Genotype-Phenotype Correlations in Neurodevelopmental Disorders

Hussein Kassir

ID 231032449

William Harvey Research Institute

Supervisors:

Prof. Damian Smedley

Dr. Pilar Cacheiro

Table of Contents

Acknowledgements:

At the end of this fulfilling journey, I would like to give my regards to several people without whom this would not have been possible. First, I would like to express my deepest gratitude to my supervisors. Professor Damian Smedley for his exceptional kindness, prompt responses and support despite his busy schedule. It showed me the importance of supportive leadership. I am profoundly thankful to Dr. Pilar Cacheiro and for her continuous feedback, encouragement, and belief in my capabilities. Her trust in my work and her guidance have been invaluable, providing me with the confidence to persevere and deliver this dissertation.

I also extend my appreciation to the course lead, Professor Panos Deloukas for his support and guidance throughout the course. It is an honour to have undertaken this program under his mentorship, given his significant contributions to the field of genomics. I am grateful to the academic staff for all the information and knowledge provided and to the admin staff for their support and for providing access to essential resources. I would also like to acknowledge my colleagues and peers for their spirit and valuable discussions.

To my amazing partner, Georges, thank you for your utmost support, your firm belief in me when I doubted myself, and for always lifting me. Your love, encouragement, and countless thoughtful gestures have been my greatest source of strength. I hope this dissertation stands as a testament to our love and here's to building our future together.

This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.(1)

Abstract:

Neurodevelopmental disorders (NDDs) include a group of conditions, such as autism spectrum disorder, intellectual disability, attention deficit hyperactivity disorder, and speech impediments. NDDs impact brain development and function and are characterised by heterogeneity in clinical presentations and aetiologies. Research continues to uncover potential pathways, genes, and mechanisms by which NDDs are affected. This dissertation uses patient data from the 100,000 Genomes Project to explore genotype-phenotype correlations in NDDs, focusing on intellectual disability (ID). The analysis focuses on 22 genes to identify correlations between genetic features and phenotypes to improve the understanding and management of NDDs. Phenotypic features are assessed using standardised Human Phenotype Ontology (HPO) terms and revealed associations between genes, variant types, and variant location on the phenotype outcomes. The analysis also highlights notable gene-specific comorbidities, including cardiovascular, ocular, endocrine, and musculoskeletal abnormalities. A comprehensive statistical analysis using Fisher's exact test was performed to assess results and a logistic regression model was implemented to predict certain outcomes. A statistically significant relationship was established for SNC2A's involvement with a delay in social development aligning with its implications in autism spectrum disorder, highlighting the importance of detailed genotype-phenotype analyses to enhance the understanding of NDDs and improve diagnostic precision, therapeutic strategies, and patient outcomes.

Introduction:

Neurodevelopmental disorders (NDDs) are a group of diseases that mainly affect brain development and its function, with a complex heterogeneous aetiology that is not fully understood (2). They affect the individual's personal, social, academic, or occupational functioning. Historically, the first mention of developmental disorder was back in the 1800s in a textbook by Étienne Jean Georget (1795-1828), a student of Philippe Pinel (1745-1826) and Jean-Étienne Esquirol (1772-1840)(3). Pionel's work mainly describes intellectual disability as a psychiatric illness in his classification of mental illnesses. Many of the terms used at the time to describe mental illnesses or people with diminished intellectual disabilities would be considered derogatory in today's world. However, we are luckily past those times, with guidelines and training currently being implemented to address the language used in healthcare settings. In his work on recommendations for healthcare providers to address bias, C.R. Herron states that written and verbal stigmatising language can fuel bias among healthcare professionals (4). Rather, we should opt for more inclusive and uplifting language to avoid any bias or negative outcomes that can arise due to the affective nature of stigmatising language. For example, instead of 'Crippled by' X condition we can sue 'Disabled person' / 'Person who has ' / 'Person with' (5).

There has been a growing debate on the categorisation of NDDs, which typically includes any disability affecting brain development (6). This broadness can be misleading, as it includes many neurological and psychiatric disorders that are clinically and causally different (6). Organisations and scientific bodies that research and update medical dilemmas have played a significant role in helping classify psychiatric disorders. For example, the American Psychiatric Association's (APA) guidebook, known as the Diagnostic and Statistical Manual of Mental Disorders (DSM), is a comprehensive diagnostic manual used by healthcare professionals to diagnose mental illnesses. According to the (DSM-V) criteria (7), NDDs include Attention Deficit Hyperactivity Disorder (ADHD), autism spectrum disorder (ASD), **intellectual disability (ID)**–which will be the focus of this dissertation–learning disabilities, communication disorders, and Neurodevelopmental Motor Disorders, including Tic Disorders (a list of specific major symptoms are shown in Table 1). The key characteristic of these disorders is their early childhood manifestations, usually before puberty. Unlike other psychiatric disorders, such as bipolar or mood disorder that can remit and relapse, NDDs follow a steady course (6).

Disorder	Major Symptoms
Intellectual Disability	• Impaired mental function in conceptual (language, reading knowledge, interpretation), Social (empathy, compassion, judgment), and practical (hobbies and personal care) aspects.

ADHD	• Impaired attention with bursts of hyperactivity/impulsivity
Autism Spectrum Disorder	• Persistent deficits in reciprocal social communication and interaction, as well as restricted and repetitive patterns of behaviour, interest, or thoughts.
Learning Disabilities	• Difficulties in learning skills such as writing, speaking, and spelling.

Table 1: NDDs and their symptoms(7)

Although helpful, such classifications can make diagnosis difficult. As with all other diseases that affect cognition and behaviour, diagnosis is primarily made in a clinical setting via questionnaires and observations, with no available biomarkers indicative of such disorders. This is particularly troublesome since it makes it difficult to differentiate between NNDs, which tend to share similar symptoms (2). Taking ADHD and ASD as an example, Rommelse et al (8) show that there is considerable overlap in the clinical, genetic, and neuropsychological link between both NDDs. Sokolova, Oerlemans (9) mention in their causality study for these 2 disorders that 22-83% of patients with ASD satisfy the criteria for ADHD, and 30–65% of children with ADHD have clinically significant symptoms of ASD. They suggest that this linkage could be explained by hereditary genetic factors. Other studies found that certain chromosomal regions harbour quantitative trait loci (QTLs)[1](10) affiliated with language and communication (9). Together, these findings suggest that detecting and annotating genetic information is essential in evaluating a comprehensive aetiology of NDDs to understand and treat such disorders effectively. Detecting and annotating genetic information will also enable the establishment of diagnostic criteria to easily differentiate between disorders, which will help in treatment selection given the difference in disease management protocols.

Many studies were performed to determine possible causes of NDDs, with results indicating that these causes are multifactorial. They range from social deprivation, genetic and metabolic diseases, nutritional factors, physical trauma, and toxic and prenatal environmental factors (11). In addition, studies also focused on the biological pathways that are involved -Figure 1.

[1] Quantitative trait loci (QTL) are genetic regions that influence phenotypic variation of a complex trait, often through genetic interactions with each other and the environment.

Neurodevelopmental disorders	Prevalence	Proteins or biological pathways
Learning disabilities	2–4%	Chromatin remodeling Metabolism Actin skeleton organization Channels Synaptogenesis Neurotransmission
Dyslexia	5–15%	Neuronal migration?
ADHD	1.7–9%	Synapses? Cortical maturation?
ASDs	0.6–1.2%	Chromatin remodeling Metabolism Actin skeleton organization Channels Synapses
Epilepsy	0.45–1%	Synapses Channels
Fetal alcohol syndrome	0.1–5%	—

Figure 1: Prevalence and Biological Pathways Involved in NDDs(11)

Prevalence-wise, NNDs affect millions of people worldwide. According to Petersen et al, 2014, approximately 15 to 20 % of children worldwide receive such a diagnosis. This is backed up by data published by the National Centre for Health Statistics (NCHS) in 2015, where they estimate that about 15 % of children in the United States are affected by NDDs (12). The estimated prevalence is as follows: ADHD = 7.9–9.5% (13, 14), ASD = 0.7–2.2% (14-16), and learning disabilities fluctuating between 1.2–24% (17, 18). However, it is estimated that the prevalence is higher than this. According to a systematic literature review conducted by Francés et al 2022 (19), the failure to include a direct assessment of NDDs in studies concerned with the positively skewed their results. Concerning ID alone, the prevalence is 10.37/1000 population (95%CI 9.55–11.18 per 1000 population), according to a meta-analysis by Maulik et.al (20). Additionally, they reported this prevalence was higher in male (in both adults and children/adolescents with a male-to-female ratio of 0.7 -0.9 in adults and 0.4-1.0 in children/adolescents), as well as in low-income countries. The prevalence of the latter is attributed to less access to prenatal care, birth-related infections, higher environmental pollution and vitamin/mineral deficiencies (e.g. zinc) caused by poor diet.

Historically, the association between environmental/chemical causes and brain development was first hypothesised by evidence linking lead exposure to altered brain development (21). Later, it expanded to other heavy metals and pesticides. More data subsequently emerged linking industrial chemical exposure to NDDs. Some epidemiological data suggests that manganese, fluoride, chlorpyrifos, dichlorodiphenyltrichloroethane,

tetrachloroethylene, and PBDEs[2] are associated with diminished intellectual functioning, learning disabilities, attention problems, aggressiveness, hyperactivity, ADHD and ASD (22). Despite the amount of data available on this subject, some gaps remain. This is pointed out by Bellinger (23), who suggests that estimating the dose-effect relationship between chemical exposure and disease is essential to establish causality accurately.

Despite these factors' contribution to the increased prevalence of NDDs in general and ID specifically, the aetiology of such disorders is complex and not fully understood. Indeed, although significant, environmental exposure cannot fully explain the aetiology of NDDs. Recent epidemiological research indicates that the hereditary component of NDDs may also be a significant factor. For example, studies assessing the heritability of normal general cognitive abilities in adults (24, 25) demonstrate that it ranges from 66% to 68%. These results suggest that genetic factors heavily influence cognition and might therefore explain undiagnosed cases of intellectual disability and other NDDs. To that end, (26) conducted a population-based cohort study in Sweden to establish the Familial risk and heritability risk of ID. They found that ID heritability is at 95%, significantly surpassing previous studies of cognition mentioned earlier. Lichtenstein et al. also found that this heritability is higher in males, with a higher relative risk for maternal half-siblings. These findings suggest that sex chromosomes' contribution to intellect cognition and brain development is crucial and that mitochondrial factors might also be an important area to focus on in future research.

Multiple approaches are used to explore the genetic causes of NDDs, such as the "forward genetic approach" (Figure 2) which moves from phenotype to genotype, to gene. This "forward genetic approach" is in contrast to the reverse genetics one, which begins with an identified gene and then studies its different expression on phenotypes (27).

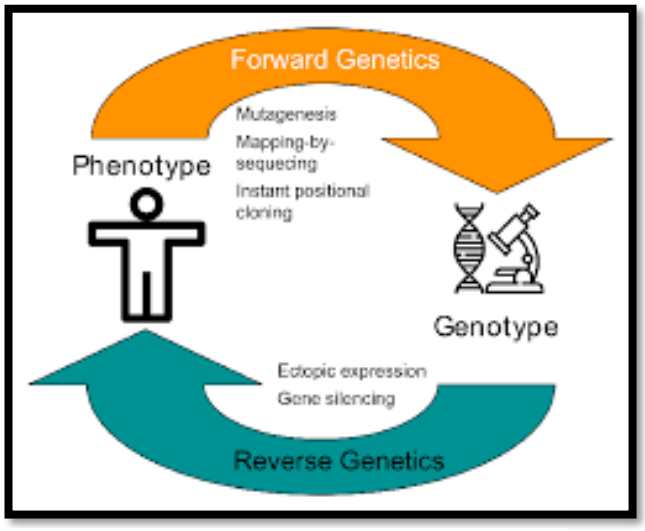

Figure 2: Forward Genetics Approach representation. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8634052/figure/f1/)

"Forward Genetics" is normally achieved via introducing random mutations and screening for the desired phenotype, which can then be mapped back to identify the causative mutation (28). The benefit of this "forward genetics" is that it offers an unbiased estimation, given the randomness of the mutagenesis. However, this approach also requires careful genetic manipulation, which can be strenuous.

[2] Polybrominated diphenyl ethers: a group of man-made chemicals used as flame retardants in various products (https://www.gov.uk/government/organisations/environment-agency) .

The above advancements in approaches to explore the genetic causes of NDDs has led to a differentiation between complex NDDs, where many mutated genes contribute to the phenotype and monogenic NDDs, where a single gene was identified to be the causal gene.

Complex NDDs include autism, ADHD, communication disorders, and some forms of intellectual disability. These complex NDDs have shown a moderate-to-high heritability pattern. According to Posthuma et.al (29), several twin studies reported that NDDs, especially ADHD and ASD showed significant heritability, while environmental factors played a more minor role. Additionally, NDDs show comorbidity with other mental disorders such as obsessive-compulsive disorder (OCD), anxiety disorders, and disruptive behaviour disorders (DBD) (30). Many studies further suggest shared molecular pathways as a causal factor. This finding will impact gene discovery by using a multivariate approach, for instance by using data from people diagnosed with mental disorders to infer genetic information on NDDs, instead of focusing on a single disorder and vice versa.

Monogenic forms of NDDs (particularly intellectual disability) have been identified and found to include genes involved in synaptic function, chromatin remodelling, and transcription regulation (31-33). According to Cardoso et.al (34), many genes not previously associated with NDDs (GRM7, STX1A, CCAR2, EEF1D, GALNT2, SLC44A1, LRRIQ3, AMZ2, CLMN, SEC23IP, INIP, NARG2, FAM234B, and TRAP1) have been identified through developing sequencing technology, such as whole exome (WES) and whole genome sequencing (WGS). Additionally, assessing familial NDDs was useful in studying the contribution of non-genetic factors (i.e. environmental factors) to the pathogenesis of such diseases in the presence of a shared genetic background (2). Moreover, researching inherited NDDs found that gene vulnerability and mutational burden are the two factors that determine the phenotypic outcome (2). Mutational load is defined as the “genetic burden given by the total number of disruptive mutations”, while gene vulnerability is the capability of a given gene to tolerate potentially disruptive mutations. Put differently, the lower the gene’s tolerance to a mutation, the higher the gene’s vulnerability. Therefore, the higher these mentioned factors, the more complex/severe the phenotype.

Discovering genes required a thorough approach that relied on multiple factors, such as technological advances, experienced personnel, and funding. With the introduction of NGS (Figure 3), there has been a marked increase in the amount of data produced and a dramatic decrease in the cost. For example, at the time of the introduction of Next Generation Sequencing (NGS) in 2005, the cost of sequencing 1 human genome was about $10,000,000. This figure began falling rapidly, reaching an all-time low of less than $1000 (35)-(Figure 4). A list of sequencing platforms and some descriptions of each can be found in Table 2. Continuous advances are made in sequencing, with platform updates and new concepts being introduced, such as long-read sequencing or “Third Generating Sequencing” devices. Their main advantage is portability,

which could be useful in resource-scarce areas with limited infrastructure. This was the case during the Ebola virus outbreak in Guinea, where portable WGS was used (MinION) to sequence and detect the virus (36) bypassing logistical hurdles.

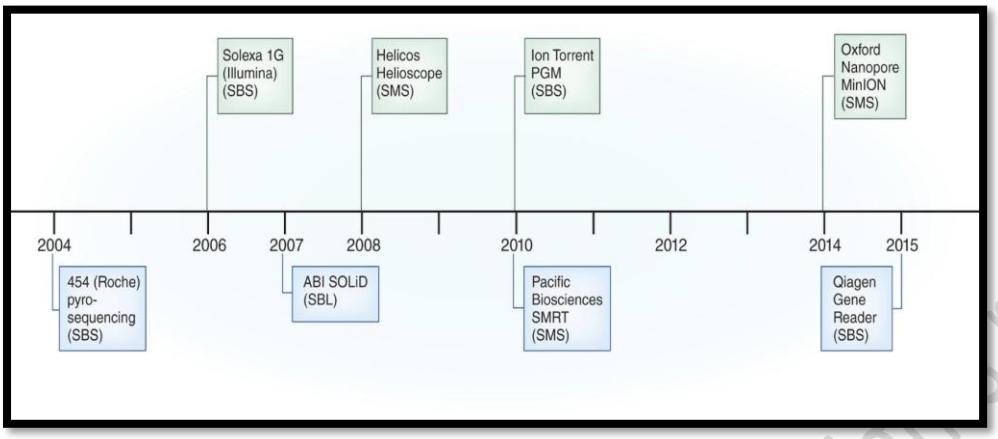

Figure 3: Next Generation Sequencing between 2004-2015(37)

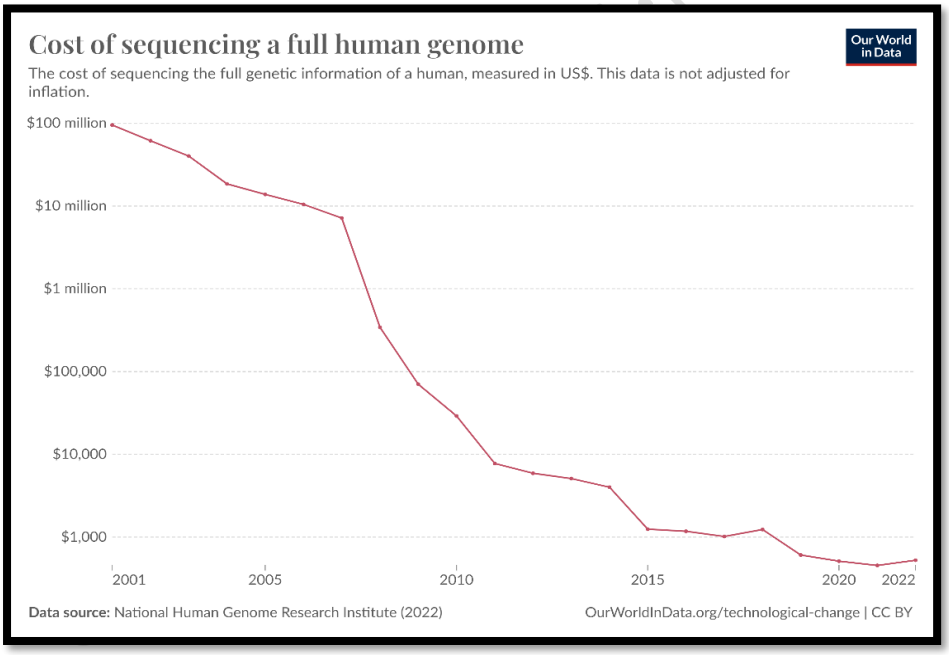

Figure 4:Cost of sequencing a human genome over the years (Institute, 2022).

Platform	Use	Principle	Read Length (bp)	Limitation

454 pyrosequencing	Short-read Sequencing	Detection of pyrophosphate released during nucleotide incorporation. (38-40)	400-1000	Error rate from insertion and deletion.
Ion Torrent	Short-read Sequencing	directly converts nucleotide sequence into digital information on a semiconductor chip(39-41)	200-400	Might have a loss of signal length due to homopolymer.
Illumina	Short-read Sequencing	Solid-phase sequencing on immobilized surface(39, 40)	36-300	Overlapping signal in overloaded samples.
Oxford Nanopore Technology (ONT)	Long-read Sequencing	Protein Nanopore in an electrically resistant polymer membrane through which characteristic current changes occur as each nucleotide passes through the detector(39, 40, 42)	10000-30000	High error rate of up to 15 %.

Table 2: Sequencing platform description (39, 40).

These advancements led to multiple global genomic projects focused on identifying, annotating, and understanding the genetic information encoded in our DNA. Such initiatives include the Human Genome Project, the International HapMap Project, the UK Biobank, & the 100,000 Genome Project, to name a few. Each project had its aim and objectives but shared a similar goal of better understanding genetic pieces of information. For example, the HapMap aimed to "determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain" (43). The UK Biobank, which is a continuous long-term study, investigates the role of genetic predispositions and environmental effects on different diseases (44) by making the data accessible to qualified researchers.

Launched by the UK government, the 100,000 genome project was one of the initiatives aimed at implementing WGS in investigating rare diseases, cancer, and infectious diseases (45). The project was run by Genomics England, a company that is fully owned by the UK's "Department of Health and Social Care," with Sir Mark Caulfield (Vice-Principal for Health at Queen Mary, University of London & Warden of Bart's and The

London School of Medicine and Dentistry) acting as chief scientist. The 100,000 genome project aimed at curating a database of 100,000 whole genome sequences permanently linked to patient's health records, allowing researchers to access such data for medical research and discoveries. A preliminary report by Smedley et al (46) examining the effect of WGS on diagnosing rare diseases in the NHS found that WGS was able to increase the diagnostic yield of patients with different disorders (Mendelian and complex disorders). This shows the importance of including genomic sequencing in routine public health protocols, with the ability to revolutionize scientific research and medicine and by extension, benefitting patients, their families, and the public health system. In the NDD case, the introduction of genomic screening proved to be influential in minimising costs and maximising diagnostic yields. According to Monroe et.al (47), while examining the cost-effectiveness (CE) of WES on the diagnosis of children with intellectual disability, they reported a diagnostic yield that reached 50 % in some cohorts–significantly above traditional methods that rely on observations and metabolic markers. Additionally, WGS can reduce the cost associated with genetic testing. Runheim et.al (48) found that when using WGS as a first-line diagnostic method, the costs were lower by $2339 against the standard of care, all while increasing the diagnostic yield by 23%.

Despite extensive work on deciphering the genetics of NDDs, some gaps persist, which will be elaborated on in the following literature review. The dissertation will then use data from the 100,000 genome project by examining patients in that programme with a molecular diagnosis of NDDs with and without ID to establish a correlation between genotypes and certain phenotypes, alongside any co-occurrence with other conditions. Furthermore, we will investigate gene variants and their phenotypes in specific gene panels to examine which variant features (zygosity, molecular consequence, protein domain) may be correlated with phenotypic severity. This can help us predict the effect of new variants discovered and allow for a better understanding of NDDs, which will help inform treatment options and, consequently, improve the quality of life of patients and their guardians.

Literature Review:

1- What resources do we have that are contributing to this research?

Initiatives that were established to contribute to a better understanding of NDDs include the Deciphering Developmental Disorders (DDD) study. DDD is a project that combines the clinical expertise of the United Kingdom's National Health Service (NHS) with the scientific research and bioinformatics power of the Wellcome Trust Sanger Institute (49). DDD has collected the genomic and phenotypic information of over 14000 trios: children with severe undiagnosed developmental disorders and their parents. The study uses the *Human Phenotype Ontology (HPO)*, a flagship component of the Monarch Initiative (50), which "provides a standardised vocabulary of phenotypic abnormalities encountered in human disease"(50). The Monarch Initiative currently contains over 13,000 terms and 156,000 annotations to hereditary diseases, which facilitates analysis by linking genomic and phenotypic information. Because most cases of NDD are seldom diagnosed and occur sporadically in children with unaffected parents, NDDs have typically been attributed to de novo mutation originating in the germline (51). Projects such as the Deciphering Developmental Disorders study and the Monarch Initiative are therefore needed to shorten the time of diagnosis, known as the " Diagnostic Odyssey" (52). Together, these projects contribute to uncovering genotype-phenotype associations. This will facilitate the further discovery of genes linked to NDDs and increase the diagnostic yield in the process.

Another hurdle that scientists face in deciphering the genetics of a particular disease is the lack of a consensus on causal mechanisms when assessing genotype-phenotype correlations. (53) states that curating a list of genes that contains the level of evidence available for disease causation can aid in prioritising and ranking variants, which will help in diagnosis by identifying causal variants for reporting. For this reason, researchers at Genomics England sought to create a publicly available database that contains different virtual gene panels related to human disease known as PanelAPP (https://panelapp.genomicsengland.co.uk/). Currently, there are over 400 gene panels for different diseases focusing on rare diseases and cancer. For example, the intellectual disability panel contains over 2700 genes with different gene ratings (Green, Amber, Red) that correspond to evidence of gene-disease association, with a green rating denoting a high level of evidence. These panels facilitate research by aggregating information based on common features and by using direct links with each gene to other databases, such as OMIM (54) and Ensembl (55).

Autism spectrum disorder (ASD) are areas where NDDs were explored heavily to uncover their genetic basis. Many resources and databases were developed, such as the AutDB (http://www.mindspec.org/autdb.html), where genes linked to AD are publicly available for collection, annotation, and visualization (56). The content is entirely formulated using data from published scientific literature, with a focus on molecular mechanisms involved in the disease. Despite the many databases curated, the Simons Foundation Autism Research Initiative (SFARI) remains the largest database that focuses on ASD, with the SFARI-Gene database listing over 1200 genes associated with autism. It relies on a scoring system that assesses evidence strength by using a set of criteria, with over 200 genes having a score of 1, which is correlated with the highest available evidence.

According to Abrahams, Arking (57), the advantage of SFARI-GENE lies in its explicitly defined scoring criteria and its interactive mechanism, which allows researchers to input their genes, offer alternate scorings, or suggest modifications. Similarly to the Deciphering Developmental Disorders study and the Monarch Initiative, use of the SFARI-Gene database shows that universal agreement between research bodies is essential for accelerating scientific discoveries by using well-defined criteria for evidence presentation.

Other genomic resources include ClinGen, which is funded by the National Institutes of Health (NIH). ClinGen's key goals are to aggregate genomic knowledge by curating public resources and continuously evaluating and improving metrics used in evaluations (58). It follows the same principle as panels, with genes given an annotation ranging from definitive, strong, moderate, or limited– depending on the evidence available (59). ClinGen stands out by its partnership with ClinVar (60), which is a database on genetic variations and their impact on human health that is aimed at sharing and archiving data to enhance genetic information curation. The intellectual disability panel currently holds over 300 genes with pathogenic variants associated with intellectual disability and/or autism.

In addition to these databases, bioinformatic expertise and software are needed to facilitate the handling of such large datasets and enhance the analysis and interpretation process of the research results. For example, different software for variant prioritisation and interpretation were developed. Exomiser©, which uses VCF[3] files and phenotypic information in HPO terms to annotate, filter and prioritise likely causative variants (61). The Variant effect predictor (VEP) from Ensembl, as the name suggests, predicts the effect of certain variations on genes, transcripts, and protein sequence, as well as regulatory regions (62).

With these tools, we can translate genomic knowledge into the public health system by enhancing management and treatment discoveries for NDDs.

2- <u>What do we know about the genetic forms of NDD?</u>

A- Complex NDDs:

[3] Variant Calling format files.

As previously mentioned, NDDs is an umbrella term for many diseases. However, Intellectual disability is a distinguished marking criterion of many developmental disorders, with a formal diagnosis being made when the intellectual coefficient (IQ) is less than 70 (63). This was known as the Phenotype-First Approach, where obtaining medical history and conducting physical examinations were necessary to make a diagnosis (64). One limitation of the Phenotype-First Approach is its dependence on the clinician's expertise and the availability of certain tests and medical equipment, which can affect diagnostic yield and lead to worse outcomes. The introduction of NGS technologies has allowed for the development of a 'genotype-first approach' to identify NDDs. More causal genes for NDDs are now being identified, shifting the diagnosis from being purely clinical examination-based to relying more on molecular analysis. In "*The Genetics of Neurodevelopmental Disorders*" (65), Dr Kevin J Mitchell divides NDDs into complex and mendelian, depending on their genetic architecture. Complex or multifactorial NDDs manifest because of genetic and non-genetic causes, which might include environmental exposure to toxins and chemicals, hormones, and infections. The heritability of complex NDDs is therefore due to multiple genes of small effect size[4] since linkage studies have failed to determine a single causal gene of many forms of NDDs (66). Despite this complexity, research has shown that the heritability of diseases and their persistence in the population depends on 2 independent parameters: the number of causal alleles per individual, and the frequency of risk alleles. The fact that multiple variants contribute to specific diseases helps aggregate data for specific diseases into a liability score that explains these disorders (66). Mitchell's model proposes that many risk variants are present in a population and are segregated independently, which leads to a normal distribution of risk alleles burden. However, the fact that people with NDDs and ID specifically demonstrate reduced fitness, higher mortality, and lower-than-average fecundity (67) does not explain why genetic variants associated with such disorders persist in the population, given negative selection pressures. According to Mitchell, individuals with diseases present at the end of the risk allele burden distribution bypass a threshold. This explains why NDDs persist in a population and why their frequency is rising. Indeed, although individuals can have a certain number of risk alleles/variants that do not cross this threshold and therefore do not manifest the associated diseases, further offspring might be at higher risk. This is seen in consanguine marriages, where a higher prevalence of ID and other NDDs are present. Patients with schizophrenia,

[4] The contribution of the SNP to genetic variance of the trait

another complex disease, were also found at a higher percentage in families with a higher rate of cousin marriage (68), which further supports the risk alleles threshold hypothesis.

- *GWAS in complex NDD:*

As mentioned above, complex NDDs are caused by multiple disease-causing variants that could be present at different loci or genes, with their cumulative effect leading to the phenotype and clinical presentations. To uncover these genes, scientists can perform a Genome-Wide Association study (GWAS). GWAS aims to identify genotype-phenotype associations by testing for differences in the allele frequency of genetic variants between individuals who share a common ancestry, but present with different phenotypes (69). It scans the entire genome of individuals, looking for variations (most commonly SNPs), which can help identify disease-associated genes, known as genetic risk loci. The two major drawbacks of such a technique is that it requires a large sample size and can produce biased data. Because it requires a large sample size, GWAS is often conducted using data from large research cohorts, which may not represent the true populations. For example, the UK biobank recruits participants via volunteers, which results in participants being healthier and more educated than the general population (70), in addition to predominately being of a white/European ancestry.

In the case of ASD, multiple GWAS were performed to identify risk loci associated with the disease. For instance, CNV[5], such as deletions or duplications, were identified to be linked to autism by multiple studies (71-73) - 7q11.23, 15q11–13, 16p11.2, and 22q11.2 loci, and *NRXN1, CNTN4, NLGNs,* and *SHANK3* genes. Furthermore, Weiss, Arking (74) GWAS revealed SNP[6] between *SEMA5A* and *TAS2R1* genes on chromosome 5 to be significantly associated with autism (P value= of 2 × 10−7). This finding makes sense since *SEMA5A* has been routinely associated with autism, as reflected by reduced expression in the brains of autistic patients (74). Additionally, Smoller, Craddock (75) identified through a GWAS encompassing data from 5 different psychiatric disorders including ASD, attention deficit–hyperactivity disorder, bipolar disorder, major depressive disorder, and

[5] Copy Number Variant.
[6] Single Nucleotide Polymorphism. A point mutation in a specific genetic region.

schizophrenia SNPs at four loci (2 on chromosomes 3p21 and 10q24), and SNPs in *CACNA1C* and *CACNB2*. These results reaffirm the existence of shared phenotypes among different NDDs and help explain why NDDs usually occur simultaneously. This overlap was also demonstrated in a meta-analysis conducted by Anney, Ripke (76), where genes/loci implicated in ASD (*FOXP1 ATP2B2*, and regions on chromosome 8p11.23) showed significant overlay with schizophrenia. GWAS were also able to uncover an association between structural changes in the nervous system and NDDs. Moreover, Xia, Zhang (77) concluded that common genetic variance influencing infant brain volumes has a clear link to mental illness and intellectual disability. This was further elaborated on by Bulayeva, Lesch (78), where several variants that predict structural changes were found to be in correlation with ID. For example, *KCNJ2*, which is important in the development of cleft lip and palate and *KCNJ16* and *USH2A*, which is associated with sensory neural hearing loss, were found to be implicated in ID.

B- Mendelian NDDs: Intellectual Disability

Although ID aetiology is multifactorial, ranging from genetic, non-genetic, and environmental causes, almost 50% of it is attributed to genetics (79). Genetic causes of ID include chromosomal aberrations and single genes which can be X-linked, autosomal dominant, or recessive (80). For the sake of simplicity, the focus of this dissertation will be on single-gene causes.

Parenti et.al (2) mention that most cases of inherited NDD are caused by De Novo mutations (mutations arising in the offspring of unaffected parents), with fewer autosomal recessive cases. In addition, Fitzgerald, Gerety (81) state that up to half of children with developmental disorders with a suspected genetic form remain without a diagnosis. The reason for this high number of diagnostic odysseys might be attributed to the fact that NDDs present with a variable clinical manifestation that is very similar to other diseases. However, with technological advancements in genomics and the increase in resources focused on NDDs, more children are being diagnosed and more genes contributing to the disease are being discovered.

- *Genes:*

Many genes have been identified as the causative gene in intellectual disability, with varying severity, mode of inheritance, molecular pathway, and phenotypic severity.

For example, the autosomal recessive ID (ARID) genes occur in both syndromic and non-syndromic forms, meaning that ID is present with other conditions, such as motor seizures, mood disorders, motor skills impairment, vision problems, or no other abnormalities, respectively (80). Because they only affect intelligence, these non-syndromic genes are being studied to understand variants that affect intelligence (82). According to Jamra (83), there are presumably over 2500 ARID genes, with less than 600 genes (84) being identified, with the most frequent variants reported in *GALT, VPS13B, ASPM, SPG11, MUT, GLDC, CEP290, POLG, LAMA2,* and *SMPD1*. Additionally, the ID panel from PanelAPP shows over 1000 genes that have a Biallelic mode of inheritance with varying levels of evidence. Ropers (85) states that the preferred method to uncover ARID genes given their heterogeneity is by homozygosity mapping using consanguine families. Moreover, a recent paper by Schwartz, Louie (86) where they surveyed literature from 2010 to 2021, mentioned that around 21 new genes were added to the list of genes following an X-linked mode of inheritance that causes ID. Figure 5 illustrates the position of these new genes, highlighted in orange, on the X chromosome. These genes' mechanisms vary between transcriptional process with *NKAP*, *TFE3*, *RIPPLY1*, and *BCORL1,* mitochondrial function with *COX7B, ZFP92, and APOO,* and ubiquitination associated with *PJA1 and OTUD5.* Furthermore, a DDD study by Fitzgerald et al. (2015) mentions that the most frequent mutated genes: *ARID1B* associated with Coffin-Siris syndrome, *SCN2A* and *ANKRD11* in KBG

syndrome, *SATB2* for Glass syndrome, *SYNGAP1*, *DYRK1A*, *MED13L, STXBP1, CTNNB1, KCNQ2* and *KMT2A* with Wiedemann-Steiner syndrome, *FOXP1* and PACS1 's Schuurs-Hoeijmakers syndrome, *SMARCA2* responsible for Nicolaides-Baraitser disease and *WDR45*. Except *WDR45*, all of these frequently mutated genes are autosomal dominant. Over 1000 AD genes have been discovered, with more added each year. The continued discovery of AD genes is the more striking when considering that the DDD stated in 2015 that AD ID gene discovery was almost complete. According to Wieczorek (87), 13 new genes were added in 2018 for autosomal dominant ID. All of these findings highlight the need for more funding and resource allocation for the identification of all possible genes that contribute to ID and other NDDs.

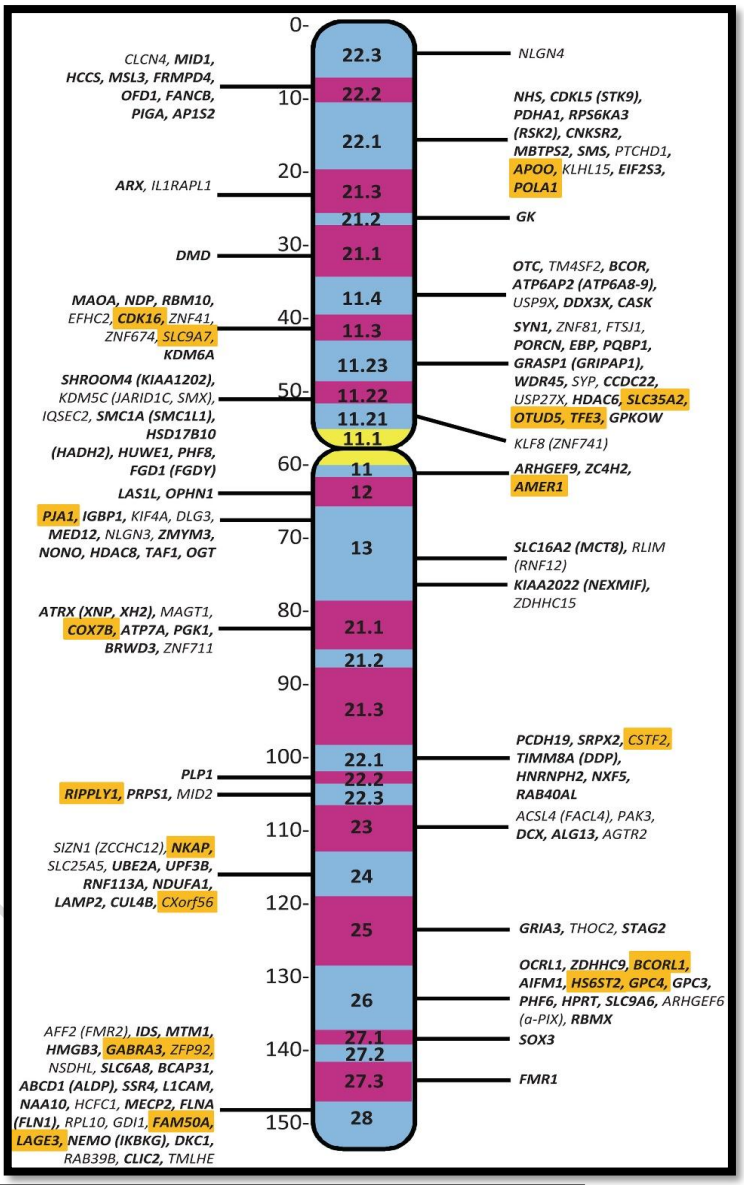

Figure 5: Gene's locations of the X chromosome associated with ID

[1]American J of Med Genetics Pt A, Volume: 191, Issue: 1, Pages: 144-159, First published: 27 October 2022, DOI: (10.1002/ajmg.a.63008.
[1]Permission to use is granted by Copyright Clearance Center's RightsLink®.

- *Molecular pathways, variations and phenotypic severity:*

The molecular functions of genes related to inherited intellectual disability are diverse, ranging from protein binding, chromatin remodelling, DNA and RNA regulatory function, and ion transport (88). For example, a search using the Ensembl database shows that *ARID1B*, *KMT2A*, *FOXP1*, SMARCA2, *SATB2*, and *CTNNB1* are involved in DNA binding activity, whereas SMARCA2, *SATB2*, *KMT2A*, *DYRK1A*, *CTNNB1* affect histone activity and *SATB2*, *FOXP1*, *CTNNB1*, SMARCA2, *ARID1B*, *DYRK1A*, *MED13L* influencing transcription. These shared molecular functions between genes indicate that certain pathways carry a superior significance in the development of diseases and phenotypes. These

shared molecular functions could also positively affect treatment options, by developing therapies targeting those common pathways, which would offer patients diagnosed with ID but under different causal genes the same therapy, which facilitates treatment and reduces costs.

In addition to having shared molecular functions, many of these genes are involved in biological functions outside of the nervous system. For example, *SATB2*, *CTNNB1*, and *FOXP1* have a role in osteoblast and osteoclast differentiation and development (89), and *CTNNB1*, *FOXP1*, and *KMT2A* in T-cell differentiation (90-92). Understanding the biological process implicated with certain genes therefore explains the phenotypes that occur in parallel with intellectual disability, since a mutated gene will have negative effects on other systems. For example, *ANKRD11* (ankyrin repeat domain containing 11- HGNC:21316) is responsible for KBG syndrome (OMIM) and 16q24.3 microdeletion syndrome (Orphanet). *ANKRD11* is mainly involved in protein binding at a biological level. Molecular pathways include odontogenesis of dentin-containing teeth (93) and skeletal system morphogenesis and face morphogenesis, which explains why KGB syndrome manifests with macrodontia, different facial features such as prominent ears, prominent nasal bridge and triangular face, and developmental delays (94). Furthermore, according to ClinVar, certain variants of *ANKRD11* phenotypes include astigmatism, heart failure, neutropenia and schizophrenia. This signals that *ANKRD11* might play an important role in cardiac remodelling and immune system regulation. For example, Devriendt, Holvoet (95) describe a mother/daughter KGB patient with KGB syndrome presenting with recurrent respiratory infections and a ventricular septal defect. Another example would be *SCN2A* (sodium voltage-gated channel alpha subunit HGNC:10588), which is a causative gene for DRAVET syndromes, which was first called severe myoclonic epilepsy of infancy (SMEI), with main symptoms of severe epilepsy that starts in the first year of life in the form of generalized or unilateral febrile-clonic seizures (96). *SCN2A*'s biological function is mainly involved in ion transport, and epilepsy is mainly caused by neurons firing as a result of abnormal excitation of neurons which is influenced by ion transport channels (97). This knowledge is crucial in disease management, since identifying the primary cause of epilepsy, via gene function study, enables us to administer the appropriate treatment such as an anti-epileptic that selectively blocks the ion channel.

Disease severity further varies with the genes or variants implicated as the cause of intellectual disability. For instance, Gilissen, Hehir-Kwa (98) compared the phenotypes of specific ID genes and highlighted that certain known and candidate ID genes have a more severe manifestation. Specifically, they reported that *SHANK3, VPS13B, SMC1A, SPTAN1, MECP2, MED13L, KCNA1, TENM3,*

and *SMC1A* were associated with a more severe form of ID, whereas, *POGZ*, *STAG1* and *SATB2* contribute to moderate disease (98). These findings could be tailored to the variants present in these genes and their penetrance. For example, Urpa, Kurki (99), while researching the additive effect of rare and common variants on ID, reported that the burden of rare loss of function (LOF) and damaging missense variants in the Finnish population was enriched with cases of severe ID with dysmorphic feature and sensory abnormality. Similarly, de novo damaging variants in known ID genes were found to be more strongly correlated with a sensory disability or epilepsy. Another example is the *IQSEC2* gene, where a splicing variant produces different grades of disease severity that are correlated with the transcript ratio of aberrant and wild-type (100). This information indicates that phenotypic severity can be predicted based on certain criteria of genes and variants, which could help in disease management. However, these predictions require a complex bioinformatics pipeline and specific software to process such a large amount of data.

3- NDDs: comorbidities, and genotype-phenotype correlations.

Historically, only a few NDDs, such as epilepsy, were described as occurring with movement disorders, intellectual affliction, and psychological conditions (101). However, a greater recognition of NDDs in recent years has increased our understanding of their “comorbidities” and “co-occurrences”. Indeed, NDDs rarely occur as a standalone disease, with more than one disorder typically being present–especially when intellectual disability is the diagnosed disease. To elaborate further, an explanation of the distinction between disease and disorder is necessary. NDDs are referred to as disorders, whereas a disease is a distinct entity with an identifiable cause (102). Comorbidities or cooccurrences might therefore refer to multiple disorders with a shared diagnostic grouping. For example, ASD with ADHD could have a homotypic relationship, meaning that a disorder predicts the occurrence of another disorder from the same overall group. On the other hand, heterotypic NDDs will occur with another disorder from a different group such as anxiety or depression (103). Another possible co-occurrence area is when NDDs, such as intellectual disability, manifest with disorders affecting other systems, such as the cardiovascular system and the ophthalmological system.

A- Comorbidities with NDDs:

Comorbidity among NDDs is widely described in both clinical and epidemiological data (104), despite the DSM-5 criteria segregating them into different entities. For instance, Dewey (102), while reviewing the importance of studying comorbidities with NDDs, found 50 % of patients diagnosed with ADHD exhibit movement disorders, including tics. Additionally, Davis and Kollins (105) reported that 30 to 50% of patients diagnosed with ASD display symptoms indicative of ADHD. This overlap of conditions indicates that these disorders share similarities in their molecular pathways, which could be the result of the same genetic factor's action. According to Parenti, Rabaneda (2), protein synthesis transcriptional/epigenetic regulation and synaptic signalling are common pathways between NDDs. Moreover, NDDs also manifest in a heterotypic form. For example, Geller, Biederman (106) noted that 30% of children and adolescents with obsessive-compulsive disorder meet the diagnostic criteria of ADHD, while Walitza, Zellmann (107) state that patients with early-onset OCD[7] that have ADHD as a comorbidity, present with more severe symptoms. Furthermore, Matson and Williams (108) while assessing depression and ASD, mention that one study by Mazzone, Postorino (109) found depression to be more severe with ASD, while Rosenberg, Kaufmann (110) reported that ASD occurred with increased rates of anxiety, depression, bipolar disorder, and ADHD. All these findings are indicative that there is a shared genetic factor between disorders. Studying one disease can therefore accelerate our understanding of other disorders and diseases.

Another example of comorbidities associated with NDDs includes afflictions that target other bodily functions and systems. In complex NDDs such as ADHD and ASD, there is no clear co-occurrence with other manifestations. However, studies exist that link certain NDDs to comorbid conditions, such as type 2 diabetes and cardiovascular disease. For example, Cortese, Moreira-Maia (111) conducted a systematic literature review on ADHD and obesity and found that the obesity prevalence increased by 70% in adults and by 40% in children with ADHD compared to the ADHD-free group. The robustness of this study was demonstrated through sensitivity analysis, which accounted for multiple covariates that might influence the results, such as gender, country, socioeconomic status and age, and medication intake. The association between ADHD and obesity can thus be explained by shared

[7] Obsessive Compulsive disorder. A mental health condition that causes people to have unwanted, repetitive thoughts and feelings, and to perform certain actions repeatedly to relieve the distress. Reference Link

genetic factors, which might influence some common molecular/biological pathways or behavioural afflictions.

Inherited form of NDDs such as Intellectual disability show a different pattern. For example, *SHANK3* causes Phelan-McDermid syndrome, which is characterised by global developmental delay, absence of severely delayed speech, and an autistic-like effect (112). There have also been multiple cases where pathogenic variants in *SHANK3* manifest with cardiovascular abnormalities. For example, Esmel-Vilomara, Dougherty-De Miguel (113) conducted a prospective evaluation involving patients with ID carrying *SHANK3* de novo variant, g.51153476G > A , g.51159361del, g.51159685_ 51159686del, g.51160126C > T, and 22q13.33 (50763701–51219009)x1, and identified cardiac abnormalities in all cases. *SHANK3*'s involvement in cardiovascular abnormalities could be attributed to its interaction with proteins, such as phospholipase Cβ1b (PLCβ1b), calcium-calmodulin-dependent protein kinase II (CaMKII), Homer, and α-actinin, as emphasised by Kim, Ko (114). On the other hand, *ANKRD11* (ankyrin repeat domain containing 11- HGNC:21316) is responsible for KBG syndrome (OMIM) and 16q24.3 microdeletion syndrome (Orphanet). According to ClinVar, certain variants of *ANKRD11* phenotypes include astigmatism, heart failure, neutropenia, and schizophrenia, which indicates that *ANKRD11* might be important in cardiac remodelling and immune system regulation. This is further evidenced by Devriendt, Holvoet (95), who show that a mother and daughter with KGB symptoms and KGB syndrome, respectively, exhibited recurrent respiratory infections and a ventricular septal defect. Applying these findings in clinical settings is crucial in the management of the disease by increasing awareness towards the possibility of finding abnormalities in other systems, which can drastically reduce mortality and improve quality of life.

B- Genotype-Phenotype correlations: Variant feature and phenotypes.

The above section explains many of the reasons why certain diseases co-occur with NDDs. However, this area of research is still relatively understudied. This dissertation's aim of expanding on the

genotype-phenotype association in NDDs is therefore a necessary step to achieve a better understanding of these disorders.

Clinical phenotypes have been hypothesised to be dependent on gene and variant features, such as variant types (de novo, missense, structural or frameshift) and its location. To elaborate on the case of comorbidity of NDDs with cardiac manifestation, different types of variation in *SHANK3* have been associated with different incidence rates of cardiovascular issues (113). According to Kim, Ko (115), the evidence of *SHANK3*'s involvement in cardiac function is mediated by its interactions with PLCβ1b, CaMKII, Homer, and α-actinin protein, which explains the presence of cardiovascular (CV) abnormalities in patients presenting with ID under that gene's influence. However, this observation does not explain the varying rates of CV issues between patients, which could either be the result of other variables or might indicate that variant features might have a strong effect on comorbidities. Esmel-Vilomara et al. (2024) observed while assessing *de novo* mutations in 5 patients (two frameshift mutations, one nonsense mutation, one splicing mutation, and a complete deletion) that patients harbouring the deletion type had a 13% risk of developing cardiac abnormalities, while those carrying the missense variant carry a 7% risk. More research is needed to find out if this difference in risk is explained by *SHANK3* expressing different transcripts in heart tissues compared to brain tissues, or that certain variants affect only certain transcripts.

The location of the variant in the gene also seems to influence phenotypes. For example, the *DHX30* gene is associated with NDDs and is a member of DExH-box helicases, which use ATP hydrolysis to unwind RNA. *DHX30* contains a highly conserved helicase core with 2 domains, which have 8 helicase core motifs (HCM) that play a role in either RNA binding or ATP binding and hydrolysis (116). *DHX30* is also reported to be one of the genes that are least tolerant to missense variation (117). Mannucci, Dang (118) studied the genotype-phenotype correlation in *DHX30* and found different gene activity depending on the coordinates of the variant. Specifically, they conducted an in-vitro experiment showcasing missense variants (p.(Gly462Glu), p.(Arg725His), p.(Ala734Asp), p.(Ser737Phe), p.(Thr739Ala), p.(Arg782Gln), and p.(Arg908Gln)) in one of the HCM with only p.(Arg908Gln)) being outside of HCM, and found that all missense variants within the HCM did not unwind the RNA, whereas the p.(Arg908Gln) acted similarly to *DHX30*- wild type. In addition to the location of the variant, the type of variant showed heterogeneity in disease severity between individuals. Mannucci, Dang (118) were able to identify four patients with variants resulting in either haploinsufficiency or a truncated protein presented with a milder form of intellectual disability. Together, these findings

show that variation location in certain genes plays a crucial role in determining certain phenotypic outcomes via their ability to tolerate variation.

Important research has been conducted on gene constituents and variant effects via location precision, the protein domain it affects, and gene tolerance, all of which might offer a means to predict the manifestation of specific phenotypes in certain diseases. Achieving this, however, requires further investments in bioinformatics pipelines and labs more broadly.

Aims & Objectives:

This dissertation focuses on finding genotype-phenotype correlations in patients with neurodevelopmental disorders (NDDs) particularly intellectual disability by studying patients recruited in the 100,000 genomes project under different disease categories. This is achieved by examining the genetic data of patients diagnosed with NDDs (including those with and without intellectual disability), to investigate any correlations between specific genotypes and clinical manifestations. These correlations are further assessed by identifying correlations between genes, variant features, and variant location to phenotypes to observe any discrepancies. The analysis then attempts to determine the potential of these correlations to influence the severity of phenotypic outcomes in NDDs by checking for patterns of comorbidities occurring with intellectual disability. To assess phenotypic severity and comorbidities, we inspect HPO terms described for each patient. Then, we evaluate whether the identified genotype-phenotype correlations found align with or expand the existing knowledge of NDD-related genes in gene panels, such as the Intellectual Disability (ID) Panel from Genomics England. The insights that are generated from this research can aid in predicting the effect of new variants on disease phenotypes and contribute to more targeted therapeutic interventions and personalised treatment plans for patients with NDDs.

Methodology:

This study's research design follows a retrospective observational research strategy using secondary genomic data from the 100,000 Genomes Project of patients recruited under different disease categories with whom a causal variant has been identified. This project's data is managed by

Genomics England and includes whole-genome sequencing (WGS) data linked to patient health records, providing detailed phenotypic and genotypic information.

Data analysis was performed using **R programming**, to tidy and organise the genomic and phenotypic data using the "tidyverse" package. The analysis was done using both a quantitative and qualitative statistical analysis. Bioinformatics pipelines were implemented to facilitate and automate the filtering, analysis, and visualization of the correlations between, certain gene presences, genetic variants and features, and phenotypic expressions described by Human Phenotype Ontology (HPO) terms. ChatGPT was also used to facilitate data analysis (119).

The initial genomic dataset included patients with various genetic diseases such as cardiovascular, ophthalmologic, cancers and neurodevelopmental disorders. The selection process involved different steps to focus on genes that have over ten patients in the intellectual disability-filtered cohort. Using the genes obtained, I filtered the original dataset containing all the disease categories to include only the patients with these genes as the diagnosed ones. This strategy resulted in a dataset that included patients with ID and other diseases, and the associated genotypic and phenotypic data, suitable for the subsequent correlation analysis.

The first step of the genotypic and phenotypic correlation descriptive analysis involved investigating the occurrence of different diseases under the filtered genes list to see if these genes are responsible for other monogenic diseases outside of the nervous system and neurodevelopmental area. Deeper correlation analysis between the genes and phenotypes was explored using the same filtered dataset, by looking into the "Human Phenotype Ontology" (HPO) terms. These terms described the signs and symptoms under which these patients were presented, which provided a structured and uniform way to link clinical symptoms with genetic data. To enhance the illustration of the findings, I created a" for loop" script to automate the visualisation process for each gene, which generated visual representations of the diseases and HPO terms that co-occurred with each gene. To gain a comprehensive understanding of the broader biological systems affected by the presence of a particular gene or genetic variants, the HPO terms were mapped to their **ancestor terms**. This process helped identify higher-level biological functions and potential comorbidities by grouping related phenotypic traits accordingly. The mapping helped reveal and enhance the insights into comorbidities and co-occurrence of phenotypes, associated with intellectual disability and other neurodevelopmental disorders. A Fisher Exact test to assess the statistical significance of the

correlation was used. In addition, a logistic regression model was implemented to establish predictability between genes and specific phenotypes using a heatmap.

Additional analyses were conducted to evaluate specific variant features such as the location in the protein domain. This effect of the variants on the protein structure and function was analysed by mapping the variants to specific protein domains and correlating these with phenotypic severity. The goal was to predict the potential impact of newly discovered variants on protein function and clinical outcomes, particularly in terms of disease severity and comorbidities. In addition, I use data from the **PanelApp** (a tool by Genomics England) to validate if the genes are present in the intellectual disability panel, and/or if it is present in other panels in certain occasions. This step allowed for the validation of known gene-phenotype associations, while also highlighting potential discovery of new correlations with specific phenotypes or diseases.

- *Ethical Considerations*

This research was conducted per ethical standards, using anonymised secondary data from the 100,000 Genomes Project. Ethical approval was not required for this study, as it involved secondary data analysis. However, strict compliance with the General Data Protection Regulation (GDPR) and guidelines set forth by Genomics England were followed to ensure the confidentiality and privacy of patient data.

Results:

REMOVED FOR PRIVACY

Discussion:

REMOVED FOR PRIVACY

Limitations:

This study has multiple limitations. First, a small sample size was used for the analysis. While the dataset includes many patients (5,735), the subset used for genotype-phenotype correlations is relatively small.This has implications for the analysis that involves comparing patients on the gen level according to certain features, such as variant type and protein domains. The relatively small sample size also limits the statistical power of the analysis and the generalisability of findings. Furthermore, the sample only encompasses 22 genes, with a minimum of 10 patients per gene. This could exclude potentially relevant genetic information that may contribute to NDDs, but is underrepresented in the dataset, which may result in information loss due to omission. Despite the standardisation of phenotypes using HPO terms, the fact that they are based on the recruiting clinician's assessment of the patient might also introduce bias. For instance, a clinician could use a specific term to describe phenotypes, while another clinician might use a lower or upper term that describes the same phenotype. This potential discrepancy may introduce inconsistencies across patients, which can affect the robustness of the analysis. Another source of potential bias comes from recruitment and data set composition. The data used for the project was extracted from the 100,000 Genomes project patients, most of whom are of white/European ancestry. Additionally, certain socioeconomic background groups could not be fully represented given that the recruitment process require a certain level of scientific knowledge or community exposure to know about the study. This could induce the overrepresentation of certain populations or conditions that affect the

analysis, while underrepresenting other groups. This can drastically diminish its generalisability given the ongoing evidence that warrant the inclusion of genetics in managing diseases, should the information be used in the clinic. Additionally, there is an exclusion of environmental and epigenetic factors which have been documented to influence diseases, especially NDDs. These effect could alter phenotypes which might not reflect the true manifestation of a certain gene's presence. Nevertheless, despite these limitations, the study provides valuable insight into the complex genetic interplay of NDDs in general and ID in particular. It highlights several phenotypic outcomes that were not reported in the literature while providing supporting evidence to already established hypotheses. The analysis proposes an approach based on creating algorithmic models to better predict outcomes, although more research is needed in this area. More broadly, this study shows that not only genes, but the specific features of variants can drastically impact a disease's manifestation, which can inform advances in precision medicine.

Conclusion:

REMOVED FOR PRIVACY

References:

1. Caulfield M, Davies J, Dennys M, Elbahy L, Fowler T, Hill S, et al. National Genomic Research Library. figshare; 2020.
2. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. Trends in Neurosciences. 2020;43(8):608-21.
3. Berrios G, Porter R. A history of clinical psychiatry: the origin and history of psychiatric disorders. Uma história da psiquiatria clínica: a origem e a história dos transtornos psiquiátricos Vol 2: As psicoses funcionais2012. p. 677-.
4. Herron CR. Inclusive language matters: Recommendations for health care providers to address implicit bias and equitable health care. Journal of Maine Medical Center. 2021;3(2):9.
5. England G. Language and terminology 2022 [Available from: https://files.genomicsengland.co.uk/documents/Genomics-England-Language-Guide.pdf.
6. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. The Lancet Psychiatry. 2017;4(4):339-46.
7. American Psychiatric Association D, American Psychiatric Association D. Diagnostic and statistical manual of mental disorders: DSM-5: American psychiatric association Washington, DC; 2013.
8. Rommelse NN, Franke B, Geurts HM, Hartman CA, Buitelaar JK. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. European child & adolescent psychiatry. 2010;19:281-95.
9. Sokolova E, Oerlemans AM, Rommelse NN, Groot P, Hartman CA, Glennon JC, et al. A Causal and Mediation Analysis of the Comorbidity Between Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). Journal of Autism and Developmental Disorders. 2017;47(6):1595-604.
10. Powder KE. Quantitative Trait Loci (QTL) Mapping. In: Shi XM, editor. eQTL Analysis: Methods and Protocols. New York, NY: Springer US; 2020. p. 211-29.
11. Bourgeron T. What Do We Know about Early Onset Neurodevelopmental Disorders? In: Nikolich K, Hyman SE, editors. Translational Neuroscience: Toward New Therapies. Cambridge (MA): MIT Press

© 2015 Massachusetts Institute of Technology and the Frankfurt Institute for Advanced Studies.; 2015.
12. He Y, Cai B, Shin H-C, Beresovsky V, Parsons V, Irimata K, et al. The National Center for Health Statistics' 2015 and 2016 research and development surveys. 2020.
13. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. International journal of epidemiology. 2014;43(2):434-42.
14. Baio J. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveillance Summaries. 2018;67.
15. Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcín C, et al. Global prevalence of autism and other pervasive developmental disorders. Autism research. 2012;5(3):160-79.
16. Fombonne E. Epidemiology of Pervasive Developmental Disorders. Pediatric Research. 2009;65(6):591-8.
17. Al-Yagon M, Cavendish W, Cornoldi C, Fawcett AJ, Grünke M, Hung L-Y, et al. The proposed changes for DSM-5 for SLD and ADHD: International perspectives—Australia, Germany, Greece, India, Israel, Italy, Spain, Taiwan, United Kingdom, and United States. Journal of learning disabilities. 2013;46(1):58-72.
18. Taanila A, Yliherva A, Kaakinen M, Moilanen I, Ebeling H. An epidemiological study on Finnish school-aged children with learning difficulties and behavioural problems. International journal of circumpolar health. 2011;70(1):59-71.
19. Francés L, Quintero J, Fernández A, Ruiz A, Caules J, Fillon G, et al. Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: a systematic review in accordance with the PRISMA criteria. Child and Adolescent Psychiatry and Mental Health. 2022;16(1):27.

20. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: A meta-analysis of population-based studies. Research in Developmental Disabilities. 2011;32(2):419-36.
21. Baghurst P, Robertson E, McMichael A, Vimpani G, Wigg N, Roberts R. The Port Pirie Cohort Study: lead effects on pregnancy outcome and early childhood development. Neurotoxicology. 1987;8(3):395-401.
22. De Felice A, Ricceri L, Venerosi A, Chiarotti F, Calamandrei G. Multifactorial Origin of Neurodevelopmental Disorders: Approaches to Understanding Complex Etiologies. Toxics. 2015;3(1):89-129.
23. Bellinger DC. Interpreting epidemiologic studies of developmental neurotoxicity: conceptual and analytic issues. Neurotoxicology and teratology. 2009;31(5):267-74.
24. Haworth CM, Wright MJ, Luciano M, Martin NG, de Geus EJ, van Beijsterveldt CE, et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. Molecular psychiatry. 2010;15(11):1112-20.
25. Panizzon MS, Vuoksimaa E, Spoon KM, Jacobson KC, Lyons MJ, Franz CE, et al. Genetic and environmental influences on general cognitive ability: Is ga valid latent construct? Intelligence. 2014;43:65-76.
26. Lichtenstein P, Tideman M, Sullivan PF, Serlachius E, Larsson H, Kuja-Halkola R, et al. Familial risk and heritability of intellectual disability: a population-based cohort study in Sweden. Journal of Child Psychology and Psychiatry. 2022;63(9):1092-102.
27. Ali Khan A, Raess M, de Angelis MH. Moving forward with forward genetics: A summary of the INFRAFRONTIER Forward Genetics Panel Discussion. F1000Res. 2021;10:456.
28. Everson JL, Eberhart JK. Chapter Three - Gene-alcohol interactions in birth defects. In: Lipinski RJ, Krauss RS, editors. Current Topics in Developmental Biology. 152: Academic Press; 2023. p. 77-113.
29. Posthuma D, Polderman TJC. What have we learned from recent twin studies about the etiology of neurodevelopmental disorders? Current Opinion in Neurology. 2013;26(2):111-21.
30. Hansen BH, Oerbeck B, Skirbekk B, Petrovski BÉ, Kristensen H. Neurodevelopmental disorders: prevalence and comorbidity in children referred to mental health services. Nordic Journal of Psychiatry. 2018;72(4):285-91.
31. Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell. 2011;146(2):247-61.
32. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209-15.
33. Iossifov I, O'roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216-21.
34. Cardoso AR, Lopes-Marques M, Silva RM, Serrano C, Amorim A, Prata MJ, et al. Essential genetic findings in neurodevelopmental disorders. Human Genomics. 2019;13(1):31.
35. National Human Genome Research Institute, "DNA Sequencing Costs" [original data] [Internet]. 2022 [cited Retrieved July 22, 2024]. Available from: https://ourworldindata.org/grapher/cost-of-sequencing-a-full-human-genome.
36. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.
37. Mardis ER. DNA sequencing technologies: 2006–2016. Nature Protocols. 2017;12(2):213-8.
38. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry. 1996;242(1):84-9.
39. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology. 2023;12(7):997.
40. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol. 2018;122(1):e59.
41. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348-52.
42. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome biology. 2016;17:1-11.
43. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The International HapMap Project. Nature. 2003;426(6968):789-96.

44. Biobank U. UK Biobank: Protocol for a large-scale prospective epidemiological resource www. ukbiobank. ac. uk/wp-content/uploads/2011/11. UK-Biobank-Protocol pdf. 2007.
45. Caulfield M, Davies J, Dennys M, Elbahy L, Fowler T, Hill S, et al. The 100,000 genomes project protocol. Genomics Engl P. 2017;3:1-112.
46. Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, Cipriani V, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. N Engl J Med. 2021;385(20):1868-80.
47. Monroe GR, Frederix GW, Savelberg SMC, de Vries TI, Duran KJ, van der Smagt JJ, et al. Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. Genetics in Medicine. 2016;18(9):949-56.
48. Runheim H, Pettersson M, Hammarsjö A, Nordgren A, Henriksson M, Lindstrand A, et al. The cost-effectiveness of whole genome sequencing in neurodevelopmental disorders. Scientific Reports. 2023;13(1):6904.
49. Firth HV, Wright CF. The Deciphering Developmental Disorders (DDD) study. Developmental Medicine and Child Neurology. 2011;53(8):702-3.
50. Gargano MA, Matentzoglu N, Coleman B, Addo-Lartey EB, Anagnostopoulos Anna V, Anderton J, et al. The Human Phenotype Ontology in 2024: phenotypes around the world. Nucleic Acids Research. 2023;52(D1):D1333-D46.
51. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet. 2015;385(9975):1305-14.
52. Miller D. The diagnostic odyssey: our family's story. Am J Hum Genet. 2021;108(2):217-8.
53. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nature Genetics. 2019;51(11):1560-5.
54. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Research. 2014;43(D1):D789-D98.
55. Harrison PW, Amode MR, Austine-Orimoloye O, Azov Andrey G, Barba M, Barnes I, et al. Ensembl 2024. Nucleic Acids Research. 2023;52(D1):D891-D9.
56. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. Nucleic Acids Res. 2009;37(Database issue):D832-6.
57. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Molecular Autism. 2013;4(1):36.
58. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen — The Clinical Genome Resource. New England Journal of Medicine. 2015;372(23):2235-42.
59. Leblond CS, Le T-L, Malesys S, Cliquet F, Tabet A-C, Delorme R, et al. Operative list of genes associated with autism and neurodevelopmental disorders based on database review. Molecular and Cellular Neuroscience. 2021;113:103623.
60. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980-5.
61. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nature Protocols. 2015;10(12):2004-15.
62. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016;17(1):122.
63. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nature Reviews Genetics. 2016;17(1):9-18.
64. Jansen S, Vissers LELM, de Vries BBA. The Genetics of Intellectual Disability. Brain Sciences. 2023;13(2):231.
65. Mitchell KJ. The Genetics of Neurodevelopmental Disorders. Newark, UNITED STATES: John Wiley & Sons, Incorporated; 2015.

66. Plomin R, Haworth CMA, Davis OSP. Common disorders are quantitative traits. Nature Reviews Genetics. 2009;10(12):872-8.
67. Keller MC, Miller G. Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? Behavioral and Brain Sciences. 2006;29(4):385-404.
68. Dobrusin M, Weitzman D, Levine J, Kremer I, Rietschel M, Maier W, et al. The rate of consanguineous marriages among parents of schizophrenic patients in the Arab Bedouin population in Southern Israel. The World Journal of Biological Psychiatry. 2009;10(4):334-6.
69. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021;1(1):59.
70. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. American journal of epidemiology. 2017;186(9):1026-34.
71. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein *SHANK3* are associated with autism spectrum disorders. Nature genetics. 2007;39(1):25-7.
72. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. 2009;459(7246):569-73.
73. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11. 23 Williams syndrome region, are strongly associated with autism. Neuron. 2011;70(5):863-85.
74. Weiss LA, Arking DE, Hopkins GDPoJ, c tACWgDMJajemcmhebCAaje, West JHUNgstADEBCWWKOCAHGTRL, 67 MOsPDMIEHMM-LKSJKIJ, et al. A genome-wide linkage and association scan reveals novel loci for autism. Nature. 2009;461(7265):802-8.
75. Smoller J, Craddock N, Kendler K, Lee P, Neale B, Nurnberger J, et al. Genetic Risk Outcome of Psychosis (GROUP) Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381(9875):1371-9.
76. Anney RJL, Ripke S, Anttila V, Grove J, Holmans P, Huang H, et al. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Molecular Autism. 2017;8(1):21.
77. Xia K, Zhang J, Ahn M, Jha S, Crowley JJ, Szatkiewicz J, et al. Genome-wide association analysis identifies common variants influencing infant brain volumes. Translational Psychiatry. 2017;7(8):e1188-e.
78. Bulayeva K, Lesch K-P, Bulayev O, Walsh C, Glatt S, Gurgenova F, et al. Genomic structural variants are linked with intellectual disability. Journal of Neural Transmission. 2015;122(9):1289-301.
79. Milani D, Ronzoni L, Esposito S. Genetic Advances in Intellectual Disability. J Pediatr Genet. 2015;04(03):125-7.
80. Ilyas M, Mir A, Efthymiou S, Houlden H. The genetics of intellectual disability: advancing technology and gene editing. F1000Res. 2020;9.
81. Fitzgerald TW, Gerety SS, Jones WD, van Kogelenberg M, King DA, McRae J, et al. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519(7542):223-8.
82. Hill WD, Davies G, Liewald D, Payton A, McNeil C, Whalley L, et al. Examining non-syndromic autosomal recessive intellectual disability (NS-ARID) genes for an enriched association with intelligence differences. Intelligence. 2016;54:80-9.
83. Jamra R. Genetics of autosomal recessive intellectual disability. medizinische genetik. 2018;30(3):323-7.
84. Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, et al. Genetics of intellectual disability in consanguineous families. Molecular Psychiatry. 2019;24(7):1027-39.
85. Ropers HH. Genetics of intellectual disability. Current Opinion in Genetics & Development. 2008;18(3):241-50.
86. Schwartz CE, Louie RJ, Toutain A, Skinner C, Friez MJ, Stevenson RE. X-Linked intellectual disability update 2022. American Journal of Medical Genetics Part A. 2023;191(1):144-59.
87. Wieczorek D. Autosomal dominant intellectual disability. Medizinische Genetik. 2018;30(3):318-22.

88. Srivastava AK, Schwartz CE. Intellectual disability and autism spectrum disorders: Causal genes and molecular mechanisms. Neuroscience & Biobehavioral Reviews. 2014;46:161-74.
89. Chan WCW, Tan Z, To MKT, Chan D. Regulation and Role of Transcription Factors in Osteogenesis. International Journal of Molecular Sciences. 2021;22(11):5445.
90. Xiao X, Mo H, Tu K. *CTNNB1* mutation suppresses infiltration of immune cells in hepatocellular carcinoma through miRNA-mediated regulation of chemokine expression. International Immunopharmacology. 2020;89:107043.
91. Kaminskiy Y, Kuznetsova V, Kudriaeva A, Zmievskaya E, Bulatov E. Neglected, yet significant role of *FOXP1* in T-cell quiescence, differentiation and exhaustion. Frontiers in Immunology. 2022;13.
92. Lamble AJ, Myers RM, Taraseviciute A, John S, Yates B, Steinberg SM, et al. *KMT2A* Rearrangements Are Associated with Lineage Switch Following CD19 Targeting CAR T-Cell Therapy. Blood. 2021;138:256.
93. Cobourne MT. The genetic control of early odontogenesis. Br J Orthod. 1999;26(1):21-8.
94. Morel Swols D, Foster J, Tekin M. KBG syndrome. Orphanet journal of rare diseases. 2017;12:1-7.
95. Devriendt K, Holvoet M, Fryns JP. Further delineation of the KBG syndrome. Genet Couns. 1998;9(3):191-4.
96. Dravet C. The core Dravet syndrome phenotype. Epilepsia. 2011;52(s2):3-9.
97. Lerche H, Jurkat-Rott K, Lehmann-Horn F. Ion channels and epilepsy. American Journal of Medical Genetics. 2001;106(2):146-59.
98. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344-7.
99. Urpa L, Kurki MI, Rahikkala E, Hämäläinen E, Salomaa V, Suvisaari J, et al. Evidence for the additivity of rare and common variant burden throughout the spectrum of intellectual disability. European Journal of Human Genetics. 2024;32(5):576-83.
100. Madrigal I, Alvarez-Mora MI, Rosell J, Rodríguez-Revenga L, Karlberg O, Sauer S, et al. A novel splicing mutation in the *IQSEC2* gene that modulates the phenotype severity in a family with intellectual disability. European Journal of Human Genetics. 2016;24(8):1117-23.
101. Johnson MR, Shorvon SD. Heredity in epilepsy: neurodevelopment, comorbidity, and the neurological trait. Epilepsy & Behavior. 2011;22(3):421-7.
102. Dewey D. What Is Comorbidity and Why Does It Matter in Neurodevelopmental Disorders? Current Developmental Disorders Reports. 2018;5(4):235-42.
103. Angold A, Costello EJ, Erkanli A. Comorbidity. The Journal of Child Psychology and Psychiatry and Allied Disciplines. 1999;40(1):57-87.
104. Cruddace S, Riddell P. Attention processes in children with movement difficulties, reading difficulties or both. Journal of Abnormal Child Psychology. 2006;34:672-80.
105. Davis NO, Kollins SH. Treatment for Co-Occurring Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder. Neurotherapeutics. 2012;9(3):518-30.
106. Geller DA, Biederman J, Faraone SV, Cradock K, Hagermoser L, Zaman N, et al. Attention-deficit/hyperactivity disorder in children and adolescents with obsessive-compulsive disorder: fact or artifact? Journal of the American Academy of Child & Adolescent Psychiatry. 2002;41(1):52-8.
107. Walitza S, Zellmann H, Irblich B, Lange K, Tucha O, Hemminger U, et al. Children and adolescents with obsessive-compulsive disorder and comorbid attention-deficit/hyperactivity disorder: preliminary results of a prospective follow-up study. Journal of Neural Transmission. 2008;115:187-90.
108. Matson JL, Williams LW. Depression and mood disorders among persons with Autism Spectrum Disorders. Research in Developmental Disabilities. 2014;35(9):2003-7.
109. Mazzone L, Postorino V, De Peppo L, Fatta L, Lucarelli V, Reale L, et al. Mood symptoms in children and adolescents with autism spectrum disorders. Research in developmental disabilities. 2013;34(11):3699-708.
110. Rosenberg RE, Kaufmann WE, Law JK, Law PA. Parent report of community psychiatric comorbid diagnoses in autism spectrum disorders. Autism Research and Treatment. 2011;2011(1):405849.

111. Cortese S, Moreira-Maia CR, Fleur DS, Morcillo-Peñalver C, Rohde LA, Faraone SV. Association Between ADHD and Obesity: A Systematic Review and Meta-Analysis. American Journal of Psychiatry. 2016;173(1):34-43.
112. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, et al. Contribution of *SHANK3* mutations to autism spectrum disorder. Am J Hum Genet. 2007;81(6):1289-97.
113. Esmel-Vilomara R, Dougherty-De Miguel L, Artigas-Baleri A, Turón-Viñas E, Cuscó I, Díaz-Gómez A, et al. Cardiovascular abnormalities in patients with *SHANK3* pathogenic variants: Beyond neurodevelopmental disorders and epilepsy. European Journal of Medical Genetics. 2024;71:104965.
114. Kim Y, Ko TH, Jin C, Zhang Y, Kang HR, Ma R, et al. The emerging roles of *SHANK3* in cardiac function and dysfunction. Frontiers in Cell and Developmental Biology. 2023;11:1191369.
115. Kim Y, Ko TH, Jin C, Zhang Y, Kang HR, Ma R, et al. The emerging roles of *SHANK3* in cardiac function and dysfunction. Frontiers in Cell and Developmental Biology. 2023;11.
116. Lessel D, Schob C, Küry S, Reijnders MRF, Harel T, Eldomery MK, et al. De Novo Missense Mutations in *DHX30* Impair Global Translation and Cause a Neurodevelopmental Disorder. The American Journal of Human Genetics. 2017;101(5):716-24.
117. Lessel D, Schob C, Küry S, Reijnders MR, Harel T, Eldomery MK, et al. De novo missense mutations in *DHX30* impair global translation and cause a neurodevelopmental disorder. The American Journal of Human Genetics. 2017;101(5):716-24.
118. Mannucci I, Dang NDP, Huber H, Murry JB, Abramson J, Althoff T, et al. Genotype–phenotype correlations and novel molecular insights into the *DHX30*-associated neurodevelopmental disorders. Genome Medicine. 2021;13(1):90.
119. OpenAI. ChatGPT (Mar 14 version) [Large language model]. 2024 [Available from: https://chat.openai.com/chat.
120. Kibalnyk Y, Afanasiev E, Noble RMN, Watson AES, Poverennaya I, Dittmann NL, et al. The chromatin regulator *ANKRD11* controls cardiac neural crest cell-mediated outflow tract remodeling and heart function. Nat Commun. 2024;15(1):4632.
121. Kibalnyk Y, Afanasiev E, Noble RMN, Watson AES, Poverennaya I, Dittmann NL, et al. The chromatin regulator *ANKRD11* controls cardiac neural crest cell-mediated outflow tract remodeling and heart function. Nature Communications. 2024;15(1):4632.
122. Rappold G, Siper P, Kostic A, Braden R, Morgan A, Koene S, et al. *FOXP1* Syndrome. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle

123. Gauthier J, Champagne N, Lafrenière RG, Xiong L, Spiegelman D, Brustein E, et al. De novo mutations in the gene encoding the synaptic scaffolding protein *SHANK3* in patients ascertained for schizophrenia. Proc Natl Acad Sci U S A. 2010;107(17):7863-8.
124. Awamleh Z, Choufani S, Cytrynbaum C, Alkuraya Fowzan S, Scherer S, Fernandes S, et al. *ANKRD11* pathogenic variants and 16q24.3 microdeletions share an altered DNA methylation signature in patients with KBG syndrome. Human Molecular Genetics. 2022;32(9):1429-38.
125. Sinha R, Dvorak M, Ganesan A, Kalesinskas L, Niemeyer CM, Flotho C, et al. Epigenetic Profiling of *PTPN11* Mutant JMML Hematopoietic Stem and Progenitor Cells Reveals an Aberrant Histone Landscape. Cancers. 2023;15(21):5204.
126. Braden RO, Amor DJ, Fisher SE, Mei C, Myers CT, Mefford H, et al. Severe speech impairment is a distinguishing feature of *FOXP1*-related disorder. Dev Med Child Neurol. 2021;63(12):1417-26.
127. Lippa NC, Barua S, Aggarwal V, Pereira E, Bain JM. A novel de novo *KDM5C* variant in a female with global developmental delay and ataxia: a case report. BMC Neurology. 2021;21(1):358.
128. Santos-Rebouças CB, Fintelman-Rodrigues N, Jensen LR, Kuss AW, Ribeiro MG, Campos M, et al. A novel nonsense mutation in *KDM5C*/JARID1C gene causing intellectual disability, short stature and speech delay. Neuroscience Letters. 2011;498(1):67-71.

129. Sanders SJ, Campbell AJ, Cottrell JR, Moller RS, Wagner FF, Auldridge AL, et al. Progress in Understanding and Treating *SCN2A*-Mediated Disorders. Trends Neurosci. 2018;41(7):442-56.
130. Pierpont EI, Pierpont ME, Mendelsohn NJ, Roberts AE, Tworog-Dube E, Seidenberg MS. Genotype differences in cognitive functioning in Noonan syndrome. Genes Brain Behav. 2009;8(3):275-82.
131. Carter DC, Kierzkowska O, Sarino K, Guo L, Marchi E, Lyon GJ. Ocular manifestations in a cohort of 43 patients with KBG syndrome. Am J Med Genet A. 2024;194(4):e63473.
132. Tørring PM, Larsen MJ, Brasch-Andersen C, Krogh LN, Kibæk M, Laulund L, et al. Is *MED13L*-related intellectual disability a recognizable syndrome? European Journal of Medical Genetics. 2019;62(2):129-36.
133. Angus SP, Nevins JR. A role for Mediator complex subunit *MED13L* in Rb/E2F-induced growth arrest. Oncogene. 2012;31(44):4709-17.
134. Iwagawa T, Kawabata R, Fukushima M, Kuribayashi H, Watanabe S. *SETD5*, but not Setd2, is indispensable for retinal cell survival and proliferation. FEBS Letters. 2023;597(3):427-36.
135. Cruciat C-M, Dolde C, De Groot RE, Ohkawara B, Reinhard C, Korswagen HC, et al. RNA helicase DDX3 is a regulatory subunit of casein kinase 1 in Wnt–β-catenin signaling. Science. 2013;339(6126):1436-41.
136. Grigoryan T, Wend P, Klaus A, Birchmeier W. Deciphering the function of canonical Wnt signals in development and disease: conditional loss-and gain-of-function mutations of β-catenin in mice. Genes & development. 2008;22(17):2308-41.
137. Chen C-Y, Chan C-H, Chen C-M, Tsai Y-S, Tsai T-Y, Wu Lee Y-H, et al. Targeted inactivation of murine *DDX3X*: essential roles of *DDX3X* in placentation and embryogenesis. Human Molecular Genetics. 2016;25(14):2905-22.
138. Liu Y, Patel L, Mills GB, Lu KH, Sood AK, Ding L, et al. Clinical significance of *CTNNB1* mutation and Wnt pathway activation in endometrioid endometrial carcinoma. J Natl Cancer Inst. 2014;106(9).
139. Kurnit KC, Kim GN, Fellman BM, Urbauer DL, Mills GB, Zhang W, et al. *CTNNB1* (beta-catenin) mutation identifies low grade, early stage endometrial cancer patients at increased risk of recurrence. Modern Pathology. 2017;30(7):1032-41.
140. Park M, Moon B, Kim J-H, Park S-J, Kim S-K, Park K, et al. Downregulation of *SETD5* Suppresses the Tumorigenicity of Hepatocellular Carcinoma Cells. Molecules and Cells. 2022;45(8):550-63.
141. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in *ADNP*. Nature Genetics. 2014;46(4):380-4.
142. Van Dijck A, Vulto-van Silfhout AT, Cappuyns E, van der Werf IM, Mancini GM, Tzschach A, et al. Clinical Presentation of a Complex Neurodevelopmental Disorder Caused by Mutations in *ADNP*. Biological Psychiatry. 2019;85(4):287-97.
143. Zhang Y, Zheng Y, Wang S, Fan Y, Ye Y, Jing Y, et al. Single-nucleus transcriptomics reveals a gatekeeper role for *FOXP1* in primate cardiac aging. Protein Cell. 2023;14(4):279-93.
144. Wang B, Weidenfeld J, Lu MM, Maika S, Kuziel WA, Morrisey EE, et al. *FOXP1* regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. Development. 2004;131(18):4477-87.
145. Liu X-M, Du S-L, Miao R, Wang L-F, Zhong J-C. Targeting the forkhead box protein P1 pathway as a novel therapeutic approach for cardiovascular diseases. Heart Failure Reviews. 2022;27(1):345-55.
146. Asadollahi R, Zweier M, Gogoll L, Schiffmann R, Sticht H, Steindl K, et al. Genotype-phenotype evaluation of *MED13L* defects in the light of a novel truncating and a recurrent missense mutation. European Journal of Medical Genetics. 2017;60(9):451-64.
147. Napoli C, Schiano C, Soricelli A. Increasing evidence of pathogenic role of the Mediator (MED) complex in the development of cardiovascular diseases. Biochimie. 2019;165:1-8.
148. Gardner RN, Oppman AM, Martins I, Grueter CE. Cardiomyocyte-specific deletion of Med13 & *MED13L* results in dysregulated gene expression and lethal systolic heart failure. The FASEB Journal. 2022;36(S1).
149. Cheung MY-Q, Roberts C, Scambler P, Stathopoulou A. *SETD5* is required in cardiopharyngeal mesoderm for heart development and its haploinsufficiency is associated with outflow tract defects in mouse. genesis. 2021;59(7-8):e23421.

150. Kauffman H, Ahrens-Nicklas RC, Calderon-Anyosa RJC, Ritter AL, Lin KY, Rossano JW, et al. Genotype–phenotype association by echocardiography offers incremental value in patients with Noonan Syndrome with Multiple Lentigines. Pediatric Research. 2021;90(2):444-51.
151. Liu S, Zhao Y, Mo H, Hua X, Chen X, Wang W, et al. Genetic variations in *PTPN11* lead to a recurrent left ventricular outflow tract obstruction phenotype in childhood hypertrophic cardiomyopathy. The Journal of Thoracic and Cardiovascular Surgery. 2025;169(1):196-207.e5.