DOI: 10.18413/2658-6533-2022-8-3-0-2

C9orf16 (BBLN) gene, encoding a member of Hero proteins, is a novel marker in ischemic stroke risk

Ksenia A. Kobzeva[1], Irina V. Shilenok[2], Andrey E. Belykh[1,3], Denis E. Gurtovoy[1], Lyubov A. Bobyleva[1], Anna B. Krapiva[1], Tatiana A. Stetskaya[1], Marina A. Bykanova[1], Anastasiya A. Mezhenskaya[4], Ekaterina A. Lysikova[5], Maxim B. Freidin[6,7], Olga Y. Bushueva[1]

[1] Kursk State Medical University,
3 Karl Marx St., Kursk, 305041, Russia
[2] Kursk City Clinical Emergency Hospital,
14 Pirogova St., Kursk, 305035, Russia
[3] Nencki Institute of Experimental Biology of the Polish Academy of Sciences,
3 Pasteur St., Warsaw, 02-093, Poland
[4] Alekseevskaya Central District Hospital,
2 Nikolskaya St., Alekseevka, 309850, Russia
[5] Institute of Gene Biology, Russian Academy of Sciences,
34/5 Vavilov St., Moscow, 119334, Russia
[6] Queen Mary University of London,
Mile End Road, London, E1 4DQ, United Kingdom
[7]Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences,
10 Nabereznaya *Ushaiki* street, *Tomsk,* 634050, Russia
Corresponding author: Olga Y. Bushueva (olga.bushueva@inbox.ru)

Abstract

Background: Ischemic stroke (IS) is the leading cause of death and disability worldwide. Chaperone proteins protect brain cells from the ischemic damage by restoring the structures of damaged proteins. Chaperone C9orf16 (also known as BBLN) belongs to the class of heat-resistant obscure (HERO) proteins, characterized by the ability to stabilize various proteins, suppress neurotoxicity and reduce proteotoxic stress. In this regard, it may play a potentially significant role in the risk of development and clinical manifestations of IS. **The aim of the study:** To investigate an association between a single nucleotide polymorphism rs2900262 in the gene encoding *C9orf16* and predisposition to IS. **Materials and methods:** A total of 897 patients with IS and 1140 healthy controls were recruited for the study. Genotyping was done using a probe-based genotyping assay. Multiple logistic regression analysis was performed to evaluate the associations of the rs2900262 genotypes with the risk of IS and ischemic events. Dominant, recessive and additive models of associations of genotypes were analyzed. Adjustment for sex, age, and smoking was done throughout. Benjamini-Hogberg false-discovery rate was used to correct for multiple comparisons. **Results:** The rs2900262*T allele was found to be associated with the increased risk of IS exclusively in females (dominant model:

OR=1.74, 95% CI=1.07-2.82, $P^{FDR=}$0.042; additive model: OR=1.69, 95% CI=1.06-2.71, $P^{FDR=}$0.042). Additional analysis showed that the rs2900262*T is associated with the increased risk of IS in smokers only (dominant model: OR=1.92, 95% CI=1.09-3.37, $P^{FDR=}$0.042; additive model: OR=1.79, 95%CI=1.04-3.08, $P^{FDR=}$0.042). Also, we demonstrated that C/T-T/T genotype carriers exhibit an earlier manifestation of IS (59.53±1.12 years) compared to the C/C genotype carriers (61.63±0.4 years); mean difference=-1.98; 95% CI=-3.61 – -0.36; P^{FDR}=0.026. **Conclusion:** This study is the first in the world to demonstrate the possible contribution of the rs2900262 *C9orf16* gene polymorphism to the risk of ischemic stroke.
Keywords: ischemic stroke; Hero; heat-resistant obscure; chaperones; rs2900262; *C9orf16; BBLN*

For citation: Kobzeva KA, Shilenok IV, Belykh AE, et al. *C9orf16 (BBLN)* gene, encoding a member of Hero proteins, is a novel marker in ischemic stroke risk. Research Results in Biomedicine. 2022;8(3):278-292. DOI: 10.18413/2658-6533-2022-8-3-0-2

Background. Despite the significant efforts made by the world's medical community in the field of prevention and treatment of stroke, this disease remains the second leading cause of death worldwide, and a leading cause of adult disability [1], cognitive decline and dementia [2]. Ischemic stroke (IS) comprises up to 85% of all cases [3]. The risk of IS is determined by a complex interplay of genetic and environmental factors partly acting through modifiable risk factors such as high blood pressure [4, 5].

The previous studies have already revealed genetic markers associated with the development of IS [6, 7, 8], as well as related phenotypes/risk factors, such as hypertension [9, 10, 11], atherosclerosis [12, 13], and thrombosis [14]. However, despite the large number of studies conducted in the world, the molecular mechanisms of IS development remain poorly understood.

One of the most promising fields in this area is the study of chaperone proteins. A critical decrease in perfusion in IS leads to hypoxic damage to brain tissues [15]. Cascades of ischemic and reperfusion damage to neurons and glia include hundreds of individual links that lead to cell death. It is the critical accumulation of damaged enzymes and structural proteins that causes irreversible changes that lead to necrosis or programmed cell death [16]. The main function of chaperones is to restore the native tertiary and quaternary structure of damaged proteins. There is already enough information that confirms the neuroprotective role of chaperones in IS, for example, the HSP class of chaperones ("heat shock" proteins) [17].

Recently, a new class of chaperones, "Hero proteins", has been discovered, which main mechanism of action is the physical shielding of other proteins. Such a function is provided by a flexible structure and a high negative or positive charge of molecules [18]. Moreover, Hero proteins can suppress some forms of pathogenic protein aggregates in cells and in animal models of neurodegenerative diseases in vivo. The described properties of Hero let us suggest their wide involvement in the physiological and pathological processes occurring in cells with ischemic damage.

The aim of this study was to analyze association of the tagger SNP rs2900262 in gene *C9orf16* encoding the member of Hero proteins with the risk and clinical manifestations of ischemic stroke.

Material and methods. A total of 1929 unrelated Russians (861 patients with IS and 1068 healthy individuals) from Central Russia were recruited into the study. The Ethical Review Committee of Kursk State Medical University approved the study protocol. All the participants gave written informed consent before the enrollment in this study, subject to the following inclusion criteria: self-declared Russian descent, a birthplace inside of Central Russia.

Baseline and clinical characteristics of the study population are listed in Table 1. The patients were enrolled in to the study in two periods: from the Regional Vascular Center of Kursk Regional Clinical Hospital between 2015 and 2017 [19, 20] and Neurology Clinics of Kursk Emergency Medicine Hospital Kursk

between 2010 and 2012 [21, 22]. All the patients were examined by qualified neurologists. The diagnosis of IS was made in the acute phase of stroke, according to the results of the neurological examination and computed tomography and/or magnetic resonance imaging of the brain. The patients were recruited consecutively. The IS patients were enrolled under the following exclusion criteria: hepatic or renal failure or endocrine, autoimmune, oncological, or other diseases that can cause an acute cerebrovascular event; intracerebral hemorrhage, hemodynamic or dissection-related stroke, traumatic brain injury. All the patients with IS had a history of hypertension and received antihypertensive therapy.

The control group was compiled from healthy volunteers with no clinical evidence of cerebrovascular, cardiovascular, or other chronic diseases and with normal blood pressure without antihypertensive therapy. Healthy individuals were included in control group if they had a systolic blood pressure less than 130 mm Hg and/or a diastolic blood pressure less than 85 mm Hg on at least 3 separate measurements. Control subjects were enrolled from Kursk hospitals during periodic medical examinations at public institutions and industrial enterprises of Kursk region. This group was recruited from the same population and during the same period [23, 24].

Table 1

Baseline and clinical characteristics of the studied groups

Baseline and clinical characteristics		**IS patients (n=897)**	**Controls (n=1140)**	**p_{raw}**
Age, Me [Q1; Q3]		63 [55; 70]	57 [53; 64]	**<0.001**
Gender	Males, N(%)	498 (55.5%)	508 (44.6%)	**<0.001**
	Females, N(%)	399 (44.5%)	632 (55.4%)	
Smoking	Yes, N (%)	442 (49.3%)	308 (27%)	**<0.001**
	No, N (%)	455 (50.7%)	832 (73%)	
Coronary artery disease	Yes, N (%)	113 (13.2%)	-	
	No, N (%)	746 (86.8%)	-	
Type 2 diabetes mellitus	Yes, N (%)	103 (12.1%)	-	
	No, N (%)	747 (87.9%)	-	
Family history of cerebrovascular diseases	Yes, N (%)	306 (35.9%)	ND	
	No, N (%)	546 (64.1%)	ND	
Age at onset of stroke, Me [Q1; Q3]		61 [54; 70], n=869	-	
Number of strokes including event in question	1, N (%)	774 (89.1%)	-	
	2, N (%)	82 (9.4%)	-	
	3, N (%)	13 (1.5%)	-	
Stroke localization	Right/left middle cerebral artery basin, N (%)	723 (83.2%)	-	
	Vertebrobasilar basin, N (%)	146 (16.8%)	-	
Area of lesion in stroke, mm^2, Me [Q1; Q3]		104 [30; 468], n=850	-	
Total cholesterol, mmol/L, Me [Q1; Q3]		5.2 [4.4; 5.9], n=583	ND	
Triglycerides, mmol/L, Me [Q1; Q3]		1.3 [1.1; 1.8], n=575	ND	
Glucose level, mmol/L, Me [Q1; Q3]		4.8 [4.3; 5.5], n=861	ND	
Prothrombin time, seconds, Me [Q1; Q3]		10.79 [10.14; 11.7], n=845	ND	
International normalized ratio, Me [Q1; Q3]		1 [0.93; 1.09], n=566	ND	
Activated partial thromboplastin time, seconds, Me [Q1; Q3]		32.6 [29; 37], n=569	ND	

Note: statistically significant differences between groups are indicated in bold; ND – no data.

The following criteria were used in the selection of SNPs: the SNP must be tagging, have a minor allele frequency of at least 0.05 in the European population and be characterized by a high regulatory potential. According to the bioinformatic tool GenePipe (https://snpinfo.niehs.nih.gov/snpinfo/selegene.html), which was used to select SNPs based on the reference haplotypic structure of the Caucasian population (CEU) of the project HapMap, the only tag SNP in the gene *C9orf16* (also known as *BBLN*, ID:79095) is rs2900262. This genetic variant is localized in the intron. Several bioinformatic resources were used to assess the regulatory potential of this SNP. Bioinformatic resource HaploReg (v4.1) established a connection of rs2900262 *C9orf16* with modifications of histones marking promoters in 15 different tissues; it established the localization of this SNP in the regions of hypersensitivity to deoxyribonuclease in 7 issues, regions of 4 regulatory motifs (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php). According to rSNPBase (http://rsnp.psych.ac.cn/index.do), it is characterized by distal transcriptional regulation, the regulation mediated by RNA-binding proteins and regulation mediated by microRNA. The instrument RegulomeDB allowed to reveal that rs2900262 *C9orf16* is characterized by a regulatory coefficient of 1b (eQTL + TF binding + any motif + DNase Footprint + DNase peak) (http://regulome.stanford.edu/). According to the date presented in Ensembl genome browser, this genetic variant is characterized by an average frequency of the minor T allele in European populations of 0.057 (https://www.ensembl.org/). Thus, rs2900262 *C9orf16* was selected for this study, which meets the necessary criteria for inclusion in the study.

Genetic Analysis. DNA analysis was carried out at the Laboratory of Genomic Research of Research Institute for Genetic and Molecular Epidemiology of Kursk State Medical University (Kursk, Russia). Approximately 5 mL of venous blood was collected from the cubital vein of each participant into EDTA-coated tubes and maintained at −20◦C until processed. Genomic DNA was extracted from thawed blood samples by the standard procedure of phenol/chloroform extraction and ethanol precipitation. Genotyping of the SNP was done using Taq-Man–based real-time polymerase chain reaction assays according to the protocol designed in the Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology. Primers and probes were designed (http://primer3.ut.ee/) using the Primer3 online program [25], and then synthesized by the Syntol company (Moscow, Russia). The 2 flanking primers such as forward 5′- ATATACCATGAGGGGGTTGGA -3′ and reverse 5′- GAGGAGCTGGAGGAGCGTAT-3′ and also allele-specific probes T-5′-FAM-CCTTTCCCAGAACACARTQ1-3′ and 5′-ROX-CCTTTCCTAGAACACA-BHQ2-3′ were used for genotyping of the polymorphism. A real-time PCR was done in a 25-mL reaction mixture containing 1.5 U of Hot Start Taq DNA polymerase (Biolabmix, Novosibirsk, Russia), about 1 μg of DNA, 0.25 μM each primer, 250 μM dNTPs, 3.0 mM $MgCl_2$, 1xPCR buffer [67 mM Tris-HCl, pH 8.8, 16.6 mM $(NH4)_2SO_4$, 0.01% Tween-20]. The amplification reaction consisted of an initial denaturation for 10 minutes at 95°C, followed by 39 cycles of 92°C for 30 seconds and 58°C for 1 minute. Figure 1 shows allelic discrimination plot for SNP rs2900262 *C9orf16* assay designed for this study. The plot shows clear separation between the signals derived from allele 1 (rs2900262-C, FAM fluorescent dye) or allele 2 rs2900262-T, ROX fluorescent dye). Genotypes CC, CT, and TT are shown as circles, triangles and squares, respectively. Primer sequences used for genotyping are available upon request. To ensure quality control, 10% of DNA samples were genotyped in duplicates blinded to the case-control status. The concordance rate was >99%.

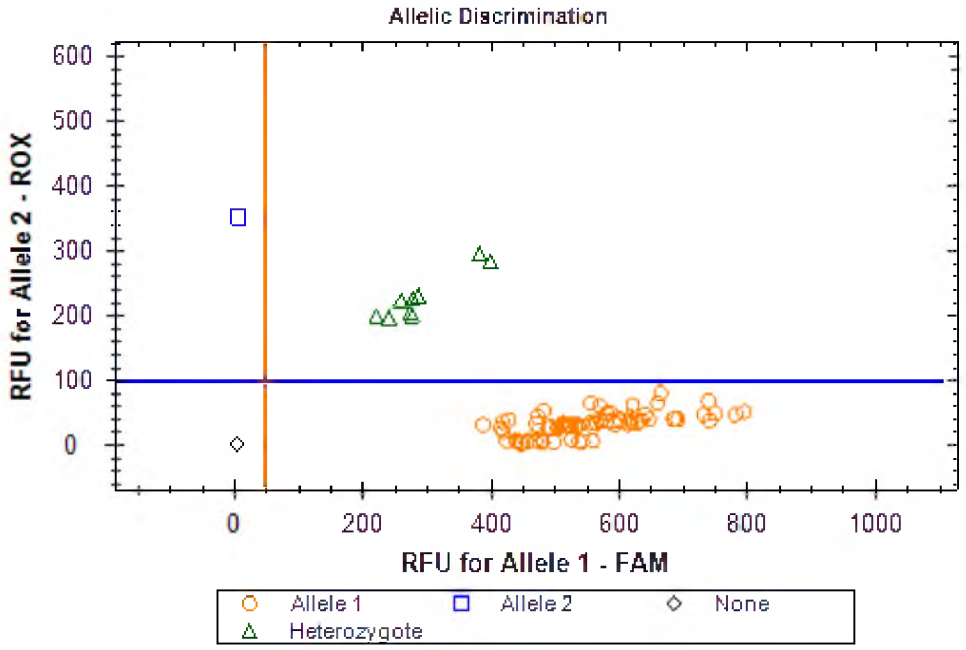

Fig. 1. Allelic discrimination plot for SNP rs2900262 *C9orf16* assay designed for this study.

Allele and genotype frequencies in the case and control groups were counted and compared by the Fisher exact test to identify significant departures from Hardy–Weinberg equilibrium. Genotype frequencies in the study groups and their associations with disease risk were analyzed using the SNPStats software (https://www.snpstats.net/start.htm) [26]. For the analysis of associations of genotypes, dominant, recessive, and additive models were considered; taking into account the number of models of associations, a correction for multiple testing was applied using Benjamini-Hogberg false-discovery rate approach.

Multiple logistic regression analysis was performed to evaluate the associations of rs2900262 *C9orf16* genotypes with the risk of ischemic events. Continuous variables were analyzed for normal distribution by the Kolmogorov–Smirnov test. Linear regression analysis was used to evaluate the association between rs4644832 and age of ischemic events, which was previously transformed to a normal variable through the procedure of inverse transformation of ranks. All associations were adjusted for age, gender and smoking status.

The following bioinformatics recourses were used for analysis of regulatory potential of rs2900262 *C9orf16*:

- The rSNPBase tool (http://rsnp.psych.ac.cn/index.do) was used to evaluate the effect of SNPs on proximal transcriptional regulation, distal transcriptional regulation, RNA-binding protein-mediated regulation, and miRNA-mediated regulation (accessed on 7 May 2022) [27].

- RegulomeDB (Version 1.1) (http://regulome.stanford.edu/) was used to estimate the regulatory coefficient of SNP (accessed on 7 May 2022) [28].

- GTExportal (http://www.gtexportal.org/) was used to analyze the expression levels of the studied genes in brain, whole blood, and blood vessels, as well as to analyze the binding of SNPs to quantitative expression trait loci (eQTLs) (accessed on 7 May 2022) [29].

- Additionally, to analyze the binding of rs2900262 *C9orf16* to quantitative expression trait loci (eQTL) in peripheral blood, the eQTLGen resource was used (https://www.eqtlgen.org/) (accessed on 7 May 2022) [30].

- Bioinformatics resource HaploReg (v4.1) (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php) was used to assess the association of rs2900262 *C9orf16* with histone modifications marking promoters and enhancers: mono-methylation at the 4th lysine residue of the histone H3 protein (H3K4me1),

tri-methylation at the 4th lysine residue of the histone H3 protein (H3K4me3), the acetylation at the 9th lysine residues of the histone H3 protein (H3K9ac), acetylation of the lysine residues at N-terminal position 27 of the histone H3 protein (H3K27ac). This resource has also been used to analyze the localization of SNPs in regions of DNase hypersensitivity, regions of regulatory motifs, and sites binding with regulatory proteins (accessed on 7 May 2022) [31].

- To assess the effect of rs2900262 *C9orf16* on the binding of transcription factors (TFs) to DNA, depending on the carriage of the reference/alternative alleles, the atSNP Function Prediction online tool was used (http://atsnp.biostat.wisc.edu/search; accessed on 14 May 2022) [32]. TFs were included in the analysis only under the condition of a high and very high degree of influence of SNP on the interaction of TFs with DNA, calculated on the basis of a positional weight matrix. The subsequent analysis of the possible joint participation of TFs associated with the reference/SNP alleles in common biological processes directly related to the pathogenesis of IS was carried out using the Enrichr online tool (https://maayanlab.cloud/Enrichr/) [33]. Biological functions controlled by transcription factors associated with SNP rs2900262 *C9orf16* were used as functional groups (accessed on 14 May 2022). The level of significance of gene ontologies, taking into account multiple tests, was assessed by the Benjamini–Hochberg method (False Discovery Rate) [34].

- For bioinformatics analysis of interactions between the *C9orf16* gene and cigarette smoke components, the Comparative Toxicogenomics Database resource (CTD) available online, was used (http://ctdbase.org) (accessed on 16 May 2022) [35]. CTD provides the ability to analyze specific interactions between genes and chemicals in vertebrates and invertebrates based on data obtained from published scientific studies worldwide. This tool was used to analyze binary interactions involving one chemical and one gene/protein.

The Cerebrovascular Disease Knowledge Portal (CDKP) (https://cd.hugeamp.org/) was used for bioinformatic analyze of the associations of rs2900262 *C9orf16* with atherosclerosis-associated diseases, intermediate phenotypes, and risk factors for ischemic stroke (such as blood pressure, the level of triglycerides, the level of low density lipoproteins, etc.) (accessed on 18 May 2022) [36].

Results and discussion. The allele frequencies in the study patients according to the data published in the 1000 Genomes Project, Phase 3 (http://www.ensembl.org (accessed on 30 May 2022)) were comparable with those reported in following European populations: British in England and Scotland (GBR), Iberian populations in Spain (IBS), Toscani in Italy (TSI) (Table 2).

Table 2

Comparative analysis of allele frequencies between the population investigated in the present study and other European populations

Population	Minor allele	Present study (control)	European populations	χ^2	P-level
CEU	T	0.038	0.071	4.94	**0.03**
FIN	T	0.038	0.076	6.53	**0.01**
GBR	T	0.038	0.060	2.19	0.14
IBS	T	0.038	0.047	0.38	0.54
TSI	T	0.038	0.033	0.16	0.69

Note: Data on the frequency of the minor allele in European populations were obtained using the Ensembl genome browser (https://www.ensembl.org/index.html); statistically significant differences in the frequency of the minor allele are indicated in bold

The genotype frequencies of rs2900262 *C9orf16* in study groups are presented in Table 3. The distribution of genotypes frequencies corresponded to the Hardy-Weinberg equilibrium both in the group of patients and in the control group (P>0.05).

Analysis of the general group of patients/controls did not show an association of rs2900262 *C9orf16* with the risk of IS (Table 3).

Table 3

Genetic models of associations of rs2900262 *C9orf16* with the risk of ischemic stroke

M[1]	Genotypes	Controls (N=1140)	Patients (N=897)	corOR (95% CI)[2]	Padj[3]	P[FDR]
D	C/C	1054 (92.5%)	817 (91.1%)	1.00	0.33	0.44
	C/T-T/T	86 (7.5%)	80 (8.9%)	1.18 (0.84-1.66)		
R	C/C-C/T	1139 (99.9%)	895 (99.8%)	1.00	0.44	0.44
	T/T	1 (0.1%)	2 (0.2%)	2.52 (0.22-28.50)		
A	-	---	---	1.19 (0.86-1.65)	0.29	0.44

Note: 1 – genetic models of associations of genotypes: D – dominant; R – recessive; A-additive; 2 – odds ratio and 95% confidence interval adjusted by sex, age, smoking; 3-P – level adjusted by sex, age, smoking; P^{FDR} – P-level, after the correction on the number of models (calculated by the Benjamini-Hochberg method); All calculations were performed relative to the minor allele T; Statistically significant differences are marked in bold.

Due to a possible sex dimorphism in the associations of genetic variants with the risk of IS, an analysis of the relationship of rs2900262 *C9orf16* with the development of IS separately in males and females was carried out. It turned out that this genetic variant does not affect the risk of the disease in males. As shown in Table 4, SNP rs2900262 *C9orf16* was found to be associated with a increased risk of IS in females. In particular, according to the dominant model, genotypes C/T-T/T was associated with an increased risk of IS (corOR=1.74, 95%CI 1.07–2.82, Padj=0.025, P^{FDR}=0.042). According to the additive model, the rs2900262 genotypes were also associated with an increased risk of IS (corOR=1.69, 95%CI 1.06–2.71, Padj=0.028, P^{FDR}=0.042).

Table 4

Sex-specific associations of rs2900262 *C9orf16* with the development of ischemic stroke

Males						
M[1]	**Genotypes**	**Controls (N=508)**	**Patients (N=498)**	**corOR (95% CI)[2]**	**Padj[3]**	**P[FDR4]**
D	C/C	461 (90.8%)	457 (91.8%)	1.00	0.57	0.65
	C/T-T/T	47 (9.2%)	41 (8.2%)	0.88 (0.57-1.36)		
R	C/C-C/T	508 (100%)	497 (99.8%)	1.00	0.24	0.65
	T/T	0 (0%)	1 (0.2%)	NA (0.00-NA)		
A	-	-	-	0.91 (0.59-1.39)	0.65	0.65
Females						
M[1]	**Genotypes**	**Controls (N=632)**	**Patients (N=399)**	**corOR (95% CI)[2]**	**P[3]**	**P[FDR]**
D	C/C	593 (93.8%)	360 (90.2%)	1.00	0.025	**0.042**
	C/T-T/T	39 (6.2%)	39 (9.8%)	1.74 (1.07-2.82)		
R	C/C-C/T	631 (99.8%)	398 (99.8%)	1.00	0.83	0.83
	T/T	1 (0.2%)	1 (0.2%)	1.40 (0.06-32.79)		
A	-	-	-	1.69 (1.06-2.71)	0.028	**0.042**

Note: 1 – genetic models of associations of genotypes: D – dominant; R – recessive; A-additive; 2 – odds ratio and 95% confidence interval adjusted by age, smoking; 3-P – level adjusted by age, smoking; P^{FDR} – P-level, after the correction on the number of models (calculated by the Benjamini-Hochberg method); All calculations were performed relative to the minor allele T; Statistically significant differences are marked in bold.

The analysis of associations of rs2900262 *C9orf16* with of IS risk depending on the smoking status revealed that this genetic variant is associated with an increased risk of IS exclusively in smoking individuals according to the dominant model (OR=1.92, 95%CI 1.09–3.37, Padj=0.019, P^{FDR}=0.042). Analysis of the log-additive model also revealed the association of rs2900262 *C9ORF16* with the increased risk of IS in smokers (OR=1.79, 95%CI 1.04–3.08, Padj=0.028, P^{FDR}=0.042) (Table 5).

Table 5

The effect of smoking on the association of rs2900262 *C9orf16* with the risk of ischemic stroke

Non-smokers						
M[1]	**Genotypes**	**Controls (N=832)**	**Patients (N=455)**	**OR (95% CI)[2]**	**Padj[3]**	**P[FDR]**
D	C/C	764 (91.8%)	422 (92.8%)	1.00	0.56	0.66
	C/T-T/T	68 (8.2%)	33 (7.2%)	0.88 (0.57-1.35)		
R	C/C-C/T	832 (100%)	454 (99.8%)	1.00	0.15	0.45
	T/T	0 (0%)	1 (0.2%)	NA (0.00-NA)		
A	-	-	-	0.91 (0.59-1.39)	0.66	0.66
Smokers						
M[1]	**Genotypes**	**Controls (N=308)**	**Patients (N=442)**	**OR (95% CI)[2]**	**P[3]**	**P[FDR]**
D	C/C	290 (94.2%)	395 (89.4%)	1.00	0.019	**0.042**
	C/T-T/T	18 (5.8%)	47 (10.6%)	1.92 (1.09-3.37)		
R	C/C-C/T	307 (99.7%)	441 (99.8%)	1.00	0.8	0.8
	T/T	1 (0.3%)	1 (0.2%)	0.70 (0.04-11.17)		
A	-	-	-	1.79 (1.04-3.08)	0.028	**0.042**

Note: 1 – genetic models of associations of genotypes: D – dominant; R – recessive; A-additive; 2 – odds ratio and 95% confidence interval adjusted by sex, age; 3-P – level adjusted by sex, age; P^{FDR} – P-level, after the correction on the number of models (calculated by the Benjamini-Hochberg method); All calculations were performed relative to the minor allele T; Statistically significant differences are marked in bold.

Analysis of the effects of rs2900262 *C9orf16* on clinical manifestations of IS revealed that in carriers of genotypes C/T-T/T, ischemic stroke manifests on average 2 years earlier than in those of genotype C/C (Recessive model: Padj=0.017; Difference= -1.98, 95% CI=-3.61 – -0.36, Padj= 0.017, $P^{FDR=}$0.026). According to the additive model, rs2900262 *C9ORF16* is also associated with a decrease in the age of manifestation of IS (Difference= -1.82, 95% CI=-3.38 – -0.26, Padj= 0.015, $P^{FDR=}$0.026) (Table 6).

Table 6

Analysis of the impact of rs2900262 *C9orf16* on the age of ischemic stroke manifestation

M[1]	**Genotypes**	**N**	**Response mean (s.e.)**	**Difference (95% CI)**	**Padj**	**P[FDR]**
D	C/C	792	61.63 (0.4)	0.00	**0.017**	**0.026**
	C/T-T/T	78	59.53 (1.12)	-1.98 (-3.61 – -0.36)		
R	C/C-C/T	868	61.45 (0.37)	0.00	0.9	0.9
	T/T	2	60 (19)	0.59 (-9.11 – 10.29)		
A	-	---	---	-1.82 (-3.38 – -0.26)	**0.015**	**0.026**

Note: 1 – genetic models of associations of genotypes: D – dominant; R – recessive; A – additive; Padj – P-level adjusted by sex, age, smoking; P^{FDR} – P-level, after the correction on the number of models (calculated by the Benjamini-Hochberg method); All calculations were performed relative to the minor allele T; Statistically significant differences are marked in bold.

The *C9orf16* gene is expressed in brain tissues, blood vessels, and whole blood (Figure 2). In brain tissues, MeTPM varies from 50.20 to 172.2; in blood vessels – from 64.62 to 79.12; in whole blood MeTPM=50.32 (Fig. 2).

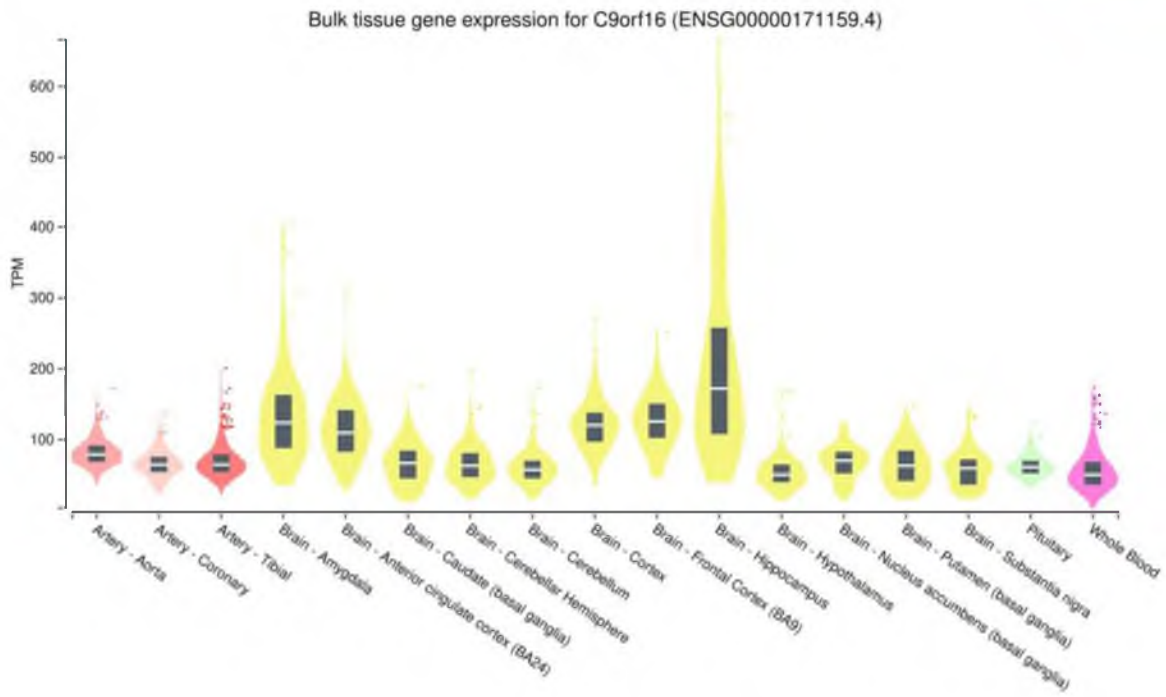

Fig. 2. Level of the expression of the *C9orf16* gene in brain, whole blood, and vessels (https://gtexportal.org)

C9orf16 (also known as *BBLN*), a gene with unknown functions, has not been investigated for the risk of IS. However, in a previous study, a change in the expression of this gene in the blood of patients with Parkinson's Disease was found, which indicates a possible significant role of this gene in neuroprotection [37].

Due to the fact that the use of PubMed and Google Scholar search resources did not allow us to detect studies aimed at studying the associations of the studied Tag SNP rs2900262 *C9orf16* with the development of IS, we used the bioinformatic resource Cardiovascular Disease Knowledge Portal to interpret the phenotypic effects of this genetic variant (CVDKP), (https://cvd.hugeamp.org), which combine and analyze the results of genetic associations of the largest consortia for the study of cardio- and cerebrovascular diseases, as well as their intermediate phenotypes. It was found that the protective C allele is associated with a decrease in BMI, Body fat percentage, Systolic blood pressure, Diastolic blood pressure, Mean arterial pressure, Pulse pressure, Creatine kinase. At the same time, this allele is associated with an increase in levels of HDL cholesterol и AF-SNP hypertension interaction (Table 7).

Table 7

Results of aggregated analyzes of associations between Tag SNP rs2900262 *C9orf16* cardiovascular diseases and their and their intermediate phenotypes (https://cvd.hugeamp.org/)

N	SNP	Phenotype	P-Value	Beta (OR)	Sample Size
1.	rs2900262 *C9orf16* (T/C)	BMI	6.38×10^{-5}	$_{Beta}$▼-0.0122	3 133 750
2.		Systolic blood pressure	5.44×10^{-4}	$_{Beta}$▼-0.0102	1 269 390
3.		Pulse pressure	0.001	$_{Beta}$▼-0.0097	858 587
4.		Creatine kinase	0.008	$_{Beta}$▼-0.0122	111 424
5.		HDL cholesterol	0.02	$_{Beta}$▲0.0052	1 726 890
6.		Diastolic blood pressure	0.02	$_{Beta}$▼-0.0079	1 282 710
7.		Mean arterial pressure	0.03	$_{Beta}$▼-0.0086	136 482
8.		AF-SNP hypertension interaction	0.04	$_{OR}$▲1.2460	20 050
9.		Body fat percentage	0.047	$_{Beta}$▼-0.0800	26 639

Note: Data obtained using the bioinformatic resource Cardiovascular Disease Knowledge Portal (https://cvd.hugeamp.org/); All presented calculations were performed with respect to the C allele.

Thus, this genetic variant may play an important role in the regulation of blood pressure, lipid metabolism, and may also influence the association of SNPs with the development of atrial fibrillation. Taking into account that atrial fibrillation acts as a significant risk factor for ischemic stroke of cardioembolic type, the influence of rs2900262 *C9orf16* on the risk of IS may include a range of molecular mechanisms involved in various links in the pathogenesis of the disease.

Bioinformatic analysis was additionally performed to interpret the functional effects of rs2900262 *C9orf16*. The use of the GTEx Portal resource and the eQTL Calculator search query showed that the C allele (with a protective effect on the risk of IS development established in our study) affects an increase in gene expression in blood vessels. At the same time, the C allele is characterized by multidirectional effects relative to the effect on the level of expression in brain tissues: carriers of the C allele showed an increase in the expression of the *C9orf16* gene in Brain – Caudate (basal ganglia) and Brain – Cerebellum and, conversely, a decrease in gene expression in Brain – Cortex и Brain – Frontal Cortex (BA9) (Table 8).

Table 8

Impact of rs2900262 on *C9orf16* expression by cis-eQTL-effects (GTEx Portal data)

Tissue	P-Value	NES	T-statistic
Brain – Caudate (basal ganglia)	0.024	0.17	2.03
Brain – Cerebellum	0.0062	0.22	2.04
Brain – Cortex	0.018	-0.19	-2.4
Brain – Frontal Cortex (BA9)	0.037	-0.18	-2.1
Artery – Tibial	0.00059	0.13	3.05

Note: effective allele – C

Resource eQTLGen Consortium (https://www.eqtlgen.org/cis-eqtls.html) showed that the risk allele T affects the decrease in expression of gene *URM1* (Z-score = -10.6884, P =1.15×10^{-26}) and *DNM1* (Z-score = -4.67, P = 2.9×10^{-6}), as well as an increase in gene expression *SH3GLB2* (Z-score = 4.44, P = 8.8×10^{-6}) и *RP11-395P17.3* (Z-score = 5.78, P = 7.41×10^{-9}) through cis-eQTL effects in blood cells. Gene *URM1* acts as a ubiquitin-like protein (UBL) that is covalently conjugated via an isopeptide bond to lysine residues of target proteins such as MOCS3, ATPBD3, CTU2, USP15 and CAS. The thiocarboxylated form serves as substrate for conjugation and oxidative stress specifically induces the formation of UBL-protein conjugates (https://www.genecards.org/cgi-bin/carddisp.pl?gene=URM1&keywords=URM1). *DNM1* gene encodes a member of the dynamin subfamily of GTP-binding proteins. The encoded protein possesses unique mechanochemical properties used to tubulate and sever membranes, and is involved in clathrin-mediated endocytosis and other vesicular trafficking processes. Actin and other cytoskeletal proteins act as binding partners for the encoded protein, which can also self-assemble leading to stimulation of GTPase activity (https://www.genecards.org/cgi-bin/carddisp.pl?gene=DNM1&keywords=DNM1). *SH3GLB2* is a Protein Coding gene. Enables identical protein binding activity (https://www.genecards.org/cgi-bin/carddisp.pl?gene=SH3GLB2&keywords=SH3GLB2).

Subsequent analysis revealed a significant effect of rs2900262 *C9orf16* on histone modifications. According to the HaploReg database (v4.1), this genetic variant is located in the region of DNA binding to histone H3, characterized by mono-methylation at the 4th lysine residue of the histone H3 protein (H3K4me1) and marking enhancers in all brain tissues represented by HaploReg (v4.1), peripheral blood cells, as well as tri-methylation at the 4th lysine residue of the histone H3 protein (H3K4me3) and marking promoters in the brain tissues and blood. The effect of this histone mark is enhanced by acetylation of the lysine residues at N-terminal position 27 of the histone H3 protein (H3K27ac), marking enhancers in peripheral blood cells and all brain tissues, as well as the acetylation at the 9th lysine residues of the histone H3 protein (H3K9ac), marking promoters in blood cells and all brain tissues, except Brain Hippocampus Middle (Table 9).

Table 9

Tissue-specific effects of rs2900262 *C9orf16* on histone modifications (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php)

Histone marks / Tissues	H3K4me1	H3K4me3	H3K27ac	H3K9ac
Brain Hippocampus Middle	Enh	Pro	Enh	-
Brain Substantia Nigra	Enh	-	Enh	Pro
Brain Anterior Caudate	Enh	-	Enh	Pro
Brain Cingulate Gyrus	Enh	Pro	Enh	Pro
Brain Inferior Temporal Lobe	Enh	Pro	Enh	Pro
Brain Angular Gyrus	Enh	Pro	Enh	Pro
Brain_Dorsolateral_Prefrontal_Cortex	Enh	Pro	Enh	Pro
Cells from peripheral blood (any)	Enh	Pro	Enh	Pro

Note: H3K4me1 – mono-methylation at the 4th lysine residue of the histone H3 protein; H3K4me3 – tri-methylation at the 4th lysine residue of the histone H3 protein; H3K9ac – the acetylation at the 9th lysine residues of the histone H3 protein; H3K27ac – acetylation of the lysine residues at N-terminal position 27 of the histone H3 protein; Enh – histone modification in the enhancer region; Pro – histone modification at the promoter region.

Table 10

Analysis of the impact of rs2900262 *C9ORF16* on the binding of transcription factors with DNA

Ref/SNP allele[1]	TF[2]	GAIN /LOSS[3]	Motif[4]	P-Value SNP impact[5]	P-Value Ref[6]	P-Value SNP[7]
T/C	NFKB	GAIN	NFKB_disc2	$<1\times10^{-5}$	4.71×10^{-6}	0.18
T/C	IKZF2	GAIN	IKZF2_2	2.09×10^{-3}	1.11×10^{-2}	0.40
T/C	RAD21	GAIN	RAD21_disc10	2.28×10^{-3}	1.99×10^{-2}	0.28
T/C	SMC3	GAIN	SMC3_disc3	2.53×10^{-3}	3.10×10^{-2}	0.35
T/C	NR3C1	GAIN	NR3C1_disc6	3.27×10^{-3}	1.22×10^{-2}	0.14
T/C	HAND1	GAIN	HAND1_2	5.87×10^{-3}	1.99×10^{-2}	0.48
T/C	BCL	GAIN	BCL_disc9	7.63×10^{-3}	4.01×10^{-2}	0.30
T/C	MGA	GAIN	MGA_3	9.46×10^{-3}	3.52×10^{-2}	0.30
T/C	NANOG	GAIN	NANOG_1	1.00×10^{-2}	2.40×10^{-2}	0.71
T/C	ZNF410	GAIN	ZNF410_1	1.45×10^{-2}	2.67×10^{-2}	0.29
T/C	MZF1_5-13	GAIN	MA0057.1	1.50×10^{-2}	4.93×10^{-2}	0.46
T/C	THAP1	GAIN	MA0597.1	1.53×10^{-2}	2.31×10^{-2}	0.35
T/C	RFX1	GAIN	RFX1_4	1.84×10^{-2}	3.83×10^{-2}	0.40
T/C	SOX2	GAIN	SOX2_1	2.02×10^{-2}	9.14×10^{-3}	6.18×10^{-2}
T/C	NR2C2	GAIN	MA0504.1	4.40×10^{-2}	4.67×10^{-2}	0.23
T/C	NR1H	LOSS	NR1H_3	2.37×10^{-5}	0.15	1.09×10^{-4}
T/C	IKZF2	LOSS	IKZF2_3	3.91×10^{-3}	0.38	1.10×10^{-2}
T/C	ETS1	LOSS	ETS1_2	6.90×10^{-3}	4.20	4.23×10^{-2}
T[8]	-					
C[9]	regulation of leukocyte adhesion to vascular endothelial cell (GO:1904994; Padj=0.02); regulation of vasculature development (GO:1901342; Padj=0.02); regulation of blood vessel endothelial cell migration (GO:0043535; Padj=0.02); positive regulation of defense response (GO:0031349; Padj=0.03); regulation of angiogenesis (GO:0045765; Padj=0.045); regulation of inflammatory response (GO:0050727; Padj=0.045)					

Note: 1 – reference (Ref) / alternative (SNP) allele; 2 – TF – transcription factor; 3 – TF binding to the reference (GAIN) / alternative (LOSS) allele; 4 – binding sites with high affinity for TF; 5 – p value, statistically confirming the potential increase or loss of function of the genomic region with SNP in terms of transcription factor binding; 6 – p-value for assessing the binding of TF to the Ref allele; 7 – p-value for assessing the binding of TF to the SNP allele; 8 – biological processes in which TFs that bind to the reference allele are jointly involved (data from the Enrichr resource; https://maayanlab.cloud/Enrichr/); 9 – biological processes in which TFs that bind to an alternative allele are jointly involved (data from the Enrichr resource; https://maayanlab.cloud/Enrichr/)

The analysis of TFs using atSNP source revealed that TFs binding to the protective allele C, are involved in joint regulation of leukocyte adhesion to vascular endothelial cell (GO:1904994; Padj=0.02), regulation of vasculature development (GO:1901342; Padj=0.02), regulation of blood vessel endothelial cell migration (GO:0043535; Padj=0.02), positive regulation of defense response (GO:0031349; Padj=0.03), regulation of angiogenesis (GO:0045765; Padj=0.045), regulation of inflammatory response (GO:0050727; Padj=0.045) (Table 10).

It is noteworthy that the association of rs2900262 *C9orf16* with the risk of IS was found in women and was not found in men. Previously, Smith et al. showed that *C9orf16* expression may be negatively regulated by estradiol [38]. Thereby decreased expression of *C9orf16* in T allele carriers may manifest female-specifically being more impeded in women.

Similarly, in contrast to non-smokers, smoking patients were characterized by more strong association between mutant allele and the risk of ischemic stroke. This finding may have the same ground since some tobacco components apparently may decrease *C9orf16* expression. Using bioinformatic resourse Comparative Toxicogenomics Database we revealed that tobacco smoke and its components are able to down-regulate expression of *C9orf16* as well as up-regulate methylation of *C9orf16* promoter (Table 11).

Table 11

Proven mechanisms of interaction between the gene *C9orf16* and cigarette smoke (data obtained using a bioinformatics resource Comparative Toxicogenomics Database)

N	Interacting chemical	Interactiong	References
1.	Tobacco Smoke Pollution	Tobacco Smoke Pollution results in increased methylation of *BBLN* promoter	[39]
2.	Tobacco tar	Tobacco tar results in decreased expression of *BBLN* mRNA	[40]
3.	Benzo(a)pyrene	Benzo(a)pyrene results in decreased expression of *BBLN* mRNA	[41]
4.	Benzo(a)pyrene	Benzo(a)pyrene results in increased methylation of *BBLN* promoter	[42]

Considering that an increase in methylation leads to a decrease in gene expression, studies conducted in this direction provide additional evidence of the effect of smoking on a decrease in the expression of the *C9orf16* gene and an increase in "risk" effects regarding the development of IS in carriers of variant genotypes C/C-C/T.

Conclusion. To the best of our knowledge, this is the first study that showed the association of rs2900262 *C9orf16* with the risk of IS, sex-specific effect in this association, and the trigger effect of smoking on the association of this genetic variant with the development of the disease. Bioinformatic analysis revealed a high regulatory potential of rs2900262 *C9orf16*, particular cis-eQTL-mediated effects on the regulation of gene expression in blood, blood vessels and brain, binding to transcription factors involved in biological processes with significant role in the pathogenesis of IS, as well as epigenetic regulation of this SNP. Future studies have yet to accumulate the data disclosing detailed networks of its probable implication in ischemic stroke risk and course.

Financial support

This research was funded by Russian Science Foundation (No. 22-15-00288, https://rscf.ru/en/project/22-15-00288/).

Conflict of interests

The authors have no conflict of interest to declare.

References

1. Young MJ, Regenhardt RW, Leslie-Mazwi TM, et al. Disabling stroke in persons al-

ready with a disability: Ethical dimensions and directives. Neurology. 2020;94(7):306-310. DOI: https://doi.org/10.1212/WNL.0000000000008964

2. Johansen, MC, Langton-Frost N, Gottesman RF. The role of cardiovascular disease in cognitive impairment. Current Geriatrics Reports. 2020;9(1):1-9. DOI: https://doi.org/10.1007/s13670-020-00309-7

3. Murphy SJX, Werring DJ. Stroke: causes and clinical features. Medicine. 2020;48(9):561-566. DOI: https://doi.org/10.1016/j.mpmed.2020.06.002

4. Abraham G, Malik R, Yonova-Doing E, et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nature Communications. 2019;10:5819. DOI: https://doi.org/10.1038/s41467-019-13848-1

5. Wang C, Du Z, Ye N, et al. Hyperlipidemia and hypertension have synergistic interaction on ischemic stroke: insights from a general population survey in China. BMC Cardiovascular Disorders. 2022;22:47. DOI: https://doi.org/10.1186/s12872-022-02491-2

6. Appunni S, Rubens M, Ramamoorthy V, et al. Stroke Genomics: Current Knowledge, Clinical Applications and Future Possibilities. Brain Sciences. 2022;12(3):302. DOI: https://doi.org/10.3390/brainsci12030302

7. Polonikov A, Bocharova I, Azarova I, et al. The Impact of Genetic Polymorphisms in Glutamate-Cysteine Ligase, a Key Enzyme of Glutathione Biosynthesis, on Ischemic Stroke Risk and Brain Infarct Size. Life. 2022;12(4):602. DOI: https://doi.org/10.3390/life12040602

8. Bushueva O, Barysheva E, Markov A, et al. DNA Hypomethylation of the MPO Gene in Peripheral Blood Leukocytes Is Associated with Cerebral Stroke in the Acute Phase. Journal of Molecular Neuroscience. 2021;71:1914-1932. DOI: https://doi.org/10.1007/s12031-021-01840-8

9. Cipolla MJ, Liebeskind DS, Chan SL. The importance of comorbidities in ischemic stroke: Impact of hypertension on the cerebral circulation. Journal of Cerebral Blood Flow and Metabolism. 2018;38(12):2129-2149. DOI: https://doi.org/10.1177/0271678X18800589

10. Bushueva O, Solodilova M, Ivanov V, et al. Gender-specific protective effect of the -463G>A polymorphism of myeloperoxidase gene against the risk of essential hypertension in Russians. Journal of the American Society of Hypertension. 2015;9(11):902-906. DOI: https://doi.org/10.1016/j.jash.2015.08.006

11. Moskalenko MI. The involvement of genes of matrix metalloproteinases in the development of arteial hypertension and its complication (review). Research Result. Medicine and Pharmacy. 2018;4(1):53-69. Russian. DOI: https://doi.org/10.18413/2313-8955-2018-4-1-53-69

12. Song X, Zhao X, Liebeskind DS, et al. Incremental value of plaque enhancement in predicting stroke recurrence in symptomatic intracranial atherosclerosis. Neuroradiology. 2020;62:1123-1131. DOI: https://doi.org/10.1007/s00234-020-02418-8

13. Sorokin A, Kotani K, Bushueva O, et al. The cardio-ankle vascular index and ankle-brachial index in young russians. Journal of Atherosclerosis and Thrombosis. 2015;22(2):211-218. DOI: https://doi.org/10.5551/jat.26104

14. Nimjee SM, Akhter AS, Zakeri A, et al. Sex differences in thrombosis as it affects acute ischemic stroke. Neurobiology of Disease. 2022;165:105647. DOI: https://doi.org/10.1016/j.nbd.2022.105647

15. Dong S, Maniar S, Manole MD, et al. Cerebral Hypoperfusion and Other Shared Brain Pathologies in Ischemic Stroke and Alzheimer's Disease. Translational Stroke Research. 2018;9:238-250. DOI: https://doi.org/10.1007/s12975-017-0570-2

16. Datta A, Sarmah D, Mounica L, et al. Cell Death Pathways in Ischemic Stroke and Targeted Pharmacotherapy. Translational Stroke Research. 2020;11:1185-1202. DOI: https://doi.org/10.1007/s12975-020-00806-z

17. Demyanenko S, Nikul V, Rodkin S, et al. Exogenous recombinant Hsp70 mediates neuroprotection after photothrombotic stroke. Cell Stress and Chaperones. 2021;26:103-114. DOI: https://doi.org/10.1007/s12192-020-01159-0

18. Tsuboyama K, Osaki T, Matsuura-Suzuki E, et al. A widespread family of heat-resistant obscure (Hero) proteins protect against protein instability and aggregation. PLoS Biology. 2020;18(3):e3000632. DOI: https://doi.org/10.1371/journal.pbio.3000632

19. Polonikov A, Rymarova L, Klyosova E, et al. Matrix metalloproteinases as target genes for gene regulatory networks driving molecular and cellular pathways related to a multistep pathogenesis of cerebrovascular disease. Journal of Cellular Biochemistry. 2019;120(10):16467-16482. DOI: https://doi.org/10.1002/jcb.28815

20. Bushueva OY, Stetskaya TA, Polonikov AV, et al. The relationship between polymorphism 640A>G of the CYBA gene with the risk of

ischemic stroke in the population of the Central Russia. Zhurnal Nevrologii i Psikhiatrii imeni S.S. Korsakova. 2015;115(9 Pt 2):38-41. Russian. DOI: https://doi.org/10.17116/jnevro20151159238-41

21. Polonikov A, Vialykh E, Vasil'eva O, et al. Genetic variation in glutathione S-transferase genes and risk of nonfatal cerebral stroke in patients suffering from essential hypertension. Journal of Molecular Neuroscience. 2012;47(3):511-3. DOI: https://doi.org/10.1007/s12031-012-9764-y

22. Vialykh EK, Solidolova MA, Bushueva OIu, et al. Catalase gene polymorphism is associated with increased risk of cerebral stroke in hypertensive patients. Zhurnal Nevrologii i Psikhiatrii imeni S.S. Korsakova. 2012;112(8 Pt 2):3-7. Russian.

23. Bushueva OY, Bulgakova IV, Ivanov VP, et al. Association of Flavin Monooxygenase Gene E158K Polymorphism with Chronic Heart Disease Risk. Bulletin of Experimental Biology and Medicine. 2015;159(6):776-778. DOI: https://doi.org/10.1007/s10517-015-3073-8

24. Bushueva OYu. Single nucleotide polymorphisms in genes encoding xenobiotic metabolizing enzymes are associated with predisposition to arterial hypertension. Research Results in Biomedicine. 2020;6(4):447-456. DOI: https://doi.org/10.18413/2658-6533-2020-6-4-0-1

25. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007;23(10):1289-1291. DOI: https://doi.org/10.1093/bioinformatics/btm091

26. Solé X, Guinó E, Valls J, et al. SNPStats: a web tool for the analysis of association studies. Bioinformatics. 2006;22(15):1928-1929. DOI: https://doi.org/10.1093/bioinformatics/btl268

27. Guo L, Wang J. rSNPBase 3.0: an updated database of SNP-related regulatory elements, element-gene pairs and SNP-based gene regulatory networks. Nucleic Acids Research. 2018;46(D1):D1111-D1116. DOI: https://doi.org/10.1093/nar/gkx1101

28. Dong S, Boyle AP. Predicting functional variants in enhancer and promoter elements using RegulomeDB. Human Mutation. 2019;40(9):1292-1298. DOI: https://doi.org/10.1002/humu.23791

29. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-1330. DOI: https://doi.org/10.1126/science.aaz1776

30. Võsa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nature Genetics. 2021;53:1300-1310. DOI: https://doi.org/10.1038/s41588-021-00913-z

31. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Research. 2016;44(D1):D877-81. DOI: https://doi.org/10.1093/nar/gkv1340

32. Shin S, Hudson R, Harrison C, et al. atSNP Search: a web resource for statistically evaluating influence of human genetic variation on transcription factor binding. Bioinformatics. 2019;35(15):2657-2659. DOI: https://doi.org/10.1093/bioinformatics/bty1010

33. Xie Z, Bailey A, Kuleshov MV, et al. Gene Set Knowledge Discovery with Enrichr. Current Protocols. 2021;1(3):e90. DOI: https://doi.org/10.1002/cpz1.90

34. Jafari M, Ansari-Pour N. Why, When and How to Adjust Your P Values? Cell Journal. 2019;20(4):604-607. DOI: https://doi.org/10.22074/cellj.2019.5992

35. Davis AP, Grondin CJ, Johnson RJ, et al. Comparative Toxicogenomics Database (CTD): update 2021. Nucleic Acids Research. 2021;49(D1):D1138-D1143. DOI: https://doi.org/10.1093/nar/gkaa891

36. Crawford KM, Gallego-Fabrega C, Kourkoulis C, et al. Cerebrovascular Disease Knowledge Portal: An Open-Access Data Resource to Accelerate Genomic Discoveries in Stroke. Stroke. 2018;49(2):470-475. DOI: https://doi.org/10.1161/STROKEAHA.117.018922Stroke

37. Pantaleo E, Monaco A, Amoroso N, et al. A Machine Learning Approach to Parkinson's Disease Blood Transcriptomics. Genes. 2022;13(5):727. DOI: https://doi.org/10.3390/genes13050727

38. Smith LC, Moreno S, Robertson L, et al. Transforming growth factor beta1 targets estrogen receptor signaling in bronchial epithelial cells. Respiratory Research. 2018;19:160. DOI: https://doi.org/10.1186/s12931-018-0861-5

39. Howe CG, Zhou M, Wang X, et al. Associations between Maternal Tobacco Smoke Exposure and the Cord Blood CD4+ DNA Methylome. Environmental Health Perspectives. 2019;127(4):47009. DOI: https://doi.org/10.1289/EHP3398

40. Hu YC, Yang ZH, Zhong KJ, et al. Alteration of transcriptional profile in human bronchial epithelial cells induced by cigarette smoke condensate. Toxicology Letters. 2009;190(1):23-31. DOI: https://doi.org/10.1016/j.toxlet.2009.06.860

41. Jennen DGJ, Magkoufopoulou C, Ketelslegers HB, et al. Comparison of HepG2 and HepaRG by whole-genome gene expression analysis for the purpose of chemical hazard identification. Toxicological Sciences. 2010;115(1):66-79. DOI: https://doi.org/10.1093/toxsci/kfq026

42. Jiang CL, He SW, Zhang YD, et al. Air pollution and DNA methylation alterations in lung cancer: A systematic and comparative study. Oncotarget. 2017;8(1):1369-1391. DOI: https://doi.org/10.18632/oncotarget.13622

Received 11 April 2022
Revised 19 May 2022
Accepted 25 May 2022

Information about the authors

Ksenia A. Kobzeva, Student, Kursk State Medical University, Kursk, Russia, E-mail: kseniya.kobzeva.0246@gmail.com, ORCID: https://orcid.org/0000-0002-0652-2219.

Irina V. Shilenok, Neurologist, Kursk City Clinical Emergency Hospital, Kursk, Russia, E-mail: irinka.zharikova.96@mail.ru, ORCID: https://orcid.org/0000-0001-7894-8385.

Andrey E. Belykh, Cand. Sci. (Medicine), Associate Professor at the Department of Pathophysiology, Kursk State Medical University, Kursk, Russia, Postdoctoral Researcher at the Dioscuri Centre for Metabolic Diseases, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, Warsaw, Poland, E-mail: and-white@yandex.ru, ORCID: https://orcid.org/0000-0001-9766-2104.

Denis E. Gurtovoy, Student, Kursk State Medical University, Kursk, Russia, E-mail: dne.ar@mail.ru, ORCID: https://orcid.org/0000-0001-8632-4708.

Lyubov A. Bobyleva, Student, Kursk State Medical University, Kursk, Russia, E-mail: labobuleva@mail.ru, ORCID: https://orcid.org/0000-0002-2014-5827.

Anna B. Krapiva, Student, Kursk State Medical University, Kursk, Russia, E-mail: anny1999.10@mail.ru, ORCID: https://orcid.org/0000-0001-5163-575X.

Tatiana A. Stetskaya, Cand. Sci. (Medicine), Junior Researcher at the Laboratory of Statistical Genetics and Bioinformatics, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia, E-mail: steckaya_tatiana@mail.ru, ORCID: https://orcid.org/0000-0002-4729-4654.

Marina A. Bykanova, Researcher at the Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia, E-mail: marina.bickanova@yandex.ru, ORCID: https://orcid.org/0000-0001-5420-3557.

Anastasiya A. Mezhenskaya, General Practitioner, Alekseevskaya Central District Hospital, Alekseevska, Russia, E-mail: mezhenskaya_a.a@mail.ru, ORCID: https://orcid.org/0000-0002-5811-2537.

Ekaterina A. Lysikova, Research Scientist at the Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia, E-mail: lysikova_katerina@mail.ru, ORCID: https://orcid.org/0000-0001-6521-3793.

Maxim B. Freidin, Doct. Sci. (Biology), Lecturer at the Department of Biology, School of Biological and Behavioural Sciences, Queen Mary University of London, London, United Kingdom, Senior Researcher at the Laboratory of Population Genetics, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia, E-mail: m.freydin@qmul.ac.uk, ORCID: https://orcid.org/0000-0002-1439-6259.

Olga Y. Bushueva, Cand. Sci. (Medicine), Associate Professor at the Department of Biology, Medical Genetics and Ecology, Head of the Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia, E-mail: olga.bushueva@inbox.ru, ORCID: https://orcid.org/0000-0003-3333-0623.