



## OPEN ACCESS

## EDITED AND REVIEWED BY

Paul Alan Nyquist,  
Johns Hopkins Medicine, United States

## \*CORRESPONDENCE

Myriam Fornage  
Myriam.Fornage@uth.tmc.edu

## SPECIALTY SECTION

This article was submitted to  
Genetics and Omics of Stroke,  
a section of the journal  
Frontiers in Stroke

RECEIVED 01 July 2022

ACCEPTED 04 August 2022

PUBLISHED 24 August 2022

## CITATION

Fornage M (2022) Grand challenges in  
stroke genomics.  
*Front. Stroke* 1:984176.  
doi: 10.3389/fstro.2022.984176

## COPYRIGHT

© 2022 Fornage. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Grand challenges in stroke genomics

Myriam Fornage<sup>1,2\*</sup>

<sup>1</sup>Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, United States, <sup>2</sup>Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, United States

## KEYWORDS

genomics, stroke, genome-wide association, genetics, omics

Stroke, a leading cause of death, disability and dementia, is a significant public health burden with enormous financial and human costs that are projected to rise over the next decades due to demographic shifts in populations around the globe (Donkor, 2018). Stroke is a complex and heterogeneous disease encompassing multiple etiological subtypes. Recent genetics and omics studies have yielded some exciting novel insights into stroke pathophysiology and the promise of new advances in stroke prevention, diagnosis, treatment, and outcome. In particular, genome-wide association studies (GWAS) have proven an essential tool to identify common variants associated with stroke, thus unveiling previously unsuspected biological mechanisms underlying stroke and stroke subtypes (Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SiGN), and the International Stroke Genetics Consortium (ISGC), 2016; NINDS Stroke Genetics Network and International Stroke Genetics Consortium, 2016; Malik et al., 2018; Keene et al., 2020). For example, the discovery of a common variant in the *HDAC9* gene as the strongest genetic risk factor for large vessel stroke identified to date has shed light on a potential mechanism related to atherosclerotic plaque vulnerability (Asare et al., 2020). With a growing number of stroke gene discoveries owing to sustained increases in sample sizes and development of high-throughput omics discovery methods, the opportunities for genomic medicine applications in stroke have never been so great. Yet, several enduring challenges in stroke genomics remain to be addressed.

## Increasing diversity in stroke gene discovery

Most stroke risk loci known to date have been identified in populations of European ancestry. With few exceptions (Keene et al., 2020; Kumar et al., 2021), populations of African, Hispanic, and South-Asian ancestry who bear the largest burden of stroke worldwide are underrepresented in current GWAS. There are both scientific and ethical arguments to support the need for greater diversity in stroke genomic studies. Diversity may be defined in terms of ancestral/genetic background, environmental exposures, and socio-cultural/ socio-demographic factors. Increasing ancestral/genetic diversity is essential for the fine-mapping of functional variants (Asimit et al., 2016; Graham et al., 2021). Moreover, while effect size estimates of identified stroke risk loci have generally shown good correlations across ancestries (Malik et al., 2018), there may exist population-specific variants with allele frequencies differing across ancestry groups. Increasing diversity in environmental influences is particularly relevant for omics approaches that may capture the effect of a variety of non-genetic factors, such as lifestyle factors or the built environment (e.g., air pollution), which play a role

in stroke etiology. Increasing socio-cultural diversity is important because social and cultural measures of identity, such as self-reported “race” or “ethnicity,” are intricately tied to historical, political and traditional contexts that often shape institutional and structural inequalities and result in health disparities (Mersha and Abebe, 2015; Williams et al., 2016). Identifying and characterizing the effect of genes on stroke risk in the context of environmental and socio-cultural factors is an essential step toward developing accurate and effective tools for precision medicine, such as polygenic risk scores and pharmacogenetic testing. For these tools to equitably benefit people from all ancestral and socio-cultural backgrounds, there needs to be vastly greater representation of global populations and communities in genetics and omics studies of stroke.

## Looking beyond common genetic variation

As next generation sequencing is steadily gaining in throughput and accuracy while simultaneously decreasing costs, well-powered whole genome sequencing (WGS) studies will provide an opportunity to examine the impact of rare and structural variation on stroke risk, which are not well captured with array-based technologies, and to expand genomic coverage for fine mapping of known stroke loci. The first WGS association analysis of stroke identified 5 novel low-frequency variants in a multi-ancestry sample of 6,833 stroke cases and 27,116 controls (66% European ancestry) from the TOPMed consortium (Hu et al., 2022). No rare variants were identified. Compared to the MEGASTROKE GWAS (Malik et al., 2018), this study afforded a greatly expanded genome coverage (40 million vs. 8 million variants tested) but its sample size remained limited, underscoring the need for larger studies. In parallel, WGS resources generated by the TOPMed program are enabling considerable improvements in genotype imputation of rare variants for GWAS, hence enhancing their power, which should facilitate novel gene discoveries (Taliun et al., 2021).

The continuous advances in high throughput biotechnologies are opening new avenues for generating high-dimensional omics data across multiple biomolecular levels, such as transcriptomics, epigenetics, proteomics, and metabolomics, from bulk tissues or at the single-cell resolution. Analyses of single omics data for biomarker and therapeutic target discovery in stroke are underway but remain plagued by small sample sizes, technical and methodological limitations, including different detection platforms (Montaner et al., 2020). Moreover, single omics analyses cannot capture the complex interactions among the diverse omics layers, which underlie the biological processes influencing health and disease. Integrating these various omics data to extract meaningful knowledge about stroke etiology and identify biomarkers and therapeutic targets remains a daunting task owing to the heterogeneity

of data types and data structures. Machine learning and deep learning approaches are likely to play a critical role in meeting the challenges of multi-omics data analysis (Li et al., 2022).

## Moving from loci to genes to mechanisms

Although many stroke risk loci have been robustly identified, our understanding of their underlying molecular mechanisms remains limited. Most GWAS discoveries map within non-coding regions of the genome, such as intronic or promoter regions and intergenic regions, which harbor a variety of regulatory sequences. A long-standing challenge in genetic studies is to characterize the functional effects of non-coding variants and assign causality to variants within a discovered GWAS locus. While prediction of variants effects is steadily improving through the application of machine and deep learning methodologies (Zhou and Troyanskaya, 2015), the complex regulatory code of the human genome, which may be cell type-specific, remains poorly understood. Hence, current functional annotations have limited performance and practical application in post-GWAS analyses.

A more recent approach is to leverage information from common genetic effects on gene regulation that can be inferred from mapping quantitative trait loci (QTL) for various molecular features. The most commonly used are QTLs associated with gene expression levels (eQTLs) because comprehensive catalogs have been made available through large consortia resources such as GTEx and eQTLGen (GTEx Consortium, 2020; Vosa et al., 2021), but QTLs for protein and metabolites levels, DNA methylation, and other molecular features are being generated. These data can prove valuable in pinpointing potential causal genes in individual GWAS loci. For example, on chromosome 1q22, *SLC25A44*, encoding a mitochondrial carrier protein, was prioritized using this strategy (Yang et al., 2019). A current limitation is the inadequate representation of a wide variety of tissue sources derived from diverse populations. The development of induced pluripotent stem cells (iPSC) that can be differentiated into specific cell types or even organoids has the potential to ease this constraint (Song et al., 2021). These model systems can be used for scalable experimental introduction of genetic variants using CRISPR/Cas9 technology (Findlay, 2021). Combined with multi-omics and high throughput imaging or other deep-phenotyping techniques, these cellular models provide important tools for in-depth measures of diverse cellular functions. Although the transformative potential of these multimodal cellular approaches to understand stroke mechanisms is likely to be profound, the biological complexity of the disease may still require studies in model organisms. To date, only two genes identified in stroke GWAS have been functionally studied in animal models

(Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SiGN), and the International Stroke Genetics Consortium (ISGC), 2016; Asare et al., 2020). Development of humanized mouse models such as that developed for Alzheimer's Disease (Espuny-Camacho et al., 2017) through transplantation in the mouse brain of human iPSC-derived cortical neurons may provide key insights for disease mechanism and therapeutic discovery.

## Translating genomic discoveries for population health and precision medicine applications

It is widely anticipated that genomic discoveries will have a transformative impact on personal and population health through implementation of tailored medical treatment and prevention strategies (Guttmacher and Collins, 2005). Although translating stroke genomics discoveries into clinical practice remains challenging, several applications show some promise including polygenic risk scores, diagnostic biomarker discovery, pharmacogenomics and drug development.

Polygenic risk scores (PRSs) are estimates of a person's genetic risk for a trait or disease reflecting the cumulative effect of risk alleles across the genome. Their potential clinical applications include early detection of individuals or populations at risk, who could be targeted for lifestyle or therapeutic interventions; refinement of diagnosis; and guidance on therapeutic modalities and patient management. Recently, a PRS for stroke was developed by combining PRSs for each of 19 stroke and stroke-related vascular risk factors into a single meta-genetic risk score ("metaGRS") (Abraham et al., 2019). In the UK Biobank, the metaGRS was as predictive as or more predictive than any single traditional risk factor, including blood pressure; and individuals in the 0.25% of the metaGRS distribution had a level of risk of ischemic stroke almost as high as that imparted by monogenic stroke (three-fold). In the Atherosclerosis Risk In Communities (ARIC) study, individuals aged 45 years with a high metaGRS had a remaining lifetime risk of stroke of 23% vs. 10% for those with a low metaGRS, corresponding to ~3 years less lived (Thomas et al., 2022). In a retrospective analysis of 5 cardiovascular clinical trials, a stroke PRS developed from 32 genetic variants identified in the MEGASTROKE GWAS was a significant and independent predictor of ischemic stroke risk but not recurrent stroke (Marston et al., 2021). In addition, among patients with atrial fibrillation, those with a high PRS were at significantly elevated risk of developing of stroke despite having a low CHA<sub>2</sub>DS<sub>2</sub>-VASc clinical score, suggesting that the PRS may help in guiding management of a subgroup patients who may benefit from anticoagulation therapy. While the translational applications of PRSs are promising, there

currently remains considerable barriers to their widespread clinical deployment and additional progress must be made to achieve an equitable deployment and use of PRSs so as not to exacerbate existing health disparities (Martin et al., 2019). In particular, improvements in PRS development and evaluation are needed. Indeed, most PRS have a much greater predictive power in populations of European ancestry than others, most notably African ancestry (Mars et al., 2022). For example, in the ARIC study, the predictive power of the stroke metaGRS was ten-fold greater in Whites than in Blacks (Thomas et al., 2022). Beyond ancestry biases, there may be other factors that influence performance of PRSs and that remain addressing, including context-dependent effects (e.g., gene-environment interactions, sex-specific effects), stroke subtype specificity, incorporation of rare variants, as well as ways for selecting the best performing PRS. As methods to develop and evaluate PRSs vary extensively, there is a need for standardization and transparency in PRS curation and reporting. To this end, the Polygenic Score Catalog provides an important open resource (Lambert et al., 2021).

Blood-based biomarkers that aid in the diagnosis of several cardiovascular diseases have been available for many years (e.g., D-dimer for venous thromboembolism; troponin for myocardial infarction). For stroke, no such biomarkers are available. Despite longstanding efforts to identify blood-based biomarkers that can distinguish between ischemic and hemorrhagic stroke subtypes, none has yet achieved sufficient sensitivity or specificity for clinical use (Kamtchum-Tatuene and Jickling, 2019). A biomarker panel including Glial fibrillary acid protein (GFAP), retinol binding protein 4 (RBP-4), N-terminal pro-B-type natriuretic peptide (NT-proBNP) developed in a small sample of patients was shown to distinguish IS from ICH with moderate accuracy (Bustamante et al., 2021) but larger multicenter studies are warranted. The advent of large omics studies, most notably proteomics studies is likely to accelerate this quest, especially when combined with powerful Mendelian Randomization (MR) techniques (Davey Smith and Hemani, 2014). A recent systematic MR screen of the circulating proteome identified the scavenger receptor class A5 (SCARA5) as a promising biomarker of cardioembolic stroke. Prediction of side-effect profiles showed that it had no predicted adverse side effects, suggesting that it may be prioritized as a new therapeutic target (Chong et al., 2019). Although MR has known limitations, this and other studies suggest that integrating multi-omics data through MR techniques will facilitate drug discovery through identification of new targets, prediction of drug effects, including adverse effects.

Indeed, drug repurposing provides opportunities for stroke genomics applications, because previously approved drugs can be more readily carried forward to clinical practice. Databases such as the DrugBank (Wishart et al., 2006) or the Therapeutic Targets Database (Wang et al., 2020) hold information on immediate biological targets of drugs that can be leveraged for translation of genetic discoveries. For example, linking findings

from the MEGASTROKE GWAS, it was shown that identified stroke risk loci were significantly enriched for genes encoding targets of alteplase, an approved thrombolytic drug, and cilostazol, an antiplatelet agent approved in Asia (Malik et al., 2018). In MR analyses using genetic proxies for antihypertensive drug targets, calcium channel blockers, but not  $\beta$ -blockers, were associated with lower risk of stroke, especially small vessel stroke and white matter hyperintensities on neuroimaging. Hence, calcium channel blockade may represent a promising strategy for preventing cerebral small vessel disease (Georgakis et al., 2020).

Pharmacogenomics, the study of the contribution of genomic variation on individual responses to drug treatment, is an active area of research in stroke. A prime example is that of clopidogrel, a commonly prescribed antiplatelet drug, which exhibits a wide variation in patient responsiveness (Angiolillo et al., 2007). Carriers of specific loss-of-function alleles in *CYP2C19* have a significantly reduced metabolism of clopidogrel into its active form and, when treated with clopidogrel, have an increased risk of major adverse cardiovascular events, especially after percutaneous coronary intervention (PCI) (Mega et al., 2010). Clinical trials conducted to date show some promises of a genotype-informed antiplatelet selection after PCI compared with a conventional clopidogrel therapy (Pereira et al., 2020; Wang et al., 2021). These include an improved efficacy (reduction in ischemic events) and reduced toxicity (reduction in bleeding complications), and a greater cost-effectiveness (Liu et al., 2020; Narasimhalu et al., 2020).

## Understanding post-stroke outcomes and VCID

While tremendous progress has been made in understanding stroke risk, the functional and cognitive outcomes after stroke are only now being tackled (Lee et al., 2021). Critical to meeting this challenge is the development of informative phenotypes and endophenotypes from the acute, subacute and chronic phases. These phenotypes must capture the dynamic processes that underlie the spectrum of injury and recovery after stroke. The change in National Institutes of Health Stroke Scale (NIHSS) during the first 24 h post stroke ( $\Delta$ NIHSS<sub>24h</sub>) has been used as a measure of evolution of neurological deficits in the acute phase. A GWAS of  $\Delta$ NIHSS<sub>24h</sub> has identified 7 genetic loci that harbor genes influencing excitotoxicity (Ibanez et al., 2022). Neuroimaging endophenotypes that capture hemorrhagic transformation or edema formation are also of great interest. Longer term outcome using the modified Rankin Scale (mRS) have also been used in GWAS studies (Mola-Caminal et al., 2019; Soderholm et al., 2019). Future progress will require large standardized and harmonized datasets. The Global Alliance of Acute and Long-Term Outcome is an initiative of the International Stroke Genetics Consortium (ISGC) that provides

recommendations and leadership in this area. Post-stroke cognitive impairment and dementia (PSCID) is an important area of opportunities for genomic studies. The DISCOVERY study (Determinants of Incident Stroke Cognitive Outcomes and vascular Effects on Recovery) a large collaborative network among 30 sites will address the critical need for understanding the factors that influence PSCID risk (Rost et al., 2021). Because injury and recovery after stroke are influenced by dynamic processes controlled by changes in gene expression, analyses of omics data that are plastic and sensitive to interventions and changes overtime are particularly well suited to identifying key genes and biological pathways for resilience and recovery after stroke.

## Conclusion

In conclusion, genomics has the potential to revolutionize stroke diagnosis, prevention, treatment, and outcome. Rapid advances in stroke genomics are fueled by investments in technologies and population resources that are brought together in large national and international programs and collaborations, such as NeuroCHARGE and the ISGC, and facilitated by the wide sharing of data through open portals, including the Cerebrovascular Disease Knowledge Portal (Crawford et al., 2018). The contribution of biobanks, including the UK Biobank is likely to accelerate progress. Novel methods that exploit multi-omics and machine learning hold great promise to discovering novel pathways for stroke risk and outcome.

## Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Abraham, G., Malik, R., Yonova-Doing, E., Salim, A., Wang, T., Danesh, J., et al. (2019). Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.* 10, 5819. doi: 10.1038/s41467-019-13848-1
- Angiolillo, D. J., Fernandez-Ortiz, A., Bernardo, E., Alfonso, F., Macaya, C., Bass, T. A., et al. (2007). Variability in individual responsiveness to clopidogrel: clinical implications, management, and future perspectives. *J. Am. Coll. Cardiol.* 49, 1505–1516. doi: 10.1016/j.jacc.2006.11.044
- Asare, Y., Campbell-James, T. A., Bokov, Y., Yu, L. L., Prestel, M., El Bounkari, O., et al. (2020). Histone deacetylase 9 activates IKK to regulate atherosclerotic plaque vulnerability. *Circ. Res.* 127, 811–823. doi: 10.1161/CIRCRESAHA.120.316743
- Asimit, J. L., Hatzikotoulas, K., McCarthy, M., Morris, A. P., and Zeggini, E. (2016). Trans-ethnic study design approaches for fine-mapping. *Eur. J. Hum. Genet.* 24, 1330–1336. doi: 10.1038/ejhg.2016.1
- Bustamante, A., Penalba, A., Orset, C., Azurmendi, L., Llombart, V., Simats, A., et al. (2021). Blood biomarkers to differentiate ischemic and hemorrhagic strokes. *Neurology* 96, e1928–e1939. doi: 10.1212/WNL.00000000000011742
- Chong, M., Sjaarda, J., Pigeyre, M., Mohammadi-Shemirani, P., Lali, R., Shoamanesh, A., et al. (2019). Novel drug targets for ischemic stroke identified through mendelian randomization analysis of the blood proteome. *Circulation* 140, 819–830. doi: 10.1161/CIRCULATIONAHA.119.040180
- Crawford, K. M., Gallego-Fabrega, C., Kourkoulis, C., Miyares, L., Marini, S., Flannick, J., et al. (2018). Cerebrovascular disease knowledge portal: an open-access data resource to accelerate genomic discoveries in stroke. *Stroke* 49, 470–475. doi: 10.1161/STROKEAHA.117.018922
- Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–98. doi: 10.1093/hmg/ddu328
- Donkor, E. S. (2018). Stroke in the 21(st) century: a snapshot of the burden, epidemiology, and quality of life. *Stroke Res. Treat* 2018, 3238165. doi: 10.9734/JEMT/2018/40292
- Espuny-Camacho, I., Arranz, A. M., Fiers, M., Snellinx, A., Ando, K., Munck, S., et al. (2017). Hallmarks of Alzheimer's Disease in stem-cell-derived human neurons transplanted into mouse brain. *Neuron* 93, 1066–1081 e1068. doi: 10.1016/j.neuron.2017.02.001
- Findlay, G. M. (2021). Linking genome variants to disease: scalable approaches to test the functional impact of human mutations. *Hum. Mol. Genet.* 30, R187–R197. doi: 10.1093/hmg/ddab219
- Georgakis, M. K., Gill, D., Webb, A. J. S., Evangelou, E., Elliott, P., Sudlow, C. L. M., et al. (2020). Genetically determined blood pressure, antihypertensive drug classes, and risk of stroke subtypes. *Neurology* 95, e353–e361. doi: 10.1212/WNL.00000000000009814
- Graham, S. E., Clarke, S. L., Wu, K. H., Kanoni, S., Zajac, G. J. M., Ramdas, S., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature* 600, 675–679. doi: 10.1038/s41586-021-04064-3
- GTEx Consortium (2020). The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. doi: 10.1126/science.aaz1776
- Guttmacher, A. E., and Collins, F. S. (2005). Realizing the promise of genomics in biomedical research. *JAMA* 294, 1399–1402. doi: 10.1001/jama.294.11.1399
- Hu, Y., Haessler, J. W., Manansala, R., Wiggins, K. L., Moscati, A., Beiser, A., et al. (2022). Whole-genome sequencing association analyses of stroke and its subtypes in ancestrally diverse populations from trans-omics for precision medicine project. *Stroke* 53, 875–885. doi: 10.1161/STROKEAHA.120.031792
- Ibanez, L., Heitsch, L., Carrera, C., Farias, F. H. G., Del Aguila, J. L., Dhar, R., et al. (2022). Multi-ancestry GWAS reveals excitotoxicity associated with outcome after ischaemic stroke. *Brain*. 145, 2394–406. doi: 10.1093/brain/awac080
- Kamtchum-Tatuene, J., and Jickling, G. C. (2019). Blood biomarkers for stroke diagnosis and management. *Neuromol. Med.* 21, 344–368. doi: 10.1007/s12017-019-08530-0
- Keene, K. L., Hyacinth, H. I., Bis, J. C., Kittner, S. J., Mitchell, B. D., Cheng, Y. C., et al. (2020). Genome-wide association study meta-analysis of stroke in 22 000 individuals of African descent identifies novel associations with stroke. *Stroke* 51, 2454–2463. doi: 10.1161/STROKEAHA.120.029123
- Kumar, A., Chauhan, G., Sharma, S., Dabla, S., Sylaja, P. N., Chaudhary, N., et al. (2021). Association of SUMOylation pathway genes with stroke in a genome-wide association study in India. *Neurology* 97, e345–e356. doi: 10.1212/WNL.00000000000012258
- Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., et al. (2021). The polygenic score catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* 53, 420–425. doi: 10.1038/s41588-021-00783-5
- Lee, J. M., Fernandez-Cadenas, I., and Lindgren, A. G. (2021). Using human genetics to understand mechanisms in ischemic stroke outcome: from early brain injury to long-term recovery. *Stroke* 52, 3013–3024. doi: 10.1161/STROKEAHA.121.032622
- Li, R., Li, L., Xu, Y., and Yang, J. (2022). Machine learning meets omics: applications and perspectives. *Brief Bioinform.* 23:bbab460. doi: 10.1093/bib/bbab460
- Liu, D., Zhao Li, Y., Wu, H., Zhou, G., Yang, J., Yang, J., et al. (2020). Efficacy and safety of genotype-guided antiplatelet therapy versus standard treatment in 4,604 patients with CAD after PCI: a meta-analysis of randomized controlled trials. *Pharmazie* 75, 651–655. doi: 10.1691/ph.2020.0755
- Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., et al. (2018). Multiethnic genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* 50, 524–537. doi: 10.1038/s41588-018-0058-3
- Mars, N., Kerminen, S., Feng, Y. A., Kanai, M., Lall, K., Thomas, L. F., et al. (2022). Genome-wide risk prediction of common diseases across ancestries in one million people. *Cell Genom.* 2. doi: 10.1016/j.xgen.2022.100118
- Marston, N. A., Patel, P. N., Kamanu, F. K., Nordio, F., Melloni, G. M., Roselli, C., et al. (2021). Clinical application of a novel genetic risk score for ischemic stroke in patients with cardiometabolic disease. *Circulation* 143, 470–478. doi: 10.1161/CIRCULATIONAHA.120.051927
- Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591. doi: 10.1038/s41588-019-0379-x
- Mega, J. L., Simon, T., Collet, J. P., Anderson, J. L., Antman, E. M., Bliden, K., et al. (2010). Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA* 304, 1821–1830. doi: 10.1001/jama.2010.1543
- Mersha, T. B., and Abebe, T. (2015). Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. *Hum. Genomics* 9, 1. doi: 10.1186/s40246-014-0023-x
- Mola-Caminal, M., Carrera, C., Soriano-Tarraga, C., Giralt-Steinhauer, E., Diaz-Navarro, R. M., Tur, S., et al. (2019). PATJ low frequency variants are associated with worse ischemic stroke functional outcome. *Circ. Res.* 124, 114–120. doi: 10.1161/CIRCRESAHA.118.313533
- Montaner, J., Ramiro, L., Simats, A., Tiedt, S., Makris, K., Jickling, G. C., et al. (2020). Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat. Rev. Neurol.* 16, 247–264. doi: 10.1038/s41582-020-0350-6
- Narasimhalu, K., Ang, Y. K., Tan, D. S. Y., De Silva, D. A., and Tan, K. B. (2020). Cost effectiveness of genotype-guided antiplatelet therapy in asian ischemic stroke patients: ticagrelor as an alternative to clopidogrel in patients with CYP2C19 loss of function mutations. *Clin. Drug Investig.* 40, 1063–1070. doi: 10.1007/s40261-020-00970-y
- Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SiGN), and the International Stroke Genetics Consortium (ISGC). (2016) Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 15, 695–707. doi: 10.1016/S1474-4422(16)00102-2
- NINDS Stroke Genetics Network and International Stroke Genetics Consortium (2016). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol.* 15, 174–184. doi: 10.1016/S1474-4422(15)00338-5
- Pereira, N. L., Farkouh, M. E., So, D., Lennon, R., Geller, N., Mathew, V., et al. (2020). Effect of genotype-guided oral P2Y12 inhibitor selection vs conventional clopidogrel therapy on ischemic outcomes after percutaneous coronary intervention: the TAILOR-PCI randomized clinical trial. *JAMA* 324, 761–771. doi: 10.1001/jama.2020.12443
- Rost, N. S., Meschia, J. F., Gottesman, R., Wruck, L., Helmer, K., Greenberg, S. M., et al. (2021). Cognitive impairment and dementia after stroke: design and rationale for the DISCOVERY study. *Stroke* 52, e499–e516. doi: 10.1161/STROKEAHA.120.031611
- Soderholm, M., Pedersen, A., Lorentzen, E., Stanne, T. M., Bevan, S., Olsson, M., et al. (2019). Genome-wide association meta-analysis of

functional outcome after ischemic stroke. *Neurology* 92, e1271–e1283. doi: 10.1212/WNL.0000000000007138

Song, G., Zhao, M., Chen, H., Zhou, X., Lenahan, C., Ou, Y., et al. (2021). The application of brain organoid technology in stroke research: challenges and prospects. *Front. Cell Neurosci.* 15, 646921. doi: 10.3389/fncel.2021.646921

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature* 590, 290–299. doi: 10.1038/s41586-021-03205-y

Thomas, E. A., Enduru, N., Tin, A., Boerwinkle, E., Griswold, M. E., Mosley, T. H., et al. (2022). Polygenic risk, midlife life's simple 7, and lifetime risk of stroke. *J. Am. Heart Assoc.* 11, e025703. doi: 10.1161/JAHA.122.025703

Vosa, U., Claringbould, A., Westra, H. J., Bonder, M. J., Deelen, P., Zeng, B., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* 53, 1300–1310. doi: 10.1038/s41588-021-00913-z

Wang, Y., Meng, X., Wang, A., Xie, X., Pan, Y., Johnston, S. C., et al. (2021). Ticagrelor versus clopidogrel in CYP2C19 loss-of-function carriers with stroke or TIA. *N. Engl. J. Med.* 385, 2520–2530. doi: 10.1056/NEJMoa2111749

Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., et al. (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* 48, D1031–D1041. doi: 10.1093/nar/gkz981

Williams, D. R., Priest, N., and Anderson, N. B. (2016). Understanding associations among race, socioeconomic status, and health: patterns and prospects. *Health Psychol.* 35, 407–411. doi: 10.1037/hea0000242

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34, D668–672. doi: 10.1093/nar/gkj067

Yang, J., Yan, B., Fan, Y., Yang, L., Zhao, B., He, X., et al. (2019). Integrative analysis of transcriptome-wide association study and gene expression profiling identifies candidate genes associated with stroke. *PeerJ* 7, e7435. doi: 10.7717/peerj.7435

Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* 12, 931–934. doi: 10.1038/nmeth.3547