DATA FORUM 2013

March 20-21, 2013

The Joseph B. Martin Conference Center
Harvard Medical School
77 Avenue Louis Pasteur
Boston, MA 02115

The LINCS Data Forum 2013 was supported by
National Institutes of Health grant 3U54HG006097-03S2

Contents

Dear Colleagues,

I am pleased to welcome you to the first LINCS Data Forum. The NIH Library of Integrated Network-based Cellular Signatures (LINCS) project is designed to measure and model the responses of diverse cell types to perturbation and to make the resulting data freely available in a useful way. Data on cellular responses to perturbation are expected to reveal the operation of networks that maintain cellular homeostasis and to uncover differences in these pathways between health and disease. When perturbation is provoked by therapeutic drugs, we expect to identify response determinants that are useful in understanding mechanisms of drug sensitivity and resistance. Knowledge of these determinants should help ultimately guide drug evaluation and clinical use. You can find much more about these goals and developments at http://www.lincsproject.org and the websites that link off it.

The LINCS project is at an early stage but we believe it is important to engage a diverse audience for LINCS data early on, to ensure that our data standards and release procedures are effective and to solicit your thoughts about where to go next. An effort such as LINCS cannot succeed if it operates in isolation and it is our intention to adopt existing data standards whenever we can and to contribute to the larger community effort when new standards need to be promulgated. This is critical because much of the data being collected in LINCS – large scale image libraries, cytokine secretion at the single-cell level, biochemical assays of signal protein state and activity, high-dimensionality transcriptional signatures – has not previously be generated and organized at a similar scale with the intention of public release.

We are particularly interested in the interface between investigators, publishers and funding agencies: who should decide what formats are required, who should house the data, how should we optimize electronic formats so that they are more than just digital versions of paper? We are fortunate that several prominent editors have agreed to join us and we look forward to a panel discussion on this topic on day two of the meeting.

We are aware that LINCS data are only now being released on a significant scale and one purpose of the current meeting is to show you what we have. We are also excited to hear about new classes of measurement that would fit well in the LINCS approach, such as those being collected using CyTOF by the Nolan Lab. Data is rarely useful in the absence of contemporaneous analysis, and it is therefore a pleasure to have Olga Troyanskaya present her approach to synthesizing complex multi-factorial data. In this meeting, as in most meetings, posters and software tool demos are often the best way to see the latest research and we encourage you to participate in these events.

We have prepared a brief questionnaire to solicit your opinions on LINCS and any related activities in written form. We are required to generate a synopsis of this feedback and greatly appreciate your spending a few minutes on it. If the questions seem wrong, feel free to write on the back or to email us after the meeting concludes. Laura Maliszewski will collect your feedback (lauram@hms.harvard.edu).

As many of you know, this meeting is contemporaneous with the VIZBI "Visualizing Biological Data" meeting being held at the Broad Institute and thirty LINCS participants may attend the joint session on Thursday afternoon. We have arranged transportation to the Broad for the session and look forward to an interesting joint session. Getting to VIZBI on time will require some alacrity with respect to transportation so if you are cross-registered for VIZBI, please do not dawdle.

Thank you again for coming,

Peter Sorger
Caroline Shamu
Laura Maliszewski
On behalf of the organizing committee

This meeting is supported by NIH Grant 3U54HG006097-03S2

Brief Description of the NIH LINCS Project

The Library of Integrated Network-based Cellular Signatures (LINCS) Project (lincsproject.org) is an NIH Common Fund program that aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents, and by using computational tools to integrate this diverse information into a comprehensive view of normal and disease states that can be applied for the development of new biomarkers and therapeutics. By generating and making public data that indicates how cells respond to various genetic and environmental stressors, the LINCS project will help us gain a more detailed understanding of cell pathways and aid efforts to develop therapies that might restore perturbed pathways and networks to their normal states.

LINCS data are being 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 and to facilitate the identification of biological targets for new disease therapies.

The LINCS program is being implemented in two phases. Phase 1 began in FY 2010 and focuses on the following activities:

- Large-scale production of perturbation-induced molecular and cellular signatures
- Creation of databases, common data standards, and public user interfaces for accessing the data
- Computational tool development and integrative data analyses
- Development of new cost-effective molecular and cellular phenotypic assays
- Integration of existing datasets into LINCS

Currently, LINCS datasets consist of assay results from cultured and primary human cells treated with bioactive small molecules, ligands such as growth factors and cytokines, or genetic perturbations. Many different assays are used to monitor cell responses, including assays measuring transcript and protein expression; cell phenotype data are captured by biochemical and imaging readouts. Assays are typically carried out on multiple cell types, and at multiple timepoints; perturbagen activity is monitored at multiple doses.

Pending the outcome of the pilot Phase 1 LINCS project, Phase 2 will begin in FY 2014.

Schedule

Wednesday, March 20, 2013

Registration and Breakfast 8:00 - 8:30 AM

Welcome Remarks 8:30 - 8:45 AM
Peter Sorger, Harvard Medical School
Ajay Pillai, National Institutes of Health

Session 1: Collaborating on Big Data
Chair: Caroline Shamu

Jason Swedlow, Dundee University 8:45 - 9:15 AM
The Open Microscopy Environment: Open Source Image Informatics for the Biological Sciences

Peter Sorger, Harvard Medical School 9:15 - 9:45 AM
Measuring and Modeling Cellular Responses to Perturbation

Jake Jaffe, Broad Institute 9:45 - 10:15 AM
Going Wide with Targeted Proteomics: the Rise of Unhole-y Data

Coffee Break 10:15 - 10:30 AM

Jeremy Jenkins, Novartis 10:30 - 11:00 AM
Integrating Chemogenomics and Interactomics in Drug Discovery

Daniela Gerhard, NCI 11:00 - 11:30 AM
Cancer Target Discovery and Development Network

Stephan Schürer, University of Miami 11:30 - 12:00 PM
The LINCS Information FramEwork (LIFE)

Poster and Tool Demonstration Flashtalks

Mario Medvedovic, University of Cincinnati Medical Center 12:00 - 12:05 PM
iLINCS: Web-portal for integrative analysis

Avi Ma'Ayan, Mt Sinai 12:05 - 12:10 PM
Network2Canvas: Network Visualization on a Canvas with Enrichment Analysis

Aris Floratos, Columbia University 12:10 - 12:15 PM
Using the open-source geWorkbench bioinformatics platform to access drug-synergy datasets and computational tools that can be used to characterize drug mechanism of action

Rumi Naik, Carnegie Mellon University 12:15 - 12:20 PM
Learning and using image-derived cell models with CellOrganizer

Lunch with Poster and Tool Demonstration Session 12:20 - 3:00 PM

Keynote Talk **Olga Troyanskaya**, Princeton University *Understanding of Human Disease in a Tissue-Specific Context Through the Integration of Diverse Functional Genomics*	3:00	-	3:45 PM
Todd Golub, Broad Institute *Connectivity Map*	3:45	-	4:10 PM
Nikesh Kotecha, Cytobank *Delivering advanced analyses to the broader community: a view from Cytometry*	4:10	-	4:35 PM
Andrew Hufton, Nature Publishing Group *Scientific Data: An open-access, peer-reviewed platform for data-focused publications*	4:35	-	5:10 PM
Adjourn			5:10 PM

Thursday, March 21, 2013

Registration and Breakfast 8:30 - 8:55 AM

Welcome Remarks 8:55 - 9:00 AM
Peter Sorger, Harvard Medical School

Keynote Talk 9:00 - 9:45 AM
Garry Nolan, Stanford University
Single Cell Systems-Structured View of Immunity & Cancer

Session 3: Tackling the Challenges of Big Data Sets
Chair: Peter Sorger

Andrea Califano, Columbia University 9:45 - 10:15 AM
Elucidating Drug Synergy from large repertoires of molecular profile data

Coffee Break 10:15 - 10:30 AM

Myles Axton, Nature Genetics 10:30 - 10:45 AM
Timing, formatting and attribution for data access

Mike Rossner, JCB, Rockefeller University Press 10:45 - 11:00 AM
The JCB DataViewer: Publishing multidimensional image data

Thomas Lemberger, EMBO 11:00 - 11:15 AM
SourceData - Towards Next Generation Open Access

Panel – Perspectives on managing large data sets 11:15 - 12:00 PM
Moderator: Peter Sorger

- **Thomas Lemberger**, EMBO
- **Mike Rossner**, JCB, Rockefeller University Press
- **Myles Axton**, Nature Genetics
- **Andrew Hufton**, Nature Publishing Group
- **Jason Swedlow**, Dundee University

Wrap Up and Closing at HMS 12:00 PM

If you are registered for the VIZBI-LINCS joint session, the bus for travel to Broad Institute lunch and sessions will depart from Avenue Louis Pasteur. You are responsible for your own travel at the conclusion of VIZBI.

LINCS Data Forum participants not attending VIZBI will be provided with lunch at HMS and are welcome to stay until 1pm.

Speaker Bios

Myles Axton

Editor, Nature Genetics

Myles Axton is the editor of Nature Genetics. He was a university lecturer in molecular and cellular biology at the University of Oxford and a Fellow of Balliol College from 1995 to 2003. He obtained his degree in genetics at Cambridge in 1985, and his doctorate at Imperial College in 1990, and between 1990 and 1995 did postdoctoral research at Dundee and at MIT's Whitehead Institute. Myles's research made use of the advanced genetics of Drosophila to study genome stability by examining the roles of cell cycle regulators in life cycle transitions. His interests broadened into human genetics, genomics and systems biology through lecturing and from tutoring biochemists, zoologists and medical students from primary research papers. Helping to establish Oxford's innovative research MSc. in Integrative Biosciences led Myles to realize the importance of the integrative overview of biomedical research. As a full time professional editor he is now in a position to use this perspective to help coordinate research in genetics.

Andrea Califano

Columbia University

Dr. Califano's research interests reside in the assembly and interrogation of gene regulatory models for the elucidation of mechanisms presiding over cell physiology and their dysregulation in disease, with specific applications to cancer, stem cells, and neurodegenerative disease. His lab, which integrates both experimental and computational research, has pioneered the first genome-wide regulatory model of human cells and a variety of methods for the identification of master regulators of aberrant transformation and physiological differentiation/maturation events and their controlling genetic alterations. These approaches have led to the discovery of several new genes in glioma, leukemia, lymphoma, and prostate cancer that can be targeted pharmacologically to abrogate tumorigenesis. He has also pioneered network-based methods for the elucidation of mechanisms of action of drugs, drug synergy, and drug sensitivity.

Daniela S. Gerhard

Director, Office of Cancer Genomics
Office of the Director, National Cancer Institute, NIH

Her research interests include the identification of somatic mutations in cancer, determination of genetic risk factors in cancer, identification which pathways are the same and which are unique in sporadic vs. inherited cancers, the integration of the results from somatic and genetic studies to elucidate all the various components of cancer etiology and the development of new scientific methods to rapidly translate large scale genomic data into patient-based therapeutics with concomitant predictive markers. These interests reflect the activities of the Office of Cancer Genomics (http://ocg.cancer.gov). OCG supports research to: improve molecular characterization methods and their throughput; evaluate the various novel DNA sequencing technologies; improve the detection of epigenetic changes; and develop new analytical protocols to correlate disease state with the intricate network of molecular interactions in a cancer. The large-scale genomic data sets require the development of new technologies to rapidly translate findings into patient-based therapies with associated predictive markers. The aim is to bridge the technology transfer gap, the "valley of death," in drug discovery and make progress possible. Already the development of novel molecular characterization technologies, including sequencing, has proven to be invaluable to the understanding of cancer-causing mechanisms. OCG is supporting the development of these processes and making results publicly available in an effort to accelerate innovations in drug discovery that will lead to better cancer treatments. Dr Gerhard is a reviewer for a number of journals and has published more than 90 manuscripts and 8 chapters and reviews.

Todd Golub

Broad Institute

Todd Golub is a founding core member of the Broad Institute and serves as the institute's chief scientific officer and director of its Cancer Program.

Golub is a world leader in understanding the basis of cancer, by creating and applying tools of genomics. He has made fundamental discoveries in the molecular basis of childhood leukemia, and laid the foundation for the diagnosis and classification of human cancers using gene expression analysis. He also pioneered the development of chemical screening approaches based on gene expression.

He joined the faculty of the Dana-Farber Cancer Institute and Harvard Medical School in 1997. At the same time he served as the leader of cancer genomics at the Whitehead Institute/MIT Center for Genome Research, which evolved into the Cancer Program at the Broad Institute, which he has directed since 2004. Golub is currently the Charles A. Dana Investigator in Human Cancer Genetics at the Dana-Farber Cancer Institute, professor of pediatrics at Harvard Medical School, and a Howard Hughes Medical Institute investigator.

He is the recipient of multiple awards, including the Richard and Hinda Rosenthal Memorial Award and the Outstanding Achievement Award from the American Association for Cancer Research, the Paul Marks Prize for Cancer Research, the E. Mead Johnson Award from the Society for Pediatric Research, and the Judson Daland Prize for Outstanding Achievement in Clinical Investigation from the American Philosophical Society.

Golub serves on the Board of Directors of the Damon Runyon Cancer Research Foundation and the American Association for Cancer Research. He also serves on the scientific advisory boards of the Sanford-Burnham Medical Research Institute, the Wistar Institute, and St. Jude Children’s Research Hospital. He also serves as Chair of the Board of Scientific Advisors of the National Cancer Institute (effective 11/11).

Golub received his B.A. from Carleton College and his M.D. from the University of Chicago Pritzker School of Medicine. He completed his internship, residency, and fellowship training at Children’s Hospital Boston and the Dana-Farber Cancer Institute.

Andrew Hufton

Nature Publishing Group

Within the Nature Publishing Group, Andrew Hufton is currently developing a new data publication concept, which will help make scientific data more available, discoverable, interpretable and reusable. Andrew was previously an editor at Molecular Systems Biology, a leading open-access systems biology journal, where he helped promote community data standards and encourage data-rich publications.

Jacob D. Jaffe

Broad Institute

Dr. Jacob D. Jaffe is the Assistant Director of the Proteomics Platform at the Broad Institute. He obtained his B.A. in Biochemistry from the University of Pennsylvania and his Ph.D. from Harvard University where he studied with George Church and Howard Berg. Dr. Jaffe has pioneered diverse problems in modern proteomics including large-scale mapping of proteomic data onto genomes, thus allowing their de novo annotation from proteomic evidence, pattern recognition for quantitative proteomics, determination and quantification of epigenetic marks on histone proteins, and high-throughput targeted phosphoproteomics.

Jeremy Jenkins

Novartis Institutes for BioMedical Research

Jeremy Jenkins is a Senior Investigator at the Novartis Institutes for BioMedical Research in Cambridge, MA in Developmental & Molecular Pathways. He has worked for 10 years at Novartis in the fields of lead discovery, cheminformatics, chemogenomics and chemical biology, knowledge engineering, and network pharmacology. Prior to 2003, he was a postdoc at the Harvard Medical School in the Center for Biochemical & Biophysical Sciences and Medicine in the lab of Dr. Robert Shapiro, where he trained in HTS, virtual screening, and computer-aided drug design. He received his PhD in Molecular Genetics from The Ohio State University in 2000.

Nikesh Kotecha

CEO & Co-founder, Cytobank, Inc.

Dr. Nikesh Kotecha is the CEO and Co-founder of Cytobank Inc., a company focused on developing approaches focused on single cell technologies and big data. The company's premier product Cytobank, a cloud based platform for cytometry and the analysis engine behind landmark papers in signaling and the new field of mass cytometry. The Cytobank project grew out of his focus in Dr. Garry Nolan's lab at Stanford University in developing a diagnostic for juvenile myelomonocytic leukemia and on analysis methods for single cell assays.

Dr. Kotecha received his PhD in Biomedical Informatics from Stanford University and his B.S. in Biomedical Engineering from Boston University. Dr. Kotecha's past experiences include work at TIBCO Spotfire Inc. and Wyeth Ayerst Inc. in multiple roles and departments including engineering, technical sales and business development.

He is also a consulting faculty member in the Computational and Systems Immunology program at Stanford University.

Thomas Lemberger

EMBO

Thomas Lemberger is Chief Editor of Molecular Systems Biology and Deputy Head of Scientific Publications at EMBO. Thomas completed his PhD at the University of Lausanne, where he studied the hormonal regulation of gene expression by nuclear receptors. He moved then to Heidelberg where his research focused on the regulation of transcription in the brain. He joined EMBO in 2005.

Garry Nolan

Stanford University School of Medicine

Dr. Nolan is the Rachford and Carlota A. Harris Professor in the Department of Microbiology and Immunology at Stanford University School of Medicine. He trained with Leonard Herzenberg (for his Ph.D.) and Dr. David Baltimore (for postdoctoral work for the first cloning/characterization of NF-□B p65/RelA and the development of rapid retroviral production systems). He has published over 160 research articles and is the holder of 17 US patents, and has been honored as one of the top 25 inventors at Stanford University. He is a recent recipient of the Teal Innovator Award from the Department of Defense and received an award for "Outstanding Research Achievement in 2011" from the Nature Publishing Group.

Dr. Nolan was the founder of Rigel Inc. (NASDAQ: RIGL), and Nodality, Inc., a diagnostics development company and serves on the Boards of Directors of several companies, as well as consults for other biotechnology companies. He is an outspoken proponent of translating public investment in basic research to serve public welfare.

His areas of research include hematopoiesis, cancer and leukemia, autoimmunity and inflammation, and computational approaches for network and systems immunology. Dr. Nolan's most recent efforts are focused on a single cell analysis advance using a mass spectrometry-flow cytometry hybrid device, the so-call "CyTOF". The approach uses an advanced ion plasma source to determine the levels of tagged reagents bound to cells—enabling a vast increase in the number of parameters that can be measured per cell. His laboratory has already begun a large scale mapping of the hematopoietic hierarchy in healthy human bone marrow at an unprecedented level of detail. Dr. Nolan's efforts are to enable a deeper understanding not only of normal immune function, but also detailed substructures of leukemias and solid cancers—which will enable wholly new understandings that will enable better management of disease and clinical outcomes.

Mike Rossner

Executive Director, The Rockefeller University Press

Mike Rossner is the Executive Director of The Rockefeller University Press, which publishes T*he Journal of Cell Biology*, *The Journal of Experimental Medicine*, and *The Journal of General Physiology*. He received his Ph.D. in molecular biology from The University of Edinburgh, under the supervision of Professor Sir Kenneth Murray, and was a post-doctoral fellow at The Walter and Eliza Hall Institute in Melbourne, Australia. He was the Managing Editor of *The Journal of Cell Biology* from 1997 to 2007.

Stephan Schürer

University of Miami

Stephan Schürer PhD currently leads the chemoinformatics program at the Center for Computational Science and is Assistant Professor at the Department of Pharmacology at the University of Miami medical school. He is adjunct Assistant Professor at the Department of Molecular Therapeutics at the Scripps Research Institute Florida where he was previously heading the HTS- and drug discovery chemoinformatics development. In prior industry positions Dr. Schürer was Sr. Director at Eidogen-Sertanty, a developer of life science information technology and scientific content products. At Libraria Inc. he was directing the content operations to develop chemistry and SAR knowledge bases. Stephan Schürer received his Ph.D. in synthetic organic chemistry from the Technical University of Berlin and studied chemistry at Humboldt University-Berlin and University of California Berkeley.

Peter Sorger

Harvard Medical School

Peter Sorger is Otto Krayer Professor of Systems Pharmacology at Harvard Medical School with a joint appointment as Visiting Professor of Biological Engineering at MIT. He is director of a new Laboratory of Systems Pharmacology at Harvard whose aim is to apply network-based approaches to studying therapeutic mechanisms and evaluating new drugs. Sorger has co-founded both pharmaceutical and software companies and serves on the board of several startups. His research focuses on cancer and inflammatory diseases, with the goal of developing and applying novel computational and measurement approaches to understanding cellular biochemistry and molecular pharmacology.

Jason Swedlow

University of Dundee

Jason Swedlow earned a BA in Chemistry from Brandeis University in 1982 and PhD in Biophysics from UCSF in 1994. After a postdoctoral fellowship with Dr T. J. Mitchison at UCSF and then Harvard Medical School, Dr Swedlow established his own laboratory in 1998 at the Wellcome Trust Biocentre, University of Dundee, as a Wellcome Trust Career Development Fellow. He was awarded a Wellcome Trust Senior Research Fellowship in 2002 and named Professor of Quantitative Cell Biology in 2007. His lab focuses on studies of mitotic chromosome structure and dynamics and has published numerous leading papers in the field. He is co-founder of the Open Microscopy Environment (OME), a community-led open source software project that develops specifications and tools for biological imaging. In 2005, he founded Glencoe Software, Inc., a commercial start-up that provides commercial licenses and customization for OME software. In 2011, Prof Swedlow and the OME Consortium were named BBSRC's Social Innovator of the Year and Overall Innovator of the Year. In 2012, He was named Fellow of the Royal Society of Edinburgh. Prof Swedlow is Co-Director of the Analytical and Quantitative Microscopy Course, an annual course that covers the latest developments in advanced quantitative light microscopy at Marine Biological Laboratory, Woods Hole, USA.

Olga Troyanskaya

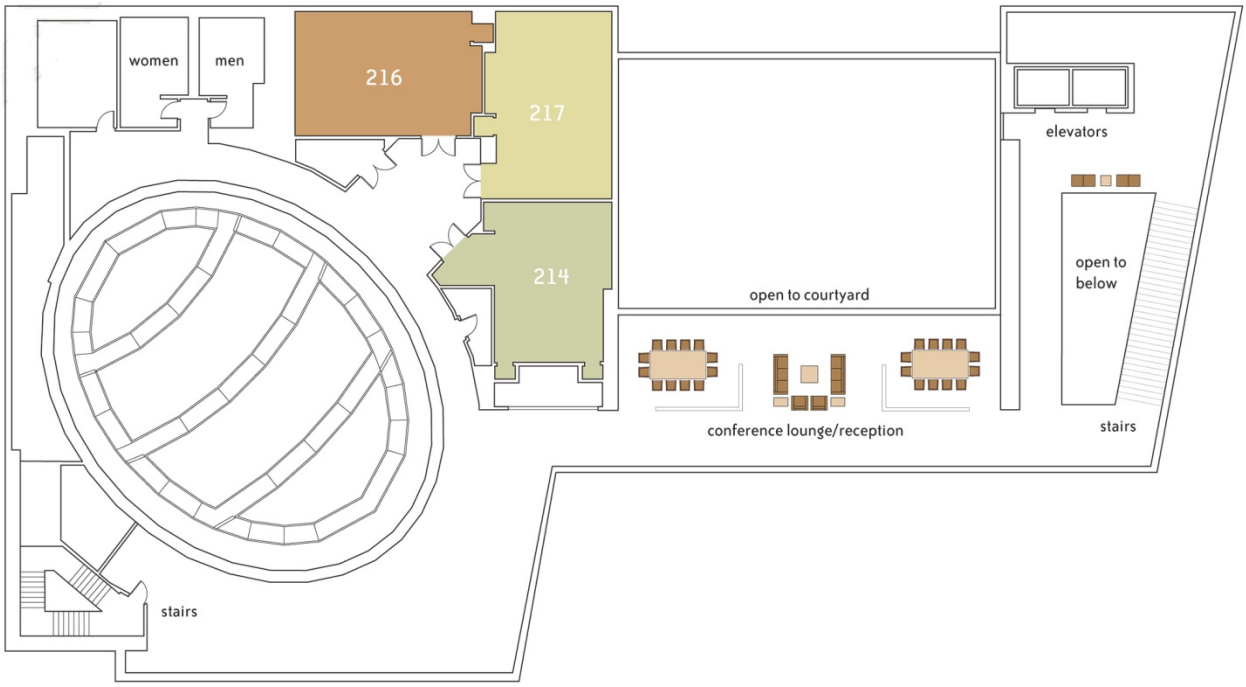
Princeton University

Olga Troyanskaya is an Associate Professor in the Lewis-Sigler Institute for Integrative Genomics and the Department of Computer Science at Princeton University, USA, where she runs the Laboratory of Bioinformatics and Functional Genomics. Her work bridges computer science and molecular biology in an effort to develop better methods for analysis of diverse genomic data with the goal of understanding and modeling protein function and interactions in biological pathways. Her group includes computational and experimental aspects, and tackles diverse questions including developing integrative technologies for pathway prediction and the study of biological networks in complex human disease. Dr. Troyanskaya is an Associate Editor for Bioinformatics, PLOS Computational Biology, and editorial board member of Journal of Biomedical Informatics, Briefings in Bioinformatics, and Biology Direct. She is also a member of the Board of Directors of the International Society for Computational Biology. She received her Ph.D. from Stanford University and is a recipient of the Sloan Research Fellowship, the NSF CAREER award, the Howard Wentz faculty award, and the Blavatnik Finalist Award. She has also been honored as one of the top young technology innovators by the MIT Technology Review and is the 2011 recipient of the Overton Prize in computational biology.

Tool Demonstrations

#	Title	Presenter	Location
1	Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool	Edward Chen	Lounge
2	Network2Canvas: Network Visualization on a Canvas with Enrichment Analysis	Christopher Tan	Lounge
3	SignedPPI : A Web Tool for Integrating Protein-Protein Interaction and Phenotype Data	Bahar Yilmazel	Lounge
4	HMS LINCS DB	Sean Erickson	Lounge
5	LINCS Canvas Browser: Web-Based Interactive Tool to Explore the L1000 Data	Qiaonan Duan	214
6	LINCS Information FramEwork (LIFE)	Christopher Mader	214
7	Explore HMS-LINCS Data and Resources	Jeremy Muhlich	216
8	Using the Open-Source Geworkbench Bioinformatics Platform to Access Drug-Synergy Datasets and Computational Tools that can be Used to Characterize Drug Mechanism of Action	Kenneth Smith	216
9	iLINCS: Web-Portal for Integrative Analysis of LINCS Perturbation, Disease-Related and Transcription Factor Binding Signatures	Mario Medvedovic	216
10	Learning and Using Image-Derived Cell Models with CellOrganizer	Armaghan Naik	216

Joseph B. Martin Conference Center 2nd Floor Map



#1: Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool

Edward Y. Chen[1], Christopher M. Tan[1], Yan Kou[1], Neil R. Clark[1], Zichen Wang[1], Avi Ma'ayan[1,*]
[1]Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York (SBCNY), Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Systematic profiling of genes and proteins in mammalian cells produce lists of genes that need to be further analyzed for their collective functions in order to extract new knowledge. Here, we present Enrichr, an integrative web-based and mobile software application that includes new gene-set libraries, an alternative approach to rank enriched terms, and various interactive visualization approaches to display enrichment results using the JavaScript library Data Driven Documents (D3). Enrichr is open source and freely available online at: http://amp.pharm.mssm.edu/Enrichr. The software can also be embedded easily into any tools that perform gene list analysis. We applied Enrichr to analyze nine cancer cell lines by comparing their enrichment signatures to the enrichment signatures of matched normal tissues. We observed a common pattern of up regulation of the polycomb group and enrichment for the histone mark H3K27me3 in many cancer cell lines, as well as alterations in Toll-like receptor and interlukin signaling in K562 cells when compared with normal myeloid CD33+ cells. Such analyses provide global visualization of critical differences between normal tissues and cancer cell lines but can be applied to many other scenarios. In conclusion, Enrichr is an easy to use intuitive interface providing various types of visualization summaries of functional analyses of gene lists. The tool is available at: http://amp.pharm.mssm.edu/Enrichr.

#2: Network2Canvas: Network Visualization on a Canvas with Enrichment Analysis

Christopher M. Tan[1], Edward Y. Chen[1], Ruth Dannenfelser[1], Neil R. Clark[1], Avi Ma'ayan[1]
[1]Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York (SBCNY), Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Networks are vital to computational systems biology research but visualizing them is a challenge. For networks larger than ~100 nodes and ~200 links, ball-and-stick diagrams fail to convey much information. To address this, we developed Network2Canvas (N2C), a web application that provides an alternative way to view networks. N2C visualizes networks by placing nodes on a square toroidal canvas. The network nodes are clustered on the canvas using simulated annealing to maximize local connections where a node's brightness is made proportional to its local fitness. The canvas is interactive, implemented in HTML5 with the JavaS-cript library D3. We applied N2C to create canvases for 30 gene-gene functional association networks connecting functional terms associated with human and mouse genes, and 6 drug-drug networks based on FDA approved drugs shared properties such side-effects and structural elements. N2C can also perform enrichment analyses. Given lists of genes or drugs, enriched terms are highlighted on the canvas, while the degree of clustering for these enriched terms on the canvas is computed. Since N2C produces visual patterns of enriched terms on canvases, a trained eye can detect signatures instantly. To help users learn the fixed location of terms on canvases we developed a game that rewards remembering the location of terms on the various gene-gene canvases. In summary, N2C provides a new flexible method to visualize large networks and can be used to perform and visualize gene- and drug-set enrichment analyses. N2C is freely available at http://www.maayanlab.net/N2C and is open source.

#3: SignedPPI : A Web Tool for Integrating Protein-Protein Interaction and Phenotype Data

Bahar Yilmazel[2,3], Arunachalam Vinayagam[1], Jonathan Zirin[1], Charles Roesel[2,3], Yanhui Hu[1,2], Anastasia Samsonova[1], Ralph Neumuller[1], Stephanie E. Mohr[1,2], and Norbert Perrimon[1,4]

[1] Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[2] *Drosophila* RNAi Screening Center, Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[3] Bioinformatics program, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

[4] Howard Hughes Medical Institute, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

SignedPPI was developed at Harvard Medical School, in the lab of Norbert Perrimon and Drosophila RNAi Screening Center (DRSC). SignedPPI is a database of protein-protein interaction networks. The current version of the SignedPPI database contains Drosophila signed PPI network, i.e. PPI network with predicted edge-signs (activation/inhibition relationships). The resource includes both literature based signs (collected from SignaLink, KEGG and STKE databases) and predicted edge-signs. The purpose of the database is to help researchers to access, build and navigate a PPI network by offering a simple and interactive web interface.

#4: HMS LINCS DB

Sean Erickson, Dave Wrobel, Jay Copeland, Gabriel Berriz, Jeremy Muhlich, Caroline Shamu
Harvard Medical School

The HMS LINCS DB is a repository for HMS LINCS metadata (perturbagens, cells studied, and proteins) and data. Investigators can use the web interface to browse and search as well as to download datasets. The website presents LINCS data and metadata using a tabular spreadsheet-like metaphor. The columns in the tabular interface are drawn from the fields defined in the data standards documents from the LINCS Data Working Group. Datasets originating in microscopy include thumbnail versions of the raw images with links to download the originals. Microscopy images are managed using the OMERO server software.

A programmatic REST like interface also provides access to the datasets and metadata. Through this interface, other systems may maintain real time data interchange with the system. The main user of the REST interface is currently the University of Miami LIFE system.

The site is built using the Django Web application framework, written in the Python programming language. It is hosted on load-balanced Apache web servers running on multiple Linux based hosts (technical specifications). The database runs on PostgresSQL 8.4 servers running on dual Linux based servers. The servers have RAID-1 disk arrays and are backed up nightly.

#5: LINCS Canvas Browser: Web-Based Interactive Tool to Explore the L1000 Data

Qiaonan Duan[1], Christopher M. Tan[1], Edward Y. Chen[1], Avi Ma'ayan[1]

[1]Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York (SBCNY), Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

The Library of Integrated Network-Based Cellular Signatures (LINCS) project produces masses of genome-wide gene expression data using the L1000 technology. Visualizing and analyzing these data by linking the gene expression signatures to pathways and regulatory networks is a challenge. For this we developed a LINCS Canvas Browser (LCB). LCB is a web-based software application that visualizes L1000 experiments and their enrichment analysis results all at once in a user-friendly interactive canvas viewer. The interface of LCB is made of two compacted layered canvases. The left canvas displays many L1000 experiments clustered based on their gene expression signature similarity, and the right canvas presents results of enrichment analyses for each experiment using 30 gene-set libraries. Both canvases are interactive. The experiment canvas allows users to perform gene list enrichment analysis by clicking on a specific experiment or a group of experiments. The results on the right canvas can be viewed with the following options: a canvas that shows only the enriched terms highlighted, a canvas that displayed the enriched terms on top of existing clusters within the gene set library, a network of enriched terms, or a table with ranked enriched terms. The canvases were created using the Data-Driven Document (D3) JavaScript library that bind data to HTML DOM SVG elements and enables their data driven manipulation making the application light weight a suitable for mobile apps. In conclusion, LCB can be used to browse and interrogate thousands of LINCS L1000 experiments at once. We demonstrate the usefulness of LCB on one of the LINCS drug perturbation datasets. Demo prototype of the application is available at: http://www.maayanlab.net/LINCS/HMS/LCBI/.

#6: LINCS Information FramEwork (LIFE)

Christopher Mader[1], Stephan Schürer[1,2]
[1]Center for Computational Science, [2]Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL.

We will demonstrate key functionality and specific use cases of the most recent release of the LIFE software system.

#7: Explore HMS-LINCS Data and Resources

Jeremy Muhlich

The HMS LINCS center's "Explore" web gallery offers focused vignettes that introduce many of the complex, high-dimensional datasets being produced and curated at the center. For example, the "Kinase inhibitor pathways" vignette is a graphic of canonical immediate-early cell signaling which contains a simple interface for browsing HMS LINCS data on kinase inhibitor selectivity. Touching a kinase in the signaling network displays a list of small molecules which inhibit its activity, along with IC50 data for those molecules collected from six canonical cell lines. The proteins, small molecules, cell lines and data are live links directly into the HMS LINCS database of reagents and datasets.

#8: Using the Open-Source Geworkbench Bioinformatics Platform to Access Drug-Synergy Datasets and Computational Tools that can be Used to Characterize Drug Mechanism of Action

Kenneth Smith

We have used geWorkbench (http://www.geworkbench.org) , the open source bioinformatics platform of Columbia University's National Center for Biomedical Computing (MAGNet, http://magnet.c2b2.columbia.edu/) to provide access to datasets and tools produced by the laboratory of Dr. Andrea Califano. The following geWorkbench modules will be demonstrated:

LINCS Data Querying Interface: the LINCS effort at Columbia focuses on the assessment of the synergistic potential of drug-pairs on a number of different cancer cell lines, using both computational and experimental approaches. A large number of compounds are being tested using multiple synergy metrics and assays. The LINCS Data Querying module allows users to explore the results of computations and experiments and to probe for predicted and validated synergistic effects of drug-pairs across all treated cancer cell lines. It also provides a graphical user interface to enable advanced visualization and interaction with the data.

VIPER Analysis and Visualization: The VIPER algorithm (Virtual Inference of Protein-activity by Regulon Readout) leverages the increasingly accurate and context specific knowledge of regulatory networks, to infer the differential activity of proteins on an individual sample basis, in proteome-wide fashion. Its central premise is that the expression of a protein's transcriptional targets, either direct ones for a transcription factor or indirect ones for a signaling protein, provides the most accurate assessment of its activity. VIPER can thus be used to transform a typical gene expression matrix into a protein activity matrix, representing the inferred activity of each protein across each sample. We will demonstrate how to use geWorkbench to invoke VIPER and visualize its analysis results.

DeMAND Analysis and Visualization: The DeMAND algorithm (Drug Mechanism of Action using Network Dysregulation) integrates gene expression profiles from drug-treated and non-treated cancer lines with tissue-specific regulatory networks (interactomes) to identify targets of drug-induced dysregulation. For each pair of genes connected through an interactome edge, DeMAND first determines the joint probability distribution of the gene expression levels both pre-and post-drug treatment. The change in the probability distribution is used to estimate the statistical significance of the dysregulation of each edge. If the first-order neighborhood of a gene in the interactome is enriched in dysregulated edges, then this is an indication that this gene may be involved in the drug's mechanism of action. We will demonstrate how to use geWorkbench to invoke DeMAND and visualize its analysis results.

#9: iLINCS: Web-Portal for Integrative Analysis of LINCS Perturbation, Disease-Related and Transcription Factor Binding Signatures

Mario Medvedovic

iLINCS (integrative LINCS) data analysis portal is a biologist-friendly web tool for using LINCS data, accessing results of LINCS data analyses. The portal facilitates searching for pre-computed genome-scale signatures of gene activity as well as constructing new signature using genomics datasets stored in the portal's backend databases. Once users identify or construct a signature of interest, they can search for signatures that are concordant with their signature of interest. Finally, the portal facilitates examination of biological underpinning of such concordances. Currently there are four libraries of pre-computed genome scale signatures:

LINCS Perturbagen Signatures: transcriptional signatures of "perturbagen" activity constructed based on Broad L1000 assay data. Each signature consists of the differential expression of each of ~1000 genes measured by the assay averaged across at least two replicated plates, and associated p-values.

Disease Related Signatures: transcriptional signatures constructed by comparing sample groups within the collection of public domain transcriptional dataset GEO GDS collection. Each signatures consists of genome scale (ie for each gene) differential expression levels and associate p-values.

ENCODE Transcription Factor Binding Signatures: transcription factor (TF) binding signatures constructed using ENCOD ChIP-seq data. Each signature consists of genome scale (ie for each gene) scores and probabilities of regulation by the given TF in the specific context (cell line and treatment). These signatures were developed using our in-house TREG methodology (paper under review).

Connectivity Map Signature: transcriptional signatures of "perturbagen" activity constructed based on the version 2 of the original "Connectivity Map" dataset using Affymetrix expression arrays.

Signatures in these libraries can be either used to find a signature of interest to start the analysis, or to compare against when searching for concordant signatures.

Genomic datasets that can be used to create custom signatures are organized in ten groups ranging from RNA-seq and CNV data generated by The Cancer Genome Atlas (TCGA) project, an almost complete collection of GEO GDS datasets, and collections of datasets associated with different diseases (eg Breast Cancer, Prostate Cancer and Juvenile Rheumatoid Arthritis). The gene lists that can be used for identifying signatures using enrichment analysis consist of list of genes associated with Gene Ontology terms, KEGG pathways, transcription factor binding targets, diseases, mammalian phenotypes, oncogenic pathways, lists of differentially expressed genes collected from peer reviewed papers, lists of genes affected by specific chemicals, and a few more custom created sets of gene lists.

#10: Learning and Using Image-Derived Cell Models with CellOrganizer

Armaghan Naik

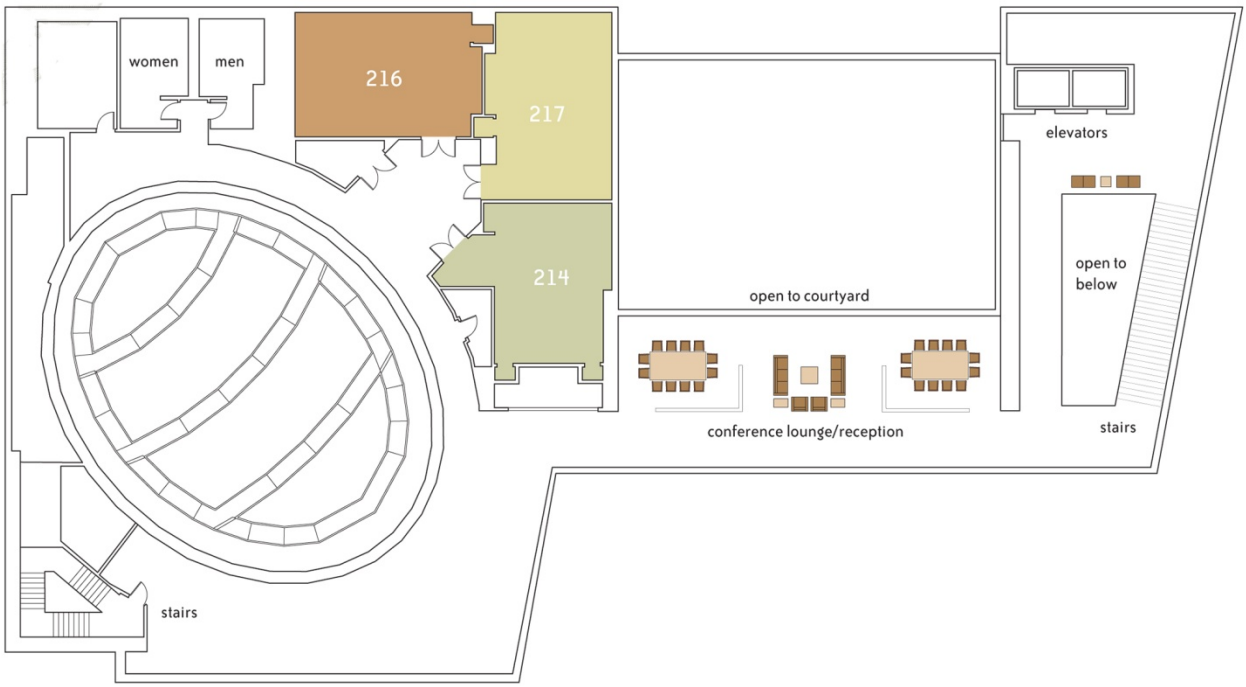
CellOrganizer is an open source software system that can learn models of the size, shape and spatial distribution of cellular components directly from images. These models are generative, which means that they can be used to synthesize new images of cells that are statistically similar to the ones they were trained on.

Models can also be combined to generate new images that can be used as the basis for spatially-realistic cell simulations using systems such as Virtual Cell and MCell. This demo will illustrate the processes of learning models and of synthesizing images from the models in a variety of formats.

Poster Abstracts

#	Title	Presenter	Location
11	From Synergistic Interaction Prediction to Experimental Assessment of Compound Combinations	Mariano Alvarez	Lounge
12	Functionalization of Cancer Genomes by High-Throughput Cell Line Screening	Cyril Benes	Lounge
13	Genome-Wide Signatures of Transcription Factor Activity: Connecting Transcription Factors, Disease, and Small Molecules	Jing Chen	Lounge
14	A Geometrical Approach to Differential Gene Expression	Neil Clark	Lounge
15	Multi-Parametric Analysis of Dose-Response Behavior in Cancer Cells	Mohammad Fallahi-Sichani	Lounge
16	Visualization, Integration, and Access for High Throughput Single-Cell Metabolic Data	Honor Glenn	Lounge
17	Viewing LINCS Data with the UCSC Cancer Genomics Browser	Mary Goldman	Lounge
18	ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode the Transcriptional Regulome in Breast Cancer Cell-Line and Patients	Yan Kou	Lounge
19	Proteoplex 1.1: A 45-Plex Single-Cell Secretion Profiling Platform to Interrogate Functional Cellular Heterogeneity	Yao Lu	Lounge
20	LINCS Information FramEwork (LIFE)	Christopher Mader	Lounge
21	Multivariate Analysis of Immediate Early Signaling Following Growth Factor Stimulation in Breast Cancer Cell Lines	Mario Niepel	214
22	The Receptor Tyrosine Kinase Layer Characterizes Breast Cancer Cell Lines and Predicts Sensitivity to Therapeutic Drugs	Mario Niepel	214
23	The Directionality of Drug-Drug Interactions	Adam Palmer	214
24	Generating Network Signatures for Cancer Using a Hybrid Search and Bayesian Modeling Approach	Evan Paull	214
25	COMPLEAT: A Protein Complex-Based Analysis Tool for High-Throughput Data Sets	Charles Roesel	214
26	Modeling Cancer Combination Therapies Reveals Systems Strategies to Overcome Drug Resistance	Jérémie Roux	214
27	Network Dysregulation Following Drug Treatment Predicts Drug Mode of Action	Yishai Shimoni	216
28	itNETZ: The Signature-based Drug Discovery Platform	Jing Su	216
29	Metadata Standards and Annotation Format to Describe the LINCS Assays	Uma Vempati	216
30	Compound Identity-Based Survey of LINCS Assays and Results	Dušica Vidović	216

Joseph B. Martin Conference Center 2nd Floor Map



#11: From Synergistic Interaction Prediction to Experimental Assessment of Compound Combinations

Mariano J. Alvarez[1], Yao Shen[1], Gabrielle Rieckhof[1], Serge Pampou[2], Charles Karan[2] and Andrea Califano[1, 2, 3, 4, 5]

[1]Columbia Initiative in Systems Biology, [2]Columbia Genome Center High Throughput Screening Core, [3]Departments of Biochemistry & Molecular Biophysics and of Biomedical Informatics, [4]Institute of Cancer Genetics and Herbert Irving Comprehensive Cancer Center, [5]Motor Neuron Center and Columbia Stem Cells Initiative, Columbia University, New York.

Our proposal aims to complement LINCS data and other publically available datasets by generating the molecular profile data necessary to transform combination therapy from a trial and error exercise to a rational discipline, by allowing identification of synergistic drug combinations on a predictive basis. To maximize the potential synergistic behavior of compound combinations, we selected a panel of small molecule compounds showing orthogonal mechanisms of action. Specifically, we developed a computational framework to infer small compound functional mode of action (fMoA) by estimating their effect on transcriptional regulators. We performed this task by first assembling relevant cell-context specific regulatory networks; and then computing the enrichment of compound-perturbation gene expression signatures on each regulator target gene-set (regulon) to obtain compound-characteristic transcriptional regulator activity signatures (compound fMoA). We previously showed that our approach is able to correctly infer the MoA for the nuclear receptor targeting compounds in MCF7-CMAP2 (connectivity map-2 dataset generated by Todd Golub's lab at the Broad Institute).

Then, we generated a panel of 100 compounds (sensitizers panel) showing maximal orthogonality on the fMoA space. We show here the results for the screen of 10 compounds with well characterized mode of action combined with the sensitizers panel. The readout of the assay was LnCap viability at 48h after compound combinations assayed at 4 different doses. We estimated non-additive interactions by the Bliss method and we plan to perform full 10x10 titration curves for the combinations showing significant deviation from the expected additive interaction. These preliminary results showed a clear separation of the sensitized compounds in 3 categories: compounds showing specific synergy interactions with few compounds, non-specific sensitizers showing synergy with most compounds, and non-specific inhibitors interacting antagonistically with most compounds.

In parallel, we performed and show the results for the first combinatorial screen of the 5 trans-centers project compounds on five cell lines. The experimental design was 4x4 dilutions and non-additive interactions were inferred by Bliss criteria. The use of five different cell lines on the screen allowed us to show that the non-additive interactions are highly cell-line specific.

#12: Functionalization of Cancer Genomes by High-Throughput Cell Line Screening

Anahita Dastur, Elena Edelman, Patricia Greninger, Daniel Haber, Sridhar Ramaswamy and Cyril Benes
HMS-LINCS. The Massachusetts General Hospital Cancer Center. Bldg. 149, 13th St. Charlestown. MA 02129.

We seek to provide mechanistic basis for the development of targeted therapies and to identify biomarkers actionable in patient selection strategies. Variations in cancer genomes are linked to specific drug responses by systematic interrogation of a large collection of genetically characterized tumor derived cell lines using small molecules. The >1,000 cell line collection is representative of a wide array of tumors and captures a broad heterogeneity of cancer genomes. Using mutational status, copy number variations and genome wide expression data we identify genomic correlates of drug sensitivity. Modeling of the drug sensitivity data can give insight into functional disease subclass, mechanism of action of drugs and cellular regulatory networks. Genome-drug sensitivity modeling is currently performed using MANOVA to identify single genetic events that might confer strong resistance or sensitivity phenotypes. Elastic Net, a linear regression modeling incorporating all the genomic data available is used to define multivariate descriptors or drug sensitivity.

#13: Genome-Wide Signatures of Transcription Factor Activity: Connecting Transcription Factors, Disease, and Small Molecules

Jing Chen, Zhen Hu, Mukta Phatak, John Reichard, Johannes M Freudenberg, Siva Sivaganesan and Mario Medvedovic

Identifying transcription factors (TF) involved in producing a genome-wide transcriptional profile is an essential step in building a mechanistic model that can explain observed gene expression data. We developed a statistical framework for constructing genome-wide signatures of TF activity, and for using such signatures in the analysis of gene expression data produced be complex transcriptional regulatory programs. Our framework integrates ChIP-seq data and appropriately matched gene expression profiles to identify True REGulatory (TREG) TF-gene interactions. It provides genome-wide quantification of the likelihood of regulatory TF-gene interaction which can be used to either identify regulated genes, or as genome-wide signature of TF activity.

To effectively use ChIP-seq data, we introduce a novel statistical model that integrates information from all binding "peaks" within 2Mb window around a gene's transcription start site (TSS), and provides gene-level binding scores and probabilities of regulatory interaction. In the second step we integrate these binding scores and probabilities with gene expression data to assess the likelihood of True REGulatory (TREG) TF-gene interactions. Through a large study of disease-related transcriptional signatures and transcriptional signatures of drug activity, we demonstrate that increase in statistical power associated with the use of TREG signatures makes the crucial difference in identifying key targets for treatment and drugs to use for treatment. We also computed TREG binding signatures for all ENCODE TF binding datasets and correlated them with LINCS perturbation signatures. Associations between perturbagens and TF binding signatures can be accessed and analyzed within the iLINCS data analysis portal (http://LincsGenomics.org). All methods are implemented in an open-source R package treg, which can be downloaded at http://GenomicsPortals.org.

#14: A Geometrical Approach to Differential Gene Expression

Neil R. Clark[1], Avi Ma'ayan[1]

[1]Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York (SBCNY), Mount Sinai School of Medicine, New York, NY 10029 USA

Genome-wide gene expression profiles, as measured using microarray or RNA-Seq experiments, have revolutionized biological and biomedical research by providing a quantitative measure of the entire mRNA transcriptome. Typically, researchers use a statistical test of difference of location, such as the t test, to identify differentially expressed genes upon which further analysis and ultimately biological discovery from such experiments is based. Here we describe a novel geometrical approach to differential expression which is capable of identifying more relevant genes. In an extension of this approach we are also able to use the principal angle, which is related to the canonical correlations, to evaluate the significance of genes sets. We evaluate our approach in two case studies from breast cancer: In the first, we use various methods to match target genes from ChIP-Seq profiling estrogen receptor alpha and beta with differentially expressed genes after estrogen simulation of the same MCF7 cells; whereas in the second case study, we evaluate the quality of enriched terms when comparing normal epithelial cells with cancer stem cells. In conclusion, we demonstrate that our approach might be better in calling significant genes and should be considered as an alternative to the widely used t-test method.

#15: Multi-Parametric Analysis of Dose-Response Behavior in Cancer Cells

Mohammad Fallahi-Sichani, Peter K Sorger
Department of Systems Biology, Harvard Medical School, Boston, MA 02115

Pharmacological sensitivity of cancer cells to anti-cancer drugs are often evaluated using in vitro assays measuring the 72 h impact of different doses of a drug on viability of a group of cancer cell lines followed by analysis of pharmacological parameters derived from dose-response curves. Despite the multi-parametric nature of dose-response behavior of cancer cells, implicit in most of the current analyses is the assumption that dose-response curves for resistant cell lines are shifted to the right as compared with sensitive cell lines. In other words, cell lines with larger half-maximal effective concentration (EC50) and half-maximal inhibitory concentration (IC50) values for a drug are considered more resistant to that drug in comparison with other cell liens, and thus these parameters are widely used as a measure for drug sensitivity in cancer cells. Interestingly, different anti-cancer drugs (which inhibit different targets) can induce totally different patterns of response across the same group of cell lines. Thus, a challenge for analysis of drug sensitivity in cancer cell lines is how to select the appropriate parameters that can be used to accurately describe relative sensitivity of different cancer cell lines to a drug. For example, high slopes of the dose-response curve or high maximal effect level (Emax) values allow for significantly higher levels of inhibition at concentrations above IC50 compared to lower slopes or lower Emax values. Therefore, if a genetic mutation in cancer cells decreases the slope or Emax, it may cause resistance beyond what is anticipated from a consideration of EC50 and IC50 alone.

To gain insight into relative usefulness of different pharmacological parameters derived from dose-response curves and to capture drug-to-drug and cell line-to-cell line variation with respect to these parameters, we perform a multi-parametric analysis of dose-response behavior of a group of breast cancer cell lines treated with a variety of anti-cancer drugs, including both targeted agents and cytotoxic chemotherapeutics. Our analysis shows that among pharmacological parameters that can be extracted from dose-response data for a drug, some might be more informative than others in describing relative sensitivity of different cell lines to that drug, and the level of information depends on drug class.

#16: Visualization, Integration, and Access for High Throughput Single-Cell Metabolic Data

Honor L. Glenn
Arizona State University

We are developing a high throughput platform for dynamic, multiparameter sensing of single-cell metabolic phenotypes. The emerging technology focuses on the metabolic analytes, O2, glucose, pH, and ATP, however this tool is extensible to any intracellular or extracellular target that can be detected by a fluorescent sensor. Ultimately, the raw data from a single experiment will consist of fluorescent intensities at multiple time points for up to four molecular species for hundreds to thousands of individual cells. This constitutes a very high ratio of data volume to biological sample material. We are exploring methods for organizing and presenting this data that facilitate interpretation while preserving information content. We are interested in discovering ways to synergistically integrate this kind of data with other LINCS contributions to aid non-intuitive extrapolation and enhance the practical value of combined data sets. We are also evaluating approaches to make this data easily accessible to a wider scientific community in order to maximize biomedical impact. We are actively seeking collaborations with other groups interested in data visualization, integration, and access.

#17: Viewing LINCS Data with the UCSC Cancer Genomics Browser

Brian Craft, Teresa Swatloski, Mary Goldman, Kyle Ellrott, Melissa Cline, Christopher Wilks, Singer Ma, Christopher Szeto, Josh Stuart, David Haussler, Jingchun Zhu

The UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu/) is a set of web-based tools to display, investigate and analyze cancer data. Genomic data is shown as a heatmap side-by-side with its associated clinical information, allowing users to group samples by either data type. Biological pathways and collections of genes can also be used to sort, aggregate and zoom into groups of samples.

Summary views, online statistical analysis, and Kaplan Meier survival plots allow easy comparisons across subgroups. Custom clinical data allow users to integrate their own sample annotations into existing datasets. We currently display an expanding set of data, including datasets from 25 TCGA (The Cancer Genome Atlas) cancers as well as data from CCLE (Cancer Cell Line Encyclopedia) and SU2C (Stand Up To Cancer). We host L1000 data consisting of a 15 cell line panel treated by ~2000 compounds and genetic perturbations on ~4000 genes (including shRNA knockout and over-expression of gene of interest) under various treatment concentration and durations. We also host LINCS KinomeScan data and protein expression data generated on RPPA arrays from 216 cell lines. Using the browser, investigators can query genes of interest and compare the gene expression change under any treatment condition. LINCS data access is currently limited to LINCS investigators until the consortium decides to make it public.

Displaying LINCS data has given us an opportunity to start working on addressing issues surrounding very large datasets.

#18: ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode the Transcriptional Regulome in Breast Cancer Cell-Line and Patients

Yan Kou[1], Qiaonan Duan[1], Christopher M. Tan[1], Edward Y. Chen[1], Neil R. Clark[1], Avi Ma'ayan[1]
[1]Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York (SBCNY), Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Gene expression in mammalian cells is regulated by transcriptional complexes that include transcription factors, histone modifiers, other chromatin and DNA modifiers, and co-regulators proteins. These cellular components are increasingly profiled using the ChIP-seq/chip technologies to identify the locations on the genome where such complexes bind. Integrating datasets from many sparse ChIP-seq/chip studies applied to mammalian cells is challenging but can lead to a new level of global understanding of gene expression regulation. In 2010 we published a database called ChIP-X Enrichment Analysis (ChEA) where we compiled data from 87 publications reporting the binding of 92 mammalian transcription factors to their putative targets. Since then we expanded this database where it is currently contains 198 transcription factors from 213 publications for a total of 441,828 transcription-factor/target interactions. In addition, we compiled data from the ENCODE project which includes 920 experiments applied in 44 cell-lines profiling 160 transcription factors for a total of ~1.4 million transcription-factor/target interactions. Moreover, we processed data from the NIH Epigenomics Roadmap project for 27 different types of histone marks in 64 different human cell-lines. All together the data was processed into four simple gene-set libraries where the set label is either a mammalian transcription factor or a histone modification mark in a particular cell line, organism and experiment. Such gene-set libraries are useful for elucidating the transcriptional networks regulating sets of genes of interest using gene-set enrichment analyses. Furthermore, processing the integrated gene-set libraries using a network inference algorithm that we developed, we constructed a regulatory network of transcription factors and histone modifications to identify groups of regulators that work together. As a case study, we applied enrichment analysis, using these novel gene-set libraries, to all the breast cancer tumor samples from TCGA. We first identified differentially expressed genes across all tumor samples, and then applied enrichment analyses for each patient, and then clustered the patients based on the results. We call this analysis, meta-signature analysis because it converts a gene expression signature to an inferred transcription factor or a histone modification signature. We show the usefulness of this meta-signature strategy by demonstrating that the clusters of patients we identify strongly classify patients based on their outcome. This means that it is predictive. In conclusion, the construction of a global regulatory network of transcription factors, histone modifiers and co-regulators allows exploration of upstream transcriptional regulatory complexes given a set of differentially expressed genes, as well as assists in elucidating the relationship among these regulators at a global scale.

#19: Proteoplex 1.1: A 45-Plex Single-Cell Secretion Profiling Platform to Interrogate Functional Cellular Heterogeneity

Yao Lu, Jonathan Chen, Qiong Xue, Kathryn Miller-Jensen* and Rong Fan*
Department of Biomedical Engineering, Yale University
(KMJ: kathryn.miller-jensen@yale.edu, RF: rong.fan@yale.edu)

Recent evidence indicates that a genetically-identical cell population can give rise to diverse phenotypic differences. Non-genetic heterogeneity is emerging as a potential barrier to accurate monitoring of cellular immunity and effective pharmacological therapies, suggesting the need for practical tools for single cell analysis of proteomic signatures. Herein we describe a microchip platform, called Proteoplex 1.1, that integrates a sub-nanoliter microchamber array and high-density antibody barcodes for simultaneous detection of a panel of cytokines from more than a thousand single cells in parallel. The chip can be executed in a simple assay "kit" with no need of sophisticated fluid control or bulky equipment. This technology builds upon prior successes in antibody barcode-based protein secretion measurement technique, but uses different schemes of cell capture, quantification, automated data analysis, and eliminates bulky fluid handling systems, resulting in a truly practical and informative tool that may find immediate use in both laboratory research and clinical cellular diagnosis. In this device, the antibody barcode array slide comprises 2 repeats of 15 stripes of different antibodies, immobilized on a poly-L-lysine-coated surface. The capture antibody stripes are 20□m in width and the pitch size of a full barcode is 1mm. The detection antibodies conjugated with three different fluorophores were synthesized to perform multi-color detection in each stipe. All-together, it resulted in a 45-plexed protein assay at the single-cell level. The microchamber array is a one-layer microchip fabricated by soft lithography from polydimethylsiloxane (PDMS), an optically transparent silicone elastomer widely used for biological microfluidics. It contains 5044 rectangular microchambers, each of which is 1.8mm, 20□m and 15□m, in length, width and depth, respectively. These two parts were manufactured independently and combined during the assay such that the barcode slide acts as a disposable test strip and the microchamber array as a reusable device. Proteoplex 1.1 has been tested with human macrophages and its results showed good correlation with Micro-ELISA and ICS flow cytometry. A wide variety of single cell polyfunctionality was observed and U937 derived macrophage cells showed more polyfunctionality upon LPS stimulation. Proteoplex 1.1 offers both high throughput and high multiplexing capability, representing a potentially transformative tool for monitoring of cellular function and immunity.

#20: LINCS Information FramEwork (LIFE)

Christopher Mader[1], Hande Küçük[1,2], Amar Koleti[1], Nakul Datar[1], Saminda Abeyruwan[1,2], Caty Chung[1], Uma Vempati[1], Ubbo Visser[1,2], Stephan Schürer[1,3]

[1]Center for Computational Science, [2]Department of Computer Science, [3]Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL.

We illustrates the LIFE software components, ontology model, and User Interface (UI) to integrate, access, navigate, and explore LINCS data from a variety of perspectives. LIFE's simple UI allows searching based on asserted and inferred annotations of LINCS assays, participants and LINCS screening results.

#21: Multivariate Analysis of Immediate Early Signaling Following Growth Factor Stimulation in Breast Cancer Cell Lines

Mario Niepel*[1], Marc Hafner*[1], Emily Pace*[2], Birgit Schoeberl[2], Peter K. Sorger[1]
[1]Department of Systems Biology, Harvard Medical School, Boston, MA
[2]Merrimack Pharmaceuticals, Cambridge, MA
*equal contribution

Intracellular signaling following growth factor stimulation is generally studied using a limited number of cell lines and ligands. Here we treat a panel of 40 breast cancer cell lines with 15 different growth factors at two concentrations and measure the activity of key kinases downstream of the receptors (ERK and AKT) at three time points. We combine this data with the expression and phosphorylation levels of 20 RTKs under serum starvation conditions to discover general rules that govern ligand stimulation in breast cancer cell lines.

We observed much variance in the ligand response both across different cell lines and ligands, but we also found many consistent patterns in the data. Breast cancer cell lines responded generally well to ErbB ligands, FGFs, and IGF/INS, while they responded only sporadically or weakly to PDGF, VEGF, or NGF. The kinetics of the response was driven predominantly by the type of ligand and only weakly influenced by cell line, dose, or the downstream target. We found that the ErbB ligands and HGF induce a large fold change on both pAKT and pERK levels, whereas the FGF ligands are specific to pERK and the INS/IGF ligands to pAKT. We related the signaling response to the RTK profile of the cell lines and found that the strength of the response to a ligand is driven by the availability of the receptor, but inhibited by the activation of the pathway in basal conditions.

Our results show that the ligand response in breast cancer cell lines varies to great extent, but there are commonalities among ligand families and the clinical subtype of the cells. This work is an important step toward a better understanding of the response of cell lines to growth factors which are the major drivers of the activity level of native breast cancers in their microenvironment.

#22: The Receptor Tyrosine Kinase Layer Characterizes Breast Cancer Cell Lines and Predicts Sensitivity to Therapeutic Drugs

Mario Niepel*[1], Marc Hafner*[1], Emily Pace*[2], Birgit Schoeberl[2], Peter K. Sorger[1]
[1]Department of Systems Biology, Harvard Medical School, Boston, MA
[2]Merrimack Pharmaceuticals, Cambridge, MA
*equal contribution

Currently, clinicians use the levels of the hormone receptors for estrogen and progesterone and amplifications of the ErbB2 receptor tyrosine kinase (RTK) to distinguish three groups of breast cancers and guide the choice of therapeutic treatments. However, even the best biomarker in breast cancer – HER2amp status – correctly predicts trastuzumab responsiveness in only ~30% of patients, suggesting that other tumor properties contribute to therapeutic efficacy. In this work we measure the level of and responsiveness of RTKs across a panel of breast cancer cell lines and show that these measurements characterize the cell lines and predict their sensitivity to therapeutic inhibitors.

We used a panel of approximately 40 cell lines, treated them with 15 different growth factors and measured the activity of key kinases downstream of the receptors (ERK and AKT). In addition, we measured the expression and phosphorylation levels of 20 RTKs under serum starvation conditions. We used these measurements to predict the drug sensitivity (defined by the GI50 value) of the cell lines in response to over 40 therapeutic inhibitors. We found that our measurements predict the sensitivity to about half of the drugs studied. The majority of predicted drugs target the ErbB receptors and the AKT pathway. For these drugs, we extracted markers of sensitivity that may be useful in a clinical setting. In particular for the drugs targeting the ErbB receptors (like Lapatinib), we found the phosphorylation level of ErbB2 to be more predictive of sensitivity than the expression level of ErbB2, which is commonly used. We also found the expression level of ErbB3 to be correlated with sensitivity to drugs targeting mTOR or PIK3CA. This correlation is particularly predictive of the sensitivity of triple negative cell lines. Finally, by combining genomic information such as the PTEN/PIK3CA mutational status with our measurements, we derived more predictive models. For example Triciribine (AKT inhibitor) sensitivity of the wild type cell lines is solely predicted by the expression level VEGFR1 and the phosphorylation level of Met: with higher levels correlating with a better sensitivity of the cells to the drug.

Our results suggest that a carefully selected set of measurements might be sufficient to classify responsiveness of breast cancers to a number of cancer drugs currently used in the clinic. This is particularly relevant in the case of triple negative breast cancers for which choosing effective treatment has proven difficult.

#23: The Directionality of Drug-Drug Interactions

Adam Palmer

Drug-drug interactions can be classified as synergistic, additive or antagonistic based on how a single phenotype (commonly growth or viability) responds to a drug combination. However, the relative contribution of each component drug to the final outcome cannot be resolved from a single phenotype. Here we present a method to study drug interactions with multiphenotype profiling, where the response to a drug combination can be decomposed into a weighted sum of the individual drugs' phenotypic signatures. Thus, the 'directionality' of drug-drug interactions is revealed by separately resolving how the first drug's effect is altered by the second drug, and how the second drug's effect is altered by the first drug. We apply this method to a high-content screen of 29 bioactive compounds applied in all combinations to HeLa cells in culture, and observe that synergistic interactions are typically 'sent' from a few compounds that increase the potency of many other compounds, while antagonistic interactions are often 'received' by a set of compounds whose effects are suppressed by many other compounds. Studying the directionality of drug-drug interactions should be valuable to the rational design of drug combinations.

#24: Generating Network Signatures for Cancer Using a Hybrid Search and Bayesian Modeling Approach

Evan Paull, Daniel Carlin, Christopher Wong and Josh Stuart
Department of Biomolecular Engineering, University of California, Santa Cruz

Samples from the same cohort are characterized by any number of genomic perturbations involving gene mutations, focal copy number gains and losses, and distinct promoter methylation events. One goal of cancer genomics is to connect these observed and imposed perturbations to the molecular changes that occur in cancer cells. Identifying genetic pathways activated in response to perturbations will lead to a mechanistic understanding of drug response and disease progression.

We have developed a method that combines a new search method (TieDIE) to perform network feature selection with a Bayesian approach to fit the observed data, generating a "molecular network signature" in the process. The TieDIE method computes a subnetwork solution that interconnects protein level data to gene expression level data using protein-protein interactions, predicted transcription factor to target connections, and curated interactions from literature. This "data-centered" subnetwork is then used to build network signatures with a probabilistic graphical model framework that learns the relationships between genes, proteins and transcriptional activity in the data.

We used publicly available data from the TCGA Breast Cancer project to generate a subnetwork summarizing the relevant genes and connections that distinguish Basal and Luminal cancer subtypes. We then used a Bayes net framework to train molecular network signatures on each subtype and show that this model can distinguish between the common molecular features in each subtype. We outline a plan to construct and use these "molecular signatures" with additional LINCS and TCGA datasets, and discuss future efforts to use these signatures in a range of prediction tasks.

#25: COMPLEAT: A Protein Complex–Based Analysis Tool for High-Throughput Data Sets

Arunachalam Vinayagam, Yanhui Hu, Meghana Kulkarni, Charles Roesel, Richelle Sopko, Stephanie E. Mohr, and Norbert Perrimon

The analysis of data sets from genome-scale screens typically involves raw data processing, such as calculating z-scores and fold changes, where genes are given a score and identified as hits. Because these screens output hundreds of genes, it is standard practice to employ gene-set enrichment analysis to interpret these screens. COMPLEAT is conceptually similar to gene-set enrichment analysis, but COMPLEAT is based on protein complexes rather than gene sets. COMPLEAT allows multiple data sets to be uploaded, analyzed, and compared in the context of these complexes. To facilitate analysis, COMPLEAT supports:

- Multiple gene identifiers including human Entrez gene ID, FlyBase gene ID, gene symbol, locus tags, and synonyms
- Mixed gene identifiers within a data set
- Multiple data sets within in a single file, or separated in multiple files or worksheets.
- RNAi screen, proteomics, and expression data sets
- Dynamic comparison of multiple conditions and time-series data

#26: Modeling Cancer Combination Therapies Reveals Systems Strategies to Overcome Drug Resistance

Jérémie Roux, Marc Hafner, Samuel Bandara, Josh Sims and Peter K. Sorger
Department of Systems Biology, Harvard Medical School, Boston, MA, USA

TRAIL, DR4, and DR5 agonistic antibodies, which bind death receptors to trigger cell death, are exciting new therapeutics currently in clinical trials. In this study, we seek to understand the determinants of the effective Death receptor Induced Signaling Complex (DISC) activation that ultimately lead to cell death. Using live cell imaging we monitor, in each single cell: the dynamics of activation of initiator caspases within the DISC and the cell death outcome. In combination with pySB, a rules-based programmatic modeling approach, we investigate (a) how DISC dynamics alone can be pharmacologically modulated to impact cell fate and (b) how pharmacological perturbations of other modules of the apoptotic pathway, such as the mitochondria and effectors caspases, can sensitize cells with a given DISC activation (combination therapy). In this presentation we emphasize the synergy between theoretical and experimental approaches to help evaluate, on a tumor cell type basis, the most predictive set of measurable biomarkers with hopes to identify optimal therapeutic combinations.

#27: Network Dysregulation Following Drug Treatment Predicts Drug Mode of Action

Jung Hoon Woo[1], Mukesh Bansal[1], Yishai Shimoni[1], Gonzalo Garci Lopez[1], Archana Iyer[1], Maria Rodriguez-Martinez[1], Andrea Califano[1,2,3,4]

[1]Columbia Initiative in Systems Biology, [2]Columbia Genome Center High Throughput Screening Core, [3]Departments of Biochemistry & Molecular Biophysics and of Biomedical Informatics, [4]Institute of Cancer Genetics and Herbert Irving Comprehensive Cancer Center, Columbia University, New York.

Our lab is developing a method to identify Drug Mechanism of Action using Network Dysregulation (DeMAND). DeMAND identifies drug mechanism of action by comparing gene expression profile following drug perturbation and control samples and computing it's effect on the individual interactions within an integrated transcriptional and post-translational regulatory model (interactome). The method is comprised of two steps. In the first step a context-specific interactome is used, which was previously reverse engineered using the ARACNe and MINDy algorithms. Then, for each edge in the interactome we determine the two-dimensional probability distribution of the gene expression levels both in the control state, and following drug treatment. Any changes in the probability distribution are estimated using the Kullback-Leibler (KL) divergence, from which we determine the statistical significance of the dysregulation of each edge. In the second step of DeMAND, we interrogate each gene independently to determine whether its interactions are enriched in dysregulated ones, suggesting that it is a candidate mechanism of action.

#28: itNETZ: The Signature-based Drug Discovery Platform

Hongwei Shao[1], Chenglin Liu[1], Dongmin Guo[1], Caty Chung[2], Tao Peng[1], Stephan Schürer[2], Jing Su[1], and Xiaobo Zhou[1]

1. Department of Radiology, The Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
2. Center for Computational Science, University of Miami School of Medicine, Miami, FL 33136, USA

The primary goal of Wake Forest-LINCS center is to address the needs of multiplex data integration by developing a signature-oriented software platform, the Integrative and Translational Network-based Cellular Signature Analyzer (itNETZ). The itNETZ platform provides pipelines for comprehensive, mechanism-based analysis of drug candidates by incorporating 1) the processing and analyzing toolkits for image data events and Luminex genomics (L1K) data of drug-triggered cellular, 2) the data integration and mining toolkits and data warehouse for cell-specific core pathway signature identification, and 3) the online phosphoprotein signaling network modeling tool for predicting drug efficacy and toxicity. The itNETZ platform thus bridges the gap between the clinical and pharmaceutical needs and the rich LINCS data, and facilitates researchers of diverse backgrounds and interests to explore potentials and mechanisms of drug candidates for specific diseases.

The itNETZ drug discovery pipeline is demonstrated in this poster by a specific pharmaceutical use case: Mechanism-based screening for alternatives of a high-efficacy but high-toxic drug against human lung adenocarcinoma (non-small lung cancer) and predicting optimal dosages. Also illustrated are the usages of the major itNETZ modules, Cell-IA (image processing and analysis), csNMF (L1K data processing and compound signature detection), KIEP (drug efficacy and toxicity modeling), and pLINDAW (pan-LINCS data warehouse for data integration and fuzzy query of similar drugs). Cell line PC9 and primary cell HPH are used as cell models for lung cancer and primary human hepatocytes, respectively, to estimate drug efficacy and toxicity. The Cell-IA image analysis of cell responses to drug candidates in terms of proliferation and apoptosis changes identifies the kinase inhibitor GW843682 of high efficacy against PC9 but also high toxicity for HPH. To find alternative drugs, the molecular signature of GW843682 is then detected using csNMF, and other drug candidates of similar signatures are screened out by fuzzy query using the pLINDAW data warehouse. An ODE model is established by inferring drug and cell specific signaling pathways using LINCS KinomeScan data as well as external knowledge bases IPA and KEGG, and parameters estimated by Cell-IA analysis results and KiNomeScan data. The efficacy and toxicity of candidates are then examined in silico. Drug PF02341066 is discovered of higher efficiency, lower optimal dosage, and better tolerance.

This demonstration shows how the signature-based drug discovery pipeline seamlessly integrates LINCS data sets and external databases using major itNETZ modules to address specific clinical needs. With molecular signature identification, signaling mechanism inferring, and optimal dosing predictions, the itNETZ thus provides researchers and physicians a centralized LINCS “home base” for disease-oriented drug exploring.

#29: Metadata Standards and Annotation Format to Describe the LINCS Assays

Uma Vempati[1], Caty Chung[1], Amar Koleti[1], Chris Mader[1], Cyril Benes[2], David Wrobel[3], Sean Erickson[3], Jeremy Muhlich[4], Gabriel Berriz[4], Caroline Shamu[3], Stephan Schürer[1,5]

[1]Center for Computational Science, University of Miami, Miami, FL; [2]Center for Molecular Therapeutics, Massachusetts General Hospital, Boston MA; [3]ICCB-Longwood Screening Facilty, Harvard Medical School Boston MA; [4]Dept. of Systems Biology, Harvard Medical School Boston MA; [5]Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL.

This poster illustrates several LINCS metadata standards that have been collaboratively developed and their use to annotate and categorize LINCS assays and results. We also introduce the JSON-based Simple Annotation Format (SAF) to exchange assay and result information via web services.

#30: Compound Identity-Based Survey of LINCS Assays and Results

Dušica Vidović[1], Amar Koleti[1], Saminda Abeyruwan[1,2], Uma Vempati[1], Stephan Schürer[1,3]
[1]Center for Computational Science, [2]Department of Computer Science, [3]Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL.

The LINCS compounds, used as biological system perturbagens, were standardized and categorized. The variety of LINCS experimental data was analyzed and integrated by joint perturbagens in order to identify their effects in different cell lines and diseases.

General Information

Some important things to keep in mind:

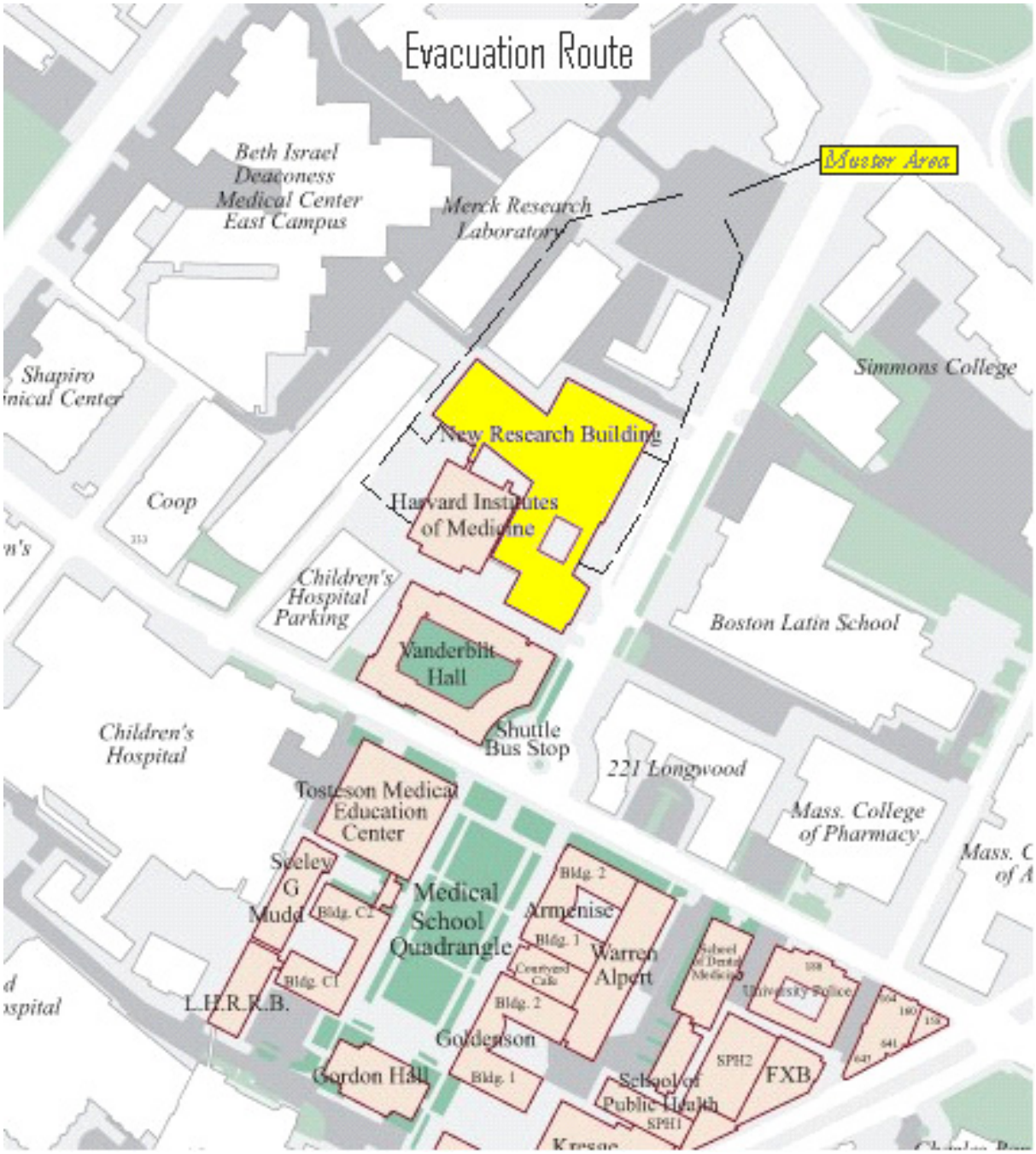
Speakers:

- Please keep an eye on the podium timer and do not go over your allotted time.
- Please bring your presentations or computers to the attention of the AV staff at the front of the room well before the session begins.

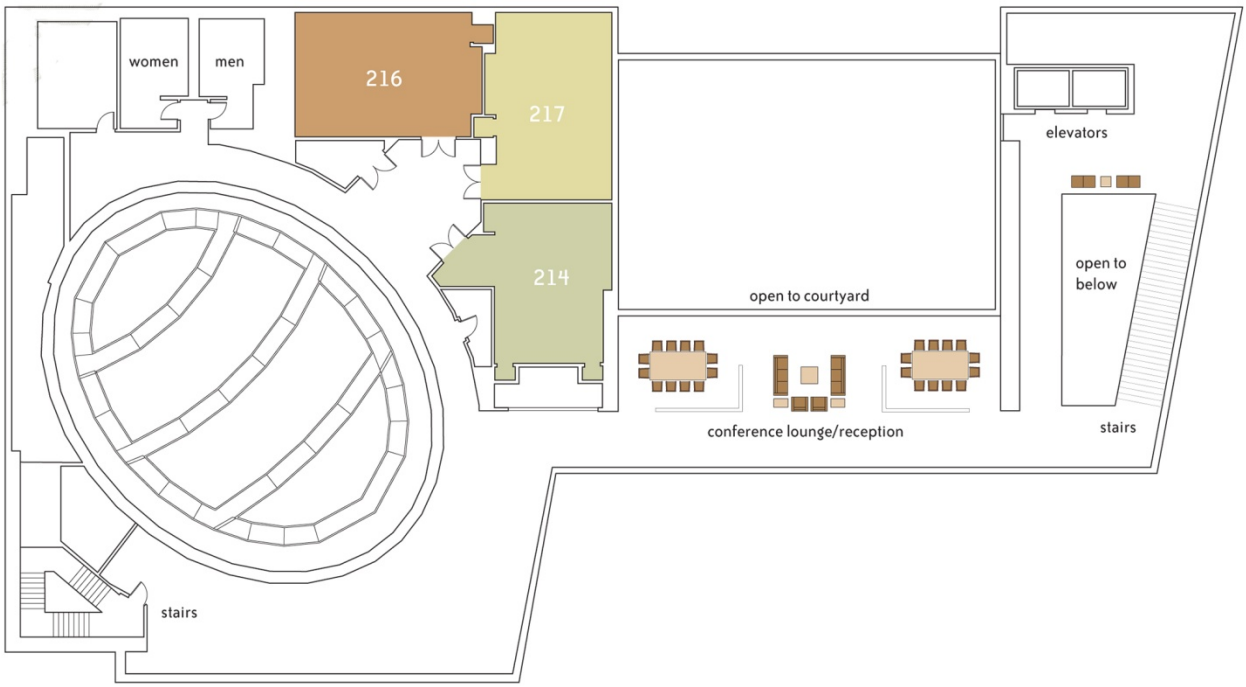
Everyone:

- Please do not stand or sit in the aisles – there is sufficient seating for everyone.
- Please use the microphone when asking questions of the speakers.
- Public wireless access is available in all the conference rooms. Please speak to an AV staff member if you experience any difficulties.
- In case of emergency:
    - Please evacuate the room by following the nearest exit signs.
    - Upon exiting, head to the conference center assembly area (see map on facing page).
    - The assembly area is the Merck parking lot that is located to the left of the building as you exit from the main entrance.
    - Please report to the person wearing the green vest and holding "The Conference Center at Harvard Medical" sign.
    - Await further instructions.

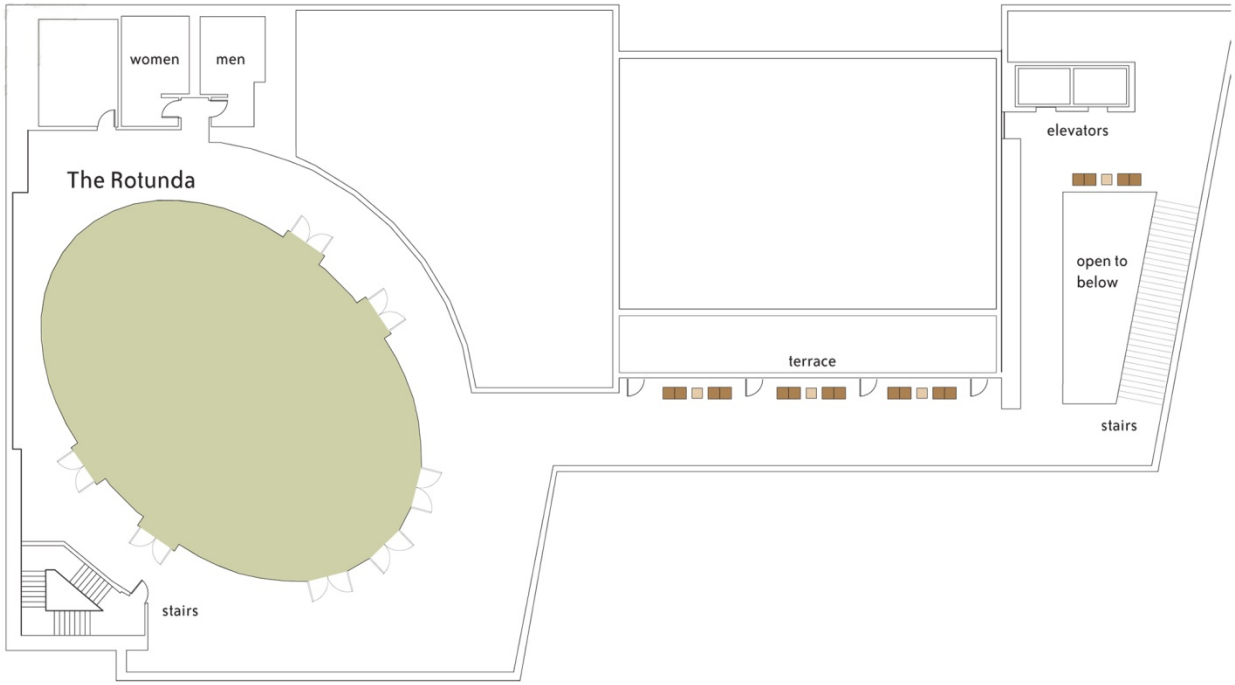
Joseph B. Martin Conference Center Evacuation Route



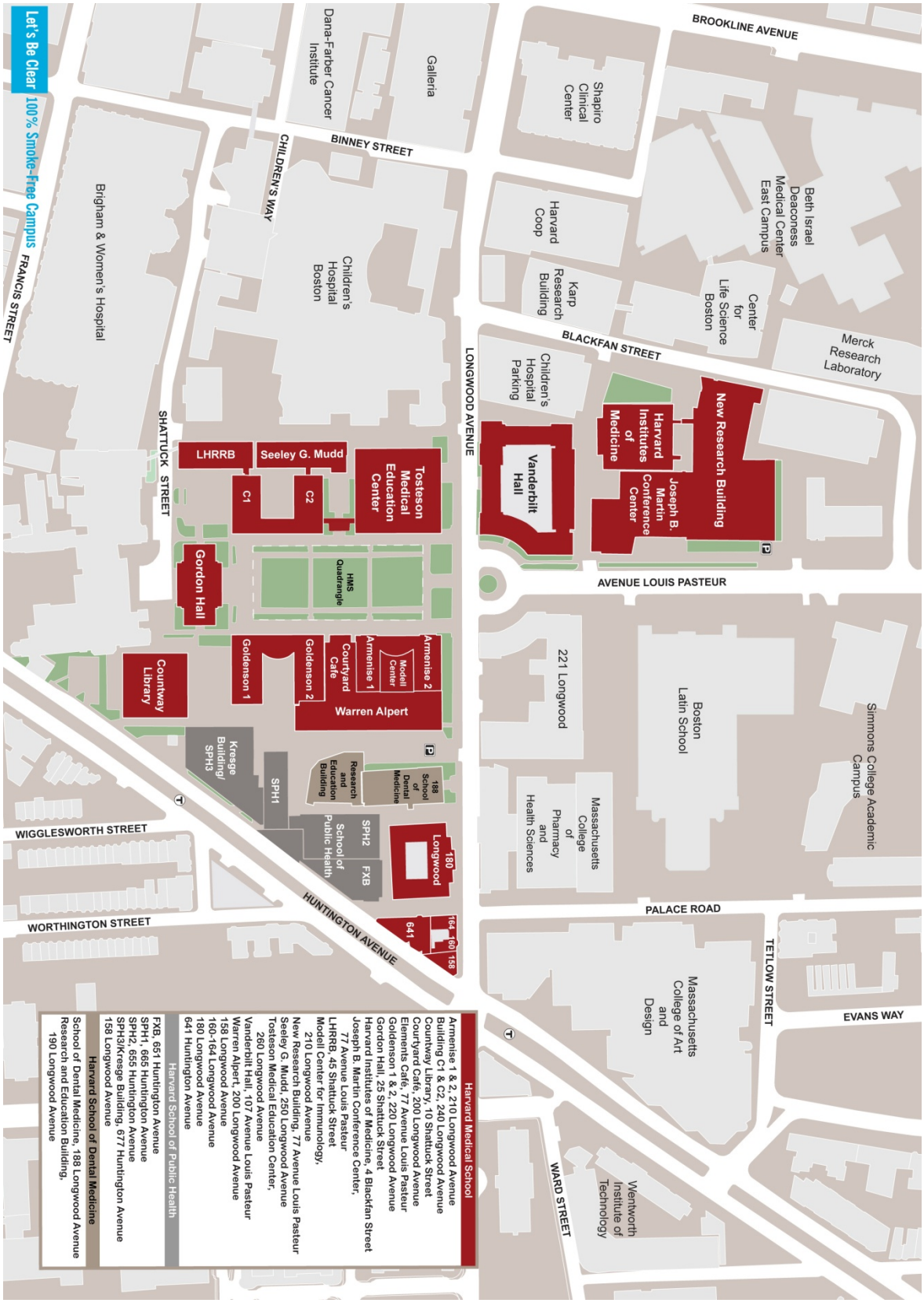
Joseph B. Martin Conference Center 2nd Floor Map



Joseph B. Martin Conference Center 3rd Floor Map



Harvard Medical School/Longwood Medical Area Street Map



Notes