

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

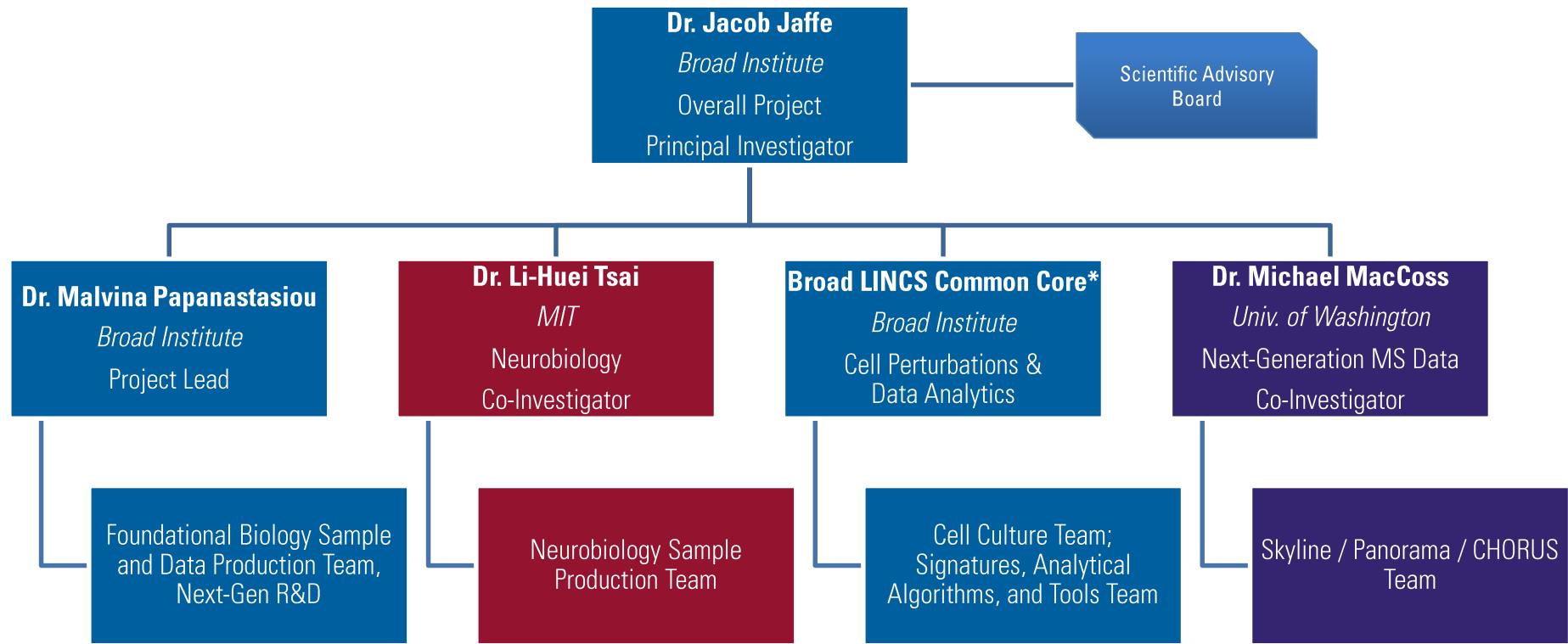
- Brief re-introduction to the LINCS PCCSE (Jake Jaffe, Li-Huei Tsai, and Mike MacCoss)
- NextGen MS Data Sharing (Mike MacCoss)
- Initial feedback from SAB (SAB Members)
- Co-investigators questions for SAB
- Discuss on-site SAB logistics (9-DECEMBER-2016)

Q & A is welcome at any point

LINCS PCCSE

A BRIEF REINTRODUCTION

Organization of the Center:



Jaffe Lab – Data Production and Analysis



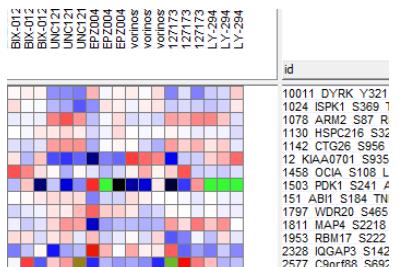
Cell Lines



Drugs



P100 / DIA
(phospho)

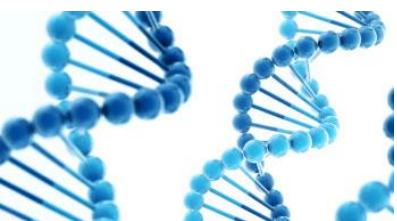


Signatures



Neural Lineages

X

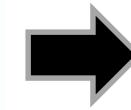


Genes

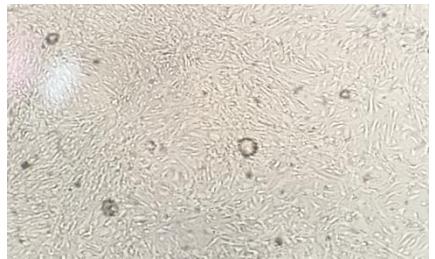
X



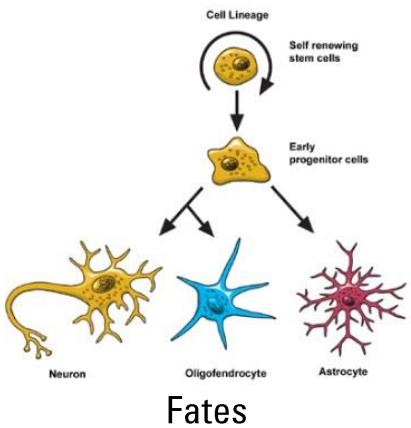
GCP
(histone marks)



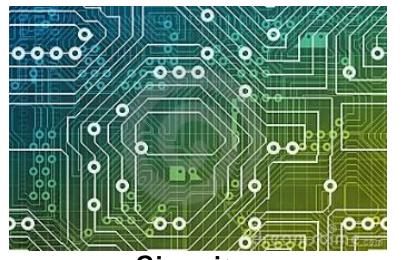
Connectivity



Primary Cells



L1000
(mRNA)



Circuitry

Tsai Lab – Neurobiology Focus

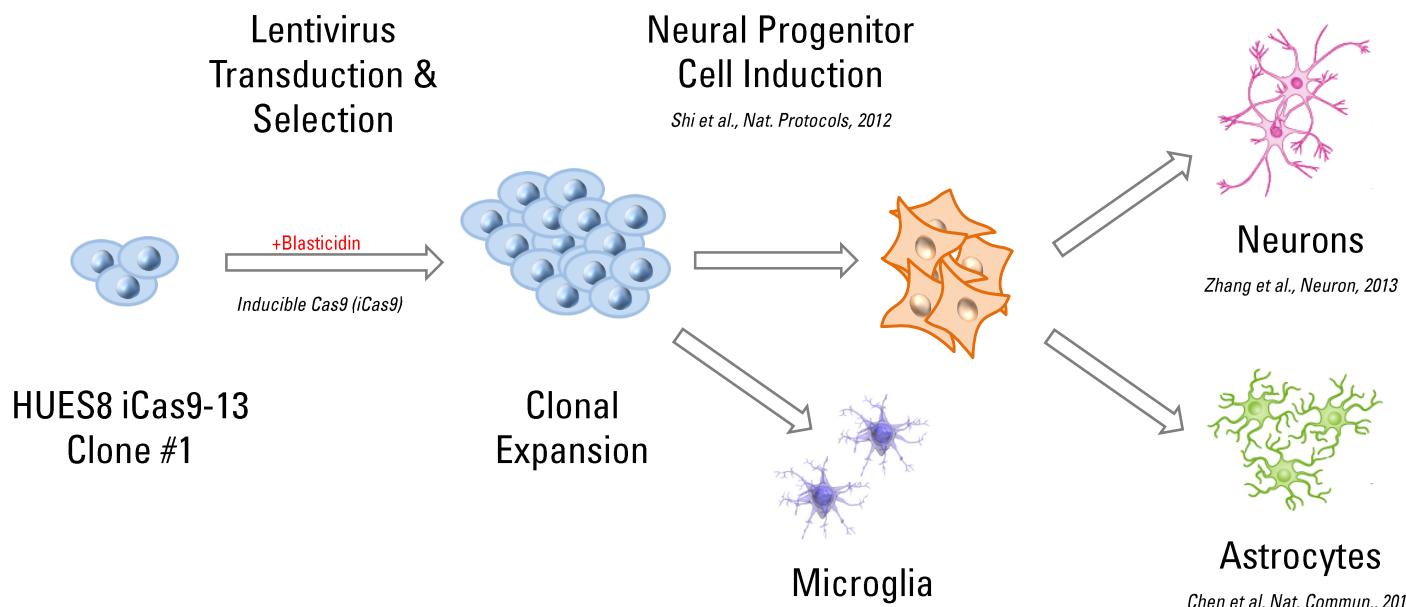
Core Epigenetic Reference Genes

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

Neural Cell Differentiation



MacCoss Lab – Next-Gen MS Focus

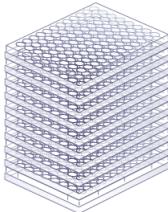
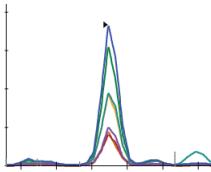
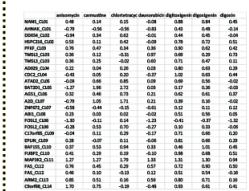
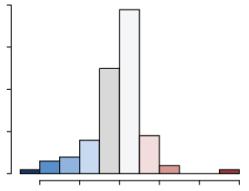
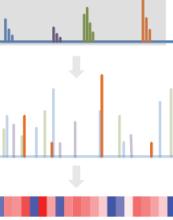
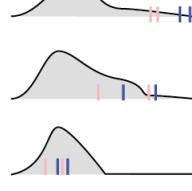
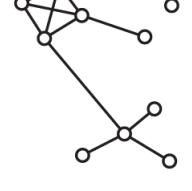
- New data independent acquisition (DIA) strategies
 - Improving the selectivity and dynamic range without sacrificing sensitivity
- New algorithms for peptide-centric analysis of DIA data.
 - Flexibility and sensitivity of discovery analyses
 - No missing data
 - PTM site localization
- Cloud tools 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	 QC/NORM	 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

My Drive > 2013-Slides					
Name	Owner	Last modified	File size	More	
ASMS	Michael MacCoss	9/2/14	Michael MacCoss	–	
ASMS Short Course	Michael MacCoss	9/2/14	Michael MacCoss	–	
Brendan	Michael MacCoss	9/2/14	Michael MacCoss	–	
CSHL Proteomics Course	Michael MacCoss	9/2/14	Michael MacCoss	–	
Jarrett	Michael MacCoss	9/2/14	Michael MacCoss	–	
MSACL	Michael MacCoss	9/2/14	Michael MacCoss	–	
P 120128-Call with Ron Moore.pptx	Michael MacCoss	9/2/14	Michael MacCoss	1 MB	
P 2012-NJ International Isotope Symposium.pptx	Michael MacCoss	9/2/14	Michael MacCoss	8 MB	
P 2013-Alien Brain Institute.pptx	Michael MacCoss	9/2/14	Michael MacCoss	16 MB	
P 2013-BAMS.pptx	Michael MacCoss	9/2/14	Michael MacCoss	14 MB	
P 2013-Biogen.pptx	Michael MacCoss	9/2/14	Michael MacCoss	18 MB	
P 2013-Campinas-1.pptx	Michael MacCoss	9/2/14	Michael MacCoss	14 MB	
P 2013-Campinas-2.pptx	Michael MacCoss	9/2/14	Michael MacCoss	16 MB	
P 2013-Celgene.pptx	Michael MacCoss	9/2/14	Michael MacCoss	15 MB	
P 2013-CST.pptx	Michael MacCoss	9/2/14	Michael MacCoss	20 MB	
P 2013-ISPP-Boston.pptx	Michael MacCoss	9/2/14	Michael MacCoss	18 MB	
P 2013-KHUPD.pptx	Michael MacCoss	9/2/14	Michael MacCoss	14 MB	
P 2013-MacCoss-Pathology.pptx	Michael MacCoss	9/2/14	Michael MacCoss	20 MB	
P 2013-MacCoss-Pathology2.pptx	Michael MacCoss	9/2/14	Michael MacCoss	18 MB	
P 2013-MacCoss-Vienna.pptx	Michael MacCoss	9/2/14	Michael MacCoss	27 MB	
P 2013-NIH MS Repositories.pdf	Michael MacCoss	9/2/14	Michael MacCoss	3 MB	
P 2013-NIH MS Repositories.pptx	Michael MacCoss	9/2/14	Michael MacCoss	4 MB	
P 2013-Pittcon.pptx	Michael MacCoss	9/2/14	Michael MacCoss	10 MB	
P 2013-Seoul National University.pptx	Michael MacCoss	9/2/14	Michael MacCoss	34 MB	
P 2013-Sept-Thermo.pptx	Michael MacCoss	9/2/14	Michael MacCoss	18 MB	
P 2013-Shock EAB.pptx	Michael MacCoss	9/2/14	Michael MacCoss	6 MB	
P 2013-TAMS.pptx	Michael MacCoss	9/2/14	Michael MacCoss	24 MB	
P 2013-Toronto.pptx	Michael MacCoss	9/2/14	Michael MacCoss	12 MB	
P 2013-USPPO-mjpm.pptx	Michael MacCoss	9/2/14	Michael MacCoss	9 MB	
P 2013-YRC November.pptx	Michael MacCoss	9/2/14	Michael MacCoss	2 MB	
P 2013-YRC-November.pptx	Michael MacCoss	9/2/14	Michael MacCoss	11 MB	
P 2013-YRC-May.pptx	Michael MacCoss	9/2/14	Michael MacCoss	192 bytes	
P 2013-YRC-May.pptx.gslides	Michael MacCoss	9/2/14	Michael MacCoss	–	
P 2013-YRC-SAB-October.pptx	Michael MacCoss	9/2/14	Michael MacCoss	4 MB	
P Big Data, Algorithmic Grand Challenges.pptx	Michael MacCoss	9/2/14	Michael MacCoss	41 KB	
P DIA focus.pptx	Michael MacCoss	9/2/14	Michael MacCoss	43 KB	
P DIA-basics.pdf	Michael MacCoss	9/2/14	Michael MacCoss	2 MB	
P DIA-basics.pptx	Michael MacCoss	9/2/14	Michael MacCoss	4 MB	
P GS retreat-2013.pptx	Michael MacCoss	9/2/14	Michael MacCoss	28 MB	

Chorus - Projects

https://chorusproject.org/pages/dashboard.html#/projects/all

Apps Chorus - Home Android Device Man... PurchasePATH Login MyFinancial.deskt... LINCS DWG MSDsPI ADT Pulse(TM) Inter... WDMyCloud.com - ... Other bookmarks

Chorus

News Blogs Application About Support Forum* Search Michael MacCoss*

+ Create Upload Files

All Projects Filter list

ID	PROJECT NAME	OWNER	LABORATORY	AREA OF RESEARCH	MODIFIED
1	724 Comparison of bottom-up and middle...	Simone St...	Garcia Lab	Epigenetics	Dec 26, 20...
2	751 Cryptococcus neofmans spor proteo...	Alex Hebert	Coon Lab - UW madison	Fungal Pathogen	Dec 22, 20...
3	747 STAT3 Interactors	Alexander...	Barry Karger Lab	Proteins identified as S...	Dec 17, 20...
4	745 UL97 influences global changes in phos...	David Perl...	Collaborative Proteom...	ICMV virology	Dec 15, 20...
5	671 iPF Proteomics	Matt Foster	Duke Proteomics Core ...	BALF Proteomics	Dec 11, 20...
6	729 DeBlasio_MPML_2015_PLRV_co-IP	Stacy DeB...	Cilia	Plant Virology	Nov 26, 2...
7	723 Proteome and secretome characterisat...	Jeroen Krij...	European Molecular Biol...	Proteomics	Nov 24, 2...
8	722 Detecting envelope stress by monitori...	didier vert...	de Duve Institute - Ma...	Molecular Microbiology	Nov 21, 2...
9	715 Detecting envelope stress by monitori...	Manuel B...	European Molecular Biol...	Cellular biology	Nov 19, 2...
10	683 mRNA binding protein	Sophia Fo...	European Molecular Biol...	RNA-binding	Oct 30, 20...
11	701 SWAT™ Analysis for Characterization ...	Simone St...	Garcia Lab	Epigenetics	Oct 29, 20...
12	670 Off-line high pH reversed-phase fractio...	Tanveer Ba...	Novo Nordisk Foundat...	Phosphoproteomics	Oct 27, 20...
13	619 Azospirillum brasiliense	Greg Hurst	Organic and Biological ...	microbiology	Oct 24, 20...
14	682 Colwellia psychrerythraea at subzero L...	Brook Nunn	Nunn Lab	environmental proteo...	Oct 23, 20...
15	679 Skyline Tutorial Webinars	Brendan ...	MacCoss Lab	Skyline Outreach	Oct 23, 20...
16	663 Worm Protein Expression (PES) Study	Michael Ri...	MacCoss Lab	C. elegans developme...	Oct 7, 2014
17	662 Wheat Target Antigens in Celiac Disease...	William Ve...	Western Regional Rese...	Wheat Proteomics	Oct 4, 2014
18	658 Yeast_Salt_Stress_MSB	Alex Hebert	Coon Lab - UW madison	yeast salt stress phospho...	Sep 30, 20...
19	650 Diatom response to allelopathy	Brook Nunn	Nunn Lab	Diatom proteomics	Sep 27, 20...
20	651 Signalling network stimulated by Beta2...	Bill Lane	Harvard Mass Spectro...	Signaling	Sep 26, 20...
21	639 Bacteria metaproteomics QExactive U...	Brook Nunn	UWPR (University of W...	environmental proteo...	Sep 12, 20...
22	638 PEACAN	Ying Sonia...	MacCoss Lab	Proteomics	Sep 9, 2014
23	634 iCAN label-free data sets	Chengjian...	Qu's Lab	Quantitative Proteomics	Sep 9, 2014
24	633 Test Analyses	Michael M...	MacCoss Lab	Proteomics	Sep 3, 2014
25	630 Proteomics of rare cells	Alexander...	Barry Karger Lab	High sensitivity Proteo...	Sep 3, 2014

Page 1 of 4

Items per page: 25 | 50 | 100 | 200

© 2015 Infolinkinc. All rights reserved. Privacy Policy Subscription Agreement Support Contact Us

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

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

537: With phosphorylation Processing Details

Create New Analysis

General Info

Analysis Name: My analysis

Workflow Type: Shotgun

Ratio: All Pairwise

Factor: Time

Using: 24

Description (optional):

Isolation Width (min 0.5): 2

Max Charge (min 1 - max 7): 7

Rt Error Max: 2

Mz Error Max: 10

Protein Search Range: Overriding the default search range to reduce execution time. However, this may affect results.

Min Rt:

Min Mz:

Cancel

Create New Analysis

General Info

Create New Analysis

General Info

Processing

Database

Mucata new

Precursor Mass Tolerance

20

Modification Parameters (Amino Acid)

M 15.994

Static Modification Parameters (Amino Acid)

C 57.021

Mass Type Fragment

monoisotopic ma...

Cancel

Create New Analysis

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

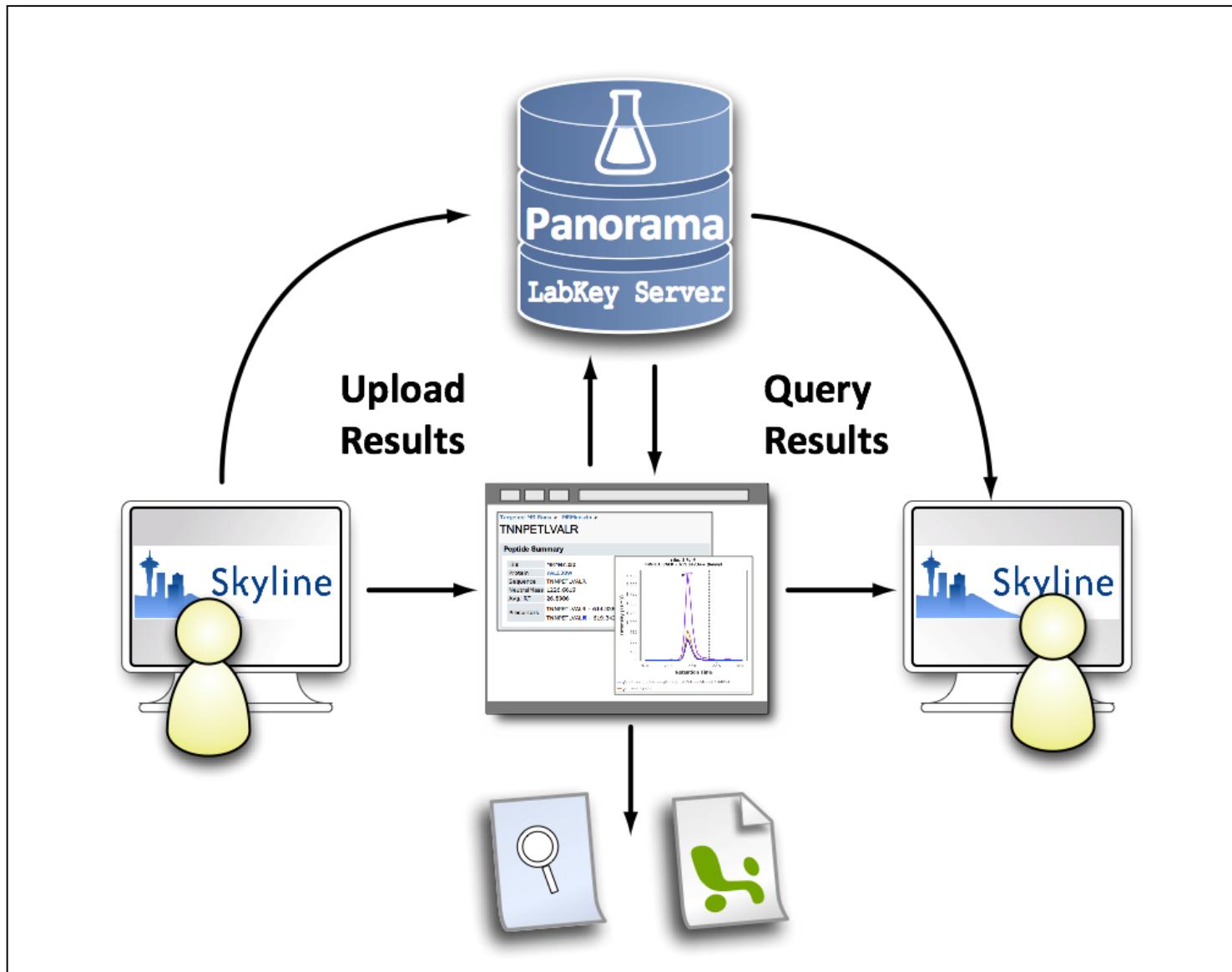
Back

Run Analysis

Database Search Engines

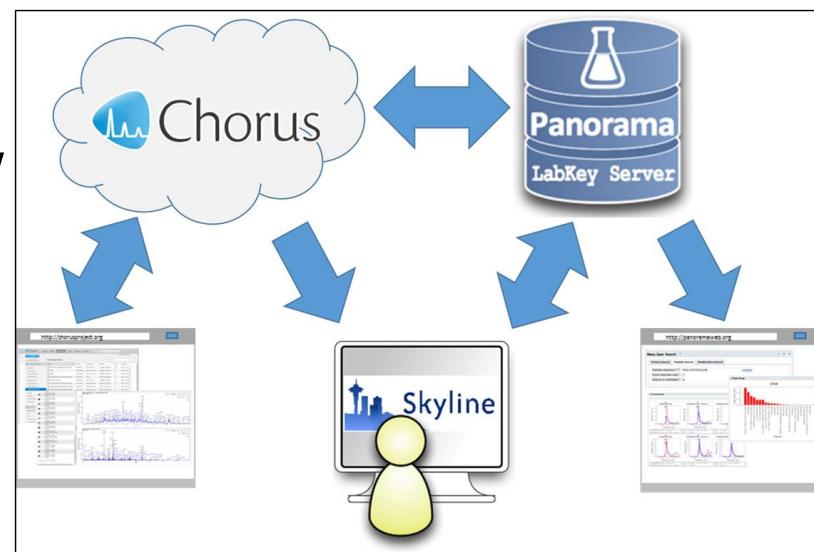
- Comet (SEQUEST) and Percolator

Sharing processed data in Panorama



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

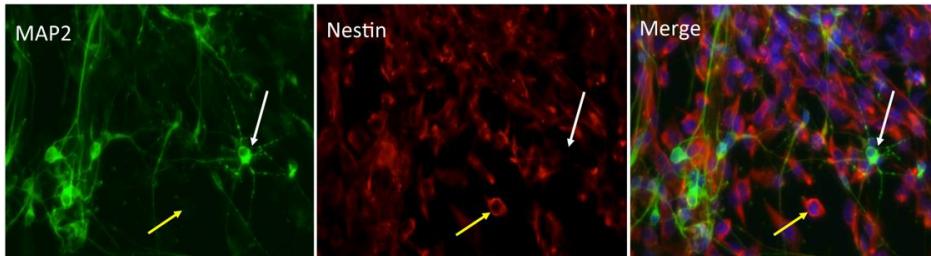
SEEKING YOUR GUIDANCE

Overall Center Guidance Questions

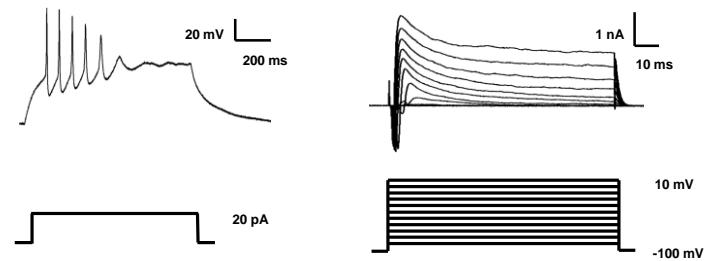
- We have completed our initial run of compound profiling (epi, neuro, and kinase active compounds), and are moving on to gene disruptions.
 - Is this the right focus?
 - Are there key major classes we should consider?
- We make data available at several levels (extracted MS profiles, raw numerical data, 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