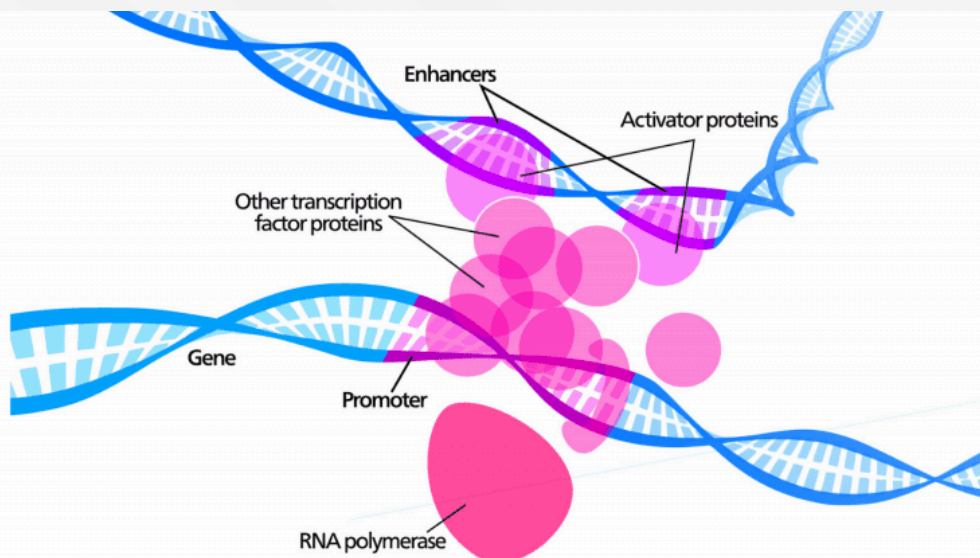


[← BACK TO STORIES](#)

Calico Stories

Improving the ability to predict how DNA sequence influences gene expression



We value your privacy

Our site may use certain essential "cookie" technologies in order to function. By clicking "Accept Optional Cookies", you agree to our use of certain optional cookies to help us understand and enhance the site's operation. You may also choose to "Reject Optional Cookies" to disable non-essential cookies when you use our site. For more information, read our full [Privacy Policy](#).

[Reject Optional Cookies](#)[Accept Optional Cookies](#)

color to disease susceptibility. The remaining base pairs are called "noncoding" and contain less well-understood instructions about when, where and how much protein-coding genes should be produced or expressed.

These instructions are critical to healthy development and lifespan, and mutations in noncoding DNA can affect physical traits like height and cause diseases such as cancer. Understanding how these noncoding sequences and their variants function could help explain if and when a person could be at risk for disease.

To unlock a deeper understanding of noncoding DNA, Principal Investigator [David Kelley](#) and scientist Vikram Agarwal started a collaboration with a team of [DeepMind](#) researchers led by Žiga Avsec. Together, they developed a new deep learning model named Enformer to better predict how DNA sequence determines gene expression.

New technology can "see" noncoding DNA further from the gene

Enformer's neural network architecture is based on Basenji, a previous program published by David's group that can predict gene regulatory activity (one main function of noncoding DNA) from DNA sequence. Basenji considers ~40,000 local base pairs to make its predictions, but many known regulatory elements occur at distances greater than that. These include "enhancers," which are noncoding DNA elements that can increase gene expression even at a million base pairs away from a gene.

The previous Basenji model could not see far enough from genes and missed many

We value your privacy

Our site may use certain essential "cookie" technologies in order to function. By clicking "Accept Optional Cookies", you agree to our use of certain optional cookies to help us understand and enhance the site's operation. You may also choose to "Reject Optional Cookies" to disable non-essential cookies when you use our site. For more information, read our full [Privacy Policy](#).

After demonstrating that the new models improve gene prediction accuracy from 0.64 to 0.71 Spearman correlation on gene sequences they've never seen before, the team studied the long-range behavior in depth, bringing in datasets generated by labs around the world to profile enhancer sequence activity to further validate the model. They went on to report more accurate predictions for how human genetic variants (or mutations) affect gene expression for both natural genetic variants in the population and designed mutations measured by reporter assays in the lab. The improvement represents a modest increase over previous methods, but it is well-supported by a diverse set of analyses.

An advance in genetic research

Enformer is a step forward toward making better use of noncoding DNA in genetic research. Since most human disease associations involve noncoding DNA, the advances the team made will help researchers determine the causal variants and the genes they affect in specific cell types. The team intends to collaborate with other researchers at Calico and beyond to ultimately help the scientific community improve their understanding of why one person might develop a disease while another doesn't, and also offer insights into effective treatments.

"There is this untapped potential of disease associations that require new tools and strategies to understand," said David. "Every association for which we can work out the details offers a seed that will hopefully grow into better understanding of diseases and how we might treat them to slow the aging process and improve quality of life."

We value your privacy

Our site may use certain essential "cookie" technologies in order to function. By clicking "Accept Optional Cookies", you agree to our use of certain optional cookies to help us understand and enhance the site's operation. You may also choose to "Reject Optional Cookies" to disable non-essential cookies when you use our site. For more information, read our full [Privacy Policy](#).

influence gene expression, including enhancers that contact the promoter from a distance.

JOIN OUR TEAM

JOIN OUR TEAM

More stories

We value your privacy

Our site may use certain essential "cookie" technologies in order to function. By clicking "Accept Optional Cookies", you agree to our use of certain optional cookies to help us understand and enhance the site's operation. You may also choose to "Reject Optional Cookies" to disable non-essential cookies when you use our site. For more information, read our full [Privacy Policy](#).



We value your privacy

Our site may use certain essential "cookie" technologies in order to function. By clicking "Accept Optional Cookies", you agree to our use of certain optional cookies to help us understand and enhance the site's operation. You may also choose to "Reject Optional Cookies" to disable non-essential cookies when you use our site. For more information, read our full [Privacy Policy](#).

We value your privacy

Our site may use certain essential "cookie" technologies in order to function. By clicking "Accept Optional Cookies", you agree to our use of certain optional cookies to help us understand and enhance the site's operation. You may also choose to "Reject Optional Cookies" to disable non-essential cookies when you use our site. For more information, read our full [Privacy Policy](#).