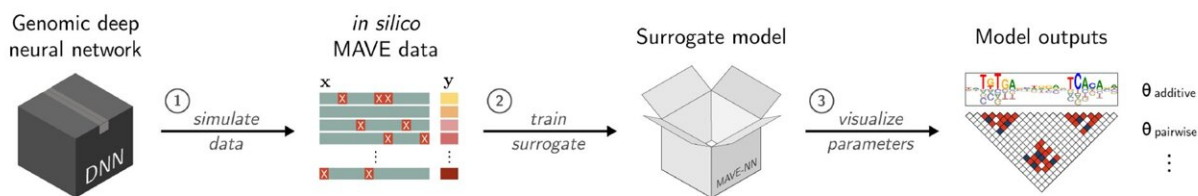# New computational tool helps interpret AI models in genomics

June 21 2024

An illustration outlining the SQUID computational pipeline. Credit: Koo and Kinney labs/ Cold Spring Harbor Laboratory

Artificial intelligence continues to squirm its way into many aspects of our lives. But what about biology, the study of life itself? AI can sift through hundreds of thousands of genome data points to identify potential new therapeutic targets. While these genomic insights may appear helpful, scientists aren't sure how today's AI models come to their conclusions in the first place. Now, a new system named SQUID arrives on the scene, armed to pry open AI's black box of murky internal logic.

SQUID, short for Surrogate Quantitative Interpretability for Deepnets, is a computational tool created by Cold Spring Harbor Laboratory (CSHL) scientists. It's designed to help interpret how AI models analyze the genome. Compared with other analysis tools, SQUID is more consistent, reduces background noise, and can lead to more accurate predictions

about the effects of genetic mutations.

How does it work so much better? The key, CSHL Assistant Professor Peter Koo says, lies in SQUID's specialized training.

"The tools that people use to try to understand these models have been largely coming from other fields like computer vision or natural language processing. While they can be useful, they're not optimal for genomics. What we did with SQUID was leverage decades of quantitative genetics knowledge to help us understand what these deep neural networks are learning," explains Koo.

SQUID works by first generating a library of over 100,000 variant DNA sequences. It then analyzes the library of mutations and their effects using a program called MAVE-NN (Multiplex Assays of Variant Effects Neural Network). This tool allows scientists to perform thousands of virtual experiments simultaneously. In effect, they can "fish out" the algorithms behind a given AI's most accurate predictions. Their computational "catch" could set the stage for experiments that are more grounded in reality.
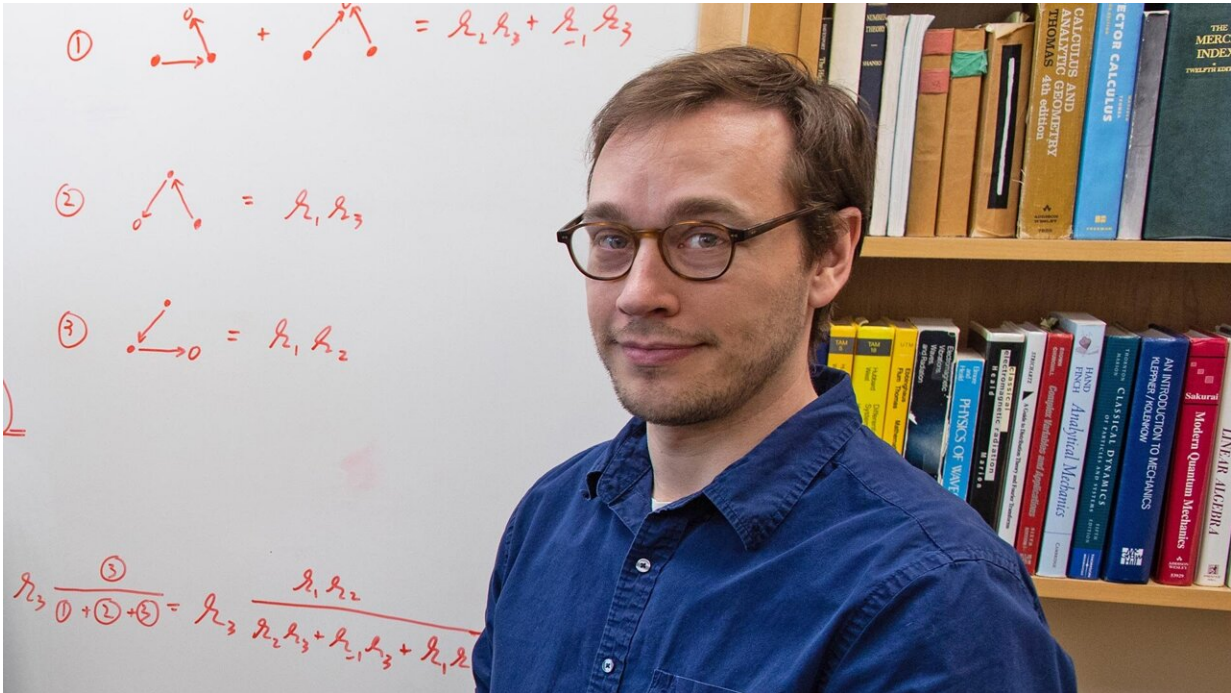
Evan E. Seitz, the lead author of this study, is a postdoc in the Kinney and Koo labs. Credit: Cold Spring Harbor Laboratory

"In silico [virtual] experiments are no replacement for actual laboratory experiments. Nevertheless, they can be very informative. They can help scientists form hypotheses for how a particular region of the genome works or how a mutation might have a clinically relevant effect," explains CSHL Associate Professor Justin Kinney, a co-author of the study.

There are tons of AI models in the sea. More enter the waters each day. Koo, Kinney, and colleagues hope that SQUID will help scientists grab hold of those that best meet their specialized needs.

Though mapped, the human genome remains an incredibly challenging terrain. SQUID could help biologists navigate the field more effectively,

bringing them closer to their findings' true medical implications.

The research is published in the journal *Nature Machine Intelligence*.

More information: Interpreting cis-regulatory mechanisms from genomic deep neural networks using surrogate models, *Nature Machine Intelligence*, DOI: 10.1038/s42256-024-00851-5

Provided by Cold Spring Harbor Laboratory

Citation: New computational tool helps interpret AI models in genomics (2024, June 21) retrieved 4 July 2024 from https://phys.org/news/2024-06-tool-ai-genomics.html