Targeted sequencing of 242 clinically important genes in the Russian population from the Ivanovo region

Vasily E. Ramensky, Alexandra I. Ershova, Marija Zaicenoka, Anna V. Kiseleva, Anastasia A. Zharikova, Yuri V. Vyatkin, Evgenia A. Sotnikova, Irina A. Efimova, Mikhail G. Divashuk, Olga V. Kurilova, Olga P. Skirko, Galina A. Muromtseva, Olga A. Belova, Svetlana A. Rachkova, Maria S. Pokrovskaya, Svetlana A. Shalnova, Alexey N. Meshkov, Oxana M. Drapkina

Table Captions

Supplementary Table 1. Overview of 242 targeted genes. Columns:
Variant and carrier counts for each gene. Path: KP variants; VUS: variants of unknown significance; PTV: protein-truncating variants; MisSDam: missense strictly damaging; MisOther: other missesne variants; Inframe: inframe indels; Other, UTR, intron, synonymous etc. variants. ACMG59: genes from ACMG59 gene list; oe_syn_upper, oe_mis_upper, oe_lof_upper: gene constraint metrics based on variant counts in gnomAD (Karczewski et al., 2020). Inheritance: inheritance model; Phenotype: associated disorders with OMIM identifier (if applicable).

Supplementary Table 2. Novel variants with allele count AC>2. Columns:

1	#Chr	Chromosome
2	Pos	Position
3	Variant	Variant id
4	Ref	Reference allele
5	Alt	Alternative allele
6	Feature	Transcript id (VEP)
7	Consequence	Consequence type: missense, splice_acceptor_variant, etc. (VEP)
8	Existing_variation	dbSNP id (VEP)
9	IMPACT	Subjective impact classification of consequence type (VEP)
10	LoF	Quality prediction for high impact variants: HC=High Quality, LC=Low Quality, OS=other splice variants (VEP, LOFTEE plugin)
11	SIFT	Effect prediction for missenses
12	PolyPhen	Effect prediction for missenses
13	STRAND	Transcript strand (VEP)
14	VARIANT_CLASS	SNV/insertion/deletion
15	SYMBOL	Gene symbol name (HGNC, etc.) (VEP)
16	BIOTYPE	Biotype of transcript or regulatory feature (VEP)
17	ACMG59	ACMG59/other
18	CANONICAL	Indicates if transcript is canonical for this gene: YES/- (VEP)

19	HGVSc	HGVS coding sequence name (VEP)
20	HGVSp	HGVS protein sequence name (VEP)
21	VarSome	VarSome classification: P, pathogenic; LP, likely pathogenic; VUS: variant of unknown significance; LB, likely benign; B, benign
22	1000G.AF	Alt.allele frequency in 1000 Genomes combined population (VEP)
23	MAX_AF_POPS	Populations in which maximum frequency was observed (VEP)
24	MAX_AF	Maximum alt.allele frequency in 1000 Genomes, ESP and ExAC/gnomAD (VEP)
25	gnomADe.AF	Alt.allele frequency in gnomAD exomes combined population (gnomAD exomes)
26	gnomADe.popmax	GnomAD exomes population in which maximum frequency was observed (gnomAD exomes)
27	gnomADe.AF_popmax	Maximum alt.allele frequency in gnomAD exomes populations (gnomAD exomes)
28	gnomADe.AF_nfe	Alt.allele frequency in European non-Finnish gnomAD exomes (gnomAD exomes)
29	gnomADg.AF	Alt.allele frequency in gnomAD exomes combined population (gnomAD exomes)
30	gnomADg.popmax	GnomAD exomes population in which maximum frequency was observed (gnomAD exomes)
31	gnomADg.AF_popmax	Maximum alt.allele frequency in gnomAD exomes populations (gnomAD exomes)
32	gnomADg.AF_nfe	Alt.allele frequency in European non-Finnish gnomAD exomes (gnomAD exomes)
33	ClinSigFull	Clinical significance (ClinVar)
34	ClinSigShort	Simplified ClinSig: Pathogenic, Benign, VUS (ClinVar), Conflict
35	PhenIDS	Phenotype IDs (ClinVar)
36	PhenList	Phenotype list (ClinVar)
37	Origin	Origin: somatic/germline (ClinVar)
38	ReviewStatus	Review status (ClinVar)
39	NumSubmitters	Number of submitters (ClinVar)
40	Inheritance	Inheritance (MedGen)
41	NWR.AF	AF in 694 NorthWest regions (Barbitoff 2019)
42	NWR.AC	AC in 694 NorthWest regions (Barbitoff 2019)
43	NWR.AN	AN in 694 NorthWest regions (Barbitoff 2019)
44	QUAL	Variant quality (GATK)
45	FILTER.hg37	Variant filter with hg37 reference (GATK)
46	FILTER.hg38	Variant filter with hg38 reference (GATK)
47	AF	Allele frequency in our sample (GATK)
48	AC	Allele count in genotypes (GATK)
49	AN	Total number of alleles in called genotypes (GATK)
50	NumHomAlt	Number of homalt samples (bcftools)

51	Region	Low complexity genome region? (SDUST)
52	MultiSite	Multiallelic site?
53	MaxGQ	Maximum GQ among alt. allele carriers (GATK)
54	AveGQ	Average GQ among alt. allele carriers (GATK)
55	Phenotype evaluation	Variant selected for phenotype evaluation of alt. allele carriers

Supplementary Table 3. Known pathogenic or likely pathogenic, novel or rare protein truncating variants. Columns are as in Supplementary Table 2.

Supplementary Table 4. Known pathogenic variants with frequencies in the Ivanovo population significantly exceeding the non-Finnish Europeans; extended version of Table 5.

Variant, dbSNP rsID; HGVS: variant description; VarSome, automated variant classification according to the ACMG guidelines; gnomAD, allele frequenct in gnomAD NFE genomes and exomes considered together; Ivanovo AC, allele counts in the Ivanovo population; Ivanovo AF, allele counts in the Ivanovo population; Ivanovo/gnomAD is the ratio of allele frequencies in our sample and gnomAD NFE genomes and exomes considered together; P-value, Fisher test on direct allele counts, raw and Bonferroni corrected; CI for Ivanovo AF; NWR, variant frequency in 694 northwest exomes (Barbitoff et al., 2019).