Published in final edited form as:

Stroke. 2020 December ; 51(12): 3751–3755. doi:10.1161/STROKEAHA.120.030260.

Replication of top loci from COL4A1/2 associated with white matter hyperintensity burden in ischemic stroke patients

Jiang Li, MD, & PhD[1], **Vida Abedi, PhD**[1,3], **Regeneron Genetic Center**[5], **Ramin Zand, MD**[2], **Christoph J. Griessenauer, MD**[2,4,*]

[1]Department of Molecular and Functional Genomics, Geisinger, Danville, PA, 17822, USA.

[2]Neuroscience Institute, Geisinger, Danville, PA 17822, USA.

[3]Biocomplexity Institute, Virginia Tech, Blacksburg, VA, USA.

[4]Research Institute of Neurointervention, Paracelsus Medical University, Salzburg, Austria.

[5]Regeneron Genetic Center, LLC, Tarrytown, NY 10591, USA.

Abstract

Background and Purpose: The purpose of this study was to replicate the top loci associated with white matter hyperintensity (WMH) phenotypes identified by large genome-wide association studies and the loci identified from the previous candidate gene studies.

Methods: A total of 946 Geisinger MyCode acute ischemic stroke (AIS) patients with validated European ancestry and MRI data were included in this study. Log transformed WMH volume (WMHv), as a quantitative trait, was calculated by a fully automated quantification process. The GWAS was carried out by a linear mixed regression model (GEMMA). A candidate-SNP analysis by including known SNPs, reported from a meta-analysis and several large GWAS for WMH, was conducted in all cases and binary converted extreme cases.

Results: No genome-wide significantly associated variants were identified. In a candidate-SNP study, rs9515201 (*COL4A2*) and rs3744028 (*TRIM65*), two known genetic loci, showed nominal or trend of association with the WMHv (β=0.13 and p=0.001 for rs9515201; β=0.094 and p=0.094 for rs3744028), and replicated in a subset of extreme cases versus controls (OR= 1.78, $p = 7.74\times10^{-4}$ for rs9515201; OR=1.53, p=0.047 for rs3744028, respectively). MTHFR677 cytosine/thymine (rs1801133) also showed an association with the binary WMH with OR=1.47 for T allele (p=0.019).

Conclusion: Replication of COL4A1/2 associated with WMH reassures that the genetic risk factors for monogenic and polygenic ischemic stroke are shared at gene level.

[*]Corresponding author: Christoph J. Griessenauer, MD, mailing address: Neuroscience Institute, Geisinger, 100 North Academy Ave. Danville, PA, 17822; phone number: 570-214-4101; christoph.griessenauer@gmail.com.

Conflict of interest: No

History of procoagulant disorder was acquired from the chart review without performing the validation testing on, for example, a Factor V Leiden mutation. The following link lists the conditions which make a patient qualified in this category if any of those conditions have been documented: https://www.ihtc.org/inherited-blood-clots/

Keywords

white matter hyperintensity; cerebral small vessel disease; genome-wide association study; stroke; COL4A2; COL4A1; TRIM65

Introduction

The genetic basis of WMH has been investigated since the early linkage studies[1, 2] as well as several large-scale genome-wide association studies (GWAS)[3–8]. Due to its high heritability, WMH burden becomes the most commonly used intermediate phenotype to access the heritability for cerebral small vessel disease (CSVD). WMH volume (WMHv) is the frequently used quantitative trait, calculated manually or automatically by machine-learning algorithms, semi-quantitative visual rating scales, and can be binarily converted to represent a subset of so-called 'exteme' cases with high or low WMH burden. The discrepancy in heritability for WMH estimated by twin/family studies and by GWAS can be at least partially explained by the contribution of rare variants with larger effect size and high penetrance in familial monogenic CSVD versus the contribution of common variants with smaller effect size and the polygenic nature in sporadic CSVD. Nevertheless, the causal mechanisms such as *COL4A1/2, NOTCH3* for WMH in monogenic and sporadic CSVD can be shared at gene and pathway level.[9] A recent Phase II GWAS on periventricular or deep WMH not only confirms these known genetic loci/genes specific to the anatomically stratified endophenotypes of WMH, but also identifies novel loci/genes[10].

The purpose of this study was to identify common variants associated with WMH burden among patients with acutely ischemic stroke (AIS) and to replicate the genetic loci associated with WMH phenotypes through a candidate-SNP approach.

Methods

The summary statistics of this Geisinger cohort may be shared with third party upon execution of data sharing agreement. Such request should be addressed to the corresponding author.

The study cohort was made up of participants of the Geisinger's MyCode® Community Health Initiative consisting of 946 AIS patients with validated European ancestry and MRI data. The informed consent was obtained for all MyCode® patients. This study was approved by the Geisinger Institutional Review Board. Patient Characteristics, Clinical Variables, and Outcome Measures were based on the neurological examination and corresponding neuroimaging[11]. Quantification of WMHv using a fully automated pipeline was conducted on clinical brain MRIs obtained at the time of the stroke. This pipeline integrated automated brain extraction, intensity normalization, and WMH segmentation[12]. The exponentially distributed WMHv (Supplementary Figure I) was converted into nearly normal distribution after a natural log transformation (ln[WMHv + 1]). Genotyping, imputation and quality control was described in Supplementary material *at* http://stroke.ahajournals.org. GEMMA(version 0.98.1), a linear mixed model, was adopted to test genetic associations with WMH volumes while accounting for covariates such as index age,

sex, and five major principal components, and cryptic relatedness between individuals with allelic dose (0, 1, or 2 copies of the reference allele) as the independent variable. We conducted a candidate-SNP analysis to validate known genetic variants based on a meta-analysis of previous candidate gene studies[13,14] and top 20 genetic loci identified by the large-scale GWAS[3–8] which were reviewed by others [9, 14, 15]. We selected patients distributed at the top (232 from 331) and bottom quantile (246 from 331) of WMHv (946 with genomic data from 1324) and subsequently convert into binary trait to represent subgroups of patients with high or low WMHv. A logistic regression model adjusted for covariates ($WMHv_{binary}$ ~ Age + Sex + $PC_{(1–5)}$ + Genotype), was considered. The p value and Odds Ratio (OR) of the SNPs in this subset of patients were also determined.

Results

Demographics and clinical characteristics of the cohort were listed in Figure 1A.

Baseline Ln_WMHv showed significant positive associations with some comorbidities such as hypertension (OR=1.07; p=0.001), diabetes (OR=1.25; p=0.007), and COPD (OR=1.31; p=0.028) (Figure 1B). It also predicted poor outcome at 90days such as mRS 3–5 (OR=1.42; p<0.001) and mRS 3–6 (OR=1.35; p=0.001), suggesting this WMHv could be considered as a surrogate biomarker for outcome prediction of IS.

No genome-wide significant variants were identified (Supplementary Figure II *at* http://stroke.ahajournals.org) suggesting the complexity of the trait and our GWAS was underpowered. We therefore conducted a candidate-SNP analysis by including several known SNPs, gathered from a meta-analysis for WMH[14]. We conducted a subgroup analysis by simulating ‘extreme’ cases and evaluated the association and the effect size of the SNPs with the binary WMHv. The frequency of MTHFR677 cytosine/thymine (rs1801133) showed difference between lower and upper quantile groups with OR=1.47 for T allele (p=0.019) (Table 1), the direction of which was consistent with previous studies.[14] Due to the nominal p value and the small effect size for T allele reported by others[14], we considered our cohort replicated this association.

Although only 15 SNPs out of 20 top GWAS loci having genetic data available in our dataset, rs9515201 (*COL4A1/2*) and rs3744028 (*TRIM65*) were replicated in this subset (OR=1.78, $p=7.74\times10^{-4}$; OR=1.53, p=0.047, respectively, Table 2) and showed nominal or trend of association with the quantitative Ln_WMHv in the entire cohort (MAF=0.318, β=0.13 and p_{wald}=0.001 for rs9515201; MAF=0.178, β=0.094 and p_{wald}=0.065 for rs3744028) (Supplementary Table I *at* http://stroke.ahajournals.org). The significance for the association of rs9515201 from *COL4A1/2* with the quantitative or binary WMHv survived the Bonferroni correction.

Discussion

Common variants (rs9521732, rs9521733, and rs9515199) from *COL4A1/A2* may contribute to the risk for sporadic CSVD and intracerebral hemorrhage (ICH) in a subtype of stroke patients and controls with European ancestry.[16] They have a moderate significance in association with lacunar stroke and WMH in symptomatic IS patients.[16] A multi-ethnic,

genome-wide meta-analyses of dementia- and stroke-free subjects also has revealed that a *COL4A2* SNP, rs9515201, is associated with WMH in community populations as well as stroke patients[3]. This SNP is in strong LD with three SNPs previously identified.[16] We also interrogated deep or periventricular WMH-specific loci, recently identified by a Phase II GWAS[10] including *COL4A2*(rs11838776) demonstrating a nominal association with WMHv(p=0.0046) and a trend towards an association with *TRIM47*(rs3744020; p=0.150) and *TRIM65*(rs35392904; p=0.189). A further analysis of the effect size and the association in an anatomically stratified subgroup would be interesting. rs1801133, also known as C677T, Ala222Val, or A222V, is the most commonly investigated common variant in the *MTHFR* gene. Individuals with this MTHFR mutation have elevated homocysteine levels and the replication of this association reconfirms that total homocysteine is associated with WMHv. Unlike baseline WMHv, these replicated variants cannot be solely considered as a proxy for WMHv and used as a surrogated biomarker to predict outcome (data not shown).

As more health care systems adopt this infrastructure by leveraging the EHR data, extensive sequencing/genotyping, in combination with automated WMH phenotyping as illustrated in Supplementary Figure III, we believe more genome-wide, replicable findings will be identified through meta-analyses. Secondary pheWAS would help to determine their pleiotropy as well as serving as clinical actionable biomarkers.

Conclusion

Replication of *COL4A1/2* associated with WMH reassures that monogenic and polygenic cerebral vascular disease are shared at the genetic level.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: VA has research support from the Defense Threat Reduction Agency (DTRA) grant No. HDTRA1–18-1–0008 and from NIH grant No. R56HL116832. RZ has research support from Bucknell University Initiative Program, Roche – Genentech Biotechnology Company, and the Geisinger Health Plan Quality fund.

Abbreviations:

WMH	white matter hyperintensity
LN_WMH	nature log transformed WMH
NIHSS	National Institute of Health Stroke Scale
NIHSS_7above	NIHSS ≥ 7
NIHSS_10above	NIHSS ≥ 10
NIHSS_16above	NIHSS ≥ 16
mRS02	modified Rankin Score from 0 to 2

mRS35	modified Rankin Score from 3 to 5
mRS36	modified Rankin Score from 3 to 6
mRS6	modified Rankin Score at 6
Thrombectomy	Mechanical thrombectomy
iv_tPA	intravenous tissue plasminogen activator
DAPT	Dual antiplatelet therapy
TAPT	Triple antiplatelet therapy
Anticoagulant	anticoagulant therapy
Procoagulant	History of procoagulant disorder

References

1. Turner ST, Fornage M, Jack CR Jr., Mosley TH, Kardia SL, Boerwinkle E, de Andrade M. Genomic susceptibility loci for brain atrophy in hypertensive sibships from the genoa study. Hypertension. 2005;45:793–798 [PubMed: 15699467]
2. Kochunov P, Glahn D, Winkler A, Duggirala R, Olvera RL, Cole S, Dyer TD, Almasy L, Fox PT, Blangero J. Analysis of genetic variability and whole genome linkage of whole-brain, subcortical, and ependymal hyperintense white matter volume. Stroke. 2009;40:3685–3690 [PubMed: 19834011]
3. Traylor M, Zhang CR, Adib-Samii P, Devan WJ, Parsons OE, Lanfranconi S, Gregory S, Cloonan L, Falcone GJ, Radmanesh F, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. Neurology. 2016;86:146–153 [PubMed: 26674333]
4. Network NSG, International Stroke Genetics C. Loci associated with ischaemic stroke and its subtypes (sign): A genome-wide association study. Lancet Neurol. 2016;15:174–184 [PubMed: 26708676]
5. Neurology Working Group of the Cohorts for H, Aging Research in Genomic Epidemiology Consortium tSGN, the International Stroke Genetics C. Identification of additional risk loci for stroke and small vessel disease: A meta-analysis of genome-wide association studies. Lancet Neurol. 2016;15:695–707 [PubMed: 27068588]
6. Rutten-Jacobs LCA, Tozer DJ, Duering M, Malik R, Dichgans M, Markus HS, Traylor M. Genetic study of white matter integrity in uk biobank (n=8448) and the overlap with stroke, depression, and dementia. Stroke. 2018;49:1340–1347 [PubMed: 29752348]
7. Verhaaren BF, Debette S, Bis JC, Smith JA, Ikram MK, Adams HH, Beecham AH, Rajan KB, Lopez LM, Barral S, et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on mri. Circ Cardiovasc Genet. 2015;8:398–409 [PubMed: 25663218]
8. Fornage M, Debette S, Bis JC, Schmidt H, Ikram MA, Dufouil C, Sigurdsson S, Lumley T, DeStefano AL, Fazekas F, et al. Genome-wide association studies of cerebral white matter lesion burden: The charge consortium. Ann Neurol. 2011;69:928–939 [PubMed: 21681796]
9. Dichgans M, Pulit SL, Rosand J. Stroke genetics: Discovery, biology, and clinical applications. Lancet Neurol. 2019;18:587–599 [PubMed: 30975520]
10. Armstrong NJ, Mather KA, Sargurupremraj M, Knol MJ, Malik R, Satizabal CL, Yanek LR, Wen W, Gudnason VG, Dueker ND, et al. Common genetic variation indicates separate causes for periventricular and deep white matter hyperintensities. Stroke. 2020;51:2111–2121 [PubMed: 32517579]
11. Hendrix P, Sofoluke N, Adams MD, Kunaprayoon S, Zand R, Kolinovsky AN, Person TN, Gupta M, Goren O, Schirmer CM, et al. Risk factors for acute ischemic stroke caused by anterior large vessel occlusion. Stroke. 2019;50:1074–1080 [PubMed: 31009355]

12. Schirmer MD, Dalca AV, Sridharan R, Giese AK, Donahue KL, Nardin MJ, Mocking SJT, McIntosh EC, Frid P, Wasselius J, et al. White matter hyperintensity quantification in large-scale clinical acute ischemic stroke cohorts - the mri-genie study. Neuroimage Clin. 2019;23:101884
13. Paternoster L, Chen W, Sudlow CL. Genetic determinants of white matter hyperintensities on brain scans: A systematic assessment of 19 candidate gene polymorphisms in 46 studies in 19,000 subjects. Stroke. 2009;40:2020–2026 [PubMed: 19407234]
14. Tran T, Cotlarciuc I, Yadav S, Hasan N, Bentley P, Levi C, Worrall BB, Meschia JF, Rost N, Sharma P. Candidate-gene analysis of white matter hyperintensities on neuroimaging. J Neurol Neurosurg Psychiatry. 2016;87:260–266 [PubMed: 25835038]
15. Cuadrado-Godia E, Dwivedi P, Sharma S, Ois Santiago A, Roquer Gonzalez J, Balcells M, Laird J, Turk M, Suri HS, Nicolaides A, et al. Cerebral small vessel disease: A review focusing on pathophysiology, biomarkers, and machine learning strategies. J Stroke. 2018;20:302–320 [PubMed: 30309226]
16. Rannikmae K, Davies G, Thomson PA, Bevan S, Devan WJ, Falcone GJ, Traylor M, Anderson CD, Battey TW, Radmanesh F, et al. Common variation in col4a1/col4a2 is associated with sporadic cerebral small vessel disease. Neurology. 2015;84:918–926 [PubMed: 25653287]
17. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–824 [PubMed: 22706312]
18. Traylor M, Malik R, Nalls MA, Cotlarciuc I, Radmanesh F, Thorleifsson G, Hanscombe KB, Langefeld C, Saleheen D, Rost NS, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. Ann Neurol. 2017;81:383–394 [PubMed: 27997041]
19. Traylor M, Tozer DJ, Croall ID, Lisiecka Ford DM, Olorunda AO, Boncoraglio G, Dichgans M, Lemmens R, Rosand J, Rost NS, et al. Genetic variation in plekhg1 is associated with white matter hyperintensities (n = 11,226). Neurology. 2019;92:e749–e757 [PubMed: 30659137]
20. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, Anderson CD, Brouwers HB, Valant V, Battey TW, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. Am J Hum Genet. 2014;94:511–521 [PubMed: 24656865]
21. Zhang Z, Xu G, Liu D, Fan X, Zhu W, Liu X. Angiotensin-converting enzyme insertion/deletion polymorphism contributes to ischemic stroke risk: A meta-analysis of 50 case-control studies. PLoS One. 2012;7:e46495
22. Wu L, Shen Y, Liu X, Ma X, Xi B, Mi J, Lindpaintner K, Tan X, Wang X. The 1425g/a snp in prkch is associated with ischemic stroke and cerebral hemorrhage in a chinese population. Stroke. 2009;40:2973–2976 [PubMed: 19520989]
23. Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, et al. A nonsynonymous snp in prkch (protein kinase c eta) increases the risk of cerebral infarction. Nat Genet. 2007;39:212–217 [PubMed: 17206144]

A

Baseline Characteristics	Value
No. of patients with MRI	1324
No. of MyCode patients with MRI	946
Demographics	
Female(%)	50.4%(477)
AGE, y	68.04(13.01)
BMI 25above	77.20%
BMI 30above	44.60%
Drinking (%)	34.1%(323/902)
Never Smoker	39.6%(375/930)
Comorbidity	
Anemia	14.6%(138)
Dementia	2%(19)
Atrial Fib	20.3%(192)
COPD	11.7%(111)
Coronary artery disease	29.8%(282)
Diabetes mellitus type 2	39.2%(371)
Dyslipidemia	71.8%(679)
Hypertension	79.2%(749)
Peripheral vascular disease	10.9%(103)
Procoagulant disorders	3.7%(35)
Sleep Apnea	11.2%(106)
Carotid stenosis	37.9%(359)
Parkinson	1.4%(13)
White Matter Hyperintensity	
WMH	5.33(6.14)
Ln WMH	1.44(0.90)
NIHSS	
NIHS Admission	4.22(5.38)
NIHSS 7above	18.9%(179/907)
NIHSS 10above	11.4%(108/907)
NIHSS 16above	5.4%(51/907)

B

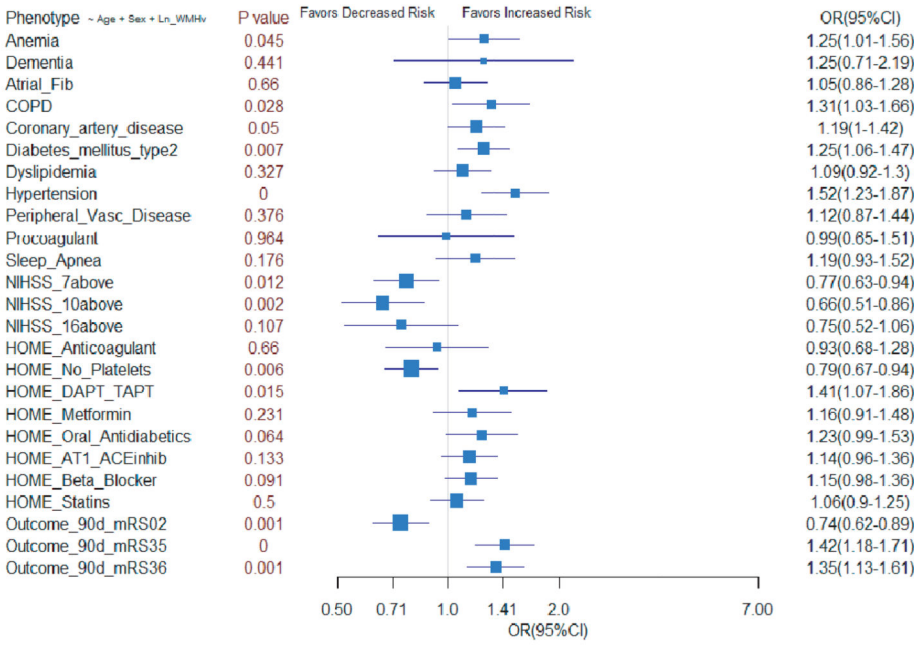

Figure 1.

Associations of WMHv with clinical variables of the Geisinger cohort. **1A.** Demographics and clinical characteristics of the Geisinger cohort. **1B.** The forest plots demonstrated the summary statistics of the associations between log transformed WMHv and demographic/clinical variables. A logistic regression model for the natural log transformed WMHv, adjusted for covariates (subphenotype ~ Age + Sex + $PC_{(1–5)}$ + Ln_WMHv), was considered.

Table 1.

Replication of candidate loci from the previous candidate-gene studies association with WMH (binary) in an extreme defined subset.

Candidate_SNPs	dbSNP	CHR	BP	A1/A2	A1_Freq	TEST	NMISS	OR(95%CI)	P	gene	feature
MTHFR677	**rs1801133**	1	11856378	T/C	0.346	ADD	478	1.47(1.06–2.02)	**0.019**	MTHFR	missense
						DOM		1.61(1.05–2.48)	**0.030**		
						REC		1.72(0.87–3.39)	0.120		
Angiotensinogen Met235Thr	rs699	1	230845794	G/A	0.456	ADD	478	0.85(0.63–1.16)	0.319	AGT	missense
						DOM		0.71(0.44–1.15)	0.163		
						REC		0.96(0.57–1.61)	0.867		
Angiotensin II receptor 1 A1166C	rs5186	3	148459988	C/A	0.311	ADD	476	1.22(0.87–1.7)	0.243	AGTR1	utr-3
						DOM		1.19(0.78–1.82)	0.425		
						REC		1.62(0.75–3.46)	0.217		
Paraoxonase 1 L55M	rs854560	7	94946084	T/A	0.362	ADD	478	0.86(0.62–1.18)	0.336	PON1	missense
						DOM		0.96(0.62–1.47)	0.844		
						REC		0.57(0.3–1.1)	0.095		
CYP11B2T(−344)C	rs1799998	8	143999600	G/A	0.458	ADD	478	1.04(0.76–1.41)	0.809	CYP11B2	near-gene-5
						DOM		1.32(0.82–2.1)	0.251		
						REC		0.78(0.46–1.34)	0.371		

Summary statistics of candidate SNPs from the previous candidate-gene studies at three modes of inheritance (additive, recessive, and dominant) for the corresponding minor alleles in a logistic regression model. The binary trait representing subgroups of patients with high or low WMHv (top and bottom quantiles of WMHv) was defined as an extreme subset.

Table 2.

Replication of candidate loci from the previous large GWAS association with WMH (binary) in an extreme defined subset.

SNP	CHR	BP (HG19)	A1/A2	A1_MAF	NMISS	OR(95%CI)	P	gene	feature
rs2984613	1	156197380	T/C	0.342	475	0.92(0.66–1.26)	0.589	PMF1	Intron
rs11679640	2	43141485	G/C	0.192	478	1.21(0.84–1.76)	0.311	~ZFP36L2	Intergenic
rs78857879	2	56135099	A/G	0.087	476	1.2(0.72–2)	0.482		Intergenic
rs72934505	2	203916487	G/T	0.113	478	0.81(0.51–1.31)	0.397		Intergenic
rs67827860	5	82860485	T/C	0.189	473	1.08(0.74–1.58)	0.702		Intergenic
rs13164785	5	82861400	G/T	0.189	473	1.08(0.74–1.58)	0.702	VCAN	Intron
rs12204590	6	1337393	A/T	0.214	477	0.67(0.46–0.98)	0.037	~FOXF2	Intergenic
rs275350	6	151016058	C/G	0.396	474	1.15(0.84–1.57)	0.373	PLEKHG1	Intron
rs10744777	12	112233018	C/T	0.322	477	0.73(0.53–1)	0.052	ALDH2	Intron
rs9515201	13	111040798	A/C	0.318	470	1.78(1.27–2.49)	7.737E-04	COL4A2	Intron
rs941898	14	100599437	G/T	0.256	477	1.25(0.89–1.75)	0.204	EVL	Intron
rs12445022	16	87575332	A/G	0.335	478	0.97(0.69–1.35)	0.841	~ZCCHC14	Intergenic
rs962888	17	43059071	A/G	0.301	433	0.91(0.64–1.3)	0.619	~C1QL1	Intergenic
rs1055129	17	73872948	G/A	0.306	478	1.19(0.85–1.67)	0.304	TRIM47	Intron
rs3744028	17	73888672	C/T	0.178	476	1.53(1.01–2.34)	0.047	TRIM65	Intron

Summary statistics of candidate SNPs from the previous large GWAS at additive mode of inheritance for the corresponding minor alleles in a logistic regression model.