This Week in Genomics
September 25,2023

AlphaFold tool pinpoints protein mutations that cause disease

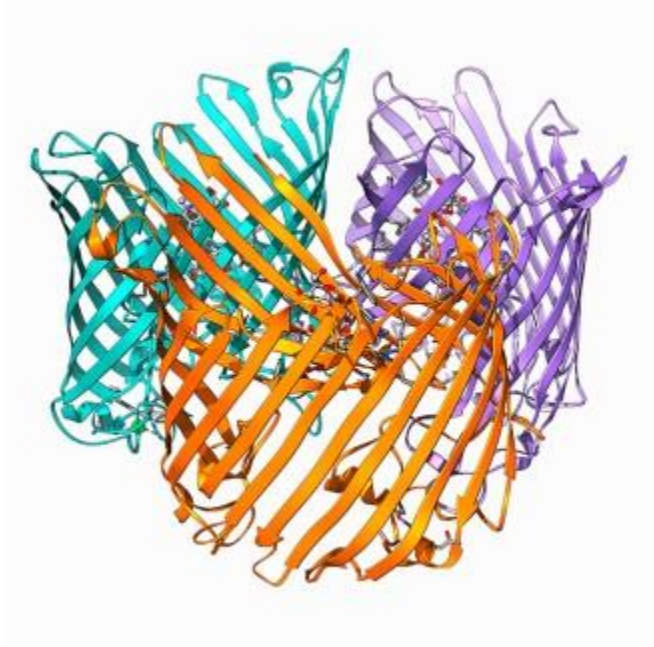
Researchers have adapted the AI network to search for genetic changes linked to ill health.

https://www.nature.com/articles/d41586-023-02943-5

Google DeepMind has wielded its revolutionary protein-structure-prediction AI in the hunt for genetic mutations that cause disease.

A new tool based on the AlphaFold network can ***accurately predict which mutations in proteins are likely to cause health conditions*** — a challenge that limits the use of genomics in healthcare.

The ***AI network*** — called ***AlphaMissense*** — is a ***step forward***, say researchers who are developing similar tools, but not necessarily a sea change. It is ***one of many techniques in development that aim to help researchers***, and ultimately physicians, to ***'interpret' people's genomes*** to ***find the cause of a disease***. But tools such as ***AlphaMissense*** — which is described in a 19 September paper in *Science*1 — will need to ***undergo thorough testing*** before they are used in the clinic.



'A Pandora's box': map of protein-structure families delights scientists

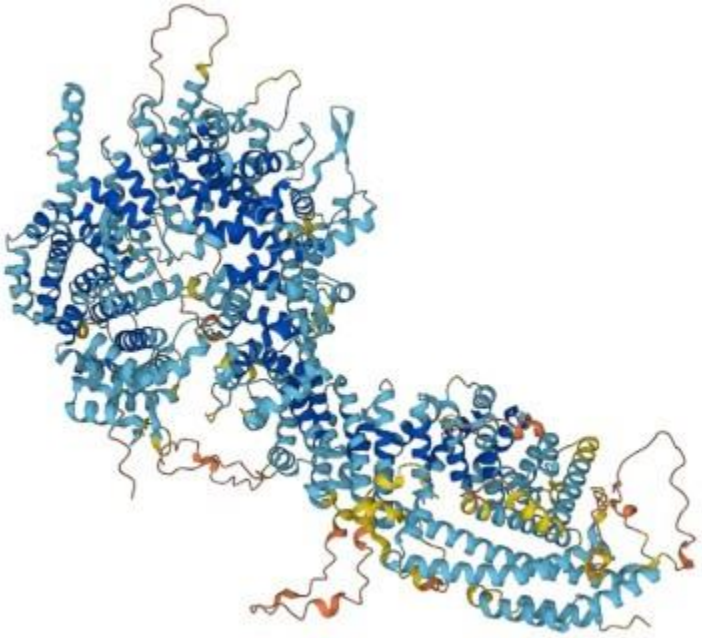
Many of the genetic mutations that directly cause a condition, such as those responsible for ***cystic fibrosis*** and ***sickle-cell disease***, tend to ***change the amino acid sequence*** of the protein they encode. But ***researchers have observed*** only a ***few million*** of these ***single-letter 'missense mutations***'. Of the more than ***70 million possible in the human genome***, only a ***sliver have been conclusively linked to disease***, and ***most seem to have no ill effect on health***.

So when researchers and ***doctors find a missense mutation*** they've never seen before, it can be ***difficult to know what to make of it***. To help ***interpret*** such ***'variants of unknown significance'***, researchers have developed dozens of different computational tools that can ***predict whether a variant is likely to cause disease***. ***AlphaMissense*** ***incorporates existing approaches*** to the problem, which ***are increasingly being addressed with machine learning***.

Locating mutations

The *network is based on AlphaFold*, which *predicts a protein structure from an amino-acid sequence*. But instead of determining the structural effects of a mutation — an open challenge in biology — *AlphaMissense uses AlphaFold's 'intuition' about structure to identify where disease-causing mutations are likely to occur within a protein*, Pushmeet Kohli, DeepMind's vice-president of Research and a study author, said at a press briefing.

AlphaMissense also ***incorporates a type of neural network*** inspired by ***large language models like ChatGPT*** that has been ***trained on millions of protein sequences*** instead of words, called a ***protein language model***. These have proven ***adept at predicting protein structures and designing new proteins***. They are ***useful for variant prediction*** because they have ***learned which sequences are plausible and which are not***, Žiga Avsec, the DeepMind research scientist who co-led the study, told journalists.



Foldseek gives AlphaFold protein database a rapid search tool

DeepMind's network seems to ***outperform other computational tools at discerning variants known to cause disease*** from those that don't. It also does well at ***spotting problem variants identified in laboratory experiments*** that measure the ***effects of thousands of mutations at once***. The researchers also ***used AlphaMissense*** to ***create a catalogue of every possible missense mutation in the human genome***, determining that ***57% are likely to be benign*** and that ***32% may cause disease***.

Clinical support

AlphaMissense is an advance over existing tools for predicting the effects of mutations, "***but not a gigantic leap forward***", says Arne Elofsson, a computational biologist at the University of Stockholm.

Its impact won't be as significant as AlphaFold, which ***ushered in a new era in computational biology***, agrees Joseph Marsh, a computational biologist at the MRC Human Genetics Unit in Edinburgh, UK. "It's exciting. ***It's probably the best predictor we have right now***. But ***will it be the best predictor in two or three years***? There's a good chance it won't be."

Computational predictions currently have a minimal role in diagnosing genetic diseases, says Marsh, and recommendations from ***physicians' groups say that these tools should provide only supporting evidence in linking a mutation to a disease***. ***AlphaMissense confidently classified a much larger proportion of missense mutations than have previous methods***, says Avsec. "As these ***models get better*** than I think ***people will be more inclined to trust them***."

Yana Bromberg, a bioinformatician at Emory University in Atlanta, Georgia, emphasizes that ***tools such as AlphaMissense must be rigorously evaluated*** — using good performance metrics — ***before ever being applied in the real-world***.

For example, an exercise called the ***Critical Assessment of Genome Interpretation (CAGI)*** has ***benchmarked the performance*** of such prediction methods for years ***against experimental data that*** has not yet been released. "***It's my worst nightmare to think of a doctor taking a prediction and running with it, as if it's a real thing, without evaluation by entities such as CAGI***," Bromberg adds.