

https://dspace.org/

The role of small genetic variants in the aetiology of developmental disorders in South Africa - a whole exome sequencing study

Molatoli, Mhlekazi Cathrine

2024

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/43295

Molatoli, Mhlekazi Cathrine. (2024). The role of small genetic variants in the aetiology of developmental disorders in South Africa - a whole exome sequencing study [PhD thesis, University of the Witwatersrand, Johannesburg]. WireDSpace.

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

THE ROLE OF SMALL GENETIC VARIANTS IN THE AETIOLOGY OF DEVELOPMENTAL DISORDERS IN SOUTH AFRICA - A WHOLE EXOME SEQUENCING STUDY

Mhlekazi Cathrine Molatoli

Student number: 1095937



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Supervised by

Associate Professor Zané Lombard

Dr Nadia Carstens

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

Declaration

I, Mhlekazi Cathrine Molatoli, declare that this Thesis is my 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.

_____ _____

Mhlekazi Cathrine Molatoli

on 31 March 2023 in Johannesburg

Dedication

This thesis is in loving memory of my dearest, mother Mrs ‘Mamolatoli Molatoli, and my uncles Keke Matsepe and Bishop Molatoli. To say I miss you every day is an understatement.

To my siblings, Thato Matsepe (you get it) and Maqaleng Molatoli, you two are not only my given family but also my chosen best friends. I don’t know how I would have survived this last year without your support and late-night calls.

To my father, thank you “Wama”, for all the wise words, even when it was hard. Above all to my grandmother, Ester ‘Matsiu Matsepe, the superhero of our family, I can never thank you enough for keeping me focused during this journey when everything around us was falling apart. No words can sum up all you have done for me. This is your thesis as much as it is mine.

Presentations from this study

Molatoli M, Mpangase P, Carstens N and Lombard Z. The evaluation of ClinVar pathogenic variants in patient with developmental disorders in South Africa. Oral presentation at SASHG Young Researchers Symposium –2021, 25th August.

Molatoli M, Mpangase P, Carstens N and Lombard Z. The evaluation of ClinVar pathogenic variants in patient with developmental disorders in South Africa. Oral presentation at 18th H3Africa-consortium meeting–2021, 1st September.

Molatoli M, Mpangase P, Carstens N and Lombard Z. Deciphering Developmental Disorders in Africa study – A genotype- driven variant prioritization pipeline in an African setting. Poster presentation at Wits Faculty of Health Sciences Research Day–2022, 15th September.

Abstract

Developmental disorders (DD) are a diverse group of chronic conditions characterized by significant limitations to both mental and physical development. Genetic variants have been identified as the underlying aetiology in about 40-50% of DD cases. Whole exome sequencing (WES) is the recommended first-line genetic test in this group of patients and is associated with diagnostic yields of 16-45%. However, in South Africa and other resource-poor settings, karyotype testing and MLPA analysis (offering low diagnostic rates of 3% and ~9% respectively) are still being utilized for genetic testing. Thus, a higher proportion of patients remain with unexplained DD due to the limitations of these diagnostic tools and limited genetic services. The main challenge facing the clinical implementation of WES in African settings is the complex data analysis and interpretation associated with the large amount of variant data produced. This is especially challenging as African ancestry individuals have been demonstrated to have a high level of genetic diversity resulting in a higher number of novel variants reported compared to European ancestry individuals. This study seeks to investigate whether the clinical utility of WES can be replicated in an African setting. Additionally, we seek to make recommendations for variant filtering and prioritization, thus making the process of WES data analysis for DD patients more efficient. To achieve these, WES was performed in 117 patients with unexplained DD and their 180 unrelated parents. Variant data was filtered and prioritized using two in-house semi-automated pipelines. The first pipeline, prioritized variants overlapping known DD genes, as identified using the G2P-DDG2P bioinformatics analysis tool. The second pipeline identified *de novo* variants in trio families using the trio-dnm bioinformatics analysis tool. Sanger sequencing was used to validate low-quality prioritized variants prior to in-house interpretation and curation, and all subsequently identified putative disease-causing variants prior to reporting. Of the 117 patients from 115 families analysed, a positive molecular diagnosis was achieved for 29 families, resulting in a diagnostic yield of 25.2% (29/115). Leveraging currently available DD data, our findings demonstrate the diagnostic and clinical utility of WES which resulted in recommendations for improving patient clinical management and surveillance. This study has also developed and made recommendations for variant filtering and prioritization strategies, which can be implemented in both research and diagnostic settings to streamline and aid in the identification of putative disease-causing variants in DD patients.

Acknowledgments

I would like to express my gratitude to the following people and institutional bodies who have contributed towards the completion of my PhD:

- My supervisors, Associate Professor Zané Lombard, and Dr Nadia Carstens thank you for the support and guidance.
- Dr Phelalani Mpangase, thank you for all your support and for answering all my data and bioinformatics questions.
- Dr Lindie Lamola and Dr Thandiswa Ngcungcu for all the wet lab assistance and talking me off the ledge when I was ready to hang it all up.

Funders:

- H3Africa NIMH/NIH: U01MH115483
- National Health Laboratory Service Research Trust (NHLSRT):94785
- Faculty of Health Science Research Council (FRC)
- The National Manpower and Development Secretariat (NMDS) Lesotho



Table of Contents

Chapter 1 - Introduction 1

1.1 Developmental disorders – definition and prevalence 1

1.2 The aetiology of DD is heterogenous in nature 1

1.2.1. Environmental factors contributing to DD 2

1.2.2. The main genetic causes of DD 3

1.3 Delineating the genetic aetiology of DD is challenging 5

1.4 Traditional genetic testing for DD 6

1.5 Genetic testing in resource limited settings 9

1.6 An overview of Next generation sequencing 9

1.7 NGS data analysis and variant calling workflows 11

1.7.1. Primary data analysis 12

1.7.2. Secondary data analysis 12

1.8 Variant analysis and interpretation 15

1.8.1. Variant annotation 15

1.8.2. Variant interpretation using the ACMG-AMP guidelines 15

1.8.3. WES is the current recommended first-line test in DD patients 21

1.9 Challenges to diagnostic implementation of NGS 23

1.10 Current initiatives to increase African genomic research 27

Deciphering Developmental Disorders (DDD) 29

1.11 DDD-UK 29

1.12.1. DDD-Africa 30

1.12 Study Rationale 30

1.13 Aim and Objectives of the current study 32

Chapter 2 - Methodology 33

2.1 Ethics and consent 33

2.2 Patient cohort 33

2.3 Patient clinical information 34

2.4 Sample collection and processing 34

2.5 WES and variant calling at the WTSI 35

2.6 Variant annotation 35

2.7 Variant filtering and prioritization workflow 36

2.7.1. DDG2P variant prioritization strategy 39

2.7.2. De novo prioritization pipeline 41

2.8 Variant validation using Sanger sequencing 44

2.9 Variant reporting 46

Chapter 3 - Results 47

3.1 Patient demographics and details 47

3.2 Clinical features of patients and indications for study inclusion 50

3.3 WES and data quality measurements 52

3.4 Variant annotation identifies a high proportion of known variants 53

3.5 African dataset indicates a high percentage of very rare variants 54

3.6 Filtering and prioritization of variants in known DD genes 55

3.6.1. Variant quality control and interpretation of variants using ACMG-AMP guidelines 56

3.7 Low-quality variants in plausible DD genes confirmed as false-positive 58

3.8 Identification and manual curation of candidate variants 60

3.9 Screening for *de novo* variants in DD 65

3.10 WES diagnostic yield 75

3.10.1. De novo inheritance 77
3.10.2. X-linked inheritance 77
3.10.3. Autosomal recessive inheritance 78

3.11 The association of recorded HPO terms and a positive diagnosis 78

3.12 Genotype-Phenotype evidence for disease-causing variants 81

3.12.1. D3S.0013 - *MTOR* associated *Smith-Kingsmore syndrome in an African ancestry patient* 81
3.12.2. D3S.0028 - *PPP2R1A* related neurodevelopmental disorder in an African patient 83
3.12.3. D3S.0037 - *C12orf57* first report of Temtamy syndrome in an African ancestry patient 86
3.12.4. D3S.0051 - *CHD7* variant in CHARGE syndrome 88
3.12.5. D3S.0052 - *MECP2* variant results in atypical Rett syndrome 90
3.12.6. D3S.0053 - *AUTS2 variant in a patient with severe feeding problems* 92
3.12.7. D3S.0056 - *PLP1* variant *associated with Pelizaeus-Merzbacher disease* 94
3.12.8. D3S.0064 - *SMAD4* variant is *associated with both cancer and ID* 96
3.12.9. D3S.0071 - *EMHT1* variant *Kleefstra syndrome no hypotonia* 98
3.12.10. D3S.0082 - *SON* variant in gene hotspot resulting in *ZTTK syndrome* 100
3.10.11. D3S.0086 - *STXBP1* variant associated with suppression bursts 102
3.12.12. D3S.0094 - *DDX3X* related ID with Snijders Blok type 104
3.12.13. D3S.0098 - *OFD1* variant confirms suspected clinical presentation 106
3.12.14. D3S.0111 - *PTEN* hamartoma tumor syndrome and cancer risk 109
3.12.15. D3S.0118 - *FLNA* variant associated with X-linked OPD 1 111
3.12.16. D3S.0124 - *PACS1* variant in Schuurs-Hoeijmakers syndrome 114
3.12.17. D3S.0129 - *MAPK1* variant associated with Noonan-like clinical presentation 116
3.12.18. D3S.0015 - *NFIX* variant associated with Malan syndrome 118
3.12.19. D3S.0033 - *UBE2A positive diagnosis of UBE2A ID with Nascimento* 120
3.12.20. D3S.0034 - *CHD7* variant in CHARGE syndrome 122
3.12.21. D3S.0035 - *OCRL* variant in Lowe syndrome 125
3.12.22. D3S.0047 - *ARID1B* variant associated with Coffin Siri syndrome (CSS) 127
3.12.23. D3S. 130
0050 - *CHD7* variant in CHARGE syndrome 130
3.12.24. D3S.0058 - *MFSD8* variant associated with CLN7 disease 131
3.12.25. D3S.0081 - *SCN2A* related DD without epilepsy 134
3.12.26. D3S.0089 - *MEF2C* haploinsufficiency syndrome 137
3.12.27. D3S.0108 - *XYLT2* variant in Spondyl-oocular Syndrome 138
3.12.28. D3S.0117 - *NSD1* report of a second Sotos syndrome patient of African ancestry 140
3.12.29. D3S.0119 - *RPS6KA3* variant confirms the suspected clinical diagnosis of Coffin-Lowry 143
3.12.30. Study diagnostic yield summary 145

3.13 Summary of patient recommendations based on a positive molecular diagnosis 146

3.14 Summary of key research results 147

Chapter 4 - Discussion 147

4.1 The diagnostic utility of WES in African DD patients 148

4.1.1. Leveraging known DD gene and variant data results in a positive molecular diagnosis 149

4.2 The benefits of a positive molecular diagnosis in DD patient 151

4.3 Phenotypic complexity increases the chance of a positive diagnosis 152

4.4 A high proportion of rare variants makes data interpretation challenging 153

4.5 WES results suggest a high contribution of *de novo* variants in DD 154
4.5.1. The identification of potential novel DD variants is challenging. 154
4.6 Strict variant quality control is sufficient to identify false-positive variants 156
4.6.1. Sanger sequencing may not always be required for NGS variant confirmation 156
4.7 Limitations of the study 159
4.8 Future work 159
Chapter 5 - Conclusion 161
Appendices 162
Appendix 162
Appendix 2 163
Appendix 3 164
Appendix 4 165
References 167

List of abbreviation

A: Adenine

C: Cytosine

G: Guanine

T: Thymine

ACMG: American College of Medical Genetics and Genomics

AMP: Association for Molecular Pathology

AD: Autosomal dominant

AGVP: African Genome project

AR: Autosomal recessive

BAM: Binary alignment map

CAP: College of American Pathologists

CMA: Chromosomal microarray

CNS: Central nervous system

CNV: Copy number variants

CRAM: Compressed Reference-oriented Alignment Map

DD: Developmental disorders

DDD: Deciphering developmental disorders

DDG2P: Developmental Disorders Gene to Phenotype database

DECIPHER: DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources

DNB: DNA nanoball

DP: Read Depth

FISH: Fluorescence In-Situ-hybridisation

FMR1: Fragile X messenger ribonucleoprotein1

FXS: Fragile X syndrome

GATK: Genome Analysis Toolkit

Gb: Gigabyte

GOF: Gain-of-function

HGP: human genome project

HPC: High-performance computing system

HPO: Human Phenotype Ontology

ID: Intellectual disability

IGV: Integrative Genome Viewer

Indel: Insertions and deletions

LB/B: likely benign/benign

LMIC: low-and middle-income countries

LOF: loss-of-function

MAF: Minor allele frequency
MLPA: Multiplex ligation-dependent probe amplification

MQ: Mapping quality

NDD: Neurodevelopmental disorders

NGS: Next Generation Sequencing

NHLS: National Health Laboratory Service

NIH: National Institutes of Health

OMIM Online Mendelian Inheritance in Man

QD: Quality by depth

SAM: Sequence alignment map

SB: Strand bias

SNP: Single Nucleotide Polymorphism

SNV: Single nucleotide variants

SV: Structural variants

TB: Terabytes

USD: United States Dollar

VAF: Variant allele fraction

VCF: Variant Call Format

VEP: Variant Effect Predictor

VQSR: Variant Quality Score Recalibration

VUS: Variants of unknown significance

WES: Whole exome Sequencing

WGS: Whole genome sequencing

XLD: X-linked dominant

XLR: X-linked recessive

List of Figures

Figure 1.1. The differences between NGS sequencing strategies. Whole genome sequencing (WGS).....10
Figure 1.2. Genome Analysis Toolkit (GATK) germline short variant discovery best practices workflow.....12
Figure 1.3. Evidence framework for each of the criteria for a benign or pathogenic assertion of a variant......17
Figure 1.4. The decreasing estimated cost of sequencing the whole human genome using current next-generation sequencing platforms......24
Figure 1.5. Population genome data across Africa. 28
Figure 2.1. Semi-automated DDG2P variant filtering and prioritization pipeline. 38
Figure 2.2. A semi-automated *de novo* variant filter and prioritization workflow.....43
Figure 2.3. Standardized variant consequence terms are represented according to gene location. 44
Figure 3. 1. The number of previous genetic tests undergone by patients in the DDD-Africa study.....48
Figure 3.2. Specific genetic tests undergone by DD patients prior to study enrolment. 49
Figure 3.3. The number of HPO terms recorded for patients with unexplained developmental disorders.....50
Figure 3.4. Top ten HPO terms observed among patients with unexplained developmental disorders.....51
Figure 3.5. The number of DD patients recruited categorized based on the DDD-Africa inclusion criteria. 51
Figure 3.6. Mean sequence depth for DDD-Africa dataset......52
Figure 3.7. Ts/Tv ratio for DDD-Africa dataset.....53
Figure 3.8. Allele frequency distribution of variants in the DDD-Africa dataset......54
Figure 3.9. DD2P Variant filtering and prioritization strategy results......55
Figure 3.10. DDG2P Variant filtering and prioritization results. 57
Figure 3.11. The number of *de novo* variants after each filter strategy.....66
Figure 3.12.The frequency of disease-causing genes reported in the DD cohort. d.....76
Figure 3.13. Frequency of identified putative disease-causing variant type......77
Figure 3.14. The number of HPO terms reported for diagnosed and undiagnosed patients.....79
Figure 3.15. Validation of *MTOR* variant in patient D3S.0013.....82
Figure 3.16. Validation of *PPP2R1A* variant in patient D3S.0028......85
Figure 3.17. Validation of *C12orf57* variant in patient D3S.003**7.** 87

Figure 3.18. Validation of *CHD7* variant in patient D3S.0051. 89
Figure 3.19. Validation of *MECP2* variant in patient D3S.0052..... 91
Figure 3.20. Validation of *AUTS2* variant in patient D3S0053. 93
Figure 3.21. Validation of *PLP1* variant in patient D3S.0056. 95
Figure 3.22. Validation of *SMAD4* variant in patient D3S.0064. 97
Figure 3.23. Validation of *EMHT1* variant in patient D3S.0071..... 100
Figure 3.24. Validation of *SON* variant in patient D3S.0082.. 102
Figure 3.25. Validation of *STXBP1* variant in patient D3S.0086. 104
Figure 3.26. Validation of *DDX3X* variant in patient D3S.0094. 106
Figure 3.27. Validation of *OFD1* variant in patient D3S.0094. 108
Figure 3.28. Validation of *PTEN* variant in patient D3S.0111..... 110
Figure 3.29. Validation of *FLNA* variant in patient D3S. 0118..... 113
Figure 3.30. Validation of *PACS1* variant in patient D3S.0124. 115
Figure 3.31. Validation of *MAPK1* variant in patient D3S.0129..... 117
Figure 3.32. Validation of *NFIX* variant in patient D3S.0015..... 119
Figure 3.33. Validation of *UBE2A* variant in patient D3S.0033. 121
Figure 3.34. Validation of *CHD7* variant in patient D3S.0034. 123
Figure 3.35. Validation of *OCRL* variant in patient D3S.0035**.**. 126
Figure 3.36. Validation of *ARID1B* variant in patient D3S.0047.. 129
Figure 3.37. Validation of *CHD7* variant in patient D3S.0050. 131
Figure 3.38. Validation of *MFSD8* variant in patient D3S.0058..... 133
Figure 3.39. Validation of *SCN2A* variant in patient D3S.0081..... 136
Figure 3.40. Validation of *MEF2C* variant in patient D3S.0089..... 138
Figure 3.41. Validation of *XYLT2* variant in patient D3S.0108 140
Figure 3.42. Validation of *NSD1* variant in patient D3S.0117.. 142
Figure 3.43. Validation of *RPS6KA3* variant in patient D3S.0119.. 145
Figure 3.44. Summary of key PhD research results……...147

List of Tables

Table 1.1.The benefits and limitations of traditional genetic tests. 8
Table 2.1. Patient recruitment eligibility criteria for the DDD-Africa study..... 34
Table 2.2. Ensembl-VEP annotation sources used in the study..... 37
Table 2.3. Polymerase Chain Reaction components along with their concentrations 45
Table 2.4. Polymerase Chain Reaction thermocycler conditions 45
Table 3.1. Demographic data for patients with unexplained developmental disorders. 47
Table 3.2. Summary of the number of variants identified in the DDD-Africa cohort..... 53
Table 3.3. DD Cohort polymorphic variants annotated in ClinVar as pathogenic.. 56
Table 3.4. Low-quality variants prioritised in the DDD-Africa study..... 59
Table 3.5. A summary of the putative disease-causing variants identified in the DD patient. 61
Table 3.6. *De novo* prioritized variants identified using the trio-dnm prioritization strategy. 67
Table 3.7. Diagnostic yield of WES based on mode of inheritance in the DDD-Africa study 78
Table 3.8. Suspected clinical Diagnosis of patient in DDD-Africa cohort..... 80
Table 3.9. Clinical phenotype for patient D3S.0013 described using Human Phenotype Ontology terms. 81
Table 3.12. Clinical phenotype for patient D3S.0051 described using Human Phenotype Ontology terms. 88
Table 3.13. Clinical phenotype for patient D3S.0052 described using Human Phenotype Ontology terms. 90
Table 3.14. Clinical phenotype for patient D3S.0053 described using Human Phenotype Ontology terms 92
Table 3.15. Clinical phenotype for patient D3S.0056 described using Human Phenotype Ontology terms. 94
Table 3.16. Clinical phenotype for patient D3S.0064 described using Human Phenotype Ontology terms. 96
Table 3.17. Clinical phenotype for patient D3S.0071 described using Human Phenotype Ontology terms. 99
Table 3.18. Clinical phenotype for patient D3S.0082 described using Human Phenotype Ontology terms. 101
Table 3.19. Clinical phenotype for patient D3S.0086 described using Human Phenotype Ontology terms 103
Table 3.20. Clinical phenotype for patient D3S.0081 described using Human Phenotype Ontology terms. 105

Table 3.21. Clinical phenotype for patient D3S.0098 described using Human Phenotype Ontology terms. 107
Table 3.22. Clinical phenotype for patient D3S.0111 described using Human Phenotype Ontology terms. 109
Table 3.23. Clinical phenotype for patient D3S.0118 described using Human Phenotype Ontology terms. 111
Table 3.24. Clinical phenotype for patient D3S.0124 described using Human Phenotype Ontology terms. 114
Table 3.25. Clinical phenotype for patient D3S.0129 described using Human Phenotype Ontology terms. 116
Table 3.26.Clinical phenotype for patient D3S.0015 described using Human Phenotype Ontology terms. 118
Table 3.27. Clinical phenotype for patient D3S.0033.01.1 and D3S.0033.04.1 described using Human Phenotype Ontology terms..... 120
Table 3.28. Clinical phenotype for patient D3S.0034 described using Human Phenotype Ontology terms. 122
Table 3.29. Clinical phenotype for patient D3S.0035 described using Human Phenotype Ontology terms. 125
Table 3.30. Clinical phenotype for patient D3S.0035 described using Human Phenotype Ontology terms. 128
Table 3.31. Clinical phenotype for patient D3S.0050 described using Human Phenotype Ontology terms. 130
Table 3.32. Clinical phenotype for patient D3S.0058 described using Human Phenotype Ontology terms. 133
Table 3.33. Clinical phenotype for patient D3S.0081 described using Human Phenotype Ontology terms. 135
Table 3.34. Clinical phenotype for patient D3S.0089 described using Human Phenotype Ontology terms. 137
Table 3.35. Clinical phenotype for patient D3S.00108 described using Human Phenotype Ontology terms. 139
Table 3.36. Clinical phenotype for patient D3S.0117 described using Human Phenotype Ontology terms. 141
Table 3.37. Clinical phenotype for patient D3S.0119 described using Human Phenotype Ontology terms. 144

Table 3.38. Summary of diagnosis and impact of WES results in patients with disease-causing variants. 146

Chapter 1 - Introduction

1.1 Developmental disorders – definition and prevalence

Developmental disorders (DD) are a diverse group of chronic conditions characterized by significant limitations to both mental and physical development (Boyle et al., 2011). These developmental impairments typically present at birth or early childhood and affect at least two functional domains (Moeschler et al., 2014). The most common DD is intellectual disability (ID), defined by substantial delays in cognitive functioning and adaptive behaviour, including language and memory, socio-behavioural skills, and self-care and management skills (Nuckols & Nuckols, 2013). Clinically, ID presents in isolation or as part of other comorbidities affecting different organ systems (Kochinke et al., 201). The term ID is used when referring to children above five years of age, while for those below this age, the term global developmental disorder is reserved (Moeschler et al., 2014).

The global prevalence of ID is estimated at 1-3% (Maulik et al., 2011; Flore and Milunsky, 2012; Moeschler et al., 2014), but in South Africa and other low-and middle-income countries (LMICs) prevalence remains relatively unknown and varied (Blasko et al., 2022; Mckenzie et al., 2013). In a 2002 study, Christianson et al (2002) reported an ID prevalence of 3.65% among children aged 2-9 years from rural South Africa (Christianson et al., 2002). In a more recent study, developmental milestone data analysis collected among children aged 3-4 years from 35 LMICs, found higher prevalence rates, with 36.8% of this population group failing to meet the age appropriate cognitive and socio-emotional developmental milestones (McCoy et al., 2016). This notable variation and paucity in prevalence data are in part due to varying definitions of ID, inconsistent surveillance, and data recording methods being utilized in these countries and studies (McMahon & Hatton, 2021). Moreover, there has been a large prioritization of other health concerns such as HIV/AIDS and malnutrition in these regions, with limited recognition of DD and other genetic disorders (Emerson et al., 2012). Thus, DD screening, diagnosis, and service accessibility in an already burdened healthcare system are limited (de Villiers, 2021; Marszalek & De Villiers, 2006; Sango, 2017).

1.2 The aetiology of DD is heterogenous in nature

A range of both environmental and genetic factors have been recognized as causes of DD (Bélanger & Caron, 2018), thus highlighting the high level of aetiological heterogeneity and the multifactorial nature associated with the disorder. Data that relates to the contribution levels

of each of these factors to DD is mostly available from high income countries with little documentation and data available from LMICs (Jimenez-Gomez and Standridge, 2014).

1.2.1. Environmental factors contributing to DD

Environmental or secondary exposures which include postnatal exposure to contaminants, (Dórea, 2019; Mendola et al., 2002; Shah-Kulkarni et al., 2016) malnutrition, and prenatal alcohol and drug exposure (Olivier et al., 2016) have been identified as some of the key causes of DD. It is estimated that secondary factors account for approximately 21% of all DD cases (Bélanger & Caron, 2018). However, the contribution of each of the secondary causes differs extremely between countries (Jimenez-Gomez & Standridge, 2014).

Alcohol is the largest secondary cause of DD (Lange et al., 2017; May et al., 2018; Roozen et al., 2016) and literature indicates that South Africa has the highest prevalence of Foetal Alcohol Spectrum Disorders in the world (Adebiyi & Mukumbang, 2021). In 2022 it was estimated that 29-290 per 1,000 children measured across three South African provinces had Foetal Alcohol Spectrum Disorders (May et al., 2022). This is much higher compared to the estimated 18.1 per 1,000 children reported in Canada (Popova et al., 2019) or 194.4 per 1,000 children in Australia (Fitzpatrick et al., 2017).

Infectious diseases have also been identified as another contributor to DD. This is especially relevant in Western and Southern Africa where HIV infection is still a major challenge with an estimated 20.6 million people living with the disease (UNAIDS, 2022). HIV has been found to result in neuronal damage caused by the virus infecting macrophages and other cells in the central nervous system and thus negatively impacting the development of the brain (Ellis et al., 2009; A. M. Kumar et al., 2007). Significant cognitive impairment was reported in 21 of the 44 HIV infected children compared to their unaffected controls in South Africa (Dorsey et al., 2014). The study and others (Banks et al., 2014; Ferguson & Jelsma, 2009; Potterton & Eales, 2001) have provided evidence for a causative link between DD and HIV infection. HIV positive mothers have also been shown to be at a higher risk for premature births and miscarriages compared to those uninfected (S. Beck et al., 2010; Zack et al., 2014). However, the overall percentage contribution of HIV to DD cases remains unknown.

1.2.2. The main genetic causes of DD

Genetic variants have been identified as the underlying cause in most of the patients with DD, accounting for about 40-50% of all solved DD cases (Rauch et al., 2012; Vorstman & Ophoff, 2013). Chromosomal aberrations, short nucleotide base insertions and deletions of ≤ 50 base pairs (bp) (Indels), and single nucleotide variants (SNVs) have all been recognized as major contributors to DD (Srour & Shevell, 2014; Wright et al., 2015).

- *Chromosomal abnormalities*

Chromosomal abnormalities are a consequence of errors during mitosis and meiosis cell division processes (Milani & Tadi, 2022). These chromosomal aberrations contribute to about 25% of DD cases (Miclea et al., 2015). These are classified into two broad groups: (i) large structural abnormalities or variants, due to genomic rearrangement of one or more chromosomes, resulting in abnormalities such as transversions and translocations. (ii) chromosome number abnormalities, also called aneuploidies, resulting from a missing or extra chromosome copy (Milani & Tadi, 2022). The latter constitutes the most common group of chromosomal abnormalities (Jacobs, 1987) associated with a range of DD. Downs syndrome is a consequence of an extra copy of chromosome 21 (Trisomy 21) and along with Trisomy 18 and 13 are the most frequently reported chromosomal abnormality associated with DD (Bélanger & Caron, 2018; Cottino et al., 2022; Lakhan, 2015).

Another group of chromosomal abnormalities is structural variants (SVs), which are defined as genomic rearrangements ranging from at least 50bp in size (Sudmant et al., 2015). As reviewed by (Carvalho and Lupski, 2016), approximately 7,000-25,000 polymorphic SVs have been identified and are considered a significant contributor to human genome variation and disease (Zhao et al., 2021). This group of genomic variations comprises mainly of inversions, translocations and CNVs (Sudmant et al., 2015). The latter arises from a loss or gain of genomic fragments >50bp resulting in a variable copy number in comparison to the reference genome (Alkan et al., 2011; Pös et al., 2021). Studies of healthy controls have demonstrated that a large percentage of CNVs have a neutral function or provide an evolutionary advantage (Perry et al., 2007) and are thus not associated with disease. Nonetheless, large pathogenic CNVs have been identified in a range of DD (Cooper et al., 2011; Kessi et al., 2018; C. T. Watson et al., 2014).

- *Monogenic contributors to DD*

Monogenic causes of DD account for about 10% of all cases (Miclea et al., 2015). Fragile X Syndrome (FXS) is the most common monogenic cause of DD and is responsible for about 5% of DD (Inaba et al., 2014; Miclea et al., 2015). The syndrome is a result of a Cytosine-Guanine-Guanine repeat sequence expansion at the 5'-UTR of the Fragile X Messenger Ribonucleoprotein1 (FMR1) gene (Jin & Warren, 2000; Santoro et al., 2012). The syndrome has been reported in one in 11,111 females and one in every 7,143 males(Clifford et al., 2007).

Literature evidence suggests that for the majority, ~96% of Fragile X cases, a positive molecular diagnosis is made in an individual presenting with ID and distinct clinical features suggestive of the disorder (Borch et al., 2020; Kengne Kamga et al., 2020; Weinstein et al., 2017). Thus, Fragile X syndrome testing is performed as part of the first-line of genetic investigations in such patient (Flore & Milunsky, 2012).

- *Single Nucleotide Variants and Small deletions and insertions*

The most common variant classes in the genome are SNVs, where one base is substituted for another in the genome sequence, and small insertions and deletions are less than 50bp (Indels) (Shigemizu et al., 2013). Both have been the focus of most research and the underpinning mechanism in both mendelian and non-mendelian diseases (Bromberg et al., 2013; Palamara et al., 2015).

Variants in the coding region of the protein, which alter the encoded amino acid and consequently impact the structure and function of the encoded protein have been identified as causative in DD (Hiraide et al., 2021; Järvelä et al., 2021; Trujillano et al., 2016; Wright et al., 2015). Several variant consequences have been associated with disease and these include,

i. Frameshifts, where SNVs or Indels which are not multiples of three base pairs induce a reading frame "frameshift" at the site of the nucleotide insertion or deletion, subsequently affecting gene expression (Culbertson, 1999) and resulting in a change in the encoded protein (Mullaney et al., 2010).

ii. Alternately, frameshifts can introduce a pre-mature stop codon resulting in mRNA degradation or translation of mRNA to a truncated protein which may not be functional (Mullaney et al., 2010). The severity associated with this variant is determined by its

location in the amino acid sequence, with variants near the beginning of the sequence potentially disrupting the entire protein (Rausell et al., 2014; Seligmann & Pollock, 2004).

iii. Missense is a single base nucleotide variant leading to a change in the resultant amino acid and likely affecting protein function. If the substitution is early in the sequence and the resultant codon translates into a stop codon, then the protein will not be synthesized and may lead to disease (Dabrowski, Bukowy-Bieryllo and Zietkiewicz, 2018).

iv. Variants at intronic regions or near the splice-sites may cause defects in splicing sites which can further lead to aberrant splicing and a potentially truncated protein (Lord & Baralle, 2021).

1.3 Delineating the genetic aetiology of DD is challenging

As evidenced in section 1.2 above, DD is associated with a high level of genetic heterogeneity with no one gene or variant type implicated as causative in the disorder. Additionally, this group of disorders demonstrates phenotype heterogeneity which is characterized by overlapping clinical features between the different disorders (Lee & Lupski, 2006; Tarailo-Graovac et al., 2016). For example, there is large phenotypic overlap between CHARGE syndrome and 22q11.2 deletion syndrome (Hefner & Fassi, 2017; Jyonouchi et al., 2009). Both syndromes are characterized by heart defects, hearing loss, thymic hypoplasia, and swallowing abnormalities (Hefner & Fassi, 2017), thus making diagnosis based on clinical presentation alone challenging.

Moreover, atypical clinical presentations have been noted in patient with a positive molecular diagnosis of well-defined disorders (Du Plessis et al., 2001; Mubungu et al., 2020; Valentino et al., 2021). A pathogenic splice site variant, c.4378+5G>A overlapping *NSD1* gene which is associated with Sotos syndrome was identified in a Japanese patient (Minatogawa et al., 2022). The disorder is characterized by overgrowth, moderate to profound learning disabilities, and characteristic dysmorphic features (Leventopoulos et al., 2009; Tatton-Brown et al., 2022). The authors noted a lack of overgrowth, and the presence of severe connective tissue related complications which included pes planus and mild skin hyperextensibility, all which had thus far not been associated with Sotos syndrome (Minatogawa et al., 2022). Thus, delineating the underlying aetiology of DD based on the patient's clinical phenotype alone is challenging.

This characteristic genetic and phenotypic heterogeneity associated with DD make delineating the underlying aetiology challenging thus a variety of genetic testing strategies are deployed to arrive at a positive molecular diagnosis.

1.4 Traditional genetic testing for DD

Historically, diagnostic testing for patients with unexplained DD and multiple congenital abnormalities has involved the identification of genome-wide chromosomal aneuploidies and structural rearrangements, using conventional karyotype (Vermeesch et al., 2007). This G-banding-based approach offers a detection resolution of >5-10Mb and is associated with a diagnostic yield of ~3% (excluding Down syndrome and other recognizable chromosomal syndromes in DD patients) (Miller et al., 2010; Fauzdar et al., 2013). However, one of the main limitations of the test is an inability to detect subtelomeric rearrangements (Teixeira et al., 2016; Vermeesch et al., 2007).

In the 1980's technology improvements led to the identification of subtelomeric rearrangements through the development of fluorescence in-situ hybridization (FISH) assay using probes complementary to specific genome regions (Langer-Safer, Levine and Ward, 1982; Lichter et al., 1993). This offers a higher detection resolution than that of conventional karyotype at 100-200Kb, allowing for the detection of subtelomeric rearrangements, mainly microdeletions at the ends of the chromosome (Cui et al., 2016). This has allowed for the identification and diagnosis of well-known syndromes such as Williams-Beuren syndrome caused by a microdeletion at loci q11.23 of chromosome 7 (Leme et al., 2013; Merla et al., 2010) which is otherwise not detectable by conventional karyotype (Leme et al., 2013). Given that FISH can only detect chromosomal aberrations at known target sites, combined use with conventional karyotype is associated with ~6-10% diagnostic yield in DD patients depending on the number of probes used, sample size, and patient inclusion criteria (Biesecker, 2002; Dubuc et al., 2016; Ravnan et al., 2006).

Another approach for the identification of subtelomeric rearrangements at chromosome ends is the PCR-based amplification of ligated probes using multiplex ligation-dependent probe amplification (MLPA) (Schouten et al., 2002). MLPA allows for multiplexing of samples and the use of 40-50 probes in one reaction, thus reducing the overall costs per sample and reaction preparation times compared to FISH (Schouten et al., 2002). It is reported that the overall cost of MLPA is 86% lower than that FISH based on reagent costs and overall hands-on time for sample preparation, processing, result analysis and interpretation (Stuppia et al., 2012). The use

of MLPA is associated with a diagnostic yield of ~9-30% depending on the number of probes and kits utilized, and more especially the level of clinical pre-screening performed prior to testing (Boggula et al., 2014; Koolen et al., 2004; Miclea et al., 2021; Pohovski et al., 2013).

The detection of genome-wide chromosomal rearrangements and CNVs (Batzir et al., 2015; Manning & Hudgins, 2010; D. T. Miller et al., 2010) can be achieved using chromosomal microarray (CMA). This refers to a group of microarray-based genomic copy-number analysis which includes, array-based comparative genomic hybridization (aCGH), and single nucleotide polymorphism (SNP) arrays. The diagnostic test allows for the detection of submicroscopic rearrangements and CNV of 5Kb in size, otherwise not possible using conventional cytogenetic tests such as Karyotype (Peterson et al., 2015).

CMA was previously the recommended first-line diagnostic test for DD patient following a systematic review of DD study outcomes by an international committee of DD experts. The review found that CMA was associated with a diagnostic yield of ~10-25% (D. T. Miller et al., 2010; Shin et al., 2015). Despite the diagnostic benefits of CMA, it is however, unable to reliably detect balanced rearrangements, such as translocation, insertions and inversions, and low-level genomic mosaicism (Waggoner et al., 2018). Thus, additional genetic tests are still requested to overcome the draw backs associated with CMA (D. T. Miller et al., 2010). Table 1.1 lists the main benefits and challenges of available diagnostic tests.

Table 1.1.The benefits and limitations of traditional genetic tests.

Genetic test	Utility	Benefits	Limitations	Yield
Karyotype	• Genome-wide detection of chromosomal number and structural abnormalities	• Accurate detection of aneuploidies and large structural rearrangement	• Long turnaround times (~ mean 13 days). • Detection resolution 5-10 Mb. • Unable to detect sub microscopic chromosomal rearrangements	• ~3% (excluding Down syndrome and other recognisable chromosomal syndromes
FISH	• Detection of clinically relevant chromosomal abnormalities in gene regions using fluorescently labelled probes	• Rapid turnaround times (at most +/-2 days2) • Higher target region specificity • Higher sensitivity compared to karyotype	• Less sensitivity for detecting balanced duplications. • Challenges in result interpretation as adjustment of detection cut-offs is required for different probes used.	• ~ 6-10% for FISH
MLPA	• Detection of CNVs in specific genomic regions in a single reaction using probes attached to fluorophores	• Allows for sample multiplexing. • Can detect CNVs up to 50 genomic regions. • Lower cost, ~ 86% lower than that FISH	• Balanced rearrangements (and contamination) cannot be detected. • minimum cell detection of > 25%.	• ~9-30% depending on the number of probes and kits
CMA	• Genome wide chromosomal rearrangements and CNVs > 400kb	• Detection resolution of < 3 - 10 Mb depending on design.	• Unable to reliably detect balanced rearrangements, such as translocation, insertions, and inversions. • Low-level genomic mosaicism	• ~10%-20%

Abbreviation; CMA, Chromosomal Microarray; CNV, Copy Number Variants; FISH, Fluorescence in Situ Hybridization; MLPA, multiplex ligation-dependent probe amplification

1.5 Genetic testing in resource limited settings

Despite the previously recommended use of CMA, its routine use as a first-line diagnostic test has remained limited in resource-poor settings. In such settings most of the population relies on the public health care system (Félix et al., 2023; Lumaka et al., 2022; Wiener et al., 2022) and the number and type of genetic tests available is thus mainly limited by financial constraints, available infrastructure, the logistics involved in sending samples for testing, and the policies mandating the national health care system (Félix et al., 2023; Lay-Son et al., 2015).

Conventional karyotype, and molecular cytogenetic techniques, FISH and MLPA are offered routinely as first-line diagnostic tests in poor resource settings. In 2015, it was reported that in Korea, only karyotype, FISH and MLPA testing for Prader-Willi/Angelman syndrome, DiGeorge syndrome, Miller-Dieker syndrome, and Williams syndrome was permitted under the national health insurance system (Shin et al., 2015) In a retrospective audit performed of cases received in 2017 at the division of Human Genetics (University of the Witwatersrand and National Health Laboratory Service), South Africa, the most commonly performed diagnostic tests among patients with unexplained non-syndromic DD, were karyotype and MLPA analysis. (Wiener et al., 2022). Thus, the available genetic tests in these settings are still limited.

The combined cost of karyotype and FISH or MLPA is 4-8x lower than that of CMA (Lay-Son et al., 2015). Additionally, CMA is not covered under most medical insurance policies in poor resource settings (Lay-Son et al., 2015). The high cost of CMA is compounded by the cost of the reagents and the initial costs involved in setting up CMA testing compared to MLPA which utilises instruments and equipment as such as thermocyclers, which are already in use in most of the diagnostic laboratories (Lay-Son et al., 2015). It is because of some of these challenges that MLPA and Karyotype have remained the recommended first-line cost-effective diagnostic tests in poor resource settings (Amr et al., 2020; Boggula et al., 2014; Fieggen et al., 2019).

1.6 An overview of Next generation sequencing

Despite the previous recommended use of CMA as a first-line genetic test in DD patients, around 75-90% of patients remain with unexplained DD (D. T. Miller et al., 2010; Shin et al., 2015; Tan et al., 2017). This high proportion of undiagnosed patients necessitated a shift to Next Generation Sequencing (NGS) technologies that offer simultaneous sequencing of multiple samples at improved sequencing resolution, reduced sequencing costs and shorter turnover times (Pervez et al., 2022; Stavropoulos et al., 2016; Vissers et al., 2017), which were

not offered by any of the available molecular genetic tests. In this current era of NGS, the human genome can now be sequenced using short read sequencing platforms in about 12-72 hours (J. Huang et al., 2017; Qin et al., 2021). This is compared to earlier years where the completion of the human genome project (HGP) took 15 years, 3 billion USD and Sanger sequencing efforts from 20 sequencing centres (Hood, 2008; Hood & Rowen, 2013).

There are three main sequencing approaches offered by NGS (Figure 1.1); (i) whole genome sequencing (WGS) which offers an unbiased approach, as approximately 92% of the entire genome (Nurk et al., 2021) is sequenced, (ii) whole exome sequencing (WES) which allows for the simultaneous sequencing of the exome region, which is the protein coding region of the genome and makes up about ~1-2% of the total genome (Posey, 2019), and (iii) gene-panels, which offer a more genotype-phenotype driven approach as a subset of specific genes known to be associated with disease and phenotype are examined (Metzker, 2009).

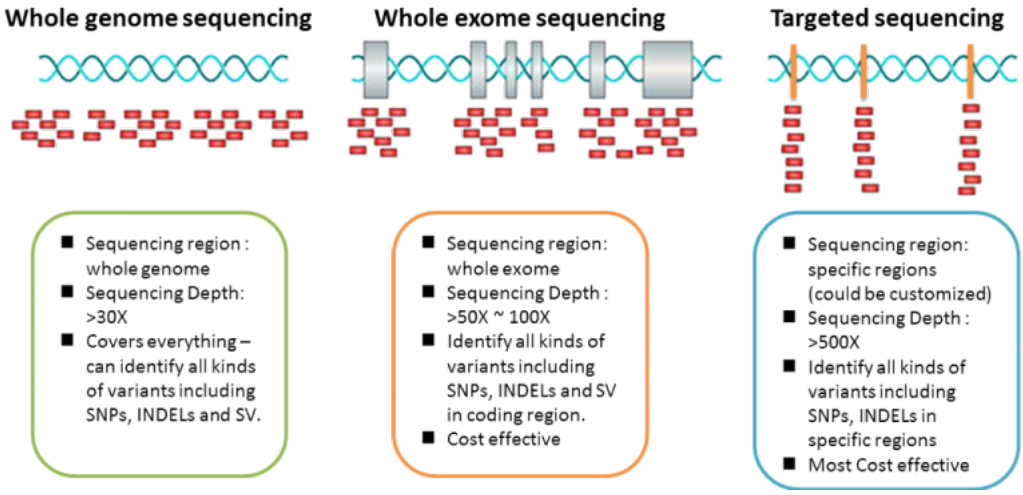

Figure 1.1. The differences between NGS sequencing strategies. Whole genome sequencing (WGS), Whole exome sequencing (WES), and Targeted sequencing using gene panels. Taken from: (http://toolsbiotech.blog.fc2.com/blog-entry-122.html?sp).

There are several different NGS platforms currently available for short read NGS, each using different sequencing chemistries. Despite this, short read sequencing across the platforms involves the clonal amplification and parallel sequencing of target DNA regions (Morey et al., 2013). Briefly, DNA is fragmented into smaller pieces, either by enzymatic or mechanical methods. This is followed by optional target enrichment aimed at capturing specific genomic regions which include exomes, or a subset of selected genes (Zhong et al., 2021). Clonal amplification of the target regions is then performed to increase the detection signal generated in the sequencing stage, which is indicative of nucleotide incorporation and a base-call (Cantu

et al., 2022; Zhong et al., 2021). The DNA fragment is then sequenced from both the forward and reverse direction, resulting in 2x150 bp or 2x100 bp paired-end reads (Jeon et al., 2021).

The sequencing chemistry is dependent on the platform utilized with differences mainly involving sequencing signal generation and detection following each nucleotide incorporation (Cantu et al., 2022; Zhong et al., 2021). Commonly used chemistries include Illumina sequencing-by-synthesis, where millions of sample DNA fragments bound to the surface of the flow cell are amplified, and reversible terminator nucleotides each bound with a characteristic fluorescent dye are incorporated one at a time into each growing DNA strand and detected during sequencing (Bentley et al., 2008).

ThermoFisher Ion Torrent uses a semiconductor to detect changes in the distinct pH levels caused by hydrogen ions released during nucleotide incorporation (Merriman et al., 2012; Rothberg et al., 2011).

BGI-MGI Tech uses DNA nanoball (DNB) sequencing technology, involving rolling circle amplification of single stranded circular DNA fragments to generate a DNA nanoball (Drmanac et al., 2010; Q. Li et al., 2019). Differential fluorescent signals are generated from uniquely fluorophore-labelled nucleotides (Heather & Chain, 2016; K. R. Kumar et al., 2019).

These signals and changes generated from the different chemistries during sequencing are measured using a camera and interpreted by a computer allowing for each nucleotide sequence in each DNA fragment to be determined (Morey et al., 2013).

1.7 NGS data analysis and variant calling workflows

The increased use of NGS in research and clinical settings has been equally met by the development of bioinformatics pipelines for analysis (Pabinger et al., 2014). There are a few workflows for variant calling available, including, DRAGEN (Goyal et al., 2017) , and Genome Analysis Toolkit (GATK) (DePristo et al., 2011; Van der Auwera et al., 2013). The use of this more streamlined workflow approach eliminates the need for selecting the best analysis tools from the large number of tools available and developing in-house parameters and thresholds for each of the analysis tools. The GATK best practices workflow is the most widely used and current gold-standard for data pre-processing and variant calling (P. J. Huang et al., 2020; Supernat et al., 2018). The toolkit makes recommendations and houses several bioinformatics tools for each data analysis step (Goyal et al., 2017).

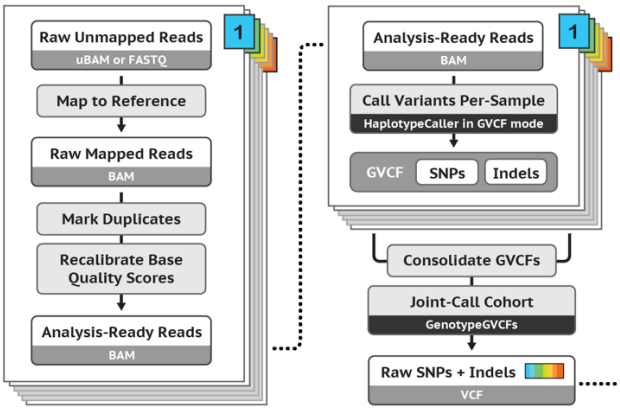

Figure 1.2. Genome Analysis Toolkit (GATK) germline short variant discovery best practices workflow. Recommended workflow for raw sequence data analysis and SNV and Indel variant calling. Taken from the GATK website (https://www.broadinstitute.org/gatk/about/. Accessed on 01/07/2022.

1.7.1. Primary data analysis

The base call data generated during sequencing is converted on-board the sequencing instrument to a FASTQ file format (Koboldt, 2020). This is a text tab-delineated file containing the raw sequence read data and associated quality scores for each incorporated nucleotide in the sequence, stored using ASCII characters (Cock et al., 2010).

The overall quality of the raw sequence reads is assessed using Phred scores, which represent the error probability that a nucleotide was incorrectly called. For example, a phred score of 30 means 1/1,000 probability of an incorrect base or a 99% accurate base call (Liao et al., 2017). The higher the Q30 phred score the more reliable the run data (Illumina, 2011). Read sequence quality relies heavily on the fluorescent signal intensity generated during base calling and, a characteristic decrease in signal intensity corresponding to poor sequence quality is observed at the 3' ends of reads (Hansen et al., 2010). Using the phred score quality thresholds, the read quality is improved by trimming these 3' sequence reads (Pereira, Oliveira and Sousa, 2020)

1.7.2. Secondary data analysis

- *Sequence read alignment*

Sequencing produces millions of short reads per DNA sample, and a challenge in data analysis is determining the correct order in which these reads occur in the human genome (Morey et al., 2013). There are several tools that have been developed to determine this order, by aligning the sequence reads to the human genome reference ((build GRCh37, hg19) or (build hg38, GRCh38)). The GATK best practices workflow recommends the use of the BWA-MEM (H. Li

& Durbin, 2009) alignment tool, which has been reported to offer a higher recall and precision rate compared to other alignment software such as BOWTIE2 (Musich et al., 2021).

Briefly, the reference sequence is fragmented, and the sequence order and location of each fragment is stored and compressed in an indexing step (Musich et al., 2021). The read sequence is then compared to the reference index and a best possible match is found within a defined margin of difference. The reference alignment and mapping information is stored in a sequence alignment map (SAM) format file.

- *Removal of duplicate reads*

Another important step in data processing is distinguishing between true sequence read and reference differences as opposed to low-quality sequence reads. Sequence reads that align or map to the same reference genome location are regarded as read duplicates. These can arise due to; (1) PCR amplification bias, which is a phenomenon that occurs when one DNA fragment is preferentially amplified more times over other and (2) random hybridisation and sequencing of the same DNA fragments to the flow cell (Ebbert et al., 2016). These pose a challenge as, the proportion of one allele may appear more frequent in comparison to others thus introducing bias and inaccurate variant calling downstream (Ebbert et al., 2016). The GATK guidelines recommend the use of Picard MarkDuplicates (reference URL can be found in Appendix 1) to remove these PCR duplicates.

- *Conversion and storage of sequencing data*

The SAM file containing the mapped reads is converted to its compressed binary version, Binary Alignment Map (BAM) format, or more preferably as a Compressed Reference-oriented Alignment Map (CRAM) file, which is around 50-70% smaller than the BAM and thus has fewer storage requirements (Bonfield 2022). The compressed binary files use less disk storage making it faster to process in subsequent analysis steps. This is usually completed using SAMtools (H. Li et al., 2009).

- *SNV and Indel variant calling*

Variant calling is the process of identifying positional differences in the genome of an individual in relation to the reference genome (Bohannan & Mitrofanova, 2019). The GATK best practices workflow for SNV and indel variant calling recommends the use of GATK HaplotypeCaller in joint calling mode (--gVCF mode) and results are outputted into a variant

calling file (VCF) file. (Poplin et al., 2017; Van der Auwera & O'Connor, 2020). Joint calling is recommended for large cohort studies, allowing for improved variant calling every time new samples from later sequencing runs are available. One benefit of joint variant calling is an increased sensitivity in calling variants at sites of low coverage. Additionally, the large genotyping data provided by joint calling allows for more accurate error models to be generated and used for the identification and filtering of possible false positive variants in subsequent steps (Poplin et al., 2017; Van der Auwera & O'Connor, 2020).

- *Increasing overall variant data quality*

In comparison to more traditional sequencing methods such as Sanger sequencing, NGS is associated with high error rates (Koboldt et al., 2010). A Variant Quality Score Recalibration (VQSR) filtering step is used in GATK to remove any false positive calls (DePristo et al., 2011). This is a variant filtering method involving the construction of an adaptive error model using annotation metrics such as mapping quality (MQ), strand bias (SB), quality by depth (QD) and variant position within reads from a set of high-confidence variants, from the HapMap (Caetano-Anolles, 2023). The model assumes that known variants that occur at high frequencies in a population are more likely to be true than novel rare variants. Using this logic, the error model is then applied to the called variants to determine the probability of a variant being true (Carson et al., 2014). Variant quality score log-odds (VQSLOD) are assigned to all variants and used to filter false positive variants. For example, variants annotated with the 99.9<sensitivity<100% VQSR have a VQSLOD score in the same range as 0.1% of the variants thought to be false positives in the variant dataset used for benchmarking. Thus, the variants are annotated as false-positive variants (Caetano-Anolles, 2023; Van der Auwera & O'Connor, 2020).

However, a few limitations of this filtering approach have been highlighted. One study showed that low quality genotypes and invalidated variants remain within the filtered dataset and are a major source of sequencing error (O'Rawe et al., 2013).

To further improve the overall quality of the dataset, additional genotype and variant filtering steps over VQSR need to be applied downstream (McKenna et al., 2010). In one study, the effectiveness of additional filtering steps following GATK VQSR was assessed (Carson et al., 2014). The authors recommended the use of variant metrics, read depth (DP) (the total number of independent reads supporting the variant) at a threshold of above 10, and genotype quality, which is the probability that a genotype call is correct expressed as a phred-scale score, (GQ)

>20. Anything below the thresholds was suggestive of sequencing bias or low coverage (Carson et al., 2014).

The effectiveness of different variant and genotype filters on the overall SNV quality can be measured using quality control parameter, transition/transversion (Ts/Tv) ratio (Wang et al., 2015). It has been calculated that the ratio of SNVs occurring in the exome region is between 2-3, with studies suggesting that high quality data would have a ratio of ~2.8 (Bainbridge et al., 2011). Ts/Tv ratios below this range are assumed to be low-quality and contain a high proportion of false-positive variant calls and additional filters and quality assessments are required (Q. Liu et al., 2012). High quality variants can then be further analysed and interpreted to identify clinically relevant variants.

1.8 Variant analysis and interpretation

The third and final step in NGS is variant analysis and interpretation, in which the variants are annotated, filtered, and prioritised using available variant information associating them to disease. This stage of analysis is less streamlined, but recommendations and guidelines are available for filtering and prioritising the identified variants.

1.8.1. Variant annotation

Annotation is the process of adding information relating to the biological and functional consequence of the identified variants to determine their genomic context (McCarthy et al., 2014). Annotation information is available from several publicly accessible databases and bioinformatics tools that integrate available variant and gene information from the different sources. Several tools are available in genomics, and the most widely used include, Ensembl-VEP, (McLaren et al., 2016), ANNOVAR (H. Yang & Wang, 2015), Vcfanno (Fuentes-Pardo & Ruzzante, 2017) and SnpEff (Cingolani et al., 2012). The available annotation information is used to filter and prioritize variants to identify candidate disease-causing variants.

1.8.2. Variant interpretation using the ACMG-AMP guidelines

Guidelines for the evaluation and interpretation of genomic variants to identify possibly disease contributing variants, are currently available from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) (Richards et al., 2015). The guidelines use a five-tier system to classify variants based on supporting evidence levels from a range of annotation sources.

Variants are classified based on two assertions of benign and pathogenic; benign (ACMG-AMP class I); likely benign (ACMG-AMP class II); variants of unknown significance (ACMG-AMP class III); likely pathogenic (ACMG-AMP class IV) and pathogenic (ACMG-AMP class V). Variants classified as both pathogenic and likely pathogenic have the same pathogenesis weight and are considered to provide sufficient evidence for a positive molecular diagnosis in clinical decision making (Richards et al., 2015).

Pathogenesis evidence is evaluated using multiple evidence categories which includes but not limited to, population frequency data, variant predicted pathogenesis based on several *in-silico* algorithms, functional studies linking the variant or gene to human phenotype and disease, and segregation analysis data (Richards et al., 2015). Despite the clear guidelines for interpretation, different laboratories use their own judgement for the interpretation of the different evidence categories. Figure 1.3. provides a list of the pathogenesis categories and the level of evidence and codes for each. Below is a more detailed discussion of some of the evidence categories that are especially important for variant interpretation in African ancestry populations.

It is assumed that potentially disease-causing variants would be under negative selection and thus would be absent or extremely rare in the general population (Rapaport et al., 2021). Variant population frequency data is available from large public databases which include but not limited to, the 1000 genome project (Auton et al., 2015), Exome Aggregation Consortium (ExAC) (Lek et al., 2016), and the Genome Aggregation Database (gnomAD) (Karczewski et al., 2020), with the latter recommended as it contains the largest population frequency data.

The ACMG-AMP guidelines recommend the prioritization of rare population variants, and the application of PM2 (“Absent from controls, or at extremely low frequency if recessive, in Exome Sequencing Project, 1000Genomes Project, or ExAC”) (Richards et al., 2015) for such variants. However, determining rarity in the African ancestry population using currently available reference population databases, is challenging.

African genomes have a high level of genetic diversity (Campbell & Tishkoff, 2008) with approximately 10-20% more genetic variation compared with European populations. This translates to around 300 million more genetic variants than in the current human reference genome build GRCh38 (Sherman et al., 2018). A more recent study by Choudhury et al., (2020) identified 3.4 million novel SNVs in genomic data from 426 individuals from 13 African countries (Choudhury et al., 2020).

This highly diverse population is underrepresented in current genetic databases (Popejoy & Fullerton, 2016). There are currently 12,487 individuals compared to 112,386 for European individuals (comprised of non-Finnish Europeans, North-Western European, Southern European, Other non-Finnish European) in the gnomAD v2.1 database (Laurent Francioli et al., 2018). This is a challenge, as rarity or the absence of a variant may be due to the small sample size in the reference population database rather than true rarity among individuals in the population.

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 1.3. Evidence framework for each of the criteria for a benign or pathogenic assertion of a variant. **Abbreviations**; BS, benign strong; BP, benign supporting; FH, family history; LOF, loss of function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong; PVS, pathogenic very strong.

The ACMG-AMP guidelines recommend the application of PP3 for missense variants predicted to be pathogenic using *in-silico* pathogenesis predictor algorithms (Richards et al., 2015). Several algorithms are available to distinguish between likely benign and probably pathogenic variants. Annotation tools and bioinformatics databases such as Ensembl-VEP (McLaren et al., 2016) and dbNSFP (X. Liu et al., 2020) facilitate access to majority of these *in-silico* tools thus removing the need to interrogate each one individually (X. Liu et al., 2016).

Majority of these pathogenesis predictor algorithms use several variant and gene characteristics to calculate the possibility of the variant being pathogenic (Djotsa Nono et al., 2018; T. Zhang et al., 2019). These include but are not limited to, variant consequence and its impact on protein structure and function, evolutionary constraint which is defined as whether the variant is in a highly conserved genomic region across all species, protein sequence homology of the impacted gene, and allele frequency in reference population databases (Djotsa Nono et al., 2018). Moreover, these algorithms are benchmarked using variant datasets of known benign and pathogenic variants, in which samples from individuals of African ancestry are limited (Krusche et al., 2019). This lack of diversity in the represented population during benchmarking, coupled with the underrepresentation of African individuals in reference population databases (Lek et al., 2016; Popejoy & Fullerton, 2016) poses a challenge in the interpretation of variants in this population.

Firstly, the accuracy and performance of *in-silico* pathogenesis predictor algorithm tools have been demonstrated to be lower for novel variants (Gunning et al., 2020; Niroula & Vihinen, 2018). The performance of 10 *in-silico* pathogenesis algorithms was evaluated among six population subsets from the ExAC database (Niroula & Vihinen, 2018). The study results reemphasized the observation that African genomes have a high level of genetic diversity by reporting 62.4% unique variants among African individuals in comparison to 6.8% among those of European ancestry (Niroula & Vihinen, 2018). Secondly, the study also found that the predictor tools had a lower performance for unique variants compared to common variants across the populations included. This observation is most likely due to *in-silico* algorithms being benchmarked using variant datasets in which African ancestry individuals are underrepresented as already mentioned. Thus, the algorithms have poor sensitivity for novel variants. This is important as a higher proportion of variants are prioritized by *in-silico* pathogenesis algorithms due to the rarity of the variants in the population reference databases.

This has been reported to result in a lower proportion of pathogenic variants among patient of African ancestry compared to their European ancestry counterparts (Lumaka et al., 2022). A WES study aimed at identifying variants in ACMG-AMP actionable genes, reported pathogenic and likely pathogenic variants in 0.7% and 1.2% of the 4,300 European ancestry individuals from the National Heart Lung and Blood Institute Exome Sequencing Project respectively. This is in comparison to 0.3% pathogenic SNV and 0.6% likely pathogenic SNVs among 2,203 African ancestry individuals (Amendola et al., 2015). The authors suggested that the lower yield in African individuals could be a result of the underrepresentation of African ancestry

participants in population reference databases, as a higher proportion of variants prioritised are considered rare based on allele frequency. However limited evidence from other pathogenesis categories is available for most of these variants to classify them as pathogenic (Amendola et al., 2015). Thus, a higher percentage of variants of unknown significance (VUS) are reported.

VUSs may occur in genes not yet associated with disease, due to a lack of adequate pathogenesis evidence (Shirts et al., 2016). These have been reported to account for about 40-50% of variants identified in patient (Federici & Soddu, 2020). It would be required that the variant or gene be identified in multiple individuals (Stoller, 2018) or genotype-phenotype associations be confirmed using animal and functional studies (Baldridge et al., 2021) for these variants to be upgraded to pathogenic in patient. However, there is a lack of functional studies and variant data and research sharing among laboratories, especially in Africa, therefore many unique variants may remain classified as VUSs (Van Der Merwe et al., 2022).

Secondly, a higher proportion of false positive or negative variants have been reported among patients of African ancestry (Lumaka et al., 2022; T. Zhang et al., 2019). A study aimed at identifying causative variants in patients with hypertrophic cardiomyopathy, identified and issued positive diagnosis results to seven patients. Conclusive results were based on variants that upon reanalysis were found to be common in the population and thus variants were reclassified as benign necessitating a need to recall reports already issued (Manrai et al., 2016). The identification and reporting of false-positive pathogenic findings negatively impacts the patient and their families (G. S. Baynam et al., 2019). Clinical management and genetic counseling recommendations are initially made based on incorrect molecular diagnosis. Additionally, families find themselves in a diagnostic odyssey yet again.

Overall, the underrepresentation of African individuals in the current reference population databases makes variant interpretation and identifying the disease-causing variants in this population group challenging.

- *Known genotype-phenotype associations*

Information relating to previously reported or known variants, along with gene-disease associations are available through publicly accessible databases such as the public archive of interpretations of clinically relevant variants (ClinVar), DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources (DECIPHER), and Online Mendelian Inheritance in Man (OMIM).

ClinVar (Landrum et al., 2018) is an open-access database of clinically relevant variants and their interpreted effect on clinical phenotype and their association to disease. The database is maintained by National Centre for Biotechnology Information, and all data and supporting pathogenesis evidence is submitted by different users from genetic testing laboratories. Currently, there are nearly 2.4 million submitted records from 2,356 submitters in ClinVar (accessed 05.11.2022). Approximately 1.56 million of the records are unique variants (accessed 05.11.2022). ClinVar also reports information relating to the submissions, such as information about the submitter and evidence for the interpretations (Landrum et al., 2018).

An evaluation of ClinVar variants between 2016 and 2017, reported a reclassification of 466 pathogenic variants to VUSs, while 855 pathogenic were downgraded to conflicting interpretation on account of conflicting pathogenesis evidence from independent submitters. In the same assessment, 492 VUSs were downgraded to benign and 2,525 were reclassified as conflicting interpretation (Shah et al., 2018). Therefore, ClinVar is a useful repository for known variants, and as the number of genomic and disease studies increases the ACMG-AMP classification of previously identified pathogenic variants and VUSs may change thus improving our understanding of the genetic architecture involved in the disease.

DECIPHER is a publicly available online repository of variant and phenotype associations in disease populations (Firth et al., 2009). The data contained in the database is primarily from the large DD study, DDD-UK (Deciphering developmental disorders in UK) study(Wright et al., 2015), but also contains submissions from different projects and users. Users can select open access thus allowing for their submitted data to be publicly accessible. This facilitates data sharing between institutes thus further improving variant analysis, interpretation, and gene discovery (Firth et al., 2009; Swaminathan et al., 2012).

The database makes available a curated gene list of ~2,500 genes known to be associated with DD via the Developmental Disorders Gene to Phenotype (DDG2P) gene panel. The genes are categorized by the level of supporting pathogenesis evidence, with “confirmed or probable annotations” given to genes with sufficient evidence of disease or phenotype association, and those annotated as “possible” having limited gene phenotype evidence.

OMIM is a database which documents human genes and associated genetic phenotypes. The database contains information on over 15,000 genes and 6,000 mendelian phenotypes. It also

provides phenotypes and a summary of publications that have reported pathogenic variants in the gene including any functional studies and evidence if available (Amberger et al., 2014).

1.8.3. WES is the current recommended first-line test in DD patients

The current guidelines for genetic testing in DD patients recommend the use of WES or WGS as first-line or second-line tests following normal results from CMA or targeted gene testing (Manickam et al., 2021). The choice of sequencing approach and whether to use the test as a first-line or second-line test is influenced by factors such as the policies mandating the health-care system, and overall cost. Reduced diagnostic costs have been reported for the use of WES or WGS as first-line tests. Monroe et al (2016), estimated a cost of 16,409 USD per patient for serial testing over a mean of 6.6 years using traditional diagnostic tests. This cost estimate included genetic tests, biochemical tests, specialist visits, and hospital admissions from disorder comorbidities. The once-off cost for trio-based WES was estimated at a mean of 3, 972 USD and resulted in a diagnostic yield of 29.4% following negative results from traditional diagnostic testing (Monroe et al., 2016). The authors suggested that the cost saving would be much higher if WES were used as a first-line test.

The recommended use of WES or WGS was made following a systematic review of evidence from studies assessing the clinical utility of the tests in patients presenting with one or more congenital anomalies, ID and or developmental delay (Manickam et al., 2021). A meta-analysis comparing the diagnostic utility of WES, WGS, and CMA across 37 studies and a total of 20,068 children with suspected genetic disorders, reported a higher diagnostic yield for WGS of 41% and WES at 36%, compared to CMA at 10% (Clark et al., 2018). In another study, a diagnostic yield of 41% using WGS in 103 patients with a range of clinical phenotypes was reported. The authors reported the identification of causative pathogenic variants in the non-coding region of the genome in nine patients that were missed by WES (Lionel et al., 2018). Additionally, a positive molecular diagnosis was reached using standard testing in 25% of the patients who would have also received a diagnosis based on WGS and WES if used as a first-line genetic test (Lionel et al., 2018).

Gene panels offer more sequence coverage, at a much lower cost compared to both WES and WGS (Klein & Foroud, 2017). This approach is however not ideally suited to disorders with a high level of heterogeneity such as DD, as it fails to provide the opportunity to identify novel disease genes. This is especially important for understudied African ancestry populations with a demonstrated high level of genetic diversity (Popejoy & Fullerton, 2016). Additionally, the

use of gene panels does not afford the opportunity for data reanalysis when more disease-causative genes are identified (Mørup et al., 2022; Saintenac, 2015). A study comparing the diagnostic utility of WGS and WES showed that the initial diagnostic yield of WES increased from 41% to 52% following reanalysis. This was in part due to the discovery of new disease genes, and improved bioinformatics tools (Ewans et al., 2022). This is important, for the reclassification of VUSs and false positive pathogenic variants that may be identified in underrepresented populations, as discussed in 1.8.2.

These results demonstrate the diagnostic superiority of both WES and WGS over gene panels and CMA which had been the previously recommended first-line genetic test in these patients (D. T. Miller et al., 2010). Moreover, they highlight the higher diagnostic yield of WGS, in undiagnosed DD patients (Stavropoulos et al., 2016; Taylor et al., 2015; Yuen et al., 2018). The higher yield is in part on account of the ability of WGS, to detect non-coding SNVs and SV at a higher sensitivity of < 5Kb, and the detection of SNV and Indels in genome regions that are difficult to sequence using WES such as homopolymer regions (Barbitoff et al., 2020; Lelieveld et al., 2015; Nolan & Carlson, 2016).

Furthermore, the clinical utility of WES and WGS, measured as a positive molecular diagnosis resulting in a change in disease management or clinical intervention has been reported (Manickam et al., 2021). Positive molecular diagnosis was reached in 70% (52/74) of critically ill Chinese infants admitted to ICU, using the combined efforts of WGS and WES. For 21.6% of the infants a change in clinical management was recommended following a positive diagnosis, while for 32.2% referrals were made to more specialised physicians (Wu et al., 2021). In another example, among 24/29 patients with a positive molecular diagnosis, recommendations for therapeutic interventions were made in 20. In eight patients genetic counseling relating to recurrence risk assessment in family members was provided, and for 17 patient family planning for future pregnancies was provided (Gubbels et al., 2020).

Genetic testing can be performed as proband only, duo or trio testing, with the latter preferred should samples from both biological parents and the patient be available. Studies have assessed and highlight the higher diagnostic yield of trio analysis compared to proband only and duo analysis (Charng et al., 2016; Helbig et al., 2016; Ji et al., 2019; Retterer et al., 2016).

Trio analysis allows for the identification of *de novo* variants known to contribute to 42-48% of DD cases (Deciphering Developmental Disorders Study, 2017). A diagnostic yield of 31%

and 24% was reported for trio and proband only analysis respectively among 3,040 patients with a range of clinical features including multiple congenital anomalies. Of these 37.7% were due to *de novo* variants (Retterer et al., 2016).

Additionally, trio analysis facilitates the identification of novel candidate disease genes, which coupled with the beneficial power of data reanalysis leads to an increase in the overall cohort diagnostic yield (Wright et al., 2018; Fung et al., 2020). A total of 17 variants were reclassified following WES data reanalysis. In majority of the cases (14/17), the variants were classified as pathogenic, due to the availability of new gene disease associations (Bowling et al., 2017). Interestingly, for three of the variants a combination of applying less stringent quality control parameters, clinical cases and variant discussions with scientists from other institutes led to classification upgrades to pathogenic (Bowling et al., 2017).

Despite the higher diagnostic yield of WGS over WES, there are challenges to its routine implementation in clinical settings. First, WGS produces around 90 GB volume of raw data per sample, which is much higher than 6-8 GB for WES at 100x coverage (Suwinski et al., 2019). Thus, WGS requires more disk storage capacity, and computational processing power. Additionally, the large data requires more time for data analysis compared to WES (Schwarze et al., 2018; Suwinski et al., 2019). Second, the implementation cost of WGS is significantly higher compared to WES. A single WGS run can cost up to 24, 800 USD compared to 5,169 USD for WES (Schwarze et al., 2018). Thus, for many institutes, the implementation of WGS is a trade-off between the financial costs and diagnostic yield, with the decision being influenced by the available funds, especially in resource-poor settings where most patients rely on the public health care system for genetic testing (Ewans et al., 2022). The above detailed challenges of WGS are in part the reason for the wider implementation of WES in diagnostic and research settings.

1.9 Challenges to diagnostic implementation of NGS

The diagnostic utility of NGS particularly WES is undoubtable, with the tool now recommended as a first-line diagnostic test for individuals with DD (Manickam et al., 2021). However, there are a few challenges associated with its translation from mainly research to clinical setting especially in resource-poor settings (Nguengang Wakap et al., 2019).

- *Relatively high implementation and running costs*

The average cost price for sequencing has significantly decreased over the past decades. Sequencing of the whole genome previously cost around 3 million USD (Hood, 2008; Hood & Rowen, 2013). This decreased and stabilised to around 1,000 USD in 2021 (NIH, 2021) as shown in figure 1.4. Recent studies claim even more reduced costs at ~USD1/GB (Almogy et al., 2022) with the development of new sequencing platforms such as AVITI by Element Biosciences (Element Biosciences AVITI:, 2022) and improvements in the sequencing chemistries (Almogy et al., 2022; Jeon et al., 2021).

Despite these cost reductions, the implementation and recurrent running costs of NGS remain relatively high and require substantial continued investment. For example, the cost of the IIlumina NovaSeq S4 V1.5 platform would require an initial cash investment of 900, 000 USD (Element Biosciences AVITI:2022). Additionally, the yearly maintenance and warranty cost can run up to 26,000 USD per year contributing to the overall cost of implementation (Vogel et al., 2021). Moreover, the costs for appropriate general laboratory infrastructure which includes for example, adequate space, a temperature-controlled environment for instrument setup and reagent storage, and a continuous power supply to ensure an uninterrupted sequencing run (Yu, 2014) are required but currently limited in most resource-poor settings.

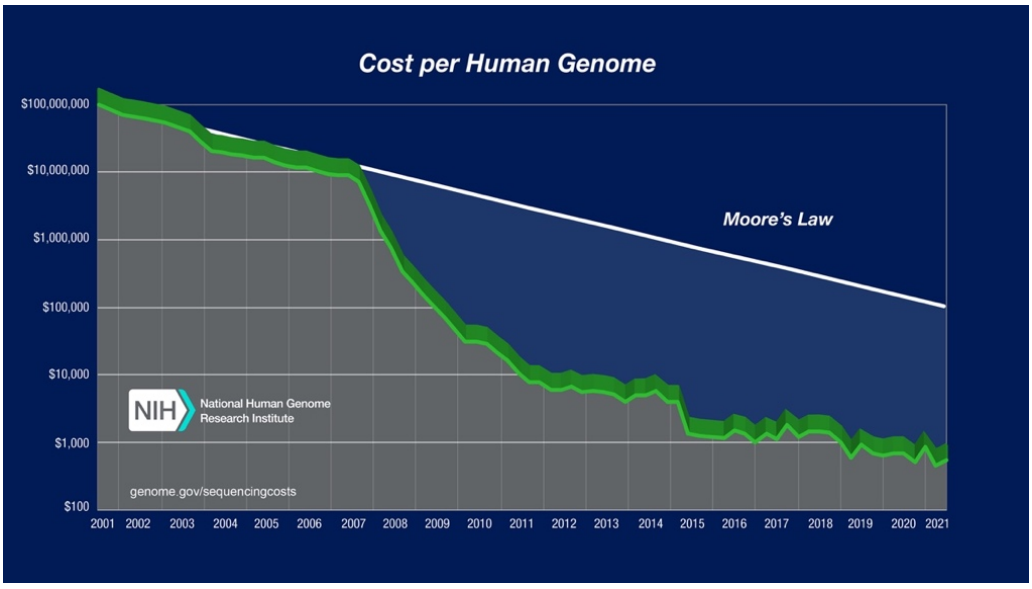

Figure 1.4. The decreasing estimated cost of sequencing the whole human genome using current next-generation sequencing platforms. This is calculated against Moore's law. Taken from Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

In resource-poor settings such as Africa there is a lack of financial support and prioritization from national government bodies funding the public health care system (Kamp et al., 2021) for

the implementation of WES/WGS. Currently, these sequencing technologies are offered by private laboratories, and research institutes where implementation and running costs are covered through individual research grants from institutes such as the National Institute of Health (NIH) (Van Der Merwe et al., 2022).

In these settings, individuals are unable to afford the cost to cover WES or even the traditional tests offered routinely. The cost of WES in South Africa is estimated at R15,000–R35,000 (1,000-2,250 USD) per sample (Wiener et al., 2022) which is about 4-9 times higher than the minimum wage of most individuals in the country at R3,500-R4,100 (South African Government, 2018).

A retrospective study of requested genetic tests and the payment systems utilized at a County hospital in the US between 2013-2018, found that 60% of patients visiting the hospital were uninsured, and genetic testing costs had to be covered under the County Hospital Financial Aid Programme. Of the 1,429 genetic tests requested, 15.5% of the tests were not performed due to financial limitations from the hospital financial aid program, while 13.3% of the tests were not performed as patients were unable to pay out-of-pocket (Erwin et al., 2020).

- *Geographic and logistic complexities*

There are complex logistics relating to the importation of instruments and sequencing reagents. Reagents are imported from European suppliers even when ordered through local providers. This can cause shipping and clearance delays at cross-border customs agencies. It was reported that such long delays in importing sequencing reagents can result in the products subsequently expiring before use by the time they arrive at the laboratory (Vogel et al., 2021).

- *A lack of adequately trained personnel*

The implementation of WES requires a workforce with highly skilled individuals with different skill sets. These include lab scientists and technologists, bioinformaticians, medical geneticists and genetic counsellors. For example, the analysis of a large volume of raw sequence reads using the recommended GATK best practices workflow would require the use of workflow managers such as Nextflow (Di Tommaso et al., 2017) and SnakeMake (Mölder et al., 2021). These however require one to have prior Linux environment and bioinformatics training for execution, a skill that is lacking in the region. There are workflows such as GALAXY(Afgan et al., 2016, 2022) and DRAGEN (Goyal et al., 2017), that require minimum user interface and

therefore no highly specialised skills are needed for execution (Afgan et al., 2016). However, these do not offer much flexibility with regards to making changes to the parameters for each of the different workflow steps (Di Tommaso et al., 2017).

Variant interpretation is a manual process that requires a multidisciplinary team of specialists which includes, genetic clinicians, genetic counsellors and scientists are limited in poor resource settings (Krause, 2019). In 2018, it was estimated that ~20 genetic counsellors were available in South Africa (Abacan et al., 2018). One would expect this number to have only increased slightly as new genetic counsellors are enrolled for training every two years at only two institutes, the University of Cape Town, and the University of the Witwatersrand (Abacan et al., 2018).

In most regions, genetic clinics and services are available exclusively in urban areas. In South Africa, genetic services are provided by the National Health Laboratory Services (NHLS) and are linked to academic institutes available mainly in four provinces: Gauteng, Western Cape, Free State, and Kwa-Zulu Natal (Kromberg et al., 2013). Thus, patients in rural areas have very limited access to genetic services and have to travel long distances to access genetic clinics and laboratories. In a retrospective study by (Erwin et al., 2020), the authors noted that each requested genetic test completed, it required two separate hospital visits; at the first visit, was an initial assessment and request for the appropriate genetic test; second visit, involved paperwork completion, and blood or saliva sample collection. This would mean that should serial testing be required due to negative or normal results from the initial requested test, then the patient must go through the same long process again. This becomes a large logistical issue especially for those travelling from rural areas (Erwin et al., 2020). Thus, most patients are lost to follow-up and the number of tests performed as part of the trajectory of diagnostic investigations is limited.

- *Increased data storage and computational infrastructure*

It is estimated that a single WES run at 40x coverage produces a BAM file of approximately 5.7 GB (Strand Life Sciences, 2016). This increases significantly to about 1.6 terabytes (TB) for 200 samples (Strand Life Sciences, 2016). Additional storage considerations also need to be made for files relating to instrument data (image files and signal intensities) and, interpretation and analytical files (quality metrics, BAM files) enabling reanalysis of data. The ACMG clinical laboratory standards for NGS (Rehm et al., 2013), recommend that laboratories

reanalyse data every two years, thus these files would need to be stored on the servers for a minimum of two years.

These data storage requirements are typically not met using local desktops especially for processing and storing large sample datasets. The solution is to use high performance computing systems (HPC), which comprises of networking several computers (known as nodes) together in a cluster system or using a cloud environment (Kurtzer et al., 2017). These allow for parallel processing of data at high speeds, large data storage and easy data transfer and sharing across different networks and research groups. Despite these highlighted benefits of HPCs, they require an initial substantial investment in infrastructure and long-term maintenance (Kurtzer et al., 2017; Mondragon et al., 2016).

- *Complex data analysis and interpretation*

There are guidelines and bioinformatic tools for the analysis and interpretation of NGS data to identify candidate variants. As discussed in section 1.7 and 1.8 above, most of the bioinformatics tools available have been benchmarked using variant data from patients of non-African ancestry. and rely heavily on allele frequency (Djotsa Nono et al., 2018; T. Zhang et al., 2019) from population reference databases in which individuals of non-European ancestry are underrepresented (Lek et al., 2016; Popejoy & Fullerton, 2016). These challenges, result in a higher proportion of VUSs and fewer pathogenic variants have been reported in African individuals compared to their European counterparts (see discussion in 1.8). Thus, a higher number of patients in the African population have inconclusive results and negative results following WES.

1.10 Current initiatives to increase African genomic research

One of the main limitations in African genomic research is the underrepresentation of African individuals in current reference databases. The availability of NGS technology has led to increased sample size and increased submissions of high-quality genomic data from mostly individuals of non-African ancestry(G. S. Baynam et al., 2019). Genomic data from African individuals sequenced using NGS (WES and WGS) is also available as seen in Figure 1.5, but it still lags behind to that of European ancestry populations (G. S. Baynam et al., 2019; Lumaka et al., 2022).

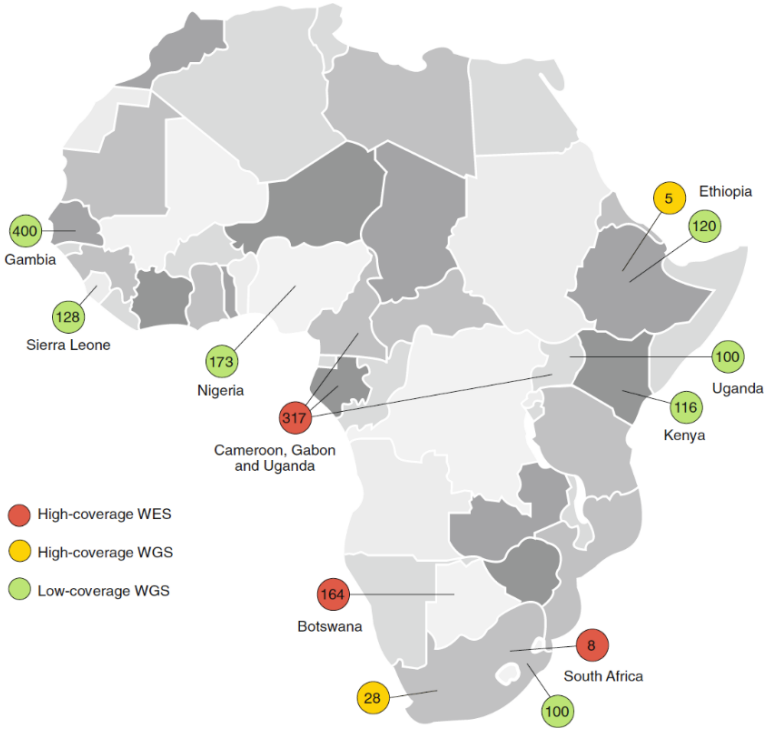

Figure 1.5. Population genome data across Africa. Numbers represent people with whole-exome and whole genome sequences published in genomics databases per country. Most individuals have had low coverage WGS performed. Taken from (G. S. Baynam et al., 2019).

Efforts to increase the overall sample size and level of diversity for African ancestry individuals in population databases have been initiated mostly through the Human Heredity and Health in Africa (H3Africa) consortium. The consortium is funded by the National Institutes of Health (NIH) and Wellcome Trust (N. Mulder et al., 2018). One of the aims of the consortium is to facilitate research on the African continent focused on identifying the genetic and environmental contributors to diseases affecting the African population (N. Mulder et al., 2018).

One of the main limitations to the implementation of NGS technologies and increased genomic research in Africa is the lack of appropriate data storage, analysis infrastructure, and a lack of adequately trained personnel for data analysis. Consortiums such as the Pan African Bioinformatics Network (H3ABioNet) that have been established within the continent and address some of these challenges. The main aim of H3ABioNet is to equip African scientists with the skills to analyse their own data (N. J. Mulder et al., 2016). Among other strategies, the consortium aims to achieve by facilitating training and support programs for biologists and clinicians across African institutes. The consortium held 49 training courses between 2013 and 2016 and provided training to 1,036 individuals during that period.

As discussed above in section 1.10, the use of HPC systems offers a solution to some of the data storage and the analysis of large sample volumes. Currently, there are at least 17 HPC

available across African research institutes (https://www.h3abionet.org/tools-and-services/computing-infrastructure). The ZA-Wits Core cluster, which is a partnership across several departments, houses nine storage servers with memory ranging from 24GB–1TB each (http://grid.core.wits.ac.za/).

Moreover, there are also a few research studies that aim to delineate the underlying genetic causes of different diseases affecting African patients.. Among the different studies, the NeuroDev group aims to "expand knowledge of the genetic architecture and environmental risk factors for neurodevelopmental disorders in Africa" (de Menil et al., 2019). To achieve this, the study intends to recruit a total of 5,600 participants from Kenya and South Africa (2,600 in Kenya and 3,000 in South Africa). This includes children of 2-17 years of age presenting with defined neurodevelopmental disorders (ID, Autism spectrum disorders, Attention-deficit/hyperactivity disorder, communication disorders, any other specific learning disabilities, and global developmental delay) (de Menil et al., 2019).

Another working group is the Deciphering Developmental Disorders in Africa (DDD-Africa) study which aims to delineate the aetiology of DD in African patients using WES. The study is discussed in more detail below.

1.10.1.Deciphering Developmental Disorders (DDD) DDD-UK

Deciphering Developmental Disorders (DDD) in UK study aimed to identify the underlying genetic aetiology in patients with unexplained DD (Wright et al., 2015). Trio families were recruited from 24 regional genetics clinics around the United Kingdom and Ireland. In the first year they had recruited 2,000 families and within three years this number had increased to 8,000 and currently 14,000 families have been recruited (Deciphering Developmental Disorders (DDD) project, 2020). Testing was performed using microarray and WES.

In the study's first report the authors reported on the diagnostic yield in 1,133 trio families. In this cohort a diagnostic yield of 27% due to pathogenic variants in known DD genes was reported. In 2018, the study reported an increased diagnostic yield of 40%
(Wright et al., 2018), in the same cohort of patients which included an additional 454 families due to the identification of variants in newly discovered DD genes. This demonstrates the diagnostic benefit of WES in gene discovery and the power of reanalysing NGS data considering new gene and disease associations.

Moreover, the study provided evidence that most DD cases are caused by *de novo* variants, with the DDD-UK reporting 345/435 (78%) of identified variants were *de novo* (Deciphering Developmental Disorders Study, 2017). The study highlighted the benefits of trio family analysis in identifying the underlying genetic aetiology of DD and significantly reducing the number of candidate variants to be considered resulting in more accurate diagnoses.

1.10.2. DDD-Africa

The DDD-Africa study was initiated in 2018 and is part of the H3Africa consortium, supported by the NIMH. This is a collaborative study between researchers in the Division of Human Genetics, NHLS and WITS, South Africa, Centre for Human Genetics, the University of Kinshasa, Democratic Republic of the Congo and the DDD-UK study. The project's overarching aim is to explore whether systematic phenotyping and detailed genomic analysis improve the prospect of identifying likely pathogenic mutations in African patient with DD. Additionally, it aims to understand the genetic contributors to DD in Africa and effectively implement WES in a diagnostic setting to allow for better precision public health in Africa. The study aims to recruit 350 families from South Africa and 150 from the DRC respectively.

The study has developed a recruitment system, phenotyping methodology, WES logistics and data analysis framework which has already yielded a significant finding. The study has thus far published a case report of the first Africa patient with Takenouchi–Kosaki syndrome (Flynn et al., 2021). This current PhD work is nested under the DDD-Africa study.

1.11 Study Rationale

DD is a large genetic and phenotypic heterogeneous group of disorders, making the identification of the underlying aetiology challenging. WES is the current first-line genetic test recommended in this group of patients. The genetic test allows for the identification of SNVs, and Indels in the exome region of the genome at a higher resolution compared to traditional genetic tests and previously recommended first-line test CMA.

However, in South Africa traditional genetic tests, mainly, karyotype, FISH and MLPA with lower associated diagnostic yields are still routinely used. Therefore, the proportion of patients with unexplained DD is potentially much higher compared to that reported from high income countries.

Studies and evidence evaluating the diagnostic utility of WES are mainly from high-income countries using European patients. Currently, no studies have replicated this benefit on a large scale in African ancestry patients. African ancestry individuals have been demonstrated to have a high level of genetic diversity with a lower proportion of pathogenic variants and a high number of VUSs reported. This is mainly due to an underrepresentation of these individuals in publicly available reference population databases.

Thus, based on this, the hypothesis is that pathogenic variants may be identified in African DD patients resulting in a lower overall diagnostic yield. Again because of the high genetic diversity in the African population we anticipate a higher number of novel variants in this group of patients.

Moreover, the challenge with WES is the large amount of variant data produced per patient, making prioritization and interpretation of variants complex. This current PhD study thus seeks to investigate whether WES is truly beneficial in the discovery of causal variants in African patients with unexplained DD. Also, we ask whether recommendations for variant prioritization and interpretation could be made, thus making the process of discovering the underlying genetics of DD more efficient for research and diagnostics in Africa

1.12 Aim and Objectives of the current study

The aim of this study was therefore to develop a DD variant analysis pipeline with filtering and prioritization criteria appropriate for African patients with unexplained DD and to use this pipeline to identify putative disease-causing small genetic variants in this DD cohort. To this end, the specific objectives of this research study were to:

1. Capture the phenotypic diversity of the cohort using human phenotype ontology terms.
2. Annotate variants detected by WES, using Ensembl-VEP annotation tool.
3. Identify putative disease-causing variants in known DD genes using the DDG2P gene list, and candidate *de novo* variants in novel DD genes using BCFtools/trio-dnm.
4. Perform Sanger sequencing and segregation analysis where possible to validate low quality candidate variants and putative disease-causing variants.

Chapter 2 - Methodology

Patients were recruited into the study, and clinical data had previously been recorded by the DDD-Africa team. This clinical data along with WES data was available to fulfil the study objectives as outlined in section 1.14 above.

2.1 Ethics and consent

Ethical approval for the DDD-Africa study (protocol M180678) (Appendix 2) and this PhD study (protocol M200688) were obtained from the Human Research Ethics Committee (Medical), at the University of the Witwatersrand (Appendix 3). Written informed consent was provided by all study participants, with parents or legal guardians consenting on behalf of minors and if possible, assent was obtained from minors.

2.2 Patient cohort

Data from a total of 117 patients with unexplained DD and 180 of their unrelated parents were available for this study. These represented a proportion of patients recruited as part of the larger DDD-Africa study (detailed in 1.12.2) which aimed to recruit a total of 350 patients with unexplained DD from South Africa.

Patients were referred from the Division of Human Genetics affiliated genetic clinics held at tertiary academic hospitals in Johannesburg. These were Charlotte Maxeke Johannesburg Academic Hospital, Chris Hani Baragwanath Academic Hospital, and the Rahima Moosa Mother and Child Hospital. Study referrals and recruitment for patients included in this study were from 2018 through 2019.

Patients were eligible for inclusion into the study if they met one of the following inclusion criteria;

- unexplained DD with moderate or profound ID.
- multiple major malformations in two, more different organs.
- mild to moderate ID with dysmorphic features.
- one major malformation and three or more well documented minor anomalies.

The detailed inclusion and exclusion patient criteria is provided in Table 2.1. Importantly, the inclusion criteria were later improved with criteria B segmented into B1 and B2 with the difference between the two reflecting the number of months the patient presented with a major malformation.

Table 2.1. Patient recruitment eligibility criteria for the DDD-Africa study.

Patient inclusion criteria		**Patient exclusion criteria**
	A. Unexplained DD with moderate to profound intellectual disability (ID)	A. Known conditions for which no strong evidence exists for a monogenic cause
B**.** Multiple major malformations in 2 or more different organs	B1. Multiple major malformations in 2 or more different organ systems > 1month)	B. Suspicion of an acquired brain lesion, with predominant neurological manifestations
	B2. Multiple major malformations in 2 or more different organ systems < 1month	C. Suspected multifactorial or suspected environmental cause
	C. One major malformation and three or more well documented minor anomalies	D. Suspected environmental cause
	D. Mild to moderate intellectual disability and dysmorphic features	E. Mild ID only

2.3 Patient clinical information

Clinical data which included growth and development parameters, family history, birth, and pregnancy information, had been collected during recruitment using a standardised questionnaire. Additional clinical history was obtained from the patient medical records, if available.

Collected clinical data pertaining to patient growth parameters, and clinical features was defined using appropriate Human Phenotype Ontology (HPO) terms and was entered and stored on the DECIPHER database (Firth et al., 2009). The use of standardised terminology through HPO terms enables phenotypic similarities to be assessed within the current DD cohort and with other subgroups of similarly affected DD patients.

2.4 Sample collection and processing

Following written informed consent, whole blood samples were collected from all enrolled patients and their parents. Genomic DNA had previously been extracted by the DDD-Africa laboratory team from peripheral blood using a modified in-house salting-out method (S. A.

Miller et al., 1988). The extracted DNA was purified using ethanol solution and re-suspended in Tris-EDTA buffer and stored at 4°c to prevent degradation. The quantity and quality of the extracted gDNA was assessed on the Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, USA). Absorbance scores greater than 1.8-2.0nm at A260/280 was considered an acceptable range of purity while a value of >2.0nm was regarded as an indicator of RNA and protein contamination which may interfere with downstream processes. Agarose gel electrophoresis was utilized as an additional measure of DNA integrity and quality by electrophoresing samples through a 0.8% gel.

2.5 WES and variant calling at the WTSI

The DNA samples were sent to the Wellcome Trust Sanger Institute (WTSI) for exome sequencing. Genomic DNA was enzymatically fragmented to 150 bp using the Index Sequence Capture (ISC) and Twist library capture kit (Twist Bioscience, San Francisco, CA, USA). Adaptor-ligated libraries were amplified and appropriately indexed via PCR. This was followed by 2 x 150 paired-end sequencing on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) at a target sequence depth of 40x.

Sequence reads were mapped to the reference human genome build hg38 (GRCh38) using the BWA (Li & Durbin, 2009). Variants were called using the GATK (v 4.1.0.0) germline short variant discovery best practices recommendations (DePristo et al., 2011; Poplin et al., 2018; van der Auwera & O'Connor, 2020). Read duplicates which arise during the PCR amplification steps during library construction were flagged using Picard-MarkDuplicates (reference URL can be found in Appendix 1).

Variants were called using the GATK v4.1.0.0 HaplotypeCaller in gVCF mode (Van der Auwera & O'Connor, 2020) generating a gVCF for each sample. The sample gVCF files were combined into a single dataset and joint-genotyped using GATK GenotypeGVCFs. GATK Variant Quality Score Recalibration (VQSR) was applied for both SNPs and INDELs to assess the accuracy of the variant calls and to identify true variants. All variant sequencing, read alignment and variant calling was performed by the WTSI team.

2.6 Variant annotation

Ensembl-VEP release v106 (McLaren et al., 2016) accessed via the ZA-Wits Core Cluster UNIX command-line interface, was used to predict the functional impact of the variants. The

annotation databases that were included in the annotation workflow were, population frequency databases (gnomAD, ExAC, 1000 genome etc), *in-silico* pathogenesis predictor bioinformatics tools (SIFT, Poly-Phen2, CADD etc), disease and variant databases (ClinVar), VEP-DDG2P plugin and bcftools trio-dnm. Table 2.2 provides a detailed list of all database sources included in the annotation. The output was limited to only the canonical transcript or longest transcript using the --canonical flag option in Ensembl-VEP.

2.7 Variant filtering and prioritization workflow

The process of prioritising variants to identify likely disease-causing variants from exome data can be performed using various approaches. In this study, variants were prioritised using two strategies. In the first strategy, Ensembl-VEP G2P prioritization tool using the DDG2P virtual gene panel (release v2.2) was used to identify and prioritise candidate variants overlapping known and well-defined DD genes. The second strategy identified and prioritised candidate de novo variants in complete family trios using bcftools trio-dnm (Figure 2.2).

The strategies were executed using a semi-automated filter and prioritization workflow that was ran using Nextflow workflow manager executed via the ZA-Wits Core Cluster UNIX command-line interface.

Table 2.2. Ensembl-VEP annotation sources used in the study.

Annotation plugin	Function	Reference
G2P-DDG2P (release v2.2)	database constructed using available literature and containing gene and variant data with the associated phenotype	(Thormann et al., 2019)
Trio-dnm	The tool identifies de novo variants by filtering out regions with an alternative allele in either parent. This is unlike in the previous versions were genotype likelihood scores were used.	(Danecek, 2020)
ClinVar (v 2021-03-28)	Public database that contains variant data submitted by different users regarding gene-disease associations and clinical significance.	(Landrum et al., 2018)
CADD (v3.0)	Prediction for deleteriousness of variants.	(Rentzsch et al., 2019)
Condel	Prediction for deleteriousness of variants	(González-Pérez & López-Bigas, 2011)
Fathmm-XF	Prediction of variant functional impact in disease.	(Rogers et al., 2018)
DANN	Prediction for deleteriousness of variants	(Quang et al., 2015)
M-CAP	Predicted deleteriousness of missense variants	(Jagadeesh et al., 2016)
PROVEAN	Prediction of variant impact on protein.	(Choi & Chan, 2015)
MutationTaster	Prediction for deleteriousness of variants	(Schwarz et al., 2010)
GERP	Measure of sequence conservation	(Davydov et al., 2010)
REVEL	Prediction of deleteriousness of missense variants	(Ioannidis et al., 2016)
dbSCSNV	Public database containing SNVs within the splicing consensus region	(Jian et al., 2014)
LoFtool	Predicted deleteriousness of loss	(Fadista et al., 2017)
SpliceAI	Predicts the impact of variants on splicing	(Jaganathan et al., 2019)
SIFT	Predicted variant impact on the protein	(Ng & Henikoff, 2003)
PolyPhen-2	Predicted variant impact on the protein	(Adzhubei et al., 2010)

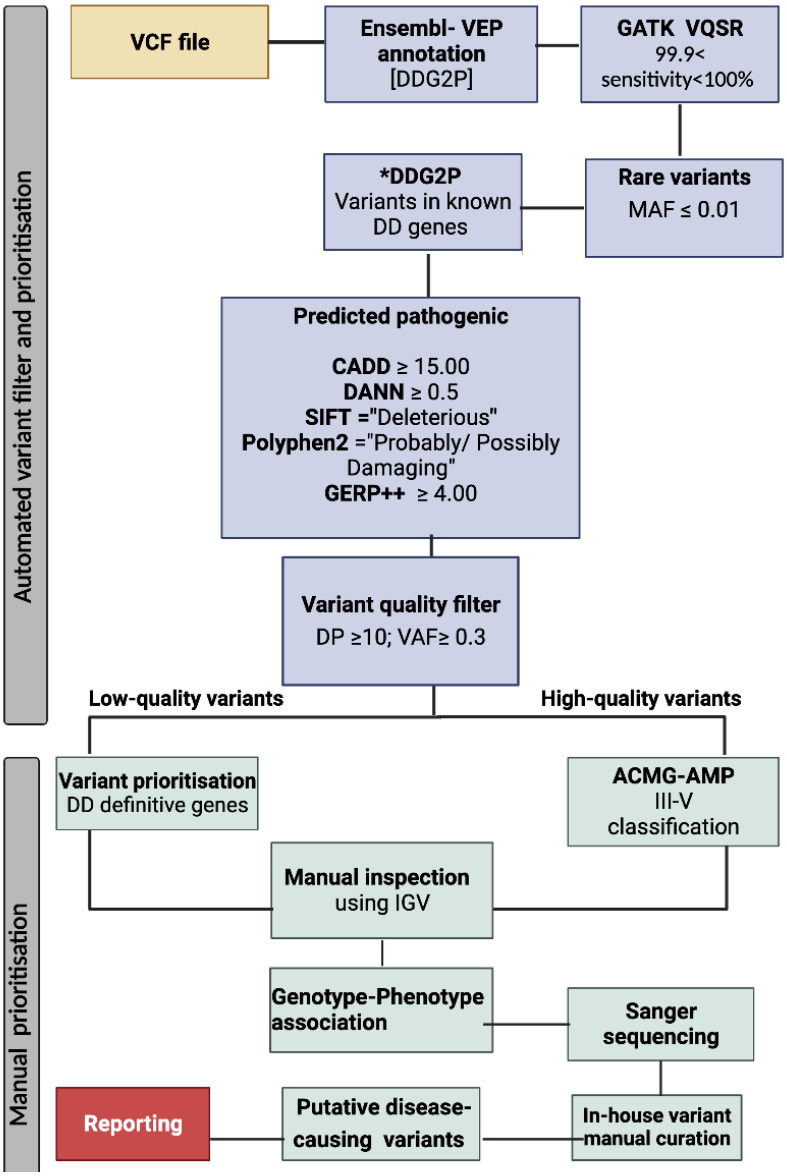

Figure 2.1. Semi-automated DDG2P variant filtering and prioritization pipeline. Variants in the VCF file were annotated using Ensembl-VEP which included information from different annotation sources, scores from different *in-silico* pathogenesis algorithm tools, reference population allele frequency databases, and the Ensembl-VEP DDG2P plugin. Variants were prioritised using different thresholds for each of the annotation sources. Prioritised variants were segregated based on quality and ACMG-AMP pathogenic high-quality variants and low-quality variants in DD definitive genes were shortlisted. Candidate variants were identified based on suggestive genotype-phenotype association and sanger confirmation. Putative disease-causing variants were prepared for patient feedback via a diagnostic report. **Abbreviations;** ACMG-AMP, American College of Medical Genetics and Genomics and the Association for Molecular Pathology; IGV, Integrative Genome Viewer. *DDG2P, Developmental Disorders Gene to Phenotype database, only variants that passed the DDG2Pplugin parameters were prioritized further.

2.7.1. DDG2P variant prioritization strategy

- *Automated variant filtering and prioritization*

To remove possible false-positive variants which may be a result of read alignment and variant calling errors, the recommended GATK 99.9<sensitivity<100% VQSR threshold for both SNP and Indel variants was applied (van der Auwera & O'Connor, 2020b). Variants that are annotated with this threshold are thought to be false-positive variants. This is because their calculated VQSLOD score is in the same range as 0.1% of the variants thought to be false positives in the variant dataset used for benchmarking and modeling.

Variants were then filtered using Minor Allele Frequency (MAF) ≤0.01, based on gnomAD allele frequency included in the Ensembl-VEP annotated VCF file. As shown in Figure 2.1 these rare variants were then prioritized to identify those overlapping known DD genes. This was achieved using annotation information from the Ensembl-VEP $G2P^{DD}$ plugin, which prioritizes possibly causative DD gene variants using defined gene parameters. The parameters used by the $G2P^{DD}$ plugin were:

The variant must overlap a DDG2P gene

1. The variant must have a severe consequence on the protein-coding function of the gene.
2. All allele frequencies from co-located variants in the human reference population databases (1000 Genomes project, ESP, gnomAD) needs to be below a given threshold. The default allele frequency values used are:
    - For a bi-allelic gene 0.005.
    - For a mono-allelic gene is 0.0001
3. The number of variants must be sufficient to meet the allelic requirements of the gene.
    - Autosomal recessive genes: at least 2 heterozygous variants or 1 homozygous variant located in the same transcript must pass all the other filters.
    - Autosomal dominant: either 1 heterozygote variant or 1 homozygote variant must pass all other filter.

Variants passing all the above parameters are annotated in the "G2P complete" column of the annotated VCF file with "YES". Variants were prioritised based on this annotation information and all other variants were removed from all further analyses.

The DD gene variants were further prioritized based on their predicted deleterious impact on the protein. *In-silico* pathogenesis algorithm models were used to predict variant pathogenicity. The variant was prioritized if it was predicted to be pathogenic by at least four of the five *in-silico* tools used. The tools and thresholds used were:

1. CADD, phred-score of ≥15.00.
2. DANN, phred-score of ≥0.5.
3. SIFT, variants annotated as “Deleterious”.
4. Polyphen2, variants annotated as "Probably Damaging" and “Possibly Damaging”.
5. GERP++, evolutionary constraint variants using a score of ≥4.00.

Variant Quality control

We identified low-quality variants using the recommended quality metrics of read depth (DP) ≤10 (Pedersen et al., 2021) and variant allele fraction (VAF) ≤ 0.3 (30%) (Song et al., 2021). The bcftools +Fill-tag plugin was used to calculate and amend a VAF score for each of the variants. This was then used along with DP to segregate the prioritized variants into low-quality and high-quality variants.

The read alignment of the prioritized variants was inspected by visualization of the CRAM file on Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011). This is a widely used open-access tool that allows for more interactive visualization of the genomic data which includes inspection of the read coverage of each identified variant (Thorvaldsdóttir et al., 2013). Any variants that showed any misalignments for example due to an occurrence in low-complexity sequence or alignment near the start or end of reads were flagged and removed from further analysis.

- *Manual variant interpretation using ACMG-AMP guidelines*

The prioritized variants were interpreted using the recommended ACMG-AMP guidelines five-tier interpretation system guidelines (Richards et al., 2015). Variants were classified as class V (pathogenic), class IV (likely pathogenic), class III (of uncertain significance), class II (likely benign), or class I (benign). Only variants classified as pathogenic and likely pathogenic were shortlisted.

The filter and prioritization pipeline used in this research work was among others developed as part of the larger DDD-Africa variant analysis workflow. In this current analysis, variants that had been previously annotated in the ClinVar variant and disease database (Landrum et al., 2018)as benign, likely benign, and VUS, were not assessed further and would be closely examined in proceeding DDD-Africa research projects.

VarSome, accessed via the web interface, was also used for variant interpretation and classification using the ACMG-AMP guidelines (Kopanos et al., 2019). This is a human genome variant search engine facilitating variant discovery and the interpretation of variants using ACMG-AMP classifications. The tool was used to aid the process of manual variant interpretation.

- *Low-quality variants prioritized to identify clinically relevant variants*

Low-quality variants were filtered to remove those previously reported in ClinVar as either likely benign/benign or VUS. The remaining variants were further prioritized and limited to those in DD definitive genes and were considered plausible based on the patient's clinical phenotype. Sanger sequencing confirmation was performed for the shortlisted plausible low-quality variants. This was due to a low WES mean coverage that could have resulted in some true variants being poorly covered and incorrectly filtered out as sequencing errors. Thus, variant confirmation with a secondary sequencing technique was performed.

2.7.2. De novo prioritization pipeline

- *Automated de novo variant filtering and prioritization*

Patients with no identified pathogenic variants explaining the clinical phenotype following analysis using the DDG2P workflow were reanalyzed to screen for *de novo* variants in candidate novel DD genes as shown in Figure 2.2. African ancestry individuals have a high level of genetic diversity thus there is a higher likelihood of identifying novel DD genes in the patient cohort. Additionally, de novo variants account for about 40%-48% of DD cases (Deciphering Developmental Disorders Study, 2017).

Complete family trios were annotated using the bcftools trio-in plugin. The tool uses genotype likelihood scores to identify heterozygous variants in the patient. *De novo* inheritance of the variant was confirmed by prioritizing those where both parents presented with a homozygous reference allele. The trio VCF files (merged parents-patient VCF files) needed to identify and

prioritize de novo variants were generated using bcftools --merge option. Variants were annotated using Ensembl-VEP and the bcftools trio-dnm (Danecek et al., 2021). Variant quality filters were applied to remove false positive de novo variants.

DNM variant quality filters applied:

- DNM score ≤10, this is the log probability that the variant is a true *de novo* mutation (DNM). calculated using genotype likelihood scores. The calculated score is found in the trio-dnm. annotated output file.
- VAF ≤0.3 (30%), the fraction of reads supporting the DNM allele at that position. The score is outputted in the trio-dnm annotated output file.
- DP ≤10, this is the total read depth at the variant site.
- 99.9< sensitivity<100% VQSR threshold, recommended by the GATK workflow for the identification of true variants for both SNP and Indels.

De novo variants were prioritized to identify those with a predicted deleterious consequence on protein function and gene expression. Ensembl-VEP classifies variants according to consequence and these are described using sequence ontology standardized annotation terms (Figure 2.3). We prioritized variants:

- impacting function (including but not limited to, in-frame deletions, in-frame insertions, initiator codon variants, missense variants, and variants resulting in loss/gain of a stop codon)
- variants with a predicted LOF, (nonsense, frameshift, or essential splice site variants).

This filter parameter was not applied in the DDG2P pipeline as it is already implemented as part of the tools variant filter criteria. The variants were further prioritized using MAF, in-silico pathogenesis predictor algorithms, using the same parameters as described in the DDG2P workflow. Novel DD gene *de novo* variants were interpreted using the ACMG-AMP guidelines, and only likely pathogenic and pathogenic variants were prioritised.

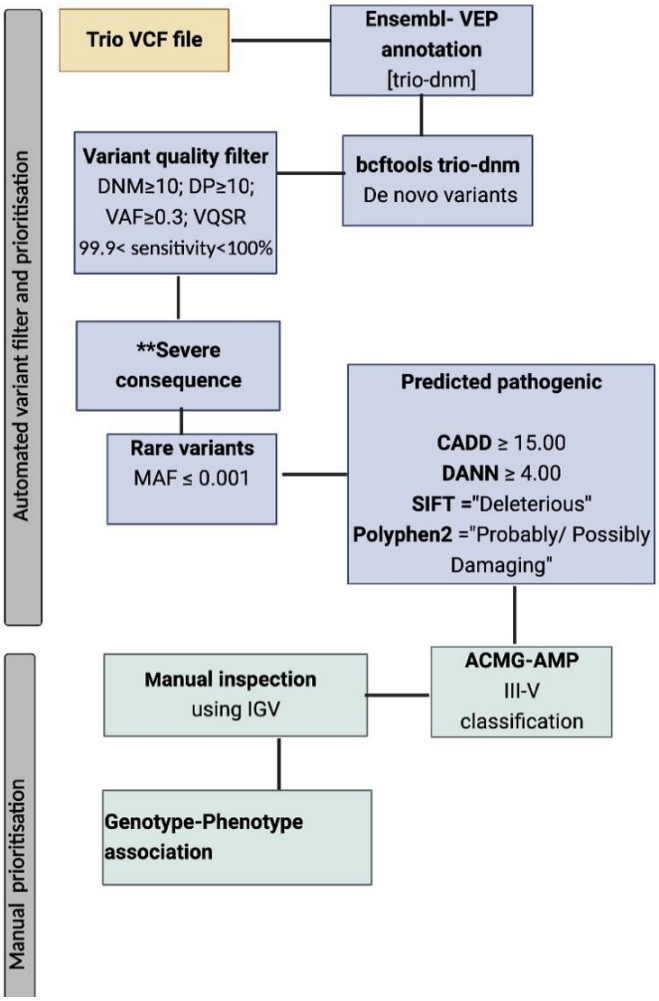

Figure 2.2. A semi-automated *de novo* variant filter and prioritization workflow. A trio VCF file containing merged variants from both parents and patient were annotated using Ensembl-VEP. De novo variants were identified using bcftools trio-dnm and possible true variants were identified using variant quality filters. Variants were prioritised to identify candidate de novo variants.**Severe consequence: variants impacting function (in-frame deletions, in-frame insertions, initiator codon variants, missense variants, and variants resulting in loss of a stop codon); variants with a predicted LOF, (nonsense, frameshift or essential splice site variants); ACMG-AMP, American College of Medical Genetics and Genomics and the Association for Molecular Pathology; IGV, Integrative Genome Viewer.

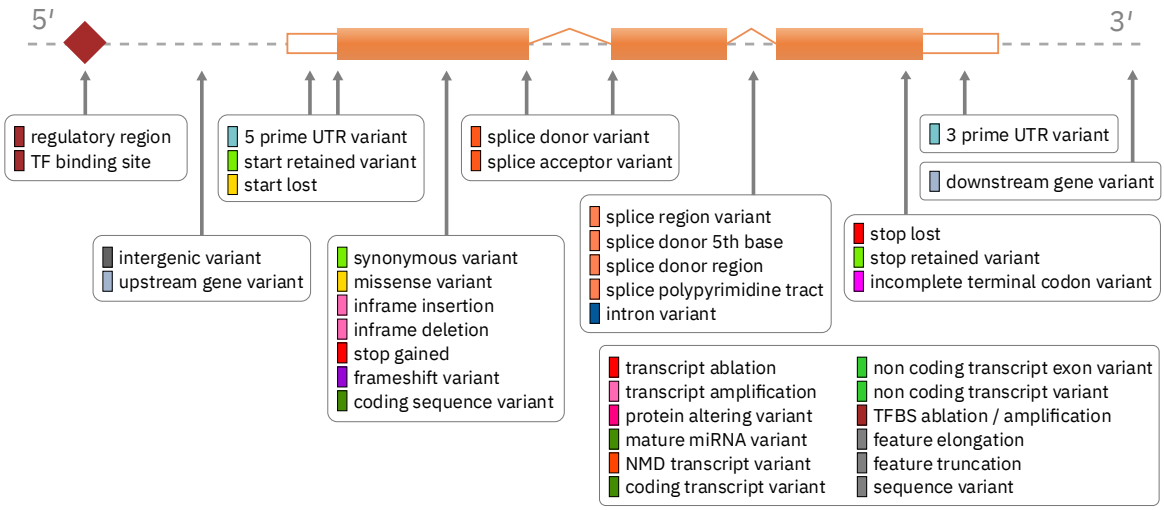

Figure 2.3. Standardised variant consequence terms represented according to gene location. Variants with a severe consequence on gene function and expression are located within the protein-coding region of the gene, and those both upstream of the start-codon (5' untranslated region (UTR) variants and start lost variants) and downstream of the stop-codon (3' untranslated region (UTR) variants). These variants result in amino acid changes; disrupt the splicing process involving pre-mRNA processing important for regulating gene and protein expression levels such as splice region variants, splice acceptor and donor variants; disrupt translation by affecting the binding or unbinding of the translational machinery. Taken from https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html, (accessed 01-07-2022).

2.8 Variant validation using Sanger sequencing

Variants prioritised as being the likely cause of the patient phenotype were confirmed using Sanger sequencing and segregation analysis was performed in family trios.

Sequencing primers were designed using Primer3web (v.4.1.0) and primer sequence specificity was evaluated using NCBI Primer-BLAST tool (Jian et al., 2012) to ensure that the primers would only bind to the target genomic location. In addition, primer sequences were assessed for the presence of common single-nucleotide polymorphism which may inhibit binding to their complementary DNA template using the web-based tool SNPCheck v3.0 (reference URL can be found in Appendix 1). Finally, the IDT OligoAnalyzer web-based tool (reference URL can be found in Appendix 1) was utilized to evaluate the presence of any predicted secondary hairpin structure and 3'-end primer-dimers that may inhibit PCR amplification. All primers were synthesised by Inqaba Biotec™ (Pretoria, South Africa).

Genomic DNA was PCR amplified using the AmpliTaq Gold™ DNA Polymerase with Buffer II and MgCl2 kit and purified using the Exo-SAP express kit (ThermoFisher Scientific, California, USA) according to the manufacturer's instructions.

Table 2.3. Polymerase Chain Reaction components along with their concentrations

PCR Reagents	1X reaction volume (µl)
ddH20	14.3
AmpliTaq Gold MgCl2 [25mM]	2.5
AmpliTaq Gold Buffer II [10X]	2.5
dNTP MIX [2.5 mM]	2.5
Forward primer [100ng/µl]	1
Reverse primer[100ng/µl]	1
DNA [100ng/ µl]	1
AmpliTaq Gold DNA Polymerase [5U/µl]	0.2
Total Volume	**25**

Cycle sequencing was performed using the BigDye v3.1 kit (ThermoFisher Scientific, California, USA), and purification and capillary sequencing using the BigDye Xterminator purification kit (ThermoFisher Scientific, California, USA) and ABI 3500xl Genetic Analyzer (ThermoFisher Scientific, California, USA). Finally, raw sequencing data was analysed using the CLC Bio software version 7.7.1 (Qiagen, Hilden, Germany).

Table 2.4. Polymerase Chain Reaction thermocycler conditions

Stage	Temperature (°C)	Time	Cycles
Initial Denaturation	95	5 minutes	1
denaturation	95	30 seconds	35
Annealing	55-62	30 seconds	
Extension	72	30 seconds	
Final extension	72	1 minutes	1
Hold	4	∞	-

Table 2.5. Cycle sequencing components

Component	[Final]	1x reaction volume (µl)
Purified PCR product	-	5
BigDye v3.1		2
BigDye buffer	0.5x	1
Primer (F or R)	100ng/µl	1
ddH2O	-	1
Total Volume	**-**	**10µl**

2.9 Variant reporting

Candidate variants and supporting gene-disease and variant-disease information were evaluated and curated at our in-house variant review meeting composed of a multidisciplinary team of clinicians, genetic counsellors, and scientists. Additionally, DECIPHER was used as a decision and clinical support tool within the DDD-Africa group. Data sharing was also facilitated using DECIPHER, as in addition to the patient clinical and demographic data, candidate variants and their associated ACMG-AMP classifications were recorded on the data repository.

Following in-depth discussions, variants determined to be disease-causing in the patient were prepared for reporting and feedback to the referring clinician in the form of a research report.

Chapter 3 - Results

3.1 Patient demographics and details

A total of 117 patients with unexplained developmental disorders and their unaffected parents were available from the larger DDD-Africa study cohort. Included were 65 complete family trios (mother, father, child), 46 parent-duos (father unavailable at recruitment), two proband-only (both parents not recruited), and two extended families (mother and two affected siblings; mother, father, and two affected siblings). No consanguinity was reported by the parents. Among the patients included in this study, 62 (53%) were male and 55 (47%) were female. The median age was seven years (range 1-49) at the time of recruitment (Table 3.1). The majority of patients identified as Black (90.6%, n=106) while 5.1% (n=6), 3.4% (n=4), and 0.9% (n=1) identified as white, admixed, and Indian respectively.

Before enrolment, all but three patients (114/117), had undergone an extensive diagnostic workup. Most patient had two diagnostic tests performed, (61.5%, n= 72/114) (Figure 3.1), with many patients in this group having undergone a combination of karyotype and MLPA testing, 76.3% (55/72) (Figure 3.2). Overall, karyotype (94/114;82.5%) and MLPA (101/114; 88.6%) testing were the most common genetic tests performed as part of routine testing in the DD cohort (Figure 3.2). These tests had not led to a diagnosis for any of the patients.

Table 3.1. Demographic data for patients with unexplained developmental disorders.

Variables	% or median (25th-75th percentile)	Value (range)	Number of patients
Gender			
Female	47.0	-	55
Male	53.0	-	62
Age (years)	7.0 (5.0 - 9.5)	1 – 49	117
Family structure			
Trio	56.5	-	65
**Duo	40.0	-	46
Proband-only	1.7	-	2
Extended	1.7	-	2
Ethnicity		-	
Black	90.6	-	106
White	5.1	-	6
Mixed	3.4	-	4
Indian	0.9	-	1

*Duo, patient and only one parental DNA sample available

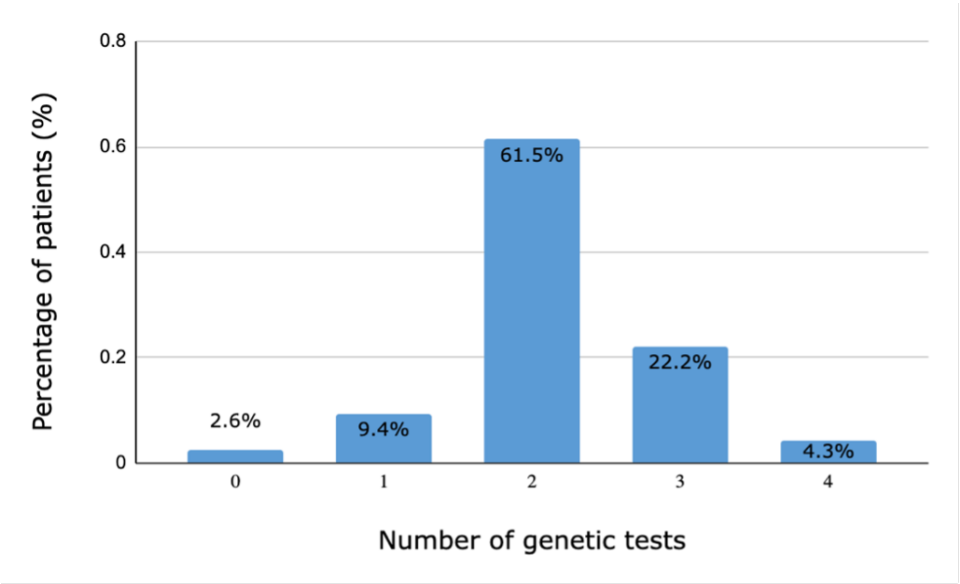

Figure 3. 1. The number of previous genetic tests undergone by patients in the DDD-Africa study. A total of 97.4% (114/117) of the patients included in the study had undergone diagnostic testing prior to study enrolment. Most patients (61.5%, 72/114) had two tests performed and 88.0% of the patient had two or more tests performed. No etiological diagnosis was achieved in any of the patients.

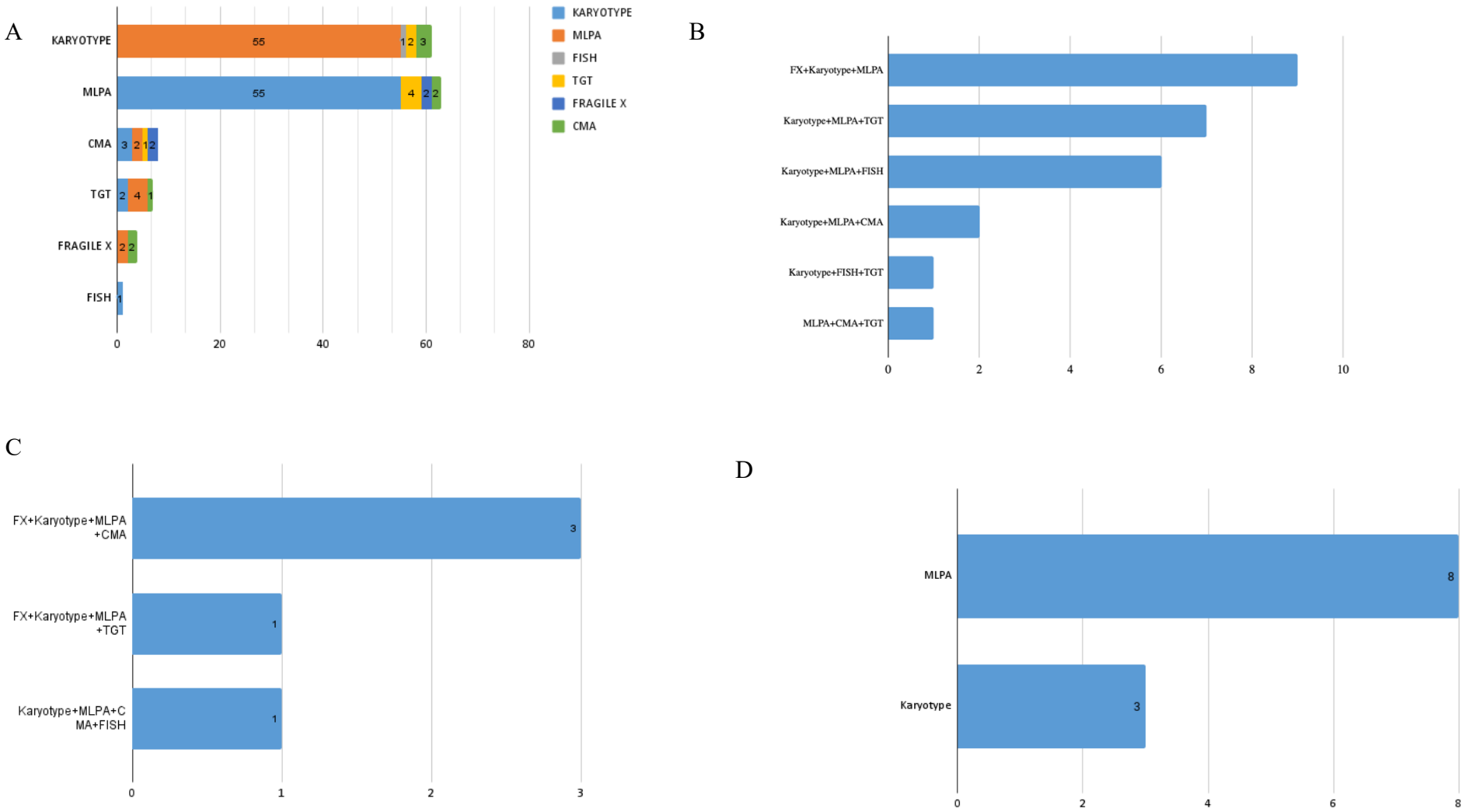

Figure 3.2. Specific genetic tests were undergone by DD patients prior to study enrolment.
Representation of the specific and number of genetic tests performed for each of the patients, (A) two genetic tests (B) three genetic tests. (C) four genetic tests (D) one genetic test. Overall, the majority (114; 48.2%) of the patients had undergone karyotype and MLPA (101/114;88.6%) testing which are routinely offered as part of first-line testing for patients with suspected DD. **Abbreviations**: FX, Fragile X; MLPA, Multiplex A ligation-dependent probe amplification; TGT, Targeted genetic test; CMA, Chromosomal microarray analysis; FISH, Fluorescence in Situ Hybridization)

3.2 Clinical features of patients and indications for study inclusion

Consistent patient clinical phenotyping was accomplished by classifying each patient's clinical features using the HPO system (Köhler et al., 2021; P. N. Robinson et al., 2008). A total of 1961 HPO terms (613 unique) across 20 organ systems were recorded for the cohort, with most recorded number of HPO terms per patient being 12 or 13 (range: 3-57) (Figure 3.3.); 84 (72%) children had developmental delay (moderate to severe), 35 (31%) had microcephaly, and 29 (25%) had delayed speech and language. Figure 3.4 provides a list of the top ten most observed clinical features in the DD cohort.

These results are in accordance with most patients, 35.0% (41/117) being recruited based on inclusion A only (Intellectual disability-moderate to profound) and 26.5% (31/117) recruited based on both inclusion A and C (one major malformation and 3 or more well-documented minor anomalies) of the DDD-Africa study (Figure 3.5). Overall, 65.0% (76/117) of the cohort were recruited based on the presence of major at least two malformations and anomalies in different organ systems.

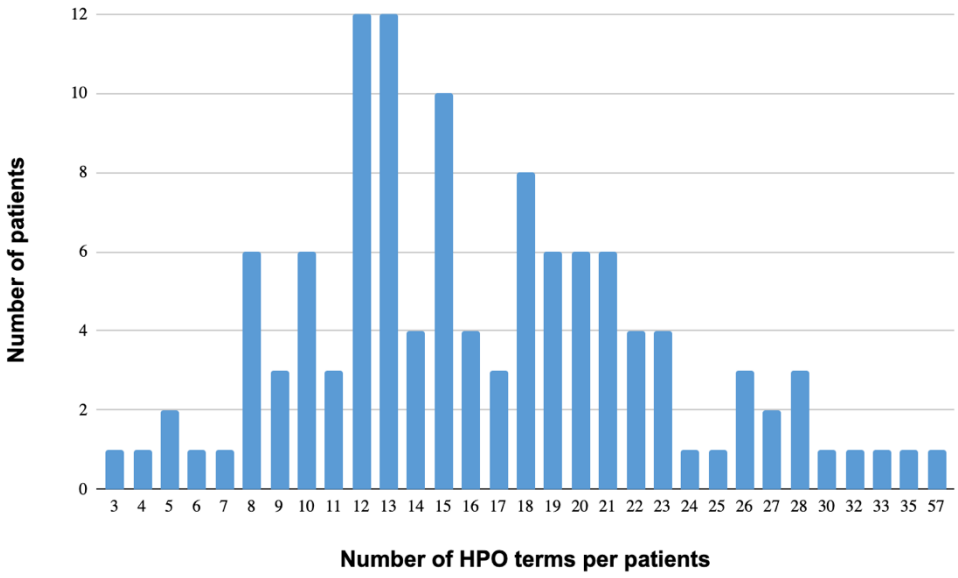

Figure 3.3. The number of HPO terms recorded for patients with unexplained developmental disorders. A mode of 12 and 13 clinical features were recorded using HPO terms for the 117 patients with unexplained developmental disorders.

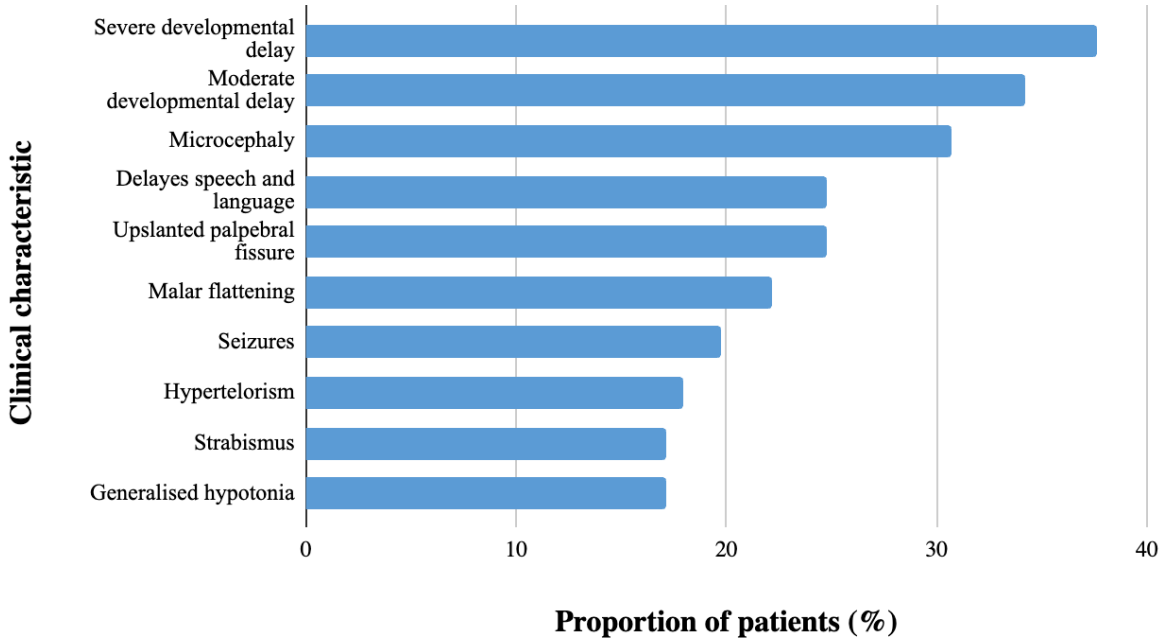

Figure 3.4. Top ten HPO terms observed among patients with unexplained developmental disorders. The most common clinical features recorded among the 117 patients included in this study were developmental delay (moderate to severe).

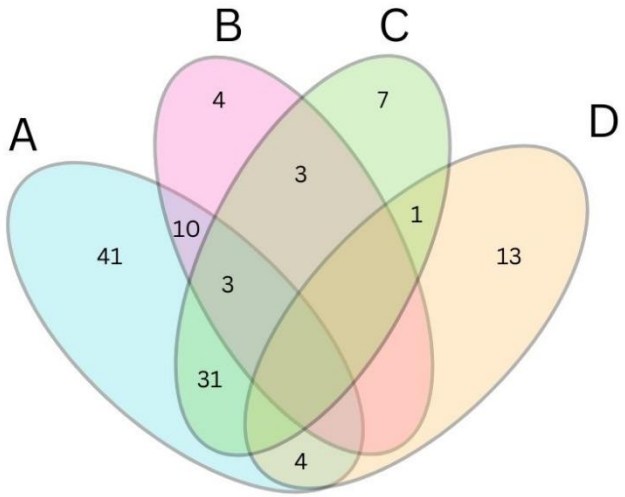

Figure 3.5. The number of DD patients recruited categorized based on the DDD-Africa inclusion criteria. Each of the patients were recruited based on a defined criteria, inclusion A (Intellectual disability (moderate to profound), inclusion B is the sum of inclusion B1 (Multiple major malformations in 2 or more different organ systems >1month) and B2 (Multiple major malformations in 2 or more different organ systems <1 month), inclusion C (One major malformation and 3 or more well-documented minor anomalies) and inclusion D (Mild to moderate intellectual disability and dysmorphic features). Most patients, 35.0% (41/117) were recruited based on inclusion A only and 26.5% (31/117) were recruited based on both inclusion A and C. Overall, 65.0% (76/117) of the cohort were recruited based on the presence of at least two malformations and anomalies in different organ systems.

3.3 WES and data quality measurements

Across the cohort, the average sequencing coverage achieved was 31.6x (Figure 3.6). This is slightly lower than the recommended mean coverage of 33x-35x which has been associated with a 95% accuracy rate in the detection of SNVs (Ahn et al., 2009; Ajay et al., 2011; Wang et al., 2008). However, this was acceptable as other cohorts have reported that even at a mean sequence coverage below the recommended, high quality true variants can still be accurately reported (Y. Li et al., 2011; Meynert et al., 2013).

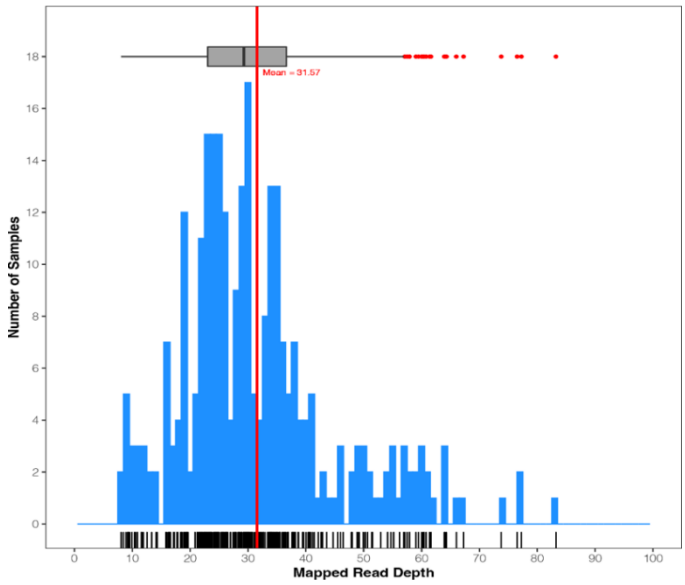

Figure 3.6. Mean sequence depth for DDD-Africa dataset. The mean depth coverage of whole exome sequencing samples included in this study from the DDD-Africa study.

The overall quality of the WES data was measured using Ts/Tv ratio, with an estimated ratio of ~2.6-2.8 expected for high quality WES sample datasets (Bainbridge et al., 2011). The quality metric indicates a level of accuracy in detecting true positive variants in the dataset (Q. Liu et al., 2012). An average Ts/Tv ratio of 2.3 (range; 2.2-2.4) was observed for the DDD-Africa WES dataset (Figure 3.7). This was following the application of the recommended GATK 99.9<sensitivity<100% VQSR threshold, aimed at removing false-positive variants in the unfiltered dataset.

These results indicated that our dataset contains possible false-positive variant calls which are likely a result of sequencing and variant calling errors, but also true variants that are poorly read covered. Taking this into consideration, stringent variant quality filters (VAF ≤0.3 and DP ≤10) to distinguish between false-positive variants and true variants were only applied downstream following variant filtering and prioritization.

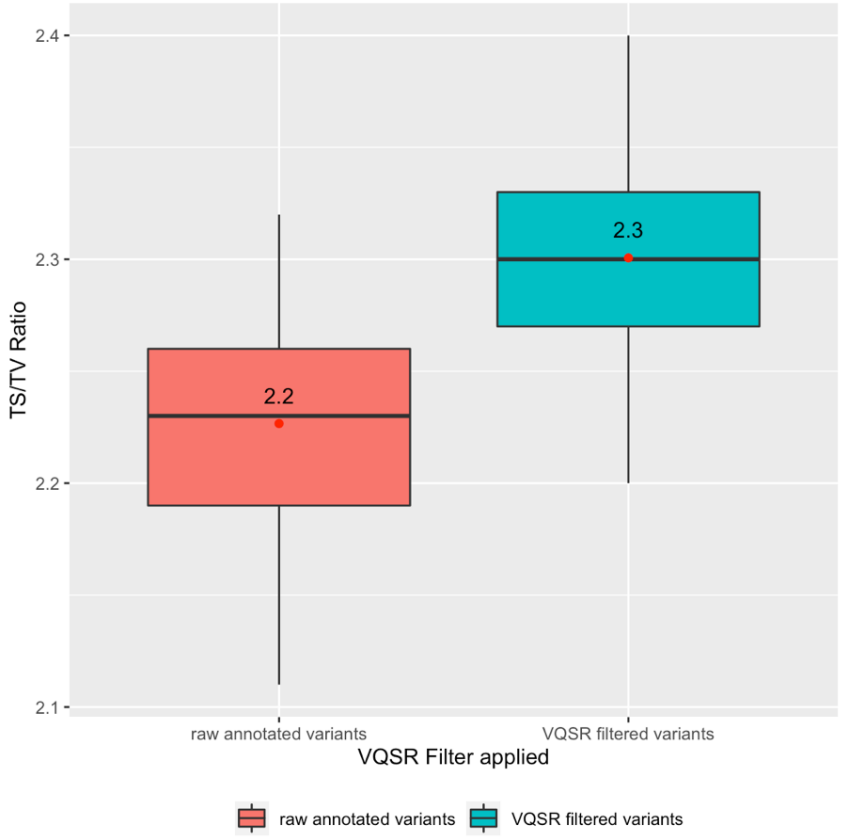

Figure 3.7. Ts/Tv ratio for DDD-Africa dataset. The overall quality of the WES dataset was evaluated using Ts/Tv ratio. We calculated the ratio for the raw dataset (Red) and for those variants passing the GATK best practices 99.9< sensitivity<100% VQSR threshold (Blue). The Ts/Tv ratio shows an improvement in the quality of the dataset. However, the value is lower than the expected ~2.6-2.8 for exome data, thus signalling the presence of false-positive variants in the dataset.

3.4 Variant annotation identifies a high proportion of known variants

Variants were annotated using the human reference genome build GRCh38/hg38. A total of 1,0693,973 genomic variants were identified across the patient cohort (mean= 91,401, range 69,466-104,879). Notably, of these, 97.5% (1,0428,603) were regarded as known as they had been reported before in an Ensembl variation dataset. Table 3.2 provides a more detailed description of the variant range.

Table 3.2. Summary of the number of variants identified in the DDD-Africa cohort. The number of variants identified using WES for 117 patients included in this study cohort.

	Total number (%)	Mean	Range (min-max)
Total Variants	10693973	91401	69466-104879
Known	10428603 (97.5)	67953	89133-101965
Novel	265370 (2.5)	2268	1513-2944

3.5 African dataset indicates a high percentage of very rare variants

We evaluated the allele frequency distribution of all variants identified among the 117 African patient samples included in this study. The allele frequency values were obtained from gnomAD reference population database accessed via Ensembl-VEP annotation. The distribution indicated that, many variants are very rare, clustering at an allele frequency of $1x10^{-6}$ as seen in figure 3.8(a). Moreover, many variants had been reported in version 147 of the dbSNP database (~9 million variants), and MAF distribution for these variants further highlighted that most variants in our African dataset are extremely rare even when previously reported (Figure 3.8(b)).

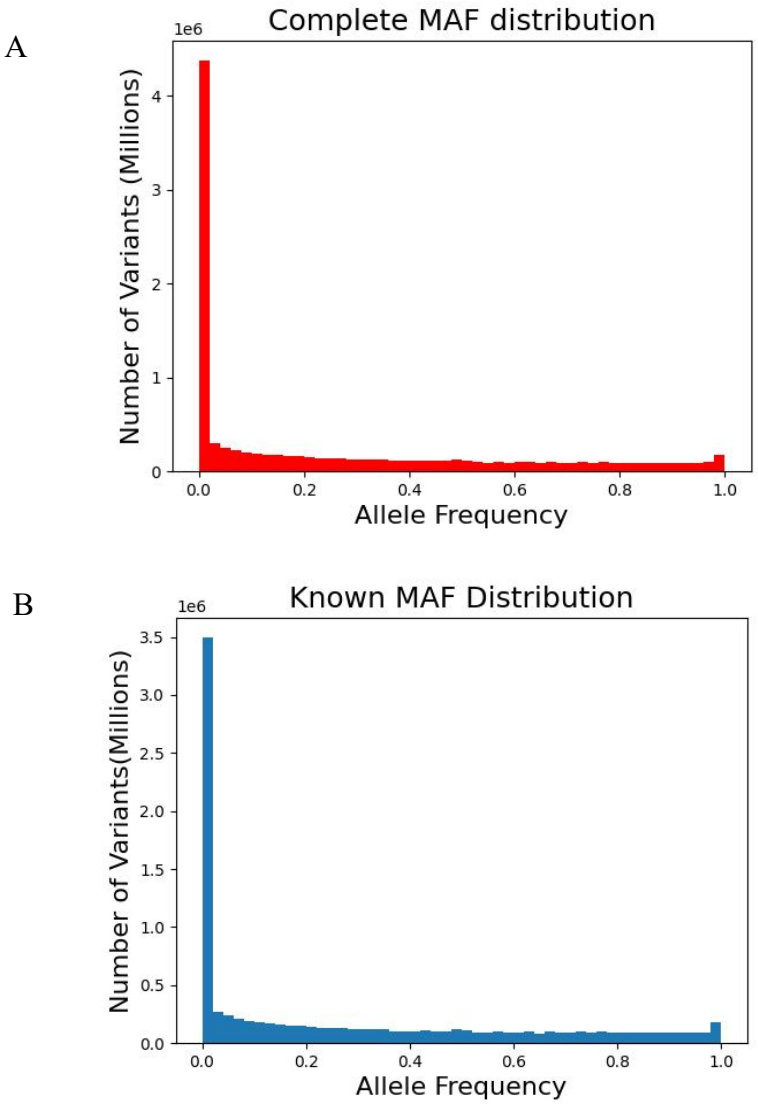

Figure 3.8. Allele frequency distribution of variants in the DDD-Africa dataset. (A) Allele frequency of all variants identified in the African dataset. (B) Allele frequency of variants reported in 147 version of dbSNP database. Most variants in the dataset are very rare clustering at approximately $1x10^{-6}$.

3.6 Filtering and prioritization of variants in known DD genes

Our first-tier study variant filtering and prioritizing strategy focused on identifying variants in known DD genes. This was accomplished using a semi-automated variant filtering and prioritization workflow to narrow down the variants to a pool of candidate variants. A total of 1,069,3973 (mean=91,401; range, 69,466-104,879) genomic variants were identified across the patient cohort, of which 408,169 (mean=34,886, range 24,369-43,636) had passed the VQSR quality control filter and were considered rare (Figure 3.9).

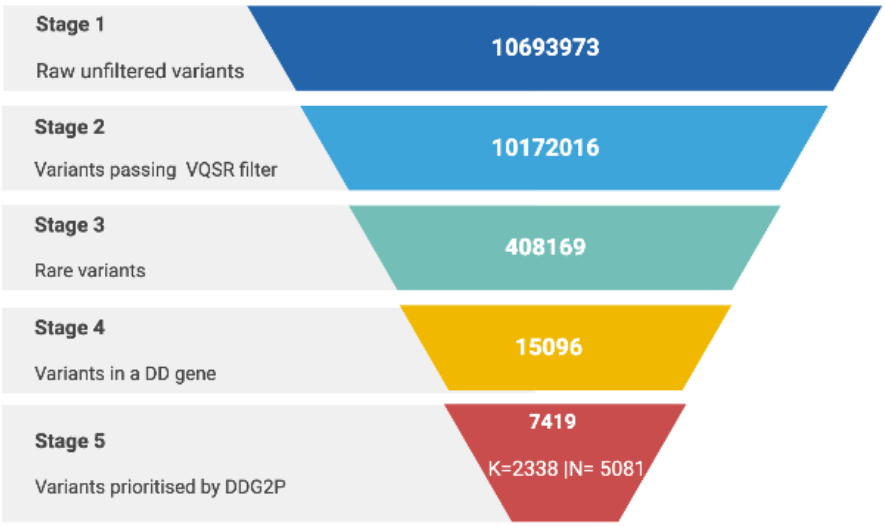

Figure 3.9. DD2P Variant filtering and prioritization strategy results. To restrict the analysis to only variants in known DD genes, variants were filtered and prioritized using Ensembl VEP-DDG2P. We first aimed to improve the quality of the unfiltered variant dataset by removing false-possible variants based on the GATK best practices 99.9< sensitivity<100% VQSR threshold. In a stepwise filtering process, we identified rare variants using MAF (≤ 0.01) and prioritized these using Ensembl-VEP DDG2P. This identified a total of 7419 variants of which 31.5% (n=2338) had been reported in ClinVar variant and disease database. These results suggested most of the variants in the dataset were novel. **Abbreviations: K**, known variants, reported in ClinVar, **N,** novel variants, not reported in ClinVar.

Furthermore, we prioritized DD-gene variants based on the Ensembl VEP-DDG2P criteria (Refer to Chapter 2, section 2.7.1 for a detailed description of the prioritization criteria). The tool assesses and prioritizes variants based on:

(i) Allele frequency to identify rare variants.

(ii) Severe consequence on protein function.

(iii) Fulfilment of the gene allelic requirement.

This yielded on average 63 variants per patient (n= 7,419, range 23-99). Interestingly, of these, 31.5% (2,338/7,419) had previously been reported in the ClinVar variant and disease database, while the remainder, 68.5% (5081/7419) (Figure 3.10) had not been previously reported.

3.6.1. Variant quality control and interpretation of variants using ACMG-AMP guidelines

Variant quality filter (VAF ≤ 0.3 and DP ≤ 10) cut-offs were applied to identify low-quality, possible false positive variants from a shortlist of variants. The majority (57.0% (n=2513/4407)) of the variants that were predicted to be pathogenic using *in-silico* pathogenesis predictor algorithms (Figure 3.10) had also passed quality control and were considered high-quality potentially true variants. A mean of 21 high-quality variants and 16 low quality variants was identified for the patient.

Manual variant interpretation of high-quality variants using the ACMG-AMP guidelines identified 88 pathogenic variants (ACMG class IV and class V) and 1,210 VUS (ClinVar variants = 567; novel variants= 643) (ACMG class III). The rest of the variants were classified as likely benign/benign (1,285 variants (ClinVar=385; novel=900)) or were considered common variants (18 variants), as they had been identified in more than four unrelated patients and were thus removed from further analysis (Figure 3.10 and Table 3.3).

Table 3.3. DD Cohort polymorphic variants annotated in ClinVar as pathogenic. Variants annotated in ClinVar as pathogenic were prioritised during variant interpretation. However, variants that were identified in four or more unrelated families were classified as polymorphic and thus not likely to cause disease.

# of patients	Gene	Variant	ClinVar	Class
5	*CEP290*	NM_025114.4: c.1666dup (p. Ile556AsnfsTer20)	Yes	Pathogenic
4	*SLC16A2*	NM_006517.5: c.407dup (p. Asn136LysfsTer31)	Yes	Pathogenic
5	*DSPP*	NM_014208.3: c.3262del (p. Ser1088ValfsTer226)	Yes	**Pathogenic
4	*ABCB6*	NM_005689.4: c.376del (p. Val126SerfsTer124)	Yes	**Pathogenic

**Pathogenic; at the time of analysis, (date accessed:(07.2021)), variants were annotated as pathogenic, however at the time of this write-up, these variant entries had been deleted from ClinVar. Class, American College of Medical Genetics and Genomics classification.

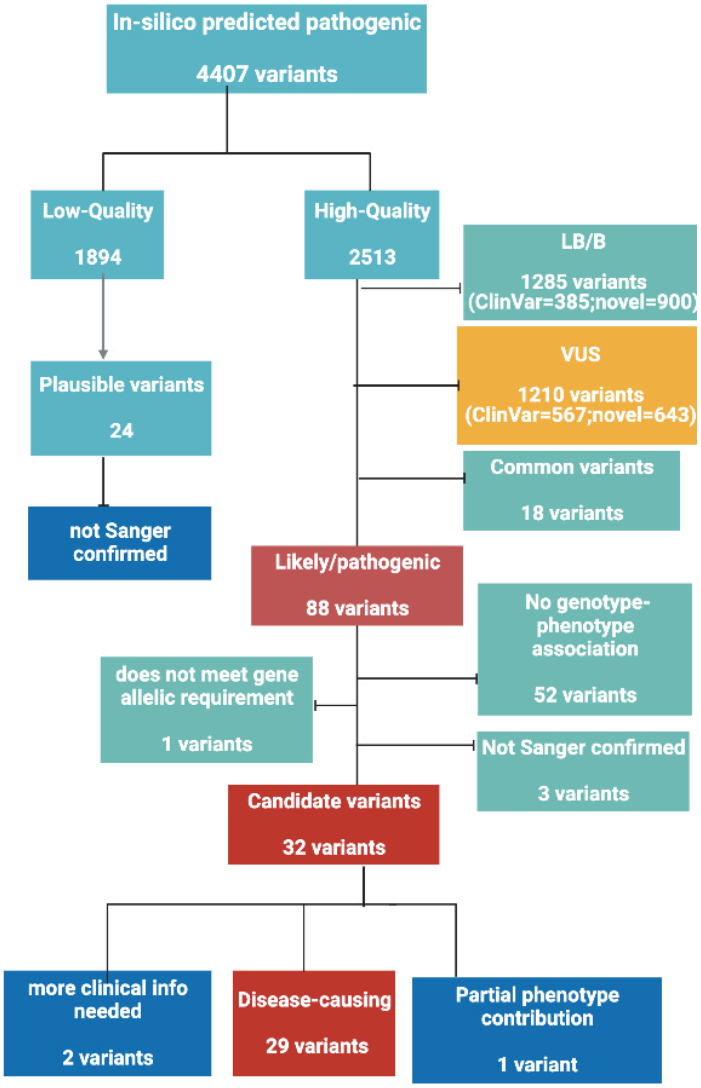

Figure 3.10. DDG2P Variant filtering and prioritization results. Variants were prioritised using *in-silico* pathogenesis predictor algorithms and variant quality filters. High-quality prioritised variants were interpreted using American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines and pathogenic variants shortlisted. Genotype-phenotype correlation identified 32 candidate variants. Further curation and interpretation by the DDD-Africa multidisciplinary team of experts (scientists, genetic counsellors, and clinicians) led to the identification of 29 disease-causing variants. Manual curation of low-quality variants based on genotype-phenotype resulted in the identification of 24 plausible variants which failed Sanger confirmation and determined to be false-positive variants.

3.7 Low-quality variants in plausible DD genes confirmed as false-positive

The results from WES mean sequence coverage analysis and the Ts/Tv quality analysis suggested that potentially true variants may accidentally be excluded as false-positive variants due to inadequate read coverage (section 3.3). Thus, we prioritised and shortlisted plausible low-quality variants in DD genes based on genotype-phenotype correlation and evaluated these using Sanger sequencing.

Overall, a total of 24 low-quality variants were prioritised as plausible based on the patient's phenotype (Table 3.4). These variants had a mean VAF of 27% (range:9-100) and DP of 22 (range:2-91). Of these, none had passed variant quality control based on both VAF and DP. None were confirmed using Sanger sequencing and thus were determined to be false-positive variants. These results were used to make an informed decision on the quality metrics and cut-offs to be used in this process. These variants were removed from further analysis.

Table 3.4. Low-quality variants prioritised in the DDD-Africa study. Variants in known DD genes determined to be low quality based on variant quality control filters (VAF ≤0.3 and DP ≤10) were prioritised as plausible contributors to the patient's phenotype. None of the variants were confirmed by Sanger sequencing and were thus false-positive variants.

ID	Gene	Variant description	DP (Ref; Alt)	VAF (%)	Confirmed
D3S.0001	*WDR37*	NM_014023.4: c.301G>T	26 (22;4)	15	No
D3S.0049	*SLC9A6*	NM_001400909.1: c.653G>T	4 (2;2)	50	No
D3S.0014	*ASH1L*	NM_001366177.2: c.8698dup	50 (44:6)	12	No
D3S.0023			81 (72;7)	9	No
D3S.0104			22 (18;4)	18	No
D3S.0040	*EYA1*	NM_001370335.1: c.620G>T	6 (4;2)	33	No
D3S.0044	*NEK1*	NM_001374418.1: c.397-1G>T	4 (2;2)	50	No
D3S.0044	*BPTF*	NM_00138777.11: c.2598del	13 (10;3)	23	No
D3S.0057			13 (11;2)	15	No
D3S.0121			11 (9;2)	18	No
D3S.0046	*CDKL5*	NM_001323289.2: c.2269G>T	12 (10;2)	17	No
D3S.0048	*BRAF*	NM_00138537.3: c.2248-1G>T	10 (8;2)	20	No
D3S.0062	*TAF1*	NM_00118085.1: c.3960del	5 (4;1)	20	No
D3S.0063	*CHD8*	NM_001170629.2: c.1600A>C	57 (44;13)	23	No
D3S.0068		NM_001170629.2: c.5173G>T	11 (9:2)	18	No
D3S.0069	*CSPP1*	NM_024790.6: c.2499+1G>C	20 (17;3)	15	No
		NM_024790.6: c.2499+2T>C	20 (16;4)	20	No
D3S.0078	*MCPH1*	NM_024596.5: c.221T>G	30 (22;8)	27	No
		NM_024596.5: c.218T>G	30 (22;8)	27	No
D3S.0087	*OPHN1*	NM_002547.3: c.201del	6 (4;2)	33	No
D3S.0095	*ASXL1*	NM_015338.6: c.1549C>T	91 (78;13)	13	No
D3S.0099	*ZNF292*	NM_015021.39: c.6145dup	13 (11;2)	15	No
D3S.0100	*NFIB*	NM_001369458.1: c.1688C>A	10 (8;2)	20	No
D3S.0101	*CSNK2A1*	NM_001895.4: c.340G>T	22 (19;3)	14	No
D3S.0106	*PAX3*	NM_181458.4: c.805A>C	25(20;5)	20	No
D3S.0112	*CDK8*	NM_001260.3: c.955G>T	20(16;4)	16	No
D3S.0121	*BLM*	NM_001287248.2: c.3267del	2(0;2)	100	No

3.8 Identification and manual curation of candidate variants

Candidate variants that failed to be confirmed using Sanger sequencing were excluded from further consideration. A total of 88 pathogenic variants (likely pathogenic and pathogenic) variants in 48 patients were shortlisted. Manual curation of these variants resulted in the identification of 52 variants which at the time of analysis were determined to not explain or contribute to the patient's clinical phenotype and were thus removed from further consideration (Figure 3.10).

Three pathogenic candidate variants in *THOC2, EMHT1* and *CHAMP1* were identified in three patients. All variants passed the internal variant quality control using VAF ≤0.3 and DP ≤10. However, none were confirmed using Sanger sequencing and were thus removed from further analysis. The WES IGV and Sanger sequencing electrogram can be found in Appendix 4.

A pathogenic heterozygous missense variant, in *NEK1* gene, was shortlisted in patient D3S.0044 and Sanger confirmed (WES IGV screenshot and Sanger sequencing electrogram in Appendix 4). The gene is associated with autosomal recessive Short-rib Polydactyly Syndrome Type Majewski (Chen et al., 2012). A second heterozygous missense variant was identified in the same patient but was determined to be of low quality due to low DP (VAF= 0.5, DP= 4). A wildtype sequence was observed at the position using Sanger sequencing and as such the variant and gene were removed from further consideration (The WES IGV screenshot and Sanger sequencing electrogram can be found in Appendix 4).

The remainder of the 32 variants were selected for further manual interpretation by the DDD-Africa multi-disciplinary team.

Table 3.5. A summary of the putative disease-causing variants identified in the DD patient.

Patient	Sex	Gene (Transcript)	Variant	Consequence	Inheritance	ClinVar ID	ACMG-AMP Code	ACMG-AMP class	Phenotype (OMIM ID)
Known gene-Known variant									
D3S.0013	M	*MTOR* (NM.004958.4)	c.5395G>A p. (Glu1799Lys)	Missense	*de novo*	VCV000217823.16	PP5, PP3, PM3, PM6	Pathogenic	Smith-Kingsmore (616638)
D3S.0028	M	*PPP2R1A* (NM_014225.6)	c.773G>A p. (Arg258His)	Missense	*de novo*	VCV000217458.4	PP5, PP3, PM2, PM6	Pathogenic	Intellectual disability (605983)
D3S.0037	F	*C12orf57* (NM_138425.4)	c.184C>T p. (Gln62Ter)	Stop-gain	Maternal	RCV000143978.3	PP5, PVS1, PM2, PM6	Pathogenic	Temtamy syndrome (218340)
D3S.0051	M	*CHD7* (NM_017780.4)	c.5833C>T p. (Arg1945Ter)	Stop-gain	*de novo*	VCV000267434.3	PP5, PVS1, PM2	Pathogenic	CHARGE syndrome (300860)
D3S.0052	M	*MECP2* (NM_001110792.2)	c.961G>A p. Arg321Trp	Missense	maternal	VCV000143749.18	PP5, PM2, PP3, PP2, PM1	Pathogenic	Rett Syndrome (312750)
D3S.0053	M	*AUTS2* (NM_015570.4)	c.376C>T p. (Arg126Ter)	Stop-gain	*de novo*	VCV000620261.1	PP5, PVS1, PP2, PM2, PM6	Pathogenic	Syndromic ID (61584)
D3S.0056	M	*PLP1* (NM_001128834.2)	c.44C>T p. Pro15Leu	Missense	maternal**	VCV000011075.1	PP5, PM1, PM2, PP2, PP3	Pathogenic	Pelizaeus-Merzbacher (312080)
D3S.0064	M	*SMAD4* (NM_005359.6)	c.1498A>G p. Ile500Val	Missense	maternal	VCV000030150.13	PP5, PP3, PP2, PM2	Pathogenic	Juvenile polyposis syndrome (175050) Myhe syndrome (139210)
D3S.0071	F	*EHMTI* (NM_024757.5)	c.2712+1G>A	Splice donor	*de novo**	VCV000374034.4	PVS1, PM2, PP5	Pathogenic	Kleefstra syndrome 1 (610253)
D3S.0082	M	*SON* (NM_138927.4)	c.5753_5756del p. (Val1918GlufsTer87)	Frameshift	*de novo**	RCV001093465.2	PVS1, PM2, PP5	Pathogenic	ZTTK syndrome (617140)

Table 3.5 (continued**).** A summary of the putative disease-causing variants identified in the DD patients.

Patient	Sex	Gene (Transcript)	Variant	Consequence	Inheritance	ClinVar ID	ACMG-AMP code	ACMG-AMP class	Phenotype (OMIM ID)
D3S.0086	M	*STXBP1* (NM_003165.6)	c.416C>T p. (Pro139Leu)	Missense	*de novo*	RCV000416131.3	PVS1, PM2, PM6, PP5	Pathogenic	Epileptic encephalopathy, early infantile (612164)
D3S.0094	F	*DDX3X* (NM_001356.5)	c.931C>T p. (Arg311Ter)	Stop-gain	*de novo**	VCV00098152.3	PP5, PVS1, PM2	Pathogenic	DDX3X syndrome (300958)
D3S.0098	F	*OFD1* (NM_003611.3)	c.710dup p. (Tyr238fs)	Frameshift	N/A	VCV000041143.1	PVS1, PP5, PM2	Pathogenic	Oral-Facial-Digital Syndrome Type 1 (311200)
D3S.0118	F	*FLNA* (NM_001110556.2)	c.620C>T p. (Pro207Leu)	Missense	maternal	VCV000011755.3	PM1, PP2, PP3, PM2, PP5	Pathogenic	Terminal Osseous Dysplasia (300244)
D3S.0124	F	*PACS1* (NM_018026.4)	c.607C>T p. (Arg203T)	Missense	*de novo**	VCV000039581.20	PVS1, PM2, PP5	Pathogenic	Schuurs-Hoeijmakers syndrome (615009)
D3S.0129	M	*MAPK1* (NM_002745.5)	c.952G>A p. (Asp318As)	Missense	*de novo**	RCV001264764.1	PVS1, PM2	Pathogenic	Noonan syndrome 13 (619087)
Known gene-novel variant									
D3S.0015	M	*NFIX* (NM_001365902.3)	c.293del p.(Lys106ArgfsTer37)	Frameshift	*de novo*	Novel	PM6, PVS1, PM2	Pathogenic	Malan syndrome (614753)
D3S.0033	M	*UBE2A* (NM_003336.4)	c.129dup p. (Glu44Ter)	Frameshift	maternal	Novel	PVS1, PM2	Pathogenic	Syndromic XL ID (300860)
D3S.0034	M	*CHD7* (NM_017780.4)	c.8528del p. (Gly2843GlufsTer46)	Frameshift	*de novo*	Novel	PM6, PVS1, PM2	Pathogenic	CHARGE syndrome (300860)

Table 3.5 (continued). A summary of the putative disease-causing variants identified in the DD patients.

Patient	Sex	Gene (Transcript)	Variant	Consequence	Inheritance	ClinVar ID	ACMG-AMP code	ACMG-AMP class	Phenotype (OMIM ID)
D3S.0035	M	*OCRL* (NM_001318784)	c.2399C>T p. (Pro799Leu)	Missense	maternal	Novel	PM1, PM2, PP3, PM5	Likely Pathogenic	Lowe Syndrome (309000)
D3S.0047	F	*ARID1B* (NM_00137482)	c.2819del p. (Pro927GlnfsTer57)	Frameshift	*de novo*	Novel	PVS1, PM2, PM6	Pathogenic	Coffin-Siris syndrome (135900)
D3S.0050	F	*CHD7* (NM_017780.4)	c.232C>T p. (Gln78Ter)	Stop-gain	*de novo*	Novel	PVS1, PM2, PM6	Pathogenic	CHARGE syndrome (300860)
D3S.0058	F	*MFSD8* (NM_152778.3)	c.1350+1del	Splice-acceptor	paternal, maternal	Novel	PVS1, PM2	Pathogenic	Ceroid-Lipofuscinosis, Neuronal 7 (610951)
D3S.0081	M	*SCN2A* (NM_001040142.2)	c.727C>T p. (Gln243Ter)	Stop-gain	*de novo*	Novel	PM6, PVS1, PM2	Pathogenic	Developmental, Epileptic Encephalopathy (613721)
D3S.0089	M	*MEF2C* (NM_001193347.1)	c.963T>A p. (Tyr275Phe)	Stop-gain	*de novo**	Novel	PVS1, PM2	Likely Pathogenic	NDD with hypotonia, and impaired language (613443)
D3S.0108	F	*XYLT2* (NM_022167.4)	c.1703_1721del p. (His568ProfsTer33)	Frameshift	paternal, maternal	Novel	PVS1, PM2	Pathogenic	Spondyloocular Syndrome (605822)
D3S.0111	F	*PTEN* (NM_001304717.5)	c.1129C>A p. (Thr534Ala)	Missense	de novo*	Novel	PVS1, PM2	Likely pathogenic	Macrocephaly/Autism syndrome (605309)
D3S.0117	M	*NSD1* (NM_022455.5)	c.4913A>G p. (His1638Arg)	Missense	*de novo*	Novel	PM6, PM1, PM2, PP3	Likely pathogenic	Sotos syndrome (117550)

D3S.0119	M	*RPS6KA3* (NM_004586.3)	c.734A>G p. (Val601Glu)	Missense	d*e novo*	Novel	PM6, PP3, PM2, PM1	Likely pathogenic	Coffin-Lowry syndrome (303600)

de novo*; Heterozygous in affected proband, however DNA of one of the unaffected parents(s) unavailable to confirm. ***maternal; Hemizygous in affected proband, however DNA of mothers DNA unavailable to confirm. N/A; both parents unavailable to confirm inheritance and multiple inheritance patterns have been associated with the disorder. **Abbreviations: Class, American College of Medical Genetics and Genomics classification; ID, intellectual disability; XL, X-linked; Known: variant has been previously reported and associated with disease in ClinVar database; Novel: variant has not previously been reported or associated with disease in ClinVar database

3.9 Screening for *de novo* variants in DD

Literature evidence suggests that *de novo* variants account for most DD cases (Deciphering Developmental Disorders Study, 2017; Vissers et al., 2010). For patients with no disease-causing variants identified, screening for *de novo* variants in novel genes, was accomplished using the trio-dnm algorithm tool through the implementation of a semi-automated *de novo* variant prioritization pipeline.

Sequencing data was available for a total of 63 complete trios from 62 families (Family D3S.0033, is an extended family, of mother, father, and two male siblings). Of these families *de novo* pathogenic variants explaining the patient phenotype had been identified in 16. These families were used as positive controls in the *de novo* trio analysis.

On average, 1,383 variants were predicted to be *de novo*, per patient. This average number was extremely high compared to the ~70 *de novo* variants expected per genome (Sasani et al., 2019). Despite this extremely high number of identified variants, the variants were filtered and prioritised to a shortlist of pathogenic variants using population allele frequency, *in-silico* pathogenesis predictor algorithms variant consequences as detailed in the methodology chapter.

An average of 25 *de novo* variants were shortlisted per family. Evaluation of the data showed that four families were outliers with a much higher number of variants identified compared to the others (Figure 3.11).

Manual interpretation using standard ACMG-AMP guidelines identified 27 *de novo* pathogenic variants in 26 patients (Table 3.6). A total of 17 of the 27 *de novo* variants (63.0%) overlapped known DD genes (defined as genes part of the DDG2P virtual gene list). All variant positive controls (*de novo* variants prioritised using the DDG2P pipeline), were also prioritised through this *de novo* pipelines (n=16/27).

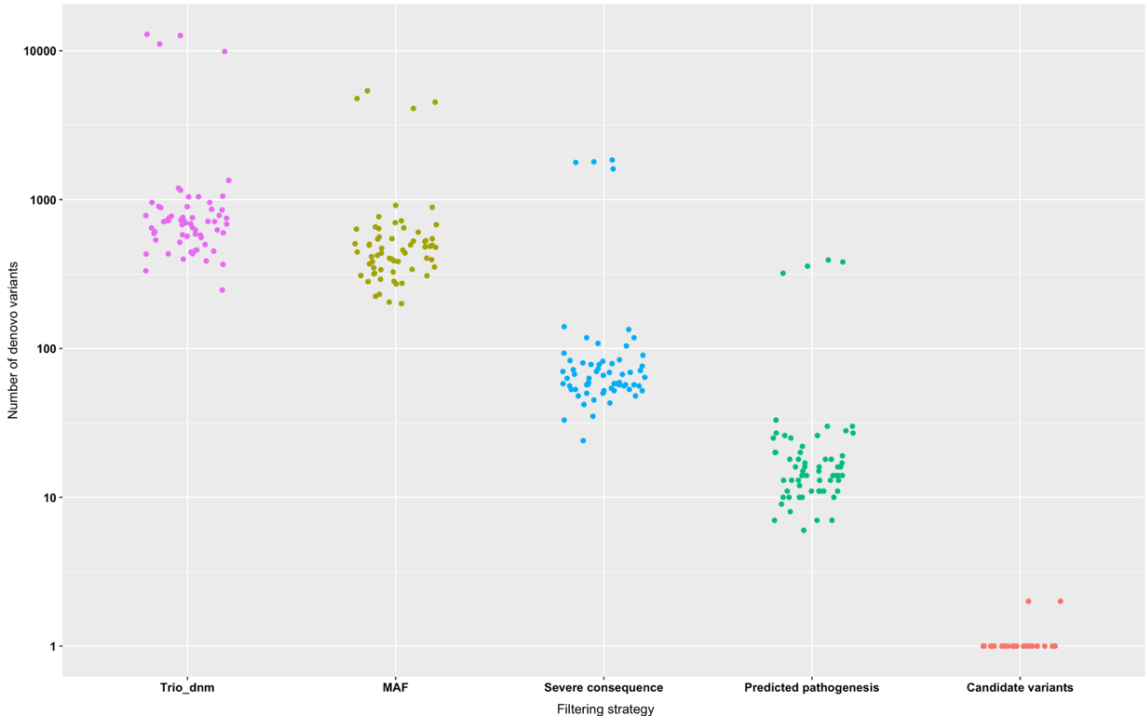

Figure 3.11. The number of *de novo* variants after each filter strategy. A total of 63 trios were available and included in the *de novo* variant analysis. A total of 27 candidate variants were identified in 26 trios. Four outlier trios with a higher number of variants in the dataset can be seen. Purple (Trio-DNM); *de novo* variants identified using trio-dnm tool and subsequent variant quality control; Yellow, variants were filtered using minor allele frequency (MAF); Blue, variants were prioritised to identify those with severe consequence on protein function; Green, variants predicted to be pathogenic using *in-silico* predictor pathogenesis tools; Red, pathogenic variants identified.

Table 3.6. *De novo* prioritised variants identified using the trio-dnm prioritization strategy. *De novo* variants were identified using trio-DNM and further prioritised to identify candidate DD variants. Literature evidence, and gene-phenotype correlation is provided for each identified candidate gene.

Patient ID	Gene (RefSeq ID)	Variant	Gene Function	*Gene constraint metrics	Relevant phenotype (OMIM ID)	**DD Research gene
Known DDG2P gene (Solved cases)						
D3S.0013	*MTOR* (NM_004958.4)	c.5395G>A p. (Glu1799Lys)	Role in mTOR signalling pathway for cellular processes such as cell proliferation, survival and apoptosis (Hua et al., 2019)	pLI = 1.00 missense Z-score= 7.02	Smith-Kingsmore syndrome (616638)	DDG2P
D3S.0015	*NFIX* (NM_001365902.3)	c.293del p. (Lys106ArgfsTer37)	Involved in transcriptional regulator and the differentiation of neural progenitor cells. Differentially expressed in brain tissue (Harris et al., 2016)	pLI = 1.00 missense Z-score= 4.08	Malan syndrome (614753)	DDG2P
D3S.0028	*PPP2R1A* (NM_014225.6)	c.773G>A p. (Arg258His)	Negative control of cell growth and cell death (Janssens & Goris, 2001)	pLI = 0.98 missense Z-score = 4.46	PPP2R1A- Intellectual disability (605983)	DDG2P
D3S.0034	*CHD7* (NM_017780.4)	c.8528del p. (Gly2843GlufsTer46)	Involved in cell cycle and development, and expressed predominantly in the nucleolus and nucleoplasm (Zentner et al., 2010)	pLI = 1.00 missense Z-score = 3.22	CHARGE syndrome (300860)	DDG2P
D3S.0047	*ARID1B* (NM_001374820.)	c.2819del p. (Pro927GlnfsTer57)	Involved in chromatin remodelling in gene expression and differentiation (Santen et al., 2012).	pLI = 1.00 missense Z-score = 2.59	Coffin-Siris syndrome (135900)	DDG2P

Table 3.6 (continued). *De novo* prioritised variants identified using the de novo prioritization strategy workflow. De novo variants were identified using the Trio-DNM and further prioritised to identify candidate DD variants. Literature evidence, and gene-phenotype correlation is provided for each identified candidate gene.

Patient ID	**Gene (RefSeq ID)**	**Variant**	**Gene Function**	***Gene constraint metrics**	**Relevant phenotype (OMIM ID)**	****DD Research gene**
Known DDG2P gene (Solved cases)						
D3S.0050	*CHD7* (NM_017780.4)	c.232C>T p. (Gln78Ter)	Involved in cell cycle and development, and expressed predominantly in the nucleolus and nucleoplasm (Zentner et al., 2010)	pLI = 1.00 missense Z-score = 3.22	CHARGE syndrome (300860)	DDG2P
D3S.0051	*CHD7* (NM_017780.4)	c.5833C>T p. (Arg1945Ter)	Involved in cell cycle and development, and expressed in the nucleolus and nucleoplasm (Zentner et al., 2010)	pLI = 1.00 missense Z-score = 3.22	CHARGE syndrome (300860)	DDG2P
D3S.0053	*AUTS2* (NM_015570.4)	c.376C>T p. (Arg126Ter)	Involved in several cell processes including, neuronal differentiation (Monderer-Rothkoff et al., 2021), cytoskeleton regulation (Hori et al., 2014) and synapse function (Hori et al., 2020)	pLI = 1.00 missense Z-score = 2.22	Syndromic ID (615834)	DDG2P
D3S.0081	*SCN2A* (NM_001040142.2)	c.727C>T p. Gln243Ter	Critical in the initiation and conduction of action potentials (Bender & Trussell, 2012)	pLI = 1.00 missense Z-score = 6.46	Developmental, Epileptic Encephalopathy (613721)	DDG2P

Table 3.6 (continued). *De novo* prioritised variants identified using the *de novo* prioritization strategy workflow. *De novo* variants were identified using the Trio-DNM and further prioritised to identify candidate DD variants. Literature evidence, and gene-phenotype correlation is provided for each identified candidate gene.

Patient ID	Gene (RefSeq ID)	Variant	Gene Function	*Gene constraint metrics (90% CI)	Relevant phenotype (OMIM ID)	**DD Research gene
Known-DD gene (Solved cases)						
D3S.0086	*STXBP1* (NM_003165.6)	c.416C>T p. (Pro139Leu)	Involved in neurotransmitter release through syntaxis regulation (Rizo & Xu, 2015)	pLI = 1.00 missense Z-score = 4.26	Epileptic encephalopathy, early infantile (612164)	DDG2P
D3S.0117	*NSD1* (NM_022455.5)	c.4913A>G p. (His1638Arg)	Acts as both a co-repressor and a co-activator transcriptional factor during post-translational modifications of the epigenome (Kurotaki et al., 2001)	pLI = 1.00 missense Z-score = 3.41	(Sotos syndrome (117550)	DDG2P
D3S.0119	*RPS6KA3* (NM_004586.3)	c.734A>G p. (Val601Glu)	Involved in cell growth and differentiation (Romeo et al., 2012).	pLI = 1.00 missense Z-score = 4.52	Coffin-Lowry syndrome (303600)	DDG2P
D3S.0014	*SETBP1*	c.2612T>C (p. Ile871Thr)	Binds DNA to increase gene expression in the production of proteins (Whitlock et al., 2023)	pLI =1.00 missense Z-score = 1.1	Schinzel-Giedion syndrome (269150)	DDG2P
D3S.0038	*CHAMP1* (NM_032436.4)	c.775_782del: p. (Pro259SerfsTer16)	Involved in the maintenance of microtubule attachment on the spindle in mitosis (Okamoto et al., 2017)	pLI = 0.99 missense Z-score = 0.14	CHAMP1- related NDD (616579)	DDG2P
D3S.0043	*CDH2* (NM_001792.4)	c.1885G>C	Essential role in the early steps of brain morphogenesis (Suzuki & Takeichi, 2008)	pLI = 0.99 missense Z-score = 2.09	Agenesis of corpus callosum, cardiac, ocular, and genital syndrome (618929)	DDG2P

D3S.0049	*PLCE1* (NM_001165979)	c.3466-1G>T	Regulation of Rho GTPases and MAPKs in podocytes (S. Yu et al., 2020)	pLI = 0.00 missense Z-score = 1.38	Early-onset nephrotic syndrome (610725)	DDG2P
D3S.0104	*KDM5B* (NM_001314042.1)	c.1725-1G>T	Role in normal embryonic development and for proliferation and differentiation	pLI = 0.00 missense Z-score = 1.78	Autosomal recessive ID (618109)	DDG2P

Table 3.6 (continued). *De novo* prioritised variants identified using the *de novo* prioritization strategy workflow. *De novo* variants were identified using the trio-dnm and further prioritised to identify candidate DD variants. Literature evidence, and gene-phenotype correlation is provided for each identified candidate gene.

Patient ID	**Gene (RefSeq ID)**	**Variant**	**Gene Function**	***Gene constraint metrics (90% CI)**	**Relevant phenotype (OMIM ID)**	****DD Research gene**
Potential DD gene						
D3S.0011	*WASHC2A* (NM_001005751)	c.2695G>T (p. Glu899Ter)	NI	pLI = 0.00 missense Z-score = 0.66	NI	No
D3S.0012	*TNRC18* (NM_001080495.3)	c.4662-2A>G	Function is unclear but may enable chromatin binding activity, with ubiquitous expression in tissue (Virdee, 2022)	pLI = 0.99 missense Z-score = 0.26	NI	Yes
D3S.0017	*MYCBP2* (NM_015057.5)	c.6788C>A (p. Ser2263Ter)	Axon guidance and synapse formation in the developing nervous system. Also regulates mTOR signalling pathway (AlAbdi et al., 2022)	pLI = 1.00 missense Z-score = 6.05	NI	Yes
D3S.0018	*DPP4* (NM_001379604.1)	c.1988-1G>T	NI	pLI = 0.00 missense Z-score = 0.79	NI	No
D3S.0023	*CLIP4* (NM_001287527.2)	c.946G>T (p. Glu316Ter)	Possible involvement in cytoplasmic microtubule organization (Howard & Hyman, 2003)	pLI = 0.00 missense Z-score = 0.45	NI	Yes

D3S.0029	*TARBP1*	c.2394+1G>A	Binds to HIV RNA polymerase II and involve in infection (Weiß et al., 2021)	pLI = 0.00 missense Z-score = 0.43	NI	No
D3S.0049	*SNX4* (NM_003794.4)	c.598-1G>T	involved in the regulation of endocytosis and intercellular trafficking and recycling.	pLI = 0.00 missense Z-score = 0.89	hereditary ataxia (not available)	Yes
D3S.0075	*BRWD*1 (NM_033656.4)	c.1659+1G>C	May play a role in the regulation of cell morphology and cytoskeletal organization (Mandal et al., 2018)	pLI = 1.00 missense Z-score = 2.85	NI	Yes

Table 3.6 (continued). *De novo* prioritised variants identified using the *de novo* prioritization strategy workflow. *De novo* variants were identified using the trio-dnm and further prioritised to identify candidate DD variants. Literature evidence, and gene-phenotype correlation is provided for each identified candidate gene.

Patient ID	**Gene (RefSeq ID)**	**Variant**	**Gene Function**	***Gene constraint metrics (90% CI)**	**Relevant phenotype (OMIM ID)**	****DD Research gene**
Potential DD gene						
D3S.0105	C*IR1* (NM_004882.4)	c.865dup (p. Ile289AsnfsTer15)	Regulates transcription and acts as a regulator for RBPJ-mediated repression of transcription (Maita et al., 2005)	pLI = 0.00 missense Z-score = 1.04	NI	No
D3S.0116	*ANO2* (NM_001278596.3)	c.939+1G>A	involved in the calcium chloride channels for olfactory signal transduction (Stephan et al., 2009)	pLI = 0.00 missense Z-score =-0.01	NI	Yes

*Gene constrains metrics (90% confidence interval). These scores are taken from the gnomAD database (https://gnomad.broadinstitute.org/); pLI, a high score (maximum 1) indicates more intolerance of protein-truncating variants (≥ 0.9 mean the gene is extremely intolerant); Missense Z-score, high (more positive) scores suggest that the transcript is more intolerant of variation. **DD Research gene. The gene is part of the DDD-UK research gene list as evidenced on the DECIPHER database. These are a group of genes that have been reported among several DDD-UK study patient but do not have enough evidence to consider gene-disease association. **Abbreviations**: NI, no information relating to human phenotype.

A pathogenic missense variant (NM_015559.3: c.2612T>C (p. Ile871Thr)) overlapping a known DD gene, *SETBP1* was identified in patient, D3S.0014 only through the *de novo* pipeline. The variant had not been prioritised using the DDG2P prioritization pipeline, as it had been reported and annotated in ClinVar as "conflicting interpretation" (VCV000001031.15), which is a classification category not evaluated in this current PhD study. A pathogenic missense variant (NM_015559.3: c.2612T>C (p. Ile871Thr)) overlapping a known DD gene, *SETBP1* was identified in patient, D3S.0014 only through the *de novo* pipeline. The variant had not been prioritised using the DDG2P prioritization pipeline, as it had been reported and annotated in ClinVar as "conflicting interpretation" (VCV000001031.15), which is a classification category not evaluated in this current PhD study. The variant was classified as pathogenic (ACMG codes; PM1, PM2, PP3, PM5). Pathogenic variants in the *SETBP1* gene have been associated with Schinzel-Giedion syndrome (Hoischen et al., 2010; Verloes et al., 1993). Important to note that this variant has not been through final variant review by the DDD-Africa team due to meeting dates only available later after the submission of this PhD work.

- *Genes warranting investigation as potential DD genes*

The remainder of the prioritised variants (10/27) were identified in genes which at the time of analysis had not yet been associated with DD or did not have enough functional evidence (Table 3.6). Analysis of this variants found that, 60% (6/10) had been previously identified in other DD patient and were available as part of the DD research gene list, accessible via the DECIPHER online repository (Table 3.6).

- *Gene constraint metric show an enrichment for loss-of-function de novo variants*

Deleterious protein truncating *de novo* variants have been shown to impact highly constrained biological pathways and can result in an almost complete elimination of reproductive fitness(Kosmicki et al., 2017). Highly constrained genes involved in neurodevelopmental disorders have been reported to have a high intolerance for loss-of-function variants and as such gene metrics for intolerance are used in most gene discovery strategies (Betancur & Buxbaum, 2020; Coe et al., 2019; Kosmicki et al., 2017). We, therefore, screened for potential DD genes using the recommended loss-of-function intolerant score of pLI (>0.9) and missense Z-score of 3.09 (Kosmicki et al., 2017).

We observed that most of the known DD genes (n = 15/17, 88.2%) had a pLI score > 0.9 indicating high intolerance for loss-of-function variants. Additionally, 10/17 (58.8%) of the DD

known genes had Z-score > 3.09 indicating a high constraint for missense variants (Table 3.6). Of note, *ARID1B* showed the highest pLI score with 1.00 and a Z -score of 1.1. *AUTS2* also did not show constraint toward missenses (Z-score of 2.22) but a pLI of 1 suggesting an intolerance to loss-of-function variants.

A further evaluation of the constraint metrics showed that all the identified potentially novel DD genes except for *MYCBP2* (missense Z-score =6.05) did not show a constraint towards missense variants. Additionally, an intolerance for loss-of-function variants was indicated in 38.7% (n=4/11) of the genes.

Further interpretation and manual curation of these variants is required by the DDD-Africa multidisciplinary team of experts but was also outside the scope of this current PhD study. A pathogenic missense variant (NM_015559.3: c.2612T>C (p. Ile871Thr)) overlapping a known DD gene, *SETBP1* was identified in patient, D3S.0014 only through the *de novo* pipeline. The variant had not been prioritized using the DDG2P prioritization pipeline, as it had been reported and annotated in ClinVar as a "conflicting interpretation" (VCV000001031.15), which is a classification category not evaluated in this current PhD study. The variant was classified as pathogenic (ACMG codes; PM1, PM2, PP3, PM5). Pathogenic variants in the *SETBP1* gene have been associated with Schinzel-Giedion syndrome (Hoischen et al., 2010; Verloes et al., 1993). Important to note that this variant has not been through final variant review by the DDD-Africa team due to meeting dates only being available later after the submission of this PhD work.

- *Genes warranting investigation as potential DD genes*

The remainder of the prioritized variants (10/27) were identified in genes that at the time of analysis had not yet been associated with DD or did not have enough functional evidence for association with any human development or function to be included as part of the DDG2P virtual panel (Table 3.6).

Any analysis of these variants found that 60% (6/10) had been previously identified in other DD patients and were available as part of the DD research gene list, accessible via the DECIPHER online repository (Table 3.6).

- *Gene constraint metric show an enrichment for loss-of-function de novo variants*

Deleterious protein truncating *de novo* variants have been shown to impact highly constrained biological pathways and can result in an almost complete elimination of reproductive fitness(Kosmicki et al., 2017). Highly constrained genes involved in neurodevelopmental disorders have been reported to have a high intolerance for loss-of-function variants and as such gene metrics for intolerance are used in most gene discovery strategies (Betancur & Buxbaum, 2020; Coe et al., 2019; Kosmicki et al., 2017). We, therefore, screened for potential DD genes using the recommended loss-of-function intolerant score of pLI (>0.9) and missense Z-score of 3.09 (Kosmicki et al., 2017).

We observed that most of the known DD genes (n = 15/17, 88.2%) had a pLI score > 0.9 indicating high intolerance for loss-of-function variants. Additionally, 10/17 (58.8%) of the DD known genes had Z-score > 3.09 indicating a high constraint for missense variants (Table 3.6). Of note, *ARID1B* showed the highest pLI score with 90.1 and a Z -score of 1.1. *AUTS2* also did not show constraint toward missenses (Z-score of 2.22) but a pLI of 1 suggesting an intolerance to loss-of function variants.

A further evaluation of the constraint metrics showed that all the identified potentially novel DD genes except for *MYCBP2* (missense Z-score =6.05) did not show a constraint towards missense variants. Additionally, an intolerance for loss-of-function variants was indicated in 38.7% (n=4/11) of the genes.

Further interpretation and manual curation of these variants is required by the DDD-Africa multidisciplinary team of experts but was also outside the scope of this current PhD study.

The variant was classified as pathogenic (ACMG codes; PM1, PM2, PP3, PM5). Pathogenic variants in the *SETBP1* gene have been associated with Schinzel-Giedion syndrome (Hoischen et al., 2010; Verloes et al., 1993). Important to note that this variant has not been through final variant review by the DDD-Africa team due to meeting dates only available later after the submission of this PhD work.

- *Genes warranting investigation as potential DD genes*

The remainder of the prioritised variants (10/27) were identified in genes which at the time of analysis had not yet been associated with DD or did not have enough functional evidence for

association with any human development or function to be included as part of the DDG2P virtual panel (Table 3.6).

Any analysis of this variants found that, 60% (6/10) had been previously identified in other DD patient and were available as part of the DD research gene list, accessible via the DECIPHER online repository (Table 3.6).

- *Gene constraint metric show an enrichment for loss-of-function de novo variants*

Deleterious protein truncating *de novo* variants have been shown to impact highly constrained biological pathways and can result in an almost complete elimination of reproductive fitness(Kosmicki et al., 2017). Highly constrained genes involved in neurodevelopmental disorders have been reported to have a high intolerance for loss-of-function variants and as such gene metrics for intolerance are used in most gene discovery strategies (Betancur & Buxbaum, 2020; Coe et al., 2019; Kosmicki et al., 2017). Therefore, screened for potential DD genes using the recommended loss-of-function intolerant score of pLI (>0.9) and missense Z-score of 3.09 (Kosmicki et al., 2017).

We observed that most of the known DD genes (n = 15/17, 88.2%) had a pLI score > 0.9 indicating high intolerance for loss-of-function variants. Additionally, 10/17 (58.8%) of the DD known genes had Z-score > 3.09 indicating a high constraint for missense variants (Table 3.6). Of note, *ARID1B* showed the highest pLI score with 90.1 and a Z -score of 1.1. *AUTS2* also did not show constraint toward missenses (Z-score of 2.22) but a pLI of 1 suggesting an intolerance to loss-of-function variants.

A further evaluation of the constraint metrics showed that all the identified potentially novel DD genes except for *MYCBP2* (missense Z-score =6.05) did not show a constraint towards missense variants. Additionally, an intolerance for loss-of-function variants was indicated in 38.7% (n=4/11) of the genes. Further interpretation and manual curation of these variants is required by the DDD-Africa multidisciplinary team of experts but was also outside the scope of this current PhD study.

3.10 WES diagnostic yield

Overall, for 30 out of the 117 patients (from 115 families) that underwent WES, likely pathogenic or pathogenic variants that explain the patient phenotype were identified in known

DD genes (Table 3.5.) This resulted in an overall WES diagnostic yield of 25.2% (29/115 families). This included at the time of data analysis and interpretation, 16 variants (55.2%) that had previously been reported in the ClinVar variant and disease database (known), and 13 (44.8%) newly reported variants (novel).

Disease-causing variants were identified in a total of 27 genes, all occurring once except for *CHD7* gene that is associated with CHARGE disorder which was observed in three of the patients (Table 3.5). The spectrum of genes containing disease-causing variants is shown in Figure 3.12.

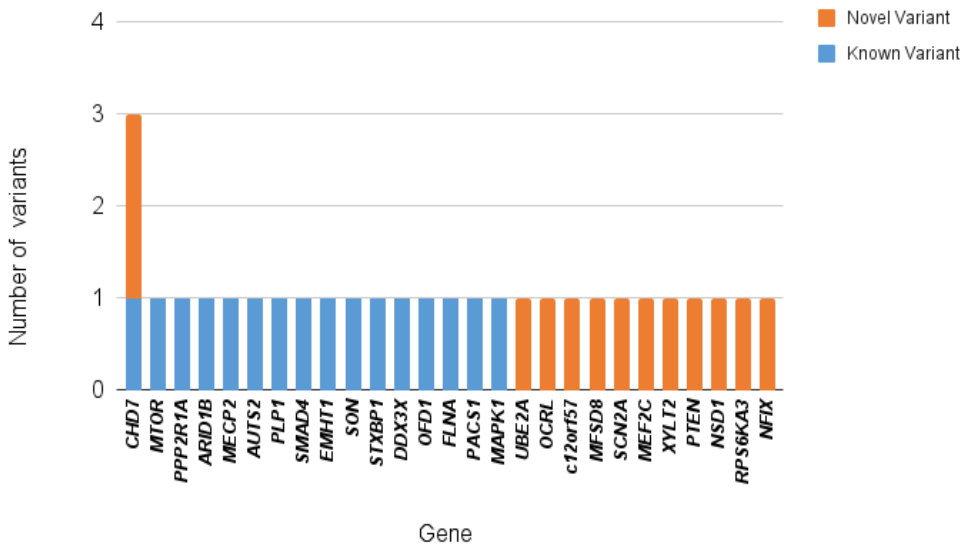

Figure 3.12.The frequency disease-causing genes reported in the DD cohort. Majority of DD disease-causing variants (55.2%, 16/29) had previously been reported in the ClinVar variant and disease database (Blue, known). Each DD gene was reported once except for *CHD7*, which had three pathogenic variants reported.

The most common variant classes were, missense variants (14/29, 48.3%) and followed by variants leading to premature stop codons (7/29, 24.1%). The remainder of the variants were loss-of-function variants, frameshift, splice site acceptor, and splice-site donor variants accounting for six (20.7%), one (3.4%), and one (3.4%) variant respectively. See Figure 3.13.

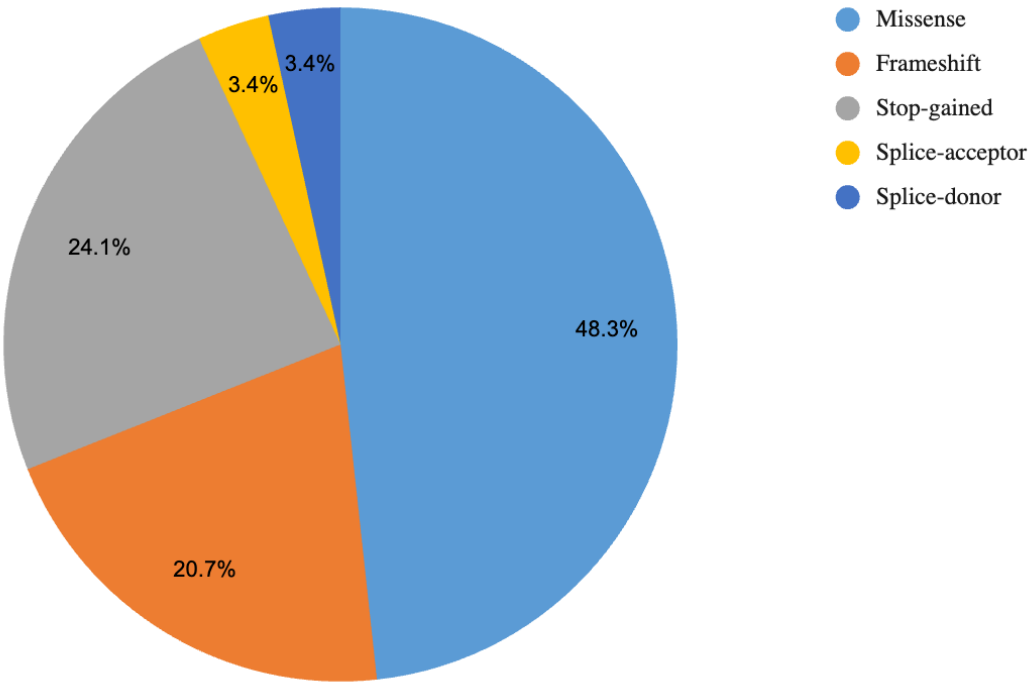

Figure 3.13. Frequency of identified putative disease-causing variant type. A total of 29 putative disease-causing variants were identified in our DD cohort, and 14/29 (48.3%) were missense variants.

3.10.1. De novo inheritance

A molecular diagnosis was reached based on autosomal dominant (AD) DD-associated genes for 17.4% of patients (n = 20/115) (Table 3.5 and Table 3.7). Of these confirmed *de novo* variants accounted for 57.1% (n = 12/21) and an overall diagnostic yield of 10.4% (n= 12/115). In an additional seven patients, the mode of inheritance was unconfirmed due to the unavailability of one parent at the time of recruitment and analysis. However, the reported genes and diseases are associated with *de novo* inheritance in literature, and the variants were not identified in the mother using either WES or Sanger sequencing, and thus *de novo* inheritance was highly likely (Table 3.5 and Table 3.7).

3.10.2. X-linked inheritance

We identified disease-causing variants in seven genes located on the X chromosome. All X-linked (XL) variants were identified in affected male patients and were associated with XL-recessive mode of inheritance. Segregation analysis demonstrated that pathogenic variants were inherited from asymptomatic mothers. However, segregation analysis was not available in patient D3S.0056 who was found to carry a hemizygous missense variant in *PLPI* gene. This was due to the unavailability of the mother's DNA sample which had failed WES quality

control and was not available for segregation analysis (Table 3.5). Overall, variants in XL genes contributed 5.2% (6/115) to the overall diagnostic yield in our DD cohort.

3.10.3. Autosomal recessive inheritance

In our cohort, 10% (3/30) of the patients with a positive molecular diagnosis had a genetic alteration associated with an autosomal recessive (AR) pattern of inheritance (Table 3.5). For one of these patients (D3S.0037), the homozygous variant was detected in AR associated *C12orf57* gene. Each allele was suspected to have been inherited from each respective parent, however, only the mother's sample was available and thus inheritance could not be confirmed. Overall, variants in AR genes contributed 2.6% (3/115) to the overall diagnostic yield in our DD cohort.

Table 3.7. Diagnostic yield of WES based on mode of inheritance in the DDD-Africa study. Confirmed *de novo* disease-causing variants accounted for a large proportion of the solved cases (41.4%, n=12/29), compared to inherited variants (34.5%, 10/29). Of the variants seven were suspected *de novo* variants but were identified in duo families thus segregation analysis could not be performed.

Mode of inheritance	Number of variants	Diagnostic yield (n=115 families)
Autosomal Dominant		16.5%
De novo	11	
Inherited	1	
*Suspected *de novo*	7	
Autosomal Recessive		2.6%
**Inherited	3	
X-linked		6.1%
De novo	1	
inherited	6	

*Suspected *de novo*, Heterozygous in affected proband, however DNA of unaffected parent(s) unavailable to confirm.
**Inherited, includes an extended family (D3S.0033, mother, father and two male siblings).

3.11 The association of recorded HPO terms and a positive diagnosis

We observed a relationship between the patient's recorded phenotype (evaluated by the number of HPO terms recorded) and a positive molecular diagnosis (Figure 3.14). Patient with less than 10 HPO terms recorded, had no molecular diagnosis reached. However, for those with 11–20 HPO terms recorded a positive diagnosis was reached in 22% and this increased to as much as 60% in cases with over 30 HPO terms.

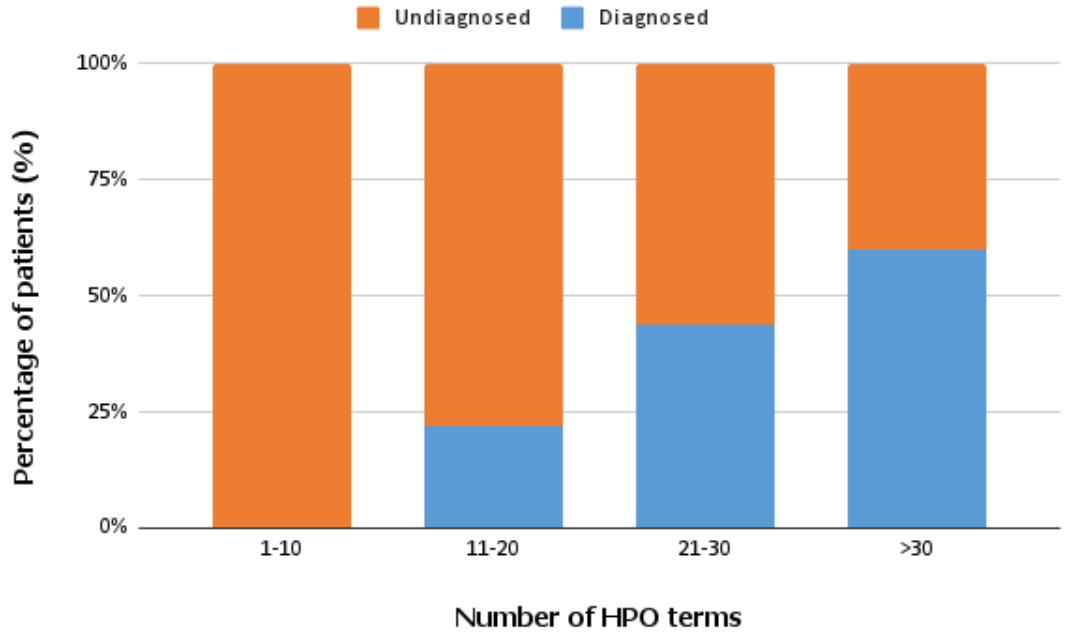

Figure 3.14. The number of HPO terms reported for diagnosed and undiagnosed patient. The data suggests that the more complex the phenotype as indicated by the higher number of HPO terms (>10) the greater the likelihood of a molecular diagnosis.

Among the 33 patient that had a suspected clinical diagnosis prior to WES testing molecular confirmation was achieved in four (n= 3/33 9.09%) (CHARGE syndrome in patient D3S.0050, Coffin-Lowry syndrome in patient D3S.0119, and Oral-Facial-Digital Syndrome 1 in patient D3S.0098). In another four patients (n=4/33, 12.1%) a different molecular diagnosis was reached compared to that initially suspected based on clinical presentation. The remainder of the reported putative disease-causing variants were identified in patient whose phenotypes failed to suggest a link to the disease prior to exome sequencing (Table 3.8). These findings further highlight the high level of phenotypic and genotypic heterogeneity associated with DD.

Table 3.8. Suspected clinical Diagnosis of patients in DDD-Africa cohort. A possible clinical diagnosis was suggested during a clinical case review by the DDD-Africa team of experts based on the patient's clinical phenotype. Suspected diagnosis was available for 33 of the117 DD patients (28.2%). Molecular testing (WES and Sanger sequencing) confirmed clinical diagnosis in 3/33 (9.09%) patients, while for 4/33 (12.1%) a different diagnosis was achieved. For the rest (26/33) no diagnosis was reached and cases remain open.

Patient ID	Suspected Diagnosis (MIM ID)	Confirmed Diagnosis (OMIM ID)	Status
D3S.0014	Coffin-Siris syndrome (135900)	N/A	Open
D3S.0015	Loeys-Dietz syndrome (609192)	Malan syndrome (614753)	Closed
D3S.0016	Coffin-Siris syndrome (135900)	N/A	Open
D3S.0019	Kleefstra syndrome (610253)	N/A	Open
D3S.0021	Fetal alcohol syndrome	N/A	Open
D3S.0033*	Intellectual disability-hypotonic facies syndrome, X-linked (309580)	Malan syndrome (614753)	Closed
D3S.0039	DOORS syndrome (220500)	N/A	Open
D3S.0043	3q29 microdeletion syndrome (609425)	N/A	Open
D3S.0047	Cornelia de Lange syndrome 2 (300590)	Coffin-Siris syndrome (135900)	Closed
D3S.0048	Langer Gideon Syndrome (150230)	N/A	Open
D3S.0050	CHARGE syndrome (214800)	CHARGE syndrome (214800)	Closed
D3S.0057	Smith Magenis syndrome (182290)	N/A	Open
D3S.0062	Noonan syndrome 1 (163950)	N/A	Open
D3S.0076	Opitz GBBB Syndrome (300000)	N/A	Open
D3S.0078	Kleefstra syndrome 1 (610253)	N/A	Open
D3S.0084	Acrofacial dysostosis (154400)	N/A	Open
D3S.0092	WAGR (194072)	N/A	Open
D3S.0096	Tatton-Brown-Rahman (615879)	N/A	Open
D3S.0098	OFD1syndrome (311200)	OFD1 syndrome (311200)	Closed
D3S.0099	PTEN hamartoma tumor syndrome (PHTS)*	N/A	Open
D3S.0105	KBG syndrome (148050)	N/A	Open
D3S.0109	Coffin-Siris (135900)	N/A	Open
D3S.0110	Townes-Brocks syndrome 1(107480)	N/A	Open
D3S.0111	Sotos syndrome (117550)	Sotos syndrome (117550)	Closed
D3S.0113	MFDGA (610536)	N/A	Open
D3S.0115	Suspected Rasopathy like	N/A	Open
D3S.0116	Catel-manzke syndrome (616145)	N/A	Open
D3S.0119	Coffin-Lowry syndrome (303600)	Coffin-Lowry syndrome (303600)	Closed
D3S.0121	Fanconi renotubular syndrome 1 (134600)	N/A	Open
D3S.0122	Hanhart syndrome (103300)	N/A	Open
D3S.0125	Coffin-Lowry (303600) or Coffin-Siris (135900) syndrome	N/A	Open
D3S.0126	Smith Magenis syndrome (182290)	N/A	Open
D3S.0127	Coffin-Lowry (303600)	N/A	Open

D3S.0033*; Extended family consisting of two affected male boys, mother, and father. **DOORS** syndrome, Deafness, Onychodystrophy, Osteodystrophy, Impaired Intellectual Development, And Seizures Syndrome; **OFD1 syndrome**, Oro-facio-digital 1 syndrome; MFDGA, Mandibulofacial Dysostosis, Guion-Almeida Type; **PTEN hamartoma tumor syndrome (PHTS)*,** includes a range of related syndromes mainly Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome, PTEN-related Proteus syndrome, and Proteus-like syndrome. **WAGR**, Wilms tumor, Aniridia, GenitoUrinary anomalies, and intellectual disability.

3.12 Genotype-Phenotype evidence for disease-causing variants

The overall reported diagnostic yield was due to 29 pathogenic variants (likely pathogenic/pathogenic) identified in 28 unrelated patients and two male siblings. Of these variants, 16 (55.2%) had previously been reported in the ClinVar database "known" and 13 (41.4%) were novel as at the time of analysis and interpretation had not been reported in ClinVar (Table 3.5).

3.12.1. D3S.0013 - *MTOR* associated *Smith-Kingsmore syndrome in an African ancestry patient*

- *Clinical manifestation*

A male patient (D3S.0013) was recruited at the age of four based on DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound) and C (moderate DD and macrocephaly). He was delivered to phenotypically normal parents by caesarean at 40 weeks of gestation. Clinical examination of the patient revealed macrocephaly (>3SD), hypotonia, and moderate global DD with only a few words spoken. The patient attends renal, ear, nose and throat physiotherapy, neurodevelopmental and speech therapy clinics. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.9.

Table 3.9. Clinical phenotype for patient D3S.0013 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0000494	Downslanted palpebral fissures	Abnormality of head or neck
HP:0000565	Esotropia	Abnormality of the eye
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0004906	Hypernatremic dehydration	Abnormality of metabolism/homeostasis
HP:0000256	Macrocephaly	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0000817	Poor eye contact	Abnormality of the nervous system
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0000486	Strabismus	Abnormality of the eye
HP:0009884	Tapered distal phalanges of finger	Abnormality of the musculoskeletal system
HP:0002781	Upper airway obstruction	Abnormality of the respiratory system

- *Molecular evaluation*

Exome trio analysis identified a *de novo* missense variant, c.5395G>A, in *MTOR* gene resulting in a glutamic acid replacement with lysine at amino acid 1799 (p. Glu1799Lys). The gene encodes a 300kDa highly conserved serine/threonine kinase that plays a critical role in the mTOR signalling pathway important for cell proliferation, cell survival, apoptosis, metabolism, and cancer (Hay & Sonenberg, 2004; Moore et al., 1996).

Segregation analysis performed following the sequencing of both parent DNA demonstrated a reference allele sequence for both, thus supporting a *de novo* change (Table 3.5, Figure 3.15). In accordance with the standard ACMG-AMP guidelines, the variant was classified as pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5.

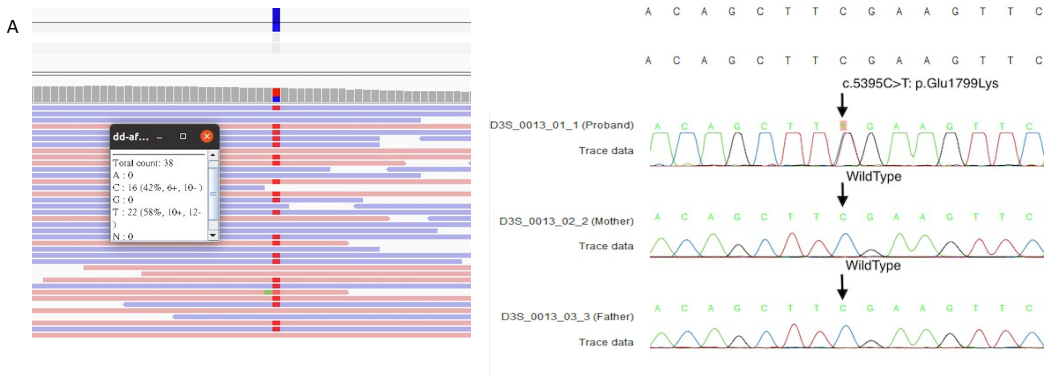

Figure 3.15. Validation of *MTOR* variant in patient D3S.0013. (A) *De novo* heterozygous *MTOR* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. The region was covered by a total of 38 reads, with 22 reads and 16 reads supporting the alternative allele and reference allele (VAF=58%). (B) Sanger sequencing analysis of the *MTOR* variant visualized by CLC bio in the proband, mother, and father confirms the *MTOR* variant is *de novo* in the proband.

- *Molecular diagnosis*

Pathogenic variants in the *MTOR* gene result in MTOR regulation dysfunction and have been associated with mTOR hyperactivity. Hyperactivity of the mTOR pathway is associated with Smith-Kingsmore syndrome (SKS) (#616638), which is mainly characterized by global DD, macrocephaly/ hemimegaloencephaly, and seizures (Gordo et al., 2018). This finding was in accordance with the patient clinical presentation as displayed in the list of HPO terms provided in Table 3.9. The pathogenic *de novo* variant (NM_004958.4: c.5395G>A (p. Glu1799Lys)) has been previously reported (RCV000255268.18) and patient clinical phenotype reported in literature (Baynam et al., 2015; Elizondo-Plazas et al., 2020; Gordo et al., 2018; Mirzaa et al.,

2016; Moosa, Böhrer-Rabel, et al., 2017; Mroske et al., 2012). To the best of our knowledge we report, the first African patient with the *de novo* variant and a diagnosis of SKS).

Unlike previous reports of SKS caused by the same variant, our patient presented with upper airway obstruction. However, asthma has been reported in other patients diagnosed with the syndrome (Besterman et al., 2021). The onset of asthma has been associated with hyperactivation of the mTOR pathway highlighting its importance in normal immune function and dysfunction, and an increased risk of patient comorbidities (Y. Zhang et al., 2017).

The mTOR inhibitor drug rapamycin has proven effective in decreasing the frequency and duration of seizures in treated patients (Hua et al., 2019; Wong, 2010). It was not noted whether our patient was at the time of recruitment on any treatment for seizures, but we recommend the use of this rapamycin to treat seizures in the patient.

In most patients diagnosed with SKS, the confirmed mode of inheritance is maternal germline mosaicism (genetic change occurring following zygote formation and therefore only present in a subset of cells) (Gordo et al., 2018; Mirzaa et al., 2016; Moosa et al., 2017; Mroske et al., 2012). Our patient has an older brother who has a similar clinical presentation (personal communication with the referring clinician) but at the time of initial recruitment was not included in the study. We suspect the same pathogenic *MTOR* variant to be causal in the brother and thus recommend targeted testing be performed using Sanger sequencing to confirm its presence.

Additionally, given the mother is asymptomatic, we suspect maternal germline mosaicism inheritance. In 8/10 of the patient in which the variant was detected, maternal germline mosaicism was reported (Gordo et al., 2018). Thus, we recommend MLPA testing be performed for the detection of any mosaicism.

3.12.2. D3S.0028 - *PPP2R1A* related neurodevelopmental disorder in an African patient

- *Clinical manifestation*

A male patient was recruited at the age of four years based on the DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound) and C (moderate DD and macrocephaly). He was delivered to phenotypically normal parents. The pregnancy and birth

were unremarkable. Clinical examination of the patient revealed microcephaly with a head circumference of 36cm (50th centile), moderate DD, seizures, micrognathia, hypotonia. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.10.

Table 3.10. Clinical phenotype for patient D3S.0028 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000370	Obstruction of Eustachian tube	Abnormality of the ear
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0007321	Deep white matter hypodensities	Abnormality of the nervous system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0002263	Exaggerated cupid's bow	Abnormality of head or neck
HP:0002373	Febrile seizure (within the age range of 3 months to 6 years)	Abnormality of the nervous system
HP:0002079	Hypoplasia of the corpus callosum	Abnormality of the nervous system
HP:0001252	Hypotonia	Abnormality of the musculoskeletal system
HP:0001388	Joint laxity	Abnormality of the musculoskeletal system
HP:0000343	Long philtrum	Abnormality of head or neck
HP:0000368	Low-set, posteriorly rotated ears	Abnormality of the ear
HP:0000347	Micrognathia	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0000341	Narrow forehead	Abnormality of head or neck
HP:0003764	Nevus	Abnormality of the integument
HP:0005487	Prominent metopic ridge	Abnormality of the musculoskeletal system
HP:0010823	Ridged cranial sutures	Abnormality of the musculoskeletal system
HP:0005484	Secondary microcephaly	Abnormality of the musculoskeletal system
HP:0001250	Seizure	Abnormality of the nervous system
HP:0009237	Short 5th finger	Abnormality of the musculoskeletal system
HP:0012745	Short palpebral fissure	Abnormality of head or neck
HP:0004322	Short stature	Growth abnormality
HP:0000506	Telecanthus	Abnormality of head or neck
HP:0001537	Umbilical hernia	Abnormality of the musculoskeletal system
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck
HP:0000431	Wide nasal bridge	Abnormality of head or neck

- *Molecular evaluation*

Trio exome sequencing analysis identified a *de novo* missense variant (c.773G>A), in *PPP2R1A* resulting in the replacement of an arginine with a histidine at amino acid 258 (p. Arg258His) (Table 3.5, Figure 3.16). The *PPP2R1A* variant was validated via Sanger sequencing of proband and both parental samples and observed only in the proband sample. The variant was classified as pathogenic in accordance with the ACMG-AMP guidelines. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5.

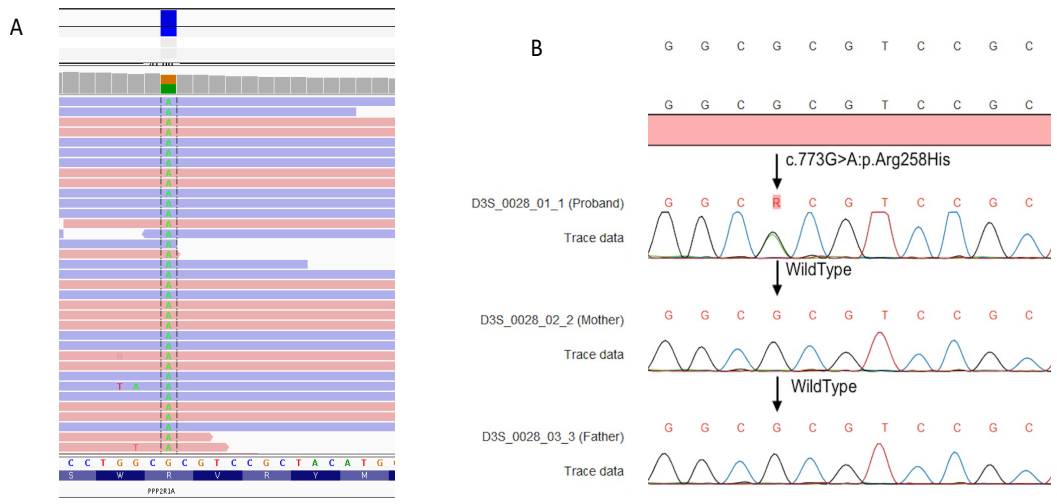

Figure 3.16. Validation of *PPP2R1A* variant in patient D3S.0028. (A) *De novo* heterozygous *PPP2R1A* variant visualized by IGV (reference base G in orange, variant base A in green) in the proband. The region was covered by a total of 93 reads, with 47 supporting the alternative allele and reference allele respectively (VAF 50.5%). (B) Sanger sequencing analysis of the *PPP2R1A* variant visualized by CLC bio in the proband, mother, and father confirms the *PPP2R1A* variant is *de novo* in the proband.

- *Molecular diagnosis*

The pathogenic variant has been previously reported (VCV000217458.12) and shown to result in decreased PP2A activity (Houge et al., 2015) which is important for the negative control of cell growth and cell death (Janssens and Goris 2001). Variants in the gene are associated with a diagnosis of PPP2R1A-related ID (#616362) (Houge et al., 2015; Lenaerts et al., 2021). This disorder is characterized by mild to severe ID, microcephaly, or macrocephaly, hypotonia, severe motor delay, and a range of brain malformations (Douzgou et al., 2022). In some cases, hearing loss, ear-shape abnormality, seizures, short stature, brain abnormality (hypoplastic corpus callosum), and significant microcephaly. This finding was in accordance with the patient's clinical presentation, of severe ID, hypotonia, microcephaly, and epilepsy (See table 3.13) which are characteristic of PPP2R1A related ID (Douzgou et al., 2022; Lenaerts et al., 2021).

The variant showed *de novo* mode of inheritance (Figure 3.20), and the risk of recurrence in siblings associated with this inheritance is estimated to be ~1% (Rahbari et al., 2016). Thus, diagnosis would not directly impact genetic counseling offered in relation to any family planning and any recurrence risk assessments.

3.12.3. D3S.0037 - *C12orf57* first report of Temtamy syndrome in an African ancestry patient

- *Clinical manifestation*

A female patient (D3S.0037) was recruited at the age of nine years based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound)). The patient was delivered to phenotypically normal parents by caesarean at 36 weeks of gestation following spontaneous pre-term labor but was otherwise unremarkable. The patient had an older female sibling with severe DD who deceased at eight years. A CT scan performed at the age of one signalled a chronic communicating hydrocephalus. Seizures were noted from age 3 months to 1 year. Clinical assessment of the patient additionally revealed global DD, short stature, microphthalmia and macrocephaly. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.11.

Table 3.11. Clinical phenotype for patient D3S.0037 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0001334	Communicating hydrocephalus	Abnormality of the nervous system
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0000565	Esotropia	Abnormality of the eye
HP:0100876	Infra-orbital crease	Abnormality of head or neck
HP:0000612	Iris coloboma	Abnormality of the eye
HP:0000272	Malar flattening	Abnormality of the musculoskeletal system
HP:0000691	Microdontia	Abnormality of head or neck
HP:0000568	Microphthalmia	Abnormality of the eye
HP:0002370	Poor coordination	Abnormality of the nervous system
HP:0000358	Posteriorly rotated ears	Abnormality of the ear
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0004482	Relative macrocephaly	Abnormality of the musculoskeletal system
HP:0000480	Retinal coloboma	Abnormality of the eye
HP:0001250	Seizure	Abnormality of the nervous system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0002000	Short columella	Abnormality of head or neck
HP:0004322	Short stature	Growth abnormality
HP:0008070	Sparse hair	Abnormality of the integument
HP:0012471	Thick vermilion border	Abnormality of head or neck
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck
HP:0012810	Wide nasal base	Abnormality of head or neck

- *Molecular evaluation*

A missense variant in *C12orf57*, c.184C>T leading to a p. Gln62Ter change and resulting in a stop codon was identified. Sanger sequencing confirmed a homozygous variant observed in the patient with the mother showing a heterozygous sequence (Figure 3.17). The father was not available at the time of study recruitment. This variant was classified using standard ACMG-AMP guidelines as pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5

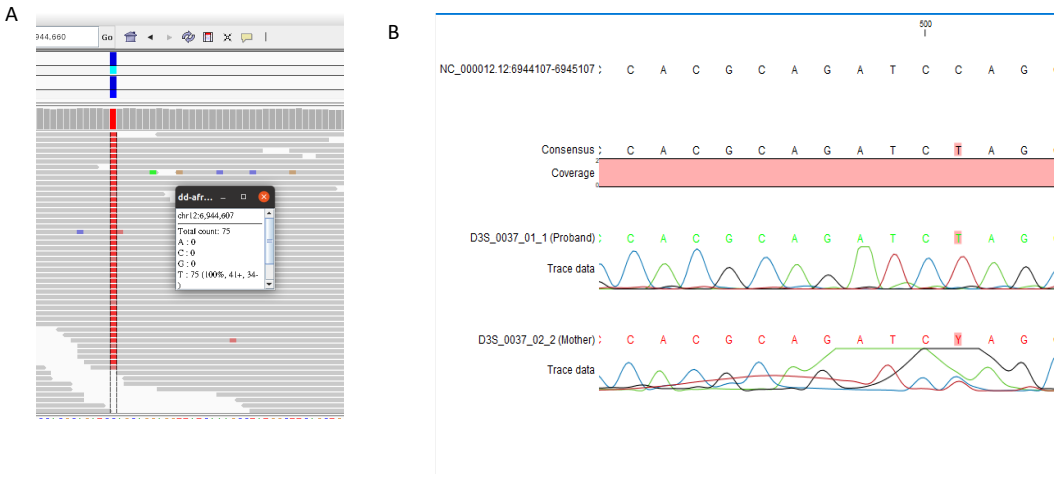

Figure 3.17. Validation of *C12orf57* variant in patient D3S.0037. (A) Homozygous *C12orf57* variant visualized by IGV (variant base T in red) in the proband. The region is covered by a total of 75 reads, with all supporting the alternative allele (VAF 100%). B) Sanger sequencing analysis of the *C12orf57* variant visualized by CLC bio in the proband, and mother confirms the variant is maternally inherited. Sequence pull-up is seen in the mother, but the analysis software was still able to detect presence of a heterozygous variant in the maternal sequence.

- *Molecular diagnosis*

The gene encodes a 126-amino acid cytoplasmic protein with uncertain function (Akizu et al., 2013) but has been reported to be important for human corpus callosum development for brain and eye function (Zahrani et al., 2013). Pathogenic mutations in the gene have been associated autosomal recessive Temtamy syndrome (#218340) (Temtamy et al., 1996) characterized by profound DD, intractable seizure and a range of eye abnormalities including microphthalmia (Akizu et al., 2013; Temtamy et al., 1996; Wang et al., 2020) which were all present in our patient as seen in Table 3.11. Most patients, develop seizures before the age of three years old and remain difficult to control (Wang et al., 2020). Our patient was reported to have had seizures from 3 months up to one year and thereafter they stopped.

To the best of our knowledge this is the first report of the syndrome in a patient of African ancestry, as literatures reports the syndrome in patients from East Asia (Wang et al., 2020) and

the Middle east (Akizu et al., 2013; Alrakaf et al., 2018; Platzer et al.,2014; Salih et al., 2013; Zahrani et al., 2013). We recommend that *C12orf57* testing and Temtamy syndrome be suspected in patients clinically presenting with DD, microphthalmia and a range of eye abnormalities.

3.12.4. D3S.0051 - *CHD7* variant in CHARGE syndrome

- *Clinical manifestation*

A male patient was recruited at the age of two years old based on DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound) and C (One major malformation and 3 or more well-documented minor anomalies). He was delivered to phenotypically normal parents. The pregnancy and birth were unremarkable. Clinical examination of the patient revealed moderate ID, aortic valve stenosis, dysmorphic features, genital hypoplasia, undescended testes and hypotonia. The full list of clinical features, the affected system anomalies and associated HPO codes is summarized in Table 3.12.

Table 3.10. Clinical phenotype for patient D3S.0051 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0001650	Aortic valve stenosis	Abnormality of the cardiovascular system
HP:0011229	Broad eyebrow	Abnormality of the integument
HP:0000028	Cryptorchidism	Abnormality of the genitourinary system
HP:0010730	Double eyebrow	Abnormality of the integument
HP:0000286	Epicanthus	Abnormality of head or neck
HP:0003241	External genital hypoplasia	Abnormality of the genitourinary system
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0000369	Low-set ears	Abnormality of the ear
HP:0001643	Patent ductus arteriosus	Abnormality of the cardiovascular system
HP:0005487	Prominent metopic ridge	Abnormality of the musculoskeletal system
HP:0002092	Pulmonary arterial hypertension	Abnormality of the cardiovascular system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0003196	Short nose	Abnormality of head or neck

- *Molecular evaluation*

Exome sequencing trio analysis identified a *de novo* missense variant (c.5833C>T) in the *CHD7* (605983) gene, resulting in a stop codon (p. Arg1945Ter) (Table 3.5, Figure 3.18). The variant was classified as pathogenic in accordance with the ACMG-AMP guidelines as displayed in table 3. 5.. The variant was validated via Sanger sequencing of proband and both parental samples and observed only in the proband sample (Figure 3.18).

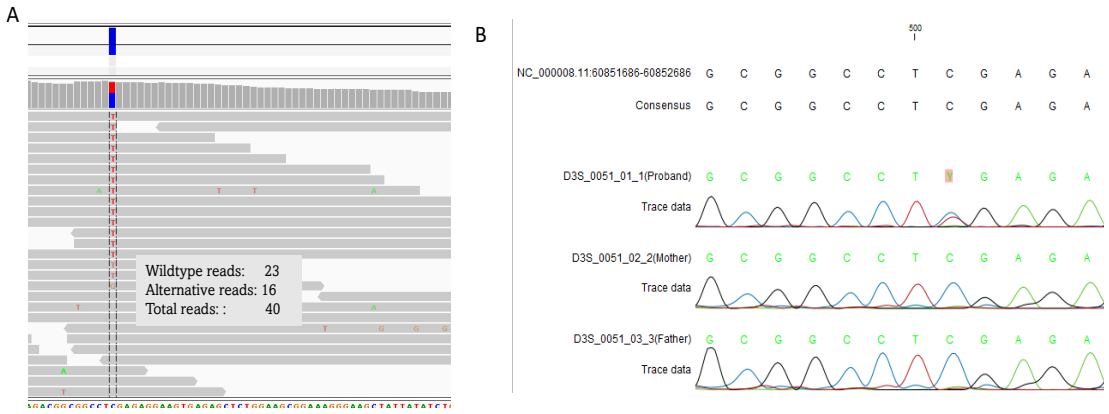

Figure 3.18. Validation of *CHD7* variant in patient D3S.0051. (A) *De novo* heterozygous *CHD7* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 40 reads covers the region and 16 support the alternative allele (VAF 40.0%). (B) Sanger sequencing analysis of the *CHD7* variant visualized by CLC bio in the proband, mother, and father confirms the variant is *de novo* in the proband.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with diagnosis of CHARGE syndrome (# 616362). Our patient presented with severe developmental delay, bilateral colobomas at optic nerve with visual loss, abnormal movements, hyperactivity, and microcephaly, profound bilateral sensorineural hearing loss fitting a CHARGE syndrome diagnosis. Scoliosis has been reported to develop in patients with CHARGE syndrome (Borysiak et al., 2020; Doyle and Blake, 2005). Thus, we recommend that the patient be monitored spine development abnormalities which may be an indicator of Scoliosis. Late diagnosis has been shown to limit treatment options for the patients with surgical interventions such as spinal fusion (REAMY and SLAKEY, 2001).

3.12.5. D3S.0052 - *MECP2* variant results in atypical Rett syndrome

- *Clinical manifestation*

A six year old male patient (D3S.0052) was recruited based on DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound) and C (moderate DD and macrocephaly). The patient was delivered to phenotypically normal parents by caesarean at 40 weeks' gestation due to pre-eclampsia. The mother suffered from chronic hypertension during the pregnancy. Clinical history revealed a three-day neonatal ICU admission and treatment using nasal prong oxygen. Additionally, two separate hospital admissions for respiratory tract infections were reported. Seizures were noted from two years and the patient was placed on Epilim treatment which was discontinued at 3 years. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.13.

Table 3.11. Clinical phenotype for patient D3S.0052 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0007018	Attention deficit hyperactivity disorder	Abnormality of the musculoskeletal system
HP:0001348	Brisk reflexes	Abnormality of the musculoskeletal system
HP:0010862	Delayed fine motor development	Abnormality of the musculoskeletal system
HP:0002194	Delayed gross motor development	Abnormality of the musculoskeletal system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0001540	Diastasis recti	Abnormality of the musculoskeletal system
HP:0025116	Fetal distress	Abnormality of prenatal development or birth
HP:0002007	Frontal bossing	Abnormality of the musculoskeletal system
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0001007	Hirsutism	Abnormality of the integument
HP:0004305	Involuntary movements	Abnormality of the musculoskeletal system
HP:0000400	Macrotia	Abnormality of the ear
HP:0007565	Multiple cafe-au-lait spots	Abnormality of the integument
HP:0000817	Poor eye contact	Abnormality of the musculoskeletal system
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0000520	Proptosis	Abnormality of the eye
HP:0004482	Relative macrocephaly	Abnormality of the musculoskeletal system
HP:0011344	Severe global developmental delay	Abnormality of the musculoskeletal system
HP:0000322	Short philtrum	Abnormality of head or neck
HP:0002781	Upper airway obstruction	Abnormality of the respiratory system

- *Molecular evaluation*

Trio exome analysis of revealed a hemizygous missense variant c.961C>T overlapping *MECP2* gene (600662) causing an amino acid change of arginine to tryptophan (p. Arg321Trp) resulting in a stop-gain was detected in the patient. The gene encodes a MECP2 protein which acts as a methylation-dependent transcription repressor and activator by interaction with different cofactors (Chahrour et al., 2008; Thambirajah et al., 2012; Yasui et al., 2007*)*. A missense variant in *MECP2*, (c.961G>A. Sanger sequencing confirmed the presence of the variant in a hemizygous state. Both parents DNA showed reference allele sequence (Figure 3.19). This variant was classified as ACMG-AMP pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5.

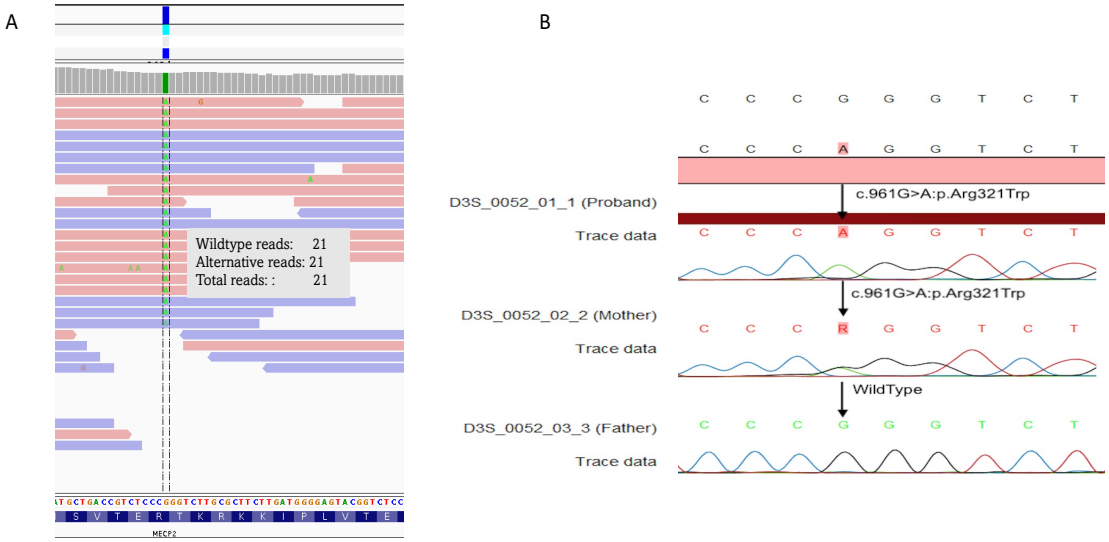

Figure 3.19. Validation of *MECP2* variant in patient D3S.0052. (A) A hemizygous *MECP2* variant visualized by IGV (variant base A in green) in the proband. A total of 21 reads covers the region and all support the alternative allele (VAF 100%). (B) Sanger sequencing analysis of the *MECP2* variant visualized by CLC bio in the proband, mother, and father confirms the variant is maternally inherited.

- *Molecular diagnosis*

Pathogenic variants in the gene affect its stability (Kucukkal et al., 2015) and affinity (Yang et al., 2016) for methyl marks in different cofactors (Kriaucionis and Bird, 2003). This is associated with neurodevelopmental disorders in Rett-like syndrome (# 312750)(Chahrour et al., 2008).

The mutation was confirmed to be maternally inherited. In all cases with a confirmed molecular diagnosis, skewed X-inactivation was observed in the mother (Ghorbel et al., 2018; Jdila et al.,

2020; Kharrat et al., 2017; Schönewolf-Greulich et al., 2016). We therefore recommend that X-chromosome inactivation assay testing be performed to confirm this in the mother.

3.12.6. D3S.0053 - *AUTS2 variant in a patient with severe feeding problems*

- *Clinical manifestation*

A two-year-old male patient (D3S.0053) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound). The patient was delivered to phenotypically normal parents at 40 weeks of gestation. Following birth, the patient was admitted to the neonatal clinic for a 21-day period due to poor suck reflex and feeding. Clinical history revealed a hospital admission due to poor feeding. Diagnostic testing performed prior to recruitment included Fragile X and karyotype all of which were normal. The full list of dysmorphic features and associated HPO codes and HPO system anomalies are represented in Table 3.14.

Table 3.12. Clinical phenotype for patient D3S.0053 described using Human Phenotype Ontology terms

HPO ID	HPO term	HPO category
HP:0008936	Axial hypotonia	Abnormality of the musculoskeletal system
HP:0001348	Brisk reflexes	Abnormality of the nervous system
HP:0002059	Cerebral atrophy	Abnormality of the nervous system
HP:0010862	Delayed fine motor development	Abnormality of the nervous system
HP:0002194	Delayed gross motor development	Abnormality of the nervous system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0006956	Dilation of lateral ventricles	Abnormality of the nervous system
HP:0002353	EEG abnormality	Abnormality of the nervous system
HP:0001298	Encephalopathy	Abnormality of the nervous system
HP:0000286	Epicanthus	Abnormality of head or neck
HP:0020045	Esodeviation	Abnormality of the eye
HP:0011471	Gastrostomy tube feeding in infancy	Abnormality of the digestive system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0001276	Hypertonia	Abnormality of the musculoskeletal system
HP:0040261	Increased size of nasopharyngeal adenoids	Abnormality of the nervous system
HP:0007105	Infantile encephalopathy	Abnormality of the nervous system
HP:0009183	Joint contracture of the 5th finger	Abnormality of the musculoskeletal system
HP:0000347	Micrognathia	Abnormality of the musculoskeletal system
HP:0000341	Narrow forehead	Abnormality of head or neck
HP:0002788	Recurrent upper respiratory tract infections	Abnormality of head or neck

HP:0000322	Short philtrum	Abnormality of head or neck
HP:0005338	Sparse lateral eyebrow	Abnormality of the integument
HP:0000098	Tall stature	Growth abnormality
HP:0012471	Thick vermilion border	Abnormality of head or neck
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck

- *Molecular evaluation*

Trio exome analysis of the patient revealed a *de novo* missense pathogenic variant (c.376C>T) in *AUTS2* causing a p. Arg126Ter resulting in a pre-mature stop-gain. The gene encodes a 1259bp protein which functions as a transcription regulator during brain development. It is also involved in cell adhesion, axon guidance, and in synaptic function (Beunders et al., 2013; Pang et al., 2021).

Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.20). Segregation analysis showed a showed *de novo* mode of inheritance. This variant was classified as ACMG-AMP pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5.

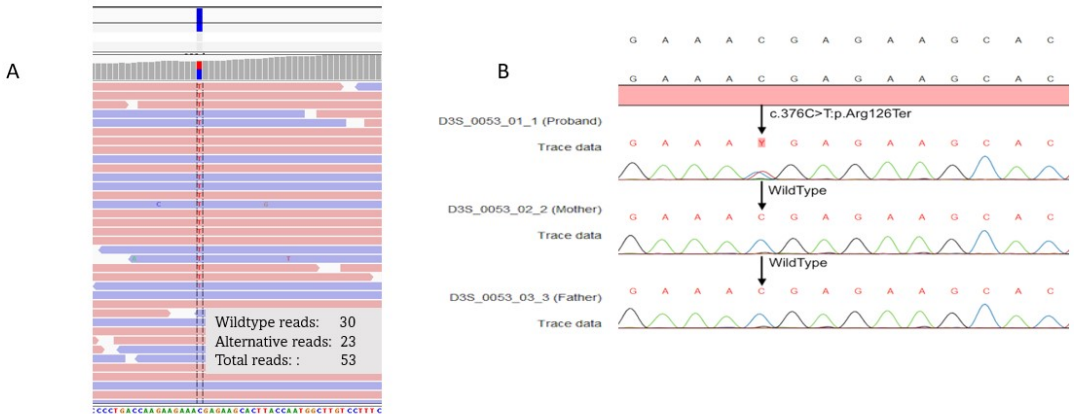

Figure 3.20. Validation of *AUTS2* variant in patient D3S0053. (A) A heterozygous *AUTS2* variant was visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 53 reads covers the region with 23 supporting the alternative allele (43.4%). B) Sanger sequencing analysis of the *AUTS2* variant visualized by CLC bio in the proband, mother, and father confirms the variant is *de novo* inheritance.

- *Molecular diagnosis*

Mutations in the gene have been associated with a diverse group of clinical features considered AUTS2 syndrome (#615834) (Beunders et al., 2016). The most common clinical characteristics

of the disorder, include ID, microcephaly, and feeding problems which were all present in our patient as shown in Table 3.14.

Severe feeding problems were reported in our patient with the use of an oral gastric feeding tube, and ongoing swallowing issues were detailed (Table 3.14). Currently, there are no guidelines for clinical surveillance or management of the disorder but in our patient, management needs to be focused on managing the noted severe feeding issues.

3.12.7. D3S.0056 - *PLP1* variant *associated with Pelizaeus-Merzbacher disease*

- *Clinical manifestation*

A two year old male patient (D3S.0056) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound). The patient was delivered to phenotypically normal parents by caesarean at 40 weeks of gestation due to poor labor progression. Additionally, Ante-Partum Haemorrhage was noted during the pregnancy. One previous miscarriage was also noted. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO terms and codes is summarized in Table 3.15.

Table 3.13. Clinical phenotype for patient D3S.0056 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0008936	Axial hypotonia	Abnormality of the musculoskeletal system
HP:0100814	Blue nevus	Abnormality of the integument
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0000028	Cryptorchidism	Abnormality of the genitourinary system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0007162	Diffuse demyelination of the cerebral white matter	Abnormality of the nervous system
HP:0001187	Hyperextensibility of the finger joints	Abnormality of the musculoskeletal system
HP:0002509	Limb hypertonia	Abnormality of the musculoskeletal system
HP:0000369	Low-set ears	Abnormality of the ear
HP:0000639	Nystagmus	Abnormality of the eye
HP:0000396	Overfolded helix	Abnormality of the ear
HP:0005490	Postnatal macrocephaly	Abnormality of the musculoskeletal system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0000977	Soft skin	Abnormality of the integument
HP:0000486	Strabismus	Abnormality of the eye
HP:0001762	Talipes equinovarus	Abnormality of limbs
HP:0000431	Wide nasal bridge	Abnormality of head or neck

- *Molecular evaluation*

A hemizygous missense variant in *PLP1* gene (c.44C>T) causing a p. Pro15Leu change was detected in the patient. The gene encodes a myelin membrane protein which is important for central nervous system (CNS) development and function (Inoue, 2017).

WES data was only available for the patient as both parents DNA failed WES quality control at the WTSI. Sanger sequencing confirmed the presence of the homozygous variant (Figure 3.21). The mothers DNA sample was also not available for segregation analysis, but Sanger sequence of the father's DNA showed a reference base at the position. This variant was classified as ACMG-AMP pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5.

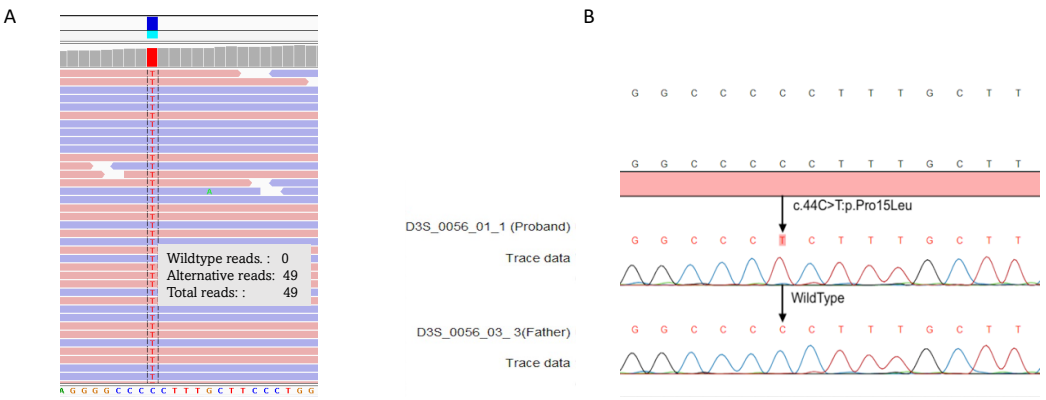

Figure 3.21. Validation of *PLP1* variant in patient D3S.0056. (A) A homozygous *PLP1* variant visualized by IGV (variant base T in red) in the proband. A total of 49 reads covers the region and all support the alternative allele (VAF 100%). (B) Sanger sequencing analysis of the *PLP1* variant visualized by CLC bio in the proband, and father shows variant is not paternally inherited.

- *Molecular diagnosis*

Pathogenic variants in the gene have been suggested to result in the accumulation of myelin membrane protein, disrupting cellular haemostasis and leading to cellular apoptosis (Inoue, 2017). These are associated with Leukodystrophy Hypo myelinating Type 1 and Spastic Paraplegia X-Linked Type 2 (Wolf et al., 2005). Our patient had severe myelination evidenced by diffuse demyelination on brain scans. Additionally, he presented with nystagmus and hypotonia (see table 3.15) which are all characteristic of PMD (Hobson and Garbern, 2012). Of note, the patient was still not able to sit up and had severe stiffness in the legs and feet (Wolf et al., 2005). This further confirmed a clinical diagnosis of PMD, in which patients have severe muscle stiffness and ataxia.

There are several therapeutics that have been suggested and shown promise in vitro. These include a dietary modification, by adding curcumin, which is a compound available in curry spice turmeric (L. H. Yu et al., 2012) as well as the use of Chloroquine an anti-malaria drug (Morimura et al., 2014). These have not been approved for use in vivo as potential deleterious side effects have been highlighted (L. H. Yu et al., 2012).

3.12.8. D3S.0064 - *SMAD4* variant is *associated with both cancer and ID*

- *Clinical manifestation*

An eight year old male patient (D3S.0064) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound)). The patient was born to phenotypically normal parents via caesarean at 30 weeks of gestation due to pre-eclampsia. At least three previous miscarriages were noted. Clinical history revealed a hospital admission for a blood transfusion at six weeks postnatal and another for a tonsillectomy and grommet insertion at three years. Additionally, a hearing test and a vision assessment both revealed vision and hearing problems. The patient was attending endocrine, ENT clinic, neurodevelopmental, psychiatry, and eye clinics. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO terms and codes is summarized in Table 3.16.

Table 3.14. Clinical phenotype for patient D3S.0064 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000878	11 pairs of ribs	Abnormality of the musculoskeletal system
HP:0000718	Aggressive behaviour	Abnormality of the nervous system
HP:0007018	Attention deficit hyperactivity disorder	Abnormality of the nervous system
HP:0000729	Autistic behaviour	Abnormality of the nervous system
HP:0001156	Brachydactyly	Abnormality of the musculoskeletal system
HP:0004209	Clinodactyly of the 5th finger	Abnormality of the musculoskeletal system
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005832	Disharmonic delayed bone age	Abnormality of the musculoskeletal system
HP:0001376	Limitation of joint mobility	Abnormality of the musculoskeletal system
HP:0002996	Limited elbow movement	Abnormality of the musculoskeletal system
HP:0000272	Malar flattening	Abnormality of the musculoskeletal system
HP:0008180	Mildly elevated creatine kinase	Abnormality of metabolism/homeostasis
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0000545	Myopia	Abnormality of the eye
HP:0001622	Premature birth	Abnormality of prenatal development or birth
HP:0000508	Ptosis	Abnormality of the eye
HP:0011110	Recurrent tonsillitis	Abnormality of the cardiovascular system
HP:0004482	Relative macrocephaly	Abnormality of the musculoskeletal system
HP:0004991	Rhizomatic arm shortening	Abnormality of the musculoskeletal system
HP:0002650	Scoliosis	Abnormality of the musculoskeletal system
HP:0004322	Short stature	Growth abnormality

HP:0005720	Shortening of all metacarpals	Abnormality of the musculoskeletal system
HP:0000274	Small face	Abnormality of head or neck
HP:0001792	Small nail	Abnormality of the integument
HP:0000486	Strabismus	Abnormality of the eye
HP:0000325	Triangular face	Abnormality of head or neck

- *Molecular evaluation*

A heterozygous missense variant in *SMAD4*, (c.1498A>G) causing a p.Ile500Val change was detected in the patient. The gene encodes a protein critical for TGF-b signalling which is critical for skeletal growth (Zhao et al., 2018)

WES data was unavailable for both parents. Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.22). Sequencing of the mother's DNA demonstrated a heterozygous variant, and a normal reference sequence was identified in the father. This variant was classified as ACMG-AMP pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.12.

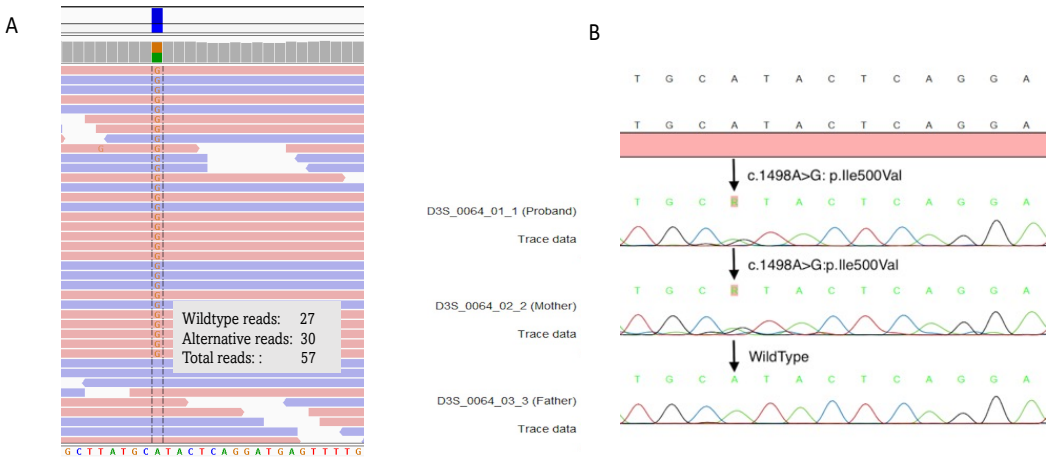

Figure 3.22. Validation of *SMAD4* variant in patient D3S.0064. (A) A heterozygous *SMAD4* variant visualized by IGV (reference base A in green, variant base G in orange) in the proband. A total of 57 reads covers the region with 30 supporting the alternative allele (VAF 52.6%). B) Sanger sequencing analysis of the *SMAD4* variant visualized by CLC bio in the proband, mother, and father confirms the variant is maternally inherited.

- *Molecular diagnosis*

The variant was previously shown to be a gain-in-function missense variant that is localized at the Mad Homology 2 (MH2) domain of SMAD4, leading to its over expression and the dysfunction of the TGF-b signalling (Goff et al., 2012). Pathogenic variants in the gene are known to be associated with Myhre Syndrome (Piccolo et al., 2014). characterized by ID, short stature, hypertropia, hearing loss, characteristic facial dysmorphisms and skeletal abnormalities

(Goff et al., 2014). An evaluation of the clinical features in patients diagnosed with the disorder revealed a large clinical phenotype overlap with those noted in our patient. These included but were not limited to delayed speech and language development, short stature, autistic behaviour, brachydactyly, and numerous skeletal abnormalities (See table). Thus, confirming a positive diagnosis of Myhre syndrome (#139210).

The mutation is also localized in a location that is additionally a hotspot for cancers (Caputo et al., 2012) and mutations associated with Juvenile Polyposis Syndrome (JPS), which predispositions patients to the development of hamartomatous polyps in the gastrointestinal system (van Hattem et al., 2008). In a report of 34 individuals with a SMAD4 causative variant reported an incidence of colonic pops in 31/32 (97%) developed colonic polyps (diagnosed between ages 4 and 51 years), 21/31 (68%) developed gastric polyps, and 76% had some feature of HHT (Wain et al 2014). JPS Patients have been shown to be at an increased risk of 11% to 86% for gastrointestinal cancer, which includes colorectal cancer with a lifetime risk of 40.0%-50% (Aytac et al., 2015; Haeger et al., 2015; Liao et al., 2019; van Hattem et al., 2008). In a study of 27 patients with JPS due to the identification of a causative mutation in SMAD4, 15% of individuals developed cancer (Aytac et al., 2015).

The National Comprehensive Cancer Network (NCCN) makes recommendations for monitoring and testing in patients with a pathogenic SMAD4 variant (Weiss et al., 2021). Based on the NCCN documentation, we recommended testing any available family member, given that the mother is heterozygous for the variant, there is a 50% risk of hereditary hemorrhagic telangiectasia (HHT) development in the siblings. The increased development of HHT and colon cancer has been reported in individuals with a pathogenic SMAD4 mutation but no JPS. Additionally, it is recommended that surveillance for both colon and stomach cancers be initiated from around the age of 15 years (Weiss et al., 2021).

3.12.9. D3S.0071 - *EMHT1* variant *Kleefstra syndrome no hypotonia*

- *Clinical manifestation*

A six year old female patient (D3S.0071) was recruited based on DDD-Africa recruitment criteria Inclusion A (Intellectual disability (moderate to profound). The patient was born via caesarean to phenotypically normal parents at 39 weeks of gestation. Additionally, CT scans and vision and hearing tests showed no abnormalities and were within normal range. The full

list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO terms and codes is summarized in Table 3.17

Table 3.15. Clinical phenotype for patient D3S.0071 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000717	Autism	Abnormality of the nervous system
HP:0000670	Carious teeth	Abnormality of head or neck
HP:0001814	Deep-set nails	Abnormality of the integument
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0000286	Epicanthus	Abnormality of head or neck
HP:0040019	Finger clinodactyly	Abnormality of the musculoskeletal system
HP:0002007	Frontal bossing	Abnormality of the musculoskeletal system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0001382	Joint hypermobility	Abnormality of the musculoskeletal system
HP:0000396	Overfolded helix	Abnormality of the ear
HP:0000817	Poor eye contact	Abnormality of the nervous system
HP:0004467	Preauricular pit	Abnormality of the integument
HP:0100023	Recurrent hand flapping	Abnormality of the nervous system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0005338	Sparse lateral eyebrow	Abnormality of the integument
HP:0000574	Thick eyebrow	Abnormality of the integument
HP:0001537	Umbilical hernia	Abnormality of the musculoskeletal system
HP:0000154	Wide mouth	Abnormality of head or neck

- *Molecular evaluation*

Exome sequencing identified a pathogenic variant splice-donor variant in *EHMT1*, c.2712+1G>A. The variant has been shown to result in exon 18 skipping but with no disruption to the open reading frame, thus a full-length EHMT1 protein is produced (Rump et al., 2013).

Sanger sequencing confirmed the presence of the variant in a hemizygous state in the patient (Figure 3.23). Segregation analysis could not be performed as DNA from the father was not available, but the heterozygous variant was detected in the mother thus suggesting maternal inheritance. This variant was classified as ACMG-AMP pathogenic. The ACMG-AMP codes and evidence for variant classification are provided in Table 3.5.

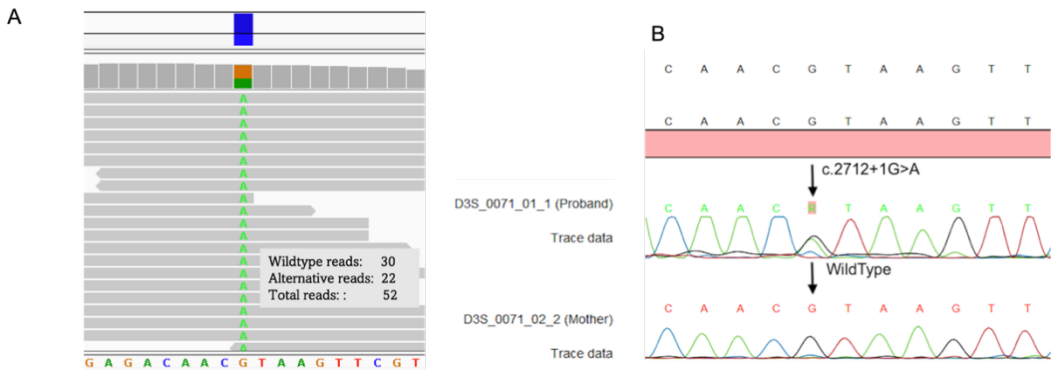

Figure 3.23. Validation of *EMHT1* variant in patient D3S.0071. (A) A heterozygous *EMHT1* variant visualized by IGV (reference base G in orange, variant base A in green) in the proband. A total of 52 reads covers the region with 22 reads supporting the alternative allele (VAF 42.3%) (B) Sanger sequencing analysis of the *EMHT1* variant visualized by CLC bio in the proband, shows the variant not to have been inherited from the mother.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with Kleefstra syndrome (610253) characterized by DD, childhood hypotonia, and distinct facial features. Our patient presented with DD, autism spectrum behaviour, umbilical hernia, and distinct facial features which included, frontal bossing, depressed nasal bridge, Hypertelorism. The syndromes characteristic childhood hypotonia was not reported in our patient.

In all patients in which the variant was identified and all other cases of EHMT1 Kleefstra syndrome, *de novo* pattern of inheritance was reported. De novo pattern of inheritance could not be confirmed due to the unavailability of the father's DNA sample, but the variant was not identified in the mother based on both exome and Sanger sequencing results thus there is a high likelihood that it is *de novo* in our patient. Germline maternal mosaicism has been reported in some patients with unaffected parents, we thus cannot exclude maternal mosaicism in this family and thus recommend the mother be tested for mosaicism using for example MLPA testing.

3.12.10. D3S.0082 - *SON* variant in gene hotspot resulting in *ZTTK syndrome*

- *Clinical manifestation*

A 10 year old male patient (D3S.0082) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound)). The patient was born to phenotypically normal parents at 40 weeks' gestation following prolonged labour and foetal nuchal cord. Diagnostic tests and scans performed prior to recruitment included Karyotype,

MLPA, Array-CGH and a cytoscan which were all normal. Family history revealed a death of a male sibling at six months old due to severe diarrhoea and vomiting. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.18.

Table 3.16. Clinical phenotype for patient D3S.0082 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000717	Autism	Abnormality of the nervous system
HP:0009768	Broad phalanges of the hand	Abnormality of the musculoskeletal system
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0000752	Hyperactivity	Abnormality of the nervous system
HP:0010485	Hyperextensibility at elbow	Abnormality of the musculoskeletal system
HP:0005722	Hyperextensible thumb	Abnormality of the musculoskeletal system
HP:0000953	Hyperpigmentation of the skin	Abnormality of the integument
HP:0000256	Macrocephaly	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0001513	Obesity	Growth abnormality
HP:0011110	Recurrent tonsillitis	Abnormality of the cardiovascular system
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck
HP:0000431	Wide nasal bridge	Abnormality of head or neck

- *Molecular evaluation*

A *de novo* variant in *SON* (c.5753_5756del) causing a p. Val1918GlufsTer87 change was detected in the patient. The gene encodes a 2426aa protein which is involved in the cell cycle process and is important for the maintenance of embryonic stem cells (Lu et al., 2013).

Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.24). Segregation analysis could not be performed as DNA from the father was not available but sequence from mother showed a normal reference sequence. This variant was classified as ACMG-AMP pathogenic as evidenced in Table 3.5.

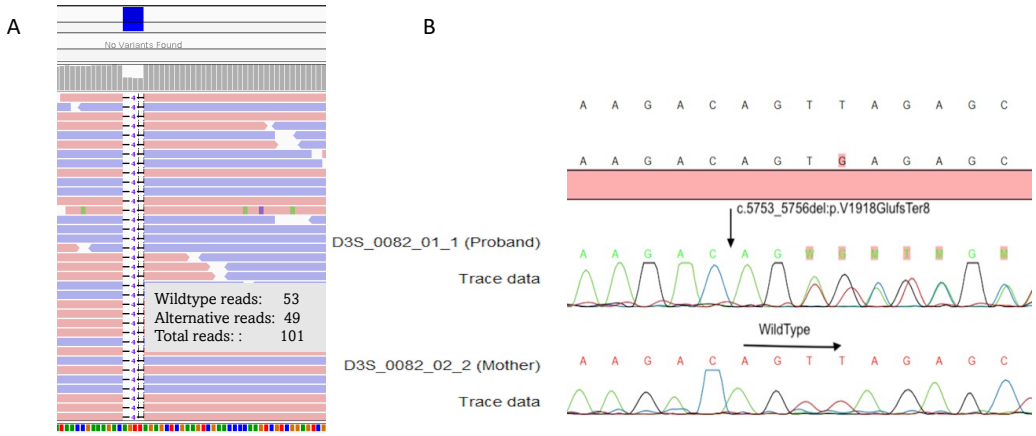

Figure 3.24. Validation of *SON* variant in patient D3S.0082. (A) A heterozygous *SON* 4-base pair deletion visualized by in the proband A total of 101 reads covers the region with 49 supporting the alternating allele (VAF 48.5%). (B) Sanger sequencing analysis of the *SON* variant visualized by CLC bio in the proband, and mother confirms the variant is not inherited from the mother.

- *Molecular diagnosis*

Pathogenic variants in the gene show a clustering in exon 3 of the gene which makes up about 82% of the entire gene's protein coding region (KHAN et al., 1994; Lu et al., 2013). The gene is expressed in a range of human tissue (Lu et al., 2013) thus disease-causing variants in the gene result in a highly diverse group of clinical features (Peng et al., 2021). A study by Peng et al (2021), identified the variant in a Chinese female in addition to another eight patients diagnosed with Zhu-Tokita-Takenouchi-Kim (ZTTT syndrome). The patient's presented exclusively with short stature, strabismus, congenital heart defects, urogenital malformations, hand, and feet abnormalities in addition to ID and facial dysmorphisms.

Our patient presented with similar distinctive dysmorphic features but did not present with any of the distinctive clinical features noted. Additionally, our patient was obese which contradicts the observation of a failure to thrive and feeding difficulties noted for most patients with pathogenic SON variants (Tokita et al., 2016).

3.10.11. D3S.0086 - *STXBP1* variant associated with suppression bursts

- *Clinical manifestation*

A 10 year old male patient (D3S.0086) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound)). He was born to phenotypically normal parents via caesarean at 40 weeks of gestation due to breech presentation. Clinical history revealed several hospital admissions for poor feeding and apenic episodes, febrile convulsions, tonsillectomy and adenoidectomy. Diagnostic testing and scanning performed prior to recruitment included Karyotype, and MLPA which were all normal. Additionally, MR

imaging results showed frontal horns and EEG results were normal. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.19.

Table 3.17. Clinical phenotype for patient D3S.0086 described using Human Phenotype Ontology terms

HPO ID	HPO term	HPO category
HP:0000164	Abnormality of the dentition	Abnormality of head or neck
HP:0000718	Aggressive behaviour	Abnormality of the nervous system
HP:0007874	Almond-shaped palpebral fissure	Abnormality of head or neck
HP:0000414	Bulbous nose	Abnormality of head or neck
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0002373	Febrile seizure (within the age range of 3 months to 6 years)	Abnormality of the nervous system
HP:0011968	Feeding difficulties	Abnormality of the digestive system
HP:0002970	Genu varum	Abnormality of the musculoskeletal system
HP:0001265	Hyporeflexia	Abnormality of the nervous system
HP:0001252	Hypotonia	Abnormality of the musculoskeletal system
HP:0003186	Inverted nipples	Abnormality of the breast
HP:0000341	Narrow forehead	Abnormality of head or neck
HP:0001513	Obesity	Growth abnormality
HP:0030276	obsolete small scrotum	Other / obsolete terms
HP:0100716	Self-injurious behaviour	Abnormality of the nervous system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0009237	Short 5th finger	Abnormality of the musculoskeletal system
HP:0001773	Short foot	Abnormality of the musculoskeletal system
HP:0001518	Small for gestational age	Growth abnormality
HP:0000046	Small scrotum	Abnormality of the genitourinary system
HP:0001182	Tapered finger	Abnormality of the musculoskeletal system
HP:0000215	Thick upper lip vermilion	Abnormality of head or neck

- *Molecular evaluation*

A missense variant in *STXBP1,* c.416C>T: p. Pro139Leu (NM_003165.6) was detected in the patient. The gene encodes MUNC18-1 protein which is involved in regulating the release of neurotransmitters at the synaptic membrane (verhang 2000, Toonan 2006). Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.25). Segregation analysis was performed following sequencing of both parents DNA. A reference allele was detected thus confirming *de novo* mode of inheritance. This variant was classified as ACMG-AMP pathogenic as shown in table 3.5.

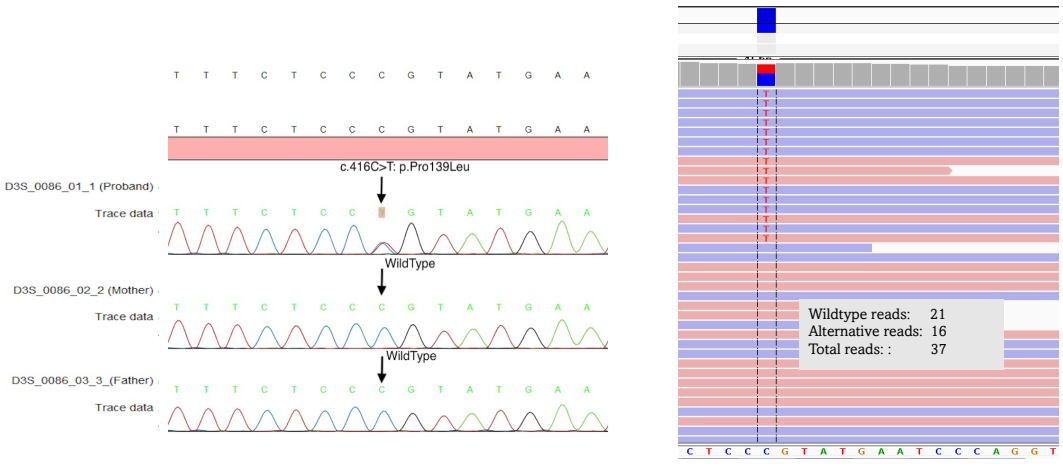

Figure 3.25. Validation of *STXBP1* variant in patient D3S.0086. (A) A heterozygous *STXBP1* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 37 reads covers the region with 16 supporting the alternative allele (VAF 43.2%). (B) Sanger sequencing analysis of the *STXBP1* variant visualized by CLC bio in the proband, mother, and father confirms the variant is *de novo* inherited.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with Developmental and epileptic encephalopathy 4 (612164) (Khaikin & Mercimek-Mahmutoglu, 2016). Our patient presented with a history of seizures which were reported to have stopped at 6 years old following medical treatment. A few cases of ID without early on set epilepsy (EEE) have been reported. In EEE patients are reported to be seizure free for a few years following medical intervention but reoccur following suppression bursts and progress to intractable seizures in patients. Seizure episodes were reported by parents and thus we suggest that epilepsy presence be re-evaluated at the follow-up clinic visit. If the present of epilepsy is confirmed, we recommend the administration of vigabatrin drug which has been proven to be effective in patients with the same type of mutation (Barcia et al., 2014).

3.12.12. D3S.0094 - *DDX3X* related ID with Snijders Blok type

- *Clinical manifestation*

A two year old female patient (D3S.0094) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound)). She was born to phenotypically normal parents at 40 weeks of gestation. The mother was on antiretroviral treatment and antimicrobial medication for HIV infection and pulmonary TB infection respectively. Additionally, alcohol use during the pregnancy was reported. Clinical history revealed two separate hospital admissions for seizure episodes and seizure management using Epilim from three months of age. Diagnostic testing performed prior to recruitment included

Karyotype, MLPA and Fragile X which were all normal. Additionally, MR imaging results at 12 months showed frontal temporal atrophy and corpus calloum hypoplasia suggestive of delayed myelination. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.20.

Table 3.18. Clinical phenotype for patient D3S.0081 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000729	Autistic behaviour	Abnormality of the nervous system
HP:0003763	Bruxism	Abnormality of the nervous system
HP:0010862	Delayed fine motor development	Abnormality of the nervous system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0001382	Joint hypermobility	Abnormality of the musculoskeletal system
HP:0001388	Joint laxity	Abnormality of the musculoskeletal system
HP:0000733	Motor stereotypy	Abnormality of the nervous system
HP:0007832	Pigmentation of the sclera	Abnormality of the eye
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0005585	Spotty hyperpigmentation	Abnormality of the integument
HP:0000486	Strabismus	Abnormality of the eye
HP:0025339	Superficial episcleral hyperaemia	Abnormality of the cardiovascular system

- *Molecular evaluation*

A missense variant in *DDX3X*, c.931C>T. p. Arg311Ter resulting in a stop-gain was detected in the patient. The gene encodes a conserved DEAD-box RNA helicase that is involved in RNA metabolism which includes transcription and mRNA export and translation, and dendrite formation and neuronal migration. Sanger sequencing confirmed the presence of the variant in a hemizygous state in the patient (Figure 3.26). Segregation analysis could not be performed as only DNA from the patient and mother was available. However, the mother's DNA showed normal reference sequence. This variant was classified as ACMG-AMP pathogenic as displayed in table 3.5.

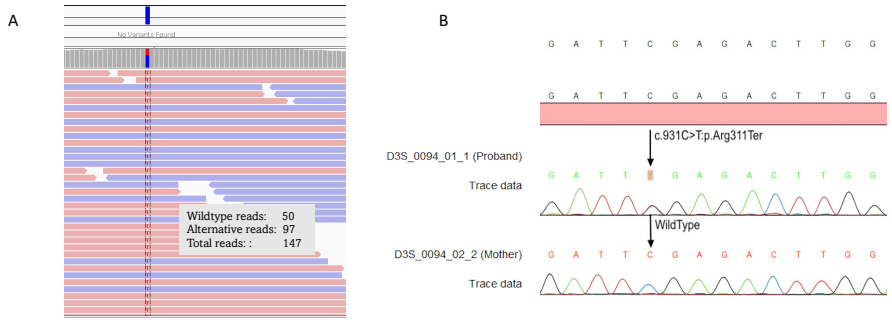

Figure 3.26. Validation of *DDX3X* variant in patient D3S.0094. (A). A hemizygous *DDX3X* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 147 reads covers the first region with 97 covering the alternative (VAF 66%). (B) Sanger sequencing analysis of the *DDX3X* variant visualized by CLC bio in the proband, and mother, shows the variant is not maternally inherited.

- *Molecular diagnosis*

Mutations in the gene have been associated with DD Intellectual developmental disorder, X-linked syndromic, Snijders Blok type (300958) (Blok et al., 2015). Most individuals present with abnormal corpus callosum, ventriculomegaly, hypotonia, microcephaly, and abnormal brain structures. Our patient's MRI results showed Frontotemporal atrophy, corpus callosum hypoplasia, delayed myelination (at one year), and visual impairment Hypotonia in the patient had led to thoracolumbar kyphosis. The variant has been reported in other similarly affected female patients (Blok et al., 2015). Streaky hypopigmentation on the arms and chest and Blaschko's lines on the back were noted, which have not been reported for any other syndrome individuals.

In all reported female individuals with a molecular confirmation of DDX3X-related ID, *de novo* variants have been identified. We were unable to confirm. *de novo* inheritance as the father was unavailable at the time of recruitment. However, we did observe a homozygous reference variant in the mother (Figure 3.30) and we suspect a de novo inheritance. Pathogenic variants in the gene associated with DDX3X-related neurodevelopmental Disorder. (Blok et al., 2015).

3.12.13. D3S.0098 - *OFD1* variant confirms suspected clinical presentation

- *Clinical manifestation*

A six year old female patient (D3S.0098) was recruited based on DDD-Africa recruitment criteria inclusion A (Intellectual disability (moderate to profound)).and C (moderate DD and macrocephaly). She was born to a phenotypically normal father and delayed mother characterized by intellectual disability, abnormalities of the tongue, median cleft lip,

brachydactyly and multiple renal cysts. Pregnancy was unremarkable. Clinical history revealed limb defects at birth which included a duplicated thumb and right 2-3 syndactyly hallux. A hospital admission for bronchitis was also noted. Results of brain MR imaging showed agenesis of corpus callosum and interhemispheric cysts variant supporting an Oro-facio-digital syndrome I clinical diagnosis. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.21

Table 3.19. Clinical phenotype for patient D3S.0098 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0001233	2-3 finger syndactyly	Abnormality of the musculoskeletal system
HP:0001274	Agenesis of corpus callosum	Abnormality of the nervous system
HP:0001156	Brachydactyly	Abnormality of the musculoskeletal system
HP:0011229	Broad eyebrow	Abnormality of the integument
HP:0040024	Clinodactyly of the 3rd finger	Abnormality of the musculoskeletal system
HP:0040025	Clinodactyly of the 4th finger	Abnormality of the musculoskeletal system
HP:0000378	Cupped ear	Abnormality of the ear
HP:0001305	Dandy-Walker malformation	Abnormality of the musculoskeletal system
HP:0001476	Delayed closure of the anterior fontanelle	Abnormality of the musculoskeletal system
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0000953	Hyperpigmentation of the skin	Abnormality of the integument
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0032327	Interhemispheric cyst	Abnormality of the nervous system
HP:0000180	Lobulated tongue	Abnormality of head or neck
HP:0005439	Maxillozygomatic hypoplasia	Abnormality of the musculoskeletal system
HP:0010747	Medial flaring of the eyebrow	Abnormality of the integument
HP:0000161	Median cleft lip	Abnormality of head or neck
HP:0004987	Mesomelic leg shortening	Abnormality of the musculoskeletal system
HP:0011800	Midface retrusion	Abnormality of head or neck
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0010101	Partial duplication of the phalanges of the hallux	Abnormality of the musculoskeletal system
HP:0009944	Partial duplication of thumb phalanx	Abnormality of the musculoskeletal system
HP:0000470	Short neck	Abnormality of head or neck
HP:0000188	Short upper lip	Abnormality of head or neck
HP:0002982	Tibial bowing	Abnormality of the musculoskeletal system
HP:0000431	Wide nasal bridge	Abnormality of head or neck

- *Molecular evaluation*

A frameshift variant in *OFD1* gene, c.702dup (p. Tyr238ValfsTer2) was detected in the patient. Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure.3.27). Segregation analysis could not be performed as DNA from each patient was unavailable. This variant was classified as ACMG-AMP, pathogenic as displayed in table 3.5.

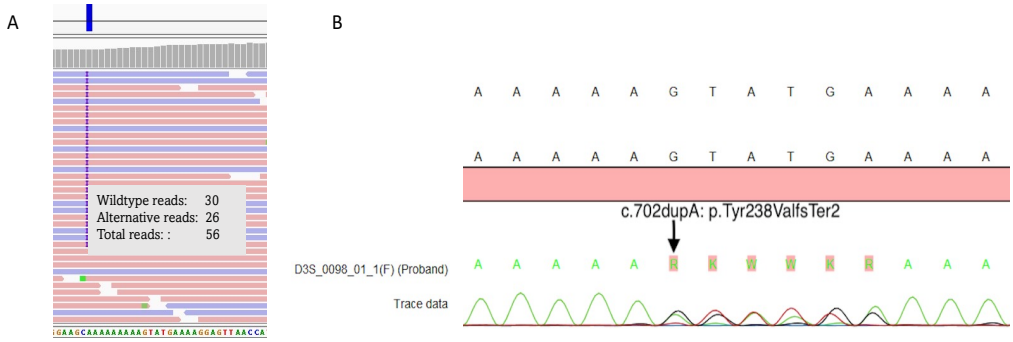

Figure 3.27. Validation of *OFD1* variant in patient D3S.0094. (A) A heterozygous *OFD1* 1-base pair insertion of A is visualized by IGV in the proband. A total of 56 reads with 26 reads supporting the alternative allele (VAF 46.4%). (B) Sanger sequencing analysis of the *OFD1* variant visualized by CLC bio in the proband confirms variant in heterozygous state.

- *Molecular diagnosis*

Our patient presented with oral, digital defects, and brain MR imaging showing agenesis of corpus callosum and interhemispheric cysts signally Dandy-Walker variant was present in our patient. Clinical phenotype fit OFD1 diagnosis, which was suspected during patient clinical review and our results provide a positive molecular confirmation. The patient was born to a phenotypically normal father and delayed mother characterized by intellectual disability, abnormalities of the tongue, median cleft lip, brachydactyly and multiple renal cysts. The clinical information for the mother was retrieved from her clinical file, as she was not available at recruitment. Previous reports of this variant have shown familial inheritance and de novo inheritance. Pattern of inheritance could not be confirmed due to the unavailability both parents DNA samples. However, we suspect clinical diagnosis of OFD1 in the mother.

It is reported that approximately 50% of OFD1 individuals will develop polycystic kidney disease which may progress to end-stage renal disease. It is recommended that affected patients be monitored from 10 years of age for kidney disease through yearly blood serum creatinine concentration assessments, and ultrasound examinations for cyst development in the kidneys, liver, ovary and pancreas. (gene reviews). Genetic counseling should cover risk assessment as

there is a 50% risk of penetrance. When the patient is older, prenatal testing are should be considered.

3.12.14. D3S.0111 - *PTEN* hamartoma tumor syndrome and cancer risk

- *Clinical manifestation*

A six year old female patient (D3S.0111) was recruited based on DDD-Africa inclusion D criteria (mild to moderate intellectual disability and dysmorphic features). She was born to phenotypically normal parents at 40 weeks' gestation. Ultrasound abnormalities were noted but no clear information given. Both parents and older brother noted to have repeated schools, but no other delays reported. Additionally, brain MR imaging revealed cerebral and vermis atrophy and macrocephaly was noted. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.22

Table 3.20. Clinical phenotype for patient D3S.0111 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000463	Anteverted nares	Abnormality of head or neck
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0001272	Cerebellar atrophy	Abnormality of the nervous system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0100876	Infra-orbital crease	Abnormality of head or neck
HP:0000256	Macrocephaly	Abnormality of the musculoskeletal system
HP:0000691	Microdontia	Abnormality of head or neck
HP:0011342	Mild global developmental delay	Abnormality of the nervous system
HP:0000396	Overfolded helix	Abnormality of the ear
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0000322	Short philtrum	Abnormality of head or neck
HP:0000687	Widely spaced teeth	Abnormality of head or neck

- *Molecular evaluation*

Exome trio analysis revealed a missense variant in *PTEN* (601728), c.1129C>A causing a change from Threonine to Alanine in amino acid 534 (p. Thr534Ala). PTEN is a modulator of the mTOR pathway which is important for cell proliferation, cell survival, apoptosis, metabolism, and cancer (Hay & Sonenberg, 2004; Moore et al., 1996). The gene acts through the dephosphorylation of PIP3 to PIP2 required for Akt activation in the PI3K/Akt pathway (Fricano-Kugler et al., 2018). Dephosphorylation by PTEN limits Akt activity, therefore, limiting cell proliferation and the propagation of cancer cells and as such is regarded as a cancer suppressor gene (Trotman et al., 2007).

Segregation analysis could not be performed as the father was not available at recruitment, but sequencing of the mothers DNA demonstrated a reference allele sequence (Figure 3.28). The variant was classified using standard ACMG-AMP guidelines as likely pathogenic.

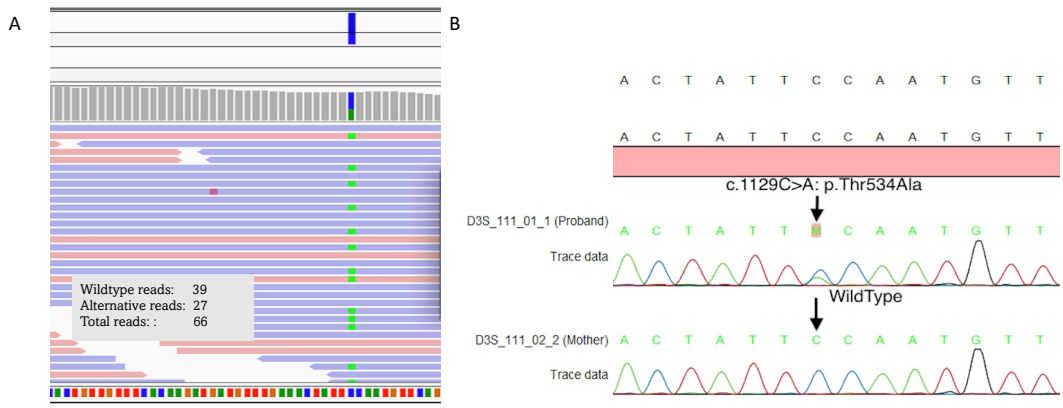

Figure 3.28. Validation of *PTEN* variant in patient D3S.0111. (A) A heterozygous *PTEN* variant visualized by IGV (reference base C in blue, variant base A in green) in the proband. A total of 66 reads covers the region with 27 supporting the alternative allele (VAF 40.9%). B) Sanger sequencing analysis of the *PTEN* variant visualized by CLC bio in the proband, and mother shows it was not inherited from the mother.

- *Molecular diagnosis*

Loss-of –function mutations in the gene have been associated with PTEN hamartoma tumor syndromes due hyperactivation of the mTOR pathway resulting in an overgrowth of the brain (macrocephaly) and ID (Hansen-Kiss et al., 2017). Our patient presented clinically with macrocephaly and a range of CNS abnormalities determined to be consistent with PTEN associated DD.

Mutations in the gene are the second most common in a range of cancers and tumor development (Hendricks et al., 2021). Our female patient is at high risk of developing a range of cancer which includes, thyroid cancer (35% risk) from as early as 7 years old which is around the same age as our female patient, and an even higher risk for breast cancer, which has been reported in ~ 85% of all women around the age of 40 (Hendricks et al., 2021; Ngeow et al., 2017). Thus, we recommend monitoring and screening for the various cancers and tumour development in the patient and any first-degree relatives.

3.12.15. D3S.0118 - *FLNA* variant associated with X-linked OPD 1

- *Clinical manifestation*

An eight year old male patient (D3S.0118) was recruited based on DDD-Africa recruitment criteria inclusion D (Suspected multifactorial or suspected environmental cause). He was born to phenotypically normal parents at 40 weeks of gestation. Two previous miscarriages were reported. Clinical history revealed several hospital admissions for seizure and was being managed using Epilim. Clinical examination revealed a range of dysmorphic features including abnormalities of the digits and characteristic facial features. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.23

Table 3.21. Clinical phenotype for patient D3S.0118 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0005807	Absent distal phalanges	Abnormality of the musculoskeletal system
HP:0012386	Absent hallux	Abnormality of the musculoskeletal system
HP:0010744	Absent metatarsal bone	Abnormality of the musculoskeletal system
HP:0000463	Anteverted nares	Abnormality of head or neck
HP:0006710	Hypoplasia of the clavicles	Abnormality of the musculoskeletal system
HP:0008362	Hypoplasia of the hallux	Abnormality of the musculoskeletal system
HP:0008936	Axial hypotonia	Abnormality of the musculoskeletal system
HP:0001348	Brisk reflexes	Abnormality of the nervous system
HP:0011304	Broad thumb	Abnormality of the musculoskeletal system
HP:0002072	Chorea	Abnormality of the nervous system
HP:0004209	Clinodactyly of the 5th finger	Abnormality of the musculoskeletal system
HP:0006184	Decreased palmar creases	Abnormality of the integument
HP:0001814	Deep-set nails	Abnormality of the integument
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0000437	Depressed nasal tip	Abnormality of head or neck
HP:0000494	Downslanted palpebral fissures	Abnormality of head or neck
HP:0002263	Exaggerated cupid's bow	Abnormality of head or neck
HP:0000348	High forehead	Abnormality of head or neck
HP:0005639	Hyperextensible hand joints	Abnormality of the musculoskeletal system
HP:0000953	Hyperpigmentation of the skin	Abnormality of the integument
HP:0001347	Hyperreflexia	Abnormality of the nervous system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0001276	Hypertonia	Abnormality of the musculoskeletal system
HP:0000601	Hypotelorism	Abnormality of the eye
HP:0004305	Involuntary movements	Abnormality of the nervous system
HP:0001377	Limited elbow extension	Abnormality of the musculoskeletal system
HP:0000637	Long palpebral fissure	Abnormality of head or neck
HP:0000369	Low-set ears	Abnormality of the ear
HP:0000368	Low-set, posteriorly rotated ears	Abnormality of the ear
HP:0000308	Micro retrognathia	Abnormality of the musculoskeletal system

HP:0008551	Microtia	Abnormality of the ear
HP:0011800	Midface retrusion	Abnormality of head or neck
HP:0000341	Narrow forehead	Abnormality of head or neck
HP:0000160	Narrow mouth	Abnormality of head or neck
HP:0000639	Nystagmus	Abnormality of the eye
HP:0000396	Overfolded helix	Abnormality of the ear
HP:0000767	Pectus excavatum	Abnormality of the musculoskeletal system
HP:0001763	Pes planus	Abnormality of limbs
HP:0001357	Plagiocephaly	Abnormality of the musculoskeletal system
HP:0002057	Prominent glabella	Abnormality of head or neck
HP:0004991	Rhizomatic arm shortening	Abnormality of the musculoskeletal system
HP:0001852	Sandal gap	Abnormality of the musculoskeletal system
HP:0001250	Seizure	Abnormality of the nervous system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0000049	Shawl scrotum	Abnormality of the genitourinary system
HP:0003196	Short nose	Abnormality of head or neck
HP:0009778	Short thumb	Abnormality of the musculoskeletal system
HP:0000319	Smooth philtrum	Abnormality of head or neck
HP:0000486	Strabismus	Abnormality of the eye
HP:0000574	Thick eyebrow	Abnormality of the integument
HP:0010539	Thin calvarium	Abnormality of the musculoskeletal system
HP:0000233	Thin vermilion border	Abnormality of head or neck
HP:0002942	Thoracic kyphosis	Abnormality of the musculoskeletal system
HP:0002943	Thoracic scoliosis	Abnormality of the musculoskeletal system
HP:0001537	Umbilical hernia	Abnormality of the musculoskeletal system
HP:0012454	Unilateral wrist flexion contracture	Abnormality of the musculoskeletal system

- *Molecular evaluation*

Exome trio analysis identified a missense variant in *FLNA* (300017) gene, c.620C>T, resulting in a change of proline to Leucine at amino acid 207 (p. Pro207Leu). The FLNA gene encodes a 280-kD dimeric Filamin A protein that is responsible for attaching membrane proteins to the cytoskeletal membrane (Nakamura et al., 2007). This function is critical for the regulation of cell functions such as signal transduction and neuron migration (Maestrini et al., 1993). The protein is widely expressed thus mutations in the gene impacting protein function result in a wide spectrum of phenotypes in the affected patients. Sanger sequencing confirmed the presence of a hemizygous variant in the patient (Figure 3.29) that was observed only in the proband and the mother. Moreover, the variant has previously been reported in ClinVar. This variant was classified using standard ACMG-AMP as pathogenic (see table 3.5).

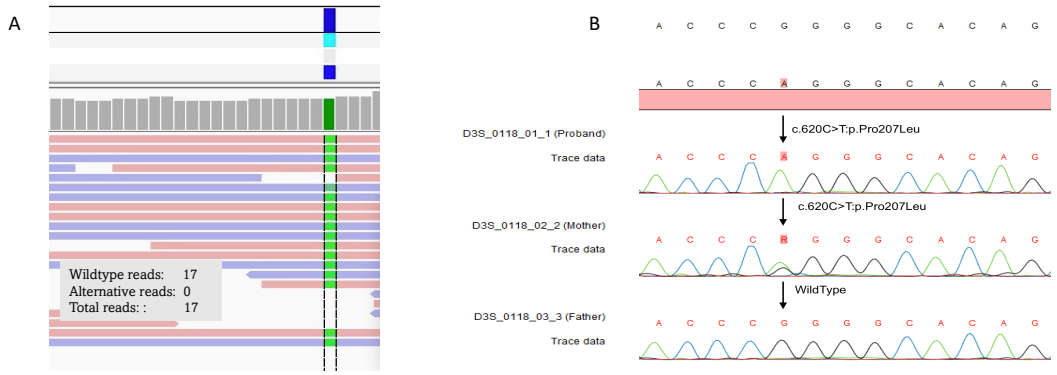

Figure 3.29. Validation of *FLNA* variant in patient D3S. 0118. (A) A hemizygous *FLNA* variant visualized by IGV (variant base A in green) in the proband. A total of 17 reads covers the region with all supporting the alternative allele (VAF 100%). (B) Sanger sequencing analysis of the *FLNA* variant visualized by CLC bio in the proband, mother, and father confirms the variant is maternally inherited.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with X-linked disorders, Melnick-Needles syndrome and, otopalatodigital (OPD) spectrum disorders type 1 and type 2 (Robertson, 2019) Pathogenic variants in the *FLNA* gene are associated with a spectrum of clinically overlapping X-linked disorders, Melnick-Needles syndrome, otopalatodigital (OPD) spectrum disorders type 1 and type 2, and a range of other cardiovascular abnormalities (Robertson, 2019).

In-vivo studies have indicated that all missense variants and small deletions identified in OPD1 are clustered in the actin-binding domain in exon 2-5 of the gene, resulting in gain-of-function effect which leads to in increased levels of FLNA protein (Robertson, 2019). However, the genotype-phenotype correlation is still not well understood.

The missense variant (c.620C>T p. (Pro207Leu) has been previously reported in four other patients (VCV000011755.3) and in all cases was associated with X-linked otopalatodigital type 1 disorder (# 311300). Our patient presented with a missense mutation in exon 3 along with clinical features corresponding to an OPD spectrum disorder which included a range of skeletal dysplasia, anomalies of the digits such as short thumb, sandal gap and limited joint mobility (limited elbow extension, Hyperextensible hand joints). However, our patient presented with mild-moderate ID which contradicts normal ID reported in all cases of OPD1 thus far. Therefore, we could not make a definitive differential diagnosis, thus, the patient was diagnosed with OPD spectrum disorder.

Hearing loss has been documented to present in patients with the disorder. No hearing problems have been reported in our patient, but we recommend that regular testing be performed to allow

for early hearing loss diagnosis and interventions such as the use of hearing aids Additionally, genetic counseling relating to family planning and future pregnancies should include recommendations for prenatal testing. The missense mutation could potentially be identified, and a prenatal diagnosis provided via amniocentesis or chorionic villus sampling (CVS).

3.12.16. D3S.0124 - *PACS1* variant in Schuurs-Hoeijmakers syndrome

- *Clinical manifestation*

A nine year old female patient (D3S.0124) was recruited based on DDD-Africa recruitment criteria inclusion D (Suspected multifactorial or suspected environmental cause). She was born to phenotypically normal parents at 40 weeks of gestation. Clinical history revealed post-natal meconium aspiration syndrome. She was also diagnosed with epilepsy at age of six years old, and prescribed Risperdal. Diagnostic tests performed prior to recruitment included Karyotype, MLPA, QFPCR which were all normal. Additionally, ECG, CT brain scan MR imaging and hearing test results showed no abnormalities and were all within normal range. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.24.

Table 3.22. Clinical phenotype for patient D3S.0124 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000752	Hyperactivity	Abnormality of the nervous system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0000637	Long palpebral fissure	Abnormality of head or neck
HP:0000343	Long philtrum	Abnormality of head or neck
HP:0002162	Low posterior hairline	Abnormality of head or neck
HP:0000272	Malar flattening	Abnormality of the musculoskeletal system
HP:0003502	Mild short stature	Growth abnormality
HP:0001763	Pes planus	Abnormality of limbs
HP:0002465	Poor speech	Abnormality of the nervous system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0000914	Shield chest	Abnormality of the musculoskeletal system
HP:0000470	Short neck	Abnormality of head or neck
HP:0003196	Short nose	Abnormality of head or neck
HP:0000319	Smooth philtrum	Abnormality of head or neck
HP:0001182	Tapered finger	Abnormality of the musculoskeletal system
HP:0000219	Thin upper lip vermilion	Abnormality of head or neck
HP:0001537	Umbilical hernia	Abnormality of the musculoskeletal system
HP:0000465	Webbed neck	Abnormality of head or neck
HP:0000154	Wide mouth	Abnormality of head or neck
HP:0000445	Wide nose	Abnormality of head or neck

- *Molecular evaluation*

WES data analysis revealed a missense variant in *PACS1*, c.607C>T. p. Arg203T (NM_018026.4). The variant has been indicated to disrupt normal PACS1 function by forming cytoplasmic aggregates and altering protein trafficking (Schuurs- Hoeijmakers et al., 2012). Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.30). Segregation analysis could not be performed as the father was not available at recruitment, but sequencing of the mother's DNA demonstrated a reference allele sequence. The variant was classified as ACMG-AMP pathogenic as shown in table 3.5.

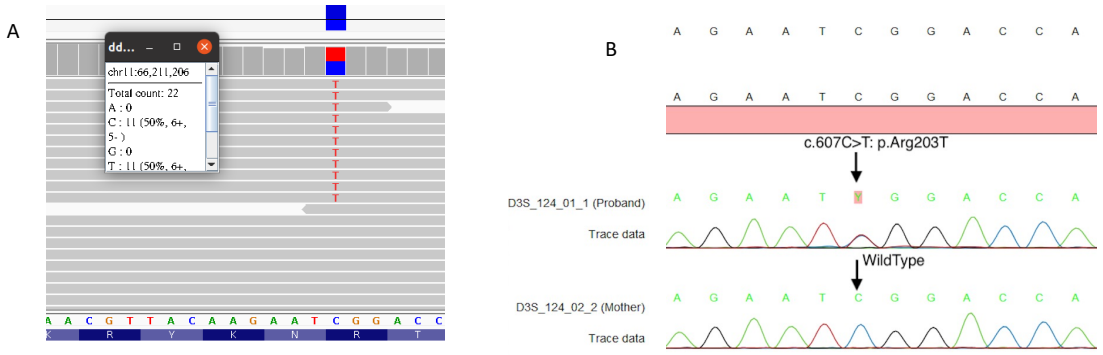

Figure 3.30. Validation of *PACS1* variant in patient D3S.0124. (A) A heterozygous *PACS1* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 22 reads covers the region and 11 support the alternative (VAF 50%). (B) Sanger sequencing analysis of the *PACS1* variant visualized by CLC bio in the proband, and mother shows that the is not inherited from the mother.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with Schuurs-Hoeijmakers syndrome (# 615009) which is characterized by ID, a range of dysmorphic features, hypotonia, seizures, and structural malformations (Schuurs-Hoeijmakers et al., 2016; Stern et al., 2017). Our patient presented with DD and characteristic dysmorphic features that were like those described for a patient with the same molecular diagnosis. These included hypertelorism, palpebral fissures, downturned corners of the mouth, a narrow upper lip, flat philtrum, a short neck and small umbilical hernia.

All cases of Schuurs-Hoeijmakers syndrome show *de novo* inheritance (Dutta, 2019; Schuurs-Hoeijmakers et al., 2012) . *De novo* pattern of inheritance could not be confirmed due to the unavailability of the father's DNA sample, but the variant was not identified in the mother based on both exome and Sanger sequencing results thus there is a high likelihood that it is *de novo* in our patient.

3.12.17. D3S.0129 - *MAPK1* variant associated with Noonan-like clinical presentation

- *Clinical manifestation*

A five year old male patient (D3S.0129) was recruited based on DDD-Africa recruitment criteria inclusion D (Suspected multifactorial or suspected environmental cause). He was born to phenotypically normal parents at 40 weeks of gestation. Clinical history revealed several hospital admissions for seizures and management using Epilim (currently discontinued) but no official epilepsy diagnosis has been made. Diagnostic tests and scans performed prior to recruitment included Karyotype, MLPA, Aneuploidy and brain MR imaging which all showed no abnormalities. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.25.

Table 3.23. Clinical phenotype for patient D3S.0129 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0000964	Eczema	Abnormality of the immune system
HP:0012168	Head-banging	Abnormality of the nervous system
HP:0000953	Hyperpigmentation of the skin	Abnormality of the integument
HP:0000272	Malar flattening	Abnormality of the musculoskeletal system
HP:0001763	Pes planus	Abnormality of limbs
HP:0001250	Seizure	Abnormality of the nervous system
HP:0009237	Short 5th finger	Abnormality of the musculoskeletal system
HP:0002000	Short columella	Abnormality of head or neck
HP:0004322	Short stature	Growth abnormality
HP:0005338	Sparse lateral eyebrow	Abnormality of the integument
HP:0001182	Tapered finger	Abnormality of the musculoskeletal system
HP:0010804	Tented upper lip vermilion	Abnormality of head or neck
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck
HP:0000154	Wide mouth	Abnormality of head or neck

- *Molecular evaluation*

WES data analysis revealed a missense variant in *MAPK1*, c.952G>A p. Asp318Asn. Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.31). The gene encodes extracellular signal-regulated kinases (ERK1/2), which forms part of the MAPK signalling cascade. important for cell proliferation, cell survival, apoptosis, metabolism, and cancer (Tidyman & Rauen, 2009). Segregation analysis could not be performed as the father was not available at recruitment, but sequencing of the mothers DNA

demonstrated a reference allele sequence. The variant was classified as ACMG-AMP pathogenic as evidenced in table 3.5.

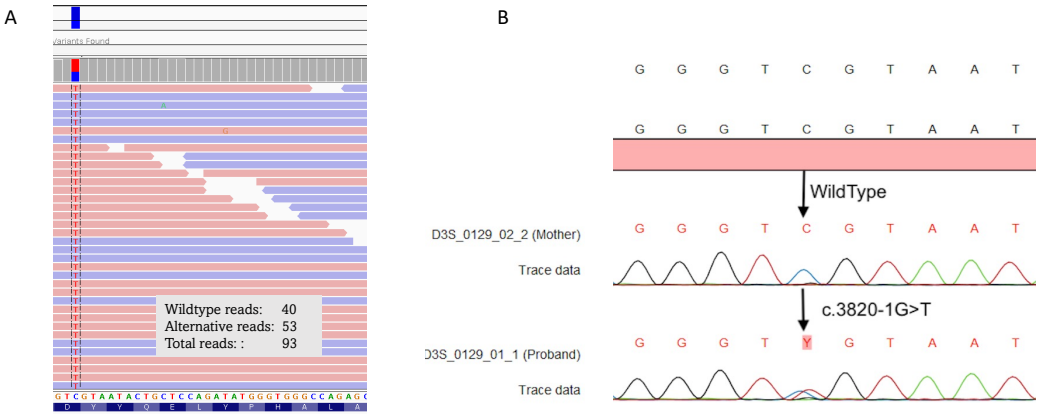

Figure 3.31. Validation of *MAPK1* variant in patient D3S.0129. (A) A heterozygous *MAPK1* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 93 reads covers the region and 53 of this support the alternative allele (VAF 57%). (B) Sanger sequencing analysis of the *MAPK1* variant visualized by CLC bio in the proband, and mother, suggests suspected *de novo* inheritance.

- *Molecular diagnosis*

The pathogenic variant has been previously reported in one other patient diagnosed with Noon syndrome (Motta et al., 2020). Our patient did not present with the classic noon syndrome craniofacial dysmorphisms such as short neck and chin, and cardiovascular abnormalities which are reported in 20%-50% of individuals (Pierpont, 2022). The patient presented with Noonan-like features, which include short statue, ID, and abnormalities of the hands. Our patient also presented seizures, which have been documented at a higher frequency in Noonan-like syndromes compared to classic Noonan syndrome (Davico et al., 2022).

MAPK1 is known to result in Noonan-like syndrome with an evolving phenotype. In a study of 26 patients diagnosed with Noonan syndrome in South Africa, a patient who was initially diagnosed with Noonan syndrome had their diagnosis revised to Costello syndrome, due to a progression of clinical features (Tekendo-Ngongang et al., 2019).

The disorder has autosomal dominant mode of inheritance with about 70% of variants showing a *de novo* inheritance pattern. At the time of recruitment, the father was not available, thus the mode of inheritance could not be definitively determined, but we suspect that the variant shows de novo pattern of inheritance. MAPK1 associated Noonan-like syndrome shows an evolving clinical phenotype, thus, it is recommended that patients be monitored for evidence of

progression in neurologic manifestation, such as neck pain, changes in tone, dizziness, or obstructive sleep apnea. Moreover, patients need to be monitored for the development of cardiovascular complications (Pierpont, 2022).

3.12.18. D3S.0015 - *NFIX* variant associated with Malan syndrome

- *Clinical manifestation*

A five year old male patient (D3S.0015) was recruited based on DDD-Africa recruitment criteria inclusion D (Mild to moderate intellectual disability and dysmorphic features). He was born to phenotypically normal parents at 40 weeks of gestation. Clinical assessment showed delayed speech, thoracic kypho-scoliosis, joint hyperextensibility, dysmorphic features and cryptorchidism. Diagnostic tests and scans performed prior to recruitment included Karyotype, MLPA, Fragile X. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.26.

Table 3.24.Clinical phenotype for patient D3S.0015 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0001555	Asymmetry of the thorax	Abnormality of the musculoskeletal system
HP:0000193	Bifid uvula	Abnormality of head or neck
HP:0000028	Cryptorchidism	Abnormality of the genitourinary system
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0031770	Epicanthus palpebralis	Abnormality of head or neck
HP:0005639	Hyperextensible hand joints	Abnormality of the musculoskeletal system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0000050	Hypoplastic male external genitalia	Abnormality of the genitourinary system
HP:0001382	Joint hypermobility	Abnormality of the musculoskeletal system
HP:0100807	Long fingers	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0030799	Scaphocephaly	Abnormality of the musculoskeletal system
HP:0000200	Short lingual frenulum	Abnormality of head or neck
HP:0045075	Sparse eyebrow	Abnormality of the integument
HP:0001182	Tapered finger	Abnormality of the musculoskeletal system
HP:0005659	Thoracic kyphoscoliosis	Abnormality of the musculoskeletal system

- *Molecular evaluation*

A frameshift variant in *NFIX*, c.293del p. (Lys106ArgfsTer37) was identified in the patient (Figure 3.32). NFIX is also a transcriptional regulator, differentially expressed in brain tissue. Animal model studies suggest that the absence or down-regulation of NFIX results in expansion enlargement of the postnatal mouse brain (Campbell et al., 2008; Harris et al., 2016). As with *NSD1*, the direct relationships between NFIX function and the developing cerebral cortex, are not fully understood. The variant was classified using standard ACMG-AMP guidelines and was classified as pathogenic as evidenced in table 3.5.

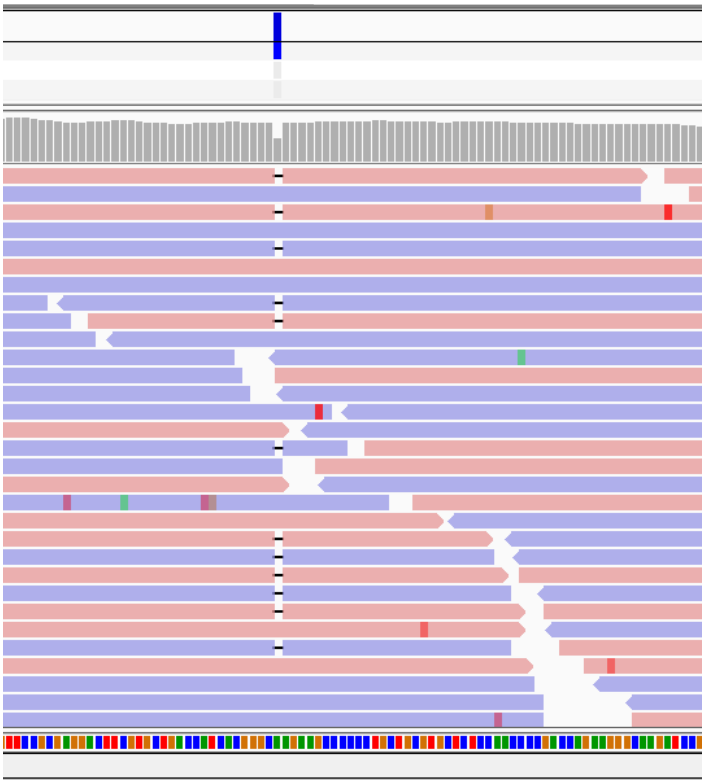

Figure 3.32. Validation of *NFIX* variant in patient D3S.0015. (A) A heterozygous 1 base pair insertion in *NFIX* visualized by IGV in the proband. Sanger confirmation not available.

- *Molecular diagnosis*

Causative variants in the gene can result in two distinct syndromes, Malan syndrome and Marshall-Smith syndrome depending on their exon location (Martinez et al., 2015; Yoneda et al., 2012) (# 614753). Malan syndrome causative variants exclusively cluster in the DNA binding and dimerization domain of exon 2, as in our patient, while variants causative in Marshall-Smith syndrome cluster in exon 6-10 of the gene (Priolo et al., 2018; Rai et al., 2018). Our patient presented with characteristic features of Malan syndrome correlating to the identified causative variant in exon 2. Prominent clinical features included, a prominent

forehead, thoracic kypho-scoliosis, joint hyperextensibility, cryptorchidism, in addition to macrocephaly, and mild global developmental delay.

Recommendations for clinical evaluation and management of Malan syndrome have been outlined by a multidisciplinary team from the Ospedale Pediatrico Bambino Gesù (Macchiaiolo et al., 2022). A high recurrence of long bone fractures has been noted in patients with NFIX mutations, thus the committee recommends monitoring and dosage with Vitamin D during puberty in patients. They also recommend that caregivers be made aware of the emergence of subtle neurological signs such as fainting, seizures, vomiting in patients (Macchiaiolo et al., 2022).

Most reports of NFIX syndromes including Malan syndrome have *de novo* inheritance pattern, but mosaicism has been reported in a few cases (Nimmakayalu et al., 2013; Priolo et al., 2018; Yoneda et al., 2012). Thus, we recommend germline mosaicism testing in the mother.

3.12.19. D3S.0033 - *UBE2A positive diagnosis of UBE2A ID with Nascimento*

- *Clinical manifestation*

Two male siblings (D3S.0033) were recruited at the age of 9 years based on DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound)). The pregnancy and birth were unremarkable. The recruited mother was phenotypically normal. Seizures were reported from 1 years and persisted until the age of 4-5 years. Clinical examination of both patients revealed shared clinical features including, moderate-severe DD, Upslanted palpebral fissure, generalised hypotonia, hypertelorism. The full list of phenotypic anomalies and associated HPO codes is summarized in Table 3.27.

Table 3.25. Clinical phenotype for patient D3S.0033.01.1 and D3S.0033.04.1 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0031770	Epicanthus palpebralis	Abnormality of head or neck
HP:0007429	Few cafe-au-lait spots	Abnormality of the integument
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0000303	Mandibular prognathia	Abnormality of the musculoskeletal system
HP:0000252	Microcephaly	Abnormality of the musculoskeletal system
HP:0008551	Microtia	Abnormality of the ear
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0001250	Seizure	Abnormality of the nervous system
HP:0012745	Short palpebral fissure	Abnormality of head or neck

HP:0045075	Sparse eyebrow	Abnormality of the integument
HP:0001537	Umbilical hernia	Abnormality of the musculoskeletal system
D3S.0033.04.1		
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0031770	Epicanthus palpebralis	Abnormality of the integument
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0000316	Hypertelorism	Abnormality of the musculoskeletal system
HP:0008551	Microtia	Abnormality of the ear
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0001250	Seizure	Abnormality of the nervous system
HP:0045075	Sparse eyebrow	Abnormality of the integument
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck
HP:0001629	Ventricular septal defect	Abnormality of the cardiovascular system
HP:0000154	Wide mouth	Abnormality of head or neck
HP:0012810	Wide nasal base	Abnormality of head or neck

- *Molecular evaluation*

A frameshift variant in *UBE2A*, c.129dup (p. Glu44Ter) was identified in both male siblings. The gene encodes a UBE2A which is critical in the degradation of proteins and DNA repair. Sanger sequencing confirmed the hemizygous variant in both patient and the unaffected mother. The father was unavailable at the time of study recruitment (Figure 3.33). This variant was classified using standard ACMG-AMP guidelines and was classified as pathogenic (see table 3.5).

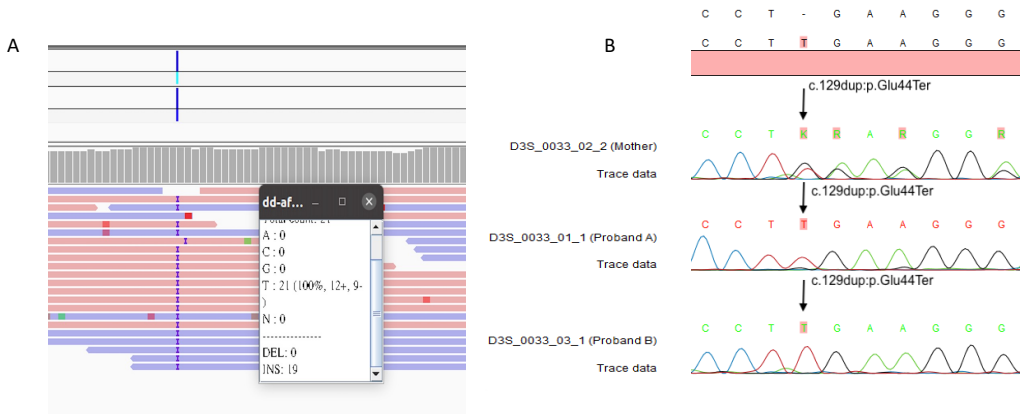

Figure 3.33. Validation of *UBE2A* variant in patient D3S.0033. (A) A homozygous 1- base pair insertion in *UBE2A* visualized by IGV (reference base G in orange, variant base A in green) in the proband. The insertion is supported by 19 reads (VAF 90.1%). (B) Sanger sequencing analysis of the *UBE2A* variant visualized by CLC bio in the proband A, proband B, and mother confirms the variant is maternally inherited.

- *Molecular diagnosis*

Pathogenic Mutations in *UBE2A* have been associated with X-linked ID with Nascimennto. Individuals with the disorder, present with moderate-severe ID, speech impairment, dysmorphic facial features, genital anomalies and skin abnormalities (Tsurusaki et al., 2017).. Both siblings presented with microcephaly, hypertelorism, microtia, epicanthus palpebral, hypotonia, and dysmorphic features which included, upslanted palpebral fissures, sparse eyebrow. The younger brother also had delayed speech, with limited vocabulary and writing at three years old (Table 3.27).

The mutation was confirmed to be maternally inherited. In all cases with a confirmed molecular diagnosis, skewed X-inactivation was observed in the mother. We therefore recommend X-chromosome inactivation assay testing to be performed to confirm this in the mother.(Tsurusaki et al., 2017).

3.12.20. D3S.0034 - *CHD7* variant in CHARGE syndrome

- *Clinical manifestation*

A male patient was recruited at the age of five years old based on DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound) and B. He was delivered to phenotypically normal parents. The pregnancy and birth were unremarkable. Clinical examination of the patient revealed moderate ID, a repaired cardiac defect (large patent ductus arteriosus and coarctation of the aorta), profound sensorineural hearing loss, dysmorphic features and global developmental delay. The full list of clinical features, the affected system anomalies and associated HPO codes is summarized in Table 3.28

Table 3.26. Clinical phenotype for patient D3S.0034 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000154	Wide mouth	Abnormality of head or neck
HP:0008619	Bilateral sensorineural hearing impairment	Abnormality of the ear
HP:0001680	Coarctation of aorta	Abnormality of the cardiovascular system
HP:0000028	Cryptorchidism	Abnormality of the genitourinary system
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000494	Downslanted palpebral fissures	Abnormality of head or neck
HP:0012368	Flat face	Abnormality of head or neck
HP:0002007	Frontal bossing	Abnormality of the musculoskeletal system

HP:0000327	Hypoplasia of the maxilla	Abnormality of the musculoskeletal system
HP:0001382	Joint hypermobility	Abnormality of the musculoskeletal system
HP:0000637	Long palpebral fissure	Abnormality of head or neck
HP:0007340	Lower limb muscle weakness	Abnormality of the musculoskeletal system
HP:0000347	Micrognathia	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0001643	Patent ductus arteriosus	Abnormality of the cardiovascular system
HP:0001212	Prominent fingertip pads	Abnormality of the musculoskeletal system
HP:0002000	Short columella	Abnormality of head or neck
HP:0000046	Small scrotum	Abnormality of the genitourinary system
HP:0000486	Strabismus	Abnormality of the eye
HP:0000430	Underdeveloped nasal alae	Abnormality of head or neck
HP:0009891	Underdeveloped supraorbital ridges	Abnormality of head or neck

- *Molecular evaluation*

Exome sequencing trio analysis identified a *de novo* missense variant (c.8527del) in the *CHD7* (605983), resulting in a frameshift (p. Gly2843GlufsTer46) (Table 3.5, Figure 3.34). The gene encodes a SHF2-like ATPase which is involved in cell cycle and development and is expressed predominantly in the nucleolus and nucleoplasm (Woodage et al., 1997). The *CHD7* variant was validated via Sanger sequencing of proband and both parental samples and observed only in the proband sample (Figure 3.34). The variant was classified as pathogenic in accordance with the ACMG-AMP guidelines as shown in table 3.5.

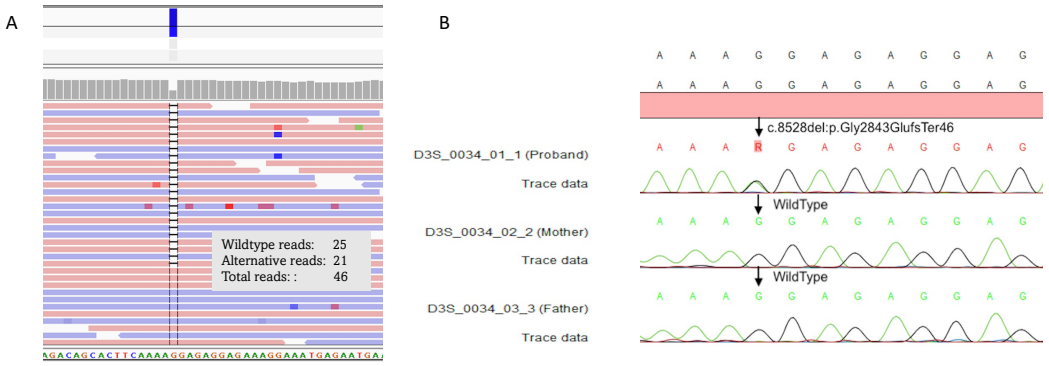

Figure 3.34. Validation of *CHD7* variant in patient D3S.0034. (A) A heterozygous 1-base pair deletion in *CHD7* variant visualized by IGV in the proband. A total of 46 reads covers the region with 21 reads supporting the alternative allele (VAF 45.7%). A reference base was detected at this position for the mother (middle panel) and from the father's sample (third panel). B) Sanger sequencing analysis of the *CHD7* variant visualized by CLC bio in the proband, mother, and father confirms *de novo* inheritance.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with diagnosis of CHARGE syndrome (# 616362) (Janssen et al., 2012), with variants overlapping *CHD7* found causative in approximately 90–95% of CHARGE cases (Qin et al., 2020). This is a well-defined syndrome with a distinctive diagnostic criterion (Blake et al., 1998; Verloes, 2005). The disorder is characterized clinically by ocular coloboma, cranial nerve defects, cardiovascular malformations, atresia of choanae, growth retardation, genitourinary defects, and ear abnormalities (Verloes, 2005). The disorder is phenotypically heterogenous with minor clinical features documented including orofacial defects, heart or esophagus malformations, hypothalamic-pituitary deficiency, and intellectual disability documented in some of the cases (Jongmans et al., 2006). Patients presenting with three major clinical features are said to have typical CHARGE syndrome, while those presenting with multiple minor clinical features in addition to either one or two of the major clinical features are considered to have “atypical” CHARGE presentation (Verloes, 2005).

Our patient presented with severe Developmental delay, bilateral colobomas at optic nerve with visual loss, abnormal movements, hyperactivity, and microcephaly, profound bilateral sensorineural hearing loss fitting a CHARGE syndrome diagnosis. Mutations overlapping *CHD7* gene have been identified as causative in approximately 90–95% of CHARGE cases (Qin et al., 2020).

3.12.21. D3S.0035 - *OCRL* variant in Lowe syndrome

- *Clinical manifestation*

A male patient (D3S.0035) was recruited at the age of 13 years based on DDD-Africa recruitment criteria inclusion B1 (Multiple major malformations in 2 or more different organ systems >1month). Pre-eclampsia during pregnancy was reported. Both recruited parents were phenotypically normal. The patient had a half-brother who was deceased due to hydrocephalus. Clinical history revealed neonatal jaundice which was treated with phototherapy. Clinical examination revealed short stature (< -3SD), congenital cataract and a range of eye abnormalities, mild ID, and hyperactive behaviour which was being treated using Risperdal. Additionally, the patient attends physiotherapy, speech therapy, and endocrine clinics (Table 3.29)

Table 3.27. Clinical phenotype for patient D3S.0035 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0001156	Brachydactyly	Abnormality of the musculoskeletal system
HP:0004209	Clinodactyly of the 5th finger	Abnormality of the musculoskeletal system
HP:0000028	Cryptorchidism	Abnormality of the genitourinary system
HP:0000490	Deeply set eye	Abnormality of the eye
HP:0000689	Dental malocclusion	Abnormality of head or neck
HP:0000519	Developmental cataract	Abnormality of the eye
HP:0000494	Downslanted palpebral fissures	Abnormality of head or neck
HP:0000666	Horizontal nystagmus	Abnormality of the eye
HP:0000752	Hyperactivity	Abnormality of the nervous system
HP:0001053	Hypopigmented skin patches	Abnormality of the integument
HP:0100876	Infra-orbital crease	Abnormality of head or neck
HP:0001382	Joint hypermobility	Abnormality of the musculoskeletal system
HP:0000369	Low-set ears	Abnormality of the ear
HP:0000272	Malar flattening	Abnormality of the musculoskeletal system
HP:0011342	Mild global developmental delay	Abnormality of the nervous system
HP:0000396	Overfolded helix	Abnormality of the ear
HP:0004482	Relative macrocephaly	Abnormality of the musculoskeletal system
HP:0009237	Short 5th finger	Abnormality of the musculoskeletal system
HP:0004322	Short stature	Growth abnormality
HP:0000046	Small scrotum	Abnormality of the genitourinary system
HP:0000319	Smooth philtrum	Abnormality of head or neck
HP:0005338	Sparse lateral eyebrow	Abnormality of the integument
HP:0000486	Strabismus	Abnormality of the eye

- *Molecular evaluation*

Exome trio analysis identified a missense variant in *OCRL* (615140) c.2399C>T, causing an amino acid change from proline to leucine (p. Pro799Leu). The gene encodes an inositol 5-phosphatase that dephosphorylates PI (4,5) P2 to PI4P, involved in lysozyme positioning important for nutrient and growth factor cycling, among its many functions (Pirruccello et al., 2014). Sanger sequencing results were not available due to failed primer optimisation (Figure 3.35). This variant was classified using standard ACMG-AMP guidelines as pathogenic.

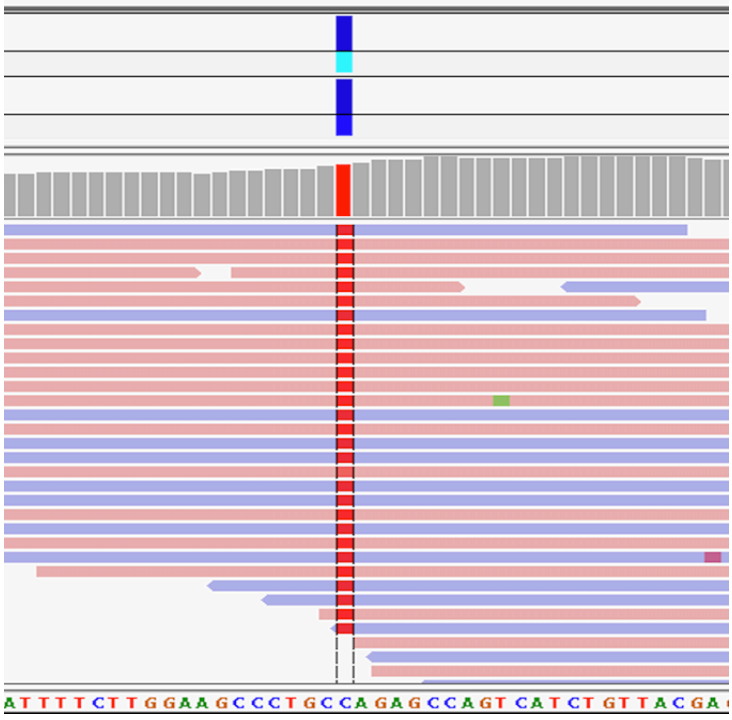

Figure 3.35. Validation of *OCRL* variant in patient D3S.0035. A homozygous *OCRL* variant was visualized by IGV (variant base T in red) in the proband. A total of 29 reads covered the region and all supported the alternative allele (VAF 100%). Sanger sequencing results were not available due to failed primer optimization.

- *Molecular diagnosis*

Mutations in the gene result in an inability for the lysosomes to correctly respond to nutrient and growth factor signalling. This is associated with a failure to thrive presenting as intellectual disability and growth retardation, eye abnormalities, behavioural abnormalities, and renal proximal tubular (Bokenkamp & Ludwig, 2016) as core clinical features in all patients with OCRL associated X-linked recessive Lowe syndrome (309000).

Our patient's clinical features were considered adequate for Lowe syndrome, however, unlike previous reports of the syndrome, at the time of the clinical assessment no renal abnormalities were reported. We therefore suggest renal testing in this patient for signs of kidney disease also referred to as Fanconi syndrome. Kidney disease has been reported to worsen with age in patients eventually leading to chronic renal disease. It has been reported that death in most patients is due to renal failure with the largest surviving patient being a 54-year-old male (Ingemi et al., 2012).

3.12.22. D3S.0047 - *ARID1B* variant associated with Coffin Siri syndrome (CSS)

- *Clinical manifestation*

A female patient (D3S.0047) was recruited at the age of 10 years old based on DDD-Africa recruitment criteria inclusion A (Unexplained DD with moderate to profound intellectual disability (ID)) and C (One major malformation and three or more well documented minor anomalies). The patient was delivered to phenotypically normal parents by caesarean at 40 weeks of gestation due to mentum posterior face presentation but was otherwise unremarkable. Clinical history revealed cardiac lesions which were treated using anti-failure medication (specific medication not provided). Seizures were noted from two years old and the patient was placed on Epilim treatment which was discontinued at 3 years old. A hearing test and cardiac test were performed which revealed mild conductive hearing loss and pericardial effusion respectively. Clinical examination revealed a range of facial anomalies, growth abnormalities, and hypotonia. The full list of dysmorphic features associated with HPO codes and HPO system anomalies is presented in Table 3.30.

Table 3.28. Clinical phenotype for patient D3S.0035 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000366	Abnormality of the nose	Abnormality of head or neck
HP:0008386	Aplasia/Hypoplasia of the nails	Abnormality of head or neck
HP:0000670	Carious teeth	Abnormality of head or neck
HP:0000405	Conductive hearing impairment	Abnormality of the ear
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0000365	Hearing impairment	Abnormality of the ear
HP:0000218	High palate	Abnormality of head or neck
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0009719	Hypomelanotic macule	Abnormality of head or neck
HP:0000821	Hypothyroidism	Abnormality of the endocrine system
HP:0001388	Joint laxity	Abnormality of the musculoskeletal system
HP:0000276	Long face	Abnormality of head or neck
HP:0000637	Long palpebral fissure	Abnormality of head or neck
HP:0000294	Low anterior hairline	Abnormality of head or neck
HP:0000369	Low-set ears	Abnormality of the ear
HP:0000272	Malar flattening	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0001698	Pericardial effusion	Abnormality of the cardiovascular system
HP:0001763	Pes planus	Abnormality of limbs
HP:0001357	Plagiocephaly	Abnormality of the musculoskeletal system
HP:0001212	Prominent fingertip pads	Abnormality of the musculoskeletal system
HP:0000426	Prominent nasal bridge	Abnormality of head or neck
HP:0005274	Prominent nasal tip	Abnormality of head or neck
HP:0010808	Protruding tongue	Abnormality of head or neck
HP:0004482	Relative macrocephaly	Abnormality of the musculoskeletal system
HP:0000470	Short neck	Abnormality of head or neck
HP:0004322	Short stature	Growth abnormality
HP:0000574	Thick eyebrow	Abnormality of head or neck
HP:0000179	Thick lower lip vermilion	Abnormality of head or neck
HP:0000430	Underdeveloped nasal alae	Abnormality of head or neck
HP:0001629	Ventricular septal defect	Abnormality of the cardiovascular system
HP:0006610	Wide intermammillary distance	Abnormality of the breast

- *Molecular evaluation*

Exome trio analysis revealed a frameshift variant in *ARID1B* (614556), (c.2819del) causing an amino acid (p. Pro927GlnfsTer57) change was detected in the patient. The gene overlaps the BAF complex which is involved in chromatin remodelling for gene expression and differentiation (Gauri deak 2020). The variant was confirmed by Sanger sequencing and segregation analysis of the proband and parental samples confirming *de novo* mode of inheritance (Figure.3.36). This variant was classified using standard ACMG-AMP guidelines as pathogenic as evidenced in Table 3.5.

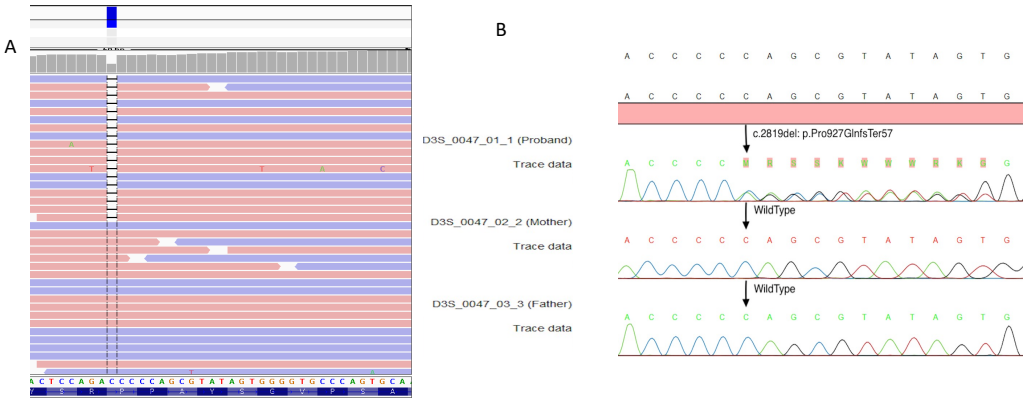

Figure 3.36. Validation of *ARID1B* variant in patient D3S.0047. (A) A heterozygous 1- base pair deletion in *ARID1B* visualized by IGV in the proband (first panel. A total of 18 reads supports the alternative allele (VAF). (B) Sanger sequencing analysis of the *ARID1B* variant visualized by CLC bio in the proband, mother, and father confirms the variant is *de novo* inheritance.

- *Molecular diagnosis*

Pathogenic mutations in AR1D1B have been associated with non-syndromic ID and Coffin Siri syndrome (CSS). It has been reported that only frameshifts, nonsense mutations and deletions in *ARID1B* are causal in CSS. When pathogenic missense variants have been reported, they are shown to result in a mild form of the disease, with patients presenting with short stature and no other DD or ID.

Coffin-Siris syndrome is characterized by varying degrees of DD, distinct facial features, hypotonia, aplasia or hypoplasia of the distal phalanx or nail of the fifth and additional digits, and sparse scalp hair (Iwamoto et al., 2021). Our patient's clinical features were considered adequate for CSS, as he presented with the key features of the syndrome which include, seizures, short statue, hypotonia, bilateral hearing loss, and distinct facial abnormalities, our patient did not present with hypertrichosis, which has been noted in about 94% of CSS patients.

Interestingly, she presented with hyperthyroidism which, following a search of CSS reported phenotypes on DECIPHER online database, has not been reported in other patients.

ARID1B gene has been associated with the development of a range of cancers with most of the causative cancer variants being somatic. However, there is one reported case of a causative germline variant being associated with the development of thyroid cancer (van der Sluijs et al., 2019). Given that our patient already presents with hyperthyroidism, the risk of thyroid cancer development is increased in this patient. As such we recommend monitoring and testing for cancer in this patient.

3.12.23. D3S.0050 - *CHD7* variant in CHARGE syndrome

- *Clinical manifestation*

A female patient was recruited at the age of five years old based on DDD-Africa recruitment criteria A (Intellectual disability (moderate to profound) B (Multiple major malformations in two or more different organs) and C (One major malformation and three or more well documented minor anomalies). She was delivered to phenotypically normal parents. The mother was on ARVs during pregnancy but otherwise unremarkable. Clinical examination of the patient revealed severe developmental delay, Bilateral colobomas at optic nerve with visual loss, abnormal movements, hyperactivity, pigmentation anomalies, microcephaly, profound bilateral sensory-neural hearing loss. The full list of clinical features, the affected system anomalies and associated HPO codes is summarized in Table 3.31

Table 3.29. Clinical phenotype for patient D3S.0050 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0100022	Abnormality of movement	Abnormality of the nervous system
HP:0007018	Attention deficit hyperactivity disorder	Abnormality of the nervous system
HP:0008619	Bilateral sensorineural hearing impairment	Abnormality of the ear
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0003191	Cleft ala nasi	Abnormality of head or neck
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0001034	Hypomelanotic macule	Abnormality of the integument
HP:0000252	Microcephaly	Abnormality of the musculoskeletal system
HP:0000588	Optic disc coloboma	Abnormality of the eye

HP:0001315	Reduced tendon reflexes	Abnormality of the nervous system
HP:0000486	Strabismus	Abnormality of the eye
HP:0000572	Visual loss	Abnormality of the eye

- *Molecular evaluation*

Exome sequencing trio analysis identified a *de novo* missense variant (c.8527del) in the *CHD7* gene, resulting in a frameshift (p. Gly2843GlufsTer46) (Table 3.5, Figure 3.37). The variant was validated via Sanger sequencing of the proband and both parental samples and observed only in the proband sample (Figure 3.37). The variant was classified as pathogenic in accordance with the ACMG-AMP guidelines, as shown in Table 3.5.

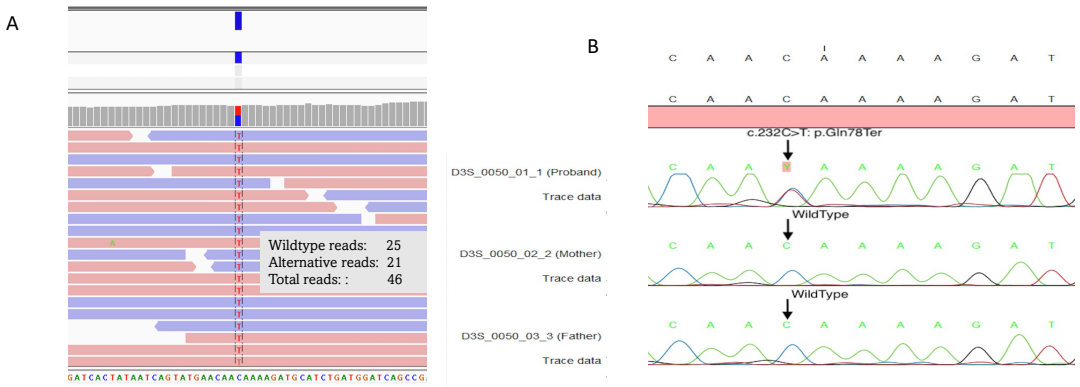

Figure 3.37. Validation of *CHD7* variant in patient D3S.0050. (A) A heterozygous *CHD7* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of 46 reads covers the region with 21 reads supporting reads the alternative allele (45.7%). B) Sanger sequencing analysis of the *CHD7* variant visualized by CLC bio in the proband, mother, and father confirms the variant is *de novo* inherited.

- *Molecular diagnosis*

Pathogenic variants in the gene are associated with diagnosis of CHARGE syndrome (#616362) (Janssen et al., 2012).

3.12.24. D3S.0058 - *MFSD8* variant associated with CLN7 disease

- *Clinical manifestation*

A six year old female patient (D3S.0058) was recruited based on DDD-Africa recruitment criteria inclusion A (Unexplained DD with moderate to profound intellectual disability (ID)). The patient was the only child of a phenotypically normal mother and a father with non-syndromic postaxial polydactyly at 40 weeks of gestation. One previous miscarriage was also noted at the first trimester and four years of infertility were noted in the mother. Clinical history revealed a hospital admission due to focal seizures following birth. Diagnostic testing and

scanning performed prior to recruitment included a metabolic test which showed elevated levels of ammonia and pyruvate. The results of MR images highlighted hypomyelination of the white matter. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO terms and codes is summarized in Table 3.32.

Table 3.30. Clinical phenotype for patient D3S.0058 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0001320	Cerebellar vermis hypoplasia	Abnormality of the nervous system
HP:0005750	Contractures of the joints of the lower limbs	Abnormality of the musculoskeletal system
HP:0100360	Contractures of the joints of the upper limbs	Abnormality of the musculoskeletal system
HP:0007665	Curly eyelashes	Abnormality of the integument
HP:0002376	Developmental regression	Abnormality of the nervous system
HP:0001332	Dystonia	Abnormality of the nervous system
HP:0010816	Epidermal nevus	Abnormality of the integument
HP:0001347	Hyperreflexia	Abnormality of the nervous system
HP:0100729	Large face	Abnormality of head or neck
HP:0002509	Limb hypertonia	Abnormality of the musculoskeletal system
HP:0032794	Myoclonic seizure	Abnormality of the nervous system
HP:0003764	Nevus	Abnormality of the integument
HP:0011231	Prominent eyelashes	Abnormality of the integument
HP:0011344	Severe global developmental delay	Abnormality of the nervous system

- *Molecular evaluation*

A splice donor variant in *MFSD8*, (c.1350+1del) causing a frameshift was detected in the patient. The gene encodes a 518aa MFSD8 protein kinase that is involved in the regulation of autophagy for neuronal development. The protein is expressed predominantly in brain and eye tissue Sanger sequencing confirmed the presence of the variant in a homozygous state in the patient (Figure 3.38). Segregation analysis performed demonstrated the variant was inherited from both unaffected parents who were heterozygous carriers. This variant was classified as ACMG-AMP pathogenic as evidenced in Table 3.5.

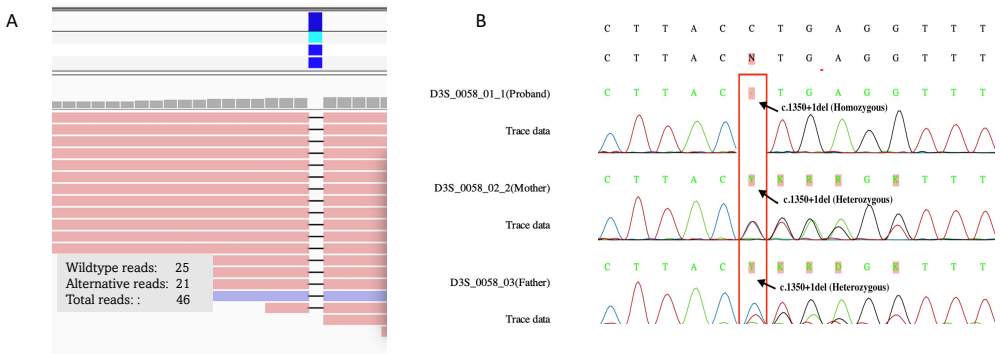

Figure 3.38. Validation of *MFSD8* variant in patient D3S.0058. (A) A homozygous 1 base pair deletion in *MFSD8* visualized by IGV in the proband. A total of 46 reads covers the regions with 21 reads supporting the alternative (VAF 45.6%). B) Sanger sequencing analysis of the *MFSD8* variant visualized by CLC bio in the proband, mother, and father confirms the variant is inherited from both parents.

- *Molecular diagnosis*

Mutations in the gene have been associated neuronal ceroid lipofuscinoses (CLN7) (#610951) caused mostly by loss-of-function mutations. The disorder is characterized by progressive motor and mental regression. Most individuals present with abnormal corpus callosum, ventriculomegaly, hypotonia, microcephaly, and abnormal brain structures. Our patient's clinical phenotype was a fit for the disorder, presenting with dystonic posturing with continuous tremors, myoclonic seizures, hypomyelination with cerebellar atrophy, and a developmental regression (Table 3.32). Similar clinical features were reported in a Chinese male with ceroid lipofuscinoses 7 (CLN7) by Yang Gu et al (2022).

Promising treatment options that are aimed at preventing any further regression in cognitive function and motor function are in both research and testing stages (Brudvig & Weimer, 2022). One individual with CLN7 disease has been treated with an antisense oligonucleotide (Morimura et al., 2014).

3.12.25. D3S.0081 - *SCN2A* related DD without epilepsy

- *Clinical manifestation*

An eight-year-old male patient (D3S.0081) was recruited based on DDD-Africa recruitment criteria inclusion A (Unexplained DD with moderate to profound intellectual disability (ID)). The patient was born via elective caesarean at 40 weeks of gestation. The mother was treated for hypertension using methyldopa during the pregnancy. Additionally, at least three previous miscarriages were noted. Diagnostic tests and scans performed prior to recruitment included Karyotype, MLPA Array-CGH, and Fragile X syndrome testing which were all normal. Additionally, renal testing and brain MR imaging results showed no abnormalities. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.33.

Table 3.31. Clinical phenotype for patient D3S.0081 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0002648	Abnormality of calvarial morphology	Abnormality of the musculoskeletal system
HP:0001488	Bilateral ptosis	Abnormality of the eye
HP:0001348	Brisk reflexes	Abnormality of the nervous system
HP:0003763	Bruxism	Abnormality of the nervous system
HP:0000414	Bulbous nose	Abnormality of head or neck
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0000437	Depressed nasal tip	Abnormality of head or neck
HP:0000494	Downslanted palpebral fissures	Abnormality of head or neck
HP:0000286	Epicanthus	Abnormality of head or neck
HP:0001508	Failure to thrive	Growth abnormality
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0005072	Hyperextensibility at wrists	Abnormality of the musculoskeletal system
HP:0010500	Hyperextensibility of the knee	Abnormality of the musculoskeletal system
HP:0006460	Increased laxity of ankles	Abnormality of the musculoskeletal system
HP:0006149	Increased laxity of fingers	Abnormality of the musculoskeletal system
HP:0000368	Low-set, posteriorly rotated ears	Abnormality of the ear
HP:0000303	Mandibular prognathia	Abnormality of the musculoskeletal system
HP:0000995	Melanocytic nevus	Abnormality of the integument
HP:0011800	Midface retrusion	Abnormality of head or neck
HP:0000733	Motor stereotypy	Abnormality of the nervous system
HP:0000341	Narrow forehead	Abnormality of head or neck
HP:0000980	Pallor	Abnormality of the integument
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0000411	Protruding ear	Abnormality of the ear
HP:0030799	Scaphocephaly	Abnormality of the musculoskeletal system
HP:0001250	Seizure	Abnormality of the nervous system
HP:0011344	Severe global developmental delay	Abnormality of the nervous system
HP:0000470	Short neck	Abnormality of head or neck
HP:0012307	Spatulate ribs	Abnormality of the musculoskeletal system
HP:0000574	Thick eyebrow	Abnormality of the integument
HP:0000179	Thick lower lip vermilion	Abnormality of head or neck
HP:0001537	Umbilical hernia	Abnormality of the musculoskeletal system
HP:0000431	Wide nasal bridge	Abnormality of head or neck

- *Molecular evaluation*

A missense variant in *SCN2A*, (c.727C>T) resulting in a pre-mature stop-gain (p. Gln243Ter) change was detected in the patient. Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.39). Segregation analysis confirmed *de novo* inheritance. This variant was classified as ACMG-AMP pathogenic as evidenced in Table 3.5.

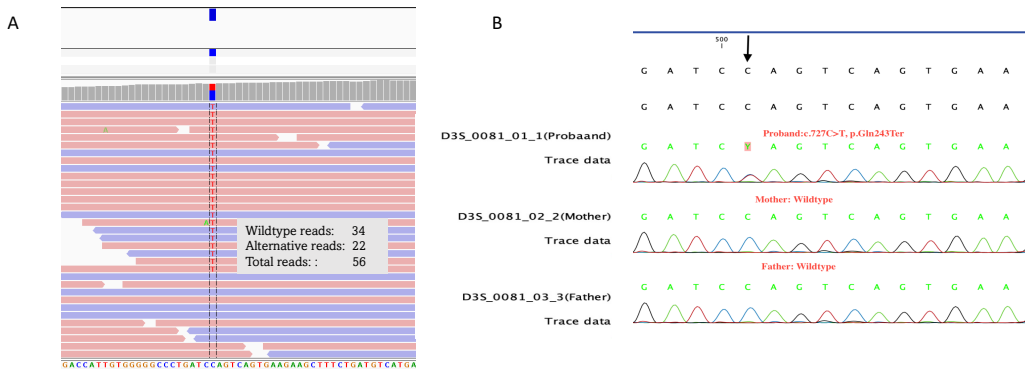

Figure 3.39. Validation of *SCN2A* variant in patient D3S.0081. (A) A heterozygous *SCN2A* variant visualized by IGV (reference base C in blue, variant base T in red) in the proband. A total of A total of 56 reads covers the region with 22 alternative reads supporting the alternative allele (VAF 39.3%) (B) Sanger sequencing analysis of the *SCN2A* variant visualized by CLC bio in the proband, mother, and father confirms the variant is *de novo* inherited.

- *Molecular diagnosis*

Mutations in the gene have been associated with a range of neurodevelopmental disorders, including developmental and epileptic encephalopathy 11 (#613721); episodic ataxia type 9 (#618924); and benign familial infantile seizures 3 (#607745). Pathogenic gain-of-function (GOF) variants in the gene are associated with benign (familial) neonatal/infantile seizures (BFNIS) with onset at < 3 months.

Patients with DEE presenting at > 3 months, have had loss -of-function (LOF) variants or mix of both LOF and GOF variants. Protein truncating, splice site and missense variants resulting in LOF and haploinsufficiency for SCN2A are reported in patients with ID and/or autism spectrum disorder (ASD). The patient (D3S.0081) presented with DD and ID, macrocephaly, ASD, obesity, hyperpigmentation but no history of seizures. The gene is associated with haplosufficiency neurodevelopmental disorder with hypotonia, impaired language and hand movements (OMIM # 613443).

Patients presenting with ID and/or autism without epilepsy has been reported in several studies. A study by Wolff et al (2017) reported a prevalence of 16% (32/201) for patients presenting

with DD/ID but no epilepsy due to a causative SCN2A mutation (Wolff et al.,2017). A diagnosis of SC2NA- related DD was made.

3.12.26. D3S.0089 - *MEF2C* haploinsufficiency syndrome

- *Clinical manifestation*

A nine year old male patient (D3S.0089) was recruited based on DDD-Africa recruitment criteria inclusion A. The patient was delivered to phenotypically normal parents at 40 weeks of gestation. Clinical history revealed a hospital admission due to seizures but currently resolved. Diagnostic testing performed prior to recruitment included Array-CGH, karyotype, MLPA, hearing test and serum and urine metabolic test, all which were normal. Additionally, CT scan revealed small area of T2 fair gradient recalled echo hyperintensity which were deemed of doubtful significance. Family history indicated two paternal relatives with learning difficulties. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.34.

Table 3.32. Clinical phenotype for patient D3S.0089 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000729	Autistic behaviour	Abnormality of the nervous system
HP:0001348	Brisk reflexes	Abnormality of the nervous system
HP:0007321	Deep white matter hypodensities	Abnormality of the nervous system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0002376	Developmental regression	Abnormality of the nervous system
HP:0001034	Hypermelanotic macule	Abnormality of the integument
HP:0001572	Macrodontia	Abnormality of head or neck
HP:0000400	Macrotia	Abnormality of the ear
HP:0000639	Nystagmus	Abnormality of the eye
HP:0012736	Profound global developmental delay	Abnormality of the nervous system
HP:0001250	Seizure	Abnormality of the nervous system
HP:0001264	Spastic diplegia	Abnormality of the nervous system
HP:0008577	Underfolded helix	Abnormality of the ear
HP:0000154	Wide mouth	Abnormality of head or neck

- *Molecular evaluation*

A stop-gain variant in *MEF2C* (c.963T>A) causing a p. Tyr275Phe change resulting in a stop-gain was detected in the patient. The gene encodes the voltage-gated sodium channel Nav1.2 protein, which has been reported to be critical in the initiation and conduction of action

potentials (Wolff et al., 2017). Sanger sequencing confirmed the presence of the variant in a heterozygous state in the patient (Figure 3.40). The father was not available at recruitment, but sequencing of the mother's DNA demonstrated the mother had a reference allele sequence. This variant was classified as ACMG-AMP, pathogenic as evidenced in Table 3.5.

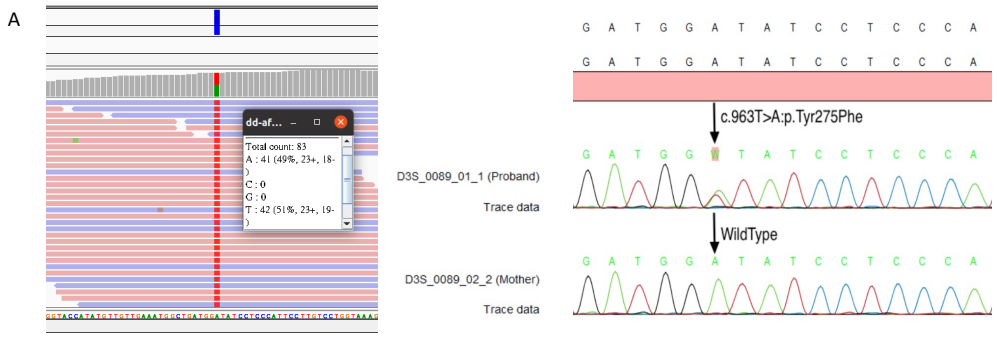

Figure 3.40. Validation of *MEF2C* variant in patient D3S.0089. (A) A hemizygous *MEF2C* variant visualized by IGV (reference base A in green, variant base T in red) in the proband. A total of 83 reads covers the region with 42 supporting the alternative read (VAF 50.6%). (B) Sanger sequencing analysis of the *MEF2C* variant visualized by CLC bio in the proband, and the mother shows that the variant was not inherited from the mother.

- *Molecular diagnosis*

Myocyte enhancer factor 2C (MEF2C) haploinsufficiency syndrome (MCHS). The syndrome presents with a diverse phenotype in affected patients, but some key clinical characteristics include, severe global developmental delay accompanied by absent speech, limited walking, seizures, and stereotypic movements (Vrečar et al., 2017). These were present in our patient as evidenced in the list of HPO terms in Table 3.34.

Low-level mosaicism has been reported in some cases and thus we recommend testing in the patient when at childbearing years.

3.12.27. D3S.0108 - *XYLT2* variant in Spondyl-oocular Syndrome

- *Clinical manifestation*

A six year old female patient (D3S.0108) was recruited based on DDD-Africa recruitment criteria inclusion A (Unexplained DD with moderate to profound intellectual disability (ID)). She was born to phenotypically normal parents at 40 weeks of gestation. The mother was treated for hypertension during the pregnancy. Clinical history revealed the patient had feet and cataracts surgery performed and was on Vitamin D3. Additionally, hospital admissions for spinal problems at 2 weeks were reported with no additional information given. Diagnostic tests

and scans performed prior to recruitment included ECG, haematological and metabolic tests which were all normal. Additionally, liver function tests showed elevated transaminases. Brain MR imaging and CT scan result both supported corpus callosum agenesis. The full list of the patient phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.35

Table 3.33. Clinical phenotype for patient D3S.00108 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0001274	Agenesis of corpus callosum	Abnormality of the nervous system
HP:0001137	Alternating esotropia	Abnormality of the eye
HP:0000518	Cataract	Abnormality of the eye
HP:0030048	Colpocephaly	Abnormality of the nervous system
HP:0007665	Curly eyelashes	Abnormality of the integument
HP:0010862	Delayed fine motor development	Abnormality of the nervous system
HP:0002194	Delayed gross motor development	Abnormality of the nervous system
HP:0000750	Delayed speech and language development	Abnormality of the nervous system
HP:0000286	Epicanthus	Abnormality of head or neck
HP:0000565	Esotropia	Abnormality of the eye
HP:0031846	Femur fracture	Abnormality of the musculoskeletal system
HP:0003961	Fractured forearm bones	Abnormality of the musculoskeletal system
HP:0000666	Horizontal nystagmus	Abnormality of the eye
HP:0002808	Kyphosis	Abnormality of the musculoskeletal system
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0002756	Pathologic fracture	Abnormality of the musculoskeletal system
HP:0000767	Pectus excavatum	Abnormality of the musculoskeletal system
HP:0005487	Prominent metopic ridge	Abnormality of the musculoskeletal system
HP:0004482	Relative macrocephaly	Abnormality of the musculoskeletal system
HP:0004322	Short stature	Growth abnormality
HP:0000486	Strabismus	Abnormality of the eye
HP:0002942	Thoracic kyphosis	Abnormality of the musculoskeletal system
HP:0002953	Vertebral compression fracture	Abnormality of the musculoskeletal system

- *Molecular evaluation*

A frameshift variant in *XYLT2* (# 608125), c.1703-1721del (p. His568ProfsTer33) was detected in the patient. The gene encodes Xylosyltransferases II that which catalyses the assembly of

proteoglycan protein central to cell homeostasis and signalling. Sanger sequencing confirmed the presence of the variant in a homozygous state in the patient (Figure 3.41. Segregation analysis performed demonstrated the variant was inherited from both unaffected parents who were heterozygous carriers. This variant was classified as ACMG-AMP pathogenic. As evidenced in Table 3.5.

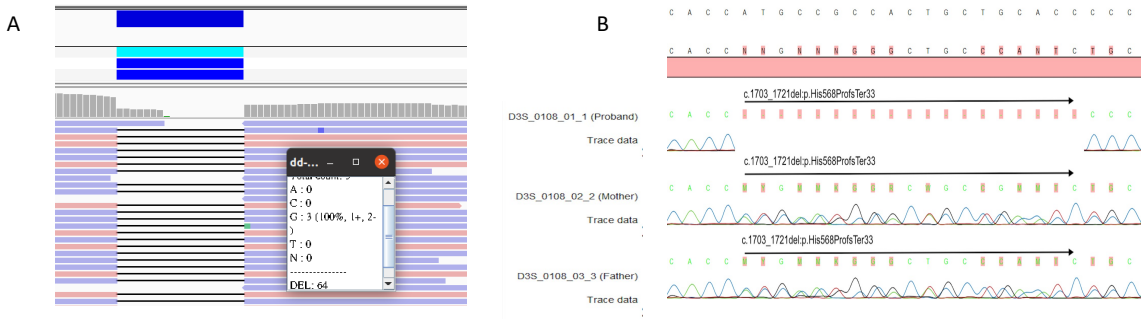

Figure 3.41. Validation of *XYLT2* variant in patient D3S.0108. (A) A homozygous 18 base pair deletion in *XYLT2* visualized by IGV in the proband. A total of 64 reads supports the alternative allele (VAF 100%). B) Sanger sequencing analysis of the *XYLT2* variant visualized by CLC bio in the proband, mother, and father confirms the variant is inherited from both parents.

- *Molecular diagnosis*

Variants in *XYLT2* result in proteoglycan-impaired function in the inner ear, heart muscle retina system, and skeletal. Thus, variants in XYTL2 are associated with autosomal recessive Spondyl-oocular Syndrome (Munns et al., 2015)., which is characterized by eye anomalies, multiple fractures consistent with the clinical features in our patient. Our patient also presented with short stature and kyphosis which have been documented in other patient cohorts.

The syndrome and all previously reported variants are inherited from consanguineous parents. Non-consanguinity was reported in the family but not confirmed. There is a high likelihood that a founder mutation in the African population. Assessment of founder effect in the South African population would require a secondary study for haplotype analysis or SNP genetic marker analysis which is beyond the scope of this current research.

3.12.28. D3S.0117 - *NSD1* report of a second Sotos syndrome patient of African ancestry

- *Clinical manifestation*

A six year old male patient (D3S.0117) was recruited based on DDD-Africa recruitment criteria inclusion D. He was born to phenotypically normal parents via caesarean at 38 weeks of

gestation due to pre-eclampsia. The mother was treated for hypertension and deep vein thrombosis using Klexane and anti-hypertension medication during the pregnancy. Clinical history revealed the patient was treated for neonatal jaundice by three days of phototherapy. Two separate hospital admissions for febrile seizures were reported. Additionally, tongue tie surgery was performed in 2016. Diagnostic tests and scans performed prior to recruitment included Fragile X, MPLA, cardiac echo, hearing test, and renal ultrasound, which were all normal. Additionally, there was notable bone aging. CT scans and MR imaging showed ventriculomegaly and scaphocephaly. The full list of the patient's phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.36.

Table 3.34.Clinical phenotype for patient D3S.0117 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0005616	Accelerated skeletal maturation	Abnormality of the musculoskeletal system
HP:0010296	Ankyloglossia	Abnormality of head or neck
HP:0000957	Cafe-au-lait spot	Abnormality of the integument
HP:0012725	Cutaneous syndactyly	Abnormality of the musculoskeletal system
HP:0005280	Depressed nasal bridge	Abnormality of head or neck
HP:0002373	Febrile seizure (within the age range of 3 months to 6 years	Abnormality of the nervous system
HP:0000316	Hypertelorism	Abnormality of the eye
HP:0020073	Hypopigmented macule	Abnormality of the integument
HP:0000952	Jaundice	Abnormality of the integument
HP:0045086	Knee joint hypermobility	Abnormality of the musculoskeletal system
HP:0000256	Macrocephaly	Abnormality of the musculoskeletal system
HP:0011342	Mild global developmental delay	Abnormality of the nervous system
HP:0000341	Narrow forehead	Abnormality of head or neck
HP:0001763	Pes planus	Abnormality of limbs
HP:0011220	Prominent forehead	Abnormality of head or neck
HP:0008070	Sparse hair	Abnormality of the integument
HP:0000098	Tall stature	Growth abnormality
HP:0002943	Thoracic scoliosis	Abnormality of the musculoskeletal system
HP:0002119	Ventriculomegaly	Abnormality of the nervous system

- *Molecular evaluation*

A missense variant in *NSD1* (# 606681) gene, c.4913A>G (p. His1638Arg) was identified following WES data analysis. The presence of the variant was confirmed using secondary results from a gene panel study within our department. The gene encodes a histone–lysine N-methyltransferase (H3K36) that can act as both a co-repressor and a co-activator transcriptional

factor during post-translational modifications of the epigenome (Karina Lucio-Eterovic et al., 2010). This variant was classified as ACMG-AMP pathogenic as evidenced in Table 3.5.

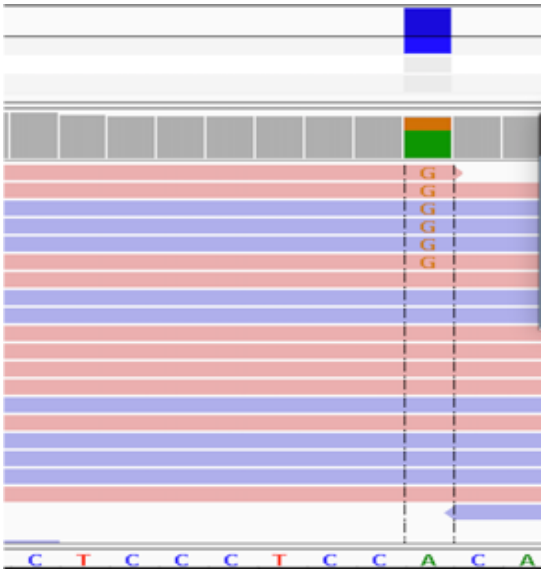

Figure 3.42. Validation of *NSD1* variant in patient D3S.0117. (A) A heterozygous *NSD1* variant visualized by IGV (reference base A in green, variant base G in orange) in the proband. A total of 19 reads covers the region with 6 supporting the alternative allele (VAF 31.2%). The variant was confirmed using a second sequencing method by a gene panel project in the department. Data not shown.

- *Molecular diagnosis*

NSD1 is expressed in various human fetal and adult tissues, including the brain and skeletal muscle (Kurotaki et al., 2001). However, the biological function and contribution of the gene in the developing cerebral cortex, is not fully understood. Missense variant and truncating mutations have been shown to occur within the functional domain of the gene and resulting in the reduced levels of the NSD1 enzyme (Kurotaki et al., 2001). Reduced enzyme production has been linked with overproduction of neurons in the cortex, and skeletal leading to an overgrowth of the brain, tall statue and ID which are characteristic of Sotos syndrome (Kurotaki et al., 2002; Pavone et al., 2017).

There is a large phenotypic overlap between the syndrome and Soto-like syndromes caused mainly by mutations in *NFIX* (Malan syndrome), *SETD2* and *DNMT3A* genes (Lu et al., 2017). However, patients with Sotos syndrome present with distinct clinical features that distinguish it from other overgrowth syndromes. Our patient presented with characteristic features of Sotos syndrome including a prominent forehead, sparse hair, thoracic scoliosis, joint hypermobility, in addition to tall stature, macrocephaly, and mild global developmental delay.

There is only one other African patient that has been reported with Sotos syndrome in the literature (Mubungu et al., 2020). The patient was from Central Africa and was noted to present with the core Sotos syndrome feature but also additional ethic-specific features, prominent lips, and deep philtrum which the authors suggest are common features in African patients (Mubungu et al., 2020). However, we did not observe any of these African-specific features in our South African patients. There are no current clinical management recommendations for Sotos syndrome. However, thoracic scoliosis in our patient could be treated by either braces or surgical correction (Corrado et al., 2011).

3.12.29. D3S.0119 - *RPS6KA3* variant confirms the suspected clinical diagnosis of Coffin-Lowry

- *Clinical manifestation*

A 14 year old male patient (D3S.0119) was recruited based on DDD-Africa recruitment criteria inclusion A (Unexplained DD with moderate to profound intellectual disability (ID)) and D (Suspected multifactorial or suspected environmental cause). He was born to phenotypically normal parents at 40 weeks of gestation. The mother contracted and was treated for pneumonia at five months gestation. Maternal smoking during pregnancy was documented, at a frequency of 4-5 cigarettes per day. Clinical history revealed several hospital admissions for seizures and was being managed using Epilim. It was also noted that he used to take Ritalin, but no additional information was provided. Diagnostic tests performed prior to recruitment included Fragile X, MPLA, Karyotype, which were all normal. Family history indicated a history of deafness and delay within the extended family. Additionally, the father has been diagnosed with dyslexia. The full list of the patient's phenotypic anomalies and affected system anomalies using appropriate HPO term and codes is summarized in Table 3.37.

Table 3.35. Clinical phenotype for patient D3S.0119 described using Human Phenotype Ontology terms.

HPO ID	HPO term	HPO category
HP:0000457	Depressed nasal ridge	Abnormality of head or neck
HP:0025383	Dorsocervical fat pad	Abnormality of the musculoskeletal system
HP:0002373	Febrile seizure (within the age range of 3 months to 6 years	Abnormality of the nervous system
HP:0002007	Frontal bossing	Abnormality of the musculoskeletal system
HP:0001290	Generalized hypotonia	Abnormality of the musculoskeletal system
HP:0002857	Genu valgum	Abnormality of the musculoskeletal system
HP:0002265	Large fleshy ears	Abnormality of the ear
HP:0011343	Moderate global developmental delay	Abnormality of the nervous system
HP:0000189	Narrow palate	Abnormality of head or neck
HP:0001513	Obesity	Growth abnormality
HP:0000629	Periorbital fullness	Abnormality of head or neck
HP:0009237	Short 5th finger	Abnormality of the musculoskeletal system
HP:0012745	Short palpebral fissure	Abnormality of head or neck
HP:0001182	Tapered finger	Abnormality of the musculoskeletal system
HP:0012471	Thick vermilion border	Abnormality of head or neck
HP:0001956	Truncal obesity	Growth abnormality
HP:0000582	Upslanted palpebral fissure	Abnormality of head or neck
HP:0012810	Wide nasal base	Abnormality of head or neck
HP:0000687	Widely spaced teeth	Abnormality of head or neck

- *Molecular evaluation*

A missense variant in *RPS6KA3* (#,300075) gene, c.734A>G. (p. Val601Glu). Sanger sequencing confirmed the presence of the variant in a hemizygous state in the patient (Figure 3.43). The gene encodes a amino acid (RSK2) kinase which is involved in the Ras-Mitogen-Activated Protein Kinase (MAPK) signalling cascade important for cell proliferation, cell survival, apoptosis, metabolism, and cancer. Segregation analysis was performed following sequencing of both parents DNA demonstrated a reference allele sequence for both thus suggesting the variant has *de novo* mode of inheritance This novel variant was classified as ACMG-AMP likely pathogenic as evidenced in Table 3.5.

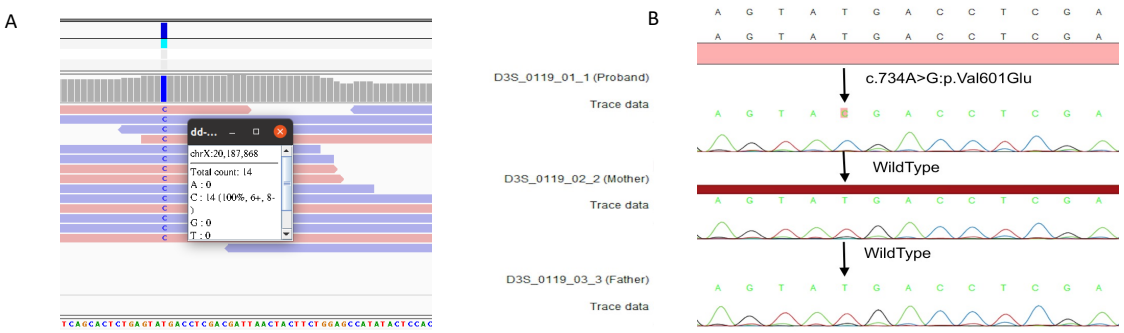

Figure 3.43. Validation of *RPS6KA3* variant in patient D3S.0119. (A) A hemizygous RPS6KA3 variant visualized by IGV (variant base C in blue) in the proband. A total of 14 reads covered this region and all supported the alternative allele (VAF 100%). B) Sanger sequencing analysis of the RPS6KA3 variant visualized by CLC bio in the proband, mother, and father confirms the variant is de novo inherited.

- *Molecular diagnosis*

Mutations of *RPS6KA3* have been associated with Coffin–Lowry syndrome (# 303600). Individuals with the disorder, present with moderate-severe ID, and distinct facial dysmorphisms become more apparent with age, hand and skeletal malformations, psychomotor retardation, and hypotonia. Our patient presented, coarse facial features, including, deeply set eyes, short palpebral fissure, high palate, buffalo hump, frontal bossing, left foot low arch, and hypotonia. This molecular diagnosis confirmed the suspected clinical diagnosis of Coffin-Lowry which was based on characteristic face and hand dysmorphisms.

Individuals with Coffin-Lowry, develop late-onset Kyphoscoliosis which leads to cardiorespiratory complications. It is recommended that individuals with a positive diagnosis of Coffin-Lowry be monitored for progressive spine abnormalities suggestive of Kyphoscoliosis.

3.12.30. Study diagnostic yield summary

Pathogenic variants explaining the patient phenotype were identified in 25.2% of the families. This was achieved following variant filtering and prioritization using the DDG2P and *de novo* pipelines. However, for 75% (86/115) of the families in the DD cohort, there remains unexplained DD.

3.13 Summary of patient recommendations based on a positive molecular diagnosis

A positive molecular diagnosis in 15 of the patients resulted in improved clinical management, disease progression management and surveillance. These recommendations are based on literature from similarly affected DD individuals from other cohorts, and a summary of these clinical recommendations have been provided in Table 3.38.

Table 3.36. Summary of diagnosis and impact of WES results in patients with disease-causing variants.

Patient ID	GENE	Diagnosis based on WES (OMIM ID)	Significant impact
D3S.0051	*CHD7*	CHARGE syndrome (300860)	Monitor spine development for scoliosis
D3S.0053	*AUTS2*	Syndromic ID (615834)	Treatment and management on severe feeding issues
D3S.0056	*PLP1*	Pelizaeus-Merzbacher (312080)	Focused physiotherapy on muscle stiffness
D3S.0064	*SAMD4*	Juvenile polyposis syndrome (175050) Myhre syndrome (139210)	Monitoring and surveillance of colon and stomach cancers > 15 years old.
D3S.0086	*STXBP1*	Epileptic encephalopathy, early infantile (612164)	Vigabatrin for seizures and epilepsy evaluation at next clinical visit.
D3S.0111	*PTEN*	Macrocephaly/Autism syndrome (605309)	Monitoring and surveillance screening for various cancers and polyps' development
D3S.0118	*FLNA*	Terminal Osseous Dysplasia (300244)	Regular testing for hearing loss
D3S.0129	*MAPK1*	Noonan-like syndrome (619087)	Regular testing for hearing
D3S.0015	*NFIX*	Malan syndrome (614753)	Vitamin D dosage, monitoring and surveillance for emergence of subtle neurological signs such as seizures and vomiting
D3S.0034	*CHD7*	CHARGE syndrome (300860)	Monitor spine development for scoliosis
D3S.0035	*OCRL*	Lowe Syndrome (309000)	monitoring and surveillance for renal abnormalities and kidney disease
D3S.0047	*ARID1B*	Coffin-Siris syndrome (135900)	Monitor and surveillance for thyroid cancer
D3S.0050	*CHD7*	CHARGE syndrome (300860)	Monitor spine development for scoliosis
D3S.0117	*NSD1*	Sotos syndrome (117550)	Monitor for thoracic scoliosis and genetic counseling
D3S.0119	*RPS6KA3*	Coffin-Lowry syndrome (303600)	Monitor for progressive spine abnormalities, Kyphoscoliosis leading to cardiorespiratory difficulties

3.14 Summary of key research results

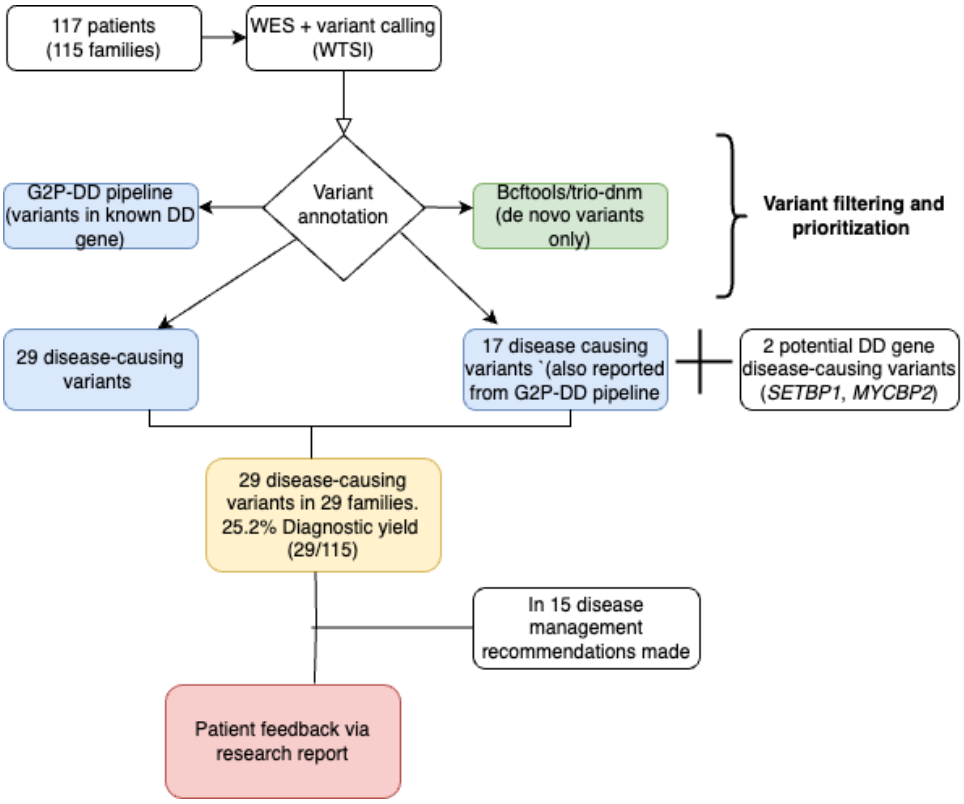

Figure 3.44. Summary of key PhD research results.

Chapter 4 - Discussion

The genetic aetiology of most patients diagnosed with DD remains to be elucidated. In South Africa, the proportion of patients with unexplained DD is potentially much higher on account of limitations in the available genetic tests, and access to genetic services (Kromberg et al., 2013; Lumaka et al., 2022; Wiener et al., 2022). The use of WES has been shown to be beneficial in delineating the underlying aetiology of the disorder and it is thus currently the recommended first-line test in this group of patients (Manickam et al., 2021; Srivastava et al., 2019).

This PhD study aimed to first, investigate whether this utility of WES can be replicated in an African setting as evidenced in non-African cohorts. Secondly, we aimed to make recommendations for a variant filtering and prioritization strategy appropriate for African patients with unexplained DD, thus making the process of WES data analysis for DD patients more efficient.

4.1 The diagnostic utility of WES in African DD patients

Pathogenic variants overlapping known DD-associated genes were found to explain the patient phenotype in 30 patients from 115 families (two pairs of siblings included). This corresponded to an overall diagnostic yield of 25.2%.

The reported diagnostic yield is in-line with that of 22% reported among 200 patients with ID and a range of clinical features predominantly, epilepsy, hypotonia, and craniofacial dysmorphisms as reported by Valentino et al., 2021. The DDD-UK study which is similar to the DDD-Africa study, reported a diagnostic yield of 27% (311/1133) (Wright et al., 2015).However, this result included a much larger sample size and the identification of both causative SNVs and CNVs in the DD cohort. Thus, a direct comparison between the WES diagnostic yields for SNVs and indels was not possible.

Disease-causing variants were reported among patients following a minimum of two genetic tests performed prior to study enrolment (Figure 3.1). These results highlight the diagnostic superiority of WES over traditional genetic tests such as karyotype and MLPA which were the most common tests performed in this DD cohort (Figure 3.2; Figure 3.3).

Several studies as evidenced by the current DD testing guidelines (Manickam et al., 2021) have also demonstrated this diagnostic benefit of WES in unveiling the underlying variant contributors of heterogeneous disorders. In one study, a diagnostic yield of 29% among 150 patients with a range of neurological symptoms was reported using WES (Vissers et al., 2015). In the same patient cohort, a diagnostic yield of 7.3% was reported for standard diagnostic testing, which included, a range of metabolic and biochemical tests, and targeted gene testing using Sanger sequencing, karyotype, and genomic microarray among others (Vissers et al., 2015). Another study by Li and colleagues evaluated the diagnostic utility of WES as a first-line test for patients with DD and a range of cognitive anomalies. The authors found that the use of WES resulted in a higher diagnostic yield of 46.6% (n=466) compared to 18.5% by traditional testing (n=185) for 1000 individuals tested (C. Li et al., 2021).

A total of 27 distinct molecular diagnoses were reported in this study, re-emphasizing the high level of genetic heterogeneity associated with DD. A different DD diagnosis from that suspected initially based on the patient clinical phenotype was reached in 12.1% of the patients (Table 3.8). For example, a positive diagnosis of Coffin-Siris (# 135900) due to a pathogenic variant in *ARID1B* was reached in the patient (D3S.0047). Initially, the patient had a clinical

suspicion of Cornelia de Lange syndrome (CdLS). There is phenotype overlap between the two syndromes, with nail hypoplasia, poor growth, and limb abnormalities reported in both syndromes (Iwamoto et al., 2021). These results reinforce that diagnosis based on clinical examination alone is challenging due to the high level of phenotypic heterogeneity known to underlay DD. Additionally, these results suggest and further support evidence demonstrating that a hypothesis-free sequencing approach, such as that offered by WES, may be more appropriate than traditional genetic tests for the identification of disease-causing variants in DD.

4.1.1. Leveraging known DD gene and variant data results in a positive molecular diagnosis

A positive molecular diagnosis in this study was reached by restricting the variant analysis to known DD genes as defined using the DDG2P gene list (release v2.2). An average of 63 variants per patient were prioritised across the cohort using DDG2P. This is ~2x higher than that reported in the much larger DDD-UK cohort (1,133 patient), which reported an average of 30 variants per patient (Wright et al., 2015).

This observation could reflect the known high genetic diversity of African ancestry individuals compared to that of European ancestry individuals. However, it is more likely a result of the presence of low-quality variants which were only identified downstream. This followed further prioritization using *in-silico* pathogenesis predictor algorithms and variant quality control. A mean of 21 high-quality variants and 16 low-quality variants were identified (section 3.6.1).

Additionally, the ClinVar variant and disease database was used to prioritise variants in known DD genes, by filtering out those previously classified as benign, likely benign and VUSs in the ClinVar database. This identified a slightly higher proportion of disease-causing variants which had previously been reported in ClinVar (16/29) among the patient with a positive diagnosis.

Studies in European ancestry patient have also reported a high prevalence of disease-causing ClinVar known variants (Hiraide et al., 2021; Järvelä et al., 2021; Stojanovic et al., 2020). In a study that reported on the utility of WES in a cohort of patient presenting with moderate-severe global developmental delay, a positive molecular diagnosis was achieved in 49 of the 88 patients. Of these 12 patients had a positive molecular diagnosis due to known ClinVar variants, corresponding to a diagnostic yield of 13.6% (Stojanovic et al., 2020). In another study, Hiraide and colleagues reported a diagnostic yield of 53.5% (53/101) in a cohort of 101 patients

presenting with DD and ID (Hiraide et al., 2021). Of these 22/53 patients had disease-causing variants previously reported in ClinVar and alone corresponded to a diagnostic yield of 21.8% (Hiraide et al., 2021).

These studies interrogated ClinVar as a source for known variants, for the interpretation of candidate variants only. This is compared to use of ClinVar as a virtual panel in this current study to prioritise variants in known DD genes. These results in addition to overall diagnostic yield of 25.2% which is due to pathogenic variants in known DD genes, suggests the benefit of using the DDG2P virtual panel of known DD genes, and the ClinVar variant list as the basis of a variant filter and prioritization strategy.

It is however, important to note that the sample size of this current DD cohort is relatively small as it is only a subset of the 350 families recruited by the larger DDD-Africa study. It is possible that a higher number of novel DD variants remains to be identified in patients from the larger DDD-Africa study that remain with unexplained DD. This is possibly true as African ancestry individuals present with a high level of novel disease-causing variants, compared to those of European ancestry mainly due to an underrepresentation of this ancestry group in population reference databases and a lack of literature and data on causative variants in African patients. It is thus not surprising that lower diagnostic yields among different African disease cohorts have been reported compared to those of European ancestry when testing for known disease variants that prove common in non-African patients (Lebeko et al., 2015). For example, genetic testing for inherited retinal disease using a genotype-based microarray, consisting of variants commonly associated with the disease in mostly European individuals resulted in a molecular diagnostic yield of 12.8% among (n=109) African ancestry patients compared to 41.25% reported in (n=279) European ancestry patients (Roberts et al., 2016). Of note, the previously reported variants had only been seen in one other patient of Africa ancestry. Therefore, the study results suggested that African patients have a high number of novel and exclusive disease-causing variants. The authors also noted that the reported diagnostic yield was lower than that of similar studies in European population cohorts, thus suggesting that novel inherited retinal disease genes may be causal in African patients (Roberts et al., 2016).

Therefore, the overall contribution of causative known variants in this African DD patient cohort can only be clearly made once WES data from of the rest of the DDD-Africa study cohort has been analysed.

4.2 The benefits of a positive molecular diagnosis in DD patient

Clinical utility in part refers to the impact of a positive diagnosis on the clinical outcome and management for the patient. In this study, for 15 patient, recommendations for improved patient clinical management, disease progression and comorbidity surveillance could be made based on available literature from similarly affected individuals (Table 3.. For example, patient D3S.0015, presented with characteristic features suggestive of Malan syndrome. Recommendations for clinical evaluation and management of Malan syndrome have been outlined by a multidisciplinary team from the Ospedale Pediatrico Bambino Gesù (Macchiaiolo et al., 2022). A high reoccurrence of long bone fractures has been noted in patient with *NFIX* mutations, thus the committee recommends dosage with Vitamin D during puberty in patient. They also recommend that caregivers be made aware of the emergence of subtle neurological signs such as fainting, seizures and vomiting in these patient (Macchiaiolo et al., 2022).

In another patient, D3S.0035, clinical features were considered adequate for Lowe syndrome, however, unlike previous reports of the syndrome, at the time of the clinical assessment, no renal abnormalities were reported. Therefore, renal testing in this patient for signs of kidney disease also referred to as Fanconi syndrome is recommended (Abdalla et al., 2018). Kidney disease has been reported to worsen with age in patient eventually leading to chronic renal disease. It has been reported that death in most patient is due to renal failure with the longest surviving patient being a 54-year-old male (Ingemi et al., 2012).

Narrow definitions of clinical utility may focus on a test's ability to produce a diagnosis, but broader definitions of clinical utility consider health and non-health related, familial and societal outcomes as proposed by the ACMG (Watson, 2015).

Reaching a positive molecular diagnosis in this study has ended the diagnostic odyssey associated with unexplained DD for this subset of patient. Diagnostic odysseys result in both financial and emotional costs to patients and their families (Chinnery, 2021; Grier et al., 2018; Lavelle et al., 2022). Monroe et al., (2016), estimated a cost of 16,409 USD per patient for serial testing using traditional diagnostic tests. This cost estimate included genetic tests, biochemical tests, specialist visits, and hospital admissions from disorder comorbidities. The once-off cost for trio-based WES was estimated at a mean of 3,972 USD and resulted in a diagnostic yield of 29.4% following negative results from traditional diagnostic testing (Monroe et al., 2016).

In a recent study by Miller et al (2021), families expressed frustration over having the child's symptoms continue to worsen yet no positive diagnosis had been made (Miller, 2021). Additionally, parents expressed continually second-guessing themselves over lifestyle and treatment choices given they could not predict how these would affect the child's overall health and disease progression. These studies highlight the potential financial cost savings related to the implementation of WES as a first-line diagnostic test and ending the diagnostic odyssey associated with a negative genetic test.

A positive molecular diagnosis also allows for the establishment of disease specific support groups which enables families and parents to share their experiences and provide emotional support for one another while managing their children's disorder.

It also provides improved genetic counseling which includes recurrence risk assessments and overall better informed decision making for future pregnancies. For example, it is suggested that the recurrence rate of *de novo* variants for parents with a child with a genetic disease caused by a *de novo* mutation is around 1-4% higher than that of the general population (Campbell et al., 2014). Invasive prenatal testing with either Chorionic Villus Sampling or Amniocentesis can be offered for such families (Peretz-Machluf et al., 2021).

4.3 Phenotypic complexity increases the chance of a positive diagnosis

WES analysis was performed for 117 patients with unexplained DD and a range of clinical features (Figure 3.1), with most patient having either 12 or 13 HPO terms recorded (Figure 3.2). Results showed that the likelihood of reaching a positive molecular diagnosis potentially increases as a higher number of dysmorphic features (represented in this case by the number of HPO terms) are observed in the patient (Figure 3.14). The current DD cohort is phenotypically well defined, as 65.0% (76/117) of the patient had at least one malformation or at least 3 well-documented minor anomalies.

Similar observations have been made in other patient cohorts (Brunet et al., 2021; Trujillano et al., 2016; Yang et al., 2014). A higher diagnostic yield (52.1%, n =111/213) was reported for patients presenting with neurodevelopmental disorders and associated clinical features compared to patients with isolated NDD (22.2%, n = 4/18) (Brunet et al., 2021). Trujilano et al (2016) noted that in a group of patients with 6-15 HPO terms the diagnostic yield was slightly higher at 33%, compared to that of 26% for those with 1-5 HPO terms recorded.

There is no clear explanation for this observation, but it is probably due to (i) the prioritization of variants in patient presenting with one or only a few clinical features can result in the prioritization of a relatively higher number of candidate genes and variants given the known phenotype overlap within DD. This makes identifying a single causative variant and gene more challenging; (ii) the more complex the clinical features of the disorder, the higher the probability of identifying distinctive clinical features that are associated with a specific disease and monogenic cause.

Therefore, for patient with a higher number of well-defined clinical malformations, there is improved candidate variant interpretation which can potentially result in a higher diagnostic yield.

4.4 A high proportion of rare variants makes data interpretation challenging

Most variants that are rare or absent in ancestry-specific populations are considered potentially pathogenic during variant analysis (Gudmundsson et al., 2022). Reference population data from individuals of African ancestry who have been shown to present with a high level of genetic diversity (Sherman et al., 2018) is limited in current large, curated population reference databases (Gudmundsson et al., 2022; Lumaka et al., 2022). The interpretation of allele frequency data in the DD patient cohort showed a high number of rare and unique variants that have not been previously observed in the gnomAD reference population database (Figure 3.8). This was evidenced by most of the variants having an extremely low allele frequency (allele frequency approximately $1x\ 10^{-6}$) thus, suggesting a high number of rare variants in the patient cohort. This could be because of an underrepresentation of African ancestry individuals in the currently available reference population databases. The gnomAD v2.1 database (version used in variant annotation) has a lower sample size of 12,487 for African ancestry individuals compared to 112,386 for European individuals (comprised of non-Finnish Europeans, North-Western European, Southern European, Other non-Finnish European) (Laurent Francioli et al., 2018).

A consequence of the low African ancestry sample size in the current reference population databases, is a relatively high number of VUSs compared to well characterized European ancestry populations (Macklin et al., 2018; Ndugga-Kabuye & Issaka, 2019; Tsai et al., 2019). Of the variants prioritised as probably pathogenic using *in-silico* pathogenesis predictor algorithms, 48.1% (1210/2513) were VUSs (Figure 3.10).

It is important to note that VUSs identified in this current study were not further assessed in an effort to identify those that could potentially prove pathogenic, and those that would likely be downgraded to benign. The analysis strategy taken for this DDD-Africa cohort was to look at VUSs only after the additional CNV screening filter was applied (performed as part a different PhD project under the larger DDD-Africa study), thus the true contribution of VUSs in this current PhD study has not been evaluated.

4.5 WES results suggest a high contribution of *de novo* variants in DD

The availability of 63 complete trios allowed for the prioritization and identification of pathogenic *de novo* variants using the *de novo* prioritization pipeline. A total of 12/29 (41.4%) confirmed *de novo* variants were identified in the cohort compared to 34.5% (10/29) that were inherited. The number of *de novo* variants is potentially higher, 65.5% due to seven suspected *de novo* variants identified in duo families. Our results suggest a high burden of causative *de novo* variants in DD and highlights the benefit of trio exome analysis for the identification of causative DD variants.

Other studies have demonstrated the high burden of *de novo* variants in DD. (Study, 2017; Wright et al., 2015). A study conducted by (Brunet et al., 2021), reported a 40.3% (93/231) contribution of *de novo* variants to the overall diagnostic yield in a cohort of patients with neurodevelopmental disorders (Brunet et al., 2021). In smaller study, a diagnostic yield of 48.4% due to *de novo* variants were reported among 87 patients with neurodevelopmental disorders (Álvarez-Mora et al., 2022). The DDD-UK study reported a diagnostic yield of 42% due *de novo* SNVs and CNVs (Wright et al., 2015).

4.5.1. The identification of potential novel DD variants is challenging.

In patient with no disease-causing variants identified using the DDG2P pipeline, potential DD genes were identified and assessed based on the identification of pathogenic variants. Five of the shortlisted genes had been previously reported in other DD cohorts but were classified as VUSs and considered as DD research genes in DECIPHER on account of limited evidence for disease association and pathogenesis (Table 3.38). There is limited patient phenotype information available on DECIPHER thus phenotype overlaps with the patients in the DDD-cohort could not be completed.

Additionally, an assessment of the genes using pLI and the missense Z-score as indicators for an intolerance to loss-of-function and missense variants which would likely result in disease (Betancur & Buxbaum, 2020; Kosmicki et al., 2017), indicated that only one of the genes, *BRWD1* showed a high intolerance (pLI =1) for loss-of function variants.

Thus, there was limited evidence to support these genes as strong candidate novel DD genes. In the absence of animal models and functional studies, these genes need to be shared with the larger community via DECIPHER and the Matchmaker exchange, to identify any other patient with similar genes and clinical phenotypes.

The *de novo* pipeline did however identify, gene *MYCBP2* in patient D3S.0017. The gene is classified as a DD-limited gene in DECIPHER, this is on account of limited pathogenesis information, which includes adequate functional studies demonstrating human phenotype association, and the reporting of pathogenic variants in a high enough number of patient with similar phenotypes. Thus, this gene is not included as part of the DDG2P variant list (release v2.2) and was therefore not identified using the DDG2P pipeline.

An evaluation of DECIPHER and literature shows that VUSs, have been identified in the gene, (Bertoli-Avella et al., 2021; BM et al., 2012; Kosmicki et al., 2017) with the gene included as part of the DD research gene list on DECIPHER.

A recent study published in 2022, provided evidence for abnormal axon development in humans through animal model and *in-vivo* gene editing (AlAbdi et al., 2022). They also reported pathogenic *de novo* variants in eights patient with overlapping clinical features. Common features among the patient were developmental delay (7/8), ID (4/8), corpus collosum (5/8) (for the other three brain imaging not available), eye and vision abnormalities (6/8) (AlAbdi et al., 2022). The authors have suggested that the disorder be referred to *MYCBP2* related DD with collosum defect (AlAbdi et al., 2022).

A missense variant (c.6788C>A (p. Ser2263Ter)) was identified in patient, D3S.0017 (Table 3.6) and in addition to a total of six missense variants (6/8), identified by AlAbdi et al (2022), support the gene constraint metric results that show that the gene is highly intolerant to missense variants, with a Z- score of 6.05.

This gene and variant will be further reviewed and evaluated by the DDD-African multi-disciplinary team together with the outcomes of the other variant analysis pipelines (CNV analysis and germline mosaicism) to be able to draw a more definitive conclusion regarding its contribution to disease in the patient.

In another patient, D3S.0052, a hemizygous missense variant, c.961C>T (Arg321Trp) overlapping *MECP2* gene. Pathogenic mutations in the gene affect its ability to recognize and interact with the methyl marks in different cofactors (Kriaucionis and Bird, 2003) and cause Rett syndrome (#312750).

4.6 Strict variant quality control is sufficient to identify false-positive variants

Adequate read coverage results in increased accuracy and confidence in variant calls for distinguishing true variants from sequencing errors (Sims et al., 2014). Despite this, sufficient read coverage at capture target regions and sequencing of repeat sequences remains a major challenge (T. F. Beck et al., 2016; McCourt et al., 2013; Strom et al., 2014). These can lead to sequencing, alignment, and variant calling errors which can result in false-positive and false-negative variant calls (T. F. Beck et al., 2016; Sikkema-Raddatz et al., 2013).

Currently, there is no consensus in the recommended average WES sequence depth. A study by Fang et al., (2014), recommended a coverage of 60x to achieve 95% accuracy in variant calls (Fang et al., 2014). This is in comparison to Ajay et al., (2011) who showed that a coverage of 35x was required to achieve an accuracy of 95% in calling SNVs and insertions and deletions (Ajay et al., 2011). The average sequence coverage of 31.6x (Figure 3.6) which was achieved in this current study is lower than previous reports (Ahn et al., 2009; Wang et al., 2008). However, we were able to demonstrate that this mean coverage, is adequate to reliably detect true variants.

Sanger sequencing failed to positively confirm any of the 24 low-quality variants identified using variant quality control filters (DP, and VAF) and prioritised as plausible variants due to genotype-phenotype correlation. An assessment of the variant characteristics showed that none of the variants passed both quality filters. Thus, the variant filters applied were sufficient to identify most false-positive variants resulting from NGS-related errors.

4.6.1. Sanger sequencing may not always be required for NGS variant confirmation

The ACMG and the College of American Pathologists (CAP) initially recommended the confirmation of identified variants to reduce the chance of reporting possible false positive variants (Aziz et al., 2015; Rehm et al., 2013).

Sanger sequencing positively confirmed 30/32 high-quality variants identified. The negative confirmation results from *CHAMP1* and *EHMT1* led us to suspect the occurrence of locus-specific allelic drop-out. This is a phenomenon that is said to be a result of selective allele amplification during PCR-based sequencing techniques such as Sanger sequencing (Martins et al., 2011). During allelic drop-out, the reference allele on one read strand could be preferentially amplified while on the other strand, the alternative/variant allele will not be amplified or detected "dropped" (Shestak et al., 2021). In such a case, the heterozygous variant will not be detected rather a homozygous reference genotype will be detected at that position (Shestak et al., 2021).

In a study by Beck et al (2016), a negative result was initially received for 19 of the 5,800 high-quality variants selected for confirmation using Sanger sequencing. For two variants, a positive result was achieved following a second Sanger sequencing run using the same primer pairs. A second sequencing run was performed for another four variants, but a negative result was again received (homozygous reference genotype at position). A new primer pair was designed, and a second sequencing run was performed for another two variants, resulting in the positive confirmation of both variants. In a study by Shestak et al (2021), Sanger sequencing was performed to confirm the presence of a heterozygous c.2091A>G variant in *ARVD16* gene (Shestak et al., 2021). Following three consecutive Sanger sequencing runs, a positive result was achieved only once, with the other two independent runs failing to confirm the presence of the variant. The authors were able to identify a common intronic variant (c.1904-49T>A) at the 3'-end of the primer binding site (Shestak et al., 2021).

In these cases, they hypothesised that there was a cryptic private polymorphism in the patient sequence which may have led to allelic drop-out. In this current study, the presence of common SNVs at the primer binding site was assessed during primer design using SNPCheck® design tool, and no common variants were found. There is however the possibility of common SNVs that have not been identified in African ancestry individuals given the underrepresentation of this ancestry group in current reference population databases, which are also utilized by tools such as SNPCheck®.

The studies by Beck et al (2016) and Shestak et al (2021) highlight the negative impact of allelic drop-out in the use of PCR-based techniques, as they led to incorrect genotype calls and false-negative variant results. Moreover, they highlight the inconsistency in the occurrence of allelic drop-out as multiple sequencing runs with the same primer pairs can lead to different genotyping results. Based on these results it is possible that the two variants in *CHAMP1* and *EMHT1* could still be confirmed with a different sequencing method.

These results contribute to the growing discussion regarding the need for confirmation of NGS-identified variants with a large body of evidence suggesting that given sufficient variant quality, confirmation of variants may not be necessary. Our reported validation rate is comparable to that reported by other studies at 94-100% (Baudhuin et al., 2015; T. F. Beck et al., 2016; McCourt et al., 2013; Sikkema-Raddatz et al., 2013; Strom et al., 2014). A recent study, Arteche- López and colleagues, assessed the need to validate high quality variants using 1109 SNV and indel in 825 clinical exomes. They reported 100% Sanger validation for these variants (Arteche-López et al., 2021).

The confirmation of variants using a secondary diagnostic test would no doubt increase the overall cost of WES. This cost is even higher as the number of variants to be confirmed increases, cascade testing is requested, and several sequencing attempts and primer pair redesigns and several optimisations are completed. It also necessitates the need for additional lab infrastructure to support large-scale Sanger workflows.

Additionally, variant confirmation can result in major prolonged turnaround times thus delaying the benefits associated with a positive diagnosis. Moreover, should the variant prove difficult to amplify following multiple sequencing attempts, then patient feedback should be delayed as the reporting of false positive variants can led to the wrong treatment and clinical management recommendations in these patients.

Currently, there are no consensus guidelines on the variant quality parameters to use to identify false-positive variants. The use of variant quality score (Strom et al., 2014), and VAF and DP (Mu et al., 2016) have been suggested. Mu et al., suggested the use DP≤100X in combination with VAF of 40-60% to identify true positive variants (Mu et al., 2016). Based on the results of this study, we recommend the combined use of GATK VQSR filter (99.9<sensitivity <100%), DP ≤10 and VAF ≤0.3 cut-offs for WES variant quality control to identify false-positive variants.

4.7 Limitations of the study

During this study, we used allele frequency from gnomAD database to filter variants based on MAF. Our data supported the knowledge that African ancestry samples are underrepresented in such public reference databases. Thus, we prioritised a large number of variants which following interpretation were classified as VUSs.

4.8 Future work

- *In-house allele frequency database*

Variant filtering using allele frequency is a critical step in the prioritization of rare potentially pathogenic variants as already mentioned. Future work to improve the use of allele frequency and identify common variants in the African population could include the creation of an in-house allele frequency database. This could be accomplished by using allele frequency data from unaffected parents as healthy controls for the identification of population-specific common variants. An increased sample size by using both the in-house allele frequency and publicly available reference population databases could improve the identification of truly rare variants and can improve variant and prioritization in African patients.

- *Data sharing within the community*

There are initiatives to increase the level of available African population ancestry genomic data and representation through projects such as the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID) (the reference homepage can be found in Appendix 1), and the African Genome project (AGVP) (Gurdasani et al., 2015). It is important that sequencing data from this current study and others be curated, shared, and submitted to the large reference population databases.

The DDD-Africa data is to be shared through ClinVar and DECIPHER databases. This increased representation can allow for improved variant analysis and interpretation especially if NGS is to be widely implemented and utilized in both research and diagnostic settings.

- *WES data reanalysis*

The reanalysis of the WES data using the *de novo* prioritization strategy in our study cohort identified a pathogenic variant (NM_015559.3(*SETBP1*) c.2612T>C (p. Ile871Thr)) which at

the time of initial analysis using the DDG2P filter and prioritization pipeline was classified in ClinVar as a "VUS". However, at the time of data reanalysis using the *de novo* prioritization pipeline, the variant was annotated as "conflicting interpretation" due to different ACMG-AMP classifications by submitters.

Additionally, we identified two variants, *DSPP* c.3262del (p. Ser1088ValfsTer226), and *ABCB6*, c.376del (p. Val126SerfsTer124) which at the time of initial analysis (date accessed:(07.2021) were annotated as pathogenic in ClinVar (RCV001336859.1, RCV001330497.1), but later during the write-up of this thesis had been deleted from the ClinVar database. This was most likely a deletion by the submitter following variant re-interpretation and a change in variant ACMG-AMP classification. Our interpretation of these variants using standard ACMG-AMP guidelines classified both variants as benign.

These results highlight the continually evolving nature of gene-disease and variant-disease discovery and associations, and the need for periodic reanalysis of exome sequencing data. Thus, the reanalysis of WES in patients that remained with unexplained DD in our cohort should be performed.

Chapter 5 - Conclusion

This study focused on the identification of the underlying genetic aetiology in African patient with unexplained DD using WES. To achieve this, we implemented a variant filtering and prioritization pipeline that identified variants in known DD genes. This resulted in the identification of disease-causing variants in 25.2% of the families in the DD cohort. This result demonstrates the utility of WES as a first-tier test for African DD patients and it's utility in African public healthcare facilities.

The reported diagnostic yield is on account of restricting the analysis to variants overlapping known DD genes, with a slightly higher proportion of variants having been previously reported in ClinVar. Therefore, we recommended that both the ClinVar variant database and the DDG2P gene panel, and the DDG2P gene panel be integration as first-line strategies for variant filtering prioritization workflows in both research and diagnostic settings integration as first-line strategies for variant filtering prioritization workflows in both research and diagnostic settings.

Additionally, we have identified the first African cases for a range of rare genetic disorders. Thus, this study contributes information relating to the genetic aetiology and clinical phenotype of African DD patients to the current pool of knowledge.

The study has also demonstrated that WES with low mean coverage at ~32x is adequate for the identification of variants. Variant quality control filters, GATK VQSR filter (99.9 < sensitivity 100%), DP ≤10, and VAF ≤0.30 should be applied for the removal of false-positive variants. Moreover, variant confirmations using Sanger sequencing may not be necessary given adequate variant quality control. The results and observations from this pipeline will be used to further improve the variant quality control in the analysis of the rest of the DDD-Africa WES data prioritization pipeline were not performed.

In conclusion, based on the results from this study, a recommendation for the implementation of WES as a first-line genetic test for DD patients in an African setting is made.

Appendices

Appendix 1. URLs for applications referred to in this thesis.

- CLC Bio Main Workbench: https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-main-workbench/
- Ensembl consequences: (https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html/)
- GnomAD database https://gnomad.broadinstitute.org/
- NeuroDev: https://www.neurodevproject.org/
- OligoAnalyzer web-based tool (https://eu.idtdna.com/calc/analyzer)
- Picard MarkDuplicates: http://broadinstitute.github.io/picard/ .
- Primer 3: http://primer3.ut.ee/
- Primer BLAST: https://www.ncbi.nlm.nih.gov/tools/primer-blast/
- SNPCheck: https://genetools.org/SNPCheck/snpcheck.htm
- UCSC *in silico* PCR: https://genome.ucsc.edu/cgi-bin/hgPcr
- Primer 3: http://primer3.ut.ee/
- VQSR: https://gatk.broadinstitute.org/hc/en-us/articles/360035531612-Variant-Quality-Score-Recalibration-VQSR-

Appendix 2. Ethical clearance certificate for DDD-Africa study

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Prof. A Krause & Drs N Carstens & Z Lombard

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M180678

NAME: (Principal Investigator)	Prof. A Krause & Drs N Carstens & Z Lombard
DEPARTMENT:	School of Pathology Department of Human Genetics National Health Laboratory Service
PROJECT TITLE:	Decyphering Development Disorders in Africa (DDD-Africa) - evaluating clinical exome sequencing in an African setting
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M160830 NIH Study No. UO1MH115483
SUPERVISOR:	Not applicable
APPROVED BY:	Professor CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	06/07/2018

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

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on 3rd floor, Phillip V Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **June** and will therefore be due in the month of **June** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature — Date: 15/07/2018

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix 3. Ethical clearance certificate for this current PhD study

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Ms M Molatoli

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M200688

NAME: (Principal Investigator)	Ms M Molatoli
DEPARTMENT:	School of Pathology Department of Human Genetics Medical School University
PROJECT TITLE:	The role of small genetic variants in the aetiology of development disorders in South Africa - a whole exome sequencing study
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M180678
SUPERVISOR:	Professor Z Lombard and Dr N Carstens
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2020/07/09

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

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the 3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **June** and will therefore reports and re-certification will be due early in the month of **June** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

10 JULY 2020
Date

Appendix 4. IGV and Sanger electrogram for high quality variants not confirmed using Sanger sequencing.

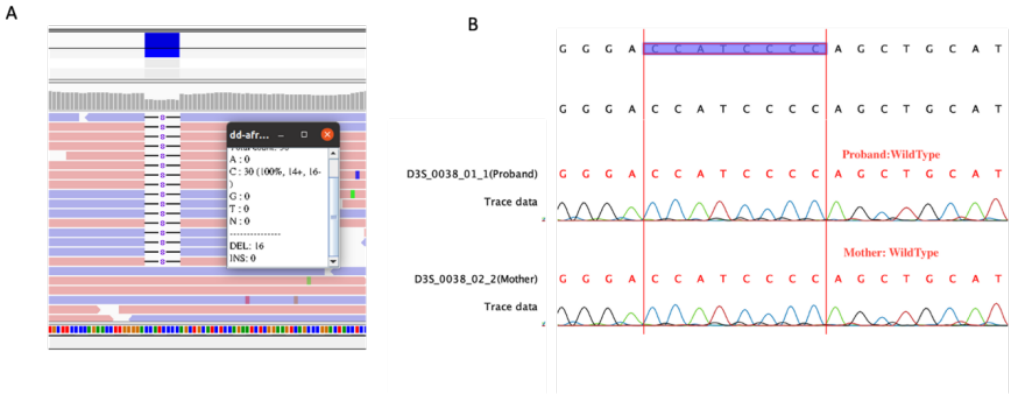

Figure 1. IGV and Sanger electrogram for *CHAMP1* variant in patient D3S.0066.(A) Heterozygous *CHAMP1* 8 base-pair deletion visualized by IGV in the proband (first panel) with reference base detected at this position for samples from mother. (B) Sanger sequencing analysis of the *CHAMP1* at this position visualized by CLC bio in the proband, and mother shows wildtype base for both samples. Thus, variant not confirmed by Sanger sequencing.

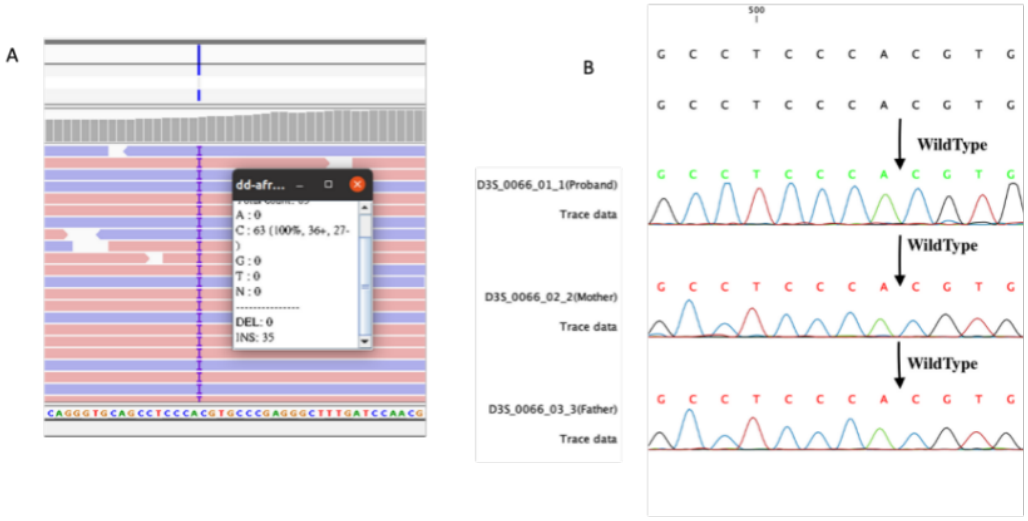

Figure 2. IGV and Sanger electrogram for *EMHT1* variant in patient D3S.0071. Heterozygous *EMHT1* insertion visualized (reference base AC, variant base ACC) in the proband (first panel) and mother (last panel), with reference base detected at this position for samples from the father (middle panel). **(B)** Sanger sequencing analysis of the *EMHT1* at this position visualized by CLC bio in the proband, father, and mother shows a wildtype base for all samples. Thus, variant not confirmed by Sanger sequencing

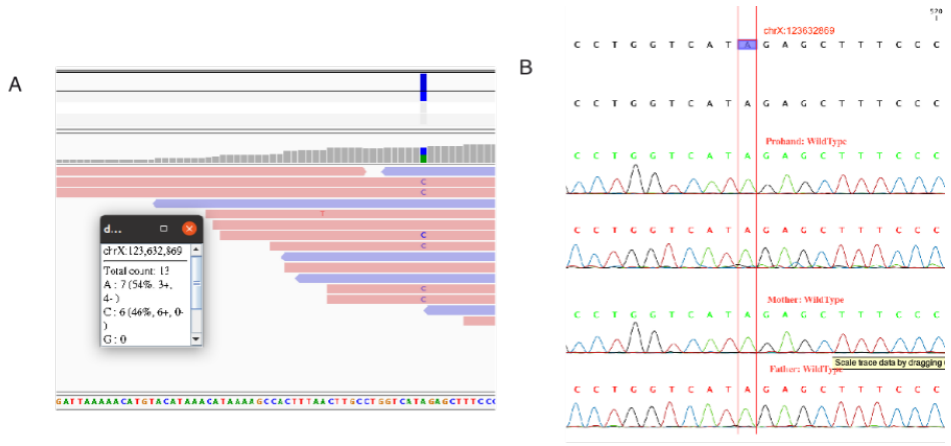

Figure 3. IGV and Sanger electrogram for THOC2 variant in patient D3S.0009.
(A) Heterozygous *THOC2* variant visualized (reference base A in green, variant base C in blue) in the proband with reference base detected at this position for samples from the mother and father. (B) Sanger sequencing analysis of the *THOC2* at this position visualized by CLC bio in the proband, father, and mother shows a wild-type base for all samples. Thus, the variant is not confirmed by Sanger sequencing.

References

Abacan, M. A., Alsubaie, L., Barlow-Stewart, K., Caanen, B., Cordier, C., Courtney, E., Davoine, E., Edwards, J., Elackatt, N. J., Gardiner, K., Guan, Y., Huang, L. H., Malmgren, C. I., Kejriwal, S., Kim, H. J., Lambert, D., Lantigua-Cruz, P. A., Lee, J. M. H., Lodahl, M., … Wicklund, C. (2018). The Global State of the Genetic Counseling Profession. *European Journal of Human Genetics 2018 27:2*, *27*(2), 183–197. https://doi.org/10.1038/s41431-018-0252-x

Abdalla, E., El-Beheiry, A., Dieterich, K., Thevenon, J., Fauré, J., & Rendu, J. (2018). Lowe syndrome: a particularly severe phenotype without clinical kidney involvement. *American Journal of Medical Genetics Part A*, *176*(2), 460–464.

Adebiyi, B. O., & Mukumbang, F. C. (2021). A pooled prevalence of fetal alcohol spectrum disorders in South Africa: a systematic review and meta-analysis protocol. *Archives of Public Health*, *79*(1), 1–6.

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

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Kuster, G. Von, Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., & Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, *44*(W1), W3–W10. https://doi.org/10.1093/NAR/GKW343

Afgan, E., Nekrutenko, A., Grüning, B. A., Blankenberg, D., Goecks, J., Schatz, M. C., Ostrovsky, A. E., Mahmoud, A., Lonie, A. J., Syme, A., Fouilloux, A., Bretaudeau, A., Nekrutenko, A., Kumar, A., Eschenlauer, A. C., Desanto, A. D., Guerler, A., Serrano-Solano, B., Batut, B., … Briggs, P. J. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, *50*(W1), W345–W351. https://doi.org/10.1093/NAR/GKAC247

Ahn, S. M., Kim, T. H., Lee, S., Kim, D., Ghang, H., Kim, D. S., Kim, B. C., Kim, S. Y., Kim, W. Y., Kim, C., Park, D., Yong, S. L., Kim, S., Reja, R., Jho, S., Chang, G. K., Cha, J. Y., Kim, K. H., Lee, B., … Kim, S. J. (2009). The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Research*, *19*(9), 1622–1629. https://doi.org/10.1101/GR.092197.109

Ajay, S. S., Parker, S. C. J., Abaan, H. O., Fuentes Fajardo, K. V., & Margulies, E. H. (2011). Accurate and comprehensive sequencing of personal genomes. *Genome Research*, *21*(9), 1498. https://doi.org/10.1101/GR.123638.111

AlAbdi, L., Desbois, M., Rusnac, D. V., Sulaiman, R. A., Rosenfeld, J. A., Lalani, S., Murdock, D. R., Burrage, L. C., Network, U. D., Billie Au, P. Y., Towner, S., Wilson, W. G., Wong, L., Brunet, T., Strobl-Wildemann, G., Burton, J. E., Hoganson, G., McWalter, K., Begtrup, A., … Alkuraya, F. S. (2022). Loss-of-function variants in MYCBP2 cause neurobehavioural phenotypes and corpus callosum defects. *Brain*. https://doi.org/10.1093/BRAIN/AWAC364

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*(5), 363–376.

Almogy, G., Pratt, M., Oberstrass, F., Lee, L., Mazur, D., Beckett, N., Barad, O., Soifer, I., Perelman, E., Etzioni, Y., Sosa, M., Jung, A., Clark, T., Lithwick-Yanai, G., Pollock, S., Hornung, G., Levy, M., Coole, M., Howd, T., … Lipson, D. (2022). *Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform*. https://doi.org/10.1101/2022.05.29.493900

Álvarez-Mora, M. I., Sánchez, A., Rodríguez-Revenga, L., Corominas, J., Rabionet, R., Puig, S., & Madrigal, I. (2022). Diagnostic yield of next-generation sequencing in 87 families with neurodevelopmental disorders. *Orphanet Journal of Rare Diseases*, *17*(1), 1–10. https://doi.org/10.1186/S13023-022-02213-Z/FIGURES/2

Amendola, L. M., Dorschner, M. O., Robertson, P. D., Salama, J. S., Hart, R., Shirts, B. H., Murray, M. L., Tokita, M. J., Gallego, C. J., Kim, D. S., Bennett, J. T., Crosslin, D. R., Ranchalis, J., Jones, K. L., Rosenthal, E. A., Jarvik, E. R., Itsara, A., Turner, E. H., Herman, D. S., … Jarvik, G. P. (2015). Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Research*, *25*(3), 305–315. https://doi.org/10.1101/GR.183483.114

Amr, K. S., El-Bassyouni, H. T., Hammad, S. A., Ashaat, E. A., Kader, R. M. A. A., El Ruby, M. O., Sayed, S. M., & Attia, H. A. (2020). Assessment of Multiplex Ligation-Dependent Probe Amplification (MLPA) as a diagnostic test for Egyptian patients with Williams-Beuren syndrome. *Gene Reports*, *20*, 100767.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68. https://doi.org/10.1038/NATURE15393

Bainbridge, M. N., Wang, M., Wu, Y., Newsham, I., Muzny, D. M., Jefferies, J. L., Albert, T. J., Burgess, D. L., & Gibbs, R. A. (2011). Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biology*, *12*(7), 1–12. https://doi.org/10.1186/GB-2011-12-7-R68/FIGURES/7

Baldridge, D., Wangler, M. F., Bowman, A. N., Yamamoto, S., Schedl, T., Pak, S. C., Postlethwait, J. H., Shin, J., Solnica-Krezel, L., Bellen, H. J., & Westerfield, M. (2021). Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: Current state and a future vision. *Orphanet Journal of Rare Diseases*, *16*(1), 1–17. https://doi.org/10.1186/S13023-021-01839-9/FIGURES/3

Banks, L. M., Zuurmond, M., Ferrand, R., & Kuper, H. (2014). The relationship between HIV and prevalence of disabilities in sub-Saharan Africa: systematic review (FA)*. *Tropical Medicine and International Health*. https://doi.org/10.1111/tmi.12449

Barbitoff, Y. A., Polev, D. E., Glotov, A. S., Serebryakova, E. A., Shcherbakova, I. V., Kiselev, A. M., Kostareva, A. A., Glotov, O. S., & Predeus, A. V. (2020). Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Scientific Reports 2020 10:1*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-59026-y

Batzir, N. A., Shohat, M., & Maya, I. (2015). Chromosomal Microarray Analysis (CMA) a Clinical Diagnostic Tool in the Prenatal and Postnatal Settings. *Pediatric Endocrinology Reviews : PER*, *13*(1), 448–454. https://europepmc.org/article/med/26540760

Baudhuin, L. M., Lagerstedt, S. A., Klee, E. W., Fadra, N., Oglesbee, D., & Ferber, M. J. (2015). Confirming variants in next-generation sequencing panel testing by Sanger sequencing. *The Journal of Molecular Diagnostics*, *17*(4), 456–461.

Baynam, G., Overkov, A., Davis, M., Mina, K., Schofield, L., Allcock, R., Laing, N., Cook, M., Dawkins, H., & Goldblatt, J. (2015). A germline MTOR mutation in Aboriginal Australian siblings with intellectual disability, dysmorphism, macrocephaly, and small thoraces. *American Journal of Medical Genetics Part A*, *167*(7), 1659–1667.

Baynam, G. S., Groft, S., van der Westhuizen, F. H., Gassman, S. D., du Plessis, K., Coles, E. P., Selebatso, E., Selebatso, M., Gaobinelwe, B., Selebatso, T., Joel, D., Llera, V. A., Vorster, B. C., Wuebbels, B., Djoudalbaye, B., Austin, C. P., Kumuthini, J., Forman, J., Kaufmann, P., … Posada, M. (2019). A call for global action for rare diseases in Africa. *Nature Genetics 2019 52:1*, *52*(1), 21–26. https://doi.org/10.1038/s41588-019-0552-2

Beck, S., Wojdyla, D., Say, L., Betran, A. P., Merialdi, M., Requejo, J. H., Rubens, C., Menon, R., & Van Look, P. F. A. (2010). The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bulletin of the World Health Organization*, *88*, 31–38.

Beck, T. F., Mullikin, J. C., & gov, N. C. S. P. B. L. G. lesb@ mail. nih. (2016). Systematic evaluation of Sanger validation of next-generation sequencing variants. *Clinical Chemistry*, *62*(4), 647–654.

Bélanger, S. A., & Caron, J. (2018). Evaluation of the child with global developmental delay and intellectual disability. *Paediatrics & Child Health*, *23*(6), 403–410.

Bender, K. J., & Trussell, L. O. (2012). The physiology of the axon initial segment. *Annual Review of Neuroscience*, *35*, 249–265.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature 2008 456:7218*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Bertoli-Avella, A. M., Kandaswamy, K. K., Khan, S., Ordonez-Herrera, N., Tripolszki, K., Beetz, C., Rocha, M. E., Urzi, A., Hotakainen, R., Leubauer, A., Al-Ali, R., Karageorgou, V., Moldovan, O., Dias, P., Alhashem, A., Tabarki, B., Albalwi, M. A., Alswaid, A. F., Al-Hassnan, Z. N., … Bauer, P. (2021). Combining exome/genome sequencing with data repository analysis reveals novel gene-disease associations for a wide range of genetic disorders. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *23*(8), 1551–1568. https://doi.org/10.1038/S41436-021-01159-0

Besterman, A. D., Althoff, T., Elfferich, P., Gutierrez-Mejia, I., Sadik, J., Bernstein, J. A., Van Ierland, Y., Kattentidt-Mouravieva, A. A., Nellist, M., Abramson, J., & Martinez-Agosto, J. A. (2021). Functional and structural analyses of novel Smith-Kingsmore Syndrome-Associated MTOR variants reveal potential new mechanisms and predictors of pathogenicity. *PLoS Genetics*, *17*(7). https://doi.org/10.1371/JOURNAL.PGEN.1009651

Betancur, C., & Buxbaum, J. D. (2020). Gene constraint and genotype–phenotype correlations in neurodevelopmental disorders. *Current Opinion in Genetics & Development*, *65*, 69–75. https://doi.org/10.1016/J.GDE.2020.05.036

Biesecker, L. G. (2002). The end of the beginning of chromosome ends†. *American Journal of Medical Genetics*, *107*(4), 263–266. https://doi.org/10.1002/AJMG.10160

Blasko, A. M., Morin, K. L., Bauer, K., Johnson, K. M., Enriquez, G. B., Hunsicker, L. E., Tasik, E. J., & Renz, T. E. (2022). A Meta-Synthesis of Disability Research in Western Africa. *Https://Doi.Org/10.1177/00224669221090231*. https://doi.org/10.1177/00224669221090231

Blok, L. S., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M. R. F., Venselaar, H., Helsmoortel, C., Cho, M. T., & Hoischen, A. (2015). Mutations in DDX3X are a common cause of unexplained intellectual disability with gender-specific effects on Wnt signaling. *The American Journal of Human Genetics*, *97*(2), 343–352.

BM, N., Y, K., L, L., A, M., KE, S., A, S., CF, L., C, S., LS, W., V, M., P, P., S, Y., J, M., EL, C., NG, C., ET, G., O, V., C, S., H, L., … MJ, D. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, *485*(7397), 242–246. https://doi.org/10.1038/NATURE11011

Boggula, V. R., Shukla, A., Danda, S., Hariharan, S. V., Nampoothiri, S., Kumar, R., & Phadke, S. R. (2014). Clinical utility of multiplex ligation-dependent probe amplification technique in identification of aetiology of unexplained mental retardation: A study in 203

Indian patients. *The Indian Journal of Medical Research*, *139*(1), 66. /pmc/articles/PMC3994742/

Bohannan, Z. S., & Mitrofanova, A. (2019). Calling Variants in the Clinic: Informed Variant Calling Decisions Based on Biological, Clinical, and Laboratory Variables. *Computational and Structural Biotechnology Journal*, *17*, 561–569. https://doi.org/10.1016/J.CSBJ.2019.04.002

Borch, L. A., Parboosingh, J., Thomas, M. A., & Veale, P. (2020). Re-evaluating the first-tier status of fragile X testing in neurodevelopmental disorders. *Genetics in Medicine 2020 22:6*, *22*(6), 1036–1039. https://doi.org/10.1038/s41436-020-0773-x

Bowling, K. M., Thompson, M. L., Amaral, M. D., Finnila, C. R., Hiatt, S. M., Engel, K. L., Cochran, J. N., Brothers, K. B., East, K. M., Gray, D. E., Kelley, W. V., Lamb, N. E., Lose, E. J., Rich, C. A., Simmons, S., Whittle, J. S., Weaver, B. T., Nesmith, A. S., Myers, R. M., … Cooper, G. M. (2017). Genomic diagnosis for children with intellectual disability and/or developmental delay. *Genome Medicine*, *9*(1), 1–11. https://doi.org/10.1186/S13073-017-0433-1/TABLES/3

Boyle, C. A., Boulet, S., Schieve, L. A., Cohen, R. A., Blumberg, S. J., Yeargin-Allsopp, M., Visser, S., & Kogan, M. D. (2011). Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*, *127*(6), 1034–1042.

Bromberg, Y., Kahn, P. C., & Rost, B. (2013). Neutral and weakly nonneutral sequence variants may define individuality. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(35), 14255–14260. https://doi.org/10.1073/PNAS.1216613110/SUPPL_FILE/SD01.XLSX

Brudvig, J. J., & Weimer, J. M. (2022). CLN7 gene therapy: hope for an ultra-rare condition. *The Journal of Clinical Investigation*, *132*(5).

Brunet, T., Jech, R., Brugger, M., Kovacs, R., Alhaddad, B., Leszinski, G., Riedhammer, K. M., Westphal, D. S., Mahle, I., Mayerhanser, K., Skorvanek, M., Weber, S., Graf, E., Berutti, R., Necpál, J., Havránková, P., Pavelekova, P., Hempel, M., Kotzaeridou, U., … Wagner, M. (2021). De novo variants in neurodevelopmental disorders—experiences from a tertiary care center. *Clinical Genetics*, *100*(1), 14–28. https://doi.org/10.1111/CGE.13946

Caetano-Anolles, D. (2023). *Variant Quality Score Recalibration (VQSR) – GATK*. https://gatk.broadinstitute.org/hc/en-us/articles/360035531612-Variant-Quality-Score-Recalibration-VQSR-

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, *9*, 403–433. https://doi.org/10.1146/ANNUREV.GENOM.9.081307.164258

Cantu, M., Morrison, M. A., & Gagan, J. (2022). Standardized Comparison of Different DNA Sequencing Platforms. *Clinical Chemistry*, *68*(7), 872–876. https://doi.org/10.1093/CLINCHEM/HVAC023

Carson, A. R., Smith, E. N., Matsui, H., Brækkan, S. K., Jepsen, K., Hansen, J. B., & Frazer, K. A. (2014). Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics*, *15*(1), 125. https://doi.org/10.1186/1471-2105-15-125

Carvalho, C. M. B., & Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews Genetics 2016 17:4*, *17*(4), 224–238. https://doi.org/10.1038/nrg.2015.25

Chahrour, M., Sung, Y. J., Shaw, C., Zhou, X., Wong, S. T. C., Qin, J., & Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, *320*(5880), 1224–1229. https://doi.org/10.1126/SCIENCE.1153252/SUPPL_FILE/CHAHROUR_SOM.PDF

Charng, W. L., Karaca, E., Coban Akdemir, Z., Gambin, T., Atik, M. M., Gu, S., Posey, J. E., Jhangiani, S. N., Muzny, D. M., Doddapaneni, H., Hu, J., Boerwinkle, E., Gibbs, R. A., Rosenfeld, J. A., Cui, H., Xia, F., Manickam, K., Yang, Y., Faqeih, E. A., … Lupski, J. R. (2016). Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Medical Genomics*, *9*(1). https://doi.org/10.1186/S12920-016-0208-3

Chen, C. P., Chang, T. Y., Chen, C. Y., Wang, T. Y., Tsai, F. J., Wu, P. C., Chern, S. R., & Wang, W. (2012). Short rib-polydactyly syndrome type II (Majewski): prenatal diagnosis, perinatal imaging findings and molecular analysis of the NEK1 gene. *Taiwanese Journal of Obstetrics & Gynecology*, *51*(1), 100–105. https://doi.org/10.1016/J.TJOG.2012.01.020

Chinnery, P. F. (2021). Shortening the diagnostic odyssey-the impact of whole genome sequencing in the NHS. *BMJ (Clinical Research Ed.)*, *375*. https://doi.org/10.1136/BMJ.N2683

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

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y. J., Ghoorah, A. W., Dareng, E., Odia, T., Falola, O., Adebiyi, E., Hazelhurst, S., Mazandu, G., Nyangiri, O. A., Mbiyavanga, M., … Hanchard, N. A. (2020). High-depth African genomes inform human migration and health. *Nature*, *586*(7831), 741–748. https://doi.org/10.1038/S41586-020-2859-7

Christianson, A. L., Zwane, M. E., Manga, P., Rosen, E., Venter, A., Downs, D., & Kromberg, J. G. R. (2002). Children with intellectual disability in rural South Africa: prevalence and associated disability. *Journal of Intellectual Disability Research*, *46*(2), 179–186.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80. https://doi.org/10.4161/FLY.19695

Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., & Kingsmore, S. F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Medicine*, *3*(1), 1–10.

Clifford, S., Dissanayake, C., Bui, Q. M., Huggins, R., Taylor, A. K., & Loesch, D. Z. (2007). Autism spectrum phenotype in males and females with fragile X full mutation and premutation. *Journal of Autism and Developmental Disorders*, *37*(4), 738–747. https://doi.org/10.1007/S10803-006-0205-Z/METRICS

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767. https://doi.org/10.1093/NAR/GKP1137

Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., Dougherty, J. D., Lein, E. S., Hormozdiari, F., & Bernier, R. A. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature Genetics*, *51*(1), 106–116.

Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorski, J., … Eichler, E. E. (2011). A copy number variation morbidity map of developmental delay. *Nature Genetics 2011 43:9*, *43*(9), 838–846. https://doi.org/10.1038/ng.909

Cottino, L., Sahibdeen, V., Mudau, M., Lekgate, N., & Krause, A. (2022). QF-PCR: a valuable first-line prenatal and postnatal test for common aneuploidies in South Africa. *Journal of Community Genetics*, 1–9.

Cui, C., Shu, W., & Li, P. (2016). Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Frontiers in Cell and Developmental Biology*, *4*(SEP), 89. https://doi.org/10.3389/FCELL.2016.00089/BIBTEX

Culbertson, M. R. (1999). RNA surveillance: unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends in Genetics*, *15*(2), 74–80. https://doi.org/10.1016/S0168-9525(98)01658-8

Dabrowski, M., Bukowy-Bieryllo, Z., & Zietkiewicz, E. (2018). Advances in therapeutic use of a drug-stimulated translational readthrough of premature termination codons. *Molecular Medicine 2018 24:1*, *24*(1), 1–15. https://doi.org/10.1186/S10020-018-0024-7

Danecek, P. (2020). *BCFtools/trio-dnm*.

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*, *6*(12). https://doi.org/10.1371/JOURNAL.PCBI.1001025

de Menil, V., Hoogenhout, M., Kipkemoi, P., Kamuya, D., Eastman, E., Galvin, A., Mwangasha, K., de Vries, J., Kariuki, S. M., Murugasen, S., Mwangi, P., Singh, I., Stein, D. J., Abubakar, A., Newton, C. R., Donald, K. A., & Robinson, E. (2019). The NeuroDev Study: Phenotypic and Genetic Characterization of Neurodevelopmental Disorders in Kenya and South Africa. *Neuron*, *101*(1), 15–19. https://doi.org/10.1016/J.NEURON.2018.12.016

de Villiers, K. (2021). Bridging the health inequality gap: an examination of South Africa's social innovation in health landscape. *Infectious Diseases of Poverty*, *10*(1), 1–7. https://doi.org/10.1186/S40249-021-00804-9/METRICS

Deciphering Developmental Disorders (DDD) project. (2020). *Deciphering Developmental Disorders (DDD) project - Project Updates*. https://www.ddduk.org/updates.html

Deciphering Developmental Disorders Study. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, *542*(7642), 433.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., & Hanna, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498.

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319.

Djotsa Nono, A., Chen, K., & Liu, X. (2018). *Comparison of different functional prediction scores using a gene-based permutation model for identifying cancer driver genes*. https://doi.org/10.1186/s12920-018-0452-9

Dórea, J. G. (2019). Environmental exposure to low-level lead (Pb) co-occurring with other neurotoxicants in early life and neurodevelopment of children. *Environmental Research*, *177*, 108641.

Douzgou, S., Janssens, V., & Houge, G. (2022). PPP2R1A-Related Neurodevelopmental Disorder. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK580243/

Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., Carnevali, P., Nazarenko, I., Nilsen, G. B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K. P., Baccash, J., Borcherding, A. P., Brownley, A., Cedeno, R., Chen, L., … Reid, C. A. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science (New York, N.Y.)*, *327*(5961), 78–81. https://doi.org/10.1126/SCIENCE.1181498

Du Plessis, J. A., Pretorius, P. J., Lippert, M. M., Knoll, D. P., Erasmus, E., van der Westhuizen, F. H., & Mienie, L. J. (2001). An atypical carbohydrate deficient glycoprotein (CDG) syndrome patient in South Africa. *South African Medical Journal*, *91*(5), 392–394.

Dubuc, A. M., Davids, M. S., Pulluqi, M., Pulluqi, O., Hoang, K., Hernandez-Sánchez, J. M., Schlich, C., Hernández-Rivas, J. M., Brown, J. R., & Dal Cin, P. (2016). FISHing in the dark: How the combination of FISH and conventional karyotyping improves the diagnostic yield in CpG-stimulated chronic lymphocytic leukemia. *American Journal of Hematology*, *91*(10), 978–983. https://doi.org/10.1002/AJH.24452

Dutta, A. K. (2019). Schuurs-Hoeijmakers syndrome in a patient from India. *American Journal of Medical Genetics Part A*, *179*(4), 522–524.

Ebbert, M. T. W., Wadsworth, M. E., Staley, L. A., Hoyt, K. L., Pickett, B., Miller, J., Duce, J., Kauwe, J. S. K., & Ridge, P. G. (2016). Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics*, *17*(7), 491–500. https://doi.org/10.1186/S12859-016-1097-3/FIGURES/6

Element Biosciences AVITI: (2022). *Element Biosciences AVITI: the new kid on the NGS block - Enseqlopedia*. https://enseqlopedia.com/2022/03/element-biosciences-aviti-the-new-kid-on-the-ngs-block/

Elizondo-Plazas, A., Ibarra-Ramírez, M., Garza-Báez, A., & Martínez-de-Villarreal, L. E. (2020). Expanding the phenotype of MTOR-related disorders and the Smith-Kingsmore syndrome. *Neurology Genetics*, *6*(3).

Ellis, R. J., Calero, P., & Stockin, M. D. (2009). HIV Infection and the Central Nervous System: A Primer. *Neuropsychology Review*, *19*(2), 144. https://doi.org/10.1007/S11065-009-9094-1

Emerson, E., Yasamy, M. T., & Saxena, S. (2012). Scaling up support for children with developmental disabilities in low- and middle-income countries. *Journal of Applied Research in Intellectual Disabilities*, *25*(2), 96–98. https://doi.org/10.1111/J.1468-3148.2011.00677.X

Erwin, D. J., LaMaire, C., Espana, A., Eble, T. N., & Dhar, S. U. (2020). Financial barriers in a county genetics clinic: Problems and solutions. *Journal of Genetic Counseling*, *29*(4), 678–688. https://doi.org/10.1002/JGC4.1279

Ewans, L. J., Minoche, A. E., Schofield, D., Shrestha, R., Puttick, C., Zhu, Y., Drew, A., Gayevskiy, V., Elakis, G., Walsh, C., Adès, L. C., Colley, A., Ellaway, C., Evans, C. A., Freckmann, M. L., Goodwin, L., Hackett, A., Kamien, B., Kirk, E. P., … Roscioli, T. (2022). Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. *European Journal of Human Genetics 2022 30:10*, *30*(10), 1121–1131. https://doi.org/10.1038/s41431-022-01162-2

Fadista, J., Oskolkov, N., Hansson, O., & Groop, L. (2017). LoFtool: a gene intolerance score based on loss-of-function variants in 60 706 individuals. *Bioinformatics*, *33*(4), 471–474.

Fang, H., Wu, Y., Narzisi, G., O'Rawe, J. A., Barrón, L. T. J., Rosenbaum, J., Ronemus, M., Iossifov, I., Schatz, M. C., & Lyon, G. J. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Medicine*, *6*(10), 1–17. https://doi.org/10.1186/S13073-014-0089-Z/FIGURES/12

Federici, G., & Soddu, S. (2020). Variants of uncertain significance in the era of high-throughput genome sequencing: A lesson from breast and ovary cancers. *Journal of Experimental and Clinical Cancer Research*, *39*(1), 1–12. https://doi.org/10.1186/S13046-020-01554-6/TABLES/6

Félix, T. M., Fischinger Moura de Souza, C., Oliveira, J. B., Rico-Restrepo, M., Zanoteli, E., Zatz, M., & Giugliani, R. (2023). Challenges and recommendations to increasing the use of exome sequencing and whole genome sequencing for diagnosing rare diseases in Brazil: an expert perspective. *International Journal for Equity in Health*, *22*(1), 1–10. https://doi.org/10.1186/S12939-022-01809-Y/FIGURES/1

Ferguson, G., & Jelsma, J. (2009). The prevalence of motor delay among HIV infected children living in Cape Town, South Africa. *International Journal of Rehabilitation Research*, *32*(2), 108–114. https://doi.org/10.1097/MRR.0B013E3283013B34

Fieggen, K. J., Lambie, L. A., & Donald, K. A. (2019). Investigating developmental delay in south africa: A pragmatic approach. *South African Medical Journal*, *109*(4), 210–213. https://doi.org/10.7196/SAMJ.2019.v109i4.13991

Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., Vooren, S. Van, Moreau, Y., Pettett, R. M., & Carter, N. P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics*, *84*(4), 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010

Fitzpatrick, J. P., Latimer, J., Olson, H. C., Carter, M., Oscar, J., Lucas, B. R., Doney, R., Salter, C., Try, J., Hawkes, G., Fitzpatrick, E., Hand, M., Watkins, R. E., Tsang, T. W., Bower, C., Ferreira, M. L., Boulton, J., & Elliott, E. J. (2017). Prevalence and profile of Neurodevelopment and Fetal Alcohol Spectrum Disorder (FASD) amongst Australian Aboriginal children living in remote communities. *Research in Developmental Disabilities*, *65*, 114–126. https://doi.org/10.1016/J.RIDD.2017.04.001

Flore, L. A., & Milunsky, J. M. (2012). Updates in the Genetic Evaluation of the Child with Global Developmental Delay or Intellectual Disability. *Seminars in Pediatric Neurology*, *19*(4), 173–180. https://doi.org/10.1016/J.SPEN.2012.09.004

Flynn, K., Feben, C., Lamola, L., Carstens, N., Krause, A., Lombard, Z., & Consortium, D. as members of the H. (2021). Ending a diagnostic odyssey—The first case of Takenouchi–Kosaki syndrome in an African patient. *Clinical Case Reports*, *9*(4), 2144–2148.

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, *26*(20), 5369–5406. https://doi.org/10.1111/MEC.14264

Fung, J. L. F., Mullin, H. C., Huang, S., Chung, C. C. Y., Chan, M. C. Y., Pajusalu, S., Mak, C. C. Y., Hui, V. C. C., Tsang, M. H. Y., & San Yeung, K. (2020). A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *NPJ Genomic Medicine*, *5*(1), 1–11.

González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *The American Journal of Human Genetics*, *88*(4), 440–449.

Gordo, G., Tenorio, J., Arias, P., Santos-Simarro, F., García-Miñaur, S., Moreno, J. C., Nevado, J., Vallespin, E., Rodriguez-Laguna, L., & de Mena, R. (2018). mTOR mutations in Smith-Kingsmore syndrome: Four additional patients and a review. *Clinical Genetics*, *93*(4), 762–775.

Goyal, A., Kwon, H. J., Lee, K., Garg, R., Yun, S. Y., Kim, Y. H., Lee, S., Lee, M. S., Goyal, A., Kwon, H. J., Lee, K., Garg, R., Yun, S. Y., Kim, Y. H., Lee, S., & Lee, M. S. (2017). Ultra-Fast Next Generation Human Genome Sequencing Data Processing Using DRAGENTM Bio-IT Processor for Precision Medicine. *Open Journal of Genetics*, *7*(1), 9–19. https://doi.org/10.4236/OJGEN.2017.71002

Grier, J., Hirano, M., Karaa, A., Shepard, E., & Thompson, J. L. P. (2018). Diagnostic odyssey of patients with mitochondrial disease: Results of a survey. *Neurology. Genetics*, *4*(2). https://doi.org/10.1212/NXG.0000000000000230

Gubbels, C. S., VanNoy, G. E., Madden, J. A., Copenheaver, D., Yang, S., Wojcik, M. H., Gold, N. B., Genetti, C. A., Stoler, J., Parad, R. B., Roumiantsev, S., Bodamer, O., Beggs, A. H., Juusola, J., Agrawal, P. B., & Yu, T. W. (2020). Prospective, phenotype-driven selection of critically ill neonates for rapid exome sequencing is associated with high diagnostic yield. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *22*(4), 736–744. https://doi.org/10.1038/S41436-019-0708-6

Gudmundsson, S., Singer-Berk, M., Watts, N. A., Phu, W., Goodrich, J. K., Solomonson, M., Rehm, H. L., MacArthur, D. G., & O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*, *43*(8), 1012. https://doi.org/10.1002/HUMU.24309

Gunning, A. C., Fryer, V., Fasham, J., Crosby, A. H., Ellard, S., Baple, E. L., & Wright, C. F. (2020). Assessing performance of pathogenicity predictors using clinically relevant variant datasets Diagnostics. *J Med Genet*, *0*, 1–9. https://doi.org/10.1136/jmedgenet-2020-107003

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M. O., Choudhury, A., Ritchie, G. R. S., Xue, Y., Asimit, J., Nsubuga, R. N., Young, E. H., Pomilla, C., Kivinen, K., Rockett, K., Kamali, A., … Sandhu, M. S. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature*, *517*(7534), 327–332. https://doi.org/10.1038/NATURE13997

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, *38*(12), e131–e131. https://doi.org/10.1093/NAR/GKQ224

Harris, L., Zalucki, O., Gobius, I., McDonald, H., Osinki, J., Harvey, T. J., Essebier, A., Vidovic, D., Gladwyn-Ng, I., & Burne, T. H. (2016). Transcriptional regulation of intermediate progenitor cell generation during hippocampal development. *Development*, *143*(24), 4620–4630.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/J.YGENO.2015.11.003

Hefner, M. A., & Fassi, E. (2017). Genetic counseling in CHARGE syndrome: Diagnostic evaluation through follow up. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *175*(4), 407–416. https://doi.org/10.1002/AJMG.C.31589

Helbig, K. L., Farwell Hagman, K. D., Shinde, D. N., Mroske, C., Powis, Z., Li, S., Tang, S., & Helbig, I. (2016). Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genetics in Medicine 2016 18:9*, *18*(9), 898–905. https://doi.org/10.1038/gim.2015.186

Hiraide, T., Yamoto, K., Masunaga, Y., Asahina, M., Endoh, Y., Ohkubo, Y., Matsubayashi, T., Tsurui, S., Yamada, H., Yanagi, K., Nakashima, M., Hirano, K., Sugimura, H., Fukuda, T., Ogata, T., & Saitsu, H. (2021). Genetic and phenotypic analysis of 101 patients with developmental delay or intellectual disability using whole-exome sequencing. *Clinical Genetics*, *100*(1), 40–50. https://doi.org/10.1111/CGE.13951

Hoischen, A., Van Bon, B. W. M., Gilissen, C., Arts, P., Van Lier, B., Steehouwer, M., De Vries, P., De Reuver, R., Wieskamp, N., Mortier, G., Devriendt, K., Amorim, M. Z., Revencu, N., Kidd, A., Barbosa, M., Turner, A., Smith, J., Oley, C., Henderson, A., … Veltman, J. A. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature Genetics 2010 42:6*, *42*(6), 483–485. https://doi.org/10.1038/ng.581

Hood, L. (2008). A Personal Journey of Discovery: Developing Technology and Changing Biology. *Https://Doi.Org/10.1146/Annurev.Anchem.1.031207.113113*, *1*, 1–43. https://doi.org/10.1146/ANNUREV.ANCHEM.1.031207.113113

Hood, L., & Rowen, L. (2013). The human genome project: Big science transforms biology and medicine. *Genome Medicine*, *5*(9), 1–8. https://doi.org/10.1186/GM483/METRICS

Hori, K., Nagai, T., Shan, W., Sakamoto, A., Taya, S., Hashimoto, R., Hayashi, T., Abe, M., Yamazaki, M., & Nakao, K. (2014). Cytoskeletal regulation by AUTS2 in neuronal migration and neuritogenesis. *Cell Reports*, *9*(6), 2166–2179.

Hori, K., Yamashiro, K., Nagai, T., Shan, W., Egusa, S. F., Shimaoka, K., Kuniishi, H., Sekiguchi, M., Go, Y., & Tatsumoto, S. (2020). AUTS2 regulation of synapses for proper synaptic inputs and social communication. *Iscience*, *23*(6).

Howard, J., & Hyman, A. A. (2003). Dynamics and mechanics of the microtubule plus end. *Nature*, *422*(6933), 753–758.

Hua, H., Kong, Q., Zhang, H., Wang, J., Luo, T., & Jiang, Y. (2019). Targeting mTOR for cancer therapy. *Journal of Hematology & Oncology 2019 12:1*, *12*(1), 1–19. https://doi.org/10.1186/S13045-019-0754-1

Huang, J., Liang, X., Xuan, Y., Geng, C., Li, Y., Lu, H., Qu, S., Mei, X., Chen, H., Yu, T., Sun, N., Rao, J., Wang, J., Zhang, W., Chen, Y., Liao, S., Jiang, H., Liu, X., Yang, Z., … Gao, S. (2017). A reference human genome dataset of the BGISEQ-500 sequencer. *GigaScience*, *6*(5), 1–9. https://doi.org/10.1093/GIGASCIENCE/GIX024

Huang, P. J., Chang, J. H., Lin, H. H., Li, Y. X., Lee, C. C., Su, C. T., Li, Y. L., Chang, M. T., Weng, S., Cheng, W. H., Chiu, C. H., & Tang, P. (2020). DeepVariant-on-Spark: Small-Scale Genome Analysis Using a Cloud-Based Computing Framework. *Computational and Mathematical Methods in Medicine*, *2020*. https://doi.org/10.1155/2020/7231205

Illumina. (2011). *Quality Scores for Next-Generation Sequencing*. http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/

Inaba, Y., Schwartz, C. E., Bui, Q. M., Li, X., Skinner, C., Field, M., Wotton, T., Hagerman, R. J., Francis, D., Amor, D. J., Hopper, J. L., Loesch, D. Z., Bretherton, L., Slater, H. R., & Godler, D. E. (2014). Early Detection of Fragile X Syndrome: Applications of a Novel Approach for Improved Quantitative Methylation Analysis in Venous Blood and Newborn Blood Spots. *Clinical Chemistry*, *60*(7), 963–973. https://doi.org/10.1373/CLINCHEM.2013.217331

Ingemi, A. I., Bota, V. M., Peguero, A., & Charpentier, M. (2012). Fanconi's syndrome and nephrogenic diabetes insipidus in an adult treated with ifosfamide. *Pharmacotherapy*, *32*(1). https://doi.org/10.1002/PHAR.1013

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., & Karyadi, D. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, *99*(4), 877–885.

Iwamoto, O., Koga, C., Matsumoto, H., Terasaki, S., Kusukawa, J., & Kameyama, T. (2021). Coffin-Siris Syndrome. *Asian Journal of Oral and Maxillofacial Surgery*, *15*(3), 205–207. https://doi.org/10.1016/S0915-6992(03)80044-4

Jacobs, P. (1987). Chromosome abnormalities origin and etiology in abortions and livebirths. *Human Genetics*.

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., Bernstein, J. A., & Bejerano, G. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature Genetics*, *48*(12), 1581–1586.

Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., & Schwartz, G. B. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, *176*(3), 535–548.

Janssen, N., Bergman, J. E. H., Swertz, M. A., Tranebjaerg, L., Lodahl, M., Schoots, J., Hofstra, R. M. W., Van Ravenswaaij-Arts, C. M. A., & Hoefsloot, L. H. (2012). Mutation update on the CHD7 gene involved in CHARGE syndrome. *Human Mutation*, *33*(8), 1149–1160. https://doi.org/10.1002/HUMU.22086

Janssens, V., & Goris, J. (2001). Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochemical Journal*, *353*(3), 417–439.

Järvelä, I., Määttä, T., Acharya, A., Leppälä, J., Jhangiani, S. N., Arvio, M., Siren, A., Kankuri-Tammilehto, M., Kokkonen, H., Palomäki, M., Varilo, T., Fang, M., Hadley, T. D., Jolly, A., Linnankivi, T., Paetau, R., Saarela, A., Kälviäinen, R., Olme, J., … Schrauwen, I. (2021). Exome sequencing reveals predominantly de novo variants in disorders with intellectual disability (ID) in the founder population of Finland. *Human Genetics*, *140*(7), 1011–1029. https://doi.org/10.1007/S00439-021-02268-1/TABLES/4

Jeon, S. A., Park, J. L., Park, S. J., Kim, J. H., Goh, S. H., Han, J. Y., & Kim, S. Y. (2021). Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. *Genes & Genomics*, *43*(7), 713–724. https://doi.org/10.1007/S13258-021-01096-X

Ji, J., Shen, L., Bootwalla, M., Quindipan, C., Tatarinova, T., Maglinte, D. T., Buckley, J., Raca, G., Saitta, S. C., Biegel, J. A., & Gai, X. (2019). A semiautomated whole-exome sequencing workflow leads to increased diagnostic yield and identification of novel candidate variants. *Cold Spring Harbor Molecular Case Studies*, *5*(2). https://doi.org/10.1101/MCS.A003756

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Research*, *42*(22), 13534–13544.

Jimenez-Gomez, A., & Standridge, S. M. (2014). A refined approach to evaluating global developmental delay for the international medical community. In *Pediatric Neurology* (Vol. 51, Issue 2, pp. 198–206). Elsevier Inc. https://doi.org/10.1016/j.pediatrneurol.2013.12.018

Jin, P., & Warren, S. T. (2000). Understanding the molecular basis of fragile X syndrome. *Human Molecular Genetics*, *9*(6), 901–908. https://doi.org/10.1093/HMG/9.6.901

Jyonouchi, S., McDonald-McGinn, D. M., Bale, S., Zackai, E. H., & Sullivan, K. E. (2009). CHARGE Syndrome and Chromosome 22q11.2 Deletion Syndrome: A Comparison of Immunologic and Non-Immunologic Phenotypic Features. *Pediatrics*, *123*(5), e871. https://doi.org/10.1542/PEDS.2008-3400

Kamp, M., Krause, A., & Ramsay, M. (2021). Has translational genomics come of age in Africa? *Human Molecular Genetics*, *30*(R2), R164–R173.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … Daly, M. J. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature 2020 581:7809*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Karina Lucio-Eterovic, A., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C., & Carpenter, P. B. (2010). *Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function*. https://doi.org/10.1073/pnas.1002653107

Kengne Kamga, K., Nguefack, S., Minka, K., Wonkam Tingang, E., Esterhuizen, A., Nchangwi Munung, S., De Vries, J., & Wonkam, A. (2020). Cascade testing for fragile X syndrome in a rural setting in Cameroon (sub-Saharan Africa). *Genes*, *11*(2), 136.

Kessi, M., Xiong, J., Wu, L., Yang, L., He, F., Chen, C., Pang, N., Duan, H., Zhang, W., Arafat, A., Yin, F., & Peng, J. (2018). Rare Copy Number Variations and Predictors in Children With Intellectual Disability and Epilepsy. *Frontiers in Neurology*, *9*(NOV), 947. https://doi.org/10.3389/FNEUR.2018.00947/BIBTEX

Khaikin, Y., & Mercimek-Mahmutoglu, S. (2016). STXBP1 Encephalopathy with Epilepsy. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK396561/

Klein, C. J., & Foroud, T. M. (2017). Neurology Individualized Medicine: When to Use Next-Generation Sequencing Panels. *Mayo Clinic Proceedings*, *92*(2), 292–305. https://doi.org/10.1016/J.MAYOCP.2016.09.008

Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine 2020 12:1*, *12*(1), 1–13. https://doi.org/10.1186/S13073-020-00791-W

Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M. A. W., Kleefstra, T., Kramer, J. M., Webber, C., Huynen, M. A., & Schenck, A. (2016). Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *The American Journal of Human Genetics*, *98*(1), 149–164. https://doi.org/10.1016/J.AJHG.2015.11.024

Köhler, S., Gargano, M., Matentzoglu, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., Danis, D., Balagura, G., Baynam, G., Brower, A. M., Callahan, T. J., Chute, C. G., Est, J. L., Galer, P. D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M.,

… Robinson, P. N. (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, *49*(D1), D1207–D1217. https://doi.org/10.1093/NAR/GKAA1043

Koolen, D. A., Nillesen, W. M., Versteeg, M. H. A., Merkx, G. F. M., Knoers, N. V. A. M., Kets, M., Vermeer, S., Van Ravenswaaij, C. M. A., De Kovel, C. G., Brunner, H. G., Smeets, D., De Vries, B. B. A., & Sistermans, E. A. (2004). Screening for subtelomeric rearrangements in 210 patients with unexplained mental retardation using multiplex ligation dependent probe amplification (MLPA). *Journal of Medical Genetics*, *41*(12), 892–899. https://doi.org/10.1136/JMG.2004.023671

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Aguilera, M. A., Meyer, R., & Massouras, A. (2019). VarSome: the human genomic variant search engine. *Bioinformatics*, *35*(11), 1978.

Kosmicki, J. A., Samocha, K. E., Howrigan, D. P., Sanders, S. J., Slowikowski, K., Lek, M., Karczewski, K. J., Cutler, D. J., Devlin, B., Roeder, K., Buxbaum, J. D., Neale, B. M., Macarthur, D. G., Wall, D. P., Robinson, E. B., & Daly, M. J. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nature Genetics 2017 49:4*, *49*(4), 504–510. https://doi.org/10.1038/ng.3789

Krause, A. (2019). New genetic testing technologies: Advantages and limitations. *SAMJ: South African Medical Journal*, *109*(4), 207–209. https://doi.org/10.7196/SAMJ.2019.V109I4.13990

Kromberg, J. G., Sizer, E. B., & Christianson, A. L. (2013). Genetic services and testing in South Africa. *Journal of Community Genetics*, *4*(3), 413. https://doi.org/10.1007/S12687-012-0101-5

Krusche, P., Trigg, L., Boutros, P. C., Mason, C. E., De La Vega, F. M., Moore, B. L., Gonzalez-Porta, M., Eberle, M. A., Tezak, Z., Lababidi, S., Truty, R., Asimenos, G., Funke, B., Fleharty, M., Chapman, B. A., Salit, M., & Zook, J. M. (2019). Best practices for benchmarking germline small-variant calls in human genomes. *Nature Biotechnology 2019 37:5*, *37*(5), 555–560. https://doi.org/10.1038/s41587-019-0054-x

Kumar, A. M., Borodowsky, I., Fernandez, B., Gonzalez, L., & Kumar, M. (2007). Human immunodeficiency virus type 1 RNA Levels in different regions of human brain: quantification using real-time reverse transcriptase-polymerase chain reaction. *Journal of Neurovirology*, *13*(3), 210–224.

Kumar, K. R., Cowley, M. J., & Davis, R. L. (2019). Next-Generation Sequencing and Emerging Technologies. *Seminars in Thrombosis and Hemostasis*, *45*(7), 661–673. https://doi.org/10.1055/S-0039-1688446

Kurotaki, N., Harada, N., Yoshiura, K. ichiro, Sugano, S., Niikawa, N., & Matsumoto, N. (2001). Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene. *Gene*, *279*(2), 197–204. https://doi.org/10.1016/S0378-1119(01)00750-8

Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017). *Singularity: Scientific containers for mobility of compute*. https://doi.org/10.1371/journal.pone.0177459

Lakhan, R. (2015). Profile of Social, Environmental and Biological Correlates in Intellectual Disability in A Resource-Poor Setting in India. *Indian Journal of Psychological Medicine*, *37*(3), 311. https://doi.org/10.4103/0253-7176.162957

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., & Jang, W. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067.

Lange, S., Probst, C., Gmel, G., Rehm, J., Burd, L., & Popova, S. (2017). Global prevalence of fetal alcohol spectrum disorder among children and youth: a systematic review and meta-analysis. *JAMA Pediatrics*, *171*(10), 948–956.

Langer-Safer, P. R., Levine, M., & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences*, *79*(14), 4381–4385. https://doi.org/10.1073/PNAS.79.14.4381

Laurent Francioli, Grace Tiao, Konrad Karczewski, Matthew Solomonson, & Nick Watts. (2018, October 18). *gnomAD v2.1*. https://gnomad.broadinstitute.org/news/2018-10-gnomad-v2-1/

Lavelle, T. A., Feng, X., Keisler, M., Cohen, J. T., Neumann, P. J., Prichard, D., Schroeder, B. E., Salyakina, D., Espinal, P. S., Weidner, S. B., & Maron, J. L. (2022). Cost-effectiveness of exome and genome sequencing for children with rare and undiagnosed conditions. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *24*(6), 1349–1361. https://doi.org/10.1016/J.GIM.2022.03.005

Lay-Son, G., Espinoza, K., Vial, C., Rivera, J. C., Guzmán, M. L., & Repetto, G. M. (2015). Chromosomal microarrays testing in children with developmental disabilities and congenital anomalies. *Jornal de Pediatria*, *91*(2), 189–195. https://doi.org/10.1016/J.JPED.2014.07.003

Lebeko, K., Bosch, J., Noubiap, J. J. N., Dandara, C., & Wonkam, A. (2015). Genetics of hearing loss in africans: use of next generation sequencing is the best way forward. *PAMJ. 2015; 20:383*, *20*(383). https://doi.org/10.11604/PAMJ.2015.20.383.5230

Lee, J. A., & Lupski, J. R. (2006). Genomic Rearrangements and Gene Copy-Number Alterations as a Cause of Nervous System Disorders. *Neuron*, *52*(1), 103–121. https://doi.org/10.1016/J.NEURON.2006.09.027

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., … Williams, A. L. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature 2016 536:7616*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A., & Gilissen, C. (2015). Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Human Mutation*, *36*(8), 815–822. https://doi.org/10.1002/HUMU.22813

Leme, D., Souza, D., Mercado, G., Pastene, E., Dias, A., & Moretti-Ferreira, D. (2013). Assessment of clinical scoring systems for the diagnosis of Williams-Beuren syndrome. *Genetics and Molecular Research*, *12*(3), 3407–3411. https://doi.org/10.4238/2013.September.4.7

Lenaerts, L., Reynhout, S., Verbinnen, I., Laumonnier, F., Toutain, A., Bonnet-Brilhault, F., Hoorne, Y., Joss, S., Chassevent, A. K., Smith-Hicks, C., Loeys, B., Joset, P., Steindl, K., Rauch, A., Mehta, S. G., Chung, W. K., Devriendt, K., Holder, S. E., Jewett, T., … Janssens, V. (2021). The broad phenotypic spectrum of PPP2R1A-related neurodevelopmental disorders correlates with the degree of biochemical dysfunction. *Genetics in Medicine*, *23*(2), 352–362. https://doi.org/10.1038/S41436-020-00981-2

Leventopoulos, G., Kitsiou-Tzeli, S., Kritikos, K., Psoni, S., Mavrou, A., Kanavakis, E., & Fryssira, H. (2009). A Clinical Study of Sotos Syndrome Patients With Review of the Literature. *Pediatric Neurology*, *40*(5), 357–364. https://doi.org/10.1016/j.pediatrneurol.2008.11.013

Li, C., Vandersluis, S., Holubowich, C., Ungar, W. J., Goh, E. S., Boycott, K. M., Sikich, N., Dhalla, I., & Ng, V. (2021). Cost-effectiveness of genome-wide sequencing for unexplained developmental disabilities and multiple congenital anomalies. *Genetics in Medicine*, *23*(3), 451–460. https://doi.org/10.1038/S41436-020-01012-W

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Li, Q., Zhao, X., Zhang, W., Wang, L., Wang, J., Xu, D., Mei, Z., Liu, Q., Du, S., Li, Z., Liang, X., Wang, X., Wei, H., Liu, P., Zou, J., Shen, H., Chen, A., Drmanac, S., Liu, J. S., … Jiang, Y. (2019). Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. *BMC Genomics*, *20*(1), 1–13. https://doi.org/10.1186/S12864-019-5569-5/TABLES/3

Li, Y., Sidore, C., Kang, H. M., Boehnke, M., & Abecasis, G. R. (2011). Low-coverage sequencing: Implications for design of complex trait association studies. *Genome Research*, *21*(6), 940. https://doi.org/10.1101/GR.117259.110

Liao, P., Satten, G. A., & Hu, Y. J. (2017). PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genetic Epidemiology*, *41*(5), 375–387. https://doi.org/10.1002/GEPI.22048

Lichter, J. B., Difilippantonio, M. J., Pakstis, A. J., Goodfellow, P. J., Ward, D. C., & Kidd, K. K. (1993). Physical and Genetic Maps for Chromosome 10. *Genomics*, *16*(2), 320–324. https://doi.org/10.1006/GENO.1993.1192

Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., Thiruvahindrapuram, B., Merico, D., Jobling, R., & Nalpathamkalam, T. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genetics in Medicine*, *20*(4), 435–443.

Liu, Q., Guo, Y., Li, J., Long, J., Zhang, B., & Shyr, Y. (2012). Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. *BMC Genomics*, *13*(Suppl 8), S8. https://doi.org/10.1186/1471-2164-13-S8-S8

Liu, X., Li, C., Mou, C., Dong, Y., & Tu, Y. (2020). dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Medicine*, *12*(1), 1–8. https://doi.org/10.1186/S13073-020-00803-9/FIGURES/4

Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation*, *37*(3), 235–241. https://doi.org/10.1002/HUMU.22932

Lord, J., & Baralle, D. (2021). Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Frontiers in Genetics*, *12*, 1146. https://doi.org/10.3389/FGENE.2021.689892/BIBTEX

Lumaka, A., Carstens, N., Devriendt, K., Krause, A., Kulohoma, B., Kumuthini, J., Mubungu, G., Mukisa, J., Nel, M., & Olanrewaju, T. O. (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 Journal of Rare Diseases*, *17*(1), 1–6.

Macchiaiolo, M., Panfili, F. M., Vecchio, D., Gonfiantini, M. V., Cortellessa, F., Caciolo, C., Zollino, M., Accadia, M., Seri, M., Chinali, M., Mammì, C., Tartaglia, M., Bartuli, A., Alfieri, P., & Priolo, M. (2022). A deep phenotyping experience: up to date in management and diagnosis of Malan syndrome in a single center surveillance report. *Orphanet Journal of Rare Diseases*, *17*(1), 1–16. https://doi.org/10.1186/S13023-022-02384-9/TABLES/4

Macklin, S. K., Jackson, J. L., Atwal, P. S., & Hines, S. L. (2018). Physician interpretation of variants of uncertain significance. *Familial Cancer 2018 18:1*, *18*(1), 121–126. https://doi.org/10.1007/S10689-018-0086-2

Maita, H., Kitaura, H., Ariga, H., & Iguchi-Ariga, S. M. M. (2005). CIR, a corepressor of CBF1, binds to PAP-1 and effects alternative splicing. *Experimental Cell Research*, *303*(2), 375–387.

Mandal, M., Maienschein-Cline, M., Maffucci, P., Veselits, M., Kennedy, D. E., McLean, K. C., Okoreeh, M. K., Karki, S., Cunningham-Rundles, C., & Clark, M. R. (2018).

BRWD1 orchestrates epigenetic landscape of late B lymphopoiesis. *Nature Communications*, *9*(1), 3888.

Manickam, K., McClain, M. R., Demmer, L. A., Biswas, S., Kearney, H. M., Malinowski, J., Massingham, L. J., Miller, D., Yu, T. W., & Hisama, F. M. (2021). Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine 2021 23:11*, *23*(11), 2029–2037. https://doi.org/10.1038/s41436-021-01242-6

Manning, M., & Hudgins, L. (2010). Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genetics in Medicine*, *12*(11), 742–745. https://doi.org/10.1097/GIM.0B013E3181F8BAAD

Manrai, A. K., Funke, B. H., Rehm, H. L., Olesen, M. S., Maron, B. A., Szolovits, P., Margulies, D. M., Loscalzo, J., & Kohane, I. S. (2016). Genetic Misdiagnoses and the Potential for Health Disparities. *New England Journal of Medicine*, *375*(7), 655–665. https://doi.org/10.1056/NEJMSA1507092/SUPPL_FILE/NEJMSA1507092_DISCLOSURES.PDF

Marszalek, J., & De Villiers, P. J. T. (2006). Morbidity profile of admissions to GF Jooste Hospital, Manenberg, Cape Town. *South African Family Practice*, *48*(6), 15–15e.

Martinez, F., Marín-Reina, P., Sanchis-Calvo, A., Perez-Aytés, A., Oltra, S., Roselló, M., Mayo, S., Monfort, S., Pantoja, J., & Orellana, C. (2015). Novel mutations of NFIX gene causing Marshall-Smith syndrome or Sotos-like syndrome: one gene, two phenotypes. *Pediatric Research 2015 78:5*, *78*(5), 533–539. https://doi.org/10.1038/pr.2015.135

Martins, E. M., Vilarinho, L., Esteves, S., Lopes-Marques, M., Amorim, A., & Azevedo, L. (2011). Consequences of primer binding-sites polymorphisms on genotyping practice. *Open Journal of Genetics*, *201*(1), 15–17. https://doi.org/10.4236/ojgen.2011.12004

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities*, *32*(2), 419–436. https://doi.org/10.1016/J.RIDD.2010.12.018

May, P. A., Chambers, C. D., Kalberg, W. O., Zellner, J., Feldman, H., Buckley, D., Kopald, D., Hasken, J. M., Xu, R., & Honerkamp-Smith, G. (2018). Prevalence of fetal alcohol spectrum disorders in 4 US communities. *Jama*, *319*(5), 474–482.

May, P. A., de Vries, M. M., Marais, A., Kalberg, W. O., Buckley, D., Hasken, J. M., Abdul-Rahman, O., Robinson, L. K., Manning, M. A., & Seedat, S. (2022). The prevalence of fetal alcohol spectrum disorders in rural communities in South Africa: A third regional sample of child characteristics and maternal risk factors. *Alcoholism: Clinical and Experimental Research*, *46*(10), 1819–1836.

McCarthy, D. J., Humburg, P., Kanapin, A., Rivas, M. A., Gaulton, K., Cazier, J. B., & Donnelly, P. (2014). Choice of transcripts and software has a large effect on variant annotation. *Genome Medicine*, *6*(3), 1–16. https://doi.org/10.1186/GM543/TABLES/2

McCourt, C. M., McArt, D. G., Mills, K., Catherwood, M. A., Maxwell, P., Waugh, D. J., Hamilton, P., O'Sullivan, J. M., & Salto-Tellez, M. (2013). Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS One*, *8*(7), e69604.

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., Fawzi, W., & Fink, G. (2016). Early childhood developmental status in low-and middle-income countries: national, regional, and global prevalence estimates using predictive modeling. *PLoS Medicine*, *13*(6), e1002034.

Mckenzie, J. A., McConkey, R., & Adnams, C. (2013). Intellectual disability in Africa: implications for research and service development.

Https://Doi.Org/10.3109/09638288.2012.751461, *35*(20), 1750–1755. https://doi.org/10.3109/09638288.2012.751461
McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology*, *17*(1), 1–14.
McMahon, M., & Hatton, C. (2021). A comparison of the prevalence of health problems among adults with and without intellectual disability: A total administrative population study. *Journal of Applied Research in Intellectual Disabilities*, *34*(1), 316–325. https://doi.org/10.1111/JAR.12785
Mendola, P., Selevan, S. G., Gutter, S., & Rice, D. (2002). Environmental factors associated with a spectrum of neurodevelopmental deficits. *Mental Retardation and Developmental Disabilities Research Reviews*, *8*(3), 188–197.
Merla, G., Brunetti-Pierri, N., Micale, L., & Fusco, C. (2010). Copy number variants at Williams-Beuren syndrome 7q11.23 region. *Human Genetics*, *128*(1), 3–26. https://doi.org/10.1007/S00439-010-0827-2
Merriman, B., Torrent, I., & Rothberg, J. M. (2012). Progress in Ion Torrent semiconductor chip based sequencing. *ELECTROPHORESIS*, *33*(23), 3397–3417. https://doi.org/10.1002/ELPS.201200424
Metzker, M. L. (2009). Sequencing technologies — the next generation. *Nature Reviews Genetics 2010 11:1*, *11*(1), 31–46. https://doi.org/10.1038/nrg2626
Meynert, A. M., Bicknell, L. S., Hurles, M. E., Jackson, A. P., & Taylor, M. S. (2013). Quantifying single nucleotide variant detection sensitivity in exome sequencing. *BMC Bioinformatics*, *14*(1), 1–10. https://doi.org/10.1186/1471-2105-14-195/TABLES/1
Miclea, D., Peca, L., Cuzmici, Z., & Pop, I. V. (2015). Genetic testing in patients with global developmental delay / intellectual disabilities. A review. *Clujul Medical*, *88*(3), 288. https://doi.org/10.15386/CJMED-461
Miclea, D., Szucs, A., Mirea, A., Stefan, D. M., Nazarie, F., Bucerzan, S., Lazea, C., Grama, A., Pop, T. L., Farcas, M., Zaharie, G., Matyas, M., Snr, M. M., Vintan, M., Popp, R., & Alkhzouz, C. (2021). Diagnostic Usefulness of MLPA Techniques for Recurrent Copy Number Variants Detection in Global Developmental Delay/Intellectual Disability. *International Journal of General Medicine*, *14*, 4511–4515. https://doi.org/10.2147/IJGM.S320033
Milani, D. A. Q., & Tadi, P. (2022). Genetics, Chromosome Abnormalities. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK557691/
Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., Church, D. M., Crolla, J. A., Eichler, E. E., Epstein, C. J., Faucett, W. A., Feuk, L., Friedman, J. M., Hamosh, A., Jackson, L., Kaminsky, E. B., Kok, K., Krantz, I. D., Kuhn, R. M., … Ledbetter, D. H. (2010). Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *The American Journal of Human Genetics*, *86*(5), 749–764. https://doi.org/10.1016/J.AJHG.2010.04.006
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: 1215. *Find This Article Online*.
Minatogawa, M., Tsuji, T., Inaba, M., Kawakami, N., Mizuno, S., & Kosho, T. (2022). Atypical Sotos syndrome caused by a novel splice site variant. *Human Genome Variation 2022 9:1*, *9*(1), 1–4. https://doi.org/10.1038/s41439-022-00219-4
Mirzaa, G. M., Campbell, C. D., Solovieff, N., Goold, C. P., Jansen, L. A., Menon, S., Timms, A. E., Conti, V., Biag, J. D., & Olds, C. (2016). Association of MTOR mutations with developmental brain disorders, including megalencephaly, focal cortical dysplasia, and pigmentary mosaicism. *JAMA Neurology*, *73*(7), 836–845.

Moeschler, J. B., Shevell, M., Genetics, C. on, Moeschler, J. B., Shevell, M., Saul, R. A., Chen, E., Freedenberg, D. L., Hamid, R., & Jones, M. C. (2014). Comprehensive evaluation of the child with intellectual disability or global developmental delays. *Pediatrics*, *134*(3), e903–e918.

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., Köster, J., & Reich, M. (2021). Sustainable data analysis with Snakemake. *F1000Research 2021 10:33*, *10*, 33. https://doi.org/10.12688/f1000research.29032.1

Monderer-Rothkoff, G., Tal, N., Risman, M., Shani, O., Nissim-Rafinia, M., Malki-Feldman, L., Medvedeva, V., Groszer, M., Meshorer, E., & Shifman, S. (2021). AUTS2 isoforms control neuronal differentiation. *Molecular Psychiatry*, *26*(2), 666–681.

Mondragon, O. H., Bridges, P. G., Levy, S., Ferreira, K. B., & Widener, P. (2016). Understanding Performance Interference in Next-Generation HPC Systems. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, *0*, 384–395. https://doi.org/10.1109/SC.2016.32

Monroe, G. R., Frederix, G. W., Savelberg, S. M. C., De Vries, T. I., Duran, K. J., Van Der Smagt, J. J., Terhal, P. A., Van Hasselt, P. M., Kroes, H. Y., Verhoeven-Duif, N. M., Nijman, I. J., Carbo, E. C., Van Gassen, K. L., Knoers, N. V., Hövels, A. M., Van Haelst, M. M., Visser, G., & Van Haaften, G. (2016). Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. *Genetics in Medicine 2016 18:9*, *18*(9), 949–956. https://doi.org/10.1038/gim.2015.200

Moosa, S., Böhrer-Rabel, H., Altmüller, J., Beleggia, F., Nürnberg, P., Li, Y., Yigit, G., & Wollnik, B. (2017). Smith–Kingsmore syndrome: a third family with the MTOR mutation c. 5395G> A p.(Glu1799Lys) and evidence for paternal gonadal mosaicism. *American Journal of Medical Genetics Part A*, *173*(1), 264–267.

Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J. M., Couce, M. L., & Cocho, J. A. (2013). A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism*, *110*(1–2), 3–24. https://doi.org/10.1016/J.YMGME.2013.04.024

Morimura, T., Numata, Y., Nakamura, S., Hirano, E., Gotoh, L., Goto, Y., Urushitani, M., & Inoue, K. (2014). Attenuation of endoplasmic reticulum stress in Pelizaeus-Merzbacher disease by an anti-malaria drug, chloroquine. *Experimental Biology and Medicine*, *239*(4), 489–501.

Mørup, S. B., Nazaryan-Petersen, L., Gabrielaite, M., Reekie, J., Marquart, H. V., Hartling, H. J., Marvig, R. L., Katzenstein, T. L., Masmas, T. N., Lundgren, J., Murray, D. D., Helleberg, M., & Borgwardt, L. (2022). Added Value of Reanalysis of Whole Exome- and Whole Genome Sequencing Data From Patients Suspected of Primary Immune Deficiency Using an Extended Gene Panel and Structural Variation Calling. *Frontiers in Immunology*, *13*, 1. https://doi.org/10.3389/FIMMU.2022.906328/FULL

Mroske, C., Rasmussen, K., Shinde, D. N., Huether, R., Powis, Z., Lu, H.-M., Baxter, R. M., Mcpherson, E., & Tang, S. (2012). *Germline activating MTOR mutation arising through gonadal mosaicism in two brothers with megalencephaly and neurodevelopmental abnormalities*. https://doi.org/10.1186/s12881-015-0240-8

Mu, W., Lu, H.-M., Chen, J., Li, S., & Elliott, A. M. (2016). Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. *The Journal of Molecular Diagnostics*, *18*(6), 923–932.

Mubungu, G., Lukute, G., Makay, P., Songo, C., Lukusa, P., Devriendt, K., & Lumaka, A. (2020). Phenotype and growth in Sotos syndrome patient from DR Congo (Central Africa). *American Journal of Medical Genetics, Part A*, *182*(7), 1572–1575. https://doi.org/10.1002/AJMG.A.61617

Mulder, N., Abimiku, A., Adebamowo, S. N., de Vries, J., Matimba, A., Olowoyo, P., Ramsay, M., Skelton, M., & Stein, D. J. (2018). H3Africa: current perspectives. *Pharmacogenomics and Personalized Medicine*, *11*, 59. https://doi.org/10.2147/PGPM.S141546

Mulder, N. J., Adebiyi, E., Alami, R., Benkahla, A., Brandful, J., Doumbia, S., Everett, D., Fadlelmola, F. M., Gaboun, F., Gaseitsiwe, S., Ghazal, H., Hazelhurst, S., Hide, W., Ibrahimi, A., Fakim, Y. J., Victor Jongeneel, C., Joubert, F., Kassim, S., Kayondo, J., … Wele, M. (2016). H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Research*, *26*(2), 271–277. https://doi.org/10.1101/GR.196295.115

Mullaney, J. M., Mills, R. E., Stephen Pittard, W., & Devine, S. E. (2010). Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics*, *19*(R2), R131–R136. https://doi.org/10.1093/HMG/DDQ400

Munns, C. F., Fahiminiya, S., Poudel, N., Munteanu, M. C., Majewski, J., Sillence, D. O., Metcalf, J. P., Biggin, A., Glorieux, F., Fassier, F., Rauch, F., & Hinsdale, M. E. (2015). Homozygosity for Frameshift Mutations in XYLT2 Result in a Spondylo-Ocular Syndrome with Bone Fragility, Cataracts, and Hearing Defects. *American Journal of Human Genetics*, *96*(6), 971. https://doi.org/10.1016/J.AJHG.2015.04.017

Musich, R., Cadle-Davidson, L., & Osier, M. V. (2021). Comparison of Short-Read Sequence Aligners Indicates Strengths and Weaknesses for Biologists to Consider. *Frontiers in Plant Science*, *12*, 692. https://doi.org/10.3389/FPLS.2021.657240/BIBTEX

Ndugga-Kabuye, M. K., & Issaka, R. B. (2019). Inequities in multi-gene hereditary cancer testing: Lower diagnostic yield and higher VUS rate in individuals who identify as Hispanic, African or Asian and Pacific Islander as compared to European. *Familial Cancer*, *18*(4), 465. https://doi.org/10.1007/S10689-019-00144-6

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. https://doi.org/DOI: 10.1093/nar/gkg509

Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., & Rath, A. (2019). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics 2019 28:2*, *28*(2), 165–173. https://doi.org/10.1038/s41431-019-0508-0

NIH. (2021). *The Cost of Sequencing a Human Genome*. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

Niroula, A., & Vihinen, M. (2018). How good are pathogenicity predictors in detecting benign variants? *BioRxiv*, 408153. https://doi.org/10.1101/408153

Nolan, D., & Carlson, M. (2016). Whole Exome Sequencing in Pediatric Neurology Patients. *Http://Dx.Doi.Org/10.1177/0883073815627880*, *31*(7), 887–894. https://doi.org/10.1177/0883073815627880

Nuckols, C. C., & Nuckols, C. C. (2013). The diagnostic and statistical manual of mental disorders,(DSM-5). *Philadelphia: American Psychiatric Association*.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2021). The complete sequence of a human genome. *BioRxiv*, 2021.05.26.445798. https://doi.org/10.1101/2021.05.26.445798

Okamoto, N., Tsuchiya, Y., Kuki, I., Yamamoto, T., Saitsu, H., Kitagawa, D., & Matsumoto, N. (2017). Disturbed chromosome segregation and multipolar spindle formation in a patient with CHAMP 1 mutation. *Molecular Genetics & Genomic Medicine*, *5*(5), 585–591.

Olivier, L., Viljoen, D. L., & Curfs, L. M. G. (2016). Fetal alcohol spectrum disorders: prevalence rates in South Africa: the new millennium. *South African Medical Journal*, *106*(sup-1), 103–106.

O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., Bodily, P., Tian, L., Hakonarson, H., Johnson, W. E., Wei, Z., Wang, K., & Lyon, G. J. (2013). Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing. *Genome Medicine*, *5*(3), 1–18. https://doi.org/10.1186/GM432/FIGURES/5

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M. R., Zschocke, J., & Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, *15*(2), 256–278. https://doi.org/10.1093/BIB/BBS086

Palamara, P. F., Francioli, L. C., Wilton, P. R., Genovese, G., Gusev, A., Finucane, H. K., Sankararaman, S., Sunyaev, S. R., De Bakker, P. I. W., Wakeley, J., Pe'Er, I., & Price, A. L. (2015). Leveraging Distant Relatedness to Quantify Human Mutation and Gene-Conversion Rates. *The American Journal of Human Genetics*, *97*(6), 775–789. https://doi.org/10.1016/J.AJHG.2015.10.006

Pereira, R., Oliveira, J., & Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of Clinical Medicine 2020, Vol. 9, Page 132*, *9*(1), 132. https://doi.org/10.3390/JCM9010132

Peretz-Machluf, R., Rabinowitz, T., & Shomron, N. (2021). Genome-Wide Noninvasive Prenatal Diagnosis of De Novo Mutations. *Methods in Molecular Biology (Clifton, N.J.)*, *2243*, 249–269. https://doi.org/10.1007/978-1-0716-1103-6_12

Perry, G. H., Yang, F., Marques-Bonet, T., Murphy, C., Fitzgerald, T., Lee, A. S., Hyland, C., Stone, A. C., Hurles, M. E., Tyler-Smith, C., Eichler, E. E., Carter, N. P., Lee, C., & Redon, R. (2007). *Copy number variation and evolution in humans and chimpanzees*. https://doi.org/10.1101/gr.082016.108

Pervez, M. T., Hasnain, M. J. U., Abbas, S. H., Moustafa, M. F., Aslam, N., & Shah, S. S. M. (2022). A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/3457806

Piccolo, P., Mithbaokar, P., Sabatino, V., Tolmie, J., Melis, D., Schiaffino, M. C., Filocamo, M., Andria, G., & Brunetti-Pierri, N. (2014). SMAD4 mutations causing Myhre syndrome result in disorganization of extracellular matrix improved by losartan. *European Journal of Human Genetics*, *22*(8), 988–994.

Pohovski, L. M., Dumic, K. K., Odak, L., & Barisic, I. (2013). Multiplex ligation-dependent probe amplification workflow for the detection of submicroscopic chromosomal abnormalities in patients with developmental delay/intellectual disability. *Molecular Cytogenetics*, *6*(1), 1–6. https://doi.org/10.1186/1755-8166-6-7/FIGURES/2

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature 2016 538:7624*, *538*(7624), 161–164. https://doi.org/10.1038/538161a

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Auwera, G. A. Van der, Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178. https://doi.org/10.1101/201178

Popova, S., Lange, S., Poznyak, V., Chudley, A. E., Shield, K. D., Reynolds, J. N., Murray, M., & Rehm, J. (2019). Population-based prevalence of fetal alcohol spectrum disorder in Canada. *BMC Public Health*, *19*(1), 1–12. https://doi.org/10.1186/S12889-019-7213-3/TABLES/2

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., & Szemes, T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal*, *44*(5), 548–559.

Posey, J. E. (2019). Genome sequencing and implications for rare disorders. *Orphanet Journal of Rare Diseases*, *14*(1), 1–10.

Potterton, J. L., & Eales, C. J. (2001). Prevalence of developmental delay in infants who are HIV positive. *South African Journal of Physiotherapy*, *57*(3), 11–15. https://doi.org/10.4102/SAJP.V57I3.507

Qin, Y., Koehler, S., Zhao Bgi-Shenzhen, S., Mai, R., Liu, Z., Lu, H., & Xing, C. (2021). *High-throughput low-cost DNA sequencing using surface-coating technology*. https://doi.org/10.21203/rs.3.rs-130381/v1

Quang, D., Chen, Y., & Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, *31*(5), 761–763. https://doi.org/DOI: 10.1093/bioinformatics/btu703

Rapaport, F., Boisson, B., Gregor, A., Béziat, V., Boisson-Dupuis, S., Bustamante, J., Jouanguy, E., Puel, A., Rosain, J., Zhang, Q., Zhang, S. Y., Gleeson, J. G., Quintana-Murci, L., Casanova, J. L., Abel, L., & Patin, E. (2021). Negative selection on human genes underlying inborn errors depends on disease outcome and both the mode and mechanism of inheritance. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3), e2001248118. https://doi.org/10.1073/PNAS.2001248118/SUPPL_FILE/PNAS.2001248118.SD01.XLSX

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., Joset, P., Röpke, A., Moog, U., Riess, A., … Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *The Lancet*, *380*(9854), 1674–1682. https://doi.org/10.1016/S0140-6736(12)61480-9

Rausell, A., Mohammadi, P., Mclaren, P. J., Xenarios, I., Fellay, J., & Telenti, A. (2014). *Analysis of stop-gain and frameshift variants in human innate immunity genes*. https://doi.org/10.1101/003376

Ravnan, J. B., Tepperberg, J. H., Papenhausen, P., Lamb, A. N., Hedrick, J., Eash, D., Ledbetter, D. H., & Martin, C. L. (2006). Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *Journal of Medical Genetics*, *43*(6), 478–489. https://doi.org/10.1136/JMG.2005.036350

Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., Friez, M. J., Funke, B. H., Hegde, M. R., & Lyon, E. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *15*(9), 733–747. https://doi.org/10.1038/GIM.2013.92

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1), D886–D894. https://doi.org/DOI: 10.1093/nar/gky1016

Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K. G., McKnight, D., Bai, R., Suchy, S., Friedman, B., Tahiliani, J., Pineda-Alvarez, D., Richard, G., Brandt, T., Haverfield, E., … Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *18*(7), 696–704. https://doi.org/10.1038/GIM.2015.148

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., & Spector, E. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–423. https://doi.org/doi: 10.1038/gim.2015.30

Rizo, J., & Xu, J. (2015). The synaptic vesicle release machinery. *Annual Review of Biophysics*, *44*, 339–367.

Roberts, L., Ratnapriya, R., du Plessis, M., Chaitankar, V., Ramesar, R. S., & Swaroop, A. (2016). Molecular Diagnosis of Inherited Retinal Diseases in Indigenous African Populations by Whole-Exome Sequencing. *Investigative Ophthalmology & Visual Science*, *57*(14), 6374. https://doi.org/10.1167/IOVS.16-19785

Robertson, S. (2019). X-Linked Otopalatodigital Spectrum Disorders. *Atlas of Genetic Diagnosis and Counseling*, 2267–2278. https://doi.org/10.1007/978-1-4939-2401-1_188

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/DOI: 10.1038/nbt.1754

Robinson, P. N., Köhler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics*, *83*(5), 610–615.

Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., & Campbell, C. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*, *34*(3), 511–513.

Romeo, Y., Zhang, X., & Roux, P. P. (2012). Regulation and function of the RSK family of protein kinases. *Biochemical Journal*, *441*(2), 553–569.

Roozen, S., Peters, G. Y., Kok, G., Townend, D., Nijhuis, J., & Curfs, L. (2016). Worldwide prevalence of fetal alcohol spectrum disorders: A systematic literature review including meta-analysis. *Alcoholism: Clinical and Experimental Research*, *40*(1), 18–32.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., … Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature 2011 475:7356*, *475*(7356), 348–352. https://doi.org/10.1038/nature10242

Rump, A., Hildebrand, L., Tzschach, A., Ullmann, R., Schrock, E., & Mitter, D. (2013). A mosaic maternal splice donor mutation in the EHMT1 gene leads to aberrant transcripts and to Kleefstra syndrome in the offspring. *European Journal of Human Genetics*, *21*(8), 887–890.

Saintenac, C. (2015). Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biology*, *16*(1), 1–14. https://doi.org/10.1186/S13059-015-0693-2/COMMENTS

Sango, P. N. (2017). Country profile: Intellectual and developmental disability in Nigeria. *Tizard Learning Disability Review*, *22*(2), 87–93. https://doi.org/10.1108/TLDR-07-2016-0019/FULL/XML

Santen, G. W. E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., Kant, S. G., Snoeck, I. N., Peeters, E. A. J., & Hilhorst-Hofstee, Y. (2012). Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nature Genetics*, *44*(4), 379–380.

Santoro, M. R., Bray, S. M., & Warren, S. T. (2012). Molecular mechanisms of fragile X syndrome: a twenty-year perspective. *Annual Review of Pathology*, *7*, 219–245. https://doi.org/10.1146/ANNUREV-PATHOL-011811-132457

Sasani, T. A., Pedersen, B. S., Gao, Z., Baird, L., Przeworski, M., Jorde, L. B., & Quinlan, A. R. (2019). Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife*, *8*, e46922.

Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F., & Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*, *30*(12), e57. https://doi.org/10.1093/NAR/GNF056

Schuurs-Hoeijmakers, J. H. M., Oh, E. C., Vissers, L. E. L. M., Swinkels, M. E. M., Gilissen, C., Willemsen, M. A., Holvoet, M., Steehouwer, M., Veltman, J. A., De Vries, B. B. A., Van Bokhoven, H., De Brouwer, A. P. M., Katsanis, N., Devriendt, K., & Brunner, H. G. (2012). Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. *American Journal of Human Genetics*, *91*(6), 1122–1127. https://doi.org/10.1016/j.ajhg.2012.10.013

Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, *7*(8), 575–576.

Schwarze, K., Buchanan, J., Taylor, J. C., & Wordsworth, S. (2018). Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in Medicine*, *20*(10), 1122–1130.

Seligmann, H., & Pollock, D. D. (2004). The Ambush Hypothesis: Hidden Stop Codons Prevent Off-Frame Gene Reading. *Https://Home.Liebertpub.Com/Dna*, *23*(10), 701–705. https://doi.org/10.1089/DNA.2004.23.701

Shah, N., Hou, Y. C. C., Yu, H. C., Sainger, R., Caskey, C. T., Venter, J. C., & Telenti, A. (2018). Identification of Misclassified ClinVar Variants via Disease Population Prevalence. *The American Journal of Human Genetics*, *102*(4), 609–619. https://doi.org/10.1016/J.AJHG.2018.02.019

Shah-Kulkarni, S., Ha, M., Kim, B.-M., Kim, E., Hong, Y.-C., Park, H., Kim, Y., Kim, B.-N., Chang, N., & Oh, S.-Y. (2016). Neurodevelopment in early childhood affected by prenatal lead exposure and iron intake. *Medicine*, *95*(4).

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M. P., Chavan, S., Vergara, C., Ortega, V. E., Levin, A. M., Eng, C., Yazdanbakhsh, M., Wilson, J. G., Marrugo, J., Lange, L. A., Williams, L. K., Watson, H., Ware, L. B., … Salzberg, S. L. (2018). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nature Genetics 2018 51:1*, *51*(1), 30–35. https://doi.org/10.1038/s41588-018-0273-y

Shestak, A. G., Bukaeva, A. A., Saber, S., & Zaklyazminskaya, E. V. (2021). Allelic Dropout Is a Common Phenomenon That Reduces the Diagnostic Yield of PCR-Based Sequencing of Targeted Gene Panels. *Frontiers in Genetics*, *12*, 41. https://doi.org/10.3389/FGENE.2021.620337/BIBTEX

Shigemizu, D., Fujimoto, A., Akiyama, S., Abe, T., Nakano, K., Boroevich, K. A., Yamamoto, Y., Furuta, M., Kubo, M., Nakagawa, H., & Tsunoda, T. (2013). A practical method to detect SNVs and indels from whole genome and exome sequencing data. *Scientific Reports*, *3*. https://doi.org/10.1038/SREP02161

Shin, S., Yu, N., Choi, J. R., Jeong, S., & Lee, K.-A. (2015). Routine Chromosomal Microarray Analysis is Necessary in Korean Patients With Unexplained Developmental Delay/Mental Retardation/Autism Spectrum Disorder. *Ann Lab Med*, *35*, 510–518. https://doi.org/10.3343/alm.2015.35.5.510

Shirts, B. H., Pritchard, C. C., & Walsh, T. (2016). Family-Specific Variants and the Limits of Human Genetics. *Trends in Molecular Medicine*, *22*(11), 925–934. https://doi.org/10.1016/J.MOLMED.2016.09.007

Sikkema-Raddatz, B., Johansson, L. F., de Boer, E. N., Almomani, R., Boven, L. G., van den Berg, M. P., van Spaendonck-Zwarts, K. Y., van Tintelen, J. P., Sijmons, R. H., & Jongbloed, J. D. H. (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Human Mutation*, *34*(7), 1035–1042. https://doi.org/DOI: 10.1002/humu.22332

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews. Genetics*, *15*(2), 121–132. https://doi.org/10.1038/NRG3642

South African Government. (2018). *Minister Thulas Nxesi hikes minimum wage of farmworkers and domestic workers | South African Government*. https://www.gov.za/speeches/minister-thulas-nxesi-hikes-minimum-wage-farmworkers-and-domestic-workers-22-feb-2023-0000

Srour, M., & Shevell, M. (2014). Genetics and the investigation of developmental delay/intellectual disability. *Archives of Disease in Childhood*, *99*(4), 386–389. https://doi.org/10.1136/ARCHDISCHILD-2013-304063

Stavropoulos, D. J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R. K. C., Szego, M. J., Hayeems, R. Z., Shaul, R. Z., Brudno, M., Girdea, M., Frey, B., Alipanahi, B., Ahmed, S., Babul-Hirji, R., Porras, R. B., … Marshall, C. R. (2016). Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine. *Npj Genomic Medicine 2016 1:1*, *1*(1), 1–9. https://doi.org/10.1038/npjgenmed.2015.12

Stephan, A. B., Shum, E. Y., Hirsh, S., Cygnar, K. D., Reisert, J., & Zhao, H. (2009). ANO2 is the cilial calcium-activated chloride channel that may mediate olfactory amplification. *Proceedings of the National Academy of Sciences*, *106*(28), 11776–11781.

Stoller, J. K. (2018). The Challenge of Rare Diseases. *CHEST*, *153*, 1309–1314. https://doi.org/10.1016/j.chest.2017.12.018

Strand Life Sciences. (2016). *Storage and Computation Requirements | Strand NGS*. https://www.strand-ngs.com/support/ngs-data-storage-requirements

Strom, S. P., Lee, H., Das, K., Vilain, E., Nelson, S. F., Grody, W. W., & Deignan, J. L. (2014). Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genetics in Medicine*, *16*(7), 510–515. https://doi.org/doi: 10.1038/gim.2013.183

Stuppia, L., Antonucci, I., Palka, G., & Gatta, V. (2012). Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. *International Journal of Molecular Sciences*, *13*(3), 3245. https://doi.org/10.3390/IJMS13033245

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., … Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature 2015 526:7571*, *526*(7571), 75–81. https://doi.org/10.1038/nature15394

Supernat, A., Vidarsson, O. V., Steen, V. M., & Stokowy, T. (2018). Comparison of three variant callers for human whole genome sequencing. *Scientific Reports 2018 8:1*, *8*(1), 1–6. https://doi.org/10.1038/s41598-018-36177-7

Suwinski, P., Ong, C. K., Ling, M. H. T., Poh, Y. M., Khan, A. M., & Ong, H. S. (2019). Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Frontiers in Genetics*, *10*(FEB), 49. https://doi.org/10.3389/FGENE.2019.00049/BIBTEX

Suzuki, S. C., & Takeichi, M. (2008). Cadherins in neuronal morphogenesis and function. *Development, Growth & Differentiation*, *50*, S119–S130.

Swaminathan, G. J., Bragin, E., Chatzimichali, E. A., Corpas, M., Bevan, A. P., Wright, C. F., Carter, N. P., Hurles, M. E., & Firth, H. V. (2012). DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Human Molecular Genetics*, *21*(R1), R37–R44.

Tan, T. Y., Dillon, O. J., Stark, Z., Schofield, D., Alam, K., Shrestha, R., Chong, B., Phelan, D., Brett, G. R., Creed, E., Jarmolowicz, A., Yap, P., Walsh, M., Downie, L., Amor, D. J., Savarirayan, R., McGillivray, G., Yeung, A., Peters, H., … White, S. M. (2017). Diagnostic Impact and Cost-effectiveness of Whole-Exome Sequencing for Ambulant Children With Suspected Monogenic Conditions. *JAMA Pediatrics*, *171*(9), 855–862. https://doi.org/10.1001/JAMAPEDIATRICS.2017.1755

Tarailo-Graovac, M., Shyr, C., Ross, C. J., Horvath, G. A., Salvarinova, R., Ye, X. C., Zhang, L.-H., Bhavsar, A. P., Lee, J. J. Y., Drögemöller, B. I., Abdelsayed, M., Alfadhel, M., Armstrong, L., Baumgartner, M. R., Burda, P., Connolly, M. B., Cameron, J., Demos, M., Dewan, T., … van Karnebeek, C. D. (2016). Exome Sequencing and the Management of Neurometabolic Disorders. *New England Journal of Medicine*, *374*(23), 2246–2255. https://doi.org/10.1056/NEJMOA1515792/SUPPL_FILE/NEJMOA1515792_DISCLOSURES.PDF

Tatton-Brown, K., Cole, T. R., & Rahman, N. (2022). Sotos Syndrome. *Anasthesiologie Und Intensivmedizin*, *63*(10), S356–S363. https://doi.org/10.19224/ai2022.S356

Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J. B., Rimmer, A., Kanapin, A., Lunter, G., Fiddy, S., Allan, C., Aricescu, A. R., Attar, M., Babbs, C., Becq, J., Beeson, D., Bento, C., Bignell, P., Blair, E., Buckle, V. J., … McVean, G. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature Genetics 2015 47:7*, *47*(7), 717–726. https://doi.org/10.1038/ng.3304

Teixeira, W. G., Marques, F. K., & Freire, M. C. M. (2016). Retrospective karyotype study in mentally retarded patients. *Revista Da Associação Médica Brasileira*, *62*(3), 262–268. https://doi.org/10.1590/1806-9282.62.03.262

Temtamy, S. A., Salam, M. A., Aboul-Ezz, E. H. A., Hussein, H. A., Helmy, S. A. H., & Shalash, B. A. (1996). New autosomal recessive multiple congenital abnormalities/mental retardation syndrome with craniofacial dysmorphism absent corpus callosum, iris colobomas and connective tissue dysplasia. *Clinical Dysmorphology*, *5*(3), 231–240. https://doi.org/10.1097/00019605-199607000-00007

Thormann, A., Halachev, M., McLaren, W., Moore, D. J., Svinti, V., Campbell, A., Kerr, S. M., Tischkowitz, M., Hunt, S. E., & Dunlop, M. G. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nature Communications*, *10*(1), 1–10. https://doi.org/doi: 10.1038/s41467-019-10016-3.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192.

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., Weiss, M. E., Köster, J., Marais, A., Paknia, O., Schröder, R., Garcia-Aznar, J. M., Werber, M., Brandau, O., Calvo Del Castillo, M., Baldi, C., Wessel, K., Kishore, S., Nahavandi, N., Eyaid, W., Al Rifai, M. T., Al-Rumayyan, A., … Abou Jamra, R. (2016). Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *European Journal of Human Genetics 2017 25:2*, *25*(2), 176–182. https://doi.org/10.1038/ejhg.2016.146

Tsai, G. J., Rañola, J. M. O., Smith, C., Garrett, L. T., Bergquist, T., Casadei, S., Bowen, D. J., & Shirts, B. H. (2019). Outcomes of 92 patient-driven family studies for reclassification of variants of uncertain significance. *Genetics in Medicine*, *21*(6), 1435–1442. https://doi.org/10.1038/s41436-018-0335-7

Tsurusaki, Y., Ohashi, I., Enomoto, Y., Naruto, T., Mitsui, J., Aida, N., & Kurosawa, K. (2017). *A novel UBE2A mutation causes X-linked intellectual disability type Nascimento*. https://doi.org/10.1038/hgv.2017.19

UNAIDS. (2022). *F A C T S H E E T 2 0 2 2*.

Valentino, F., Bruno, L. P., Doddato, G., Giliberti, A., Tita, R., Resciniti, S., Fallerini, C., Bruttini, M., Lo Rizzo, C., Mencarelli, M. A., Mari, F., Pinto, A. M., Fava, F., Baldassarri, M., Fabbiani, A., Lamacchia, V., Benetti, E., Zguro, K., Furini, S., … Ariani, F. (2021). Exome sequencing in 200 intellectual disability/autistic patients: New candidates and atypical presentations. *Brain Sciences*, *11*(7), 936. https://doi.org/10.3390/BRAINSCI11070936/S1

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.

V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, *43*(1), 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.BI1110S43

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media.

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. *Frontiers in Genetics*, *13*, 912. https://doi.org/10.3389/FGENE.2022.864822/BIBTEX

Verloes, A., Moës, D., Palumbo, L., Elmer, C., François, A., & Bricteux, G. (1993). Schinzel-Giedion syndrome. *European Journal of Pediatrics*, *152*(5), 421–423. https://doi.org/10.1007/BF01955902/METRICS

Vermeesch, J. R., Fiegler, H., de Leeuw, N., Szuhai, K., Schoumans, J., Ciccone, R., Speleman, F., Rauch, A., Clayton-Smith, J., Van Ravenswaaij, C., Sanlaville, D., Patsalis, P. C., Firth, H., Devriendt, K., & Zuffardi, O. (2007). Guidelines for molecular karyotyping in constitutional genetic diagnosis. *European Journal of Human Genetics : EJHG*, *15*(11), 1105–1114. https://doi.org/10.1038/SJ.EJHG.5201896

Virdee, S. (2022). An atypical ubiquitin ligase at the heart of neural development and programmed axon degeneration. *Neural Regeneration Research*, *17*(11), 2347.

Vissers, L. E. L. M., De Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., De Vries, P., Van Lier, B., Arts, P., Wieskamp, N., Del Rosario, M., Van Bon, B. W. M., Hoischen, A., De Vries, B. B. A., Brunner, H. G., & Veltman, J. A. (2010). A de novo paradigm for mental retardation. *Nature Genetics 2010 42:12*, *42*(12), 1109–1112. https://doi.org/10.1038/ng.712

Vissers, L. E. L. M., Gilissen, C., & Veltman, J. A. (2015). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics 2015 17:1*, *17*(1), 9–18. https://doi.org/10.1038/nrg3999

Vissers, L. E. L. M., Van Nimwegen, K. J. M., Schieving, J. H., Kamsteeg, E. J., Kleefstra, T., Yntema, H. G., Pfundt, R., Van Der Wilt, G. J., Krabbenborg, L., Brunner, H. G., Van Der Burg, S., Grutters, J., Veltman, J. A., & Willemsen, M. A. A. P. (2017). A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genetics in Medicine 2017 19:9*, *19*(9), 1055–1063. https://doi.org/10.1038/gim.2017.1

Vogel, M., Utpatel, C., Corbett, C., Kohl, T. A., Iskakova, A., Ahmedov, S., Antonenka, U., Dreyer, V., Ibrahimova, A., Kamarli, C., Kosimova, D., Mohr, V., Sahalchyk, E., Sydykova, M., Umetalieva, N., Kadyrov, A., Kalmambetova, G., Niemann, S., & Hoffmann, H. (2021). Implementation of whole genome sequencing for tuberculosis diagnostics in a low-middle income, high MDR-TB burden country. *Scientific Reports 2021 11:1*, *11*(1), 1–9. https://doi.org/10.1038/s41598-021-94297-z

Vorstman, J. A. S., & Ophoff, R. A. (2013). Genetic causes of developmental disorders. In *Current Opinion in Neurology* (Vol. 26, Issue 2, pp. 128–136). https://doi.org/10.1097/WCO.0b013e32835f1a30

Vrečar, I., Innes, J., Jones, E. A., Kingston, H., Reardon, W., Kerr, B., Clayton-Smith, J., & Douzgou, S. (2017). Further clinical delineation of the MEF2C haploinsufficiency syndrome: report on new cases and literature review of severe neurodevelopmental disorders presenting with seizures, absent speech, and involuntary movements. *Journal of Pediatric Genetics*, *6*(03), 129–141.

Waggoner, D., Wain, K. E., Dubuc, A. M., Conlin, L., Hickey, S. E., Lamb, A. N., Martin, C. L., Morton, C. C., Rasmussen, K., Schuette, J. L., Schwartz, S., & Miller, D. T. (2018). Yield of additional genetic testing after chromosomal microarray for diagnosis of neurodevelopmental disability and congenital anomalies: a clinical practice resource of

the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, *20*(10), 1105–1113. https://doi.org/10.1038/S41436-018-0040-6

Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., & Guo, Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics*, *31*(3), 318–323. https://doi.org/10.1093/BIOINFORMATICS/BTU668

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J., Guo, Y., Feng, B., Li, H., Lu, Y., Fang, X., Liang, H., Du, Z., Li, D., Zhao, Y., … Wang, J. (2008). The diploid genome sequence of an Asian individual. *Nature 2008 456:7218*, *456*(7218), 60–65. https://doi.org/10.1038/nature07484

Watson, C. T., Marques-Bonet, T., Sharp, A. J., & Mefford, H. C. (2014). The genetics of microdeletion and microduplication syndromes: An update. *Annual Review of Genomics and Human Genetics*, *15*, 215–244. https://doi.org/10.1146/ANNUREV-GENOM-091212-153408

Watson, M. (2015). Clinical utility of genetic and genomic services: a position statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *17*(6), 505–507. https://doi.org/10.1038/GIM.2015.41

Weinstein, V., Tanpaiboon, P., Chapman, K. A., Mew, N. A., & Hofherr, S. (2017). Do the data really support ordering fragile X testing as a first-tier test without clinical features? *Genet Med*, *19*(12), 1317–1322. https://doi.org/10.1038/gim.2017.64

Weiß, A. L., Meijer, M., Budeus, B., Pauper, M., Hakobjan, M., Groothuismink, J., Shi, Y., Neveling, K., Buitelaar, J. K., & Hoogman, M. (2021). DNA methylation associated with persistent ADHD suggests TARBP1 as novel candidate. *Neuropharmacology*, *184*, 108370.

Whitlock, J. H., Wilk, E. J., Howton, T. C., Clark, A. D., & Lasseigne, B. N. (2023). The landscape of SETBP1 gene expression and transcription factor activity across human tissues. *BioRxiv*.

Wiener, E., Buchanan, J., Lombard, Z., & Krause, A. (2022). *Genetic services in a resource-constrained African setting-diagnostic odysseys, limited testing strategies and the need for exome sequencing*. https://doi.org/10.21203/rs.3.rs-1738191/v1

Wolf, N. I., Sistermans, E. A., Cundall, M., Hobson, G. M., Davis-Williams, A. P., Palmer, R., Stubbs, P., Davies, S., Endziniene, M., Wu, Y., Chong, W. K., Malcolm, S., Surtees, R., Garbern, J. Y., & Woodward, K. J. (2005). Three or more copies of the proteolipid protein gene PLP1 cause severe Pelizaeus–Merzbacher disease. *Brain*, *128*(4), 743–751. https://doi.org/10.1093/BRAIN/AWH409

Wong, M. (2010). Mammalian target of rapamycin (mTOR) inhibition as a potential antiepileptogenic therapy: From tuberous sclerosis to common acquired epilepsies. *Epilepsia*, *51*(1), 27–36. https://doi.org/10.1111/j.1528-1167.2009.02341.x

Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., McRae, J. F., Van Kogelenberg, M., King, D. A., Ambridge, K., Barrett, D. M., Bayzetinova, T., Bevan, A. P., Bragin, E., Chatzimichali, E. A., Gribble, S., Jones, P., Krishnappa, N., Mason, L. E., Miller, R., Morley, K. I., … Firth, H. V. (2015). Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of genome-wide research data. *The Lancet*, *385*(9975), 1305–1314. https://doi.org/10.1016/S0140-6736(14)61705-0

Wright, C. F., McRae, J. F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T. W., Jones, P., Prigmore, E., Rajan, D., & Lord, J. (2018). Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine*, *20*(10), 1216–1223.

Wright, C. F., McRae, J. F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T. W., Jones, P., Prigmore, E., Rajan, D., Lord, J., Sifrim, A., Kelsell, R., Parker, M. J., Barrett, J. C., Hurles, M. E., FitzPatrick, D. R., & Firth, H. V. (2018). Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families

with developmental disorders. *Genetics in Medicine*, *20*(10), 1216–1223. https://doi.org/10.1038/GIM.2017.246/ATTACHMENT/BEDE2221-E535-4B2E-8308-451478B94ECD/MMC3.PDF

Wu, B., Kang, W., Wang, Y., Zhuang, D., Chen, L., Li, L., Su, Y., Pan, X., Wei, Q., Tang, Z., Li, Y., Gao, J., Cheng, R., Zhou, W., Wang, Z., Qiu, G., Wang, J., Yang, L., Zhang, P., … Zhou, W. (2021). Application of Full-Spectrum Rapid Clinical Genome Sequencing Improves Diagnostic Rate and Clinical Outcomes in Critically Ill Infants in the China Neonatal Genomes Project. *Critical Care Medicine*, *49*(10), 1674–1683. https://doi.org/10.1097/CCM.0000000000005052

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols 2015 10:10*, *10*(10), 1556–1566. https://doi.org/10.1038/nprot.2015.105

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., & Buhay, C. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Jama*, *312*(18), 1870–1879.

Yoneda, Y., Saitsu, H., Touyama, M., Makita, Y., Miyamoto, A., Hamada, K., Kurotaki, N., Tomita, H., Nishiyama, K., Tsurusaki, Y., Doi, H., Miyake, N., Ogata, K., Naritomi, K., & Matsumoto, N. (2012). Missense mutations in the DNA-binding/dimerization domain of NFIX cause Sotos-like features. *Journal of Human Genetics 2012 57:3*, *57*(3), 207–211. https://doi.org/10.1038/jhg.2012.7

Yu, L. H., Morimura, T., Numata, Y., Yamamoto, R., Inoue, N., Antalfy, B., Goto, Y. ichi, Deguchi, K., Osaka, H., & Inoue, K. (2012). Effect of curcumin in a mouse model of Pelizaeus–Merzbacher disease. *Molecular Genetics and Metabolism*, *106*(1), 108–114. https://doi.org/10.1016/J.YMGME.2012.02.016

Yu, S., Choi, W.-I., Choi, Y. J., Kim, H.-Y., Hildebrandt, F., & Gee, H. Y. (2020). PLCE1 regulates the migration, proliferation, and differentiation of podocytes. *Experimental & Molecular Medicine*, *52*(4), 594–603.

Yuen, T., Carter, M. T., Szatmari, P., & Ungar, W. J. (2018). Cost-effectiveness of Genome and Exome Sequencing in Children Diagnosed with Autism Spectrum Disorder. *Applied Health Economics and Health Policy*, *16*(4), 481–493. https://doi.org/10.1007/S40258-018-0390-X/METRICS

Zack, R. M., Golan, J., Aboud, S., Msamanga, G., Spiegelman, D., & Fawzi, W. (2014). Risk factors for preterm birth among HIV-infected Tanzanian women: a prospective study. *Obstetrics and Gynecology International*, *2014*.

Zentner, G. E., Hurd, E. A., Schnetz, M. P., Handoko, L., Wang, C., Wang, Z., Wei, C., Tesar, P. J., Hatzoglou, M., & Martin, D. M. (2010). CHD7 functions in the nucleolus as a positive regulator of ribosomal RNA biogenesis. *Human Molecular Genetics*, *19*(18), 3491–3501.

Zhang, T., Garrison, E., Marti, M. A., Bope, C. D., Domilongo Bope, C., Chimusa, E. R., Nembaware, V., Mazandu, G. K., De Vries, J., & Wonkam, A. (2019). *Dissecting in silico Mutation Prediction of Variants in African Genomes: Challenges and Perspectives*. https://doi.org/10.3389/fgene.2019.00601

Zhang, Y., Jing, Y., Qiao, J., Luan, B., Wang, X., Wang, L., & Song, Z. (2017). *Activation of the mTOR signaling pathway is required for asthma onset*. https://doi.org/10.1038/s41598-017-04826-y

Zhao, X., Collins, R., Lee, W., Weber, A., … Y. J.-T. A. J. of, & 2021, undefined. (2021). Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. *Elsevier*. https://www.sciencedirect.com/science/article/pii/S0002929721000987

Zhong, Y., Xu, F., Wu, J., Schubert, J., & Li, M. M. (2021). Application of Next Generation Sequencing in Laboratory Medicine. *Annals of Laboratory Medicine*, *41*(1), 25–43. https://doi.org/10.3343/ALM.2021.41.1.25