



## OPEN ACCESS

## EDITED BY

Farren Briggs,  
University of Miami, United States

## REVIEWED BY

Dana C. Crawford,  
Case Western Reserve University, United States

## \*CORRESPONDENCE

Lindsay Fernandez-Rhodes,  
✉ fernandez-rhodes@psu.edu

RECEIVED 15 October 2025

REVISED 29 November 2025

ACCEPTED 09 December 2025

PUBLISHED 05 January 2026

## CITATION

Tu Y and Fernandez-Rhodes L (2026)  
Monitoring diversity in genome-wide  
association studies requires measuring and  
reporting on immigration-related factors.  
*Front. Genet.* 16:1725866.  
doi: 10.3389/fgene.2025.1725866

## COPYRIGHT

© 2026 Tu and Fernandez-Rhodes. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Monitoring diversity in genome-wide association studies requires measuring and reporting on immigration-related factors

Yao Tu<sup>1</sup> and Lindsay Fernandez-Rhodes<sup>1,2\*</sup>

<sup>1</sup>Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, University Park, PA, United States, <sup>2</sup>Department of Epidemiology, University of North Carolina at Chapel Hill, Gillings School of Global Public Health, Chapel Hill, NC, United States

Genome-wide association studies (GWAS) have made remarkable progress to date in deciphering the genetic foundations of complex traits, yet persistent gaps remain in how sample heterogeneity is measured and reported. Current practices typically emphasize diversity by broad ancestry categories or stratification by country of recruitment, but these dimensions alone fail to capture the immigration-related factors that contribute to the genetic or environmental origins of heterogeneity. We argue that incorporating variables, such as country of origin, in descriptions and analyses provides essential context for interpreting genetic associations, particularly in increasingly multi-population and trans-national GWAS samples. We highlight how neglected these variables are in the literature using the GWAS Catalog. We provide suggestions for reporting on these data in future studies. By advocating for a more comprehensive view of diversity in GWAS, we aim to address the under-representation of immigrants in GWAS and thereby strengthen the validity and interpretability of future genomic studies.

## KEYWORDS

Country of birth, country of recruitment, diversity and inclusion, environment, gene-environment (G-E) interaction, genome-wide association studies, immigration

## Introduction

Genome-wide association studies (GWAS) play an important role in revealing variants that are associated with specific traits from complex genomic structures. Careful attention to study design and methodology, such as harmonizing data, adjusting or stratifying analyses by key covariates, and incorporating environmental interactors, can substantially improve the discovery and generalization of GWAS findings (Visscher et al., 2017). However, unexplained heterogeneity in the results of these studies can arise from differences in phenotype definitions, environmental exposures, and technical platforms, all of which influence the comparability of results across studies. Monitoring and correcting contributors to heterogeneity has become increasingly important with the growing size of GWAS samples and the increased use of advanced meta-analytic techniques, which enhance study power by combining GWAS summary statistics from across studies and populations (Kulminski et al., 2016; Woodward et al., 2022).

Potential sources of heterogeneity are not limited to genetic factors; non-genetic factors play a major role, as well. Immigration—the voluntary or involuntary movement of individuals across country borders—is ushering in structural changes to an individual's

environmental exposures, behavioral or cultural changes in lifestyles, socioeconomic contexts, etc., that could introduce phenotypic heterogeneity in GWAS (Peterson et al., 2019; Suarez-Pajes et al., 2021). For example, a wealth of research on immigrant populations shows that the prevalence of many diseases, such as obesity, diabetes and hypertension, increases along with the duration of time an immigrant lives in the United States (US) (Oza-Frank et al., 2011; Parikh, 2011; Commodore-Mensah et al., 2016). Describing and modeling such non-genetic sources is therefore essential to the validity of GWAS findings, as it improves our ability to distinguish statistical artifacts from true genetic associations, and provides benchmarks for us to use when examining the external validity of the results to diverse and dynamic populations like immigrants.

## Heterogeneity in GWAS by country context and other immigration-related factors

In GWAS, covariates are used to adjust for potential heterogeneity in the phenotype, and most analyses include standard covariates such as age, sex and principal components, as well as any that are phenotype specific. Inclusion of principal components (or some other global measures of genetic ancestry) can avoid false positive results due to population stratification. Furthermore, the race/ethnicity of participants is usually obtained via self-report, or when missing, inferred by similarity in principal component space, and in some cases may have been used for further batch assignment or stratification. Lastly, although often not acknowledged explicitly, practically speaking, most researchers stratify their analyses by country of recruitment, due to the heterogeneity that separate funding streams, regulations, or research teams impose on the collection of samples. For example, the Health and Retirement Study (HRS) is joined by an international family of 17 studies in different countries across the globe (Lee et al., 2021). Those sister studies are run by different teams in each country and follow similar protocols to the HRS. This research infrastructure is an invaluable resource and could be leveraged to construct thoughtful contrasts that would estimate the heterogeneity caused by different countries of recruitment, above and beyond differences in ancestry. However, their results may still be affected by differences in the countries of origin of their sampled participants, especially in cases where a large share of participants in the sample were born in another country than the country of recruitment.

There are clear disparities in which countries have been able to assemble large enough samples to conduct GWAS, to date. High-income countries, with about 17% of the world population (World Bank, 2021), contributed 84% of GWAS participants, according to data from the GWAS Catalog (Sollis et al., 2022). Low and middle-income countries have contributed a far smaller proportion of GWAS and samples. Most high-income countries also have a high immigration rate from middle-income countries (Spooner et al., 2022), which begs the question of whether GWAS from these countries (especially population representative studies) adequately represent immigrants living within their borders or account for what may be heterogeneity due to immigration-related factors. Thus, we posit that there is added value in incorporating information on immigration and related social factors into GWAS design and analysis.

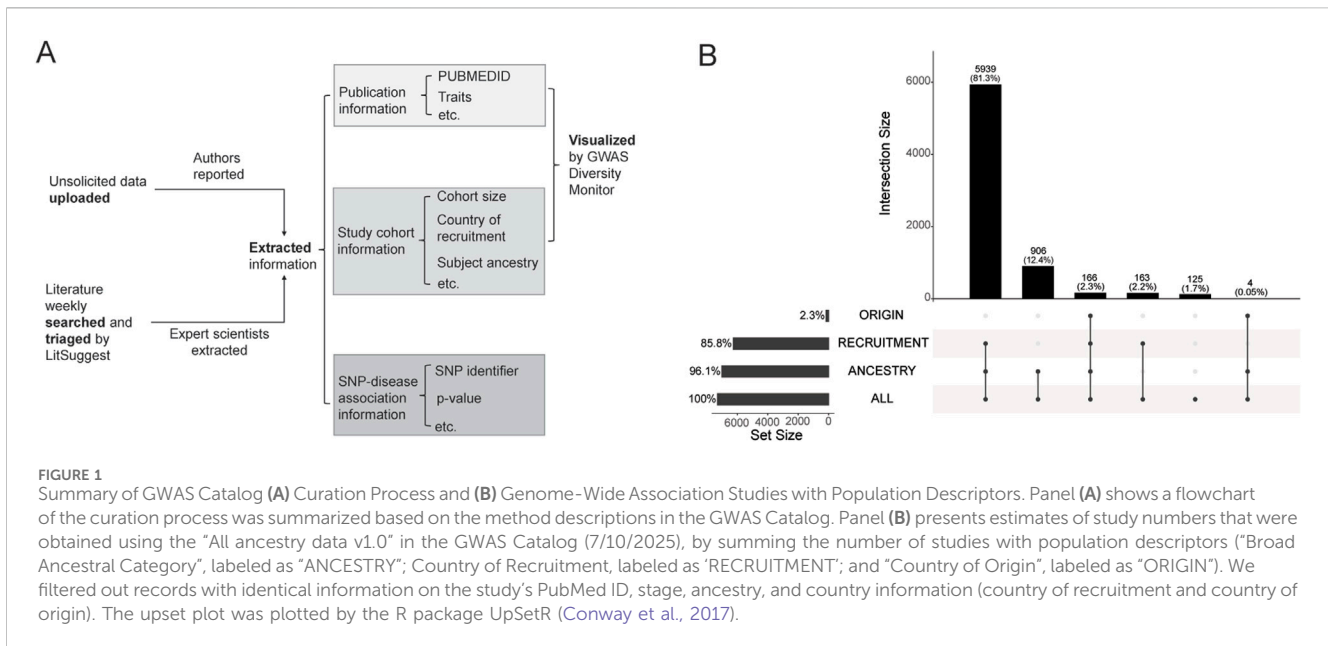
As the largest resource of GWAS summary statistics, the GWAS catalog has a well-developed curation process, including both manual and automatic data extraction and cleaning (GWAS Catalog, 2016). Eligible papers are identified through weekly automatic searches using LitSuggest (Allot et al., 2021). Then, full genome-wide GWAS statistics and complementary information are either reported by the author upload (optional) or manually extracted by expert scientists from the literature (Figure 1A). The GWAS Catalog uses 'broad' ancestral categories (either explicitly defined or inferred when information is lacking) to describe specific GWAS and the field's representation in terms of genetic ancestry (Morales et al., 2018). Data on this variable is rarely missing at a study level, as 96.1% of the 7,303 GWAS studies in the GWAS Catalog as of July 2025 reported information on the ancestry of the sample (Figure 1B). Similarly, 85.8% of all cataloged studies had 'country of recruitment' information, and when country of recruitment is present for a study, it is very common (97.4%) to also have ancestry information.

Using country of recruitment information downloaded from the GWAS Catalog, the GWAS Diversity Monitor maps out country of recruitment by publication year across all studies, traits/outcomes, and stages (Mills and Rahal, 2020). Following their methods, we have plotted out the distribution of countries of recruitment in the GWAS Catalog since 2008 (Figure 2A). Whenever multiple countries are reported for one publication, we make the strong assumption that they each had equal sample size, which may make the geographic distributions of GWAS samples to date appear more equitable than they really are. Furthermore, whenever the same samples could be included in multiple studies or meta-analyses, the estimated contribution from that country may also be overestimated, which is a limitation of current practices in monitoring diversity in GWAS.

GWAS literature's current focus on describing and monitoring 'broad' ancestral categories and country of recruitment as main indicators of GWAS diversity fails to address more nuanced aspects of ancestral and geographic diversity. Additionally, questions remain about how much residual heterogeneity exists due to unmeasured immigration-related factors, such as the life course exposures that could serve as gene-environment interactors or drivers of phenotypic variability (Fernández-Rhodes, 2023). While most GWAS include age and sex as covariates that may, at least in part, characterize varied immigration waves (e.g., as in the case of younger male guest workers), these covariates do not account for all aspects of the immigration experience. Such additional factors may include the experience of structural inclusionary/exclusionary practices towards certain countries of origin groups, socioeconomic differentials between sending versus host countries, an individual's reasons for emigration (voluntary emigration versus involuntary displacement) or others.

## Reporting immigration-related factors in GWAS

Although some immigration-related factors (e.g., time living in the US, age at first immigration, immigrant generation, reasons for or timing of most recent immigration) may be unavailable in most datasets, 'country of origin' is an easily measured demographic



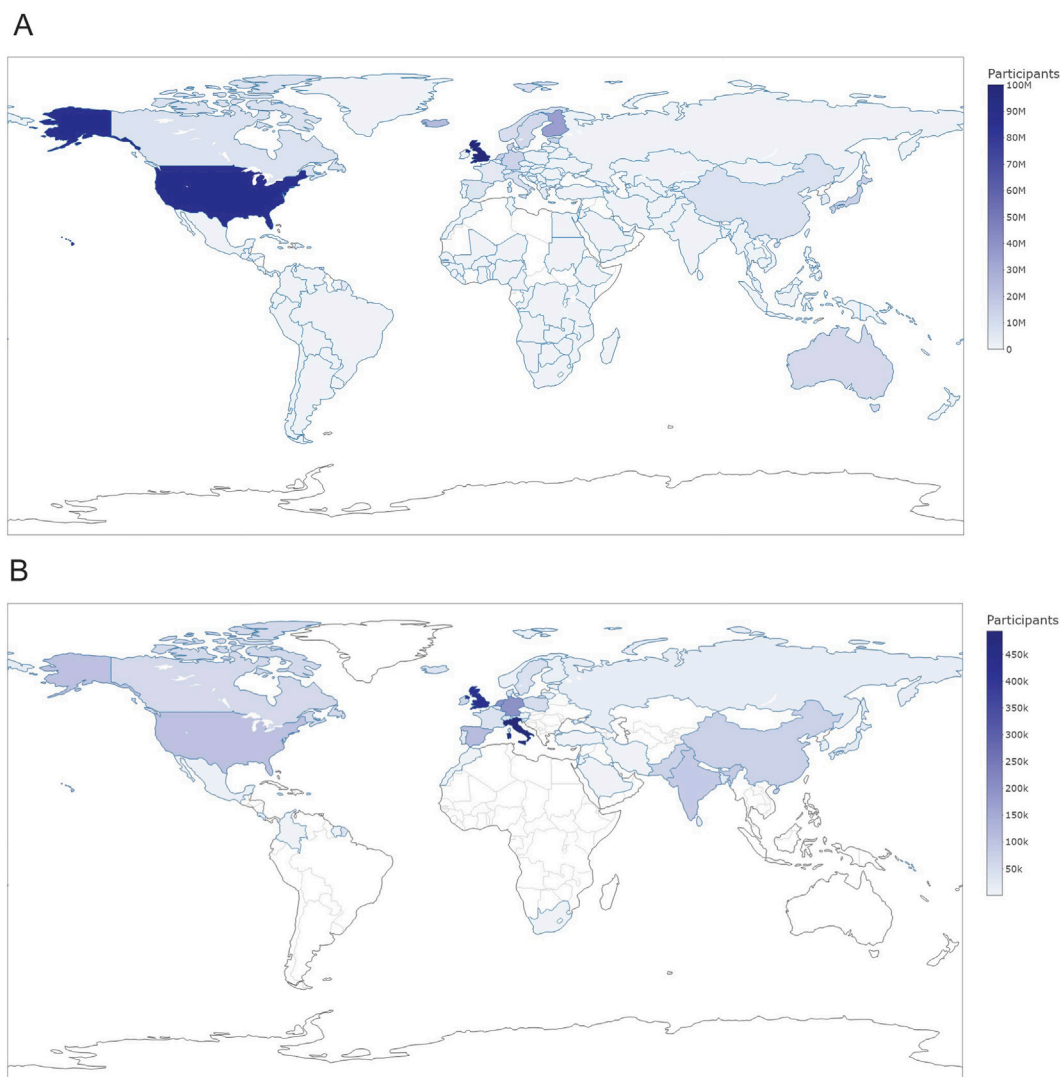
variable that we believe may capture some of the residual heterogeneity that not explained by country of recruitment, ancestry or other standard covariates alone. To answer the question of “What would it take to improve our monitoring of immigrant samples in the GWAS literature?”, we turn again to the GWAS Catalog. Of all GWAS published through July 2025, it becomes clear that remarkably few studies (2.3%) report enough information for a ‘country of origin’ designation. Whenever country of origin is provided in a GWAS publication, information on ancestry is always present (Figure 1B).

This sparse reporting of country of origin is likely due to how the GWAS Catalog defines this variable—the country of origin of the participants’ grandparents, or if lacking, the country of known genealogy (Morales et al., 2018). This definition is in contrast with the way that immigration scholars define country of origin—as being primarily related to the participant themselves. Likewise, a search for country of origin in the PhenX Toolkit suggests a demographic variable for ‘birthplace’ of the participant (Krzyzanowski et al., 2023). In cases where ‘country of birth’ is reported for just the participant or their parents, but there is no specific mention of grandparents or genealogy, based on this definition the GWAS Catalog curators would erroneously list country of origin as not reported. An example of this comes from our own work in the Hispanic Community Health Study/ Study of Latinos, where participants’ country of origin, nativity, and Hispanic/Latino background variables, or some combination of them, are routinely reported in the descriptives; of the 31 GWAS published from this study and reviewed separately (Rao et al., 2025), only one had country of origin information in the GWAS Catalog (Burkart et al., 2018). Although the curators may choose to provide an additional ancestry description with supplemental information, it does not appear to be consistently done.

Assuming information on immigration-related factors like a participant’s country of origin exists in most genomic datasets, we recommend two short-term solutions to advance the monitoring of

representation of immigrants and improve estimations of them as a source of heterogeneity in genetic studies. First, the GWAS Catalog should prospectively update its definition of country of origin to also include both participants’ and grandparental/genealogic information. Alternatively, an additional variable called ‘participant country of birth’ could be generated for all future data extractions to accomplish the goal of being able to track more information on the diversity and inclusion of the field. Either of these options would have the advantage of better distinguishing between ancestry, grandparental/genealogic origins and a participant’s own country of origin, which in many cases can be different within an individual, and can vary within and across country of recruitment.

Second, the GWAS Diversity Monitor should be redesigned to include an additional map of country of origin, just as they have done for country of recruitment (Figure 2A). Figure 2B shows the dearth of information on country of origin in the GWAS Catalog since 2008, with both a lower total participant sample overall (as compared to Figure 2A) and a different geographic distribution of common countries of origin. The United Kingdom (UK) was one of the most commonly identified countries for both recruitment (largest) and origin (second largest). The UK Biobank has participated in thousands of GWAS publications to date (UK BioBank, 2020), and is commonly in studies with country of recruitment, but not with country of origin information. In contrast, the US is the second most commonly mentioned country of recruitment (Figure 2A), but is not as common as a country of origin (sixth most common, Figure 2B). Italy was among the top countries of recruitment (the thirteen most common, Figure 2A), but surprisingly was the most common country of origin in the GWAS Catalog (Figure 2B). We hope that adding country of origin as an additional visual depiction to the GWAS Diversity Monitor will serve as a reminder for future researchers interested in the diversity of the field to strive to describe immigration-related variables in their own work.



**FIGURE 2** Total Participant Sample in Genome-Wide Association Studies of the GWAS Catalog Mapped by (A) Country of Recruitment and (B) Country of Origin. Estimates of the total participants per country were obtained by summing the number of individuals across all studies, stages and ancestral categories in the same filtered dataset used in (B). When studies listed multiple countries, we assumed that the number of individuals in this specific study from each country was uniform and calculated the country-specific sample as the number of individuals/number of countries listed. Interactive html version of these figures can be founded online: (A) [https://yao876.github.io/GWAS-Catalog-maps/Participants%20by%20Country%20of%20Recruitment\\_max100M.html](https://yao876.github.io/GWAS-Catalog-maps/Participants%20by%20Country%20of%20Recruitment_max100M.html) (B): <https://yao876.github.io/GWAS-Catalog-maps/Participants%20by%20Country%20of%20Origin.html>; if possible, as they are interactive for the user.

## Discussion

Recent commentaries on the topic of diversity in GWAS have shed light on how immigration-related factors should be considered as another aspect of diversity (Fernández-Rhodes, 2023; Riehm et al., 2023). As described above, we have advanced this work by identifying several challenges to monitoring the inclusion of immigrants in GWAS. We call for future studies, and particularly new data-generating initiatives, to measure several immigration-related factors in their participants using established methods such as those in the PhenX Toolkit (Krzyzanowski et al., 2023) and report on these factors in the GWAS descriptives, just as they would do for age, sex, etc.

Immigration-related factors may include country of birth of the participant, birthplace of parents or grandparents, years of residence, or even go beyond these measures to include variables like age at first immigration, immigrant generation, reasons for migration (forced versus voluntary), patterns of seasonal migration, and current citizenship or legal immigration status, provided that the privacy of this information can be protected. Calls have been made for the inclusion of these variables and other variables in US-based national surveys (Marouf et al., 2025) as well as in studies outside of the US context (Pathak et al., 2021). We recognize that not all datasets have information on such immigration-related factors, and even when it is collected, the data may not be missing at random and should be analyzed with this in mind.

Nonetheless, a greater availability of such data in the larger genomics ecosystem should allow researchers to better describe and possibly also account for the potential heterogeneity related to immigration-related factors. As genomics researchers begin to work with more immigration-related variables, they should have the ethical, legal and social implications of their work always in front of mind. Whenever possible, investigators should foster community-engaged and informed designs, as a means of ensuring that communities are able to actively guide the manner in which immigration-related variables are incorporated in GWAS descriptives and analyses (Lemke et al., 2022; Ravi et al., 2025; Valdez-Aguilar et al., 2025). Yet, numerous challenges remain to ensuring that all individuals and communities, immigrant and non-immigrant, feel that their privacy is adequately protected amidst changing political times.

For example, in the US research context, crowd-sourced resources estimate that more than several hundred research projects have been terminated by the National Institutes of Health (NIH) and National Science Foundation in the first months of 2025 (Kozlov, 2025; Kozlov and Ryan, 2025; Ross et al., 2021), totaling at least \$4 billion dollars of lost research support (Nienaber and Leeming, 2025). Although the exact methodology that was used to flag grants remains unclear, most of these grants were focused on diversity, equity and inclusion, health disparities, and included either investigators or participants from under-represented groups in genomics (Kozlov and Ryan, 2025; Nienaber and Leeming, 2025). Simultaneously, the NIH has seen a 6% reduction in force, delays in grant reviews, and all-time low success rates (Kozlov, 2025). In the 2025 fiscal year, the NIH funded several thousand less grants than in the previous fiscal year, by multi-year disbursements to fewer awardees in the final months of the fiscal year.

In light of long-standing disparities in the participation of women and minorities in NIH-funded research that predate such recent developments [for example, (NIH, 2023)], it may become harder to engage under-represented, marginalized groups, such as immigrants, in genomic studies in the near future. Scholars have expressed concern about the future of NIH-issued Certificates of Confidentiality (NIH, 2024) in protecting the privacy of sensitive participant data in both genomics and other areas of health research (Wolf et al., 2024; Kozlov, 2025; LoTempio and Moreno, 2025). Thus, it may be important for researchers to consult with legal scholars, bioethicists, or form other interdisciplinary connections to guard against individual harms or discuss what may be population-level sensitivities. Greater uptake of community-engaged initiatives may be one way to better understand immigration as a facet of diversity that has been largely missing in the GWAS literature to date, as well as address the historical under-representation of certain groups in genomics research (Lemke et al., 2022).

Immigration is an important contributor to environmental exposures, behavioral/cultural practices, lifestyles and the socioeconomic circumstances, in which study participants find themselves. Immigration-related variables should be explicitly described and considered as analytic variables in future GWAS and genomics initiatives, including diversity monitoring. By advocating for more comprehensive diversity metrics, we hope

that future research can address the under-representation of immigrants in GWAS, while strengthening the validity and interpretability of their genomic findings.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: We obtained information on published GWAS in the GWAS Catalog file named 'All ancestry data v1.0' here: <https://www.ebi.ac.uk/gwas/docs/file-downloads>, accessed on July 18, 2025.

## Author contributions

YT: Data curation, Visualization, Writing – original draft, Writing – review and editing. LF-R: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review and editing.

## Funding

The author(s) declared that financial support was not received for this work and/or its publication.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author LF-R declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## 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

- Allot, A., Lee, K., Chen, Q., Luo, L., and Lu, Z. (2021). LitSuggest: a web-based system for literature recommendation and curation using machine learning. *Nucleic Acids Res.* 49, W352–W358. doi:10.1093/nar/gkab326
- Burkart, K. M., Sofer, T., London, S. J., Manichaikul, A., Hartwig, F. P., Yan, Q., et al. (2018). A genome-wide association study in Hispanics/Latinos identifies novel signals for lung function. The Hispanic community health study/study of Latinos. *Am. J. Respir. Crit. Care Med.* 198, 208–219. doi:10.1164/rccm.201707-1493oc
- Commodore-Mensah, Y., Ukonu, N., Obisesan, O., Aboagye, J. K., Agyemang, C., Reilly, C. M., et al. (2016). Length of residence in the United States is associated with a higher prevalence of cardiometabolic risk factors in immigrants: a contemporary analysis of the National Health Interview survey. *J. Am. Heart Assoc.* 5, e004059. doi:10.1161/jaha.116.004059
- Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi:10.1093/bioinformatics/btx364
- Fernández-Rhodes, L. (2023). Beyond borders: a commentary on the benefit of promoting immigrant populations in genome-wide association studies. *Hum. Genet. Genom. Adv.* 4, 100205. doi:10.1016/j.xhgg.2023.100205
- GWAS Catalog (2016). GWAS catalog: the NHGRI-EBI catalog of human genome-wide association studies. Available online at: <https://www.ebi.ac.uk/gwas/docs/methods/curation> (Accessed November 28, 2025).
- Kozlov, M. (2025). NIH races to spend its 2025 grant money — but fewer projects win funding. *Nature* 646, 264–266. doi:10.1038/d41586-025-03168-4
- Kozlov, M., and Ryan, C. (2025). How trump 2.0 is slashing NIH-backed research — in charts. *Nature* 640, 863–865. doi:10.1038/d41586-025-01099-8
- Krzyzanowski, M. C., Ives, C. L., Jones, N. L., Entwisle, B., Fernandez, A., Cullen, T. A., et al. (2023). The PhenX toolkit: measurement protocols for assessment of social determinants of health. *Am. J. Prev. Med.* 65, 534–542. doi:10.1016/j.amepre.2023.03.003
- Kulminski, A. M., Loika, Y., Culminkaya, I., Arbeev, K. G., Ukraintseva, S. V., Stallard, E., et al. (2016). Explicating heterogeneity of complex traits has strong potential for improving GWAS efficiency. *Sci. Rep.* 6, 35390. doi:10.1038/srep35390
- Lee, J., Phillips, D., Wilkens, J., and Team, G. to G. A. D. (2021). Gateway to global aging data: resources for cross-national comparisons of family, social environment, and healthy aging. *J. Gerontol. Ser. B* 76, S5–S16. doi:10.1093/geronb/gbab050
- Lemke, A. A., Esplin, E. D., Goldenberg, A. J., Gonzaga-Jauregui, C., Hanchard, N. A., Harris-Wai, J., et al. (2022). Addressing underrepresentation in genomics research through community engagement. *Am. J. Hum. Genet.* 109, 1563–1571. doi:10.1016/j.ajhg.2022.08.005
- LoTempio, J. E., and Moreno, J. D. (2025). Overcoming challenges associated with broad sharing of human genomic data. *Nat. Genet.* 57, 287–294. doi:10.1038/s41588-024-02049-2
- Marouf, F. E., Perreira, K. M., and Pham, H. (2025). Adding nativity, citizenship, and immigration status to Health Monitoring and survey data. *Am. J. Public Heal.* 115, 75–82. doi:10.2105/ajph.2024.307867
- Mills, M. C., and Rahal, C. (2020). The GWAS Diversity monitor tracks diversity by disease in real time. *Nat. Genet.* 52, 242–243. doi:10.1038/s41588-020-0580-y
- Morales, J., Welter, D., Bowler, E. H., Cerezo, M., Harris, L. W., McMahon, A. C., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS catalog. *Genome Biol.* 19, 21. doi:10.1186/s13059-018-1396-2
- Nienaber, V., and Leeming, J. (2025). What research might be lost after the NIH's cuts? Nature trained a bot to find out. *Nature*. doi:10.1038/d41586-025-02748-8
- NIH (2023). Inclusion of women and minorities in clinical research. Available online at: <https://report.nih.gov/research/inclusion-women-and-minorities-clinical-research/> (Accessed November 28, 2025).
- NIH (2024). CoCs for NIH-Funded research. Available online at: <https://grants.nih.gov/policy-and-compliance/policy-topics/human-subjects/coc/nih-funded-research> (Accessed November 28, 2025).
- Oza-Frank, R., Stephenson, R., and Narayan, K. M. V. (2011). Diabetes prevalence by length of residence among US immigrants. *J. Immigr. Minor. Heal.* 13, 1–8. doi:10.1007/s10903-009-9283-2
- Parikh, N. I. (2011). Sex differences in the risk of cardiovascular disease. *BMJ* 343, d5526. doi:10.1136/bmj.d5526
- Pathak, N., Zhang, C. X., Boukari, Y., Burns, R., Mathur, R., Gonzalez-Izquierdo, A., et al. (2021). Development and validation of a primary care electronic health record phenotype to study migration and health in the UK. *Int. J. Environ. Res. Public Heal.* 18, 13304. doi:10.3390/ijerph182413304
- Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C.-Y., Popejoy, A. B., Periyasamy, S., et al. (2019). Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* 179, 589–603. doi:10.1016/j.cell.2019.08.051
- Rao, H., Weiss, M. C., Moon, J. Y., Perreira, K. M., Daviglus, M. L., Kaplan, R., et al. (2025). Advancements in genetic research by the Hispanic Community Health Study/Study of Latinos: a 10-year retrospective review. *Hum. Genet. Genom. Adv.* 6, 100376. doi:10.1016/j.xhgg.2024.100376
- Ravi, P., Smith, M., Bogović, S., Lin, C., Yaffee, A., Dudgeon, M., et al. (2025). Characterizing identifiers for immigrant populations in electronic health records. *J. Immigr. Minor. Heal.* 27, 1–8. doi:10.1007/s10903-025-01698-7
- Riehm, K. E., Keyes, K. M., and Susser, E. S. (2023). Social determinants of health and selection bias in genome-wide association studies. *World Psychiatry* 22, 160–161. doi:10.1002/wps.21047
- Ross, N., Delaney, S., Barente, A., Mairson, E., Scott, E., and Shan, M. (2021). Grant witness. Available online at: <https://grant-witness.us/> (Accessed November 28, 2025).
- Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., et al. (2022). The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* 51, D977–D985. doi:10.1093/nar/gkac1010
- Sollis, E., Spooner, F., Acisu, T., Teutem, S., van, Ritchie, H., Ortiz-Ospina, E., et al. (2022). *Migration*. Available online at: <https://www.ebi.ac.uk/gwas/docs/file-downloads> (September 13, 2025)
- Suarez-Pajes, E., Usera, A. D., Marcelino-Rodríguez, I., Guillen-Guio, B., and Flores, C. (2021). Genetic ancestry inference and its application for the genetic mapping of human diseases. *Int. J. Mol. Sci.* 22, 6962. doi:10.3390/ijms22136962
- UK Biobank (2020). Publications - UK biobank. Available online at: <https://www.ukbiobank.ac.uk/discoveries-and-impact/publications/> (Accessed November 28, 2025).
- Valdez-Aguilar, M., Conklin, J. L., and Abebe, D. S. (2025). Engaging migrants and immigrants in genetics research. *Nat. Genet.* 57, 2634–2643. doi:10.1038/s41588-025-02373-1
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. L., Brown, M. A., et al. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi:10.1016/j.ajhg.2017.06.005
- Wolf, L. E., Ram, N., Contreras, J., and Beskow, L. M. (2024). Certificates of confidentiality: privileging research data. *J. Law Biosci.* 11, lsa003. doi:10.1093/jlb/lsa003
- Woodward, A. A., Urbanowicz, R. J., Naj, A. C., and Moore, J. H. (2022). Genetic heterogeneity: challenges, impacts, and methods through an associative lens. *Genet. Epidemiol.* 46, 555–571. doi:10.1002/gepi.22497
- World Bank (2021). *Population, total for high income countries [SPPOPTOTLHC]*, retrieved from FRED. St. Louis: Federal Reserve Bank of. Available online at: <https://fred.stlouisfed.org/series/SPPOPTOTLHC> (Accessed September 1, 2025).