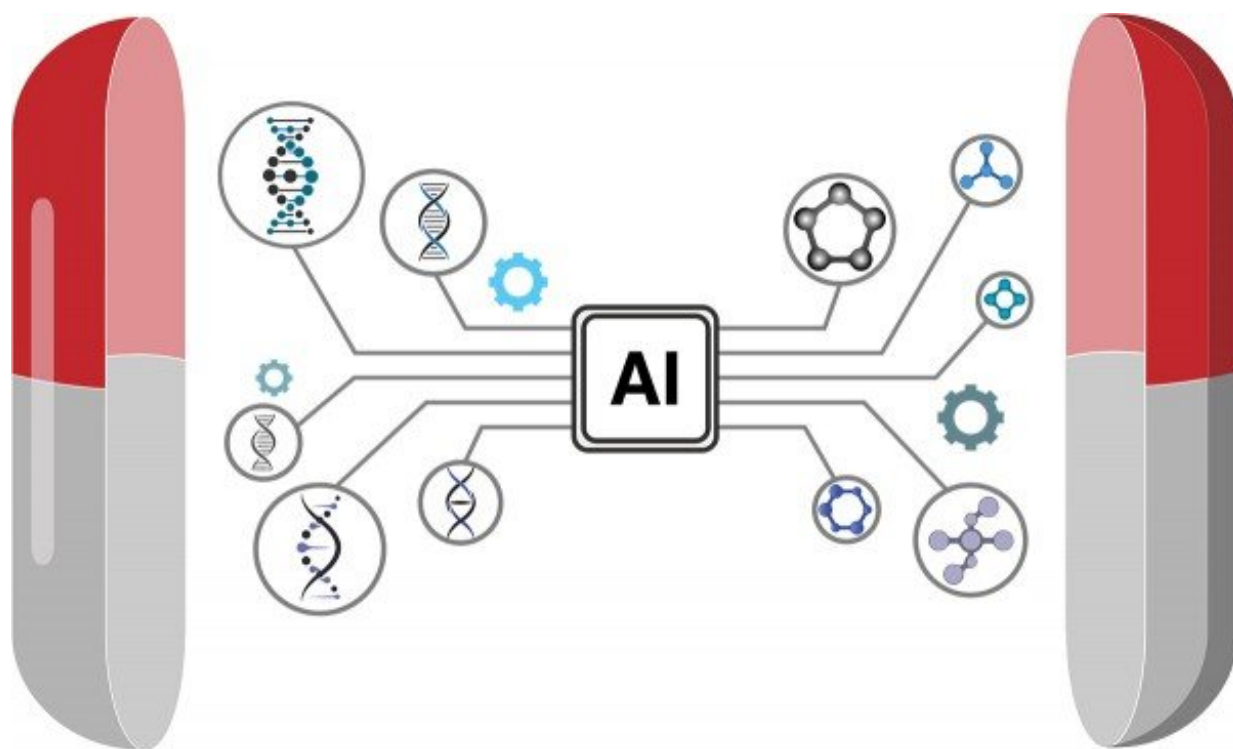# Team develops machine learning platform that mines nature for new drugs

June 2 2021, by Aaron Aupperlee

Researchers in the Computational Biology Department have developed a new process that could reinvigorate the search for natural product drugs to treat cancers, viral infections and other ailments. Credit: Carnegie Mellon University

Researchers from Carnegie Mellon University's Computational Biology Department in the School of Computer Science have developed a new process that could reinvigorate the search for natural product drugs to

treat cancers, viral infections and other ailments.

The machine learning algorithms developed by the Metabolomics and Metagenomics Lab match the signals of a microbe's metabolites with its genomic signals and identify which likely correspond to a natural product. Knowing that, researchers are better equipped to isolate the natural product to begin developing it for a possible drug.

"Natural products are still one of the most successful paths for drug discovery," said Bahar Behsaz, a project scientist in the lab and lead author of a paper about the process. "And we think we're able to take it further with an algorithm like ours. Our computational model is orders of magnitude faster and more sensitive."

In a single study, the team was able to scan the metabolomics and genomic data for about 200 strains of microbes. The algorithm not only identified the hundreds of natural product drugs the researchers expected to find, but it also discovered four novel natural products that appear promising for future drug development. The team's work was published recently in *Nature Communications*.

The paper, "Integrating Genomics and Metabolomics for Scalable Non-Ribosomal Peptide Discovery," outlines the team's development of NRPminer, an artificial intelligence tool to aid in discovering non-ribosomal peptides (NRPs). NRPs are an important type of natural product and are used to make many antibiotics, anticancer drugs and other clinically used medications. They are, however, difficult to detect and even more difficult to identify as potentially useful.

"What is unique about our approach is that our technology is very sensitive. It can detect molecules with nanograms of abundance," said Hosein Mohimani, an assistant professor and head of the lab. "We can discover things that are hidden under the grass."

Most of the antibiotic, antifungal and many antitumor medications discovered and widely used have come from natural products.

Penicillin is among the most used and well-known drugs derived from natural products. It was, in part, discovered by luck, as are many of the drugs made from natural products. But replicating that luck is difficult in the laboratory and at scale. Trying to uncover natural products is also time and labor intensive, often taking years and millions of dollars. Major pharmaceutical companies have mostly abandoned the search for new natural products in the past decades.

By applying machine learning algorithms to the study of genomics, however, researchers have created new opportunities to identify and isolate natural products that could be beneficial.

"Our hope is that we can push this forward and discover other natural drug candidates and then develop those into a phase that would be attractive to pharmaceutical companies," Mohimani said. "Bahar Behsaz and I are expanding our discovery methods to different classes of natural products at a scale suitable for commercialization."

The team is already investigating the four new natural products discovered during their study. The products are being analyzed by a team led by Helga Bode, head of the Institute for Molecular Bioscience at Goethe University in Germany, and two have been found to have potential antimalarial properties.

More information: Bahar Behsaz et al, Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery, *Nature Communications* (2021). DOI: 10.1038/s41467-021-23502-4