

Introducing EUGENE: An easy-to-use deep learning genomics software

November 16 2023



Credit: CC0 Public Domain

Deep learning—a form of artificial intelligence capable of improving itself with limited user input—has radically reshaped the landscape of biomedical research since its emergence in the early 2010s. It's been particularly impactful in genomics, a field of biology that examines how our DNA is organized into genes and how these genes are activated or deactivated in individual cells.

Despite this synergy, genomics researchers wanting to employ this technology are often challenged by the actual coding necessary to analyze vast pools of dense data.

Now, researchers at University of California San Diego have simplified this task for scientists by creating a new deep-learning platform that can be quickly and easily adapted to suit a wide variety of different genomics projects. The newly-developed software, named EUGENE, is detailed in a study published November 16, 2023 in [Nature Computational Science](#).

"Each of our cells has the same DNA, but the way that DNA is expressed changes what our cells look like and what they do," explained Hannah Carter, Ph.D., associate professor in the Department of Medicine at UC San Diego School of Medicine.

"Deep learning can provide valuable insights into the biological machinery driving this variety, but it can be challenging to implement for researchers without extensive computer science expertise. We wanted to create a platform that can help genomics researchers streamline their deep learning data analysis to make predictions from raw data."

Although genes coding for specific proteins make up only about 2% of our total genome, the remaining 98% of our DNA sequence, often referred to as "junk" DNA with no known function, plays a crucial role in determining when, where and how certain genes are activated. Unraveling the functions of these non-coding regions of the genome is a longstanding goal of genomics researchers, and deep learning has proven to be a powerful tool for achieving this goal—at least when researchers can figure out how to use it.

"A lot of existing platforms require many hours of coding and data

wrangling to use," said first author Adam Klie, a Ph.D. student in the Carter's lab. "Most projects require researchers to start from scratch, which takes expertise that not all labs interested in this stuff have access to."

Klie designed the new software to address the computing challenges he faced in his own work.

"With EUGENE, you give an algorithm a sequence of DNA and ask it to make predictions about anything you'd expect that DNA could predict, such as whether a particular DNA sequence is functional or whether it regulates a gene in a certain biological context," Klie said.

"This lets you explore properties of the DNA sequence and ask what would happen if I modified this piece here or moved this piece there. This is particularly relevant for researchers studying complex genetic disorders where many different sequences are implicated."

The researchers tested EUGENE by attempting to reproduce the results of three existing genomics studies that utilized several different types of sequencing data. Ordinarily, analyzing these different types of data would require mixing and matching multiple technology platforms. However, EUGENE proved adaptable enough to reproduce the findings of each of these studies.

"Being able to reproduce results is critically important in all [scientific research](#), but can be very difficult in genomics studies that use deep learning," said Carter.

"EUGENE is already showing a lot of promise in how adaptable it is to different types of DNA sequencing data and supporting a lot of different deep learning models. We hope it will evolve into a platform that can support collaborative tool development by the research community and

accelerate genomics research."

While the current version of EUGENE works on many types of genomic data, the researchers are working on expanding its scope to include an even wider variety of data types, such as single-cell sequencing data, which looks at the genomics of [individual cells](#) instead of in a whole tissue. They also plan to make EUGENE available to research groups around the world.

"One of the exciting things about this project is that the more people use the platform, the better we can make it over time, which will be essential as [deep learning](#) continues to evolve so rapidly," said Carter. "We hope that our platform will open many doors for researchers in this field and help them answer new questions about the complex molecular machinery that's inside all of us."

Co-authors of the study include: David Laub, James V. Talwar, Joe J. Solvason and Emma K. Farley at UC San Diego, Hayden Stites at Daniel Land High School and Tobias Jores at University of Washington.

More information: Predictive analyses of regulatory sequences with EUGENE, *Nature Computational Science* (2023). DOI: www.nature.com/articles/s43588-023-00544-w , www.nature.com/articles/s43588-023-00544-w

Provided by University of California - San Diego

Citation: Introducing EUGENE: An easy-to-use deep learning genomics software (2023, November 16) retrieved 29 April 2024 from <https://phys.org/news/2023-11-eugene-easy-to-use-deep-genomics-software.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.