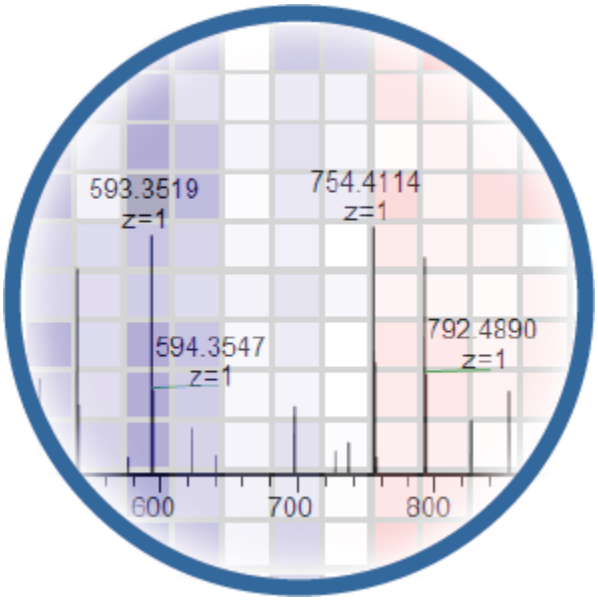
The Library of Integrated Network-based Cellular Signatures

Proteomic Characterization Center
for
Signaling and Epigenetics



Guide to the LINCS PCCSE

NIH LINCS PROGRAM

Contents

What we do

The LINCS Proteomic Characterization Center for Signaling and Epigenetics (PCCSE) is dedicated to understanding the therapeutic potential of drugs and unlocking the fundamental mechanisms of action of drugs, genes, and diseases through the lens of their cellular effects on phosphosignaling and epigenetic processes.

(***motivate***) Two extremely important cellular processes – signaling and the propagation of epigenetic information – are mediated through the deposition of modifications on proteins. These processes often manifest aberrant behavior in the setting of human disease, and modulating them back to more typical functionality is a goal of therapeutic intervention. We believe that having a systematic roadmap of the changes induced by human disease and the impact of therapeutics on signaling and epigenetics would be enabling for the scientific community. We are inspired by the revolution in molecular profiling, and recognize the opportunity to develop similar approaches and resources in the **proteomic** space.

(***build***) Our core technologies allow us to build a resource of molecular profiles – or **signatures** – of the changes induced by cellular perturbations. These signatures reflect the direct observation of changes in protein modifications using a rigorously quantitative approach. Signatures can be compared to one another for the purpose of recognizing convergent cellular mechanisms and novel therapeutic opportunities.

We are using our core technologies to systematically profile the impact of drug and genetic perturbations in a range of disease and developmental models. Our roadmap includes focused studies on cancer, neurodevelopment, neuropsychiatric disease, and cardiovascular disease. To begin building our **resource**, we have selected drugs that are commonly used in therapeutic and research indications for these models. We are also disrupting and editing genes to model loss- and gain-of-function variants that are associated with these models and processes.

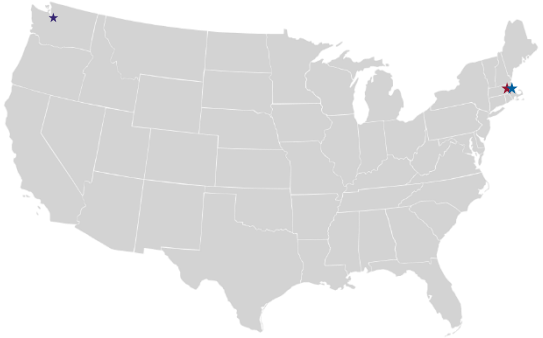
(***enable***) The outcome of these efforts will be a reference **library** of proteomic molecular signatures that can be analyzed in its own right and serve as a basis for comparison to signatures of novel disease models or perturbations. The key means of performing such comparisons is the ability to formally recognize similarity among multiple signatures, a concept we call **connectivity**. More than just correlation, connectivity lets us rank and assess the likelihood that two (or more) perturbations are driving a common cellular response. Here are some examples of questions one might research in the library:

- What is the mechanism of action of a compound?
  - Through connectivity to drugs of known mechanism
  - Through connectivity to genetic perturbations
- Does a drug or gene's function depend on disease model or biological context?
  - Through connectivity of the same perturbation across cell types
- Does a drug activate signaling or epigenetic processes?
  - Through signature strength relative to known examples
- Are there potentially novel or unrecognized therapeutic opportunities?
  - Through *anti*-connectivity of a drug to a genetic model of disease

(***extend***) We are committed to sharing our resource and have made all of our signature data **public**, even prior to publication. We have developed several collaborative models that allow others to help us build the resource, including joint projects and sharing of reagents. And we have collected data in a next-generation comprehensive mass spectrometry format that should allow even deeper interrogation of cellular phosphosignaling, by us and by third parties. We are continuously working to build the tools and infrastructure necessary to share our science with the world.

Who we are

"We are a team dedicated to building out our LINCS resource for the benefit of the scientific community. We have assembled the right core competencies and use the tools of modern scientific collaboration to execute in data production and tool development that enhance the resource's value."



BROAD INSTITUTE

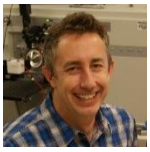
Massachusetts Institute of Technology

UNIVERSITY of WASHINGTON

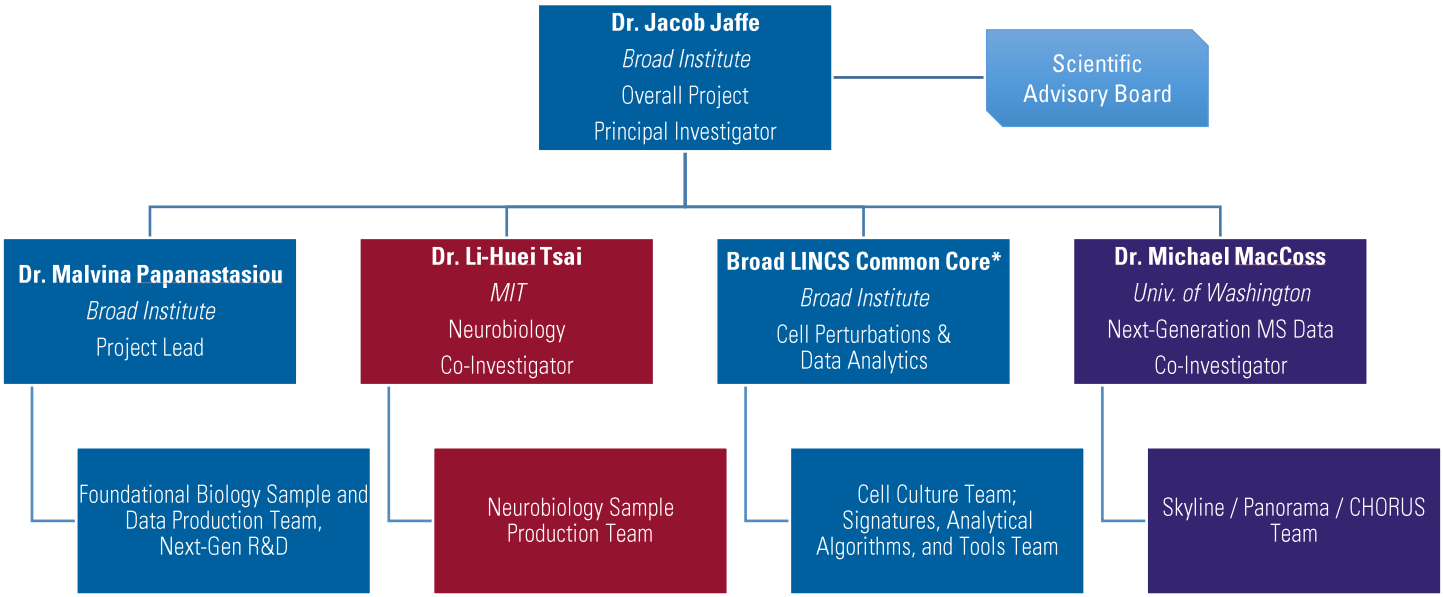
The Investigators

Jacob D. Jaffe, Broad Institute – Dr. Jaffe is the overall director of the LINCS PCCSE and leads its primary scientific direction and operations. He is also the Associate Director of Proteomics at the Broad Institute.

Li-Huei Tsai, MIT – Dr. Tsai leads the neurobiology team and is responsible for the generation and development of neurobiology models for the PCCSE. She is also the Director of the Picower Institute for Learning and Memory and a Professor at MIT.



Michael MacCoss, Univ. of Washington – Dr. MacCoss develops next-generation mass spectrometry data acquisition methodologies and software that are used by the PCCSE. He is also a Professor in the Department of Genetics at the University of Washington.



*about the common core – We take advantage of our unique proximity to the LINCS Center for Transcriptomics to share resources for laboratory work, data analysis, and data integration.

How we do it

The LINCS PCCSE unites multiple cutting-edge technologies to generate data. Bringing together:

- dedicated team members
- high performance mass spectrometry
- laboratory automation
- advanced analytical techniques
- carefully leveraged collaborations

we are able to generate targeted proteomic profiles of phosphosignaling and epigenetic activity from cells at unprecedented scale.

The Team. A truly multidisciplinary team participates in all aspects of experimental planning, sample generation, data analysis, quality control, and continuous process improvement. Team members with diverse backgrounds and skills sets are charged with discovering novel patterns and exciting connections in the data they help to generate. Regular meetings, in-person and via telepresence, keep the distributed group of scientists well-connected.

High Performance Targeted and Comprehensive Mass Spectrometry. Our center assays the signaling states and epigenetic landscapes of perturbed cells using two state-of-the-art proteomic profiling assays.

P100 Phosphoproteomics – The P100 assay measures the levels of ~100 phosphorylation sites on cellular proteins. Changes in their levels are early sentinels of bioactivity of a diverse set of drug mechanisms and altered activities of biological signaling pathways[1].

GCP Global Chromatin Profiling – The GCP assay measures the levels of nearly any post-translational modification on the core nucleosomal histone proteins. These modifications convey epigenetic information in cells and their dysregulation is associated with a wide range of diseases[2,3].

Automation. To achieve this scale of data generation, the LINCS PCCSE relies heavily on laboratory automation. We use the Agilent Bravo as our primary platform, and take special advantage of the AssayMAP microchromatography capabilities of these instruments to execute key biochemical processing steps with a high degree of reproducibility while preparing 96 samples in parallel.

Analysis. We have independently developed the complete analytical pipeline starting from Skyline software to process raw mass spectrometry data through innovative visualization and query tools (in development) that are directly linked to our 100% public data repository. Our Proteomic Signature Pipeline (PSP) ensures that data are treated reproducibly and that changes to processing methods are applied equally on past and future data.

Collaboration. We engage in strategic collaborations that allow us to push the science to new model systems and extend the scope of our data collection efforts. Our special relationship with the LINCS Center for Transcriptomics (co-located at the Broad Institute with Dr. Todd Golub as principal investigator) allows us to collect transcriptional profiles using their L1000 assay for the samples that we generate. Our work with investigators at Tufts Medical Center has allowed us to push into the cardiovascular space, and collaborations with Massachusetts General Hospital have let us study patient-derived cellular models of neuropsychiatric disease. We also make our P100 reagents freely available to academic collaborators with the goal of distributing data generation beyond our center.

1. Abelin et al., Mol. Cell. Proteomics, 2016 **2.** Jaffe et al., Nature Genetics, 2013 **3.** Creech et al., Methods, 2015

What we've accomplished: the first 2 years at a glance

Key Highlights

- Completed compound profiling roadmap goals
  - Demonstrated that the P100 assay is sensitive to perturbations of key signaling pathways
  - Established lineage dependencies for selected epigenetic and signaling functionalities
  - Profiled three compound series with epigenetic, signaling, and neuroactive indications
  - Recapitulated linkage between DNA methylation and histone H3K18 ubiquitination
  - Observed cross-talk between BET and PI3K inhibition for certain chemical matter
- Publication of complete and detailed assay descriptions
- Established CRISPR/Cas9 systems for proteomic screening
- Co-organization of the first ever LINCS Community Outreach meeting (Irvine, CA, 3/2016)

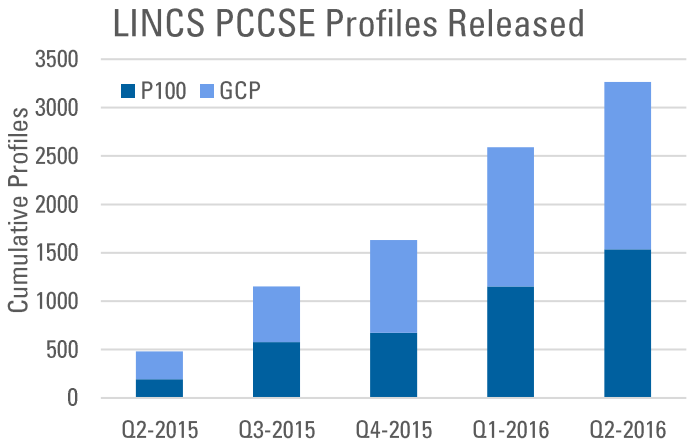
Samples generated

- Core mission: 3400+ biological samples generated spanning:
  - 92 compounds in triplicate
  - 6 cellular models (breast, lung, skin, prostate, and pancreatic cancers; neuronal precursors)
  - 2 assay platforms
- Collaborative projects:
  - Neuronal systems and cardiovascular systems with core mission compounds and other drugs
  - LINCS Consortium Common Project: Drug sensitivity of breast epithelial models

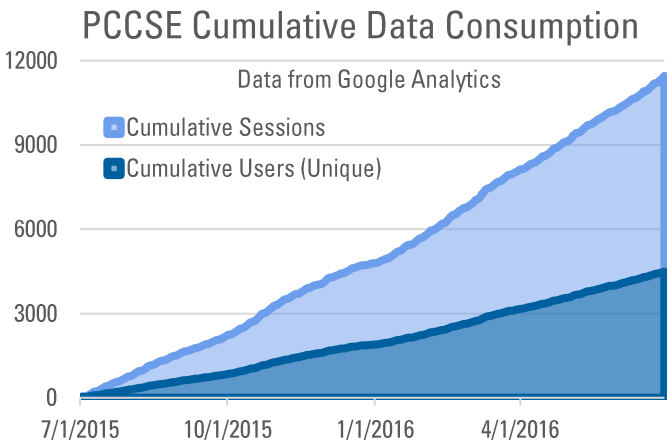
Infrastructure developed

- Completely documented SOPs of sample processing workflows
- Automated sample handling for proteomics sample preparation
- Public Panorama data repository established - http://bit.ly/PCCSEData
- Automated computational pipeline for data QC and normalization
- Interactive heatmap visualization of signature profiles

Data Released



Web-based Usage Statistics



Where we're going – Summer 2016

Our progress thus far has heightened our enthusiasm to extend the project in several key areas:

- Making genetic reference perturbations
- Developing neurobiology models
- Increasing coverage of phosphosignaling pathways monitored through next-gen MS
- Releasing interactive tools

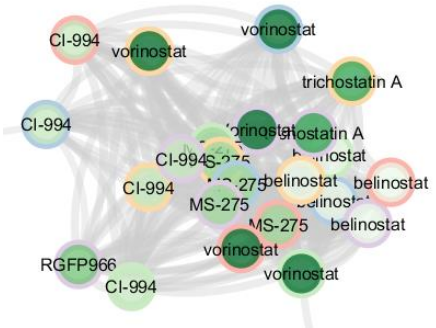
Gene disruption and editing. Through July of 2016 we have focused on studying the responses of cells to drug-induced perturbations. We have now begun to make genetic reference perturbations by gene disruption using the CRISPR/Cas9 system. Our initial focus is to disrupt genes responsible for the deposition, removal, and recognition of epigenetic modifications. The systematic disruption of nearly all of the recognized epigenetic modifier genes across multiple cell types with direct proteomic readouts will be a first-in-kind study. We will benefit from the extensive knowledgebase around CRISPR/Cas9 technology built at the Broad Institute which should help us meet the challenge of scaling a genetic screen for the input requirements of proteomics workflows.

Increased neurobiology focus. Epigenetic processes play an important role in the etiology of many neurological disorders, including autism spectrum disorders (ASDs) and intellectual disability (ID). High-throughput sequencing efforts in ASD patients and their families have found causal mutations in many chromatin-associated enzymes, but only recently has it been possible to study the biological implications of epigenetic dysregulation in human neural cells. With advances in creating human neural cells from pluripotent stem cells, and by applying the CRISPR/Cas9-mediated genomic editing technology, we will be able to disrupt and re-create the expression of curated ASD risk genes.

Our goal is to first investigate the changes in the epigenetic and phosphosignaling landscape through different stages of neural development – e.g., neural progenitors versus mature neurons - and between different cells types such as inhibitory or excitatory neurons, microglia, and astrocytes. Then we will determine the functional impact of disease-associated alterations in epigenetic factors and other genes implicated in neurodevelopmental disorders such as autism.

Next-generation MS profiling for phosphosignaling. Dr. MacCoss's group has pioneered the development of comprehensive, or "DIA" (data-independent acquisition) mass spectrometry. DIA holds great promise for being able to reproducibly characterize and quantify 1000s of analytes across large sets of samples, and avoids the stochastic nature of "shotgun" proteomics. We hope to use these techniques to extend our P100 assay to cover 1000s of phosphosites and map these sites onto existing signaling pathways, thus offering a more biologically intuitive view beyond the sentinels that are currently analyzed. *In fact most of our P100 data has already been collected in a manner that is compatible with DIA analysis.* The key challenge is to develop robust analytical methods to "unlock" these data.

Interactive tools. While all of our data have been publically available since the project's inception, we hope to enhance the value of these data by developing several tools to allow for greater interactivity. The concept of connectivity is key to our understanding of how perturbations are related to one another. To establish connectivity, we "query" the signature profiles of one perturbation vs. all others. We plan to make it easy for data consumers to obtain on-demand lookup of connectivity queries for any perturbations in our core data set, with rank ordered results and significance (*coreQuery*). We will extend this framework to allow users to "bring their own data," querying their own signatures (both real and virtual) vs. our core (*externalQuery*). Results of multiple queries – even across cell and assay types – can be integrated together into networks that lend themselves to intuitive visualizations (*connectivityMaps*). We envision these tools working seamlessly together to create a rich interactive environment to enable data exploration.



Nuts and bolts:

How to collaborate

The PCCSE is eager to collaborate. We work in several collaborative modes:

- **Traditional collaborations** – we have a limited capacity to profile compelling biological samples in a no-cost collaboration. Written proposals are required. Projects are selected by PCCSE leadership.
- **Subsidized data generation** – if you want to generate data using our assay technologies, we may be able to underwrite some of the costs provided that you utilize both GCP and P100 profiling and agree to place any data in the public domain. Degree of subsidy to be determined by PCCSE leadership.
- **Make-your-own P100 data** – we will provide reagents necessary to execute the P100 assay using a DIA MS acquisition scheme, and analyze your data provided that it becomes part of the public domain.
- **NIH grants and subcontracts** – we are happy to contribute to grants and other funding opportunity proposals that wish to take advantage of our LINCS assay technologies, provided that data will be made public in accordance with LINCS data release policies.
- **Service Model** – the LINCS PCCSE exists as a scientific Specialized Service Facility (SSF) which provides a mechanism for charge-back of costs associated with data generation and analysis.

Contact

Jacob D. Jaffe, Ph.D.
jjaffe@broadinstitute.org / @JaffeLab / +1.617.714.7638
The Broad Institute / 415 Main St. / Cambridge, MA / 02142 / USA
Administrative Assistant: Amy Galaviz / agalaviz@broadinstitute.org

URLs

- PCCSE Data repository: http://bit.ly/PCCSEData
- Information on perturbations and assays: http://bit.ly/PCCSEPoster
- NIH LINCS Program: http://www.lincsproject.org

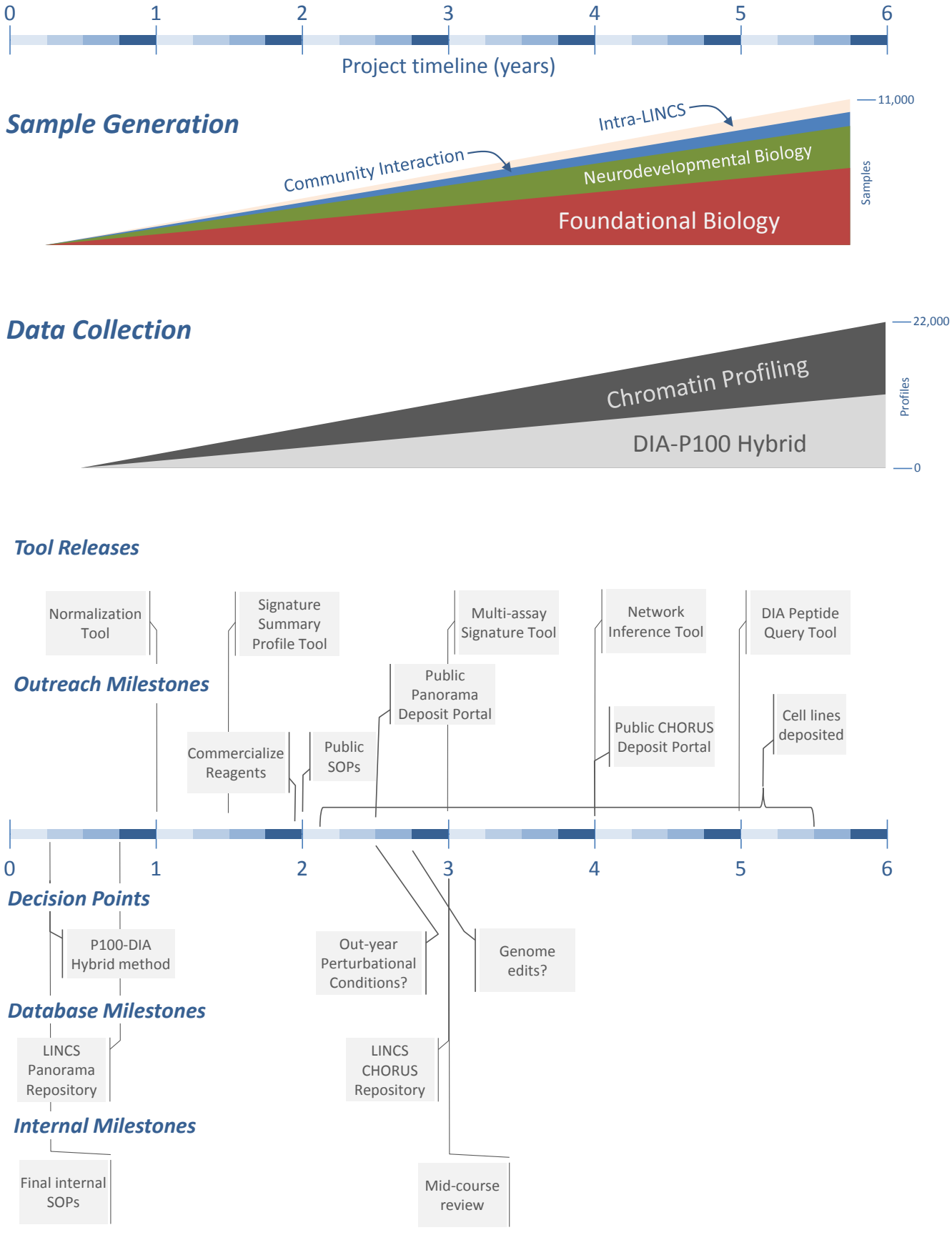
Grant funding

- NIH U54 HG008097-03 (J. Jaffe, PI)
- NIH P50 MH106933-02S1 (I. Kohane, PI)
- NIH U01 CA164186-01 (J. Jaffe, PI; completed)

Scientific Advisory Board

- Dr. Paola Picotti, ETH Zurich, Switzerland
- Dr. Hongjun Song, Johns Hopkins, Baltimore, MD USA
- Dr. Patrick Trojer, Constellation Pharmaceuticals, Cambridge, MA USA

PCCSE Roadmap:



The team:

Broad Institute – Jaffe Lab and Proteomics Platform


Amanda Creech
Data Production


Shawn Egri
Data Production


Adam Officer
Data Production


Dr. Malvina Papanastasiou
Project Lead


Dr. Ryan Peckner
Analysis


Dr. Sebastian Vaca
Methods & Analysis


Janice Williamson
Administration


Dr. Steven Carr
Special Consultant

MIT – Tsai Lab and Stem Cell Core


Fatema Abdurrob
Neurobiology Production


Tak Ko
iPS Core


Dr. Joel Blanchard
Neurobiology Analysis


Dr. Jennie Young
Neurobiology Sr. Scientist

University of Washington – MacCoss Lab


Dr. Jarrett Egertson
DIA Methods and Analysis


Brendan MacLean
Analysis Pipeline


Brian Searle
DIA Methods and Analysis


Vagisha Sharma
Analysis Pipeline


Sonia Ting
DIA Methods and Analysis

Broad Institute – LINCS Common Core


Desiree Davison
Cellular Production


Dr. David Lahr
Analysis


Daniel Lam
Cellular Production


Lev Litichevskiy
Analysis


Xiaodong Lu
Cellular Production


Ted Natoli
Analysis


Dr. Mukta Bagul
Cellular Production


Dr. Aravind Subramanian
Special Consultant