Published in final edited form as:

Stroke. 2020 November ; 51(11): 3356–3360. doi:10.1161/STROKEAHA.120.031357.

Exome Array Analysis of Early-Onset Ischemic Stroke

Thomas Jaworek, BS[1], **Kathleen A. Ryan, MPH, MS**[1], **Brady J. Gaynor, MS**[1], **Patrick F. McArdle, PhD**[1], **Oscar C. Stine, PhD**[1], **Timothy D. OConnor, PhD**[1], **Haley Lopez, BS**[1], **Hugo J. Aparicio, MD, MPH**[2], **Yan Gao, MPH**[3], **Xiaochen Lin, PhD**[4], **Megan L. Groves, MS**[5], **Matthew L. Flaherty, MD**[6], **Simin Liu, MD, ScD**[4], **Qiong Yang, PhD**[2], **James Wilson, MD**[3], **Sudha Seshadri, MD**[2], **Steven J. Kittner, MD, MPH**[1,7], **Braxton D. Mitchell, PhD**[1,8], **Huichun Xu, MD, PhD**[*,1], **John W. Cole, MD, MS**[*,1,7]

[1]University of Maryland School of Medicine, Baltimore, MD 21201

[2]Boston University School of Medicine, Boston, MA 02118

[3]University of Mississippi Medical Center, Jackson, Mississippi 39216

[4]Brown University, Providence, Rhode Island 02912

[5]The University of Texas Health Science Center at Houston, Houston, Texas 77030

[6]The University of Cincinnati College of Medicine, Cincinnati, Ohio 45267

[7]Baltimore Veteran's Affairs Medical Center, Baltimore MD 21201

[8]Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD 21201

Abstract

Background and Purpose: The genetic contribution to ischemic stroke may include rare- or low-frequency variants of high-penetrance and large-effect sizes. Analyses focusing on early-onset disease, an extreme-phenotype, and on the exome, the protein-coding portion of genes, may increase the likelihood of identifying such rare functional variants. To evaluate this hypothesis, we implemented a 2-stage discovery and replication design, and then addressed whether the identified variants also associated with older-onset disease.

Methods: Discovery was performed in UMD-GEOS Study, a biracial population-based study of first-ever ischemic stroke cases 15-49 years of age (n=723) and non-stroke controls (n=726). All participants had prior GWAS and underwent Illumina exome-chip genotyping. Logistic-regression was performed to test single-variant associations with all-ischemic stroke and TOAST subtypes in European- and African-Americans. Population level results were combined using meta-analysis. Gene-based aggregation testing and meta-analysis were performed using seqMeta. Covariates included age and gender, and principal-components for population structure. Pathway analyses

Corresponding Author: John W. Cole, M.D., M.S., Maryland Stroke Center;Department of Neurology, Baltimore VA Medical Center and University of Maryland School of Medicine, 12th Floor, Bressler-Bldg 12-006, 655 West Baltimore Street, Baltimore, MD 21201-1559, Office:410-328-6483;410-706-0414, Fax:410-328-5899, jcole@som.umaryland.edu.
[*]Both authors contributed equally to this manuscript.

Disclosures: Dr Xu has a patent to U.S. Application No. 16/454,755 (based on U.S. Application No.: 15/092,599) pending and with royalties paid.

were performed across all nominally associated genes for each stroke outcome. Replication was attempted through lookups in a previously reported meta-analysis of early-onset stroke and a large-scale stroke genetics study consisting of primarily older-onset cases.

Results: Gene burden tests identified a significant association with *NAT10* in small-vessel stroke ($p=3.79x^{-6}$). Pathway analysis of the top 517 genes ($p<0.05$) from the gene-based analysis of small-vessel stroke identified several signaling and metabolism-related pathways related to neurotransmitter, neurodevelopmental notch-signaling, and lipid/glucose metabolism. While no individual SNPs reached chip-wide significance ($p<2.05E-7$), several were near, including an intronic variant in *LEXM* (rs7549251; $p=4.08x10^{-7}$) and an exonic variant in *TRAPPC11* (rs67383011; $p=5.19x10^{-6}$).

Conclusion: Exome-based analysis in the setting of early-onset stroke is a promising strategy for identifying novel genetic risk variants, loci and pathways.

Keywords

ischemic; stroke; exome; young

Introduction.

Stroke is a common medical problem worldwide with major economic impacts, however, relatively little is known about its genetic underpinnings. The genetic contribution to ischemic stroke (IS) may include rare- or low-frequency variants with high-penetrance and large-effect sizes. Analyses focusing on early-onset disease, an extreme-phenotype, and on the exome, the protein-coding portion of genes, may increase the likelihood of identifying functional variants of large-effect size. To evaluate this hypothesis, we implemented a 2-stage discovery and replication design and then addressed whether the identified variants also associated with older-onset disease.

Materials and Methods.

Data Sharing:

The aggregated-data that support the findings described in this manuscript are available from the corresponding-author and participating studies upon reasonable request as listed in the Supplementary Data. Regarding the GEOS Discovery cohort, and in order to minimize the possibility of unintentionally sharing information that can be used to re-identify private information, a subset of the data generated for this study will be made available at International Stroke Genetics Consortium's Cerebrovascular Disease Knowledge Portal.[1] Further, regarding replication cohorts, each study can be contacted to attain their data individually, and for NIH-funded studies data is available via the database of Genotypes and Phenotypes (dbGaP).[2]

Discovery population:

The University of Maryland's Genetics of Early Onset Stroke (UMD-GEOS) Study is a population-based case-control study of men and women aged 18-49 primarily of European-American (EA) and African-American (AA) ancestry, and has previously been described.[3]

Cases were subtyped by TOAST,[4] with all-subjects genotyped on the Illumina-Human-Exome-Bead-Chip-v1.2 (see Supplementary Methods and Supplementary Table I). Ethics approval was obtained from the UMAB Institutional Review Board and written informed consent was obtained from all patients.

Single Variant Analyses:

Logistic-regression was performed using PLINK to test the association between each genotyped variant and all-stroke and TOAST-stroke-subtypes as outcomes in additive models. Covariates included five principal-components to adjust for population-structure, as well as age and sex. Population-strata results from GEOS-EA and GEOS-AA were combined with meta-analysis implementing fixed- and random-effects models.

Gene-based Aggregation Analysis:

Gene-based burden-testing implemented the seqMeta-R-package[5] to test for the association between all-stroke and TOAST-subtypes as outcomes with each gene. Covariates included age, sex, and five population-specific principal-components. Population-strata results from GEOS-EA and GEOS-AA were combined with meta-analysis. Ancestry-specific results were meta-analyzed using seqMeta. Only genes with two or more SNPs were included and with these genes further filtered to include those with a cumulative minor-allele-count >/=20 across all SNPs.

Pathway Analyses:

Network-based analysis was performed using the Ingenuity-Pathway-Analysis (IPA) tool.[6] Gene lists were used as input files for the IPA-tool inclusive of the seqMeta gene-based-aggregation meta-analyses results (combined AA and EA) that met a threshold of $p<0.05$.

Early-onset replication and extension to older-onset stroke:

Replication lookups of the top-associated SNPs and genes identified in UMD-GEOS was sought in other datasets including: 1) an independent set of early-onset-stroke studies from the Genetics-of-Early-Onset-Stroke-Consortium as previously described by Cheng *et al.*[3] ; 2) a large-scale exome-wide-association-study of primarily older-onset-stroke, MEGASTROKE.[7] (For further details see Supplementary Methods).

Results.

Characteristics of the GEOS-discovery-population are provided in Table 1 and Supplementary Table I. After exclusions, UMD-GEOS included 393 cases (mean-age stroke-onset: 41.4 years) and 428 controls of EA-ancestry, and 330 cases (mean-age stroke-onset: 42.5 years) and 298 controls of AA-ancestry.

Single Variant Analyses:

No SNP reached exome-wide-significance ($p<2.05E-7$) in the ethnicity-stratified single-variant analysis for all-stroke (genomic-inflation-measure-lambda=1.01; Supplementary Table II and Supplementary Figure I) or for TOAST-subtypes. Further, there was no overlap among the most strongly associated SNPs. Our GEOS meta-analysis combining both

ethnicities demonstrated *LEXM* rs7549251 was near chip-wide significance, as were missense-variants in *TRAPC11* and *VWDE* (Table 2 and Supplementary Table III).

Gene-Burden Analysis Results:

We performed gene-burden test-analysis inclusive of all genes for the GEOS-EA and GEOS-AA separately, and then combined results using meta-analysis. In the combined-analyses we observed a statistically-significant-association between *NAT10* and small-vessel stroke (Table 3) with 10,518 genes tested (p=3.79E-06; exome-wide p-value threshold<4.75E-06). The gene-burden results demonstrated no other significant-associations for all-stroke or its subtypes.

Pathway Analyses:

Given our findings, we then performed Pathway-Analyses on our UMD-GEOS-based seqMeta meta-analyses results (combined AA and EA) in all-stroke and small-vessel stroke. The top-ten pathway analysis results for all-stroke implementing the gene-burden test results (all genes with p<0.05) are listed in Supplementary Table IV. Similarly, for our small-vessel stroke pathway-enrichment-analysis, we implemented our gene-based association findings (all-genes with p<0.05) including 517 genes associated with small-vessel stroke. These results indicated potential important roles for several metabolism and signaling pathways, including: (1) nuclear-receptor-signaling related to lipid/glucose-metabolism; (2) neurotransmitter-glutamate-receptor-signaling, and notably; (3) neurodevelopment-notch-signaling (Table 3).

Replication of Variants or Genes from Discovery:

We evaluated replication of the top-five associations from our single-variant-analysis in the *Cheng et al.*[3] early-onset-stroke GWAS results (excluding GEOS) but were unable to detect any replication in these re-analyzed meta-analysis results (p<0.05). Additionally, we performed a lookup of these SNPs in MEGASTROKE,[7] all-stroke and small-vessel summary results, and did not observe replication (p-value<0.05).

To approximately replicate our exome-wide significant gene-burden test results for small-vessel stroke identifying *NAT10*, we then looked for small-vessel replication at a p-value<0.05 in individual common-SNPs inclusive of the Refseq gene boundary (+/−3Kb) of *NAT10* in *Cheng et al.*[3] and MEGASTROKE[7] with the results detailed in Supplementary Table V. While there was little direct overlap of SNP content between the exome-chip and the GWAS datasets, we observed several *NAT10* SNPs in both replication samples at p-value<0.05. Interestingly, further contrasting these individual SNPs across the datasets demonstrated a rare *NAT10* missense SNP (rs36006049; MAF ~1%; p=0.00069) was identified in the GEOS African-American population.

Discussion.

As demonstrated in Table 2 and Supplementary Table II, the top-hits in our all-stroke single-variant-analysis and ethnicity-stratified-analyses were primarily common-variants. The top-hit in the meta-analyses was also the top-hit in African-Americans. None of the top-hits

were the same in AA or EA stratified-analyses. Among our top-most highly associated variants for all-stroke, we identified a missense-variant in *VWDE* (rs6460939 (K(AA**G**)-->N(AA**C**)), which encodes the Von-Willebrand-Factor-D and EGF-domain-containing protein. This protein plays a role in intracellular calcium ion-binding within a variety of cell-types. Notably, the association was present in both EA (OR=1.41;p=0.0007) and AA (OR=1.41; p=0.004), and was strengthened in a meta-analysis of both ethnic groups (OR=1.41;$p=8.95\times10^{-6}$). The frequency of the effect C-allele was 0.46 in EU and 0.55 in AA. Interestingly, related to its calcium-ion binding function, VDWE appears to play a role in early-heart and -neuronal structural development. Lookups of our lead SNPs in the *Cheng et al.*[3] early-onset stroke-meta-analysis (excluding GEOS samples) demonstrated no evidence of replication (at $p<0.05$).

Most notably, gene-burden-testing in the UMD-GEOS discovery population identified NAT10 as associated with the small-vessel-subtype at an exome-wide significance-level. In an effort to replicate our findings, we used summary statistics from the same Cheng et al.[3] early-onset-meta-analyses (excluding GEOS subjects) by performing lookups of SNPs within the NAT10 gene range and identified several SNPs at $p<0.05$. Further, NAT10 replication was also seen in the larger MEGASTROKE TOAST small-vessel subtype at $p<0.05$. While there is a limitation in comparing gene-level results to single-variant results, GWAS signals may tag coding- or rare-variants in or near this gene[8]. While the GEOS samples are included in MEGASTROKE, they make-up only ~0.2% of the overall sample, hence it is unlikely that our replication findings were solely-driven by GEOS.

UMD-GEOS pathway-analysis pointed to the potential important role of nuclear-receptor-signaling-related-to-lipid/glucose-metabolism, neurotransmitter-glutamate-receptor-signaling, and neurodevelopment-notch-signaling-in-vascular-development. Notably, mutations of the *Notch3* cause a hereditary-vascular-degenerative disease known as cerebral-autosomal-dominant-arteriopathy-with-subcortical-infarcts-and-leukoencephalopathy (CADASIL) which is often associated with small-vessel stroke. *NAT10* has been shown to play a role in aging-related phenotypes and laminopathies, including progeria[9]. Such phenotypes and laminopathies can result in severe heart-disease, atherosclerosis, and stroke, thereby providing support regarding our *NAT10* findings.

Despite the hypothesis that rare-variants may be more easily detectable due to their greater effect-size, our single-variant results were likely underpowered, with the small sample-size of this study being a major limitation. However, using gene-burden testing, we were able to detect a subtype-specific association with supporting evidence in another young-onset stroke cohort and a large older-onset stroke GWA-study.

Conclusion.

The all-stroke and subtype-specific single-variant analyses performed on the UMD-GEOS-cohort failed to achieved exome-wide-significance, but several SNPs were trending. Most, but not all, top-signals were ancestry-specific. Gene burden-testing in the UMD-GEOS discovery-population identified *NAT10* as associated with the small-vessel-subtype at an exome-wide-significance level, with confirmation in another young-onset stroke cohort and

a large older-onset-stroke cohort. Notably, *NAT10* has also been shown to play a role in aging related phenotypes and laminopathies. As such, the phenotypes associated with this gene support a role in early-onset-stroke risk, warranting further study. UMD-GEOS pathway-analysis pointed to the potential important role of nuclear-receptor-signaling related to lipid/glucose-metabolism, and neurotransmitter-glutamate-receptor signaling and neurodevelopment-notch-signaling for small-vessel-stroke. Overall, exome-based analyses in the setting of early-onset stroke are promising for identifying novel-genetic-risk-variants, -loci and -pathways, and warrant additional study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding: Thomas Jaworek was partially supported by a NIH/NIA Research Training in the Epidemiology of Aging Grant (T32-AG000262). Dr. Cole was partially supported by an American Heart Association-Bayer Discovery Grant (Grant-17IBDG33700328), the AHA Cardiovascular Genome-Phenome Study (Grant-15GPSPG23770000), NIH (Grants: R01-NS114045; R01-NS100178; R01-NS105150), and the US Department of Veterans Affairs. Dr. Xu was supported by the AHA (Grant-19CDA34760258). Framingham-Heart-Study investigators including Drs. Seshadri, Yang, and Aparicio were partially supported by NIH-R01-NS017950 and NHLBI/HHS-contract:75N92019D00031.

Non-standard Abbreviations and Acronyms

GWAS	Genome Wide Association Study
SNP	single nucleotide polymorphism
TOAST	Trial of Org 10172 in Acute Stroke Treatment
UMD-GEOS Study	University of Maryland-Genetics of Early-Onset Stroke Study

References

1. Stroke Genetics Consortium's Cerebrovascular Disease Knowledge Portal. http://www.cerebrovascularportal.org/ Last accessed August 5, 2020.
2. NIH database of Genotypes and Phenotypes (dbGaP). https://www.ncbi.nlm.nih.gov/gap/ Last accessed August 5, 2020.
3. Cheng Y-C, Stanne TM, Giese A-K, Ho WK, Traylor M, Amouyel P, Holliday EG, Malik R, Xu H, Kittner SJ, et al. Genome-Wide Association Analysis of Young-Onset Stroke Identifies a Locus on Chromosome 10q25 Near HABP2. Stroke. 2016;47:307–16. [PubMed: 26732560]
4. Adams HP Jr , Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993;24:35–41. [PubMed: 7678184]
5. seqMeta-R-package. https://cran.r-project.org/web/packages/seqMeta Last accessed August 5, 2020.
6. Ingenuity-Pathway-Analysis (IPA) tool. QIAGEN Inc https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis Last accessed August 5, 2020.
7. 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. Nature Genetics. 2018;50:524–37. [PubMed: 29531354]

8. Wang K, Dickson SP, Stolle CA, Krantz ID, Goldstein DB, Hakonarson H, et al. Interpretation of Association Signals and Identification of Causal Variants from Genome-wide Association Studies. Am J Hum Genet. 2010;86: 730–742. [PubMed: 20434130]

9. Balmus G, Larrieu D, Barros AC, Collins C, Abrudan M, Demir M, Geisler NJ, Lelliott CJ, White JK, Karp NA, et al. Targeting of NAT10 enhances healthspan in a mouse model of human accelerated aging syndrome. Nature Communications. 2018;9:1700.

10. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, O'Dushlaine C, Moran JL, Chambert K, Stevens C, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics. 2012;28:2543–5. [PubMed: 22843986]

Table 1.

Characteristics of UM-GEOS-Study-Cohort.

Study	Cases			Controls		
	Subject, n	Age, mean (SD)	Male, n (%)	Subject, n	Age, mean (SD)	Male, n (%)
GEOS-EA	393	41.4 (6.9)	260 (66%)	428	39.6 (6.7)	263 (61%)
GEOS-AA	330	42.5 (6.3)	192 (58%)	298	41.3 (7.0)	176 (59%)
Total	723		452 (63%)	726		439 (60%)

Table 2.

GEOS Single-Variant All-Stroke Meta-Analysis Results (Fixed-Effect Model).

CHR	BP	SNP (type)	Gene	EA	OR	EAF (AA;Eur;Ref*)	P-value
1	55304970	rs7549251 (intron)	*LEXM* (*C1orf177*)	G	1.50	0.38;0.61;0.44	4.08E-07
4	184612553	rs67383011 (missense)	*TRAPC11*	C	0.50	0.10;0.05;0.08	5.19E-06
7	12406989	rs6460939 (missense)	*VWDE*	G	1.41	0.46;0.55;0.49	8.79E-06
8	120052238	rs6993813 (intron)	*COLEC10*	T	1.41	0.47;0.25;0.41	2.60E-05
9	139111870	rs7849585 (intron)	*QSOX2*	G	0.71	0.67;0.24;0.55	5.32E-05

*Ref indicates effect-allele-frequency in gnomAD-database[4]. EA=effect-allele and EAF=effect-allele-frequency.

Table 3.

Results of Small-Vessel-Subtype Gene-Burden and Pathway-Analyses.

Top five associations in Gene-based Analyses					
Gene	**P**	**Beta**	**SE**	**SNPs per Gene**	***NAT10* SNPs included**
NAT10 *	3.79E-06	2.074	0.45	23	rs140188192, rs201730594, rs35674959, rs139800295, rs148211973, rs2957516, rs146685334, rs138988892, rs142148595, rs139767479, rs149555377, rs199661193, rs200962843, rs145242316, rs145482727, rs137942423, rs151223396, rs140934116, rs142960948, rs36006049, rs143930117, rs200149938, rs139546360
CHST5	7.90E-05	−0.650	0.17	12	
APOPT1	1.29E-04	−1.297	0.34	3	
PIKFYVE	2.35 E-04	−0.2904	0.08	31	
KDM4C	3.39 E-04	−0.338	0.09	26	

Pathway Analysis Results: Based on 517 genes from small-vessel gene-burden test results as filtered by p-value threshold <0.05.		
Signaling Pathways	**P**	**Molecules**
FXR/RXR-Activation	2.9E-03	FOXA1; SERPINF1; MLXIPL; FETUB; APOF; A1BG; PLTP; MTTP
Glutamate-Receptor-Signaling	2.3E-02	GRID2; GRM5; SLC17A2; GRIN3A
LXR/RXR-Activation	2.8E-02	NOS2; SERPINF1; MLXIPL; APOF; A1BG; PLTP
Fcγ-Receptor-mediated-Phagocytosis-in-Macrophages-and-Monocytes	3.2E-02	CSF2; PRKCQ; FYB1; DOCK1; VAV2
Antigen-Presentation-Pathway	3.5E-02	TAPBP; HLA-DRA; PSMB8
Notch-Signaling	3.5E-02	APH1B; MAML2; FURIN
CardiolipinBiosynthesisII	3.8E-02	PGS1

* Significant at exome-wide p-value-threshold<4.75E-6.