

DeepMind launches AlphaGenome to predict how DNA variants affect gene regulation

Source copy kept by wheresthe.ai for the story "DeepMind launches AlphaGenome to predict how DNA variants affect gene regulation" (<https://wheresthe.ai/d/gb-deepmind-alphagenome-2025-06/>).

Source: <https://deepmind.google/blog/alphagenome-ai-for-better-understanding-the-genome/>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 739731a5c16dcc9713353f0c3500bf626e604f6a2e7803cfe06b246dbb2e624e

June 25, 2025 Science AlphaGenome: AI for better understanding the genome Ziga Avsec and Natasha Latysheva

Share

Introducing a new, unifying DNA sequence model that advances regulatory variant-effect prediction and promises to shed new light on genome function — now available via API. Update January 2026: This research has been published in Nature. You can read the full paper [here](#) and access the model [here](#). The genome is our cellular instruction manual. It's the complete set of DNA which guides nearly every part of a living organism, from appearance and function to growth and reproduction. Small variations in a genome's DNA sequence can alter an organism's response to its environment or its susceptibility to disease. But deciphering how the genome's instructions are read at the molecular level — and what happens when a small DNA variation occurs — is still one of biology's greatest mysteries. Today, we introduce AlphaGenome, a new artificial intelligence (AI) tool that more comprehensively and accurately predicts how single variants or mutations in human DNA sequences impact a wide range of biological processes regulating genes. This was enabled, among other factors, by technical advances allowing the model to process long DNA sequences and output high-resolution predictions. To advance scientific research, we're making AlphaGenome available in preview via our AlphaGenome API for non-commercial research, and planning to release the model in the future. We believe AlphaGenome can be a valuable resource for the scientific community, helping scientists better understand genome function, disease biology, and ultimately, drive new biological discoveries and the development of new treatments. How AlphaGenome works Our AlphaGenome model takes a long DNA sequence as input — up to 1 million letters, also known as base-pairs — and predicts thousands of molecular properties characterising its regulatory activity. It can also score the effects of genetic variants or mutations by comparing predictions of mutated sequences with unmutated ones. Predicted properties include where genes start and where they end in different cell types and tissues, where they get spliced, the amount of RNA being produced, and also which DNA bases are accessible, close to one another, or bound by certain proteins. Training data was sourced from large public consortia including ENCODE, GTEx, 4D Nucleome and FANTOM5, which experimentally measured these properties covering important modalities of gene regulation across hundreds of human and mouse cell types and tissues.

Your browser does not support the video tag.

Animation showing AlphaGenome taking one million DNA letters as input and predicting diverse molecular properties across different tissues and cell types.

The AlphaGenome architecture uses convolutional layers to initially detect short patterns in the genome sequence, transformers to communicate information across all positions in the sequence, and a final series of layers to turn the detected patterns into predictions for different modalities. During training, this computation is distributed across multiple interconnected Tensor Processing Units (TPUs) for a single sequence. This model builds on our previous genomics model, Enformer and is complementary to AlphaMissense, which specializes in categorizing the effects of variants within protein-coding regions. These regions cover 2% of the genome. The remaining 98%, called non-coding regions, are crucial for orchestrating gene activity and contain many variants linked to diseases. AlphaGenome offers a new perspective for interpreting these expansive sequences and the variants within them. AlphaGenome's distinctive features AlphaGenome offers several distinctive features compared to existing DNA sequence models: Long sequence-context at high resolution Our model analyzes up to 1 million DNA letters and makes predictions at the resolution of individual letters. Long sequence context is important for covering regions regulating genes from far away and base-resolution is important for capturing fine-grained biological details. Previous models had to trade off sequence length and resolution, which limited the range of modalities they could jointly model and accurately predict. Our technical advances address this limitation

without significantly increasing the training resources — training a single AlphaGenome model (without distillation) took four hours and required half of the compute budget used to train our original Enformer model.

Comprehensive multimodal prediction By unlocking high resolution prediction for long input sequences, AlphaGenome can predict the most diverse range of modalities. In doing so, AlphaGenome provides scientists with more comprehensive information about the complex steps of gene regulation.

Efficient variant scoring In addition to predicting a diverse range of molecular properties, AlphaGenome can efficiently score the impact of a genetic variant on all of these properties in a second. It does this by contrasting predictions of mutated sequences with unmutated ones, and efficiently summarising that contrast using different approaches for different modalities.

Novel splice-junction modeling Many rare genetic diseases, such as spinal muscular atrophy and some forms of cystic fibrosis, can be caused by errors in RNA splicing — a process where parts of the RNA molecule are removed, or “spliced out”, and the remaining ends rejoined. For the first time, AlphaGenome can explicitly model the location and expression level of these junctions directly from sequence, offering deeper insights about the consequences of genetic variants on RNA splicing.

State-of-the-art performance across benchmarks AlphaGenome achieves state-of-the-art performance across a wide range of genomic prediction benchmarks, such as predicting which parts of the DNA molecule will be in close proximity, whether a genetic variant will increase or decrease expression of a gene, or whether it will change the gene’s splicing pattern.

Bar graph showing AlphaGenome’s relative improvements on selected DNA sequence and variant effect tasks, compared against results for the current best methods in each category.

When producing predictions for single DNA sequences, AlphaGenome outperformed the best external models on 22 out of 24 evaluations. And when predicting the regulatory effect of a variant, it matched or exceeded the top-performing external models on 24 out of 26 evaluations. This comparison included models specialized for individual tasks. AlphaGenome was the only model that could jointly predict all of the assessed modalities, highlighting its generality. Read more in our preprint . The benefits of a unifying model AlphaGenome’s generality allows scientists to simultaneously explore a variant’s impact on a number of modalities with a single API call. This means that scientists can generate and test hypotheses more rapidly, without having to use multiple models to investigate different modalities. Moreover AlphaGenome’s strong performance indicates it has learned a relatively general representation of DNA sequence in the context of gene regulation. This makes it a strong foundation for the wider community to build upon. Once the model is fully released, scientists will be able to adapt and fine-tune it on their own datasets to better tackle their unique research questions. Finally, this approach provides a flexible and scalable architecture for the future. By extending the training data, AlphaGenome’s capabilities could be extended to yield better performance, cover more species, or include additional modalities to make the model even more comprehensive.

It’s a milestone for the field. For the first time, we have a single model that unifies long-range context, base-level precision and state-of-the-art performance across a whole spectrum of genomic tasks.

Dr. Caleb Lareau Memorial Sloan Kettering Cancer Center

A powerful research tool AlphaGenome’s predictive capabilities could help several research avenues:

- Disease understanding:** By more accurately predicting genetic disruptions, AlphaGenome could help researchers pinpoint the potential causes of disease more precisely, and better interpret the functional impact of variants linked to certain traits, potentially uncovering new therapeutic targets. We think the model is especially suitable for studying rare variants with potentially large effects, such as those causing rare Mendelian disorders.
- Synthetic biology:** Its predictions could be used to guide the design of synthetic DNA with specific regulatory function — for example, only activating a gene in nerve cells but not muscle cells.
- Fundamental research:** It could accelerate our understanding of the genome by assisting in mapping its crucial functional elements and defining their roles, identifying the most essential DNA instructions for regulating a specific cell type’s function. For example, we used AlphaGenome to investigate the potential mechanism of a cancer-associated mutation. In an existing study of patients with T-cell acute lymphoblastic leukemia (T-ALL) , researchers observed mutations at particular locations in the genome. Using AlphaGenome, we predicted that the mutations would activate a nearby gene called TAL1 by introducing a MYB DNA binding motif, which replicated the known disease mechanism and highlighted AlphaGenome’s ability to link specific non-coding variants to disease genes.

AlphaGenome will be a powerful tool for the field. Determining the relevance of different non-coding variants can be extremely challenging, particularly to do at scale. This tool will provide a crucial piece of the puzzle, allowing us to make better connections to understand diseases like cancer.

Professor Marc Mansour University College London

Current limitations AlphaGenome marks a significant step forward, but it's important to acknowledge its current limitations. Like other sequence-based models, accurately capturing the influence of very distant regulatory elements, like those over 100,000 DNA letters away, is still an ongoing challenge. Another priority for future work is further increasing the model's ability to capture cell- and tissue-specific patterns. We haven't designed or validated AlphaGenome for personal genome prediction, a known challenge for AI models. Instead, we focused more on characterising the performance on individual genetic variants. And while AlphaGenome can predict molecular outcomes, it doesn't give the full picture of how genetic variations lead to complex traits or diseases. These often involve broader biological processes, like developmental and environmental factors, that are beyond the direct scope of our model. We're continuing to improve our models and gathering feedback to help us address these gaps. Enabling the community to unlock AlphaGenome's potential AlphaGenome is now available for non-commercial use via our AlphaGenome API . Please note that our model's predictions are intended only for research use and haven't been designed or validated for direct clinical purposes. Researchers worldwide are invited to get in touch with potential use-cases for AlphaGenome and to ask questions or share feedback through the community forum . We hope AlphaGenome will be an important tool for better understanding the genome and we're committed to working alongside external experts across academia, industry, and government organizations to ensure AlphaGenome benefits as many people as possible. Together with the collective efforts of the wider scientific community, we hope it will deepen our understanding of the complex cellular processes encoded in the DNA sequence and the effects of variants, and drive exciting new discoveries in genomics and healthcare.

Learn more about AlphaGenome

Read our paper in Nature Use the AlphaGenome API Join the community forum Express interest for commercial access

Acknowledgements We would like to thank Juanita Bawagan, Arielle Bier, Stephanie Booth, Irina Andronic, Armin Senoner, Dhavanthi Hariharan, Rob Ashley, Agata Laydon and Kathryn Tunyasuvunakool for their help with the text and figures. This work was done thanks to the contributions of the AlphaGenome co-authors: Žiga Avsec, Natasha Latysheva, Jun Cheng, Guido Novati, Kyle R. Taylor, Tom Ward, Clare Bycroft, Lauren Nicolaisen, Eirini Arvaniti, Joshua Pan, Raina Thomas, Vincent Dutordoir, Matteo Perino, Soham De, Alexander Karollus, Adam Gayoso, Toby Sargeant, Anne Mottram, Lai Hong Wong, Pavol Drotár, Adam Kosiorek, Andrew Senior, Richard Tanburn, Taylor Applebaum, Souradeep Basu, Demis Hassabis and Pushmeet Kohli. We would also like to thank Dhavanthi Hariharan, Charlie Taylor, Ottavia Bertolli, Yannis Assael, Alex Botev, Anna Trostanetski, Lucas Tenório, Victoria Johnston, Richard Green, Kathryn Tunyasuvunakool, Molly Beck, Uchechi Okereke, Rachael Tremlett, Sarah Chakera, Ibrahim I. Taskiran, Andreea-Alexandra Muşat, Raiyan Khan, Ren Yi and the greater Google DeepMind team for their support, help and feedback.