

HIT OR MISS?

Genome-wide association studies have identified hundreds of genetic clues to disease.

Kelly Rae Chi looks at three to see just how on-target the approach seems to be.

Five years ago human geneticists rallied around an emerging concept. Technology had granted the ability to compare the genomes of individuals by looking at tens of thousands of known single-letter differences scattered across them. These differences, called single nucleotide polymorphisms or SNPs, served as reference points or signposts of common variation between individuals. The idea was that common variants in the genome might contribute to the genetics of common diseases.

Genome-wide association studies (GWAS) could scan SNPs in thousands of people, with and without a disease. When a DNA variant can be associated with the risk of developing a disease, it signals that something in that area of the genome might be partly responsible. With such 'hits' would presumably flow a better mechanistic understanding of disease, genetic-testing abilities and even treatment.

"Many researchers really grabbed on to the common variant hypothesis, and in some cases it worked," says Jonathan Haines, director of the Vanderbilt University Medical Center's Center for Human Genetics Research in Nashville, Tennessee. But, he adds, "it hasn't panned out to be as pervasive an explanation as we thought".

Here are the stories of three hits. One provides a near perfect example of the positive outcome that this sort of unbiased approach can have. One reveals that without biological context the findings can be hard to interpret. And the third demonstrates that GWAS in their current form can't cope well with some common traits.

A direct hit in haemoglobin

In 2007, researchers reported on genome-wide scans of healthy adults looking for SNPs associated with very high or

very low levels of fetal haemoglobin. Among several hits were variants of a gene on chromosome 2 called *BCL11A* (ref. 1). This finding, quickly replicated in multiple populations, generated a lot of excitement.

Fetal haemoglobin is a remnant of embryonic development. For most people, the fetal version of this crucial oxygen-carrying protein drops off after birth as the adult version kicks in. Some people retain relatively high expression, which seems to have no effect in healthy adults. But for patients with blood disorders such as sickle-cell disease and β -thalassaemia, those expressing high fetal-haemoglobin levels can be protected from some of the nastier ravages of the disease, such as leg ulcers, severe pain and even death.

GWAS findings often just provide the signpost, a rough coordinate for a causal gene. The SNP signal is often outside gene sequences. The variants in *BCL11A* were a direct hit in a gene, however, and a surprising gene at that. The protein it codes for, which controls the expression of other genes, had been associated with cancer progression, but never with haemoglobin production. A mouse model had even been made in which the gene had been knocked out, but until the GWAS no one had looked at its regulation of blood. "Nobody would have ever dreamed that a gene like this would have any regulatory role in fetal haemoglobin," says Martin Steinberg, a haematologist at the Boston University School of Medicine in Massachusetts. Last year, he and his colleagues replicated the *BCL11A* finding in three different populations with haemoglobin disorders².

Of course, there was functional work to be done. Stuart Orkin's lab at Harvard Medical School in Boston reduced expression of *BCL11A* in cultured blood progenitor cells from humans. Fetal haemoglobin expression went up,

ILLUSTRATIONS BY J. TAYLOR

suggesting that the gene normally acts as a repressor³. In a follow-up study⁴, the group showed that the gene controls the silencing of fetal haemoglobin during development in mice.

But the exact switch mechanism has not been solved, says Orkin. His group has gone on to look for proteins and other genes that the product of *BCL11A* binds to or influences. Results from these studies will inform research that looks for molecules that would interfere with the gene's expression or function and thus serve as potential therapies to activate synthesis of fetal haemoglobin in people with blood disorders.

Steinberg, for his part, hopes to use these and other GWAS findings to refine a computational tool that predicts disease severity or death in people with sickle-cell disease. The hope is to intervene earlier and with more specific treatments.

The verdict: even those who have been generally pessimistic about the outcomes of GWAS consider the *BCL11A* find a win for science. "It's a tour de force illustration of the value of GWAS," says David Goldstein, a geneticist at the Duke Institute for Genome Sciences & Policy in Durham, North Carolina. "You learn something new, you understand the mechanism, and it's biologically and clinically important." As winners go, however, Goldstein says it is on a short list.

In some ways, says Steinberg, the GWAS provided a lucky hit. The researchers built on years of evidence that fetal haemoglobin has a powerful effect on the severity of sickle-cell disease and β -thalassaemia. It was the clear physiological signal — quantity of fetal haemoglobin — that helped researchers to design the GWAS. Other genes and pathways will be found to affect the severity of the disorders, but probably none with the same force as fetal haemoglobin. "We're not going to find another fetal haemoglobin," Steinberg says.

Scoping schizophrenia

Schizophrenia genetics has been a mire of false starts. Scores of candidate gene association studies had identified promising targets, but few held up to further scrutiny. So the excitement around approaching the disease in an unbiased genome-wide study was high. But the first four schizophrenia GWAS reported no statistically significant associations. Then, in research published last year, researchers performed scans in roughly 500 people with the disorder and 3,000 healthy controls. When 12 of the hits that turned up were examined in 16,000 more individuals, a signal started to emerge. Three variants were significant, but only one of them was in a gene, *ZNF804A*, which encodes a protein with unknown function⁵.

Having a potential candidate gave researchers something concrete to work with. One group took

115 healthy people, a little less than half of whom had two copies of the high-risk *ZNF804A* variant, and compared their brain activity using functional magnetic resonance imaging (fMRI), a method that reveals local blood oxygenation and presumably electrical activity in the brain in real time. Those with the variant, they found, had abnormal connectivity between certain brain areas, impairing "the degree in which they talk to one another", says Andreas Meyer-Lindenberg, the director of the Central Institute of Mental Health in Mannheim, Germany, who led the study⁶. Healthy adults with the variant were showing schizophrenia-like brain activity even though they showed no outward signs of disease.

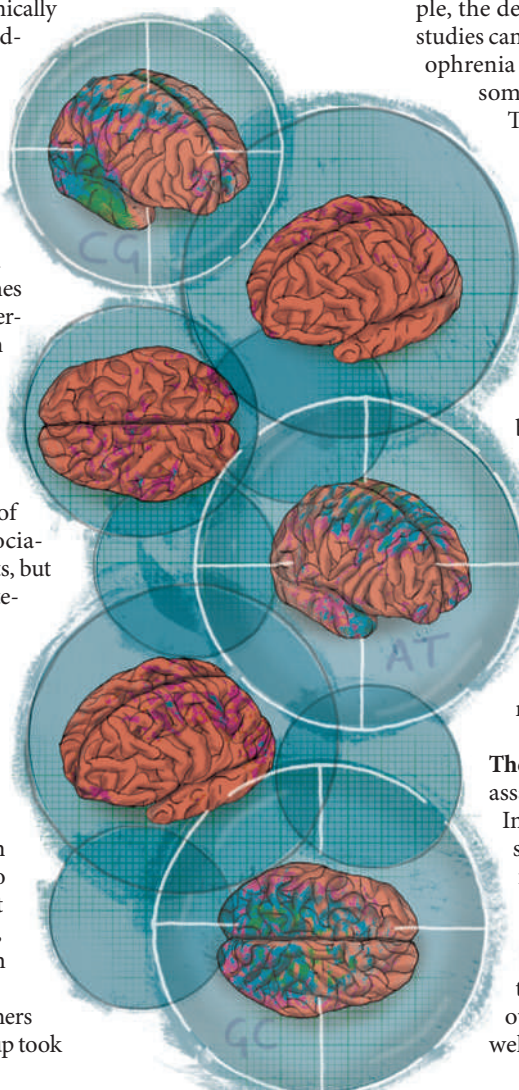
Combining genetic and brain-imaging data to study psychiatric disease is not new. Since 2001, researchers have used the strategy to link imaging data to candidate gene findings in schizophrenia, depression and autism. Meyer-Lindenberg's study is the first to use a genetic loci identified through GWAS for follow-up with fMRI. "We've now applied [the technique] to a variant that has definitive support as being a schizophrenia risk gene. That wasn't available before," says Meyer-Lindenberg.

Part of the problem when seeking schizophrenia-related genes is that, unlike fetal haemoglobin levels for example, the definition of the trait both within and between studies can differ. Also the spectrum and severity of schizophrenia symptoms varies between individuals and are sometimes subjective from a clinical perspective.

That's why fMRI is attractive. The researchers hope to get closer to quantitative measures of psychiatric disorders. "It makes sense to have a biological level of analysis on which these genetic associations can be studied," says Daniel Weinberger, the director of the genes, cognition and psychosis programme at the National Institute of Mental Health in Bethesda, Maryland, who pioneered the method in the 1990s.

The *ZNF804A* association from GWAS has been replicated in some studies but not others. And there are few clues to the mechanism by which this gene might contribute to brain connectivity. It was the group of Michael O'Donovan, a professor of psychological medicine at Cardiff University, UK, that made the initial discovery using GWAS. The team is now carrying out a series of experiments to determine which DNA sequences and other proteins it binds, and how variants might alter gene expression.

The verdict: some have doubts about the combined assault of GWAS and fMRI on psychiatric illness. Imaging data itself isn't the best quantitative trait, says Goldstein, because one three-dimensional fMRI image can contain more than 50,000 picture elements of data, a single trait can be defined in multiple ways. "There hasn't been a sufficient consistency in how those phenotypes are defined," he says. Weinberger and others contend that the imaging paradigms are well established before they are used in imaging



genetics research. "I think it's very important that the phenotypes are well validated — that they are themselves heritable, and that they're related in some way to the underlying neural circuitry," says Weinberger.

Reviews from others studying schizophrenia are somewhat lukewarm. Kári Stefánsson, chief executive of the Icelandic biopharmaceutical company deCODE Genetics in Reykjavik, says he's not completely convinced. "In schizophrenia, the imaging differences are subtle," he says. Nevertheless, he plans to study differences in brain morphology using imaging and GWAS, in people with and without the disorder.

Given the shortage of standout GWAS hits, should researchers continue to use the candidate-gene approach to form the basis for hypothesis-driven imaging genetics work? "It is still a point of debate," Meyer-Lindenberg says.

Sight set on height

Height has produced clearer hits than schizophrenia, but with a less than satisfying punch. In 2007, by analysing the genomes of nearly 5,000 people, researchers were able to see that a variant in a gene called *HMGA2* explains some of height's variability — about 0.3% (ref. 7). Since then, additional GWAS have revealed more than 40 loci involved in height. Added together, these variants account for 5% of the trait's variation. Even a clear quantitative trait doesn't necessarily provide simple answers.

Genes are thought to contribute to roughly 60–80% of the variation in stature, leaving much of the heritability of height unaccounted for by GWAS findings. This 'missing heritability' has been a thorn in the side of the common-disease-common-variant hypothesis (see page 747). "In the field of height," says Haines, "obviously that hypothesis is not completely correct."

But the news isn't all bad. "Optimistic people like me say we didn't know anything about the genetics of height before 2007," says Guillaume Lettre, a geneticist at the Montreal Heart Institute in Quebec, Canada. "Now we have more than 40 loci."

Researchers may have loci, but they have little idea how these contribute to height. As with other traits, many of the associated SNPs fall within the vast regions between genes or within genes whose function is unknown. And with little funding for understanding height variation and scant biological footholds, the field sees very little follow-up of its GWAS leads.

Lettre is collaborating with others with the hope of tying mystery SNPs to genes through animal models. "Basically what we are doing is taking the genes near these markers and looking at the expression of these genes in tissues that are relevant to height," he says. "There are not so many: bone, cartilage and pituitary gland."

He and others are also trying to coax existing height data into revealing stronger associations by grouping hits based on a single molecular pathway. Hong-Wen Deng, at the

University of Missouri in Kansas City, is planning to analyse pathways involved in either bone health or stature. "Many genes which may have small effects for height may not be detected if you analyse them individually," he says. But jointly their effects may be detected. Others are looking at height at various points with the hope that differences in growth curves will reveal larger genetic associations. Researchers have already examined height and growth rates in about 3,500 people from Northern Finland. Of 48 height-associated variants that they tested, 12 were linked with the rate of growth during infancy or puberty⁸.

The verdict: some of the loci implicate molecular pathways already known to be involved in growth and development. A 1995 study had shown that a gene related to *HMGA2* could influence height: mice lacking the gene were shorter, whereas mice with a truncated version developed gigantism⁹.

The *HMGA2* association has been further confirmed by most, but not all, GWAS.

Predicting how hits outside genes will fare is more difficult, and depends to some extent on how close the hit is to the nearest gene. "If you look at the height loci, they are much more likely to be near a gene that causes abnormal skeletal growth, than a similarly sized random set of loci," says Joel Hirschhorn, a geneticist at the Broad Institute in Cambridge, Massachusetts.

But the nearest gene is a poor marker for what is likely to be causal says Goldstein. "Depending on the genetic model for what is causing the association, it could be nearby or not nearby," he says. In some instances changes in DNA act on genes a million bases away. "It really is remarkable that there are hundreds of reported associations, and the number that you can actually track down to an actual cause of the association is probably countable on one hand."

Researchers point to height as a 'model trait' because it is simple to measure and relatively constant compared with phenotypes such as blood pressure or glucose level. Then again, in GWAS of height, tens of thousands of people have been necessary to see the slightest associations. As a model trait, that could be problematic. "In some ways, it is showing us the future for other traits," says Karen Mohlke, a geneticist at the University of North Carolina at Chapel Hill who was involved in some of the initial height GWAS work. "What it means for many other complex traits is that there will be as many loci found, or more."

Kelly Rae Chi is a freelance writer in Cary, North Carolina.

1. Menzel, S. *et al. Nature Genet.* **39**, 1197–1199 (2007).
2. Sedgewick, A. E. *et al. Blood Cells Mol. Dis.* **41**, 255–258 (2008).
3. Sankaran, V. G. *et al. Science* **322**, 1839–1842 (2008).
4. Sankaran, V. G. *et al. Nature* **460**, 1093–1097 (2009).
5. O'Donovan, M. C. *et al. Nature Genet.* **40**, 1053–1055 (2008).
6. Esslinger, C. *et al. Science* **324**, 605 (2009).
7. Weedon, M. N. *et al. Nature Genet.* **39**, 1245–1250 (2007).
8. Sovio, U. *et al. PLoS Genet.* **5**, e1000409 (2009).
9. Zhou, X., Benson, K. F., Ashar, H. R. & Chada, K. *Nature* **376**, 771–774 (1995).

See Editorial page 697, and the human genetics online special at <http://go.nature.com/VqPUE2>.