scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Targeted Metabolomics in the REasons for Geographic and Racial Differences in Stroke (REGARDS) Study

Zsuzsanna Ament[1,2], Naruchorn Kijpaisalratana[1,2,3], Varun M. Bhave[4], Catharine A. Couch[5], Ana-Lucia Garcia Guarniz[1,2], Amit Patki[6], Mary Cushman[7], Suzanne E. Judd[6], M. Ryan Irvin[5] & W. Taylor Kimberly[1,2] ✉

Targeted metabolomics was conducted on plasma samples from a nested case-cohort study within the biracial REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort. This longitudinal study investigates health outcomes with a focus on stroke disparities across the United States, particularly in the Southeastern "Stroke Belt," where stroke risk and mortality are 2–4 times higher in the Black population. The REGARDS study recruited 30,239 Black and White participants aged 45 years or older. This dataset includes 2,377 baseline plasma samples collected between 2003 and 2007 from a stroke case-cohort sub-study, with 1,056 randomly selected cohort participants and 1,321 stroke cases. The resulting data provides a resource for investigating metabolic profiles and their potential implications for health outcomes and disparities.

Background & Summary

Changes in metabolite levels across the health and disease continuum can reveal biomarkers with significant implications for identifying drug targets and informing preventative intervention strategies. Metabolomics offers a valuable tool for monitoring these changes and identifying markers of disease.

A large number of studies support the importance of a metabolic basis for stroke risk and pathology. For example, research on dietary patterns and nutrient intake has identified diet related metabolic changes that can influence stroke risk. The cluster of health issues which influence metabolic processes and contribute to the development of metabolic syndrome have been strongly associated with an increased risk of stroke[1–3] including hypertension, diabetes, excess body fat around the waist, elevated triglyceride levels, and low levels of high-density lipoprotein cholesterol. Metabolomic profiling studies conducted in multiple large-scale epidemiological studies and other populations have also identified specific metabolites associated with stroke and stroke risk[4,5].

Given the evidence connecting metabolic changes to stroke risks, the objective of this study was to examine metabolite levels among participants of the REasons for Geographic and Racial Differences in Stroke (REGARDS) Study[6]. The REGARDS study is a large, biracial, longitudinal study designed to explore health outcomes and disparities, particularly stroke, across the United States. The study recruited 30,239 Black and White participants aged 45 years or older, specifically oversampling of Black Americans and persons living in the Southeastern states known as the "Stroke Belt," where stroke risk and stroke-related mortality are significantly higher[6].

[1]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [2]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. [3]Division of Neurology, Department of Medicine and Division of Academic Affairs, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand. [4]Harvard Medical School, Boston, MA, USA. [5]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. [6]Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. [7]Department of Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT, USA. ✉e-mail: wtkimberly@mgh.harvard.edu

Targeted metabolomics using liquid chromatography-tandem mass spectrometry (LC-MS/MS) was conducted in the REGARDS stroke case-cohort sub-study. Plasma samples were analyzed from 2,377 participants, comprising 1,321 stoke cases and 1,056 randomly selected cohort participants. At the time of data collection, follow-up information was available up to April 1, 2019, and stroke cases adjudicated by this date were included. Notably, the unique design of the case-cohort study allows for the inclusion of stroke cases among participants who were also selected as part of the initial random cohort sample.

The plasma samples were collected from each participant at a baseline study visit between 2003 and 2007, and the baseline metabolic profiles were assessed to identify associations with the risk of stroke. Our previous studies have identified metabolites associated with incident ischemic stroke and linked these to environmental factors and health behaviours, such as dietary habits and changes in gut microbial metabolites[7–14].

This metabolomics dataset provides a resource for facilitating further research into the metabolic basis of stroke to inform health strategies, mitigate stroke disparities, and identify therapeutic targets to reduce stroke risk.

Methods

Study design and participants. The REGARDS study was designed with the objective of determining the causes of excess stroke mortality in the Southeastern US and among Black Americans and recruited 30,239 individuals from the contiguous US between January 2003 and October 2007[6].

Recruitment. REGARDS participants were selected from commercially available lists (Genesys), and a letter and brochure were sent to inform selected participants of the study and an upcoming phone call. During the call, verbal consent was obtained, and a 45-minute questionnaire was administered. The telephone response rate was 33%, and the cooperation rate was 49%. Race was determined by participant self-report as Black or White.

Exclusion/inclusion criteria. Exclusion criteria included race other than Black or White, medical conditions preventing long-term participation, active cancer or active treatment for cancer, being a resident or waiting placement in a nursing home, or inability to communicate in English. After the survey, qualifying participants were asked if they would be interested in participating in an in-home examination. Approximately 70% of people agreed to allow a research nurse to visit their home to collect blood pressure, height, and weight measures, blood samples, and to perform an EKG. Written informed consent was obtained from all participants who completed the baseline in-home visit. Participants were considered to be enrolled in the study if they completed both the 45-minute telephone questionnaire and consented at the in-person baseline study visit.

By design, the overall study cohort included 55% female and 41% Black American participants, and 21% of subjects were enrolled from the Stroke Buckle (coastal plain region of North Carolina, South Carolina, and Georgia), 35% from the Stroke Belt states (remainder of North Carolina, South Carolina, and Georgia, plus Alabama, Mississippi, Tennessee, Arkansas, and Louisiana), and the remaining 44% of participants from the other 40 contiguous states.

Plasma sample collection. Blood samples were collected at the time of the baseline study visit[15]. All participants were asked to fast at least 8 hours before venous blood drawing in the early morning. K_2EDTA plasma sample tubes were shipped overnight to a central laboratory, and the plasma was separated, aliquoted, and stored at −80 °C until metabolomics analysis.

Outcome assessment. The REGARDS study conducts active surveillance of cohort members, and participants or their proxies were contacted every six months by telephone to ascertain stroke or stroke symptoms. Medical records, including neuroimaging and other diagnostic reports, are centrally reviewed by physicians to confirm the diagnosis, stroke type, and possible etiology. Adjudication is based on the World Health Organization's definition of stroke and/or imaging results consistent with a stroke[16]. For this analysis, stroke cases adjudicated until April 1, 2019, were included.

Ethics. The study was approved by the Institutional Review Boards of all participating institutions. Informed consent was obtained from all participants included in the study. Procedures were performed in accordance with the ethical standards of the institutional and national research committees and with the 1975 Helsinki Declaration and its later amendments. Metabolomics analysis was conducted under IRB approval by the Mass General Brigham Human Research Committee (MGBHRC, protocol number: IRB 2016P001801).

Selection of samples for metabolomics. Study samples for metabolomics were selected by including all adjudicated stroke cases and a random subset of the cohort. The random cohort was sampled based on demographic strata, including age (20% 45–54, 20% 55–64, 25% 65–74, 25% 75–84, and 10% ≥ 85), race (50% black, 50% white), and gender (50% male, 50% female)[17].

After the stratified selection process, samples were blinded and went through a freeze-thaw cycle for the purpose of sub-aliquoting and shipping for metabolomics. Targeted metabolomics data was collected for 1,321 cases and 1,056 random cohort samples.

Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS). Data collection procedures have been stored with the corresponding data on the Brain Data Science Platform repository (https://doi.org/10.60508/wt9r-9z95)[18] and are detailed below.

Analytical reagents. Acetonitrile and methanol were LC-MS grade or higher, and water was ultrapure (Type 1). Ammonium acetate and ammonium hydroxide (10%) were LC-MS grade.

Reference standards were purchased from a range of manufacturers and suppliers, including Sigma-Aldrich (St. Louis, MO), TRC Chemicals (Toronto, ON, Canada), Cayman Chemicals (Ann Arbor, MI), Thermo-Fisher Scientific (Fairlawn, NJ), Cambridge Isotope Laboratories (Tewksbury, MA, USA). A full list is provided in Supplementary Table 1. The internal solution mixture contained stable isotope-labeled proline (13C5, 15 N), glutamine (13C5, 15N2), leucine-d10, phenylalanine-d8, and valine-d8. Each compound was prepared at 15 µM final concentration in methanol: water (50:50) and stored in amber glass at nominal 4 °C.

Metabolite extraction. The sample aliquots were randomly subdivided across 32 sample batches. Each batch of samples went through a second freeze-thaw cycle immediately before metabolite extraction. All procedures requiring manual sample handling were conducted on wet ice.

Samples were extracted using protein precipitation. First, 30 µL of K_2EDTA plasma was aliquoted onto a 96-well plate. Before quenching, each sample received 30 µL of internal solution mixture, and samples were then briefly vortexed for 1 minute at 1400 rpm to ensure homogeneity. Proteins were precipitated with 110 µL of ice-cold acetonitrile: methanol (75:25), and samples were vortexed for a further 5 min at 1400 rpm. Samples were centrifuged at 3200 g for 10 min at 4 °C (Allegra X-15R Centrifuge, Beckman Coulter), and 100 µL of the clean supernatants were transferred to clean well plates for LC-MS/MS injection.

Chromatography. Metabolites were chromatographed using dual Infinity II 1290 high-performance liquid chromatography pumps with automated alternating column regeneration (ACR). The analytical columns were the XBridge Amide 2.1 × 100 mm, 3.5 µm (Waters; Part# 186004860), and the pre-columns were XBridge BEH Amide 3.5 µm VanGuard Cartridge. 2.1 × 5 mm. (Waters Part # 186007757).

Mobile phase A was water: acetonitrile (95:5) with 20 mM ammonium acetate and 20 mM ammonium hydroxide; mobile phase B was acetonitrile. Mobile phases were prepared and stored in amber glass bottles at room temperature. Mobile phase A was stored for a maximum of 7 days.

Gradient conditions. The flow rate of the analytical pump was 0.25 mL/min, and the starting conditions were 10%A and 90%B. Mobile phase B was decreased to 5%B over 6 min. At 6.1 min, %B was returned to 90 and held until 6.5 min. The regeneration pump started at 90%B with a flow rate of 0.25 mL/min, increased to 5%B over 0.1 min, and held for 1 min. Mobile phase B was increased to 90% over 0.5 min, and the flow rate was increased to 0.5 mL/min over 1 min. The column was washed under these conditions (90%B at 0.5 mL/min) for 3 minutes. For re-equilibration, the flow rate was decreased to 0.25 mL/min in 0.5 min. These conditions (90%B at 0.25 mL/min) were held for the remainder of the gradient time until 6.5 min. The runtime was 6.5 min, and the cycle time (injection start to next injection start) was 7.05 min. Each sample was injected twice to acquire data sequentially in negative and positive ionization modes. The ACR gradient table may also be accessed as a downloadable pdf document through the repository.

Metabolites were measured using an Agilent 6495 triple quadrupole mass spectrometer in dynamic multiple reaction monitoring (MRM) mode. The electrospray ionization (ESI) source settings were as follows: gas temperature: 230 °C, gas flow: 15 L/min, nebulizer pressure: 30 psi, sheath gas temperature: 400 °C, sheath gas flow: 12 L/min, capillary voltage in negative ionization mode: 2500 V, capillary voltage in positive ionization mode: 2000 V, and nozzle voltage: 1000 V. Further details of the corresponding dynamic MRM transitions with collision energies are provided within the Brain Data Science Platform repository.

Metabolomics data processing. Blinding was employed throughout all data collection and LC-MS/MS peak integration processes.

Data were processed using Agilent MassHunter QQQ Quantitative Analysis software utilizing the Agile2 integrator, Quartic/Quintic Savitzky-Golay, and the Root Mean Square (RMS) noise algorithms. Records of corresponding data tables can be accessed through the Brain Data Science Platform. For further processing, peak area values were normalized using the nearest pooled plasma normalization approach, where each analyte's peak area values were divided by the average of the corresponding nearest human pooled plasma (HPP) signals (see below).

Statistical analyses. Statistical analyses were conducted using the normalized area ratio values in Stata (v 17.).

Data Records

This study is available at the Brain Data Science Platform, https://bdsp.io/. The data can be accessed via the Project: https://doi.org/10.60508/wt9r-9z95[18]. Researchers are invited to access the data after accepting the Data Use Agreement (DUA).

Overview of the data files and their formats. Raw files were acquired as Agilent.d files. Prior to data upload, files were converted to an open-source mzML format using ProteoWizard[19,20]. Each data file was assigned a unique number reflecting the order of acquisition. Each study sample has one unique identifier and two associated data files: one acquired in negative ionization mode and one in positive ionization mode. Files with odd numbers contain data from negative ionization mode, while files with even numbers contain data from positive ionization mode. For example, files 19.mzML and 20.mzML contain information about the same sample extract for metabolites ionizing in negative and positive ionization modes, respectively. If a study sample represents an HPP (used for quality control or normalization), the assigned unique identifier includes both a number and a letter, which matches the HPP's negative ionization mode data file name. A separate tabulated Study Design table includes the unique identifiers, the corresponding data file names, and the case-cohort designation.

Outputs from the Agilent MassHunter processed analyte peak area values are provided in two separate files, one for the negative and one for the positive ionization mode compounds. Values may be normalized before further processing to account for systematic differences across each batch using the nearest pooled plasma normalization approach. Custom code to compute area ratio values from the deposited data is provided as a Stata code (v.17) and available for download from Figshare (https://doi.org/10.6084/m9.figshare.26755777)[21].

Reference standards of optimized metabolites and the resulting mzML files are provided with the dataset as an additional tool for annotation. Metabolite names and their associated reference files are listed in the 'Reference standard files' table. High-concentration solutions represent 10 µM, while low-concentration solutions represent approximately 0.5 µM. Please note that this data is not intended for quantitative analysis, and that the lower limit of detection for some analytes may fall below 0.5 µM.

Technical Validation

Quality assurance and quality control samples. Matrix-matched phenotypic QC samples are referred to as HPPs. For this study, HPP samples were prepared as unfiltered gender-unspecified human plasma in K_2EDTA in bulk and stored as multiple aliquots (550 µL) at nominal −80 °C.

Each sample batch was prepared using a 96-well plate format. A typical analytical plate contained 15 HPP samples and used one aliquot of the stored HPP vials to ensure the same number of freeze-thaw cycles. At the beginning of each analytical run, system-suitability samples, including a blank and two HPP samples, were injected to assess chromatographic performance. Additional HPP samples were used for quality control and normalization. HPP samples were included following every 10th injection for the nearest pooled plasma normalization approach. To calculate variability (%CV), additional HPP samples were included after every 20th sample. Because multiple individual HPP aliquots were prepared, %CV values represent the variability of the extraction protocols and the instrumental variability.

The intra-assay %CV had a median of 14.60% (IQR: 12.67%–19.53%) as measured across 32 batches. Values for each metabolite are included in Supplementary Table 2.

Code availability

Custom code created in Stata (v17) for nearest pooled plasma normalization and %CV calculation can be accessed from Figshare (https://doi.org/10.6084/m9.figshare.26755777)[21] with no restrictions.

Received: 11 September 2024; Accepted: 3 March 2025;
Published online: 06 March 2025

References

1. Boehme, A. K., Esenwa, C. & Elkind, M. S. V. Stroke Risk Factors, Genetics, and Prevention. *Circ Res* **120**, 472–495 (2017).
2. Bhatt, H., Safford, M. & Glasser, S. Coronary heart disease risk factors and outcomes in the twenty-first century: findings from the REasons for Geographic and Racial Differences in Stroke (REGARDS) Study. *Curr Hypertens Rep* **17**, 541 (2015).
3. Rodriguez-Colon, S. M. *et al*. Metabolic syndrome clusters and the risk of incident stroke: the atherosclerosis risk in communities (ARIC) study. *Stroke* **40**, 200–205 (2009).
4. Chumachenko, M. S., Waseem, T. V. & Fedorovich, S. V. Metabolomics and metabolites in ischemic stroke. *Rev Neurosci* **33**, 181–205 (2022).
5. Au, A. Metabolomics and Lipidomics of Ischemic Stroke. *Advances in clinical chemistry* **85**, 31–69 (2018).
6. Howard, V. J. *et al*. The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and Design. *Neuroepidemiology* **25**, 135–143 (2005).
7. Kijpaisalratana, N. *et al*. Plasma Metabolites and Life's Simple 7 in REGARDS. *Stroke* https://doi.org/10.1161/STROKEAHA.123.044714 (2024).
8. Jones, A. C. *et al*. Metabolite profiles and DNA methylation in metabolic syndrome: a two-sample, bidirectional MendHyperelian randomization. *Frontiers in Genetics* **14** (2023).
9. Kijpaisalratana, N. *et al*. Acetylglutamine Differentially Associated with First-Time Versus Recurrent Stroke. *Transl Stroke Res* https://doi.org/10.1007/s12975-023-01181-1 (2023).
10. Kijpaisalratana, N. *et al*. Trimethylamine N-Oxide and White Matter Hyperintensity Volume Among Patients With Acute Ischemic Stroke. *JAMA Netw Open* **6**, e2330446 (2023).
11. Ament, Z. *et al*. Gut microbiota-associated metabolites and risk of ischemic stroke in REGARDS. *J Cereb Blood Flow Metab* **43**, 1089–1098 (2023).
12. Kijpaisalratana, N. *et al*. Association of Circulating Metabolites With Racial Disparities in Hypertension and Stroke in the REGARDS Study. *Neurology* https://doi.org/10.1212/WNL.0000000000207264 (2023).
13. Bhave, V. M. *et al*. Plasma Metabolites Link Dietary Patterns to Stroke Risk. *Ann Neurol* **93**, 500–510 (2023).
14. Ament, Z. *et al*. Nucleosides Associated With Incident Ischemic Stroke in the REGARDS and JHS Cohorts. *Neurology* **98**, e2097–e2107 (2022).
15. Gillett, S. R. *et al*. Validating laboratory results in a national observational cohort study without field centers: the Reasons for Geographic and Racial Differences in Stroke cohort. *Clin Biochem* **47**, 243–246 (2014).
16. Stroke–1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. *Stroke* **20**, 1407–1431 (1989).
17. Zakai, N. A. *et al*. ABO Blood Type and Stroke Risk: The REasons for Geographic and Racial Differences in Stroke (REGARDS) Study. *J Thromb Haemost* **12**, 564–570 (2014).
18. Ament, Z. *et al*. Targeted Metabolomics in the REGARDS Stroke Case-Cohort Study. *Brain Data Science Platform* https://doi.org/10.60508/wt9r-9z95 (2025).
19. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536 (2008).
20. Chambers, M. C. *et al*. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918–920 (2012).
21. Ament, Z. & Kimberly, W. T. Stata Code for Metabolomics Data Analysis in the REGARDS Stroke Case-Cohort Study: Data Transformation and Quality Control. *figshare* https://figshare.com/account/projects/217678/articles/26755777?file=48609832.

Acknowledgements
This research project was supported by R01 NS099209 and cooperative agreement U01 NS041588, co-funded by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NINDS or the NIA. Representatives of the NINDS were involved in the review of the manuscript but were not directly involved in the collection, management, analysis or interpretation of the data. The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at: https://www.uab.edu/soph/regardsstudy/.

Author contributions
All authors played an instrumental role in the creation of this manuscript. Kimberly, Irvin, Cushman, and Judd designed the conceptual framework for the study. Ament and Kimberly contributed to data acquisition and analysis and drafted key sections of the manuscript. Kijpaisalratana, Bhave, Couch, Garcia Guarniz, Patki, Cushman, Judd, Irvin, and Kimberly provided valuable feedback during the revision process and critically reviewed the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04746-x.

Correspondence and requests for materials should be addressed to W.T.K.