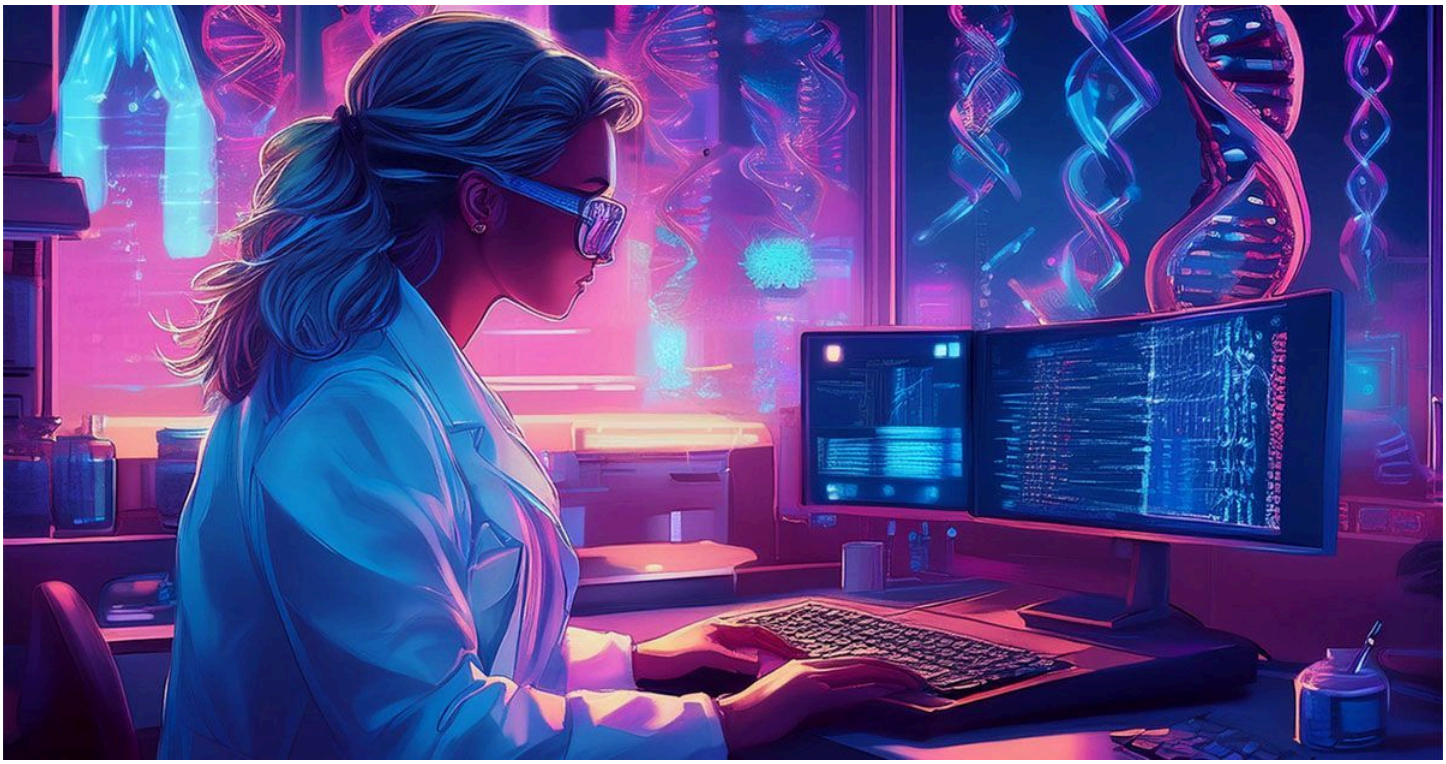
RESEARCH ALERTS

How Artificial Intelligence Could Automate Genomics Research



Functional genomics seeks to identify what genes do and how they interact. Researchers at UC San Diego have demonstrated that large language models such as GPT-4 could make functional genomics research significantly faster and less laborious than current, non-AI approaches. Image credit: UC San Diego Health Sciences (generated with Adobe Firefly)

Story by:

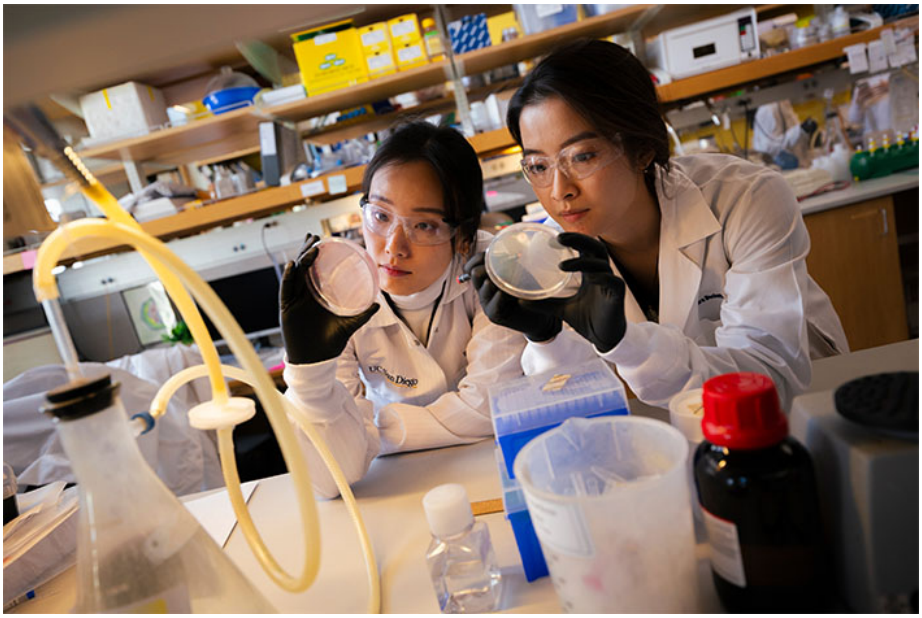
Miles Martin - milesmartin@ucsd.edu

Research Alerts

December 02, 2024

Researchers at University of California San Diego School of Medicine have demonstrated that large language models

(LLMs), such as GPT-4, could help automate functional genomics research, which seeks to determine what genes do and how they interact. The most frequently-used approach in functional genomics, called gene set enrichment, aims to determine the function of experimentally-identified gene sets by comparing them to existing genomics databases. However, more interesting and novel biology is often beyond the scope of established databases. Using artificial intelligence (AI) to analyze gene sets could save scientists many hours of intensive labor and bring science one step closer to automating one of the most widely used methods for understanding how genes work together to influence biology.



This image shows first study author Clara Hu (left) with fellow doctoral candidate Scarlett Qian (right). Both work in Trey Ideker's lab, which combines computational methods and wet-lab experiments to answer fundamental questions in biology and develop new approaches for precision medicine. Photo credit: Erik Jepsen/University Communications

Testing five different LLMs, the researchers found that GPT-4 was the most successful, achieving a 73% accuracy rate in identifying common functions of curated gene sets from a commonly used genomics database. When asked to analyze random gene sets, GPT-4 refused to provide a name in 87% of cases, demonstrating the potential of GPT-4 to analyze gene sets with minimal hallucination. GPT-4 was also capable of providing detailed narratives to support its naming process.

While further research is needed to fully explore the potential of LLMs in automating functional genomics, the study highlights the need for continued investment in the development of LLMs and their applications in genomics and precision medicine. To support this, the researchers created a web portal to help other researchers incorporate LLMs into their functional genomics workflows. More broadly, the findings also demonstrate the power of AI to revolutionize the scientific process by synthesizing complex information to generate new, testable hypotheses in a fraction of the time.

The study, published in Nature Methods, was led by Trey Ideker, Ph.D., a professor at UC San Diego School of Medicine and UC San Diego Jacobs School of Engineering, Dexter Pratt, Ph.D., a software architect in Ideker's group, and Clara Hu, a biomedical sciences doctoral candidate in Ideker's group. The study was funded, in part, by the National Institutes of Health.

Topics covered:

AI, Genomics, Genetics

Share This:

YOU MAY ALSO LIKE