Library of Integrated Network-based Cellular Signatures (LINCS) Virtual Symposium

National Institutes of Health

Office of Strategic Coordination—NIH Common Fund

November 19-20, 2020

February 3, 2021

This meeting summary was prepared by Caroline Sferrazza, Rose Li and Associates, Inc., under contract to the National Institutes of Health's (NIH) Office of Strategic Coordination (OSC). Additional writing support for Sessions III and VI was provided by Bethany Stokes, Rose Li and Associates, Inc. The views expressed in this document reflect both individual and collective opinions of the meeting participants and not necessarily those of NIH OSC. Review of earlier versions of this meeting summary by the following individuals is gratefully acknowledged: Lucas Smalldon, Nancy Tuvesson.

Acronym Definitions

AD	Alzheimer's disease
AE	adverse event
ALS	amyotrophic lateral sclerosis
API	application programming interface
AUD	Alcohol Use Disorder
CMap	Connectivity Map
DCIC	Data Coordination and Integration Center
DGE	differential gene expression
DIA	data independent acquisition
DToxS	Drug Toxicity Signature Generation Center
ECM	extracellular matrix
FAIR	findable, accessible, interoperable, reusable
FDA	Food and Drug Administration
GAN	generative adversarial network
GBM	glioblastoma multiforme
GCP	global chromatin profiling
GEO	gene expression omnibus
GR	Growth Rate inhibition
GREIN	GEO RNA-seq Experiments Interactive Navigator
(h)iPSC	(human) induced pluripotent stem cell
HMS	Harvard Medical School
KE	knowledge environment
LINCS	Library of Integrated Network-based Cellular Signatures
MEP	Microenvironment Perturbagen
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MSDW	Mount Sinai Data Warehouse
NCATS	National Center for Advancing Translational Sciences
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
NIH	National Institutes of Health
OSC	Office of Strategic Coordination
PBMC	peripheral blood mononuclear cell
PTM	post-translational modification
PRISM	Profiling Relative Inhibition Simultaneously in Mixtures
R&D	research and development
SMA	spinal muscular atrophy
TAS	transcriptional activity score
WGS	whole genome sequencing

Table of Contents

Meeting Summary

This 2-day symposium highlighted the impact of the Library of Integrated Network-based Cellular Signatures (LINCS) Program on the research community by featuring work from 25 LINCS and external investigators who have leveraged LINCS resources. The symposium topics spanned drug action and prediction of drug-related adverse events, integration of multiple data types, methodologies for rigorous reproducible biological research, computational tools for data integration and data FAIRness, and future challenges in perturbational biology. Each day consisted of three sessions, including four interactive workshops about ways to access and utilize LINCS resources; these workshops were conducted in parallel with poster sessions that featured 28 LINCS investigators. Including NIH staff, 475 people attended on November 19 and 334 people attended on November 20, 2020.

Introduction

Ajay Pillai and Albert Lee, NIH

LINCS is a National Institutes of Health (NIH) Common Fund consortium that aims to support a better understanding of human disease through high-quality, large-scale perturbation biology assays in complex human cellular models. Since the program's launch in 2010, 15 LINCS institutions throughout the United States have developed tools, cell lines, and datasets for small molecule, genetic, antibody, and microenvironment perturbagens that researchers can utilize to answer a wide array of biological questions. The LINCS consortium has generated more than 200 publications, and LINCS assets have been cited in many publications outside of the consortium—a sign of their growing utility in studying diverse aspects of human biology. A steadily increasing number of NIH grant applications use LINCS data, most of which are not submitted by LINCS consortium members. Private-sector companies and government organizations have also used LINCS tools and data to develop their approaches for drug discovery and development.

Session I: Undertaking Large-Scale Perturbation Studies: New Biology and Lessons

Generating and Mining Signatures of Drug-Induced Perturbations

Peter Sorger, Harvard Medical School, HMS LINCS

Harvard Medical School (HMS) LINCS was charged with generating resources that can derive signatures from cell perturbation assay-based molecular profiles, and assessing the phenotypes (e.g., growth, motility, cell death) produced by these signatures in the context of their molecular networks. Among these resources are the Small Molecule Suite, an open-access library for comparison of drug and target combinations, and Kinome.org, which maintains information on human protein kinases as therapeutic targets; both resources employ automated discovery tools. HMS LINCS also developed Growth Rate inhibition (GR) Metrics, a new way of computing dose–response relationships by correcting for differences in cell division. These metrics led to the generation of the online GR Calculator, which has been adopted almost universally in industry-based, preclinical pharmacology studies. In addition, HMS LINCS developed DyeDrop, a high-throughput, low-cost, imaging-based assay of perturbagen phenotypes with single-cell resolution.

HMS LINCS has also released large-scale "data cubes" illustrating the response matrices of many drugs over hundreds of cell lines. These data are housed within the Data Coordination and Integration Center

(DCIC) dataset and HMS LINCS Database. A wide variety of assays are used to collect these data, including DGE-RNA/Drug-Seq (i.e., rapid and inexpensive RNA sequencing tools that can be connected directly to the LINCS Connectivity Map [CMap] to query RNA signatures against other related molecules). HMS LINCS also participated in a consortium-wide reproducibility study that developed a series of specific guidelines for conducting reproducible drug–response assays in mammalian cell lines.

Transcriptomic Signatures for Kinase Inhibitors Drug Adverse Event Prediction in Human iPSC-derived Cardiomyocytes

Ravi Iyengar, Mount Sinai School of Medicine, DToxS

The Drug Toxicity Signature Generation Center (DToxS) focuses on the generation of transcriptomic signatures for Food and Drug Administration (FDA)-approved kinase inhibitors to predict cardiotoxicity. DToxS began this effort by testing a range of kinase inhibitors in commercially available human cardiomyocyte-like cells and performing 3'-DGE mRNAseq to produce transcriptomic signatures. These signatures were then matched with clinical risk scores for cardiac adverse events (AEs) and found to predict cardiotoxicity risk.

DToxS has produced 40 fully characterized induced pluripotent stem cell (iPSC) lines derived from a diverse pool of healthy individuals; these cell lines were deposited into WiCell for distribution to the academic community, and whole genome sequencing data are now available on Database of Genotypes and Phenotypes (dbGaP). Six of these iPSC lines were differentiated into cardiomyocytes to generate transcriptomic signatures for 54 FDA-approved drugs (primarily kinase inhibitors). A singular value decomposition–based analysis was used to deconvolute differential gene expression (DGE) data into components that are drug- and cell line-specific and into those components that describe the perturbational stress response more generally. DToxS has identified 54 drug-specific subspaces that allow for clustering of DEGs by drug across all tested cell lines; these subspaces will be analyzed in relation to clinical risk for cardiac AEs to determine the pathways that contribute to the at-risk phenotype and the genomic variants within individual cell lines that drive the expression patterns.

The Impact of Microenvironmental Signals on Molecular and Cellular Phenotypes of Mammary Epithelial Cells

Laura Heiser, Oregon Health & Science University, MEP-LINCS

This LINCS Program sought to identify linkages between cellular phenotypes and molecular states by leveraging the resources and expertise of the full LINCS consortium to systematically generate large-scale perturbation datasets that include a broad range of readouts (i.e., changes in signaling, epigenetics, transcription, protein expression, and behavior). Following a perturbation by extracellular growth factors and cytokines, researchers performed bulk molecular assays and imaging-based assays of behavioral phenotypes on MCF10A mammary epithelial cells to assess molecular responses. Multiple analytical approaches (e.g., pathway-based, statistical, and machine learning) were used to assess the relationships among these data types, and to determine how the data types were linked to observed cellular phenotypes. All project data are available on Synapse.

Individual extracellular ligands were found to induce unique varieties of cellular phenotypes (e.g., proliferation, differentiation, collective behaviors, and migration). Molecular responses that varied by ligand and stimulus duration were also observed at all molecular levels (i.e., protein, spatial protein, RNA, and chromatin). Consortium investigators performed a causal path network analysis to integrate data across these different assays and infer a network structure to explore stimulus-induced responses.

Proteins identified as likely mediators of a cellular response are interrogated in subsequent experiments to identify causal mechanisms and relationships between target protein activity and cellular response to the perturbagen. This study provides a framework for using a powerful perturbation biology approach (e.g., ligand combinations) to investigate complex stimuli.

LINCS Connectivity Map

Todd Golub, Broad Transcriptomics

LINCS CMap is an effort to represent, through mRNA signatures, relationships between biological disease states and perturbation (e.g., drug- or gene-induced) states. The resulting library of more than 3 million gene expression profiles is comprehensive, information-rich, easily searchable, and readily comparable to non-LINCS data, and accessible to bench and computational scientists. Because both genetic and pharmacologic perturbations are included in the dataset, this library can be used to interrogate the extent to which various perturbagens of either type are biologically equivalent (i.e., induce the same genetic signature) to facilitate drug discovery. To make the construction of such a large-scale library feasible, the L1000 assay was developed to produce a simultaneous transcriptional readout of 1,000 landmark genes.

Other readouts of perturbation are also available, including chromatin and phosphorylation readouts through the proteomics platform, cell morphology readouts through the imaging platform, and cell viability readouts through PRISM (i.e., Profiling Relative Inhibition Simultaneously in Mixtures)—an assay that enables rapid compound cytotoxicity profiling against more than 800 barcoded cancer cell lines. Data from all stages of analysis are publicly available, allowing scientists to access raw or processed data; more than 800 publications have cited LINCS CMap as a resource. Several crowd challenges have also been staged to solicit insight from the international community into the best data processing methods (e.g., how best to infer a full transcriptome from 1,000 measurements).

Establishing an iPSC-based Multi-omics and Cell-based Pipeline for Motor Neuron Disease

Clive Svendsen, Cedars Sinai, NeuroLINCS

Human iPSCs are a powerful tool for investigating the pathophysiologies of motor neuron diseases such as amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (SMA). NeuroLINCS utilizes a range of transcriptomic, proteomic, epigenomic, and imaging approaches to understand early molecular changes in motor neuron diseases and discover potential drug targets. Unlike other LINCS studies that screen many small molecules, NeuroLINCS focused on the downstream molecular cascades of a single genetic perturbagen—the C9ORF72 mutation associated with familial ALS—and found numerous causal and compensatory pathways that could lead to new therapeutic targets. The same multi-omics approach was applied to identify a genetic perturbagen for sporadic ALS; this effort spawned the Answer ALS project, which leveraged direct iPSC-derived motor neurons to identify ALS subtypes driven by genetic perturbagens among 1,000 patients. An experimental paradigm was developed to allow multiple NeuroLINCS sites, according to their expertise, to contribute omics data for each sample in this project, utilizing batch technical and differentiation controls to understand and quantify data quality.

Proteomic Characterization of APOE Phenotype and Drug Response

Maeve Bonner, MIT/Broad Proteomics

LINCS researchers leveraged the proteomics platform to identify the cellular changes that lead to a "risk state" for Alzheimer's disease (AD). Human iPSC models were generated to characterize cellular phenotypes associated with the presence or absence of AD risk variants, including the strongest validated risk factor for sporadic AD, APOE; isogenic cell lines were created by CRISPR/Cas9 using iPSCs from APOE3+ (control) and APOE4+ (at risk) donors. Unbiased epigenetic and phosphorylation profiling revealed a distinct APOE4 risk signature at both levels. The APOE4 risk signature has begun to provide novel insights into the biology of APOE4 and has been used to query the LINCS database to identify candidate therapeutics that successfully promoted non-risk (i.e., APOE3-like) profiles in APOE4 cells and vice versa (i.e., induce a risk profile in APOE3 cells). A newly available data independent acquisition (DIA) cloud-based pipeline enables deeper mapping of the genetic risk phosphoproteomic landscape, as well as identification of more key kinase pathways, which will ultimately foster deeper insights into APOE4 biology.

Where to Find and How to Use the LINCS Data and Tools

Dušica Vidović and Mario Medvedović, DCIC

The LINCS Consortium established two main metadata categories—reagents and experiments (each with specified subcategories)—and developed standard specifications to annotate the wide variety of LINCS data and enable data integration; these standards are available at FAIRsharing.org. After data and metadata are received, they are validated, processed, and packaged for release on the LINCS Data Portal, where they can be filtered by various annotations (e.g., assay method, LINCS center) within the Datasets application. Every dataset landing page lists annotations, reagents, and corresponding metadata along with data packages organized by data level for download. To improve accessibility of LINCS signatures, high-level data are parsed from each data package so that signatures may be queried by metadata, gene, and chemical structure. The LINCS Data Portal has also begun to integrate Google BigQuery functionality to enable rapid/granular queries; provide long-term, no-cost access to LINCS data; and allow future integration with other BigQuery datasets.

The online data analysis portal—iLINCS—is a user-friendly platform that integrates omics datasets and signatures with analytical tools and interactive visualizations (both LINCS and non-LINCS signatures and datasets are available). Users can submit a signature or gene list, select from more than 200,000 pre-computed signatures, or generate a signature *de novo* by selecting data from more than 14,000 omics datasets for various analyses. LINCS tools—including iLINCS—are connected to the LINCS Data Portal, GEO RNA-seq Experiments Interactive Navigator (GREIN), piNET, and other data annotation resources. All LINCS tools created by LINCS centers and DCIC can be found at LINCSproject.org.

Session II: Impact of LINCS on the Community: Short Community Vignettes

PredicTox Knowledge Environment: LINC-ing Data to Cardiotoxicity

Rebecca Racz, FDA

PredicTox is an international public-private partnership to build a precompetitive research and development (R&D) platform for drug safety analysis, as well as to create both a "knowledge environment" (KE) for sharing data and an analysis platform to test hypotheses. The PredicTox KE integrates traditional AE data from clinical trials and pharmacovigilance with new sources (e.g.,

biomarkers, signatures, in vitro and in vivo models, and drug target/absorption, distribution, metabolism, and excretion data). The LINCS center DToxS contributes transcriptomic DGE data for 47 FDA-approved drugs—primarily drugs such as tyrosine kinase inhibitors and monoclonal antibodies that produce unanticipated cardiac side effects—and measurements from iPSC-derived cardiomyocytes. The overlap between drug-induced DGE and known disease-related genes allows for identification of cardiotoxicity risk. The KE also integrates molecular pathway data from public databases to build gene and protein networks relevant to cardiotoxicity. The main challenge for the KE hinges on data integration and harmonization, because data require varying levels of cleaning and curation based on their various formats and sources.

Generating Hit-like Molecules from L1000 Profiles Using Artificial Intelligence

Oscar Méndez-Lucio, The Janssen Pharmaceutical Companies of Johnson & Johnson

The CMap L1000 dataset contains more than 1.3 million induced gene expression profiles that can be used for many purposes, including predictive and generative modeling. Whereas predictive models associate known molecules with their resulting effects (e.g., DGE) to help predict the effects of a new molecule, generative models associate known molecules with effects *and* vice versa, such that a desired gene expression profile (e.g., one without toxic effects) can be entered into the model to produce a new molecule that is likely to induce the desired profile. In a proof-of-concept test, a generative adversarial network (GAN) (one type of generative model) was trained on 20,000 compounds from the L1000 dataset to predict new molecules likely to knockout a genetic target. The GAN produced new molecules, for each of 10 knockout profiles, that were structurally similar to known inhibitors. This artificial intelligence method can also be used with other phenotypic data (e.g., morphology) to inspire molecule optimization and library design.

Identifying Novel Addiction Treatment Strategies Through Gene Expression and LINCS Analysis

R. Dayne Mayfield, University of Texas at Austin

Only four medications for alcohol use disorder (AUD) have been made publicly available since 1951, highlighting an urgent need to pursue new drug discovery approaches. Signatures from the LINCS L1000 dataset were integrated with DGE data from rodent models of voluntary alcohol drinking to generate a list of candidate drugs predicted to decrease alcohol intake. In one study, two candidate drugs were selected based on signatures that were negatively correlated with genomic signatures from rodent models of AUD risk. These drugs were administered to alcohol-naïve mice of the same genetic model, and drinking behaviors were assessed after treatment. Both candidate drugs were highly effective in reducing alcohol intake and blood alcohol levels in the rodent models. In a larger study that drew from 86 unique DGE signatures from a variety of alcohol-exposed models of drinking, 78 percent of compounds predicted by LINCS analysis to decrease voluntary alcohol consumption produced this behavioral phenotype. The addition of more brain-relevant cell lines and brain-related compounds to the LINCS dataset, as well as the utilization of high-throughput behavioral screening, machine learning (ML), and single-cell based methods, will further strengthen the capacity for LINCS analyses to predict successful drug candidates.

Session III: Hands-On Workshop/Poster Session

LINCS Transcriptomics: Data, Tools, and Workflows for Connectivity Map Analysis

Rajiv Narayan, Ted Natoli, Anup Jonchhe, and Jacob Asiedu, Broad Transcriptomics

CMap Resource Updates

CMap uses mRNA signatures of cell state to link genes, drugs, and diseases. To make this resource as rich and diverse as possible, the L1000 assay was used to experimentally measure a reduced representation of the transcriptome and then computationally infer thousands of other genes, which allows queries to identify approximately 80 percent of the gene–drug–disease associations that would be identified by measuring the full transcriptome. The previous CMap contained approximately 1.3 million profiles from 28,000 perturbagens and 80 cell lines. The 2020 CMap release has expanded into new perturbations (e.g., CRISPR) and cell lines (e.g., hematopoietic cell lines, non-cancer cell lines); the updated resource contains 81, 979 perturbagens, 33,609 compounds, 657 mechanisms of action, 9,288 genes, 240 cell contexts, 3.18 million profiles, and 1.16 million signatures. For greater accessibility, the data are being migrated to Google BigQuery to facilitate indexing and a controlled vocabulary as well as efficient access to arbitrary slices of data without requiring users to download the entire matrix. Other data access tools include web apps on clue.io; data application programming interface (API) via clue.io; the Python CMap BigQuery toolkit; and the BigQuery console; data can also be downloaded from DCIC, clue.io, and gene expression omnibus (GEO). The Python CMap BigQuery toolkit is a Python package (cmapBQ) that allows for targeted retrieval of relevant gene expression data from the CMap dataset on BigQuery.

Sample Use Cases

Users can leverage CMap to query signatures for a broad range of connections to an inputted gene set (e.g., compounds, targets, mechanisms of action); a curated visualization of exemplar signatures are returned to the user and full results are available for download via the History app. To support drug repurposing efforts, the new CMap release contains 1,500 compounds profiled at a 6-point dose series in 10 cell lines; these data can be used to identify dose responses through transcriptional activity by plotting a dose–response curve of the transcriptional activity score (TAS), which computes both the strength of a signature (i.e., the number of genes significantly changed by a treatment) and the correlation across the biological replicates that comprise the signature to assess the impact of a treatment. For CRISPR projects, the new CMap release contains 5,116 unique genes in 10 Cas9-derivatized cell lines, enabling users to determine whether CRISPR reduces the expression of a targeted gene and how CRISPR signatures correspond with shRNA and compound signatures. A subset of data containing 140 drugs in 11 hematopoietic cell lines can be used to investigate mechanisms that are active in hematopoietic cell lines but not in solid tumor lines, and to identify distinctions between the various diffuse large B-cell lymphoma subtypes represented in L1000 data.

Proteomics: Roundtable Discussion Focused on Extending Use of Available LINCS Data

Andrea Matlock, Cedars-Sinai, and Mike MacCross, University of Washington

Future Directions for LINCS Data

LINCS generates data that can be re-analyzed in new ways to reveal novel insights, which increases the utility of these datasets; however, such re-analyses have not yet occurred. Participants recommended encouraging LINCS investigators and external researchers (via the LINCS Request for Applications) to begin these re-analyses, particularly with P100 or Answer ALS data. Participants also recommended harmonizing analytical pipelines—including the quality control procedures (e.g., common reference

controls)—across proteomics laboratories to facilitate easy transfer of protocols between laboratories for re-analyses and validation studies. One participant suggested requesting proteomics journals to require stringent quality control criteria before publishing proteomics studies.

The Broad Institute's LINCS Center has developed a spectra-based algorithm to identify post-translational modifications (PTMs) by comparing detected spectra with a known library of isomer spectra. This method has worked previously to identify phosphorylation sites; however, the PTM can only be identified if the spectra corresponds to an isomer within the library. Thus, participants agreed that development of new proteomic analysis tools and methods, or refinement of existing approaches, is required to better localize PTMs within LINCS datasets.

Rapid High-Content Measurement of Perturbagen Response in Living Cells: DyeDrop Assays, GR Metrics, and Multi-Center Reproducibility in the Face of Confounding Variables

Caitlin Mills, Harvard Medical School

The HMS LINCS Center developed GR metrics for perturbation response studies that will allow researchers to better interrogate underlying perturbagen phenotypes (e.g., cytotoxicity) in a cell division–independent manner. GR metrics compare cell counts at baseline, time of treatment, and post-treatment and involve summary statistics (i.e., GR_{50}, GR_{max}, and GR_{AOC}) comparable to traditional methods. Analyses using GR metrics result in net cell count curves that can be decoupled to show the individual normalized cell growth and death rates, which in turn illustrate how many viable or dead cells are present at different timepoints and can identify confounding biases. The HMS LINCS Center uses DyeDrop assays for studies generating GR metrics; these assays perform best when using monolayer cell lines that are seeded using densities to induce exponential growth. Cells can be marked with dead/alive markers or more descriptive markers (which can be specific to the study) before being classified based on their Hoechst signals (which indicate whether a cell is viable) and analyzed to generate a normalized cell count table, as well as GR metrics and summary plots. The GR Calculator can also be used to generate GR metrics on already existing datasets. To ensure that different studies using GR metrics can be compared, all elements of the study should be automated and optimized.

MCF10A: Building an Integrative Data Matrix of Perturbation Responses

Laura Heiser, Sean Gross, and Mark Dane, Oregon Health & Science University

Critical Decision Points in a Large Multi-Center, Assay, Timepoint, and Ligand Experiment

The goal of the MCF10A project is to identify high-impact ligands that can be used to phenotype cells and ultimately to elucidate the molecular basis of those phenotypes. To achieve this goal, the study team evaluated multiple decision points that would create the most robust dataset. It selected MCF10A cells because they are well-studied, diploid, dynamic, and growth factor–dependent. Six ligands (i.e., TGFB, BMP2, IFNG, EGF, OSM, HGF) were selected from the MCF10A LINCS Receptome because of their large effect sizes and high expression patterns; ligand doses were selected according to each ligand's strongest perturbagen response. Investigators designed a treatment paradigm that included one PBS wash, one media exchange, and six timepoints of data collection (i.e., 0, 1, 4, 8, 24, and 48 hours), with each sample being collected in triplicate. The study team selected seven assays (i.e., RPPA, cycIF, IF, RNASeq, L1000, global chromatin profiling [GCP], and ATACSeq) to collect key information on molecular signaling, epigenetics, and genetics. These assays were executed in three collection batches in order to ensure sufficient cell growth. Collection 1 included RPPA, IF, RNASeq, ATACSeq, and L1000; Collection 2 included IF, GCP, and L1000; and Collection 3 included cycIF and L1000. Collections 1 and 2 occurred at

Oregon Health & Science University, whereas Collection 3 occurred at HMS and the Broad Institute and thus many procedures had to be optimized and coordinated to ensure consistent cell handling.

MCF10A Data Access

Data from the MCF10A study are hosted on the Synapse platform's LINCS MCF10A Molecular Deep Dive webpage. Upon accessing the webpage, users can click on the "Data Files" tab, which will display multiple folders for each assay used during the MCF10A project. Each assay folder includes multiple data types (from raw data to median-summarized data, or level 1 to level 4, respectively), as well as metadata (for replicate treatments and multiple timepoints for that assay), a README file, and descriptions of the various levels of generated data; imaging data are not yet available on the Synapse platform. Users can view the provenance of each data type to assess the steps taken to transform data into levels 1 through 4. Users can also filter the LINCS MCF10A datasets according to cell type, data level, or assay type, and all data available on the platform are downloadable.

Session IV: Integrative Data Analysis within Perturbational Studies

LINCS and Connectivity Map Datasets, Tools, and Challenges

Aravind Subramanian, Broad Transcriptomics

CMap is a functional lookup table for biology in which signatures (e.g., of drugs, genes, or diseases) can be analyzed to derive information about mechanisms of action, targets, pathways, and disease indications. The LINCS Program has expanded the CMap dataset to include more than 3 million L1000 profiles based on more than 80,000 perturbagens with greater coverage of various cell lines and mechanisms of action; this expansion is likely to improve analyses that utilize CMap data, and integration of multi-dimensional data can help to prioritize results. The full CMap data matrix is available for download, but Google BigQuery also allows users to select subsets of data to access, and online apps can be used to query the dataset without downloading it (Clue Query is the core app that allows users to input a gene set to identify correlations within the CMap dataset). The platform supports use of APIs to download query results for post-processing. Every CMap signature is labeled with a HiQ metric, which represents the relative confidence that the signature and any constituent profiles are of adequate quality for future analysis; the dataset is not yet optimized for every cell line or gene of interest. Signatures are also assigned a TAS to indicate the number and magnitude of fold changes induced by a perturbagen, which can be used to determine dose responses. Consensus gene knockdown signatures can be used to enable discovery of connections to the consensus signature, which helps users to adjust for off-target effects of genetic perturbations (e.g., shRNA). Crowd challenges have improved parts of the analysis pipeline, including the inference algorithm, as well as the query implementation and gene deconvolution processes.

A Library of Induced Pluripotent Stem Cells from Clinically Well-Characterized, Racially Diverse Healthy Human Individuals

Christoph Schaniel and Nicole C. Dubois, Mount Sinai School of Medicine, DToxS

The DToxS LINCS Center created the first library of hiPSCs from a cohort of 40 racially and ethnically diverse healthy individuals; health status was defined by a set of rigorous inclusion and exclusion criteria. All hiPSCs have been well characterized in terms of karyotype, whole genome sequencing (WGS), and pluripotency, and are being banked at WiCell; WGS identified some hiPSC donors as carriers of disease risk variants. The gene signature of each hiPSC line is independent of the age or sex of the donor. To assess variability with single-cell resolution, two hiPSC lines were differentiated to atrial and

ventricular cardiomyocytes and assessed on multiple levels. Line-to-line and clone-to-clone variability was evident in the efficiency of differentiation, although purification of cultures was equally efficient regardless of differentiation efficiency. Expected physiological differences between atrial and ventricular cardiomyocytes were reliably observed across lines and clones; variability between individual cells did not cluster by technical replicate, which indicated absence of batch effects.

Interactions Between Drug, Genomic, and Environmental Perturbations

Jim Korkola, Oregon Health & Science University, MEP-LINCS

Although multiple mechanisms that confer therapeutic resistance in cancer cells have been identified, the role of the tumor microenvironment in developing resistance has been understudied. Microenvironment microarrays allow researchers to explore the effects of the microenvironment by culturing cells in wells containing "growth pads," which are comprised of extracellular matrix (ECM) proteins that can be combined in more than 2,500 ways to create unique microenvironments. Microenvironment Perturbagen (MEP)-LINCS uses high-throughput imaging to process these samples for integrative analysis and network discovery. The microenvironment microarray has been used to demonstrate cell subtype–specific and mutation-associated responses reflecting fundamental differences in receptor expression, usage, and downstream signaling pathways that influence resistance phenotypes.

NeuroLINCS: A Collaborative Multi-omic Approach to Define Disease Signatures

Leslie Thompson, University of California, Irvine, NeuroLINCS

NeuroLINCS is a multi-site effort to generate signatures from hiPSC-derived neural cell types in response to genetic perturbagens associated with motor neuron disease. Integrated signatures are produced by multi-omic analysis (e.g., of RNA-seq, proteomic, and ATAC-seq data). These integrated data can be used to infer new networks and previously undetected protein interactions, even with small sample sizes. Incorporating ML methods can expand the use of these data to identify new variant combinations that may give rise to motor neuron disease. Inferred networks can be experimentally validated; for example, inferred networks associated with an ALS mutation were genetically perturbed in a fly model (i.e., loss-of-function alleles for network nodes were crossed with a C9ORF72 mutant) to designate network-level changes as either compensatory or causal based on the resulting phenotype. Networks were validated across lines and differentiation protocols; pathways overlapped with RNA-seq data from postmortem patient samples.

In a second disease-related perturbation, inferred networks were significantly enriched for known genes associated with SMA. As expected, many genes exhibited alternative splicing, and RNA-seq data validated the upregulation of an SMA-associated splice variant in hiPSC models. NeuroLINCS has laid the groundwork to characterize genetic perturbations in individual patients that can enable personalized medicine in the future, including patient stratification and identification of disease progression clusters to predict rate of decline and survival outcome.

Integrating and Mining Multi-omic Data on Cell State and Phenotype

Caitlin Mills, Harvard Medical School, HMS LINCS

Predictive models of drug response in breast cancer cell lines were generated from baseline molecular expression (i.e., RNA, protein, phosphoprotein), drug sensitivity, and inferred kinase activity profiles. LINCS researchers performed DyeDrop assays to measure the dose response to 70 kinase inhibitors in 70

breast cancer cell lines and to confirm the predictive accuracy of the models. Predictive accuracy for each model depended on the baseline data type used to build that model, as well as on the drug class (e.g., kinase activity was the most predictive data input for CDK inhibitor sensitivity but poorly predicted other drug responses). The predictive models demonstrated the ability to differentiate among inhibitors with the same nominal targets, identifying inhibitor-specific characteristics that can be used to predict off targets, candidate targets for combination treatments, and candidate biomarkers.

Another multi-omics approach was also used to describe the cell phenotypes produced by two kinase inhibitors with the same nominal target (CDK4/6). Bulk RNA sequencing produced a signature unique to one of the inhibitors (abemaciclib) that revealed this inhibitor's pan-CDK activity when queried in CMap. Phosphoproteomics profiling identified numerous candidate kinase targets, which were confirmed as abemaciclib targets by in vitro kinase profiling, indicating that abemaciclib was less selective than some other CDK4/6 inhibitors. Because it results in mixed cell cycle arrest and can be observed in vivo (demonstrating functional consequence), this polypharmacology may explain cells' phenotypic differences.

A Mass Spectrometry Cloud-based Pipeline Enables the Accurate Analysis of Thousands of Phosphosites in P100 Datasets

Karen Christianson, Broad Proteomics

Broad Proteomics built an unprecedented library of more than 7,000 phosphoproteomic and epigenetic profiles from several cancer and neural cell lines for 120 drug perturbations that are classified as epigenetically active, neuroactive, or cardiotoxic, or as kinase/pathway inhibitors; all profiles are available on PanoramaWeb. LINCS researchers used the P100 and GCP assays, which quantitatively measure a reduced representation of the phosphoproteome and chromatin modifications, respectively, to generate corresponding signatures; chromatographs for each peptide were extracted using Skyline. While the P100 assay targets only 96 phosphopeptides, a new cloud-based DIA analysis pipeline was developed to enable high-throughput quantification of thousands of phosphopeptides from raw P100 data. Phosphopeptides are identified by Specter, a novel tool that deconvolutes complex DIA spectra and can distinguish between peptides with similar sequences, such as alternate phosphosite localizations. Skyline and Avant-garde (another novel tool that chooses the most suitable peaks for accurate quantification and eliminates manual peak validation) are then used, respectively, to extract and refine signals, which reduces quality control timelines from weeks to hours; Avant-garde is available as an external tool in Skyline. The increased coverage enabled by DIA enables deeper and more nuanced analysis of the phosphosignaling landscape than was previously possible.

The Many Ways the Community Utilized the LINCS Resources

Avi Ma'ayan, Mount Sinai School of Medicine, DCIC

At least 56 tools and databases have been produced by LINCS; these tools have been utilized by more than 1.2 million unique users and are accessed by more than 3,000 unique users daily. During 2020, more than 5,000 papers cited the DCIC grant. In one application, a series of LINCS tools—including BioJupies and Enrichr—were used to identify signatures from SARS-CoV-2 and hydroxychloroquine, a well-publicized drug that blocked SARS-CoV-2 infection in a cell-based assay but has not demonstrated clinical efficacy. Signatures were projected onto the L1000FWD Map—a display capturing more than 17,000 signatures for approximately 5,000 drugs—which revealed a region of overlap indicating that SARS-CoV-2 downregulates pathways that hydroxychloroquine upregulates. With this knowledge, seven additional drugs were chosen that target this region, and all prevented infection by SARS-CoV-2 in

pancreatic and lung organoids, emphasizing the potential role of these pathways in SARS-CoV-2 infection mechanisms.

In a related effort, signatures related to COVID-19 collected from the research community were published in the COVID-19 Drug and Gene Set Library. Conversion of data from 12 drug screens to L1000 signatures revealed enrichment for drugs in the same pharmacological space as hydroxychloroquine. This finding was visualized by the Drugmonizome Consensus Terms Appyter (the more than 40 available Appyters enable rapid creation of bioinformatics web-based applications from Jupyter notebooks).

Session V: Impact of LINCS on the Community: Short Community Vignettes

gDR: A Software Suite for Drug Response Data

Marc Hafner, Oncology Bioinformatics, gRED, Genentech

When experiments are performed carefully and data are handled properly, drug response data are generally reproducible. The Genentech screening facility relies on standardized protocols, barcoded plates, registered reagents, and established pipelines to generate reproducible data. Using LINCS as a proof of concept to address the common challenge of siloed data, Genentech built the gDR drug response software suite. This suite, which provides a central repository for all in vitro data as well as a self-service tool for data processing and a series of standard analyses, enables easy access to data with duplicates removed. Response data are also converted to a data frame matrix —a list of data frames shaped as a matrix and embedded in a "SummarizedExperiment" with data levels as multiple assays—to allow for simpler analysis and storage of a single object throughout processing without risk of loss. The GR method for quantifying drug efficacy, standardized LINCS protocols, and the GR Calculator were especially important LINCS resources for the gDR blueprint. The suite was designed according to FAIR (i.e., findable, accessible, interoperable, reusable) principles and includes an underlying local database, enforced unique identifiers for metadata, traceable data processing, and modular design, as well as an open-source release. The core gDR database can reference external databases, and outputs can be processed and packaged for public release.

Invasion of Homogeneous and Polyploid Populations in Nutrient-limiting Environments

Noemi Andor, Integrated Mathematical Oncology Department, Moffitt Cancer Center

The Andor lab studies intra-tumor heterogeneity by interpreting genomic differences and quantifying phenotypic differences among cells. One of the lab's projects focused on the "tip-over" hypothesis, which states that cells are sensitive to DNA-damaging agents when therapy-induced DNA damage pushes a cell's genomic instability beyond a tolerated limit. This hypothesis is related to a tumor's fitness, which may decline as copy number variation abundance (i.e., genomic instability) rises and crosses a threshold because of the energetic demands of associated high ploidy levels. The lab calibrated a mathematical model that can be used to understand the impact of resource limitations on heterogeneous cell populations using LINCS data from microenvironment microarrays. Changes to all five model parameters—chemotactic coefficient, energy diffusion, energy consumption, chemotactic inability, and sensitivity to low energy—resulted in ECM-specific differences. In polyploid populations, the model revealed spatial segregation of growth patterns modulated by the chemotactic threshold (i.e., energy deficit inducing chemotaxis) of the high ploidy population. Researchers can use this model to determine what levels of ploidy different tissues can afford (in terms of energy) as well as to understand

the potential benefits of a given ploidy level (e.g., high ploidy may mask deleterious mutations). The model is currently being used to analyze data from primary and recurring glioblastoma.

LINCS-based Approach to Identify Anti-Atrophogenic Compounds to Protect Skin from Glucocorticoid-induced Atrophy

Irina Budunova, Northwestern University

Glucocorticoids are among the most frequently used drugs in dermatology, but their therapeutic effects are often accompanied by pronounced AEs. A major side effect of topical glucocorticoid treatment is skin atrophy. Researchers have discovered and validated several atrophogenes. Among these atrophogenes, REDD1 was identified as one of the central drivers of steroid-induced skin atrophy. Because no pharmacological REDD1 inhibitors are known, researchers used bioinformatics approaches via the LINCS database to search for REDD1 inhibitors among FDA-approved and experimental drugs. Nearly half of the first 20 REDD1 suppressors identified in the LINCS database are PI3K/AKT/mTOR inhibitors. All tested PI3K/AKT/mTOR inhibitors prevented activation of REDD1 by glucocorticoids in vitro and in vivo*, and* pre-treatment of mice with the PI3K inhibitor LY-294002 did not influence the anti-inflammatory effect of glucocorticoids but did protect skin against atrophy induced by systemic dexamethasone treatment. The findings suggest that inhibition of REDD1 can dissociate the therapeutic and adverse atrophogenic effects of glucocorticoids, and that co-administration of a REDD1 inhibitor may allow patients to receive glucocorticoid treatment while also protecting their skin.

Single-cell-driven Drug Repurposing in Atherosclerosis

Chiara Giannarelli, Mount Sinai School of Medicine

Atherosclerosis is a systemic disease initiated by lipid accumulation in the arterial wall and sustained by cardiovascular inflammation that eventually causes clinical cardiovascular events (e.g., stroke or heart attack). LINCS resources were leveraged to understand how immune cells respond to lipid accumulation and chronic inflammation, as well as how that progression leads to cardiovascular events. Researchers treated peripheral blood mononuclear cells (PBMCs) from cardiovascular patients and healthy donors with plasma either from healthy patients or from patients with cardiovascular conditions. They then generated proteomic and transcriptional signatures with single-cell resolution. The resulting signatures showed an inflammatory response to atherosclerotic plasma. These inflammatory gene signatures were queried against the LINCS database and, based on those queries, eight Phase 2a–ready drugs were chosen for screening.

Candidate drugs were validated ex vivo (by PhosphoCyTOF, the same method used in the discovery phase) to determine which were most effective at reversing the inflammatory signature produced in response to atherosclerotic plasma. Further in vivo validation showed that administering the most promising drug resulted in reduced macrophage accumulation and aortic lesions, and chronic administration of the candidate drug in rabbits reduced previously observed signs of atherosclerosis. Analysis of the candidate's transcriptional signature revealed an upregulation of genes known to be anti-atherosclerotic, providing some insight into the mechanism of action.

Utilizing LINCS Data to Identify Synergistic Combinations in Glioblastoma

Nagi Ayad, University of Miami Brain Tumor Initiative

Glioblastoma multiforme (GBM) is the most common and aggressive adult brain tumor type, but no new and effective treatments have been approved during the past 20 years, and no monotherapies have

ever been successful. Investigators have identified BRD4 as one potential therapeutic target. BRD4 regulates the noncoding RNA HOTAIR; because levels of HOTAIR are higher in the serum of GBM patients than they are in healthy controls, levels of HOTAIR may provide a useful biomarker for responsiveness of GBM and other cancers to epigenetic pathway inhibitors. The SynergySeq platform was developed to identify small molecule combinations with predicted synergy, and in this case, researchers used it to identify compounds that would target GBM and synergize with bromodomain inhibitors.

Researchers generated L1000 data from GBM cells treated with the JQ1 bromodomain inhibitor, and then derived transcriptional signatures, which allowed them to identify a highly selective Aurora kinase inhibitor based on its orthogonal signature to JQ1. Combination treatment with both the Aurora kinase inhibitor and JQ1 demonstrated a synergistic effect (i.e., reduced GBM tumor growth) when administered using xenograft and in vivo models. Proteomic information from GBM cells revealed that newly diagnosed and recurrent GBM tumors have distinct active kinases; this result was surprising given the transcriptomic similarities observed between the two cancers and implies that different compounds may be more or less synergistic with JQ1 in newly diagnosed versus recurrent GBM. LINCS resources are also being used to analyze proteomic and transcriptomic signatures of treated tumor samples collected from patients, in an effort to calculate the remaining disease signature and identify therapeutic agents based on their elimination of that signature, as well as to identify compounds that may restore MLH1 function (loss of which is common in recurrent GBM) and thereby renew sensitivity to the standard of care GBM treatment TMZ.

Session VI: Hands-on Workshop/Poster Session

LINCS Transcriptomics: Data, Tools, and Workflows

Daniel Clarke, Minji Jeon, and Avi Ma'ayan, Mount Sinai School of Medicine, DCIC

BioJupies

BioJupies is an interactive RNA sequencing pipeline for researchers without advanced coding knowledge. Users can analyze their own data in the cloud environment or work with published data. The user-friendly BioJupies interface prompts users to input information about their data (e.g., metadata) and select from a series of analyses to execute a cloud-based Jupyter notebook that produces a publication-style report; Jupyter notebooks can also be downloaded for customization and local execution. BioJupies can only compare two groups at once (e.g., control versus perturbation), so multiple comparisons require the generation of multiple Jupyter notebooks.

Appyters

The Appyters platform enables rapid creation of bioinformatics web-based applications from Jupyter notebooks. Each of the 40 currently available Appyters is designed to perform multiple analyses (e.g., the Bulk RNA-seq Analysis Appyter allows non-computational researchers to analyze and visualize RNA-seq datasets). To run an Appyter, users must first prepare an RNA-seq dataset, which may involve specifying parameters such as normalization methods. Analysis outputs include various interactive plots (e.g., 3D UMAP, clustered heatmap). Another Appyter, the Drugmonizome Consensus Terms Appyter, performs drug set enrichment analysis for uploaded drug sets against a curated set of drug set libraries. The bioinformatics community can contribute Appyters based on their own Jupyter notebooks by adding annotations to the workflows that will allow users to easily work with the code even without computational experience.

Use of iPSC-derived Cell Types for Perturbation Biology

Dhruv Sareen and Arun Sharma, Cedars Sinai, NeuroLINCS; Priyanka Narayan, NIDDK; Nicole Dubois and Mustafa Siddiq, Mount Sinai School of Medicine, DToxS

Using iPSC-derived Neurons for Assessing Drug Perturbation-induced Signatures

Unexpanded PBMC-derived iPSCs provide a platform for assessing drug perturbation biology in relevant cells for characterization of cell type–specific signatures. Given the inherent variability of iPSC lines (due to factors such as variability in culture methods and reagents), iPSCs from two institutions—Cedars Sinai and Mount Sinai—were standardized and characterized by an independent collaborator at the National Center for Advancing Translational Sciences (NCATS). All LINCS iPSCs are derived from PBMCs to reduce the incidence of karyotype abnormalities. The LINCS iPSC common project derived neurons and cardiomyocytes from iPSCs for perturbation with FDA-approved drugs and generation of resulting transcriptional signatures by L1000 to establish relationships between short-term and long-term pathways for eight drugs known to impact these cells. Researchers sorted the connectivity results by the average across drug and vehicle-treated samples at this timepoint and observed a large degree of similarity across most samples. At 48 hours, however, distinct clusters of compounds began to produce distinct perturbations, indicating a separation of perturbation signatures that was consistent across cell lines. Researchers can use such analyses to identify distinct cell type–specific responses and unique drug perturbation signatures.

Using iPSC-derived Neural Cell Types to Investigate Alzheimer's Disease Risk

Large-scale genomic studies have identified single point mutation risk factors for AD. Because the expected effect size of these disease-promoting (rather than disease-causing) mutations is small, isogenic iPSC pairs are generated by CRISPR/Cas9 to control for the genetic background of patient donors. Using these iPSC lines, researchers can derive a range of neuronal and glial subtypes that can in turn be studied in various formats (e.g., purified culture, co-culture, organoid). High-throughput and targeted phenotyping of these lines/formats can facilitate therapeutic discovery and provide insight into the mechanisms by which these mutations influence the fundamental biology of different cell types.

Use of iPSC-derived Cardiovascular Cell Types for Perturbation Biology

DToxS derived highly enriched cultures of ventricular cardiomyocytes from hiPSCs on a large scale. To generate these cells in large numbers from different hiPSC lines with high purity, several challenges had to be addressed. First, to produce sufficient numbers of cardiomyocytes for perturbation experiments—approximately 60-80 million per experiment from the same differentiation to minimize batch effects—DToxS scaled its manpower and formed teams to maintain a labor-intensive, established suspension protocol. Second, DToxS optimized its purification strategies to ensure consistent purity and reduce line-to-line variability. Although fluorescence-activated cell sorting (FACS) is the most effective way to achieve high-purity cultures, it is too costly to be feasible on this scale. Instead, DToxS used metabolic selection with lactate to generate enriched cardiomyocyte cultures regardless of original differentiation efficiency with a cutoff of 95 percent purity. Third, a protocol to modulate PPAR signaling was described to enhance the maturity of hiPSC-derived cardiomyocytes by inducing more mature metabolic activity.

Human iPSC-Cardiomyocytes Are Susceptible to SARS-CoV-2 Infection

Because human cardiac tissue samples are difficult to obtain, DToxS used hiPSC-derived cardiomyocytes to investigate the course of SARS-CoV-2 infection in cardiac tissue. DToxS used an image-based analysis to demonstrate that SARS-CoV-2 infects cardiomyocytes directly and induces apoptosis. Treatment with an ACE-2 antibody reduced overall Spike protein expression, suggesting that SARS-CoV-2 infects

cardiomyocytes via an ACE-2-dependent mechanism. Transcriptomic data confirmed that SARS-CoV-2 actively infected cardiomyocytes and upregulated pathways involved in innate immune response.

Pathophysiology of Cardiomyocyte Infection by SARS-CoV-2

The Mount Sinai Data Warehouse (MSDW) is a resource for researchers to access anonymized COVID-19 patient information. It includes data on patients without prior history of cardiac disease who develop cardiac dysfunction after SARS-CoV-2 infection. Researchers categorized patients based on levels of Troponin I, which are indicative of cardiac damage when elevated. Nearly half of patients who developed reduced left ventricular ejection fraction demonstrated elevated troponin levels. Infected hiPSC-derived cardiomyocytes showed disrupted expression of Troponin I that was exacerbated by addition of interleukins. Interleukins were found to enlarge cardiomyocytes, and the combination of interleukins with SARS-CoV-2 infection resulted in greater troponin release (as measured by ELISA) and more attenuated beating; interleukins appeared to have no impact on the infectivity of SARS-CoV-2.

Accessing and Integrating LINCS Data with iLINCS and LDP

Vasileois Stathias and Amar Koleti, University of Miami, and Mario Medvedović, University of Cincinnati

LINCS Portal Capabilities and Resources

LINCS launched the first iteration of the LINCS Portal, which aimed to store data for download by the broader research community, in 2016. Then, in 2019, it launched Portal 2.0, which allowed users to better filter LINCS data and more quickly identify molecular signatures of interest for further analyses. Portal 2.0's main feature is a search bar, which has three tabs: (1) Metadata Search, (2) Signature Search, and (3) Chemical Structure Search. The Metadata Search queries across all perturbations, model system, and signatures in LINCS datasets. The Signature Search allows users to input specific genes (as well as whether they are upregulated or downregulated in a preferred signature) and to identify LINCs datasets (characterized by assays, perturbagens, and model systems used) that contain the queried genetic signature. The Chemical Structure Search allows users to query using simplified molecular-input line-entry system (SMILES) images and identify signatures that either involve the input structure or are similar to that structure. After identifying signatures or perturbagen datasets of interest, users can download data either directly from the Portal or through an API (provided on the LINCS Swagger page). LINCS data are also hosted within Google's public BigQuery datasets. LINCS BigQuery datasets include the same metadata, reagent descriptions, and signature IDs available on the LINCS Portal; however, BigQuery hosts only normalized and processed LINCS data, not raw data. Users can access the LINCS BigQuery datasets by searching through the Google BigQuery website or by selecting the BigQuery tab on the LINCS Portal (which also includes multiple walkthrough scenarios that can be launched directly from the Portal into BigQuery).

iLINCS

iLINCS is an integrative database that allows users to analyze both LINCS- and non-LINCS-generated omics and perturbation signatures. Using iLINCS, users can search for signatures specific to a pharmacological action, mechanism of action, or genetic or proteomic target. Users can also search available datasets for specific signatures of interest. Once a user finds a signature of interest, iLINCS will provide information on all corresponding signatures and present both external tools (e.g., Enrichr) and iLINCS-based applications (e.g., GREIN) that can be used to further analyze and visualize the selected data. Users can click the iLINCS Help menu to review use cases and workflows, which can be used as guidelines and tutorials on how to use iLINCS to analyze various data types.

LINCS Proteomics Analysis with piNET

Jarek Meller and Behrouz Shamsaei, University of Cincinnati

The piNET platform facilitates integrated analysis and visualization of LINCS proteomics data via three main workflows: Peptides2Proteins, PTMs2Modifiers, and Proteins2Pathways. Each workflow identifies piNET data that matches a user's input query, whether it be related to peptide, PTM, or protein signatures., Users can then filter the generated correlation maps according to their interests, and then download high-resolution images of the filtered maps for publication. Dr. Shamsaei summarized the input and output data involved in each workflow, shown below:

- The Peptides2Proteins workflow uses inputted peptides and PTMs to identify connected proteomic signatures.
- The PTM2Modifiers workflow uses PTM signatures to identify correlated enzymatic signatures (which can be analyzed in the future via tools such as PhosphoSitePlus).
- The Protein2Pathways workflow maps protein-level signatures onto associated pathways or perturbation networks to identify significant connections.

Users can initiate piNET analyses with LINCS data on the LINCS Portal or on piNET directly. Descriptions and guidelines detailing how to use piNET can be found within the Help menu; additional information can be found in the original *Nucleic Acids Research* publication on piNET.[1]

[1] Behrouz Shamsaei, Szymon Chojnacki, Marcin Pilarczyk, Mehdi Najafabadi, Wen Niu, Chuming Chen, Karen Ross, Andrea Matlock, Jeremy Muhlich, Somchai Chutipongtanate, Jie Zheng, John Turner, Dušica Vidović, Jake Jaffe, Michael MacCoss, Cathy Wu, Ajay Pillai, Avi Ma'ayan, Stephan Schürer, Michal Kouril, Mario Medvedovic, Jarek Meller, piNET: a versatile web platform for downstream analysis and visualization of proteomics data, Nucleic Acids Research, Volume 48, Issue W1, 02 July 2020, Pages W85–W93, https://doi.org/10.1093/nar/gkaa436

Appendix A: Agenda

Day 1: November 19, 2020

11:00 – 11:10 am **Welcome and Introduction**
Ajay Pillai and Albert Lee (NIH)

11:10 am – 2:00 pm **Session I: Undertaking Large-Scale Perturbation Studies: New Biology and Lessons**
Moderator: Steve Finkbeiner (University of California, San Francisco)

Speakers:

- Peter Sorger (Harvard Medical School, HMS LINCS)
- Ravi Iyengar (Mt. Sinai School of Medicine, DToxS)
- Laura Heiser (Oregon Health & Science University, MEP-LINCS)
- Todd Golub (Broad Transcriptomics)
- Clive Svendsen (Cedars Sinai, NeuroLINCS)
- Maeve Bonner (MIT/Broad Proteomics)
- Dušica Vidović and Mario Medvedović (DCIC)

2:00 – 2:30 pm **BREAK**

2:30 – 4:10 pm **Session II: Impact of LINCS on the Community: Short Community Vignettes**
Moderator: Mario Medvedović (University of Cincinnati, DCIC)

Speakers:

- Rebecca Racz (FDA)
- Oscar Méndez-Lucio (The Janssen Pharmaceutical Companies of Johnson & Johnson)
- Dayne Mayfield (University of Texas at Austin)
- Sikander Hayat (Bayer Pharmaceuticals)

4:10 – 4:25 pm **BREAK**

4:25 – 5:30 pm **Session III: Hands-On Workshop/Poster Session (Parallel Sessions)**

Workshop Presenters:

- Rajiv Narayan, Ted Natali, Anup Jonchhe, and Jacob Asiedu (Broad Transcriptomics)
- Andrea Matlock (Cedars Sinai NeuroLINCS) and Mike MacCoss (University of Washington, Broad Proteomics)
- Caitlin Mills (Harvard Medical School, HMS LINCS)
- Laura Heiser, Sean Gross, and Mark Dane (Oregon Health & Science University, MEP-LINCS)

Poster Presenters:

- Jens Hansen (Mt. Sinai School of Medicine, DToxS)
- Chiara Victor (Harvard Medical School, HMS LINCS)
- Mirra Chung (Harvard Medical School, HMS LINCS)
- Luke Terne (Oregon Health & Science University, MEP-LINCS)

- Ian McLean (Oregon Health & Science University, MEP-LINCS)
- Daniel Clarke (Mt. Sinai School of Medicine, DCIC)
- Amar Koleti (University of Miami, DCIC)
- Tanya Kelley (University of Miami, DCIC)
- Jim Reigel (University of Cincinnati College of Medicine, DCIC)
- Yan Ren (University of Cincinnati College of Medicine, DCIC)
- Andrea Blasco (Broad Transcriptomics)
- Johnny Li (NeuroLINCS)
- Julie Kaye (NeuroLINCS)

Day 2: November 20, 2020

11:00 – 11:10 am **Day 2 Introduction**
Ajay Pillai and Albert Lee (NIH)

11:10 am – 2:00 pm **Session IV: Integrative Data Analysis within Perturbational Studies**
Moderator: Eric Sobie (Mt. Sinai School of Medicine, DToxS)

Speakers:

- Aravind Subramanian (Broad Transcriptomics)
- Christoph Schaniel and Nicole Dubois (Mt. Sinai School of Medicine, DToxS)
- Jim Korkola (Oregon Health & Science University, MEP-LINCS)
- Leslie Thompson (University of California, Irvine, NeuroLINCS)
- Caitlin Mills (Harvard Medical School, HMS LINCS)
- Karen Christianson (Broad Proteomics)
- Avi Ma'ayan (Mt. Sinai School of Medicine, DCIC)

2:00 – 2:30 pm **BREAK**

2:30 – 4:30 pm **Session V: Impact of LINCS on the Community: Short Community Vignettes**
Moderator: Kathleen Jagodnik (Mt. Sinai School of Medicine, DCIC)

Speakers:

- Marc Hafner (Genentech)
- Noemi Andor (Moffitt Cancer Center)
- Irina Budunova (Northwestern University)
- Chiara Giannarelli (Mt. Sinai School of Medicine)
- Nagi Ayad (University of Miami Brain Tumor Initiative)

4:30 – 5:30 pm **Session VI: Hands-On Workshop/Poster Session (Parallel Sessions)**

Workshop Presenters:

- Daniel Clarke, Minji Jeon, and Avi Ma'ayan (Mt. Sinai School of Medicine, DCIC)
- Dhruv Sareen (Cedars Sinai, NeuroLINCS), Arun Sharma (Cedars Sinai, NeuroLINCS), Priyanka Narayan (NIDDK), Nicole Dubois (Mt. Sinai School of Medicine, DToxS), and Mustafa Siddiq (Mt. Sinai School of Medicine (DToxS)
- Vasileios Stathias (University of Miami), Amar Koleti (University of Miami), and Mario Medvedović (University of Cincinnati, DCIC)
- Jarek Meller and Behrouz Shamsaei (University of Cincinnati, DCIC)

Poster Presenters:

- Rayees Rahman (Mt. Sinai School of Medicine, DToxS)
- Maulik Nariya (Harvard Medical School, HMS LINCS)
- Shu Wang (Harvard Medical School, HMS LINCS)
- Karen Christianson (Broad Proteomics)
- Erol Evangelista (Mt. Sinai School of Medicine, DCIC)
- Megan Wojciechowicz (Mt. Sinai School of Medicine, DCIC)
- Alexander Lachmann (Mt. Sinai School of Medicine, DCIC)
- Dušica Vidović (University of Miami, DCIC)
- Ted Natoli (Broad Transcriptomics)
- Rajiv Narayan (Broad Transcriptomics)
- Andrea Matlock (Cedars Sinai, NeuroLINCS)
- Leandro Lima and Julie Kaye (NeuroLINCS)
- Jenny Wu and Ryan Lim (NeuroLINCS)