의학박사 학위논문

Systematic Analysis of Genetic Etiology in Pediatric Neurodevelopmental Disorders

소아 신경발달 질환에 대한

유전적 병인의 체계적 분석

2023 년 2 월

서울대학교 대학원

의학과 중개의학전공

박 수 진

소아 신경발달 질환에 대한

유전적 병인의 체계적 분석

지도 교수 채 종 희

이 논문을 의학박사 학위논문으로 제출함

2022년 10월

서울대학교 대학원

의학과 중개의학전공

박 수 진

박수진의 의학박사 학위논문을 인준함

2023년 1월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Abstract

Systematic Analysis of Genetic Etiology in Pediatric Neurodevelopmental Disorders

Soojin Park
Translational Medicine
The Graduate School
Seoul National University

Neurodevelopmental disorders (NDDs) are caused by various alterations in gene function. Since a large number of genes are known to cause NDDs, a comprehensive understanding of genes and pathogenic variants is critical. However, a systematic understanding on the contribution of variants to Mendelian diseases is still insufficient. To overcome this limitation, whole exome sequencing (WES) analysis was conducted to characterize genetic variants of 1180 patients with neurological symptoms. The diagnostic yield for definitive pathogenic variant findings was 50.8%, after including increased diagnostic yield (5.9%) identified by the reanalysis. Among the diagnosed patients, 33.4% of them carried inherited variants. Gene ontology (GO) analysis revealed that autosomal recessive-inherited

genes were characterized by metabolic process, muscle organization and metal ion homeostasis pathways. The majority of autosomal recessive genes had the probability of being loss-of-function intolerant(pLI) of 0 and functional prediction scores for recessive variants tend to be lower than dominantly inherited variants. Transcriptome and interactome profiling revealed differences in the tissue-specific expression and protein-protein interaction by different inheritance patterns. Furthermore, the rate of carriers for recessive variants was predicted using gnomAD and Korean Variant Archive (KOVA) databases. The results show that genes responsible for NDDs harbor different molecular mechanisms and expression patterns by inheritance patterns, which may be associated with disease manifestation. Also, calculated frequency rate for recessive variants can be utilized to pre-screen rare severe neurodevelopmental disorder carriers.

*This thesis is based on published article; Park *et al*., 2022 *Front Genet*, 13, 99015.

Keywords: neurodevelopmental disorder, inheritance pattern, carrier prediction, whole exome sequencing, recessive disorders

---

Student number: 2016-38160

Table of Contents

List of Figures

List of Tables

Chapter 1

Whole Exome Sequencing Diagnosis and Re-analysis in Patients with Neurodevelopmental Disorders

1.1 Introduction

Neurodevelopmental disorders (NDDs) are a group of disorders that affect the development of the nervous system leading to abnormal brain function, which may affect motor function, learning ability, development, language, and other brain activities. NDDs are not only caused by aberrant brain development but also strong genetic causes and risk factors. So, affected individuals have various clinical symptoms and signs including intellectual disability (ID), global developmental delay (GDD), muscle weakness, epilepsy, and thereby challenging to diagnosis because of the extreme genetic heterogeneity and rare occurrence [1-7]. NDDs affect more than 3% of children worldwide but each individual NDDs are rare [8]. Furthermore, pediatric patients have the possibility of an evolving phenotype that might at some point alter, making its diagnosis more difficult. The conventional tests aren't comprehensive enough to pick up every diagnosis. So, diagnosis may still not be achieved if the causal gene or mechanism of action is missing from the analysis undertaken. Numerous patients had undergone multiple clinical evaluations and diagnostic tests over the years, with no definitive diagnosis.

Whole exome sequencing can help provide a molecular diagnosis especially for disorders that are genetically heterogeneous or patients with overlapping symptoms are difficult to diagnose. Whole exome sequencing as a primary clinical test provided a higher diagnostic yield than conventional genetic testing in a clinically

heterogeneous cohort. Due to the correlation of genetic disorders with mutations in protein-coding genes, the cheaper and quicker whole-exome sequencing (WES) is preferred as a diagnostic tool to the more informative whole-genome sequencing [9, 10]. The ACMG guidelines highlighted exome and genome sequencing that could be considered as first or second-tier in patients with congenital anomalies, developmental delay, or intellectual disability [11]. Many studies have found a high diagnostic yield of whole exome sequencing, having a diagnostic yield up to >40% in patients with NDDs, especially when both biological parents are considered [12-14]. To facilitate diagnosis and discovery of novel disease pathophysiology, large-scale systematic efforts have been conducted at regional or national scales [15-18]. Recent efforts into patient genome sequencing, diagnosis, and the discovery novel genes have enhanced the yield of NDDs molecular diagnosis in clinical practice. Lee *et al.* shows Notable cases where WES-based analysis conferred correct diagnoses or changed medical treatment strategies (Table1) [19]. So, patients who are searching for a diagnosis to explain symptoms are able to avoid unnecessary diagnostic tests by undergoing exome sequencing. Given the rapid pace of disease gene discovery, it is recommended to periodically reanalyze the exome data for individuals without a definitive diagnosis, because evidence suggests that doing so may increase the diagnostic yield by 10% or more [20, 21].

Our cohort is Korean patients who were left undiagnosed with complex neurodevelopmental disorders (KND cohort) Notably, the medical system in Korea

Table1. Notable cases where WES-based analysis provided correct diagnoses or changed medical treatment strategies

Initial clinical problem	Causal gene	Modified clinical interpretation (MIM number)	Significance of WES-based patient evaluation (treatment)	References
Developmental regression with Rett syndrome-like phenotype	*ST3GAL5*	Salt and pepper developmental regression syndrome (#609056)	Identified the molecular defect and established an accurate diagnosis	50,51
Hypotonia and motor delay followed by lower extremity weakness	*DYNC1H1*	Spinal muscular atrophy, lower extremity-predominant 1, AD (#158600)	Diagnosed a case with pleiotropic and evolving symptoms	52
Early onset hypotonia, sacral mass, congenital heart disease, and facial dysmorphism	*ASAH1*	Farber lipogranulomatosis (#228000)	Corrected a misdiagnosis	53
Ataxia followed by generalized dystonia	*ANO3*	Expanded spectrum of dystonia 24 (#615034)	Suggested a treatment strategy that resulted in gradual improvement within one year (deep brain stimulation)	54
Focal lower leg dystonia, dystonic gait	*SLC2A1*	GLUT1 deficiency syndrome 2 (#612126)	Identified disease-specific treatment that resulted in near-elimination of dystonia (ketogenic diet)	55
Leigh syndrome	*SLC19A3*	Thiamine metabolism dysfunction syndrome 2 (#606152)	Identified disease-specific treatment that resulted in clinical improvements in dystonia, spasticity, and cognitive function (supplements of thiamine and biotin)	56
Recurrent infections, telangiectatic skin mottling, and brain infarctions	*TMEM173*	STING-associated vasculopathy, infantile-onset (#615934)	Provided a rationale for a new treatment strategy that improved the skin lesions	44

			(tofacitinib treatment)	
Initial clinical problem	**Causal gene**	**Modified clinical interpretation (MIM number)**	**Significance of WES-based patient evaluation (treatment)**	**References**
Severe global developmental delay, seizures, and acanthotic skin lesions	*RAB11B*	Neurodevelopmental disorder with ataxic gait, absent speech, and decreased cortical white matter (#617807)	Identified a new disease gene leading to a neurodevelopmental syndrome	57

From: Lee, Y., Park, S., Lee, J.S. *et al.* Genomic profiling of 553 uncharacterized neurodevelopment patients reveals a high proportion of recessive pathogenic variant carriers in an outbred population. *Sci Rep* **10**, 1413 (2020). https://doi.org/10.1038/s41598-020-58101

provides a unique opportunity to conduct a systematic survey of rare disorders at a large scale. With a nationwide referral system focused on a handful of major tertiary clinical institutions, Seoul National University Children's Hospital (SNUCH) covers a large portion of the 51-million population, allowing for consistent evaluation and treatment of the patient cohort. Taking advantage of the extensive coverage of the patient pool maintained by Korea's centralized medical system and our analysis results of patients with severe neurodevelopmental disorders, it is feasible to infer the probability of recessive variant assembly in Koreans.

A reanalysis can be initiated to check for updates on new information specifically for what was already reported such as a VUS or a previously reported candidate gene with an uncertain link to human disease. The reanalysis of previously generated exome sequence data provides an opportunity to identify additional genetic causes with the patient's phenotype and increase the diagnostic yield of this testing. The ACMG recently reported some important factors to consider regarding the re-evaluation and reanalysis of genomic test results [22]. The evolving phenotype of pediatric patients over time and recognition of the phenotypic spectrum of a condition may also expand as more individuals with specific genotypes are identified. Therefore, re-review and routine evaluation may be considered when the patient's phenotype or family history has been developed in the interim. Multiple studies have shown that the diagnostic rate of reanalysis

ranges from 5 to 12% [21, 23-26]. Furthermore, exome reanalysis as a routine clinical practice may yield additional diagnoses due to improved knowledge of phenotype-genotype correlation [27, 28]. With advances in genomic sequencing technologies, the number of reported gene-disease relationships has rapidly expanded[29]. New entries are uploaded daily to OMIM or ClinVar, which is one of the most important databases of diseases and variants. According to the Online Mendelian Inheritance in Men (OMIM) compendium, 4,617 genes and their variants are associated with human disease as of April 2022 [29, 30](Fig1.1).

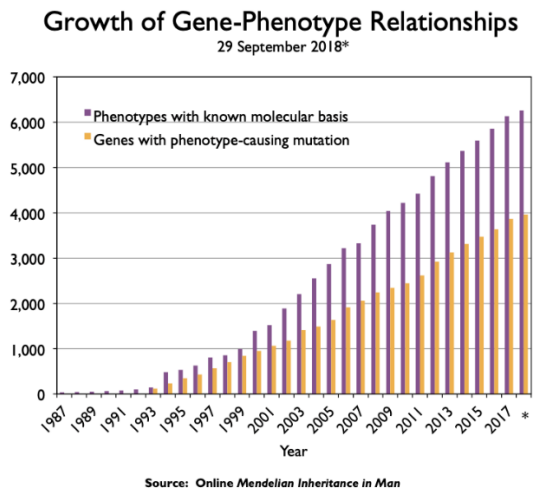


Figure 1.1 The pace of disease gene discovery as cataloged by the OMIM Morbid Map Scorecard.

From : Amberger, J.S., *et al*., OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res, 2015. **43**(Database issue): p. D789-98

Recent developments in NGS technology have led to the expanded carrier screening assessing hundreds of mutations associated with genetic diseases providing the opportunity to successfully screen different human population for carrier status [31, 32]. Autosomal recessive (AR) diseases make up a significant portion of Mendelian disorders, estimating to occur in 1.7-5 in 1,000 neonates [33]. Also, some recessive diseases are quite common in certain populations and for some of these disorders, Carriers having an affected allele have a risk of passing inherited disorders to their children, which is why the study of genes that cause recessive diseases is important. β-thalassaemia and Tay–Sachs disease is population-specific pilot carrier screening programs for recessive disease that has been proven successful reduction of the birth of affected individuals [34, 35]. However, those diseases are single gene disorders and display ethnic biases, making the process of variant curation and evaluation for pathogenicity and also the prediction of patients more efficient. The populations of these research targeted only specific regions or specific ethnic groups. Carrier frequency estimates depend on the variant evaluated and the population studied. Successful application of NGS in genetic screening overcomes these limitations by expanding the number of diseases covered and applies the testing to whole populations

The Genome Aggregation Database (gnomAD) is a popular genomic database used worldwide, which contains exome data collected from 9197 East Asians, including 1909 Koreans, and is suitable for East Asian studies. Korean Variant

Archive (KOVA) is a reference database of genetic variations in the Korean population [36]. By calculating the carrier frequency obtained from gnomAD in East Asians and KOVA, actual incidence of NDDs can be estimated. Thus, it would be feasible to predict potential rare recessive variants from genomic data of healthy parents with the help of large patients and control genomic data in the near future.

1.2 Materials and Methods

1.2.1 Patients and study criteria

Patients who were undiagnosed with complex neurological symptoms of suspected genetic origin were recruited for whole exome sequencing. The majority of patients visited the Seoul National University Children's Hospital (SNUCH) pediatric neurology clinic that is a tertiary referral center for pediatric neurology. A total of 1,180 patients who initial conventional tests such as candidate gene sequencing, microarray, metabolic work-up or muscle biopsy has failed were selected by specialist. Informed consent and blood samples for genomic DNA were obtained under the approval of the Seoul National University Hospital (SNUH) internal review board (#1406-081-588).

1.2.2 Whole exome sequencing

Genomic DNA was extracted from whole blood using the QIAamp DNA Blood Midi Kit according to the manufacturer's instructions (Qiagen, Valencia, CA). WES including library preparation, sequencing, and converting raw data to FastQ was performed at Theragen Etex Bio Institute (Suwon, Korea) following the manufacturer's recommendation. The data were analyzed by In-house bioinformatic pipeline using Picard software (v.2.8.0; [37]), samtools (v.1.8; [38])

and Genome Analysis Toolkit (GATK, v.4.1.4; [39]) which included read alignment, removal of PCR duplicates, base recalibration and variant qulity control (Fig1.2).

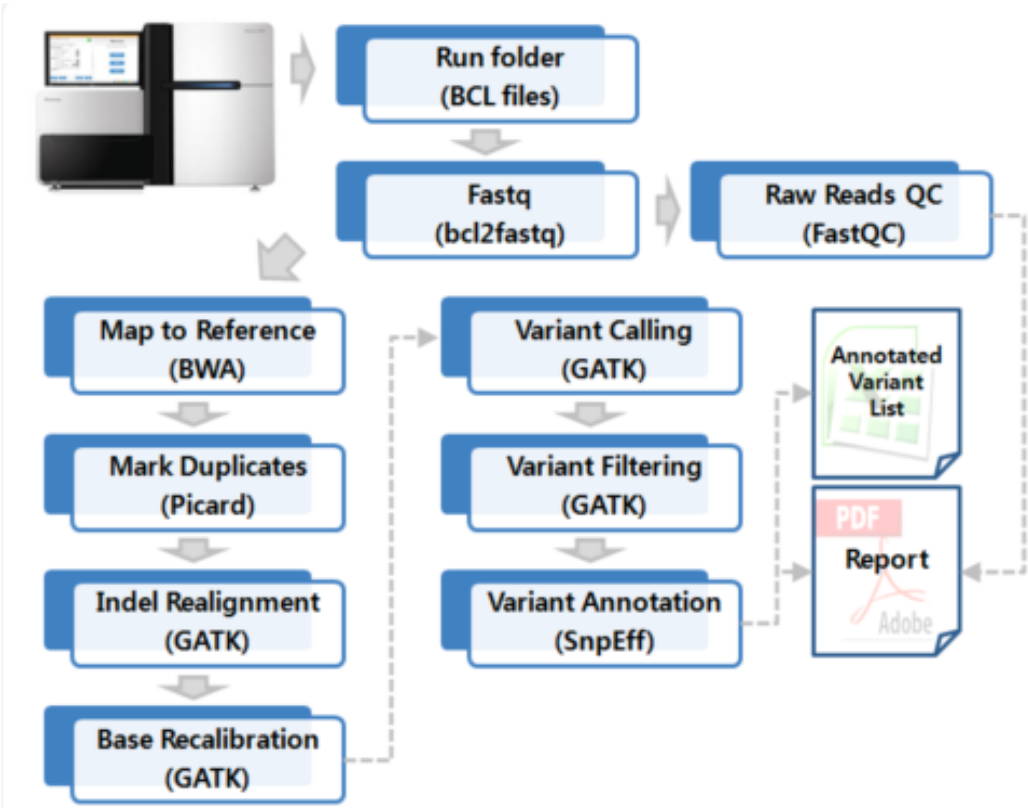


Fig1.2 Pipeline overview for whole exome sequencing. First, DNA sequence data from biological sample is produced. Then, Genotypes for all positions with an SNV were called through read alignment and variant calling. Finally, a variant list through variant annotation and classification were prepared for analysis step.

1.2.3 Evaluation of pathogenic variants

Annotated variants using ANNOVAR [40] and SnpEff [41] were classified with reference database including: (1) Normal population database such as gnomAD [42], ExAC, and 1000 Genomes. (2) In silico prediction scores such as CADD [43], SIFT [44], and phyloP [45] (3) Disease database such as OMIM [30, 46], HGMD [47], ClinVar [48]. Dominant variants that were never seen as heterozygous and recessive variants that were never seen as homozygous or hemizygous when filtered by allele frequency of 0.001 in heterozygous status were filtered. Subsequent filtering of remaining variants was considered based on information regarding whether variants were evolutionarily well conserved at amino acid level, genotype-phenotype associations, inheritance pattern such as de novo, compound heterozygous, homozygous and hemizygous by comparing to genotypes in the parents (Fig1.4). Copy number variation (CNV) analysis through WES was carried out by comparing the mean coverage depth of each captured interval to the mean coverage depth of parental samples as described previously (Fig1.3) [19].

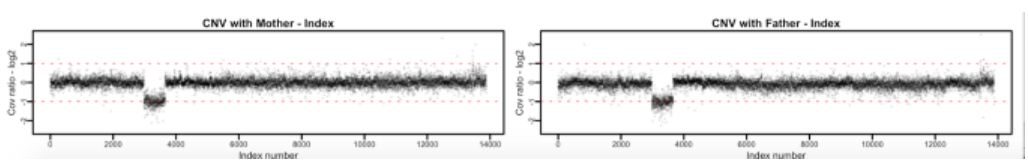


Fig1.3 Example of copy number variation (CNV) analysis by whole exome sequencing

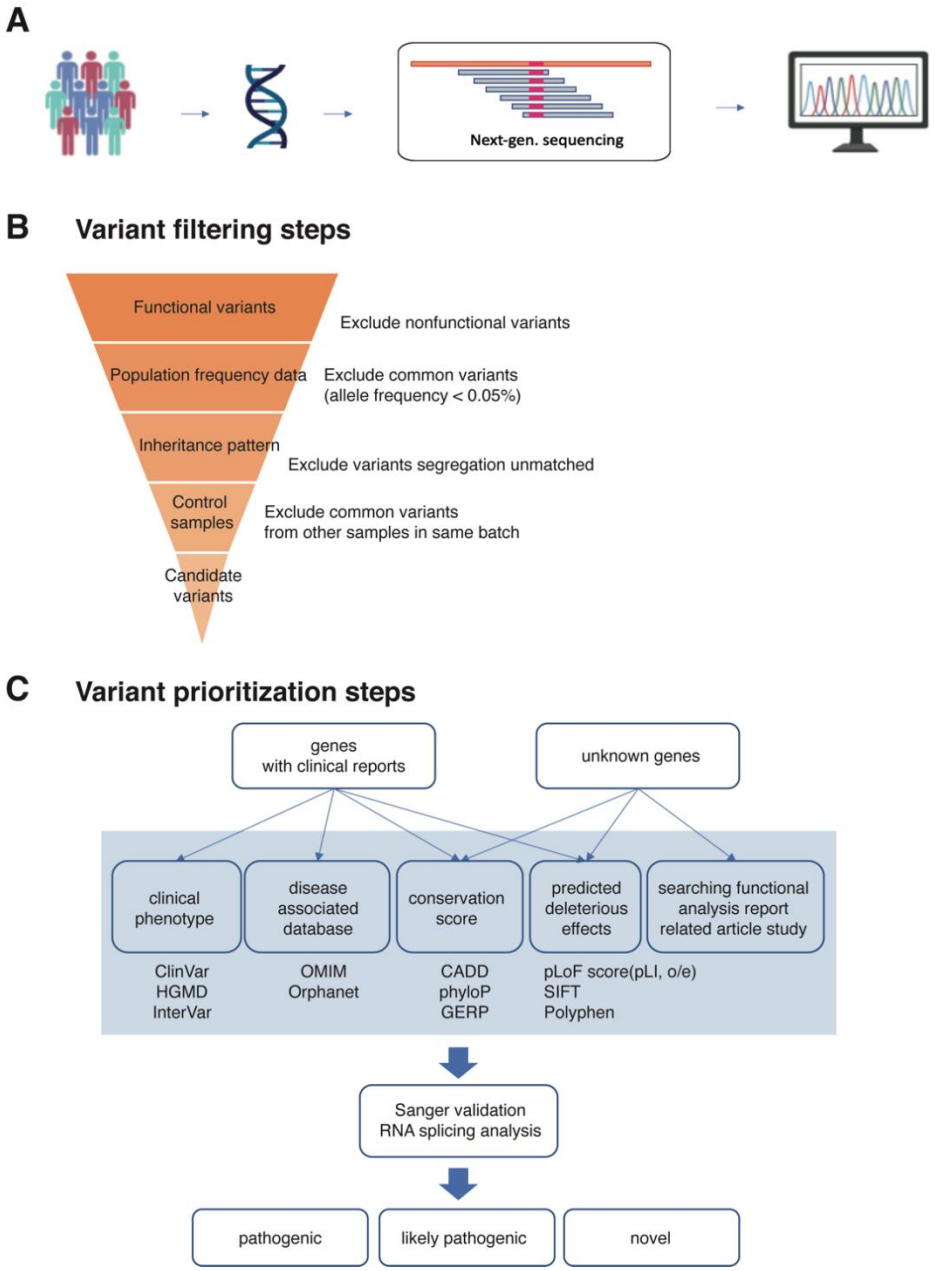


Figure 1.4 Experimental overview of whole-exome sequencing. (A) Preparing for NGS workflow. (B) Illustrating the steps for rare variants. (C) multiple tools for bioinformatic analysis and rare disease database to find causative gene.

1.3 Results

1.3.1 Genetic analysis of 1,180 patients with neurodevelopmental disorders

KND cohort consisted of pediatric patients (mean age = 11.27, range 1-62) with suspected genetic conditions displaying one or more neurological symptoms including developmental delay, intellectual disability, intractable seizure, involuntary movements, or muscle weakness who visited SNUCH during a 7-year period (2014-2020). For genetic analysis, WES was performed on DNA samples from 1,180 patients and the parents of 711 patients (Fig1.5). For detected variants in the patients with no parents, segregations were validated by sanger sequencing. The majority of KND patients are sporadic (1098/1180 = 93.1%; Fig1.5), which could be either de novo or inherited from carrier parents. Patients with diagnosed neurodevelopmental disorders were the most common (66.8%), followed by neuromuscular disease (21.3%) (Fig1.6). The resulting genome data were processed, and pathogenic variants called with a standard process (Subjects and Methods). The pathogenic variants were identified in the 284 disease-causing genes in 1,180 KND patients. Overall, 41.9% of the patients carried known variants that is highly scored or previously reported in known phenotype associated genes and 4.7% carried known variants but displayed symptoms different from those previously reported, possibly extending the disease spectra of these variants

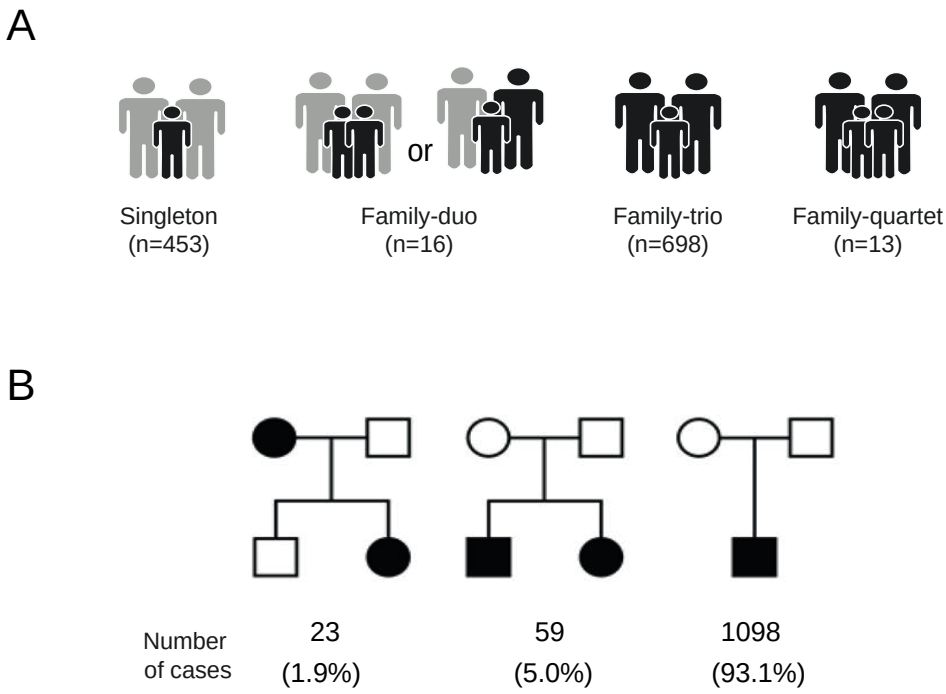


Fig1.5 Classification of KND cohort. (A) Family number applied to a case study (B) Genetic pedigree of a family with disease.

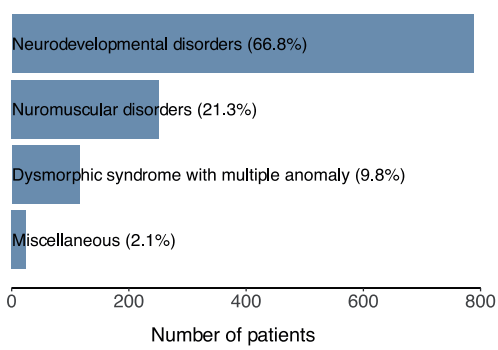


Fig1.6 Major clinical features of the KND cohort.

(Fig1.7A and Table1.2). Including the 4.2% of the patients with known CNVs (Table1.3), the result shows overall diagnostic yield of 50.8% with high confidence (Fig1.7A). This group with known variants was further divided according to variant inheritance pattern (Fig1.7B). De novo variants were identified in more than half of diagnosed patients (62.9%) and recessive variants in about a quarter (24.7%). Lastly, 8.7% carried variants on the X chromosome with hemizygous status, making the proportion of inherited variants 33.4%. 3.7% were miscellaneous which contains a shared dominant pathogenic variant from a parent, large deletion, loss of heterozygous (LOH), and mosaicism. There is no substantial difference in the proportion of variants for missense and loss-of-function by the dominant and recessive inheritance pattern (Fig1.7C). We identified 49 rare CNVs from whole exome sequencing read depth, in which 41 CNVs originated de novo (35 autosomal de novo CNV, 6 de novo CNV on chromosome X) and 7 was inherited maternal CNVs at chromosome X. Of the CNVs detected, CNVs with large del/dup (>10 Mb) was 5 and CNVs smaller that 1Mb in size was 44 and contained at least one morbid gene associated with phenotype (Table3). Pathogenic and likely pathogenic CNVs were obtained from DECIPHER CNV database (https://decipher.sanger.ac.uk/). This distribution of pathogenic variant inheritance is comparable to those reported in other studies using rare disease patients from outbred populations [1, 49-51].

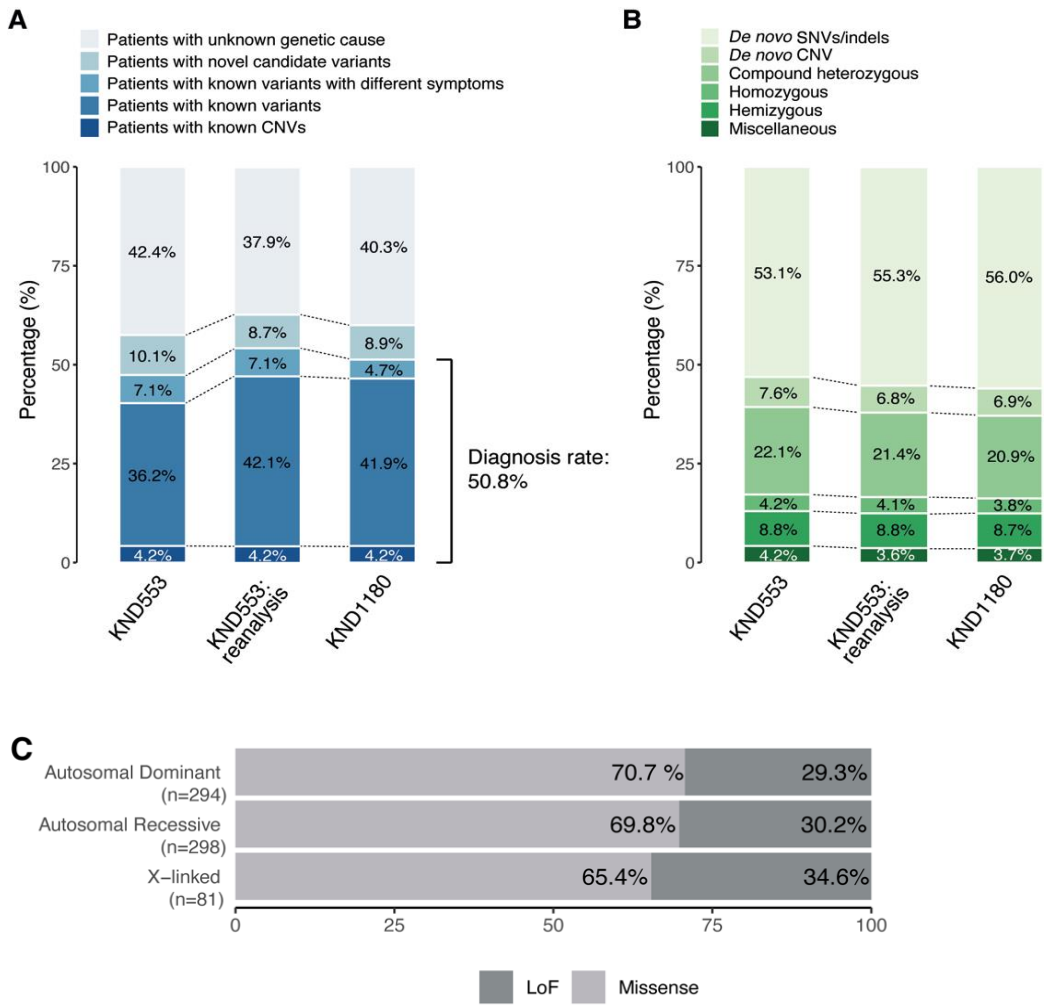


Fig1.7 Genetic diagnosis of 1,180 KND patients (KND1180). (A) Diagnostic yields of the 553 KND patients in 2020 (KND553), reanalysis of KND553, and KND1180. (B) Breakdown of diagnosed patients by mode of inheritance. (C) Breakdown of pathogenic variants by action mode.

Table2. List of variants with different symptoms

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
HSP3	Hereditary spastic paraplegia	*SLC25A15*	Missense	p.T39K	Com het	chr13	41373253	238970
-	-	*SLC25A15*	Nonsense	p.R179*	Com het	chr13	41381512	
HSP4	Hereditary spastic paraplegia	*RARB*	Missense	p.F167S	De novo	chr3	25635043	615524
HSP16	Hereditary spastic paraplegia	*RARB*	Missense	p.G289S	De novo	chr3	25635051	615524
HSP25	Hereditary spastic paraplegia	*PLA2G6*	Missense	p.K545R	Com het	chr22	38516874	256600, 610217, 612953
-	-	*PLA2G6*	Missense	p.P223L	Com het	chr22	38536118	
HSP31	Hereditary spastic paraplegia	*SPTBN2*	Missense	p.S315L	Shared variant	chr11	66478182	600224, 615386
LD10	Leukodystrophy	*CHD7*	Missense	p.S527C	De novo	chr8	61655571	214800, 612370
PM9	Pelizaeus-Merzbacher disease	*GLB1*	Missense	p.D317V	Hom	chr3	33059944	230500, 230600,

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
								230650, 253010
RS5	Rett syndrome like	*UBE3A*	Frameshift	p.L619fs	De novo	chr15	25601952	105830
RS8	Rett syndrome like	*GRIA2*	Missense	p.F644L	De novo	chr4	158262503	618917
RS14	Rett syndrome like	*ATRX*	Missense	p.G211D	Hemi	chrX	76940002	300448, 301040, 309580
RS18	Rett syndrome like	*SMC1A*	Indel (inframe)	p.K88del	De novo	chrX	53441788	300590, 301044
RS53	Rett syndrome like	*KMT2A*	Missense	p.G1168D	De novo	chr11	118348850	605130
RS60	Epileptic encephalopathy	*TUBA8*	Missense	p.D46G	Hom	chr22	18604379	619840
RS61	Rett syndrome like	*CLTC*	Missense	p.C1260R	De novo	chr17	57760280	617854
RS62	Epileptic encephalopathy	*SCN2A*	Missense	p.A240P	De novo	chr2	166166853	613721, 618924, 607745
SME1	Metabolic myopathy	*NSDHL*	Missense	p.S365R	Hemi	chrX	152037631	308050, 300831

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
TWE22	Neurometabolic disorder, mitochondrial cytopathy	*NARS2*	Missense	p.A244G	Com het	chr11	78204200	618434, 616239
-	-	*NARS2*	Missense	p.R451C	Com het	chr11	78147799	
TWE26	Multiple anomaly with facial dysmorphism	*HCFC1*	Missense	p.R344H	Hemi	chrX	153225739	309541
TWE28	Epileptic encephalopathy	*NARS2*	Ex-In boundary	NA	Com het	chr11	78180291	618434, 616239
-	-	*NARS2*	Missense	p.S101G	Com het	chr11	78279749	
TWE29	Mitochondrial cytopathy	*FIG4*	Missense	p.V295F	Com het	chr6	110064319	612691, 612577, 611228, 216340
-	-	*FIG4*	Frameshift	p.S583fs	Com het	chr6	110088097	
YAT6	Developmental delay	*ANO3*	Missense	p.S712N	De novo	chr11	26655829	615034
TWE80	Rett syndrome like	*ANKRD11*	Frameshift	p.E1075fs	De novo	chr16	89349723	148050

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
TWE87	Spastic quadriplegia and severe retardation	*COL1A1*	Nonsense	p.R120*	De novo	chr17	48276790	114000, 619115, 130060, 166200, 166210, 259420, 166220, 166710
TWE102	Developmental and epileptic encephalopathy	*PLA2G6*	Missense	p.R600Q	Com het	chr22	38512162	256600, 610217, 612953
-	-	*PLA2G6*	Missense	p.P223L	Com het	chr22	38536118	
TWE123	Developmental delay with leukodystrophy	*FOXG1*	Missense	p.R188P	De novo	chr14	29237047	613454
-	-	*FOXG1*	Missense	p.L189F	De novo	chr14	29237050	
TWE144	Global developmental delay with multiple anomaly	*NIPBL*	Missense	p.K697E	De novo	chr5	36985371	122470
TWE165	Mitochondrial disorder	*WDR81*	Missense	p.Q1695E	Com het	chr17	1637414	610185, 617967

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
-	-	*WDR81*	Missense	p.R1765H	Com het	chr17	1638980	
TWE168	Global developmental delay with syndromic face	*ASXL3*	Nonsense	p.R1117*	De novo	chr18	31323161	615485
TWE175	Nonketotic hyperglycinemia	*SCN1A*	Missense	p.G891R	De novo	chr2	166894561	619317, 607208, 604403, 609634
TWE176	Global developmental delay with congenital corneal clouding	*RARB*	Missense	p.R394L	De novo	chr3	25637920	615524
TWE178	Spinoal Muscular Atrophy	*MYH7*	Missense	p.L1822P	De novo	chr14	23884298	613426, 192600, 160500, 613426, 608358, 255160, 181430
TWE188	Global DD without dysmorphism	*KMT2C*	Missense	p.T1636M	Com het	chr7	151884448	617768
-	-	*KMT2C*	Missense	p.R2497H	Com het	chr7	151875048	

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
TWE195	Russell Silver syndrome like	*ASXL3*	Frameshift	p.V1200fs	De novo	chr18	31323408	615485
TWE196	Epileptic encephalopathy	*CASK*	Ex-In boundary	NA	De novo	chrX	41646537	300422, 300749, 300422
TWE248	Multiple anomaly	*SMC1A*	Ex-In boundary	NA	De novo	chrX	53409426	300590, 301044
TWE257	Noonan -RASopathy	*ARID2*	Frameshift	p.E402fs	De novo	chr12	46231360	617808
TWE320	Severe combined immunodeficiency (SCID)	*USP9X*	Missense	p.A2539V	Hom	chrX	41091680	300919, 300968
TWE330	Autism spectrum disorder (ASD)	*CACNA1C*	Missense	p.R860Q	Com het	chr12	2702427	611875, 618447, 601005
-	-	*CACNA1C*	Missense	p.R1851Q	Com het	chr12	2788926	
TWE331	Central hypotonia with microcephaly	*PIK3C2A*	Missense	p.G1660R	Com het	chr11	17111368	618440

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
-	-	*PIK3C2A*	Splice	NA	Com het	chr11	17112868	
TWE334	Hereditary spastic paraplegia	*RARB*	Missense	p.A300D	De novo	chr3	25635085	615524
TWE360	Neurometabolic disorder	*GRIA2*	Missense	p.Q607L	De novo	chr4	158257875	618917
TWE460	Multiple congenital anomalies	*PIEZO1*	Missense	p.R953H	Com het	chr16	88798876	194380, 616843
-	-	*PIEZO1*	Frameshift	largedel	Com het	chr16	88783317-88875970 (92,653)	
TWE478	Global developmental delay with facial dysmorphism	*MAN1B1*	Nonsense	p.Q82*	Com het	chr9	139982551	614202
-	-	*MAN1B1*	Missense	p.E570L	Com het	chr9	140001843	
TWE553	Focal epilepsy with mental retardation	*KIF4A*	Missense	p.R442V	Hemi	chrX	69563611	923
TWE567	Autism syndrome with	*PTEN*	Frameshift	p.E6fs	De novo	chr10	89623713	158350, 605309,

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
	macrocephaly							176807, 613028, 607174
TWE587	Neurodegenerative disease	*SLC6A19*	Splice	NA	Com het	chr5	1219227	234500, 138500, 242600
-	-	*SLC6A19*	Missense	p.M628T	Com het	chr5	1221997	
U1	Russel-Silver syndrome like	*COL1A1*	Missense	p.G1127D	De novo	chr17	48264888	114000, 619115, 130060, 166200, 166210, 259420, 166220, 166710
KBRI-16	Global developmental dealy with facial dysmorphism	*ZC4H2*	Missense	p.Q82P	Hemi	chrX	64140114	314580, 301041
KBRI-66	Global developmental delay with static leukodystrophy	*USP9X*	Missense	p.R425P	Hemi	chrX	41002655	300919, 300968
CGS36	Epileptic encephalopathy with	*ATP2B2*	Frameshift	p.V1113fs	De novo	chr3	10379941	619804, 601386

ID	Simple Dx	Gene name	Status	Amino acid change	Mutation type	Chr	Position (hg19)	OMIM association
	macrocephaly							
CGS73	Nonspecific mental retardation and developmental delay	*ZIC1*	Missense	p.H260N	De novo	chr3	147128677	616602, 618736
CGS88	Global developmental delay with facial dysmorphism	*GRIP1*	Missense	p.P910L	Hom	chr12	66765601	617667
CGS119	Global developmental delay with facial dysmorphism	*WDR26*	Frameshift	p.E353fs	De novo	chr1	224599230	617616
CGS124	Developmental and epileptic encephalopathy	*PUF60*	Frameshift	p.I521fs	De novo	chr8	144898810	615583
CGS137	Motor developmental delay with microcephaly	*TUBB*	Missense	p.A303P	De novo	chr6	30691686	615771, 156610

Table3. List of known CNVs.

ID	Simple Dx	Mutation type	CNV interval	Length	Del/Dup
HSP10	Hereditary spastic paraplegia	Shared CNV	chr2:32.3-32.5	0.2 Mb	Del
HSP21	Hereditary spastic paraplegia	De novo CNV	chr22:21.1-21.6	0.5 Mb	Del
HSP29	Hereditary spastic paraplegia	Inherited CNV	chrX:153.6-153.8	0.2 Mb	Dup
RS24	Rett syndrome like	De novo CNV	chr4:0.0-2.8	2.8 Mb	Del
RS42	Rett syndrome like	De novo CNV	chr1:0.6-3.1	2.6 Mb	Del
RS43	Rett syndrome like	De novo CNV	chr7:119.7-119.9	0.2 Mb	Del
RS46	Rett syndrome like	De novo CNV	chr1:107.0-113.1	6.1 Mb	Del
RS55	Rett syndrome like	De novo CNV	chr2:234.8-242.8	8.0 Mb	Del
RS57	Rett syndrome like	De novo CNV	chrX:153.3-153.3	0.01 Mb	Del
RS66	Rett syndrome like	De novo CNV	chr3:9.0-13.0	4.0 Mb	Del
RS71	Rett syndrome like	De novo CNV	chr22:42.8-51.3	8.6 Mb	Del
SME15	Developmental delay	De novo CNV	chrX:41.4-41.7	0.3 Mb	Del

ID	Simple Dx	Mutation type	CNV interval	Length	Del/Dup
TWE35	Neurodegenerative disease	De novo CNV	chrX:102.5-103.2	0.7 Mb	Dup
TWE72	Developmental dealy with congenital retinal dystrophies	De novo CNV	chr15:20.3-23.7	3.4 Mb	Del
TWE124	Epileptic encephalopathy	De novo CNV	chr9:0-47.3	47.3 Mb	Dup
TWE240	Autism spectrum disorder with facial dysmorphism	De novo CNV	chr16:29.6-30.2	0.6 Mb	Del
TWE293	Leukodystrophy with macrocephaly	De novo CNV	chr5:0.1-19.8	19.7 Mb	Dup
		De novo CNV	chr11:120.7-134.9	14.1 Mb	Del
TWE303	Mobius syndrome like face	De novo CNV	chr16:21.8-22.4	0.6 Mb	Del
TWE310	Developmental delay	De novo CNV	chr22:18.9-21.4	2.5 Mb	Del
TWE325	Global developmental delay with facial dysmorphism	De novo CNV	chr2:46.1-48.1	2.1 Mb	Del
TWE362	Global developmental delay with microcephlay	De novo CNV	chr17:43.9-44.2	0.3 Mb	Del
TWE419	Multiple anomaly with severe intellectual disability	De novo CNV	chr16:46.5-49.7	3.2 Mb	Del
		De novo CNV	chr16:52.1-55.6	3.6 Mb	Dup
TWE450	Gloval developmental delay	De novo CNV	chr3:41.3-42.0	0.7 Mb	Del

ID	Simple Dx	Mutation type	CNV interval	Length	Del/Dup
TWE465	Autism spectrum disorder	De novo CNV	chrX:141.0-154.7	13.8 Mb	Dup
TWE495	Hereditary spastic paraplegia	Inherited CNV	chrX:153.6-153.8	0.2 Mb	Dup
TWE506	Global developmental delay with facial dysmorphism	De novo CNV	chr15:20.6-28.8	8.2 Mb	Del
TWE518	Global developmental delay with multiple joint contracture	De novo CNV	chrX:63.5-64.7	1.2 Mb	Del
TWE588	Global developmental delay with facial dysmorphism	De novo CNV	chr9:0.3-39.2	38.9 Mb	Dup
TWE659	Neonatal seizure	De novo CNV	chr2:165.4-167.3	1.9 Mb	Dup
U2	Epileptic encephalopathy	Inherited CNV	chrX:100.1-107.3	7.2 Mb	Dup
U8	congenital hypotonia	De novo CNV	chr5:139.0-139.6	0.6 Mb	Del
YAT7	Familial spinocerebellar ataxia	Inherited CNV	chrX:62.9-63.6	0.8 Mb	Dup
KBRI-3	Global developmental delay with facial dysmorphism	De novo CNV	chr9:139.6-141.1	1.6 Mb	Del
KBRI-6	Global developmental delay with facial dysmorphism	De novo CNV	chr3:44.2-48.0	3.9 Mb	Del
KBRI-41	Global developmental delay	De novo CNV	chr18:52.5-53.3	0.8 Mb	Del
KBRI-53	Global developmental delay with facial dysmorphism	De novo CNV	chr12:53.6-54.1	0.6 Mb	Dup

ID	Simple Dx	Mutation type	CNV interval	Length	Del/Dup
KBRI-54	Epileptic encephalopathy	De novo CNV	chr2:171.2-175.1	3.8 Mb	Del
CGS13	Epileptic encephalopathy	Inherited CNV	chrX:152.8-153.4	0.6 Mb	Dup
CGS21	Global developmental delay	De novo CNV	chr16:87.9-89.9	2.0 Mb	Del
CGS35	Global developmental delay with facial dysmorphism	De novo CNV	chr4:83.2-85.8	2.6 Mb	Del
CGS43	Global developmental delay with microcephaly	De novo CNV	chr17:34.8-36.2	1.4 Mb	Dup
CGS67	Developmental delay and dandy walker malformation	De novo CNV	chr8:1-7.3	7.3 Mb	Del
		De novo CNV	chr8:12.6-38.2	25.7 Mb	Dup
CGS100	Intellectual disability with microcephaly	De novo CNV	chr16:0.07-1.7	1.6 Mb	Del
		De novo CNV	chr22:51.0-51.2	0.2 Mb	Dup
CGS112	Global developmental delay	Inherited CNV	chrX:153.0-153.6	0.6 Mb	Dup
CGS113	Global developmental delay with facial dysmorphism	De novo CNV	chr20:61.5 - 62.3	0.8 Mb	Del
CGS114	Global developmental delay	De novo CNV	chr22:18.8-21.6	2.8 Mb	Del
CGS116	Epilepsy and mental retardation	Inherited CNV	chrX:153.1-153.6	0.4 Mb	Dup

ID	Simple Dx	Mutation type	CNV interval	Length	Del/Dup
CGS120	Global developmental delay with microcephaly	De novo CNV	chrX:41.4-41.4	0.07 Mb	Del
CGS129	Global developmental delay with microcephaly	De novo CNV	chr18:63.4-78.0	14.6 Mb	Del

1.3.2 Reanalysis improved diagnostic yield

We applied reanalysis to the 553 cases that were previously analyzed in 2020 (KND553; Fig1.7) [19]. Re-assessment of existing exome sequence data with updated pipeline and clinical phenotypes provides discovery of new variants [27, 52-54]. Improved diagnostic yield obtained by reanalysis have been reported [27, 52-54]. With the reanalysis of the KND553, an additional 33 patients was diagnosed and a 5.9% increase in diagnostic yield was achieved. These 33 cases were reclassified as patients with known variants and they can be broadly divided into two groups, (1) variants for which new entries in OMIM allowed defining them as pathogenic (n = 16; Table4) and (2) pathogenic calls previously missed during the bioinformatic process (n = 17; Table5).

Table4. List of newly diagnosed cases due to new gene entry into OMIM.

Index	Gene	Variant type	Variant	Phenotype**	OMIM entry number and date of creation
1	*GEMIN5*	Comp het	c.3857A>G; p.Tyr1286Cys c.2510-2A>T	Progressive cerebellar atrophy with severe developmental arrest	#619333; 05/19/2021
2	*STAG2*	De novo het	c.3724C>T; p.Arg1242*	Holoprosencephaly	#301043; 04/07/2020
3	*MED12L*	De novo het	c.1895C>T; p.Ser632Leu	Global DD with FD	#618872; 05/02/2020
4	*DHX16*	De novo het	c.2021C>T; p.Thr674Met	Congenital myopathy	#618733; 01/09/2020
5	*HK1*	De novo het	c.1475C>T; p.Thr492Met	Severe brain atrophy, deep cortex disruption	#618547; 08/20/2019
6	*ADH5*	Comp het	c.678delA; p.Asp227fs	Global DD, myelodysplastic syndrome	#619151; 01/13/2021
7	*SIAH1*	De novo het	c.613G>C; p.Gly205Arg	Global DD and FD	#619314; 05/06/2021
8	*MN1*	De novo het	c.3850delC; p.His1284fs	CHARGE syndrome	#618774; 02/11/2020

Variant is pathogenic in ClinVar, but parental samples were not available.

** Abbreviations: ID, intellectual disability; DD, developmental delay; FD, facial dysmorphism; EE, epileptic encephalopathy.

Table5. List of newly diagnosed cases by data re-analysis.

Index	Gene	Variant type	Variant	Phenotype	Reason
1	*PMM2*	Comp het	c.194A>G; p.Asp65Gly c.713G>C; p.Arg238Pro	Progressive cerebellar atrophy	Not clear
2	*PDHA1*	De novo het	c.613G>A; p.Val205Met	Rett syndrome-like	Not clear
3	*ZEB2*	De novo het	c.2083C>T; p.Arg695*	Rett syndrome-like	Not clear
4	*KAT6B*	De novo het	c.3147G>A; p.Pro1049Pro	Rett syndrome-like	A synonymous variant; called during re-evaluation
5	*ACTB*	De novo het	c.547C>T; p.Arg183Trp	Severe dystonia, ID, and SNHL	Not clear
6	*CASK*	Hemizygous	c.1667T>G; p.Leu556Arg	Autism spectrum disorder with FD	Not clear
7	*MAGEL2*	De novo het	c.2873G>A; p.Trp958*	ID with FD and multiple anomaly	Not clear
8	*NARS2*	Comp het	c.1163C>T; p.Thr388Met c.88G>C; p.Val30Leu	EE	Not clear

Index	Gene	Variant type	Variant	Phenotype	Reason
9	*FOXG1*	pending	c.460dupG; p.Glu154fs	EE and microcephaly	Not clear
10	*DYRK1A*	De novo het	c.520G>T; p.Val174Leu c.521T>A; p.Val174Glu	DD with microcephaly and FD	Not clear
11	*SLC16A2*	Hemizygous	c.1265T>G; p.Leu422Arg	Neurodegenerative disorder	Not clear
12	*UGDH*	Comp het	c.1183G>A; p.Val395Met c.1038-2A>G	Familial EE	Not clear
13	*PDHA1*	Hemizygous	c.761T>C; p.Leu254Ser	Leigh Syndrome	Initially missed due to coverage depth < 10
14	*TBR1*	De novo het	c.1588_1594dupGGCTGCA; p.Thr532fs	Rett syndrome-like	Initially missed due to coverage depth < 10
15	*IQSEC2*	De novo hemi	c.2139delC; p.Gly714fs	Rett syndrome-like	Initially missed due to coverage depth < 10
16	*SMC1A*	Possible de novo het*	c.2923C>T; p.Arg975*	Rett syndrome-like	Initially missed due to coverage depth < 10

Index	Gene	Variant type	Variant	Phenotype	Reason
17	*AHDC1*	Possible de novo het*	c.2389G>T; p.Glu797*	Rett syndrome-like	Initially missed due to coverage depth < 10
18	*IRF2BPL*	De novo het	c.562C>T; p.Arg188*	Neurodegenerative disease	Called during phenotype re-evaluation
19	*UBAP1*	De novo het	c.529dupA; p.Met177fs	Hereditary spastic paraplegia	Called during phenotype re-evaluation
20	*GABRB3*	Possible de novo het*	c.554C>T; p.Thr185Ile	Rett syndrome-like	Called during phenotype re-evaluation
21	*CDK13*	Possible de novo het*	c.2149G>A; p.Gly717Arg	Rett syndrome-like	Called during phenotype re-evaluation
22	*TRAPPC11*	Hom	c.302A>G; p.Tyr101Cys	Unknown muscular dystrophy, most likely calpainopathy	Called during phenotype re-evaluation
23	*SLC35A2*	De novo hemi	c.1A>C; p.Met1?	Ullrich disease or Bethlem myopathy suspected	Called during phenotype re-evaluation

Index	Gene	Variant type	Variant	Phenotype	Reason
24	*DHDDS*	De novo het	c.110G>A; p.Arg37His	EE	Called during phenotype re-evaluation
25	*KMT2C*	De novo het	c.5716C>T; p.Arg1906*	Female ID, microcephaly	Called during phenotype re-evaluation

Variant is pathogenic in ClinVar, but parental samples were not available.

** Abbreviations: ID, intellectual disability; DD, developmental delay; FD, facial dysmorphism; EE, epileptic encephalopathy.

1.3.3 Estimating carrier frequencies of variants that cause recessive neurodevelopmental disorders

We Recent developments in genomics enables identifying carriers of hundreds of causal mutations. Population frequencies of gnomAD East Asian and Korean Variant Archive (KOVA 2; 5,305 healthy Korean individual set [55]) was used for calculating carrier frequencies. Allele frequencies of all lof variants in gnomAD and KOVA, as well as missense variants expected to be pathogenic through ClinVar and KND, were analyzed to predict the number of carriers. This provided us with an estimation of the number of carriers of recessive NDD in the general Korean population (Fig1.8A). The estimation yields were variable by gene. For example, the most common gene, *VPS13B*, was present at a frequency of 1.2%. Adding up all the estimations of known NDD genes in KND1180, it is predicted that 25.4% of healthy individuals will carry one of the recessive variants, under the assumption that each carrier harbors a single recessive variant in one gene. Finally, on average, the estimation yield for KND1180 variants were 3.92-fold higher than those determined for KND5533 variants, implying that larger cohort size is critical for increased sensitivity (Fig1.8B).

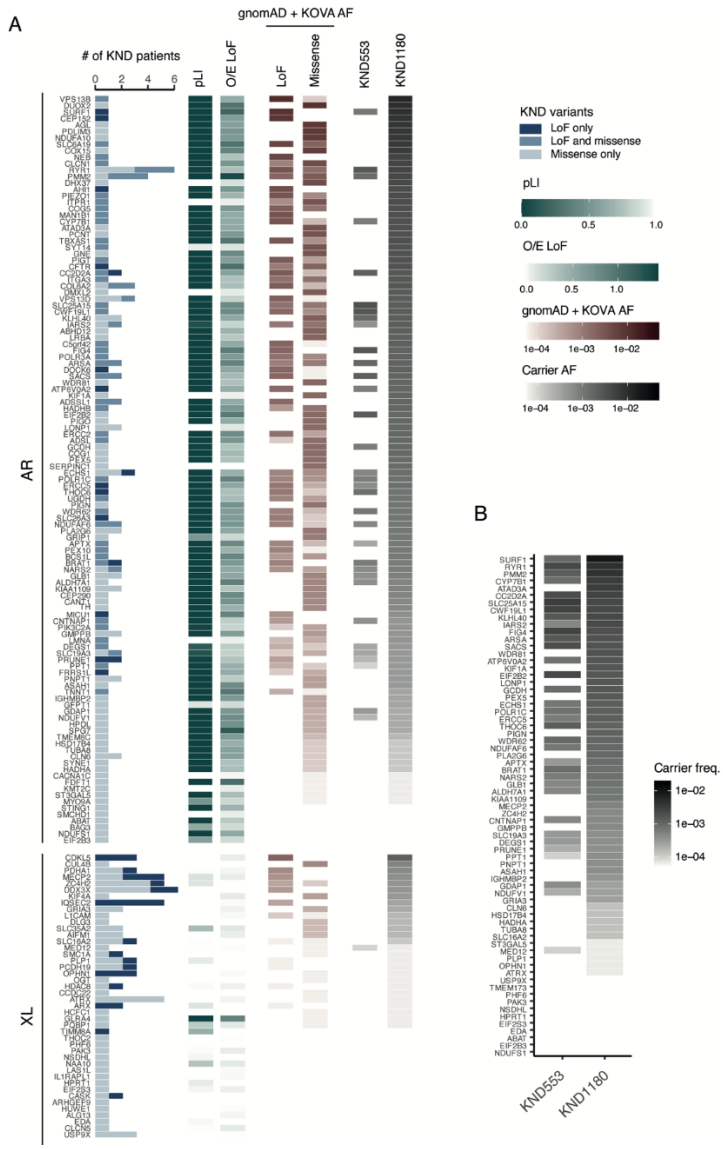


Fig1.8 Carrier frequency of pathogenic variants based on KND. (A) Heatmap represents the number of KND patients carrying a causal variant, the carrier frequency of pathogenic variants and the aggregation of carrier frequency. (B) Comparison of the carrier frequency based on KND553 or KND1180 on the 34 overlapping genes.

1.4 Discussion

NDDs are characterized by considerable genetic and clinical variability. WES from 1,180 undiagnosed patients with NDD was conducted to elucidate genetic etiology. Previously reported diagnostic rates of WES vary substantially among studies, ranging from 25% to 56% [1-3, 5, 6, 49, 51, 54]. Here, the diagnostic yield for definitive pathogenic variant findings in 1,180 KND patients was 50.8%. Among the diagnosed patients, the majority of the patients is sporadic cause (376/598 = 62.9%) and inherited variants is 33.4% (200/598), demonstrating that a substantial portion of KND patients inherited pathogenic variants from healthy parents (Fig1.7). Estimates of recessive manner contribution remain variable for study populations and disorders. DDD study revealed a small contribution of recessive disorders (3.6%) to patients of European ancestry [56], whereas recessive inheritance in patients with neuromuscular disorders was 57% [57]. This likely reflects differences in disorder subtypes.

It is expected that exome reanalysis applying the latest versions of databases and using improved bioinformatic tools would increase diagnostic yield [27, 52-54]. In the initial reporting using 553 patients in 2020, a diagnostic yield of 47.5% was achieved. With the re-analysis of 291 patients from the KND553 set who remained without clear pathogenic variants, the diagnostic yield increased to 53.4%. and this was due to some reasons: 8 cases of these novel diagnoses were attributed to newly

discovered and deposited gene-disease associations in OMIM. 5 cases were resulted from increased coverage allowing identification of variants that may previously have been missed. Synonymous variant affecting gene splicing of KAT6B (p.Pro1049=) was detected as synonymous variants were filtered out by automated steps in the initial analysis. Previously analyzed variants for 11 probands were re-evaluated and re-classified as pathogenic (Not clear) and 7 probands were diagnosed during re-evaluation of the phenotype due to evolving clinical phenotype in pediatric patients (Fig1.7; Table4; Table5). Therefore, those results suggest that exome sequencing data should be a routine clinical practice.

The frequency of carriers varies among population groups and specific genetic conditions could be skewed toward particular ethnic groups [35, 58-60]. In isolated populations, carrier frequencies of rare disorders in the general population may be very high[61]. Ethnic Koreans are a relatively outbred population, and a marriage is not allowed between men and women who have the same surnames and blood relatives. for more than 500 years [62]. As a major tertiary clinical institution, SNUCH covers a large portion of rare NDD patients in the country. Therefore, this study provides an unprecedented opportunity to study the occurrence of recessive diseases in an outbred population. We estimated that 33.4% of patients in the KND1180 cohort were affected by recessive conditions, which allows us to use population-specific databases such as gnomAD East Asian and KOVA to calculate carrier frequencies for reported and predicted pathogenic variants in the general

population. As expected, the larger sample size of this cohort relative to the KND553 cohort resulted in a greater number of pathogenic genes and an increase in the reported disease-associated variants enrolled in ClinVar and OMIM. Although calculated carrier frequencies may differ from those observed in clinical practices, the findings from this study will provide genetic evidence for the utility of preconception carrier screening.

Chapter 2

Genetic characterization of Inheritance Pattern in Genes Cause Neurodevelopmental Disorders

2.1 Introduction

1.2.1 Protein function based on inheritance pattern

Group of disorders characterized by genetic heterogeneity and complex phenotypes is difficult to establish genotype-phenotype correlations. However, it's possible to predict disease etiologies by identifying genetic traits. Genetic features such as innate function and expression pattern of the gene was used to predict novel risk genes. Hence, comparing their genetic properties will help to gain a more comprehensive view of NDD. Genetic disorders are caused from gene alterations with autosomal-dominant, autosomal-recessive, and X-linked patterns of inheritance. Also, disease-causing genes can be categorized in gain-of-function (GoF) and loss-of-function (LoF) based on biological function of protein. GoF conditions are typically dominant, which can confer a new function. However, LoF conditions may be either dominant or recessive depending whether a gene carrying variant is haploinsufficient [63, 64]. These biological properties can impact both inheritance mode and the phenotype when gene function or expression was altered. However, it is not yet clear what drives genes to carry variants that inherited in dominant and recessive patterns. Molecular and clinical characterization of NDD is strongly needed to get further insight into the genotype–phenotype correlation.

A lot of disease is typically classified using various clinical symptoms and signs. However, overlapping pathological phenotype could be determined by

different mutations in different genes, which needs to re-evaluation and proper clinical diagnosis. SETD5 gene, cause mental retardation autosomal dominant 23 (MRD23), and ANKRD11 gene, cause KBG syndrome, were considered overlapping phenotype that features delayed speech and motor development as well as mild to moderate intellectual disabilities. Both genes follow an autosomal dominant inheritance pattern, functionally act as chromatin regulator recruiting histone deacetylases (HDACs) and highly expressed in the brain. Crippa *et al.* proposed that phenotypic overlap may be the functional role and interactions of gene pathways [65]. In HSP that is a group(subset) of NDD, clinical phenotype already has been tried to classify based on inheritance mode, in which dominant HSP mainly present pure form, while recessive HSP is often associated with complex form according to additional neurological and extraneurological signs [66, 67]. Given the clinical and mechanistic overlap, HSP reflects the contribution of diverse cellular pathways to their pathogenesis [68-70]. A number of molecular pathways have been implicated in heterogeneity of HSP including mitochondrial functions, microtubule trafficking and lipid metabolism [68, 71, 72]. In the same context, Systematic classification of several recessive multisystemic or complex metabolic disorders present with ataxia has been previously published [73, 74]. These suggested that there may be a shared genetic architecture distinguishing dominance and recessiveness, as well as mechanistic overlap can lead to phenotypic overlap.

1.2.2 Data resources

2.1.2.1 The disease-associated variant database

Gene-disease associations are continually growing. The Online Mendelian Inheritance in Men (OMIM) that shows curated information on genetic disorders and genes listed 4,681 genes associated with human diseases, 4,313 of these are classified as single gene disorders and traits (Updated August 31st, 2022) [29, 30]. ClinVar also provide information about relationship among sequence variation and associated trait [48]. New entries are uploaded daily to OMIM or ClinVar, which is one of important databases of diseases and variant.

2.1.2.2 The ASD/ID-associated variant databases

The application of WES in clinical practice can enable a large-scale research study. the Simons Foundation Autism Research Initiative (SFARI) represents the largest autism spectrum disorder (ASD) cohort in the world [75, 76]. Over the past few decades, genetic studies of ASD have supported a significant genetic contribution to its etiology. SFARI-gene database, which comprises lists of curated genes considered to have causative roles in ASD when mutated in patients.

The Deciphering Developmental Disorders (DDD) Study has recruited nearly 14,000 children with severe undiagnosed developmental disorders in the UK and has performed genome-wide microarray and whole exome sequencing [49].

2.1.2.3 Expression atlas

Gene expression is one of the main molecular processes regulating differentiation, development, and functioning of cells and tissues. Genome-wide expression profiling in a number of human tissues from the Genotype-Tissue Expression (GTEx) project can be used to identify gene expression pattern across tissues [77, 78].

BrainSpan is accumulating data on the spatial and temporal dynamics of the transcriptome, which is profiling 16 anatomically distinct brain structures across the 13 developmental stages. The rapid change of gene expression patterns in the developing brain during the prenatal and neonatal stages of its development is associated with major neurodevelopmental trajectories [79].

2.1.2.4 Functional significance of variants

Assessing the pathogenicity of variants is critically important in genetic studies and clinical diagnosis. However, accurate pathogenicity prediction of variants has many challenges in, especially, predicting the pathogenicity of missense variants. Several in silico methods such as CADD [43], SIFT [44], PolyPhen-2 [80] are commonly used to see if a given variant is damaging or benign considering sequence conservation and biochemical features. The variants with higher scores are more likely to be deleterious. pLI [81], a measure widely used to identify genes that are intolerant to a single copy of a truncating mutation. It has been widely

adopted to classify genes into recessive and haploinsufficiency. High pLI scores indicates a very high likelihood for haploinsufficiency of the gene. These are measures to gain a more comprehensive view of variants functional effect.

2.2 Materials and methods

2.2.1 Gene expression analysis

The Genotype-Tissue Expression (GTEx) project (v8;[78]) were used to extract the normalized transcript level (TPM) of each gene. To determine relative brain expression vs body expression, gene-level relative values were calculated by comparing the median TPM value of brain region and other regions. and then plotted to visualize the distribution of the gene set. RNA-seq data and exon microarray data from BrainSpan (http://www.brainspan.org) were used to analyze brain spatial and temporal gene expression in the brain [79]. The prenatal time points include 8, 9, 12, 13, 16, 17, 19, 21, 24, 25, 26, 35, and 37 weeks and postnatal time points include 4 months, 10 months, 1, 2, 3, 4, 8, 11, 13, 15, 18, 19, 21, 23, 30, 36, 27, and 40 years. This dataset contains expression from 16 cortical and subcortical structures along the full course of human brain development.

2.2.2 Protein-protein interaction analysis

Tissue-specific protein-protein interaction information was investigated using experimentally validated PPI data that is an in vivo mapping of protein–protein interactions of seven mouse tissues that Miachael A. Skinnider recently reported in cell [82]. The dataset contained more than 190,000 high-confidence PPIs

identified with stable isotope labelling of tissues was downloaded. Protein pairs that contained causal genes identified in the KND, DDD and SFARI cohort was extracted which means physical interaction.

2.2.3 Gene ontology analysis

To analyze disease associations and biological pathways that are enriched in a selected genes, a web-based analysis tool from Metascape (https://metascape.org/gp/index.html) was used. Conventional GO sources were used: biological process (BP), Cellular Component (CC) and Molecular Function (MF). Disease Gene Network (DisGeNET) was used for disease ontology. Results were collected and grouped into clusters for comparative analyses of biological process and disease association between gene groups.

2.2.4 Statistical evaluation

Wilcox test was used to determine the statistical significance of the observed differences in functional scores for genes with different inheritance modes. The statistical significance of the expression level in boxplots was measured by a two-sample t-test. Statistical analyses were performed with R version 3.6.2.

1.3 Results

2.3.1 Genetic characteristics of genes that follow a dominant or recessive pattern

NDDs are characterized by considerable clinical and biological heterogeneity. An understanding of heterogeneity of NDDs need a deep knowledge of clinical and biological mechanisms. The innate function and expression pattern of a gene can impact both its inheritance mode and the phenotype observed when its function or expression is altered. Identifying common pathological mechanisms in heterogeneous disorders may be helpful to understand the relation between the phenotype and genotype. To find out biological meaning from pathogenic gene sets that follow a dominant or recessive pattern in NDD, we first performed functional enrichment analysis and compared them with corresponding gene sets from DDD or SFARI (Fig2.1). We used Gene ontology (GO) terms about specific biological process (BP), molecular function (MF), or cellular components (CC) within the gene set. This method shows the classification of gene function at the molecular and cellular level. The results showed KND genes are annotated to terms relating to brain developmental progression, such as regulation of membrane potential, chromatin organization, head development, and pyrophosphatase activity. In addition, the published variants from DDD and SFARI shared biological mechanisms involved in brain development (Fig2.1). Each cohort can be divided into inheritance pattern to see if there is a significant annotation bias. For each gene set, we systematically identified enriched GO terms. Interestingly, we find that

dominant and recessive gene sets have different terms which indicates that they are functionally associated with other types of biological mechanism. In particular, KND dominant genes were strongly enriched for synaptic functions, while KND recessive genes were characterized by metabolic process, mitochondrial function, and muscular disease terms. Remarkably, KND X-linked genes shared more terms with dominant genes than with recessive genes (Fig2.1B). This is unexpected because the majority of X-linked genes follow a hemizygous pattern, where disease manifests when the only X-chromosome allele in a male patient is mutated and may follow a recessive pattern. These results are based on enrichment relative to the set of all genes. But there is a risk of misinterpretation of these data due to the incorrect choice of the background gene set. A whole gene background might be inappropriate for enrichment analysis of X-linked genes. For enrichment analysis of X-linked genes, 850 genes encoded on chromosome X were used as a background gene set, and analysis was rerun. The result showed that GO terms for X-linked genes were still enriched for synaptic function (Table6). We also identified enriched GO terms for the DDD and SFARI gene sets divided into inheritance patterns in the same manner (Fig2.1C, D). We found differential enrichments between DDD dominant genes and DDD recessive genes as well as DDD X-linked genes, though SFARI showed little change in GO enrichment by inheritance pattern.

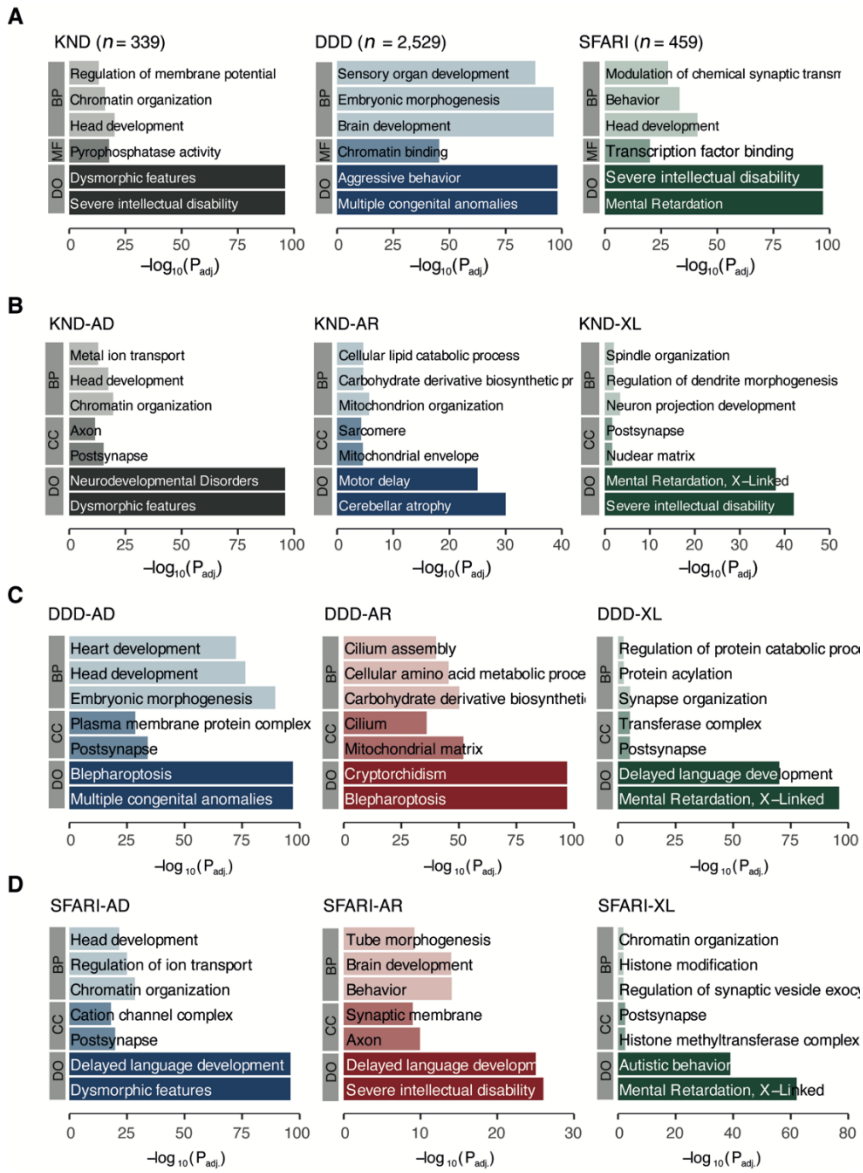


Fig2.1 Gene Ontology enrichment analysis of genes that cause NDDs. (A) Genes in KND, DDD, and SFARI. (B)(C)(D) Breakdown of KND, DDD, and SFARI genes by inheritance patterns. BP, biological process; CC, cellular component; MF, molecular function; DO, disease ontology. AD, autosomal dominant; AR, autosomal recessive; XL, X-linked.

KND-XL

Category	Term	Description	Log(q-value)	InTerm_InList
GO Biological Processes	GO:0051276	chromosome organization	-3.299	8/13
GO Biological Processes	GO:0031175	neuron projection development	-3.093	10/25
GO Biological Processes	GO:0042063	gliogenesis	-1.686	4/4
GO Cellular Components	GO:0016363	nuclear matrix	-0.711	4/7
GO Cellular Components	GO:0098794	postsynapse	-0.711	7/27

DDD-XL

Category	Term	Description	Log(q-value)	InTerm_InList
GO Biological Processes	GO:0050808	synapse organization	-3.703	15/19
GO Biological Processes	GO:0031175	neuron projection development	-1.936	15/25
GO Biological Processes	GO:0016570	histone modification	-1.877	14/23
GO Cellular Components	GO:0098794	postsynapse	-4.197	18/27
GO Cellular Components	GO:0030424	axon	-2.613	15/25

SFARI-XL

Category	Term	Description	Log(q-value)	InTerm_InList
GO Biological Processes	GO:0050804	modulation of chemical synaptic transmission	-2.255	7/12
GO Biological Processes	GO:0016570	histone modification	-1.142	8/23
GO Biological Processes	GO:0006417	regulation of translation	-0.847	5/10
GO Cellular Components	GO:0098685	Schaffer collateral - CA1 synapse	-1.768	4/4
GO Cellular Components	GO:0035097	histone methyltransferase complex	-1.460	4/5

Table6. Re-run the GO enrichment analysis using chromosome X gene set as a reference.

2.3.2 Expression patterns of genes that follow dominant or recessive inheritance

In this study, we used GTEx data to determine if the expression profiles of KND genes reflected different biological process functions based on inheritance pattern. First, to view the genetic variation on gene expression levels, the expression distribution of all 56,200 genes in 54 tissues collected from GTEx was investigated. All of the GTEx protein-coding genes exhibit a wide range of expression and bimodal distribution in each tissue (Fig2.2). And then the difference in brain-specific expression distribution of KND genes by inheritance pattern was compared. As a control, the expression distribution of all database genes and annotated disease genes (OMIM) was used based on a random selection of genes that was equal in size to the number of KND genes obtained. Interestingly, both OMIM and KND genes revealed high levels of expression in both brain and non-brain tissue compared to all genes. Furthermore, the proportions of genes with higher expression in KND dominant genes are greater than the OMIM gene expression levels in the brain (Fig2.3). Further examination was carried out with the dominant gene set and recessive gene set of KND, as shown in Fig2.4, Fig2.5 and Fig2.6. The autosomal dominant gene set exhibited a higher gene expression levels in the brain, especially the cerebellum, compared to other tissues. Meanwhile, the recessive genes are widely expressed in most tissues and some of them displayed clearly distinct expression on the muscle (Fig2.4). The tissue-specific expression

pattern of KND dominant genes is highly expressed in cerebellar hemisphere, cerebellum, cortex, and frontal cortex (Fig2.5). By comparing non-brain tissue expression, relative brain expression was calculated. A value 0.5 indicates that there are no significant differences in expression levels between brain and non-brain tissues. By comparing all 41 non-brain tissues, the KND dominant gene set is found to have lower (< 0.5) or higher (> 0.5) brain expression. In contrast, relative brain expression of the KND recessive gene set is abundant in the midline in most of the 41 tissues (Fig2.6A). The violin plot showed that KND genes having brain-specific expression were more enriched in the dominant gene set compared to the recessive gene set ($P = 1.4 \times 10^{-7}$; Fig. 2.6B). Interestingly, the relative brain expression of X-linked genes was more similar to dominant genes than to recessive genes.

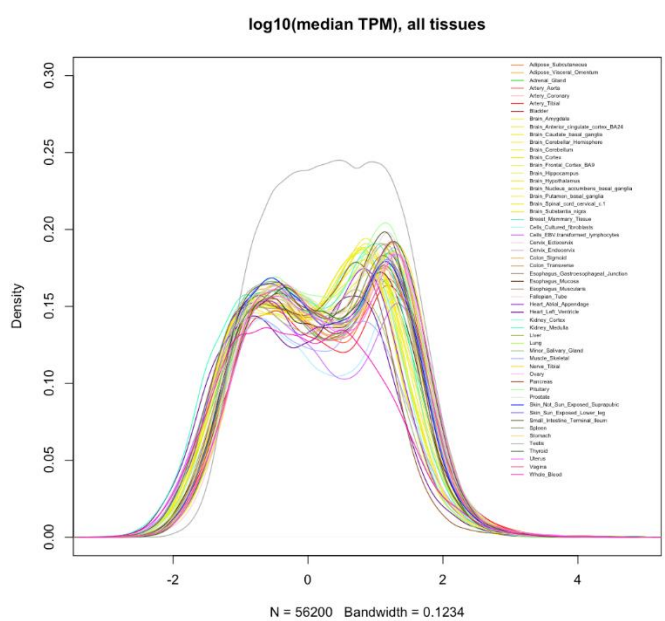


Fig2.2 Density plot of expression levels for all GTEx gene in 54 tissues, obtained from GTEx. The Y-axis shows the expression density of genes measured as transcript per million(TPM). The color was assigned to each tissue. Yellow color indicates brain tissue.

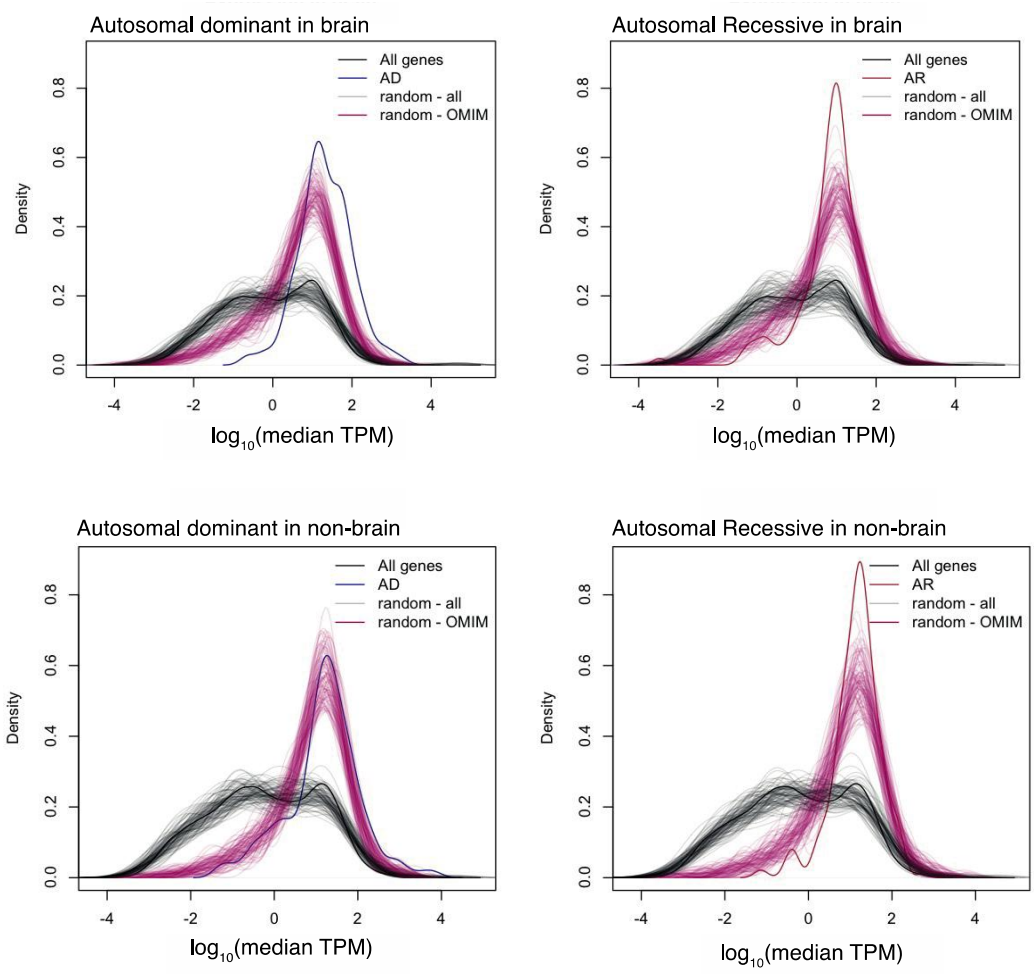


Fig2.3 Expression pattern of KND genes and random genes in brain and non-brain. The expression distribution of observed autosomal dominant and recessive KND genes. As a control, all genes, random all genes and random OMIM genes are used.

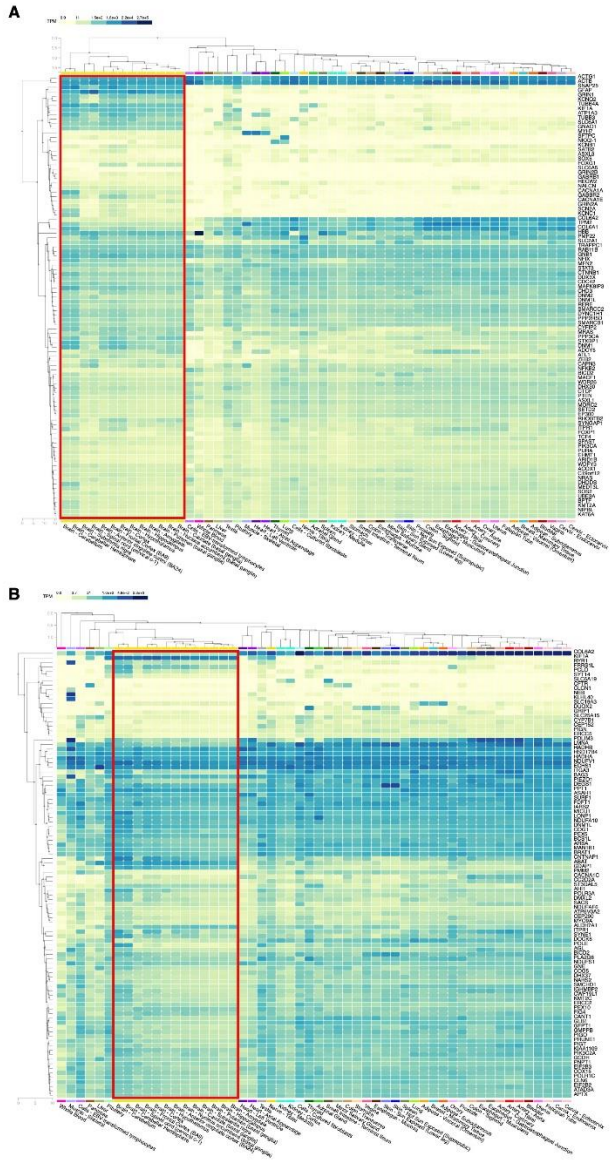


Fig2.4 Profile of global expression by the inheritance patterns. KND gene expression heatmap was constructed with GTEx tissues. The brain tissues are highlighted in red box. (A)autosomal dominant KND genes (B)autosomal recessive KND genes.

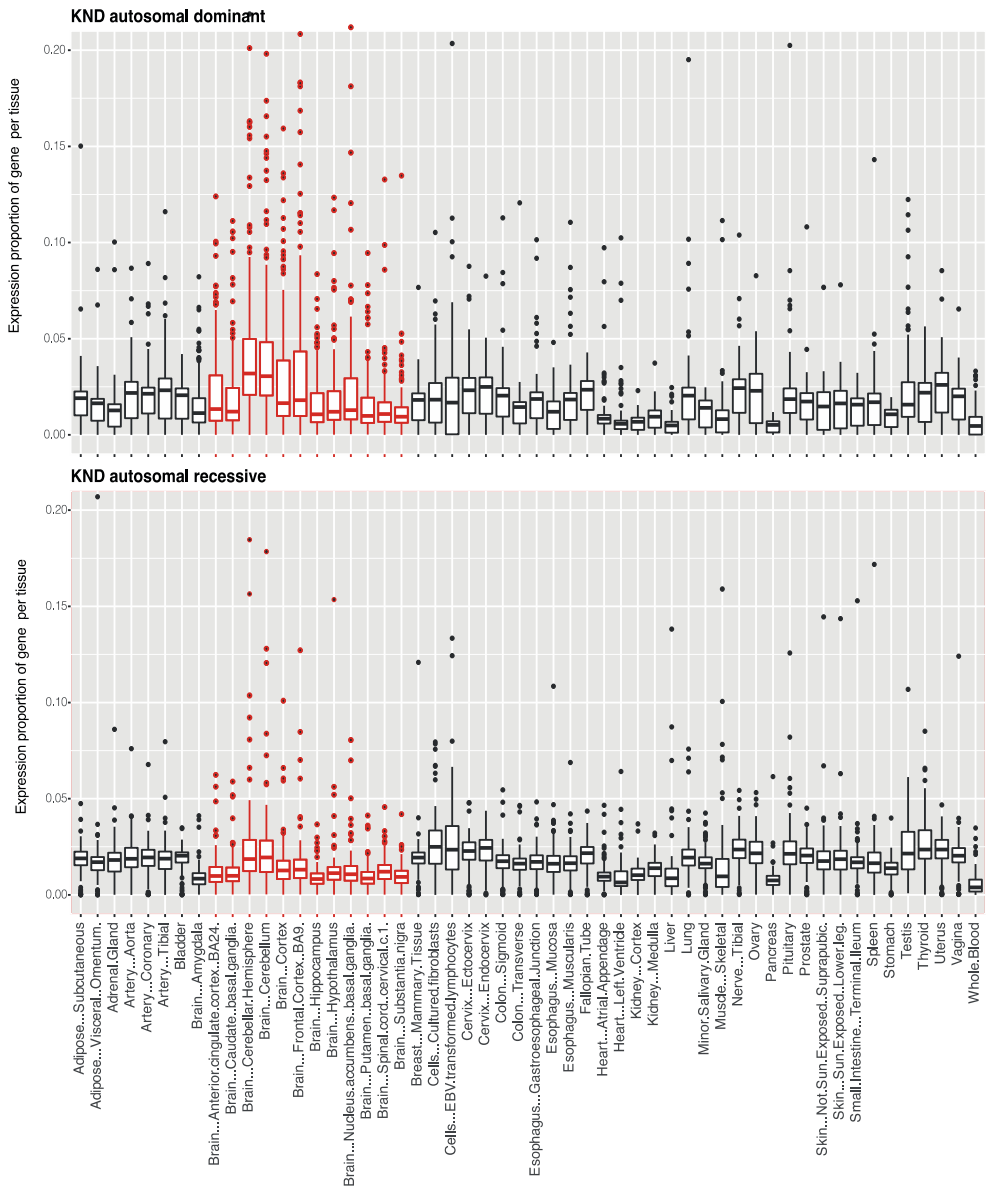


Fig2.5 expression proportions of KND genes based on total 54 GTEx tissue.

Each tissue's gene expression level was divided by the sum of the 54 total tissue expression levels. Brain tissues were highlighted by the red.

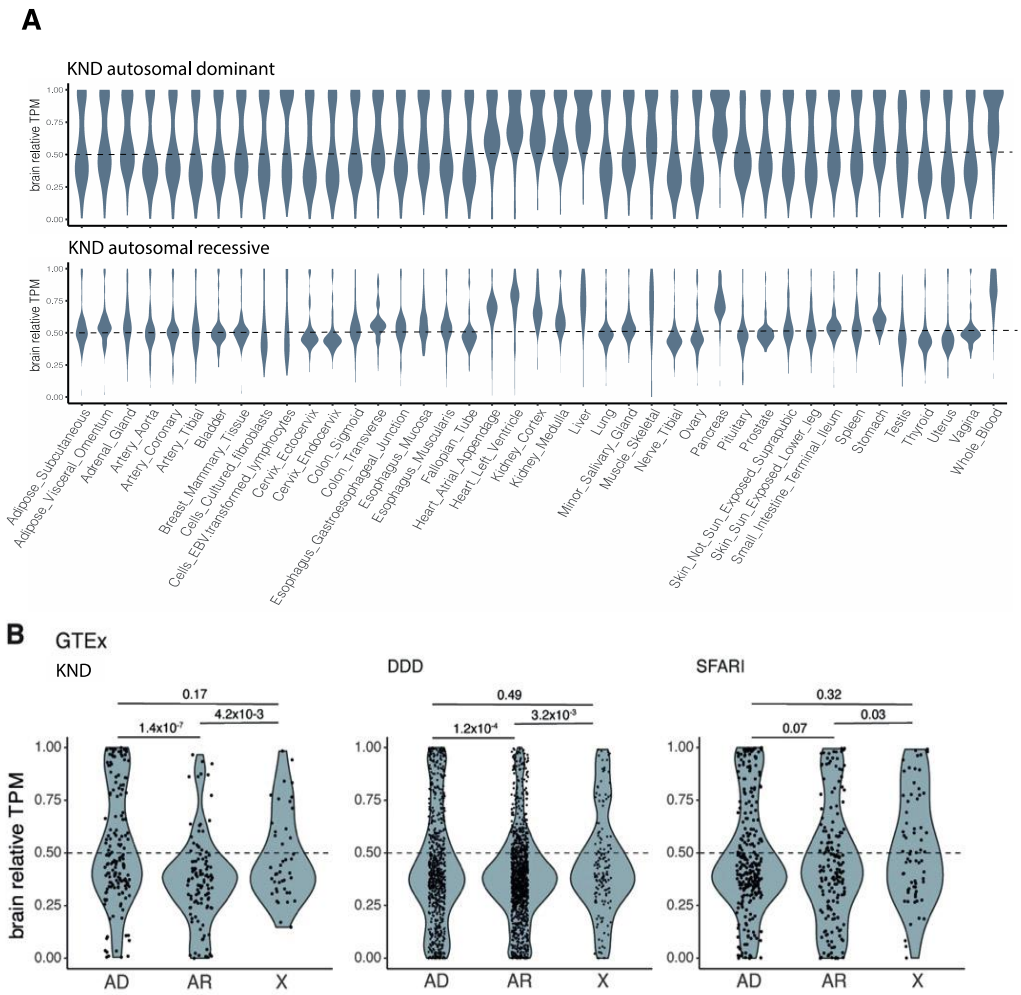
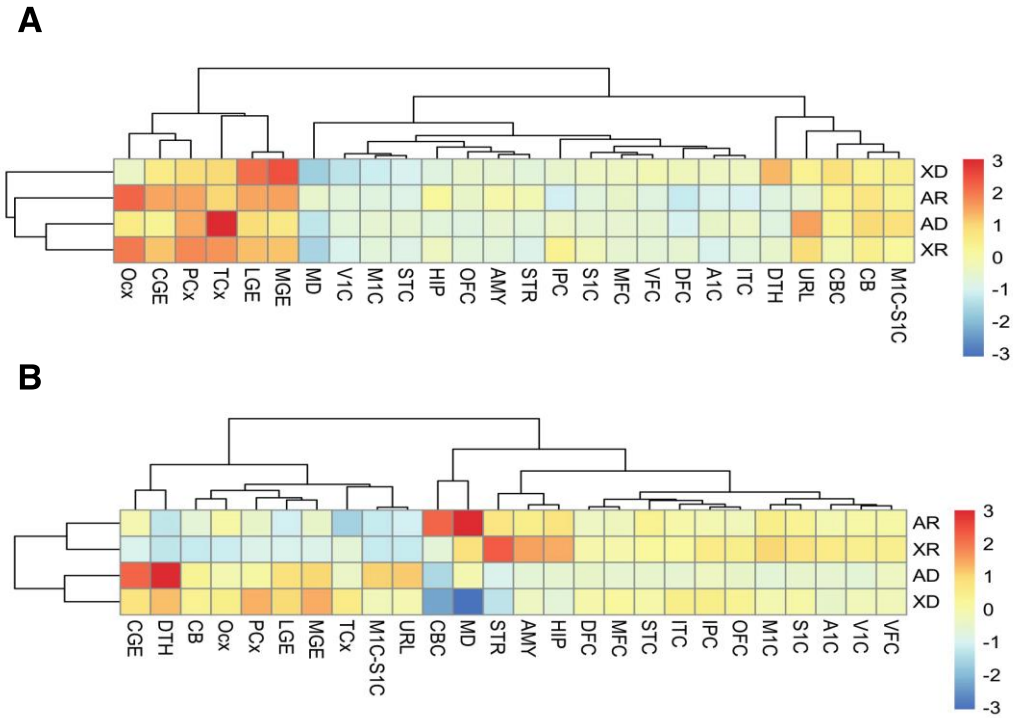


Fig2.6 Brain expression profiling of NDD causal genes by inheritance patterns.

(A)Brain expression value compared to each tissue expression value (B)Difference in relative enrichment in brain expression compared to other tissue types.

BrainSpan comprises a comprehensive survey of gene expression across brain regions and at different developmental time points. We could look into spatial and temporal dynamics; the average expression quantifications of 16 brain regions by age stages obtained from post-conception to adulthood are presented in heatmap form. The expression data were log normalized. According to the mean expression heatmap, genes in autosomal and X-linked dominant are expressed in caudal ganglionic eminence (CGE), dorsal thalamus (DTH), cerebellum (CB), occipital neocortex (Ocx), parietal neocortex (PCx), lateral ganglionic eminence (LGE), and medial ganglionic eminence (MGE), whereas the autosomal and X-linked recessive genes are expressed in cerebellar cortex (CBC), mediodorsal nucleus of thalamus (MD), striatum (STR), amygdaloid complex (AMY), hippocampus (HIP) (Fig2.7). In temporal expression data, dominant genes show constitutively higher expression across all brain developmental stages (Fig2.8A). In the KND RNA-seq dataset, dominant genes displayed increased expression level relative to both recessive and X-linked genes (between dominant and recessive genes, $P = 1.4 \times 10^{-9}$ for the prenatal period and $P = 3.5 \times 10^{-6}$ for the postnatal period: Fig2.8B). There is no obvious pattern of transition between the prenatal and postnatal stages. Temporal expression data showed significant differences in the expression patterns for KND genes of different inheritance modes, which imply distinct biological functions.



A1C	primary auditory cortex (core)	MD	mediodorsal nucleus of thalamus
AMY	amygdaloid complex	MFC	anterior (rostral) cingulate (medial prefrontal) cortex
CB	cerebellum	MGE	medial ganglionic eminence
CBC	cerebellar cortex	Ocx	occipital neocortex
CGE	caudal ganglionic eminence	OFC	orbital frontal cortex
DFC	dorsolateral prefrontal cortex	PCx	parietal neocortex
DTH	dorsal thalamus	S1C	primary somatosensory cortex (area S1, areas 3,1,2)
HIP	hippocampus (hippocampal formation)	STC	posterior (caudal) superior temporal cortex (area 22c)
IPC	posteroventral (inferior) parietal cortex	STR	striatum
ITC	inferolateral temporal cortex (area TEv, area 20)	TCx	temporal neocortex
LGE	lateral ganglionic eminence	URL	upper (rostral) rhombic lip
M1C	primary motor cortex (area M1, area 4)	V1C	primary visual cortex (striate cortex, area V1/17)
M1C-S1C	primary motor-sensory cortex (samples)	VFC	ventrolateral prefrontal cortex

Fig2.7 Differential spatial expression from BrainSpan 26 regions of brain.

KND gene expression variability by inheritance mode on different brain anatomical structures using median TPM (A) and mean TPM (B).

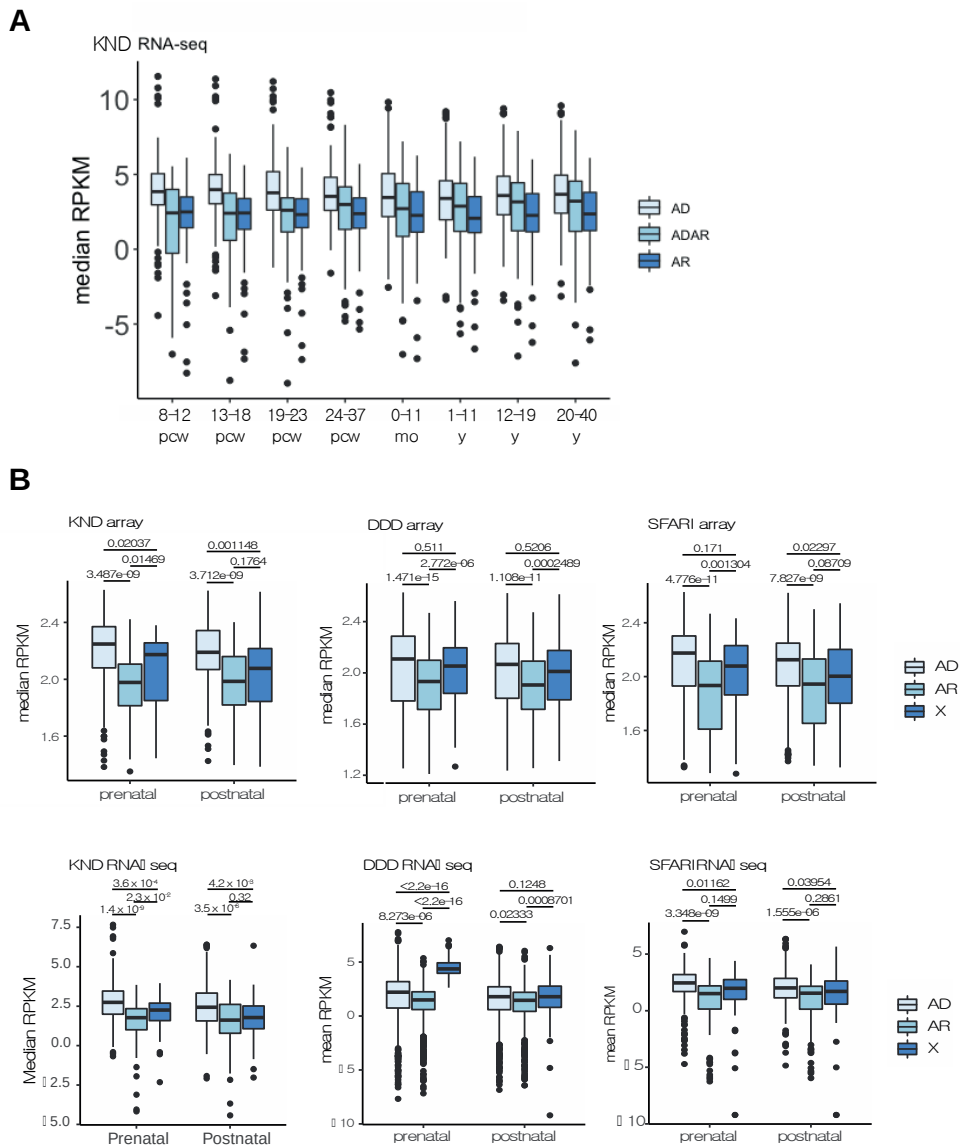


Fig2.8 Difference in median expression in brain regions divided by inheritance patterns. (A) KND gene expression in different developmental stages of brain. (B) Gene expression pattern compared with the DDD and SFARI databse genes.

2.3.3 Tissue-specific PPI networks

PPI information plays key role in predicting the biological function by exploring proteins that physically interact with other proteins. A recent study provided data on protein pairs that interacted in seven mouse tissues, which we used to identify PPIs for KND genes in tissue-specific context based on the idea that proteins can physically interact in tissue they express [82]. Although a number of interactions for the KND genes have not been identified in the brain, we observed that the fraction of genes with PPIs was greater among dominant genes (59/163 = 36.2%) than for recessive genes (26/117 = 22.2%), and the mean number of interactions was also higher (7.2 for dominant genes vs 4.5 for recessive genes) (Fig2.9A). Among those genes having PPIs in the brain, genes with interaction detected in only brain tissue comprise 30.5% of dominant genes (18/59), greater than recessive genes (3/26 = 11.5%) (Fig2.9B). On the other hand, most of interacting proteins for recessive genes were identified in various tissue and more than half interacted with other proteins in all seven tissues (14/26 = 53.8%; Fig2.9B), consistent with broadly expressed pattern, implying these genes to have a more ubiquitous functional pattern. PPIs from DDD and SFARI were also compared for validation and yielded similar patterns of brain-specific PPIs for dominant gene products and broader PPIs for recessive gene products.

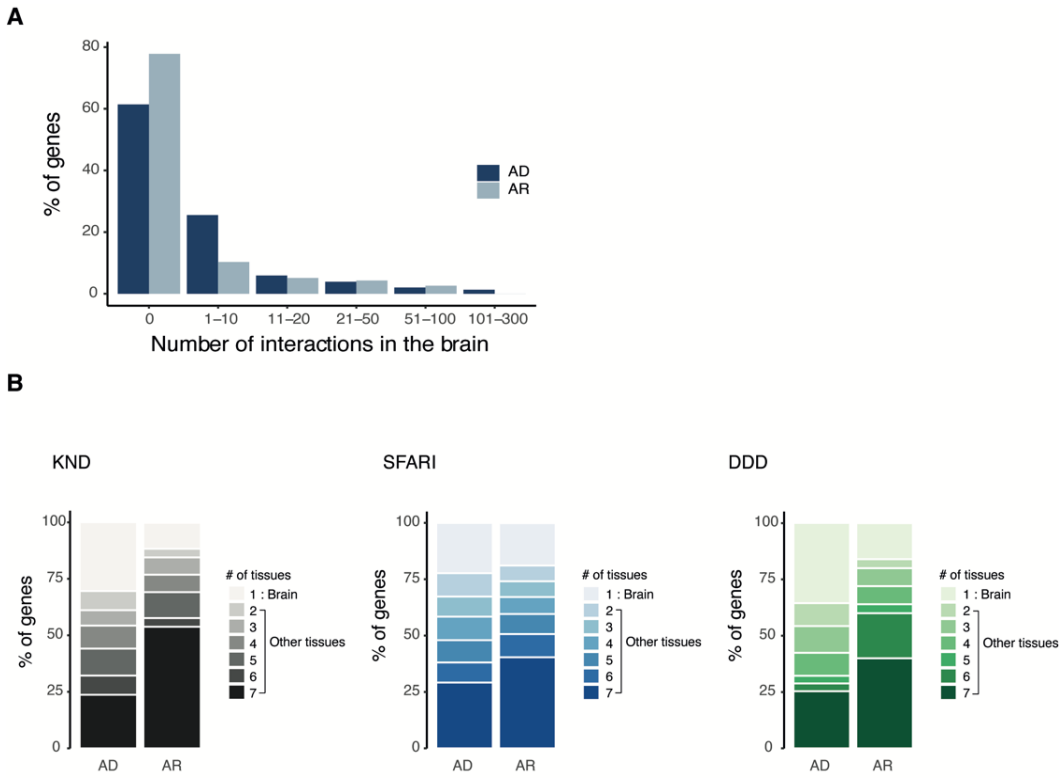


Fig2.9 Comparison of brain PPI network of NDD causal genes by inheritance patterns. (A) Number of genes with PPI events in the brain tissue and (B) proportion of genes having number of PPI-positive tissues. Tissues that correspond to two to seven are heart, kidney, liver, lung, muscle, and thymus.

2.3.4 Tolerance to pathogenic variants

Predicting haploinsufficient gene is essential for interpretation of genetic architecture in genetic studies. loss of function intolerance (pLI) and the observed/expected (O/E) constraint ratio scores in gnomAD represent the tolerance of a a given gene to protein-truncatin variation [42]. In KND, pLI is ~1 for autosomal domianant genes, representing strong constraint, while pLI is ~0 for autosomal recessive genes that shows the opposite trend (e.g., 75.0% of autosomal dominant genes are > 0.9, and 84.6% of autosomal recessive genes are < 0.1; Fig2.10). Consistent patterns were also observed for DDD and SFARI. Meanwhile, similar to the GO analysis findings, X-linked recessive genes exhibited patterns akin to dominant genes. These observations were recapitulated when using O/E values (Fig2.10). The dominant genes are significantly more constrained than the recessive genes for missense (o/e_mis) and LoF (o/e_lof). The Recessive genes are significantly shifted towards high values for both missense and LoF (Wilcoxon P < 10^{-16}), but synonymous variants created a mirrored distribution (oe_syn) (Fig2.10). All told, these findings suggest that genes responsible for NDDs harbor different functions according to their inheritance patterns, and they share little in terms of the molecular pathways leading to disease phenotype.

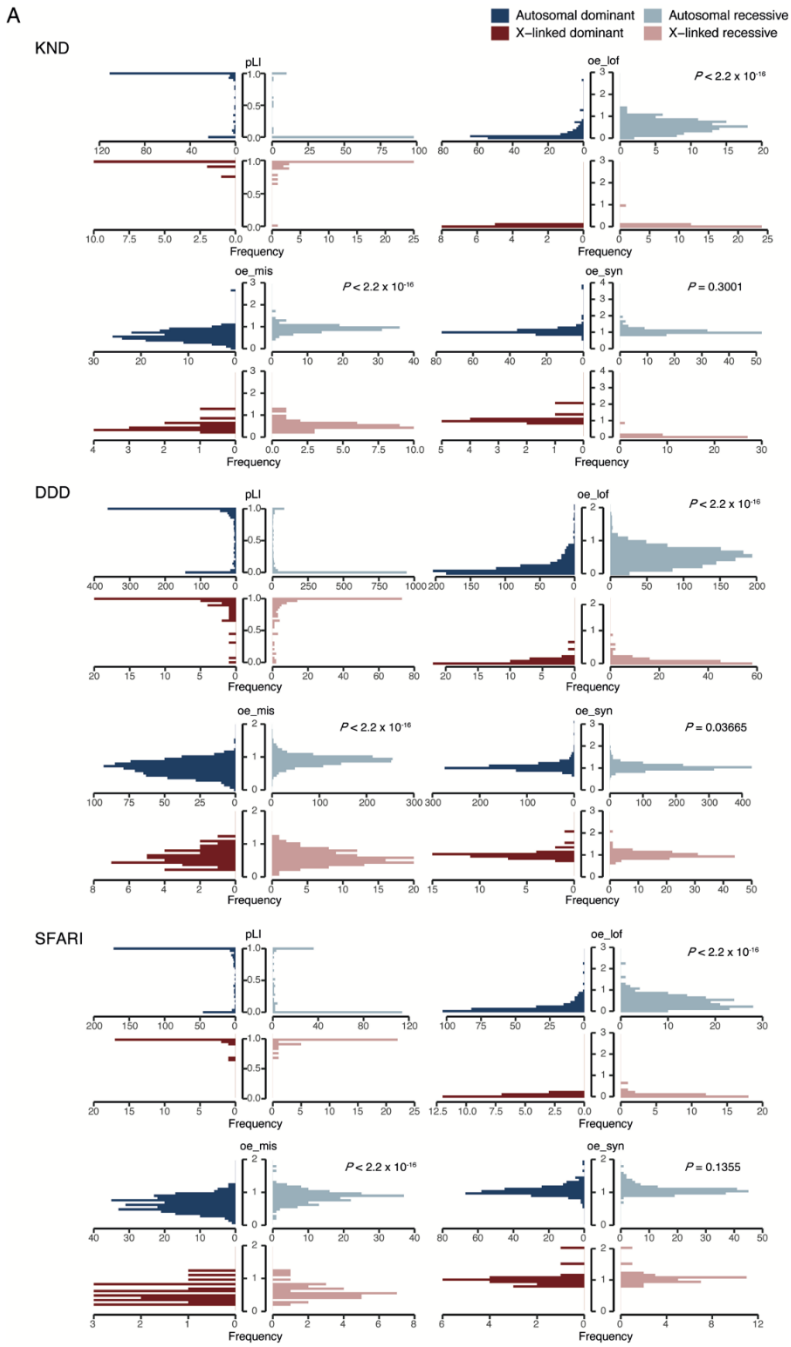


Fig2.10 Constraint score for KND, DDD and SFARI genes by inheritance patterns.

2.3.5 Characteristics of variants that follow dominant or recessive inheritance

Functional prediction scores like CADD, SIFT, and PhyloP was used to predict the impact of variants in KND patients according to their inheritance pattern. Those are very helpful in predicting to be harmful for protein structure and function. This analysis revealed that functional prediction scores for recessive variants tend to be lower than those of dominant variants (between dominant and recessive variants, P = 0.11 for CADD, P = 2.2 x 10^{-4} for SIFT, P = 9.6 x 10^{-6} for PhyloP, and P = 6.6 x 10^{-7} for AA conservation; Fig2.11). This finding indicates that variants under dominant inheritance with higher scores are more likely to be deleterious and variants under recessive inheritance are less damaging and less critical in function, hence demonstrate little physiological effect on carriers.

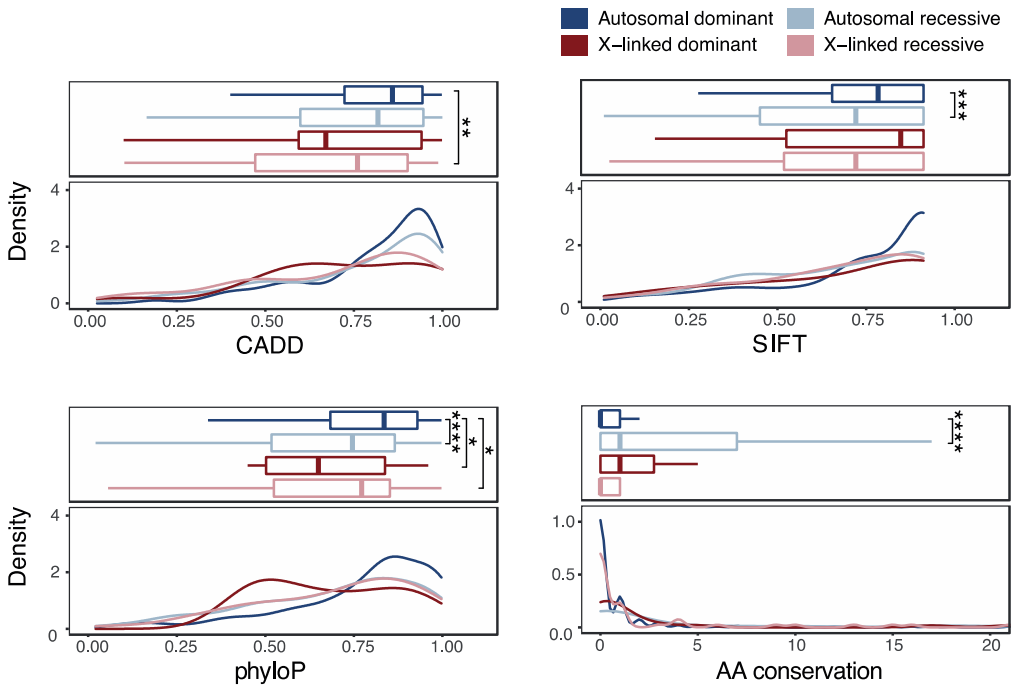


Fig2.11 Functional scores (CADD, SIFT and phyloP) and conservation scores among NDD causal variants, divided by inheritance patterns.

1.4 Discussion

Since NDDs may be caused by various alterations in genes with autosomal dominant, autosomal recessive, or X-linked inheritance modes, genotype-phenotype correlations are often difficult to establish. Furthermore, many disorders of the brain lack any clear unifying pathology at the molecular, cellular, or systems level. So, genetic profiling in NDDs is regarded as crucial for understanding the pathogenic mechanisms associated with the reported mutations and also offers an elucidation of the complexities associated with the wide genetic variability. In molecular terms, the same mutation may lead to different clinical manifestations (phenotypes), and mutations in different genes in the same or related pathways may lead to the same disorder. Since phenotype-genotype correlation has important implications for gene discovery, understanding the mechanisms will lead to the identification of novel genes and candidates that accelerate clinically actionable treatment. For example, copy number mutations in patients with autism may disrupt genes involved in the neuronal signaling pathway, which is an important network for brain development [83]. In addition, rare severe mutations in multiple genes important for brain development have been identified in patients with autism spectrum disorders [84]. Dominant and recessive, as medical terms might be able to be regarded as categories that help to understand or predict a phenotype on the cellular level as there may be fundamental differences. Analysis of the Deciphering

Developmental Disorders (DDD) study of 7448 intellectual disability (ID) cases revealed a small contribution of recessive disorders (3.6%) in outbred populations [56]. Unlike metabolic disorders, autosomal recessive ID is less prevalent in outbred populations. Metabolic disorders follow mostly autosomal recessive (AR) and less commonly autosomal dominant patterns of inheritance [85, 86]. There are hundreds of different genetic metabolic disorders, and most people with inherited metabolic disorders have a defective gene that results in an enzyme deficiency. Jimenez-Sanchez, G., et al., tried to identify correlations between the function of the gene product and features of disease like age of onset and mode of inheritance mode [87]. They categorized each disease gene according to the function of its protein product. Online Mendelian Inheritance in Men (OMIM) compendium, 4,617 genes and their variants are associated with human disease as of April 2022 [29, 30] Interestingly, comparison of the inheritance patterns shows that disorders caused by genes encoding enzymes are primarily recessive, whereas disorders caused by genes encoding transcription factors are more likely to be dominant. These correlations provide biological support for the validity of the functional characterization, and they hint at additional principles of disease.

A rare mutation that disrupts the function of genes operating in critical neurodevelopmental pathways may lead to a wide range of pleiotropic effects. There are thousands of candidate genes for NDDs, given that most human genes are expressed in the brain. The characterization of the functional consequences of

these mutations will help to elucidate both normal brain development and neurodevelopmental processes leading to disease [88]. As a first step, biological pathways enriched for KND, DDD and SFARI genes represent brain developmental progression, but gene groups associated with distinct inheritance modes reveal different biological processes (Fig2.1). The results revealed that dominant and recessive genes are most strongly associated with synaptic function and metabolic processes, respectively, implying that diseases can be caused through different molecular mechanisms according to their inheritance patterns. Moreover, we observed dominant and recessive gene sets to have opposite trends in pLI and O/E scores: haploinsufficient genes are highly vulnerable to disruptive mutations, which proved the differences in genetic architecture between these inheritance patterns (Fig2.10). These genes commonly exhibit heterozygous loss-of-function mutations that are sufficient to elicit NDDs. Recent large-scale sequencing efforts to identify human genes that are intolerant to heterozygous loss-of-function have yielded valuable insight into their role in disease etiology [42, 89, 90]. In addition, information about multiple differences in the expression of a set of genes provides understanding of the molecular mechanisms underlying their function. Many changes in gene expression have been associated with developmental and behavioral disorders in different tissue types [91-93]. Many studies have explored gene expression profiles in neurological disorders [94-96], but none of them focuses on function in a framework of disease-associate genes

based on inheritance. In this study, the comparisons were made either between brain and non-brain tissues or between dominant and recessive inheritance. As a result, gene expression profiles reflected this fundamental difference. The profiling of brain expression patterns in GTEx and BrainSpan revealed that the expression profile of annotated disease genes (OMIM genes) differs significantly from the expression profile of all genes. All genes show broad patterns of gene expression. More highly expressed genes are found in OMIM genes than in all genes in both brain and non-brain. These observations mean OMIM genes might be implicated in core signaling pathways that might have large effects on disease status. Importantly, the switch in expression level distribution of the KND dominant genes demonstrates elevated expression in the brain compared to other tissue types. (Fig2.3, Fig2.5, Fig2.6). The dominant gene set to exhibit specific and increased expression in the brain compared to the recessive gene set, suggesting dominant genes to be more brain-specific (Fig2.6). This result could be used to decipher the link between biological mechanisms and the regulation of gene expression. Besides, these genes are expressed consistently across different brain regions and developmental time points. The brain is a complex organ that is comprised of several anatomical substructures. Transcriptome studies across brain regions suggest that gene expression patterns across brain regions are related to both functional and anatomical differences in their substructures. The cerebellum, in particular, has the most distinct gene expression pattern [97, 98]. Brain regional

expression data shows that KND dominant genes are especially enriched in their expression in the cerebellum and cortex (Fig2.5, Fig2.7). PPI data supported an association of tissue-specific expression and function with inheritance mode. Tissue-specific PPI networks based on direct interactions have previously demonstrated biological relevance [82]. Here, brain-specific interactions predominantly arise from an autosomal dominant gene set, consistent with the earlier gene expression profile results. In contrast, recessive genes tended to have interactions that were ubiquitous across all seven tissues (Fig2.9). Therefore, combined biological studies, including PPI networks, functional pathways, and phenotype data, will strengthen our understanding of disease progression in NDD. Also, variant functional impact that may lead to molecular changes was investigated, and it was found that variants with recessive pathogenic alleles were less deleterious than those with dominant alleles (Fig2.11). This is well supported by the fact that variants occur in both copies of the gene to contribute to a phenotype, and parental carriers are mostly healthy, although recent large-scale analyses have revealed heterozygous carriers of rare diseases to harbor subtle effects in various aspects of individual health and reproductivity [99, 100]. Taken together, a comprehensive functional analysis of tissue provides more detailed clues on gene function.

The findings from this study, of course, will have a potential problem of false positive diagnosis in analyzing variants causing diseases. But pathogenic variants

were carefully evaluated by our clinical and scientific teams following the recommendation of the ACMG guidelines. Therefore, the result was interpreted based on a confident diagnostic decision.

References

1. Yang, Y., et al., *Clinical whole-exome sequencing for the diagnosis of mendelian disorders.* N Engl J Med, 2013. **369**(16): p. 1502-11.

2. Iglesias, A., et al., *The usefulness of whole-exome sequencing in routine clinical practice.* Genet Med, 2014. **16**(12): p. 922-31.

3. Lee, H., et al., *Clinical exome sequencing for genetic identification of rare Mendelian disorders.* Jama, 2014. **312**(18): p. 1880-7.

4. Srivastava, S., et al., *Clinical whole exome sequencing in child neurology practice.* Ann Neurol, 2014. **76**(4): p. 473-83.

5. Yang, Y., et al., *Molecular findings among patients referred for clinical whole-exome sequencing.* Jama, 2014. **312**(18): p. 1870-9.

6. Retterer, K., et al., *Clinical application of whole-exome sequencing across clinical indications.* Genet Med, 2016. **18**(7): p. 696-704.

7. Trujillano, D., et al., *Clinical exome sequencing: results from 2819 samples reflecting 1000 families.* Eur J Hum Genet, 2017. **25**(2): p. 176-182.

8. Gilissen, C., et al., *Genome sequencing identifies major causes of severe intellectual disability.* Nature, 2014. **511**(7509): p. 344-7.

9. Clark, M.M., et al., *Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases.* NPJ Genom Med, 2018. **3**: p. 16.

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

11. Manickam, K., et al., *Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG).* Genet Med, 2021. **23**(11): p. 2029-2037.

12. Blesson, A. and J.S. Cohen, *Genetic counseling in neurodevelopmental disorders.* Cold Spring Harbor Perspectives in Medicine, 2020. **10**(4): p. a036533.

13. Srivastava, S., et al., *Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders.* Genetics in Medicine, 2019. **21**(11): p. 2413-

2421.

14. Wright, C.F., et al., *Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders.* Genetics in Medicine, 2018. **20**(10): p. 1216-1223.

15. McRae, J., et al., *Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders.* Nature, 2017. **542**(7642): p. 433-438.

16. Gahl, W.A., et al., *The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine.* Mol Genet Metab, 2016. **117**(4): p. 393-400.

17. Beaulieu, C.L., et al., *FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project.* Am J Hum Genet, 2014. **94**(6): p. 809-17.

18. Baynam, G., et al., *The rare and undiagnosed diseases diagnostic service - application of massively parallel sequencing in a state-wide clinical service.* Orphanet J Rare Dis, 2016. **11**(1): p. 77.

19. Lee, Y., et al., *Genomic profiling of 553 uncharacterized neurodevelopment patients reveals a high proportion of recessive pathogenic variant carriers in an outbred population.* Scientific Reports, 2020. **10**(1): p. 1413.

20. Al-Nabhani, M., et al., *Reanalysis of exome sequencing data of intellectual disability samples: yields and benefits.* Clinical genetics, 2018. **94**(6): p. 495-501.

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

22. Deignan, J.L., et al., *Points to consider in the reevaluation and reanalysis of genomic test results: a statement of the American College of Medical Genetics and Genomics (ACMG).* Genetics in Medicine, 2019. **21**(6): p. 1267-1270.

23. Al-Nabhani, M., et al., *Reanalysis of exome sequencing data of intellectual disability samples: Yields and benefits.* Clin Genet, 2018. **94**(6): p. 495-501.

24. Li, J., et al., *Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation.* Gene, 2019. **700**: p. 168-175.

25. Shashi, V., et al., *A comprehensive iterative approach is highly effective in diagnosing individuals who are exome negative.* Genet Med, 2019. **21**(1): p. 161-172.

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

27. Fung, J.L.F., et al., *A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis.* NPJ Genom Med, 2020. **5**(1): p. 37.

28. Liu, P., et al., *Reanalysis of Clinical Exome Sequencing Data.* N Engl J Med, 2019. **380**(25): p. 2478-2480.

29. Amberger, J.S., et al., *OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders.* Nucleic Acids Res, 2015. **43**(Database issue): p. D789-98.

30. Amberger, J.S. and A. Hamosh, *Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes.* Curr Protoc Bioinformatics, 2017. **58**: p. 1.2.1-1.2.12.

31. Chan, O.Y.M., et al., *Expanded carrier screening using next-generation sequencing of 123 Hong Kong Chinese families: a pilot study.* Hong Kong Med J, 2021. **27**(1): p. 177-183.

32. Singh, K., et al., *NGS-based expanded carrier screening for genetic disorders in North Indian population reveals unexpected results – a pilot study.* BMC Medical Genetics, 2020. **21**(1): p. 216.

33. Xiao, Q. and V.M. Lauschke, *The prevalence, genetic complexity and population-specific founder effects of human autosomal recessive disorders.* npj Genomic Medicine, 2021. **6**(1): p. 41.

34. Kaback, M.M., *Screening and prevention in Tay-Sachs disease: origins, update, and impact.* Adv Genet, 2001. **44**: p. 253-65.

35. Cao, A. and Y.W. Kan, *The prevention of thalassemia.* Cold Spring Harb Perspect Med, 2013. **3**(2): p. a011775.

36. Lee, S., et al., *Korean Variant Archive (KOVA): a reference database of genetic variations in the Korean population.* Scientific Reports, 2017. **7**(1): p. 4287.

37. Broad Institute. *Picard toolkit*. 2019; Available from: https://broadinstitute.github.io/picard/.

38. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

39. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce*

framework for analyzing next-generation DNA sequencing data. Genome Res, 2010. **20**(9): p. 1297-303.

40. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

41. Cingolani, P., et al., *Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift.* Front Genet, 2012. **3**: p. 35.

42. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

43. Rentzsch, P., et al., *CADD: predicting the deleteriousness of variants throughout the human genome.* Nucleic Acids Res, 2019. **47**(D1): p. D886-d894.

44. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.

45. Siepel, A., et al., *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes.* Genome Res, 2005. **15**(8): p. 1034-50.

46. Amberger, J.S., et al., *OMIM.org: leveraging knowledge across phenotype-gene relationships.* Nucleic Acids Res, 2019. **47**(D1): p. D1038-d1043.

47. Stenson, P.D., et al., *Human Gene Mutation Database (HGMD): 2003 update.* Hum Mutat, 2003. **21**(6): p. 577-81.

48. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants.* Nucleic Acids Res, 2016. **44**(D1): p. D862-8.

49. Wright, C.F., et al., *Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data.* Lancet, 2015. **385**(9975): p. 1305-14.

50. Kuperberg, M., et al., *Utility of Whole Exome Sequencing for Genetic Diagnosis of Previously Undiagnosed Pediatric Neurology Patients.* J Child Neurol, 2016. **31**(14): p. 1534-1539.

51. Marinakis, N.M., et al., *Phenotype-driven variant filtration strategy in exome sequencing toward a high diagnostic yield and identification of 85 novel variants in 400 patients with rare Mendelian disorders.* Am J Med Genet A, 2021. **185**(8): p. 2561-2571.

52. Ewans, L.J., et al., *Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders.*

Genet Med, 2018. **20**(12): p. 1564-1574.

53. Epilepsy Genetics Initiative, *The Epilepsy Genetics Initiative: Systematic reanalysis of diagnostic exomes increases yield.* Epilepsia, 2019. **60**(5): p. 797-806.

54. Jalkh, N., et al., *The added value of WES reanalysis in the field of genetic diagnosis: lessons learned from 200 exomes in the Lebanese population.* BMC Med Genomics, 2019. **12**(1): p. 11.

55. Lee, S., et al., *Korean Variant Archive (KOVA): a reference database of genetic variations in the Korean population.* Sci Rep, 2017. **7**(1): p. 4287.

56. Martin, H.C., et al., *Quantifying the contribution of recessive coding variation to developmental disorders.* Science, 2018. **362**(6419): p. 1161-1164.

57. Herman, I., et al., *Clinical exome sequencing in the diagnosis of pediatric neuromuscular disease.* Muscle Nerve, 2021. **63**(3): p. 304-310.

58. Rozen, P., et al., *Prevalence of the I1307K APC gene variant in Israeli Jews of differing ethnic origin and risk for colorectal cancer.* Gastroenterology, 1999. **116**(1): p. 54-7.

59. Lynch, H.T., W.S. Rubinstein, and G.Y. Locker, *Cancer in Jews: introduction and overview.* Fam Cancer, 2004. **3**(3-4): p. 177-92.

60. Lazarin, G.A., et al., *An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals.* Genet Med, 2013. **15**(3): p. 178-86.

61. Zlotogora, J., *Multiple mutations responsible for frequent genetic diseases in isolated populations.* European Journal of Human Genetics, 2007. **15**(3): p. 272-278.

62. Deuchler, M., *The Confucian Transformation of Korea : A Study of Society and Ideology*. 1 ed. Vol. 36. 1992: Harvard University Asia Center.

63. Jimenez-Sanchez, G., B. Childs, and D. Valle, *Human disease genes.* Nature, 2001. **409**(6822): p. 853-5.

64. Schuster-Böckler, B. and A. Bateman, *Protein interactions in human genetic diseases.* Genome Biol, 2008. **9**(1): p. R9.

65. Crippa, M., et al., *SETD5 Gene Haploinsufficiency in Three Patients With Suspected KBG Syndrome.* Frontiers in Neurology, 2020. **11**.

66. Coutinho, P., et al., *Clinical heterogeneity of autosomal recessive spastic paraplegias: analysis of 106 patients in 46 families.* Arch Neurol, 1999. **56**(8): p.

943-9.

67. Fink, J.K., *Hereditary spastic paraplegia.* Current Neurology and Neuroscience Reports, 2006. **6**(1): p. 65-76.

68. Salinas, S., et al., *Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms.* Lancet Neurol, 2008. **7**(12): p. 1127-38.

69. Blackstone, C., C.J. O'Kane, and E. Reid, *Hereditary spastic paraplegias: membrane traffic and the motor pathway.* Nat Rev Neurosci, 2011. **12**(1): p. 31-42.

70. Crosby, A.H. and C. Proukakis, *Is the transportation highway the right road for hereditary spastic paraplegia?* Am J Hum Genet, 2002. **71**(5): p. 1009-16.

71. Blackstone, C., *Cellular pathways of hereditary spastic paraplegia.* Annu Rev Neurosci, 2012. **35**: p. 25-47.

72. Schapira, A.H., *Mitochondrial involvement in Parkinson's disease, Huntington's disease, hereditary spastic paraplegia and Friedreich's ataxia.* Biochim Biophys Acta, 1999. **1410**(2): p. 159-70.

73. Beaudin, M., et al., *Systematic review of autosomal recessive ataxias and proposal for a classification.* Cerebellum & Ataxias, 2017. **4**(1): p. 3.

74. Beaudin, M., et al., *The Classification of Autosomal Recessive Cerebellar Ataxias: a Consensus Statement from the Society for Research on the Cerebellum and Ataxias Task Force.* The Cerebellum, 2019. **18**(6): p. 1098-1125.

75. Banerjee-Basu, S. and A. Packer, *SFARI Gene: an evolving database for the autism research community.* Dis Model Mech, 2010. **3**(3-4): p. 133-5.

76. Abrahams, B.S., et al., *SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs).* Mol Autism, 2013. **4**(1): p. 36.

77. *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.

78. GTEx Consortium, *The GTEx Consortium atlas of genetic regulatory effects across human tissues.* Science, 2020. **369**(6509): p. 1318-1330.

79. Kang, H.J., et al., *Spatio-temporal transcriptome of the human brain.* Nature, 2011. **478**(7370): p. 483-9.

80. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

81. Fuller, Z.L., et al., *Measuring intolerance to mutation in human genetics.* Nat Genet, 2019. **51**(5): p. 772-776.

82. Skinnider, M.A., et al., *An atlas of protein-protein interactions across mouse tissues.* Cell, 2021. **184**(15): p. 4073-4089.e17.

83. Bucan, M., et al., *Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes.* PLoS Genet, 2009. **5**(6): p. e1000536.

84. Guilmatre, A., et al., *Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation.* Arch Gen Psychiatry, 2009. **66**(9): p. 947-56.

85. Wertheim-Tysarowska, K., et al., *Genetic analysis in inherited metabolic disorders--from diagnosis to treatment. Own experience, current state of knowledge and perspectives.* Dev Period Med, 2015. **19**(4): p. 413-31.

86. Zschocke, J., *Dominant versus recessive: molecular mechanisms in metabolic disease.* J Inherit Metab Dis, 2008. **31**(5): p. 599-618.

87. Jimenez-Sanchez, G., B. Childs, and D. Valle, *Human disease genes.* Nature, 2001. **409**(6822): p. 853-855.

88. Geschwind, D.H., *Autism: many genes, common pathways?* Cell, 2008. **135**(3): p. 391-5.

89. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* Nature, 2020. **581**(7809): p. 434-443.

90. Kaplanis, J., et al., *Evidence for 28 genetic disorders discovered by combining healthcare and research data.* Nature, 2020. **586**(7831): p. 757-762.

91. Mudge, J., et al., *Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum.* PLoS One, 2008. **3**(11): p. e3625.

92. Gregg, J.P., et al., *Gene expression changes in children with autism.* Genomics, 2008. **91**(1): p. 22-9.

93. Buechel, H.M., et al., *Deep sleep and parietal cortex gene expression changes are related to cognitive deficits with age.* PLoS One, 2011. **6**(4): p. e18387.

94. Winkler, J.M. and H.S. Fox, *Transcriptome meta-analysis reveals a central role for sex steroids in the degeneration of hippocampal neurons in Alzheimer's disease.* BMC Syst Biol, 2013. **7**: p. 51.

95. de Jong, S., et al., *A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes.* PLoS One, 2012. **7**(6): p. e39498.

96. Campbell, M.G., I.S. Kohane, and S.W. Kong, *Pathway-based outlier method reveals heterogeneous genomic structure of autism in blood transcriptome.* BMC Med Genomics, 2013. **6**: p. 34.

97. Strand, A.D., et al., *Conservation of regional gene expression in mouse and human brain.* PLoS Genet, 2007. **3**(4): p. e59.

98. Roth, R.B., et al., *Gene expression analyses reveal molecular relationships among 20 regions of the human CNS.* Neurogenetics, 2006. **7**(2): p. 67-80.

99. Barton, A.R., et al., *A spectrum of recessiveness among Mendelian disease variants in UK Biobank.* MedRxiv, 2021.

100. Gardner, E.J., et al., *Reduced reproductive success is associated with selective constraint on human genes.* Nature, 2022. **603**(7903): p. 858-863.

List of abbreviations

OMIM, the Online Mendelian Inheritance in Men

GoF, gain-of-function

LoF, loss-of-function

NDD, neurodevelopmental disorder

WES, whole exome sequencing

KND, Korean Neurodevelopmental Disorder

DDD, Deciphering Developmental Delay

SFARI, Simons Foundation Autism Research Initiative

SNUH, the Seoul National University Hospital

BWA, Burrows-wheeler Aligner

GATK, Genome Analysis Toolkit

TPM, Transcripts Per Million

GTEx, the Genotype-Tissue Expression

PPI, Protein-protein interaction

CNV, Copy number variation

GO, Gene ontology

pLI, the probability of loss of function intolerance

O/E, the observed/expected

KOVA, Korean Variant Archive

SNUCH, the Seoul National University Children's Hospital

ID, intellectual disability

GDD, global developmental delay

CHARGE, coloboma, heart defects, atresia choanae, growth retardation, genital abnormalities, and ear abnormalities

FD, facial dysmorphism

EE, epileptic encephalopathy

BP, biological process

CC, cellular component

MF, molecular function

DO, disease ontology

AD, autosomal dominant

AR, autosomal recessive

XL, X-linked

국문초록

신경발달질환은 뇌발달 과정의 주요 유전자 기능의 다양한 변형으로 인해 유발된다. 수많은 유전자들이 신경발달질환을 일으키는 원인으로 작용하기 때문에 유전자와 발병변이에 대한 포괄적인 이해가 굉장히 중요하다. 그러나 어떻게 변이가 멘델성 유전질환에 기여하는지에 대한 체계적인 이해는 여전히 부족하다. 이에 본 연구에서는 신경학적 증상을 가지는 환자들의 유전적 변이를 밝히기 위해 전장 엑솜 염기서열 분석을 수행하였다. 확실한 발병변이를 가지는 환자들의 진단율은 50.8% 였으며, 이는 재분석 과정을 거친 후 확인된 추가진단율 5.9% 를 포함한 수치이다. 진단된 환자들 중 33.4% 는 물려받은 유전변이를 가지고 있었다. 유전자 온톨로지 분석에서 상염색체 열성유전자는 대사과정, 근육구성, 금속이온 항상성 패스웨이 등의 특징을 나타내고 있었다. 대부분의 상염색체 열성유전자는 pLI 가 0 이었고 열성변이의 기능예측 스코어는 우성유전변이들보다 낮은 편이었다. 전사체와 상호작용체의 프로파일링은 유전양상에 따른 조직 특이적 발현과 단백질 간의 상호작용의 차이를 밝혔다. 게다가, gnomAD 와 KOVA 데이터베이스를 이용하여 열성유전변이를 가지는 보인자의 비율을 예측하였다. 그 결과 신경발달질환의 원일이 되는 유전자가 유전양상에 따라 다른 분자메커니

증과 발현패턴을 가지고 있다는 것을 보여준다. 또한 신경발달질환 보인자들을 미리 스크리닝 하는데 열성변이들의 계산된 빈도율을 활용할 수 있다. 이 연구는 분자생물학적 경로, 조직특이적 발현양상, 단백질간 상호작용등의 종합적인 관점에서의 신경발달질환과 표현형간의 관계를 이해하는데 목표를 두었다.

*본 학위 논문은 출판된 논문 (Park *et al*., 2022 *Front Genet*, 13, 99015)을 기반으로 작성되었음.

주요어: 신경발달질환, 유전패턴, 보인자 예측, 전장 엑솜 시퀀싱, 열성 질환

학 번: 2016-38160

선택과 통합적 임상 표현형 및 유전형 분석이 매우 중요하다.