Designing and evaluating the utility of a panel of *de novo* mutation enriched genes for diagnosing South African patients with developmental delay.

Patracia Livhuhani Nevondwe
Student Number: 1819460
Supervisor: Prof Amanda Krause
Supervisor: Dr Nadia Carstens

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

March 2024

DECLARATION

I, Patracia Livhuhani Nevondwe, declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

…… 18 September 2024 …….

Patracia Livhuhani Nevondwe Date

DEDICATION

In loving memory of my mother

Emily Malixu Nevondwe

1969 – 2016

PRESENTATIONS

Patracia Nevondwe, Zane Lombard, DDD-Africa Consortium, Amanda Krause and Nadia Carstens: A DNM enriched Gene Panel: An Acceptable Targeted Diagnostic Approach for unexplained DDs in Resource-Limited Settings. Poster Presentation. International Congress of Human Genetics 2023 -Young Investigator Forum, 21 February 2023, Cape Town.

Patracia Nevondwe, Zane Lombard, DDD-Africa Consortium, Amanda Krause and Nadia Carstens: A DNM enriched Gene Panel: An Acceptable Targeted Diagnostic Approach for unexplained DDs in Resource-Limited Settings. Poster Presentation. International Congress of Human Genetics 2023, 22 – 26 February 2023, Cape Town.

Patracia Nevondwe, Zane Lombard, DDD-Africa Consortium, Amanda Krause and Nadia Carstens: Designing and evaluating the utility of a *de novo* mutation enriched gene panel in diagnosing patients with DDs in South Africa. Oral presentation at the Wits Faculty of Health Sciences, 2022 Research day and postgraduate expo, 15 September 2022, Johannesburg.

Patracia Nevondwe, Zane Lombard, DDD-Africa Consortium1, Amanda Krause and Nadia Carstens: Designing and evaluating the utility of a *de novo* mutation enriched gene panel in diagnosing patients with DDs in South Africa. Oral presentation at the Division of Human Genetics, Seminar, 26 October 2022, Johannesburg.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following individuals and/or institutions:

- **Supervisors:** Prof Amanda Krause and Dr Nadia Carstens, I extend my sincere thanks for providing me with the opportunity to work on this intriguing project. Your support, guidance, encouragement, patience, and wisdom over the years have been invaluable. I could not have asked for better supervisors.
- **Collaborators:** Prof Zane Lombard, thank you for granting me access to clinical and WES data for the DDD-Africa project. I also appreciate Ms Fahmida Essop for her assistance with samples referred for FXS testing in the Division. A special thank you is extended to Mr Phelelani Mpangase for his invaluable assistance in developing bioinformatics scripts to extract the study's region of interest from WES data.
- **Division of Human Genetics (NHLS):** To the postgraduate students and staff, thank you for your help, support, and ongoing encouragement during my time in the Division.
- **Thermo Fisher Scientific:** I want to express my gratitude to Dewaldt Engelbrecht for his technical and emotional support during the lab work of this study.
- **Family:** To my family, thank you for your unwavering support, love, and encouragement. I know it was not an easy journey, and I appreciate your steadfast support.
- **Funders:** A heartfelt thank you to NHLS Research Trust, Faculty Research Committee, and the National Research Foundation, South Africa, for their financial support.
- **Friends:** Lastly, I want to extend my thanks to my friends for their unwavering support.

Your contributions have been instrumental in the success of this project, and I am truly grateful to each one of you.

NOMENCLATURE

ACMG: American College of Medical Genetics

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

AD: Autosomal Dominant

AR: Autosomal Recessive

BAM: Binary Alignment Map

CADD: Combined Annotation Dependent Depletion

CMA: Chromosomal Microarray

CNV: Copy Number Variant

DD: Developmental Disorder

DDD-Africa: Deciphering Developmental Disorders in Africa

DDD-UK: Deciphering Developmental Disorders in the United Kingdom

DNA: Deoxyribonucleotide Acid

DNM: *De Novo* Mutation

FISH: Fluorescent in Situ Hybridization

FXS: Fragile X Syndrome

gDNA: Genomic Deoxyribonucleic Acid

GnomAD: The Genome Aggregation Database

GRCh37: Genome Reference Consortium Human Build 37

GRCh38: Genome Reference Consortium Human Build 38

HPO: Human Phenotype Ontology

HREC: Human Research Ethics Committee

HumGen: Division of Human Genetics, Wits & NHLS

ID: Intellectual Disability

IGV: Interactive Genome Viewer

Indels: Insertion–Deletion Mutations

LMIC: Low- and Middle-Income Countries

MAF: Minor Allele Frequency

MLPA: Multiplex Ligation Probe Dependent Amplification

NDD: Neurodevelopmental Disorder

NGS: Next Generation Sequencing

NHLS: National Health Laboratory Service

PCR: Polymerase Chain Reaction

POLYPHEN:2: Polymorphism Phenotyping Version 2

QC: Quality Control

SA: South Africa

SNP: Single Nucleotide Polymorphism

SNV: Single Nucleotide Variant

SV: Structural Variant

VCF: Variant Call Format

VEP: Ensembl Variant Effect Predictor

VUS: Variant of Uncertain/Unknown Significance

WES: Whole Exome Sequencing

WGS: Whole Genome Sequencing

ABSTRACT

Developmental disorders (DDs), are diverse and life-altering conditions, with a higher prevalence in sub-Saharan Africa. Approximately 50% of DDs have a genetic basis, but many patients remain undiagnosed due to limitations in current testing methods. Whole exome sequencing (WES) and whole genome sequencing (WGS) are recommended due to their high diagnostic rates, but their routine use is impractical in many low- and middle-income countries (LMICs), including South Africa. Thus, cost-effective alternatives are needed. Targeted gene panels, despite declining global use, might still be relevant in LMICs. This study explored a targeted *de novo* mutation (DNM)-enriched gene panel in two cohorts: a well-phenotyped group and a group with developmental delay but less phenotypic detail. The analysis included WES on samples from 96 patients in each cohort, followed by a virtual DNM-enriched gene panel analysis. Results showed that 15% of the well-phenotyped cohort and 5% of the unphenotyped cohort had causal variants, which aligns with global diagnostic yields for targeted gene panels in DDs. Among these variants, 47% were *de novo* mutations, 5% were maternally inherited, and 47% had indeterminate inheritance due to the absence of parental samples. Significant clinical interventions were identified for the majority of patients. Approximately 50% of the causative variants were novel, thus expanding the mutation profile of DDs. Additionally, we identified key factors influencing diagnostic yield, such as the availability of phenotypic data and the use of trio analysis. These findings provide valuable insights into the implementation of routine genetic testing in LMICs, highlighting the importance of comprehensive clinical data submission, trio sequencing and/or analysis, and the evaluation of copy number variants in the diagnostic workup of individuals with DDs. We developed a DNM-enriched gene panel as a starting point for creating practical, resource-efficient diagnostic strategies for DDs. We conclude that this approach would be beneficial for the diagnostic evaluation of DDs and developmental delays in resource-constrained settings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1. INTRODUCTION

Developmental disorders (DDs) encompass a range of conditions marked by abnormal growth and development, typically manifesting during infancy to early childhood. These disorders are characterized by a variety of physical issues, including congenital malformations, neurological or muscular defects, growth delays, skeletal anomalies, sensory impairments, cardiovascular complications, dermatological conditions, and abnormalities in internal organs, along with delayed developmental milestones (Reiss, 2009). While DDs exhibit diverse aetiologies, at least 50% are known to have a genetic origin (Srour & Shevell, 2014). DDs are characterised by both phenotypic and genetic heterogeneity, making deciphering their genetic aetiology challenging. A retrospective file audit at the Division of Human Genetics, Wits, and NHLS (HumGen), showed that approximately 83% of patients at the participating genetic clinics of the Division present with clinical features, including delays or atypical developmental milestones, difficulties with motor skills, cognitive impairments, communication challenges, behavioural issues, sensory processing issues, physical abnormalities, social interaction difficulties, learning difficulties suggestive of DDs. Despite undergoing standard genetic testing established for DDs at the Division, the majority of this patient group do not receive a molecular diagnosis (Wiener *et al*., 2023).

Globally, next-generation sequencing (NGS), specifically whole exome sequencing (WES) and/or whole genome sequencing (WGS), have emerged as powerful tools for diagnosing rare genetic disorders, including DDs (Satam *et al*., 2023). International recommendations and guidelines endorse these as primary diagnostic tools for assessing children with DDs (Manickam *et al*., 2021; Srivastava *et al*., 2019). However, their adoption in routine diagnostic laboratories faces obstacles such as cost considerations and the need for specialized infrastructure, particularly in resource-constrained settings like South Africa (Kamp *et al*., 2021; Krause, 2019; Lumaka *et al*., 2022).

Acknowledging the challenges associated with implementing recommended testing strategies for patients with DD is imperative for fostering innovation and harnessing existing resources to devise more efficient testing approaches. This is crucial in addressing prevailing healthcare disparities faced by African populations in accessing personalised healthcare, where genomic medicine needs to become integrated into routine healthcare practices. Acknowledging the urgent need for resource-effective testing strategies, our study focused on evaluating the utility of a targeted gene panel in South African patients with DDs. This gene panel encompasses

genes enriched for de novo mutations (DNMs), as previous studies have reported a substantial contribution of causal variants in these genes to DDs (Wilfert *et al*., 2017).

This introductory chapter begins with an overview of DDs, followed by an exploration of common and historical genetic testing strategies for DDs. Additionally, international guidelines and recommendations for genetic testing in individuals with DDs are presented. The subsequent section discusses the challenges encountered in implementing these guidelines in resource-constrained settings. The study employed a targeted gene panel strategy, and thus, the design and curation of the targeted gene panel are outlined. A subsequent section provides a description of genes enriched for DNMs in DDs. Finally, the chapter details two cohorts with developmental delay and DDs under investigation in this study, followed by the rationale, aims, and objectives of the study.

1.1. Developmental Disorders

Developmental disorders are a group of clinically and genetically heterogeneous disorders characterised by abnormal delayed physical and/or mental development (Boyle *et al*., 2011). These developmental impairments typically result in severe and chronic disabilities that impact a patient's daily functioning. Physical abnormalities and malformations such as organ malformations or limb anomalies are evident from embryonic development or infancy. Abnormal neurological development usually manifests in early childhood as delayed developmental milestones, or later in life as intellectual disability (Reiss, 2009).

The prevalence of DDs is estimated to be more than 15% in children 3–17 years of age worldwide (Boyle *et al.*, 2011). The documentation of the prevalence and epidemiology of DDs in South Africa is lacking. The dearth of comprehensive population-based data regarding childhood disabilities, including DDs in the country, has resulted in inadequate funding and reduced prioritisation of services aimed at diagnosing and treating these conditions (Kromberg *et al.*, 2013; Malherbe *et al.*, 2021). Addressing this critical data gap is imperative for the development of effective strategies, optimal resource allocation, and enhancement of support and services for individuals grappling with DDs in the country. There have been very few studies to establish the overall prevalence and dominant aetiologies of DDs in South Africa. In 2002, a study conducted within a rural South African community reported a minimum prevalence of intellectual and developmental disabilities at 35.6 per 1000 children (Christianson *et al.*, 2002). Similarly, within the same year and province, KwaZulu-Natal, Couper reported a prevalence of DDs at 83 per 1000 children (Couper, 2002).

A common symptom of many patients with DD is developmental delay which is defined as a phenomenon in which a child fails to attain age-related developmental milestones as compared to their peers from the same population and is a common reason for diagnostic assessment by paediatricians (Srour & Shevell, 2014). Developmental delay is estimated to affect 1% – 3% of the population of children under 5 years of age (Srour & Shevell, 2014). In an effort to establish the prevalence of developmental delay, Grantham-McGregor and colleagues estimated that around 200 million children worldwide were at risk for developmental delay with a considerable proportion in sub-Saharan Africa and South Asia. They suggested that these delays are often linked to a combination of factors: limited access to healthcare, which can hinder the management of treatable conditions; nutritional deficiencies, which affect brain and physical development; a high prevalence of infectious diseases that can interfere with overall health; socioeconomic challenges that restrict access to essential resources and opportunities; and prenatal and perinatal conditions that impact development from the earliest stages (Grantham-McGregor *et al.*, 2007). In a separate study, McCoy and colleagues estimated that around 80.8 million children ages 3 and 4 years in LMICs experienced low cognitive and/or socioemotional development, with the largest number of affected children in sub-Saharan Africa (43.8% of children ages 3 and 4 years) (McCoy *et al.*, 2016).

1.2. Aetiology of Developmental Disorders

The aetiology of DDs exhibits a high degree of heterogeneity and, for the majority of cases, remains idiopathic (Musante & Ropers, 2014; Vissers *et al.*, 2016). When the aetiology is known, it involves a complex interplay of genetic and non-genetic factors (Miclea *et al.*, 2015; Strømme *et al.*, 2000). When the aetiology of DD is known, it involves a complex interplay of two primary factors: genetic and non-genetic (Miclea *et al*., 2015; Strømme *et al*., 2000). Non-genetic factors encompass a wide range of influences, including infectious agents (such as cytomegalovirus, rubella virus, and Trichomonas), exposure to toxic substances (such as alcohol or lead), perinatal events (e.g., periventricular haemorrhage in extreme prematurity, hypoxia-ischemia at preterm gestation, and congenital hypothyroidism), and postnatal events (e.g., meningitis and traumatic brain injury) (Huang et al., 2016; Meyer, 2019; Smith & Brown, 2014). At least 50% of DDs have a genetic aetiology (Srour & Shevell, 2014). Given that this project focuses on the genetic aetiology of developmental DDs, the following section will discuss the most common, well-characterised, and diagnostically evaluated genetic alterations. This includes chromosomal abnormalities, copy number variations, single nucleotide variants (SNVs), and insertions/deletions (indels).

1.2.1. Chromosomal Abnormalities

Chromosomal abnormalities encompass structural and numerical variations in chromosomes (Shaffer & Theisen, 2010). Structural variations (SVs) involve genomic rearrangements involving at least 50 nucleotides altering the structure of at least one chromosome through deletions, translocations, inversions, duplications, ring chromosomes, and isochromosomes (Alkan *et al.,* 2011; Shaffer, 2005). SVs may contribute to human genomic disorders through disruption of protein-coding genes or interactions with cis-regulatory elements (Collins *et al.,* 2017; Pócza *et al.,* 2021). Contrastingly, numerical abnormalities result in changes in the number of chromosomes, leading to aneuploidy (Shaffer & Theisen, 2010). Numerical disorders are more prevalent than structural ones and can be categorised into groups where individual chromosomes are missing or duplicated, or entire haploid sets of chromosomes are added or lost (McFeely, 1993). Notable numerical chromosomal disorders include Trisomy 13, Trisomy 18, Trisomy 21, XXY syndrome, Turner syndrome, and XXX syndrome (Witters *et al.,* 2011). Among these, Trisomy 13, 18 and 21 are particularly prominent in their association with DDs (Cottino *et al.,* 2022).

1.2.2. Single Nucleotide Variants (SNVs) and Indels

Monogenic causes of DDs involve SNVs and indels, contributing to more than 30% of the broader spectrum of neurodevelopmental disorders (Srivastava *et al*., 2019). Based on the location and nature of the nucleotide change, SNVs can be classified as synonymous (silent), missense (resulting in a change in the encoded amino acid), or nonsense (leading to premature protein termination). In the context of DDs, causal SNVs and indels can occur in genes essential for normal neurodevelopment and cognitive function. These genetic variations can disrupt developmental pathways involved in brain development, neuronal connectivity, and synaptic function, leading to DDs and intellectual disabilities (Forrest *et al.*, 2018; Liaci *et al.,* 2021; Moretto *et al.,* 2018). The severity and clinical manifestations of DDs caused by these variations can vary widely, depending on the specific gene affected, the location of the variant, and its functional consequences. Monogenic variants associated with DDs can arise de novo, be inherited from parents, or occur in a mosaic pattern (where only a subset of cells carries the variant) (Wilfert *et al.*, 2017)

1.2.3. Copy Number Variants

Copy number variants (CNVs) are genomic alterations that result in differences in the number of copies of specific DNA segments compared to the reference genome (Freeman *et al*., 2006). CNVs are defined as DNA segments that vary in copy number and range from as small as 50 base pairs to several megabases in size (Alkan *et al*., 2011; Feuk *et al*., 2006; MacDonald *et al*., 2014). These variations arise from structural changes within the genome, including duplications, deletions, insertions, and translocations (Park *et al*., 2019; Shaikh, 2017). The mechanisms underlying CNVs involve processes such as DNA recombination, replication, and repair (Pös *et al*., 2021). Initially, CNV detection relied on techniques such as comparative genomic hybridization (CGH), multiple ligation probe amplification (MLPA) and karyotyping. However, technological advancements, particularly the introduction of microarray technology and WES, have significantly enhanced CNV detection (Coe *et al*., 2019; Miller *et al*., 2010; Uddin *et al*., 2016). These modern technologies have progressively revealed the diverse roles of CNVs in various disorders, including DDs.

1.3. Genetic Testing of Developmental Disorders

The investigation into the underlying genetic aetiology of DDs involves a variety of molecular and cytogenetic tests (Bélanger & Caron, 2018). The evolution of genetic testing technologies has notably altered the landscape of diagnostic procedures for DDs (Manickam *et al.*, 2021). Considering the heterogeneous nature of DDs, testing methodologies encompass both CNVs and SNVs. Given that this thesis is focused on developing a resource-effective diagnostic strategy for DDs, the coming sections will focus on historic diagnostic testing strategies for DDs at HumGen. Additionally, the discussion will extend to NGS strategies, as the proposed diagnostic approach in this thesis is NGS-based.

1.3.1. Conventional Cytogenetics (Karyotyping)

Conventional cytogenetics, commonly known as karyotyping, emerged as one of the earliest genetic tests employed in the initial assessment of suspected developmental or intellectual disabilities (Moeschler *et al.*, 2006; Srour & Shevell, 2014). This technique is proficient in detecting chromosomal anomalies, such as large translocations, aneuploidies, as well as duplications, deletions, or insertions, and can identify the underlying genetic cause in up to 3%-5% of cases (excluding common aneuploidies) (Sadek & Mohamed, 2018). However, karyotyping has inherent limitations, particularly in its resolution, which is confined to the range of 5 Mb-10 Mb (Berisha *et al.*, 2020). Furthermore, karyotyping is labour-intensive,

involving a multi-step process encompassing cell culturing, harvesting, slide preparation, banding, and chromosomal analysis (Howe *et al.*, 2014).

1.3.2. Fluorescent *in situ* Hybridization (FISH)

Since the 1990s, Fluorescent in situ hybridization (FISH) technology has played a crucial role in molecular diagnostics. Offering a targeted approach, FISH allows for the visualisation and of specific chromosomal regions in metaphase or interphase (Shakoori, 2017). This technique uses fluorescently labelled DNA probes that bind to precise target sequences within chromosomes, facilitating the identification and visualisation of genetic material under a microscope (Shakoori, 2017). The resolution achieved by FISH, typically in the 100-200kb range, is greater than that of karyotyping (Cui *et al.*, 2016). Notably, FISH has shown a diagnostic yield of at least 1.3% in patients with developmental delays (Ravanan et al., 2006: Raunch et al., 2006). However, a significant limitation of FISH lies in its dependence on pre-designated probes for specific targets, making it unsuitable for exploring and identifying novel pathogenic loci. Despite this constraint, FISH becomes highly effective when a suspected cause with a corresponding probe is well-established, as it targets a single locus rather than conducting a comprehensive examination of the entire genome (Manning & Hudgins, 2010).

1.3.3. Multiple Ligation Probe Amplification (MLPA)

Multiplex ligation-dependent probe amplification (MLPA) is a molecular genetic technique employed for screening CNVs (Schouten *et al.*, 2002).. MLPA focuses on specific genome regions by employing multiple DNA probes that are ligated to form an elongated DNA molecule. Subsequently, through multiplex PCR and capillary electrophoresis, fluorescently labelled amplicons generate a distinctive peak pattern. Analysing this pattern in comparison to reference probes facilitates the determination of relative quantities or ratios, aiding in genetic analysis (Schouten *et al.*, 2002). Despite MLPA not providing the same comprehensive genome coverage as CMA, it remains a crucial technique for assessing cases involving DDs, particularly in settings where routine CMA is not yet feasible (Miclea *et al.*, 2021).

1.3.4. Chromosomal Microarray (CMA)

Chromosomal microarray technology (CMA) remains the gold standard for CNV detection. It has a higher resolution than conventional cytogenetic analysis allowing for the identification of chromosomal imbalances with greater precision, accuracy, and technical sensitivity (Manning & Hudgins, 2010; Shaffer, 2005). While MLPA and CMA target the same types of genetic variants, MLPA exhibits lower resolution compared to CMA (Capkova *et al.*, 2019) There are two common types of microarrays. Array comparative hybridisation (aCGH) and

single nucleotide polymorphisms (SNP) array. aCGH is a microarray technique employed for exclusively detecting CNVs across the entire genome (Lantieri *et al.,* 2017) This technique involves hybridising test and reference samples, each labelled with a different fluorescent dye, to a microarray slide containing DNA probes representing specific genomic regions (Bejjani & Shaffer, 2006). CNVs are then determined by the differences in hybridization pattern intensities between patient DNA and control DNA (Freeman *et al.,* 2006). Similarly, SNP microarray is also employed for the detection of CNVs. However, in addition, SNP microarray can detect regions of homozygosity attributed to uniparental disomy or parental consanguinity that can give rise to disease (Bernardini *et al.,* 2010; Bruno *et al.,* 2011; McCarroll *et al.,* 2008). Although CMA technologies have significantly transformed the detection of CNVs across the genome, their effectiveness is limited in detecting SNVs or smaller insertions and deletions. As a result, relying solely on CMA in resource-limited settings does not facilitate efficient resource utilisation (Park *et al.,* 2019). Adding to the complexity, the substantial cost linked to CMA poses a challenge, making its universal application for all individuals experiencing delayed development impractical in resource-constrained settings. To render the integration of routine CMA technology economically feasible, it becomes essential to implement rigorous gatekeeping procedures, which may involve trained geneticists and/or genetic counsellors. Regrettably, in resource-limited environments such as the African continent, the shortage of such expertise is apparent (Abacan *et al.,* 2019; Kamp *et al.,* 2021).

1.3.5. Sequencing Techniques

DNA sequencing involves identifying the order of nucleotides (adenine, guanine, cytosine, and thymine) in a DNA molecule. The two primary sequencing methods widely used are Sanger sequencing and next-generation sequencing, briefly outlined below.

i. Sanger Sequencing

Sanger sequencing, a pioneering method in DNA sequencing, relies on the enzymatic synthesis of complementary DNA strands to reveal the sequence of a target DNA fragment. The technique employs a DNA primer, DNA polymerase, and chain-terminating dideoxynucleosides, enabling the identification of nucleotide sequences through the termination of DNA synthesis at specific bases (Sanger *et al.,* 1977). Its strength lies in its ability to provide accurate and reliable sequence data, making it well-suited for sequencing short to medium-sized DNA fragments. This positions Sanger sequencing as an effective choice for phenotype-driven testing strategies, especially when focusing on specific genome segments. However, Sanger sequencing has notable limitations, such as cost, limited

scalability, and efficiency for large-scale genomic projects due to its labour-intensive and time-consuming nature. Nonetheless, in the context of DDs, Sanger sequencing proves valuable, particularly when clinicians can refine a patient's phenotype to identify a probable syndrome. An example of this is demonstrated at the Division, where Sanger sequencing is employed for the testing of single gene DDs and/or disabilities including Rett and Fragile X syndromes.

ii. Next-Generation Sequencing (NGS)

Next-generation sequencing is a massively parallel sequencing technology that offers ultra-high throughput, scalability, and speed. NGS signifies a ground-breaking leap in genomics, enabling researchers to sequence and analyse substantial volumes of genetic material within a shorter period than Sanger (Satam *et al.,* 2023). NGS can sequence complete genomes, encompassing the entirety of genetic material. Alternatively, it can be customised for specific regions of interest, such as sequencing of all protein-coding regions of the genome (whole exome) or focusing on a designated subset of individual genes (targeted gene panels) (Behjati & Tarpey, 2013).

Whole-genome sequencing (WGS)

Whole-genome sequencing (WGS) is a comprehensive method for sequencing all DNA materials within a sample (Wang *et al.,* 2020). Its significance lies in the comprehensive identification of disease candidate genes and their roles by detecting a broad spectrum of variants in both coding and non-coding regions (Lappalainen *et al.,* 2019) Predominantly employed in cases of diagnostic uncertainty, WGS offers a comprehensive analysis of an individual's entire genome, uncovering rare and novel variants often overlooked by more focused testing approaches. Beyond its capability to pinpoint disease-related genes, WGS plays a pivotal role in detecting various genetic variations, including SNVs, indels, CNVs, structural variants, and short tandem repeats (Coutelier *et al.,* 2022; Pálla *et al.,* 2023; Rajan-Babu *et al.,* 2021; Wheeler *et al.,* 2022). While WGS offers numerous advantages, challenges persist, primarily related to its cost. Despite significant decreases in per-sample costs over the years (Austin-Tse *et al.,* 2022) WGS remains prohibitively expensive and inaccessible for many individuals, particularly those in LMICs. Additionally, the analysis of WGS data is intricate and time-consuming, demanding the expertise of highly qualified bioinformaticians for accurate interpretation (Austin-Tse *et al.,* 2022; Oakeson *et al.,* 2017) Nevertheless, the extensive genomic coverage provided by WGS renders it an invaluable tool for advancing our understanding of genomics and its applications in medical research and diagnostics (Lin *et al.,* 2022).

Whole-exome sequencing (WES)

Whole exome sequencing (WES) is a genomic technique that selectively sequences the protein-coding regions of an individual's DNA, offering a comprehensive insight into the exome (Goh & Choi, 2012; Warr *et al.,* 2015). These protein-coding regions, constituting approximately 1% - 2% of the genome, but encompass approximately 85% of known Mendelian disease-causing variants (Rabbani *et al.,* 2014). Akin to WGS, WES can detect both SNVs and CNVs. Initially focused on SNV detection due to their simplicity, recent computational advancements have facilitated the inclusion of CNV analysis from WES data, enhancing practicality and cost-effectiveness, particularly in disorders where both SNVs and CNVs contribute to disease aetiology, such as DDs (Srivastava *et al.,* 2019; Yao *et al.,* 2019). Consequently, WES is recommended as a first-line investigation for patients with developmental delays or DDs. However, WES remains confined to research settings in most resource-constrained settings, including the South African State healthcare system, due to the absence of established infrastructure, the prohibitive cost of reagents, and the need for personnel training (Van Der Merwe *et al.,* 2022; Wiener *et al.,* 2023). Despite being relatively more cost-effective and generating less data than WGS, WES still yields a substantial output, necessitating additional time, advanced bioinformatics skills, and resources for data storage, interpretation, and analysis (Bartha & Győrffy, 2019).

Targeted NGS gene panels

Targeted gene panels scrutinize a specific set of genes with known functions and established associations with diseases (Brunelli *et al.,* 2019). There are two types of NGS targeted gene panels which are discussed below:

Capture-Based Targeted Gene Panels

Targeted capture based NGS gene panels are designed to sequence and analyse specific genes associated with phenotypes (Stoddard *et al.,* 2014). This approach generates data specific to the genes of interest only. Contrastingly to WES and WGS, targeted sequencing panels yield considerably smaller datasets rendering their output more manageable, easier to analyse and interpret (Rehm, 2013). Moreover, targeted capture-based sequencing panels offer a relatively cost-effective solution and typically yield higher sequencing coverage compared to the broader approaches of WES and WGS (Brunelli *et al.,* 2019; Rehm, 2013). However, this approach may miss significant variants in genes not included on the panel or fail to cover the entire gene region (Reid *et al.,* 2016) Additionally, the gene content of a capture-based targeted panels

varies significantly between laboratories, this may result in variable diagnostic yields (Kwan *et al.*, 2022).

Virtual Targeted Gene Panels

Virtual gene panels dynamically analyse a subset of clinically relevant genes from WES or WGS data (Wang *et al.*, 2019). In contrast to predefined capture-based gene panels, virtual panels are flexible, constructed from data obtained through WES and/or WGS. Virtual gene panels minimise initial time and analysis costs while preserving the option for a comprehensive examination of additional genes later on, without additional sequencing expenses (Molina-Ramírez *et al.*, 2022). This adaptability is crucial in resource-constrained settings where cost efficiency is paramount. However, it is important to acknowledge that the initial sequencing cost associated with WES/WGS could pose a challenge for adopting virtual gene panels, particularly in LMICs.

1.4. Guidelines and Recommendations for Genetic Testing in Individuals with DDs

Given the challenges associated with unravelling the genetic aetiology of DDs, a range of diagnostic guidelines and recommendations have emerged over time as genetic testing technology evolved and improved (Figure 1.1). These guidelines aim to improve the diagnostic rate of DDs.

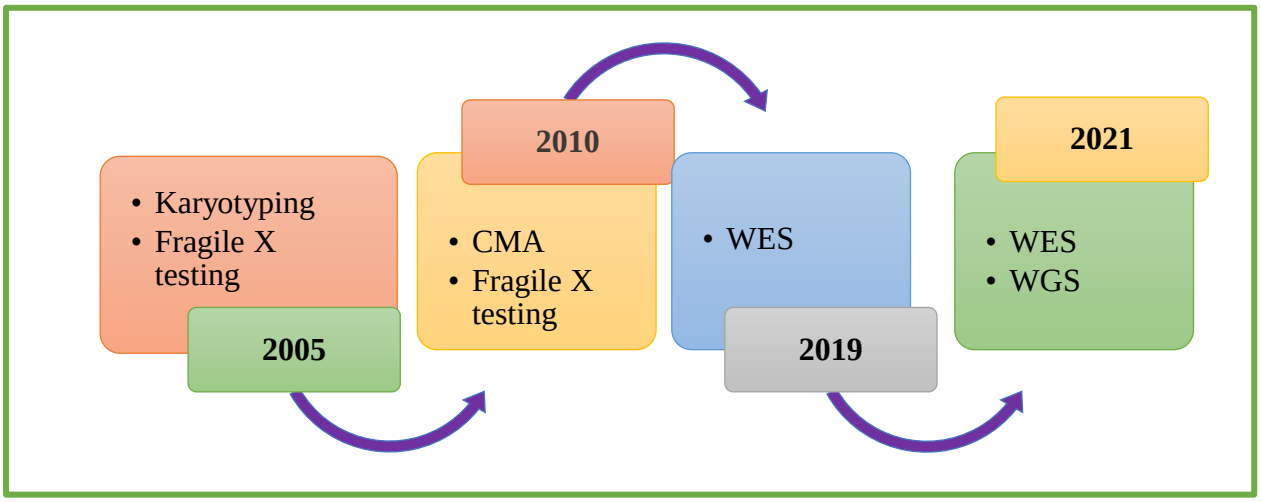


Figure 1.1. A schematic diagram of guidelines and recommendations for molecular diagnostic tests recommended to investigate the genetic aetiology of DDs between 2005 and 2023 (Manickam *et al.*, 2021; Miller *et al.*, 2010; Srivastava *et al.*, 2019).

In 2005, the standard approach to investigating DDs involved karyotyping and FXS analysis (Shaffer, 2005) However, a significant shift in genetic testing practices occurred in 2010 with the release of updated guidelines recommending the integration of genome-wide CMA along with FXS testing as the primary genetic investigation for DDs. This transition was substantiated by literature suggesting that CMA testing yielded a diagnostic rate of approximately 12%, surpassing the 5% yield of karyotyping in cases of unexplained DDs (Miller *et al.,* 2010).

Over subsequent years, technological advancements further refined diagnostic recommendations. In 2019, a multidisciplinary expert working group conducted a comprehensive assessment of WES for neurodevelopmental disorders. Following a literature scoping review, they found that WES exhibited a diagnostic yield ranging from 30% - 53%, outperforming CMA. Moreover, the ability to detect both SNVs and CNVs offered by WES makes it superior to CMA technology. Consequently, the expert group concluded that WES should be the preferred first-line investigation for neurodevelopmental disorders (Srivastava *et al.,* 2019). In 2021, the ACMG reinforced this perspective by publishing an evidence-based guideline for congenital anomalies, developmental delay, and intellectual disability. They recommended WES or WGS as the primary first-line tests for these conditions (Manickam *et al.,* 2021).

While these guidelines are widely adopted in diagnostic laboratories globally, resource-constrained settings encounter challenges in routinely adopting new and advanced technologies. In such settings, primary genetic tests employed in the investigation of DDs may still include karyotyping, FXS analysis, CMA, single-gene disorder tests, and MLPA. An example of this can be seen at the Division of Human Genetics, a major genetic service provider in Southern Africa. Here the standard testing protocol for patients with developmental delay and DDs historically involved karyotyping, MLPA, single gene disorder testing and FXS analysis, which collectively yielded a diagnosis in approximately 15% of the cases (personal communication., 2019), at the initiation of this study in 2019.

In 2020, aCGH was introduced as an additional test replacing routine karyotyping and increasing the diagnostic yield for this patient group to approximately 20% (Personal communication., 2019). Despite these advancements, a notable number of patients remain undiagnosed, underscoring the imperative for concerted efforts to implement international guidelines within the South African State Healthcare system. Comprehensive studies to assess

the feasibility and utility of locally viable options are essential to identify the most cost-effective and sustainable diagnostic strategy (Wiener *et al*., 2023).

1.5. Challenges in Implementing WES and/or WGS in Resource-Constrained Settings

At present, the ACMG recommends the use of WES and/or WGS as a primary diagnostic test in individuals with DDs. While WES and/or WGS have seen widespread adoption globally, resource-constrained settings like SA limit their application to research settings (Van Der Merwe *et al*., 2022). The implementation of routine diagnostic WES and/or WGS in these settings faces complex challenges including the cost of establishment of new infrastructure, greater data storage and security requirements, validation of laboratory processes, adaptation of policies, poor representation of African data in reference population databases and the provision of necessary expertise to diagnostic laboratory personnel for proficient operation and result interpretation (Gudmundsson *et al.,* 2022; Kamp *et al*., 2021; Krause, 2019; Lumaka *et al.,* 2022). The advantages of local infrastructure would include knowledge generation, skill development, student training, job creation, and improvement of local resources. Nevertheless, given the country's economic challenges, the feasibility of implementation is contingent on demonstrating cost-effectiveness, especially in a State public healthcare system where most resources are directed to infectious and communicable disease.

Moreover, widely employed public repositories for data predominantly consist of well-characterised reference datasets derived from populations of European ancestry, lacking substantial representation from individuals of continental African descent despite their status as the world's most diverse populations (Campbell & Tishkoff, 2008). Notably, reports indicate that a fraction less than 2% of the human genomes analysed to date pertain to African individuals (Sirugo *et al*., 2019). In 2019, Africans accounted for only 3% of the genome data employed in genome-wide association studies, a percentage that had alarmingly decreased to 1.1% in 2021 (Omotoso *et al*., 2022). Similarly, individuals of African descent constitute only 1.6% of the genotype data encompassing 487,000 individuals within the United Kingdom Biobank resource (Fatumo *et al.*, 2022). Additionally, research on polygenic risk scores reveals that individuals with African ancestry make up 17% (Duncan *et al*., 2019). The dearth of African population genetic data within these repositories presents a substantial obstacle in interpreting data from WES and/or WGS, leading to the classification of numerous variants as VUSs. The prevailing Eurocentric bias in genomics studies further hinders individuals of African descent from fully benefiting from crucial genomic findings that may not translate

uniformly across diverse populations, thereby contributing to existing health disparities (George *et al*., 2023; Ju *et al*., 2022).

Considering the challenges posed by resource constraints, particularly in settings like the South African State healthcare system, it appears crucial to explore cost and resource-effective diagnostic strategies. Among potential solutions, targeted gene panels may emerge as one viable option. Through focusing on specific genes associated with prevalent conditions, targeted gene panels not only mitigate the financial strain but also address the intricate issues of infrastructure establishment, policy adaptation, and personnel training. The streamlined nature of these panels not only simplifies the validation of laboratory processes but also ensures that the implementation aligns with the priorities of a state public healthcare system, where resources are often directed towards combating infectious and communicable diseases. With a focused and locally relevant approach, targeted gene panels may emerge as a beacon of feasibility, offering a pathway for South African patients with DDs to leverage the benefits of genetic diagnostics.

1.6. Targeted Gene Panel Design Considerations and Challenges.

Over the course of many years, there was a lack of international consensus or standardised criteria concerning the genes to be featured in phenotype-related panels. However, in 2017, the ClinGen consortium proposed recommendations and an evidence-centric framework to assess the clinical validity of gene-disease correlations (https://clinicalgenome.org/) (Strande & Berg, 2016). In an effort to establish consistency regarding the genes included in diagnostic panels, the ACMG has provided guidelines and recommendations (Bean *et al.*, 2020). These standards address considerations such as clinical validity, analytic validity, and clinical utility. Assessing clinical validity involves a thorough evaluation of the evidence supporting the association between genetic variants in a given gene and the corresponding phenotype. This evaluation process is an ongoing process and incorporates factors like clinical studies and segregation analysis. Various gene-disease curation working groups and platforms have been established to assess gene-disease validity, most notably the ClinGen gene-disease validity curation resource (Strande & Berg, 2016) and PanelApp (Martin *et al.*, 2019).

1.6.1. ClinGen Gene-Disease Validity Curation and PanelApp

The ClinGen gene-disease validity curation resource (https://clinicalgenome.org/curation-activities/gene-disease-validity/), first published in 2017, is managed by the Clinical Genome

Resource (ClinGen) to assess and curate the association between specific genes and diseases (Strande *et al.*, 2017). During the gene-disease validity curation process, experts systematically review existing evidence related to a particular gene-disease association. The curation outcomes are categorized into various levels of evidence, ranging from "Definitive" or "Strong" to "Limited" or "Disputed," providing an understanding of the current scientific consensus regarding the gene-disease association of a specific gene in a particular clinical indication. The ACMG recommends that a diagnostic gene panel should consist of genes associated with Mendelian disorders. These genes should meet the criteria for definitive, strong, or moderate evidence for association with the disease, as outlined by ClinGen (Bean *et al.*, 2020).

PanelApp (https://panelapp.genomicsengland.co.uk/), launched November 2019, serves as a platform designed to facilitate gene panel curation and establish gene-disease association pertaining to a specific clinical indication through crowdsourcing of expert reviews (Martin *et al.*, 2019). PanelApp employs a traffic light system to categorize genes within a panel. This system utilizes a combination of case-level and experimental data to ascertain which genes possess sufficient evidence for analysis in specific clinical indications. A green rating within PanelApp signifies genes with substantial clinical evidence, suitable for diagnostic reporting (Diagnostic genes). This green rating aligns with the a "Definitive" or "Strong" ClinGen rating. The amber rating is assigned to genes considered "borderline," lacking sufficient evidence for confident clinical use, while the red rating denotes genes with only low levels of evidence.

1.7. Genes Intolerant to Variation and their Contribution to DDs

The evolution of NGS has increased the rate at which the genetic basis of a patient's disease is determined (Sun *et al.*, 2021). However, in cases where a patient lacks a well-established pathogenic variant, identifying which among the numerous variants identified is contributing to their condition becomes challenging. To address this challenge, prioritisation of variants is commonly undertaken, often by ranking the genes associated with these variants (Marian, 2020). Within the entire human genome, certain genes are more prone to harbour functional variations beyond what is expected by chance, termed tolerant genes (Petrovski *et.al*., 2013: Karczewski *et al*., 2020). Conversely, there are genes, referred to as intolerant genes, that exhibit fewer functional variations (Petrovski *et al.*, 2013). This notion has motivated the development of several measures of gene variation intolerance.

There are several measures developed to assess whether a gene is tolerant or intolerant of common functional variation including gnomAD pLI, gnomAD missense Z-scores, residual

variation intolerance score (RVIS), missense tolerance ratio (MTR) and loss-of-function observed/expected upper bound fraction (LOEUF). The majority of these measures are based on the number of unique variants observed in a gene and the number expected under a mutation model for the gene. Expected values are calculated differently based on the size of the gene and frequency of variation, the amount of putatively neutral variation, or the overall number of observed variants. Most Mendelian and severe complex phenotype diseases are known to be caused by variants in variation/mutation intolerant genes (Samocha *et al.*, 2014; Karczewski *et al.*, 2020; Chopra *et al.,* 2021; Agarwal *et al.,* 2023). Variants in genes intolerant to variation are construed as more likely to cause disease, while those in genes tolerant to variation are deemed less predisposed to causation.

The RVIS ranks genes by probability of harbouring more, or less, functional genetic variation than expected highlighting genes intolerant to common functional variation. Genes with positive scores have more common functional variation, while negative scoring genes are less tolerant having reduced associated common functional variation (Petrovski *et al.,* 2013). The pLI, quantifies the likelihood that a gene is intolerant to a mutation that produces LoF in the protein: genes with scores above 0.9 are considered the most intolerant of LoF variation and most evolutionarily constrained (Lek *et al*., 2016; Karczewski *et al*., 2020). The Missense Z score assesses intolerance to missense mutations. Genes with a Z score >3.09 are considered intolerant to missense variation. Additionally, the MTR compares observed and expected missense variants, with a lower MTR suggesting reduced tolerance and potential functional consequences (Silk *et al.,* 2019).

Petrovski and colleagues conducted an analysis of the RVIS across genes responsible for different classes of genetic diseases. Their findings indicated that genes intolerant of variation are more inclined to contribute to cardiovascular disease, skeletal issues, and DDs. Notably, genes associated with DDs exhibited greater constraint, with approximately half of all Online Mendelian Inheritance in Man (OMIM) genes causing DDs being situated within the 25th percentile of intolerance, while only 10% are located among genes surpassing the 75th percentile (Petrovski *et al.,* 2013). This observation has been corroborated by several other studies, demonstrating causal variation in genes intolerant to variation in DDs (Beck *et al.,* 2021; Gillentine *et al.,* 2021; Hamanaka *et al.,* 2022).

Moreover, DDs encompass a spectrum of complex conditions impacting various aspects of development, manifesting from the embryonic period to childhood. Any disruption to the

intricate sequence of events in normal development can lead to dysfunction in various systems, resulting in a DD phenotype (Cardoso-Moreira *et al.,* 2019; Griffin *et al.,* 2022; Parenti *et al.,* 2020). Genes associated with DDs play crucial roles in orchestrating complex developmental biological processes vital for the formation and functioning of an organism (Iossifov *et al.,* 2015; Parenti *et al.,* 2020). Their functional networks involve multiple pathways including chromatin modification, organ development, signal transduction pathways, synaptogenesis, cell proliferation and differentiation (Griffin *et al.,* 2022; Hsueh, 2019; Jiang *et al.,* 2022; Mossink *et al.,* 2021; Stessman *et al.,* 2017). This group of genes are evolutionarily conserved and exhibit lower tolerance to variation (Cardoso-Moreira *et al.,* 2019; Petrovski *et al.,* 2013).

Notable, within this group of genes (intolerant to variation and associated with DDs), there exists a subset of genes enriched with disease-causing *de novo* mutations (DNMs), and literature reports that causal variants in DNM enriched genes are implicated in at least 42% of DD cases (McRae *et al.,* 2017). In light of this, screening for causal variation-both inherited and *de novo* in these genes could present a potentially feasible DD diagnostic strategy, especially in resource-constrained settings.

1.8. Genes Enriched for DNMs in DDs and Developmental Delay

Proband-parent trio-based investigations have unveiled genes exhibiting substantial enrichment for DNMs in individuals with DDs and/or intellectual disabilities. In 2012, de Ligt *et al* . identified 24 genes enriched for DNMs in severe DD cases (de Ligt *et al.,* 2012), while a meta-analysis conducted in 2016 identified 30 genes enriched for DNMs in DD cases (Lelieveld *et al.,* 2016). Notably, 20 genes overlapped between the 2012 and 2016 studies. In a 2017 study, Lelieveld and colleagues identified 15 genes enriched for DNMs in a DD/ID cohort, with five genes overlapping de Ligt *et al* .'s findings and seven overlapping with Lelieveld *et al* .'s 2016 study (Lelieveld *et al.,* 2017). To further discover genes enriched for DNMs in DD and neurodevelopmental disorder cohorts, McRae *et al* . analysed 4,224 published DNMs from WES and/or WGS studies, identifying 94 genes with damaging DNMs. Eighty of these genes were previously reported (McRae *et al.,* 2017). In the largest meta-analysis of DNMs in DD cohorts, Coe and colleagues examined 12,172 DNMs from the denovo-db database across 17 exome sequencing studies. They identified 253 genes enriched for DNMs, with 80 genes overlapping McRae *et al* .'s 2017 findings (Coe *et al.,* 2019). While considerable overlap exists between genes enriched for DNMs in these studies, a core set emerges in patient cohorts with DDs. Thus, targeting variation in genes enriched for DNMs and intolerant to variation in the context of DDs could present an appealing diagnostic strategy,

particularly in settings where the practicality of comprehensive WES/WGS analysis may be limited.

1.9. Developmental Delay and DDs in the Division of HumGen

The Division of Human Genetics is one of the major service providers of clinical and laboratory-based genetic tests in South Africa. Over the years, it has actively engaged in diverse research endeavours related to developmental delay and DDs (Essop and Krause, 2013; Flynn *et al.,* 2021; Cottino *et al.,* 2022; Baine-Savanhu *et al.,* 2023; Wiener *et al.,* 2023). Building upon these research initiatives, the current PhD thesis focuses on two cohorts affiliated with the Division: The DDD-Africa cohort and the HumGen referral cohort which are explained in detail below:

1.9.1. The Phenotyped DDD-Africa Cohort A

The Deciphering Developmental Disorders in Africa (DDD-Africa) study project (https://h3africa.org/index.php/ddd-africa/), is one of the biggest DDs initiatives in Africa. This project is focused on enhancing the scope and delivery of diagnostic and clinical genetic services in sub-Saharan Africa. It aims to achieve this by systematically describing the phenotypes and delineating the underlying genetic aetiology of DDs with a probable genetic origin in a cohort of 500 patients who presented with DDs and their parents. These efforts are anticipated to contribute to a more comprehensive understanding of the role played by genetic factors in DDs within the African context. The DDD-Africa cohort consists of South African patients with DDs of unknown genetic aetiology who tested negative for the standard DD diagnostic tests at the Division. This group of patients consists of individuals already integrated into our healthcare system, receiving care at the participating genetic clinics within the Division. As they undergo assessments at these clinics, our in-house medical geneticists and/or genetic counsellors conduct thorough and comprehensive systemic clinical phenotyping. Consequently, detailed phenotypic data, represented in the form of Human Phenotype Ontology (HPO) terms, is readily accessible for these patients. This information serves as a valuable resource for facilitating variant interpretation and correlating clinical phenotypes.

1.9.2. The Unphenotyped HumGen Referral Cohort B

This cohort consists of DNA samples of singletons patients who presented with developmental delay and were referred for FXS molecular testing at the Division, by external doctors between 2017 and 2018. Following FXS testing, all patients in this cohort tested negative, and some also received negative results for additional reflex tests conducted after the initial FXS testing including karyotyping and MLPA in some cases. It noteworthy to mention that this cohort of patients had not been assessed by a medical geneticist or attended a genetic counselling prior to their FXS testing; instead, they were selected as they were referred by an external attending doctor. As a result, detailed phenotypic data and family history information for this patient group were not obtained prior to their FXS testing. In terms of their indication for testing: patients included in this cohort have limited clinical features written on the request form. On average, two clinical features would be listed and usually generic. Attending doctors normally refer patients who present with unexplained developmental delay for FXS testing as a routine practice, aiming to exclude FXS as a first line test. However, the global diagnostic yield of FXS testing is generally low ranging from 0.5% to 5% (Weinstein *et al.,* 2017). This comparable rate is mirrored in a South African context, as evidenced by (Essop & Krause (2013) who reported a diagnostic rate of 5.7% in a comprehensive 20-year review study involving individuals referred for FMR1-related disorders at the Division of Human Genetics. In many cases, patients receive negative results for FXS testing and consequently lack a molecular diagnosis. This is especially prevalent in resource-constrained settings where additional diagnostic assessments may either be prohibitively expensive or unavailable. A comparison of the two cohorts is shown in Table 1.1 below.

Table 1.1. Overview of DDD-Africa cohort A and HumGen referral cohort B characteristics.

	DDD-Africa cohort A	**HumGen referral cohort B**
Description	DDD-Africa study patients	Random FXS molecular testing referrals
Family type	Trios, Duos & singletons- predominantly trios	Singletons
Indication for testing	Developmental disorders	Developmental delay
Referral status	Patients already receiving genetic services and clinical phenotyping offered by the Division	Patients not in the Division system – external referrals
Clinical assessment	Reviewed at the participating genetic clinics as part of the genetic workup offered to patients	No prior assessment at genetic clinics
Clinical phenotyping	Thorough clinical phenotyping by medical geneticists	Limited clinical data from referring doctors
Previous genetic testing	Previously tested negative for DD using in-house standard diagnostic tests	Previously tested negative for FXS
Availability of phenotypic and family history data	Detailed clinical features (HPO terms), and family history information available in the majority of patients	Limited clinical features and no family history information

1.10. Study Rationale

Developmental disorders are life altering and debilitating conditions estimated to be present at high rates in sub-Saharan Africa. At least 50% of DDs have a genetic origin yet discerning their genetic origin proves challenging mainly due to their nonspecific presentation and substantial genetic heterogeneity (Srour and Shevell, 2014). While international guidelines advocate for WES and/or WGS as first-tier genetic diagnostic tests for DDs, the practical implementation in resource-constrained settings, such as the South African State Healthcare, faces hindrances related to cost and the demand for specialized infrastructure and resources (Kamp *et al.*, 2021; Krause, 2019). In light of this, there arises a pressing need for cost-effective diagnostic strategies tailored to DDs within resource-limited settings. Among these strategies, a targeted gene panel may present a viable option. Targeted gene panels are known for their cost-effectiveness, requiring fewer resources and quicker turnaround times as compared to WES and/or WGS. The adoption of such a strategy could potentially bridge the diagnostic gap for DDs in regions with limited resources and financial constraints.

At least 42% of disease-causing variants in DDs are in genes intolerant to variation and enriched for DNMs (McRae *et al.*, 2017). Despite their known association with DNMs, these core genes also harbour inherited causal variations which significantly contribute to DDs. Considering this, targeting variants within these genes could present a potentially cost-effective and feasible diagnostic strategy, particularly in resource-constrained settings where routine WES and/or WGS may yet not be practical. Therefore, the research questions for this study are as follows: (1) What is the role of mutations in DNM-enriched genes in the aetiology of DDs in South African patients? and (2) How effective and feasible is a diagnostic strategy employing a panel of genes enriched for DNMs in identifying the genetic aetiology of DDs?

1.10.1. Aims of the Study

1. To determine the diagnostic yield of a DNM-enriched gene panel in undiagnosed South African patients with DDs and developmental delay

2. To evaluate the feasibility and utility of a targeted DNM enriched gene panel diagnostic testing strategy for DDs in resource-constrained settings

1.10.2. Objectives of the Study

1. To design a custom virtual NGS panel of known genes enriched for DNMs that could be used for disease-causing variant screening in undiagnosed South African patients with DDs.

2. To analyse a panel of DNM enriched genes designed in objective 1, in a cohort of patients with DDs that has undergone systematic clinical phenotyping and WES as part of the DDD-Africa project.

3. To perform WES in 96 SA patients referred for FXS testing at the Division who tested negative for FXS and selectively analyse a panel of DNM enriched genes designed in objective 1.

4. To explore clinical factors influencing the diagnostic yield of DNM enriched genes in South African patients with DDs.

5. To explore technical factors influencing the diagnostic yield of DNM enriched genes in South African patients with DDs.

Chapter 2

Patients & Methods

2. METHODS

This chapter will describe methods that were employed to achieve the aims and objectives outlined in Chapter One. The chapter starts by describing the two cohorts chosen for this study and samples and NGS data utilised. This is followed by the design of the virtual DNM-enriched gene panel, as well as the analysis of the DNM-enriched gene panel in two cohorts with DDs: DDD-Africa and HumGen referral cohorts Additionally, the chapter describes assessments employed to explore technical and clinical factors influencing the diagnostic yield of DNM in cohorts with DDs. Lastly, we evaluate the utility and feasibility of a DNM-enriched gene panel strategy in a routine diagnostic setting in a LMICs.

2.1. Samples and NGS Data Used in this Study

2.1.1. DDD-Africa Cohort A

For this cohort, the first 96 patients with DDs (consisting of 49 trios, 38 duos and 9 singletons), from the DDD-Africa project cohort (refer to Section 1.9.1) were included. These individuals already had both WES data and comprehensive phenotypic data. In Appendix A, the inclusion and exclusion criteria for DDD-Africa are presented.

2.1.2. HumGen Referral Cohort B

For this cohort, DNA samples were obtained from a randomly selected group of 96 South African patients with developmental delay. These individuals had been referred for FXS molecular testing at the Division between 2017 and 2018, as detailed in the cohort description Section 1.9.2.. Notably, the cohort comprised only singletons, with minimal available clinical information, typically an average of two clinical features noted, with developmental delay being the primary indicator for FXS testing in this cohort.

2.1.3. Anonymity and Confidentiality

All DNA and WES data samples for patients included in both cohorts A and B were anonymised and a unique bar code was assigned to each sample. The assigned code only retained relevant information and not the participant's name. Patients' identifying, and personal information for both cohorts were kept in a password-protected REDCap-and Microsoft Excel databases. The two databases were only accessible to individuals directly involved in this study and the umbrella projects

2.2. Ethics Clearance

The present study was conducted as a sub-study of three larger projects within the Division (Ethics clearance numbers M160830, M180506, and M180678). Ethical clearance (M200440)

for the current study was obtained from the University of the Witwatersrand Human Research Ethics Committee (Medical). Ethical clearance certificate for this study can be found in Appendix B

2.3. Designing a Custom Virtual DNM-Enriched Gene Panel

The workflow employed for the design and curation of a custom virtual DNM-Enriched gene panel is illustrated in Figure 2.1 and elaborated upon in detail below:

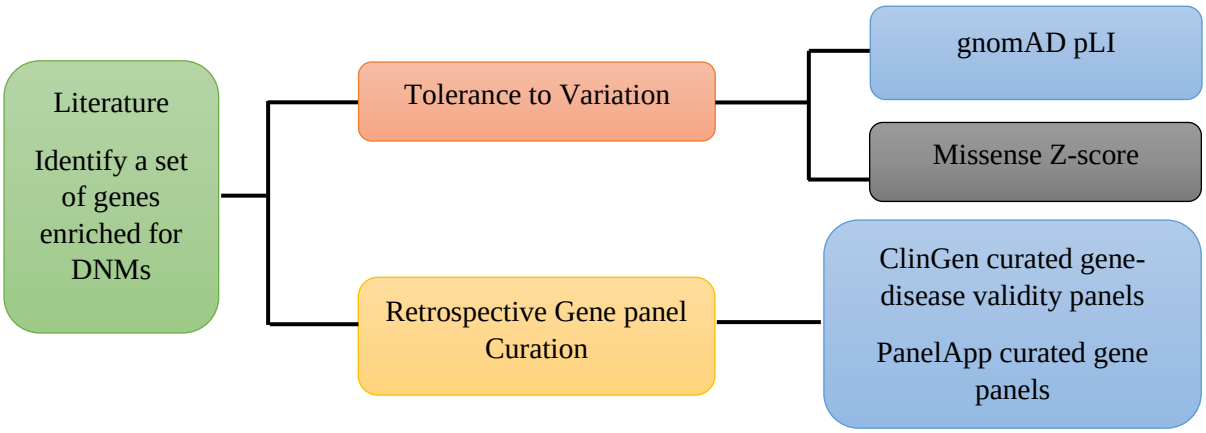


Figure 2.1. A workflow employed for the design and curation of a custom virtual targeted gene panel designed and analysed in this study.

2.3.1. Literature Review

A comprehensive literature review was conducted to compile a gene panel enriched for DNMs in cohorts with DDs, developmental delay, and/or intellectual disability. The gene list comprised genes that demonstrated enrichment for DNMs in DD cohorts, as reported in 11 studies conducted between 2012 and 2019: (Coe *et al.*, 2019; de Ligt *et al.*, 2012; De Rubeis *et al.*, 2014; Iossifov *et al.*, 2012, 2015; Lelieveld *et al.*, 2016, 2017; O'Roak *et al.*, 2011; Rauch *et al.*, 2012; Sanders *et al.*, 2012; Wilfert *et al.*, 2017; Yuen *et al.*, 2016). In total, these studies identified 416 genes enriched for DNMs. However, there was considerable overlap of genes reported in multiple studies. After removing the duplicates, a final set of 285 genes enriched for DNMs was identified.

2.3.2. Gene Tolerance to Variation Considerations

The assessment of each identified gene's tolerance to functional genetic variation was conducted using two distinct metric scores obtained from gnomAD: the gnomAD pLI score and the gnomAD missense Z-score (Lek *et al.*, 2016; Karczewski *et al.*, 2020). Genes with a high pLI score (pLI $\geq$ 0.9) are deemed most likely to exhibit intolerance to loss-of-function

(LoF) variations, while those with a pLI score ≤ 0.1 are considered tolerant to such variation. Furthermore, genes with a missense Z-score ≥ 3.09 are classified as intolerant to missense variation (Lek *et al.*, 2016; Karczewski *et al.*, 2020).

2.3.3. Gene-Disease Association Considerations

The selection of genes for inclusion in a diagnostic gene panel adheres to the guidelines outlined by ACMG-AMP (Bean *et al.,* 2020). A pivotal criterion in this process is ensuring there is sufficient evidence linking the gene to an associated disorder and/or phenotype. In this study, we evaluated the gene-disease association by examining a gene's gene-disease validity curation status on ClinGen (https://clinicalgenome.org/curation-activities/gene-disease-validity/) (Strande & Berg, 2016), and PanelApp (https://panelapp.genomicsengland.co.uk/) (Martin *et al.,* 2019) (accessed 15th October 2023). The specific gene panels in ClinGen and PanelApp against which the 285 DNM-enriched genes were cross-checked against are shown in Table 2.1 below.

Table 2.1. ClinGen and PanelApp gene panels used to assess the gene-disease validity curation status of the 285 identified DNM-enriched gene

ClinGen	**PanelApp**
• Hearing Loss • Inborn Errors of Metabolism • Neurodevelopmental Disorders • Intellectual Disability and Autism • Rett and Angelman-Like Disorders • Skeletal Disorders • Syndromic Disorders • Brain Malformations	Paediatric disorders panel (consisting of the following panels) • The Development Disorder Genotype - Phenotype Database • Intellectual disability - microarray and sequencing • Likely inborn error of metabolism - targeted testing not possible • Skeletal dysplasia • Monogenic hearing loss • Paediatric disorders - additional genes • Ophthalmological ciliopathies • Renal ciliopathies • Neurological ciliopathies • Limb disorders and Clefting • Skeletal ciliopathies

2.4. Data Generation and Processing of the Phenotyped DDD-Africa Cohort A

2.4.1. WES Data Generation of the Phenotyped DDD-Africa Cohort A

Whole exome sequencing of this cohort was performed at the Wellcome Sanger Institute (United Kingdom), as part of the DDD-Africa study project. Briefly, WES was performed using ISC-Twist library preparation and sequenced using Illumina NovaSeq 6000 paired-end sequencing (Illumina, San Diego, CA, USA), aiming for a targeted sequencing coverage of 40x to optimise cost-effectiveness.

2.4.2. Alignment and Variant Calling of Data Generated from the DDD-Africa Cohort A

Mapping of short-read sequences to the human reference genome, hg38 (GRCh38) was performed using the Burrows-Wheeler Alignment tool according to the genome analysis toolkit (GATK) best practices. Variants were called from BAM files using GATK (https://gatk.broadinstitute.org/hc/en-us) and SAMtools (https://www.htslib.org/) by the DDD-Africa team.

2.4.3. Extraction of Variants within the DNM Enriched Genes from WES Data in the DDD-Africa Cohort A

The candidate compiled a list of the 285 DNM enriched genes and gave it to the DDD-Africa bioinformatician. Subsequently, the DDD-Africa bioinformatician employed a bioinformatics script to extract called variants within the identified 285 DNM-enriched genes for each patient. The output comprised variants within these genes presented in a VCF file. This VCF file was then provided to the candidate for subsequent downstream analysis. A comprehensive description of the data analysis workflow can be found in Section 2.6.

2.5. Data Generation and Processing of the Unphenotyped HumGen Referral Cohort B

A schematic representation of the laboratory and data processing procedures carried out to achieve this objective is provided in Figure 2.2 below.

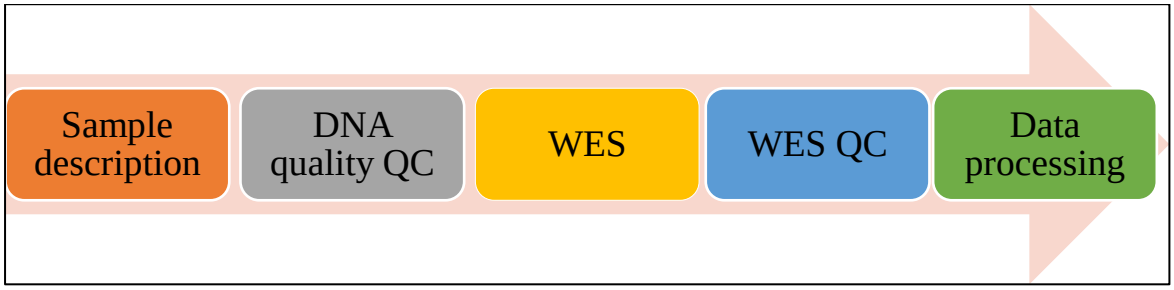


Figure 2.2. A schematic workflow employed for performing WES in the HumGen referral cohort B.

2.5.1. Sample Description

Archived DNA samples of patients who were referred for FXS testing at the Division were utilised for this cohort. The DNA samples were retrieved from the archive DNA storage at the Division, and a total of 10 µl was aliquoted for each sample.

2.5.2. DNA Quantification and Quality Assessment

The quality and integrity of genomic DNA were assessed using multiple techniques. Initially, the NanoDrop® ND-2000 spectrophotometer (Thermo Fisher Scientific, California, USA) and gel electrophoresis were employed. The NanoDrop® ND-2000 spectrophotometer measured DNA purity using A260/230 and A260/280nm absorbance ratios, where pure nucleic acid exhibits ratios of approximately 1.8 and 1.8-2.2, respectively, with low values indicating contamination. Agarose gel electrophoresis was used to evaluate the integrity of double-stranded DNA, ensuring its non-degraded state. This involved loading DNA and loading dye along with a 1-kilobase ladder into precast wells on a 1% agarose gel, which was subsequently immersed in 1x EDTA buffer and run at 100 volts until separation was achieved. Fragments were visualised and imaged using an ultraviolet transilluminator (Omega Fluor Plus Systems, Aplegen Inc., California, USA). Samples with intact DNA observed on the agarose gel and concentration of at least 100 ng/µl, along with A260/230 and A260/280 ratios within the pure range, were selected for further processing.

DNA quality was again, reassessed during the WES library preparation step. Genomic DNA was quantified using double-stranded DNA Broad Range (BR) and High Sensitivity (HS) assay kits for the Invitrogen Qubit® 4.0 Fluorometric quantification system (Invitrogen by Thermo Fisher Scientific, South Africa). The HS kit accurately quantifies samples with concentrations ranging from 10 pg/µl to 100 ng/µl, while the BR kit accurately quantifies samples with concentrations from 100 pg/µl to 1000 ng/µl. Subsequently, all DNA samples were normalised to a concentration of approximately 10 ng/µl. A total of 10 µl (range: 80ng – 100ng), of each sample was used as a template to prepare WES libraries for the 96 patients.

2.5.3. Whole Exome Sequencing

WES of this cohort was carried out by the candidate at the Division. The strategy of WES followed by selective analysis of the DNM-Enriched virtual gene panel was decided upon after a cost analysis. WES and a research custom targeted DNM-Enriched gene panel would have cost similar; it was therefore decided on the WES-based strategy for cohort B so that additional analyses could be performed on this dataset at a later date. WES library preparation was

performed using the Ion AmpliSeq™ Exome RDY Kit 1x8, Ion AmpliSeq™ Exome RDY Kit 4x2 and the Ion AmpliSeq™ Exome Library Preparation Kit Plus (Thermo Fisher Scientific, California, USA), according to the manufacturer's protocol. The WES wet-lab workflow employed in this cohort is shown in Figure 2.3 below.

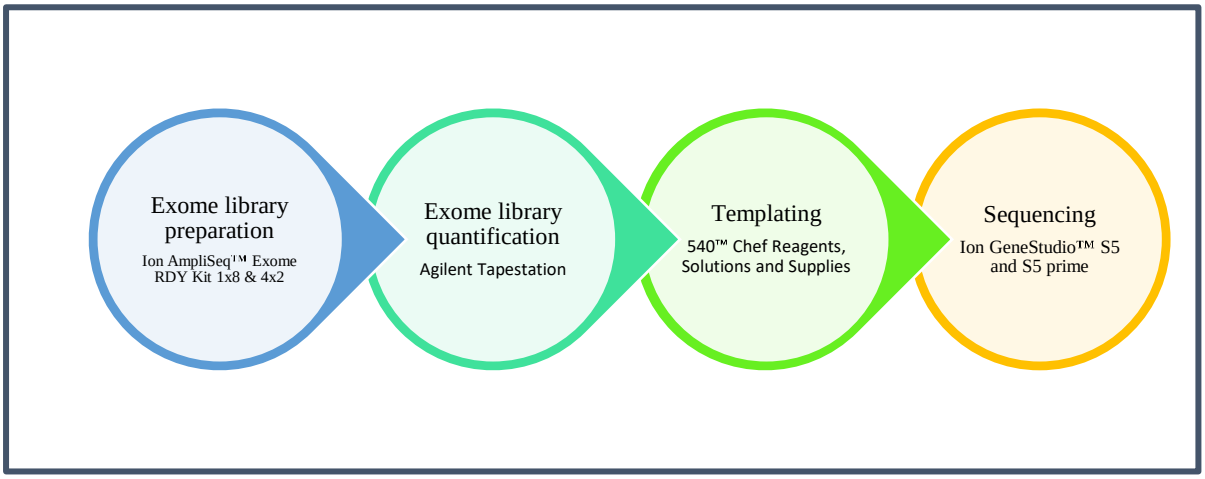


Figure 2.3. Laboratory workflow employed for the WES of the HumGen referral cohort B.

i. Whole Exome Library Preparation

A total of 80ng-100ng of each DNA sample (equivalent to approximately 10ul of a 10 ng/µl sample), was used as a template in the exome library preparation PCR protocol. Briefly, DNA was fragmented through a thermal cycler step on the SureCycler 8800 Thermal Cycler (Agilent Technologies, CA, USA). Amplicons were then treated with FuPa Reagent (Thermo Fisher Scientific, California, USA) to partially digest the primers and phosphorylate the 3'-ends through a thermal cycler step on the SureCycler 8800 Thermal Cycler (Agilent t, CA, USA), and ligated to the Ion Xpress™ adapter 1-16 barcodes (Thermo Fisher Scientific, California, USA). Barcoded libraries were purified using Agencourt™ AMPure™ XP Reagent (Beckman Coulter Life Sciences, USA). Purified barcoded libraries were amplified and eluted in low TE buffer (Thermo Fisher Scientific, California, USA). Barcoded amplicons were quantified, and their fragments analysed using the HS kit of the Qubit®3.0 Fluorometer (Thermo Fisher Scientific, California, USA), and the Agilent 4150 Tape Station System (Agilent Technologies, California, USA), respectively.

ii. Library Templating and Sequencing

The addition of barcoded adapters during the library preparation step allows multiple samples to be sequenced together in a single sequencing run. Two exome libraries were combined in equimolar amounts (approximately 50pM), and loaded onto the Ion Chef™ liquid handler for

automated template preparation and enrichment using Ion 540™ Chef Reagents, Solutions and Supplies (Thermo Fisher Scientific, California, USA). During this step, the multiple copies of a DNA fragment are created on an Ion Sphere Particle or ISP. Enriched ion sphere particles were loaded onto an Ion 540™ Chip and sequencing was performed on the Ion GeneStudio S5 and Ion GeneStudio S5 Prime platforms (Thermo Fisher Scientific, California, USA). The sequencer was first initialised. This process involved placing all the buffers (wash buffer and cleaning solution) and sequencing cartridges required for the sequencing in the sequencing machine. The machine then used those reagents to pre-run with the previously used sequencing chip while checking if enough appropriate reagents were loaded for the sequencing run. Post-initialisation, the previously used Ion 540™ sequencing chip was removed, and the newly prepared Ion 540™ chip (Thermo Fisher Scientific, California, USA) was loaded onto the sequencer and the sequencing was initiated and run to completion. The Ion Torrent sequencing platforms utilize various sequencing chips with different capacities for sequencing purposes. In this study, we used the Ion 540™ chip. The higher capacity of this chip enabled us to sequence two samples simultaneously. Consequently, we conducted a total of 48 sequencing runs using 48 Ion 540™ chips to sequence 96 patient samples in this study.

iii. WES Quality and Run Results

Quality control and data management are critical in ensuring data integrity and consistency. Upon the completion of the sequencing runs, the Torrent Suite™ Software generates a comprehensive 'Run Report' that encompasses essential statistics for QC and sequencing runs. This report can be divided into two distinct sections, namely the QC and run statistics prior to alignment (Table 2.2) and the QC and run statistics after alignment (Table 2.3). The pre-alignment statistics include amongst others the chip loading statistics, the number of total bases generated per run and read length. Tables 2.2 below shows the pre- and post-alignment metrics (and their descriptions), evaluated in this study.

Table 2.2. Pre-alignment QC metrics and run statistics evaluated in this study.

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

Post-alignment of sequencing data to the human reference sequence resulted in an alignment summary found within the 'Run Report'. This summary provided alignment QC metrics including the total number of aligned bases, number of aligned and unaligned reads and showed

the number of reads of a certain read length obtained under different quality filters. Evaluated post-alignment QC metrics and their descriptions are shown in Table 2.3.

Table 2.3. Post-alignment QC metrics and run statistics evaluated in this study.

Metric	**Description**
Reference Coverage	The ratio of the total aligned bases divided by the number of bases in the reference sequence. Reference coverage does not account for enrichment done to selectively amplify a subset of the reference sequence.
Total Aligned Bases	The number of filtered and trimmed aligned base pairs reported in the output BAM file that are aligned to the reference sequence, excluding the library key, barcodes, and 3' adapter sequences.
Alignment plot	A plot of the number of aligned reads (blue) and unaligned (purple) by position in an aligned sequence...
Total Reads	The total number of reads after filtering
Aligned Reads	The number of reads that align to the reference sequence expressed as a total count and percentage of the total aligned reads
Unaligned Reads	The number of reads that do not align to the reference sequence expressed as a total count and percentage of the total reads.
AQ17 Total Bases	The total number of bases over all positions that align with an error rate of 2% or less
Total Number of Bases (bp)	The total number of bases over all positions that align with a given error rate. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)
Mean Length (bp)	The average length, in base pairs, for aligned reads at a given error rate. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)
Longest Alignment (bp)	The maximum sequence read length for a given error rate. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)
Mean Coverage Depth (x)	The ratio of the total aligned bases at a given error rate to the size

2.5.4. Data Processing of HumGen Referral Cohort B

Data processing for this dataset was performed using the Ion Torrent Suite™ software (version 5.18). The Ion Reporter™ software is a collection of bioinformatics tools that allow for data analysis, annotation, filtering, and the export of sequencing data. Data processing in this cohort was divided into two major steps: primary and secondary processing. Primary data processing included signal analysis, base calling and scoring base quality.

2.5.5. Alignment and Variant Calling of the HumGen Referral Cohort B

Post primary data processing, raw sequencing reads were aligned to the human genome assembly (GRCh37/hg19) using the Torrent Mapping Alignment Program (TMAP) module (version 5.18), a sequence alignment software program optimised specifically for Ion Torrent data. The TMAP uses the Burrow Wheeler alignment fast map routine to map reads and applies post-processing quality control checks. Post alignment, the Torrent Suite performed variant calling using the Torrent Variant Caller (version 5.8), optimised for Ion Torrent data. Following variant calling, the variant caller generated a report of SNVs, and INDELS obtained from the dataset and the output was a list of variants in BAM and VCF files. The BAM files were downloaded and stored in a hard disk for visualisation through the IGV and the VCF files were uploaded onto the Ion Reporter™ for filtering.

2.5.6. Extraction of Variants Within the DNM Enriched Genes from WES Data in the HumGen Referral Cohort B

A step-by-step data analysis pipeline was developed for the processing of WES data generated from the HumGen referral cohort B which underwent WES on the in-house Ion GeneStudio™ S5 System. To accomplish this, the use of an Ion Reporter™ custom filter chain was employed (Figure 2.6 below). A filter chain consists of a series of filters and analysis steps that can be applied to the variants identified in an analysis that are used to narrow the analysis results to only the variants of interest. By employing a filter chain, users can save time and ensure consistency by leveraging preconfigured settings tailored to their specific needs

This involved uploading a text file containing the 285 gene names to the "Create Filter Chain" dialog box of the Ion Reporter™ software. Figure 2.4 below illustrate a flow diagram of how the DNM-Enriched gene panel filter chain was created. Within this filter chain, the software was instructed to filter variants solely within the uploaded gene names, catering to the study's specific interest. After designing the custom filter chain for DNM-enriched genes, the VCF

files generated by the Ion Torrent Suite were uploaded to the Ion Reporter™ software (version 5.18) for filtering. During this step, the custom-designed DNM-Enriched gene filter chain was employed to filter variants of the study's interest. The resulting output was a VCF file containing variants specifically found within the 285 DNM-enriched genes. This VCF file would undergo further annotation using the web browser version of the Ensembl variant effect predictor tool (McLaren *et al.*, 2016) (https://www.ensembl.org/info/docs/tools/vep/index.html).

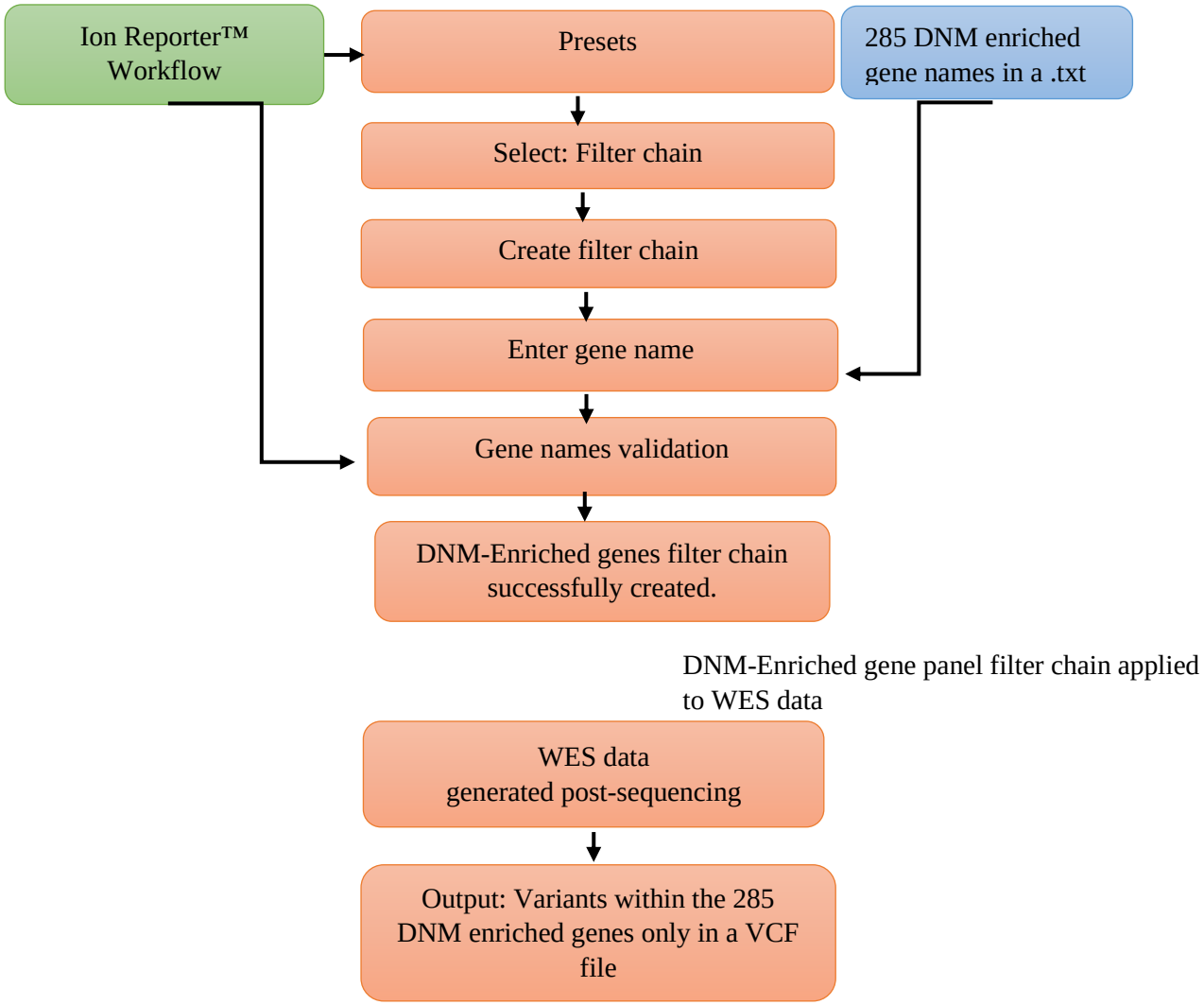


Figure 2.4. Workflow for designing a custom Ion Reporter™ filter chain to identify variants within the DNM-enriched genes from WES data.

2.6. Data Analysis for Both Cohorts A and B

The identical data analysis workflow was systematically applied to scrutinize data from both cohorts, ensuring uniformity and comparability throughout the analytical process. The workflow employed for analysing data from both cohorts is illustrated in Figure 2.5 below.

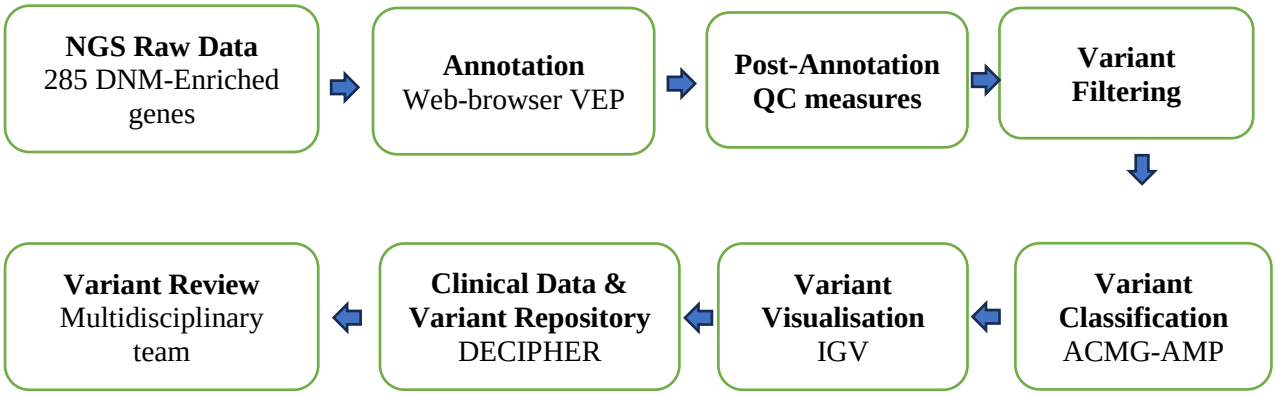


Figure 2.5. A schematic workflow employed for the analysis of data generated from cohorts A and B.

2.6.1. Variant Annotation of Data Generated from Cohorts A and B

Variant annotation involves adding supporting metadata and knowledge to quality-filtered raw putative variant calls to enhance the assessment of variants likely to impact function. Annotation aids in reducing the number of variants needing further prioritisation. Information added may include allele population frequencies, known associations with human diseases and phenotypes as well as predicted deleterious effects. Furthermore, information on the variant including the nucleotide changes affecting the gene, location of the variants (for example, upstream of a transcript in coding sequence or in non-coding RNA or in regulatory regions), information on chromosome location, reference allele, alternative allele and the consequence of the variants on the protein sequence, for example, stop-gain, missense, stop-loss or frameshift is also generated when the VCF file is annotated (Tollefson *et al.*, 2019). The VCF files obtained from the two cohorts were annotated using the web-browser version of the VEP (Ensembl release 104 – 108. Release date: May 2021- Oct 2022) (McLaren *et al.*, 2016).

2.6.2. Pre-Annotation Data QC

Low quality variants were removed. Low quality variants were defined as variants with:

- Read depth: variants with a read depth of ≤15 (number of reads per variant base≤15)

- Unbalanced allelic ratios: variants whose allelic ratios were not within the expected range as given below:
    - Heterozygous variants -approximately 50/50 (30%-65%)
    - Homozygous variants - approximately 100% (90%-100%)
    - Hemizygous variants- approximately 100% (90%-100%)

2.6.3. Variant Filtering of Data Generated from Cohorts A and B

Variant filtering aims to narrow down a list of potential putative disease-causing variants from an extensive pool of benign variants. For both datasets, a tiered variant filtering approach was employed to identify such variants. Initially, a VEP variant filtering strategy was implemented in conjunction with variant annotation using the web-browser version of the VEP. In this strategy, we excluded common variants with a minor allele frequency (MAF) of ≥1% in the 1000 Genomes combined population database and gnomAD databases integrated on the web-browser version of the VEP. The output resulted in a VCF file comprising 100-150 variants per sample. Subsequently, the VCF file was converted into Microsoft Excel format for subsequent manual filtering. VEP predicts the functional effects of genomic variants and classifies them into four categories: high, moderate, low, and modifier. This classification represents the likelihood of the variants having a disruptive or non-disruptive effect on the protein. Variants with a VEP predicted low and/or modifier functional impact and synonymous variants were manually filtered out using Microsoft Excel. This step led to a reduction in the number of variants from 100-150 variants per sample to 30-60 variants per sample.

A total of 30-60 variants still poses a challenge for manual investigations; thus, an additional filtering strategy was implemented. In this strategy, the variants were compared against the allele population frequency in the Exomes- and Genomes All datasets of the Genome Aggregation Database (gnomAD) (Karczewski *et al.*, 2020). The gnomAD database comprises data from over 100,000 individuals, making it larger than the 1000 Genomes combined population database utilised during the annotation and the initial filtering strategy on the web-browser version of the VEP. Common variants with an MAF exceeding 1% in the gnomAD Exomes- and Genomes databases were filtered out. Consequently, the number of variants available for prioritization, subsequent manual investigations, and classification was reduced to 5-10 variants per sample.

In our third step we employed multiple *in silico* prediction tools to evaluate the predicted impact of the remaining variants at the protein level and assess the evolutionary conservation of their respective locations. These tools utilise amino acid properties, functional domains, and structural attributes of the sequence variant to determine whether it has a deleterious effect or is tolerated in terms of protein function. Multiple *in silico* predictive tools that were incorporated into Ensembl VEP were used in combination for prediction of variant pathogenicity. *In silico* tools predict the pathogenicity of genetic variants by considering factors such as the type of variant, its impact at the nucleotide level, and its effect on protein structure and function (Richards *et al.,* 2015). The prediction of variant pathogenicity also considers whether the variant affects a functional domain and/or a region that is conserved throughout evolution. Variants that were predicted to have a damaging effect on the protein and/or were in evolutionarily conserved regions were given priority, while variants predicted to be non-deleterious by computational pathogenicity prediction tools were removed. This last filtering step yielded 1-3 candidate variants per sample. An overview of the variant filtering strategy employed is presented in Figure 2.6 and a list of *in silico* prediction tools used in this study are listed in Table 2.4.

Table 2.4. List of *in silico* prediction tools used in this study

Tool	**Protein function/conservation**	**Reference**
CADD	Predicted protein function	(Rentzsch *et al.,* 2019)
SIFT	Predicted protein function	(Kumar *et al.,* 2009)
REVEL	Predicted protein function	(Ioannidis *et al.,* 2016)
DANN	Predicted protein function	(Quang *et al.,* 2015)
PolyPhen2	Predicted protein function	(Adzhubei *et al.,* 2013)
Mutation Taster	Predicted protein function	(Schwarz *et al.,* 2010)
Mutation Assessor	Predicted protein function	(Reva *et al.,* 2011)
GERP	Conservation	(Cooper *et al.,* 2005)
FATHMM	Predicted protein function	(Shihab *et al.,* 2013)
PROVEAN	Predicted protein function	(Choi and Chan, 2014)

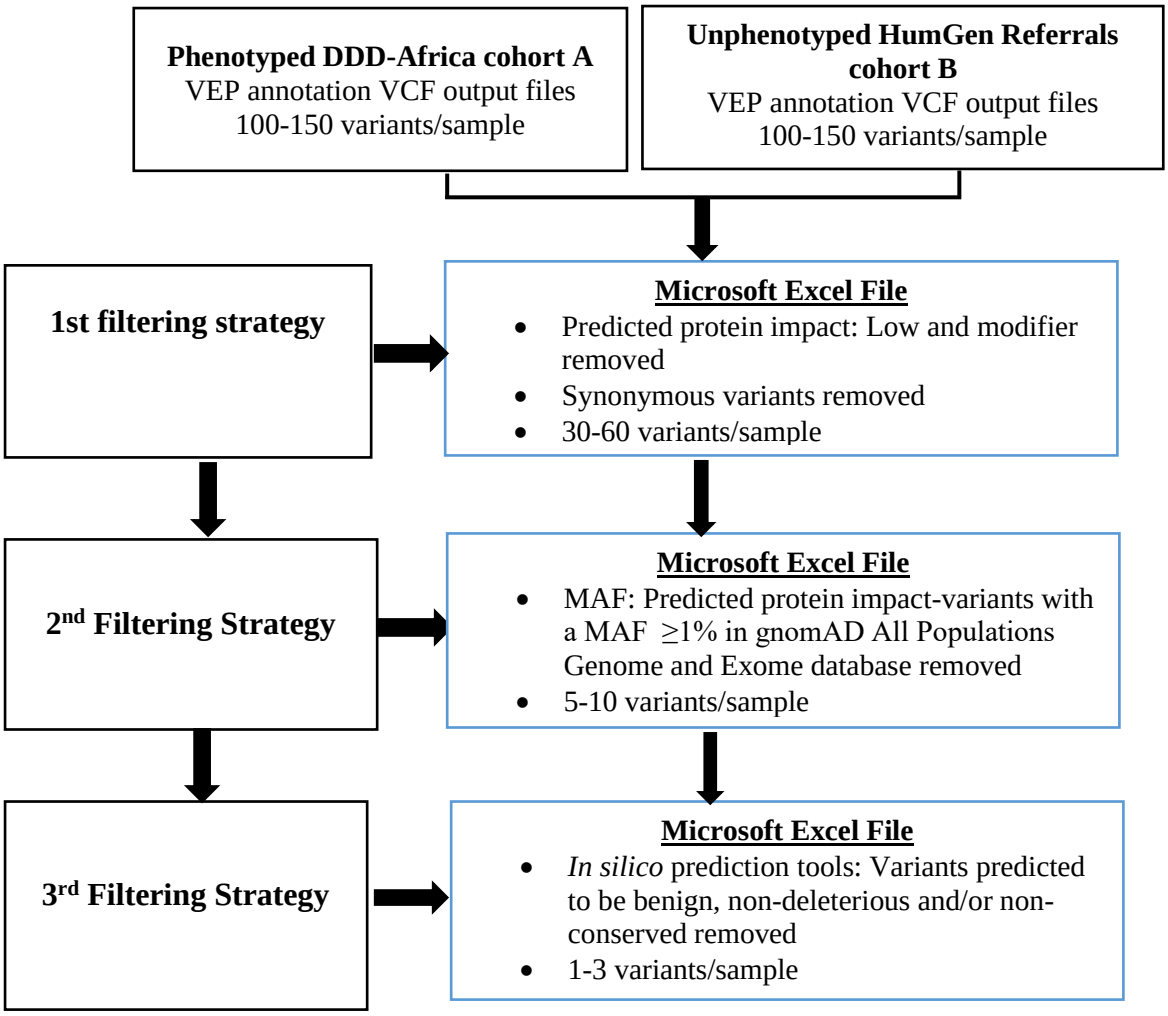


Figure 2.6. A variant filtering pipeline employed for the filtering of variants identified from both cohorts A and B.

2.6.4. Interpretation and Classification of Variants Following the ACMG-AMP Guidelines

The pathogenicity of each prioritised variant in both datasets was assessed by considering known information about the variant from the literature and the VEP annotation output. Additionally, variants were cross-referenced with ClinVar (Landrum *et al.*, 2018), an accessible repository of human genetic variants and their disease-related interpretations. Drawing upon the knowledge obtained from these sources, variants were manually classified according to the ACMG-AMP guidelines for sequence variant interpretation (Richards *et al.*, 2015). This classification process resulted in five groups: (i) benign, (ii) likely benign, (iii) VUS, (iv) likely pathogenic, and (v) pathogenic. Subsequently, the Varsome human genome variant search engine (Kopanos *et al.*, 2019), was employed to compare the manually assigned ACMG-AMP codes with the online automated ACMG codes application. Variants classified

as benign or likely benign were excluded as potential causative factors and were removed from further consideration. In most cases, variants classified as VUS were also eliminated, unless there was compelling evidence from the implicated gene indicating its potential as the causative gene. In such instances, a more thorough investigation was conducted to determine if the variant could be upgraded to the category of likely pathogenic or pathogenic. The filtering process excluded variants classified as benign, likely benign, and VUS, effectively reducing the number of variants in some patients to at most two high-priority variants—classified as 'hot' VUS, likely pathogenic, or pathogenic. However, in some patients, all variants were filtered out, leaving no high-priority variants for further assessment.

2.6.5. Variant Visualisation

The BAM files of all prioritised variants (classified as pathogenic or likely pathogenic after the application of the ACMG guideline), were uploaded onto the IGV v2.12.2 (Thorvaldsdóttir *et al.,* 2013), which enabled the visualisation of variant's read depth to determine whether base calling was consistent, or it was poorly covered. This is an important post alignment quality control step given that poorly covered regions can yield sequencing artefacts during variant calling.

2.6.6. Clinical Data and Shortlisted Variant Repository

All shortlisted variants that passed the post-classification quality control check were uploaded to DECIPHER (https://www.deciphergenomics.org) (Foreman *et al.,* 2022), a genetic and clinical data repository which was utilised for variant interpretation and the association of variants with the patient phenotype during the in-depth variant review process.

2.6.7. Variant Review and Clinical Phenotypic Correlation for Cohorts A and B

Prioritised candidate variants identified from both cohorts underwent a manual review and were correlated with the clinical phenotypes of the patients. Prioritised candidate variants from the DDD-Africa cohort A underwent an in-depth broad manual review and were correlated with the clinical phenotype of the patients by a multidisciplinary team in collaboration with DDD-Africa. This team comprised genetic counsellors, medical scientists, researchers, and medical geneticists. During these review meetings, the team evaluated and compared the variants with the patient's clinical presentations. They also examined the detailed clinical information and family history of the patients, comparing them to published clinical features associated with each gene containing the prioritised variants. The objective was to assess the

relevance of these variants in the context of each specific patient. When a significant correlation between the phenotype and variant was observed, a molecular diagnosis was established. However, if there was an insufficient correlation between the phenotype and variant or inadequate phenotypic information, the patient was considered negative within the project's pipeline.

Conversely, prioritised candidate variants from the HumGen referral cohort B underwent a less extensive review process compared to cohort A. Despite the involvement of a multidisciplinary team, the review in cohort B was constrained by the limited nature of available clinical presentations, with an average of two clinical features noted and a lack of family history information. Establishing a robust correlation between the identified variants and phenotypes in this cohort proved challenging due to the dearth or limited nature of phenotypic data.

2.7. Diagnostic Yield Calculation for Cohorts A and B

The diagnostic yield of both cohorts was determined by dividing the number of positive results (defined as variants classified as likely pathogenic and/or pathogenic), by the total number of patients tested. A positive diagnosis was confirmed by a multidisciplinary team and indicated a phenotype that was consistent with that of the patient. The calculation for the diagnostic yield analysis was as follows:

$$\text{Diagnostic Yield} = \frac{\text{Number of positive results}}{\text{Total number of patients tested}} \text{ X } 100\%$$

2.8. Exploring Factors Influencing the Yield of DNM-Enriched Genes in DDs.

2.8.1. Availability of Comprehensive Phenotypic Data

We explored the influence of thorough and precise clinical phenotyping on the diagnostic yield of the DNM-enriched gene panel strategy. Our hypothesis suggests that employing a more comprehensive and accurate phenotyping approach would result in an increased diagnostic yield. To assess this hypothesis, we employed an odds ratio (OR)analysis. The OR will be calculated using the formula OR=(a/b)/(c/d), where a and b will be the number of events and non-events in the group with phenotypic data available, and c and d will represent the number of events and non-events in the group without phenotypic data available (adapted from

Szumilas (2010). The confidence interval for the odds ratio will be calculated using the epitools package on R studio

2.8.2. Trios versus Singleton Analysis

The likelihood of attaining a molecular diagnosis across various family types was assessed through an odds ratio analysis, where the odds ratio of achieving a molecular diagnosis in trios was compared to that in duos and singleton cases.

2.9. Comparison of Diagnostic yields of Genetic Testing Strategies of DDs

A descriptive comparison was conducted to compare the diagnostic rates obtained from historical and/or recommended diagnostic tests for DDs with those achieved through a DNM-enriched gene panel strategy employed in this study.

2.10. Assessing the Impact of a DNM-Enriched Gene Panel Approach in Reducing Diagnostic Odysseys

The assessment of the impact of the DNM-Enriched Gene Panel Approach involved a detailed examination of diagnostic outcomes among patients from the DDD-Africa cohort mainly because this cohort was already receiving genetic services in our participating genetic clinics. The analysis focused on key parameters, including the age at the first genetic clinic visit, the age at which a molecular diagnosis was established, and the overall duration of the diagnostic odyssey for each participant.

Chapter 3

Results

3. RESULTS

This section presents the results obtained from the methodologies implemented in Chapter 2. The design and curation of a DNM enriched gene panel is presented, followed by patient-specific results obtained from screening of disease-causing variants in 285 DNM-enriched genes across two cohorts. Furthermore, the cost-effectiveness and feasibility of implementing a DNM enriched gene panel strategy in a diagnostic setting are explored.

3.1. DNM-Enriched Gene Panel Design

3.1.1. Intolerance to Variation Metrics Results

The tolerance of each gene to genetic variation was evaluated using the gnomAD pLI and missense Z-score (https://gnomad.broadinstitute.org/ - accessed July 2020). Of the 285 genes, a total of 78 genes were predicted to be intolerant to LoF variations based on the pLI score of ≥ 0.9. Similarly, using the missense Z-score, 25 genes were predicted to be intolerant to missense variations. Additionally, 126 genes were found to be intolerant to both missense and LoF variations. Conversely, 56 genes were predicted to be tolerant to both missense and loss-of-function variations.

3.1.2. Gene-Disease Validity Curation Results

Among the 285 identified genes, 156 were curated in the ClinGen gene-disease validity list. Within this curated list, 150 genes had a definitive association with NDDs, two genes showed a moderate association, and a single gene had limited association. Additionally, two genes had conflicting association with NDDs (https://clinicalgenome.org/ - accessed: 15th October 2023). Evaluating the gene-disease validity curation on PanelApp, 236 genes were curated across various gene panels, with 228 classified as diagnostic-grade, two as borderline, and six as minimal evidence genes(https://panelapp.genomicsengland.co.uk/ - accessed: 15th October 2023). A list of the 285 genes and their ClinGen and PanelApp curation statuses together with their intolerance status is shown in Appendix C.

3.1.3. Final DNM-Enriched Custom Gene Panel Design

The final custom DNM-enriched gene panel consisted of all 285 DNM-enriched genes, regardless of their association with disease (as curated on ClinGen and PanelApp gene-disease curated gene lists), and regardless of their predicted variation tolerance. The decision to include genes lacking established direct associations or predicted tolerance in the gene panel, especially

when focusing on African data, is grounded in several considerations. Firstly, African populations exhibit extensive genetic diversity (Pfennig *et al*., 2023), highlighting the importance of incorporating genes that facilitate the exploration of novel associations and population-specific variants. Secondly, acknowledging that our understanding of gene-disease associations and genetic variation is continually evolving, genes with limited or inconclusive evidence cannot be dismissed, as they may prove to be significant in the future (McGlaughon *et al.*, 2018). However, it is crucial to note that while the gene panel integrates well-known disease-causing genes and a sub-group of a few candidate genes, the ACMG recommends that caution should be exercised in interpreting variants within genes with limited and/or conflicting gene-disease validity associations, particularly if the results are intended for use in a diagnostic context (Bean *et al.*, 2020). Thus, in adherence to these recommendations, we diligently reviewed the shortlisted variants, giving due consideration to the limited and/or conflicting gene-disease validity associations when performing our final classification.

3.2. Phenotyped DDD-Africa Cohort A Results

3.2.1. DDD-Africa cohort A demographics

As previously mentioned in section 2.3, this cohort underwent WES at the Sanger Institute as part of the DDD-Africa project. A total of 96 WES data samples were selected from the first 100 cases of the DDD-Africa study and analysed in the present study. In addition to patient data, parental data was available in most cases, resulting in the cohort consisting of 49 parents-child trios, 38 parent-child duos and nine singletons (proband only) (Figure 3.1).

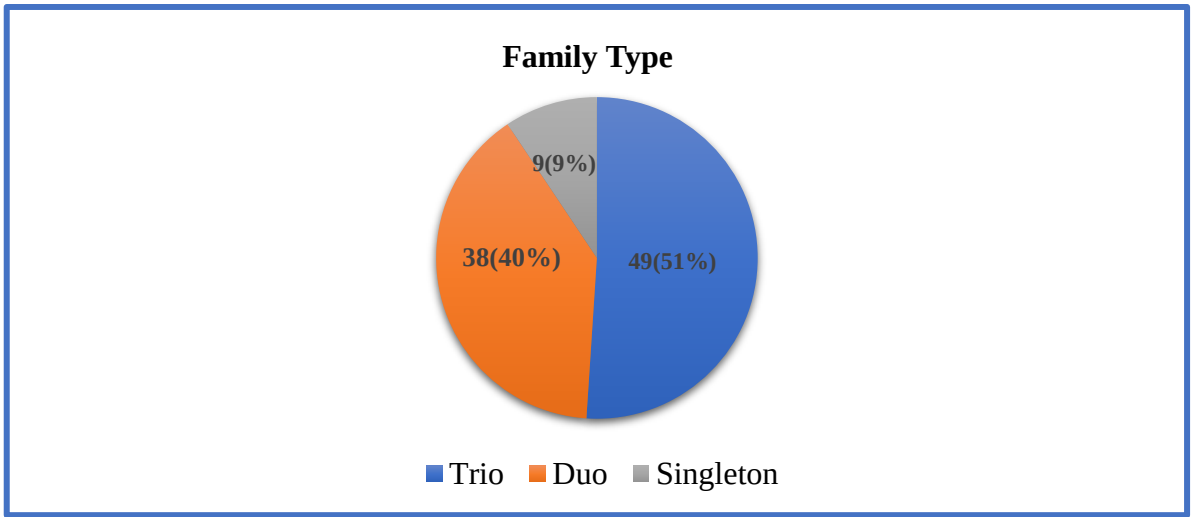


Figure 3.1. DDD-Africa cohort A family types. The cohort consisted of proband-parent trios, proband-parent duos, and proband-only cases. The majority of the cases (51%) were proband-parent trios.

3.2.2. DDD-Africa cohort A NGS Results

The average WES depth coverage for this cohort was 40x. The analysis of 285 DNM enriched genes in 96 patients included in this cohort resulted in up to three variants identified in each patient. Variants were filtered according to their predicted protein impact, MAF and *in silico* prediction tools. Subsequently, a comprehensive analysis led to the identification of 18 distinct candidate disease-causing variants among 18 unrelated patients, which were shortlisted for further consideration. Shortlisted candidate variants were subsequently uploaded onto DECIPHER and presented during a multidisciplinary variant review meeting within the DDD-Africa team. The details of the shortlisted candidate variants and the outcomes of the review meeting are summarised in Table 3.1.

3.2.3. DDD-Africa cohort A Diagnostic Yield

Fourteen shortlisted candidate variants in 14 patients exhibited variants consistent with their clinical phenotype. These variants were subsequently interpreted and classified according to the ACMG-AMP guidelines, establishing a molecular diagnosis in this patient group (Table 3.2). The diagnostic yield was calculated to be 14.6%.

Table 3.1. Summary of 18 shortlisted candidate variants identified in the DDD-Africa cohort A

	Patient code	HGVS	Coding impact	Zygosity	Inheritance	Variant Review decision
1	D3S_0028	NM_014225.6(*PPP2R1A*):c.773G>A: p.Arg258His	Missense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
2	D3S_0034	NM_017780.4(*CHD7*):c.8528del: p.Gly2843GlufsTer46	Frameshift	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
3	D3S_0038	(NM_032436.4(*CHAMP1*):c.775_782del:p.Pro259SerfsTer16	Frameshift	Heterozygous	Unknown	Exclude: False positive-not detected by Sanger
4	D3S_0047	NM_001374820.1(*ARID1B):*c.2819del: p.Pro927GlnfsTer57	Frameshift	Heterozygous	*De novo*	Include: Gene is consistent with phenotype
5	D3S_0066	NM_024757.5(*EHMT*1):c.2073dup:p.Val692ArgfsTer6	Frameshift	Heterozygous	Maternally inherited	Exclude: False positive-not detected by Sanger
6	D3S_0050	NM_017780.4(*CHD7*):c.232C>T:p.Gln78Ter	Nonsense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
7	D3S_0051	NM_017780.4(*CHD7*): c.5833C>T:p.Arg1945Ter	Nonsense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
8	D3S_0052	NM_001110792.2(*MECP2*):c.961G>A:p.Arg321Trp	Missense	Hemizygous	Maternally inherited	Include: Gene is consistent with the phenotype
9	D3S_0053	NM_015570.4(*AUTS2*):c.376C>T: p.Arg126Ter	Missense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
10	D3S_0064	NM_005359.6(*SMAD4*):c.1498A>G:p.Ile500Val	Missense	Heterozygous	Maternally inherited	Exclude: Gene fits phenotype, in depth discussion pending
11	D3S_0082	NM_138927.4(*SON*):c.5753_5756del:p.Val1918GlufsTer87	Frameshift	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
12	D3S_0086	NM_003165.6(STXBP1): c.416C>T:p.Pro139Leu	Missense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
13	D3S_0089	NM_001193347.1(*MEF2C*):c.963T>A:p.Tyr275Phe	Nonsense	Heterozygous	Unknown	Include: Gene is consistent with the phenotype
14	D3S_0094	NM_001356.5(*DDX3X*): c.931C>T:p.Arg311Ter	Nonsense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
15	D3S_0111	NM_001304717.5*(PTEN*):c.1129C>A:p.Thr534Ala	Missense	Heterozygous	Unknown	Include: Gene is consistent with the phenotype
16	D3S_0114	NM_001205293.3(*CACNA1A*):c.4190T>G:p.Met1397Arg	Missense	Heterozygous	Unknown	Exclude: Gene is not consistent with the phenotype
17	D3S_0117	NM_022455.5(*NSD1*): c.4913A>G:p.His1638Arg	Missense	Heterozygous	*De novo*	Include: Gene is consistent with the phenotype
18	D3S_0124	NM_018026.4(*PACS1*): c.607C>T:p.Arg203T	Missense	Heterozygous	Unknown	Include : Gene is consistent with the phenotype

Table 3.2. Summary of 14 clinically relevant variants leading to molecular diagnoses in the DDD-Africa cohort A

	Patient	Variant	Type of variant	Known/Novel	ACMG-AMP codes and Strength applied	Classification	Disease association
1	D3S_0028	NM_014225.6(PPP2R1A):c.773G>A	Missense	Known (VCV000217458.4)	PP5_strong, PP3_supporting, PM2_supporting, PS3_supporting, PM6_moderate & PM1_moderate	Pathogenic	PPP2R1A-related neurodevelopmental disorders
2	D3S_0034	NM_017780.4(CHD7):c.8528delG	Frameshift	Novel	PVS1_very strong, PM2_supporting, PM6_moderate	Pathogenic	CHD7-related disorder
3	D3S_0047	NM_001374820.1(*ARID1B):*c.2819delC	Frameshift	Novel	PVS1_very strong, PM2_supporting, PM6_moderate	Pathogenic	Coffin-Siris syndrome
4	D3S_0050	NM_017780.4(*CHD7*):c.232C>T	Nonsense	Novel	PVS1_very strong, PM2_supporting, PM6_moderate, PP3_supporting	Pathogenic	CHD7-related disorder
5	D3S_0051	NM_017780.4(*CHD7*): c.5833C>T	Nonsense	Known (VCV000217458.4)	PVS1_very strong, PM2_supporting, PM6_moderate, PP5_very strong	Pathogenic	CHD7-related disorder
6	D3S_0052	NM_001110792.2(*MECP2*):c.961C>T	Missense	Known (VCV000143749.18)	PP5_very strong, PM2_supporting, PM1_strong, PM5_moderate	Pathogenic	Rett syndrome
7	D3S_0053	NM_015570.4(*AUTS2*):c.376C>T	Nonsense	Known (VCV000620261)	PVS1_very strong, PM2_supporting, PM6_moderate, PP5_very strong	Pathogenic	AUTS2-related disorder
8	D3S_0082	NM_138927.4(*SON*):c.5753_5756delAGTT	Frameshift	Known (VCV000252929.48)	PVS1_very strong, PM2_supporting, PP5_very strong	Pathogenic	Zhu–Tokita–Takenouchi–Kim syndrome
9	D3S_0086	NM_003165.6(STXBP1): c.416C>T	Missense	Known (RCV000416131.3)	PP5_very strong, PP3_supporting, PM2_supporting, PM5_moderate, PM6_moderate	Pathogenic	STXBP1-related disorder
10	D3S_0089	NM_001193347.1(*MEF2C*):c.963T>A	Nonsense	Novel	PVS1_very strong, PM2_supporting	Likely pathogenic	MEF2C haploinsufficiency syndrome
11	D3S_0094	NM_001356.5(*DDX3X*): c.931C>T	Nonsense	Known (VCV000981252.3)	PVS1_very strong, PM2_supporting, PP5_very strong, PM6_moderate	Pathogenic	DDX3X-related disorder
12	D3S_0111	NM_001304717.5(*PTEN*): c.1129C>A	Missense	Novel	PM2_supporting, PM5_moderate, PM1_moderate, PP3_moderate	Likely pathogenic	PTEN hamartoma tumour syndrome
13	D3S_0117	NM_022455.5(*NSD1*): c.4913A>G	Missense	Novel	PM1_moderate, PM2_supporting, PP5_supporting	Pathogenic	Sotos syndrome
14	D3S_0124	NM_018026.4(*PACS1*): c.607C>T	Missense	Known (VCV000039581.20)	PM2_supporting, PS3_very strong, PP2_supporting, PP5_ very strong	Pathogenic	PACS1-related disorder

3.2.4. DDD-Africa cohort A Individual Patient Results

i. Patient D3S_0028

Before enrolling in the DDD-Africa study, the 4-year-old male patient had undergone karyotyping and MLPA screenings to detect microdeletion/duplication syndromes and subtelomeric deletions/duplications, with no abnormalities detected. A missense variant in the *PPP2R1A* gene was prioritised in this patient: *PPP2R1A*:c.773G>A. Pathogenic variants in the *PPP2R1A* are associated with PPP2R1A-related neurodevelopmental disorders (Qian *et al.,* 2023). The *PPP2R1A:*c.773G>A variant identified in this patient is not documented in public allele frequency data repositories such as gnomAD (PM2). This variant is located within in the HEAT domain of the PPP2R1A protein (PM1). Previous reports have identified this variant in individuals with varying NDDs (Houge *et al.,* 2015; Lenaerts *et al.,* 2021; Lei *et al.,* 2023). Several lines of computational evidence strongly suggest a deleterious impact gene product (PP3). Additionally, both parents were tested and found to be homozygous for the wild type allele – suggesting *de novo* occurrence (PM6). Experimental studies conducted by Houge and colleagues demonstrate that this missense alteration adversely affects PPP2R1A function (PS3) (Houge *et al.,* 2015). Furthermore, the variant is classified as likely pathogenic/pathogenic in ClinVar, with a 2-star rating (VCV000217458.4) (PP5). Based on the ACMG-AMP codes PP5_strong, PP3_supporting, PM2_supporting, PS3_supporting, PM6_moderate and PM1_moderate, the *PPP2R1A*:c.773G>A variant is classified as pathogenic.

ii. Patient D3S_0034

The 5-year-old male patient had undergone multiple genetic tests before participating in the DDD-Africa study. These tests included MLPA for microdeletion/duplication syndromes and subtelomeric deletions/duplications, karyotyping, common aneuploidy screening, and FISH (Di-George), all of which detected no abnormalities. A *CHD7*:c.8528delG variant was prioritised in this patient. Disease-causing variants in the *CHD7* gene are associated with CHD7-related disorder. The *CHD7:*c.8528delG variant identified in this patient is a frameshift variant within a gene where LoF is an established mechanism of disease (PVS1) (Zentner *et al.,* 2010). This variant is absent from gnomAD databases (PM2). Testing of both parents revealed homozygosity for the wild-type allele, pointing to a *de novo* origin (PM6). Based on the ACMG-AMP codes PVS1_very strong, PM2_supporting and PM6_moderate, the variant is classified as pathogenic.

iii. Patient D3S_0047

Before enrolling in the DDD-Africa study, the patient, a 9-year-old female at the time of recruitment, had undergone MLPA screenings for microdeletion/duplication syndromes and subtelomeric deletions/duplications and karyotyping, and no abnormalities were detected. A frameshift variant within the *ARID1B* gene was prioritised in this patient: *ARID1B*:c.2819delC. Pathogenic variants in the *ARID1B* gene are the predominant cause of Coffin-Siris syndrome. (Tsurusaki *et al.*, 2014). The *ARID1B*:c.2819delC variant identified in this patient is a frameshift variant in a gene where LoF is an established mechanism of disease (Sluijs *et al.*, 2018) (PVS1). The variant is not documented in gnomAD databases (PM2). Homozygosity for the wild-type allele was observed in both parents upon testing, implying a *de novo* origin (PM6). In accordance with the ACMG-AMP codes PVS1_very strong, PM2_supporting and PM6_moderate, this variant is classified as pathogenic.

iv. Patient D3S_0050

Before enrolling into the DDD-Africa study, the patient, a 4-year-old female at the time of recruitment, had undergone prior genetic testing, including MLPA for microdeletion/duplication syndromes and karyotyping and no abnormalities were detected. The heterozygous *CHD7*:c.232C>T variant identified in this patient is not observed in gnomAD databases (PM2). Furthermore, this variant is a nonsense variant introducing a premature stop codon at amino acid 78 of the CHD7 protein. The resulting protein will be shortened and lack all functional domains (PVS1). The *CHD7* variant identified in this patient is absent from gnomAD databases (PM2) and assumed *de novo* with no parental confirmation (PM6). Based on the ACMG-AMP codes PVS1_very strong, PM2_supporting, and PM6_moderate, the variant is classified as pathogenic.

v. Patient D3S_0051

Prior to his participation in the DDD-Africa study, the 2-year-old male patient had undergone testing for microdeletion/duplication syndromes, karyotyping, and FISH 22q, and no abnormalities were detected. The *CHD7*: c.5833C>T variant identified in this patient is not observed in gnomAD databases (PM2). Furthermore, Furthermore, this variant is a nonsense variant introducing a premature stop codon at amino acid 1945 of the CHD7 protein in a gene where LoF is an established mechanism of disease (PVS1). Testing confirmed that both parents harbour a homozygous wild-type allele at this locus, suggesting a *de novo* origin (PM6). This variant is extensively documented in literature and is associated with CHARGE syndrome

(Bartels *et al.,* 2010; Janssen *et al.,* 2012; Jongmans *et al.,* 2006; Lalani *et al.,* 2006) Furthermore, the variant is a documented and classified as pathogenic in ClinVar, with a 2-star rating (VCV000217458.4) (PP5). Based on the ACMG-AMP codes PVS1_very strong, PM2_supporting, PM6_moderate and PP5_very strong, the variant is classified as pathogenic.

vi. Patient D3S_0052

At the time of recruitment, the patient was a 5-year-old male and had undergone multiple genetic investigations including MLPA for microdeletion/duplication syndromes; FXS, karyotyping, Glutaryl-Coa Dehydrogenas c.877G>A test, and OPTIMA + Cytoscan high density array. Notably, Notably, all previous tests yielded no abnormalities. A hemizygous variant, *MECP2*:c.961C>T, was prioritised in this patient. Causal variants in *MECP2* are a major cause of Rett syndrome (RTT), a severe and well-characterized neurological disorder primarily affecting females (Hagberg *et al* . 1983)

In males, causal variants in *MECP2* can lead to a broad spectrum of clinical presentations, including neonatal encephalopathy (OMIM #300673), autism susceptibility (OMIM #300496), X-linked mental retardation syndrome (OMIM #300055), MECP2 duplication syndrome (MDS) (OMIM #300260), and late-onset parkinsonism and spasticity (Neul *et al.,* 2020). Consequently, diagnosing and classifying males can be more challenging than classical RTT females. Recognising these complexities, a 2021 systematic review by Inuzuka and colleagues explored the diverse clinical presentations in hemizygous males carrying a single copy of the *MECP2* causal variant. Their findings underscored that *MECP2* causal variants associated with X-linked mental retardation syndrome and X-linked psychosis syndrome often manifest as de novo variants or are inherited from a normal or mildly affected mother. They also emphasised that the typical Rett syndrome phenotype is not anticipated in males, except in those with Klinefelter syndrome or somatic mosaicism for *MECP2* (Inuzuka *et al.,* 2021).

The *MECP2*:c.961C>T variant identified in our male patient is not found in the gnomAD database (PM2). This variant, previously reported in the literature among male individuals without typical RTT features (Hengel *et al.,* 2020; Neul *et al.,* 2020), is classified as likely pathogenic/pathogenic in ClinVar, earning a 3-star rating (VCV000143749.18) (PP5). Despite the absence of typical RTT features in this patient, the variant is inherited from an asymptomatic mother. Additionally, a different amino acid change, chrX:154030902:C>T

(p.Arg321Gln), is classified as likely pathogenic with a 1-star rating in ClinVar (RCV001235683) (PM5). Following the ClinGen Rett/Angelman-like Expert Panel recommendations (McKnight *et al.,* 2022), and considering the ACMG-AMP codes (PP5_very strong, PM2_supporting, PM1_strong, PM5_moderate), the variant is classified as pathogenic.

vii. Patient D3S_0053

At the time of recruitment, the 2-year-old male participant had already undergone karyotyping, which showed no abnormalities. In this patient, a nonsense variant in the *AUTS2* gene was prioritised: *AUTS2:*c.376C>T. Disease-causing variants in the A*UTS2* gene are associated with a wide range of neuropsychological disorders, such as ASD, ID, schizophrenia, and epilepsy (Amarillo *et al.*, 2014; Beunders *et al.,* 2015) The *AUTS2*:c.376C>T variant identified in this patient is a nonsense variant in a gene where LoF is a known mechanism of disease (Oksenberg *et al.,* 2013) (PVS1). The variant is predicted to introduce a premature stop codon, leading in the resulting protein product lacking the Auts2 functional domain of the protein. This variant is absent from gnomAD database (PM2) and is a known causative variant reported in previous studies in individuals with ASD (Beunders *et al*., 2015, Beunders *et al*., 2016; Turner *et al.,* 2019). The variant is classified as pathogenic in ClinVar, with a 2-star rating (VCV000620261) (PP5). Additionally, both parents underwent testing and were identified as homozygous for the wild-type allele (PM6). Based on the ACMG-AMP codes PVS1_very strong, PM2_supporting, PM6_moderate and PP5_very strong, the *AUTS2*:c.376C>T variant in this patient is classified as pathogenic.

viii. Patient D3S_0082

The patient, a 10-year-old male at the time of recruitment, had received negative results obtained a series genetic diagnostic tests including MLPA for microdeletion/duplication syndromes and subtelomeric deletions/duplications, karyotyping, high density cytoscan and array-CGH before enrolling into the DDD-Africa study. We prioritised a single frameshift variant in this patient: *SON*:c.5753_5756delAGTT. Disease-causing variants in the SON gene are associated with Zhu–Tokita–Takenouchi–Kim syndrome (Kim *et al.,* 2016; Takenouchi *et al.,* 2016; Tokita *et al.,* 2016; Zhu *et al.,* 2015). The SON:c.5753_5756delAGTT variant identified in this patient is a frameshift variant in a gene in which LoF is an established mechanism of disease (Kim *et al.,* 2016). The affected exon 3/12 truncate the protein at the RSRP functional domain, supporting the ACMG-AMP code PVS1. This variant is absent from

gnomAD database (PM2). Furthermore, it is extensively documented in the literature (Kim *et al.,* 2016; Tokita *et al.,* 2016), and is classified as likely pathogenic/pathogenic, with a 2-star rating (PP5) (VCV000252929.48). The inheritance status of this variant remains undetermined because of lack of parental data. Based on the ACMG-AMP codes PVS1_very strong, PM2_supporting and PP5_very strong, the variant is classified as pathogenic.

ix. Patient D3S_0086

Before participating in the DDD-Africa study, the 9-year-old male patient had obtained negative results from Prader-Willi MLPA, BBS10 founder mutation analysis, and karyotyping. A single missense variant in the *STXBP1* gene was prioritised in this patient. *STXBP1* disease-causing variants are linked to severe early epileptic encephalopathies and neurodevelopmental disorders (Carvill *et al.,* 2014; Otsuka *et al.,* 2010; Saitsu *et al.,* 2010). The *STXBP1*:c.416C>T variant is not found in gnomAD population databases (PM2). Multiple computational prediction tools predict a deleterious effect on the gene (PP3). Homozygosity for the wild-type allele was observed in both parents upon testing, implying a *de novo* origin (PM6). Furthermore, the variant has been previously documented in the literature by Barcia *et al* . (2014) and Keogh *et al* . (2015) and is classified as likely pathogenic/pathogenic with a 2-star rating ClinVar (RCV000416131.3) (PP5). Additionally, an alternative variant, chr9:127661192:C:A (Pro139Gln), is classified as pathogenic with a 1-star rating in ClinVar (RCV001201587) (PM5). Based on the ACMG-AMP codes PP5_very strong, PP3_supporting, PM2_supporting, PM5_moderate and PM6_moderate, the *STXBP1*:c.416C>T variant is classified as pathogenic.

x. Patient D3S_0089

At the time of recruitment, the patient, an 8-year-old female, had undergone Array-CGH, MECP2 molecular testing, and Angelman syndrome MLPA testing, all yielding negative results. In this patient, a heterozygous nonsense variant in the *MEF2C* gene, was prioritised: *MEF2C*:c.963T>A. Disease-causing variants in the *MEF2C* gene cause MEF2C haploinsufficiency syndrome, characterised by severe global developmental delay accompanied by absent speech, limited walking, seizures, and stereotypic movements (Le Meur *et al.,* 2010; Zweier and Rauch, 2011). The *MEF2C*:c.963T>A variant is a nonsense variant in a gene where haploinsufficiency is a known disease mechanism (Engels *et al.,* 2009). It is predicted to introduce a stop codon at amino acid 275 of the MEF2C protein resulting in a

shortened protein lacking the polar residues region of the protein (PVS1). The *MEF2C*:c.963T>A variant is not present in control databases, including gnomAD (PM2). Regrettably, due to the unavailability of samples from both parents for testing purposes, it was not feasible to determine the mode of inheritance. Based on the ACMG-AMP codes PVS1_very strong and PM2_supporting, the variant is classified as likely pathogenic.

xi. Patient D3S_0094

The patient, a 1-year-old female at the time of recruitment, had received negative results from MLPA for microdeletion/duplication syndromes and karyotyping prior to her participation in the DDD-Africa study. A heterozygous variant, *DDX3X*:c.931C>T, was prioritised in this patient. Sporadic *DDX3X* variants have been associated with intellectual disability and are considered one of the most common causes of intellectual disability in females (Blok *et al.,* 2015; Nicola *et al.,* 2019). Predominantly affecting females, the reported causal *DDX3X* variants are largely *de novo* loss-of-function, resulting in haploinsufficiency (Nicola *et al.,* 2019; Snijders Blok *et al.,* 2015; Tang *et al.,* 2021; X. Wang *et al.,* 2018). Remarkably, females harbouring missense variants exhibit a severe phenotype compared to those with truncating variants (Lennox *et al.,* 2020). The *DDX3X*:c.931C>T variant identified in this patient is a nonsense variant located within the DEAD domain of DDX3X protein and is predicted to introduce a stop codon at amino acid 311 of the protein. Resulting in the protein product without the helicase C functional domain of the DDX3X protein. Haploinsufficiency of the *DDX3X* is a known mechanism of disease in female individuals (PVS1) (Snijders Blok *et al.,* 2015). Additionally, the variant is not documented in control databases including gnomAD (PM2). This variant has been documented in the literature and is classified as pathogenic with a 2-star rating in ClinVar (VCV000981252.3) (PP5). Homozygosity for the wild-type allele was observed in both parents upon testing, implying a *de novo* origin (PM6). Based on the ACMG-AMP codes PVS1_very strong, PM2_supporting, PM6_moderate and PP5_very strong, the variant is classified as pathogenic.

xii. Patient D3S_0111

Prior to her enrolment to the DDD-Africa, the 6-year-old female patient had previously undergone MLPA testing for microdeletion/duplication syndromes and karyotyping, both of which yielded negative results. A heterozygous nonsense variant within *PTEN* gene was prioritised in this patient: *PTEN*:c.1129C>A. Causal variants in the *PTEN* gene are associated

with PTEN hamartoma tumour syndrome (PHTS) (Pilarski *et al.,* 2013). The *PTEN*:c.1129C>A variant identified in this patient is a missense variant occurring within a gene characterised by a low incidence of benign missense mutations, where missense mutation serves as a prevalent mechanism underlying the associated disease (PP2). This variant is absent in gnomAD population databases (PM2) and is located within the PTEN_C2 domain of the PTEN protein (PM1). Additionally, an alternative variant (*PTEN*:c.1129: C>G) is classified likely pathogenic in ClinVar (PM5), with a 3-star rating. Computational prediction tools predict a deleterious effect on the gene product (PP3). Due to the absence of parental information, the inheritance of this variant cannot be confirmed. Based on the ClinGen PTEN Expert Panel guidelines (Mester *et al.*, 2018), and the ACMG-AMP codes PM2_supporting, PM5_moderate, PM1_moderate and PP3_moderate, the variant is classified as pathogenic.

xiii. Patient D3S_0117

This patient had received negative results from MLPA screening for microdeletion/duplication and FXS testing. A heterozygous missense variant within the *NSD1* gene was prioritised in this patient. Causal variants in the *NSD1* gene are associated with Sotos syndrome (Douglas *et al.*, 2003). The *NSD1*:c.4913A>G variant, prioritised in this patient, is located within missense variant hot spot - exon 13 (PM1) (Testa *et al.*, 2023). This variant is not documented in gnomAD database (PM2). Computational prediction tools consistently predict a deleterious impact on the gene impact (PP3). This variant is classified in ClinVar as a VUS, with a 1-star rating, indicating a preliminary review (RCV003227346). The 1-star rating suggests that the variant has been submitted by a single submitter with an assertion criterion, and caution is advised in its interpretation due to the preliminary nature of the review. The inheritance of the *NSD1* variant in this study could not be established due to the unavailability of parental data. Based on the ACMG-AMP codes PM1_moderate, PM2_supporting, PP3_supporting, the variant is classified as likely pathogenic.

xiv. Patient D3S_0124

Prior to the DDD-Africa Study, the patient had undergone multiple genetic assessments, including MLPA for microdeletion/duplication syndromes, QF-PCR for common aneuploidies, and karyotyping. All conducted tests revealed no abnormalities. A heterozygous missense variant in the *PACS1* gene was prioritised in this patient: *PACS1*:c.607C>T. Causal variants within the *PACS1* gene cause PACS1 neurodevelopmental disorder-NDD (Schuurs-

Hoeijmakers *et al.,* 2012). Only two missense variants in *PACS1* gene, which affect the same amino acid, have been reported as the cause of PACS1-NDD (c.607C>T (p.Arg203Trp) and c.608G>A (p.Arg203Gln) (Bruno *et al.,* 2023; Miyake *et al.,* 2018). The *PACS1*:c.607C>T variant is not present in gnomAD population databases (PM2). It is classified as pathogenic in ClinVar, (VCV000039581.20) (PP5). Moreover, the *PACS1* gene has a missense Z-score of 3.71, which is greater than 3.12, indicating missense variants are a common cause of disease in this gene (PP2). Notably, this variant (acquired *de novo*), is the most frequent recurring variant in nearly all cases of PACS1-related NDD cases reported in the literature (Colak *et al.,* 2020; Seto *et al.,* 2021; Van Nuland *et al.,* 2021). However, in this case, inheritance could not be established due to the unavailability of parental sample for testing. Functional studies revealed that this variant alters the normal function pf the PACS1 protein (PS3). Based on the ACMG-AMP codes PM2_supporting, PS3_very strong, PP2_supporting and PP5_ very strong, this variant is classified as pathogenic.

3.3. Unphenotyped HumGen Referral Cohort B Results

3.3.1. Cohort B demographics

This cohort consisted of 96 proband cases, primarily comprising male individuals (90/96). The age of the cohort ranged from one to 17 years, with a median age of 5 years. Furthermore, the interquartile range (IQR), representing the middle 50% of the age distribution, was calculated as 5 years (Q1 = 3, Q3 = 8).

3.3.2. WES data generation

A total of 96 samples underwent WES at the Division. The WES of the cohort was completed in 48 sequencing runs, with two samples loaded per chip and per WES run. The sequencing process was followed by the generation of a sequencing run report by the Torrent Suite software. This report included useful QC metrics, and sequencing run statistics, which are divided into pre-alignment and post-alignment QC metrics and run statistics.

3.3.3. NGS data quality assessment

The results from the 48 WES runs yielded a total number of bases generated per run ranging from 4.4Gb to 18.9Gb. In the 48 WES runs performed in this cohort, the ISP loading density exhibited variability, ranging from 50% to 95%. The median read length spanned from 147bp to 210bp. The percentage of polyclonal reads was observed to fluctuate between 15% and 41%

between runs. Table 3.3 below shows some of the pre-alignment QC metrics and run statistics for all 48 WES runs. For the runs defined as optimal, we considered a read length within the range of 180bp to 220bp. The expected output data for WES on the 540 chip is between 10 Gb and 15 Gb, with a low-quality read percentage of less than 20%, and an ISP loading of above 80%. Additionally, we aimed for a minimum of 30% usable reads and polyclonal of less than 35%. Figure 3.2 below shows pre-alignment QC metrics for two sequencing runs: Run 1, representing an example of runs defined as optimal runs, and Run 45, illustrating suboptimal performance. The suboptimal performance of Run 45 could be attributed to several factors, such as the use of substandard starting DNA material, excessive fragmentation during library preparation, challenges associated with fragment size selection, inadequate DNA input, potential PCR bias, and the formation of adapters. Despite this, the generated data remains suitable for analysis. However, some genes within the panel may exhibit low coverage, which subsequently could result in the inability to detect disease-causing variants in these regions.

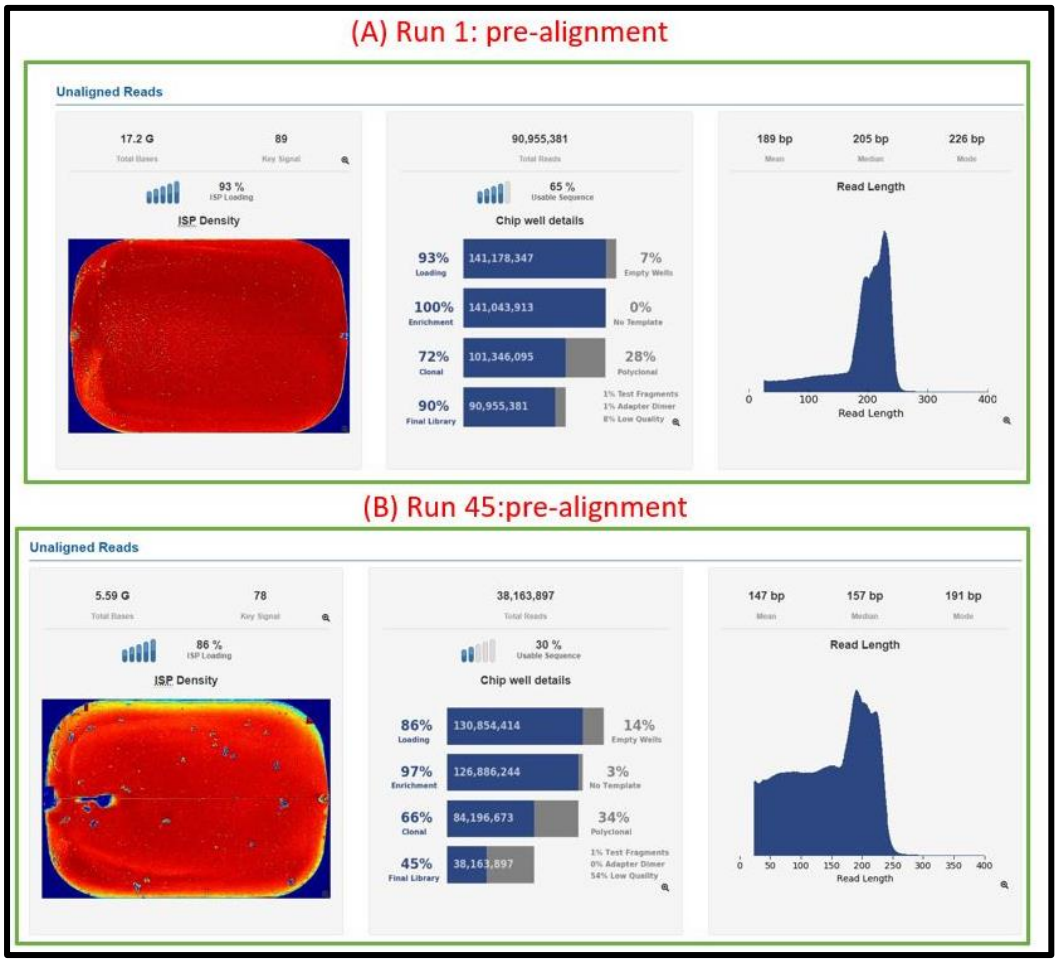


Figure 3.2. Examples of high quality and suboptimal runs. Image displaying the pre-alignment QC metrics and run statistics for Run1/chip1 and Run45/chip45. Run1/chip1 serves as an example of a run that generated high-quality pre-alignment sequencing data, while Run45/chip45 exemplifies a run that produced suboptimal, yet still analysable data

Table 3.3. Quality control metrics and run statistics for all 48 WES runs performed on the HumGen referral cohort B.

Chip#	Ion express barcodes used	Read length (bp)	Run total bases	% of low-quality reads	ISP loading	% usable reads	% polyclonal
1	1&2	205 bp	17.2Gb	8%	93%	65%	28%
2	3&4	197 bp	8.3Gb	20%	50%	52%	16%
3	5&6	201 bp	16.1Gb	10%	94%	63%	29%
4	7&8	204 bp	13.2Gb	22%	84%	42%	35%
5	9&10	207bp	10.8Gb	20%	86%	63%	31%
6	11&12	206 bp	12.6Gb	14%	94%	47%	41%
7	13&14	202 bp	11.8Gb	17%	93%	45%	33%
8	15&16	199 bp	11.6Gb	15%	94%	49%	32%
9	1&2	80 bp	4.4Gb	52%	91%	35%	26%
10	3&4	201 bp	15Gb	12%	93%	58%	33%
11	5&6	207 bp	16.1Gb	11%	92%	60%	32%
12	7&8	208 bp	16.8Gb	10%	90%	64%	29%
13	9&10	203 bp	13.7Gb	9%%	88%	55%	31%
14	11&12	206 bp	14.7Gb	8%	92%	57%	35%
15	13&14	207 bp	17.7Gb	8%	94%	65%	29%
16	15&16	199 bp	17.2Gb	12%	86%	66%	28%
17	1&2	203 bp	15.2Gb	13%	94%	47%	29%
18	3&4	210 bp	14.3Gb	12%	93%	42%	16%
19	5&6	198 bp	13Gb	9%	89%	53%	23%
20	7&8	199 bp	14.7Gb	12%	85%	58%	31%
21	9&10	172 bp	9.9Gb	15%	84%	44%	33%
22	11&12	198 bp	11.4Gb	23%	81%	41%	32%
23	13&14	201 bp	8.2Gb	19%	79%	39%	15%
24	15&16	165 bp	8.3Gb	22%	82%	42%	16%
25	1&2	201 bp	11.2Gb	21%	94%	52%	21%
26	3&4	189 bp	10.1Gb	23%	90%	45%	22%
27	5&6	180 bp	9.4Gb	17%	89%	43%	23%
28	7&8	200 bp	15.6Gb	11%	92%	47%	33%
30	9&10	199 bp	14.8Gb	23%	84%	34%	35%
31	11&12	207 bp	10.8Gb	20%	86%	63%	33%
32	13&14	208 bp	11.2Gb	11%	94%	67%	28%
33	15&16	197 bp	8.3Gb	20%	50%	52%	16%
34	1&2	147 bp	9.67Gb	23%	81%	51%	33%
35	3&4	152 bp	8.11Gb	30%	92%	45%	36%
36	5&6	207 bp	18.9Gb	7%	95%	69%	26%
37	7&8	206 bp	18.4Gb	10%	93%	67%	24%
38	9&10	206 bp	5.7Gb	49%	70%	29%	23%
39	11&12	206 bp	7.0Gb	43%	75%	33%	24%
40	13&14	212 bp	17Gb	16%	91%	63%	25%
41	15&16	212 bp	11.6Gb	31%	80%	43%	25%
42	1&2	148 bp	7.4Gb	34%	85%	42%	35%
43	3&4	208 bp	13.8Gb	23%	79%	60%	20%
44	5&6	206 bp	17.9Gb	9%	91%	67%	27%
45	7&8	157 bp	5.59Gb	54%	86%	30%	34%
46	9&10	207 bp	16.6Gb	11%	89%	64%	27%
47	11&12	188 bp	17.1Gb	19%	90%	58%	23%
48	13&14	207 bp	15.2Gb	20%	86%	61%	23%

Chip#: run number out of 48 WES performed in this study, live Ion Sphere Particle, % usable reads that are sequenced in a single sequencing run, Run total bases: total number of DNA bases that are sequenced during a single sequencing run, % of low-quality reads: proportion or percentage of sequencing reads in a data set that are deemed to be of low quality, ISP loading: percentage of wells in the chip that contain a live Ion Sphere Particle, % usable reads: percentage of sequencing reads that are considered usable or of sufficient quality for downstream analysis in a sequencing run, % polyclonal: percentage of polyclonal sequences

within a sample or dataset. Runs shaded in mustard were deemed optimal, but all data sets could be used for variant screening

3.3.4. HumGen Referral Cohort B NGS Results

The analysis of 285 DNM enriched genes in 96 proband-only cases included in this cohort resulted in a total of 12 unique shortlisted candidate variants identified in 12 unrelated patients. The shortlisted candidate variants were presented to a multidisciplinary team. The details of the shortlisted candidate variants and the outcomes of the review meeting are summarized in Table 3.4.

3.3.5. HumGen Referral Cohort B Diagnostic Yield

Five shortlisted candidate variants in six patients exhibited variants consistent with the limited clinical presentation. These variants were subsequently interpreted and classified according to the ACMG-AMP guidelines, establishing a molecular diagnosis in this patient group (Table 3.5). Of the six variants, five are classified as likely pathogenic/pathogenic and a single variant is classified as a VUS. The diagnostic yield of this cohort is calculated as follows:

Diagnostic Yield = 5.2%

Table 3.4. Summary of 12 shortlisted candidate variants identified in the HumGen referral cohort B

	Patient ID	Variant	Coding Impact	Zygosity	Variant Review decision
1	FRAX3026	NM_022455.5(*NSD1*):c.7567_561insC, p.(Ala2521ArgfsTer11)	Frameshift	Heterozygous	Include: Consistent with limited phenotypic data
2	FRAX3107	NM_023035.3(*CACNA1A*):c.6953_6971delTGCTGCTGCTGCTGCTGCG, p.(Pro2318ArgfsTer290)	Frameshift	Heterozygous	Include: Consistent with limited phenotypic data
3	FRAX3177	NM_006245.4(*PPP2R5D*):c.592G>A, p.(Glu198Lys)	Missense	Heterozygous	Include: Consistent with limited phenotypic data
4	FRAX3234	NM_015125.5(*CIC*):c.1721delC, p.(Pro574HisfsTer154)	Frameshift	Heterozygous	Include: Consistent with limited phenotypic data
5	FRAX3272	NM_004380.3(*CREBBP*):c.5599C>T, p.(Arg1867Trp)	Missense	Heterozygous	Include: Consistent with limited phenotypic data
6	FRAX2963	NM_001374820.1(*ARID1B*):c.3578_3579insGG, p.(Ile1193MetfsTer21)	Frameshift	Heterozygous	Exclude: Variant in homopolymer region. Needs Sanger validation
7	FRAX3168	NM_017780.4(*CHD7*):c.8738delT, p.(Leu2913ArgfsTer4)	Frameshift	Heterozygous	Exclude: Unbalanced allele ratios
8	FRAX3357	NM_001273.5(*CHD4*):c.220delC, p.(Glu74SerfsTer128)	Frameshift	Heterozygous	Exclude: Poor coverage around the variant -needs Sanger validation
9	FRAX3208	NM_017780.4(*CHD7*):c.2407delG, p.(Glu803LysfsTer4)	Frameshift	Heterozygous	Exclude: Poor coverage around the variant -needs Sanger validation
10	FRAX3166	NM_022455.5(*NSD1*):c.2093delC, p.(Ala698GlufsTer15)	Frameshift	Heterozygous	Exclude: Unbalanced allele ratios
11	FRAX3371	NM_015155.3(*LARP4B*):c.2146delG, p.(Arg716GlyfsTer37)	Frameshift	Heterozygous	Exclude: Poor coverage around the variant
		NM_017519.3(*ARID1B*):c.5922delC, p.(Phe1975LeufsTer69	Frameshift	Heterozygous	Exclude: Poor coverage around the variant

Table 3.5. Summary of five clinically relevant variants leading to molecular diagnoses in the HumGen referral cohort B

	Patient	**Variant**	**Variant Impact**	**Known/Novel and ClinVar ID**	**ACMG-AMP codes and Strength applied**	**Classification**	**Disease Association**
1	FRAX3026	NM_022455.5(*NSD1*):c.7567_561insC	Frameshift	Novel	PVS1_strong and PM2_supporting	Likely pathogenic	Sotos syndrome
2	FRAX3107	NM_023035.3(*CACNA1A*):c.6953_6971delTGCTGCTGCTGCTGCTGCG	Frameshift	Novel	PVS1_strong and PM2_supporting	Likely pathogenic	CACNA1A-related disorder
3	FRAX3177	NM_006245.4(*PPP2R5D*):c.592G>A,	Missense	Known (VCV000190286.45)	PP5_very strong, PM2_supporting, PM1_moderate and PS3_moderate	Pathogenic	PPP2R5D-related neurodevelopmental disorder
4	FRAX3234	NM_015125.5(*CIC*):c.1721delC	Frameshift	Novel	PVS1_very strong and PM2_supporting	Likely pathogenic	CIC-related disorder
5	FRAX3272	NM_004380.3(*CREBBP*):c.5599C>T	Missense	Known (RCV001091530.17)	PP5_strong, PP3_supporting PM2_supporting, PS3_strong, PM1_moderate	Pathogenic	Rubinstein-Taybi syndrome-1 Menke-Hennekam syndrome 1

3.3.6. HumGen Referral Cohort B Diagnostic Yield

Five shortlisted candidate variants in six patients exhibited variants consistent with the limited clinical presentation. These variants were subsequently interpreted and classified according to the ACMG-AMP guidelines, establishing a molecular diagnosis in this patient group (Table 3.5 above). Of the six variants, five are classified as likely pathogenic/pathogenic and a single variant is classified as a VUS. The diagnostic yield of this cohort is calculated as follows:

Diagnostic Yield = 5.2%

3.3.7. HumGen Referral Cohort B Individual Patient Result

i. FRAX3026

The patient is a 5-year-old male child who is reported to exhibit dysmorphic features, behavioural and learning problems. In this patient, we prioritised a frameshift variant in the *NSD1* gene. Disease-causing variants in the *NSD1* gene cause Sotos syndrome (OMIM 117550). The *NSD*1:c.7560_7561insC variant identified in this patient is a frameshift variant in a gene where LoF is a known mechanism of disease (Mencarelli *et al.*, 2018) (PVS1). This variant is absent in gnomAD databases (PM2). Based on the ACMG-AMP codes PVS1_strong and PM2_supporting, this variant is classified as likely pathogenic.

ii. FRAX3107

Is an 8-year-old male child who is reported to present with epilepsy and DD. We prioritised a frameshift variant in the *CACNA1A* gene in this patient. Initially, mutations in the *CACNA1A* gene were primarily associated with three disorders: episodic ataxia type 2, familial hemiplegic migraine type 1, and spinocerebellar ataxia type (Martínez-monseny *et al.*, 2021). Nonetheless, the application of cutting-edge sequencing techniques has revealed correlations between *CACNA1A* variants and an expanded phenotypic spectrum, comprising global developmental delay, intellectual disability, epileptic encephalopathy, and ASD (Kessi *et al.*, 2021; Niu *et al.*, 2022). The *CACNA1A* variant identified in this patient is a 36-bp frameshift variant in a gene where LoF is a known mechanism of disease. Additionally, the exon affects the polar residues of the CACNA1A protein (PVS1). This variant is absent in gnomAD population databases (PM2). Based on the ACMG-AMP codes PVS1_strong and PM2_supporting, this variant is classified as likely pathogenic.

iii. FRAX3177

The patient is a 2-year-old male child who is reported to exhibit developmental delay, macrocephaly and dysmorphism. A missense variant in the P*PP2R5D* gene was prioritised in this patient. Disease-causing variants in the *PPP2R5D* gene are known to cause PPP2R5D-related neurodevelopmental disorder (Houge *et al.*, 2015; Shang *et al.*, 2016; Biswas *et al*., 2020). The *PPP2R5D*:c.592G>A variant identified in this patient has been previously reported in the literature as de novo in cases with variable developmental delay, ASD, and intellectual disabilities (Biswas *et al.*, 2020; Reynhout *et al.*, 2019). This variant is classified pathogenic, with a 2-star rating on ClinVar (VCV000190286.45) (PP5). The variant is absent from gnomAD population databases (PM2) and is located within the regulatory B subunit domain of the PPP2R5D protein (PM1). In addition, functional studies show damaging effect on the gene product (PS3) (Houge *et al.*, 2015). Based on ACMG-AMP codes PP5_very strong, PM2_supporting, PM1_moderate and PS3_moderate, the variant is classified as pathogenic.

iv. FRAX3234

An 11-year-old male child who is reported to exhibit intellectual disability. A frameshift variant in *CIC* gene was prioritised in this patient. Disease-causing variants in this gene were primarily associated with a wide range of NDDs (Iossifov *et al.*, 2012; Singh *et al.*, 2021). More recently, they have also been associated with cerebral folate deficiency syndrome (Cao *et al.*, 2021). The *CIC*: c.1721delC variant identified in this patient is a frameshift variant in a gene where LoF is a known mechanism of disease (PVS1). Additionally, this variant is absent from gnomAD population databases (PM2). Based on the ACMG-AMP codes PVS1_very strong and PM2_supporting, the variant is classified as likely pathogenic.

v. FRAX3272

Is a 9-year-old male child who is reported to exhibit speech and language delay. A frameshift variant was identified in the *CREBBP* gene in this patient. Causal heterozygous variants in the *CREBBP* gene are associated Rubinstein-Taybi syndrome 1 and Menke-Hennekam syndrome 1 (Menke *et al.*, 2018; Sima *et al.*, 2022). The *CREBBP*:c5599C>T variant identified in this patient is classified pathogenic in ClinVar, with a 2-star rating (RCV001091530.17) (PP5). Computational tools predict a deleterious effect on the gene product (PP3). Furthermore, this variant is absent from gnomAD population databases (PM2). Based on the ACMG-AMP codes PP5_strong, PP3_supporting PM2_supporting, PS3_strong, and PM1_moderate, the *CREBBP* variant identified in this patient is classified as pathogenic.

3.4. Genetic Variants Landscape: Insights from Dual Cohorts

3.4.1. Variants Documented in ClinVar Observed in this Study

A total of 10 /19 variants (53%), classified as likely pathogenic/pathogenic and correlating with the clinical presentation of the patients, are known variants reported in ClinVar and 47% of the variants are novel variants expanding the mutation spectrum of *ARID1B, CACNA1A, CHD7, CIC, MEF2C, NSD1* and *PTEN* genes.

3.4.2. Distribution of Clinically Relevant Disease-Causing Variants Across Genes

Disease-causing variants were identified in a total of 19 patients in 15 genes from the 285-gene panel. Most genes had a single variant identified, with *CHD7* having three variants and *NSD1* having two variants. Among the 16 mutated genes, there were two X-linked genes and 14 AD genes. Interestingly, 13/15 genes were curated as definitive on ClinGen with only two genes (*CACNA1A* and *CIC*), remaining with pending curations. Moreover, all 15 genes were rated green on PanelApp, indicating an established gene-disease association and recommendation for inclusion in diagnostic patient workup. Table 3.6 below summarises the number of disease-causing variants identified in each gene.

Table 3.6. Distribution of disease-causing variants across genes

	Gene	**OMIM**	**Inheritance**	**# of Disease-causing variants**	**ClinGen Curation[1]**	**PanelApp Curation[1]**
1	*ARID1B*	614556	AD	1	Definitive	Green
2	*AUTS2*	607270	AD	1	Definitive	Green
3	*CACNA1A*	601011	AD	1	Not curated	Green
4	*CHD7*	608892	AD	3	Definitive	Green
5	*CIC*	612082	AD	1	Not curated	Green
6	*CREBBP*	600140	AD	1	Definitive	Green
7	*DDX3X*	300160	XL	1	Definitive	Green
8	*MECP2*	300005	XL	1	Definitive	Green
9	*NSD1*	606681	AD	2	Definitive	Green
10	*PACS1*	607492	AD	1	Definitive	Green
11	*PPP2R1A*	605983	AD	1	Definitive	Green
12	*PPP2R5D*	601646	AD	1	Definitive	Green
13	*PTEN*	601728	AD	1	Definitive	Green
14	*SON*	182465	AD	1	Definitive	Green
15	*STXBP1*	602926	AD and AR	1	Definitive	Green

[1]ClinGen and PanelApp curation status evaluated on the 15th of October 2023

3.4.3. *De novo* and Inherited Disease-Causing Variants Observed in this Study

Although DNM-enriched genes are known for their enrichment of *de novo* disease-causing variants in individuals with DDs, they also encompass inherited variations. Expectedly, the majority of causal variants (47%, 9/19), identifed in this study are *de novo*. A small percentage (5%, n=1/19) were inherited maternally, while for 47% (n=9/19), determination of inheritance was was not feasible due to the absence of parental samples or data necessary for segregation analysis.

3.4.4. Variants of Uncertain Significance Observed from this Study

Amongst candidate variants shortlisted in this study. A single variant consistent with the clinical phenotype of the patient is classified as a “hot” VUS (*CDK13*:c.2620G>A). The classification of variants as VUSs arises from a lack of conclusive evidence regarding their involvement in the disease during the initial discovery phase. Resolving VUSs is feasible through periodic reassessment, driven by emerging evidence (Chen *et al.*, 2023), and/or the adoption of trio-based testing. This comprehensive approach involves sequencing both parents along with the affected individual, presenting a robust strategy to elucidate the variant's inheritance status. Consequently, this approach may facilitate the potential reclassification of the variant, either upgrading or downgrading it to pathogenic or benign categories, respectively.

3.5. Factors Influencing the Diagnostic Yield of DDs

3.5.1. Assessing the Impact of Phenotypic Data Availability on Molecular Diagnosis

An odds ratio was calculated to assess the influence of the availability of phenotypic data on establishing a molecular diagnosis

$$\text{Odds Ratio (OR)} = \frac{\text{Odds of diagnostic yield in cohort without clinical phenotype data}}{\text{Odds of diagnostic yield in cohort with clinical phenotype data}}$$

$$\text{OR} = \frac{14/82}{5/91}$$

$$= 3.11 \text{ (95\% CI: 1.486 - 6.563)}$$

Therefore, these results indicate that the odds of establishing a molecular diagnosis in a cohort with detailed phenotypic data are approximately three times higher than in a cohort without such data

3.5.2. Assessing the likelihood of establishing a diagnosis in Trios, Duos, and Singletons Cohorts

i. Trios vs. Duos

The odds ratio (Trios vs. Duos) was calculated by comparing the odds of establishing a diagnosis in trios from Cohort A to duos in the same cohort

$$\text{Odds Ratio (OR)} = \frac{\text{Odds of establishing a molecular diagnosis in trios}}{\text{Odds of establishing a molecular diagnosis in singletons}}$$

$$\text{OR} = \frac{8/41}{1/37}$$

OR = 7.22 (95% CI: 1.00-8.38)

The results indicate that the odds of establishing a diagnosis in trios are approximately seven times higher than the odds of establishing a diagnosis in duos.

ii. Trios vs. Singletons

The odds ratio (Trios vs. Singletons) was calculated by comparing the odds of establishing a diagnosis in trios from cohort A to singletons from the same cohort

Odds of Diagnosis in Trios in Cohort A (n=8/48)

Odds of Diagnosis in Singletons in Cohort A (n=5/9)

$$\text{OR} = \frac{8/41}{5/4}$$

OR = 0.156 (95% CI: 0.24-13.25)

These results suggest that the odds of establishing a diagnosis in trios are approximately 0.156 times the odds of establishing a diagnosis in singletons. Thus, patients from singletons are more likely to receive a diagnosis compared to patients from trios in this cohort. However, it's important to note that the 95% confidence interval (CI) ranges from 0.24 to 13.25, indicating a wide range of uncertainty in this estimation.

3.5.3. Variants in Complex Genomic Regions May need Secondary Confirmation

Determining the impact of variants within intricate genomic regions presents formidable challenges in genomics. A notable example is the complexity introduced by variants in homopolymer regions. The intricacies arise during the sequencing of regions containing repeating identical bases, as the incorporation of these bases occurs simultaneously in the same synthesis cycle, leading to uncertainties in base calling. Homopolymer regions can yield imprecise results, presenting as strand bias and low-quality. In this study, there were five variants in the *CHD4*, *ARID1B*, *CHD7* and *LARP4B* genes. These variants were excluded from further analysis due to uncertainties regarding their existence and financial constraints, preventing their confirmation through a secondary conformation method. While these cases were excluded in these cases, it is imperative to highlight that future endeavours should incorporate confirmatory analyses to resolve such cases and measures to address such cases should be considered when contemplating the integration of this panel into clinical settings. Below, two representative cases exemplifying variants located within homopolymer region are presented.

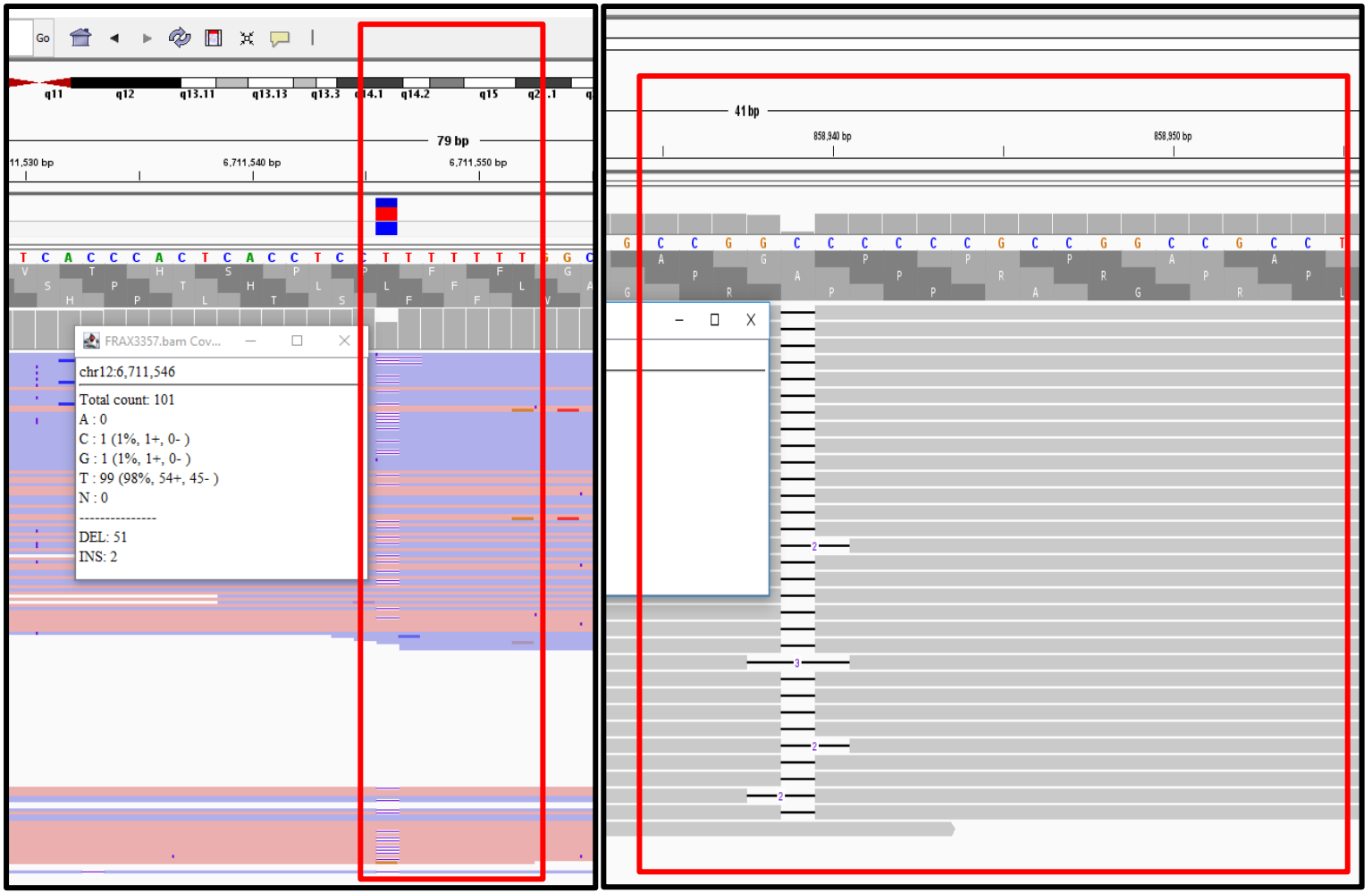


Figure 3.3. Examples of variants in homopolymer regions. A: Depicts the *CHD4*: c.220del variant identified in FRAX3357. B: Shows the *LARP4B*: c.2146del variant identified in FRAX3371. These variants were excluded from subsequent analysis due to their presence within repetitive cytosine and thymine stretches, respectively.

3.6. Diagnostic Yield Comparison of Genetic Testing Strategies for DDs

In Table 3.7 below, the diagnostic yield of various genetic testing methods employed in the diagnostics of DDs is summarised. Notably, the DNM-Enriched gene panel, designed in this study, demonstrated a diagnostic yield of 14.6% for DDs and 5.2% for developmental delay, positioning it as a promising approach in deciphering the genetic aetiology of DDs and developmental delay in resource-constrained settings where the currently recommended full WES and/or WGS may not be feasible.

Table 3.7. Genetic testing strategies and diagnostic yields in DDs

Diagnostic test	**Types of variants detected**	**Diagnostic Yield**
Karyotyping	Chromosomal aberrations	< 5% (Hochstenbach *et al.*, 2009)
FISH	sub microscopic deletions and duplications	< 5% (Hochstenbach *et al.*, 2009 ; Ravnan *et al.*, 2006)
MLPA	CNVs	2%- 13% (Boggula *et al*., 2014 ; Shin *et al.* 2015)
FXS testing	*FMR1* triplet repeat expansion	0.5% - 2% (Essop and Krause, 2013; Weinstein *et al.*, 2017)
CMA	CNVs	11% to 36% (Jang *et al.*, 2019 ; Kim *et al.*, 2018 ; Miller *et al.*, 2010)
WES	SNVs and CNVs	21% - 57% (da Cunha Leite Leite *et al.*, 2022 ; Gao *et al.*, 2019 ; Ko & Chen, 2023)
WGS	SNVs and CNVs	21% - 63% (Abe-Hatano *et al.*, 2021 ; Ko & Chen, 2023 ; Sun *et al.*, 2021 ; Zahir *et al.*, 2017)
DNM-enriched gene panel designed in this study	SNVs	14.6% (DDs) and 5.2% (limited phenotype, developmental delay)

3.7. The Contribution of a DNM-Enriched Gene Panel in Ending a Diagnostic Odyssey

Among the 14 patients from the phenotyped DDD-Africa cohort A with an established molecular diagnosis, the age at the first genetic clinic varied between less than a year and 6 years, with a molecular diagnosis being established at an average age of 6.64 years (range: 6-10 years). Additionally, these patients received a molecular diagnosis after an average period of 5.07 years (range: 3-10 years) from the start of their diagnostic odyssey (Table 3.8). This underscores the crucial role played by the DNM-enriched gene panel in effectively concluding complex diagnostic journeys in these individuals.

Table 3.8. Timeline of Molecular Diagnoses in the DDD-Africa cohort A.

Patient #	**Patient ID**	**Year of birth**	**Year of first genetic clinic**	**Age at first genetic clinic**	**Year in which a molecular diagnosis was established through this study**	**Age of molecular diagnosis**	**Period of diagnostic odyssey**
1	D3S_0028	2015	2015	< 1 year	2021	6 years	6 years
2	D3S_0034	2015	2015	< 1 year	2021	6 years	6 years
3	D3S_0047	2013	2015	2 years	2021	6 years	6 years
4	D3S_0050	2014	2017	3 years	2021	7 years	4 years
5	D3S_0051	2016	2018	2 years	2021	6 years	3 years
6	D3S_0052	2013	2015	2 years	2021	6 years	6 years
7	D3S_0053	2016	2016	< 1 year	2021	5 years	5 years
8	D3S_0082	2008	2011	3 years	2021	7 years	10 years
9	D3S_0086	2010	2016	6 years	2021	10 years	5 years
10	D3S_0089	2010	2015	3 years	2021	7 years	6 years
11	D3S_0094	2017	2019	2 years	2021	6 years	2 years
12	D3S_0111	2010	2015	5 years	2021	9 years	6 years
13	D3S_0117	2013	2017	4 years	2021	8 years	4 years
14	D3S_0124	2011	2017	6 years	2021	10 years	4 years

Chapter 4

Discussion

4. DISCUSSION

In the evolving landscape of genomics, there is a notable shift towards translational applications, aiming to bring genomic discoveries into clinical use, fostering precision medicine. Nevertheless, this paradigm shift poses a challenge for African populations due to the limited availability of genomic and implementation data in this region, hindering their complete engagement with the benefits of these advancements. To harness genomics in African healthcare fully, it is essential to generate appropriate datasets specific to African populations, capturing both disease-causing and non-pathogenic genetic variations linked to genes causing various disorders, including DDs (Lumaka *et al.,* 2022; Omotoso *et al.,* 2022; Owolabi *et al.,* 2023).

In terms of the genetic diagnostic evaluation of DDs, international guidelines recommend WES or WGS as first-tier genetic analysis (Manickam *et al.,* 2021; Srivastava *et al.,* 2019). While genomics has made considerable advancements, leading to notable cost reductions for WES and WGS per sample, implementing these techniques in LMIC settings, such as the South African State Healthcare system, is hindered by additional challenges that impair widespread adaptation including acquiring key sequencing and appropriate equipment, computational infrastructure, bioinformatics expertise, and clinical referral.

Given these challenges, there is a need to explore alternative, innovative, and resource-effective strategies to address the diagnostic challenges associated with DDs in these regions. The development of cost-effective NGS panels, strategically curated with a focused selection of genes associated with the diseases of interest, may emerge as a practical compromise. This project aimed to evaluate the feasibility and utility of a DNM-enriched gene panel as a routine diagnostic strategy for DDs and developmental delay in a resource-constrained setting. We explore the practical application of targeted sequencing encompassing 285 DNM-enriched genes in a well-phenotyped and an unphenotyped cohort with DDs and developmental delay, respectively.

This chapter will begin with a discussion on the design, curation, and optimisation of the DNM-enriched gene panel. Subsequently, it delves into the diagnostic yield observed within the two cohorts investigated in this study. Additionally, factors influencing the diagnostic yield of the DNM-enriched strategy will be discussed. The subsequent sections will address the diagnostic feasibility of a DNM-enriched strategy, followed by a discussion on the personal- and clinical

utility of molecular diagnoses in our patient population. The chapter will draw to a close with an exploration of project limitations, specific considerations pertinent to the implementation of genetic testing in LMICs, and an overall conclusion section.

4.1. Design, Curation and Optimisation of the DNM-Enriched Gene Panel

Various comprehensive guidelines, recommendations, and standards have been published to direct the development, validation, and implementation of NGS-based clinical genetic testing (Matthijs *et al.,* 2016; Rehm, 2013), facilitating the integration of NGS into standardized frameworks for clinical genetic tests. There are various strategies used in designing a gene panel. In our study, our objective was to create a gene panel targeting fewer than 300 genes, ensuring compatibility with any benchtop NGS system. With hundreds of genes associated with DDs and developmental delay, selecting a subset for inclusion in a custom gene panel posed a challenge. To address this, we focused on 285 genes previously identified as enriched for DNM in studies of DDs and/or developmental delay. Utilising variation intolerance as a curation tool, we hypothesized that genes enriched for DNMs would demonstrate intolerance to variation, thereby increasing their likelihood of contributing to DDs.

The majority (n = 229/285), of the genes were predicted to be intolerant to variation using the pLI and/or missense z-score. Aligning with previous findings suggesting that most genes enriched for DNMs in cohorts with DDs, autism, and intellectual disabilities exhibit intolerance to variation (Hormozdiari *et al.*, 2015; Nguyen *et al.*, 2018; Nguyen *et al.*, 2020). The observed intolerance is likely due to the involvement of these genes in important developmental pathways such as embryonic brain development, chromatin organisation, nucleoplasm, chromosome organization, histone methyltransferase complex and regulation of gene expression (Guo *et al.*, 2022) . Notable, 21% (n=56) were not predicted to be intolerant to any variation based on the pLI and missense Z-score metrics. Despite this, these genes were not excluded from the gene panel, as studies have indicated that the pLI score may not be suitably adapted to genes associated with AR disorders (Betancur & Buxbaum, 2020). Higher pLI scores correlate with AD genes, while low pLI scores are associated with AR genes (Fridman *et al.,* 2021). The decision to retain these 56 genes in the panel recognises the significance of AR variations, even though the majority of DDs arise from causal variations in AD and X-linked genes.

At the onset of this study, the ClinGen and PanelApp gene curation platforms were not publicly accessible. However, as they became available during the study period, we retrospectively

evaluated and incorporated their gene curation assessments and classifications for the 285 DNM-enriched genes. The majority of genes of the 285 DNM enriched genes are curated as definitively associated and/or diagnostic genes in ClinGen and PanelApp. Few genes are showed moderate and/or limited gene-disease associations evidence of association with DDs and/or developmental delay. A study conducted by McGlaughon *et al.* (2018) examined the progression of ClinGen gene clinical validity classifications for 30 genes over time. Their findings revealed that gene-disease associations typically spend minimal time in the "moderate classification" before progressing to "Strong" or "Definitive" classifications. The study also highlighted that a majority of genes initially classified as having a moderate association eventually advance to the strong or definitive category, supporting the rationale for including genes with moderate statuses in targeted gene panels. Furthermore, the research indicated that genes with limited gene-disease associations for at least five years are more likely to remain in the "limited" category or downgrade to "Disputed/Refuted" (McGlaughon *et al.,* 2018).

The utilisation of the DNM-enriched gene panel strategy in this study showed promising results, but further refinement and optimisation could enhance its effectiveness. Improvements may involve excluding genes identified as tolerant to variation based on pLI and/or z-scores, as well as those categorised as amber or red on PanelApp, or showing limited or moderate evidence of gene-disease validity on ClinGen, given that no causal variants were found in these gene groups in our study. Nonetheless, it's crucial to acknowledge the limitation of our small sample size. Additionally, leveraging curation resources like ClinGen, PanelApp, and the developmental disorder genotype-phenotype database could facilitate the optimisation process and enhance gene selection. Moreover, insights from WES and/or WGS research initiatives on DDs could provide valuable guidance for further refining the gene panel

4.2. Diagnostic yield of DNM-Enriched Gene Panel in the two Cohorts

This study focused on two cohorts: DDD-Africa cohort A, comprised of individuals with DDs, and HumGen referral cohort B, consisting of patients with developmental delay. Cohort A, benefiting from ongoing genetic services in our participating clinics, is under the care of our in-house medical geneticists and genetic counsellors, providing comprehensive phenotypic data recorded in HPO terms. This cohort included trios, duos, and singletons. In contrast, the HumGen referral cohort B, consisting predominantly of singleton cases referred for FXS testing at our Division, lacks extensive phenotypic and family history data. This cohort mirrors real-world conditions in many LMICs, where a shortage of trained medical geneticists poses challenges to appropriate clinical data collection.

The diagnostic yield in the phenotyped DDD-Africa cohort A was 14.6% (n = 14/96). A positive case was defined as one in which a possible candidate variant was identified in a gene consistent with the clinical presentation of the patient and classified as either likely pathogenic or pathogenic according to the ACMG-AMP variant interpretation guidelines. The diagnostic yield of this cohort aligns with the global diagnostic rate of targeted gene panels in neurodevelopmental disorders and DDs ranging between 8% to 27% (Aspromonte *et al.,* 2019; Mellone *et al.,* 2022).

The diagnostic yield in the unphenotyped HumGen referral cohort B was 5.2%, falling significantly below the reported literature range of 9.8% - 32% (Grozeva *et al.,* 2015; Redin *et al.,* 2014; Yan *et al.,* 2019). The lack of pre-screening by a medical geneticist in this cohort resulted in limited phenotypic information. Testing our panel in both well-phenotyped (cohort A) and unphenotyped (cohort B) settings allowed us to assess its performance under varying conditions, considering the scarcity of clinical genetics expertise in many LMIC settings. The findings from cohort B underscore the importance of clinical gatekeeping in optimising the effectiveness of a panel-based intervention for improving DD diagnostic rates. Gatekeeping strategies, recommending NGS-based tests based on clinical indications, prove crucial in resource allocation, minimising financial burdens, and enabling a focused and cost-effective approach, particularly in settings with limited resources. Additionally, cohort B consisted solely of singletons, while research indicates that sequencing/analysing additional family members and considering family history increases the likelihood of establishing a molecular diagnosis (Farwell *et al.,* 2015; Wright *et al.,* 2023).

4.3. Factors Affecting the Diagnostic Yield Obtained from this Study

4.3.1. Availability of detailed phenotypic data

As previously mentioned, this study investigated two cohorts (DDD-Africa cohort A- with phenotypic data and the HumGen referral cohort B –without detailed phenotypic data). The odd ratio analysis revealed a significant correlation between the availability of detailed clinical phenotype data and the likelihood of obtaining a molecular diagnosis, with a substantial increase by a factor of 3.35 between the two cohorts. These findings align with existing research that underscores the paramount importance of deep phenotyping in elevating diagnostic rates for rare diseases, including DDs. The integration of phenotypic data, represented through HPO terms in the analysis, proves to be effective in linking phenotypes to genes and diseases. This approach significantly aids in narrowing down potential disease-

causing variants, thereby enhancing the overall diagnostic rate (Köhler *et al.*, 2019; Zhao *et al.*, 2020). Moreover, detailed phenotypic data plays a pivotal role in guiding the selection of the most appropriate genetic tests and identifying patients who would derive the greatest benefit from genetic testing (Seyhan & Carini, 2019). This is particularly crucial as some patients may be excluded from genetic testing based on clinical information, thereby preventing the inappropriate use of resources This aspect is particularly crucial when considering high-cost genetic tests such as NGS-based tests, especially in LMICs. The utilisation of detailed phenotypic information allows for a more cautious allocation of scarce resources for testing, ensuring optimal use and efficiency in these resource-constrained settings. Highlighting the need to develop a base level of clinical phenotypes in these settings.

4.3.2. Trios Versus Singleton Analysis

The odds of establishing a molecular diagnosis within the cohort with trios (DDD-Africa cohort A), were assessed across different family structures. When comparing trios to duos in the phenotyped DDD-Africa cohort A, the odds ratio was 7.22 (95% CI: 1.00-8.38), indicating that the likelihood of achieving a diagnosis 7.22 times higher than the odds in duos. Contrasting trios with singletons yields an odds ratio of 0.156 (95% CI: 0.24-13.25), suggesting lower odds of diagnosis in trios compared to singletons. However, the wide confidence interval underscores uncertainty in this estimation, prompting further investigation. Discrepancies in cohort sizes, with trios (n=49) outnumbering singletons (n=9), may introduce bias and widen confidence intervals, diminishing precision in odds ratio estimation.

Trio-based testing has the potential to streamline NGS data analysis and significantly enhance the chances of achieving a conclusive molecular diagnosis. This was evidenced in a recent study conducted by DDD-UK researchers, which found that children undergoing family trio analysis were nearly five times more likely to receive a definitive diagnosis compared to those without neither parent available (Wright *et al.*, 2023). The significance of parental testing in the diagnostic workflow is also reflected in the ACMG/AMP guidelines on variant classification criteria (Richards *et al.*, 2015) . Parental analysis allows prioritised variants to be classified as *de novo* or bi-allelic respectively, thus enabling application of codes PS2/PM6 and PM3 criteria respectively aiding variant classification and impacting the final diagnostic yield. While trio-sequencing and analysis are acknowledged for their diagnostic superiority, their cost poses a significant challenge, particularly in LMICs where limited financial resources need to cater to large populations (Baine-Savanhu *et al.*, 2023).

While trio sequencing/analysis is acknowledged for improving diagnostic rates, it may not be necessary and resource-effective for all cases. For example, among the 14 causal variants identified in the DDD-Africa cohort A, only 21% (n=3/14) required *de novo* ACMG-AMP codes for classification as likely pathogenic or pathogenic. In contrast, the remaining 78% (n=11/14) were variants that would have been classified as likely pathogenic or pathogenic regardless of their *de novo* status. This finding demonstrates that singleton analysis of well-curated genes can be effective in low resource settings. This cost-effective approach optimises budget allocation by identifying causal variants in most patients, with trio sequencing reserved for a small subset of the cohort and/or as a reflex test where understanding the inheritance pattern could assist in establishing a diagnosis in the proband. In settings where trio NGS is impractical, screening only the proband using a well-curated gene panel can establish a molecular diagnosis in most cases. This approach would then allow for targeted additional family testing in a small subset of cases, which can be easily conducted through less-costly targeted genetic testing strategies.

4.3.3. Sequencing Platform Influence on Diagnostic Yield

In this study, the two cohorts underwent sequencing using two different platforms, Illumina NovaSeq (DDD-Africa cohort A) and Ion Torrent S5 (HumGen referrals cohort B). Our evaluation focused on the error profiles associated with each platform, considering their potential impact on diagnostic rates. Despite the Ion Torrent system achieving an average coverage of 120x in the HumGen referral cohort B, data generated from this platform exhibited a higher incidence of indel calls, especially in homopolymer regions. The consistency of these indel calls across multiple samples raised concerns about potential false positives. Conversely, the Illumina NovaSeq platform, with an average coverage of 40x in the DDD-Africa cohort A, did not show similar problematic indel patterns. Indels from the Ion Torrent system with questionable quality control metrics and situated in complex regions, were excluded as likely false positive. It is crucial to highlight that both platforms demonstrated effectiveness. Therefore, laboratories in LMICs can design panel-based diagnostic strategies based on locally available resources without the necessity for special capital acquisitions. However, careful consideration of technology-specific issues is crucial in ensuring accurate and reliable results.

4.4. Genetic Profiles Observed from Patients in this Study

Establishing a molecular diagnosis in DDs poses a significant challenge due to the pronounced phenotypic and genetic diversity observed among affected individuals. This complexity is believed to be further compounded in Africa, where the scarcity of dysmorphology and genomic data in public repositories adds to the challenges (Lumaka *et al.,* 2022). This heterogeneity makes it challenging to establish a specific diagnostic process, often resulting in individuals with DDs undergoing multiple genetic tests in pursuit to delineate the underlying genetic aetiology.

Regrettably, this extensive testing process does not always result in a conclusive diagnosis in the majority of patients. In both cohorts studied, 19 patients presented with 19 distinct causal variants in 16 genes. Of these 16 genes, 13 are associated with autosomal dominant inheritance *(CHD7, PPP2R1A, ARID1B, AUTS2, SON, PTEN, NSD1, PACS1, CACNA1A, PPP2R5D, CIC, CDK13, STXBP1* and *CREBBP*) and two genes (*MECP2* and *DDX3X*) are X-linked. Notably, some of these genes are linked to specific syndromic phenotypes, while others are associated with non-specific DD phenotypes. These findings underscore the importance of implementing well-curated multigene panels for the genetic evaluation of DDs, as single gene testing will only be effective in establishing a molecular diagnosis in a very limited number of cases. Furthermore, among the 19 unique causal variants, the majority are LoF variants (frameshift and nonsense), accounting for a cumulative 58% (n=11/19) of causal variants identified. Additionally, out of the 19 variants, 10 have been previously reported and documented in ClinVar, while nine variants are novel, expanding the mutations spectrum, especially for *ARID1B, CACNA1A, CHD7, CIC, NSD1*, *PTEN* and *SON*.

4.5. Validation of NGS Data

The validation of variants identified through NGS remains a topic of debate in the diagnostic and research communities. Despite the ongoing discourse, advancements in NGS technology have significantly reduced error rates in newer technologies (Cheng *et al.,* 2023; Satam *et al.,* 2023). Nevertheless, disagreements persist on whether validation of NGS data through Sanger sequencing, is necessary, considering the associated costs and increased turnaround time. In the study by Strom and colleagues, validation of 94 unique variants from 110 genes identified through NGS in 144 exomes, yielded only one discordant low-quality variant. This study concluded that Sanger confirmation remains essential for low-quality SNVs and indels <10 bp.

The study also concluded that, under conditions of high coverage and high-quality scores, the likelihood of detecting false positives from NGS might be lower than that of detecting false negatives by Sanger sequencing (Strom *et al.,* 2014) . In 2016, Beck and colleagues evaluated Sanger-based (Cheng *et al.,* 2023; Satam *et al.,* 2023), based validation of over 5,800 NGS-derived SNVs from 19 genes in 684 samples, revealing a notably high validation rate of 99.965%. A notable finding from this study underscored the constraints of Sanger sequencing, as 19 out of over 5,800 variants derived from NGS were not initially confirmed. This prompted the design of new Sanger primers, and subsequently, 17 out of the 19 NGS variants were successfully validated. Consequently, the authors concluded that relying solely on a single round of Sanger sequencing is more prone to incorrectly rejecting a true positive variant identified by NGS than accurately identifying a false positive variant (Beck *et al.,* 2016).

Mu and colleagues observed a 99/7845 (1.3%) false positive rate and concluded that Sanger confirmation is needed to maintain high accuracy, particularly in difficult-to-sequence regions (Mu *et al.,* 2016). In contrast, Zheng and colleagues argued against the necessity of validating all NGS variants. Their study concluded that newly identified variants obtained through a well-validated capture based NGS workflow, meeting stringent quality standards (≥35× depth coverage and ≥35% heterozygous ratio), do not require secondary validation and can be directly reported (Zheng *et al.,* 2019).

In a 2021 study, researchers assessed the need to confirm NGS results with Sanger sequencing in 1109 variants of 425 genes from 825 clinical exomes, achieving a 100% concordance rate. The study concluded that Sanger confirmation is no longer always necessary but can be especially useful as an internal quality control. The findings from these studies suggest that confirmation of NGS results using Sanger sequencing may be redundant in many cases. They further highlighted the importance of establishing stringent NGS QC criteria (Arteche-López *et al.,* 2021).

In this study, all causal variants identified from the DDD-Africa cohort A underwent Sanger confirmation. However, this validation was not applied to all cases recruited in the DDD-Africa study; instead, it was selectively performed on a subset of the first 100 case variants. This subset served as both an internal QC measure and contributed to the evaluation of the necessity of Sanger validation, aiding in the establishment of appropriate NGS QC criteria in our setting. In contrast, causal variants identified in the HumGen referral cohort B were not subjected to

Sanger sequencing validation. This decision was because the identified causal variants in this cohort met the recommended NGS QC criteria outlined in the studies by Strom *et al* and Zheng *et al* (Strom *et al.,* 2014 and Zheng *et al.,* 2019), as well as those established from the validation of variants obtained in the DDD-Africa cohort A. In routine diagnostic NGS tests, Sanger validation is often deemed optional, given its significant impact on the implementation and associated cost of NGS-based testing, particularly in resource-limited settings. Nevertheless, ensuring strict accuracy and reproducibility of NGS-based genetic testing metrics remains imperative. Therefore, in LMICs, it is recommended to initially validate the first few batches of NGS data and establish stringent QC measures to abide by. Subsequently, it may be prudent to limit validation efforts to clinically relevant variants that do not meet the established QC metrics. This approach ensures cost-effectiveness while maintaining the accuracy and reliability of diagnostic outcomes

4.6. Diagnostic Feasibility of the DNM-Enriched Gene Panel as Designed in this Study

4.6.1. Focused Data Analysis and Reduced Turnaround Time

Targeted gene panels generally yield less candidate variants for review (Dillon *et al.,* 2018; Xue *et al.,* 2015). The DNM-enriched gene panel analysis strategy employed in this study yielded an average of two rare (MAF ≤1%), nonsynonymous and/or indels variants (range: 1-3 variants per individual). This reduced the amount of time spent on data analysis. On average, each patient's data analysis required approximately 1.5 hours, and during our variant review meetings, an extra 45 minutes was spent reviewing each shortlisted/prioritised variant. Targeted gene panels not only enhance the speed of analysis but also contribute to quicker clinical decision-making, providing a more efficient and timely diagnostic process for patients with DDs. Automation and bioinformatics pipelines play a crucial role in NGS data analysis (Cabello-Aguilar *et al.,* 2023). In this study, filtering out synonymous variants and common nonsynonymous variants (defined as variants with a MAF of ≥1%) from the annotation output file was conducted manually. However, implementation of bioinformatics scripts to automate this process could have significantly reduced the already short TAT for variant analysis of the cohorts in this study. While the analysis of data from a targeted gene panel can be performed without extensive bioinformatics skills, handling larger datasets, such as those generated from WES and WGS, requires advanced analysis and bioinformatics proficiency. The limited availability of these skills poses a significant hurdle to implementing WES and/or WGS, particularly in routine settings with resource constraints.

Furthermore, precision in clinical variant interpretation and classification is of utmost importance. Achieving this precision relies on adopting standardised approaches to scrutinise genomic content and correlate clinical findings with existing medical literature (Giles *et al.*, 2021). The emergence of genomic curation and analysis platforms like DECIPHER (https://www.deciphergenomics.org/), ClinGen variant curation interface (https://curation.clinicalgenome.org/), and Franklin (https://franklin.genoox.com/clinical-db/home), is crucial. These platforms serve as essential decision support tools for variant-classification and sharing, contributing significantly to the timely interpretation of variants, and expediting the identification of and recording of causal variants. Moreover, the integration of artificial intelligence and machine learning technologies into genomics is transforming the landscape. This evolution, highlighted by researchers such as Alharbi & Rashid (2022), Caudai *et al.* (2021), Jovčevska (2020), and Quazi (2022), equips genomics with the capability to process vast amounts of data swiftly (Alharbi & Rashid, 2022; Caudai *et al.*, 2021; Jovčevska, 2020; Quazi, 2022). This technological advancement is instrumental in the rapid identification of causal variants, ultimately reducing TAT. Healthcare professionals can leverage this efficiency to devise timely and effective management plans, highlighting the evolving role of technology in enhancing patient care.

4.6.2. Low Yield of Variants of Uncertain Significance (VUSs)

Targeted gene panels yield fewer variants classified as VUS as compared to WES and WGS. The management of VUS presents complexities for both patients and clinicians, giving rise to considerations about disclosure, patient counselling, implications for clinical management, and the approach to subsequent studies (Hoffman-Andrews, 2017). To streamline our data analysis, we systematically excluded variants exhibiting evidence of benign or likely benign status. Our primary focus was directed towards variants falling into distinct categories, which include strong VUS, as well as likely pathogenic and pathogenic variants. Among the 30 patients with unique candidate variants, only one variant in a single patient received a classification of a strong VUS. It is noteworthy that the ACMG guidelines recommend that a VUS should not influence clinical decision-making, highlighting the need to pursue efforts to resolve its classification (Richards *et al.*, 2015).

Therefore, as we strive to develop effective diagnostic approaches for DDs in resource-constrained settings, prioritising strategies that could result in fewer VUSs becomes crucial. The utilisation of well-curated targeted gene panels emerges as one such strategic option. However, it is important to highlight that while targeted gene panels may be suitable for routine

diagnostics in resource-constrained settings, research endeavours involving WES and WGS should persist. These research initiatives would assess the efficacy of these strategies, as their effectiveness in limited resource settings remains incompletely explored. Such studies may uncover novel genes associated with DDs, contributing valuable insights for optimising gene panels. Considering the dynamic nature of genomics knowledge, regular updates, and adaptation of panels to emerging genetic discoveries should be prioritised to enhance their diagnostic efficacy over time. Research involving WES and WGS plays a crucial role in achieving this goal. It is pertinent to note that the persistence of VUS, attributed to insufficient information and the lack of comprehensive population data, poses a challenge across various NGS-based strategies, including WES and WGS. Even with the comprehensive nature of WES and WGS, the complexities of the human genome and evolving genomic knowledge may contribute to encountering VUS.

4.6.3. High Sequencing Coverage

Whole-exome sequencing of the HumGen referral cohort B yielded an average coverage of 120x (range: 43x-162x), whereas WES of the DDD-Africa cohort A achieved an average coverage of 40x. While our study did not involve the use of capture-based targeted gene panels, the high coverage obtained through the WES approach is noteworthy. Studies have documented that high sequencing coverage, consistently exceeding ≥100x, is achievable through capture-based targeted gene panel sequencing on both the Ion Torrent and Illumina platforms (Lorenz *et al.,* 2020; Malapelle *et al.,* 2015). The substantial coverage provided by the WES approach in this study played a pivotal role in minimising errors in variant calling, thereby reducing the likelihood of both false positives and false negatives. Consequently, this eliminates the need to validate most variants, mitigating the costs associated with validating potential disease-causing variants through additional methods. This is especially significant in resource-constrained settings, where limited resources for validation could be more efficiently allocated, specifically towards validating variants in challenging regions such as homopolymer regions.

4.7. The Contribution of DNM-Enriched Gene Panel in Reducing the Diagnostic Odyssey

The DDD-Africa cohort A consisted of patients already receiving genetic services at the Division's participating genetic clinics. Patients in this cohort underwent their initial genetic assessments at an average age of 2.8 years, ranging from 1 to 6 years. Subsequently, they were enrolled in the DDD-Africa research at an average age of 7.8 years (range: 5.3 - 11 years). The

process of receiving a conclusive diagnosis occurred later, with patients obtaining a molecular diagnosis after an average period of 7.2 years (range: 4 - 12 years), from the initial assessment. The diagnostic odyssey period can be an enduring journey. Previous studies using WES in DDs have reported concluding the diagnostic odyssey for rare genetic disorders after 10 to 40 years from the initial date of genetic assessment (Flynn *et al.*, 2021; Lam, 2021). The prolonged duration of the diagnostic odyssey underscores the complexity and challenges of establishing a molecular aetiology in DDs and the shortcoming of the traditional genetic testing strategies. Our findings highlight the effectiveness of a targeted gene panel in expeditiously resolving diagnostic odysseys, potentially alleviating years of uncertainty, anxiety, and unnecessary tests and procedures, while also addressing delays in receiving effective care.

4.8. The Utility of Multigene Panels in Uncertain DD Cases

Diagnosing DDs is a complex challenge due to their extensive phenotypic and genetic diversity (Manickam et al., 2021). In South Africa, this challenge is intensified by a range of factors, including resource limitations, healthcare access issues, and systemic constraints. The shortage of specialised genetic healthcare professionals and diagnostic test, especially in rural areas, significantly hinders the ability to deliver timely and accurate assessments (Kromberg *et al*., 2013). Geographic and socio-economic disparities further exacerbate these difficulties, as families in underserved regions face limited access to specialized care. Additionally, broader systemic issues, such as inadequate funding and infrastructure within the public healthcare system, undermine the availability and quality of diagnostic services.

Traditional diagnostic methods, specifically single gene tests, necessitate clinicians to thoroughly evaluate clinical presentations, and narrow down suspected diagnosis for targeted molecular testing. However, in cases where patients may present with a non-specific and variable clinical presentation lacking predetermined clinical suspicions, clinicians encounter a complex diagnostic landscape, rendering single gene testing of limited effectiveness. Contrastingly, multigene panels, which evaluate multiple genes associated with a particular condition or phenotype are recognised for their superiority over single gene tests and may prove valuable in cases without clinical suspected diagnosis or those with a suspected diagnosis that later establishes a different molecular diagnosis (Reid & Pal, 2020). Below, we present three representative cases from this study - two without a clinically suspected diagnosis and one with a clinically suspected diagnosis, which was later established as a different molecular diagnosis.

These cases highlight the effectiveness of a multigene approach in broadening the diagnostic scope in the diagnostic testing of rare DDs.

4.8.1. Patient with a *PACS1* Causal Variant

Causal variants in *PACS1* cause PACS1-neurodevelopmental disorder (MIM 615009), characterised by a range of complications such as mild-to-severe neurodevelopmental delays, hypotonia, feeding difficulties, dysmorphism, short stature, constipation, seizures, behavioural issues, congenital heart anomalies, renal anomalies, short stature, heart disease and microcephaly (Schuurs-Hoeijmakers *et al.,* 2012; Van Nuland *et al.,* 2021). The prevalence of PACS1-NDD is unknown, however, 87 cases had been reported in the literature by 2022 and more than 90% of these cases share the same c.607C>T;p.R203W causal variant (Arnedo *et al.,* 2022). In this study, the patient presented with severe developmental delay with absent speech, minor anomalies (tapered fingers, umbilical hernia, short fingers, single umbilical artery), short stature, behavioural problems, seizures and dysmorphic features and had an initial clinical suspicion diagnosis of Coffin-Lowry or Coffin-Siris syndromes. This study identified the recurrent PACS1-NDD causal c.607C>T; p.R203W variant in this patient. Initially, a molecular diagnosis of PACS1-NDD was not suspected, primarily due to clinician unfamiliarity with this diagnosis, stemming from limited access to molecular testing in our setting and many LMICs, compounded by relatively non-specific features that overlapped with other DDs. However, further clinical evaluation conclusively identified clinical features consistent with PACS1-NDD. The c.607C>T;p.R203W variant identified in this patient is confirmed as *de novo*, indicating that the parents carry a baseline population average recurrence risk of approximately 1%–2% (Kay *et al*., 2023).

4.8.2. Patient with an *AUTS2* Causal Variant

Causal variants in *AUTS2* are a known cause of AUTS2 syndrome (MIM 615834) (Beunders *et al.,* 2015, 2016; Hori *et al.,* 2022). The clinical features associated with AUTS2 syndrome include low birth weight, developmental delay, a small head, delayed speech, and language development, feeding difficulties, hypotonia, learning disability, hyperactivity, microcephaly, mild dysmorphic features, limb abnormalities, and skeletal system abnormalities (Beunders *et al.,* 2016). In this study, the patient presented with moderate developmental delay, scoliosis, generalised hypotonia, and feeding problems. Initial clinical suspicion for a diagnosis in this patient was absent owing to constraints in access to comprehensive genetic testing. Nevertheless, through this study, a known causal *AUTS2*:c.376C>T variant was identified,

thereby establishing a definitive diagnosis of AUTS2 syndrome in this patient. The presence of the *de novo* AUTS2 variant in this patient implies that the parents harbor a basal population average recurrence risk estimated at approximately 1%–2% (Kay *et al*., 2023).

4.8.3. Patient with *STXBP1* causal variant

Disease causing variants within the *STXBP1* gene are associated with a rare disorder known as STXBP1-related disorder (MIM 612164). This condition is characterised by early onset encephalopathy, exhibiting a spectrum of manifestations. Affected individuals may present with moderate to severe intellectual disability, hypotonia, seizures, and various movement disorders, including ataxia, dystonia, and tremors (Abramov *et al.*, 2021). The complexity of this disorder is further compounded by behavioural challenges, such as hyperactivity, self-aggressiveness, and autism, along with feeding difficulties. Notably, approximately 95% of patients with the *STXBP1* gene variant are diagnosed with epilepsy (Stamberger *et al.*, 2016). In this study, the patient, characterised by profound intellectual disability, hyperactivity, movement disorders, and seizures, posed a diagnostic challenge initially lacking clinical suspicion. The gene panel strategy proved crucial in identifying a previously documented causal *STXBP1*:c.416C>T variant, ultimately leading to a diagnosis of STXBP1-related disorder. As part of clinical management, the patient will be referred to neurology for an electroencephalogram assessment aiming to establish control over seizures and effectively manage associated behaviours. Furthermore, regular genetic and neurodevelopment follow-up clinics will be instituted to monitor developmental progress and address any emerging movement disorders. Moreover, conducting periodic family reassessments will be crucial.

4.9. Long-term impact of the DNM-Enriched Gene panel strategy in DDs

While this study's findings present immediate and patient-level benefits, the potential long-term impacts are equally significant. The research has the potential to improve patient care by enabling the development of more personalised clinical management plans tailored to individual genetic profiles. By introducing early screening for DDs through a rapid TAT gene panel approach, the strategy proposed in this study could lead to earlier diagnoses and timely interventions, which are critical for enhancing long-term health outcomes. Moreover, the cost-effective nature of the screening strategy could substantially reduce healthcare expenses by streamlining the diagnostic process into a single test, as opposed to the current practice of multiple genetic tests commonly employed in LMICs. In addition to these practical

implications, this research contributes valuable data to the growing body of literature on the genetics of DDs, particularly within understudied populations such as those in Africa. This, in turn, enhances our understanding of these disorders and supports the development of improved clinical management and genetic counselling strategies for these communities.

4.10. Clinical- and Personal Utility of an Established Molecular Diagnosis

The majority DDs with a genetic origin lack a cure or specific treatment (Reiss, 2009). However, early diagnosis and intervention can significantly improve symptoms, reduce complications, and enhance the overall quality of life for patients and their families (Aldharman *et al.*, 2023). Research, such as the work done Carmichael *et al.* (2015), emphasizes the importance of establishing a definitive genetic diagnosis for DDs, even when a cure is not possible. Such a diagnosis aids healthcare providers in making informed decisions about treatment for associated co-morbidities, pre-emptive management, and guides the need for prenatal testing, providing crucial information on relative's risks for subsequent pregnancies (Carmichael *et al.*, 2015). Moreover, a confirmed diagnosis holds significant value for parents, enabling them to better understand prognosis and actively participate in relevant support groups and advocacy groups and make informed reproductive decisions (Jeffrey *et al.*, 2021). Additionally, it puts an end to the diagnostic odyssey that families often endure, involving successive investigations and causing psychosocial and financial stress due to the uncertainty and indirect costs over time (Dragojlovic *et al.*, 2021). Moreover, elucidating the genetic aetiology of a disorder can unveil the associated biological pathway and, in certain instances, may inform personalized pharmacological and behavioural interventions (Henderson *et al.*, 2014). Therefore, it is imperative to consistently pursue optimal strategies for the implementation of genomic medicine, particularly in settings characterised by limited resources. Ensuring accessibility to genomic medicine for individuals in these specific geographies and populations is an essential component of mitigating healthcare disparities and ensuring inclusivity for all. Below we present representative cases highlighting the utility of a molecular diagnosis in devising management plans in patients with *PTEN* and *CHD7* causal variants.

4.10.1. Management of Patient with *PTEN* Causal Variant

Pathogenic variants in PTEN are known to cause a group of related genetic disorders called the PTEN hamartoma tumour syndrome (PHTS) (Innella *et al.*, 2021) This syndrome encompasses four major clinically distinct syndromes: Cowden syndrome, Bannayan-Riley-Ruvalcaba

syndrome, Proteus syndrome, and Proteus-like syndrome (Hobert & Eng, 2009). Individuals with *PTEN* causal variants face an increased risk of brain, breast, uterine, thyroid, and colon cancers, as well as Lhermitte-Duclos disease (Cummings *et al.,* 2023; Pilarski, 2019). In our case, the patient exhibited macrocephaly, café-au-lait spots, and central nervous system abnormalities, consistent with PHTS features. Several promising therapies are currently in progress for patients with causal *PTEN* variants such as the mammalian target of rapamycin (mTOR) inhibitor sirolimus and Sirolimus treatment (Marsh *et al.,* 2008; Şahin *et al.,* 2022; Schmid *et al.,* 2014). It is crucial to note that while these therapies show potential, they are undergoing trials in developed countries, and their availability in LMICs, including SA, may be constrained by cost factors. In our current setting, our primary management focus for this patient would be devising effective management strategies, and the most crucial aspect would involve vigilant monitoring for signs of malignancies due to the elevated and unsuspected cancer risk associated with *PTEN* mutations. Additionally, a neurology referral is recommended to address observed neurodevelopmental disorders. As both parents were unavailable for testing at recruitment, the management plan would also include conducting segregation analysis to establish the inheritance pattern of the identified variant. This step is crucial for understanding familial implications and guiding interventions. Family members with a familial *PTEN* pathogenic variant, indicative of PHTS, require initial evaluation and ongoing cancer surveillance. To facilitate informed decision-making, genetic cancer risk assessment and counselling will be offered to the patient and their family members.

4.10.2. Management of Patients with *CHD7* Causal Variants

Pathogenic variants in *CHD7* cause CHARGE syndrome (MIM 214800). The acronym CHARGE is based on the cardinal features identified when the syndrome was delineated: coloboma, heart malformation, choanal atresia, retardation of growth and / or development, genital anomalies, and ear anomalies (Pagon *et al.,* 1981). Over time, additional clinical findings such as dysmorphic features and hypoplasia of the semi-circular canals have been recognised, resulting in a highly variable clinical phenotype. Clinical diagnostic criteria, updated by Blake and later Verloes, emphasise major and minor clinical features (Blake & Prasad, 2006; Verloes, 2005). The major criteria include coloboma, choanal atresia, and hypoplastic semicircular canals, while the minor criteria encompass malformations of the middle and/or external ear, mental retardation, anomalies of the heart and oesophagus, hypothalamo-hypophyseal and rhombencephalic dysfunction (Verloes, 2005). Verloes's

criteria further classify CHARGE syndrome cases into three categories: typical, partial, or incomplete, and atypical diagnoses based on the presence of major and minor clinical features.

Diagnosing CHARGE syndrome exclusively relying on symptoms presents a formidable challenge due to the overlapping nature of these features with numerous other genetic disorders. Additionally, the absence of specialised magnetic resonance imaging in our setting further complicates the diagnostic process as major features canot be reliably diagnosed or excluded. As a result, genetic testing, specifically screening for *CHD7* variants, becomes an essential component of the diagnostic workup of CHARGE syndrome (Corsten-Janssen *et al.,* 2013). In this study, *CHD7* causal variants were identified in three patients: Patient D3S_0034, D3S_0050, and D3S_0051. Patient D3S_0034 presented with atypical CHARGE, while patient D3S_0050 displayed classic features. Individualised management plans for patients Patient D3S_0034 and D3S_0050 would include referrals to cardiology for heart defects and endocrine clinics for the evaluation of pituitary hormone deficiency (Blake & Prasad, 2006). The molecular confirmation of CHARGE syndrome is pivotal, as it not only guides specific medical interventions but also provides crucial insights into the overall management and care of these individuals. Additionally, X-rays of the spine and renal ultrasounds will be conducted to detect bone abnormalities, scoliosis, and potential structural issues. Continuous monitoring at the neurodevelopmental and genetic clinic would be planned to ensure comprehensive care. Regrettably, Patient D3S_0051 passed away at the age of 2 years due to heart-related complications. The patient's family will undergo counselling to discuss the risks and management of potential future pregnancies. It is noteworthy that individuals with CHARGE syndrome commonly present with de novo *CHD7* causal variants (Talkowski *et al.,* 2012), which typically leads to a diminished probability of recurrence in subsequent pregnancies.

4.11. Patients with no Causal Variants Identified from this Study

A significant portion of the patient cohort, 85.4% (n=82/96) DDs and 95% (n = 91/96) with developmental delay, did not receive a molecular diagnosis during this study. Despite the high proportion without a molecular diagnosis, the strategy holds promise in terms of cost-effectiveness and expediating diagnosis. The relatively small size of the gene panel offers a unique opportunity to pool multiple samples into a single run, thereby expediting the testing process, particularly in high-throughput settings like the Division of HumGen, where this study was conducted. However, in settings with lower patient volumes, pooling could

potentially delay diagnosis as testing might be postponed while waiting for a sufficient number of samples to be pooled. Therefore, laboratories would need to carefully evaluate their patient throughput to optimise efficiency. This pooling of samples not only streamlines laboratory processes but also reduces the per-patient testing cost. By adopting a cost-effective first-tier strategy for local screening, the potential financial impact is noteworthy. The ability to analyse multiple samples simultaneously not only optimises resources but also positions the DNM-enriched gene panel as an economically viable option for widespread screening initiatives. This approach may significantly contribute to reducing overall diagnostic costs, addressing the economic considerations associated with comprehensive genetic testing in large patient cohorts.

To improve the diagnostic yield for both cohorts, additional investigations, such as analysis of the entire exome, CNV analysis and further clinical characterisation could be carried out. The DNM-enriched gene panel strategy employed in both cohorts investigated in this study was a WES-virtual gene panel strategy. Thus, this allows for further analysis of the entire WES data set if indicated. WES analysis has demonstrated a diagnostic rate superior to that of targeted gene panels, with studies in cohorts with DDs and developmental delay reporting rates of at least 40% (Wright *et al.*, 2018). Nevertheless, it is crucial to highlight that this study was conducted within a research setting, where the resources and feasibility for conducting WES were available. The cost of WES and other complexities such as bioinformatics expertise may not be readily available in routine diagnostic settings of many LMICs, including the South African State Healthcare system.

Additionally, this group of patients could also benefit from re-analysis of their sequencing data. New and novel insights into the causal genetic factors behind various disorders, including DDs and developmental delay, are consistently surfacing (Gialluisi *et al.*, 2021; Hamanaka *et al.*, 2022; Nott & Holtman, 2023). Reanalysis of NGS data at periodic intervals has been shown to improve the diagnostic yield (Schobers *et al.*, 2022). There is no universally agreed-upon time for how often NGS data should be reanalysed. Nonetheless, numerous studies have suggested specific periods for this purpose. The ACMG recommends reassessing genomic data when there are significant updates in the knowledge base or changes in clinical guidelines (Richards *et al.*, 2015). A review and meta-analysis study looking at the re-analysis of NGS data in unsolved cases with suspected Mendelian disorders reported an overall yield of 10% for reanalysis. They report that the majority of the new diagnoses are attributable to updates in the

literature and annotated disease databases cataloguing disease-causing variants and genes. They proposed that NGS data reanalysis should be delayed to ≥24 months unless there is an urgent clinical need to reanalyse earlier (Dai *et al.,* 2022).

Wright and colleagues reported that subsequent to the initial assessment, an additional 16% of patients (n = 182/1133) attained a diagnostic outcome through systematic reanalysis conducted three years later. Won and colleagues reanalysed targeted gene panel data in a cohort with DDs and reported a reanalysis diagnostic yield of 5.2%. Based on these findings, they suggested that the reanalysis of NGS data should be implemented as a routine practice in clinical laboratories (Won *et al.,* 2020). While reanalysis of data obtained from capture-based gene panels designed for known disease genes will not lead to new diagnoses through identifying novel genes, it remains beneficial as it can improve the interpretation of genetic variants (Sun *et al.,* 2019). Moreover, this study employed a WES-virtual gene panel analysis in both cohorts. Thus, causal variant – negative patients in this study could also benefit from the analysis of additional diagnostic genes associated with DDs without incurring additional sequencing costs.

Furthermore, it is well-established that CNVs play a substantial role in DDs (Park *et al.*, 2019). This study did not evaluate CNVs. The integration of CNV analysis from the original NGS data, as highlighted in the findings by Louw and colleagues, presents a promising avenue for enhancing the diagnostic rate (Louw *et al.,* 2023). Notably, advancements in bioinformatics have made CNV analysis feasible even from primary data obtained through targeted gene panels, providing an additional opportunity for a relatively comprehensive investigation Singh *et al.*, 2021; Yao *et al.*, 2019). Moreover, there was no clinical pre-screening gatekeeping strategy applied to the HumGen referral cohort B. The intentional decision to forgo gatekeeping in the HumGen referral cohort B was designed to observe outcomes in an unphenotyped setting aimed at observing outcomes in an unphenotyped setting, mimicking scenarios without readily available medical genetics expertise for pre-screening observing outcomes in an unphenotyped setting, mimicking scenarios without readily available medical genetics expertise for pre-screening (Hosen *et al.,* 2021; Zhong *et al.,* 2021). The absence of gatekeeping in this cohort mirrors our deliberate decision to understand the implications and challenges in environments where resources are constrained. This deliberate choice provides insights into unphenotyped cohorts, offering valuable information for addressing scenarios where gatekeeping practices may be impractical due to resource constraints, a common challenge in LMICs.

In such settings, the integration of phenotyping tools, such as Face2Gene (www.face2gene.com) and GestaltMatcher (Hsieh et al., 2022), could offer an additional layer of phenotypic assessment. These AI-driven applications analyse facial phenotypes and suggest potential genetic syndromes based on photographic input. While traditional genetic expertise may be limited in resource-poor settings, these tools can serve as valuable adjuncts to clinical decision-making. They could help bridge the gap in diagnostic capabilities where genetic specialists are scarce. However, it is important to recognise the limitations of relying solely on these digital tools, as their effectiveness may vary across diverse populations due to limitations in training data. Despite these challenges, these AI-driven tools offer a promising avenue to supplement local clinical expertise, particularly in settings where medical genetics expertise is limited.

Gatekeeping, involving the assessment of clinical presentations to prioritise specific tests, is crucial for optimising limited testing resources. The lack of adequately trained genetic personnel in many LMICs underscores the significance of this approach (Chou *et al.,* 2020: Kromberg *et al.,* 2013b; Kromberg *et al.,* 2013; Nembaware & Mulder, 2019; Quinonez *et al.,* 2021). A more comprehensive phenotyping of the HumGen referral cohort B could have identified suitable patients for screening using the DNM-enriched gene panel, potentially improving the diagnostic yield. Focusing on cases with higher pre-test probabilities through gatekeeping strategies can enhance the diagnostic effectiveness of testing. However, it is noteworthy that the HumGen cohort B was specifically referred for FXS due to developmental delay as the sole clinical presentation. In contrast to the typical global diagnostic rate of >5% (Weinstein *et al.,* 2017), this study was able to identify the genetic aetiology of developmental delay in 5.2% of this patient group. This diagnostic yield of 5.2% remains significant, signifying progress in understanding genetic factors underlying developmental delay. Additionally, further improvement in diagnostic yield may be achievable through patient reassessment and the inclusion of additional phenotypic data.

4.12. Capture-Based Gene Panel versus WES-Virtual gene panel

When guiding the selection between a capture-based gene panel and a WES-virtual gene panel, clinical laboratories should conduct their resource-effectiveness analyses to determine the most practical and efficient approach in their specific settings. Capture-based panels offer targeted sequencing, allowing for the specific capture of predetermined genes associated with a particular phenotype/condition. This approach is relatively cost-effective (depending on the number of genes included), generates less data and ensuring quicker turnaround times for

results - a critical consideration in clinical settings (Jennings *et al.,* 2017). However, its limitation lies in the potential to miss causative variants outside the targeted regions, and the need for regular updates to include new genetic knowledge can be challenging (Reid *et al.,* 2016; Vinkšel *et al.,* 2021). Notably, capture-based panels can be executed with minimal NGS infrastructure (small benchtop NGS platforms and less data storage avenues).

In contrast, WES-virtual gene panels take a comprehensive approach by sequencing the entire exome and focusing on specific genes associated with a phenotype, allowing adaptability to emerging genetic knowledge without additional sequencing costs (Wang *et al.,* 2019) This tiered approach involves a virtual panel first, followed by expanded analysis if required. Nevertheless, a notable advantage of a WES-virtual panel is their provision of complete exome data, enabling additional diagnostic or research analyses. This comprehensive data could mitigate or eliminate costs associated with conducting additional genetic tests for patients who yield negative results on the initial panel screening.

In this study, estimations of reagent costs linked to a capture-based DNM-enriched gene panel and a WES-virtual gene panel strategy, were obtained from Thermo Fisher Scientific. The quotation for reagents for the capture-based targeting sequencing of 96 samples was about R1,300,535 and for WES for 96 patients about R494,456. Notably, the reagents for the capture-based gene panel in this study were higher in cost than those for WES. In addition to reagent costs, additional cost considerations must be considered. These encompass expenditures associated with data analysis, data storage, and specific NGS platform requisites, particularly pertinent to WES. The supplementary costs linked to WES may be higher due to its propensity for generating more extensive data, necessitating advanced data storage solutions, and requiring the expertise of trained bioinformatics professionals for data analysis and medical geneticists for identifying appropriate patient to test and reviewing the gene-phenotype correlations. Consequently, when deciding between these two approaches, laboratories would need to carefully assess feasibility within their settings to identify the most suitable compromise.

4.13. Special Considerations for DD Genetic Testing Implementation in LMICs

Based on the insights from this study, we propose the following considerations for the implementation of routine genetic testing for DDs in LMICs:

4.13.1. Comprehensive Clinical Data Submission:

Primary healthcare providers referring patients for genetic testing should actively provide comprehensive phenotypic and family history information to the testing laboratory. These data are critical for improving variant interpretation and establishing meaningful associations between genes and patient phenotypes, thereby avoiding inappropriate testing (Foreman *et al.,* 2023; Johnson *et al.,* 2022; Kipkemoi *et al.,* 2023). However, barriers such as a lack of genetic knowledge among non-geneticist primary healthcare providers may hinder this effort (Seibel *et al.,* 2022; Suther & Goodson, 2003).

Global studies highlight widespread deficiencies in genetic knowledge among non-geneticist healthcare providers, who play a pivotal role in ordering tests, interpreting results, referring patients to genetic centers, and discussing genetic issues with patients (Baars *et al.,* 2005; Cornel, 2019; Pasquier *et al.,* 2022). Addressing these knowledge gaps through education is vital for integrating genetics into mainstream healthcare practices. Recognising and bridging these gaps is essential for ensuring effective utilisation and understanding of genetic testing within the broader context of clinical practice, particularly significant in the era of precision medicine, where clinical genomics significantly shapes tailored treatments for individual patients. Thus, it becomes paramount to equip primary healthcare providers with adequate training on the appropriate selection of genetic tests, interpretation of data, and feedbacking findings.

Efforts to bridge these knowledge gaps are underway, even in LMIC settings like the African continent. The African genomic medicine training initiative is a noteworthy example, aiming to implement a genomic medicine training program primarily designed for non-geneticist healthcare professionals in Africa (Nembaware & Mulder, 2019). Furthermore, adequate healthcare training can enhance diagnostic yield through clinical pre-screening gatekeeping. By selectively testing suitable patients, it optimises the use of scarce testing resources. We propose that clinical diagnostic laboratories should develop effective gatekeeping strategies to ensure the screening of suitable individuals who would likely benefit from an expensive diagnostic test such as an NGS-based testing strategy.

4.13.2. Trio-Sequencing and/or Analysis:

Trio sequencing or analysis may not be universally possible or necessary, particularly when a molecular diagnosis can be attained through publicly available data without confirming

inheritance from parents. This is especially applicable when the identified variant is extensively documented in the literature, possesses functional evidence of pathogenicity, and is well-documented in disease databases. Thus, adopting a proband-only first-tier NGS strategy could proficiently identify the causal variant in most patients. The testing of additional family members can be selectively performed using Sanger sequencing in cases where a putative disease-causing variant is consistent with the phenotype of the patient and applying the ACMG/AMP criteria codes PS2 or PM6, would upgrade the classification of the variant to a likely pathogenic or a pathogenic class. This strategic approach may present a cost-effective alternative, especially in resource-constrained settings catering to large populations.

4.13.3. Fixed Targeted Gene Panel versus WES-Virtual gene panel

Diagnostic laboratories should systematically evaluate cost-effective strategies tailored to their specific settings. If routine WES is feasible, prioritising WES with a range of virtual panels is proposed as a first-tier strategy for patients with DDs and/or developmental delays. The virtual gene panel approach offers a notable advantage, allowing for the expansion of the gene panel or a transition to full exome analysis without incurring additional expenses. This contrasts with the financial implications associated with capture-based targeted gene panels. In cases where routine WES is not viable, it is recommended to implement a well-curated panel consisting of 200 - 500 genes. The selection of DNM-enriched genes is considered an appropriate starting point for developing a locally relevant gene panel.

4.13.4. Copy Number Variants Evaluation

CNVs contribute significantly to the genetic aetiology of DDs (Park *et al.*, 2019). However, this study did not include CNV analysis in its scope. We propose that diagnostic laboratories should establish appropriate bioinformatics pipelines, to allow for the concurrent evaluation of SNVs and CNVs in cohorts with DDs and/or developmental delay. Prior research has consistently demonstrated the utility of combined CNV and SNV analysis in improving diagnostic rates for DDs (Shaikh, 2017; Uddin *et al.*, 2016; Wayhelova *et al.*, 2024).

4.14. Study Limitations

Several limitations are acknowledged in this study. Firstly, an overrepresentation of male individuals in the HumGen referral cohort B was observed, comprising 94.8% (n= 91/96) of the cases, potentially introducing a gender bias. This skewness can be attributed to the specific nature of this cohort, as individuals were referred to the Division for FXS genetic testing. It is well-established that FXS is more prevalent in males than females (Hagerman, 2008; Hunter

et al., 2014), thereby explaining the higher proportion of males in the cohort. Previous studies on DDs and neurodevelopmental disorders have reported high numbers of male individuals, attributing this skewness to the male-biased contribution of X-linked (Bölte *et al.,* 2023; Polyak *et al.,* 2015). However, in this study only two patients (n=2/19 = 10.5%) of patients with causal variants have causal variants in X-linked genes. Another specific limitation within HumGen referral cohort B is the limited availability of clinical information, posing challenges in establishing gene-phenotype correlations during the variant review process. An additional limitation arises from the relatively smaller sample size in both cohorts (n=96 per cohort), compared to previous studies, impacting the generalisability of findings.

Moreover, the limited availability of African population frequency databases represents another limitation commonly encountered in genetic research focused on African populations (Bentley *et al.,* 2020). The underrepresentation of African populations in global databases poses challenges in interpreting NGS data and establishing diagnoses for African patients. This gap increases the risk of false positives, as variants may be misclassified as pathogenic due to their novelty or rarity (Lumaka *et al.,* 2022). Despite using the gnomAD and the 1000 Genomes data, both of which offer limited African genetic information, the lack of dedicated databases for African populations hinders a comprehensive analysis. To illustrate, the ACMG-AMP PM2 code, indicating rarity or absence from controls or an extremely low frequency if recessive, relies on data from the Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium has different strength/weights: supporting, moderate, strong, and very strong. In our data interpretation of rare variants, we applied the PM2 with a supporting strength instead of a moderate strength, recognising that these widely used databases lack a comprehensive representation of African populations. This practical example underscores the critical need for databases specifically designed to capture the genetic diversity present in African populations (Pfennig *et al.,* 2023). Access to such tailored databases would significantly enhance our ability to understand the prevalence and contribution of identified variants within this genetically diverse population.

5. CONCLUSION

Developmental disorders are debilitating conditions marked by non-specific manifestations, genetic diversity, and often accompanied by comorbidities (Boyle *et al.,* 2011). Establishing a molecular diagnosis in DDs is a challenge, given the considerable clinical and genetic heterogeneity associated with these conditions (Wright *et al.,* 2015). This challenge often leads to a substantial number of DD cases remaining undiagnosed, initiating an extended diagnostic journey that becomes a considerable health burden. This burden could be heightened in sub-Saharan Africa, where inadequate infrastructure, limited characterisation of DDs, and overall resource constraints hamper effective evaluation (Kamp *et al.,* 2021; Lumaka *et al.,* 2022). This hypothesis is substantiated by insights from the DDD-UK study, indicating a notably low diagnostic rate among individuals of African ancestry. This is attributed in part to the absence of ancestry-matched controls for accurate allele frequency estimation and, additionally, a lower likelihood of being recruited as a trio compared to individuals from other populations. (Wright *et al.,* 2023).

Prior research indicates that timely identification of medical conditions contributes to enhanced patient care and favourable outcomes (Aldharman *et al.,* 2023; (Carmichael *et al.,* 2015). This is achieved through the facilitation of adjustments in medical treatment, early implementation of targeted therapies, heightened surveillance, appropriate initiation of comfort care, diminished reliance on costly, repetitive, and occasionally invasive investigations, and the initiation of cascade testing for family members. and occasionally invasive investigations, and the initiation of cascade testing for family members. Additionally, it supports counselling for reproductive planning (Rabbani *et al.,* 2014; Wright *et al.,* 2018). Current international guidelines advocate for WES and/or WGS as the first-tier diagnostic tests for DDs, given their high diagnostic yield and capability to explore various genetic variations (Manickam *et al.,* 2021; Srivastava *et al.,* 2019). However, the routine implementation of WES/WGS encounters significant barriers in LMIC settings, such as the South African State Healthcare system. Despite the global focus on precision medicine, individuals in many LMICs, including SA, experience a further widening of existing healthcare disparities (Wright *et al.,* 2023). This disparity is exacerbated by the disproportionate allocation of resources, with LMICs struggling to keep pace with genomics and precision medicine advancements compared to wealthier counterparts. As the global health community focuses on tailored medical approaches, individuals in LMICs find themselves on the margins, unable to benefit from these advancements, perpetuating the existing healthcare divide (Owolabi *et al.,* 2023).

Given the substantial genetic diversity in African populations and unique local challenges, implementing precision medicine becomes important to tackle region-specific health challenges (Hussein *et al.*, 2022). The question now becomes: How do we ensure that populations in LMICs, benefit from precision medicine without being left behind in resource-scarce environments? The solution may lie in developing innovative and resource-efficient genetic diagnostic strategies tailored to each unique setting. One such strategy could be the use of targeted gene panels. Employing targeted sequencing may hold promise as an efficient strategy applicable in clinical contexts, contingent upon the utilisation of carefully curated gene panels. This approach could present a focused testing paradigm, limited to predetermined gene sets, thereby facilitating accurate variant identification with potentially increased sensitivity, and potentially decreased overall resource requirements relative to WES and WGS. Integrating the proposed strategy into the current diagnostic workflow would involve several key steps. First, collaborating with healthcare providers to ensure the panel fits seamlessly into clinical practice and develop training programs to educate them on its use. Adapting laboratory processes and bioinformatics pipelines to incorporate the gene panel, and conduct pilot studies to evaluate its real-world effectiveness. Finally, establishing quality assurance measures to monitor and refine the panel's performance continually.

The adoption of NGS by diagnostic laboratories necessitates a focus on resource-efficient strategies in settings where limited resources serve a large population. Our experience serves as a starting point to develop practical resource-efficient diagnostic strategies for DDs, ensuring clinically relevant results within a suitable time. In this study we demonstrated that the use of gene intolerance metrics and gene-disease validity curation platforms can be used as an effective strategy to define a smaller set of DD genes to screen in resource-constrained settings. Additionally, we identified key factors influencing the panel's yield. Addressing these factors is imperative for routine diagnostic implementation. While this study made use of the Ion Torrent and Illumina platforms, the diagnostic strategy proposed in this thesis may be adaptable to any other NGS platform This adaptability provides clinical diagnostic laboratories with the flexibility to utilise existing local or regional NGS infrastructure without the need for additional capital investment, thereby reducing barriers to the implementation of genomic medicine.

6. REFERENCES

Abacan, M. A., Alsubaie, L., Barlow-Stewart, K., Caanen, B., Cordier, C., Courtney, E., *et al*. (2019). The Global State of the Genetic Counseling Profession. *European Journal of Human Genetics*, 27(2), pp. 183–197.

Abe-Hatano, C., Iida, A., Kosugi, S., Momozawa, Y., Terao, C., Ishikawa, K., *et al*. (2021). Whole genome sequencing of 45 Japanese patients with intellectual disability. *American Journal of Medical Genetics, 185*(5), pp. 1468–1480.

Abramov, D., Guiberson, N. G. L., & Burré, J. (2021). STXBP1 encephalopathies: Clinical spectrum, disease mechanisms, and therapeutic strategies. *Journal of Neurochemistry*, *157*(2), pp. 165–178.

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, Chapter 7, Unit7.20.

Agarwal, I., Fuller, Z. L., Myers, S. R., & Przeworski, M. (2023). Relating pathogenic loss-of-function mutations in humans to their evolutionary fitness costs. *E Life Sciences*, 12, pp. e83172-e83180.

Aldharman, S. S., Al-jabr, K. H., Alharbi, Y. S., Alnajar, N. K., Alkhanani, J. J., Alghamdi, A, *et al*. (2023). Implications of Early Diagnosis and Intervention in the Management of Neurodevelopmental Delay (NDD) in Children: A Systematic Review and Meta-Analysis. *Cureus Journal of Medical Science*, 15(5), pp. e38745-e38753.

Alharbi, W. S., & Rashid, M. (2022). A review of deep learning applications in human genomics using next-generation sequencing data. *Human Genomics*, 16(1), pp. 26-335.

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

Amarillo, I. E., Li, W. L., Li, X., Vilain, E., & Kantarci, S. (2014). De novo single exon deletion of AUTS2 in a patient with speech and language disorder: A review of disrupted AUTS2 and further evidence for its role in neurodevelopmental disorders. *American Journal of Medical Genetics*, 164(4), pp. 958–965.

Andreassen, O. A., Hindley, G. F. L., Frei, O., & Smeland, O. B. (2023). New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. *World Psychiatry*, 22(1), pp. 4–24.

Arnedo, M., Ascaso, Á., Latorre-Pellicer, A., Lucia-Campos, C., Gil-Salvador, M., Ayerza-Casas, A., *et al*. (2022). Molecular Basis of the Schuurs–Hoeijmakers Syndrome: What We Know about the Gene and the PACS-1 Protein and Novel Therapeutic Approaches. In *International Journal of Molecular Sciences*, 23(17), pp. 9649-9660.

Arteche-López, A., Ávila-Fernández, A., Romero, R., Riveiro-Álvarez, R., López-Martínez, M. A., Giménez-Pardo, *et al*. (2021). Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Scientific Reports*, 11(1), pp. 5697-5708.

Aspromonte, M. C., Bellini, M., Gasparini, A., Carraro, M., Bettella, E., Polli, R., *et al*. (2019). Characterization of intellectual disability and autism comorbidity through gene panel sequencing. *Human Mutation*, *40*(9), pp. 1346–1363.

Austin-Tse, C. A., Jobanputra, V., Perry, D. L., Bick, D., Taft, R. J., Venner, *et al*. (2022). Best practices for the interpretation and reporting of clinical whole genome sequencing. *NPJ Genomic Medicine*, 7(1), pp. 27-41.

Baars, M. J. H., Henneman, L., & Ten Kate, L. P. (2005). Deficiency of knowledge of genetics and genetic tests among general practitioners, gynecologists, and pediatricians: A global problem. *Genetics in Medicine*, *7*(9), pp. 605–610.

Baine-Savanhu, F., Macaulay, S., Louw, N., Bollweg, A., Flynn, K., Molatoli, M., *et al*. (2023). Identifying the genetic causes of developmental disorders and intellectual disability in Africa: a systematic literature review. *Frontiers in Genetics*, 14, pp. 7922-7940.

Bartels, C. F., Scacheri, C., White, L., Scacheri, P. C., & Bale, S. (2010). Mutations in the CHD7 gene: The experience of a commercial laboratory. *Genetic Testing and Molecular Biomarkers*, *14*(6), pp. 881–891.

Bartha, Á., & Győrffy, B. (2019). Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. *Cancers*, *11*(11), pp. 1–20.

Bean, L. J. H., Funke, B., Carlston, C. M., Gannon, J. L., Kantarci, S., Krock, B. L., *et al*. (2020). Diagnostic gene sequencing panels: from design to report—a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 22(3), pp. 453–461.

Bear, M. F., Dölen, G., Osterweil, E., & Nagarajan, N. (2008). Fragile X: Translation in action. *Neuropsychopharmacology*, 33(1), pp. 84–87.

Beck, D. B., Basar, M. A., Asmar, A. J., Thompson, J. J., Oda, H., Uehara, D. T., *et al*. (2021). Linkage-specific deubiquitylation by OTUD5 defines an embryonic pathway intolerant to genomic variation. *Science Advances*, 7(4), pp. 2116-2124.

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

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, 98(6), pp. 236–238.

Bejjani, B. A., & Shaffer, L. G. (2006). Application of array-based comparative genomic hybridization to clinical diagnostics. *Journal of Molecular Diagnostics*, 8(5), pp. 528–533.

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

Bentley, A. R., Callier, S. L., & Rotimi, C. N. (2020). Evaluating the promise of inclusion of African ancestry populations in genomics. *Genomic Medicine*, 5(1), pp. 550-562.

Berisha, S. Z., Shetty, S., Prior, T. W., & Mitchell, A. L. (2020). Cytogenetic and molecular diagnostic testing associated with prenatal and postnatal birth defects. *Birth Defects Research*, 112(4), pp. 293–306.

Bernardini, L., Alesi, V., Loddo, S., Novelli, A., Bottillo, I., Battaglia, A., *et al*. (2010). High-resolution SNP arrays in mental retardation diagnostics: How much do we gain. *European Journal of Human Genetics*, *18*(2), pp. 178–185.

Betancur, C., & Buxbaum, J. D. (2020). Gene constraint and genotype–phenotype correlations in neurodevelopmental disorders. *Current Opinion in Genetics and Development*, 3(65), pp. 69–75.

Beunders, G., De Munnik, S. A., Van Der Aa, N., Ceulemans, B., Voorhoeve, E., Groffen, A. J., *et al*. (2015). Two male adults with pathogenic AUTS2 variants, including a two-base pair deletion, further delineate the AUTS2 syndrome. *European Journal of Human Genetics*, 23(6), pp. 803–807.

Beunders, G., van de Kamp, J., Vasudevan, P., Morton, J., Smets, K., Kleefstra, T., *et al*. (2016). A detailed clinical analysis of 13 patients with AUTS2 syndrome further delineates the phenotypic spectrum and underscores the behavioural phenotype. *Journal of Medical Genetics*, *53*(8), pp. 523–532.

Biswas, D., Cary, W., & Nolta, J. A. (2020). PPP2R5D-related intellectual disability and neurodevelopmental delay: A review of the current understanding of the genetics and biochemical basis of the disorder. *International Journal of Molecular Sciences*, 21(4), pp. 1286.

Blake, K. D., & Prasad, C. (2006). CHARGE syndrome. *Orphanet Journal of Rare Diseases*, 1(34), pp. 1150-1172.

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), pp. 66–75.

Bölte, S., Neufeld, J., Marschik, P. B., Williams, Z. J., Gallagher, L., & Lai, M. C. (2023). Sex and gender in neurodevelopmental conditions. *Nature Reviews Neurology*, 19(3), pp. 136–159.

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*, *22*(6), pp. 1036–1039.

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

Brunelli, L., Jenkins, S. M., Gudgeon, J. M., Bleyl, S. B., Miller, C. E., Tvrdik, T., *et al*. (2019). Targeted gene panel sequencing for the rapid diagnosis of acutely ill infants. *Molecular Genetics and Genomic Medicine*, *7*(7), pp. 1–10.

Bruno, D. L., White, S. M., Ganesamoorthy, D., Burgess, T., Butler, K., Corrie, S., *et al*. (2011). Pathogenic aberrations revealed exclusively by single nucleotide polymorphism (SNP) genotyping data in 5000 samples tested by molecular karyotyping. *Journal of Medical Genetics*, *48*(12), pp. 831–839.

Bruno, L. P., Doddato, G., Baldassarri, M., Rizzo, C. L., Resciniti, S., Bruttini, M., *et al*. (2023). Expanding the clinical spectrum associated with the PACS1 p.Arg203Trp

mutational hot-spot: Two additional Italian patients. *American Journal of Medical Genetics*, *191*(1), pp. 284–288.

Cabello-Aguilar, S., Vendrell, J. A., & Solassol, J. (2023). A Bioinformatics Toolkit for Next-Generation Sequencing in Clinical Oncology. *Current Issues in Molecular Biology*, 45(12), pp. 9737–9752.

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(1), pp. 403–433.

Cao, X., Wolf, A., Kim, S. E., Cabrera, R. M., Wlodarczyk, B. J., Zhu, H., *et al*. (2021). CIC de novo loss of function variants contribute to cerebral folate deficiency by downregulating FOLR1 expression. *Journal of Medical Genetics*, *58*(7), pp. 484–494.

Capkova, P., Srovnal, J., Capkova, Z., Staffova, K., Becvarova, V., Trkova, M., *et al*. (2019). MLPA is a practical and complementary alternative to CMA for diagnostic testing in patients with autism spectrum disorders and identifying new candidate CNVs associated with autism. *PeerJ*, *2019*(1), pp. 456-483.

Cardoso, A. R., Lopes-Marques, M., Silva, R. M., Serrano, C., Amorim, A., *et al*. (2019). Essential genetic findings in neurodevelopmental disorders. *Human Genomics*, 13(1), pp. 31-38.

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y, *et al*. (2019). Gene expression across mammalian organ development. *Nature*, *571*(7766), pp. 505–509.

Carmichael, N., Tsipis, J., Windmueller, G., Mandel, L., & Estrella, E. (2015). "Is it Going to Hurt?": The Impact of the Diagnostic Odyssey on Children and Their Families. *Journal of Genetic Counseling*, *24*(2), pp. 325–335.

Carvill, G. L., Weckhuysen, S., McMahon, J. M., Hartmann Rikke Møller, C. S., Hjalgrim, H., Cook, J., *et al*. (2014). GABRA1 and STXBP1: Novel genetic causes of Dravet syndrome. *Neurology*, *82*(14), pp. 1245–1253.

Caudai, C., Galizia, A., Geraci, F., Le Pera, L., Morea, V., Salerno, E., *et al*. (2021). AI applications in functional genomics. *Computational and Structural Biotechnology Journal*, 19(1), pp. 5762–5790.

Cheng, C., Fei, Z., & Xiao, P. (2023). Methods to improve the accuracy of next-generation sequencing. *Frontiers in Bioengineering and Biotechnology*, 11(1), pp. 2111-2129.

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

Chopra, M., McEntagart, M., Clayton-Smith, J., Platzer, K., Shukla, A., Girisha, K. M., *et al*. (2021). Heterozygous ANKRD17 loss-of-function variants cause a syndrome with intellectual disability, speech delay, and dysmorphism. *American Journal of Human Genetics*, *108*(6), pp. 1138–1150.

Chou, A. F., Duncan, A. R., Hallford, G., Kelley, D. M., & Dean, L. W. (2021). Barriers and strategies to integrate medical genetics and primary care in underserved populations: a scoping review. *Journal of Community Genetics*, *12*(3), pp. 291–309.

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), pp. 179–186.

Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., *et al*. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature Genetics*, *51*(1), pp. 106–116.

Collins, R. L., Brand, H., Redin, C. E., Hanscom, C., Antolik, C., Stone, M. R., *et al*. (2017). Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biology*, *18*(1), pp. 36-52.

Cooper, G. M., Stone, E. A., Asimenos, G., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15*(7), pp. 901–913.

Cornel, M. C. (2019). Evidence-based genetic education of non-genetic-expert physicians: Experiences over three decades in Amsterdam. *Frontiers in Genetics, 10(2)*, pp. 712-723.

Corsten-Janssen, N., Kerstjens-Frederikse, W. S., Du Marchie Sarvaas, G. J., Baardman, M. E., Bakker, M. K., Bergman, J. E. H., *et al*. (2013). The cardiac phenotype in patients with a CHD7 mutation. *Circulation: Cardiovascular Genetics*, *6*(3), pp. 248–254.

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*, *13*(3), pp. 355–363.

Couper, J. (2002). Prevalence of childhood disability in rural KwaZulu-Natal. *South African Medical Journal*, *92*(7), pp. 549–552.

Coutelier, M., Holtgrewe, M., Jäger, M., Flöttman, R., Mensah, M. A., Spielmann, M., *et al*. (2022). Combining callers improves the detection of copy number variants from whole-genome sequencing. *European Journal of Human Genetics*, *30*(2), pp. 178–186.

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(3)*, 89-110.

Cummings, S., Alfonso, A., Hughes, E., Kucera, M., Mabey, B., Singh, N., & Eng, C. (2023). Cancer Risk Associated With *PTEN* Pathogenic Variants Identified Using Multigene Hereditary Cancer Panel Testing. *Journal of Clinical Oncology: Precision Oncology*, *7(3)*, pp. 567-580.

da Cunha Leite, A. J., Pinto, I. P., Leijsten, N., Ruiterkamp-Versteeg, M., Pfundt, R., de Leeuw, N., *et al*. (2022). Diagnostic yield of patients with undiagnosed intellectual disability, global developmental delay and multiples congenital anomalies using karyotype, microarray analysis, whole exome sequencing from Central Brazil. *The Public Library of Science ONE*, *17*(4), pp. 13-21.

Dai, P., Honda, A., Ewans, L., McGaughran, J., Burnett, L., Law, M., *et al*. (2022). Recommendations for next generation sequencing data reanalysis of unsolved cases with suspected Mendelian disorders: A systematic review and meta-analysis. *Genetics in Medicine 24*(8), pp. 1618–1629.

de Ligt, J., Willemsen, M. H., van Bon, B. W. M., Kleefstra, T., Yntema, H. G., Kroes, T., *et al* (2012). Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *New England Journal of Medicine*, *367*(20), pp. 1921–1929.

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., *et al*. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), pp. 209–215.

Dillon, O. J., Lunke, S., Stark, Z., Yeung, A., Thorne, N., Gaff, C., White, S. M., & Tan, T. Y. (2018). Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *European Journal of Human Genetics*, *26*(5), pp. 644–651.

Dragojlovic, N., Van Karnebeek, C. D. M., Ghani, A., Genereaux, D., Kim, E., Birch, P., *et al*. (2020). The cost trajectory of the diagnostic care pathway for children with suspected genetic disorders. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 22(2), pp. 292–300.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., *et al*. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*(1), pp. 231-253.

Engels, H., Wohlleber, E., Zink, A., Hoyer, J., Ludwig, K. U., Brockschmidt, F. F., *et al*. (2009). A novel microdeletion syndrome involving 5q14.3-q15: Clinical and molecular cytogenetic characterization of three patients. *European Journal of Human Genetics*, *17*(12), pp. 1592–1599.

Essop, F. B., & Krause, A. (2013). Diagnostic, carrier and prenatal genetic testing for fragile X syndrome and other FMR-1-related disorders in Johannesburg, South Africa: A 20-year review. *South African Medical Journal*, *103*(1), pp. 994–998.

Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., *et al*. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: Results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine*, *17*(7), pp. 578–586.

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, *7*(2), pp. 85–97.

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

Foreman, J., Brent, S., Perrett, D., Bevan, A. P., Hunt, S. E., Cunningham, F., *et al*. (2022). DECIPHER: Supporting the interpretation and sharing of rare disease phenotype-linked variant data to advance diagnosis and research. *Human Mutation*, *43*(6), pp. 682–697

Foreman, J., Perrett, D., Mazaika, E., Hunt, S. E., Ware, J. S., & Firth, H. V. (2023). Downloaded from www.annualreviews.org Access provided by 196. *Annual Review of Genomics and Human Genetics. 2023*, *24(2)*, pp. 151–176.

Forrest, M. P., Hill, M. J., Kavanagh, D. H., Tansey, K. E., Waite, A. J., & Blake, D. J. (2018). The psychiatric risk gene transcription factor 4 (TCF4) regulates neurodevelopmental pathways associated with schizophrenia, Autism, and intellectual disability. *Schizophrenia Bulletin*, *44*(5), pp. 1100–1110.

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., *et al*. (2006). Copy number variation: New insights in genome diversity. *Genome Research, 16*(8), pp. 949–961.

Gao, C., Wang, X., Mei, S., Li, D., Duan, J., Zhang, P., *et al*. (2019). Diagnostic yields of trio-wes accompanied by CnVSeq for rare neurodevelopmental disorders. *Frontiers in Genetics, 10*(5), pp. 461-482.

Garber, K. B., Visootsak, J., & Warren, S. T. (2008). Fragile X syndrome. *European Journal of Human Genetics*, *16*(6), pp. 666–672.

George, S. H. L., Medina-Rivera, A., Idaghdour, Y., Lappalainen, T., & Gallego Romero, I. (2023). Increasing diversity of functional genetics studies to advance biological discovery and human health. *American Journal of Human Genetics*, *110*(12), pp.1996–2002.

Gialluisi, A., Andlauer, T. F. M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., *et al*. (2021). Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. *Molecular Psychiatry*, *26*(7), pp. 3004–3017.

Giles, H. H., Hegde, M. R., Lyon, E., Stanley, C. M., Kerr, I. D., Garlapow, *et al*. (2021). The Science and Art of Clinical Genetic Variant Classification and Its Impact on Test Accuracy. *Annual Review of Genomics and Human Genetics*, *22(3)*, pp. 285–307.

Gillentine, M. A., Wang, T., Hoekzema, K., Rosenfeld, J., Liu, P., Guo, H., *et al*. (2021). Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders. *Genome Medicine, 13*(1). pp. 740-752.

Goh, G., & Choi, M. (2012). Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics & Informatics*, *10*(4), pp. 214-226.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet*, *369*(9555), pp. 60–70.

Griffin, A., Mahesh, A., & Tiwari, V. K. (2022). Disruption of the gene regulatory programme in neurodevelopmental disorders. *Biochimica et Biophysica Acta Gene Regulatory Mechanisms, 1865*(7), pp. 194860 -184885.

Grozeva, D., Carss, K., Spasic-Boskovic, O., Tejada, M. I., Gecz, J., Shaw, M., *et al*. (2015). Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human Mutation*, *36*(12), pp. 1197–1204.

Gudmundsson, S., Singer-Berk, M., Watts, N. A., Phu, W., Goodrich, J. K., Solomonson, M., *et al*. (2022). Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*, *43*(8), pp. 1012–1030.

Guo, H., Hou, L., Shi, Y., Jin, S. C., Zeng, X., Li, B., Lifton, R. P., Brueckner, M., Zhao, H., & Lu, Q. (2022). Quantifying concordant genetic effects of de novo mutations on multiple disorders. *E Life Sciences*, *11(2)*, pp. e75551-e75559.

Hagberg, B., Aicardi, J., Dias, K., & Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Annals of Neurology*, *14*(4), pp. 471–479

Hagerman, P. J. (2008). The fragile X prevalence paradox. *Journal of Medical Genetics*, 45(8), pp. 498–499.

Hamanaka, K., Miyake, N., Mizuguchi, T., Miyatake, S., Uchiyama, Y., Tsuchida, N., *et al*. (2022). Large-scale discovery of novel neurodevelopmental disorder-related genes through a unified analysis of single-nucleotide and copy number variants. *Genome Medicine*, *14*(1), pp. 40-58.

Haraksingh, R. R., & Snyder, M. P. (2013). Impacts of variation in the human genome on gene regulation. *Journal of Molecular Biology*, *425*(21), pp. 3970–3977.

Henderson, L. B., Applegate, C. D., Wohler, E., Sheridan, M. B., Hoover-Fong, J., & Batista, D. A. S. (2014). The impact of chromosomal microarray on clinical management: A retrospective analysis. *Genetics in Medicine*, *16*(9), pp. 657–664.

Hengel, H., Buchert, R., Sturm, M., Haack, T. B., Schelling, Y., Mahajnah, M., *et al*. (2020). First-line exome sequencing in Palestinian and Israeli Arabs with neurological disorders is efficient and facilitates disease gene discovery. *European Journal of Human Genetics*, *28*(8), pp. 1034–1043.

Hobert, J. A., & Eng, C. (2009). PTEN hamartoma tumor syndrome: An overview. *Genetics in Medicine*, 11(10), pp. 687–694.

Hochstenbach, R., van Binsbergen, E., Engelen, J., Nieuwint, A., Polstra, A., Poddighe, P., *et al*. (2009). Array analysis and karyotyping: Workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. *European Journal of Medical Genetics*, *52*(4), pp. 161–169.

Hoffman-Andrews, L. (2017). The known unknown: The challenges of genetic variants of uncertain significance in clinical practice. *Journal of Law and the Biosciences*, *4*(3), pp. 648–657.

Hori, K., Shimaoka, K., & Hoshino, M. (2022). Auts2 gene: Keys to understanding the pathogenesis of neurodevelopmental disorders. *Cells*, *11*(1), pp. 11-24

Hormozdiari, F., Penn, O., Borenstein, E., & Eichler, E. E. (2015). The discovery of integrated gene networks for autism and related disorders. *Genome Research*, *25*(1), pp. 142–154.

Hosen, M. J., Anwar, S., Taslem Mourosi, J., Chakraborty, S., Miah, M. F., & Vanakker, O. M. (2021). Genetic counseling in the context of Bangladesh: current scenario, challenges, and a framework for genetic service implementation. *Orphanet Journal of Rare Diseases*, *16*(1), pp. 168 -184.

Houge, G., Haesen, D., Vissers, L. E. L. M., Mehta, S., Parker, M. J., Wright, M., *et al*. (2015). B56δ-related protein phosphatase 2A dysfunction identified in patients with intellectual disability. *Journal of Clinical Investigation*, *125*(8), pp. 3051–3062.

Howe, B., Umrigar, A., & Tsien, F. (2014). Chromosome preparation from cultured cells. *Journal of Visualized Experiments*, (83), pp. e50203.

Hsueh, Y. P. (2019). Synaptic Formation, Neural Circuits and Neurodevelopmental Disorders Controlled by Signaling, Translation, and Epigenetic Regulation. *Developmental Neurobiology*, 79 (1), pp. 2–7.

Hunter, J., Rivero-Arias, O., Angelov, A., Kim, E., Fotheringham, I., & Leal, J. (2014). Epidemiology of fragile X syndrome: A systematic review and meta-analysis. *American Journal of Medical Genetics,164*(7), pp. 1648–1658.

Hussein, A. A., Hamad, R., Newport, M. J., & Ibrahim, M. E. (2022). Individualized Medicine in Africa: Bringing the Practice Into the Realms of Population Heterogeneity. *Frontiers in Genetics*, 13 (3), pp. 132-153.

Innella, G., Bonora, E., Neri, I., Virdi, A., Guglielmo, A., Pradella, L. M., *et al*. (2021). PTEN Hamartoma Tumor Syndrome: Skin Manifestations and Insights Into Their Molecular Pathogenesis. *Frontiers in Medicine*, 8(2), pp. 31-48.

Inuzuka, L. M., Guerra-Peixe, M., Macedo-Souza, L. I., Pedreira, C. C., Gurgel-Gianetti, J., Monteiro, F. P., *et al*. (2021). MECP2-related conditions in males: A systematic literature review and 8 additional cases. *European Journal of Paediatric Neurology*, *34*, pp. 7–13.

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

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y. H., *et al*. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), pp. E5600–E5607.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., *et al*. (2012). De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron*, *74*(2), pp. 285–299.

Jang, W., Kim, Y., Han, E., Park, J., Chae, H., Kwon, A., *et al*. (2019). Chromosomal microarray analysis as a first-tier clinical diagnostic test in patients with developmental delay/intellectual disability, autism spectrum disorders, and multiple congenital anomalies: A prospective multicenter study in korea. *Annals of Laboratory Medicine*, *39*(3), pp. 299–310.

Janssen, N., Bergman, J. E., Swertz, M. A., Tranebjaerg, L., Lodahl, M., Schoots, J., *et al*. (2012). Mutation update on the CHD7 gene involved in CHARGE syndrome. *Human Mutation*, *33*(8), pp. 1149–1160.

Jeffrey, J. S., Leathem, J., King, C., Mefford, H. C., Ross, K., & Sadleir, L. G. (2021). Developmental and epileptic encephalopathy: Personal utility of a genetic diagnosis for families. *Epilepsia Open*, *6*(1), pp. 149–159.

Jennings, L. J., Arcila, M. E., Corless, C., Kamel-Reid, S., Lubin, I. M., Pfeifer, J., *et al*. (2017). Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics*, *19*(3), pp. 341–365.

Jiang, C. C., Lin, L. S., Long, S., Ke, X. Y., Fukunaga, K., Lu, Y. M., *et al*. (2022). Signalling pathways in autism spectrum disorder: mechanisms and therapeutic implications. *Signal Transduction and Targeted Therapy*, *7*(1), pp. 229-243.

Johnson, B., Ouyang, K., Frank, L., Truty, R., Rojahn, S., Morales, A., *et al*. (2022). Systematic use of phenotype evidence in clinical genetic testing reduces the frequency of variants of uncertain significance. *American Journal of Medical Genetics*, *188*(9), pp. 2642–2651.

Jongmans, M. C. J., Admiraal, R. J., Van Der Donk, K. P., Vissers, L. E. L. M., Baas, A. F., Kapusta, L., *et al*. (2006). CHARGE syndrome: The phenotypic spectrum of mutations in the CHD7 gene. *Journal of Medical Genetics*, *43*(4), pp. 306–314.

Jovčevska, I. (2020). Next Generation Sequencing and Machine Learning Technologies Are Painting the Epigenetic Portrait of Glioblastoma. *Frontiers in Oncology*, *10(20)*, pp. 798-812.

Ju, D., Hui, D., Hammond, D. A., Wonkam, A., & Tishkoff, S. A. (2022). Importance of Including Non-European Populations in Large Human Genetic Studies to Enhance Precision Medicine. *Annual Review of Biomedical Data Science*. https://doi.org/10.1146/annurev-biodatasci-122220

Kamp, M., Krause, A., & Ramsay, M. (2021). Has translational genomics come of age in Africa?. *Human Molecular Genetics*, *30*(20), pp.164-173.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., *et al*. (2020a). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), pp. 434–443.

Kay, A. C., Wells, J., Hallowell, N., & Goriely, A. (2023). Providing recurrence risk counselling for parents after diagnosis of a serious genetic condition caused by an apparently de novo mutation in their child: a qualitative investigation of the PREGCARE strategy with UK clinical genetics practitioners. *Journal of Medical Genetics*, *60*(9), pp. 925-931.

Kessi, M., Chen, B., Peng, J., Yan, F., Yang, L., & Yin, F. (2021). Calcium channelopathies and intellectual disability: a systematic review. *Orphanet Journal of Rare Diseases*, *16*(1), pp. 219-231.

Kim, H. J., Park, C. Il, Lim, J. W., Lee, G. M., Cho, E., & Kim, H. J. (2018). Phenotypic analysis of Korean patients with abnormal chromosomal microarray in patients with unexplained developmental delay/intellectual disability. *Yonsei Medical Journal*, *59*(3), pp. 431–437.

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

Kipkemoi, P., Kim, H. A., Christ, B., O'Heir, E., Allen, J., Austin-Tse, C., *et al*. (2023). Phenotype and genetic analysis of data collected within the first year of NeuroDev. *Neuron*, *111*(18), pp. 2800-2810.

Ko, M. H. J., & Chen, H. J. (2023). Genome-Wide Sequencing Modalities for Children with Unexplained Global Developmental Delay and Intellectual Disabilities—A Narrative Review. *Children*, 10(3), pp. 72-95.

Köhler, S., Øien, N. C., Buske, O. J., Groza, T., Jacobsen, J. O. B., McNamara, C., *et al*. (2019). Encoding Clinical Data with the Human Phenotype Ontology for Computational Differential Diagnostics. *Current Protocols in Human Genetics*, *103*(1), pp. e92-e117.

Krause, A. (2019). New genetic testing technologies: Advantages and limitations. *South African Medical Journal*, *109*(4), pp. 207–209.

Kromberg, J. G. R., Wessels, T. M., & Krause, A. (2013a). Roles of genetic counselors in South Africa. *Journal of Genetic Counseling*, *22*(6), pp. 753–761.

Kromberg, J. G., Sizer, E. B., & Christianson, A. L. (2013). Genetic services and testing in South Africa. *Journal of Community Genetics*, *4*(3b), pp. 413–423.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), pp. 1073–1082.

Kurt Colak, F., Eyerci, N., Aytekin, C., & Eksioglu, A. S. (2020). Renpenning Syndrome in a Turkish Patient: De novo Variant c.607C>T in PACS1 and Hypogammaglobulinemia Phenotype. *Molecular Syndromology*, *11*(3), pp. 157–161.

Lalani, S. R., Safiullah, A. M., Fernbach, S. D., Harutyunyan, K. G., Thaller, C., Peterson, L. E., *et al*. (2006). Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation. *American Journal of Human gGenetics*, *78*(2), pp. 303–314.

Lam, C. W. (2021). Ending diagnostic odyssey using clinical whole-exome sequencing (CWES). *Journal of Laboratory Medicine*, *45*(6), pp. 259–266.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., *et al*. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), pp. D1062–D1067.

Lantieri, F., Malacarne, M., Gimelli, S., Santamaria, G., Coviello, D., & Ceccherini, I. (2017). Custom array comparative genomic hybridization: The importance of DNA quality, an expert eye, and variant validation. *International Journal of Molecular Sciences*, 18(3), pp. 609-617.

Lappalainen, T., Scott, A. J., Brandt, M., & Hall, I. M. (2019). Genomic Analysis in the Age of Human Genome Sequencing. *Cell*, *177*(1), pp. 70–84.

Le Meur, N., Holder-Espinasse, M., Jaillard, S., Goldenberg, A., Joriot, S., Amati-Bonneau, P., *et al*. (2010). MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *Journal of Medical Genetics*, *47*(1), pp. 22–29.

Lei, T., Zhen, L., Yang, X., Pan, M., Fu, F., Han, J., *et al*., (2023). Prenatal Diagnosis of PPP2R1A-Related Neurodevelopmental Disorders Using Whole Exome Sequencing: Clinical Report and Review of Literature. *Genes*, *14*(1), pp. 126-142.

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

Lelieveld, S. H., Reijnders, M. R. F., Pfundt, R., Yntema, H. G., Kamsteeg, E. J., De Vries, *et al*., (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nature Neuroscience*, *19*(9), pp. 1194–1196.

Lelieveld, S. H., Wiel, L., Venselaar, H., Pfundt, R., Vriend, G., Veltman, J. A., *et al*.,. (2017). Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *American Journal of Human Genetics*, *101*(3), pp. 478–484.

Lenaerts, L., Reynhout, S., Verbinnen, I., Laumonnier, F., Toutain, A., Bonnet-Brilhault, F., *et al* (2021). The broad phenotypic spectrum of PPP2R1A-related neurodevelopmental disorders correlates with the degree of biochemical dysfunction. *Genetics in Medicine*, *23*, pp. 352-362.

Lennox, A. L., Hoye, M. L., Jiang, R., Johnson-Kerner, B. L., Suit, L. A., *et al*. (2020). Pathogenic DDX3X Mutations Impair RNA Metabolism and Neurogenesis during Fetal Cortical Development. *Neuron*, *106*(3), pp. 404-420

Liaci, C., Camera, M., Caslini, G., Rando, S., Contino, S., Romano, V., *et al*. (2021). Neuronal cytoskeleton in intellectual disability: From systems biology and modeling to therapeutic opportunities. *International Journal of Molecular Sciences*, *22*(11), pp. 6167-6181.

Lin, X., Yang, Y., Melton, P. E., Singh, V., Simpson-Yap, S., *et al*. (2022). Integrating Genetic Structural Variations and Whole-Genome Sequencing Into Clinical Neurology. *Neurology. Genetics*, *8*(4), pp. e200005.

Lorenz, J., Rothhammer-Hampl, T., Zoubaa, S., Bumes, E., Pukrop, T., Kölbl, O., *et al*. (2020). A comprehensive DNA panel next generation sequencing approach supporting diagnostics and therapy prediction in neurooncology. *Acta Neuropathologica Communications*, *8*(1), pp. 124-142.

Louw, N., Carstens, N., & Lombard, Z. (2023). Incorporating CNV analysis improves the yield of exome sequencing for rare monogenic disorders—an important consideration for resource-constrained settings. *Frontiers in Genetics*, 14(2), pp 103-122.

Lumaka, A., Carstens, N., Devriendt, K., Krause, A., Kulohoma, B., Kumuthini, J., *et al*. (2022). Increasing African genomic data generation and sharing to resolve rare and undiagnosed diseases in Africa: a call-to-action by the H3Africa rare diseases working group. *Orphanet Journal of Rare Diseases*, *17*(1), pp. 47-63.

MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, 42, pp. D986–D992.

Malapelle, U., Vigliar, E., Sgariglia, R., Bellevicine, C., Colarossi, L., Vitale, D., *et al*. (2015). Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *Journal of Clinical Pathology*, *68*(1), pp. 64–68.

Malherbe, H. L., Aldous, C., Christianson, A. L., Darlison, M. W., & Modell, B. (2021). Modelled epidemiological data for selected congenital disorders in South Africa. *Journal of Community Genetics*, *12*(3), pp. 357–376

Manickam, K., McClain, M. R., Demmer, L. A., Biswas, S., Kearney, H. M., Malinowski, J., *et al*. (2021). Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American

College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, *23*(11), pp. 2029–2037.

Manning, M., & Hudgins, L. (2010). Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. In *Genetics in Medicine 12*(11), pp. 742–745.

Marian A. J. (2020). Clinical Interpretation and Management of Genetic Variants. *JACC. Basic to Translational Science*, *5*(10), pp. 1029–1042.

Marsh, D. J., Trahair, T. N., Martin, J. L., Chee, W. Y., Walker, J., *et al*. (2008). Rapamycin treatment for a child with germline PTEN mutation. *Nature Clinical Practice Oncology*, *5*(6), pp. 357–361.

Martin, A. R., Williams, E., Foulger, R. E., Leigh, S., Daugherty, L. C., Niblock, O., *et al*. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics*, 51(11), pp. 1560–1565.

Martínez-monseny, A. F., Edo, A., Casas-alba, D., Izquierdo-serra, M., Bolasell, M., Conejo, D., *et al*. (2021). Cacna1a mutations causing early onset ataxia: Profiling clinical, dysmorphic, and structural-functional findings. *International Journal of Molecular Sciences*, *22*(10), pp. 5180-5199.

Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., *et al*. (2016). Guidelines for diagnostic next-generation sequencing. In *European Journal of Human Genetics*, *24*(10), pp. 1515-1530.

McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A., *et al*. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics*, *40*(10), pp. 1166–1174.

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., *et al*. (2016). Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modeling. *The Public Library of Science Medicine*, *13*(6), pp. 1–18.

McFeely, R. A. (1993). Chromosome abnormalities. *The Veterinary Clinics of North America. Food Animal Practice*, *9*(1), pp. 11–22.

McGlaughon, J. L., Goldstein, J. L., Thaxton, C., Hemphill, S. E., & Berg, J. S. (2018). The progression of the ClinGen gene clinical validity classification over time. *Human Mutation*, *39*(11), pp. 1494–1504.

McKnight, D., Bean, L., Karbassi, I., Beattie, K., Bienvenu, T., Bonin, H., *et al*. (2022). Recommendations by the ClinGen Rett/Angelman-like expert panel for gene-specific variant interpretation methods. *Human Mutation*, *43*(8), pp. 1097–1113.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., *et al*. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), pp. 122-145.

McRae, J. F., Clayton, S., Fitzgerald, T. W., Kaplanis, J., Prigmore, E., Rajan, D., *et al*. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, *542*(7642), pp. 433–438.

Mellone, S., Puricelli, C., Vurchio, D., Ronzani, S., Favini, S., Maruzzi, A., *et al*. (2022). The Usefulness of a Targeted Next Generation Sequencing Gene Panel in Providing Molecular

Diagnosis to Patients With a Broad Spectrum of Neurodevelopmental Disorders. *Frontiers in Genetics*, *13(3), pp. 45-63*.

Mencarelli, A., Prontera, P., Mencarelli, A., Rogaia, D., Stangoni, G., Cecconi, M., *et al*. (2018). Expanding the clinical spectrum of sotos syndrome in a patient with the new "c.[5867T>A]+[=]"; "p.[Leu1956Gln]+[=]" NSD1 missense mutation and complex skin hamartoma. *International Journal of Molecular Sciences*, *19*(10), pp. 3189 -3207.

Menke, L. A., Gardeitchik, T., Hammond, P., Heimdal, K. R., Houge, G., Hufnagel, S. B., *et al*. (2018). Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein–Taybi syndrome. *American Journal of Medical Genetics, 176*(4), pp. 862–876.

Mester, J. L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K. S., *et al*. (2018). Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Human Mutation*, *39*(11), pp. 1581–1592.

Miclea, D., Peca, L., Cuzmici, Z., & Pop, I. V. (2015). Genetic testing in patients with global developmental delay / intellectual disabilities. A review. *Medicine and Pharmacy Reports*, *88*(3), pp. 288–292. https://doi.org/10.15386/cjmed-461

Miclea, D., Szucs, A., Mirea, A., Stefan, D. M., Nazarie, F., Bucerzan, S., Lazea, C., *et al*. (2021). Diagnostic usefulness of MLPA techniques for recurrent copy number variants detection in global developmental delay/intellectual disability. *International Journal of General Medicine*, *14*, pp. 4511–4515.

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., *et al*. (2010). Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *American Journal of Human Genetics*, *86*(5), pp. 749–764.

Miyake, N., Ozasa, S., Mabe, H., Kimura, S., Shiina, M., Imagawa, E., *et al*. (2018). A novel missense mutation affecting the same amino acid as the recurrent PACS1 mutation in Schuurs-Hoeijmakers syndrome *Clinical Genetics*, *93*(4), pp. 929–930.

Moeschler, J. B., Shevell, M., Saul, R. A., Chen, E., Freedenberg, D. L., Hamid, R., *et al*. (2014). Comprehensive evaluation of the child with intellectual disability or global developmental delays. *Pediatrics*, *134*(3), pp. e903–e918.

Moeschler, J. B., Shevell, M., Schaefer, G. B., Bull, M. J., Enns, G. M., Gruen, J. R., *et al*. (2006). Clinical genetic evaluation of the child with mental retardation or developmental delays. *Pediatrics*, *117*(6), pp. 2304–2316.

Molina-Ramírez, L. P., Kyle, C., Ellingford, J. M., Wright, R., Taylor, A., Bhaskar, S. S., *et al*. (2022). Personalised virtual gene panels reduce interpretation workload and maintain diagnostic rates of proband-only clinical exome sequencing for rare disorders. *Journal of Medical Genetics*, *59*(4), pp. 393–398.

Moretto, E., Murru, L., Martano, G., Sassone, J., & Passafaro, M. (2018). Glutamatergic synapses in neurodevelopmental disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *84*(5), pp. 328–342.

Mossink, B., Negwer, M., Schubert, D., & Nadif Kasri, N. (2021). The emerging role of chromatin remodelers in neurodevelopmental disorders: a developmental perspective. *Cellular and Molecular Life Sciences 78*(6), pp. 2517–2563.

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. *Journal of Molecular Diagnostics, 18*(6), pp. 923–932.

Musante, L., & Ropers, H. H. (2014). Genetics of recessive cognitive disorders. *Trends in Genetics, 30*(1), pp. 32–39.

Nembaware, V., & Mulder, N. (2019). The African Genomic Medicine Training Initiative (AGMT): Showcasing a Community and Framework Driven Genomic Medicine Training for Nurses in Africa. *Frontiers in Genetics, 10(2),* pp. 147-154

Neul, J. L., Skinner, S. A., Annese, F., Lane, J., Heydemann, P., Jones, M., *et al*. (2020). Metabolic Signatures Differentiate Rett Syndrome From Unaffected Siblings. *Frontiers in Integrative Neuroscience*, 14(2), pp. 43-56.

Nguyen, L. S., Fregeac, J., Bole-Feysot, C., Cagnard, N., Iyer, A., Anink, J., *et al*. (2018). Role of miR-146a in neural stem cell differentiation and neural lineage determination: Relevance for neurodevelopmental disorders. *Molecular Autism*, *9*(1), pp. 451-461.

Nguyen, T. A., Lehr, A. W., & Roche, K. W. (2020). Neuroligins and Neurodevelopmental Disorders: X-Linked Genetics. *Frontiers in Synaptic Neuroscience* 12(2), pp. 33-49.

Nicola, P., Blackburn, P. R., Rasmussen, K. J., Bertsch, N. L., Klee, E. W., Hasadsri, L., *et al*. (2019). De novo DDX3X missense variants in males appear viable and contribute to syndromic intellectual disability. *American Journal of Medical Genetics,179*(4), pp. 570–578.

Niu, X., Yang, Y., Chen, Y., Cheng, M., Liu, M., Ding, C., *et al*. (2022). Genotype–phenotype correlation of CACNA1A variants in children with epilepsy. *Developmental Medicine and Child Neurology*, *64*(1), pp. 105–111.

Nott, A., & Holtman, I. R. (2023). Genetic insights into immune mechanisms of Alzheimer's and Parkinson's disease. In *Frontiers in Immunology*, 14(1), pp. 1168-1183.

O'Brien, S., Ng-Cordell, E., Astle, D. E., Scerif, G., & Baker, K. (2019). STXBP1-associated neurodevelopmental disorder: A comparative study of behavioural characteristics. *Journal of Neurodevelopmental Disorders*, *11*(1), pp. 653-667.

O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., *et al*. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, *43*(6), pp. 585–589.

Oakeson, K. F., Wagner, J. M., Mendenhall, M., Rohrwasser, A., & Atkinson-Dunn, R. (2017). Bioinformatic analyses of whole-genome sequence data in a public health laboratory. *Emerging Infectious Diseases*, *23*(9), pp. 1441–1445.

Oksenberg, N., Stevison, L., Wall, J. D., & Ahituv, N. (2013). Function and Regulation of AUTS2, a Gene Implicated in Autism and Human Evolution. *The Public Library of Science Genetics*, *9*(1), pp. 1723-1739.

Omotoso, O. E., Teibo, J. O., Atiba, F. A., Oladimeji, T., Adebesin, A. O., & Babalghith, A. O. (2022). Bridging the genomic data gap in Africa: implications for global disease burdens. *Globalization and Health*, 18(1), pp. 103.

Otsuka, M., Oguni, H., Liang, J. S., Ikeda, H., Imai, K., Hirasawa, K., *et al*. (2010). STXBP1 mutations cause not only Ohtahara syndrome but also West Syndrome-Result of Japanese cohort study. *Epilepsia*, *51*(12), pp. 2449–2452.

Owolabi, P., Adam, Y., & Adebiyi, E. (2023). Personalizing medicine in Africa: current state, progress, and challenges. *Frontiers in Genetics*, 14(2), pp. 472-479.

Pagon, R. A., Graham, J. M., Jr, Zonana, J., & Yong, S. L. (1981). Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *The Journal of Pediatrics*, *99*(2), pp. 223–227.

Pálla, S., Tőke, J., Bozsik, A., Butz, H., Papp, J., Likó, I., *et al*. (2023). Whole genome sequencing resolves 10 years diagnostic odyssey in familiar myxoma. *Scientific Reports*, *13*(1), pp. 451-462.

Parenti, I., Rabaneda, L. G., Schoen, H., & Novarino, G. (2020). Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in Neurosciences*, 43(8) pp. 608–621.

Park, K. B., Nam, K. E., Cho, A. R., Jang, W., Kim, M., & Park, J. H. (2019). Effects of copy number variations on developmental aspects of children with delayed development. *Annals of Rehabilitation Medicine*, *43*(2), pp. 215–223.

Pasquier, L., Minguet, G., Moisdon-Chataigner, S., Jarno, P., Denizeau, P., Volf, G., *et al*. (2022). How do non-geneticist physicians deal with genetic tests? A qualitative analysis. *European Journal of Human Genetics*, *30*(3), pp. 320–331.

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., & Goldstein, D. B. (2013). Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *The Public Library of Science Genetics*, *9*(8), pp. 56-63.

Pfennig, A., Petersen, L. N., Kachambwa, P., & Lachance, J. (2023). Evolutionary Genetics and Admixture in African Populations. *Genome Biology and Evolution*, 15(4), pp. 54-61.

Pilarski, R. (2019). PTEN hamartoma tumor syndrome: A clinical overview. *Cancers*, *11*(6), pp. 844.

Pilarski, R., Burt, R., Kohlman, W., Pho, L., Shannon, K. M., & Swisher, E. (2013). Cowden syndrome and the PTEN hamartoma tumor syndrome: Systematic review and revised diagnostic criteria. *Journal of the National Cancer Institute*, 105 (21), pp. 1607–1616.

Pócza, T., Grolmusz, V. K., Papp, J., Butz, H., Patócs, A., & Bozsik, A. (2021). Germline Structural Variations in Cancer Predisposition Genes. *Frontiers in Genetics*, 12(2), pp. 112-123.

Polyak, A., Rosenfeld, J. A., & Girirajan, S. (2015). An assessment of sex bias in neurodevelopmental disorders. *Genome Medicine*, *7*(1), pp. 341-351.

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., *et al*. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal*, 44(5), pp. 548–559.

Qian, Y., Jiang, Y., Wang, J., Li, G., Wu, B., Zhou, Y., *et al*. (2023). Novel Variants of PPP2R1A in Catalytic Subunit Binding Domain and Genotype–Phenotype Analysis in Neurodevelopmentally Delayed Patients. *Genes*, *14*(9), pp. 1750.

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

Quazi, S. (2022). Artificial intelligence and machine learning in precision and genomic medicine. *Medical Oncology*, 39(8), pp. 120.

Quinonez, S. C., O'Connor, B. C., Jacobs, M. F., Tekleab, A. M., Marye, A., Bekele, D., *et al*. (2021). The introduction of genetic counseling in Ethiopia: Results of a training workshop and lessons learned. *The Public Library of Science ONE*, *16*(7), pp. 567-580.

Rabbani, B., Tekin, M., & Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, *59*(1), pp. 5–15.

Rajan-Babu, I. S., Peng, J. J., Chiu, R., Adam, S., Souich, C. Du, Elliott, A., *et al*. (2021). Genome-wide sequencing as a first-tier screening test for short tandem repeat expansions. *Genome Medicine*, *13*(1), pp. 126-132.

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., et al. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American Journal of Medical Genetics*, 140(19), pp. 2063–2074

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., *et al*. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *The Lancet*, *380*(9854), pp. 1674–1682.

Ravnan, J. B., Tepperberg, J. H., Papenhausen, P., Lamb, A. N., Hedrick, J., Eash, D., *et al*. (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), pp. 478–489.

Redin, C., Gérard, B., Lauer, J., Herenger, Y., Muller, J., Quartier, A., *et al*. (2014). Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of Medical Genetics*, *51*(11), pp. 724–736.

Rehm, H. L. (2013). Disease-targeted sequencing: A cornerstone in the clinic. *Nature Reviews Genetics*, *14*(4), pp. 295–300.

Reid, E. S., Papandreou, A., Drury, S., Boustred, C., Yue, W. W., Wedatilake, Y., *et al*. (2016). Advantages and pitfalls of an extended gene panel for investigating complex neurometabolic phenotypes. *Brain*, *139*(11), pp. 2844–2854.

Reid, S., & Pal, T. (2020). Update on multi-gene panel testing and communication of genetic test results. *Breast Journal*, *26*(8), pp. 1513–1519.

Reiss, A. L. (2009). Childhood developmental disorders: An academic and clinical convergence point for psychiatry, neurology, psychology, and pediatrics. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *50*(1–2), pp. 87–98.

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*(1), pp. 886–894.

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*, *39*(17), pp. 118-128.

Reynhout, S., Jansen, S., Haesen, D., van Belle, S., de Munnik, S. A., Bongers, E. M. H. F., *et al*. (2019). De Novo Mutations Affecting the Catalytic Cα Subunit of PP2A, PPP2CA, Cause Syndromic Intellectual Disability Resembling Other PP2A-Related Neurodevelopmental Disorders. *American Journal of Human Genetics*, *104*(1), pp. 139–156.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., *et al*. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), pp. 405–423.

Şahin, G. E., Hoşnut, F. Ö., Yeşil, Ş., Lafcı, N. G., Gül, A. E., & Şahin, G. (2022). Sirolimus treatment of a PTEN hamartoma tumor syndrome presenting with melena. *Turkish Journal of Pediatrics*, *64*(4), pp. 766–774.

Saitsu, H., Kato, M., Okada, I., Orii, K. E., Higuchi, T., Hoshino, H., *et al*. (2010). STXBP1 mutations in early infantile epileptic encephalopathy with suppression-burst pattern. *Epilepsia*, *51*(12), pp. 2397-2405.

Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., *et al*. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, *46*(9), pp. 944–950.

Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., *et al*. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, *485*(7397), pp. 237–241.

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

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., *et al*. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, *12*(7), pp. 997.

Schmid, G. L., Kässner, F., Uhlig, H. H., Körner, A., Kratzsch, J., Händel, N., *et al*. (2014). Sirolimus treatment of severe PTEN hamartoma tumor syndrome: Case report and in vitro studies. *Pediatric Research*, *75*(4), pp. 527–534.

Schobers, G., Schieving, J. H., Yntema, H. G., Pennings, M., Pfundt, R., Derks, R., *et al*. (2022). Reanalysis of exome negative patients with rare disease: a pragmatic workflow for diagnostic applications. *Genome Medicine*,*14*(1), pp. 66.

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

Schuurs-Hoeijmakers, J. H. M., Oh, E. C., Vissers, L. E. L. M., Swinkels, M. E. M., Gilissen, C., Willemsen, M. A., Holvoet, M., *et al*. (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), pp. 1122–1127

Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, *7*(8), pp. 575–576.

Seibel, E., Gunn, G., Ali, N., Jordan, E., & Kenneson, A. (2022). Primary Care Providers' Use of Genetic Services in the Southeast United States: Barriers, Facilitators, and Strategies. *Journal of Primary Care and Community Health*, *13(2)*, pp. 1319-1332.

Seto, M. T. Y., Bertoli-Avella, A. M., Cheung, K. W., Chan, K. Y. K., Yeung, K. S., Fung, J. L. F., *et al*. (2021). Prenatal and postnatal diagnosis of Schuurs-Hoeijmakers syndrome: Case series and review of the literature. *American Journal of Medical Genetics*, *185*(2), pp. 384–389.

Seyhan, A. A., & Carini, C. (2019). Are innovation and new technologies in precision medicine paving a new era in patients centric care? *Journal of Translational Medicine*, 17(1), pp. 114.

Shaffer, L. G. (2005). American College of Medical Genetics guideline on the cytogenetic evaluation of the individual with developmental delay or mental retardation. *Genetics in Medicine*, *7*(9), pp. 650–654.

Shaffer, L., & Theisen. (2010). Disorders caused by chromosome abnormalities. *The Application of Clinical Genetics*, 3(1), pp. 159–174.

Shaikh, T. H. (2017). Copy Number Variation Disorders. *Current Genetic Medicine Reports*, *5*(4), pp. 183–190.

Shakoori, A. R. (2017). Fluorescence In Situ hybridization (FISH) and its applications. *Chromosome Structure and Aberrations*, 3(12), pp. 343–367.

Shang, L., Henderson, L. B., Cho, M. T., Petrey, D. S., Fong, C. T., Haude, K. M., *et al*. (2016). De novo missense variants in PPP2R5D are associated with intellectual disability, macrocephaly, hypotonia, and autism. *Neurogenetics*, *17*(1), pp. 43–49.

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., *et al*. (2013). Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Human Mutation*, *34*(1), pp. 57–65.

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. *Annals of Laboratory Medicine*, *35*(5), pp. 510–518.

Silk, M., Petrovski, S., & Ascher, D. B. (2019). MTR-Viewer: identifying regions within genes under purifying selection. *Nucleic Acids Research*, *47*(1), pp. 121-126.

Sima, A., Smădeanu, R. E., Simionescu, A. A., Nedelea, F., Vlad, A. M., & Becheanu, C. (2022). Menke–Hennekam Syndrome: A Literature Review and a New Case Report. *Children*, 9(5), pp. 759.

Singh, A. K., Olsen, M. F., Lavik, L. A. S., Vold, T., Drabløs, F., & Sjursen, W. (2021). Detecting copy number variation in next generation sequencing data from diagnostic gene panels. *BMC Medical Genomics*, *14*(1), pp.214.

Singh, R., Cohen, A. S. A., Poulton, C., Hjortshøj, T. D., Akahira-Azuma, M., Mendiratta, G., *et al*. (2021). Deletion of ERF and CIC causes abnormal skull morphology and global developmental delay. *Cold Spring Harbor Molecular Case Studies*, *7*(3), pp. 5991-6010.

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

Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M. R. F., *et al*. (2015). Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *American Journal of Human Genetics*, *97*(2), pp. 343–352.

Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., *et al*. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genetics in Medicine*, *21*(11), pp. 2413–2421.

Srour, M., & Shevell, M. (2014). Genetics and the investigation of developmental delay/intellectual disability. *Archives of Disease in Childhood*, *99*(4), pp. 386–389.

Stamberger, H., Nikanorova, M., Willemsen, M. H., Accorsi, P., Angriman, M., Baier, H., *et al*. (2016). STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology*, *86*(10), pp. 954–962.

Stankiewicz, P., Shaw, C. J., Withers, M., Inoue, K., & Lupski, J. R. (2004). Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Research*, *14*(11), pp. 2209–2220.

Stessman, H. A. F., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., *et al*. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nature Genetics*, *49*(4), pp. 515–526.

Stoddard, J. L., Niemela, J. E., Fleisher, T. A., & Rosenzweig, S. D. (2014). Targeted NGS: A cost-effective approach to molecular diagnosis of PIDs. *Frontiers in Immunology*, *5*(11), pp. 1–7.

Strande, N. T., & Berg, J. S. (2016). Defining the Clinical Value of a Genomic Diagnosis in the Era of Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, 17(1), pp. 303–332.

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), pp. 510–515.

Strømme P. (2000). Aetiology in severe and mild mental retardation: a population-based study of Norwegian children. *Developmental Medicine and Child Neurology*, *42*(2), pp. 76–86.

Sun, H., Shen, X. R., Fang, Z. B., Jiang, Z. Z., Wei, X. J., Wang, Z. Y., *et al*. (2021). Next-generation sequencing technologies and neurogenetic diseases. *Life*, *11*(4), pp. 361.

Sun, Y., Xiang, J., Liu, Y., Chen, S., Yu, J., Peng, J., *et al*. (2019). Increased diagnostic yield by reanalysis of data from a hearing loss gene panel. *BMC Medical Genomics*, *12*(1), pp. 76.

Suther, S., & Goodson, P. (2003). Barriers to the provision of genetic services by primary care physicians: A systematic review of the literature. *Genetics in Medicine*, 5(2), pp. 70–76.

Takenouchi, T., Miura, K., Uehara, T., Mizuno, S., & Kosaki, K. (2016). Establishing SON in 21q22.11 as a cause a new syndromic form of intellectual disability: Possible contribution to Braddock–Carey syndrome phenotype. *American Journal of Medical Genetics*, *170*(10), pp. 2587–2590.

Talkowski, M. E., Ordulu, Z., Pillalamarri, V., Benson, C. B., Blumenthal, I., Connolly, S., *et al*. (2012). Clinical Diagnosis by Whole-Genome Sequencing of a Prenatal Sample. *New England Journal of Medicine*, *367*(23), pp. 2226–2232.

Tang, L., Levy, T., Guillory, S., Halpern, D., Zweifach, J., Giserman-Kiss, I., *et al*. (2021). Prospective and detailed behavioral phenotyping in DDX3X syndrome. *Molecular Autism*, *12*(1), pp. 36.

Testa, B., Conteduca, G., Grasso, M., Cecconi, M., Lantieri, F., Baldo, C., *et al*. (2023). Molecular Analysis and Reclassification of NSD1 Gene Variants in a Cohort of Patients with Clinical Suspicion of Sotos Syndrome. *Genes*, *14*(2), pp. 295.

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), pp. 178–192.

Tokita, M. J. J., Braxton, A. A. A., Shao, Y., Lewis, A. M. M., Vincent, M., Küry, S., e. (2016). De Novo Truncating Variants in SON Cause Intellectual Disability, Congenital Malformations, and Failure to Thrive. *American Journal of Human Genetics*, *99*(3), pp. 720–727.

Tollefson, G. A., Schuster, J., Gelin, F., Agudelo, A., Ragavendran, A., Restrepo, I., *et al*. (2019). VIVA (VIsualization of VAriants): A VCF File Visualization Tool. *Scientific Reports*, *9*(1), pp. 1–7.

Tsurusaki, Y., Okamoto, N., Ohashi, H., Mizuno, S., Matsumoto, N., Makita, Y., *et al*. (2014). Coffin-Siris syndrome is a SWI/SNF complex disorder. *Clinical Genetics*, *85*(6), pp. 548–554.

Turner, T. N., Wilfert, A. B., Bakken, T. E., Bernier, R. A., Pepper, M. R., Zhang, Z., *et al*. (2019). Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders. *American Journal of Human Genetics*, *105*(6), pp. 1274–1285.

Uddin, M., Pellecchia, G., Thiruvahindrapuram, B., D'Abate, L., Merico, D., Chan, A., *et al*. (2016). Indexing Effects of Copy Number Variation on Genes Involved in Developmental Delay. *Scientific Reports*, 6(3), pp. 28663-28669.

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*(6), pp. 1–12.

Van der Sluijs, P. J., Jansen, S., Vergano, S. A., Adachi-Fukuda, M., Alanay, Y., AlKindy, A., *et al*. (2019). The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to Coffin-Siris syndrome. *Genetics in Medicine*, 21(6), pp. 1295–1307.

Van Nuland, A., Reddy, T., Quassem, F., Vassalli, J. D., & Berg, A. T. (2021). PACS1-Neurodevelopmental disorder: clinical features and trial readiness. *Orphanet Journal of Rare Diseases*, *16*(1), pp. 1–10.

Verloes, A. (2005). Updated diagnostic criteria for CHARGE syndrome: A proposal. *American Journal of Medical Genetics, 133(*3), pp. 306–308.

Vinkšel, M., Writzl, K., Maver, A., & Peterlin, B. (2021). Improving diagnostics of rare genetic diseases with NGS approaches. *Journal of Community Genetics*, 12(2), pp. 247–256.

Vissers, L. E. L. M., Gilissen, C., & Veltman, J. A. (2016). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*, *17*(1), pp. 9–18.

Wang, X. J., Jiao, Y., Ma, S., Yang, J. T., & Wang, Z. X. (2020). Whole-Genome Sequencing: An Effective Strategy for Insertion Information Analysis of Foreign Genes in Transgenic Plants. *Frontiers in Plant Science*, *11*(12), pp. 1–10.

Wang, X., Posey, J. E., Rosenfeld, J. A., Bacino, C. A., Scaglia, F., Immken, L. D., *et al*. (2018). Phenotypic expansion in DDX3X – a common cause of intellectual disability in females. *Annals of Clinical and Translational Neurology*, *5*(10), pp. 1277–1285.

Wang, X., Shen, X., Fang, F., Ding, C. H., Zhang, H., Cao, Z. H., *et al*. (2019). Phenotype-driven virtual panel is an effective method to analyze Wes data of neurological disease. *Frontiers in Pharmacology*, *9*(1), pp. 1529 -1538.

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., & Watson, M. (2015). Exome sequencing: Current and future perspectives. *G3: Genes, Genomes, Genetics*, *5*(8), pp. 1543–1550.

Wayhelova, M., Vallova, V., Broz, P., Mikulasova, A., Smetana, J., Dynkova Filkova, H., *et al*. (2024). Exome sequencing improves the molecular diagnostics of paediatric unexplained neurodevelopmental disorders. *Orphanet Journal of Rare Diseases*, *19*(1), pp. 41.

Weinstein, V., Tanpaiboon, P., Chapman, K. A., Mew, N. A., & Hofherr, S. (2017a). Do the data really support ordering fragile X testing as a first-tier test without clinical features? *Genetics in Medicine*, *19*(12), pp. 1317–1322.

Weinstein, V., Tanpaiboon, P., Chapman, K. A., Mew, N. A., & Hofherr, S. (2017b). Do the data really support ordering fragile X testing as a first-tier test without clinical features? *Genetics in Medicine*, *19*(12), pp. 1317–1322.

Wheeler, M. M., Stilp, A. M., Rao, S., Halldórsson, B. V., Beyter, D., Wen, J., *et al*. (2022). Whole genome sequencing identifies structural variants contributing to hematologic traits in the NHLBI TOPMed program. *Nature Communications*, *13*(1), pp. 1734 -1739.

Wiener, E. K., Buchanan, J., Krause, A., & Lombard, Z. (2023). Retrospective file review shows limited genetic services fails most patients – an argument for the implementation of exome sequencing as a first-tier test in resource-constraint settings. *Orphanet Journal of Rare Diseases, 18(1), pp. 81-86.*

Wilfert, A. B., Sulovari, A., Turner, T. N., Coe, B. P., & Eichler, E. E. (2017). Recurrent de novo mutations in neurodevelopmental disorders: Properties and clinical implications. *Genome Medicine*, *9*(1), pp. 1–16.

Witters, G., Van Robays, J., Willekes, C., Coumans, A., Peeters, H., Gyselaers, W., *et al*. (2011). Trisomy 13, 18, 21, Triploidy and Turner syndrome: the 5T's. Look at the hands. *Facts, Views & Vision in ObGyn*, *3*(1), pp. 15–21.

Won, D., Kim, S. H., Kim, B., Lee, S. T., Kang, H. C., & Choi, J. R. (2020). Reanalysis of Genomic Sequencing Results in a Clinical Laboratory: Advantages and Limitations. *Frontiers in Neurology, 11(3), pp. 612-615.*

Wright, C. F., Campbell, P., Eberhardt, R. Y., Aitken, S., Perrett, D., Brent, S., *et al*. (2023a). Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. *New England Journal of Medicine*, *388*(17), pp. 1559–1571.

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

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

Xue, Y., Ankala, A., Wilcox, W. R., & Hegde, M. R. (2015). Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: Single-gene, gene panel, or exome/genome sequencing. *Genetics in Medicine*, 17(6), pp. 444–451.

Yan, H., Shi, Z., Wu, Y., Xiao, J., Gu, Q., Yang, Y., *et al*. (2019). Targeted next generation sequencing in 112 Chinese patients with intellectual disability/developmental delay: Novel mutations and candidate gene. *BMC Medical Genetics*, *20*(1), pp. 80-91.

Yao, R., Yu, T., Qing, Y., Wang, J., & Shen, Y. (2019). Evaluation of copy number variant detection from panel-based next-generation sequencing data. *Molecular Genetics and Genomic Medicine*, *7*(1), pp. e00513-e00527.

Yuen, R. K. C., Merico, D., Cao, H., Pellecchia, G., Alipanahi, B., Thiruvahindrapuram, B., *et al*. (2016). Genome-wide characteristics of de novo mutations in autism. *Nature Portfolio journal Genomic Medicine, 1 (1), pp.* 160271–1602710.

Zahir, F. R., Mwenifumbo, J. C., Chun, H. J. E., Lim, E. L., Van Karnebeek, C. D. M., Couse, M., *et al*. (2017). Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. *BMC Genomics*, *18*(1), pp. 403.

Zentner, G. E., Layman, W. S., Martin, D. M., & Scacheri, P. C. (2010). Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *American Journal of Medical Genetics*,152(3), pp. 674–686.

Zhang, F., & Lupski, J. R. (2015). Non-coding genetic variants in human disease. *Human Molecular Genetics*, 24(1), pp. 102–110.

Zhao, M., Havrilla, J. M., Fang, L., Chen, Y., Peng, J., Liu, C., *et al*. (2020). Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genomics and Bioinformatics*, *2*(2), pp. 32-38.

Zhao, X., Li, C., Guillermo Paez, J., Chin, K., Jänne, P. A., Chen, T.-H., *et al*. (2004). An Integrated View of Copy Number and Allelic Alterations in the Cancer Genome Using Single Nucleotide Polymorphism Arrays. *Cancer Research*, *64*(9), pp. 3060–3071.

Zheng, J., Zhang, H., Banerjee, S., Li, Y., Zhou, J., Yang, Q., *et al*. (2019). A comprehensive assessment of Next-Generation Sequencing variants validation using a secondary technology. *Molecular Genetics and Genomic Medicine*, *7*(7), pp. 748-757.

Zhong, A., Xia, K., Hadjis, Z., Lifman, G., Njambi, L., & Dimaras, H. (2021). Opportunities and barriers for genetic service delivery in Kenya from a health personnel perspective. *Journal of Community Genetics*, *12*(4), pp. 525–538.

Zhu, X., Petrovski, S., Xie, P., Ruzzo, E. K., Lu, Y. F., McSweeney, K. M., *et al*. (2015). Whole-exome sequencing in undiagnosed genetic diseases: Interpreting 119 trios. *Genetics in Medicine*, *17*(10), pp. 774–781.

Zweier, M., & Rauch, A. (2011). The MEF2C-related and 5q14.3q15 microdeletion syndrome. *Molecular Syndromology*, *2*(5), pp. 164–170.

Appendix A: DDD-Africa Study Inclusion and Exclusion Criteria

DDD-AFRICA: INCLUSION & EXCLUSION CRITERIA

Doctor Name: _____ Date: _____

Patient Name: _____

INCLUSION:

A. **Intellectual disability (moderate to profound)** ☐

B. **Multiple major malformations[1] in 2 or more different organ systems** ☐

C. **One major malformation AND 3 or more well-documented minor anomalies** ☐

D. **Mild to moderate intellectual disability AND dysmorphic features** ☐

Organ Systems affected:

1. _____
2. _____
3. _____
4. _____

*Major Malformations:	Minor Anomalies:
1. _____	1. _____
2. _____	2. _____
3.	3.
4. _____	4. _____
5. _____	5. _____

Diagnostic tests done:

☐ **MLPA** ☐ **Chromosomes** ☐ **FRAX** ☐ **Array CGH (normal)**

EXCLUSION:

☐ Known conditions for which no strong evidence exist for a monogenic cause

☐ Suspicion of an acquired brain lesion, with predominant neurological manifestations

☐ Suspected multifactorial cause

☐ Suspected environmental cause

☐ Mild ID **only**

a _____

[1] *Microcephaly and macrocephaly are major malformations if ±3SD.*

Page | 23

Appendix B: Ethical Clearance Certificate for the Current PhD Study

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Miles P Nevondwe & F Essop; Prof Z Lombard

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M200440

NAME: (Principal Investigator)	Miles P Nevondwe & F Essop; Prof Z Lombard
DEPARTMENT:	School of Pathology Department of Human Genetics Medical School University
PROJECT TITLE:	Designing and evaluating the utility of a panel of de novo mutation (DNM) enriched genes for diagnosing South African patients with developmental delay (DD) Sub-study under M160830, M 180506 and M 180678
DATE CONSIDERED:	2020/04/24
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Professor A Krause and Dr N Carstens
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2020/07/31

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 **April** and will therefore reports and re-certification will be due early in the month of **April** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date

Appendix C: A list of the 285 DNM-Enriched Genes

Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
ABI2	606442	Not curated	Not curated				√
ACHE	100740	Not curated	Green	√			
ACTC1	102540	Definitive	Green		√		
ADAP1	608114	Not curated	Not curated				√
ADCY3	600291	Not curated	Green			√	
ADNP	611386	Definitive	Green	√			
AFF2	300806	Definitive	Green	√			
AGAP2	605476	Not curated	Green				√
AGO1	606228	Not curated	Green			√	
AGO3	607355	Not curated	Not curated			√	
AGO4	607356	Not curated	Not curated			√	
AHDC1	615790	Definitive	Green	√			
AKAP9	604001	Disputed	Green				√
ALG13	300776	Not curated	Green	√			
ANK2	106410	Definitive	Green	√			
ANKRD11	611192	Definitive	Green	√			
ANP32A	600832	Not curated	Not curated	√			
AQP10	606578	Not curated	Not curated		√		
ARID1B	614556	Definitive	Green	√			
ARID2	609539	Not curated	Green	√			
ASH1L	607999	Not curated	Green				√
ASXL1	612990	Definitive	Green				
ASXL3	615115	Definitive	Green	√			
AUTS2	607270	Definitive	Green	√			
BAZ2B	605683	Limited	Green	√			
BCL11A	606557	Definitive	Green			√	
BRAF	164757	Definitive	Green			√	
BRPF1	602410	Not curated	Green			√	
BTF3	602542	Not curated	Not curated			√	
C2ORF42	-	Not curated	Not curated			√	
CABP7	618759	Not curated	Not curated			√	
CACNA1A	601011	Not curated	Green			√	
CACNA1C	114205	Moderate	Green			√	
CACNA1E	601013	Not curated	Green				√
CACNA2D3	606399	Not curated	Green				√
CAPN15	603267	Not curated	Green	√			
CAPRIN1	601178	Not curated	Green				√
CASK	300172	Definitive	Green			√	
CASZ1	609895	Not curated	Not curated	√			
CBL	165360	Definitive	Green		√		
CDC42BPB	614062	Not curated	Green				√
CDK13	603309	Definitive	Green	√			
CDKL5	300203	Definitive	Green	√			
CHAMP1	616327	Not curated	Green	√			
CHD2	602119	Definitive	Green			√	
CHD3	602120	Definitive	Green			√	

Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
CHD4	603277	Not curated	Green			√	
CHD7	608892	Definitive	Green			√	
CHD8	610528	Definitive	Green			√	
CIC	612082	Definitive	Green			√	
CLASP1	605852	Not curated	Green			√	
CLTC	118955	Not curated	Green			√	
CNKSR2	300724	Definitive	Green				√
CNOT3	604910	Definitive	Green			√	
COL4A3BP	-	Not curated	Green				√
CREBBP	600140	Definitive	Green			√	
CSNK2A1	115440	Definitive	Green			√	
CTCF	604167	Definitive	Green			√	
CTNNB1	116806	Definitive	Green				√
CUL3	603136	Definitive	Green			√	
DDX3X	300160	Definitive	Green			√	
DEAF1	602635	Not curated	Green			√	
DIP2A	607711	Not curated	Green		√		
DIP2C	611380	Not curated	Green			√	
DLG4	602887	Definitive	Green				√
DLX3	600525	Not curated	Green			√	
DNM1	602377	Definitive	Green			√	
DNMT3A	602769	Definitive	Green		√		
DSCAM	602523	Definitive	Green			√	
DVL3	601368	Not curated	Green			√	
DYNC1H1	600112	Not curated	Green	√			
DYRK1A	600855	Definitive	Green				√
EBF3	607407	Not curated	Green			√	
EEF1A2	602959	Definitive	Green				√
EFTUD2	603892	Definitive	Green			√	
EGLN2	606424	Not curated	Green		√		
EHMT1	607001	Definitive	Green	√			
ENO3	131370	Definitive	Green			√	
EP300	602700	Definitive	Green	√			
FAM104A	-	Not curated	Not curated				√
FAM200A	-	Not curated	Not curated				√
FAM200B	-	Not curated	Not curated				√
FAM47A	-	Not curated	Green				√
FOSL2	601575	Not curated	Not curated	√			
FOXG1	164874	Definitive	Green			√	
FOXP1	605515	Definitive	Green	√			
FOXP2	605317	Definitive	Not curated	√			
FRYL	-	Not curated	Not curated				√
GABRB2	600232	Not curated	Green				√
GABRB3	137192	Definitive	Green	√			
GATAD2B	614998	Definitive	Green				√
GIGYF1	612064	Not curated	Green			√	

Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
GLRA2	305990	Not curated	Green				
GNAI1	139310	Definitive	Green	√			
GNAO1	139311	Definitive	Green	√			
GRIN2B	138252	Definitive	Green			√	
HDAC8	300269	Definitive	Green	√			
HECTD4	-	Not curated	Green			√	
HECW2	617245	Not curated	Green				√
HIST1H1E	-	Not curated	Green				√
HIST1H2AC	-	Not curated	Red			√	
HIVEP2	143054	Not curated	Green	√			
HIVEP3	606649	Not curated	Red	√			
HMGXB3	619800	Not curated	Not curated				√
HNRNPD	601324	Not curated	Amber	√			
HNRNPU	602869	Definitive	Green			√	
HUWE1	300697	Definitive	Green			√	
IQSEC2	300522	Definitive	Green			√	
IRF2BPL	611720	Not curated	Green				√
ITPR1	147265	Definitive	Green				√
KANSL1	612452	Definitive	Green	√			
KAT6A	601408	Not curated	Green	√			
KAT6B	605880	Definitive	Green	√			
KATNAL2	614697	Disputed	Red				√
KCNC1	176258	Definitive	Green			√	
KCND3	605411	Disputed	Red				√
KCNH1	603305	Definitive	Green			√	
KCNJ6	600877	Not curated	Green			√	
KCNQ2	602235	Definitive	Green			√	
KCNQ3	602232	Not curated	Green				√
KCNS3	603888	Not curated	Not curated				√
KDM5B	605393	Definitive	Green			√	
KDM6A	300128	Definitive	Green	√			
KDM6B	611577	Definitive	Green	√			
KIAA2022	-	Definitive	Not curated	√			
KIF1A	601255	Definitive	Green			√	
KIF5C	604593	Moderate	Green			√	
KMT2A	159555	Definitive	Green			√	
KMT2C	606833	Definitive	Green				√
KMT2E	608444	Definitive	Green				√
KMT5B	610881	Definitive	Green			√	
LAMC3	604349	Definitive	Not curated				√
LARP4B	616513	Not curated	Not curated	√			
LEO1	610507	Not curated	Amber	√			
LRP2	600073	Not curated	Not curated	√			
MAP2K1	176872	Definitive	Not curated			√	
MAPK3	601795	Not curated	Red		√		
MBD5	611472	Definitive	Green	√			

Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
MECP2	300005	Definitive	Green	√			
MED12	300188	Definitive	Green			√	
MED12	300188	Definitive	Green		√		
MED13L	608771	Definitive	Green			√	
MEF2C	600662	Definitive	Green			√	
MEIS2	601740	Definitive	Green	√			
MSL3	300609	Definitive	Green	√			
MTF2	609882	Not curated	Not curated	√			
MYO1E	601479	Not curated	Green				√
MYO5A	160777	Definitive	Green			√	
MYT1L	613084	Definitive	Green			√	
NAA10	300013	Definitive	Green				√
NAA15	608000	Definitive	Green			√	
NCKAP1	604891	Not curated	Green			√	
NFE2L3	604135	Not curated	Not curated				√
NFIX	164005	Definitive	Green			√	
NLGN3	300336	Definitive	Green			√	
NLGN4X	300427	Definitive	Green				√
NONO	300084	Not curated	Green			√	
NR2F1	132890	Definitive	Green			√	
NR4A2	601828	Definitive	Green				√
NSD1	606681	Definitive	Green			√	
NTNG1	608818	Not curated	Green				√
ODC1	165640	Not curated	Green		√		
PACS1	607492	Definitive	Green			√	
PACS2	610423	Not curated	Green	√			
PAPOLG	616865	Not curated	Not curated	√			
PAX5	167414	Moderate	Red	√			
PBX1	176310	Definitive	Green				√
PCDH11X	300246	Not curated	Green		√		
PDHA1	300502	Definitive	Green	√			
PDK2	602525	Not curated	Green				√
PER2	603426	Not curated	Green		√		
PHF2	604351	Not curated	Green			√	
PHF21A	608325	Definitive	Green	√			
PHF3	607789	Not curated	Green	√			
PHIP	612870	Definitive	Green			√	
PIK3CA	171834	Definitive	Green			√	
PLAC8L1	-	Not curated	Not curated				√
PLK5	-	Not curated	Not curated				√
POGZ	614787	Definitive	Green			√	
POU3F3	602480	Definitive	Green			√	
PPM1D	605100	Definitive	Green			√	
PPP1CB	600590	Definitive	Green			√	
PPP2R1A	605983	Definitive	Green			√	
PPP2R5D	601646	Definitive	Green			√	

Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
PRKCA	176960	Not curated	Not curated		√		
PRKD1	605435	Not curated	Green			√	
PRPF18	604993	Definitive	Not curated				√
PRPF40A	612941	Not curated	Not curated		√		
PRR12	616633	Not curated	Green				√
PSMG4	617550	Not curated	Not curated		√		
PTCHD1	300828	Definitive	Green	√			
PTEN	601728	Definitive	Green		√		
PTPN11	176876	Definitive	Green			√	
PUF60	604819	Definitive	Green				√
PURA	600473	Definitive	Green			√	
QRICH1	617387	Definitive	Green			√	
RAI1	607642	Definitive	Green	√		√	
RFX8	-	Not curated	Not curated			√	
RIMS1	606629	Definitive	Green				
RPL26	603704	Not curated	Green	√			
RRP8	615818	Not curated	Not curated				√
SATB2	608148	Definitive	Green			√	
SCN1A	182389	Definitive	Green			√	
SCN2A	182390	Definitive	Green			√	
SCN8A	600702	Definitive	Green				√
SEPT10	-	Not curated	Not curated			√	
SET	600960	Definitive	Green	√			
SETBP1	611060	Definitive	Green	√			
SETD1B	611055	Not curated	Green			√	
SETD2	612778	Not curated	Green				√
SETD5	615743	Not curated	Green	√			
SF3B1	605590	Not curated	Green			√	
SHANK2	603290	Definitive	Green			√	
SHANK3	606230	Definitive	Green			√	
SHISA6	617327	Not curated	Not curated				√
SIN3A	607776	Definitive	Green			√	
SKIDA1	-	Not curated	Not curated	√			
SLC35A2	314375	Definitive	Green		√		
SLC6A1	137165	Not curated	Green			√	
SMAD4	-	Definitive	Green			√	
SMAD6	602931	Not curated	Green		√		
SMARCA2	600014	Definitive	Green			√	
SMARCA4	603254	Definitive	Green			√	
SMARCC2	601734	Definitive	Green			√	
SMARCD1	601735	Not curated	Green			√	
SMC1A	300040	Not curated	Green			√	
SMC3	606062	Definitive	Green	√			
SNAPC5	605979	Not curated	Not curated				√
SOX5	605937	Definitive	Not curated			√	
SON	182465	Not curated	Green			√	

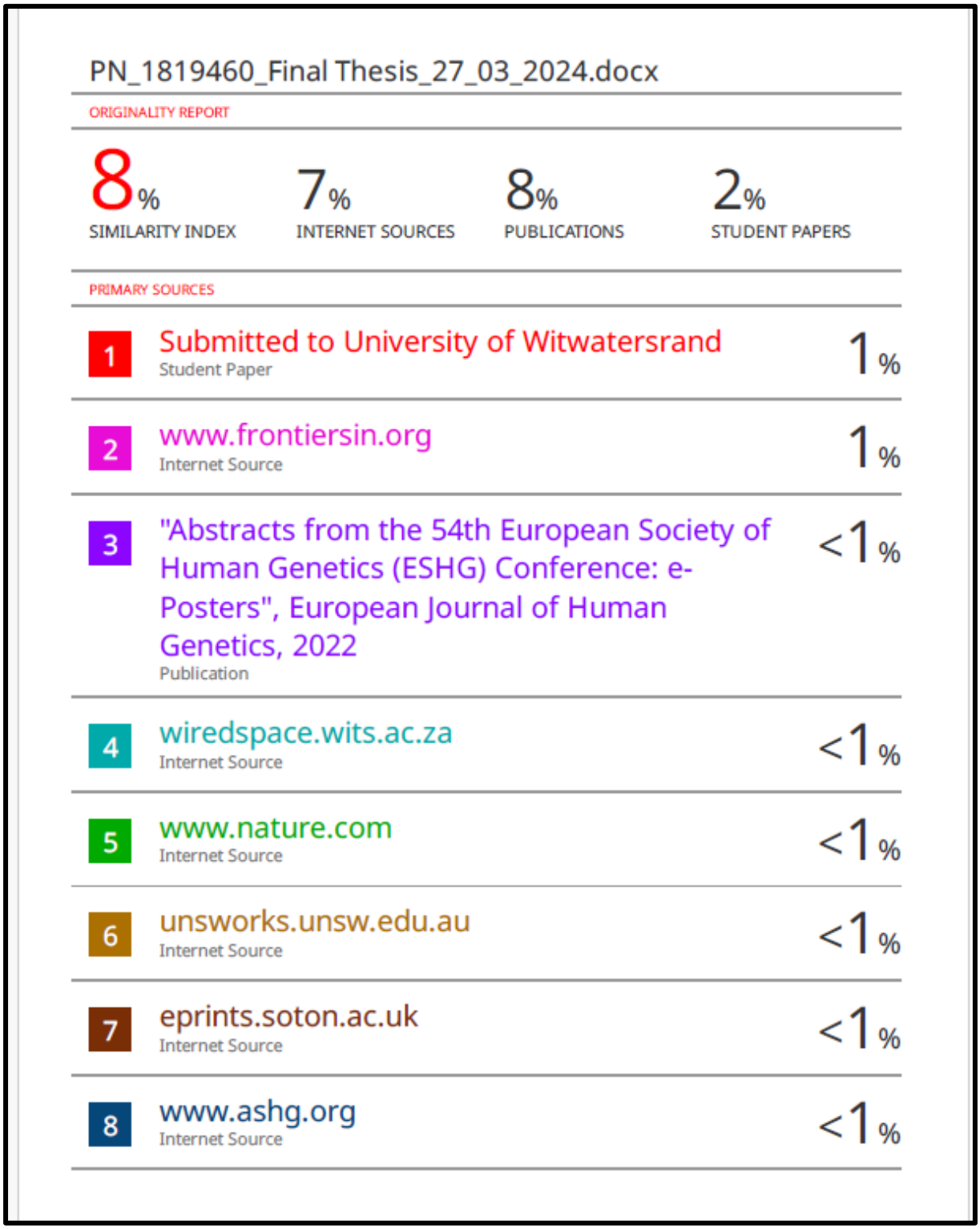
Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
SOX5	604975	Definitive	Green			√	
SPAST	604277	Not curated	Green	√			
SPEN	613484	Not curated	Green	√			
SPRY2	602466	Not curated	Green	√			
SRCAP	611421	Definitive	Green	√			
SRRM2	606032	Not curated	Green	√			
SRSF11	602010	Definitive	Green	√			
STARD9	614642	Not curated	Not curated		√		
STC1	601185	Not curated	Not curated			√	
STXBP1	602926	Not curated	Green			√	
SUSD4	615827	Not curated	Not curated				√
SYNCRIP	616686	Not curated	Green			√	
SYNGAP1	603384	Definitive	Green			√	
SYT1	185605	Not curated	Green				√
TAB2	605101	Definitive	Green	√			
TAF1	313650	Not curated	Green	√			
TAF6	602955	Not curated	Green		√		
TAOK1	610266	Definitive	Green			√	
TBL1XR1	608628	Definitive	Green			√	
TBR1	604616	Definitive	Green			√	
TCF20	603107	Definitive	Green	√			
TCF4	602272	Definitive	Green			√	
TCF7L2	602228	Definitive	Green	√			
TLK2	608439	Not curated	Green				
TMEM178A	-	Not curated	Not curated			√	
TMEM42	-	Not curated	Not curated				√
TNPO2	603002	Not curated	Green			√	
TNPO3	610032	Not curated	Green		√		
TNRC14	604520	Not curated	Not curated			√	
TNRC6B	610740	Definitive	Green	√			
TRA2B	602719	Not curated	Green				
TRAF7	606692	Not curated	Green		√		
TRIO	601893	Definitive	Green			√	
TRIP12	604506	Definitive	Green			√	
TRRAP	603015	Definitive	Green			√	
TTN	188840	Definitive	Green				√
U2AF2	191318	Not curated	Green			√	
UBN2	613841	Not curated	Green		√		
UNC80	612636	Definitive	Green		√	√	
UPF3B	300298	Definitive	Green	√			
USP9X	300072	Definitive	Green			√	
VAMP2	185881	Not curated	Green	√			
VEZF1	606747	Not curated	Not curated	√			
WAC	615049	Definitive	Green			√	
WDFY3	617485	Definitive	Green			√	
WDFY4	613316	Not curated	Not curated		√		
WDR26	617424	Definitive	Green			√	

Gene	OMIM number	ClinGen curation status[1]	PanelApp curation status[1]	Intolerant to LoF variation only (pLI ≥ 0.9)[2]	Intolerant to missense variation only[2]	Intolerant to both LoF and missense variation[2]	Not predicted to be intolerant to both variation[2]
WDR45	300526	Definitive	Green	√			
WDR87	-	Not curated	Not curated		√		
WHSC1	-	Definitive	Not curated			√	
YTHDF3	618669	Not curated	Not curated	√			
YWHAG	605356	Not curated	Green	√			
ZBTB18	608433	Not curated	Green			√	
ZBTB7A	605878	Not curated	Green			√	
ZC4H2	300897	Definitive	Green				√
ZMYND11	608668	Definitive	Green			√	
ZNF292	616213	Definitive	Green			√	

[1] Evaluated 15th October 2023

[2] Evaluated July 2020

Appendix D: Turnitin Report



PN_1819460_Final Thesis_27_03_2024.docx

ORIGINALITY REPORT

8%	7%	8%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Witwatersrand Student Paper	1%
2	www.frontiersin.org Internet Source	1%
3	"Abstracts from the 54th European Society of Human Genetics (ESHG) Conference: e-Posters", European Journal of Human Genetics, 2022 Publication	<1%
4	wiredspace.wits.ac.za Internet Source	<1%
5	www.nature.com Internet Source	<1%
6	unsworks.unsw.edu.au Internet Source	<1%
7	eprints.soton.ac.uk Internet Source	<1%
8	www.ashg.org Internet Source	<1%