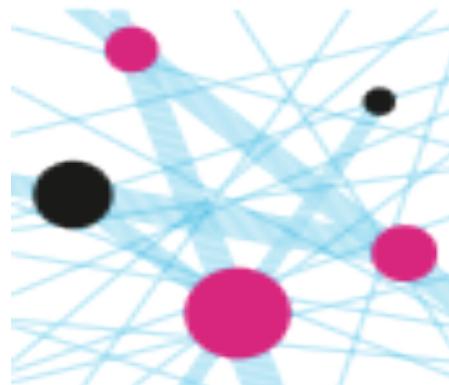


Reproducibility of research results in proteomics

Ruedi Aebersold, Ph.D

*Institute of Molecular Systems Biology, ETH-Zürich;
Faculty of Science, University of Zürich*



Outline

- We need to be able to generate reproducible proteomic research results:
 - for political reasons
 - for scientific reasons
- If two or more people have the same proteomic data, can they generate the same results?
- If two or more people have the same sample, can they generate the same proteomic results?
- If two or more people do the same experiment in a cell line, do they get the same results?

Lack of Reproducibility: A serious threat to Science

Studies show only 10% of published science articles are reproducible. What is happening?

Posted on [May 3, 2012](#) by [Moshe Pritsker](#)

Studies show a very low reproducibility for articles published in scientific journals, often as low as 10-30%. Here is a partial list:

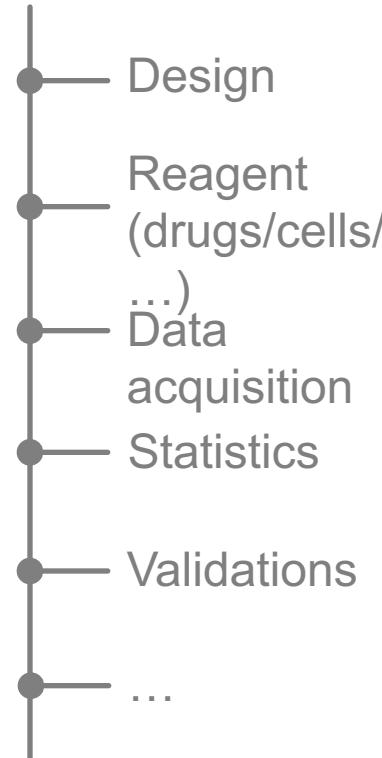
- The biotech company Amgen had a team of about 100 scientists trying to reproduce the findings of 53 “landmark” articles in cancer research published by reputable labs in top journals.
[Only 6 of the 53 studies were reproduced](#) (about 10%).
- Scientists at the pharmaceutical company, Bayer, examined 67 target-validation projects in oncology, women’s health, and cardiovascular medicine. Published results were reproduced in only
[14 out of 67 projects](#) (about 21%).
- The project, PsychFileDrawer, dedicated to replication of published articles in experimental psychology, shows a
[replication rate 3 out of 9](#) (33%) so far.

On a practical level, the US government gives nearly [\\$31 billion every year in science funding through NIH](#) only, which is mainly distributed in research grants to academic scientists. The 10% reproducibility rate means that 90% of this money (\$28 billion) is wasted. That’s a lot. How are the tax-payers supposed to respond to the scientist plight for more research funding given these numbers? Would you give more of your own money to someone who delivered you such a result?

Reproducibility challenges standardization also in proteomics

The screenshot shows the 'nature' website with a dark header. The main navigation bar includes 'Home', 'News & Comment', 'Research', 'Careers & Jobs', 'Current Issue', 'Archive', 'Audio & Video', and 'For...'. Below this, a breadcrumb trail shows 'Archive > Specials & supplements archive > Challenges in irreproducible research'. A 'SPECIAL' section is highlighted with three blue pipette icons. A link 'See all specials' is visible. The main content area features a large image of three blue pipettes above a row of three petri dishes. Below this is a dark banner with the text 'CHALLENGES IN IRREPRODUCIBLE RESEARCH'. A detailed paragraph discusses the lack of reproducibility in research and the steps taken by journals like Nature to address this issue. At the bottom left, a 'Free full access' button is shown.

Reproducibility



<http://www.nature.com/news/reproducibility-1.17552>

The screenshot shows a journal article from the 'Journal of Proteome Research'. The top right corner has a red 'ANALYSIS' label. The article title is 'A HUPO test sample study reveals common problems in mass spectrometry-based proteomics'. The authors listed are Alexander W Bell¹, Eric W Deutscher², Catherine E Au¹, Robert E Kearney³, Ron Beavis⁴, Salvatore Sechi⁵, Tommy Nilsson⁶, John J M Bergeron¹ & HUPO Test Sample Working Group⁷. The text discusses a test sample study to identify errors leading to irreproducibility in mass spectrometry-based proteomics. It mentions 20 highly purified recombinant human proteins, 27 laboratories, and the analysis of raw data. The journal logo 'Journal of proteome research' is at the bottom.

bioRxiv preprint doi: https://doi.org/10.1101/2016.09.22.040602; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

Abstract

We performed a test sample study to try to identify errors leading to irreproducibility, including incompleteness of peptide sampling, in liquid chromatography-mass spectrometry-based proteomics. We distributed an equimolar test sample, comprising 20 highly purified recombinant human proteins, to 27 laboratories. Each protein contained one or more unique tryptic peptides of 1,250 Da to test for ion selection and sampling in the mass spectrometer. Of the 27 labs, members of only 7 labs initially reported all 20 proteins correctly, and members of only 1 lab reported all tryptic peptides of 1,250 Da. Centralized analysis of the raw data, however, revealed that all 20 proteins and most of the 1,250 Da peptides had been detected in all 27 laboratories as well as a general inability to identify purified proteins in samples of low complexity¹ (<http://www.abf.org/ResearchGroups/ProteomicsStandardsResearchGroup/HUPOs/ABFPRGStudy2006center.pdf>). This is in part due to the stochastic nature of peptide sampling by the mass spectrometer and the inherent bias toward peptides of higher concentrations, which also confounds the statistical challenges and pitfalls associated with MS-based analyses, particularly when samples are rich in protein complexity. Protein solubilization, protein separation, protease digestion, peptide separation and peptide selection, all involve steps and protocols that vary greatly among labs, and different commercially available tandem mass

Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography–Tandem Mass Spectrometry

David L. Tabb,^{*,†} Lorenzo Vega-Montoto,^{‡,✉} Paul A. Rudnick,[§] Asokan Mulayath Variyath,^{†,✉} Amy-Joan L. Ham,[†] David M. Bunk,[§] Lisa E. Kilpatrick,^{||} Dean D. Billheimer,[†] Ronald K. Blackman,^{*,¶} Helene L. Cardozo,^{¶,§} Steven A. Carr,[¶] Karl R. Clauer,^{*} Jacob D. Jaffe,^{*} Kevin A. Kowalski,[○] Thomas A. Neubert,[¶] Fred E. Regnier,[○] Birgit Schilling,^{*} Tony J. Tegeler,[¶] Mu Wang,[¶] Pei Wang,[¶] Jeffrey R. Whiteaker,[¶] Lisa J. Zimmerman,[¶] Susan J. Fisher,[¶] Bradford W. Gibson,^{*} Christopher R. Kinslinger,^{*} Mehdi Mesri,^{*} Henry Rodriguez,^{*} Stephen E. Stein,[§] Paul Tempst,[○] Amanda G. Paulovich,^{*} Daniel C. Liebler,^{*,†} and Cliff Splegleman,[¶]

Why is reproducibility important?

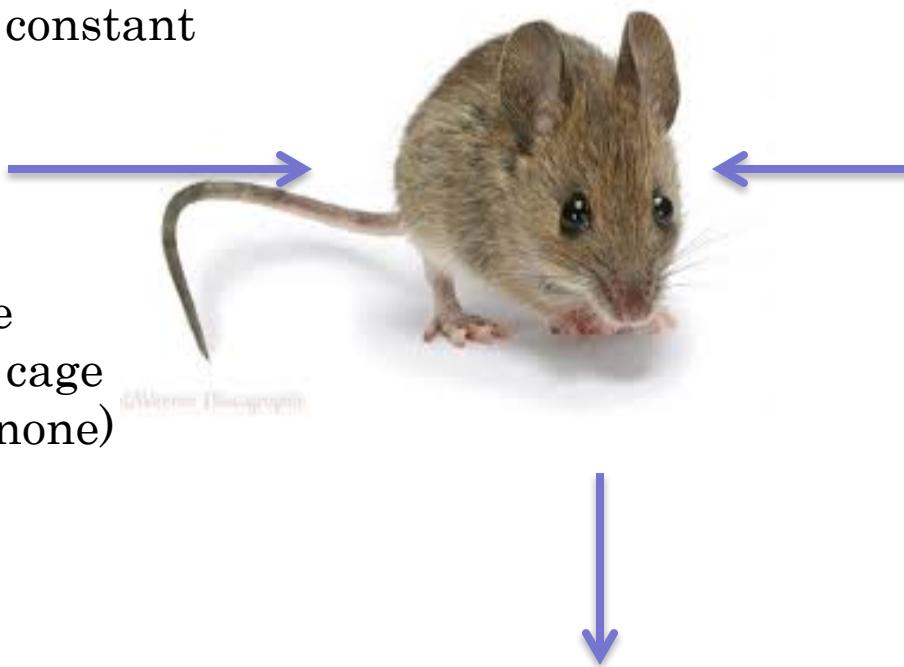
- Reproducibility of results is a bedrock requirement of experimental science
- Clinical/biomarker studies depend on the comparison of multiple samples
- Systems biology studies depend on the correlative analysis of multiple datasets (Big Data)
- **Perceived or real poor reproducibility is construed with poor competency or, worse, as poor ethical standards**

We learn as students: Do a controlled experiment

Recipe:

Keep all variables constant

- strain,
- age,
- time of day,
- food state
- temperature
- Microbes in cage
(preferably none)
- etc



Change one variable: e.g.

- Add a drug, or
- Delete a gene
- Mutate a gene

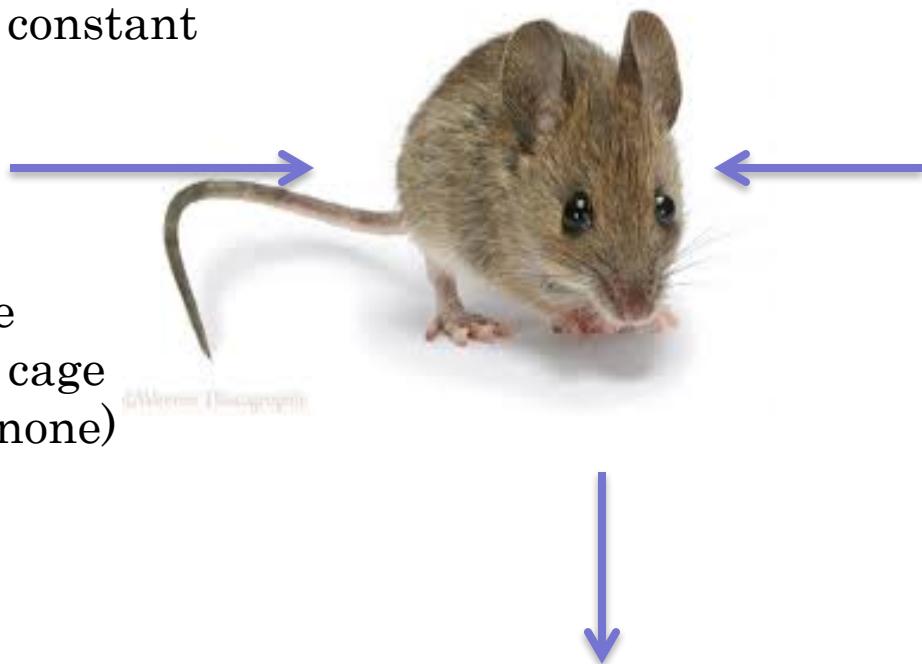
- Do a measurement (preferably proteomic)
- Generalize the finding to all mice in all conditions (e.g. gene XX causes cancer)
- Write paper

We learn as students: Do a controlled experiment

Recipe:

Keep all variables constant

- strain,
- age,
- time of day,
- food state
- temperature
- Microbes in cage
(preferably none)
- etc

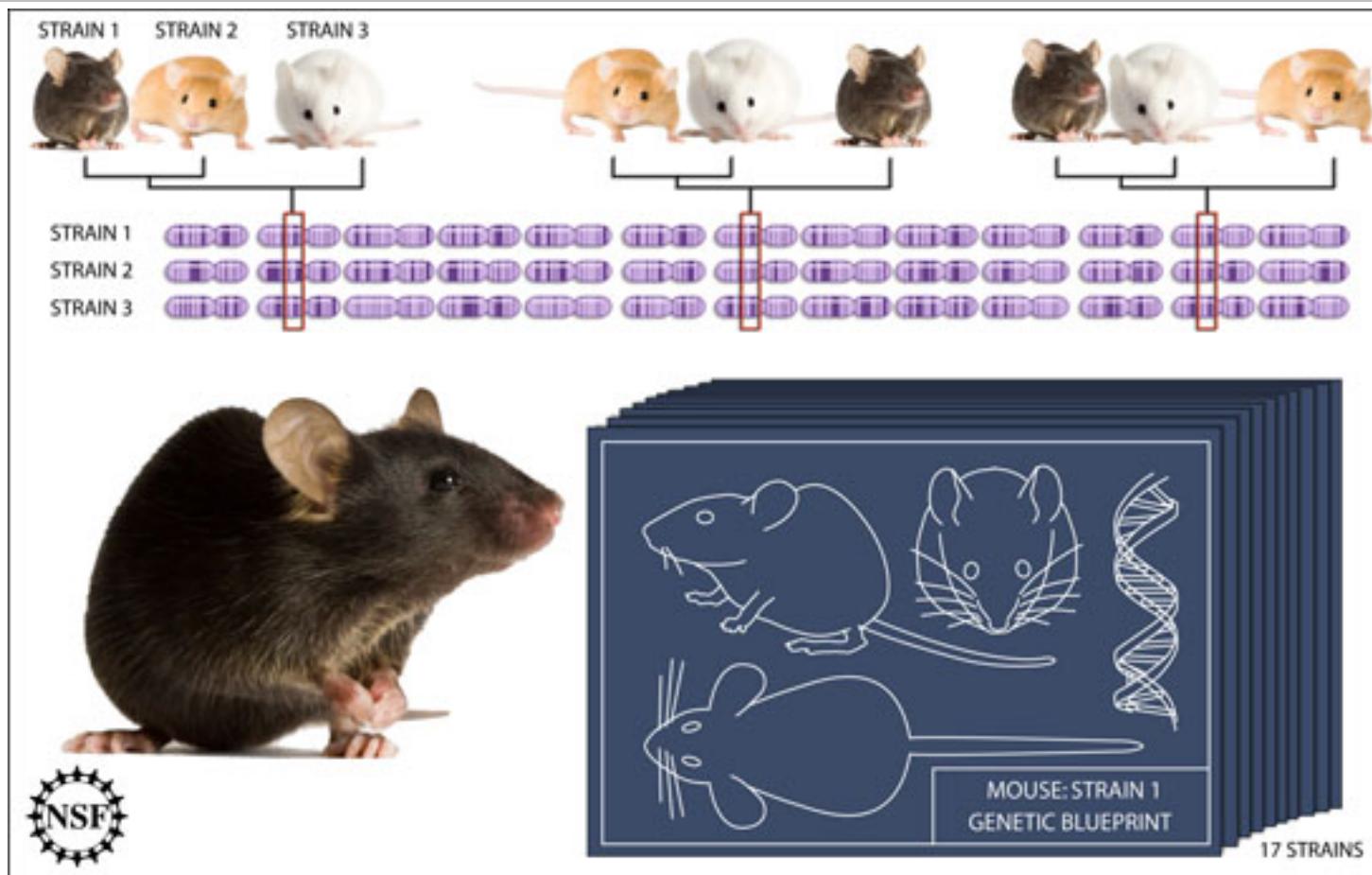


Change one variable: e.g.

- Add a drug, or
- Delete a gene

- Do a measurement (preferably proteomic)
- Generalize the finding to all mice in all conditions (e.g. gene XX causes cancer)
- Write paper

Variability is an essential element of biology



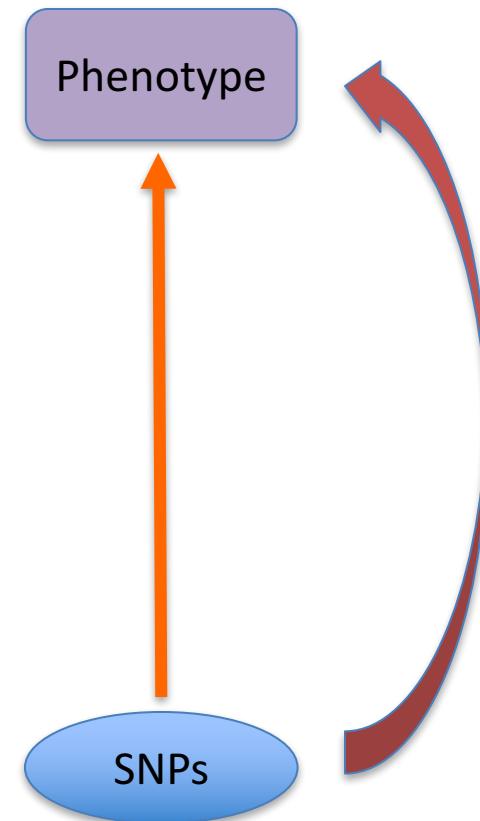
There is no reference mouse

There is no reference state

All results are conditional

Mice are different likely they act differently

Mapping genetic variation to phenotypic variation

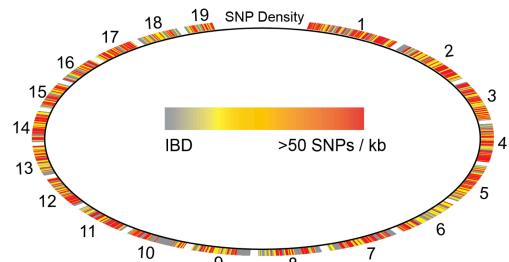


Complex Trait Analysis in a mouse genetic reference panel

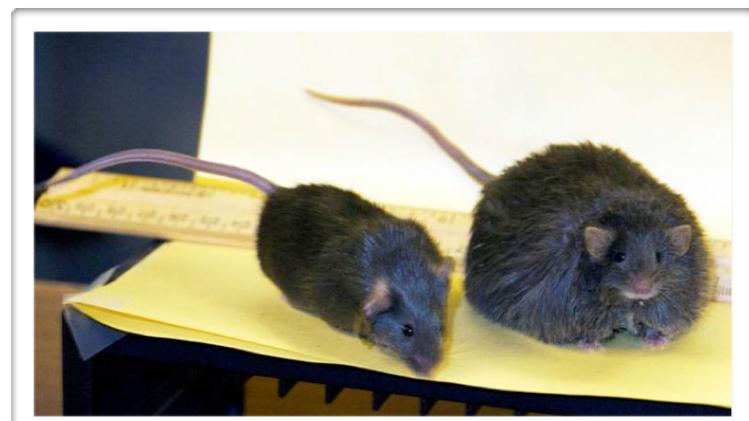
BXD Mice: ~150 Inbred Lines

- Descended from C57BL/6 (B) and DBA/2 (D)
- Between strains mice are like brothers and sisters

~5 million sequence variants

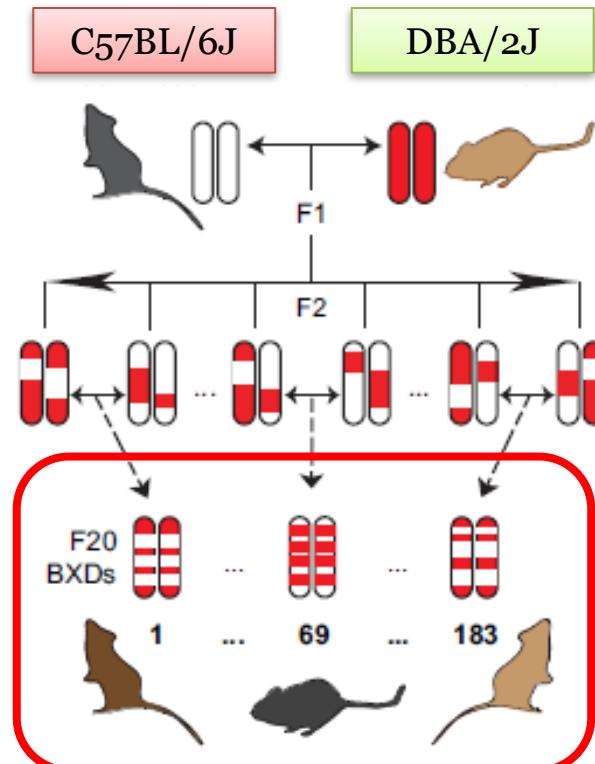


100's of phenotypes



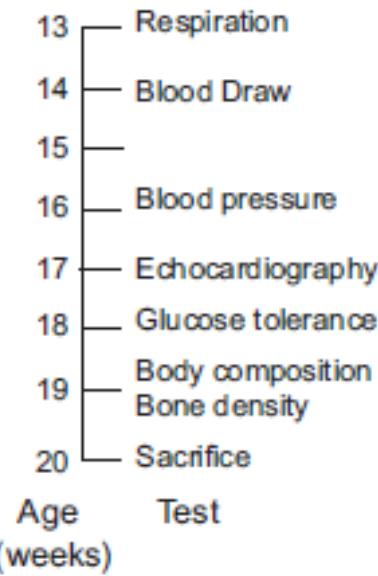
1
0

The BXD family: Currently the largest and best characterized mouse genetic reference populations(GRPs)



Phenotyping Program

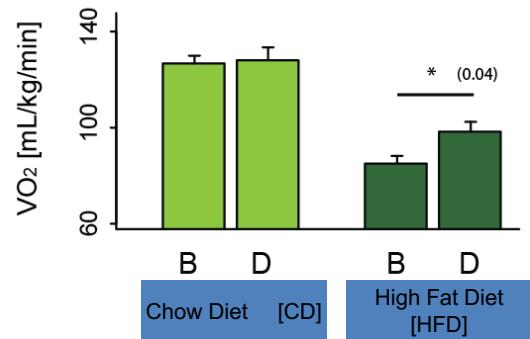
42 ♂ strains
25 ♀ strains
2-4 animals per group



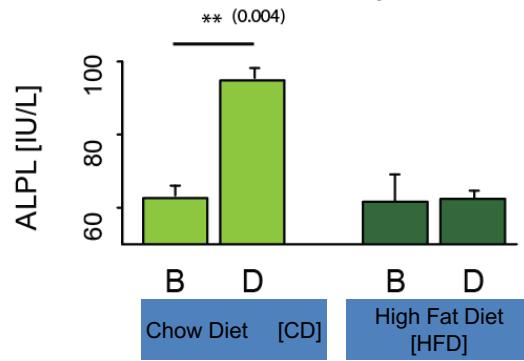
Cell 150, 1287-1299, September 14, 2012

Two Metabolic Traits in the BXD Parents (B & D)

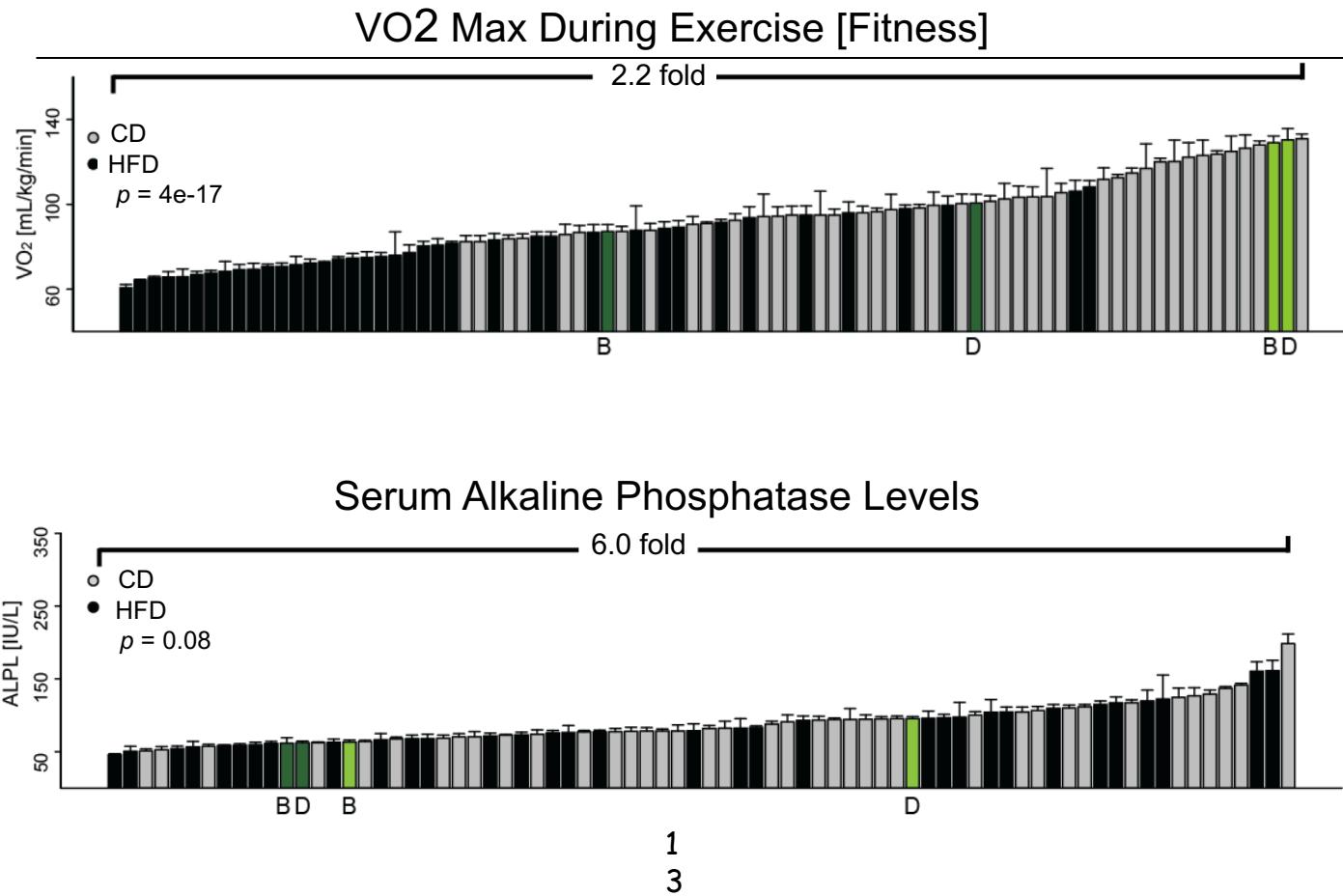
VO₂ Max During Exercise [Fitness]



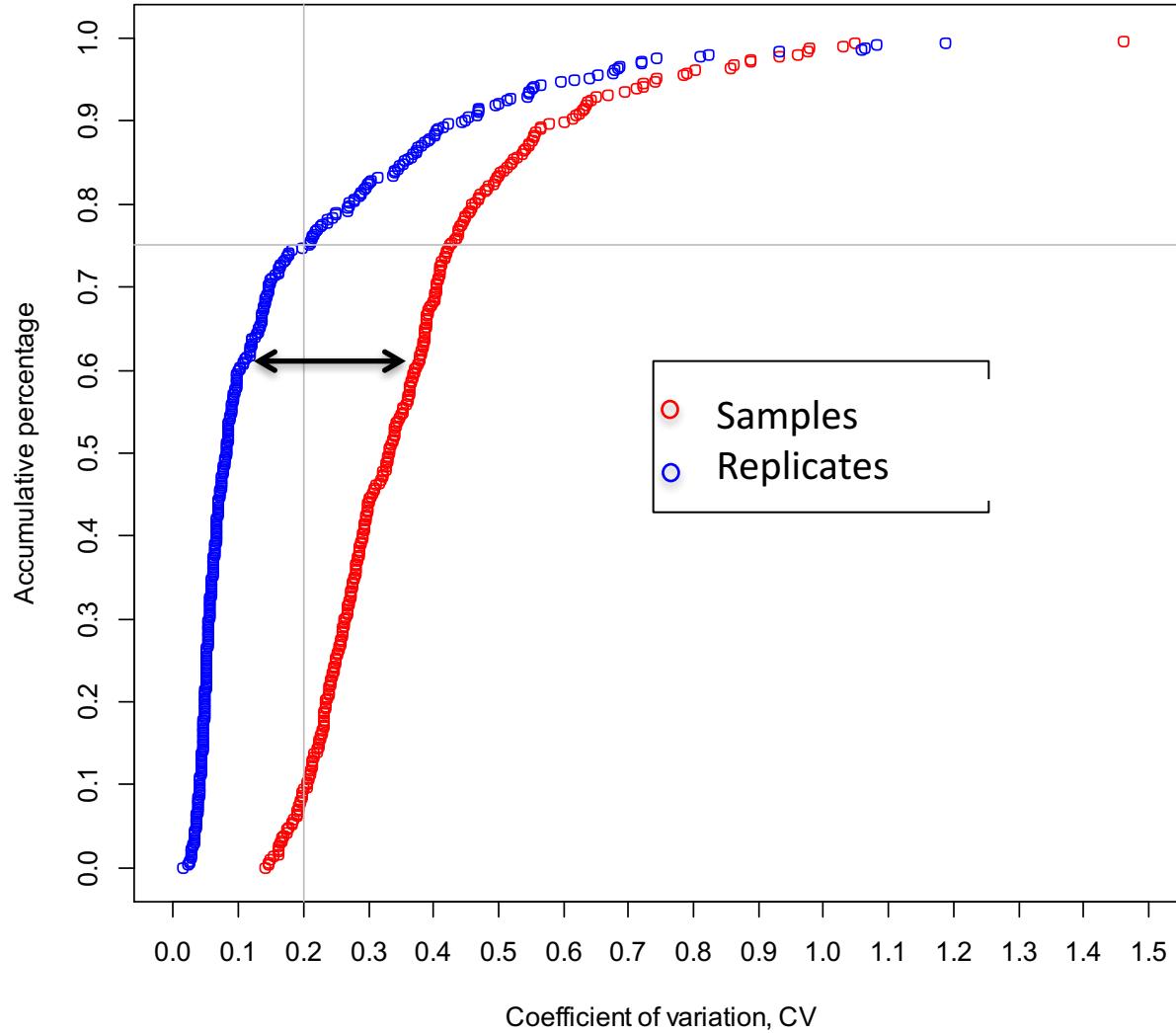
Serum Alkaline Phosphatase Levels



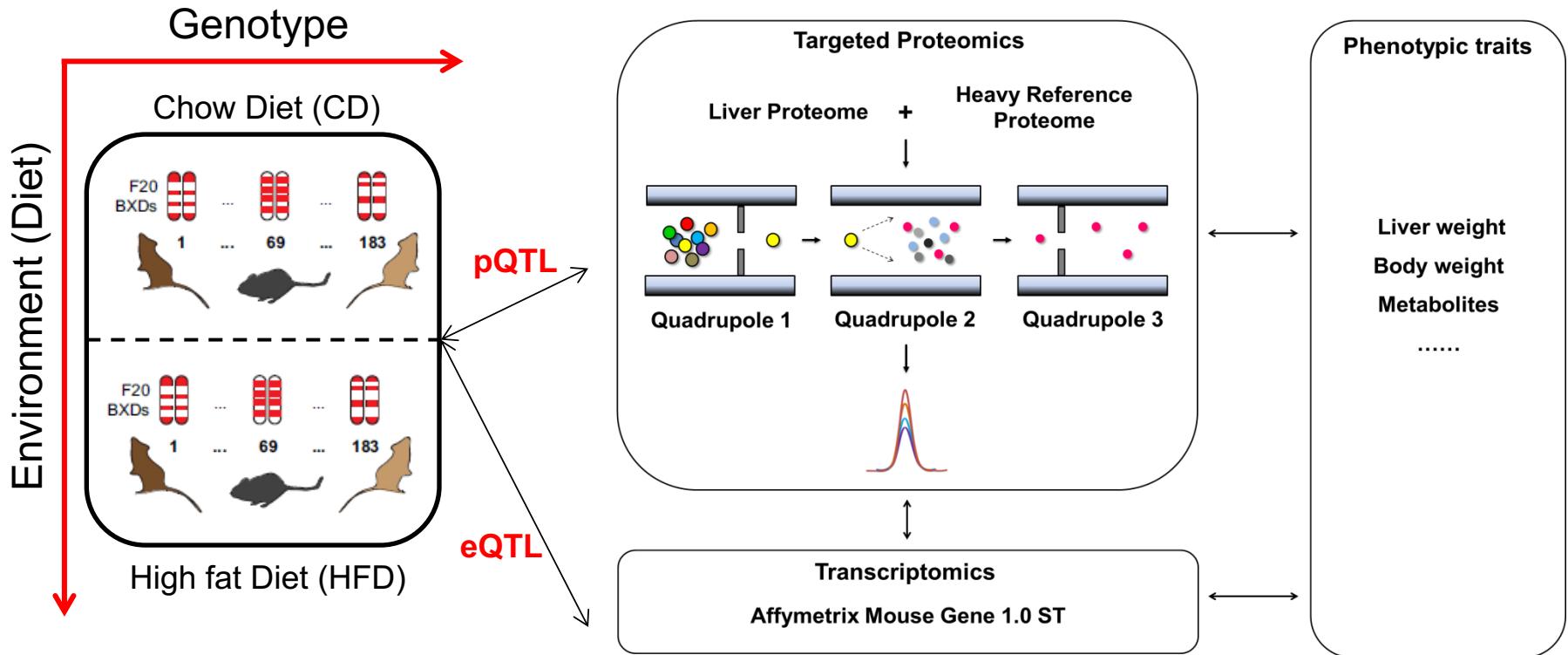
Two Metabolic Traits in the BXD Family



Reproducible measurements distinguish biological vs. technical variability



Experiment Scheme



Linking genotype and proteotype: pQTLs

	Total QTLs	<i>cis</i> -QTL	<i>trans</i> -QTL	Shared with CD and HFD
pQTL	44	14	30 (77%)	10 (23%)

- 
- Protein level regulated by other genes than itself
 - Environment(diet) appeared to have impact on protein level

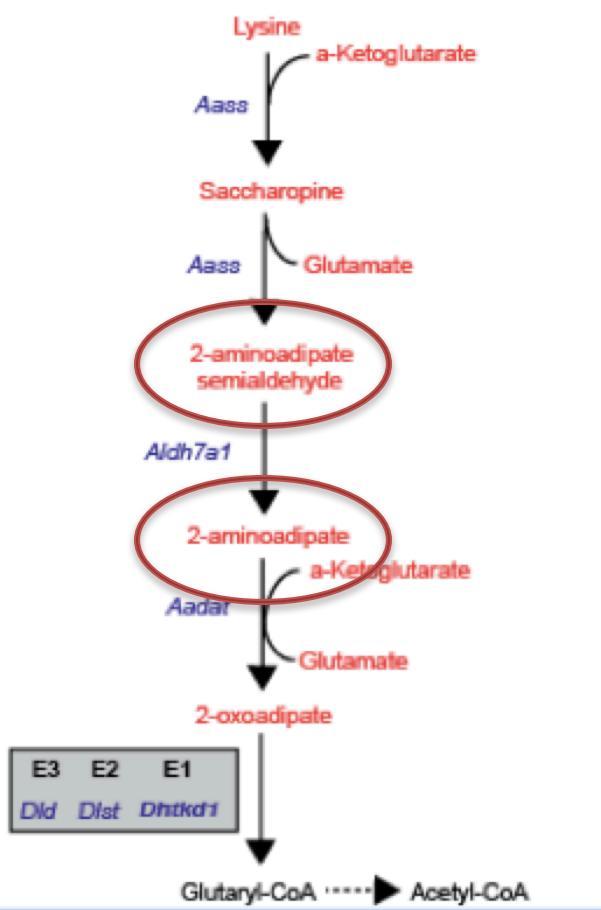
Comparison of eQTLs and pQTLs

	Total QTLs	<i>cis</i> -QTL	<i>trans</i> -QTL	Shared with CD and HFD
eQTL	47	37 (79%)	10	24 (50%)
pQTL	44	14 (32%)	30	10 (23%)
Overlap of eQTL and pQTL		10		

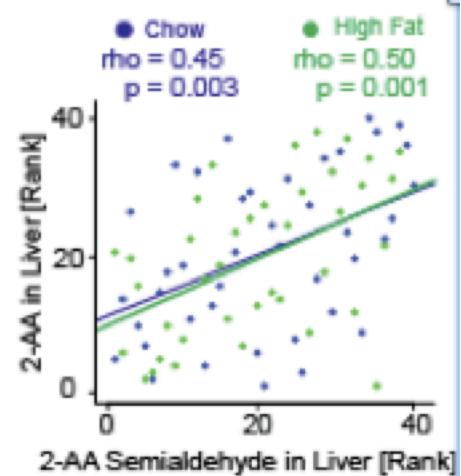
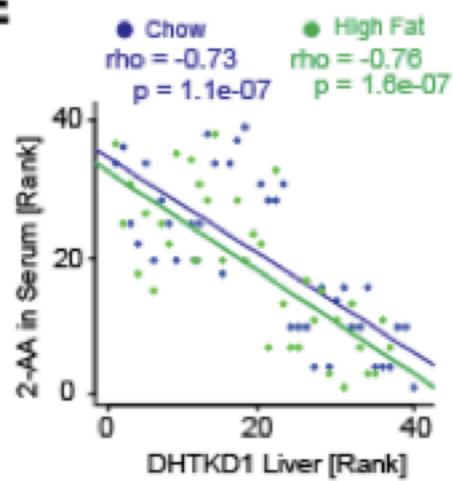
- Transcripts are more likely than proteins to be directly *cis*-regulated
- HFD appeared to have more impact on protein level
- 80% od QTL's are observed at transcript OR protein level – small overlap

From association to biochemical mechanism

D DHTKD1 Pathway



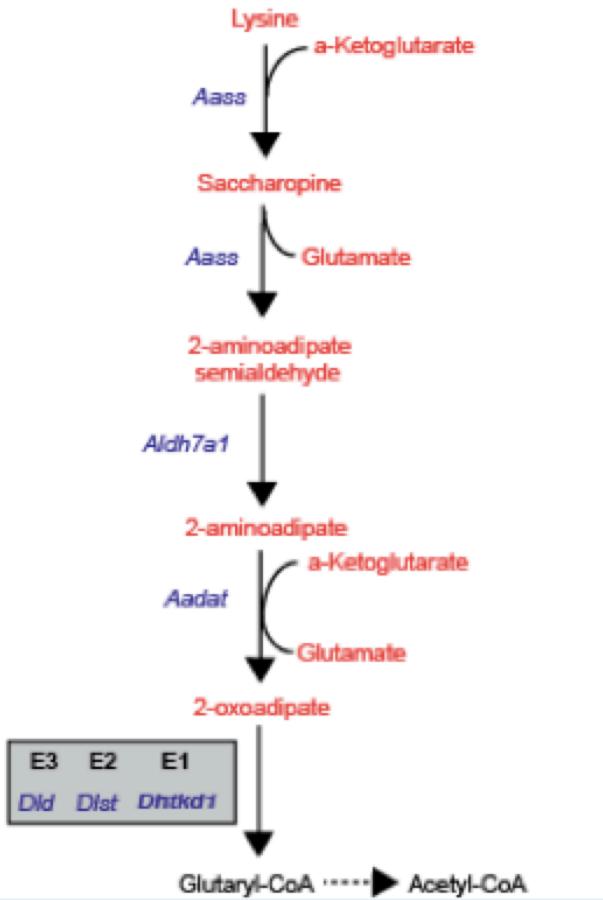
E



DHTKD1: strong pQTL
2AA/2AA semialdehyde: strong mQTL's

From biochemical mechanism to actionable targets

D DHTKD1 Pathway



Downloaded on October 27, 2013. *The Journal of Clinical Investigation*. More information at www.jci.org/articles/view/64801



Research article

2-Amino adipic acid is a biomarker for diabetes risk

Thomas J. Wang,^{1,2,3,4} Debby Ngo,^{1,5} Nikolaos Psychogios,¹ Andre Dejam,¹ Martin G. Larson,^{3,6} Ramachandran S. Vasan,^{3,7} Anahita Ghorbani,^{2,3} John O'Sullivan,¹ Susan Cheng,^{3,8} Eugene P. Rhee,^{1,9,10} Sumita Sinha,¹ Elizabeth McCabe,¹¹ Caroline S. Fox,^{3,12,13} Christopher J. O'Donnell,^{2,3,13} Jennifer E. Ho,^{3,7} Jose C. Florez,^{10,14,15} Martin Magnusson,^{16,17} Kerry A. Pierce,¹⁰ Amanda L. Souza,¹⁰ Yi Yu,¹⁸ Christian Carter,¹⁸ Peter E. Light,¹⁸ Olle Melander,^{17,19} Clary B. Clish,¹⁰ and Robert E. Gerszten^{1,2,10}

Why is reproducibility important?

- Reproducibility of results is a bedrock requirement of experimental science
- Clinical/biomarker studies depend on the comparison of multiple samples
- Systems biology studies depend on the correlative analysis of multiple datasets (Big Data)
- Perceived or real poor reproducibility is construed with poor competency or, worse, as poor ethical standards
- **Understanding biological reasons for variability uncovers interesting biology (e.g. QTL's)**

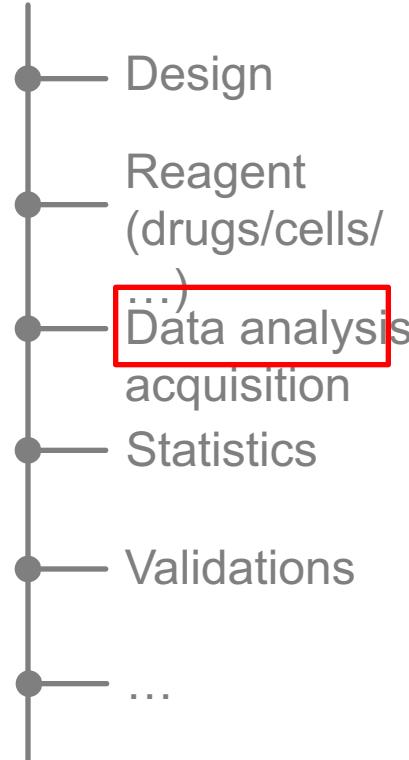
Outline

- We need to be able to generate reproducible proteomic research results:
 - for political reasons
 - for scientific reasons
- **If two or more people have the same proteomic data, can they generate the same results?**
- If two or more people have the same sample, can they generate the same proteomic results?
- If two or more people do the same experiment in a cell line, do they get the same results?

Reproducibility challenges standardization in proteomics

The screenshot shows the 'nature' website with a dark header. The top navigation bar includes links for Home, News & Comment, Research, Careers & Jobs, Current Issue, Archive, Audio & Video, and Forum. Below this, a secondary navigation bar shows 'Archive' and 'Specials & supplements archive' with a link to 'Challenges in irreproducible research'. The main content area is titled 'SPECIAL' and features a large image of three test tubes and their corresponding petri dishes. A button labeled 'See all specials' is visible. At the bottom of the special issue section, there is a dark banner with the text 'CHALLENGES IN IRREPRODUCIBLE RESEARCH'. Below this banner, a detailed paragraph discusses the nature of reproducibility challenges in research, mentioning the complexity of systems and the need for transparency. A 'Free full access' link is at the very bottom.

Reproducibility



<http://www.nature.com/news/reproducibility-1.17552>

A HUPO test sample study reveals common problems in mass spectrometry-based proteomics

Alexander W Bell¹, Eric W Deutscher², Catherine E Au¹, Robert E Kearney³, Ron Beavis⁴, Salvatore Sechi⁵, Tommy Nilsson⁶, John J M Bergeron¹ & HUPO Test Sample Working Group⁷

We performed a test sample study to try to identify errors leading to irreproducibility, including incompleteness of peptide sampling, in liquid chromatography-mass spectrometry-based proteomics. We distributed an equimolar test sample, comprising 20 highly purified recombinant human proteins, to 27 laboratories. Each protein contained one or more unique tryptic peptides of 1,250 Da to test for ion selection and sampling in the mass spectrometer. Of the 27 labs, members of only 7 labs initially reported all 20 proteins correctly, and members of only 1 lab reported all tryptic peptides of 1,250 Da. Centralized analysis of the raw data, however, revealed that all 20 proteins and most of the 1,250 Da peptides had been detected in all 27

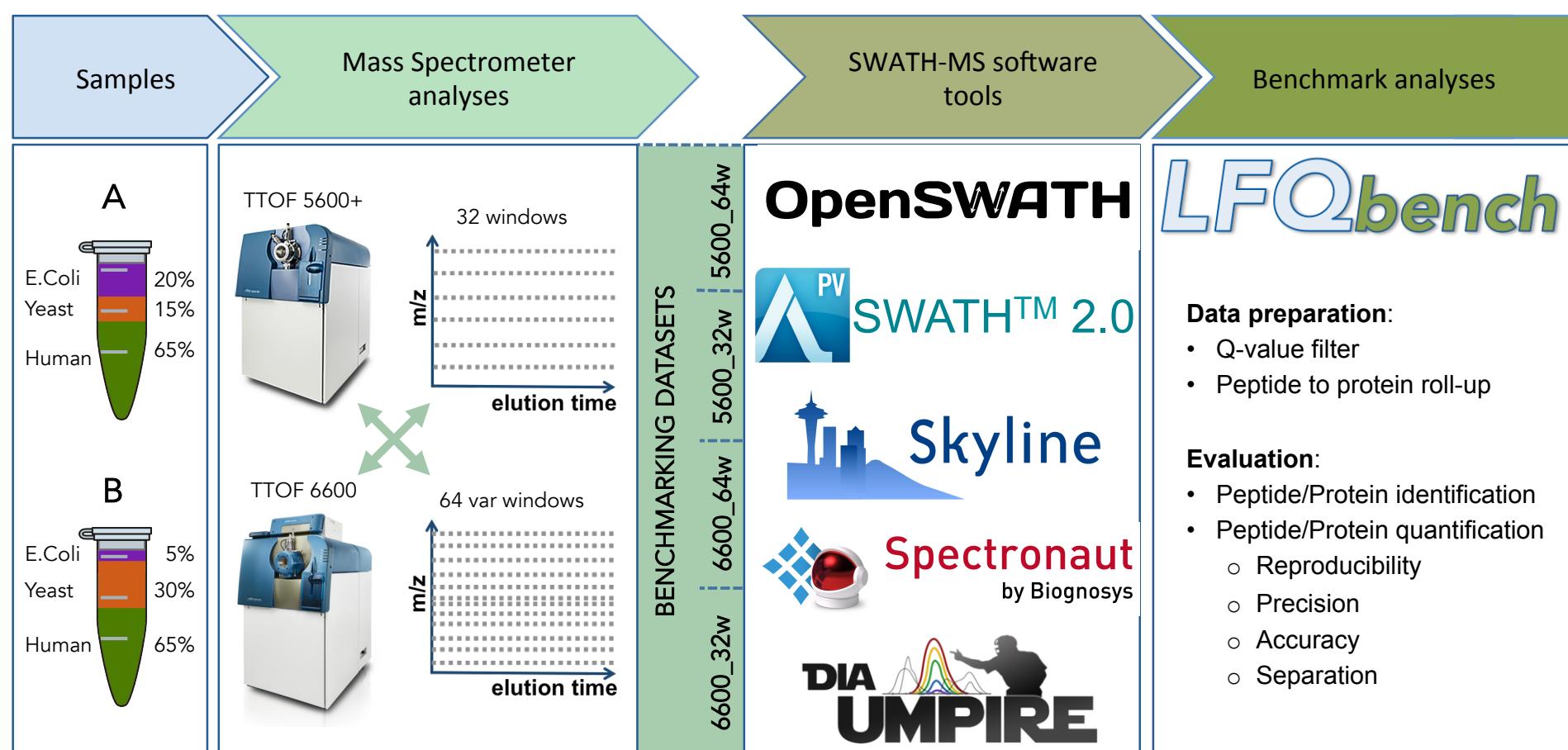
that there is both a lack of reproducibility between different laboratories as well as a general inability to identify purified proteins in samples of low complexity¹ (<http://www.abf.org/ResearchGroups/ProteomicsStandardsResearchGroup/Testers/ABRF/PRGStudy2006center.pdf>). This is in part due to the stochastic nature of peptide sampling by the mass spectrometer and the inherent bias toward peptides of higher concentrations, which also confounds the statistical challenges and pitfalls associated with MS-based analyses, particularly when samples are rich in protein complexity. Protein solubilization, protein separation, protease digestion, peptide separation and peptide selection, all involve steps and protocols that vary greatly among labs, and different commercially available tandem mass

Journal of
research articles
proteome
research

Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography–Tandem Mass Spectrometry

David L. Tabb,^{*,†} Lorenzo Vega-Montoto,^{‡,✉} Paul A. Rudnick,[§] Asokan Mulayath Variyath,^{‡,✉} Amy-Joan L. Ham,[†] David M. Bunk,[§] Lisa E. Kilpatrick,^{||} Dean D. Billheimer,[†] Ronald K. Blackman,^{*,¶} Helene L. Cardozo,[¶] Steven A. Carr,[†] Karl R. Clauer,^{*} Jacob D. Jaffe,^{*} Kevin A. Kowalski,[○] Thomas A. Neubert,[†] Fred E. Regnier,[○] Birgit Schilling,^{*} Tony J. Tegeler,^{*} Mu Wang,[¶] Pei Wang,[†] Jeffrey R. Whiteaker,[†] Lisa J. Zimmerman,[†] Susan J. Fisher,[○] Bradford W. Gibson,^{*} Christopher R. Kinslinger,^{*} Mehdi Mesri,^{*} Henry Rodriguez,^{*} Stephen E. Stein,[§] Paul Tempst,[○] Amanda G. Paulovich,^{*} Daniel C. Liebler,^{*,†} and Cliff Splegelman,^{*,†}

SWATH benchmark: Project design



Navarro, P, Gillet L, Tenzer S et al (software development teams),. *Nature Biotechnol.*.
2016

SWATH benchmark: Project design

Sample

Sample Composition

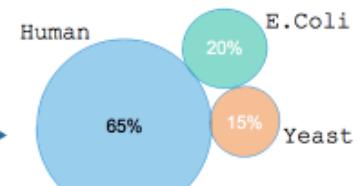
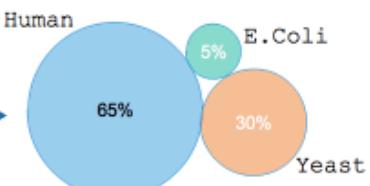
Instrument

Instrument Configuration

Datasets

Workflow Description

HYE124



TripleTOF 5600

TripleTOF 6600

32 fixed windows

64 variable windows

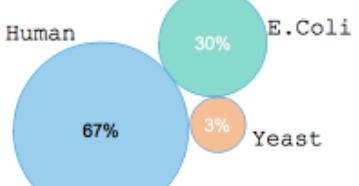
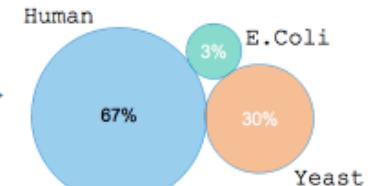
HYE124_5600_32fix

HYE124_5600_64var

HYE124_6600_32fix

HYE124_6600_64var

HYE110



TripleTOF 6600

32 fixed windows
32 variable windows

64 fixed windows
64 variable windows

HYE110_6600_32fix

HYE110_6600_32var

HYE110_6600_64fix

HYE110_6600_64var

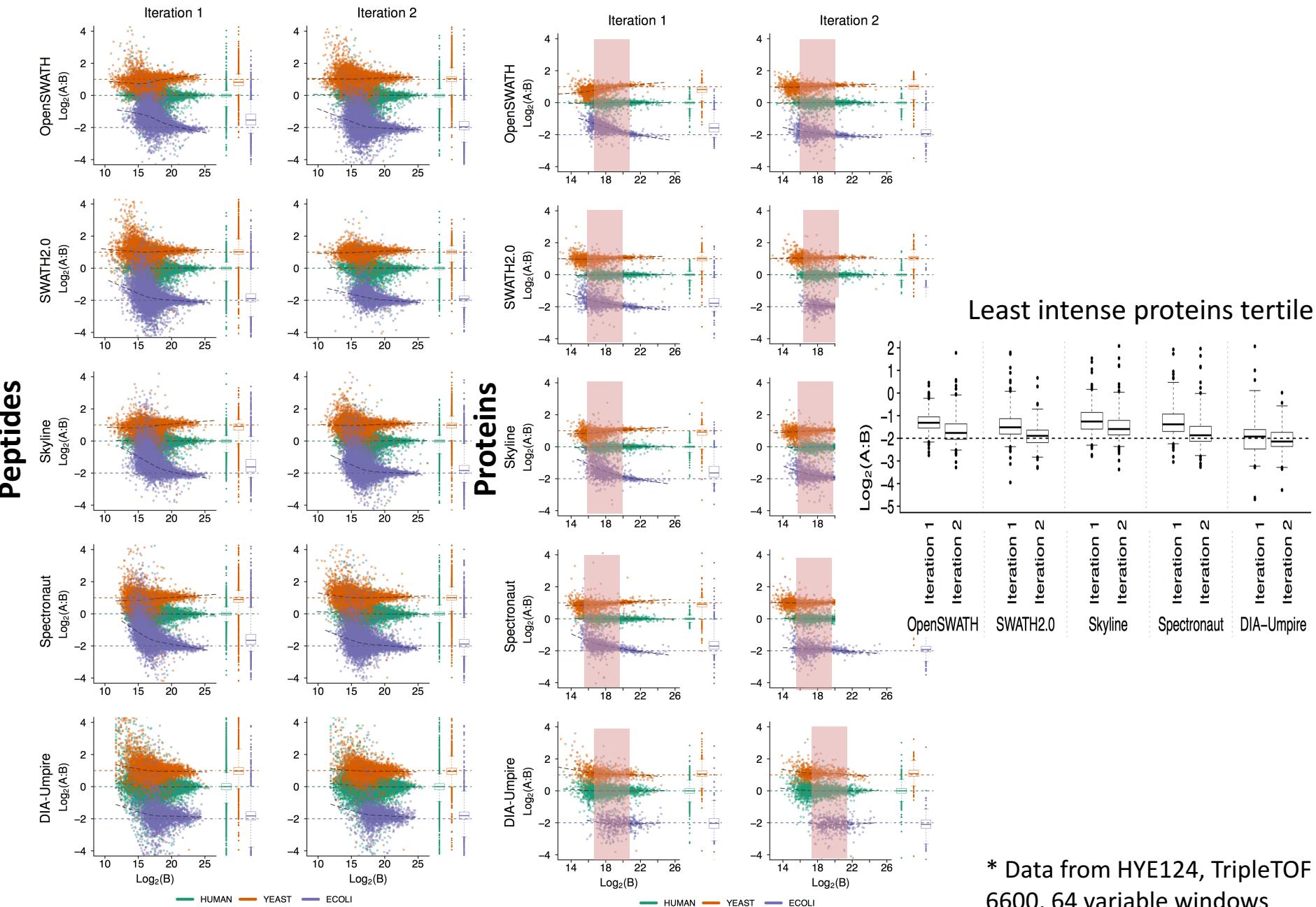
Sample A →

Sample B →

Sample A →

Sample B →

SWATH benchmark: Improvement of software tools



SWATH benchmark: Improvement of software tools

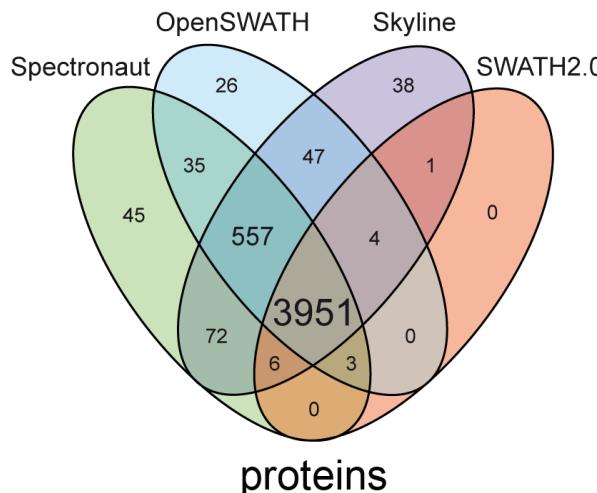
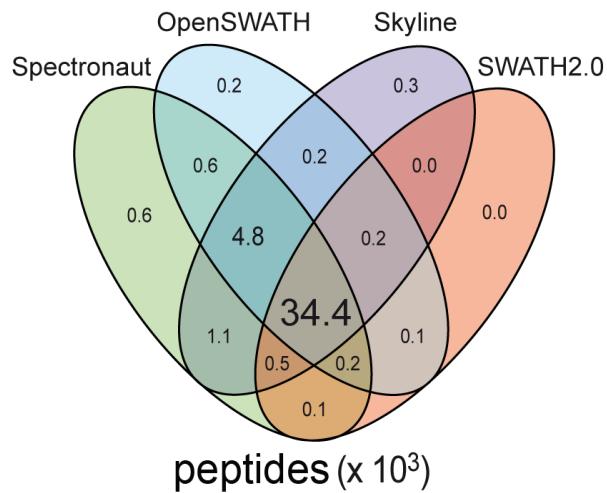
		Iteration 1					Iteration 2				
		Median CV of Human	Identifications	Valid log ratios	Overlap Yeast - Human (arctanh)	Overlap Ecoli - Human (arctanh)	Median CV of Human	Identifications	Valid log ratios	Overlap Yeast - Human (arctanh)	Overlap Ecoli - Human (arctanh)
peptides	OpenSWATH	0.06	40726	36098	2.12	2.24	0.08	40728	35944	2.10	2.26
	SWATH2.0	0.07	35517	35517	2.14	2.11	0.06	35489	26303	2.49	2.52
	Skyline	0.07	40804	34103	1.91	1.85	0.07	42517	37977	2.13	2.14
	Spectronaut	0.06	42439	37120	1.97	1.90	0.06	42325	36292	2.11	2.26
	DIA-Umpire	0.13	36332	28785	1.74	1.98	0.13	36249	25677	1.82	2.18
proteins	OpenSWATH	0.05	4724	4369	2.27	2.56	0.06	4732	4372	2.42	2.55
	SWATH2.0	0.06	4323	4323	2.37	2.36	0.06	4247	3407	2.40	2.56
	SWATH2.0 (built-in)	0.06	6178	6178	2.23	2.03					
	Skyline	0.06	4692	4172	1.99	2.15	0.05	4824	4483	2.35	2.44
	Spectronaut	0.03	4810	4363	2.09	2.18	0.03	4801	4314	2.30	2.46
	DIA-Umpire	0.12	4077	3466	2.09	2.30	0.12	4038	3194	2.11	2.85
	DIA-Umpire (built-in)	0.13	4849	4489	1.78	1.94					

Iteration 2 improved:

- populations separation (Yeast to Human and E.Coli to Human)
- Technical replication (CVs)

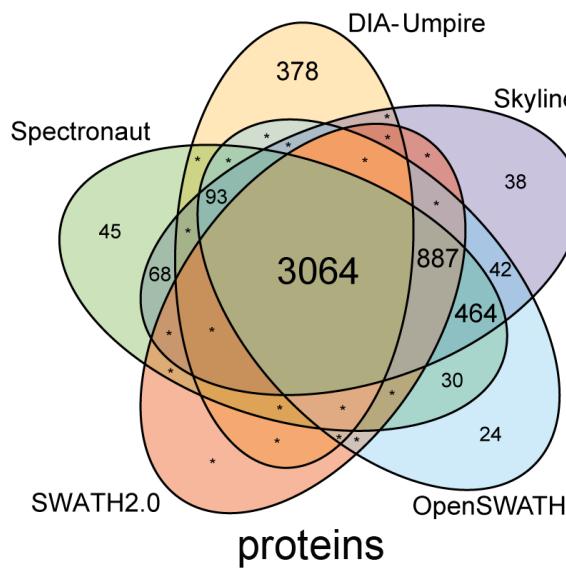
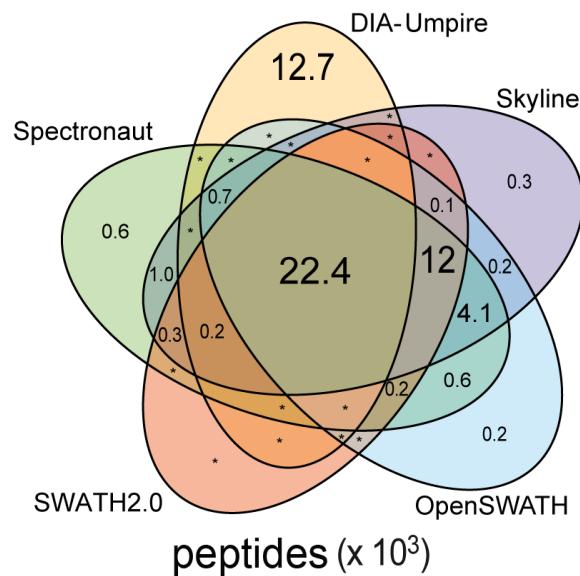
SWATH benchmark: Protein and Peptide identifications

Library-based tools



Total: 4,785 Proteins

All tools



Total: >5,100 Proteins

* Data from HYE124, TripleTOF 6600, 64 variable windows, Iteration 2

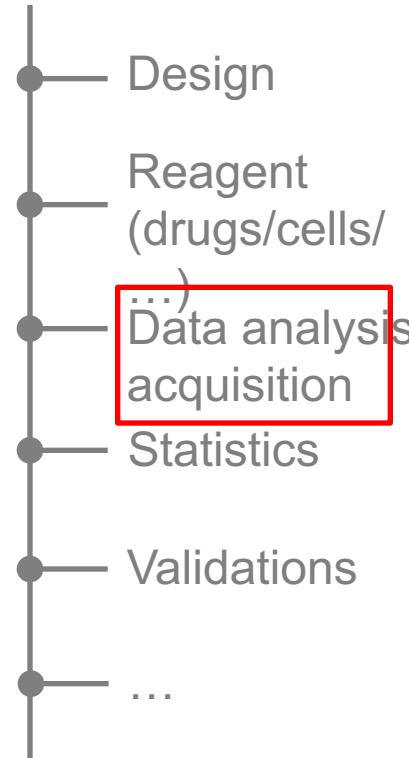
Outline

- We need to be able to generate reproducible proteomic research results:
 - for political reasons
 - for scientific reasons
- If two or more people have the same proteomic data, can they generate the same results?
- **If two or more people have the same sample, can they generate the same proteomic results?**
- If two or more people do the same experiment in a cell line, do they get the same results?

Reproducibility challenges standardization in proteomics

The screenshot shows the homepage of the journal 'nature' with a dark header. The header includes the word 'nature' in white, followed by 'International weekly journal of science'. Below the header are several navigation links: Home, News & Comment, Research, Careers & Jobs, Current Issue, Archive, Audio & Video, and Forum. Under the 'Archive' section, there is a link to 'Specials & supplements archive' and 'Challenges in irreproducible research'. A 'SPECIAL' banner is visible on the left, and a 'See all specials' link is on the right. Below the banner, there is an image of three test tubes and three petri dishes, with the text 'CHALLENGES IN IRREPRODUCIBLE RESEARCH' overlaid. A detailed text box discusses the challenges of reproducibility in research, mentioning the complexity of systems and techniques, and the need for transparency and robustness. At the bottom, there is a 'Free full access' link.

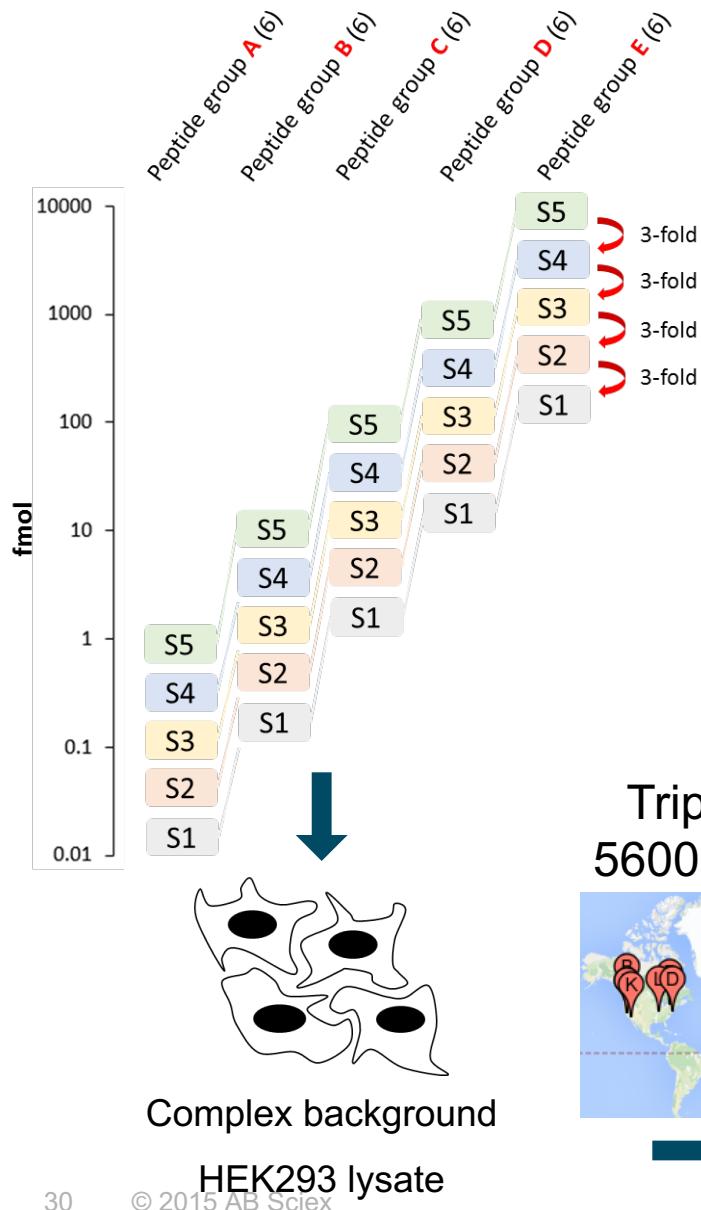
Reproducibility



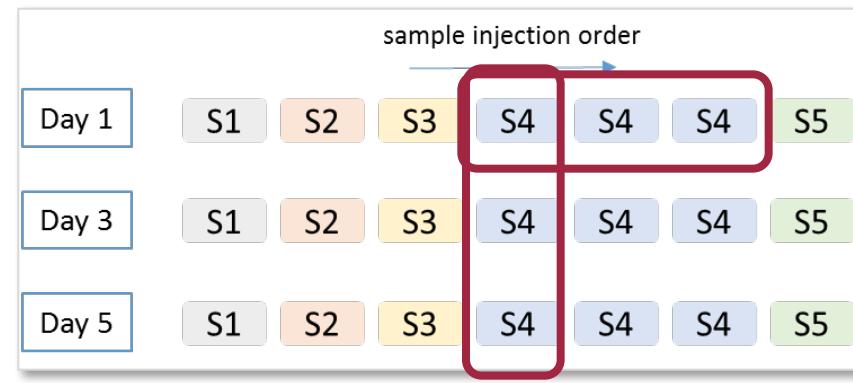
<http://www.nature.com/news/reproducibility-1.17552>

The screenshot shows a journal article from the 'Journal of Proteome Research'. The title of the article is 'A HUPO test sample study reveals common problems in mass spectrometry-based proteomics'. The authors listed are Alexander W Bell¹, Eric W Deutscher², Catherine E Au¹, Robert E Kearney³, Ron Beavis⁴, Salvatore Sechi⁵, Tommy Nilsson⁶, John J M Bergeron¹ & HUPO Test Sample Working Group⁷. The article discusses a test sample study to identify errors leading to irreproducibility in mass spectrometry-based proteomics. It details the distribution of 20 highly purified recombinant human proteins among 27 laboratories. The study found that while most proteins were identified correctly, there was significant variability in peptide sampling and ion selection. The text also notes the challenges of MS-based analyses, particularly with complex samples and the inherent bias toward peptides of higher concentrations. The journal logo 'Journal of proteome research' is at the bottom.

Study of cross-lab reproducibility of targeting MS



- 30 SIS peptides spiked into 1 µg HEK293 whole cell lysate background
- **3-fold** dilution series with 5 levels starting from starting from 5 different top concentrations (6 orders LDR -12 amol to 10 pmol)
- **5 samples** (S1-S5) distributed to 11 labs for SWATH MS analysis
- Analyze the sample set 3 times over the course of a week



Questions

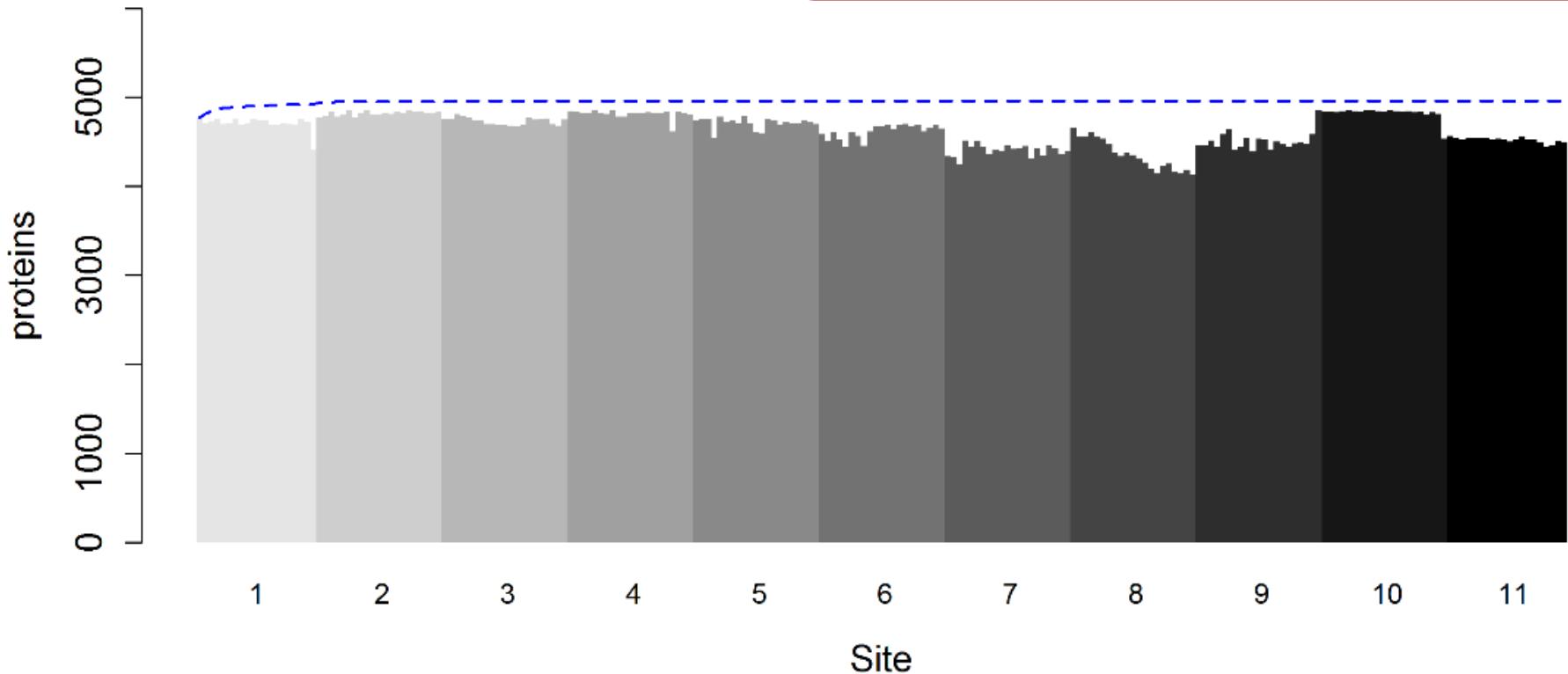
- Do we detect a consistent set of proteins across labs?
 - Proteins from the HEK293 lysate
 - Data extracted by automated analysis (OpenSWATH) at ETH
- What are the quantitative characteristics across labs?
 - Dynamic range, limit of detection, reproducibility
 - Determined by heavy synthetic peptide dilution series
 - Data extracted by targeted manual analysis

Peptide/protein detection rates (HEK293 lysate)

-- accumulated proteins detected

Total -- 4,960 proteins / 39,928 peptide PGs
Site median – 4,691 / 34,286 PGs

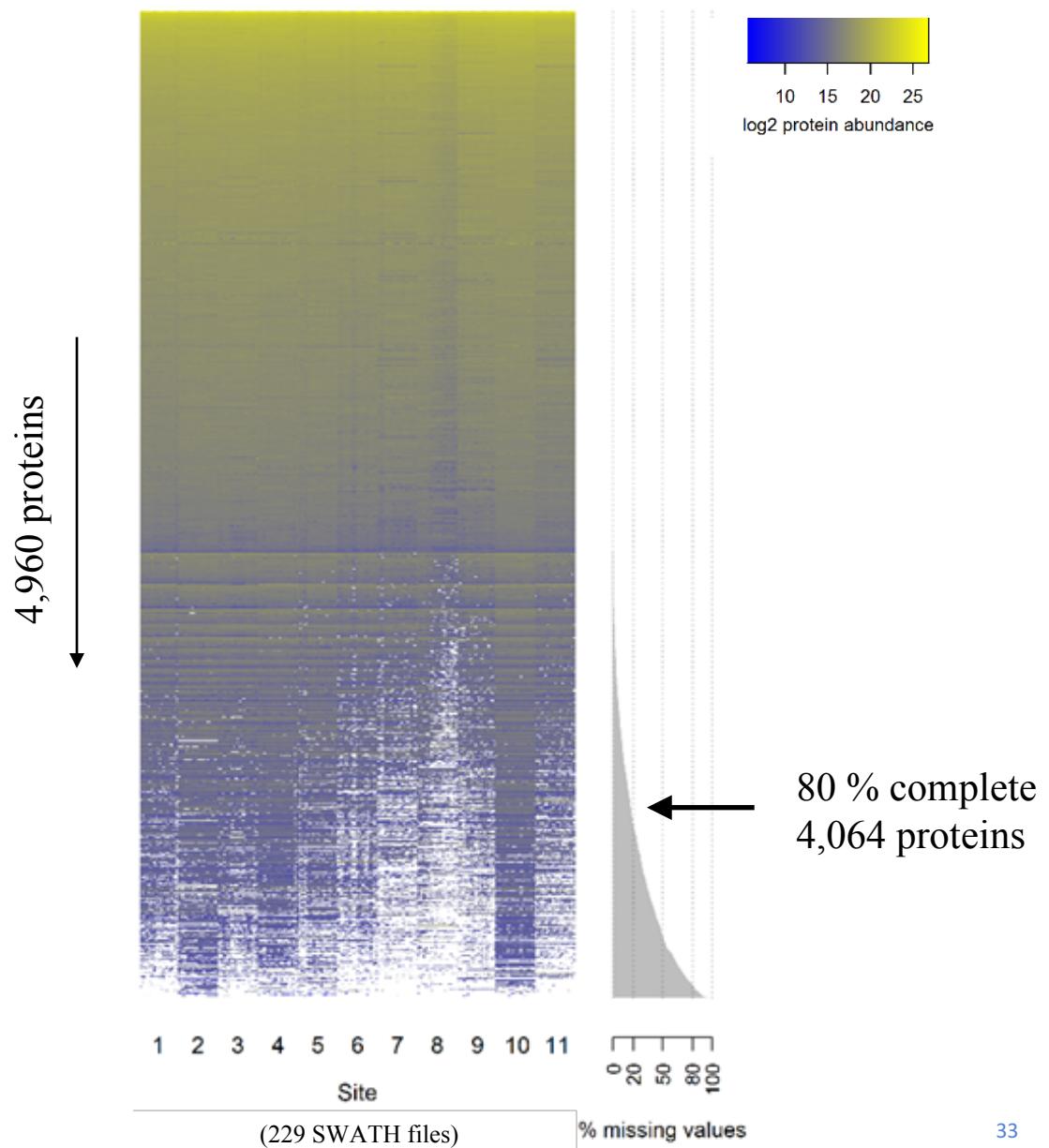
(1% protein FDR / 1% peptide PG FDR)



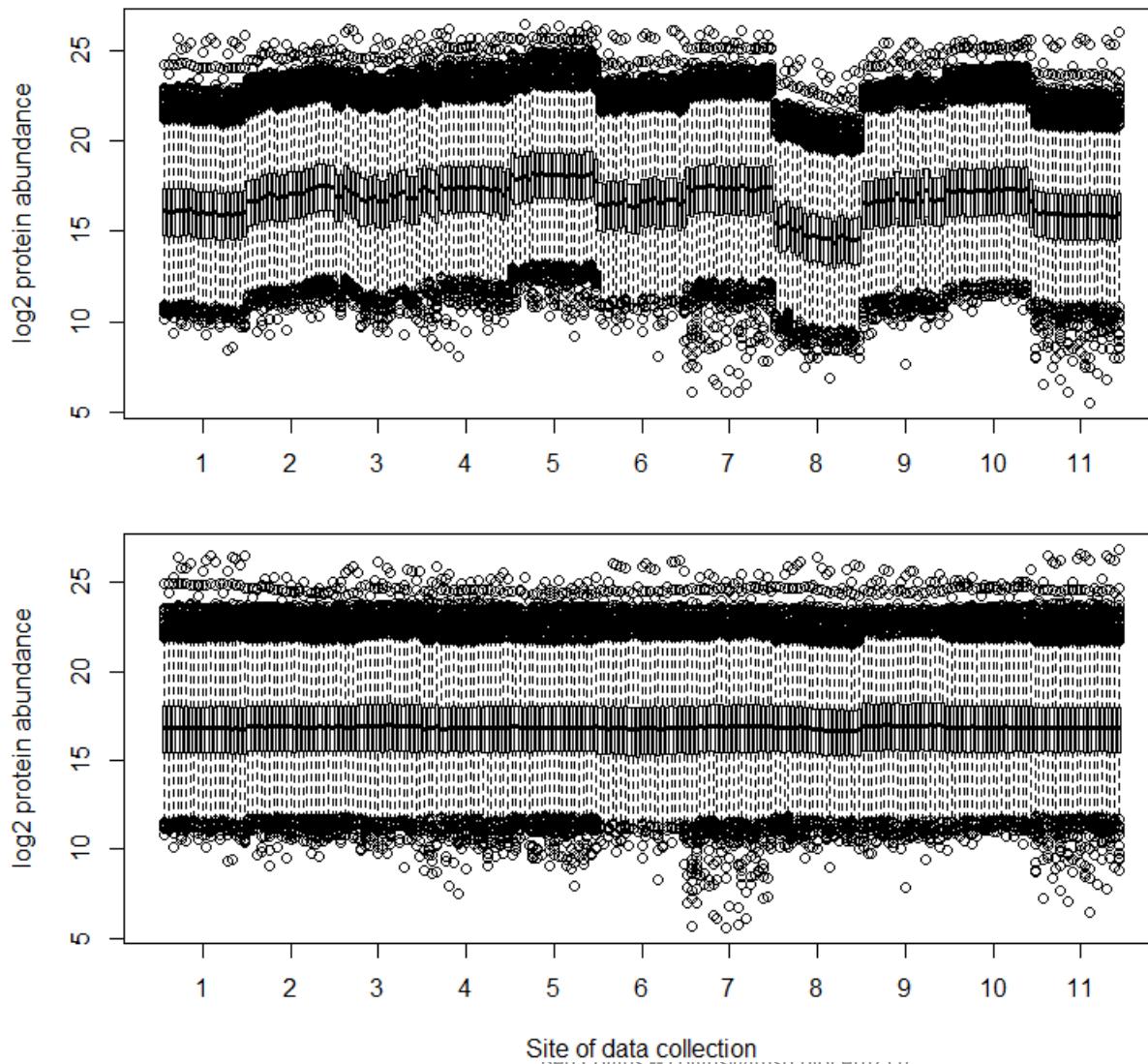
Data completeness

No alignment

No ID propagation
between runs

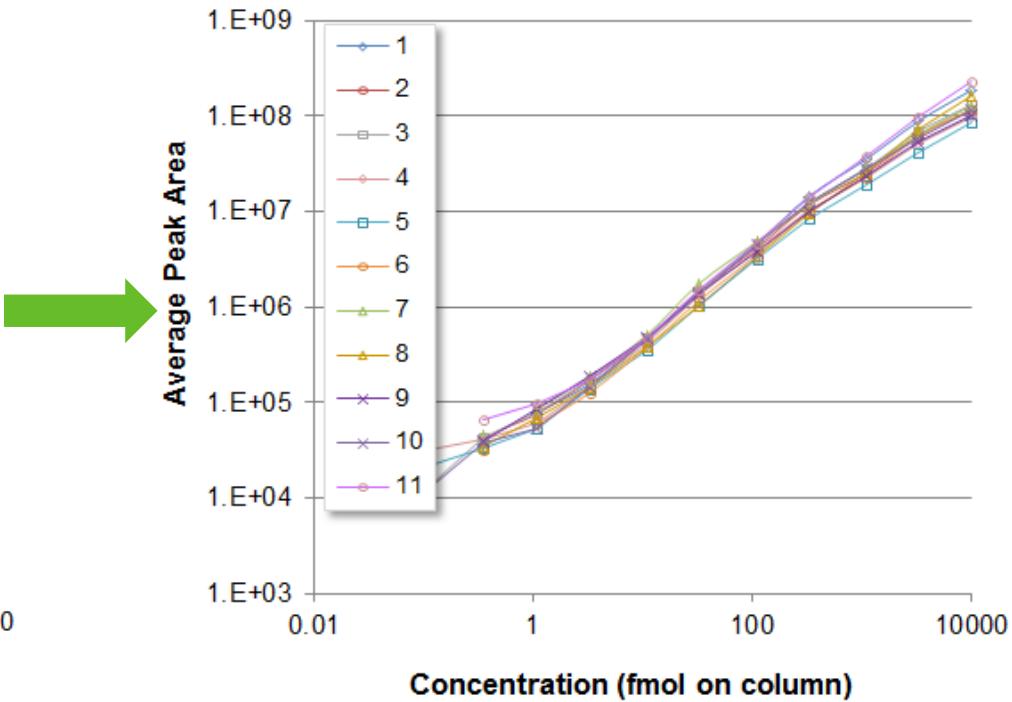
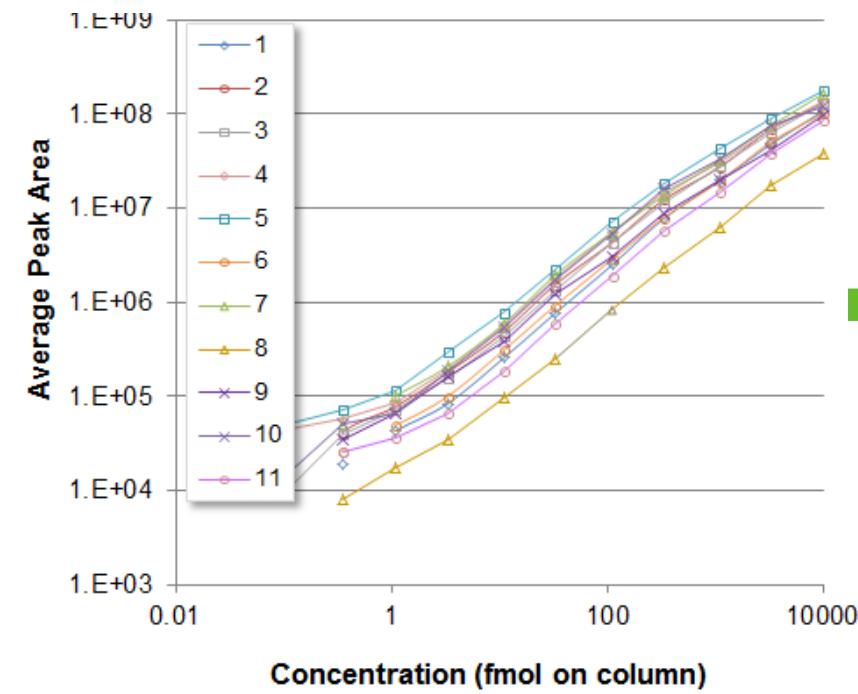


Median normalization is effective for rescaling

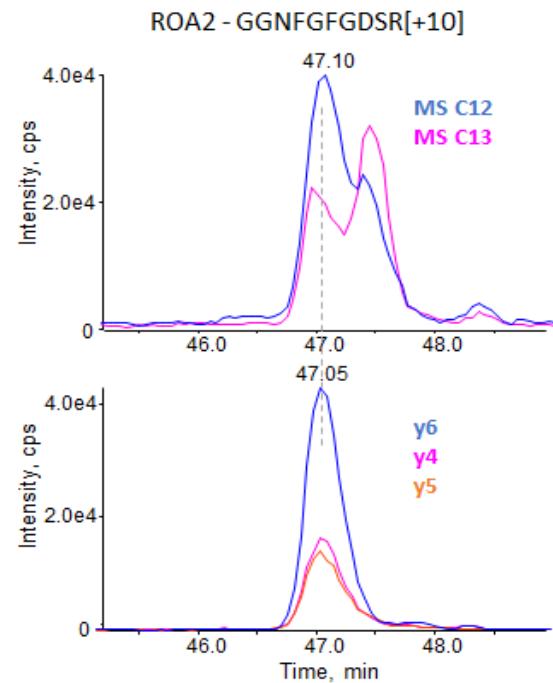
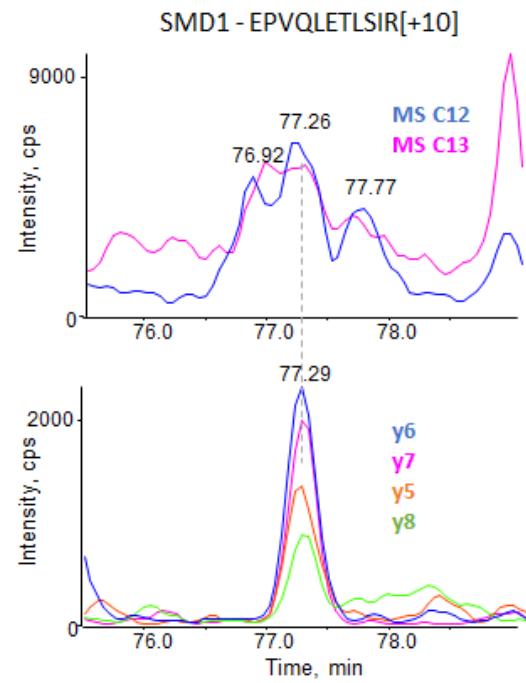
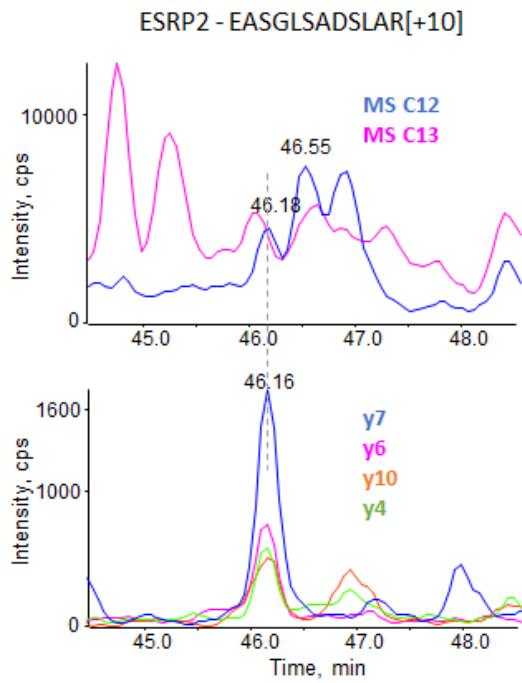


Dynamic range (30 x SIL peptides)

- ~4.5 orders from all sites
- Similar slope of response from all sites
- Y-axis (response) is offset between sites
- Simple median normalization based on HEK293 background

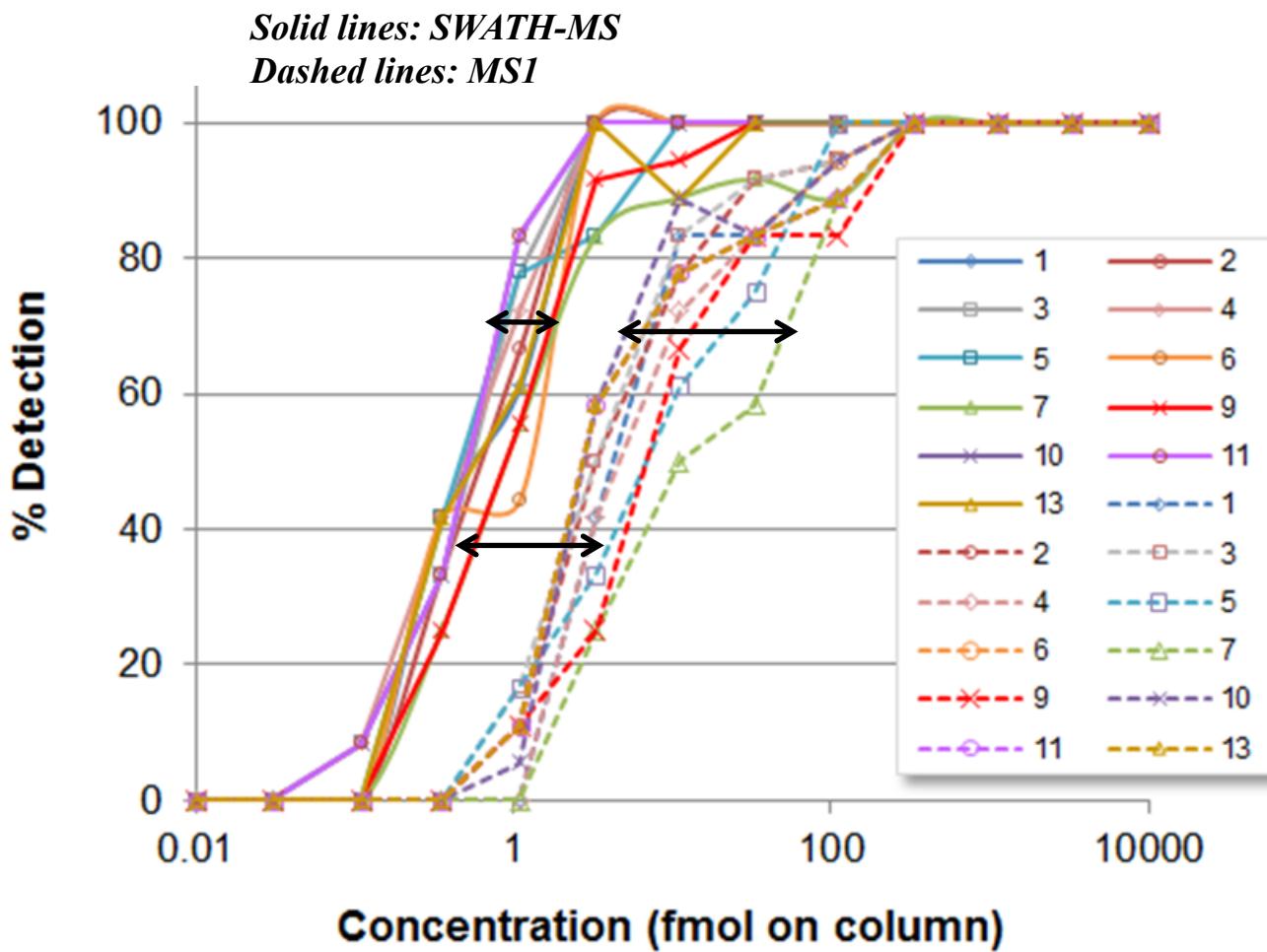


Comparison with MS 1 analysis; Anecdotal cases



MS1 - 0.02 m/z XIC width
SWATH - 0.05 m/z XIC width

Comparison with MS 1 analysis: Systematic analysis



- Sensitivity difference
- Consistent across labs
- Detection rate has higher stability for SWATH-MS

LLOQ

- Accuracy 80-120%
- < 20% CV
- S/N > 20
- Manual inspection for correct peak detection

At this point SWATH-MS is an intensely validated and well supported technique

- Validation of software tools against "ground truth" datasets (*Reiter et al, Nat. Methods, 2011; Roest et al Nat. Biotechnol. 2014; Roest et al , Nat. Meth. 2016*)  
- Comparison and validation of software tools (*Navarro et al, Nat. , 2016*)  
- Reference spectral libraries (*Rosenberger et al, Sci Data, 2015; Kusebauch et al, Cell, 2016, Schubert et al, Cell host and Microbe 2015*)  
- Methods and protocols for library generation (*Schubert et al 2015, Nature Protocol*)  
-  
- Cross-lab benchmarking and comparison study (*Collins, Hunter et al, Nat. Comm, in press*) 
- Error propagation analysis in large datasets (many samples) (*Rosenberger et al, under revision*) 
- Extension to PTM's (*Rosenberger et al, NBT, accepted*) 

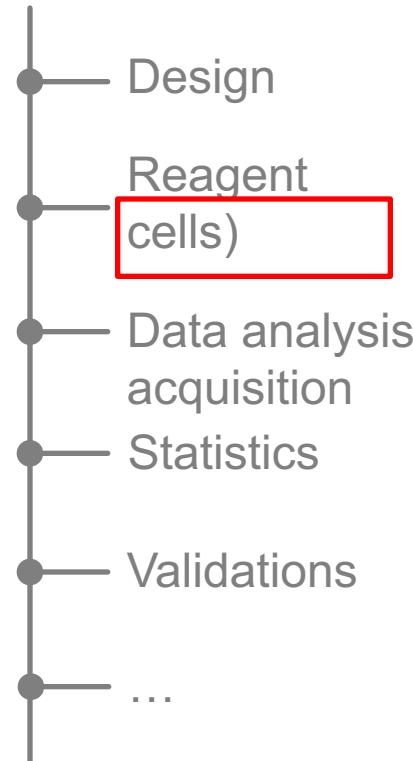
Outline

- We need to be able to generate reproducible proteomic research results:
 - for political reasons
 - for scientific reasons
- If two or more people have the same proteomic data, can they generate the same results?
- If two or more people have the same sample, can they generate the same proteomic results?
- **If two or more people do the same experiment in a cell line, do they get the same results?**

Reproducibility challenges standardization in proteomics

The screenshot shows the **nature** website, International weekly journal of science. The header includes links for Home, News & Comment, Research, Careers & Jobs, Current Issue, Archive, Audio & Video, and Forum. Below the header, a breadcrumb navigation shows Archive > Specials & supplements archive > Challenges in irreproducible research. A large banner on the left is labeled **SPECIAL** and features three test tubes. The text "CHALLENGES IN IRREPRODUCIBLE RESEARCH" is visible. A small link "See all specials" is at the top right of the banner. The main content area has a dark background with white text, discussing the challenges of reproducibility in research.

Reproducibility



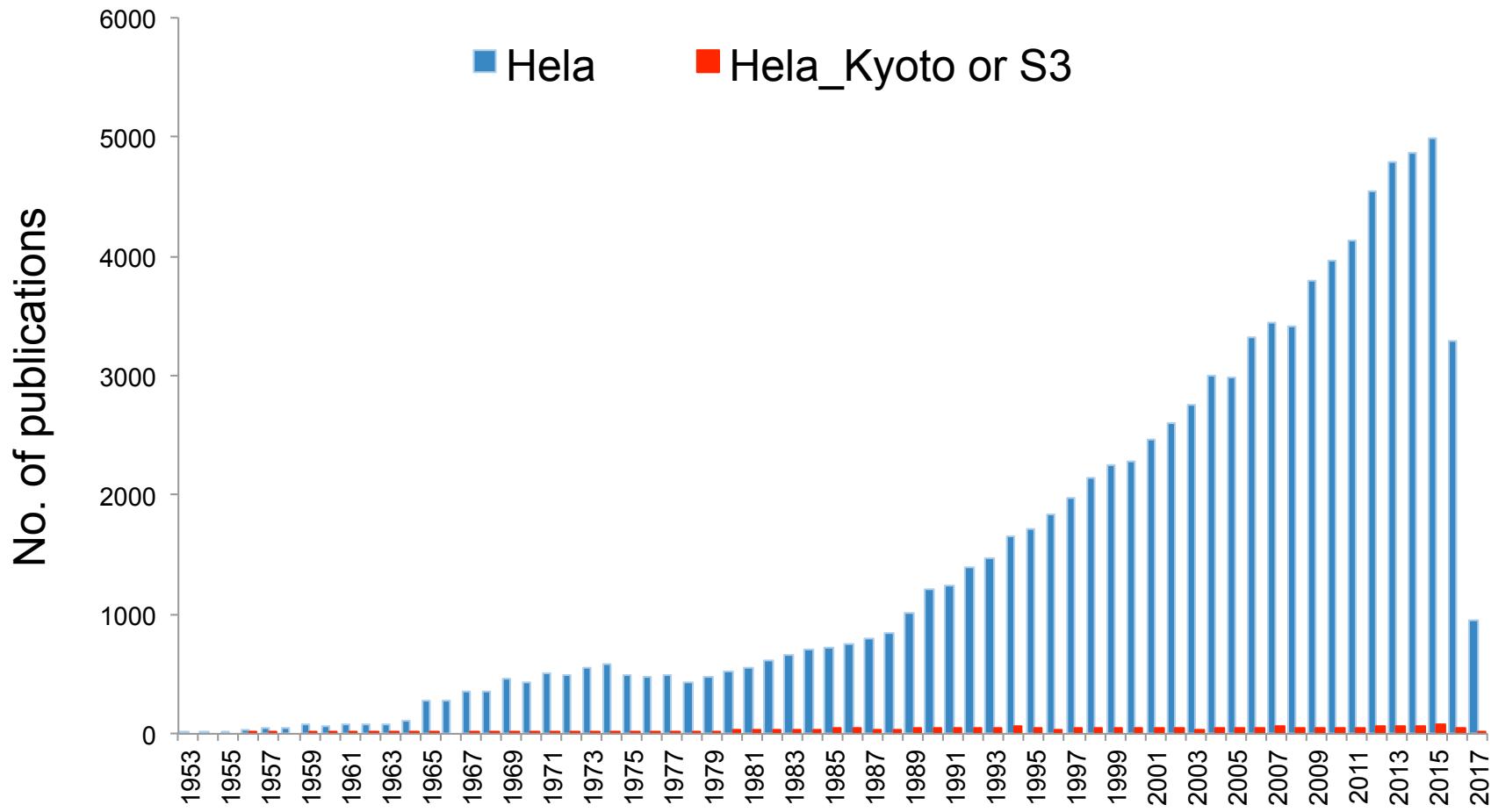
<http://www.nature.com/news/reproducibility-1.17552>

The article is titled "A HUPO test sample study reveals common problems in mass spectrometry-based proteomics". It is categorized under ANALYSIS. The authors listed are Alexander W Bell¹, Eric W Deutscher², Catherine E Au¹, Robert E Kearney³, Ron Beavis⁴, Salvatore Sechi⁵, Tommy Nilsson⁶, John J M Bergeron¹ & HUPO Test Sample Working Group⁷. The text discusses a test sample study to identify errors leading to irreproducibility in peptide sampling. It mentions 20 highly purified recombinant human proteins, 27 laboratories, and 20 tryptic peptides of 1,250 Da. The study found that while most proteins were detected, there was a lack of reproducibility between laboratories and a general inability to identify purified proteins in samples of low complexity. The journal logo "Journal of proteome research" is visible at the bottom.

Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography–Tandem Mass Spectrometry

David L. Tabb,^{*,†} Lorenzo Vega-Montoto,^{‡,§} Paul A. Rudnick,[§] Asokan Mulayath Variyath,^{†,¶} Amy-Joan L. Ham,[†] David M. Bunk,[§] Lisa E. Kilpatrick,^{||} Dean D. Billheimer,[†] Ronald K. Blackman,^{*,**} Helene L. Cardozo,^{*,§} Steven A. Carr,[†] Karl R. Clauer,^{*} Jacob D. Jaffe,^{*} Kevin A. Kowalski,[○] Thomas A. Neubert,[†] Fred E. Regnier,[○] Birgit Schilling,^{*} Tony J. Tegeler,^{*} Mu Wang,[†] Pei Wang,[†] Jeffrey R. Whiteaker,[†] Lisa J. Zimmerman,^{*} Susan J. Fisher,[○] Bradford W. Gibson,^{*} Christopher R. Kinslinger,^{*} Mehdi Mesri,^{*} Henry Rodriguez,^{*} Stephen E. Stein,[§] Paul Tempst,[○] Amanda G. Paulovich,^{*} Daniel C. Liebler,^{*,†} and Cliff Splegleman,^{*,†}

Which Hela cell?



A total of 93,000 publications in Pubmed (~0.3%)

HeLa: CCL2/ S3(CCL2.2)/ Kyoto

- **HeLa S3 (CCL2.2)**: they were **the third clone isolated** and propagated from the heterogeneous HeLa culture with the letter S in honor of Dr. Florence Rena Sabin, for whom the building housing Puck's new laboratory was named in December, 1951.
- **HeLa Kyoto**: carcinoma From: Narumiya S.; Kyoto University; Kyoto; Japan. “strongly adherent”...

HeLa Cell Line	Modal Chr. Number	Reference
HeLa	82; 70-164	(ATCC)
HeLa	78; 76-80	(Macville et al. 1999)
HeLa	60; 57-63	(Francke, Hammond, and Schneider 1973)
HeLa	65; 62-67	(Ash et al. 1984)
HeLa	74; 69-77	(Mincheva, Gissmann, and Zur Hausen 1987)
HeLa	84; 58-179	(Lavappa, Macy, and Shannon 1976)
HeLa	80; 79-81	(Bottomley, Trainer, and Griffin 1969)
HeLa	51	(Hughes 1965)
HeLa	60	(Obara et al. 1974)
HeLa	67	(Duesberg et al. 2011)
HeLa	69	(Ghosh and Ghosh 1975)
HeLa	71	(Czaker 1973)
HeLa	67	(Popescu and DiPaolo 1989)
HeLa	77	(Heneen 1976)
HeLa	69	(Singer and Fishman 1974)
HeLa	57; 54-59	(Harris et al. 1965)
HeLa	69; 67-70	(Cireli, Frimmel, and Schwarzacher 1966)
HeLa	61; 59-62	(Cireli, Frimmel, and Schwarzacher 1966)

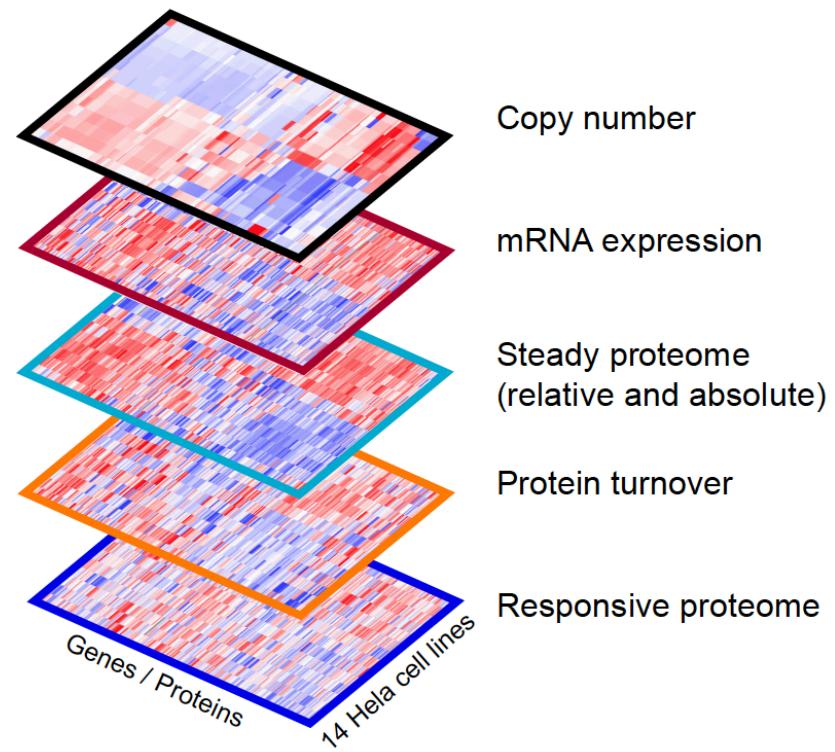
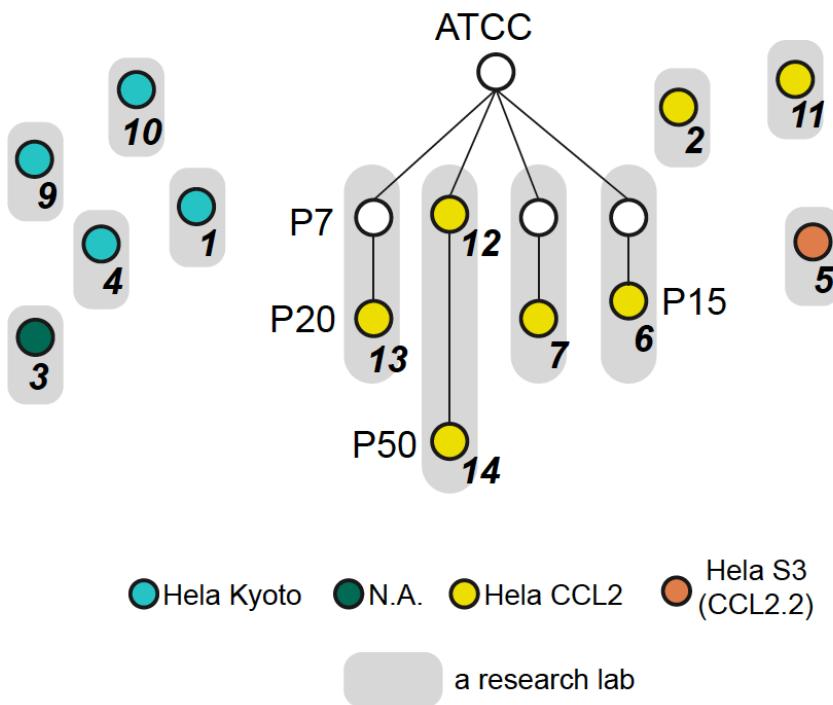
HeLa	69; 66-71	(Spurna and Hill 1967)
HeLa	70; 68-72	(Cerny, Korych, and Soukup 1964)
HeLa Kyoto	65; 62-68	(Landry et al. 2013)
HeLa St1	77	(Vogt 1959)
HeLa F8	70	(Vogt 1959)
HeLa D98/AH-2	62; 58-65	(Stanbridge et al. 1981)
HeLa-S3	68; 51-74	(Lavappa, Macy, and Shannon 1976)
HeLa-S3	68	(Gille and Joenje 1989)
HeLa-20	112	(Gille and Joenje 1989)
HeLa-80	84	(Gille and Joenje 1989)
M-HeLa-76	49	(Savelyeva and Mamaeva 1987)
HeLa-20	50; 49-50	(Mamaeva, Litvinchuk, and Pinaev 1983)
HeLa Ep. 1	72; (68-76)	(Norrby 1959)
HeLa	75; (73-76)	(Heneen 1976)
HeLa D98/AG	63	(Heneen 1976)

<https://thewinnow.com/>

Aims of study

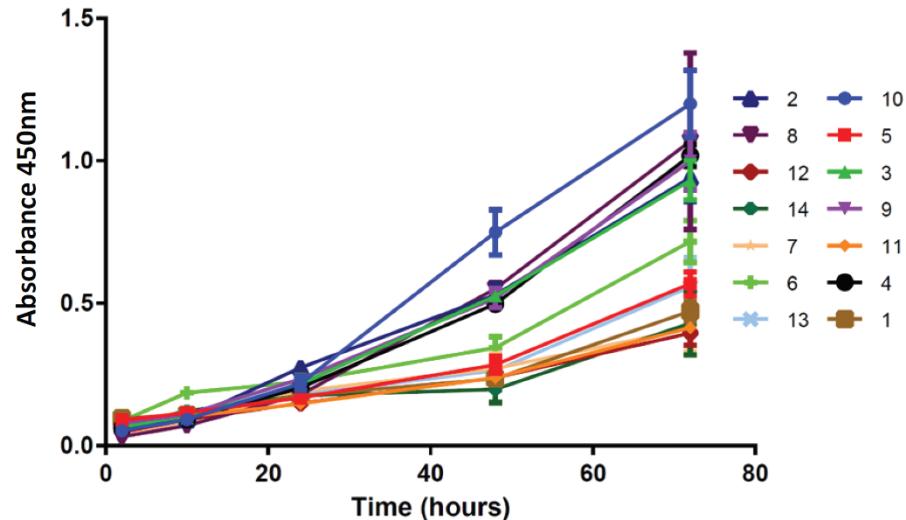
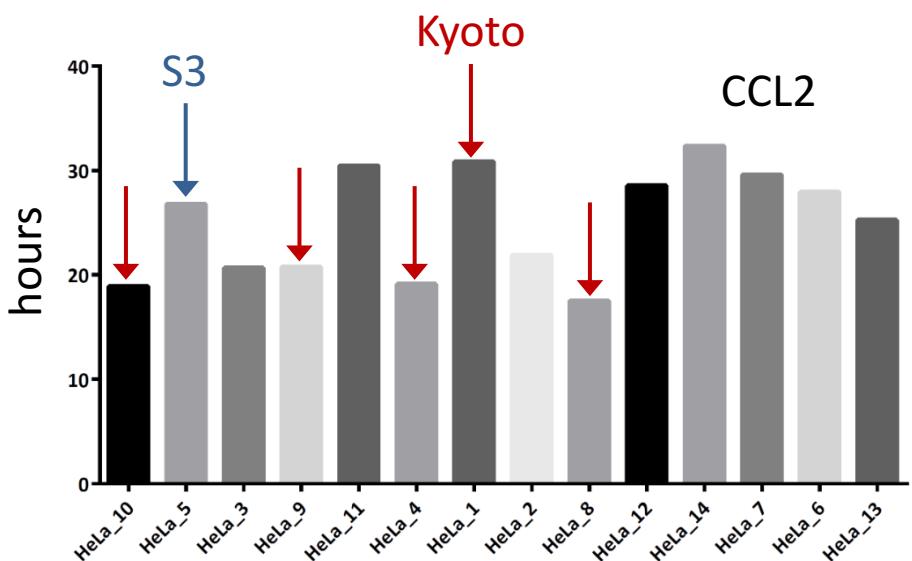
- A cross lab study addressing biological reproducibility.
- Similarity and heterogeneity of a cell line of the same name (Hela) in different labs.
- Can Hela cells across labs derive identical biological conclusions in a simple/routine experiment?
- How can we use multilayered data including proteomics to understand the dosage control of gene expression?

Experimental design



Phenotypes of centrally cultured cell

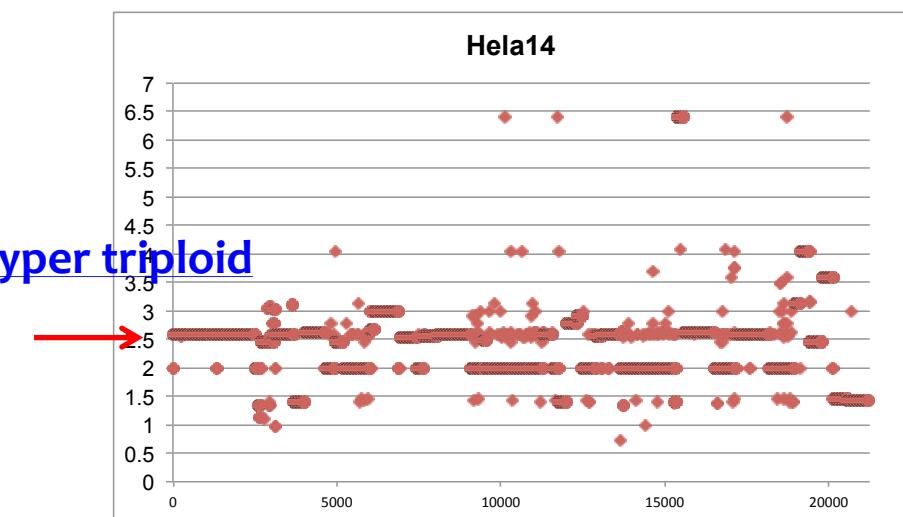
Cell doubling time



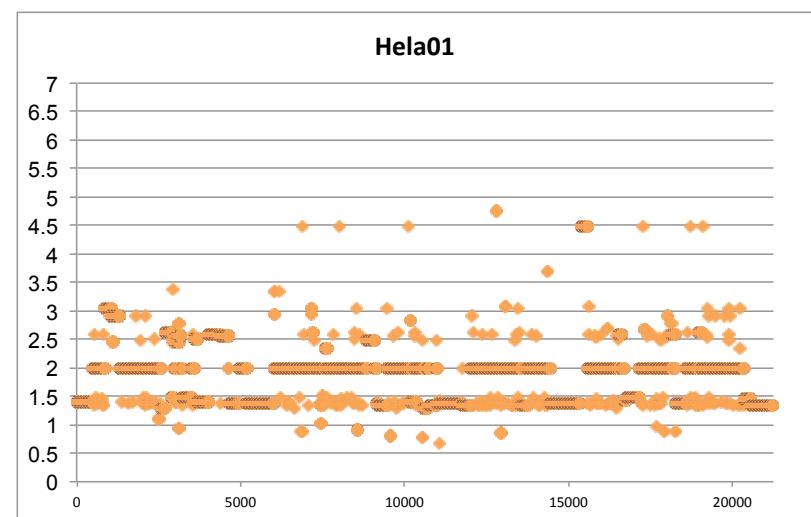
(Cell counting kit, Dojindo Laboratories)

Copy number variation (CNV) by arrayCGH

Hela CCL2

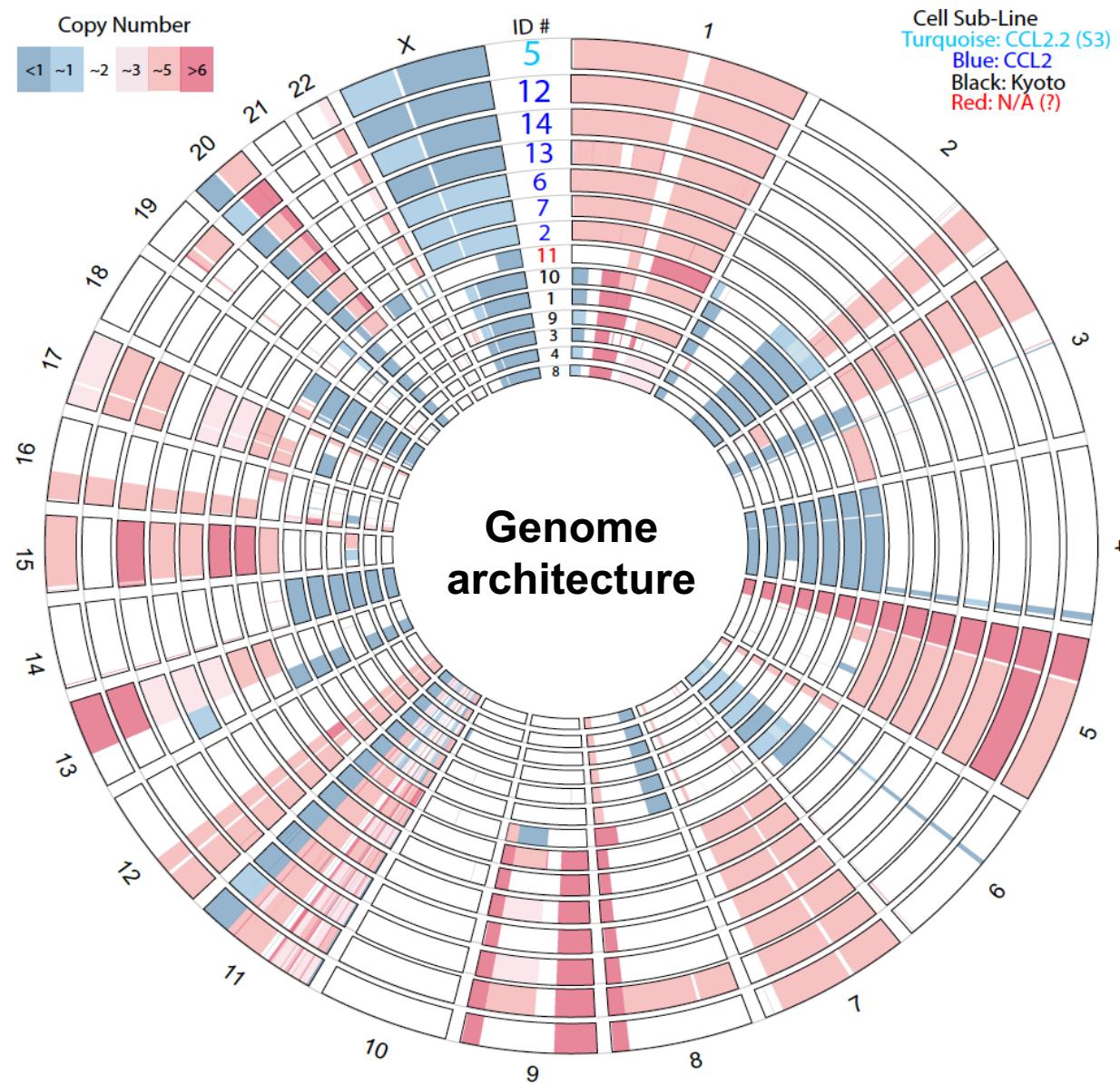


Hela
Kyoto



Agilent Human Genome CGH Microarray Kit G3 180K (Agilent Technologies, Palo Alto, USA);
(Control is DNA pool of 7 normal individuals)

CNV heterogeneity of Hela cells across labs



Proteome by SWATH-MS

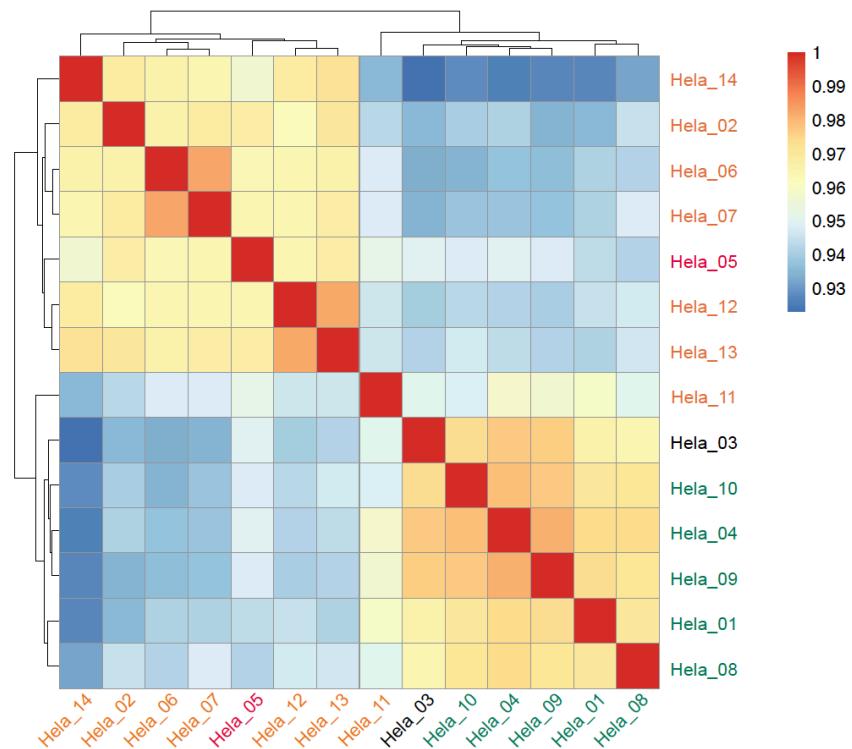
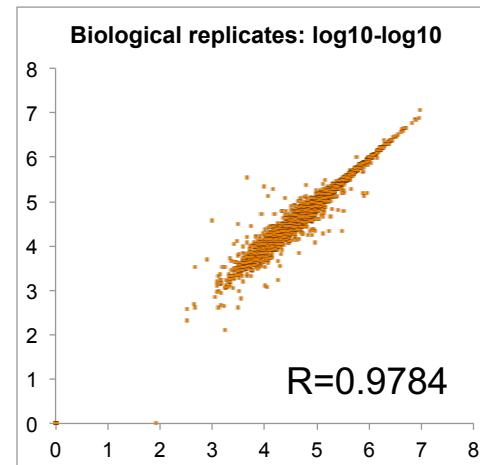
OpenSWATH + JumboProphet

- 64 windows, 90mins for steady state; 60mins for responsive experiment.
 - Two biological replicates
 - Centroid data
 - **Data procession:** weight.txt based on pyprophet2.0 → fixed → assay level FDR < 001

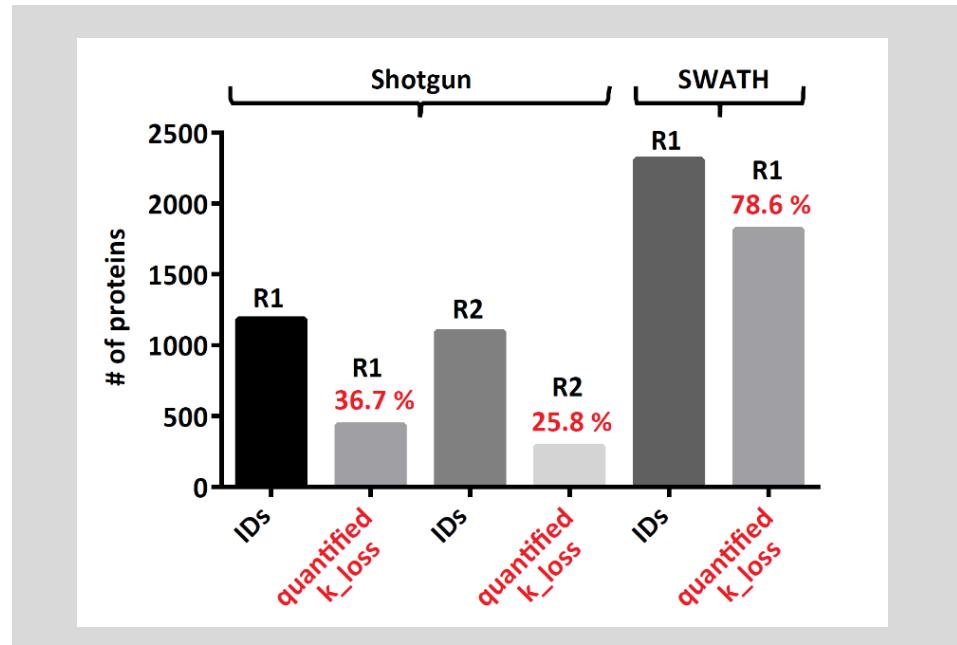
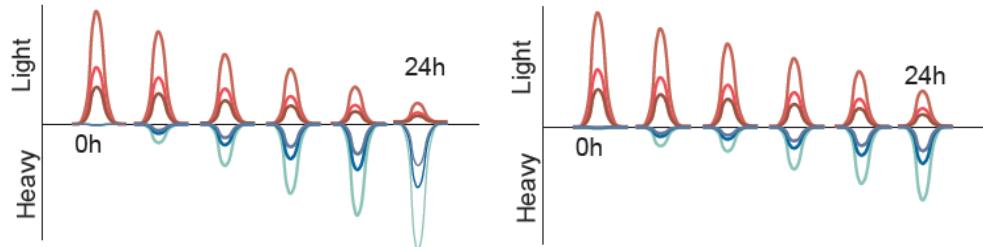
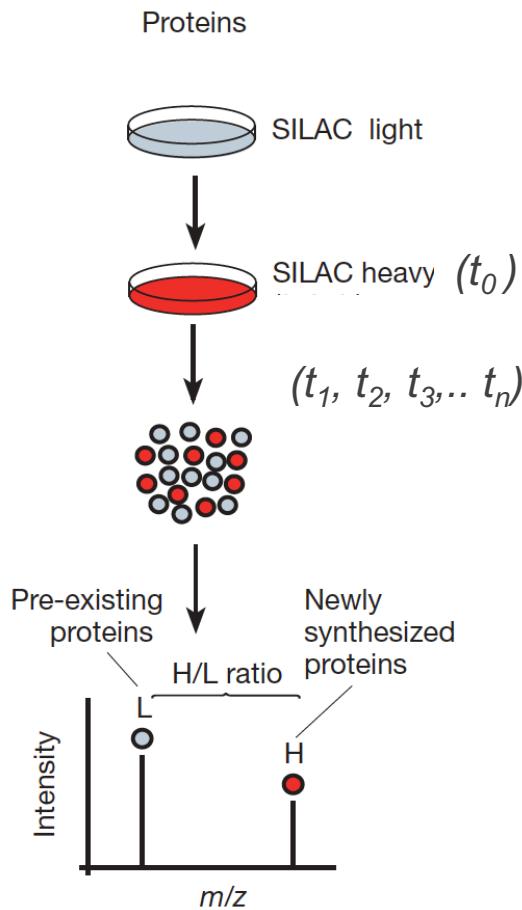
Filter 1: protein FDR<0.01 (Pyprophet 2.0): 4335 proteins.

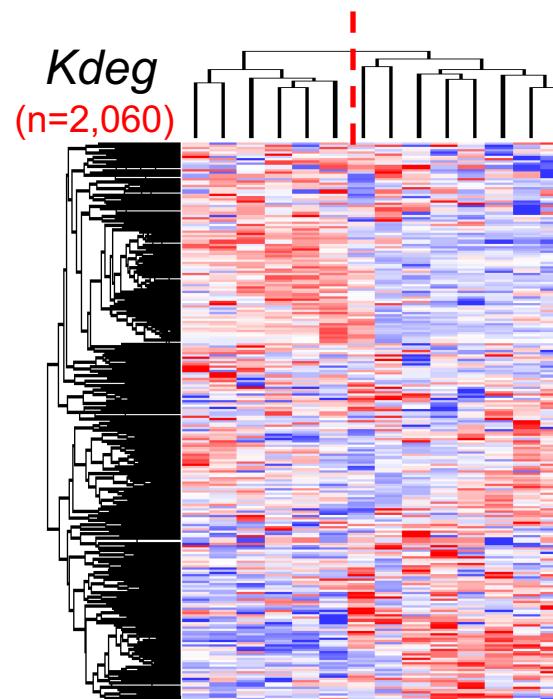
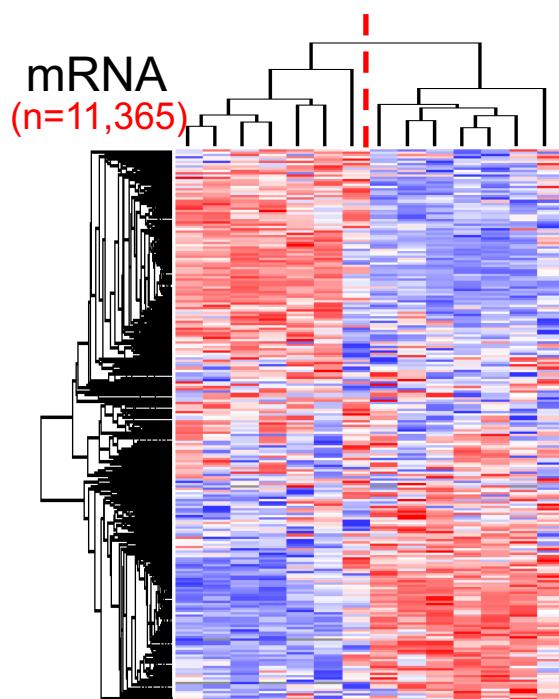
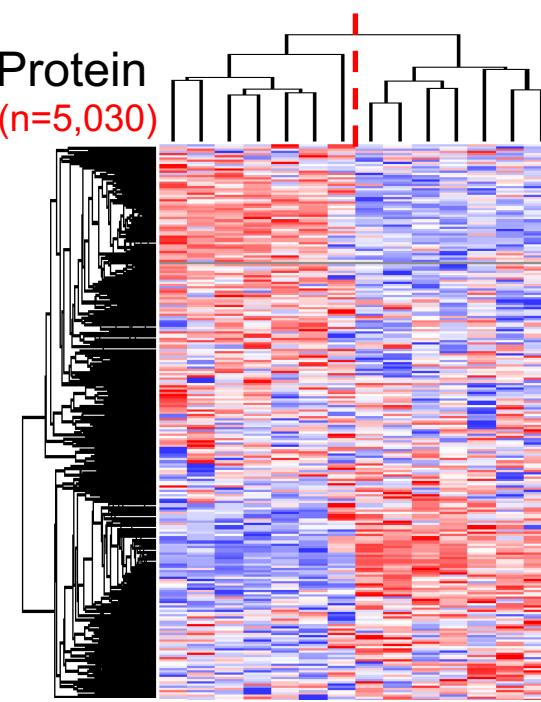
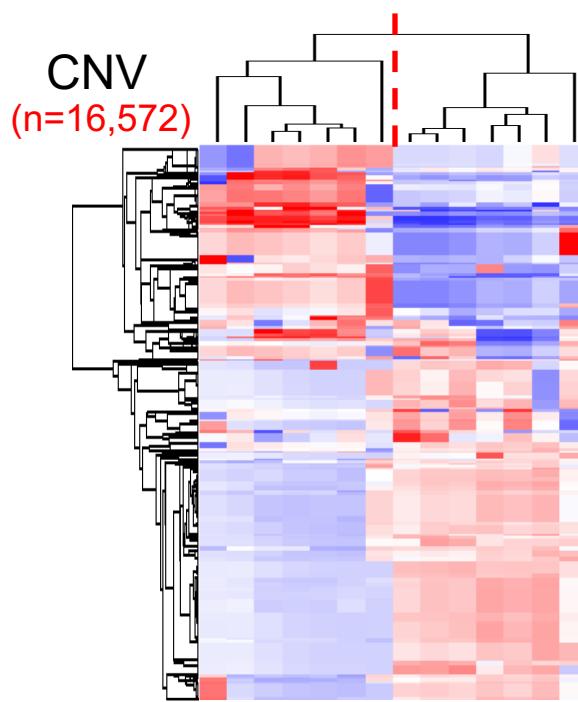
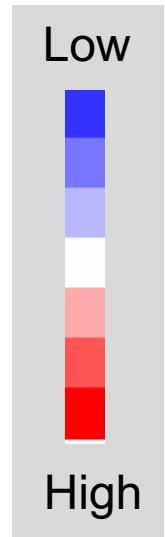
Filter 2: protein FDR<005 but identified in >25% of all 96 MS runs (i.e., requantified values less than 72 runs.)

- Totally 46951 unique peptides (43521 peptides are proteotypic, belong to 1/SW)
 - Totally 50225 precursors accepted
 - Totally 5032 unique SwissProt proteins accepted!!



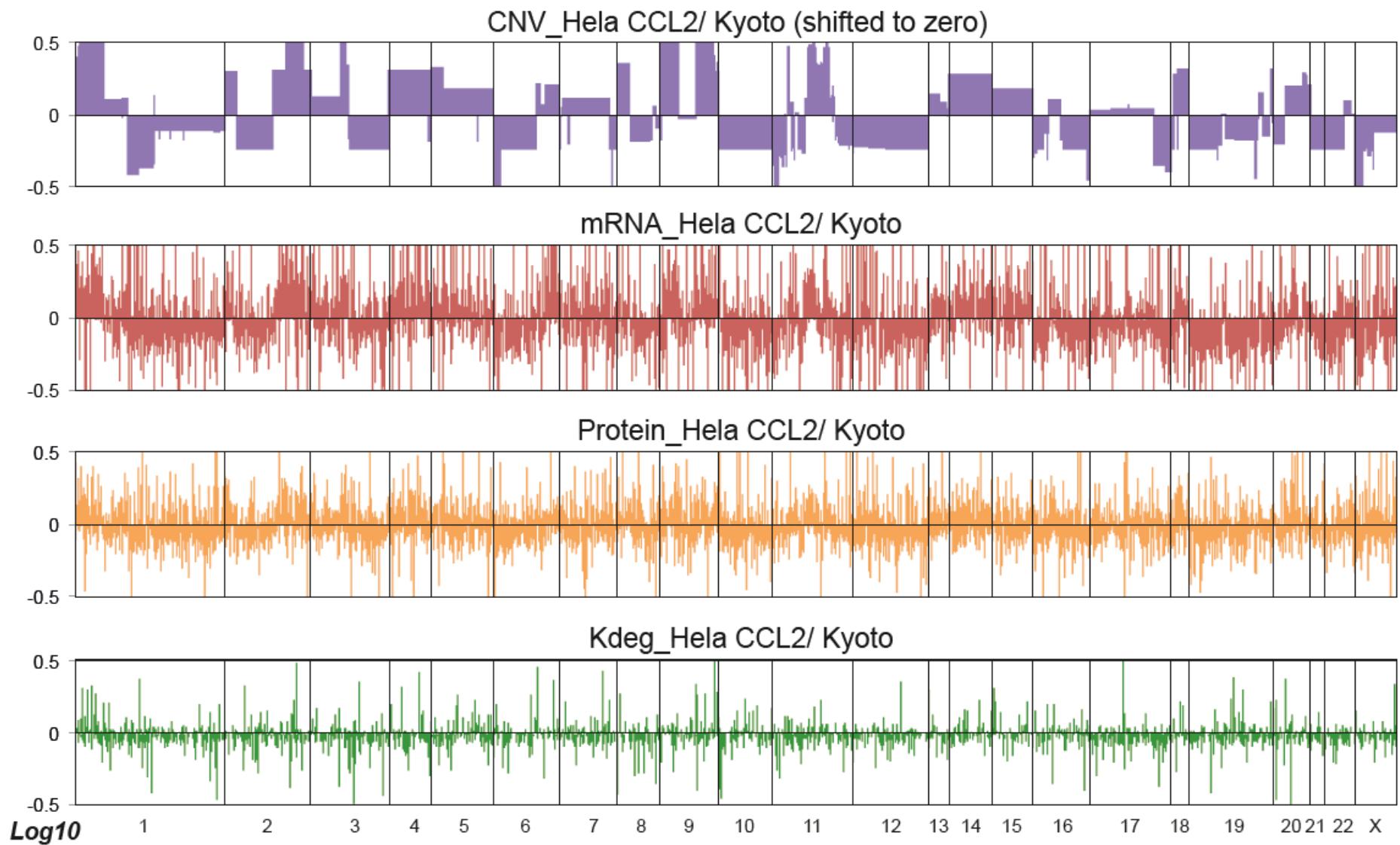
Protein turnover measurement by pSILAC-SWATH



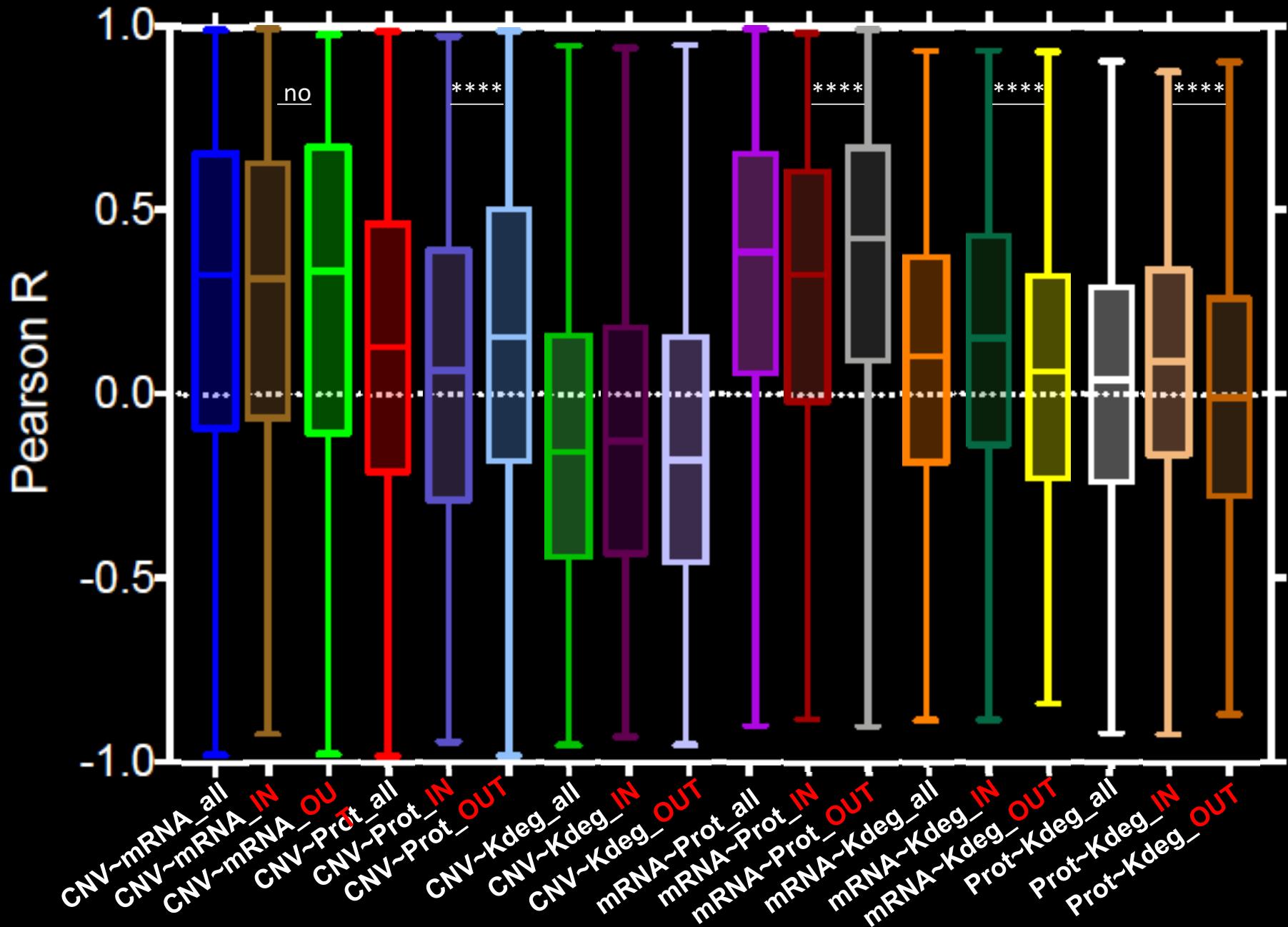


2D-centered

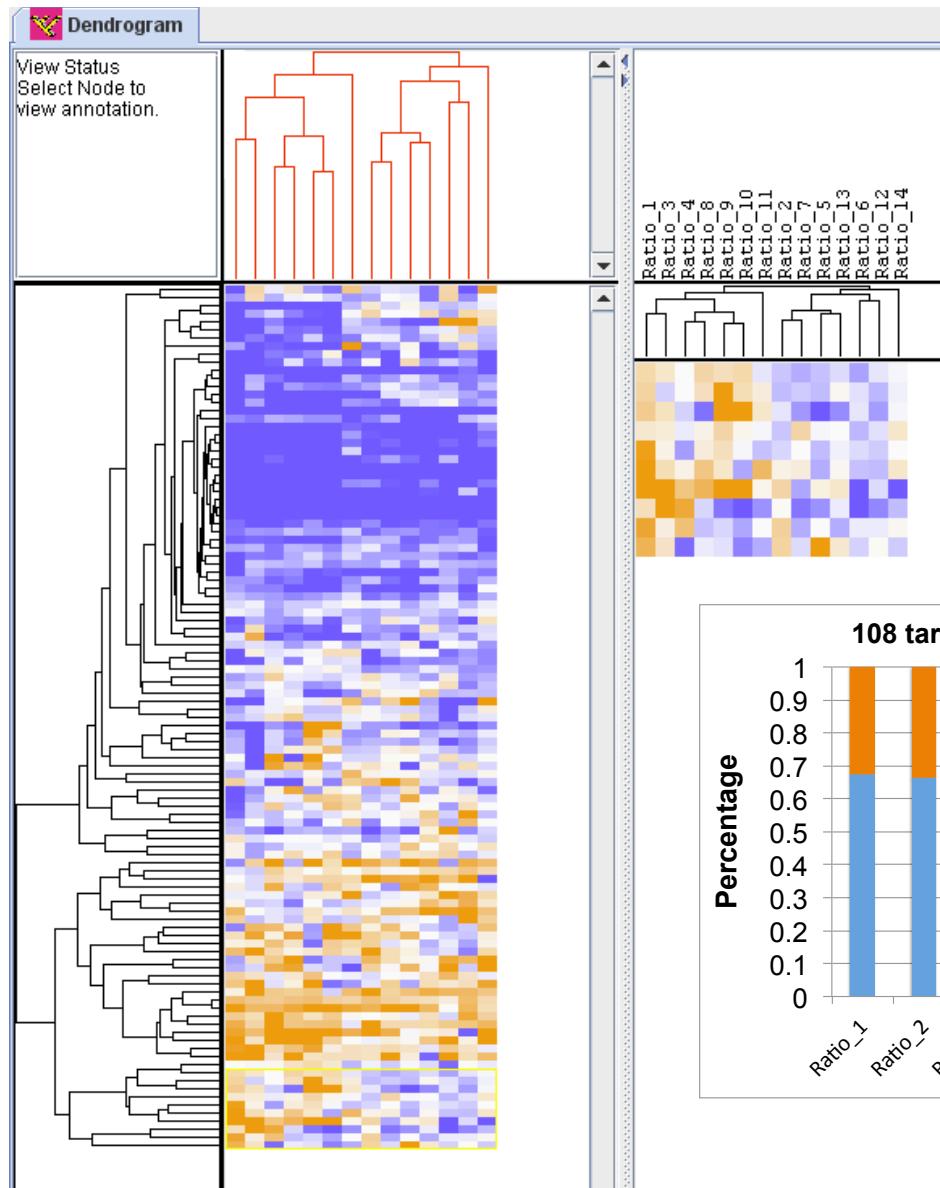
Genetic impact at gene expression layers



Correlation analysis across 14 Hela cell lines

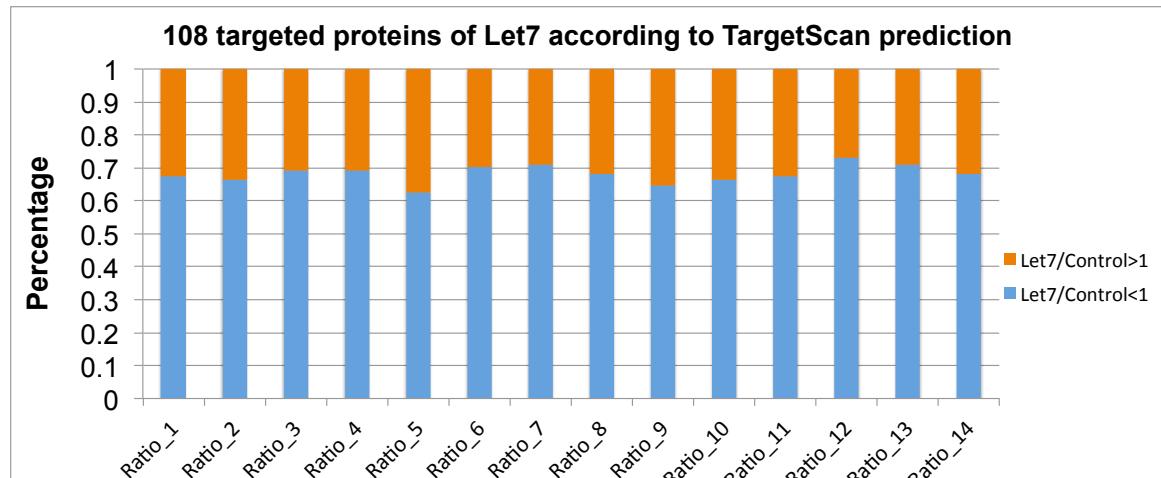


A routine experiment characterizing let7 mimics effect on Hela and S.Tm infection

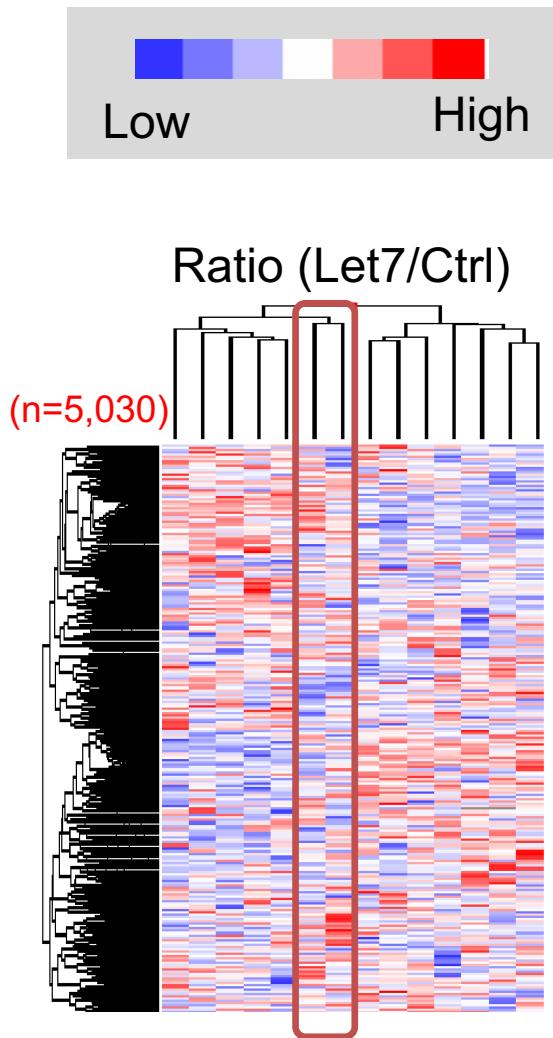


Note card for let7

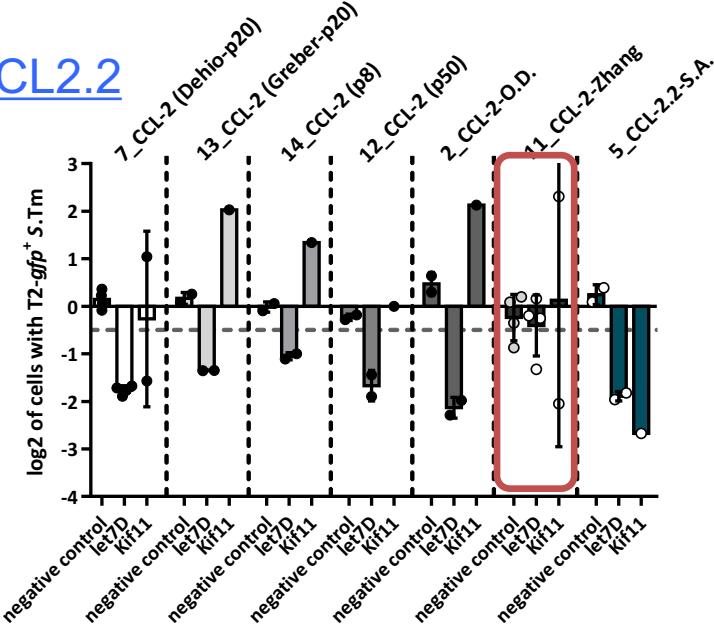
- Let-7 miRNAs have been predicted or experimentally confirmed.
- let-7 is closely associated with human cancer and acts as a tumor suppressor.
- Let-7 has been implicated in post-transcriptional control of innate immune responses to pathogenic agents.



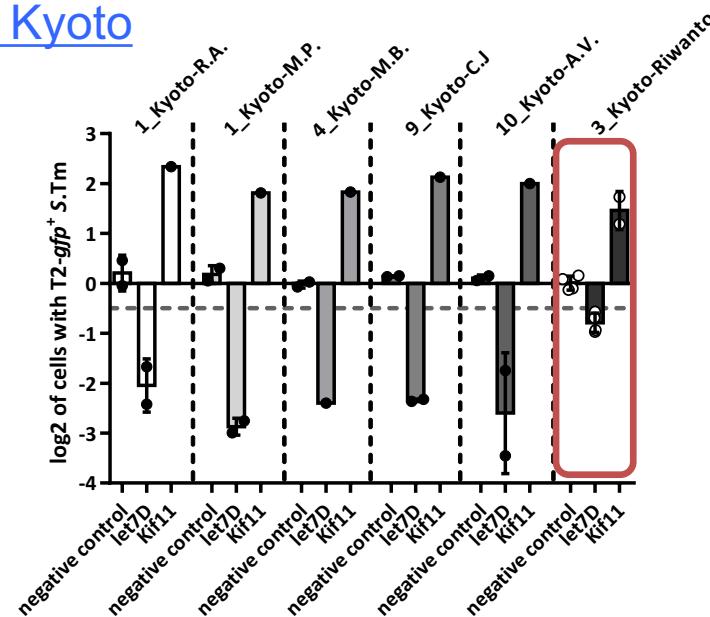
Phenotype heterogeneity



Hela CCL2.2

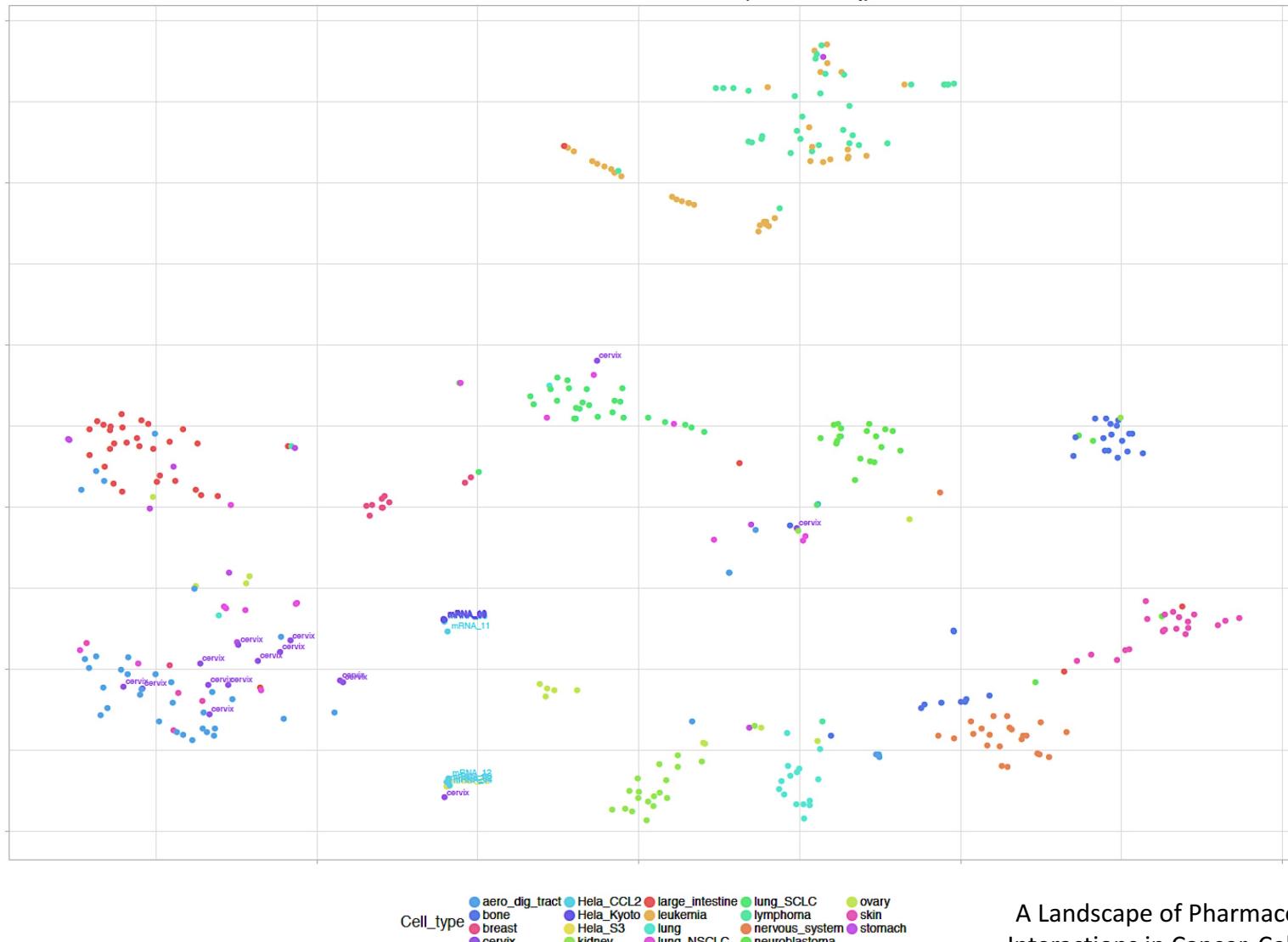


Hela Kyoto



Discussion (1): How to understand the difference between CCL2 and Kyoto

t-SNE Hela vs GDSC (RNAseq)



Summary

- Hela CCL2 and Kyoto are extremely different and should be reported clearly in the future.
- Multilayered expression data (especially turnover data) reveals significant dosage compensation through the control of protein complex stoichiometry.
- mRNA, protein, and protein degradation vary with different extent and impact divergent pathways among Hela cells.
- Different Hela cells show varied responsive proteome for Let7 treatment, leading to S.Tm infection divergence, interrogating the “reproducibility” issue from the angle of cell line used.
- Implication for reporting research results??

Thank you for your attention!!

Best success with your research