

GenomeSpace 'recipes' help biologists interpret genomic data

January 20 2016

Many biomedical researchers are striving to make sense of the flood of data that has followed recent advances in genomic sequencing technologies. In particular, researchers are often limited by the challenge of getting multiple bioinformatics tools to "talk" to one another. To help address this need, researchers at University of California, San Diego School of Medicine, in collaboration with labs at the Broad Institute of MIT and Harvard, Stanford University, Weizmann Institute and Pennsylvania State University, developed GenomeSpace, a cloud-based, biologist-friendly platform that connects more than 20 bioinformatics software packages and resources for genomic data analysis.

The team is now developing and crowdsourcing "recipes"—step-by-step workflows—to better enable non-programming researchers to interpret their genomic data. The work is described in a paper published January 18, 2016 in *Nature Methods*.

"Now that new sequencing technologies can produce significantly greater amounts of data than they could a decade ago, the methods required to analyze that data must be correspondingly more powerful," said Jill Mesirov, PhD, associate vice chancellor for computational health sciences and professor of medicine at UC San Diego School of Medicine and Moores Cancer Center. "The problem is that only a small portion of the biomedical research community has the expertise to know the right method, or combination of methods, to solve their research questions and the best way to apply those methods to their data."

Before GenomeSpace, it was extraordinarily difficult for researchers, especially without programming skills, to get many of the available analysis tools to work together. Users needed to know how to write short computer programs in order to transform and transfer data between platforms. GenomeSpace now performs this service seamlessly with a user-friendly interface, connecting popular genomic data analysis tools such as Cytoscape, Galaxy, GenePattern and the Integrative Genomics Viewer (IGV). Several of these tools are themselves "tool aggregators," so in linking them, GenomeSpace provides access to hundreds of bioinformatics analyses.

What's more, GenomeSpace doesn't just leave users on their own to determine the best tools for their particular research questions. The site also provides "recipes"—easy-to-follow example workflows that clearly demonstrate the sequence of tools researchers should use to get the information they are looking to extract from their raw data. GenomeSpace currently provides 13 recipes. The platform's developers are now inviting the user community to contribute their own additional recipes.

"No individual lab can possibly develop all the right useful recipes—crowdsourcing will help make GenomeSpace even more useful to non-programming researchers," said Michael Reich at UC San Diego School of Medicine, who leads the GenomeSpace development team.

Here's how an example GenomeSpace recipe works: A researcher wonders if there is a specific set of genes that leukemia stem cells express differently than normal white blood cell precursors. She also wants to better understand the biological mechanism underlying those differentially expressed genes but doesn't know where to start. With GenomeSpace, the researcher can simply upload the gene expression data and other information about the two cell types (the "ingredients") and follow a GenomeSpace recipe, designed specifically for these types

of research questions. In this case, the recipe tells the researcher how to run the data ingredients through two tools available in GenomeSpace: 1) GenePattern, which finds a list of the 50 genes that differ the most between the two cell types and 2) Cytoscape, which identifies how proteins associated with these genes interact in networks, thus providing clues to the roles that tumor-specific or normal cell-specific genes play in the body.

This type of information provided by GenomeSpace could help the researcher better understand how leukemia develops and help identify possible targets for new therapeutics, said Reich.

"Our recipe resource was modeled on Tom Maniatis' classic, *Molecular Cloning: A Laboratory Manual*. We hope, with a combination of our own development and crowdsourcing, to grow the resource and increase its breadth," Mesirov said. "It's our long-term goal to convert these descriptive workflows into more dynamic, interactive interfaces making them even easier to follow."

More information: — genomespace.org/

— Kun Qu et al. Integrative genomic analysis by interoperation of bioinformatics tools in GenomeSpace, *Nature Methods* (2016). [DOI: 10.1038/nmeth.3732](https://doi.org/10.1038/nmeth.3732)

Provided by University of California - San Diego

Citation: GenomeSpace 'recipes' help biologists interpret genomic data (2016, January 20) retrieved 24 April 2024 from <https://phys.org/news/2016-01-genomespace-recipes-biologists-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.