


Matrix metalloproteinases as target genes for gene regulatory networks driving molecular and cellular pathways related to a multistep pathogenesis of cerebrovascular disease

Alexey Polonikov^{1,2}  | Larisa Rymarova² | Elena Klysova³ | Anastasia Volkova¹ | Iuliia Azarova^{3,4} | Olga Bushueva^{1,5} | Marina Bykanova^{1,5} | Iuliia Bocharova¹ | Sergey Zhabin⁶ | Mikhail Churnosov⁷ | Vitaliy Laskov⁸ | Maria Solodilova¹

¹Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russian Federation

²Laboratory of Statistical Genetics and Bioinformatics, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russian Federation

³Laboratory of Biochemical Genetics and Metabolomics, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russian Federation

⁴Department of Biological Chemistry, Kursk State Medical University, Kursk, Russian Federation

⁵Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russian Federation

⁶Department of Surgical Diseases, Kursk State Medical University, Kursk, Russian Federation

⁷Department of Medical Biological Disciplines, Belgorod State University, Belgorod, Russian Federation

⁸Department of Neurology and Neurosurgery, Kursk State Medical University, Kursk, Russian Federation

Correspondence

Alexey Polonikov, MD, PhD, Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, 3 Karl Marx St., Kursk 305041, Russian Federation.

Email: polonikov@rambler.ru.

Funding information

Russian Science Foundation, Grant/Award Number: 15-15-10010

Abstract

The present study investigated a joint contribution of matrix metalloproteinases (MMPs) genes to ischemic stroke (IS) development and analyzed interactions between MMP genes and genome-wide associated loci for IS. A total of 1288 unrelated Russians (600 IS patients and 688 healthy individuals) from Central Russia were recruited for the study. Genotyping of seven single nucleotide polymorphisms (SNPs) of MMP genes (rs1799750, rs243865, rs3025058, rs11225395, rs17576, rs486055, and rs2276109) and eight genome-wide associated loci for IS were done using Taq-Man-based assays and MALDI-TOF mass spectrometry iPLEX platform, respectively. Allele – 799T at rs11225395 of the *MMP8* gene was significantly associated with a decreased risk of IS after adjustment for sex and age (OR = 0.82; 95%CI, 0.70-0.96; $P = 0.016$). The model-based multifactor dimensionality reduction method has revealed 21 two-order, 124 three-order, and 474 four-order gene-gene (G×G) interactions models meaningfully ($P_{\text{perm}} < 0.05$) associated with the IS risk. The bioinformatic analysis enabled establishing the studied MMP gene polymorphisms possess a clear regulatory potential and may be targeted by gene regulatory networks driving molecular and cellular pathways related to the pathogenesis of IS. In conclusion, the present study was the first to identify an association between polymorphism rs11225395 of the *MMP8* gene and IS risk. The study

findings also indicate that MMPs deserve special attention as a potential class of genes influencing the multistep mechanisms of cerebrovascular disease including atherosclerosis in cerebral arteries, acute cerebral artery occlusion as well as the ischemic injury of the brain and its recovery.

KEYWORDS

gene regulatory networks., gene-gene interactions, ischemic stroke, matrix metalloproteinases, pathogenesis, single nucleotide polymorphisms

1 | INTRODUCTION

Stroke is the third most common cause of death and chronic disability in developed countries.¹ An important point is that the overall burden of stroke in terms of an absolute number of people affected or remained disabled; the disease has dramatically increased across the globe in both men and women.¹ Ischemic stroke (IS) (also known as a cerebral infarction) is the most common type of cerebrovascular disease that occurs as a result of an obstruction of the artery supplying blood to the brain. Atherothrombotic stroke, a predominant type of IS, occurs when a blood clot forms on an atherosclerotic plaque within a cerebral artery.

Numerous studies demonstrated that genetic factors play an essential role in the development of IS. In the vast majority of cases, IS is a polygenic multifactorial disorder resulting from complex interactions between genetic and environmental factors which are jointly involved in the disease development.^{2,3} Candidate gene and genome-wide association studies (GWAS) have identified a broad spectrum of polymorphic genes whose effects on the susceptibility to IS are mediated through abnormalities in homocysteine metabolism, renin-angiotensin-aldosterone system, hemostatic system, leukotriene metabolism and other pathways.² In recent years, there has been increased interest in the study of matrix metalloproteinases (MMPs) and their relation to pathogenesis of cerebrovascular disease and brain recovery after IS.⁴⁻⁶

MMPs represent a group of proteolytic zinc-dependent enzymes synthesized by endothelial cells, neurons, glial cells, and astrocytes and responsible for the collagen and other protein degradation in extracellular matrix and implicated in various pathological conditions such as inflammation, tissue destruction, atherosclerosis, and many other processes playing a role in diseases of the central nervous system.^{4,7,8} An increased activity of MMPs has been shown to be linked with the susceptibility and severity of IS whereas a blockage in the activity of MMPs was found to reduce disease risk or brain injury after stroke.⁹ Thus, polymorphic genes encoding MMPs become attractive candidates for serving as molecular

markers for susceptibility and clinical course of IS. Almost three dozens of studies have been conducted to assess associations between polymorphisms of genes encoding individual MMPs and the risk of IS.^{6,10} However, these studies have been designed on populations with a small sample size that decreased the power to detect genotype-phenotype associations and, therefore, yielded inconsistent and/or inconclusive results. A recent systematic review and meta-analysis of the 29 studies investigated association between single nucleotide polymorphisms (SNP) in MMP family genes and risk of IS has shown that SNPs $-1562C/T$ of *MMP9* and $-1082A/G$ of *MMP12* could be considered as risk factors for IS, whereas polymorphisms $-16071G/2G$ of *MMP1*, $-1306C/T$ and $-735C/T$ of *MMP2* and also $-16125A/6A$ of *MMP3* are not associated with disease risk.¹⁰ No studies so far have been done to evaluate a joint contribution of various MMPs to the development of IS on populations with greater sample size than investigated before. Therefore, the present study was designed (1) to investigate a joint contribution of MMPs genes to IS development; (2) to analyze interactions between genes encoding MMPs and well-known candidate genes for IS susceptibility, and (3) to perform a comprehensive bioinformatic analysis with a purpose to identify both the regulatory potential of SNPs related to disease risk and functional partners and regulators of MMPs involved in different steps of pathogenesis of IS.

2 | MATERIALS AND METHODS

2.1 | Study participants

Ethical Review Committee of Kursk State Medical University has approved study protocol. All participants gave written informed consent before enrollment for this study. A total of 1288 unrelated Russians (600 patients with IS and 688 healthy individuals) from Central Russia were recruited for the study. The patients were enrolled from the Regional Cardiovascular Centre of Kursk Regional Clinical Hospital and Neurology Division of Kursk Emergency Hospital during two periods, between

2007 and 2013 as described previously,^{11–13} and between 2015 and 2017. All patients were examined by qualified neurologists. The diagnosis of cerebral infarction was based on the results of the neurological examination and computed tomography and/or magnetic resonance imaging of the brain. Stroke patients with a history of cardiac arrhythmias were not included in the study. To be eligible, the subjects for the control group were required to be in good health. This group was recruited from the same population and for the same time periods and included healthy blood volunteers and hospital-based patients with no clinical evidence for cerebrovascular, cardiovascular, or other chronic diseases. Baseline and clinical characteristics of the study population are listed in Table 1. The control group was matched to the IS group for sex, age and body mass index ($P > 0.05$).

2.2 | DNA analysis

Whole blood samples (5 mL) were collected from all study participants into EDTA-coated tubes and maintained at -20°C until processed. Genomic DNA was isolated using the standard procedure of phenol/chloroform extraction. We have selected seven common frequently investigated SNPs of MMP genes, namely $-1607\text{ G}/\text{A}$ (rs1799750) *MMP1*, $-1306\text{ C}>\text{T}$ (rs243865) *MMP2*, $-1612\text{ A}/\text{G}$ (rs3025058) *MMP3*, $-799\text{ C}>\text{T}$ (rs11225395) *MMP8*, $\text{c.836 A}>\text{G}$ (rs17576) *MMP9*, $\text{c.158 C}>\text{T}$ (rs486055) *MMP10*, and $-82\text{ A}>\text{G}$ (rs2276109) *MMP12*. Genotyping of the SNPs was done using Taq-Man-based real-time polymerase chain reaction assays according to the protocols published in the literature. Primer and probe sequences are shown in Table S1. The polymerase chain reaction was performed on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA). In addition, we have selected eight SNPs which have been found to be strongly associated with IS risk in European GWAS. These GWAS loci included rs2417957 of *SLCO1B1*,¹⁴ rs6511720

of *LDLR*,¹⁵ rs4322086 of *RASEF*,¹⁶ rs12449964 of *PMT1*,¹⁷ rs12646447 of *PITX2*,¹⁸ rs899997 of *LOC105370913*,¹⁷ rs11556924 of *ZC3HC1*,¹⁷ and rs783396 of *AIM1*.¹⁹ Genotyping of these genes was performed using a MALDI-TOF mass spectrometry iPLEX platform (Agena Bioscience Inc, San Diego, CA). Primer sequences used for genotyping are shown in Table S2. Blind replicates were included to ensure quality control of genotyping. Laboratory personnel conducting genotyping were blinded to patients' information. The repeatability test was conducted for 5% of randomly selected samples, yielded 100% reproducibility. All molecular genetic investigations were performed at the Research Institute of Genetic and Molecular Epidemiology of Kursk State Medical University (Kursk, Russia).

2.3 | Statistical and bioinformatic analysis

Allele frequencies were estimated by the gene counting method. The χ^2 -test was used to identify significant departures from Hardy-Weinberg equilibrium. P value ≤ 0.05 was considered statistically significant. Allele and genotype frequencies in the study groups were analyzed using the SNPStats software (<http://bioinfo.iconcologia.net/snpstats/start.htm>).²⁰ Associations between MMPs genotypes and the risk of IS were evaluated by multiple logistic regression assuming codominant, dominant, recessive, overdominant and log-additive genetic models and adjusted for covariates such as age and gender. AIC value (Akaike's Information Criterion) was used to select the best genetic model for SNP-disease association.

A joint contribution of genes to the risk of IS was evaluated by method multifactor dimensionality reduction (MDR), a data-mining and model-free bioinformatic approach to identifying high-order gene-gene and gene-environment interactions determining susceptibility to complex multifactorial disorders.^{21,22} In the present study,

TABLE 1 Baseline and clinical characteristics of the study population

Baseline and clinical characteristics	Controls (N = 688)	IS patients (N = 600)	P value
Age, M \pm SD	60.8 \pm 7.5	61.1 \pm 9.8	0.59
Sex, N (%)			
Males	366 (53.2)	330 (55.0)	0.52
Females	322 (46.8)	270 (45.0)	
BMI (kg/m ²), M \pm SD	24.6 \pm 3.8	25.2 \pm 4.2	0.11
Hypertension	0 (0.0)	586 (97.7)	-
Coronary artery disease	0 (0.0)	49 (8.2)	-
Diabetes mellitus	0 (0.0)	52 (8.7)	-

Abbreviations: BMI, body mass index; M, mean; N, number; SD, standard deviation.

we utilized the model-based multifactor dimensionality reduction method (MB-MDR), an extension of the MDR method, which measures the association between multi-locus genotypes and the phenotype, providing a set of statistically significant interactions instead of a single best model.²³ The *mbmdr* package for R was used to perform modeling gene-gene interactions underlying stroke susceptibility.²⁴ A detailed algorithm of the usage of MB-MDR method was described previously.²⁵ We evaluated two-, three-, and four-order gene-gene interactions associated with IS risk, and statistical significance (P_{perm}) for each n -order model was evaluated through 1000 permutation tests.

The regulatory potential of SNPs of MMP genes was evaluated in silico using two bioinformatics tools: (1) the SNP Function Prediction (FuncPred) utilizing the TRANSFAC database of potential transcription factor (TF) recognition sites and available online at <https://snpinfo.nih.gov/snpinfo/snpfunc.html>,²⁶ (2) the rSNPBase database of curated regulatory SNPs available at <http://rsnp.psych.ac.cn>.²⁷ Those transcription factor binding sites falling into the investigated SNPs and whose core or matrix match scores were impacted, or which were eliminated or created by variant sequences are considered to be regulatory within a particular SNP. Thus, the FuncPred tool enabled identification transcription factors binding sites (TFBS) and other regulatory proteins at SNP regions, thereby gaining important information on the gene regulatory networks having the potential to influence expression and/or activity of the MMP genes.

To better understand the mechanisms by which MMPs are involved in the polygenic basis of IS, we performed an in silico search for biological processes, molecular functions, and metabolic pathways up or downregulating the functions of MMPs through the binding with TFBS at their SNPs and potentially related with disease pathogenesis. To accomplish this purpose, enrichment analysis on a gene set was performed using online tool hosted by Gene Ontology (GO, www.geneontology.org). The enrichment analysis enabled finding GO terms which are over-represented (or under-represented) using annotations for a gene set including TFBS identified in the studied MMPs. To overcome the multiple hypothesis testing issue, significance levels for GO terms sharing with a gene set were evaluated by the false discovery rate (FDR) and calculation of Q -values, which is the expected proportion of true null hypotheses rejected out of the total number of null hypotheses rejected²⁸

Bioinformatic tool of the STRING Database v10.5 database (<https://string-db.org/>)²⁹ was used for visualization of complex protein-protein interactions between the disease-associated MMPs, their functional partners, gene

regulatory networks, and metabolic pathways which could be involved in the molecular mechanisms of IS.

3 | RESULTS

3.1 | Allele and genotype frequencies of MMPs SNPs: Interpopulation differences and relationship with IS risk

The genotype and allele frequencies of matrix metallo-proteinases SNPs are shown in Tables 1 and 2. The observed genotype frequency for polymorphisms rs1799750 of *MMP1*, rs243865 of *MMP2*, rs3025058 of *MMP3*, rs486055 of *MMP10* and rs2276109 of *MMP12* were in Hardy-Weinberg equilibrium ($P > 0.05$). A deviation from Hardy-Weinberg equilibrium was found for SNPs rs11225395 of *MMP8* and rs17576 of *MMP9* due to decreased values of observed heterozygosity (0.385 and 0.254, respectively), as compared with expected heterozygosity values (0.480 and 0.295, respectively). Minor allele frequencies (MAF) for SNPs rs1799750 of *MMP1*, rs243865 of *MMP2*, rs3025058 of *MMP3* and rs11225395 of *MMP8* were in accordance with those reported in European populations (www.ensembl.org), according to the data published in 1000 Genomes Project, Phase 3. Meantime, MAF for SNPs rs17576 of *MMP9*, rs486055 of *MMP10* and rs2276109 of *MMP12* differed significantly ($P < 0.05$) from those reported in Europeans. For instance, MAF for SNP rs17576 of the *MMP9* gene was 0.18 in our Russian population and 0.38 in Europeans. Notably, a relatively low MAF for this SNP also occurs in Peruvians from Lima (0.14), Mexicans from California (0.27), Colombians from Medellin (0.26), Puerto Ricans (0.25), and Gambians from the Western Division (0.27). MAF for SNPs rs486055 of *MMP10* and rs2276109 of *MMP12* in Russians were compatible with those reported in European populations (0.17 vs 0.14 and 0.15 vs 0.12, respectively).

The association analysis between polymorphic genes encoding MMPs and the risk of IS is shown in Table 2. Allele -799T at rs11225395 of the *MMP8* gene was significantly associated with decreased risk of IS after adjustment for sex and age (OR = 0.82; 95%CI, 0.70-0.96; $P = 0.016$). As can be seen from Table 2, no statistically significant associations between the MMPs genotypes and IS risk were found ($P > 0.05$). Then we performed a replication study of eight genome-wide associated loci for IS established in European populations (Table 3). We found that allele G (OR = 1.29; 95%CI, 1.10-1.51; $P = 0.002$) and genotype A/G (OR = 1.44; 95%CI, 1.03-2.03; $P < 0.0001$) for SNP rs4322086 of the *RASEF* gene were significantly associated with increased risk of IS. Furthermore, the C/T genotype of *ZC3HC1* (SNP

TABLE 2 Association analysis between polymorphic genes for matrix metalloproteinases and the risk of ischemic stroke

Genes (SNP ID)	Genotype, allele	N (%)		P value	corOR (95% CI)
		Controls (N = 688)	IS patients (N = 600)		
<i>MMP1</i> -1607 2G/1G (rs1799750)	1G/1G	172 (25.3)	159 (26.5)	0.79	1.00
	1G/2G	345 (50.7)	305 (50.8)		0.96 (0.73-1.25)
	2G/2G	164 (24.1)	136 (22.7)		0.90 (0.66-1.23)
	1G	0.506	0.519		1.05 (0.90-1.29)
<i>MMP2</i> -1306C > T (rs243865)	C/C	401 (59.9)	366 (61.6)	0.43	1.00
	C/T	224 (33.4)	198 (33.3)		0.97 (0.76-1.22)
	T/T	45 (6.7)	30 (5.0)		0.73 (0.45-1.18)
	T	0.234	0.247		0.30
<i>MMP3</i> -1612 5A/6A (rs3025058)	6A/6A	186 (27.5)	172 (28.7)	0.61	1.00
	5A/6A	349 (51.5)	293 (48.8)		0.91 (0.70-1.17)
	5A/5A	142 (21.0)	135 (22.5)		1.02 (0.75-1.40)
	5A	0.469	0.468		0.93
<i>MMP8</i> -799C > T (rs11225395)	C/C	251 (38)	261 (43.9)	0.076	1.00
	C/T	261 (39.5)	222 (37.3)		0.82 (0.64-1.05)
	T/T	148 (22.4)	112 (18.8)		0.72 (0.54-0.98)
	T	0.422	0.375		0.016
<i>MMP9</i> c.836A > G (rs17576)	A/A	477 (70.3)	404 (68.1)	0.055	1.00
	A/G	158 (23.3)	165 (27.8)		1.23 (0.95-1.59)
	G/G	43 (6.3)	24 (4.0)		0.66 (0.40-1.11)
	G	0.180	0.180		0.98
<i>MMP10</i> c.158C > T (rs486055)	C/C	456 (67.7)	415 (69.2)	0.24	1.00
	C/T	191 (28.3)	171 (28.5)		0.98 (0.77-1.26)
	T/T	27 (4.0)	14 (2.3)		0.57 (0.30-1.11)
	T	0.182	0.166		0.29
<i>MMP12</i> -82A > G (rs2276109)	A/A	507 (74.9)	419 (69.8)	0.11	1.00
	A/G	155 (22.9)	168 (28.0)		1.31 (1.02-1.69)
	G/G	15 (2.2)	13 (2.2)		1.04 (0.49-2.21)
	G	0.137	0.162		0.076

rs11556924) was significantly associated with increased risk of IS (OR = 1.38; 95%CI, 1.08-1.77; $P = 0.037$). No significant associations were found between other genome-wide associated loci and disease risk ($P > 0.05$). Thus, the present study was able to validate two of eight investigated SNPs such as rs4322086 of *RASEF* and of rs11556924 of *ZC3HC1* as IS susceptibility loci in the Russian population.

3.2 | High-order interactions between the studied genes and the risk of IS

The MB-MDR method was applied to the data set aiming to identify whether epistatic interactions between polymorphic genes for MMPs and genome-wide associated loci for IS jointly contribute to the disease risk. Totally, 21 two-order, 124 three-order, and 474 four-order gene-gene (G×G) interactions models showed significant associations ($P_{\text{perm}} < 0.05$) with the risk of IS (a detailed description of all gene-gene models are presented in Tables S3-S5). Notably, 14 of 21 two-order models (67%), 104 of 124 three-order models (84%), and 445 of 474 four-order (94%) comprised MMP genes interacting to each other or to the IS loci identified by GWAS. Table 4 summarizes the best n -order G×G interactions models underlying susceptibility

to IS (only the top 4 G×G models per an order are shown in Table 4). The best two-order G×G models comprised interactions between gene polymorphisms such as rs3025058 of *MMP3*, rs11225395 of *MMP8*, rs17576 of *MMP9*, rs4322086 of *RASEF*, and rs11556924 of *ZC3HC1* ($P_{\text{perm}} < 0.001$). SNPs rs11225395 of *MMP8*, rs1799750 of *MMP1*, rs243865 of *MMP2* or rs3025058 of *MMP3* were presented in each of the four three-order G×G models in combinations with the GWAS loci such as rs4322086 of *RASEF*, rs11556924 of *ZC3HC1*, rs899997 of *SLCO1B1*, and rs12449964 of *PEMT* ($P_{\text{perm}} < 0.002$). Furthermore, each of the best three four-order G×G interactions shown in Table 4, included at least two genes for MMPs such as rs3025058 of *MMP3*, rs11225395 of *MMP8*, rs1799750 of *MMP1*, or rs17576 of *MMP9* in combinations with genome-wide associated loci for IS such as rs4322086 of *RASEF*, rs11556924 of *ZC3HC1*, rs899997 of *SLCO1B1*, or rs12449964 of *PEMT* ($P_{\text{perm}} < 0.005$).

3.3 | Bioinformatic analysis for the regulatory potential of MMP SNPs

Table 5 summarizes the results of an in silico analysis for the regulatory potential of SNPs of MMP genes using two bioinformatic tools such as FuncPred (<https://snpinfo>).

TABLE 3 SNPs associated with ischemic stroke in genome-wide association studies in European populations: a replication study in population from the Central Russia

Genes (SNP ID)	Genotype, allele	N (%)		P value	corOR (95% CI)
		Controls (N = 688)	IS patients (N = 600)		
<i>ZC3HC1</i> (rs11556924)	C/C	280 (44.0)	195 (37.1)	0.037	1.00
	C/T	281 (44.2)	270 (51.4)		1.38 (1.08-1.77)
	T/T	75 (11.8)	60 (11.4)	0.10	1.16 (0.79-1.70)
	T	0.339	0.371		1.15 (0.97-1.37)
<i>PEMT</i> (rs12449964)	C/C	251 (37.2)	221 (38.9)	0.75	1.00
	C/T	336 (49.8)	279 (49.1)		0.95 (0.74-1.20)
	T/T	88 (13.0)	68 (12.0)	0.47	0.87 (0.61-1.26)
	T	0.379	0.365		0.94 (0.80-1.11)
<i>PITX2</i> (rs12646447)	T/T	424 (78.7)	332 (79)	0.50	1.00
	T/C	100 (18.6)	71 (16.9)		0.91 (0.65-1.28)
	C/C	15 (2.8)	17 (4.0)	0.77	1.43 (0.70-2.91)
	C	0.121	0.125		1.04 (0.79-1.37)
<i>SLCO1B1</i> (rs2417957)	C/C	564 (98.8)	556 (98.1)	0.37	1.00
	C/T	7 (1.2)	10 (1.8)		1.54 (0.58-4.09)
	T/T	0 (0.0)	1 (0.2)	0.35	-
	T	0.006	0.011		1.69 (0.68-4.18)
<i>RASEF</i> (rs4322086)	A/A	260 (39.8)	162 (27.9)	< 0.0001	1.00
	A/G	288 (44.0)	323 (55.6)		1.81 (1.41-2.33)
	G/G	106 (16.2)	96 (16.5)	0.002	1.44 (1.03-2.03)
	G	0.382	0.443		1.29 (1.10-1.51)
<i>LDLR</i> (rs6511720)	G/G	606 (95.0)	520 (93.9)	0.57	1.00
	G/T	30 (4.7)	33 (6.0)		1.28 (0.77-2.13)
	T/T	2 (0.3)	1 (0.2)	0.47	0.58 (0.05-6.37)
	T	0.027	0.032		1.19 (0.74-1.92)
<i>AIM1</i> (rs783396)	C/C	603 (91.8)	518 (90.1)	0.16	1.00
	C/A	54 (8.2)	55 (9.6)		1.18 (0.79-1.75)
	A/A	0 (0.0)	2 (0.4)	0.28	-
	A	0.041	0.051		1.26 (0.86-1.84)
LOC105370913 (rs899997)	T/T	327 (51.9)	302 (56.3)	0.23	1.00
	T/G	253 (40.2)	190 (35.5)		0.81 (0.63-1.03)
	G/G	50 (7.9)	44 (8.2)	0.26	0.96 (0.62-1.49)
	G	0.280	0.259		0.90 (0.75-1.08)

niehs.nih.gov/snpinfo/snpfunc.php) and rSNPBase database (<http://rsnp.psych.ac.cn>). We revealed that polymorphisms rs243865 of *MMP2*, rs11225395 of *MMP8*, and rs2276109 of *MMP12* are located within binding sites for multiple transcription factors. Detailed information on TFBS is shown in Tables S6 to S8). In particular, intronic polymorphism – 1306C > T (rs243865) of the *MMP2* gene (Table S6) contains a functional motif related to the regulatory effects of numerous TFBS such as ZEB1, AR, ATF6, CACD, VDR, TEAD2, NR3C1, HAND1, ZNF354C, NF1, NKX2-5, PAX6, PAX8, PLZF, SP3, SPZ1, SCAP, SCAP, TBX5, VDR, WT1, and ZBTB14. As shown in Table S7, an intronic SNP – 799C > T (rs11225395) of the *MMP8* gene is located within binding sites for transcription factors such as TFAP2A, CUTL1, CRX, DBP, RXRB, NR3C1, HAND1, HNF4A, NF1, POU2F1, POU3F1, PAX5, PAX6, PAX8, PLZF, POU3F2, POU6F1, PPARA, PPARG, RFX1, TBP, and VDR. As can be seen from Table S8, a SNP rs2276109 in the promoter of the *MMP12* gene falls into a recognition site for multiple TFBS such as

JUN, ZEB1, BACH2, BRCA1, CUX1, CUX1, CEBPG, ESR1, NR1H4, FOXQ1, MAF, HOXA9, MYB, MYOD1, NKX2-5, POU2F1, PAX2, PAX3, PAX5, and PAX6. Thus, polymorphisms rs243865 of *MMP2*, rs11225395 of *MMP8*, and rs2276109 of *MMP12* are located within functional motifs shared with multiple signaling pathways and driven by gene regulatory networks.

Additionally, many SNPs were found to possess own regulatory potential, as predicted by the FuncPred tool, or are in linkage disequilibrium with known regulatory SNPs (Table 4). It should be noted that almost all investigated MMPs polymorphisms (except of rs3025058 of *MMP3*) are involved in proximal regulation of the promoter activity of corresponding genes (Tables S9-S15). Moreover, two coding SNPs such as rs17576 of *MMP9* and rs486055 of *MMP10* represent the targets for posttranscriptional regulation of gene expression and/or activity through an RNA-binding protein mechanism (Table S15). Interestingly, several SNPs represent subjects to epigenetic modifications influencing gene activity in various cell types and tissues. In particular,

TABLE 4 The best gene-gene (G×G) interactions significantly associated with the risk of ischemic stroke

Genes and SNPs in interactions	NH	β H	WH	NL	β L	WL	P_{perm}
Two-order G×G interactions							
1 <i>MMP3</i> rs3025058 × <i>RASEF</i> rs4322086	2	0.113	13.17	3	-0.182	34.54	< 0.001
2 <i>MMP9</i> rs17576 × <i>RASEF</i> rs4322086	3	0.160	31.96	2	-0.145	20.25	< 0.001
3 <i>RASEF</i> rs4322086 × <i>ZC3HC1</i> rs11556924	1	0.183	26.66	2	-0.123	14.73	< 0.001
4 <i>MMP8</i> rs11225395 × <i>ZC3HC1</i> rs11556924	2	0.173	25.79	2	-0.119	7.47	< 0.001
Three-order G×G interactions							
1 <i>MMP8</i> rs11225395 × <i>RASEF</i> rs4322086 × <i>ZC3HC1</i> rs11556924	4	0.253	43.39	4	-0.185	18.20	< 0.002
2 <i>MMP1</i> rs1799750 × <i>SLCO1B1</i> rs899997 × <i>RASEF</i> rs4322086	3	0.221	38.21	4	-0.177	28.32	< 0.002
3 <i>MMP2</i> rs243865 × <i>RASEF</i> rs4322086 × <i>ZC3HC1</i> rs11556924	4	0.206	37.42	4	-0.163	24.65	< 0.002
4 <i>MMP3</i> rs3025058 × <i>RASEF</i> rs4322086 × <i>PEMT</i> rs12449964	2	0.252	14.24	6	-0.206	37.39	< 0.002
Four-order G×G interactions							
1 <i>MMP3</i> rs3025058 × <i>MMP8</i> rs11225395 × <i>RASEF</i> rs4322086 × <i>ZC3HC1</i> rs11556924	9	0.366	66.04	7	-0.329	43.20	< 0.005
2 <i>MMP1</i> rs1799750 × <i>MMP9</i> rs17576 × <i>SLCO1B1</i> rs899997 × <i>RASEF</i> rs4322086	7	0.280	58.04	6	-0.233	32.89	< 0.005
3 <i>MMP3</i> rs3025058 × <i>MMP8</i> rs11225395 × <i>RASEF</i> rs4322086 × <i>PEMT</i> rs12449964	9	0.301	57.70	10	-0.290	50.46	< 0.005
4 <i>MMP8</i> rs11225395 × <i>LDLR</i> rs6511720 × <i>RASEF</i> rs4322086 × <i>ZC3HC1</i> rs11556924	6	0.283	55.90	6	-0.192	22.43	< 0.005

Abbreviations: β H, regression coefficient for high-risk exposition in the step 2 analysis; β L, regression coefficient for low-risk exposition in the step 2 analysis; NH, number of significant high-risk genotypes in the interaction; NL, number of significant low-risk genotypes in the interaction; WH, Wald statistic for the high-risk category; WL, Wald statistic for the low-risk category.

Models are obtained using the model-based multifactor dimensionality reduction method, MB-MDR package for R.

P_{perm} , permutation P value for the interaction model. The models were adjusted for age and sex.

The full list of statistically significant models for gene-gene interactions is presented in Tables S3-S5.

various histone marked regions have been found to be associated with polymorphisms such as rs1799750 of *MMP1*, rs243865 of *MMP2*, rs11225395 of *MMP8*, rs17576 of *MMP9*, rs486055 of *MMP10*, and rs2276109 of *MMP12* (Tables S9-S15). We also found that polymorphism rs17576 of the *MMP9* gene spans the region of an exon splice enhancer pointing out that the motif represents a critical site for tissue-specific regulation of gene expression (Table S16). SNPs such as rs243865 of *MMP2*, rs11225395 of *MMP8*, rs17576 of *MMP9*, and rs2276109 of *MMP12* represent expression quantitative trait loci (eQTLs), i.e genomic loci explaining all or a fraction of variation in expression levels of mRNAs.³⁰ For instance, meaningful eQTLs were found for SNP rs11225395 in whole blood ($P = 4.8 \times 10^{-9}$), skeletal muscle ($P = 0.0000092$), subcutaneous adipose tissue ($P = 1.1 \times 10^{-8}$), and also in skin exposed ($P = 2.4 \times 10^{-10}$) and not exposed to sun ($P = 0.0000022$) (Data of GTEx portal, v7, <https://gtexportal.org>).

3.4 | In silico search for potential regulators of MMPs and biological processes related to their effects

GO enrichment analysis on a gene set including transcription factors targeting SNPs of MMPs such as rs11225395 of *MMP8*, rs243865 of *MMP2*, and

rs2276109 of *MMP12* (Tables S6-S8) enabled identification of biological processes, molecular functions and metabolic pathways that could up or down-regulate the functions of the MMPs. Table S17 shows biological processes related to the functions of the given MMPs. Biological processes and metabolic pathways that may play roles in the pathogenesis of IS were of great interest. In particular, the following GO terms related to the regulation of lipid metabolism and atherosclerosis formation were identified for a gene set including TF binding to an SNP rs11225395 of *MMP8*: regulation of sequestering of triglyceride (GO:0010889, $Q = 0.01$), negative regulation of sequestering of triglyceride (GO:0010891, $Q = 0.003$), regulation of cholesterol storage (GO:0010885, $Q = 0.01$), negative regulation of cholesterol storage (GO:0010887, $Q = 0.005$), regulation of fatty acid oxidation (GO:0046320, $Q = 0.02$), positive regulation of fatty acid oxidation (GO:0046321, $Q = 0.01$), negative regulation of lipid storage (GO:0010888, $Q = 0.02$), regulation of lipid metabolic process (GO:0019216, $Q = 0.05$), regulation of macrophage derived foam cell differentiation (GO:0010743, $Q = 0.04$), and negative regulation of macrophage derived foam cell differentiation (GO:0010745, $Q = 0.01$). GO terms related with apoptosis playing a role in neuronal cell death included: positive

TABLE 5 Bioinformatic analysis for the regulatory potential of single nucleotide polymorphisms of genes for matrix metalloproteinases

Gene (SNP ID)	SNP Allele	Location within gene		SNP Function Prediction		Regulatory annotations on SNPs (rSNPBase)						TFBS and other factors associated with SNP				
		intron	exon	TFBS	miRNA	Regulatory potential	rSNP	LD-proxy of rSNP ($r^2 > 0.8$)	Proximal regulation	Distal regulation	miRNA regulation	RNA-binding protein mediated regulation	eQTL	TRANSFAC Database	rSNPBase	
<i>MMP1</i> (rs1799750)	2 G/1 G	intron		No	No	0.006	Yes	No	Yes	No	No	No	No	No	FOS, GATA-2, CEBPB, Table S9	
<i>MMP2</i> (rs243865)	C > T	intron		Yes	No	0.015	Yes	Yes	Yes	No	No	No	Yes	Yes	Table S6	Table S10
<i>MMP3</i> (rs3025058)	5 A/6 A	promoter		No	No	No	No	No	No	No	No	No	No	No	-	-
<i>MMP8</i> (r-s11225395)	C > T	intron		Yes	No	0.200	Yes	Yes	Yes	No	No	No	Yes	Yes	Table S7	Tables S11,S14
<i>MMP9</i> (rs17576)	A > G	exon		No	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	-	POLR2A, TAF1, TBP, ZNF143, Tables S15, S16	
<i>MMP10</i> (rs486055)	C > T	exon		No	No	0.206	Yes	Yes	Yes	No	No	Yes	No	-	Supporting Information 12	
<i>MMP12</i> (rs2276109)	A > G	promoter		Yes	No	0.064	Yes	Yes	Yes	No	No	No	Yes	Yes	Table S8	GATA3, JUN, FOS, Supporting Information 13

TFBS, transcription factor binding sites predicted by the TRANSFAC database on potential transcription factors, their genomic binding sites and DNA binding profiles (BIOBASE Corporation, Wolfenbuettel, Germany, <http://gene-regulation.com>) and available via the SNP Function Prediction (FuncPred) tool hosted by National Institute of Environmental Health Sciences (<https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.php>).

The only TFBS whose core or matrix match score was impacted, or which were eliminated or created by variant sequences were included in this table to be regulatory within a particular SNP (all TFBS identified by TRANSFAC are listed in Tables S6-S8).

rSNPBase is a database of curated regulatory SNPs (<http://rsnp.psych.ac.cn>).

regulation of apoptotic process (GO:0043065, $Q = 0.04$), positive regulation of programmed cell death (GO:0043068, $Q = 0.04$) and positive regulation of cell death (GO:0010942, $Q = 0.05$). GO terms related to biological processes participating in ischemic injury of the brain and its recovery after stroke included: negative regulation of neural precursor cell proliferation (GO:2000178, $Q = 0.02$), regulation of neuroblast proliferation (GO:1902692, $Q = 0.04$), astrocyte development (GO:0014002, $Q = 0.04$), astrocyte differentiation (GO:0048708, $Q = 0.003$), regulation of oligodendrocyte differentiation (GO:0048713, $Q = 0.05$), glial cell development (GO:0021782, $Q = 0.01$), myelination (GO:0042552, $Q = 0.02$), glial cell differentiation (GO:0010001, $Q = 0.003$), gliogenesis (GO:0042063, $Q = 0.008$), cerebral cortex development (GO:0021987, $Q = 0.001$), ensheathment of neurons (GO:0007272, $Q = 0.02$), axon ensheathment (GO:0008366, $Q = 0.02$), central nervous system development (GO:0007417, $Q = 0.00006$), and brain development (GO:0007420, $Q = 0.01$).

For a gene set including TFBS at SNP rs243865 of *MMP2*, we identified GO terms related to a development of the cardiovascular system and regulation of apoptosis: vasculogenesis (GO:0001570, $Q = 0.0001$), blood vessel development (GO:0001568, $Q = 0.0000002$), vasculature development (GO:0001944, $Q = 0.0000002$), blood vessel morphogenesis (GO:0048514, $Q = 0.00001$), cardiovascular system development (GO:0072358, $Q = 0.0000003$), regulation of apoptotic process (GO:0042981, $Q = 0.001$), positive regulation of apoptotic process (GO:0043065, $Q = 0.02$), regulation of programmed cell death (GO:0043067, $Q = 0.001$), positive regulation of programmed cell death (GO:0043068, $Q = 0.02$), regulation of cell death (GO:0010941, $Q = 0.0002$), positive regulation of cell death (GO:0010942, $Q = 0.02$), and negative regulation of cell death (GO:0060548, $Q = 0.02$).

GO terms related with the development of nervous system and processes playing a role in stroke pathogenesis such as apoptosis and oxidative stress were identified for a gene set including TFBS at SNP rs2276109 of *MMP12*: nervous system development (GO:0007399, $Q = 0.008$), regulation of nervous system development (GO:0051960, $Q = 0.009$), positive regulation of nervous system development (GO:0051962, $Q = 0.04$), neurogenesis (GO:0022008, $Q = 0.03$), regulation of neurogenesis (GO:0050767, $Q = 0.04$), positive regulation of neurogenesis (GO:0050769, $Q = 0.004$), regulation of cell death (GO:0010941, $Q = 0.04$), regulation of apoptotic process (GO:0042981, $Q = 0.03$), regulation of programmed cell death (GO:0043067, $Q = 0.03$), cellular response to reactive oxygen species (GO:0034614, $Q = 0.02$), and regulation of reactive oxygen species metabolic process (GO:2000377, $Q = 0.04$).

3.5 | Bioinformatic analysis for functional partners of MMPs involved in the molecular basis of IS

An important objective of the present study was to analyze molecular interactions between MMPs, metabolic pathways, and gene regulatory networks potentially involved in the pathogenesis of IS. The bioinformatic tool of the STRING database enabled identification protein-protein interactions between MMPs associated with IS in our study and their functional partners in the form of interactome network representing the structural, spatial, and functional conjugation of MMPs with other protein components of the network (Figure 1). The structure of an interactome network indicates that the interacting proteins are at least a biologically related group of molecules.²⁹ The present interactome network included 176 protein-protein interactions (PPI) comprising meaningful interactions between 25 proteins at PPI enrichment P value = 1.0×10^{-16} . As indicated in Figure 1, the 10 functional partners that are most tightly interacting with IS-related MMPs represent a core part of the interactome network. These functional partners included: TIMP metalloproteinase inhibitors such as type 1 (TIMP1, Gene ID: 7076, OMIM: 305370), type 2 (TIMP2, Gene ID: 7077, OMIM: 188825), and type 3 (TIMP3, Gene ID: 7078, OMIM: 188826), plasminogen (PLG, Gene ID: 5340, OMIM: 173350), Jun proto-oncogene (JUN, Gene ID: 3725, OMIM: 165160), CD44 molecule (CD44, Gene ID: 960, OMIM: 107269), vascular endothelial growth factor A (VEGFA, Gene ID: 7422, OMIM: 192240), basigin or Ok blood group (BSG, Gene ID: 682, OMIM: 109480), nuclear factor kappa B subunit 1 (NFKB1, Gene ID: 4790, OMIM: 164011), and tumor protein p53 (TP53, Gene ID: 7157, OMIM: 191170). GO enrichment analysis performed by bioinformatic tools of the STRING database enabled identification of GO biological processes terms associated with IS-related MMPs (Table S17), their molecular functions and cellular components (Table S18) as well as the KEGG Pathways (Table S19) in which the MMPs are potentially involved.

Finally, biological processes reflecting the functions of MMPs and related with pathogenesis of IS were selected from the GO list (GO terms listed below are ranked by descending a degree of involvement of the MMPs in the biological processes): extracellular matrix disassembly (GO:0022617, $Q = 5 \times 10^{-15}$), collagen catabolic process (GO:0030574, $Q = 1 \times 10^{-7}$), response to oxidative stress (GO:0006979, $Q = 1 \times 10^{-6}$), regulation of proteolysis (GO:0030162, $Q = 1 \times 10^{-6}$), regulation of response to stress (GO:0080134, $Q = 2 \times 10^{-6}$), regulation of cell death (GO:0010941, $Q = 3 \times 10^{-6}$), response to toxic substance (GO:0009636, $Q = 6 \times 10^{-6}$), cell

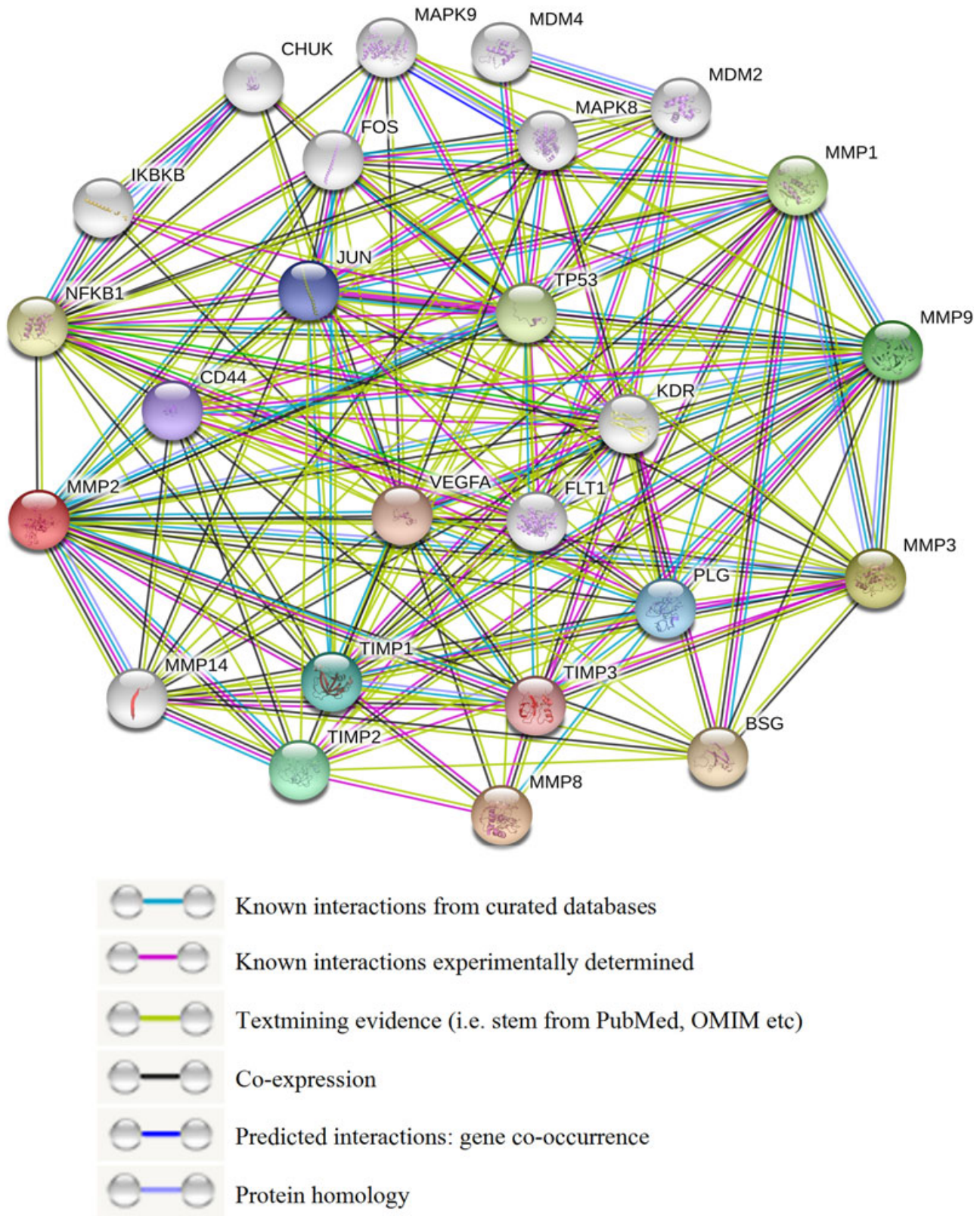


FIGURE 1 Interactome network representing protein-protein interactions between matrix metalloproteinases associated with the risk of ischemic stroke and other proteins. Network nodes represent proteins: splice isoforms or posttranslational modifications are collapsed, that is, each node represents all the proteins produced by a single, protein-coding gene locus. Colored nodes: query proteins and first shell of interactors; white nodes: second shell of interactors; empty nodes: proteins of unknown 3D structure; filled nodes: some 3D structure is known or predicted. Edges represent protein-protein associations: associations are meant to be specific and meaningful, that is, proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other

migration (GO:0016477, $Q = 1 \times 10^{-5}$), response to mechanical stimulus (GO:0009612, $Q = 2 \times 10^{-5}$), aging (GO:0007568, $Q = 5 \times 10^{-5}$), angiogenesis (GO:0001525, $Q = 0.0001$), blood coagulation (GO:0007596, $Q = 0.0002$), positive regulation of cell death (GO:0010942, $Q = 0.0003$), regulation of cell migration (GO:0030334, $Q = 0.0003$), blood vessel morphogenesis (GO:0048514, $Q = 0.0003$), response to chemical (GO:0042221, $Q = 0.0006$), blood vessel development (GO:0001568, $Q = 0.0008$), stress-activated MAPK cascade (GO:0051403, $Q = 0.0008$), response to hypoxia (GO:0001666, $Q = 0.0009$), vasculature development (GO:0001944, $Q = 0.001$), proteolysis (GO:0006508, $Q = 0.001$), positive regulation of cell proliferation (GO:0008284, $Q = 0.001$), positive regulation of cell migration (GO:0030335, $Q = 0.002$), regulation of cell-substrate adhesion (GO:0010810, $Q = 0.002$), cellular response to oxidative stress (GO:0034599, $Q = 0.002$), regulation of MAPK cascade (GO:0043408, $Q = 0.003$), response to reactive oxygen species (GO:0000302, $Q = 0.003$), positive regulation of neuron death (GO:1901216, $Q = 0.003$), positive regulation of apoptotic signaling pathway (GO:2001235, $Q = 0.003$), positive regulation of endothelial cell proliferation (GO:0001938, $Q = 0.003$), positive regulation of MAPK cascade (GO:0043410, $Q = 0.005$), inflammatory response (GO:0006954, $Q = 0.005$), regulation of cellular response to stress (GO:0080135, $Q = 0.006$), regulation of cell proliferation (GO:0042127, $Q = 0.006$), apoptotic process (GO:0006915, $Q = 0.006$), platelet degranulation (GO:0002576, $Q = 0.007$), positive regulation of macrophage derived foam cell differentiation (GO:0010744, $Q = 0.007$), cellular response to reactive oxygen species (GO:0034614, $Q = 0.007$), cellular response to hypoxia (GO:0071456, $Q = 0.01$), positive regulation of apoptotic process (GO:0043065, $Q = 0.01$), regulation of cell adhesion (GO:0030155, $Q = 0.01$), regulation of reactive oxygen species metabolic process (GO:2000377, $Q = 0.02$), programmed necrotic cell death (GO:0097300, $Q = 0.02$), positive regulation of cytokine production (GO:0001819, $Q = 0.02$), regulation of apoptotic signaling pathway (GO:2001233, $Q = 0.03$), programmed cell death (GO:0012501, $Q = 0.03$), positive regulation of neuron apoptotic process (GO:0043525, $Q = 0.04$), and cellular response to stress (GO:0033554, $Q = 0.05$).

4 | DISCUSSION

4.1 | Joint contribution of MMP genes to the pathogenesis of IS

To date, three dozens of association studies have been conducted to elucidate associations of polymorphic genes encoding MMPs with the risk of IS in various populations of the world. Despite these studies have yielded incon-

sistent results, recent meta-analysis¹⁰ enabled identification two polymorphisms such as $-1562C/T$ of *MMP9* and $-1082A/G$ of *MMP12* as meaningful markers for IS susceptibility. It is important to note, that all the studies investigated associations between particular MMP genes and IS risk have been conducted on populations with small sample size, reducing the power to obtain reliable estimates for genetic effects of MMPs on disease risk. The present study is the first attempt to assess the joint involvement of genes encoding various MMPs in the multistep molecular mechanisms of cerebrovascular disease on a large population of the Central Russia. The present study demonstrated, for the first time, that MMPs is an important class of genes collectively involved in various stages of cerebrovascular disease, starting from the formation of atherosclerosis in carotid arteries, their occlusion to ischemic injury of the brain and its recovery after stroke.

Our study was the first to identify an association between polymorphism rs11225395 ($-799C > T$) of the *MMP8* gene and IS risk. A recent study of Han et al³¹ conducted on mice has identified that *MMP8* represents a critical factor for brain damage in transient focal cerebral ischemia via modulating neuroinflammation likely microglial activation and production of tumor necrosis factor alpha. This means the rs11225395 polymorphism of the *MMP8* gene may be considered as a novel genetic marker of susceptibility to IS however this association should be confirmed by independent studies. The present study has confirmed the results of meta-analysis 27 studies¹⁰ that polymorphisms rs1799750 ($-1607 G/1G$) of *MMP1*, rs243865 ($-1306C > T$) of *MMP2* and rs3025058 ($-1612 5A/6A$) of *MMP3* are not associated with the risk for IS. However a recent meta-analysis conducted by Zhang et al⁶ and including ethnically diverse populations has observed that rs1799750 of *MMP1* and also rs2276109 of *MMP12* are associated with the risk IS, although a relationship between the polymorphisms and disease risk varied substantially across different races. Clear tendencies in the relationship between SNPs rs2276109 of *MMP12* and rs17576 of *MMP9* with IS risk was also seen in our study, but the association did not reach statistical significance. In addition, our study has successfully confirmed associations of polymorphisms rs4322086 of *RASEF* and rs11556924 of *ZC3HC1* with increased risk of IS, SNPs originally established by GWAS in European populations.^{16,17}

It is well known that gene-gene interaction (or epistasis) is an important genetic mechanism underlying susceptibility to complex polygenic disorders.³² Since gene-gene interaction represents the rule rather than the exception for genes determining complex disease

phenotypes like IS, the present study was also designed to evaluate whether epistatic interactions between MMP genes and the GWAS loci interact to each other to produce the disease phenotype. We found that rs4322086 of *RASEF* and rs11556924 of *ZC3HC1* and some other GWAS loci interact tightly to MMP genes pointing out that the genes are collectively involved in the molecular basis of IS. The MB-MDR enabled identification of numerous high-order gene-gene interaction models, associated with IS risk. Notably, a large portion of the G×G interaction models comprised various MMP genes pointing out that these genes are characterized by a greater contribution to the polygenic mechanisms of IS risk than disease susceptibility loci established by GWAS. Although one-way statistical analysis failed to show associations of disease risk with a majority of the investigated MMPs, the observed gene-gene interactions demonstrate a greater contribution of these genes to the development of IS when they act in cooperation than alone. To the best of our knowledge, this is the first study reporting that complex interactions between MMP genes and known disease susceptibility genes constitute the molecular basis for the development of IS.

The involvement of MMPs in the complex pathophysiology of IS can be supported by numerous publications in the literature.^{4,7-9} The GO enrichment analysis of a set of transcription factors associated with MMP genes has identified biological functions of MMPs that may play roles in the development of IS. These functions include: (1) angiogenesis and induction of cell proliferation in the vascular wall; (2) increased cell migration and adhesion; (3) destruction of the extracellular matrix around the atherosclerotic plaque leading to its destabilization and rupture; (4) activation of blood coagulation and intravascular thrombus formation; (5) stimulation of vascular inflammation; and (6) apoptosis of cells including neurons. Based on the results of the GO enrichment analysis and PubMed publications on biological functions of MMPs, we summarized the potential mechanisms by which MMPs are involved in the multistep pathogenesis of the cerebrovascular disease (Figure 2). As indicated in Figure 2, there are at least three stages of the development of cerebrovascular disease in which MMPs are involved: (1) atherosclerotic lesions of cerebral arteries, (2) acute cerebral artery occlusion, and (3) ischemic injury of the brain. First, MMPs is a part of molecular and

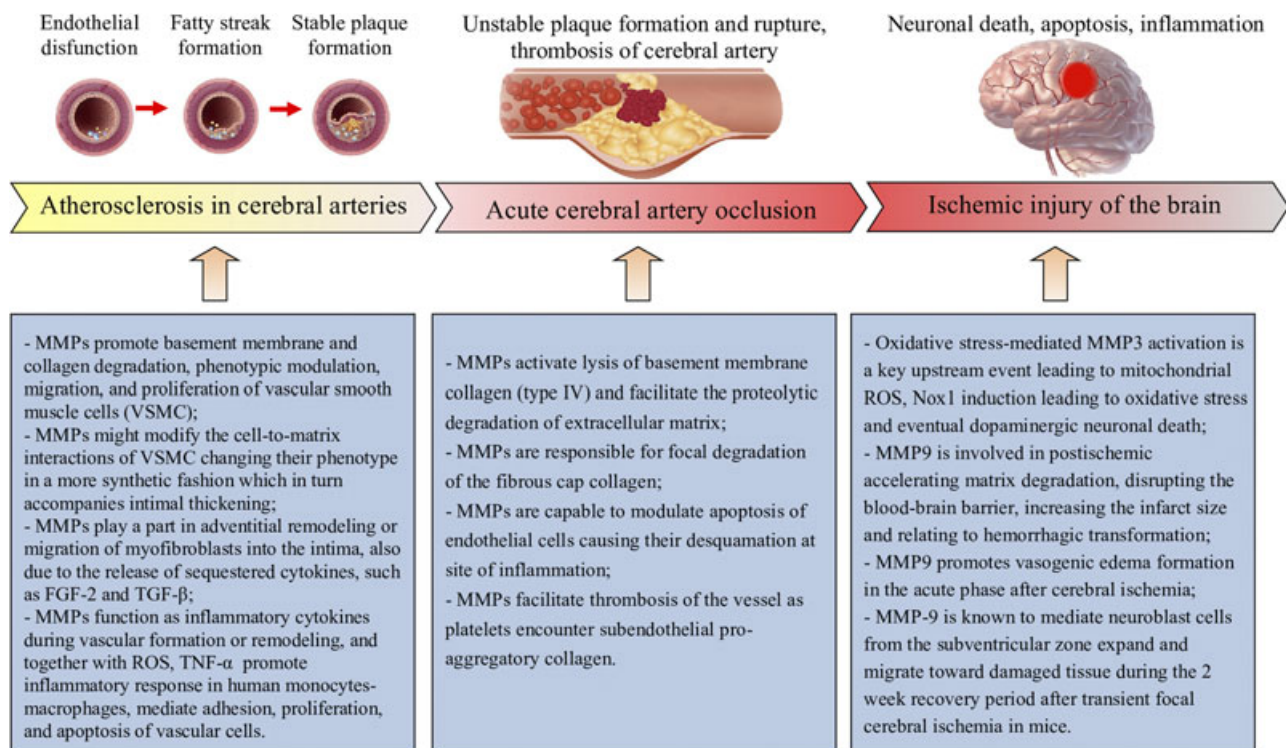


FIGURE 2 Pleiotropic effects of matrix metalloproteinases on multistep pathogenesis of the cerebrovascular disease. At least three points of application for effects of matrix metalloproteinases (MMP1, MMP2, MMP3, MMP8, and MMP9 associated with the ischemic stroke risk) to the pathogenesis of ischemic stroke are suggested: (1) the involvement in multiple stages of atherosclerosis of cerebral arteries; (2) acute cerebral artery occlusion resulting from a rupture of instable atherosclerotic plaque and/or thrombosis of cerebral artery; (3) the involvement in pathological processes during and after ischemic injury of the brain. Conclusions on the effects of these MMPs on the pathogenesis of ischemic stroke are based on the bioinformatics analysis of Gene Ontology (GO) terms significantly associated with the gene set (false discovery rate < 0.05). A detailed description of GO terms is present in the body of text

cellular pathways driving the atherosclerotic process in the vessels. In particular, MMPs are involved in the formation and progression of the atherosclerotic process through the positive regulation of biological functions such as cell migration and proliferation (including endothelial cells), differentiation of foam cells from macrophages, MAPK cascade activation, enhanced generation of reactive oxygen species, cytokine production, inflammation response, and angiogenesis of blood vessels.⁷ Second, MMPs are known to be responsible for both the rupture of unstable plaque and intraluminal thrombus formation in the artery. Interestingly, MMPs may be activated in response to an exposure to various environmental stressors such as tobacco smoking and drinking alcohol,^{33,34} well recognized etiological factors of cerebrovascular disease. The activated enzymes are known to be involved in the degradation of extracellular matrix, activation of collagen catabolism, induction of proteolysis, inflammation, platelet degranulation, and plasminogen activation, the processes leading to the destabilization and rupture of atherosclerotic plaque and thrombosis of the cerebral artery.³⁵⁻³⁷ Finally, MMPs mediate many biological and pathological processes during and after ischemic injury of the brain. In particular, MMPs participate in the regulation of many processes in ischemic brain tissue such as induction of neuronal apoptosis, activation of cytokine production and inflammation, thereby influencing the severity of brain damage, and activity of repairing the damaged tissue.³⁸⁻⁴⁰ Moreover, many functional partners of MMPs and their coregulated transcription factors such as members of the FOS and JUN gene families, that have been identified in the present study by GO enrichment analysis, play a crucial role in controlling the cell cycle, development, growth, and cell differentiation in the central nervous system, thereby demonstrating an importance of MMPs in the regulation of recovery processes in brain tissues after IS.

4.2 | MMP genes as targets for gene regulatory networks

A comprehensive bioinformatic analysis enabled establishing that the studied MMP gene polymorphisms possess a clear regulatory potential and may be targeted by gene regulatory networks driving molecular and cellular pathways related to the pathogenesis of the cerebrovascular disease. For instance, we found that several SNPs of MMP genes fall into a recognition site for multiple transcription factors. TFBS are short sequences on the DNA representing targets for transcription factors driving and controlling gene expression and orchestrating gene activity.⁴¹ In other words, transcription factors can

be considered as effector nodes affecting the behavior of their downstream target genes, and a gene regulatory network represents a collection of regulatory interactions between transcription factors and their target genes.⁴²

Polymorphisms of MMP genes (ie *MMP8*, *MMP9*, *MMP1*, *MMP2*, and *MMP3*) which alone or in combination with other genes were found to be associated with the risk of IS in our study represent a great interest, as they give important information on how transcriptional regulation of disease-associated genes is organized at the genome-wide level. For instance, we found transcription factors such as FOS and GATA-2 and CEBPB to be responsible for proximal regulation of activity/expression of the *MMP1* gene through the binding to an SNP rs1799750. Interestingly, CEBPB is an important transcription factor that binds to the promoter regions of numerous genes involved in inflammatory processes, synergistically enhancing and supporting their expression in response to pro-inflammatory stimulation.⁴³ In addition, polymorphisms such as rs1799750 of *MMP1*, rs243865 of *MMP2*, rs11225395 of *MMP8*, rs17576 of *MMP9* represent the targets for epigenetic regulation of gene expression through posttranslational modifications to histone proteins.

Association of SNP rs11225395 of the *MMP8* gene with IS risk was a major finding of our study. This polymorphism contains functional motifs related to the regulatory effects of transcription factors which, on the one hand, regulate lipid and carbohydrate metabolism, inflammation, and other pathways that drive the progression of atherosclerosis; on the other hand, such TFs are responsible for the regulation of neurogenesis and other important biological processes in the brain. It is reasonable to explore which gene regulatory networks are capable to control the expression of the *MMP8* gene through the binding of TFs to a motif at SNP rs11225395. NR3C1 is a transcription factor binding to glucocorticoid response elements at targeted genes, affecting inflammatory responses, cellular proliferation, and differentiation in target tissues.⁴⁴ HNF4A is a transcriptionally controlled TF binding to DNA sites required for the transcription of numerous genes including apolipoprotein CIII, a multifaceted protein involved in the metabolism of triglyceride-rich lipoproteins, associating with hypertriglyceridemia and formation of atherosclerotic lesions and other pathological processes playing a role in atherosclerosis.⁴⁵ VDR is a nuclear receptor for calcitriol, the active form of vitamin D3 which mediates the action of this vitamin on cells. Study of Szeto et al⁴⁶ provided strong evidence supporting an antiatherosclerotic role of the VDR signaling through the mechanism of suppressing the activation of the local renin-angiotensin system in macrophages/leukocytes and within the

atherosclerotic lesion. PLZF encodes a finger-type transcription factor, involved in physiological development, proliferation, differentiation and apoptosis, processes contributing to atherosclerosis.⁴⁷ PPARA and PPARG (peroxisome proliferator-activated receptors) are members of the nuclear receptor superfamily of ligand-inducible transcription factors, key regulators of peroxisomal and mitochondrial β -oxidation of fatty acids, ketone body synthesis and systemic lipid metabolism and also possess anti-inflammatory activities.⁴⁸ PPARA has the potential to affect atherosclerosis via the regulation of lipid metabolism and impact on the vascular inflammation.⁴⁸ PPARG is also crucial for controlling gene networks involved in glucose homeostasis.⁴⁹ Peroxisome proliferator-activated receptors are expressed in multiple brain regions especially in neurons.⁵⁰ Notably, PPAR agonists have well-documented anti-inflammatory and neuroprotective roles in the central nervous system.⁵⁰ We also identified several transcription factors possessing to mediate effects of the gene regulatory networks on neuronal development and remodeling in the brain through the control of MMP expression. For instance, POU6F1 is required for proper dendrite specification and patterning, as well as synapse development and function in adult-born neurons.⁵¹ Interestingly, POU6F1 promotes transcription of neuronal genes and in cooperation with transcription factors ASCL1 and MYT1L is capable to reprogram fibroblasts and other somatic cells into induced neuronal cells,⁵² making the gene regulatory network operating by transcription factor POU6F1 an attractive target to induce regeneration of the damaged brain and restore its function after an ischemic attack. In addition, the enrichment analysis for the gene set including TFBS identified at SNP rs11225395 of *MMP8* enabled finding biological processes which can be related to disease pathogenesis through the regulation of lipid metabolism, formation of atherosclerotic lesions and cellular apoptosis as well as processes playing an important role in ischemic injury of the brain and its recovery after stroke attack.

4.3 | Study limitations

Our study has some limitations. First, many important polymorphic genes for MMPs were not included in the study. In particular, our study had no opportunity to confirm robust associations of polymorphisms 1562 C/T of *MMP9* and -1082A/G of *MMP12* with the risk IS, since these SNPs were not included in the study. Second, the associations observed in our study need to be confirmed in independent populations. Third, the roles of interactions between MMPs and environmental factors such as smoking, heavy drinking and lifestyle factors in

determining the risk of IS were not investigated in the study. Although we have explored comprehensively the functional potential of SNPs using bioinformatic methods, further studies are also required to investigate how the polymorphisms influence the concentration of MMPs in carotid arteries and the brain, thereby substantiating their effects on disease pathogenesis.

5 | CONCLUSIONS

Despite the discrepancies observed between the conducted studies polymorphic genes for MMPs are related with susceptibility to IS, but their effects on disease risk are ethnically dependent. This means that differences in the genetic architecture between races and ethnicities as well as population-specific environmental exposures may influence the results of genetic studies of MMP genes in IS. A joint contribution of MMPs to IS susceptibility has been little studied, and so the findings of our study clearly indicate that this class of genes deserves special attention as potential genetic factors influencing the multistep mechanisms of cerebrovascular disease. The pleiotropic biological effects of MMPs on various pathogenetic stages of the cerebrovascular disease may include the involvement of MMPs in the formation of atherosclerosis, atherothrombosis of cerebral vessels and the contribution of the enzymes to brain damage and its subsequent recovery. Complex interactions between various MMPs as well as with the known disease susceptibility genes demonstrate that MMPs may be collectively involved in the polygenic mechanisms of cerebrovascular disease. Further studies of independent ethnic populations are needed to substantiate the role of MMPs genes in the pathogenesis of IS and to investigate the mechanisms by which the polymorphisms influence the concentration of these proteins in carotid arteries and the brain. Moreover, genetic studies focusing on a wider spectrum of polymorphic genes encoding MMPs, their gene-gene, and gene-environment interactions can provide a better understanding the molecular mechanisms by which the enzymes are involved in the complex pathophysiology of cerebrovascular disease and end up in the discovery of novel drug targets for treatment and prevention of the disease.

ACKNOWLEDGMENTS

The study was supported by the Russian Science Foundation (No.15-15-10010).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA ACCESSIBILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Alexey Polonikov  <http://orcid.org/0000-0001-6280-247X>

REFERENCES

1. Feigin VL, Norrving B, Mensah GA. Global burden of stroke. *Circ Res*. 2017;120:439-448. <https://doi.org/10.1161/CIRCRESAHA.116.308413>
2. Dichgans M. Genetics of ischaemic stroke. *Lancet Neurol*. 2007;6:149-161. [https://doi.org/10.1016/S1474-4422\(07\)70028-5](https://doi.org/10.1016/S1474-4422(07)70028-5)
3. Hassanm A, Markusm HS. Genetics and ischaemic stroke. *Brain*. 2000;123:1784-1812.
4. Gasche Y, Soccia PM, Kanemitsu M, Copin JC. Matrix metalloproteinases and diseases of the central nervous system with a special emphasis on ischemic brain. *Front Biosci*. 2006;11:1289-1301.
5. Wang B, Wang Y, Zhao L. MMP-9 gene rs3918242 polymorphism increases risk of stroke: a meta-analysis. *J Cell Biochem*. 2018;119(9801-9808):9801-9808. <https://doi.org/10.1002/jcb.27299>
6. Zhang G, Li W, Guo Y, Li D, Liu Y, Xu S. MMP gene polymorphisms, MMP-1 -1607 1G/2G, -519 A/G, and MMP-12 -82 A/G, and ischemic stroke: a meta-analysis. *J Stroke Cerebrovasc Dis*. 2018;27:140-152. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2017.08.021>
7. Lin J, Kakkar V, Lu X. Impact of matrix metalloproteinases on atherosclerosis. *Curr Drug Targets*. 2014;15:442-453.
8. Moskalenko MI. The involvement of genes of matrix metalloproteinases in the development of arterial hypertension and its complication (review). *Research Result. Med Pharmacy*. 2018;4:53-69. <https://doi.org/10.18413/2313-8955-2018-4-1-53-69>
9. Rosenberg GA. Matrix metalloproteinases in neuroinflammation. *GLIA*. 2002;39:279-291. <https://doi.org/10.1002/glia.10108>
10. Misra S, Talwar P, Kumar A, et al. Association between matrix metalloproteinase family gene polymorphisms and risk of ischemic stroke: A systematic review and meta-analysis of 29 studies. *Gene*. 2018;672:180-194. <https://doi.org/10.1016/j.gene.2018.06.027>
11. 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. *J Mol Neurosci*. 2012;47:511-513. <https://doi.org/10.1007/s12031-012-9764-y>
12. Polonikov AV, Vialykh EK, Churnosov MI, et al. The C718T polymorphism in the 3'-untranslated region of glutathione peroxidase-4 gene is a predictor of cerebral stroke in patients with essential hypertension. *Hypertens Res*. 2012;35:507-512. <https://doi.org/10.1038/hr.2011.213>
13. Bushueva OY, Stetskaya TA, Polonikov AV, Ivanov VP. The relationship between polymorphism 640A>G of the CYBA gene with the risk of ischemic stroke in the population of the Central Russia. *Zh Nevrol Psikhiatr Im S S Korsakova*. 2015;115(9(2)):38-41. <https://doi.org/10.17116/jnevro20151159238-41>
14. Cheng YC, Stanne TM, Giese AK, et al. Genome-wide association analysis of young-onset stroke identifies a locus on chromosome 10q25 near HABP2. *Stroke*. 2016;47:307-316. <https://doi.org/10.1161/STROKEAHA.115.011328>
15. Hindy G, Engström G, Larsson SC, et al. Role of blood lipids in the development of ischemic stroke and its subtypes: a Mendelian Randomization Study. *Stroke*. 2018;49:820-827. <https://doi.org/10.1161/STROKEAHA.117.019653>
16. Traylor M, Malik R, Nalls MA, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. *Ann Neurol*. 81, 2017:383-394. <https://doi.org/10.1002/ana.24840>
17. Dichgans M, Malik R, König IR, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke*. 2014;45:24-36. <https://doi.org/10.1161/STROKEAHA.113.002707>
18. Chauhan G, Arnold CR, Chu AY, et al. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2016; 15:695-707. [https://doi.org/10.1016/S1474-4422\(16\)00102-2](https://doi.org/10.1016/S1474-4422(16)00102-2)
19. Carty CL, Buzková P, Fornage M, et al. Associations between incident ischemic stroke events and stroke and cardiovascular disease-related genome-wide association studies single nucleotide polymorphisms in the population architecture using genomics and epidemiology study. *Circ Cardiovasc Genet*. 2012;5:210-216. <https://doi.org/10.1161/CIRCGENETICS.111.962191>
20. Solé X, Guinó E, Valls J, Iniesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics*. 2006;22:1928-1929. <https://doi.org/10.1093/bioinformatics/btl268>
21. Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*. 2001;69:138-147. <https://doi.org/10.1086/321276>
22. Ponomarenko IV. Using the method of multifactor dimensionality reduction (MDR) and its modifications for analysis of gene-gene and gene-environment interactions in genetic-epidemiological studies (review). *Res Result Biomed*. 2019;5:4-21. <https://doi.org/10.18413/2313-8955-2019-5-1-0-1>
23. Calle ML, Urrea V, Vellalta G, Malats N, Steen KV. Improving strategies for detecting genetic patterns of disease susceptibility in association studies. *Stat Med*. 2008;27:6532-6546. <https://doi.org/10.1002/sim.3431>
24. Calle ML, Urrea V, Malats N, Van Steen K. mbmdr: an R package for exploring gene-gene interactions associated with binary or quantitative traits. *Bioinformatics*. 2010;26:2198-2199. <https://doi.org/10.1093/bioinformatics/btq352>
25. Polonikov AV, Bushueva OY, Bulgakova IV, et al. A comprehensive contribution of genes for aryl hydrocarbon receptor signaling pathway to hypertension susceptibility. *Pharmacogenet Genom*. 2017;27:57-69. <https://doi.org/10.1097/FPC.0000000000000261>
26. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37:600-605. <https://doi.org/10.1093/nar/gkp290>

27. Guo L, Du Y, Chang S, Zhang K, Wang J. rSNPBase: a database for curated regulatory SNPs. *Nucleic Acids Res.* 2014;42:1033-1039. <https://doi.org/10.1093/nar/gkt1167>
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B.* 1995;57:289-300.
29. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45:362-368. <https://doi.org/10.1093/nar/gkw937>
30. Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet.* 2006;7:862-872. <https://doi.org/10.1038/nrg1964>
31. Han JE, Lee EJ, Moon E, Ryu JH, Choi JW, Kim HS. Matrix metalloproteinase-8 is a novel pathogenetic factor in focal cerebral ischemia. *Mol Neurobiol.* 2016;53:231-239. <https://doi.org/10.1007/s12035-014-8996-y>
32. Carlborg O, Haley CS. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet.* 2004;5:618-625. <https://doi.org/10.1038/nrg1407>
33. Koken T, Gursoy F, Kahraman A. Long-term alcohol consumption increases pro-matrix metalloproteinase-9 levels via oxidative stress. *J Med Toxicol.* 2010;6:126-130. <https://doi.org/10.1007/s13181-010-0081-y>
34. Sivaraman SK, Zachariah G, Annamala P. Effect of smoking on metalloproteinases (MMPs) activity in patients with acute myocardial infarction (AMI). *J Clin Diagn Res.* 2014;8:27-30. <https://doi.org/10.7860/JCDR/2014/7052.3998>
35. Ketelhuth DF, Bäck M. The role of matrix metalloproteinases in atherothrombosis. *Curr Atheroscler Rep.* 2011;13:162-169. <https://doi.org/10.1007/s11883-010-0159-7>
36. Ruddy JM, Ikonomidis JS, Jones JA. Multidimensional contribution of matrix metalloproteinases to atherosclerotic plaque vulnerability: multiple mechanisms of inhibition to promote stability. *J Vasc Res.* 2016;53:1-16. <https://doi.org/10.1159/000446703>
37. Toutouzias K, Synetos A, Nikolaou C, Tsiamis E, Tousoulis D, Stefanadis C. Matrix metalloproteinases and vulnerable atheromatous plaque. *Curr Top Med Chem.* 2012;12:1166-1180.
38. Morancho A, Rosell A, García-Bonilla L, Montaner J. Metalloproteinase and stroke infarct size: role for anti-inflammatory treatment? *Ann NY Acad Sci.* 2010;1207:123-133. <https://doi.org/10.1111/j.1749-6632.2010.05734.x>
39. Rodríguez JA, Sobrino T, Orbe J, et al. proMetalloproteinase-10 is associated with brain damage and clinical outcome in acute ischemic stroke. *J Thromb Haemost.* 2013;11:1464-1473. <https://doi.org/10.1111/jth.12312>
40. Ulrich NH, Dehmel T, Wittsack HJ, Kieseier BC, Seitz RJ. Peripheral blood levels of matrix metalloproteinase-9 predict lesion volume in acute stroke. *Neurol Sci.* 2013;34:379-382. <https://doi.org/10.1007/s10072-012-0999-8>
41. Inukai S, Kock KH, Bulyk ML. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev.* 2017;4:110-119. <https://doi.org/10.1016/j.gde.2017.02.007>
42. Li Y, Varala K, Coruzzi GM. From milliseconds to lifetimes: tracking the dynamic behavior of transcription factors in gene networks. *Trends Genet.* 2015;31:509-515. <https://doi.org/10.1016/j.tig.2015.05.005>
43. Poli V. The role of C/EBP isoforms in the control of inflammatory and native immunity functions. *J Biol Chem.* 1998;273:29279-29282.
44. Vitellius G, Fagart J, Delemer B, et al. Three novel heterozygous point mutations of NR3C1 causing glucocorticoid resistance. *Hum Mutat.* 2016;37:794-803. <https://doi.org/10.1002/humu.23008>
45. Luo M, Peng D. The emerging role of apolipoprotein C-III: beyond effects on triglyceride metabolism. *Lipids Health Dis.* 2016;15:184. <https://doi.org/10.1186/s12944-016-0352-y>
46. Szeto FL, Reardon CA, Yoon D, et al. Vitamin D receptor signaling inhibits atherosclerosis in mice. *Mol Endocrinol.* 2012;26:1091-1101. <https://doi.org/10.1210/me.2011-1329>
47. Bernardo MV, Yelo E, Gimeno L, Campillo JA, Parrado A. Identification of apoptosis-related PLZF target genes. *Biochem Biophys Res Commun.* 2007;359:317-322. <https://doi.org/10.1016/j.bbrc.2007.05.085>
48. Varga T, Czimmerer Z, Nagy L. PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation. *Biochim Biophys Acta.* 2011;1812:1007-1022. <https://doi.org/10.1016/j.bbadis.2011.02.014>
49. Ahmadian M, Suh JM, Hah N, et al. PPAR γ signaling and metabolism: the good, the bad and the future. *Nature Med.* 2013;19:557-566. <https://doi.org/10.1038/nm.3159>
50. Warden A, Truitt J, Merriman M, et al. Localization of PPAR isotypes in the adult mouse and human brain. *Sci Rep.* 2016;6:27618. <https://doi.org/10.1038/srep27618>
51. McClard CK, Kochukov MY, Herman I, et al. POU6f1 mediates neuropeptide-dependent plasticity in the adult brain. *J Neurosci.* 2018;38:1443-1461. <https://doi.org/10.1523/JNEUROSCI.1641-17.2017>
52. Wapinski OL, Vierbuchen T, Qu K, et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell.* 2013;155:621-635. <https://doi.org/10.1016/j.cell.2013.09.028>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: 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. *J Cell Biochem.* 2019;1-16. <https://doi.org/10.1002/jcb.28815>