

The Importance of Increasing Diversity in Genomics Research

Sachi Badola¹ and Bethany Zettler[#]

¹Clemsford High School, Clemsford, MA, USA

[#]Advisor

ABSTRACT

Purpose: Genomic diversity is important in medical research and precision medicine, as appropriate disease management and drug development heavily utilizes genomic data. This article sought out to present a narrative review of limitations of genomic medicine utilization in historically underrepresented populations. **Method:** A literature review was conducted, and relevant publications written on the lack of genomic diversity in research and medicine were chosen from PubMed. Additionally, the ClinVar, ExAc, and gnomAD databases were used to review variants that were reported as pathogenic or non-pathogenic in different ethnicities, with ethnicity dependent minor allele frequencies. A total of twenty-five publications were used and examined to identify common themes. **Results:** Twenty-seven publications were chosen to be a part of this review, and based on the content of the publications, three different categories of genomics utilization with a problematic lack of diversity were developed: “Genes and Disease Correlation”, “Disease Management”, and “Drug Development”. The literature review also identified possible solutions and interventions for researchers and physicians to increase genomic diversity in order to ensure that the future of precision medicine is equitable for all populations. **Conclusion:** The results of this paper highlight that there are currently significant limitations to genomic medicine for underrepresented populations. By engaging, educating, and building trust with underrepresented communities, precision medicine can overcome the current lack of genomic diversity. This review recognizes the need for large multi-ethnic population studies, community representation in research, and understanding the interplay of genes, lifestyle, and environment for minorities.

Introduction

Within the United States healthcare system, existing racial disparities cannot go unaddressed. The root of racial disparities is racism, which constructs unequal access to healthcare for certain communities. The unprecedented coronavirus pandemic proves the urgency of the matter, as historically marginalized groups and lower socioeconomic status groups are disproportionately burdened by the health and social impacts of the disease. Studies from the National Academy of Medicine emphasize that for almost all therapeutic interventions, including diagnostic and treatment interventions, African American and other minority groups receive a lower quality of healthcare in comparison to white communities (Williams et al 2019). This discrepancy applies to precision medicine as well: if genomics fails to incorporate underrepresented population’s genetic data, the current advancements in genetics may disproportionately impact ethnic minority groups. 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 world-wide 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 alleviate racial disparities in healthcare.

Methods

A literature review was conducted, and relevant publications written on the lack of genomic diversity in research and medicine were chosen from PubMed. Additionally, the ClinVar, ExAc, and gNOMAD databases were used to review variants that were reported as pathogenic or non-pathogenic in different ethnicities, with ethnicity dependent minor allele frequencies. In PubMed, there were 51,970 papers about human genomics medicine from 2013 to the present. By narrowing the search to clinical trials, commentaries, editorials, journal articles, meta-analyses, and reviews using the keywords “human genomics medicine” and “genomic diversity” from 2013 to the present, there were 3,224 relevant publications with free full text available. After adding the keywords “lack of genomic diversity for diverse and minority populations” to the search, 151 publications were selected. From those 151, a total of twenty-four publications were used and examined to identify common themes. This review developed three different categories of genomics utilization: “Genes and Disease Correlation” (10 publications), “Disease Management” (9 publications), and “Drug Development” (5 publications).

Results

The literature review identified the genomic data utilization across medical research and the impacts of a lack of genomic diversity. The literature review also identified various opportunities to increase the diversity and inclusiveness of all ethnicities in medical research and precision medicine by advancing education, awareness and support for various consortiums as well as introducing genomics literacy as early as high school. These genomics data barriers for minority and possible solutions will be discussed in this review with supportive literature evidence.

Genes and Disease Correlation

In this category, there are 10 relevant publications that explore how a lack of genomic diversity impacts gene and disease correlation, revealing the disadvantage for research participants and patients who come from a historically underrepresented background. The papers include examples of current limitations of genomic medicine and pathogenic variant distribution across populations.

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 (Gurdasani et al. 2019). 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 (Strande et al. 2017). The ClinVar database classified 55.8% of observations from the clinically relevant variants among European ancestral populations as pathogenic or likely pathogenic (Popejoy et al. 2018). However, in an ExAC database

of 61,486 individuals, only seven individuals of South Asian origin were identified with a mutation in *MUTYH*. The *MUTYH* gene provides instructions to make the enzyme MYH glycosylase, and this mutation causes *MUTYH* associated polyposis, increasing the risk for colorectal cancer. 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 (Wright et al. 2019). Patients who belong to underrepresented groups in genomic data, including African American, Latino and other minority populations, face ambiguous genetic test results and interpretation, including many variants of unknown significance (Strande et al. 2017).

Table 1. Examples of Genes with Pathogenic Variant Distribution Across Populations

Gene	Germline Variant SNP rsIDS	Disease Association based on European Data	Increased or Decreased Risk for a Minority Population	Minor Allele Frequency of Affected Population as compared to Caucasian Populations
ALDH2	rs671	Variant almost absent	Increased risk of hypertension for East Asians	East Asian: 0.2554 Caucasian: 0.00002404
APOL1	rs73885319	Kidney Disease and End Stage Renal Disease	7-10 fold increased risk for African Americans	African/African American: 0.2276 Caucasian:: 0.0001084
KCNQ1	rs2237897	Risk for Type 2 Diabetes Mellitus(T2DM)	Increased Risk for Type 2 Diabetes Mellitus(T2DM)	South East Asian: 0.3342 Caucasian: 0.04244
	rs2237892			South East Asian: 0.3297 Caucasian: 0.06291
PCSK9	rs67608943	Increased Risk for Hypertrophic Cardiomyopathy	Lower risk of coronary heart disease in African Americans	African American: 0.002364 Caucasian: 0.000

Table 1 summarizes four gene examples found in the literature review, in which germline variants in the gene are associated with disease for a Caucasian population. However, this literature review highlights that only after analyzing data from underrepresented populations (African American, East Asian, and Latino), different correlations were made between specific variants and their effect on the respective minority population. Certain variants were more prevalent in underrepresented populations, increasing risk or decreasing risk for associated diseases as compared to Caucasian populations. The first example is the missense variant on the gene *ALDH2*, aldehyde dehydrogenase 2 family member, the second enzyme of the major oxidative pathway of

alcohol metabolism. This single nucleotide variant is almost absent in European populations, as the minor allele frequency(MAF) is 0.0001084. However, for East Asian populations(MAF: 0.2554), the minor allele frequency is significantly larger, resulting in the variant causing an increased risk of hypertension for East Asian populations after alcohol consumption. The second example includes a single nucleotide polymorphisms(SNPs) on the APOL1 gene, which encodes a high density lipoprotein which binds to apolipoprotein, involved in the formation of cholesteryl esters in plasma. The SNP is a missense variant on APOL1 that increases risk for kidney disease and end stage renal disease. However, the MAF for African American populations(0.2276) is greater than the MAF for European populations(0.0001084), and African American Populations have a significantly increased risk for kidney disease and end stage renal disease with this variant. The third example is the KCNQ1 gene, which provides instructions for the potassium voltage-gated channel subfamily Q member 1. Two SNV's in this gene(rs2237897 and rs2237892) are associated with risk for Type 2 Diabetes Mellitus for European populations with respective MAFs(0.04244 and 0.06291). However, the MAFs(0.3342 and 0.3297) for South East Asian populations are significantly greater, resulting in a greater risk for Type 2 Diabetes Mellitus.

The final example looks at the gene PCSK9, which provides instructions for creating proteins that regulate cholesterol levels in the blood and controls the amount of low-density lipoprotein receptors on the cell surface. The SNV rs67608943 is a stop-gain mutation that increases risk for hypertrophic cardiomyopathy in European populations. The MAF is very low in European populations(0.000), meaning that this variant is very rare. However, for African American populations, the MAF(0.002364) is considerably greater, and the variant is more common. In contrast to European populations, the variant has been associated with lower risk of coronary heart disease in African Americans.

These examples emphasize the importance of diversity in genomics research because without studies done on the minority populations, there would be a lack of accurate risk association and minor allele frequency distribution. Misinterpreting gene validity in the absence of curated health data results in clinical consequences for non-European patients. As seen in the above table, 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(Hindorff et al. 2018). 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(Gurdasani et al. 2019). 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.

Disease Management

Certain diseases have strong underlying genetic associations. A lack of genetic diversity results in an inaccurate assessment of risk and lack of effective interventions in under-studied populations. In this category, there are 9 relevant publications that emphasize the correlation between genomic diversity and effective disease management for Mendelian diseases and polygenic risk scores for underrepresented populations.

Mendelian Disease

There are many variants in genes that are commonly found in minority populations with a mendelian inheritance pattern. CFTR is an example of a Mendelian(single-gene) disorder that is often underdiagnosed in African Americans. In European populations, the most common allele associated with more than 70% of cystic fibrosis cases is $\Delta F508$ in the *CFTR* gene, but this allele causes only 29% of CF cases in populations of African ancestry(Sirugo et al. 2019). In African populations, the point mutation 3120+1G→A in the *CFTR* gene causes between 15-65% of CF cases (Sirugo et al. 2019). This proves that without more genomic information on diverse populations, it is difficult to accurately associate variants with disease for global populations. If there was no genomic information on people of African American descent, cystic fibrosis would be underdiagnosed because they have a different pathogenic variant in the *CFTR* gene. Disease management plans are often made based on pathogenic variants, which may differ among populations.

Another example of a mendelian disease that is affected by the lack of genomic diversity is sickle cell disease. However, sickle cell disease is most commonly caused by the same mutation in populations. The difference is the clinical presentation variability due to the allelic spectrum of other genetic modifiers across populations. One example of this is the population-specific variability of alpha-thalassemia, which can impact SCD presentation. The phenotypic presentation of alpha thalassemia is directly correlated to the number of alpha globin genes affected by various mutations/deletions. There are many different mutations and deletions that cause the changes in the alpha chain in the hemoglobin gene. This variability in the number of alpha chains in hemoglobin gene can present variability in clinical phenotype and thus affects the severity of the SCD. Sickle cell disease is a condition with autosomal recessive Mendelian inheritance that affects the red blood cells in the body, and it is caused by a missense mutation (Glu6Val) in the *HBB* gene. The *HBB* gene encodes a red blood cell protein hemoglobin, which binds oxygen and allows the cell to travel through the bloodstream. However, the pathogenic variant in the *HBB* gene results in sickle shaped red blood cells leading to decreased oxygen delivery and anemia. Sickle cell anemia disproportionately affects African Americans in comparison to other populations.

Unfortunately, sickle cell disease can have a mortality rate of up to 90% for African children below the age of 5, but the mutation is often maintained at a high frequency due to heterozygous individuals with sickle cell trait being protected from malaria (Sirugo et al. 2019). More genomic research needs to be done on diverse populations to understand the contributing factors that generate variability in clinical presentation of SCD to better manage the disease.

The Utilization of Polygenic Risk Scores

Unlike mendelian diseases and isolated candidate genes, polygenic risk scores (PRS) are utilized as predictive and reliable scoring methods for complex genetic traits, including diabetes and schizophrenia(Slunecka et al. 2021). Disease understanding is based on multiple variants with PRS, and the scores often offer more quantifiable genetic risk information on a patient than subjective family history alone (1000 Genomes Project Consortium, Auton et al. 2015). These scores are based on variants with a low effect size and calculated by “summing risk alleles, which are weighted by effect sizes derived from GWAS results”(Duncan et al. 2019). Non-European populations have different “variant frequencies and linkage disequilibrium patterns”, so as the effect size of a variant decreases, it is less transferable to other populations (Duncan et al. 2019). Thus, polygenic risk scores calculated from a GWAS of a largely European population may not be applicable to other populations and may underestimate or overestimate disease risk in understudied populations. A systematic evaluation of the lack of diversity in PRS usage highlights that only 3.8% of polygenic studies from the first 10 years of polygenic scoring research included African, Latino/Hispanic, or Indigenous people. (Duncan et al. 2019).

Drug Development

In this category, five relevant publications were reviewed and selected with common themes of the need for more diverse genomic data in drug development. These papers also offered different solutions for increasing diversity in clinical trials and the need for broader GWAS analysis.

For more than 20 years, genomics has been an important tool in the process of drug development and approval, as it allows for gene target identification and validation. Genetically validated targets are more successful in clinical drug trials, and there are various genetic data sources being utilized for the target identification and drug development. These data sources include the UK biobank, OMIM, gnomAD, GTEX, Human Cell atlas, CRISPR KO, LINCS, and Open target platform (Spreafico et al. 2020). However, a lack of diversity in these data sources results in challenges to globalize drug utilization for all populations.

Target Identification

In the context of genomics, the process of target identification includes utilizing genomics and proteomics data and correlating it with the associated disease models and tissues to identify drug targets. During the target identification, a process of accumulating sufficient evidence for the target validation is needed. These data sources provide the genetic evidence of association between gene and target and further validate whether the gene target is pertinent to the disease tissue and disease pathogenesis. As part of the drug discovery process, the gene target needs to be evaluated for druggability and consequences of long-term action and safety. The data sources that are utilized for a target characterization and validation are supported by a significantly high percentage of European data. Some examples of the lack of diversity in these commonly used data sources include: the UK Biobank has 94.6% White and 1.6% Black genetic data. FINNGEN has 99% White and <0.1% Black.

NHGRI-EBI GWAS Catalog provides clear insights into the ancestry distribution among individuals, studies, and disease associations. GWAS is utilized to identify a gene target for the drug development process. As shown in this GWAS analysis, approximately 78% of data sources are from individuals of European ancestry, and only 2.4% from African ancestry. Most studies with gene association analysis are done with these data sources, meaning that 49% studies are from European datasets and only 2.8% studies are from African datasets. In this study the disease associations were performed with the data sets showing relevant associations in 54% of the European population. They also observed the disproportionate contribution of associations from African (7%) and Hispanic/Latin American (4%) categories, when compared to the percentage of individuals (2.4%) and studies (2.8%).(Spreafico et al. 2020)

Based on the genomics data and catalogs created by the US National Human Genome Research Institute in partnership with the European Bioinformatics Institute, GWAS analysis identified 25% of the variants in European Americans as being associated with body mass index, type 2 diabetes, and lipid levels. But the strength of the association differs in at least one out of five populations of non-European ancestry. This means that a variant that is associated with diabetes may confer a different risk of disease in someone of European ancestry than in an individual of African ancestry. These examples continue to reiterate the importance of increasing diversity in data collection from all populations. Assessing the accuracy and broader relevance of findings with GWAS analysis in other populations is crucial. One of the major challenges in GWAS analysis with existing data sources is the false positive in underrepresented populations. It is important to replicate a study across various populations for validating a gene/variant disease association and identifying easily overlooked insights (Popejoy et al. 2018).

Case Study : Pharmacogenetic drug response in ethnically diverse patients

A whole-genome sequencing pharmacogenomics study was conducted by the New York Genome Center on 1,441 racially diverse children with asthma to identify genetic variants influencing bronchodilator drug response (BDR). (6) Asthma-related deaths are around five times higher in individuals with African, Puerto Rican, and Mexican ancestry. By studying genetic variants in these populations, researchers found that these

individuals had a decreased sensitivity to a common inhaler and bronchodilator drug called albuterol(6). Complex disease traits are influenced by environmental, social, and genetic factors. The example of BDR as a complex disease trait shows heritability estimates ranging from 47-92%. Several Genome-wide association analyses shows the correlation of BDR with several common SNPs in the European populations. There was only one GWAS analysis done with the African American population that identified a novel SNP which did not replicate in the European population. The authors made significant efforts to utilize age and ethnic-matched cohorts for the analysis. Lack of such cohorts in minority populations posed significant challenges to the analysis performed to identify rare and common SNPs. Authors identified BDR associated common and rare variants in three racially and ethnically diverse populations of children with asthma. They argue that in the era of precision medicine, the impact of genetic variations on drug response, treatment regimen and disease prevention has to be studied across all ethnic populations. It is important to realize the potential of precision medicine globally (Schärfe et al. 2017).

Clinical Development

In the US, clinical trials that are designed to test the efficacy, adverse events and long-term benefits are skewed with the homogenous white population. There is a lack of diversity in these trials, as generalizing the findings poses a significant risk to ethnic populations. According to the US census bureau, the population race-ethnic profile is comprised of ~40% of individuals who can be categorized as non-white or people of color and ~60% are non-Hispanic white. Based on the data reported from FDA clinical trials, 80-90% of the participants are white. It is clear from these datasets that the drug design to development is skewed towards the white population and based on our understanding of the underpinnings of genetics. According to pivotal trials on PubMed and Clinical Trials portal, most cardiovascular and diabetes medications have been approved with predominantly white participants, including Clevidipine, a hypertension drug, and Saxagliptin, a diabetes drug.

Discussion

In an era of precision medicine, it is essential to improve diversity in genomic research and medicine. Precision medicine allows for a personalized approach to disease treatment and prevention that identifies variability in genes, environment, and lifestyle for each patient. Therefore, doctors will determine what effective treatments and prevention strategies for a particular disease will work in which groups of people with greater accuracy. Contrasting the one-size fits all approach, where disease treatment and prevention strategies are developed with less consideration for variability between individuals, precision medicine involves the omics profile of a patient. Currently, precision medicine may not be equitable for historically underrepresented populations. In this review, three categories of genomics utilization affected by a lack of genomic diversity were created: “genes and disease correlation”, “disease management”, and “drug development. The relevant selected literature offered solutions to improve genomic diversity in genomics research through recruitment of diverse participants, education and training, building trust, and considering the gene-environment interplay.

Recruiting Diverse Participants

An essential solution to increasing diversity within genomics research is recruiting diverse participants. To conduct this ethically, researchers must include an ethics review board consisting of community leaders and stakeholders to review at all stages of research. This will allow researchers to gain trust of historically underrepresented populations and collect more diverse data and more accurate gene and disease correlation. Furthermore, researchers must increase access to research tools for populations of low-income economies in remote areas by allowing satellite infrastructure, such as covering transportation fees for traveling and training local healthcare professionals working directly with participants.

Considering Gene-Environment Interplay

There is a causal correlation of social factors with ancestry which influences the genetically dependent complex diseases. Genetic research remains very focused on the genotypic correlation with disease pathogenesis and often lacks the gene-environment interplay. The solution to address this major drawback is to engage with behavioral scientists and geneticists, to benefit diverse individuals from the accurate interpretation of polygenic risk scores. Considering the social factors related to ancestry, researchers must run analysis utilizing GWAS and PheWAS studies for diverse populations. It is evident that environmental factors, such as socioeconomic status and education, influence phenotype. There have been various approaches taken to support the implementation and utilization of polygenic risk scores across the diverse populations. Well-powered GWAS approaches with specific populations means that the databases with large numbers of each ethnicity need to be built, including ad-mix and ancestrally homogenous populations.

The coming decades will likely see the further expansion of genetic and phenotypic data collection to improve and expand PRSs for multiple ancestry populations and the diseases within those populations. Potentially, ancestry agnostic PRSs will be developed given enough subjects from diverse ancestry cohorts. This set of universal PRSs may perform better than ancestry-specific scores because they more closely approach the true genetic risk and reduce the amount of biases, such as overfitting.

Education and Training

Physician training is essential when working with PheWAS and GWAS data and PRS. Continuing medical education (CME) courses on PRS and bioinformatics tools can be an excellent supplement to medical education, aiding physicians to improve prevention and disease management for patients. Additionally, by enhancing the roles and responsibilities of genetic counselors, they can work in partnership with physicians on risk-assessment, discussing how the technology works, how the PRS should be interpreted, and how the data should be used to optimize disease management. In addition, researchers and key stakeholders must communicate with communities and create educational resources for the public. By improving genomic literacy, especially for communities with limited resources and scientific knowledge, researchers can break educational barriers and improve public health.

Limitations and Implications

While the field of genetics and genomics offer possible solutions to limit racial health disparities, further efforts outside of genomics must be made to reform the healthcare system. Precision medicine includes genetics, environment, and lifestyle. By teaching medical students about health equity and population health, future physicians will be better equipped on how to care for specific communities and ethnicities and provide equitable care for all. Furthermore, hospitals and clinics across the nation should implement training programs and workshops that discuss ways to eliminate implicit bias among healthcare providers. Additional research must be done on the implications of diverse datasets on current patient populations, such as Caucasians. There may be limitations created by adding diverse data into existing datasets, making them more heterogeneous, as well as the risk of less treatment reliability for Caucasian populations. More research must be done on if creating separate homogeneous genomic databases for all ancestries can be a solution.

Conclusion

While genetics and genomics offers many benefits for population health and precision medicine, a lack of significant efforts to eliminate racial health disparities will put minority groups at a further disadvantage. The fields of genetics and genomics have a responsibility to ensure that the benefits of precision medicine are equitable and significant for all ethnicities within the United States. 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. This review article explores impact of a lack of genomic diversity on the utilization of genomic medicine across three different categories of genomics utilization. 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 in the United States.

Limitations

This review paper has many limitations, including a one-year time restraint and basing conclusions off of chosen articles found in the literature and public PubMed domain. Additionally, it is important to acknowledge that there are other social determinants of health that affect underrepresented minorities beyond the scope of this paper.

Acknowledgments

I would like to thank my mentor at Genomes2People Research Program: Bethany Zettler, MS, CGC for guiding me throughout the entire process of this research and helping me advance my review article. I would also like to thank Dr. Robert Green, the director of the Genomes2People Research Program, for enhancing my understanding on this research topic and giving me the opportunity to work as a trainee in his lab. This study has not received any funding, and the author and contributors do not have any conflict of interest.

References

- Bentley, A. R., Callier, S., & Rotimi, C. N. (2017). Diversity and inclusion in genomic research: why the uneven progress?. *Journal of community genetics*, 8(4), 255–266. <https://doi.org/10.1007/s12687-017-0316-6>
- Clark, L. T., Watkins, L., Piña, I. L., Elmer, M., Akinboboye, O., Gorham, M., Jamerson, B., McCullough, C., Pierre, C., Polis, A. B., Puckrein, G., & Regnante, J. M. (2019). Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. *Current problems in cardiology*, 44(5), 148–172. <https://doi.org/10.1016/j.cpcardiol.2018.11.002>
- 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
- Cooke Bailey, J. N., Bush, W. S., & Crawford, D. C. (2020). Editorial: The Importance of Diversity in Precision Medicine Research. *Frontiers in genetics*, 11, 875. <https://doi.org/10.3389/fgene.2020.00875>

- Duncan, L., Shen, H., Gelaye, B. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* 10, 3328 (2019). <https://doi.org/10.1038/s41467-019-11112-0>
- 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
- 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
- Luther T. Clark, Laurence Watkins, Ileana L. Piña, Mary Elmer, Ola Akinboboye, Millicent Gorham, Brenda Jamerson, Cassandra McCullough, Christine Pierre, Adam B. Polis, Gary Puckrein, Jeanne M. Regnante, Increasing Diversity in Clinical Trials: Overcoming Critical Barriers, Current Problems in Cardiology, Volume 44, Issue 5, 2019, Pages 148-172, ISSN 0146-2806, <https://doi.org/10.1016/j.cpcardiol.2018.11.002>.
- 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
- 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
- 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
- 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
- Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, 538(7624), 161–164. <https://doi.org/10.1038/538161a>
- 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
- Schärfe, C.P.I., Tremmel, R., Schwab, M. *et al.* Genetic variation in human drug-related genes. *Genome Med* 9, 117 (2017). <https://doi.org/10.1186/s13073-017-0502-5>
- Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*, 177(1), 26–31. <https://doi.org/10.1016/j.cell.2019.02.048>
- 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

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

Slunecka, J. L., van der Zee, M. D., Beck, J. J., Johnson, B. N., Finnicum, C. T., Pool, R., Hottenga, J. J., de Geus, E., & Ehli, E. A. (2021). Implementation and implications for polygenic risk scores in healthcare. *Human genomics*, 15(1), 46. <https://doi.org/10.1186/s40246-021-00339-y>
Spreatico, R., Soriaga, L. B., Grosse, J., Virgin, H. W., & Telenti, A. (2020). Advances in Genomics for Drug Development. *Genes*, 11(8), 942. doi:10.3390/genes11080942

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

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

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

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.

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>