Neuron

CellPress

NeuroView

Missing in Action: African Ancestry Brain Research

Daniel R. Weinberger,[1,4,*] Kafui Dzirasa,[2,5] and Lesia L. Crumpton-Young[3]
[1]Lieber Institute for Brain Development, Maltz Research Laboratories, Departments of Psychiatry, Neurology, Neuroscience and Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[2]Department of Psychiatry and Behavioral Sciences, Departments of Neurobiology, Biomedical Engineering, Neurosurgery, Center for Neuroengineering, Duke University Medical Center, Durham, NC 27710, USA
[3]Morgan State University, Provost and Senior Vice President for Academic Affairs, Professor of Engineering, Baltimore, MD 21251, USA
[4]Twitter: @LieberInstitute
[5]Twitter: @KafuiDzirasa
*Correspondence: drweinberger@libd.org
https://doi.org/10.1016/j.neuron.2020.07.008

Individuals of African ancestry have been starkly underrepresented in the pursuit of personalized medicine for brain illnesses. The African Ancestry Neuroscience Research Initiative will seek to generate much-needed brain gene and protein expression profiles for people of African ancestry.

SARS-CoV-2 has laid bare the disparities in health care that people of color have endured for centuries (Bailey et al., 2017). As stark as the health contrasts have been, the disparity in research focused on understanding biological variations in individuals of African ancestry (AA) are even greater. This is especially the case in neuroscience research. Because Africa was the birthplace of all modern humans, the phrase "African ancestry" used here refers to the majority of people living in sub-Sahara Africa today and those with lineage tracing to the region within the past 500 years.

The explosion of data related to genetic and environmental factors that influence disease has profoundly changed the landscape of medical research and underwritten the promise of precision medicine, but little of this promise is based on studies of individuals of AA. In research studies of brain disorders, individuals of African descent make up, on average, less than 5% of research cohorts. In large-scale genetic studies of CNS disorders, the percentage of individuals of AA is even smaller (Gurdasani et al., 2019). This is particularly troubling because AA Americans are 20% more likely to experience serious mental health problems than the general population and are arguably up to twice as likely to develop Alzheimer's disease. Suicide rates for AA children aged 5–11 are twice as high as comparable individuals of European ancestry even after controlling for socioeconomic factors (https://blog.samhsa.gov/2019/07/23/alarming-suicide-trends-in-african-american-children-an-urgent-issue). Genetic differences have also been linked to divergent responses to antipsychotics, lithium, and other CNS drugs in Black individuals compared with European ancestry individuals. While much of this is related to the impact of genetic variation on drug metabolism, targets in brain are also implicated. Clearly, the time for a brain research initiative focused on AA is long overdue.

Genetic variation traces the timeline of human migrations and explains many of the differences in traits across populations, including susceptibility and resilience to illness and to environmental exposures. It is clinical lore that among individuals of AA, mutation of the beta globin gene responsible for sickle cell anemia likely arose as protection from malaria. Unique variants in the *APOL1* gene, which are associated with increased risk of kidney disease in individuals of AA, likely were positively selected to protect against African trypanosomiasis. Variants not unique to recent differences in ancestry may still show dramatic heterogeneity of effects. So, for example, the APOe4 haplotype that is the principle risk factor for late-onset Alzheimer's disease (LOAD) is less penetrant in individuals of AA. Homozygotes for the risk haplotype are three times less likely to be at risk of LOAD as are individuals of European ancestry with the same haplotype, despite AD being potentially twice as common in AA people (Rajabli et al., 2018). Overall, individuals of AA have the most diverse genotypes and phenotypes of any population (Bentley et al., 2020).

The Missing Persons in Personalized Medicine

Genes and the interacting environment conspire to personalize human health. When President Obama launched the personalized ("precision") medicine initiative in January 2015, the idea was "an innovative approach that takes into account individual differences in people's genes, environments, and lifestyles." Personalized medicine begins with the genetic mosaic of an individual's ancestors. For more than 20 years, from the start of the dbSNP project to catalog single nucleotide polymorphisms (SNPs) in the human genome across ancestries, it has been apparent that allele frequencies at many variants differ substantially between individuals of AA and those of European ancestry. The human reference genome, the signature deliverable from the Human Genome Project, is the backbone of personalized medicine. It is the map universally used to assemble newly sequenced genomes, epigenomes, and transcriptomes and to calculate risk and predict treatment response based on genetic variation mapped to this genome. While derived from pieces of genomes of multiple volunteers, the current "build" of the human reference genome (GRCh38) is primarily (i.e., 70%) assembled from the genome of one individual of mixed European/African ancestry. Recent revisions of

this reference have involved updating most SNPs based on European ancestry major alleles. This ancestry bias explains the startling results of a recent analysis of the DNA sequences of 910 individuals of African descent that found that approximately 10% of the pangenome constructed from these individuals (roughly 300 million DNA base pairs) was literally missing from the reference human genome (Sherman et al., 2019).

Large-scale population studies of common variants associated with common disease, the genome-wide association study (GWAS) revolution, have identified genetic associations with literally thousands of disease and common traits, mostly in European ancestry populations. The Psychiatric Genetics Consortium current GWAS of schizophrenia, involving 69,369 cases and 236,642 controls, includes not a single subject of AA. The latest GWAS of autism spectrum disorder, involving more than 190,000 individuals, included no one of AA. The recent state-of-the-art GWAS of Alzheimer's disease, involving over 600,000 people, also included no individuals of AA. The largest GWAS of Parkinson's disease, with upward of 1.9 million subjects, again, has no individuals of AA. A meta-analysis of GWASs of depression involving more than 2 million people is based exclusively on individuals of European ancestry. Of the 418 GWASs of neurological disorders in the NHGRI catalog, only 4% contain any minority groups, which include African, Native American, and Latinx ancestry. Polygene risk scores, which are based on summing risk-associated alleles at SNPs across the genomes from these GWAS datasets, predict illness risk in individuals of AA, on average, only one-fifth as strongly as predictions in people of European ancestry (Martin et al., 2019). The inclusion of primarily European ancestry individuals in current genetic research limits understanding of how genetics influences disease (Bentley et al., 2020) and obscures a significant portion of the potential for scientific advancements for personalizing medicine because emerging therapeutics may fail to apply equally to individuals of AA.

There is finally, however, momentum building to lessen these stark disparities in genetic research (Bentley et al., 2020). The effort jump started with the 1000 Genomes Project beginning in 2008, which endeavored to characterize genetic variation across diverse populations. The Personal Genome project, established in the United States in 2005 and recently expanded to include centers in Canada, Europe, and Asia, is an effort to eventually collect and make public the personal information and genome sequences of 100,000 individuals. It is unclear how many (if any) individuals of AA have volunteered so far. The NIH has launched the "All of Us" project, an ambitious program to acquire health information and sequence the genomes of 1 million individuals, with a target of 50% from minority populations. Some GWASs of individuals of AA with psychiatric disorders are finally ongoing in the United States and in Africa (Bigdeli et al., 2019). Also in Africa, The Human Heredity and Health in Africa (H3Africa) consortium is an effort to build partnerships of scientists across several African countries and study primarily general medical and infectious diseases. 23andME, the consumer genetics company, has made public whole-genome sequence data from 2,300 AA customers. The Genome Aggregation Database (gnomAD; https://gnomad.broadinstitute.org/), an ambitious effort to identify rare coding and noncoding variants as well as structural and repeat variations in large samples, has released data from more than 125,000 exome sequences and over 71,000 whole-genome sequences. AA genome data are available for 8,100 of the exome sequences and 12,100 of the whole-genome sequences, though as these sequences are currently mapped to the existing reference genome, it is unclear whether or how many variations unique to the AA genome may have been dropped because of mapping errors. The NIH also has committed to constructing multiple reference genomes of AA, which may be available within the next few years. This, in addition to reference panels for imputation of genotypes from AA populations becoming more inclusive of genomic diversity (e.g., the NHLBI TOPMed project), though short of closing the gap, is a major step toward bringing the missing AA populations into the personalized medicine agenda.

The Missing Brains in Brain Research

Population-based genetic associations do not directly identify potential mechanisms of disease, do not elucidate how variation in a gene matters to the development and function of the brain, and do not by themselves identify a specific causal gene. To bridge these critical gaps in CNS research, it is necessary to study genetic variation in the molecular context of a relevant biological tissue or organism. Since there are no animal models of human ancestral variation, the only high-fidelity biological tissue is the human brain. To understand the biology of unique ancestral variations and the heterogeneity of effects of common variations on gene networks and pathways requires brain tissue from diverse ancestries. Ancestry differences in gene expression, splicing, and epigenetic regulation are largely unknown, particularly in brain.

The work horse resource for research about genetic regulation of gene expression is the Gene-Tissue Expression Project (GTEx), a publicly available dataset of RNA-sequenced and SNP-genotyped tissues from nearly 1,000 individuals, including gross brain samples from 13 brain regions, with sample sizes per region varying from 114 to 209. Mixed ancestry tissue is found, on average, in only 7% of these various samples. In addition to the GTEx dataset being under representative of non-European ancestry tissue, concerns have been raised about RNA quality and instances of cross-sample contamination. The database representing participants from all Alzheimer's Disease Centers, as maintained by the National Alzheimer's Coordinating Center, lists 18 centers as having neuropathologic data from 5,283 brains, of which 321 (6.1%) are from individuals of AA. It is unclear how many of these brain samples are actually available for ongoing research.

The NIH neuroBioBank, a loose aggregation of brain repositories in eight academic centers, claims to have specimens from an aggregate sum of 1,364 brains from donors of AA, of which 1,091 are said to be neurotypical, 136 to have schizophrenia, and 57 to have Alzheimer's disease, among other less prevalent diagnoses. It is unclear what specific regions and samples remain and are

available for ongoing research from these tissues, and diagnostic criteria, dissection, and tissue processing are not uniform across sites. The BrainSeq Consortium, a precompetitive collaboration between seven pharmaceutical companies and the Lieber Institute for Brain Development (LIBD, see below), has published studies on RNA sequencing and SNP genotyping in up to 495 donors from prefrontal cortex, hippocampus, and granule cells of the dentate gyrus across normal development and in psychiatric disorders, principally schizophrenia, with up to 213 samples from donors of AA (Jaffe et al., 2020). PsychENCODE, a major NIMH initiative to characterize genetic and epigenetic regulation of gene expression in brain in relation to psychiatric disorders, released RNA-seq data from 1,695 brains, mostly prefrontal cortex samples, 392 of which were from AA donors (Gandal et al., 2018). The majority of the AA brain samples were from LIBD (n = 254). Interestingly, in the initial reports from these consortia, inclusion of AA brain tissue did not lead to an ancestry-specific analysis of how genetic variation across ancestry differentially influenced gene processing and expression. Rather, the samples were adjusted for global genomic ancestry in expression and in expression quantitative trait loci (eQTLs) analyses, a standard analytic approach in most studies of tissue from diverse ancestries in an effort to "pseudo-homogenize" the samples and identify biological effects that do not segregate with ancestry.

To the extent that environments also differ across populations and cultures, epigenetic programming, which in turn regulates gene expression and splicing, also likely differs (Gurdasani et al., 2019). Since genetic sequence is a major driver of epigenetic programming, different populations will show variable sensitivity and resilience to environmental risk factors based on their genomes. Brain tissue is necessary to understand the biological ramifications of these possibilities. As an example, extreme stress in early childhood has been identified as a potentially increasing risk for many common disorders throughout life, including depression, post-traumatic stress disorder (PTSD), and drug abuse, as well as diabetes, cancer, and heart disease. Individuals of AA, particularly in urban environments, may be differentially exposed to extreme early childhood stress. A recent report based on whole-genome bisulfite sequencing of neurons and glia sorted from prefrontal cortex of 24 mixed ancestry neurotypical individuals between the ages of infancy and 20 years showed that the rate of DNA methylation (i.e., the epigenetic plasticity of the neuronal genome) was dramatically greater during the first 5 years of life than during the following years (Price et al., 2019). The biggest bias in this rate of change across age was ancestry. Clearly, there is a need for examination of how AA critically modulates the expression of disease and of environmental exposures.

The Missing Minority Inclusion in Research

Low inclusion of minority groups in basic and clinical research is based on a number of factors, one of which reflects scientific opportunism. Since the AA genome has a more expanded population history, it is a more complex genome with more genetic diversity. This presents a greater statistical challenge for scientists to identify genes that are associated with common medical illnesses, in contrast to the European ancestry genome in which the statistical challenges are easier. This being said, the potential to discover novel loci and variants not common in European ancestry genomes is increasingly being realized in studies of admixed samples and those of diverse ancestries.

Another reason for the missing minority groups in research is a fundamental distrust fueled by a history of medical abuses and ethical violations. The Black community is keenly aware of the unfortunate legacy of the medical establishment's violation of its Hippocratic Oath, most egregiously, for example, the eugenics agenda in which Black women were disproportionality sterilized; the notorious Tuskegee study that withheld treatment for Black men with syphilis; the falsehoods promulgated about IQ data; and the shameful story of Henrietta Lacks, a Black woman whose cancer cells are the source of the HeLa cell line, the first immortalized cell line and one of the most important cell lines in medical research. Mrs. Lacks was the unwitting source of these cells from a tumor biopsied during treatment for cervical cancer at Johns Hopkins Hospital in Baltimore, Maryland, United States, in 1951. No consent was obtained to culture her cells, nor were she, her husband, or her five children compensated for their extraction or use. It is incumbent on the medical research establishment to restore trust, enhance the involvement of the minority population, and increase the confidence of minority communities in the beneficence of human research.

The African Ancestry Neuroscience Research Initiative (AANRI)

Today, in Baltimore, a city plagued by poverty and the psychological sequelae of multiple systematic traumas, a group of community leaders has partnered with the LIBD, an independent not-for-profit research institute on the Johns Hopkins Medical Campus, to transform the neuroscience research paradigm so that individuals of AA are no longer missing from studies of genes and the brain and actively involved in conducting this critical research. The partnership is committed to ensuring that minority communities are engaged from the start, when the scientific agenda is being set, in order to provide crucial trust building and for the results of the research to be understood and accessible to the community. The African Ancestry Neuroscience Research Initiative (AANRI) is led by Rev. Dr. Alvin Hathaway, pastor of the historic Union Baptist Church of Baltimore, and includes a formal partnership with Morgan State University, a historic black college, and participation by senior faculty from the Departments of African History and Anthropology and The Alexander Williams Law Center at the University of Maryland. A scientific advisory board is headed by Yasmin Hurd, PhD, of the Icahn School of Medicine at Mount Sinai and Kafui Dzirasa, MD, PhD, of Duke University. The initiative aims to galvanize civil rights leaders, scientists, affected communities, celebrities, philanthropists, business leaders, and government officials to advance an inclusive roadmap for neuroscience research as well as the creation of a future workforce of diverse scientists.

In little more than a year since its inception, the AANRI has secured local, state,

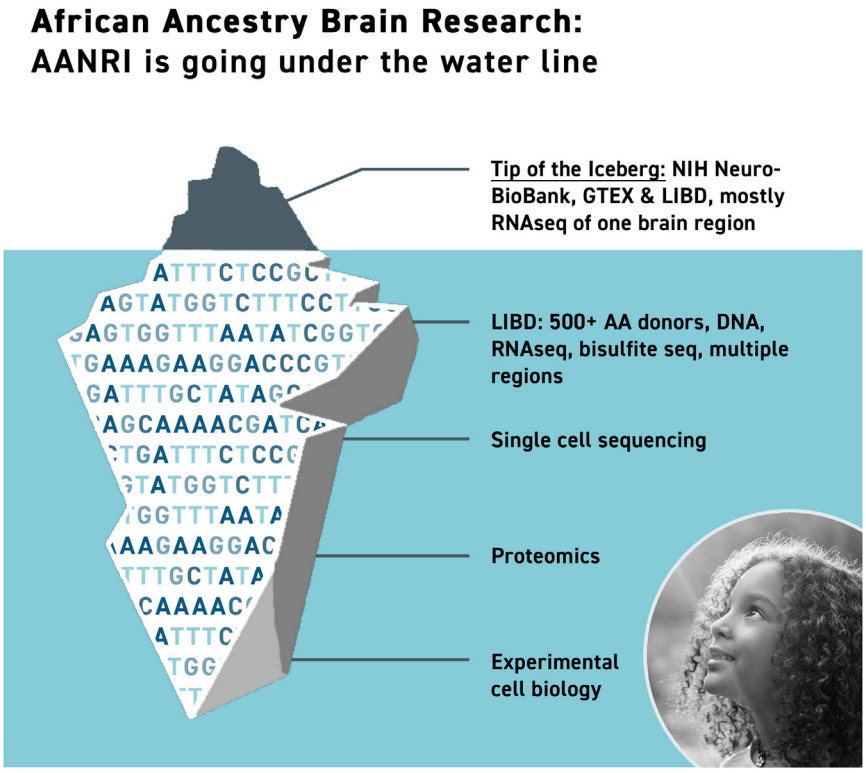

Figure 1. What's Needed Is below the Water Line

The African Ancestry Neuroscience Research Initiative (AANRI) is a partnership between prominent community leaders in Baltimore, Maryland, the Lieber Institute for Brain Development, Morgan State University, Brown Capital Management, The Abell Foundation, and the State of Maryland to generate a public functional genomic dataset focused on human brain from donors of AA. The first phase will go beyond earlier studies (i.e., the tip of the iceberg) to measure gene and protein expression of multiple brain regions and individual cells from 500 donors of AA. Future work will involve expanded brain datasets and pluripotent cell lines from brain donors. Abbreviations are explained in the text.

and federal government support to close the gap in brain research. Last year, Senator Ben Cardin (D-Md) championed authorizing language in the HHS Senate Appropriations Bill that solicited unprecedented bipartisan backing for an appropriation of $20 million in additional funds to the NIMH and the NIMHD targeting neuroscience research for ethnic minority groups. This past March, Maryland Governor Larry Hogan authored a supplemental budget for Fiscal Year 2021 that included $1.25 million in support specifically for the AANRI. In May, Brown Capital Financial Management, a philanthropic group in Baltimore, committed $1 million and the Abell Foundation awarded a grant of $275,000 for administration. The LIBD has committed $500,000 for the first phase of the research. Thus, with startup support in excess of $3 million, the AANRI has begun the first phase of brain genomics research that will lead to the public release in real time of data over the coming 3 years. The initiative will produce the largest dataset in the world of AA brain tissue that can be used by researchers globally to make medicine more personalized (Figure 1).

The critical resource necessary to achieve the goals of the AANRI is the availability of high-quality human brain tissue of individuals of AA. With brain samples from over 3,100 individuals, including donations from more than 700 African American families, and a donation rate of over 400 new brains each year, the LIBD brain repository is one of the most extensive and well-curated collections of postmortem, multi-ancestry human brain material in the world. Brains are donated to the LIBD repository by next of kin, and the cases consented are then diagnosed as neurotypical or with a spectrum of neuropsychiatric disorders including schizophrenia, bipolar disorder, autism, depression, and anxiety disorders, with neurodegenerative disorders such as Alzheimer's disease, and with traumatic brain injuries, suicide, PTSD, and drug abuse. All samples have been collected and processed by the same team for more than 20 years, and all dissections are performed by the same neuroanatomist. Each case has a detailed clinical history collected via interviews with the next of kin, treating physicians, and available medical records and includes comprehensive toxicology testing. The extensive clinical database contains over 100 fields of detailed demographic, clinical, and laboratory information on every sample. Brain tissue donations are collected from the Gift of Life program in Michigan as well as from four medical examiner sites around the country: The Office of the Chief Medical Examiner in Baltimore, Maryland; WMed-Homer Stryker School of Medicine, Kalamazoo, Western Michigan; University of North Dakota Pathology Office, Grand Forks, North Dakota; and the Medical Examiner's Office in Santa Clara County, California.

The first phase of the AANRI will involve whole-genome DNA sequencing of each brain, followed by RNA sequencing, bisulfite sequencing, and peptide sequencing of bulk homogenate tissue from multiple cortical and subcortical brain regions from 500 donors of AA, combined neurotypical samples, and several neuropsychiatric diagnoses. These initial brain samples have been selected because of the high quality of the tissue and RNA, the availability of most of the brain, and access for future investigation to consented living dural fibroblasts cultured from most cases. As additional support for AANRI is secured, a second phase of the project will proceed, focused on single-cell sequencing of subsamples from phase I, with analogous "multi-omics" on individual nuclei and cells isolated using droplet techniques as well as laser capture and spatial transcriptomics technology. Phase III will involve additional samples to fill in gaps in the prior phases around specific diagnostic groups to generate high-dimensional analyses as contrast to other LIBD and public data from similar assays in brain tissue of European ancestry.

Outlook

The AANRI project will generate a detailed functional genomic resource, a compendium of how genetic and epigenetic variation influence gene and protein expression and splicing in brains of individuals of AA, how specific regions and cells parse these associations, and how genetic variations show heterogeneous molecular effects across ancestries. The ancestry-dependent stratification of gene and protein expression in brain will be essential for translating genetic risk loci associated with CNS disorders into causative genes, pathogenic mechanisms, and potentially "druggable" pathways and targets that are heterogeneous across populations. But there will be many challenges. Remapping based on multiple reference sequences will require novel computational approaches. Translating expression into disease association will require larger ancestry-based genetic association datasets and more human brain samples from individuals of AA with brain disorders. Validating the mechanism of illness and new therapeutic targets requires experimental biology, which cannot be done in postmortem brain tissue, but the availability of living cells from the majority of brains in the AANRI offers a unique opportunity to do the necessary research. Community involvement at many levels will be critical to advancing this project, as will bringing other sites and scientists into the AANRI. The inclusion of individuals of AA in the mainstream of genomic and neuroscience research will not only proverbially level the playing field, it will lead to an understanding of human illness at much greater depth and breadth and will help create a precision medicine paradigm that is truly personalized. The AANRI is committed to achieving these goals while preparing the next generation of diverse scientists to lead this critical research for future generations.

ACKNOWLEDGMENTS

The AANRI would not have happened were it not for the tireless dedication and commitment of Kari Stoever, Becky Oldham, and Martin Knott. The Lieber Institute brain repository was funded by philanthropic gifts from the families of Connie and Stephen Lieber and Milton and Tamar Maltz and was created and is led by Tom Hyde and Joel Kleinman, who, it should be noted, also were responsible for establishing and maintaining for more than 20 years the NIMH Clinical Brain Disorder Branch brain repository, renamed in 2013 to the NIMH Human Brain Collection Core.

REFERENCES

Bailey, Z.D., Krieger, N., Agénor, M., Graves, J., Linos, N., and Bassett, M.T. (2017). Structural racism and health inequities in the USA: evidence and interventions. Lancet *389*, 1453–1463.

Bentley, A.R., Callier, S.L., and Rotimi, C.N. (2020). Evaluating the promise of inclusion of African ancestry populations in genomics. NPJ Genom. Med. *5*, 5.

Bigdeli, T.B., Genovese, G., Georgakopoulos, P., Meyers, J.L., Peterson, R.E., Iyegbe, C.O., Medeiros, H., Valderrama, J., Achtyes, E.D., Kotov, R., et al.; Consortium on the Genetics of Schizophrenia (COGS) Investigators; Genomic Psychiatry Cohort (GPC) Consortium (2019). Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. Mol. Psychiatry. Published online October 7, 2019. https://doi.org/10.1038/s41380-019-0517-y.

Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al.; PsychENCODE Consortium (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*, eaat8127.

Gurdasani, D., Barroso, I., Zeggini, E., and Sandhu, M.S. (2019). Genomics of disease risk in globally diverse populations. Nat. Rev. Genet. *20*, 520–535.

Jaffe, A.E., Hoeppner, D.J., Saito, T., Blanpain, L., Ukaigwe, J., Burke, E.E., Collado-Torres, L., Tao, R., Tajinda, K., Maynard, K.R., et al. (2020). Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. Nat. Neurosci. *23*, 510–519.

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.

Price, A.J., Collado-Torres, L., Ivanov, N.A., Xia, W., Burke, E.E., Shin, J.H., Tao, R., Ma, L., Jia, Y., Hyde, T.M., et al. (2019). Divergent neuronal DNA methylation patterns across human cortical development reveal critical periods and a unique role of CpH methylation. Genome Biol. *20*, 196.

Rajabli, F., Feliciano, B.E., Celis, K., Hamilton-Nelson, K.L., Whitehead, P.L., Adams, L.D., Bussies, P.L., Manrique, C.P., Rodriguez, A., Rodriguez, V., et al. (2018). Ancestral origin of *ApoE ε4* Alzheimer disease risk in Puerto Rican and African American populations. PLoS Genet. *14*, e1007791.

Sherman, R.M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M.P., Chavan, S., Vergara, C., Ortega, V.E., et al. (2019). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nat. Genet. *51*, 30–35.