Different Strokes for Different Folks: Presenting Two New Datasets of Phospho- and Epigenetic Signatures

Lev Litichevskiy[1,2,3], Ryan Peckner[1,2], Aravind Subramanian[1,3], and Jake Jaffe[1,2] on behalf of the Broad PCCSE and Connectivity Map teams

[1] BROAD INSTITUTE [2] PCCSE [3] ConnectivityMap

ABSTRACT

High-throughput genetic profiling of cell lines subjected to perturbation enables the discovery of functional connections among diseases, genes, and drug action. Other readouts of cellular response to perturbation could also be immensely valuable, but it has previously been economically unfeasible to produce such resources. Here, we present two datasets of proteomic readouts created using the GCP and P100 assays. We demonstrate high replicate consistency in both assays, and we highlight examples in which the two assays complement each other.

OVERVIEW OF DATASETS

GCP

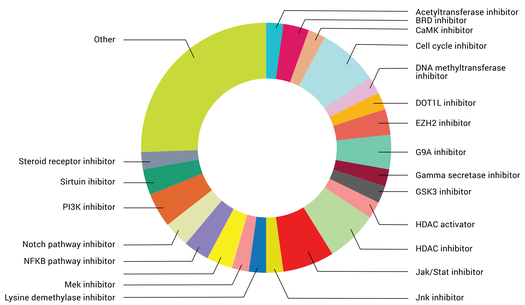
GCP is a targeted proteomic assay that uses ~60 probes to monitor combinations of post-translational modifications on histones

P100

P100 is a targeted proteomic assay that uses 96 phosphopeptide probes that are commonly observed and modulated in diverse cell types

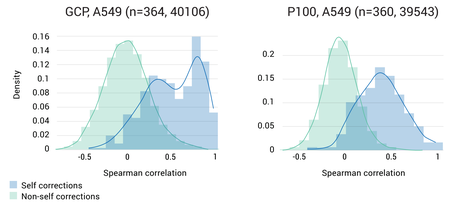
We profiled 90 compounds in 6 cell lines in triplicate. DMSO (negative control) and staurosporine and vorinostat (positive controls) were included on each plate of data. Descriptions of cell lines and an overview of the mechanisms of action (MOAs) that our compounds cover are below.

CELL LINE	DESCRIPTION
A375	Skin cancer (malignant melanoma)
A549	Lung cancer (non-small cell carcinoma)
MCF7	Breast cancer (carcinoma)
NPC	Neural progenitor cell (ESC-derived)
PC3	Prostate cancer (carcinoma)
YAPC	Pancreatic cancer (carcinoma)

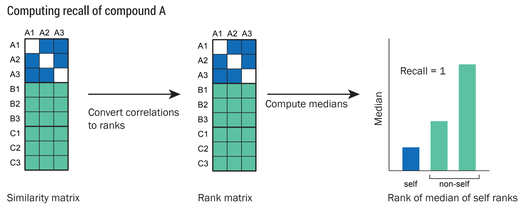


BOTH ASSAYS SHOW HIGH REPLICATE CONSISTENCY

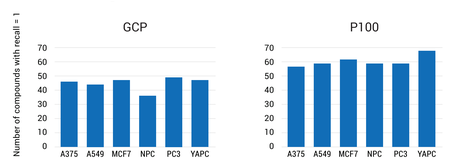
An expectation of a robust biological dataset is replicate consistency; in other words, we expect the replicates of a given compound to correlate better with themselves (self correlation) than with replicates of other compounds (non-self correlation). We demonstrate that self correlations are considerably higher than non-self correlations in both assays and all cell lines. Two examples of A549 in GCP and P100 are shown below.



Next, we sought to evaluate the consistency of individual compounds. We computed a metric called replicate recall for each compound in each assay.

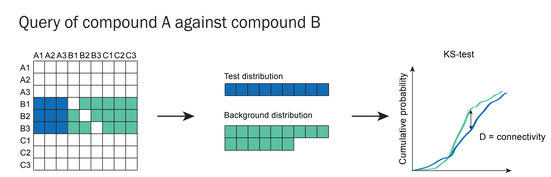


At least 30 (GCP) and 50 (P100) compounds out of 90 have replicate recall of 1. Aggregating across cell lines, we found that 60% (GCP) and 78% (P100) of compounds exhibit replicate recall of 1.

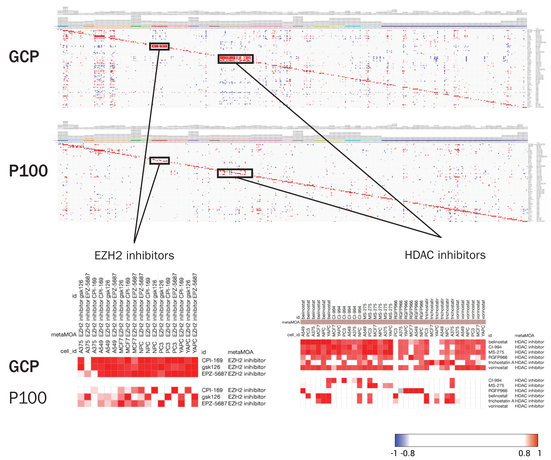


EXPLORING CONNECTIONS

Next, we wanted to explore connections between compounds and MOAs in our datasets. We compute connectivity between two compounds using the Kolgomorov-Smirnov (KS) two-sample test.



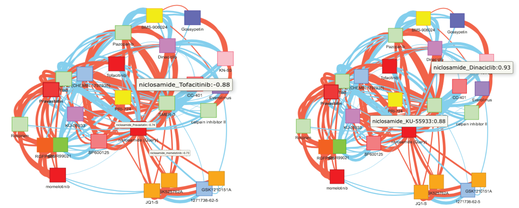
We analyzed all within-cell connectivities in our datasets using Morpheus (a heatmap visualization tool). Our initial observations were that 1) there is high self-connectivity (diagonal line across the heatmap) and 2) we can recover expected connections of MOAs.



SPECIFIC QUERIES YIELD NEW INSIGHT INTO MOAS

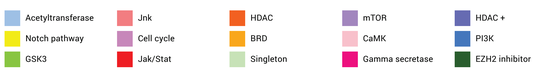
Vignette 1: Niclosamide

Although niclosamide is typically categorized as a Stat inhibitor, querying niclosamide in P100 shows that it has a strong negative connection to other Jak/Stat inhibitors. Instead, it connects strongly with cyclin-dependent kinase inhibitors, suggesting previously unrecognized activity.



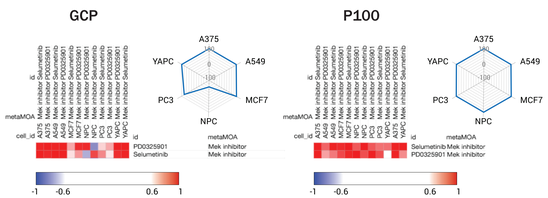
Negative connections to niclosamide

Positive connections to niclosamide



Vignette 2: Mek inhibitors

While all cell lines demonstrate consistent phosphosignaling activity (P100) of Mek inhibitors, NPC cells have distinct epigenetic events (GCP) that differentiate two closely related class members.



EXTENSION TO NOVEL MODELS USING QUERY FRAMEWORK

In addition to analyzing connections within our resource, we have developed our computational framework to allow for querying with external data.

Vascular endothelial cells (aka HUVEC) provided by Iris Jaffe (Tufts)

LINCS PCCSE resource

Connectivity results (analysis pending)

DATA AND COMPUTATIONAL TOOLS AVAILABLE ONLINE

We created a computational pipeline that enables processing of data all the way from raw mass spectrometry readings to connectivities between compounds. Our framework was built to be adaptable for integration with other LINCS datasets.

Panorama Data: https://bit.ly/PCCSEdata

Parsing data: https://github.com/cmap/l1ktools
Analyzing data: https://github.com/cmap/proteomics-signature-pipeline

Morpheus (for visualizing heatmaps): https://www.broadinstitute.org/cancer/software/morpheus/

Force-directed networks: https://rpeckner.shinyapps.io/lincs_network_app/

CONCLUSIONS

- We have validated the GCP and P100 assays with high replicate consistency
- We have discovered both expected and novel connections
- Our connectivity framework allows for flexible assay integration, including non-proteomic data

NIH LINCS PROGRAM