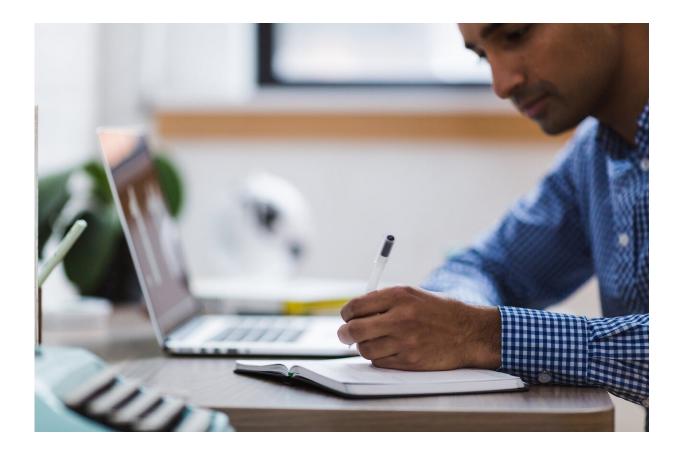# Next-gen bioinformatics tool enables big data analysis without programming expertise

September 24 2020

Credit: CC0 Public Domain

A new data analysis tool developed by researchers at The University of Texas MD Anderson Cancer Center incorporates a user-friendly, natural-language interface to allow biomedical researchers without specialized

expertise in bioinformatics or programming languages to conduct intuitive analysis of large datasets.

The open-access, artificial intelligence (AI)-driven program, called DrBioRight, was created to lower barriers for all researchers to make full use of the increasingly large amounts of data generated in modern research methods. A report of this platform was published today in *Cancer Cell.*

"We felt that we could improve the current model for conducting routine bioinformatics analysis and greatly speed up turnaround time by creating a tool that any researcher could use," said Han Liang, Ph.D., professor of Bioinformatics and Computational Biology. "Our long-term goal for DrBioRight is to be an intelligent collaborator for every researcher."

High-throughput technologies used in modern biomedical research generate large, complex datasets that provide comprehensive information about patients, animal models or cell lines being studied. These may include, for example, studying the whole of genetic information (genomics), gene expression (transcriptomics), or protein expression (proteomics).

Because these "omics" datasets are so complex, it can be challenging to answer specific biological questions without specialized analytical approaches, explained Liang. These analyses are usually done with using a computer script written in a variety of programming languages, which requires some understanding of both programming and bioinformatics.

Bioinformaticians can help to navigate and process these complex datasets, but the work can be time consuming. Therefore, the research team developed DrBioRight to enable researchers to more easily conduct routine analyses of their own data through a user-friendly chat interface with natural-language interactions.

The natural language-oriented program allows users to ask questions of the program as if they were speaking naturally rather than in complex programming languages, explained Liang.

DrBioRight is freely available to academic researchers. Initially, the program has a number of modules ready-built to handle the most common types of bioinformatics questions and includes some of most frequently used public cancer datasets available, such as The Cancer Genome Atlas and Cancer Cell Line Encyclopedia.

As a confirmation of the approach, the researchers replicated the analysis of a classic cancer genomics paper using DrBioRight and found it to accurately reproduce the previously published results.

Because the program is driven by AI, it also has the ability to learn from each inquiry and improve analysis, becoming a more useful tool over time. Going forward, the researchers hope to improve DrBioRight to enable users to analyze their own datasets as well as allow open development for new modules.

"As we work to improve the program, we also want to enable other bioinformaticians to contribute their algorithms and teach DrBioRight," said Liang. "Involvement from the entire research community will help to create a tool that is useful in answering complex research questions more efficiently."