

VILNIUS UNIVERSITY

Evelina
SIAVRIENĖ

A Molecular and Functional Evaluation of Coding and Non- Coding Genome Sequence Variants and Copy Number Variants

SUMMARY OF DOCTORAL DISSERTATION

Medical and Health Sciences,
Medicine (M 001)

VILNIUS 2020

This dissertation was written between 2016 and 2020 at the Department of Human and Medical Genetics, Biomedical Sciences Institute, Faculty of Medicine, Vilnius University. The research was supported by the Research Council of Lithuania (No. S-MIP-17-19/LSS-150000-1179, INGENES project). A scholarship was granted for academic accomplishments in 2019. A scholarship was granted for study results in 2020.

Academic supervisor:

Acad. Prof. Habil. Dr. Vaidutis Kučinskas (Vilnius University, Medical and Health Sciences, Medicine, M 001).

Academic consultant:

Assoc. Prof. Dr. Eglė Preikšaitienė (Vilnius University, Medical and Health Sciences, Medicine, M 001).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Loreta Cimbalistienė (Vilnius University, Medical and Health Sciences, Medicine, M 001).

Members:

Prof. Dr. Arvydas Kaminskas (Vilnius University, Medical and Health Sciences, Medicine, M 001);

Prof. Dr. Vytautas Kasiulevičius (Vilnius University, Medical and Health Sciences, Medicine, M 001);

Prof. Dr. Jānis Kloviņš (Latvian Biomedical Research and Study Centre, Natural Sciences, Biology, N 010);

Prof. Habil. Dr. Limas Kupčinskas (Vilnius University, Medical and Health Sciences, Medicine, M 001).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 14:00 PM on October 2, 2020 in the Red Auditorium of the Vilnius University Hospital Santaros Clinics, 2 Santariškių St., Vilnius, Lithuania.

The text of this dissertation can be accessed at the libraries of Vilnius University as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

Evelina
SIAVRIENĖ

Koduojančios ir nekoduojančios genomo sekos variantų bei kopijų skaičiaus pokyčių molekulinis ir funkcinis vertinimas

DAKTARO DISERTACIJOS SANTRAUKA

Medicinos ir sveikatos mokslai,
Medicina (M 001)

VILNIUS 2020

Disertacija rengta 2016–2020 metais Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedroje. Mokslinius tyrimus rėmė Lietuvos mokslo taryba (Nr. S-MIP-17-19/LSS-150000-1179; INGENES projektas). 2019 metais gauta stipendija už akademinis pasiekimus. 2020 metais gauta stipendija už studijų rezultatus.

Mokslinis vadovas:

Akad. prof. habil. dr. Vaidutis Kučinskas (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Mokslinė konsultantė:

Doc. dr. Eglė Preikšaitienė (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Gynimo taryba:

Pirmininkė – **prof. dr. Loreta Cimbalistienė** (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Nariai:

prof. dr. Arvydas Kaminskas (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001);

prof. dr. Vytautas Kasiulevičius (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M001);

prof. dr. Jānis Kloviņš (Latvijos biomedicininių tyrimų ir studijų centras; gamtos mokslai, biologija – N 010);

prof. habil. dr. Limas Kupčinskas (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Disertacija ginama viešame Gynimo tarybos posėdyje 2020 m. spalio mėn. 2 d. 14 val. Vilniaus universiteto ligoninės Santaros klinikų Raudonojoje auditorijoje, Santariškių g. 2, Vilnius, Lietuva. Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: <https://www.vu.lt/naujienos/ivykiu-kalendorius>

ABBREVIATIONS

3'UTR	– 3' untranslated region
aCGH	– array comparative genome hybridization
ACMG	– American College of Medical Genetics and Genomics
CA	– congenital anomalies
cDNA	– complementary (copy) deoxyribonucleic acid
CHERISH	– research project “Improving Diagnoses of Mental retardation in Children in Eastern Europe and Central Asia through Genetic Characterisation and Bioinformatics/Statistics”, 2009-2012
CNVs	– copy number variants
CRISPR-Cas9	– Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR associated protein 9
ENCODE	– Encyclopedia of DNA Elements
FC	– fold change
FISH	– fluorescent <i>in situ</i> hybridization
gDNA	– genomic deoxyribonucleic acid
ID	– intellectual disability
INGENES	– research project “Deciphering the genetic architecture of intellectual disability and congenital anomalies,” 2017–2020
LITGEN	– research project “Genetic diversity of the population of Lithuania and changes of its genetic structure associated with evolution and common diseases,” 2011–2015

MAF	– minor allele frequency
NGS	– next generation sequencing
NMD	– nonsense-mediated decay
PCR	– polymerase chain reaction
PROGENET	– research project “Evaluation of Pathogenic Copy Number Variation in Etiopathology of Mental Retardation,” 2010–2011
qPCR	– quantitative real-time polymerase chain reaction
SNP-CGH	– single nucleotide polymorphism-comparative genome hybridization
TPM	– transcripts per million
UNIGENE	– research project “Unique genome variants in congenital neurodevelopmental disorders: origin, genomic mechanisms, functional and clinical consequences,” 2012–2016.
VU MF BMI DHMG	– Vilnius University Faculty of Medicine Institute of Biomedical Sciences Department of Human and Medical Genetics
WES	– whole exome sequencing

1. INTRODUCTION

The Human Genome Project is one of the most famous and significant international research projects in history. During this scientific project, an International Human Genome Sequencing Consortium published the complete sequence of the human genome and found that human genomes are about 99.9% identical, while the remaining 0.1% is the reason of difference between people, caused by different DNA sequence variants, the spectrum of which varies from point variants in specific genes to changes in the number and structure of chromosomes [1].

The rapid advancement of molecular technologies, such as molecular karyotyping via array comparative genome hybridization (aCGH) and whole exome/genome sequencing (WES/WGS), has provided the necessary conditions for the identification of disease-causing variants in the human coding and non-coding DNA sequence. Previously it was thought that genetic alterations were functionally important only in protein coding regions; however, during the ENCODE project it has been found that more than 80% of untranslated DNA sequence is transcribed and regulates gene expression [2–4].

According to data from the 1000 Genomes Project, a typical human genome contains about 20,000–23,000 variants in synonymous and nonsynonymous regions. Approximately 530–610 of the variants have functional impact by causing in-frame deletions and insertions, premature stop codons, frameshifts, or by disrupting splice sites [5, 6]. Nevertheless, the number of identified variants and CNVs is constantly increasing. Based on *ClinVar* [7], to date the human genome contains about 670,416 genetic variants, more than 40,000 of which are associated with intellectual disability (ID) [8].

ID is one of the most common neurodevelopmental disability that affects 1–3% of the general population and represents one of the major health care socioeconomic issue. ID encompasses a wide spectrum of clinical phenotypes, including various congenital

anomalies (CA). The evaluation of the genetic causes of ID/CA remains challenging because of their highly heterogeneous nature. For this reason, the causes and molecular basis of many ID/CA are still unknown [9–12].

ID/CA is a relevant health care issue in Lithuania as well. Since 2009, more than 400 individuals with syndromic and non-syndromic ID and/or CA have been investigated during LITGEN, CHERISH, PROGENET, and UNIGENE projects. These studies revealed novel DNA sequence variants, CNVs, and candidate genes for ID/CA by the use of the most advanced techniques.

Although the techniques for detecting variants are now becoming a routine procedure all over the world, the key question concerns the function of detected variants. According to *ClinVar* [7] data, only 12% of DNA sequence variants and CNVs are pathogenic and 17% are benign. This means that the significance of the remaining genetic alterations is unknown [8]. Therefore, the challenge of performing a detailed analysis and determining the function of the altered gene is still open [13]. In order to describe the pathogenicity of such DNA sequence variants and CNVs, it is necessary to perform a large-scale molecular and functional genome analysis that involves different fields of study: genomics, epigenomics, transcriptomics, proteomics, and/or interactomics. In this study, DNA sequence variants and CNVs of unknown significance were examined at the mRNA level for the assessment of the etiopathogenesis of ID/CA, which was chosen as a disease model.

Aim of the study

This study aims to evaluate the pathogenicity of splicing variants and copy number variants that cause intellectual disability and/or congenital anomalies through extensive clinical, molecular, and functional characterization.

Main tasks of the study

1. To analyze next-generation sequencing data and select coding and non-coding DNA sequence variants and CNVs of uncertain significance and which may be associated with individuals' intellectual disability and/or congenital anomalies.
2. To investigate DNA sequence variants and CNVs at the mRNA level using qPCR and Sanger sequencing technology.
3. To evaluate the pathogenicity of the DNA sequence variants and CNVs after the analysis of scientific literature and databases.
4. To predict the effects of changes in DNA sequence at the protein level using bioinformatics tools.
5. To confirm the pathogenicity of selected genetic variant using CRISPR-Cas9 technology in human fibroblast culture.
6. To characterize the relationship between genotype and phenotype based on clinical, molecular, and functional data.

Relevance and novelty of the study

In this study, previously undescribed novel DNA sequence variants and CNVs have been selected for a molecular and functional characterization at the mRNA level. The analysis of cDNA confirmed the pathogenicity of splice site variants, the unique consequences of which were revealed at the protein level using a detailed *in silico* analysis. The investigation of CNVs provided the information about the origin, size, and gene content of the chromosomal alteration, thus allowing to identify a new syndrome, to narrow the critical regions of known syndromes, or even to identify the critical genes for syndromes. In order to confirm the pathogenicity of a genetic variant, for the first time, the novel CRISPR-Cas9 technology was used for the *MED13L* gene silencing experiment in the fibroblasts culture.

Functional genomic approaches, such as mRNA expression analysis and genome editing tools, provide an unique possibility to understand the etiopathogenetic mechanisms of many diseases and conditions, including ID/CA. The expansion of the scientific knowledge and molecular diagnostic capability may contribute for the development of a novel diagnostic and therapeutic strategies in the future.

Statements to be defended

1. mRNA and/or cDNA assays by qPCR, Sanger sequencing of cDNA and cell culture genome editing via CRISPR-Cas9 are effective tools for assessing the pathogenicity of genetic variants.
2. 5' donor splice site variants tend to cause skipping of the upstream exon, while 3' acceptor splice site variants usually lead to downstream exon skipping.
3. Pathogenic heterozygous variants of *MED13L* affect the viability and senescence of fibroblast culture as well as alter the expression of genes that are involved in the cell cycle control.
4. A molecular, functional, and clinical characterization of CNVs is crucial for the assessment of the genetic disease critical regions and the identification of new syndromes.

2. MATERIALS AND METHODS

This research was funded by a grant (No. S-MIP-17-19/LSS-150000-1179) from the Research Council of Lithuania. Ten individuals (#1-#10) with ID, psychomotor developmental delay, and/or at least one CA and who were exposed with DNA sequence variants and CNVs of uncertain significance during previous research projects (CHERISH, PROGENET, LITGEN, and UNIGENE) have been selected for this study. Phenotype data of the individuals were obtained from medical records. The biological material (blood and skin biopsy) has been collected from the individuals, after obtaining written informed consent. All procedures performed in this study were in accordance with the 1964 Helsinki declaration and its later amendments. The study was approved by the Vilnius Regional Biomedical Research Ethics Committee of Lithuania (Permission No 158200-17-962-469). The strategy of the dissertational work is shown in Figure 1.

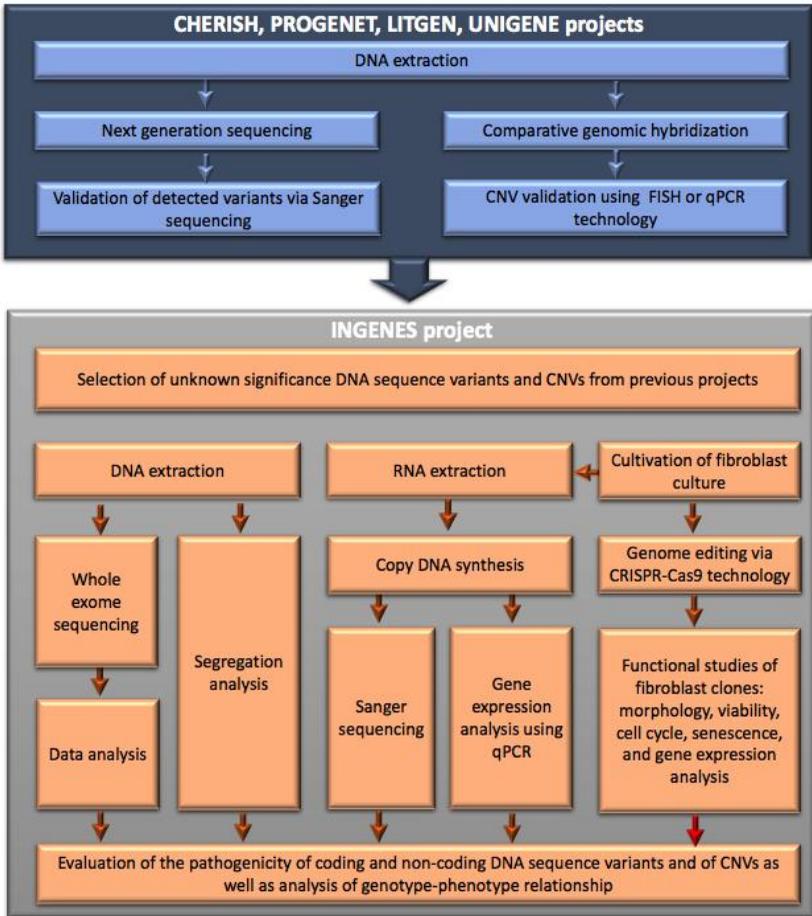


Figure 1. The strategy of the genetic research

DNA extraction

The gDNA, extracted from peripheral blood of the individuals in VU MF BMI DHMG following the standard phenol-chloroform extraction protocol or using the magnetic beads on the automatic robotic system TECAN Freedom EVO® 200 (Tecan Schweiz AG).

Cell culture

A primary fibroblast cell line was obtained from individuals (#2 and #10) and control individual skin biopsy. The growing cell line was cultured in AmnioMAX C-100 Basal Medium (Thermo Fisher Scientific), which was supplemented with AmnioMAX C-100 Supplement (Thermo Fisher Scientific) and Amphotericin B (Gibco) according to the standard laboratory procedures for human cell cultures. The final cell pellets were subsequently used for gene expression and/or gene editing assay.

RNA extraction and reverse transcription reaction

Total RNA of the individual #8 and #9 was isolated from whole blood using a PureLink Total RNA Blood Kit (Invitrogen), while whole blood RNA of the rest individuals (#1-#7, #10) as well as control individuals was extracted using Tempus™ Blood RNA Tube and Tempus™ Spin RNA Isolation Kit (Thermo Fisher Scientific) according to the optimized manufacturers' protocols. Total RNA was isolated from the fibroblast cell line using RNeasy Mini Kit (Qiagen). cDNA was synthesized from total RNA using a High-Capacity RNA-to-cDNA Kit (Thermo Fisher Scientific) following manufacturer's protocol. Reverse transcription reactions were performed in a ProFlex PCR system (Thermo Fisher Scientific) under the recommended conditions.

Gene expression analysis

The gene (*TGFBR2*, *BLM*, *GPC5*, *HNRNPD*, *HNRNPDL*, *ENOPH1*, *RASGEF1B*, *PRKG2*, *MED13L*, *RB1*, *E2F1*, *CCNC*) expression assays of the individual #5, #6, #8, #9, #10 and control individual have been performed by quantitative real-time PCR (qPCR). The PCR reaction mix was prepared by mixing cDNA with 2x TaqMan Gene Expression Master Mix (Applied Biosystems) and 20x TaqMan Gene Expression Assays (Applied Biosystems). PCR

assays were performed using a real-time PCR 7900HT system (Applied Biosystems) under standard conditions. The data were analyzed using SDS v.2.3 (Applied Biosystems), ExpressionSuite v.1.1 (ThermoFisher Scientific), qbase+ (Biogazelle), Microsoft Excel 2010 (Microsoft Corporation), and R v.3.4.0. Before the analysis, the Cq values of each sample were averaged over the three technical replicates. The fold change (FC) in target gene expression was calculated using the $2^{-\Delta\Delta CT}$ method and normalized to the *ACTB* (MIM #102630) and/or *GAPDH* (MIM #102630) housekeeping gene. For the comparative gene expression analysis, the sample from the proband was compared to the samples from random unrelated and self-reported healthy individuals and/or healthy parents, who were included as the external biological group. During the *TGFBR2* and *MED13L* gene expression analysis the internal control, in which probes were designed using different sites of target gene, was used. A more than two-fold change in the gene expression level was considered up- or down-regulation, respectively.

Sanger sequencing

PCR of gDNA and cDNA sequences were performed using specific primers designed with Primer Blast tool. PCR products were fractionated according to standard agarose gel electrophoresis. The PCR products were sequenced with BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and ABI 3130xL Genetic Analyser (Thermo Fisher Scientific). The sequences were aligned with the reference sequence of the *ARID1B* (NM_020732.3), *GLI3* (NM_000168.6), *SLC9A6* (NM_001042537.1), *CHD7* (NM_017780.4), *TGFBR2* (NM_001024847.2), *BLM* (NM_000057.2), *CAPN3* (NM_000070.3), *MED13L* (NM_015335.5) gene.

Whole exome sequencing

The gDNA sample of the individual #8 was sequenced using the high throughput whole exome sequencing (WES) technique (Illumina, Inc.), to uncover genetic variants possibly associated with the phenotype of the proband. DNA libraries generated using TruSeq Rapid Exome Library Prep kit (8x3plex) (Illumina, Inc.). The concentration of DNA libraries was measured with Qubit dsDNA BR Assay kit (Thermo Fisher Scientific) and Qubit fluorimeter (Thermo Fisher Scientific). Clusters amplified using cBot system (Illumina, Inc.), TruSeq PE Cluster Kit v3-HS (Illumina, Inc.), and TruSeq Dual Index Sequencing Primer Box – Paired End (Illumina, Inc.). WES was performed using TruSeq SBS Kit v3-HS (Illumina, Inc.) with the HiScanSQ (Illumina, Inc.) Genetic Analyzer.

The sequencing data was aligned against The Human NCBI Build GRCh37 (hg19/2009) reference genome. The annotation of WES data was performed with the ANNOVAR v.2018Apr16 program. The pathogenicity of variants was assessed using American College of Medical Genetics and Genomics (ACMG) criteria, taking into account the data provided by the ANNOVAR program, the available databases, and the scientific literature. The pathogenic and probably pathogenic sequence variants were checked by analysing individual's #8 BAM files using the visualization tool Integrative Genomics Viewer.

Gene editing via CRISPR-Cas9 technique

CRISPR-Cas9 genome editing technology has been applied to knock-out the *MEDI3L* gene in a culture of fibroblasts of healthy individual. Gene editing experiment performed using guide RNA (No CRISPR927871_SGM; TrueGuide™, Invitrogen), Cas9 protein (TrueCut™, Invitrogen, USA), and specific transfection reagent (Lipofectamine™, CRISPRMAX™, Invitrogen) according to the optimized manufacturer protocols. In order to detect the locus specific cleavage of gDNA, GeneArt™ Genomic Cleavage Detection

Kit (Invitrogen) has been used following manufacturer protocol. PCR products were fractioned according to standard agarose gel electrophoresis. ImageJ v.1.52t gel analysis software was used to determine the relative proportion of DNA contained in each band. The following equation was used to calculate the cleavage efficiency:

$$\text{Cleavage Efficiency} = 1 - [(1 - \text{fraction cleaved})^{1/2}]$$

For further functional analyses, 16 single cell clones have been cultivated after limiting dilution cloning in ten different 96-well plates. Due to a *force majeure* (quarantine due to the SARS-CoV2 virus), the experiment has been interrupted in the safe phase. After the end of the quarantine, the functional assays of cultured fibroblast have been performed.

Functional assays of fibroblast cultures

The viability of the cultured fibroblasts was assessed using a Bürker counting chamber (Heinz-Herenz) according to the manufacturer's instructions.

Cell morphology was determined using an inverted phase contrast microscope Olympus CKX41 (Olympus Life Science). Image analysis was performed in early (III-IV) and late (X-XV) passages using JuLI™ (NanoEnTek) analyzer.

Senescence of cultivated fibroblasts was investigated using Senescence Cells Histochemical Staining Kit (Sigma-Aldrich) according to the manufacturer's protocol. At the end of the staining procedure, ten pictures were taken from random areas of each culture. The percentage of senescent cells was calculated using the following formula: (number of cells with intracellular blue deposits/ total number of cells) x 100%.

Cell cycle analysis has been performed using propidium iodide and RNase (BD Bioscience) staining solution according to standard instructions as described by Kim and Sederstrom (2015) [14].

Statistical analysis

Statistical analysis was performed using SDS v.2.3 (Applied Biosystems), ExpressionSuite v.1.1 (ThermoFisher Scientific), qbase+ (Biogazelle), Microsoft Excel 2010 (Microsoft Corporation), and R v.3.4.0 software. The obtained data were displayed as an arithmetic mean \pm standard deviation.

In silico analysis

In order to evaluate the pathogenicity of detected variant/CNV and its impact on the structure and function of the protein, an *in silico* analysis was performed using different databases (e.g, gnomAD, dbSNP, dbVAR, DGV, DECIPHER) and bioinformatics tools (e.g., Mutation Taster, Human Splicing Finder, ExpASy, Pfam). In addition, a detailed analysis and review of scientific literature was performed.

3. RESULTS

In this study, ten individuals/families (#1-#10) with ID, psychomotor developmental delay, and/or at least one CA and who were exposed with DNA sequence variants and CNVs of uncertain significance have been investigated. The pathogenicity of pre-mRNA splicing variants in the *ARID1B*, *GLI3*, *SLC9A6*, *CHD7*, *TGFBR2*, *BLM*, and *CAPN3* genes has been evaluated in seven different individuals (#1-#7). The *TGFBR2*, *BLM*, *GPC5*, *HNRNPD*, *HNRNPDL*, *ENOPH1*, *RASGEF1B*, *PRKG2*, *MED13L*, *RB1*, *E2F1*, and *CCNC* gene expression was assessed for five different individuals. The size of intragenic *MED13L* deletion has been specified in individual #10. Additionally, CRISPR-Cas9 technology was used for the *MED13L* gene silencing experiment in the culture of skin fibroblasts of healthy individual. In order to evaluate the potential effects of splicing variants and CNVs at the protein level as

well as to understand the complex relationship between the genotype and phenotype, a detailed *in silico* analysis has been performed.

Molecular and functional analysis of splicing variants

In seven individuals (#1–#7) enrolled in this study, gene panel next generation sequencing (NGS), whole exome sequencing (WES), or Sanger sequencing was performed during previous scientific projects (CHERISH, PROGENET, LITGEN, UNIGENE). Three variants (*ARID1B* NM_020732.3:c.4986+2T>C, *GLI3* NM_000168.6:c.473+3A>T, NM_001042537.1:c.899+1G>A) have been detected in 5' donor splice site, three variants (*CHD7* NM_017780.4:c.5535-1G>A, *TGFBR2* NM_001024847.2:c.1600-2A>G, *BLM* NM_000057.2:c.2308-2A>G) have been detected in 3' acceptor splice site, while *CAPN3* NM_000070:c.1746-20C>G DNA sequence variant has been detected 20 nucleotides upstream of the 3' acceptor site. All these variants were in a heterozygous state. The variants were confirmed by Sanger sequencing on the probands' gDNA sample. Moreover, family segregation analysis confirmed the *de novo* origin of the *ARID1B* c.4986+2T>C and *CHD7* c.5535-1G>A variant, whereas the *GLI3* c.473+3A>T, *SLC9A6* c.899+1G>A, *BLM* c.2308-2A>G, and *CAPN3* c.1746-20C>G variants were inherited from one of the individuals' parent. The origin of the *TGFBR2* c.1600-2A>G variant has not been determined, because there was no possibility to investigate the parents' gDNA sample (Table 1).

Table 1. Splicing variants detected by NGS or Sanger sequencing technique

Individual	Detection method	Splicing variant	Inheritance	Gene
#1	WES	NM_020732.3: c.4986+2T>C	<i>De novo</i>	<i>ARID1B</i> MIM#614556
#2	Sanger	NM_000168.6: c.473+3A>T	Segregation in III generations	<i>GLI3</i> MIM#165240
#3	WES	NM_001042537.1: c.899+1G>A	Inherited from mother	<i>SLC9A6</i> MIM#300231
#4	WES	NM_017780.4: c.5535-1G>A	<i>De novo</i>	<i>CHD7</i> MIM#608892
#5	Gene panel NGS	NM_001024847.2: c.1600-2A>G	NA	<i>TGFBR2</i> MIM#190182
#6	WES	NM_000057.2: c.2308-2A>G	Inherited from father	<i>BLM</i> MIM#210900
#7	Gene panel NGS	NM_000070: c.1746-20C>G	Inherited from mother	<i>CAPN3</i> MIM#114240

NA – data is not available

The splicing variants of *ARID1B*, *GLI3*, *SLC9A6*, *CHD7*, and *TGFBR2* are novel and not recorded in the 1000 Genomes Project, gnomAD, HGMD, ClinVar or other databases. Although the *BLM* c.2308-2A>G and *CAPN3* c.1746-20C>G variants are recorded in scientific databases, both of them are rare (MAF<0,01) and to date there are conflicting interpretations of their pathogenicity. So far, none of these splicing variants have been published in scientific literature.

Computational algorithms were used to predict the effect of splicing variants on pre-mRNA. Mutation Taster predicted the variants to be likely pathogenic and disease causing. The alterations of the splice sites was also predicted by the Human Splicing Finder database. All these variants were predicted to affect pre-mRNA splicing.

In order to confirm the pathogenicity of the splicing variants, molecular and/or functional analysis of individuals' #1–#7 cDNA

synthesized from blood RNA were performed using Sanger sequencing or qPCR techniques.

For the individual #2, primarily, cDNA, synthesized from RNA, which was extracted from a total blood sample, was analyzed. To avoid gDNA contamination, the primer pair was designed on the cDNA sequence to anneal on the junction of exons 3 and 4 (forward primer) and on exon 5 (reverse primer). The expected amplification fragment size of this primer pair was 166 nt in length (amplicon #I). However, agarose gel electrophoresis did not fraction any PCR product in both individual's #2 and control sample, thus suggesting that this may be due to a low *GLI3* expression in the whole blood sample (Fig. 2 A). Further analysis of cDNA synthesized from individual's #2 RNA sample, extracted from a fibroblast cell line, was performed. To get a longer amplicon, another forward primer annealing on the junction of exons 2 and 3 with the same reverse primer on exon 5 was used. The expected size of the amplification fragment of this primer pair was 419 nt (amplicon #II). The results of gel electrophoresis revealed an amplification product of 166 nt (amplicon #I), thus confirming successful amplification of the patient's cDNA but not excluding the possibility of only wild type allele amplification. The forward primer of amplicon #I was designed to be complementary to the junction of exons 3 and 4 in the cDNA. For this reason, if exon 4 was skipped, there would be no possibility for the mutated allele to be amplified. Further analysis of the longer amplicon #II, which is 419 nt in length, confirmed this assumption, because the mutated cDNA produced an additional amplicon of about 313 nt in length (Fig. 2 B). The additional band was about 106 nt shorter than the band of the wild type allele, thus corresponding to the length of exon 4.

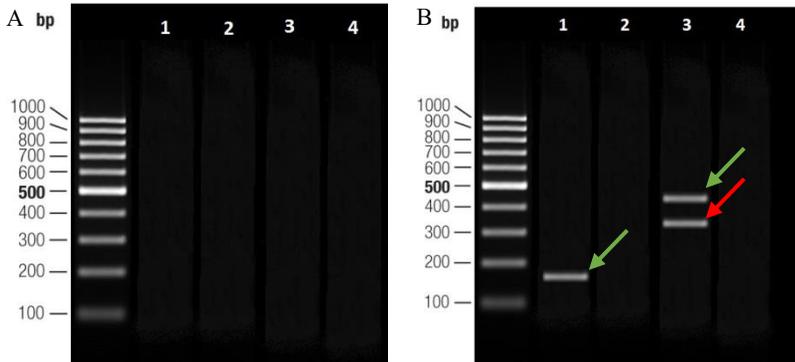


Figure 2. PCR amplifications of individual's #2 cDNA to verify the *GLI3* splice site c.473+3A>T variant visualized on agarose gel and the consequences of this variant; A) Lanes 1–2 show the results of the amplification of amplicon #I of cDNA synthesized from the RNA sample extracted from individual's #2 total blood sample. Lanes 3–4 correspond to the amplification of amplicon #I performed on the blood cDNA of the healthy control individual without c.473+3A>T variant (control sample). Lanes 2 and 4 correspond to the PCR reaction products of negative control. A 100 bp DNA ladder was used as the size marker. B) Lanes 1–4 show the results of the amplification of cDNA synthesized from the RNA sample that was extracted from individual's #2 fibroblast cell line. Lane 1 shows one band corresponding to a shorter amplicon (#I) of wild type allele of individual's #2 cDNA sample (166 nt); lane 3 shows the two bands derived from the amplification of the cDNA of individual's #2 (313 nt and 419 nt reflecting normal and mutated alleles [amplicon #II]); and lanes 2 and 4 correspond to the PCR reaction products of negative control. The green arrow indicates the amplicon of the wild type allele, while the red one indicates the amplicon of the mutated allele.

To verify these findings, PCR products were analyzed by Sanger sequencing. The analysis of the *GLI3* coding sequence revealed that the donor splice site variant led to the skipping of exon 4.

In case of individual #1, #3–#7, gene expression level in blood samples was sufficient to perform molecular and functional analyses. The analysis of the *ARID1B* coding sequence revealed that the donor

splice site c.4986+2T>C variant led to the skipping of exon 19. Similarly, the skipplings of exon 4, 6, 11, 14 were detected in the *GLI3*, *SLC9A6*, *BLM*, and *CAPN3* coding sequence, respectively. Sanger sequencing of the individual's #4 cDNA sample revealed that the c.5535-1G>A variant causes the loss of an original acceptor splice site at position - 1 in the intron 26 of *CHD7* and consequently activates a cryptic splice site only one nucleotide downstream of the pathogenic variant site.

The c.1600-2A>G splicing variant is located next to exon 8, which is the last exon of the *TGFBR2* gene. For this reason, the effect of this splicing variant on pre-mRNA was investigated by qPCR using individual's #5 cDNA sample. For the comparative gene expression analysis, the sample from the proband was compared to the samples from two random unrelated and self-reported healthy individuals (biological group, n=2) and the samples of internal control representing TaqMan probes designed on *TGFBR2* exons' 6-7 and 7-8 junctions. The relative quantification revealed that the expression level of *TGFBR2* exon 7-8 junction is approximately two-fold lower than the expression of exon 6-7 junction, thus providing the evidence that the exon 8 is skipped (Fig. 3). To confirm the exact sites of the disrupted pre-mRNA splicing process, the Sanger sequencing was performed using reverse primer designed on 3' untranslated region (3'UTR). The analysis revealed that only a wild type allele of the *TGFBR2* gene was sequenced. Based on these results, it is predicted that not only the eighth exon but also the 3'UTR is lost due to the c.1600-2A>G variant.

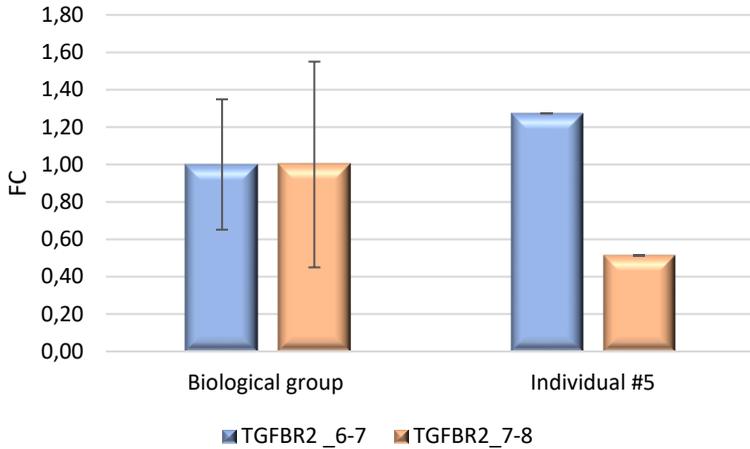


Figure 3. *TGFBR2* gene expression analysis. Histogram represents the FC values of the individual's #5 sample compared to the FC of samples of two random unrelated and self-reported healthy individuals merged into one biological group.

Although the *BLM* c.2308-2A>G splicing variant leads to exon 11 skipping and is probably pathogenic, only one heterozygous variant is insufficient to explain the pathogenesis of the suspected Bloom syndrome (OMIM #210900; ORPHA #125). Therefore, an additional *BLM* expression analysis and sister chromatid exchange (SCE) assay was performed. Compared to the biological group (n=4), the expression of *BLM* in individual's #6 blood cells was found to be within the normal range (Fig. 4 A). Also any abnormalities were revealed by SCE assay as the number of all SCEs in the metaphase was determined at the control level (Fig. 4 B).

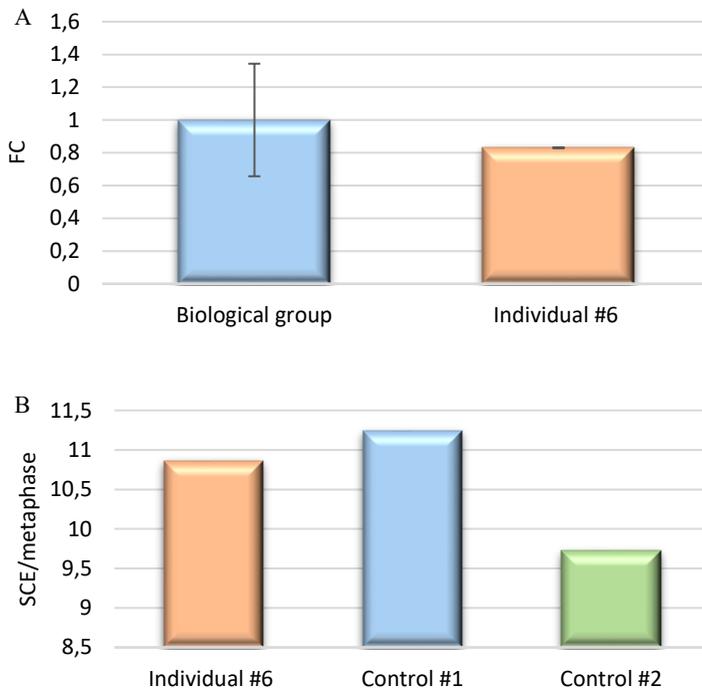


Figure 4. A) *BLM* gene expression analysis. The histogram represents the FC values of the individual's #6 sample compared to the FC of samples of four random unrelated and self-reported healthy individuals merged into one biological group; B) Sister chromatid exchange assay of individual's #6 and two healthy controls' samples.

A further *in silico* analysis revealed the consequences of aberrant splicing at the protein level. The abnormal splicing of the *ARID1B*, *GLI3*, *SLC9A6*, *CHD7*, and *CAPN3* gene is predicted to cause a translational frameshift and formation of premature termination codon, thus most probably resulting in protein truncation. The truncated *ARID1B*, *GLI3*, *SLC9A6*, *CHD7*, and *CAPN3* proteins are predicted to lose functional sequences. *ARID1B* (UniProtKB #Q8NFD5) truncation likely results in the loss of the BAF250 domain, which is part of the SWI/SNF-like ATP-dependent chromatin remodeling complex, which regulates gene expression

[16, 17]. The truncated GLI3 (UniProtKB #P10071) protein lacks almost all functionally important domains: part of repressor domain, zinc finger domain, proteolytic cleavage site, transactivation domain 2, and transactivation domain 1 [18]. The truncated SLC9A6 (UniProtKB #Q9258) protein does not contain several transmembrane helices and a C-terminal domain, which are crucial for proper Na⁺/H⁺ exchange [19, 20]. The premature truncation of the CHD7 protein (UniProtKB #Q9P2D1) is predicted to result in the loss of two BRK domains, which are especially important to higher eukaryotes and contribute to the regulation of gene expression [21, 22]. The truncated CAPN3 (UniProtKB #P20807) protein lacks a part of the calpain-type β -sandwich domain, a specific insertion region 2, and the penta E-F hand domain, which binds four calcium ions and is hypothesized to contribute to CAPN3 dimerization [25] (table 2).

The other splicing variants of *TGFBR2* and *BLM* do not result in a frameshift, but these changes lead to *TGFBR2* (UniProtKB #P37173) haploinsufficiency and to loss of the ATPase domain of *BLM* (UniProtKB #P54132), respectively. Therefore, the function of both *TGFBR2* and *BLM* protein is likely to be affected (table 2).

Table 2. *In silico* predicted consequences of aberrant splicing at protein level

Individual	Gene	Splicing variant	Consequence at protein level
#1	<i>ARID1B</i> MIM#614556	NM_020732.3: c.4986+2T>C	NP_065783.3: p.(Thr1633Valfs*11)
#2	<i>GLI3</i> MIM#165240	NM_000168.6: c.473+3A>T	NP_000159.3: p.(His123Argfs*57)
#3	<i>SLC9A6</i> MIM#300231	NM_001042537.1: c.899+1G>A	NP_001036002.1: p.(Val264Alafs*3)
#4	<i>CHD7</i> MIM#608892	NM_017780.4: c.5535-1G>A	NP_060250.2: p.(Gly1846Glu fs*23)
#5	<i>TGFBR2</i> MIM#190182	NM_001024847.2: c.1600-2A>G	<i>TGFBR2</i> haploinsufficiency
#6	<i>BLM</i> MIM#210900	NM_000057.2: c.2308-2A>G	NP_000048.1: p.(Ile700_Gln802del)
#7	<i>CAPN3</i> MIM#114240	NM_000070: c.1746-20C>G	NP_000061.1: p.(Glu582Aspfs*62)

Molecular and functional analysis of copy number variants

CNVs of coding and non-coding DNA sequence are one of the most common causes of IN/CA [26]. During previous projects (CHERISH, PROGENET, UNIGENE), three CNVs (#8, #9, #10) of unknown clinical significance have been detected via the molecular karyotyping technique. The analysis of the individuals' #8 and #9 gDNA by aCGH revealed an 844 kb duplication of the 13q31.3 cytoband spanning from 91,166,748 to 92,010,901 bp (GRCh37/hg19) and an 824 kb deletion of the 4q21.22 cytoband spanning from 83,273,844 to 84,097,897 bp (GRCh36/hg18), respectively. A 97.88 kb intragenic deletion in the cytoband 12q24.21 including exon 3 and 4 of the *MED13L* gene was detected by SNP-CGH. These novel CNVs were confirmed using the standard qPCR or fluorescent *in situ* hybridization (FISH) method. Moreover, the *de novo* origin of these deletions and duplication has been confirmed, as they were absent from individuals' #8, #9, and #10 parents' genomes. According to the Database of Genomic Variants, these deleted/duplicated genomic regions have never been described as copy number polymorphism and have not been found in the general population.

The 13q31.3 duplication encompasses the *MIR17HG* (MIM #609415) gene encoding miR-17~92 cluster, but not the *GPC5* gene. *GPC5* was suggested as playing a role in the phenotype presented by five individuals of two unrelated families with overlapping duplication within the 13q31.3 region (Fig. 5) [27, 28].

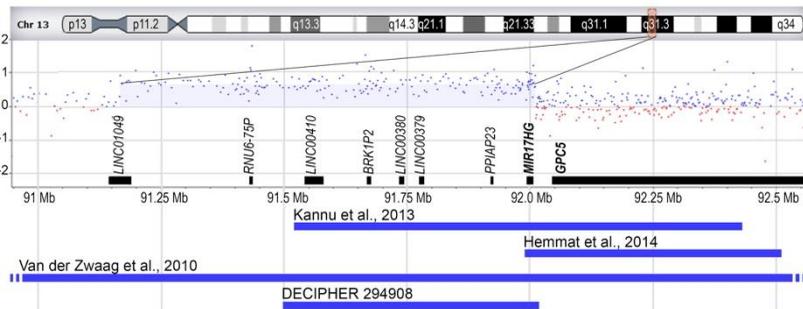


Figure 5. Ideogram of chromosome 13, individual's #8 aCGH results arr[GRCh37/hg19] 13q31.3(91,166,748-92,010,901)×3 *dn*, and schematic view of gene content are presented in the top portion. OMIM genes are highlighted in bold letters. Horizontal lines below represent previously reported overlapping duplications (blue) of 13q31.3

The proximity of *MIR17HG* and *GPC5* (about 40 kb), combined with the inclusion of at least one *GPC5* regulatory element (GH13J091346; GeneCards) in the individual's #8 duplication region, induced us to assess the possibility that the expression of *GPC5* would be perturbed in her cells. In this study the expression level of *GPC5* was measured and it was compared to that of two unrelated controls and both healthy parents (biological group, n=4). The analysis of qPCR data indicated that *GPC5* expression was highly variable in the population and not altered in the proband (Fig. 6).

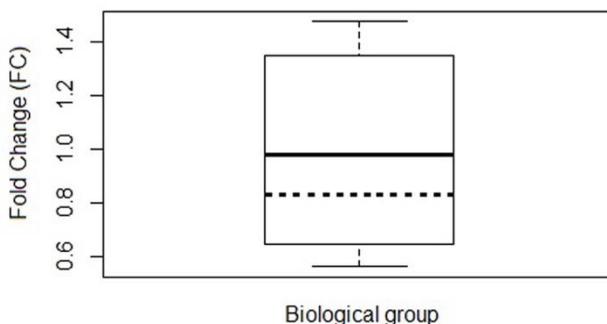


Figure 6. *GPC5* gene expression analysis. Boxplot diagram represents the FC values of the individual's #8 sample compared to the FC of four control samples merged into one biological group. The horizontal solid line inside the box is the mean value of the FC of the biological group (FC=1), while the dashed line indicates the FC of the individual's #8 sample (FC=0.83). The bottoms and tops of the boxes define the 25th and 75th percentile, and the whiskers below and above the box define the lowest and the highest values of the FC of the controls' samples, respectively.

In order to uncover the genetic variants possibly associated with the phenotype of the individual #8, WES has been performed on the probands gDNA sample. The analysis of WES data, including a special focus on genes associated with a marfanoid phenotype, failed to reveal any variants that could be associated with features of individual #8.

Another CNV, which was investigated in this study is 4q21.22 deletion, encompassing *HNRNPD* (MIM#601324), *HNRNPD*L (MIM #607137), *ENOPH1*, *TMEM150C* (MIM #617292), *SCD5* (MIM #608370), *SEC31A* (MIM #610257), *THAP9* (MIM #612537), *LINC00575*, miR575, and a part of the *LIN54* (MIM #613367) gene. Bonnet *et al.* [29] suggested *HNRNPD*, *HNRNPD*L, *ENOPH1*, *RASGEF1B* (MIM #614532), and *PRKG2* (MIM#601591) as candidate genes for 4q21 deletion syndrome (MIM #613509), which has been suspected to individual #9 (Fig. 7).

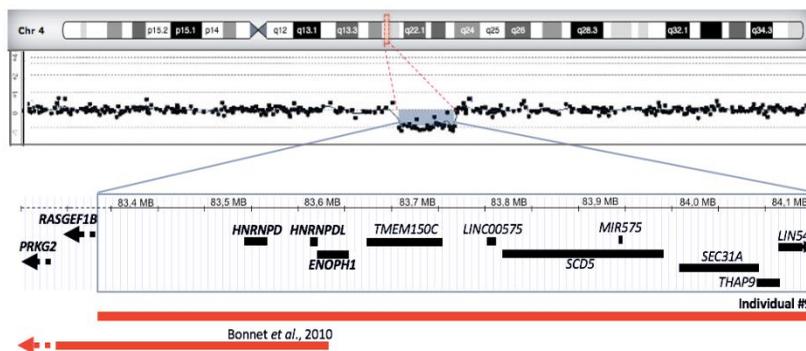


Figure 7. Ideogram of chromosome 4, individual's #9 aCGH results $\text{arr}[\text{GRCh36/hg18}] 4\text{q}21.22(83\ 273\ 844\text{--}84\ 097\ 897)\times 1\ \text{dn}$, and schematic view of gene content. OMIM genes are highlighted in bold letters. Horizontal lines below represent the overlapping 4q21.22 deletion of individual #9 and reported by Bonnet *et al.* (2009).

As CNVs influence transcriptomes by modifying the levels and timing of expression of genes within and adjacent to them [30–32], thus having an effect that can extend over the entire length of the

affected chromosome [33], we assessed the expression level of *HNRNPD*, *HNRNPDL*, and *ENOPH1* as well as neighboring *RASGEF1B* and *PRKG2* in cells of the individual #9. Gene expression level in individuals' #9 cells was evaluated comparing to the control samples of healthy parents merged into one biological group. Gene expression analysis revealed that the expression level of the *HNRNPD*, *HNRNPDL*, *ENOPH1*, and *PRKG2* is approximately two-fold lower in the sample of individual #9. However, the fold change (FC) values of *RASGEF1B* was within the normal range ($0.5 < FC < 2$; Fig. 8).

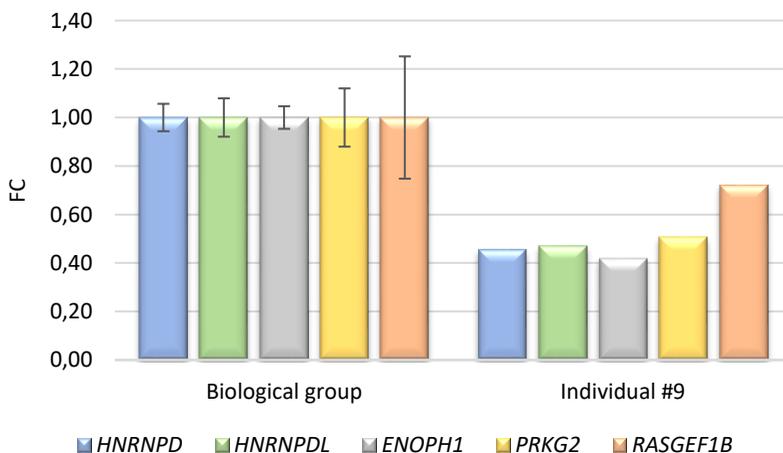


Figure 8. The *HNRNPD*, *HNRNPDL*, *ENOPH1*, *RASGEF1B*, and *PRKG2* gene expression analysis. Histograms represent the FC values of the individual's #9 sample compared to the FC of two control samples merged into one biological group. The whiskers define the standard deviation of the mean value.

Last but not least, in this study the size of intragenic *MEDI3L* deletion in 12q24.21 cytoband has been specified in individual's #10 cDNA sample (Fig. 9).

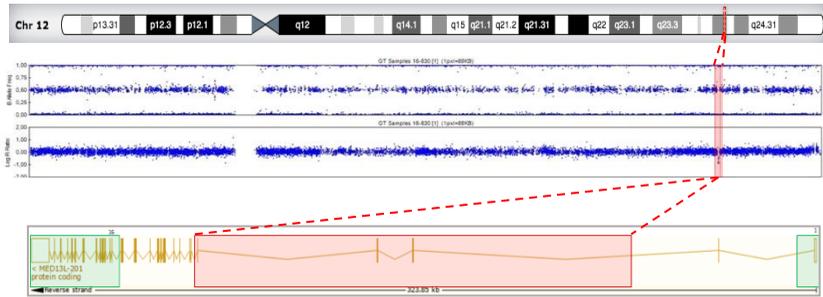


Figure 9. Ideogram of chromosome 12, individual's #10 SNP-CGH results arr[GRCh37/hg19] 12q24.21(116 523 305-116 621 185)x1 *dn*, and schematic view of the *MED13L* gene. The deleted exons are indicated in the red rectangle, while not deleted exons – in green rectangles.

Quantitative expression analysis of several *MED13L* exons disclosed the deletion of exons 3–4 and excluded the deletion of exons 1–2 and 16–17. This gene expression study indicated that expression level of exons 3–4 of *MED13L* is approximately two-fold lower comparing to samples of three random unrelated healthy individuals, while the expression level of exons 1–2 as well as exons 16–17 was similar and in normal range (Fig. 10).

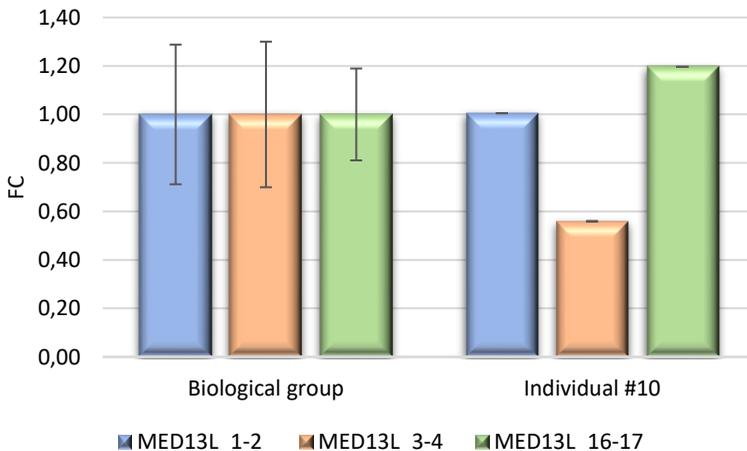


Figure 10. The *MED13L* gene expression analysis. The results indicate a mean fold change in target gene expression values of the control group (n=3) versus individual #10 by three different locations along *MED13L*, which span: exons 1–2 (marked as MED13L_1 in blue box), exons 3–4 (MED13L_2; orange), and exons 16–17 (MED13L_3; green).

The borders of the deletion in the coding region of the *MED13L* gene were narrowed by Sanger sequencing. For this purpose, different primer sets were designed. In case of wild type allele, the expected amplification fragment size of the primer pairs spanning exons: (1) from 2 to 8 was 953 nt in length (amplicon #A); (2) from 2 to 10 was 1262 nt (amplicon #B); (3) from 2 to 15 was 2524 nt (amplicon #C). The results of gel electrophoresis firstly revealed an amplification product of approximately 953 nt, thus confirming the successful amplification of patient's cDNA and suggesting the amplification of only wild-type allele. The analysis of the longer amplicon #B disclosed two bands, which were approximately 1262 nt and 292 nt in length, thus corresponding to wild-type and mutated alleles, respectively. The agarose gel electrophoresis of amplicon #C did not fraction any PCR product, thus suggesting that the expected products of both wild-type as well as mutated alleles were too long (Fig. 11 A). In order to verify these findings, PCR products of individual's #10 cDNA sample were analyzed by standard Sanger sequencing. The analysis of the *MED13L* coding sequence revealed the deletion, which spanned from exon 3 to 9 (Fig. 11 B).

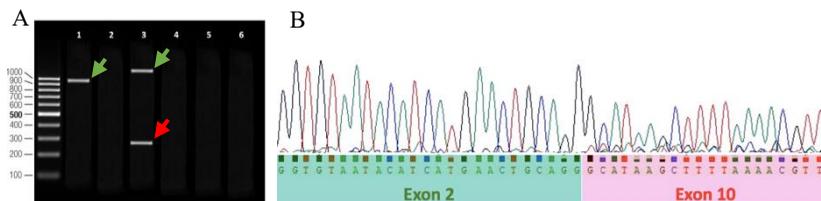


Figure 11. A) The gel electrophoresis of individual's #10 cDNA synthesized from the RNA sample. Lanes 1, 3, and 5 show the results

corresponding to amplicon #A, #B, and #C, respectively. Lanes 2, 4, and 6 correspond to PCR reaction products of negative control. The green arrows indicate the wild-type alleles, while the red arrow indicates the allele with intragenic *MED13L* deletion; B) The chromatogram represents the Sanger sequencing results of the *MED13L* coding sequence.

In silico, NG_023366.1(NM_015335.5):c.(310+1_311-1)(1280+1_1281-1)del was predicted to result in truncated MED13L (UniProtKB #Q71F56) protein, in which Med13_N domain is partly altered and medPIWI as well as Med13_C domains are lost.

Although the pathogenic variants of *MED13L* are known to be causative of the *MED13L* haploinsufficiency syndrome (also known as Intellectual disability and distinctive facial features with or without cardiac defects; MIM #616789; ORPHA #369891), the detailed mechanism of *MED13L* etiopathogenesis at the human cellular level have not been fully investigated. For this reason, CRISPR-Cas9 genome editing technology has been applied to knock-out the *MED13L* gene in a culture of skin fibroblasts of the healthy individual. The cleavage efficiency was approximately 44%. For further functional analyses, 16 single cell clones have been cultivated after limiting the dilution cloning procedure. However, the experiment has been interrupted in the safe phase because of quarantine due to the SARS-CoV2 virus. After the end of the quarantine, the functional assays of 11 cultured fibroblasts have been performed.

In order to identify *MED13L* variants in fibroblasts of a control healthy individual, Sanger sequencing of PCR products from modified fibroblast clones have been performed. Different *MED13L* variants have been detected in 6 out of 11 fibroblast clones (#A–#F; table 3). A heterozygous deletion of the second exon has been determined in three fibroblast clones (#A–#C), and a heterozygous deletion of two G nucleotides has been found in one cell culture (#D). In the other two fibroblast clones (#E, #F), different genetic

variants were found in both alleles (table 3). All these genetic variants *in silico* have been predicted to cause frameshift and premature termination codon, thus resulting in MED13L haploinsufficiency.

Table 3. *MED13L* variants detected by Sanger sequencing after genome editing via CRISPR-Cas9 technology.

Cell culture	Genotype	<i>MED13L</i> variants caused by CRISPR-Cas9	<i>In silico</i> predicted effect at the protein level
#A	Heterozygous	NG_023366.1(NM_015335.5): c.(72+1_73-1)_(310+1_311-1)del	NP_056150.1: p.(Ala25Cysfs*13)
#B	Heterozygous	NG_023366.1(NM_015335.5): c.(72+1_73-1)_(310+1_311-1)del	NP_056150.1: p.(Ala25Cysfs*13)
#C	Heterozygous	NG_023366.1(NM_015335.5): c.(72+1_73-1)_(310+1_311-1)del	NP_056150.1: p.(Ala25Cysfs*13)
#D	Heterozygous	NM_015335.5:c.95_96del	NP_056150.1: p.(Trp32Serfs*1)
#E	Different variants in both alleles	NG_023366.1(NM_015335.5): c.(72+1_73-1)_(310+1_311-1)del	NP_056150.1: p.(Ala25Cysfs*13)
		NM_015335.5:c.95_96del	NP_056150.1: p.(Trp32Serfs*1)
#F	Different variants in both alleles	NM_015335.5:c.95_96del	NP_056150.1: p.(Trp32Serfs*1)
		NM_015335.5:c.96_97insG	NP_056150.1: p.(Arg33Alafs*1)

The effect of these genetic variants on expression of genes involved in cell cycle regulation (*RBI* MIM #614041, *E2F1* MIM #189971, *CCNC* MIM #123838) was investigated. For the comparative gene expression analysis, the samples were compared to internal control representing TaqMan probes designed on *MED13L* exons' 1–2 junction (*MED13L_1*), as in the previous experiment it has been found that the expression levels of exons 1–2 are normal in samples of both individual #10 and three control individuals (Fig.10). Even though, according to GTE_x data, the expression level of investigated genes (*RBI* TPM = 38,1; *E2F1* TPM = 9,1; *CCNC*

TPM = 29,8) in cultured fibroblasts is similar to expression level of *MEDI3L* (TPM = 28,9), in this study qPCR analysis revealed that expression level of *RBI*, *E2F1*, and *CCNC* is more than seven-fold lower in fibroblasts of both individual #10 and control individual (Fig. 12).

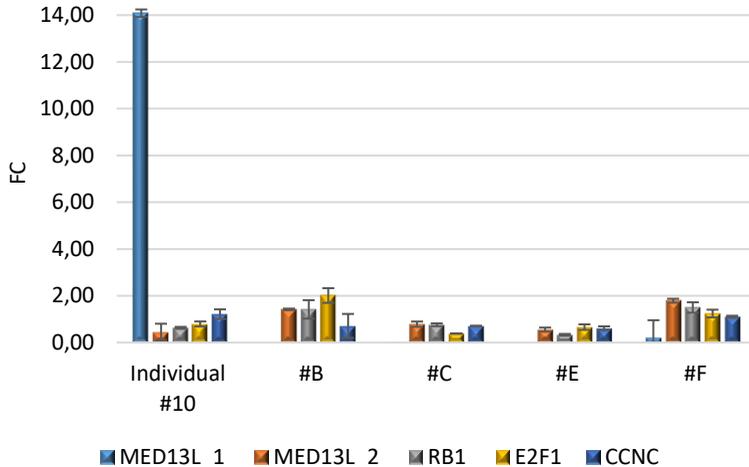


Figure 12. *MEDI3L*, *RBI*, *E2F1*, and *CCNC* expression analysis. The results indicate a mean fold change in target gene expression values of the modified fibroblast clones (#B, #C, #E, #F) of control individual, versus individual #10 by two different locations along *MEDI3L*. The RNA concentration of fibroblast clones #A and #D was insufficient for qPCR analysis.

Cultured fibroblasts of both individual #10 and control individual were observed microscopically at every passage. Before the gene editing experiment, adherent long spindle-shaped or flat cells were detected in both cultures at early passages (P3-P4; Fig.13 A, B). Later on, the proportion of enlarged cells with an altered morphology obviously increased also in both unmodified cells of individual #10 as well as the edited fibroblasts of the control individual.

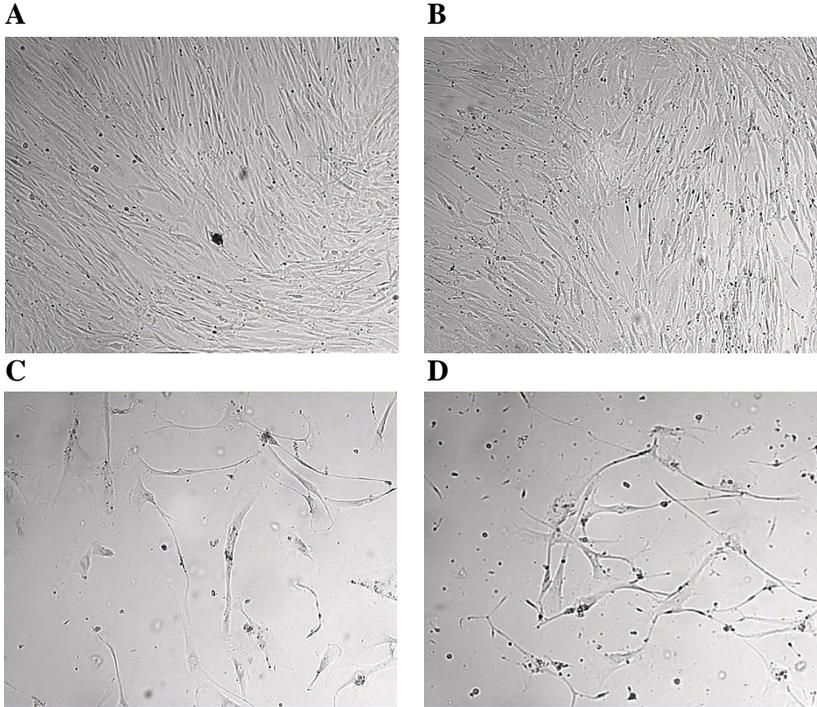


Figure 13. Early passages (P3–P4) of fibroblast culture of individual #10 (A) and control individual before gene editing experiment (B). Late passages (P10–P15) of fibroblast culture of individual #10 (C) and control individual after gene editing experiment (D). Original magnification x4, scale represent 500 μm .

Despite the fact that only single cells were morphologically changed in the early passage of individual's #10 fibroblast culture, cell dyeing for SA- β -gal showed almost three-fold (27.6%) more aging cells compared to unmodified fibroblasts of control individual (Fig. 14 A, B). Comparing long-term fibroblast cultures of control individual after a CRISPR-Cas9 genome editing experiment, it was found that the fibroblast culture, in which any *MED13L* variants have been detected, contained approximately 55.2% aging cells, while the fibroblast culture, in which the heterozygous *MED13L*

variant was found, accounted for 76.8% of aging fibroblasts (Fig. 14 C, D).

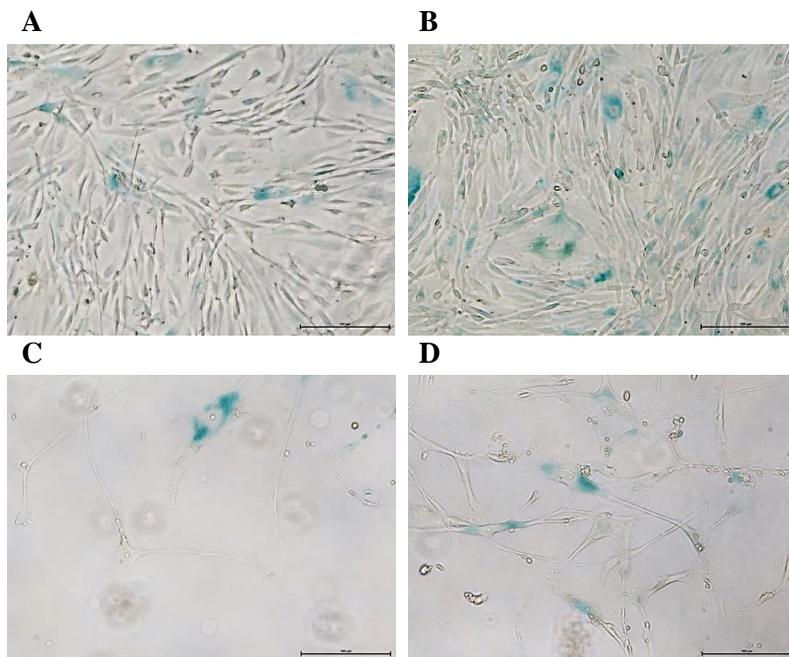


Figure 14. Early passages (P3–P4) of fibroblast culture of control individual before gene editing experiment (A) and individual #10 (B). Late passages (P10–P15) of fibroblast culture of control individual after gene editing experiment: fibroblast culture, in which any *MED13L* variants have been detected (C); fibroblast culture, in which the heterozygous *MED13L* variant was found (D). Original magnification x40, scale represent 100 μ m.

Prior to the genome editing experiment, the viability of control individual's and individual's #10 fibroblast culture was 96% and 89%, respectively. In addition, in modified fibroblast culture with a heterozygous *MED13L* variant, approximately two-fold more dead cells were found comparing to a fibroblast culture without the *MED13L* variant. Furthermore, assessing cell cycle changes of control individual's and individual's #10 fibroblast culture, in both fibroblast cultures the majority of cells (82–85%) were found in G1

(growth 1) phase. Some cells (3–5%) were in the S (synthesis) phase, and the remaining cells (12–15%) were detected in the G2 (growth 2) phase (Fig. 15 A, B).

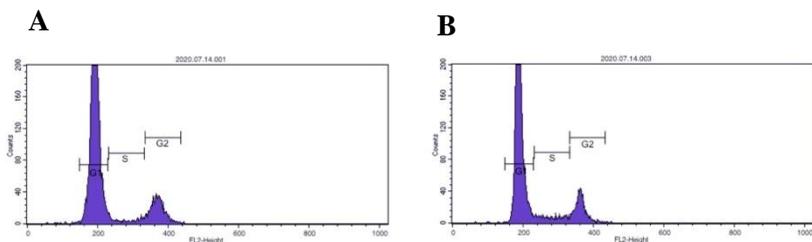


Figure 15. Results of cell cycle analysis of: unmodified control individual’s fibroblast culture at early passage (A); individual’s #10 fibroblast culture at early passage (B).

4. DISCUSSION

The rapid advancement of high-throughput NGS and molecular karyotyping techniques has provided the necessary conditions for the identification of disease-causing variants in the human genome [34]. Previous studies have shown that many genetic diseases have been caused by coding and non-coding CNVs as well as DNA sequence variants, which disrupt splice sites. Variants altering pre-mRNA splicing account for at least 7–10% of genetic alterations causing different disorders, including ID/CA, while CNVs are known to be one of the most common causes of neurodegenerative disorders [26, 35, 36].

Even though the technique for detecting variants is now becoming routine, most of detected variants are of unknown functional and clinical significance. Therefore, the challenge of performing a detailed analysis and determining the function of the altered gene is still open [37]. For this reason the aim of this study was to evaluate the pathogenicity of splicing variants and CNVs, which cause ID/CA through extensive clinical, molecular, and functional characterization. In order to achieve this aim, the consequences of

seven previously unreported splicing variants (of individual #1–#7) and three novel CNVs (of individual #8–#10) were assessed at the mRNA level.

Splicing variants usually refer to the point mutation at canonical splice sites in mRNA and result in the generation of an aberrant transcript of the mutated gene. The most common consequence of splicing variants is skipping of one or more exons, but in certain cases the activation of cryptic 5' (donor) or 3' (acceptor) splice site and partial or full retention of specific introns may occur [38–40]. In this study the splicing variants in the *ARID1B*, *GLI3*, *SLC9A6*, *TGFBR2*, *BLM*, and *CAPN3* gene led to exon skipping. *ARID1B* c.5025+2T>C, *GLI3* c.473+3A>T, and *SLC9A6* c.899+1G>A variants at 5' donor splice sites caused skipping of the upstream exon, while *BLM* c.2308-2A>G splicing variant at 3' acceptor splice site led to downstream exon skipping. *TGFBR2* c.1600-2A>G splicing variant, which is also located at the 3' acceptor splice site, resulted in a loss of downstream exon and 3'UTR. Although deep intronic variants usually create new splice sites resulting in the inclusion of cryptic exons [38], in this study, *CAPN3* c.1746-20C>G splicing variant also led to exon skipping.

Unexpectedly, the data obtained on the individual's #4 cDNA showed that the *CHD7* c.5535-1G>A variant does not lead to adjacent exon skipping; instead, it activates a cryptic 3' splice site, only one nucleotide downstream from the pathogenic variant site. According to reviewed literature and the results of this experiment, the variant at - 1 position in the acceptor splice site and the first G nucleotide in the adjacent exon sequence is quite frequent [41, 42]. However, such variants usually disturb the reading frame by exon skipping [38] instead of the slippage of the splicing mechanism only 1 nt downstream as in our case. There is only one similar study presented by Eng *et al.* [43], in which the acceptor splice site c.4437-1G>A variant in the *ATM* (MIM #607585) gene causes the same consequences to the cDNA sequence. In both cases, the first nucleotide of a particular exon is a G. The cryptic 3' splice site has

therefore been created using the mutated A of the intron and the G of the exon. These findings could suggest that if the original 3' splice site is mutated the subunits of U2 spliceosomal snRNAs recognise the next AG site, which is located downstream of the original acceptor splice site. However, Moon *et al.* [44] reported two abnormal transcripts of the *HRPT2* (MIM #145001) gene, also known as the *CDC73* gene, resulting from the same c.238-1G>A variant of intron 2. The first transcript eliminated the whole exon 3, while the another lacked the first 23 nucleotides of exon 3 due to the use of a cryptic acceptor splice site in exon 3. For this reason, the exact mechanism for how one cryptic splice site is selected over another within the genome remains unclear.

Further bioinformatic analysis revealed that *ARID1B* c.5025+2T>C, *GLI3* c.473+3A>T, *SLC9A6* c.899+1G>A, *CHD7* c.5535-1G>A, and *CAPN3* c.1746-20C>G splicing variant causes a translational frameshift and formation of premature termination codon, thus resulting either in protein truncation or mRNA degradation due to a nonsense-mediated decay. The Nonsense-mediated mRNA decay (NMD) pathway efficiently degrades mRNAs harboring a premature termination codon at least 200 nucleotides downstream of the start codon and 50–55 nucleotides upstream of the last exon–exon junction [45–47]. A premature termination codon might trigger the NMD pathway with a variable proportion of the mutated allele. If the NMD affects almost all of the transcripts, haploinsufficiency is the leading pathomechanism. On the other hand, the production of a truncated protein due to complete NMD escape, can impact its function [48]. Premature termination codon, induced by *ARID1B* c.5025+2T>C splicing variant, reside less than 50 nucleotides upstream of the last exon–exon junction; therefore, it is unable to trigger NMD. However, the premature termination codons induced by the splicing variants of the *GLI3*, *SLC9A6*, *CHD7*, and *CAPN3* gene conform the conditions for NMD; therefore, mRNA cleavage mechanisms could be initiated. However, the presence of both wild type and abnormal transcript resulting from

these splicing variants indicates that the abnormal transcripts do not undergo full NMD, but instead result in the production of the truncated proteins. The other splicing variants of *TGFBR2* and *BLM* cleavage do not result in a frameshift, but these changes lead to *TGFBR2* haploinsufficiency and the loss of some functionally important *BLM* domains, respectively. Therefore, the function of both *TGFBR2* and *BLM* protein is likely to be affected.

In this study, the *ARID1B* c.5025+2T>C, *GLI3* c.473+3A>T, *SLC9A6* c.899+1G>A, *CHD7* c.5535-1G>A, and *CAPN3* c.1746-20C>G splicing variants clearly showed their pathogenicity at both the RNA and protein levels, affecting the mRNA splicing and resulting in protein truncation, which is predicted to be the leading pathomechanism for clinical phenotypes: Coffin-Siris syndrome 1 (MIM #135900; ORPHA #1465), Greig Cephalopolysyndactyly (MIM #175700; ORPHA #380), Christianson type intellectual disability (MIM #300243; ORPHA #85278), CHARGE syndrome (MIM #214800; ORPHA #138), and Limb-girdle muscular dystrophy type 2A (MIM #253600; ORPHA #267), respectively. Haploinsufficiency of *TGFBR2* caused by splicing variant c.1600-2A>G of this gene leads to Loeys-Dietz syndrome 2 (MIM #610168; ORPHA #60030) (Table 4).

The *BLM* c.2308-2A>G splicing variant is not sufficient to explain the pathogenesis of suspected Bloom syndrome (OMIM #210900; ORPHA #125). The hypothesis of a compound heterozygous variant leading to Bloom syndrome has not been confirmed, as an analysis of the *BLM* gene expression as well as SCE did not revealed any other genetic alterations. Even though the c.2308-2A>G splicing variant does not determine the clinical phenotype of individual #6, this variant might be classified as likely pathogenic according to its frequency (MAF <0.01) in the general population and the predicted consequences in the conserved protein region (Table 4).

Table 4. Summarized results of molecular and/or functional studies of non-coding DNA sequence splicing variant.

Individual	Splicing variant	Gene	Consequences at the mRNA level	Predicted consequences at the protein level	Confirmed diagnosis	Pathogenicity
#1	NM_020732.3: c.4986+2T>C	<i>ARID1B</i> MIM #614556	19 exon skipping	NP_065783.3: p.(Thr1633Valfs*11)	Coffin-Siris syndrome 1 (MIM #135900)	Pathogenic
#2	NM_000168.6: c.473+3A>T	<i>GLI3</i> MIM #165240	4 exon skipping	NP_000159.3: p.(His123Argfs*57)	Greig Cephalo- polysyndactyly (MIM #175700)	Pathogenic
#3	NM_001042537.1: c.899+1G>A	<i>SLC9A6</i> MIM #300231	6 exon skipping	NP_001036002.1: p.(Val264Alafs*3)	Christianson type intellectual disability (MIM #300243)	Pathogenic
#4	NM_017780.4: c.5535-1G>A	<i>CHD7</i> MIM #608892	Activation of a cryptic 3' splice site	NP_060250.2: p.(Gly1846Glu fs*23)	CHARGE syndrome (MIM #214800)	Pathogenic
#5	NM_001024847.2: c.1600-2A>G	<i>TGFBR2</i> MIM #190182	8 exon and 3'UTR skipping	TGFBR2 (UniProtKB #P37173) haploinsufficiency	Loeys-Dietz syndrome 2 (MIM #610168)	Pathogenic
#6	NM_000057.2: c.2308-2A>G, rs1248548542	<i>BLM</i> MIM #210900	11 exon skipping	NP_000048.1: p.(Ile700_Gln802del)	-	Likely pathogenic
#7	NM_000070.3: c.1746-20C>G, rs201892814	<i>CAPN3</i> MIM #114240	14 exon skipping	NP_000061.1: p.(Glu582Aspfs*62)	Limb-girdle muscular dystrophia type 2A (MIM #253600)	Pathogenic

NA – data is not available

Additionally, a genotype-phenotype relationship analysis was performed in cases of novel CNVs: 13q31.3 duplication, 4q21.22 deletion, and intragenic deletion in *MED13L* gene.

Published reports of individuals from two unrelated families with the duplication of 13q31.3, partly involving *GPC5*, suggested that gene expression changes of both *GPC5* and *MIR17HG* could play a role in the individuals' abnormal growth and skeletal development [27, 28]. However, the limited extent of the individuals' rearrangement to the *MIR17HG* gene and the corresponding normal expression level of the neighboring *GPC5* contradict this statement and allow to define a new syndrome characterized by features mirroring those of Feingold syndrome 2 (OMIM #614326; ORPHA #391646).

According to reviewed literature, 4q21 deletion syndrome is caused by deletion of five genes: *HNRNPD*, *HNRNPDL*, *ENOPHI*, *RASGEF1B*, and *PRKG2* [29]. In this study, one of the smallest deletion, which encompassed only *HNRNPD*, *HNRNPDL*, and *ENOPHI*, has been revealed. The following expression analysis of adjacent two genes showed that the *RASGEF1B* gene expression was normal, while the expression level of *PRKG2* was reduced. Therefore, the critical region of 4q21 deletion syndrome has been narrowed to four candidate genes: *HNRNPD*, *HNRNPDL*, *ENOPHI*, and *PRKG2*.

In this dissertational work, the size of intragenic *MED13L* deletion has been specified in individual #10. An *in silico* analysis revealed that NG_023366.1(NM_015335.5):c.(310+1_311-1)(1280+1_1281-1)del results in a translational frameshift and formation of premature termination codon (NP_056150.1:p.(Val104Glyfs*5)). The premature termination codon is sufficiently distant from the start codon and the last exon-exon junction. Therefore, this region could be recognized by the NMD. However, the presence of both wild type and abnormal transcript resulting from this genetic alteration could suggest partial or full NMD avoidance. Although the pathogenicity of intragenic *MED13L*

deletion has been revealed, the fundamental molecular mechanism underlying the pathogenesis of this CNV was still not fully understood. In order to confirm these findings at the cellular level, CRISPR-Cas9 technology was used for the experiment of *MEDI3L* gene silencing in the culture of control individual's skin fibroblasts. After the *MEDI3L* gene editing experiment, reduced viability, accelerated aging process, and an inhibition of certain gene expression was assessed due to heterozygous *MEDI3L* variants.

In conclusion, despite a growing number of reported splicing variants and CNVs, the effect on the phenotype persists difficult to demonstrate, and therefore it remains one of the principal challenges in human molecular genetics. The results obtained herein enhance the importance of studying the pathogenicity of the genetic alterations at the mRNA/cDNA level. Moreover, the clinical, molecular, and/or functional characterization of genetic variants and successfully managed genome editing via CRISPR-Cas9 technology in this study provided a unique possibility to understand the etiology and pathophysiology of many hereditary diseases and conditions, including those associated with ID/CA. This fundamental scientific knowledge, together with the expansion of the capabilities of molecular diagnostic and functional investigations may contribute to the development of novel diagnostic and therapeutic strategies in the future. A straightforward benefit of the obtained results is an accurate diagnosis, which has been provided to multiple individuals and their families and gave an opportunity to improve their family planning, clinical care, and follow-up.

Despite the evidences proving the pathogenicity of detected splicing variants and CNVs, further proteomic assays and/or functional experiments of model organisms could be suggested to clarify and validate the molecular effects caused by these genetic variants on the protein and/or model organism level.

5. CONCLUSIONS

1. Seven splicing variants in the *ARID1B*, *GLI3*, *SLC9A6*, *CHD7*, *TGFBR2*, *BLM*, and *CAPN3* gene as well as three CNVs of 13q31.3, 4q21.22, and 12q24.21 genetic region were selected using bioinformatics tools for human genome sequence analysis to unravel their clinical significance.
2. In order to assess the pathogenicity of the selected DNA sequence variants and CNVs at the mRNA level, cDNA samples of ten unrelated individuals were analyzed.
3. DNA sequence variants/CNVs of unknown or unconfirmed clinical significance are classified as pathogenic (9) or likely pathogenic (1).
4. *In silico*, *ARID1B* c.5025+2T>C, *GLI3* c.473+3A>T, *SLC9A6* c.899+1G>A, *CHD7* c.5535-1G>A, *CAPN3* c.1746-20C>G and *MED13L* NG_023366.1(NM_015335.5):c.(310+1_311-1)(1280+1_1281-1)del cause a translational frameshift and formation of premature termination codon, while *BLM* c.2308-2A>G and *TGFBR2* c.1600-2A>G splicing variants lead to the loss of some functionally important protein domains. These genetic alterations most probably result in the production of the truncated proteins, which are the leading pathomechanism.
5. CRISPR-Cas9 technology, which has been used for the *MED13L* gene silencing experiment in the culture of skin fibroblasts, edited the target fragment with 44% efficiency and provided evidence that heterozygous *MED13L* variants lead to the reduction of cell viability, the acceleration of the aging process, and the inhibition of certain genes' expression.
6. A molecular, functional, and clinical characterization of splicing variants (*ARID1B*, *GLI3*, *SLC9A6*, *CHD7*, *TGFBR2*, *BLM* and *CAPN3*) and CNVs (13q31.3 duplication, 4q21.22 deletion, intragenic *MED13L* deletion) provided an opportunity to unravel the complex relationship between the genotype and phenotype:

- 6.1. Splicing variants at 5' donor splice sites of the *ARID1B*, *GLI3*, and *SLC9A6* gene lead to Coffin-Siris syndrome 1, Greig Cephalopolysyndactyly, Christianson type intellectual disability, respectively, while splicing variants at 3' acceptor splice sites of the *CHD7* and *TGFBR2* causes CHARGE syndrome and Loeys-Dietz syndrome 2, respectively. *CAPN3* c.1746-20C>G splicing variant results in Limb-girdle muscular dystrophia type 2A;
- 6.2. *BLM* c.2308-2A> G splicing variant at 3' acceptor splice site is insufficient to explain the pathogenesis of autosomal recessive Bloom syndrome;
- 6.3. The duplication of 13q31.3, encompassing only *MIR17HG*, results in a new syndrome characterized by features mirroring those of Feingold syndrome 2;
- 6.4. The normal *RASGEF1B* expression level and reduced expression level of *PRKG2* allow to narrow the critical region of 4q21 deletion syndrome to four candidate genes: *HNRNPD*, *HNRNPDL*, *ENOPH1*, and *PRKG2*;
- 6.5. Intragenic *MED13L* deletion results in *MED13L* haploinsufficiency syndrome.

REFERENCES

1. Collins FS, Lander ES, Rogers J, Waterson RH. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931–45. doi:10.1038/nature03001.
2. Feingold EA, Good PJ, Guyer MS, Kamholz S, Liefer L, Wetterstrand K, et al. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* (80-). 2004;306:636–40. doi:10.1126/science.1105136.
3. Tragante V, Moore JH, Asselbergs FW. The ENCODE project and perspectives on pathways. *Genet Epidemiol*. 2014;38:275–80. doi:10.1002/gepi.21802.
4. Gloss BS, Dinger ME. Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Med*. 2018;50:97. doi:10.1038/s12276-018-0087-0.
5. Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–73. doi:10.1038/nature09534.
6. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65. doi:10.1038/nature11632.
7. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980–5. doi:10.1093/nar/gkt1113.
8. ClinVar. ClinVar (NCBI). <https://www.ncbi.nlm.nih.gov/clinvar>. Accessed 27 Apr 2020.
9. Hagberg B, Kyllerman M. Epidemiology of mental retardation—A Swedish survey. *Brain Dev*. 1983;5:441–9. doi:10.1016/S0387-7604(83)80072-2.
10. Association AP, others. *Diagnostic and Statistical Manual of Mental Disorders*. 4th text revision ed. Washington, DC Am Psychiatr Assoc. 2000.
11. Moeschler JB, Shevell M. Comprehensive Evaluation of the Child With Intellectual Disability or Global Developmental Delays. *Pediatrics*. 2014;134:e903–18. doi:10.1542/peds.2014-1839.

12. Preikšaitienė E, Kasnauskienė J, Utkus A, Kučinskas V. Asmenu su intelektine negalia genetinio ištyrimo gairės. *Neurol Semin.* 2012;16:283–8.
13. ClinVar. ClinVar (NCBI). 2019. <https://www.ncbi.nlm.nih.gov/clinvar/>. Accessed 27 Apr 2020.
14. Kim KH, Sederstrom JM. Assaying cell cycle status using flow cytometry. *Curr Protoc Mol Biol.* 2015;2015:28.6.1-28.6.11.
15. Laimer M, Klausegger A, Aberer W, Oender K, Steinhuber M, Lanschuetzer CM, et al. Haploinsufficiency due to deletion within the 3'-UTR of C1-INH-gene associated with hereditary angioedema. *Genet Med.* 2006;8:249–54.
16. Jung E-M, Moffat JJ, Liu J, Dravid SM, Gurumurthy CB, Kim W-Y. *Arid1b* haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat Neurosci.* 2017;20:1694–707. doi:10.1038/s41593-017-0013-0.
17. Hassan AH, Neely KE, Workman JL. Histone acetyltransferase complexes stabilize SWI/SNF binding to promoter nucleosomes. *Cell.* 2001;104:817–27. doi:10.1016/S0092-8674(01)00279-3.
18. Ito S, Kitazawa R, Haraguchi R, Kondo T, Ouchi A, Ueda Y, et al. Novel *GLI3* variant causing overlapped Greig cephalopolysyndactyly syndrome (GCPS) and Pallister-Hall syndrome (PHS) phenotype with agenesis of gallbladder and pancreas. *Diagn Pathol.* 2018;13:1. doi:10.1186/s13000-017-0682-8.
19. Donowitz M, Ming Tse C, Fuster D. *SLC9/NHE* gene family, a plasma membrane and organellar family of Na⁺/H⁺ exchangers. *Mol Aspects Med.* 2013;34:236–51. doi:10.1016/j.mam.2012.05.001.
20. Ouyang Q, Lizarraga SB, Schmidt M, Yang U, Gong J, Ellis D, et al. Christianson Syndrome Protein NHE6 Modulates TrkB Endosomal Signaling Required for Neuronal Circuit Development. *Neuron.* 2013;80:97–112.
21. Daubresse G, Dearing R, Moore L, Papoulas O, Zakrajsek I, Waldrip WR, et al. The *Drosophila* *kismet* gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. *Development.* 1999;126:1175–87. <http://www.ncbi.nlm.nih.gov/pubmed/10021337>.
22. Allen MD, Religa TL, Freund SMV, Bycroft M. Solution

- Structure of the BRK Domains from CHD7. *J Mol Biol.* 2007;371:1135–40. doi:10.1016/j.jmb.2007.06.007.
23. Heldin CH, Moustakas A. Signaling receptors for TGF- β family members. *Cold Spring Harb Perspect Biol.* 2016;8.
 24. Kitano K. Structural mechanisms of human RecQ helicases WRN and BLM. *Front Genet.* 2014;5 OCT. doi:10.3389/fgene.2014.00366.
 25. Lasa-Elgarresta J, Mosqueira-Martín L, Naldaiz-Gastesi N, Sáenz A, López de Munain A, Vallejo-Illarramendi A. Calcium Mechanisms in Limb-Girdle Muscular Dystrophy with CAPN3 Mutations. *Int J Mol Sci.* 2019;20. doi:10.3390/ijms20184548.
 26. Zahir F, Friedman J. The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility. *Clin Genet.* 2007;72:271–87. doi:10.1111/j.1399-0004.2007.00847.x.
 27. Kanna P, Campos-Xavier AB, Hull D, Martinet D, Ballhausen D, Bonafé L. Post-axial polydactyly type A2, overgrowth and autistic traits associated with a chromosome 13q31.3 microduplication encompassing miR-17-92 and GPC5. *Eur J Med Genet.* 2013;56:452–7.
 28. Hemmat M, Rumble MJ, Mahon LW, Strom CM, Anguiano A, Talai M, et al. Short stature, digit anomalies and dysmorphic facial features are associated with the duplication of miR-17 ~ 92 cluster. *Mol Cytogenet.* 2014;7:1–20.
 29. Bonnet C, Andrieux J, Béri-Dexheimer M, Leheup B B, Boute O, Manouvrier S, et al. Microdeletion at chromosome 4q21 defines a new emerging syndrome with marked growth restriction, mental retardation and absent or severely delayed speech. *J Med Genet.* 2010;47:377–84. doi:10.1136/jmg.2009.071902.
 30. Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zobot MT, et al. Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am J Hum Genet.* 2006;79:332–41.
 31. Reymond A, Henrichsen CN, Harewood L, Merla G. Side effects of genome structural changes. *Curr Opin Genet Dev.* 2007;17:381–6. doi:10.1016/j.gde.2007.08.009.
 32. Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, et al. Segmental copy number variation

- shapes tissue transcriptomes. *Nat Genet.* 2009;41:424–9.
33. Ricard G, Molina J, Chrast J, Gu W, Gheldof N, Pradervand S, et al. Phenotypic consequences of copy number variation: Insights from smith-magenis and Potocki-Lupski syndrome mouse models. *PLoS Biol.* 2010;8.
 34. Kircher M, Kelso J. High-throughput DNA sequencing - concepts and limitations. *BioEssays.* 2010;32:524–36. doi:10.1002/bies.200900181.
 35. López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 2005;579:1900–3. doi:10.1016/j.febslet.2005.02.047.
 36. Wang G-S, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet.* 2007;8:749–61. doi:10.1038/nrg2164.
 37. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* 2013;132:1077–130. doi:10.1007/s00439-013-1331-2.
 38. Abramowicz A, Gos M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet.* 2018;59:253–68. doi:10.1007/s13353-018-0444-7.
 39. Krawczak M, Reiss J, Cooper D. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum Genet.* 1992;90:41–54. doi:10.1007/BF00210743.
 40. Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat.* 2007;28:150–8. doi:10.1002/humu.20400.
 41. Simpson MA, Hsu R, Keir LS, Hao J, Sivapalan G, Ernst LM, et al. Mutations in FAM20C are associated with lethal osteosclerotic bone dysplasia (Raine syndrome), highlighting a crucial molecule in bone development. *Am J Hum Genet.* 2007;81:906–12.
 42. Gelincik A, Demir S, Olgaç M, Karaman V, Toksoy G, Çolakoğlu B, et al. Idiopathic angioedema with F12 mutation: is it a new entity? *Ann Allergy, Asthma Immunol.* 2015;114:154–6. doi:10.1016/j.anai.2014.11.018.

43. Eng L, Coutinho G, Nahas S, Yeo G, Tanouye R, Babaei M, et al. Nonclassical Splicing Mutations in the Coding and Noncoding Regions of the ATM Gene: Maximum Entropy Estimates of Splice Junction Strengths. *Hum Mutat.* 2004;23:67–76. doi:10.1002/humu.10295.
44. Moon SD, Park JH, Kim EM, Kim JH, Han JH, Yoo SJ, et al. A novel IVS2-1G> A mutation causes aberrant splicing of the HRPT2 gene in a family with hyperparathyroidism-jaw tumor syndrome. *J Clin Endocrinol Metab.* 2005;90:878–83. doi:10.1210/jc.2004-0991.
45. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural and Molecular Biology.* 2009.
46. Hug N, Longman D, Cáceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 2016;44:1483–95. doi:10.1093/nar/gkw010.
47. Lindeboom RGH, Supek F, Lehner B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet.* 2016.
48. Fusco C, Morlino S, Micale L, Ferraris A, Grammatico P, Castori M. Characterization of Two Novel Intronic Variants Affecting Splicing in FBN1-Related Disorders. *Genes (Basel).* 2019;10:442. doi:10.3390/genes10060442.

CURRICULUM VITAE

Evelina Siavrienė was born on February 11, 1992 in Vilnius, Lithuania. In 2010, she graduated from the the John Paul II Gymnasium. In 2014, she graduated with a bachelor's degree in genetics from Vilnius University, Faculty of Natural Sciences; in 2016, she graduated with a master's degree in medical genetics from Vilnius University, Institute of Biomedical Sciences, Faculty of Medicine, Department of Human and Medical Genetics, and began her PhD studies in medicine under the supervision of Prof. Habil. Dr. Vaidutis Kučinskas. Since 2016, Evelina is working in the Department of Human and Medical Genetics as a junior research associate. From 2017 to 2020, she participated in the research project "Deciphering the genetic architecture of intellectual disability and congenital anomalies (INGENES)," which has been supervised by Assoc. Prof. Dr. Eglė Preišaitienė. A fundamental question in current biomedical research is the study of functional genomics using human cell lines. Therefore, Evelina extended her knowledge and professional skills in cell culturing and functional human genome analysis during her traineeships at the The Cell Factory in Belgium and Center for Integrative Genomics hosted at the University of Lausanne in Switzerland. She is a member of Lithuanian Stem Cell Research Association and Lithuanian Society of Human Genetics.

LIST OF PUBLICATIONS

Published articles:

1. **Siavrienė E.**, Preikšaitienė E., Maldžienė Ž., Mikštienė V., Rančelis T., Ambrozaitytė L., Gueneau L., Reymond A., Kučinskas V. A *de novo* 13q31.3 microduplication encompassing the miR-17 ~ 92 cluster results in features mirroring those associated with Feingold syndrome 2. *Gene*. 2020 May 27;753:144816. IF=2.319.

2. Pranckėnienė L., **Siavrienė E.**, Gueneau L., Preikšaitienė E., Mikštienė V., Reymond A., Kučinskas V. *De novo* splice site variant of *ARID1B* associated with pathogenesis of Coffin-Siris syndrome. *Mol Genet Genomic Med.* 2019 Oct 19:e1006. doi: 10.1002/mgg3.1006. IF=2.448.
3. **Siavrienė E.**, Mikštienė V., Radzevičius D., Maldžienė Ž., Rančelis T., Petraitytė G., Tamulytė G., Kavaliauskienė I., Šarkinas L., Utkus A., Kučinskas V., Preikšaitienė E. Novel *GLI3* variant causes Greig cephalopolysyndactyly syndrome in three generations of a Lithuanian family. *Mol Genet Genomic Med.* 2019 Jul 20:e878. doi: 10.1002/mgg3.878. IF=2.448.
4. **Siavrienė E.**, Petraitytė G., Mikštienė V., Rančelis T., Maldžienė Ž., Morkūnienė A., Byčkova J., Utkus A., Kučinskas V., Preikšaitienė E. A novel *CHD7* variant disrupting acceptor splice site in a patient with mild features of CHARGE syndrome: a case report. *BMC Med Genet.* 2019 Jul 17;20(1):127. doi: 10.1186/s12881-019-0859-y. IF=2.051.
5. **Gasperskaja (Siavrienė) E.**, Kučinskas V. The most common technologies and tools for functional genome analysis. *Acta Med Litu.* 2017; 24 (1): 1-11. doi: 10.6001/actamedica.v24i1.3457.

Oral presentations:

1. **Siavrienė E.**, Preikšaitienė E., Kučinskas V. Su intelektine negalia ir įgimtomis anomalijomis siejamų genų variantų funkcinė analizė. LSA Conference of Young Scientists “Bioateitis.” Diploma for Best Oral presentation, 1st place. 11 12 2019, Kaunas, Lithuania.
2. **Siavrienė E.** iRNR splaisingo variantų, susijusių su intelektine negalia ir įgimtomis anomalijomis, funkcinė analizė. LSA Conference of Young Scientists “Bioateitis.” 14 12 2018, Vilnius, Lithuania.

3. **Siavrienė E.**, Preikšaitienė E., Kučinskas V. Functional genome analysis aiming to define the genetic cause of intellectual disability. 56th Polish and 14th International Conference “Juvenes Pro Medicina.” Diploma for Best Oral presentation, 2nd place. 25–26 05 2018, Lodz, Poland. ISBN 978-83-947627-1-1.
4. **Siavrienė E.**, Preikšaitienė E., Kučinskas V. Functional Genome Analysis in Patients with Intellectual Disability. XIV Baltic Congress of Laboratory Medicine. 10–12 05 2018, Vilnius, Lithuania.

Posters:

1. **Siavrienė E.**, Mikštienė V., Radzevičius D., Maldžienė Ž., Rančelis T., Petraitytė G., Kavaliauskienė I., Šarkinas L., Utkus A., Kučinskas V., Preikšaitienė E. Functional analysis of a novel splice site variant in *GLI3* gene causing Greig cephalopolysyndactyly syndrome. Conference of American Society of Human Genetics, 15–19 10 2019, Houston, USA.
2. **Siavrienė E.**, Mikštienė V., Maldžienė Ž., Petraitytė G., Rančelis T., Utkus A., Preikšaitienė E., Kučinskas V. Functional delineation of de novo heterozygous intragenic deletion in *MED13L*. Conference of European Society of Human Genetics, 15-18 06 2019, Gothenburg, Sweden.
3. Petraitytė G., **Siavrienė E.**, Mikštienė V., Maldžienė Ž., Rančelis T., Utkus A., Preikšaitienė E., Kučinskas V. Functional analysis of a novel c.899+1G>A variant in *SLC9A6* gene. Conference of European Society of Human Genetics, 15–18 06 2019, Gothenburg, Sweden.
4. **Siavrienė E.**, Petraitytė G., Mikštienė V., Maldžienė Ž., Rančelis T., Preikšaitienė E., Kučinskas V. Novel c.5535-1G>A variant in a patient with a mild features of CHARGE syndrome. Conference of American Society of Human Genetics, 16–20 10 2018, San Diego, USA.

5. **Siavrienė E.**, Preikšaitienė E., Maldžienė Ž., Ambrozaitytė L., Gueneau L., Reymond A., Kučinskas V. Microduplication of the 13q31.3 miR17~92 cluster results in a syndrome with features opposite to those associated with Feingold syndrome 2. Conference of European Society of Human Genetics, 16–19 06 2018, Milan, Italy.

E-poster:

1. Petraitytė G., **Siavrienė E.**, Mikštienė V., Norvilas R., Burnytė B., Utkus A., Preikšaitienė E. Analysis of the effect of a novel splice site variant in *TGFBR2* on mRNA level. Virtual Conference of European Society of Human Genetics, 6–9 06 2020.

Traineeships:

1. Center for Integrative Genomics hosted at the University of Lausanne, 08 01 2018–06 02 2018, Lausanne, Switzerland.
2. The Cell Factory, Department of Research and Development, 01 02 2016–31 04 2016, Niel, Belgium.

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
Print run copies 30