

# Human Complex Trait Genetics in the 21st Century

Peter M. Visscher<sup>1</sup>

ORCID ID: 0000-0002-2143-8760 (P.M.V.)

*Making predictions about the next 100 years of genetics research is bound to be imprecise and biased, but I am going to stick my neck out anyway and make some shorter-term predictions, some no doubt controversial, about research and likely discoveries in human complex trait genetics.*

**KEYWORDS** complex trait; genomics; population genetics; quantitative genetics

I moved into the field of human complex trait genetics less than 20 years ago, from a background in quantitative genetics and animal breeding. Even in this period of time, major changes have occurred that were hard to predict back in the 1990s. Driven by enormous advances in DNA sequencing technologies, one can now sequence and analyze an entire human genome for a few thousand dollars. Some may argue that the cost of a sequenced genome is much lower than that, but that usually ignores the expense of storage, analysis, and interpretation. Sequencing technology has facilitated easy and fast discovery of Mendelian disease mutations and coding variants with high penetrance (a high probability of disease given genotype), and has led to precise estimates of the per-generation mutation rate (1000 Genomes Project Consortium *et al.* 2010). In the same period, development of array genotyping technology has made it possible to genotype hundreds of thousands of DNA variants for less than \$100. Millions of samples have been genotyped using such arrays to study the genetic basis of complex traits such as common disease and quantitative traits, which has led to the discovery of many thousands of genes, gene variants, and biological pathways that are associated with one or more complex traits (Visscher *et al.* 2012). The traits vary widely, from psychiatric disorders to autoimmune disease, cancer, anthropometric traits such as height and weight, traits measured in blood such as platelet size and counts, and behavioral traits such as intelligence and years of schooling. In addition to trait-variant discovery, the

technologies have led to new discoveries in human evolution and population genetics.

These mostly unpredicted rapid developments were not just taking place in human complex trait genetics. In plant and animal breeding, a revolution has been taking place in the last 15 years. In 2001, a theoretical paper in this journal showed that with a sufficiently dense marker map, linkage disequilibrium could be exploited to predict breeding values and speed up genetic gain by radically changing the structure of breeding programs (Meuwissen *et al.* 2001). This paper was published well before the first commercial SNP chips were available, and within 10 years of publication, the method, called “genomic selection” (or “genomic prediction”), was implemented in dairy cattle breeding programs around the world; breeders of other livestock species and crops are following the same route. The update of this technology has led to a doubling of the rate of genetic gain in dairy cattle (Veerkamp 2015), an astounding increase and an incredibly rapid update of new technology.

My main thesis is that the relentless pace of technological innovation will cause a change in how science is conducted. Instead of the model-based hypothesis-testing science that dominated the last century, the next will be hypothesis-generating-discovery science that is driven by data. I believe that this change will not be confined to human complex trait genetics, but will apply to all areas of research in genetics. Genomics will become synonymous with biology, a trend already occurring.

## Genetic Data Will Not Be Limiting

A conservative prediction is that genetic data will not be a limiting factor in answering fundamental questions about the evolution and nature of complex trait variation in human

populations. The cost of generating a whole genome sequence is still going down, and it is not inconceivable that a majority of people on the planet will have their genome sequenced in 50 years' time. What will limit our ability to find answers to questions about genome–phenome relationships is the availability of high-quality, in-depth phenotypic and environmental information to link to the genetic data. But even the phenome might become tractable with better technologies, such as smart sensors and devices that track behavior, physiology, and the environment in real time. With gigantic sample sizes, it will be possible to explain most, if not all, additive genetic variation for a range of traits and to tackle old questions about the nature of mutational variance, the maintenance of genetic variation, the genetic control of variability, and the elusive quantification of variation due to nonadditive and genotype-by-environment ( $G \times E$ ) interactions.

I predict that the tens of millions of single nucleotide variants and the many copy number variants that currently segregate in the population will be whittled down to a much more manageable credible set of plausible causal variants. I am agnostic as to what the size of that set is going to be (ten thousand? a hundred thousand? one million?). The phenome will not only consist of continuous measurements on individuals such as physical activity, heart rate, and blood pressure, but will also include genome-wide nonsequenced-based “omics” data such as gene expression and epigenetic modifications. Sophisticated data-driven multivariate algorithms are likely to be developed that will enable prediction of the consequence (if any) on the phenome of a *de novo* mutation in the context of a person's genome. A credible set of causal variants is likely to provide new insight into pleiotropy, for example, by quantifying the contributions to genetic covariance by functional annotation and by quantifying the joint distribution of effect sizes on different traits, even when their genome-wide genetic correlation is zero.

If most additive genetic variation is accounted for by known variants, then additive by additive variance can be quantified, and similarly the interaction (or lack thereof) between identified environmental factors and additive genetic values. Differentiating between genotypic and additive variation will remain problematic for highly polygenic traits because there will be too many unique genotypes for their values to be estimated accurately, and theory predicts that for highly polygenic traits most genetic variation will be additive anyway (Maki-Tanila and Hill 2014).

### **In Osteo Population Genetics Studies?**

Population genetics studies, including those applied to human populations, were founded on sophisticated mathematical models of changes in gene frequencies geographically and over time. Until recently, genetic data were limiting and largely constrained to observed allele frequencies between and within populations. This has changed drastically in the last 10 years because of the availability of SNP arrays and

genome sequences, leading to the identification of several loci and variants that have been under natural selection. DNA extraction and sequencing technology have improved to the extent that partial genome sequences of Neanderthals have been generated, and SNP data have been acquired from recent ancestors living in Europe 3000 to 8000 years ago (Haak *et al.* 2015), drawing inference about natural selection in the past 8000 years (Mathieson *et al.* 2015). I predict that the technologies will develop further and that, in principle, it will be possible to take bone samples from a number of individuals who lived 100, 200, ... 10,000 years ago and infer recent natural selection as if it was in real time by tracking changes in allele frequencies of variants that are known (from modern day studies) to be associated with complex traits and fitness. It might even be possible to study  $G \times E$  interaction by performing gene mapping on ancestral samples, for example on femur lengths (which is a highly heritable complex trait). Dig up the bodies!

### **Modeling Human Complex Traits in Experimental Organisms Will Become Obsolete**

Model organisms such as fruit flies, mice, and worms have been at the forefront of major discoveries in genetics over the last century. Many if not most of these discoveries were about mechanisms, *e.g.*, mechanisms of natural selection, speciation, recombination, imprinting, response to selection, and gene function. Experimental organisms have been less successful in modeling human disease (in the sense of leading to successful prevention or treatment), even, for example, when engineered mutations in mice are identical to those discovered in human patients. My prediction for future research into human disease causes and drug discovery is that humans will become a “model organism” through exploiting new technologies such as tissue-specific cell lines and gene editing.

I would also argue that model organisms have been largely unsuccessful in modeling complex traits in general, whether for proposed applications in human health or for potential applications in plant and animal breeding. Progress in livestock genetics has come from studying complex traits in cattle, pigs, and poultry, not from studying crosses between inbred lines of mice. Similarly, progress in understanding disease in humans has largely come from studying those diseases in humans and not from building models of them in other species. Indeed, the rapid developments in human complex trait genetics over the last 10 years have outshone those in, *e.g.*, mice or flies. There are exceptions, of course, but they are not common.

### **Personalized Genetics and Genomics Will Become an Integral Part of Health Care and Clinical Practice**

One major application of studying complex traits in humans is in medicine. Indeed, most of the public funding to study complex traits in human populations has come from medical

research funding bodies such as the National Institutes of Health, the Wellcome Trust, and the Medical Research Council. Genetic technologies, including genome sequencing, have already led to changes in clinical practice, for example by personalizing drug advice for cancer depending on the tumor's genomes. I believe the very near future will see this extended to diagnosis of Mendelian disease and to providing more refined personalized treatment advice for cancer.

The bigger question for the future is how to extend this to common diseases and traits, which provide the largest personal, health, and economic burden on society. I predict major changes in how health care will be managed through a person's lifetime, using personalized genetic- and genomic-based information (including metabolomic, proteomic, and microbiome data) combined with phenome-tracking information from smart electronic devices. The bottleneck to make this happen will be in the collection and analysis of relevant data. It is telling that in 2015, both Google and Apple are seeing health and medicine as a major field of interest.

## Literature Cited

- 1000 Genomes Project Consortium; G. R. Abecasis, D. Altshuler, A. Auton, and L. D. Brooks *et al.*, 2010 A map of human genome variation from population-scale sequencing. *Nature* 467: 1061–1073.
- Haak, W., I. Lazaridis, N. Patterson, N. Rohland, and S. Mallick *et al.*, 2015 Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207–211.
- Maki-Tanila, A., and W. G. Hill, 2014 Influence of gene interaction on complex trait variation with multilocus models. *Genetics* 198: 355–367.
- Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, and N. Patterson *et al.*, 2015 Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* DOI: 10.1038/nature16152.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Veerkamp, R. F., 2015 Abstracts of the XVIth Meeting of the Eucarpia Section Biometrics in Plant Breeding, September, 2015, Wageningen, The Netherlands.
- Visscher, P. M., M. A. Brown, M. I. McCarthy, and J. Yang, 2012 Five years of GWAS discovery. *Am. J. Hum. Genet.* 90: 7–24.

*Communicating editor: M. Johnston*