Genetic information: potential uses for economic and social research

New collaborations between genetic scientists and social scientists are exploring links between people's physical characteristics and their behaviour and economic outcomes. *Stephanie von Hinke Kessler Scholder* and *Neil Davies* describe the potential uses of genetic information for social science.

Recent developments in the science of genetics have dramatically reduced the cost of obtaining genetic data. The first draft of the human genome project was published in 2003 and cost $3 billion, but since then the cost of sequencing DNA has fallen dramatically. This has led many cohort studies and other surveys, including those often used by social scientists, to collect bio-samples and extract genetic data. But what can genetic information offer social scientists?

Our genome influences many human features – from physical characteristics, such as height (Visscher, 2008), to traits that are more difficult to measure, such as risk preferences (Cesarini et al, 2010). One way to study the influence of our genome on these characteristics is to calculate their *heritability*. This is defined as the proportion of the total variation in the characteristic that is due to genetic differences.

Heritability studies use statistical techniques to estimate this proportion. If certain characteristics of interest to social scientists are found to be heritable, we may then wish to examine and understand the specific underlying differences in DNA and the genetic mechanism that can explain this heterogeneity across individuals.

But how do scientists find genetic variants or specific locations on the DNA sequence that are related to characteristics such as height, weight or risk preferences? Until recently, genetics researchers used a 'candidate gene approach' to examine the association between a specific genetic variant and a particular characteristic.

Based on biological knowledge, researchers examined associations between the variant (the 'candidate genetic variant') and the characteristic. But because the genome is so large, this was a very slow process. Improvements in technology and computing, however, have led to the development of new methods and techniques.

One of the most successful methods has been genome wide association studies, which can relate up to 2.5 million genetic variants to the characteristic of interest. For example, this has allowed researchers to pinpoint specific parts on the DNA sequence that are associated with individual fat mass or 'body mass index' (BMI, calculated as an individual's weight in kilograms divided by their height in meters squared).

Genome wide association studies can relate up to 2.5 million genetic variants to a single human characteristic

The FTO gene was discovered using this approach. This gene shows a clear and robust association with BMI, obesity and fat mass. People who carry the rare version of *FTO* are on average slightly heavier, have a larger fat mass and are more likely to be overweight or obese, compared with people who carry the common version.

FTO is often called the 'obesity gene', though this is somewhat misleading. Not all carriers of the rare *FTO* variant are obese, and obesity is influenced by many other factors, some genetic and some environmental. But the rare variant of *FTO* does increase adult body weight by 2-3 kilograms on average (Frayling et al, 2007).

Genome wide association studies can be applied to characteristics that are of interest to social scientists. For example, the heritability of educational attainment is typically estimated to be around 40% (Beauchamp et al, 2010). Genome wide association studies can attempt to find the exact locations on the DNA sequence that contribute to this heritability.

Indeed, researchers are already attempting to relate educational attainment to specific genetic variants in genome wide data (Beauchamp et al, 2010). Understanding the underlying differences that explain the overall heritability could improve our understanding of the variation that is observed across individuals.

Genetic information can also be used to test broader causal hypotheses. We know that the allocation of genes from parents to offspring is random, as shown by Gregor Mendel (1822-84) in his study of the inheritance of traits in pea plants.

Studies have also shown that individuals' genes are unlikely to be related to their background characteristics, such as socio-economic position, life expectancy or income. In fact, the term 'Mendelian randomization'

The variation caused by genes can be used to study the causal effects of fat mass on academic performance

refers to studies that exploit this random assignment of individuals' genomes (Davey Smith and Ebrahim, 2003).

Mendelian randomization uses the random assignment of genes to study the effects of a particular characteristic on an outcome of interest. For example, the *FTO* gene can be used to explain variation in fat mass or BMI. This variation can then be related to individuals' academic outcomes to investigate the effect of fat mass on academic outcomes. This is the methodology that economists and genetic epidemiologists at the University of Bristol are using, as described in more detail in the next article.

It may seem counterintuitive to use *FTO* to predict differences in fat mass, rather than measuring fat mass itself. But there are several crucial advantages to this approach. First, unlike fat mass, *FTO* is unlikely to be related to other behavioural, social and physiological factors that may confound the association between fat mass and academic achievement. This means that if *FTO* is used to predict fat mass, it is unlikely to be confounded with such factors.

Second, any association between fat mass and academic achievement may be due to 'reverse causation'. For example, children who perform poorly in school may decide to over- or under-eat in response. This would imply that differences in school outcomes cause differences in fat mass rather than vice versa. In contrast, *FTO* is not affected by the child's performance in school and cannot be biased by reverse causation.

The method of 'instrumental variables', a statistical approach commonly used in econometrics, can be used to study causality. In this case, it would use the variation caused by genes to study the effects of fat mass on academic performance. The use of genetic information in this way is new, and makes this a novel and exciting area for social scientists.

It allows researchers to explore causal research questions that are difficult or perhaps even impossible to answer in observational studies. In addition, collaboration between social scientists and genetic epidemiologists will benefit both: social scientists will learn about the science behind genetics, and genetic epidemiologists will increase their understanding of the latest statistical and methodological techniques.

The advantage of Mendelian randomization lies in avoiding the confounding factors that often complicate the interpretation of observational studies. But like all research methods, it has limitations that must be taken into account. One potential problem is that the frequencies with which genetic variants occur may differ across different populations, such as ethnic groups. For example, *FTO* is more common among Europeans than among Chinese and Japanese people.

Another potential problem is that when genetic variants are passed on from one generation to the next, they may be linked (or co-inherited) with other variants. Similarly, variants may influence more than one characteristic (Davey Smith, 2010).

These potential problems may therefore re-introduce confounding factors. But rapid developments in our understanding of the genome and in the functions of specific variants, as well as technological and statistical advances, may alleviate some of these problems and make Mendelian randomization an increasingly powerful method for social science as well as medical and epidemiological research.

Stephanie von Hinke Kessler Scholder is a post-doctoral research fellow at Imperial College Business School and a CMPO research associate; Neil Davies is a post-graduate research student in the University of Bristol's School of Social and Community Medicine

Further reading

Jonathan Beauchamp, David Cesarini, Matthijs van der Loos, Philipp Koellinger, Patrick Groenen, James Fowler, Niels Rosenquist, Roy Thurik and Nicholas Christakis (2010) 'A Genome-Wide Association Study of Educational Attainment' (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1655023)

David Cesarini, Christopher Dawes, Magnus Johannesson, Paul Lichtenstein and Bjorn Wallace (2009) 'Genetic Variation in Preferences for Giving and Risk Taking', *Quarterly Journal of Economics* 124: 809-42

George Davey Smith and Shah Ebrahim (2003) 'Mendelian Randomization: Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?', *International Journal of Epidemiology* 32: 1-22

George Davey Smith (2010) 'Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene x Environment Interactions', *Perspectives on Psychological Science* 5: 527-45

Tim Frayling et al (2007) 'A Common Variant in the *FTO* Gene is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity', *Science* 316: 889-94

Peter Visscher (2008) 'Sizing Up Human Height Variation', *Nature Genetics* 40: 489-90