

Proteomics & Protein-Protein Interactions

Jesse Rinehart, PhD

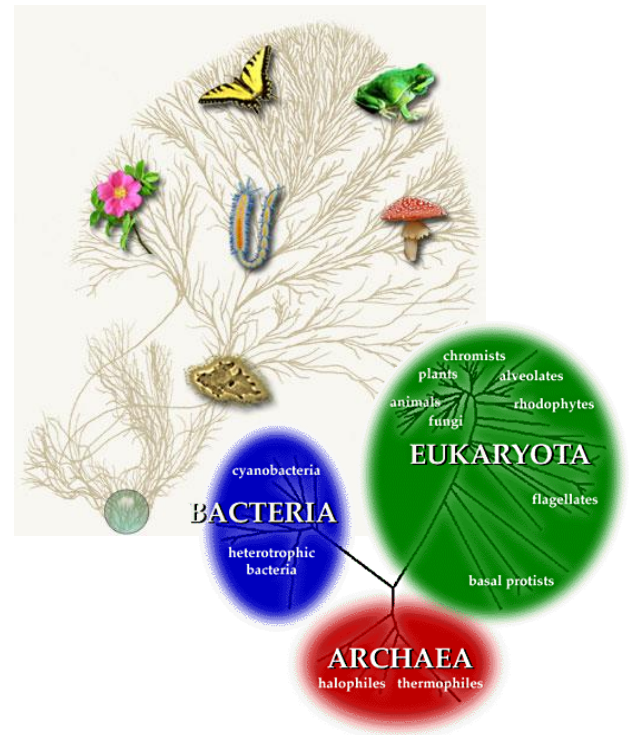
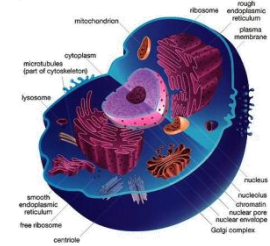
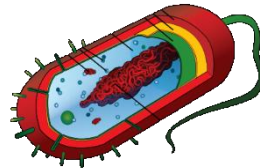
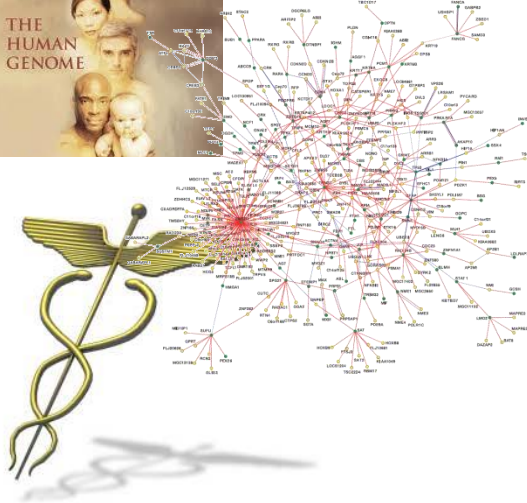
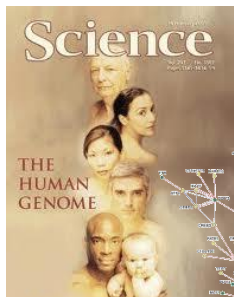
**Biomedical Data Science: Mining & Modeling
CBB 752, Spring 2018**



**Cellular & Molecular Physiology
Yale University School of Medicine**



DNA → RNA → PROTEIN



DNA → RNA → PROTEIN

2007

Cell

Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors

Kazutoshi Takahashi,¹ Koji Tanabe,¹ Mari Ohnuki,¹ Megumi Narita,^{1,2} Tomoko Ichisaka,^{1,2} Kiichiro Tomoda,³
and Shinya Yamanaka^{1,2,3,4,*}

2012

The Nobel Prize in Physiology or Medicine 2012 jointly to :

John B. Gurdon and Shinya Yamanaka

“for the discovery that mature cells can be reprogrammed to
become pluripotent”

DNA → RNA → PROTEIN

RNA-Guided Human Genome Engineering via Cas9

2013

Prashant Mali,^{1*} Luhan Yang,^{1,3*} Kevin M. Esvelt,² John Aach,¹ Marc Guell,¹ James E. DiCarlo,⁴ Julie E. Norville,¹ George M. Church^{1,2†}

Multiplex Genome Engineering Using CRISPR/Cas Systems

2013

Le Cong,^{1,2*} F. Ann Ran,^{1,4*} David Cox,^{1,3} Shuailiang Lin,^{1,5} Robert Barretto,⁶ Naomi Habib,¹ Patrick D. Hsu,^{1,4} Xuebing Wu,⁷ Wenyan Jiang,⁸ Luciano A. Marraffini,⁸ Feng Zhang^{1†}

[Home](#) ▶ [About Us](#) ▶ [Archives](#) ▶ [January 2016](#) ▶

U.S. Summit Draws Attention to Technology with Potential, Peril

By Karen Pallarito (HealthDay News)
Uploaded on December 21, 2015

Dec 2015

NATURE | NEWS

Chinese scientists genetically modify human embryos

Rumours of germline modification prove true — and look set to reignite an ethical debate.

David Cyranoski & Sara Reardon

22 April 2015

April 2015

ARTICLE

Aug. 2017

doi:10.1038/nature23305

Correction of a pathogenic gene mutation in human embryos

Hong Ma^{1*}, Nuria Marti-Gutierrez^{1*}, Tomonari Hayama¹, Riffat Ahmed¹, Sang-Tae Kim², Jianhui Gong^{2,6,7,8}, Don P. Wolf¹, Stephen B. Heitner¹⁰, J Shoukhrat Mitalipov^{1,10§}

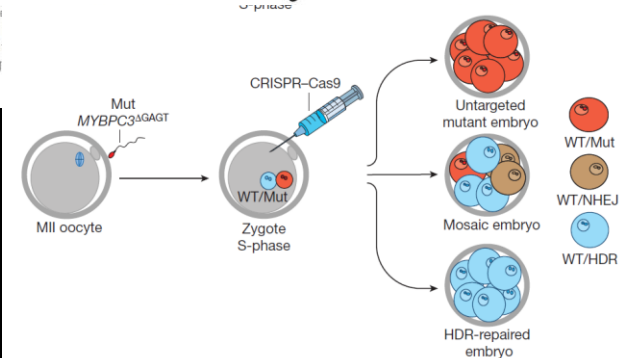
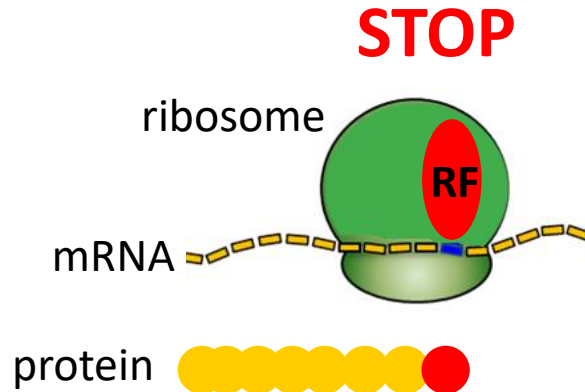


Figure 1 | Gene correction in S-phase-injected human embryos.

SYNTHETIC BIOLOGY

DNA → RNA → PROTEIN

E. coli genome editing technologies to change 321 native UAG stop codons to UAA and produced the **First Whole Genome Edited Organism**

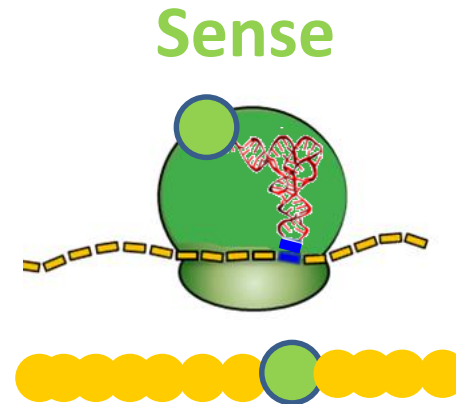
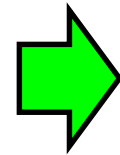


UAG=STOP

Native Genome

Uses UAG, UAA, UGA

RF1= STOP at UAG



UAG=Sense

Native Genome

Uses only UAA, UGA and RF2

RF1 deletion

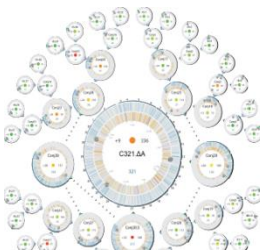
Can install new amino acid 

Genomically Recoded Organisms Expand Biological Functions

Marc J. Lajoie,^{1,2} Alexis J. Rovner,^{3,4} Daniel B. Goodman,^{1,5} Hans-Rudolf Aerni,^{4,6} Adrian D. Haimovich,^{3,4} Gleb Kuznetsov,¹ Jaron A. Mercer,⁷ Harris H. Wang,⁸ Peter A. Carr,⁹ Joshua A. Mosberg,^{1,2} Nadin Rohland,¹ Peter G. Schultz,¹⁰ Joseph M. Jacobson,^{11,12} Jesse Rinehart,^{4,6} George M. Church,^{1,13*} Farren J. Isaacs^{3,4*}

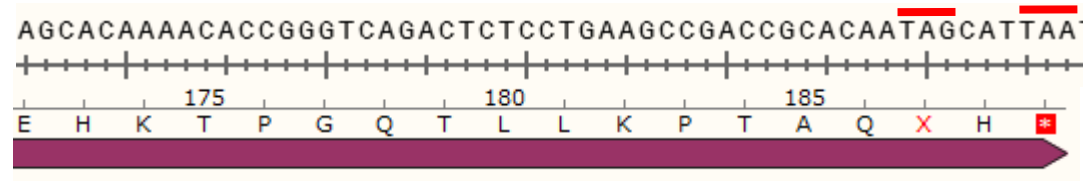
SCIENCE VOL 342 18 OCTOBER 2013

(Lajoie et al. Science 2013:PMID: 24136966)

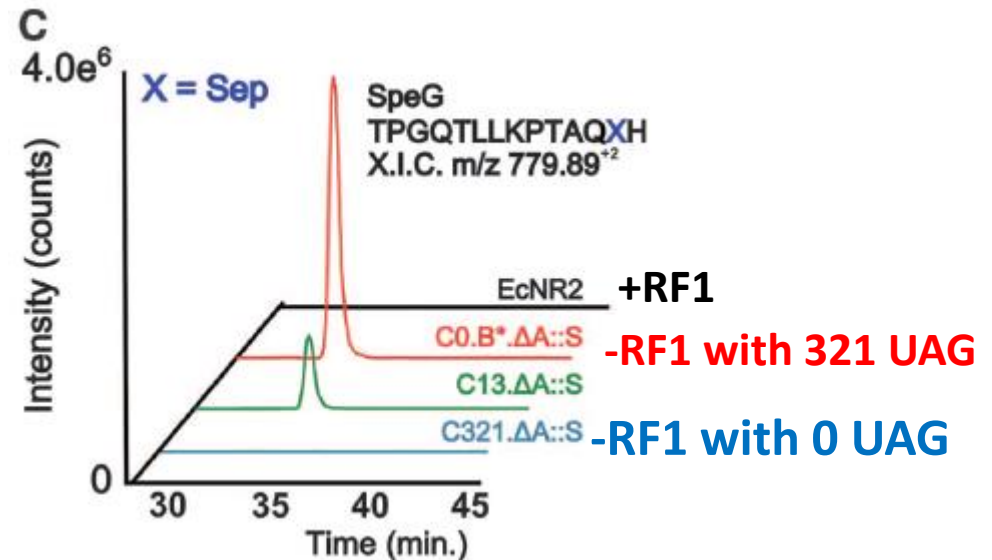


Whole genome editing = Whole *proteome* editing

STOP at native UAG or translation to next in-frame TAA



Translation through 321 native UAG **STOP** codons was ablated with genome editing



Proteomics

The study of the expression, location, modification, interaction, function, and structure of all the proteins in a given cell, organelle, tissue, organ, or whole organism.

Proteomics & Protein-Protein Interactions

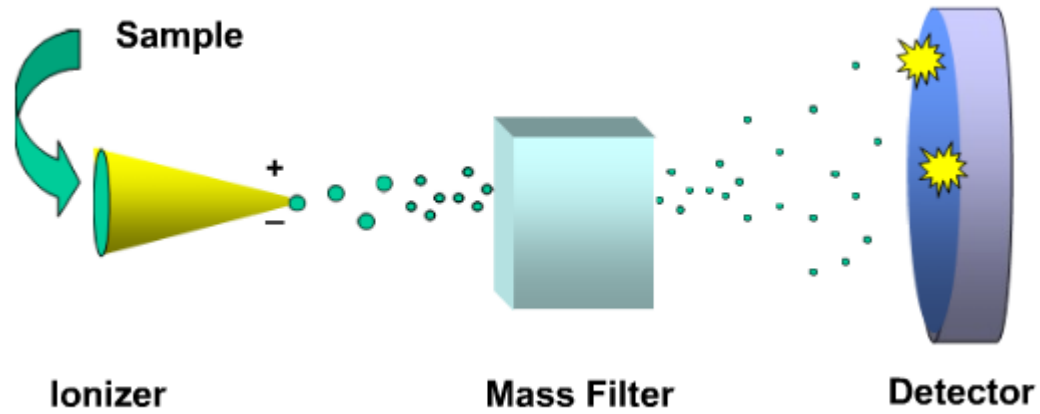
Overview

- **Techniques & Technologies**
 - Mass Spectrometry
 - Protein-Protein Interactions
 - Quantitative Proteomics
- **Applications**
 - Representative Studies
- **Putting it all together....**
 - Databases & Pathways

Principles of Mass Spectrometry (MS)

- In a mass spectrum we measure m/z (mass-to-charge)
- For proteins we measure peptide m/z
- A sample has to be ionizable in order to be analyzed

Basic Components of a Mass Spectrometer



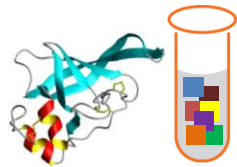
Two major ionization techniques enabled the success of mass spectrometry in the life sciences.

- Electrospray Ionization (ESI)
Fenn JB, *Mann M, Meng CK, Wong SF, Whitehouse CM. *Science*. 1989
- Matrix Assisted Laser Desorption Ionization (MALDI)
Tanaka K, Waki H, Ido Y, et al. *Rapid Commun Mass Spectrom* 1988
- 2002 Nobel Prize in Chemistry awarded to
John B. Fenn & Koichi Tanaka
- Enabled direct measurement and “sequencing” of intact peptides & MS based Proteomics is born

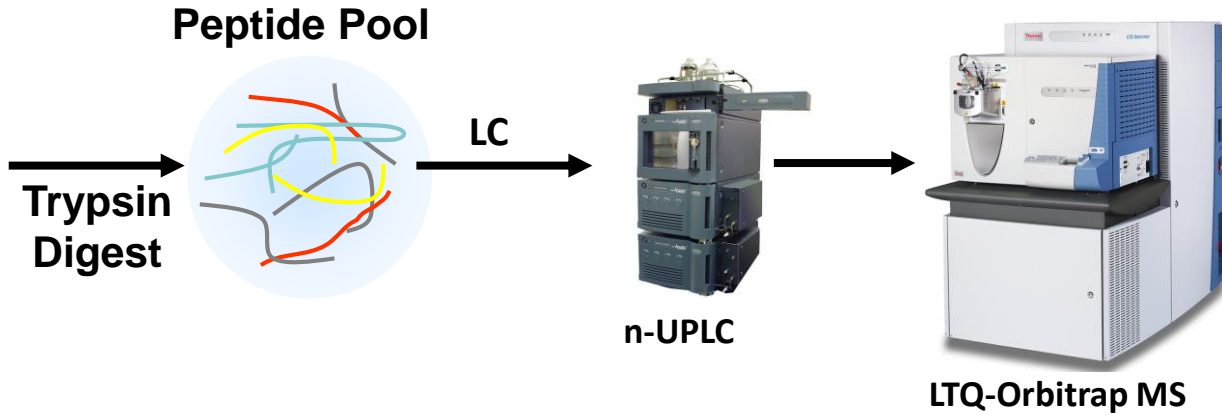
*

Matthias Mann (Yale University; Ph.D.; 1988; Chemical Engineering) trained with John Fenn during some of the breakthrough work at Yale

Typical work flow for LC-MS "shotgun proteomics"



Protein mixture



Peptide Pool

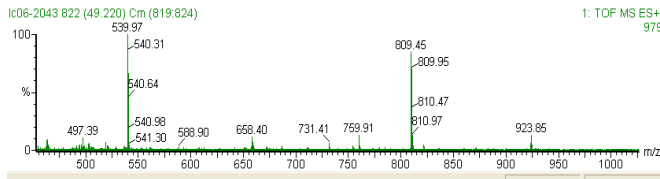
Trypsin Digest

LC

n-UPLC

LTQ-Orbitrap MS

MS



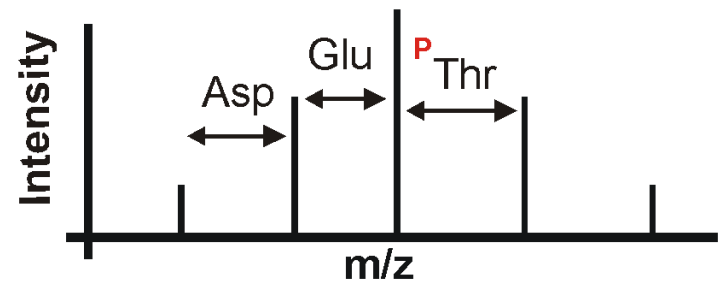
peptide

peptide

isolate & fragment

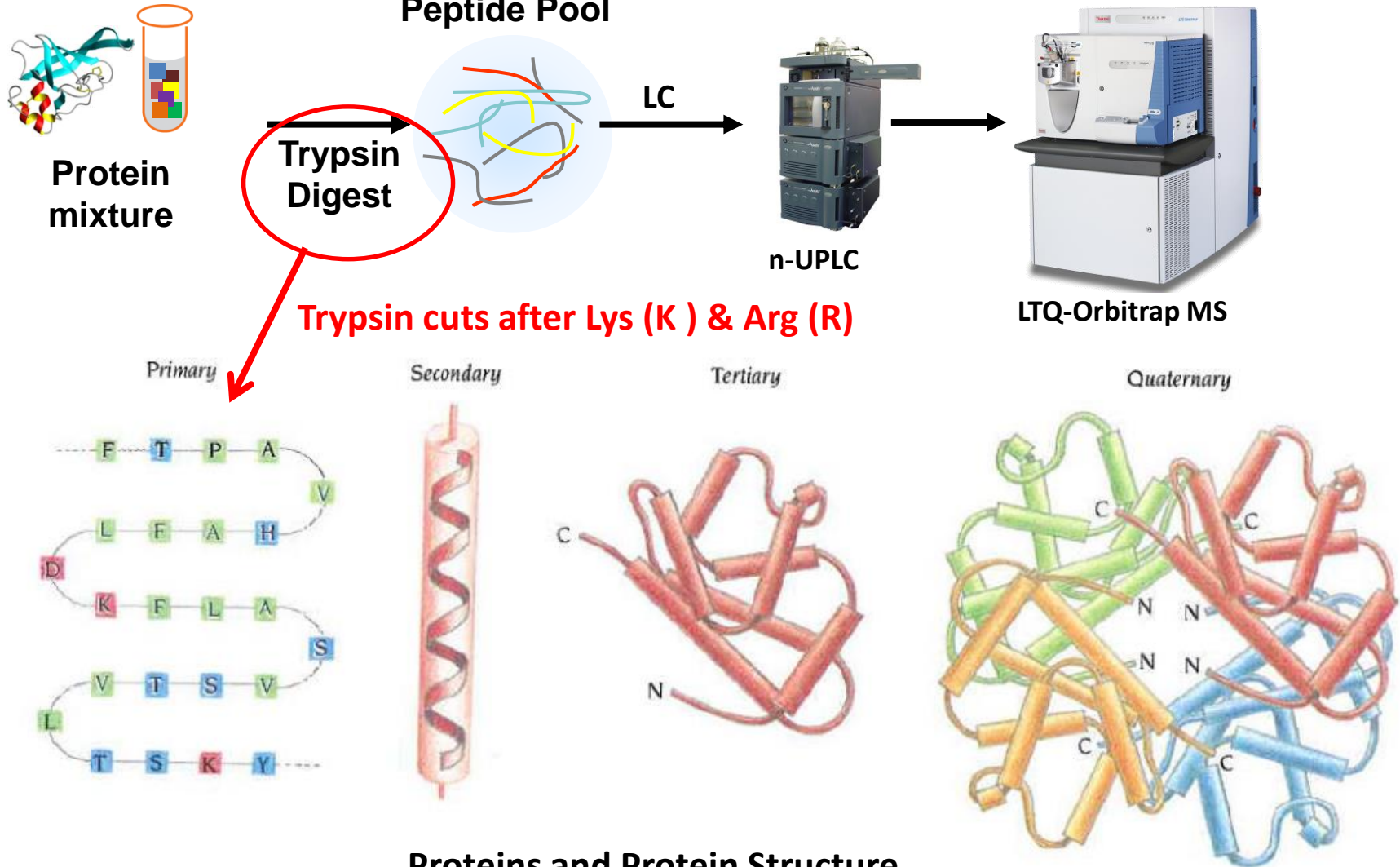
peptide

MS/MS



peptide fragments

Typical work flow for LC-MS “shotgun proteomics”



Proteins and Protein Structure

(Branden, C. and Tooze, J. *Introduction to Protein Structure*)

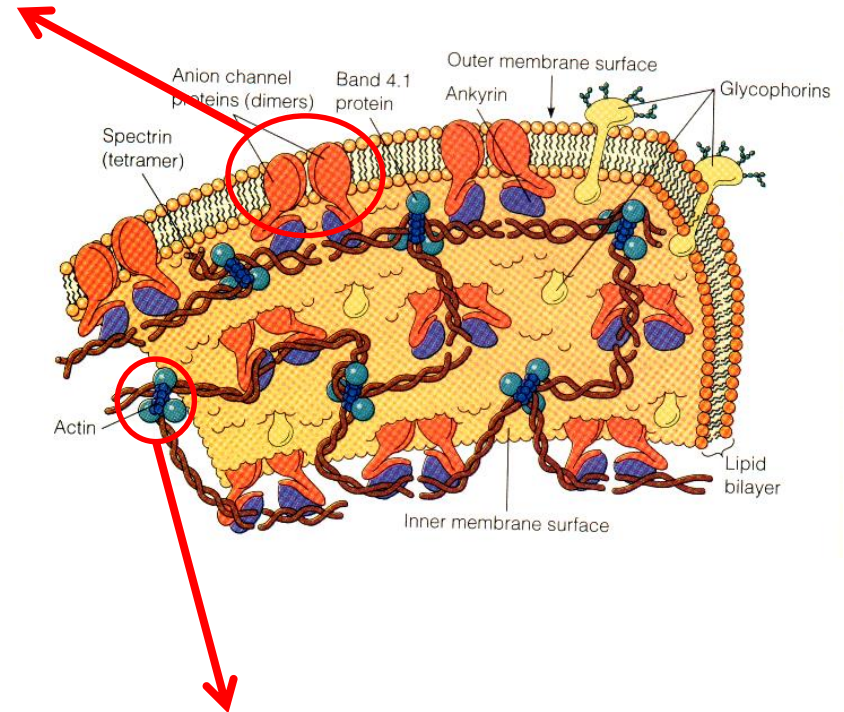
Trypsin digest followed by LC-MS: Examples of “Sequence Coverage”

Band 3 Anion Transporter

Matched peptides shown in Bold Red

```

1 MEELQDDYED MMEENLEQEE YEDPDIPESQ MEEPAAHDE ATATDYHTS
51 HPGTHKVYVE LQELVMDEKN QELRWMEEAR WVQLEENLGE NGAWGRPHLS
101 HLTFSWLEL RRVFTKGTVL LDLEQETSLAG VANQLLDRFI FEDQIRPQDR
151 EELLRALLLK HSHAGELEAL GGVKPAVLTR SGDPSQPLL QHSSLETQLF
201 CEQGDGGTEG HSPSGILEKI PPDSEATLVL VGRADFLEQP VLGFVRLQEA
251 AELEAVELPV PIRFLFVLLG PEAPHIDYTQ LGRAAATLMS ERVFRIDAYM
301 AQSRGELLHS LEGFLDCSLV LPPTDAPSEQ ALLSLVPVQR ELLRRRYQSS
351 PAKPDSSFYK GLDLNGGPDD PLQQTGQLFG GLVRDIRRRY PYLSDITDA
401 FSPQVLAAVI FIYFAALSPA ITFGGLLGEK TRNQMGVSEL LISTAVQGIL
451 FALLGAQPLL VVGFSGPLLV FEEAFFSFCE TNGLEYIVGR VWIGFWLILL
501 VVLVAFEGS FLVRFISRYT QEIFSFLISL IFIYETFSKL IKIFQDHPLQ
551 KTYNYNVLV PKPQGPLENT ALLSLVLMAG TFFFAMMLRK FKNSSYFP GK
601 LRRVIGDFGV PISILIMVLV DFFIQDTYTQ KLSVPDGFVK SNSSARGWVI
651 HPLGLRSEFP IWMMFASALP ALLVFILIFL ESQITTLIVS KPERKMKV KGS
701 GFHLDLLL V GMGGVAALFG MPWLSATTVR SVTHANALTV MGKASTPGAA
751 AQIQEVKEQR ISGLLVAVLV GLSILMEPIL SRIPLAVLFG IFLYMGVTSL
801 SGIQLFDRIL LLFKPPKYHP DVPYVKRVKT WRMHLFTGIQ IICLAVLWV
851 KSTPASLALP FVLILT VPLR RVLLPLIFRN VELQCLDADD AKATFDEEEG
901 RDEYDEVAMP V
    
```

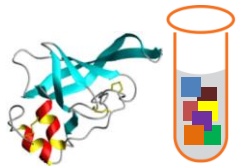


β -actin

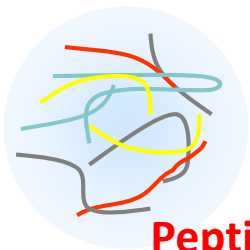
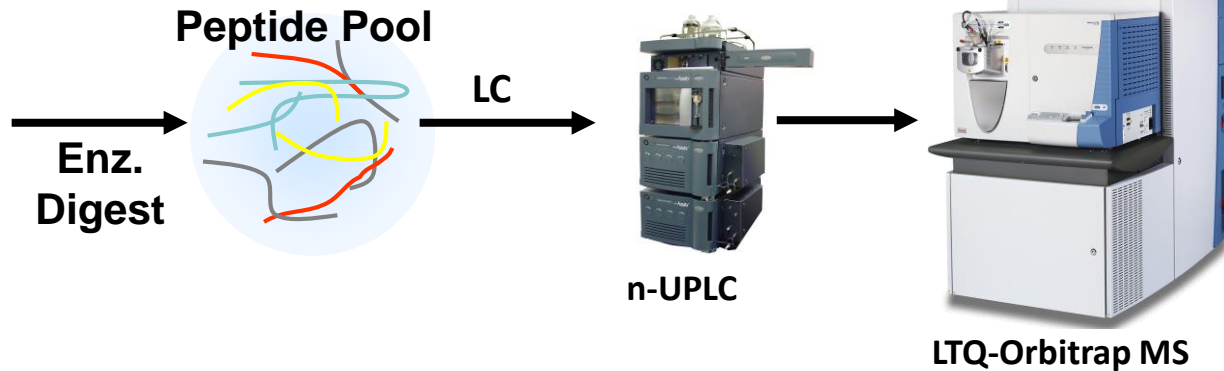
Matched peptides shown in Bold Red

```

1 MDDDIAALVV DNGSGMCKAG FAGDDAPRAV FPSIVGRPRH QGVVMGMGQK
51 DSYVGDEAQS KRGILTLYP IEHGIVTNWD DMEKIWHHTF YNELRVAPEE
101 HPVLLTEAPL NPKANREKMT QIMPETFNTF AMYVAIQAVL SLYASGRTTG
151 IVMDSGDGVT HTVPIYEGYA LPHAILRLDL AGRDLTDYLM KILTERGYSF
201 TTTAEREIVR DIKEKLCYVA LDPEQEMATA ASSSSLEKSY ELPDQVITI
251 GNERFRCPEA LFQPSFLGME SCGIHETTFN SIMKCDVDIR KDLYANTVLS
301 GGTMYPGIA DRMQKEITAL APSTMKIKII APPERKYSVW IGGASILASLS
351 TFQQMWISKQ EYDESGPSIV HRKCF
    
```



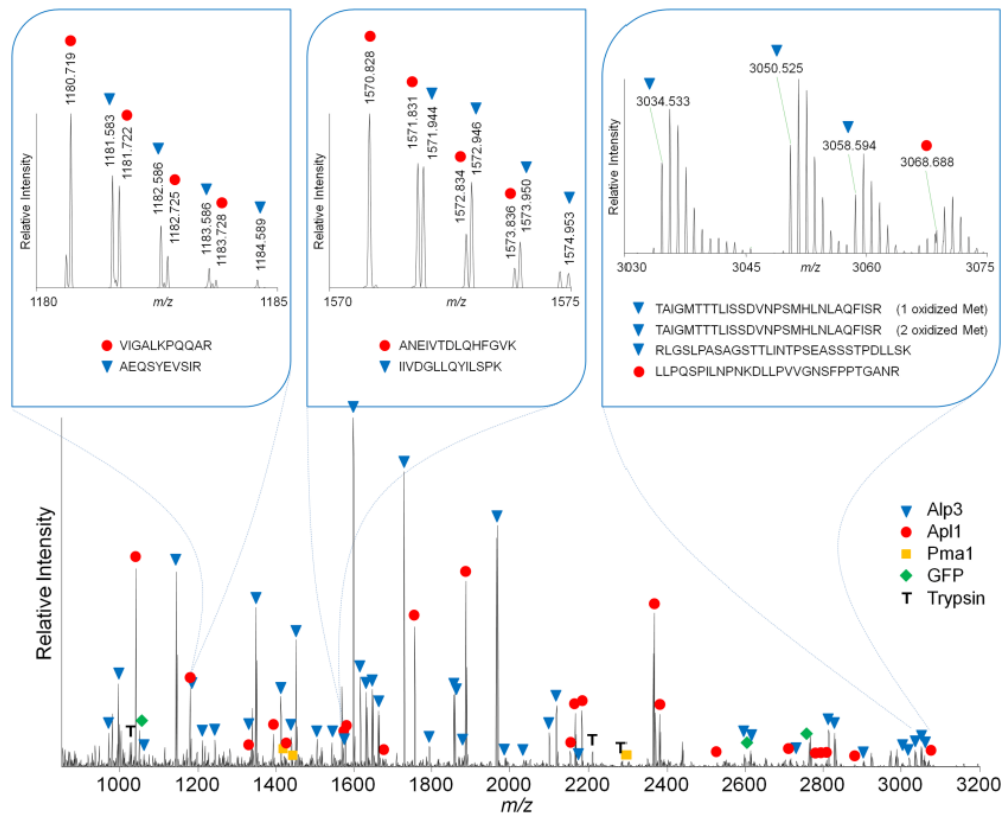
Protein mixture

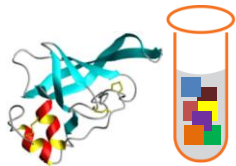


Peptide ions have a mass (m) and a charge (z).

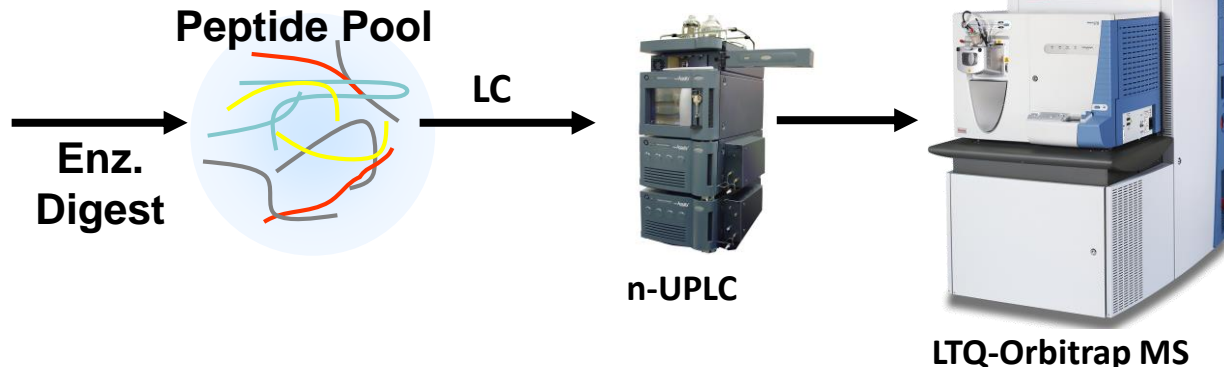
100 Da peptide:
 +1 = 100 m/z
 +2 = 50 m/z
 +3 = 33.3 m/z

Mass Spectrum

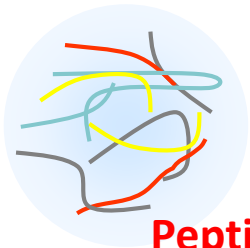




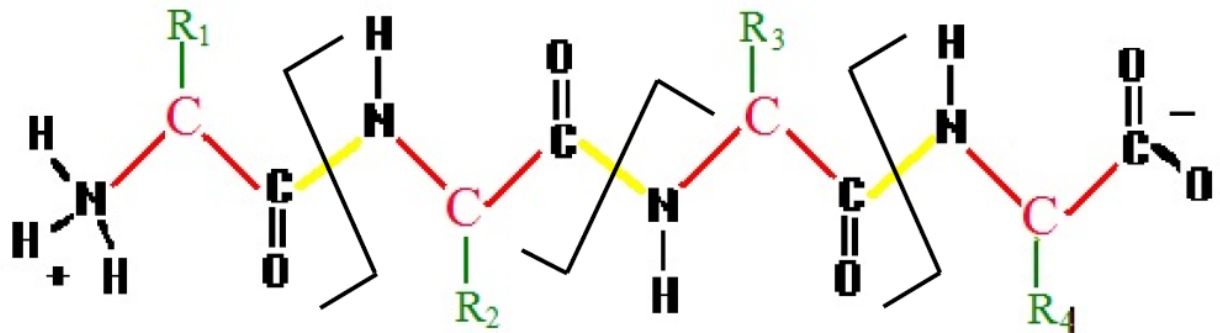
Protein mixture



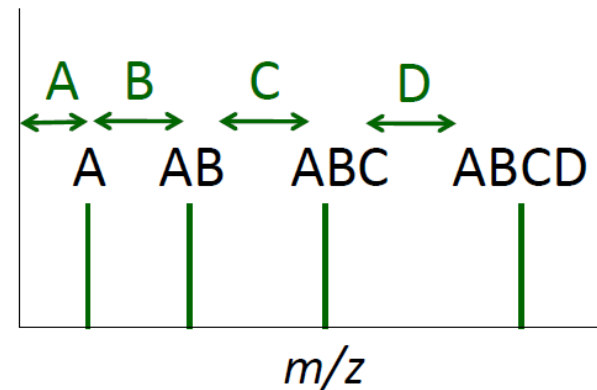
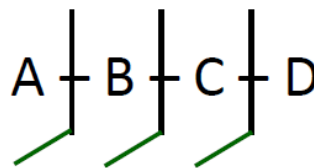
Peptide sequencing

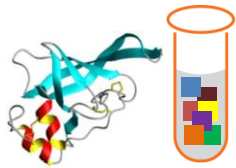


Peptide ions are isolated and "sequenced"

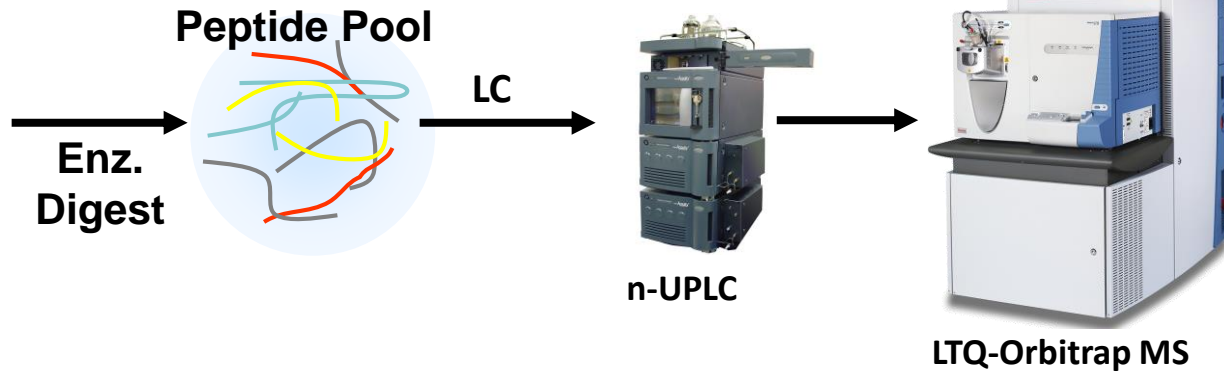


Simplified concept of peptide fragmentation

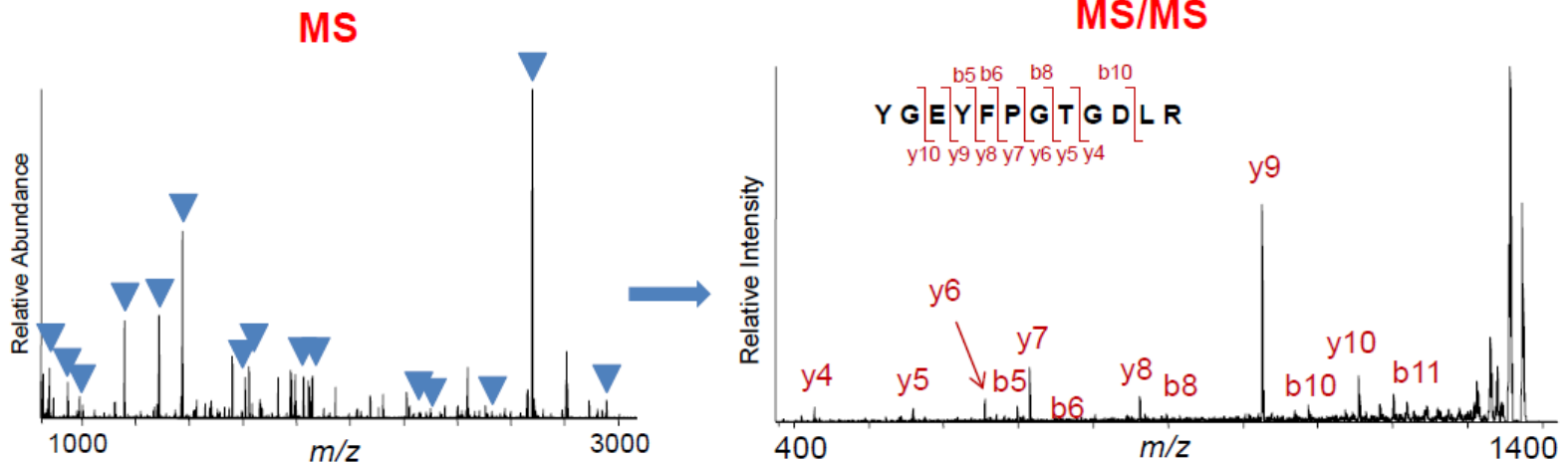




Protein mixture



Database searching - at MS or MS/MS level



Computational Steps: massive amounts of MS data are read & interpreted. Databases searched to match peptide sequences.

The ***pace of proteomics is set by a combination of techniques and technological advances.**

***orders of magnitude behind genome technologies (*sequencing*)**

Yeast proteome reported in Washburn et al. *Nature Biotech* 2001:

~82 hours* = 1,484 proteins → ~0.3 proteins/ min

***estimates from paper: 3 fractions @ 15 X 110 minute “runs” for each fraction**

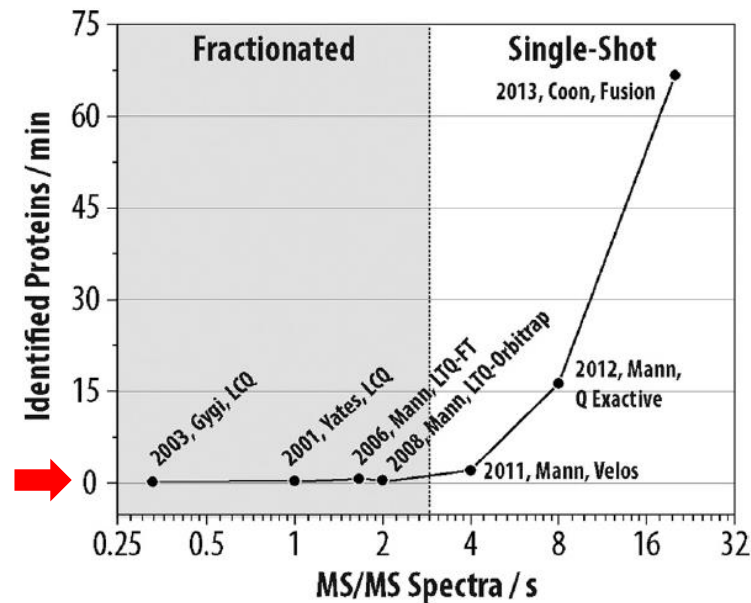


FIG. 5. Rate of protein identifications as a function of mass spectrometer scan rate for selected large-scale yeast proteome analyses over the past decade. Each data point is annotated with the year, corresponding author, type of MS system used, and reference number.

The one hour yeast proteome. Hebert AS, et al, Coon JJ.

Mol Cell Proteomics. 2014 PMID: 24143002 & *Nat Protoc*. 2015. PMID: 25855955

The ***pace of proteomics is set by a combination of techniques and technological advances.**

***orders of magnitude behind genome technologies (*sequencing*)**

Yeast proteome reported in Washburn et al. *Nature Biotech* 2001:

~82 hours* = 1,484 proteins ~0.3 proteins/ min

***estimates from paper: 3 fractions @ 15 X 110 minute “runs” for each fraction**

Technological Innovation and Resources

✦ Author's Choice

© 2014 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at <http://www.mcponline.org>

The One Hour Yeast Proteome*

Alexander S. Hebert^{‡§**}, Alicia L. Richards^{§¶**}, Derek J. Bailey^{§¶}, Arne Ulbrich[¶],
Emma E. Coughlin[§], Michael S. Westphall[§], and Joshua J. Coon^{‡§¶||}

On average, each **one hour** analysis achieved detection of **3,977 proteins**

PROTOCOL

One-hour proteome analysis in yeast

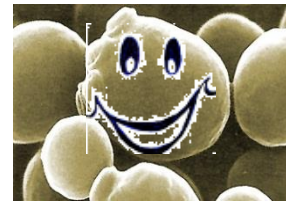
Alicia L Richards^{1,2,4}, Alexander S Hebert^{1,3,4}, Arne Ulbrich^{1,2}, Derek J Bailey^{1,2}, Emma E Coughlin¹,
Michael S Westphall¹ & Joshua J Coon¹⁻³

“ ...the identification of up to **4,002 proteins**, This protocol, which includes cell lysis, overnight tryptic digestion, sample analysis and database searching, **takes ~24 h to complete.**”

The one hour yeast proteome. Hebert AS, et a, Coon JJ.

Mol Cell Proteomics. 2014 PMID: 24143002 & *Nat Protoc*. 2015. PMID: 25855955

A tour of proteomics: Studies with the budding yeast *Saccharomyces cerevisiae*



2000 & 2001

Uetz et al, A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* .
& Ito et al, A comprehensive two-hybrid analysis to explore the yeast protein interactome . *PNAS*.

➔ **Large scale yeast two hybrid screens to map proteome wide interactions.**

2001

Washburn, et al. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.*

➔ **Established the 'shotgun' technology by showing that many proteins in a yeast-cell lysate could be identified in a single experiment.**

2002

Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*.

& Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* .

➔ **Protein-protein interaction maps can be obtained by MS; the yeast cell is organized into protein complexes.**

2003

Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature*. & Huh, W. K. et al. Global analysis of protein localization in budding yeast. *Nature*.

➔ **TAP-Tag and expression studies & GFP-Tag and localization studies**

2006

Krogan NJ, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*.

➔ **TAP-Tag and Protein-Protein Interaction**

2008

de Godoy LM, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*.

➔ **SILAC based quantitation of an entire proteome.**

2009

Picotti P, et al. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*.

➔ **Towards proteome wide targeted proteomics.**

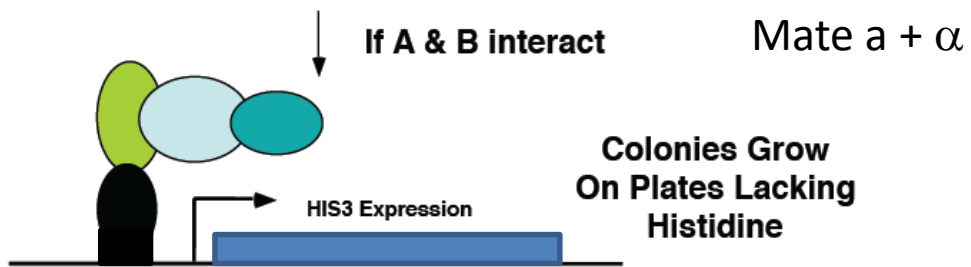
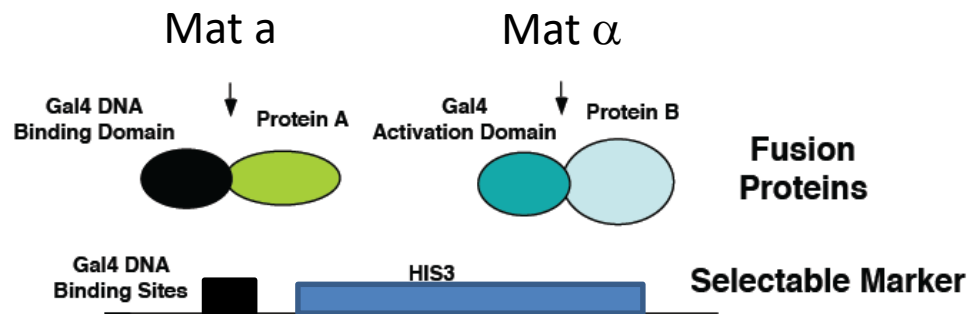
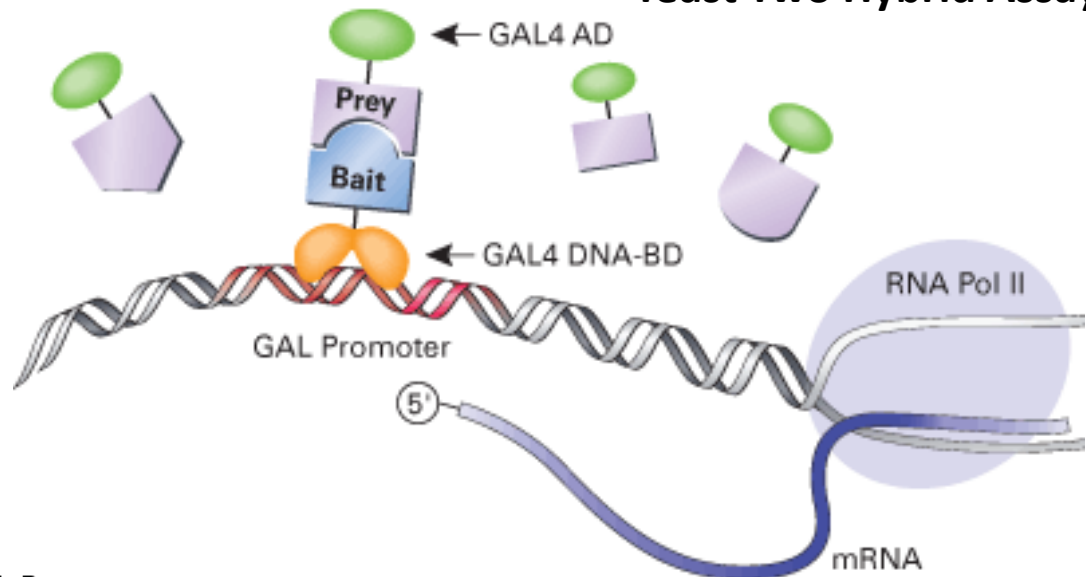
A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.

Uetz et al, Nature 2000

Ito et al, PNAS 2001

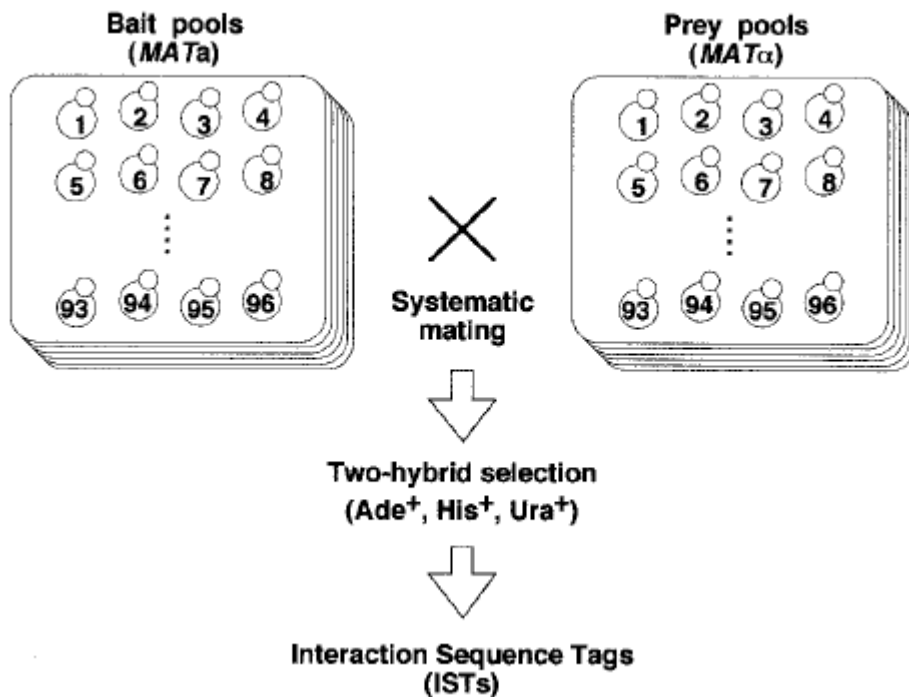
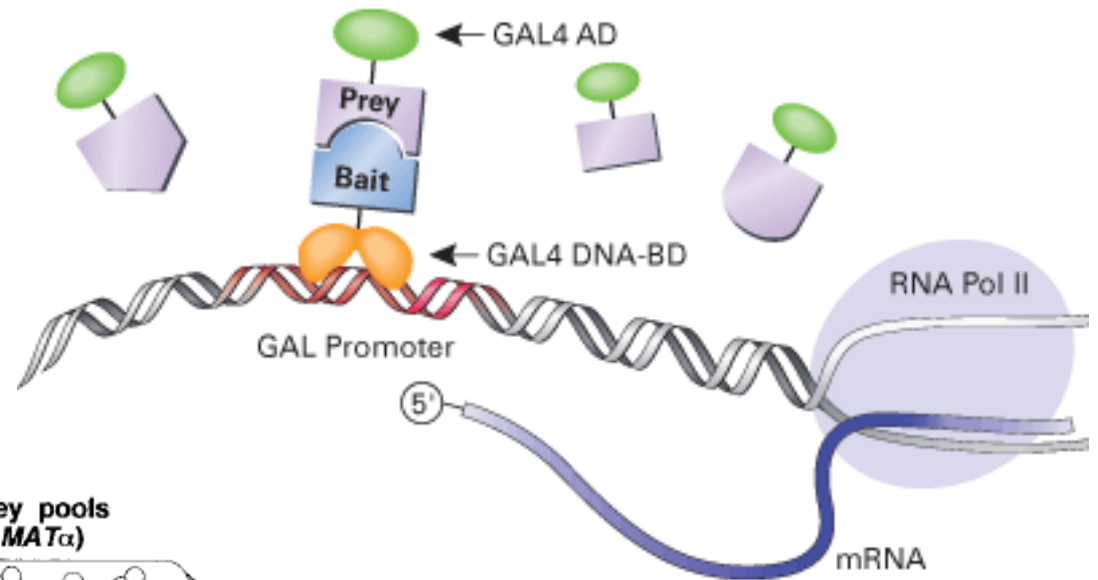
Yeast Two Hybrid Assay

Clone bait and prey constructs and place in separate strains.



Uetz et al, Nature 2000

Ito et al, PNAS 2001



Results of Two Studies

- 1) 4,549 Interactions Among 3,278 Proteins (Ito et al.)
- 2) 957 Interactions 1004 proteins (Uetz et al.)

A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.

Uetz et al, Nature 2000

Ito et al, PNAS 2001

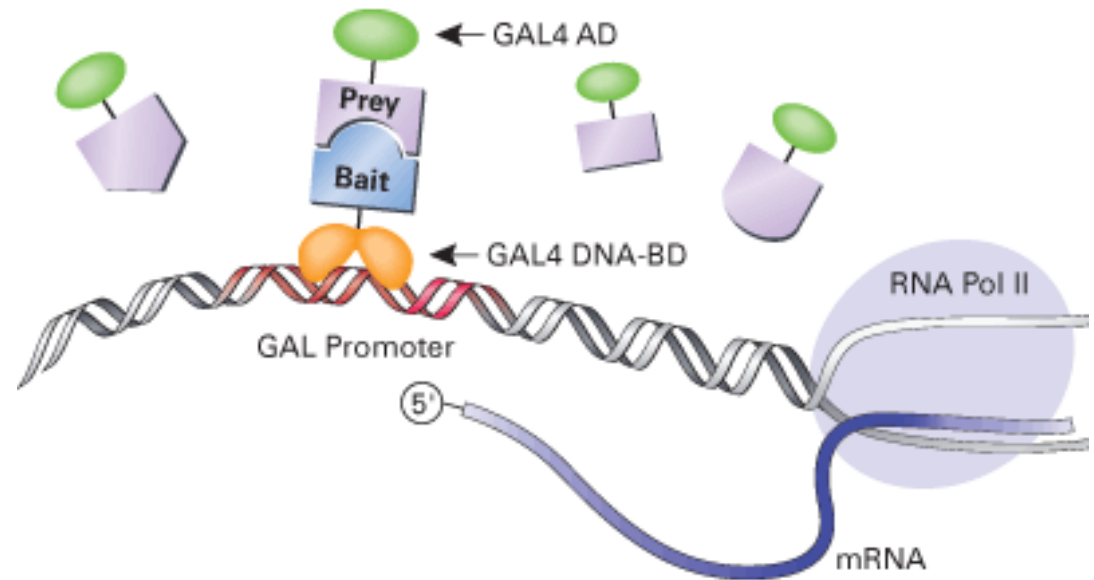
Yeast Two Hybrid Assay

Advantages:

- *In vivo* assay
- Simple

Some Disadvantages

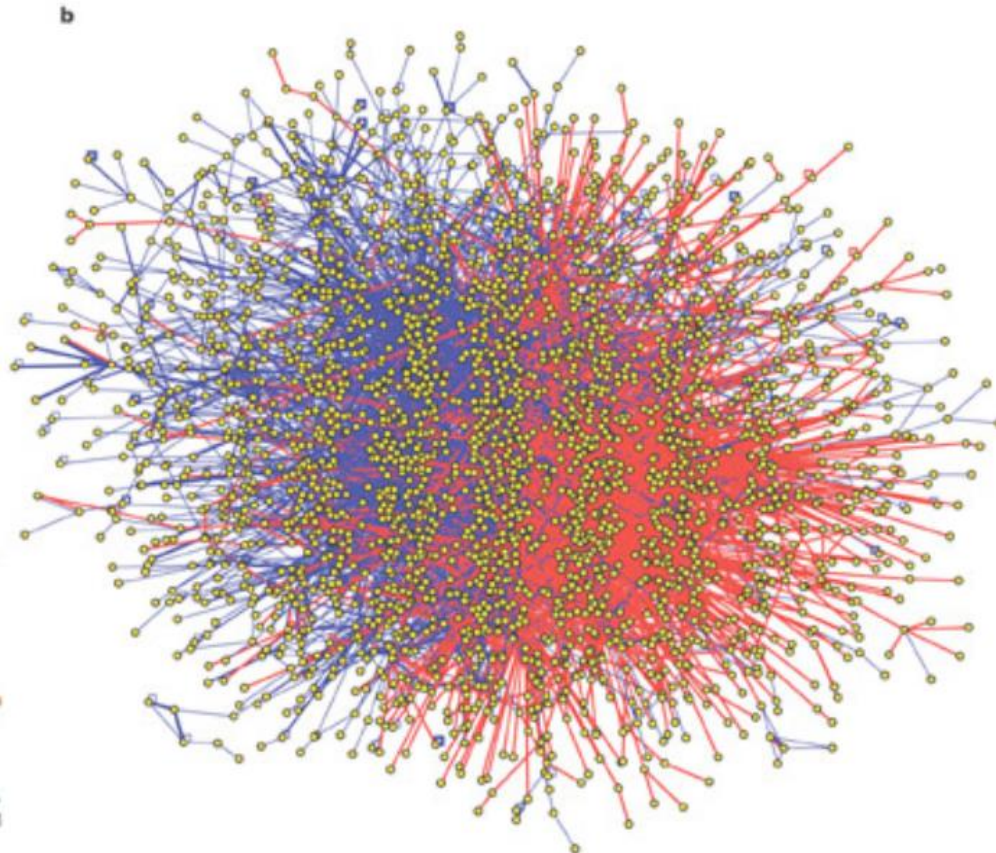
- Hard to execute on large scale
- False positives: a real interaction or “possible” interaction
- Interaction in nucleus (required for GAL system)
- Clones are fusion proteins and sometimes “partial” proteins
- Multiple protein complexes not “captured”



Human Two Hybrid Map

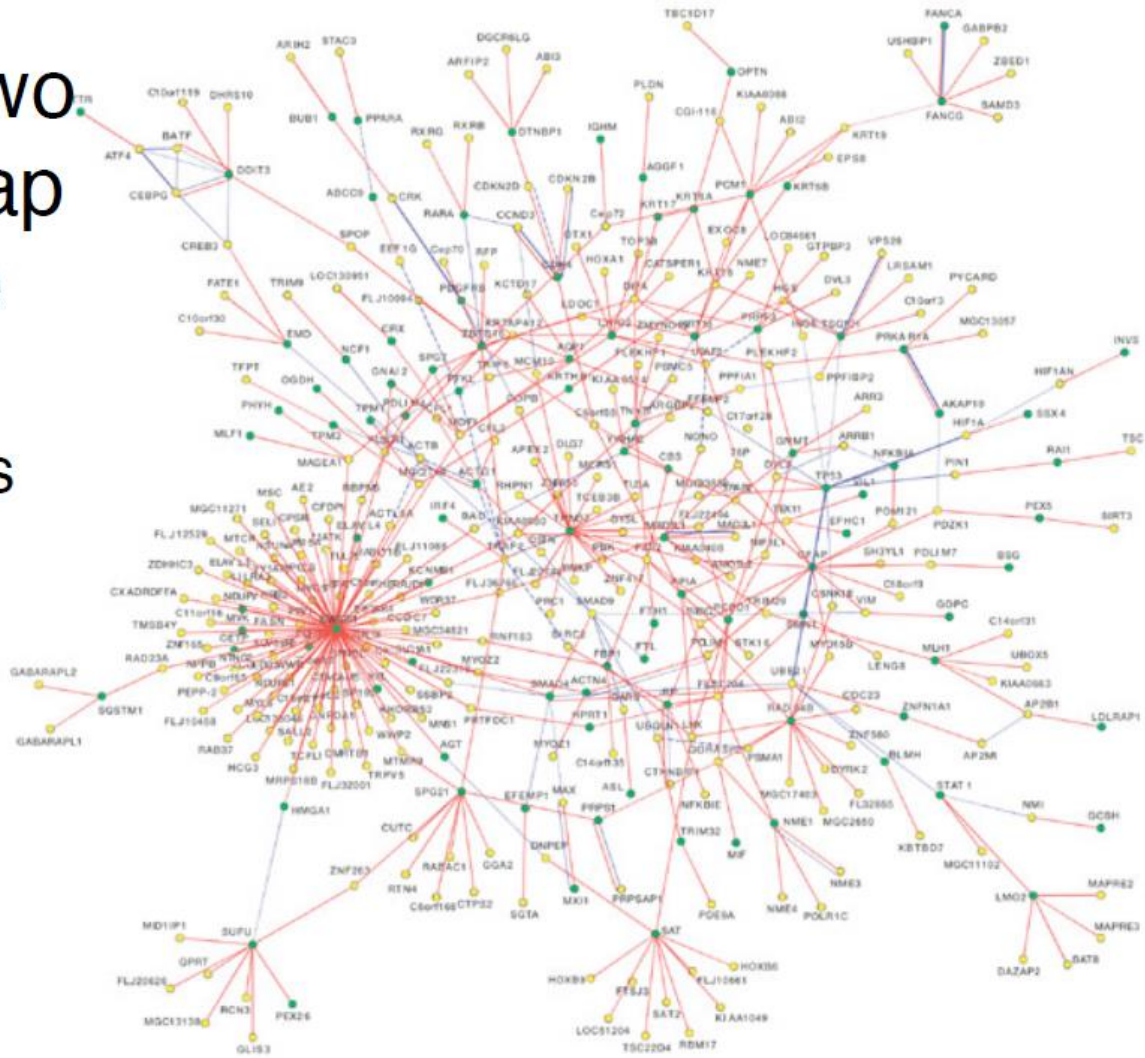
8,100 ORFs (~7,200 genes)

10,597 interactions



Rual et al. Nature 2005

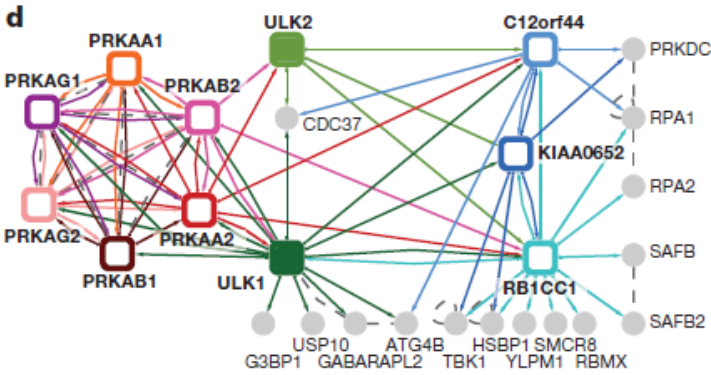
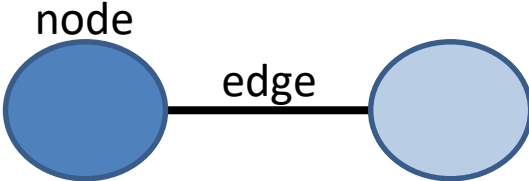
Human Two Hybrid Map Disease Genes (121 genes (green))



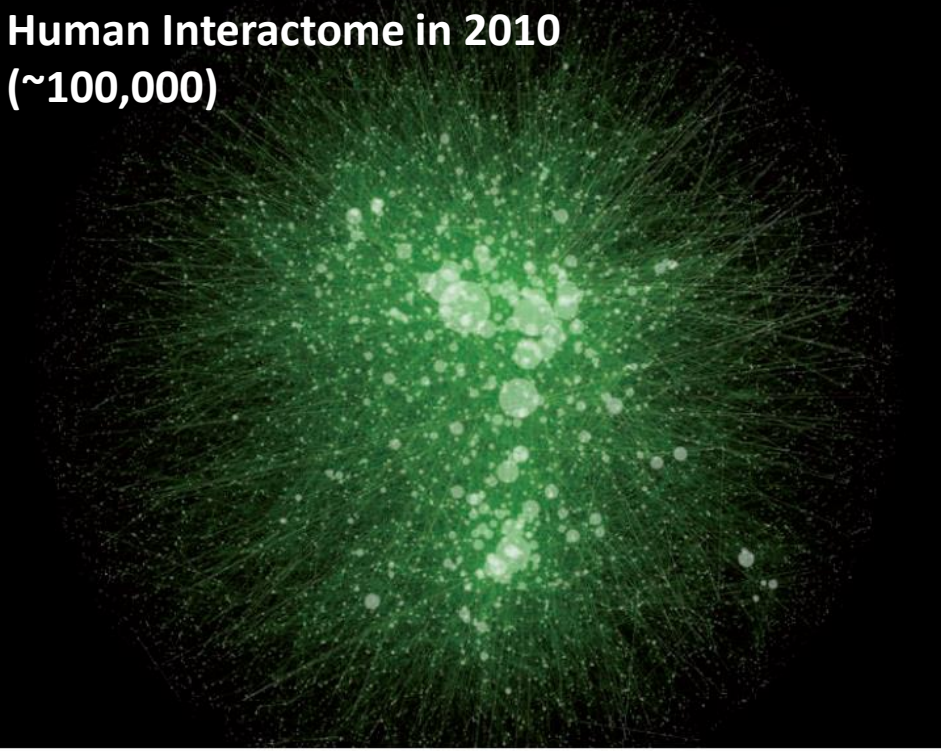
Rual et al. Nature 2005 Vol 437

Protein-Protein interaction maps:

Proteins are represented by nodes and interactions are represented by edges between nodes.

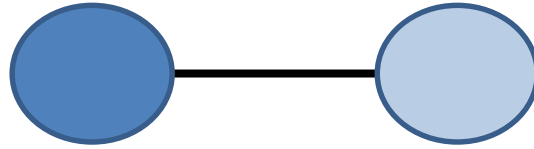


K. ONO/UC SAN DIEGO/CYTOSCAPE



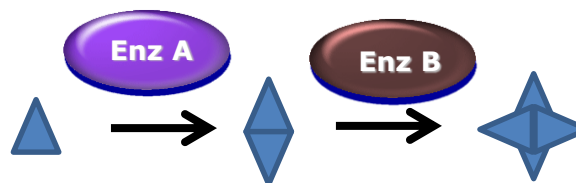
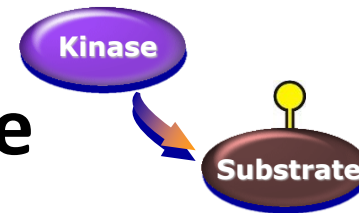
Bonetta, *Nature* 2010

Protein-Protein interactions:

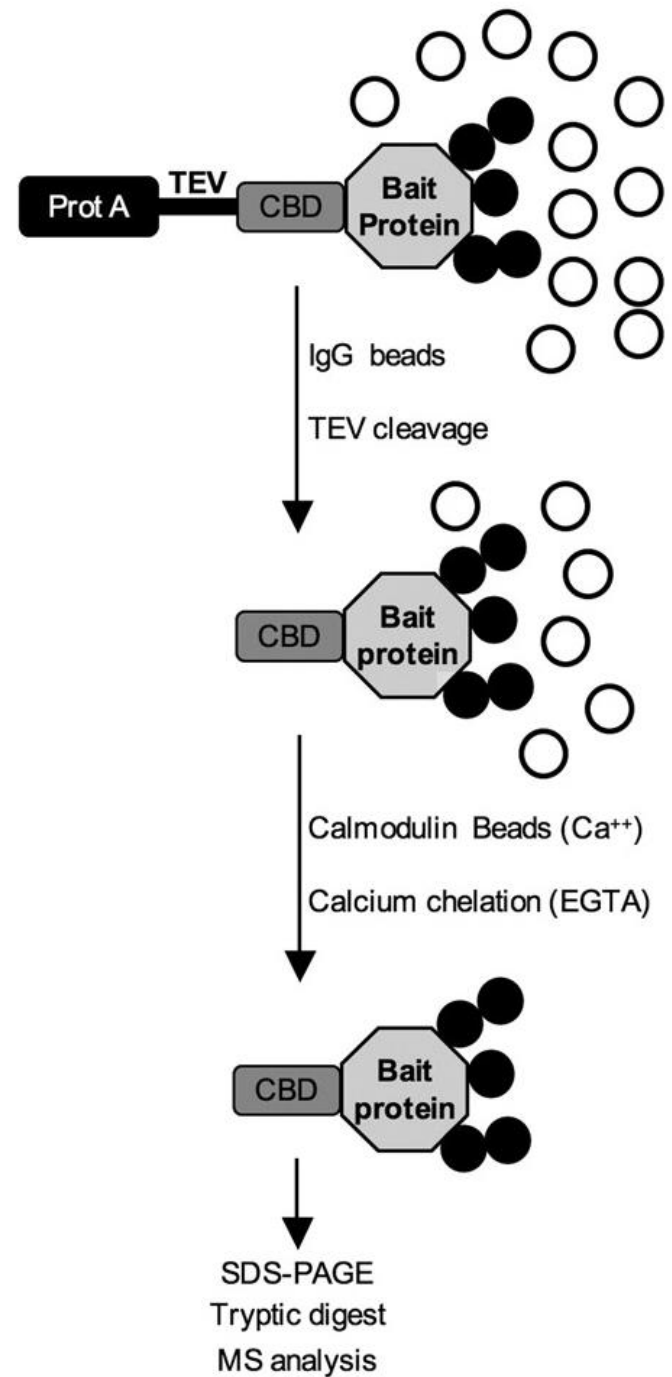
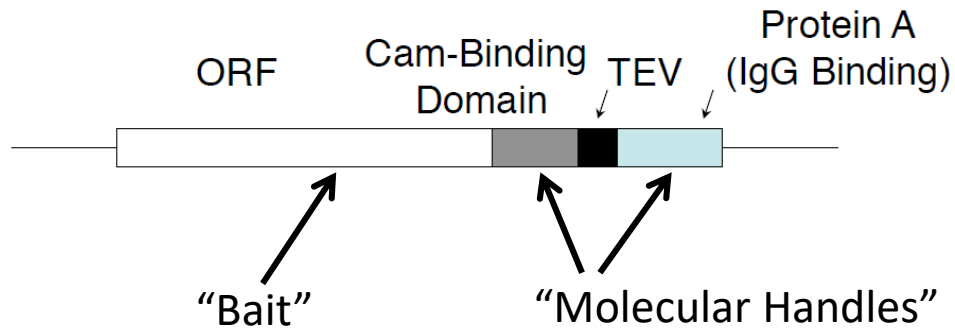


Some examples:

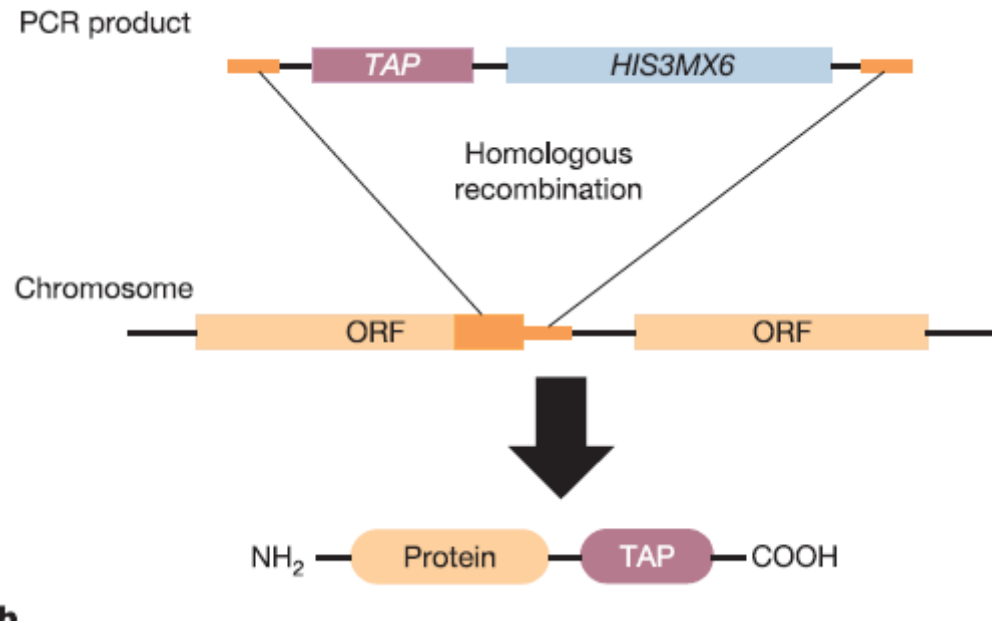
- Physical and direct
- Physical and indirect
- Multi-protein complexes
- Scaffolds
- Transient
- Kinase & substrate
- Metabolic



Tandem Affinity Purification (TAP) Tagging



Global TAP Tagging in yeast



2003

Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature*. & Huh, W. K. et al. Global analysis of protein localization in budding yeast. *Nature*.

➡ TAP-Tag and expression studies & GFP-Tag and localization studies

2002

Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*.

& Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*.

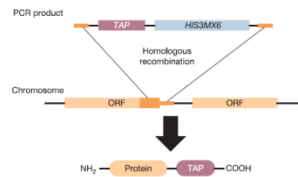
➔ Protein–protein interaction maps can be obtained by MS; the yeast cell is organized into protein complexes.

2006

Krogan NJ, *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*.

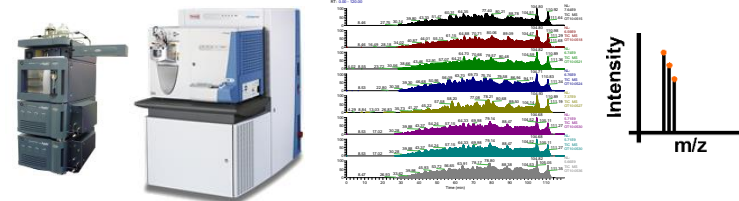
➔ TAP-Tag and Protein-Protein Interaction

Collection of tagged “bait”
expression strains



TAP bait + Interacting proteins

Multiple runs of “shotgun” MS
& SDS-PAGE with MS on individual proteins



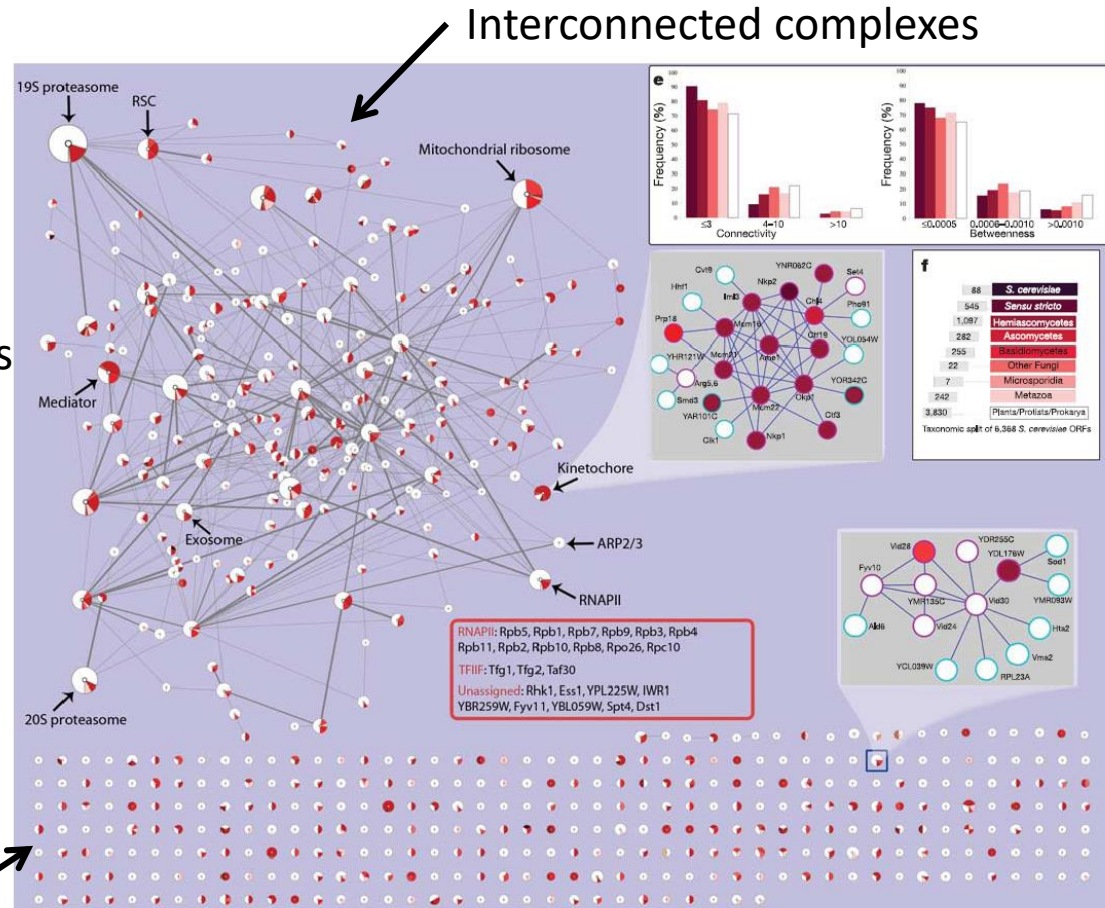
Krogan et al. observed 7,123 protein–protein interactions:

Important aspects:

- Tagged the native genes and did not overexpress the fusion proteins
- Could immediately validate partners (reciprocal purification in data set)
- Complementary MS techniques, deeper coverage of complexes
- Authors state, “...rigorous computational procedures to assign confidence values to our predictions...”

Cellular proteins are organized into complexes

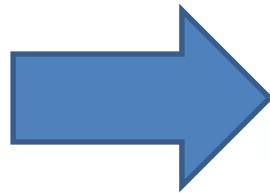
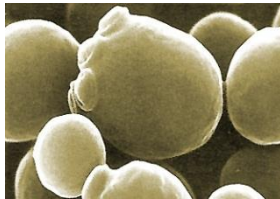
- 4,562 tagged proteins
- 2,357 successful purifications
- Identified 4,087 interacting proteins
~72 % proteome
- Majority of the yeast proteome is organized into complexes
- Many complexes are conserved in other species



Complexes with little or no interconnectivity

Krogan NJ, et al. *Nature*. 2006

How do we learn more about the organization of the human proteome?

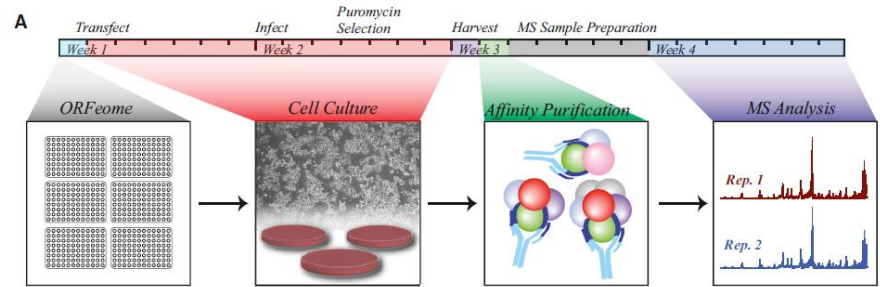


BioPlex (Biophysical Interactions of ORFeome-derived complexes)

~25% of human genes used as baits

5,891 IP-MS experiments

56,553 interactions from 10,961 proteins



The BioPlex Network: A Systematic Exploration of the Human Interactome

Edward L. Huttlin,¹ Lily Ting,¹ Raphael J. Bruckner,¹ Fana Gebreab,¹ Melanie P. Gygi,¹ John Szpyt,¹ Stanley Tam,¹ Gabriela Zarraga,¹ Greg Colby,¹ Kurt Baltier,¹ Rui Dong,² Virginia Guarani,¹ Laura Pontano Vaites,¹ Alban Ordureau,¹ Ramin Rad,¹ Brian K. Erickson,¹ Martin Wüthrich,¹ Joel Chick,¹ Bo Zhai,¹ Deepak Kolippakkam,¹ Julian Mintseris,¹ Robert A. Obar,^{1,3} Tim Harris,³ Spyros Artavanis-Tsakonas,^{1,3} Mathew E. Sowa,¹ Pietro De Camilli,² Joao A. Paulo,¹ J. Wade Harper,^{1,*} and Steven P. Gygi^{1,*}

BioPlex 1.0 Huttlin et al, *Cell*. 2015, PMID: 26186194

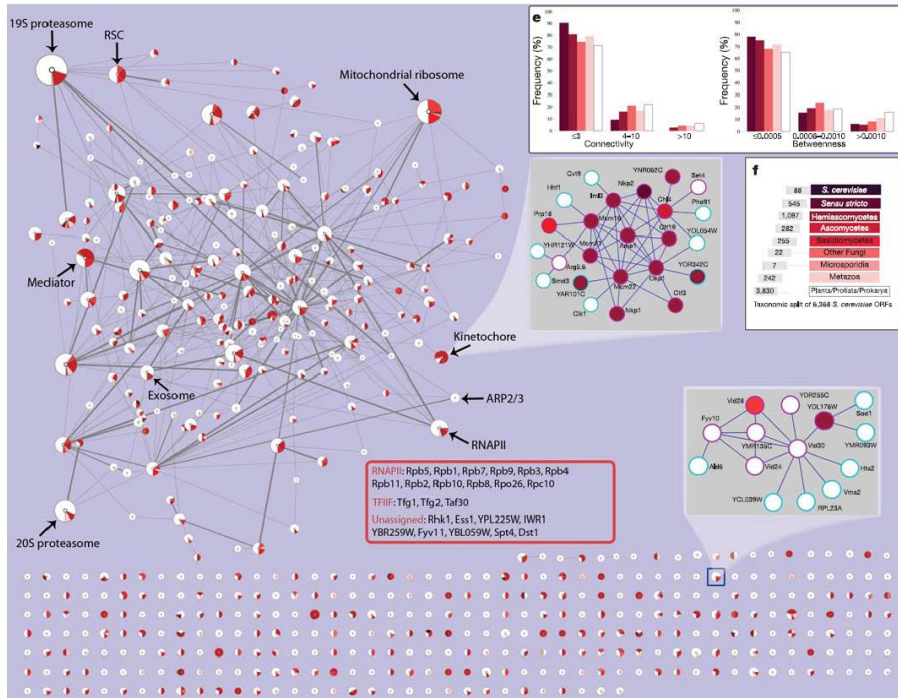
Architecture of the human interactome defines protein communities and disease networks

Edward L. Huttlin¹, Raphael J. Bruckner¹, Joao A. Paulo¹, Joe R. Cannon¹, Lily Ting¹, Kurt Baltier¹, Greg Colby¹, Fana Gebreab¹, Melanie P. Gygi¹, Hannah Parzen¹, John Szpyt¹, Stanley Tam¹, Gabriela Zarraga¹, Laura Pontano-Vaites¹, Sharan Swarup¹, Anne E. White¹, Devin K. Schweppe¹, Ramin Rad¹, Brian K. Erickson¹, Robert A. Obar^{1,2}, K. G. Guruharsha², Kejie Li², Spyros Artavanis-Tsakonas^{1,2}, Steven P. Gygi¹ & J. Wade Harper¹

BioPlex 2.0 Huttlin et al, *Nature*. 2017 PMID: 28514442

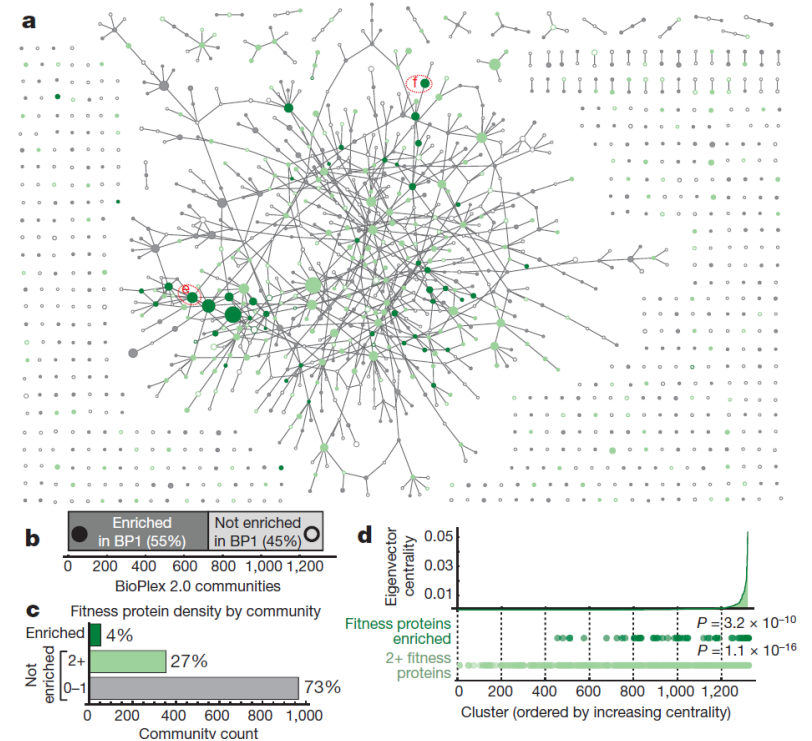
Cellular proteins are organized into complexes and this proteome organization is conserved

Yeast: Interaction Network of Complexes



Krogan NJ, et al. *Nature*. 2006 PMID: 16554755

Human: Protein Complex "Communities"

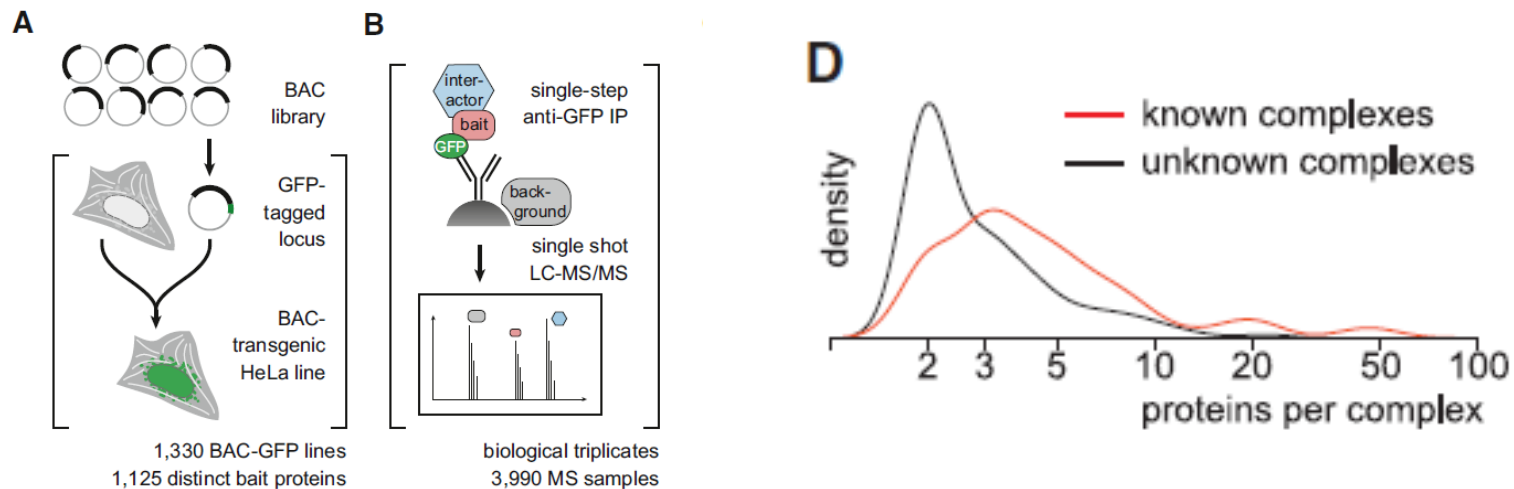


Huttlin et al, *Nature*. 2017 PMID: 28514442

A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances

Marco Y. Hein,^{1,6,8} Nina C. Hubner,^{1,6,9} Ina Poser,² Jürgen Cox,¹ Nagarjuna Nagaraj,¹ Yusuke Toyoda,^{2,10} Igor A. Gak,³ Ina Weisswange,^{4,5} Jörg Mansfeld,³ Frank Buchholz,^{2,4} Anthony A. Hyman,^{2,7,*} and Matthias Mann^{1,7,*}

- GFP-tagged proteins are expressed in mammalian cell lines from BAC transgenes with near-endogenous expression patterns
- Human interactome dataset connecting **5,400** proteins with **28,500** interactions

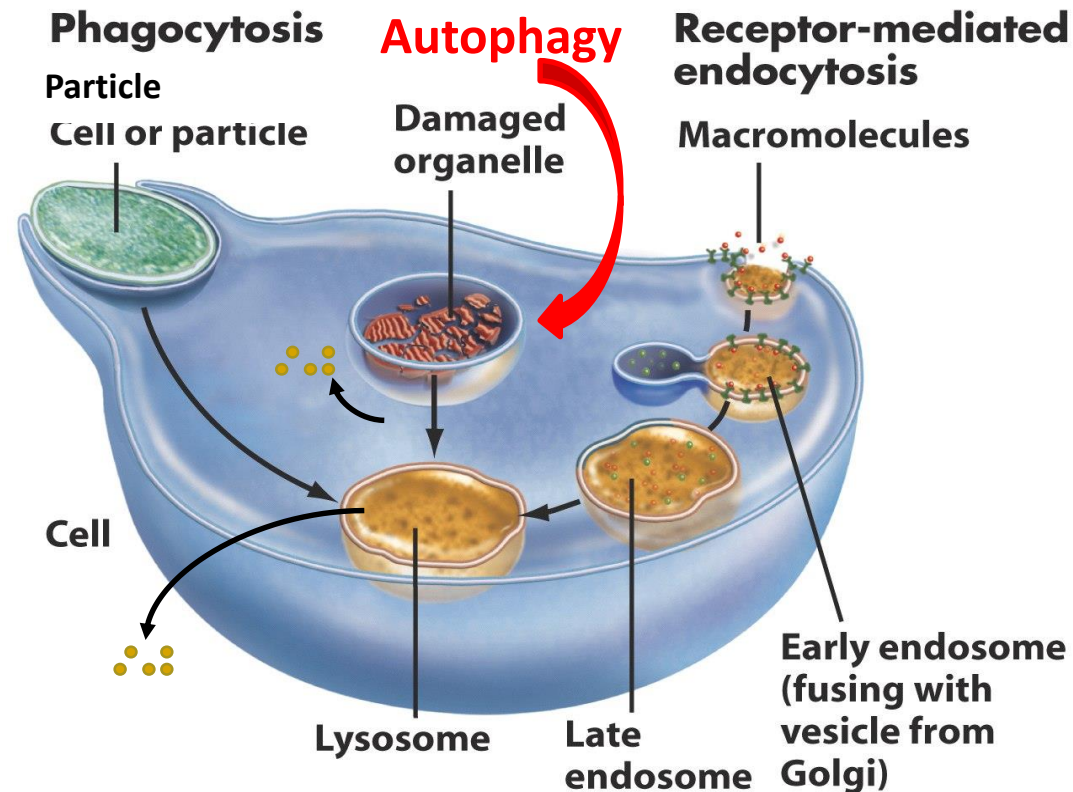


- Three quantitative dimensions measure specificities, stoichiometries, and abundances
- Stable complexes are rare but stand out by a signature of balanced stoichiometries
- Weak interactions dominate the network and have critical topological properties

ARTICLES

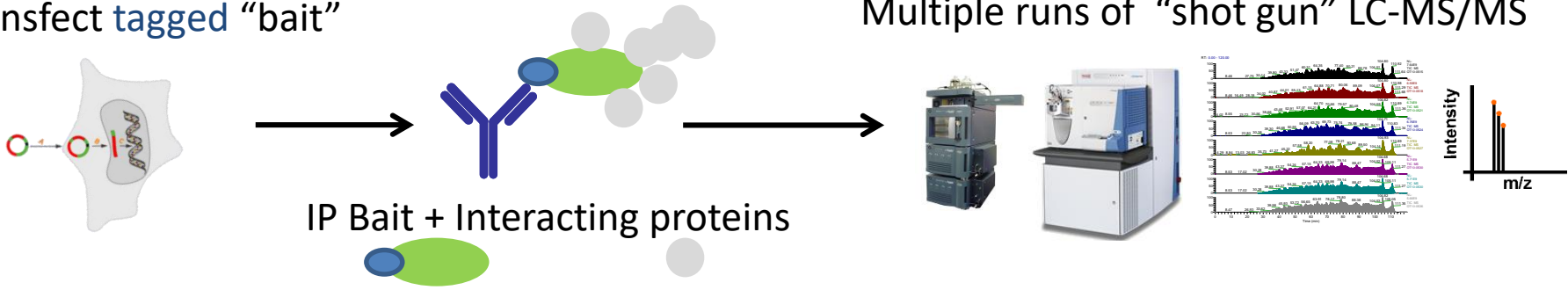
Network organization of the human autophagy system

Christian Behrends¹, Mathew E. Sowa¹, Steven P. Gygi² & J. Wade Harper¹



Transfect tagged "bait"

Multiple runs of "shotgun" LC-MS/MS



IP Bait + Interacting proteins

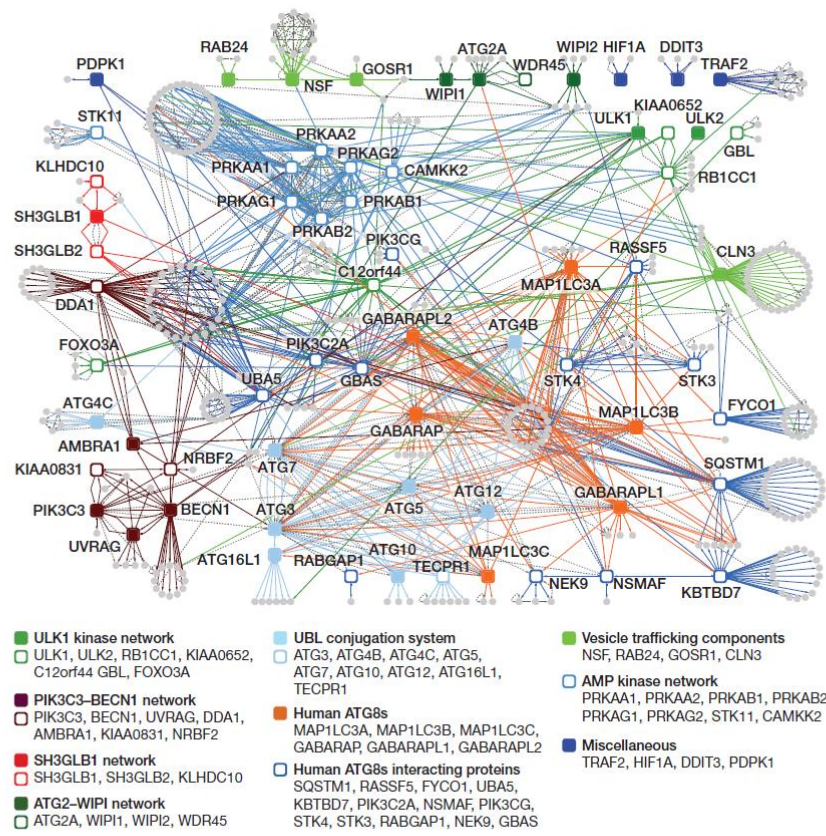
~65 bait proteins
LC-MS/MS identifies
2553 proteins

Data analysis to sort out real
interaction from background

Authors use CompPASS
to identify High-Confidence
Interacting Proteins (HCIP)

763 HCIPs identified that compose
The Autophagy Interaction Network

Autophagy Interaction Network



Behreands et al, Nature 2010

Figure 1 | Overview of the autophagy interaction network (AIN). HCIPs within the autophagy network are shown for 32 primary baits (filled squares) and 33 secondary baits (open squares). Subnetworks are colour-coded. Interacting proteins are indicated by grey circles.

The Hippo Signaling Pathway Interactome

Young Kwon,¹ Arunachalam Vinayagam,^{1*} Xiaoyun Sun,^{3*} Noah Dephoure,⁴ Steven P. Gygi,⁴ Pengyu Hong,³ Norbert Perrimon^{1,2,†}

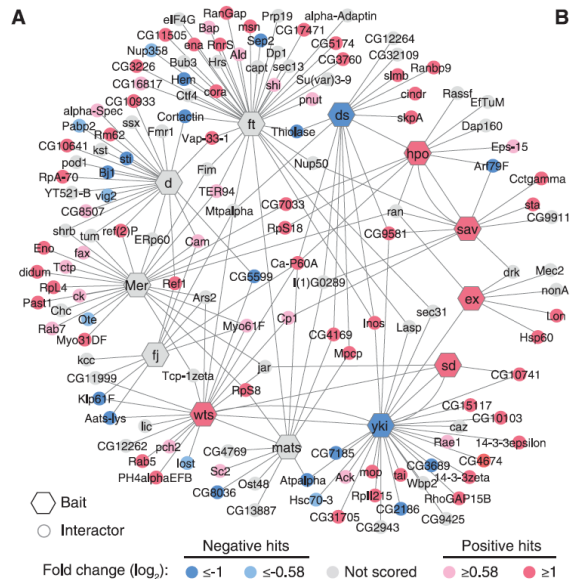
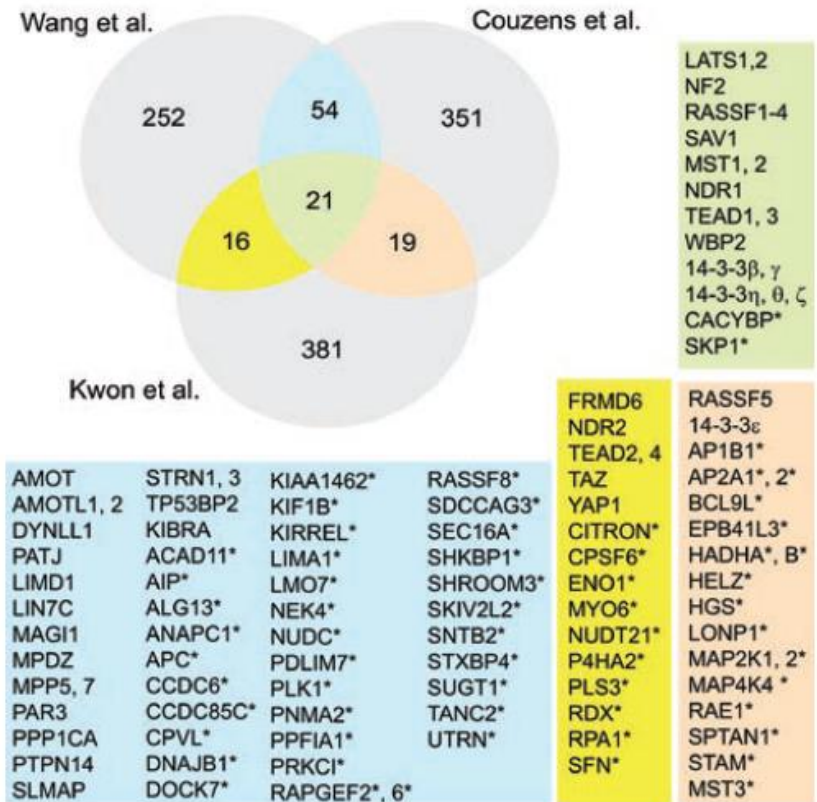


Fig. 2. Validation of Hippo-PPIN with functional RNAi screen and co-IP. (A) Distribution of Yki-reporter values for individual double-stranded RNAs (dsRNAs) in our focused RNAi screen. About 70% of genes are covered by two dsRNAs. **(B)** Recovery of Hippo pathway components from RNAi screen [fold-change (\log_2) cutoff ± 1]. **(C)** The positive

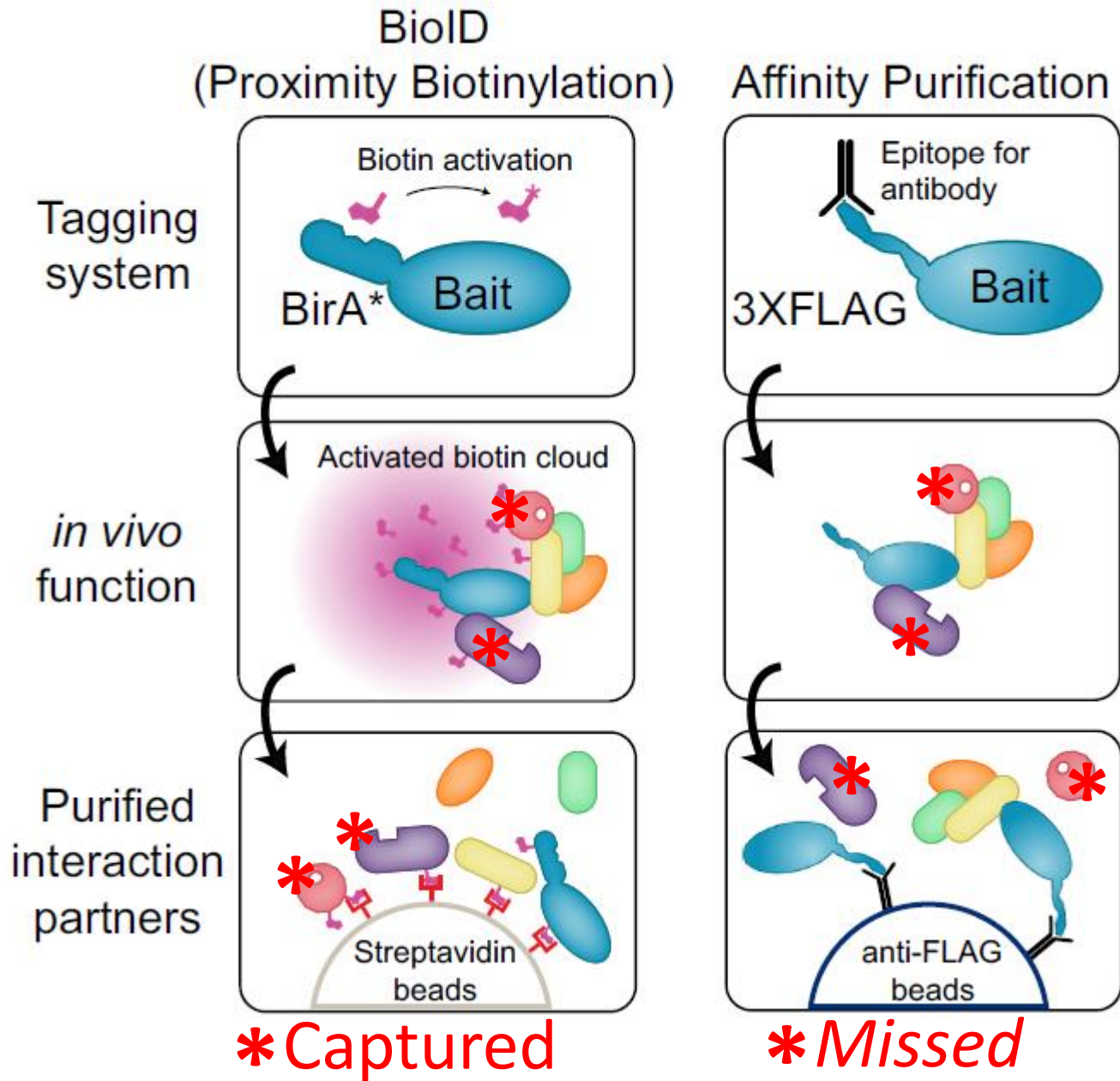


Cell Research (2014) 24:137-138.
 © 2014 IBCB, SIBS, CAS. All rights reserved 1001-0602/14 \$ 32.00
www.nature.com/cr

RESEARCH HIGHLIGHT

Discovering the Hippo pathway protein-protein interactome

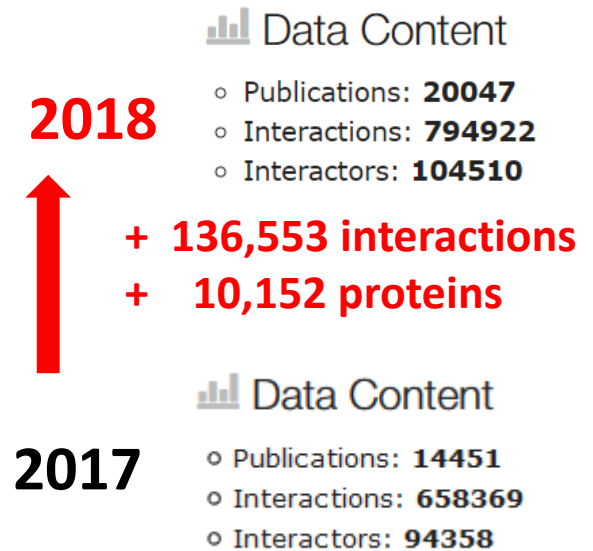
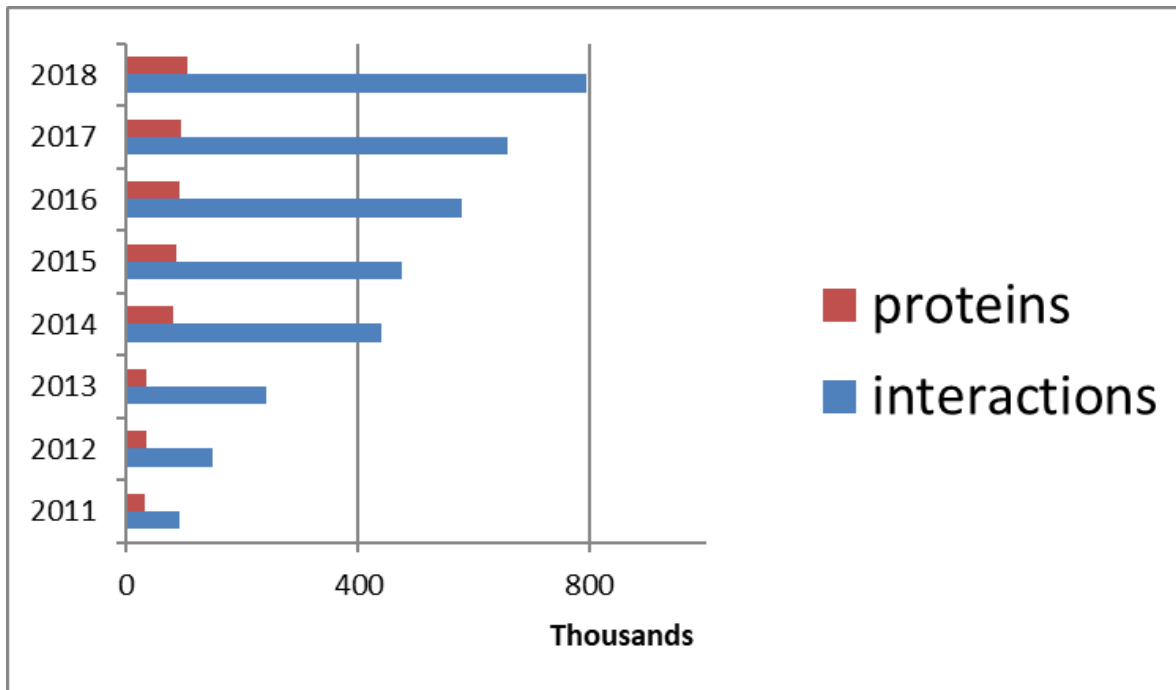
Cell Research (2014) 24:137-138. doi:10.1038/cr.2014.6; published online 14 January 2014



Protein-Protein Interaction Databases



<http://www.ebi.ac.uk/intact/>



Proteomics & Protein-Protein Interactions

Overview

- **Techniques & Technologies**
 - Mass Spectrometry
 - Protein-Protein Interactions
 - Quantitative Proteomics
- **Applications**
 - Representative Studies
- **Putting it all together....**
 - Databases & Pathways

Protein interaction networks:

Some of the many important aspects:

- Parts List
- Organization and assembly
- Biological function can be inferred



However:

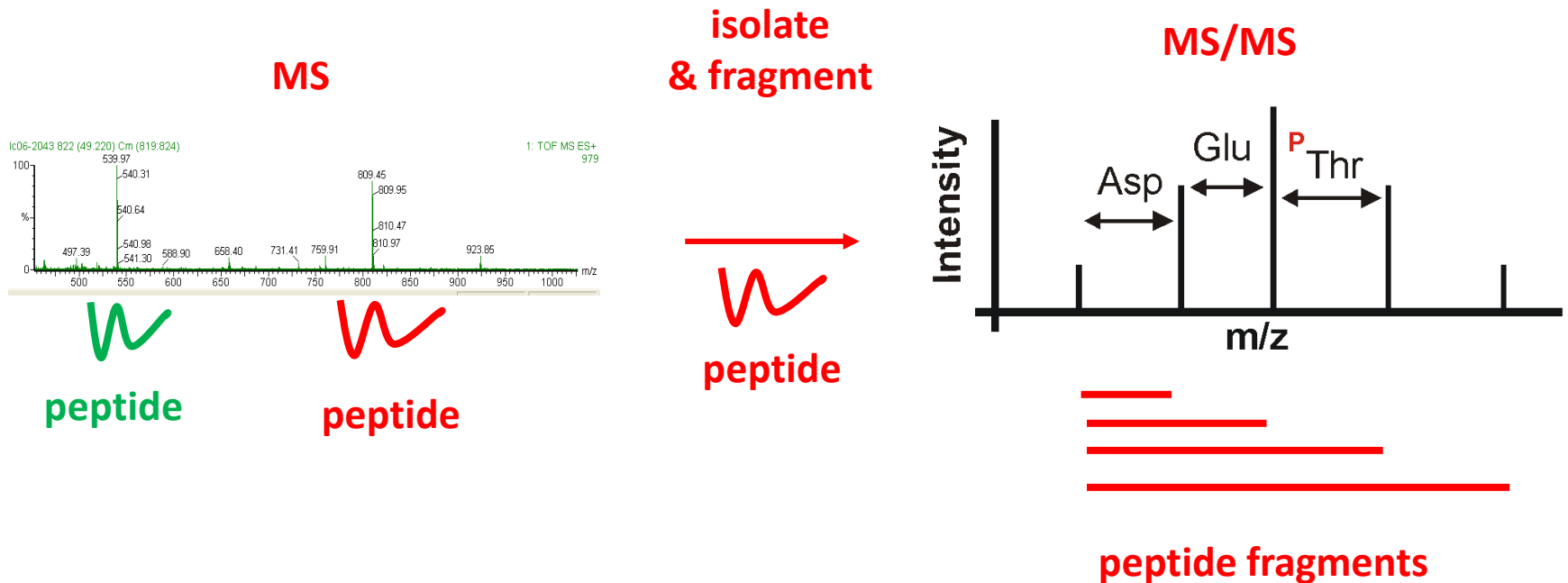
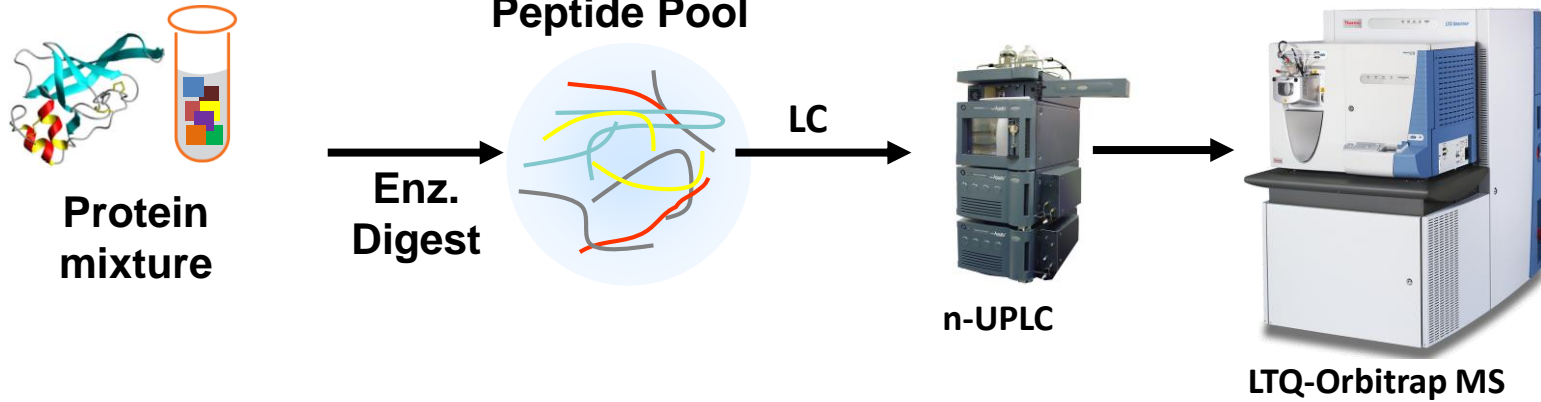
- Interaction data is largely static

Next Step:

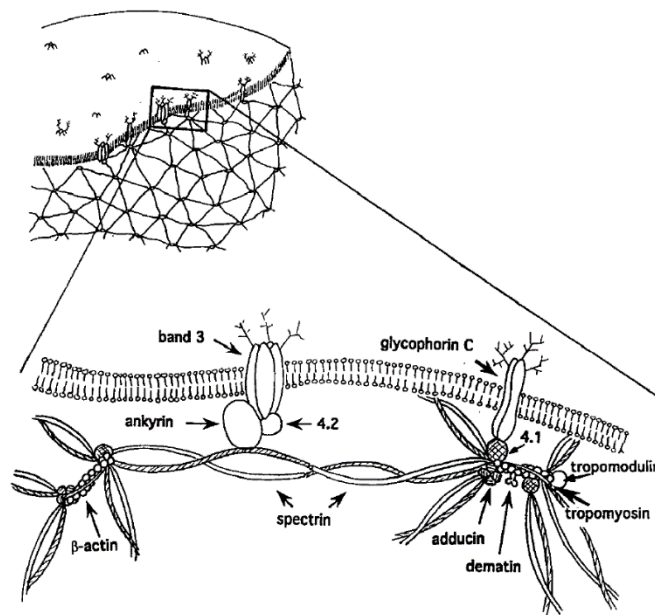
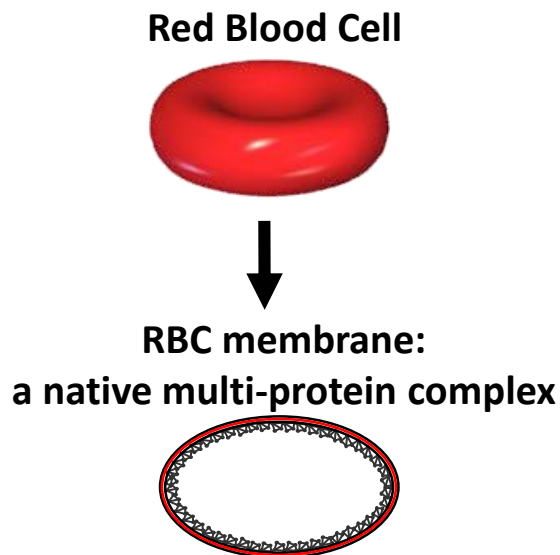
- **How do protein interaction networks change over time?**



Typical work flow for LC-MS "shotgun proteomics"

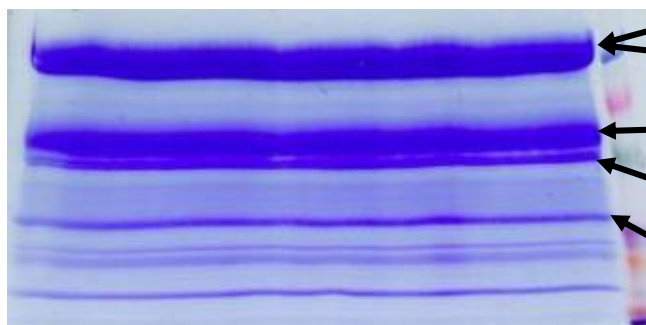


MS Data is not inherently quantitative, *but ...*



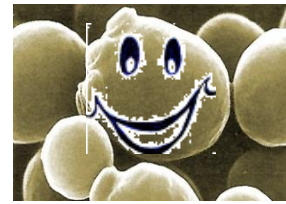
RBC membrane proteome
Coomassie Stained
SDS-PAGE (250 ug Protein)
~16 bands

RBC membrane proteome
Shotgun Proteomics
1ug Peptides (242 Proteins)



	# peptides (unique)	
Spectrin α	352 (291)	Spectrin alpha chain, erythrocyte OS=Homo sapiens GN=SPTA1 PE=1 SV=5
Spectrin β	291 (233)	Spectrin beta chain, erythrocyte OS=Homo sapiens GN=SPTB PE=1 SV=5
	172 (134)	Ankyrin-1 OS=Homo sapiens GN=ANK1 PE=1 SV=3
Band 3	57 (46)	Band 3 anion transport protein OS=Homo sapiens GN=SLC4A1 PE=1 SV=3
	52 (39)	Erythrocyte membrane protein band 4.2 OS=Homo sapiens GN=EPB42 PE=1 SV=3
Band 4.1	43 (34)	Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1
	30 (20)	Actin, alpha cardiac muscle 1 OS=Homo sapiens GN=ACTC1 PE=1 SV=1
	22 (9)	Beta-actin-like protein 2 OS=Homo sapiens GN=ACTBL2 PE=1 SV=2
β -actin	28 (6)	POTE ankyrin domain family member J OS=Homo sapiens GN=POTEJ PE=3 SV=1
	68 (49)	Protein 4.1 OS=Homo sapiens GN=EPB41 PE=1 SV=4

A tour of proteomics: Studies with the budding yeast *Saccharomyces cerevisiae*



2000 & 2001

Uetz et al, A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* .
& Ito et al, A comprehensive two-hybrid analysis to explore the yeast protein interactome . *PNAS*.

➔ **Large scale yeast two hybrid screens to map proteome wide interactions.**

2001

Washburn, et al. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.*

➔ **Established the 'shotgun' technology by showing that many proteins in a yeast-cell lysate could be identified in a single experiment.**

2002

Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*.

& Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* .

➔ **Protein-protein interaction maps can be obtained by MS; the yeast cell is organized into protein complexes.**

2003

Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature*. & Huh, W. K. et al. Global analysis of protein localization in budding yeast. *Nature*.

➔ **TAP-Tag and expression studies & GFP-Tag and localization studies**

2006

Krogan NJ, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*.

➔ **TAP-Tag and Protein-Protein Interaction**

2008

de Godoy LM, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*.

➔ **SILAC based quantitation of an entire proteome.**

2009

Picotti P, et al. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*.

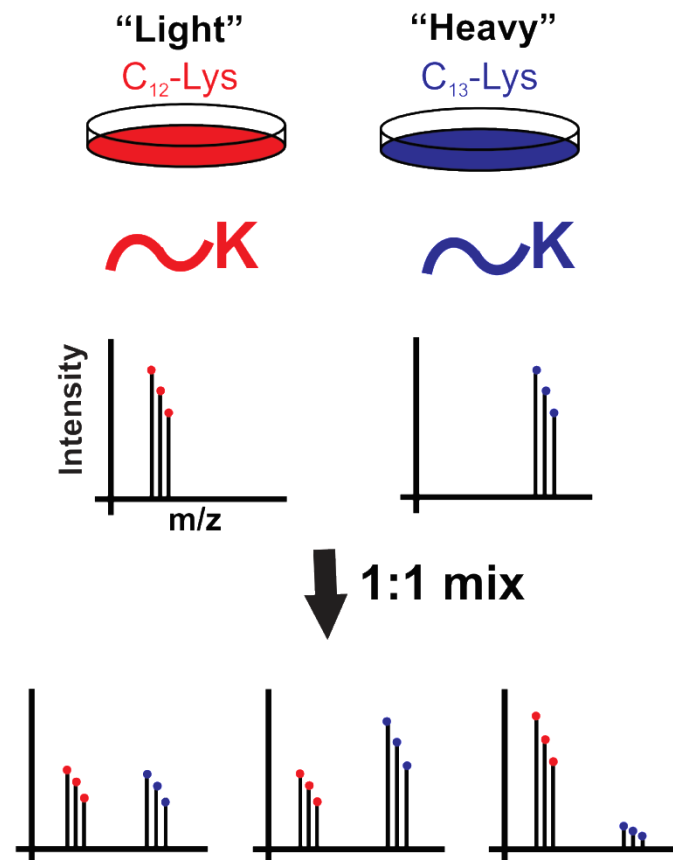
➔ **Towards proteome wide targeted proteomics.**

Quantitative Proteomics

S.I.L.A.C. - Stable isotope labeling with amino acids in cell culture

-Ong S.E. et al. *Molecular & Cell Proteomics* 2002

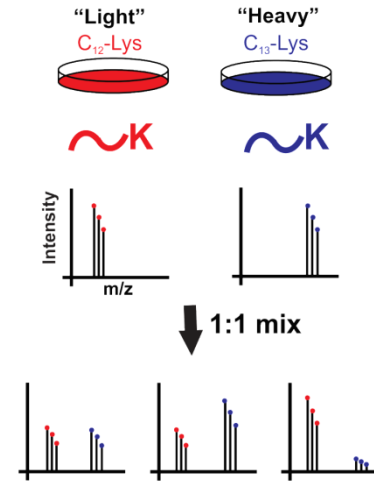
- Stable isotopes are *not radioactive*, and they occur naturally in nature. For example, 99% of all carbon in the world is carbon-12 (^{12}C) and 1% is carbon-13 (^{13}C).
- SILAC reagents have enriched stable isotopes that have been placed into compounds in abundances much greater than their natural abundance.
- We can obtain labeled compounds with ~95-99% ^{13}C .
- Because a mass spectrometer separates ions by mass, we use mass spectrometry to distinguish isotopes in compounds by their mass.
- Simultaneous comparison in the same MS run is key



2008

de Godoy LM, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*.

➔ SILAC based quantitation of an entire proteome.



S.I.L.A.C. - Stable isotope labeling with amino acids in cell culture

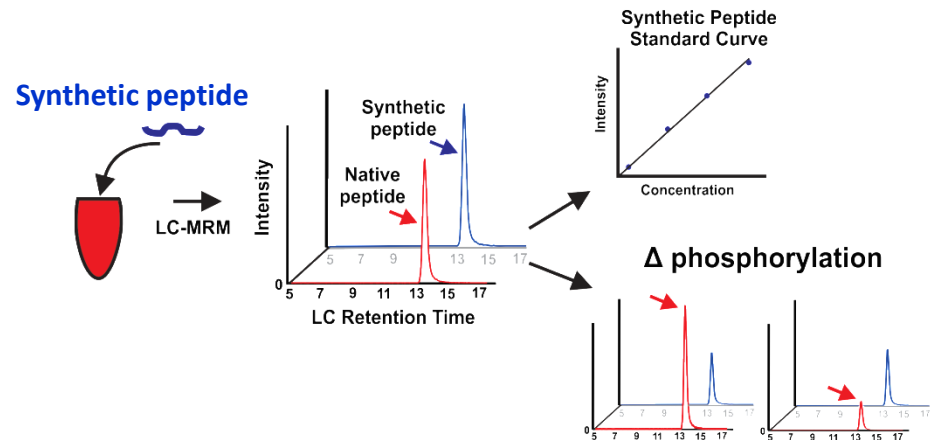
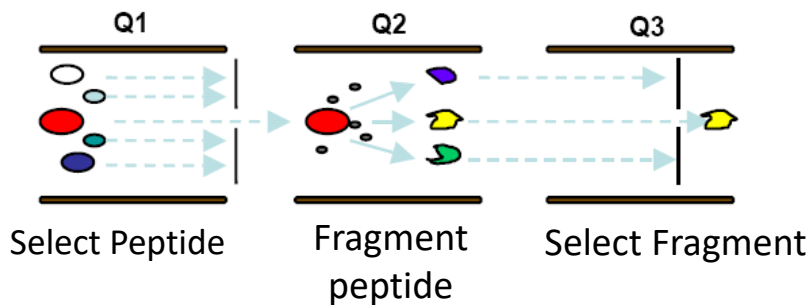
-Ong SE et al. *Molecular & Cell Proteomics* 2002.

2009

Picotti P, et al. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*.

➔ Towards proteome wide targeted proteomics.

Multiple Reaction Monitoring (MRM)



2008

de Godoy LM, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*. 30;455(7217):1251-4.

➡ SILAC based quantitation of an entire proteome.

Table 1 | Yeast ORFs identified by SILAC-based quantitative proteomics

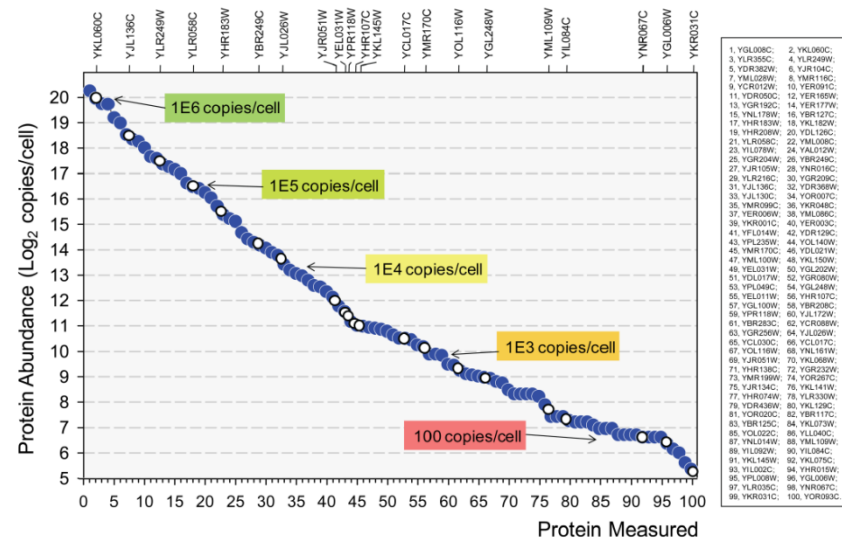
	Number of ORFs	TAP	GFP	nanoLC-MS
Total yeast ORFs	6,608	4,251	4,154	4,399
Characterized yeast ORFs	4,666	3,629	3,581	3,824
Uncharacterized yeast ORFs	1,128	581	539	572
Dubious yeast ORFs	814	26 (3%)	23 (3%)	3 (<1%)
Not present in ORF database		15	11	0

Comparative sequencing shows that 814 of the 6,608 yeast ORFs are never expressed (dubious ORFs, <http://www.yeastgenome.org>). Of these only six were identified in this experiment and three were validated by SILAC-assisted *de novo* sequencing of several peptides (Supplementary Table 5 and Supplementary Figs 2-4). Two of the three validated ones were reclassified as genuine yeast genes during writing of this manuscript (YGL041W-A and YPR170W-B). This leaves three potential false-positives (0.37% of 815) and suggests that our estimate of a false-positive identification rate of maximally 1% is conservative.

2009

Picotti P, et al. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*.

➡ Towards proteome wide targeted proteomics.



2008

de Godoy LM, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*.

➔ SILAC based quantitation of an entire proteome.

Pheromone signaling is required for mating of haploid cells and is absent from diploid cells.

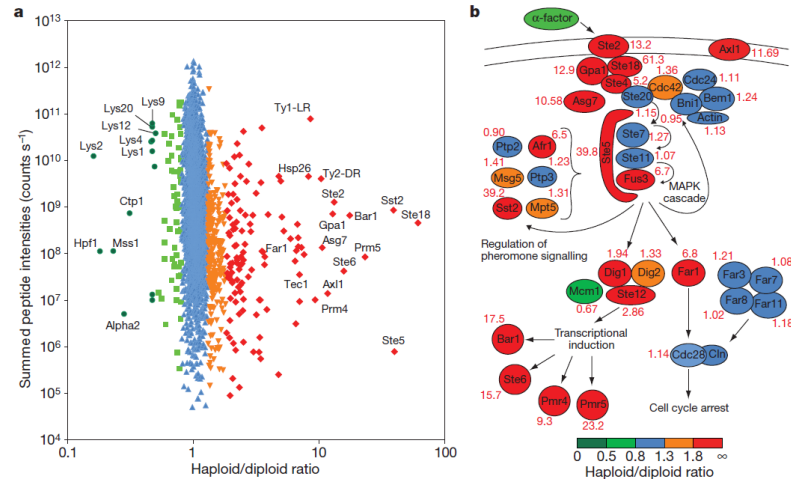


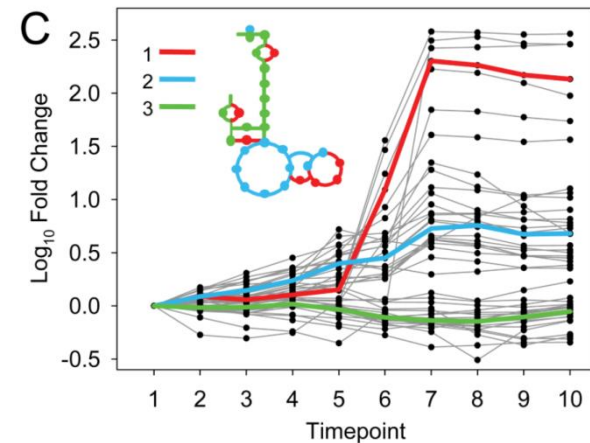
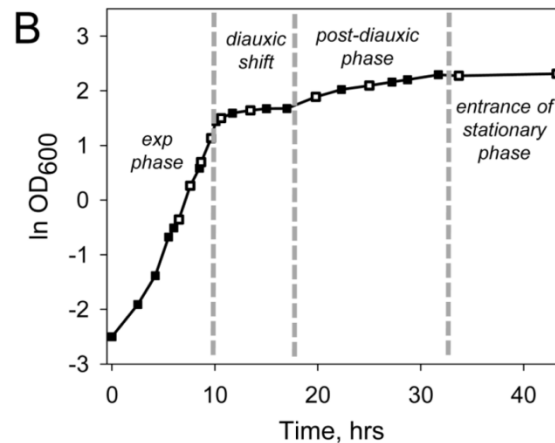
Figure 3 | Quantitative differences between the haploid and diploid yeast proteome. **a**, Overall fold change for the yeast proteome. **b**, Members of the yeast pheromone response are colour-coded according to fold change. The diploid to haploid ratio as determined by SILAC is indicated for each protein. Figure is adapted from ref. 13.

2009

Picotti P, et al. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*.

➔ Towards proteome wide targeted proteomics.

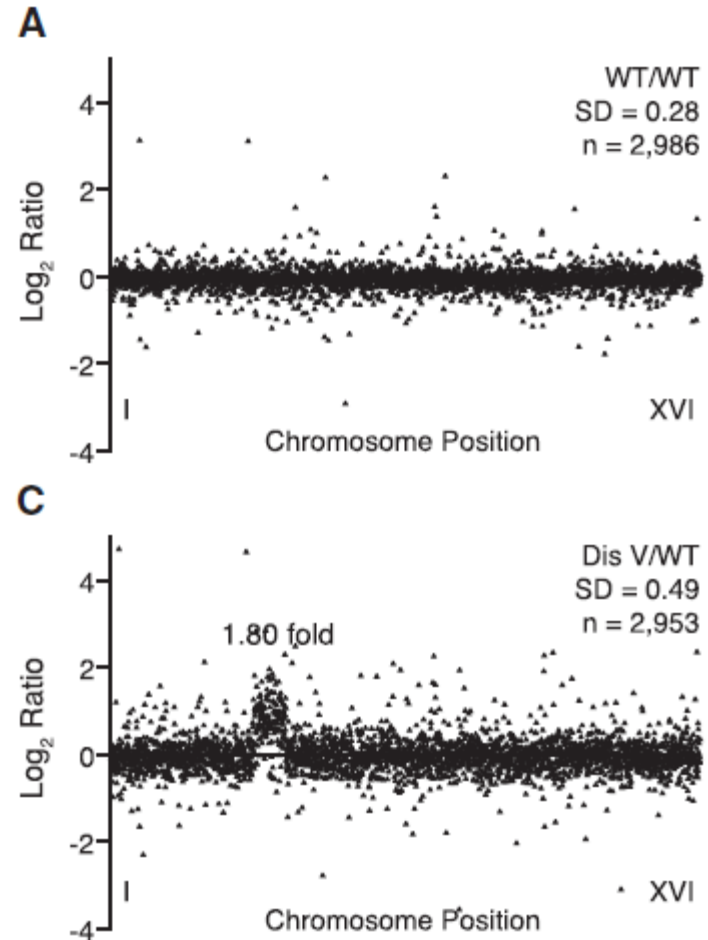
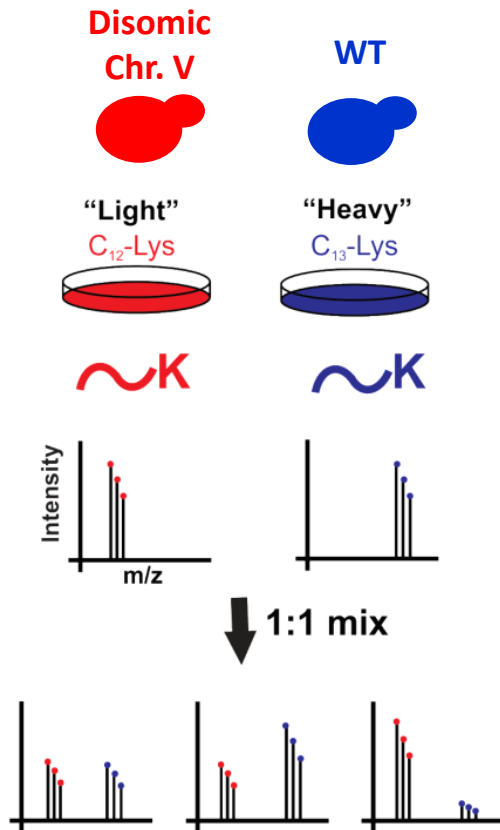
Network expression dynamics



Identification of Aneuploidy-Tolerating Mutations

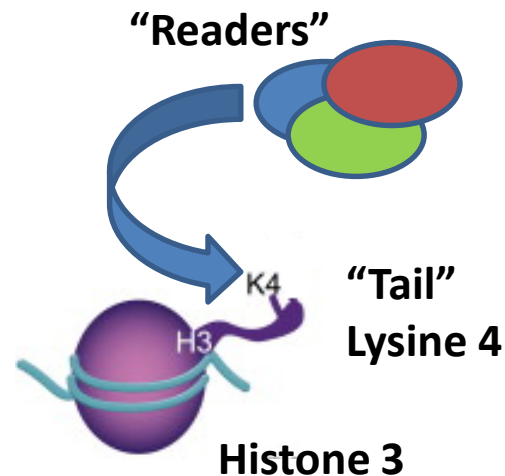
Cell 143, 71–83, October 1, 2010

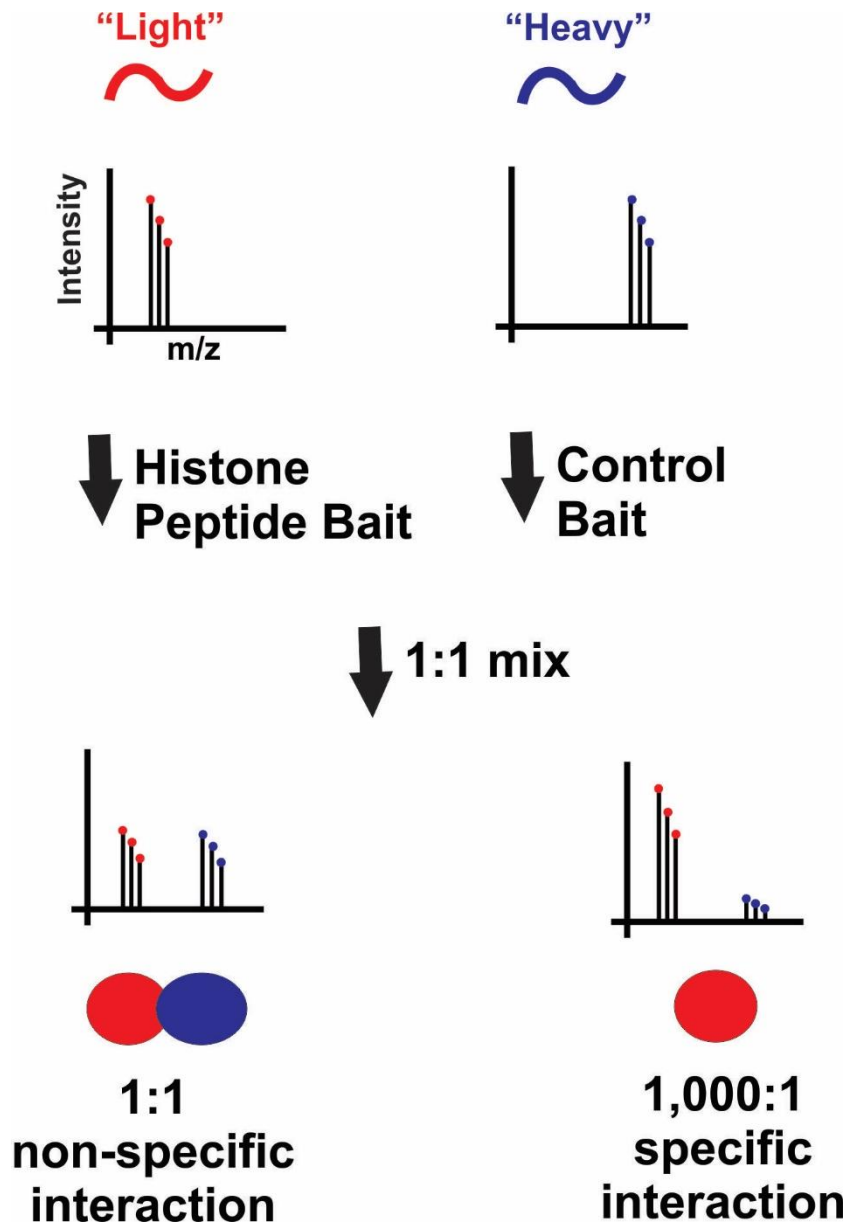
Eduardo M. Torres,^{1,2} Noah Dephoure,³ Amudha Panneerselvam,¹ Cheryl M. Tucker,⁴ Charles A. Whittaker,¹ Steven P. Gygi,³ Maitreya J. Dunham,⁵ and Angelika Amon^{1,2,*}



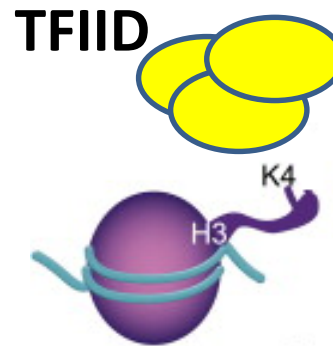
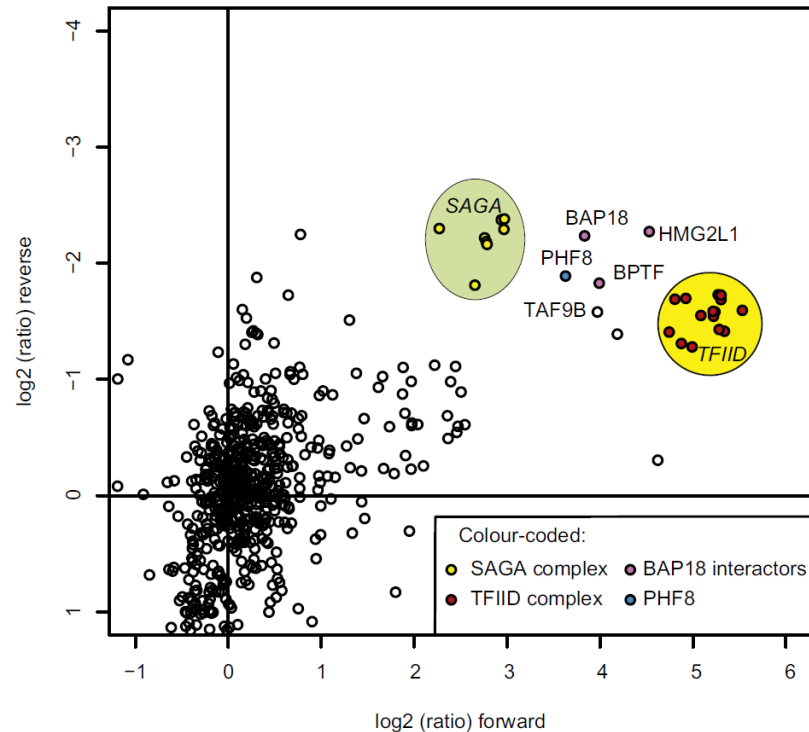
Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers

Michiel Vermeulen,^{1,6,7,*} H. Christian Eberl,^{1,6} Filomena Matarese,^{2,6} Hendrik Marks,² Sergei Denissov,² Falk Butter,¹ Kenneth K. Lee,³ Jesper V. Olsen,^{1,5} Anthony A. Hyman,⁴ Henk G. Stunnenberg,^{2,*} and Matthias Mann^{1,*}



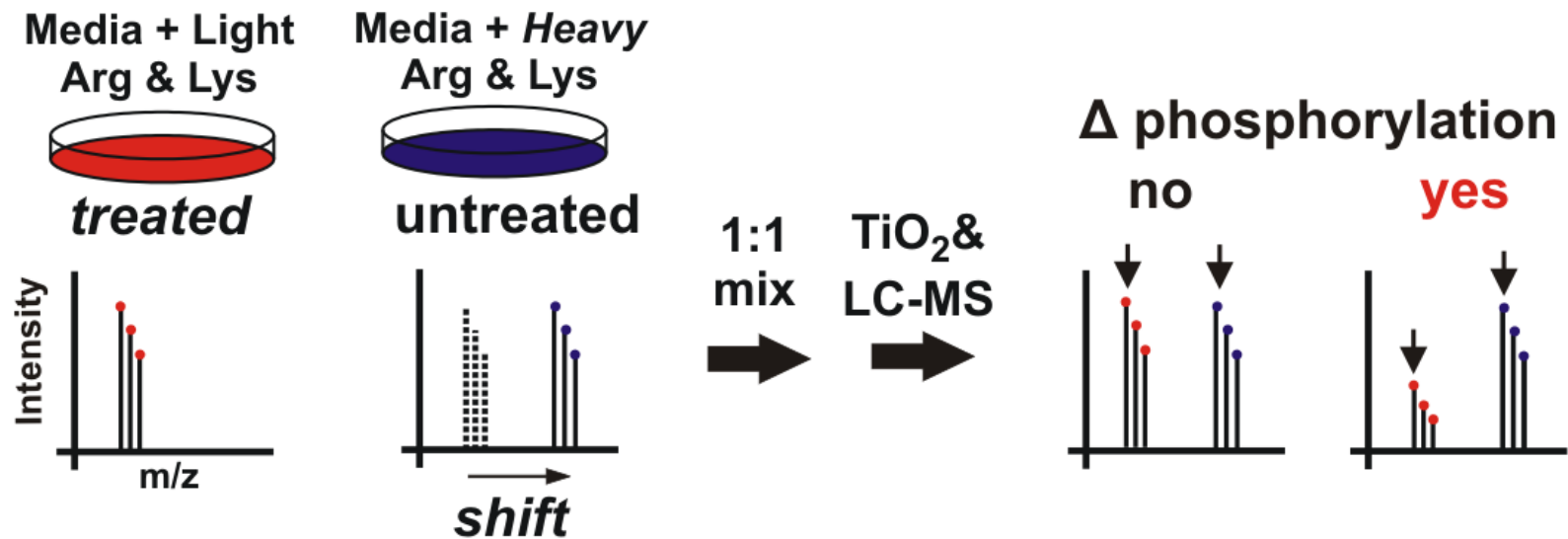


H3K4me3 interactors

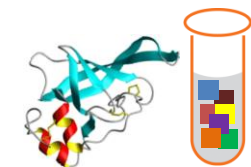


Active Genes →

A SILAC approach to study protein phosphorylation dynamics



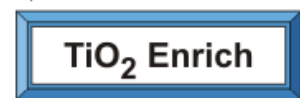
Major technological advances in mass spectrometers and phosphopeptide enrichment



Protein mixture



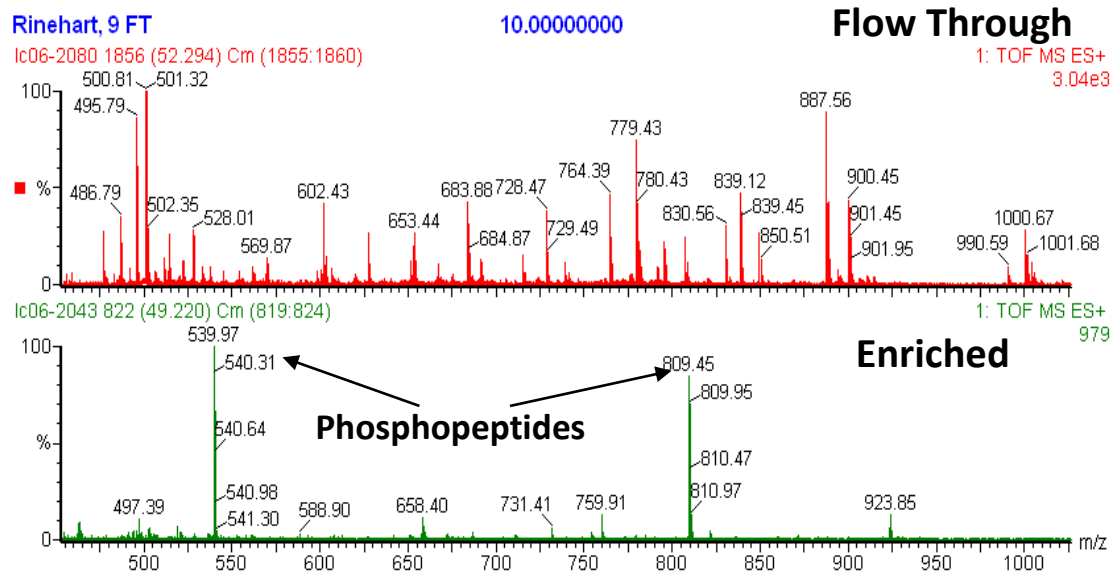
Peptides



Phosphopeptides



TiO₂ Enrichment

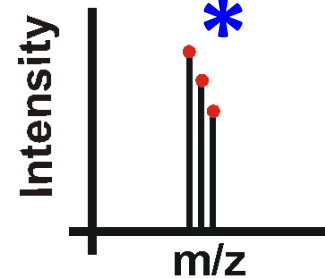


* Phosphopeptide signatures in MS

Phosphopeptide

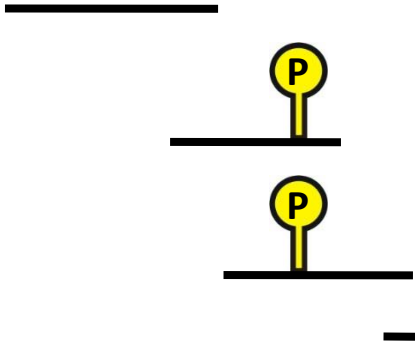


MS

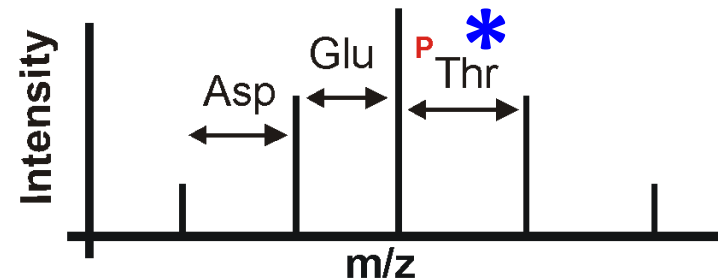


+80 Da
in precursor

isolate
& fragment



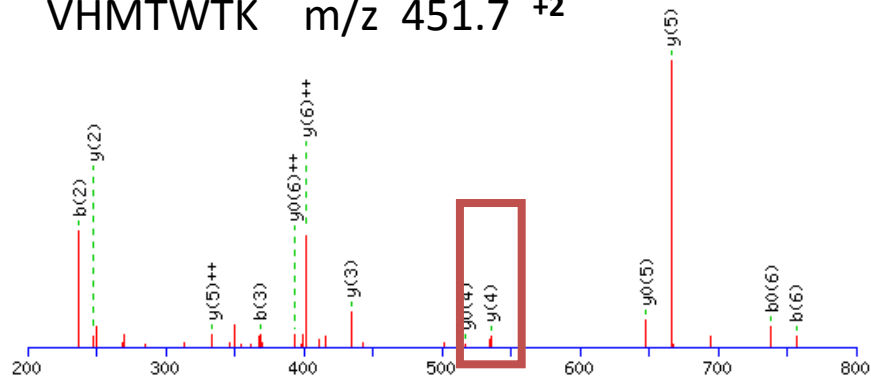
MS/MS



-98 Da loss of phosphoric acid H_3PO_4
during fragmentation

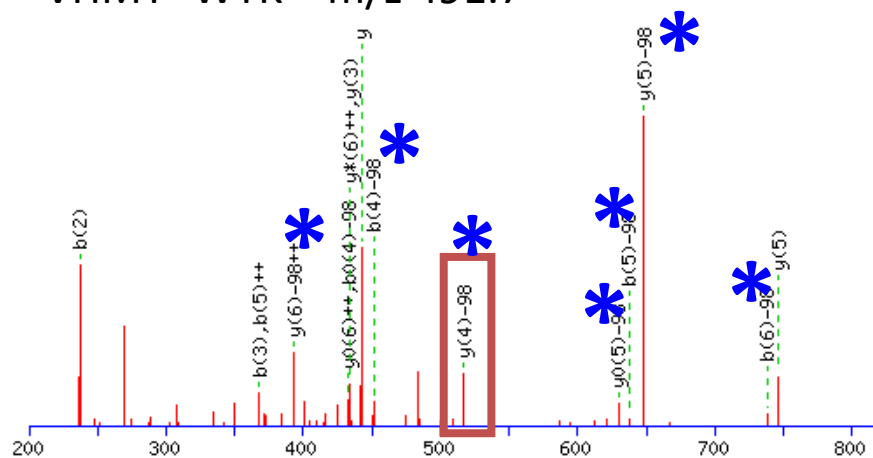
V
H 803.3869
M 666.3280
T 535.2875
W 434.2398
T 248.1605
K 147.1128

VHMTWTK m/z 451.7 +2



v
H 785.3763
M 648.3174
-18 **T 517.2769**
W 434.2398
T 248.1605
K 147.1128

VHMT^PWTK m/z 491.7 +2



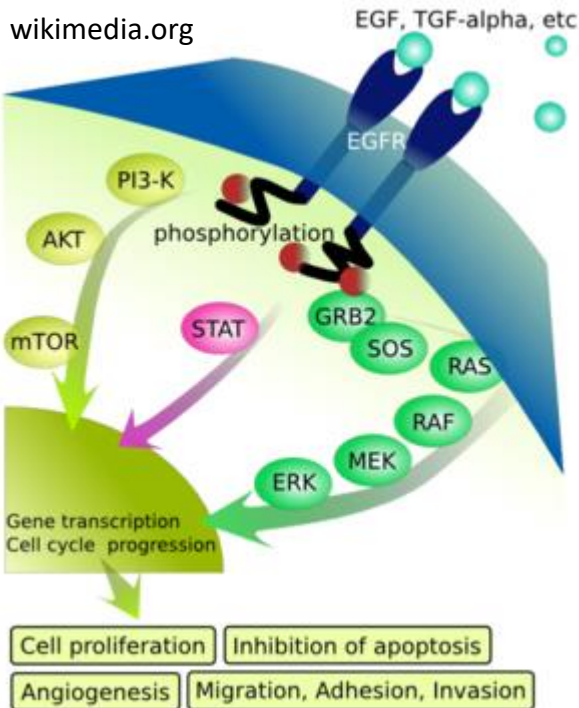
P +80 Da
in precursor

***** - H₃PO₄, 98 Da

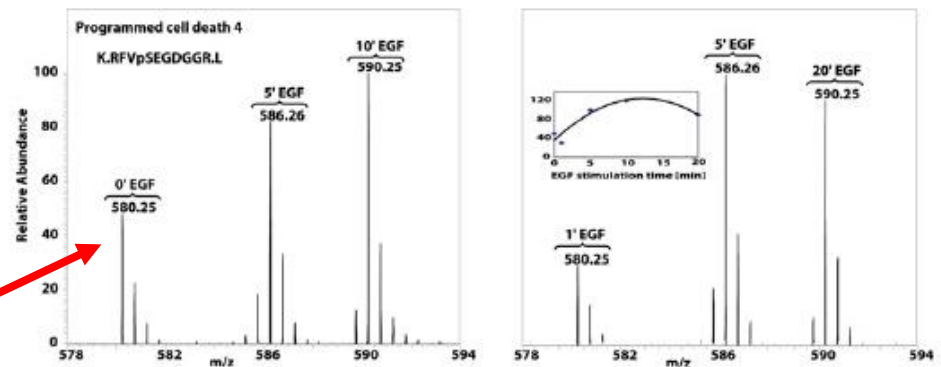
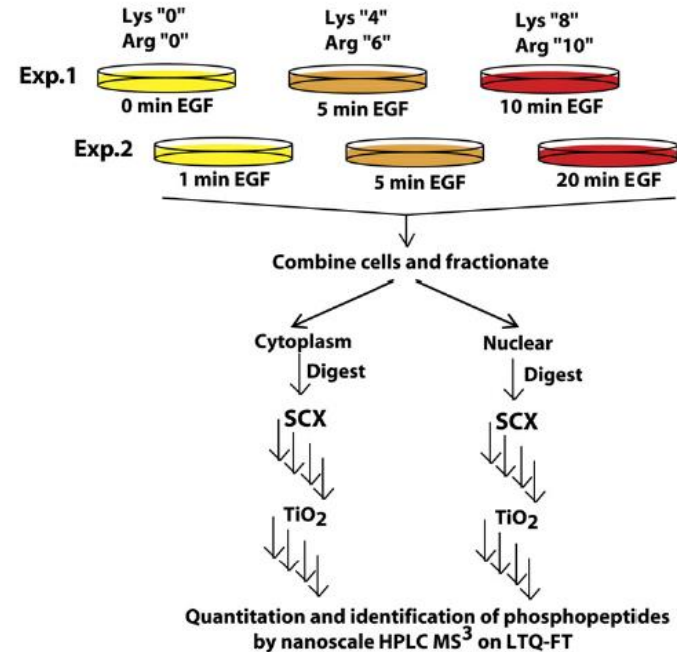
(Threonine changes to 2-aminodehydrobutyric acid, -18 Da)

Quantitative Proteomics Reveals Dynamics in Signaling Networks

Phosphorylation dynamics after EGF stimulation



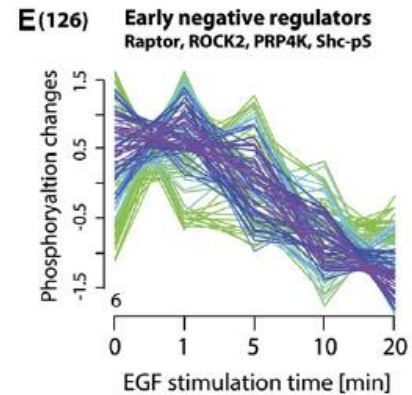
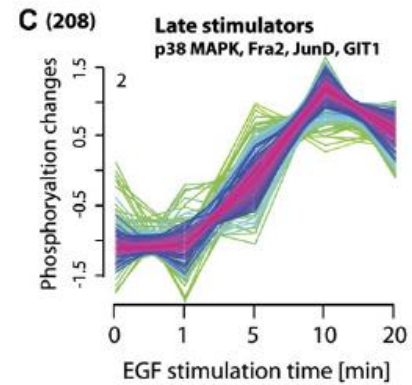
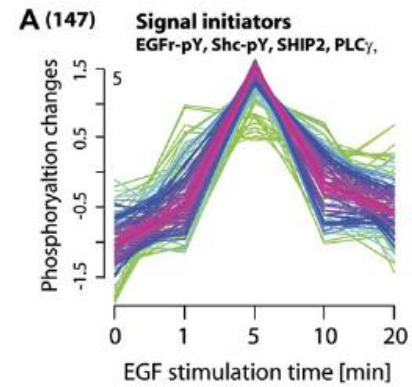
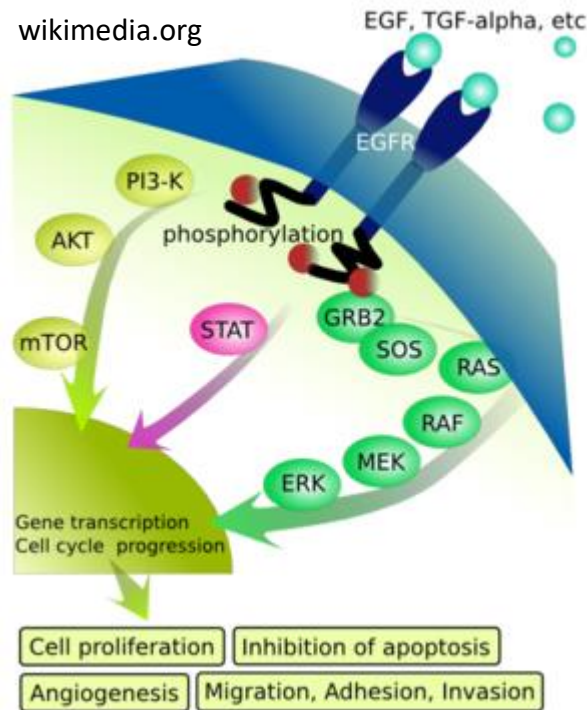
SILAC approach enables dynamic analysis



MS spectra triplets

Phosphorylation dynamics after EGF stimulation

wikimedia.org



Proteomics & Protein-Protein Interactions

Overview

- **Techniques & Technologies**
 - Mass Spectrometry
 - Protein-Protein Interactions
 - Quantitative Proteomics
- **Applications**
 - Representative Studies
- **Putting it all together....**
 - Databases & Pathways

DNA → RNA → PROTEIN



2001

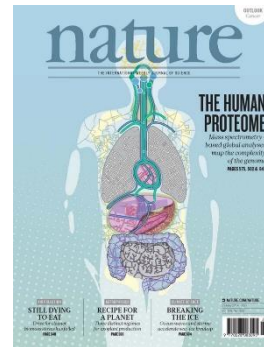
The Sequence of the Human Genome

J. Craig Venter,^{1*} Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹ Granger G. Sutton,¹ Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium¹



2014

ARTICLE

doi:10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm^{1,2*}, Judith Schlegel^{3*}, Hannes Hahne^{4*}, Amin Moghaddas Gholami^{4*}, Marcus Lieberenz², Mikhail M. Savitski⁵, Emanuel Ziegler², Lars Butzmann², Siegfried Gesualdo², Harald Marx¹, Toby Mathison¹, Simone Lemeier², Karsten Schmittaumann², Ulf Reimer², Holger Wenschuh², Martin Mollenhauer², Julia Siotta-Husperina², Joos Hendrik Boese², Marcus Bantscheff², Anja Gerstmaier², Franz Faerber² & Bernhard Kuster^{1,6}

ARTICLE

doi:10.1038/nature13302

A draft map of the human proteome

Min-Sik Kim^{1,2}, Sneha M. Pinto³, Derese Getnet^{1,4}, Raja Sekhar Nirujogi³, Srikanth S. Manda³, Raghobama Chaerkady^{1,2}, Anil K. Madugundu³, Dhanashree S. Kelkar³, Ruth Isserlin⁵, Shobhit Jain³, Joji K. Thomas³, Babylakshmi Muthusamy³, Pamela Leal-Rojas^{1,6}, Praveen Kumar³, Nandini A. Sahasrabudhe³, Lavanya Balakrishnan³, Jayshree Advani³, Bijesh George³, Santosh Renuse³, Lakshmi Dhevi N. Selvan³, Arun H. Patil³, Vishalakshi Nanjappa³, Aneesh Radhakrishnan³, Sumarjet Prasad³

The Sequence of the Human Genome. PMID: 11181995

Initial sequencing and analysis of the human genome. PMID: 11237011

A draft map of the human proteome. PMID: 24870542

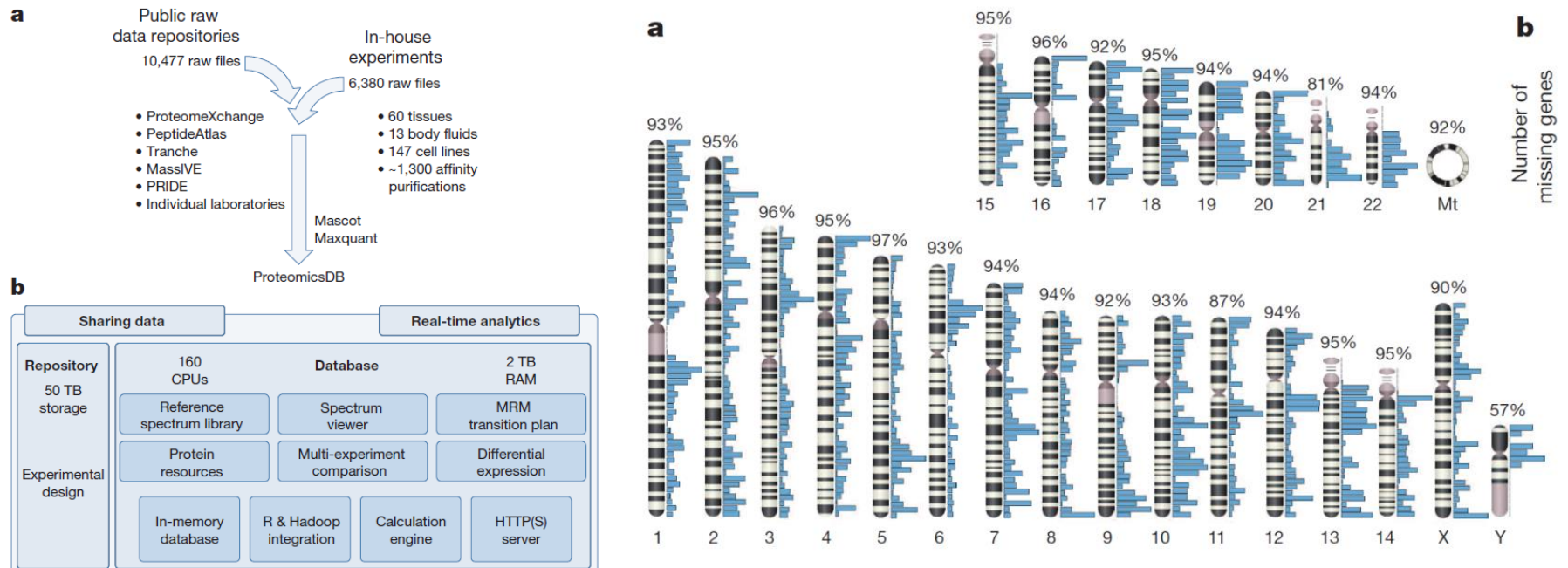
Mass-spectrometry-based draft of the human proteome. PMID: 24870543

Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm^{1,2*}, Judith Schlegl^{2*}, Hannes Hahne^{1*}, Amin Moghaddas Gholami^{1*}, Marcus Lieberenz², Mikhail M. Savitski³, Emanuel Ziegler², Lars Butzmann², Siegfried Gessulat², Harald Marx¹, Toby Mathieson³, Simone Lemeer¹, Karsten Schnatbaum⁴, Ulf Reimer², Holger Wenschuh⁴, Martin Mollenhauer⁵, Julia Slotta-Huspenina⁵, Joos-Hendrik Boese², Marcus Bantscheff³, Anja Gerstmair², Franz Paerber² & Bernhard Kuster^{1,6}

- Large Assembly of new and existing data:
- ProteomicsDB, database designed for the real-time analysis of big data

<https://www.proteomicsdb.org>



Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm^{1,2*}, Judith Schlegl^{2*}, Hannes Hahne^{1*}, Amin Moghaddas Gholami^{1*}, Marcus Lieberenz², Mikhail M. Savitski³, Emanuel Ziegler², Lars Butzmann², Siegfried Gessulat², Harald Marx¹, Toby Mathieson³, Simone Lemeer¹, Karsten Schnatbaum⁴, Ulf Reimer⁴, Holger Wenschuh⁴, Martin Mollenhauer⁵, Julia Slotta-Huspenina⁵, Joos-Hendrik Boese², Marcus Bantscheff³, Anja Gerstmaier², Franz Paerber² & Bernhard Kuster^{1,6}

- Large Assembly of new and existing data:
- ProteomicsDB, database designed for the real-time analysis of big data

<https://www.proteomicsdb.org>



Wilhelm *et al.* carried out 6,380 LC-MS experiments (or runs):

How long would it take to get the same data?

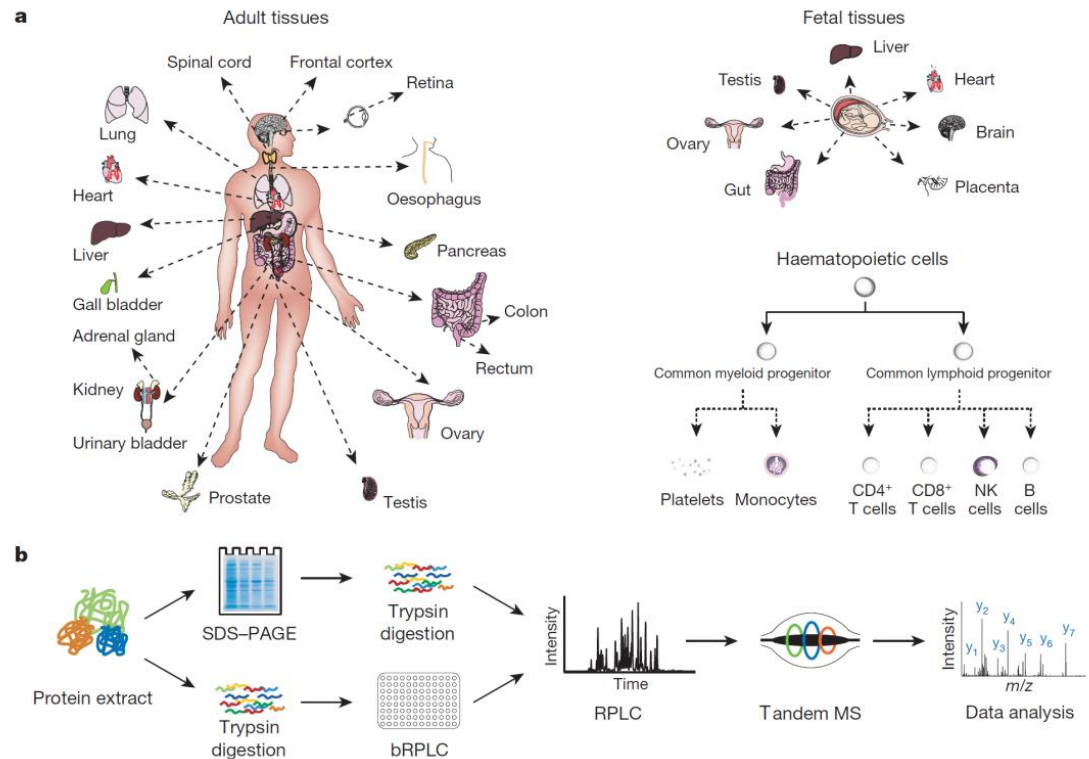
In 2001? ~61 years

In 2014? ~265 Days

A draft map of the human proteome

Min-Sik Kim^{1,2}, Sneha M. Pinto³, Derese Getnet^{1,4}, Raja Sekhar Nirujogi³, Srikanth S. Manda³, Raghothama Chaerkady^{1,2}, Anil K. Madugundu³, Dhanashree S. Kelkar³, Ruth Isserlin⁵, Shobhit Jain⁵, Joji K. Thomas³, Babylakshmi Muthusamy³, Pamela Leal-Rojas^{1,6}, Praveen Kumar³, Nandini A. Sahasrabudhe³, Lavanya Balakrishnan³, Jayshree Advani³, Bijesh George³, Santosh Renuse³, Lakshmi Dhevi N. Selvan³, Arun H. Patil³, Vishalakshi Nanjappa³, Aneesh Radhakrishnan³, Samarjeet Prasad¹,

- New, large collection of proteomics data
 - 30 histologically normal human samples
 - 17 adult tissues,
 - 7 fetal tissues
 - 6 purified primary haematopoietic cells
- 17,294 genes accounting for approximately 84% of the total annotated protein-coding genes in humans.



Proteomics Databases: Peptide depositories



PeptideAtlas

PEPTIDEATLAS HOME

Seattle Proteome Center

PeptideAtlas Builds – Bulk Downloads

<http://www.peptideatlas.org/builds/>

TaxID	Date	Number of Samples	Peptide Inclusion Cutoff	Number of Peptide-Spectrum Matches (PSMs)	Number of Distinct Peptides	Reference Database	Peptide Sequences	Peptide CDS Coordinates	Peptide CDS and Chromosomal Coordinates	Database Tables
9606	Mar 2015	1011	PSM FDR = 0.0002	133,638,335	1,025,698	Ensembl v78+UPSP+Trembl201412+14IPI 3.87+cRAP+nextprotSNP	APD_Hs_all.fasta	prot_map	chrom_map	MYSQL.XML

Protein Identification Terminology used in PeptideAtlas

http://www.peptideatlas.org/docs/protein_ident_terms.php

- Each PeptideAtlas build is associated with a reference database usually a combination of several protein sequence databases (Swiss-Prot, IPI, Ensembl ...)
- From the reference database, any protein that contains any observed peptide is considered to be a member of the Atlas.
- It is easy to see that the entire list of proteins in an Atlas is going to be highly redundant. Thus, we label each Atlas protein using the terminology below.
 - The term "observed peptides" in this context refers to the set of peptides in the PeptideAtlas build.
 - These peptides are selected using a PSM (peptide spectrum match)

Proteomics Databases: Peptide depositories

<http://thegpm.org/GPMDB/index.html>



The Global Proteome Machine

Proteomics data analysis, reuse and validation for biological and biomedical research.

The GPMDB Project

gpmDB: Design

gpmDB was designed to be a simplification and extension of the MIAPE scheme proposed by the PSI committee of HUPO. Rather than being a complete record of a proteomics experiment, this database holds the minimum amount of information necessary for certain bioinformatics-related tasks, such as sequence assignment validation. Most of the data is held in a set of XML files: the database serves as an index to those files, allowing for very rapid lookups and reduced database storage requirements. We call this combination of a relational database with XML data XIAPE (Xml Information About a Proteomics Experiment).

The Minimum Information About a Proteomics Experiment (MIAPE)

<http://www.psidev.info/node/91>

Nature Biotechnology 25, 887 - 893 (2007) PMID: 17687369

Methods Mol Biol. 2014;1072:765-80. PMID: 24136562

Proteomics Databases: Peptide depositories



HUMAN PROTEOME MAP

[Home](#)[Query](#)[Download](#)[FAQs](#)[Contact us](#)

About Human Proteome Map

The Human Proteome Map (HPM) portal is an interactive resource to the scientific community by integrating the massive peptide sequencing result from the draft map of the human proteome project. The project was based on LC-MS/MS by utilizing of high resolution and high accuracy Fourier transform mass spectrometry. All mass spectrometry data including precursors and HCD-derived fragments were acquired on the Orbitrap mass analyzers in the high-high mode. Currently, the HPM contains direct evidence of translation of a number of protein products derived from over 17,000 human genes covering >84% of the annotated protein-coding genes in humans based on >290,000 non-redundant peptide identifications of multiple organs/tissues and cell types from individuals with clinically defined healthy tissues. This includes 17 adult tissues, 6 primary hematopoietic cells and 7 fetal tissues. The HPM portal provides an interactive web resource by reorganizing the label-free quantitative proteomic data set in a simple graphical view. In addition, the portal provides selected reaction monitoring (SRM) information for all peptides identified.

Statistics

Organs/cell types	30
Genes identified	17,294
Proteins identified	30,057
Peptide sequences	293,700
N-terminal peptides	4,297
Splice junctional peptides	66,947
Samples	85
Adult tissues	17
Fetal tissues	7
Cell types	6

ARTICLE

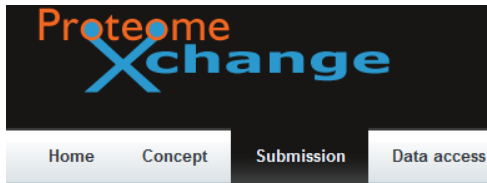
doi:10.1038/nature13302

A draft map of the human proteome

Min-Sik Kim^{1,2}, Sneha M. Pinto³, Derese Getnet^{1,4}, Raja Sekhar Nirujogi³, Srikanth S. Manda³, Raghothama Chaerkady^{1,2}, Anil K. Madugundu³, Dhanashree S. Kelkar³, Ruth Isserlin⁵, Shobhit Jain², Joji K. Thomas³, Babylakshmi Muthusamy³, Pamela Leal-Rojas^{1,6}, Praveen Kumar³, Nandini A. Sahasrabudhe³, Lavanya Balakrishnan³, Jayshree Advani³, Bijesh George³, Santosh Renuse³, Lakshmi Dhevi N. Selvan³, Arun H. Patil³, Vishalakshi Nanjappa³, Aneesh Radhakrishnan³, Samarjeet Prasad¹,

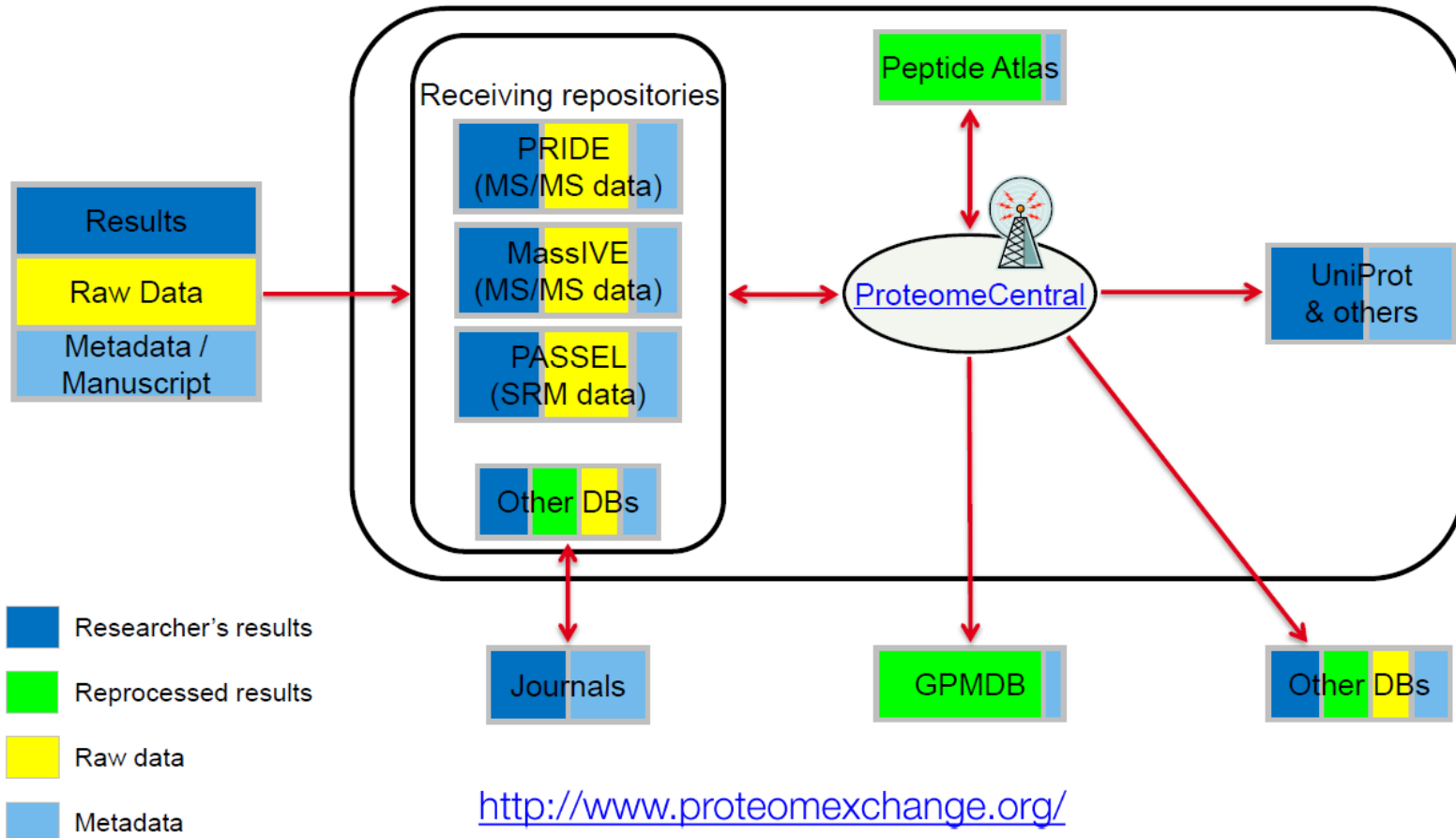
Kim & Akhilesh Pandey et al., *Nature*, 2014. PMID: 24870542

Proteomics Databases: Integrated Resources



<http://www.proteomexchange.org/>

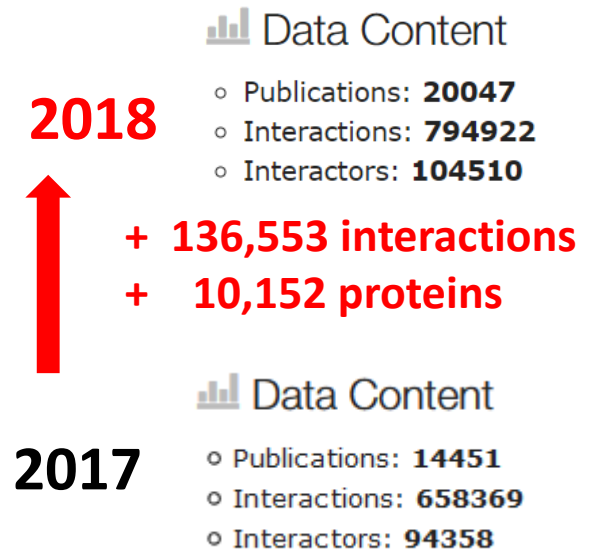
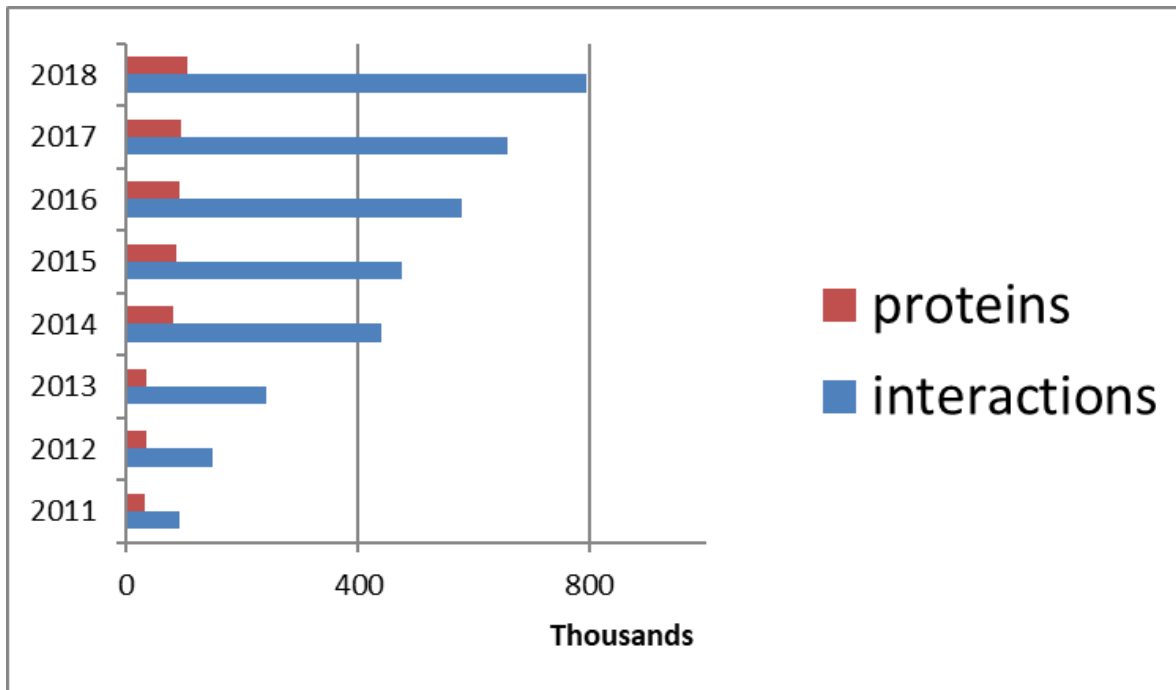
ProteomeXchange (PX) consortium



Protein-Protein Interaction Databases




<http://www.ebi.ac.uk/intact/>



Proteomics Databases: Integrated Resources Beyond Mass Spectrometry

<http://www.proteinatlas.org/>



A Tissue-Based Map of the Human Proteome

Here, we summarize our current knowledge regarding the human proteome mainly achieved through antibody-based methods combined with transcriptomics analysis across all major tissues and organs of the human body. A large number of lists can be accessed with direct links to gene-specific images of the corresponding proteins in the different tissues and organs. [Read more](#)

THE HUMAN PROTEIN ATLAS

Search [Fields »](#)

THE HUMAN PROTEOME

THE HUMAN PROTEOME

PROTEIN CLASSES

PROTEIN EVIDENCE

LEARN

DICTIONARIES

METHODS

CELL LINES

EVENTS

MEDIA & DATA

BLOG

MEDIA

DOWNLOADABLE DATA

INTRODUCTION

The Human Protein Atlas portal is a publicly available database with millions of high-resolution images showing the spatial distribution of proteins in 44 different normal human tissues and 20 different cancer types, as well as 46 different human cell lines. The data is released together with application-specific validation performed for each antibody, including immunohistochemistry, Western blot analysis and, for a large fraction, a protein array assay and immunofluorescent-based confocal microscopy. The database has been developed in a gene-centric manner with the inclusion of all human genes predicted from genome efforts. [Search functionalities](#) allow for complex queries regarding protein expression profiles, protein classes and chromosome location.

Uhlén et al (2015). **Tissue-based map of the human proteome.** *Science*
PubMed: 25613900 DOI: 10.1126/science.1260419

Uhlen et al (2010). **Towards a knowledge-based Human Protein Atlas.** *Nat Biotechnol.*
PubMed: 21139605 DOI: 10.1038/nbt1210-1248

Berglund et al (2008). **A genecentric Human Protein Atlas for expression profiles based on antibodies.** *Mol Cell Proteomics.*
PubMed: 18669619 DOI: 10.1074/mcp.R800013-MCP200

Uhlén et al (2005). **A human protein atlas for normal and cancer tissues based on antibody proteomics.** *Mol Cell Proteomics.*
PubMed: 16127175 DOI: 10.1074/mcp.M500279-MCP200

Pontén et al (2008). **The Human Protein Atlas - a tool for pathology.** *J Pathol.*
PubMed: 18853439 DOI: 10.1002/path.2440

Proteomics & Protein-Protein Interactions

Jesse Rinehart, PhD

**Biomedical Data Science: Mining & Modeling
CBB 752, Spring 2018**



**Cellular & Molecular Physiology
Yale University School of Medicine**



