

Identification of genetic causes of congenital neurodevelopmental disorders using genome wide molecular technologies

Eglė Preikšaitienė,

Laima Ambrozaitytė,

Živilė Maldžienė,

Aušra Morkūnienė,

Loreta Cimbalistienė,

Tautvydas Rančelis,

Algirdas Utkus,

Vaidutis Kučinskas*

*Department of Human
and Medical Genetics,
Faculty of Medicine,
Vilnius University,
Vilnius, Lithuania*

Background. Intellectual disability affects about 1–2% of the general population worldwide, and this is the leading socio-economic problem of health care. The evaluation of the genetic causes of intellectual disability is challenging because these conditions are genetically heterogeneous with many different genetic alterations resulting in clinically indistinguishable phenotypes. Genome wide molecular technologies are effective in a research setting for establishing the new genetic basis of a disease. We describe the first Lithuanian experience in genome-wide CNV detection and whole exome sequencing, presenting the results obtained in the research project UNIGENE.

Materials and methods. The patients with developmental delay/intellectual disability have been investigated ($n = 66$). Diagnostic screening was performed using array-CGH technology. FISH and real time-PCR were used for the confirmation of gene-dose imbalances and investigation of parental samples. Whole exome sequencing using the next generation high throughput NGS technique was used to sequence the samples of 12 selected families.

Results. 14 out of 66 patients had pathogenic copy number variants, and one patient had novel likely pathogenic aberration (microdeletion at 4p15.2). Twelve families have been processed for whole exome sequencing. Two identified sequence variants could be classified as pathogenic (in *MECP2*, *CREBBP* genes). The other families had several candidate intellectual disability gene variants that are of unclear clinical significance and must be further investigated for possible effect on the molecular pathways of intellectual disability.

Conclusions. The genetic heterogeneity of intellectual disability requires genome wide approaches, including detection of chromosomal aberrations by chromosomal microarrays and whole exome sequencing capable of uncovering single gene mutations. This study demonstrates the benefits and challenges that accompany the use of genome wide molecular technologies and provides genotype-phenotype information on 32 patients with chromosomal imbalances and ID candidate sequence variants.

Keywords: intellectual disability, developmental delay, array-CGH, whole exome sequencing

* Principal investigator.

Correspondence to: Eglė Preikšaitienė, Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, 2 Santariškių Street, LT-08661 Vilnius, Lithuania. E-mail: egle.preiksaitiene@mf.vu.lt

INTRODUCTION

Intellectual disability affects about 1–2% of the general population worldwide, and this is the leading socio-economic problem of health care (1). It is a broad diagnosis encompassing a wide variety of overlapping phenotypes and severities. It can range from mild to profound and can occur as an isolated phenotype or can be associated with other clinical symptoms as part of a syndrome.

Despite the technical progress that has been achieved over the last few years, the evaluation of the genetic causes of intellectual disability remains challenging because these conditions are genetically heterogeneous with many different genetic alterations resulting in clinically indistinguishable phenotypes. It is known that non-genetic factors can cause intellectual disability in about 1/3 of patients (2). The remaining are mostly genetic causes. During a routine cytogenetic investigation it is possible to identify chromosomal alterations, most common of which is trisomy 21. The subtelomeric MLPA or FISH analysis may aid diagnosis in 1–2% of patients (3). The diagnosis of clinically recognizable syndromes can be confirmed using FISH/MLPA in patients with submicroscopic chromosomal alterations, or the Sanger sequencing and other molecular genetic tests, when intellectual disability is a part of a clinically recognizable monogenic disorder. In more than 60% of patients the genetic diagnosis remains unknown using this approach of genetic investigation, because many known genetic disorders are not always easily diagnosed or their genetic basis is not known yet (4). However, the advent of the technologies for high resolution whole genome analysis into medical practice has simplified the estimation of genetic diagnosis.

Genome wide molecular technologies are effective in a research setting for establishing the new genetic basis of a disease and are significantly accelerating the pace of discovery of novel Mendelian disease genes. About 100 genes are described as implicated in X-linked intellectual disability, and more associated to autosomal-recessive or autosomal-dominant forms (5). There are more than 500 genes proposed to cause intellectual disability with high penetrance when mutated (6). The molecular basis of even more named Mendelian or suspected Mendelian disorders is unknown and currently awaits elucidation.

Herein, we describe the first Lithuanian experience in genome-wide CNV detection of copy number variants (CNVs) and whole exome sequencing, presenting the results obtained in the course of the Lithuanian–Swiss cooperation programme “Research and Development”, the joint research project “Unique Genome Variants in Congenital Neurodevelopmental Disorders: Origin, Genomic Mechanisms, Functional and Clinical Consequences” (UNIGENE). This study provides detailed information of the structural variation affecting many base pairs, resulting in the copy number variation, and the DNA sequence variation in the form of DNA base-pair substitutions and short indels in a cohort of Lithuanian patients with the unknown cause of developmental delay or intellectual disability (ID/DD). We also discuss the challenges that accompany the use of high resolution array-CGH and the whole exome sequencing technologies.

MATERIALS AND METHODS

Patients

Patients with both syndromic and non-syndromic DD/ID were included in the study ($n = 66$). The DD/ID levels of patients ranged from mild to profound. Previously performed conventional karyotyping, targeted FISH, molecular tests, and investigations for metabolic disorders revealed no causative anomaly. The detailed clinical history and data of physical examinations of all patients were reviewed. DNA from the patients' blood samples was isolated using the phenol-chloroform extraction method (The protocol is available on request). Informed consents for genetic investigations were obtained from the participating families. The study was approved by the Vilnius Regional Biomedical Research Ethics Committee.

Array-CGH

Diagnostic screening was performed using array-CGH with 60K, AMADID 031746 ($n = 32$), 105K, AMADID 031750 ($n = 4$) and 1M ($n = 34$, 4 of them after 60K) oligo chips (Agilent Technologies, USA) according to the recommendations of the manufacturer. Patient and reference DNA were digested by AluI and RsaI enzymes and labelled by random priming using an Agilent Genomic

DNA labeling kit (Agilent Technologies). Patient and reference DNA were labelled with Cy5 and Cy3, respectively, and were co-hybridised to arrays for 24 h at 67 °C in a rotating oven (Agilent Technologies) at 20 rpm. The arrays were washed and scanned with an Agilent SureScan Microarray Scanner (model G4900DA, Agilent Technologies). Data were extracted using the Feature Extraction 10.7.3.1 software (Agilent Technologies) and analysed using the CytoGenomics 3.0.0.27 software (Agilent Technologies). Genomic copy number changes were identified with the assistance of the Aberration Detection Method 2 algorithm with a threshold of 6. The array data was analysed according to the annotation GRCh37/hg19. Copy number changes identified in the sample were compared with the dataset of previous array-CGH/SNP-CGH results in the Lithuanian population and the copy number variants (CNVs) from the Database of Genomic Variants (<http://projects.tcag.ca/variation/>) to exclude the specific variants of the Lithuanian population or the previously reported polymorphisms. The non-polymorphic CNVs were compared with the entries in DECIPHER (<http://decipher.sanger.ac.uk/>), ECARUCA (www.ecaruca.net), ISCA (<http://www.iccg.org/>) or ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>) databases, evaluated against the existing literature for the known syndromes and overlapping causal aberrations and were further analysed according to the type and size of aberration, the function, and expression profile of genes.

FISH and real time-PCR

FISH and real time-PCR were used for the confirmation of gene-dose imbalances and investigation of parental samples. FISH analysis was performed on interphase/metaphase cell slides prepared from peripheral blood lymphocytes obtained by standard protocols. Custom FISH probes that are located in the duplicated or deleted region were used. The FISH slides were pretreated with $2 \times$ SSC for 2 minutes at 72 °C, 0.5 mg/ml pepsin in 0.01 HCl for 5 minutes at 37 °C, and 10% phosphate buffered formalin for 5 minutes at room temperature (RT), followed by sequential dehydration in 70, 85, and 96% ethanol for 1 minute at RT. The FISH slides and labelled probes were denatured separately for 5 minutes at

72 °C, followed by sequential dehydration in 70, 85, and 96% ethanol for 1 minute at RT. The denatured probes were added on the denatured FISH slides and co-hybridized overnight at 37 °C. After the hybridization, the slides were washed in $0.4 \times$ SSC/0.3% NP40 for 2 minutes at 72 °C and $2 \times$ SSC/0.1% NP40 for 1 minute at RT. The slides were counterstained with DAPI II (Abbot Molecular, USA), and the signals were analysed using the CytoVision system (Applied Imaging, UK).

Real time-PCR using two different pairs of primers within the deleted or duplicated region and one pair outside the region was performed. Primers were designed using the Primer3 programme (7). The realtime PCR amplifications were carried out in optical grade 96-well plates (Applied Biosystems, USA) in a 25 μ l reaction volume containing 12.5 μ l SYBR GREEN qPCR Master Mix (Applied Biosystems, USA), 2.5 μ M of each primer and 16 ng of DNA. All samples were analysed in duplicate for both reference and test fragment. In each assay normal control and no-template control were included. PCR reactions were performed on the 7900 HT Fast Real-Time PCR System (Applied Biosystems, USA). Fluorescence intensity was measured once per cycle at the end of the elongation phase. Amplification was followed by the melting curve analysis performed according to the dissociation stage data. Reactions with a single peak at expected T_m were considered for further analysis. The data analysis was performed using the SDS v2.3 software (Applied Biosystems, USA). Gene dosage ratios for each fragment were calculated using the $2^{-\Delta\Delta CT}$ method.

Whole exome sequencing (WES)

Whole exome sequencing using a next generation 5500 SOLiD™ system (Applied Biosystems, Thermo Fisher Scientific, USA) after the in-solution capture enrichment was used to sequence the samples. The 5500 SOLiD™ system sequencing strategy is ligation based and uses colour-space encoding. The sequencing protocol was carried out according to the optimized manufacturer's protocols (Life Technologies, Thermo Fisher Scientific, USA) using Agilent SureSelectXT Target Enrichment System pooling six samples on one Flowchip. 2 ng of genomic DNA was used for Fragment Library preparation. The shearing of the genomic DNA was performed by the Covaris S220 system with a mean fragment size of 160 bp

and a fragment size range of 100–250 bp. The sequencing fragment length on the 5500 SOLiD™ system is 75 bp. An average of 30x coverage was achieved.

The sequence alignment and primary data analysis was performed using the LifeScope™ Genomic Analysis Software v2.5.1. Several different mappings in separate .bam files were generated, compared and analysed further by a variant calling pipeline. On the average 30 to 50K genomic variants have been identified for each individual exome. GATK (Genome Analysis Toolkit), a widely accessed sequencing data analysis toolkit, was used for the identification and filtration of genomic variants. The exome data annotation was applied for each individual exome. For annotation .vcf files generated by the LifeScope™ Genomic Analysis Software v2.5.1 (Applied Biosystems, Thermo Fisher Scientific, USA) were used. The Annovar v.2015mar22 programme that combines SIFT v5, PolyPhen v2.2.2, MutationTaster v2, MutationAssessor v2, LRT, FATHMM v2.3, GERP++ v2, SiPhy v0.5, PhyloP v1 CADD v1.1, PhastCons algorithmic annotation tools data, ge-

nome variants frequency data from 1000 genome and 6500 exome projects data, as well as information from the already known pathogenic variants databases, COSMIC v70 ir ClinVar (30.03.2015), was used for annotation (8). Due to annotation tools filtered genome variants were checked for associations with the already known (published) brain malformations in the DisGeNET database, using Human–Mouse: Disease Connection (HMDC) and Phenolyzer tools. The final list of intellectual disability candidate genomic variants have been run through the genome visualisation programme Integrative Genomics Viewer (IGV) (<https://www.broadinstitute.org/igv/>), reanalysing .BAM files and filtering out incorrectly identified variants that have been included due to errors of variant calling or not-identified DNA fragment duplications (9).

RESULTS

Table 1 summarizes the array-CGH findings and clinical data of investigated patients. In 21 out of 66 patients CNVs not previously reported in

Table 1. Genomic and clinical data of patients with pathogenic, likely pathogenic array-CGH findings and CNVs of uncertain clinical significance

No. Family – patient	Nomenclature according to ISCN 2013 (NCBI build GRCh37/hg 19)	Size, Mb	Age, sex, phenotype	Syndrome/ references
Known microdeletion syndromes				
1_1	arr[hg19] 19p13.3 (739,176-1,258,185)x1 <i>dn</i>	0.5	26 years, ♂. ID, freckling of the lips and perioral region, dysmorphic features, kyphoscoliosis, broad feet, pes equinus, congenital enlargement of kidneys, axonopathy, spastic paraparesis.	<i>Peutz–Jeghers s.</i>
2_1	arr[hg19] 17p13.3p13.2 (2,414,600-3,587,961)x1 <i>dn</i>	1.2	1 year, ♀. DD, epilepsy, lissencephaly, microcephaly, minor facial anomalies.	<i>Miller–Dieker s.</i>
3_1	arr[hg19] 9q34.3(140,503,430-141,008,915)x1, arr[hg19] 19q13.43 (57,724,877-59,030,051)x3 <i>pat</i>	0.5 1.3	35 years, ♀. ID, microcephaly, obesity, mandibular prognathia, synophrys, diabetes mellitus.	<i>Kleefstra s.</i>
4_1	arr[hg19] 4p16.2(1-4,370,244)x1 <i>dn</i>	4.4	11 months, ♀. DD, prenatal and postnatal growth retardation, muscular hypotonia, seizures, minor facial anomalies.	<i>Wolf–Hirschhorn s.</i>
5_1	arr[hg19] 15q13.2q13.3 (31,014,508-32,438,943)x1 <i>mat</i>	1.4	10 years, ♀. ID, muscular hypotonia, obesity, minor facial anomalies.	15q13 microdeletion syndrome

Table 1. (continued)

Known pathogenic copy number variants				
6_1	arr[hg19] 6q16.1q22.31 (96,076,878-123,209,593)x1 <i>dn</i>	27.1	2 years, ♀. DD, muscular hypotonia, epilepsy, hydrocephalus, microcephaly, congenital heart defects, short stature, strabismus, minor facial anomalies.	Rosenfeld JA et al. 2012 (10)
7_1	arr[hg19] 18q21.2q23 (54,473,192-77,982,126)x1 <i>dn</i>	23.5	14 months, ♀. DD, minor facial anomalies, congenital heart disease, thickened nails.	Linnankivi T et al. 2006 (11)
8_1	arr[hg19] 10p15.3p15.1(1-4,427,945)x3, 10q26.12q26.3 (121,938,683-135,534,747)x1 <i>mat</i>	4,4 13.6	13 years, ♀. ID, short stature, microcephaly, minor facial anomalies, transversal palmar crease.	Ciuladaite Z et al. 2014 (12)
9_1	arr[hg19] 8p11.22q11.23 (39,416,498-53,184,507)x3 <i>dn</i>	13.8	5 years, ♂. DD, short stature, hypoplasia of the corpus callosum.	Eyüpoğlu FC et al. 2014 (13)
10_1	arr[hg19] 10q22.1q22.3 (74,236,933-79,422,266)x1 <i>dn</i>	5.2	4 years, ♀. DD, muscular hypotonia, minor facial anomalies.	Tzschach A et al. 2010 (14)
11_1	arr[hg19] 6q21q22.33(113,381,924- 129,540,205)x1 <i>dn</i>	16.2	15 years, ♂. ID disability, dolichostenomelia, microcephaly, minor facial anomalies, pterigium colli, cryptorchidism.	Rosenfeld JA et al. 2012 (10)
12_1	arr[hg19] Xp11.23p11.21 (48,791,150-58,051,765)x3 <i>dn</i>	9.3	15 years, ♀. ID, obesity, hyperopic astigmatism, minor facial anomalies.	Nizon M. 2015 (15)
13_1	arr[hg19] 3q26.1q29 (166,659,726-197,803,820)x3, arr[hg19] 5p15.33p13.3 (1-33,683,173)x1 <i>mat</i>	31.1 33.7	1 month, ♂. DD, muscular hypotonia, short stature, minor facial anomalies, spina bifida, long thumbs, micropenis.	Rossi M et al. 2002 (16)
14_1	arr[hg19] 13q31.3 (91,166,748-92,010,901)x3 <i>dn</i>	0.8	19 years, ♀. ID, Arnold-Chiari deformation, strabismus, nystagmus, tall stature, facial asymmetry, arachnodactyly, and rough hair.	Hemmat M et al. 2014 (17)
Novel likely pathogenic copy number variant genes				
15_1	arr[hg19] 4p15.2(22,324,464- 24,776,373)x1 <i>dn</i>	2.5	4 years, ♀. DD, dysgenesis of the right iris, strabismus, thin upper lip, dens invaginatus, redundant periumbilical skin, sacral dimple.	<i>GPR125, GBA3, PPARGC1A, MIR573, DHX15</i>
CNVs of uncertain clinical significance				
16_1	arr[hg19] 10q21.3 (64,638,002-65,385,071)x1 <i>mat</i>	0.8	3 years, ♂. DD, atrial septal defect, Duane anomaly, strabismus.	<i>NRBF2, JMJD1C, MIR1296, LOC84989, REEP3</i>
17_1	arr[hg19] 12p13.31 (6,409,765-6,442,276)x3 <i>pat</i>	0.03	18 months, ♂. DD, megalencephaly.	<i>PLEKHG6, TNFRSF1A</i>
18_1	arr[hg19] 7q21.2 (91,615,893-91,689,470)x1 <i>pat</i>	0.3	15 years, ♂. ID, sleep disturbance, behaviour problems, cerebellar hypoplasia, cerebellar vermis partial dysplasia, enlarged cerebral ventricles, short stature, dolichocephaly.	<i>AKAP9</i>
19_1	arr[hg19] 4q22.1 (93,236,129-93,459,373)x1 <i>pat</i>	0.2	4 years, ♂. DD, stereotypic movements, regression, autism.	<i>GRID2</i>

Table 1. (continued)

No. Family _patient	Nomenclature according to ISCN 2013 (NCBI build GRCh37/hg 19)	Size, Mb	Age, sex, phenotype	Syndrome/ references
20_1	arr[hg19] 14q23.3 (67,177,677-67,204,920)×3	0.03	3 months, ♂. DD, hypotony, seizures, thin corpus callosum, mild hepatomegaly, hypoplasia of optic nerves, short stature, umbilical hernia.	<i>GPHN</i>
21_1	arr[hg19] 10q23.33 (95,089,475-95,661,691)×3, 10q24.1(97,548 459-98,011,349)×3 <i>mat</i>	0.6 0.5	3 years, ♂. DD, microcephaly, ocular hyper- telorism.	<i>MYOF, CEP55, O3FAR1, RBP4, PDE6C, FRA10AC1, LGII, TMEM20 ENTPDI, C10orf131, CC2D2B, CCNJ, MIR3157, ZNF518A, BLNK</i>

the Database of Genomic Variants have been detected. 14 patients of them had causal aberrations, 1 patient had novel likely pathogenic CNV and in 6 patients the clinical significance of detected rare CNVs was uncertain. Each CNV was confirmed either by the FISH or by real time-PCR followed by the parental analysis using the same molecular method. 12 CNVs were *de novo*, 9 were familial, and the inheritance of 2 aberrations was uncertain. The causal aberrations involved 3 duplications, 8 deletions, and 3 chromosomal rearrangements with partial deletion and partial duplication. The size of pathogenic CNVs varied between 0.5 and 33.7 Mb. Six patients with the CNVs of uncertain clinical significance had 3 deletions and 4 duplications. The size of uncertain clinical CNVs varied between 27 and 747 kb.

The results of whole exome sequencing are presented in Table 2. Ten trios of a proband and both parents, and two tetrads with two affected siblings and both parents have been processed for WES. One proband with the pathogenic variant of *MECP2* gene has been confirmed with the Rett syndrome diagnosis that complies with the clinical features, whereas the affected sibling has not a single potential candidate ID variant. One proband has a possible molecular confirmation of the Rubinstein–Taybi syndrome with a variant in the *CREB-BP* gene. The other families have several candidate ID gene variants each (3 to 9) that are of unclear

clinical significance and must be further investigated for pathogenic effect and possible effect on the ID molecular pathways.

DISCUSSION

Genomic disorders compose a family of genetic diseases that are characterized by large genomic alterations, mostly deletions and duplications. They result in the loss or gain of genomic segments (CNVs), which can contain multiple genes. The phenotypic features of the disorders associated with CNVs are thought to be related to altered gene dosage effects in most patients (18). It is important to know which of the genes located within the CNV are associated with which of the clinical features. During our study one of the challenges was to identify potential candidate genes through searches of the literature and online resources.

The terminal 10p15.3 deletion identified in one of the investigated patients (patient 8_1) comprises a clinically recognizable syndrome with an invariable neurodevelopmental disorder (19). The minimal critical region of deletions overlap pointed to two genes, *ZMYND11* and *DIP2C*, as the main candidate genes determining clinical features of the patients. Recently, few patients with a *de novo* mutation in *ZMYND11* were described (20). As the phenotype of patients with loss-of-function *ZMYND11* mutations closely resembled

Table 2. Genomic and clinical data of patients with candidate pathogenic sequence variants

No. Family_patient	Gene	Chromosomal position	Nucleotide	dbSNP (snp142)	Age, sex, phenotype	Syndrome/references
22_1	<i>HUWE1</i>	X:53627159	T>C	rs145758265	18 years, ♂. ID, behaviour problems, poor coordination, minor facial anomalies, arachnodactyly, abnormality of the aortic valve.	Isrie M et al. (29)
	<i>OCRL</i>	X:128674722	C>T	rs61752970		
	<i>PRDX2</i>	19:12911055	C>T	NA		
23_1	<i>NTRK1</i>	1:156848946	G>T	rs6339	20 years, ♂. ID, epilepsy, behaviour problems, muscular hypotonia, poor coordination, microcephaly, minor facial anomalies.	Robertson SP (30)
	<i>FLNA</i>	X:153593579	C>T	NA		
24_1	<i>CDKL5</i>	X:18638082	A>C		10 years, ♂. ID, stereotypy, asymmetric ventricles, minor facial anomalies.	<i>Rubinstein–Taybi syndrome</i>
	<i>CREBBP</i>	16:3786782	T>G			
	<i>LHX5</i>	12:113906140	G>A			
25_1	<i>SLC9A6</i>	X:135084373	G>A		6 years, ♂. ID, seizures, stereotypy, short stature, microcephaly, hypermetropia, minor facial anomalies.	Gilillan GD et al. (31), Garbern JY et al. (32) Kutsche K et al. (33), Ramakers GJA et al. (34)
	<i>ARHGEF6</i>	X:135862972	T>C			
25_2	<i>SLC9A6</i>	X:135084373	G>A		10 years, ♂. ID, seizures, stereotypy, hypermetropia, minor facial anomalies.	Gilillan GD et al. (31), Garbern JY et al. (32) Kutsche K et al. (33), Ramakers GJA et al. (34)
	<i>ARHGEF6</i>	X:135862972	T>C			
26_1	<i>PLCG2</i>	16:81922813	C>T	rs1143687	3 years, ♂. DD, epilepsy, pachygyria, polymicrogyria, minor facial anomalies, tall stature.	Freeze HH et al. (35)
	<i>ALG6</i>	1:63902524	C>G	rs41285372		
27_1	<i>DSE</i>	6:116720514	C>T		19 years, ♂. ID, behaviour problems, epilepsy, multiple demyelination in CNS, retinal nonattachment, cataract, myopia, minor facial anomalies.	
	<i>KIAA0586</i>	14:59014563	G>A			
	<i>TMEM8B</i>	9:35853517	C>T			
	<i>CC2D1A</i>	19:14040896	G>A			
	<i>ZMAT2</i>	5:140081674	A>G			
28_1	<i>MECP2</i>	X:153296516	G>A	rs61749721	3 years, ♀. DD, developmental regression, epilepsy, hyperactivity, microcephaly, stereotypy, muscular hypotonia.	<i>Rett syndrome</i>

Table 2. (continued)

No. Family_patient	Gene	Chromosomal position	Nucleotide	dbSNP (snp142)	Age, sex, phenotype	Syndrome/references
28_2	No candidate variants				5 years, ♀. DD, short stature, microcephaly, transverse palmar crease, café-au-lait spots.	
29_1	CACNA1A	19:13372340	C>T		9 years, ♀. ID, seizures, stereotypy, minor facial anomalies.	Guerin AA et al. (36), Blumkin L et al. (37), Damaj L et al. (38)
30_1	HSPD1	2:198363406	C>T	rs200514123	7 years, ♀. ID, muscular hypotonia, minor facial anomalies.	
	PSPH	7:56087423	G>A	rs79451216		
	SGSH	17:78184679	C>T	rs9894254		
	RAB-3GAP1	2:135887597; 2:135893372	C>T; A>G	rs15047834; rs10445686		
	FGF10	5:44305104; 5:44388783	T>G; A>C	rs14771550; NA		
	CASR	3:122003757	G>T	rs1801725		
	MBL2	10:54531235	C>T	rs1800450		
30_2	ALDH4A1	1:19201928; 1:19199448	C>T; G>T	rs2230709; rs61757683	8 years, ♀. ID, muscular hypotonia, hypopigmented spots, minor facial anomalies, strabismus, arachnodactyly.	
	FZD9	7:72849221	A>C	rs73134914		
	VPS13B	8:100654723; 8:100847866	A>G; C>T	rs13964022; NA		
	MED12	X:70345554	A>G	NA		
30_3	ZNF81	X:47775540	A>G		18 years, ♂. ID, behaviour problems, cryptorchidism, minor facial anomalies, kyphosis, pterygium colli.	
	MED12	X:70351463	C>T			
30_4	CC2D1A	19:14038791	C>T	rs2305777	10 years, ♂. ID, macrocephaly, obesity, minor facial anomalies.	
	ORC1	1:52838992	A>G	rs34521609		
	ARSH	X:2933233	G>A	rs142811205		
	L1CAM	X:153137616	T>C	NA		
	ANK3	10:61802477; 10:61828954	C>T; G>A	rs14193931; rs7923682		

the 10p15.3 microdeletion syndrome, the contribution of *ZMYND11* haploinsufficiency to the phenotype of patients with the 10p15.3 microdeletion syndrome is most reliable.

Similarly, we show a candidate gene for blepharophimosis and other clinical features of a patient with the 10q22.1q22.3 microdeletion (patient 10_1). Besides blepharophimosis, the patient had global

developmental delay, hypotony, and facial minor anomalies, including ptosis, epicanthus inversus, and microtia. The deletion involves 44 *RefSeq* genes. The overlapping region of reported deletions contains 10 genes, and one of them is *KAT6B* gene. The dominant mutations of the *KAT6B* gene have been detected by whole-exome sequencing in patients with the Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS) (21). The clinical overlap in the *KAT6B* gene mutations causing SBBYSS and 10q22 deletions confirms the hypothesis that haploinsufficiency of the *KAT6B* gene is associated with a specific phenotype.

However, many patients have genomic alterations that are described in only few individuals and the evaluation of clinical significance of these chromosomal alterations is very challenging. Nevertheless, detection of novel pathogenic copy number variants and delineation of novel genomic disorders is an important part of genomic studies. A girl with developmental delay, clinical features of the Axenfeld–Rieger syndrome, and the *de novo* pericentric inversion of chromosome 4 was included in the study for more detailed investigation (patient 15_1). The FISH analysis confirmed that the breakpoint at 4q25 interrupted the gene or its regulatory elements and is a cause of the clinical features of the Axenfeld–Rieger syndrome. The array-CGH analysis revealed a 2.45 Mb deletion at the short arm of chromosome 4. Few possibly pathogenic microdeletions have been reported in the databases, which were associated with developmental delay. We define a common region, about 800 kb in size, which encompasses two genes: *PPARGC1A* encoded protein is a transcriptional coactivator that regulates the genes involved in energy metabolism; *DHX15* encoded protein is a putative ATP-dependent RNA helicase implicated in pre-mRNA splicing which is highly expressed in brain (22–23). Haploinsufficiency of these genes could contribute to the manifestation of developmental delay in the proband.

The pathogenic aberration detection rates were different between different resolutions of microarray platforms in our study. Higher causal CNVs detection rate was obtained with lower resolution microarrays (60, 105K). These differences can be explained by the fact that patients with syndromic and more severe DD/ID have been selected for a lower resolution microarray. As expected, the applica-

tion of high resolution (1M) microarrays revealed significantly more variation of uncertain clinical significance (VOUS). Some of VOUS involved genes with potential impact on the function of central nervous system: *JMJD1C* (21), *GRID2* (22), *GPHN* (23). Interpretation of the clinical significance of small rare CNVs was difficult because most of them (6 out of 7) were inherited from a healthy parent and 2 out of 7 VOUS encompassed only an intron of the gene. More studies are required to assess the real significance of CNVs with uncertain clinical relevance.

The application of sequence variation detection methods (i. e. whole exome or whole genome sequencing) for the identification of molecular ID causes is a powerful tool for the diagnostic evaluation of patients with highly heterogeneous DD/ID phenotypes (27–28). The introduction of these next generation sequencing methods has advanced the detection of new ID genes though the confirmation of the candidate genomic loci still remains complicated and requires many more research studies and functional studies of the genomic variants.

The missense variant of the *HUWE1* gene detected in this study (patient 22_1) supports its role in the ID cases. *HUWE1* encodes an E3 ubiquitin ligase, which is involved in cancer as well as neuronal development. An essential role of *HUWE1* in the development of the cerebellum and neurogenesis is demonstrated by studying the targeted *Huwe1* inactivation in the mouse brain (29).

Null mutations in *FLNA*, the gene that encodes filamin A, lead to defects in neuronal migration, vascular function, and connective tissue integrity. By contrast, missense mutations in this same gene produce a spectrum of malformations in multiple organ systems, especially the skeleton (30). *NTRK1* is a protein coding gene, receptor of Neurotrophic Tyrosine Kinase. Diseases associated with *NTRK1* include congenital insensitivity to pain with anhidrosis (CIPA). CIPA is caused by homozygous or compound heterozygous mutation in the *NTRK1* gene on chromosome 1q23 and is characterized by insensitivity to pain, anhidrosis (the inability to sweat), and intellectual disability. By bioinformatics tools filtered genome variants in both genes (patient 23_1) might contribute to the clinical features of the patient though further investigation must be applied to confirm the hypothesis.

Mutations in *SLC9A6* cause the X-linked ID leading to a syndrome associated with microcephaly, seizures, ataxia, and absent speech (32–33). Some patients, particularly those who were in a younger age group, displayed symptoms resembling those of the Angelman syndrome. Thus, males with the findings suggestive of an unexplained Angelman syndrome should be considered as potential candidates for *SLC9A6* mutations (32). Mutations in the *ARHGEF6* gene, encoding the guanine nucleotide exchange factor aPIX/Cool-2 for the Rho GTPases Rac1 and Cdc42, cause X-linked ID (33–34). Imbalance in the activity of different Rho GTPases may underlie altered neuronal connectivity and impaired synaptic function and cognition in aPix/Arhgef6 knockout mice (34). Both *SLC9A6* and *ARHGEF6* gene variants (patients 25_1, 25_2) might be contributors to the specific clinical phenotype of the two affected males in the family of this study.

One of the clinical features in the broad spectrum of inborn errors of metabolism is neurodevelopmental delay. ALG6-CDG is the second most common N-linked congenital disorder of glycosylation. Clinical features include developmental delay, axial hypotonia, strabismus, and seizures (35). Most patients are compound heterozygotes. Carriers of the mutations present no clinical symptoms. *PLCG2* (*Phospholipase C, Gamma 2 (Phosphatidylinositol-Specific)*) is a protein coding gene. Diseases associated with *PLCG2* include familial cold auto-inflammatory syndrome 3 and autoinflammation, antibody deficiency, and immune dysregulation syndrome. Despite the relatively high variant prioritisation scores, a pathogenic variant might be missed by the bioinformatic filters. More in-depth sequencing experiments and periodic re-evaluation of data in the databases and scientific literature is of high importance.

The *CACNA1A* gene encodes the pore forming alpha-1A subunit of neuronal voltage-dependent P/Q-type Ca(2+) channels. Mutations in this gene result in clinical heterogeneity, described in familial hemiplegic migraine, episodic ataxia type 2, and spinocerebellar ataxia type 6 (36–37). *CACNA1A* loss-of-function mutations are of the phenotypic heterogeneity, and the cognitive and epileptic manifestations caused by the loss of Ca_v2.1 channels function should be stressed, presumably affecting cerebellar, cortical and limbic networks (38).

The complex of the clinical features of the patient might be explained by the *de novo* mutation of the *CACNA1A* gene (patients 29_1).

In this UNIGENE study, out of 12 whole-exome sequenced families 5 identified ID candidate sequence variants can be classified as pathogenic or most likely pathogenic. The most likely reasons for not identifying the causal ID gene variants for many ID families are the following: lack of sequence coverage of the variant, bioinformatics variant calling issues, and misinterpretation of variants. Moreover, it may be that the cause of the disease is located outside the coding sequences or being a large indel or structural genomic variant missed by exome sequencing. In addition, for some projects too many candidate variants remained after filtering and no independent recurrence or functional proof has been obtained so far (39).

CONCLUSIONS

The genetic heterogeneity of intellectual disability requires genome wide approaches, including detection of chromosomal aberrations by chromosomal microarrays and whole exome sequencing capable of uncovering single gene mutations. This study demonstrates the benefits and challenges that accompany the use of genome wide molecular technologies and provides genotype-phenotype information on 32 patients with chromosomal imbalances and ID candidate sequence variants. Unique chromosomal alteration (microdeletion at 4p15.2) and 14 pathogenic chromosomal rearrangements have been detected by array-CGH. Rare causative variants or novel loci causing intellectual disability using the high throughput NGS technique were discovered. New phenotypic observations of recognizable syndromes added new insights for further delineation of the known disorders.

ACKNOWLEDGEMENTS

This work was funded by the Lithuanian–Swiss cooperation programme to reduce economic and social disparities within the enlarged European Union under Project Agreement No. CH-3-ŠMM-01/04, UNIGENE project.

Received 1 April 2016

Accepted 7 June 2016

References

- Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev.* 2002; 8(3): 117–34.
- Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, et al. Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. *Am J Med Genet.* 1997; 72: 468–77.
- Yu S, Baker E, Hinton L, Eyre HJ, Waters W, Higgins S, et al. Frequency of truly cryptic subtelomere abnormalities – a study of 534 patients and literature review. *Clin Genet.* 2005; 68: 436–41.
- Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A.* 2006; 140: 2063–74.
- Niranjan TS, Skinner C, May M, Turner T, Rose R, Stevenson R, et al. Affected kindred analysis of human X chromosome exomes to identify novel X-linked intellectual disability genes. *PloS One.* 2015; 10(2): e0116454.
- Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature.* 2011; 478: 57–63.
- Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics.* 2007; 23: 1289–91.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38: e164.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013; 14: 178–92.
- Rosenfeld JA, Amrom D, Andermann E, Andermann F, Veilleux M, Curry C, et al. Genotype-phenotype correlation in interstitial 6q deletions: a report of 12 new cases. *Neurogenetics.* 2012; 13: 31–47.
- Linnankivi T, Tienari P, Somer M, Kahkonen M, Lonnqvist T, Valanne L, et al. 18q deletions: clinical, molecular, and brain MRI findings of 14 individuals. *Am J Med Genet A.* 2006; 140: 331–9.
- Ciuladaite Z, Preiksaitiene E, Utkus A, Kucinskas V. Relatives with opposite chromosome constitutions, *rec(10)dup(10p)inv(10)(p15.1q26.12)* and *rec(10)dup(10q)inv(10)(p15.1q26.12)*, due to a familial pericentric inversion. *Cytogenet Genome Res.* 2014; 144: 109–13.
- Eyupoglu FC, Sunnetci D, Cine N, Savli H, Okten A, Acikgoz EG, et al. Array-CGH and clinical findings in a patient with a small supernumerary *r(8)* mosaicism. *Genet Couns.* 2014; 25: 305–13.
- Tzschach A, Bisgaard AM, Kirchhoff M, Graul-Neumann LM, Neitzel H, Page S, et al. Chromosome aberrations involving 10q22: report of three overlapping interstitial deletions and a balanced translocation disrupting *C10orf11*. *Eur J Hum Genet.* 2010; 18: 291–5.
- Nizon M, Andrieux J, Rooryck C, de Blois MC, Bourel-Ponchel E, Bourgois B, et al. Phenotype-genotype correlations in 17 new patients with an *Xp11.23p11.22* microduplication and review of the literature. *Am J Med Genet A.* 2015; 167A: 111–22.
- Rossi M, Di Micco P, Perone L, De Brasi D, Guzzetta V, Andreucci MV, et al. Unbalanced translocation (3;5)(q26.1;p14): a clinical report. *Am J Med Genet.* 2002; 110: 353–8.
- 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: 27.
- Doelken SC, Kohler S, Mungall CJ, Gkoutos GV, Ruef BJ, Smith C, et al. Phenotypic overlap in the contribution of individual genes to CNV pathogenicity revealed by cross-species computational analysis of single-gene mutations in humans, mice and zebrafish. *Dis Model Mech.* 2013; 6: 358–72.
- DeScipio C, Conlin L, Rosenfeld J, Tepperberg J, Pasion R, Patel A, et al. Subtelomeric deletion of chromosome 10p15.3: clinical findings and molecular cytogenetic characterization. *Am J Med Genet A.* 2012; 158A: 2152–61.
- Cobben JM, Weiss MM, van Dijk FS, De Reuver R, de Kruiff C, Pondaag W, et al. A de novo mutation in *ZMYND11*, a candidate gene for 10p15.3 deletion syndrome, is associated with syndromic intellectual disability. *Eur J Med Genet.* 2014; 57: 636–8.

21. Campeau PM, Lu JT, Dawson BC, Fokkema IF, Robertson SP, Gibbs RA, et al. The KAT6B-related disorders genitopatellar syndrome and Ohdo/SBBYS syndrome have distinct clinical features reflecting distinct molecular mechanisms. *Hum Mutat.* 2012; 33: 1520–5.
22. Imamura O, Sugawara M, Furuichi Y. Cloning and characterization of a putative human RNA helicase gene of the DEAH-box protein family. *Biochem Biophys Res Commun.* 1997; 240: 335–40.
23. Lin J, Wu PH, Tarr PT, Lindenberg KS, St-Pierre J, Zhang CY, et al. Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1 α null mice. *Cell.* 2004; 119: 121–35.
24. Saez MA, Fernandez-Rodriguez J, Moutinho C, Sanchez-Mut JV, Gomez A, Vidal E, et al. Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. *Genet Med.* 2015; 16.
25. Coutelier M, Burglen L, Mundwiller E, Abada-Bendib M, Rodriguez D, Chantot-Bastaraud S, et al. GRID2 mutations span from congenital to mild adult-onset cerebellar ataxia. *Neurology.* 2015; 84: 1751–9.
26. Dejanovic B, Djemie T, Grunewald N, Suls A, Kress V, Hetsch F, et al. Simultaneous impairment of neuronal and metabolic function of mutated gephyrin in a patient with epileptic encephalopathy. *EMBO Mol Med.* 2015; 7: 1580–94.
27. Rump P, Jazayeri O, van Dijk-Bos KK, Johansson LF, van Essen AJ, Verheij JB, et al. Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Med Genomics.* 2016; 9: 7.
28. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 2012; 367: 1921–9.
29. Isrie M, Kalscheuer VM, Holvoet M, Kleefstra T, Yntema HG, Kroes T, et al. HUWE1 mutation explains phenotypic severity in a case of familial idiopathic intellectual disability. *Eur J Med Genet.* 2013; 56: 379–82.
30. Robertson SP. Filamin A: phenotypic diversity. *Curr Opin Genetics Dev.* 2005; 15: 301–7.
31. Gilillan GD, Selmer KK, Roxrud I, Smith R, Kyllerman M, Eiklid K, et al. SLC9A6 mutations cause X-linked mental retardation, microcephaly, epilepsy, and ataxia, a phenotype mimicking angelman syndrome. *Am J Hum Genet.* 2008; 82: 1003–10.
32. Garbern JY, Neumann M, Trojanovski JQ, Lee VM, Feldman G, Norris JW, et al. A mutation affecting the sodium/proton exchanger, SLC9A6, causes mental retardation with tau deposition. *Brain.* 2010; 133: 1391–402.
33. Kutsche K, Yntema H, Brandt A, Jantke I, Nothwang HG, Orth U, et al. Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat Genet.* 2000; 26: 247–50.
34. Ramakers GJ, Wolfer D, Rosenberger G, Kuchenbecker K, Kreienkamp HJ, Prange-Kiel J, et al. Dysregulation of Rho GTPases in the aPix/Arhgef6 mouse model of X-linked intellectual disability is paralleled by impaired structural and synaptic plasticity and cognitive deficits. *Hum Mol Genet.* 2011; 21: 268–86.
35. Freeze HH, Eklund EA, Ng BG, Patterson MC. Neurology of inherited glycosylation disorders. *Lancet Neurol.* 2012; 11: 453–66.
36. Guerin AA, Feigenbaum A, Donner EJ, Yoon G. Stepwise developmental regression associated with novel CACNA1A mutation. *Pediatr Neurol.* 2008; 39: 363–4.
37. Blumkin L, Michelson M, Leshinsky-Silver E, Kivity S, Lev D, Lerman-Sagie T. Congenital ataxia, mental retardation, and dyskinesia associated with a novel CACNA1A mutation. *J Child Neurol.* 2010; 25: 892–7.
38. Damaj L, Lupien-Meilleur A, Lortie A, Riou É, Ospina LH, Gagnon L, et al. CACNA1A haploinsufficiency causes cognitive impairment, autism and epileptic encephalopathy with mild cerebellar symptoms. *Eur J Hum Genet.* 2015; 23: 1505–12.
39. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet.* 2012; 20: 490–7.

Eglė Preikšaitienė, Laima Ambrozaitytė,
Živilė Maldžienė, Aušra Morkūnienė,
Loreta Cimbalistienė, Tautvydas Rančelis,
Algirdas Utkus, Vaidutis Kučinskas

INTELEKTINĖS NEGALIOS GENETINIŲ PRIEŽASČIŲ NUSTATYMAS NAUDOJANT VISO GENOMO ANALIZĖS MOLEKULINĖS TECHNOLOGIJAS

Santrauka

Įžanga. Populiacijoje intelektinė negalia diagnozuojama 1–2 % žmonių. Tai labai svarbi socialinė bei ekonominė sveikatos priežiūros problema visame pasaulyje. Intelektinei negaliai būdingas platus klinikinių požymių pasireiškimas ir didelis genetinis heterogeniškumas, todėl jos genetinių priežasčių nustatymas yra sudėtingas. Paveldimų ligų naujoms genetinėms priežastims nustatyti efektyviai taikomos viso genomo analizės molekulinės technologijos. Šiame straipsnyje aprašoma pirmoji Lietuvos tyrėjų patirtis atliekant molekulinio kariotipavimo ir viso egzomo sekoskaitos tyrimus, pateikiant mokslinio projekto UNIGENE rezultatus.

Tiriamieji ir metodai. Į projektą įtraukti pacientai, turintys psichomotorinės raidos atsilikimą / intelektinę negalią ($n = 66$). Tiriamiesiems atliktas molekulinis kariotipavimas. FISH ir tikro laiko PGR tyrimai naudoti

nustatytiems nesubalansuotiems chromosominiams pokyčiams patvirtinti ir kilmei nustatyti. Dvylikai šeimų atlikta viso egzomo sekoskaita naudojant naujos kartos sekoskaitos technologijas.

Rezultatai. 14 iš 66 pacientų nustatyti patogeniniai DNR kopijų skaičiaus pokyčiai, vienam pacientui patvirtinta galimai patogeninė 4p15.2 mikrodelecija. Iš viso egzomo sekoskaitai atrinkta dvylika šeimų. Dviems tiriamiesiems identifikuoti patogeniniai sekos pokyčiai (*MECP2*, *CREBBP* genuose). Kitose šeimose nustatytų pokyčių patogeniškumui įvertinti atliekama bioinformacinė ir funkcinė analizė.

Išvados. Esant dideliame intelektinės negalios priežasčių heterogeniškumui, juos identifikuojant svarbi viso genomo analizė: molekulinis kariotipavimas DNR kopijų skaičiaus pokyčiams nustatyti ir viso egzomo sekoskaita detaliems geno pokyčiams identifikuoti. Šiuo tyrimu atskleidžiami viso genomo tyrimams naudojamų molekulinų technologijų privalumai ir rezultatų analizės sudėtingumas bei pateikiama 32 tiriamųjų su chromosominiu nesubalansuotumu ir kandidatiniiais su intelektine negalia susijusiais sekos pokyčiais bei genotipo ir fenotipo informacija.

Raktažodžiai: intelektinė negalia, psichomotorinės raidos atsilikimas, vektorinė lyginamoji genomo hibridizacija, viso egzomo sekoskaita