Genetics

Missing heritability found for height

Karoline Kuchenbaecker

An combined analysis of 281 genome-wide association studies finds 12,111 common DNA variants associated with a person's height — and shows that larger studies will not yield more variants in populations of European ancestry.

The largest human genetic study to date has just been completed. Writing in *Nature*, Yengo *et al.*[1] describe how they have used genetic data from almost 5.4 million people to identify more than 12,000 DNA variants associated with differences in height between individuals. Remarkably, they show that they have reached the point at which more data will not reveal more variants — at least for some populations. This feat marks a milestone in our understanding of the contribution of genetics to complex traits. It also as highlights the essential work still to do to close the diversity gap in existing genetic data.

Some disorders, such as Huntingtin's disease, are caused by changes in one or only a few specific genes. Once the genes have been identified, the cause of the disorder is seen as resolved (although there can be more complexity[2]). Genome-wide association studies (GWAS) were designed in 2007 to identify the genetic causes of more-complex diseases, such as coronary artery disease or bipolar disorder. GWAS use an approach called genotyping to compare the DNA of thousands of individuals, enabling researchers to identify genetic variants associated with differences in a trait of interest. Indeed, GWAS have identified many disease-associated variants and helped uncover the underlying biology.

However, variants from these early GWAS could explain only 5% of the variation in height between people, even though height is largely determined by genetic factors, sparking a debate about the 'missing heritability. Hypotheses to explain this included undetected common variants of small effects, rare and structural variants as well as gene–gene interactions and overestimation of the heritability estimates[4]. Part of the explanation is that extremely large numbers of hitherto unidentified genetic variants with small effects contribute to each complex trait[3]. Some of the genetic effects are so tiny that it seemed unrealistic to ever accumulate

sufficient sample sizes to pinpoint these variants. Consequently, a full picture of the genetics underlying any complex trait seemed out of reach.

Since then, technological advances have resulted in genotyping of millions of human genomes across thousands of GWAS. Often, these studies involved collecting a range of information about participants – commonly including their height. This wealth of data makes height a good model for studying the genetics of complex traits. Yengo *et al.* therefore asked whether a large enough sample size could reveal the missing heritability for height. In an immense collaborative effort, they combined data from 281 different studies, predominantly including participants of European ancestry, and identified 12,111 genetic variants that affect human height.

Yengo and colleagues showed that these variants are clustered into small genomic segments that cover 21% of the genome. This confirms earlier predictions that a large proportion of the human genome is functionally involved in shaping height. The authors demonstrated that the GWAS variants can explain about 40% of the heritability of height. They also found that the genomic segments they identified account for the entire heritability of height in participants of European ancestry — at least, for the common variants assessed using genotyping (some rare variants can only be found using sequencing approaches). Finally, the group used their variants to develop a genetic risk score, which they showed could explain 40% of the variation in height of a separate cohort of individuals of European ancestry, from the UK Biobank. This matches their heritability figure of 40%, meaning the score cannot be further improved.

Hence, the missing heritability of height has finally been found. More genetic studies in people of European ancestry will not contribute any more information — saturation has been achieved for this trait in this ancestry group. This astonishing result will now motivate efforts to achieve saturation for other traits and diseases. Once achieved, attention may shift more widely to the momentous task of identifying the genes that underlie these associations and inferring mechanisms for some of them.

Efforts to achieve saturation for complex diseases will face challenges. Height is a relatively stable trait that can be measured easily and reliably. By contrast, it can be difficult to determine complex disease outcomes, such as major depressive disorder. Low disease prevalence can make it hard to generate large-enough sample sizes, and some diseases might be affected by more genes than is height. But more concerted international efforts to overcome these challenges are a likely outcome of Yengo and colleagues' work.

It is important to acknowledge, as the authors do, that saturation is currently limited to only one ancestral group. Yengo *et al.* did not collect their own data, but made use of existing datasets, in which diversity is limited. Out of the 5.4 million samples that they included in their study, 4 million were from participants of European ancestry, and less than 78,000 from participants of South Asian ancestry.

There is also a lack of global diversity. Amongst the studies contributing data from participants of African descent, to the best of my knowledge only one is from the African continent, with the rest coming from the diaspora. This single study, conducted in Nigeria, had 1,188 participants, implying that only 0.4% of the participants of African descent included in the GWAS were from Africa. The genetic and linguistic diversity of Africa is immense[5–7]. Although participants from the African diaspora add valuable data, these groups do not capture Africa's diversity. Thus, this study, like other GWAS, did not comprehensively cover human genetic variation[8]. Some of the genetic variants associated with height will have been missed.

Yengo *et al.* showed that, for participants of African ancestry from the USA and the UK, the identified genetic variants could explain only 5-12% of variation in height. This is in line with other research demonstrating that GWAS based on people of European descent can less accurately predict gene–trait relationships in other groups[10]. Moreover, a study published earlier this year demonstrated that across different sub-Saharan African population, the performance of genetic risk scores based on African American GWAS varied widely [9]. Understanding how environmental factors might affect gene–trait relationships[11] will require more global diversity in genetic research.

Yengo *et al.* have demonstrated that it is possible to achieve saturation for complex traits. Now, ancestrally, ethnically, globally and socioeconomically diverse samples are needed to truly reap the full benefits of GWAS.

Karoline Kuchenbaecker is in the Division of Psychiatry and the UCL Genetics Institute, University College London, London WC1E 6BT, UK.
e-mail: k.kuchenbaecker@ucl.ac.uk

1. Yengo, L. *et al. Nature* XXX (2022).
2. Barton, A. R., Hujoel, M. L. A., Mukamel, R. E., Sherman, M. A. & Loh, P.-R. A spectrum of recessiveness among Mendelian disease variants in UK Biobank. *Am.*

J. Hum. Genet. (2022) doi:10.1016/j.ajhg.2022.05.008.

3. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
4. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
5. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
6. *Assessing rare variation in complex traits: Design and analysis of genetic studies*. (Springer, 2015).
7. Fan, S., Hansen, M. E. B., Lo, Y. & Tishkoff, S. A. Going global by adapting local: A review of recent human adaptation. *Science* vol. 354 54–59 (2016).
8. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nature Medicine* vol. 28 243–250 (2022).
9. Kamiza, A. B. *et al.* Transferability of genetic risk scores in African populations. *Nat. Med.* **28**, 1163–1166 (2022).
10. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics* vol. 51 584–591 (2019).
11. Giannakopoulou, O. *et al.* The Genetic Architecture of Depression in Individuals of East Asian Ancestry: A Genome-Wide Association Study. *JAMA Psychiatry* **78**, 1258–1269 (2021).

The author declares no competing interests.