Ph.D. THESIS

COMPARISON OF RECURRENT AND NON-RECURRENT COPY NUMBER VARIATIONS FOUND ON CHROMOSOME 4

BALÁZS DUGA

PROJECT LEADER: PROF. DR. MELEGH BÉLA

PROGRAM LEADER: PROF. DR. MELEGH BÉLA

DOKTORAL SCHOOL LEADER: PROF. DR. SÜMEGI BALÁZS

PÉCS, 2015

UNIVERSITY OF PÉCS

MEDICAL SCHOOL, DEPARTMENT OF MEDICAL GENETICS

INTRODUCTION

Genomic disorders

At the beginning of the 1990's, when scientists started the Human Genome Project (HGP), they assumed that variations in the human genome sequence (single nucleotide polymorphism - SNP) were behind most disorders. With the end of the HGP and the rapid development of hybridization technology scientists started to view things from a differently. With the advancements in sequencing technologies the human genome could be sequenced within a week. However, the raw data still didn't give answers as to why most diseases developed. The approach today is unchanged, but with geneticists focusing more on gemonics.

CNVs (Copy Number Variations)

In the past decade regions with several copies of one or more regions of DNA were discovered during genome analysis. These regions range in size from one kilobase to several megabases. The copy number changes in these regions of the DNA are called copy number variations or CNVs. The occurrence of CNVs in the genome is a driving force of diversity in population genetics. However, in many cases they are also the underlying cause for genomic diseases. There are two main groups of CNVs. The first is called CNP or copy number polymorphisms. These are found in more than 1% of the population. The second group includes rare CNVs, where the size of CNVs is greater than 100 kb but they occur in less than 1% of the population.

Rare CNVs

There are two groups of rare CNVs: recurrent and non-recurrent CNVs. Recurrent CNVs are the same size and have the same breakpoints. The latter are connected to certain LCR regions. The breakpoints in non-recurrent CNVs rarely fall in the same position and they greatly differ in size as well. Non recurrent CNVs of different sizes may overlap and thus maybe include the same genomic region. The degree of expression of the dosage sensitive genes found in these overlapping regions could be the cause for the observed phenotypes.

Array Comparative Genome Hybridization (aCGH)

Microarray technologies first took off at the beginning of the 2000's. Thanks to the HGP the precise position of each sequence in the genome was known, so chromosome preparations weren't necessary to pinpoint the location of labeled and bound DNA fragments.

The resolution of traditional CGH is 5-10 Mb, however, due to databases containing human genome data and the rapid development of microtechnology it was possible to develop a higher resolution genome hybridization technology, array CGH.

Probes representing the entire genome are bound to a glass slide. The position and sequence of these probes are known. We can bind 2, 4 or even 8 copies of the probes to the slide, which allows us to analyze several samples at one time, however by increasing the number of samples, per slide, the number of probes per sample decreases. During the analysis, just like with traditional CGH, it is required to use a reference DNA of the same sex. During isolation the genomic DNA is fragmented using restriction enzymes. The sample and DNA fragments are labeled with different fluorescence dyes. The sample DNA is labeled with fluorescence dye Cyanine (Cy5), while the reference DNA is labeled with Cyanine (Cy3). At some manufacturers it is possible to interchange the two dyes. The sample and reference DNA are applied to the slide simultaneously and hybridized to the probes on the slide. The slide is then scanned by a laser scanner. The laser generated by the scanner detects the fluorescence dye labeled fragments hybridized to the probes on the slide. The scanner detects the intensity of the emitted light signal generated by the laser. The scanner performs this scan on the entire slide and detects the fluorescent signals at each probe. If the intensity of the two dyes match, then the genomic fragment hybridized to the probe have the same dosage. If the signal intensity of one of the dye is pushed in either direction then we have either a duplication or deletion. It is important to not forget that this technology is not suitable for the detection of genomic variations. Array CGH is able to detect copy number variations and dose alterations in the genome.

AIMS

For my research we collected over 140 samples from patients with severe developmental delay at the Genetic Counseling at the Department of Medical Genetics at the University of Pécs for array CGH studies. For the study I chose patients who had psychomotor delay, muscle hypotonia and congenital malformation and/or cardiac defects as well as delay growth. Absent speech, epilepsy, and stereotypical hand movements were also an important feature. Routine diagnostics revealed no pathological variants responsible for the phenotype in the chosen patients.

Aims of the study:

- To detect expected genomic alterations in patients with complex malformations and abnormal phenotypes, where traditional chromosomal analysis revealed a normal karyotype.
- The analysis of the function of genes and genomic regions affected by the abnormalities using the available literature and public databases.
- Analysis of the literature data to the genes involved in term to assess the impact of the changed gene dosage;
- Examining the phenotype genotype correlation of the detected variants, comparing them to the literature and other cases with similar phenotypes and variants.
- Based on the above mentioned comparison, further collection of data is needed in order to determine which region, which affected gene could cause in similar cases the phenotypic differences.
- Specification of the detected variants (CNVs), comparison with cases found in the literature and comparison of the characteristic features of the variants respectively.
- Comparing the types of rare CNVs based on clinical features in with data available in the literature.
- Based on the genomic alterations and the observed phenotype, more precise definition of array CGH testing indication within the group of patients with multiple malformations.

MATERIALS AND METHODS

PATIENTS

First Patient

The index patient was born as the first child to nonconsanguineous healthy parents at 39th week of gestation, with a birth weight of 2160 g by caesarean section, Apgar scores were 8/9. Because of heart murmur cardiologic examination was performed which revealed ventricular septum defect in the subaortic region, Patent Foramen Ovale and patent ductus arteriosus. Abdominal ultrasound showed smaller kidneys with normal structure and cranial ultrasound showed dysgenesis of the corpus callosum. At the age of two month she was hospitalized due to feeding problems and hypoglycaemia (blood glucose value: 1.2 mmol/L). Severe dystrophy, muscle hypotonia and dysmorphic facial features including remarkable asymmetry of the face (hemihypertrophy of the left side), short right palpebral fissure, long eyelashes, asymmetry of the ears with a dysplastic, smaller and low-set right ear, short philtrum and high palate were detected. During hospitalization impairment of consciousness and short atonic periods were present, which were presumed to be convulsion equivalent, but the performed EEG did not indicate paroxysmal signs. At the age of 3 months her congenital heart defect required surgical correction (closure of the ventricular septum defect, foramen ovale and ductus arteriosus). The results of laboratory tests showed no abnormality besides hyponatremia. After the heart operation there were no residual defects, the pulmonary pressure was normal. During the 4th week of the postoperative period, a severe septic state developed. Because of an advanced AV-block a temporary pacemaker was needed. She was dismissed from the hospital and was regularly checked on by an outpatient cardiology clinic. Four months following heart surgery she was admitted to the clinic again because of pulmonary hypertension. Radiological examinations (Computed Tomography (CT) and CT-angiography) were performed in order to exclude a possible pulmonary cause which revealed a lower arborisation of the pulmonary arteries supplying the left lower lobe. No branches were detected to the left upper lobe and a congenital anomaly was suggested, not excluding the possibility of a persistent non-recanalized thrombosis. Cardiomegaly was reported (primarily to the right part of the heart) with an enlarged pulmonary trunk as a result of elevated pulmonary arterial pressure. A left sided vena cava superior and also an atypically localized (left side) undefined venous vessel was present on the axial and the reconstructed CT slices. The vena cava inferior and also the hepatic veins were enlarged due to elevated venous

pressure in the major blood circle. In the thorax cavity hydrothorax was found on the right side with consolidation on both sides and no normal lung parenchyma. The severe recurrent infections raised the possibility of an immunodeficiency, which was confirmed by flow cytometry of the white blood cells in which significantly decreased lymphocytes were detected. Although hypocalcaemic episodes and high parathyroid hormone levels were reported in the neonatal period, microdeletion of the DiGeorge region of chromosome 22 was excluded. Generalized oedema, deterioration of right ventricular function and recurrent infections progressed, the patient needed continuous mechanical ventilation and despite therapy died at the age of nine month. In search of metabolic diseases urine organic acids, serum amino acids, ammonia and serum transferrin isoelectric focusing were performed to exclude congenital disorders of glycosylation, but all yielded normal results. Routine karyotyping revealed no visible chromosomal aberration and Fluorescence In Situ Hybridization (FISH) analysis of the DiGeorge syndrome critical region (22q11.2) gave also a normal result (Figure 1).

Second Patient

The patient was a 5 year old girl born by caesarean section at 39th week of gestation as the second child of non-consanguineous healthy Hungarian parents, the family history was unremarkable. Her birth weight was 2750 g (25-50 pc), her length 49 cm (5-10 pc), the head circumference 36 cm (+1SD). Her 5 and 10 minute Apgar scores were 9/10. In the perinatal period mild icterus, joint laxity in the hips, axial hypotonia and poor feeding was noted. At 1 week of age severe axial hypotonia and spasticity in the lower limbs was recognized and there was only slight improvement following extensive neurohabilitation. After 3 months her somatic and psychomotor development slowed down and has been very slow ever since. At 6 months of age the patient was hospitalized with severe obstructive bronchitis and during her first year she suffered several upper airway infections with dense mucous and chronic diarrhea, but CFTR-related diseases were excluded. At 14 months of age brain MRI revealed significantly widened and abnormally structured ventricles, diminished periventricular white matter and hypoplasia of the corpus callosum. At the age of 18 month the patient was referred to our genetic counseling unit because of severe hypotonia and developmental delay. Postnatal growth delay: weight was 9.5 kg (5-10 pc), height 68 cm (<3 pc) and head circumference 48.5 cm (+1 SD) and a distinctive facies including broad forehead, frontal bossing, downward slanting palpebral fissures, hypertelorism, hypoplastic ear-lobes, anteverted nostrils, short philtrum, small mouth, higharched palate as well as short, small

hands and feet, distally narrowing fingers, clinodactyly and joint laxity were noted. Neurological examination revealed severe generalized hypotonia and absent speech development. Gross motor milestones were severely delayed despite of extensive neurohabilitation: at the age of 2.5 years she was unable to sit alone, she did not crawl and was unable to stand alone. At the age of 5 years she was able to walk, sit alone, but had no speech. She had good receptive language and used signs and gestures to communicate but had no speech. Stereotypical movements such as hand clapping and flapping and a behavioral disturbance, including occasional self-injurious behavior and aggression toward others were observed. Epilepsy has not been noted so far and repeated EEGs gave negative results. Extensive metabolic (carnitine-ester profiling, amino acids, urine organic acids, isoelectric focusing for CDGs) and genetic testing (routine karyotyping, CFTR sequencing, mitochondrial mutation screening) yielded negative results.

METHODS

DNA isolation

The patients and their families underwent a thorough physical examination and samples were taken with their consent during genetic counseling sessions. For laboratory testing 8-12 ml of venous blood in EDTA tubes was obtained, which were placed in the Biobank following processing. DNA isolation was performed with Omega E.Z.N.A. Blood Maxiprep kit according to manufacturer's directions, which makes it possible to isolate DNA from greater amounts of blood (up to 20 ml). The capacity of DNA binding filter columns is 1.5 mg DNA. During isolation the purity and concentration of the DNA was verified with NanoDrop. If the DNA purity didn't meet the requirement we further purified the DNA with Macherey-Nagel NucleoSpin gDNA Clean-up Purification Kit.

G-banding

The cultivation process: Steril, peripheral blood treated with Na-Heparinate was used for the chromosomal analysis. Four – four ml medium was placed in two tubes to which 5-5 drops of blood were added. The tubes were then incubated in a 37 ˚C thermostat for 72 hours. Two hours before processing 2 drops of Colcemide was added to the cell culture to block cell division in metaphase and it wass placed in a 37 ˚C thermostat for 2 more hours. At the end of the 2 hours the samples were placed in a centrifuge for 10 minutes at 2000 rpm at room temperature and the supernatant was removed. A hypotonic solution was added to the

sediment and incubated for 30 minutes at 37 ˚C. After the half an hour was up, the sample was centrifuged again, and most of the supernatant is removed, and fixing solution was added to the sediment which was then centrifuged at 2000 rpm. The process was repeated three times, until we get a clear suspension. The sample was then placed in -20 ˚C for 20 minutes. After another session in the centrifuge the supernatant was removed again. The sediment was resuspended and 1 drop was placed on 4 clean slides.

Evaluation The cell cultures on the slide were treated with Giemsa staining solution. (80 ml Giemsa staining solution was placed in a dye cuvette, the slides were placed in this cuevette for 15 minutes and then rinsed with distilled water), after air drying the cell cultures were viewed under a microscope: 15 chromosomes in metphase were counted and grouped them. G banding: After Giemsa staining drops of cell culture were placed on another slide and allowed to rest for 24 hours. The slides were then placed in a dye cuvette with 2 x SSC solution for 2.5 hours at 65°. The cells were then treated with Leishmann's staining solution. Banding was assessed with a microscope and each sample yielded 5 karyotypes.

Metaphase FISH

Human peripheral blood was used for FISH. Five drops of blood were added to medium (Chromosome Medium 1A) containing 2 x 4 ml phytohemagglutinin. After vortexing the cell culture was incubated in a thermostat at 37 °C for 72 hours. Two hours before evaluation colcemide (0.1 μg/ml) was added to the cultures. Cellular hypotonization took place at 37°C for 30 minutes with 0.075 M KCl solution, and fixed by rinsing several times with 3:1 ratio mixture of methanol and acetic acid glacial. The cell culture then rests for 30 minutes at -20°C- before drops are placed on slides. The slides are stored at -20°C until analysis. The first step is bringing the slides to room temperature.

Preparation: The slides underwent a pepsin treatment to remove proteins following which they are fixed. Those sections of the slides which were suitable for hybridization were examined with a phase contrast microscope, and the areas of interested were marked.

Denaturation and hybridization: The codenaturation method used required the addition of a hybridization buffer to fluorochrome or hapten probes, then enough distilled water was added to have a solution of 10 μl. Drops of the solution was placed on the area chosen for hybridization. After covering the slides, the probe and chromosomal DNA was denatured on a hot plate at 80 °C for 3 minutes. Finally the slide was placed in a cuvette in a humidity chamber ar 37 °C overnight.

Post hybridization: The freshly made solutions and cuvettes were placed in a 37 °C warm bath. The pH of the solutions were set at 7.0. Slides labeled with fluorochrome were rinsed with distilled water and left to dry at room temperature. The hybridized region was covered with DAPI stain. After denaturation the slides were protected from light. Probes labeled with hapten were treated with 100 µl developer solution (antibody blocker reagent had been added according to the probe's specifications) after the first 4T washes, then covered and placed in a humidity chamber for 30 minutes at 37 °C. After incubation the procedure continued from the 4T protocol all the while protecting the slides from light. The slides were then washed with distilled water and left to dry at room temperature and the hybridized areas were covered with DAPI staining solution. The slides were assessed under a fluorescence microscope with the appropriate filters.

Array CGH

Agilent Human Genome G3 SurePrint 8x60K array was used for the studies.

Preparation of samples and references for the study: For the analysis we used DNA isolated from peripheral blood. The concentration and purity of the isolated DNA was verified using NanoDrop 2000 (Thermo Fisher Scientific, NanoDrop Products, 3411 Silverside Road, Bancroft Building, Wilmington, DE 19810 USA). The elution buffer used during isolation was used as a negative control during assessment. 1.5 µl of DNA per sample were placed on the surface of the NanoDrop 2000 detector, and the lid was closed to being the process. The program detects the absorbance differences between the negative control and the samples, and include the DNA concentration and the purity of the samples.

The DNA concentration of the samples have to be at least 100 ng/µl for the required DNA concentration in the first step of the protocol to reach 1 µg in 10.1 µl volume. The purity of DNA is denoted with 260/280 and 260/230 ratios. The ratio of 260/280 is the ratio of the absorbency of DNA and RNA. Under normal circumstances the double stranded DNA absorbs UV light at 260 nm. Single stranded RNA absorbs UV light at 280 nm. The ratio of the two values gives the contamination of RNA in the DNA. An RNA contamination above 1.8 is acceptable. The second ratio, 260/230 is the contamination of the samples DNA with an organic solvent. This is a higher ration than 260/280. Here a value above 1.9 – 2.0 is preferred. After preparation of the samples according to array CGH protocol the volume and concentration of the samples are set at 10.1 µl and 1µg/10.1µl. DNA with the wrong concentration and purity was purified using Macherey-Nagel gDNA Clean-up Purfication Kit.

Fragmenting the samples and references with restriction enzymes: For the analysis 1 µg of fragmented reference DNA and sample DNA is necessary. The initial sample and reference DNA was maximized in volume at 10.1 µl. The final volume of the samples and digestive master mix together was 13µl. The mixture was incubated in a PCR with the following program: 37 degrees – 2 hours, 65 degrees – 10 minutes, 4 degrees (or ice). Thanks to the restriction enzymes the genomic DNA had fragmented. Agilent BioAnalyzer 2100 was used to determine the size of the fragments. After digestion 1 µl of the sample and reference sample was assessed with gel electrophoresis. The used amount was replaced with distilled water.

Amplification and labeling of sample and reference fragments: 2.5 µl of random primer was added to the sample and reference and then vortexed. The samples were incubated in the PCR at the following program: 95 degrees – 5 minutes, 4 degrees – 3 minutes. The samples were centrifuged for 1 minute at 6000g. Afterwards a labeling master mix was added (9.5 µl). It was important the sample and references are labeled with a different fluorescence dye (samples: Cy5/Cyanin5, references: Cy3/Cyanin3). 9.5 µl of labeling mix was added to each tube, Cy5 mix was added to the samples and Cy3 was added to the references. The tubes were incubated in the PCR at the following program: 37 degrees – 2 hours, 65 degrees – 10 minutes, 4 degrees.

Purification of samples and references: The samples were centrifuged for 1 minutes at 6000 g. For purification Amicon AU-30 filter tubes were used. 430 µl TE buffer was added to these tubes, centrifuged for 10 minutes at 14000 g. The flow through was removed and 480 µl TE buffer was added. Once again the samples were centrifuged and the flow through removed. At this point the labeled DNA was bound to the filter while unbound DNA is washed away. The last step was to place the filters upside down in a new tube and centrifuged for 1 minute at 1000 g. The volume of obtained labeled DNA was between 20-32 µl. In order to have the required volume (9.5 µl), the samples were concentrated with thermo block. If the volume of the samples didn't reach 9.5 µl TE buffer was added. The integration of the dye was measured with the NanoDrop. 1.5 µl was used for this. During the measurement the concentration, and absorbency of the dye was used to calculate the integration of the dye. In the case of samples the optimal value for integration is 20-35, for the references 25-40. The correct sample pairs were measured together (16 µl).

Hybridization of reference pairs onto the slide: The labeled DNA was added to the hybridization mix (29 µl/minta), and incubated in a PCR at 95 degrees for 3 minutes, 37 degrees for 30 minutes. 40 µl of the sample was pipetted on the correct position of a gasket

slide („drag and dispense”), the active surface of the array slide was placed the gasket slide. The hybridization chamber was closed and incubated at 65 degrees at 20 rpm for 24 hours.

Washing the slides: After hybridization the chamber was taken apart, the gasket slide – array slide removed, placed in a tray with washing buffer number 1 and the two slides separated under the buffer. The slide was placed in washing tub on a magnetic stirrer in buffer number one for 5 minutes. In the meantime buffer number two had been stored at 37 degrees overnight on a magnetic stirrer. After 5 minutes in buffer number 1 the slide was placed in buffer number 2 for 1 minute and dried in a dark chamber.

Scanning the slides: The dried slides are placed in an array scanner container and placed in the scanner. Agilent ScanControl is used for the scan and the created image file (.TIFF) is processed with the help of Agilent Feature Extraction program. The program places a grid file on top of the image, which determines the position of the probes on the slide. The output files are viewed with Agilent Cytogenomics program.

Data evaluation: Evaluation of the data was performed with Agilent Cytogenomics program. The program analyses files generated by the Feature Extraction program makes it possible for visual evaluation. It is possible to determine the affected genes and precise position of the genomic aberrations and breakpoints with the program. The program is in contact with several databases so it is possible to determine pathogenic variations with the program.

RESULTS

We analyzed 140 patients with complex malformations by array CGH. In my thesis I would like to present two cases, in which we found similar variations, namely on chromosome 4. One is a recurrent CNV while the other is non-recurrent., which based on size and genes are responsible for the development of the phenotype.

First case

Based on the clinical features of the female patient we assumed that she had some sort of genomic disorder, so at first we performed a chromosome analysis, which yielded normal karyotype. As traditional banding showed no variations our next step was to perform array CGH analysis. Array CGH analysis of our patient with complex malformations revealed a 14.56 Mb deletion on the long arm of chromosome 4 (4q28.3q31.23; 136,127,048 - 150,690,325). We identified 47 genes in the deleted region: *PCDH18, LOC641365, SLC7A11, CCRN4L, ELF2, C4orf49, NDUFC1, NAA15, RAB33B, SETD7, MGST2, MAML3, SCOC, LOC100129858, CLGN, ELMOD2, TBC1D9, RNF150, ZNF330, IL15, INPP4B, USP38, GAB1, SMARCA5, LOC441046, FREM3, GYPE, GYPB, GYPA, LOC646576, HHIP, ANAPC10, ABCE1, OTUD4, SMAD1, MMAA, C4orf51, ZNF827, LSM6, SLC10A7, POU4F2, TTC29, EDNRA, TMEM184C, PREMT10, ARHGAP10, NR3C2.* In a detailed analysis of the genes affected by the deletion we highlighted 8 genes (PCDH18, SETD7, ELMOD2, IL15, GAB1, HHIP, SMAD1, NR3C2) with possible contributions to the phenotype We performed array CGH analysis on the patients parents in order to determine if CNV was inherited or a de novo mutation. Neither of the parents had the deletion that we detected in the child. Array CGH results were confirmed in the parents via metaphase FISH.

Comparison of first patient to similar cases in literature

The detected 4q28 deletion in the first patient is rarely mentioned in literature, especially in publications with detailed phenotype description of the patient. Although the detected CNV breakpoints are fairly close to each other the phenotypic features give a heterogenic picture. The reason for this that the phenotypic data in databases are insufficient. This can be seen in case of large deletions which partially or completely overlap the deleted region in our case, however, the phenotypic variation in patients are poor when compared to the extent of the deletion, which is not justified with the gene density and dosage density. The variations listed in DECIPHER also made evaluation of the data more difficult because in most cases the

patients age was not listed, so it was impossible to determine if the patient didn't exhibit symptom or if the patent was too young and symptoms would have appeared later on or not.

Second case

After extensive metabolic tests and exclusion of subtelomeric deletions array CGH analysis was performed using the Agilent Human Genome G3 SurePrint 8x60K Microarray, which detected a 4.85 Mb de novo interstitial deletion of 4q21.21-4q21.23 (ch4:81 408 980–86 261 953) The deletion in our patient involved the following genes: PRKG2 (MIM 601591), RASGEF1B (MIM 614532), HNRNPD (MIM 607137), HNRPDL, ENOPH1, COQ2, MRPS18C, THAP9, HPSE, and CDS1. Except for known CNVs, no copy number alterations were observed in other chromosomes (data not shown). Based on the normal CGH array profile of the parents this deletion proved to be de novo. The results were confirmed with metaphase FISH.

Comparison of second patient to similar cases in literature

Unlike in the first patient, the detected abnormality in the second patient is a microdeletion syndrome that has been known since 2010. However, even with this case there are very few cases in the literature which affect the 4q21 region.

Comparison of recurrent and non-recurrent CNVs

Comparison of the genomic region affected by the 14.56 Mb deletion detected in the first patient and the patients phenotype to the available literature we can say that the cases show no great similarities in breakpoints or in clinical symptoms. However, comparison of the phenotypes in the second case caused by the 4.85 Mb deletion to similar cases with similar breakpoints in the literature we can say that the symptoms are more uniform. The majority of symptoms in patients 1 and 2 are similar, however the incident rate differs. In case of the second patient and its group short stature was registered in more than 80% of the cases, in the case of the first patient and its group this ration didn't exceed 50%. The most common symptoms in the second group had a ratio of more than 50%. This is possibly due to the different mechanisms which play a role in the development.

Comparison of cases with 4q28.3 deletions revealed a common region which could be found in all deletions. This region is called the smallest overlapping region. The deletion breakpoints are not near the LCR region and it is obvious that there are no hotspots, where the formation of a CNV could be predicted. This is why we can state that the deletion detected in the 4q28.3

region is a non-recurrent CNV, thus, in our patient that detected deletion developed thanks to the NHEJ or FoSTeS mechanism. After examining the different CNVs in our two cases we can state that the phenotypic features caused by non-recurrent CNVs do not produce uniform clinical symptoms. The reason for this is that there is no determined breakpoint in non-recurrent CNVs, so the size and genes affected cannot be predicted, so the clinical features vary phenotypically. On the other hand the recurrent CNVs linked to the 4q21 microdeletion syndrome are determined by the LCR mediate breakpoints. Thus the clinical features are more homogenous and their detection based on phenotype is more likely.

NEW RESULTS

1. We have identified a de novo interstitial deletion in a 9 month-old girl with growth failure, developmental delay, ventricular septum defect in the subaortic region, patent foramen ovale and patent ductus arteriosus, vascular malformation of the lung, dysgenesis of the corpus callosum and craniofacial dysmorphism using array-comparative genomic hybridization. This de novo deletion is located at 4q28.3-31.23 (136,127,048 - 150,690,325), its size is 14.56 Mb, and contains 8 relevant genes (PCDH18, SETD7, ELMOD2, IL15, GAB1, HHIP, SMAD1, NR3C2) with possible contributions to the phenotype
2. Functions lost with the deleted genes could give us an answer for the clinical features in case of the first patient (cognitive functions: *PCDH18, SETD7*; immunodeficincies: *IL15*; low Na ion levels: *NR3C2*), especially pulmonary vascular disorders (*ELMOD2, GAB1, HHIP*), pulmonary hypertension (*SMAD1*). These results lead to the further understanding and discovery of 4q CNVs genetic spectrum.
3. The second patient was a 5 year old girl here in Hungary with typical 4q21 microdelection syndrome (second case). The clinical features partially overlap with symptoms mentioned in similar cases in literature for 4q21 microdeletion syndrome (severe developmental delay, absent speech, behavior problems). We detected a *de novo* 4.85 Mb deletion with array CGH on the long arm of chromosome 4 (4q21.21-4q21.23). The deleted region involves 10 genes: PRKG2 (MIM 601591), RASGEF1B (MIM 614532), HNRNPD (MIM 607137), HNRPDL, ENOPH1, COQ2, MRPS18C, THAP9, HPSE, and CDS1.
4. Out of the 10 above mentioned genes, there is a 1.37 Mb minimal critical region with 5 genes: *PRKG2, RASGEF1B, HNRNPD, HNRPDL, ENOPH1*. From these genes *PRKG2* (severe growth delay) and *RASGEF1B* (cognitive functions) are the major determinants of 4q21 phenotype. BMP3 gene haploinsufficiency could play a role in the development of bone deformity, such as frontal bossing and broad forehead. The other 5 genes *BMP3, COQ2, MRPS18C, THAP9, HPSE*, and *CDS1* no link could be found between the proteins encoded by the genes and the patients clinical features. Our results contribute to the future treatment of patients suffering from 4q21 microdeletion syndrome, furthermore we emphasis the importance of array CGH in the determination of the genetic cause behind intellectual disabilities.

5. Based on the results of the array CGH analysis of our patients we concluded that in the first case the breakpoint of the deletions do not coincide with segmental duplications, while in the second case, the breakpoint of the detected aberration is oriented towards the LCR hotspot.
6. In the case of the first patient the detected aberrations were compared to the published literature data where we concluded that breakpoints of 4q28.3-31.23 and its region do not give a uniform picture and the breakpoints of the deletions do not coincide with segmental duplications. However there is a common region in the deleted regions, the minimal critical region. We concluded that in the case of the first patient the detected copy number variation is a non-recurrent CNV.
7. Comparison of the second patients to cases in the published literature we discovered that the breakpoints are found in the LCR region. These breakpoints formed through the NAHR mediated segmental duplication. Based on this we can conclude that in the case of the second patient the aberration is a recurrent CNV.
8. Comparison of the two cases with the published literature revealed that in the case of non-recurrent CNVs there are no uniform clinical symptoms due to the variable genetic region. (the ratio of the most common symptoms are maximum 50 %). With recurrent CNVs and determined breakpoints the clinical symptoms are more uniform which is also due to the determined breakpoints. (the ratio of the most common symptoms are above 80%).
9. We were the first in Hungary to identify indications in patients with complex malformations with the help of array CGH.

PUBLICATIONS

Publications related to thesis:

1. Phenotypic variability in a Hungarian patient with the 4q21 microdeletion syndrome.

Komlósi K, **Duga B**, Hadzsiev K, Czakó M, Kosztolányi G, Fogarasi A, Melegh B. Mol Cytogenet. 2015 Mar 3;8:16. doi: 10.1186/s13039-015-0118-7. eCollection 2015. Impact factor: 2,140

2. Deletion of 4q28.3-31.23 in the background of multiple malformations with pulmonary hypertension.

Duga B, Czako M, Komlosi K, Hadzsiev K, Torok K, Sumegi K, Kisfali P, Kosztolanyi G, Melegh B. Mol Cytogenet. 2014 Jun 5;7:36. doi: 10.1186/1755-8166-7-36. eCollection 2014. Impact factor: 2,140

Publications not related to thesis:

Published papers:

1. Common functional variants of APOA5 and GCKR accumulate gradually in association with triglyceride increase in metabolic syndrome patients.

Hadarits F, Kisfali P, Mohás M, Maász A, **Duga B**, Janicsek I, Wittmann I, Melegh B. Mol Biol Rep. 2012 Feb;39(2):1949-55. doi: 10.1007/s11033-011-0942-8. Epub 2011 Jun 4. Impact factor: 2,506

2. Mutations of the apolipoprotein A5 gene with inherited hypertriglyceridaemia: review of the current literature.

Melegh BI, **Duga B**, Sümegi K, Kisfali P, Maász A, Komlósi K, Hadzsiev K, Komoly S, Kosztolányi G, Melegh B. Curr Med Chem. 2012;19(36):6163-70. Impact factor: 4,070

3. Hodgkin disease therapy induced second malignancy susceptibility 6q21 functional variants in roma and hungarian population samples.

Varszegi D, **Duga B**, Melegh BI, Sumegi K, Kisfali P, Maasz A, Melegh B.

Pathol Oncol Res. 2014 Jul;20(3):529-33. doi: 10.1007/s12253-013-9724-z. Epub 2013 Dec 5. Impact factor: 1,855

4. Marked differences of haplotype tagging SNP distribution, linkage, and haplotype profile of IL23 receptor gene in Roma and Hungarian population samples.

Magyari L, Varszegi D, Sarlos P, Jaromi L, Melegh BI, **Duga B**, Kisfali P, Kovesdi E, Matyas P, Szabo A, Szalai R, Melegh B. Cytokine. 2014 Feb;65(2):148-52. doi: 10.1016/j.cyto.2013.11.011. Epub 2013 Dec 11. Impact factor: 2,664

5. [Identifying rare genomic disorders with array comparative genomic hybridization in Hungary].

Duga B, Czakó M, Hadzsiev K, Komlósi K, Sümegi K, Kisfali P, Kosztolányi G, Melegh B. Orv Hetil. 2014 Mar 2;155(9):358-61. doi:10.1556/OH.2014.29825. Hungarian. Impact factor: 0

6. Admixture of beneficial and unfavourable variants of GLCCI1 and FCER2 in Roma samples can implicate different clinical response to corticosteroids.

Szalai R, Matyas P, Varszegi D, Melegh M, Magyari L, Jaromi L, Sumegi K, **Duga B**, Kovesdi E, Hadzsiev K, Melegh B.

Mol Biol Rep. 2014 Nov;41(11):7665-9. doi: 10.1007/s11033-014-3659-7. Epub 2014 Aug 5. Impact factor: 2,024

7. [Attention deficit hyperactivity disorder analyzed with array comparative genome hybridization method. Case report].

Duga B, Czakó M, Komlósi K, Hadzsiev K, Sümegi K, Kisfali P, Melegh M, Melegh B. Orv Hetil. 2014 Oct 5;155(40):1598-601. doi: 10.1556/OH.2014.30006. Hungarian. Impact factor: 0

8. Significant interethnic differencies in functional variants of PON1 and P2RY12 genes in Roma and Hungarian population samples.

Janicsek I, Sipeky C, Bene J, **Duga B**, Melegh B, Sümegi K, Jaromi L, Magyari L, Melegh B. Mol Biol Rep. 2015 Jan;42(1):227-32. doi: 10.1007/s11033-014-3762-9. Epub 2014 Oct 9. Impact factor: 2,024

9. Genetic polymorphisms in promoter and intronic regions of CYP1A2 gene in Roma and Hungarian population samples.

Szalai R, Magyari L, Matyas P, **Duga B**, Banfai Z, Szabo A, Kovesdi E, Melegh B. Environ Toxicol Pharmacol. 2014 Nov;38(3):814-20. doi: 10.1016/j.etap.2014.09.012. Epub 2014 Sep 28. Impact factor: 2,084

10. Functional Variants of Lipid Level Modifier MLXIPL, GCKR, GALNT2, CILP2, ANGPTL3 and TRIB1 Genes in Healthy Roma and Hungarian Populations.

Sumegi K, Jaromi L, Magyari L, Kovesdi E, **Duga B**, Szalai R, Maasz A, Matyas P, Janicsek I, Melegh B. Pathol Oncol Res. 2015 Jan 9. Impact factor: 1,855

11. Partial tetrasomy of the proximal long arm of chromosome 15 in two patients: the significance of the gene dosage in terms of phenotype

Szabo A, Czako M, Hadzsiev K, **Duga B**, Komlosi K, Melegh B. Mol Cytogenet. 2015 Jun 25;8:41. doi: 10.1186/s13039-015-0137-4 Impact factor: 2,14

12. Increased prevalence of functional minor allele variants of drug metabolizing CYP2B6 and CYP2D6 genes in Roma population samples.

Weber A, Szalai R, Sipeky C, Magyari L, Melegh M, Jaromi L, Matyas P, **Duga B**, Kovesdi E, Hadzsiev K, Melegh B. Pharmacological Reports 67:(3) pp. 460-464. (2015 June) doi: 10.1016/j.pharep.2014.11.006. Impact Factor: 1,928

13. Erratum to: Significant interethnic differences in functional variants of PON1 and P2RY12 genes in Roma and Hungarian population samples.

Janicsek I, Sipeky C, Bene J, **Duga B**, Melegh BI, Sümegi K, Jaromi L, Magyari L, Melegh B., Mol Biol Rep. 2015 Jan;42(1):317. doi: 10.1007/s11033-014-3798-x. Impakt Faktor: 2,024

Book Chapters:

1. Role of Functional Variants and Mutations of the Apolipoprotein A5 Gene in Human Pathology

Balázs Duga, Béla I.Melegh, Katalin Sümegi, Anita Maász, Péter Kisfali, Katalin Komlósi, Béla Melegh B.: Apolipoproteins: Regulatory Functions, Health Effects and Role in Disease, New York: **Nova Science Publishers Inc.**, 2012. pp. 75-92, (ISBN:978-1-62257-484-1)

2. Ischemic Stroke Susceptibility Gene Research: Lessons We Learned

Bela I. Melegh, Anita Maasz, Peter Kisfali, Katalin Sumegi, **Balazs Duga**, Gyorgy Kosztolanyi, Samuel Komoly, Bela Melegh: Ischemic Stroke: Symptoms, Prevention and Recovery, New York: **Nova Science Publishers Inc.**, 2013. pp. 117-144. (ISBN:978-1-62257-799-6)

3. Role of Triglyceride Modifier Genetic Variants in Development of Metabolic Syndrome

Anita Maasz, Peter Kisfali, Bela I. Melegh, **Balazs Duga**, Katalin Sümegi, Bela Melegh: Handbook on Metabolic Syndrome: Classification, Risk Factors and Health Impact, New York: **Nova Science Publishers Inc.**, 2012. pp.95-119. (ISBN**:** 978-1-62257-025-6)

Cumulative Impact Factor: 29,454