BENIGN AND UNKNOWN COPY NUMBER VARIATIONS IN BULGARIAN PATIENTS WITH INTELLECTUAL DISABILITY AND CONGENITAL MALFORMATIONS

Savina Petrova Hadjidekova[1], Daniela Mircheva Avdjieva-Tzavella[2], Blaga Borisova Rukova[1], Desislava Valentinova Nesheva[1], Radka Stefanova Tincheva[2], Draga Ivanova Toncheva[1]
[1]Medical University of Sofia, Faculty of Medicine, Sofia, Bulgaria
[2]Medical University of Sofia, University Hospital of Paediatrics "Knyaginya Evdokia", Sofia, Bulgaria
Correspondence to: Savina Hadjidekova
E-mail: savinaagova@yahoo.com

ABSTRACT

Molecular karyotyping is an extremely suitable method for genetic diagnosis of patients with unclear dysmorphic syndromes and intellectual disability. In this study we present our results from microarray analysis of 52 patients with developmental delay and congenital malformations. Our data revealed definite etiology in 9 out of 52 patients tested. Fifteen pathological aberrations were found in them. All pathological findings were validated by fluorescent in situ hybridization (FISH) analysis. In addition, the majority of the patients tested (41 patients) showed normal variations in the number of copies and variations of unknown clinical significance (34 patients). Analyses of the type and distribution of the different variations was performed and the clinical significance of variants of unknown nature was discussed. Our results show the advantages of high resolution microarrays for clinical diagnosis of patients with intellectual disability and congenital malformations, and also highlight the need for extensive population studies revealing the molecular nature and clinical significance of different copy number variations and for creation of detailed maps of variations in the Bulgarian population.

Biotechnol. & Biotechnol. Eq. 2013, **27**(6), 4304-4307

Keywords: array CGH, unknown CNVs, benign CNVs

Abbreviations: BAC: bacterial artificial chromosomes; CNVs: copy number variations; FISH: fluorescent in situ hybridization; OMIM: Online Mendelian Inheritance in Man

Introduction

"Each human's genome is distinguished by extra, and sometimes missing, DNA that can powerfully impact everything from development to disease" (2). After an initial delay of the microarray analysis application for routine diagnosis due to some technical flaws, molecular karyotyping is now widely used in almost every molecular cytogenetic laboratory. The evergrowing resolution of the DNA microarrays and the optimization of this technique allow detection of larger (> 1 Mb) aberrations and a variety of small copy number variations (CNVs) whose clinical significance is unclear in some cases. The development of the microarray technology to whole genome screening revealed an unexpectedly large number of deletions and duplications in the human genome. These findings require a thorough study of CNVs in healthy individuals and patients.

Currently, it is known that CNVs are ubiquitous in the human genome. They can be polymorphic (frequency >1%) or rare (<1%), inherited or *de novo*, biallelic or multiallelic (1). A negative correlation between the size of CNVs and their frequency has been established: the smaller the size, the greater the frequency (6). In 2004 two large studies provided data on the analysis of CNVs in healthy populations. They found that the genome of each person has 12 CNVs on the average (7, 17). Another study showed that 12 % (about 360 Mb) of the human genome is covered by CNVs; they are found mostly in low copy repeat (LCRs) areas and cover at least around 2900 genes (10 % of known so far) (16).

Data from the sequencing of the human genome suggest that insertions and deletions are responsible for 22 % of the observed variation, while they comprise 74 % of the affected nucleotides (11). More recent data from Venter (20) revealed that the genomes of two individuals may differ between 1 % and 3 %.

Itsara et al. (8) studied 2500 individuals for CNV by mining data from an Illumina array and observed that 65 % to 80 % of the individuals have a CNV > 100 Kb[5], 5 % to 10% of the individuals have CNVs > 500 Kb, while 1 % to 2 % have CNVs > 1 Mb. The average amount of CNVs per person is estimated to be between three and seven variants. Another important observation has been made, that the majority of the genomic variations are present at ~ 0.02 % to 1 % frequency and span 6 % of the human genome, whereas polymorphic CNVs encompass 0.09 % of the genome. Another study also reported that large CNVs affect much less of the genome than previously thought (i.e. the estimation by Redon et al. [16] of 12 %). This overestimation has been explained by the usage of large insert BAC clones, which are characterized by decreased sensitivity. That is, the initial estimation that 12 % of the genome is encompassed by CNVs has not been supported by subsequent studies, rather, the new data suggest rates closer to 5 %. This rate was also further supported by Pinto et al.

(15), who studied a healthy control population and observed that 160 Mb of the genome (~ 5 %) is covered by CNVs; 96 % of these CNVs are rare with a frequency of < 2 %, and the rest are common.

Materials and Methods

We studied a group of 52 patients of both sexes with congenital anomalies and developmental delay or intellectual disability. Whole-genome oligo-array CGH was performed using the BlueGnome CytoChip oligo 2X105K microarray, v1.1. This array contained ~ 105 000 oligonucleotide probes spaced at an average distance of 35 kb based on the NCBI build (36) of the human genome. Arrays were scanned at 532 nm (Cy3) and 635 nm (Cy5), using a GenePix 4100A two-color fluorescent laser scanner (Axon Instruments, Union City, CA, U.S.A.). The images were analyzed by BlueFuse Multi, version 2.2. (BlueGnome, Cambridge, UK).

Results and Discussion

Our data revealed definite etiology in 9 out of 52 patients tested. Fifteen pathological aberrations were found in them. All pathological findings were validated by FISH analysis.

We detected a total of 247 CNVs, of which 15 pathogenic (7 deletions, 8 duplications), 124 benign (62 deletions, 62 duplications) and 108 with unknown clinical significance (68 deletions, 40 duplications). Variations were found in all chromosomes except chromosome 20. Only one variation was found in chromosome 19. It is known that both chromosomes have relatively high concentration of genes. Fifteen pathogenic genomic aberrations were detected in nine of the patients. Six of the normal variations occurred in more than 10 % of the investigated group. Three variants of unknown clinical significance were found in more than 5 % of the patients.

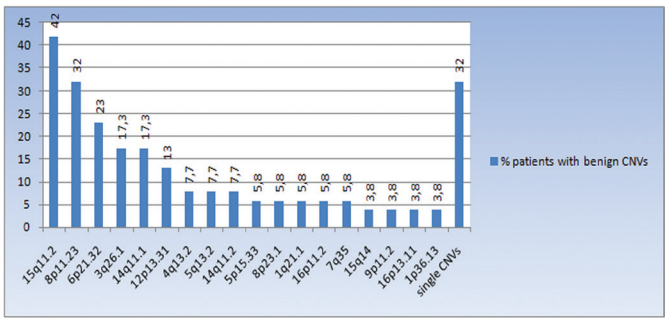

Fig. 1. Frequency of different types of benign CNVs in the investigated group. Variations in 15q11.2, 8p11.23, 6p21.32, 3q26.1, 14q11.1 and 12p13.31 loci occurring in over 10 % of the patients.

The results revealed 124 benign CNVs distributed among 41 patients. Eighteen loci occurred in more than one patient

Fig. 2. Benign CNVs observed in over 10 % of the selected patients. Losses (red), duplications (green).

Fig. 3. Unknown CNVs detected in over 5 % of the investigated group. (red), duplications (green). The variation in the 10q11.22 locus in patients 4, 8 and 12 is benign.

TABLE 1

Unknown CNVs detected in the patients

Patient	Locus	Number of OMIM genes	Size	Type	OMIM disorder
22	1p36.32	2 (OMIM 601990, 601883)	886 175	duplication	
41	1p21.3	1 (OMIM 274270),	60 977	deletion	dihydropyrimidine dehydrogenase (274270)
45	1q31.3	2 (OMIM 134371, 605337)	84 096	duplication	
48	1q31.3	2 (OMIM 605336, 134371)	101 856	deletion	
17	1q43	1 (OMIM 180902)	110 704	deletion	ventricular tachycardia, catecholaminergic polymorphic, 1 (604772), arrhythmogenic right ventricular dysplasia, familial, 2 (600996), ventricular tachycardia, familial (192605)
51	2p11.2	1 (OMIM 173340)	666 238	deletion	
34	2q32.1	2 (OMIM 114190, 152310)	85 584	deletion	
52	4p13	1 (OMIM 604604)	99 545	deletion	
51	5q13.2	5 (OMIM 603011, 600354, 601627, 600355, 601748)	957 476	deletion	spinal muscular atrophy, type II (253550), spinal muscular atrophy, type IV (271150), spinal muscular atrophy, type I (253300), spinal muscular atrophy, type III (253400)
26	7p14.1	1 (OMIM 603309)	45 75	duplication	
50	7q31.1	1 (OMIM 605977)	35 762	deletion	
10	9p13.1	1 (OMIM 610517)	104 834	deletion	
10	10q11.22	1 (OMIM 601790)	136 905	deletion	
10	10q11.22	2 (OMIM 610630, 610631)	466 265	deletion	
13	10q11.22	3 (OMIM 608081, 611240, 601790)	404 978	deletion	
25	10q11.22	2 (OMIM 608081, 611240)	60 352	deletion	
50	10q11.22	2 (OMIM 610630, 610631)	527 093	duplication	
50	12p13.31	2 (OMIM 611039, 138170)	114 563	duplication	
40	12q24.33	1 (OMIM 611257)	54 372	duplication	
45	14q13.2	1 (OMIM 605680)	87 634	duplication	
44	15q13.1	2 (OMIM 602712, 608243)	715 018	duplication	
44	15q13.1	1 (OMIM 601009)	171 664	duplication	
16	17q11.1	1 610091	36 63	deletion	
38	22q11.21	1 (OMIM 601279)	155 585	duplication	
39	Xp22.13	1 (OMIM 300208)	100 922	duplication	
17	Xq13.3	1 (OMIM 300135)	73 685	deletion	anemia, sideroblastic, and spinocerebellar ataxia (301310)
52	Xq21.1	1 (OMIM 603121)	63 337	deletion	
52	Xq22.1	1 (OMIM 300642)	47 438	deletion	polymicrogyria, bilateral perisylvian (300388), rolandic epilepsy, mental retardation, and speech dyspraxia, x-linked; (300643)
52	Xq21.31	1 (OMIM 603121)	85 753	deletion	

and 17 loci in only one patient. The most common normal variations in our study were in 15q11.2, 8p11.23, 6p21.32, 3q26.1, 14q11.1 and 12p13.31 loci, occurring in over 10 % of the patients (**Fig. 1**). It is worth noting that in some patients there was a deletion, while in others there was a duplication, which indicates that these regions are considerably variable (**Fig. 2**). It is necessary to perform genomic screening of a larger group of individuals to determine the real frequency of the uncovered benign variants.

The rates of unknown CNVs in our study were notably high: 108 CNVs of unknown clinical significance distributed among 34 patients. This indicates that some of these variations are probably normal for the Bulgarian population and cannot be found in the studied foreign populations. It is well known that the Bulgarians are characterized by high genetic heterogeneity. The contemporary Bulgarian gene pool is a result of a change in allele frequencies in the course of evolution, which occurred under the influence of a number of natural and demographic events. The territory of Bulgaria has a key geographical position and served as the front door to the people entering Europe in the Middle Paleolithic. Bulgaria used to be the entrance of the migration flow from the Middle East to Central and Western Europe and that fact seriously affected its population's genetic history.

The unknown CNVs were with the following size distribution: 57 (52.7 %) less than 100 Kb in size; 40 (37 %)

from 100 Kb to 500 Kb; 10 (9.2 %) in the range of 500 Kb to 1 Mb; 1 (0.9 %) larger than 1 Mb. Sixty-eight of these were deletions and 40 were duplications.

We applied the following algorithm in the interpretation of the CNVs of unknown clinical significance. Unknown CNVs should be cross-referenced to catalogs of CNVs detected in healthy controls and affected individuals to assess the likelihood of pathogenicity of the genomic imbalance. Unfortunately, in Bulgaria there has been no intensive research on the structure, frequency and distribution of the CNVs in the Bulgarian population, which is why we made the interpretation based on data from other populations (3, 4, 5, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 23, 24). For the CNVs with no information, we made individual assessment of each single variant.

In some cases it is possible to include data from patients that have clinical manifestation. The hypothesis is that a particular, large CNV is deemed pathogenic and that the other variations detected in the affected patient, typically smaller in size, are less likely to contribute to the reason for the genetic disorder.

In our study, 25 of the unknown CNVs in the following loci: 18p11.21, 21q22.3, Xq21.2, 1p34.2, 1p31.1, 1p21.1, 15q26.2, 16p13.11, Xp22.2, 10q23.2, Xq23, 22q11.23, Xq21.1, 10q11.22, 8p23.1, 1q43, and 2p11.2, occurred in patients with large pathogenic variations, based on which we accepted them as probably normal. Nineteen unknown variations in three loci (2q37.3, 10q11.22, Xp22.33) occurred with a frequency higher than 5 %, suggesting that they were most probably benign for our sample (**Fig. 3**). For the other 64 variants we made the interpretation based on their size, nature (deletions or duplications) and gene content. Thirty-five of these did not contain OMIM genes and we, therefore, classified them as probably benign. Out of the remaining 29 variants containing OMIM genes, 24 were without OMIM disorder loci and were also considered as likely benign. The remaining five genomic aberrations associated with OMIM disorders were deletions. Of these, only one aberration could be directly related to the clinical phenotype of the patient, a 14-year-old boy with intellectual disability, epilepsy and developmental delay. This gave us a reason to propose that the deletion in Xq22.1 is potentially pathogenic (**Table 1**).

The obtained results demonstrate that there is an obvious need for large population studies and detailed maps of variations in the Bulgarian population. This would facilitate an extremely precise interpretation of genomic imbalances of unknown nature in a clinical aspect. Moreover, it would help microchip-based diagnostic practices to become widely introduced not only in postnatal diagnosis of individuals with developmental delay and dysmorphism, but also in prenatal genetic diagnosis.

Conclusions

In a selected group of 52 Bulgarian patients with congenital malformations, 232 CNVs were proven: 124 benign and 108 with unknown clinical significance (mean number of CNVs per patient - 4.5). Only one aberration, the deletion in Xq22.1, could be directly related to a particular clinical phenotype: intellectual disability, epilepsy and developmental delay in a 14-year-old boy. Larger population studies and detailed maps of variations in the Bulgarian population are needed for more precise interpretation of genomic imbalances of unknown nature and for promoting the widespread introduction of microchip-based diagnostic practices not only in postnatal diagnosis of individuals with developmental delay and dysmorphism, but also in prenatal genetic diagnosis.

Acknowledgements

This study was supported by Grant 02/76-21.12.2009, National Science Fund of Bulgaria.

REFERENCES

1. **Alaerts M. Del-Favero J.** (2009) Hum. Mutat., **30**(8), 1139-1152.
2. **Cohen J.** (2007) Science, **317**(5843), 1315-1317.
3. **Conrad D.F., Andrews T.D., Carter N.P., Hurles M.E., Pritchard J.K.** (2006) Nat. Genet., **38**(1), 75-81.
4. **Cooper G.M., Zerr T., Kidd J.M., Eichler E.E., Nickerson D.A.** (2008) Nat. Genet., **40**(10), 1199-1203.
5. **de Smith A.J., Tsalenko A., Sampas N., Scheffer A., et al.** (2007) Hum. Mol. Genet., **16**(23), 2783-2794.
6. **Estivill X., Armengol L.** (2007) PLoS Genet., **3**(10), 1787-1799.
7. **Iafrate A.J., Feuk L., Rivera M.N., Listewnik M.L., et al.** (2004) Nat. Genet., **36**(9), 949-951.
8. **Itsara A., Cooper G.M., Baker C., Girirajan S., et al.** (2009) Am. J. Hum. Genet., **84**(2), 148-161.
9. **Kidd J.M., Cooper G.M., Donahue W.F., Hayden H.S., et al.** (2008) Nature, **453**(7191), 56-64.
10. **Kim K.Y., Lee G.Y., Kim J., Jeung H.C., et al.** (2009) Genomics, **94**(5), 317-323.
11. **Levy S., Sutton G., Ng P.C., Feuk L., et al.** (2007) PLoS Biol., **5**(10), e254.
12. **Locke D.P., Sharp A.J., McCarroll S.A., McGrath S.D., et al.** (2006) Am. J. Hum. Genet., **79**(2), 275-290.
13. **McCarroll S.A., Kuruvilla F.G., Korn J.M., Cawley S., et al.** (2008) Nat. Genet., **40**(10), 1166-1174.
14. **Perry G.H., Yang F., Marques-Bonet T., Murphy C., et al.** (2008) Genome Res., **18**(11), 1698-1710.
15. **Pinto D., Marshall C., Feuk L., Scherer S.W.** (2007) Hum. Mol. Genet., 16 (Spec. 2), R168-R173.
16. **Redon R., Ishikawa S., Fitch K.R., Feuk L., et al.** (2006) Nature, **444**(7118), 444-454.
17. **Sebat J., Lakshmi B., Troge J., Alexander J., et al.** (2004) Science, **305**(5683), 525-528.
18. **Shaikh T.H., Gai X., Perin J.C., Glessner J.T., et al.** (2009) Genome Res., **19**(9), 1682-1690.
19. **Sharp A.J., Locke D.P., McGrath S.D., Cheng Z., et al.** (2005) Am. J. Hum. Genet., **77**(1), 78-88.
20. **Venter J.C.** (2010) Nature, **464**(7289), 676-677.
21. **Wang K., Li M., Hadley D., Liu R., et al.** (2007) Genome Res., **17**(11), 1665-1674.
22. **Wheeler D.A., Srinivasan M., Egholm M., Shen Y., et al.** (2008) Nature, **452**(7189), 872-876.
23. **Wong K.K., deLeeuw R.J., Dosanjh N.S., Kimm L.R., et al.** (2007) Am. J. Hum. Genet., **80**(1), 91-104.
24. **Zogopoulos G., Ha K.C., Naqib F., Moore S., et al.** (2007) Hum. Genet., **122**(3-4), 345-353.