	C	Yes	1	Yes	Metabolic
EDA	Ectodermal dysplasia, hypohidrotic	Definitive	XLR	High	A	Yes		No	Dermatological
EDAR	Ectodermal dysplasia, hypohidrotic	Definitive	AR	High	A	Yes		No	Dermatological
EDARADD	Ectodermal dysplasia, hypohidrotic	Strong	AD	High	A	Yes		No	Dermatological
EDN3	Waardenburg syndrome, type 4B	Moderate, Limited	AD, AR	High, Moderate	C	Yes	1	Yes	Multisystem
EDNRB	Waardenburg syndrome, type 4A	Moderate, Limited	AD, AR	High, Moderate	C	Yes	1	Yes	Multisystem
EFEMP2	Cutis laxa, AR, type IB	Moderate	AR	High	C	Yes	2	Yes	Dermatological

EFHC1	Myoclonic epilepsy	Strong	AD	High	A	Yes		No	Neurological
EFTUD2	Mandibulofacial dysostosis with microcephaly	Definitive	AD	High	A	Yes		No	Developmental
EGR2	Charcot-Marie-Tooth disease	Definitive	AD, AR	High	A	Yes		No	Neuromuscular
EIF2AK3	Wolcott-Rallison syndrome	Definitive	AR	High	A	Yes		No	Multisystem
ELANE	Neutropenia, congenital	Definitive	AD	High	A	Yes		No	Immunological
ELN	Cutis laxa AD, Supravalvar aortic stenosis	Definitive	AD	High	A	Yes	1, 2	Yes	Dermatological
EMD	Muscular dystrophy, Emery-Dreifuss	Definitive	XLR	High	A	Yes		No	Neuromuscular
ENG	Telangiectasia, hereditary hemorrhagic, type 1	Definitive	AD	High	A	Yes	1	Yes	Haematological
ENPP1	Arterial calcification, generalized, of infancy, 1	Strong	AR	High	A	Yes		No	Skeletal
EPM2A	Epilepsy, progressive myoclonic 2A (Lafora)	Definitive	AR	High	A	Yes		No	Neurological
ERCC2	Xeroderma pigmentosum	Definitive	AR	High	A	Yes		No	Dermatological
ERCC4	Xeroderma pigmentosum	Definitive	AR	High		No	2	Yes	Dermatological
ERCC5	Xeroderma pigmentosum	Definitive	AR	High	A	Yes		No	Dermatological
ERCC6	Cockayne syndrome	Definitive	AR	High	A	Yes		No	Neurological
ERCC8	Cockayne syndrome	Definitive	AR	High	A	Yes		No	Neurological
ESCO2	Roberts syndrome	Definitive	AR	High	A	Yes		No	Skeletal
ESPN	Deafness, autosomal recessive and dominant	Definitive, Limited	AD, AR	High	C	Yes	1, 2	Yes	Hearing and vision loss
ESRRB	Hearing loss	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
ETFA	Glutaric acidemia IIA	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ETFB	Glutaric acidemia IIB	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ETFDH	Glutaric acidemia IIC	Definitive	AR	High	A	Yes	1	Yes	Metabolic
ETHE1	Leigh syndrome	Definitive	AR	High	A	Yes		No	Neurological
EVC	Ellis-van Creveld syndrome	Definitive	AR	High	A	Yes		No	Skeletal
EVC2	Ellis-van Creveld syndrome	Definitive	AR	High	A	Yes		No	Skeletal
EXT1	Exostoses, multiple, type 1	Definitive	AD	High	A	Yes		No	Skeletal
EXT2	Exostoses, multiple, type 2	Definitive	AD	High	A	Yes		No	Skeletal
EYA1	Branchiootorenal syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
EYA4	Nonsyndromic hearing loss, Dilated cardiomyopathy 1J	Definitive, Limited	AD	High	A	Yes	2	Yes	Multisystem

EZH2	Weaver syndrome 2	Definitive	AD	High	A	Yes		No	Multisystem
F10	Factor X deficiency	Definitve	AR	High		No	1	Yes	Haematological
F11	Factor XI deficiency	Definitive	AD, AR	High	A	Yes	1	Yes	Haematological
F12	Congenital factor XII deficiency, Angioedema, hereditary, type III	Definitive, Moderate	AD	Moderate		No	1	Yes	Haematological
F13A1	Factor XIIIA deficiency	Definitive	AR	Moderate		No	1	Yes	Haematological
F13B	Factor XIIIB deficiency	Definitive	AR	Moderate		No	1	Yes	Haematological
F2	Prothrombin deficiency	Definitive	AR	High	A	Yes	1	Yes	Haematological
F5	Factor V deficiency, Thrombophilia due to activated protein C resistance (Homozygote)	Definitive	AD, AR	High		No	1	Yes	Haematological
F7	Factor VII deficiency	Definitive	AR	Moderate		No	1	Yes	Haematological
F8	Hemophilia A	Definitive	XLR	High	A	Yes	1	Yes	Haematological
F9	Hemophilia B	Definitive	XLR	High	A	Yes	1	Yes	Haematological
FAH	Tyrosinemia, type I	Definitive	AR	High	A	Yes	1	Yes	Metabolic
FAM161A	Retinal dystrophy	Strong	AR	High	A	Yes		No	Hearing and vision loss
FAM20C	Osteosclerotic bone dysplasia	Strong	AR	High	A	Yes		No	Skeletal
FANCA	Fanconi anaemia	Definitive	AR	High	A	Yes	1	Yes	Multisystem
FANCB	Fanconi anaemia	Definitive	XLR	High	A	Yes	1	Yes	Multisystem
FANCC	Fanconi anaemia	Definitive	AR	High	A	Yes	1	Yes	Multisystem
FANCD2	Fanconi anaemia	Definitive	AR	High	A	Yes	1	Yes	Multisystem
FANCE	Fanconi anemia, complementation group E	Definitive	AR	High	C	Yes	1	Yes	Multisystem
FANCF	Fanconi anemia, complementation group F	Definitive	AR	High	C	Yes	1	Yes	Multisystem
FANCG	Fanconi anaemia	Definitive	AR	High	A	Yes	1	Yes	Multisystem
FANCI	Fanconi anaemia	Definitive	AR	High	A	Yes	1	Yes	Multisystem
FANCL	Fanconi anemia, complementation group L	Definitive	AR	High	C	Yes	1	Yes	Multisystem
FANCM	Fanconi-Anemia-Associated Polypeptide	Unknown	AR	High	C	Yes	2	Yes	Multisystem
FBLN5	Cutis laxa	Strong	AR	High	A	Yes	2	Yes	Dermatological
FBN1	Marfan's syndrome	Definitive	AD	High	A	Yes	1	Yes	Connective tissue
FBN2	Contractural arachnodactyly	Limited	AD	High	A	Yes		No	Connective tissue

FBP1	Fructose bisphosphatase deficiency	Limited	AR	Moderate		No	1	Yes	Metabolic
FGA	Afibrinogenaemia	Definitive	AR	High	A	Yes		No	Haematological
FGB	Afibrinogenaemia	Definitive	AR	High	A	Yes		No	Haematological
FGD1	Aarskog-Scott syndrome	Definitive	XLR	High	A	Yes		No	Developmental
FGD4	Charcot-Marie-Tooth disease	Definitive	AR	High	A	Yes		No	Neuromuscular
FGF3	Deafness, congenital with inner ear agenesis, microtia, and microdontia	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
FGFR1	Kallmann syndrome	Definitive	AD	High	A	Yes		No	Sex development
FGFR3	Achondroplasia, Hypochondroplasia, Crouzon syndrome with acanthosis nigricans, Thanatophoric dysplasia type 1, Muenke syndrome	Definitive	AD, AR	High	A	Yes	1	Yes	Skeletal
FGG	Afibrinogenaemia	Definitive	AD, AR	High	A	Yes		No	Haematological
FH	Fumarase deficiency	Definitive	AR	High	A	Yes	2	Yes	Metabolic
FHL1	Emery-Dreifuss muscular dystrophy	Strong	XLR	High	A	Yes		No	Neuromuscular
FKBP10	Osteogenesis imperfecta III	Strong	AR	Unknown		No		No	Skeletal
FKTN	Muscular dystrophy	Definitive	AR	High	A	Yes	2	Yes	Neuromuscular
FLCN	Birt-Hogg-Dube syndrome	Definitive	AD	High	A	Yes	4	Yes	Dermatological
FLNA	Otopalatodigital spectrum disorder	Definitive	XLR	High	A	Yes		No	Skeletal
FOLR1	Neurodegeneration due to cerebral folate transport deficiency	Definitive	AR	High		No	1	Yes	Neurological
FOXC1	Axenfeld-Rieger syndrome	Definitive	AD	High	A	Yes		No	Developmental
FOXC2	Lymphoedema, primary	Strong	AD	High	A	Yes		No	Developmental
FOXE1	Bamforth-Lazarus syndrome	Limited	AR	High	C	Yes	1	Yes	Metabolic
FOXF1	Alveolar capillary dysplasia with misalignment of pulmonary veins	Definitive	AD	High	A	Yes	2	Yes	Pulmonary
FOXP3	IPEX syndrome	Definitive	XLR	High	A	Yes		No	Immunological
FRAS1	Fraser syndrome	Definitive	AR	High	A	Yes		No	Developmental
FTL	Neuroferritinopathy	Strong	AD	High	A	Yes		No	Neurological
FUCA1	Fucosidosis	Definitive	AR	High	A	Yes	2	Yes	Metabolic
FXN	Friedreich ataxia	Definitive	AR	High	A	Yes		No	Neurological
G6PC	Glycogen storage disease Ia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
G6PC3	Neutropenia, congenital	Strong	AR	High	A	Yes		No	Immunological
G6PD	Glucose-6-phosphate dehydrogenase deficiency	Definitive	XLR	High	A	Yes	1	Yes	Haematological

GAA	Glycogen storage disease II (Pompe disease)	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GALC	Krabbe disease	Definitive	AR	High	A	Yes	2	Yes	Neurological
GALE	Galactose epimerase deficiency	Definitive	AR	Unknown		No	2	Yes	Metabolic
GALK1	Galactokinase deficiency with cataracts	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GALNS	Mucopolysaccharidosis IVA	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GALT	Galactosaemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GAMT	Guanidinoacetate methyltransferase deficiency	Definitive	AR	High		No	1	Yes	Metabolic
GAN	Giant axonal neuropathy	Definitive	AR	High	A	Yes		No	Neurological
GAS8	Ciliary dyskinesia, primary, 33	Limited	AR	Moderate		No	1	Yes	Pulmonary
GATA1	Dyserythropoietic anemia with thrombocytopenia	Strong	XLR	High	A	Yes	1	Yes	Haematological
GATA2	Immunodeficiency 21	Definitive	AD	High		No	1	Yes	Immunological
GATA3	Hypoparathyroidism, sensorineural deafness, and renal dysplasia	Definitive	AD	High		No	1	Yes	Multisystem
GATA4	Congenital heart defects	Definitive	AD	High	A	Yes		No	Cardiac
GATAD1	Cardiomyopathy, dilated, 2B	Limited	AR	Unknown	C	Yes	2	Yes	Cardiac
GATM	L-arginine:glycine amidinotransferase deficiency	Definitive	AR	High		No	1	Yes	Intellectual disability
GBA	Gaucher disease 1	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GBE1	Glycogen storage disease IV	Definitive	AR	High	A	Yes	2	Yes	Metabolic
GCDH	Glutaricaciduria, type I	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GCH1	Dystonia, dopa-responsive	Definitive	AD, AR	Moderate	B	Yes	1	Yes	Neuromuscular
GCK	Monogenic diabetes	Definitive	AD	High	A	Yes	1	Yes	Metabolic
GDAP1	Charcot-Marie-Tooth disease	Definitive	AR	High	A	Yes		No	Neuromuscular
GDF2	Pulmonary hypertension, Telangiectasia, hereditary hemorrhagic, type 5	Definitive, Moderate	AD	Unknown		No	2	Yes	Haematological
GDNF	Central hypoventilation syndrome	Unknown	AD	Unknown	C	Yes	2	Yes	Neurological
GFAP	Alexander disease	Definitive	AD	High	A	Yes		No	Neurological
GFM1	Combined oxidative phosphorylation deficiency 1, Leigh syndrome	Moderate	AR	High	A	Yes		No	Neurological
GFPT1	Congenital myasthenic syndrome, limb-girdle	Strong	AR	High	A	Yes		No	Neuromuscular
GGCX	Vitamin K-dependent coagulation defect, pulmonary arterial hypertension	Definitive, Moderate	AR	High		No	1	Yes	Haematological
GH1	Growth hormone deficiency, isolated, type IA recessive and dominant	Strong	AD, AR	High		No	1	Yes	Growth

GIPC3	Hearing loss	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
GJA1	Oculodentodigital dysplasia	Definitive	AD	Moderate	A	Yes		No	Multisystem
GJA5	Atrial fibrillation	Strong	AD	Moderate	B	Yes	3	Yes	Cardiac
GJB1	Charcot-Marie-Tooth neuropathy	Definitive	XLD	High	A	Yes		No	Neurological
GJB2	Deafness and palmoplantar keratoderma	Definitive	AR, AD	High	A	Yes	1	Yes	Multisystem
GJB6	Clouston syndrome	Definitive	AR	High		No	1	Yes	Dermatological
GJC2	Pelizaeus-Merzbacher-like disease	Strong	AR	High	A	Yes		No	Neurological
GLA	Fabry disease	Definitive	XLR	High	A	Yes	1	Yes	Metabolic
GLB1	Gangliosidosis GM1	Definitive	AR	High	A	Yes	2	Yes	Neurological
GLDC	Glycine encephalopathy	Definitive	AR	High	A	Yes		No	Neurological
GLI3	Greig cephalopolysyndactyly syndrome	Definitive	AD	High	A	Yes		No	Skeletal
GLRA1	Hyperekplexia, hereditary 1, autosomal dominant or recessive	Strong	AD	High	A	Yes		No	Neurological
GLUD1	Hyperinsulinism	Definitive	AD	High	A	Yes		No	Metabolic
GM2A	GM2-gangliosidosis, AB variant	Definitive	AR	High		No	2	Yes	Neurological
GNE	Inclusion body myopathy	Definitive	AR	High	A	Yes		No	Neuromuscular
GNMT	Glycine N-Methyltransferase Deficiency	Limited	AR	High		No	2	Yes	Metabolic
GNPTAB	Mucolipidosis II	Definitive	AR	High	A	Yes	2	Yes	Multisystem disorder
GNPTG	Mucolipidosis III gamma	Definitive	AR	High	A	Yes	2	Yes	Multisystem disorder
GNS	Mucopolysaccharidosis IIId	Definitive	AR	High	A	Yes	2	Yes	Multisystem disorder
GP1BA	Bernard-Soulier Syndrome, type A1	Definitive	AD, AR	High		No	2	Yes	Haematological
GP1BB	Bernard-Soulier Syndrome, type B	Definitive	AR	High		No	1	Yes	Haematological
GP6	Bleeding disorder, platelet-type, 11	Definitive	AR	High		No	1	Yes	Haematological
GP9	Bernard-Soulier Syndrome, type C	Definitive	AR	High		No	1	Yes	Haematological
GPC3	Simpson-Golabi-Behmel syndrome	Definitive	XLR	High	A	Yes		No	Multisystem
GPD1L	Brugada syndrome	Disputed	AD	Moderate	B	Yes	1	Yes	Cardiac
GPR143	Ocular albinism, type I	Definitive	XLR	High	A	Yes		No	Dermatological
GPR56	Polymicrogyria, bilateral frontoparietal	Definitive	AR	High	A	Yes		No	Neurological

GPR98	Usher syndrome	Definitive	AR	High	A	Yes		No	Hearing and vision loss
GPSM2	Chudley-McCullough syndrome	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
GRHL2	Nonsyndromic hearing loss	Strong	AD	Moderate	C	Yes	2	Yes	Hearing and vision loss
GRHPR	Hyperoxaluria, primary, type II	Strong	AR	High	A	Yes		No	Renal
GRXCR1	Deafness, autosomal recessive 25	Definitive	AR	High	C	Yes	1	Yes	Hearing and vision loss
GSDME	Hearing loss	Definitive	AD	High	A	Yes	2	Yes	Hearing and vision loss
GSS	Glutathione synthetase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
GUSB	Mucopolysaccharidosis VII	Definitive	AR	High	A	Yes	2	Yes	Metabolic
GYS1	Glycogen storage disease 0, muscle	Unknown	AR	Unknwon		No	2	Yes	Neuromuscular
GYS2	Glycogen storage disease 0	Definitive	AR	High	A	Yes	1	Yes	Neuromuscular
H19	Beckwith-Wiedemann Syndrome	Definitive	IMP	High	A	Yes		No	Growth
HADH	Hyperinsulinemic hypoglycemia, familial, 4	Definitive	AR	High	A	Yes	1	Yes	Metabolic
HADHA	Mitochondrial trifunctional protein deficiency	Definitive	AR	High	A	Yes	2	Yes	Metabolic
HADHB	Mitochondrial trifunctional protein deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
HAMP	Hemochromatosis, type 2B	Moderate	AR	High	C	Yes	1	Yes	Haematological
HARS	Usher Syndrome, type 3B	Refuted	AR	Moderate	B	Yes	2	Yes	Hearing and vision loss
HARS2	Perrault Syndrome 2	Limited	AR	High	C	Yes	2	Yes	Neurological
HBA1	Thalassaemia alpha	Definitive	AR	High	A	Yes		No	Haematological
HBA2	Thalassemia, alpha	Definitive	AR	High	A	Yes		No	Haematological
HBB	Beta-thalassemia	Definitive	AR	High	A	Yes	1	Yes	Haematological
HBG2	Cyanosis, transient neonatal	Moderate	AD	High		No	1	Yes	Cardiac
HCN4	Familialthoracic aortic aneurysm and aortic disssection	Limited	AD	Unknown	C	Yes	2	Yes	Cardiac
HDAC8	Cornelia de Lange syndrome-like features, ocular hypertelorism & large fontanelle	Definitive	XLD	High	A	Yes		No	Developmental
HEXA	Tay-Sachs disease	Definitive	AR	High	A	Yes	2	Yes	Neurological
HEXB	Sandhoff disease, infantile, juvenile, and adult forms	Definitive	AR	High	A	Yes	2	Yes	Neurological
HGD	Alkaptonuria	Definitive	AR	High	A	Yes	4	Yes	Metabolic

HGF	Deafness, autosomal recessive 39	Moderate	AR	High	C	Yes	1	Yes	Hearing and vision loss
HGSNAT	Mucopolysaccharidosis IIIC	Definitive	AR	High	A	Yes	2	Yes	Metabolic
HINT1	Charcot-Marie-Tooth Disease	Definitive	AR	High	A	Yes		No	Neurological
HJV	Hemochromatosis, type 2A	Strong	AR	High	C	Yes	1	Yes	Haematological
HLCS	Holocarboxylase synthetase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
HMGCL	3-hydroxy-3-methylglutaric aciduria	Definitive	AR	High	A	Yes	1	Yes	Metabolic
HMGCS2	HMG-CoA synthase-2 deficiency	Definitive	AR	High		No	1	Yes	Metabolic
HNF1A	MODY, type III (MODY3)	Definitive	AD	High		No	1	Yes	Metabolic
HNF1B	Renal cysts and diabetes syndrome (MODY5)	Definitive	AD	High	C	Yes	2	Yes	Metabolic
HNF4A	MODY, type I (MODY1)	Definitive	AD	High		No	1	Yes	Metabolic
HPD	Tyrosinemia, type III	Definitive	AR	Unknown	C	Yes	2	Yes	Metabolic
HPRT1	Lesch-Nyhan syndrome 1	Definitive	XLR	High	A	Yes		No	Metabolic
HPS1	Hermansky-Pudlak syndrome 1	Definitive	AR	High	A	Yes	1	Yes	Multisystem
HPS3	Hermansky-Pudlak syndrome 3	Definitive	AR	High	A	Yes	1	Yes	Multisystem
HPS4	Hermansky-Pudlak syndrome 4	Definitive	AR	High	A	Yes	1	Yes	Multisystem
HPS5	Hermansky-Pudlak syndrome 5	Definitive	AR	High	A	Yes	1	Yes	Multisystem
HPS6	Hermansky-Pudlak syndrome 6	Definitive	AR	High	C	Yes	2	Yes	Multisystem
HRAS	Costello syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
HSD11B2	Apparent mineralocorticoid excess	Moderate	AR	Moderate		No	1	Yes	Renal
HSD17B10	17-beta-hydroxysteroid dehydrogenase X deficiency	Definitive	XLD	High	A	Yes	2	Yes	Metabolic
HSD17B3	Pseudohermaphroditism, male, with gynecomastia	Definitive	AR	High	A	Yes		No	Sex development
HSD17B4	D-bifunctional protein deficiency, Perrault syndrome	Definitive	AR	High	A	Yes	2	Yes	Neurological
HSD3B2	3-beta-hydroxysteroid dehydrogenase, type II, deficiency	Moderate	AR	Moderate		No	1	Yes	Sex development
HSD3B7	3 beta-hydroxysteroid dehydrogenase deficiency	Definitive	AR	High	A	Yes		No	Sex development
HSPB8	Charcot-Marie-Tooth disease, axonal, type 2L	Definitive	AD	High	A	Yes		No	Neuromuscular
HSPG2	Schwartz-Jampel syndrome, Silverman-Handmaker dyssegmental dysplasia	Definitive	AR	High	A	Yes		No	Skeletal
HTRA1	CARASIL syndrome	Strong	AR	High	A	Yes		No	Neurological

HTT	Huntington disease	Definitive	AD	High		No		No	Neurological
HYAL1	Mucopolysaccharidosis type IX	Moderate	AR	Moderate		No	2	Yes	Metabolic
HYCC1	Hypomyelination and congenital cataract	Strong	AR	High	A	Yes		No	Neurological
HYDIN	Ciliary dyskinesia, primary, 5	Definitive	AR	High	C	Yes	1	Yes	Pulmonary
IDS	Mucopolysaccharidosis II	Definitive	XLR	High	A	Yes	1	Yes	Multisystem
IDUA	Mucopolysaccharidosis Ih	Definitive	AR	High	A	Yes	1	Yes	Multisystem
IGHMBP2	Spinal muscular atrophy with respiratory distress	Definitve	AR	High	A	Yes		No	Neuromuscular
IGSF1	Central hypothyroidism and testicular enlargement	Strong	XLR	High	A	Yes		No	Metabolic
IKBKAP	Dysautonomia, familial	Definitive	AR	High	A	Yes		No	Neurological
IKBKG	Incontinentia pigmenti 1	Definitive	XLD	High	A	Yes	2	Yes	Dermatological
IL10RA	Inflammatory bowel disease	Strong	AR	High	A	Yes		No	Gastrointestinal
IL21R	Immunodeficiency, primary, autosomal recessive, IL21R-related	Definitive	AR	High		No	2	Yes	Immunological
IL2RA	Interleukin-2 receptor, alpha chain, deficiency of	Limited	AR	High		No	1	Yes	Immunological
IL2RG	Severe combined immunodeficiency, X-linked	Definitive	XLR	High	A	Yes	1	Yes	Immunological
IL7R	Severe Combined Immunodeficiency, T-cell Negative, B-cell / Natural Killer Cell-Positive Type	Definitive	AR	High		No	1	Yes	Immunological
ILDR1	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
INS	Diabetes mellitus, permanent neonatal	Definitive	AD, AR	High		No	1	Yes	Metabolic
INSR	Leprechaunism	Definitive	AR	High	A	Yes		No	Metabolic
INVS	Nephronophthisis 2	Definitive	AR	High	A	Yes		No	Renal
IQCB1	Senior-Loken syndrome 5	Moderate	AR	High	A	Yes		No	Renal
IRF6	van der Woude syndrome	Definitive	AD	High	A	Yes		No	Developmental
ITGA2	Glycoprotein Ia deficiency / Bleeding disorder, platelet-type, 9	Limited	AD	High		No	2	Yes	Haematological
ITGA2B	Glanzmann thrombasthenia, Bleeding disorder, platelet-type, 16	Definitive, Moderate	AD, AR	High, Moderate		No	2	Yes	Haematological
ITGB3	Glanzmann thrombasthenia,Bleeding disorder, platelet-type, 16	Definitive, Moderate	AD, AR	High, Moderate		No	1	Yes	Haematological
ITGB4	Epidermolysis bullosa, junctional, with pyloric atresia	Definitive	AR	High	A	Yes		No	Dermatological
ITK	Lymphoproliferative syndrome 1	Definitive	AR	High		No	1	Yes	Metabolic
IVD	Isovaleric acidemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
IYD	Thyroid dyshormonogenesis 4	Moderate	AR	High	C	Yes	1	Yes	Metabolic

JAG1	Alagille syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
JAK3	SCID, autosomal recessive, T-negative/B-positive type	Strong	AR	High	A	Yes	1	Yes	Immunological
JPH2	Cardiomyopathy, Familial Hypertrophic, 17	Moderate	AD	Unknown	C	Yes	2	Yes	Cardiac
JPH3	Huntington disease-like 2	Strong	AD	High		No		No	Neurological
JUP	Arrhythmogenic right ventricular dysplasia 12, Naxos disease	Strong	AD, AR	Moderate, High	A, B	Yes	1, 2	Yes	Multisystem
KANSL1	Koolen-De Vries syndrome	Definitive	AD	High	A	Yes		No	Multisystem
KARS1	Deafness, autosomal recessive 89	Limited	AR	High	C	Yes	1	Yes	Hearing and vision loss
KAT6B	Genitopatellar syndrome and Say-Barber-Biesecker-Yooung-Simpson syndrome	Definitive	AD	High	A	Yes		No	Multisystem
KBTBD13	Nemaline myopathy	Moderate	AD	High	A	Yes		No	Neuromuscular
KCNA1	Episodic ataxia type 1	Definitive	AD	High	A	Yes		No	Neurological
KCNA5	Atrial fibrillation	Strong	AD	High	B	Yes	1	Yes	Cardiac
KCND3	Brugada Syndrome 9	Dsiputed	AD	Unknown	C	Yes	2	Yes	Cardiac
KCNE1	Long QT syndrome-5, Jervell and Lange-Nielsen syndrome	Definitive	AD, AR	High, Moderate	A, B	Yes	1	Yes	Cardiac
KCNE2	Long QT syndrome-6	Disputed	AD	Moderate	B	Yes	2	Yes	Cardiac
KCNE3	Brugada Syndrome 6	Disputed	AD	Unknown	C	Yes	2	Yes	Cardiac
KCNH2	Long QT syndrome-2	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
KCNJ1	Bartter syndrome	Strong	AR	High	A	Yes		No	Renal
KCNJ10	SESAME syndrome	Moderate	AR	High		No	1	Yes	Multisystem
KCNJ11	Monogenic diabetes	Definitive	AD	High	A	Yes	1	Yes	Metabolic
KCNJ2	Andersen cardiodysrhythmic periodic paralysis	Definitive	AD	High	A	Yes	1	Yes	Neuromuscular
KCNJ5	Long QT syndrome 13	Disputed	AD	High	C	Yes	2	Yes	Cardiac
KCNK3	Pulmonary hypertension, primary, 4	Definitive	AD	High		No	1	Yes	Pulmonary
KCNQ1	Long QT syndrome-1, Jervell and Lange-Nielsen syndrome, Short QT syndrome	Definitive, Strong	AD, AR	Moderate	A, B	Yes	1	Yes	Cardiac
KCNQ4	Deafness, autosomal dominant	Strong	AD	High	A	Yes	1	Yes	Hearing and vision loss
KCTD7	Epilepsy, progressive myoclonic	Strong	AR	High	A	Yes		No	Neurological
KDM6A	Kabuki syndrome 2	Strong	XLD	High	A	Yes		No	Multisystem
KIF21A	Fibrosis of extraocular muscles, congenital	Strong	AD	High	A	Yes		No	Neuromuscular

KIT	Piebaldism	Definitive	AD	High	A	Yes		No	Dermatological
KLHL40	Nemaline myopathy	Strong	AR	High	A	Yes		No	Neuromuscular
KLHL41	Nemaline myopathy	Strong	AR	High	A	Yes		No	Neuromuscular
KMT2D	Kabuki syndrome 1	Definitive	AD	High	A	Yes		No	Multisystem
KRAS	Noonan syndrome	Definitive	AD	High	A	Yes		No	Multisystem
KRT14	Epidermolysis bullosa simplex	Definitive	AD	High	A	Yes		No	Dermatological
KRT16	Pachyonychia congenita	Strong	AD	High	A	Yes		No	Dermatological
KRT17	Pachyonychia congenita	Strong	AD	High	A	Yes		No	Dermatological
KRT5	Epidermolysis bullosa simplex	Definitive	AD	High	A	Yes		No	Dermatological
KRT6A	Pachyonychia congenita	Strong	AD	High	A	Yes		No	Dermatological
L1CAM	X-linked hydrocephalus syndrome	Definitive	XLR	High	A	Yes		No	Neurological
LAMA2	Muscular dystrophy, congenital merosin-deficient	Definitive	AR	High	A	Yes		No	Neuromuscular
LAMA3	Epidermolysis bullosa, junctional	Definitive	AR	High	A	Yes		No	Dermatological
LAMA4	Cardiomyopathy, dilated, 1JJ	Limited	AD	Unknown	C	Yes	2	Yes	Cardiac
LAMB2	Pierson syndrome	Definitive	AR	High	A	Yes		No	Renal
LAMB3	Epidermolysis bullosa, junctional	Definitive	AR	High	A	Yes		No	Dermatological
LAMC2	Epidermolysis bullosa, junctional	Definitive	AR	High	A	Yes		No	Dermatological
LAMP2	Danon disease	Definitive	XLD	High	A	Yes	1	Yes	Metabolic
LARGE1	Walker-Warburg syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
LARS2	Perrault Syndrome 4	Strong	AR	High	C	Yes	1	Yes	Neurological
LBR	Pelger-Huet anomaly	Strong	AR	High	A	Yes		No	Haematological
LCK	Immunodeficiency 22	Moderate	AR	Unknown		No	2	Yes	Immunological
LDB3	Dilated cardiomyopathy	Limited	AD	High	C	Yes	1	Yes	Cardiac
LDLR	Hypercholesterolemia	Definitive	AD, AR	High	A	Yes	1	Yes	Metabolic
LDLRAP1	Hypercholesterolemia, familial, autosomal recessive	Definitive	AR	High		No	1	Yes	Metabolic
LEPR	Obesity, morbid, due to leptin receptor deficiency	Strong	AR	High	A	Yes		No	Metabolic
LHFPL5	Nonsyndromic hearing loss	Definitive	AR	High	A	Yes	2	Yes	Hearing and vision loss
LHX3	Pituitary hormone deficiency, combined	Strong	AR	High	A	Yes	1	Yes	Metabolic
LIFR	Stuve-Wiedemann syndrome	Definitive	AR	High	A	Yes		No	Multisystem

LIG4	Severe combined immunodeficiency with sensitivity to ionizing radiation (DNA ligase IV deficiency)	Definitive	AR	High	A	Yes	2	Yes	Immunological
LIPA	Wolman syndrome	Definitive	AR	High	A	Yes	2	Yes	Metabolic
LITAF	Charcot-Marie-Tooth disease	Moderate, Strong	AD	High	A	Yes		No	Neuromuscular
LMBRD1	Methylmalonic aciduria and homocystinuria	Definitive	AR	High	A	Yes	1	Yes	Metabolic
LMNA	Emery-Dreifuss muscular dystrophy 2, Dilated cardiomyopathy	Definitive, Strong	AD, AR	High, Moderate	A, B	Yes	1	Yes	Neuromuscular
LMOD3	Nemaline myopathy	Strong	AR	High	A	Yes		No	Neuromuscular
LMX1B	Nail patella syndrome	Definitive	AD	High	A	Yes		No	Multisystem
LOXHD1	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
LRP2	Donnai-Barrow syndrome	Strong	AR	High	A	Yes		No	Multisystem
LRP4	Cenani-Lenz syndactyly syndrome	Strong	AR	High	A	Yes		No	Skeletal
LRPPRC	Leigh syndrome	Definitive	AR	High	A	Yes		No	Neurological
LRSAM1	Charcot-Marie-Tooth disease	Strong	AD	High	A	Yes		No	Neuromuscular
LRTOMT	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
LTBP4	Cutis laxa, autosomal recessive, type IC	Strong	AR	High	A	Yes	2	Yes	Dermatological
LYST	Chediak-Higashi syndrome	Definitive	AR	High	A	Yes	2	Yes	Immunological
MAFB	Multicentric carpotarsal osteolysis syndrome	Strong	AD	High	A	Yes		No	Skeletal
MAGI2	Infantile spasms	Strong	AD	High	A	Yes		No	Neurological
MAGT1	X-linked immunodeficiency with magnesium defect, Epstein-Barr virus infection, and neoplasia (XMEN)	Disputed	X-linked	Unknown		No	2	Yes	Immunological
MAN2B1	Mannosidosis, alpha	Definitive	AR	High	A	Yes	2	Yes	Metabolic
MANBA	Beta Manosidosis	Definitive	AR	High		No	2	Yes	Metabolic
MAP2K1	Cardiofaciocutaneous syndrome	Definitive	AD	High	A	Yes		No	Multisystem
MAP2K2	Cardiofaciocutaneous syndrome	Definitive	AD	High	A	Yes		No	Multisystem
MARVELD2	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
MAT1A	Methionine Adenosyltransferase Deficiency, AR	Definitive	AR	High	C	Yes	2	Yes	Metabolic
MAX	Pheochromocytoma susceptibility	Definitive	AD	Moderate		No	1	Yes	Cancer
MBTPS2	Ichthyosis follicularis, alopecia & photophobia	Definitive	XLR	High	A	Yes		No	Dermatological

MCCC1	3-Methylcrotonyl-CoA carboxylase 1 deficiency	Definitive	AR	Moderate	B	Yes	2	Yes	Metabolic
MCCC2	3-Methylcrotonyl-CoA carboxylase 2 deficiency	Definitive	AR	Moderate	B	Yes	2	Yes	Metabolic
MCFD2	Factor V and Factor VIII deficiency, combined	Definitive	AR	High	A	Yes		No	Haematological
MCIDAS	Ciliary dyskinesia, primary	Definitive	AR	High		No	1	Yes	Pulmonary
MCOLN1	Mucolipidosis IV	Definitive	AR	High	A	Yes	2	Yes	Multisystem
MCPH1	Microcephaly 1, primary, autosomal recessive	Definitive	AR	High	A	Yes		No	Skeletal
MECP2	Rett syndrome	Definitive	XLD	High	A	Yes		No	Neurological
MED12	Intellectual disability	Strong	XLR	High	A	Yes		No	Intellectual disability
MEFV	Mediterranean fever, familial	Definitive	AD, AR	High	A	Yes	1	Yes	Immunological
MEGF10	Myopathy, areflexia, respiratory distress, and dysphagia, early-onset	Definitive	AR	High	A	Yes		No	Neuromuscular
MEN1	Multiple endocrine neoplasia I	Definitive	AD	High	A	Yes	1	Yes	Cancer
MFN2	Charcot-Marie-Tooth disease	Definitive	AD	High	A	Yes		No	Neuromuscular
MFSD8	Ceroid lipofuscinosis, neuronal	Definitive	AR	High	A	Yes	2	Yes	Neurological
MGP	Keutel syndrome	Strong	AR	High	A	Yes		No	Skeletal
MIR96	Deafness, autosomal dominant 50	Moderate	AD	High	C	Yes	2	Yes	Hearing and vision loss
MITF	Waardenburg syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
MKKS	Bardet-Biedl syndrome	Strong	AR	High	A	Yes		No	Multisystem
MKS1	Meckel syndrome	Strong	AR	High	A	Yes		No	Multisystem
MLC1	Megalencephalic leukoencephalopathy	Definitive	AR	High	A	Yes		No	Neurological
MLH1	Mismatch repair cancer syndrome, Lynch syndrome	Definitive	AR	High		No	2	Yes	Cancer
MLYCD	Malonyl-CoA decarboxylase deficiency	Definitive	AR	High	A	Yes	2	Yes	Metabolic
MMAA	Methylmalonic aciduria, vitamin B12-responsive	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MMAB	Methylmalonic aciduria, vitamin B12-responsive, due to defect in synthesis of adenosylcobalamin, cblB complementation type	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MMACHC	Methylmalonic aciduria and homocystinuria, cblC type	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MMADHC	Methylmalonic aciduria and homocystinuria, cblD type	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MOCS1	Molybdenum cofactor deficiency	Definitive	AR	High	A	Yes		No	Neurological
MOCS2	Molybdenum cofactor deficiency	Definitive	AR	High	A	Yes		No	Neurological
MPI	Congenital disorder of glycosylation 1b	Strong	AR	High	A	Yes	1	Yes	Metabolic

MPL	Amegakaryocytic thrombocytopaenia, congenital and thrombocytopenia 2	Definitive	AR	High	A	Yes		No	Haematological
MPV17	Mitochondrial DNA depletion syndrome, hepatic	Definitive	AR	High	A	Yes		No	Hepatic
MPZ	Charcot-Marie-Tooth disease	Definitive	AD	High	A	Yes		No	Neuromuscular
MSH2	Mismatch repair cancer syndrome, Lynch syndrome	Definitive	AR	High		No	2	Yes	Cancer
MSH6	Mismatch repair cancer syndrome, Lynch syndrome	Definitive	AR	High		No	2	Yes	Cancer
MSRB3	Deafness, autosomal recessive 74	Definitive	AR	High	C	Yes	1	Yes	Hearing and vision loss
MSX2	Parietal foramina 1, craniosynostosis	Definitive	AD	High	A	Yes		No	Skeletal
MTHFR	Homocystinuria due to MTHFR deficiency	Definitive	AR	Moderate	B	Yes	1	Yes	Metabolic
MTM1	Myotubular myopathy, X-linked	Definitive	XLR	High	A	Yes		No	Neuromuscular
MTR	Methylmalonic aciduria and homocystinuria	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MTRR	Methylmalonic aciduria and homocystinuria	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MTTP	Abetalipoproteinaemia	Strong	AR	High	A	Yes		No	metabolic
MUSK	Congenital myasthenic syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
MUT	Methylmalonic aciduria, mut(0) type	Definitive	AR	High	A	Yes	1	Yes	Metabolic
MUTYH	MUTYH-associated polyposis	Definitive	AR	High	A	Yes	3	Yes	Gastrointestinal
MVK	Hyperimmunoglobulin D and periodic fever syndrome	Strong	AR	High	A	Yes		No	Immunological
MYBPC3	Cardiomyopathy, familial hypertrophic	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
MYCN	Feingold syndrome	Strong	AD	High	A	Yes		No	Multisystem
MYH11	Aortic aneurysm, familial thoracic 4	Strong	AD	Moderate	B	Yes	1	Yes	Cardiac
MYH14	Deafness, autosomal dominant	Strong	AD	High	A	Yes	2	Yes	Hearing and vision loss
MYH2	Proximal myopathy and ophthalmoplegia	Strong	AR	High	A	Yes		No	Neuromuscular
MYH3	Arthrogryposis, distal	Definitive	AD	High	A	Yes		No	Neuromuscular
MYH6	Cardiomyopathy (Hypertrophic 14, dilated, 1EE)	Limited	AD	Unknown	C	Yes	2	Yes	Cardiac
MYH7	Hypertrophic cardiomyopathy, Dilated cardiomyopathy, Skeletal myopathy	Definitive	AD	Moderate, High	A, B	Yes	1	Yes	Cardiac
MYH9	Macrothrombocytopenia and progressive sensorineural deafness	Definitive	AD	High	A	Yes		No	Haematological
MYL2	Cardiomyopathy, familial hypertrophic, 10	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
MYL3	Cardiomyopathy, familial hypertrophic, 8	Definitive	AD	Moderate	B	Yes		No	Cardiac
MYLK	Aortic aneurysm, familial thoracic 7	Moderate	AD	Moderate	B	Yes	4	Yes	Cardiac

MYLK2	Hypertrophic cardiomyopathy 1, digenic	Disputed	AD	Unknown	C	Yes	2	Yes	Cardiac
MYO15A	Sensorineural hearing loss	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
MYO3A	Sensorineural hearing loss	Strong	AR	High	A	Yes	2	Yes	Hearing and vision loss
MYO6	Deafness	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
MYO7A	Usher syndrome and nonsyndromic hearing loss	Definitive	AD, AR	High	A	Yes	1, 2	Yes	Hearing and vision loss
MYOZ2	Cardiomyopathy, Familial Hypertrophic, 16	Disputed	AD	Unknown	C	Yes	2	Yes	Cardiac
MYPN	Congenital myopathy, Dilated cardiomyopathy	Definitive, Limited	AD, AR	Unknown	C	Yes	2	Yes	Neuromuscular
NAA10	Ogden syndrome	Definitive	X-linked	High	C	Yes	2	Yes	Developmental
NADK2	Encepahlopathy due to 2,4-dienoyl-CoA reductase deficiency	Moderate	AR	High		No	2	Yes	Neurological
NAGA	N-acetylgalactosaminidase alpha deficiency	Definitive	AR	High	A	Yes	4	Yes	Metabolic
NAGLU	Sanfilippo syndrome type B	Definitive	AR	High	A	Yes	2	Yes	Metabolic
NAGS	N-acetylglutamate synthetase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
NARS2	Leigh syndrome	Limited	AR	High		No	2	Yes	Neurological
NBEAL2	Gray Platelet syndrome	Definitive	AR	High		No	1	Yes	Haematological
NBN	Nijmegen breakage syndrome	Definitive	AR	High	A	Yes		No	Immunological
NCF1	Chronic granulomatous disease	Definitive	AR	High	A	Yes		No	Immunological
NCF2	Chronic granulomatous disease	Definitive	AR	High	A	Yes		No	Immunological
NDP	Norrie disease	Definitive	XLR	High	A	Yes		No	Hearing and vision loss
NEB	Nemaline myopathy	Definitive	AR	High	A	Yes		No	Neuromuscular
NEFL	Charcot-Marie-Tooth disease	Definitive	AD, AR	High	A	Yes		No	Neuromuscular
NEU1	Sialidosis	Definitive	AR	High	A	Yes	2	Yes	Metabolic
NEUROD1	Monogenic diabetes	Moderate, Limited	AD, AR	High		No	2	Yes	Metabolic
NEXN	Cardiomyopathy (Familial Hypertrophic 20, Dilated 1CC)	Moderate, Limited	AD	High	C	Yes	1	Yes	Cardiac
NF1	Neurofibromatosis, type 1	Definitive	AD	High	A	Yes	1	Yes	Multisystem
NF2	Neurofibromatosis 2	Definitive	AD	High	A	Yes	3	Yes	Multisystem

NFKB2	Immunodeficiency, common variable, 10	Definitive	AD	High		No	1	Yes	Immunological
NGLY1	Congenital disorder of glycosylation	Definitive	AR	High	A	Yes		No	Metabolic
NHEJ1	Severe combined immunodeficiency with microcephaly, growth retardation, and sensitivity to ionizing radiation (Cernunnos-XLF-deficiency)	Definitive	AR	High	A	Yes	2	Yes	Metabolic
NHLRC1	Myoclonic epilepsy of Lafora	Definitive	AR	High	A	Yes		No	Neurological
NIPAL4	Ichthyosis, autosomal recessive	Strong	AR	High	A	Yes		No	Dermatological
NIPBL	Cornelia de Lange syndrome	Definitive	AD	High	A	Yes		No	Developmental
NKX2-1	Choreoathetosis, hypothyroidism, and neonatal respiratory distress	Definitive	AD	High	A	Yes	1	Yes	Multisystem
NKX2-5	Congenital heart disease - conduction and myopathic	Definitive	AD	Moderate	B	Yes	2	Yes	Cardiac
NKX2-6	Persistent truncus arteriosus / Conotruncal heart defects	Limited	AR	High		No	1	Yes	Cardiac
NLRP3	Muckle-Wells syndrome / CINCA syndrome / Familial cold-induced inflammatory syndrome 1	Moderate	AD	High		No	1	Yes	Immunological
NME8	Ciliary dyskinesia, primary, 6	Limited	AR	High	C	Yes	2	Yes	Pulmonary
NODAL	Heterotaxy, Visceral	Limited	AD	Moderate		No	2	Yes	Cardiac
NOG	Symphalangism, proximal, 1A	Strong	AD	High	A	Yes		No	Skeletal
NOTCH2	Hajdu-Cheney syndrome and Alagile syndrome	Strong	AD	High	A	Yes	1	Yes	Skeletal
NOTCH3	Cerebral arteriopathy with subcortical infarcts and leukoencephalopathy	Definitive	AD	High	A	Yes		No	Neurological
NPC1	Niemann-Pick disease type C1	Definitive	AR	High	A	Yes	1	Yes	Metabolic
NPC2	Niemann-Pick disease type C2	Definitive	AR	High	A	Yes	1	Yes	Metabolic
NPHP1	Nephronophthisis	Definitive	AR	High	A	Yes		No	Renal
NPHP3	Nephronophthisis	Definitive	AR	High	A	Yes		No	Renal
NPHP4	Nephronophthisis	Definitive	AR	High	A	Yes		No	Renal
NPHS1	Congenital nephrotic syndrome, Finnish type	Definitive	AR	High	A	Yes		No	Renal
NR0B1	Congenital adrenal hypoplasia	Definitive	XLR	High	A	Yes		No	Developmental
NSD1	Sotos syndrome	Definitive	AD	High	A	Yes		No	Growth
NSDHL	CHILD syndrome	Strong	XLD	High	A	Yes		No	Developmental
NTRK1	Congenital insensitivity to pain with anhidrosis	Definitive	AR	High	A	Yes		No	Neuromuscular
OAT	Gyrate Atrophy of the Choroid and Retina	Definitive	AR	High		No	1	Yes	Hearing and vision loss
OBSL1	3-M syndrome	Strong	AR	High	A	Yes		No	Skeletal

OCA2	Albinism, oculocutaneous	Definitive	AR	High	A	Yes		No	Dermatological
OCRL	Lowe oculocerebrorenal syndrome	Definitive	XLR	High	A	Yes		No	Multisystem
ODAD4	Ciliary dyskinesia, primary, 35	Definitive	AR	High		No	1	Yes	Pulmonary
OFD1	Oral-facial-digital syndrome	Definitive	XLD	High	A	Yes	2	Yes	Developmental
OPA1	Optic atrophy 1	Definitive	AD	High	A	Yes	2	Yes	Hearing and vision loss
OPA3	3-methylglutaconic aciduria, type III	Strong	AR	High	A	Yes	2	Yes	Metabolic
OPLAH	5-oxoprolinase deficiency	Limited	AD, AR	High		No	2	Yes	Metabolic
ORC1	Meier-Gorlin syndrome	Strong	AR	High	A	Yes		No	Growth
OSBPL2	Nonsyndromic hearing loss	Moderate	AD	High		No	2	Yes	Hearing and vision loss
OSMR	Amyloidosis, primary cutaneous	Strong	AD	High	A	Yes		No	Dermatological
OSTM1	Osteopetrosis	Strong	AR	High	A	Yes		No	Skeletal
OTC	Ornithine transcarbamylase deficiency	Definitive	XLR	High	A	Yes	1	Yes	Metabolic
OTOA	Nonsyndromic hearing loss	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
OTOF	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
OTOG	Deafness, autosomal recessive 18B	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
OTOGL	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
P2RX2	Deafness, autosomal dominant 41	Moderate	AD	High		No	2	Yes	Hearing and vision loss
P2RY12	Bleeding disorder, platelet-type, 8	Moderate	AD, AR	High		No	1	Yes	Haematological
PAH	Phenylketonuria	Definitive	AR	High	A	Yes	1	Yes	Metabolic
PAK3	Mental retardation syndrome, X-linked	Definitive	XLR	High	A	Yes		No	Intellectual disability
PALB2	Fanconi anemia, complementation group N	Definitive	AR	High		No	1	Yes	Multisystem
PANK2	Neurodegeneration with brain iron accumulation 1	Definitive	AR	High	A	Yes		No	Neurological
PAX3	Waardenburg syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
PAX6	Aniridia	Definitive	AD	High	A	Yes		No	Hearing and vision loss

PAX8	Hypothyroidism, congenital, due to thyroid dysgenesis or hypoplasia	Strong	AD	High	A	Yes	1	Yes	Metabolic
PC	Pyruvate carboxylase deficiency	Definitive	AR	High	A	Yes		No	Neurological
PCBD1	Hyperphenylalaninemia, BH4-deficient, D	Definitive	AR	High		No	2	Yes	Metabolic
PCCA	Propionicacidemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
PCCB	Propionicacidemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
PCDH15	Usher syndrome	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
PCNT	Microcephalic osteodysplastic primordial dwarfism type 2	Definitive	AR	High	A	Yes		No	Skeletal
PCSK9	Hypercholesterolemia	Definitive	AD	Moderate	B	Yes	1	Yes	Metabolic
PDE4D	Acrodysostosis 2, with or without hormone resistance	Strong	AD	High	A	Yes		No	Skeletal
PDHA1	Pyruvate dehydrogenase deficiency	Definitive	XLD	High	A	Yes		No	Neurological
PDHX	Leigh syndrome	Definitive	AR	High	A	Yes		No	Neurological
PDX1	Pancreatic agenesis 1, and monogenic diabetes	Definitive, Moderate	AD, AR	High		No	1, 2	Yes	Multisystem
PDZD7	Autosomal recessive hearing loss	Definitive	AR	Unknown		No	2	Yes	Hearing and vision loss
PEX1	Zellweger syndrome	Definitive	AR	High	A	Yes	2	Yes	Multisystem
PEX10	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX12	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX13	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX2	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX26	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX3	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX5	Zellweger syndrome	Strong	AR	High	A	Yes		No	Multisystem
PEX6	Zellweger syndrome	Definitive	AR	High	A	Yes	2	Yes	Multisystem
PFKM	Glycogen storage disease 7	Definitive	AR	High	A	Yes	2	Yes	Metabolic
PGAM2	Glycogen storage disease X	Limited	AR	Moderate		No	2	Yes	Metabolic
PHF6	Borjeson-Forssman-Lehmann syndrome	Definitive	XLD	High	A	Yes		No	Developmental
PHKA2	Phosphorylase kinase deficiency	Strong	XLR	High	A	Yes	1	Yes	Metabolic
PHKB	Phosphorylase kinase deficiency	Strong	AR	High	A	Yes		No	Metabolic
PHKG2	Phosphorylase kinase deficiency	Strong	AR	High	A	Yes	1	Yes	Metabolic

PHOX2B	Central hypoventilation syndrome	Definitive	AD	Moderate	B	Yes	1	Yes	Neurological
PHYH	Refsum disease	Strong	AR	High	A	Yes		No	Neurological
PIEZO2	Arthrogryposis, distal, type 5	Strong	AD	High	A	Yes		No	Neuromuscular
PIK3CD	Immunodeficiency 14	Definitive	AD, AR	High		No	1	Yes	Immunological
PINK1	Parkinson disease 6, early onset	Definitive	AR	High	A	Yes		No	Neurological
PJVK	Hearing loss	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
PKD1	Polycystic kidney disease	Definitive	AD	High	A	Yes		No	Renal
PKD2	Polycystic kidney disease	Definitive	AD	High	A	Yes		No	Renal
PKHD1	Polycystic kidney and hepatic disease	Definitive	AR	High	A	Yes		No	Renal
PKLR	Pyruvate kinase deficiency	Definitive	AR	High	A	Yes		No	Neurological
PKP2	Arrhythmogenic right ventricular dysplasia 9	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
PLA2G6	Infantile neuroaxonal dystrophy 1	Strong	AR	High	A	Yes		No	Neurological
PLAU	Quebec Platelet Disorder	Moderate	AD	High		No	1	Yes	Haematological
PLCE1	Nephrotic syndrome	Strong	AR	High	A	Yes		No	Renal
PLCG2	Autoinflammation and PLCG2-associated antibody deficiency and immune dysregulation (APLAID)	Unknown	AD	Unknown		No	2	Yes	Immunological
PLG	Plasminogen deficiency	Definitive	AR	High	A	Yes		No	Haematological
PLN	Cardiomyopathy, dilated	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
PLOD1	Ehlers-Danlos syndrome, kyphoscoliotic type	Strong	AR	High	A	Yes		No	Connnective tissue
PMM2	Congenital disorder of glycosylation, type Ia	Definitive	AR	High	A	Yes		No	Metabolic
PMP22	Charcot-Marie-Tooth disease	Definitive	AD	High	A	Yes		No	Neuromuscular
PMS2	Mismatch repair cancer syndrome, Lynch syndrome	Definitive	AR	High		No	2	Yes	Cancer
PNKD	Paroxysmal nonkinesiogenic dyskinesia	Definitive	AD	High	A	Yes		No	Neurological
PNKP	Microcephaly - seizures - developmental delay	Strong	AR	High	A	Yes		No	Multisystem
PNP	Immunodeficiency due to purine nucleoside phosphorylase deficiency	Moderate	AR	High		No	2	Yes	Immunological
PNPO	Epileptic encephalopathy, neonatal	Strong	AR	High	A	Yes	1	Yes	Neurological
POLG	POLG-Related Ataxia Neuropathy Spectrum Disorders	Definitive	AR	High	A	Yes		No	Neurological
POLH	Xeroderma pigmentosum	Definitive	AR	High	A	Yes		No	Dermatological
POLR1C	Leukodystrophy, hypomyelinating, 11, Treacher Collins syndrome 3	Limited	AR	High		No	2	Yes	Developmental

POLR1D	Treacher Collins syndrome 2	Limited	AD, AR	High		No	1	Yes	Developmental
POMT2	Muscular dystrophy-dystroglycanopathy (limb-girdle), type C, 2	Definitive	AR	High	A	Yes		No	Neuromuscular
POR	Disordered steroidogenesis with and without Antley-Bixler syndrome	Definitive	AR	High	A	Yes		No	Skeletal
PORCN	Focal dermal hypoplasia	Definitive	XLD	High	A	Yes		No	Dermatological
POU1F1	Pituitary hormone deficiency	Strong	AR	High	A	Yes	1	Yes	Metabolic
POU3F4	Deafness, X-linked	Definitive	XLR	High	A	Yes	1	Yes	Hearing and vision loss
POU4F3	Deafness, autosomal dominant	Definitive	AD	High	A	Yes	1	Yes	Hearing and vision loss
PPT1	Neuronal ceroid lipofuscinosis	Definitive	AR	High	A	Yes	2	Yes	Neurological
PQBP1	Mental retardation	Definitive	XLR	High	A	Yes		No	Intellectual disability
PRDM16	Cardiomyopathy, dilated 1LL / Left Ventricular Noncompaction 8	Limited	AD	Unknown	C	Yes	2	Yes	Cardiac
PRF1	Hemophagocytic lymphohistiocytosis, familial, 2	Definitive	AR	High	A	Yes		No	Immunological
PRKACG	Bleeding disorder, platelet-type, 19	Unknown	AR	High		No	2	Yes	Haematological
PRKAG2	Cardiomyopathy, hypertrophic, Wolff-Parkinson-White syndrome	Definitive, Strong	AD	Moderate, High	B	Yes	1, 2	Yes	Cardiac
PRKAR1A	Carney complex	Definitive	AD	High	A	Yes		No	Cancer
PRKDC	SCID	Definitive	AR	Unknown		No	2	Yes	Immunological
PROC	Thrombophilia due to protein C deficiency	Definitive	AD, AR	High	A	Yes	1	Yes	Haematological
PROKR2	Hypogonadotropic hypogonadism	Definitive	AD	High	A	Yes		No	Sex development
PROP1	Pituitary hormone deficiency, combined, 2	Strong	AR	High	A	Yes	1	Yes	Metabolic
PROS1	Protein S deficiency	Definitive	AD, AR	High	A	Yes	1	Yes	Haematological
PRPS1	Deafness, X-linked 1	Definitive	X-linked	High		No	1	Yes	Hearing and vision loss
PRX	Charcot-Marie-Tooth disease	Strong	AR	High	A	Yes		No	Neuromuscular
PSAP	Metachromatic leukodystrophy	Strong	AR	High	A	Yes		No	Neurological
PSEN1	Cardiomyopathy, dilated, 1U	Disputed	AD	Unknown		No	2	Yes	Cardiac
PSEN2	Cardiomyopathy, dilated, 1V	Limited	AD	Unknown		No	2	Yes	Cardiac
PTCH1	Nevoid basal cell carcinoma syndrome	Definitive	AD	High	A	Yes	2	Yes	Cancer
PTEN	Cowden disease, Bannayan-Riley-Ruvalcaba syndrome	Definitive	AD	High	A	Yes	1	Yes	Cancer

PTH1R	Metaphyseal chondrodysplasia	Strong	AD	High	A	Yes		No	Skeletal
PTPN11	Noonan syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
PTPRC	Severe Combined Immunodeficiency, T cell-negative, B-cell / Natural Killer-Cell Positive	Definitive	AR	High		No	1	Yes	Immunological
PTPRQ	Deafness, autosomal recessive 84A	Definitive	AR	High		No	1	Yes	Hearing and vision loss
PTS	Hyperphenylalaninemia, BH4-deficient, A	Definitive	AR	High	A	Yes	1	Yes	Metabolic
PYGL	Glycogen storage disease VI	Strong	AR	High	A	Yes	1	Yes	Metabolic
PYGM	McArdle Disease	Strong	AR	High		No	1	Yes	Skeletal
QDPR	Dihydropteridine reductase deficiency	Strong	AR	High	A	Yes	1	Yes	Metabolic
RAB23	Carpenter syndrome	Definitive	AR	High	A	Yes	1	Yes	Skeletal
RAB27A	Griscelli syndrome	Definitive	AR	High	A	Yes		No	Dermatological
RAB3GAP1	Warburg micro syndrome	Strong	AR	High	A	Yes		No	Developmental
RAB7A	Charcot-Marie-Tooth disease	Strong	AD	High	A	Yes		No	Neuromuscular
RAD51	Fanconi anemia, complementation group R	Unknown	AD	High		No	2	Yes	Multisystem
RAD51C	Fanconi anemia, complementation group O	Unknown	AR	High		No	2	Yes	Multisystem
RAF1	Noonan syndrome	Definitive	AD	High	A	Yes		No	Multisystem
RAG1	Omenn syndrome	Definitive	AR	High	A	Yes	1	Yes	Immunological
RAG2	Omenn syndrome	Definitive	AR	High	A	Yes	1	Yes	Immunological
RAPSN	Congenital myasthenic syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
RASA1	Capillary malformation-arteriovenous malformation	Moderate	AD	High	A	Yes	1	Yes	Haematological
RASGRP2	Bleeding disorder, platelet-type, 18	Definitive	AR	High		No	2	Yes	Haematological
RB1	Retinoblastoma	Definitive	AD	High	A	Yes	1	Yes	Cancer
RBM20	Cardiomyopathy, dilated, 1DD	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
RBM8A	Thrombocytopaenia-absent radius syndrome	Definitive	AR	High	A	Yes		No	Haematological
RDX	Deafness, autosomal recessive 24	Definitive	AR	High	C	Yes	1	Yes	Hearing and vision loss
REN	Renal tubular dysgenesis and juvenile hyperuricemic nephropathy	Definitive	AR	High	A	Yes		No	Renal
RET	Multiple endocrine neoplasia IIA and IIB	Definitive	AD	High	A	Yes	1	Yes	Cancer
RIT1	Noonan syndrome 8	Definitive	AD	High		No	1	Yes	Multisystem
RMRP	Cartilage-hair hypoplasia	Definitive	AR	High	A	Yes		No	Skeletal

RNASEH2A	Aicardi-Goutieres syndrome	Strong	AR	High	A	Yes		No	Neurological
RNASEH2B	Aicardi-Goutieres syndrome	Strong	AR	High	A	Yes		No	Neurological
RNASEH2C	Aicardi-Goutieres syndrome	Strong	AR	High	A	Yes		No	Neurological
RPGR	Retinitis pigmentosa	Definitive	XLR	High	A	Yes		No	Hearing and vision loss
RPL11	Diamond-Blackfan anemia	Strong	AD	High	A	Yes	1	Yes	Haematological
RPL15	Diamond-Blackfan anemia 12	Unknown	AD	High		No	2	Yes	Haematological
RPL26	Diamond-Blackfan anemia 11	Unknown	AD	High		No	2	Yes	Haematological
RPL27	Diamond-Blackfan anemia 16	Unknown	AD	High		No	2	Yes	Haematological
RPL31	Diamond-Blackfan anemia	Unknown	AD	High		No	2	Yes	Haematological
RPL35A	Diamond-Blackfan anemia 5	Strong	AD	High	C	Yes	1	Yes	Haematological
RPL5	Diamond-Blackfan anemia	Definitive	AD	High	A	Yes	1	Yes	Haematological
RPS10	Diamond-Blackfan anemia 9	Definitive	AD	High	C	Yes	1	Yes	Haematological
RPS15	Diamond-Blackfan anemia	Strong	AD	High	A	Yes		No	Haematological
RPS17	Diamond-Blackfan anemia	Strong	AD	High	A	Yes	1	Yes	Haematological
RPS19	Diamond-Blackfan anemia	Definitive	AD	High	A	Yes	1	Yes	Haematological
RPS24	Diamond-Blackfan anemia	Definitive	AD	High	A	Yes	1	Yes	Haematological
RPS26	Diamond-Blackfan anemia	Strong	AD	High	A	Yes	1	Yes	Haematological
RPS27	Diamond-Blackfan anemia 17	Unknown	AD	High		No	2	Yes	Haematological
RPS28	Diamond Blackfan anemia 15 with mandibulofacial dysostosis	Unknown	AD	High		No	2	Yes	Haematological
RPS29	Diamond-Blackfan anemia 13	Limited	AD	High		No	1	Yes	Haematological
RPS6KA3	Coffin-Lowry syndrome	Definitive	XLD	High	A	Yes		No	Skeletal
RPS7	Diamond-Blackfan anemia 8	Unknown	AD	High	C	Yes	2	Yes	Haematological
RRM2B	Mitochondrial DNA depletion syndrome	Strong	AR	High	A	Yes		No	Multisystem
RS1	Retinoschisis, X linked	Definitive	XLD	High	A	Yes		No	Hearing and vision loss
RSPH1	Ciliary dyskinesia, primary, 24	Definitive	AR	High		No	1	Yes	Pulmonary
RSPH3	Ciliary dyskinesia, primary, 32	Limited	AR	High		No	1	Yes	Pulmonary
RSPH4A	Ciliary dyskinesia, primary	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
RSPH9	Ciliary dyskinesia, primary	Definitive	AR	High	A	Yes	1	Yes	Pulmonary
RUNX1	Platelet disorder, familial, with associated myeloid malignancy	Definitive	AD	Moderate		No	2	Yes	Haematological

RUNX2	Cleidocranial dysostosis	Definitive	AD	High	A	Yes		No	Skeletal
RYR1	Central core disease, Centronuclear myopathy, Malignant hyperthermia, Multiminicore disease	Definitive	AD, AR	High	A, B	Yes	1	Yes	Neuromuscular
RYR2	Catecholaminergic polymorphic ventricular tachycardia, hypertrophic cardioomyopathy	Definitive	AD	High	A	Yes	1	Yes	Cardiac
S1PR2	Deafness, autosomal recessive 68	Strong	AR	High		No	1	Yes	Hearing and vision loss
SACS	Spastic ataxia Charlevoix-Saguenay type	Strong	AR	High	A	Yes		No	Neuromuscular
SALL1	Townes-Brocks syndrome	Definitive	AD	High	A	Yes		No	Multisystem
SAMHD1	Aicardi-Goutieres syndrome	Strong	AR	High	A	Yes		No	Neurological
SBDS	Shwachman-Bodian-Diamond syndrome	Definitive	AR	High	A	Yes		No	Multisystem
SCN11A	Episodic pain syndrome	Definitive	AD	High	A	Yes		No	Neurological
SCN1A	Dravet syndrome, generalised epilepsy with febrile seizures, developmental and epileptic encephalopathy, familial hemiplegic migraine	Definitive, Strong, Moderate	AD	High	A	Yes		No	Neurological
SCN1B	Developmental and epileptic encephalopathy, generalised epilepsy with febrile seizures	Definitive	AD	High	C	Yes	1	Yes	Neurological
SCN3B	Brugada syndrome 1	Disputed	AD	Unknown	C	Yes	2	Yes	Cardiac
SCN4A	Hyperkalemic periodic paralysis, type 2	Definitive	AR	High	A	Yes	1	Yes	Neuromuscular
SCN4B	Long QT syndrome-10	Disputed	AD	High	C	Yes	1	Yes	Cardiac
SCN5A	Brugada syndrome, dilated cardiomyopahy,familial long QT syndrome, Arrhthmogenic right ventricular cardiomyopathy	Definitive, Limited	AD	Moderate	B	Yes	1	Yes	Cardiac
SCNN1A	Pseudohypoaldosteronism type 1	Strong	AR	High	A	Yes	1	Yes	Metabolic
SCNN1B	Pseudohypoaldosteronism	Definitive	AR	High	A	Yes	1	Yes	Metabolic
SCNN1G	Liddle Syndrome	Definitive	AD	High	C	Yes	1	Yes	Renal
SCO2	Leigh syndrome	Definitive	AR	High	A	Yes		No	Neurological
SDHA	Paragangliomas 5	Definitive	AD	High		No	2	Yes	Neurological
SDHAF2	Hereditary Paraganglioma-Pheochromocytoma Syndromes	Definitive	AD	Moderate	B	Yes	2	Yes	Cancer
SDHB	Hereditary Paraganglioma-Pheochromocytoma Syndromes	Definitive	AD	Moderate	B	Yes	1	Yes	Cancer
SDHC	Hereditary Paraganglioma-Pheochromocytoma Syndromes	Definitive	AD	Moderate	B	Yes	1	Yes	Cancer
SDHD	Hereditary Paraganglioma-Pheochromocytoma Syndromes	Definitive	AD	High	A	Yes	1	Yes	Cancer
SEPTIN9	Amyotrophy, hereditary neuralgic	Moderate	AD	High	A	Yes		No	Neurological
SERPINC1	Thrombophilia due to antithrombin III deficiency	Definitive	AD, AR	High	C	Yes	1	Yes	Haematological

SETBP1	Schinzel-Giedion syndrome	Definitive	AD	High	A	Yes		No	Multisystem
SETX	Ataxia-ocular apraxia 2	Definitive	AD	High	A	Yes		No	Neurological
SFTPB	Surfactant metabolism dysfunction, pulmonary	Definitive	AR	High	A	Yes	2	Yes	Pulmonary
SFTPC	Surfactant metabolism dysfunction, pulmonary, 2	Moderate	AD	High	C	Yes	2	Yes	Pulmonary
SGCA	Muscular dystrophy, limb-girdle, type 2D	Definitive	AR	High	A	Yes		No	Neuromuscular
SGCB	Muscular dystrophy, limb-girdle, type 2E	Definitive	AR	High	A	Yes		No	Neuromuscular
SGCD	Muscular dystrophy, limb-girdle, type 2F	Definitive	AR	High	A	Yes	2	Yes	Neuromuscular
SGCG	Muscular dystrophy, limb-girdle, type 2C	Definitive	AR	High	A	Yes		No	Neuromuscular
SGSH	Mucopolysaccharidisis type IIIA (Sanfilippo A)	Definitive	AR	High	A	Yes	2	Yes	Metabolic
SH2D1A	Lymphoproliferative syndrome	Definitive	XLR	High	A	Yes		No	Metabolic
SH3TC2	Charcot-Marie-Tooth disease	Strong	AR	High	A	Yes		No	Neuromuscular
SHANK3	Phelan-McDermid syndrome	Strong	AD	High	A	Yes		No	Developmental
SHH	Holoprosencephaly-3	Definitive	AD	High	A	Yes		No	Neurological
SIL1	Marinesco-Sjogren syndrome	Strong	AR	High	A	Yes		No	Multisystem
SIX1	Branchiootorenal syndrome	Definitive	AD	High	A	Yes	1	Yes	Multisystem
SIX3	Holoprosencephaly-2	Definitive	AD	High	A	Yes		No	Neurological
SKI	Shprintzen-Goldberg syndrome	Definitive	AD	High	A	Yes		No	Connective tissue
SKIC3	Trichohepatoenteric syndrome	Strong	AR	High	A	Yes		No	Multisystem
SLC12A1	Bartter syndrome	Definitive	AR	High	A	Yes		No	Renal
SLC12A3	Gitelman syndrome	Definitive	AR	High	A	Yes		No	Renal
SLC12A6	Agenesis of the corpus callosum with peripheral neuropathy	Strong	AR	High	A	Yes		No	Neurological
SLC16A2	Allan-Herndon-Dudley syndrome	Strong	XLD	High	A	Yes		No	Neurological
SLC17A5	Sialic acid storage disorder, infantile	Strong	AR	High	A	Yes	2	Yes	Neurological
SLC17A8	Deafness, autosomal dominant 25	Strong	AD	High		No	2	Yes	Hearing and vision loss
SLC19A2	Thiamine-responsive megaloblastic anemia syndrome	Strong	AR	High	A	Yes	1	Yes	Haematological
SLC19A3	Basal ganglia disease, biotin-responsive	Strong	AR	High	A	Yes	1	Yes	Neurological
SLC22A4	Deafness, autosomal recessive	Limited	AR	High		No	2	Yes	Hearing and vision loss
SLC22A5	Carnitine deficiency, systemic primary	Definitive	AR	High	A	Yes	1	Yes	Metabolic

SLC25A13	Citrullinemia	Definitive	AR	High	A	Yes	1	Yes	Metabolic
SLC25A15	Hyperornithinemia-hyperammonemia-homocitrullinemia syndrome	Definitive	AR	High	A	Yes	1	Yes	Metabolic
SLC25A20	Carnitine-acylcarnitine translocase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
SLC25A38	Anemia, sideroblastic, pyridoxine-refractory, autosomal recessive	Definitive	AR	High	A	Yes		No	Haematological
SLC25A4	Progressive external ophthalmoplegia	Definitive	AD	High	A	Yes		No	Hearing and vision loss
SLC26A2	Achondrogenesis 1B, multiple epiphyseal dysplasia, diastrophic dysplasia, atelosteogenesis type 2	Definitive	AR	High	A	Yes		No	Skeletal
SLC26A3	Chloride diarrhea, congenital, Finnish type	Definitive	AR	High	A	Yes		No	Gastrointestinal
SLC26A4	Pendred syndrome	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
SLC27A4	Ichthyosis prematurity syndrome	Strong	AR	High	A	Yes		No	Dermatological
SLC2A1	GLUT1 deficiency syndrome 1	Definitive	AD	High	A	Yes	1	Yes	Multisystem
SLC2A10	Arterial tortuosity syndrome	Definitive	AR	High	A	Yes		No	Haematological
SLC2A9	Hypouricemia, renal, 2	Moderate	AR	Moderate		No	1	Yes	Renal
SLC34A2	Pulmonary alveolar microlithiasis	Strong	AR	High	A	Yes		No	Pulmonary
SLC34A3	Hypophosphatemic rickets with hypercalciuria	Definitive	AR	High	A	Yes		No	Skeletal
SLC35D1	Schneckenbecken dysplasia	Strong	AR	High	A	Yes		No	Skeletal
SLC37A4	Glycogen storage disease Ib	Definitive	AR	High	A	Yes	1	Yes	Multisystem
SLC39A4	Acrodermatitis enteropathica	Definitive	AR	High	A	Yes	1	Yes	Dermatological
SLC3A1	Cystinuria	Definitive	AR	High	A	Yes	2	Yes	Renal
SLC45A2	Oculocutaneous albinism, type IV	Definitive	AR	High	A	Yes		No	Dermatological
SLC46A1	Folate malabsorption, hereditary	Strong	AR	High	A	Yes	1	Yes	Metabolic
SLC4A1	Spherocytosis	Strong	AD	High	A	Yes		No	Haematological
SLC4A11	Corneal endothelial dystrophy	Definitive	AR	High	A	Yes		No	Hearing and vision loss
SLC5A2	Renal glucosuria	Definitive	AR	High	A	Yes		No	Renal
SLC5A5	Thyroid dyshormonogenesis 1	Strong	AR	High	A	Yes	1	Yes	Metabolic
SLC6A5	Hyperekplexia 3	Strong	AR	High	A	Yes		No	Neurological
SLC6A8	Creatine deficiency syndrome, X-linked	Definitive	XLR	High	A	Yes	2	Yes	Metabolism
SLC7A7	Lysinuric protein intolerance	Definitive	AR	High	A	Yes	1	Yes	Multisystem
SLC7A9	Cystinuria	Definitive	AR	High	A	Yes	2	Yes	Renal

SLC9A6	Christianson syndrome	Definitive	XLD	High	A	Yes		No	Neurological
SLCO2A1	Hypertrophic osteoarthropathy, primary, autosomal recessive 2	Strong	AR	High	A	Yes		No	Skeletal
SLITRK6	Deafness and myopia	Definitive	AR	High		No	1	Yes	Hearing and vision loss
SLX4	Fanconi anemia, complementation group P	Limited	AR	High		No	2	Yes	Multisystem
SMAD3	Loeys-Dietz syndrome	Definitive	AD	High	A	Yes	1	Yes	Connective tissue
SMAD4	Juvenile polyposis syndrome	Definitive	AD	High	A	Yes	1	Yes	Cancer
SMAD9	Pulmonary hypertension, primary, 2	Definitive	AD	Unknown	C	Yes	2	Yes	Pulmonary
SMARCAL1	Schimke immunoosseous dysplasia	Definitive	AR	High	A	Yes		No	Multisystem
SMC1A	Cornelia de Lange syndrome	Definitive	AD	High	A	Yes		No	Developmental
SMN1	Spinal muscular atrophy	Definitive	AR	High	A	Yes		No	Neuromuscular
SMPD1	Niemann-Pick disease, type A and B	Definitive	AR	High	A	Yes	2	Yes	Metabolic
SMPX	Deafness, X-linked	Definitive	XLR	High	A	Yes	1, 2	Yes	Hearing and vision loss
SNAI2	Waardenburg Syndrome, type 2D	Limited	AR	High		No	1	Yes	Multisystem
SNTA1	Long QT syndrome	Moderate	AD	High	B	Yes	2	Yes	Cardiac
SOX10	Shah-Waardenburg syndrome	Definitive	AD	High	A	Yes	1, 2	Yes	Multisystem
SOX9	Campomelic dysplasia	Definitive	AD	High	A	Yes		No	Skeletal
SP110	Hepatic venoocclusive disease with immunodeficiency	Strong	AR	High	A	Yes		No	Hepatic
SPAG1	Ciliary dyskinesia, primary, 28	Definitive	AR	High		No	1	Yes	Pulmonary
SPINK5	Netherton syndrome	Strong	AR	High	A	Yes		No	Dermatological
SPR	Sepiapterin reductase deficiency (dopa responsive dystonia)	Definitive	AR	High	A	Yes	1	Yes	Neurological
SPRED1	Legius syndrome	Definitive	AD	High	A	Yes		No	Dermatological
SPTA1	Elliptocytosis	Strong	AD	High	A	Yes		No	Haematological
SPTB	Spherocytosis	Definitive	AD	High	A	Yes		No	Haematological
SPTLC1	Neuropathy, hereditary sensory and autonomic, type IA	Strong	AD	High	A	Yes		No	Neurological
SRCAP	Floating-Harbor syndrome	Definitive	AD	High	A	Yes		No	Multisystem
SRY	46XY sex reversal 1	Strong	YL	Moderate		No	1	Yes	Sex development
STAC3	STAC3 disorder (Bailey-Bloch congenital myopathy)	Definitive	AR	High		No		No	Neuromuscular
STAR	Congenital lipoid adrenal hyperplasia	Strong	AR	High	A	Yes	1	Yes	Metabolic

STAT3	Hyper-IgE recurrent infection syndrome	Definitive	AD	High	A	Yes	1	Yes	Immunological
STK11	Peutz-Jeghers syndrome	Definitive	AD	High	A	Yes	1	Yes	Gastrointestinal
STRA6	Microphthalmia, syndromic	Strong	AR	High	A	Yes		No	Developmental
STRC	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
STS	Ichthyosis, X-linked	Definitive	XLR	High	A	Yes		No	Dermatological
STX11	Hemophagocytic lymphohistiocytosis, familial, 4	Definitive	AR	High	A	Yes		No	Immunological
STXBP1	Epileptic encephalopathy, early infantile	Definitive	AD	High	A	Yes		No	Neurological
STXBP2	Hemophagocytic lymphohistiocytosis	Definitive	AR	High	A	Yes		No	Immunological
SUCLA2	Mitochondrial DNA depletion syndrome 5 (encephalomyopathic with methylmalonic aciduria)	Definitive	AR	High	A	Yes	2	Yes	Multisystem
SUCLG1	Mitochondrial DNA depletion syndrome 9 (encephalomyopathic type with methylmalonic aciduria)	Moderate	AR	High	A	Yes	2	Yes	Multisystem
SUMF1	Multiple Sulfatase Deficiency	Definitive	AR	High		No	2	Yes	Metabolic
SUOX	Sulphite oxidase deficiency	Definitive	AR	High	A	Yes		No	Metabolic
SURF1	Leigh syndrome, due to COX deficiency	Definitive	AR	High	A	Yes		No	Neurological
SYNE4	Deafness, autosomal recessive 76	Moderate	AR	High	C	Yes	2	Yes	Hearing and vision loss
TAT	Tyrosinemia, type II	Definitive	AR	High	A	Yes	1	Yes	Metabolic
TAZ	Barth syndrome	Definitive	XLR	High	A	Yes	1	Yes	Multisystem
TBC1D24	Deafness, onychodystrophy, osteodystrophy, mental retardation, and seizures syndrome	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
TBX1	DiGeorge syndrome	Strong	AD	High	A	Yes		No	Multisystem
TBX19	Adrenocorticotropic hormone deficiency	Strong	AR	High		No	1	Yes	Metabolic
TBX5	Holt-Oram syndrome	Definitive	AD	High	A	Yes		No	Multisystem
TCAP	Muscular dystrophy, limb-girdle, type 2G	Strong	AR	High	A	Yes	1	Yes	Neuromuscular
TCIRG1	Osteopetrosis, infantile malignant	Definitive	AR	High	A	Yes		No	Skeletal
TCN2	Transcobalamin II deficiency	Moderate	AR	Moderate		No	1	Yes	Metabolic
TCOF1	Treacher Collins syndrome 1	Definitive	AD	High	A	Yes	1	Yes	Developmental
TECTA	Deafness	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
TERC	Dyskeratosis congenita	Strong	AD	Moderate	B	Yes	2	Yes	Multisystem
TERT	Dyskeratosis congenita	Definitive	AD	Moderate	B	Yes	4	Yes	Multisystem

TFAP2A	Branchiooculofacial syndrome	Strong	AD	High	A	Yes		No	Multisystem
TFAP2B	Char syndrome	Strong	AD	High	A	Yes		No	Developmental
TFG	Hereditary motor and sensory neuropathy	Strong	AD	High	A	Yes		No	Neurological
TFR2	Hemochromatosis, type 3	Moderate	AR	High	C	Yes	1	Yes	Haematological
TG	Thyroid dyshormonogenesis 3	Definitive	AR	High	A	Yes	1	Yes	Metabolic
TGFB2	Loeys-Dietz Syndrome 4	Definitive	AD	High		No	1	Yes	Connective tissue
TGFB3	Arrhythmogenic right ventricular cardiomyopathy 1, Loeys-Dietz Syndrome 5	Limited	AD	Moderate	C	Yes	2, 1	Yes	Cardiac, Connective tissue
TGFBR1	Loeys-Dietz syndrome	Definitive	AD	High	A	Yes	1	Yes	Connective tissue
TGFBR2	Loeys-Dietz syndrome	Definitive	AD	High	A	Yes	1	Yes	Connective tissue
TGM1	Ichthyosis, congenital, autosomal recessive	Strong	AR	High	A	Yes		No	Dermatological
TGM5	Peeling skin syndrome, acral type	Strong	AR	High	A	Yes		No	Dermatological
TH	Tyrosine hydroxylase deficiency	Definitive	AR	High	A	Yes	1	Yes	Metabolic
THRA	Hypothyroidism, congenital, nongoitrous, 6	Strong	AD	High	A	Yes	1	Yes	Metabolic
THRB	Thyroid hormone resistance	Definitive	AD	High	A	Yes		No	Metabolic
TIMM8A	Mohr-Tranebjaerg syndrome	Definitive	XLR	High	A	Yes	2	Yes	Multisystem
TINF2	Dyskeratosis congenita	Strong	AD	Moderate	B	Yes	2	Yes	Multisystem
TK2	Mitochondrial DNA depletion syndrome	Strong	AR	High	A	Yes		No	Neuromuscular
TMC1	Deafness	Strong	AR	High	A	Yes	1, 2	Yes	Hearing and vision loss
TMEM127	Hereditary pheochromocytoma-paraganglioma	Definitive	AD	High		No	2	Yes	Cancer
TMEM43	Arrhythmogenic right ventricular dysplasia 5	Definitive	AD	High	A	Yes	1	Yes	Cardiac
TMIE	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
TMPO	Cardiomyopathy, dilated, 1T	Refuted	AD	Unknown	C	Yes	2	Yes	Cardiac
TMPRSS3	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
TNFRSF11B	Paget disease	Strong	AR	High	A	Yes		No	Skeletal
TNFSF11	Osteopetrosis, autosomal recessive 2	Strong	AR	High	A	Yes		No	Skeletal

TNNC1	Cardiomyopathy, dilated and hypertrophic	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
TNNI2	Distal arthrogryposis syndrome 2b	Definitive	AD	High	A	Yes		No	Neuromuscular
TNNI3	Cardiomyopathy, dilated, Familial hypertrophic cardiomyopathy	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
TNNT1	Nemaline myopathy, Amish type	Strong	AR	High	A	Yes		No	Neuromuscular
TNNT2	Cardiomyopathy, dilated, Familial hypertrophic cardiomyopathy	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
TNNT3	Nemaline myopathy	Definitive	AD	High	A	Yes		No	Neuromuscular
TP53	Li-Fraumeni syndrome	Definitive	AD	High	A	Yes	1	Yes	Cancer
TPM1	Cardiomyopathy, hypertrophic	Definitive	AD	Moderate	B	Yes	1	Yes	Cardiac
TPO	Thyroid dyshormonogenesis 2A	Strong	AR	High	A	Yes	1	Yes	Metabolic
TPP1	Neuronal ceroid lipofuscinosis	Definitive	AR	High	A	Yes	2	Yes	Neurological
TPRN	Deafness, autosomal recessive 79	Definitive	AR	High	C	Yes	1	Yes	Hearing and vision loss
TRAPPC2	Spondyloepiphyseal dysplasia tarda	Definitive	XLR	High	A	Yes		No	Skeletal
TRDN	Ventricular tachycardia, catecholaminergic polymorphic, 2 and Long QT syndrome	Definitive	AR	Moderate	C	Yes	1	Yes	Cardiac
TREX1	Aicardi-Goutieres syndrome 1	Strong	AR	High	A	Yes		No	Neurological
TRHR	Thyrotropin-releasing hormone resistance, generalized	Unknown	AR	High	C	Yes	2	Yes	Metabolic
TRIM32	Muscular dystrophy, limb-girdle, type 2H	Definitive	AR	High	A	Yes		No	Neuromuscular
TRIM37	Mulibrey nanism syndrome	Definitive	AR	High	A	Yes		No	Growth
TRIOBP	Deafness, autosomal recessive	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
TRMU	Liver failure, transient infantile	Moderate	AR	High	A	Yes		No	Hepatic
TRPM4	Cardiac conduction disease, Brugada syndrome	Limited	AD	High	A	Yes		No	Cardiac
TSC1	Tuberous sclerosis 1	Definitive	AD	High	A	Yes	1	Yes	Multisystem
TSC2	Tuberous sclerosis 2	Definitive	AD	High	A	Yes	1	Yes	Multisystem
TSEN54	Pontocerebellar hypoplasia type 4	Strong	AR	High	A	Yes		No	Neurological
TSHB	Hypothryoidism, congenital, nongoitrous 4	Strong	AR	High	A	Yes	1	Yes	Metabolic
TSHR	Hypothyroidism	Definitive	AR	High	A	Yes	1	Yes	Metabolic
TSPEAR	Deafness, autosomal recessive 98	Disputed	AR	High	C	Yes	2	Yes	Hearing and vision loss
TSR2	Diamond-Blackfan anemia 14 with mandibulofacial dysostosis	Unknown	X-linked	High		No	2	Yes	Haematological
TTC7A	Immunodeficiency, combined, with intestinal atresias	Definitive	AR	High	A	Yes		No	Gastrointestinal

TTN	Centronuclear myopathy, cardiomyopathy dilated, tibial muscular dystrophy, myofibrillar myopathy	Definitive, Moderate	AR, AD	High, Moderate	A, B	Yes	1	Yes	Neuromuscular
TTPA	Ataxia with isolated vitamin E deficiency	Strong	AR	High	A	Yes	1	Yes	Neurological
TTR	Amyloidosis, hereditary, transthyretin-related	Definitive	AD	High	A	Yes		No	Multisystem
TWIST1	Saethre-Chotzen syndrome	Definitive	AD	High	A	Yes		No	Skeletal
TWNK	Perrault syndrome 5	Unknown	AR	High		No	2	Yes	Neurological
TYMP	Mitochondrial DNA depletion syndrome	Definitive	AR	High	A	Yes		No	Neuromuscular
TYR	Albinism, oculocutaneous 1	Definitive	AR	High	A	Yes		No	Dermatological
TYRP1	Oculocutaneous albinism type 3	Definitive	AR	High		No		No	Dermatological
UBE2T	Fanconi anemia, complementation group T	Unknown	AR	High		No	2	Yes	Multisystem
UBR1	Johanson-Blizzard syndrome	Strong	AR	High	A	Yes		No	Multisystem
UGT1A1	Crigler-Najjar syndrome	Definitive	AR	High	A	Yes		No	Hepatic
UMOD	Nephropathy	Definitive	AD	High	A	Yes		No	Renal
UNC13D	Hemophagocytic lymphohistiocytosis, familial, 3	Strong	AR	High	A	Yes	1	Yes	Immunological
UROD	Porphyria, hepatoerythropoietic	Definitive	AR	High	A	Yes		No	Dermatological
UROS	Porphyria, congenital erythropoietic	Strong	AR	High	A	Yes		No	Dermatological
USH1C	Usher syndrome 1	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
USH1G	Usher syndrome 1	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
USH2A	Usher syndrome 2	Definitive	AR	High	A	Yes	1	Yes	Hearing and vision loss
VCAN	Wagner syndrome	Definitive	AD	High	A	Yes		No	Hearing and vision loss
VCL	Cardiomyopathy, dilated	Moderate	AD	Moderate	B	Yes	1	Yes	Cardiac
VCP	Inclusion body myopathy with early-onset paget disease and frontotemporal dementia	Definitive	AD	High	A	Yes		No	Neuromuscular
VDR	Vitamin D-dependent rickets	Definitive	AR	High	A	Yes		No	Skeletal
VHL	von Hippel-Lindau syndrome	Definitive	AD	High	A	Yes	1	Yes	Cancer
VIPAS39	Arthrogryposis, renal dysfunction and cholestasis	Definitive	AR	High	A	Yes		No	Multisystem
VLDLR	Cerebellar hypoplasia and mental retardation with or without quadrupedal locomotion 1	Definitive	AR	High	A	Yes		No	Neurological
VPS13A	Choreoacanthocytosis	Definitive	AR	High	A	Yes		No	Neurological

VPS13B	Cohen syndrome	Definitive	AR	High	A	Yes		No	Multisystem
VPS33B	Arthrogryposis renal dysfunction cholestasis syndrome	Definitive	AR	High	A	Yes		No	Multisystem
VWF	von Willebrand disease	Definitive	AD	Moderate	B	Yes	1	Yes	Haematological
WAS	Wiskott-Aldrich syndrome	Definitive	XLR	High	A	Yes		No	Immunological
WDR62	Microcephaly 2, primary, autosomal recessive, with or without cortical malformations	Definitive	AR	High	A	Yes		No	Neurological
WFS1	Wolfram syndrome	Definitive	AR	High	A	Yes	2	Yes	Multisystem
WHRN	Usher Syndrome, type 2D, Nonsyndromic hearing loss	Definitive	AR	High		No	1	Yes	Hearing and vision loss
WNT10A	Ectodermal dysplasia	Definitive	AR	High	A	Yes		No	Dermatological
WRN	Werner syndrome	Definitive	AR	High	A	Yes		No	Growth
WT1	Wilms tumor, type 1, Denys-Drash syndrome, Frasier syndrome	Definitive	AD	Moderate, High	B, A	Yes	1	Yes	Cancer
XPA	Xeroderma pigmentosum	Definitive	AR	High	A	Yes		No	Dermatological
XPC	Xeroderma pigmentosum	Definitive	AR	High	A	Yes		No	Dermatological
ZAP70	ZAP70-related severe combined immunodeficiency	Definitive	AR	High	A	Yes	1	Yes	Immunological
ZEB2	Mowat-Wilson syndrome	Definitive	AD	High	A	Yes		No	Multisystem
ZIC2	Holoprosencephaly-5	Strong	AD	High	A	Yes		No	Neurological
ZIC3	Heterotaxy	Strong	XLR	High	A	Yes		No	Multisystem
ZMPSTE24	Manibular dysplasia type B with lipodystrophy	Definitive	AR	High	A	Yes		No	Growth
ZMYND10	Ciliary Dyskinesia, Primary, 22	Definitive	AR	High		No	1	Yes	Pulmonary
ZNF469	Brittle cornea syndrome	Strong	AR	High	A	Yes		No	Connective tissue

Appendix E. Table of additional genes included in the DDG2P virtual gene panel

Table E.1. List of 1704 genes included in the DDG2P virtual gene panel that are not in the NICU/African virtual gene panel.

AASS	*ABAT*	*ABCB6*	*ABCB7*	*ABCD4*	*ABHD16A*	*ABHD5*	*ABL1*	*ACAN*	*ACBD5*	*ACBD6*	*ACER3*
ACO2	*ACP5*	*ACSL4*	*ACTL6B*	*ACVR2B*	*ACY1*	*ADAM22*	*ADAMTS18*	*ADAMTS9*	*ADAR*	*ADARB1*	*ADCY5*
ADGRG1	*ADGRG6*	*ADNP*	*ADPRS*	*ADRA2B*	*ADSL*	*AFF2*	*AFF3*	*AFF4*	*AFG3L2*	*AGK*	*AGO1*
AGPS	*AGTPBP1*	*AHDC1*	*AIMP1*	*AIPL1*	*AKT1*	*AKT2*	*AKT3*	*ALAD*	*ALDH1A2*	*ALDH1A3*	*ALDH4A1*
ALDOA	*ALG11*	*ALG13*	*ALG2*	*ALG9*	*ALKBH8*	*ALX1*	*ALX3*	*AMER1*	*AMOTL1*	*AMPD2*	*ANAPC1*
ANKRD11	*ANKRD17*	*ANO1*	*ANTXR1*	*AP1B1*	*AP1G1*	*AP1S2*	*AP2M1*	*AP2S1*	*AP3B2*	*AP4B1*	*AP4E1*
AP4M1	*AP4S1*	*APC2*	*ARCN1*	*ARF1*	*ARFGEF1*	*ARHGAP31*	*ARHGAP35*	*ARHGEF9*	*ARID1A*	*ARID2*	*ARL14EP*
ARL3	*ARL6*	*ARMC9*	*ARNT2*	*ARPC4*	*ARSL*	*ASCC1*	*ASCC3*	*ASH1L*	*ASNS*	*ASPH*	*ASPM*
ASXL1	*ASXL2*	*ASXL3*	*ATAD3A*	*ATG4D*	*ATG7*	*ATIC*	*ATL1*	*ATN1*	*ATOH7*	*ATP13A2*	*ATP1A1*
ATP1A3	*ATP2B1*	*ATP5F1A*	*ATP5F1D*	*ATP6AP2*	*ATP6V0A1*	*ATP6V0C*	*ATP6V1A*	*ATP6V1B2*	*ATP6V1E1*	*ATP9A*	*ATR*
AUTS2	*AXIN1*	*B3GALNT2*	*B3GALT6*	*B3GAT3*	*B4GALT1*	*B4GALT7*	*B9D1*	*BANF1*	*BAP1*	*BAZ2B*	*BCAP31*
BCAS3	*BCL11A*	*BCL11B*	*BCOR*	*BCORL1*	*BFSP2*	*BGN*	*BHLHA9*	*BICRA*	*KIAA1109*	*BMP2*	*BMP4*
BMPER	*BMPR1B*	*BNC2*	*BOLA3*	*BPNT2*	*BPTF*	*BRAT1*	*BRCA1*	*BRD4*	*BRF1*	*BRPF1*	*BRSK2*
BRWD3	*BUB1*	*BUB1B*	*C12orf57*	*C1QBP*	*C2CD3*	*CA5A*	*CA8*	*CACNA1B*	*CACNA1E*	*CACNA1G*	*CACNA1H*
CACNB4	*CAD*	*CAMK2A*	*CAMK2B*	*CAMK2G*	*CAMSAP1*	*CAMTA1*	*CANT1*	*CAPN10*	*CAPRIN1*	*CARS1*	*CARS2*
CASP2	*CC2D1A*	*CCBE1*	*CCDC115*	*CCDC22*	*CCDC32*	*CCDC47*	*CCDC78*	*CCDC8*	*CCDC88A*	*CCDC88C*	*CCNA2*
CCND2	*CCNK*	*CD151*	*CD96*	*CDC40*	*CDC42*	*CDC42BPB*	*CDC45*	*CDC6*	*CDH11*	*CDH15*	*CDH2*
CDH3	*CDK10*	*CDK13*	*CDK16*	*CDK19*	*CDK5RAP2*	*CDK8*	*CDON*	*CDT1*	*CELF2*	*CENPJ*	*CEP104*
CEP135	*CEP41*	*CEP57*	*CEP63*	*CEP83*	*CEP85L*	*CERT1*	*CFAP300*	*CFAP410*	*C8orf37*	*CHAMP1*	*CHD1*
CHD3	*CHD4*	*CHD5*	*CHD8*	*CHKA*	*CHMP1A*	*CHRDL1*	*CHRM1*	*CHRNA2*	*CHRNA3*	*CHRNA4*	*CHRNB1*
CHRNB2	*CHST14*	*CHST3*	*CHSY1*	*CHUK*	*CIC*	*CIT*	*CKAP2L*	*CLCN3*	*CLCN4*	*CLCN6*	*CLCN7*
CLCNKA	*CLDN5*	*CLIC2*	*CLMP*	*CLP1*	*CLPB*	*CLTC*	*CNKSR1*	*CNKSR2*	*CNNM2*	*CNOT1*	*CNOT3*
CNPY3	*CNTNAP1*	*CNTNAP2*	*COA5*	*COA8*	*COASY*	*COG1*	*COG4*	*COG5*	*COG7*	*COG8*	*COL10A1*
COL13A1	*COL18A1*	*COL25A1*	*COL27A1*	*COL4A1*	*COL4A2*	*COL9A2*	*COL9A3*	*COLEC10*	*COLEC11*	*COMP*	*COPB1*
COPB2	*COQ2*	*COQ4*	*COQ5*	*COQ9*	*COX10*	*COX14*	*COX15*	*COX16*	*COX6B1*	*COX7B*	*CPAMD8*
CPLANE1	*CPSF3*	*CRADD*	*CRB1*	*CRB2*	*CRBN*	*CRELD1*	*CRIM1*	*CRIPT*	*CRKL*	*CRX*	*CRYAA*

CRYBA1	*CRYBA4*	*CRYBB1*	*CRYBB2*	*CRYBB3*	*CRYGC*	*CRYGD*	*CSDE1*	*CSF1R*	*CSNK1G1*	*CSNK2A1*	*CSNK2B*
CSPP1	*CSTA*	*CTBP1*	*CTCF*	*CTDP1*	*CTNNA2*	*CTNNB1*	*CTNND1*	*CTNND2*	*CTU2*	*CUL3*	*CUL4B*
CUX1	*CUX2*	*CWC27*	*CYB5R3*	*CYC1*	*CYFIP2*	*CYP1B1*	*CYP24A1*	*CYP2U1*	*DACT1*	*DAG1*	*DARS1*
DARS2	*DAW1*	*DCAF17*	*DCC*	*DCDC2*	*DCHS1*	*DDB1*	*DDHD1*	*DDHD2*	*DDOST*	*DDR2*	*DDX11*
DDX23	*DDX3X*	*DDX54*	*DDX59*	*DDX6*	*DEAF1*	*DEGS1*	*DENND5A*	*DEPDC5*	*DGAT1*	*DHCR24*	*DHDDS*
DHFR	*DHODH*	*DHPS*	*DHRS3*	*DHTKD1*	*DHX16*	*DHX30*	*DHX34*	*DHX37*	*DIP2B*	*DIS3L2*	*DISP1*
DLAT	*DLG3*	*DLG4*	*DLG5*	*DLL1*	*DLL4*	*DLX5*	*DNA2*	*LRRC6*	*DNAAF6*	*DNAH14*	*DNAH9*
DNAJB4	*DNAJC12*	*DNM1*	*DNM1L*	*DNMT3A*	*DOCK6*	*DOCK7*	*DOHH*	*DOLK*	*DPF2*	*DPH5*	*DPM1*
DPM3	*DPYSL5*	*DSE*	*DSG1*	*DSTYK*	*DVL1*	*DVL3*	*DYM*	*DYNC1H1*	*DYNC1I2*	*DYNC2H1*	*DYNC2I1*
DYNC2I2	*DYNC2LI1*	*DYRK1A*	*EBF3*	*EBP*	*ECEL1*	*ECHS1*	*ECM1*	*EDEM3*	*EDN1*	*EDNRA*	*EED*
EEF1A2	*EEF1B2*	*EEF2*	*EFNB1*	*EHMT1*	*EIF2AK1*	*EIF2AK2*	*EIF2B4*	*EIF2B5*	*EIF2S3*	*EIF3F*	*EIF4A3*
EIF5A	*ELAC2*	*ELFN1*	*ELMO2*	*ELOVL4*	*ELP2*	*EMC1*	*EMC10*	*EMG1*	*EMX2*	*ENTPD1*	*EOGT*
EOMES	*EP300*	*EPB41L1*	*EPCAM*	*EPG5*	*EPHB4*	*EPRS1*	*ERBB3*	*ERCC1*	*ERCC3*	*ERCC6L2*	*ERF*
ERLIN2	*ERMARD*	*EXOSC2*	*EXOSC3*	*EXOSC9*	*EXPH5*	*EXTL3*	*FAM111A*	*FAM149B1*	*FAM20A*	*FAR1*	*FARS2*
FASN	*FAT4*	*FBLN1*	*FBXL4*	*FBXO11*	*FBXO28*	*FBXW11*	*FBXW4*	*FBXW7*	*FCSK*	*FDFT1*	*FEZF1*
FGF10	*FGF12*	*FGF13*	*FGF14*	*FGF9*	*FGFR2*	*FIG4*	*FKBP14*	*FKRP*	*FLAD1*	*FLG*	*FLNB*
FLT4	*FLVCR1*	*FLVCR2*	*FMN2*	*FMR1*	*FN1*	*FOXE3*	*FOXG1*	*FOXI3*	*FOXJ1*	*FOXL2*	*FOXN1*
FOXP1	*FOXP2*	*FOXP4*	*FOXRED1*	*FRA10AC1*	*FREM1*	*FREM2*	*FRMD5*	*FRMD7*	*FRMPD4*	*FRRS1L*	*FRY*
FTCD	*FTO*	*FTSJ1*	*FUT8*	*FXR1*	*FYCO1*	*FZD5*	*FZD6*	*FZR1*	*GABBR1*	*GABBR2*	*GABRA1*
GABRB2	*GABRB3*	*GABRG1*	*GABRG2*	*GAD1*	*GAS2L2*	*GATA6*	*GATAD2B*	*GBA2*	*GCSH*	*GDF1*	*GDF11*
GDF3	*GDF5*	*GDF6*	*GDI1*	*GEMIN4*	*GEMIN5*	*GFER*	*GHR*	*GIGYF1*	*GJA3*	*GJA8*	*GJB3*
GK	*GLDN*	*GLE1*	*GLI2*	*GLIS2*	*GLIS3*	*GLMN*	*GLRB*	*GLUL*	*GMNN*	*GMPPA*	*GMPPB*
GNA11	*GNA14*	*GNAI1*	*GNAI3*	*GNAO1*	*GNAQ*	*GNAS*	*GNB1*	*GNB2*	*GNB3*	*GNB5*	*GNPAT*
GOLGA2	*GON4L*	*GORAB*	*GOT2*	*GPAA1*	*GPC4*	*GPC6*	*GPHN*	*GPX4*	*GREB1L*	*GRHL3*	*GRIA1*
GRIA2	*GRIA3*	*GRIA4*	*GRID2*	*GRIN1*	*GRIN2A*	*GRIN2B*	*GRIN2D*	*GRM1*	*GRM6*	*GRM7*	*GSPT2*
GTF2E2	*GTF2H5*	*GTF2IRD1*	*GTPBP2*	*GTPBP3*	*GUCY2C*	*GZF1*	*H1-4*	*H2AC6*	*H3-3A*	*H3-3B*	*H3-4*
H4C11	*H4C2*	*H4C3*	*HAAO*	*HACD1*	*HACE1*	*HARS1*	*HAX1*	*HCCS*	*HCFC1*	*HCN1*	*HDAC4*
HECW2	*HERC1*	*HERC2*	*HESX1*	*HIBCH*	*HIRA*	*HIVEP2*	*HK1*	*HMGB1*	*HMGB3*	*HMX1*	*HNRNPA2B1*
HNRNPD	*HNRNPH1*	*HNRNPH2*	*HNRNPK*	*HNRNPR*	*HNRNPU*	*HOXA1*	*HOXA11*	*HOXA13*	*HOXB1*	*HOXC13*	*HOXD13*
HPDL	*HPGD*	*HPSE2*	*HR*	*HS2ST1*	*HSF4*	*HSPD1*	*HTRA2*	*HUWE1*	*HYAL2*	*FAM126A*	*HYLS1*

IARS1	*IARS2*	*IFIH1*	*IFITM5*	*IFT122*	*IFT140*	*IFT172*	*IFT43*	*IFT74*	*IFT80*	*IGBP1*	*IGF1*
IGF1R	*IGF2*	*IGFBP7*	*IHH*	*IL11*	*IL11RA*	*IL1RAPL1*	*INPP4A*	*INPP5E*	*INPP5K*	*INPPL1*	*IPO8*
IQSEC1	*IQSEC2*	*IREB2*	*IRF2BPL*	*IRX5*	*ITCH*	*ITGA3*	*ITGA6*	*ITGA7*	*ITGA8*	*ITPR1*	*JAG2*
JAGN1	*JAM3*	*JARID2*	*JMJD1C*	*KANK1*	*KAT5*	*KAT6A*	*KATNB1*	*KCNA2*	*KCNA4*	*KCNB1*	*KCNC1*
KCNC3	*KCNH1*	*KCNH5*	*KCNJ6*	*KCNJ8*	*KCNK4*	*KCNK9*	*KCNMA1*	*KCNN3*	*KCNQ2*	*KCNQ3*	*KCNQ5*
KCNT1	*KCNT2*	*KCTD1*	*KDELR2*	*KDM1A*	*KDM2B*	*KDM3B*	*KDM4B*	*KDM5A*	*KDM5B*	*KDM5C*	*KDM6B*
KIAA0586	*KIDINS220*	*KIF11*	*KIF14*	*KIF1A*	*KIF22*	*KIF2A*	*KIF3B*	*KIF4A*	*KIF5A*	*KIF5C*	*KIF7*
KIFBP	*KIRREL3*	*KITLG*	*KLF1*	*KLF7*	*KLF8*	*KLHL15*	*KLHL7*	*KMT2A*	*KMT2B*	*KMT2C*	*KMT2E*
KMT5B	*KPNA7*	*KPTN*	*KRIT1*	*KRT74*	*L2HGDH*	*LAGE3*	*LAMA1*	*LAMB1*	*LAMC3*	*LARP7*	*LAS1L*
LEFTY2	*LEMD2*	*LEMD3*	*LETM1*	*LFNG*	*LGI4*	*LHX4*	*LIAS*	*LINGO1*	*LINS1*	*LIPN*	*LIPT1*
LIPT2	*LMBRD2*	*LMNB1*	*LMNB2*	*LNPK*	*LONP1*	*LRAT*	*LRBA*	*LRIG2*	*LRIT3*	*LRP5*	*LRP6*
LRPAP1	*LRRC56*	*LTBP1*	*LTBP2*	*LTBP3*	*LZTR1*	*MAB21L1*	*MAB21L2*	*MACF1*	*MADD*	*MAF*	*MAGEL2*
MAMLD1	*MAN1B1*	*MAN2A2*	*MAN2C1*	*MAOA*	*MAP3K1*	*MAP3K7*	*MAPK1*	*MAPK10*	*MAPK8IP3*	*MAPKAPK5*	*MAPRE2*
MASP1	*MAST1*	*MATN3*	*MAU2*	*MBD5*	*MBOAT7*	*MC2R*	*MCEE*	*MDH2*	*MECOM*	*MECR*	*MED11*
MED13	*MED13L*	*MED17*	*MED23*	*MED25*	*MED27*	*MEF2C*	*MEGF8*	*MEIS2*	*MEOX1*	*MESD*	*MESP2*
METTL23	*METTL5*	*MFF*	*MFRP*	*MFSD2A*	*MGAT2*	*MIB1*	*MICU1*	*MID1*	*MIR17HG*	*MIR184*	*MMGT1*
MMP13	*MMP14*	*MMP15*	*MMP21*	*MMUT*	*MN1*	*MNX1*	*MOGS*	*MORC2*	*MPDU1*	*MPDZ*	*MPLKIP*
MRAS	*MRE11*	*MRPS2*	*MRPS22*	*MRPS34*	*MSI1*	*MSL2*	*MSL3*	*MSX1*	*MTF1*	*MTO1*	*MTOR*
MTRFR	*MTSS2*	*MT-TL1*	*MT-TP*	*MYBPC1*	*MYCBP2*	*MYF5*	*MYH10*	*MYH8*	*MYL1*	*MYLPF*	*MYO18B*
MYO5A	*MYO5B*	*MYOC*	*MYOCD*	*MYRF*	*MYSM1*	*MYT1*	*MYT1L*	*NAA15*	*NAA20*	*NACC1*	*NADSYN1*
NAE1	*NALCN*	*NANS*	*NAPB*	*NARS1*	*NAXD*	*NAXE*	*NBAS*	*NBEA*	*NCAPD2*	*NCAPD3*	*NCAPG2*
NCAPH	*NCDN*	*NCKAP1*	*NCOR1*	*NDE1*	*NDNF*	*NDST1*	*NDUFA1*	*NDUFA10*	*NDUFA12*	*NDUFA6*	*NDUFA8*
NDUFA9	*NDUFAF2*	*NDUFAF8*	*NDUFB11*	*NDUFB3*	*NDUFB7*	*NDUFB8*	*NDUFS1*	*NDUFS4*	*NDUFS7*	*NDUFS8*	*NDUFV1*
NDUFV2	*NECTIN1*	*NECTIN4*	*NEDD4L*	*NEK1*	*NEK8*	*NEXMIF*	*NFE2L2*	*NFIA*	*NFIB*	*NFIX*	*NFU1*
NHLRC2	*NHP2*	*NHS*	*NKAP*	*NKX3-2*	*NKX6-2*	*NLGN3*	*NLGN4X*	*NMNAT1*	*NONO*	*NOP10*	*NOTCH1*
NOVA2	*NPHS2*	*NPM1*	*NPR2*	*NPR3*	*NR1I3*	*NR2F1*	*NR2F2*	*NR4A2*	*NRAS*	*NRCAM*	*NRROS*
NRXN1	*NRXN2*	*NRXN3*	*NSD2*	*NSMCE3*	*NSRP1*	*NSUN2*	*NT5C3A*	*NTNG2*	*NTRK2*	*NUBPL*	*NUDT2*
NUP107	*NUP133*	*NUP214*	*NUP54*	*NUP62*	*NUS1*	*NYX*	*OCLN*	*TTC25*	*ODAPH*	*ODC1*	*OGDH*
OGDHL	*OGT*	*ONECUT1*	*OPHN1*	*ORC4*	*ORC6*	*OSGEP*	*OTUD5*	*OTUD6B*	*OTUD7A*	*OTULIN*	*OTX2*
OXCT1	*OXR1*	*P3H1*	*P4HB*	*P4HTM*	*PACS1*	*PACS2*	*PAFAH1B1*	*PAK1*	*PAN2*	*PAPSS2*	*PARN*

PARP1	*PAX1*	*PAX2*	*PAX9*	*PBX1*	*PCARE*	*PCBP2*	*PCDH12*	*PCDH19*	*PCDHGC4*	*PCGF2*	*PCYT1A*
PCYT2	*PDCD10*	*PDE10A*	*PDE6G*	*PDE6H*	*PDGFRB*	*PDIA6*	*PDSS1*	*PDSS2*	*PECR*	*PEPD*	*PET100*
PEX11B	*PEX14*	*PEX16*	*PEX19*	*PEX7*	*PGAP1*	*PGAP2*	*PGAP3*	*PGK1*	*PGM1*	*PGM2L1*	*PGM3*
PHACTR1	*PHC1*	*PHF21A*	*PHF8*	*PHGDH*	*PHIP*	*PI4KA*	*PIBF1*	*PIDD1*	*PIEZO1*	*PIGA*	*PIGB*
PIGG	*PIGH*	*PIGK*	*PIGL*	*PIGM*	*PIGN*	*PIGO*	*PIGQ*	*PIGS*	*PIGT*	*PIGU*	*PIGV*
PIGW	*PIGY*	*PIK3CA*	*PIK3R1*	*PIK3R2*	*PIP5K1C*	*PITX1*	*PITX2*	*PITX3*	*PKD1L1*	*PLAA*	*PLCB1*
PLCB4	*PLCH1*	*PLEC*	*PLK4*	*PLOD2*	*PLOD3*	*PLP1*	*PLPBP*	*PLXNA1*	*PLXND1*	*PMPCB*	*PNPLA1*
PNPLA2	*PNPLA6*	*PNPT1*	*POC1A*	*POC1B*	*POGZ*	*POLA1*	*POLD1*	*POLE*	*POLR1A*	*POLR2A*	*POLR3A*
POLR3B	*POLR3GL*	*POMGNT1*	*POMGNT2*	*POMK*	*POMP*	*POMT1*	*POT1*	*POU3F3*	*POU4F1*	*PPA2*	*PPFIBP1*
PPIL1	*PPM1D*	*PPP1CB*	*PPP1R12A*	*PPP1R13L*	*PPP1R15B*	*PPP1R21*	*PPP2CA*	*PPP2R1A*	*PPP2R5D*	*PPP3CA*	*PRDM12*
PRDM13	*PRDM15*	*PRDM6*	*PRDX3*	*PREPL*	*PRIM1*	*PRKACA*	*PRKACB*	*PRKAR1B*	*PRKD1*	*PRKG2*	*PRMT7*
PRMT9	*PRORP*	*PROSER1*	*PRPF8*	*PRR12*	*PRRT2*	*PRRX1*	*PRSS12*	*PRSS56*	*PRUNE1*	*PSAT1*	*PSMB8*
PSMC1	*PSMC5*	*PSMD12*	*PSPH*	*PTCHD1*	*PTDSS1*	*PTF1A*	*PTH*	*PTHLH*	*PTPN14*	*PTPRF*	*PTRH2*
PUF60	*PURA*	*PUS1*	*PUS3*	*PUS7*	*PXDN*	*PYCR1*	*PYCR2*	*PYROXD1*	*QARS1*	*QKI*	*QRICH1*
RAB11A	*RAB11B*	*RAB14*	*RAB18*	*RAB39B*	*RAB3GAP2*	*RABL6*	*RAC1*	*RAC3*	*RAD21*	*RAD50*	*RAI1*
RALA	*RALGAPA1*	*RALGDS*	*RANBP2*	*RAP1B*	*RARB*	*RARS1*	*RARS2*	*RAX*	*RBBP8*	*RBFOX1*	*RBM10*
RBM28	*RBPJ*	*RECQL4*	*RELN*	*RERE*	*RETREG1*	*RFT1*	*RFX6*	*RGS7*	*RHOBTB2*	*DDX58*	*RIMS2*
RIN2	*RINT1*	*RIPK4*	*RLIM*	*RMI1*	*RMND1*	*RNASET2*	*RNF113A*	*RNF125*	*RNF13*	*RNF135*	*RNF168*
RNPC3	*RNU12*	*RNU4ATAC*	*ROBO3*	*ROBO4*	*ROGDI*	*ROR2*	*RORA*	*RORB*	*RPE65*	*RPGRIP1*	*RPGRIP1L*
RPL10	*RPL13*	*RPS23*	*RRAS*	*RRAS2*	*RRM1*	*RSPO2*	*RSPO4*	*RSPRY1*	*RSRC1*	*RTEL1*	*RTN4IP1*
RTTN	*RUBCN*	*RXYLT1*	*RYR3*	*SALL4*	*SAMD9*	*SAMD9L*	*SARS1*	*SARS2*	*SATB1*	*SATB2*	*SC5D*
SCAF4	*SCAPER*	*SCARF2*	*SCN2A*	*SCN3A*	*SCN8A*	*SCNM1*	*SCO1*	*SCRIB*	*SCUBE3*	*SCYL1*	*SDCCAG8*
SDHAF1	*SEC23A*	*SEC23B*	*SEC24D*	*SEC61A1*	*SECISBP2*	*SELENOI*	*SELENON*	*SEMA3A*	*SEMA6B*	*SEPSECS*	*SERAC1*
SET	*SETD1A*	*SETD1B*	*SETD2*	*SETD5*	*SF3B4*	*SH3BP2*	*SH3PXD2B*	*SHANK1*	*SHANK2*	*SHMT2*	*SHOC2*
SHOX	*SHROOM3*	*SIAH1*	*SIK1*	*SIM1*	*SIN3A*	*SIN3B*	*SIX5*	*SIX6*	*SKIV2L*	*TTC37*	*SLC10A7*
SLC12A5	*SLC13A1*	*SLC13A5*	*SLC1A2*	*SLC1A4*	*SLC24A1*	*SLC24A4*	*SLC25A1*	*SLC25A19*	*SLC25A22*	*SLC25A24*	*SLC25A26*
SLC25A42	*SLC2A2*	*SLC30A7*	*SLC31A1*	*SLC32A1*	*SLC33A1*	*SLC35A1*	*SLC35A2*	*SLC35B2*	*SLC35C1*	*SLC38A3*	*SLC39A13*
SLC39A8	*SLC45A1*	*SLC4A4*	*SLC52A2*	*SLC52A3*	*SLC5A6*	*SLC5A7*	*SLC6A1*	*SLC6A17*	*SLC6A3*	*SLC6A9*	*SLC9A7*
SLC9A9	*SLIRP*	*SMAD2*	*SMAD6*	*SMARCA2*	*SMARCA4*	*SMARCB1*	*SMARCC2*	*SMARCD1*	*SMARCE1*	*SMC3*	*SMCHD1*
SMG8	*SMG9*	*SMO*	*SMOC1*	*SMOC2*	*SMPD4*	*SMS*	*SNAP25*	*SNAP29*	*SNIP1*	*SNORD118*	*SNRPB*

SNRPE	*SNX14*	*SNX3*	*SOBP*	*SON*	*SOS1*	*SOS2*	*SOX2*	*SOX3*	*SOX4*	*SOX6*	*SPARC*
SPATA5	*SPATA5L1*	*SPECC1L*	*SPEG*	*SPEN*	*SPG11*	*SPOP*	*SPRED2*	*SPRY1*	*SPTAN1*	*SPTBN1*	*SPTBN2*
SPTBN4	*SPTLC2*	*SRD5A3*	*SRP54*	*SRPX2*	*SRRM2*	*ST14*	*ST3GAL3*	*ST3GAL5*	*STAG1*	*STAG2*	*STAMBP*
STAT2	*STAT5B*	*STIL*	*STIM1*	*STN1*	*STRADA*	*STT3A*	*STT3B*	*STX1B*	*SUFU*	*SUMO1*	*SUPT16H*
SUZ12	*SYN1*	*SYNCRIP*	*SYNE1*	*SYNGAP1*	*SYP*	*SYT1*	*SYT2*	*SZT2*	*TAB2*	*TAC3*	*TACO1*
TACR3	*TAF1*	*TAF13*	*TAF2*	*TAF4*	*TAF8*	*TANC2*	*TANGO2*	*TAOK1*	*TAPT1*	*TARS1*	*TASP1*
TBC1D20	*TBC1D23*	*TBC1D2B*	*TBCD*	*TBCE*	*TBCK*	*TBL1XR1*	*TBR1*	*TBX15*	*TBX18*	*TBX20*	*TBX22*
TBX3	*TBX4*	*TBXAS1*	*TCF12*	*TCF20*	*TCF4*	*TCF7L2*	*TCTN1*	*TCTN2*	*TCTN3*	*TDRD7*	*TECPR2*
TEK	*TELO2*	*TET3*	*TFE3*	*TFRC*	*TGDS*	*TGFB1*	*TGIF1*	*THAP1*	*THG1L*	*THOC2*	*THOC6*
THUMPD1	*TKFC*	*TKT*	*TLK2*	*TLL1*	*TM4SF20*	*TMCO1*	*TMEM106B*	*TMEM114*	*TMEM126B*	*TMEM135*	*TMEM147*
TMEM163	*TMEM165*	*TMEM199*	*TMEM216*	*TMEM218*	*TMEM222*	*TMEM237*	*TMEM240*	*TMEM251*	*TMEM260*	*TMEM63A*	*TMEM63C*
TMEM67	*TMEM70*	*TMEM94*	*TMPRSS6*	*TMTC3*	*TMX2*	*TNFRSF13B*	*TNPO2*	*TNRC6B*	*TOE1*	*TOGARAM1*	*TOP3A*
TP53RK	*TP63*	*TP73*	*TPM2*	*TPM3*	*TPP2*	*TPRKB*	*TRA2B*	*TRAF7*	*TRAIP*	*TRAPPC10*	*TRAPPC11*
TRAPPC12	*TRAPPC2L*	*TRAPPC4*	*TRAPPC9*	*TRIM8*	*TRIO*	*TRIP11*	*TRIP12*	*TRIP13*	*TRIP4*	*TRIT1*	*TRMT1*
TRMT10A	*TRMT10C*	*TRNT1*	*TRPC5*	*TRPM1*	*TRPM3*	*TRPS1*	*TRPV3*	*TRPV4*	*TRPV6*	*TRRAP*	*TSEN15*
TSEN2	*TSEN34*	*TSHZ1*	*TSPAN7*	*TTC12*	*TTC19*	*TTC5*	*TTC8*	*TTI2*	*TUBA1A*	*TUBA8*	*TUBB*
TUBB2A	*TUBB2B*	*TUBB3*	*TUBB4A*	*TUBG1*	*TUBGCP2*	*TUBGCP4*	*TUBGCP6*	*TUFM*	*TUSC3*	*TWIST2*	*TXNL4A*
U2AF2	*UBA5*	*UBAP2L*	*UBE2A*	*UBE3A*	*UBE3B*	*UBE4A*	*UBR7*	*UBTF*	*UFC1*	*UFM1*	*UFSP2*
UGP2	*UMPS*	*UNC45A*	*UNC45B*	*UNC80*	*UPF1*	*UPF3B*	*UQCRB*	*UQCRFS1*	*UQCRQ*	*UROC1*	*USB1*
USP14	*USP18*	*USP27X*	*USP7*	*USP9X*	*UTP4*	*UVSSA*	*VAC14*	*VAMP2*	*VANGL1*	*VIP*	*VPS4A*
VPS53	*VRK1*	*VSX2*	*WAC*	*WARS1*	*WASF1*	*WASHC5*	*WDFY3*	*WDPCP*	*WDR11*	*WDR19*	*WDR26*
WDR35	*WDR37*	*WDR4*	*WDR45*	*WDR45B*	*WDR5*	*WDR73*	*WDR81*	*WNK3*	*WNT1*	*WNT10B*	*WNT3*
WNT4	*WNT5A*	*WNT7A*	*WRAP53*	*WWOX*	*XPNPEP3*	*XRCC4*	*XYLT1*	*XYLT2*	*YAP1*	*YARS2*	*YRDC*
YWHAG	*YY1*	*ZBTB16*	*ZBTB18*	*ZBTB20*	*ZBTB40*	*ZBTB7A*	*ZC4H2*	*ZCCHC8*	*ZDHHC15*	*ZDHHC9*	*ZEB1*
ZFHX3	*ZFHX4*	*ZFP57*	*ZFPM2*	*ZFYVE19*	*ZFYVE26*	*ZIC1*	*ZMIZ1*	*ZMYM2*	*ZMYM3*	*ZMYM6*	*ZMYND11*
ZMYND8	*ZNF142*	*ZNF148*	*ZNF292*	*ZNF407*	*ZNF462*	*ZNF526*	*ZNF599*	*ZNF711*	*ZNF713*	*ZNF750*	*ZSWIM6*

Appendix F. Age of infant cohort at enrolment

Table F.1. Age of participants in months at enrolment

Study ID	Age at Recruitment (months)
NE001	0.07
NE002	18.09
NE004	5.07
NE005	18.62
NE006	0.59
NE007	0.30
NE009	16.35
NE010	12.70
NE011	12.63
NE012	0.16
NE013	0.30
NE014	0.26
NE015	9.93
NE016	18.62
NE017	3.45
NE018	0.10
NE019	4.31
NE020	5.86
NE021	0.03
NE022	13.82
NE023	0.56
NE025	11.51
NE026	23.72
NE027	3.78
NE028	10.92
NE029	14.44
NE030	0.13
NE031	1.64
NE032	0.46
NE033	2.86
NE034	0.20
NE035	11.15
Average	6.96
Interquartile Range	12.35
Standard Deviation	7.21

Appendix G. Run and sample statistics from the 17 exome sequencing runs

Table G.1. Run statistics from the 17 exome sequencing runs on the Ion GeneStudio S5 sequencer.

Run	Samples	Data yield (GB)	ISP loading (%)	Total number of reads	Final Library (%)	Polyclonality (%)	Low quality (%)	Usable Reads (%)	Mean Read Length (bp)
1	NE001	18.1	91	94038707	90	25	9	68	192
2	NE002, NE004	17.3	90	88987574	86	24	13	66	194
3	NE005, NE006	17.5	92	89727730	87	26	12	65	195
4	NE007, NE009	15.3	88	77867999	80	27	19	59	197
5	NE010, NE011	18.02	93	91303928	90	28	9	65	197
6	NE012, NE013	18.3	94	93063091	90	26	9	66	197
7	NE014, NE016	18.5	93	92683037	92	28	7	66	199
8	NE017, NE018	16.4	90	84450797	97	29	12	63	195
9	NE019, NE021	15.9	88	80329419	82	27	17	60	198
10	NE015, NE022	17.2	94	88688621	90	31	9	63	194
11	NE028, NE029	18.2	95	95247647	90	27	9	67	191
12	NE020, NE023	17.9	95	91844302	91	29	8	65	195
13	NE026, NE027	13.2	83	67829072	76	29	23	54	195
14	NE030, NE031	15.03	88	76411423	83	30	16	58	197
15	NE032, NE033	17.2	93	87354058	91	32	9	62	197
16	NE025, NE034	15.3	87	78143942	85	30	14	60	196
17	NE035	17.6	91	89458702	90	29	9	65	197
Average		16.9	90.9	86319414.6	87.6	28.1	12.0	63.1	195.6

Table G.2. Coverage statistics for the 32 clinical exome sequencing samples.

Study ID	Mapped Reads	On Target (%)	Mean Depth	Uniformity (%)
NE001	49 470 618	96.32	156.6	89.71
NE002	56 805 794	94.77	178.8	82.85
NE004	30 287 064	95	95.61	85.19
NE005	33 597 312	95.4	107.2	87.65
NE006	54 592 171	95.34	173.4	87.74
NE007	39 916 408	96.46	129	89.08
NE009	36 445 562	97.04	119.3	86.04
NE010	47 423 389	96.67	154.3	92.4
NE011	42 049 603	96.16	135.8	91.7
NE012	56 298 860	96.4	181.4	90.35
NE013	34 712 751	96.24	112	90.86
NE014	34 196 839	94.42	108.2	89.49
NE015	39 399 562	95.08	123.1	92.54
NE016	57 314 186	96.16	187.8	88.06
NE017	29 379 983	94.73	93.39	89.94
NE018	53719696	94.94	168.3	89.6
NE019	29 425 600	94.85	93.95	91.77
NE020	44 219 954	95.04	139.7	91.3
NE021	49 523 729	96.07	160.7	91.12
NE022	46 731 780	96.15	148.9	92.2
NE023	45 132 921	94.95	143.1	94.57
NE025	42 275 215	92.8	131.2	94.26
NE026	39 754 450	95.79	126.7	84.83
NE027	26 882 445	95.7	85.88	84.14
NE028	49 048 383	94.82	149.6	88.94
NE029	33 577 813	94.96	140.1	88.3
NE030	45 858 951	95.92	148.1	90.16
NE031	29 371 129	95.54	93.93	88.97
NE032	43 448 426	95.06	139	90.48
NE033	42 205 350	95.41	135.5	90.92
NE034	34 257 746	95.32	109.7	91.98
NE035	37 290 100	95.69	120.1	91.28
Average	41 706 681	95	134	90

Appendix H. Statistical Analysis

Table H.1. Fisher's Exact Test examining the influence of cohort characteristics (covariates) on positive diagnostic yield.

Covariates			p-values
Sex	Male (0), Female (1)		1.0000
Referring Clinician	Geneticist (0), Neonatologist (1)		0.3702
Indication for Testing	Neuromuscular	No (0), Yes (1)	0.1129
	Skeletal	No (0), Yes (1)	0.0423
	Cardiac	No (0), Yes (1)	0.2964
	Renal	No (0), Yes (1)	0.3952
	Multiple Congenital Anomalies	No (0), Yes (1)	0.6401
	Hepatic	No (0), Yes (1)	1.0000
	Developmental	No (0), Yes (1)	0.3952
	Abnormal Growth	No (0), Yes (1)	1.0000
	Dysmorphic	No (0), Yes (1)	1.0000
	Neurological	No (0), Yes (1)	1.0000
	Gastrointestinal	No (0), Yes (1)	1.0000

Table H.2. Logistic regression examining the relationship between positive diagnosis and age at enrolment in months.

Estimate (Log odds ratio)	0.003991
Standard Error	0.010604
T-value	0.376
P-value	0.709

Appendix I. Clinical Description of Participants Without A Causative Variant Identified

Participant NE002

Participant NE002 was a male participant referred for testing by a medical geneticist at 18 months of age. He was referred with a broad forehead, bilateral exotropia, epicanthic folds, low set ears, a deviated nasal septum, pointed chin, laryngomalacia, microcephaly, multiple cardiac defects (a double outlet right ventricle, patent ductus arteriosus, ventricular septal defect, atrial septal defect), a surgically corrected bowel malrotation and bowel obstruction, congenital dermal melanocytosis, two cafe au lait spots, and weight and height below the third centile. This participant had multiple hospital admissions during the first 18 months of life for lower respiratory tract infections, indicative of a potential T-cell deficiency. A microarray was performed prior to enrolment; however, no significant chromosomal abnormalities were detected.

Participant NE004

Participant NE004 was a male participant who was referred with developmental delay and in whom a diagnosis of a RASopathy was suspected. This participant had vertical nystagmus, epicanthic folds, partial ptosis, bilateral down-slanted palpebral fissures, low set ears, a prominent nasal bridge, enlarged occipital horn ventricles, pectus excavatum, brachydactyly, bilateral contractures of the interphalangeal joints of the third and fourth digits, and a history of respiratory distress. Participant NE004 had undergone previous genetic testing for chromosomal abnormalities, receiving a negative result on CMA. This participant was referred to the study at five months of age after hospital admission for respiratory distress.

Participant NE005

Participant NE005 was a female patient with congenital heart defects referred by a medical geneticist for exome sequencing at 19 months of age. Participant NE005 was noted by her clinician to have epicanthic folds, a flat nasal bridge, malar hypoplasia, microcephaly, a ventricular septal defect, tricuspid valve defect, a history of feeding difficulties, global developmental delay and failure to thrive. The participant was recruited after surgery to correct her cardiac defects and presented with mild peripheral hypotonia at the time of recruitment, however, this was likely due to her post-operative state.

Participant NE006

Participant NE006 was a female patient referred for testing by her treating neonatologist at two weeks old due to her dysmorphic features. An aneuploidy was performed prior to enrolment; however, quantitative fluorescence PCR (qf-PCR) was negative. This participant was noted to have hypodensities in her left occipital lobe, a low hairline, low set ears, a submucosal cleft palate, small chin, microcephaly, short neck, a horseshoe kidney, bilateral talipes equinovarus, contractures, global brisk reflexes, and a history of feeding and swallowing difficulties.

Participant NE007

Participant NE007 was a male patient who was referred for testing with multiple congenital anomalies. This participant had a flat nasal bridge, low set ears, microcephaly, a short neck and chest, an imperforate anus, undescended testes, a micropenis, bilateral talipes equinovarus and bilateral genu recurvatum. Trisomy 18 was suspected in this participant; however, a negative aneuploidy result was received prior to enrolment.

Participant NE009

Participant NE009 was a male participant referred for testing by his treating paediatrician with a suspicion of a congenital myopathy. This participant was referred with up-slanted palpebral fissures, blue sclera, posteriorly rotated ears, glossoptosis, contractures at the wrists and distal interphalangeal joints, arthrogryposis of all joints, kyphoscoliosis, hyporeflexia, bilateral talipes equinovarus, global decreased tone and reflexes, developmental delay and slightly raised creatine kinase levels (6.5ug/L, reference range 0-6.22).

Participant NE010

Participant NE010 was a male participant referred by his paediatrician due to a cardiac anomaly and subtle dysmorphic features. This participant had undergone corrective surgery for a Tetralogy of Fallot (a combined ventricular septal defect, pulmonary stenosis, misplaced aorta and thickened right ventricle wall) prior to enrolment. Participant NE010 was also noted to have a broad forehead, down-slanted eyes and long palpebral fissures, hypotelorism, a depressed nasal bridge, a short nose, large cup-shaped ears, a small chin, knee and finger laxity, bilateral clinodactyly of the fifth

digit, and bilaterally absent toenails on two toes. A CMA was performed prior to enrolment; however, no significant chromosomal aberrations were detected

Participant NE011

Participant NE011 was a male participant referred by his neonatologist due to his dysmorphic features. Participant NE011 presented with significant microcephaly at birth (six standard deviations below the mean) and with weight, height and head circumference measuring below the third centile at enrolment. Bifrontal narrowing, sparse arched eyebrows, hypotelorism, up-slanted palpebral fissures, prominent alae nasi, upper ear helical notches, retromicrognathia, a short neck, mild joint restriction at the elbows, a café au lait spot on the upper right limb and hyperactive behaviour were noted by the referring clinician. The participant had a postnatal NICU admission history for respiratory distress.

Participant NE012

Participant NE012 was a female participant who was referred to the study due to her dysmorphic features. This participant had a six-year-old male sibling who was developmentally delayed and non-verbal, also suggestive of a genetic condition. Participant NE012 presented with severe bilateral microphthalmia, hypertelorism, posteriorly rotated low set ears with a prominent tragus, a high arched palate, retromicrognathia, an atrial septal defect, patent ductus arteriosus, sandal gap feet and neonatal jaundice.

Participant NE013

Participant NE013 was a female participant who presented with abnormal growth parameters and dysmorphism. This participant presented with macrosomia, posteriorly rotated low set ears, macroglossia, omphalocele and low muscle tone.

Participant NE014

Participant NE014 was a female referred for testing by her neonatologist due to multiple congenital anomalies. A trisomy was suspected in this participant; however, her qf-PCR result was negative. She presented with agenesis of the corpus callosum, hydrocephalus, hypertelorism, a flat nasal bridge, a tight jaw, and a short neck.

Participant NE015

Participant NE015 was a female participant with multiple congenital anomalies, who was a twin to a stillborn female who also presented with multiple congenital anomalies. Participant NE015 presented with bilateral exotropia, a depressed nasal bridge, low set ears with hypoplastic lobes, a short neck, hyperlaxity of the finger joints, omphalocele, cloacal exstrophy, an imperforate anus, a spinal cord lesion at birth, talipes equinovarus and mild truncal hypotonia. A diagnosis of OEIS complex, a combination of defects comprising omphalocele, exstrophy of the cloaca, imperforate anus, and spinal defects, was suspected in this participant (Bohring, 2002). The contribution of genetics to the presentation of OEIS complex is unclear (El-Hattab et al., 2010, Reutter et al., 2016). This participant was referred to elucidate if any genetic factors may be contributing to her OEIS-like clinical presentation

Participant NE016

Participant NE016 was a male participant referred for exome sequencing by his neonatologist due to his severe neurological features. Participant NE016 was a dizygotic twin to a well and healthy sister. Participant NE016 presented with cortical thickening, bilateral pachygyria of the frontal lobes, hypodensities in the left basal ganglia and right temporal area, dilated third ventricles, severe epileptic encephalopathy, bilateral lissencephaly, cortical visual impairment with structurally normal eyes, a head lag, dysmorphic head shape and facial features, a wide inter-mamillary space, scrotal dysplasia and generalised hypotonia. This participant displayed severe drug refractory epilepsy and underwent a functional hemispherectomy to manage his seizures prior to enrolment.

Participant NE017

Participant NE017 was a male patient who presented with renal anomalies, referred for testing by his neonatologist. This participant was reported to have bilateral pelvic-utero obstruction, bilateral hydronephrosis and was failing to thrive.

Participant NE018

Participant NE018 was a female patient referred to the study by her neonatologist at three days old for multiple congenital heart defects. This participant presented with Ebstein’s anomaly, pulmonary atresia, and small pulmonary arteries.

Participant NE019

Participant NE019 was a female participant referred by her paediatrician at four months of age as she presented with multiple congenital anomalies. She had large anterior fontanelle, a short neck, patent ductus arteriosus, a dilated left heart, severe systolic dysfunction, reduced right ventricular function, severe pulmonary hypertension, short limbs, bilateral talipes equinovarus, bilateral genu recurvatum, hyperpigmentation on the lower face and left body, Mongolian spot, metabolic acidosis, hypocalcaemia, and hypophosphatemia. This participant had a history of multiple PICU admissions over her short life span.

Participant NE020

Participant NE020 was a male participant referred by his paediatrician due to his hepatic abnormalities. A history of low gamma-GT intrahepatic cholestasis, with giant cell formation and hepatocyte degeneration were reported. The participant had undergone a liver biopsy to confirm that native bile ducts were present in the liver.

Participant NE023

Participant NE023 was a female patient who presented with multiple congenital anomalies. The participant's mother was noted to have had a previous loss of a twin pregnancy at six months gestation. This participant was referred by her neonatologist who noted the following features: a prominent occiput, hydrocephalus, a wide forehead, epicanthic folds, low set ears, retrognathia, patent ductus arteriosus, wide inter-mamillary space, myelomeningocele, flaccid lower limbs, respiratory distress, and failure to thrive.

Participant NE025

Participant NE025 was a female participant referred for exome sequencing by her paediatrician at 11 months of age due to her gastrointestinal abnormalities. This participant presented with meconium cysts and peritonitis, failure to thrive, vomiting and recurrent diarrhoea, volvulus and hyponatremia. A diagnosis of CF was suspected. A sweat test showed normal chloride levels in the participant and did not corroborate a CF diagnosis. A genetic test prior to enrolment assessing 50 common CF variants found that Participant NE025 harboured a heterozygous pathogenic variant in the *CFTR* gene, c.3718-2477C>T (3849+10kbC>T). This test suggested that the participant was a carrier for CF but could not confirm a diagnosis.

Participant NE027

Participant NE027 was a female patient who presented with hepatic anomalies, including low gamma-GT intrahepatic cholestasis, as well as hepatomegaly and splenomegaly.

Participant NE028

Participant NE028 was a male patient referred for testing by his medical geneticist at 11 months old as he presented with multiple congenital anomalies. This participant was noted to have deep set eyes, low set ears, microcephaly, a right sided aortic arch, a left subclavian artery with vascular sling, patent ductus arteriosus, a ventricular septal defect, spatulate fingertips, a hypoplastic scrotum, non-palpated testes in the inguinal canal, mild truncal hypotonia, severe growth restriction and moderate developmental delay. Participant NE028 had a brother with developmental delay and a half-brother with intellectual delay, strengthening the case for the involvement of a hereditary factor in his clinical presentation.

Participant NE030

Participant NE030 was a male participant who also presented with multiple congenital anomalies. This participant had a flat nasal bridge, low set ears, long tapering fingers with hypoplastic nails, omphalocele, bladder exstrophy and a neural tube defect. A diagnosis of Trisomy 18 was suspected in this participant; however, qf-PCR was negative.

Participant NE031

Participant NE031 was a female patient referred for exome sequencing because of a cardiac malformation. This participant was referred with coarctation of the aorta, pericardial effusion, a tri-lobar left lung, and had experienced generalised tonic clonic seizures.

Participant NE032

Participant NE032 was a female participant who presented with cardiac anomalies and a suspicion of Trisomy 21, but in whom qf-PCR testing yielded negative results. This participant was referred with transposition of the great arteries, patent ductus arteriosus, pulmonary hypertension, collapse of the left lung resulting in respiratory failure and a cardiac murmur, short stubby fingers, a history of feeding difficulties and a history of hypocalcaemia.

Participant NE035

Participant NE035 was a male participant referred by a medical geneticist for clinical exome panels at 11 months old. This participant was the third child to consanguineous parents and had two healthy siblings. Participant NE035 presented with multiple congenital anomalies, including a broad forehead, sparse eyebrows, down-slanted palpebral fissures, vertical nystagmus, low set posteriorly rotated ears, microretrognathia, malar hypoplasia, a high arched palate, head lag, tracheal tug, inspiratory stridor, a wide inter-mamillary space, right sided rib crowding, bilateral dysplastic kidneys and renal dysfunction, bilateral upper and lower limb reduction defects with absent digits and constriction rings, truncal hypotonia, failure to thrive and a history of poor feeding.

Appendix J. Published Paper - Campbell et al. (2023)

frontiers | Frontiers in Genetics

TYPE Brief Research Report
PUBLISHED 09 November 2023
DOI 10.3389/fgene.2023.1277948

Check for updates

OPEN ACCESS

EDITED BY
Atanu Kumar Dutta,
All India Institute of Medical Sciences, Kalyani (AIIMS Kalyani), India

REVIEWED BY
Volkan Okur,
New York Genome Center, United States
Ferran Casals,
University of Barcelona, Spain

*CORRESPONDENCE
N. Carstens,
nadia.carstens@mrc.ac.za

RECEIVED 15 August 2023
ACCEPTED 23 October 2023
PUBLISHED 09 November 2023

CITATION
Campbell L, Fredericks J, Mathivha K, Moshesh P, Coovadia A, Chirwa P, Dillon B, Ghoor A, Lawrence D, Nair L, Mabaso N, Mokwele D, Novellie M, Krause A and Carstens N (2023), The implementation and utility of clinical exome sequencing in a South African infant cohort.
Front. Genet. 14:1277948.
doi: 10.3389/fgene.2023.1277948

COPYRIGHT
© 2023 Campbell, Fredericks, Mathivha, Moshesh, Coovadia, Chirwa, Dillon, Ghoor, Lawrence, Nair, Mabaso, Mokwele, Novellie, Krause and Carstens. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The implementation and utility of clinical exome sequencing in a South African infant cohort

L. Campbell[1], J. Fredericks[2], K. Mathivha[3], P. Moshesh[3], A. Coovadia[2], P. Chirwa[4], B. Dillon[1], A. Ghoor[2], D. Lawrence[2], L. Nair[2], N. Mabaso[1], D. Mokwele[1], M. Novellie[1], A. Krause[1] and N. Carstens[1,5]*

[1]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [2]Department of Paediatrics and Child Health, School of Clinical Medicine, Rahima Moosa Mother and Child Hospital, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [3]Department of Paediatrics and Child Health, School of Clinical Medicine, Nelson Mandela Children's Hospital, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [4]Nelson Mandela Children's Hospital, Johannesburg, South Africa, [5]Genomics Platform, South African Medical Research Council, Cape Town, South Africa

Genetic disorders are significant contributors to infant hospitalization and mortality globally. The early diagnosis of these conditions in infants remains a considerable challenge. Clinical exome sequencing (CES) has shown to be a successful tool for the early diagnosis of genetic conditions, however, its utility in African infant populations has not been investigated. The impact of the under-representation of African genomic data, the cost of testing, and genomic workforce shortages, need to be investigated and evidence-based implementation strategies accounting for locally available genetics expertise and diagnostic infrastructure need to be developed. We evaluated the diagnostic utility of singleton CES in a cohort of 32 ill, South African infants from two State hospitals in Johannesburg, South Africa. We analysed the data using a series of filtering approaches, including a curated virtual gene panel consisting of genes implicated in neonatal-and early childhood-onset conditions and genes with known founder and common variants in African populations. We reported a diagnostic yield of 22% and identified seven pathogenic variants in the NPHS1, COL2A1, OCRL, SHOC2, TPRV4, MTM1 and STAC3 genes. This study demonstrates the utility value of CES in the South African State healthcare setting, providing a diagnosis to patients who would otherwise not receive one and allowing for directed management. We anticipate an increase in the diagnostic yield of our workflow with further refinement of the study inclusion criteria. This study highlights important considerations for the implementation of genomic medicine in under-resourced settings and in under-represented African populations where variant interpretation remains a challenge.

KEYWORDS

clinical-exome, infant, diagnostic, South Africa, LMIC, implementation

1 Introduction

Genetic conditions are significant contributors to infant mortality, morbidity, and hospitalisations globally (Kingsmore et al., 2020). Despite advances in diagnostics with next-generation sequencing (NGS) and microarrays, it remains challenging to make diagnoses in infant populations. Infants often present with atypical or non-specific disease symptoms; many genetic disease phenotypes cannot be distinguished during the neonatal period; and disease progression may often be very rapid, making the identification and diagnosis of genetic conditions difficult (Wilkinson et al., 2016; van Diemen et al., 2017; French et al., 2019).

Clinical and whole exome sequencing (CES and WES), have been widely investigated for their use in diagnosing ill infants, with many studies evidencing their benefit (Petrikin et al., 2015; Smith et al., 2015; Willig et al., 2015; Stark et al., 2016; Meng et al., 2017; van Diemen et al., 2017; Farnaes et al., 2018; French et al., 2019; Kingsmore et al., 2019; Lunke et al., 2020; Wang et al., 2020). While WES targets all protein-coding regions of the genome, CES targets select genes with known disease associations and their flanking splice regions, and has been shown to provide an early, definitive diagnosis, preventing a diagnostic odyssey, and allowing for targeted clinical management (Smith et al., 2015). Virtual panels examining a subset of these clinically relevant genes from WES/CES data can be utilized to minimise the initial time and cost of analysis while offering the potential of comprehensive analysis of additional genes at a later stage with no additional laboratory costs, a key cost consideration for resource-constrained settings. Previous studies suggest that earlier diagnoses promote improved patient care and outcomes by facilitating changes in medical treatment, the early introduction of targeted therapies, increased surveillance, the appropriate initiation of comfort care, the reduction of costly, repeat, and sometimes invasive investigations, and the instigation of cascade testing for family members and counselling for reproductive planning–all measurable utility for the use of diagnostic testing (Rabbani et al., 2012; Splinter et al., 2018; Wright et al., 2018; Lunke et al., 2020).

Despite evidence from numerous research studies illustrating the benefits of CES for the diagnosis and management of ill infants, CES is not accessible in many parts of the world, particularly in low- and middle-income countries (LMICs). In many LMICs, CES is only offered in limited research settings (or not at all) creating larger health disparities in regions where access to primary healthcare and diagnostic services are already poor. This is true across the African continent, with few countries having established genetic services (Kamp et al., 2021). In South Africa, only three of the nine provinces offer genetic services in the State healthcare system, which services more than 80% of the population (Kromberg et al., 2013a; Stats SA, 2022). With limited access to genetic services, many affected by genetic conditions go undiagnosed and untreated. The lack of funding and resources for CES implementation is widely acknowledged by stakeholders, particularly shortages in infrastructure and trained genetics professionals (Kamp et al., 2021; Lumaka et al., 2022). The implementation of CES in the South African State healthcare system needs to be informed by international guidelines and standards, but requires optimisation in the local context to accommodate the scarcity of resources and the limited bioinformatics and genomics capacity (Kamp et al., 2021). The implementation of CES is further complicated by challenges in genetic data interpretation due to the underrepresentation of African populations in genomic databases and literature, African genetic diversity, and the lack of disease registries to document genetic disease prevalence (Kromberg et al., 2013b; Baynam et al., 2020; Lumaka et al., 2022). Despite these challenges, CES may offer significant improvements in the clinical management of vulnerable, underserved South African populations, including ill infants in the neonatal intensive care unit (NICU). Without a genetic diagnosis, infants with genetic conditions are constrained by less accurate risk assessments, prognoses, and specialist referrals; minimal assistance from support organisations and government welfare; and no access to emerging therapies and clinical trials, widening the existing healthcare disparities African populations face in accessing personalised healthcare.

With the shift in focus of genomic medicine to implementation science and translation into clinical practice, studies are needed to determine the benefits and challenges faced in implementing CES in real-world settings (Kingsmore and Cole, 2022). There are many barriers to the implementation of genomic medicine in LMICs which need to be investigated and addressed for genomic medicine and CES to be successfully integrated into global healthcare systems. To meet the World Health Organisation's sustainable development goal to reduce neonatal and children under-5 mortality, the implementation of adequate genetic services to address the burden of genetic disease, often overshadowed and masked by infectious and communicable disease, is necessary in LMICs. Implementation should address the various barriers faced, including limited infrastructure and technology, shortages in the genetics workforce, poor genomic literacy amongst healthcare providers, poor literacy and education amongst the general population, language barriers and the lack of standardised genomics terminology in local languages, cultural and societal nuances around family structure, cultural beliefs around disease causality, and paucity of information around genetic disease burden in African populations (Kamp et al., 2021). These challenges will vary across LMICs due to population, economic, political, and social diversity, rendering a one-size-fits all implementation approach inappropriate for most settings and making investigations into unique, country-specific challenges necessary (Tiffin, 2014; Kamp et al., 2021; Lumaka et al., 2022).

We performed a scoping study to evaluate the diagnostic utility of singleton CES in a cohort of ill infants suspected of having a genetic condition in the South African State healthcare system through a series of three virtual gene panels. This study provides insight into the implementation of CES for ill infants in an under-resourced LMIC setting, where access to genetic services is limited, the burden of genetic disease is unknown and the interpretation of genomic data remains a considerable challenge.

2 Methods

The study workflow is summarised in Figure 1.

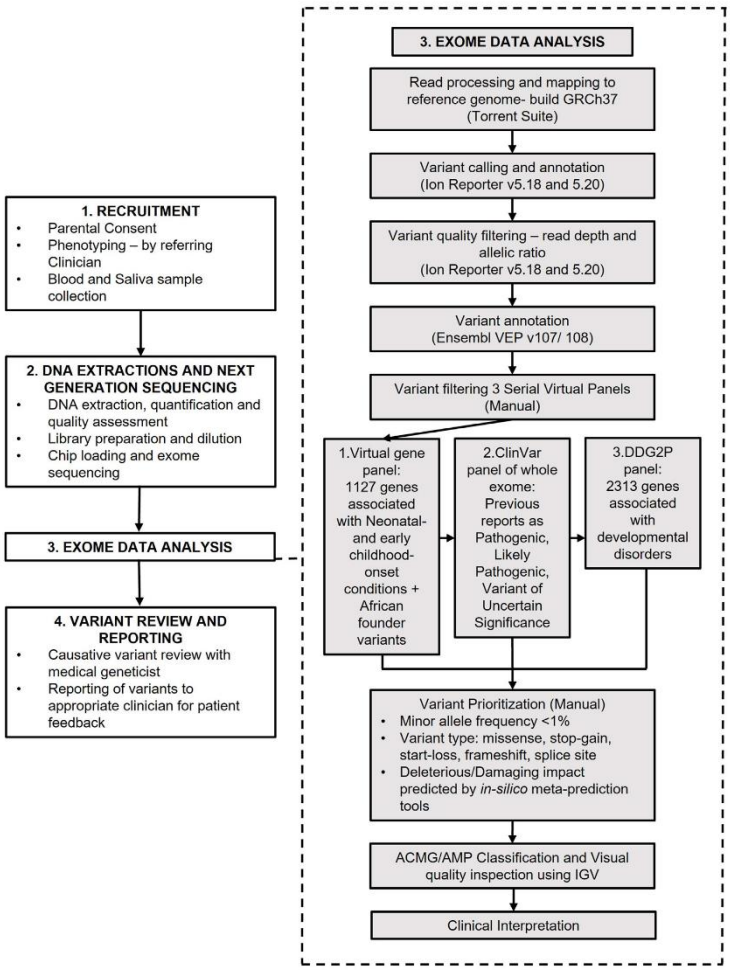


FIGURE 1
Study overview and analysis workflow.

2.1 Recruitment

A cohort of thirty-two ill infants, suspected of having a genetic condition, were recruited from two State hospitals in Johannesburg, South Africa: Rahima Moosa Mother and Child Hospital (RMMCH) in Coronationville, and Nelson Mandela Children's Hospital (NMCH) in Parktown. These hospitals are regional and quaternary hospitals respectively and serve mostly State patients. Participants were referred for genetic testing either directly by neonatologists and paediatricians without consultation with medical geneticists, or by medical geneticists at call out consultations or referral. Clinicians were provided a list of recruitment criteria, highlighting a broad range of phenotypic features that may be suggestive of a genetic cause for an ill infant's condition. Infants were considered eligible for recruitment if they presented with any of the following: multiple congenital anomalies, a concern for their neurological status, metabolic abnormalities of uncertain cause, dysmorphic features or if they had abnormal growth parameters. Infants were excluded if they had significant teratogen exposure during pregnancy, experienced birthing trauma or asphyxia, or had a diagnosed infection that could be the primary cause of illness. Premature patients were considered eligible for inclusion if their clinical features were not primarily explained by prematurity.

Parents and guardians provided informed consent on behalf of infants for participation in the study after the study was explained to them by a principal investigator or genetic counsellor. A blood sample was collected for DNA extraction and exome testing from each participant. A saliva sample from available parents was banked at consent, however, this study employed exome sequencing of probands only.

Phenotypic information regarding the condition and medical history of the participants was provided by the referring clinician. Clinicians were provided a short format phenotype collection rubric to assist with this as a Redcap form (Supplementary Table S1) (Harris et al., 2009; Harris et al., 2019), and requested to provide any information they believed was relevant to a possible genetic diagnosis.

2.2 DNA extraction and next-generation sequencing

DNA extraction was performed on the stored participant blood samples using a modified salting-out method (Miller et al., 1988). DNA quality and quantity was assessed using the NanoDrop 2000 (Thermo Fisher Scientific, United States), gel electrophoresis and the Qubit 4.0 system (Thermo Fisher Scientific, United States). Library preparation was performed manually using the Ion Ampliseq Exome RDY library preparation kit, as per the manufacturer's instruction. DNA libraries were quantified and sized using the TapeStation High Sensitivity D5000 ScreenTape assay (Agilent Technologies, Germany). Diluted libraries were loaded onto Ion 540 sequencing chips for exome sequencing using the automated Ion Chef instrument (Thermo Fisher Scientific, United States). Loaded chips were sequenced in-house on an Ion GeneStudio S5 sequencer (Thermo Fisher Scientific, United States). The laboratory processing time for four samples was approximately 5 days.

2.3 Exome data analysis

Read alignment to the GRCh37 human reference genome was performed on the Ion Torrent Suite software (Thermo Fisher Scientific, United States). Variant calling was subsequently performed using Ion Reporter software (Thermo Fisher Scientific, United States). Variants were annotated using the Ensembl Variant Effect Predictor (VEP), versions 107 and 108 (McLaren et al., 2016).

Variants were filtered and prioritised for manual curation using three sequential filtering approaches, summarised in Figure 1. The first variant filtering approach retained variants in a virtual gene panel consisting of 1,127 genes, curated for their association with neonatal- and early-childhood-onset conditions (Ceyhan-Birsoy et al., 2017; Milko et al., 2019) and genes with known founder variants and common variants in African populations (Krause et al., 2018) (NICU/African) (Supplementary Table S2). If no causative variant was identified, the second variant filtering approach, retaining variants with previous pathogenic, likely pathogenic and variant of uncertain significance (VUS) classifications on ClinVar, including variants with conflicting interpretations, was employed (Landrum et al., 2018). The third filtering strategy, employed if no causative variant was identified from the first two strategies, retained variants in 2,313 genes on the Deciphering Developmental Disorders Genotype to Phenotype (DDG2P) gene panel (Thormann et al., 2019). Filtered variants were then prioritised for manual curation based on the type of variant, population minor allele frequencies, previous pathogenic and likely pathogenic interpretations on ClinVar and predicted deleterious impacts on protein function from *in silico* prediction tools CADD, REVEL, BayesDel and MetaRNN (Ioannidis et al., 2016; Feng, 2017; Rentzsch et al., 2019; Li et al., 2022). Variants were visually inspected to determine quality using Integrated Genomics Viewer software (IGV) (Robinson et al., 2011).

Prioritised variants were classified according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines to determine their pathogenicity (Richards et al., 2015). Filtering, prioritization, manual curation and classification took approximately 6 h to perform per filtering strategy per sample. Pathogenic and likely pathogenic variants were then reviewed by a multidisciplinary team to determine if the diagnosis was appropriate for the participant. Findings were reported in a research report and returned to the referring clinicians and participants. VUSs and incidental findings were not reported in this study. Positive molecular findings were returned to parents by a genetic counsellor or medical geneticist, and appropriate downstream medical intervention and referrals initiated where necessary. Parents of participants with negative findings were invited for further consultation with a medical geneticist to discuss the implications of the negative result and determine an appropriate clinical management route going forward.

3 Results

3.1 Cohort characteristics

Thirty-two infants under the age of two, with a suspicion of a genetic condition were recruited for CES between May 2021 and December 2022 (Table 1). A single participant was enrolled and excluded as the participant had received a blood transfusion and demised before an appropriate amount of time had passed for a blood sample to be drawn. Most participants were referred for genetic testing directly by their treating neonatologist or paediatrician (23) and were not formally assessed by a medical geneticist prior to recruitment. An equal proportion of male and female participants were recruited. Most participants were of African ancestry (26). Only two participants, NE025 and NE035, were born to self-reported consanguineous parents. The most frequent indications for CES were multiple congenital anomalies (9), cardiac defects and abnormalities (6), general dysmorphism (3) and neuromuscular abnormalities (3). No participants had abnormalities detected prenatally on ultrasound or had prenatal genetic investigations, however, fetal anomaly scans are infrequently performed as standard obstetric care in the State healthcare system.

3.2 Genetic diagnoses from singleton CES

Singleton CES was performed on all participants with an average coverage of 134%X and 90% uniformity. A molecular diagnosis was

TABLE 1 Demographics and clinical characteristics of the patient cohort referred for clinical exome sequencing.

Characteristic		Number of individuals	Percentage of cohort (%)	Number of positive diagnoses	Percentage of positive diagnoses (%)
Total number of participants		32	—	7	21.9
Hospital recruitment site	RMMCH	17	53.1	6	35.3
	NMCH	15	46.9	1	6.7
Sex	Female	16	50.0	3	18.8
	Male	16	50.0	4	25.0
Referring clinician	Neonatologist/ Paediatrician	23	71.9	4	17.4
	Geneticist	9	28.1	3	33.3
Ancestry	African	26	81.3	5	19.2
	European	1	3.1	—	—
	Mixed	1	3.1	1	100.0
	Other	4	12.5	1	25.0
Consanguinity	Consanguineous parents	2	6.3	—	0.0
	Non-consanguineous parents	30	93.8	7	~23.3
Indication for genetic testing	Neuromuscular	3	9.4	2	66.7
	Skeletal	2	6.3	2	100.0
	Cardiac	6	18.8	—	—
	Renal	2	6.3	1	50.0
	Multiple congenital anomalies	9	28.1	1	11.1
	Hepatic	2	6.3	—	—
	Developmental	2	6.3	1	50.0
	Abnormal growth	1	3.1	—	—
	Dysmorphic	3	9.4	—	—
	Neurological	1	3.1	—	—
	Gastrointestinal	1	3.1	—	—

Abbreviations: RMMCH, Rahima Moosa Mother and Child Hospital; NMCH, Nelson Mandela Children's Hospital.

confirmed for 7 of the 32 recruited participants, resulting in a diagnostic yield of 22% (Table 2). Five of the identified causative variants were in genes in our NICU/African virtual panel.

Two participants received a severe myopathy diagnosis. A homozygous *STAC3* variant (c.851G>C; p.Trp284Ser) was identified in participant NE001. This variant has been associated with a diagnosis of *STAC3* myopathy (Horstick et al., 2013; Telegrafi et al., 2017; Waldrop et al., 2017; Zaharieva et al., 2018; Schoonen et al., 2019; Ravenscroft et al., 2021; Saleh et al., 2021). Multiple *in silico* prediction tools predict this variant to be deleterious to STAC3 function, confirmed in zebrafish models, which show a deficiency in skeletal muscle excitation-contraction coupling as a result of this missense mutation (Horstick et al., 2013; Linsley et al., 2017). A hemizygous, nonsense *MTM1* variant (c.664C>T; p.Arg222Ter), affecting the functional myotubularin phosphatase domain of the MTM1 protein, was identified in participant NE021. This variant has been associated with severe X-linked myotubular myopathy (Laporte et al., 1997; Tanner et al., 1999; Biancalana et al., 2003; Longo et al., 2016). Both diagnosed myopathies had severe disease presentations, resulting in the demise of both infants within a week of life. Parents of these infants were offered prenatal testing for future pregnancies.

A pathogenic, heterozygous missense variant in *COL2A1* (c.3589G > A; p.Gly1197Ser) was identified in participant NE022. This variant occurs in a predicted missense constrained gene and has predicted deleterious effects on type II collagen stability and structure (Beck et al., 2000). Other amino acid substitutions at the same position, p.Gly1197Ala and p.Gly1197Arg, have previously been reported as pathogenic. Mutations in the *COL2A1* gene have been associated with a spectrum of overlapping skeletal dysplasias,

TABLE 2 Positive infant diagnoses through clinical exome sequencing using three filtering strategies.

Study ID	Hospital	Referring clinician type	Sex	Ancestry	Indication for testing	Phenotypic features	Variant HGVS; gene OMIM number	Variant type	Filtering strategy	ACMG/AMP classification and codes	ClinVar accession ID and classification	Variant inheritance, variant coverage	Diagnosed condition
NB001	RMMCH	Neonatologist/ Paediatrician	M	African	Neuromuscular	Myopathic facies, generalised hypotonia, reduced tendon reflexes, low set, posteriorly rotated ears, high arched palate, cryptorchidism, wide intermammillary distance, genu recurvatum, rocker bottom foot, single transverse palmar crease	*STAC3*:c.851G > C (p.Trp284Ser) 615521	Missense	NICU/ African Panel	Pathogenic PM3_strong, PS3, PP3, PP5_supporting	VCV000088744.39 2 star Pathogenic	Autosomal recessive, 124X	STAC3 (Bailey Bloch) myopathy
NB021	RMMCH	Neonatologist/ Paediatrician	M	Other Pakistani	Neuromuscular	General hypotonia, myopathic facies, frog like posture, areflexia, low power, breathing difficulties, open anterior and posterior fontanelle, generalised hypertrichosis, retrognathia, narrow chest, hypoplastic scrotum, cryptorchidism	*MTM1*:c.664C > T (p.Arg222Ter) 300415	Stop gain	NICU/ African Panel	Pathogenic PVS1, PM2_supporting, PP5_supporting	VCV000158994.14 2 star Pathogenic	X linked recessive, 47X	X linked myotubular myopathy
NB022	NMCH	Geneticist	F	African	Skeletal	Mild rhizomelic shortening, restricted elbow extension and hip abduction, truncal hypotonia, short sternal prominence, pectus carinatum, short and narrow chest, Harrison's sulcus, flared ribs, protuberant abdomen with palpable soft liver and spleen, exaggerated lumbar lordosis, relative macrocephaly, flat nasal bridge,	*COL2A1*: c.3589G > A (p.Gly1197Ser) 120140	Missense	NICU/ African Panel	Likely Pathogenic PM5, PM1, PM2_supporting, PP3, PP2, PP5_supporting	VCV000017361.21 2 star Pathogenic/ Likely Pathogenic	Autosomal dominant, 85X	COL2A1 collagen disorder

(Continued on following page)

TABLE 2 (*Continued*) **Positive infant diagnoses through clinical exome sequencing using three filtering strategies.**

Study ID	Hospital	Referring clinician type	Sex	Ancestry	Indication for testing	Phenotypic features	Variant HGVS; gene OMIM number	Variant type	Filtering strategy	ACMG/AMP classification and codes	ClinVar accession ID and classification	Variant inheritance, variant coverage	Diagnosed condition
						prominent forehead, epicanthal folds, infraorbital creases, grey sclera, malar hypoplasia, short neck							
NE026	RMMCH	Neonatologist/ Paediatrician	M	African	Developmental	Neonatal seizures, hypocalcaemia, failure to thrive, alopecia, sparse eyebrows, frontal bossing, scaphocephaly, malar flattening, flat nasal bridge, retromicrognathia, mild abdomen distention, scrotal dysplasia, mild hypotonia with normal power, dry skin, feeding difficulties due to fragile teeth and bleeding gums	*SHOC2*c.4A > G (p.Ser2Gly) 602775	Missense	ClinVar Filter	Pathogenic PS2_very strong, PS4, PS3_moderate, PM2_supporting, PP2, PP5_supporting	VCV000006821.64 3 star Pathogenic	Autosomal dominant, 34X	Noonan like syndrome with loose anagen hair
NE029	RMMCH	Geneticist	M	Mixed	Multiple congenital anomalies	Cataracts, blocked tear ducts, bilateral eye opacity, bilateral nystagmus, retrognathia, open anterior fontanelle without craniotabes, redundant skin at shoulder joints, brachydactyly, tapered fingers, hyperlaxity of finger joints, short stature, global hypotonia with reduced reflexes and power, head lag, osteomalacia on x ray, proteinuria, stagnating weight	*OCRL*c.1621C > T (p.Arg541Ter) 300535	Stop gain	NICU/ African Panel	Pathogenic PVS1_strong, PM2_supporting, PP5_supporting	VCV000521093.9 2 star Pathogenic	X linked recessive, 1.6X	Lowe syndrome

(Continued on following page)

TABLE 2 (*Continued*) Positive infant diagnoses through clinical exome sequencing using three filtering strategies.

Study ID	Hospital	Referring clinician type	Sex	Ancestry	Indication for testing	Phenotypic features	Variant HGVS; gene OMIM number	Variant type	Filtering strategy	ACMG/AMP classification and codes	ClinVar accession ID and classification	Variant inheritance, variant coverage	Diagnosed condition
NE033	RMMCH	Geneticist	F	African	Renal	Congenital nephrotic syndrome, left ventricular hypertrophy, effusion of right ventricle, peripheral pulmonary stenosis, anasarca, proteinuria, electrolyte abnormalities, hypothyroidism	*NPHS1*: c.1379G>A (p.Arg460Gln) 602716	Missense	NICU/ African Panel	Likely Pathogenic PM3_strong, PM1, PP5_supporting	VCV000056438.17 2 star Pathogenic	Autosomal recessive 89X	Congenital nephrotic syndrome
NE034	RMMCH	Neonatologist/ Paediatrician	F	African	Skeletal	Metaphyseal widening left femur and midshaft fracture of right femur, hypoplastic left tibia, restricted elbow extension, proximally inserted thumb, tapered fingers bilaterally, fixed hip flexion bilateral hip and knee extension, bilateral talipes equinovarus, hypoplastic nails both feet, prominent occiput, Mongolian blue spot over buttocks and sacral, dimple over medial malleoli, reduced plantar creases	*TRPV4*:c.806G>A (p.Arg269His) 605427	Missense	ClinVar Filter	Pathogenic PS4, PS3, PM5, PM1, PP5_supporting	VCV000005000.37 2 star Pathogenic	Autosomal dominant, 73X	Spondylometaphyseal dysplasia

Abbreviations: RMMCH, Rahima Moosa Mother and Child Hospital; NMCH, Nelson Mandela Children's Hospital; M, Male; F, female; NICU/African, Neonatal- and early childhood onset and African Founder Variant.

therefore, this participant received a broad diagnosis of a *COL2A1* collagen disorder (Cole et al., 1993; Nishimura et al., 2005; Terhal et al., 2015). Further assessments are needed to refine this diagnosis. This participant will need to be monitored for eye abnormalities, cervical instability and have their respiratory function assessed by the appropriate medical specialties.

Participant NE026 was diagnosed with Noonan-like syndrome with loose anagen hair caused by a heterozygous *SHOC2* variant (c.4A > G, p.Ser2Gly). This variant, carrying a three-star ClinVar classification, has been shown to alter MAPK activation in *in vitro* studies and animal models (Cordeddu et al., 2009; Hoban et al., 2012; Gripp et al., 2013; Gargano et al., 2014; Choi et al., 2015). This participant will be referred for neurodevelopmental and cardiac assessments, as cardiac defects and psychomotor delay are frequently observed in this condition.

A hemizygous, nonsense *OCRL* variant (c.1621C > T, p.Arg541Ter) was identified in participant NE029, associated with X-linked recessive Lowe (Oculo-cerebro-renal) syndrome. This specific premature termination variant has been observed in other Lowe syndrome patients with similar clinical presentations (Hichri et al., 2011; Yang et al., 2014; Recker et al., 2015). This participant will require ophthalmology treatment for cataracts and glaucoma, as well as electrolyte monitoring for renal tubular acidosis. The mother of Participant NE029 received carrier screening by Sanger sequencing, confirming that she was not a carrier for the identified pathogenic variant.

Participant NE033 received a diagnosis of severe congenital nephrotic syndrome (Finnish type), harbouring a homozygous variant in the *NPHS1* gene (c.1379G > A, p.Arg460Gln). This variant has been reported in a homozygous state in multiple congenital nephrotic syndrome cases (Beltcheva et al., 2001; Schoeb et al., 2010; Warejko et al., 2018; Mann et al., 2019). Participant NE033 presented with end stage renal disease and was unable to access renal replacement therapy in South Africa. Due to a poor prognosis and limited available medical intervention, this family received counselling on comfort care.

A heterozygous missense variant in *TRPV4* (c.806G > A, p.Arg269His) was identified in participant NE034. This variant has been extensively reported in the literature for its association with autosomal dominant neuromuscular disease and skeletal dysplasia (Auer-Grumbach et al., 2010; Deng et al., 2010; Landoure et al., 2010; Zimon et al., 2010; Echaniz-Laguna et al., 2014; Louis et al., 2014; Biasini et al., 2016; Fleming and Quan, 2016; Jedrzejowska et al., 2019). Multiple functional studies support this variant's gain of function role in increasing calcium channel activity, resulting in cell death and axonal degeneration (Deng et al., 2010; Landoure et al., 2010; Fecto et al., 2011; Klein et al., 2011; Takahashi et al., 2014). Participant NE034 presented clinically with only the skeletal features of *TRPV4*-associated disease and received a diagnosis of spondylometaphyseal dysplasia. As the neuropathic symptoms of this disease have been reported to occur after the neonatal period, this participant will need to be monitored for the development of these symptoms (Fleming and Quan, 2016). This participant will require physical therapy to maintain lower limb function. Pulmonary function and cervical spine stability will also be monitored in this participant.

4 Discussion

This study generated valuable insights into the practicalities and utility of implementing CES in a NICU setting in South Africa. As the developed world moves toward rapid testing for ill infants, challenges in the implementation of genomic medicine in African and LMICs exacerbates existing healthcare disparities due to the slow uptake of these newer technologies, particularly in Sub-Saharan Africa, where most NGS testing, even in the research setting, is performed outside of the African continent (Baine-Savanhu et al., 2023). These challenges require careful consideration for the implementation of cutting-edge genomic services and for health equity to be achieved (Kingsmore and Cole, 2022).

We identified a genetic cause in seven ill infants with a range of clinical disease presentations, achieving a minimum diagnostic yield of 22%. Many studies globally have assessed the utility of CES in infant populations, showing impressive diagnostic yields of up to 70% (Clark et al., 2018), however, there is an absence of African individuals in these studies and low representation from LMICs. Patients of African ancestry are less likely to receive a diagnosis using current analytical strategies, as shown in the Deciphering Developmental Disorders study, creating a significant healthcare disparity for these populations (Wright et al., 2023). A tailored implementation strategy is needed for successful uptake of CES in Africa and in LMICs.

Comprehensive analysis strategies offer a higher probability of identifying a causative variant, however, these strategies are more time and labour intensive and have higher VUS rates (Stark and Ellard, 2022). A curated virtual panel of carefully selected genes, as demonstrated in this study, is an effective first approach for exome data analysis. Five of the seven causative variants identified in this study were in genes in our custom NICU/African panel, illustrating its utility for infant cohorts. The simplification of analysis pipelines and strategies in LMIC settings is necessary to deliver a result in a timely manner and with a low dependence on computational infrastructure and bioinformatics capacity. More comprehensive analysis pipelines, such as whole exome analysis, and the investigation of CNVs and structural variants could be deployed next for the remaining undiagnosed participants (Splinter et al., 2018). Undiagnosed participants may also benefit from reanalysis to incorporate updates to the virtual gene panel, new variant information, novel gene-disease associations, updated phenotypic data and bioinformatic advancements (Schobers et al., 2022).

The implementation of CES across the African continent and in LMICs is significantly under-investigated and many of the challenges associated with CES implementation into genetic service are exacerbated in an African setting. The significant barriers to the implementation of genomic medicine include a lack of infrastructure and support for clinical translation, the scarcity of information regarding African genetic epidemiology, poor genomic literacy among healthcare workers, few formally trained genetics professionals, a lack of investment by governments and international funders, the low cost effectiveness in establishing genetic services, and the fear of widening the existing disparities by only benefiting those who can afford access to these technologies (Jongeneel et al., 2022). These challenges, many seen in this study, need to be addressed when planning implementation strategies for CES in our context.

More than 70% of our infant cohort was referred for genetic testing directly by a neonatologist or paediatrician. Appropriate phenotyping and a predictive clinical diagnosis from a medical geneticist can direct CES data interpretation and contribute to a higher diagnostic yield (Trujillano et al., 2017; Gubbels et al., 2020). The limited formally trained genetics workforce in South Africa makes input from trained geneticists prior to genetic testing a virtually impossible task (Kromberg et al., 2013a). The engagement and upskilling of primary care medical specialities is necessary for CES implementation in under-resourced settings, to allow the few medical geneticists and genetic counsellors to be most efficiently utilised (Dragojlovic et al., 2020; Chou et al., 2021).

The lack of representation of African and Sub-Saharan African populations in publicly available databases and limited understanding of African genetic disease epidemiology makes CES data analysis and interpretation complex. A representative reference genome which considers African population diversity and disease epidemiology is essential in identifying disease-causing variants in a timely manner. A better understanding of genetic disease in African populations is key to improved management and service of our populations. This study allowed us to investigate genetic disease epidemiology in an underrepresented African population. The variant identified in participant NE001, diagnosed with *STAC3* myopathy, has the highest carrier rate in the Genome Aggregation Database (GnomAD, accessed June 2023) in individuals of African/African American ancestry (Zaharieva et al., 2018; Gromand et al., 2022). Unpublished research from the South African National Health Laboratory Service revealed a frequency of 20% for this variant in a homozygous or compound heterozygous state in *SMN1* negative African patients with myopathic phenotypes (Mhlongu et al., 2022). Despite its prevalence in African-ancestry patients, *STAC3* testing has only been recognized as a common myopathy and introduced into diagnostic service in the South African State system within the past year.

Singleton CES was utilised in this study to elucidate the potential cause of disease in our infant cohort. Despite its recognised diagnostic superiority, trio-sequencing is costly to consider in LMICs, where limited financial resources must effectively serve large populations {Baine-Savanhu et al., 2023 #416} (Clark et al., 2018; Kingsmore et al., 2019). Trio-sequencing is thus not affordable at present in LMICs and the benefit versus cost requires further investigation. Complex family dynamics and social issues are other key considerations for African populations. In this cohort, only 31% of participants had both parents available at enrolment due to a variety of social factors, including: admission at hospitals far away from their residence with no access to transportation; the inability to get time off work for hospital visits; some participants were African immigrants who were not in South Africa as a complete family unit; and many participants were born to single-parent households and families. A diagnosis was still achievable in some participants in this study with singleton sequencing, illustrating the effectiveness and utility of this tool as a first-tier strategy, despite difficulties in documenting *de novo* mutations and reoccurrence risk. For effective implementation of genomic services in Africa, these social challenges need to be embraced to prevent the exclusion of patients who would greatly benefit from access to these services.

Approximately 80% of participants did not receive a molecular diagnosis in this study. The workflow offers a starting strategy in the diagnostic trajectory of these participants, who would benefit from further investigations to fill in gaps not addressed in our virtual gene panels. Refinement of the inclusion and exclusion criteria may provide more clarity to referring clinicians as to which infants would most benefit from CES, allowing for better utilisation of the scarce resources for testing. More comprehensive phenotyping may also provide more guidance in participant screening and in variant interpretation for a more cost- and time-effective implementation strategy.

In conclusion, this study yielded a definitive diagnosis for seven families affected by a genetic condition with minimal access to genetic testing in the South African State healthcare system, providing the first evidence of the diagnostic utility of CES for ill infants in an under-resourced NICU setting in Africa. Our experience can be used as a point of departure to develop feasible, local CES implementation strategies, utilising appropriately curated virtual panels to provide clinically actionable findings within an appropriate timeframe. Optimisation of an appropriate gatekeeping strategy that leverages locally available genetics capacity may increase our diagnostic yield and minimise inappropriate testing.

Data availability statement

The data presented in the study are deposited into ClinVar, Organisation ID 508172 and can be accessed at https://www.ncbi.nlm.nih.gov/clinvar/submitters/508172/.

Ethics statement

The studies involving humans were approved by the University of the Human Research Ethical Committee (Medical) at the University of the Witwatersrand, Johannesburg, South Africa. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

LC: Data curation, Formal Analysis, Investigation, Methodology, Project administration, Writing–original draft, Writing–review and editing. JF: Conceptualization, Investigation, Project administration, Writing–review and editing. KM: Investigation, Project administration, Writing–review and editing. PM: Investigation, Project administration, Writing–review and editing. AC: Conceptualization, Project administration, Writing–review and editing. PC: Conceptualization, Project administration, Writing–review and editing. BD: Investigation, Writing–review and editing. AG: Investigation, Writing–review and editing. DL: Investigation, Writing–review and editing. LN: Investigation, Writing–review and editing. NM: Investigation, Writing–review and editing. DM: Investigation, Writing–review

and editing. MN: Investigation, Writing–review and editing. AK: Conceptualization, Data curation, Formal Analysis, Investigation, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing. NC: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by a South African National Research Foundation Thuthuka Grant (Grant Number: 1903204240019). LC was supported by a National Research Foundation Block Grant.

Acknowledgments

We acknowledge and thank the participants and their families and caregivers; the staff of the neonatal and paediatric intensive care units and wards at the recruiting hospitals, Rahima Moosa Mother and Child Hospital and Nelson Mandela Children's Hospital; and the medical geneticists and genetic counselling interns of the Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1277948/full#supplementary-material

References

Auer-Grumbach, M., Olschewski, A., Papic, L., Kremer, H., Mcentagart, M. E., Uhrig, S., et al. (2010). Alterations in the ankyrin domain of TRPV4 cause congenital distal SMA, scapuloperoneal SMA and HMSN2C. *Nat. Genet.* 42, 160–164. doi:10.1038/ng.508

Baine-Savanhu, F., Macaulay, S., Louw, N., Bollweg, A., Flynn, K., Molatoli, M., et al. (2023). Identifying the genetic causes of developmental disorders and intellectual disability in Africa: a systematic literature review. *Front. Genet.* 14, 1137922. doi:10.3389/fgene.2023.1137922

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

Beck, K., Chan, V. C., Shenoy, N., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (2000). Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4273–4278. doi:10.1073/pnas.070050097

Beltcheva, O., Martin, P., Lenkkeri, U., and Tryggvason, K. (2001). Mutation spectrum in the nephrin gene (NPHS1) in congenital nephrotic syndrome. *Hum. Mutat.* 17, 368–373. doi:10.1002/humu.1111

Biancalana, V., Caron, O., Gallati, S., Baas, F., Kress, W., Novelli, G., et al. (2003). Characterisation of mutations in 77 patients with X-linked myotubular myopathy, including a family with a very mild phenotype. *Hum. Genet.* 112, 135–142. doi:10.1007/s00439-002-0869-1

Biasini, F., Portaro, S., Mazzeo, A., Vita, G., Fabrizi, G. M., Taioli, F., et al. (2016). TRPV4 related scapuloperoneal spinal muscular atrophy: report of an Italian family and review of the literature. *Neuromuscul. Disord.* 26, 312–315. doi:10.1016/j.nmd.2016.02.010

Ceyhan-Birsoy, O., Machini, K., Lebo, M. S., Yu, T. W., Agrawal, P. B., Parad, R. B., et al. (2017). A curated gene list for reporting results of newborn genomic sequencing. *Genet. Med.* 19, 809–818. doi:10.1038/gim.2016.193

Choi, J. H., Oh, M. Y., Yum, M. S., Lee, B. H., Kim, G. H., and Yoo, H. W. (2015). Moyamoya syndrome in a patient with Noonan-like syndrome with loose anagen hair. *Pediatr. Neurol.* 52, 352–355. doi:10.1016/j.pediatrneurol.2014.11.017

Chou, A. F., Duncan, A. R., Hallford, G., Kelley, D. M., and Dean, L. W. (2021). Barriers and strategies to integrate medical genetics and primary care in underserved populations: a scoping review. *J. Community Genet.* 12, 291–309. doi:10.1007/s12687-021-00508-5

Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., et al. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med.* 3, 16. doi:10.1038/s41525-018-0053-8

Cole, W. G., Hall, R. K., and Rogers, J. G. (1993). The clinical features of spondyloepiphyseal dysplasia congenita resulting from the substitution of glycine 997 by serine in the alpha 1(II) chain of type II collagen. *J. Med. Genet.* 30, 27–35. doi:10.1136/jmg.30.1.27

Cordeddu, V., Di Schiavi, E., Pennacchio, L. A., Ma'ayan, A., Sarkozy, A., Fodale, V., et al. (2009). Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nat. Genet.* 41, 1022–1026. doi:10.1038/ng.425

Deng, H. X., Klein, C. J., Yan, J., Shi, Y., Wu, Y., Fecto, F., et al. (2010). Scapuloperoneal spinal muscular atrophy and CMT2C are allelic disorders caused by alterations in TRPV4. *Nat. Genet.* 42, 165–169. doi:10.1038/ng.509

Dragojlovic, N., Borle, K., Kopac, N., Ellis, U., Birch, P., Adam, S., et al. (2020). The composition and capacity of the clinical genetics workforce in high-income countries: a scoping review. *Genet. Med.* 22, 1437–1449. doi:10.1038/s41436-020-0825-2

Echaniz-Laguna, A., Dubourg, O., Carlier, P., Carlier, R. Y., Sabouraud, P., Péréon, Y., et al. (2014). Phenotypic spectrum and incidence of TRPV4 mutations in patients with inherited axonal neuropathy. *Neurology* 82, 1919–1926. doi:10.1212/WNL.0000000000000450

Farnaes, L., Hildreth, A., Sweeney, N. M., Clark, M. M., Chowdhury, S., Nahas, S., et al. (2018). Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *NPJ Genom Med.* 3, 10. doi:10.1038/s41525-018-0049-4

Fecto, F., Shi, Y., Huda, R., Martina, M., Siddique, T., and Deng, H. X. (2011). Mutant TRPV4-mediated toxicity is linked to increased constitutive function in axonal neuropathies. *J. Biol. Chem.* 286, 17281–17291. doi:10.1074/jbc.M111.237685

Feng, B. J. (2017). PERCH: a unified framework for disease gene prioritization. *Hum. Mutat.* 38, 243–251. doi:10.1002/humu.23158

Fleming, J., and Quan, D. (2016). A case of congenital spinal muscular atrophy with pain due to a mutation in TRPV4. *Neuromuscul. Disord.* 26, 841–843. doi:10.1016/j.nmd.2016.09.013

French, C. E., Delon, I., Dolling, H., Sanchis-Juan, A., Shamardina, O., Mégy, K., et al. (2019). Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med.* 45, 627–636. doi:10.1007/s00134-019-05552-x

Gargano, G., Guidotti, I., Balestri, E., Vagnarelli, F., Rosato, S., Comitini, G., et al. (2014). Hydrops fetalis in a preterm newborn heterozygous for the c.4A>G SHOC2 mutation. *Am. J. Med. Genet. A* 164A, 1015–1020. doi:10.1002/ajmg.a.36376

Gripp, K. W., Zand, D. J., Demmer, L., Anderson, C. E., Dobyns, W. B., Zackai, E. H., et al. (2013). Expanding the SHOC2 mutation associated phenotype of Noonan syndrome with loose anagen hair: structural brain anomalies and myelofibrosis. *Am. J. Med. Genet. A* 161A, 2420–2430. doi:10.1002/ajmg.a.36098

Gromand, M., Gueguen, P., Pervillé, A., Ferroul, F., Morel, G., Harouna, A., et al. (2022). STAC3 related congenital myopathy: a case series of seven Comorian patients. *Eur. J. Med. Genet.* 65, 104598. doi:10.1016/j.ejmg.2022.104598

Gubbels, C. S., VanNoy, G. E., Madden, J. A., Copenheaver, D., Yang, S., Wojcik, M. H., et al. (2020). Prospective, phenotype-driven selection of critically ill neonates for rapid exome sequencing is associated with high diagnostic yield. *Genet. Med.* 22, 736–744. doi:10.1038/s41436-019-0708-6

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., et al. (2019). The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inf.* 95, 103208. doi:10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inf.* 42, 377–381. doi:10.1016/j.jbi.2008.08.010

Hichri, H., Rendu, J., Monnier, N., Coutton, C., Dorseuil, O., Poussou, R. V., et al. (2011). From Lowe syndrome to Dent disease: correlations between mutations of the OCRL1 gene and clinical and biochemical phenotypes. *Hum. Mutat.* 32, 379–388. doi:10.1002/humu.21391

Hoban, R., Roberts, A. E., Demmer, L., Jethva, R., and Shephard, B. (2012). Noonan syndrome due to a SHOC2 mutation presenting with fetal distress and fatal hypertrophic cardiomyopathy in a premature infant. *Am. J. Med. Genet. A* 158A, 1411–1413. doi:10.1002/ajmg.a.35318

Horstick, E. J., Linsley, J. W., Dowling, J. J., Hauser, M. A., McDonald, K. K., Ashley-Koch, A., et al. (2013). Stac3 is a component of the excitation-contraction coupling machinery and mutated in Native American myopathy. *Nat. Commun.* 4, 1952. doi:10.1038/ncomms2952

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., et al. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* 99, 877–885. doi:10.1016/j.ajhg.2016.08.016

Jedrzejowska, M., Dębek, E., Kowalczyk, B., Halat, P., Kostera-Pruszczyk, A., Ciara, E., et al. (2019). The remarkable phenotypic variability of the p.Arg269HiS variant in the TRPV4 gene. *Muscle Nerve* 59, 129–133. doi:10.1002/mus.26346

Jongeneel, C. V., Kotze, M. J., Bhaw-Luximon, A., Fadlelmola, F. M., Fakim, Y. J., Hamdi, Y., et al. (2022). A view on genomic medicine activities in Africa: implications for policy. *Front. Genet.* 13, 769919. doi:10.3389/fgene.2022.769919

Kamp, M., Krause, A., and Ramsay, M. (2021). Has translational genomics come of age in Africa? *Hum. Mol. Genet.* 30, R164–R173. doi:10.1093/hmg/ddab180

Kingsmore, S. F., Cakici, J. A., Clark, M. M., Gaughran, M., Feddock, M., Batalov, S., et al. (2019). A randomized, controlled trial of the analytic and diagnostic performance of singleton and trio, rapid genome and exome sequencing in ill infants. *Am. J. Hum. Genet.* 105, 719–733. doi:10.1016/j.ajhg.2019.08.009

Kingsmore, S. F., and Cole, F. S. (2022). The role of genome sequencing in neonatal intensive care units. *Annu. Rev. Genomics Hum. Genet.* 23, 427–448. doi:10.1146/annurev-genom-120921-103442

Kingsmore, S. F., Henderson, A., Owen, M. J., Clark, M. M., Hansen, C., Dimmock, D., et al. (2020). Measurement of genetic diseases as a cause of mortality in infants receiving whole genome sequencing. *NPJ Genom Med.* 5, 49. doi:10.1038/s41525-020-00155-8

Klein, C. J., Shi, Y., Fecto, F., Donaghy, M., Nicholson, G., McEntagart, M. E., et al. (2011). TRPV4 mutations and cytotoxic hypercalcemia in axonal Charcot-Marie-Tooth neuropathies. *Neurology* 76, 887–894. doi:10.1212/WNL.0b013e31820f2de3

Krause, A., Seymour, H., and Ramsay, M. (2018). Common and founder mutations for monogenic traits in sub-saharan african populations. *Annu. Rev. Genomics Hum. Genet.* 19, 149–175. doi:10.1146/annurev-genom-083117-021256

Kromberg, J. G., Sizer, E. B., and Christianson, A. L. (2013a). Genetic services and testing in South Africa. *J. Community Genet.* 4, 413–423. doi:10.1007/s12687-012-0101-5

Kromberg, J. G., Wessels, T. M., and Krause, A. (2013b). Roles of genetic counselors in South Africa. *J. Genet. Couns.* 22, 753–761. doi:10.1007/s10897-013-9606-2

Landoure, G., Zdebik, A. A., Martinez, T. L., Burnett, B. G., Stanescu, H. C., Inada, H., et al. (2010). Mutations in TRPV4 cause Charcot-Marie-Tooth disease type 2C. *Nat. Genet.* 42, 170–174. doi:10.1038/ng.512

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067. doi:10.1093/nar/gkx1153

Laporte, J., Guiraud-Chaumeil, C., Vincent, M. C., Mandel, J. L., Tanner, S. M., Liechti-Gallati, S., et al. (1997). Mutations in the MTM1 gene implicated in X-linked myotubular myopathy. ENMC international consortium on myotubular myopathy. European neuromuscular center. *Hum. Mol. Genet.* 6, 1505–1511. doi:10.1093/hmg/6.9.1505

Li, C., Zhi, D., Wang, K., and Liu, X. (2022). MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. *Genome Med.* 14, 115. doi:10.1186/s13073-022-01120-z

Linsley, J. W., Hsu, I. U., Groom, L., Yarotskyy, V., Lavorato, M., Horstick, E. J., et al. (2017). Congenital myopathy results from misregulation of a muscle Ca2+ channel by mutant Stac3. *Proc. Natl. Acad. Sci. U. S. A.* 114, E228–E236. doi:10.1073/pnas.1619238114

Longo, G., Russo, S., Novelli, G., Sangiuolo, F., and D'Apice, M. R. (2016). Mutation spectrum of the MTM1 gene in XLMTM patients: 10 years of experience in prenatal and postnatal diagnosis. *Clin. Genet.* 89, 93–98. doi:10.1111/cge.12674

Louis, E. D., Hernandez, N., Ottman, R., Ionita-Laza, I., and Clark, L. N. (2014). Essential tremor in a charcot-marie-tooth type 2C kindred does not segregate with the TRPV4 R269H mutation. *Case Rep. Neurol.* 6, 1–6. doi:10.1159/000357665

Lumaka, A., Carstens, N., Devriendt, K., Krause, A., Kulohoma, B., Kumuthini, J., et al. (2022). Increasing African genomic data generation and sharing to resolve rare and undiagnosed diseases in Africa: a call-to-action by the H3Africa rare diseases working group. *Orphanet J. Rare Dis.* 17, 230. doi:10.1186/s13023-022-02391-w

Lunke, S., Eggers, S., Wilson, M., Patel, C., Barnett, C. P., Pinner, J., et al. (2020). Feasibility of ultra-rapid exome sequencing in critically ill infants and children with suspected monogenic conditions in the australian public health care system. *JAMA* 323, 2503–2511. doi:10.1001/jama.2020.7671

Mann, N., Braun, D. A., Amann, K., Tan, W., Shril, S., Connaughton, D. M., et al. (2019). Whole-exome sequencing enables a precision medicine approach for kidney transplant recipients. *J. Am. Soc. Nephrol.* 30, 201–215. doi:10.1681/ASN.2018060575

Mclaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl variant effect predictor. *Genome Biol.* 17, 122. doi:10.1186/s13059-016-0974-4

Meng, L., Pammi, M., Saronwala, A., Magoulas, P., Ghazi, A. R., Vetrini, F., et al. (2017). Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. *JAMA Pediatr.* 171, e173438. doi:10.1001/jamapediatrics.2017.3438

Mhlongu, F., Bhengu, L., Dillon, B., Essop, F., and Krause, A. (2022). *STAC3 disorder: frequent genetic cause of congenital hypotonia in African patients*. Johannesburg, South Africa: University of the Witwatersrand Faculty of Health Sciences Research Day.

Milko, L. V., O'Daniel, J. M., DeCristo, D. M., Crowley, S. B., Foreman, A. K. M., Wallace, K. E., et al. (2019). An age-based framework for evaluating genome-scale sequencing results in newborn screening. *J. Pediatr.* 209, 68–76. doi:10.1016/j.jpeds.2018.12.027

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 Res.* 16, 1215. doi:10.1093/nar/16.3.1215

Nishimura, G., Haga, N., Kitoh, H., Tanaka, Y., Sonoda, T., Kitamura, M., et al. (2005). The phenotypic spectrum of COL2A1 mutations. *Hum. Mutat.* 26, 36–43. doi:10.1002/humu.20179

Petrikin, J. E., Willig, L. K., Smith, L. D., and Kingsmore, S. F. (2015). Rapid whole genome sequencing and precision neonatology. *Semin. Perinatol.* 39, 623–631. doi:10.1053/j.semperi.2015.09.009

Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H., and Inoue, I. (2012). Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J. Hum. Genet.* 57, 621–632. doi:10.1038/jhg.2012.91

Ravenscroft, G., Clayton, J. S., Faiz, F., Sivadorai, P., Milnes, D., Cincotta, R., et al. (2021). Neurogenetic fetal akinesia and arthrogryposis: genetics, expanding genotype-phenotypes and functional genomics. *J. Med. Genet.* 58, 609–618. doi:10.1136/jmedgenet-2020-106901

Recker, F., Zaniew, M., Böckenhauer, D., Miglietti, N., Bökenkamp, A., Moczulska, A., et al. (2015). Characterization of 28 novel patients expands the mutational and phenotypic spectrum of Lowe syndrome. *Pediatr. Nephrol.* 30, 931–943. doi:10.1007/s00467-014-3013-2

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47, D886–D894. doi:10.1093/nar/gky1016

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., 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. *Genet. Med.* 17, 405–424. doi:10.1038/gim.2015.30

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi:10.1038/nbt.1754

Saleh, S., Beyyumi, E., Al Kaabi, A., Hertecant, J., Barakat, D., Al Dhaheri, N. S., et al. (2021). Spectrum of neuro-genetic disorders in the United Arab Emirates national population. *Clin. Genet.* 100, 573–600. doi:10.1111/cge.14044

Schobers, G., Schieving, J. H., Yntema, H. G., Pennings, M., Pfundt, R., Derks, R., et al. (2022). Reanalysis of exome negative patients with rare disease: a pragmatic workflow for diagnostic applications. *Genome Med.* 14, 66. doi:10.1186/s13073-022-01069-z

Schoeb, D. S., Chernin, G., Heeringa, S. F., Matejas, V., Held, S., Vega-Warner, V., et al. (2010). Nineteen novel NPHS1 mutations in a worldwide cohort of patients with congenital nephrotic syndrome (CNS). *Nephrol. Dial. Transpl.* 25, 2970–2976. doi:10.1093/ndt/gfq088

Schoonen, M., Smuts, I., Louw, R., Elson, J. L., van Dyk, E., Jonck, L. M., et al. (2019). Panel-based nuclear and mitochondrial next-generation sequencing outcomes of an ethnically diverse pediatric patient cohort with mitochondrial disease. *J. Mol. Diagn* 21, 503–513. doi:10.1016/j.jmoldx.2019.02.002

Smith, L. D., Willig, L. K., and Kingsmore, S. F. (2015). Whole-exome sequencing and whole-genome sequencing in critically ill neonates suspected to have single-gene disorders. *Cold Spring Harb. Perspect. Med.* 6, a023168. doi:10.1101/cshperspecta023168

Splinter, K., Adams, D. R., Bacino, C. A., Bellen, H. J., Bernstein, J. A., Cheatle-Jarvela, A. M., et al. (2018). Effect of genetic diagnosis on patients with previously undiagnosed disease. *N. Engl. J. Med.* 379, 2131–2139. doi:10.1056/NEJMoa1714458

Stark, Z., and Ellard, S. (2022). Rapid genomic testing for critically ill children: time to become standard of care? *Eur. J. Hum. Genet.* 30, 142–149. doi:10.1038/s41431-021-00990-y

Stark, Z., Tan, T. Y., Chong, B., Brett, G. R., Yap, P., Walsh, M., et al. (2016). A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet. Med.* 18, 1090–1096. doi:10.1038/gim.2016.1

Stats, S. A. (2022). *General household survey 2021*. South Africa: Pretoria.

Takahashi, N., Hamada-Nakahara, S., Itoh, Y., Takemura, K., Shimada, A., Ueda, Y., et al. (2014). TRPV4 channel activity is modulated by direct interaction of the ankyrin domain to PI(4,5)P2. *Nat. Commun.* 5, 4994. doi:10.1038/ncomms5994

Tanner, S. M., Schneider, V., Thomas, N. S., Clarke, A., Lazarou, L., and Liechti-Gallati, S. (1999). Characterization of 34 novel and six known MTM1 gene mutations in 47 unrelated X-linked myotubular myopathy patients. *Neuromuscul. Disord.* 9, 41–49. doi:10.1016/s0960-8966(98)00090-x

Telegrafi, A., Webb, B. D., Robbins, S. M., Speck-Martins, C. E., FitzPatrick, D., Fleming, L., et al. (2017). Identification of STAC3 variants in non-Native American families with overlapping features of Carey-Fineman-Ziter syndrome and Moebius syndrome. *Am. J. Med. Genet. A* 173, 2763–2771. doi:10.1002/ajmg.a.38375

Terhal, P. A., Nievelstein, R. J. A. J., Verver, E. J. J., Topsakal, V., van Dommelen, P., Hoornaert, K., et al. (2015). A study of the clinical and radiological features in a cohort of 93 patients with a COL2A1 mutation causing spondyloepiphyseal dysplasia congenita or a related phenotype. *Am. J. Med. Genet. A* 167A, 461–475. doi:10.1002/ajmg.a.36922

Thormann, A., Halachev, M., McLaren, W., Moore, D. J., Svinti, V., Campbell, A., et al. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat. Commun.* 10, 2373. doi:10.1038/s41467-019-10016-3

Tiffin, N. (2014). Unique considerations for advancing genomic medicine in African populations. *Per Med.* 11, 187–196. doi:10.2217/pme.13.105

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., Weiss, M. E., Köster, J., Marais, A., et al. (2017). Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur. J. Hum. Genet.* 25, 176–182. doi:10.1038/ejhg.2016.146

Van Diemen, C. C., Kerstjens-Frederikse, W. S., Bergman, K. A., de Koning, T. J., Sikkema-Raddatz, B., van der Velde, J. K., et al. (2017). Rapid targeted genomics in critically ill newborns. *Pediatrics* 140, e20162854. doi:10.1542/peds.2016-2854

Waldrop, M. A., Boue, D. R., Sites, E., Flanigan, K. M., and Shell, R. (2017). Clinicopathologic conference: a newborn with hypotonia, cleft palate, micrognathia, and bilateral club feet. *Pediatr. Neurol.* 74, 11–14. doi:10.1016/j.pediatrneurol.2017.01.026

Wang, H., Qian, Y., Lu, Y., Qin, Q., Lu, G., Cheng, G., et al. (2020). Clinical utility of 24-h rapid trio-exome sequencing for critically ill infants. *NPJ Genom Med.* 5, 20. doi:10.1038/s41525-020-0129-0

Warejko, J. K., Tan, W., Daga, A., Schapiro, D., Lawson, J. A., Shril, S., et al. (2018). Whole exome sequencing of patients with steroid-resistant nephrotic syndrome. *Clin. J. Am. Soc. Nephrol.* 13, 53–62. doi:10.2215/CJN.04120417

Wilkinson, D. J., Barnett, C., Savulescu, J., and Newson, A. J. (2016). Genomic intensive care: should we perform genome testing in critically ill newborns? *Arch. Dis. Child. Fetal Neonatal Ed.* 101, F94–F98. doi:10.1136/archdischild-2015-308568

Willig, L. K., Petrikin, J. E., Smith, L. D., Saunders, C. J., Thiffault, I., Miller, N. A., et al. (2015). Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir. Med.* 3, 377–387. doi:10.1016/S2213-2600(15)00139-3

Wright, C. F., Campbell, P., Eberhardt, R. Y., Aitken, S., Perrett, D., Brent, S., et al. (2023). Genomic diagnosis of rare pediatric disease in the United Kingdom and Ireland. *N. Engl. J. Med.* 388, 1559–1571. doi:10.1056/NEJMoa2209046

Wright, C. F., Fitzpatrick, D. R., and Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nat. Rev. Genet.* 19, 325. doi:10.1038/nrg.2018.12

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312, 1870–1879. doi:10.1001/jama.2014.14601

Zaharieva, I. T., Sarkozy, A., Munot, P., Manzur, A., O'Grady, G., Rendu, J., et al. (2018). STAC3 variants cause a congenital myopathy with distinctive dysmorphic features and malignant hyperthermia susceptibility. *Hum. Mutat.* 39, 1980–1994. doi:10.1002/humu.23635

Zimon, M., Baets, J., Auer-Grumbach, M., Berciano, J., Garcia, A., Lopez-Laso, E., et al. (2010). Dominant mutations in the cation channel gene transient receptor potential vanilloid 4 cause an unusual spectrum of neuropathies. *Brain* 133, 1798–1809. doi:10.1093/brain/awq109

Appendix J: Plagiarism Declaration

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I LISA CAMPBELL (Student number: 808632) am a student registered for the degree of DOCTOR OF PHILOSOPHY in the academic year 2024.

I hereby declare the following:

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

Signature: L Campbell Date: 14 AUGUST 2024

Appendix K: Turnitin Report

Lisa_Thesis_250324_Turnitin.docx

ORIGINALITY REPORT

11%	8%	9%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.frontiersin.org Internet Source	4%
2	Claudia Azuelos, Marc-Antoine Marquis, Anne-Marie Laberge. "A systematic review of the assessment of the clinical utility of genomic sequencing: Implications of the lack of standard definitions and measures of clinical utility", European Journal of Medical Genetics, 2024 Publication	1%
3	L. Campbell, J. Fredericks, K. Mathivha, P. Moshesh et al. "The implementation and utility of clinical exome sequencing in a South African infant cohort", Frontiers in Genetics, 2023 Publication	1%
4	"Abstracts from the 52nd European Society of Human Genetics (ESHG) Conference: Posters", European Journal of Human Genetics, 2019 Publication	<1%