

Original Research

Integration of Mendelian randomisation and systems biology models to identify novel blood-based biomarkers for stroke

Tania Islam^a, Md Rezanur Rahman^b, Asaduzzaman Khan^a, Mohammad Ali Moni^{a,*}

^a School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, St Lucia, QLD 4067, Australia

^b Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia



ARTICLE INFO

Keywords:

Stroke
Genome-wide association studies
Expression quantitative trait loci
DNA methylation quantitative trait loci
Summary data-based Mendelian randomisation
DNA methylation
Gene expression

ABSTRACT

Stroke is the second largest cause of mortality in the world. Genome-wide association studies (GWAS) have identified some genetic variants associated with stroke risk, but their putative functional causal genes are unknown. Hence, we aimed to identify putative functional causal gene biomarkers of stroke risk. We used a summary-based Mendelian randomisation (SMR) approach to identify the pleiotropic associations of genetically regulated traits (i.e., gene expression and DNA methylation) with stroke risk. Using SMR approach, we integrated *cis*-expression quantitative loci (*cis*-eQTLs) and *cis*-methylation quantitative loci (*cis*-mQTLs) data with GWAS summary statistics of stroke. We also utilised heterogeneity in dependent instruments (HEIDI) test to distinguish pleiotropy from linkage from the observed associations identified through SMR analysis. Our integrative SMR analyses and HEIDI test revealed 45 candidate biomarker genes ($FDR < 0.05$; $P_{HEIDI} > 0.01$) that were pleiotropically or potentially causally associated with stroke risk. Of those candidate biomarker genes, 10 genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *COL4A1*, *BAG4*, *NEK6*, *SH2B3*, *SH3PXD2A*, *ACAD10*) were differentially expressed in genome-wide blood transcriptomics data from stroke and healthy individuals ($FDR < 0.05$). Functional enrichment analysis of the identified candidate biomarker genes revealed gene ontologies and pathways involved in stroke, including “cell aging”, “metal ion binding” and “oxidative damage”. Based on the evidence of genetically regulated expression of genes through SMR and directly measured expression of genes in blood, our integrative analysis suggests ten genes as blood biomarkers of stroke risk. Furthermore, our study provides a better understanding of the influence of DNA methylation on the expression of genes linked to stroke risk.

1. Introduction

Stroke is the second largest cause of mortality in the world and one of the major causes of long-term disability [1]. It affects over 15 million individuals worldwide, kills over 5.7 million people, and leaves 5 million people permanently disabled [2]. Stroke is a neurological condition characterised by obstruction of blood circulation in the affected regions of the brain, depriving brain tissues of receiving oxygen and nutrients, which results in the death of brain cells and damage to the central nervous system via cerebral infarction, intracerebral haemorrhage or subarachnoid haemorrhage [3]. Stroke is a heterogeneous disease with multiple subtypes, each with its own aetiology and risk factors [1]. The pathophysiological processes of stroke are very complex, involving numerous molecular events, including energy production failure, excitotoxicity, free-radical mediated toxicity, disruption of

blood–brain barriers and immune system dysfunction, which lead to the death of neuronal cells in the brain [4]. The biomarkers for stroke risk assessment of individuals are currently unavailable, warranting the identification of potential biomarkers for stroke risk, which may enhance the stroke risk assessment and develop novel therapeutics.

Past genome-wide association studies (GWAS) have identified several risk loci associated with stroke, including rs2304556 (located in *FMNL2*) and rs1986743 (located in *ARL6IP6*) [5–10]. The latest GWAS meta-analysis of stroke has revealed 89 risk loci associated with stroke [11]. In addition, following GWAS meta-analysis, post-GWAS gene-based association analyses, including MAGMA [12], transcriptome-wide association analysis (TWAS) [13], identified putative genes (e.g., *SH3PXD2A*, *FURIN*) associated with stroke and stroke subtypes [11]. However, the detection of putative causal genes is very challenging due to complex linkage disequilibrium [14] and the occurrence of the

* Corresponding author.

E-mail address: m.moni@uq.edu.au (M.A. Moni).

<https://doi.org/10.1016/j.jbi.2023.104345>

Received 29 October 2022; Received in revised form 4 February 2023; Accepted 16 March 2023

Available online 21 March 2023

1532-0464/© 2023 Elsevier Inc. This article is made available under the Elsevier license (<http://www.elsevier.com/open-access/userlicense/1.0/>).

majority of genetic variants associated with complex traits in non-coding regions, which are presumed to affect disease phenotype through gene expression regulation having no direct effect on protein structure or function [15]. Past reports have highlighted that trait-linked genetic variants are more enriched to expression quantitative trait loci (eQTLs) or DNA methylation quantitative loci (mQTLs), suggesting genetic variant-trait association could act through gene expression regulation [16–19].

Several analytical approaches have proposed to integrate the expression quantitative trait loci (eQTLs) data with GWAS, such as transcriptome-wide association study (TWAS) framework and colocalization analysis [20,21], but inferring putative causal variants remains difficult. TWAS framework integrates eQTLs data with GWAS to reveal associations between genes and traits [13]. TWAS studies determine the association of genetically predicted gene expression levels of genes with complex traits; the TWAS framework is unable to provide the effect of causal strength of the variants and horizontal pleiotropy [22]. Colocalisation is also an integrative gene-prioritisation method that integrates eQTLs data with GWAS signals to identify the co-occurrence of variants between pairs of traits [23]. The detected genetic variations (SNPs) are more likely to be functional when GWAS signals colocalise with the eQTL signal [24]. The colocalisation method, however, is not suggested to identify causal associations between pairs of traits.

Summary based Mendelian randomisation (SMR) approach integrates GWAS, expression quantitative trait loci (eQTL), and DNA methylation quantitative trait loci (mQTL) datasets to predict the functional genes which are either pleiotropically (i.e., single genetic variant affecting gene expression and disease phenotype) or potential causally (i.e., single variant affecting gene expression) associated with complex disease [25]. SMR has been used to identify functional genes for many complex diseases and traits, demonstrating SMR as a promising tool for discovering candidate genes linked to complex disorders [25–30].

Multi-omics analyses are well-established to discover functional genes, drug targets, and biomarkers in various complex diseases, including neurological disorders [22,31–34], cancers [35–39], and cardiovascular disease [40]. Multi-omics data integration provides an innovative avenue to bring multi-layer biological information to systematically identify novel insights into the complex pathobiology of diseases [41,42]. However, even with the crucial findings from past GWAS studies, the functional implications and biomarker potentials of the causal genes and loci in stroke risk are largely unknown. Therefore, an approach to systematically integrate post-GWAS analysis with gene expression regulation is desirable for discovering potential biomarkers for stroke risk that may be useful for risk assessment and monitoring. Our bioinformatics approach, which integrates the largest multi-stage multi-omics data integration for stroke, including eQTLs, mQTLs, GWAS, genome-wide gene expression profiling, and functional enrichment analysis, is substantially different than other studies.

Herein, our bioinformatics approach combines SMR method with multi-omics systems biology approaches to identify potential biomarkers, pathways, and therapeutic targets for stroke. We performed an integrative systems biology analysis by combining GWAS summary data for stroke and stroke subtypes with blood *cis*-eQTL and *cis*-mQTL data to identify pleiotropic or potential causal genes associated with stroke risk. We have identified putative functional genes and DNA methylation sites associated with stroke risk. Of those candidate functional genes, ten genes were differentially expressed in stroke versus controls, suggesting potential biomarkers of stroke risk assessment. Functional enrichment analysis suggests gene ontologies and pathways involved in stroke, which enhances our understanding of additional biological routes of stroke.

2. Materials and methods

2.1. Study design

Fig. 1 illustrates the overall analytical approach of this study. To identify pleiotropic or potential causal genes and DNA methylation sites associated with stroke risk, we performed network-based integrative SMR analysis by integrating GWAS summary statistics of stroke from the MEGASTROKE consortium (<https://megastroke.org/>) with *cis*-methylation quantitative trait loci (*cis*-mQTLs) and *cis*-expression quantitative trait loci (*cis*-eQTLs) data. Our study included three distinct SMR tests. Firstly, we utilised *cis*-mQTL data as instrumental variables (IVs), DNA methylation as the exposure, and stroke risk as the outcome to identify the potential pleiotropic or causal association between DNA methylation and stroke risk. Secondly, we used *cis*-eQTL data as the instrumental variables (IVs), gene expression as the exposure, and stroke as the outcome to determine the potential pleiotropic or causal association of gene expression with stroke risk. Thirdly, we used two molecular traits SMR to assess the putative pleiotropic association between DNA methylation and gene expression, where *cis*-mQTL acted as the instrumental variables (IV), DNA methylation (mQTL) as exposure, and gene expression (eQTL) as the outcome. Fig. 2. represents the potential pathways linking genetic variants to stroke risk.

2.2. Data description

Table 1 provides the summary of all the datasets used in this analysis.

2.2.1. Expression quantitative trait loci data

We used the largest whole blood expression quantitative trait loci (eQTL) dataset from the eQTLGen consortium, which includes *cis*-eQTLs data for 19,250 genes expressed in whole blood obtained from 31,684 individuals. The eQTLs data (i.e., SNP-gene pair) were available for ≥ 2 cohorts where the SNP-gene distance of ≤ 1 MB was tested as described previously [43,44]. In addition to this blood eQTL data, we also used another blood eQTL data from the CAGE consortium, which had a total of 2765 participants [45]. The eQTL CAGE dataset used in this study served as the replication of findings obtained from the eQTLGen dataset. The eQTLGen and CAGE eQTL datasets were downloaded on 14 April 2022. The eQTL datasets used in this study are publicly available in the SMR software data resource section (<https://yanglab.westlake.edu.cn/software/smr/#eQTLsummarydata>).

2.2.2. DNA methylation quantitative trait loci data

DNA methylation quantitative loci (mQTL) are genetic variants that affect DNA methylation levels of a particular transcript. We obtained mQTL summary data of blood from a *meta*-analysis by McRae et al of the Brisbane Systems Genetics Study cohorts with 614 individuals [46] and Lothian Birth Cohort with 1366 individuals [47,48]. The mQTL data were limited to DNA methylation probes with ≥ 1 *cis*-mQTL with at least a *cis*-mQTL at $P < 5e-8$ and only SNPs within 2 Mb distance from each probe were available. The mQTL data was downloaded on 14 April 2022 from <https://yanglab.westlake.edu.cn/software/smr/#mQTLsummarydata>.

2.2.3. Stroke GWAS summary statistics data

We downloaded GWAS summary statistics for stroke and stroke subtypes from a *meta*-analysis of GWAS data by Malik et al. [49] released by the MEGASTROKE consortium (<https://megastroke.org/>). The details of the study have been described previously [49]. Briefly, a fixed-effects *meta*-analysis was done for GWAS from only European participants consisting of 40,585 cases and 406,111 controls. Then, a fixed-effects *trans*-ethnic *meta*-analysis was done including all samples (67,162 S cases and 454,450 controls) which consisted of participants of European (40,585 cases; 406,111 controls), East Asian (17,369 cases; 28,195 controls), African (5,541 cases; 15,154 controls), South Asian (2,437 cases; 6,707 controls), mixed Asian (365 cases; 333 controls), and

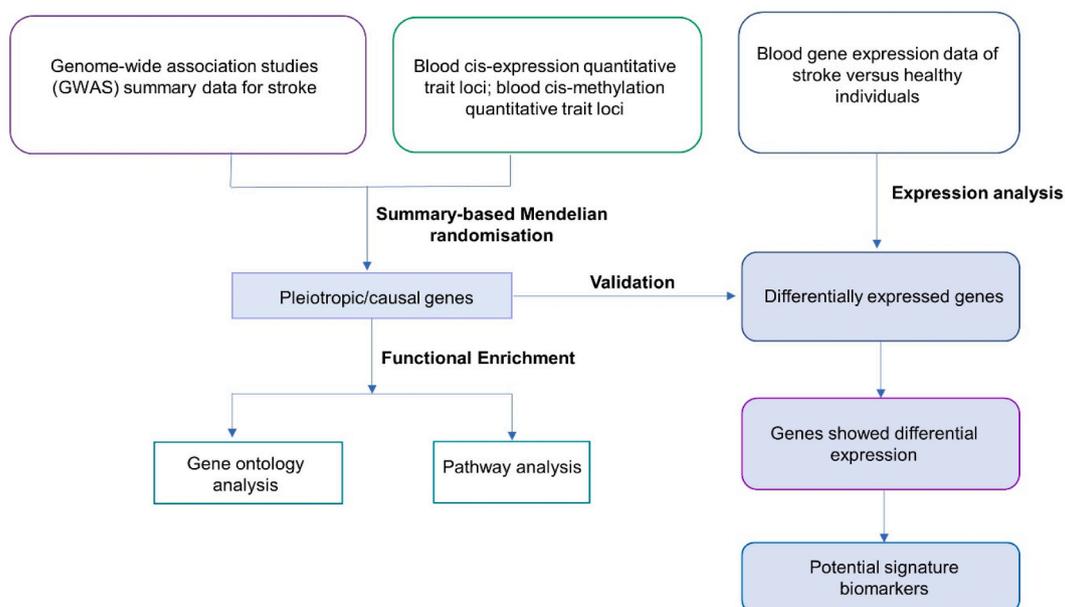


Fig. 1. The present study employed a multi-stage analysis methodology. The integrative summary based Mendelian randomisation (SMR) analysis of GWAS summary data of stroke with blood *cis*-expression quantitative trait loci (*cis*-eQTLs)/*cis*-methylation quantitative trait loci (*cis*-mQTL) data was performed to identify potential causal or pleiotropic gene biomarkers of stroke. Functional enrichment analysis of genes was performed to identify significant gene ontologies and molecular pathways. Differential expression analysis was done on transcriptomic gene expression data of stroke versus healthy individuals. Several SMR-identified genes were differentially expressed in the blood of stroke patients.

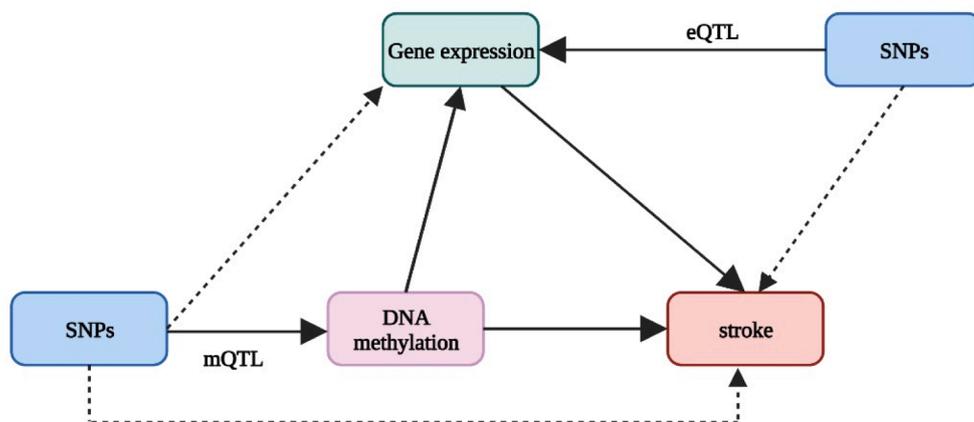


Fig. 2. A schematic illustration showing the pleiotropic or causal associations between genetic variants and stroke risk. Firstly, genetic variants can influence stroke risk mediated by DNA methylation. Secondly, genetic variants can influence stroke risk mediated by gene expression. Thirdly, genetic variants can influence stroke risk mediated through genetic regulation of gene expression by DNA methylation. The DNA methylation quantitative loci (mQTL) and expression quantitative loci (eQTL) show genetic variants linked to DNA methylation and gene expression, respectively. The figure was created with [BioRender.com](https://www.biorender.com).

Table 1
Basic information on the eQTL, mQTL, and stroke GWAS data.

Data source	Total participants or cases/controls	Number of genetic variants or probes
Blood-eQTL(Gen)	31,684	19,250
Blood-eQTL(CAGE)	2,765	36,754
Blood-mQTL	1,980	52,916
Stroke GWAS data (Europeans)	40,585 cases; 406,111 controls	-
Stroke GWAS data (<i>trans</i> -ethnic)	67,162 cases; 454,450 controls	-

GWAS: Genome-wide association studies; eQTL: Expression quantitative trait locus; mQTL: Methylation quantitative trait locus.

Latin American (865 cases; 692 controls) ancestry [49]. The results of any stroke, as well as the subtypes of stroke, were reported in distinct files which consist of any stroke (AS) cases (67,162), any ischemic stroke (AIS) cases (60,341), and large artery stroke (LAS) cases (6,688),

cardioembolic stroke (CES) cases (9,006), and small vessel stroke (SVS) cases (11,710). The datasets were downloaded on 14 April 2022.

2.3. Summary-based Mendelian randomisation and heterogeneity in dependent instruments analysis

We used the SMR method described by Zhu et al. [25] and details of the methods described in the original publication by Zhu and co-workers [25]. Briefly, SMR is a data-driven approach used to determine the associations between genetically regulated traits- gene expression or DNA methylation and outcomes of interest- disease phenotypes using genetic variants such as single nucleotide polymorphisms (SNPs) as instrumental variables (IV) [25]. In principle, SMR uses the SNPs association statistics linked to disease phenotype (outcome) regressed on SNPs association statistics linked to gene expression (exposure) to estimate the effect of an increase in gene expression of a particular transcript on the disease phenotype. So, let *y* be the outcome (e.g., disease phenotype), *x* be the exposure (i.e., gene expression or DNA methylation), and *z* be the

instrumental variable (i.e., SNPs). SMR determines the effect of gene expression or DNA methylation (exposure) on disease phenotype (outcome) as the ratio of the calculated effect of instrumental variables (SNPs) on disease phenotype (b_{zy}) and calculated effect of instrumental variable (SNPs) on gene expression (b_{zx}), which can be expressed as ($b_{xy} = b_{zy}/b_{zx}$) with no non-genetic confounders [25]. SMR predicts the putative functional genes either pleiotropically (i.e., single genetic variant has effect on both gene expression and disease phenotype), potentially causally (i.e., single variant has direct effect on disease phenotype mediated by gene expression) or linkage (two shared genetic variants where one variant has effect on gene expression and another variant has effect on disease phenotypes) [50]. As stated above, an observed association in SMR test could be due to two separate underlying causal variants due to linkage disequilibrium (LD) where one variant has effect on gene expression and another variant has effect on phenotype. This scenario is called linkage. To identify an actual underlying single causal/pleiotropic genetic variant/SNPs from linkage in SMR analysis, we applied a heterogeneity in dependent instruments (HEIDI) test to distinguish the linkage from pleiotropy in the observed association [25]. It was implemented against the null hypothesis that there is a single causal variant deriving the association between exposure (i.e., gene expression/DNA methylation) and outcome (disease phenotype). We excluded the SMR association results when HEIDI test determined the significant heterogeneity of top *cis*-eQTLs in LD with two distinct causal variant ($P_{HEIDI} < 0.01$), suggesting the presence of linkage in the observed associations in SMR analysis [25,51]. Therefore, we used $P_{HEIDI} > 0.01$ as a cut-off to exclude linkage from pleiotropy [51]. The pleiotropic and causal genes are of biological interest for functional characterisation in follow-up studies to develop mechanistic insights of genes in stroke pathogenesis. We did three SMR analyses including (1) SMR analysis with *cis*-eQTL and stroke GWAS to identify association of gene expression change of each gene and stroke risk; (2) SMR analysis with *cis*-mQTL and stroke GWAS to identify association of change of DNA methylation levels and stroke risk; (3) two molecular traits SMR analysis using *cis*-mQTL (exposure) and *cis*-eQTL (outcome) to identify association of change of DNA methylation level with gene expression. Therefore, our SMR analyses yielded pleiotropic associations between DNA methylation, gene expression, and stroke risk. We used the default parameters in SMR analysis (e.g., minor allele frequency [MAF] > 0.01 , $PeQTL < 5 \times 10^{-8}$, eliminating SNPs having low LD or not in LD [$r2 < 0.05$] and high linkage disequilibrium [LD, $r2 > 0.9$] with the top associated eQTL. We adjusted the raw *p*-value yielded in SMR analyses with Benjamini-Hochberg [52] method implemented in the R environment to control the false discovery rate (FDR) for multiple testing error. We employed the SMR method in command line Linux environment (<https://cnsgenomics.com/software/smr/>). To note, we annotated the CpG sites with the closest genes using Illumina HumanMethylation450 BeadChip (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13534>).

2.4. mRNA expression levels of SMR-identified genes leveraging blood transcriptomics of stroke cases versus healthy individuals

We have obtained a microarray gene expression dataset (accession no: GSE58294) containing human peripheral whole blood gene expression of cardioembolic stroke and healthy controls from the NCBI gene expression omnibus (NCBI-GEO) database [53]. GSE58294 was deposited in NCBI-GEO by Stamova et al. [54] [PMID: 25036109]. GSE58294 dataset contained 23 healthy controls and 23 cardioembolic stroke patients using Affymetrix U133 Plus 2.0 microarray technology [54]. The blood samples were taken at three time points following stroke ($n = 23$), at 3 h, 5 h, and 24 h. The first blood was drawn before treatment (≤ 3 h) and consecutive blood samples were drawn following treatment with 5 h and 24 h of post-onset of stroke. Herein, we analysed differential gene expression patterns at all three time points. We used GEO2R online software integrated in NCBI-GEO database to identify

differential expression of genes in stroke cases versus healthy individuals. The differentially expressed genes which passed adjusted *p*-value < 0.05 and absolute log fold change > 0.5 as statistically significant.

2.5. Functional enrichment analysis of the SMR identified genes

We performed gene set enrichment and pathway analysis using bioinformatics tools called ConsensusPath DB [55], accessed on 5 July 2022, to identify significant gene ontology (GO), including biological processes, molecular functions, and potential pathways associated with stroke. ConsensusPathDB is the largest database for pathways, including Wikipathways, Reactome, BioCarta, and the KEGG pathway. The *p*-value obtained from the hypergeometric test was adjusted for the correction of multiple testing errors in the enrichment analysis. GO terms and pathways that passed the adjusted *p*-value < 0.05 were considered significant.

3. Results

3.1. Expression of pleiotropic genes is modulated by stroke risk loci in Europeans

To test the association of genetically determined gene expression and stroke risk in the European population, we used summary based Mendelian randomisation (SMR) analysis, which integrated stroke GWAS summary statistics data with the largest blood eQTL summary data from eQTLGen consortium data (eQTLGen hereafter) and another blood eQTL data from CAGE consortium (CAGE eQTL hereafter), to identify pleiotropic or potentially causal genes associated with stroke. SMR and heterogeneity in dependent instruments (HEIDI) tests demonstrated pleiotropic association of 8 genes with stroke (any stroke; AS), 7 genes with ischemic stroke (any ischemic stroke; AIS), one gene for small vessel stroke (SVS) (Table 2), which passed the SMR ($FDR < 0.05$) and HEIDI test ($P_{HEIDI} > 0.01$). Notably, we have also obtained three genes (*PMF1*, *RERE*, and *OBFC1*) in SMR analysis with CAGE eQTL, where *PMF1* and *RERE* were found in SMR analysis using eQTLGen (Table 2). The low number of genes were identified in CAGE eQTL due to relatively smaller size data than eQTLGen.

3.2. Expression of SMR identified pleiotropic genes is modulated by stroke risk loci in trans-ancestral participants

We used SMR to identify potential genetically determined blood gene expression associated with stroke in *trans*-ethnic populations and identified several genes with significant pleiotropic or potential causal links with stroke after multiple testing corrections ($FDR < 0.05$) and the HEIDI test ($P_{HEIDI} > 0.01$). Specifically, we found that the expression of 9 genes was pleiotropic or causal for stroke (AS), 5 genes with AIS, and 5 genes with SVS (Table 3). Next, in SMR analysis with CAGE eQTL datasets, we found two genes (2 mRNA probes) associated with the stroke while two genes showed potential pleiotropic/potential causal association with AIS. Notably, rs2758600 and rs887953 SNP tagging genes *PMF1* showed significant association with stroke in both SMR analyses using eQTLGen and CAGE eQTLs (Table 3).

3.3. Genetically determined expression of common genes overlaps between Europeans and transethnic ancestry

We compared the SMR-identified genes in European and transethnic ancestries to find overlap genes. In SMR with eQTLGen, 4 genes (*PMF1*, *SH3PXD2A*, *TSPOAP1*, and *PHETA1*) associated with any stroke were overlapped between European and *trans*-ethnic populations. In addition, and 2 genes (*PMF1*, *PHETA1*) associated with ischemic stroke were overlapped between European and *trans*-ethnic populations. In SMR with CAGE eQTL, we found *PMF1* overlapped in both European and

Table 2

Pleiotropic association between gene expression and stroke risk using blood *cis*-eQTLs and GWAS of Europeans.

eQTL data	Type of stroke	probeID	Gene	topSNP	CHR	p_GWAS	p_eQTL	b_SMR	se_SMR	p_SMR	p_HEIDI	p_SMR_adj
eQTLGen	AS	ENSG00000160783	PMF1	rs2758600	1	2.50E-11	1.32E-38	5.39E-01	9.11E-02	3.23E-09	2.51E-01	5.07E-05
		ENSG00000142599	<i>RERE</i>	rs301802	1	1.08E-05	0	9.31E-02	2.12E-02	1.18E-05	6.23E-01	3.09E-02
		ENSG00000175164	<i>ABO</i>	rs550057	9	1.54E-05	0	-7.36E-02	1.71E-02	1.72E-05	1.33E-01	3.67E-02
		ENSG00000107957	SH3PXD2A	rs10786772	10	5.70E-07	1.11E-186	2.03E-01	4.12E-02	8.74E-07	3.27E-01	6.85E-03
		ENSG00000198324	PHETA1	rs11065884	12	1.55E-06	9.12E-60	-3.48E-01	7.52E-02	3.83E-06	4.20E-01	1.50E-02
		ENSG00000005379	TSPOAP1	rs12948345	17	2.92E-07	1.14E-39	4.27E-01	8.94E-02	1.83E-06	1.76E-01	9.56E-03
		ENSG00000108375	<i>RNF43</i>	rs12948345	17	2.92E-07	6.08E-18	6.53E-01	1.48E-01	1.07E-05	2.66E-02	3.09E-02
		ENSG00000005381	<i>MPO</i>	rs12940119	17	1.72E-05	1.27E-295	1.30E-01	3.03E-02	1.87E-05	9.07E-02	3.67E-02
	AIS	ENSG00000160783	PMF1	rs2758600	1	4.34E-08	1.32E-38	4.80E-01	9.52E-02	4.68E-07	4.69E-01	3.67E-03
		ENSG00000198324	PHETA1	rs11065884	12	2.79E-07	9.12E-60	-4.05E-01	8.25E-02	9.41E-07	4.37E-01	4.92E-03
		ENSG00000111252	<i>SH2B3</i>	rs587914	12	2.96E-06	5.82E-126	2.53E-01	5.51E-02	4.48E-06	3.51E-02	1.76E-02
		ENSG00000198270	<i>TMEM116</i>	rs4767068	12	9.09E-06	0	-1.20E-01	2.72E-02	1.06E-05	1.60E-01	2.77E-02
		ENSG00000111271	<i>ACAD10</i>	rs10774634	12	2.50E-06	1.05E-23	-5.87E-01	1.38E-01	2.11E-05	4.45E-02	3.69E-02
		ENSG00000182511	<i>FPS</i>	rs8027450	15	3.25E-05	0	-1.17E-01	2.83E-02	3.34E-05	1.80E-01	4.98E-02
		ENSG00000005379	TSPOAP1	rs12948345	17	1.28E-05	1.14E-39	3.94E-01	9.52E-02	3.49E-05	2.10E-01	4.98E-02
		SVS	ENSG00000043591	<i>ADRB1</i>	rs4917675	10	1.82E-06	4.82E-137	5.00E-01	1.07E-01	2.70E-06	1.97E-02
CAGE eQTL	AS	ILMN_1756445	<i>PMF1</i>	rs887953	1	2.26E-11	5.02E-10	3.71E-01	8.15E-02	5.18E-06	2.68E-01	2.21E-02
		ILMN_1802380	<i>RERE</i>	rs301802	1	1.08E-05	8.58E-116	6.43E-02	1.49E-02	1.60E-05	4.59E-01	3.26E-02
		ILMN_2327795	<i>RERE</i>	rs301802	1	1.08E-05	8.50E-77	7.95E-02	1.86E-02	1.91E-05	1.26E-01	3.26E-02
		ILMN_1789186	<i>OBFC1</i>	rs6584579	10	1.34E-07	8.29E-24	1.78E-01	3.81E-02	3.19E-06	3.48E-01	2.21E-02

eQTL: Expression quantitative trait loci; SNP: Single nucleotide polymorphism; CHR: Chromosome; p_GWAS: Genome wide association study *P*-value; p_mQTL: DNA methylation quantitative trait loci *P*-value; p_eQTL: Expression quantitative trait loci *P*-value; P_SMR: *P*-value Summary based Mendelian randomisation; P_HEIDI: Heterogeneity in dependent instruments *P*-value; LogFC: Log fold change; DEG: Differentially expressed genes; p_SMR_adj: Adjusted *P*-value from summary based Mendelian randomisation. The genes overlapped between European and *trans*-ancestral participants were bolded.

trans-ethnic ancestry (Supplementary Table 1).

3.4. Methylation at CpG sites is modulated by stroke risk loci in European population

To identify CpG methylation sites associated with stroke in European populations, we performed SMR analysis of GWAS summary data for stroke and blood *cis*-DNA methylation quantitative trait loci (*cis*-mQTL) data using McRae et al. mQTL summary data [48]. We interrogated the relationship between DNA methylation (i.e., CpG probes) and five different types of strokes. Our analysis identified pleiotropic or potential associations of 22 CpGs with any stroke (AS), 8 CpGs with any ischemic stroke (AIS), and 1 CpG with cardioembolic stroke (CES) ($FDR < 0.05$ and $P_{HEIDI} > 0.01$) (Supplementary Table 2). We did not find any significant associations of CpGs with large artery stroke (LAS) and small vessel stroke (SVS) after correction of multiple testing and HEIDI test.

3.5. Methylation at CpG sites is modulated by stroke risk loci in *trans*-ethnic participants

To identify CpG methylation sites associated with stroke in *trans*-ethnic populations, we conducted SMR analysis of *trans*-ethnic GWAS summary data of stroke with blood *cis*-mQTL data as mentioned above.

We assessed the association between CpG probes with five types of stroke. We detected 47 CpG methylation sites that showed pleiotropic or potential causal association with AS (Supplementary Table 3). In addition, we found that 41 CpG methylation sites and 4 CpG methylation sites had significant associations with AIS and CES, respectively, after correction for multiple testing and HEIDI test ($FDR < 0.05$ and $P_{HEIDI} > 0.01$). We did not find any significant CpG sites in LAS and SVS after the correction of multiple tests and HEIDI test (Supplementary Table 3).

3.6. Multi-omics SMR to prioritise genes mediated by DNA methylation regulating gene expression in stroke

To identify potential associations between DNA methylation and gene expression, we performed two molecular traits SMR analyses by integrating *cis*-mQTL data with *cis*-eQTL data. We identified 46,698 associations between 59,975 CpGs and 13,171 mRNA probes after multiple testing corrections ($FDR < 0.05$) and the HEIDI test ($P_{HEIDI} > 0.01$) (Supplementary Table 4). To determine the association between DNA methylation sites that regulate gene expression, we performed an overlapping analysis of CpG methylation sites identified from SMR analysis of mQTL and stroke with the CpG methylation sites identified from SMR analysis of mQTL and eQTL. We found 16 DNA methylated probes regulating the expression of 8 genes (*CAZ1*, *ACAD10*, *SLC44A2*,

Table 3

Pleiotropic association between gene expression and stroke risk using blood *cis*-eQTLs and GWAS of transethnic populations.

eQTL data	Type of stroke	probeID	Gene	topSNP	CHR	p_GWAS	p_eQTL	b_SMR	se_SMR	p_SMR	p_HEIDI	p_SMR_adj	
eQTLGen	AS	ENSG00000160783	PMF1	rs2758600	1	9.62E-12	1.32E-38	0.489799	0.081165	1.59E-09	0.389283	1.24E-05	
		ENSG00000146215	<i>CRIP3</i>	rs2242416	6	1.25E-06	6.77E-107	0.222992	0.047372	2.51E-06	0.510285	1.31E-02	
		ENSG00000112658	<i>SRF</i>	rs4398731	6	1.72E-08	2.05E-10	-0.84693	0.200524	2.40E-05	0.159261	4.00E-02	
		ENSG00000119408	<i>NEK6</i>	rs72759285	9	3.34E-05	0	-0.06601	0.015854	3.13E-05	0.367791	4.00E-02	
		ENSG00000107957	SH3PXD2A	rs10786772	10	7.10E-10	1.11E-186	0.214956	0.035552	1.48E-09	0.345716	1.24E-05	
		ENSG00000107960	<i>STN1</i>	rs11191865	10	1.16E-05	9.82E-70	0.165804	0.038857	1.98E-05	0.052539	4.00E-02	
		ENSG00000187486	<i>KCNJ11</i>	rs2074310	11	8.78E-06	4.37E-38	0.239204	0.057165	2.86E-05	0.951468	4.00E-02	
		ENSG00000198324	PHETA1	rs11065884	12	1.34E-05	9.12E-60	-0.26632	0.063152	2.48E-05	0.138374	4.00E-02	
		ENSG00000005379	TSPOAPI	rs12948345	17	8.46E-07	1.14E-39	0.399789	0.086708	4.01E-06	0.031304	1.57E-02	
		AIS	ENSG00000160783	PMF1	rs2758600	1	9.29E-09	1.32E-38	0.438813	0.083221	1.34E-07	0.600982	2.10E-03
	ENSG00000107957		<i>SH3PXD2A</i>	rs10786772	10	3.27E-07	1.11E-186	0.188558	0.037437	4.74E-07	0.396084	3.70E-03	
	ENSG00000187486		<i>KCNJ11</i>	rs2074310	11	9.40E-06	4.37E-38	0.249383	0.059227	2.55E-05	0.931033	3.98E-02	
	ENSG00000198324		PHETA1	rs11065884	12	5.82E-06	9.12E-60	-0.29452	0.067404	1.25E-05	0.080327	2.78E-02	
	SVS		ENSG00000156735	<i>BAG4</i>	rs11779986	8	5.12E-06	1.24E-177	-0.27591	0.061264	6.68E-06	0.053734	3.24E-02
		ENSG00000147548	<i>NSD3</i>	rs2234555	8	7.51E-06	3.51E-84	-0.44622	0.102186	1.26E-05	0.929188	3.24E-02	
		ENSG00000107957	<i>SH3PXD2A</i>	rs10786772	10	8.01E-06	1.11E-186	0.327253	0.074183	1.03E-05	0.421115	3.24E-02	
		ENSG00000120798	<i>NR2C1</i>	rs35583	12	1.59E-06	3.21E-82	0.550962	0.118234	3.16E-06	0.424414	3.24E-02	
		ENSG00000180263	<i>FGD6</i>	rs10128863	12	5.47E-06	3.27E-209	-0.22033	0.049042	7.03E-06	0.060097	3.24E-02	
	CAGE eQTL	AS	ILMN_1785175	<i>SWAP70</i>	rs173396	11	9.51E-07	1.84E-69	-0.08469	0.01796	2.41E-06	0.155065	2.05E-02
			ILMN_1756445	<i>PMF1</i>	rs887953	1	1.45E-11	5.02E-10	0.32905	0.071773	4.55E-06	0.224843	9.69E-03
AIS		ILMN_1785175	<i>SWAP70</i>	rs173396	11	6.38E-06	1.84E-69	-0.07309	0.016807	1.37E-05	0.395793	2.34E-02	
		ILMN_1792305	<i>ZNF318</i>	rs1563788	6	4.68E-06	3.75E-52	-0.0909	0.020637	1.06E-05	0.032915	4.51E-02	

eQTL: Expression quantitative trait loci; SNP: Single nucleotide polymorphism; CHR: Chromosome; p_GWAS: Genome wide association study *P-value*; p_mQTL: DNA methylation quantitative trait loci *P-value*; p_eQTL: Expression quantitative trait loci *P-value*; P_SMR: *P-value* from summary based Mendelian randomisation; P_HEIDI: Heterogeneity in dependent instruments *P-value*; LogFC: Log fold change; DEG: Differentially expressed genes; p_SMR_adj: Adjusted *P-value* from summary based Mendelian randomisation. The genes overlapped between European and *trans*-ancestral participants were bolded.

SH3PXD2A, *PMF1*, *SUPT4H1*, *ABO*, *SLC25A44*) in any stroke (AS), and 5 DNA methylated probes regulating the expression of 4 genes (*ACAD10*, *SLC44A2*, *PMF1*, *SLC25A44*) in any ischemic stroke (AIS) in Europeans (SNP → DNA methylation → Gene expression → Stroke) (Supplementary Table 5).

Next, we also determined the association of DNA methylation regulating gene expression in *trans*-ethnic ancestry. Similarly, we performed the overlap analysis of CpG methylated sites identified from SMR analysis of mQTL and stroke with the CpG methylated sites identified from SMR analysis of mQTL and eQTL. We identified 34 DNA methylated probes regulating the expression of 16 genes (*CAZ1*, *SLC38A3*, *SLC44A2*, *ILF3*, *ACAD10*, *CDKN1A*, *SH3PXD2A*, *PMF1*, *QTRT1*, *SUPT4H1*, *HTRA1*, *SLC25A44*, *KCNJ11*, *C9orf84*, *FBN2*, *TTBK1*) in AS and 27 DNA methylated probe regulating the expression of 12 genes (*CAZ1*, *SLC44A2*, *ILF3*, *ACAD10*, *DUS3L*, *SH3PXD2A*, *PMF1*, *QTRT1*, *SCARF1*, *SLC25A44*, *KCNJ11*, *FAM109A*) in AIS in *trans*-ethnic ancestry (SNP → DNA methylation → Gene expression → Stroke) (Supplementary Table 6).

3.7. Prioritisation of genes associated with stroke risk

Using blood-eQTL and GWAS summary data of stroke, we performed SMR analysis to identify potential pleiotropic or causal associations between gene expression and stroke. Our study found 14 unique gene from Europeans (Table 2) and 15 unique genes from *trans*-ethnic ancestry (Table 3). Similarly, using blood-mQTL and GWAS summary data of stroke, we also performed SMR analysis to identify pleiotropic or causal association between DNA methylation and stroke. We found total 23 unique CpGs that code for 11 unique genes in European ancestry (Supplementary Table 2) and 58 unique CpGs that code for 24 unique genes in *trans*-ethnic ancestry (Supplementary Table 3). Considering both gene expression (SMR analysis with eQTL and stroke) and DNA methylation (SMR analysis with mQTL and stroke), we found total 45 unique genes that were pleiotropically or causally associated with stroke risk (Table 4 and Table 5).

Table 4
Total unique genes prioritised using summary based Mendelian randomisation.

Category	European ancestry unique gene	Trans-ethnic ancestry unique gene	Total unique gene between European and trans-ethnic ancestry
eQTL prioritised gene	14	15	26
mQTL prioritised gene	11	24	25
Total unique gene (eQTL + mQTL) between European and trans-ethnic ancestry			45

eQTL: Expression quantitative trait loci; mQTL: Methylation quantitative trait loci.

3.8. mRNA expression of SMR prioritised gene in the blood of stroke cases versus controls

We directly examined the expression of SMR identified genes in genome-wide blood transcriptomics, leveraging GSE58294 data of stroke versus controls. we identified ten genes (*HTRA1*, *PMF1*, *FBN2*,

C9orf84, *COL4A1*, *BAG4*, *NEK6*, *SH2B3*, *SH3PXD2A*, *ACAD10*) were differentially expressed between healthy versus stroke individuals at (≤ 3 h) time point. Of those genes, seven genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *COL4A1*, *BAG4*, *NEK6*,) were up regulated and *SH2B3*, *SH3PXD2A*, and *ACAD10* were downregulated in stroke versus controls ($FDR < 0.05$; absolute log fold change > 0.50) (Table 6) and Fig. 3. Considering gene expression at 5 h time point, we found nine genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *BAG4*, *SUPT4H1*, *RERE*, *SH3PXD2A*, *ACAD10*) were differentially expressed in stroke where seven genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *BAG4*, *SUPT4H1*, *RERE*) were up regulated and two genes (*SH3PXD2A*, *ACAD10*) were down regulated (Table 6). To note, we found seven common genes were differentially expressed in stroke in two time-points (i.e., ≤ 3 h and 5 h, respectively). Similarly, we also considered gene expression in 24 h time point and found ten genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *COL4A1*, *BAG4*, *NEK6*, *RERE*, *SH3PXD2A*, *ACAD10*) were differentially expressed in stroke where eight genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *COL4A1*, *BAG4*, *NEK6*, *RERE*) were up regulated and two genes (*SH3PXD2A*, *ACAD10*) were down regulated (Table 6). To note, of those ten genes identified at 24 h time point, we found nine common genes were also differentially expressed in stroke at ≤ 3 h time point.

Table 5
Summary statistics of multi-omics prioritised genes associated with stroke risk.

Gene	topSNP	CHR	A1	A2	Freq	b_GWAS	p_GWAS	p_SMR	p_HEIDI	p_SMR_adj
<i>PMF1</i>	rs2758600	1	T	C	0.343936	-0.0574	4.34E-08	4.68E-07	0.469249	0.003672
<i>PHETA1</i>	rs11065884	12	G	A	0.224652	-0.0617	2.79E-07	9.41E-07	0.437127	0.004919
<i>SH2B3</i>	rs587914	12	C	T	0.202783	-0.0575	2.96E-06	4.48E-06	0.035064	0.017577
<i>TMEM116</i>	rs4767068	12	G	A	0.167992	-0.0575	9.09E-06	1.06E-05	0.15958	0.02767
<i>ACAD10</i>	rs10774634	12	G	A	0.16998	-0.0615	2.50E-06	2.11E-05	0.044532	0.036853
<i>FPS</i>	rs8027450	15	T	C	0.323062	0.0454	3.25E-05	3.34E-05	0.179592	0.049774
<i>TSPOAP1</i>	rs12948345	17	T	C	0.213718	0.0558	1.28E-05	3.49E-05	0.210455	0.049774
<i>RERE</i>	rs301802	1	T	A	0.471173	-0.0413	1.08E-05	1.18E-05	0.623319	0.030876
<i>ABO</i>	rs550057	9	T	C	0.282306	0.0484	1.54E-05	1.72E-05	0.133303	0.036698
<i>SH3PXD2A</i>	rs10786772	10	A	G	0.309145	-0.0484	5.70E-07	8.74E-07	0.3274	0.006852
<i>RNF43</i>	rs12948345	17	T	C	0.213718	0.0604	2.92E-07	1.07E-05	0.026604	0.030876
<i>MPO</i>	rs12940119	17	C	T	0.172962	0.053	1.72E-05	1.87E-05	0.090682	0.036698
<i>ADRB1</i>	rs4917675	10	C	T	0.241551	0.1247	1.82E-06	2.70E-06	0.019732	0.042355
<i>OBFC1</i>	rs6584579	10	G	A	0.362823	-0.0494	1.34E-07	3.19E-06	0.348092	0.022136
<i>CRIP3</i>	rs2242416	6	A	G	0.423459	0.04	1.25E-06	2.51E-06	0.510285	0.013072
<i>SRF</i>	rs4398731	6	A	G	0.329026	0.0469	1.72E-08	2.40E-05	0.159261	0.039991
<i>NEK6</i>	rs72759285	9	G	A	0.187873	0.0413	3.34E-05	3.13E-05	0.367791	0.039991
<i>STN1</i>	rs11191865	10	A	G	0.537773	0.0343	1.16E-05	1.98E-05	0.052539	0.039991
<i>BAG4</i>	rs11779986	8	G	A	0.258449	0.0894	5.12E-06	6.68E-06	0.053734	0.03241
<i>NSD3</i>	rs2234555	8	A	C	0.242545	0.0838	7.51E-06	1.26E-05	0.929188	0.03241
<i>NR2C1</i>	rs35583	12	G	A	0.147117	0.1124	1.59E-06	3.16E-06	0.424414	0.03241
<i>FGD6</i>	rs10128863	12	C	T	0.148111	0.1049	5.47E-06	7.03E-06	0.060097	0.03241
<i>ZNF318</i>	rs1563788	6	T	C	0.333002	0.0405	4.68E-06	1.06E-05	0.032915	0.045126
<i>SWAP70</i>	rs173396	11	G	A	0.405567	-0.0416	9.51E-07	2.41E-06	0.155065	0.020541
<i>SLC25A44</i>	rs2842857	1	C	T	0.351889	-0.057	8.72E-08	1.71E-07	0.341619	0.005076
<i>CASZ1</i>	rs880315	1	C	T	0.356859	0.0518	9.99E-07	2.01E-06	0.789051	0.023493
<i>LRCH1</i>	rs912426	13	C	T	0.183897	-0.0629	6.22E-07	1.01E-06	0.028712	0.015497
<i>HTRA1</i>	rs3793917	10	G	C	0.195825	-0.0531	2.06E-06	3.80E-06	0.058166	0.019935
<i>SUPT4H1</i>	rs2526375	17	G	A	0.213718	0.0597	3.62E-07	9.37E-06	0.213157	0.039818
<i>SLC44A2</i>	rs2060236	19	G	C	0.399602	-0.0465	1.54E-06	2.15E-06	0.631647	0.013684
<i>TMEM51</i>	rs10927727	1	T	C	0.161034	-0.0491	6.47E-06	7.18E-06	0.357166	0.023023
<i>CDKN1A</i>	rs762624	6	C	A	0.237575	-0.0483	2.36E-07	8.60E-07	0.054201	0.006601
<i>ARMS2</i>	rs11200630	10	C	T	0.195825	-0.0453	5.35E-06	1.87E-05	0.301292	0.044843
<i>KCNJ11</i>	rs5219	11	T	C	0.352883	0.0416	1.65E-05	2.29E-05	0.872022	0.049515
<i>FAM109A</i>	rs7398833	12	C	T	0.246521	-0.0431	1.78E-05	1.72E-05	0.030243	0.044531
<i>SCARF1</i>	rs28749238	17	G	A	0.069583	0.0808	6.25E-07	1.38E-05	0.13406	0.036701
<i>ILF3</i>	rs4435370	19	G	A	0.370775	-0.0445	9.84E-07	1.90E-06	0.521175	0.009819
<i>DUS3L</i>	rs45459297	19	A	G	0.019881	0.1283	3.13E-06	6.39E-06	0.033327	0.021201
<i>QTRT1</i>	rs2229383	19	G	T	0.400596	-0.0476	4.72E-08	1.24E-05	0.085246	0.033934
<i>SLC38A3</i>	rs4688745	3	C	T	0.441352	0.0341	1.83E-05	2.56E-05	0.046776	0.044901
<i>FBN2</i>	rs73350117	5	C	A	0.171968	0.056	5.77E-07	1.09E-05	0.717204	0.025797
<i>TTBK1</i>	rs6930689	6	C	T	0.33002	0.0455	5.68E-08	8.58E-06	0.517744	0.022191
<i>C9orf84</i>	rs823643	9	C	A	0.439364	0.0378	6.58E-06	1.90E-05	0.017133	0.034724
<i>COL4A1</i>	rs2131938	13	T	C	0.354871	0.0386	1.85E-06	1.11E-05	0.287121	0.025797
<i>PITX2</i>	rs2739206	4	G	C	0.302187	0.1203	1.76E-10	1.81E-06	0.445129	0.028187

SNP: Single nucleotide polymorphism; CHR: Chromosome; A1: Effect allele; A2: Other allele; Freq: Frequency of the effect allele; b_GWAS: effect size from GWAS; p_GWAS: Genome wide association study *P-value*; p_SMR: *P-value* from summary based Mendelian randomisation using eQTL/mQTL; p_HEIDI: *P-value* from Heterogeneity in dependent instruments test; p_SMR_adj: Adjusted *P-value* of summary based Mendelian randomisation.

Table 6

Ten genes from SMR study were differentially expressed in stroke cases and control subjects.

Gene				SMR Study				DEGs(3 h)		DEGs(5 h)		DEGs(24 h)	
	topSNP	CpG	CHR	P_GWAS	P_mQTL /P_eQTL	P_SMR	P_HEIDI	Log2FC	FDR	Log2FC	FDR	Log2FC	FDR
<i>PMF1</i>	rs3768276	cg25465065	1	1.36E-07	0	1.37E-07	0.014	1.038	9.23E-10	1.084	4.78E-12	0.926	5.08E-10
<i>FBN2</i>	rs73350117	cg23213887	5	5.77E-07	2.11E-20	1.09E-05	0.717	1.028	1.23E-05	1.122	2.05E-06	1.384	1.56E-07
<i>C9orf84</i>	rs823643	cg01822570	9	6.58E-06	7.03E-43	1.90E-05	0.017	0.925	4.34E-03	1.215	3.62E-06	1.630	1.80E-06
<i>COL4A1</i>	rs2131938	cg13424422	13	1.85E-06	7.66E-30	1.11E-05	0.287	0.594	7.29E-03	–	–	0.583	2.38E-02
<i>NEK6</i>	rs72759285	–	9	3.34E-05	0	3.13E-05	0.367	0.550	2.47E-04	–	–	0.521	1.63E-04
<i>HTRA1</i>	rs3793917	cg25446361	10	1.25E-06	8.39E-96	9.04E-03	0.383	1.355	1.39E-04	1.511	1.15E-05	1.930	5.50E-07
<i>BAG4</i>	rs11779986	–	8	5.12E-06	1.2E-177	6.68E-06	0.053	0.590	4.22E-03	0.713	2.20E-04	0.631	1.63E-03
<i>SH2B3</i>	rs587914	–	12	2.96E-06	5.8E-126	4.48E-06	0.035	–0.510	2.13E-05	–	–	–	–
<i>SH3PXD2A</i>	rs11191833	cg01727419	10	1.66E-10	0	1.77E-10	0.396	–0.588	3.89E-04	–0.710	1.04E-06	–0.639	6.29E-06
<i>ACAD10</i>	rs642898	cg08577424	12	1.37E-07	3.3E-187	2.17E-07	0.039	–0.837	5.76E-05	–0.502	1.61E-02	–0.524	6.20E-03
<i>SUPT4H1</i>	rs2526375	cg14996805	17	6.68E-07	3.99E-19	1.42E-05	0.026	–	–	0.553	1.2E-04	–	–
<i>RERE</i>	rs301802	–	1	1.08E-05	0	3.23E-09	0.251	–	–	0.679	3.59E-04	0.644	7.52E-04

SNP: Single nucleotide polymorphism; CHR: Chromosome; p_GWAS: Genome wide association study *P-value*; p_mQTL: DNA methylation quantitative trait loci *P-value*; p_eQTL: Expression quantitative trait loci *P-value*; P_SMR: *P-value* from summary based Mendelian randomisation; P_HEIDI: *P-value* from Heterogeneity in dependent instruments test; DEG: Differentially expressed genes; LogFC: Log fold change; 3 h, 5 h, 24 h: Blood samples taken at three time points 3 h(3 h), 5 h(5 h), and 24 h(24 h) following stroke; FDR: False discovery control rate.

3.9. Functional enrichment

To better understand the potential biological roles of the identified candidate genes in this study, we performed functional enrichment analysis of SMR identified genes. Our functional enrichment analysis of the SMR prioritised genes revealed some significant gene ontology and pathways. They were enriched in some gene ontology terms such as “response to extracellular stimulus”, “cell aging”, “metal ion binding” (Fig. 4A and Supplementary Table 7) and pathways, particularly “oxidative damage”, “extracellular matrix organisation”, “transcriptional regulation by the AP-2 (TFAP2) family of transcription factors”, “heart development”, and “coregulation of androgen receptor activity” (Fig. 4B and Supplementary Table 8).

4. Discussion

In this present study, we comprehensively integrated multi-omics data, particularly GWAS of stroke, eQTLs, and mQTLs using SMR to identify gene biomarkers of stroke risk. Our network SMR analyses aimed to identify the pleiotropic association of genetically regulated traits- DNA methylation and gene expression with the phenotype of interest. We identified a total of 45 pleiotropic or potential causal gene biomarkers associated with stroke risk. Of those 45 genes, 10 genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *COL4A1*, *BAG4*, *NEK6*, *SH2B3*, *SH3PXD2A*, *ACAD10*) were differentially expressed in the blood of stroke patients versus controls, suggesting potential biomarkers of stroke risk. Despite many efforts to identify stroke risk loci, the putative functional gene biomarkers of stroke risk remain unknown. Past studies attempted to explore the risk genes associated with stroke, but the dearth of systems biological approaches undermined the potential to uncover functional risk biomarkers. Therefore, we conducted the largest study to date to systematically search for blood gene biomarkers for stroke risk and validate them in the blood transcriptomics of stroke patients compared to controls. Of those identified putative gene biomarkers, *PMF1*, *SH3PXD2A*, *SH2B3*, *COL4A1*, and *HTRA1* were previously known as stroke risk genes according to GWAS catalogues, while the causal associations of *FBN2*, *ACAD10*, *C9orf84*, *BAG4*, and *NEK6* were not previously known, suggesting novel gene biomarkers of stroke. The identified genes have been implicated in stroke pathogenesis. *PMF1* is associated with intracerebral haemorrhagic stroke [56,57]. *PMF1* encodes polyamine-modulated factor 1 that plays a significant role in chromosomal alignment and kinetochores formation during mitosis [57]. A previous study found genetic variants of the *SH2B3* gene were significantly associated with the pathogenesis of ischemic stroke, large artery stroke, and coronary artery disease [58]. Past studies reported

that mutations in *COL4A1* gene were linked to a wide range of disorders, including haemorrhagic stroke, cerebral small vessel disease, myopathy, glaucoma, and among others [59,60]. Mutation in *HTRA1* is responsible for cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL) and cerebral small vessel disease [61]. CARASIL is a hereditary condition in which small blood vessels in the brain are damaged, resulting in stroke and other disabilities [62]. Moreover, the *HTRA1* gene is also associated with lacunar stroke [63]. *FBN2* gene encodes fibrillin-2 protein, which is involved in tissue elasticity, providing connective tissues with the strength and flexibility to maintain joints and organs in the body [64]. *FBN2* is strongly associated with age-related macular degeneration (AMD) [65], and AMD increases the risk of stroke by 2-fold more than patients without AMD [66]. *SH3PXD2A* (SH3 And PX Domains 2A) is a protein coding gene and was previously associated with ischemic stroke and lacunar stroke [63]. *ACAD10* encodes a component of acyl-CoA dehydrogenase enzymes family, which participate beta-oxidation of fatty acids. The variant of *ACAD10* gene was previously been found associated with ischemic stroke [67], coronary artery disease [68], type 2 diabetes, insulin resistance, and lipid oxidation [69]. The role of *C9orf84* in stroke risk is still unknown. *BAG4* gene encodes for BAG cochaperone 4 proteins which is involved in different biological processes such as cell proliferation, survival, and apoptosis processes. Overexpression of *BAG4* protein was observed in different types of cancers [70]. However, the role of *BAG4* gene in stroke risk is still elusive. The risk gene *NEK6* encodes protein kinase which plays a role in apoptosis, cell cycle regulation, chemoresistance, and telomere maintenance [71]. *NEK6* genes promote cell proliferation and develop head neck squamous cell carcinoma [72]. In addition, our study also highlighted that among the ten differentially expressed genes, three genes (*PMF1*, *SH3PXD2A*, *ACAD10*) were modulated by DNA methylation and gene expression in Europeans (SNP → DNA methylation → Gene expression → Stroke) and six genes (*HTRA1*, *PMF1*, *FBN2*, *C9orf84*, *SH3PXD2A*, *ACAD10*) were modulated by DNA methylation and gene expression in *trans*-ethnic ancestry group (SNP → DNA methylation → Gene expression → Stroke). However, we demonstrated their causal and functional insights as they could be modulated through DNA methylations, which could influence gene expression.

Our study demonstrated the pleiotropic association between gene expression and stroke. Our analysis revealed four genes (*PMF1*, *SH3PXD2A*, *TPSPOAPI*, *PHETA1*) were overlapped between Europeans and *trans*-ethnic ancestry. The *trans*-ethnic ancestral GWAS meta-analysis consisted of substantial European participants (~60%) than other ancestries. Next, pleiotropic association between DNA methylation and stroke revealed ten genes (*ARMS2*, *ACAD10*, *CASZ1*, *SH3PXD2A*,

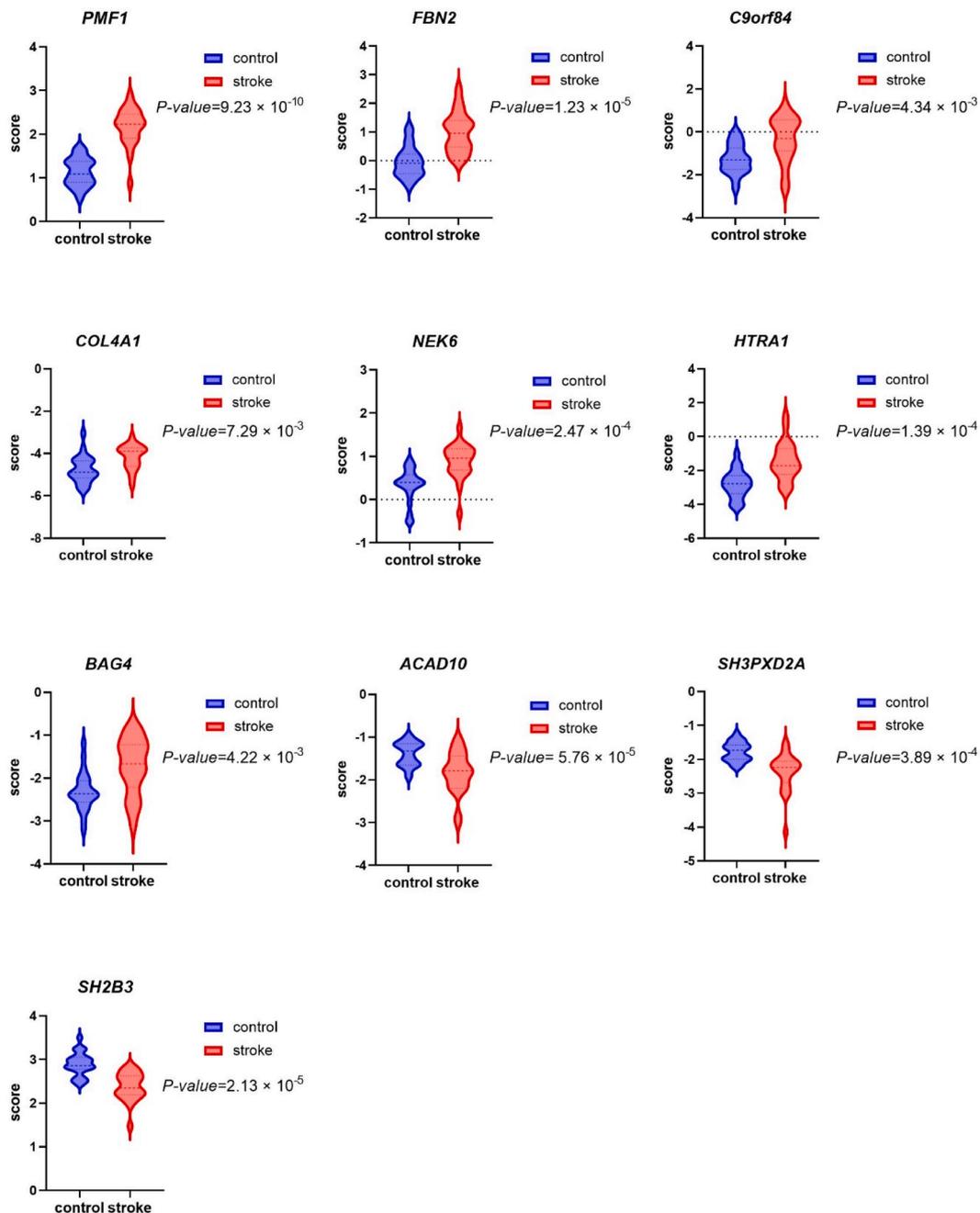


Fig. 3. The violin plot represents the significant differential expression of the genes in blood of stroke compared to healthy individuals at ≤ 3 h.

HTRA1, *PMF1*, *SUPT4H1*, *SLC25A44*, *LRCH1*, *SLC44A2*) were overlapped between European ancestry and *trans*-ethnic ancestry. To note, we identified *PMF1*, *HTRA1*, *ACAD10* and *SH3PXD2A* were differentially expressed in the blood of stroke versus healthy individuals, suggesting potential common biomarkers of stroke risk in both ancestries.

Next, our results revealed important biological insights of the identified genes, including “oxidative damage”, “extracellular matrix organisation”, and “coregulation of androgen receptor activation, suggesting potential mechanisms and pathways involved in stroke pathogenesis [73,74]. It is believed that oxidative stress along with chronic neuroinflammation results in neuronal damage associated with stroke [75]. Oxidative stress develops due to the imbalance between the formation and clearance of free radicals during pathological conditions [76]. It is also linked to the aging processes that increase risk of stroke [77]. The brain cells are more susceptible to oxidative damage due to their excessive metabolic rate and sensitivity to ischemia damage [75,78]. In

addition, several pioneering next-generation sequencing studies, particularly transcriptomic and epitranscriptomic studies, have suggested substantial contribution of oxidative damage, neuroinflammation, and angiogenesis regulation pathways in response to physiological alterations in brain arteriovenous malformation, cerebral cavernous malformation, and retinal degeneration [79–82]. The extracellular matrix (ECM) contains collagen, fibrous proteins, elastin, laminins, and microfibrils that give structural support, integrity, and elasticity to the tissue [83]. The structure and expression of ECM alter after a stroke [84,85]. Our study also highlighted some molecular functions, particularly metal ion binding and RNA polymerase II transcription factor binding. Some metal ions are important nutrients for brain function and development [86]. Evidence shows that excessive amounts of some metal ions (e.g., Ca, Zn, Fe, or Cu) contribute to brain injury after stroke. Moreover, numerous exogenous metal ions have also been associated with stroke risk, including cadmium, nickel, arsenic,

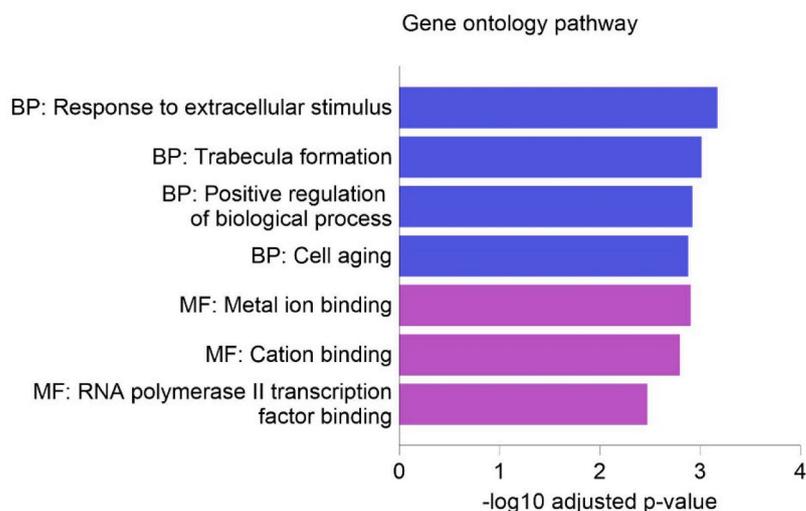
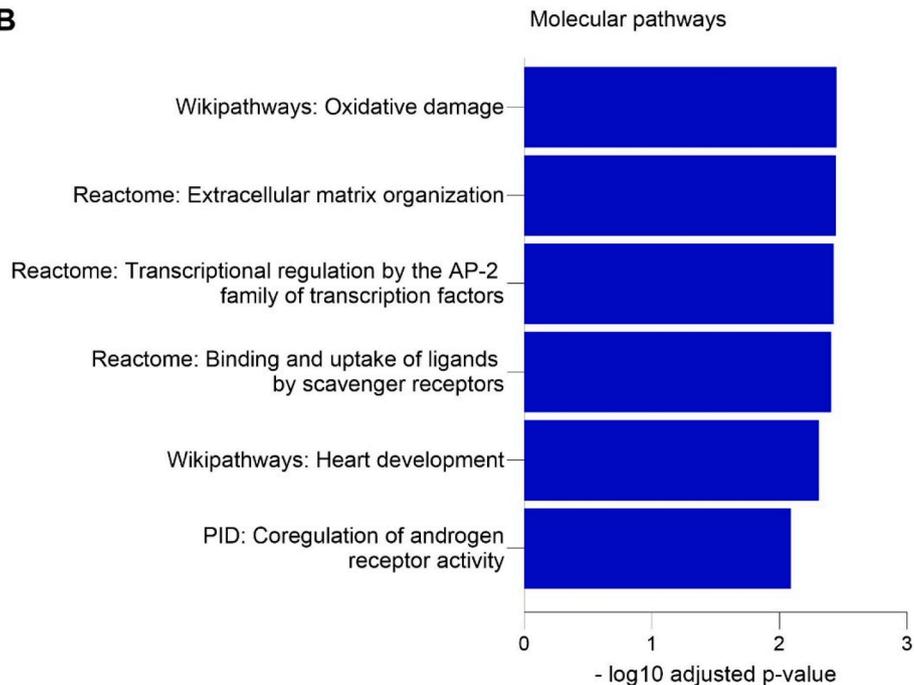
A**B**

Fig. 4. (A) Gene enrichment analysis (blue colour represent biological process and purple colour represents molecular function). (B) The pathway enrichment analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mercury, and aluminium [86].

There are several strengths in this study. We did the largest integration of multi-omics datasets for stroke to identify candidate biomarkers, including the largest eQTL datasets, mQTL datasets, GWAS datasets, followed by validation of the candidate biomarkers as differentially expressed in blood transcriptomic in stroke compared to healthy individuals [25]. In addition, we validated our candidate biomarkers in real human blood gene expression datasets of stroke versus healthy individuals, suggesting support for blood biomarkers of stroke.

Like other multi-omics approaches, it is important to acknowledge a few limitations as various bioinformatics analyses have been provided in this study, which need to be considered when interpreting the results. We used multiple corrections to limit the false positive rates. As with other SMR investigations, we cannot rule out the potential that any of

the associations (crucial SNPs, genes, and DNAm sites) identified in this study were false positives. In addition, we used eQTLs and mQTLs data to investigate the impact of genetic variant on gene expression and DNA methylation using SMR. Although the majority of eQTLs and mQTLs can reveal certain causal variants, the presence of complex linkage disequilibrium structure and secondary signals either from GWAS or eQTLs make it challenging to identify actual causal variant. Although we tried to guard against potential biases in our analyses with the use of *cis*-eQTL and *cis*-mQTLs, assessment for heterogeneity (HEIDI test), and multiple error corrections, the actual causal associations could still come from remote variants [28,87,88]. We analysed an independent blood transcriptomic dataset of 23 S and 23 controls for the differential expression of potential gene biomarkers. This dataset has a limited sample size. Therefore, we propose future work with a larger sample size

for further validation of ten candidate blood biomarkers. There is a lack of publicly available data on global methylation profiling on stroke, therefore, we could not assess whether the DNA methylation level of candidate biomarkers was altered in stroke patients versus healthy controls. Future studies should be conducted to validate these candidates in large-scale human DNA methylation profiling and gene expression profiling datasets to gain a better understanding of the functions of those genes in the aetiology of stroke risk and establish them as biomarkers of stroke before clinical use. We propose further experimental validation of candidate biomarkers using relevant clinical samples in future studies since our approach focused on identifying potential candidate biomarkers for stroke risk.

5. Conclusions

Overall, we leveraged large-scale multi-omics datasets using a systems biology approach to identify putative pleiotropic and potential causal gene biomarkers of stroke risk. We identified 45 candidate gene biomarkers by network SMR approach. Of those genes, ten genes were differentially expressed in the blood of stroke patients compared to healthy individuals. Our approach showed the importance of gene expression modulation by DNA methylation in stroke risk. Furthermore, our results suggest potential biological processes and pathways underlying the pathophysiology of stroke. The genetic variants, DNA methylation sites, and genes associated with stroke risk identified in this study demand further wet-lab experimental validation to evaluate their potential utility in stroke assessment.

CRedit authorship contribution statement

Tania Islam: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Md Rezanur Rahman:** Methodology, Software, Writing - review & editing. **Asaduzzaman Khan:** Supervision, Writing – review & editing. **Mohammad Ali Moni:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

“The MEGASTROKE project received funding from sources specified at <http://www.megastroke.org/acknowledgments.html>.

The author list of MEGASTROKE is available in Supplementary Table 9.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jbi.2023.104345>.

References

- [1] C. Hajat, R. Dundas, J.A. Stewart, E. Lawrence, A.G. Rudd, R. Howard, C.D. A. Wolfe, Cerebrovascular risk factors and stroke subtypes: differences between ethnic groups, *Stroke* 32 (2001) 37–42.
- [2] C. Feng, Y. Yang, S. Yang, X. Tu, Y. Wang, Y. Song, R. Hui, W. Zhang, Effect of gene–gene and gene–environment interaction on the risk of first-ever stroke and poststroke death, *Mol. Genet. Genomic Med.* 7 (2019) e846.
- [3] R.L. Sacco, S.E. Kasner, J.P. Broderick, L.R. Caplan, J.J. Connors, A. Culebras, M.S. V. Elkind, M.G. George, A.D. Hamdan, R.T. Higashida, An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association, *Stroke* 44 (2013) 2064–2089.
- [4] T.M. Woodruff, J. Thundiyil, S.-C. Tang, C.G. Sobey, S.M. Taylor, T.V. Arumugam, Pathophysiology, treatment, and animal and cellular models of human ischemic stroke, *Mol. Neurodegener.* 6 (2011) 1–19.
- [5] Y.-C. Cheng, J.R. O’Connell, J.W. Cole, O.C. Stine, N. Dueker, P.F. McArdle, M. J. Sparks, J. Shen, C.C. Laurie, S. Nelson, Genome-wide association analysis of ischemic stroke in young adults, *G3 Genes| Genomes| Genet.* 1 (2011) 505–514.
- [6] Y. Cao, X. Fan, W. Zhu, G. Ni, Y. Qian, Y. Han, M. Ma, Association of C7673T polymorphism in apolipoprotein B gene with ischemic stroke in the Chinese population: a meta-analysis, *Int. J. Neurosci.* 126 (2016) 584–592.
- [7] M. Traylor, M. Farrall, E.G. Holliday, C. Sudlow, J.C. Hopewell, Y.-C. Cheng, M. Fornage, M.A. Ikram, R. Malik, S. Bevan, Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies, *Lancet Neurol.* 11 (2012) 951–962.
- [8] S. Gretarsdottir, G. Thorleifsson, A. Manolescu, U. Styrkarsdottir, A. Helgadóttir, A. Gschwendtner, K. Kostulas, G. Kuhlenbäumer, S. Bevan, T. Jonsdóttir, Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke, *Ann. Neurol.* 64 (2008) 402–409.
- [9] D.F. Gudbjartsson, H. Holm, S. Gretarsdottir, G. Thorleifsson, G.B. Walters, G. Thorgeirsson, J. Gulcher, E.B. Mathiesen, I. Njåstad, A. Nyrnes, A sequence variant in ZFXH3 on 16q22 associates with atrial fibrillation and ischemic stroke, *Nat. Genet.* 41 (2009) 876–878.
- [10] A. Gschwendtner, S. Bevan, J.W. Cole, A. Plourde, M. Matarin, H. Ross-Adams, T. Meitinger, E. Wichmann, B.D. Mitchell, K. Furie, Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke, *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* 65 (2009) 531–539.
- [11] A. Mishra, R. Malik, T. Hachiya, T. Jürgenson, S. Namba, D.C. Posner, F. K. Kamani, M. Koido, Q. Le Grand, M. Shi, Stroke genetics informs drug discovery and risk prediction across ancestries, *Nature* 611 (2022) 115–123.
- [12] C.A. de Leeuw, J.M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data, *PLoS Comput. Biol.* 11 (2015) e1004219.
- [13] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E.J. C. De Geus, D.I. Boomsma, F.A. Wright, Integrative approaches for large-scale transcriptome-wide association studies, *Nat. Genet.* 48 (2016) 245–252.
- [14] Z. He, L. Liu, C. Wang, Y. Le Guen, J. Lee, S. Gogarten, F. Lu, S. Montgomery, H. Tang, E.K. Silverman, Identification of putative causal loci in whole-genome sequencing data via knockoff statistics, *Nat. Commun.* 12 (2021) 3152.
- [15] P.M. Visscher, M.A. Brown, M.I. McCarthy, J. Yang, Five years of GWAS discovery, *Am. J. Hum. Genet.* 90 (2012) 7–24.
- [16] R.S.N. Fehrmann, R.C. Jansen, J.H. Veldink, H.-J. Westra, D. Arends, M.J. Bonder, J. Fu, P. Deelen, H.J.M. Groen, A. Smolonska, Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA, *PLoS Genet.* 7 (2011) e1002197.
- [17] D.G. Hernandez, M.A. Nalls, M. Moore, S. Chong, A. Dillman, D. Trabzuni, J. R. Gibbs, M. Ryten, S. Arepalli, M.E. Weale, Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain, *Neurobiol. Dis.* 47 (2012) 20–28.
- [18] D.L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M.E. Dolan, N.J. Cox, Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS, *PLoS Genet.* 6 (2010) e1000888.
- [19] G.A. Wray, The evolutionary significance of cis-regulatory mutations, *Nat. Rev. Genet.* 8 (2007) 206–216.
- [20] C. Giambartolomei, D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, *PLoS Genet.* 10 (2014) e1004383.
- [21] H. Guo, M.D. Fortune, O.S. Burren, E. Schofield, J.A. Todd, C. Wallace, Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases, *Hum. Mol. Genet.* 24 (2015) 3305–3313.
- [22] Y. Sun, D. Zhou, M.R. Rahman, J. Zhu, D. Ghoneim, N.J. Cox, T.G. Beach, C. Wu, E. R. Gamazon, L. Wu, A transcriptome-wide association study identifies novel blood-based gene biomarker candidates for Alzheimer’s disease risk, *Hum. Mol. Genet.* 31 (2022) 289–299.
- [23] A. Hukku, M. Pividori, F. Luca, R. Pique-Regi, H.K. Im, X. Wen, Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations, *Am. J. Hum. Genet.* 108 (2021) 25–35.
- [24] F. Hormozdizari, M. Van De Bunt, A.V. Segre, X. Li, J.W.J. Joo, M. Bilow, J.H. Sul, S. Sankararaman, B. Pasaniuc, E. Eskin, Colocalization of GWAS and eQTL signals detects target genes, *Am. J. Hum. Genet.* 99 (2016) 1245–1260.
- [25] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M.R. Robinson, J.E. Powell, G.W. Montgomery, M.E. Goddard, N.R. Wray, P.M. Visscher, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets, *Nat. Genet.* 48 (2016) 481–487.
- [26] S. Zhao, H. Jiang, Z.-H. Liang, H. Ju, Integrating multi-omics data to identify novel disease genes and single-nucleotide polymorphisms, *Front. Genet.* 10 (2020) 1336.
- [27] J.M.W. Pavlides, Z. Zhu, J. Gratten, A.F. McRae, N.R. Wray, J. Yang, Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits, *Genome Med.* 8 (2016) 1–6.
- [28] E. Porcu, S. Rüeger, K. Lepik, F.A. Santoni, A. Reymond, Z. Kutalik, Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits, *Nat. Commun.* 10 (2019) 1–12.
- [29] X. Mo, Y. Guo, Q. Qian, M. Fu, S. Lei, Y. Zhang, H. Zhang, Mendelian randomization analysis revealed potential causal factors for systemic lupus erythematosus, *Immunology* 159 (2020) 279–288.

- [30] Y. Wu, J. Zeng, F. Zhang, Z. Zhu, T. Qi, Z. Zheng, L.R. Lloyd-Jones, R.E. Marioni, N. G. Martin, G.W. Montgomery, Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits, *Nat. Commun.* 9 (2018) 1–14.
- [31] Y. Sun, J. Zhu, D. Zhou, S. Canchi, C. Wu, N.J. Cox, R.A. Rissman, E.R. Gamazon, L. Wu, A transcriptome-wide association study of Alzheimer's disease using prediction models of relevant tissues identifies novel candidate susceptibility genes, *Genome Med.* 13 (2021) 1–11.
- [32] M.R. Rahman, M.C. Petralia, R. Ciurleo, A. Bramanti, P. Fagone, M. Shahjaman, L. Wu, Y. Sun, B. Turanli, K.Y. Arga, Comprehensive analysis of RNA-seq gene expression profiling of brain transcripts reveals novel genes, regulators, and pathways in autism spectrum disorder, *Brain Sci.* 10 (2020) 747.
- [33] M.R. Rahman, T. Islam, T. Zaman, M. Shahjaman, M.R. Karim, F. Huq, J.M. W. Quinn, R.M.D. Holsinger, E. Gov, M.A. Moni, Identification of molecular signatures and pathways to identify novel therapeutic targets in Alzheimer's disease: Insights from a systems biomedicine perspective, *Genomics* 112 (2020) 1290–1299.
- [34] D. Liu, Y. Wang, H. Jing, Q. Meng, J. Yang, Novel DNA methylation loci and genes showing pleiotropic association with Alzheimer's dementia: a network Mendelian randomization analysis, *Epigenetics* 17 (2022) 746–758.
- [35] Y. Xiao, M. Bi, H. Guo, M. Li, Multi-omics approaches for biomarker discovery in early ovarian cancer diagnosis, *EBioMedicine* 79 (2022), 104001.
- [36] M.A. Hossain, S.M.S. Islam, J.M.W. Quinn, F. Huq, M.A. Moni, Machine learning and bioinformatics models to identify gene expression patterns of ovarian cancer associated with disease progression and mortality, *J. Biomed. Inform.* 100 (2019), 103313.
- [37] T. Islam, R. Rahman, E. Gov, B. Turanli, G. Gulfidan, A. Haque, K.Y. Arga, N. Haque Mollah, Drug targeting and biomarkers in head and neck cancers: insights from systems biology analyses, *Omi. a J. Integr. Biol.* 22 (2018) 422–436.
- [38] M.R. Rahman, T. Islam, E. Gov, B. Turanli, G. Gulfidan, M. Shahjaman, N. Akhter Banu, M.N.H. Mollah, K.Y. Arga, M.A. Moni, Identification of prognostic biomarker signatures and candidate drugs in colorectal cancer: insights from systems biology analysis, *Medicina (B, Aires)* 55 (2019) 20.
- [39] T.Y. Demirtas, M.R. Rahman, M.C. Yurtsever, E. Gov, Forecasting Gastric Cancer Diagnosis, Prognosis, and Drug Repurposing with Novel Gene Expression Signatures, *Omi. A J. Integr. Biol.* 26 (2022) 64–74.
- [40] S. Doran, M. Arif, S. Lam, A. Bayraktar, H. Turkez, M. Uhlen, J. Boren, A. Mardinoglu, Multi-omics approaches for revealing the complexity of cardiovascular disease, *Brief. Bioinform.* 22 (2021) bbab061.
- [41] I. Subramanian, S. Verma, S. Kumar, A. Jere, K. Anamika, Multi-omics data integration, interpretation, and its application, *Bioinform. Biol. Insights.* 14 (2020) 1177932219899051.
- [42] J. Yan, S.L. Risacher, L. Shen, A.J. Saykin, Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data, *Brief. Bioinform.* 19 (2018) 1370–1381.
- [43] U. Vösa, A. Claringbould, H.-J. Westra, M.J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Kasela, Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis, *BioRxiv.* (2018) 447367.
- [44] U. Vösa, A. Claringbould, H.-J. Westra, M.J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Yazar, Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression, *Nat. Genet.* 53 (2021) 1300–1310.
- [45] L.R. Lloyd-Jones, A. Holloway, A. McRae, J. Yang, K. Small, J. Zhao, B. Zeng, A. Bakshi, A. Metspalu, M. Dermizakis, The genetic architecture of gene expression in peripheral blood, *Am. J. Hum. Genet.* 100 (2017) 228–237.
- [46] B.H. Chen, R.E. Marioni, E. Colicino, M.J. Peters, C.K. Ward-Caviness, P.-C. Tsai, N. S. Roetker, A.C. Just, E.W. Demerath, W. Guan, DNA methylation-based measures of biological age: meta-analysis predicting time to death, *Aging (Albany NY)* 8 (2016) 1844.
- [47] V.K. Rakyan, T.A. Down, D.J. Balding, S. Beck, Epigenome-wide association studies for common human diseases, *Nat. Rev. Genet.* 12 (2011) 529–541.
- [48] A.F. McRae, R.E. Marioni, S. Shah, J. Yang, J.E. Powell, S.E. Harris, J. Gibson, A. K. Henders, L. Bowdler, J.N. Painter, Identification of 55,000 replicated DNA methylation QTL, *Sci. Rep.* 8 (2018) 1–9.
- [49] R. Malik, G. Chauhan, M. Traylor, M. Sargurupremraj, Y. Okada, A. Mishra, L. Rutten-Jacobs, A.-K. Giese, S.W. Van Der Laan, S. Gretarsdottir, Multiethnicity genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes, *Nat. Genet.* 50 (2018) 524–537.
- [50] G. Davey Smith, G. Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, *Hum. Mol. Genet.* 23 (2014) R89–R98.
- [51] Y. Wu, J. Zeng, F. Zhang, Z. Zhu, T. Qi, Z. Zheng, L.R. Lloyd-Jones, R.E. Marioni, N. G. Martin, G.W. Montgomery, Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits, *Nat. Commun.* (2018, 2018), 9.
- [52] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B.* 57 (1995) 289–300.
- [53] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K. A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, NCBI GEO: archive for functional genomics data sets—update, *Nucleic Acids Res.* 41 (2012) D991–D995.
- [54] B. Stamova, G.C. Jickling, B.P. Ander, X. Zhan, D. Liu, R. Turner, C. Ho, J. C. Khoury, C. Bushnell, A. Pancioli, Gene expression in peripheral immune cells following cardioembolic stroke is sexually dimorphic, *PLoS One* 9 (2014) e102550.
- [55] A. Kamburov, C. Wierling, H. Lehrach, R. Herwig, ConsensusPathDB—a database for integrating human functional interaction networks, *Nucleic Acids Res.* 37 (2009) D623–D628.
- [56] G. Chauhan, S. DeBette, Genetic risk factors for ischemic and hemorrhagic stroke, *Curr. Cardiol. Rep.* 18 (2016) 1–11.
- [57] G.J. Falcone, D. Woo, Genetics of spontaneous intracerebral hemorrhage, *Stroke* 48 (2017) 3420–3424.
- [58] M. Dichgans, R. Malik, I.R. König, J. Rosand, R. Clarke, S. Gretarsdottir, G. Thorleifsson, B.D. Mitchell, T.L. Assimes, C. Levi, Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants, *Stroke* 45 (2014) 24–36.
- [59] S. Lanfranconi, H.S. Markus, COL4A1 mutations as a monogenic cause of cerebral small vessel disease: a systematic review, *Stroke* 41 (2010) e513–e518.
- [60] B.N. Nandeesh, P.S. Bindu, G. Narayanappa, T. Chickabasaviah Yasha, A. Mahadevan, K. Kulanthaivelu, V. Santosh, Cerebral small vessel disease with hemorrhagic stroke related to COL4A1 mutation: A case report, *Neuropathology* 40 (2020) 93–98.
- [61] K. Hara, A. Shiga, T. Fukutake, H. Nozaki, A. Miyashita, A. Yokoseki, H. Kawata, A. Koyama, K. Arima, T. Takahashi, Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease, *N. Engl. J. Med.* 360 (2009) 1729–1739.
- [62] T. Fukutake, Cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL): from discovery to gene identification, *J. Stroke Cerebrovasc. Dis.* 20 (2011) 85–93.
- [63] M. Traylor, E. Persyn, L. Tomppo, S. Klasson, V. Abedi, M.K. Bakker, N. Torres, L. Li, S. Bell, L. Rutten-Jacobs, Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies, *Lancet Neurol.* 20 (2021) 351–361.
- [64] M.Y. Frédéric, C. Monino, C. Marschall, D. Hamroun, L. Faivre, G. Jondeau, H. Klein, L. Neumann, E. Gautier, C. Binquet, The FBN2 gene: new mutations, locus-specific database (Universal Mutation Database FBN2), and genotype-phenotype correlations, *Hum. Mutat.* 30 (2009) 181–190.
- [65] R. Ratnapriya, X. Zhan, R.N. Fariss, K.E. Branham, D. Zipprer, C.F. Chakarova, Y. V. Sergeev, M.M. Campos, M. Othman, J.S. Friedman, Rare and common variants in extracellular matrix gene Fibrillin 2 (FBN2) are associated with macular degeneration, *Hum. Mol. Genet.* 23 (2014) 5827–5837.
- [66] L.S. Lim, P. Mitchell, J.M. Seddon, F.G. Holz, T.Y. Wong, Age-related macular degeneration, *Lancet* 379 (2012) 1728–1738.
- [67] X.-B. Mo, H. Zhang, A.-L. Wang, T. Xu, Y.-H. Zhang, Integrative analysis identifies the association between CASZ1 methylation and ischemic stroke, *Neurol. Genet.* 6 (2020).
- [68] J.Y. Lee, B.S. Lee, D.J. Shin, K. Woo Park, Y.A. Shin, K. Joong Kim, L. Heo, J.Y. Lee, Y.K. Kim, Y. Jin Kim, C.B. Hong, S.H. Lee, D. Yoon, H. Jung Ku, I.Y. Oh, B.J. Kim, J. Lee, S.J. Park, J. Kim, H.K. Kawk, J.E. Lee, H.K. Park, J.E. Lee, H.Y. Nam, H. Y. Park, C. Shin, M. Yokota, H. Asano, M. Nakatochi, T. Matsubara, H. Kitajima, K. Yamamoto, H.L. Kim, B.G. Han, M.C. Cho, Y. Jang, H.S. Kim, J. Euy Park, J. Y. Lee, A genome-wide association study of a coronary artery disease risk variant, *J. Hum. Genet.* 58 (2013) 120–126, <https://doi.org/10.1038/jhg.2012.124>.
- [69] L. Bian, R.L. Hanson, Y.L. Muller, L. Ma, S. Kobes, W.C. Knowler, C. Bogardus, L. J. Baier, Variants in ACAD10 are associated with type 2 diabetes, insulin resistance and lipid oxidation in Pima Indians, *Diabetologia* 53 (2010) 1349–1353.
- [70] H. Doong, A. Vrailes, E.C. Kohn, What's in the 'BAG'?—a functional domain analysis of the BAG-family proteins, *Cancer Lett.* 188 (2002) 25–32.
- [71] Z. He, X. Ni, L. Xia, Z. Shao, Overexpression of NIMA-related kinase 6 (NEK6) contributes to malignant growth and dismal prognosis in Human Breast Cancer, *Pathol. Pract.* 214 (2018) 1648–1654.
- [72] Z.-M. Yang, B. Liao, S.-S. Yang, T. Su, J. Zhang, W.-M. Wang, Predictive Role of NEK6 in Prognosis and Immune Infiltration in Head and Neck Squamous Cell Carcinoma, *Front. Endocrinol. (Lausanne)* 13 (2022).
- [73] C.L. Allen, U. Bayraktutan, Oxidative stress and its role in the pathogenesis of ischaemic stroke, *Int. J. Stroke.* 4 (2009) 461–470.
- [74] D.N. Edwards, G.J. Bix, Roles of blood-brain barrier integrins and extracellular matrix in stroke, *Am. J. Physiol. Physiol.* 316 (2019) C252–C263.
- [75] S.A. Saeed, K.F. Shad, T. Saleem, F. Javed, M.U. Khan, Some new prospects in the understanding of the molecular basis of the pathogenesis of stroke, *Exp. Brain Res.* 182 (2007) 1–10.
- [76] G. Pizzino, N. Irrera, M. Cucinotta, G. Pallio, F. Mannino, V. Arcoraci, F. Squadrito, D. Altavilla, A. Bitto, Oxidative stress: harms and benefits for human health, *Oxid. Med. Cell. Longev.* 2017 (2017).
- [77] M.A. Birch-Machin, A. Bowman, Oxidative stress and ageing, *Br. J. Dermatol.* 175 (2016) 26–29.
- [78] M. Jelinek, M. Jurajda, K. Duris, Oxidative Stress in the Brain: Basic Concepts and Treatment Strategies in Stroke, *Antioxidants.* 10 (2021) 1886.
- [79] C. Scimone, L. Donato, S. Alibrandi, T. Esposito, C. Alafaci, R. D'Angelo, A. Sidoti, Transcriptome analysis provides new molecular signatures in sporadic Cerebral Cavernous Malformation endothelial cells, *Biochim. Biophys. Acta (BBA)-Molecular Basis Dis.* (2020, 1866.), 165956.
- [80] C. Scimone, L. Donato, C. Alafaci, F. Granata, C. Rinaldi, M. Longo, R. D'Angelo, A. Sidoti, High-throughput sequencing to detect novel likely gene-disrupting variants in pathogenesis of sporadic brain arteriovenous malformations, *Front. Genet.* 11 (2020) 146.
- [81] C. Scimone, F. Granata, M. Longo, E. Mormina, C. Turiaco, A.A. Caragliano, L. Donato, A. Sidoti, R. D'Angelo, Germline mutation enrichment in pathways controlling endothelial cell homeostasis in patients with brain arteriovenous malformation: implication for molecular diagnosis, *Int. J. Mol. Sci.* 21 (2020) 4321.
- [82] L. Donato, C. Scimone, S. Alibrandi, S.Z. Scalinci, C. Rinaldi, R. D'Angelo, A. Sidoti, Epitranscriptome Analysis of Oxidative Stressed Retinal Epithelial Cells Depicted a Possible RNA Editing Landscape of Retinal Degeneration, *Antioxidants.* 11 (2022) 1967.
- [83] A.D. Theocharis, S.S. Skandalis, C. Gialeli, N.K. Karamanos, Extracellular matrix structure, *Adv. Drug Deliv. Rev.* 97 (2016) 4–27.

- [84] Y. Yao, in: *Extracellular Matrix in Stroke BT - Cerebral Ischemic Reperfusion Injuries (CIRI): Bench Research and Clinical Implications*, Springer International Publishing, Cham, 2018, pp. 121–144, https://doi.org/10.1007/978-3-319-90194-7_8.
- [85] Y. Yao, *Extracellular matrix in stroke*, in: *Cereb. Ischemic Reperfus, Inj.*, Springer, 2018, pp. 121–144.
- [86] Y. V Li, J.H. Zhang, Metal ions in stroke pathophysiology, in: *Met. Ion Stroke*, Springer, 2012: pp. 1–12.
- [87] A. Battle, S.B. Montgomery, Determining causality and consequence of expression quantitative trait loci, *Hum. Genet.* 133 (2014) 727–735.
- [88] H. Lin, F. Wang, A.J. Rosato, L.A. Farrer, D.C. Henderson, H. Zhang, Prefrontal cortex eQTLs/mQTLs enriched in genetic variants associated with alcohol use disorder and other diseases, *Epigenomics* 12 (2020) 789–800.