Library of Integrated Network-based Cellular Signatures

Proteomic Characterization Center for Signaling and Epigenetics

Scientific Advisory Board Meeting

October 24, 2016

Jacob D. Jaffe – Broad Institute, PI

Li-Huei Tsai – MIT, co-investigator

Michael MacCoss – U. of Washington, co-investigator

Agenda

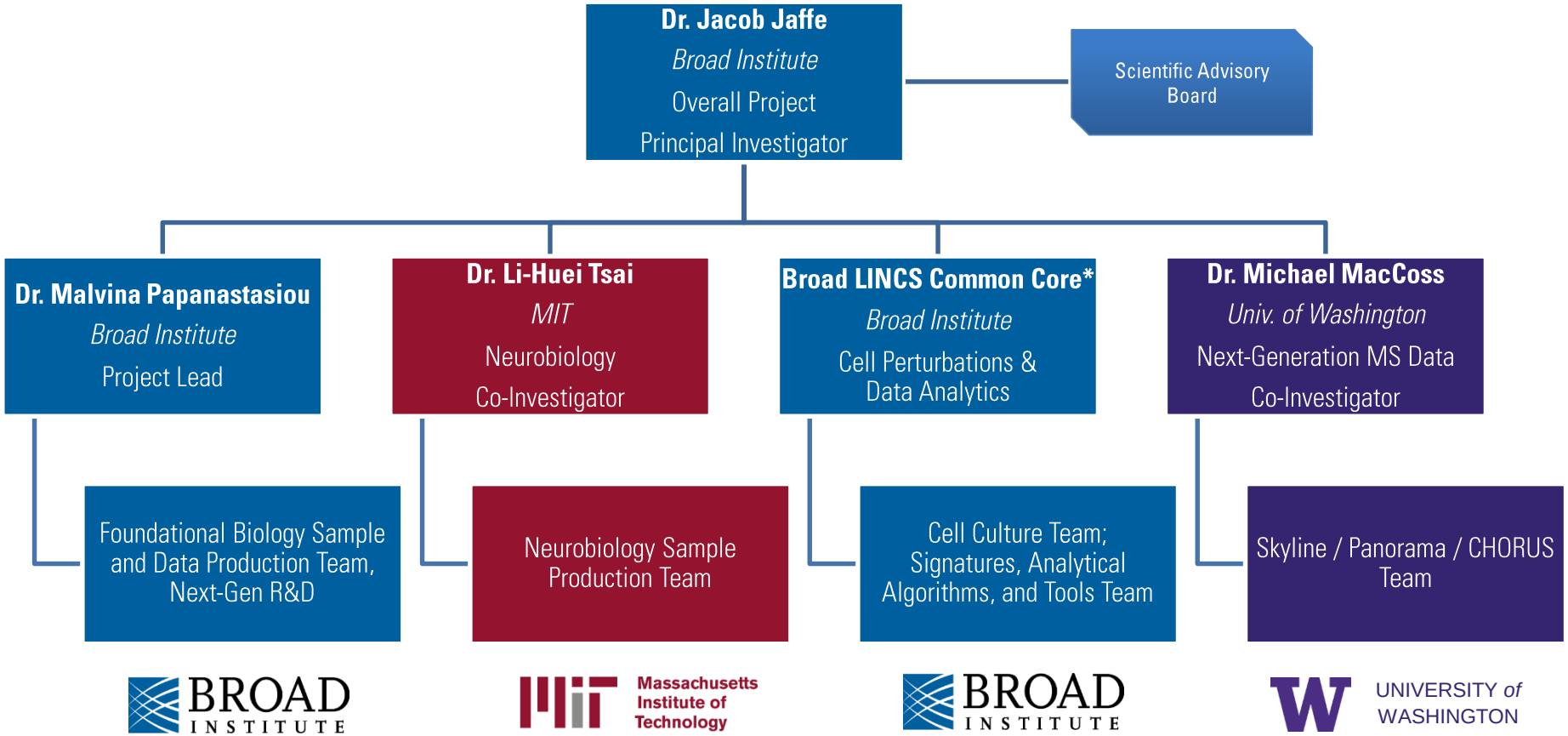
- Brief re-introduction to the LINCS PCCSE (Jake Jaffe, Li-Huei Tsai, and Mike MacCoss)
- NextGen MS Data Sharing (Mike MacCoss)
- Initial feedback from SAB (SAB Members)
- Co-investigators questions for SAB
- Discuss on-site SAB logistics (9-DECEMBER-2016)

Q & A is welcome at any point

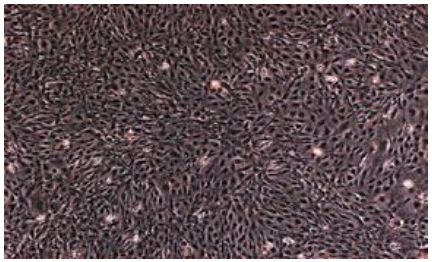
LINCS PCCSE

A BRIEF REINTRODUCTION

Organization of the Center:



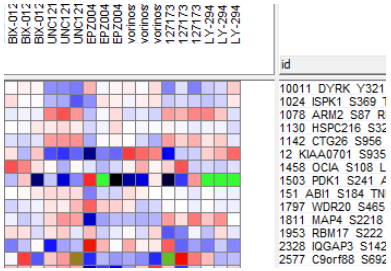
Jaffe Lab – Data Production and Analysis



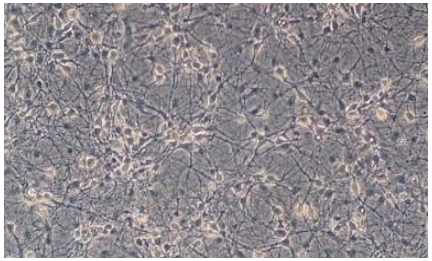
Cell Lines

Drugs

P100 / DIA
(phospho)



Signatures



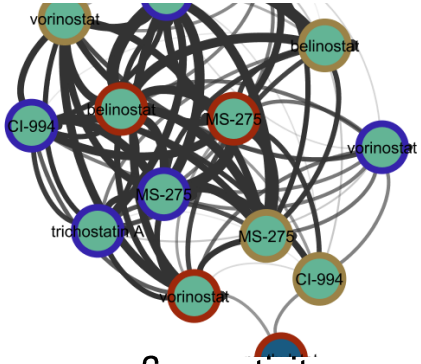
Neural Lineages

X

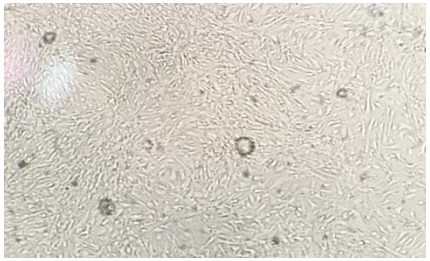
Genes

X

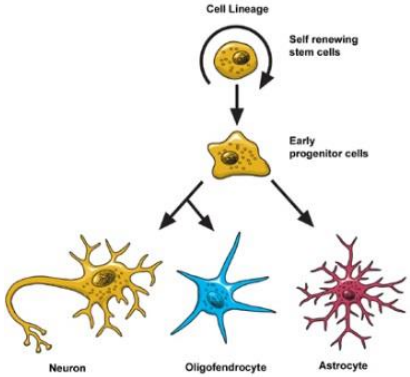
GCP
(histone marks)



Connectivity

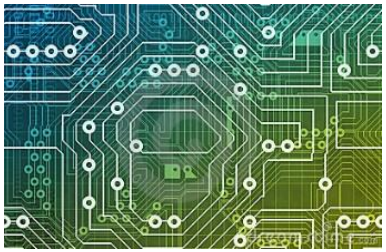


Primary Cells



Fates

L1000
(mRNA)



Circuitry

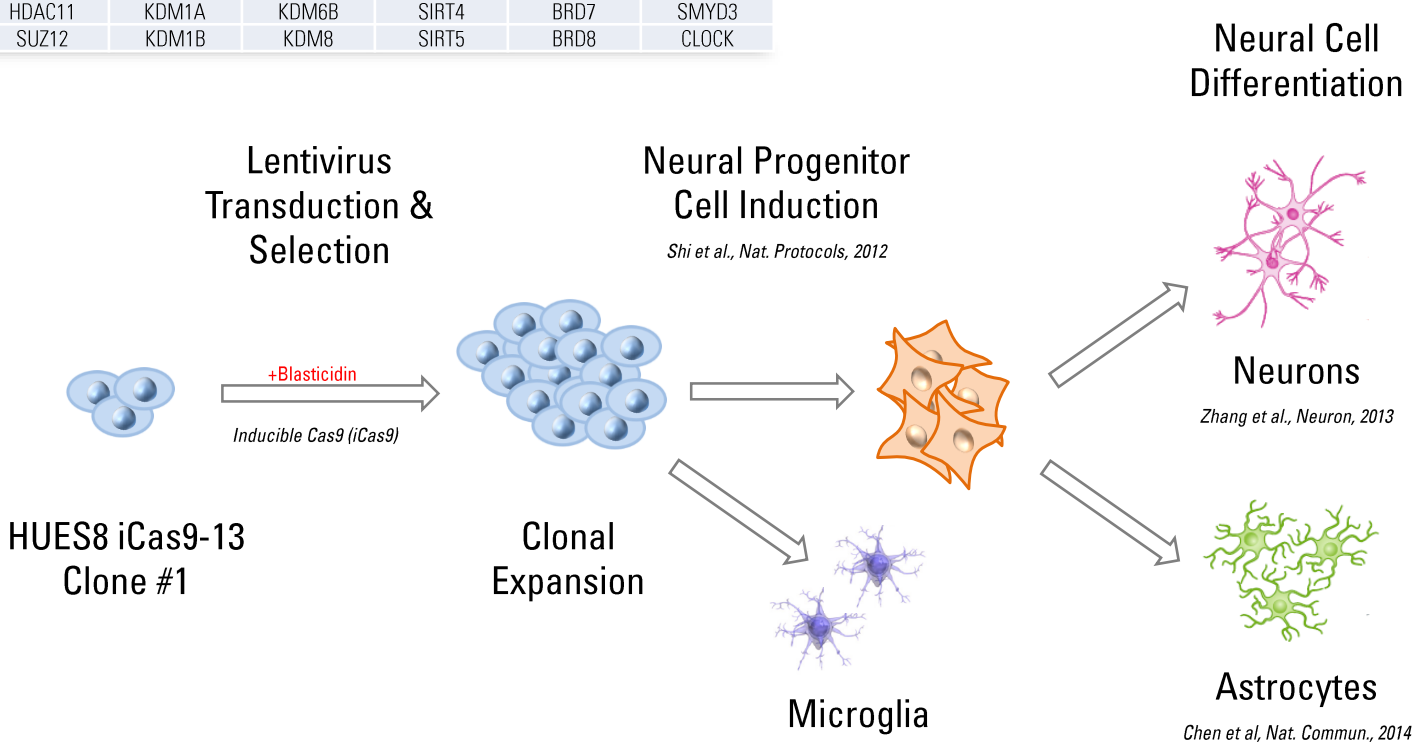
Tsai Lab – Neurobiology Focus

Core Epigenetic Reference Genes

CHD8	HDAC2	KDM2A	KMT2A	SIRT6	BRD9
DNMT1	HDAC3	KDM2B	KMT2D	SIRT7	CBX1
DNMT3A	HDAC4	KDM3A	KMT2C	SUV39H1	CBX2
DNMT3AP1	HDAC5	KDM3B	KMT2E	SUV39H2	CBX3
DNMT3B	HDAC6	KDM4A	SETD1A	SUV420H1	CBX4
DNMT3L	HDAC7	KDM4B	SETD1B	SUV420H2	CBX5
DOT1L	HDAC8	KDM4C	SETD2	WHSC1	CBX6
EHMT1	KAT2A	KDM4D	SETD7	WHSC1L1	CBX7
EHMT2	KAT2B	KDM4E	SETD8	NELFA	CBX8
EZH1	KAT5	KDM5A	SETDB1	ASXL1	EED
EZH2	KAT6A	KDM5B	SETDB2	BRD1	EP300
HAT1	KAT6B	KDM5C	SIRT1	BRD2	NSD1
HDAC1	KAT7	KDM5D	SIRT2	BRD3	SMYD1
HDAC10	KAT8	KDM6A	SIRT3	BRD4	SMYD2
HDAC11	KDM1A	KDM6B	SIRT4	BRD7	SMYD3
SUZ12	KDM1B	KDM8	SIRT5	BRD8	CLOCK

Autism-Spectrum Disorder & Other Neurobiology Genes

ADNP	DYRK1a	MECP2	SEMA5A
ARID1B	EHMT1	NLGN3	SHANK3
CACNA1C	EIF4G1	NLGN4X	SYNGAP1
CNTN4	EPC2	NRXN1	TLK2
CNTNAP2	FMR1	PTCHD1	TSC1
CREBBP	GRIN2B	PTEN	TSC2
CTCF	GRIP1	SCN1A	UBE3a
CTNNB1	GRM5	SCN2a	UBE3b



MacCoss Lab – Next-Gen MS Focus

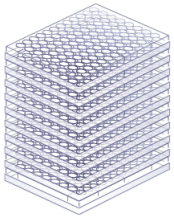
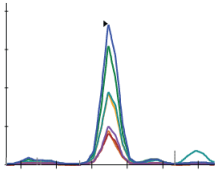
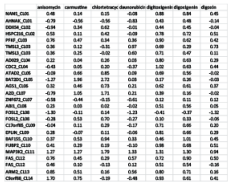
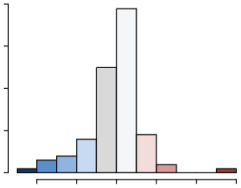
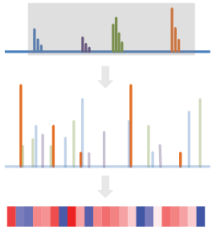
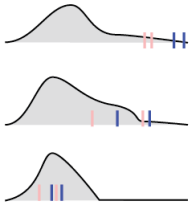
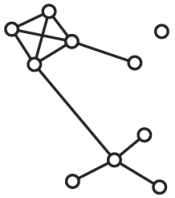
- New data independent acquisition (DIA) strategies
  - Improving the selectivity and dynamic range without sacrificing sensitivity
- New algorithms for peptide-centric analysis of DIA data.
  - Flexibility and sensitivity of discovery analyses
  - No missing data
  - PTM site localization
- Cloud tools for the sharing and dissemination of RAW data
  - http://chorusproject.org as a place to support both storage and analysis of RAW data. A critical component for data reuse and accessibility.
  - http://panoramaweb.org as a location for processed Skyline documents

LINCS PCCSE – MacCoss Lab

NEXT-GEN MS DATA SHARING

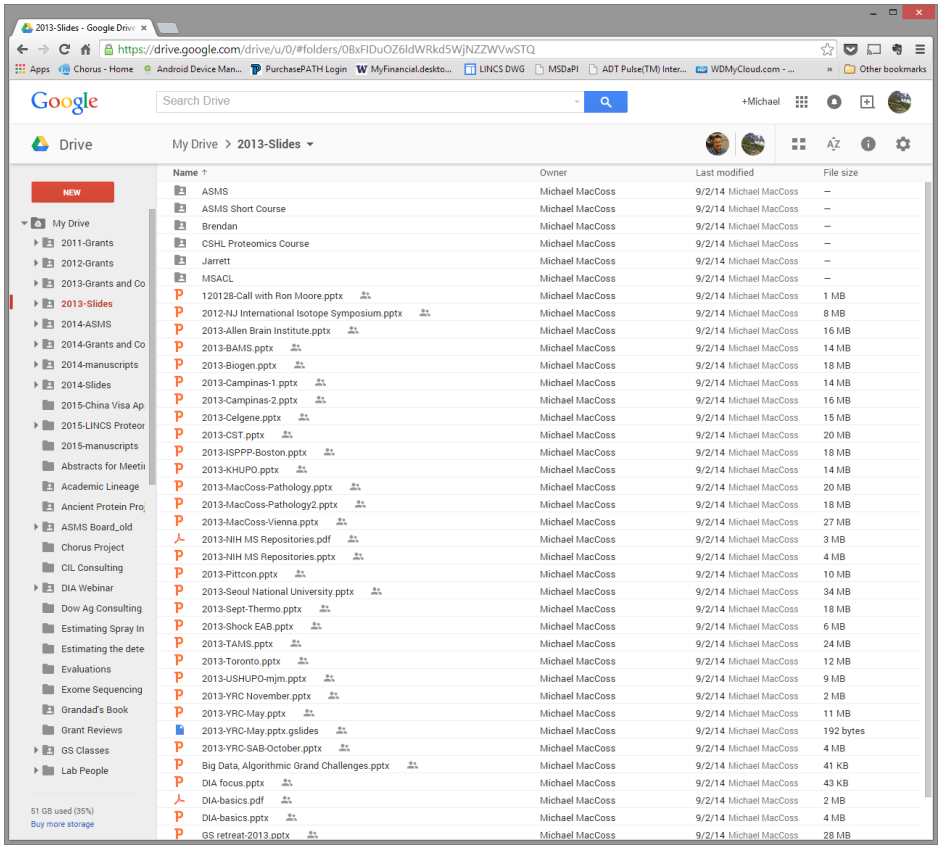
Proteomics Data Sharing Levels

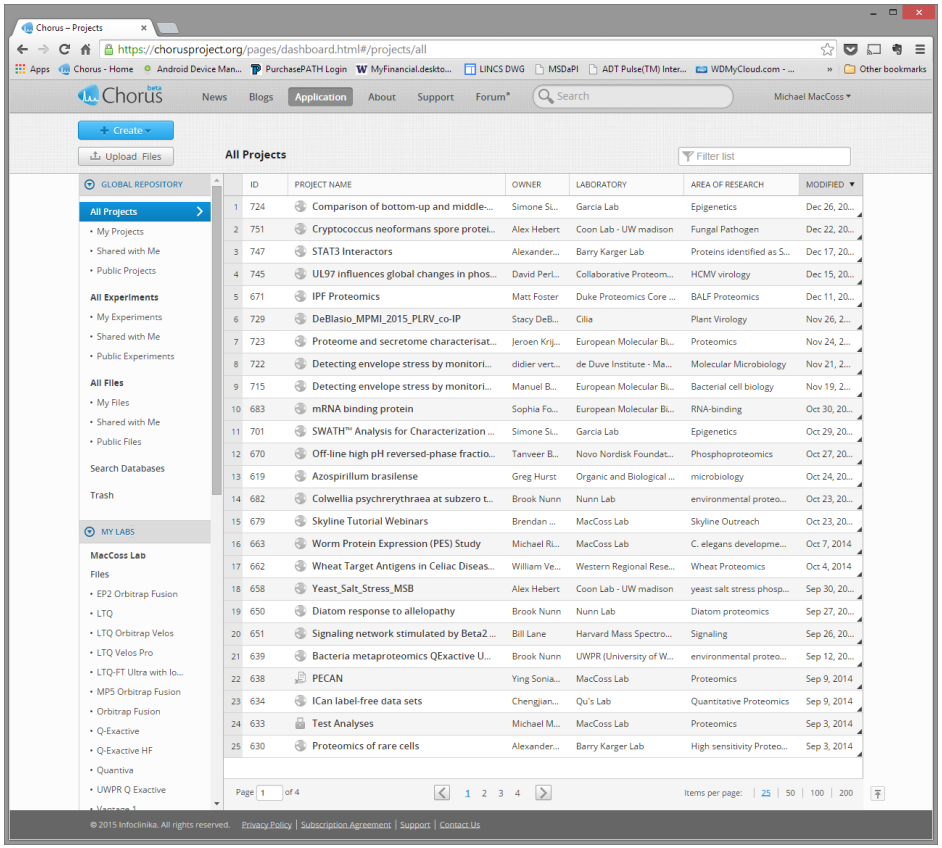
DATA LEVELS AND QUERIES OF LINCS PCCSE DATA

LCMS	PEAK DETECTION	SUMMARIZATION	QC + NORM	DIA QUERY	SIGNATURE QUERY	RELATIONSHIPS
						
LCMS from 96-well plates	Skyline software-mediated peak integration	Combination and reporting of data	Probe and sample outlier removal, normalization	Mine DIA MS data for novel analytes (P100)	Use replicate profiles to identify connections	Compute connectivity across entire signature data set
LCMS	SKY	RPT	QCNORM	DIAQ	SIGQ	GRAPH
LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3	COMING SOON	COMING SOON	COMING SOON
CHORUS	Panorama	Panorama; Morpheus	Panorama; Morpheus			

Chorus: Sharing and Dissemination of RAW Data

- The proteomics data needs to be “un-siloed”
  - A single place for all data
- Data is too large to move around.
- We need to bring algorithms to the data





Chorus allows extensive data organization

Chorus beta | News | Blogs | Application | About | Support | Forum | Search | Nathan Yates

Create | Upload Files

Public Projects | Filter list

GLOBAL REPOSITORY
- All Projects
- My Projects
- Shared with Me
- Public Projects
- All Experiments
- My Experiments
- Shared with Me
- Public Experiments
- All Files
- My Files
- Shared with Me
- Public Files
- Search Databases
- Trash

MY LABS | MY GROUPS | ADMINISTRATION TOOLS | REQUESTS

	ID	PROJECT NAME	OWNER	LABORATORY	AREA OF RESEARCH	MODIFIED
1	599	Bering Sea Sedimentary Proteins	Brook Nunn	UWPR (University of Washi...	environmental proteomics	Aug 18, 2014
2	438	Target Engagement Markers in Cerebrospinal Fluid - JPR ...	Nathan Yates	CHORUS Demo Lab	Proteomics	Aug 14, 2014
3	568	Arabidopsis WT and QR mutant Abscisic Acid Phosphopr...	Benjamin Mink...		Plant Proteomics	Jul 31, 2014
4	522	NeuCode SILAC	Anna Merrill	Coon Lab - UW madison	Proteomics	Jun 30, 2014
5	492	Fu				
6	457	St				
7	303	En				
8	366	Pr				
9	350	Al				
10	385	Al				
11	365	Hi				
12	351	Su				
13	352	Heterogeneous				
14	331	Growth Factor				
15	349	OSU_SIP				
16	345	hippocampi his				
17	325	Regulation of F				
18	134	WRN proteomic				
19	336	Disulfide bond				
20	96	Exhaustive imm				

Public Projects ›

NeuCode SILAC | Filter list

	ID	EXPERIMENT NAME	OWNER	LABORATORY	PROJECT	FILES
1	673	Multiplexing	Anna Merrill	Coon Lab - UW madison	NeuCode SILAC	14
2	672	NeuCode-TMT Comparison	Anna Merrill	Coon Lab - UW madison	NeuCode SILAC	24
3	671	Mixed Ratios	Anna Merrill	Coon Lab - UW madison	NeuCode SILAC	13

Public Projects › NeuCode SILAC ›

NeuCode-TMT Comparison | Filter

	ID	NAME	SIZE	INSTRUMENT	LABORATORY	UPLOAD DATE
1	39648	19Sep2013_4plexNeuCod...	667.91 MB	Charger	Coon Lab - UW madis...	Jun 30, 2014
2	39649	19Sep2013_4plexNeuCod...	654.67 MB	Charger	Coon Lab - UW madis...	Jun 30, 2014
3	39650	19Sep2013_4plexNeuCod...	637.12 MB	Charger	Coon Lab - UW madis...	Jun 30, 2014
4	39645	19Sep2013_4plexNeuCod...	742.07 MB	Charger	Coon Lab - UW madis...	Jun 30, 2014
5	39646	19Sep2013_4plexNeuCod...	700.27 MB	Charger	Coon Lab - UW madis...	Jun 30, 2014
6	39647	19Sep2013_4plexNeuCod...	675.21 MB	Charger	Coon Lab - UW madis...	Jun 30, 2014

Chorus to allow running pipelines in the cloud – nearly ready

- Database Search Engines
  - Comet (SEQUEST) and Percolator

537: With phosphorylation Processing Details

Create New Analysis

1 General Info

Analysis Name
My analysis

Workflow Type
Shotgun

Ratio
All Pairwise
Partitioning b
Factor
Time
Using
24

Description (optio

Create New Analysis

1 General Info

Isolation Width (min 0.5
2

Max Charge (min 1 - max
7

Rt Error Max
2

Mz Error Max
10

Protein Search Range
Overriding the default s
execution time. Howev

Min Rt

Min Mz

Cancel

Create New Analysis

1 General Info

Processi

Database
Mucata new

Precursor Mass Tolerance
20

Modification Parameters (Amino A
M 15.994 --...

Static Modification Parameters (A
C 57.021 --...

Mass Type Fragment
monoisotopic ma...

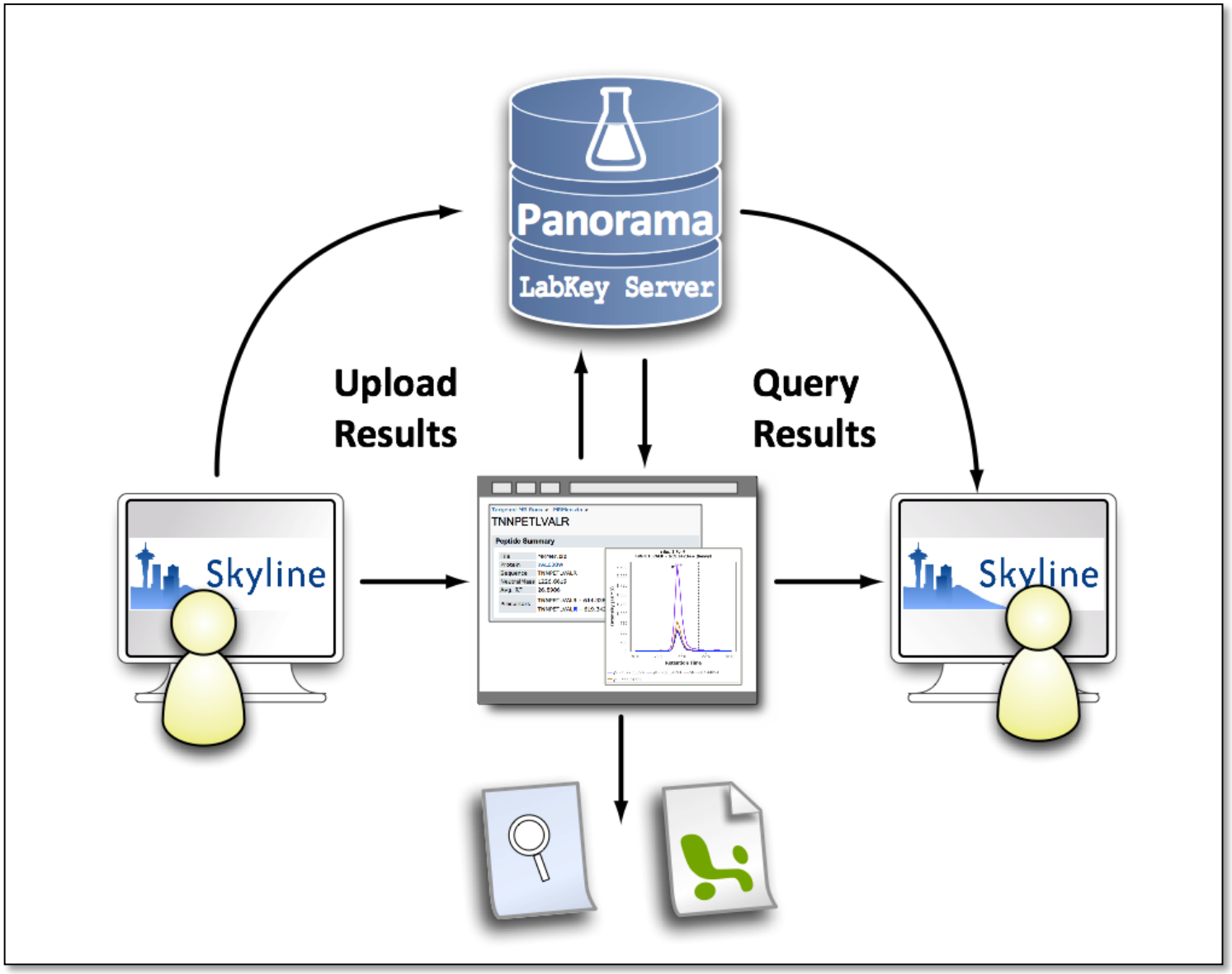
Create New Analysis

1 General Info 2 Processing Params 3 Comet Params 4 Confirm

General	Name	My analysis
	Ratio	All pairwise
	Description	
	Workflow Type	Shotgun
Processing Params	Max Charge	7
	Isolation Width	2
	Search Bounds	Rt (from start - till end), Mz (from start - till end)
Comet Params	Database	Mucata new
	Mass Type Parent	monoisotopic masses
	Precursor Mass Tolerance	20
	Peptide Mass Units	ppm
	Dynamic Modifications	15.9949 M 0 3
	Static Modifications	57.021464 C
	Mass Type Fragment	monoisotopic masses

Cancel Back Run Analysis

Sharing processed data in Panorama



http://bit.ly/PCCSEData

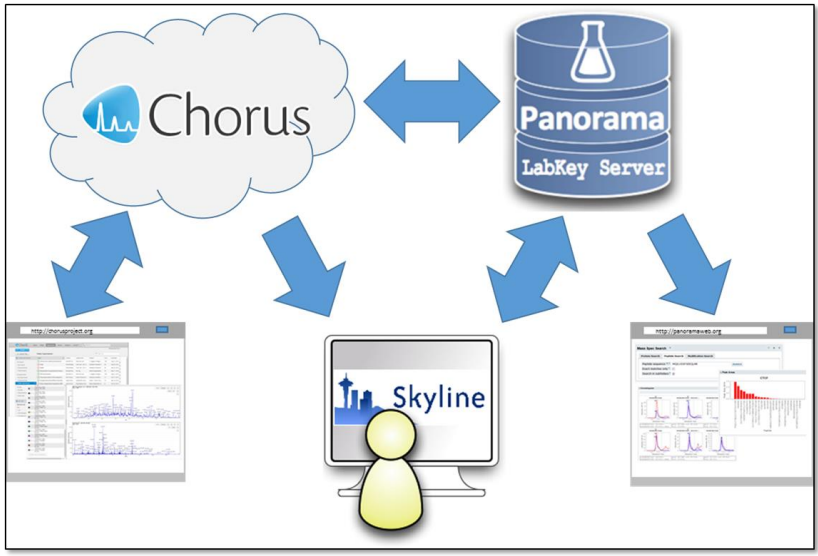
Next-Gen MS summary and outlook

- Comprehensive MS (**DIA**) is already here
  - Acquisition methods will continue to improve further

- We have **promising algorithms** to monitor unprecedented numbers of analytes in LINCS P100 DIA data

- Our Chorus cloud infrastructure will help us to **share** our data and promote **reuse**.

- Chorus will also enable use to efficiently **bring compute to the data** rather than vice versa.



LINCS PCCSE

INITIAL SAB FEEDBACK

Live Notes from Open Discussion

LINCS PCCSE

SEEKING YOUR GUIDANCE

Overall Center Guidance Questions

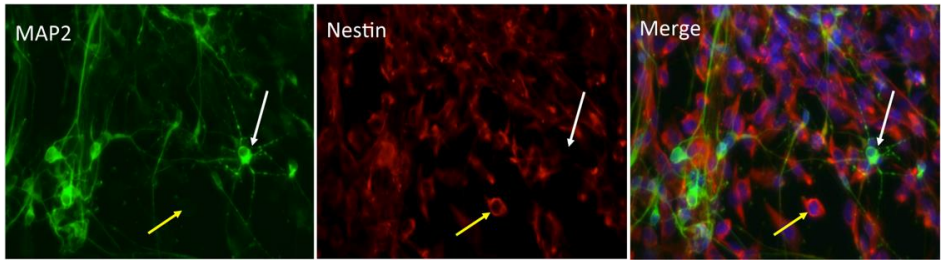
- We have completed our initial run of compound profiling (epi, neuro, and kinase active compounds), and are moving on to gene disruptions.
  - Is this the right focus?
  - Are there key major classes we should consider?

- We make data available at several levels (extracted MS profiles, raw numerical data, and QC'd numerical data) with interactive heatmap visualizations.
  - Coming soon: raw MS data, a profile query tool, and a network viz tool
  - What other types of data availability and tools would be impactful?

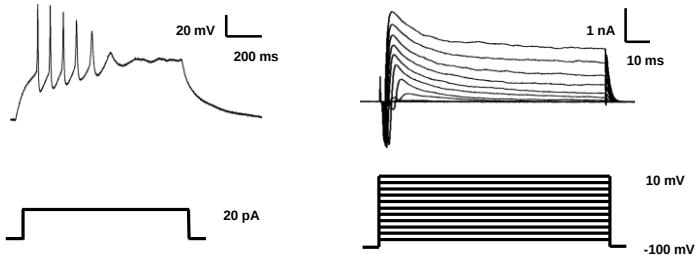
- We are preparing a manuscript describing the project and initial compound data.
  - Suggested focus, message, and journal target?
  - Other publication suggestions?

Neurobiology Guidance Questions

- Creating neurons for high-content proteomics assays
  - Heterogeneous cultures of neurons and NPCs



  - How mature should the neurons be?



- Continue to pursue the CRISPR/Cas9 strategy or focus on perturbagens?
  - Selecting a Cas9-expressing ES cell line
  - Optimizing lentivirus transduction of HUES8 iCas9 NPCs

Next-Gen Proteomics and Dissemination Questions

- We have focused on improving selectivity and dynamic range for DIA data. Are these the correct focus areas?

- What are critical attributes of software tools to drive adoption?
  - Sensitivity? Speed? Ease of use? Support?

- Because of the scale of our data, we believe it is important to enable consumers of our data to bring tools to the data. Are there limitations to this thinking?

- What sort of queries of our data should be enabled? What do you anticipate users of our data doing with our data?

LINCS PCCSE

ON-SITE SAB MEETING (9-DEC-2016)

Mid-course review criteria

1. Feedback on Data Generation & Analysis
2. Feedback on Data Release and Accessibility
3. Feedback on Community Outreach
4. Feedback on major adaptations going forward
5. Other comments/advice

Logistics

- Key details:

  December 9, 2016

  9:00 AM – 2:00 PM

  The Broad Institute

  415 Main St.

  Cambridge, MA 02139

  Dinner to be held the evening prior (December 8th)
- Paola Picotti will attend remotely (Zoom meeting)
- Amy Galaviz can help with any travel arrangements and other questions

  agalaviz@broadinstitute.org

  +1 617 714 7631
- Agenda will mirror review criteria
- We will provide time to draft your feedback report