The human genome holds the key to who we are. In the era of precision medicine, it is essential to unlock this key and determine what genetic sequences can reveal about a person and their future. Recent advances in genomic assay technologies allow us to identify a range of diseases and disorders, including Mendelian, chromosomal, and multifactorial. However, scientists rely on available genetic and healthcare data to interpret this information and draw conclusions. People with well-represented lineages are more likely to get a correct diagnosis and a better treatment regimen based on their genomic markers. Currently, the dominantly European genomic dataset limits the accuracy of gene validity and variant interpretation, hindering our use of genomic medicine for worldwide populations. Without greater diversity in this genomic data, healthcare system disparities may be further heightened. By including diverse populations in research through initiatives like the All of Us Research Program and engaging with other countries and ethnicities to generate genomic data, precision medicine holds the potential to become more beneficial globally.

Evidence-Based Variant Classification with Predominantly European Data

Scientists interpret genetic findings by comparing them to the prevalence of specific variants in the population through genetic studies, including genome-wide association studies (GWAS) and other experimental evidence. A majority of genomic data comes from research participants and patients of European ancestry; about 78% of GWAS participants and 54% of disease associations come from European descent(2). Although primarily beneficial to populations with European ancestry, these genetic findings have been useful overall: 3,000 genes have been reported in association with at least one Mendelian disease(12). The ClinVar database classified 55.8% of observations from the clinically relevant variants among European ancestral

populations as pathogenic or likely pathogenic (11). However, in an ExAC database of 61,486 individuals, only seven individuals of South Asian origin were identified with a mutation in *MUTYH*. This variant was classified as a variant of unknown significance due to the predominantly European-descent dataset. Without the South Asian population genomic data, it is unclear if the variant is a pathogenic founder mutation for this specific population(14). Patients who belong to underrepresented groups in genomic data face ambiguous genetic test results and interpretation, including many variants of unknown significance(12).

Benefits of Increasing Diversity in Genomic Data

Increasing diversity in genomic data holds the potential to benefit future genetic research on many levels, from more accurate disease-gene associations to more equitable preventive healthcare. Misinterpreting gene validity in the absence of curated health data results in clinical consequences for non-European patients. One example of this is the association of PCSK9 loss of function mutations with lower cholesterol levels and low coronary heart disease risk in African Americans. In contrast, data from individuals of mainly European descent classified the same mutations as highly pathogenic for hypertrophic cardiomyopathy, a clinically actionable disease. This data suggests that limiting studies to a single ancestry group restricts the utility of findings for non-European populations(1). Furthermore, it restricts the identification of new disease-variant associations, which are often dependent on allele frequencies in specific populations, as seen with the association of variants in the gene KCNQ1 and Type 2 Diabetes Mellitus(T2DM) in a South East Asian population. The identified pathogenic variants(rs2237897 and rs2237892) have a higher minor allele frequency(0.39 and 0.38) in comparison to European populations(0.04 and 0.06). Researchers would need a larger cohort to identify the association

based on the minor allele frequency of European populations for this gene-disease association(2). Additionally, asthma-related deaths are around five times higher in individuals with African, Puerto Rican, and Mexican ancestry. By studying genetic variants in these populations, scientists found that these individuals had a decreased sensitivity to a common inhaler drug called albuterol(6). Considering this, genomic research must include more diverse populations, as studying and including their data results in more equitable clinical care, identification of novel drug targets, and better prediction of disease risks in populations.

Call to Action

While there has been ongoing progress to incorporate more diverse data sets in genomics, there is still a significant lack of representation for various populations(15). Researchers across the globe should follow the lead of the *All of Us Research* Program, a NIH program with an ambitious plan to build one of the most diverse databases in history by sequencing one million people in the United States. Learning from the participant engagement strategies of this program and building focused consortiums on minority populations can help other groups globally.

References:

1. Hindorff, Lucia A et al. “Prioritizing diversity in human genomics research.” *Nature reviews. Genetics* vol. 19,3 (2018): 175-185. doi:10.1038/nrg.2017.89

2. Gurdasani, Deepti et al. “Genomics of disease risk in globally diverse populations.” *Nature reviews. Genetics* vol. 20,9 (2019): 520-535. doi:10.1038/s41576-019-0144-0
3. Wojcik, Genevieve L et al. “Genetic analyses of diverse populations improves discovery for complex traits.” *Nature* vol. 570,7762 (2019): 514-518. doi:10.1038/s41586-019-1310-4
4. Clyde, Dorothy. “Making the case for more inclusive GWAS.” *Nature reviews. Genetics* vol. 20,9 (2019): 500-501. doi:10.1038/s41576-019-0160-0
5. Tsosie, Krystal S et al. “Overvaluing individual consent ignores risks to tribal participants.” *Nature reviews. Genetics* vol. 20,9 (2019): 497-498. doi:10.1038/s41576-019-0161-z
6. Mak, Angel C Y et al. “Whole-Genome Sequencing of Pharmacogenetic Drug Response in Racially Diverse Children with Asthma.” *American journal of respiratory and critical care medicine* vol. 197,12 (2018): 1552-1564. doi:10.1164/rccm.201712-2529OC
7. Mallick, Swapan et al. “The Simons Genome Diversity Project: 300 genomes from 142 diverse populations.” *Nature* vol. 538,7624 (2016): 201-206. doi:10.1038/nature18964
8. Visscher, Peter M et al. “10 Years of GWAS Discovery: Biology, Function, and Translation.” *American journal of human genetics* vol. 101,1 (2017): 5-22. doi:10.1016/j.ajhg.2017.06.005
9. Strande, Natasha T et al. “Navigating the nuances of clinical sequence variant interpretation in Mendelian disease.” *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 20,9 (2018): 918-926. doi:10.1038/s41436-018-0100-y
10. Green, Robert C et al. “ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.” *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 15,7 (2013): 565-74. doi:10.1038/gim.2013.73
11. Popejoy, Alice B et al. “The clinical imperative for inclusivity: Race, ethnicity, and ancestry (REA) in genomics.” *Human mutation* vol. 39,11 (2018): 1713-1720. doi:10.1002/humu.23644
12. Strande, Natasha T et al. “Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource.” *American journal of human genetics* vol. 100,6 (2017): 895-906. doi:10.1016/j.ajhg.2017.04.015
13. Riggs, Erin R et al. “Copy number variant discrepancy resolution using the ClinGen dosage sensitivity map results in updated clinical interpretations in ClinVar.” *Human mutation* vol. 39,11 (2018): 1650-1659. doi:10.1002/humu.23610
14. Wright, Caroline F., et al. "Genomic Variant Sharing: A Position Statement." *Wellcome Open Research*, vol. 4, 4 Dec. 2019, p. 22, doi:10.12688 wellcomeopenres.15090.2.
15. McGuire, Amy L et al. “The road ahead in genetics and genomics.” *Nature reviews. Genetics* vol. 21,10 (2020): 581-596. doi:10.1038/s41576-020-0272-6

16. Laurencin, Cato T, and Aneesah McClinton. “The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities.” *Journal of racial and ethnic health disparities* vol. 7,3 (2020): 398-402. doi:10.1007/s40615-020-00756-0
17. Williams, David R, and Lisa A Cooper. “Reducing Racial Inequities in Health: Using What We Already Know to Take Action.” *International journal of environmental research and public health* vol. 16,4 606. 19 Feb. 2019, doi:10.3390/ijerph16040606