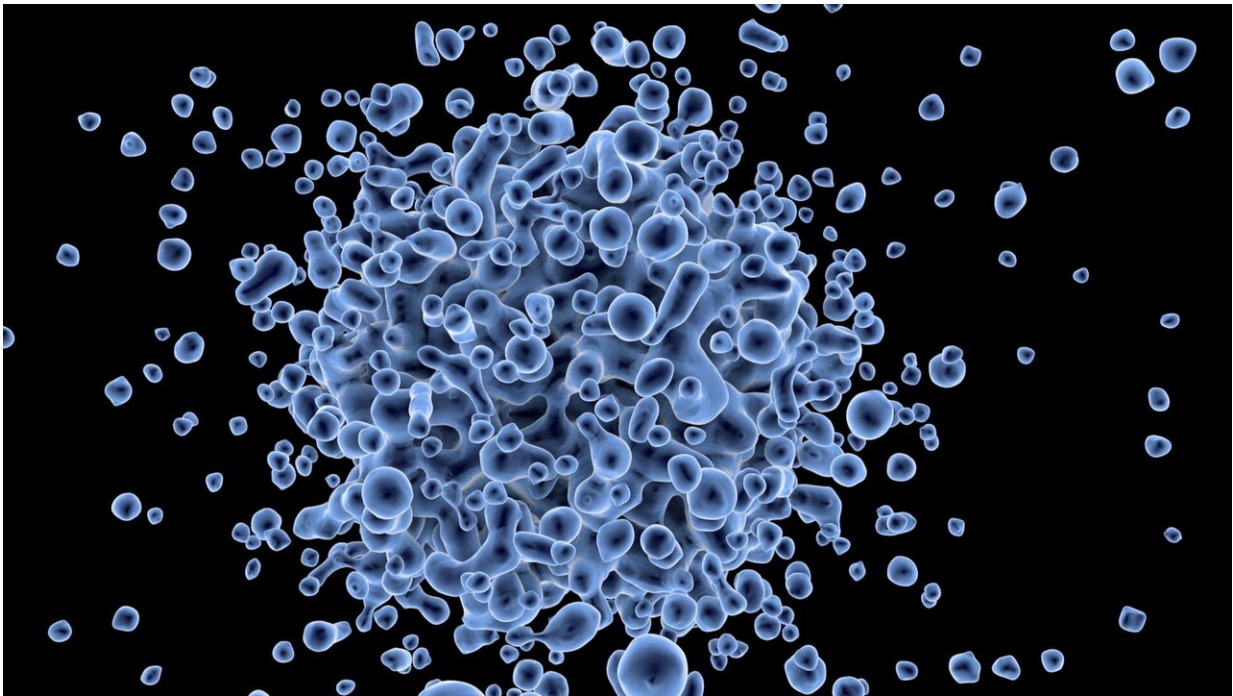


Machine learning tool pinpoints disease-related genes, functions

February 2 2021, by Stephanie Dutchen



Credit: CC0 Public Domain

The idea struck Robert Ietswaart, a research fellow in genetics at Harvard Medical School, while he was trying to determine how an experimental drug slowed the growth of lung cancer cells.

He saw that the drug triggered a cascade of molecular and genetic changes in the cells, but he needed to narrow down which of the many

activated genes were actually beating back the cancer rather than doing unrelated jobs. And, given that individual genes often do more than one thing—some even perform more than 100 different tasks—he needed to figure out which jobs the key genes were doing in these cells.

There were so many options that Ietswaart didn't know where to start.

Researchers in this position normally rely on experience, and sometimes software, to sift through the sludge of candidate genes and identify the gold nuggets that cause or contribute to a disease or amplify the effects of a drug. Then they research how those genes may be operating by poring over archives of scientific literature. This helps them build a better springboard from which to dive into experiments.

Ietswaart, however, who trained in [computational biology](#), had a better idea: create a tool that would search for and identify the most important genes and gene functions automatically. Existing tools could gauge which [biological processes](#) were relevant for an experiment but didn't rank individual genes or functions.

"I realized that many researchers struggle with the same questions," said Ietswaart. "So, I decided to build something that would be useful not only for me but for the broader scientific community."

The fruits of that labor—a collaboration between the labs of geneticist Stirling Churchman and systems pharmacologist Peter Sorger at HMS—were published Feb. 2 in *Genome Biology*.

The tool, dubbed GeneWalk, uses a combination of machine learning and automated literature analyses to indicate which genes and functions are most likely relevant to a researcher's project.

"It's the conundrum of so many biology labs these days: We have a list of

1,000 genes and we need to figure out what to do next," said Churchman, associate professor of genetics in the Blavatnik Institute at HMS and senior author of the paper. "We have a tool that helps you figure out not only which genes to follow up on but also what those [genes](#) are doing in the system you're studying."

By crunching through vast amounts of data and providing evidence-based guidance before users embark on costly, time-consuming experiments, GeneWalk promises to increase the speed and efficiency with which researchers can gain new insights into the genetics of disease and devise treatments, the authors say.

"It generates gene-specific mechanistic hypotheses that you can test," said Ietswaart, who is first author of the paper. "It should save people a lot of time and money."

Stand out from the crowd

Part of what distinguishes GeneWalk from other available tools is its use of the [INDRA](#) Database, which contains information synthesized from a vast automated literature search.

INDRA, short for Integrated Network and Dynamical Reasoning Assembler, accumulates findings from all published [biomedical literature](#), analyzes the texts to extract causal relationships and generate models and predictions, and transforms that wealth of information into the [searchable database](#).

INDRA and its database were developed by Benjamin Gyori and John Bachman, research associates in therapeutic science in the Laboratory of Systems Pharmacology at HMS; Sorger, the HMS Otto Kraye Professor of Systems Pharmacology; and colleagues.

"What Peter's group has done with INDRA is amazing and transformative," said Churchman. "It's been a special experience to use their remarkable piece of biomedical engineering in a new way that helps get relevant biomedical knowledge into more people's hands."

Leveraging the power of the INDRA Database, GeneWalk is the first tool that helps researchers home in on the most relevant gene functions for the biological context they're studying—in Ietswaart's case, lung cancer.

Most researchers are not aware that it's possible to automate gene function searches, eliminating the need to spend countless hours reading papers, the authors said.

"We're filling a gap that a lot of people didn't think was possible to fill," said Churchman.

"The value of machine learning in biomedical research is very much about making each step along the way a little easier," added Ietswaart.

The team wrote the GeneWalk software as open-source code and has made the tool [available for free](#). It is also designed to be easy for scientists to use. Churchman and Ietswaart have already heard from numerous labs at HMS and beyond who have jumped on GeneWalk for their own projects.

"I like that GeneWalk can be of broad general use," said Ietswaart. "It's not every day that you get to think of something that will be helpful for the scientific community."

More information: Robert Ietswaart et al. GeneWalk identifies relevant gene functions for a biological context using network representation learning, *Genome Biology* (2021). [DOI](#):

[10.1186/s13059-021-02264-8](https://doi.org/10.1186/s13059-021-02264-8)

Provided by Harvard Medical School

Citation: Machine learning tool pinpoints disease-related genes, functions (2021, February 2)
retrieved 15 May 2024 from <https://phys.org/news/2021-02-machine-tool-disease-related-genes-functions.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.