GENETIC EXAMINATION OF CONGENITAL HEART DEFECT AND NON-SYNDROMIC INTELLECTUAL DISABILITY

Ph.D. Thesis

Gloria Kafui Esi Zodanu

Faculty of Medicine

Department of Medical Genetics

University of Szeged

Szeged, Hungary

2021

ABBREVIATIONS

ACGM	American College of Medical Genetics
AoS	Congenital stenosis of the aorta
array-CGH	Array-comparative genome hybridization
ASD	Atrial septal defect
ASVD	Atrioventricular septal defect
ATP	Adenosine triphosphate
CHD	Congenital heart defect
CMA	Chromosomal microarray analysis
CNV	Copy number variation
CoA	Coarctation of the aorta
DD	Developmental delay
DdPCR	Digital droplet polymerase chain reaction
DEL.	Deletion
DGS	DiGeorge Syndrome
DNA	Deoxyribonucleic acid
DUPL.	Duplication
FISH	Fluorescence in situ hybridization
GTP	Guanosine Triphosphate
ID	Intellectual disability
LCRs	Low copy repeats
MLPA	Multiplex ligation-dependent probe amplification
NS-XLID	Non-syndromic X-linked intellectual disability
PAKs	p21-activated kinases
PBD	p21 binding domain
PCR	Polymerase chain reaction

RD	Rare disease
SNV	Single nucleotide variant
TBX1	T-box transcription factor 1 gene
TDR	Typically deleted region
TOF	Tetralogy of Fallot
TOP3B	Topoisomerase III Beta gene
VSD	Ventricular septal defect
VUS	Variant of unknown significance
XLID	X-linked intellectual disability
YWHAE	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Epsilon gene

INTRODUCTION

Rare diseases are conditions that affect a smaller number of people in the entire population, less than 1 in 2000 people, but collectively these diseases affect a sizeable portion of our population making it an important public health issue. Early diagnosis is essential could provide suitable treatment and medical care to improve the life quality of these individuals. Due to advancements in molecular techniques, the understanding of the genetic causes of these diseases has been brought into focus.

These rare genetic diseases could be the result of single-nucleotide variants (SNV), small insertions or deletions, trinucleotide or hexanucleotide repeat expansions, transcriptional, translational alterations of the gene products or deletions and duplications, these two latest are collectively referred to as copy number variations.
Two main rare disease categories were in focus of our research: ***congenital heart defects*** (CHDs) and ***non-syndromic intellectual disability***.

CNVs play an important role in the development of **CHDs** and the most common CNVs are found on chromosome 22q11.2. Low copy repeats (LCRs) consist of repetitive blocks of sequences that provide the structural basis for a wide range of genomic variation and are described as the main factor in the formation of CNVs. The proximal region of chromosome 22q11 contains eight LCRs (LCR22-A to LCR22-H). Copy number errors that might occur during crossing-over in meiosis, may result in several recurrent and/or rare microdeletions and duplication. Most often, these microdeletions result in the classical 22q11.2 deletion syndrome also known as DiGeorge syndrome. 85-90% of individuals with 22q11 microdeletion syndrome have the classical ~2.5-3-Mb large LCR22 A-D deletion, whereas 8-10% have a nested ~1.5-Mb LCR22A-B deletion.

One of the main genes in this region implicated for the pathogenicity of 22q11.2 deletion syndrome is T-box transcription factor 1 gene (*TBX1)*; a transcription factor that plays a vital role during embryogenesis. Haploinsufficiency of *Tbx1* gene is sufficient to generate at least one important component of the DiGeorge syndrome phenotype.

Intellectual disability (ID) refers to a person with a significantly reduced ability in understanding information and applying skills. The onset of ID is usually before 18 years of age. The X-chromosome is enriched with genes that are involved in the development and maturation of the nervous system, which may cause syndromic or non-syndromic X-linked intellectual disability (NS-XLID).

The p21-activated kinase 3 (*PAK3)* gene, was the fourth gene to be associated with non-syndromic NS-XLID. PAK3 is an essential downstream effector in the Rho-GTPase signaling pathway. It has been reported to play a vital role in dendritic spine morphogenesis, synaptic network dynamics and neuronal plasticity. Hence, *PAK3* mutation, can cause non-syndromic X-linked intellectual disability in combination with neuropsychiatric disorders and mildly dysmorphic features in affected individuals.

AIMS

In light of the diagnostic challenges with 22q11.2 CNVs and considering the relatively low referral number of patients with 22q11.2 CNV to the Department of Medical Genetics, University of Szeged in the last decade, we hypothesized that some patients with 22q11.2 CNVs — especially in the adult population — may have remained undiagnosed. Therefore, we aimed to investigate the 22q11.2 CNVs and *TBX1* gene variants in the paediatric and adult CHD patients in the Southern-Hungarian region, to perform an in-depth genotype-phenotype comparison and to carry out variant segregation analysis in the positive cases.

The second major aim of our study was to investigate the role of a novel *PAK3*-mutation in a male patient with non-syndromic intellectual disability.

PATIENTS AND METHODS

212 unrelated patients (110 females and 112 males, average mean age: 26.9 years and age range: 2 months-74 years) who were previously cardiologically diagnosed with non-syndromic congenital heart defects were enrolled in the study. 211 individuals with no CHDs and no family history of CHD (144 females, 67 males, mean age: 37 years, age range: 8-73 years) were included as controls for comparative analyses. In positive cases, genetic testing was offered to all first-degree family members.

We also report the case of a 14-years-old boy visiting our genetic counselling unit. Upon examination, our patient presented with cognitive impairment, autistic features, temper tantrums, episodic aggression, prior episodes of convulsions, spina bifida occulta, mildly dysmorphic facial features and microcephaly without structural brain abnormalities.

Written informed consent was obtained from all participants and/or legal guardians/parents before enrolment in the study. Investigations in this study were performed according to the Helsinki Declaration 2008 and approved by the National Medical Research Council and the Local Ethical Committee of the University of Szeged.

Peripheral blood samples from all subjects were collected into EDTA tubes. DNA was extracted from the peripheral blood with QIAamp DNA Blood Mini kit (QIAGEN, Gödöllő, Hungary).

To identify CNVs in the 22q11.2 region***, multiplex ligation-dependent probe amplification (MLPA)*** technique was used. All samples were screened using P250-B2 DiGeorge SALSA MLPA Probemix (IVD, MRC-Holland, Amsterdam). This MLPA kit was suitable for the detection of deletions or duplications in the 22q11.2 region. MLPA was repeated for all samples in which copy number variants were identified. Deletions and duplications were subsequently confirmed with ***FISH***, a supplementary MLPA kit or by a ***chromosomal microarray analysis***. In cases where one probe was deleted, the probe region was sequenced with bidirectional capillary sequencing. These validation methods confirmed all positive MLPA results.

During analysis, recurrent single-probe CNVs in *TOP3B* gene from the CHD patient samples was identified hence ***droplet digital polymerase chain reaction*** (ddPCR) technique was designed to confirm the variations and to assess their frequency in the control cohort as well. Primers and probes were designed for *TOP3B* exon7 and for the *PRMD15* as reference region

on chromosome 21. *TOP3B* CNVs found in the controls with ddPCR were further confirmed with MLPA.

Tbx1 sequencing: The entire coding region and the flanking introns of the *TBX1* gene were analysed and the nucleotide sequence of PCR amplified exons was determined using bidirectional Sanger sequencing in all CHD patients.

For the ***x-linked non-syndromic intellectual disability study,*** clinical exome analysis was performed on whole-exome sequence data using Illumina NextSeq500 sequencer after library preparation with Roche KAPA HyperPrep library kit and SeqCap EZ MedExome capture kit. Reads were then aligned to the human reference genome (GRCh37) using BWA (v.0.7.12). The library preparation, sequencing and related bioinformatics was carried out in QGenomics Laboratories, Barcelona.

CNV and SNV interpretation: Identified CNVs and SNVs in the CHD cohort were classified according to the standards and guidelines of the American College of Medical Genetics, and were prioritized on the basis of the functional relevance, inheritance models and minor allele frequency (MAF) in the general population (GnomAD and in-house databases) and control cohort.

Statistical analysis: GraphPad Prism (GraphPad Software, San Diego, California, USA) version 9.00 for Windows, was used for statistical analysis. The frequency of *TOP3B* CNVs and *TBX1* variants in the patient cohort was compared with the frequency in the control cohort and also with the frequency in the global dataset of GnomAD using Fisher exact test and χ^2 test. $P<0.05$ was considered to be statistically significant.

RESULTS

Results in the CHD-cohort

In our CHD cohort, eighteen different types of CHDs were identified. The four most common CHD types (higher than 10%) were ventricular septal defect (VSD), atrial septal defect (ASD), congenital aorta stenosis (AoS), and tetralogy of Fallot (TOF). In 81% of the patients, only one cardiac entity was diagnosed, while in 19% of the patients, two or more CHDs occurred together.

In total, ***17 cases out of the 212*** patients ***(8%),*** previously diagnosed with non-syndromic CHD, were ***positive for MLPA analysis***. After evaluation, ***11 (5.2%)*** of these copy number changes were interpreted as ***pathogenic variant***, ***two as variant of unknown significance (VUS, 0.9%) and four (1.9%) as benign***. However, the most frequent CNVs of the positive MLPA results were microdeletions (8/17) though, microduplications (7/17) and a combination of deletions and duplications (2/17) were also observed.

Among pathogenic CNVs, 7 microdeletions, 3 duplications and 1 combination of a deletion and a duplication, whereas among the VUS one duplication (duplication of LCR F-H) and one combined CNV (combination of a central deletion of LCR C-D with duplication of LCR D-E) was detected. Also, within the benign variants one deletion and three duplications was found. Among the 11 pathogenic variants 6 were typical deletions, 1 proximal nested deletion of LCR A-B, 2 duplications of LCR A-D, a combination of the proximal deletion of LCR A-D with duplication of LCR E-H and a duplication of LCR D-H. Pathogenic results were highest in the TOF group of the CHD cohort ***(17% of all TOF patients)***, followed by the bicuspid aortic valve group (10%). All four benign CNVs were detected in the TOP3B gene. They were subsequently confirmed with chromosomal microarray analysis.

Since *TOP3B* CNVs were identified in a relatively high proportion in our patient cohort, in overall 4 out of the 212 (1.9%) of the CHD-patients, we performed an independent analysis to determine the frequency of *TOP3B* CNVs in the healthy controls. Using ddPCR, we detected *TOP3B* deletion in 1 control sample (0.5%) and a duplication in 4 control samples (1.9%). Altogether, CNVs were identified in 2.4 % of the controls (5/211). The difference in the CNV frequency between patients and controls was not significant (p=0.751). Thus, we ultimately classified *TOP3B* CNVs as rare benign variants, which are more frequent in the Hungarian population than in the global database (frequency in DECIPHER: 0.36%).

Segregation analysis was performed for 14 of the 17 positive CNVs cases. Six cases proved to be familial, two of these were found in patients with pathogenic CNVs, one for a patient with VUS and three for patients with benign *TOP3B* variants.

Genotype-phenotype comparison was performed to compare the clinical spectrum of the 22q11.2 CNVs identified in our cohort.

The probands' age at the genetic diagnosis with pathogenic or VUS 22q11.2 CNVs ranged from 2months to 52 years (median age: 21 years). Three patients out of 13 were diagnosed in childhood, one child in the first year of life. The two oldest patients and the affected family members were born before the molecular diagnostic era. Nevertheless, there was no correlation observed between the severity of the phenotype and the age at diagnosis.

We compared the prevalence of the common clinical features in the different CNVs to previously reported prevalence data in the literature. Patients presented with more marked phenotypic features for 22q11.2 microdeletions than those with microduplications in the same region. Besides the CHDs, the typical microdeletions of LCR A-D with or without accompanying CNVs resulted in the classical phenotype of 22q11.2 deletion syndrome**.**

Also, among patients with deletions, Fallot-tetralogy was the most occurring CHD. For patients with duplications: congenital aorta stenosis, coarctation of the aorta and bicuspid aortic valve were the most common CHD types.

CHDs were overrepresented in our 22q11.2 CNV patients, and their affected family members, compared to the data in the literature (94% *vs* 74%), which could be the result of the patient enrolment criteria. Neuropsychiatric disorders were also underrepresented among our patients (19% *vs* 60%). However, other characteristics including facial features, velopharyngeal insufficiency, immunodeficiency, hypocalcemia, skeletal anomalies, developmental delay and learning difficulties had a distribution in our cohort similar to that reported in the literature. Hypocalcemia (average serum calcium level: 1.84 mmol/l, normal range: 2.2-2.55 mmol/l) was often present in patients with the typical 22q11.2 microdeletions with or without clinical symptoms.

Tbx1 sequencing analysis found no pathogenic variant, although three missense variants were found in exon 9: c.1189A>C; p. Asn397His with a 21% minor allele frequency (MAF), c.1049G>A; p.Gly350Asp with 0.48% MAF and c.1341_1342insCCGCACGCGCAT; p.Ala450_His453dup with 0.24% MAF.

Result of the patient with intellectual disability

A novel variant; NM_001128167.2:c.976G>C;p.(Val326Leu) (ClinVar submission number: SCV000927119; LOVD accession number: #0000578234, DB-ID: PAK3_000063) was identified with exome analysis in exon 10 **of the *PAK3* gene**.

The variant was present in the proband in a hemizygous form and his unaffected mother in a heterozygous form but not in any other healthy family members tested or in the control population databases. The Val326Leu variant was predicted to be likely damaging by the *in silico* predictions. The variant occurred *de novo* in the proband's mother.

Based on the ACMG criteria and the detailed clinical comparison of the typical phenotypic features including intellectual disability, microcephaly, characteristic facial features, anxiety and autistic behaviour with previously described patients, the results hence, supported the ethiopathogenicity of the novel Val326Leu*PAK3*-variant identified.

DISCUSSION and CONCLUSION

Copy number variants in the CHD cohort

Our study was the first systemic, large-scale genetic screening study conducted in Hungarian CHD patients. Although the enrolled patients were cardiologically diagnosed with non-syndromic CHDs prior to this study, 13 were found to be syndromic after the genetic screening.

- ✓ Our study further highlights that the frequency of CHDs in our cohort, was representative and corresponded to the frequency described in large epidemiological studies.
- ✓ All types of CNVs in the 22q11.2 chromosomal region were present in 8% of the CHD cohort, while pathogenic CNVs in 5.2%, VUS in 0.9% and benign CNVs in 1.9%.
- ✓ We observed a higher median age (21 years) and a similar or wider age range (0.17-52 years) at the genetic diagnosis in our cohort as compared to previously described cohorts (median age: 17.3 years, range: 0.1-59.4 years in Canadian patients: median age: 2.9 years, range: 0-17.6 years in American patients)
- ✓ Our patients presented pathogenic 22q11.2 CNVs more often compared to previously described cohorts, such as 1.27% in Brazilian, 2.8% in Cameroonian and 2.9% in Chinese population. However, this difference may also be explained by the fact that most of these studies focused on the detection of 22q11.2 deletion but not on duplications.
- ✓ For TOF, the proportion of pathogenic CNVs was significantly higher, 17% in our cohort, which corresponds with literature on the basis that 22q11.2 CNVs are common in conotruncal heart defects.
- ✓ The most common pathogenic CNV (64%) was the typical microdeletion of the LCR A-D region on chromosome 22q11.2. However, nested (LCR A-B) and central deletions (LCR C-D), proximal duplication (LCR A-D) and distal duplications (LCR D-E, LCR D-H, LCR E-H, LCR F-H) and rare combinations of deletions and duplications were also identified.
- ✓ Segregation analysis detected familial occurrence in 18% (2/11) of the pathogenic variants.
- ✓ We found no pathogenic variant with *TBX1* sequence analysis.

Our findings confirmed the previously described large phenotypic diversity in the 22q11.2 CNVs and also highlighted the benefits for large-scale genetic screening of CHD-patients and the importance of early genetic diagnosis.

PAK3 mutation

In the frame of genetic examinations in X-linked non-syndromic intellectual disabilities we identified a novel hemizygous mutation in *PAK3* gene (c.976G>C; p.Val326Leu) with exome sequencing in a 14-years-old boy.

- ✓ The Val326Leu variant was predicted to be likely damaging by the *in silico* predictions. Segregation analysis in the family revealed, the variant was present in the proband´s unaffected mother in a heterozygous form but not in any other healthy family members tested or in the control population.
- ✓ The variant occurred *de novo* in the proband's mother. Based on the ACMG criteria and the detailed clinical comparison with previously described patients, the ethiopathogenicity of the novel Val326Leu *PAK3*-variant was confirmed.

Our detailed clinical findings together with the data from the few reported families allow further insight in the phenotype of the disease, to expand the mutation spectrum of *PAK3* and support the importance of PAK3-protein in the neural synaptic function.

ACKNOWLEDGEMENTS

I consider myself fortunate to have had the opportunity as the ***first Stipendium Hungaricum student chosen to pursue my Ph.D. at the Faculty of Medicine, University of Szeged***.
This journey has indeed been a life-changing experience and would not have been possible without all the support and contributions I received throughout my study.
I owe a deep sense of gratitude to my Head of Department and amazing supervisors; Professor ***Márta Széll*** & ***Dr. Dóra Nagy*** for their keen interest, tremendous guidance and support throughout my study. No words of thanks can sum up the gratitude in my heart.

The valuable input of ***Dr. Margit Pál,*** that helped shape this work cannot be left unappreciated.

I also extend my gratitude to all my co-authors, the 2nd Department of Internal Medicine and Cardiology Center, the Department of Pediatrics and Pediatric Health Center, Faculty of Medicine, University of Szeged & the Department of Laboratory Medicine, Faculty of Medicine, University of Debrecen for your numerous contributions towards this work.

A special thanks as well to all my colleagues at the Department of Medical Genetics, Faculty of Medicine, University of Szeged: Zsuzsanna Horváth-Gárgyán, Blanka Godza, Dóra Isaszegi and Anikó Gárgyán for their skilled technical assistance.

I am also highly thankful to Dr. Martin-Paul Agbaga, Assistant Professor at the Department of Cell Biology, University of Oklahoma Health Sciences Center for guiding, proofreading and editing this work.

Through it all, this work would not be achievable if not for God; strength and grace was indeed abundant. To my ever-loving family most especially my dad and mum; thanks for the love, care and support. Daddy, I really wished you had waited just a little longer to see this day.
Hmmmm, but God knows best. Rest well in the Lord.
To the love of my life, Clifford Selasi Kekeshie, thank you for holding on and tagging along with me to the end.

I am very thankful also for the Hungarian Scientific Research Fund (grant No 5S441-A202), GINOP-2.3.2-15-2016-00039 grant and the University of Szeged Open Access Fund offered me for this research.

PUBLICATIONS RELATED TO THE THESIS

I. **Gloria Kafui Esi Zodanu**, Mónika Oszlánczi, Kálmán Havasi, Anita Kalapos, Gergely Rácz, Márta Katona, Anikó Ujfalusi, Orsolya Nagy, Márta Széll, Dóra Nagy.
Systemic screening for 22q11.2 copy number variations in Hungarian pediatric and adult patients with congenital heart diseases identified rare pathogenic patterns in the region.
Front. Genet. 12:635480. doi: 10.3389/fgene.2021.635480 IF:3.258

II. Dóra Nagy, Katalin Farkas, Luís Armengol, Maria Segura, **Gloria Kafui Esi Zodanu**, Bernadett Csányi, Alíz Zimmermann, Barbara Vámos, Márta Széll. Further delineation of the phenotype of PAK3-associated X-linked intellectual disability: identification of a novel missense mutation.
Eur J Med Genet. 63(4):103800. 2020.
doi: https://doi.org/10.1016/j.ejmg.2019.103800 IF:2.368

III. Dóra Nagy, **Gloria Kafui Esi Zodanu**, Kálmán Havasi, Anita Kalapos, Mónika Oszlánczi, Márta Katona, Márta Széll.
A 22q11 kromoszomális régió kópiaszám eltéréseinek széles klinikai spektruma congenitalis vitiumos betegekben; Wide clinical spectrum of chromosome 22q11 copy number variations in patients with congenital heart defects.
CARDIOLOGIA HUNGARICA 49 (SupplA) pp. C113-C113 2019.
Journal Article/Abstract /Scientific [30711379].