



## The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

If you want to predict how tall your children might one day be, a good bet would be to look in the mirror, and at your mate. Studies going back almost a century have estimated that height is 80–90% heritable. So if 29 centimetres separate the tallest 5% of a population from the shortest, then genetics would account for as many as 27 of them<sup>1</sup>.

This year, three groups of researchers<sup>2–4</sup> scoured the genomes of huge populations (the largest study<sup>4</sup> looked at more than 30,000 people) for genetic variants associated with the height differences. More than 40 turned up.

But there was a problem: the variants had tiny effects. Altogether, they accounted for little more than 5% of height's heritability — just 6 centimetres by the calculations above.



Even though these genome-wide association studies (GWAS) turned up dozens of variants, they did “very little of the prediction that you would do just by asking people how tall their parents are”, says Joel Hirschhorn at the Broad Institute in Cambridge, Massachusetts, who led one of the studies<sup>5</sup>.

Height isn't the only trait in which genes have gone missing, nor is it the most important. Studies looking at similarities between identical and fraternal twins estimate heritability at more than 90% for autism<sup>5</sup> and more than 80% for schizophrenia<sup>6</sup>. And genetics makes a major contribution to disorders such as obesity, diabetes and heart disease. GWAS, one of the most celebrated techniques of the past five years, promised to deliver many of the genes involved (see ‘Where's the reward?’, page 20). And to some extent they have, identifying more than 400 genetic variants that

contribute to a variety of traits and common diseases. But even when dozens of genes have been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. “It is the big topic in the genetics of common disease right now,” says Francis Collins, former head of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The unexpected results left researchers at a point “where we all had to scratch our heads and say, ‘Huh?’”, he says.

Although flummoxed by this missing heritability, geneticists remain optimistic that they can find more of it. “These are very early days, and there are things that are doable in the next year or two that may well explain another sizeable chunk of heritability,” says Hirschhorn. So where might it be hiding?

ILLUSTRATIONS BY D. PARKINS



### Right under everyone's noses

The inability to find some genes could be explained by the limitations of GWAS. These studies have identified numerous one-letter variations in DNA called single nucleotide polymorphisms (SNPs) that co-occur with a disease or other trait in thousands of people. But a given SNP represents a much bigger block of genetic material. So, for example, if two people share one of these variants at a key location, both may be scored as having the same version of any height-related gene in that area, even though one person actually has a relatively rare mutation that has a huge effect on height. The association study might identify a variant responsible for the height difference, says Teri Manolio, director of the Office of Population Genomics at the NHGRI, but averaging across hundreds of people could give the appearance that its effects are pretty wimpy. "It's going to be diluted," she says.

Finding this type of missing heritability is conceptually easy, because it involves closer scrutiny of the genes already in hand. "Just exploring, in a very dense way, genetic variation at the loci that have been discovered is probably going to [explain] another increment of missing heritability," Hirschhorn says.

Researchers will need to sequence candidate genes and their surrounding regions in thousands of people if they are to unearth more associations with the disease.

Helen Hobbs and Jonathan Cohen of the University of Texas Southwestern Medical Center in Dallas did this in an attempt to capture all the variation in *ANGPTL4*, a gene their studies had linked to cholesterol and triglyceride concentrations. They sequenced the gene in around 3,500 individuals from the Dallas Heart Study and found that some previously unknown variants had dramatic effects on the concentration of these lipids in the blood<sup>7</sup>. Mark McCarthy of Britain's Oxford Centre for Diabetes, Endocrinology and Metabolism says that such studies could reveal much of the missing heritability, but not a lot of people have had the enthusiasm to do them. This could change as the cost of sequencing falls.

### Out of sight

Other variants, for which GWAS haven't even begun to provide clues, will prove even harder to find. In the past, conventional genetic studies for inherited diseases such as cystic fibrosis identified rare, mutated genes that have a high penetrance, meaning that the gene has an effect in almost everyone who carries it. But it quickly became apparent that high-penetrance variants would not underlie most common diseases because evolution largely keeps them in check.

What powered the push into genome-wide association was a hypothesis that common diseases would be caused by common, low-penetrance variants when enough of them showed up in the same unlucky person. Now that hypothesis is being questioned. "A lot of people are recognizing that screening for common variation has delivered less than we had hoped," says David Goldstein, professor of genetics at Duke University in Durham, North Carolina.

But between those variants that stick out like a sore thumb, and those common enough to be dredged up by the wide net of GWAS, there is a potential middle ground of variants that are moderately penetrant but are rare enough that they are missed by the net. There's also the possibility that there are many more-frequent variants that have such a low pen-

etrance that GWAS can't statistically link them to a disease.

These very-low-penetrance variants pose some problems, says Leonid Kruglyak professor of ecology and evolutionary biology at Princeton University in New Jersey. "You're talking about thousands of variants that you would have to invoke to get near 80% or 90% heritability." Taken to the extreme, practically every gene in the genome could have a variant that affects height, for example. "You don't like to think about models like that," Kruglyak says.

If rare, moderately penetrant or common, weakly penetrant variants are the culprits, then bumping up the number of people in existing association studies could help find previously missed genetic associations. Peter Visscher of the Queensland Institute of Medical Research in Brisbane, Australia, says that a meta-analysis of height studies covering roughly 100,000 people is in the works. Lowering the stringency with which an association is made could drag up more, but confidence in the hits would drop.

At some point it might make sense to stop using SNPs, and start sequencing whole genomes. Collins suggests that the NHGRI's 1,000 genomes project, which aims to sequence the genomes of at least 1,000 people from all over the world, could go a long way towards finding hidden heritability, and many more genomes may become possible as the price of sequencing falls.

Not everyone supports an all-out sequencing onslaught. Goldstein warns against





## Where's the reward?

There is more riding on the case of missing heritability than academic satisfaction. By finding variants related to common disease, genome-wide association studies promised to deliver meaningful medical information and justify the US\$3 billion spent on the human genome and the multimillion-dollar effort to map human variation. "The reason for spending so much money was that the bulk of the heritability would be discovered," says Joseph Nadeau, a geneticist at Case Western Reserve University in Cleveland, Ohio.

The ability to predict someone's height from their genes would be a pretty trivial carnival trick, but it represents a mastery over the language of life that could potentially spill into most areas of medicine. Aside from some surprises, though, such as mutations in immune-system genes being tied to an eye disorder called age-related macular degeneration, many of

the variants found have only modest effects on human characteristics. For now, genetics rarely provides a clearer predictive answer than a good family history. And the path to therapy is not straightforward, says David Goldstein of Duke University in Durham, North Carolina. "This talk about personalized risk profiles, using genetics, for most common diseases, and this talk about a whole flood of new drug targets. I think that that's now pretty clearly wishful thinking."

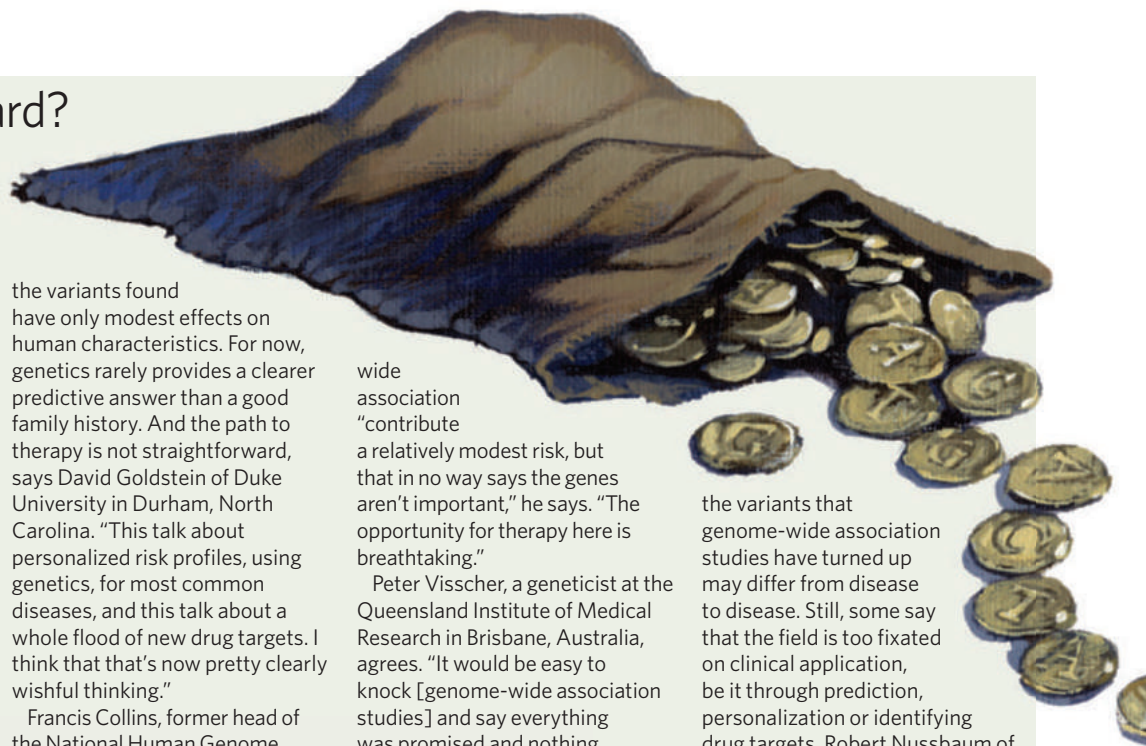
Francis Collins, former head of the National Human Genome Research Institute in Bethesda, Maryland, agrees that the picture for disease prediction remains bleak, but is still optimistic about therapeutic intervention. Most genetic variants found by genome-

wide association "contribute a relatively modest risk, but that in no way says the genes aren't important," he says. "The opportunity for therapy here is breathtaking."

Peter Visscher, a geneticist at the Queensland Institute of Medical Research in Brisbane, Australia, agrees. "It would be easy to knock [genome-wide association studies] and say everything was promised and nothing was delivered. But in terms of identifying genes and pathways for disease, it's been very successful. I would feel it's moved the field forwards tremendously."

Ultimately, the clinical value of

the variants that genome-wide association studies have turned up may differ from disease to disease. Still, some say that the field is too fixated on clinical application, be it through prediction, personalization or identifying drug targets. Robert Nussbaum of the University of California, San Francisco, puts it bluntly: "Human genetics research always assumes too quickly that it has to be translational. They're doing basic research." **B.M.**



continuing to "turn the crank" without devising a more rational approach, such as sequencing the genomes of people who exhibit extreme manifestations of diseases. "I'm not really sold on doing the sequencing version of what we did with [GWAS]," he says. "It's a big enough, costly enough job, that I think we want to think a little bit harder about exactly who gets re-sequenced."

### In the architecture

Some researchers are now homing in on copy-number variations (CNVs), stretches of DNA tens or hundreds of base pairs long that are deleted or duplicated between individuals. Variations in these features could begin to explain missing heritability in disorders such as schizophrenia and autism, for which GWAS have turned up almost nothing. Two recent studies looked at hundreds of CNVs in normal people and in those with schizophrenia, and found strong associations between the disease and several CNVs<sup>8,9</sup>. They commonly arise *de novo* — in an individual without any family history of the mutation.

These structural variants might account for a lot of the genetic variability from person to person and could account for some of those rare 'out-of-sight' mutations with moderate penetrance that GWAS can't pick up. Many

CNVs go undetected because they don't alter SNP sequences. Duplicated regions can also be difficult to sequence.

A standard technology for uncovering CNVs is array comparative genomic hybridization, in which scientists examine how genetic material from different individuals hybridizes to a microarray. If certain spots on an array pick up more or less DNA, it could indicate that there's a CNV. This and several other techniques are being tested by a consortium called the Copy Number Variation Project, run out of the Wellcome Trust Sanger Institute in Cambridge, UK. The consortium is dedicated to characterizing as many CNVs as possible so that associations can be made between them and diseases. McCarthy says that the role hidden CNVs have in heritability "should play out in the next six months to a year". But Goldstein argues that current technologies will miss many of the smaller CNVs, from 50 base pairs down to repeats of just two bases. "All we'll have verification of is the big whopping CNVs that are identifiable, and they clearly do not account for much of the missing heritability."

### In underground networks

Most genes work together with close partners, and it is possible that the effects of one on

heritability cannot be found without knowing the effects of the others. This is an example of epistasis, in which one gene masks the effect of another, or where several genes work together. Two genes may each add a centimetre to height on their own, for example, but together they could add five. GWAS don't cope with epistasis very well, and efforts to find these interactions usually require good up-front guesses about the interacting partners.

Joseph Nadeau, a geneticist at Case Western Reserve University in Cleveland, Ohio, says that 'modifier' genes act even in some straightforward single-gene diseases. "That's a simple kind of epistasis," he says. Cystic fibrosis, for example, is usually caused by mutations in one gene, *CFTR*, yet can vary greatly in symptoms and severity. The suspicion has been that modifier genes are one cause of this variability.

But despite the years of study, researchers still struggle to pin down these genes. "People haven't modelled truly the effect of epistasis," says population geneticist Sarah Tishkoff at the University of Pennsylvania in Philadelphia.

It's no surprise that genetics is more complicated than one gene, one phenotype, or even several genes, one phenotype, but it's humbling to realize how much more complex

things are starting to look. In a now classic study<sup>10</sup>, Kruglyak and his colleagues found that expression of most yeast genes is controlled by several variants, often more than five. To fill in all the heritability blanks, researchers may need better and more varied models of the entire network of genes and regulatory sequences, and of how they act together to produce a phenotype. At some point this process starts to look more like systems biology, and researchers are already applying systems methods to humans and other organisms (see page 26). “What we’re learning from these studies is that we need to think about the more complex of the complex models rather than the more simple of the complex models,” Kruglyak says.

### The great beyond

What if heritability estimates were wrong in the first place? Heritability of height was initially measured by taking the mean height of parents and comparing that value to the adult height of their offspring. As the average heights of parents increase, researchers found, so too does the average height of their children, hence the calculated 80–90% heritability.

Environment, especially factors such as nutrients or toxins present during important growth phases, can affect the mean height of a population considerably — but researchers have controlled for environment in estimates of heritability by, for example, comparing genetically identical twins raised together with those raised apart. Most researchers are confident that the heritability estimates are sound. “I don’t think anyone’s going to say that the heritability of height is 10% and let environment get you closer to the answer,” Kruglyak says. “I don’t think you can explain it away.”

But there are lingering doubts about how precisely environment has been accounted for in heritability studies. Adverse experiences *in utero* could lead to lifelong health disparities, according to David Barker from the University of Southampton, UK, and yet a shared womb is an aspect of the environment that would not be factored into such studies. “Heritability estimates are basi-

cally what clusters in families, and environment clusters in families,” says Manolio.

Epigenetics, changes in gene expression that are inherited but not caused by changes in genetic sequence, confuses things further. Feeding a mouse a certain diet, for example, can alter the coat colour not only in its children, but also in its children’s children<sup>11</sup>. Here, the expression of a coat-colour gene is controlled by a type of DNA modification called methylation, but it’s not completely clear how that methylation pattern is ‘remembered’ by the next generation. The idea that grandma’s environment could affect future generations is controversial — and such effects would have been included in the heritability normally attributed to genes.

“This complicates everything,” says Nadeau. “How do we sort out what great-grandfather and great-grandmother were exposed to when they were young and having children?” Model organisms might help. Nadeau has investigated testicular germ-cell tumours in mice that are analogous to a highly heritable cancer in humans. His group found that the effects of one weak, cancer-promoting gene, *Dnd1<sup>Ter</sup>*, are greatly enhanced by several other gene variants, and the boosted effects are passed on even if the genes that cause them are not<sup>12</sup>.

“It’s presumably transmitting its presence in some epigenetic way,” says Nadeau. The mechanisms by which epigenetic inheritance might work are still disputed, though; marks such as methylation that direct gene expression during someone’s life seem to be wiped clean in a new embryo.

One possible explanation for Nadeau’s observation, he says, is that RNA is being inherited alongside DNA through sperm or eggs.

Collins is not convinced that epigenetics will play a big part in missing heritability in humans. “It just doesn’t look likely outside of one or two examples to suggest that this is the case.” Nadeau disagrees. “It’s hard to imagine that every other organism works one way and humans are the exception,” he says.

### Lost in diagnosis

There is a nagging worry as researchers hunt for heritability: that common diseases might not, in fact, be common. Medicine tries hard to lump together a complex collection of symptoms and call it a disease. But if thousands of rare genetic variants contribute to a single disease, and the genetic underpinnings can vary radically for different people, how common is it? Are these, in fact, different diseases?

GWAS could actually be proving so difficult because researchers are seeking shared susceptibility genes in a group of people who may share few, if any. And yet without a more refined understanding of genetics, it could be impossible to categorize them any better. “It may be rare variants, common disease. And that’s kind of scary to people because it’s much, much harder to find those,” says Tishkoff.

There could be scarier and more intractable reasons for unaccounted-for heritability that are not even being discussed. “It’s a possibility that there’s something we just don’t fundamentally understand,” Kruglyak says. “That it’s so different from what we’re thinking about that we’re not thinking about it yet.”

Still the mystery continues to draw its sleuths, for Kruglyak as for many other basic-research scientists. “You have this clear, tangible phenomenon in which children resemble their parents,” he says. “Despite what students get told in elementary-school science, we just don’t know how that works.”

**Brendan Maher is a Features editor for Nature.**

1. Visscher, P. M. *Nature Genet.* **40**, 489–490 (2008).
2. Weedon, M. N. *et al. Nature Genet.* **40**, 575–583 (2008).
3. Lettre, G. *et al. Nature Genet.* **40**, 584–591 (2008).
4. Gudbjartsson, D. F. *et al. Nature Genet.* **40**, 609–615 (2008).
5. Sullivan, P. F. *PLoS Med.* **2**, e212 (2005).
6. Freitag, C. M. *Mol. Psychiatr.* **12**, 2–22 (2007).
7. Romeo, S. *et al. Nature Genet.* **39**, 513–516 (2007).
8. Stefansson, H. *et al. Nature* **455**, 232–237 (2008).
9. The International Schizophrenia Consortium *Nature* **455**, 237–241 (2008).
10. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. *Science* **296**, 752–755 (2002).
11. Waterland, R. A. & Jirtle, R. L. *Mol. Cell. Biol.* **23**, 5293–5300 (2003).
12. Lam, M. Y., Heaney, J. D., Youngren, K. K., Kawasoe, J. H. & Nadeau, J. H. *Hum. Mol. Genet.* **16**, 2233–2240 (2007).

See Editorial, page 1, and News Feature, page 26.

