Researchers Are Using A.I. to Decode the Human Genome

AlphaGenome *is a leap forward in the ability to study the human blueprint. But the fine workings of our DNA are still largely a mystery.*   **By Carl Zimmer**   NYT Jan. 28, 2026

In 2024, two scientists from Google DeepMind shared the Nobel Prize in Chemistry for an artificial-intelligence program called AlphaFold2.

For decades, scientists had struggled to understand how strings of molecular building blocks fold into the complex, three-dimensional structures of proteins. Demis Hassabis, John Jumper and their colleagues at Google DeepMind trained a program to predict the shapes; when AlphaFold2 was introduced in 2020, it performed so well at this task that scientists around the world adopted it.

"Everyone's using AlphaFold," said Alex Palazzo, a geneticist at the University of Toronto.

Scientists used the program to study how proteins normally work — and how the failure to work can lead to disease. It helped them build entirely new proteins, some of which will soon be tested in clinical trials.

Now another team of researchers at Google DeepMind is trying to do for DNA what the company did for proteins. AlphaFold, meet AlphaGenome.

On Wednesday, the researchers unveiled AlphaGenome in the journal Nature. They trained their A.I. on a vast wealth of molecular data, enabling it to make predictions about thousands of genes. For instance, AlphaGenome can predict whether a mutation will shut off a gene or switch it on at the wrong time — a crucial question for understanding cancer and other diseases.

Peter Koo, a computational biologist at Cold Spring Harbor Laboratory in New York who was not involved in the project, said that AlphaGenome represented an important step forward in applying artificial intelligence to the genome. "It's an engineering marvel," he said.

But Dr. Koo and other outside experts cautioned that it represented just one step on a long road ahead. "This is not AlphaFold, and it's not going to win the Nobel Prize," said Mark Gerstein, a computational biologist at Yale.

AlphaGenome will be useful. Dr. Gerstein said that he would probably add it to his toolbox for exploring DNA, and others expect to follow suit. But not all scientists trust A.I. programs like AlphaGenome to help them understand the genome.

"I see no value in them at all right now," said Steven Salzberg, a computational biologist at Johns Hopkins University. "I think there are a lot of smart people wasting their time."

Before the era of computers, biologists conducted painstaking experiments to uncover the rules that govern our genes. They discovered that genes are spelled out in a four-letter genetic alphabet called bases. To make a protein, a cell reads the sequence in a gene, which can run thousands of bases long.

But the more scientists studied the human genome, the more complicated and messy it turned out to be.

As cells read a gene, for example, they often skip over sections of its sequence. Through this process, known as splicing, cells can create hundreds of different proteins from a single gene.

A number of diseases occur when cells splice their genes incorrectly. But there is no simple signature for the spots in genes where they should be spliced, so scientists have spent decades building up a catalog of them.

Another profound question about the genome is how cells choose which genes they use to make proteins. Scientists have discovered special molecules that grab hold of DNA and stretch it into intricate loops. In some cases, the loops expose a gene to the cell's protein-making machinery. In other cases, the gene ends up tucked away in a coil.

Those molecules have to land precisely on tiny stretches of DNA to control genes. And these genetic locks can be hard to find since they often lie thousands or millions of bases away from the genes they control. In 2019, researchers at Google DeepMind embarked on a project that would evolve into AlphaGenome. By then, biologists had amassed huge amounts of data, including not only the three billion base pairs of the human genome but also the results of thousands of experiments measuring the activity of genes in many types of cells.

The researchers at Google DeepMind hoped that by training A.I. on these existing results, they could develop a program that could make accurate predictions about stretches of DNA it had never seen before.

"It was the right target for us," said Ziga Avsec, a research scientist at Google DeepMind.

In 2021, Dr. Avsec and his colleagues unveiled a preliminary A.I. called Enformer, which they have since expanded into AlphaGenome. They trained the program on an even greater expanse of biological data. "It's really an industrial scale," Dr. Gerstein said.

Many A.I. programs built to study the genome tackle only one aspect of it, like splicing. But AlphaGenome was trained to make predictions about 11 different processes. In the report on Wednesday, Dr. Avsec and his colleagues noted that AlphaGenome had performed as well or better than other programs across the board.

"It's state of the art," said Katherine Pollard, a data scientist at Gladstone Institutes, a research organization in San Francisco, who was not involved in the study.

Dr. Pollard and other researchers said that AlphaGenome was particularly adept with mutations, capable of predicting their effects, such as shutting down a nearby gene. In one performance test, the researchers added mutations to the stretch of DNA that includes a gene called TAL1.

In healthy people, TAL1 helps immune cells mature until they can fight pathogens. Once the cells have developed, the gene shuts down. But scientists have [discovered](#) that mutations 8,000 bases away from TAL1 can lead the gene to switch on permanently. That change can ultimately cause immune cells to multiply out of control, causing leukemia.

Dr. Avsec and his colleagues found that AlphaGenome had accurately predicted the impact of these mutations on TAL1. "It has been really exciting to see, when these models work," he said. "It feels like magic sometimes."

The AlphaGenome researchers shared their TAL1 predictions with Dr. Marc Mansour, a hematologist at University College London who spent years uncovering the leukemia-driving mutations with lab experiments.

"It was quite mind-blowing," Dr. Mansour said. "It really showed how powerful this is."

But, Dr. Mansour noted, AlphaGenome's predictive powers fade the farther its gaze strays from a particular gene. He is now using AlphaGenome in his cancer research but does not blindly accept its results.

"These prediction tools are still prediction tools," he said. "We still need to go to the lab."

Dr. Salzberg of Johns Hopkins is less sanguine about AlphaGenome, in part because he thinks its creators put too much trust in the data they trained it on. Scientists who study splice sites don't agree on which sites are real and which are genetic mirages. As a result, they have created databases that contain different catalogs of splice sites.

"The community has been working for 25 years to try to figure out what are all the splice sites in the human genome, and we're still not really there," Dr. Salzberg said. "We don't have an agreed-upon gold-standard set."

Dr. Pollard also cautioned that AlphaGenome was a long way from being a tool that doctors could use to scan the genomes of patients for threats to their health. It predicts only the effects of a single mutation on one standard human genome.

In reality, any two people have millions of genetic differences in their DNA. Assessing the effects of all those variations throughout a patient's body remains far beyond AlphaGenome's industrial-strength power.

"It is a much, much harder problem — and yet that's the problem we need to solve if we want to use a model like this for health care," Dr. Pollard said.