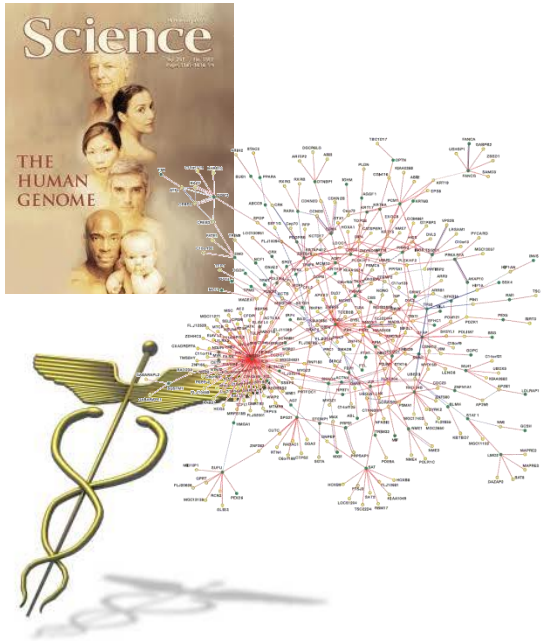
Proteomics & Protein-Protein Interactions

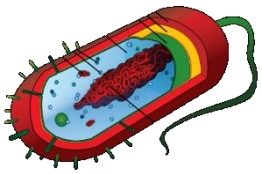
Jesse Rinehart, PhD

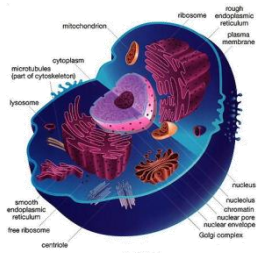
CBB 752, Spring 2017

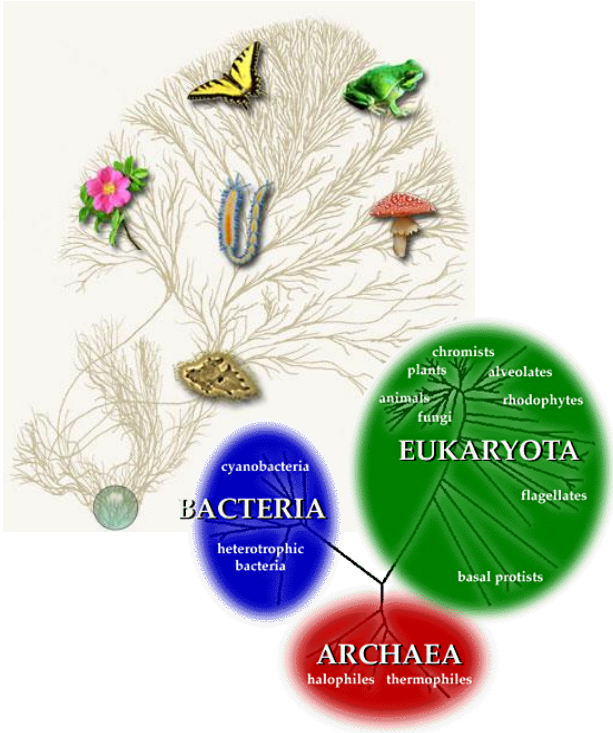
Cellular & Molecular Physiology
Yale University School of Medicine

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

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†]

Research Article
Protein & Cell
May 2015, Volume 6, Issue 5, pp 363-372
First online: 18 April 2015

April 2015

CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes

Puping Liang , Yanwen Xu , Xiya Zhang , Chenhui Ding , Rui Huang, Zhen Zhang, Jie Lv, Xiaowei Xie, Yuxi Chen and 7 more

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

THE WASHINGTON DIPLOMAT

Home News Culture Events Countries Directories Resources

Home ▸ About Us ▸ Archives ▸ January 2016 ▸

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

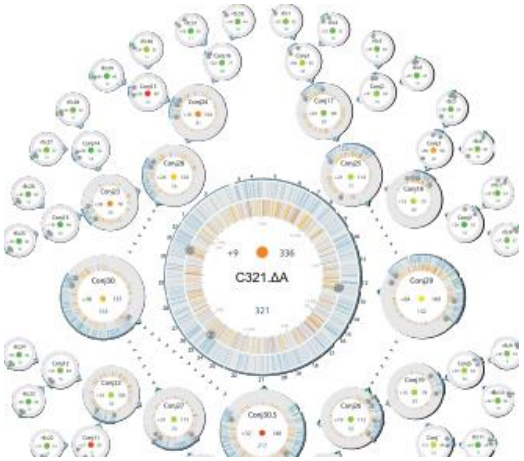
By Karen Pallarito (HealthDay News)

Uploaded on December 21, 2015

Dec 2015

SYNTHETIC BIOLOGY

DNA → RNA → PROTEIN



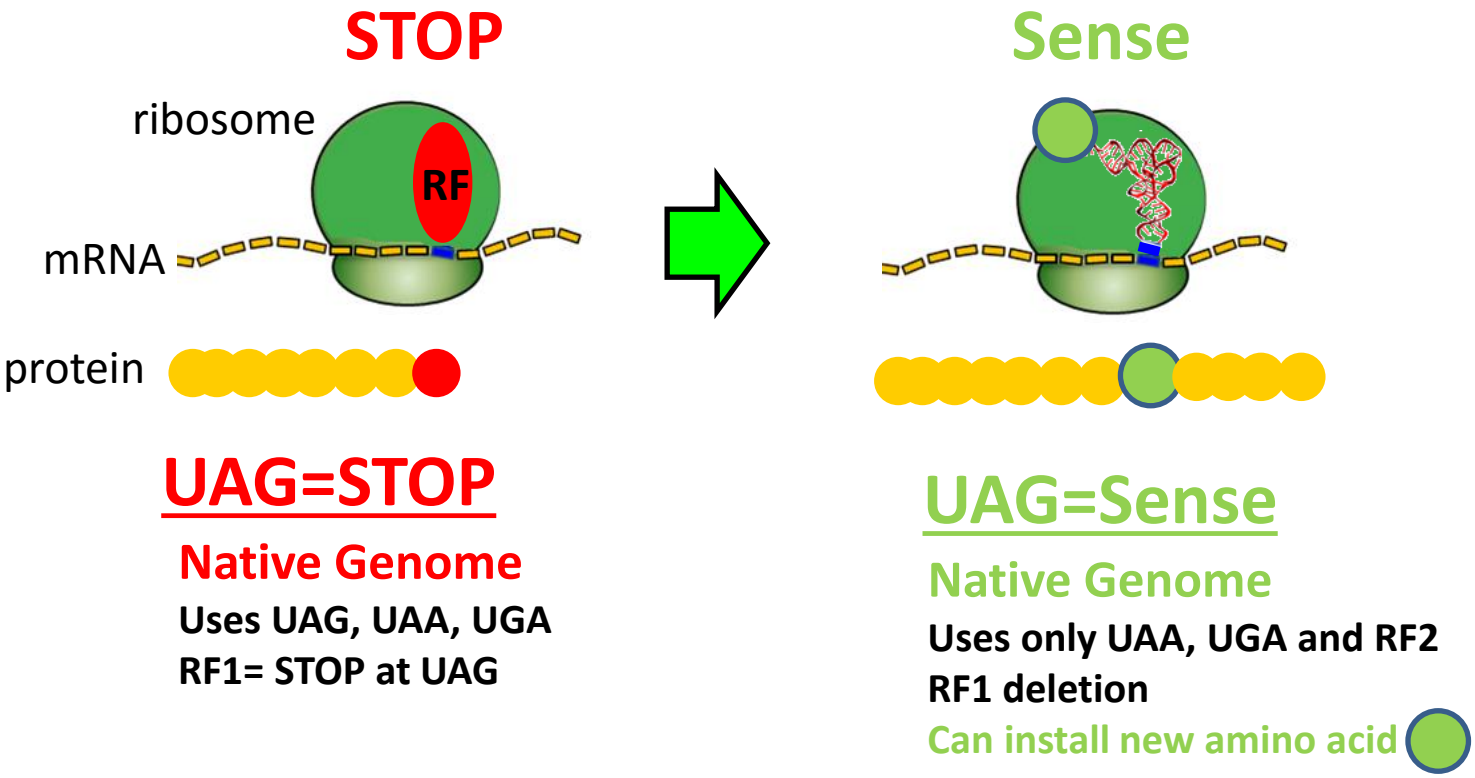
Genomically Recoded Organisms Expand Biological Functions

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)

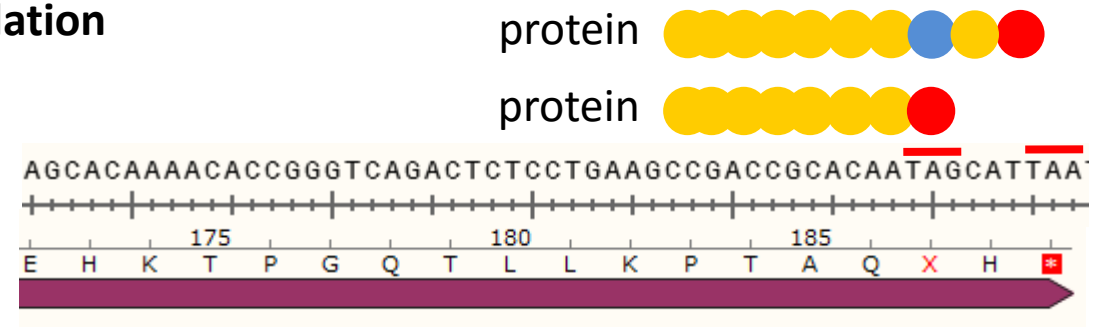
Lajoie *et al* used *E. coli* genome editing technology to change 321 native UAG stop codons to UAA and produced the *First Whole Genome Edited Organisim*



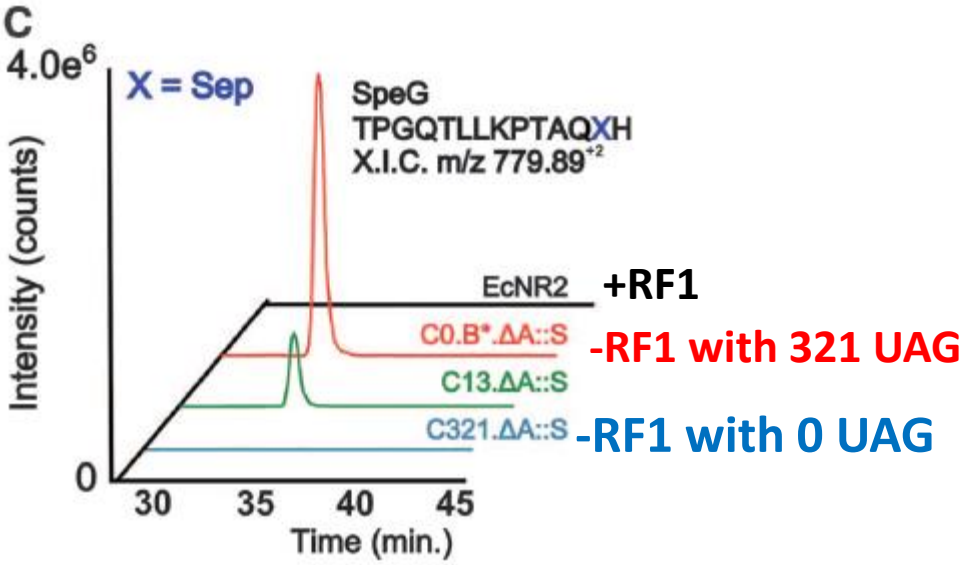
(Lajoie et al. Science 2013)

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



(Lajoie et al. Science 2013)

Proteomics & Protein-Protein Interactions

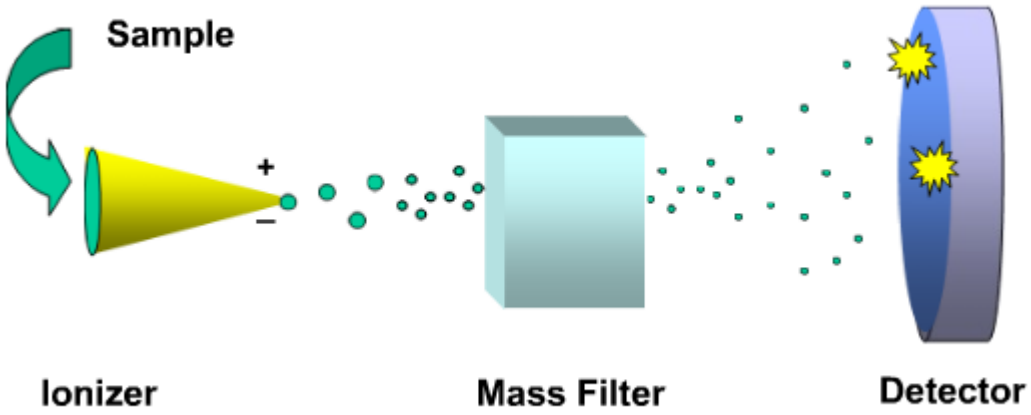
Overview

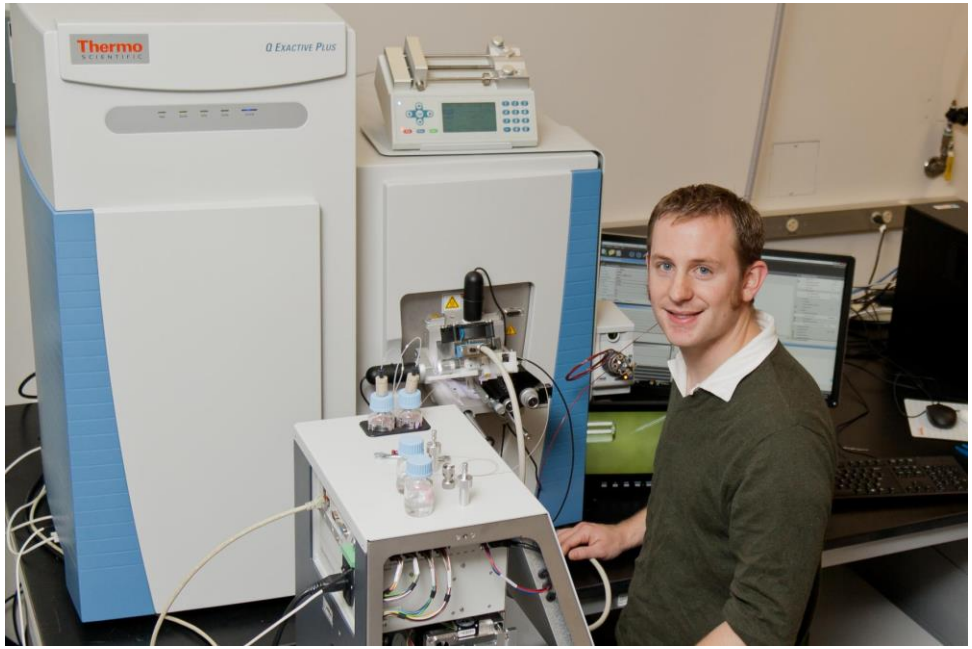
- **Techniques & Technologies**
  - Mass Spectrometry
  - Protein-Protein Interactions
  - Quantitative Proteomics
- **Applications**
  - Representative Studies
- **Putting it all together….**
  - Databases & Pathways

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

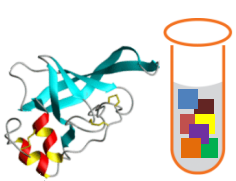
- Matrix Assisted Laser Desorption Ionization (MALDI)
  Tanaka K, Waki H, Ido Y, et al. *Rapid Commun Mass Spectrom* 1988

- 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

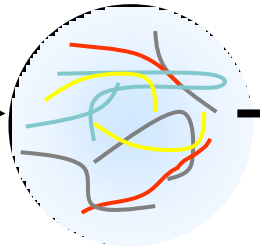
Typical work flow for LC-MS "shotgun proteomics"



Protein mixture

Trypsin Digest

Peptide Pool

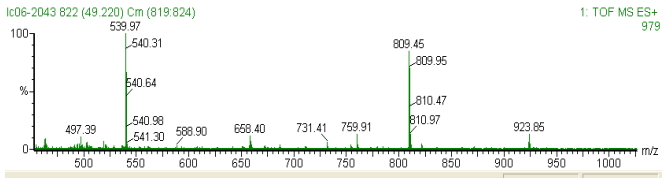


LC

n-UPLC

LTQ-Orbitrap MS

MS



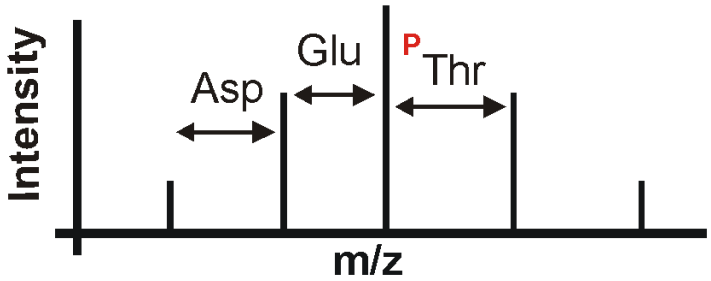
peptide

peptide

isolate & fragment

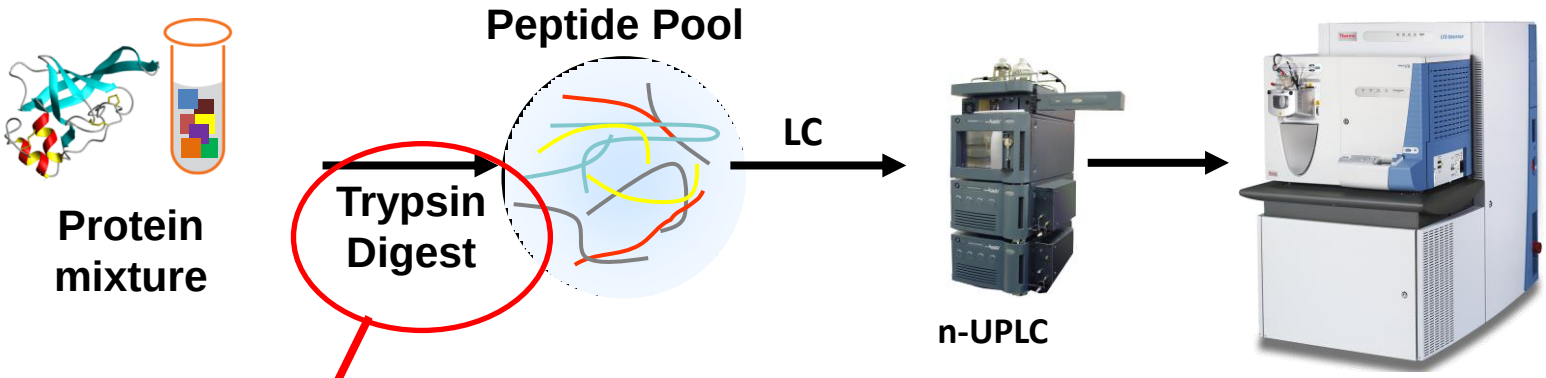
peptide

MS/MS

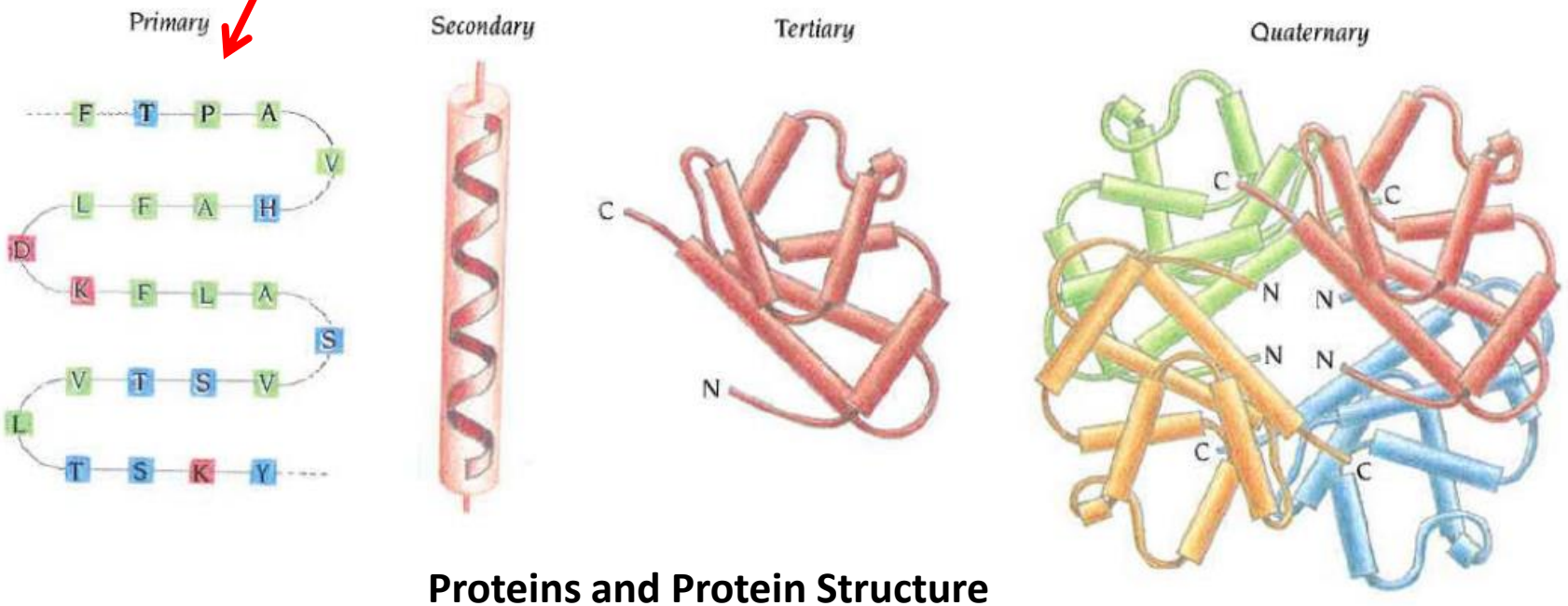


peptide fragments

Typical work flow for LC-MS "shotgun proteomics"



Trypsin cuts after Lys (K) & Arg (R)



Proteins and Protein Structure

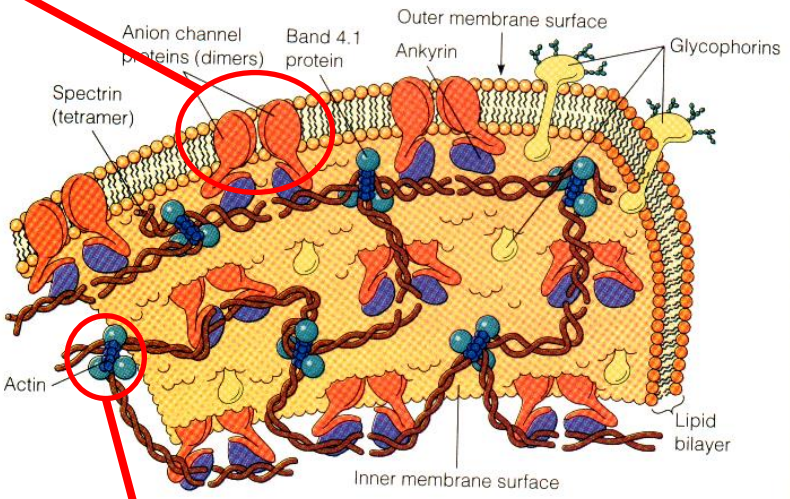
(Branden, C. and Tooze, J. ***Introduction to Protein Structure***)

Trypsin digest followed by LC-MS: Examples of "***Sequence Coverage***"

Band 3 Anion Transporter

Matched peptides shown in **Bold Red**

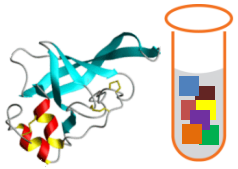
```
   1 MEELQDDYED MMEENLEQEE YEDPDIPESQ MEEPAAHDTE ATATDYHTTS
  51 HPGTHKVYVE LQELVMDEKN QELRWMEAAR WVQLEENLGE NGAWGRPHLS
 101 HLTFWSLLEL RRVFTKGTVL LDLQETSLAG VANQLLDRFI FEDQIRPQDR
 151 EELLRALLLK HSHAGELEAL GGVKPAVLTR SGDPSQPLLP QHSSLETQLF
 201 CEQGDGGTEG HSPSGILEKI PPDSEATLVL VGRADFLEQP VLGFVRLQEA
 251 AELEAVELPV PIRFLFVLLG PEAPHIDYTQ LGRAAATLMS ERVFRIDAYM
 301 AQSRGELLHS LEGFLDCSLV LPPTDAPSEQ ALLSLVPVQR ELLRRRYQSS
 351 PAKPDSSFYK GLDLNGGPDD PLQQTGQLFG GLVRDIRRRY PYYLSDITDA
 401 FSPQVLAAVI FIYFAALSPA ITFGGLLGEK TRNQMGVSEL LISTAVQGIL
 451 FALLGAQPLL VVGFSGPLLV FEEAFFSFCE TNGLEYIVGR VWIGFWLILL
 501 VVLVVAFEGS FLVRFISRYT QEIFSFLISL IFIYETFSKL IKIFQDHPLQ
 551 KTYNYNVLMV PKPQGPLPNT ALLSLVLMAG TFFFAMMLRK FKNSSYFPGK
 601 LRRVIGDFGV PISILIMVLV DFFIQDTYTQ KLSVPDGFKV SNSSARGWVI
 651 HPLGLRSEFP IWMMFASALP ALLVFILIFL ESQITTLIVS KPERKMVKGS
 701 GFHLDLLLVV GMGGVAALFG MPWLSATTVR SVTHANALTV MGKASTPGAA
 751 AQIQEVKEQR ISGLLVAVLV GLSILMEPIL SRIPLAVLFG IFLYMGVTSL
 801 SGIQLFDRIL LLFKPPKYHP DVPYVKRVKT WRMHLFTGIQ IICLAVLWVV
 851 KSTPASLALP FVLILTVPLR RVLLPLIFRN VELQCLDADD AKATFDEEEG
 901 RDEYDEVAMP V
```



β-actin

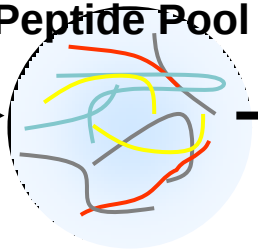
Matched peptides shown in **Bold Red**

```
   1 MDDDIAALVV DNGSGMCKAG FAGDDAPRAV FPSIVGRPRH QGVMVGMGQK
  51 DSYVGDEAQS KRGILTLKYP IEHGIVTNWD DMEKIWHHTF YNELRVAPEE
 101 HPVLLTEAPL NPKANREKMT QIMFETFNTP AMYVAIQAVL SLYASGRTTG
 151 IVMDSGDGVT HTVPIYEGYA LPHAILRLDL AGRDLTDYLM KILTERGYSF
 201 TTTAEREIVR DIKEKLCYVA LDFEQEMATA ASSSSLEKSY ELPDGQVITI
 251 GNERFRCPEA LFQPSFLGME SCGIHETTFN SIMKCDVDIR KDLYANTVLS
 301 GGTTMYPGIA DRMQKEITAL APSTMKIKII APPERKYSVW IGGSILASLS
 351 TFQQMWISKQ EYDESGPSIV HRKCF
```



Protein mixture

Enz. Digest

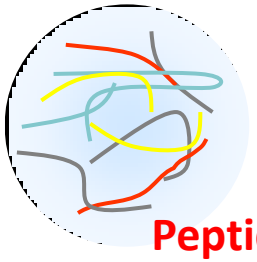


Peptide Pool

LC

n-UPLC

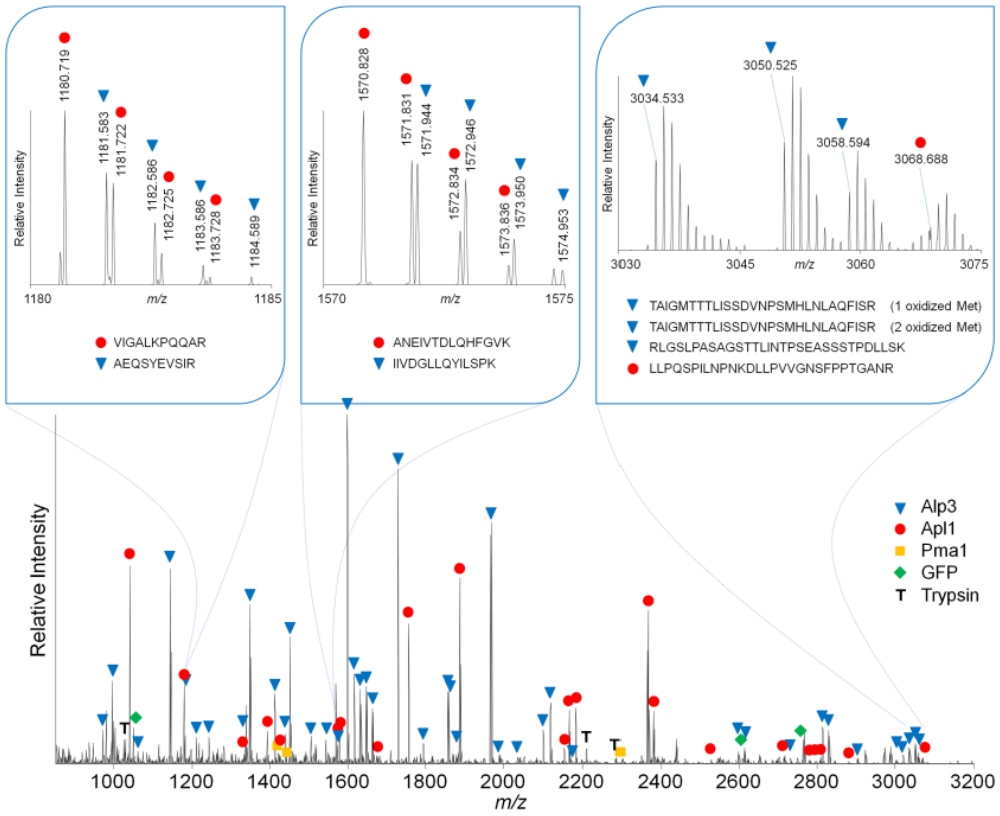
LTQ-Orbitrap MS

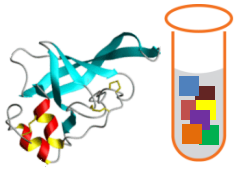


Peptide ions have a mass (m) and a charge (z).

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

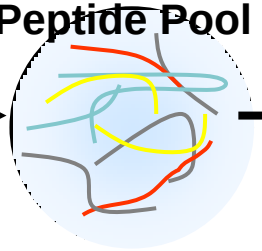
Mass Spectrum





Protein mixture

Enz. Digest

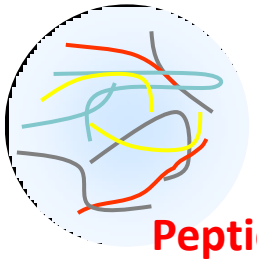


Peptide Pool

LC

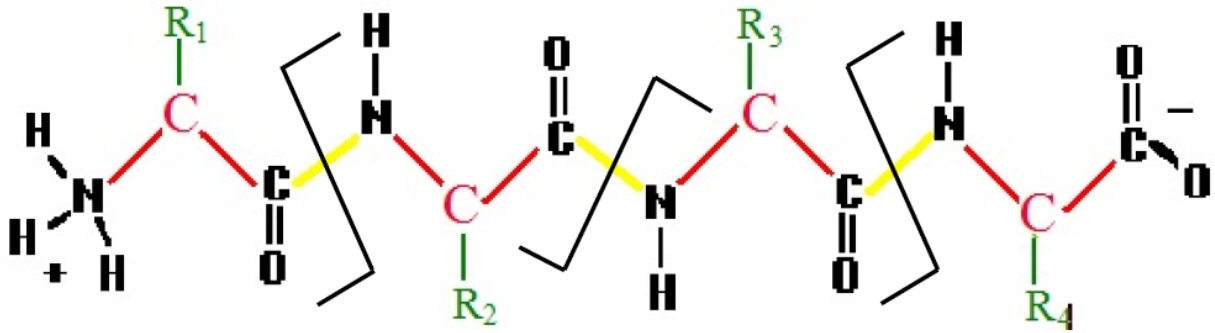
n-UPLC

LTQ-Orbitrap MS

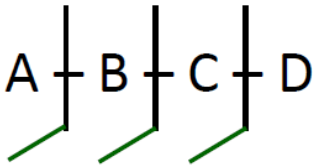


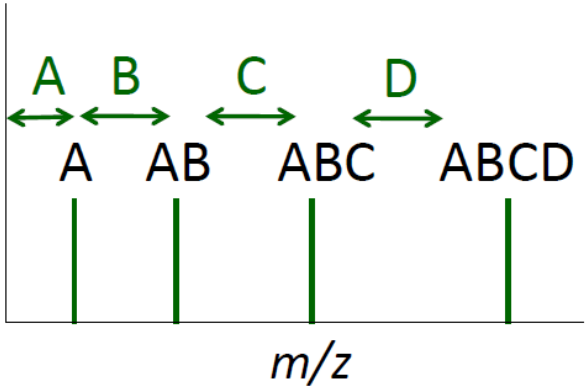
Peptide ions are isolated and "sequenced"

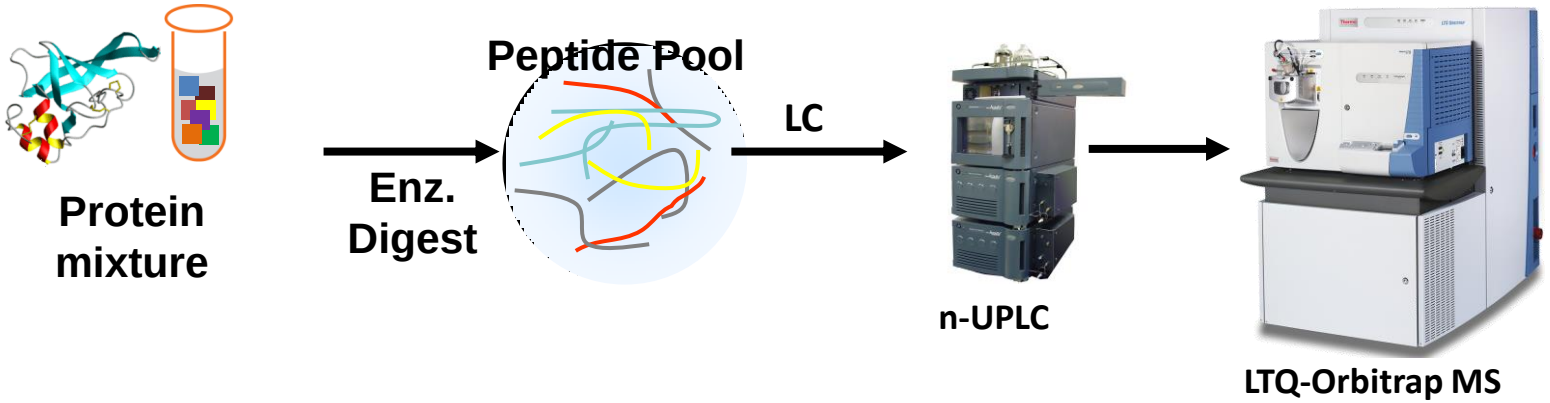
Peptide sequencing



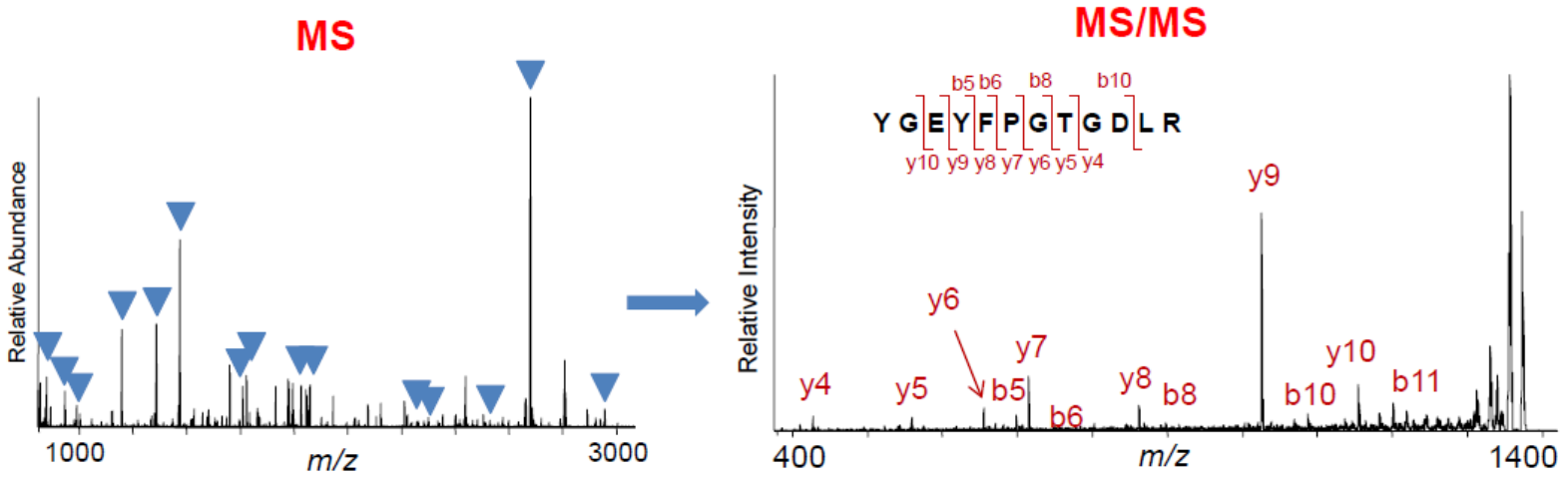
Simplified concept of peptide fragmentation







Database searching – at MS or MS/MS level



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

Proteomics

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

[Study of post-translational modifications (protein phosphorylation, acetylation, glycosylation …) via MS has grown in recent years to dramatically expand the field of Proteomics]

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**

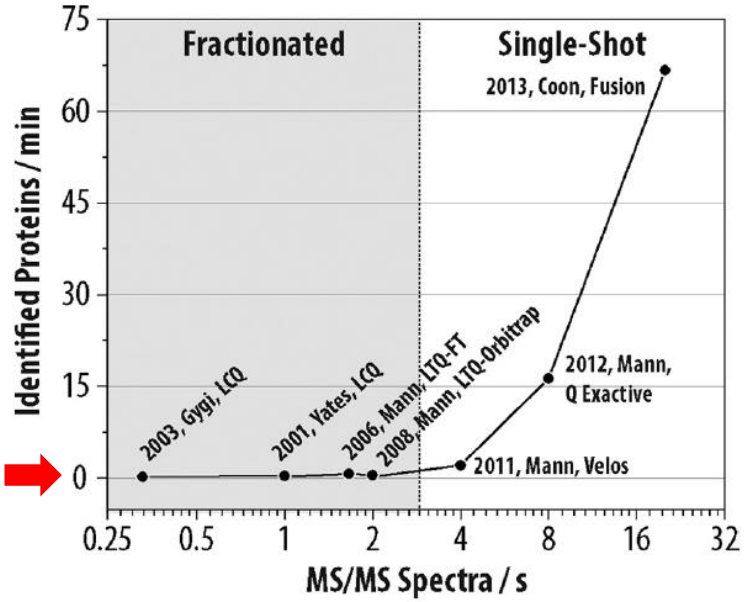


FIG. 5. **Rate of protein identifications as a function of mass spectrometer scan rate for selected large-scale yeast proteome analyses over the past decade.** Each data point is annotated with the year, corresponding author, type of MS system used, and reference number.

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

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

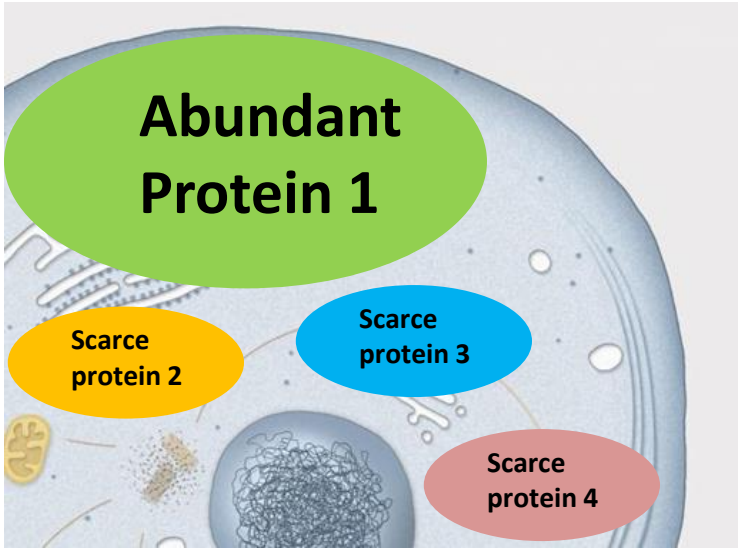
One-hour proteome analysis in yeast

Alicia L Richards[1,2,4], Alexander S Hebert[1,3,4], Arne Ulbrich[1,2], Derek J Bailey[1,2], Emma E Coughlin[1], Michael S Westphall[1] & Joshua J Coon[1–3]

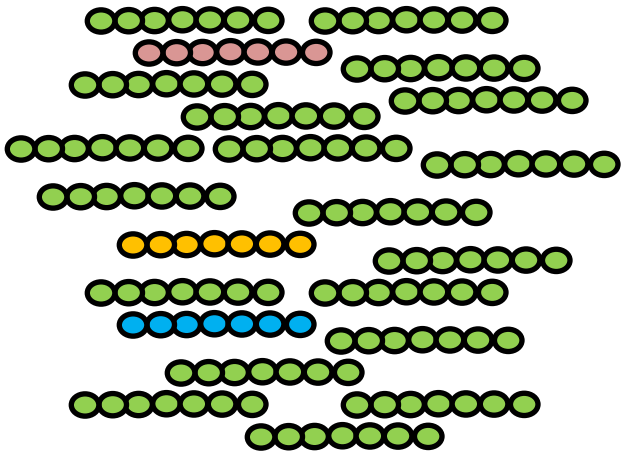
" …the identification of up to **4,002 proteins**, This protocol, which includes cell lysis, overnight tryptic digestion, sample analysis and database searching, **takes ~24 h to complete**."

The one hour yeast proteome. Hebert AS, et a, Coon JJ.
Mol Cell Proteomics. 2014 PMID: 24143002 & *Nat Protoc.* 2015. PMID: 25855955

Cell with a 4 protein proteome



Whole Proteome Tryptic Digest

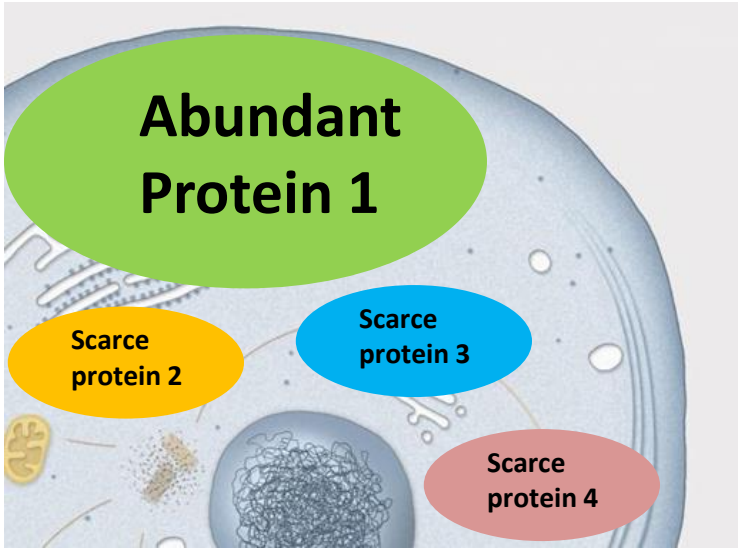


One LC-MS run

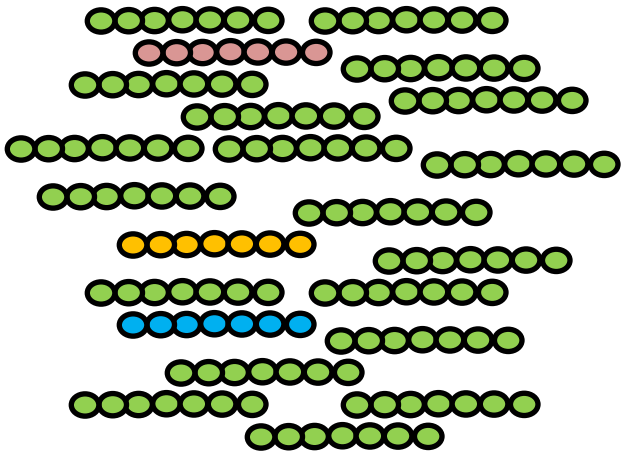
(Hypothetical MS that can only identify one peptide)

Protein 1 Identified

Cell with a 4 protein proteome



Whole Proteome Tryptic Digest

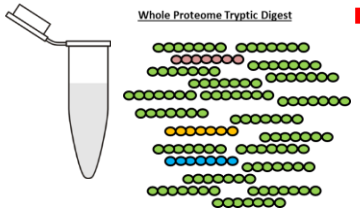


One LC-MS run

(Hypothetical MS that can only identify one peptide)

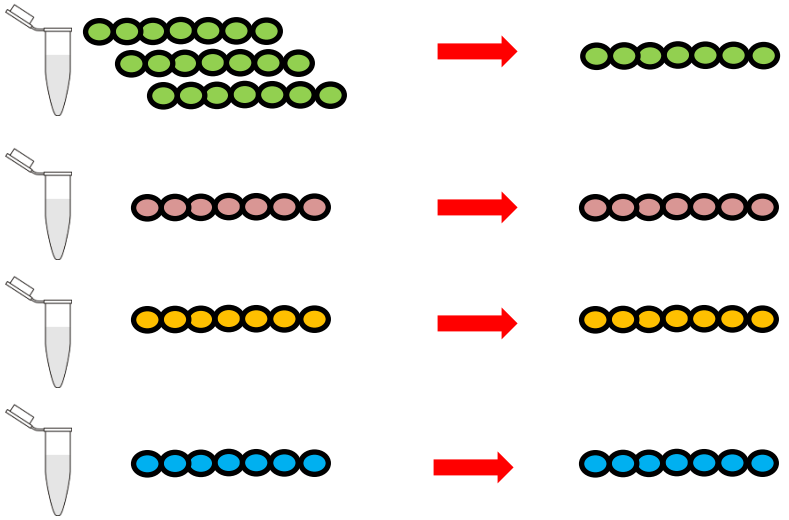
Protein 1 Identified

Option #1: Peptide Fractionation



Chromatography + fractionation

4 separate LC-MS runs



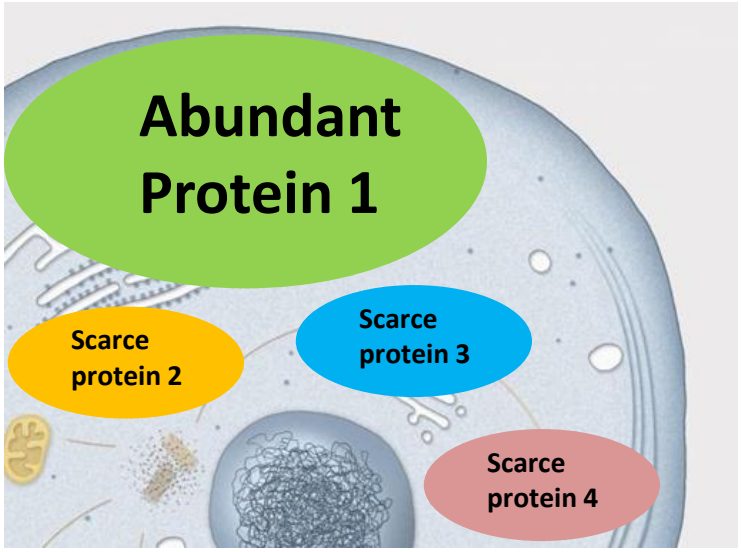
Protein 1 Identified

Protein 2 Identified

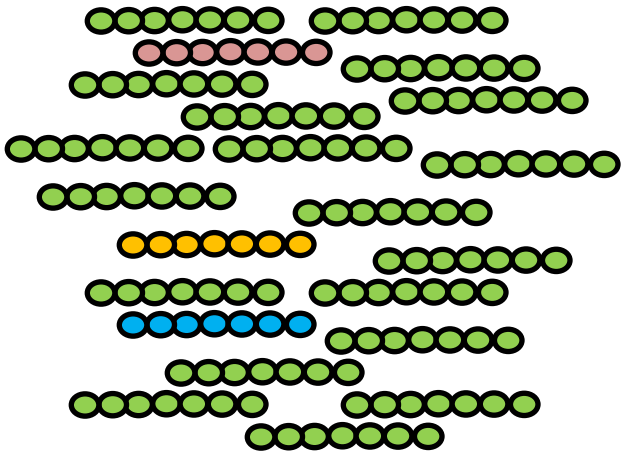
Protein 3 Identified

Protein 4 Identified

Cell with a 4 protein proteome



Whole Proteome Tryptic Digest

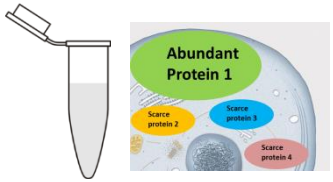


One LC-MS run

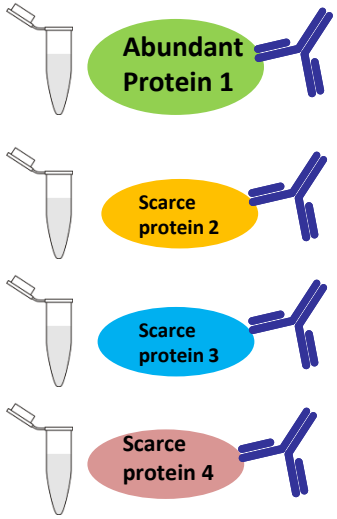
(Hypothetical MS that can only identify one peptide)

Protein 1 Identified

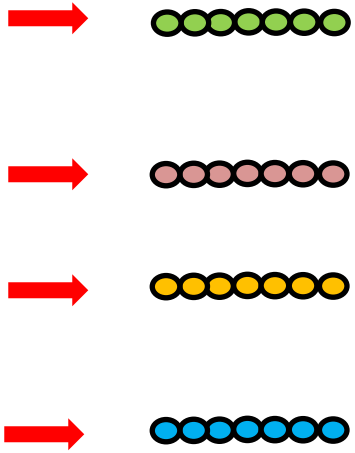
Option #2: Proteome Fractionation (e.g. Immunoprecipitation)



Separate IP Tryptic Digest



4 separate LC-MS runs



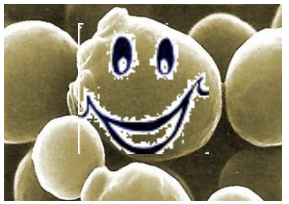
Protein 1 Identified

Protein 2 Identified

Protein 3 Identified

Protein 4 Identified

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

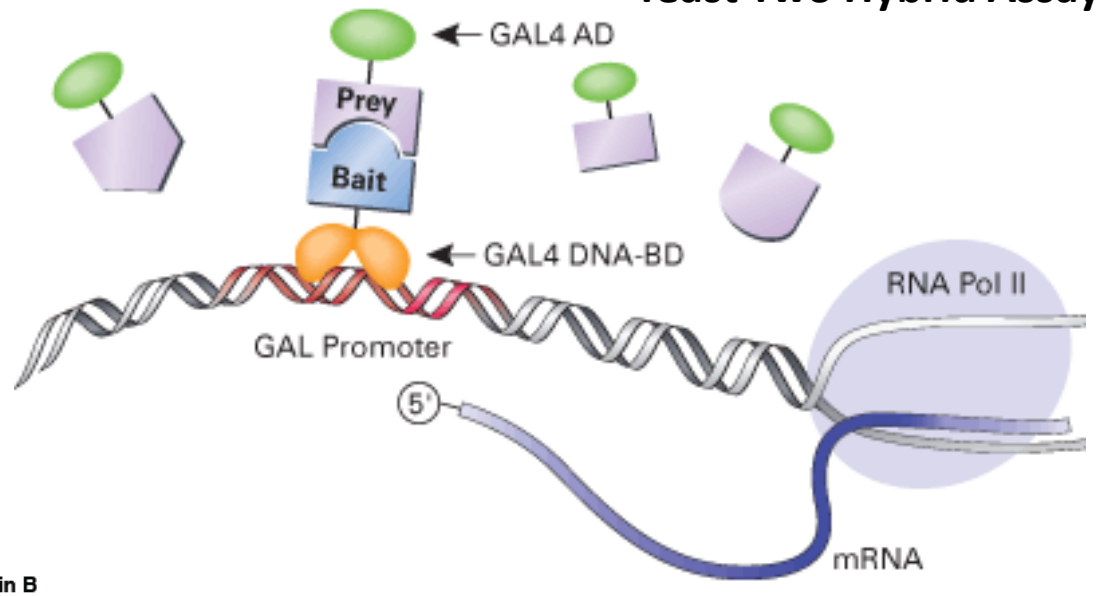
Picotti P, et al. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell.*

➲ **Towards proteome wide targeted proteomics.**

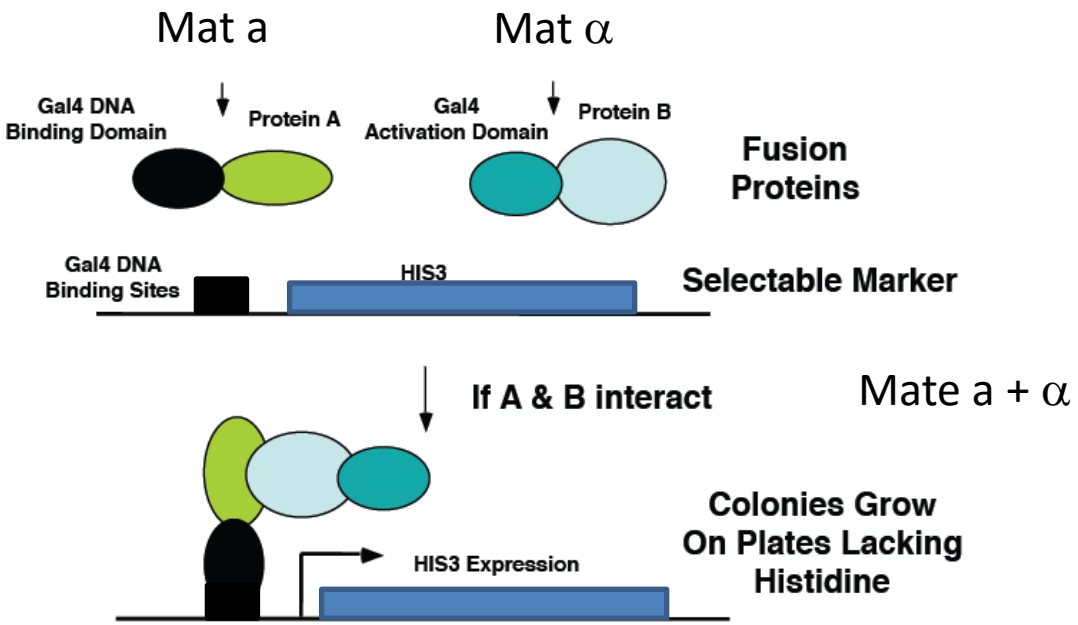
A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.

Uetz et al, Nature 2000
Ito et al, PNAS 2001

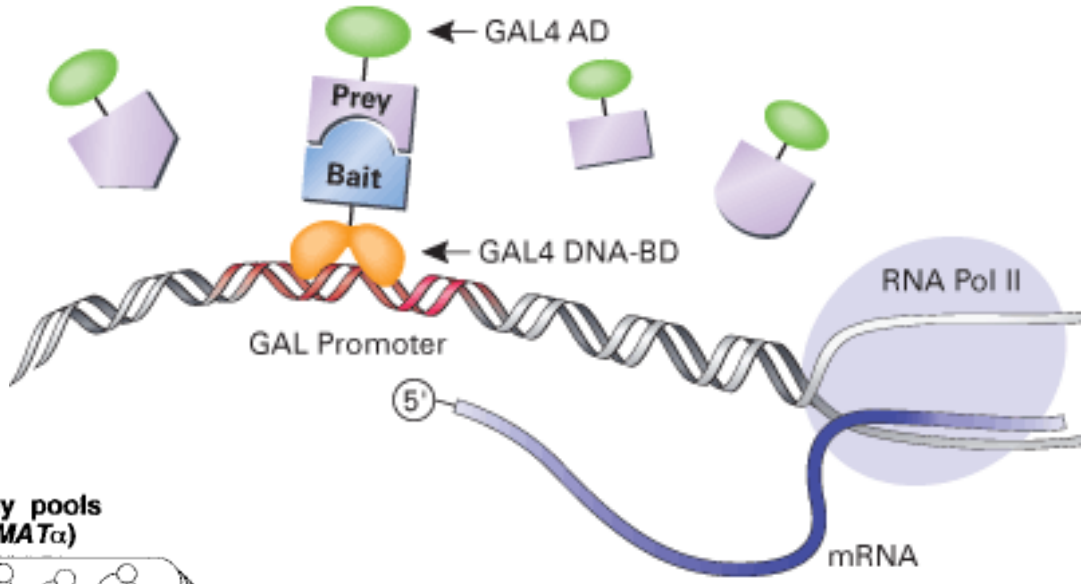
Yeast Two Hybrid Assay

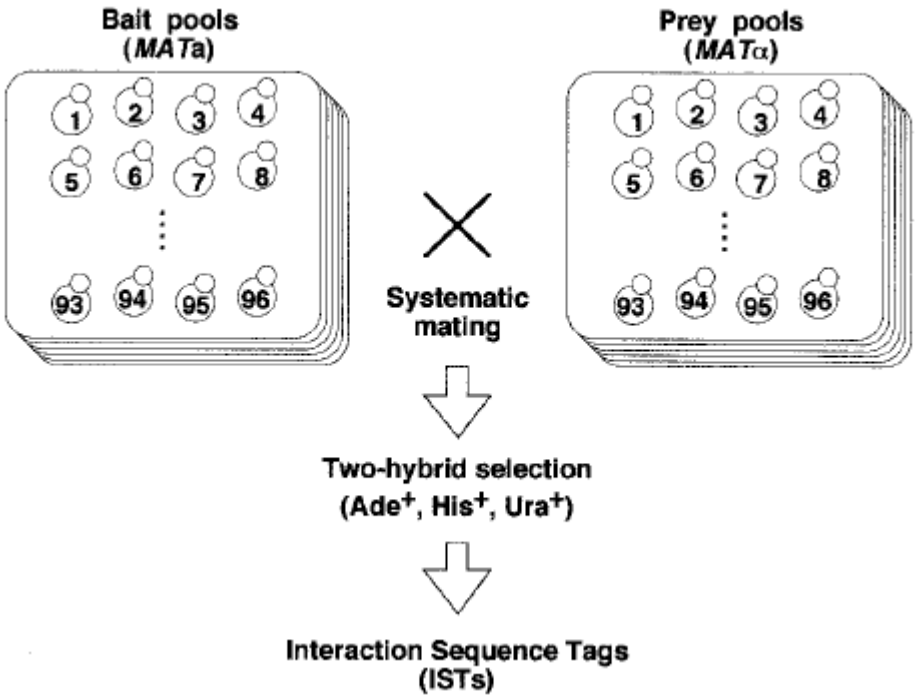


Clone bait and prey constructs and place in separate strains.



Uetz et al, Nature 2000
Ito et al, PNAS 2001





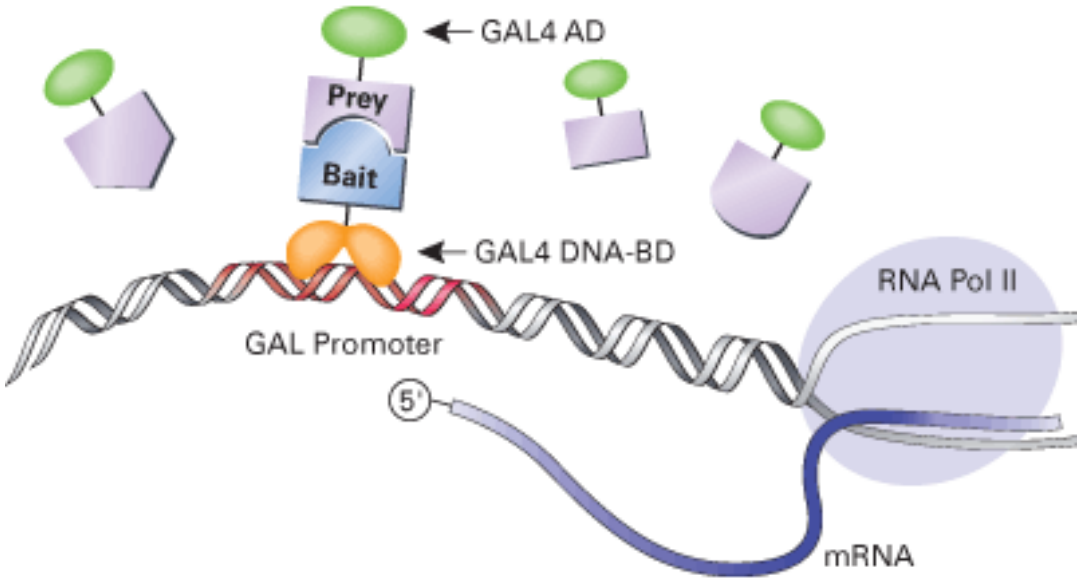
Results of Two Studies

1) 4,549 Interactions Among 3,278 Proteins (Ito et al.)

2) 957 Interactions 1004 proteins (Utez 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**

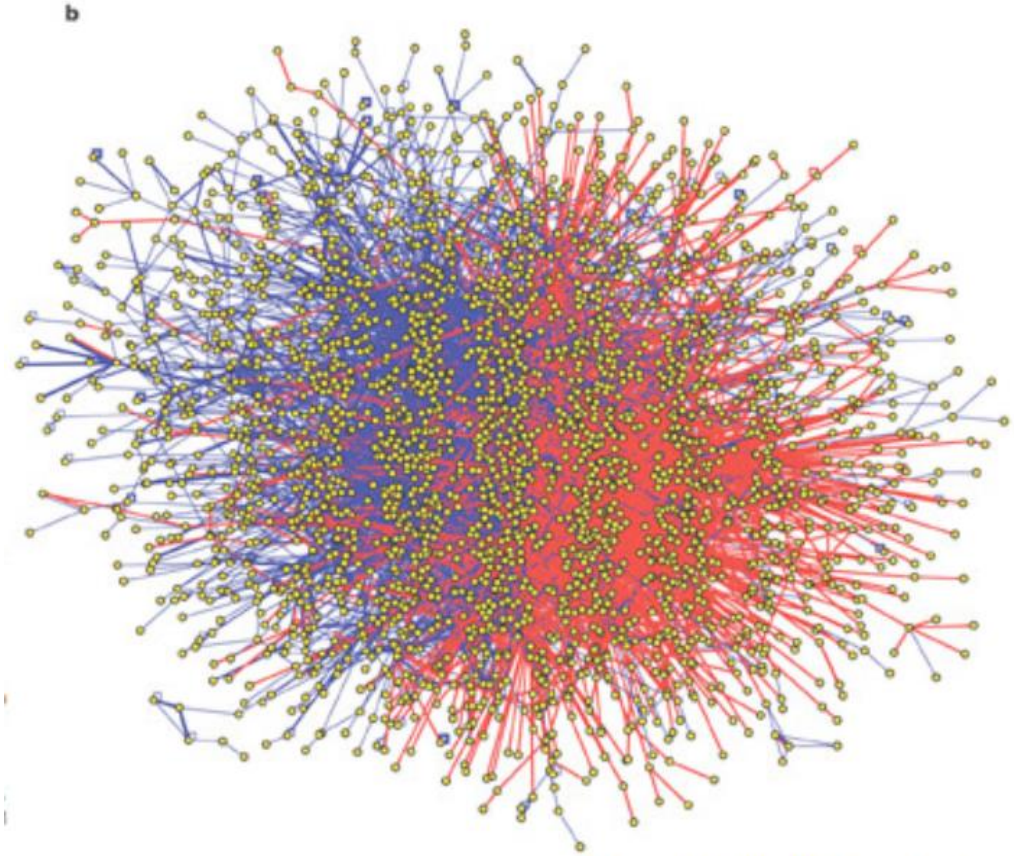
Some Disadvantages

- **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)

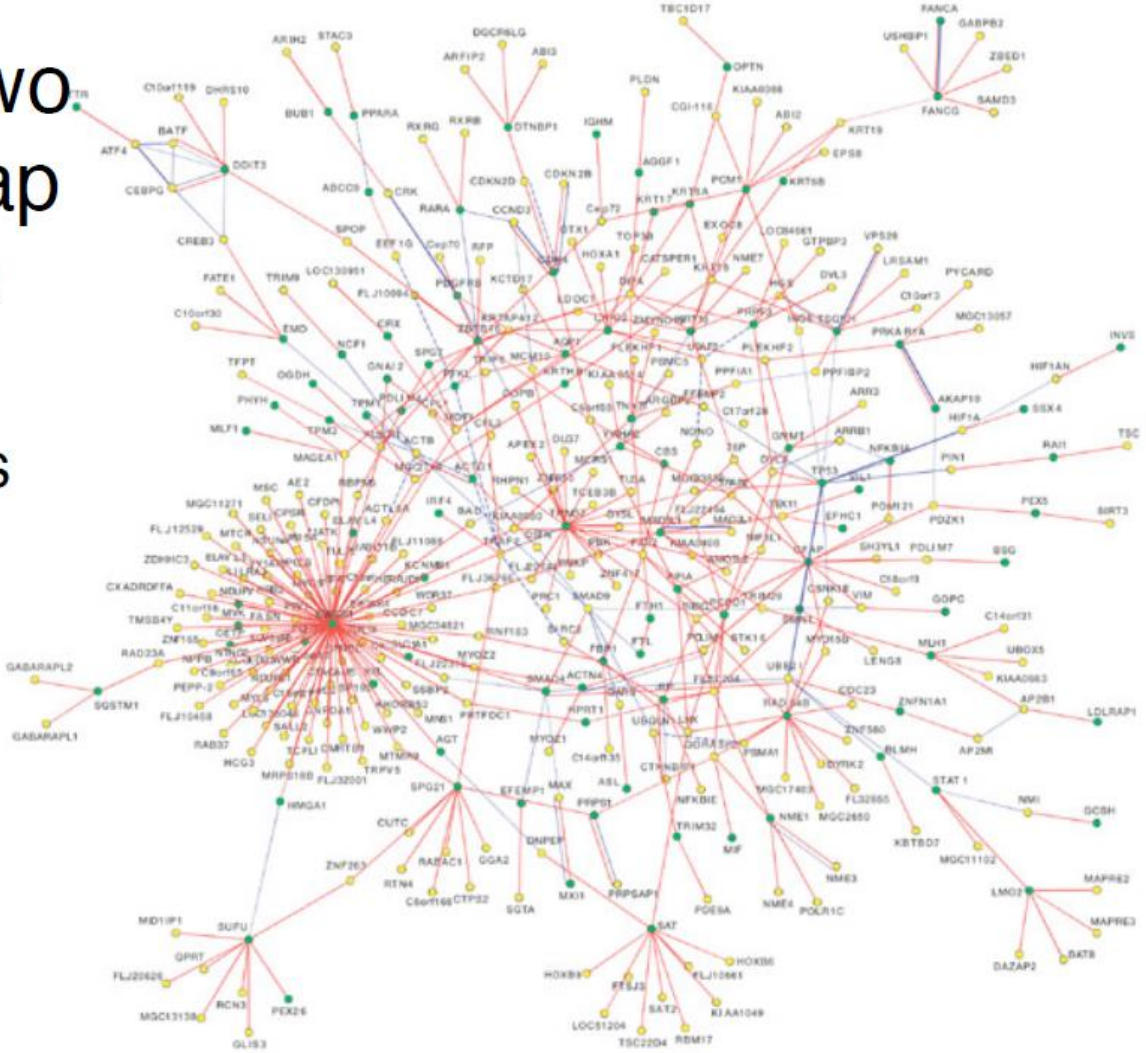
10,597 interactions



Rual et al. Nature 2005

Human Two Hybrid Map Disease Genes

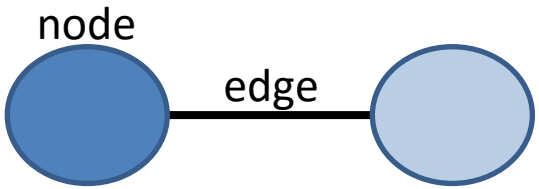
(121 genes (green))

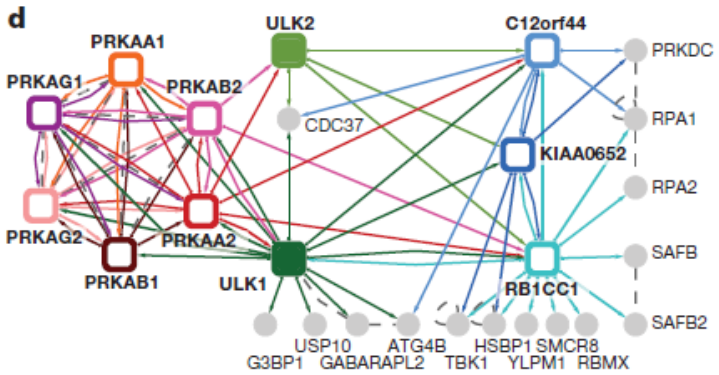


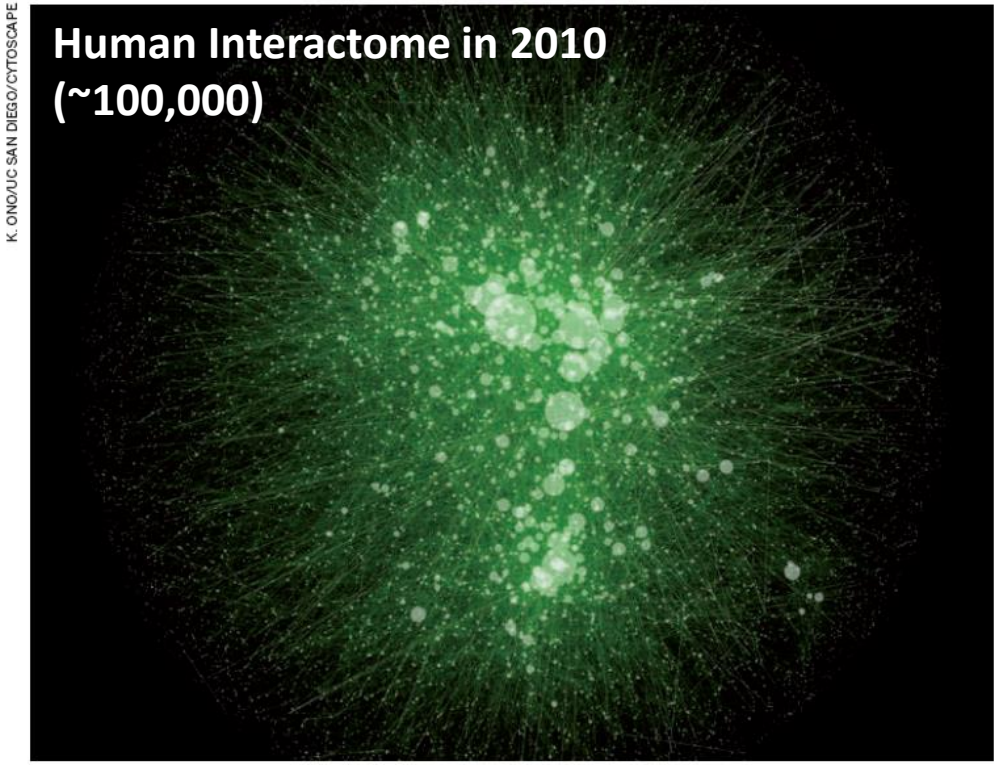
Rual et al. Nature 2005 Vol 437

Protein-Protein interaction maps:

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

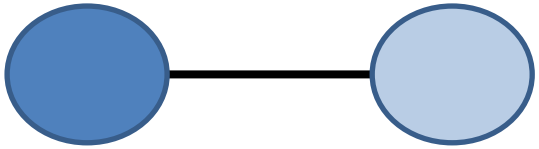






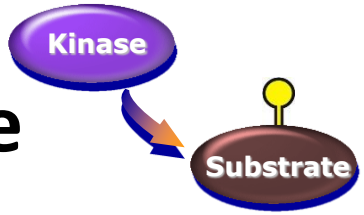
Bonetta, *Nature* 2010

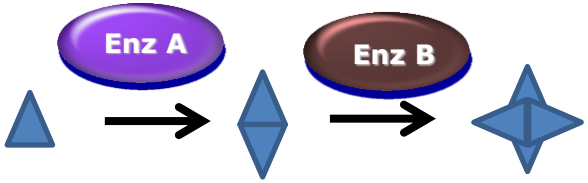
Protein-Protein interactions:



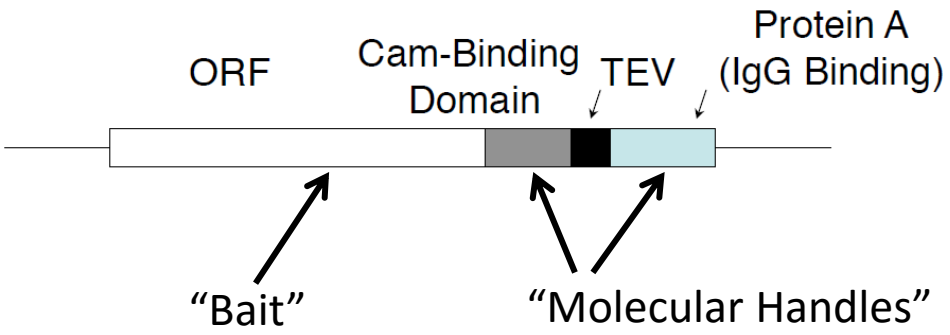
Some examples:

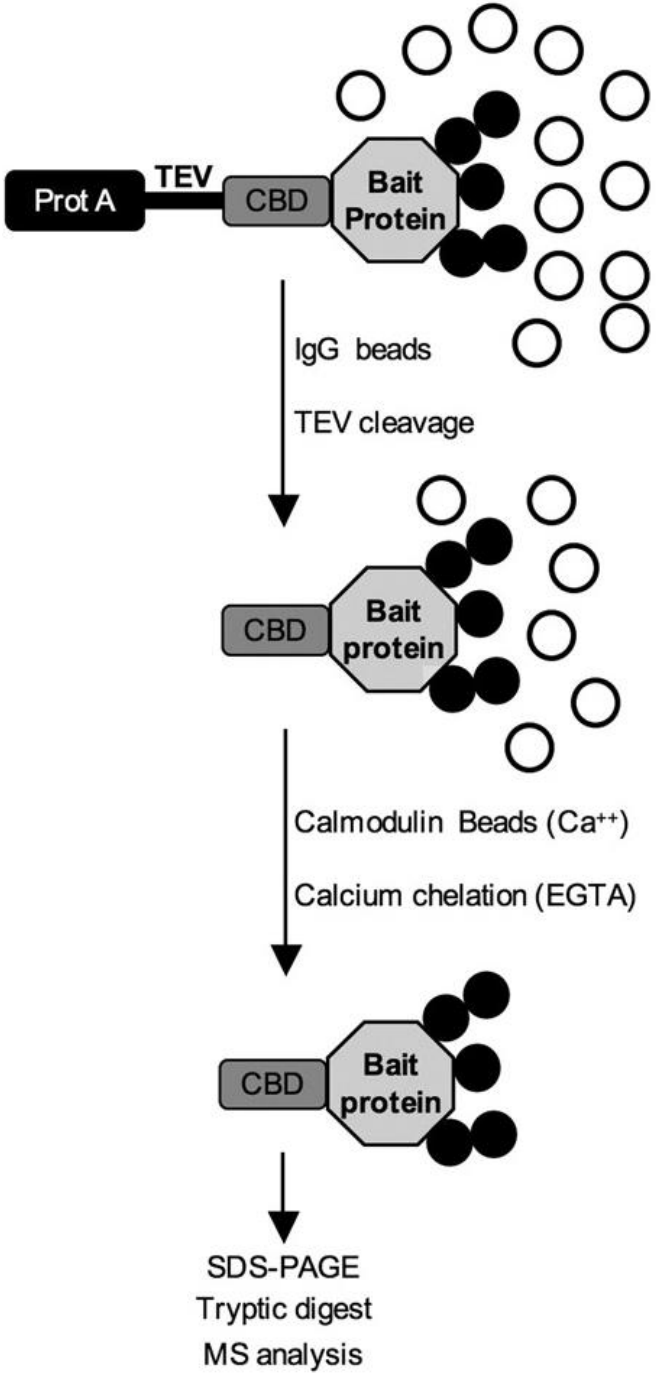
- Physical and direct
- Physical and indirect
  - Multi-protein complexes
  - Scaffolds
- Transient
  - Kinase & substrate
- Metabolic





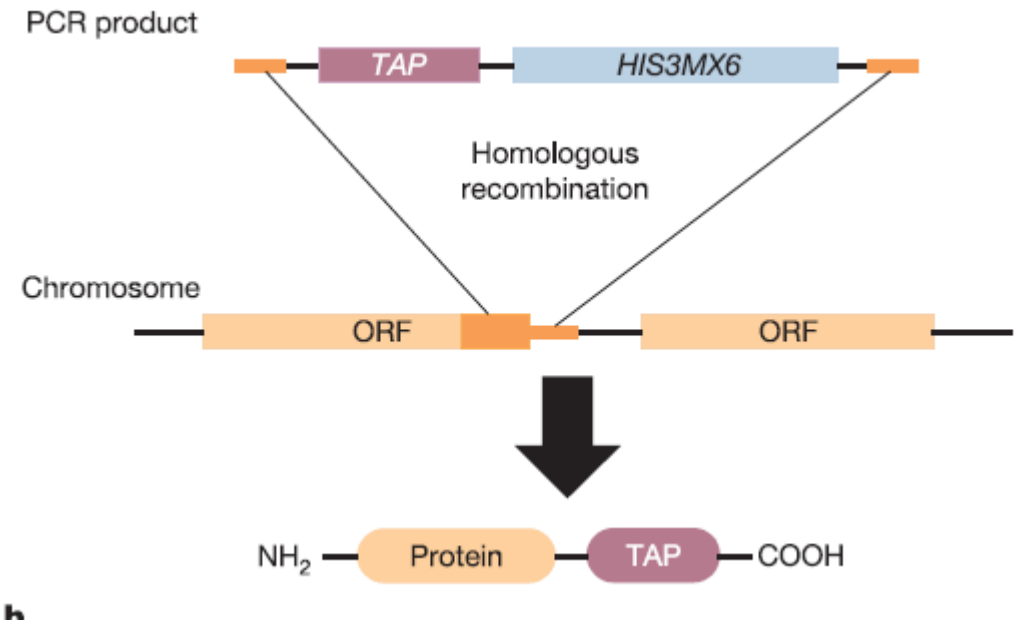
Tandem Affinity Purification (TAP) Tagging





Cannavo E et al. J. Biol. Chem. 2007

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* .

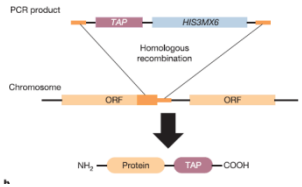
➲ Protein–protein interaction maps can be obtained by MS; the yeast cell is organized into protein complexes.

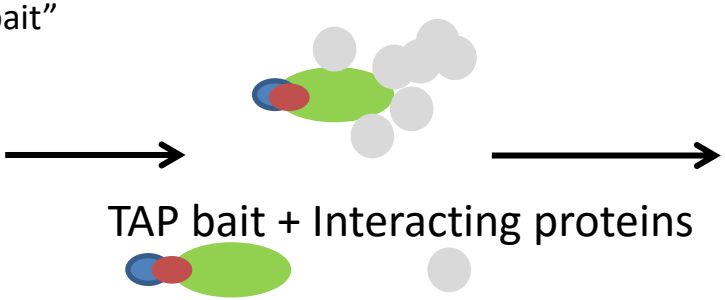
2006

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

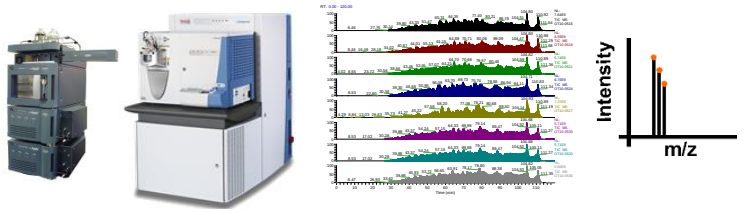
➲ TAP-Tag and Protein-Protein Interaction

Collection of tagged "bait" expression strains





Multiple runs of "shot gun" MS & SDS-PAGE with MS on individual proteins

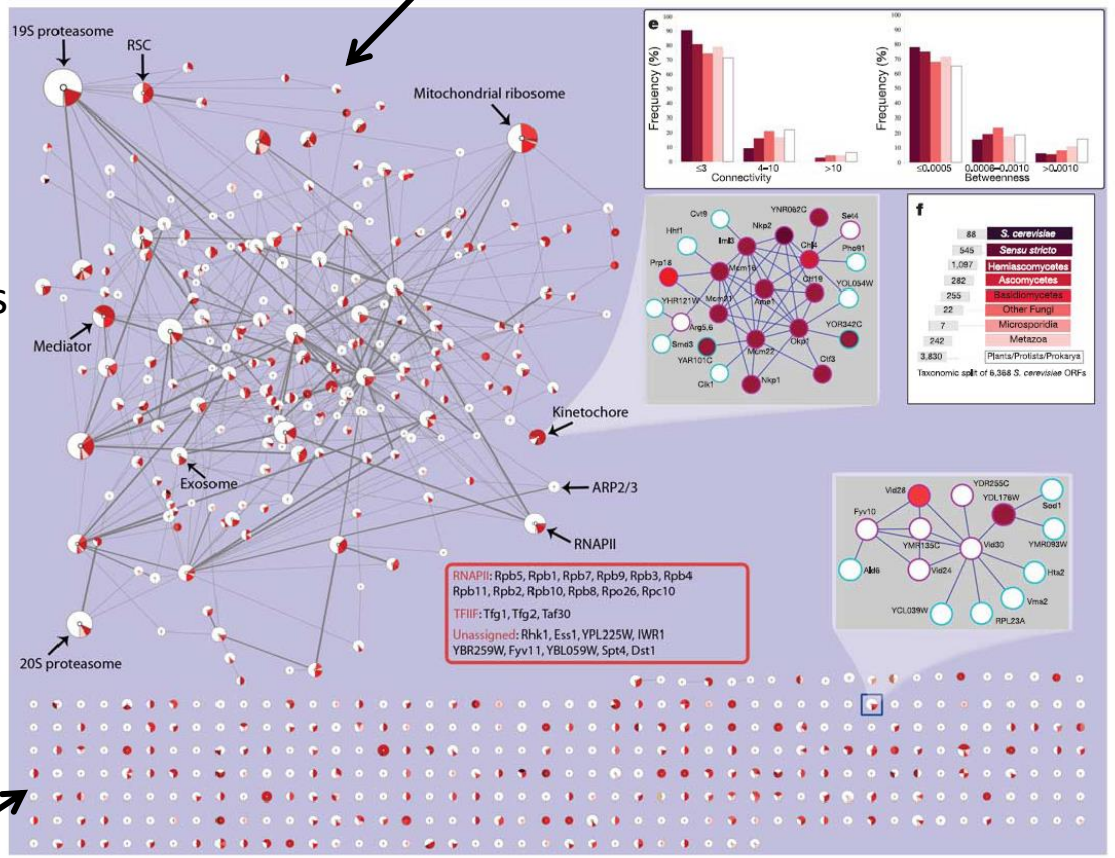


Krogan et al. observed 7,123 protein–protein interactions:

<u>Important aspects:</u>

- **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…"**

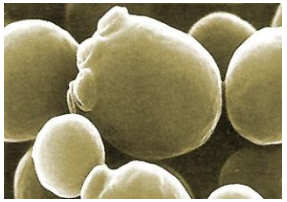
- 4,562 tagged proteins
- 2,357 successful purifications
- Identified 4,087 interacting proteins ~72 % proteome
- Majority of the yeast proteome is organized into complexes
- Many complexes are conserved in other species



Krogan NJ, et al. *Nature.* 2006

Complexes with little or no interconnectivity

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

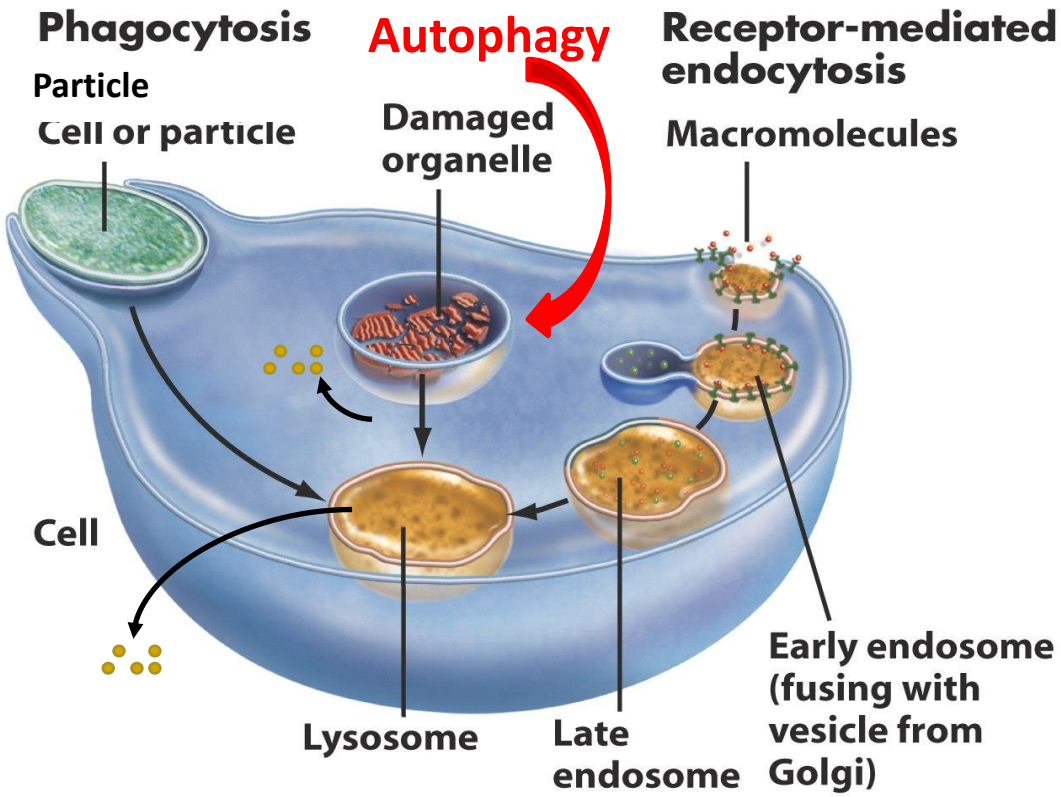


nature Vol 466|1 July 2010|doi:10.1038/nature09204

ARTICLES

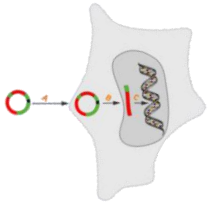
Network organization of the human autophagy system

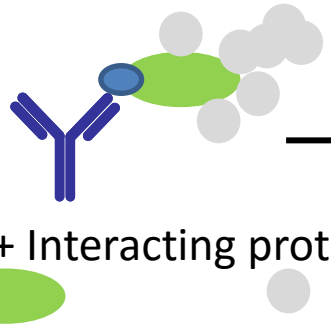
Christian Behrends[1], Mathew E. Sowa[1], Steven P. Gygi[2] & J. Wade Harper[1]



Pearson Prentice Hall, Inc. 2005
www.stolaf.edu/people/giannini/cell/lys.htm

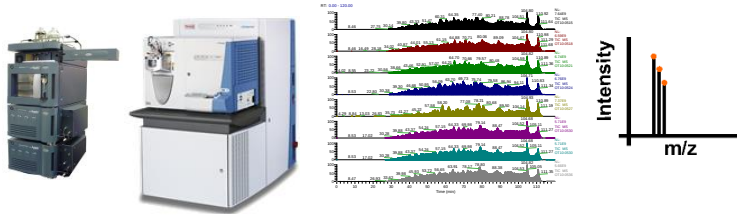
Transfect tagged "bait"





IP Bait + Interacting proteins

Multiple runs of "shot gun" LC-MS/MS



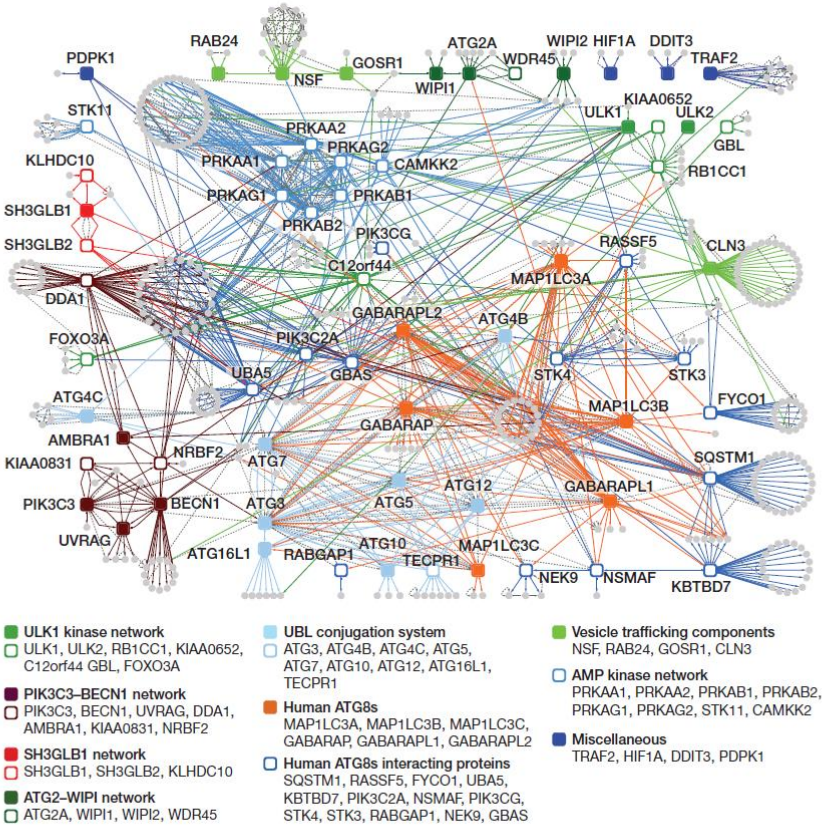
~65 bait proteins
LC-MS/MS identifies
2553 proteins

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



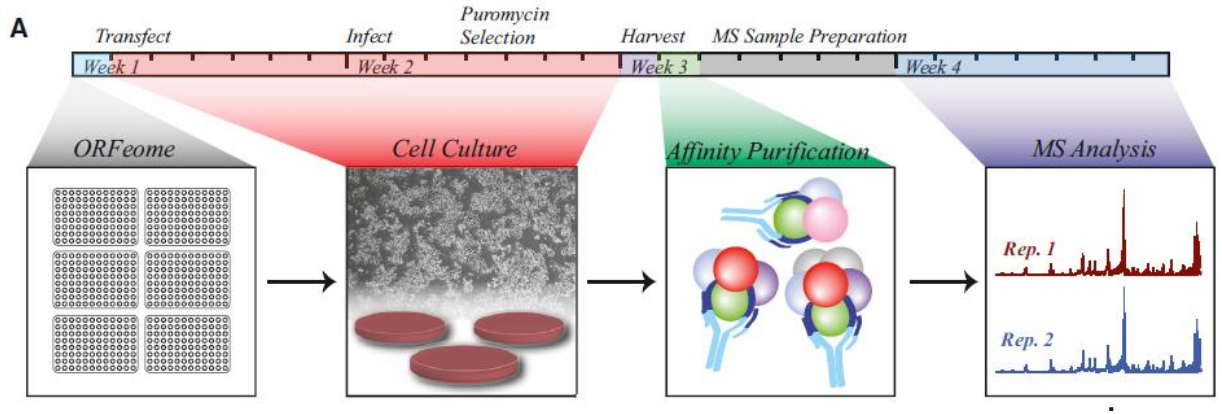
Behreands et al, Nature 2010

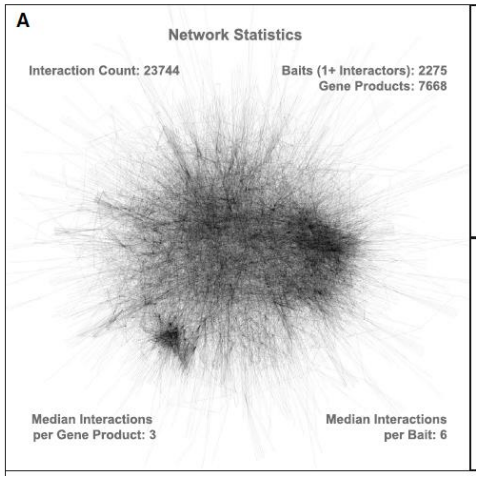
Figure 1 | Overview of the autophagy interaction network (AIN). HCIPs within the autophagy network are shown for 32 primary baits (filled squares) and 33 secondary baits (open squares). Subnetworks are colour-coded. Interacting proteins are indicated by grey circles.

Cell

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 Wühr,[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,*]





http://wren.hms.harvard.edu/bioplex/

A first paper in *Cell* reports the first ~2,500 experiments (~23,000 interactions). Our current release with more than 5,000 human proteins as baits (~50,000 interactions) is also now available.

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

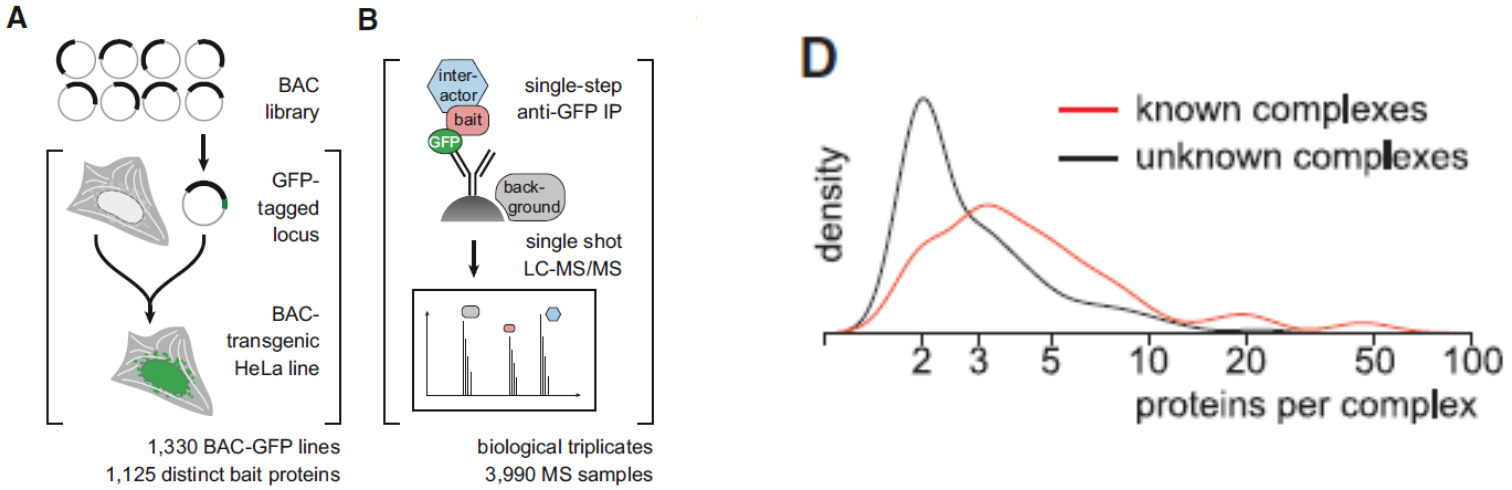
Cell

Article

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

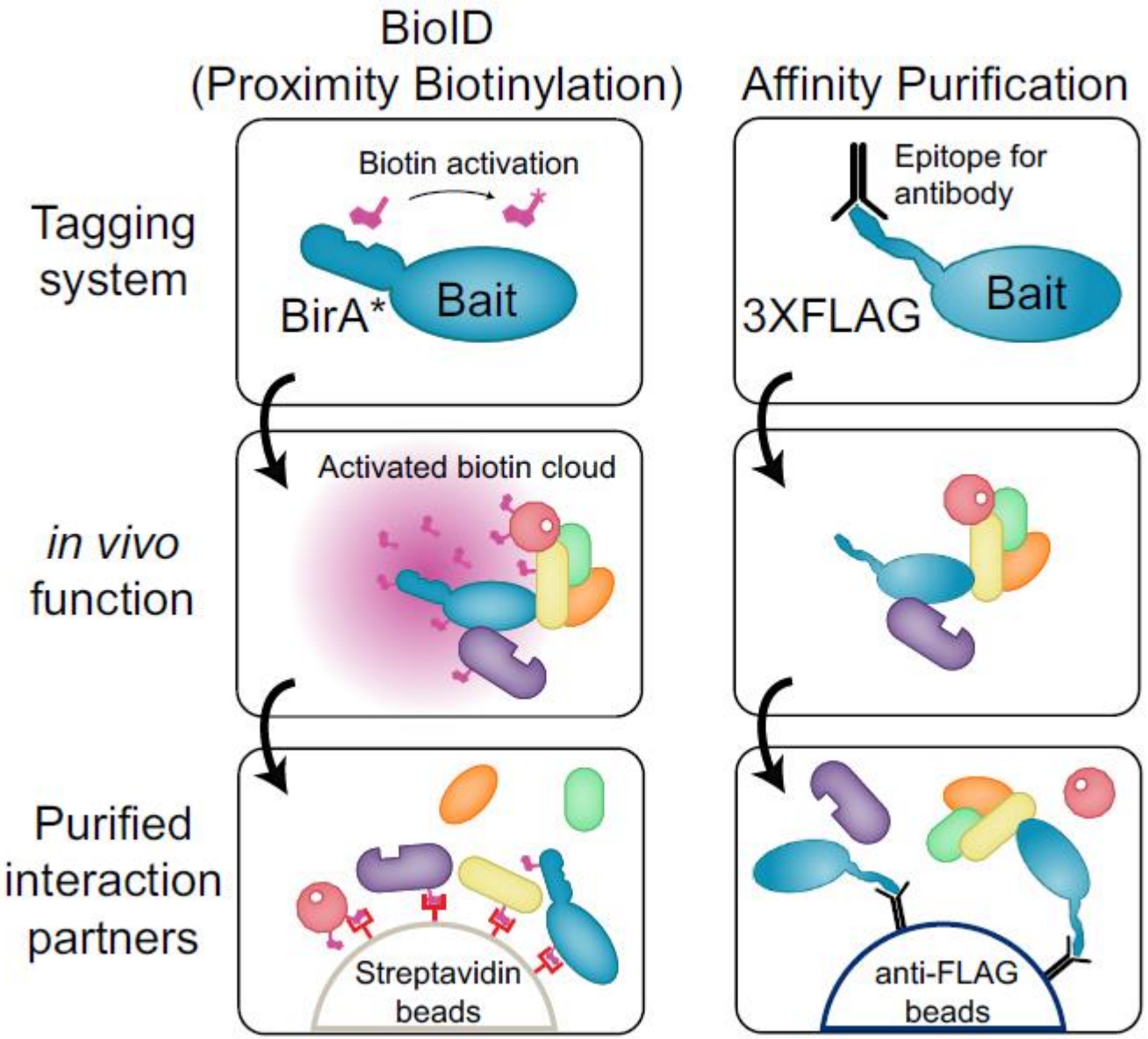
Marco Y. Hein,[1,6,8] Nina C. Hubner,[1,6,9] Ina Poser,[2] Jürgen Cox,[1] Nagarjuna Nagaraj,[1] Yusuke Toyoda,[2,10] Igor A. Gak,[3] Ina Weisswange,[4,5] Jörg Mansfeld,[3] Frank Buchholz,[2,4] Anthony A. Hyman,[2,7,*] and Matthias Mann[1,7,*]

- 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



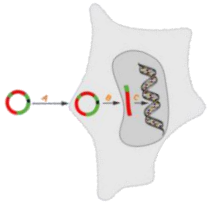
- Three quantitative dimensions measure specificities, stoichiometries, and abundances
- Stable complexes are rare but stand out by a signature of balanced stoichiometries
- Weak interactions dominate the network and have critical topological properties

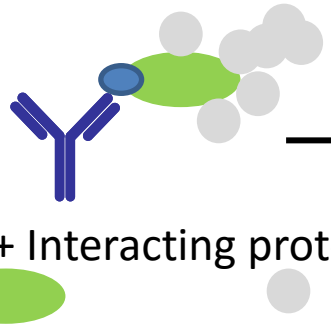
Hein MY, et al. Mann M, *Cell*. 2015 PMID: 26496610



Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes Lambert JP, et al., Gingras AC. J Proteomics. 2015 PMID: 25281560

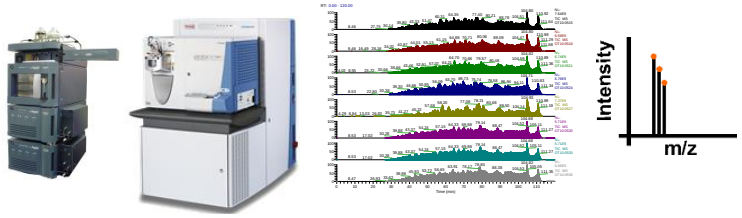
Transfect tagged "bait"





IP Bait + Interacting proteins

Multiple runs of "shot gun" LC-MS/MS



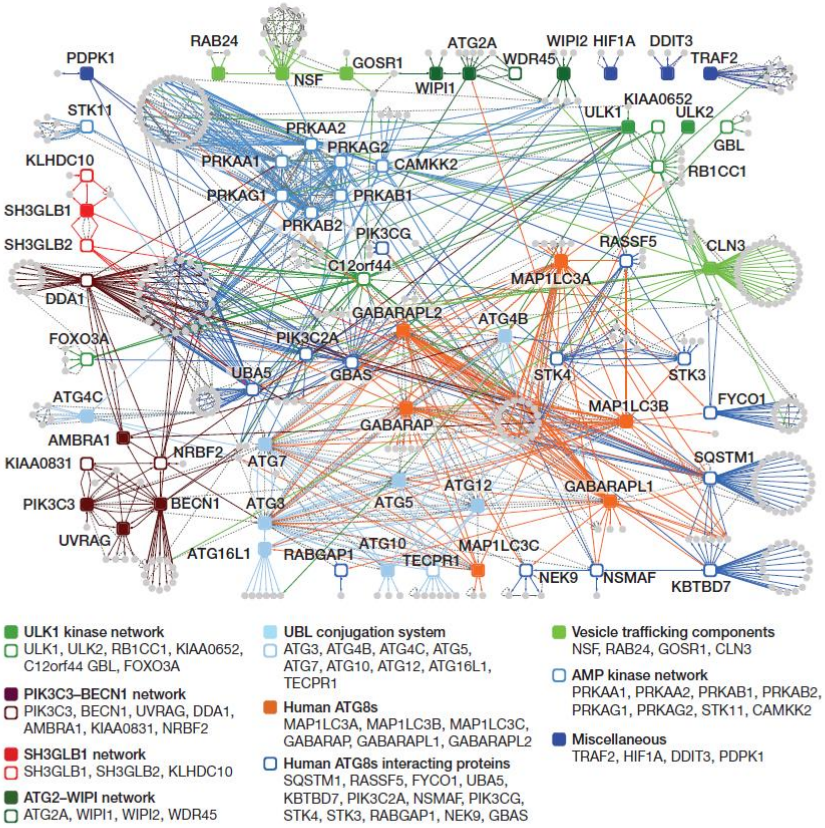
~65 bait proteins
LC-MS/MS identifies
2553 proteins

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]†

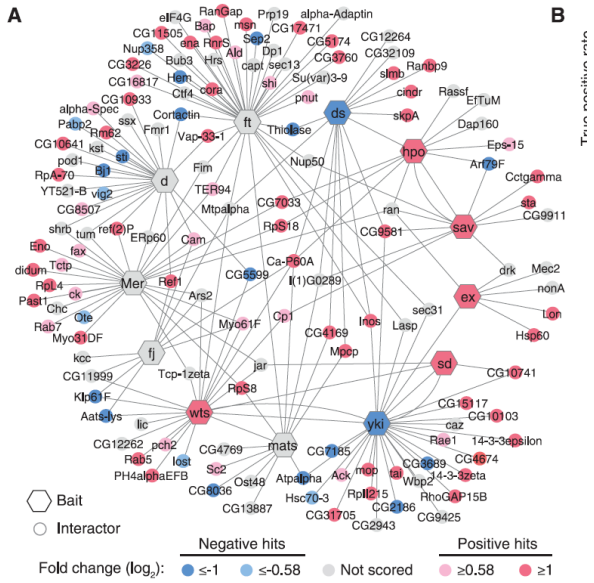
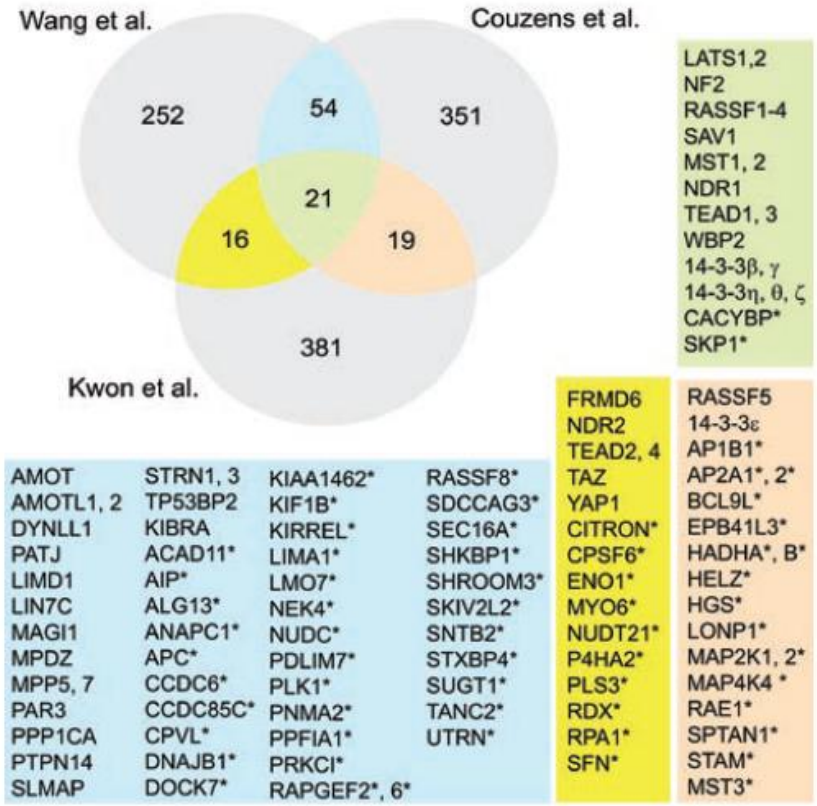


Fig. 2. Validation of Hippo-PPIN with functional RNAi screen and co-IP. (**A**) Distribution of Yki-reporter values for individual double-stranded RNAs (dsRNAs) in our focused RNAi screen. About 70% of genes are covered by two dsRNAs. (**B**) Recovery of Hippo pathway components from RNAi screen [fold-change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



Cell Research (2014) 24:137-138.

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

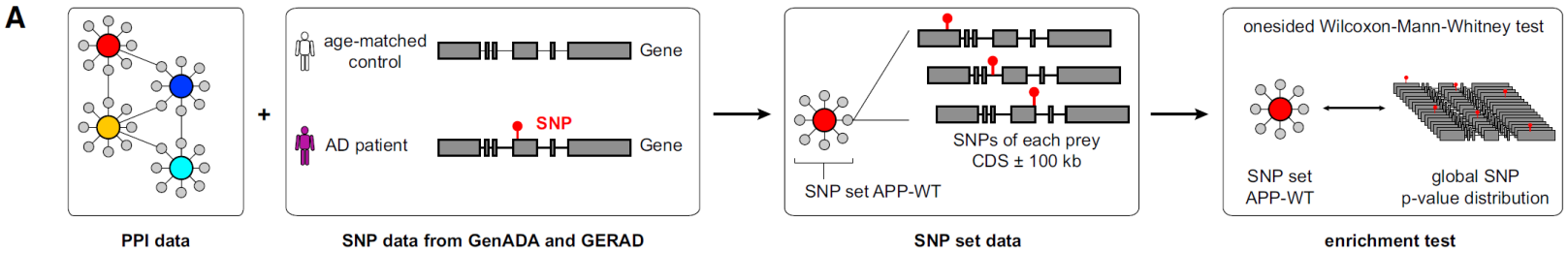
Kwon Y, *et al*. *Science* 2013; Couzens AL, *et al*. *Sci Signal* 2013; Wang W, *et al*. *Mol Cell Proteomics* 2014

OPEN ACCESS
CellPress

Cell Reports
Resource

Quantitative Interaction Proteomics of Neurodegenerative Disease Proteins

Fabian Hosp,[1,6] Hannes Vossfeldt,[2] Matthias Heinig,[1,3] Djordje Vasiljevic,[1] Anup Arumughan,[1] Emanuel Wyler,[1] the Genetic and Environmental Risk for Alzheimer's Disease (GERAD1) Consortium, Markus Landthaler,[1] Norbert Hubner,[1] Erich E. Wanker,[1] Lars Lannfelt,[4] Martin Ingelsson,[4] Maciej Lalowski,[5] Aaron Voigt,[2] and Matthias Selbach[1,*]
[1]Max Delbrück Center for Molecular Medicine, Robert-Rössle-Straße 10, 13092 Berlin, Germany



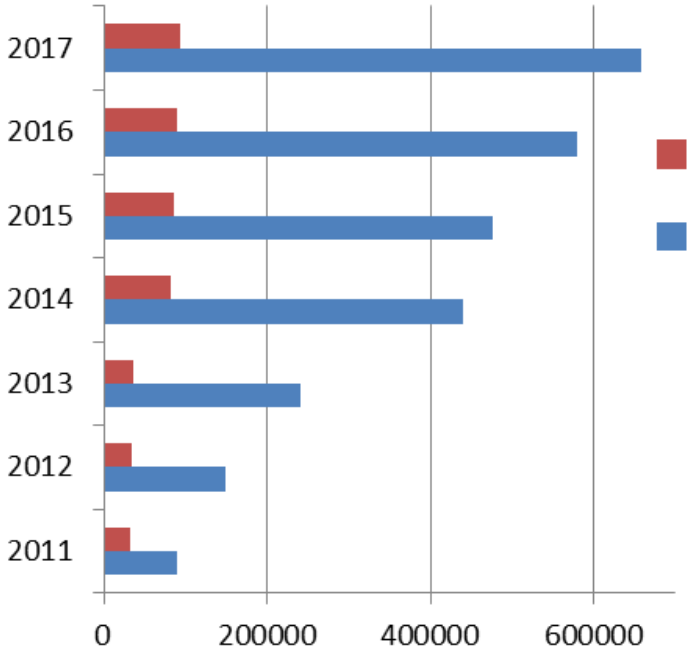
Protein-Protein Interaction Databases

http://thebiogrid.org/

BioGRID 3.4 home help wiki

Welcome to the Biological General Repository for Interaction Datasets

version **3.4.132** = **55,519** publications .
980,467 protein and genetic interactions from major model organism species.



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

EMBL-EBI

IntAct

2017

Data Content

- Publications: **14451**
- Interactions: **658369**
- Interactors: **94358**

+ 79,490 interactions
+ 4,433 proteins

2016

Data Content

- Publications: **14010**
- Interactions: **578879**
- Interactors: **89925**

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

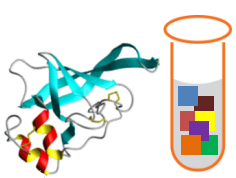
However:

- Interaction data is largely static

Next Step:

- How do protein interaction networks change over time?

Typical work flow for LC-MS "shotgun proteomics"



Protein mixture

Enz. Digest

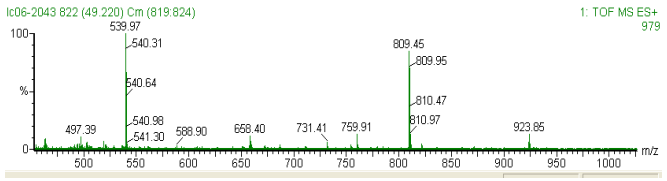
Peptide Pool

LC

n-UPLC

LTQ-Orbitrap MS

MS



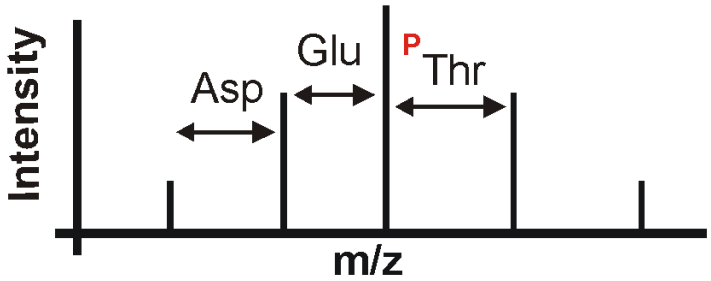
peptide

peptide

isolate & fragment

peptide

MS/MS

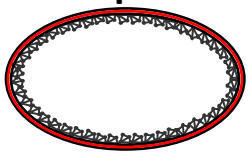


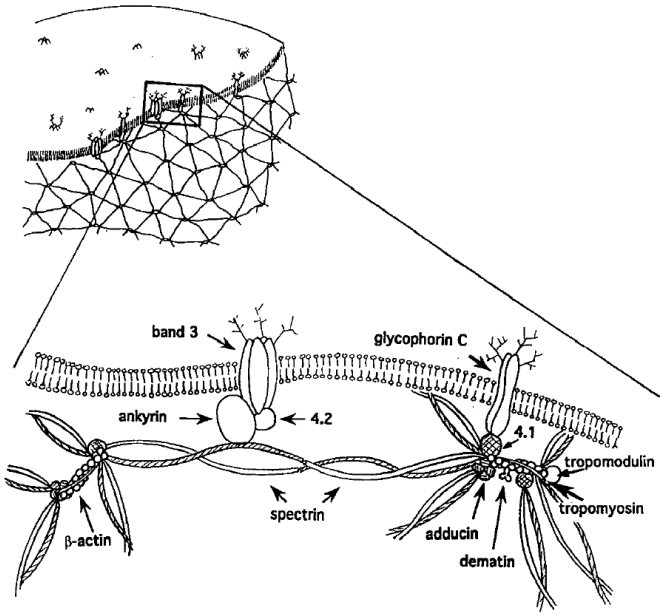
peptide fragments

MS Data is <u>not</u> inherently quantitative, *but …*

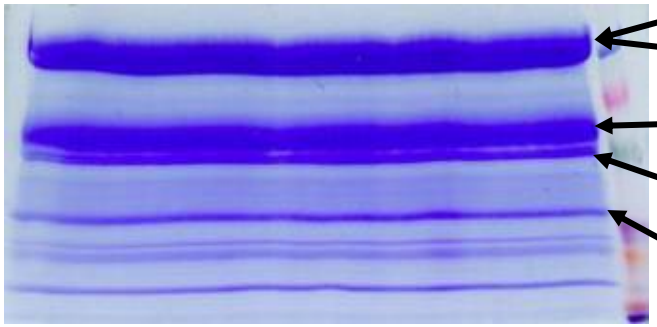
Red Blood Cell

RBC membrane:
a native multi-protein complex





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



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

# peptides (unique)	
352 (291)	Spectrin alpha chain, erythrocyte OS=Homo sapiens GN=SPTA1 PE=1 SV=5
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
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
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
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

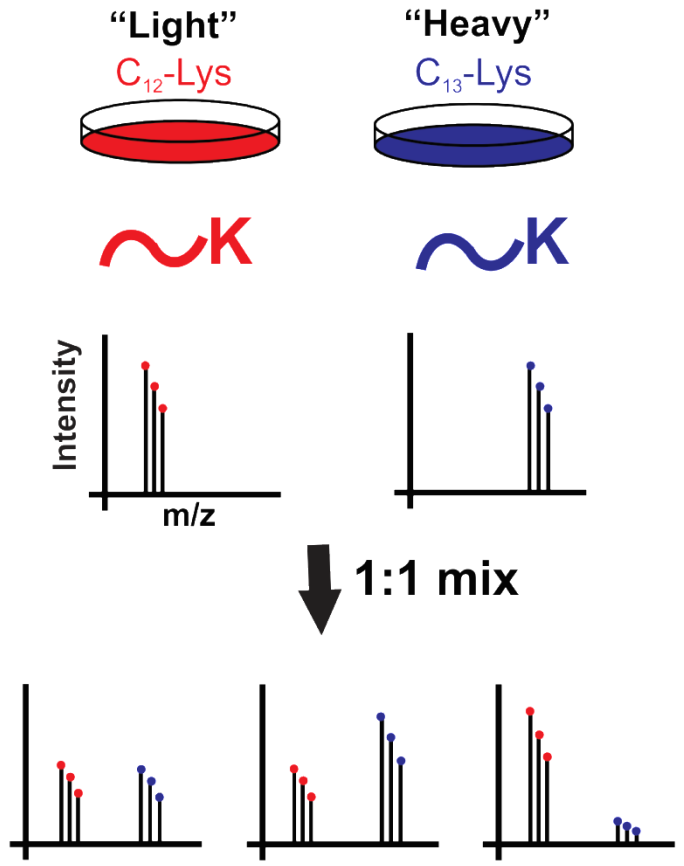
Rinehart et al., *unpublished*

Quantitative Proteomics

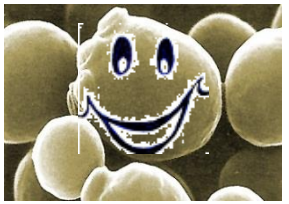
S.I.L.A.C. - Stable isotope labeling with amino acids in cell culture

-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



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.**

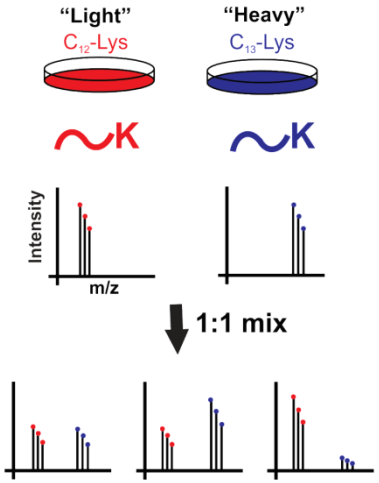
2008

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

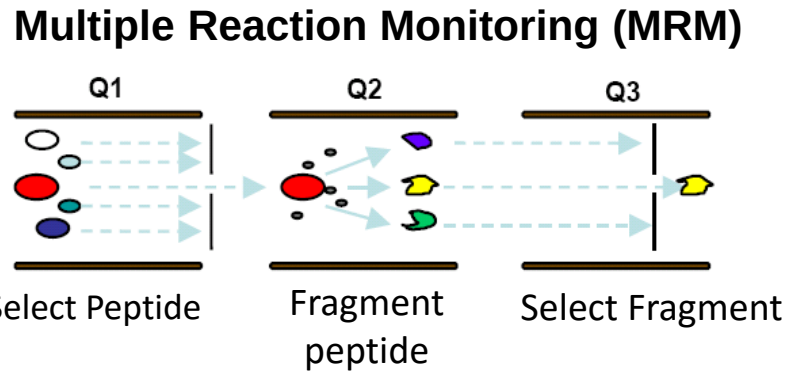
-Ong SE et al. *Molecular & Cell Proteomics* 2002.

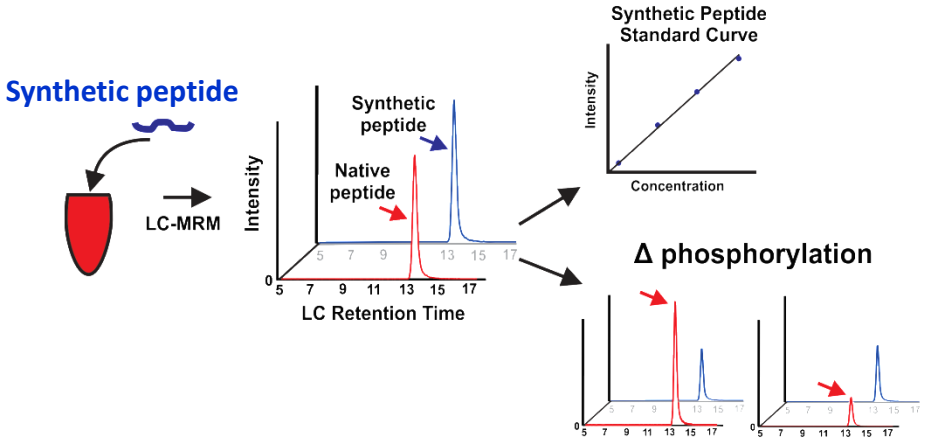


2009

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

➲ **Towards proteome wide targeted proteomics.**





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

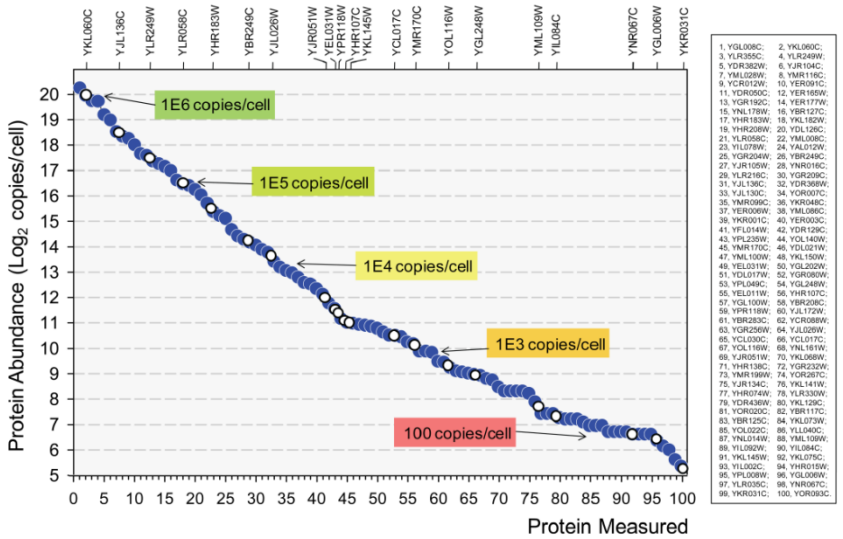
	Number of ORFs	TAP	GFP	nanoLC–MS
Total yeast ORFs	6,608	4,251	4,154	4,399
Characterized yeast ORFs	4,666	3,629	3,581	3,824
Uncharacterized yeast ORFs	1,128	581	539	572
Dubious yeast ORFs	814	26 (3%)	23 (3%)	3 (<1%)
Not present in ORF database		15	11	0

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.

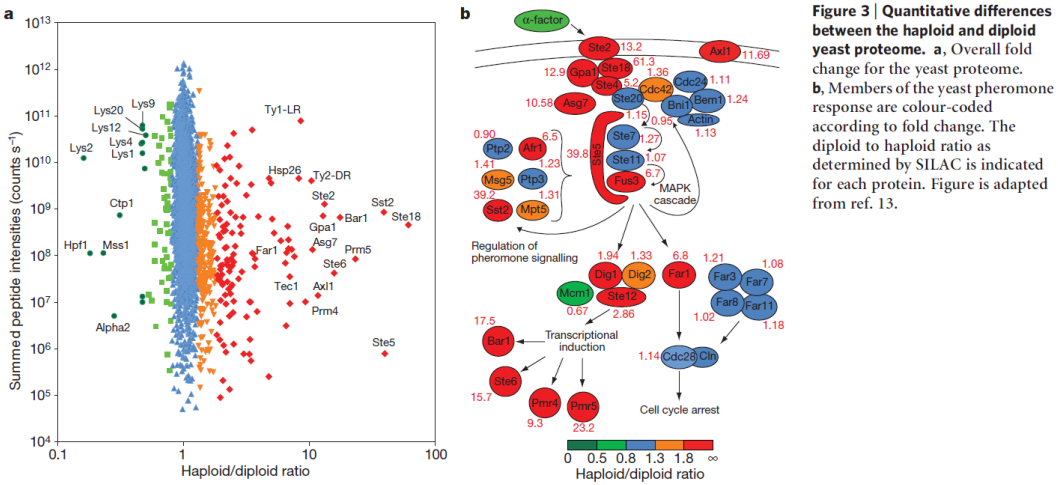


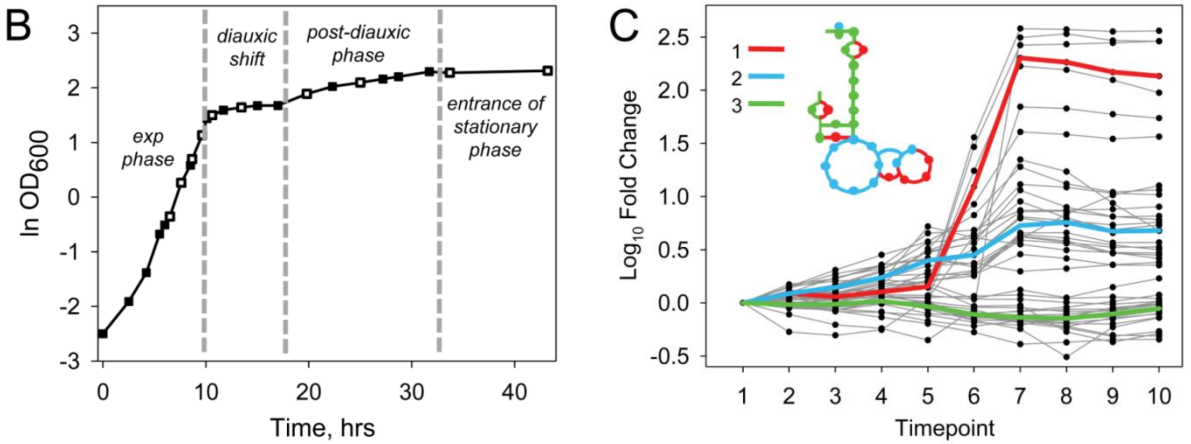
Figure 3 | Quantitative differences between the haploid and diploid yeast proteome. **a**, Overall fold change for the yeast proteome. **b**, Members of the yeast pheromone response are colour-coded according to fold change. The diploid to haploid ratio as determined by SILAC is indicated for each protein.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

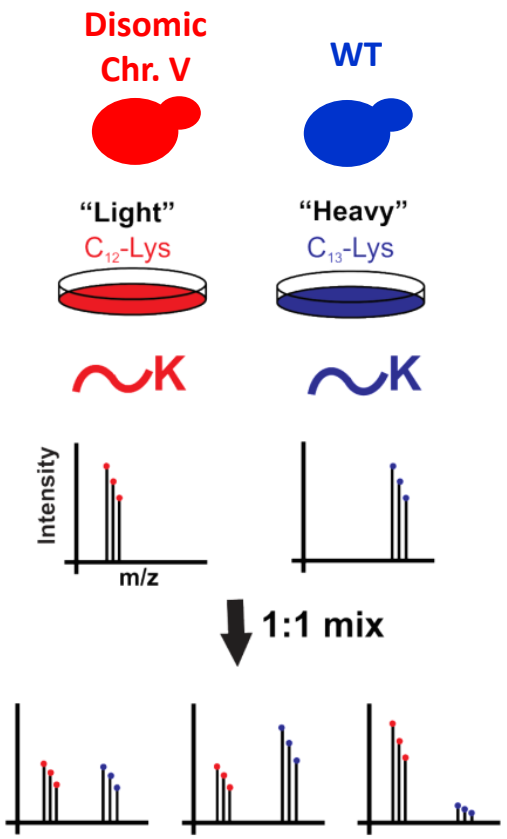
Network expression dynamics

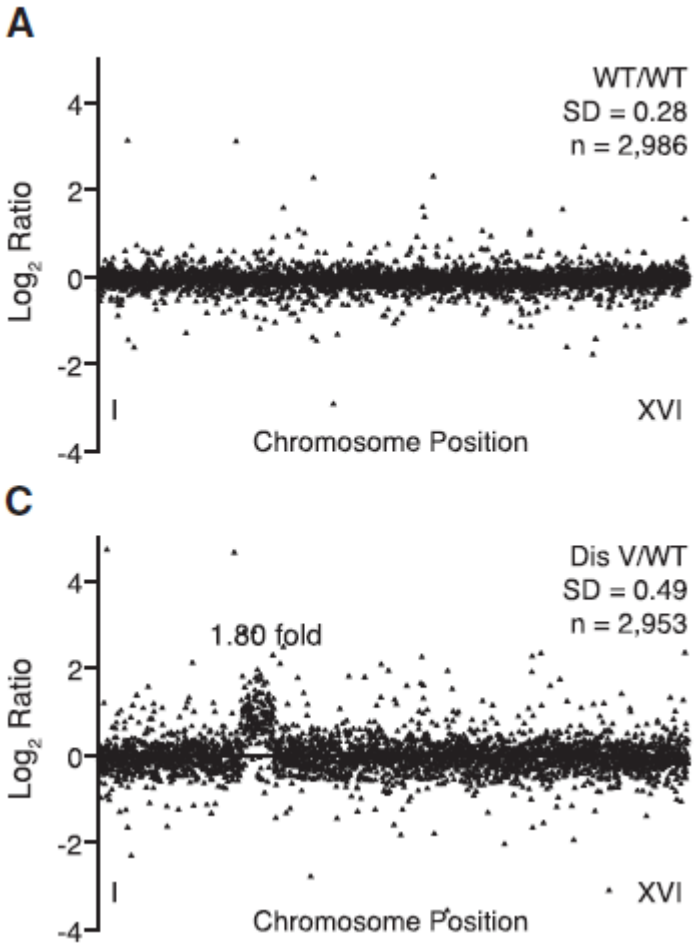


Identification of Aneuploidy-Tolerating Mutations

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,*]





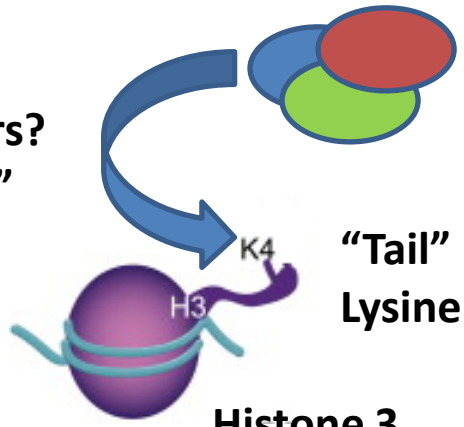
Resource

Cell

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

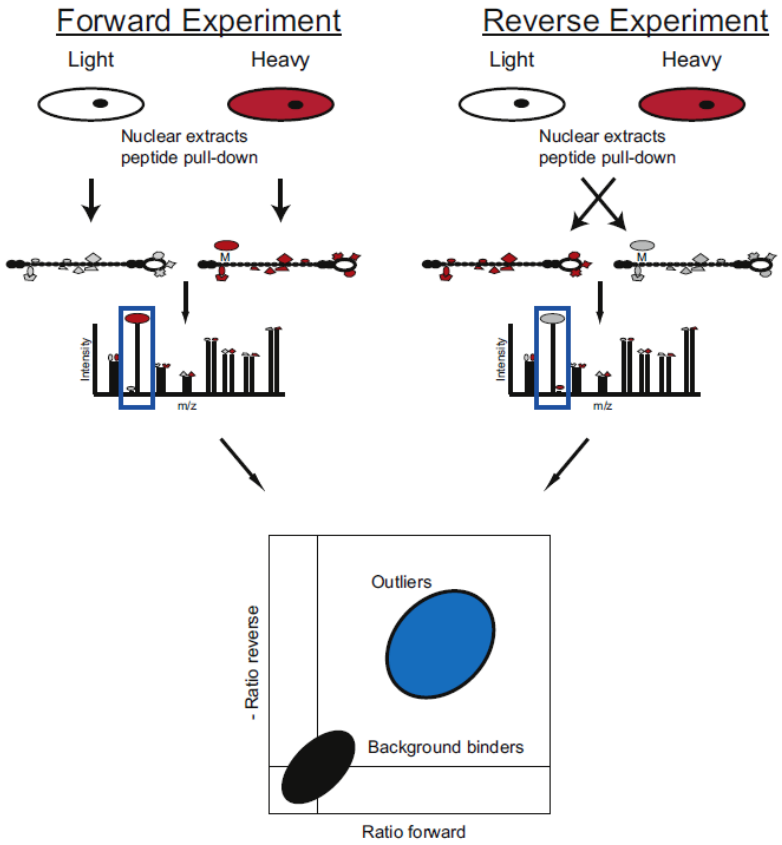
Michiel Vermeulen,[1,6,7,*] H. Christian Eberl,[1,6] Filomena Matarese,[2,6] Hendrik Marks,[2] Sergei Denissov,[2] Falk Butter,[1] Kenneth K. Lee,[3] Jesper V. Olsen,[1,5] Anthony A. Hyman,[4] Henk G. Stunnenberg,[2,*] and Matthias Mann[1,*]

Protein Regulators? "Readers"



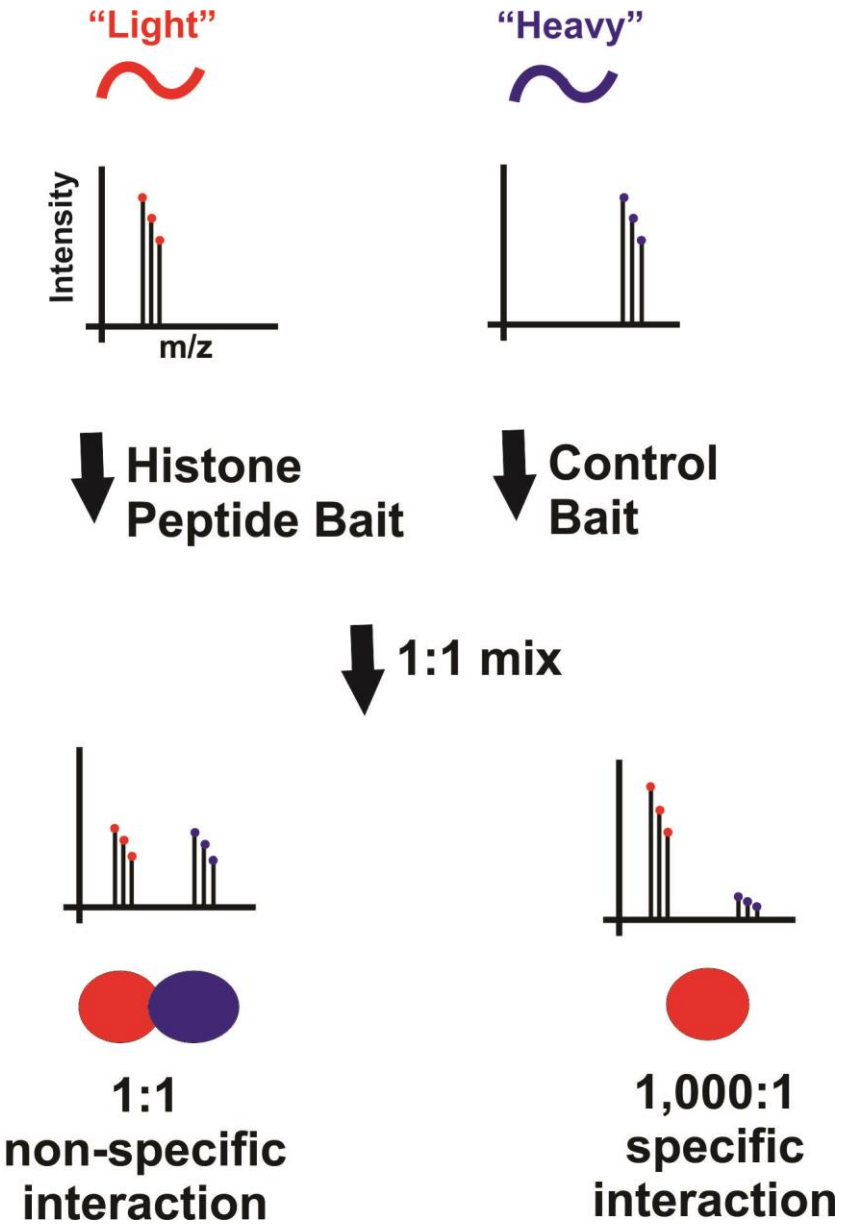
"Tail" Lysine 4

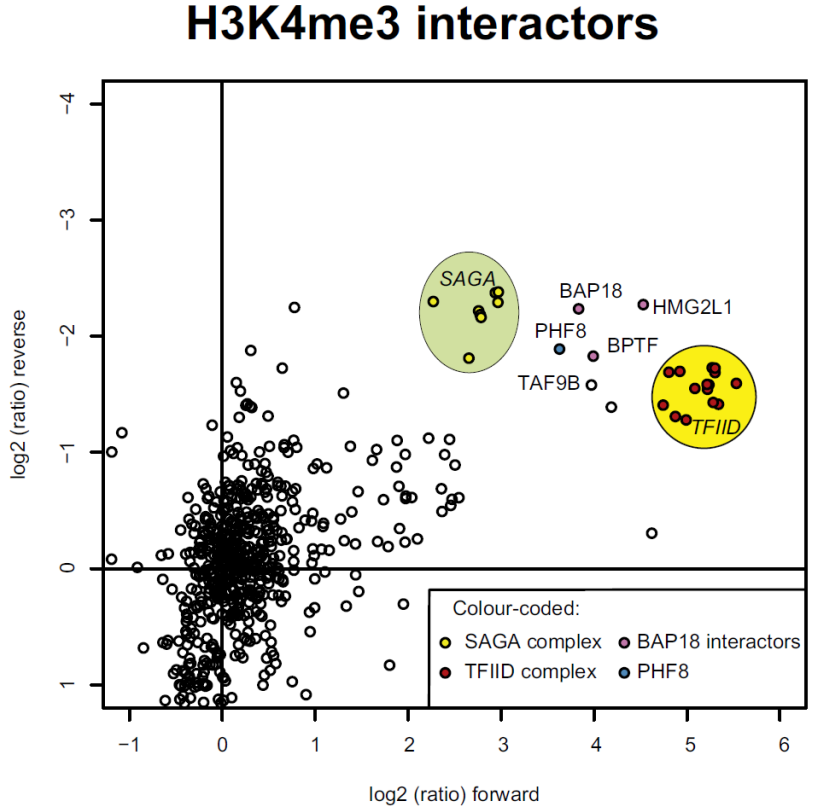
Histone 3

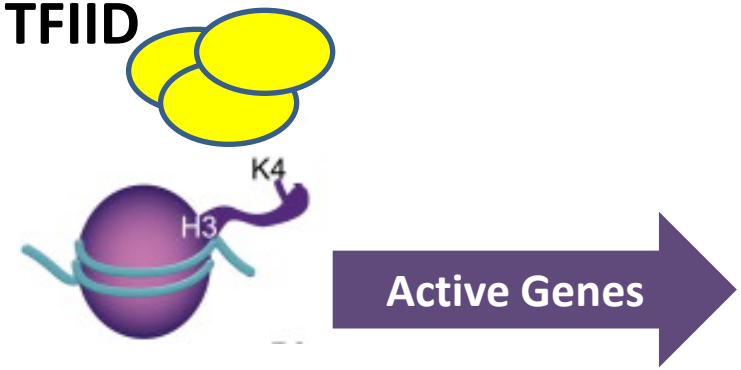


Vermeulen et al., **Cell** 2010

The major lysine methylation sites on the N terminus of histone H3 and histone H4 with a clearly defined biological function are H3K4me3, H3K9me3, H3K27me3, H3K36me3, and H4K20me3, which are associated with different functional states of chromatin. H3K4me3 is almost exclusively found on promoter regions of actively transcribed genes while H3K36me3 is linked to transcription elongation. H3K9me3, H3K27me3, and H4K20me3 are generally found on silent heterochromatic regions of the genome. Part of the functional distinction between these methylation sites relates to the proteins interacting with them. A number of these "chromatin readers" for various histone methyl lysine sites have already been identified and characterized (Kouzarides, 2007; Shilatifard, 2006; Taverna et al., 2007),

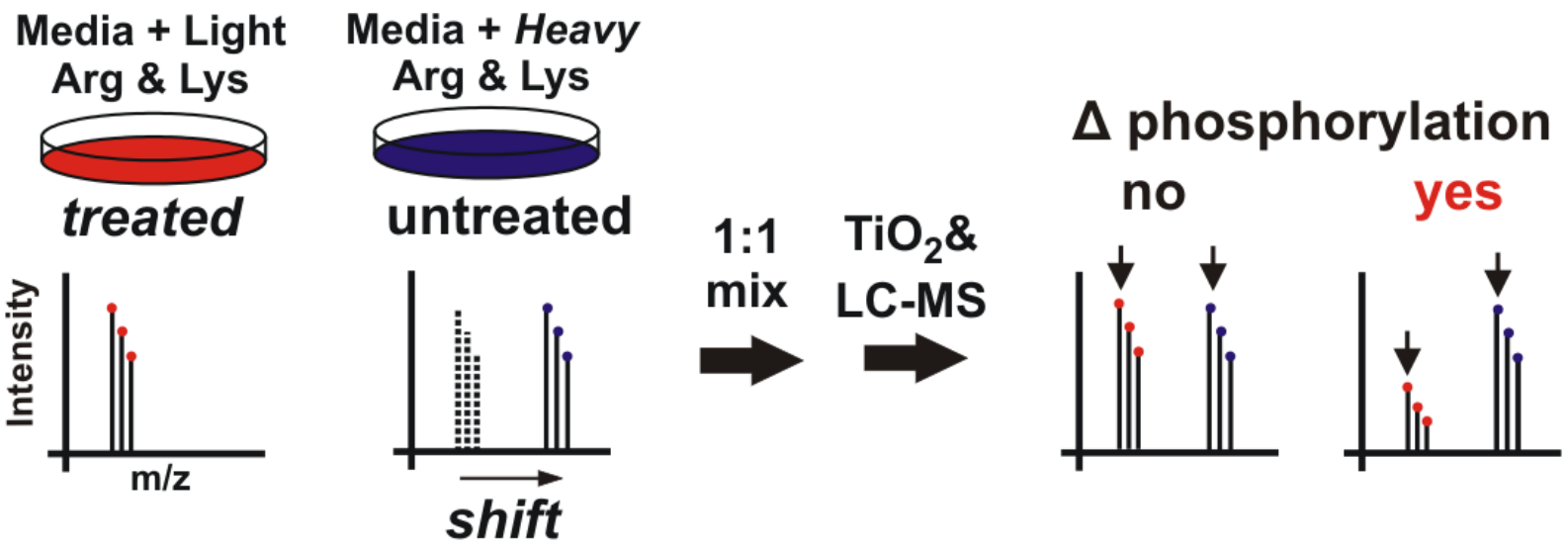




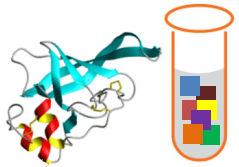


Vermeulen et al., **Cell** 2010

A SILAC approach to study protein phosphorylation dynamics



Major technological advances in mass spectrometers and phosphopeptide enrichment



Protein mixture

Digest

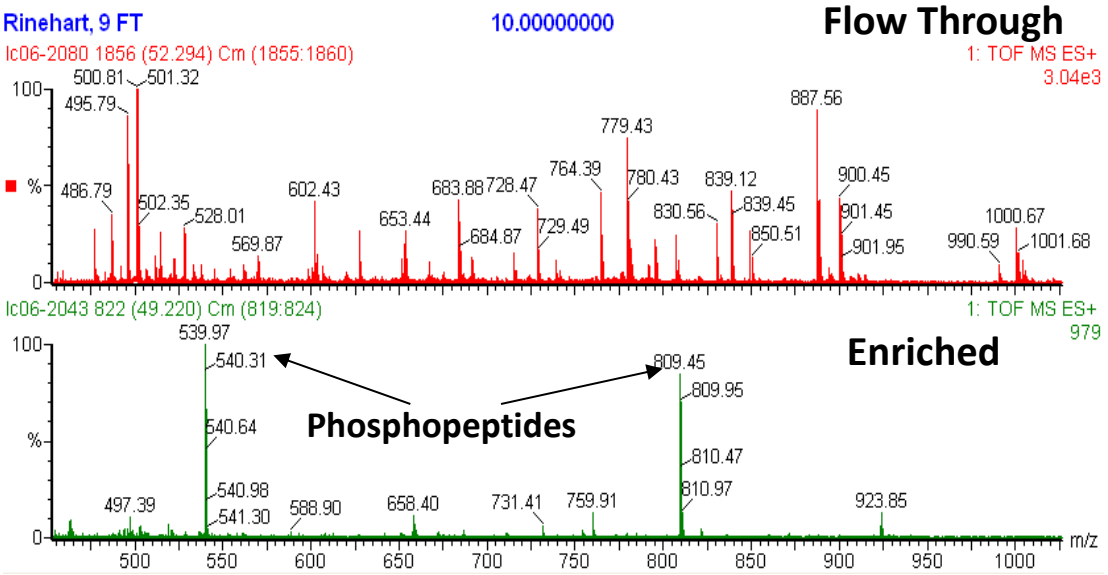
Peptides

TiO_2 Enrich

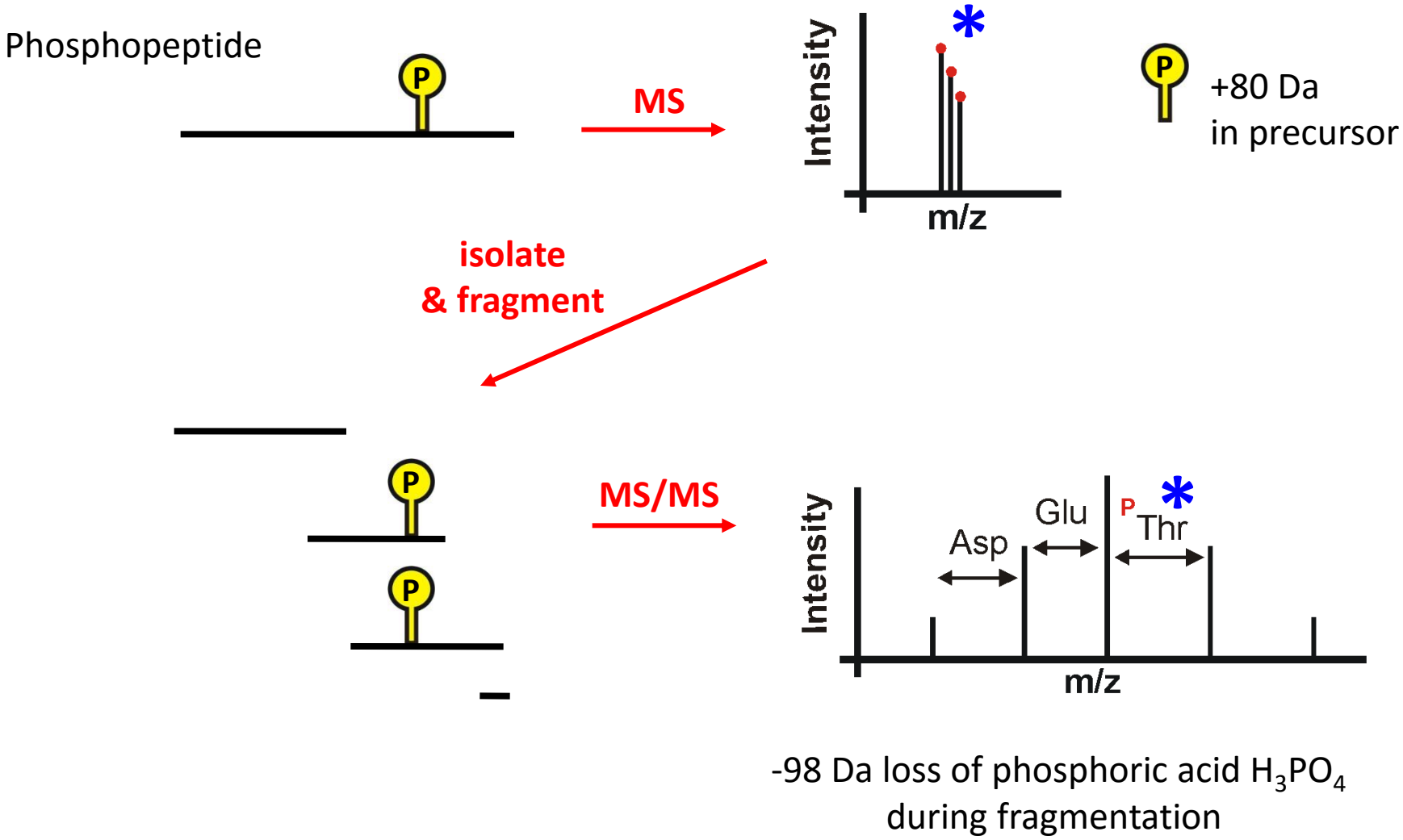
Phosphopeptides

MS

TiO_2 Enrichment

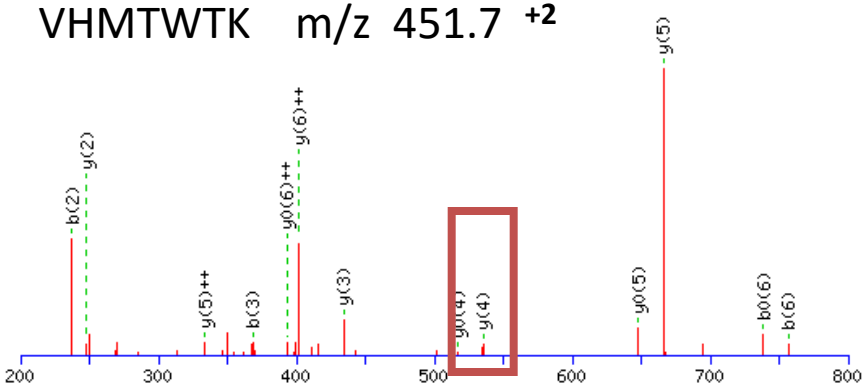


*Phosphopeptide signatures in MS



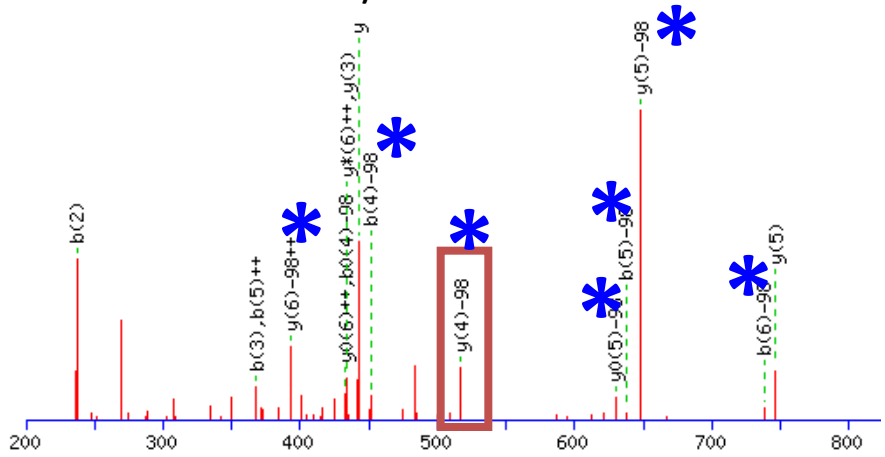
V	
H	803.3869
M	666.3280
T	535.2875
W	434.2398
T	248.1605
K	147.1128

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^{P}WTK m/z 491.7 +2



P +80 Da in precursor

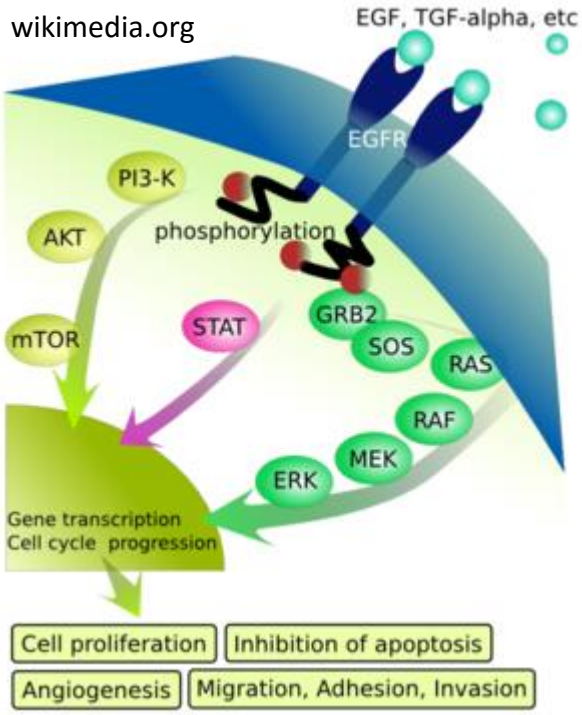
* - H_3PO_4, 98 Da

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

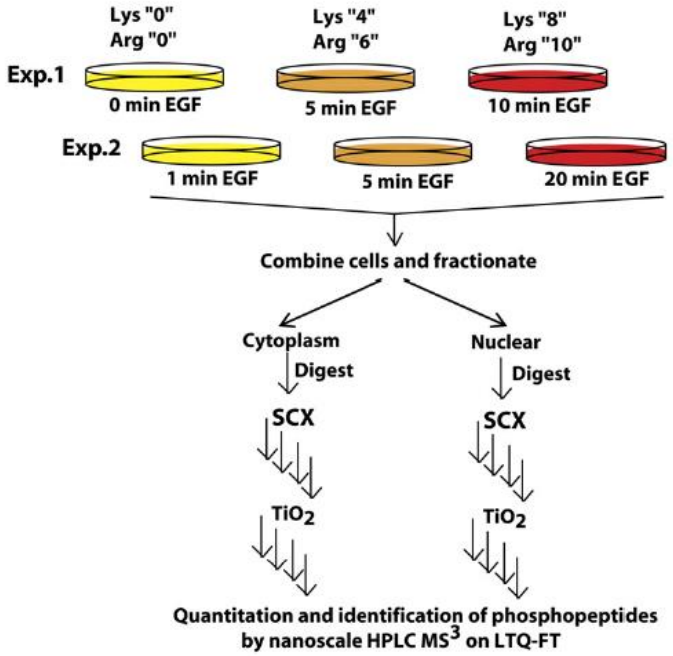
Quantitative Proteomics Reveals Dynamics in Signaling Networks

Phosphorylation dynamics after EGF stimulation

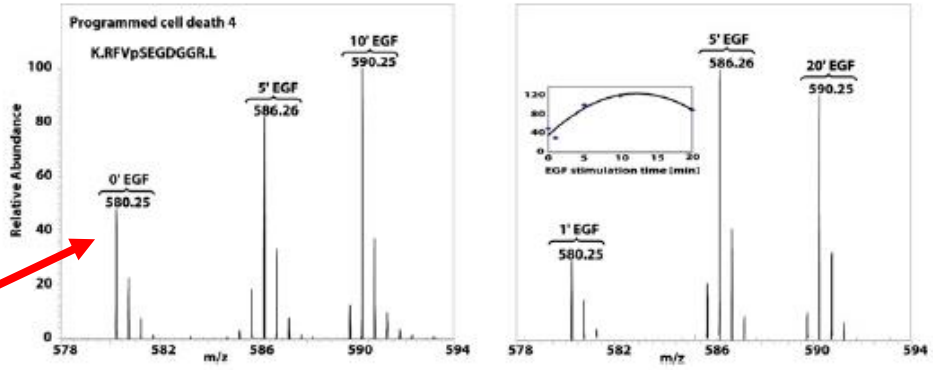
wikimedia.org



SILAC approach enables dynamic analysis



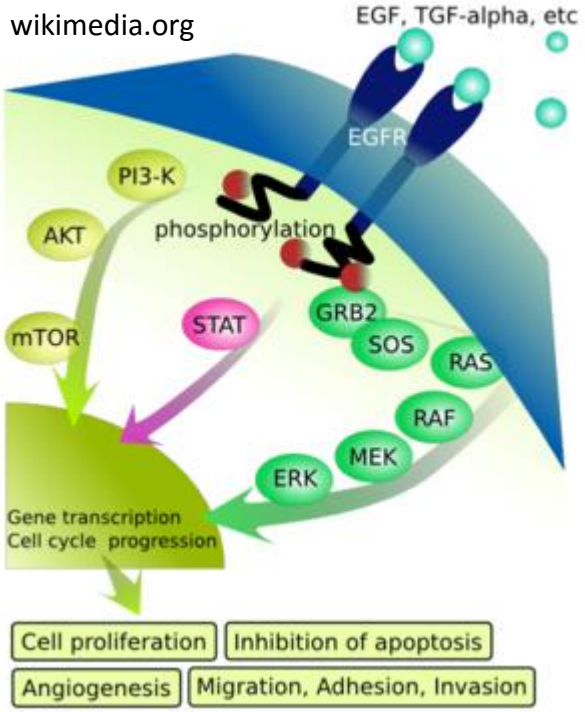
MS spectra triplets

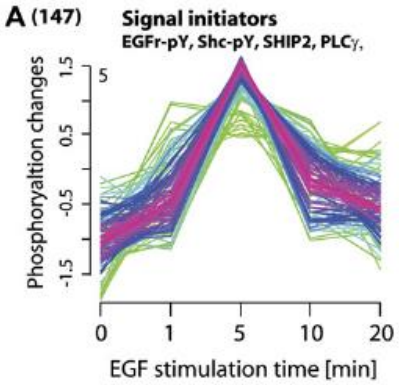


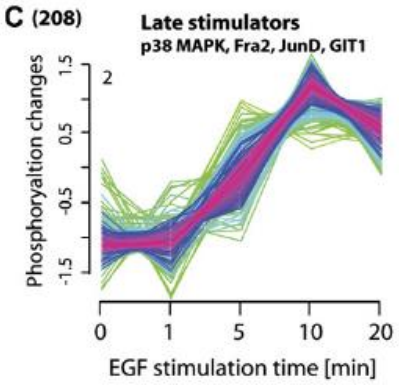
Olsen, et al. *Cell*, 2006

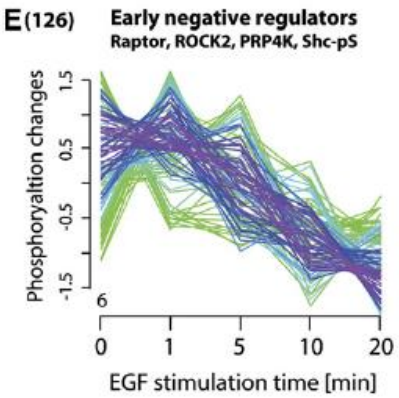
Phosphorylation dynamics after EGF stimulation

wikimedia.org









Olsen, et al. *Cell*, 2006

Proteomics & Protein-Protein Interactions

Overview

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

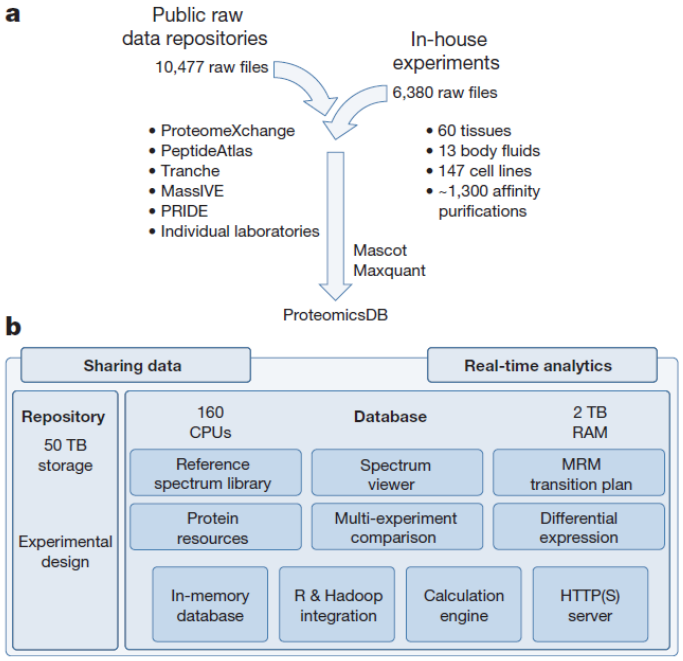
ARTICLE

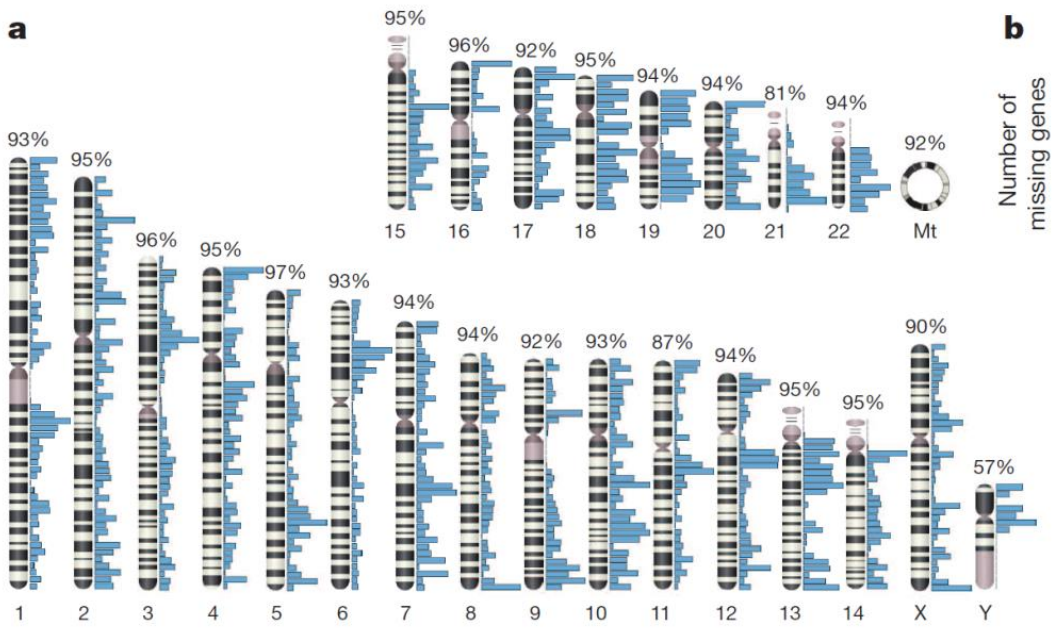
doi:10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm[1,2]*, Judith Schlegl[2]*, Hannes Hahne[1]*, Amin Moghaddas Gholami[1]*, Marcus Lieberenz[2], Mikhail M. Savitski[3], Emanuel Ziegler[2], Lars Butzmann[2], Siegfried Gessulat[2], Harald Marx[1], Toby Mathieson[3], Simone Lemeer[1], Karsten Schnatbaum[4], Ulf Reimer[4], Holger Wenschuh[4], Martin Mollenhauer[5], Julia Slotta-Huspenina[5], Joos-Hendrik Boese[2], Marcus Bantscheff[3], Anja Gerstmair[2], Franz Faerber[2] & 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; Wilhelm & Bernhard Kuster et al., **PMID: 24870543**

ARTICLE

doi:10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm[1,2]*, Judith Schlegl[2]*, Hannes Hahne[1]*, Amin Moghaddas Gholami[1]*, Marcus Lieberenz[2], Mikhail M. Savitski[3], Emanuel Ziegler[2], Lars Butzmann[2], Siegfried Gessulat[2], Harald Marx[1], Toby Mathieson[3], Simone Lemeer[1], Karsten Schnatbaum[4], Ulf Reimer[4], Holger Wenschuh[4], Martin Mollenhauer[5], Julia Slotta-Huspenina[5], Joos-Hendrik Boese[2], Marcus Bantscheff[3], Anja Gerstmair[2], Franz Faerber[2] & 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

Mass-spectrometry-based draft of the human proteome; Wilhelm & Bernhard Kuster et al., **PMID: 24870543**

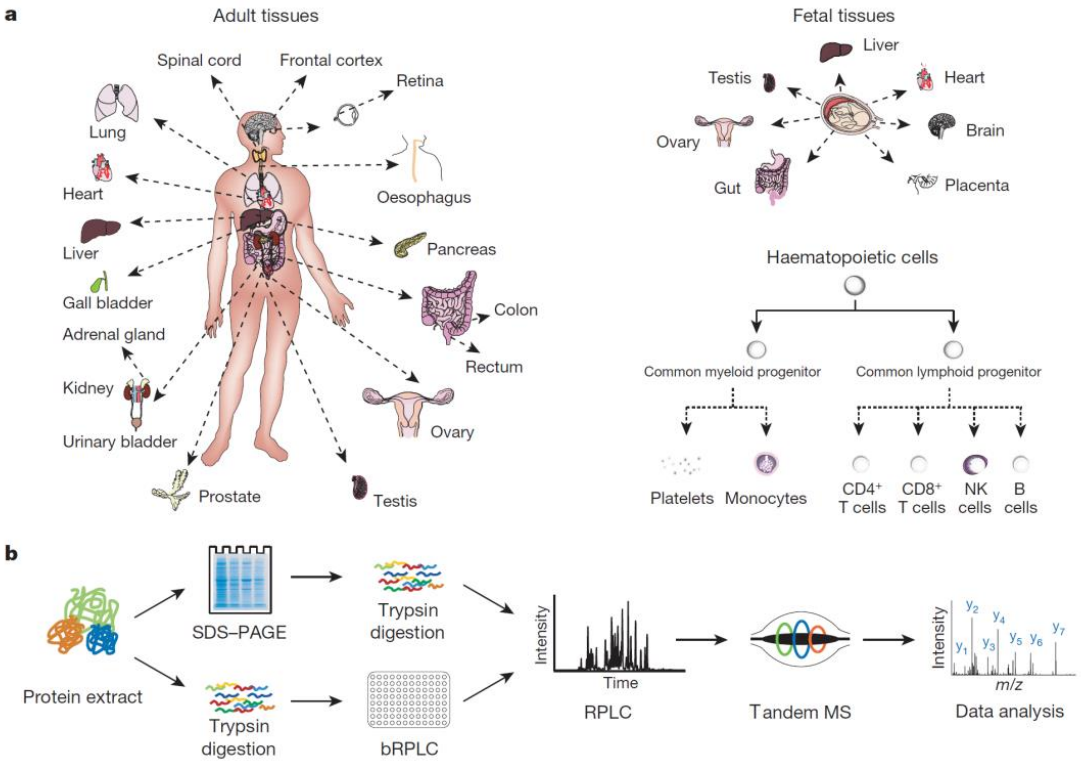
ARTICLE

doi:10.1038/nature13302

A draft map of the human proteome

Min-Sik Kim[1,2], Sneha M. Pinto[3], Derese Getnet[1,4], Raja Sekhar Nirujogi[3], Srikanth S. Manda[3], Raghothama Chaerkady[1,2], Anil K. Madugundu[3], Dhanashree S. Kelkar[3], Ruth Isserlin[5], Shobhit Jain[5], Joji K. Thomas[3], Babylakshmi Muthusamy[3], Pamela Leal-Rojas[1,6], Praveen Kumar[3], Nandini A. Sahasrabuddhe[3], Lavanya Balakrishnan[3], Jayshree Advani[3], Bijesh George[3], Santosh Renuse[3], Lakshmi Dhevi N. Selvan[3], Arun H. Patil[3], Vishalakshi Nanjappa[3], Aneesha Radhakrishnan[3], Samarjeet Prasad[1],

- 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.



A draft map of the human proteome; Kim & Akhilesh Pandey et al., PMID: 24870542

Proteomics Databases: Peptide depositories

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.

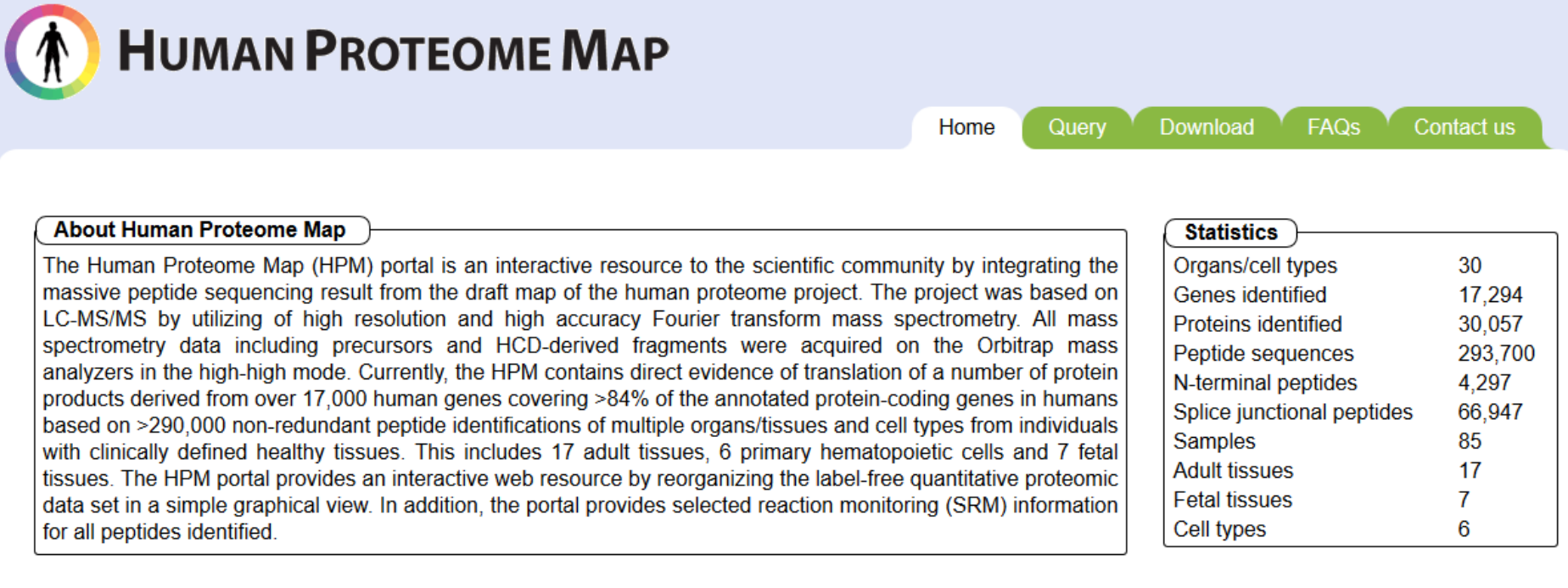
The GPMDB Project

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



ARTICLE

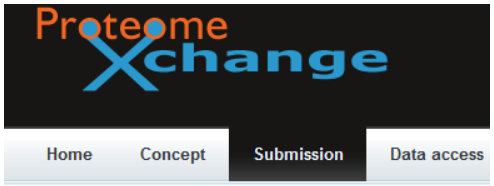
doi:10.1038/nature13302

A draft map of the human proteome

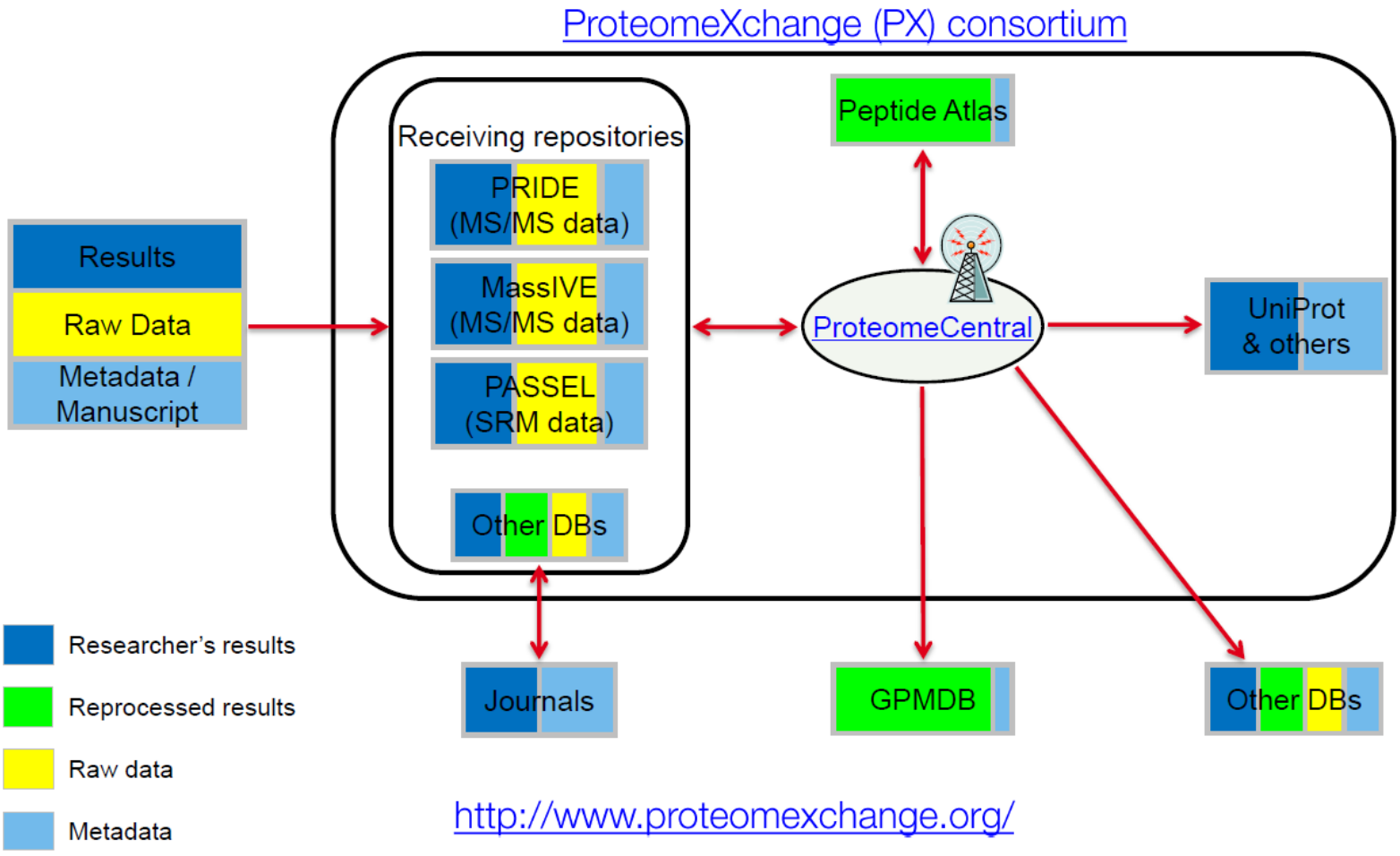
Min-Sik Kim[1,2], Sneha M. Pinto[3], Derese Getnet[1,4], Raja Sekhar Nirujogi[3], Srikanth S. Manda[3], Raghothama Chaerkady[1,2], Anil K. Madugundu[3], Dhanashree S. Kelkar[3], Ruth Isserlin[5], Shobhit Jain[5], Joji K. Thomas[3], Babylakshmi Muthusamy[3], Pamela Leal-Rojas[1,6], Praveen Kumar[3], Nandini A. Sahasrabuddhe[3], Lavanya Balakrishnan[3], Jayshree Advani[3], Bijesh George[3], Santosh Renuse[3], Lakshmi Dhevi N. Selvan[3], Arun H. Patil[3], Vishalakshi Nanjappa[3], Aneesha Radhakrishnan[3], Samarjeet Prasad[1],

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

Proteomics Databases: Integrated Resources



http://www.proteomexchange.org/



Slide modified from "Computational Mass Spectrometry-Based Proteomics 6th Maxquant Summer School" 21-25 July 2014
Emanuele Alpi, UniProt and PRIDE Development

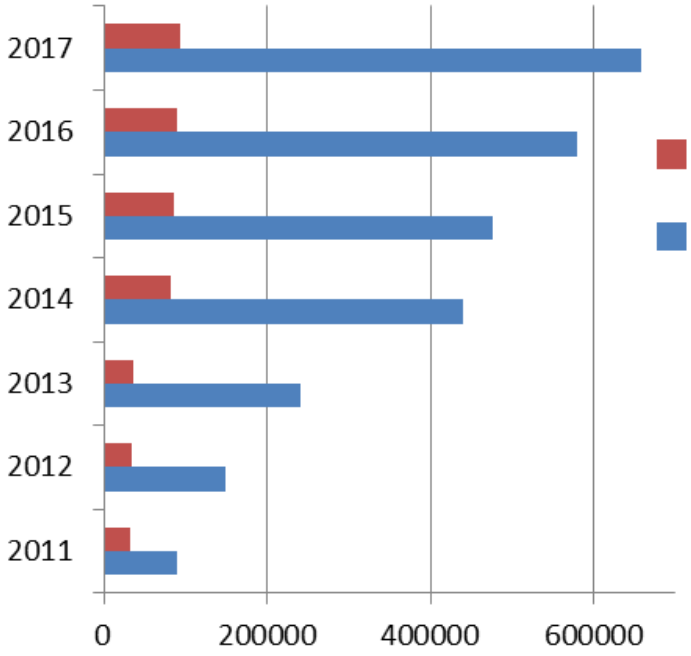
Protein-Protein Interaction Databases

http://thebiogrid.org/

BioGRID 3.4 home help wiki

Welcome to the Biological General Repository for Interaction Datasets

version **3.4.132** = **55,519** publications .
980,467 protein and genetic interactions from major model organism species.



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

EMBL-EBI

IntAct

2017

Data Content

- Publications: **14451**
- Interactions: **658369**
- Interactors: **94358**

+ 79,490 interactions
+ 4,433 proteins

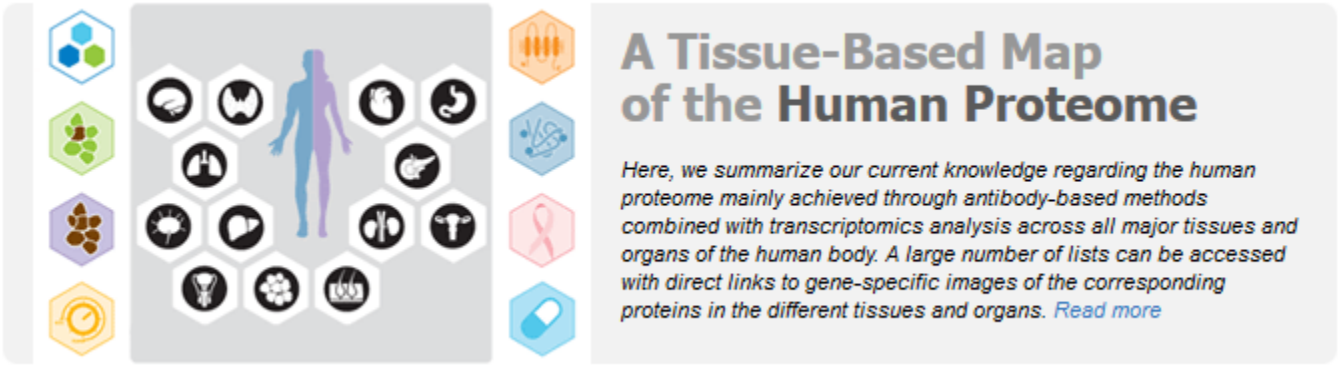
2016

Data Content

- Publications: **14010**
- Interactions: **578879**
- Interactors: **89925**

Proteomics Databases: Integrated Resources Beyond Mass Spectrometry

http://www.proteinatlas.org/



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