Stroke

CLINICAL AND POPULATION SCIENCES

Whole-Exome Sequencing Analyses Support a Role of Vitamin D Metabolism in Ischemic Stroke

Yuhan Xie, MPhil*; Julián N. Acosta, MD*; Yixuan Ye, PhD; Zachariah S. Demarais, BS; Carolyn J. Conlon, MS; Ming Chen, PhD; Hongyu Zhao, PhD; Guido J. Falcone, MD, ScD, MPH

BACKGROUND: Ischemic stroke (IS) is a highly heritable trait, and genome-wide association studies have identified several commonly occurring susceptibility risk loci for this condition. However, there are limited data on the contribution of rare genetic variation to IS.

METHODS: We conducted an exome-wide study using whole-exome sequencing data from 152 058 UK Biobank participants, including 1777 IS cases. We performed single-variant analyses for rare variants and gene-based analyses for loss-of-function and deleterious missense rare variants. We validated these results through (1) gene-based testing using summary statistics from MEGASTROKE—a genome-wide association study of IS that included 67 162 IS cases and 454 450 controls, (2) gene-based testing using individual-level data from 1706 IS survivors, including 142 recurrent IS cases, enrolled in the VISP trial (Vitamin Intervention for Stroke Prevention); and (3) gene-based testing against neuroimaging phenotypes related to cerebrovascular disease using summary-level data from 42 310 UK Biobank participants with available magnetic resonance imaging data.

RESULTS: In single-variant association analyses, none of the evaluated variants were associated with IS at genome-wide significance levels ($P<5\times10^{-8}$). In the gene-based analysis focused on loss-of-function and deleterious missense variants, rare genetic variation at *CYP2R1* was significantly associated with IS risk ($P=2.6\times10^{-6}$), exceeding the Bonferroni-corrected threshold for 16 074 tests ($P<3.1\times10^{-6}$). Validations analyses indicated that *CYP2R1* was associated with IS risk in MEGASTROKE (gene-based test, $P=0.003$), with IS recurrence in the VISP trial (gene-based test, $P=0.001$) and with neuroimaging traits (white matter hyperintensity, mean diffusivity, and fractional anisotropy) in the UK Biobank neuroimaging study (all gene-based tests, $P<0.05$).

CONCLUSIONS: Because *CYP2R1* plays an important role in vitamin D metabolism and existing observational evidence suggests an association between vitamin D levels and cerebrovascular disease, our results support a role of this pathway in the occurrence of IS.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: genomics ■ ischemic stroke ■ vitamin D ■ whole exome sequencing

Ischemic stroke (IS) is one of the leading causes of death and disability worldwide.[1] In addition, mounting evidence indicates that IS significantly contributes to cognitive decline and dementia.[2,3] Despite significant advances in stroke prevention and treatment, a substantial proportion of patients with IS are not eligible for reperfusion therapies.[4] Importantly, the overall impact of IS seems to be significantly modified by minority status[5–8] and is much higher in low- and middle-income countries.[1,9]

Mounting evidence indicates that common genetic variation substantially contributes to the occurrence of IS. This genetic contribution influences not only risk but

Correspondence to: Guido J. Falcone, MD, ScD, MPH, Department of Neurology, Yale School of Medicine, Office 111, 100 York St, New Haven, CT 06511. Email guido.falcone@yale.edu
*Y. Xie and J.N. Acosta contributed equally.
Preprint posted on medRxiv June 2, 2022. doi: https://doi.org/10.1101/2022.05.31.22275825.
Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.122.040883.
For Sources of Funding and Disclosures, see page 807.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

25(OH)D	25-hydroxyvitamin D
Dmis	deleterious missense
GARNET	Genomics and Randomized Trials Network
GO	gene ontology
GWAS	genome-wide association study
IS	ischemic stroke
LoF	loss of function
MAF	minor allele frequency
VISP	Vitamin Intervention for Stroke Prevention
WES	whole-exome sequencing

also the severity, outcome, and recurrence of stroke.[10] The largest published genome-wide association study (GWAS) of IS to date identified 32 susceptibility risk loci that modify all stroke subtypes, including cardioembolic, small vessel, and large artery stroke.[11] In addition, follow-up studies have estimated the heritability of IS at ≈38%.[12] However, due to the lack of large studies with available sequencing data, most efforts to date have focused on common and low-frequency genetic variants, with only a few addressing the contribution of rare variants.

Leveraging the availability of whole-exome sequencing (WES) data from 200 632 participants enrolled in the UK Biobank,[13] we aimed to identify rare genetic risk factors for IS.

METHODS

Data Availability

Because of the sensitive nature of the individual-level data collected for this study, requests to access the data set from qualified researchers trained in human subject confidentiality protocols may be sent to the UK Biobank at https://www.ukbiobank.ac.uk/register-apply. The summary-level data are available from the corresponding author upon reasonable request.

Study Design

We conducted a genetic association study using a complete case approach, which included individuals with available WES and clinical data. The UK Biobank is a population-based study that enrolled 502 618 participants across the United Kingdom. Ethics approval for the study was obtained from the North West Multi-Centre Research Ethics Committee as a Research Tissue Bank approval. Participants in this study provided electronic signed consent; filled questionnaires on sociodemographic, lifestyle, and health-related factors; completed physical measures and medical tests; and provided DNA samples for future analysis. For this study, we included 200 632 study participants with available WES data, GWAS data, and clinical data under project application number 29900. Our study adheres to the Strengthening the Reporting of Observational Studies in Epidemiology reporting guideline.

DNA Sampling and Genomic Data

DNA was extracted from stored blood samples that had been collected at the time of enrollment, and genotyping was carried out by the Affymetrix Research Services Laboratory. Genome-wide array genotyping was completed using the Applied Biosystems UK BiLEVE Axiom Array, which contains 807 411 markers, for a subset of 49 960 participants, and the closely related Applied Biosystems UK Biobank Axiom Array, which contains 825 927 markers (95% of which were shared with UK BiLEVE Axiom array), for 438 427 participants. Standard quality control procedures for genome-wide data were performed centrally by the UK Biobank research team including genotype-level filters, subject-level quality checks, and imputation. The pipeline yielded a final result of 93 095 623 autosomal single nucleotide polymorphisms, short indels, and large structural variants in 487 442 individuals. A detailed description of DNA sampling, genotyping, and quality control procedures is given elsewhere.[14] The BGEN format of genome-wide genotyping data for the result were downloaded from the website of the UK Biobank.

Exomes were captured using the IDT xGen Exome Research Panel v1.0 including supplemental probes.[15] The samples were sequenced with dual-indexed 75×75-bp paired-end reads on the Illumina NovaSequence 6000 platform. The first ~50 000 tranche was sequenced with IDT v1.0 oligo lot using S2 flow cells, and the subsequent ~150 000 samples were processed with the 150 000 oligo lot using S4 flow cells. The PLINK version of WES data under the Original Quality Functional Equivalent protocol for 200 632 individuals was downloaded from the website of the UK Biobank. More information on the WES Original Quality Functional Equivalent protocol was introduced in the original release paper.[13]

Outcome Ascertainment

The present study focuses on IS—the most frequent stroke subtype. To maximize power, we included both prevalent (present upon enrollment) and incident (identified during follow-up) IS cases. IS cases were identified using a previously validated algorithm that integrates the *International Classification of Diseases* codes, self-reported data, and information from national death registries. Definitions for baseline characteristics are provided in Table S1.

Vitamin D Levels

The UK Biobank measured 36 biochemistry markers in all study participants, including serum 25-hydroxyvitamin D (25[OH]D) levels. Serum 25(OH)D levels were measured in nmol/L using the Chemiluminescence ImmunoAssay test through LIASON XL by DiaSorin, Ltd (https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/serum_biochemistry.pdf). A series of robust and detailed quality procedures were used to minimize and mitigate the effects of systematic bias and random error.[16] The phenotype data of the serum 25(OH)D level were acquired from the UK Biobank data field ID 30890.

Quality Control

We only kept variants from autosomal chromosomes in our analyses. Variants with missingness >10% and Hardy-Weinberg

equilibrium $<1\times10^{-6}$ were removed. We defined the relatedness between samples using the kinship coefficient score. For samples with kinship coefficient score >0.0442, we first removed related samples who had ≥2 relatives and iteratively removed those individuals until none remained. Then for each pair of remaining related individuals, we only removed a single sample at random. A subset of the European ancestry samples was selected based on the UK Biobank data field ID 22006. Participants who did not have IS but had other stroke subtypes were further removed from the samples.

Variant-Level Functional Annotations

We annotated the retained variants in our study cohort with minor allele frequencies (MAFs) and functional categories using the Genome Reference Consortium Human Build 38 (GRCh38). Loss-of-function (LoF) variants were those predicted to cause frameshift insertion/deletion, splice-site alteration, stop gain, and stop loss, and deleterious missense (Dmis) variants were defined as those predicted consistently to be deleterious by 9 in silico prediction tools including SIFT,[17] Polyphen2_HVAR,[18] Polyphen2_HDIV,[18] M-CAP,[18,19] MetaLR,[20] MetaSVM,[20] LRT,[21] PROVEAN,[22] and MutationTaster.[22,23]

Single-Variant Association Analysis

We conducted single-variant analysis for rare variants with MAF <0.01 using the software package REGENIE. For step 1, we used the genotype array data from the UK Biobank to fit the linear mixed model and followed the quality control steps in their documentation. For step 2, we applied the approximate firth logistic regression, used 400 as the block size, and the default settings for other parameters. Sex, age, age^2, the interaction of age and sex, and the first 4 genomic principal components were adjusted as covariates in the analyses. Genome-wide significance was set as 5×10^{-8} in our analysis.

Gene-Based Rare Variant Association Analysis

We only included rare LoF and Dmis variants with MAF <0.01 in our gene-based association test. Furthermore, LoF and Dmis variants were collapsed as the damaging variants and included in our gene-based test. We applied the robust method of the Optimal Unified Test (SKAT-O)[24] to conduct the gene-based test. Further, ultrarare variants with MAF $<1\times10^{-5}$ were removed to mitigate the potential inflation because of low minor allele counts.[25] The same set of covariates as the single-variant analysis was adjusted in the gene-based test. Exome-level significance was set as 0.05 divided by the total number of genes tested.

Pathway-Based Rare Variant Association Analysis

To evaluate whether the aggregate burden of rare variants in related pathways further increases the risk of stroke, we implemented the SKAT-O test in 2 gene sets identified in gene ontology (GO) Biological Process: vitamin D metabolic process and vitamin D3 metabolic process.[26] We used the same filters and covariates that were used for the gene-based analysis. Pathway-level significance was set as 0.05 divided by the total number of pathways in GO Biological Process.

Replication Using Individual-Level Data

We replicated observed associations using individual-level data from GARNET (Genomics and Randomized Trials Network)—a genetic substudy of the VISP clinical trial (Vitamin Intervention for Stroke Prevention). Detailed descriptions of GARNET and VISP can be found elsewhere.[27] Briefly, VISP evaluated whether high doses of folic acid, pyridoxine (vitamin B_6), and cobalamin (vitamin B_{12}), given to lower total homocysteine levels, reduce the risk of recurrent stroke over a 2-year period compared with low doses of these vitamins. GARNET genotyped a portion of the study participants enrolled in the parent clinical trial using the Illumina HumanOmni1-Quad-v1 array (Illumina, Inc). Phenotypic and genotypic data from GARNET were acquired through the database of Genotypes and Phenotypes (accession number: phs000343.v3.p1). Genome-wide data were quality controlled using standard filters, population structure was accounted for via multidimensional scaling,[28] and data were prephased and imputed to 1000 Genomes integrated reference panels (phase 3 integrated variant set release in The National Center for Biotechnology Information [NCBI] build 37).[29] A detailed preprocessing procedure is presented in Table S2. We used logistic regression to perform single-variant association tests with recurrent stroke, adjusting for the treatment group (ie, high- versus low-dose vitamin supplementation), sex, age, age^2, the interaction of age and sex, and the first 4 genomic principal components. Variant-level significance was set as 0.05 divided by the total number of variants tested. We used the GATES[30] test, a rapid gene-based association test that uses extended Simes procedure, to also conduct gene-based tests.

Replication Using Summary Statistics

We performed replication analyses using summary results from MEGASTROKE[11] (the largest GWAS of all stroke types conducted to date), a recent GWAS of small vessel IS,[31] a recent GWAS of magnetic resonance imaging markers for cerebral small vessel disease in UK Biobank,[32] and the Biobank Japan.[33,34] We downloaded the summary statistics from the Cerebrovascular Disease Knowledge Portal[35] and Biobank Japan PheWeb.[36] Summary statistics of markers within the 50 kb of our targeting gene and linkage disequilibrium structure from reference panels were set as input for the GATES test.[30] To model the linkage disequilibrium structure, we used European and East Asian panels from the 1000 Genomes Project. Single nucleotide polymorphisms with INFO Score <0.8, MAF <0.001, and deviated from the Hardy-Weinberg equilibrium ($P<1\times10^{-10}$) were removed. We only considered the overlapping markers between the GWAS summary statistics and the linkage disequilibrium reference panel in the GATES tests.

Software and Packages

For UK Biobank WES, quality control on variants was conducted with PLINK 1.90,[28] and quality control on individuals was conducted with PLINK 1.90 and R 3.5.0. Variant-level MAFs were annotated by PLINK 1.90, and functional categories of variants were annotated by ANNOVAR (ANNOtate VARiation).[37] For the gene-based association test, we applied the robust function implemented in the R package SKAT that can deal with unbalanced cases and controls.[24] Manhattan plots were generated using an adapted function from the R package qqman.[38] For GARNET, quality control was conducted with PLINK 1.90, SHAPEIT,[39] and IMPUTE2.[40] Regional GWAS results were

AQ9

visualized using LocusZoom.[41] For gene-based analysis, we applied the GATES[30] test implemented in the R package aSPU.

RESULTS

Of a total of 502 618 study participants enrolled in the UK Biobank, 200 632 underwent WES. A comparison of study participants with and without WES data was included in Table S3. Of the 200 632 participants, 16 335 were removed from the analysis when applying subject-level filters, including 16 121 for relatedness and 214 for sex mismatch or aneuploidy. Within this population of study participants who passed subject-level quality control filters, 152 710 were of European ancestry. Among the European ancestry subset, 1777 individuals developed an IS. Six hundred fifty-two individuals who had sustained hemorrhagic strokes were removed from the analysis. A total of 152 058 participants (mean age, 56.8 years; female sex, 83 141) were included in our analysis. Of a total of 17 549 650 genotyped genetic variants, 428 399 were removed from the analysis when applying variant-level filters, including 289 101 for missingness and 139 298 for the Hardy-Weinberg equilibrium (Table 1).

Single-Variant Association Analysis

In the exome-wide, single-variant association analysis, none of the evaluated genetic variants was significantly associated with the risk of IS at the genome-wide significance level ($P<5\times10^{-8}$). There was 1 noncoding variant on gene *ANK2* at the margin of significance ($P=7.9\times10^{-8}$; odds ratio, 0.02; SE, 1.84).

Gene-Based Rare Variant Association Analysis

We performed a gene-based, exome-wide association analysis using the SKAT-O robust method implemented in the SKAT software. The Manhattan plot and Quantile-Quantile plot for a total number of 16 074 genes tested across the human genome are presented in Figure 1. *CYP2R1* was the only gene significantly associated with IS ($P=2.6\times10^{-6}$, surpassing the Bonferroni-corrected threshold of 3.1×10^{-6}, adjusted for 16 074 tests). In total, there were 21 LoF and Dmis variants included in the test for *CYP2R1*. Two additional genes *CCDC74B* ($P=2.3\times10^{-5}$) and *PLOD2* ($P=6.1\times10^{-5}$) showed suggestive associations (Figure 1).

Given the central role of *CYP2R1* in the metabolism of vitamin D, we tested for association between rare genetic variation at *CYP2R1* and serum 25(OH)D levels in study participants with available measurements (139 015 participants), finding a highly significant association ($P=1.3\times10^{-10}$). Next, in this same group of patients with available vitamin D measurements, we investigated serum 25(OH)D levels in all participants, participants with any rare *CYP2R1* variants, participants with rare *CYP2R1* variants not observed in patients with stroke, and participants with rare *CYP2R1* variants observed in patients with stroke. The median concentration of serum levels of 25(OH)D in these 4 groups was 48.2, 38.0, 39.7, and 35.9 ($P<0.001$; Table 2; Figure 2). T2 F2

Pathway-Based Rare Variant Association Analysis

To evaluate whether vitamin D–related pathways containing *CYP2R1* in the GO Biological Process are also

Table 1. Baseline Characteristics of the Study Population

Variables	Overall (N=152 058)	Non-cases (n=150 281)	Cases (n=1777)
Demographics			
Age, y; mean (SD)	56.8 (8.0)	56.8 (8.0)	61.4 (6.6)
Female sex, n (%)	83 141 (54.7%)	82 475 (54.9%)	666 (37.5%)
Vascular risk factors			
Hypertension, n (%)	82 952 (54.5%)	81 573 (54.3%)	1379 (77.6%)
Hyperlipidemia, n (%)	25 120 (16.5%)	24 351 (16.2%)	769 (43.3%)
Type 2 diabetes, n (%)	8130 (5.3%)	7804 (5.2%)	326 (18.3%)
Smoking, n (%)			
Current smokers, n (%)	13 852 (9.1%)	13 592 (9.0%)	260 (14.6%)
Previous smokers, n (%)	53 746 (35.3%)	53 014 (35.3%)	732 (41.2%)
BMI, mean (SD)	27.3 (4.7)	27.3 (4.7)	28.4 (5.1)
Other cardiovascular diseases			
Coronary artery disease, n (%)	7029 (4.6%)	6733 (4.5%)	296 (16.7%)
Atrial fibrillation, n (%)	6677 (4.4%)	6216 (4.1%)	461 (25.9%)
Medications			
Antiaggregant, n (%)	21 135 (13.9%)	20 373 (13.6%)	762 (42.9%)
Statins, n (%)	25 102 (16.5%)	24 333 (16.2%)	769 (43.3%)

BMI indicates body mass index.

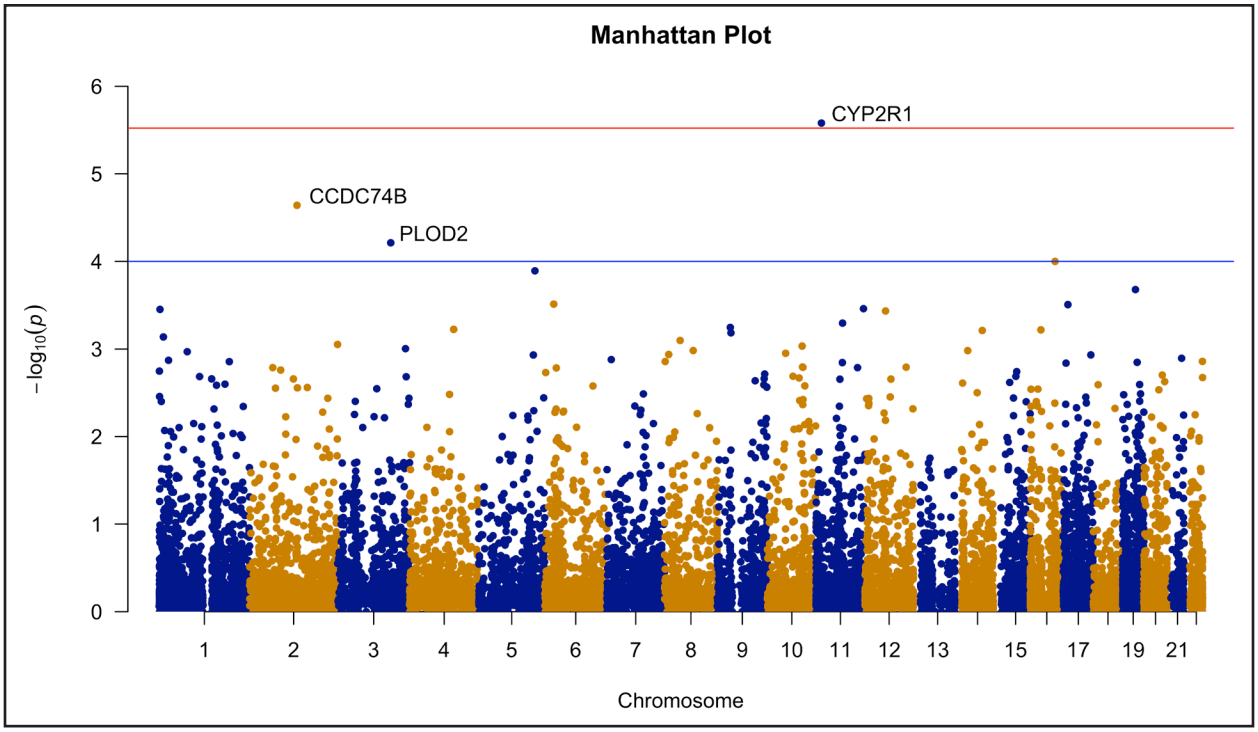

Figure 1. Manhattan plot for the exome-wide gene-based association analysis.
The genome coordinates of each variant are displayed along the *x* axis, and the negative logarithm of the association *P* values for each variant are displayed on the *y* axis. The blue line represents the threshold of 1×10^{-4}, and the red line represents the threshold of 2.5×10^{-6}. Genes that passed the thresholds in our gene-based tests are annotated with gene symbols.

associated with the risk of IS, we aggregated rare variants in the vitamin D metabolic process (GO: 0042359) and the vitamin D3 metabolic process (GO: 0070640), respectively, to conduct pathway-level association analysis using the SKAT-O robust method. For vitamin D metabolic process, there were in total 312 LoF and Dmis markers tested (P=0.40). For vitamin D3 metabolic process, there were in total 75 LoF and Dmis markers tested (P=0.006).

Replication in the GARNET Study Using Individual-Level Data

We tested for association between *CYP2R1* and recurrent stroke using both gene-based tests and single-variant analysis. When implementing gene-based testing, we replicated the association between *CYP2R1* and IS in the entire cohort (171 evaluated variants, $P=1.2\times10^{-3}$), the subgroup randomized to low-dose vitamin supplementation (161 evaluated variants, $P=6.0\times10^{-5}$), and the subgroup randomized to high-dose vitamin supplementation (166 evaluated markers, P=0.04). When implementing single-variant analyses, the makers rs149261869 ($P=5.6\times10^{-5}$) and rs117335120 ($P=9.2\times10^{-5}$) located near a recombination region were significantly associated with the risk of IS in the subgroup randomized to low-dose vitamin supplementation (Figure 3).

Replication Using Summary Statistics

We conducted gene-based testing using summary statistics from MEGASTROKE—a large GWAS of IS, separately evaluating all stroke types combined (labeled any stroke), any IS, large artery IS, cardioembolic IS, and small vessel IS. We also completed similar gene-based tests using summary statistics from a recent, large GWAS focused on small vessel IS, GWAS of neuroimaging traits related to clinically silent cerebrovascular disease (including white matter hyperintensity volume, fractional anisotropy, and mean diffusivity), and GWAS summary statistics for IS in Biobank Japan. All gene-based test *P* values except for MEGASTROKE small vessel IS (P=0.08) were significant at the significance level of 0.05 (Table 3).

DISCUSSION

In this study, we leveraged data from the UK Biobank to conduct an exome-wide association analysis of IS using both single-variant and gene-based testing. The gene-based analyses revealed that rare genetic variation at *CYP2R1* was associated with a higher risk of IS. This gene codes for the cytochrome P450 2R1—the enzyme in charge of the first step of the activation of vitamin D. We, therefore, showed that rare genetic variation at *CYP2R1* was strongly associated with

 Stroke. 2023;54:800–809. DOI: 10.1161/STROKEAHA.122.040883

Table 2. Mean and Median Vitamin D Levels in All Studied Participants and Carriers of Rare *CYP2R1* Variants

Evaluated group	No. of participants	Mean	Median
All	139015	49.7	48.2
Carriers of any rare variants	181	40.1	38.0
Carriers of variants observed in patients with stroke	54	37.8	35.9
Carriers of other rare variants	127	41.0	39.7

lower levels of circulating vitamin D. Importantly, these results were replicated using both individual-level data and summary statistics from several different studies of stroke and other phenotypes related to cerebrovascular diseases.

It has been widely demonstrated that common genetic variation contributes to stroke risk and severity.[10] The largest GWAS of IS conducted to date identified 32 susceptibility risk loci for the different types of IS that are seen in clinical practice. In addition, other published GWASs focused on functional outcome and preclinical studies using animal models also indicate that common and rare genetic variation contributes to stroke recovery as well.[42,43] Beyond IS, common genetic variation also influences intraparenchymal hemorrhage and subarachnoid hemorrhage—the most common subtypes of hemorrhagic stroke.[44,45]

Despite the compelling evidence for the role of genetic variation in stroke risk and outcome, most studies to date have focused on common genetic variation, with only a few studies investigating the role of rare variants. A recent large study from the Trans-Omics Precision Medicine program conducted a whole-genome association analysis and found 4 susceptibility loci driven by rare variants. One of these loci was specifically related to large artery IS (13q33, top associated variant rs181401679), 2 were associated with stroke of any mechanism (*RAP1GAP2*-rs60380775 and *AUTS2*-rs150022429), and the fourth was associated with hemorrhagic stroke (7q22-rs141857337).[46] Additionally, the study found 1 locus (*TEX13C*-rs145400922) associated with cardioembolic stroke at the genome-wide significance level in analysis restricted to Black participants. However, replication is still needed for these results, and the sample size represents a challenge given most of these variants are rare. Further, a whole-exome-wide analysis of white matter hyperintensities burden—a neuroimaging trait that represents clinically silent cerebrovascular disease—identified one susceptibility risk locus driven by rare genetic variation at *HTRA1*. Of note, carriers of rare mutations in this gene showed a larger effect than clinical risk factors such as hypertension, diabetes, obesity, and smoking.[47]

Our study provides significant new evidence to the field of stroke genomics research focused on rare genetic variation. In an exome-wide analysis using gene-based testing, we identified *CYP2R1* as a novel susceptibility risk locus for IS. *CYP2R1* is located at 11p15.2 and encodes the cytochrome P450 2R1—a vitamin D 25-hydroxylase involved in the first step

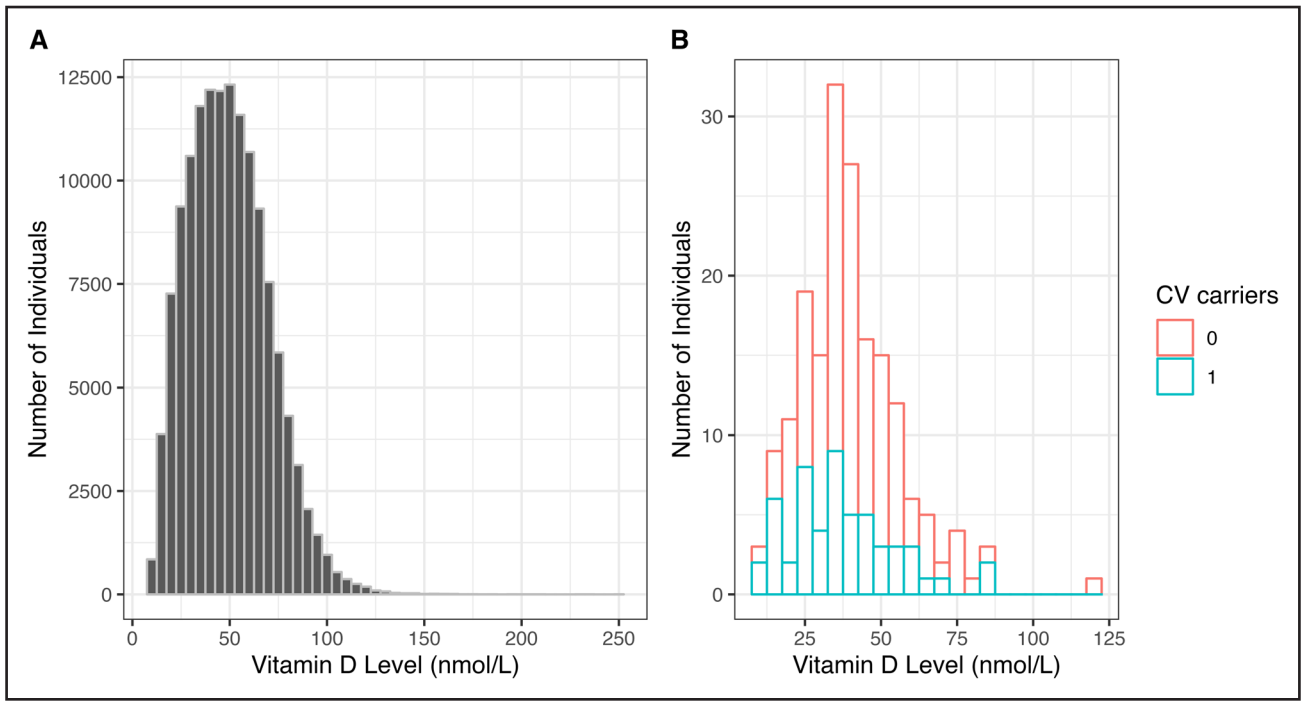

Figure 2. Vitamin D levels in all participants and *CYP2R1* carriers.
A, Histogram of serum vitamin D levels in all participants included in the analysis. **B**, Histogram of serum vitamin D levels in carriers of *CYP2R1* variants. CV, *CYP2R1* variants. Red, non-CV carriers; blue, CV carriers.

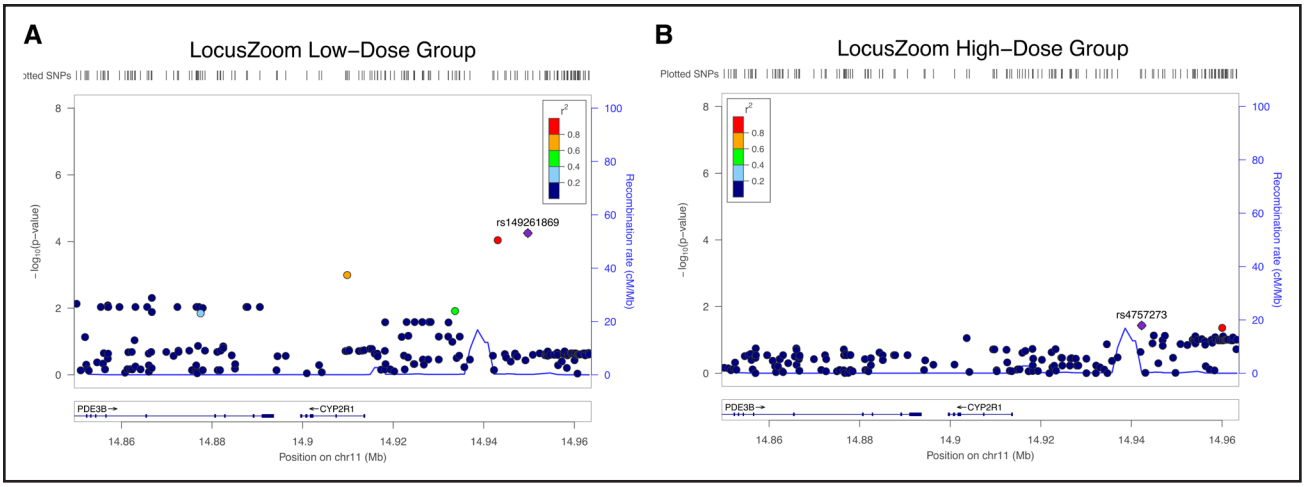

Figure 3. Locus zoom plots of the association analysis of *CYP2R1* variants in GARNET (Genomics and Randomized Trials Network) by treatment arm.

The genome coordinates of each variant are displayed along the *x* axis, and the negative logarithm of the association *P* values for each variant are displayed on the *y* axis. **A**, Low-dose group. **B**, High-dose group. IS indicates ischemic stroke; and VISP, Vitamin Intervention for Stroke Prevention.

of the activation of vitamin D.[48,49] Of note, a previous study demonstrated that one low-frequency variant in this gene leads to a 2-fold increase in risk of vitamin D insufficiency,[50] in line with our results showing that carriers of *CYP2R1* variants had, on average, lower vitamin D levels in the UK Biobank. Importantly, we pursued replication of these findings using summary statistics from several well-powered GWAS of IS (including several clinically relevant subtypes) and neuroimaging traits of clinically silent cerebrovascular disease,[51] with all but one of these analyses achieving nominal significance. Furthermore, we also replicated the association between *CYP2R1* and IS using gene-based testing and individual-level data from GARNET—a genetic substudy of the VISP clinical trial, which evaluated the role of high versus low multivitamin supplementation in stroke recurrence after a first stroke.

Many observational, experimental, and genetic studies have investigated the role of vitamin D on cardiovascular end points.[52,53] While many observational studies have suggested a relationship between lower vitamin D levels and worse cardiovascular health,[54–58] most randomized clinical trials and Mendelian randomization studies have failed to confirm a causal effect.[59–65] However, many of these studies assumed a linear relationship between vitamin D levels and disease risk, and recent studies suggest that a J-shaped relationship is more likely to accurately reflect the biological relationship between vitamin D levels and cardiovascular disease. Along these lines, a recent observational and genetic analysis of vitamin D levels and a wide range of clinical end points demonstrated a significant causal association between vitamin D and all-cause mortality in patients with vitamin D deficiency.[66] Similarly, a recent review concluded that vitamin D supplementation likely carries no benefit in people with normal levels of this vitamin but advocated for routine supplementation in people with low levels.[67]

Table 3. Results of Replication Analyses Using Summary Statistics From GWAS and Meta-Analysis From Several IS–Related Phenotypes

Study/phenotype	No. of total markers	No. of markers tested	*P* value
MEGASTROKE, any stroke (ischemic or hemorrhagic)	8 255 860	158	0.01
MEGASTROKE, any IS	8 340 184	158	3.6×10^{-3}
MEGASTROKE, large artery IS	8 451 005	162	0.003
MEGASTROKE, cardioembolic IS	8 306 090	158	3.6×10^{-3}
MEGASTROKE, small vessel IS	8 765 828	159	0.08
GWAS of small vessel IS	6 939 580	144	3.2×10^{-3}
GWAS of white matter hyperintensities	9 733 965	196	0.03
GWAS of DTI metrics fractional anisotropy	9 735 097	196	0.004
GWAS of DTI metrics mean diffusivity	9 735 314	195	0.05
GWAS of IS Biobank Japan	8 678 632	176	0.02

DTI indicates diffusion tensor imaging; GWAS, genome-wide association study; and IS, ischemic stroke.

Beyond the existing research focused on cardiovascular disease outlined above, a number of prior studies specifically evaluated the role of vitamin D in stroke.[68] Zhou et al[69] investigated the relationship between vitamin D and stroke in a meta-analysis that pooled results from 19 cohort and case/control studies, finding that lower vitamin D levels were associated with a higher risk of IS. Similarly, a single-center, prospective observational study that evaluated 349 patients with first-ever IS found that lower vitamin D levels were associated with higher risk of early stroke recurrence. In terms of possible biological mechanisms that could mediate these associations, studies using animal models found evidence for 3 possible pathways: increased risk of local thrombosis, as vitamin D appears to inhibit this process[70]; increased activity of the rennin-angiotensin system leading to elevated blood pressure, one of the most important risk factors for stroke[71,72]; and increased inflammation, another important biological state known increase stroke risk, as vitamin D seems to inhibit the production of inflammation factors, including interleukin 6 and tumor necrosis factor alpha.[73]

In future steps, additional replication of our findings in other cohorts is warranted. Specifically, analysis of WES data from individuals from other ancestries is urgently needed, as our study only included participants of the European ancestry. Further steps include the analysis of gene-environment interactions that could play a role in IS risk. In addition, we only tested the association between genes and stroke risk, but several studies have highlighted that many exposures that influence risk also modify disease severity and trajectories in patients with stroke. In that context, additional studies exploring the role of *CYP2R1* in stroke severity, outcome, and recurrence are needed. Of note, given the huge sample sizes needed to obtain enough statistical power to identify rare genetic risk factors,[74] the importance of large international consortia and collaborations is paramount.

Our study has several limitations. First, replication of our results using individual-level data from WES or Whole Genome Sequencing studies is still needed. Second, IS is a heterogeneous condition with many subtypes that share an underlying biology and risk factors but also have distinct features.[75] Unfortunately, information on IS subtypes in the UK Biobank is not available, and, therefore, these subtypes could not be analyzed separately. Finally, the ascertainment of IS in the UK Biobank is based on self-reported, electronic health records, and death reports data, which could introduce inaccuracies. However, most ascertainment was done based on electronic health records validated codes, which has shown to have a positive predictive value of 80% to 90%.[76,77]

In summary, we leveraged WES data from 152 058 participants from the UK Biobank and found that *CYP2R1*, a gene encoding the vitamin D 25-hydroxylase, a key enzyme for vitamin D activation, is associated with IS risk in people from the European ancestry. Future research to validate our findings in other cohorts, especially including participants from other ancestries, is warranted.

ARTICLE INFORMATION

Received August 10, 2022; final revision received December 13, 2022; accepted December 22, 2022.

Affiliations

Department of Biostatistics, Yale School of Public Health, New Haven, CT (Y.X., M.C., H.Z.). Department of Neurology (J.N.A., G.J.F.) and Program of Computational Biology and Bioinformatics (Y.Y., H.Z.), Yale School of Medicine, New Haven, CT. Frank H. Netter MD School of Medicine, Quinnipiac University, North Haven, CT (Z.S.D., C.J.C.).

Acknowledgments

The authors conducted the research using the UK Biobank resource under an approved data request (reference number: 29900). The authors sincerely thank the UK Biobank and its participants for their contribution to science and many genome-wide association studies (GWAS) consortia for making their GWAS summary data publicly accessible (https://www.kp4cd.org/dataset_downloads/stroke and https://pheweb.jp). Drs Zhao and Falcone jointly supervised this work.

Sources of Funding

Dr Acosta is supported by the American Heart Association Bugher Research Fellowship. Dr Falcone is supported by the National Institutes of Health (K76AG059992, R03NS112859, and P30AG021342), the American Heart Association (18IDDG34280056 and 817874), the Yale Pepper Scholar Award (P30AG021342) and the Neurocritical Care Society Research Fellowship.

Disclosures

Dr Acosta was a postdoc at Yale when this work was conducted, and he has since transitioned to RadAI. Dr Ye was a PhD student at Yale when this work was conducted, and she has since transitioned to Google. Dr Chen was a PhD student at Yale when this work was conducted, and she has since transitioned to Johnson & Johnson. The other authors report no conflicts.

Supplemental Material

Tables S1–S3

REFERENCES

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol.* 2021;20:795–820. doi: 10.1016/S1474-4422(21)00252-0
2. Mok VCT, Lam BYK, Wong A, Ko H, Markus HS, Wong LKS. Early-onset and delayed-onset poststroke dementia - revisiting the mechanisms. *Nat Rev Neurol.* 2017;13:148–159. doi: 10.1038/nrneurol.2017.16
3. Hu G-C, Chen Y-M. Post-stroke dementia: epidemiology, mechanisms and management. *Int J Gerontol.* 2017;11:210–214. doi: 10.1016/j.ijge.2017.07.004
4. Mendelson SJ, Prabhakaran S. Diagnosis and management of transient ischemic attack and acute ischemic stroke: a review. *JAMA.* 2021;325:1088–1098. doi: 10.1001/jama.2020.26867
5. Leasure AC, Acosta JN, Both C, Szejko N, Brown SC, Sheth KN, Falcone GJ. Stroke disparities among nonracial minorities in the all of US research program. *Stroke.* 2021;52:e488–e490. doi: 10.1161/STROKEAHA.121.034903
6. Trimble B, Morgenstern LB. Stroke in minorities. *Neurol Clin.* 2008;26:1177–1190. doi: 10.1016/j.ncl.2008.05.010
7. Levine DA, Duncan PW, Nguyen-Huynh MN, Ogedegbe OG. Interventions targeting racial/ethnic disparities in stroke prevention and treatment. *Stroke.* 2020;51:3425–3432. doi: 10.1161/strokeaha.120.030427
8. Leasure AC, King ZA, Torres-Lopez V, Murthy SB, Kamel H, Shoamanesh A, Salman RA-S, Rosand J, Ziai WC, Hanley DF, et al. Racial/ethnic disparities in the risk of intracerebral hemorrhage recurrence. *Neurology.* 2020;94:e314–e322. doi: 10.1212/WNL.0000000000008737
9. Lanas F, Seron P. Facing the stroke burden worldwide. *Lancet Glob Health.* 2021;9:e235–e236. doi: 10.1016/S2214-109X(20)30520-9

10. Dichgans M, Pulit SL, Rosand J. Stroke genetics: discovery, biology, and clinical applications. *Lancet Neurol*. 2019;18:587–599. doi: 10.1016/s1474-4422(19)30043-2
11. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese A-K, van der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3
12. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NLM, Jackson C, Farrall M, Rothwell PM, Sudlow C, Dichgans M, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*. 2012;43:3161–3167. doi: 10.1161/strokeaha.112.665760
13. Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, Sasson A, Wong E, Liu D, Wade Davis J, Haefliger C, et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet*. 2021;53:942–948. doi: 10.1038/s41588-021-00885-0
14. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–209. doi: 10.1038/s41586-018-0579-z
15. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, Gonzaga-Jauregui C, Khalid S, Ye B, Banerjee N, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*. 2020;586:749–756. doi: 10.1038/s41586-020-2853-0
16. Biomarker assay quality procedures: approaches used to minimise systematic and random errors (and the wider epidemiological implications) [Internet]. UK Biobank; 2019. Accessed December 15, 2021. https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/biomarker_issues.pdf.
17. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11:1–9. doi: 10.1038/nprot.2015.123
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249. doi: 10.1038/nmeth0410-248
19. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet*. 2016;48:1581–1586. doi: 10.1038/ng.3703
20. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24:2125–2137. doi: 10.1093/hmg/ddu733
21. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009;19:1553–1561. doi: 10.1101/gr.092619.109
22. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7:e46688. doi: 10.1371/journal.pone.0046688
23. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7:575–576. doi: 10.1038/nmeth0810-575
24. Zhao Z, Bi W, Zhou W, VandeHaar P, Fritsche LG, Lee S. UK Biobank whole-exome sequence binary phenome analysis with robust region-based rare-variant test. *Am J Hum Genet*. 2020;106:3–12. doi: 10.1016/j.ajhg.2019.11.012
25. Zhou W, Bi W, Zhao Z, Dey KK, Jagadeesh KA, Karczewski KJ, Daly MJ, Neale BM, Lee S. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet*. 2022;54:1466–1469. doi: 10.1038/s41588-022-01178-w
26. Gene Ontology Consortium. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res*. 2021;49:D325–D334. doi: 10.1093/nar/gkaa1113
27. Williams SR, Hsu F-C, Keene KL, Chen W-M, Nelson S, Southerland AM, Madden EB, Coull B, Gogarten SM, Furie KL, et al. Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. *Neurology*. 2016;86:351–359. doi: 10.1212/WNL.0000000000002319
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795
29. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA; 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–1073. doi: 10.1038/nature09534
30. Li M-X, Gui H-S, Kwan JSH, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet*. 2011;88:283–293. doi: 10.1016/j.ajhg.2011.01.019
31. Traylor M, Persyn E, Tomppo L, Klasson S, Abedi V, Bakker MK, Torres N, Li L, Bell S, Rutten-Jacobs L, et al. Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol*. 2021;20:351–361. doi: 10.1016/s1474-4422(21)00031-4
32. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun*. 2020;11:2175. doi: 10.1038/s41467-020-15932-3
33. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, Ninomiya T, Tamakoshi A, Yamagata Z, Mushiroda T, et al. Overview of the BioBank Japan Project: study design and profile. *J Epidemiol*. 2017;27:S2–S8. doi: 10.1016/j.je.2016.12.005
34. Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue S, Matoba N, Low S-K, Okada Y, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*. 2020;52:669–679. doi: 10.1038/s41588-020-0640-3
35. Crawford KM, Gallego-Fabrega C, Kourkoulis C, Miyares L, Marini S, Flannick J, Burtt NP, von Grotthuss M, Alexander B, Costanzo MC, et al. Cerebrovascular disease knowledge portal: an open-access data resource to accelerate genomic discoveries in stroke. *Stroke*. 2018;49:470–475. doi: 10.1161/STROKEAHA.117.018922
36. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53:1415–1424. doi: 10.1038/s41588-021-00931-x
37. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR [Internet]. *Nat Protocols*. 2015;10:1556–1566. doi: 10.1038/nprot.2015.105
38. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *J Open Source Softw*. 2018;3:731.
39. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9:179–181. doi: 10.1038/nmeth.1785
40. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529. doi: 10.1371/journal.pgen.1000529
41. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–2337. doi: 10.1093/bioinformatics/btq419
42. Söderholm M, Pedersen A, Lorentzen E, Stanne TM, Bevan S, Olsson M, Cole JW, Fernandez-Cadenas I, Hankey GJ, Jimenez-Conde J, et al. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology*. 2019;92:e1271–e1283. doi: 10.1212/wnl.0000000000007138
43. Mola-Caminal M, Carrera C, Soriano-Tárraga C, Giralt-Steinhauer E, Díaz-Navarro RM, Tur S, Jiménez C, Medina-Dols A, Cullell N, Torres-Aguila NP, et al. PATJ low frequency variants are associated with worse ischemic stroke functional outcome. *Circ Res*. 2019;124:114–120. doi: 10.1161/circresaha.118.313533
44. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, Anderson CD, Brouwers HB, Valant V, Battey TWK, 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. doi: 10.1016/j.ajhg.2014.02.012
45. Bakker MK, van der Spek RAA, van Rheenen W, Morel S, Bourcier R, Hostettler IC, Alg VS, van Eijk KR, Koido M, Akiyama M, et al. Genome-wide association study of intracranial aneurysms identifies 17 risk loci and genetic overlap with clinical risk factors. *Nat Genet*. 2020;52:1303–1313. doi: 10.1038/s41588-020-00725-7
46. Hu Y, Haessler JW, Manansala R, Wiggins KL, Moscati A, Beiser A, Heard-Costa NL, Sarnowski C, Raffield LM, Chung J, et al. Whole-genome sequencing association analyses of stroke and its subtypes in ancestrally diverse populations from trans-omics for precision medicine project. *Stroke*. 2021;53:875–885. doi: 10.1161/STROKEAHA.120.031792
47. Malik R, Beaufort N, Frerich S, Gesierich B, Georgakis MK, Rannikmäe K, Ferguson AC, Haffner C, Traylor M, Ehrmann M, et al. Whole-exome sequencing reveals a role of HTRA1 and EGFL8 in brain white matter hyperintensities. *Brain*. 2021;144:2670–2682. doi: 10.1093/brain/awab253
48. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci USA*. 2004;101:7711–7715. doi: 10.1073/pnas.0402490101

49. Bouillon R, Bikle D. Vitamin D metabolism revised: fall of dogmas. *J Bone Miner Res*. 2019;34:1985–1992. doi: 10.1002/jbmr.3884
50. Manousaki D, Dudding T, Haworth S, Hsu Y-H, Liu C-T, Medina-Gómez C, Voortman T, van der Velde N, Melhus H, Robinson-Cohen C, et al. Low-frequency synonymous coding variation in CYP2R1 has large effects on vitamin D levels and risk of multiple sclerosis. *Am J Hum Genet*. 2017;101:227–238. doi: 10.1016/j.ajhg.2017.06.014
51. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ*. 2010;341:c3666–c3666. doi: 10.1136/bmj.c3666
52. Chowdhury R, Kunutsor S, Vitezova A, Oliver-Williams C, Chowdhury S, Kiefte-de-Jong JC, Khan H, Baena CP, Prabhakaran D, Hoshen MB, et al. Vitamin D and risk of cause specific death: systematic review and meta-analysis of observational cohort and randomised intervention studies. *BMJ*. 2014;348:g1903. doi: 10.1136/bmj.g1903
53. Bouillon R, Marcocci C, Carmeliet G, Bikle D, White JH, Dawson-Hughes B, Lips P, Munns CF, Lazaretti-Castro M, Giustina A, et al. Skeletal and extraskeletal actions of vitamin D: current evidence and outstanding questions. *Endocr Rev*. 2019;40:1109–1151. doi: 10.1210/er.2018-00126
54. Wang L, Song Y, Manson JE, Pilz S, März W, Michaëlsson K, Lundqvist A, Jassal SK, Barrett-Connor E, Zhang C, et al. Circulating 25-hydroxy-vitamin D and risk of cardiovascular disease: a meta-analysis of prospective studies. *Circ Cardiovasc Qual Outcomes*. 2012;5:819–829. doi: 10.1161/CIRCOUTCOMES.112.967604
55. Judd SE, Tangpricha V. Vitamin D deficiency and risk for cardiovascular disease. *Am J Med Sci*. 2009;338:40–44. doi: 10.1097/maj.0b013e3181aaee91
56. Kim DH, Sabour S, Sagar UN, Adams S, Whellan DJ. Prevalence of hypovitaminosis D in cardiovascular diseases (from the National Health and Nutrition Examination Survey 2001 to 2004). *Am J Cardiol*. 2008;102:1540–1544. doi: 10.1016/j.amjcard.2008.06.067
57. Kendrick J, Targher G, Smits G, Chonchol M. 25-hydroxyvitamin D deficiency is independently associated with cardiovascular disease in the Third National Health and Nutrition Examination Survey. *Atherosclerosis*. 2009;205:255–260. doi: 10.1016/j.atherosclerosis.2008.10.033
58. Berghout BP, Fani L, Heshmatollah A, Koudstaal PJ, Ikram MA, Zillikens MC, Ikram MK. Vitamin D status and risk of stroke: the Rotterdam study. *Stroke*. 2019;50:2293–2298. doi: 10.1161/STROKEAHA.119.025449
59. Manson JE, Cook NR, Lee I-M, Christen W, Bassuk SS, Mora S, Gibson H, Gordon D, Copeland T, D'Agostino D, et al. Vitamin D supplements and prevention of cancer and cardiovascular disease. *N Engl J Med*. 2019;380:33–44. doi: 10.1056/NEJMoa1809944
60. Scragg R, Stewart AW, Waayer D, Lawes CMM, Toop L, Sluyter J, Murphy J, Khaw K-T, Camargo CA Jr. Effect of monthly high-dose vitamin D supplementation on cardiovascular disease in the vitamin D assessment study: a randomized clinical trial. *JAMA Cardiol*. 2017;2:608–616. doi: 10.1001/jamacardio.2017.0175
61. Barbarawi M, Kheiri B, Zayed Y, Barbarawi O, Dhillon H, Swaid B, Yelangi A, Sundus S, Bachuwa G, Alkotob ML, et al. Vitamin D supplementation and cardiovascular disease risks in more than 83 000 individuals in 21 randomized clinical trials: a meta-analysis. *JAMA Cardiol*. 2019;4:765–776. doi: 10.1001/jamacardio.2019.1870
62. Manousaki D, Mokry LE, Ross S, Goltzman D, Richards JB. Mendelian randomization studies do not support a role for vitamin D in coronary artery disease. *Circ Cardiovasc Genet*. 2016;9:349–356. doi: 10.1161/circgenetics.116.001396
63. Larsson SC, Traylor M, Mishra A, Howson JMM, Michaëlsson K, Markus HS, MEGASTROKE Project of the International Stroke Genetics Consortium. Serum 25-hydroxyvitamin D concentrations and ischemic stroke and its subtypes. *Stroke*. 2018;49:2508–2511. doi: 10.1161/STROKEAHA.118.022242
64. Huang T, Afzal S, Yu C, Guo Y, Bian Z, Yang L, Millwood IY, Walters RG, Chen Y, Chen N, et al. Vitamin D and cause-specific vascular disease and mortality: a Mendelian randomisation study involving 99,012 Chinese and 106,911 European adults. *BMC Med*. 2019;17:160. doi: 10.1186/s12916-019-1401-y
65. Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, Zeng J, Wang H, Sidorenko J, Kemper KE, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat Commun*. 2020;11:1–12.
66. Sofianopoulou E, Kaptoge SK, Afzal S, Jiang T, Gill D, Gundersen TE, Bolton TR, Allara E, Arnold MG, Mason AM, et al. Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. *Lancet Diabetes Endocrinol*. 2021;9:837–846. doi: 10.1016/S2213-8587(21)00263-1
67. Bouillon R, Manousaki D, Rosen C, Trajanoska K, Rivadeneira F, Richards JB. The health effects of vitamin D supplementation: evidence from human studies. *Nat Rev Endocrinol*. 2021;18:96–110. doi: 10.1038/s41574-021-00593-z
68. Marek K, Cichoń N, Saluk-Bijak J, Bijak M, Miller E. The role of vitamin D in stroke prevention and the effects of its supplementation for post-stroke rehabilitation: a narrative review. *Nutrients*. 2022;14:2761. doi: 10.3390/nu14132761
69. Zhou R, Wang M, Huang H, Li W, Hu Y, Wu T. Lower vitamin D status is associated with an increased risk of ischemic stroke: a systematic review and meta-analysis. *Nutrients*. 2018;10:277. doi: 10.3390/nu10030277
70. Bouillon R, Carmeliet G, Verlinden L, van Etten E, Verstuyf A, Luderer HF, Lieben L, Mathieu C, Demay M. Vitamin D and human health: lessons from vitamin D receptor null mice. *Endocr Rev*. 2008;29:726–776. doi: 10.1210/er.2008-0004
71. Ohsawa M, Koyama T, Yamamoto K, Hirosawa S, Kamei S, Kamiyama R. 1alpha,25-dihydroxyvitamin D(3) and its potent synthetic analogs downregulate tissue factor and upregulate thrombomodulin expression in monocytic cells, counteracting the effects of tumor necrosis factor and oxidized LDL. *Circulation*. 2000;102:2867–2872. doi: 10.1161/01.cir.102.23.2867
72. Forman JP, Williams JS, Fisher NDL. Plasma 25-hydroxyvitamin D and regulation of the renin-angiotensin system in humans. *Hypertension*. 2010;55:1283–1288. doi: 10.1161/HYPERTENSIONAHA.109.148619
73. Zhang Y, Leung DYM, Richers BN, Liu Y, Remigio LK, Riches DW, Goleva E. Vitamin D inhibits monocyte/macrophage proinflammatory cytokine production by targeting MAPK phosphatase-1. *J Immunol*. 2012;188:2127–2135. doi: 10.4049/jimmunol.1102412
74. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet*. 2014;95:5–23. doi: 10.1016/j.ajhg.2014.06.009
75. Radu RA, Terecoasă EO, Băjenaru OA, Tiu C. Etiologic classification of ischemic stroke: where do we stand? *Clin Neurol Neurosurg*. 2017;159:93–106. doi: 10.1016/j.clineuro.2017.05.019
76. Woodfield R, Grant I, Sudlow CLM; UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group, Sudlow CLM. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK Biobank Stroke Outcomes Group. *PLoS One*. 2015;10:e0140533. doi: 10.1371/journal.pone.0140533
77. Rannikmäe K, Ngoh K, Bush K, Al-Shahi Salman R, Doubal F, Flaig R, Henshall DE, Hutchison A, Nolan J, Osborne S, et al. Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology*. 2020;95:e697–e707. doi: 10.1212/wnl.0000000000009924