Multiple occurring copy number variants that matter

Hýblová Michaela[1,2,3], Andrej Gnip[3], Kucharík Marcel[4], Budiš Jaroslav[4], Minárik Gabriel[1,2,3]

[1]Medirex Group Academy, Bratislava, Slovakia; [2]Trisomy test, Bratislava, Slovakia; [3]Medirex, Bratislava, Slovakia, [4]Geneton, Bratislava, Slovakia

Introduction

Non invasive prenatal testing (NIPT) has proven its efficacy, reliability, and high positive predictive value (PPV) in common aneuploidy screening, sex chromosomal aneuploidy, or even copy number variants (CNVs). Simultaneously, it can be a precious source of information on maternal genome variation. Large CNVs are usually not an issue to interpret, however, many small and frequent CNVs remain ambiguous. CNVs from 200bp repeating two or more times have been referred to as multiple occurring variants (MOVs).

Aim

We analyzed data from more than 6400 NIPT results with a resolution of 200 Kb for maternal genome and identified 2484 CNVs, while 322 occurred repeatedly. We scrutinized the 20 most common ones, compared them with frequencies in population databases gnomAD SVs v2.1, dbVAR, DGV, and analyzed them with an emphasis on genomic content. The frequency of observed variants per megabase was calculated for each chromosome. The normal distribution of this value among chromosomes was assumed, mean and standard deviation were estimated.

Materials

This retrospective study identified recurrent maternal findings on a NIPT basis. Pregnant women from Slovakia, Czechia, and Hungary undertook our in-house NIPT

Methods

Blood drawn (from 10 gestational weeks)

plasma DNA isolation (Qiagen)

DNA library construction via modified TruSeq NaNo protocol (Illumina) (ref. 1,2)

Pair-ended whole-genome sequencing with low coverage on NextSeq 500/550 (Illumina)

Bioinformatic analysis (Trisomy test SW) (ref. 3)

Statistics - chi-squared statistical test

Tools: X-CNV (ref. 4), ACMG, Lift-Over (UCSC Genome Browser)

Databases: gnomAD SVs v2.1, dbVAR, DGV, DECIPHER

Results

We identified 322 distinct MOVs. We have aimed at the 20 most recurrent MOVs in our dataset which occurred 12 times (dup 4q35.2; del 4q35.2) to 126 times (dup 6q27) (Table 1). The frequency of the most common dup 6q27 is 1,85% to 0,18% for dup/del 4q35.2. The first three MOVs 6q27, dup 22q11.22 and dup Xp22.31 are polymorphisms as their frequency is higher than 1%: 1.85%, 1.52%, and 1.13%, respectively. The European gnomAD SVs v2.1 frequency is comparable to 4 CNVs. We found no match in gnomAD SV in 4 cases for dup 22q11.22, dup 15q11.2, dup 12q24.13-q24.21, and dup 3p26.3. The chromosome with the largest number of variants was chr2 (71) and the least chr19 (8) (Table 2).

This work was supported by the OPII programme as the project - Center for biomedical research - BIOMEDIRES II. phase, ITMS 313011W428, co-financed by the ERDF.

michaela.hyblova@medirexgroup.sk

Conflict of interest to declare: All co-authors are employees of the respective companies, but not shareholders

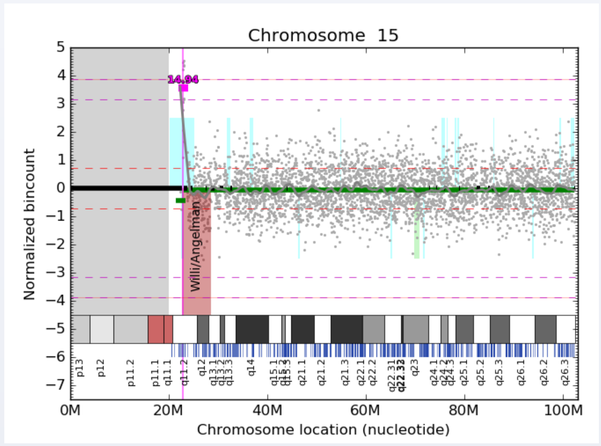

Figure 1. Visualization of CNV for chromosome 15. Normalized read counts per bin are depicted as gray dots. The dup 15q11.2 microduplication approximately of 320 kb is shown by the magenta horizontal line. The light gray vertical band depicts an unmappable centromere region. Black horizontal band signify bins that did not pass quality metrics (centromere) and are thus excluded from the analysis. The approximated z-score of duplication is displayed over the magenta segment. The estimated level of maternal aberration is 92% visualized as a magenta dashed line while the red dashed line represents the estimated level of fetal aberration detection.

Table 1. List of the most common MOVs, frequencies and statistics

			occurence	frequency (%)			Chi-squared comparison p-value	
MOVs	size (kb)	protein coding genes	n	our NIPT database (N 6422)	dbVAR/ DGV (NF Europe) (N 11222)	gnomAD SV v2.1 European % (N 7642)	*Our data vs. dbVAR / DGV*	*Our data vs. gnomAD*
dup 6q27	240	*AFDN, FRMD1, KIF25*	119	1.85	3.1	1,4	8.77E-07	4.11E-02
dup 22q11.22	260	*IGLV11-55, IGLV4-60, IGLV4-69, IGLV6-57, IGLV8-61, PPM1F, TOP3D (7)*	98	1.53	0.7	no match	1.98E-07	8.00E-27
dup Xp22.31	580	*PUDP, STS, VCX*	73	1.14	0.1	0	2.56E-21	3.09E-20
dup 8p23.2	200	-	58	0.9	1	0,48	5.90E-01	3.65E-03
dup 11q25	360	-	55	0.86	0.9	0,39	8.36E-01	6.38E-04
dup 15q13.3	480	*OTUD7A, CHRNA7*	38	0.59	0.8	0,23	1.40E-01	1.39E-03
dup 15q11.2	320	*NIPA1, NIPA2, CYFIP1, TUBGCP5*	35	0.54	0.3	no match	1.54E-02	3.30E-10
dup 1q25.1	320	*TNR*	30	0.47	0.1	0.06	2.55E-06	4.53E-06
del 17q22	340	-	28	0.44	0	0.026	1.26E-11	4.26E-07
del 9p23	260	-	27	0.42	1.2	0.01	3.02E-07	1.98E-07
dup 12p11.1	500	-	24	0.37	0.1	0.026	1.74E-04	2.79E-07
dup 12q24.13-q24.21	280	*RBM19*	20	0.31	no match	no match	1.55E-08	3.29E-06
dup 19q13.41	320	*FPR1, FPR3, ZNF577, ZNF649, ZNF613, ZNF350, ZNF615, ZNF614, ZNF432, ZNF841*	20	0.31	0.4	0.026	4.47E-01	2.70E-03
dup 6p11.2	600	-	20	0.31	0.1	0.026	2.35E-03	5.26E-05
dup 7q11.21	200	*ZNF92*	19	0.30	0.4	0.26	3.51E-01	8.30E-01
dup2p22.3	680	*BIRC6, TTC27, LTBP1*	18	0.28	0.1	0.18	7.91E-03	1.14E-05
dup 14q21.2	440	-	17	0.26	0.4	0.15	1.99E-01	2.27E-01
dup 3p26.3	360	*CNTN4*	16	0.25	no match	no match	5.53E-07	3.96E-05
del 15q11.2	240	*NIPA1, NIPA2, CYFIP1, TUBGCP5*	13	0.20	0.1	0.18	1.15E-01	9.52E-01
del 4q35.2	1600	*ZFP42, TRIML2, TRIML1*	12	0.19	0	0.026	2.04E-05	6.20E-03
dup 4q35.2	480	*ZFP42, TRIML2, TRIML2*	12	0.19	0	0	2.04E-05	4.90E-04

MOVs - multiple occuring variants; dup - duplication; del - deletion; NF - non Finnish; n - number of CNVs in our database; N - number of participants in each database/cohorts (NIPT, dbVAR/DGV European, and gnomAD SVs v2.1); p-value in red means statistically significant

Table 2. Number of MOVs according to the chromosomes, p-values

Chromosome	Effective length in bp (hg38)	Variants	Variants per megabase	p-value
1	231223641	64	0.276788306	0.309486126
2	240863511	71	0.294772752	0.39856174
3	198255541	67	0.337947679	0.625386241
4	189962376	58	0.305323618	0.453773416
5	181358067	55	0.303267458	0.442908357
6	170078524	60	0.352778226	0.697688611
7	158970135	60	0.377429383	0.801517446
8	144768136	54	0.37301026	0.78465846
9	122084564	41	0.335832792	0.614630158
10	133263006	45	0.33767811	0.624020457
11	134634058	37	0.274819021	0.300273823
12	133137821	34	0.255374466	0.216709904
13	97983128	29	0.295969322	0.404743508
14	91660769	31	0.338203578	0.626681306
15	85089576	26	0.305560343	0.455026684
16	83378703	32	0.38379105	0.824340228
17	83481871	24	0.287487567	0.36152445
18	80089650	24	0.299664189	0.4239761
19	58440758	8	0.136890764	0.008989368
20	63944268	15	0.234579275	0.144314325
21	40088623	9	0.224502598	0.115899249
22	40181019	14	0.348423219	0.677094783
X	154893034	84	0.542309734	0.998853971
		Mean	0.314017553	
		Standard deviation	0.074861628	
p ≤ 0.05 statistically significant				

Conclusion

We utilized NIPT analyses of pregnant women as a source of population data on recurrent CNVs (MOVs). Compared with notorious population databases, the frequency often differs significantly even among non-Finnish Europeans. Some MOVs seem to be candidates for population CNVs specific to the central Europe region. Based on the mappable chromosomal size, we did not record a significant difference in the occurrence of MOVs except for chromosome 19. Some MOVs have a potential for medical consequences, however with low penetrance and expressivity.

References:

1. Minarik G, Repiska G, Hyblova M, Nagyova E, Soltys K, Budis J, Duris F, Sysak R, Gerykova Bujalkova M, Vlkova-Izrael B, Biro O, Nagy B, Szemes T. Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance. PLoS One. 2015 Dec 15;10(12):e0144811.
2. Hyblova M, Harsanyova M, Nikulenkov-Grochova D, Kadlecova J, Kucharik M, Budis J, Minarik G. Validation of Copy Number Variants Detection from Pregnant Plasma Using Low-Pass Whole-Genome Sequencing in Noninvasive Prenatal Testing-Like Settings. Diagnostics (Basel). 2020 Aug 8;10(8):569. doi: 10.3390/diagnostics10080569. PMID: 32784382; PMCID: PMC7460070.
3. Zhao C, Tynan J, Ehrich M, Hannum G, McCullough R, Saldivar JS, Oeth P, van den Boom D, Deciu C. Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. Clin Chem. 2015 Apr;61(4):608-16
4. Zhang L, Shi J, Ouyang J, Zhang R, Tao Y, Yuan D, Lv C, Wang R, Ning B, Roberts R, Tong W, Liu Z, Shi T. X-CNV: genome-wide prediction of the pathogenicity of copy number variations. Genome Med. 2021 Aug 18;13(1):132. doi: 10.1186/s13073-021-00945-4. PMID: 34407882; PMCID: PMC8375180.