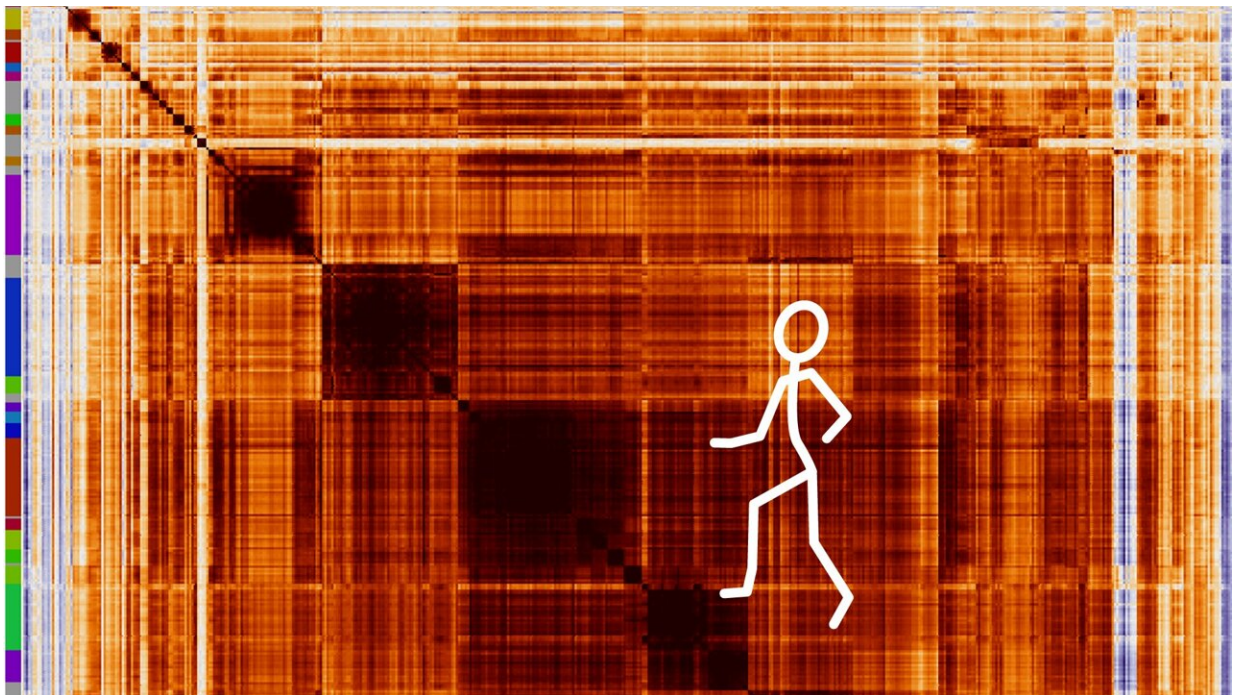


# Researchers use CRISPR technology to create map tying every human gene to its function

June 9 2022

---



A person climbs a set of "stairs" on a gene perturbation heat map. Credit: Jen Cook-Chrysos/Joseph Replogle/Whitehead Institute

The Human Genome Project was an ambitious initiative to sequence every piece of human DNA. The project drew together collaborators from research institutions around the world, including Whitehead

Institute, and was finally completed in 2003. Now, over two decades later, Whitehead Institute Member Jonathan Weissman and colleagues have gone beyond the sequence to present the first comprehensive functional map of genes that are expressed in human cells. The data from this project, published online June 9 in *Cell*, ties each gene to its job in the cell, and is the culmination of years of collaboration on the single-cell sequencing method Perturb-seq.

The data is available on the Weissman Lab website for other scientists to use. "It's a big resource in the way the [human genome](#) is a big resource, in that you can go in and do discovery-based research," said Weissman, who is also a professor of biology at Massachusetts Institute of Technology (MIT) and investigator with the Howard Hughes Medical Institute. "Rather than defining ahead of time what biology you're going to be looking at, you have this map of the genotype-phenotype relationships and you can go in and screen the database without having to do any experiments."

The screen allowed the researchers to delve into diverse biological questions. They used it to explore the cellular effects of genes with unknown functions, to investigate the response of mitochondria to stress, and to screen for genes that cause chromosomes to be lost or gained, a phenotype that has proved difficult to study in the past. "I think this dataset is going to enable all sorts of analyses that we haven't even thought up yet by people who come from other parts of biology and suddenly they just have this available to draw on," said former Weissman Lab postdoc Tom Norman, a co-senior author of the paper.

## **Pioneering Perturb-seq**

The project takes advantage of the Perturb-seq approach which makes it possible to follow the impact of turning on or off genes with unprecedented depth. This method [was first published in 2016](#) by a group

of researchers including Weissman and fellow MIT professor Aviv Regev, but could only be used on small sets of genes and at great expense.

The massive Perturb-seq map was made possible by foundational work from Joseph Replogle, an MD-Ph.D. student in Weissman's lab and co-first author of the present paper. Replogle, in collaboration with Norman, who now leads a lab at Memorial Sloan Kettering Cancer Center, Britt Adamson, an assistant professor in the Department of Molecular Biology at Princeton University, and a group at 10x Genomics, set out to create a new version of Perturb-seq that could be scaled up. The researchers published a [proof of concept paper](#) in *Nature Biotechnology* in 2020.

The Perturb-seq method uses CRISPR/Cas9 genome editing to introduce genetic changes into cells, and then uses single-cell RNA sequencing to capture information about the RNAs that are expressed resulting from a given genetic change. Because RNAs control all aspects of how cells behave, this method can help decode the many cellular effects of genetic changes.

Since their initial proof of concept paper, Weissman, Regev and others have used this sequencing method on smaller scales. For example, the researchers [used Perturb-seq in 2021](#) to explore how human and viral genes interact over the course of an infection with HCMV, a common herpesvirus.

In the new study, Replogle and collaborators including Reuben Saunders, a graduate student in Weissman's lab and co-first author of the paper, scaled up the method to the entire genome. Using human blood cancer cell lines as well noncancerous cells derived from the retina, he performed Perturb-seq across more than 2.5 million cells, and used the data to build a comprehensive map tying genotypes to phenotypes.

## Delving into the data

Upon completing the screen, the researchers decided to put their new dataset to use and examine a few biological questions. "The advantage of Perturb-seq is it lets you get a big dataset in an unbiased way," said Tom Norman. "No one knows entirely what the limits of what you can get out of that kind of dataset. Now, the question is, what do you actually do with it?"

The first, most obvious application was to look into genes with unknown functions. Because the screen also read out phenotypes of many known genes, the researchers could use the data to compare unknown genes to known ones and look for similar transcriptional outcomes, which could suggest the gene products worked together as part of a larger complex.

The mutation of one gene called *C7orf26* in particular stood out. Researchers noticed that genes whose removal led to a similar phenotype were part of a protein complex called Integrator that played a role in creating small nuclear RNAs. The Integrator complex is made up of many smaller subunits—previous studies had suggested 14 individual proteins—and the researchers were able to confirm that *C7orf26* made up a fifteenth component of the complex.

They also discovered that the 15 subunits worked together in smaller modules to perform specific functions within the Integrator complex. "Absent this thousand-foot-high view of the situation, it was not so clear that these different modules were so functionally distinct," said Saunders.

Another perk of Perturb-seq is that because the assay focuses on single cells, the researchers could use the data to look at more complex phenotypes that become muddled when they are studied together with data from other cells. "We often take all the cells where 'gene X' is

knocked down and average them together to look at how they changed," Weissman said. "But sometimes when you knock down a gene, different cells that are losing that same gene behave differently, and that behavior may be missed by the average."

The researchers found that a subset of genes whose removal led to different outcomes from cell to cell were responsible for chromosome segregation. Their removal was causing cells to lose a chromosome or pick up an extra one, a condition known as aneuploidy. "You couldn't predict what the transcriptional response to losing this gene was because it depended on the secondary effect of what chromosome you gained or lost," Weissman said. "We realized we could then turn this around and create this composite phenotype looking for signatures of chromosomes being gained and lost. In this way, we've done the first genome-wide screen for factors that are required for the correct segregation of DNA."

"I think the aneuploidy study is the most interesting application of this data so far," Norman said. "It captures a phenotype that you can only get using a single cell readout. You can't go after it any other way."

The researchers also used their dataset to study how mitochondria responded to stress. Mitochondria, which evolved from free-living bacteria, carry 13 genes in their genomes. Within the nuclear DNA, around 1000 genes are somehow related to mitochondrial function. "People have been interested for a long time in how nuclear and mitochondrial DNA are coordinated and regulated in different cellular conditions, especially when a cell is stressed," Replogle said.

The researchers found that when they perturbed different mitochondria-related [genes](#), the nuclear genome responded similarly to many different genetic changes. However, the mitochondrial genome responses were much more variable.

"There's still an open question of why mitochondria still have their own DNA," said Replogle. "A big-picture takeaway from our work is that one benefit of having a separate mitochondrial genome might be having localized or very specific genetic regulation in response to different stressors.

"If you have one mitochondria that's broken, and another one that is broken in a different way, those mitochondria could be responding differentially," Weissman said.

In the future, the researchers hope to use Perturb-seq on different types of [cells](#) besides the cancer cell line they started in. They also hope to continue to explore their map of gene functions, and hope others will do the same. "This really is the culmination of many years of work by the authors and other collaborators, and I'm really pleased to see it continue to succeed and expand," said Norman.

**More information:** Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq, *Cell* (2022).

[doi.org/10.1016/j.cell.2022.05.013](https://doi.org/10.1016/j.cell.2022.05.013)

Provided by Whitehead Institute for Biomedical Research

Citation: Researchers use CRISPR technology to create map tying every human gene to its function (2022, June 9) retrieved 28 June 2024 from <https://phys.org/news/2022-06-crispr-technology-tying-human-gene.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.