

Open-source machine learning tool connects drug targets with adverse reactions

June 18 2020



Credit: CC0 Public Domain

A multi-institutional group of researchers led by Harvard Medical School and the Novartis Institutes for BioMedical Research has created an open-source machine learning tool that identifies proteins associated with drug side effects.

The work, published June 18 in the *Lancet* journal *EBioMedicine*, offers a new method for developing safer medicines by identifying potential adverse reactions before [drug candidates](#) reach human clinical trials or enter the market as approved medicines.

The findings also offer insights into how the human body responds to [drug](#) compounds at the molecular level in both desired and unintended ways.

"Machine learning is not a silver bullet for [drug discovery](#), but I do believe it can accelerate many different aspects in the difficult and long process of developing new medicines," said paper co-first author Robert Ietswaart, research fellow in genetics in the lab of Stirling Churchman in the Blavatnik Institute at HMS. Churchman was not involved in the study.

"Although it cannot predict all possible adverse effects, we hope that our work will help researchers spot potential trouble early on and develop safer drugs in the future," Ietswaart said.

Drug side effects, technically known as adverse drug reactions, range from mild to fatal. They may occur either when taking a drug as prescribed or as a result of incorrect dosages, interaction of multiple medicines or off-label use (taking a drug for something other than what it was approved for). Adverse drug reactions are responsible for 2 million U.S. hospitalizations each year, according to the Department of Health and Human Services, and occur during 10 to 20 percent of hospitalizations, according to the Merck Manuals.

Researchers and health care providers have applied many tactics over the decades to avoid or at least minimize adverse drug reactions. But because a single drug often interacts with multiple proteins in the body—not always limited to the intended targets—it can be hard to

predict what, if any, side effects a medicine may generate. And if a drug does end up causing an adverse reaction, it can be hard to identify which of its protein targets could be responsible.

In the new study, researchers took one existing database of reported adverse drug reactions and another database of 184 proteins that specific drugs are known to often interact with. Then they constructed a [computer algorithm](#) to connect the dots.

"Learning" from the data, the algorithm unearthed 221 associations between individual proteins and specific adverse drug reactions. Some were known and some were new.

The associations indicated which proteins likely represent drug targets that contribute to particular side effects and which others may be innocent bystanders.

Based on what it has already "learned," and strengthened by any new data that researchers feed it, the program may help doctors and scientists predict whether a new drug candidate is likely to cause a certain side effect on its own or when combined with particular medicines. The algorithm can help with these predictions before a drug is tested in humans, based on lab experiments that reveal which proteins the drug interacts with.

The hope is to raise the likelihood that a drug candidate will prove safe for patients before and after it reaches the market.

"This could reduce the risks that study participants face during the first in-[human clinical trials](#) and minimize risks for patients if a drug gains FDA approval and enters clinical use," said Ietswaart.

Hack your side effects

The project was born at a quantitative science hackathon organized by Novartis Institutes for BioMedical Research (NIBR) in 2018.

Laszlo Urban, global head of preclinical secondary pharmacology at NIBR, presented on some of the problems his team faces when assessing the safety of new drug candidates. A group of Boston-area graduate students and postdocs at the hackathon jumped to apply their knowledge of data science and machine learning.

Most of the time, projects from the hackathon end as learning exercises, said Urban. On this rare occasion, however, a strong and lasting interaction of inspired scientists from different institutions resulted in a novel application published in a highly respected journal, he said.

Four members of the original hackathon group became co-first authors of the paper: Ietswaart at HMS, Seda Arat from The Jackson Laboratory, Amanda Chen of MIT and Saman Farahmand from the University of Massachusetts Boston. Arat is now at Pfizer. Another team member, Bumjun Kim of Northeastern University, is a co-author. Urban became senior author of the paper.

To tackle the problem, the team constructed its [machine learning](#) algorithm and applied it to two large data sets: one from Novartis with information about the proteins that each of 2,000 drugs interact with and one from the FDA with 600,000 physician reports of adverse drug reactions in patients.

The algorithm generated statistically robust information about how individual proteins contribute to documented adverse reactions, said Ietswaart.

"It suggests the physiological response to perturbing a particular protein—or the gene that makes it—at the molecular level," he said.

Many of the results supported previous observations, such as that binding to the protein hERG can cause cardiac arrhythmias. Findings like this strengthened the researchers' confidence that the algorithm was performing well.

Other results, however, were unexpected.

For instance, the algorithm suggested that the [protein](#) PDE3 is associated with over 40 adverse drug reactions. Doctors and researchers have known for years that PDE3 inhibitors—common anti-clotting treatments for acute heart failure, stroke prevention and a heart attack complication known as cardiogenic shock—can cause arrhythmias, low platelet counts and elevated levels of enzymes called transaminases, a possible indicator of liver damage. But it wasn't known that targeting PDE3 might raise the risk of so many other side effects, including some related to the muscles, bones, connective tissue, kidneys, urinary tract and ear.

Into the future

The algorithm also offered predictions on the likelihood that a particular drug would cause a certain adverse reaction.

How accurate were those new predictions? To find out, the researchers fed their algorithm updated information. Until then, the program had learned from [adverse drug reactions](#) reported through 2014. The team added reports gathered from 2014 through 2019, some of which revealed side effects that hadn't been observed before from particular drugs.

Sure enough, many of the algorithm's previously unproven predictions matched the recent real-world reports.

"What seemed like false-positive predictions proved not to be false at all

when the new reports became available," said Ietswaart.

To make extra certain that the algorithm is reliable, the team compared its results to drug labels, conducted text mining of the scientific literature and used other validation techniques.

Although the researchers strengthened the model as much as they could, it still assesses less than 1 percent of the 20,000 genes in the human genome.

"Our work is by no means a complete understanding of adverse drug events because many other genes and proteins might contribute for which no assay is available or no drugs have been tested," said Ietswaart.

Scientists can use, improve and build upon the model, which is posted for free online at <https://github.com/samanfrm/ADRtarget>.

"This work has been a collaborative 'open science' spirit and team effort," said Ietswaart and Urban.

More information: *EBioMedicine* (2020). [DOI: 10.1016/j.ebiom.2020.102837](https://doi.org/10.1016/j.ebiom.2020.102837)

Provided by Harvard Medical School

Citation: Open-source machine learning tool connects drug targets with adverse reactions (2020, June 18) retrieved 4 May 2024 from <https://medicalxpress.com/news/2020-06-open-source-machine-tool-drug-adverse.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.