Assay of nearly 5,000 mutations reveals roots of genetic splicing errors

April 17 2017

A depiction of the double helical structure of DNA. Its four coding units (A, T, C, G) are color-coded in pink, orange, purple and yellow. Credit: NHGRI

It's not so hard anymore to find genetic variations in patients, said Brown University genomics expert William Fairbrother, but it remains difficult to understand whether and how those mutations undermine health.

In a new study in *Nature Genetics*, his research team used a new assay technology called "MaPSy" to sort through nearly 5,000 mutations and identify about 500 that led to errors in how cells processed genes. The system also showed precisely how and why the processing failed.

"Today, because we can, we're getting tens of thousands of variants from each individual that could be relevant," said Fairbrother, an associate professor of biology. "We can sequence everything. But we want to know which variants are causing diseases—that's the beginning of precision medicine. How you respond to a therapy is going to be determined by which variant is causing your disease and how."

To accelerate that knowledge, Fairbrother has dedicated his lab to developing a variety of tools and techniques, including software and biophysical systems such as MaPSy, to study gene splicing. Genes are sections of DNA sequence that provide cells with the instructions, or code, for making proteins the body needs for its functions. During this manufacturing process, useful protein coding sequences need to be cut out and reconnected—spliced—from the longer sequences, much as usable movie scenes are cut from longer reels of raw footage when making a film.

Genes are often viewed as the blueprint of proteins. Sometimes mutations in genes affect not the code of the proteins themselves, but instead the splicing sites and instructions that govern how the gene sequence should be read. That can be a big problem—while the former kind of problem might affect a component of a protein, the latter kind of error can affect whether the protein is made at all. It's therefore important to understand how an individual's genetic variation could alter

[gene splicing](#), Fairbrother said.

"Splicing errors can be very deleterious because instead of just changing one amino acid [the building block of a protein], it can take out a stretch of 40 or 50 amino acids," he said.

In 2012, Fairbrother's lab unveiled free web-based software, Spliceman, which analyzes DNA sequences to determine if mutations are likely to cause errors in splicing. Later that year, the lab was part of a team that won the CLARITY contest in which scientists analyzed the whole genomes of three families to find the mutations causing a disease in children from each family.

Meet MaPSy

In the new project, Fairbrother and co-lead authors Rachel Soemedi, a postdoctoral researcher at Brown, and Kamil Cygan, a graduate student, developed a "Massively Parallel Splicing Assay," or "MaPSy," for rapidly screening the splicing implications of 4,964 variations in the Human Gene Mutation Database (HGMD) of disease-causing genetic problems.

MaPSy works by making thousands of artificial genes that can model the effects of disease-causing mutations. The researchers synthesized artificial genes that correspond to "normal" and disease-carrying versions of thousands of genes. These "pooled" [artificial genes](#) are processed in large batches in two modes. In the "in vivo" mode, the scientists introduced both healthy and mutant versions of the synthesized genes into living cells to see how often the normal or mutant genes would be successfully processed.

"We're putting thousands of genes into the cell and seeing which of those genes get processed correctly," Fairbrother said.

In the "in vitro" mode, they focused more directly on splicing by extracting the splicing machinery from the cell nuclei and feeding it synthesized RNA—again both normal and with HGMD mutations—to assess how often errors occurred when mutations were present.

In the in vivo mode, about 18 percent of the HGMD mutations led to splicing errors. In the in vitro mode, about 24 percent did. But most importantly, Fairbrother said, about 10 percent of mutations produced splicing errors in both modes, suggesting an especially strong likelihood that they were indeed sources of splicing error.

Patterns and predictions of problems

The screen did more than implicate nearly 500 hundred disease-causing mutations as splicing error sources. With detailed sequence information on every mutation and every splicing result, the team was able to observe the nature of the different splicing problems mutations cause. They discovered patterns that show which kinds of genes are most vulnerable to splicing problems, and they were able to predict and even fix some splicing errors arising from specific mutations.

For example, the researchers were able to quantify and rank the features of genes and mutations that were most commonly associated with splicing errors. Not surprisingly, variations that affected how recognizable the splicing sites in genes were (e.g. "cut here" areas where splicing is supposed to occur) rated high on the list. They also found that splicing errors were notably common among genes where if only one of a person's two copies had a mutation they'd end up with disease.

In some experiments highlighted in the paper, they demonstrated that they could predict and address particular splicing-related mutations. In one example they looked at a specific variation in a specific region of the gene COL1A2, which has to do with collagen and bone growth. They

predicted that the mutation would create an unwanted binding site for a protein that prevents a splicing that would normally occur. When they intervened in cells with that mutation by knocking out the binding site, they rescued the splicing process.

Eager to ensure their findings were valid beyond the lab bench, they also sought out tissues from patients with any of the mutations of interest. In many cases they were able to find in those real-life samples evidence of the splicing errors predicted in MaPSy.

They conclude that MaPSy "is a powerful tool for characterization of the sequence variation underlying splicing aberrations."

  More information: Pathogenic variants that alter protein code often disrupt splicing, *Nature Genetics* (2017). nature.com/articles/doi:10.1038/ng.3837

Provided by Brown University

Citation: Assay of nearly 5,000 mutations reveals roots of genetic splicing errors (2017, April 17) retrieved 25 April 2024 from https://medicalxpress.com/news/2017-04-assay-mutations-reveals-roots-genetic.html