

# Public resource boosts drug discovery and offers insights into protein function

November 30 2017

---



Credit: CC0 Public Domain

Researchers at the Broad Institute of MIT and Harvard have taken the Connectivity Map—a widely used resource of tools and data—to new heights with a massively scaled-up version. For this new platform, the researchers have also improved its accessibility for the scientific community, enabling studies of small molecule and gene function and informing clinical trials.

The Connectivity Map, or CMap, is a functional reference dataset of the human genome that connects [genes](#), drugs, and disease states through common patterns of gene activity, known as [gene-expression](#) signatures. The original version of CMap included only few hundred [gene expression](#) profiles in a few cell lines, produced using costly DNA microarrays. Broad scientists have now developed a low-cost, high-throughput method for gene expression, whereby the expression of a subset of genes is measured, and the expression of remaining non-measured genes are computationally inferred. This method, called L1000, allowed the Broad team to expand the existing CMap by more than 1,000-fold, making it a much more comprehensive and useful resource for the [scientific community](#).

The work is described in the November 30 issue of *Cell*.

Containing more than 1.3 million profiles of gene expression and developed through the NIH LINCS Consortium, this "next generation" CMap can be explored by researchers around the globe through a new cloud-based analysis environment, allowing them to functionally annotate genetic variants of disease genes, discover the mechanism of action of uncharacterized small molecules, and generate new therapeutic hypotheses.

"The expanded Connectivity Map is an example of new directions in genomics research. We're excited to see that our data and tools are already being used by scientists throughout academia and industry, to support basic science and [drug](#) discovery," said Todd Golub, senior author of the study and chief scientific officer of the Broad, where he is also director of the institute's Cancer Program. "It's likely that the data and tools will be used in ways that we haven't even imagined, and we hope that users will help us improve the tools and make CMap even more useful as we continue to expand the resource."

The utility of the pilot version of CMap was limited by its small size. CMap scientists knew that to build a truly comprehensive resource that could yield mechanistic and circuit-level biological insight, they would need to greatly expand the compendium with many chemical and genetic perturbations in diverse cell types.

Because doing so using microarrays or even RNA sequencing would be too costly, the CMap team developed a new profiling method known as "L1000." Instead of profiling the expression of every protein-coding gene in the genome, the method generates a genome-scale look at expression by measuring the activity of 1,000 "landmark" genes and using those measurements to infer the activity of most non-measured genes. The researchers analyzed existing data on [gene expression patterns](#) to choose those landmark genes that can serve as accurate representatives for the entire transcriptome.

This approach enabled the team to dramatically increase the scale of the experiment so that CMap now includes more than 1 million [gene expression profiles](#) from multiple cell lines treated with chemical or genetic perturbations. Compared to the 164 drugs profiled in the CMap pilot, the new dataset includes expression profiles from cells treated with 42,080 perturbagens, including small molecule drugs, tool compounds, and unoptimized compounds of previously unknown mechanisms of action.

To demonstrate the resource's utility, the team successfully showed that CMap can help predict how a small molecule or drug works, which can accelerate drug discovery efforts. If the expression profile of cells perturbed by a small molecule matches the expression signature from cells perturbed with compounds of known function, it suggests that the small molecule may work through the same cellular pathway and gives scientists an experimental head start when exploring the function of unstudied compounds or potential therapeutics.

The team also showed that CMap can help researchers discover compounds with specific, desired activities. In one instance, they used it to discover a compound that inhibits Casein Kinase 1 alpha, a protein involved in certain leukemias and that also confers resistance to a class of lung cancer drugs called EGFR inhibitors.

This underscores the power of the expanded Connectivity Map as a valuable starting point for drug discovery.

In a test of CMap's potential to inform clinical research, the researchers analyzed tumor samples obtained before and after treatment from cancer drug trials. The results showed changes in the tumor cells' patterns of gene activity due to cancer therapy, and comparison to CMap perturbagens suggested the involvement of known drug resistance pathways.

The Connectivity Map is constantly being curated with new data generated by the Broad team. The new version contains expression signatures from compounds that have previously been studied, but also those that haven't yet been characterized.

All the data and tools are now available in a cloud-based analysis environment developed by Broad researchers and known as [CLUE](#), which the CMap team encourages users to access and explore. The CMap team is planning to expand the resource to include more cell types, more perturbations, and more types of data, including proteomic and cellular imaging data.

The next generation CMap was made possible through close collaborations between the CMap team, other members of the NIH LINCS Consortium, and several other groups at the Broad Institute, including the Center for the Development of Therapeutics (CDoT), the Genetic Perturbation Platform, the PRISM Team, and the Proteomics

Platform.

"This effort was only possible with the combined expertise of many Broad programs and platforms, requiring an incredible amount of teamwork," said Aravind Subramanian, a co-first author of the paper along with Broad researchers Steven Corsello and Rajiv Narayan. "Our aspiration is that CMap becomes a routine part of [drug discovery](#), providing helpful clues as targets and molecules pass through the various stages of therapeutic development. We're pleased to be able to share the results of our efforts with the scientific community. Importantly, we're not done yet—we invite drug hunters from academia and industry to use the resource and reach out to us with your feedback."

Provided by Broad Institute of MIT and Harvard

Citation: Public resource boosts drug discovery and offers insights into protein function (2017, November 30) retrieved 7 August 2024 from <https://phys.org/news/2017-11-resource-boosts-drug-discovery-insights.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.