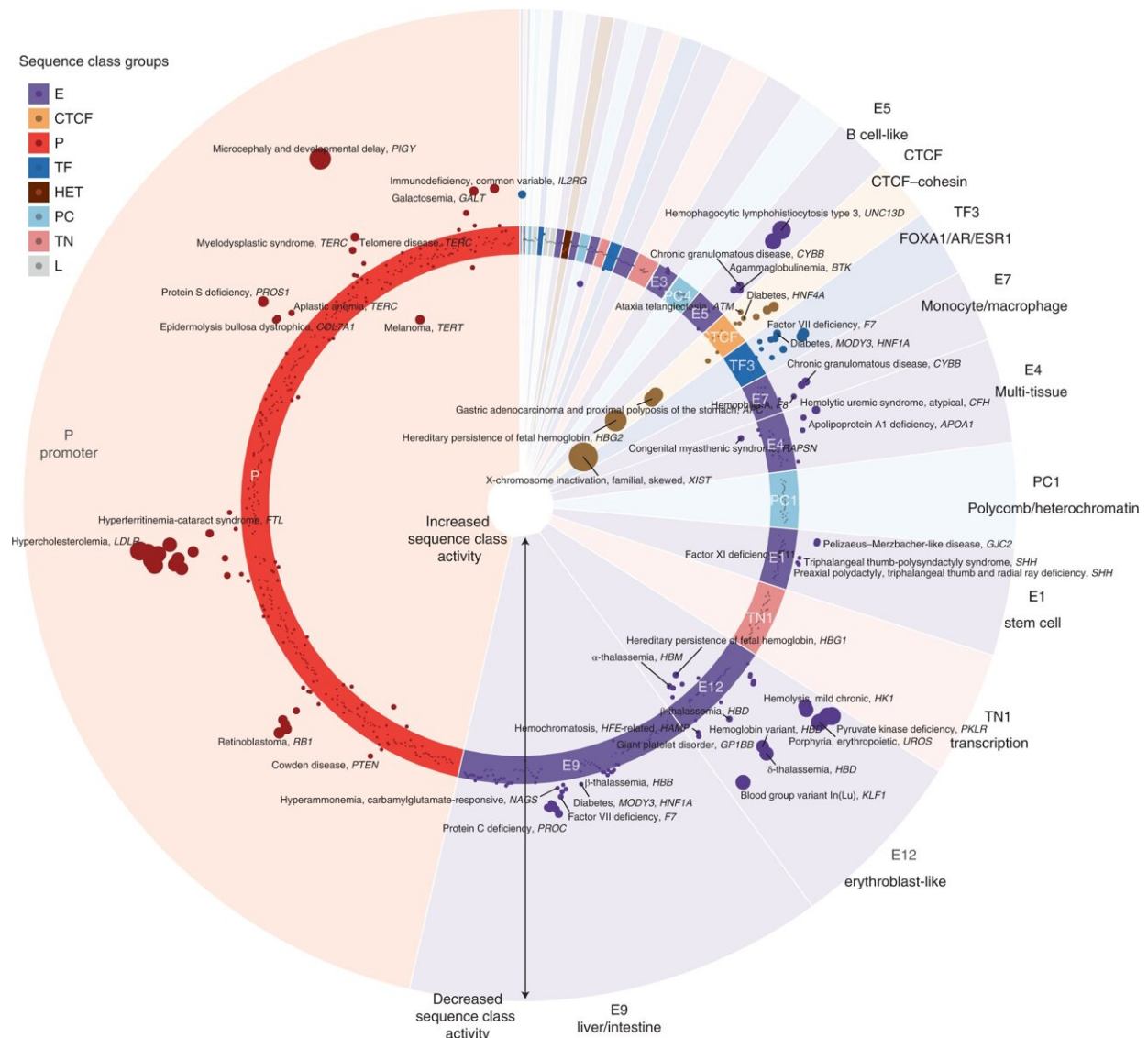


Artificial intelligence tools predict DNA's regulatory role and 3D structure

August 8 2022



Disease regulatory mutations are predicted to disrupt promoter, CTCF and tissue-specific enhancer sequence classes. Sequence class-level mutation effects of

pathogenic noncoding HGMD mutations were plotted. A polar coordinate system was used, where the radial coordinate indicates the sequence class-level effects. Each dot represents a mutation and mutations inside the circle are predicted to have positive effects (increased activity of sequence class); mutations outside the circle are predicted to have negative effects (decreased activity of sequence class). Dot size indicates the absolute value of the effect. Mutations were assigned to sequence classes based on their sequences and predicted effects. Within each sequence class, mutations were ordered by chromosomal coordinates. The associated disease and gene name were annotated for each mutation and only the strongest mutation was annotated if there were multiple mutations associated with the same disease, gene and sequence class. Credit: *Nature Genetics* (2022). DOI: 10.1038/s41588-022-01102-2

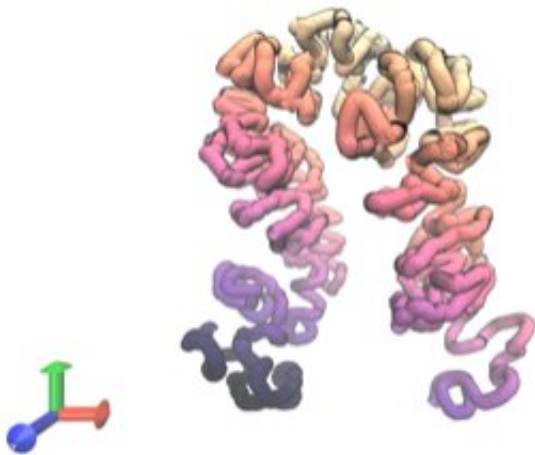
Newly developed artificial intelligence (AI) programs have accurately predicted the role of DNA's regulatory elements and three-dimensional (3D) structure based solely on its raw sequence, according to two recent studies in *Nature Genetics*. These tools could eventually shed new light on how genetic mutations lead to disease and could lead to new understanding of how genetic sequence influences the spatial organization and function of chromosomal DNA in the nucleus, said study author Jian Zhou, Ph.D., Assistant Professor in the Lyda Hill Department of Bioinformatics at UTSW.

"Taken together, these two programs provide a more complete picture of how changes in DNA sequence, even in noncoding regions, can have dramatic effects on its spatial organization and function," said Dr. Zhou, a member of the Harold C. Simmons Comprehensive Cancer Center, a Lupe Murchison Foundation Scholar in Medical Research, and a Cancer Prevention and Research Institute of Texas (CPRIT) Scholar.

Only about 1% of human DNA encodes instructions for making proteins. Research in recent decades has shown that much of the

remaining noncoding [genetic material](#) holds regulatory elements—such as promoters, enhancers, silencers, and insulators—that control how the coding DNA is expressed. How sequence controls the functions of most of these regulatory elements is not well understood, Dr. Zhou explained.

To better understand these regulatory components, he and colleagues at Princeton University and the Flatiron Institute developed a deep learning model they named Sei, which accurately sorts these snippets of noncoding DNA into 40 "sequence classes" or jobs—for example, as an enhancer for stem cell or brain cell gene activity. These 40 sequence classes, developed using nearly 22,000 data sets from previous studies studying genome regulation, cover more than 97% of the human genome. Moreover, Sei can score any sequence by its predicted activity in each of the 40 sequence classes and predict how mutations impact such activities.



Predicted 3D structure for a segment of human genomic DNA. Credit: UT Southwestern Medical Center

By applying Sei to human genetics data, the researchers were able to characterize the regulatory architecture of 47 traits and diseases recorded in the UK Biobank database and explain how mutations in [regulatory elements](#) cause specific pathologies. Such capabilities can help gain a more systematic understanding of how genomic sequence changes are linked to diseases and other traits. The findings were published this month.

In May, Dr. Zhou reported the development of a different tool, called Orca, which predicts the 3D architecture of DNA in chromosomes based on its sequence. Using existing data sets of DNA sequences and structural data derived from previous studies that revealed the molecule's folds, twists, and turns, Dr. Zhou trained the model to make connections and evaluated the model's ability to predict structure at various length scales.

The findings showed that Orca predicted DNA structures both small and large based on their sequences with high accuracy, including for sequences carrying mutations associated with various health conditions including a form of leukemia and limb malformations. Orca also enabled the researchers to generate new hypotheses about how DNA sequence controls its local and large-scale 3D structure.

Dr. Zhou said that he and his colleagues plan to use Sei and Orca, which are both publicly available on web servers and as open-source code, to further explore the role of [genetic mutations](#) in causing the molecular and physical manifestations of diseases—research that could eventually lead to new ways to treat these conditions.

More information: Kathleen M. Chen et al, A sequence-based global map of regulatory activity for deciphering human genetics, *Nature Genetics* (2022). [DOI: 10.1038/s41588-022-01102-2](https://doi.org/10.1038/s41588-022-01102-2)

Jian Zhou, Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale, *Nature Genetics* (2022).
[DOI: 10.1038/s41588-022-01065-4](https://doi.org/10.1038/s41588-022-01065-4)

Provided by UT Southwestern Medical Center

Citation: Artificial intelligence tools predict DNA's regulatory role and 3D structure (2022, August 8) retrieved 25 April 2024 from <https://phys.org/news/2022-08-artificial-intelligence-tools-dna-regulatory.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.