New maps link thousands of genetic variants to disease genes

April 8 2021, by Claudia Lopez Lloreda

Genome-wide association studies (GWAS) have identified thousands of genetic variants that are associated with a wide range of diseases. But these studies often fall short of establishing what a genetic variant does to bring about a particular trait or disease.

"The challenge has been that many of these variants are difficult to interpret," said Jesse Engreitz, a former postdoctoral fellow at the Broad Institute of MIT and Harvard and now an assistant professor at Stanford University. Genetic variants are even harder to interpret when they are in regions of the genome that don't code for proteins, such as in regions called enhancers. These noncoding elements modulate the activity of genes that are often located far away from the enhancer, and they do so differently in different cell types.

To address this problem, Engreitz, while at the Broad, along with Broad collaborators, developed a model that predicts the genes that enhancers regulate in specific cell types. Now, a collaborative effort, published today in *Nature* and led by Engreitz and Broad founding director Eric Lander (on leave), has used that model to create maps that connect enhancers to their target genes in 131 cell types and tissues. Using those maps, the team linked more than 5,000 signals from GWAS studies to nearly 2,250 genes across 72 traits and diseases including cancer and heart disease.

"We now have the ability to look comprehensively across many cell types and, for the first time, make reasonably accurate predictions of what these non-coding enhancers and variants do," said Engreitz.

The team also showed that their model, called the activity-by-contact

(ABC) model, accurately predicted which enhancers contain risk variants for inflammatory bowel disease (IBD). The study suggests a possible biological mechanism underlying IBD, and sets the stage for further study of regulatory elements in other diseases.

Enhancing the model

The researchers' ABC model predicts which enhancer regulates which gene based on two relatively simple factors: how often an enhancer and its target gene's promoter come into contact, and the enhancer's level of activity. In their study, the team applied their model to 131 samples representing different cell types, tissues, and cell lines, to create more than 100 maps that predict enhancers and their potential gene targets in each sample.

The scientists soon realized that these maps could help them identify genetic variants involved in certain traits or diseases. They saw that variants for IBD previously identified in GWASs often overlapped with predicted enhancers. Moreover, the IBD risk variants occurred in cell types that were relevant to IBD, such as immune cells of the gut.

The team next applied their model to 72 diseases and traits, and linked 5,036 GWAS signals to 2,249 genes. Of these genes, 577 appeared to have roles in multiple traits or diseases through variants in enhancers that act in different cell types. The findings suggest the ABC model can help researchers identify variants within enhancers that could be relevant to disease and pick which cell types to focus on when studying disease mechanisms.

"Our model could be really useful for us to narrow down options," said Joseph Nasser, formerly a staff scientist at Broad, now a graduate student at the California Institute of Technology, and one of five first co-authors of the study. "If a variant we're interested in overlaps an

enhancer that's predicted to regulate a gene, but does this in only a few cell types, that's a huge help in trying to figure out how that variant mediates disease risk."

Investigating IBD

To put their model to the test, the researchers looked at one specific enhancer, which the model predicted was connected to a gene called PPIF. This gene encodes for cyclophilin D, a protein involved in the activity of the cell's energy-producing mitochondria. The team found an IBD-associated variant within their predicted enhancer, suggesting that the enhancer could be important in IBD. By modulating this enhancer in a type of immune cell in the lab, the team found that the enhancer tuned the cells' mitochondrial function. The results effectively linked one of the team's predictions to an actual activity in the cell.

The study is a step forward in studying IBD, which affects about 3 million adults in the US. More broadly, it also helps researchers zero in on the parts of the vast human genome that might be important for a particular [disease](link) or trait. "This should really accelerate the process of understanding variant function because we can now home in on the exact cell type and gene combination that we should study," said Engreitz. "This study brought together people from around the Broad with different expertise. It was a really amazing collaboration with many people."

The other four co-first authors are Drew Bergman of the Broad, Charles Fulco, now at Bristol Myers Squibb, Philine Guckelberger, now at the Free University of Berlin, and Benjamin Doughty, now at Stanford.

The scientists generated more than 100 enhancer-gene maps using data collected from one person. They hope to continue refining their model and that more such maps will be generated from different people and

across hundreds of cell types, said Nasser. "This study is more about opening the book on this problem, instead of closing it."

More information: Joseph Nasser et al. Genome-wide enhancer maps link risk variants to disease genes, *Nature* (2021). DOI: 10.1038/s41586-021-03446-x

Provided by Broad Institute of MIT and Harvard