MEDICAL UNIVERSITY – PLOVDIV

DEPARTMENT OF PEDIATRICS AND MEDICAL GENETICS

Dr. IGLIKA GALENOVA SOTKOVA-IVANOVA

MUTATIONAL AND EXPRESSION PROFILE OF PROTEIN-CODING GENES IN CHILDREN WITH SPECIFIC LANGUAGE IMPAIRMENT

AUTOREFERAT

On a dissertation for award of educational and scientific degree "DOCTOR"

Professional field: "Medicine"

Doctoral Program "Medical Genetics"

Scientific supervisor: : prof. Dr. Vili K. Stoyanova, MD

PLOVDIV, 2023

The dissertation work has been approved and scheduled for defense by the extended departmental council of the Department of Pediatrics and Medical Genetics "Prof. Ivan Andreev" at the Faculty of Medicine of the Medical University of Plovdiv.

The dissertation is written on 148 typewritten pages and illustrated with 20 tables, 19 figures. The bibliographic reference includes 158 titles, of which 4 are in Cyrillic and 154 are in Latin.

Scientific jury:

Chairman:

Associate Professor Dr. Ina Evgenieva Geneva, MD

Members:

Prof. Dr. Ivanka Istilianova Dimova, MD

Prof. Dr. Daniela Mircheva Avdzhieva-Tzavella

Prof. Dr. Radka Stefanova Tincheva, MD

Associate Professor Dr. Radoslava Vazharova, MD

The public defense of the dissertation work will take place on 04.12.2023. from 2 p.m. in the Second Auditorium of the Auditorium Complex of MU-Plovdiv, 15 A "Vasil Aprilov" Blvd.

The defense materials are available in the Scientific Department and are published on the website of MU-Plovdiv

Content

I. Introduction:

Specific Language Impairment (SLI) is a socially significant neurodevelopmental disorder with a prevalence of approximately 7-10% among preschool and school-age children. It is characterized by disproportionately delayed language development with normal nonverbal intelligence. The language deficit in these children cannot be attributed to cognitive, motor, sensory, or socio-emotional factors(1).

To date, there is no unified consensus on common concepts and differentiation of phenotypic subcategories. The latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) from 2013 describes the criteria for this disorder, which is part of the broader category known as communication disorders. In recent years, some authors have used the more general term Developmental Language Disorder (DLD), encompassing language impairment and some frequently associated conditions, assumed to have a common etiology, such as speech disorders, fluency disorders, and dyslexia.

In this study, we have adopted the term Specific Language Impairment (SLI) as the most accurate descriptor of the profile of most children participating in our research. In children with SLI, first words and phrases appear later, their vocabulary is limited and repetitive, sentences are shorter and simpler with grammatical errors. They often struggle to find the right words and use synonyms. By the time they reach early school age, these children have lower academic achievement compared to their peers. SLI is often accompanied by other communication disorders, mental disorders, and dyslexia.

In 50% of cases, the problem is overcome in mature age. However, for the remaining individuals, language disorders significantly impact their entire lives. Communication deficits disrupt social, emotional, and educational development, increasing the risk of behavioral disorders that can ultimately lead to unemployment and mental health problems in adulthood.

Despite the exceptionally high social and economic burden of language disorders, we still know little about their biological basis. In recent years, numerous experimental studies using various

modern imaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) aim to provide a more accurate and detailed description of all brain structures involved in the perception, understanding, and production of language. Another subset of experiments using the same modern imaging techniques describes presumed deviations in neuroanatomical development associated with SLI in the search for etiology.

Both genetic and environmental factors influence a child's language development, exerting positive or negative effects. Complications during pregnancy, maternal smoking, fetal prematurity, maternal education, and socio-economic status, as well as family history, are important predictive risk factors for language deficit, especially in the presence of genetic predisposition.

In the quest for the genetic etiology of SLI, numerous studies employing various genetic approaches have been conducted with the aim of identifying the "gene for SLI." However, it has recently become clear that none of the candidate genes can independently be linked to SLI, and in most cases, the disorder has a more complex basis involving the interaction between several genetic and environmental risk factors.

The objective of this study is to characterize the mutational and expression profile of protein-coding genes in children with Specific Language Impairment. Quantitative analysis of gene expression in the analyzed samples (SLI group compared to a control group of healthy children) is conducted using RNA sequencing of pooled samples from both groups, followed by individual validation of differentially expressed genes and whole exome sequencing.

II. Aim and tasks

1.Aim:

The aim of this study is to characterize the mutational and expression profiles of protein-coding genes in peripheral blood in children with Specific Language Impairment (SLI).

2.Tasks:

1. To extract DNA and RNA from samples of children with SLI and healthy controls.

2. To determine the total expression profile through comprehensive RNA sequencing of children with SLI and healthy controls.

3. To identify the mutational profile of protein-coding genes in children with SLI with whole exome sequencing.

4. To conduct bioinformatic analysis of the sequencing data and differentially expressed genes.

5. ToeEnrich the DNA collection from children with SLI for genetic and genomic studies at the Department of Pediatrics and Medical Genetics, "Prof. Ivan Andreev," Faculty of Medicine, Medical University, Plovdiv.

III. Materials and Methods:

1.Subjects:

In this study, 40 children with confirmed Specific Language Impairment (SLI), aged between 3 and 12 years, were included. Additionally, 24 healthy controls matched for gender and age were selected from patients at the Pediatric Neurology Department of the Clinic of Pediatrics and Medical Genetics at the University Multiprofile Hospital for Active Treatment "St. George."

1.1.The following criteria were considered during sample selection for inclusion in the SLI group:

1. Signed informed consent from parents, approved by the ethical committee of the Medical University of Plovdiv, protocol T.2_KNE_session No. 3/2020, within the intramural project NO/18–2020 "Analysis of Gene Expression in Children with Specific Language Impairment."
2. Delay in language development using a detailed methodology for investigating receptive and expressive language, adapted in Bulgaria (e.g., TERB by A. Georgieva).
3. Delay in a non-word repetition test (e.g., Mottier).
4. Normal intelligence - HAWIK-R (Wechsler Intelligence Scales for Children) for children over 5 years and 6 months, with the use of Raven's matrices for children aged 3-5.5 years.
5. Non-verbal quotient from HAWIK-R is above 85% and greater than the verbal quotient.
6. Absence of hearing deficit - no history of ear diseases, normal hearing screening based on documentation.
7. Absence of generalized or oral motor disorders that may cause articulation problems.

8.Exclusion of Autism Spectrum Disorder with operationalized criteria based on DSM-5 and ADOS.

1.2 The following exclusion criteria were considered during sample selection:

1. Epilepsy.
2. Cerebral Palsy (CP), even without established structural brain lesion on imaging.
3. Presence of focal neurological symptoms (indicative of structural brain lesion).

4. Proven structural brain lesion on imaging (CT, MRI), even without CP or focal neurological symptoms.
5. Suspicious historical or clinical data of chronic somatic, neurological, or other psychiatric disorders: Autism Spectrum Disorder, ADHD, etc.
6. Presence of acute infectious disease (Acute infectious diseases usually lead to leukocytosis, i.e., an increase in the volume of the leukocyte mass, i.e., a change in the quantity of RNA).
7. Family history burdened with hereditary metabolic disorders.
8. Familial association with schizophrenia or other psychiatric pathology.
9. Refusal to participate in the study.
10. Presence of intellectual disability.
11. Bilingualism.
12. Refusal to participate in the study

2. Methods:

2.1. Venous Blood Collection Molecular-biological investigations, the subject of this study, were conducted using venous blood collected and stored in Pax Gene Tube systems according to the manufacturer's specifications, for total RNA and monovettes with EDTA for genomic DNA isolation. Thin needles, venocaths, or "butterflies" were used for venous blood collection. Monovette filling was carried out in a strictly vertical position to prevent blood from flowing along the tube walls. Literature review indicates that sufficient cell lysis and the highest RNA yield from the solution are achieved after the monovettes have been kept at room temperature for more than 2 hours and less than 24 hours (Wang, 2004). Therefore, in the present study, after collection, venous blood was maintained at room temperature for between 4 and 8 hours. Subsequently, the samples were frozen and stored at -20 °C until RNA isolation. If sample transportation was required beyond the 8-hour period, they were stored in a cool bag at 2 - 4 °C, covered with ice.

2.2. Isolation of Total RNA and Genomic DNA from Blood Samples in Children with ASD and the Control Group 2.2.1 Extraction of Total RNA from Peripheral Blood Using PAXgene Blood miRNA Kit Isolation, labeling, and storage of total RNA samples were carried out using the PAXgene Blood miRNA Kit following the manufacturer's instructions, which include the following general steps:

- Isolation begins after thawing the PAX Gene Tube.
- The blood in the PAXgene tube is centrifuged for 10 minutes at 3000-5000 xg in a swinging bucket rotor.
- The supernatant is removed by pipetting, and 4 ml of RNase-free water is added.
- After vortexing until the pellet is visibly dissolved, it is centrifuged for another 10 minutes at 3000-5000 xg.
- The supernatant is removed, and 350 µl of BM1 buffer is added, followed by careful mixing.
- The sample is pipetted into a 1.5 ml microcentrifuge tube, and 300 µl of BM2 buffer and 40 µl of proteinase K are added.
- The components are mixed by vortexing for 5 seconds, and the mixture is incubated for 10 minutes at 55 °C in a shaking incubator.
- The sample is pipetted into a Shredder PAXgene column placed in a 2 ml microcentrifuge tube and centrifuged for 3 minutes at maximum speed.
- The supernatant is transferred carefully to the Shredder PAXgene column without disrupting the integrity of the pellet.
- 700 µl of isopropanol (100% pure, for analytical purposes) is added, and the solution is mixed by vortexing.
- 700 µl of the sample is pipetted into the Spin PAXgene RNA column placed in a 2 ml microcentrifuge tube and centrifuged for 1 minute at 8,000 - 20,000 xg.
- The Spin PAXgene RNA column is placed for centrifugation in a new 2 ml microcentrifuge tube.
- 50 µl of BM3 buffer is added to the PAXgene RNA column and centrifuged for 15 seconds at 8,000 - 20,000 xg.
- 10 µl of DNase I stock solution and 70 µl of Buffer RDD are added to a 1.5 ml microcentrifuge tube. The mixture is carefully stirred and centrifuged briefly.
- 80 µl of DNase I mix is pipetted directly onto the RNA PAXgene column and incubated for 15 minutes at (20-30°C).
- 350 µl of BM3 buffer is added to the PAXgene RNA column, and it is centrifuged for 15 seconds at 8,000 - 20,000 xg.
- 500 µl of BM4 buffer is added to the PAXgene RNA column, and it is centrifuged for 15 seconds at 8,000 - 20,000 xg.

- The PAXgene RNA column is placed in a new 1.5 ml microcentrifuge tube, and 40 µl of BR5 buffer is pipetted directly onto the center of the membrane. It is then centrifuged for 1 minute at 8,000 - 20,000 xg for elution of the column-bound RNA.
- Step 18 (elution of the column-bound RNA) is repeated using 40 µl of BR5 buffer in the same microcentrifuge tube.
- The eluted RNA is incubated for 5 minutes at 65°C in a shaking incubator without shaking. Incubation at 65°C results in RNA denaturation.

2.2.2 Extraction of Genomic DNA from Peripheral Blood Using QIAamp® DNA Mini and Blood Mini per the manufacturer's instructions

- 20 µl of QIAGEN protease (or proteinase K) is pipetted onto the bottom of a 1.5 ml microcentrifuge tube.
- 200 µl of the sample (whole blood) is added to the microcentrifuge tube.
- 200 µl of buffer AL is added to the sample.
- To ensure effective lysis, it is crucial to mix the sample and buffer AL well to obtain a homogeneous solution.
- Incubation is carried out at 56°C for 10 minutes.
- The 1.5 ml microcentrifuge tube is briefly centrifuged to remove droplets from the inside of the lid.
- 200 µl of ethanol (100%) is added to the sample, and it is mixed again by pulsing vortexing for 15 s.
- The mixture is carefully applied to the QIAamp Mini centrifuge column (in a 2 ml tube) and centrifuged at 8000 rpm for 1 min.
- 500 µl of buffer AW1 is added to the QIAamp Mini centrifuge column, and it is centrifuged for 1 minute.
- 500 µl of buffer AW2 is added, and the column is centrifuged at full speed (20,000 x g; 14,000 rpm) for 3 minutes.
- The QIAamp Mini centrifuge column is placed in a clean 1.5 ml microcentrifuge tube, and 200 µl of buffer AE or distilled water is added.
- The mixture is incubated at room temperature (15–25°C) for 1 minute and then centrifuged at 6000 x g (8000 rpm) for 1 min.
- Incubating the rotating QIAamp Mini column loaded with buffer AE or water for 5 minutes at room temperature before centrifugation usually enhances DNA yield.
- Elution with an additional 200 µl of buffer AE also increases DNA production by up to 15%.

2.3. Quantitative Analysis of Isolated Samples 2.3.1 Quantitative Analysis of Isolated Total RNA Samples The spectrophotometric determination of the concentration of isolated total RNA was performed using the Epoch Micro-Volume Spectrophotometer System (BioTec).

2.3. Quantitative Analysis of Isolated Samples:

2.3.1 Quantitative Analysis of Isolated Total RNA Samples:

The concentration of isolated total RNA was determined using the Epoch Micro-Volume Spectrophotometer System (BioTec). Two microliters of each sample were used for spectrophotometric measurement at wavelengths λ=260 and λ=280 nm. The measured A260/A280 ratio demonstrated that all isolated samples were of sufficient quality for the planned analyses (1.93 - 2.10).

2.3.2 PCR-Based Control Detection for DNA Rejection and Purification:

To remove genomic DNA contamination in samples of extracted total RNA showing the presence of genomic DNA after PCR control detection, samples were treated with DNAse I. Subsequently, they were retested for the presence of DNA. Following this testing, all RNA samples exhibited the absence of genomic DNA and were utilized to create a complex RNA pool.

2.3.3 Quantitative Analysis of Isolated Genomic DNA:

Quantification of genomic DNA was conducted using the QFX fluorometer Denovix.

2.4. Analysis of the Expression Profile of Informational RNA (RNA Sequencing):

2.4.1 Creation of Complex Samples (Pool) and Preparation of Samples for Analysis (RNA Sequencing):

After measuring the concentrations of all isolated samples, the necessary quantity for each sample was calculated to create a complex pool from all samples of children diagnosed with specific language impairment and corresponding controls. Sample mixing was performed on dry ice to minimize the possibility of RNA degradation. For successful transportation and storage of isolated RNA, samples underwent precipitation as follows:

- Addition of 3 volumes of absolute ethanol.
- Addition of 1/10 volume of 3M NaOAc, pH 5.2.

2.4.2 RNA Sequencing:

RNA sequencing, also known as "Whole Transcriptome Shotgun Sequencing," is a revolutionary tool for transcriptomic studies. The choice of a suitable platform for quantitative RNA analysis through RNA sequencing was dictated by the good reproducibility of results from the BGI platform for transcriptome analysis, as well as the low detection limit (sensitivity) of the proposed analysis. Total RNA samples initially underwent treatment with DNAse I to degrade any possible contamination of total RNA samples with genomic DNA. Subsequently, enrichment of informational RNA molecules was performed using oligo(dT) magnetic beads. By mixing informational RNA molecules with a fragmentation buffer, informational RNA was fragmented into short fragments of about 200 base pairs. A protocol for synthesizing a single-stranded DNA chain was then carried out using non-specific (random) hexamer primers, followed by the addition of buffer, dNTPs, RNase H, and DNA polymerase I for the synthesis of the second DNA strand. The resulting double-stranded DNA was purified using magnetic beads. After purification of the double-stranded DNA, end repair and addition of a single nucleotide A (adenine) to the 3'-end were performed. Subsequently, adapters for sequencing were added to the obtained fragments. The fragments were enriched by PCR amplification. During the quality assessment (QC) stage, Agilent 2100 Bioanalyzer and ABI StepOne Plus Real-Time PCR system were used for quantitative and qualitative analysis of the libraries. The obtained libraries were then subjected to sequencing using Illumina HiSeqTM 2000.

2.4.3 Quality Assessment of RNA Sequencing:

Table 1 provides a summary of the results of sequencing, mapped to reference genes.

Sample ID	Total Reads	Total BasePairs	Total Mapped Reads	Perfect Match	<=2bp Mismatch	Unique Match	Multi-position Match	Total Unmapped Reads
CT_RNAS	28,358,958(100.00%)	1,389,588,942(100.00%)	17,620,664(62.13%)	13,832,018(48.77%)	3,788,646(13.36%)	9,108,179(32.12%)	8,512,485(30.02%)	10,738,294(37.87%)
SL_RNAS	24,905,731(100.00%)	1,220,380,819(100.00%)	15,124,070(60.73%)	12,097,003(48.57%)	3,027,067(12.15%)	7,821,425(31.40%)	7,302,645(29.32%)	9,781,661(39.27%)

2.4.4 Filtering Sequencing Data:

Due to the presence of adapter sequences and/or sequenced fragments (reads) with low quality in the raw data, data filtering was performed to obtain high-quality sequencing data. The filtering procedure included the following steps:

1. Removal of sequenced fragments with adapter sequences.
2. Removal of sequenced fragments in which the percentage of unknown bases (N) was greater than 10%.
3. Removal of sequenced fragments with low quality. If the percentage of bases with low quality (base with a value ≤ 5) was greater than 50% in the fragments, these fragments were identified as low-quality.

2.4.5 Quantitative Analysis of Gene Expression:

The expression level of each analyzed gene was determined by the number of sequenced fragments mapped to the sequence of the specific gene relative to the total number of fragments mapped to the respective gene in the analyzed sample.

2.4.6 Investigation of Gene Coverage:

Gene coverage, representing the percentage of the gene covered by sequenced fragments, was determined by the ratio of the number of bases covered by specifically mapped fragments to the total number of bases in the respective gene.

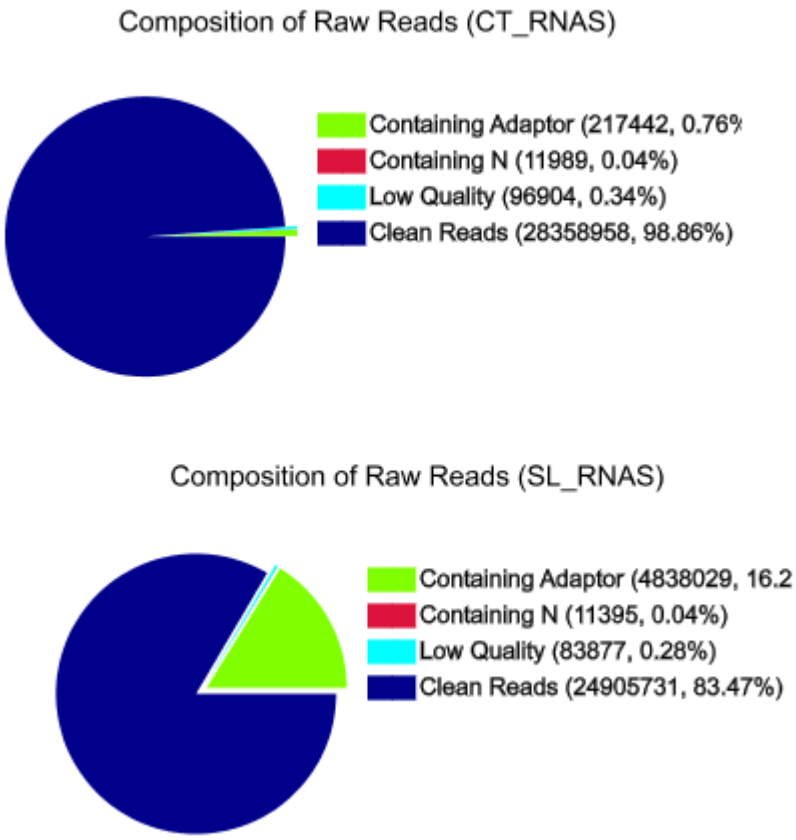


Figure 1: Quality Assessment of Sequenced Fragments from RNA Libraries

2.4.7 Investigation of Gene Expression Levels:

Gene expression levels are calculated using the RPKM (Reads Per kb per Million reads) method (Mortazavi, 2008), representing sequenced fragments per kilobase per million fragments. The formula is as follows:

$$RPKM = \frac{10^6 C}{NL/10^3}$$

Where:

- (A) is the expression level of gene A.
- C is the number of sequenced fragments uniquely mapped to gene A.
- N is the total number of sequenced fragments uniquely mapped to all genes.
- L is the length of gene A in bases.

The RPKM method eliminates the influence of variations in gene length and sequencing discrepancies when calculating gene expression levels. Therefore, RPKM values can be directly used to compare differences in gene expression between analyzed samples.

2.4.8 Identification of Differentially Expressed Genes (DEGs):

For the identification of differentially expressed genes, the p-value serves as an indicator of differential gene expression, and the False Discovery Rate (FDR) assesses the threshold of the p-value. In this work, FDR ≤ 0.001 and an absolute value of the log2 ratio ≥ 1 were used as criteria to evaluate differences in gene expression.

2.4.9 Gene Ontology Analysis of Differentially Expressed Genes:

Gene Ontology (GO) is an internationally standardized classification of gene function that provides a controlled and dynamically updated database with strictly defined gene and product properties. GO encompasses three aspects: cellular component, molecular function, and biological process. The basic unit of GO is the GO term (element). The analysis presents all GO terms containing significantly differentially expressed genes compared to the genome and filters those that correspond to biological functions.

2.4.10 REACTOME Analysis of Pathways Enriched with Differentially Expressed Genes:

Genes typically interact with each other to carry out specific biological functions. Reactome is a primary publicly accessible database related to biological pathways and gene interactions. The analysis identifies metabolic and signaling pathways associated with differentially expressed genes compared to the entire genome.

2.5. Analysis of Individual Expression Profiles of Differentially Expressed Genes in Specific Language Impairment (SLI):

2.5.1 Generation of cDNA from Total RNA:

Reverse transcription of previously isolated total RNA was performed to obtain cDNA using the Maxima® First Strand cDNA Synthesis Kit from Thermo Scientific, according to the manufacturer's instructions.

1. In a sterile, RNase-free Eppendorf tube, the following reagents were added in the specified order:
    - 5X reaction mix: 4 µl
    - Maxima Enzyme Mix: 2 µl
    - RNA (1 pg – 5 µg): Variable
    - Nuclease-free water: Up to 20 µl
    - Total reaction volume: 20 µl
2. Gently mixed and centrifuged.
3. Incubated for 10 minutes at 25 °C, followed by 15 minutes at 50 °C.
4. Reaction termination by heating to 85 °C for 5 minutes.

   The synthesized product is the first strand of cDNA, which can be used directly in quantitative PCR reactions.

2.6. Whole Exome Sequencing

Figure 2: Whole Exome Sequencing

In our study, genomic DNA was isolated from peripheral blood using the described protocol. The DNA was sent to Novogene Corporation Inc., where sequencing was performed. DNA was fragmented and prepared as "libraries" containing double-indexed sequencing barcodes. Libraries were enriched using SureSelect Human AII Exon V6 (Agilent Technologies) capture probes targeting coding regions and sites of 20,000 human genes. After capture, libraries were sequenced using the NovaSeq 6000 platform (Illumina).

Sequence data were analyzed using a specially developed bioinformatics pipeline, aligning the sequences against the reference human genome (GRCh37/GRCh38) for variant calling and annotation. The pipeline conducted quality control (QC) analysis on sequence data to ensure the quality of reported sequence data.

Variants were annotated using the Variant Effect Predictor (VEP) tool. VEP can utilize various annotation sources to extract transcript models used to predict variant types. Primary databases used include RefSeq, dbSNP, COSMIC, ClinVar, 1000 Genomes, NHLBI-ESP, genomAD, SIFT, PolyPhen, HGMD-PUBLIC, among others. Variant annotations followed the recommendations of the Human Genome Variation Society (HGVS).

Within each proband, we generated a set of genes corresponding to transcripts harboring novel or rare (≤ 1% population frequency) nonsynonymous SNVs and stop codon variants predicted as damaging by SIFT (Sorting Intolerant From Tolerant) and Polyphen (Polymorphism Phenotyping). SIFT and Polyphen are widely used computational tools for predicting the potential impact of amino acid substitutions on protein function. They are frequently employed in bioinformatics and genetic research to aid in identifying variants that may be pathogenic. SIFT is a program that predicts whether amino acid substitution in a protein will be tolerated or damaging based on sequence homology and physical properties of the involved amino acids. Results from SIFT range from 0 to 1, with results below 0.05 indicating potential harmful effects. PolyPhen-2 (Polymorphism Phenotyping v2) is another program that predicts the potential impact of amino acid substitutions on protein function. It uses a combination of sequence- and structure-based features to calculate a score ranging from 0 to 1, with scores greater than 0.85 indicating potential harmful effects.

2.7. Gene Ontology Analysis of Genes with Variants

Gene Ontology (GO) is an internationally standardized classification of gene function, offering a controlled and dynamically updated database with rigorously defined properties of genes and their products. GO encompasses three aspects: cellular component, molecular function, and biological process. The fundamental unit of GO is the GO term (element). Each GO term belongs to a type of ontology. The analysis provides all GO terms containing significant genes with variants compared to the genome and filters those corresponding to biological functions. This analysis first maps

genes to GO elements in the database, calculates the number of genes in each GO element, and then, through geometric exploration, identifies GO elements significantly enriched in our list compared to the genome.

III. Results

1.Patient Description, Gender, IQ, Risk Factors for Specific Language Impairment (SLI)

1.1 Distribution of patients by gender is presented in the figure.

Out of 40 SLI patients, 25 are boys and 15 are girls, with a ratio of 1:64, a distribution consistent with the population's sex distribution.

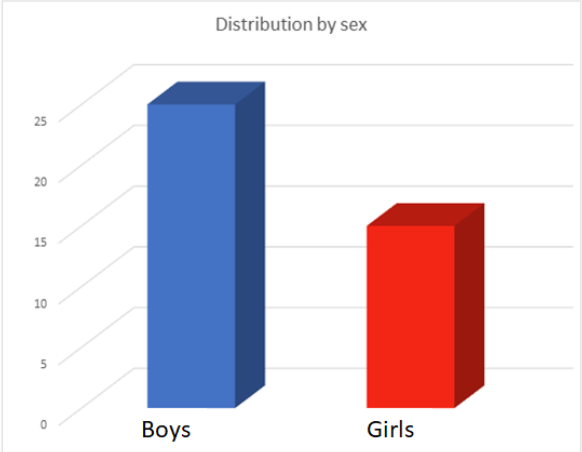


Figure 3: Distribution by Gender of Patients

1.2 Figure 13 graphically presents the overall Intelligence Quotient (IQ) of all patients assessed using the Wechsler Intelligence Scale for Children (HAWIK-R Bulgarian version).

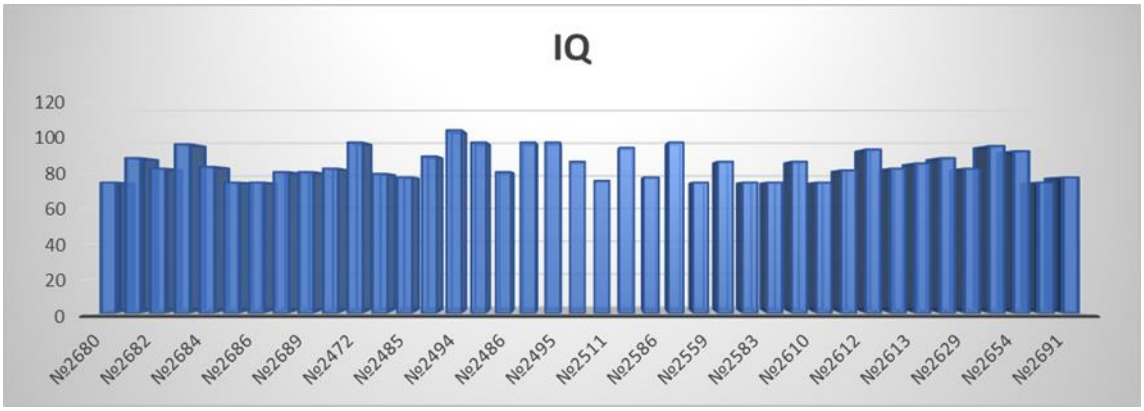


Figure 4: Graphical Representation of Overall Intelligence Quotient (IQ)

1.3 Figure 14 graphically represents verbal and non-verbal intelligence. In all cases, non-verbal intelligence is above 80 and at least 1 standard deviation greater than verbal intelligence.

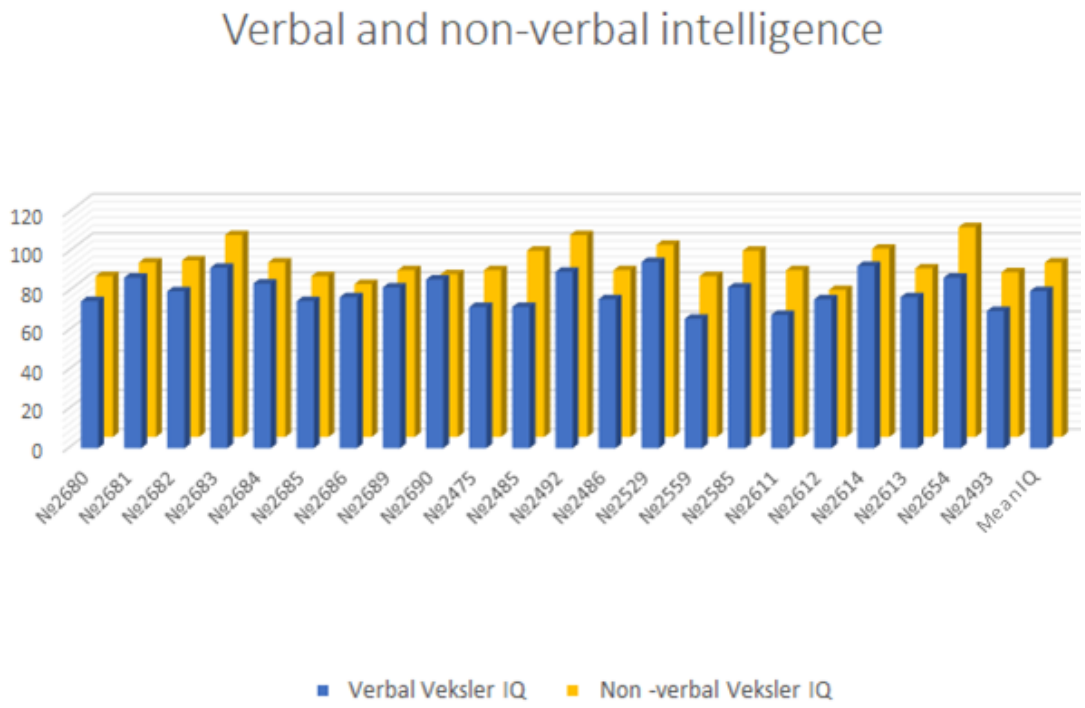
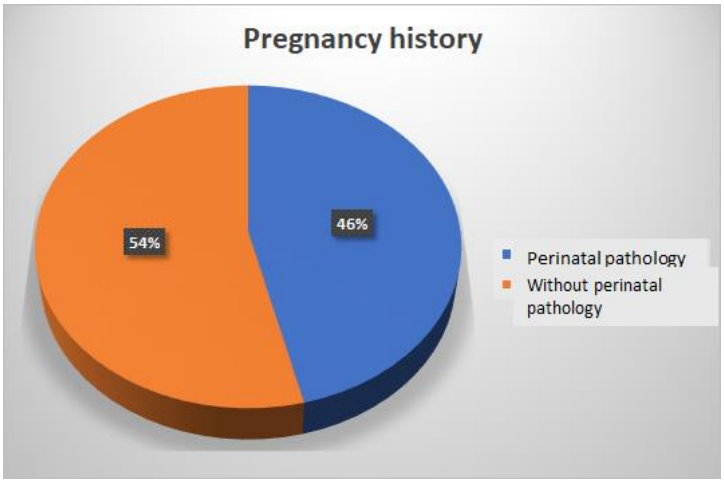


Figure 5: Verbal and Non-verbal Intelligence

1.4 Risk Factors Several risk factors are graphically presented, namely: pathology during pregnancy, mother's age at childbirth, and mother's education.



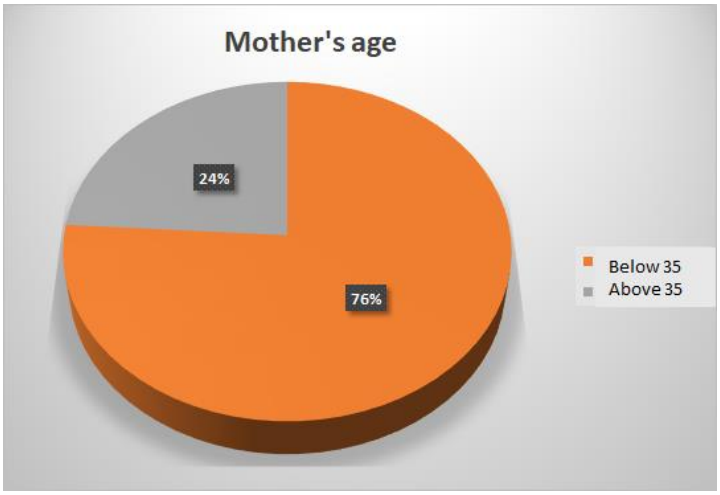


Figure 6: Presence of Perinatal Pathology Figure 7: Age of Mothers at Childbirth

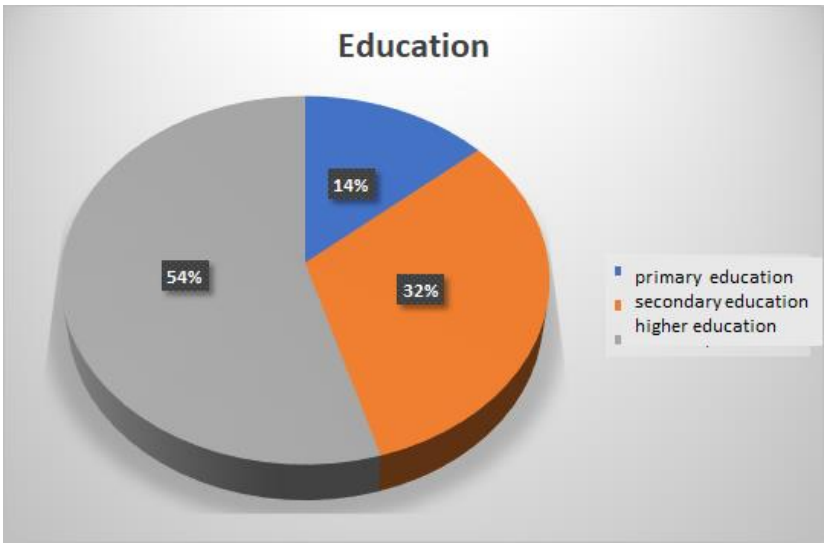


Figure 8: Education of Mothers

2.Differentially Expressed Genes Between Analyzed Groups Obtained from Transcriptomic Analysis

All analyzed information RNA molecules (transcripts) in the pooled samples of children with Specific Language Impairment (SLI) are compared with those of the control group. Sixty-one genes with statistical significance ($p < 0.001$, FDR ≤ 0.05, and $\log2 \geq 1$) are identified—47 genes with increased expression and 14 with decreased expression.

Genes with increased expression include*: GINS2, GLDC, ISG15, RSAD2, CD177, IFI44L, IGJ, BATF2, ANKRD22, RAP1GAP, IGLL5, CMPK2, SERPING1, IFIT1, TPX2, HERC5, MRPL40, C22orf23, CD38, MX1, SMTNL1, USP18, MBNL1-AS1, IGLL3P, OAS3, ETV7, CHPF, IFI44, TXNDC5, MIR650, IFIT3, OASL, BUB1, EPSTI1, SIGLEC1, MYBL2, IFI6, OTOF, ELL2, IKZF3, LINC00852, SEC61G, MZB1, TOP2A, SPAG5, LY6E, OAS1*.

Genes with decreased expression include: *FLNC, LOC100505716, LOC284379, LINC00470, RND2, HSD17B13, DET1, MORN4, NKX3-1, TMEM254-AS1, HEXA-AS1, SLITRK4, PRELP, LOC440335.*

Table 2: Genes with Increased Expression

Gene	Description	Regulation	Log ratio	P-value	False Discovery Rate
GINS2	GINS complex subunit 2 (Psf2 homolog)	Up	2.28589686401175	3.4729e-07	1.25125482328482e-05
GLDC	glycine dehydrogenase (decarboxylating)	Up	1.97906734029827	1.201978e-09	6.71944475483871e-08
ISG15	ISG15 ubiquitin-like modifier	Up	1.77892772580895	4.6833e-58	2.31890254285714e-55
RSAD2	radical S-adenosyl methionine domain containing 2	Up	1.75395924345426	5.64896e-99	4.66173699047619e-96
CD177	CD177 molecule	Up	1.69613337702678	1.106724e-19	1.37982208057554e-17
IFI44L	interferon-induced protein 44-like	Up	1.67671979422668	3.08142e-178	4.45008405e-175
IGJ	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	Up	1.60574097565126	1.445492e-260	3.57862519428571e-257
BATF2	basic leucine zipper transcription factor, ATF-like 2	Up	1.58654403881194	4.20116e-25	7.20852502970297e-23
ANKRD22	ankyrin repeat domain 22	Up	1.55659963554394	4.18644e-08	1.83673430886076e-06
RAP1GAP	RAP1 GTPase activating protein	Up	1.4914154667838	1.917426e-10	1.16185288741259e-08
CMPK2	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	Up	1.42760389359811	9.35746e-22	1.2973182544e-19
SERPING1	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1	Up	1.41371049188973	6.7187e-138	7.2771919375e-135
IFIT1	interferon-induced protein with tetratricopeptide repeats 1	Up	1.40289259973935	1.610524e-63	9.00334868387097e-61
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	Up	1.40072872465918	1.047702e-11	7.262670264e-10
HERC5	hect domain and RLD 5	Up	1.38888517744452	7.35194e-53	2.96300279534884e-50
MRPL40	mitochondrial ribosomal protein L40	Up	1.38185885468285	1.032932e-08	5.08542942045455e-07
C22orf23	chromosome 22 open reading frame 23	Up	1.34757015563595	2.2084e-07	8.21278369098712e-06
CD38	CD38 molecule	Up	1.33982228811375	2.36186e-11	1.58647417829457e-09
MX1	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)	Up	1.29589900536642	1.053512e-293	3.651472592e-290
SMTNL1	smoothelin-like 1	Up	1.25480918622765	2.79226e-07	1.01659381932773e-05
USP18	ubiquitin specific peptidase 18	Up	1.23269905318882	8.33882e-13	6.42274447111111e-11
MBNL1-AS1	hypothetical LOC401093	Up	1.23218627726699	2.62906e-06	7.68323942664418e-05
IGLL3P	immunoglobulin lambda-like polypeptide 3, pseudogene	Up	1.22785385647095	9.74422e-07	3.21040556273764e-05
OAS3	2'-5'-oligoadenylate synthetase 3, 100kDa	Up	1.22208131440826	3.89738e-170	5.19550733846154e-167

ETV7	ets variant 7	Up	1.2042802806971	2.12686e-08	9.9349012938005 4e-07
CHPF	chondroitin polymerizing factor	Up	1.20202430675673	3.20046e-05	0.0007175158059 50841
IFI44	interferon-induced protein 44	Up	1.20202430675673	5.16808e-13	4.0526165791855 2e-11
TXNDC5	thioredoxin domain containing 5 (endoplasmic reticulum)	Up	1.20092552906395	1.93737e-05	0.0004624603595 04132
MIR650	microRNA 650	Up	1.19963177711904	1.724092e-14	1.6063717397849 5e-12
IFIT3	interferon-induced protein with tetratricopeptide repeats 3	Up	1.19684676888425	2.35152e-248	5.0939802e-245
OASL	2'-5'-oligoadenylate synthetase-like	Up	1.18585607349785	1.313512e-54	5.9903060421052 6e-52
BUB1	budding uninhibited by benzimidazoles 1 homolog (yeast)	Up	1.18156020419702	1.029832e-05	0.0002655801869 04762
EPSTI1	epithelial stromal interaction 1 (breast)	Up	1.17667798281727	2.92296e-32	6.665118e-30
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin	Up	1.17624640430055	1.173004e-15	1.1682850183908 e-13
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	Up	1.17507617062681	1.24828e-24	2.0602564190476 2e-22
IFI6	interferon, alpha-inducible protein 6	Up	1.1706931515316	1.10776e-184	1.7452255272727 3e-181
OTOF	otoferlin	Up	1.10094174398873	1.76221e-05	0.0004253356448 46797
ELL2	elongation factor, RNA polymerase II, 2	Up	1.09497251920092	1.634406e-10	1.0044062404255 3e-08
IKZF3	IKAROS family zinc finger 3 (Aiolos)	Up	1.0855911456663	3.72566e-10	2.1666338187919 5e-08
LINC00852	ghrelin opposite strand RNA 2 (non-protein coding)	Up	1.0842630028421	2.49128e-05	0.0005826434871 79487
SEC61G	Sec61 gamma subunit	Up	1.0842630028421	2.36888e-07	8.7907259957173 4e-06
MZB1	marginal zone B and B1 cell-specific protein	Up	1.05525762108332	9.60798e-29	1.9138654413793 1e-26
TOP2A	topoisomerase (DNA) II alpha 170kDa	Up	1.04829367333355	8.28892e-07	2.7571398003838 8e-05
SPAG5	sperm associated antigen 5	Up	1.02445924311228	3.91632e-06	0.0001105371752 443
LY6E	lymphocyte antigen 6 complex, locus E	Up	1.02202780669915	5.91234e-129	6.0271089529411 7e-126
OAS1	2'-5'-oligoadenylate synthetase 1, 40/46kDa	Up	1.00677689556142	6.2094e-89	4.48370425e-86

Table 3 Down-expressed genes

Gene	Description	Regulation	Log ratio	P-value	False Discovery Rate
FLNC	filamin C, gamma	Down	-2.277871414440028	2.24692e-06	6.67909495711835e-05
LOC100505716	hypothetical LOC100505716	Down	-2.11960933052355	1.205976e-08	5.88720114929577e-07
LOC284379	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3 pseudogene	Down	-1.9253549997195	1.481446e-06	4.6175286294964e-05
LINC00470	chromosome 18 open reading frame 2	Down	-1.373028647480628	8.46448e-07	2.80476937667304e-05
RND2	Rho family GTPase 2	Down	-1.254538910609665	5.50496e-06	0.000151912351592357
HSD17B13	hydroxysteroid (17-beta) dehydrogenase 13	Down	-1.191898191330421	1.13958e-06	3.68450026119403e-05
DET1	de-etiolated homolog 1 (Arabidopsis)	Down	-1.172076750417691	1.414638e-06	4.44930608711434e-05
MORN4	MORN repeat containing 4	Down	-1.164890209824412	2.4907e-05	0.000583295013513514
NKX3-1	NK3 homeobox 1	Down	-1.1549652669033	5.55544e-08	2.38306374257426e-06
TMEM254-AS1	hypothetical LOC219347	Down	-1.125868320995231	1.491772e-05	0.000367744079089616
HEXA-AS1	chromosome 15 open reading frame 34	Down	-1.116619287917741	1.63405e-11	1.11929195652174e-09
SLITRK4	SLIT and NTRK-like family, member 4	Down	-1.078440141547144	4.42018e-35	1.12649587352941e-32
PRELP	proline/arginine-rich end leucine-rich repeat protein	Down	-1.058490883635941	1.208684e-06	3.87898031851852e-05
LOC440335	hypothetical LOC440335	Down	-1.024300493249544	4.12262e-15	3.9255497032967e-13

3. Reactome Analysis of Biologic Pathways Enriched with Differentially Expressed Genes

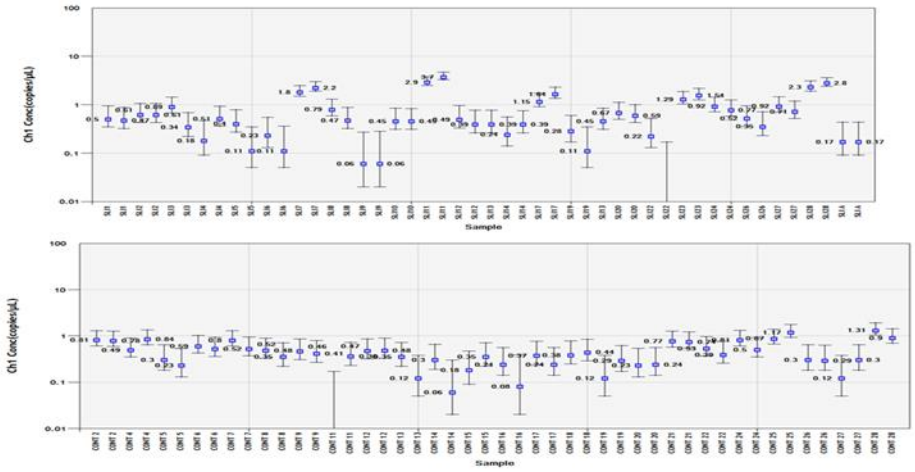
Table 4 presents statistically significant biological pathways involving differentially expressed genes.

Biological Pathway	Discovered Genes/Total Genes	P-value	FDR
Interferon Alpha/Beta Signaling	19/190	1.11e-16	2.31e-14
Interferon Signaling	27/397	4.44e-16	4.62e-14
Antiviral Mechanism by IFN-Stimulated Genes	9/94	1.27e-09	8.75e-08
Cytokine Signaling in the Immune System	21/1115	7.81e-08	4.06e-06
ISG15 Antiviral Mechanism	6/83	4.12e-06	1.69e-04
G0 and Early G1	4/38	4.30e-05	0.001
Immune System	28/2703	6.04e-05	0.002
Antiviral Response by OAS	3/15	6.35e-05	0.002
TFAP2A Acts as a Transcriptional Repressor during Retinoic Acid-Induced Cell Differentiation	2/9	9.65e-04	0.022
Interferon Gamma Signaling	6/252	0.002	0.033

Table 4: Biological Pathways Involving Differentially Expressed Genes Compared to the Total Number of Genes in These Pathways Abbreviations: G0 and Early G1 - Cell cycle stages

4. Validation of Differentially Expressed Genes from Transcriptomic Pool Analysis through ddPCR Analysis

To validate the transcriptomic analysis data, ddPCR analysis was conducted on GINS2, which exhibited the most significant differential expression across all individual samples from children diagnosed with Specific Language Impairment (SLI) and healthy controls (Figure 1).



5.1.2.. Rare missense variants

Tab.5 Rare missense variants

Gene	Transcript	Amino Acid Substitution	Nucleotide Substitution	Zygosity	Global Frequency in GnomAD	Effect
BAP1	*NM_004656.4*	*p.Tyr401Asp*	*c.1201T>G*	*Het*	*1.3940915E-4*	*Non Synonymous*
CACNA1A	*NM_001127222.2*	*p.Val604Ile*	*c.1810G>A*	*Het*	*1.2074757E-5*	*Non Synonymous*
CACNA1C	*NM_001129829.2*	*p.His1321Arg*	*c.3962A>G*	*Het*	*3.0E-5*	*Non Synonymous*
CACNA1D	*NM_001128840.3*	*p.Tyr1972His*	*c.5914T>C*	*Het*	*3.1938067E-5*	*Non Synonymous*

CACNA1D	*NM_001128840.3*	*p.Tyr1972His*	*c.5914T>C*	*Het*	*3.1938067E-5*	*Non Synonymous*
CACNA1E	*NM_001205293.3*	*p.Gly900Val*	*c.2699G>T*	*Het*	*4.320774E-6*	*Non Synonymous*
CACNA1G	*NM_018896.5*	*p.Arg1210Gln*	*c.3629G>A*	*Het*	*3.2436186E-5*	*Non Synonymous*
CACNA1G	*NM_018896.5*	*p.Thr2219Met*	*c.6656C>T*	*Het*	*8.045959E-6*	*Non Synonymous*
CACNA1H	*NM_021098.3*	*p.Glu150Asp*	*c.450G>T*	*Het*	*2.0664003E-5*	*Non Synonymous*
CAMK2B	*NM_001220.5*	*p.Ala660Thr*	*c.1978G>A*	*Het*	*8.058167E-5*	*Non Synonymous*
CAMTA1	*NM_015215.4*	*p.Leu744Ile*	*c.2230C>A*	*Het*	*4.000576E-6*	*Non Synonymous*
CAPN12	*NM_144691.4*	*p.Ala112Thr*	*c.334G>A*	*Het*	*8.919825E-4*	*Non Synonymous*
CCDC103	*NM_213607.3*	*p.Asp138Ala*	*c.413A>C*	*Hom Alt*	*1.6007463E-5*	*Non Synonymous*
CDH23	*NM_022124.6*	*p.Arg2029Gln*	*c.6086G>A*	*Het*	*9.36029E-5*	*Non Synonymous*
CDK13	*NM_003718.5*	*p.Ala487Ser*	*c.1459G>T*	*Het*	*3.4680768E-4*	*Non Synonymous*
CDK19	*NM_015076.5*	*p.Gln476Glu*	*c.1426C>G*	*Het*	*3.9767756E-6*	*Non Synonymous*
CDK19	*NM_015076.5*	*p.Gln476Glu*	*c.1426C>G*	*Het*	*3.9767756E-6*	*Non Synonymous*
CHD5	*NM_015557.3*	*p.Thr478Met*	*c.1433C>T*	*Het*	*1.04267136E-4*	*Non Synonymous*

CHD5	*NM_015557.3*	*p.Val680Leu*	*c.2038G>T*	*Het*	*7.3030905E-4*	*Non Synonymous*
CHD8	*NM_001170629.2*	*p.Ile887Val*	*c.2659A>G*	*Het*	*N/A*	*Non Synonymous*
CHD8	*NM_001170629.2*	*p.Asn1967Lys*	*c.5901C>A*	*Het*	*1.2126646E-5*	*Non Synonymous*
CHD9	*NM_001308319.2*	*p.Asp648Gly*	*c.1943A>G*	*Het*	*1.0370639E-4*	*Non Synonymous*
CIC	*NM_001386298.1*	*p.Lys94Arg*	*c.281A>G*	*Het*	*3.5875095E-5*	*Non Synonymous*
CLCN6	*NM_001286.5*	*p.Ile258Val*	*c.772A>G*	*Het*	*8.385751E-5*	*Non Synonymous*
CLPB	*NM_001258392.3*	*p.Arg90His*	*c.269G>A*	*Het*	*1.20054665E-5*	*Non Synonymous*
CLPB	*NM_001258392.3*	*p.Asn194Asp*	*c.580A>G*	*Het*	*3.992499E-6*	*Non Synonymous*
CMIP	*NM_198390.3*	*p.Ser394Ala*	*c.1180T>G*	*Het*	*1.2374851E-4*	*Non Synonymous*
CREBBP	*NM_004380.3*	*p.Gly1316Cys*	*c.3946G>T*	*Het*	*N/A*	*Non Synonymous*
CREBBP	*NM_004380.3*	*p.Pro2255Ser*	*c.6763C>T*	*Het*	*3.6041292E-5*	*Non Synonymous*
CYP1B1	*NM_000104.4*	*p.Gly61Glu*	*c.182G>A*	*Het*	*2.9268535E-4*	*Non Synonymous*
FLCN	*NM_144997.7*	*p.Arg239Cys*	*c.715C>T*	*Het*	*2.617104E-4*	*Non Synonymous*
FLNA	*NM_001110556.2*	*p.Gly20Ala*	*c.59G>C*	*Hemi Alt*	*4.5804325E-5*	*Non Synonymous*

FOXP1	*NM_001349338.3*	*p.Val445Leu*	*c.1333G>T*	*Het*	*7.991803E-6*	*Non Synonymous*
GLI2	*NM_001374353.1*	*p.Gln394His*	*c.1182G>C*	*Het*	*N/A*	*Non Synonymous*
KCNJ2	*NM_000891.3*	*p.Gly65Arg*	*c.193G>A*	*Het*	*N/A*	*Non Synonymous*
KCNMA1	*NM_001322839.2*	*p.Arg215His*	*c.644G>A*	*Het*	*8.4597E-5*	*Non Synonymous*
KCNMA1	*NM_001161352.2*	*p.Asn302Lys*	*c.906T>A*	*Het*	*N/A*	*Non Synonymous*
KMT2A	*NM_001197104.2*	*p.His2140Gln*	*c.6420T>A*	*Het*	*3.991121E-6*	*Non Synonymous*
KMT2B	*NM_014727.3*	*p.Asp1698Asn*	*c.5092G>A*	*Het*	*2.8566794E-5*	*Non Synonymous*
KMT2C	*NM_170606.3*	*p.Asp457Glu*	*c.1371C>A*	*Het*	*1.9956104E-5*	*Non Synonymous*
KMT2C	*NM_170606.3*	*p.Ser3742Asn*	*c.11225G>A*	*Het*	*N/A*	*Non Synonymous*
L1CAM	*NM_001278116.2*	*p.Glu660Lys*	*c.1978G>A*	*Hemi Alt*	*4.3598764E-5*	*Non Synonymous*
SHANK1	*NM_016148.5*	*p.Trp1944Arg*	*c.5830T>C*	*Het*	*2.0579525E-4*	*Non Synonymous*
SHANK3	*NM_001372044.2*	*p.Arg1648Gln*	*c.4943G>A*	*Het*	*N/A*	*Non Synonymous*
UNC80	*NM_001371986.1*	*p.Asp2904Val*	*c.8711A>T*	*Het*	*N/A*	*Non Synonymous*
UNC80	*NM_001371986.1*	*p.Asp2904Asn*	*c.8710G>A*	*Het*	*3.8209746E-5*	*Non Synonymous*

UROD	*NM_000374.5*	*p.Pro235Ser*	*c.703C>T*	*Het*	*4.0131104E-6*	*Non Synonymous*
USP9X	*NM_001039591.3*	*p.Ala2468Gly*	*c.7403C>G*	*Het*	*N/A*	*Non Synonymous*
VCP	*NM_007126.5*	*p.Met508Leu*	*c.1522A>C*	*Het*	*3.976839E-6*	*Non Synonymous*
WFS1	*NM_006005.3*	*p.Pro504Leu*	*c.1511C>T*	*Het*	*3.91988E-5*	*Non Synonymous*
ZBTB20	*NM_001348800.3*	*p.Pro697Leu*	*c.2090C>T*	*Het*	*6.389405E-5*	*Non Synonymous*
ZBTB7A	*NM_015898.4*	*p.Tyr351Cys*	*c.1052A>G*	*Het*	*8.39959E-6*	*Non Synonymous*
ZBTB7A	*NM_015898.4*	*p.Tyr351Cys*	*c.1052A>G*	*Het*	*8.39959E-6*	*Non Synonymous*

5.1.3 Rare nonsense variants.

Table 6 truncating variants found in children with SLI

Gene	Transcript	Amino Acid Substitution	Nucleotide Substitution	Zygosity	Global Frequency in GnomAD	Effect
ABCA13	NM_152701.5	p.Pro269ThrfsTer6	c.804_805del	Het	7.1265677E-6	Frameshift
ADGRV1	NM_032119.4	p.Asp409IlefsTer27	c.1224del	Het	N/A	Frameshift
ALG2	NM_033087.4	p.Gln41Ter	c.121C>T	Het	4.4540257E-6	Stop Gain
BAP1	NM_004656.4	p.Tyr401Ter	c.1203T>G	Het	1.393729E-4	Stop Gain
BST1	NM_004334.3	p.Arg221Ter	c.661C>T	Het	1.4891373E-5	Stop Gain

CENPJ	NM_018451.5	p.Gln657Ter	c.1969C>T	Het	7.971685E-6	Stop Gain
CFAP57	NM_001378189.1	p.Arg15Ter	c.43C>T	Het	3.0E-5	Stop Gain
CLCN1	NM_000083.3	p.Arg894Ter	c.2680C>T	Het	0.0031781709	Stop Gain
CNTNAP4	NM_033401.5	p.Trp79*	c.237G>A	Het	1.6460432E-5	Stop Gain
CPZ	NM_001014447.3	p.Gln317Ter	c.949C>T	Het	3.9686152E-4	Stop Gain
CREB3L3	NM_032607.3	p.Lys245GlufsTer130	c.732dup	Het	4.4700474E-4	Frameshift
CYFIP2	NM_001037333.3	p.Arg524Ter	c.1570C>T	Het	N/A	Stop Gain
DIABLO	NM_001371333.1	p.Trp7Ter	c.20G>A	Het	N/A	Stop Gain
DLX6	NM_005222.4	p.Tyr118LeufsTer37	c.351dup	Het	N/A	Frameshift
DPYD	NM_000110.4	p.Pro705SerfsTer12	c.2112dup	Het	4.002465E-6	Frameshift
EDN3	NM_207034.3	p.Asp190GlnfsTer8	c.568_569del	HomAlt	1.2391349E-4	Frameshift
FBXO40	NM_016298.4	p.Arg692Ter	c.2074C>T	Het	8.50295E-5	Stop Gain
GCNT2	NM_001491.3	p.Asp168Ter	c.501dup	Het	2.789902E-5	Frameshift
GJB2	NM_004004.6	p.Gly12ValfsTer2	c.35del	Het	0.006186484	Frameshift
HSPA9	NM_004134.7	p.Val296Ter	c.882_883del	Het	1.2346091E-4	Frameshift
IQGAP3	NM_178229.5	p.Arg324Ter	c.970C>T	Het	2.129638E-5	Stop Gain

KCNJ15	NM_170736.3	p.Leu114Hisfs Ter43	c.341_342del	Het	N/A	Frameshift
KCNK18	NM_181840.1	p.Phe139Trpfs Ter25	c.414_415del	Het	4.3505413E-4	Frameshift
KIAA0556	NM_015202.5	p.Asp791Glufs Ter206	c.2373del	Het	1.5932026E-4	Frameshift
KIAA0556	NM_015202.5	p.Trp577Ter	c.1730G>A	Het	2.6908875E-4	Stop Gain
KIAA0556	NM_015202.5	p.Trp577Ter	c.1730G>A	Het	2.6908875E-4	Stop Gain
LOXHD1	NM_001384474.1	p.Arg1494Ter	c.4480C>T	Het	6.516206E-4	Stop Gain
LPA	NM_005577.4	p.Pro1542Lysf sTer61	c.4624_4625del	Het	N/A	Frameshift
MCM9	NM_017696.3	p.Arg574Ter	c.1720C>T	Het	6.648848E-6	Stop Gain
MSTO1	NM_018116.4	p.Gly420Valfs Ter2	c.1259del	Het	0.0	Frameshift
MUC3A	NM_005960.2	p.His413fs	c.1236_1237insTC	HomAlt	0.0017497803	Frameshift
MYH7	NM_000257.4	p.Glu1934Argf sTer8	c.5800del	Het	N/A	Frameshift
MYO1F	NM_012335.4	p.Arg982Profs Ter80	c.2943_2944dup	Het	N/A	Frameshift
MYORG	NM_020702.5	p.Trp163Ter	c.489_490del	Het	N/A	Frameshift
NFE2L3	NM_004289.7	p.Asp601Thrfs Ter2	c.1801del	Het	N/A	Frameshift
NLRC4	NM_001199138.2	p.Arg310Ter	c.928C>T	Het	1.6984668E-4	Stop Gain
OR2T10	NM_001004693.2	p.Lys227Argfs Ter2	c.680del	Het	N/A	Frameshift

OTOG	NM_001292063.2	p.Gln822Ter	c.2464C>T	Het	4.2709042E-4	Stop Gain
OXR1	NM_001198534.1	p.Trp5Ter	c.15G>A	Het	0.004232118	Stop Gain
PAH	NM_000277.3	p.Phe55LeufsTer6	c.165del	Het	1.1963892E-5	Frameshift
PALB2	NM_024675.4	p.Ser380ArgfsTer43	c.1140_1143del	Het	7.96115E-6	Frameshift
PCDHA13	NM_018904.3	p.Gly163Ter	c.487G>T	Het	N/A	Stop Gain
PCDHA4	NM_018907.4	p.Val587TrpfsTer65	c.1759del	Het	7.971817E-6	Frameshift
PEX1	NM_000466.3	p.Ile700TyrfsTer42	c.2097dup	Het	4.853175E-4	Frameshift
PREX1	NM_020820.4	p.Pro1052GlnfsTer39	c.3153del	Het	N/A	Frameshift
PRUNE2	NM_015225.3	p.Arg14Ter	c.40C>T	Het	4.1212315E-5	Stop Gain
PTPRQ	NM_001145026.2	p.Arg2007*	c.6019C>T	Het	3.197817E-5	Stop Gain
RANBP17	NM_022897.5	p.Ser288ArgfsTer19	c.862del	Het	N/A	Frameshift
RNASEL	NM_021133.4	p.Glu265Ter	c.793G>T	Het	0.0035566823	Stop Gain
RYR1	NM_000540.3	p.Asp2389GlyfsTer16	c.7166_7176del	Het	N/A	Frameshift
SACS	NM_014363.6	p.Tyr4392Ter	c.13175dup	Het	N/A	Frameshift
SCFD2	NM_152540.4	p.Asn305MetfsTer3	c.914del	Het	1.0233499E-4	Frameshift
SLC12A6	NM_001365088.1	p.Arg1026Ter	c.3076C>T	Het	3.9770603E-6	Stop Gain

SQSTM1	NM_003900.5	p.Glu352Ter	c.1054G>T	Het	N/A	Stop Gain
TMEM134	NM_025124.4	p.Arg84Ter	c.250C>T	Het	0.0045104236	Stop Gain
TMEM134	NM_025124.4	p.Arg84Ter	c.250C>T	Het	0.0045104236	Stop Gain
TMEM134	NM_025124.4	p.Arg84Ter	c.250C>T	Het	0.0045104236	Stop Gain
TPP1	NM_000391.4	p.Arg208*	c.622C>T	Het	2.5112412E-4	Stop Gain
TRANK1	NM_001329998.2	p.Arg1568Ter	c.4702C>T	Het	8.215174E-6	Stop Gain
TTC7B	NM_001010854.2	p.Glu219Ter	c.655G>T	Het	N/A	Stop Gain
WAC	NM_016628.5	p.Thr460Hisfs Ter20	c.1378_1382del	Het	N/A	Frameshift
ZNF804A	NM_194250.2	p.Gly350Valfs Ter7	c.1049del	Het	4.0335194E-6	Frameshift

6.Analysis of biological pathways through KEGG and Reactome

6.1GO - anthology

6.1.1 Cellular component

Tab.8 The functional classification of genes with non-synonymous SNVs in the gene ontological aspect cellular component

Cellular component	Overlap	P-value	Adjusted P-value
collagen-containing extracellular matrix (GO:0062023)	104/380	1.66757E-10	6.77032E-08
actin cytoskeleton (GO:0015629)	82/316	1.75047E-07	3.55345E-05
cytoskeleton (GO:0005856)	131/600	2.47773E-06	0.000335319
spectrin-associated cytoskeleton (GO:0014731)	7/8	1.11743E-05	0.001134195
sodium channel complex (GO:0034706)	12/25	8.97735E-05	0.006549403

basement membrane (GO:0005604)	19/52	9.67892E-05	0.006549403
endoplasmic reticulum lumen (GO:0005788)	66/285	0.000127664	0.006835647
brush border membrane (GO:0031526)	15/37	0.000134693	0.006835647
integral component of plasma membrane (GO:0005887)	262/1454	0.000364587	0.016446917
intermediate filament (GO:0005882)	17/50	0.000591676	0.024022064
polymeric cytoskeletal fiber (GO:0099513)	57/256	0.000983309	0.03629304
sarcoplasmic reticulum (GO:0016529)	15/45	0.001536594	0.047536155
cell projection membrane (GO:0031253)	25/92	0.001606906	0.047536155
voltage-gated sodium channel complex (GO:0001518)	8/17	0.001639178	0.047536155
intracellular organelle lumen (GO:0070013)	157/848	0.001783499	0.048273383

6.1.2 Molecular function

Table 9. The functional classification of the differentially expressed genes in the gene ontological aspect molecular function

Molecular function	Overlap	P-value	Adjusted P-value
ion channel activity (GO:0005216)	31/84	5.41025E-07	0.000550223
actin binding (GO:0003779)	49/177	7.62329E-06	0.003876441
inorganic cation transmembrane transporter activity (GO:0022890)	20/51	1.93117E-05	0.005429547
cation channel activity (GO:0005261)	31/98	2.13552E-05	0.005429547
microtubule motor activity (GO:0003777)	21/56	2.68645E-05	0.00546423
motor activity (GO:0003774)	23/66	4.43721E-05	0.007521077
mechanosensitive ion channel activity (GO:0008381)	8/12	6.71158E-05	0.009063436
endonuclease activity (GO:0004519)	19/51	7.12955E-05	0.009063436
gated channel activity (GO:0022836)	8/14	0.000307615	0.034760482

6.1.3 Biological processes

Table 10 The functional classification of the differentially expressed genes in the gene ontological aspect biological process

Biological processes	Overlap	P-value	Adjusted P-value
extracellular matrix organization (GO:0030198)	88/300	9.22452E-11	4.64916E-07
extracellular structure organization (GO:0043062)	65/216	8.4592E-09	1.73763E-05
external encapsulating structure organization (GO:0045229)	65/217	1.0343E-08	1.73763E-05
collagen fibril organization (GO:0030199)	35/89	1.55872E-08	1.96399E-05
supramolecular fiber organization (GO:0097435)	87/351	6.42355E-07	0.000647494
cilium movement (GO:0003341)	22/52	1.67268E-06	0.001405052
actin filament-based transport (GO:0099515)	12/21	9.08249E-06	0.006539395
steroid metabolic process (GO:0008202)	31/104	7.64908E-05	0.048189221
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)	21/60	8.74631E-05	0.048979324

6.2 KEGG and Reactome analysis of biological pathways enriched in genes with non-synonymous mutations

Table.11 Statistically significant KEGG pathways

Pathway	Overlap	P-value
ECM-receptor interaction	32/88	5.25653E-07
ABC transporters	19/45	8.79031E-06
Protein digestion and absorption	30/103	0.000157029
Bile secretion	26/90	0.000480881
Fanconi anemia pathway	17/54	0.001572584
Arginine and proline metabolism	16/50	0.001765931
Cortisol synthesis and secretion	19/65	0.002272363
Human papillomavirus infection	68/331	0.003114177

Focal adhesion	44/201	0.004824498
Taste transduction	22/86	0.006517323
Metabolism of xenobiotics by cytochrome P450	20/76	0.006621791
Glyoxylate and dicarboxylate metabolism	10/30	0.009081789
beta-Alanine metabolism	10/30	0.009081789
Axon guidance	39/182	0.010863441
Staphylococcus aureus infection	23/95	0.010992334

Table 12 Statistically significant Reactome pathways

Pathway	**Overlap**	**P-value**
Extracellular Matrix Organization R-HSA-1474244	85/291	2.42638E-10
Collagen Formation R-HSA-1474290	35/90	2.1877E-08
Interaction Between L1 And Ankyrins R-HSA-445095	17/29	7.07198E-08
Assembly Of Collagen Fibrils And Other Multimeric Structures R-HSA-2022090	24/57	6.32902E-07
Collagen Biosynthesis And Modifying Enzymes R-HSA-1650814	26/67	1.44644E-06
Sensory Perception R-HSA-9709957	130/616	1.589E-05
SARS-CoV-2 Activates/Modulates Innate And Adaptive Immune Responses R-HSA-9705671	36/121	2.20304E-05
Collagen Chain Trimerization R-HSA-8948216	18/44	2.54055E-05
Diseases Associated With O-glycosylation Of Proteins R-HSA-3906995	24/69	3.16305E-05
Anchoring Fibril Formation R-HSA-2214320	6/7	6.60214E-05
HDR Thru Single Strand Annealing (SSA) R-HSA-5685938	15/37	0.000134693
SLC Transporter Disorders R-HSA-5619102	29/99	0.000179803
Diseases Of DNA Repair R-HSA-9675135	18/51	0.000243228
Ion Channel Transport R-HSA-983712	44/175	0.000248966
Sensory Perception Of Taste R-HSA-9717189	17/47	0.000256475
Impaired BRCA2 Binding To PALB2 R-HSA-9709603	11/24	0.000295121
Synthesis Of PIPs At Late Endosome Membrane R-HSA-1660517	7/11	0.000304248

Antigen Presentation: Folding, Assembly, Peptide Loading Of Class I MHC R-HSA-983170	12/28	0.000340011
Defective TPR May Confer Susceptibility Towards Thyroid Papillary Carcinoma (TPC) R-HSA-5619107	13/32	0.000365675
Sensory Processing Of Sound By Outer Hair Cells Of Cochlea R-HSA-9662361	18/53	0.000418858
Defective HDR Thru Homologous Recombination (HRR) Due To BRCA1 Loss-Of-Function R-HSA-9701192	11/25	0.000456467
Degradation Of Extracellular Matrix R-HSA-1474228	30/109	0.000461581
CDC42 GTPase Cycle R-HSA-9013148	38/149	0.000471457
Type I Hemidesmosome Assembly R-HSA-446107	6/9	0.000604162
Diseases Of Metabolism R-HSA-5668914	56/247	0.000688175
O-glycosylation Of TSR Domain-Containing Proteins R-HSA-5173214	14/38	0.000713938
Cytochrome P450 - Arranged By Substrate Type R-HSA-211897	20/65	0.000874872

KEGG and Reactome analysis of biological pathways enriched in genes with non-synonymous mutations

Table.13 Statistically significant KEGG pathways

Patwat	Overlap	P-value
ECM-receptor interaction	32/88	5.25653E-07
ABC transporters	19/45	8.79031E-06
Protein digestion and absorption	30/103	0.000157029
Bile secretion	26/90	0.000480881
Fanconi anemia pathway	17/54	0.001572584
Arginine and proline metabolism	16/50	0.001765931
Cortisol synthesis and secretion	19/65	0.002272363
Human papillomavirus infection	68/331	0.003114177
Focal adhesion	44/201	0.004824498
Taste transduction	22/86	0.006517323
Metabolism of xenobiotics by cytochrome P450	20/76	0.006621791
Glyoxylate and dicarboxylate metabolism	10/30	0.009081789

beta-Alanine metabolism	10/30	0.009081789
Axon guidance	39/182	0.010863441
Staphylococcus aureus infection	23/95	0.010992334

Table 14 Statistically significant Reactome pathways

Път	**Overlap**	**P-value**
Extracellular Matrix Organization R-HSA-1474244	85/291	2.42638E-10
Collagen Formation R-HSA-1474290	35/90	2.1877E-08
Interaction Between L1 And Ankyrins R-HSA-445095	17/29	7.07198E-08
Assembly Of Collagen Fibrils And Other Multimeric Structures R-HSA-2022090	24/57	6.32902E-07
Collagen Biosynthesis And Modifying Enzymes R-HSA-1650814	26/67	1.44644E-06
Sensory Perception R-HSA-9709957	130/616	1.589E-05
SARS-CoV-2 Activates/Modulates Innate And Adaptive Immune Responses R-HSA-9705671	36/121	2.20304E-05
Collagen Chain Trimerization R-HSA-8948216	18/44	2.54055E-05
Diseases Associated With O-glycosylation Of Proteins R-HSA-3906995	24/69	3.16305E-05
Anchoring Fibril Formation R-HSA-2214320	6/7	6.60214E-05
HDR Thru Single Strand Annealing (SSA) R-HSA-5685938	15/37	0.000134693
SLC Transporter Disorders R-HSA-5619102	29/99	0.000179803
Diseases Of DNA Repair R-HSA-9675135	18/51	0.000243228
Ion Channel Transport R-HSA-983712	44/175	0.000248966
Sensory Perception Of Taste R-HSA-9717189	17/47	0.000256475
Impaired BRCA2 Binding To PALB2 R-HSA-9709603	11/24	0.000295121
Synthesis Of PIPs At Late Endosome Membrane R-HSA-1660517	7/11	0.000304248
Antigen Presentation: Folding, Assembly, Peptide Loading Of Class I MHC R-HSA-983170	12/28	0.000340011
Defective TPR May Confer Susceptibility Towards Thyroid Papillary Carcinoma (TPC) R-HSA-5619107	13/32	0.000365675
Sensory Processing Of Sound By Outer Hair Cells Of Cochlea R-HSA-9662361	18/53	0.000418858

Defective HDR Thru Homologous Recombination (HRR) Due To BRCA1 Loss-Of-Function R-HSA-9701192	11/25	0.000456467
Degradation Of Extracellular Matrix R-HSA-1474228	30/109	0.000461581
CDC42 GTPase Cycle R-HSA-9013148	38/149	0.000471457
Type I Hemidesmosome Assembly R-HSA-446107	6/9	0.000604162
Diseases Of Metabolism R-HSA-5668914	56/247	0.000688175
O-glycosylation Of TSR Domain-Containing Proteins R-HSA-14	14/38	0.000713938
Cytochrome P450 - Arranged By Substrate Type R-HSA-211897	20/65	0.000874872

IV. Discussion

1.Discussion of Transcriptional Analysis Data

The initial phase of this scientific study entails a comprehensive transcriptomic analysis of all protein-coding genes, revealing differences in the expression of these genes in children with Specific Language Impairment (SLI) and healthy controls. The genes with altered expression are involved in processes related to cellular signaling, the immune system, and various metabolic pathways. Our results indicate that synaptic plasticity, disruptions in axonal orientation, neuronal experience, differentiation, and inflammation play a role in the pathophysiology of SLI.

1.1. Genes with Altered Expression from Transcriptomic Analysis

1.1.1. GINS2 (GINS Complex Subunit 2)

GINS2 is a gene encoding a protein located on chromosome 16q24.1. The GINS complex, identified in 2000, is known to interact with Cdc45 and Mcm2-7 in a helicase complex for DNA leading strand synthesis in human cells. The GINS complex is crucial for normal human physiology in cell proliferation, and it has been implicated in various pathologies, primarily cancer (17) source.

Several authors report speech delay in patients with 16q24.1 duplications, involving genes C16orf74, COX4I1, EMC8, GINS2, GSE1, and IRF8 (90). Changes in the gene dosage of these genes may be involved in speech development. In the SLI group, we found average values of 15.4 copies/20mcl compared to 8.8 copies/20mcl in the control group. The difference is statistically significant with a p-value of 0.03, demonstrating that the GINS2 gene is overexpressed in the children with language impairment.

A recent study by Chunhua Liu et al. establishes a connection between GINS2 expression and neuroinflammation. Their results suggest a protective role for GINS2 in microglia when its expression is induced (91). We can hypothesize that the higher expression of GINS2 in children with SLI is due to induced pro-inflammatory signaling.

1.1.2. RAP1GAP (RAP1 GTPase Activating Protein)

RAP1GAP encodes a GTPase-activating protein (GAP) that suppresses the activity of RAS-associated Rap1 proteins. Rap1 acts as a molecular switch moving between an inactive GDP-bound form and an active GTP-bound form. Overall, the RAS protein superfamily, including Rap1, plays a key role in receptor-linked signaling pathways that control cell growth and differentiation. Rap1 contributes to various processes such as cell proliferation, differentiation, adhesion, and embryogenesis by regulating the ERK/MAPK signaling pathway.

Within the nervous system, the regulation of Rap1 through Rap1GAP contributes to axonal growth, dendritic development, and dendritic plasticity (92). Rap1 influences the morphology of dendritic processes, and the Rap1GAP associated with dendrites (SPAR) is localized in the postsynaptic membrane linked with PSD-95, and SPAR expression leads to an increase in the size of dendritic processes .

RAP1GAP is also associated with tuberous sclerosis. Among its associated pathways are nervous system development and angiotensin activation of ERK source.

Phosphorylation-mediated regulation of Rap1GAP likely contributes to the control of Rap1 in various parts of the brain. In neurons, Rap1 regulates structural and functional processes underlying synaptic plasticity, learning, and memory.

1.1.3. ANKRD22 (Ankyrin Repeat Domain 22)

ANKRD22 is a gene encoding a 22-kilodalton protein with three tandem ankyrin repeat motifs. Proteins containing ankyrin repeats are associated with the cell cycle and signaling pathways related to intercellular signaling (source).

Interessestingly, the association of ANKRD22 with IL1B is noteworthy. IL1B (Interleukin 1 beta) is an inflammatory cytokine expressed early in the immune response. In tissues, IL-1B propagates inflammation by activating immune cells and local vascular endothelium. Systemically, IL-1B stimulates the production of IL-6. IL-1B can cross the blood-brain barrier and stimulate its own expression in the hypothalamus (94). Dysregulation of IL-1B may have various neurological consequences related to the RAS (Renin-Angiotensin System) and other neurodevelopmental disorders. The cytokine and its receptors are found in the nervous system during critical periods of development. IL-1B induces the proliferation of neuronal progenitor cells in some regions of the central nervous system and inhibits their proliferation in other areas (95). This may contribute to the specific regional overgrowth observed in the brains of patients with RAS and other neurodevelopmental disorders. The formation of some excitatory synapses is also partially mediated by the IL-1 receptor and receptor-associated proteins (96).

1.1.4. IFI6 (Interferon Alpha Inducible Protein 6) IFI6 was first identified as one of many genes induced by interferon. The encoded protein may play a critical role in regulating apoptosis with pro-apoptotic effects [source]. A wide-scale association study conducted by Benchet P. and colleagues in 2016 identified several candidate variants for communication disorders. They hypothesized that genomic factors associated with variations in speech production, phonological representation, and language skills may be linked to neural pathways common to speech, language, reading, and spelling. One of the candidate variants is IFI6, which is associated with autism. Overexpression of IFI6 in the brain is also observed in chronic neurodegenerative processes .

1.1.5. OTOF (Otoferlin) Mutations in this gene cause non-syndromic recessive deafness. Pathways associated with OTOF include sensory processing of sound and olfactory signaling [source]. Otoferlin is necessary for calcium-dependent exocytosis/neurosecretion at some synapses (97). The observed result of increased expression in our study may be considered a potential mechanism contributing to the imbalance of synapses involved in the pathogenesis of RAS and other neurodevelopmental disorders such as SLI.

1.1.6. RND2 - Rho Family GTPase 2 This gene encodes a member of the Rho GTPase family, whose members play a key role in regulating the organization of the actin cytoskeleton in response to extracellular growth factors. This particular member of the Rho subfamily, Rnd2, is involved in regulating neuronal morphology. Among Rnd subfamily members, only Rnd2 and Rnd3 show strong expression in the developing cerebral cortex. Rnd2 is found in cells of the preplate in the early stages of cortical development. The distribution of Rnd2 and Rnd3 transcripts in exceptionally cortical domains suggests that they may have individual and essential functions in different phases of cortical development and neuronal migration.

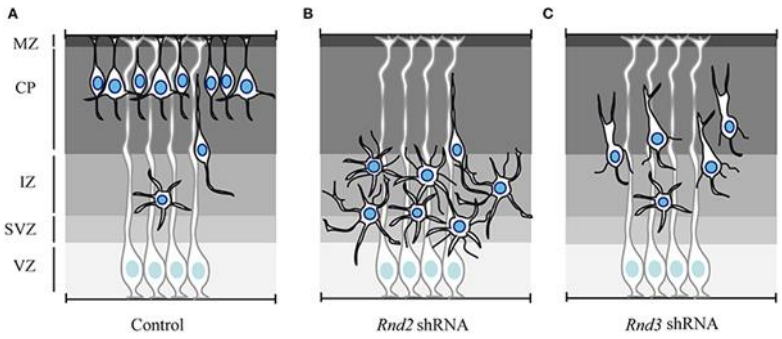


Fig. 9 Effect of loss of function of RND2 and RND3 on the migration of cortical neurons.

(A) Schematic representation of cortical radial migration in a control state. Newborn projection neurons undergo sequential steps of radial migration characterized by different morphologies. At the mid-end of corticogenesis, most neurons have reached the cortical plate (CP). Only a few cells are still migrating, showing multipolar morphology in the intermediate zone (IZ) and bipolar form in the CP.

(B) shRNA-mediated loss of function and reduced expression of Rnd2 in the embryonic cortex cause the accumulation of multipolar cells in the IZ, which exhibit more and longer neuronal processes.

C) Knockdown of Rnd3 in the embryonic cortex hinders the migration phase in the CP. Neurons with suppressed Rnd3 exhibit abnormal morphologies characterized by excessively enlarged and branched leading processes and thin processes extending from the cell body. MZ - Marginal Zone; CP - Cortical Plate; IZ - Intermediate Zone; SVZ - Subventricular Zone; VZ - Ventricular Zone.

In most of our Specific Language Impairment (SLI) patients, the gene shows reduced expression, suggesting discrete changes in cortical development or impaired neuronal migration, leading to prolonged procedural processing.

2.Discussion on gene ontological analysis of variants with altered expression and Reactome analysis of biological pathways

Gene Ontology (GO) is an internationally standardized classification for gene function, providing a controlled and dynamically updated database with strictly defined gene and product properties. GO encompasses three aspects: cellular component, molecular function, and biological process. The basic unit of GO is the GO term (element). Each GO term belongs to an ontology type. The analysis provides all GO terms containing significant genes with variants compared to the genome and filters those corresponding to biological functions. This analysis first maps genes to GO terms in the database, calculates the number of genes in each GO term, and, through geometric analysis, identifies GO terms with significant content from our list compared to the genome.

Results from our genetic analyses indicate that genes with altered expression in children with SLI participate in pathways primarily related to the immune system and cell signaling.

2.1. Cytokine Signaling in the Immune System

The immune system and the nervous system closely interact, explaining the common occurrence of immune dysfunction in various neurological and neurodevelopmental disorders. Cytokines, known as interleukins (IL) and interferons (IFN), serve as key signaling molecules for intercellular and intersystem communication. Cytokines share structural similarities and signaling pathways with neurotrophins and neurologically significant growth factors, acting as a common language between the immune and nervous systems.

Cytokines influence both the development and function of the nervous system, with their significance varying based on the timing, duration, and intensity of neuroimmune interactions. Cytokines affect the developing brain differently from the adult brain. Neurotrophic cytokines, using various combinations of receptor subunits and signaling pathways, play crucial roles in normal aspects of neurological development, including the differentiation of progenitor cells, cellular localization/migration in the nervous system, and the formation of synaptic networks.

The development of the nervous system is characterized by phases of cellular proliferation and differentiation, neuronal migration, axonal growth, and the formation and stabilization of synapses. Each of these processes results from the modulation of genetic programs through extracellular signals.

2.2. Interferon Signaling

Several epidemiological and animal studies support the role of the activated antiviral response in the etiology of neurodevelopmental disorders. The molecular mechanisms underlying this association remain unclear, but it is believed that inflammatory cytokines play a central role.

A study by Katherine Warre-Cornish et al. investigates the impact of IFN-γ exposure during the differentiation of human neurons. RNA sequencing analysis reveals morphological and transcriptomic changes associated with neurodevelopmental disorders. The study suggests that IFN-γ stimulates neuronal growth through the hyperexpression of genes from the major histocompatibility complex class I (MHCI), a effect mediated by promyelocytic leukemia protein (PML). According to this study, IFN-γ signaling inversely proportional to the expression of genes associated with schizophrenia and autism, among other neurodevelopmental disorders. This study represents the first exploration of the mechanisms by which transient immune activation induces lasting changes in neuronal phenotypes.

Figure 10: Interferon Signaling

2.3. Biosynthesis of Chondroitin Sulfate

Proteoglycans are of two main types: chondroitin sulfate (CSPG) and heparin sulfate (HSPG). During development, CSPG shapes cellular migration, axon growth pathways, and their terminal endings. In mature age, CSPG binds to certain classes of neurons, controlling their plasticity. Following damage to the nervous system, CSPG becomes the primary inhibitory component of axon growth in glial scar tissue, blocking successful regeneration. CSPG plays various roles in the nervous system, including

binding to molecules and blocking their actions, presenting molecules to different cells and axons, localizing active molecules at specific sites, and presenting growth factors to their receptors. In essence, the biosynthesis of chondroitin sulfate is critical in constructing a functionally healthy brain.

3.Discussion of Results

Sequencing the 40 samples from children with Specific Language Impairment (SLI) identified a set of different genetic variants (Table 12). We found a total number of Single Nucleotide Variants (SNV) ranging from 56,731 to 47,492 across different samples. SNVs are variations in a single nucleotide in the DNA sequence that can impact gene expression and protein function. Additionally, we identified in/dels or small deletions/insertions ranging from 3,374 to 4,987 among different patients. Indels can cause frame shifts in the coding sequence, altering the amino acid sequence of the protein and potentially disrupting its function.

The number of non-synonymous variants (missense mutations) that change one amino acid in the protein sequence and may affect protein function varied from 8,135 to 9,524 in some children. Furthermore, we observed nonsense or stop-codon variants causing premature termination of protein synthesis between positions 50 and 73. We identified frame-shifting deletions and insertions ranging from 86 to 113 and 95 to 132, respectively. These mutations lead to frame shifts in the reading frame of the coding protein sequence, altering the amino acid sequence downstream of the mutation.

We also identified a set of silent or synonymous mutations with a total count ranging from 9,665 to 10,906, which do not change the amino acid sequence of the protein.

3.1 Non-synonymous Mutations in Known "Candidate Genes"

According to current guidelines for assessing causal relationships in whole exome datasets, genes associated with SLI should be evaluated before exploring potential new candidates. Therefore, before conducting bioinformatic analyses on our exome data, we conducted a literature search to identify a set of candidate genes with accumulated evidence of involvement

in the pathogenesis of SLI in previous studies. We focused on 30 candidate genes (Table 12), including CMIP, ATP2C2, CNTNAP2, and NFXL1, previously linked to SLI; FOXP2, involved in monogenic forms of childhood apraxia of speech, and its ortholog FOXP1, also associated with neurological development disorders; DYX1C1, KIAA0319, DCDC2, and ROBO1, candidate genes in developmental dyslexia; SRPX2 and GRIN2A implicated in speech apraxia and epileptic aphasia, along with the closely related candidate GRIN2B3; and ERC1, SETBP1, CNTNAP5, DOCK4, SEMA6D, and AUTS2, each with proven rare deletions or translocations leading to speech, language, and/or reading disorders.

Known candidate genes					
Gene	Variant	Population frequency	Genotype	Type of mutation	Number of children
APLP2	c.1459C>G	<0,01%	heterozygous	Missense	1
APLP2	c.211del	new variant	heterozygous	Missense	1
ATP2C2	c.1811T>A	4,91%	heterozygous	Missense	2
ATP2C2	c.2482-2A>C	0,87%	heterozygous	Splice Acceptor	1
AUTS2	c.2345G>T	new variant	heterozygous	Missense	1
AUTS2	c.2630G>A	new variant	heterozygous	Missense	2
AUTS2	c.385C>G	0,15%	heterozygous	Missense	1
CMIP	c.1180T>G	1,00%	heterozygous	Missense	1
CMIP	c.1389-4C>G	0,21%	heterozygous	splaice site	1
CNTNAP5	c.1358C>T	3,6	heterozygous	Missense	5
CNTNAP5	c.1971C>G	0,14%	heterozygous	Missense	1

CNTNAP5	c.2153T>C	0,78%	heterozygous	Missense	1
DCDC2	c.454C>G	3,75%	heterozygous	Missense	4
DNAAF4	c.271G>A	4,70%	heterozygous	Missense	4
DOCK4	c.1280C>T	0,10%	heterozygous	Missense	1
DOCK4	c.5767G>A	4,89%	heterozygous	Missense	5
DOCK4	c.5804C>T	2,64%	heterozygous	Missense	1
DOCK4	c.976A>G	1,94%	heterozygous	Missense	1
DOCK4	c.5224C>G	0,95%	heterozygous	Missense	1
DOCK7	c.1872-9del	3,18%	heterozygous	Splice Region	1
DOCK7	c.527T>C	<0,01%	heterozygous	Missense	1
DOCK7	c.541A>G	2,08%	heterozygous	Missense	1
ERC1	c.148A>G	2,65%	heterozygous	Missense	3
ERC1	c.1879C>T	<0,01%	heterozygous	Missense	1
FOXP1	c.1333G>T	<0,01%	heterozygous	Missense	1
KIAA0319	c.1699G>A	3,39%	heterozygous	Missense	1
KIAA0319	c.2293-5C>T		heterozygous	Splice Region	1
KIAA0319	c.2317A>G		heterozygous	Missense	1
KIAA0319	c.2321T>C		heterozygous	Missense	1
KIAA0319	c.541G>A	0,02%	heterozygous	Missense	1

KMT2D	c.10192A>G	0,88%	heterozygous	Missense	1
KMT2D	c.12028T>C	2,86%	heterozygous	Missense	3
KMT2D	c.2438C>T	4,27%	heterozygous	Missense	2
KMT2D	c.248G>A	2,65%	heterozygous	Missense	3
KMT2D	c.4021-11_4021-10del	3,89%	heterozygous	Splice Region	2
KMT2D	c.6629C>T	0,27%	heterozygous	Missense	1
KMT2D	c.7670C>T	0,83%	heterozygous	Missense	1
OXR1	c.15G>A	0,42%	heterozygous	Stop Gain	1
ROBO1	c.3272G>A	3,66%	heterozygous	Missense	1
ROBO1	c.3403A>G	0,03%	heterozygous	Missense	1
ROBO1	c.3757G>A	<0,01%	heterozygous	Missense	2
SCN9A	c.1828C>A	3,05%	heterozygous	Missense	2
SCN9A	c.2137T>A	<0,01%	heterozygous	Missense	1
SCN9A	c.2248A>G	0,25%	heterozygous	Missense	1
SCN9A	c.2437A>G	0,24%	heterozygous	Missense	1
SCN9A	c.2758G>A	new variant	heterozygous	Missense	1
SCN9A	c.4645T>C	0,20%	heterozygous	Missense	1
SCN9A	c.5711G>A	0,03%	heterozygous	Missense	1
SEMA6D	c.140G>A	<0,01%	heterozygous	Missense	1
SEMA6D	c.1647-5T>A	3,24%	heterozygous	Missense	4
SETBP1	c.4398G>T	1,09%	heterozygous	Missense	1
SRPX2	c.287G>A	new variant	heterozygous	Missense	1

ZNF277	c.931G>C	<0,01%	heterozygous	Missense	1

Table 20 presents the number of different non-synonymous substitutions in 32 literature-known "candidate genes," categorized by their frequency in the population.

3.1.1.Similar to most neurodevelopmental disorders, SLI and dyslexia have a complex genetic etiology involving multiple interactions between various genetic factors and environmental factors. Dyslexia and specific language impairment show significant comorbidity and clinical overlap, suggesting shared genetic etiology. Several genetic risk variants for SLI and dyslexia have been identified, allowing for a direct assessment of possible shared genetic influences between these disorders. A study by D. F. Newbury et al. in 2010 investigated ROBO1, DCDC2, KIAA0319, DYX1C1, CNTNAP2, ATP2C2, and CMIP in the etiology of SLI and dyslexia.

DYX1C1, KIAA0319, DCDC2, and ROBO1 are key common candidate genes associated with dyslexia. The proteins encoded by these genes, although functionally distinct, are connected to pathways involved in neuronal migration and axon growth (Galaburda et al., 2006).

Similar to most neurodevelopmental disorders, SLI and dyslexia have a complex genetic etiology involving intricate interactions between various genetic factors and environmental influences. Dyslexia and specific language impairment (SLI) exhibit significant comorbidity and clinical overlap, suggesting a shared genetic etiology. Several genetic risk variants for SLI and dyslexia have been identified, allowing for a direct assessment of potential shared genetic influences between these disorders. In a study by D. F. Newbury et al. in 2010, *ROBO1, DCDC2, KIAA0319, DYX1C1, CNTNAP2, ATP2C2, and CMIP* were investigated in the etiology of SLI and dyslexia

DYX1C1, KIAA0319, DCDC2, and ROBO1 are the main shared candidate genes associated with dyslexia. The proteins encoded by these genes, although functionally diverse, are functionally linked to pathways involved in neuronal migration and axon growth (Galaburda et al., 2006).

DCDC2 (Doublecortin Domain-Containing Protein, OMIM_id: 605755) is a gene encoding a ciliary protein highly expressed in the entorhinal cortex, lower temporal cortex, medial

hypothalamus, amygdala, and hippocampus. DCDC2 enhances microtubule polymerization by binding to tubulin. RNA interference studies show that reduced expression of this gene alters neuronal migration (102). Several studies have established a strong association between the *DCDC2* gene and susceptibility to dyslexia, linked to abnormal neuronal migration and maturation (103). In a large Chinese study in 2017, polymorphisms in the DCDC2 gene were found to be associated with developmental dyslexia in Uighur children. In 36 of our patients, a non-synonymous variant DCDC2:c.661A>T was identified, classified as benign with a frequency in GnomAD of about 65%. This variant has been associated with severe spelling deficits (105). Another non-synonymous variant, DCDC2:c.454C>G, was identified in 4 children with a lower population frequency of about 3%, classified as benign.

However, it is important to note that the clinical significance of a variant may depend on the context. These variants may not cause disease on their own but could potentially contribute to certain symptoms or conditions when combined with other genetic factors or environmental factors. In the context of specific language impairment, variants in the DCDC2 gene may contribute to its manifestation.

DNAAF4 **-** This gene encodes a protein that is a dynein axonemal assembly factor, necessary for ciliary motility. It participates in neuronal migration during cerebral neocortex development and may regulate the stability and proteasomal degradation of estrogen receptors, playing a crucial role in neuronal differentiation, survival, and plasticity. Variants in this gene are associated with reading and spelling deficits .

In 4 of our patients, a variant *DNAAF4:c.271G>A* was identified with a population frequency of 4.7%. Marino et al. found that both *DNAAF4 and DCDC2* are associated with short-term and working memory processes, with *only DCDC2* being associated with word and nonsense word reading. Both genes influence individual variations in information extraction and storage, with *DCDC2* specifically linked to phonological task performance. This confirms the pleiotropic action of these two genes, which are also related to language skills and different language disorders (106). We believe that the numerous variants in DCDC2 and DNAAF4 found in our patients may be associated with specific language impairment.

ROBO1 (Roundabout Guidance Receptor 1, OMIM_id: 602430) - The gene is associated with language deficits affecting normal variations in language abilities. It is involved in axon guidance and migration of neuronal precursor cells. This receptor is activated by proteins from

the SLIT family, leading to a repulsive effect on glioma cells in the developing brain. It is the first gene associated with language deficits affecting normal variations in language abilities.

In 2 of our patients, a rare nonsynonymous mutation *ROBO1:c.3757G>A* was identified in a heterozygous state, with a frequency in GnomAD of -0.004061%, described in 11 individuals worldwide. Two more variants, *ROBO1:c.3272G>A and ROBO1:c.3403A>G,* were identified in one child each. The ROBO1 gene is associated with language acquisition, participating in neuronal migration, which is crucial for the development of bilateral symmetry and lateralization of neuronal function .

KIAA0319 encodes a plasma membrane protein that plays a role in adhesion, attachment, and migration of neurons in the developing brain, with a significant role in intra- and extracellular signaling. The encoded protein is involved in the development of the cerebral cortex by regulating neuronal migration and cell adhesion. Single nucleotide polymorphisms in this gene are associated with dyslexia. Variants in this gene interact with environmental factors, such as parental education levels, and can affect reading abilities. In 35 patients, a variant *KIAA0319:c.910A>T* was identified, classified as benign.

In conclusion, the identified genetic variants in *DYX1C1, KIAA0319, DCDC2, ROBO1, and DNAAF4,* along with their known associations with dyslexia and language disorders, suggest their potential contribution to specific language impairment in our patient cohort. Further functional studies and larger-scale genetic analyses are warranted to elucidate the precise mechanisms through which these genetic factors contribute to the complex etiology of specific language impairment.

In our sample, we identified five variants in the KIAA0319 gene - KIAA0319c.1699G>A; KIAA0319c.2293-5C>T; KIAA0319c.2317A>G; KIAA0319 c.2321T>C; KIAA0319 c.541G>A. The KIAA0319 gene seems to play a role in language acquisition, but further studies are needed to fully understand the impact of specific KIAA0319 variants on this condition .

3.1.2 Candidate Genes in ASD Associated with Autism CNTNAP5 (Contactin Associated Protein Family Member 5) belongs to the neurexin family and plays a role in the development and functioning of the peripheral and central nervous systems. It is involved in cellular adhesion, intercellular communication, the formation of myelinated

axons, and the maintenance of synaptic stability (108). CNTNAP5 and CNTNAP2 are reported as candidate genes in studies on dyslexia, autism, and ASD (56, 109).

In our study, we identified several variants in CNTNAP5 - CNTNAP5:c.1358C>T; CNTNAP5:c.1971C>G; CNTNAP5:c.2153T>C, with CNTNAP5:c.1358C>T identified in 5 children with a frequency in the population according to GnomAD-3.6%.

In 2010, Pagnamenta et al. studied a Dutch family with autism to demonstrate the role of deletions in CNTNAP5 in the etiology of Autism Spectrum Disorder (ASD). They identified a 227 kb deletion removing exons 4–11 of CNTNAP5. This led them to sequence all 24 exons and intron-exon boundaries of CNTNAP5 in 143 probands in families with ASD, identifying three additional rare nonsynonymous variants. The researchers concluded that rare variants in CNTNAP5 contribute to a predisposition to ASD.

In conclusion, although there is currently no direct evidence linking CNTNAP5 variants to language impairments, considering its role in neuronal function and its association with other neuropsychiatric conditions, it is possible to speculate that this gene plays a role in language impairments.

In 2017, Chen X. et al. reported 6 rare single nucleotide variants (SNVs) in ATP2C2, AUTS2, CNTNAP5, ROBO1, and SRPX2 in a pre-constructed set of candidate genes for ASD .

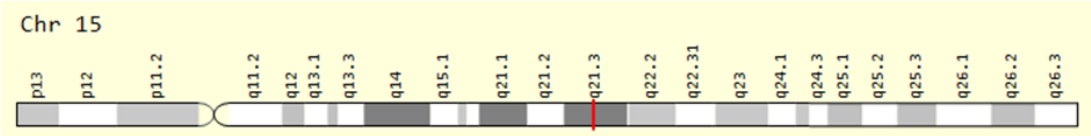


Fig. 23: Location of CNTNAP5 on Chromosome 2

AUTS2 - The AUTS2 gene, also known as the autism susceptibility 2 gene, plays a crucial role in the development and functioning of the nervous system. It is associated with neurological developmental disorders, including ASD and intellectual disabilities. These disorders often involve difficulties in language development and communication (103). Therefore, variants in the AUTS2 gene may contribute to delays in language development.

In our sample, we identified two new variants reported for the first time in AUTS2 - AUTS2: c.2345G>T and AUTS2: c.2630G>A identified in two children. The AUTS2: c.385C>G variant has a frequency in GnomAD of 0.15%.

It is important to emphasize that genetic variations can have a range of effects, and not all variants in the AUTS2 gene will necessarily lead to ASD or intellectual disability. In fact, mild changes in this gene may not result in autism but could potentially contribute to finer effects on neurological development, such as delayed language development.

The de novo variants with unclear clinical significance, AUTS2: c.2345G>T and AUTS2: c.2630G>A identified in two children with ASD, may be examples of such mild changes. These variants may subtly alter the function of the AUTS2 protein, affecting the development and function of neurons in a way that leads to delayed language development rather than autism. However, this is speculative, and the exact impact of these specific variants on language development is not known. Additional studies are needed to confirm these possibilities and understand the full spectrum of effects on neurological development associated with AUTS2 gene variants.

SRPX2 - The SRPX2 gene, also known as Sushi Repeat-containing Protein, X-linked, 2 gene, is associated with language development and neurological disorders. This gene is a target of the transcription factor FOXP2, which is known to play a key role in the development of speech and language (104).

Studies show that FOXP2 modulates synapse formation by regulating SRPX2 levels. Synapses are crucial for signal transmission between neurons, and their formation and function are essential for various cognitive processes, including language development (104). Reduced expression of SRPX2 impairs the development of ultrasonic vocalization in mice, suggesting that this gene may play a role in vocal communication. In humans, changes in the SRPX2 gene are associated with epilepsy and language-related disorders (105).

In our sample, we identified a new variant not present in GnomAD - SRPX2:c.287G>A in a homozygous genotype in one child with ASD.

Given the role of SRPX2 in neuronal function and its association with language-related disorders, it is possible to hypothesize that the SRPX2:c.287G>A variant in a homozygous genotype may potentially contribute to ASD.

APLP2 - The APLP2 gene encodes a protein known as amyloid beta precursor-like protein 2, which participates in the development and functioning of the nervous system. Variants in this gene are associated with various neurological conditions.

In our sample, we identified two APLP2 variants - APLP2:c.1459C>G with a frequency below 0.1% in GnomAD, and APLP2:c.211del – a new variant.

APLP2 plays a key role in neuronal function; therefore, we assume that these variants may be associated with delayed language development.

ZNF277 (Zinc Finger Protein 277; OMIM id 605465) is a gene expressed in several tissues, including the brain, especially in the neocortex and hippocampus, even in early fetal development. Gene ontology (GO) includes specific DNA binding to the RNA polymerase II core promoter. In most phenotype databases, the gene is associated with specific language disorders and autism.

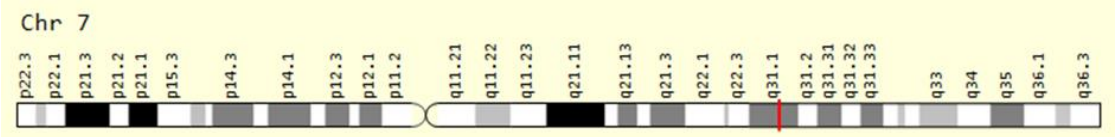


Fig. 24: Location of the gene on Chromosome 7"

In 2014, Pembley M. and colleagues described a homozygous exon microdeletion in ZNF277 in a girl with specific language impairment. This microdeletion was identified in a heterozygous form in eight families from the SEN cohort and in four families with cases of ASD. There was an increased allelic frequency of ZNF277 microdeletions in SEN probands (1.1%) compared to family members with ASD (0.3%) and unrelated controls (0.4%), suggesting that these microdeletions may be a risk factor for SEN. However, ZNF277 microdeletions exhibited incomplete segregation in the SEN phenotype and were identified in unaffected family members, with some cases not inherited by affected children (reverse discordance). We hypothesize that these CNVs may contribute to SLI susceptibility in a complex manner, acting as a risk factor with reduced penetrance. Pembrey and colleagues suggest that an epigenetic mechanism, termed Meiotic Nondisjunctional Methylation (3M), during maternal meiosis I, increases the chances of abnormal methylation due to chromosomal misalignment through improper pairing, similar to what might occur with transposon silencing.

In one of the children with , we identified the variant *ZNF277:c.931G>C* with a frequency in GnomAD below 0.01%, and in seven of the children, *ZNF277:c.1090G>A* was found with a frequency in the population according to GnomAD around 5%.

3.1.3 Genes Related to Brain Function These are genes associated with calcium homeostasis (*ATP2C2*), embryonic development (CMIP), regulation of RNA polymerase II transcription (NFXL1, ERC1), synaptogenesis (SRPX2, SETBP1, and SEMA6D). After analyzing these genes for significant mutations, we found several

important facts. NFXL1 - a candidate gene for specific language impairment. The protein encoded by this gene is presumed to be a transcription factor based on its similarity to the domain of NFX1, a repressor of HLA class II genes implicated in specific language impairment. However, there is limited literature on the function of NFXL1. High expression is detected in the cerebellar hemispheres. In our study, we identified the variant NFXL1:c.737C>T in 11 of our patients.

3.1.4 Genes Related to Brain Function These genes are associated with calcium homeostasis (ATP2C2), embryonic development (CMIP), transcriptional regulation by RNA polymerase II (NFLX1, ERC1), synaptogenesis (SPRX2, SETBP1, and SEMA6D). After analyzing for significant mutations in these genes, several important findings were established.

NFXL1 is a candidate gene for specific language impairment. The protein encoded by this gene is believed to be a transcription factor based on its similarities to the NFX1 domain, a repressor of HLA class II genes implicated in specific language impairment. However, there is limited literature on the function of NFXL1. In the study, the variant NFXL1:c.737C>T was identified in 11 of the patients.

3.2. STOP Codon (Nonsense) Variants Associated with SLI

In our dataset (Table 13), we identified and validated stop-codon variants that are rare (frequency in GnomAD < 5%) or newly described for the first time. Stop-codon variants lead to truncated proteins and have the potential for more severe consequences than most single amino acid substitutions. Within our sample, this approach allowed us to derive several variants that may be associated with the disorder. As recommended by MacArthur, we focused on rare and novel variants, using large, ethnically matched control data and employing multiple bioinformatic algorithms to predict the mutation's effect and assess potential pathogenicity.

Table 13 presents some of the stop-codon variants and their frequency in our sample.

3.2.1 ***PABPC1 Stop-Codon Variant PABPC1:c.1033G>T***

This gene encodes poly(A)-binding protein. It is part of a small gene family, including three protein-coding genes and several pseudogenes. The product of the gene binds the poly(A) tail of mRNA, including its own transcript, and regulates processes in mRNA metabolism, such as splicing and stability. Its function in translation initiation can be enhanced by PAIP1 or suppressed by PAIP2. The protein can also bind to cytoplasmic RNA sequences other than poly(A) in vivo. It participates in the regulation of nonsense-

mediated decay (NMD) of mRNAs containing premature stop codons, recognizing premature termination codons (PTCs). Competitive interaction between UPF1 and PABPC1 with ribosome-associated release factors is proposed for the recognition of PTCs and initiation of NMD. By binding to long poly(A) tails, it may protect them from uridylation by ZCCHC6/ZCCHC11, contributing to mRNA stability.

To our knowledge, two studies in the literature link PABPC1 to neurodevelopmental disorders and expressive speech delay. Meret Wegler and colleagues report four cases of dyslexia with two new missense variants grouped in PABPC1. The authors find that PABPC1 variants reduce the binding affinity to proteins involved in mRNA metabolism, and knockdown of PABPC1 in mice reduces the proliferation of neural progenitor cells, supporting our discovery of PABPC1 stop-codon mutations in children with SEN.

A study by Tianyun Wang and collaborators analyzes de novo variants from over 46,000 autism and developmental disorder trio cases. The analysis identifies this variant as one of the candidate genes due to the presence of multiple de novo mutations. The authors speculate that disruptions in PABPC1 function may be related to developmental disorders.

In our study, we identified the same stop-codon mutation (PABPC1 stop-codon mutation or *PABPC1:c.1033G>T* in 17 children (42.5%). The variant frequency in GnomAD is 6.8%. After conducting statistical analysis (Z score: Z value is 8.8786. The p-value is <0.00001. The result is significant at $p < 0.01$, and Fisher Exact Test: The p-value is <0.00001) for comparing frequencies, a statistically significant difference was found. We can assume that PABPC1 is a strong candidate gene for SEN, as well as for other neurodevelopmental disorders. We can speculate that the severity of the variants may be associated with the severity of the clinical manifestation ranging from speech developmental delay to severe delay.

3.2.2. *PRODH* is located on chromosome 22 (22q11.21) and encodes the POX enzyme, which converts proline to pyrroline-5-carboxylate (P5C) in mitochondria. In a homozygous state, this variant is associated with hyperprolinemia. The clinical phenotype of this condition is not clearly characterized. Patients with hyperprolinemia may experience seizures, intellectual deficit, language delay, autism spectrum disorder (ASD), schizophrenia, and/or bipolar disorder. The gene is also associated with"

Please note that due to the length of the text, the translation has been truncated. If you have specific sections you'd like further clarification on or if you have additional requests, feel free to let me know.

In 2014, Pembley M. et al. described a homozygous exon microdeletion in ZNF277 in a girl with specific language impairment. This microdeletion was identified in a heterozygous form in eight families from the SEN cohort and in four families with cases of ASD. There was an increased allelic frequency of ZNF277 microdeletions in probands with SEN (1.1%) compared to members of the RAS family (0.3%) and unrelated controls (0.4%), suggesting that these microdeletions may be a risk factor for SEN. However, ZNF277 microdeletions showed incomplete segregation in the SEN phenotype and were identified in unaffected family members. The authors hypothesized that an epigenetic mechanism called meiotic nondisjunctional methylation during maternal meiosis I could explain the observed lack of segregation for ZNF277 microdeletions. According to this hypothesis, the pairing of the chromosome carrying the microdeletion with a wild-type homolog during meiosis I increases the chance of abnormal methylation due to chromosomal misalignment, similar to what would occur with transposon suppression.

In one of the SEN children, a variant ZNF277:c.931G>C with a frequency below 0.01% in GnomAD was identified, and in seven children, ZNF277:c.1090G>A was found with a population frequency of around 5%.

***3.2.3. BAP1:c.1203T*>G, chr3-52438516 A>C, p.Tyr401Ter** was identified as a pathogenic variant leading to a premature stop codon in the BAP1 gene. BAP1 is associated with neurodevelopmental disorders and Kury-Isidor syndrome, characterized by developmental delay and speech difficulties.

***3.2.4. MSTO1*:c.1259del chr1-155582996 AG>A p.Gly420ValfsTer2** was found to cause a frameshift mutation leading to a premature stop codon in the MSTO1 gene. MSTO1 is associated with mitochondrial function, and mutations in this gene may contribute to mitochondrial dysfunction affecting neurological functions.

Another variant MSTO1:c.1284-9T>C was classified as having unclear clinical significance.

3.3 Non-synonymous Variants in Genes Associated with Language Deficits

3.3.1 The CACNA gene family, encoding voltage-gated calcium channels, was investigated. Variants in CACNA1C, *CACNA1D, CACNA1F, and CACNB2* were associated with enhanced channel function, while variants in *CACNA1A and*

CACNA1H were linked to loss of function, affecting calcium signaling pathways involved in cortical development, transcription, and synaptogenesis.

This summary provides an overview of the key findings in the study regarding genetic variants associated with specific language impairment and other neurodevelopmental disorders. The identified genes such as ZNF277, NFXL1, PABPC1, PRODH, BAP1, and MSTO1, along with the *CACNA* gene family, are suggested as potential candidates for further investigation in the context of language deficits and related disorders.

Subunits CACNB1 and CACNB2, encoding auxiliary β-subunits, and CACNA2D3 and CACNA2D4 for auxiliary α2-δ-subunits, are also associated with ASD (110). In addition to ASD, calcium channelopathies are linked to intellectual disability (ID) and generalized developmental delay. M. Kessi and colleagues in 2021 reported variants enhancing function in 10 genes associated with these conditions: CACNA1A, CACNA1C, CACNA1I, CACNA1H, CACNA1D, CACNA2D1, CACNA2D2, CACNA1E, CACNA1F. Pathogenic variants in CACNA1C, CACNA1F, CACNA1I, CACNA2D1, and CACNA2D2 were described in association with epilepsy. According to the authors, CACNA2D1 is linked to autism spectrum disorder, while CACNA1A, CACNA1C, and CACNA2D1 are candidate genes for attention-deficit hyperactivity disorder (ADHD) .

In nine of our cases, we identified likely pathogenic or pathogenic variants as follows: CACNA1D: c.5914T>C chr3-53844047 T>C p.Tyr1972His in two cases; CACNA1H: c.450G>T chr16-1245470 G>T p.Glu150Asp; CACNA1E: c.2699G>T chr1-181701921 G>T p.Gly900Val; CACNA1A: c.1810G>A chr19-13419037 C>T p.Val604Ile; CACNA1G: c.3629G>A chr17-48677159 G>A p.Arg1210Gln; CACNA1C: c.3962A>G chr12-2760756 A>G p.His1321Arg; CACNA2D3: c.2820G>T chr3-55043415 G>T p.Leu940Phe in two cases. Considering the critical role of the CACNA gene family in neurodevelopmental and other neurological disorders, it is possible that the identified likely pathogenic variants in our cases may lead to a milder clinical phenotype characterized primarily by language deficits.

3.3.2. *CHD genes* are a family of genes encoding ATP-dependent chromatin remodeling complexes consisting of nine members: CHD1–CHD9. Cadherins are cell adhesion molecules and play a crucial role in neuronal development, the formation, and maintenance of synaptic connections between them. CHD proteins perform multiple functions crucial for cell survival and embryonic development, chromatin remodeling,

transcriptional regulation, and DNA repair (112). Pathogenic variants in CHD1 lead to developmental disruption associated with apraxia of speech, autism, hypotonia, and facial dysmorphia (113). Pathogenic variants in CHD2 cause developmental and epileptic encephalopathy (114). Pathogenic variants in CHD7 and CHD8 are associated, respectively, with CHARGE syndrome and a syndromic form of autism spectrum disorder (115). Recently, pathogenic variants in CHD3 and CHD4 have been described in patients with delayed intellectual development, macrocephaly, speech disorders, and characteristic dysmorphia.

CHD5 is located on chromosome 1p36.31. Patients with pathogenic variants in CHD5 share nonspecific clinical features with 1p36 deletion syndrome, a disorder characterized by moderate to severe intellectual impairment, language deficits, hypotonia, and seizures . In two of our patients, we identified likely pathogenic variants in CDH5: CHD5:c.2038G>T chr1-6203888 C>A p.Val680Leu and CHD5:c.1433C>T chr1-6206882 G>A p.Thr478Met. Two variants in CHD8: CHD8:c.5901C>A chr14-21862053 G>T p.Asn1967Lys and one each in CHD6 and 9. Based on the aforementioned literature data, we believe that mutations in these patients in the CHD genes may be associated with specific language disorders. Candidate genes for SEN associated with dyslexia

4.Analysis of Biological Pathways

Analysis of biological pathways through KEGG and Reactome showed pathways with the most interactions in both analyses related to the Organization of the Extracellular Matrix (Reactome) and Extracellular Matrix - Receptor Interaction (KEGG). From the gene bioinformatics analysis, it is observed that genes with nonsynonymous variants are associated with the extracellular matrix containing collagen (GO:0062023), actin cytoskeleton (GO:0015629), cytoskeleton (GO:0005856), spectrin-associated cytoskeleton complex (GO:0014731), sodium channel complex (GO:0034706), basement membrane (GO:0005604), lumen of the endoplasmic reticulum (GO:0005788), brush border membrane (GO:0031526), integral component of plasma membrane (GO:0005887), intermediate filament (GO:0005882), polymerizing cytoskeletal fiber (GO:0099513), sarcoplasmic reticulum (GO:0016529), cellular projection membrane (GO:0031253), and voltage-gated sodium channel complex (GO:0001518), intracellular organelle lumen (GO:0070013).

We found that genes with nonsynonymous variants are associated with ion channel activity (GO:0005216), actin binding (GO:0003779), transmembrane transport of inorganic cations (GO:0022890), cation channel activity (GO:0005261), microtubule motor activity (GO:0003777), motor activity (GO:0003774), mechanosensitive ion channel activity (GO:0008381), and endonuclease activity (GO:0004519).

4.1 Extracellular Matrix (ECM)

The extracellular matrix (ECM) is the 'skeleton' in which cellular components of all tissues are embedded. ECM constitutes approximately 40% of the developing and 20% of the total volume of the adult brain. Several recent studies reveal the dynamic nature of the composition and regulation of the ECM and its active role in the development, maturity, and pathogenesis of neurodevelopmental disorders.

ECM consists of a mixture of proteins and carbohydrates. The ECM of the central nervous system (CNS) differs from the ECM of other organs by the predominance of non-fibrillar components and two specialized forms of ECM. The first specialized form is perineuronal nets (PNNs), which primarily envelop inhibitory parvalbumin (PV)-expressing interneurons. While the composition of PNNs may vary in different brain regions and change with age, their backbone consists of chondroitin sulfate proteoglycans (CSPGs) lecticans, including aggrecan, brevican, and neurocan. The second CNS-specific form of ECM is the basement membrane (BM), including the meningeal BM, which envelop.

Perineuronal nets (PNNs) appear to reduce synaptic plasticity following critical periods of heightened neuroplasticity. Interestingly, PNNs are formed and mature at different times in various brain regions during development. For instance, PNNs are formed on postnatal day 4 (PD 4) in the brainstem, PD 14 in the cortex, and PD 21 in the amygdala. The region-specific developmental timing of PNN formation and maturation suggests that critical periods may also be specific to particular brain regions.

The extracellular matrix (ECM) regulates neuronal activity, partly by controlling extracellular ion homeostasis, the expression of neurotransmitter (NT) receptors and ion channels, and the maturation of neuronal processes (138, 139). Neuronal activity can be altered by binding ECM components such as reelin and fibronectin to their receptors on nerve cells through signaling pathways and kinases that enhance the activity of neurotransmitter receptors (e.g., NMDA) and voltage-gated calcium channels. The high negative charge of PNNs, attributed to sulfated glycosaminoglycan

residues (GAG), can bind ions and signaling molecules, including growth factors, and provide neuroprotection against oxidative stress. Due to its high hydration capacity, ECM can also regulate the extracellular space volume, thereby controlling ion and neurotransmitter levels and diffusion, and consequently, brain activity. The vascular basement membrane (BM) regulates which fluids and soluble molecules can enter and leave the brain .

Numerous studies link changes in ECM, including degradation, overproduction, and altered composition, to brain conditions such as Alzheimer's disease (AD), Huntington's disease (HD), Parkinson's disease (PD), schizophrenia, autism spectrum disorders (ASD), and epilepsy Enzymes, including matrix metalloproteinases (MMPs), disintegrin and metalloproteinase with thrombospondin motif type 1 (ADAMTs), hyaluronidases/chondroitinases, and plasmin, and their regulators, including tissue inhibitors of metalloproteinases (TIMPs), tissue plasminogen activator (tPA), and plasminogen activator inhibitor, control ECM degradation. MMPs, a zinc-dependent endopeptidase family consisting of 28 individual members, participate in various physiological processes, including tissue morphogenesis and cell migration, as well as pathophysiological processes like inflammation and cancer. MMP activity is regulated by their expression, processing, and the expression of their regulators. MMPs promote turnover of collagen, elastin, gelatin, and other matrix glycoproteins and proteoglycans constituting ECM. Importantly for PNNs, gelatinases such as MMP-2 and MMP-9 degrade aggrecan, laminin, and fibronectin (146). MMPs can be released by various cell types, including microglia.

In our study, nonsynonymous variants were identified in three members of matrix metalloproteinases (MMP8, MMP1, MMP9) and members of the ADAMTs family (*ADAMTLSL 4 ADAMTSL3, ADAMTSL5, ADAMTS15,12,14, ADAM 7, ADAM15, ADAM21*). Degradation of PNNs by MMPs in disease may contribute to parvalbumin (PV) interneuron dysfunction and consequently alter the balance of excitation and inhibition in affected brain regions. Loss of PNNs may also lead to reduced levels of growth factors, loss of neuroprotection against oxidative stress, altered ion concentration in the neuronal microenvironment, and disrupted expression of NT receptors and ion channels, causing maladaptive neuroplasticity .

Nonsynonymous mutations were found in the studied children, indicating an in silico damaging effect on the protein and a frequency below 0.01% in gnomAD in 100 genes

involved in the organization of the extracellular matrix and extracellular matrix-receptor interaction.

We can speculate that disturbances in ECM organization and function can significantly impact language development and other neurological functions. Mutations in genes encoding ECM proteins may be associated with various neurological disorders, including neurodevelopmental disorders. In the context of language development, the organization of ECM is particularly important for the formation and maintenance of synaptic connections between neurons in language-related brain areas. Additionally, changes in the composition and structure of ECM molecules have been observed in individuals with autism spectrum disorders, which are often characterized by language and communication impairments.

Ankyrins are a family of adaptor proteins that connect membrane proteins such as voltage-gated Na+ channels and Na+/K+ pumps to the spectrin-actin cytoskeleton. Ankyrins are encoded by three genes *(ANK3, ANK2, ANK1*), of which ankyrin-3 and -2 are the main forms expressed in the developing nervous system. Ankyrins bind to the cytoplasmic domain of L1 CAMs and connect them to ion channel proteins and the spectrin cytoskeleton. This binding enhances the homophilic adhesive activity of L1 and reduces its mobility in the plasma membrane. L1 interaction with ankyrin (Fig16.) mediates the branching and synaptogenesis of cortical inhibitory neurons .

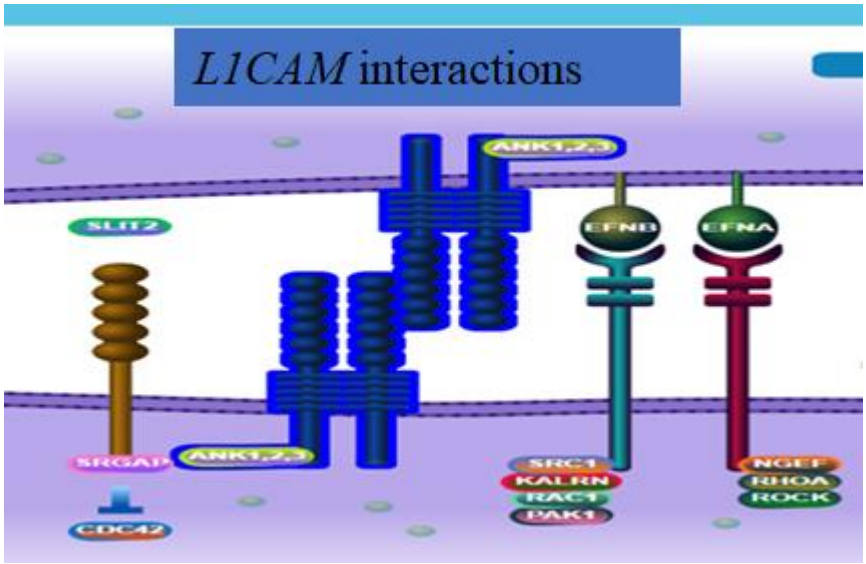


Fig. *16 L1CAM* interactions

ANK3 variants and polymorphisms have been described in children with ASD , but the detailed mechanisms underlying the neurological symptoms associated with ANK3 are

unknown. Interestingly, interactions between L1 and ankyrins have been established in a study on the role of rare variants in biological processes and molecular pathways associated with Alzheimer's disease pathogenesis, indicating prospects for further investigation in the context of neurological disorders and neurodevelopmental disorders .

The L1-type cell adhesion molecules (CAMs) play a crucial role as mediators in neuronal differentiation, axonal growth and guidance, as well as in the formation and maintenance of synapses. Their interactions with cytoskeletal components are highly conserved and tightly regulated. The L1 family of cell adhesion molecules (L1CAMs) constitutes a subfamily within the immunoglobulin superfamily of transmembrane receptors, consisting of four structurally related proteins: L1, close homolog of L1 (CHL1), NrCAM, and Neurofascin. These CAM molecules encompass six immunoglobulin-like domains, five or six fibronectin-like repeats, a transmembrane region, and a cytoplasmic domain. The L1CAM family is associated with processes integral to the development of the nervous system, including neurite outgrowth, neurite fasciculation, and interneuronal adhesion.

Members of the L1CAM family are predominantly expressed by neuronal cells, as well as some non-neuronal cells, during development. With the exception of CHL1, all other members of the L1 family contain an alternatively spliced 12-nucleotide exon that encodes the amino acid residues RSLE in neuronal splice forms, but is absent in non-neuronal cells. The extracellular regions of L1CAM members exhibit diversity and differ in their abilities to interact with extracellular heterophilic ligands. L1 ligands include other Ig-domain CAMs (such as NCAM, TAG-1/axonin, and F11), proteoglycan molecules (neurocan), beta1 integrins, and extracellular matrix protein laminin, Neuropilin-1, FGF, and EGF receptors. Some of these proteins that interact with L1 also bind to other members of the L1CAM family. For instance, TAG-1/axonin interacts with L1 and NrCAM; L1, neurofascin, and CHL1 bind to contact members of the family. The cytoplasmic domains of L1CAM members are the most strongly conserved. Nevertheless, they possess different cytoplasmic binding partners, and even those with similar binding partners can be involved in various signaling complexes and mechanisms. The most conservative characteristic of L1CAMs is their ability to interact with the actin cytoskeletal adapter protein ankyrin. The cytoplasmic ankyrin-binding domain exhibits the highest degree of amino acid conservation in the L1 family.

In our study, we identified variants with a frequency below 0.001% in gnomAD and a likely pathogenic effect on protein function through in silico analysis for: L1CAM (1 child), ANK1 (1 child), ANK2 (4 children), and ANK3 (1 child). Based on the available information, we can speculate that variants in the genes L1CAM, ANK1, ANK2, and ANK3 may be associated with the pathogenesis of sensorineural hearing loss (SNHL). Although the involvement of these genes in SNHL is not fully understood, there is evidence suggesting their potential role in neurodevelopmental disorders such as RAS (149, 151, 152, 157) and Alzheimer's disease.

L1CAM is a gene that encodes a protein participating in the development of the nervous system. Mutations in this gene are associated with various neurological conditions, including X-linked intellectual disabilities and hydrocephalus. While a well-established connection between SNHL and L1CAM mutations is lacking, we identified a variant in L1CAM (156).

Regarding the genetic variants found in these genes, the gnomAD frequency below 0.001 suggests that the variants are rare in the general population, and therefore, more likely to have a deleterious effect. Additionally, classifications by PolyPhen and SIFT imply that the variants are predicted to impair protein function. Nevertheless, it is crucial to note that the presence of a rare variant predicted to be deleterious does not necessarily imply it is the cause of the condition.

In conclusion, additional studies are necessary to fully comprehend the genetic factors and environmental factors contributing to SNHL, including the potential involvement of L1CAM, ANK1, ANK2, and ANK3 mutations.

4.2. Sensory Processing of Sound by Outer Hair Cells of the Cochlea

Outer hair cells (OHCs) amplify sound waves in the cochlea through contraction and elongation in response to sound, a phenomenon known as electromotility ((158-161). Similar to inner hair cells, OHCs possess apical stereocilia arranged in rows of increasing height. A taller stereocilium is connected to a shorter stereocilium by a top link, involving a CDH23 dimer on the taller stereocilium's side and a PCDH15 dimer at the top of the shorter stereocilium. PCDH15 interacts with LHFPL5, a subunit of the mechanoelectrical transduction (MET) channel complex, which includes TMC1 or TMC2, TMIE, CIB2, and LHFPL5 (158). Deflection of stereocilia in one direction creates tension on the top link, increasing the likelihood of MET channel opening,

leading to OHC depolarization. Deflection of stereocilia in the opposite direction causes compression of the top link, decreasing the likelihood of MET channel opening, leading to OHC hyperpolarization. This process plays a crucial role in the sensory processing of sound in the cochlea.

The study explores the mechanosensory processes leading to hyperpolarization of the cochlear outer hair cells (OHC). Sound induces micromechanical movements in the cochlear organ, resulting in variable voltage and compression in the tip link. This, in turn, generates excitatory-inhibitory cycles of opening and closing of MET channels based on the probability of the channel being in the open state. These cycles cause directed alternating flows of K+ and Ca2+, leading to depolarization-hyperpolarization cycles, inducing conformational changes in prestin (SLC26A5). These cycles are asymmetrical, with contraction caused by depolarization dominating over elongation caused by hyperpolarization due to asymmetry in the open probability of MET channels. Stereociliary ATP2B2 (PMCA2) extrudes calcium ions, and basally located KCNQ4 extrudes potassium ions to repolarize the cochlear outer hair cell. OHC depolarization reduces the length of the cochlear outer hair cell due to a rapid, voltage-sensitive conformational change in the membrane protein prestin (SLC26A5), an unusual member of the anion transporter family located in the lateral membrane, which appears to respond to cytosolic chloride by altering its conformation in the plane of the plasma membrane (162). Prestin also acts as a weak chloride-bicarbonate antiporter. Changes in the length of the cochlear outer hair cell cause movement of the reticular lamina towards and away from the basilar membrane.

In our study, we identified three variants in three genes, TMC1, TMC2, and PCDH15, related to the sensory processing of sound by external hair cells in the cochlea. We found two children with Specific Language Impairment (SLI) who have the TMC1 variant rs11143384 (allele frequency 0.05, compared to the allele frequency in the Genome Aggregation Database (gnomaAD) - 0.007012). We also found a variant in TMC2, rs145587686 c.2038G>A, and two variants in PCDH15, rs41274622 3 and rs111033436, with allele frequencies of 0.075 and 0.025 in our cohort, compared to allele frequencies in gnomaAD of 0.00767 and 0.003803, respectively. In silico analysis using SIFT and PolyPhen predicted that all identified variants are pathogenic or likely pathogenic. Although it is known that heterozygous mutations in TMC1, TMC2, and PCDH15 do not cause hearing impairment, we suggest that these variants

may alter hearing ability without leading to deafness, potentially resulting in delayed language development.

Our findings suggest that these genetic variants may be in a combined heterozygous state with other benign variants, leading to subtle changes in hearing ability, ultimately contributing to SLI. It is also worth noting that the allele frequencies of the identified variants in our cohort are higher than those in the general population, suggesting a potential connection between these genetic variants and SLI.

In conclusion, our study provides preliminary evidence for a potential association between genetic variants in *TMC1, TMC2, and PCDH15* and SLI. Future studies with larger sample sizes and functional investigations are needed to confirm these findings and further explore the potential mechanisms underlying the relationship between these genetic variants and language impairments.

V. Conclusion

For the first time in Bulgaria, we conducted a comprehensive transcriptional analysis and whole-exome sequencing in Bulgarian patients with Specific Language Impairment (SLI). The data from our study support the heterogeneous nature of SLI. Various genetic, epigenetic, and environmental factors play a role in the etiology of this disorder. Genes with altered expression and identified variants in children with SLI from the transcriptomic analysis and whole-exome sequencing participate in pathways primarily related to the immune system, cellular signaling, and the extracellular matrix. The connection between the immune system and neurodevelopmental disorders, including SLI, is confirmed. Many of the genes involved in SLI participate in the development and maintenance of neural networks, axon guidance, and synaptic formation and plasticity. We hypothesize that variants in these genes are "mild" and likely discretely disrupt normal development and functioning of neural networks in areas related to language skills, leading to language impairments. Furthermore, the specific nature of the mutation and how it affects the function of the affected gene may also play a role in determining the severity of the phenotype. Functional studies are needed to fully understand the relationship between mild mutations in specific genes and SLI. This hypothesis provides a potentially useful framework for investigating the underlying mechanisms of this complex disorder. Among the known "candidate genes" in the literature, we identified variants with statistical significance in most of our patients in genes associated with dyslexia (DCDC2, DYX1C, KIAA0319, and ROBO1). This suggests potential overlap between the genetic basis of these two disorders, explaining their frequent comorbidity. It would be interesting to further clinically follow these patients and establish the manifestation of dyslexia in later life. Significant mutations were also found in a considerable number of patients in known "candidate genes" associated with autism spectrum disorder, such as CNTNAP5 and ZNF277. RAS and SLI may represent different points on a continuum of disorders sharing similar genetic and neurobiological foundations. Different combinations of genetic factors and environmental factors may contribute to different phenotypes along this continuum, with some individuals primarily showing language impairments (SLI), while others exhibit both language and social communication difficulties (RAS). Through whole-exome sequencing, we identified candidate genes that may play a role

in the development of SLI: PMABP1, PRODH, BAP1, genes from the CACNA family, and CDH. However, additional studies are needed to confirm the role of these genes in the development of SLI. Whole-exome sequencing represents a powerful tool for identifying genetic variants and mutations associated with SLI, and it may help clarify the fundamental mechanisms contributing to language impairments in affected individuals.

VI. Conclusions

1. mRNA isolated from peripheral whole blood can confirm the expression changes in brain tissue and the differentially expressed genes can be considered as biomarker molecules in SLI

2. Based on the bioinformatic analyzes of the differentially expressed genes and those established by WES, it can be assumed that their dysregulation leads to a change in key signaling pathways related to synaptic plasticity, disturbances in axon orientation, neuronal experience, extracellular matrix, and the sensory processing of sound.

3.Using whole exome sequencing, we identified new candidate genes that may play a role in the development of this condition: MSTO1, BAP1, PMABP1, and PRODH However, further studies are needed to confirm the role of these genes in the development of SEN

4. Our study confirms the hypothesis that despite the wide variety of genetic changes found in SEN, they all converge in key signaling pathways related to the development and functions of both the nervous and immune systems.

VII. Contributions

1.Scientific Contributions:

1. Candidate genes related to SLI development were identified.
2. Biological pathways related to the pathogenesis of SLI were identified.
3. The study demonstrates peripheral blood mRNA as a biomarker in SLI.
4. Evidence supporting the hypothesis that, despite the diversity of genetic changes found in SLI, most converge into key signaling pathways.

2.Applied Contributions:

5. The first comprehensive transcriptional analysis and whole-exome sequencing were conducted on children with SLI in Bulgaria.
6. Candidate genes were identified as potential biomarkers.
7. The DNA bank of SLI patients was expanded, providing opportunities for larger future studies.

Scientific Publications Related to the Dissertation Work:

- Two articles in refereed journals.
- One conference paper in the proceedings of the European Society of Human Genetics (ESHG) Conference.

1.Iglika Sotkova-Ivanova in: "Language Disorders: Clinical and Genetic Aspects," Ivan Stefanov Ivanov and Vladimir Borisov Pilosof, Nutrition and Development, Medical University – Plovdiv, ISBN 978-619-237-065-7.

2.I. Sotkova, R. Yordanova, A. Georgieva, E. Timova, I. Pacheva, V. Stoyanova, N. Popov, T. Vachev, I. Ivanov, Analysis of gene expression in children with Specific Language Impairment - initial results, Pediatry Journal, 2023;1:2-

3.I. Sotkova, H. Ivanov, Genetic and molecular basis of language impairment – a review, Pediatry Journal, 2023;1:7-9

Participation in Scientific Forums:

In Bulgaria:

1. Initial results from genetic studies in children with Specific Language Impairment (SLI), presented at the Fifth Multifaceted Conference (April 2-3, 2021).
2. Therapeutic possibilities in children with Specific Language Impairment, presented at the National Scientific Conference on Child Neurology, Psychiatry, and Developmental Psychology in 2021.
3. Genomic studies in children with Specific Language Impairment (SLI), presented at the Sixth Multifaceted Conference (April 7-8, 2023).

Abroad:

1.Presentation of the study "Increased gene expression of GINS2 in patients with Specific Language Impairment" at the European Society of Human Genetics (ESHG) Conference in 2020.