Pinpointing How Genes Interact

Your DNA—the biological instruction manual in all of your cells—contains a mind-boggling amount of information represented in roughly 20,000 genes that encode proteins, plus a similar number of genes with other functions. As the cost of analyzing an individual’s DNA has plummeted, it has become possible to search the entire human genome for genetic variants that are associated with traits such as height or susceptibility to certain diseases. Sometimes, one gene has a straightforward impact on the trait.

But in many cases, the effect of one gene variant depends on which variants of other genes are present, a phenomenon called “epistasis.” Studying such interactions involves huge datasets encompassing the DNA of hundreds of thousands of people. Mathematically, that requires time-intensive calculations with massive matrices and a good working knowledge of statistics.

Statisticians and computer scientists are developing a variety of methods to more efficiently analyze the interactions between variants across the entire genome. For instance, rather than testing all possible pairs of gene variants for epistasis, researchers might pick a single variant and examine the combined effects of all other genes. Not only does this speed up the computation time, it also improves researchers’ ability to distinguish true genetic effects from random chance. And just as important as using the right statistical tools is gathering a dataset that reflects the full diversity of humanity. Recent research drawing upon the genomes of people from many ancestral groups revealed examples of epistasis that had not previously been found when studying only people of self-identified European ancestry.