Understanding the genetics of height—and why it matters in the fight against many human disorders

H oping to help researchers find links between diseases and mutations, the UK Biobank opened its vault last summer, allowing access to genetic data on 500,000 people.

Among the resulting research projects is a genetic model that can predict most people's height to within centimeters. The work, recently [published](#) as a preprint on [BioRxiv](#), incorporates about 20,000 single-nucleotide polymorphisms (SNPs), one-letter variations in the genome. While some of these SNPs were previously known to be associated with height, others were not. Often, the biological function of these SNPs (if any) is unknown.

It turns out, we haven't been very good at using genetics to predict height. A [2009 paper](#) published in the *European Journal of Human Genetics* used 54 SNPs to predict height could only account for about 4 percent of the variability in height, doing considerably worse than a method developed by Sir Francis Galton in the 1800s, which was able to predict about 40 percent of the variability in height.

While predicting height from genetics may have some practical applications, such as in childhood growth disorders and in forensic science, the importance of the results comes from the fact that the model captures nearly all the predictive heritability possible from the SNPs contained in the Biobank.

height 2 14 18 2

image not found or type unknown

Like all complex traits, height has both a strong genetic and a strong environmental factor. It's been estimated that about [60 to 80 percent](#) of the difference in height between two people is due to genetics. The rest is due to environmental factors, primarily nutrition. This genetic height predictor captures most of the height variance due to genetics, meaning that it may theoretically be the best possible genetic-based predictor.

According to Stephen Hsu, the lead author on the study and Vice President for Research and Graduate Studies at Michigan State University, "There is a perception that we'll never be able to predict height at this accuracy. I've been saying it's just statistical power. If we were willing to invest the money, we would be able to capture the known heritability [for complex diseases]."

In other words, given enough sequences, patients, and investment, similar models can be developed for other complex traits. Using similar methods, Hsu has also developed a model that to predict educational attainment and heel bone density. He predicts that accurate genetic prediction models for diseases such as Alzheimer's disease, type 1 diabetes, ovarian cancer, and schizophrenia can also be created, given a large enough dataset. The genetic contribution for these conditions ranges from about [40 to 70 percent](#), similar to that for height. In theory, such models could be used to generate a polygenic score of a person's risk for developing a given disease. Similar to single-gene genetic tests, such as sequencing the *BRCA1/2* gene to determine breast cancer, this information could be used to guide prevention or treatment options. Some, however, warn that it could also be used to select for disease-related as well as nondisease-related traits [in embryos](#). While it's important to consider these ethical implications, we should also keep in

mind that we already have this ability for a host of single-gene diseases.

Image not found or type unknown
The authors estimated that to create genetic prediction

models for complex diseases, they would need data from about 100,000 individuals with the disease and about 100,000 control individuals without the disease to have enough data to generate an accurate genetic prediction model. According to Hsu, this would require an investment on the order of tens of millions of dollars – a sizable sum and monumental effort to be sure, but perhaps well worth the cost.

Getting funding for such a project through traditional channels, such as government grants, would be extremely difficult for the average researcher. In addition, the resources required to store, analyze, and maintain the vast amount of data collected is likely outside of the capacity of many labs. Therefore, the onus of such large endeavors have fallen upon two sources: commercial companies, such as 23andme, and nonprofit organizations, such as DNA.land, an initiative of the New York Genome Center.

The UK Biobank was established by several national foundations in the UK, including the Welsh and Scottish governments. Over 500,000 individuals have been recruited across the UK. The amount of data being collected is massive. In addition to genetic data, each individual provided blood and saliva samples and agreed to have their health followed over time. About 20 percent of the recruits also provided data on their level of activity by wearing a 24-hour monitor for one week; information on diet, cognition, and work history are being collected, and 100,000 participants are being included in an imaging study, in which their major organs, bones, and carotid artery are being scanned.

If this study is a harbinger of things to come from the UK Biobank data, it seems that great things are to come from this resource.

Jennifer Cable has a PhD in biochemistry and is a freelance science and medical writer. Follow her at her website or on Twitter @jlcphd.