Check for updates

Identification of genetic correlates of coronary artery disease in diverse ancestral populations

The largest and most genetically diverse genome-wide association study thus far for coronary artery disease, a leading cause of death worldwide, identifies many new susceptibility loci and characterizes the genetic architecture in Black and Hispanic populations for the first time.

The problem

Considerable progress has been made in understanding the genetic factors that contribute to coronary artery disease (CAD) among people of European ancestry[1]. Several relevant loci discovered in this population have also reached genome-wide significance in South and East Asian populations, suggestive of an overlap in the genetic architecture of CAD across these three populations[2,3]. However, 15 years after the discovery of the first CAD susceptibility locus at 9p21, no genome-wide associations have been reported among Black or Hispanic populations. These discrepancies in discovery are largely a consequence of the lack of well-powered studies. This disparity in genomic studies of CAD has the potential to further hinder the benefit of precision medicine for a population already disproportionately affected by the disease burden[4]. New DNA biobanks that have enrolled diverse populations are poised to fill this knowledge gap.

The solution

We analyzed large-scale genetic data available from over 400,000 participants in the Million Veteran Program (MVP), including nearly 118,000 people with CAD. The MVP is a nationwide cohort drawn from an integrated healthcare system that serves a diverse population, including many Black and Hispanic people (Fig. 1a). These data were compared and combined with existing large-scale genome-wide association studies (GWASs) of white and Japanese populations, as well as smaller-scale studies in Black and Hispanic populations (Fig. 1a). We first estimated and compared the heritability for CAD across multiple ancestral populations. We then combined data through meta-analysis to discover new loci both within each population and across all populations. These analyses enabled us to document similarities and differences in the genetic architecture of CAD across populations. Finally, we tested the performance of existing and new polygenic risk scores derived from the combined dataset.

We found that the heritability of CAD was similar across four ancestral populations, including white, East Asian, African and Indigenous American populations, which suggested a roughly equivalent ratio of genetic determinants of disease to non-genetic determinants of disease in all populations. In addition, we found evidence that the biological underpinning of CAD is identical across all populations, through the observation that the first eight CAD susceptibility loci identified among non-Hispanic Black and Hispanic cohorts all overlapped established loci in European or East Asian populations. One important difference between African and non-African populations we identified involved the well-known 9p21 susceptibility locus, for which we found a nearly complete absence of the risk-stratifying alleles among African chromosomes. Meta-analysis of all GWASs revealed 95 new loci that reached genome-wide significance, and approximately half of these still seemed to mediate risk through well-established clinical risk factors. Existing polygenic risk scores for CAD contain strong readouts for all traditional risk factors and track with burden of disease but perform poorly when transferred to Black populations. Although performance of scores improved in all populations through the addition of non-European GWAS data, a substantial gap persisted between Black populations and non-Black populations (Fig. 1b).

Future directions

Our findings reinforce those of other multi-population or multi-ancestry genetic studies that demonstrate the remarkable value of the diversification of genetic datasets[5]. This value comes in multiple forms, including the more reliable identification of causal variants within established loci, the identification of novel loci or novel regions of susceptibility within established loci, and the improved performance of polygenic risk scores. However, further large studies of under-represented populations are needed to overcome the inequities that have arisen over the past 15 years of GWASs. We plan to meet this challenge by prioritizing the careful analysis and reporting of additional participants in the MVP with genetic data, including thousands of Black and Hispanic veterans. We welcome continued collaboration with researchers with access to diverse biobanks around the world and/or who have developed novel statistical methods that help to both overcome the challenges and take advantage of the study of admixed populations.

Catherine Tcheandjieu[1] and Themistocles L. Assimes[2]
[1]Gladstone Institutes, San Francisco, CA, USA.
[2]Stanford University School of Medicine, Stanford, CA, USA.

EXPERT OPINION

> The authors have performed the largest multi-ancestry GWAS of coronary disease to date. I think this paper makes a significant advance on previous GWASs of coronary artery disease (CAD) with identification of 95 new CAD-associated loci, representing an approximately 30% increase in CAD loci, and report the first CAD associations on chromosome X, which may in the future help explain sex differences in CAD rates." **Joanna Howson, Novo Nordisk, Oxford, UK.**
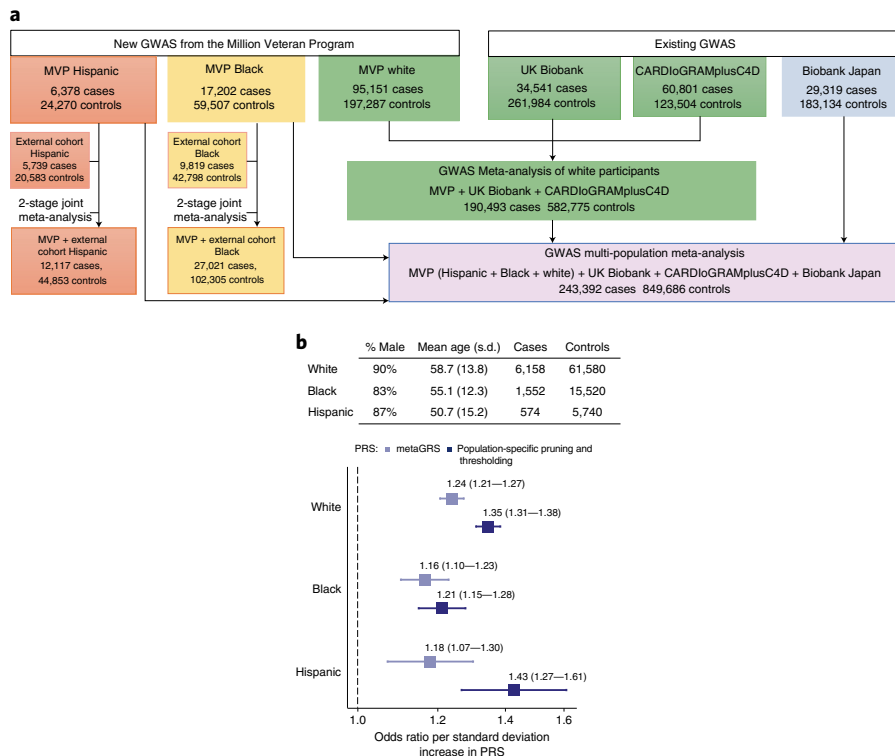
FIGURE

Fig. 1 | Design of and findings from a large-scale genetically diverse GWAS of CAD. a, Cohorts that contribute to discovery through GWASs, meta-analyses, and two-stage analyses. **b**, Performance of polygenic risk scores (PRS) before and after the addition of GWASs for Black, Hispanic and Japanese populations. © 2022, Tcheandjieu, C. et al.

BEHIND THE PAPER

This study was initiated in 2017, at the time of the first release of genetic data by MVP to funded researchers. Substantial challenges were imposed by the initial computing environment for the analysis of large-scale genetic data, followed by the COVID-19 pandemic. However, a strong collaborative spirit and the availability of exceptional multi-disciplinary expertise in bioinformatics, statistical genetics, genetic epidemiology and electronic health record phenotyping both within and outside the US Department of Veterans Affairs were instrumental in successful troubleshooting and carrying this project to its completion. The absence of a signal at 9p21 in both Black cohorts and Hispanic cohorts perplexed many of us. The breakthrough came with the lead author's idea to implement an 'old' approach to shed light on a new enigmatic observation. When an optimal set of genetic variants for building haplotypes was combined with local ancestry analysis and haplotype trend regression, the reasons behind the absence of a signal became very clear to all involved. **T.L.A.**

REFERENCES

1. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018). **This paper represents the largest published GWAS of CAD so far, with the vast majority of participants studied being of European ancestry.**

2. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–344 (2011). **This paper reports the largest GWAS of CAD among a South Asian population and demonstrates high concordance with findings in European populations.**

3. Koyama, S. et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177 (2020). **This paper reports the largest GWAS of CAD among people of East Asian ancestry so far and demonstrates high concordance with findings in European populations.**

4. Clarke, S. L., Assimes, T. L. & Tcheandjieu, C. The propagation of racial disparities in cardiovascular genomics research. *Circ. Genom. Precis. Med.* **14**, e003178 (2021). **This review summarizes the disparities in cardiovascular genomics research so far, including CAD.**

5. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021). **This paper shows how the study of diverse populations improves the understanding of the biology of blood lipid levels, as well as the ability to predict the levels of 'bad' cholesterol.**

FROM THE EDITOR

> Because GWASs have historically focused on European populations, there is a pressing need for large-scale genetic studies of other populations. This work is very timely in helping to fill this gap by providing a large-scale GWAS of CAD in a cohort of people with a large fraction of Black and Hispanic participants, through the use of the MVP biobank." **Editorial Team,** *Nature Medicine*