

Knowledge Engine is ready to accelerate genomic research

February 3 2020, by Claudia Lutz



Credit: CC0 Public Domain

Five years ago, a team of computer scientists, biomedical researchers, and bioinformaticians set out to bring the power of collective knowledge to genomic research. Their new publication in *PLOS Biology* shares the

culmination of that effort, an analytical platform that guides researchers through the process of interpreting complex genomic datasets.

The group was awarded funding by the National Institutes of Health to form a Big Data to Knowledge Center of Excellence. The center, led by Professor of Computer Science and Willett Faculty Scholar Saurabh Sinha at the Carl R. Woese Institute for Genomic Biology (IGB) at the University of Illinois at Urbana-Champaign and including numerous collaborators at Illinois and Mayo Clinic, created a first-of-its-kind analytical platform, the Knowledge Engine for Genomics (KnowEnG). Charles Blatti and Amin Emad, who were both postdoctoral researchers within the center, are co-first authors of the new publication.

"It is exhilarating to see the efforts of such an amazing group of talented and dedicated people—researchers, [software engineers](#), user experience designers, project managers, faculty, postdocs, grad students, undergrads and even high school students—over a period of years, culminate in a single product that we can all be proud of and that can hopefully help genomics researchers world-wide," Sinha said.

To understand KnowEnG's potential to impact genomic research, it's important to know that the initial outcome of many genomic studies is a set of genes of interest: genes that have different activity levels in two different experimental conditions, that carry mutations distinguishing healthy cells from tumor cells, or that exhibit sequence variations in individuals with different health conditions.

Biomedical researchers must find ways to translate a list of not obviously related genes into a comprehensive interpretation: Does the disease state affect the metabolic rate of the affected tissue? Does the experimental therapy seem likely to slow the rate of tumor cell division? A common approach to this challenge is to relate an experimental dataset with existing knowledge of the biological importance of different genes and

their relationships with one another. Like a curious internet user with a set of search terms, the researcher hopes to leverage the totality of what is already known.

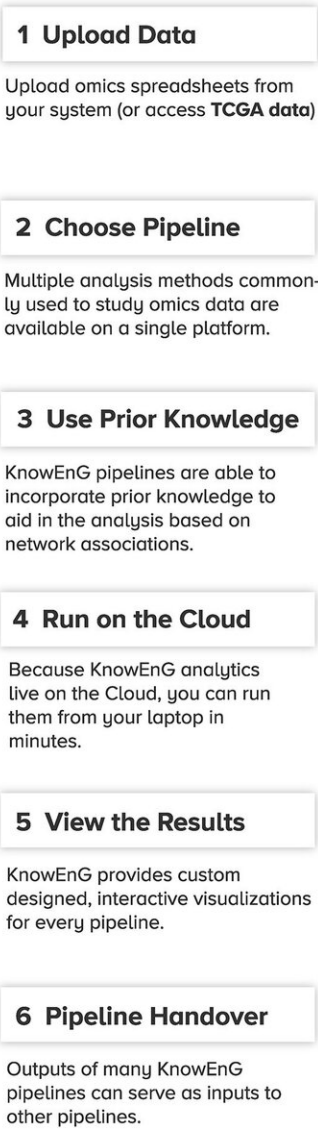
Unlike an internet user, however, an individual conducting [genomic research](#) in recent years hasn't had the flexibility like that offered by a search engine like Google to seamlessly pull together myriad sources of information; nor could they easily apply that information in many different types of analyses. Instead, each analysis had to be done piecemeal, jumping from one analytical tool to another, each offering a limited interpretation. This is an obstacle that KnowEnG removes.

"A lot of times you start with one analysis and then you want to follow up with more analyses," Emad, who is now an assistant professor of electrical and computer engineering at McGill University, said. "One thing about KnowEnG that is very useful is that you can pipeline these different analyses one after the other. You may run one [analytical] pipeline in KnowEnG, get the results, and automatically generate data in the format that can be plugged into the next pipeline . . . So you can set up different types of analyses one after the other."

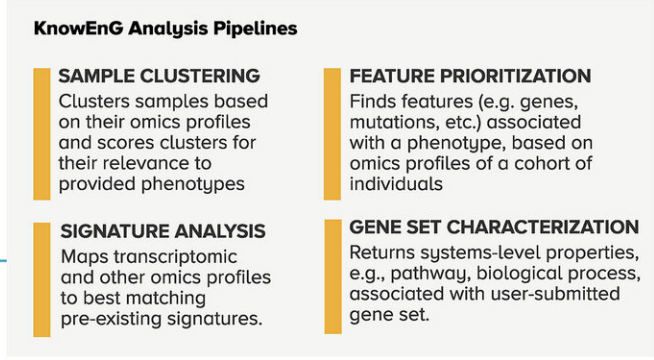
KnowEnG is also uniquely able to draw upon and synthesize diverse sources of existing genomic information, combining them into a vast "knowledge network" that can continue to expand over time as research continues and new forms of data emerge from new genomic technologies.

To highlight the functionality the center has achieved, a team led by Blatti and Emad used previously published data as case studies, re-analyzing the results within the KnowEnG platform and sharing the novel insights revealed.

A



B



C



(A) Typical workflow steps for KnowEnG: genomic analysis on the cloud. (B) Analytical functionalities are organized as “pipelines” for common tasks such as clustering, gene prioritization, gene set analysis, and signature analysis. Each pipeline offers various options to customize the analysis, including use of prior knowledge. (C) The KnowEnG Knowledge Network represents prior knowledge that may be used during analysis. Nodes in the network represent genes and biological properties, and edges represent either annotations of gene properties or gene-gene relationships. Network contents are summarized on the left with logos of data sources shown on the right. KnowEnG, Knowledge Engine for Genomics; TCGA, the Cancer Genome Atlas. Credit: Carl R. Woese Institute for

Genomic Biology, University of Illinois at Urbana-Champaign

"I worked very closely with Amin [Emad] to try to design a study that mimics previous studies that don't use prior knowledge, and then try to show where we can go beyond those studies with a knowledge-guided analysis," Blatti said. Blatti is now a research scientist at Illinois' National Center for Supercomputing Applications (NCSA).

Genomics is an incredibly broad field, and KnowEnG's capabilities span that breadth. The platform enables users to upload their data and customize step-by-step analyses for a wide variety of human genomic data forms, or for genomic data from any of 19 model organisms.

"Even though the paper is a paper about human cancer studies . . . we've also tried to target the model organisms that are studied in Illinois," Blatti said. "It's not just a cancer only platform—it's a wider platform."

Experts in biological data management have recommended that datasets and tools should be findable, accessible, interoperable, and usable (FAIR) within the research community. Center members ensured that KnowEnG would address those goals, making it be freely available through a web portal, as well as facilitating a variety of other modes of access. They also worked directly with test groups of users in biomedical research via the partnership with Mayo Clinic.

"We try to thoroughly understand the questions the user is asking of the data," said Colleen Bushell, a coauthor and Associate Director for Healthcare Innovation at NCSA. "Our approach is to then design a way to display answers to those questions clearly, and anticipate the next set of questions. When we simplify views of data, it's really just to answer some of those questions concisely, but then as they dive deeper into

trying to understand the data, we provide more and more detail, more and more explanation." Bushell provided oversight of KnowEng implementation and guided the center's visualization team, led by co-authors Lisa Gatzke and Matthew Berry, in developing innovative ways to represent complex data and analytical processes within the platform, and to create user-experiences that makes it easy for biologist to manage their data, set up data science experiments, and execute them in a cloud environment.

The work process relied heavily on feedback from [biomedical researchers](#) at Mayo and Illinois at every stage of development.

"For me, the postdoc at the KnowEnG Center and what it involved was a unique opportunity and a unique environment," Emad said. "Talking to people at Mayo Clinic and other researchers that were quite knowledgeable in the biomedical domain allowed me to learn a lot. So every task for me was a learning opportunity."

Although the NIH funding for the center has ended, the Cancer Center at Illinois is providing funding to enable access to the KnowEnG this year. Development will continue through the efforts of the NCSA Healthcare Innovation program office. NCSA works in part to support the longevity of software developed at Illinois; the structure of KnowEnG was designed to accommodate the addition of new forms of data, new analytical processes, and new visualization strategies over time.

"NCSA is committed to really continue this platform," Bushell said. "This falls into NCSA's mission to give software a life beyond funding... we're focusing on tools related to healthcare and data analysis, and we collaborate closely with IGB researchers. We want people to know that these tools will be around."

More information: Charles Blatti et al, Knowledge-guided analysis of

"omics" data using the KnowEnG cloud platform, *PLOS Biology* (2020).
[DOI: 10.1371/journal.pbio.3000583](https://doi.org/10.1371/journal.pbio.3000583)

Provided by University of Illinois at Urbana-Champaign

Citation: Knowledge Engine is ready to accelerate genomic research (2020, February 3) retrieved 22 March 2023 from <https://phys.org/news/2020-02-knowledge-ready-genomic.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.