Cell Systems

Perspective

CellPress

The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations

Alexandra B. Keenan,[1] Sherry L. Jenkins,[1] Kathleen M. Jagodnik,[1] Simon Koplev,[1] Edward He,[1] Denis Torre,[1] Zichen Wang,[1] Anders B. Dohlman,[1] Moshe C. Silverstein,[1] Alexander Lachmann,[1] Maxim V. Kuleshov,[1] Avi Ma'ayan,[1,*] Vasileios Stathias,[2] Raymond Terryn,[2] Daniel Cooper,[2] Michele Forlin,[2] Amar Koleti,[2] Dusica Vidovic,[2] Caty Chung,[2] Stephan C. Schürer,[2] Jouzas Vasiliauskas,[3] Marcin Pilarczyk,[3] Behrouz Shamsaei,[3] Mehdi Fazel,[3] Yan Ren,[3] Wen Niu,[3] Nicholas A. Clark,[3] Shana White,[3] Naim Mahi,[3] Lixia Zhang,[3] Michal Kouril,[3] John F. Reichard,[3] Siva Sivaganesan,[3] Mario Medvedovic,[3] Jaroslaw Meller,[3] Rick J. Koch,[4] Marc R. Birtwistle,[4] Ravi Iyengar,[4] Eric A. Sobie,[4] Evren U. Azeloglu,[4] Julia Kaye,[5] Jeannette Osterloh,[5] Kelly Haston,[5] Jaslin Kalra,[5] Steve Finkbiener,[5] Jonathan Li,[6] Pamela Milani,[6] Miriam Adam,[6] Renan Escalante-Chong,[6] Karen Sachs,[6] Alex Lenail,[6] Divya Ramamoorthy,[6] Ernest Fraenkel,[6]

(Author list continued on next page)

[1]BD2K-LINCS DCIC, Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]BD2K-LINCS DCIC, Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL 33146, USA
[3]BD2K-LINCS DCIC, Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45220, USA
[4]DToxS, Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[5]NeuroLINCS, Gladstone Institute of Neurological Disease and the Departments of Neurology and Physiology, University of California San Francisco, San Francisco, CA 94158, USA
[6]NeuroLINCS, Department of Biological Engineering, MIT, Cambridge, MA 02142, USA
[7]NeuroLINCS, Department of Neuroscience, Johns Hopkins University, Baltimore, MD 21205, USA
[8]NeuroLINCS, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[9]NeuroLINCS, Departments of Psychiatry and Human Behavior and Neurobiology and Behavior, University of California Irvine, Irvine, CA 92697, USA
[10]LINCS PCCSE, The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[11]LINCS Center for Transcriptomics, The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[12]Dana-Farber Cancer Institute, Boston, MA 02215, USA
[13]HMS LINCS Center, Harvard Medical School, Boston, MA 02115, USA
[14]MEP-LINCS Center, Oregon Health & Science University, Portland, OR 97239, USA
[15]MEP-LINCS Center, Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[16]MEP-LINCS Center, Department of Population Sciences, Beckman Research Institute at City of Hope, Duarte, CA 91011, USA

(Affiliations continued on next page)

The Library of Integrated Network-Based Cellular Signatures (LINCS) is an NIH Common Fund program that catalogs how human cells globally respond to chemical, genetic, and disease perturbations. Resources generated by LINCS include experimental and computational methods, visualization tools, molecular and imaging data, and signatures. By assembling an integrated picture of the range of responses of human cells exposed to many perturbations, the LINCS program aims to better understand human disease and to advance the development of new therapies. Perturbations under study include drugs, genetic perturbations, tissue micro-environments, antibodies, and disease-causing mutations. Responses to perturbations are measured by transcript profiling, mass spectrometry, cell imaging, and biochemical methods, among other assays. The LINCS program focuses on cellular physiology shared among tissues and cell types relevant to an array of diseases, including cancer, heart disease, and neurodegenerative disorders. This Perspective describes LINCS technologies, datasets, tools, and approaches to data accessibility and reusability.

Overview of the NIH LINCS Common Fund Program

The Library of Integrated Network-Based Cellular Signatures (LINCS) program aims to create a network-based understanding of human biology by cataloging changes in gene and protein expression, signaling processes, cell morphology, and epigenetic states, which occur when cells are exposed to a variety of perturbing agents. By generating and providing publicly available data on how human cells respond to various genetic and environmental stressors, the LINCS program is collecting the data required for detailed understanding of cell signaling and gene regulatory pathways involved in human disease. This will aid efforts to develop therapies that restore disease-perturbed pathways and networks to their normal physiological state. Several LINCS projects are based on the premise that disrupting

Gavin Daigle,[7] Uzma Hussain,[7] Alyssa Coye,[7] Jeffrey Rothstein,[7] Dhruv Sareen,[8] Loren Ornelas,[8] Maria Banuelos,[8] Berhan Mandefro,[8] Ritchie Ho,[8] Clive N. Svendsen,[8] Ryan G. Lim,[9] Jennifer Stocksdale,[9] Malcolm S. Casale,[9] Terri G. Thompson,[9] Jie Wu,[9] Leslie M. Thompson,[9] Victoria Dardov,[8] Vidya Venkatraman,[8] Andrea Matlock,[8] Jennifer E. Van Eyk,[8] Jacob D. Jaffe,[10] Malvina Papanastasiou,[10] Aravind Subramanian,[11] Todd R. Golub,[11,12,20,21] Sean D. Erickson,[13] Mohammad Fallahi-Sichani,[13] Marc Hafner,[13] Nathanael S. Gray,[13] Jia-Ren Lin,[13] Caitlin E. Mills,[13] Jeremy L. Muhlich,[13] Mario Niepel,[13] Caroline E. Shamu,[13] Elizabeth H. Williams,[13] David Wrobel,[13] Peter K. Sorger,[13] Laura M. Heiser,[14] Joe W. Gray,[14] James E. Korkola,[14] Gordon B. Mills,[15] Mark LaBarge,[16,17] Heidi S. Feiler,[14] Mark A. Dane,[14] Elmar Bucher,[14] Michel Nederlof,[14,18] Damir Sudar,[14,18] Sean Gross,[14] David F. Kilburn,[14] Rebecca Smith,[14] Kaylyn Devlin,[14] Ron Margolis,[19] Leslie Derr,[19] Albert Lee,[19] and Ajay Pillai[19]

[17]MEP-LINCS Center, Center for Cancer Biomarkers Research, University of Bergen, Bergen 5009, Norway
[18]MEP-LINCS Center, Quantitative Imaging Systems LLC, Portland, OR 97239, USA
[19]NIH, Bethesda, MD 20892, USA
[20]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
[21]Harvard Medical School, Boston, MA 02115, USA
*Correspondence: avi.maayan@mssm.edu
https://doi.org/10.1016/j.cels.2017.11.001

any one of the many components of a biological process causes related changes to molecular characteristics and functions of the cell: the observable composite of which is the cellular phenotype. Observing how and when cellular phenotypes are altered by specific stressors can provide clues about underlying mechanisms of disease while facilitating the identification of new therapeutic targets.

LINCS data are made openly available as a community resource through a series of data releases to enable scientists to address a broad range of basic research questions. Results are obtained in cultured and primary cells whose state has been perturbed experimentally, with the term "perturbagen" used to refer to any condition that can alter the cellular state. LINCS datasets therefore consist of assay results from cells treated with bioactive small molecules, antibodies, and ligands, such as growth factors and cytokines, microenvironment proteins, genetic perturbations, and comparisons of disease versus normal primary cells from patients and healthy control subjects. Many different assays are used to measure cell responses, including measurements of mRNA and protein expression; epigenomic status; and cellular, molecular, and morphological phenotypes captured by biochemical and imaging readouts (Figures 1 and 2). Assays typically are carried out on multiple cell types at multiple time points, and perturbagens are applied at multiple doses. The LINCS program has been implemented in two phases. The pilot phase of the program was completed in 2013 and was focused on initial production of perturbation-induced molecular and cellular signatures, assay development, development of data standards, as well as tools and databases for accessing the data. Phase 2, which began in 2014, and is the focus of this Perspective, supports six LINCS Data and Signature Generation Centers (DSGCs) and one Data Coordination and Integration Center (DCIC). The data coordination center synergizes LINCS-related activities with the NIH Big Data to Knowledge (BD2K) program (Margolis et al., 2014). The focus in phase 2 remains on the large-scale production of perturbation-induced molecular and cellular signatures, as well as computational tool development, integrative data analyses, integration of external public datasets with data generated by LINCS, metadata annotation that strictly follows the Findable, Accessible, Interoperable, and Reusable (FAIR) guidelines (Wilkinson et al., 2016), and outreach activities to engage the next generation of biomedical data scientists and promote the LINCS

Consortium. A summary of the LINCS program by the numbers is provided in Figure 1. Box 1 contains web URLs to major LINCS program and project websites. To aid the reader, abbreviations used in this Perspective may be found in Box 2.

While there are many published works using LINCS data, particularly the L1000 small-molecule perturbation dataset (Fallahi-Sichani et al., 2017; Iwata et al., 2017; Mirza et al., 2017; Wang et al., 2016b), a review of the transcriptomic efforts of the LINCS program in the context of connectivity map (CMAP) (Musa et al., 2017), and specifications for LINCS metadata (Vempati et al., 2014), there is no comprehensive description of the LINCS program in the literature to date. Given the pre-publication availability of many of the LINCS datasets, with associated tools designed to make the data accessible to both computational and experimental researchers, LINCS resources as a whole represent a tremendous community asset for scientists from a variety of disciplines. This Perspective is targeted toward computational systems biologists from academia, pharma, and biotech who could use LINCS resources to enrich their own research programs given the wide-ranging applicability of LINCS program resources. Readers of this article should come away with a better understanding of what the NIH LINCS program encompasses, the LINCS program vision, what datasets are available, and where to access them, how these datasets were generated, including the novel technologies developed as part of the LINCS program, and the tools available to aid in the analysis of these datasets.

The LINCS DSGCs
The LINCS Consortium supports six DSGCs and the BD2K-LINCS DCIC. Below we provide a summary of the DSGCs. Each center has distinct experimental approaches and strategies, which, taken together, provide multiple, complementary facets of molecular perturbations.

The Drug Toxicity Signature LINCS Center
The Drug Toxicity Signature (DToxS) Generation LINCS Center at the Icahn School of Medicine at Mount Sinai in New York aims to generate cellular signatures that predict adverse effects of Food and Drug Administration (FDA)-approved drugs, particularly cardiotoxicity, hepatotoxicity, and peripheral neuropathy. These signatures may also be used to predict how co-administration of drug pairs may either exacerbate or mitigate adverse effects. Drugs and drug pairs are selected as perturbagens based on
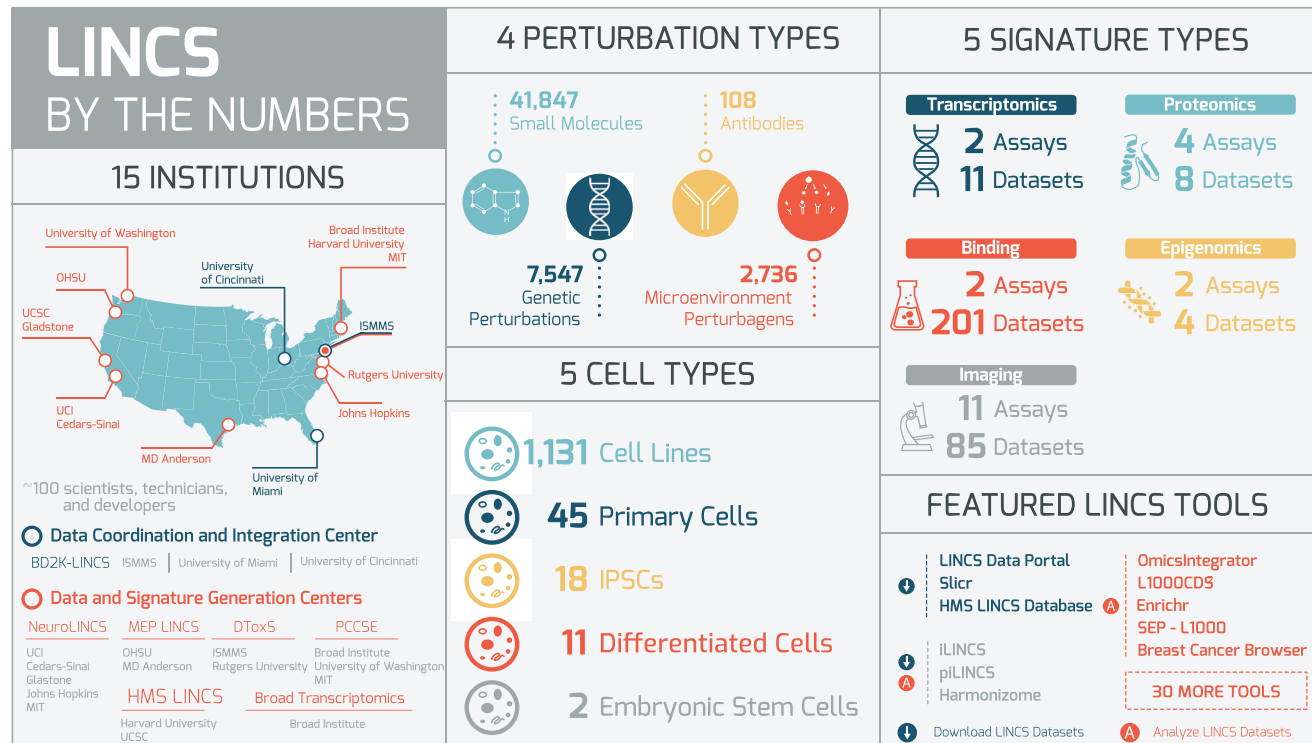
Figure 1. An Overview of the Multi-institutional LINCS Program

adverse events reporting of commonly used FDA-approved drugs. The initial efforts of DToxS are focused on cardiotoxicity caused by cancer therapeutics, particularly heart failure and reduced ventricular function resulting from treatment with kinase inhibitors. The center uses genomic and high-throughput proteomic measurements (with proteomics experiments being conducted at the Center for Advanced Proteomics at Rutgers New Jersey Medical School) and medium-throughput experimental measurements of perturbagen-induced changes in protein state. Multiple levels of data analysis are performed, including: (1) correlating expression changes with clinical risk scores; (2) identifying common protein structural motifs; (3) predicting the biological processes involved in toxicity through network analysis; and (4) developing mechanistic understanding by integrating data with models that describe dynamics. The different levels of analysis can help to identify signatures for both toxicity and toxicity mitigation.

DToxS Unique Featured Assay: Microwestern Array. The microwestern array is a scalable miniaturization of a western blot, performing either 24, 48, or 96 traditional western blots at a time. It is employed for measuring protein levels and post-translational modifications in a medium-throughput format that allows for multiple conditions to be tested in parallel (Ciaccio and Jones, 2017). Similar to the reverse-phase protein array (RPPA) (Tibes et al., 2006), in which whole-cell lysates are applied to nitrocellulose slides, probed with antibodies, and measured via a chemiluminescent reporter, the microwestern has the additional step of electrophoretic protein separation by molecular weight. In brief, cell lysates are spotted onto an acrylamide gel, subjected to electrophoresis, transferred to nitrocellulose, blotted with

antibodies in a gasket apparatus to isolate samples, and then imaged and quantified via infrared fluorescent secondary antibodies.

The Harvard Medical School LINCS Center
The Harvard Medical School (HMS) LINCS Center uses multiple measurement methods to collect data and compute signatures from cells exposed to small-molecule drugs and naturally occurring ligands. The Center emphasizes integration of imaging, mass spectrometry, and transcript profiling approaches, and collects both single-cell and population-average data. The Center also develops data analysis pipelines and informatics systems, which are particularly important in the case of microscopy data for which few such systems exist. The research products of the HMS Center comprise new measurement technologies, multi-dimensional datasets, and open source software.

The HMS Center focuses on kinase inhibitors and epigenome-modifying compounds as perturbagens because they are prototypical of drugs that aim for selective targeting of members of multi-gene families. The great majority of such drugs are active against several targets, and a key goal of the HMS Center is to characterize such poly-selectivity and understand the therapeutic implications. The majority of data in the HMS Center is collected from transformed cell lines, in part because they are simple to grow in large numbers; however human primary cells (e.g., fibroblasts for example) and iPSC-derived, *trans*-differentiated cells (e.g., cardiomyocytes) are also studied. Signatures and tools developed by the HMS Center are designed to help others analyze drug and ligand dose-response relationships, biomarkers of drug resistance, and the origins and consequences of cell-to-cell variability. Signatures and systematic
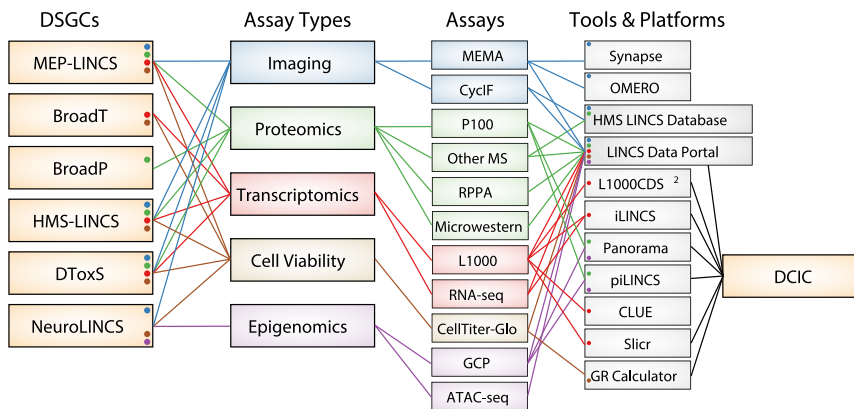
Figure 2. An Overview of the LINCS Centers, Assays, Tools, and Platforms (Not All Tools Are Listed)

MEMA, microenvironment microarrays; CycIF, cyclic immunofluorescence; MS, mass spectrometry; RPPA, reverse-phase protein array; GCP, global chromatin profiling; ATAC, assay for transposase accessible chromatin; OMERO, Open Microscopy Environment; CLUE, CMAP and LINCS Unified Environment; GR, growth response.

perturbagen-response data are frequently combined with follow-up studies in Center publications as a means to illustrate at least one use of LINCS data.

HMS LINCS Unique Featured Assay: Cyclic Immunofluorescence. Developed by the HMS LINCS Center, cyclic immunofluorescence (CycIF) is an open source single-cell imaging method that enables 20- to 30-plex immunofluorescence imaging of cells grown in culture and 60-plex imaging of formalin-fixed, paraffin-embedded tissues. Multiplexing is achieved by successive cycles of four- to six-color staining and imaging followed by chemical inactivation of fluorophores. CycIF uses standard reagents and instrumentation and is simple to implement (Lin et al., 2015, 2016), making it well-suited to high-throughput characterization of perturbagen response. CycIF data can be analyzed using computational algorithms developed for other multiplex singe-cell methods, cytometry by time of flight mass spectrometry for example.

The LINCS Center for Transcriptomics

The Broad Institute's LINCS Center for Transcriptomics and the CMAP team aims to develop a comprehensive resource of cellular state expression signatures spanning millions of genetic and small-molecule perturbations. Genetic perturbations include CRISPR knockout, small hairpin RNA knockdown, and open reading frame overexpression. Each genetic or small-molecule perturbation readout includes approximately 1,000 genes and uses the L1000 assay, a high-throughput, low-cost gene expression profiling platform. These data are publicly available via databases and applications with attention to features that enable meaningful user interaction with the data to foster biological discovery. The CMAP team catalogs the connections between these gene expression signatures to find drugs, diseases, pathways, and targets (Lamb et al., 2006). Systematic determination of cellular effects of application of a small molecule and monitoring of the downstream consequences of perturbing a gene of interest on a massive scale allows the research community to elucidate protein function, determine small-molecule mechanism of action, and dissect biological pathways in physiological and disease states.

LINCS Center for Transcriptomics Unique Featured Assay: L1000 Transcriptomics. Developed by the LINCS Center for Transcriptomics, the L1000 transcriptomics assay directly measures 978 landmark transcripts from crude cell lysates with a ligation-mediated amplification method coupled with Luminex-based detection on 384-well plates (Subramanian et al., 2017). Given the highly correlated structure of gene expression, the non-measured transcriptome can be computationally inferred from these landmark genes, which were selected to enable faithful reconstruction using an algorithm trained on a large set of complete transcriptomes. In addition, 80 invariant genes are measured to allow for scaling and normalization. The method has been shown to be comparable with RNA sequencing (RNA-seq) at a fraction of the cost and has produced a compendium of over a million publicly available cellular perturbation profiles.

The LINCS Proteomic Characterization Center for Signaling and Epigenetics at the Broad Institute

The Broad Institute's LINCS Proteomic Characterization Center for Signaling and Epigenetics (PCCSE) seeks to understand how changes in cellular signaling, transcriptional, and epigenetic states influence one another via feedforward and feedback processes. Perturbations to these states are induced by drug treatments or genetic manipulations (via CRISPR) of cancer cell models, neuronal lineages, or primary vascular cells. The PCCSE uses mass spectrometry-based targeted proteomics to assay cellular phosphosignaling and histone modification responses elicited by these perturbations. Given the strong focus on signaling and epigenetics, the drug and genetic perturbations being selected include a number of kinase inhibitors, epigenetically active drugs, and knockdowns of core chromatin modifying genes. These experiments, coupled with matched L1000 data obtained via collaboration with the LINCS Center for Transcriptomics, are designed to test the hypothesis that early cell signaling responses to perturbation may establish new cellular states by altering epigenetic landscapes.

LINCS PCCSE Unique Featured Assays: P100 and GCP Proteomics. The LINCS PCCSE employs two unique targeted proteomics assays: the P100 and global chromatin profiling (GCP) proteomics platforms. P100 measures the level of 96 phosphopeptides that are commonly observed and modulated in diverse cell types (Abelin et al., 2016), and serves as a "sentinel assay" for the phosphosignaling state of the cell (Soste et al., 2014). By comparing P100 profiles of uncharacterized perturbations with those induced by drugs known to inhibit distinct signaling pathways, this method can infer specific alterations in cellular signaling states. GCP is a similar targeted mass spectrometry-based assay that profiles global histone post-translational modifications in bulk chromatin (Gopal et al., 2016). The probes that monitor combinations of post-translational histone modifications serve to generate epigenetic signatures of genetic or small-molecule disruption of the epigenetic machinery. Data

Cell Systems
Perspective

CellPress

Box 1. URLs

GENERAL PROGRAM INFORMATION SITES

LINCS program informational site: http://www.lincsproject.org
LINCS program on the Common Fund site: https://commonfund.nih.gov/lincs
LINCS program on Twitter: https://twitter.com/lincsprogram
LINCS mobile app: http://lincsproject.org/LINCS/mobile

LINCS DATA AND SIGNATURE CENTERS

DToxS Center: http://dtoxs.org
HMS LINCS Center: http://lincs.hms.harvard.edu/
LINCS Center for Transcriptomics: https://clue.io/
LINCS PCCSE: https://panoramaweb.org/labkey/lincs.url
MEP LINCS Center: https://www.synapse.org/mep_lincs
NeuroLINCS Center: http://www.neurolincs.org/
BD2K-LINCS Data Coordination and Integration Center: http://lincs-dcic.org

---

from both assays are compatible with the CMAP framework developed by the LINCS Center for Transcriptomics to interrogate and integrate large-scale datasets.

The Microenvironment Perturbagen LINCS Center

The Microenvironment Perturbagen (MEP) LINCS Center aims to generate datasets and computational strategies to illuminate how combinatorial signals from the microenvironment (ME) affect intracellular molecular networks and their resultant cellular phenotypes. This is accomplished by profiling cells after treatment with select soluble and insoluble proteins (extracellular matrix proteins, growth factors, and cytokines) identified to be variably enriched in primary tissues. The MEP LINCS Center studies both transformed cell lines as well as normal human primary cells to learn how cell phenotypes change in response to growth with >2,500 different microenvironmental perturbagen combinations. Regulatory relationships are inferred by extracting quantitative ME response phenotype features through multi-color imaging of biomarkers associated with specific cellular phenotypes, measurement of protein expression using RPPA (Tibes et al., 2006), and measurement of gene transcription using the L1000 platform (Subramanian et al., 2017). The development of analytical pipelines and data management systems for large-scale, high-content imaging data is a key component of the MEP LINCS Center, and this effort is coordinated with the HMS LINCS and NeuroLINCS Centers, both of which have large imaging components. Examples of the use of MEP LINCS data include identification of microenvironmental perturbagens that elicit specific phenotypes, understanding how microenvironmental signals modulate response to therapy, and identification of cellular networks associated with particular cellular phenotypes.

MEP LINCS Unique Featured Assay: MEMA Platform. The MEP LINCS Center has developed a microenvironment microarray (MEMA) platform to systematically interrogate ME effects on cellular phenotypes (Lin et al., 2012; Watson et al., 2014). MEMA consists of thousands of unique combinations of soluble and insoluble ME-associated proteins. Insoluble proteins are printed on a solid substrate to form pads upon which cells can be grown while soluble factors are added to the culture medium within each well of a multi-well plate, yielding thousands of unique combinations of MEPs. MEMA includes factors that are secreted by multiple different cell types, including macrophages, infiltrating lymphocytes, stromal fibroblasts, and vascular endothelial cells, although the platform could be customized to assess nearly any soluble or insoluble protein. Following incubation, cells are fixed, stained for endpoints targeting morphology, metabolism, cell cycle, nuclear activity, and differentiation status, and subjected to high-content imaging.

The NeuroLINCS Center

The NeuroLINCS Center currently studies the motor neuron diseases amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (SMA) by characterizing the molecular networks within patient-derived induced pluripotent stem cells (iPSCs) and their differentiated motor neuron progeny. However, it is expanding to develop a global cortical neuron assay from iPSCs that will be of interest to groups working in other neurodegenerative diseases such as Alzheimer's disease. Assays used include high-throughput imaging, RNA-seq, ATAC-seq, and SWATH mass spectrometry to inspect differences in the cellular physiological, transcriptomic, epigenetic, and proteomic landscapes, respectively, of ALS and SMA patient neurons compared with unaffected controls. Perturbations include ALS- and SMA-relevant mutations in patient-derived iPSCs as well as additional chemical perturbations. NeuroLINCS is a collaborative effort between multiple research groups at University of California, Irvine, Cedars-Sinai Medical Center, the Gladstone Institute, MIT, and Johns Hopkins University. It combines expertise in iPSC technology, disease modeling, transcriptomics, epigenomics, metabolomics, proteomics, whole-genome sequencing, cell-based assays, bioinformatics, statistics, and computational biology. A unique feature of NeuroLINCS is the bioinformatics integration of the diverse high-throughput "omic" datasets to provide a network-based understanding of underlying pathways. In addition, the NeuroLINCS Center is collaborating with Google to integrate signatures across platforms into highly predictive models of responses to perturbagens using machine learning.

NeuroLINCS Unique Featured Assay: Automated Imaging System to Track Live Neurons. The NeuroLINCS Center uses automated robotic microscopy to individually track live neurons to

Box 2. Abbreviations

Abbreviations and Acronyms Glossary

ALS	amyotrophic lateral sclerosis
API	application program interface
BD2K	Big Data to Knowledge
CCA	Consortium Coordination and Administration
CEDAR	Center for Expanded Data Annotation and Retrieval
CLUE	CMAP and LINCS unified environment
CMAP	connectivity map
CTO	Community Training and Outreach
CycIF	cyclic immunofluorescence
CyTOF	cytometry by time of flight
DCIC	Data Coordination and Integration Center
DSGC	Data and Signature Generation Center
DSR	Data Science Research
DToxS	Drug Toxicity Signatures
DWG	Data Working Group
ECM	extracellular matrix
FAIR	Findable, Accessible, Interoperable, and Reusable
FFPE	formalin-fixed, paraffin-embedded
FI	Fluorescence Imaging
GCP	global chromatin profiling
GR	growth rate inhibition
HMS	Harvard Medical School
ICV	Integrated Connectivity Viewer
IKE	Integrated Knowledge Environment
L1000CDS2	L1000 Characteristic Direction Search Engine
LDP	LINCS Data Portal
LINCS	Library of Integrated Cellular Signatures
LPDP	LINCS Proteomics Data Portal
ME	microenvironment
MEMA	microenvironment microarray
MEP	Microenvironment Perturbagen
MIBBI	Minimum Information for Biological and Biomedical Investigations
MIC	Microscopy Imaging Commons
MOOC	Massive Online Open Course
MWA	microwestern array
OMERO	Open Microscopy Environment
PCCSE	Proteomic Characterization Center for Signaling and Epigenetics
QC	quality control
RPPA	reverse-phase protein array
SMA	spinal muscle atrophy

relate physiological changes in the cell over time to the fate of the cell (Arrasate and Finkbeiner, 2005; Finkbeiner et al., 2015). This high-throughput, high-content imaging technique allows for the

generation of predictive dynamic models of cell fate. The models function as LINCS signatures used to understand disease pathways and to help understand the relationship between -omics signatures and cellular phenotypes.

The BD2K-LINCS DCIC
The BD2K DCIC for LINCS consists of four major components: Integrated Knowledge Environment (IKE), Data Science Research (DSR), Community Training and Outreach (CTO), and Consortium Coordination and Administration. The IKE is enabling federated access, intuitive querying, and integrative analysis and visualization across all LINCS resources, and many additional external data types from other relevant resources. The IKE resources are built on the infrastructure, analysis tools, and data that were established in the LINCS pilot phase. For the DSR component, the DCIC is managing several internal research projects and supports several external data science research projects, addressing various data integration and intracellular molecular regulatory network challenges. The CTO efforts of the DCIC have established several educational resources, including a LINCS massive online open course (MOOC) on Coursera (https://www.coursera.org/learn/bd2k-lincs), a new PhD track in Big Data Biostatistics at the University of Cincinnati, and an intensive 10-week summer research training program for graduate and undergraduate students. In addition, the DCIC is initiating and supporting diverse collaborative projects that leverage LINCS resources and disseminate LINCS data and tools. One effort central within LINCS is to standardize workflows and pipelines to ensure that they can be version controlled, sharable, and evaluated. This is achieved by publishing pipelines as Notebooks (Shen, 2014; Wang and Ma'ayan, 2016), and containerizing pipelines using platforms such as Docker (Merkel, 2014). The Center also aims to develop and deploy a next-generation computational infrastructure and novel analysis tools and methods that enable researchers to glean new insights from integrative models of biological systems, while linking complex diseases/phenotypes with drugs and the pathways targeted by those drugs in different cells and tissues.

LINCS Program Websites, Portals, and Databases
Sites that Cover Content from the Entire LINCS Consortium
lincsproject.org: Informational Site about the LINCS Program. The entry point to access data and information about the LINCS program is http://www.lincsproject.org, a central hub for both the research community and general public. This website, along with the LINCS Data Portal (LDP), contains details about the assays, cell types, and perturbagens that are currently part of the library, as well as the LINCS DSGCs and DCIC, LINCS-related publications, news, events, video tutorials, workflows, and software that can be used for analyzing LINCS data. The LINCS mobile app, available on iTunes and Google Play, is a streamlined version of the lincsproject.org website.
The LDP: Consolidated Access to LINCS Data and Metadata. The LDP, available at http://lincsportal.ccs.miami.edu/, provide comprehensive access to LINCS data, including transcriptomics, binding, imaging, proteomics, and epigenomics datasets. Users can browse available datasets by experimental method, type of data collected, LINCS Center name, individual projects,

Cell Systems
Perspective

CellPress

Table 1. LINCS Data and Signature Generation Centers Assays and Datasets

Data and Signature Generation Center	Assay	Cell Types	Perturbations	Signatures	Data Points	Data Level	Latest Release[a]
HMS LINCS Center	KINOMEscan	0	147 small molecule	67,000	82,000	2,3	April 2015
	KiNativ	8	29 small molecule	50	8,000	2,3	Jan 2016
	ELISA protein state	39	15 growth factor	1,200	9,000	3	Sep 2013
	bead-based immunoassay	7	12 small molecule, 10 ligand	230	66,000	3	Aug 2015
	RPPA protein state	10	5 small molecule	140	720,000	2,3,4	April 2015
	FI cell count	54	159 small molecule	5,000	150,000	2,3,4	Dec 2015
	FI cell viability	7	17 small molecule	700	3.4×10^7	3	May 2017
	FI apoptosis	112	167 small molecule	200	145,000	1,2,3,4	July 2015
	FI protein state	42	8 small molecule, 5 ligand	2,400	128,000	2,3	Sep 2015
	FI cell-cycle state	18	9 small molecule	1,200	152,000	1,2	July 2011
	FI morphology	6	2 small molecule	36	3,000	3	Dec 2015
DToxS Center	RNA-seq	4	44 small molecule, 1 antibody	120	9.2×10^6	1,2,3	June 2017
	mass spectrometry	4	15 small molecule	20	155,000	1,2,3	June 2017
LINCS Center for Transcriptomics	L1000	53	25,581 small molecule, 978 genetic	476,000	1.3×10^9	1,2,3,4	Dec 2015
LINCS PCCSE	P100	8	105 small molecule	700	200,000	3,4	July 2016
	GCP	7	90 small molecule	475	84,000	3	Aug 2016
MEP LINCS	MEMA	3	2,945 microenvironment	27,000	3.4×10^6	1,2,3,4	March 2016
NeuroLINCS	RNA-seq	20	disease background	6	3.2×10^6	3,4	May 2017
	SWATH-MS	19	disease background	6	197,000	1,2,3,4	Sep 2016
	ATAC-seq	18	disease background	6	2.0×10^7	2,3,4	May 2017

FI, Fluorescence imaging.
[a]As of September 1, 2017.

and relevant biological processes. Users can also conduct searches, for example, by cell line, small molecule, gene, or protein of interest. All data are publicly available and no account is required for access. Datasets may be added to a checkout basket to be later consolidated for a downloadable package in a compressed file containing both the data and the relevant metadata. Each LINCS dataset has its own landing page with metadata details, analysis methods, and a data level designation. Data level designations range from 1 to 4, with raw data assigned level 1, and fully processed signatures are level 4. The data level concept was adopted from The Cancer Genome Atlas project (Akbani et al., 2014), with data level definitions adjusted and customized for data originating from different assays. Table 1 summarizes LINCS DSGC assays and datasets available on the LDP as of September 1, 2017.

DSGCs Official Center Websites
In addition to accessing LINCS data and tools through lincsproject.org and the LDP, each center has an official website providing additional center-specific information. Centers have made tools available on these sites to allow users to uniquely explore the data through web-based applications, command line interfaces, and application program interfaces (APIs).

Clue.io: Platform to Interface with the L1000, P100, and GCP Data. CMAP and LINCS unified environment (CLUE), accessible at http://clue.io, is a computational environment designed to be executed on the cloud with pre-loaded gene expression profiles and perturbation datasets, analytical tools, and web applications. CLUE serves data generated, and applications developed, by the two LINCS Broad DSGCs. The aim of CLUE is to facilitate engagement with the highly dimensional L1000, P100, and GCP data by users who lack a substantial computational background. Access to this data is provided by user-friendly web-based applications, as well as command line interfaces and APIs. Free accounts are available to academic and not-for-profit users.

HMS LINCS Database: Access to Data and Tools Produced by the HMS LINCS Center. The HMS LINCS Website, and its associated database (http://lincs.hms.harvard.edu/), provide a public point of access to all data generated by the HMS LINCS Center, as well as experimental protocols and computational tools developed by the Center. Users can search by cellular context, perturbagens, proteins, antibodies, and other reagents, as well as by PubMed identification number, for datasets relevant to specific HMS LINCS publications. For selected datasets, specialized browsers and visualization tools are available. Programmatic access to all data is possible via an API.

MEP LINCS Synapse Site: Access to Workflows, Tools, and Analysis from the MEP LINCS Center. The MEP LINCS Synapse site (https://www.synapse.org/mep_lincs) offers access to raw and processed imaging data, molecular profiling data, data descriptions, assay overviews, and protocols. Interactive dashboards are available for several profiled cancer cell lines, as well as interactive reports for each staining set and cell line combination. There is also an MEP data explorer that allows

Table 2. Selected LINCS Software Tools

Tool	Summary	Website
iLINCS	transcriptomics and proteomics dataset analysis tools for a biologist audience	http://www.ilincs.org/
piLINCS	facilitates proteomics data access, integrated with iLINCS	http://www.pilincs.org
GR Calculator, GR Metrics, GR Browser	access and visualize LINCS dose-response data, upload user data for analysis	http://www.grcalculator.org/grbrowser/
LINCS Proteomics Data Portal	proteomics data access and analysis, integrated with iLINCS	http://lincsproteomics.org
L1000CDS2	LINCS perturbation search engine that returns signatures that either reverse or mimic a user-input signature	http://amp.pharm.mssm.edu/L1000CDS2/
HMS Breast Cancer Browser	access to LINCS data relevant to breast cancer	http://www.cancerbrowser.org/
CLUE Platform	a collection of apps that query, visualize, analyze LINCS transcriptomics and proteomics data	https://clue.io/
OMICS Integrator	identify candidate molecular pathways underlying integrated omics data	http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/
Enrichr	gene set search engine	http://amp.pharm.mssm.edu/Enrichr
Harmonizome	omics integration tool	http://amp.pharm.mssm.edu/Harmonizome/
Slicr	L1000 signature store	http://amp.pharm.mssm.edu/Slicr
LINCS Canvas Browser	L1000 visualization	http://www.maayanlab.net/LINCS/LCB/
SEP-L1000	side effect predictions for LINCS small molecules	http://maayanlab.net/SEP-L1000

for box and scatterplot visualizations for experiments involving each cell line. A free account is required to access data and tools.

The DToxS Web Portal: Access to Data and Tools Produced by the DToxS LINCS Center. The DToxS web portal (http://dtoxs.org) provides access to experimental datasets by data type (including both transcriptomic and proteomic), and allows users to browse metadata by cell lines and drugs, which are classified as either offending drugs or toxicity-mitigating drugs. Users can also access the DToxS Center standard operating procedures for cell culture, assays, and computational analysis. Registration is required and is open to all users.

The NeuroLINCS Center Site: People, Technologies, Publications, Data, and Tools. The NeuroLINCS informational website (http://www.neurolincs.org/) provides an overview of the entire NeuroLINCS pipeline, which starts with the development of iPSC cell lines from patients and the propagation for analysis by the various NeuroLINCS Center laboratories. The site features the investigators, the unique technologies the center employs, publications, and links to datasets and tools. Instructions for principle investigators wishing to request access to NeuroLINCS data not publicly released through the LDP may be found through this site.

The LINCS Panorama Repository: Access to Targeted Proteomics Data from the PCCSE. The LINCS Panorama Repository (https://panoramaweb.org/labkey/lincs.url) allows users to access P100 and GCP profiles in matrix form and download custom subsets of profile data (i.e., by drug[s], cell type[s], etc.). Users may also access the primary mass spectrometry data in the form of Skyline documents, the standard in the field for analysis of targeted proteomics. Panorama serves as PCCSE's cloud computing infrastructure, and complete data quality control and normalization pipeline for targeted proteomics data generated by PCCSE is executed programmatically in Panorama to ensure uniform data treatment.

LINCS Tools and Workflows
The LINCS Consortium has developed software tools and analysis platforms to facilitate interaction with LINCS data, including visualization and analysis of LINCS data across LINCS data types, and in the context of other data (Table 2). Below we highlight some of the LINCS tools and platforms. Each of these tools and platforms may also be accessed through lincsproject.org.

iLINCS: An Integrated System to Analyze LINCS and Other Data
iLINCS (http://www.ilincs.org/) is a web portal aimed at providing LINCS transcriptomic and proteomic dataset analysis tools to a biologist audience. Users are guided through four pipelines: Genes, Datasets, Signatures, and Maps. The Genes pipeline allows users to input a gene set of interest, and then to select a LINCS or non-LINCS dataset and signatures to query, analyze, and visualize. The Datasets pipeline permits users to construct a signature from a selected dataset, perform functional enrichment and pathway analysis on that signature, and search for concordant (or discordant) LINCS and non-LINCS signatures. The Signatures pipeline also allows users to upload or select a signature, identify similar signatures, and analyze their common features, including the biological pathways that may underly groups of similar signatures. The Maps pipeline permits users to select a 2D map of two signature libraries, mine relationships among these signatures, and analyze the features of similar signatures.

piLINCS: An Integrated System to Analyze LINCS Proteomics Data
piLINCS (http://www.pilincs.org) provides both interactive and API access to the proteomics data generated by the LINCS program, including P100 and GCP profiles, which are automatically imported from Panorama. Users can filter profiles based on assay, cell type, perturbation, and dose; merge multiple profiles; download datasets in multiple formats; and send selected profiles to iLINCS for further analysis. In addition to processed

Cell Systems
Perspective

protein profiles, raw data are also available with links to the associated Panorama chromatograms.

GR Metrics, GR Calculator, and GR Browser: Visualize and Calculate Dose-Dependent Sensitivities of Cancer Cell Lines

The GR Calculator (http://www.grcalculator.org/grbrowser/) allows users to browse, visualize, and download LINCS dose-response datasets and to upload their own data for analysis. The GR Browser supports analysis of dose-response data using growth rate inhibition (GR) metrics, which correct for the effects of variation in growth rate in the evaluation of drug response (Hafner et al., 2016, 2017a, 2017b; Niepel et al., 2017). Growth rate-corrected response metrics such as GR_{50}, GR_{max}, and GR_{AOC} are more robust and biologically informative than traditional metrics such as IC_{50} and E_{max}, and are used to evaluate drug response in the LINCS MCF10A Common Project, the MEP-HMS LINCS Joint Project, the HMS LINCS Seeding Density Project, and the Broad-HMS LINCS Joint Project datasets (Hafner et al., 2016). GR and traditional response metrics can be visualized at the GR Calculator as scatterplots and dose-response grids sorted for drug or cell types. LINCS dose-response data can also be downloaded by users and GR tools are available as stand-alone algorithms for integration into BioConductor and other offline pipeline applications.

LINCS Proteomics Data Portal

LINCS Proteomics Data Portal (http://lincsproteomics.org) has been developed by DCIC in coordination with the LINCS proteomics community to facilitate exploration of LINCS proteomic data and related resources. The portal facilitates finding and downloading proteomic data, provides tailored "data-centric" and "assay (or annotation)-centric" views, and integrates tailored tools, e.g., piLINCS and piNET, to facilitate navigation, search, and interpretation of LINCS proteomics data.

L1000CDS2: LINCS L1000 Characteristic Direction Signatures Search Engine

L1000CDS2 (http://amp.pharm.mssm.edu/L1000CDS2/) is a search engine that returns LINCS L1000 drug perturbation signatures that either mimic or reverse a user-input signature of gene symbols (Duan et al., 2016). The search is completed against a subset of the L1000 data in which the differentially expressed genes were calculated using the characteristic direction method (Clark et al., 2014). The user either enters a list of gene symbols that are up- and downregulated, or a list of gene symbols and associated expression values. The search engine then returns the top 50 perturbation conditions that either mimic or reverse the input.

HMS Breast Cancer Browser

The HMS LINCS Breast Cancer Browser (http://www.cancerbrowser.org/) is a web interface that provides users access to LINCS data pertinent to breast cancer biology and breast cancer drug response. Users may easily filter datasets based on cell line, receptor status, molecular subtype, and mutation status. The user may also filter datasets based on perturbagen development phase, target gene, target gene class, target pathway, and target biological function. Available datasets include growth factor-induced pAKT/pERK response assays, drug dose-response profiles, and basal RTK phosphorylation profiles, as well as total protein mass spectrometry and phosphoprotein mass spectrometry datasets.

CLUE Platform: An Integrated System to Analyze LINCS Data Collected by the Broad Institute DSGCs

The Clue.io platform (https://clue.io/) includes Query app, Touchstone app, Repurposing app, Morpheus app, and Integrated Connectivity Viewer (ICV) app. The Touchstone app provides access to approximately 5,000 genetic and small-molecule perturbagens that are well characterized. The Touchstone dataset can thus serve as a benchmark for exploring connectivities among perturbagens. The ICV app visualizes connectivity data as an interactive heatmap so that users can intuitively explore relationships within the data. The Morpheus app provides an additional level of interaction to ICV by allowing users to manipulate and annotate a user-provided or existing dataset. The Repurposing app is a drug-repurposing tool that accesses the Broad Institute's collection of over 5,000 compounds that have approved clinical indications and known safety profiles. Users can filter these compounds by mechanism of action, target, disease area, clinical phase, and vendor. The Query app enables users to find positive and negative connections between a user gene signature and all signatures in CMAP.

Other Tools

Other tools and interfaces that have been developed by the consortium or customized for LINCS data include OMICSIntegrator (http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/), which can be used to identify putative molecular pathways underlying integrated omics data (Tuncbag et al., 2016); Enrichr (http://amp.pharm.mssm.edu/Enrichr), which is a search engine for gene sets (Chen et al., 2013; Kuleshov et al., 2016); and Harmonizome (http://amp.pharm.mssm.edu/Harmonizome/), a database that serves integrative omics data from 66 resources, including LINCS (Rouillard et al., 2016). Slicr (http://amp.pharm.mssm.edu/Slicr), is a signature store for L1000 signatures, whereas LINCS Canvas Browser (http://www.maayanlab.net/LINCS/LCB/) visualizes L1000 on canvases that cluster signatures based on their expression vector similarity (Duan et al., 2014). Additional LINCS-specific tools include SEP-L1000 (http://maayanlab.net/SEP-L1000), an interface for predictions of side effects for >20,000 LINCS-profiled small molecules (Wang et al., 2016a); Panorama, a web-based platform for storing, sharing, and analyzing proteomics data (Sharma et al., 2014) analyzed by Skyline (MacLean et al., 2010), a Windows-based client commonly used for processing data from proteomics experiments; and the Open Microscopy Environment (OMERO), an image data management platform that allows users to view, organize, analyze, and share imaging data securely with collaborators using a variety of different permission levels (Allan et al., 2012). The OMERO platform has been adopted by the LINCS Consortium to manage the images that are collected at various DSGCs; and several members of the consortium are working on developing a Microscopy Imaging Commons, an open environment to store and share microscopy images.

LINCS Data Findability, Accessibility, Interoperability, Reproducibility, and Reusability
Metadata and Data Standards

The LINCS Data Working Group (DWG) has established metadata standards for LINCS reagents, assays, and experiments to ensure cross-consortium data compatibility. Annotations for perturbagens, such as small molecules, small interfering RNAs,

growth factors and other ligands, cells, and some elements of experimental metadata, are standardized across all LINCS DSGCs. Harmonization of these annotations facilitates data analysis, data formatting, and data visualization strategies being used by the LINCS community. The standardized metadata is also important for the development of databases and data repositories that store and share LINCS data. Led by the DCIC, and in collaboration with the Center for Expanded Data Annotation and Retrieval (CEDAR) (Musen et al., 2015), the consortium is developing automated mechanisms to capture metadata when it is produced through customized web forms. Furthermore, the LINCS DWG developed standards to annotate all LINCS assay protocols, data analysis strategies, and datasets; these standards are deposited into BioSharing (McQuilton et al., 2016), a repository for metadata standards. The DWG efforts are aligned with standards developed by other groups, such as the Investigation/Study/Assay ISA infrastructure project (Rocca-Serra et al., 2010), Minimum Information for Biological and Biomedical Investigations efforts (Taylor et al., 2008), and NCBI PubChem (Bolton et al., 2008).

Data Processing and Analysis Pipelines
In an effort to make the process of generating LINCS data completely transparent, DCIC is leading an effort in validating, documenting, and releasing all computational processing pipelines used by DSGSs to process and analyze LINCS data. Two main objectives of the effort are to (1) facilitate de novo reconstruction of computational pipelines by completely describing all processing steps; and (2) facilitate the re-use of computational pipelines developed and used by DSGCs. To facilitate *de novo* reconstruction, DCIC is leading an effort to provide assay-specific standards for documenting all details of data processing and analysis. To facilitate the re-use of existing pipelines, DCIC works closely with DSGCs to test and validate all computational pipelines. Whenever possible, DCIC is also creating and releasing ready-to-use Docker containers implementing individual processing and analysis pipelines.

Data Availability
In general, all LINCS data are available to all, pre-publication, with no restrictions. The only exception is the raw RNA-seq data collected from patients, which will be made available for download from dbGAP to adhere to patient privacy protocols. Some websites developed by the data-generation centers, such as Clue.io, which is hosting the L1000 data and related tools, the DToxS data portal, or Synapse, which is hosting the MEP LINCS data and tools, require login and an account. However, versions of the data are made openly available by the DCIC, and through public repositories such as GEO. The LINCS Consortium established a data release policy which is available at: http://lincsproject.org/LINCS/data/release-policy.

The MCF10A Common Project
As the biomedical research community is greatly concerned with the reproducibility of molecular and cell biology studies (Begley, 2013), the LINCS investigators jointly work on cross-center projects. The consortium is currently working toward identifying the molecular networks that determine how MCF10A cells receive and integrate external signals in ways that influence MCF10A physiology, such as proliferation, differentiation, and motility. This is a dynamic process that can best be approached by integrating information from the unique assays, which are con-

ducted by several LINCS DSGCs. It is likely that responses involve transcriptional and proteomic network reconfigurations in the first few hours, followed by epigenomic changes that are induced in hours to days. The consortium has decided to test responses of eight diverse drugs (those that are known to target different canonical pathways) and apply these in different concentrations while measuring cell viability, mRNA levels, protein levels, epigenetic markers, and cell morphology under different concentrations and time points. Understanding the extent of agreement among these assays in perturbing canonical pathways will allow the consortium to evaluate the extent to which the assays and data across centers allows for reproducible and cohesive findings.

Conclusions
The six LINCS DSGCs are producing unique datasets and resources using state-of-the-art high-throughput transcriptomics, proteomics, epigenomics, biochemical assays, and imaging data. These data are collected by the unique assays such as L1000 and RNA-seq for the transcriptomics; P100, GCP, RPPA, microwestern, and other expression proteomics techniques; and MEP, CycIF, and other methods for cell phenotypic and morphological imaging. These data are organized, analyzed, visualized, and integrated jointly by the DSGCs and the DCIC, who actively develop metadata standards, tools, and portals (Figure 2). Along these lines the LINCS Consortium is centrally involved in making LINCS datasets and tools adhere to the FAIR principles (Wilkinson et al., 2016). The consortium is working on developing specific evaluations for LINCS datasets and tools so those are maximally FAIR. The overall aim of LINCS is to produce a long-term resource that will assist individual investigators to form novel hypotheses about the inner workings of human cells and tissues in normal physiology and in disease.

REFERENCES

Abelin, J.G., Patel, J., Lu, X., Feeney, C.M., Fagbami, L., Creech, A.L., Hu, R., Lam, D., Davison, D., and Pino, L. (2016). Reduced-representation phospho-signatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. Mol. Cell. Proteomics *15*, 1622–1641.

Akbani, R., Ng, P.K.S., Werner, H.M., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J.-Y., Yoshihara, K., and Li, J. (2014). A pan-cancer proteomic perspective on the cancer genome Atlas. Nat. Commun. *5*, 3887.

Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., MacDonald, D., Moore, W.J., Neves, C., and Patterson, A. (2012). OMERO: flexible, model-driven data management for experimental biology. Nat. Methods *9*, 245–253.

Arrasate, M., and Finkbeiner, S. (2005). Automated microscope system for determining factors that predict neuronal fate. Proc. Natl. Acad. Sci. USA *102*, 3840–3845.

Begley, C.G. (2013). Reproducibility: six red flags for suspect work. Nature *497*, 433–434.

Cell Systems
Perspective

Bolton, E.E., Wang, Y., Thiessen, P.A., and Bryant, S.H. (2008). PubChem: integrated platform of small molecules and biological activities. Annu. Rep. Comput. Chem. 4, 217–241.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.

Ciaccio, M.F., and Jones, R.B. (2017). Microwestern arrays for systems-level analysis of SH2 domain-containing proteins. Methods Mol. Biol. 1555, 453–473.

Clark, N.R., Hu, K.S., Feldmann, A.S., Kou, Y., Chen, E.Y., Duan, Q., and Ma'ayan, A. (2014). The characteristic direction: a geometrical approach to identify differentially expressed genes. BMC Bioinformatics 15, 79.

Duan, Q., Flynn, C., Niepel, M., Hafner, M., Muhlich, J.L., Fernandez, N.F., Rouillard, A.D., Tan, C.M., Chen, E.Y., Golub, T.R., et al. (2014). LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. Nucleic Acids Res. 42, W449–W460.

Duan, Q., Reid, S.P., Clark, N.R., Wang, Z., Fernandez, N.F., Rouillard, A.D., Readhead, B., Hodos, R., Tritsch, S., Hafner, M., et al. (2016). L1000CDS2: LINCS L1000 characteristic direction signatures search engine. NPJ Syst. Biol. Appl. 2, 16015.

Fallahi-Sichani, M., Becker, V., Izar, B., Baker, G.J., Lin, J.R., Boswell, S.A., Shah, P., Rotem, A., Garraway, L.A., and Sorger, P.K. (2017). Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. Mol. Syst. Biol. 13, 905.

Finkbeiner, S.M., Ando, D.M., and Daub, A.C.. (2015). Automated robotic microscopy systems. US patent US20150278625, filed June 11, 2015, and published October 1, 2015.

Gopal, S., Creech, A., Officer, A., Egri, S., Davison, D., Jaffe, J.D., and Jaffe, I.Z. (2016). The impact of chemotherapeutic agents on signaling and epigenetics in vascular endothelial cells. Blood 128, 1353.

Hafner, M., Niepel, M., Chung, M., and Sorger, P.K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods 13, 521–527.

Hafner, M., Niepel, M., and Sorger, P.K. (2017a). Alternative drug sensitivity metrics improve preclinical cancer pharmacogenomics. Nat. Biotechnol. 35, 500–502.

Hafner, M., Niepel, M., Subramanian, K., and Sorger, P.K. (2017b). Designing drug-response experiments and quantifying their results. Curr. Protoc. Chem. Biol. 9, 96–116.

Iwata, M., Sawada, R., Iwata, H., Kotera, M., and Yamanishi, Y. (2017). Elucidating the modes of action for bioactive compounds in a cell-specific manner by large-scale chemically-induced transcriptomics. Sci. Rep. 7, 40164.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K.N., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935.

Lin, C.-H., Lee, J.K., and LaBarge, M.A. (2012). Fabrication and use of microenvironment microarrays (MEArrays). J. Vis. Exp. https://doi.org/10.3791/4152.

Lin, J.-R., Fallahi-Sichani, M., and Sorger, P.K. (2015). Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. Nat. Commun. 6, 8390.

Lin, J.R., Fallahi-Sichani, M., Chen, J.Y., and Sorger, P.K. (2016). Cyclic immunofluorescence (CycIF), a highly multiplexed method for single-cell imaging. Curr. Protoc. Chem. Biol. 8, 251–264.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968.

Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., Guyer, M., and Green, E.D. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. J. Am. Med. Inform. Assoc. 21, 957–958.

McQuilton, P., Gonzalez-Beltran, A., Rocca-Serra, P., Thurston, M., Lister, A., Maguire, E., and Sansone, S.-A. (2016). BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. Database (Oxford) 2016, https://doi.org/10.1093/database/baw075.

Merkel, D. (2014). Docker: lightweight Linux containers for consistent development and deployment. Linux J. 2014, 2.

Mirza, N., Sills, G.J., Pirmohamed, M., and Marson, A.G. (2017). Identifying new antiepileptic drugs through genomics-based drug repurposing. Hum. Mol. Genet. 26, 527–537.

Musa, A., Ghoraie, L.S., Zhang, S.D., Glazko, G., Yli-Harja, O., Dehmer, M., Haibe-Kains, B., and Emmert-Streib, F. (2017). A review of connectivity map and computational approaches in pharmacogenomics. Brief Bioinform. https://doi.org/10.1093/bib/bbw112.

Musen, M.A., Bean, C.A., Cheung, K.-H., Dumontier, M., Durante, K.A., Gevaert, O., Gonzalez-Beltran, A., Khatri, P., Kleinstein, S.H., and O'Connor, M.J. (2015). The center for expanded data annotation and retrieval. J. Am. Med. Inform. Assoc. 22, 1148–1152.

Niepel, M., Hafner, M., Chung, M., and Sorger, P.K. (2017). Measuring cancer drug sensitivity and resistance in cultured cells. Curr. Protoc. Chem. Biol. 9, 55–74.

Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., and Hofmann, O. (2010). ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics 26, 2354–2356.

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) 2016, https://doi.org/10.1093/database/baw100.

Sharma, V., Eckels, J., Taylor, G.K., Shulman, N.J., Stergachis, A.B., Joyner, S.A., Yan, P., Whiteaker, J.R., Halusa, G.N., and Schilling, B. (2014). Panorama: a targeted proteomics knowledge base. J. Proteome Res. 13, 4205–4210.

Shen, H. (2014). Interactive notebooks: sharing the code. Nature 515, 151.

Soste, M., Hrabakova, R., Wanka, S., Melnik, A., Boersema, P., Maiolica, A., Wernas, T., Tognetti, M., von Mering, C., and Picotti, P. (2014). A sentinel protein assay for simultaneously quantifying cellular processes. Nat. Methods 11, 1045–1048.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., and Asiedu, J.K. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. bioRxiv. https://doi.org/10.1101/136168.

Taylor, C.F., Field, D., Sansone, S.-A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C.A., Binz, P.-A., Bogue, M., and Booth, T. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat. Biotechnol. 26, 889–896.

Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G.B., and Kornblau, S.M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol. Cancer Ther. 5, 2512–2521.

Tuncbag, N., Gosline, S.J., Kedaigle, A., Soltis, A.R., Gitter, A., and Fraenkel, E. (2016). Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. PLoS Comput. Biol. 12, e1004879.

Vempati, U.D., Chung, C., Mader, C., Koleti, A., Datar, N., Vidovic, D., Wrobel, D., Erickson, S., Muhlich, J.L., Berriz, G., et al. (2014). Metadata standard and data exchange specifications to describe, model, and integrate complex and diverse high-throughput screening data from the library of integrated network-based cellular signatures (LINCS). J. Biomol. Screen. 19, 803–816.

Wang, Z., Clark, N.R., and Ma'ayan, A. (2016a). Drug-induced adverse events prediction with the LINCS L1000 data. Bioinformatics 32, 2338–2345.

Wang, Z., and Ma'ayan, A. (2016). An open RNA-Seq data analysis pipeline tutorial with an example of reprocessing data from a recent Zika virus study. F1000Res. *5*, 1574.

Wang, Z., Monteiro, C.D., Jagodnik, K.M., Fernandez, N.F., Gundersen, G.W., Rouillard, A.D., Jenkins, S.L., Feldmann, A.S., Hu, K.S., McDermott, M.G., et al. (2016b). Extraction and analysis of signatures from the gene expression omnibus by the crowd. Nat. Commun. *7*, 12846.

Watson, S., Korkola, J., Rantala, J., and Gray, J. (2014). Interrogating HER2+ plasticity and lapatinib resistance with MicroEnvironment MicroArrays (AACR). Cancer Res. *74* (19 Suppl.), 1831.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., and Bourne, P.E. (2016). The FAIR guiding principles for scientific data management and stewardship. Sci. Data *3*, 160018.