Looking Beyond Suspect Genes in Cancer

Fred Hutch is among 10 institutions in the U.S., the U.K. and Europe collaborating to find the function of every protein-coding gene in the human genome

September 11, 2025 By Fred Hutch News Service and John Higgins

---

Near the end of the classic 1942 movie Casablanca when Major Strasser is shot, Captain Renault famously pretends not to know who fired the gun and orders his officers to "round up the usual suspects."

Since the completion of the Human Genome Project in 2003, biologists have taken a similar approach to understanding the functions of more than 20,000 genes that instruct cells how to make proteins, the molecules that do all the cell's work.

They round up the usual suspects involved in cancer and other diseases, devoting 75% of the research to less than 10% of the cell's proteins. That narrow focus biases knowledge about how genes regulate molecular and cellular functions, according to a paper published earlier this year in the journal Nature.

The paper's authors include researchers from Fred Hutch Cancer Center and the University of Washington Tacoma, which are among 10 institutions in the U.S., the U.K. and Europe collaborating to rectify that bias.

They want to find the function of every protein-coding gene in the human genome by discovering what happens or doesn't happen in a cell when that gene is turned off using cell culture models that mimic the function of real tissues and organs.

The Molecular Phenotypes of Null Alleles in Cells, or MorPhiC project is expected to produce an enormous flow of data that will be shared in a publicly available catalogue, beginning with 1,000 protein-coding genes over the first five years of the project.

"This type of consortium enables achievements that would be impossible to accomplish within one institute, and involvement in such consortia also helps enhance the reputation of Fred Hutch," said Wei Sun, PhD, a biostatistician in Fred Hutch's Public Health Sciences Division and a co-author on the Nature paper.

Knocking out genes at scale

Much of what we know about how protein-coding genes influence the structure and behavior of cells comes from detailed investigations of individual genes in a specific context, especially cancer.

Half of human genes have received scant attention, but more than 13,000 research articles have been published on just one of the usual suspects — the TP53 gene — which codes for a protein that plays a key role in suppressing tumors but may cause cancer cells to grow and spread when the gene is mutated. For example, TP53 mutations often are found in Li-Fraumeni syndrome, or LFS, and can drive multiple cancers including sarcomas, brain tumors, breast cancer and leukemia, among others.

Figuring out the cellular and molecular functions of all 20,319 human protein-coding genes requires leveraging recent advances in gene-editing and sequencing as well as new computational approaches to make sense of big datasets generated in different labs using different techniques.

The MorPhiC project uses a proven strategy to understand gene functions that involves removing the gene or its relevant protein and then conducting a series of tests to observe and measure the resulting biological activity to see what changes in the cell.

Modern gene-editing tools such as CRISPR-Cas9 make it feasible to create null alleles at scale, which are variants of a protein-coding gene that prevent the production of a functional protein, knocking it out (or mostly out).

“Previously if you hope to knock out all the genes in the human genome one by one, it’s a huge amount of work and likely infeasible,” Sun said. “Now the CRISPR system can make it doable.”

The National Institutes of Health launched MorPhiC in 2022 and manages the project, which is divided among institutions that produce the data, validate the data and publicize the data.

The UW Tacoma School of Engineering & Technology is part of MorPhiC’s coordinating center, which standardizes data processing and makes results quickly available to the public.

Once data from data production centers are uploaded to the coordinating center, they are sent to three data analysis and validation centers at Fred Hutch/UW, Stanford University and Jackson Laboratory in Maine, which validate and analyze the data, improving its usefulness for the research community.

The data analysis center at Fred Hutch/UW are co-led by Sun and Li Hsu, PhD, a biostatistician who also works in Fred Hutch’s Public Health Sciences Division, and Ali Shojae, PhD, who is a professor of biostatistics and statistics at UW.

A new method to analyze genes in batches rather than one at a time

To understand how genes work together, Fred Hutch and the other data validation centers will use existing computer tools and possibly new ones they invent to study patterns of gene activity and how they regulate each other in key processes such as cell signaling, the cell cycle, metabolism, the immune system, and the development and specialization of cells.

“We want to not just understand the gene functions, but also how it connects with human disease,” Sun said.

The data validation centers may also develop computer models trained on MorPhiC data that find causal relationships between genes that can be generalized to other cell types or tissue microenvironments.

“The causal part is really interesting, and this is something people do not have in the traditional gene expression studies,” Sun said. “You have the expression of two genes, and you find that their expression goes up or down at the same time, but you do not know whether they are causally related.”

Fred Hutch and the other validation centers may develop new computer tools as well.

Sun and Zhexiao Lin, a former graduate student at UW, developed a deep learning method to find gene sets that can be used to distinguish cells by condition (for example, immune cells from patients with mild versus severe COVID-19).

Their method, explained in a recent preprint, helps analyze data produced by single-cell RNA sequencing, which measures gene expression in each cell individually to see how different types of cells behave.

The first release of data reported in the Nature study comprises 11 studies that knock out 71 genes or proteins using techniques that include single-cell RNA sequencing.

Typically, researchers probe that data to see which genes are expressed differently in different groups of cells. Then they look for biological processes that are especially common among the genes that stick out.

The goal is to find meaningful genes that can act like markers to classify which cell is which.

But trudging through the data gene by gene looking for meaningful differences produces long lists of genes that are statistically significant, but not useful because the differences are tiny and make little impact.

To narrow that list to meaningful genes takes time and money, making it impractical for a project

as big as MorPhiC.

Sun's method simplifies the process by identifying sets of genes that turn on and work together and then probing the data set by set instead of gene by gene.

With this approach, Sun's team trained a computer model to accurately classify cells by analyzing batches of genes instead of individual genes.

They used the method to identify gene sets associated with severe COVID-19, dementia, and cancer patients' responses to immunotherapy.

Former Fred Hutch senior staff scientist heads institute managing MorPhiC

The specific institute at NIH managing the MorPhiC project, the National Human Genome Research Institute, has gone through a significant shakeup since President Trump's administration took over in January.

Visitors to MorPhiC's web page are informed in a banner across the top: "Due to reduction in workforce efforts, the information on this website may not be up to date, transactions submitted via the website may not be processed, and the agency may not be able to respond to inquiries."

In March, the director who had served 16 years abruptly left, and his replacement was soon placed unexpectedly on administrative leave.

However, NIH announced in April that that Carolyn Hutter, PhD, had been named the new acting director of NHGRI. Hutter earned her doctorate in epidemiology at the University of Washington and was a senior staff scientist at Fred Hutch before joining NIH in 2012.

An NIH council meeting concerning the next phase of MorPhiC has been scheduled for September, which the participants hope will be, as Humphrey Bogart says at the end of Casablanca, "the beginning of a beautiful friendship" to continue the work.

Federal Funding Helps Us Save Lives

At Fred Hutch, 70% of our research funding comes from federal grants, which are awarded based on scientific merit. This level of federal funding is a direct reflection of our proven ability to make landmark breakthroughs and our trusted ability to lead large scientific collaborations.

Learn more

John Higgins, a staff writer at Fred Hutch Cancer Center, was an education reporter at The Seattle

Times and the Akron Beacon Journal. He was a Knight Science Journalism Fellow at MIT, where he studied the emerging science of teaching. Reach him at jhiggin2@fredhutch.org or @jhigginswriter.bsky.social.

This article was originally published August 21, 2025, by Fred Hutch News Service. It is republished with permission.

© 2026 Smart + Strong All Rights Reserved.
http://beta.docker.cancerhealth.com/article/looking-beyond-suspect-genes-cancer