# Google unveils AI tool probing mysteries of human genome

GOOGLE has unveiled an artificial intelligence that its scientists say would help unravel the mysteries of the human genome — and could one day lead to new treatments for diseases.

The deep-learning model AlphaGenome was hailed by outside researchers as a "breakthrough" that would let scientists study and even simulate the roots of difficult-to-treat genetic diseases.

While the first complete map of the human genome in 2003 "gave us the book of life, reading it remained a challenge", Pushmeet Kohli, vice-president of research at Google DeepMind, told journalists.

"We have the text," he said, which is a sequence of three billion nucleotide pairs represented by the letters A, T, C and G that make up DNA.

However, "understanding the grammar of this genome — what is encoded in our DNA and how it governs life — is the next critical frontier for research," said Kohli, co-author of a new study in the journal 'Nature'.

Only around two per cent of our DNA contains instructions for making proteins, which are the molecules that build and run the body.

The other 98 per cent was long dismissed as "junk DNA" as scientists struggled to understand what it was for.

However this "non-coding DNA" is now believed to act like a conductor, directing how genetic information works in each of our cells.

PICTURE CREDIT:
KJPARGETER — FREEPIK

Bots
*By AFP*



*The model has already been tested by 3,000 scientists across 160 countries.*
PICTURE CREDIT:
FREEPIK

These sequences also contain many variants that have been associated with diseases. It is these sequences that AlphaGenome is aiming to understand.

**A MILLION LETTERS**

The project is just one part of Google's AI-powered scientific work, which also includes AlphaFold, the winner of 2024's chemistry Nobel.

AlphaGenome's model was trained on data from public projects that measured non-coding DNA across hundreds of different cell and tissue types in humans and mice.

The tool is able to analyse long DNA sequences then predict how each nucleotide pair will influence different biological processes within the cell.

This includes whether genes start and stop and how much RNA — molecules which transmit genetic instructions inside cells — is produced.

Other models already exist that have a similar aim. However, they have to compromise, either by analysing far shorter DNA sequences or decreasing how detailed their predictions are, known as resolution.

DeepMind scientist and lead study author Ziga Avsec said that long sequences — up to a million DNA letters long — were "required to understand the full regulatory environment of a single gene".

And the high resolution of the model allows scientists to study the impact of genetic variants by comparing the differences between mutated and non-mutated sequences.

"AlphaGenome can accelerate our understanding of the genome by helping to map where the functional elements are and what their roles are on a molecular level," study co-author Natasha Latysheva said.

The model has already been tested by 3,000 scientists across 160 countries and is open for anyone to use for non-commercial reasons, Google said.

"We hope researchers will extend it with more data," Kohli added.

**'BREAKTHROUGH'**

Ben Lehner, a researcher at Cambridge University who was not involved in developing AlphaGenome but did test it, said the model "does indeed perform very well".

"Identifying the precise differences in our genomes that make us more or less likely to develop thousands of diseases is a key step towards developing better therapeutics," he explained.

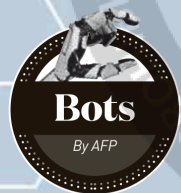However AlphaGenome "is far from perfect and there is still a lot of work to do.

"AI models are only as good as the data used to train them" and the existing data is not very suitable, he said.

Robert Goldstone, head of genomics at the UK's Francis Crick Institute, cautioned that AlphaGenome was "not a magic bullet for all biological questions".

This was partly because "gene expression is influenced by complex environmental factors that the model cannot see," he said.

However the tool still represented a "breakthrough" that would allow scientists to "study and simulate the genetic roots of complex disease", Goldstone added.