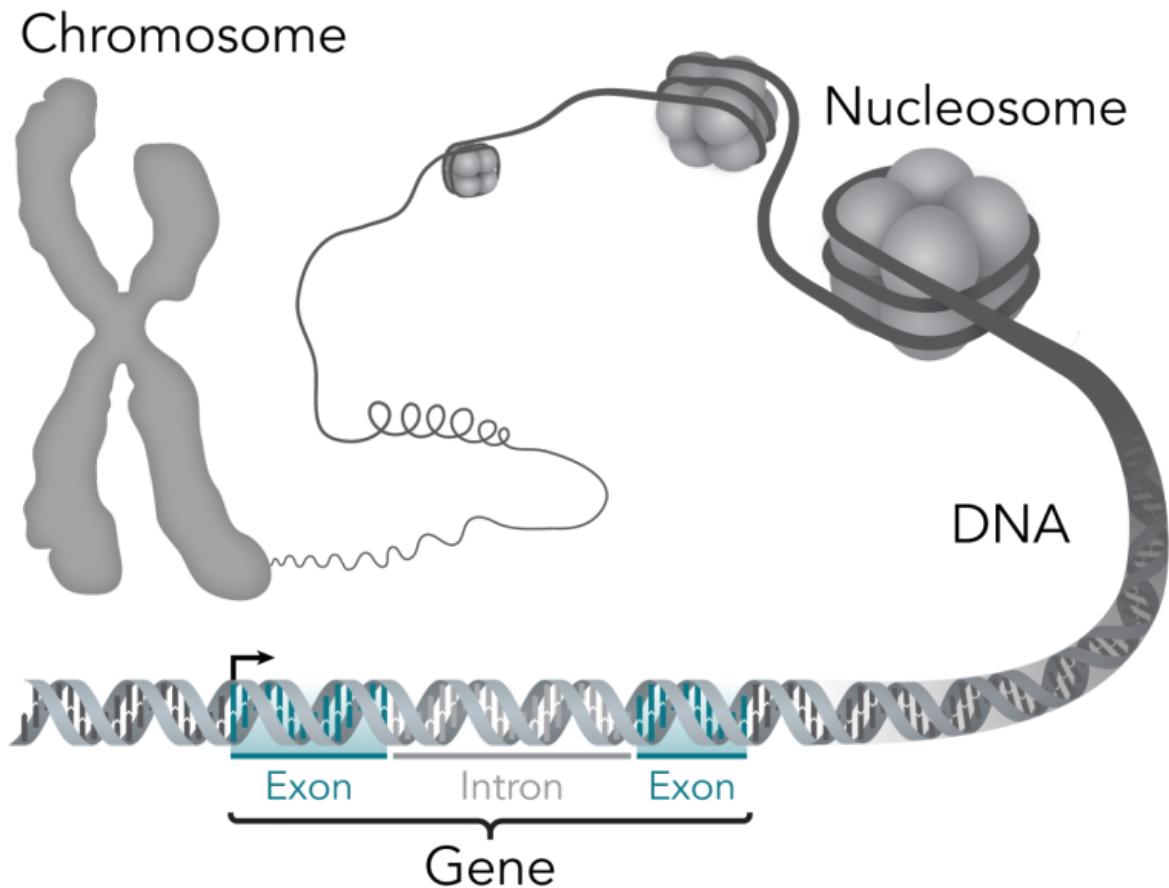# Knowing exactly what genes are saying – and where

December 3 2018

This stylistic diagram shows a gene in relation to the double helix structure of DNA and to a chromosome (right). The chromosome is X-shaped because it is dividing. Introns are regions often found in eukaryote genes that are removed in the splicing process (after the DNA is transcribed into RNA): Only the exons encode the protein. The diagram labels a region of only 55 or so bases as a gene. In reality, most genes are hundreds of times longer. Credit: Thomas

Scientists can now discover how the fine details of gene activity differ from one cell type to another in a tissue sample, thanks to a technique invented by Weill Cornell Medicine researchers.

The technique, described in a paper published Oct. 15 in *Nature Biotechnology*, will enable biologists to better understand the distinct molecular workings of different cell types in the body. It may also enable the improved understanding and treatment of diseases caused by abnormal gene activity.

"An individual gene can 'say' different things, and the true meaning often requires listening to entire phrases, rather than single words," said senior study author Dr. Hagen U. Tilgner, assistant professor of neuroscience in the Feil Family Brain and Mind Research Institute at Weill Cornell Medicine. "Our new method essentially allows us to record complete phrases, called isoforms, that each gene expresses in each cell."

A key function of genes is to store the codes for making proteins, the workhorse molecules of cells. When a protein-coding gene is active in a cell, enzymes repeatedly copy it out into individual RNA molecules called transcripts. These are processed further to become messenger RNA molecules (mRNAs), which are meant to carry the specific instructions for building a given protein to the protein-making machinery of the cell. The catch is that a given gene does not always produce the same mRNA. Depending on the circumstances of the cell or the cell type, the gene's RNA transcript may be processed—sliced up and respliced—into different mRNA isoforms, which in turn encode different proteins. This is one of nature's ways of doing more with less.

Scientists know that the hundreds of different cell types in the body vary not just in their patterns of active genes, but also in the specific mRNA isoforms those genes produce. Yet there haven't been efficient techniques for distinguishing the different mRNA isoforms produced in different cell types, especially when the sample of cells is a bulk tissue sample containing multiple, mixed-together cell types. "We haven't had a technology to record, in thousands of cells, the exact mRNA isoforms the cells are producing," said co-first author Dr. Ishaan Gupta, a postdoctoral researcher in Dr. Tilgner's lab.

In the new technique, cells in a large sample are trapped, one by one, within tiny droplets of fluid. In each trapped cell the mRNAs are converted into more stable DNA molecules, which are then tagged with a unique DNA marker—a "barcode"—identifying that cell. These tens of thousands of converted mRNAs can then be sequenced using so-called short-read and long-read techniques. The short-read sequencing data provide the big picture of gene activity and allow an identification of the cell type, such as a neuron or an immune cell. The long-read sequencing data reveal the specific mRNA isoforms being produced by each active gene in the cell. The barcodes tie these mRNA sequence data to individual cells.

Using the technique, known as ScISOr-Seq (Single-cell ISOform RNA-Sequencing), the scientists were able to take a sample of mouse brain tissue containing about 6,000 cells, group the cells into different cell types by their gene activity patterns, and then identify the different mRNA isoforms produced in each cell type. The findings just from this initial demonstration revealed thousands of mRNA isoforms never described in mouse brains before.

Dr. Tilgner and colleagues now plan to extend their ScISOr-Seq-based studies to more tissues and cell types. They also intend to use the technique to compare mRNA isoforms in diseased cells to those in

healthy [cells](). Abnormal mRNA isoforms are increasingly recognized as causes of diseases, including some cancers.

"ScISOr-Seq has the potential to reveal many more disease-causing isoforms and the cell types in which they act," said co-first author Paul Collier, a staff associate in neuroscience in Dr. Tilgner's lab, "which could lead in turn to new disease treatments."