



# Integrating Multi-Omics Data to Identify Novel Disease Genes and Single-Nucleotide Polymorphisms

Sheng Zhao<sup>1</sup>, Huijie Jiang<sup>1\*</sup>, Zong-Hui Liang<sup>2\*</sup> and Hong Ju<sup>3\*</sup>

<sup>1</sup> Department of Radiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, <sup>2</sup> Department of Radiology, Jian'an District Centre Hospital of Fudan University, Shanghai, China, <sup>3</sup> Department of Information Engineering, Heilongjiang Biological Science and Technology Career Academy, Harbin, China

## OPEN ACCESS

### Edited by:

Lei Deng,  
Central South University, China

### Reviewed by:

Tianyi Zhao,  
Harvard University, United States  
Hao Lin,  
University of Electronic Science and  
Technology of China, China

### \*Correspondence:

Huijie Jiang  
jianghuijie@hrbmu.edu.cn  
Zong-Hui Liang  
liangzh@vip.163.com  
Hong Ju  
hongju.hit@hotmail.com

### Specialty section:

This article was submitted to  
Statistical Genetics  
and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 15 October 2019

**Accepted:** 06 December 2019

**Published:** 24 January 2020

### Citation:

Zhao S, Jiang H, Liang Z-H and Ju H  
(2020) Integrating Multi-Omics Data to  
Identify Novel Disease Genes and  
Single-Nucleotide Polymorphisms.  
*Front. Genet.* 10:1336.  
doi: 10.3389/fgene.2019.01336

Stroke ranks the second leading cause of death among people over the age of 60 in the world. Stroke is widely regarded as a complex disease that is affected by genetic and environmental factors. Evidence from twin and family studies suggests that genetic factors may play an important role in its pathogenesis. Therefore, research on the genetic association of susceptibility genes can help understand the mechanism of stroke. Genome-wide association study (GWAS) has found a large number of stroke-related loci, but their mechanism is unknown. In order to explore the function of single-nucleotide polymorphisms (SNPs) at the molecular level, in this paper, we integrated 8 GWAS datasets with brain expression quantitative trait loci (eQTL) dataset to identify SNPs and genes which are related to four types of stroke (ischemic stroke, large artery stroke, cardioembolic stroke, small vessel stroke). Thirty-eight SNPs which can affect 14 genes expression are found to be associated with stroke. Among these 14 genes, 10 genes expression are associated with ischemic stroke, one gene for large artery stroke, six genes for cardioembolic stroke and eight genes for small vessel stroke. To explore the effects of environmental factors on stroke, we identified methylation susceptibility loci associated with stroke using methylation quantitative trait loci (MQTL). Thirty-one of these 38 SNPs are at greater risk of methylation and can significantly change gene expression level. Overall, the genetic pathogenesis of stroke is explored from locus to gene, gene to gene expression and gene expression to phenotype.

**Keywords:** stroke, genome-wide association study, expression quantitative trait loci, mQTL, SMR, single-nucleotide polymorphisms

## INTRODUCTION

Stroke is a major cerebrovascular disease caused by a transient or permanent decrease of local cerebral blood flow. It is characterized by arterial obstruction (Krishnamurthi et al., 2018), so it is also called cerebral infarction (Dargazanli et al., 2018). According to the World Health Organization, stroke affects more than 15 million people worldwide and directly kills about 5.7 million people. It also causes approximately 5 million people to have a lifelong disability, while

about 4.3 million people died due to disability. At present, thrombolytic therapy (Castellanos et al., 2018) (recombinant tissue plasminogen activator) is the only acute treatment for ischemic stroke with a narrow time window (3–4.5 hours). Therefore, only 3.4%–5.2% of patients were treated within the short time window. Researchers have been focusing on how to improve the clinical diagnosis and treatment of cerebral infarction beyond the time window of thrombolysis (Feil et al., 2019).

The occurrence and development of ischemic stroke is affected by a variety of risk factors, such as family history of stroke (Zheng et al., 2019), history of heart disease (Beck et al., 2018), history of diabetes (Zou et al., 2018), history of hypertension, etc. According to the investigation and analysis of Li et al. (2019), the prevalence rate of the family with a family history of stroke is 10.52%. In recent years, a number of genetic association studies have suggested that there are multiple genetic risk factors for ischemic stroke, and multiple risk loci were found to affect the susceptibility to ischemic stroke.

Cacabelos et al. (2018) and Yee et al. (2019) showed that the C7673T polymorphism of APOB gene was significantly associated with the risk of ischemic stroke. Chen et al. (2019), Nordestgaard et al. (2018) confirmed that the polymorphism of  $\epsilon$ 2, $\epsilon$ 3, $\epsilon$ 4 of APOE gene was associated with ischemic stroke. APOB gene and APOE gene are both known ischemic stroke susceptibility genes because of blood lipid level. In addition, many studies have shown that the SG13S114 (rs10507391) polymorphism of ALOX5AP gene and SG13S32 (rs9551963) polymorphism are associated with susceptibility to ischemic stroke. Zheng et al., (2018) found that carriers of SG13S114 polymorphism TT/TA genotype of ALOX5AP gene had a higher risk of acute cerebral infarction. Naderi et al. (2019) showed that SG13S114 polymorphism of ALOX5AP gene was associated with acute cerebral infarction. Previous genetic studies have found that some ischemic stroke susceptibility genes on chromosome 14, such as GCH1 gene (Wei et al., 2018), MEG3 gene (Han et al., 2018), MMP-14 gene (Elgebaly et al., 2019), PRKCH gene (Krupinski et al., 2018), are associated with the risk of ischemic stroke.

Genome-wide association study (GWAS) reveals candidate loci, susceptible genes and their loci related to the occurrence, development and treatment of diseases by genome-wide high-density genetic markers (Pei Li and Wang, 2015; Cheng et al., 2019a; Cheng et al., 2019b). Since 2009, GWAS has been widely used to explore and excavate candidate gene loci related to new types of stroke. GWAS is generally believed to be able to identify some previously undetected or identified biological markers related to stroke (Ye et al., 2018; Cheng et al., 2019c), and because of its large sample size, it can minimize false positive results. The National Institute of Neurological Diseases (NIND) has conducted the largest and most comprehensive GWAS to explore the genetic loci of stroke and its subtypes. The results supported the previously established genetic association of ischemic stroke. New loci on chromosome 1p13 (such as rs12122341 of TSPAN2 gene) have been found to be associated with ischemic stroke. Although GWAS has many advantages and

is widely used, it is still very hard to understand the role of nucleotide polymorphism (SNP) loci in diseases from the huge results of GWAS.

Therefore, recently many researchers have tried to integrate GWAS with expression quantitative trait loci (eQTL) to mine the disease-related genes (Cheng et al., 2018a; Cheng et al., 2018b). Since eQTL conveys gene expression information and GWAS conveys disease-related SNPs information, combining the two datasets, we could know the loci which are associated with diseases because of affecting other genes expression. Zhao et al. (2019) found many Alzheimer's disease-related genes and SNPs by GWAS and eQTL. Asthma-related genes were identified by Li et al. (2015). by integrating GWAS and eQTL. Systematic integration of Brain eQTL and GWAS were done by Luo et al. (2015) and they identified ZNF323 as a novel Schizophrenia risk gene.

Zhu et al. (2016) generalized Mendelian randomization to SMR. SMR is used to test the association between a trait and the expression level of each gene across the whole genome using summary data from GWAS and eQTL studies. SMR is a common tool to identify the genes whose expression levels are associated with a complex trait because of pleiotropy. Twenty-eight GWAS datasets are used by Pavlides et al. (2016) to find genes whose expression levels were associated with complex phenotype. Bone mineral density (BMD)-related genes are studied by Meng et al. (2018) using SMR. SMR is also used to identify genes and pathways for Amyotrophic Lateral Sclerosis by Du et al. (2017). Fan et al. (2017) found 6 genes are associated with neuroticism by SMR. Liu et al. (2018) used SMR on doing research on Obesity and found 20 BMI associated genes. Veturi and Ritchie (2018) compared two popular methods: MP and SMR by different datasets. Though these scholars' researches, we could judge that SMR is an effective tool. In this paper, summary-level data mendelian randomization (SMR) is used to integrate GWAS and eQTL datasets. In this way, the most functionally relevant genes at the loci identified in GWAS for stroke are found.

## METHODS

### Work Frame

As shown in **Figure 1**, since GWAS has identified SNPs which are related to stroke, and eQTL has identified SNPs which can affect genes expression, SMR is used to identify SNPs that can change gene expression and this should be the reason that they are associated with stroke. Therefore, firstly, we should obtain GWAS and eQTL data. Then, we checked the overlap between these two datasets. Finally, SMR is used to screen SNPs.

### SMR

$z$  in summary data level Mendelian Randomization (SMR) is a genetic variant (SNP),  $x$  is the expression level of a gene and  $y$  denotes the trait, then the two-step least-squares estimate of the effect of  $x$  on  $y$  from an MR analysis is:

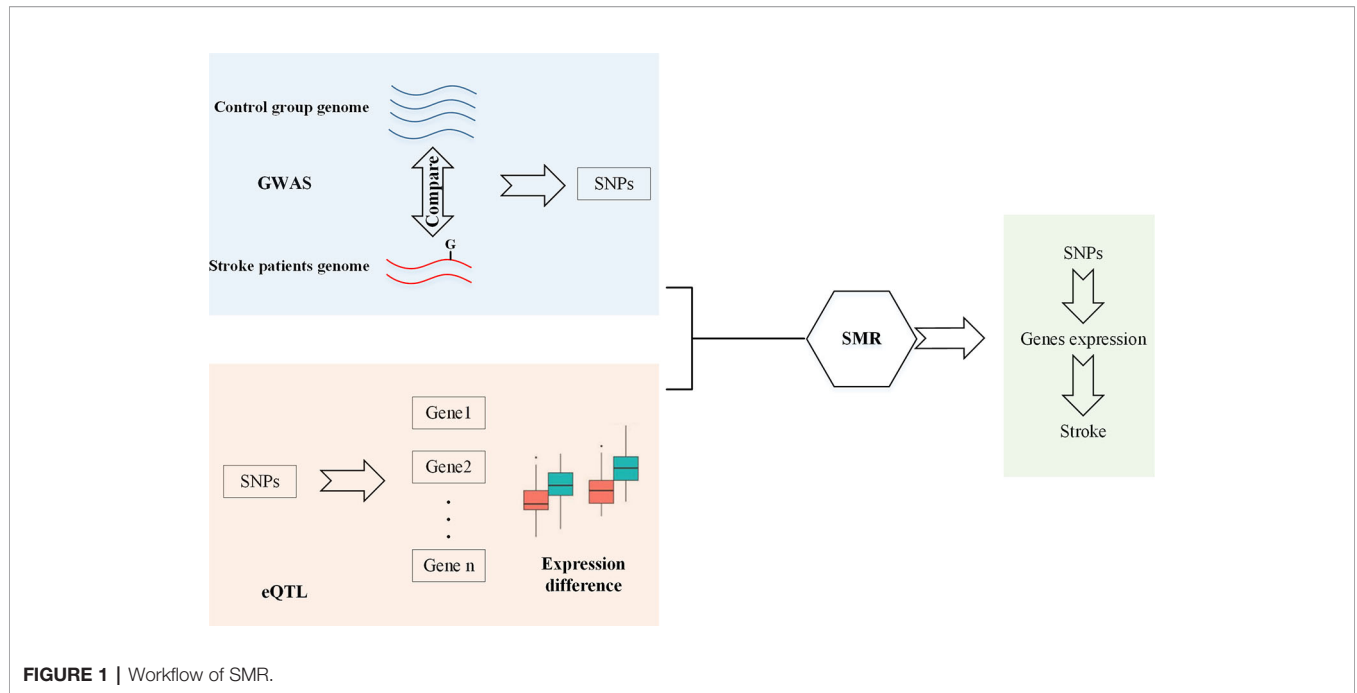


FIGURE 1 | Workflow of SMR.

$$\hat{b}_{xy} = \hat{b}_{zy} / \hat{b}_{zx} \tag{1}$$

$\hat{b}_{zy}$  and  $\hat{b}_{zx}$  are the least-squares estimates of y and x on z, respectively. Then,  $\hat{b}_{xy}$  denotes the effect size of x on y without confounding from non-genetic factors. The variance of  $\hat{b}_{xy}$  is:

$$T_{MR} = \hat{b}_{xy}^2 / \text{var}(\hat{b}_{xy}) \tag{2}$$

Here,  $T_{MR}$  obeys a chi-square distribution with a degree of freedom of 1. As we can see in equation (Dargazanli et al., 2018), MR requires genotype, gene expression and phenotype to be measured on the same sample. However, Zhu et al. have proved that the power of detecting  $\hat{b}_{xy}$  can be greatly increased using a two-sample MR analysis. Therefore, the  $T_{MR}$  can be replaced by  $T_{SMR}$ .

$$T_{SMR} = \hat{b}_{xy}^2 / \text{var}(\hat{b}_{xy}) \approx \frac{z_{zy}^2 z_{zx}^2}{z_{zy}^2 + z_{zx}^2} \tag{3}$$

$z_{zy}$  is the z statistics from GWAS and  $z_{zx}$  is the z statistics from eQTL.

## RESULTS

### Data Description

#### GWAS

We used the data from Malik et al.'s research. Eight GWAS datasets are used. **Table 1** shows the detailed information about these data.

We collected GWAS data for four different types of stroke (ischemic stroke, large artery stroke, cardioembolic stroke, small vessel stroke).

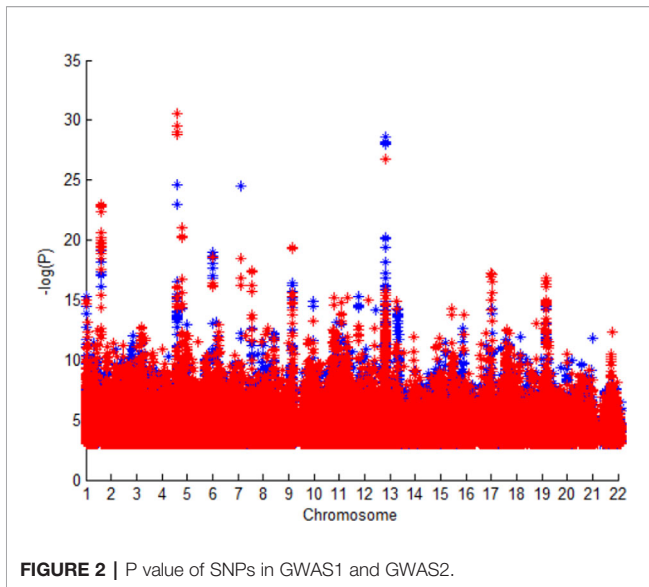
**Figure 2** shows P value of SNPs in GWAS1 and GWAS2. The SNPs are almost same in these GWAS dataset, but difference races cause the difference of P value. We could know different races have different stroke susceptibility genes.

#### eQTL

eQTL data is from a meta-analysis of GTEx brain (Consortium G, 2017), CMC (Fromer et al., 2016), and ROSMAP (Ng et al., 2017). All the data are from brain. Only SNPs within 1Mb distance from each probe are available. The estimated effective n is 1,194.

TABLE 1 | GWAS data description.

Dataset	Disease	Sample
GWAS 1	ischemic stroke	Europeans (40,585 cases; 406,111 controls)
GWAS 2	ischemic stroke	trans-ethnic meta-analysis (67,162 cases; 454,450 controls)
GWAS 3	large artery stroke	Europeans (40,585 cases; 406,111 controls)
GWAS 4	large artery stroke	trans-ethnic meta-analysis (67,162 cases; 454,450 controls)
GWAS 5	cardioembolic stroke	Europeans (40,585 cases; 406,111 controls)
GWAS 6	cardioembolic stroke	trans-ethnic meta-analysis (67,162 cases; 454,450 controls)
GWAS 7	small vessel stroke	Europeans (40,585 cases; 406,111 controls)
GWAS 8	small vessel stroke	trans-ethnic meta-analysis (67,162 cases; 454,450 controls)



**mQTL**

mQTL used in this paper is a set of brain data from a meta-analysis of ROSMAP (Ng et al., 2017), Hannon et al. (2016) and Jaffe et al. (2016). In the ROSMAP data, only SNPs within 5Kb of each DNA methylation probe are available. In the Hannon et al. data, only SNPs within 500Kb distance from each probe and with PmQTL < 1.0e-10 are available. In the Jaffe et al. data, only SNPs within 20Kb distance from each probe and with FDR < 0.1 are available. The estimated effective n is 1,160.

**Four Kinds of Stroke**

Ischemic stroke is a kind of stroke which caused by arterial obstruction. It accounts for approximately 85% of the total. large artery stroke and cardioembolic stroke are the subgroup of this kind of this stroke.

Large artery stroke is caused by blood clots (thrombus) which are formed in the neck or cerebral arteries. There may be accumulation of fatty deposits (often referred to as plaques) in these arteries.

Cardioembolic stroke is caused by blood clots that reach the brain and blocks the blood vessels. A common cause is the formation of blood clots in the two upper atrial rhythm abnormalities of the heart (atrial fibrillation).

Small vessel stroke is actually a transient stroke symptom that usually lasts only a few minutes. small vessel stroke is caused by transient blood supply to specific parts of the brain and does not cause significant persistent effects on patients. However, it is generally believed that the risk of stroke after small vessel stroke is higher.

**SNPs and Genes for Ischemic Stroke**

10 SNPs which change six genes expression are screened by Europeans dataset and 11 SNPs which change five genes expression are screened by trans-ethnic dataset.

As we can see in **Table 2**, HSD17B12 is overlapped in the two tests. Moreno et al. (2018) found upregulation of HSD17B12 is

**TABLE 2 |** SMR results of ischemic stroke.

SNP	P-value	Gene
Europeans dataset		
rs9651613	4.17E-06	HSD17B12
rs648997	5.72E-06	ALDH2
rs11065976	6.36E-06	ALDH2
rs4286007	6.70E-06	CKAP2
rs847892	7.79E-06	ALDH2
rs66480035	7.97E-06	ALDH2
rs532436	7.99E-06	SURF1
rs487399	8.21E-06	CEP192
rs11618716	8.80E-06	CKAP2
rs11620062	9.24E-06	CKAP2
Trans-ethnic dataset		
rs9651613	3.58E-07	HSD17B12
rs10838185	5.14E-06	HSD17B12
rs6599175	5.42E-06	ULK4
rs6801343	5.55E-06	ULK4
rs9874975	5.70E-06	ULK4
rs12774577	7.99E-06	C10orf32
rs10400343	8.38E-06	HSD17B12
rs3087681	8.47E-06	C10orf32
rs2371623	8.81E-06	ULK4
rs9825741	9.00E-06	ULK4
rs11191606	9.04E-06	C10orf32

associated ischemic stroke using 82 cases and 67 controls. ALDH2 is generally considered as a gene (Guo et al., 2013) which can protect against ischemic stroke, because overexpression of ALDH2 rescued neuronal survival against 4-HNE treatment in PC12 cells (Lee et al., 2012). These two genes show the accuracy of our results.

**SNPs and Genes for Large Artery Stroke**

None SNP is screened by Europeans dataset for large artery stroke. Three SNPs which correspond one gene ‘C3orf18’ are screened by trans-ethnic dataset.

Phenotypes for C3orf18 Gene include Decreased homologous recombination repair frequency, Decreased ionizing radiation sensitivity, Upregulation of Wnt pathway, Increased vaccinia virus (VACV) infection, Mildly decreased CFP-tsO45G cell surface transport. It is considered to be associated with cognitive function measurement.

**SNPs and Genes for Cardioembolic Stroke**

11 SNPs are significant in Europeans dataset and trans-ethnic dataset. rs3807989 is screened more than one time in Europeans dataset because it can affect more than one gene expression. Both CAV1 and CAV2’s expression can be changed by this SNP.

As we can see in **Table 3**, 6 genes and 3 genes are screened by SMR in Europeans dataset and Trans-ethnic dataset, respectively. Three of them are overlapped.

**SNPs and Genes for Small Vessel Stroke**

13 SNPs and 4 SNPs are significant in Europeans dataset and trans-ethnic dataset, respectively. None of these SNPs or their corresponding genes are overlapped in these two tests. As we can see in **Table 4**, although no overlap is found between these two

**TABLE 3** | SMR results of cardioembolic stroke.

SNP	P-value	Gene
Europeans dataset		
rs3807989	2.03E-05	CAV1
rs532436	4.03E-05	SURF1
rs72790984	4.68E-05	PLEKHH2
rs11773845	4.96E-05	CAV1
rs4745721	4.96E-05	ECD
rs1997571	5.62E-05	CAV1
rs507666	6.08E-05	SURF1
rs1997572	6.20E-05	CAV1
rs9313620	6.27E-05	BNIP1
rs76192127	6.35E-05	ECD
rs3807989	6.58E-05	CAV2
rs2519093	7.40E-05	SURF1
rs600038	9.67E-05	SURF1
Trans-ethnic dataset		
rs4745721	2.21E-05	ECD
rs76192127	2.87E-05	ECD
rs532436	3.37E-05	SURF1
rs507666	4.15E-05	SURF1
rs616154	5.26E-05	SURF1
rs72790984	5.60E-05	PLEKHH2
rs2519093	5.92E-05	SURF1
rs72790983	6.37E-05	PLEKHH2
rs559723	7.26E-05	SURF1
rs183153921	7.34E-05	ECD
rs3878005	9.62E-05	ECD

**TABLE 4** | SMR results of small vessel stroke.

SNP	P-value	Gene
Europeans dataset		
rs3807989	2.03E-05	CAV1
rs532436	4.03E-05	SURF1
rs72790984	4.68E-05	PLEKHH2
rs11773845	4.96E-05	CAV1
rs4745721	4.96E-05	ECD
rs1997571	5.62E-05	CAV1
rs507666	6.08E-05	SURF1
rs1997572	6.20E-05	CAV1
rs9313620	6.27E-05	BNIP1
rs76192127	6.35E-05	ECD
rs3807989	6.58E-05	CAV2
rs2519093	7.40E-05	SURF1
rs600038	9.67E-05	SURF1
Trans-ethnic dataset		
rs2501966	3.53E-06	CENPQ
rs6599175	4.49E-06	ULK4
rs2501965	4.77E-06	CENPQ
rs9874975	6.07E-06	ULK4

tests, some genes are overlapped between cardioembolic stroke and small vessel stroke.

## SNPs Changes Gene Expression Level by Methylation

Since both genetic and environmental factors are key to cause stroke, while methylation plays an important role in the interaction between environmental factors and genetic expression, we assumed that some of the SNPs identified above are at greater risk of methylation and can change gene expression levels.

Therefore, we integrated the SNPs found above with mQTL data for research. Thirty-eight unique SNPs are found in four different types of stroke. Thirty-one of these 38 SNPs are significant in mQTL dataset. We draw the P value of these 31 SNPs as **Figure 2**. As shown in **Figure 3**, most of these SNPs are associated with several genes expression. In addition, most of SNPs have a quite low P value, which means that they can significantly change the expression of genes.

## Case Study ULK4

Guo et al. (2016) have found that genetic variants in LRP1 and ULK4 are associated with acute aortic dissections. In their paper, they also mentioned that ULK4 may contribute stroke.

### CAV1

Shyu et al. (2017) discussed association of eNOS and CAV1 gene polymorphisms with susceptibility risk of large artery atherosclerotic stroke. A tendency toward an increased LAA stroke risk was significant in carriers with the eNOS Glu298Asp variant in conjunction with the G14713 A and T29107A polymorphisms of the CAV1 (aOR = 2.03, P-trend = 0.002).

### CAV2

Jolobe (2012) found that recurrent stroke is because of a novel voltage sensor mutation in CAV2. They compared stroke mouse and normal mouse to obtain this conclusion.

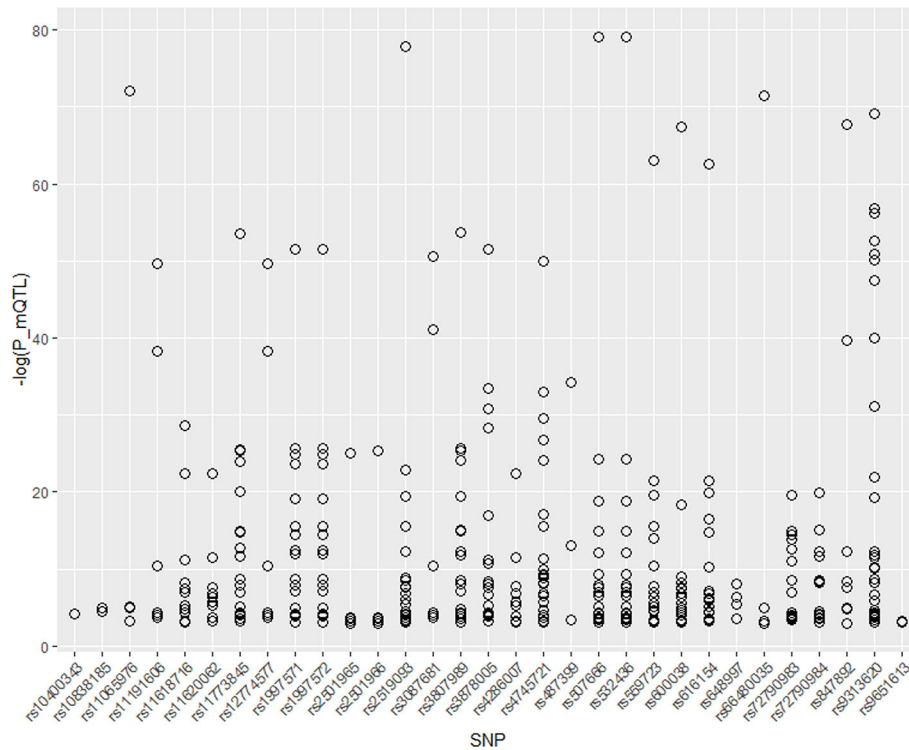
## CONCLUSIONS

Stroke is the primary cause of disability in adults, which constitutes a serious public health burden. Stroke is generally believed to be caused by genetic and environmental factors. Therefore, in this paper, we identified stroke-related genes and loci from both genetic and environmental aspects.

GWAS identified a large number of stroke-related SNPs, which were difficult to explain. We tried to identify the pathogenesis of significant SNPs by combining SMR with eQTL data. Since eQTL shows the SNPs that can significantly change genes expression and GWAS shows the SNPs that are significant related to stroke, we combined these two data to identify the genes whose expression levels are associated with stroke because of pleiotropy.

38 SNPs which cause changes in 14 genes expression were found by 8 GWAS data and brain eQTL. Those 8 GWAS data are from two different races sample and include four types of stroke (ischemic stroke, large artery stroke, cardioembolic stroke, small vessel stroke). CAV1, SURF1, PLEKHH2, ECD, BNIP1, CAV2 are found to be associated with cardioembolic stroke and Small vessel stroke in Europeans. ULK4 is a susceptibility gene for ischemic stroke and small vessel stroke.

Since methylation (Lv et al., 2019) plays an important role in the interaction between environmental factors and genetic expression, we tried to find out whether 38 SNPs are affected by methylation and lead to the changes in other genes expression levels. Thirty-one of these 38 SNPs are significant in mQTL data and most of them can affect more than one gene expression.



**FIGURE 3** | P value of 31 significant SNPs in mQTL.

Overall, integrating GWAS with eQTL, we found 38 SNPs and 14 genes are related to stroke by SMR. Thirty-one of 38 SNPs are at high risk of methylation which can also cause changes in gene expression. These findings serve as a guide to understanding the pathogenesis of stroke at the molecular level.

## DATA AVAILABILITY STATEMENT

All the datasets used in this paper could be downloaded from GWAS: [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST006908/MEGASTROKE.2.AIS.EU](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST006908/MEGASTROKE.2.AIS.EU)

R.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST005843/MEGASTROKE.2.AIS.TR](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST005843/MEGASTROKE.2.AIS.TR)

ANS.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST006907/MEGASTROKE.3.LAS.EU](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST006907/MEGASTROKE.3.LAS.EU)

R.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST005840/MEGASTROKE.3.LAS.TR](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST005840/MEGASTROKE.3.LAS.TR)

ANS.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST005842/MEGASTROKE.4.CES.TR](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST005842/MEGASTROKE.4.CES.TR)

ANS.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST006910/MEGASTROKE.4.CES.EU](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST006910/MEGASTROKE.4.CES.EU)

R.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST005841/MEGASTROKE.5.SVS.TR](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST005841/MEGASTROKE.5.SVS.TR)

ANS.out [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/MalikR\\_29531354\\_GCST006909/MEGASTROKE.5.SVS.EU](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST006909/MEGASTROKE.5.SVS.EU)

R.out eQTL: <https://cnsgenomics.com/software/smr/#eQTLsummarydata>

mQTL: <https://cnsgenomics.com/software/smr/#mQTLsummarydata>.

## AUTHOR CONTRIBUTIONS

HuJ, Z-HL, and HoJ conceived and designed the experiments. SZ analyzed data. SZ, HuJ, Z-HL, and HoJ wrote this manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was supported by grants from the National Natural Science Foundation of China (81671760 and 81873910), Scientific Research Transformation Special Fund of Heilongjiang Academy of Medical Sciences (2018415); Scientific Research Project of Health and Family Planning Commission of Heilongjiang Province (201812 and 201622), National Natural Science Foundation of China (81871423), and Shanghai Municipal Commission of Health and Family Planning (20160064).

## REFERENCES

- Beck, J. D., Moss, K. L., Morelli, T., and Offenbacher, S. (2018). Periodontal profile class is associated with prevalent diabetes, coronary heart disease, stroke, and systemic markers of C-reactive protein and interleukin-6. *J. periodontol.* 89 (2), 157–165. doi: 10.1002/jper.17-0426
- Cacabelos, R., Lombardi, V., Fernández-Novoa, L., Carrera, I., Cacabelos, P., Corzo, L., et al. (2018). “Basic and Clinical Studies With Marine LipoFishins and Vegetal Favallins in Neurodegeneration and Age-Related Disorders,” in *Studies in Natural Products Chemistry*, vol. 59. (Netherlands: Elsevier), 195–225.
- Castellanos, M., van Eendenburg, C., Gubern, C., Kádár, E., Huguet, G., Puig, J., et al. (2018). Low levels of caveolin-1 predict symptomatic bleeding after thrombolytic therapy in patients with acute ischemic stroke. *Stroke* 49 (6), 1525–1527. doi: 10.1161/strokeaha.118.020683
- Chen, J.-X., Liu, J., Hu, F., Bi, Y., Li, M., and Zhao, L. (2019). Genetic variants on chromosome 9p21 confer risks of cerebral infarction in the Chinese population: a meta-analysis. *Int. J. immunopathol. Pharmacol.* 33, 2058738419847852. doi: 10.1177/2058738419847852
- Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018a). DincRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* 34 (11), 1953–1956. doi: 10.1093/bioinformatics/bty002
- Cheng, L., Zhuang, H., Yang, S., Jiang, H., Wang, S., and Zhang, J. (2018b). Exposing the causal effect of C-reactive protein on the risk of type 2 diabetes mellitus: a mendelian randomization study. *Front. In Genet.* 9, 657. doi: 10.3389/fgene.2018.00657
- Cheng, L., Qi, C., Zhuang, H., Fu, T., and Zhang, X. (2019a). gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.* 48 (D1), D554–D560 doi: 10.1093/nar/gkz843
- Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2019b). LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* 47 (D1), D140–D144. doi: 10.1093/nar/gky1051
- Cheng, L., Zhuang, H., Ju, H., Yang, S., Han, J., Tan, R., et al. (2019c). Exposing the causal effect of body mass index on the risk of type 2 diabetes mellitus: a mendelian randomization study. *Front. In Genet.* 10, 94. doi: 10.3389/fgene.2019.00094
- Consortium G (2017). Genetic effects on gene expression across human tissues. *Nature* 550 (7675), 204. doi: 10.1038/nature24277.
- Dargazanli, C., Fahed, R., Blanc, R., Gory, B., Labreuche, J., Duhamel, A., et al. (2018). Modified thrombolysis in cerebral infarction 2c/thrombolysis in cerebral infarction 3 reperfusion should be the aim of mechanical thrombectomy: insights from the ASTER Trial (Contact Aspiration Versus Stent Retriever for Successful Revascularization). *Stroke* 49 (5), 1189–1196. doi: 10.1161/strokeaha.118.020700
- Du, Y., Yan, W., Guo, X., Hao, J., Wang, W., He, A., et al. (2017). A genome-wide expression association analysis identifies genes and pathways associated with amyotrophic lateral sclerosis. *Cell. Mol. Neurobiol.* 38 (3), 1–5. doi: 10.1007/s10571-017-0512-2
- Elgebaly, M. M., Arreguin, J., and Storke, N. (2019). Targets, treatments, and outcomes updates in diabetic stroke. *J. Stroke Cerebrovasc. Dis.* 28 (6), 1413–1420 doi: 10.1016/j.jstrokecerebrovasdis.2019.02.005
- Fan, Q., Wang, W., Hao, J., He, A., Wen, Y., Guo, X., et al. (2017). Integrating genome-wide association study and expression quantitative trait loci data identifies multiple genes and gene set associated with neuroticism. *Prog. In Neuropsychopharmacol. Biol. Psychiatry* 28 (6), 1413–1420. doi: 10.1016/j.pnpbp.2017.05.017
- Feil, K., Reidler, P., Kunz, W. G., Küpper, C., Heinrich, J., Laub, C., et al. (2019). Addressing a real life problem: treatment with intravenous thrombolysis and mechanical thrombectomy in acute stroke patients with an extended time window beyond 4.5 hours based on computed tomography perfusion imaging. *Eur. J. Neurol.* 27 (1), 168–174 doi: 10.7861/clinmedicine.17-2-161
- Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* 19 (11), 1442. doi: 10.1038/nn.4399
- Guo, J.-M., Liu, A.-J., Zang, P., Dong, W.-Z., Ying, L., Wang, W., et al. (2013). ALDH2 protects against stroke by clearing 4-HNE. *Cell Res.* 23 (7), 915. doi: 10.1038/cr.2013.69
- Guo, D.-c, Grove, M. L., Prakash, S. K., Eriksson, P., Hostetler, E. M., LeMaire, S. A., et al. (2016). Genetic variants in LRP1 and ULK4 are associated with acute aortic dissections. *Am. J. Hum. Genet.* 99 (3), 762–769. doi: 10.1016/j.ajhg.2016.06.034
- Han, X., Zheng, Z., Wang, C., and Wang, L. (2018). Association between MEG3/miR-181b polymorphisms and risk of ischemic stroke. *Lipids In Health Dis.* 17 ((1)), 292. doi: 10.1186/s12944-018-0941-z
- Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., et al. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* 19 (1), 48. doi: 10.1038/nn.4182
- Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R., et al. (2016). Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* 19 (1), 40. doi: 10.1038/nn.4181
- Jolobe, O. M. (2012). Stroke and familial hemiplegic migraine. *Lancet Neurol.* 11 (6), 484. doi: 10.1016/s1474-4422(12)70123-0
- Krishnamurthi, R. V., Barker-Collo, S., Parag, V., Parmar, P., Witt, E., Jones, A., et al. (2018). Stroke incidence by major pathological type and ischemic subtypes in the Auckland regional community stroke studies: changes between 2002 and 2011. *Stroke* 49 (1), 3–10. doi: 10.1161/strokeaha.117.019358
- Krupinski, J., Carrera, C., Muiño, E., Torres, N., Al-Baradie, R., Culler, N., et al. (2018). DNA methylation in stroke. Update of latest advances. *Comput. Struct. Biotechnol. J.* 16, 1–5. doi: 10.1016/j.csbj.2017.12.001
- Lee, W.-C., Wong, H.-Y., Chai, Y.-Y., Shi, C.-W., Amino, N., Kikuchi, S., et al. (2012). Lipid peroxidation dysregulation in ischemic stroke: plasma 4-HNE as a potential biomarker? *Biochem. Biophys. Res. Commun.* 425 (4), 842–847. doi: 10.1016/j.bbrc.2012.08.002
- Li, X., Hastie, A. T., Hawkins, G. A., Moore, W. C., Ampleford, E. J., Milosevic, J., et al. (2015). eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy* 70 (10), 1309–1318. doi: 10.1111/all.12683
- Li, W., Wang, D., Wang, X., Gong, Y., Cao, S., Yin, X., et al. (2019). The association of metabolic syndrome components and diabetes mellitus: evidence from China National Stroke Screening and Prevention Project. *BMC Public Health* 19 (1), 192. doi: 10.1186/s12889-019-6415-z
- Liu, L., Fan, Q., Zhang, F., Guo, X., Liang, X., Du, Y., et al. (2018). A genome-wide integrative analysis of GWAS and eQTLs Data identifies multiple genes and gene sets associated with obesity. *BioMed. Res. Int.* 2018, 3848560. doi: 10.1155/2018/3848560
- Luo, X.-J., Mattheisen, M., Li, M., Huang, L., Rietschel, M., Børglum, A. D., et al. (2015). Systematic integration of brain eQTL and GWAS identifies ZNF323 as a novel schizophrenia risk gene and suggests recent positive selection based on compensatory advantage on pulmonary function. *Schizophr. Bull.* 41 (6), 1294–1308. doi: 10.1093/schbul/sbv017
- Lv, H., Zhang, Z. M., Li, S. H., Tan, J. X., Chen, W., and Lin, H. (2019). Evaluation of different computational methods on 5-methylcytosine sites identification. *Briefings In Bioinf.* doi: 10.1093/bib/bb2048
- Meng, X. H., Chen, X. D., Greenbaum, J., Zeng, Q., You, S. L., Xiao, H. M., et al. (2018). Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions. *Bone* 113, 41–48. doi: 10.1016/j.bone.2018.05.012
- Moreno-Ramírez, C. E., Gutiérrez-Garzón, E., Barreto, G. E., and Forero, D. A. (2018). Genome-wide expression profiles for ischemic stroke: a meta-analysis. *J. Stroke Cerebrovasc. Dis.* 27 (11), 3336–3341. doi: 10.1016/j.jstrokecerebrovasdis.2018.07.035
- Naderi, N., Yousefi, H., Mollazadeh, S., Seyed Mikaeili, A., Keshavarz Norouzpour, M., Jazebi, M., et al. (2019). Inflammatory and immune response genes: a genetic analysis of inhibitor development in Iranian hemophilia A patients. *Pediatr. Hematol. Oncol.* 36 (1), 28–39. doi: 10.1080/08880018.2019.1585503
- Ng, B., White, C. C., Klein, H.-U., Sieberts, S. K., McCabe, C., Patrick, E., et al. (2017). An xQTL map integrates the genetic architecture of the human brain’s transcriptome and epigenome. *Nat. Neurosci.* 20 (10), 1418. doi: 10.1038/nn.4632
- Nordestgaard, L. T., Tybjaerg-Hansen, A., Rasmussen, K. L., Nordestgaard, B. G., and Frikke-Schmidt, R. (2018). Genetic variation in clusterin and risk of dementia and ischemic vascular disease in the general population: cohort studies and meta-analyses of 362,338 individuals. *BMC Med.* 16 (1), 39. doi: 10.1016/j.atherosclerosis.2018.06.075

- Pavlidis, J. M. W., Zhu, Z., Gratten, J., Mcrae, A. F., Wray, N. R., and Yang, J. (2016). Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* 8 (1), 84. doi: 10.1186/s13073-016-0338-4
- Pei Li, M. G., and Wang, C. (2015). Xiaoyan Liu, Quan Zou: An overview of SNP interactions in genome-wide association studies. *Briefings In Funct. Genomics* 14 (2), 143–155. doi: 10.1093/bfgp/elu036
- Shyu, H.-Y., Chen, M.-H., Hsieh, Y.-H., Shieh, J.-C., Yen, L.-R., Wang, H.-W., et al. (2017). Association of eNOS and Cav-1 gene polymorphisms with susceptibility risk of large artery atherosclerotic stroke. *PLoS One* 12 (3), e0174110. doi: 10.1371/journal.pone.0174110
- Sun, W., Han, Y., Yang, S., Zhuang, H., Zhang, J., Cheng, L., et al. (2019). The assessment of interleukin-18 on the risk of coronary heart disease. *Med. Chem.* doi: 10.2174/1573406415666191004115128
- Veturi, Y., and Ritchie, M. D. (2018). How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures? *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 23, 228–239. doi: 10.1142/9789813235533\_0021
- Wei, J., Zhang, Y., Li, Z., Wang, X., Chen, L., Du, J., et al. (2018). GCH1 attenuates cardiac autonomic nervous remodeling in canines with atrial-tachypacing via tetrahydrobiopterin pathway regulated by microRNA-206. *Pacing Clin. Electrophysiol.* 41 (5), 459–471. doi: 10.1111/pace.13289
- Ye, Z., Zhang, H., Sun, L., Cai, H., Hao, Y., Xu, Z., et al. (2018). GWAS-supported CRP gene polymorphisms and functional outcome of large artery atherosclerotic stroke in Han Chinese. *Neuromol. Med.* 20 (2), 225–232. doi: 10.1007/s12017-018-8485-y
- Yee, J., Kim, W., Chang, B. C., Chung, J. E., Lee, K. E., and Gwak, H. S. (2019). APOB gene polymorphisms may affect the risk of minor or minimal bleeding complications in patients on warfarin maintaining therapeutic INR. *Eur. J. Hum. Genet.* 27 (10), 1542–1549. doi: 10.1038/s41431-019-0450-1
- Zhao, T., Hu, Y., Zang, T., Wang, Y., and Integrate, G. W. A. S. (2019). eQTL, and mQTL data to identify alzheimer's disease-related genes. *Front. In Genet.* 10, 1021. doi: 10.3389/fgene.2019.01021
- Zheng, Z., Liu, S., Wang, C., and Han, X. (2018). A functional polymorphism rs145204276 in the promoter of long noncoding RNA GAS5 is associated with an increased risk of ischemic stroke. *J. Stroke Cerebrovasc. Dis.* 27 (12), 3535–3541. doi: 10.1016/j.jstrokecerebrovasdis.2018.08.016
- Zheng, X., Zeng, N., Wang, A., Zhu, Z., Peng, H., Zhong, C., et al. (2019). Family history of stroke and death or vascular events within one year after ischemic stroke. *Neurol. Res.* 41 (5), 466–472. doi: 10.1080/01616412.2019.1577342
- Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48 (5), 481. doi: 10.1038/ng.3538
- Zhuang, H., Zhang, Y., Yang, S., Cheng, L., and Liu, S. L. (2019). A mendelian randomization study of infant length and type 2 diabetes mellitus risk. *Curr. Gene Ther.* 19 (4), 224–231(8) doi: 10.2174/1566523219666190925115535
- Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., and Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Front. In Genet.* 9, 515. doi: 10.3389/fgene.2018.00515

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhao, Jiang, Liang and Ju. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.