의학박사 학위논문

Genetic and Functional Analysis on the Variants of Undiagnosed Pediatric Neurodevelopmental Disorders

미진단 소아 신경발달질환의 변이에 관한 유전학적 분석과 기능 연구

2020 년 8 월

서울대학교 대학원

의과학과 의과학전공

이 영 하

Genetic and Functional Analysis on the Variants of Undiagnosed Pediatric Neurodevelopmental Disorders

미진단 소아 신경발달질환의 변이에 관한 유전학적 분석과 기능 연구

지도교수 최 무 림

이 논문을 의학박사 학위논문으로 제출함

2020 년 5 월

서울대학교 대학원

의과학과 의과학전공

이 영 하

이영하의 박사 학위논문을 인준함

2020 년 7 월

위 원 장 [illegible]

부 위 원 장 최 무 림

위 원 [illegible]

위 원 [illegible]

위 원 [illegible]

Genetic and Functional Analysis on the Variants of Undiagnosed Pediatric Neurodevelopmental Disorders

by

Youngha Lee

A thesis submitted to the Department of Biomedical Sciences
in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy in Medical Science
at Seoul National University College of Medicine

July 2020

Approved by Thesis Committee:

Professor Chae Jongho Chairman
Professor Mirim Choi Vice Chairman
Professor Yong-Seok Lee
Professor JI-YEOB CHOI
Professor Lee Jin Sook

ABSTRACT

Genetic and Functional Analysis on the Variants of Undiagnosed Pediatric Neurodevelopmental Disorders

Youngha Lee
Biomedical Sciences Major
Department of Biomedical Sciences
Seoul National University

Recent advances in next-generation sequencing (NGS) technology led to many advances in understanding the pathogenicity of rare diseases. Whole exome sequencing (WES), based on NGS, has now become an indispensable tool in clinical diagnosis of mendelian disorders. The purpose of this thesis is to analyze 553 undiagnosed pediatric neurodevelopmental disorder patients, to identify disease-causing variants and apply them to clinical treatment.

A large portion of rare Mendelian disorder patients suffers from genetic variants that are inherited in a recessive manner. This study focuses on a detailed understanding of these recessive variants. Although there have been many studies on de novo variants that cause rare diseases, a systematic understanding of the contribution of recessive variants to Mendelian diseases is still lacking. Therefore, genetic diagnosis and variant discovery of 553 undiagnosed Korean patients with complex neurodevelopmental problems (Korean NeuroDevelopmental cohort; KND) were performed using WES of patients and their families. The diagnostic yield of our WES analysis, including previously known pathogenic genes and CNVs, was 47.4 %. In addition, the newly discovered causative

genes in this study were closely associated to the known genes in the brain developmental process. Among the patients with the previously reported pathogenic variants, 35.1% inherited these variants in a recessive manner. Genes that cause recessive disease in KND cohort tend to be less constrained by loss-of-function variants and were enriched in lipid metabolism and mitochondrial components. These observations and some assumptions were applied to an estimation that approximately 1 in 17 healthy Korean carry at least one of these pathogenic variants that develop neurodevelopmental troubles in a recessive manner. Furthermore, the feasibility of these genes for carrier screening was evaluated. Our results will be the basis for recessive variant screening for appropriate diagnosis and treatment of rare Mendelian disorder patients.

In this thesis, genetic and functional studies have enabled us to explain the genetic causes of rare neurodevelopmental disorders, additionally, the results of this study underline the utility of WES-based clinical diagnostics for improving patient care.

* This thesis is based on published article; Lee *et al*., *Sci Rep*. 10(1):1413 (2020).

Keywords: Whole exome sequencing, Developmental disorders, Pediatric disease, Neurology, Recessive variants, Variant discovery

Student Number: 2013-31180

TABLE OF CONTENTS

ABSTRACT ……...……...……i

TABLE OF CONTENTS ……..…….. iii

LIST OF FIGURES AND TABLES …….. iv

INTRODUCTION ……... 1

MATERIALS AND METHODS ……. 6

RESULT …… 13

DISCUSSION ……..…….. 48

REFERENCES …….…… 55

LIST OF ABBREVIATIONS ……..……... 65

국문초록 ……..…….. 66

감사의 글 ……..…… 69

LIST OF FIGURES AND TABLES

FIGURES

Figure 1. Whole exome sequencing (WES) pipeline …….. 5

Figure 2. Categorization of patient's clinical diagnosis (n = 553 patients) ……. 8

Figure 3. The age distribution of patients when their symptom began and the time differences between age of onset and WES analysis (n = 553 patients) ……. 14

Figure 4. Location of Seoul National University Children's Hospital (SNUCH) ……. 15

Figure 5. Classification of the KND cohort by inheritance patterns and major symptoms …… 16

Figure 6. Diagnostic yield of 553 patients with undiagnosed symptoms using WES … 18

Figure 7. Identification of an inherited deletion at 2p22.3 in the family with hereditary spastic paraplegia (HSP) …… 21

Figure 8. Verification of the deleted regions in HSP family through qPCR and Sanger sequencing …….. 22

Figure 9. Pathogenic variants divided by inheritance patterns ……..…….. 25

Figure 10. Pathogenic variants categorized by their function (n = 298 variants) ……. 26

Figure 11. Disease and GO enrichment analysis of 164 known genes ……..…… 27

Figure 12. Disease and GO enrichment analysis between male and female ……..……. 28

Figure 13. Verification of the association of new genes to known gene's network in various brain developmental stage ……..…… 30

Figure 14. Burden of recessive variants in KND patients (Pt) and their parents as controls (Ct) ……. 32

Figure 15. Comparison in the number of recessive variants for neurodevelopment related gene sets between patients and controls …….…… 33

Figure 16. Venn diagrams displaying high correlations of recessive or dominant inheritance patterns with their known inheritance patterns …….…… 36

Figure 17. ACMG code distribution of variants that are in recessive or dominant inheritance pattern …….. 37

Figure 18. Comparison of genetic properties in recessive and dominant variants …..... 38

Figure 19. LoF tolerance of genes from KND cohort against all or haploinsufficiency genes ……..…….. 39

Figure 20. Relative position of LoF variants in genes ……..…….. 40

Figure 21. Functional differences between recessive and dominant genes from GO analysis …….…….. 41

Figure 22. Estimation of the proportion of Korean rare neurodevelopmental disorder carriers …… 44

Figure 23. Comparison of various parameters between pathogenic recessive variants from KND patients and gnomAD variants from the same genes that were found in KND patients ……... 45

Figure 24. Summary of the frequency and type of recessive variants in the KND cohort. …… 46

Figure 25. Validation of case that WES-based analysis altered clinical courses ……... 52

TABLES

Table 1. Clinical information of 553 KND patients …….…… 7

Table 2. List of copy number variations discovered in this study …….. 19

Table 3. List of known neurologic disorder associated genes …… 23

Table 4. Notable case where WES-based analysis provided correct diagnoses or changed medical treatment strategies ……... 53

INTRODUCTION

Recent advances in next-generation sequencing (NGS) technology have revolutionized genetic research. NGS technology has enabled researchers to study the entire human genome more effectively. Therefore, various high-throughput sequencing techniques have been applied to clinical study as well as basic research. Whole exome sequencing (WES), one of various NGS techniques, has been widely used in clinical fields since its introduction in 2009. While whole genome sequencing (WGS) technique sequencing the entire genome, WES only covers the coding region of the genome, so it is relatively inexpensive and requires less effort to analyze (Fig. 1) (*1*). WES is now an indispensable tool in clinical diagnosis.

Clinical WES is a method of linking and analyzing the clinical information and sequencing data of patients and parents for diseases that are difficult to diagnose with conventional biochemical and radiological tests. Genotype-phenotyping analysis of disease using WES provides an accurate diagnosis and treatment strategy for the patient. The trio-based whole exome sequencing analysis provides higher diagnostic accuracy than patient-only analysis and enables the discovery of novel genetic variants that were difficult to find with conventional targeted gene sequencing panel method.

Rare disease is defined as a disease that affects less than 200,000 people in the United States, depending on its prevalence. Although numerous patients suffer from rare diseases worldwide, a large number of rare diseases have yet to be elucidated due to their genetic complexity and rarity. To date, more than 4,200 disease-causing genes have been registered in the Online Mendelian Inheritance in Man (OMIM), but the genetic cause of

many rare Mendelian disorders is still unknown. Although WES has recently been applied to various common complex diseases, WES is still more effective in studying rare diseases. Thanks to its low cost and easy of analysis, WES has been a useful tool for research on rare Mendelian disorders.

Undiagnosed rare diseases are an important challenge to overcome because they cause not only loss of happiness of patients and their families, but also constant social burden. To overcome these undiagnosed rare disorders, there have been large-scale national projects and international cooperation based on comprehensive genetic analysis, such as Undiagnosed Diseases Program (UDP) in United States, Deciphering Developmental Disorders (DDD) in UK, Finding of Rare Disease Genes (FORGE) in Canada. Through these large-scale projects, many rare Mendelian disease patients were diagnosed and new causative genes were discovered. In Korea, the need for large-scale study on rare Mendelian diseases has emerged and related projects are underway. As part of these projects, we conducted a genetic analysis of patients with pediatric rare neuro-developmental disease using WES.

A large fraction of rare Mendelian disorders follow a recessive inheritance pattern (*2*, *3*). The Online Mendelian Inheritance in Men (OMIM) lists 5,317 disorders and 3,077 of these are categorized as recessive (as of April 2019). However, the contribution of recessive variants to the pathogenesis of common complex disease is lower than expected (*4–7*). For rare diseases, the contribution of recessive variants in inbred populations has been well studied (*8–10*). However, the precise contribution of recessive variants to rare Mendelian disorders in an outbred population is still not well understood.

Due to the complexity of brain developmental processes, there are many patients with diverse neurological problems that are difficult to diagnose by conventional criteria.

Diagnosis of neurological diseases affecting children is frequently hampered by overlapping clinical features, making it difficult for clinicians to easily recognize the nature of disease and find appropriate treatment. This makes pediatric neurologic patients a target for genome-wide genetic studies (*11–13*). To facilitate diagnosis and discovery of novel disease pathology, systematic efforts on a regional and national scale have been carried out (*14–17*). Since many rare pediatric neurological disorders follow Mendelian inheritance, variant discovery by trio-based whole-exome sequencing (WES) has proved to be the most effective method, yielding diagnosis rates of 25-41% (*11–13*, *15*, *17*).

Notably, the medical system in Korea provides a unique opportunity to conduct a systematic analysis of rare disease and study the contribution of recessive variants on a large-scale. Since the nation-wide referral system focuses on a few major tertiary clinical institutions, Seoul National University Children's Hospital (SNUCH) covers a large portion of the 51-million Korean population, enabling consistent evaluation and treatment of the patient cohort. For example, we recently reported on genetic analyses of large cohorts of Duchenne muscular dystrophy (n = 507) and Rett-like syndrome without MECP2 mutations (n = 34) (*18*, *19*). Genetically, Koreans are a good example of an outbred population where marriages between relatives and even between individuals with the same surnames has been banned since the 17th century (*20*). Our study represents the largest of what was conducted in a single clinic, and highlights the careful integration of clinical and genetic analysis.

In this study, WES was used to analyze a cohort of 553 patients (KND cohort) with diverse neurodevelopmental disorders. We characterized the genotype-phenotype relationships of patients with identified molecular defects, and explored the potential association of genes not previously associated with disorder. We demonstrate that a high

proportion of recessively inherited variants are associated with patients with rare neurodevelopmental disorders. Variants that were inherited in a recessive manner were analyzed and their genetic properties were evaluated, aiming to understand their distribution in healthy populations for appropriate diagnosis and treatment of patients with rare Mendelian disorder. Together, this thesis describes the establishment of a system that efficiently integrates genetic techniques with clinical diagnosis to maximize benefits for pediatric rare disease patients and their families.

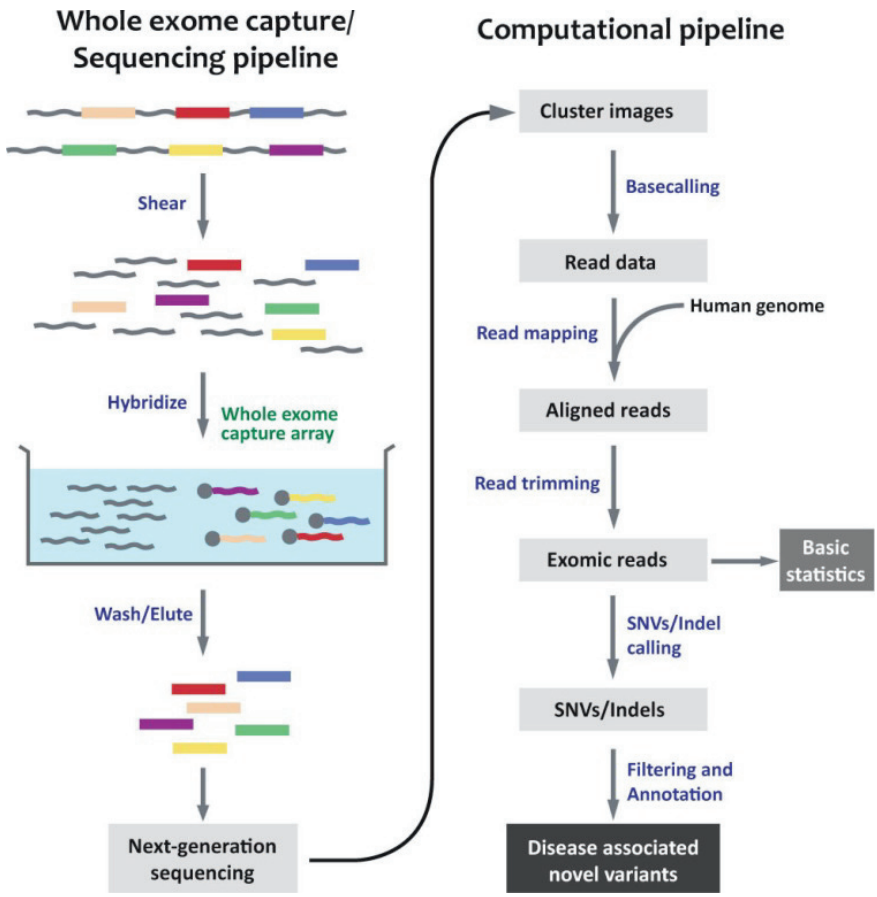

Figure 1. Whole exome sequencing (WES) pipeline. SNV, single nucleotide variant; Indel, insertion or deletion.

MATERIALS AND METHODS

Subjects

Blood samples were obtained from patients and their parents, who provided informed consent. WES was performed on 553 patients who visited the SNUCH pediatric neurology clinic from July 2014 to January 2019 and displayed various neurodevelopmental problems, such as Rett syndrome-like encephalopathy, mitochondrial encephalopathy, epileptic encephalopathy, neuromuscular disorder, leukodystrophy, hereditary spastic paraplegia (Table 1). The patients can be categorized into two groups: (i) clinically diagnosable but with genetic heterogeneity (270/553, 48.8%) or (ii) heterogeneous and nonspecific clinical features without definite diagnosis (283/553, 51.2%; Fig. 2). Prior to WES analysis, thorough clinical and laboratory evaluation of patients was performed to identify possible genetic causes. These included genetic tests with candidate gene sequencing, targeted gene panels, trinucleotide repeat analysis, metabolic testing, brain MRI, or muscle biopsy. All patients were evaluated by three pediatric neurologists, two pediatric neuroradiologists, and a pathologist.

Table 1. Clinical information of 553 KND patients.

Sex (*n* (%))	
Male	265 (47.9)
Female	288 (52.1)
Age at symptom onset (years)	1.4 (0-21)
Age at first access to a tertiary hospital (years)	1.8 (0-22)
Interval between symptom onset and first medical access (months)	3.9 (0-238)
Number of visited tertiary hospitals for diagnosis (*n* (%))	
1	62 (11.2)
2	277 (50.1)
3	178 (32.2)
4	32 (5.8)
5	4 (0.7)
Age at WES (years)	7.4 (0-37)
Interval between the first access and WES (months)	
Patients aged 0-10 years	34.0 (0-100)
Patients aged >10 years	114.5 (7-434)
Primary clinical diagnosis (*n* (%))	
Rett syndrome-like encephalopathy	72 (13.0)
Mitochondrial encephalopathy	49 (8.9)
Epileptic encephalopathy	51 (9.2)
Neuromuscular disorder	37 (6.7)
Leukodystrophy	27 (4.9)
Hereditary spastic paraplegia	34 (6.1)
Others	283 (51.2)
Number of involved specialists for diagnosis (*n* (%))	
1-2	378 (68.4)
3-5	152 (27.5)
> 5	23 (4.2)
Straight-line distance from home to the clinic, km (*n* (%))	
< 20	186 (33.6)
20-100	180 (32.5)
> 100	187 (33.8)

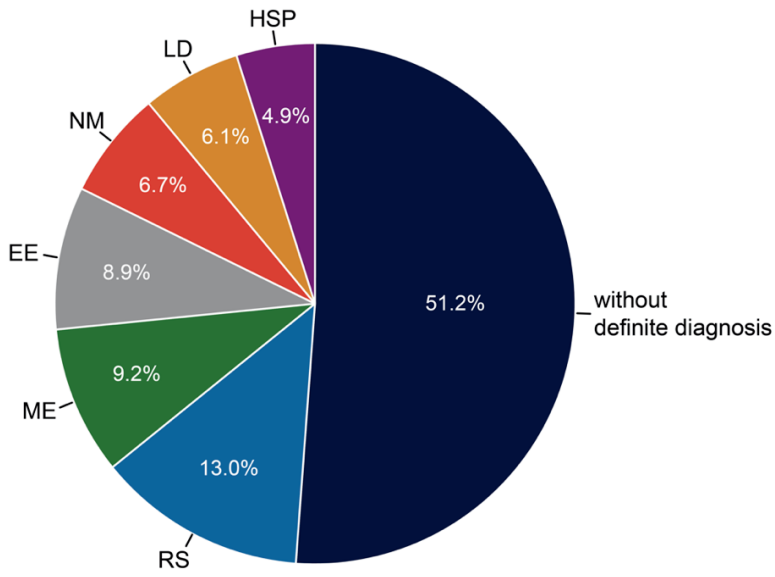

Figure 2. Categorization of patient's clinical diagnosis (n = 553 patients). RS, Rett syndrome-like encephalopathy; ME, Mitochondrial encephalopathy; EE, Epileptic encephalopathy; NM, Neuromuscular disorder; LD, Leukodystrophy; HSP, Hereditary spastic paraplegia

Whole Exome Sequencing

WES was performed at Theragen Etex Bio Institute (Suwon, Korea) following the standard protocol and the sequencing data were analyzed as described previously (*19*). Depending on the genetic analysis result, each patient was categorized as one of the following: category 1: known disease-causing genes were found; category 2: causative gene for other diseases were found; category 3: potentially pathogenic gene, but without prior disease association, was found; category 4: no disease-causing candidates were found; and category 5: known pathogenic copy number variation (CNV) was found.

Our variant assessment procedures were as follows: firstly, patient-specific CNVs were checked and samples with CNVs were classified as category 5. CNV was called by comparing the normalized coverage depth between the patient and their parents at each capture interval. Then, patient-specific variants such as de novo, compound heterozygous (CH; autosomal), and rare homozygous (RHo; autosomal) and hemizygous (RHe; X-linked) variants were selected from patients by comparing against their parents and prioritized based on the inheritance pattern (Fig. 5). Variants with low in-house quality score (< 60) or low coverage depths (< 10) were excluded. Variants in intron regions and pseudogene were also filtered out.

In associating the patient's clinical features with discovered variants, if patients carried a known pathogenic variant in OMIM or ClinVar, they were categorized as category 1 or 2, depending on similarity with reported clinical characteristic. For variants not previously reported, if they were not seen in normal population (Genome Aggregation Database (gnomAD) (*21*), Korean Variant Archive (KOVA) (*22*) and in-house database) and were evolutionarily well-conserved, they were classified as potentially pathogenic

variants. ACMG evidence codes for the variants were annotated using a script provided by InterVar (*23*).

Human Brain Transcriptome Data

The BrainSpan transcriptome database (http://www.brainspan.org) was used to build networks of developing human brain (*24*). Data from 8 post-conceptual weeks (PCWs) to 40 years of age were analyzed. A total of 385 samples were used for the network analysis after combining the multiple values by taking average. Probes with TPM (transcripts per million) > 5 in at least one sample were used, yielding 23,943 probes as "brain-expressed transcripts".

Brain Transcriptome Network Analysis

Using the above brain-expressed transcripts, we created 8 known gene co-expression networks by selecting genes that are highly associated to our set of genes (n = 164; Pearson's correlation r > 0.7) that have already been reported to have disease relevance at each brain developmental period (Fig. 13). Then, we asked whether our novel genes can be successfully integrated into the known gene co-expression network. We randomly selected 53 genes (equal to the number of our novel genes) from brain-expressed transcripts and counted how many edges they formed with the known genes at each period. The 10^5 random gene selections were performed and the edge distribution was constructed using the number of edges with known genes. The number of edges from observed novel genes was evaluated against this distribution. P-values were calculated using Z-score, assuming normal distribution.

Recessive variant analysis

Among the variants capable of altering the protein sequence, those with the gnomAD allele frequency less than 0.001 were selected. Then CH variants were called on a trio-based setting. If a gene contains more than one filtered variant and each variant was inherited from mother and father separately (for proband), or at least one but not all of the filtered variants of a gene were found in the offspring (for parent), the variants were called as CH. RHo variants were called if filtered variants are inherited in a homozygous manner in autosomes and never seen in gnomAD as homozygous. RHe variants were called if filtered variants are in the X chromosome and never seen in gnomAD as hemizygous or homozygous. Various functionality scores were extracted from dbNSFP database (*25*).

Statistical evaluation

Statistical analysis of this study was conducted using the R package (version 3.6.0). When comparing difference of two groups, normality of the distribution was first evaluated by Shapiro-Wilks test and then either Student's t-test or a non-parametric Mann-Whitney U test was performed. P-value < 0.05 was considered as statistically significant, and adjusted FDR P-values were used when correcting for multiple tests. Correlation between variables was determined by using the Pearson's correlation coefficient test.

Ethical approval and informed consent

This study was approved by the Seoul National University Hospital Institutional Review Board (No. 1406-081-588), and was performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from all enrolled patients or their

legal representatives.

RESULT

Clinical features of KND cohort

The symptoms of KND patients were mostly of pediatric onset (mean 1.4 years of age). The patients had neurodevelopmental problems and were soon referred to tertiary hospitals (mean 1.8 years of age). The majority of the patients visited multiple tertiary hospitals for diagnosis (88.8% visited more than one hospital, mean of 2.3 hospitals). Patients required an average of 2.3 specialists (31.6% required more than two) and it took an average of 5.6 years for WES analysis after visiting SNUCH (Fig. 3, Table 1). The distribution of straight distances from home to the hospital strongly correlates with the population distribution of Korea, suggesting that our cohort covers the entire Korean population (Fig. 4, Table 1).

KND Patients were classified into three groups by disease inheritance pattern; Class 1: autosomal dominant families; Class 2: families with affected siblings; Class 3: affected individuals without family history. The majority of the patients is sporadic origin (504/553 = 91.1%; Fig. 5a), making them suitable for trio-based WES analysis. Some patients showed various neuromuscular problems or multiple anomaly, but major clinical symptom of the KND cohort was neurodevelopmental disorder such as Rett syndrome-like encephalopathy, mitochondrial encephalopathy, epileptic encephalopathy, leukodystrophy, hereditary spastic paraplegia, Leigh Syndrome (84.1%; Fig. 2, Fig. 5b).

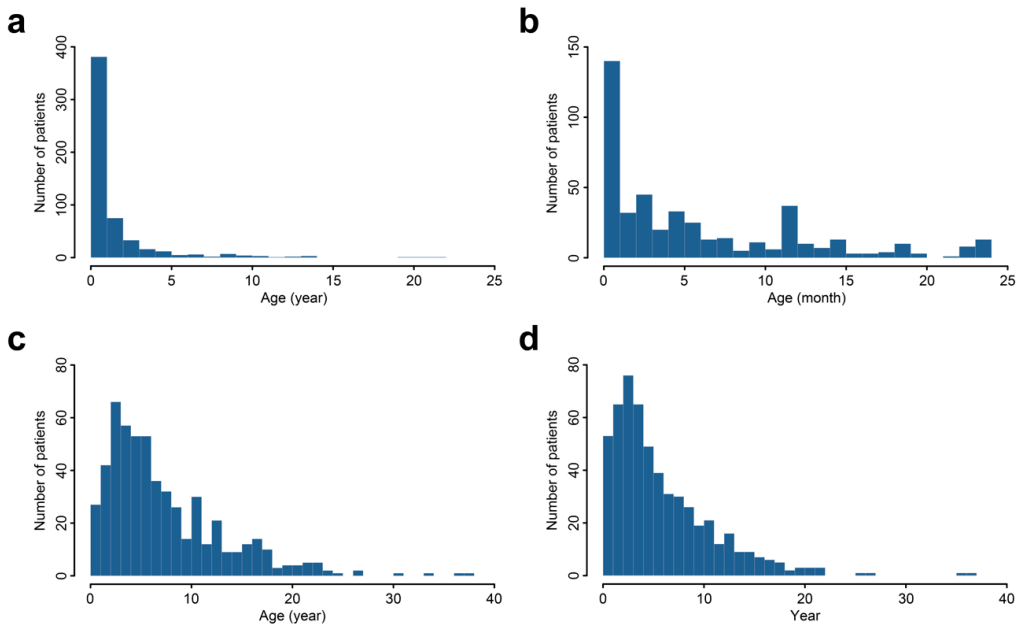

Figure 3. The age distribution of patients when their symptom began and the time differences between age of onset and WES analysis (n = 553 patients). (a) The age of onset of all patients (year). **(b)** The age of onset of patients whose symptoms began before 24 months. **(c)** The Age distribution of patients when the WES analysis was performed. **(d)** Distribution of time differences between patient's age of onset and the time of WES analysis.

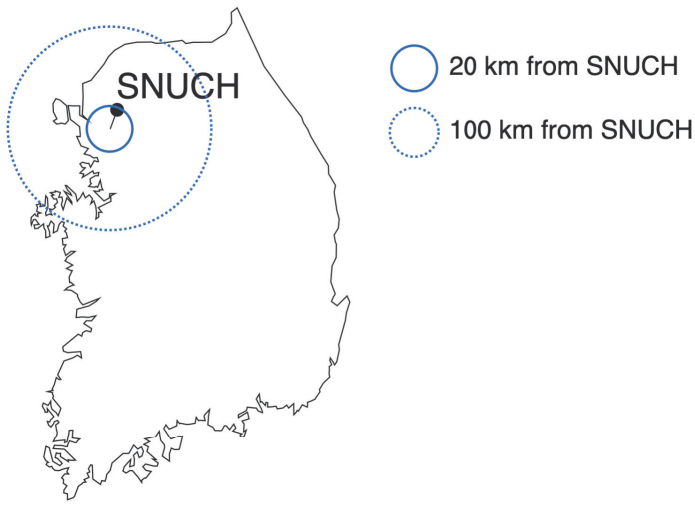

Figure 4. Location of Seoul National University Children's Hospital (SNUCH). The straight distances from SNUCH were displayed. The 20 km radius circle includes most of Seoul, where about a quarter of entire population resides. The 100 km radius circle encompasses most of the Gyeonggi province, where another quarter of entire population resides.

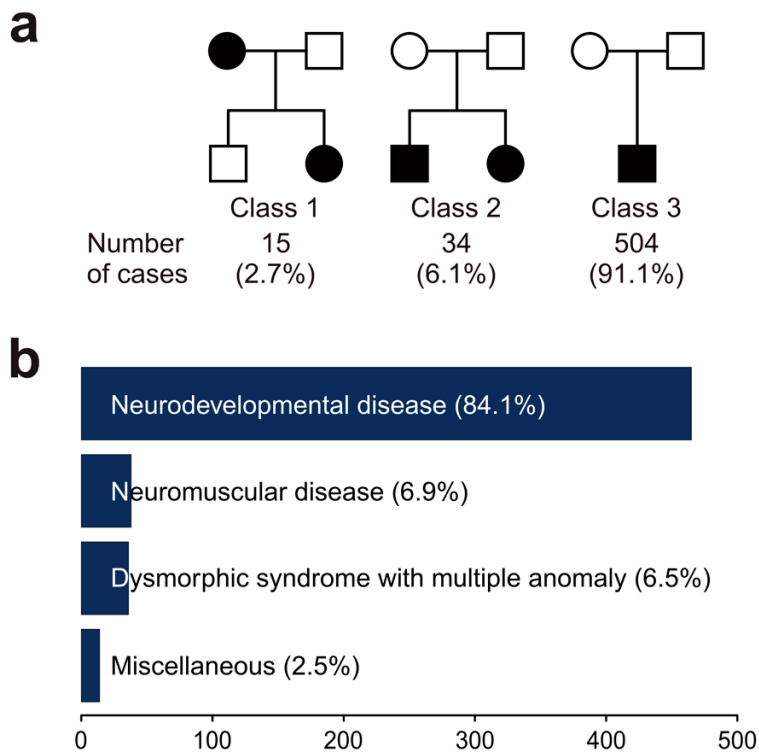

Figure 5. Classification of the KND cohort by inheritance patterns and major symptoms. (a) Subjects were categorized into 3 groups according to disease inheritance patterns. Class 1: autosomal dominant families; Class 2: families with affected siblings; Class 3: affected individuals with no family history. **(b)** Major clinical features of the KND cohort (n = 553).

Diagnostic success rate of WES analyses

Integrative evaluation of genetic variants and patient symptoms allowed us to diagnose 40.3% (223/553) of the KND cohort with high confidence (Fig. 6). This patient group included CNV carriers (23/553 = 4.2%; 16 heterozygous deletions and 7 duplications; Fig. 6, Table 2), 20 of which were de novo (3.6% of the entire cohort), which was slightly lower but similar to a previous study (*26–29*). Three inherited pathogenic CNVs were identified: 7.2 Mb and 203.2 kb hemizygous duplications on the X chromosome that were transmitted from healthy moms to their affected sons (Table 2) and a 165.5 kb deletion in a large family containing multiple affected individuals showing hereditary spastic paraplegia (HSP) symptom (Fig. 7a-7b). Actually, the 165.5 kb deletion consisted of three small deletions which were associated with *Alu* elements (Fig. 7c). Quantitative PCR (qPCR) of patient's genomic DNA showed decreased copy number in deleted regions, and breakpoint of deletions were also confirmed by Sanger sequencing (Fig. 8).

In addition to the high confidence group, 7.1% of the cohort (39 patients) harbored variants in genes previously reported to cause disease but showed distinct phenotypes, potentially expanding the phenotypic spectrum associated with these genes. For example, two patients that carried a pathogenic heterozygous missense or nonsense variant in *COL1A1*, known to cause osteogenesis imperfecta (*30*), were initially diagnosed with muscle hypotonia. These two patients did not show skeletal problems, but displayed blue sclera (*31*). Adding this group to the high confidence group yielded a diagnostic rate of 47.4% ("known genes"; Table 3). Finally, an additional 10.1% of the cohort (56 patients, 53 genes) harbored variants that are highly likely to be pathogenic but their disease associations are elusive ("novel genes"; Fig. 6).

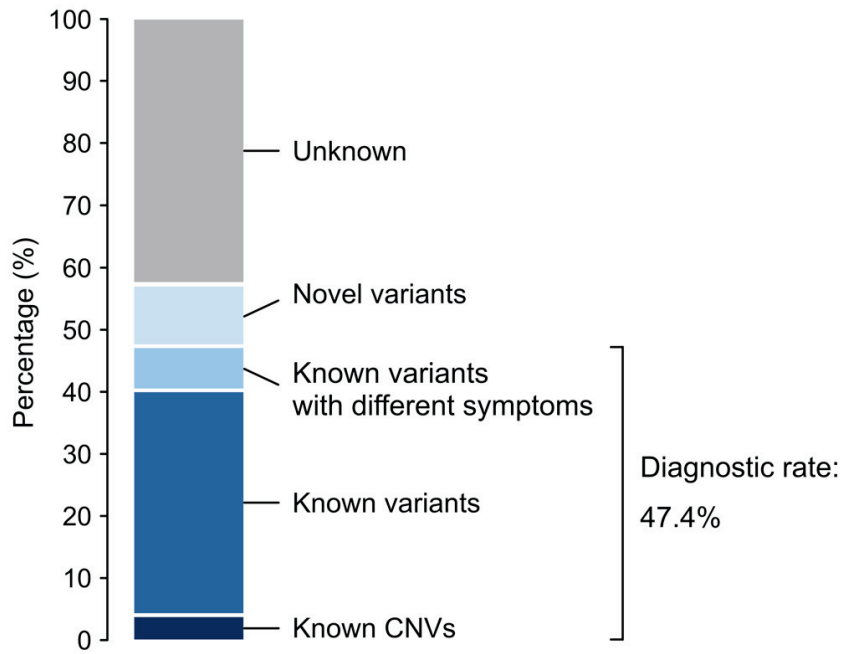

Figure 6. Diagnostic yield of 553 patients with undiagnosed symptoms using WES.

Known CNV = 4.2%, Known variants = 36.1%, Known variants with different symptoms = 7.1%, Novel variants = 10.1%, Unknown = 42.5%.

Table 2. List of copy number variations discovered in this study.

No.	Simple Diagnosis	CNV interval (Mb, hg19)	Length	Inheritance	No. of genes*	Notable gene(s) in the interval	Ref
1	Rett syndrome-like	chr1:0.6-3.1	2.6 Mb del	de novo	60	*MMP23B, GABRD, SKI, PRDM16*	(*32*)
2	Rett syndrome-like	chr1:107.0-113.1	6.1 Mb del	de novo	69	*WNT2B, NTNG1*	(*33*)
3	Hereditary spastic paraplegia	chr2:32.3-32.5	165.5 kb del	Inherited	3	*SPAST*	(*34*)
4	Epileptic encephalopathy	chr2:171.2-175.1	3.8 Mb del	de novo	18	*GAD1*	(*35*)
5	Unknown encephalopathy	chr2:234.8-242.8	8.0 Mb del	de novo	61	*HDAC4, PRLH, PER2, TWIST2, CAPN10, KIF1A, FARP2, D2HGDH, PDCD1*	(*36*, *37*)
6	Rett syndrome-like	chr3:9.0-13.0	4.0 Mb del	de novo	47	*SETD5, SLC6A1, SLC6A11*	(*38*, *39*)
7	Developmental delay with facial dysmorphism	chr3:44.2-48.0	3.9 Mb del	de novo	55	*SETD2, CSPG5, PTH1R, SMARCC1*	(*40*)
8	Rett syndrome-like	chr4:0.0-2.8	2.8 Mb del	de novo	42	*PIGG, CPLX2, FGFRL1, CTBP1, SLBP, LETM1*	(*41*)
9	Congenital hypotonia	chr5:139.0-139.6	567.7 kb del	de novo	7	*NRG2, PURA*	(*42*)
10	Epileptic encephalopathy	chr7:119.7-119.9	171.3 kb dup	de novo	0	*KCND2* upstream	(*43*)
11	Epileptic encephalopathy	chr9:0-47.3	47.3 Mb dup	de novo	202	n.d.	(*44*)

12	Developmental delay with facial dysmorphism	chr9:139.6-141.1	1.6 Mb del	de novo	56	*EHMT1*	(*45*)
13	Developmental delay with facial dysmorphism	chr12:53.6-54.1	537.9 kb dup	de novo	18	*AAAS, AMHR2, RARG*	(*46*)
14	Developmental delay, Joubert syndrome	chr15:20.3-23.7	3.4 Mb del	de novo	11	*TUBGCP5, NIPA1, NIPA2, CYFIP1*	(*47*)
15	Intellectual disability, Facial dysmorphism	chr16:29.6-30.2	635.4 kb del	de novo	26	*MVP, CDIPT1, SEZ6L2, ASPHD1*	(*48*)
16	Global Developmental delay	chr18:52.5-53.3	787.2 kb del	de novo	3	*TCF4*	(*49*)
17	Hereditary spastic paraplegia	chr22:21.1-21.6	508.7 kb del	de novo	10	*CRKL*	(*50*)
18	Unknown severe retardation	chr22:42.8-51.3	8.6 Mb del	de novo	81	*SHANK3, IGF1*	(*51*)
19	Developmental delay, Microcephaly	chrX:41.4-41.7	299.5 kb del	de novo	3	*CASK*	(*52*)
20	Unknown encephalopathy	chrX:100.1-107.3	7.2 Mb dup	hemizygous	76	*PLP1*	(*53*)
21	Neurodegeneration with motor developmental arrest	chrX:102.5-103.2	692.2 kb dup	de novo	15	*PLP1*	(*54*)
22	Epileptic encephalopathy	chrX:152.8-153.4	620.3 kb dup	de novo	21	*MECP2*	(*55*)
23	Hereditary spastic paraplegia	chrX:153.6-153.8	203.2 kb dup	hemizygous	15	*RPL10, ATP6P1, GDI1, IKBKG*	(*56*, *57*)

* Based on NCBI RefSeq database. Only protein-coding genes were counted.

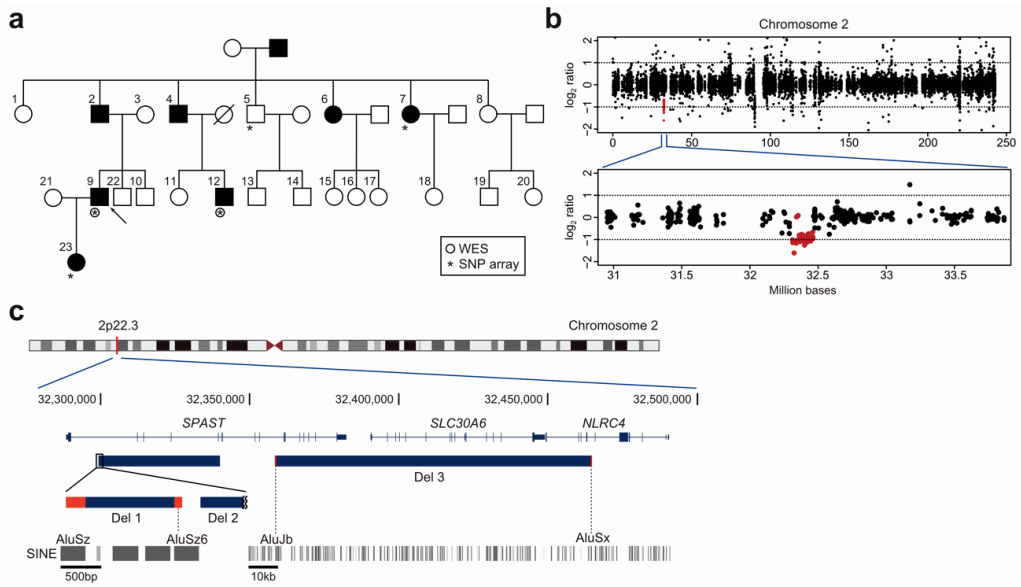

Figure 7. Identification of an inherited deletion at 2p22.3 in the family with hereditary spastic paraplegia (HSP). (a) Pedigree of an HSP family with 8 affected individuals across four generations. **(b)** Log2-based copy-number values of subject HSP-9 compared with an unrelated normal subject show the presence of heterozygous microdeletion on chromosome 2. Captured intervals of copy-number loss are indicated by red dots. **(c)** Enlarged view of deleted regions at 2p22.3 including SPAST gene. Blue solid bars represent the deleted intervals. *Alu* repeat elements at the deletion breakpoints are indicated by red solid bars and aligned with the RepeatMasker of the UCSC Genome Browser, represented by dotted lines.

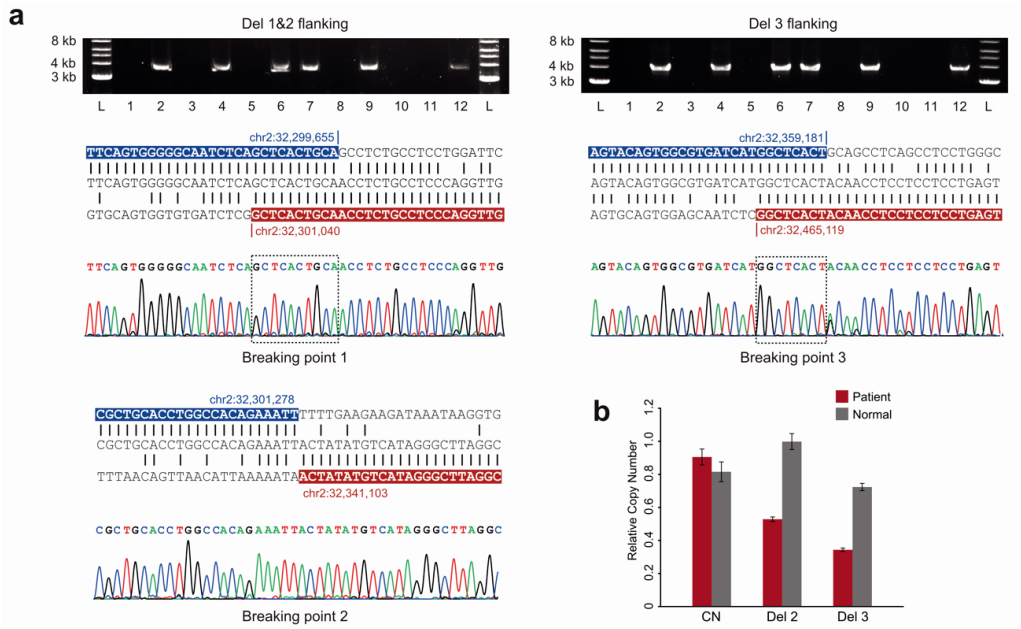

Figure 8. Verification of the deleted regions in HSP family through qPCR and Sanger sequencing. (a) DNA sequence analysis of the deleted regions. The DNA fragments containing the deletion was amplified by the deletion-specific primer pairs. The deletion-specific PCR products of 3.8 kb (Del 1 & 2) and 4.0 kb (Del 3) are observed in each affected individual. Reference sequences surrounding the breaking points are indicated in blue and red color. **(b)** Validation of the deleted regions by quantitative PCR of genomic DNA. The red bars represent an average copy-number of five patients and the gray bars represent an average copy-number of five normal individuals in the family. Error bars represent standard error. CN: copy-neutral region.

Table 3. List of known neurologic disorder associated genes.

Inheritance pattern	Gene	No.
Autosomal-dominant	*ACOX1, ACTB, ACTG1, ADCY5, ANKRD11, ANO3, ARID1B(3), ARID2, ASXL3(4), ATL1(4), ATP1A3, BICD2, BPTF(2), C19orf12, CHD3, CHD7, CLTC, COL1A1(2), CTNNB1(5), DHX30, DNM1(2), DNM1L, DYNC1H1(2), EHMT1, FOXG1(2), FOXP1, GABBR2, GABRB1, GFAP, GNAO1(5), GNB1, GRIA2, GRIN1(3), GRIN2A, GRIN2B(3), HECW2, ITPR1(3), KAT6A(2), KCNC1, KCNQ2, KIF1A(3), KMT2A, LRP5, MED13L, MRAS, MYH7, NALCN(2), NFIX, NFKB2, NIPBL, NRAS, PIK3CA(3), PMP22, POGZ, PPP2R5D(2), PPP3CA, PTEN, PURA(2), RAB11B, RARB(3), REEP1(2), RHOBTB2(2), SATB2, SCN1A, SCN2A(3), SCN4A, SETD2, SFTPC, SLC2A1(2), SLC6A1, SLC6A5, SMARCB1, SOX10, SOX5, SPAST, SPTBN2, STAT3, STXBP12(2), SYNGAP1, TCF4(3), TUBB3, TUBB4A(4), UBE3A, WASHC5, WDFY3, ZEB2*	86
Autosomal-recessive	*ABAT, ALDH7A1, APTX, ARSA(2), ASAH1, ATAD3A, ATP6V0A2, BRAT1(2), CC2D2A, CLN6(2), CNTNAP1, CWF19L1, CYP7B1, DEGS1, ECHS1(2), EIF2B2, EIF2B3, ERCC5, FIG4, GCDH, GDAP1, GLB1(2), GMPPB(2), HADHA, HSD17B4, IARS2(2), IGHMBP2, KIAA1109, KIF1A, KLHL40, LONP1(2), MICU1, NARS2(2), NDUFAF6(2), NDUFS1, NDUFV1, PEX5, PIGN, PLA2G6(2), PMM2, PNPT1(2), POLR1C, PPT1, PRUNE1, RYR1, SACS(2), SLC19A3(2), SLC25A15, ST3GAL5, SURF1, THOC6, TMEM173, TUBA8, WDR62, WDR81*	55
X-linked	*ATRX(3), CASK(2), CDKL5(3), EDA, EIF2S3, GRIA3(2), HDAC8, HPRT1, IQSEC2(2), MECP2(3), MED12, NAA10, NSDHL, OPHN1(2), PAK3, PCDH19, PDHA1(2), PHF6, PLP1, SLC16A2, SMC1A(2), USP9X, ZC4H2(4)*	23

( ): Number of patients (more than one patient only)

Characteristics of genes and variants found in the KND cohort

Among the patients with definite diagnosis, as expected, approximately 60% of variants were dominant including de novo CNV and 35.1% of variants showed recessive inheritance pattern (compound heterozygous, homozygous and hemizygous) (Fig. 9). In terms of variant function, a considerable number of pathogenic variants were missense, 29.9% harbored loss-of-function (LoF) variants (frameshift, nonsense and splicing variants). And in very few patients, we also found the damaging in-frame (Fig. 10). As expected, the known genes showed strong enrichment in disease categories such as and intellectual disability, global developmental delay. As a result of gene ontology (GO) analysis, these known genes also exhibited strong enrichment in brain development or function such as CNS development and synaptic signaling (Fig. 11). There were no distinct differences between male and female in both disease and GO analysis (Fig. 12).

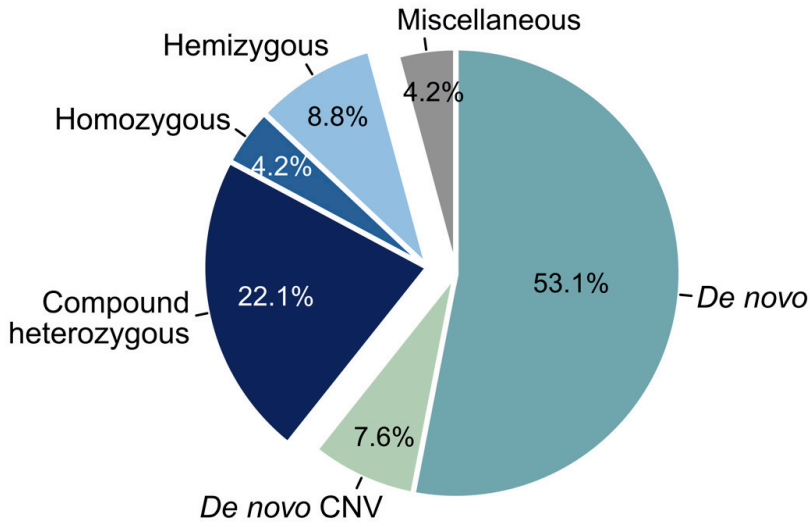

Figure 9. Pathogenic variants divided by inheritance patterns. Miscellaneous variants contain shared dominant variants and mosaic variants.

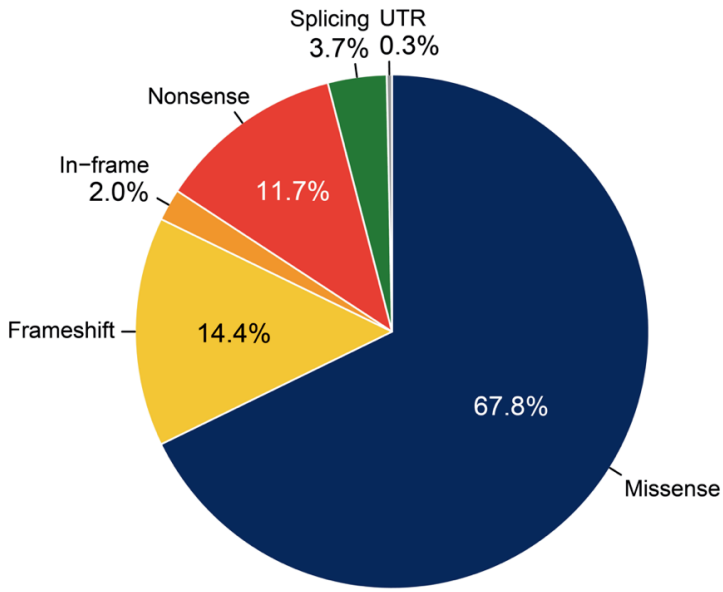

Figure 10. Pathogenic variants categorized by their function (*n* = 298 variants).

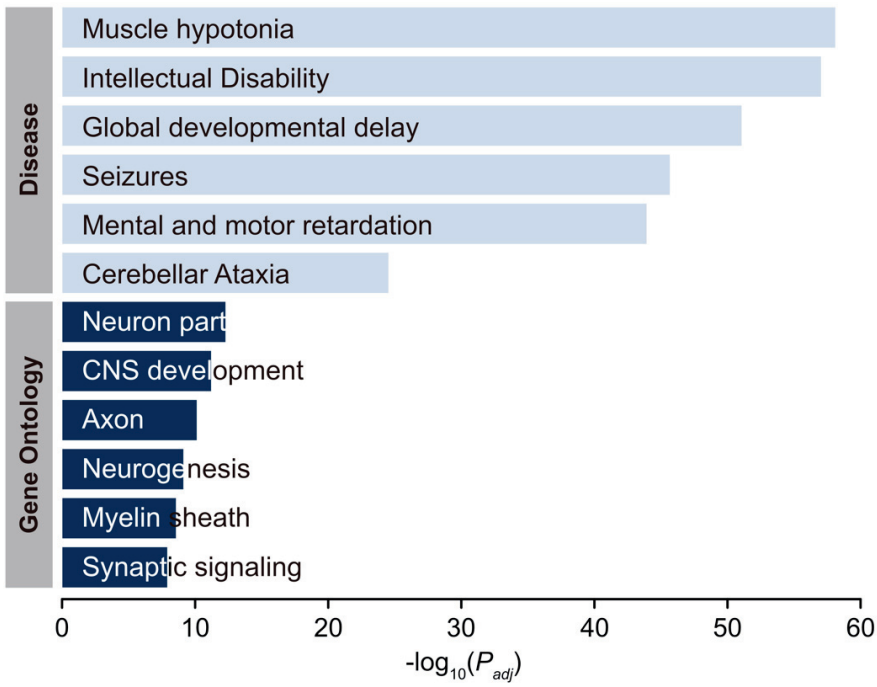

Figure 11. Disease and GO enrichment analysis of 164 known genes. Similar categories were excluded and both disease and GO categories were sorted by adjusted P value.

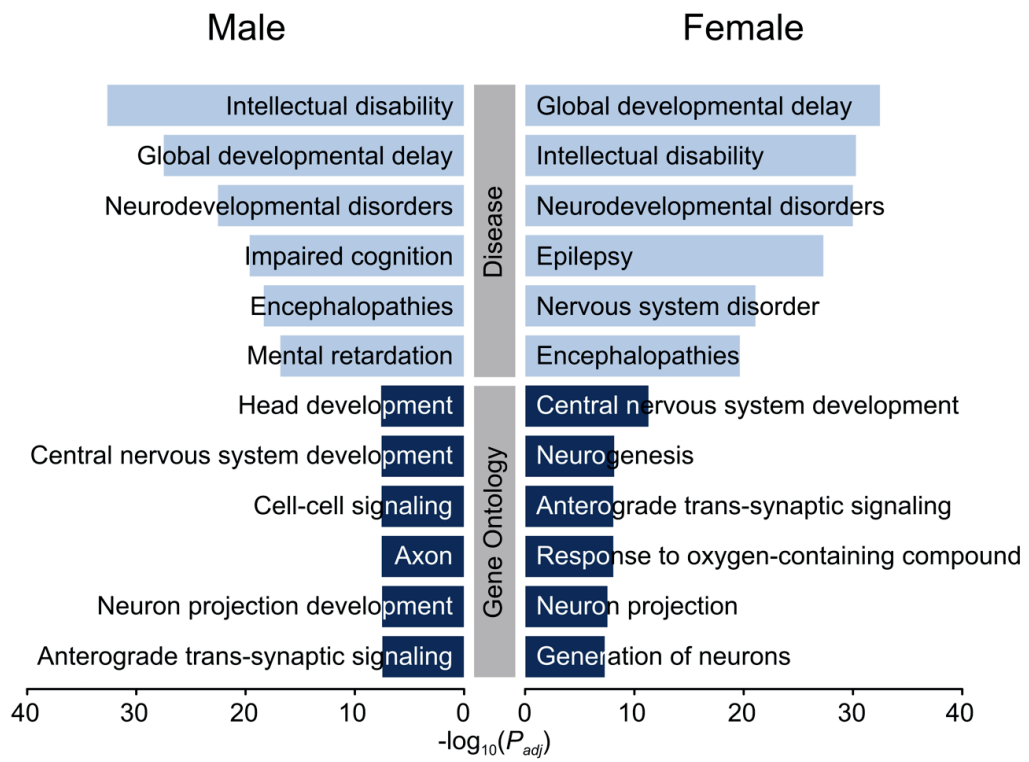

Figure 12. Disease and GO enrichment analysis between male and female. Similar categories were excluded and both disease and GO categories were sorted by adjusted P value.

Novel genes display potential enrichment in developing human brain

We evaluated whether the 53 novel genes newly discovered in KND cohort possess a neurodevelopmental disease-causing function. The novel genes were simulated using the BrainSpan data (Materials and Methods) to assess whether the expression of these novel genes was strongly correlated with known disease-associated genes during brain development. After 10^5 permutations, we found that the observed involvement of the novel genes was significantly stronger than the randomly selected gene sets across eight brain developmental windows ($P_{adj} < 0.05$ from Z-score for all periods; Fig. 13a-13b). Furthermore, this test was expanded to the four anatomical regions of the brain in each period (Fig. 13c), creating a total of 32 spatiotemporal windows. It is notable that the most highly enriched windows are concentrated in the frontal cortex area (R1 x P1-4; Fig. 13b). These results suggest that expression of the novel genes is closely related to that of known disease-causing genes in developing brains and this phenomenon is most pronounced in the frontal cortex region.

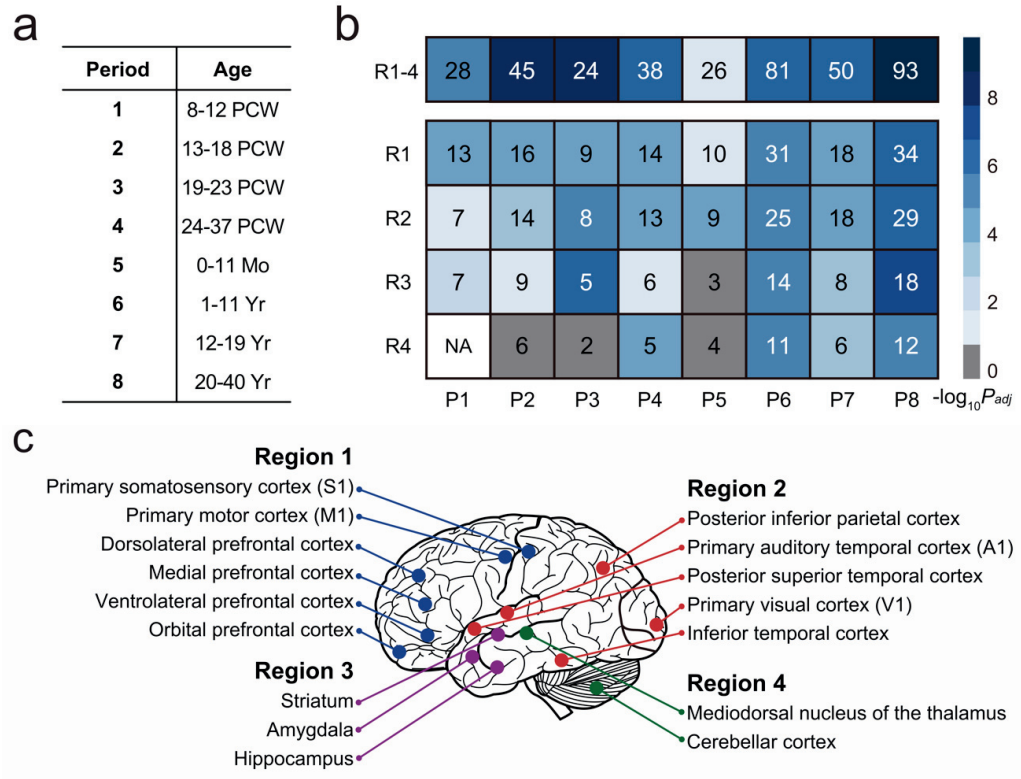

Figure 13. Verification of the association of new genes to known gene's network in various brain developmental stage. (a) Eight brain developmental periods used in the network analysis (from 8 PCWs to 40 years). **(b)** Strength of the co-expression network composed of our known/novel genes compared to random networks as measured by 10^5 permutations. The number of samples from BrainSpan RNA-seq data which were used in analysis is displayed in each window. Developmental period and brain region of each window were shown in (a) and (c), respectively. **(c)** Anatomical components of four brain regions.

Comparison of recessive variants burden between patients and normal controls

Using the set of defined pathogenic variants, we explored the genetic properties of the variants that caused disease in a recessive manner. Both dominant and recessive variants displayed similar proportions by functionality (Fig. 10) and carriers of dominant or recessive variants experienced similar ages of onset (data not shown). Next, to test if recessive variants (i.e., compound heterozygous (CH), rare homozygous (RHo) and rare hemizygous (RHe) variants) are more frequently found in patients as compared to healthy individuals, we counted the number of recessive variants in our cohort and compared these values between patients and their healthy parents as controls. Counting all recessive variants from patients and controls, we observed that there is no substantial difference in the number of recessive variants (Fig. 14). LoF variants, variants in OMIM-listed genes or variants in neurodevelopment-related genes also did not show any difference in burden (Fig. 14, Fig. 15), implying the presence of overwhelming non-pathogenic recessive variants in the patients.

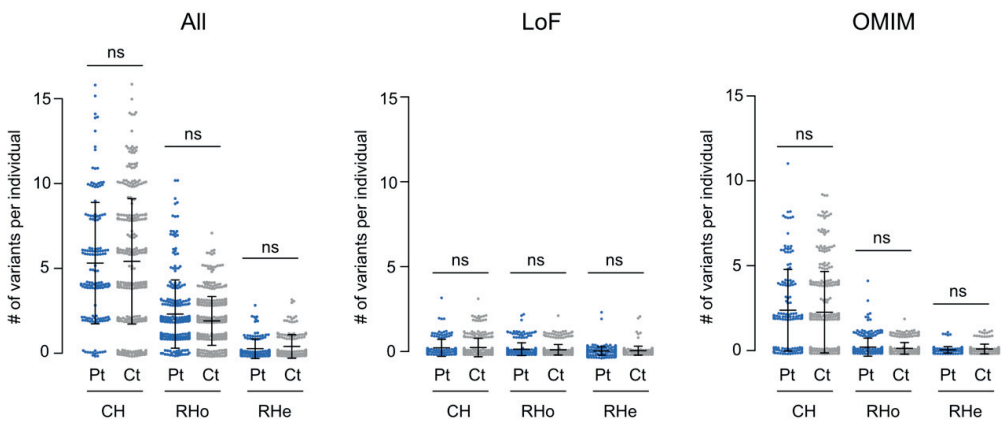

Figure 14. Burden of recessive variants in KND patients (Pt) and their parents as controls (Ct). Recessive variants are divided into three groups, compound heterozygous (CH), rare homozygous (RHo) and rare hemizygous (RHe). Numbers of all variants from all genes ("All"), LoF variants from all genes ("LoF") and all variants from OMIM-listed genes ("OMIM") are plotted. Numbers of samples used for each category are as following: patients for CH = 145; controls for CH = 290; patients for RHo = 247; controls for RHo = 341; patients for RHe = 134; controls for RHe = 168. Data are mean ± standard deviation.

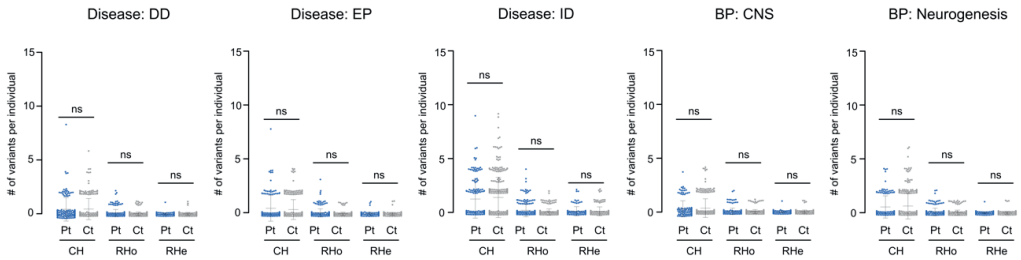

Figure 15. Comparison in the number of recessive variants for neurodevelopment-related gene sets between patients and controls. Numbers of all variants from Disease associated and neurodevelopment-related genes are plotted. DD, Developmental delay; EP, Epilepsy; ID, Intellectual disability; BP, GO biological process; CNS, central nervous system development. Disease gene sets were obtained from DisGeNET database.

Genetic properties of pathogenic recessive variants against dominant variants in KND cohort.

The majority of genes with definite diagnosis had already been previously documented in OMIM, and had a good concordance with previously known recessive or dominant inheritance patterns (Fig. 16). There was one exceptional case in which the gene is listed as recessive in OMIM but was dominantly inherited in our cohort. Until recently, only the recessive ACOX1 phenotype was identified (*58*), but recently we reported a case describing this dominant ACOX1 variant (*59*). Most of the variants in the two groups were classified as likely pathogenic or pathogenic according to the American College of Medical Genetics and Genomics (ACMG) guideline (98.4% for recessive and 96.7% for dominant variant group), but the dominant variants tended to be more pathogenic (Fig. 17; Fisher's exact test $P = 3.7 \times 10^{-22}$). There was no significant difference in basic clinical parameters (Table 1) between the recessive and dominant patient groups (data not shown).

Next, to test the difference of genetic properties between recessive variants and dominant variants, several parameters were compared. Dominant variants (mean allele frequency = 6.2×10^{-7}) were found less frequently than recessive variants in gnomAD (mean allele frequency = 1.6×10^{-5}; Mann-Whitney U test $P = 1.7 \times 10^{-13}$), because most of the dominant variants originated de novo whereas recessive variants were inherited from healthy parents (Fig. 18a). Compared to dominant variants, recessive variants were slightly less evolutionarily conserved, based on PhyloP score or amino acid conservation in vertebrates (Mann-Whitney U test P = 0.034 and 0.048, respectively; Fig. 18b). Other prediction scores did not show significant difference between the two groups (CADD P = 0.50, GERP P = 0.15 and SIFT P = 0.17, Mann-Whitney U tests).

Through comparison of observed/expected ratio (o/e) and pLI (probability of

being loss-of-function intolerant) score in gnomAD, it was found that genes containing recessive variants exhibited more lenient constraints compared to the dominant genes or known haploinsufficiency genes (*21*). However, the recessive genes still displayed a similar or slightly more constrained pattern compared to the genes in OMIM (Fig. 19). The relative position of LoF variants in recessive tended to slightly more enriched in the C-terminal region similar to all gnomAD genes, compared to the dominant genes (Fig. 20). This tendency is because the recessive variants are inherited from normal parents. Recessive genes from KND cohort were enriched in lipid metabolism and mitochondrial components, in addition to the expected enrichment in CNS development related processes or neuronal components (Fig. 21).

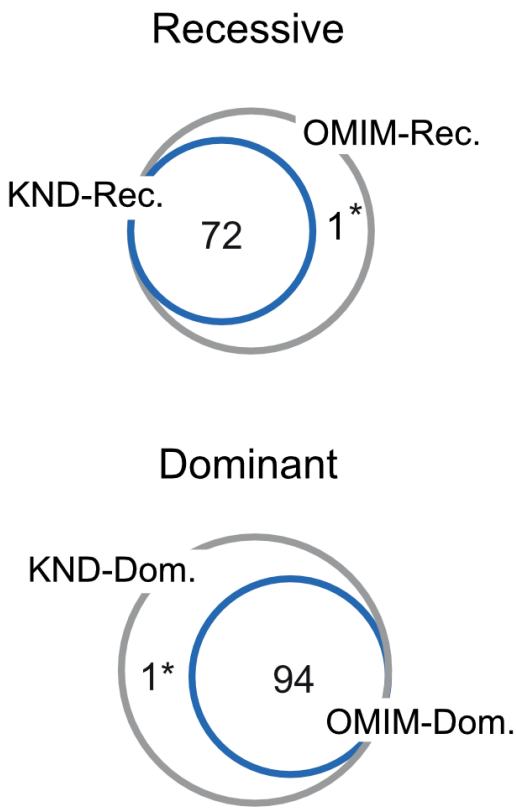

Figure 16. Venn diagrams displaying high correlations of recessive or dominant inheritance patterns with their known inheritance patterns. The asterisks denote one exceptional case, ACOX1 (see text).

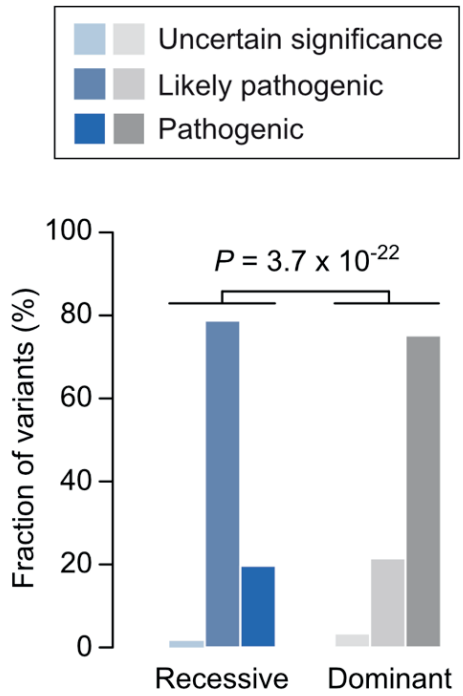

Figure 17. ACMG code distribution of variants that are in recessive or dominant inheritance pattern. P values was calculated using Fisher's exact test.

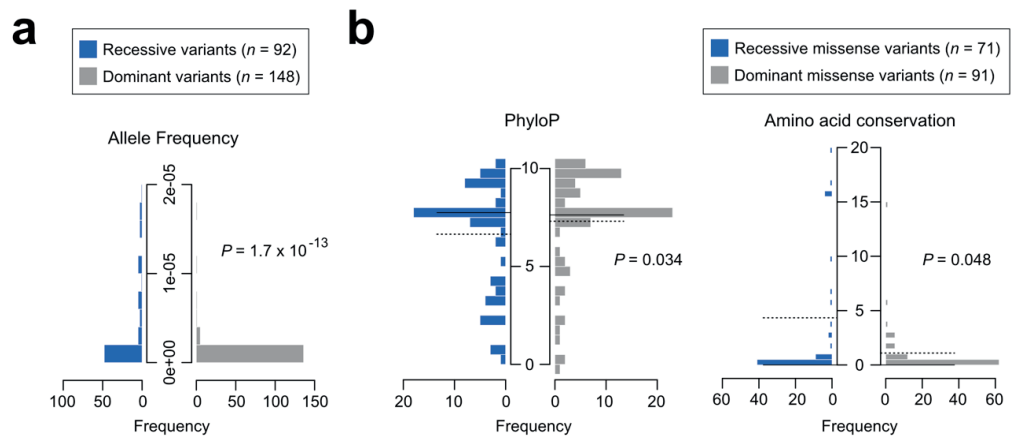

Figure 18. Comparison of genetic properties in recessive and dominant variants. (a) Allele frequency distribution of dominant and recessive variants. **(b)** PhyloP and amino acid conservation differences between dominant and recessive missense variants. Amino acid conservation is determined by the number of vertebrate species that contain an amino acid that is different from its human orthologous residue. The solid lines denote medians and the dotted lines denote means.

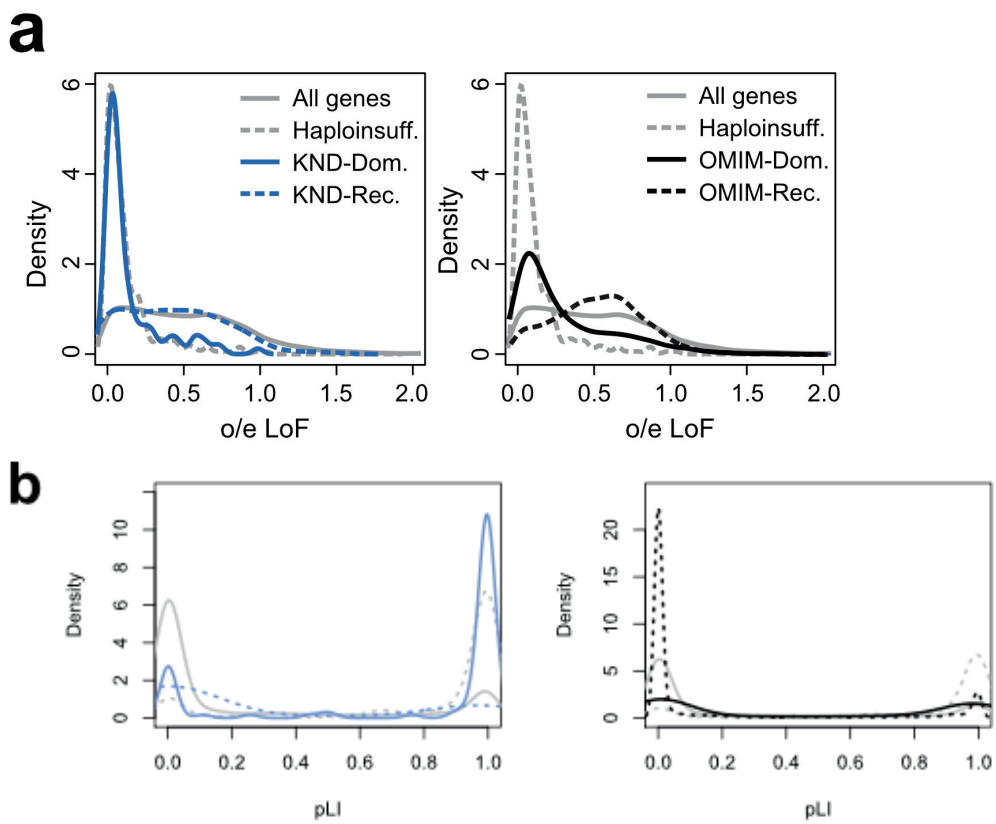

Figure 19. LoF tolerance of genes from KND cohort against all or haploinsufficiency genes. (a) Distributions of o/e LoF values for dominant and recessive genes found from KND patients (left) and dominant and recessive genes from OMIM (right) plotted against all genes and known haploinsufficiency genes (n = 291) (*60*). **(b)** Distributions of pLI score. Legend of plots is the same as (a).

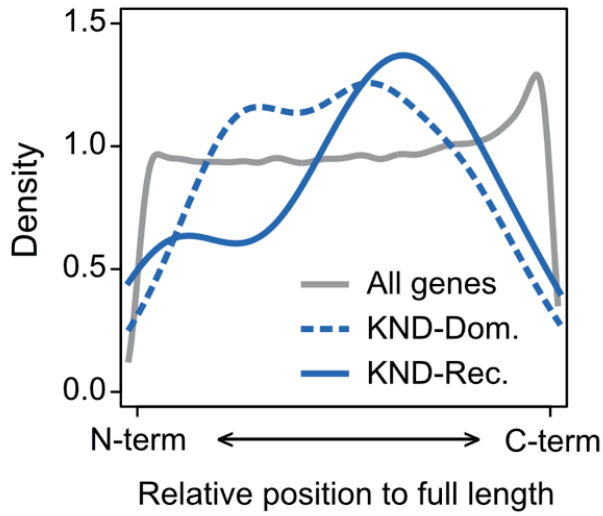

Figure 20. Relative position of LoF variants in genes. Positions of pathogenic LoF variants in genes from KND patients are plotted against those LoF variant positions from all genes in gnomAD.

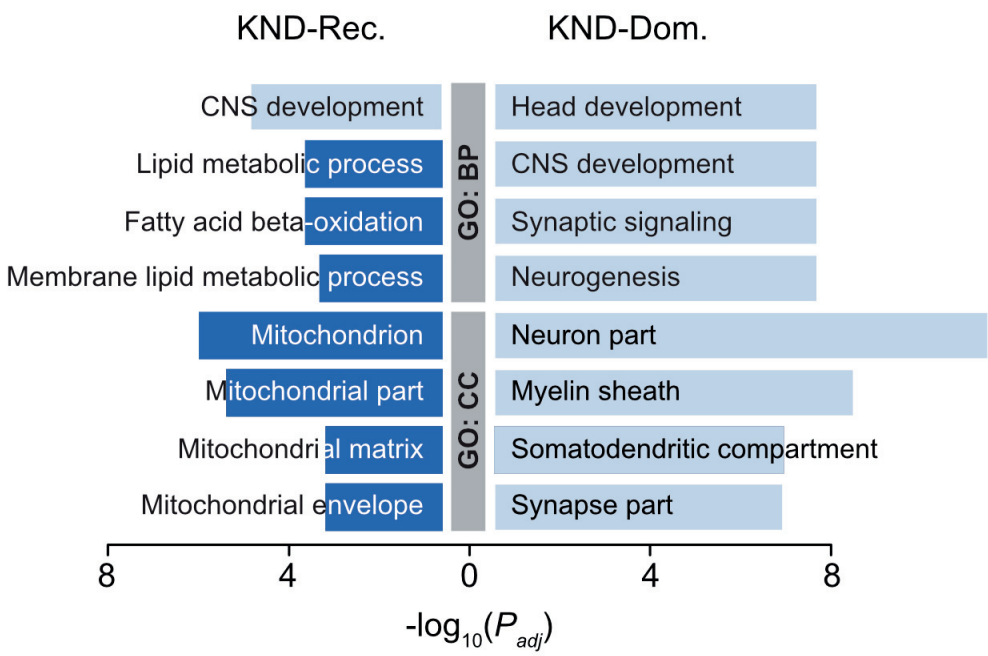

Figure 21. Functional differences between recessive and dominant genes from GO analysis. The recessive gene-specific clusters are marked with darker color. Similar categories were excluded and sorted by adjusted P value. BP, Biological Process; CC, Cellular Component.

Profile of pathogenic recessive variants carried in healthy individuals

Unlike de novo variants, which occur mostly at random, recessive variants can be pre-screened and avoided if the variants can be identified in parents. Taking advantage of comprehensive genetic analysis of KND cohort based on Korea's centralized medical system, it is possible to estimate the possibility of a combination of recessive variants in Koreans. Several assumptions are required for this estimation: (i) Approximately 400,000 babies are born every year in Korea as of 2016 (*61*). (ii) Approximately 1,000 patients with neurodevelopmental disorders newly enroll in our clinic every year (Jong Hee Chae, personal communication, May 28, 2019). (iii) These patients cover the majority of the Korean population. This is sufficiently convincing given the geographical distribution of the KND patients and the examples of our DMD and Rett Syndrome cohort (*18*, *19*) (Fig. 4, Table 1). (iv) Our result from 553 KND patients displayed a recessive genetic origin in approximately one-third of the patients (Fig. 9) and (v) Koreans generally marry an individual with minimal genetic similarity. These observations indicate that the incidence of patients with neurodevelopmental disorder in a recessive manner is about 1/1200, suggesting that 1 in 17 healthy people can be a carrier of pathogenic variants (1/1,156; Fig. 22). This is a conservative estimate, and the proportion of carriers may actually be slightly higher, given some patients we do not cover.

Next, we tried to understand the genetic properties of pathogenic recessive variants in KND by looking at all of the variants of 69 recessive genes in gnomAD. As expected, KND recessive variants were found less frequently (P = 4.2 x 10^{-10}, Mann-Whitney U test; Fig. 23a), showed stronger evolutionary conservation (P = 3.0 x 10^{-5} for PhyloP and P = 3.2 x 10^{-7} for amino acid conservation, Mann-Whitney U tests; Fig. 23b) compared to all gnomAD variants on the same genes. As a result of comparing various

damage prediction scores, KND recessive variants also displayed more deleterious effect (CADD $P = 2.5 \times 10^{-5}$, GERP $P = 4.0 \times 10^{-3}$ and SIFT $P = 3.2 \times 10^{-6}$, Mann-Whitney U tests; Fig. 23c) compared to gnomAD variants.

To test the feasibility of accurate pre-screening in a healthy population, we considered gnomAD-originated heterozygous LoF and ClinVar variants found in our recessive genes as a first-tier culprit for pathogenic recessive variants among many variants of obscure functional significances. And we observed that the portion attributed to the LoF and ClinVar variants by healthy carriers was variable among the genes, and this portion is correlated with the o/e LoF value (Pearson's correlation r = 0.33; Fig. 24).

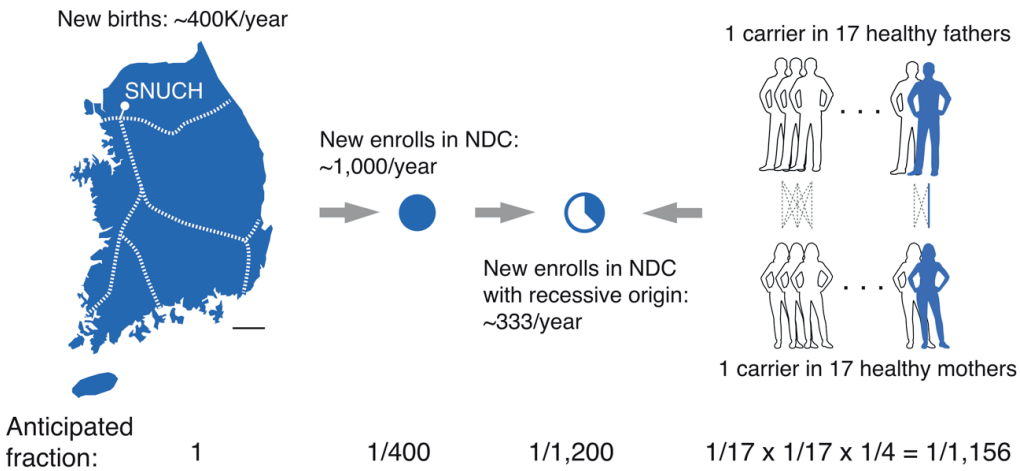

Figure 22. Estimation of the proportion of Korean rare neurodevelopmental disorder carriers. A schematic diagram describing processes used to estimate neurodevelopmental disorder carrier frequency in the Korean population. The dotted lines in the map denote the Korea Train Express network, the high-speed railway system of Korea. The healthy carriers are displayed in blue color. The expected carrier proportions are shown below the figure.

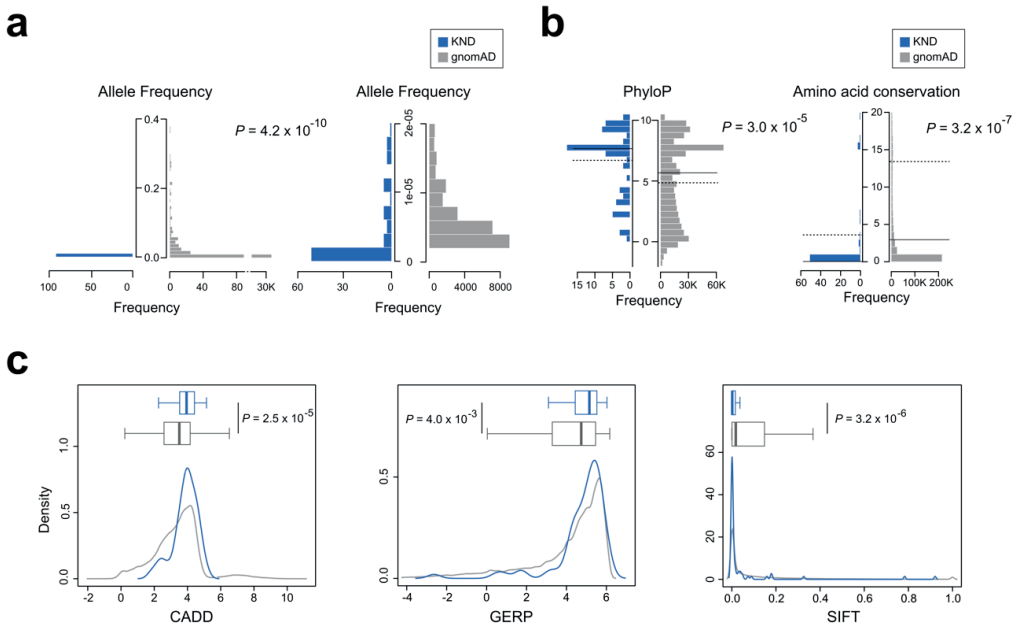

Figure 23. Comparison of various parameters between pathogenic recessive variants from KND patients and gnomAD variants from the same genes that were found in KND patients. (a) Allele frequency. The rare frequency portion of the left panel is separately plotted in the right panel. **(b)** PhyloP score and amino acid conservation. The solid lines mean medians and the dotted lines mean average values. **(c)** The distribution of CADD, GERP and SIFT scores. P values are calculated using Mann-Whitney U tests.

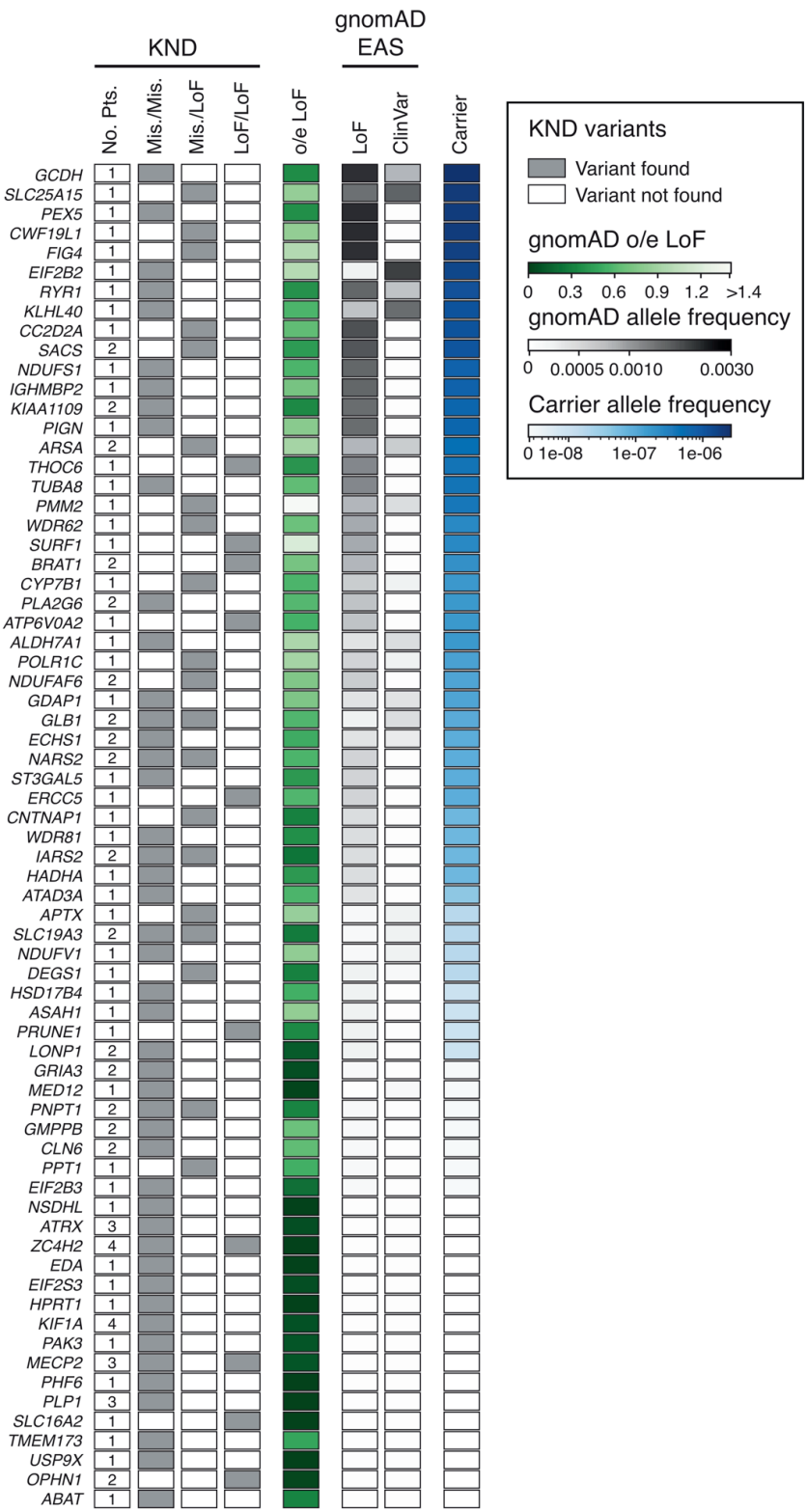
KND
gnomAD EAS
No. Pts.
Mis./Mis.
Mis./LoF
LoF/LoF
o/e LoF
LoF
ClinVar
Carrier
GCDH 1
SLC25A15 1
PEX5 1
CWF19L1 1
FIG4 1
EIF2B2 1
RYR1 1
KLHL40 1
CC2D2A 1
SACS 2
NDUFS1 1
IGHMBP2 1
KIAA1109 2
PIGN 1
ARSA 2
THOC6 1
TUBA8 1
PMM2 1
WDR62 1
SURF1 1
BRAT1 2
CYP7B1 1
PLA2G6 2
ATP6V0A2 1
ALDH7A1 1
POLR1C 1
NDUFAF6 2
GDAP1 1
GLB1 2
ECHS1 2
NARS2 2
ST3GAL5 1
ERCC5 1
CNTNAP1 1
WDR81 1
IARS2 2
HADHA 1
ATAD3A 1
APTX 1
SLC19A3 2
NDUFV1 1
DEGS1 1
HSD17B4 1
ASAH1 1
PRUNE1 1
LONP1 2
GRIA3 2
MED12 1
PNPT1 2
GMPPB 2
CLN6 2
PPT1 1
EIF2B3 1
NSDHL 1
ATRX 3
ZC4H2 4
EDA 1
EIF2S3 1
HPRT1 1
KIF1A 4
PAK3 1
MECP2 3
PHF6 1
PLP1 3
SLC16A2 1
TMEM173 1
USP9X 1
OPHN1 2
ABAT 1
KND variants
Variant found
Variant not found
gnomAD o/e LoF
0 0.3 0.6 0.9 1.2 >1.4
gnomAD allele frequency
0 0.0005 0.0010 0.0030
Carrier allele frequency
0 1e-08 1e-07 1e-06

Figure 24. Summary of the frequency and type of recessive variants in the KND cohort. The number of KND patients with recessive variants in each gene and the types of those recessive variants were displayed. The o/e LoF values, and accumulated frequencies of LoF and ClinVar variants from gnomAD East Asians (EAS) for genes that harbor known pathogenic recessive variants in KND cohort were also displayed. Finally, portion that were attributable to LoF or ClinVar variants for pre-screening parents for each recessive gene are shown.

DISCUSSION

This study demonstrates the clinical utility of applying WES to pediatric patients with various neurodevelopmental disorders. we identified genetic causes in 47.4% of the patients and evaluated the characteristics of the variants that caused the disorders in a recessive manner.

Consistent with previous studies, we were able to diagnose approximately half of the KND patients with WES (Fig. 6) (*11–13*, *15*, *17*). Our pathogenic genes and variants showed good correlation with OMIM inheritance pattern and ACMG guideline (Fig. 16, Fig. 17). The novel genes formed a strong co-expression networks with known pathogenic genes during neurodevelopment processes, which was most prominent in frontal cortex regions (Fig. 13). There was no significant difference in the number of recessive variants between the patient and the healthy control group, as was the comparison of disease-related genes (Fig. 14, Fig. 15). This result suggested that the patients carry more non-pathogenic or non-functional recessive variants than we expected.

Next, the variants found in the KND cohort were divided by inheritance patterns, and the genetic characteristics of the variants were compared and analyzed. The pathogenic recessive variants displayed slightly increased allele frequency and decreased evolutionary conservation compared to dominant variants, suggesting that the differences between these two groups of variants were not dramatically different (Fig. 18). The characteristics of variants divided by their inheritance patterns did not show significant difference, but from the viewpoint of genes causing such diseases, some parameters showed noticeable differences. First, the recessive genes showed an increased o/e LoF

values compared to the dominant genes, implying relatively low genetic constraints (Fig. 19). These results can also be seen through the difference in the distribution of the relative locations of LoF variants in genes (Fig. 20). Gene ontology (GO) analysis further supports that – while both groups are predominantly composed of neurodevelopment-related genes – recessive genes are enriched in lipid metabolism and mitochondrial componenets (Fig. 21). This is a reasonable result considering that these pathways are essential for normal brain development (*62*, *63*). In summary, these observations suggest that gene traits are a stronger determinant than that of variant in determining whether a disease adopts a recessive or dominant inheritance pattern.

It is important to predict the occurrence of recessive disorders. Carrier estimates have been traditionally performed primarily for single-gene diseases such as β-thalassaemia, Tay-Sachs disease and cystic fibrosis, and have effectively reduced the incidence of these patients (*64–66*). However, even after aggressively introducing an analysis of genetic disorder using whole exome or whole genome sequencing, the estimation of the contribution of recessive inheritance to rare Mendelian diseases varies considerably depending on the study subjects and diseases (*3*). For example, the result of the Deciphering Developmental Disorders (DDD) study revealed a small contribution of recessive inheritance (3.6%) in European patients, whereas studies in Pakistani patients showed a relatively large contribution of recessive inheritance (30.9%) (*5*). Genetic analysis of schizophrenia patients did not detect a substantial contribution of recessive variants (*6*, *7*). These observations differ from our results, where 35.1% of patients who are clearly diagnosed follow recessive inheritance (Fig. 9), which show good agreement with previous clinical WES studies (*11*, *67*, *68*). Notably, a recent study using a large autism cohort revealed that a significant proportion of the patients were attributed to the

recessive LoF variants (*69*).

Most of these pathogenic recessive variants were inherited from healthy parents, and Koreans are composed of relatively isolated populations with a centralized medical system, so one in 17 individuals can be estimated to be a healthy carrier of pathogenic recessive variants against severe neurodevelopmental diseases (Fig. 22). The contribution of known LoF and ClinVar variants varies by genes and is correlated with o/e LoF values (Fig. 24), and pathogenic recessive variants exhibit a systematic difference distinguishing them from gnomAD variants (Fig. 23). Thus, it would be feasible to predict potential rare recessive variants from genomic data of healthy parents with the help of large genomic data on patients and controls in the near future.

Our approach expanded the phenotypic spectrum of known genes (39 cases, 7.1%), and suggested novel genes that could better understand the mechanisms of neurodevelopmental disorder (56 patients, 10.1%). Nevertheless, even after our thorough WES analysis, 42.5% of the cases (235/533) remained undiagnosed, suggesting opportunities for further improvement (Fig. 6). In this regard, the efforts of systematic re-analysis through additional bioinformatics pipelines increased the diagnostic rate by 4.2% (*70*). In addition, although the improvement of the diagnostic rates through WGS was not dramatic due to our limited understanding of the function of non-coding variants, it would be beneficial to explore functional non-coding variants through WGS and to evaluate the role of multi-variants in disease occurrence (*71*, *72*). Although preparing the tissue of the patient remains a practical challenge, the integration of genome and transcriptome data to identify cryptic genetic variation can be an alternative approach (*73*, *74*).

Our study addresses the clinical challenges of evolving phenotypes and how to overcome them, facilitating the identification of treatable or actionable cases (Fig. 25,

Table 4). Our study also included a successful drug repositioning case for a rare neurological disorder (*75*) (Table 4). These cases are expected to increase as more genotype-phenotype relationships are discovered and more drugs become available. This study demonstrates that application of WES and subsequent analysis can provide clinical benefits to patients and their families. Finally, we demonstrated the successful establishment of this approach in Korea, and the need for this approach for patients with various undiagnosed neurodevelopmental disorders.

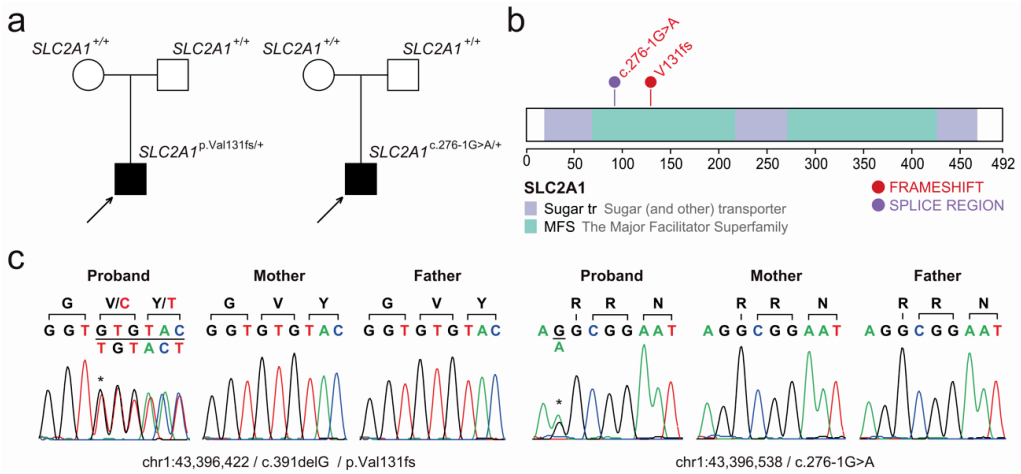

Figure 25. Validation of case that WES-based analysis altered clinical courses. (a) Two patients with dystonia and delayed motor development harbored LoF de novo variants in *SLC2A1*. **(b)** The position of variants found in two patients and domains of *SLC2A1* are displayed. **(c)** Sanger traces validating the de novo variants.

Table 4. 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	(*76*, *77*)
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	(*78*)
Early onset hypotonia, sacral mass, congenital heart disease, and facial dysmorphism	*ASAH1*	Farber lipogranulomatosis (#228000)	Corrected a misdiagnosis	(*79*)
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)	(*80*)
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)	(*81*, *82*)
Leigh syndrome	*SLC19A3*	Thiamine metabolism dysfunction syndrome 2 (#606152)	Identified disease-specific treatment that resulted in clinical improvements in dystonia, spasticity, and cognitive function	(*83*, *84*)

			(supplements of thiamine and biotin)	
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 (tofacitinib treatment)	(*75*)
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	(*85*)

REFERENCES

1. G. Goh, M. Choi, Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics Inform.* **10**, 214 (2012).
2. K. M. Boycott, M. R. Vanstone, D. E. Bulman, A. E. MacKenzie, Rare-disease genetics in the era of next-generation sequencing: Discovery to translation. *Nat. Rev. Genet.* **14** (2013), pp. 681–691.
3. S. E. Antonarakis, Carrier screening for recessive disorders. *Nat. Rev. Genet.* (2019), doi:10.1038/s41576-019-0134-2.
4. S. C. Jin *et al.*, Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601 (2017).
5. H. C. Martin *et al.*, Quantifying the contribution of recessive coding variation to developmental disorders. *Science (80-.).* **362**, 1161–1164 (2018).
6. E. Rees *et al.*, Analysis of exome sequence in 604 trios for recessive genotypes in schizophrenia. *Transl. Psychiatry*. **5**, e607 (2015).
7. D. M. Ruderfer *et al.*, No evidence for rare recessive and compound heterozygous disruptive variants in schizophrenia. *Eur. J. Hum. Genet.* **23**, 555–557 (2015).
8. D. Monies *et al.*, Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. *Am. J. Hum. Genet.*, 1–20 (2019).
9. F. S. Alkuraya, Discovery of mutations for Mendelian disorders. *Hum. Genet.* **135**, 615–623 (2016).
10. R. P. Lifton, A. G. Gharavi, D. S. Geller, H. Hughes, Molecular Mechanisms

Review of Human Hypertension. **104**, 545–556 (2001).

11. S. Srivastava *et al.*, Clinical whole exome sequencing in child neurology practice. *Ann. Neurol.* **76**, 473–483 (2014).
12. D. Trujillano *et al.*, Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur. J. Hum. Genet.* **25**, 176–182 (2016).
13. Y. Yang *et al.*, Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *Jama*. **312**, 1870 (2014).
14. G. Baynam *et al.*, The rare and undiagnosed diseases diagnostic service - application of massively parallel sequencing in a state-wide clinical service. *Orphanet J. Rare Dis.* **11**, 77 (2016).
15. C. L. Beaulieu *et al.*, FORGE Canada consortium: Outcomes of a 2-year national rare-disease gene-discovery project. *Am. J. Hum. Genet.* **94**, 809–817 (2014).
16. W. A. Gahl *et al.*, The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. *Mol. Genet. Metab.* **117**, 393–400 (2016).
17. J. F. McRae *et al.*, Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. **542**, 433–438 (2017).
18. A. Cho *et al.*, Consecutive analysis of mutation spectrum in the dystrophin gene of 507 Korean boys with Duchenne/Becker muscular dystrophy in a single center. *Muscle and Nerve*. **55**, 727–734 (2017).
19. Y. Yoo *et al.*, GABBR2 mutations determine phenotype in rett syndrome and epileptic encephalopathy. *Ann. Neurol.* **82**, 466–478 (2017).
20. M. Deuchler, *The Confucian Transformation of Korea: A Study of Society and Ideology* (Harvard University Asia Center, Harvard University, ed. 1st, 1992; https://www.jstor.org/stable/j.ctt1dnn8zj).

21. K. J. Karczewski *et al.*, Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210 (2019).

22. S. Lee *et al.*, Korean Variant Archive (KOVA): A reference database of genetic variations in the Korean population. *Sci. Rep.* **7**, 1–9 (2017).

23. Q. Li, K. Wang, InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* **100**, 267–280 (2017).

24. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature*. **478**, 483–9 (2011).

25. X. Liu, X. Jian, E. Boerwinkle, dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **32**, 894–899 (2011).

26. S. J. Sanders *et al.*, Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron*. **70**, 863–885 (2011).

27. M. Uddin *et al.*, Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat. Genet.* **46**, 742–747 (2014).

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

29. P. Feliciano *et al.*, Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *npj Genomic Med.* **4** (2019), doi:10.1038/s41525-019-0093-8.

30. D. H. Cohn, P. H. Byers, B. Steinmann, R. E. Gelinas11, Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human proal(I)

collagen allele (mutation/gene cloning/DNA sequence analysis/extracellular matrix). *Genetics*. **83**, 6045–6047 (1986).

31. J. S. Lee *et al.*, Severe hypotonia and postnatal growth impairment in a girl with a missense mutation in COL1A1: Implication of expanded phenotypic spectrum of type I collagenopathy. *Brain Dev.* **39**, 799–803 (2017).

32. V. K. Jordan, H. P. Zaveri, D. A. Scott, 1p36 deletion syndrome: An update. *Appl. Clin. Genet.* **8**, 189–200 (2015).

33. A.-M. Bisgaard, L. N. Rasmussen, H. U. Moller, M. Kirchhoff, T. Bryndorf, Interstitial deletion of the short arm of chromosome 1 (1p13.1p21.1) in a girl with mental retardation, short stature and colobomata. *Clin. Dysmorphol.* **16**, 109–112 (2007).

34. P. M. Boone *et al.*, The alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. *Am. J. Hum. Genet.* **95**, 143–161 (2014).

35. R. Bharadwaj *et al.*, Conserved Chromosome 2q31 Conformations Are Associated with Transcriptional Regulation of GAD1 GABA Synthesis Enzyme and Altered in Prefrontal Cortex of Subjects with Schizophrenia. *J. Neurosci.* **33**, 11839–11851 (2013).

36. C. Leroy *et al.*, The 2q37-deletion syndrome: an update of the clinical spectrum including overweight, brachydactyly and behavioural features in 14 new patients. *Eur. J. Hum. Genet.* **21**, 602–612 (2013).

37. S. R. Williams *et al.*, Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am. J. Hum. Genet.* **87**, 219–228 (2010).

38. A. Kuechler *et al.*, Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur. J. Hum. Genet.* **23**, 753–760 (2015).

39. N. Dikow *et al.*, 3p25.3 microdeletion of GABA transporters SLC6A1 and SLC6A11 results in intellectual disability, epilepsy and stereotypic behavior. *Am. J. Med. Genet. Part A*. **164**, 3061–3068 (2014).

40. L. Lovrecic, S. Bertok, M. Žerjav Tanšek, A New Case of an Extremely Rare 3p21.31 Interstitial Deletion. *Mol. Syndromol.* **7**, 93–98 (2016).

41. A. Battaglia, J. C. Carey, S. T. South, Wolf-Hirschhorn syndrome: A review and update. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **169**, 216–223 (2015).

42. K. Hosoki *et al.*, Clinical phenotype and candidate genes for the 5q31.3 microdeletion syndrome. *Am. J. Med. Genet. Part A*. **158 A**, 1891–1896 (2012).

43. J. Zhao, S. E. Noon, I. D. Krantz, Y. Wu, A de novo interstitial deletion of 7q31.2q31.31 identified in a girl with developmental delay and hearing loss. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **172**, 102–108 (2016).

44. K. K. Abu-Amero *et al.*, A de novo marker chromosome derived from 9p in a patient with 9p partial duplication syndrome and autism features: genotype-phenotype correlation. *BMC Med. Genet.* **11**, 135 (2010).

45. T. Kleefstra, W. M. Nillesen, H. G. Yntema, *Kleefstra Syndrome* (1993; http://www.ncbi.nlm.nih.gov/pubmed/20945554).

46. D. I. Jonsson *et al.*, A de novo 1.13 Mb microdeletion in 12q13.13 associated with congenital distal arthrogryposis, intellectual disability and mild dysmorphism. *Eur. J. Med. Genet.* **55**, 437–440 (2012).

47. M. Doornbos *et al.*, Nine patients with a microdeletion 15q11.2 between

breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. *Eur. J. Med. Genet.* **52**, 108–115 (2009).

48. A. Crepel *et al.*, Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **156**, 243–245 (2011).

49. J. A. Rosenfeld *et al.*, Genotype-phenotype analysis of TCF4 mutations causing Pitt-Hopkins syndrome shows increased seizure activity with missense mutations. *Genet. Med.* **11**, 797–805 (2009).

50. R. D. Burnside, 22q11.21 deletion syndromes: A review of proximal, central, and distal deletions and their associated features. *Cytogenet. Genome Res.* **146** (2015), pp. 89–99.

51. A. Shcheglovitov *et al.*, SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients. *Nature*. **503**, 267–271 (2013).

52. D. Atasoy *et al.*, Deletion of CASK in mice is lethal and impairs synaptic function. *Proc. Natl. Acad. Sci.* **104**, 2525–2530 (2007).

53. K. Woodward, E. Kendall, D. Vetrie, S. Malcolm, Pelizaeus-Merzbacher disease: identification of Xq22 proteolipid-protein duplications and characterization of breakpoints by interphase FISH. *Am J Hum.Genet.* **63**, 207–217 (1998).

54. K. Shimojima *et al.*, Reduced PLP1 expression in induced pluripotent stem cells derived from a Pelizaeus-Merzbacher disease patient with a partial PLP1 duplication. *J. Hum. Genet.* **57**, 580–586 (2012).

55. H. Van Esch *et al.*, Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* **77**, 442–53 (2005).

56. J. Vandewalle *et al.*, Dosage-Dependent Severity of the Phenotype in Patients with Mental Retardation Due to a Recurrent Copy-Number Gain at Xq28 Mediated by an Unusual Recombination. *Am. J. Hum. Genet.* **85**, 809–822 (2009).

57. F. Fusco, M. D'Urso, M. G. Miano, M. V. Ursini, The LCR at the IKBKG Locus Is Prone to Recombine. *Am. J. Hum. Genet.* **86**, 650–652 (2010).

58. S. Ferdinandusse *et al.*, Clinical, biochemical, and mutational spectrum of peroxisomal acyl-coenzynie A oxidase deficiency. *Hum. Mutat.* **28**, 904–912 (2007).

59. H. lok Chung *et al.*, Loss- or Gain-of-Function Mutations in ACOX1 Cause Axonal Loss via Different Mechanisms. *Neuron*. **106**, 589-606.e6 (2020).

60. H. L. Rehm *et al.*, ClinGen — The Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).

61. Statistics Korea, Final Results of Birth Statistics in 2016 (2016), (available at http://kostat.go.kr/portal/eng/pressReleases/8/1/index.board?bmode=read&bSeq=&aSeq=363582&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=srch&sTarget=title&sTxt=birth).

62. M. C. McKenna, P. F. Schuck, G. C. Ferreira, Fundamentals of CNS energy metabolism and alterations in lysosomal storage diseases. *J. Neurochem.* **148**, 590–599 (2019).

63. T. J. Tracey, F. J. Steyn, E. J. Wolvetang, S. T. Ngo, Neuronal Lipid Metabolism: Multiple Pathways Driving Functional Outcomes in Health and Disease. *Front. Mol. Neurosci.* **11**, 1–25 (2018).

64. A. Cao, Y. W. Kan, The prevention of thalassemia. *Cold Spring Harb. Perspect. Med.* **3**, 1–15 (2013).

65. M. Kaback *et al.*, Tay-sachs disease—carrier screening, prenatal diagnosis and the molecular era: An international perspective, 1970 to 1993. *Obstet. Gynecol. Surv.* **49**, 321–322 (1994).

66. M. S. Watson *et al.*, Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet. Med.* **6**, 387–91 (2004).

67. C. Evers *et al.*, Impact of clinical exomes in neurodevelopmental and neurometabolic disorders. *Mol. Genet. Metab.* **121**, 297–307 (2017).

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

69. R. N. Doan *et al.*, Recessive gene disruptions in autism spectrum disorder. *Nat. Genet.* (2019), doi:10.1038/s41588-019-0433-8.

70. G. Bergant *et al.*, Comprehensive use of extended exome analysis improves diagnostic yield in rare disease: A retrospective survey in 1,059 cases. *Genet. Med.* **20**, 303–312 (2018).

71. D. M. Werling *et al.*, An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018).

72. M. M. Clark *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 Genomic Med.* **3**, 1–10 (2018).

73. B. B. Cummings *et al.*, Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* (2017),

doi:10.1126/scitranslmed.aal5209.

74. H. D. Gonorazky *et al.*, Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019).

75. J. Seo *et al.*, Tofacitinib relieves symptoms of stimulator of interferon genes (STING)???associated vasculopathy with onset in infancy caused by 2 de novo variants in TMEM173. *J. Allergy Clin. Immunol.* **139**, 1396-1399.e12 (2017).

76. L. Boccuto *et al.*, A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. *Hum. Mol. Genet.* **23**, 418–433 (2014).

77. J. S. Lee *et al.*, GM3 synthase deficiency due to ST3GAL5 variants in two Korean female siblings: Masquerading as Rett syndrome-like phenotype. *Am. J. Med. Genet. Part A*. **170**, 2200–2205 (2016).

78. A. V. Strickland *et al.*, Mutation screen reveals novel variants and expands the phenotypes associated with DYNC1H1. *J. Neurol.* **262**, 2124–2134 (2015).

79. S. Y. Kim *et al.*, Atypical presentation of infantile-onset farber disease with novel ASAH1 mutations. *Am. J. Med. Genet. Part A*. **170**, 3023–3027 (2016).

80. D. Yoo *et al.*, Early-onset generalized dystonia starting in the lower extremities in a patient with a novel ANO3 variant. *Park. Relat. Disord.* (2018), , doi:10.1016/j.parkreldis.2018.02.012.

81. H. Kim *et al.*, Diagnostic challenges associated with GLUT1 deficiency: Phenotypic variabilities and evolving clinical features. *Yonsei Med. J.* **60**, 1209–1215 (2019).

82. M. Hully *et al.*, From splitting GLUT1 deficiency syndromes to overlapping

phenotypes. *Eur. J. Med. Genet.* **58**, 443–454 (2015).

83. J. S. Lee *et al.*, Genetic heterogeneity in Leigh syndrome: Highlighting treatable and novel genetic causes. *Clin. Genet.* **97**, 586–594 (2020).

84. M. Alfadhel, B. Tabarki, SLC19A3 Gene Defects Sorting the Phenotype and Acronyms: Review. *Neuropediatrics*. **49**, 83–92 (2018).

85. I. J. C. Lamers *et al.*, Recurrent De Novo Mutations Disturbing the GTP/GDP Binding Pocket of RAB11B Cause Intellectual Disability and a Distinctive Brain Phenotype. *Am. J. Hum. Genet.* **101**, 824–832 (2017).

LIST OF ABBREVIATIONS

CADD	Combined Annotation Dependent Depletion
CH	Compound heterozygous
CNV	Copy number variation
DDD	Deciphering Developmental Disorders
GERP	Genomic Evolutionary Rate Profiling
gnomAD	Genome Aggregation Database
GO	Gene Ontology
KND	Korean neurodevelopmental disorder
LoF	Loss-of-function
NGS	Next Generation Sequencing
OMIM	Online Mendelian Inheritance in Man
phyloP	phylogenetic P-values
RHe	Rare hemizygous
Rho	Rare homozygous
SIFT	Sorting Intolerant From Tolerant
WES	Whole exome sequencing
WGS	Whole genome sequencing

국문초록

최근 차세대 염기서열 분석(NGS) 기술의 발전으로 희귀 질환의 병원성을 연구하는 데 많은 발전이 있었다. 주요 차세대 염기서열 분석 기술 중 하나인 전장 엑솜 염기서열 분석(WES)은 이제 희귀 질환의 임상 진단에 있어 없어서는 안 되는 중요한 도구가 되었다. 본 연구의 목적은 553 명의 미진단 소아 신경발달장애 환자를 분석하여 질병을 일으키는 변이를 식별하고 이를 임상 치료에 적용하는 것이다.

희귀질환 환자의 상당 부분은 열성 방식으로 유전되는 변이들로 인해 고통받고 있다. 본 논문은 이러한 열성 변이에 대한 자세한 이해와 분석에 중점을 두었다. 지금까지 희귀 질환을 일으키는 데 있어 de novo 변이에 대한 많은 선행 연구가 있었지만, 희귀 질환에 대한 열성 변이의 기여도에 대한 체계적인 이해는 여전히 부족한 실정이다. 이에 신경발달장애가 있는 553 명의 미진단 한국인 환자의 유전자 진단 및 변이의 발굴을 전장 엑솜 염기서열 분석 기술을 사용하여 수행하였다.

이전에 알려진 병원성 유전자와 유전자 복제 수 변이(CNV)를 포함한 우리의 WES 분석의 진단율은 47.4%였다. 또한, 본 연구에서 새롭게 발견된 질병 원인 유전자는 뇌 발달 과정에서 기존에 알려진 병원성 유전자와 밀접하게 연관되어

있었다. 기존에 보고된 병원성 변이를 가진 환자 중 35.1%가 열성 유전 방식으로 이러한 변이들을 부모에게서 물려받았다. 우리 코호트에서 열성 유전 방식으로 질환을 일으키는 유전자들은 기능 상실 돌연변이(Loss-of-function mutation)에 상대적으로 덜 취약하고 지질 대사 및 미토콘드리아 관련 경로와 깊은 연관성을 보였다. 이 같은 관찰 결과와 몇 가지 가정을 통해 우리는 건강한 한국인 17 명 중 약 1 명이 심각한 신경 발달 문제를 일으키는 병원성 열성 변이 중 적어도 하나를 가지고 있다고 추정해볼 수 있었다. 또한, 이번 열성 변이에 관한 분석은 변이 보인자 스크리닝에 대한 이들 유전자의 타당성을 평가하는 데 도움이 될 것으로 생각된다. 우리의 연구 결과는 희귀 신경발달질환 환자의 적절한 진단과 치료를 위한 열성 변이 검사의 기초가 될 수 있을 것이다.

본 연구에서는 유전적 분석과 기능적 연구를 통해 희귀 신경발달질환의 유전적 원인과 병원성을 규명할 수 있었으며, 또한 우리의 결과는 중앙 집중식 의료 시스템을 가진 국가에서 환자 관리를 개선하기 위한 전장 엑솜 염기서열 분석 기반 임상 진단의 유용성을 강조한다.

* 본 학위논문은 출판된 논문 (Lee *et al*., *Sci Rep*. 10(1):1413 (2020)) 을 기반으로 작성되었음.

주요어: 전장 엑솜 염기서열분석, 발달장애, 소아 질환, 신경학, 열성 변이, 변이 발굴

학 번: 2013-31180

감사의 글

박사과정을 마치고 마침내 학위논문을 제출하게 되었습니다. 이 자리를 빌려 그동안 도움을 주신 많은 분께 감사의 말씀 드립니다. 우선 학위 과정 동안 저를 지도해주신 최무림 교수님께 깊이 감사드립니다. 제게 연구의 방향을 제시해 주시고 항상 아낌없는 조언을 해주셔서 감사합니다. 박사과정 동안 늘 창의적이고 완벽을 추구하시는 교수님과 함께 연구할 수 있어서 영광이었습니다. 덕분에 연구자로서 한 단계 더 성장할 수 있었습니다. 그리고 바쁘신 와중에도 부족한 제 논문을 심사해주시고 좋은 조언을 해주신 채종희 교수님, 이용석 교수님, 최지엽 교수님, 이진숙 교수님께 감사드립니다. 또한, 제게 분자생물학 실험의 기초를 지도해주신 유해용 교수님께도 감사의 말씀 드립니다.

기능유전체학 연구실 동료 여러분께 정말 감사합니다. 여러분의 도움이 없었다면 박사과정을 마치기 힘들었을 겁니다. 연구실의 시작부터 함께했던 용진이 형, 상문이 형, 태경이, 재소 형에게 감사드립니다. 덕분에 즐겁게 연구실 생활을 할 수 있었습니다. 그동안 부족한 선배를 많이 도와준 연구실 후배분들께도 고맙습니다. 여러분께는 모범적인 선배가 되지 못해 항상 미안한 마음이 있습니다. 앞으로 우리 연구실이 더 발전할 수 있도록 모두 좋은 연구 하길 바랍니다.

긴 학위과정 동안 항상 저를 믿고 응원해주신 우리 가족들에게 감사의 말씀 드립니다. 오랜 시간 묵묵히 저를 지켜봐 주신 아버지, 어머니께 감사드립니다. 앞으로도 부모님께서 주신 은혜에 보답할 수 있는 자랑스러운 아들이 되도록 노력하겠습니다. 동생 화영이에게도 이 자리를 빌려 고마움을 전합니다. 또 항상 제게 물심양면으로 지원을 아끼지 않으신 장인어른, 장모님, 형님, 처형께 감사드립니다. 힘들 때마다 항상 아빠 기운 나게 해준 예쁜 우리 딸, 지효에게 고맙습니다. 마지막으로, 육아와 학업을 병행하느라 힘들었을 텐데 묵묵히 학위를 마친 자랑스러운 아내 이옥상 박사에게 깊은 감사의 말을 전합니다. 당신의 응원과 배려 덕분에 항상 더 힘을 낼 수 있었습니다. 사랑합니다.