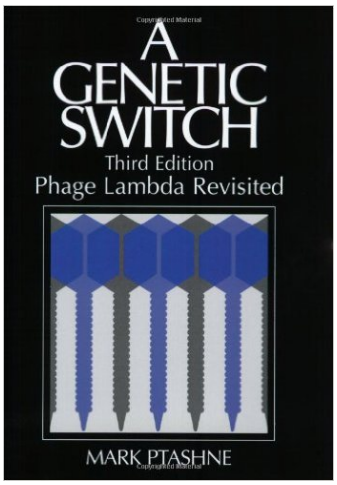
Challenges in Integrative Modeling of Gene Regulation, Protein Signaling, and Metabolism

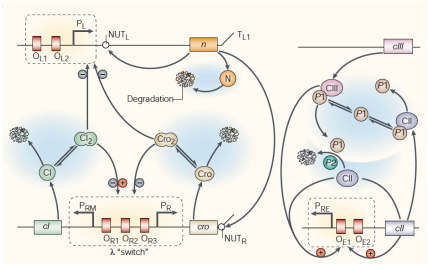
Ernest Fraenkel
Department of Biological Engineering
Massachusetts Institute of Technology

Dynamic Modeling

STEP 0: Define the system

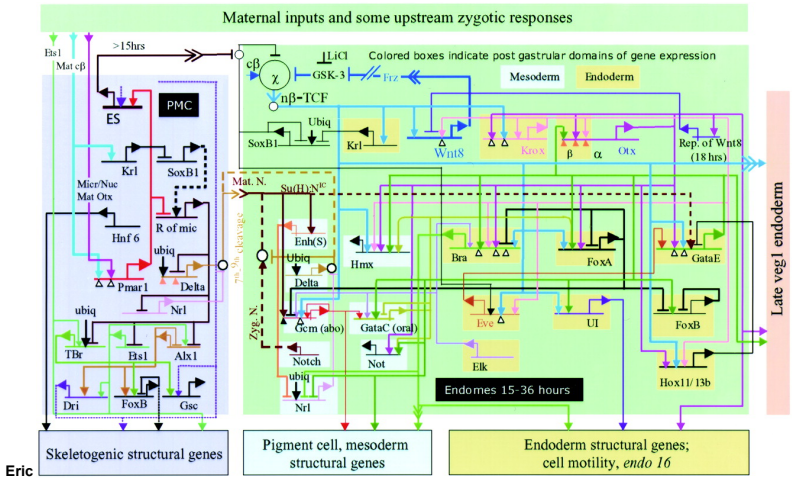
The Genetic Switch (Then)





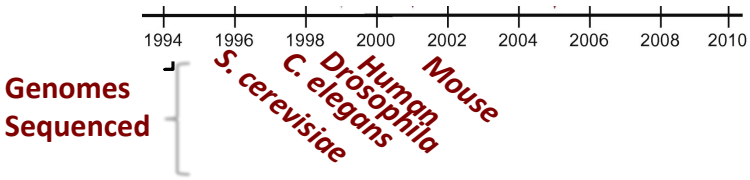
Nature Reviews Genetics **4**, 471-477 (June 2003) | doi:10.1038/nrg1089

Even Hand-built Models are Getting Complex!



Eric

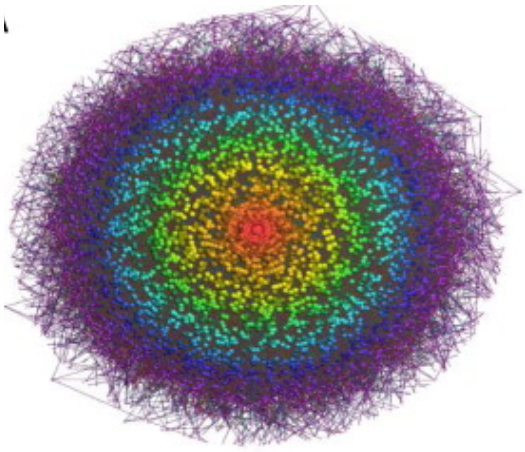
PNAS



Source: Snyder andGallagher
FEBS Letters 2009; 583:3895-3899 (DOI:10.1016/j.febslet.2009.11.011)

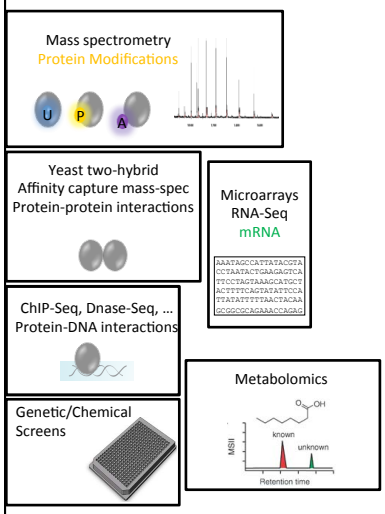
5

Yikes!



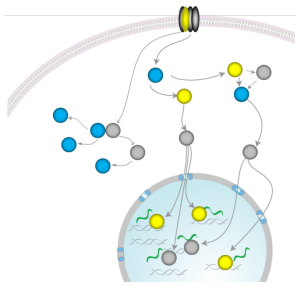
Conditional MI network of miR modulators
248,000 interactions
http //www.sciencedirect.com/science/article/pii/S0092867411011524

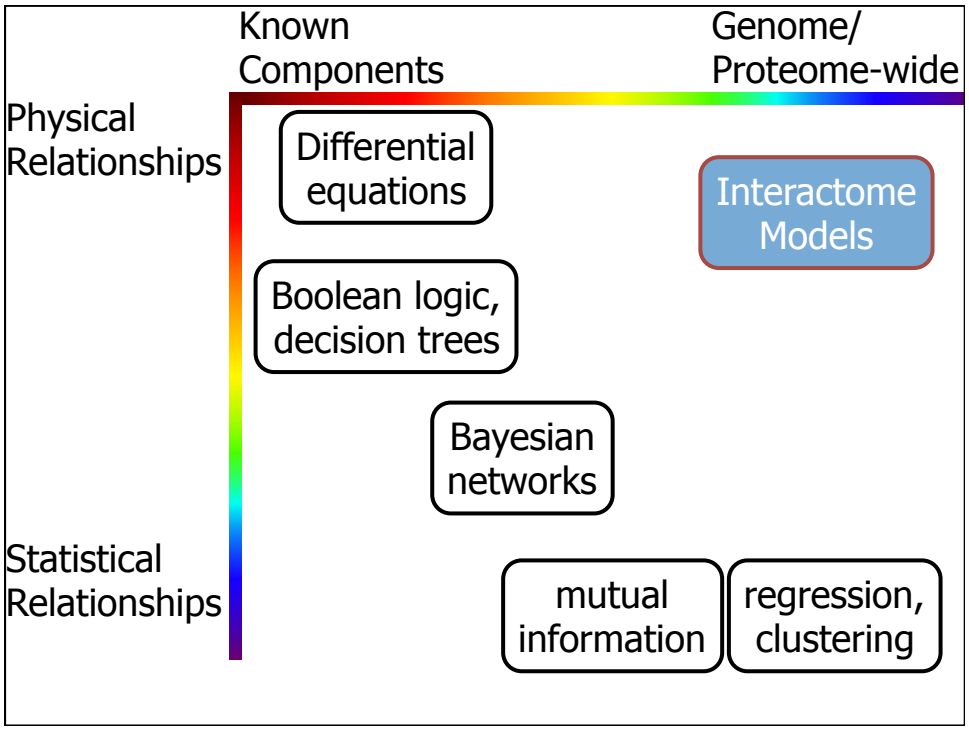
Multi-omics

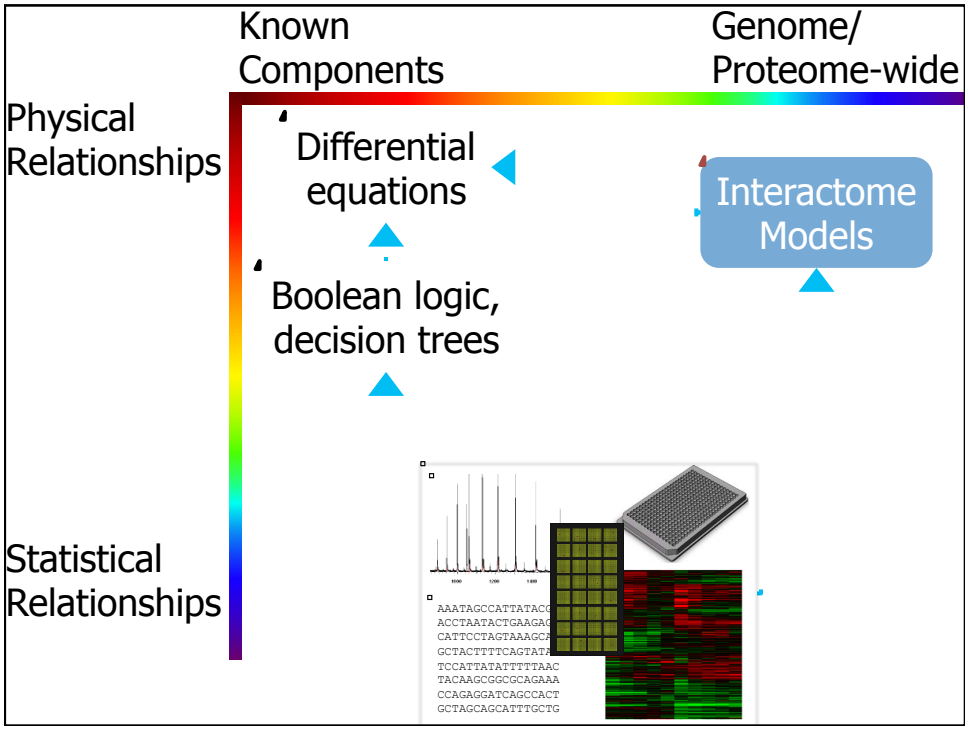


?

Appropriate for Dynamic Modeling



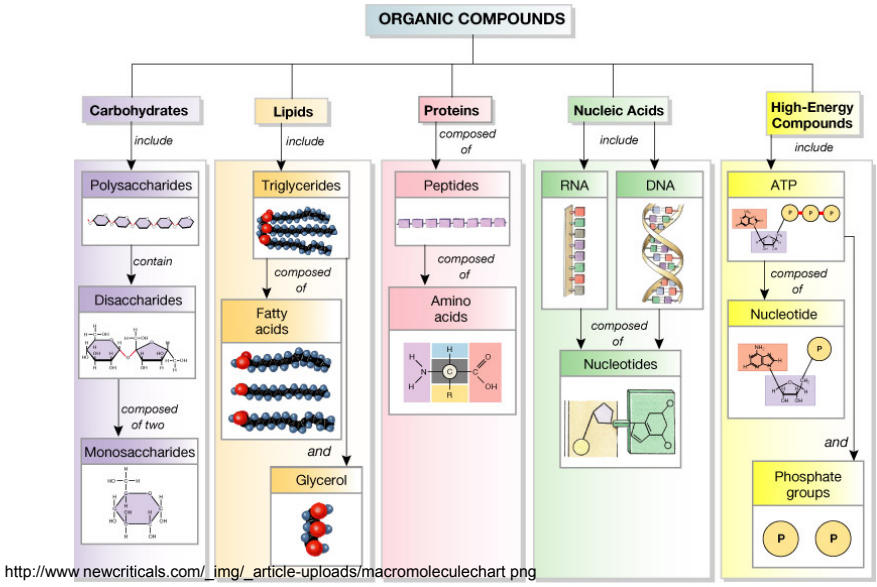




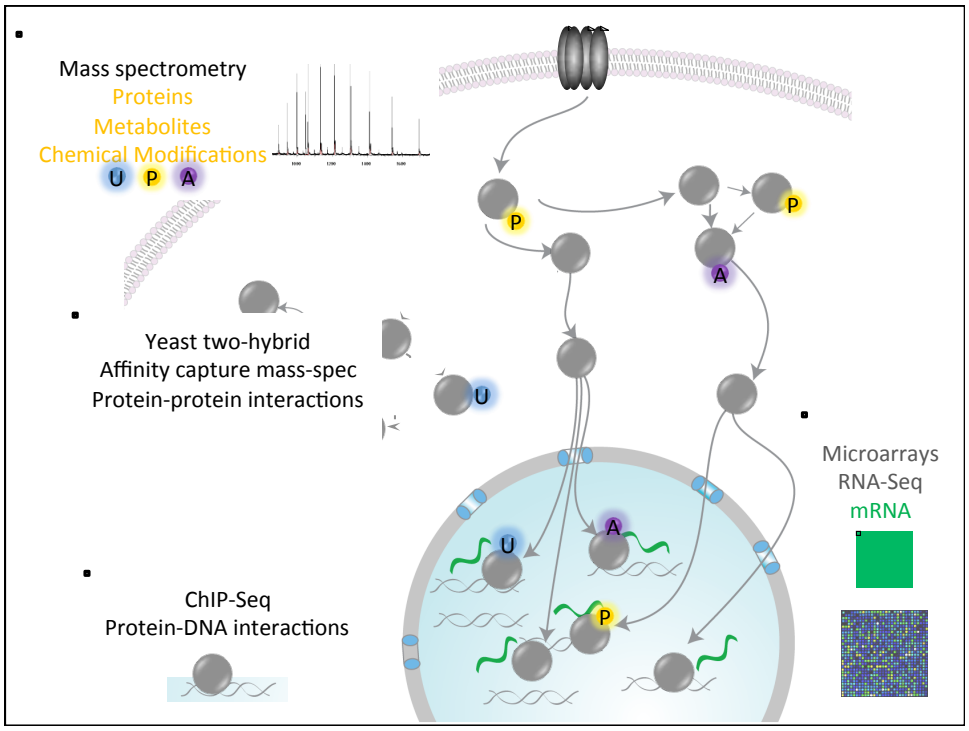
Outline

- What molecules we can measure?
- How do we know which interact?
- How do we learn anything from these data?
  - Standard Approaches
  - Challenges
  - Network Methods
  - Toward Dynamic Models

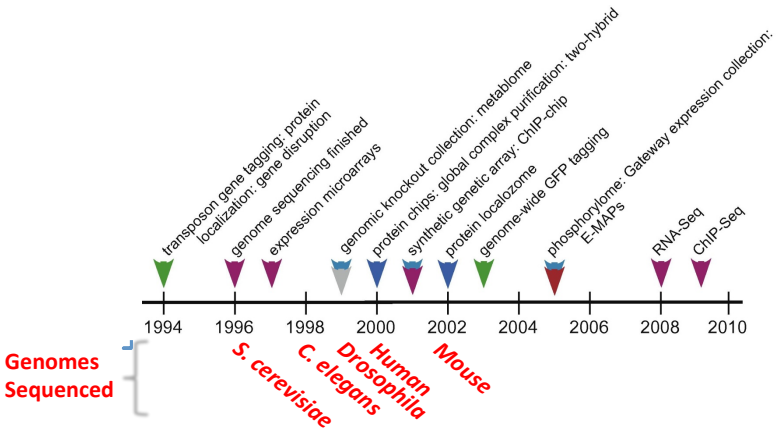
What's inside a cell?



http://www.newcriticals.com/_img/_article-uploads/macromoleculechart.png



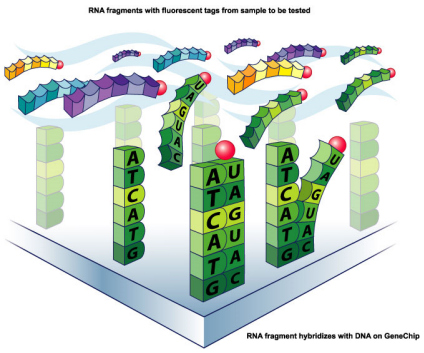
Timeline of Omic Methods



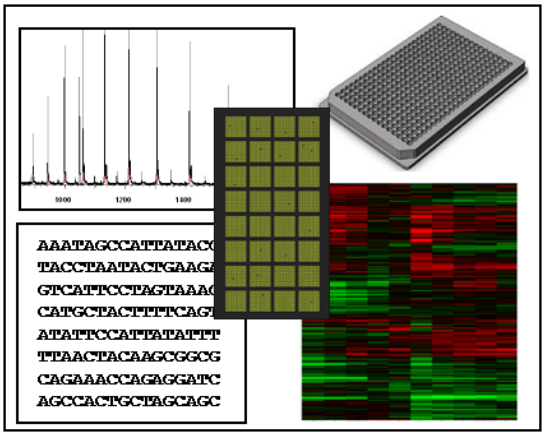
Source Snyder andGallagher
FEBS Letters 2009 583 3895-3899 (DOI 10.1016/j.febslet.2009.11.011)

13

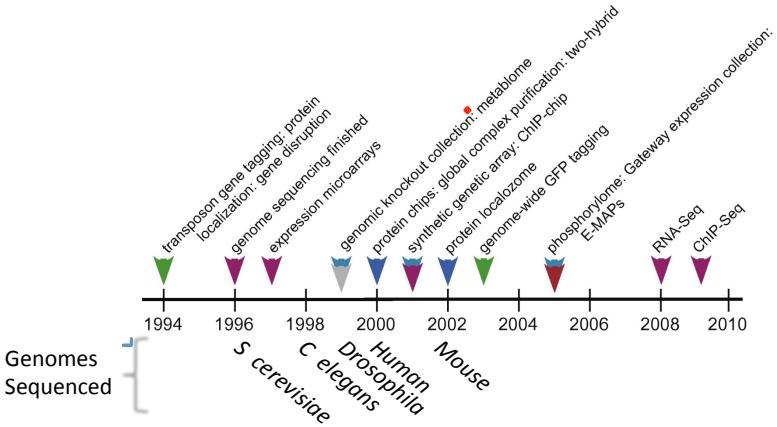
mRNA levels



What else can we measure?



15



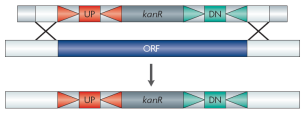
Source: Snyder andGallagher
FEBS Letters 2009: 583:3895-3899 (DOI:10.1016/j.febslet.2009.11.011)

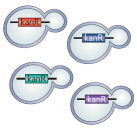
16

Homologous Recombination

Homologous recombination

Novel DNA
Chromosome
New chromosome





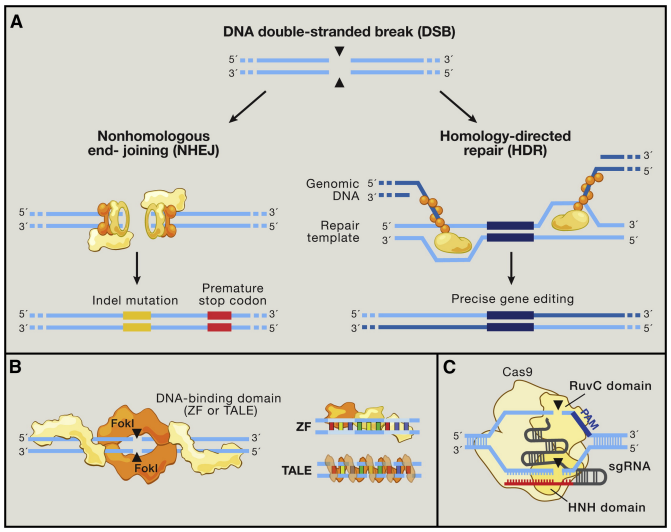
Deletion Library

~80% of yeast deletions are viable under standard lab conditions

Boone *et* al. (2007) Nature Reviews Genetics

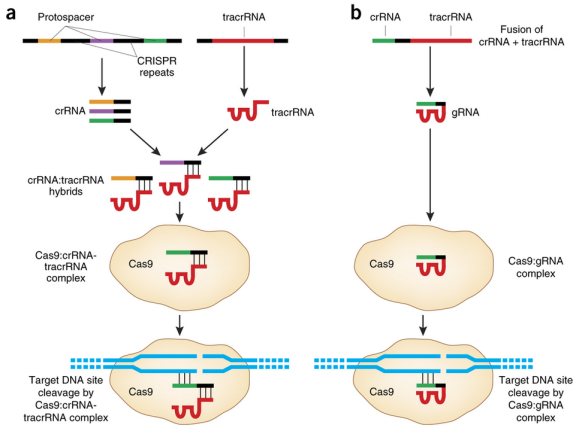
17

Strategies for Genome Editing



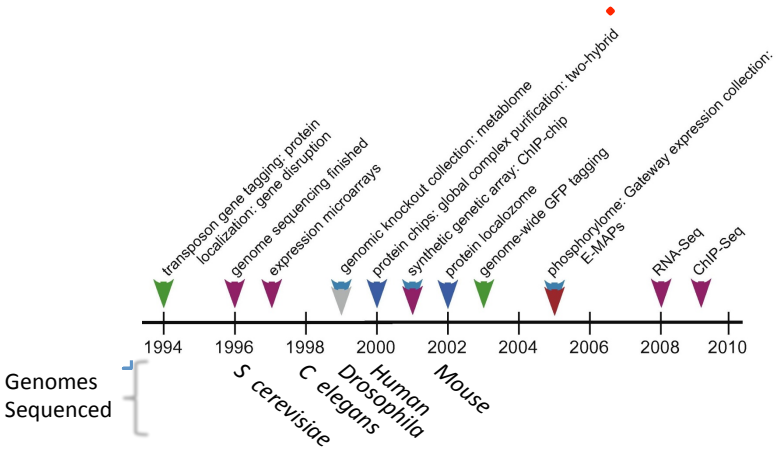
Cell 2014 157, 1262-1278DOI: (10.1016/j.cell.2014.05.010)

CRISPR/Cas9 is an RNA-guided DNA nuclease



CRISPR-Cas systems for editing, regulating and targeting genomes
Jeffry D Sander & J Keith Joung
Nature Biotechnology **32**, 347–355 (2014) doi:10.1038/nbt.2842

19



Source: Snyder andGallagher
FEBS Letters 2009: 583:3895-3899 (DOI:10.1016/j.febslet.2009.11.011)

20

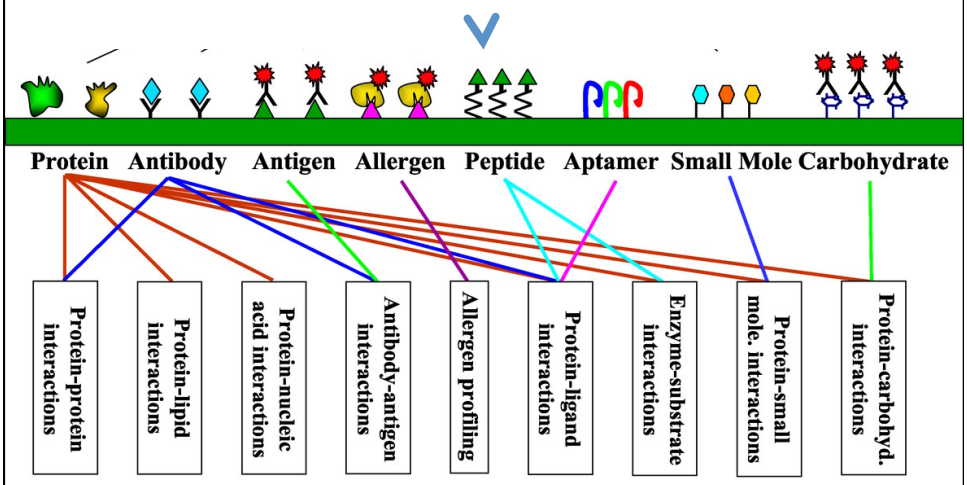
Proteomic Arrays

- Can we extend the DNA array platform for proteomics?
  - What would you put on the array?
  - What would you use to probe the array?
  - How would you detect it?
  - What are the technical challenges?

21

Detect bound proteins using:
•Fluorescently labeled proteins
•Radioactive proteins
•Antibody

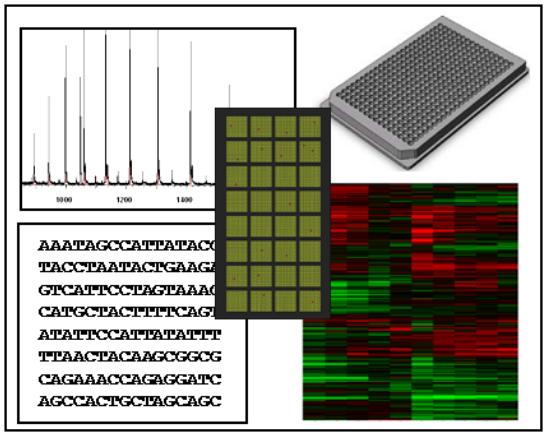
Add lysate, serum, etc



Zhu, et al. Annual Review of Biochemistry
Vol. 72: 783-812 (Volume publication date July 2003)

22

What else can we measure?



23

Post-translational modifications

Acetylation
ADP-ribosylation
Allysine
Amidation
S-archaeol
Beta-methylthiolation
Biotin
Bromination
N6-1-carboxyethyl lysine
Cholesterol
Cis-14-hydroxy-10,13-dioxo-7-heptadecenoic
acid aspartate ester
Citrullination
C-Mannosylation
Cysteine sulfenic acid (-SOH)
Cysteine sulfinic acid (-SO_2H)
Cysteine persulfide
Deamidation
Deamidation followed by a methylation
n-Decanoate
2,3-didehydroalanine (Ser)
2,3-didehydrobutyrine
(Z)-2,3-didehydrotyrosine
S-diacylglycerol cysteine
Dihydroxylation
Dimethylation
Dimethylation of proline
Diphthamide
FAD
S-farnesyl cysteine
S-12-hydroxyfarnesyl cysteine
3-phenyllactic acid
FMN conjugation (Cys)
FMN conjugation (Ser/Thr)
FMN conjugation (His)
Formylation
Geranyl-geranylation
Gamma-carboxyglutamic acid
O-GlcNAc
Glucosylation (Glycation)
Glutathionylation
Hydroxylation
Hypusine
Lipoyl
Methylation
Methionine sulfone
Myristoylation
S-Nitrosylation
n-Octanoate
Omega-hydroxyceramide glutamate ester
3-oxoalanine (Cys)
3-oxoalanine (Ser)
2-oxobutanoic acid
Palmitoylation
S-palmitoleyl cysteine
Phosphatidylethanolamine amidated glycine
Phosphorylation
Pyridoxal phosphate
N6-poly(methylaminopropyl)lysine
Phosphopantetheine
Pyrrolidone carboxylic acid (Glu)
Pyrrolysine
Pyrrolidone carboxylic acid
Pyruvic acid (Cys)
Pyruvic acid (Ser)
Sulfation
1-thioglycine
Thyroxine
2',4',5'-topaquinone
Triiodothyronine
Trimethylation
N6,N6,N6-trimethyl-5-hydroxylysine

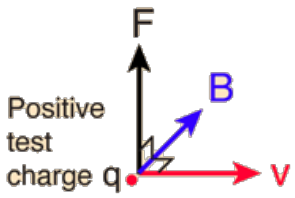
http://ca.expasy org/tools/findmod/findmod_masses.html

Mass Spectrometry

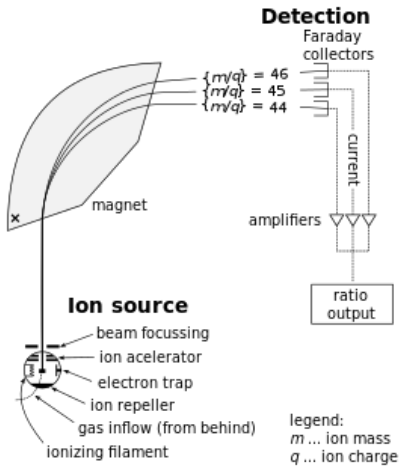
- Can identify proteins and other molecules by extremely accurate measurement of mass

Force on a charge in a magnetic field

$$\vec{F} = q\vec{v} \ x \ \vec{B}$$

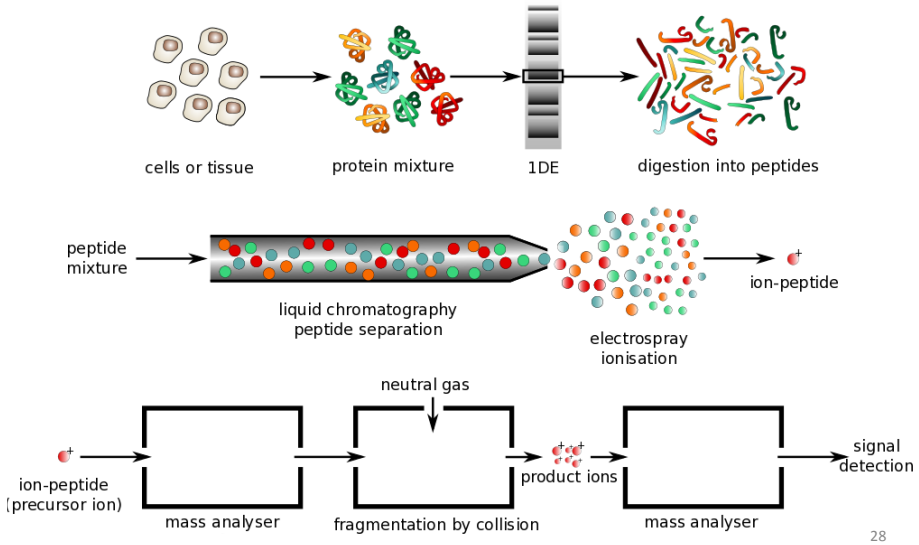


MS separates by m/q



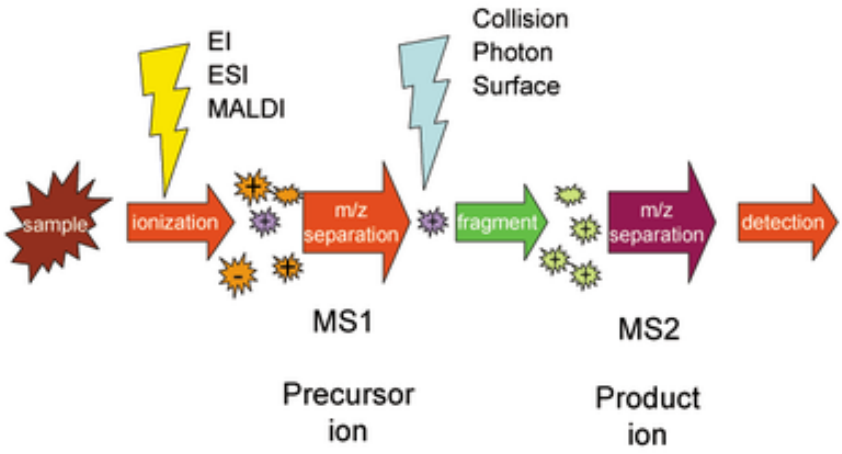
http://en.wikipedia.org/wiki/Mass_spectrometry

MS cannot analyze very complex samples



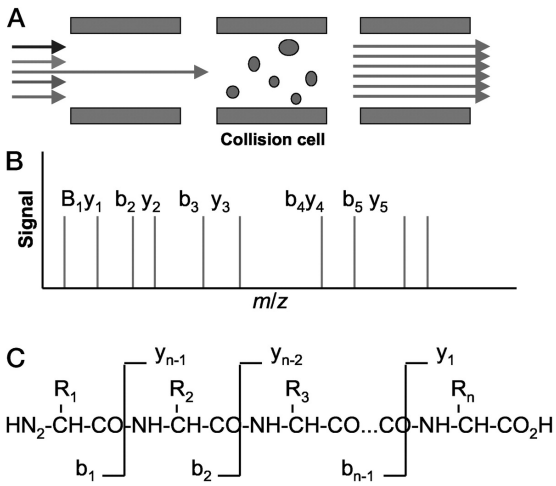
http://upload.wikimedia.org/wikipedia/commons/1/1f/Mass_spectrometry_protocol.png

Tandem Mass Spec Increases Accuracy



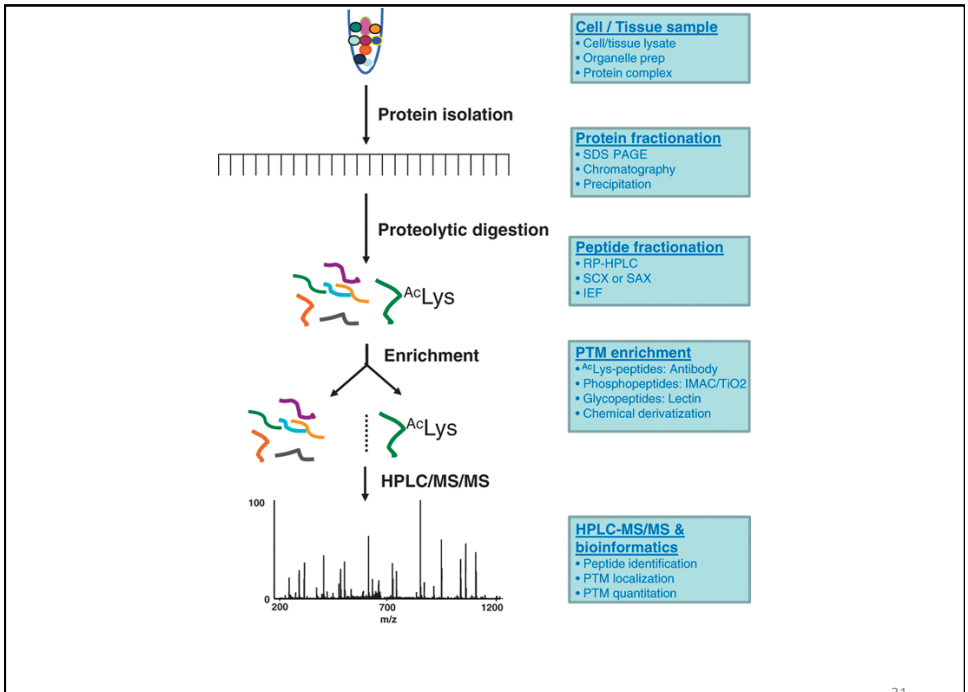
http://en.wikipedia.org/wiki/Tandem_mass_spectrometry

Peptide MS/MS

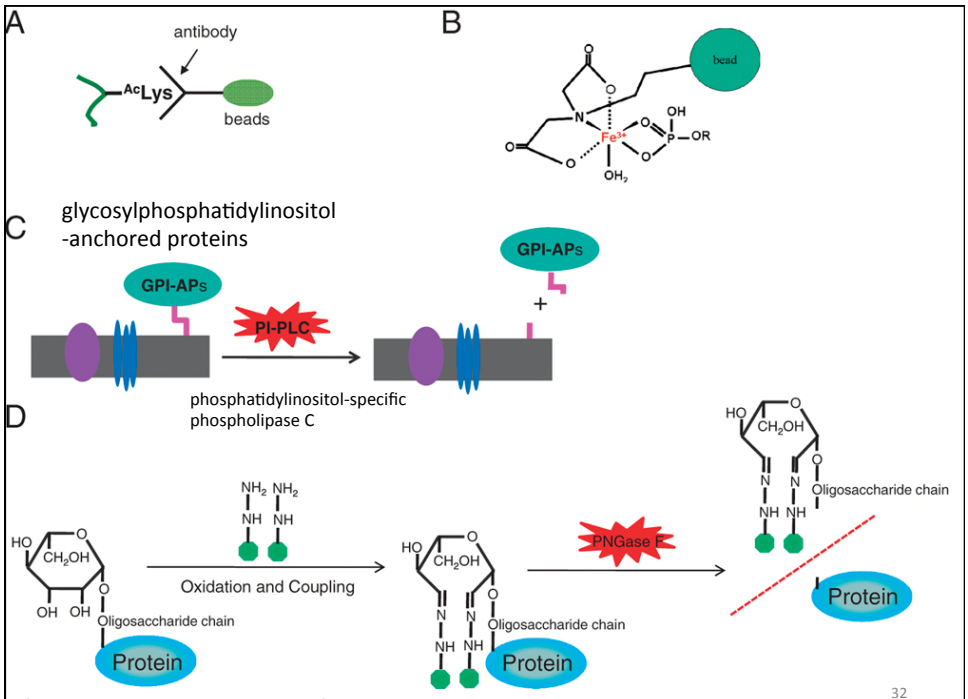


Zhu, et al. Annual Review of Biochemistry
Vol. 72: 783-812 (Volume publication date July 2003)

30

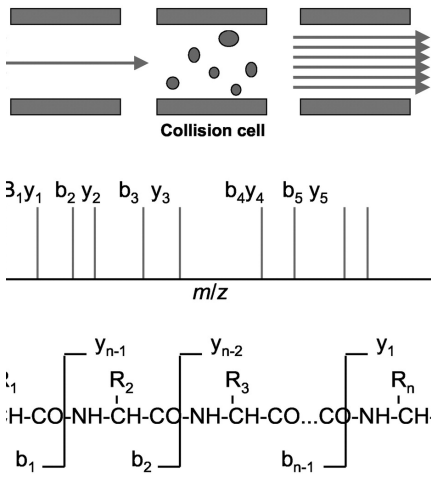


Zhao & Jensen. Proteomics Vol.9, 20 Pages: 4632-4641



Zhao & Jensen. Proteomics Vol.9, 20 Pages: 4632-4641

Peptide MS/MS

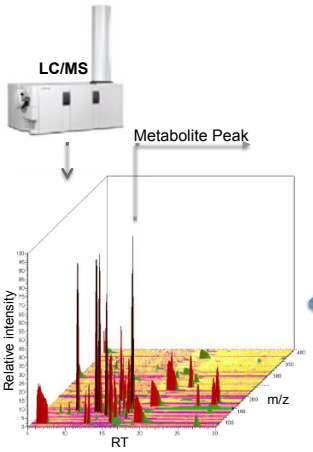


- Analyzing peptide MS/MS is a discrete matching problem
- Which peptide encoded in the genome is the most likely match?
- Leverage knowledge of peptide cutting preferences, possible modifications.

Zhu, et al. Annual Review of Biochemistry
Vol. 72: 783-812 (Volume publication date July 2003)

33

No Map of the Metabolome



Protein:
cleavage products are almost always peptides
easy to compare mass to all possible peptides encoded by the genome

Metabolites: not directly encoded in the genome.
May be > 40,000 different species.

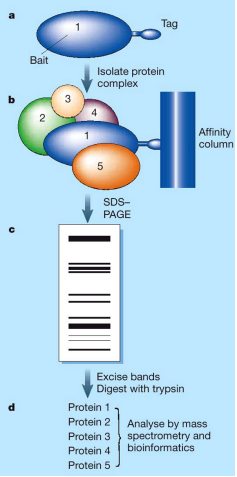
Outline

- What molecules we can measure
- How do we know which interact?
- How do we learn anything from these data?

Detecting protein-protein interactions

What are the likely false positives?

What are the likely false negatives?



Gavin, A.-C. *et al*. *Nature* **415**, 141-147 (2002).

Ho, Y. *et al*. *Nature* **415**, 180-183 (2002).

Proteomics: Protein complexes take the bait
Anuj Kumar and Michael Snyder
Nature 415, 123-124(10 January 2002)
doi:10.1038/415123a

36

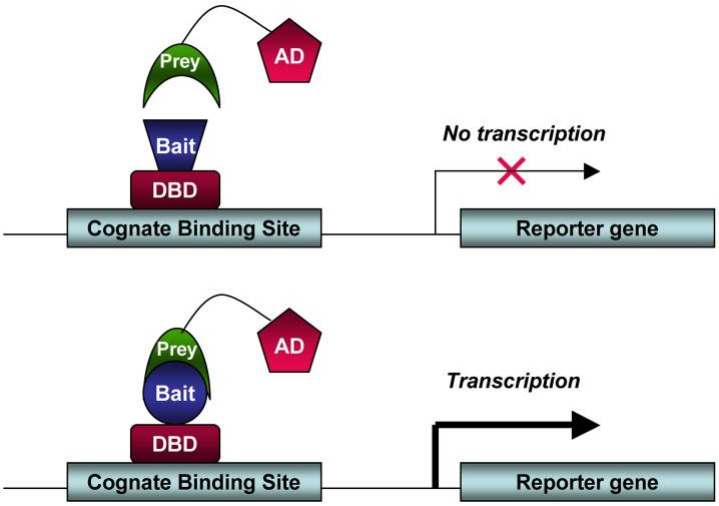
Mass-spec for protein-protein interactions

- Extremely efficient method for detecting interactions
- Proteins are in their correct subcellular location.

Limitations?

- overexpression/tagging can influence results
- only long-lived complexes will be detected

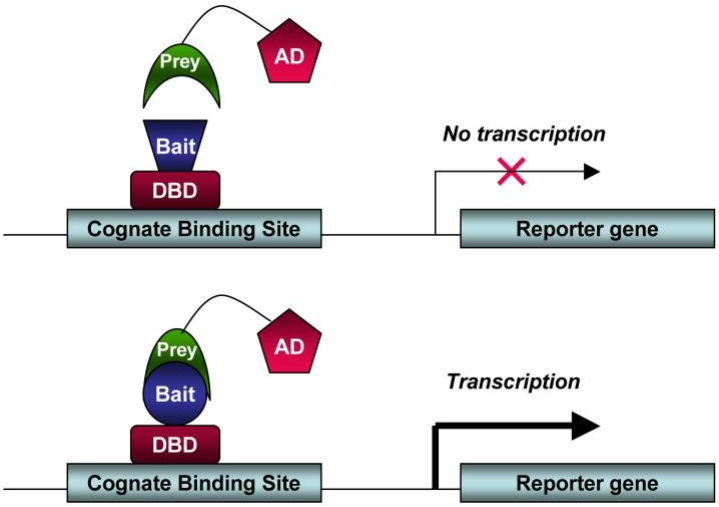
37



Biotechniques. 2008 Apr;44(5):655-62.
Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system.
Ratushny V, Golemis E.

How does this compare to mass-spec based approaches

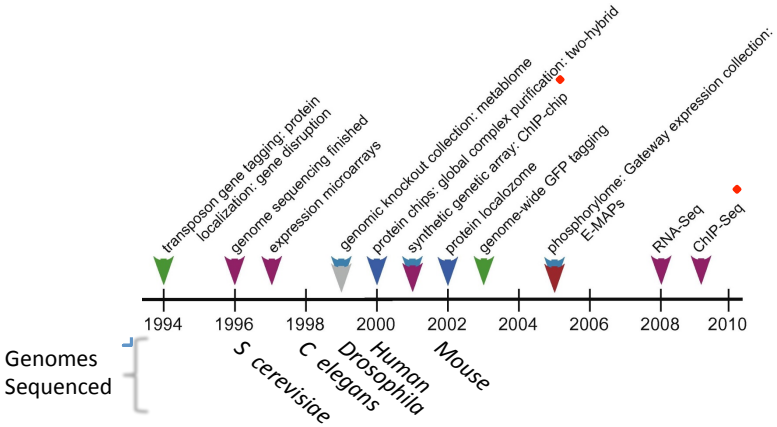
38



Biotechniques. 2008 Apr;44(5):655-62.
Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system.
Ratushny V, Golemis E.

- **Does not require purification – will pick up more transient interactions.**
- **Biased against proteins that do not express well, or are incompatible with the nucleus**

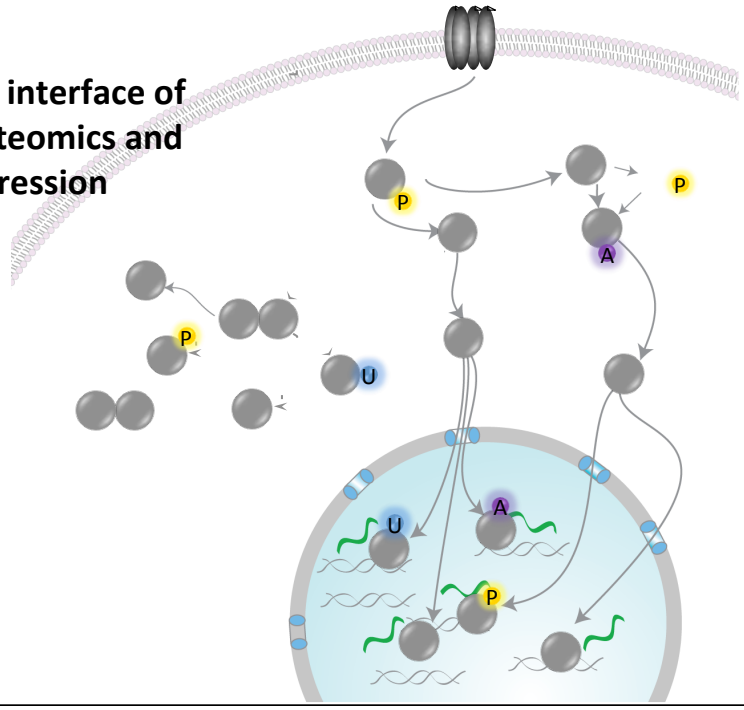
39



Source: Snyder andGallagher
FEBS Letters 2009; 583:3895-3899 (DOI:10.1016/j.febslet.2009.11.011)

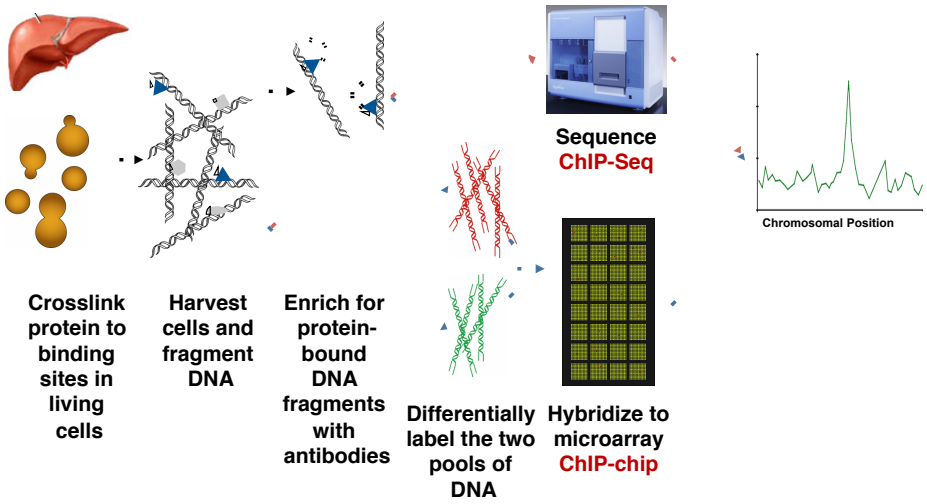
40

The interface of proteomics and expression

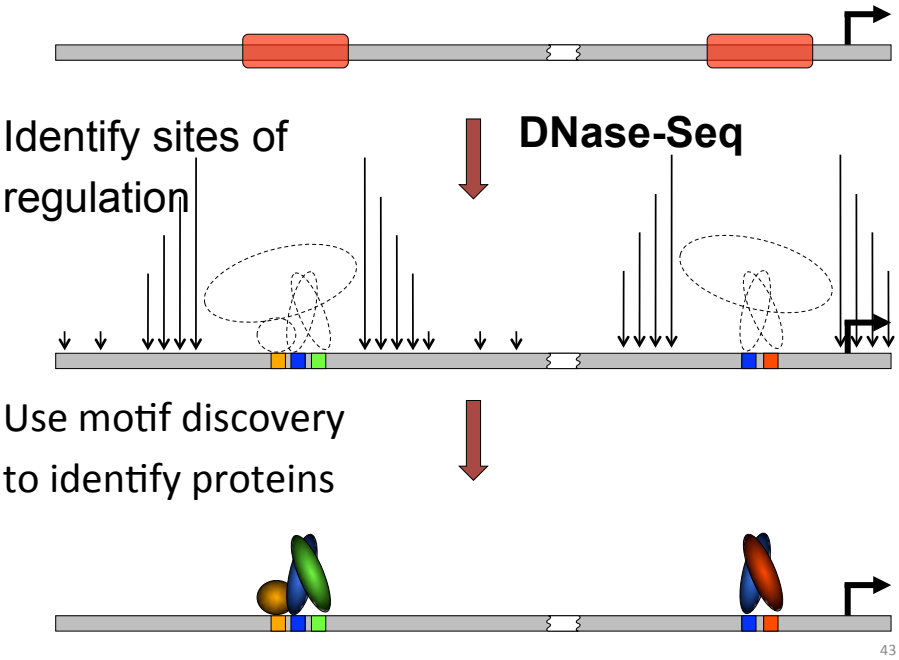


41

Genome-wide Chromatin Immunoprecipitation

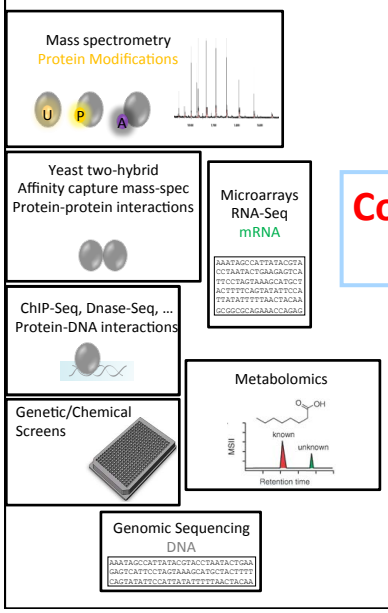


42



Outline

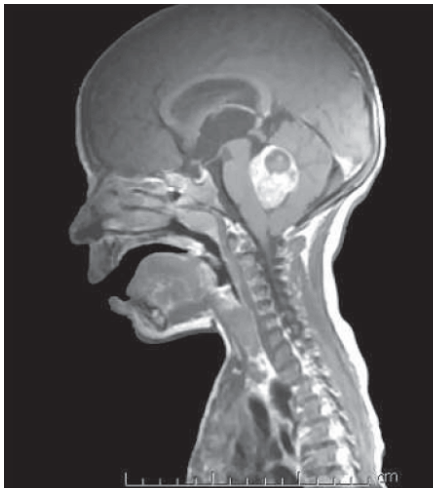
- What molecules we can measure?
- How do we know which interact?
- How do we learn anything from these data?
  - Standard Approaches
  - Challenges
  - Network Methods
  - Toward Dynamic Models



Computational Models

- Basic biology
  - transcription
  - signaling
  - microbiome
  - ...
- Tumor classification

The most common pediatric malignant brain tumor



Polkinghorn and Tarbell. Medulloblastoma: tumorigenesis, current clinical paradigm, and efforts to improve risk stratification. Nat Clin Pract Oncol (2007) vol. 4 (5) pp. 295-304

- 70% survive but only 10% live independently as adults due to neurologic disability from the tumor and treatment

Mutations identified in genomics era of Medulloblastoma

LETTER

doi:10.1038/nature11329

Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations

Trevor J. Pugh[1,2,3], Shyamal Dilhan Weeraratne[3,4], Tenley C. Archer[3,4], Daniel A. Pomeranz Krummel[5], Daniel Auclair[1], James Bochicchio[1], Mauricio O. Carneiro[1], Scott L. Carter[1], Kristian Cibulskis[1], Rachel L. Erlich[1], Heidi Greulich[1,2,3], Michael S. Lawrence[1], Niall J. Lennon[1], Aaron McKenna[1], James Meldrim[1], Alex H. Ramos[1,2,3], Michael G. Ross[1], Carsten Russ[1], Erica Shefler[1], Andrey Sivachenko[1], Brian Sogoloff[1], Petar Stojanov[1], Pablo Tamayo[1], Jill P. Mesirov[1], Vladimir Amani[3,4], Natalia Teider[3,4], Soma Sengupta[3,4], Jessica Pierre Francois[3,4], Paul A. Northcott[6], Michael D. Taylor[6], Furong Yu[7], Gerald R. Crabtree[7,8], Amanda G. Kautzman[7], Stacey B. Gabriel[1], Gad Getz[1], Natalie Jäger[9], David T. W. Jones[9], Peter Lichter[9], Stefan M. Pfister[9], Thomas M. Roberts[2,3], Matthew Meyerson[1,2,3,10], Scott L. Pomeroy[3,4,8] & Yoon-Jae Cho[1,3,4,7]

LETTER

doi:10.1038/nature11284

Dissecting the genomic complexity underlying medulloblastoma

A list of authors and their affiliations appears at the end

ARTICLE

doi:10.1038/nature11213

Novel mutations target distinct subgroups of medulloblastoma

Giles Robinson[1,2,3]*, Matthew Parker[1,4]*, Tanya A. Kranenburg[1,2]*, Charles Lu[1,5], Xiang Chen[1,4], Li Ding[1,5,6], Timothy N. Phoenix[1,2], Erin Hedlund[1,4], Lei Wei[1,4,7], Xiaoyan Zhu[1,2], Nader Chalhoub[1,2], Suzanne J. Baker[1,2], Robert Huether[1,4,8], Radhika Thiruvenkatam[1,2], Jianmin Wang[1,9], Gang Wu[1,4], Michael Rusch[1,4], Xin Hong[1,5], Jian Ma[1,7], John Easton[1,4], Bhavin Vadodaria[1,5], Arzu Onar-Thomas[1,10], Tong Lin[1,10], Sue Paugh[1,11], David Zhao[1,9], Daisuke Kawauchi[1,12], Martine F. Roussel[1,12], Ching C. Lau[1,13], Eric Bouffet[1,14], Tim Hassall[1,15], Sridharan Gururangan[1,16], Lucinda L. Fulton[1,5,6], David J. Dooling[1,5,8], Kerri Ochoa[1,5,6], Amar Gajjar[1,3], Jon[1,5,6,19], James R. Downing[1,7], Jinghui Zhang[1,4] & Richard J. Gilbertson[1,2,3]

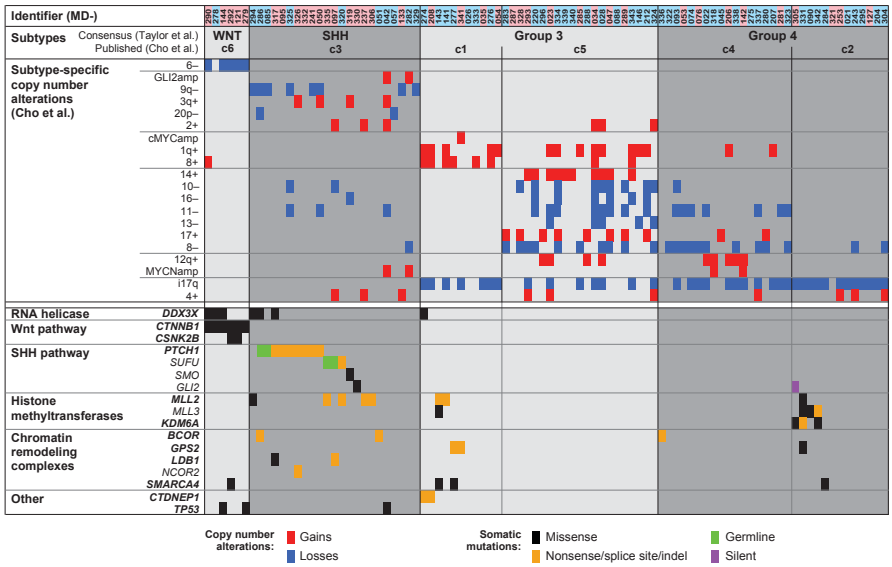
ARTICLE

doi:10.1038/nature11327

Subgroup-specific structural variation across 1,000 medulloblastoma genomes

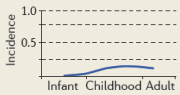
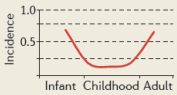
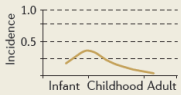
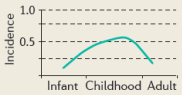
A list of authors and their affiliations appears at the end of the paper

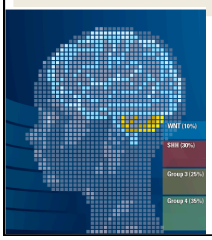
Genomic Data Identifies Subtypes



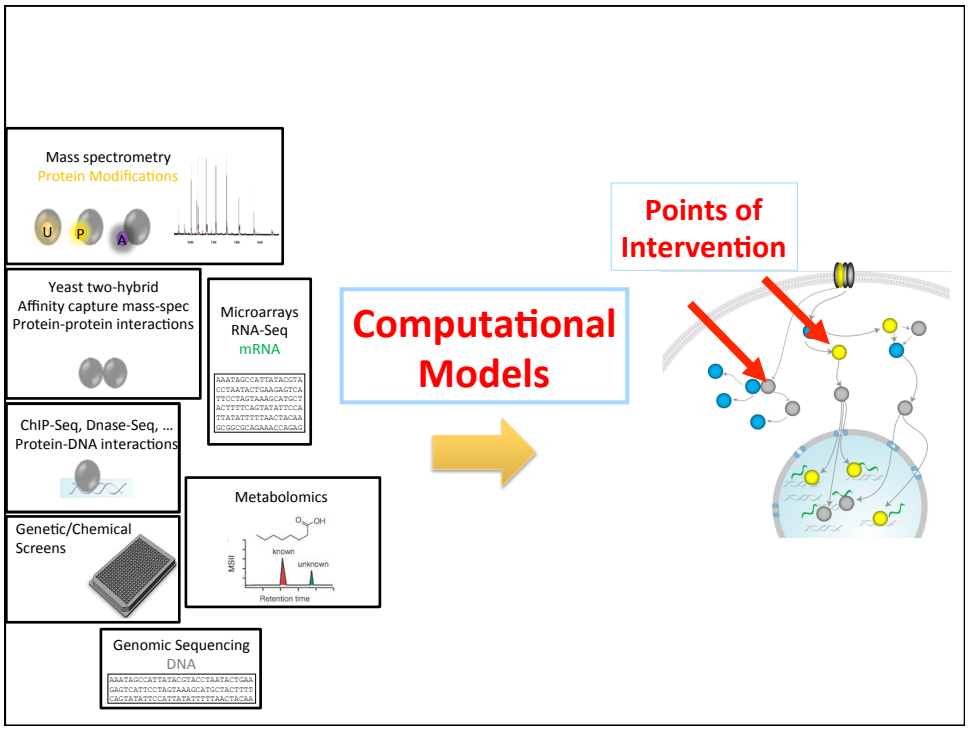
Pugh, T, Weeraratne, SD, Archer T, *et al*. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).

Four Types of Medulloblastoma

	WNT (~10%)	SHH (~30%)	Group 3 (~25%)	Group 4 (~35%)
Clinical features				
Gender ratio (M/F)	~1/1	~1.5/1	~2/1	~3/1
Age distribution				
Histology	Classic; very rare LCA	Classic > desmoplastic/ nodular > LCA > MBEN	Classic > LCA	Classic; rarely LCA
Metastasis at diagnosis	~5–10%	~15–20%	~40–45%	~35–40%
Overall survival (5 years)	~95%	~75%	~50%	~75%
Proposed cell of origin	Lower rhombic lip progenitor cells	CGNPs of the EGL and cochlear nucleus; neural stem cells of the SVZ	Prominin 1⁺, lineage⁻ neural stem cells; CGNPs of the EGL	Unknown



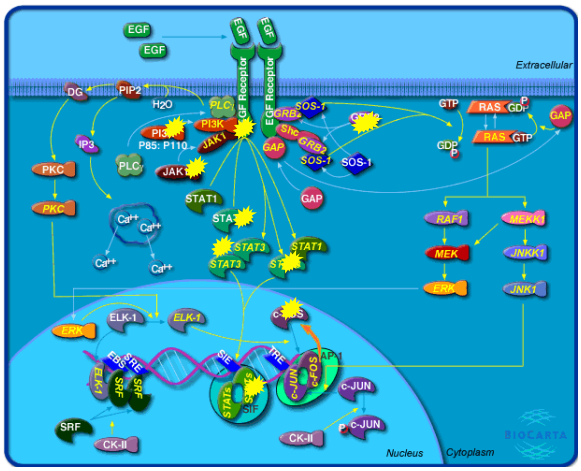
Northcott et al. Nat Rev Cancer. 2012 Dec;12(12):818-34. doi: 10.1038/nrc3410.



Outline

- What molecules we can measure?
- How do we know which interact?
- How do we learn anything from these data?
  - Standard Approaches
  - Challenges
  - Network Methods
  - Toward Dynamic Models

Map to Known Pathways



Name	Availability	Reference
ORA tools		
Onto-Express	Web (http://vortex.cs.wayne.edu)	[4,5]
GenMAPP	Standalone (http://www.genmapp.org)	[11,71]
GoMiner	Standalone, Web (http://discover.nci.nih.gov/gominer)	[72,73]
FatiGO	Web (http://babelomics.bioinfo.cipf.es)	[74]
GOstat	Web (http://gostat.wehi.edu.au)	[7]
FuncAssociate	Web (http://llama.mshri.on.ca/funcassociate/)	[6]
GOToolBox	Web (http://genome.crg.es/GOToolBox/)	[10]
GeneMerge	Standalone, Web (http://genemerge.cbcb.umd.edu/)	[9]
GOEAST	Web (http://omicslab.genetics.ac.cn/GOEAST/)	[75]
ClueGO	Standalone (http://www.ici.upmc.fr/cluego/)	[76]
FunSpec	Web (http://funspec.med.utoronto.ca/)	[77]
GARBAN	Web	[78]
GO:TermFinder	Standalone (http://search.cpan.org/dist/GO-TermFinder/)	[8]
WebGestalt	Web (http://bioinfo.vanderbilt.edu/webgestalt/)	[79]
agriGO	Web (http://bioinfo.cau.edu.cn/agriGO/)	[80]
GOFFA	Standalone, Web (http://edkb.fda.gov/webstart/arraytrack/)	[81]
WEGO	Web (http://wego.genomics.org.cn/cgi-bin/wego/index.pl)	[82]
FCS tools		
GSEA	Standalone (http://www.broadinstitute.org/gsea/)	[21,29]
sigPathway	Standalone (BioConductor)	[22]
Category	Standalone (BioConductor)	[24]
SAFE	Standalone (BioConductor)	[30]
GlobalTest	Standalone (BioConductor)	[15]
PCOT2	Standalone (BioConductor)	[17]
SAM-GS	Standalone (http://www.ualberta.ca/~yyasui/software.html)	[83]
Catmap	Standalone (http://bioinfo.thep.lu.se/catmap.html)	[84]
T-profiler	Web (http://www.t-profiler.org)	[85]
FunCluster	Standalone (http://comeliu.henegar.info/FunCluster.htm)	[86]
GeneTrail	Web (http://genetrail.bioinf.uni-sb.de)	[87]
GAzer	Web	[88]
PT-based tools		
ScorePAGE	No implementation available	[37]
Pathway-Express	Web (http://vortex.cs.wayne.edu)	[38,39]
SPIA	Standalone (BioConductor)	[40]
NetGSA	No implementation available	[43]

doi:10.1371/journal.pcbi.1002375.t001

Khatri P, Sirota M, Butte AJ (2012) Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. PLoS Comput Biol 8(2): e1002375. doi:10.1371/journal.pcbi.1002375
http://127.0.0.1:8081/ploscompbiol/article?id=info:doi/10.1371/journal.pcbi.1002375

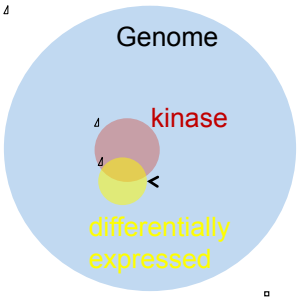
PLOS | COMPUTATIONAL BIOLOGY

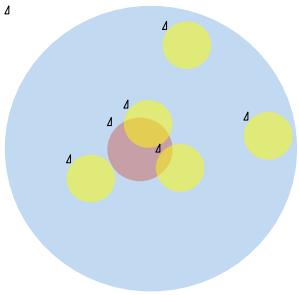
Known Pathways

Controlled vocabulary to describe genes:

- Biological process
- Cellular component
- Molecular function

Statistical Significance

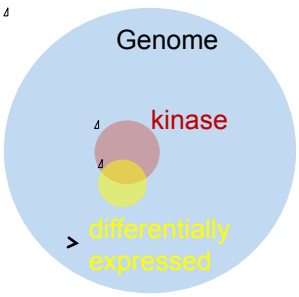


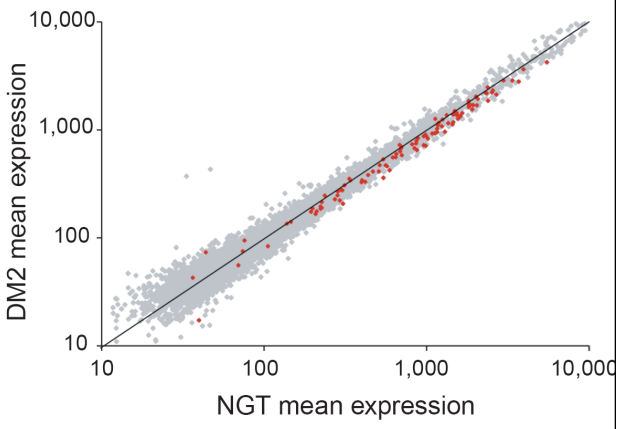


- Compute significance of overlap using the hypergeometric distribution.

Aggregate score statistics

- My results depend on how I defined "differentially expressed"



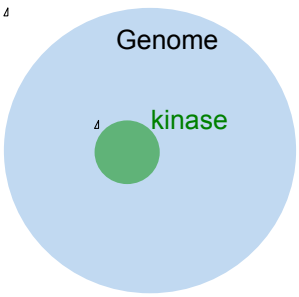


Mootha *et al.* (2003). *Nature Genetics* **34**, 267 – 273. doi:10.1038/ng1180

Aggregate score statistics

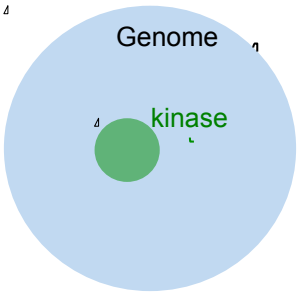
http://www.broadinstitute.org/gsea/

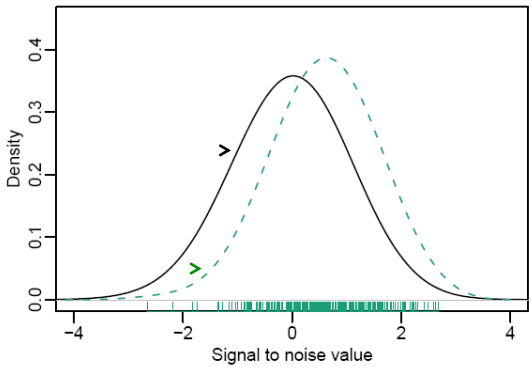
Aggregate score statistics



- For each list of genes in a database (such as GO)
  - compute distribution of t-statistics
  - compare this distribution to the overall distribution

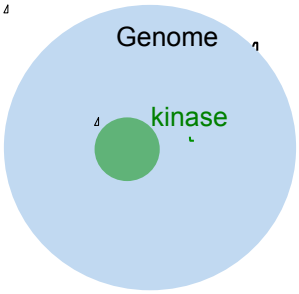
Aggregate score statistics

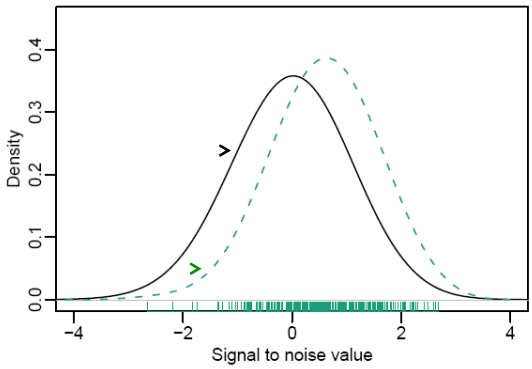




GSEA uses a Kolmogorov-Smirnov statistic to compare the distributions of t-statistics

Aggregate score statistics





Irizarry, et al. argue for X^2 and z-test

Gene set enrichment analysis made simple. (2009) Stat Methods Med Res
http://www.bepress.com/jhubiostat/paper185/

Outline

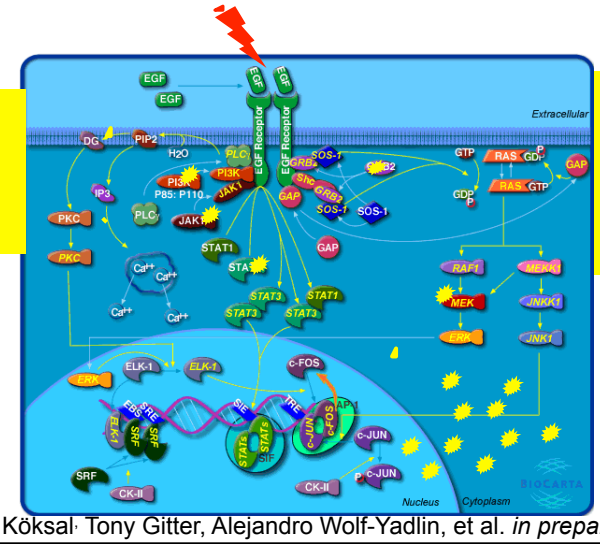
- What molecules we can measure?
- How do we know which interact?
- How do we learn anything from these data?
  - Standard Approaches
  - Challenges
  - Network Methods
  - Toward Dynamic Models

End of Omics Review

"Getting an Education from MIT is like taking a drink from a Fire Hose."
Former MIT President Jerome Weisner

Most 'Omic Hits Don't Lie in Known Pathways

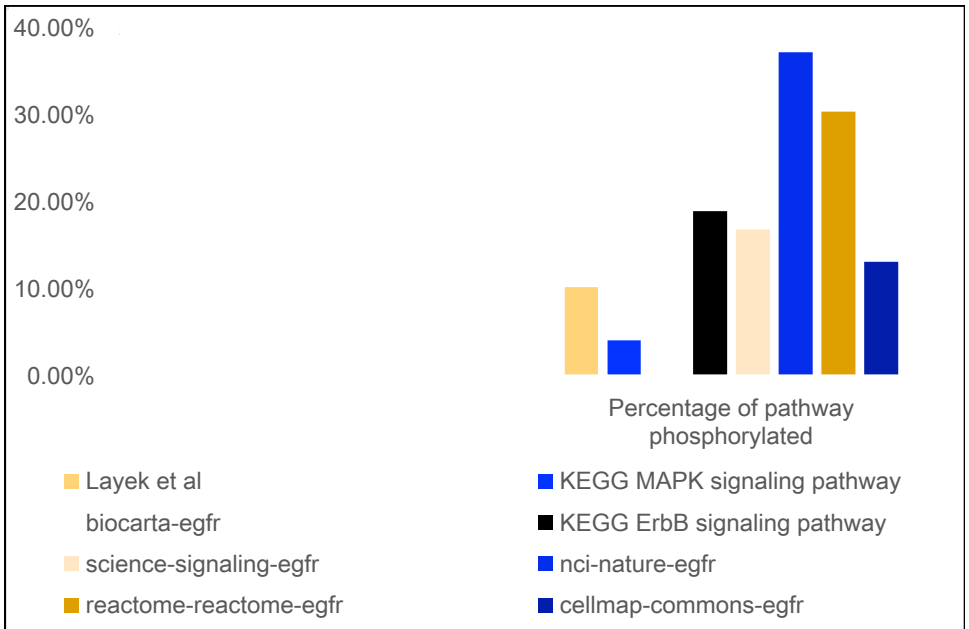
What % of expected pathway is activated?



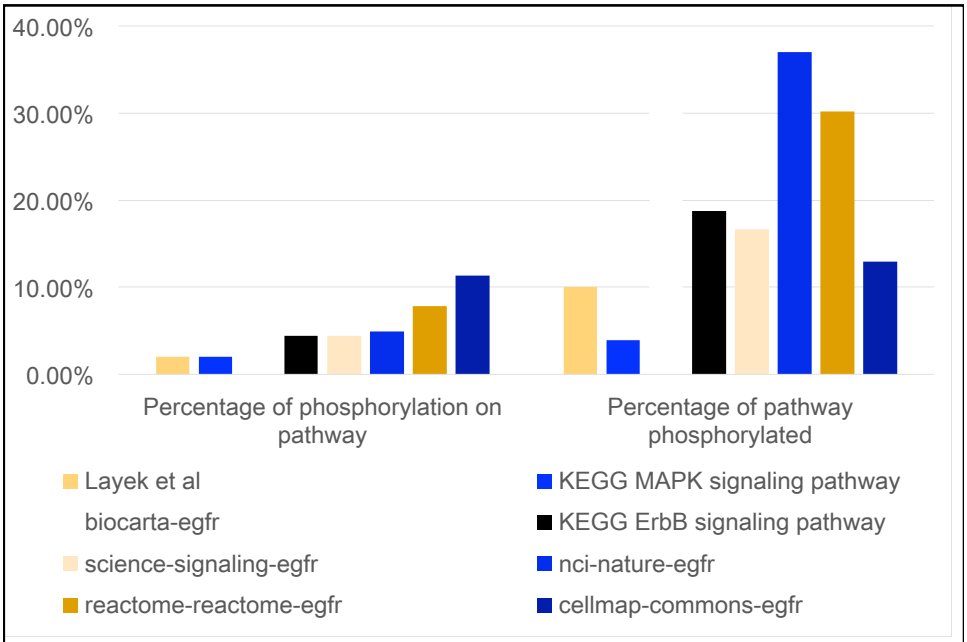
What % of activated proteins are off the pathway?

Biocarta EGF signaling pathway

Ali Sinan Köksal, Tony Gitter, Alejandro Wolf-Yadlin, et al. *in preparation.*

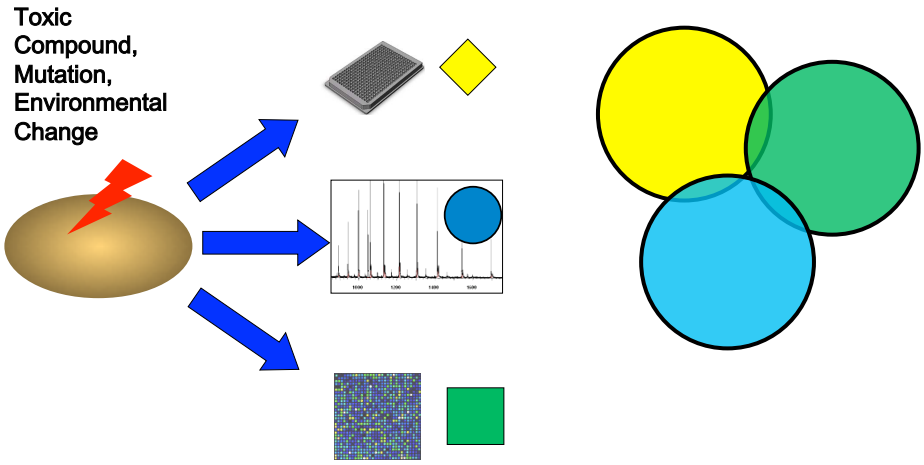


Ali Sinan Köksal, Tony Gitter, Alejandro Wolf-Yadlin, et al. *in preparation.*



Ali Sinan Köksal, Tony Gitter, Alejandro Wolf-Yadlin, et al. *in preparation.*

‘Omic data don’t agree



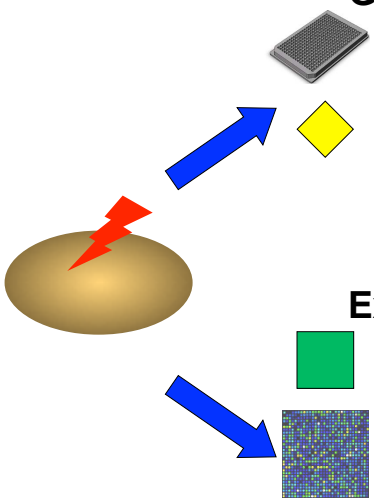
Esti Yeger-Lotem

Senior Lecturer
Ben-Gurion University
National Institute for Biotechnology in the Negev
Israel

Laura Riva

Team Leader
Center for Genomic Science
Istituto Italiano di Tecnologia
Italy

For 156 perturbations:

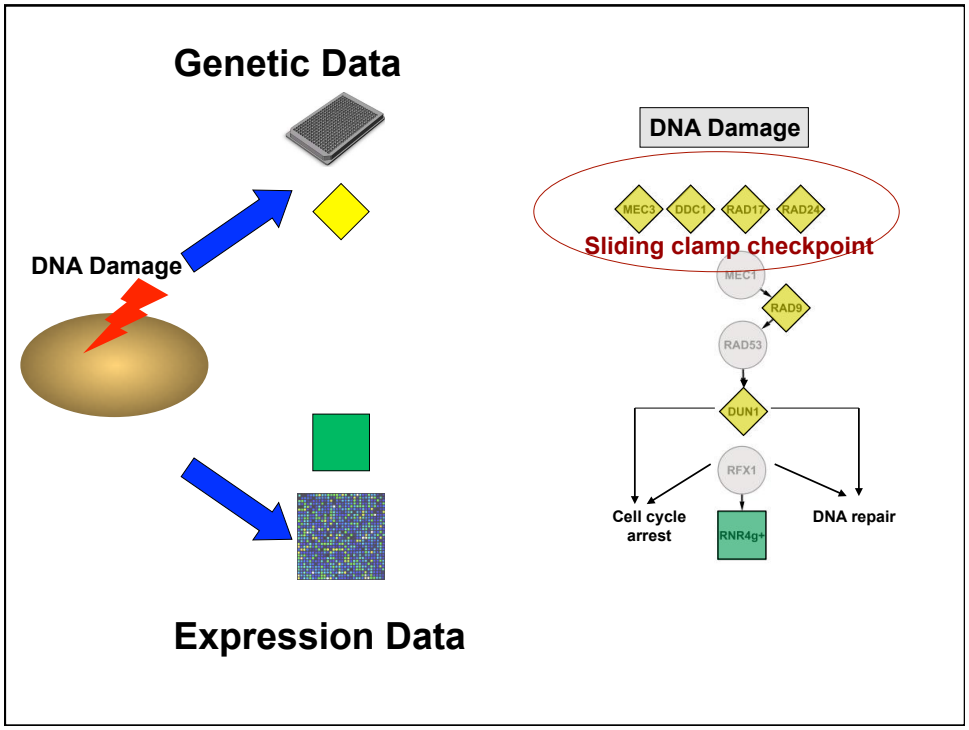


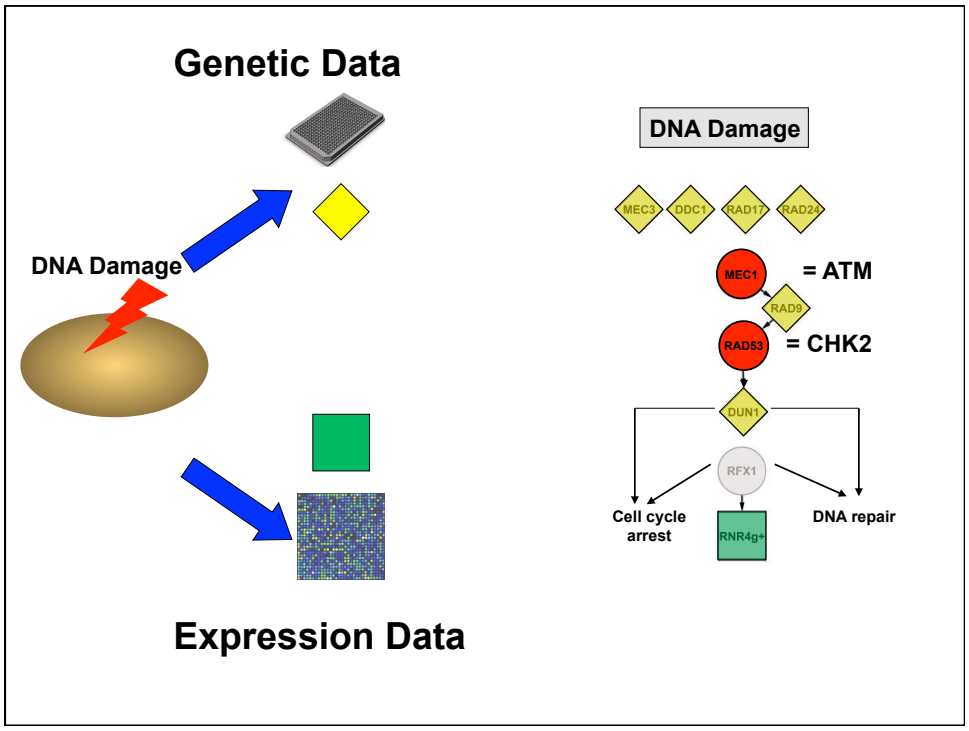
Genetic Data Enriched for:

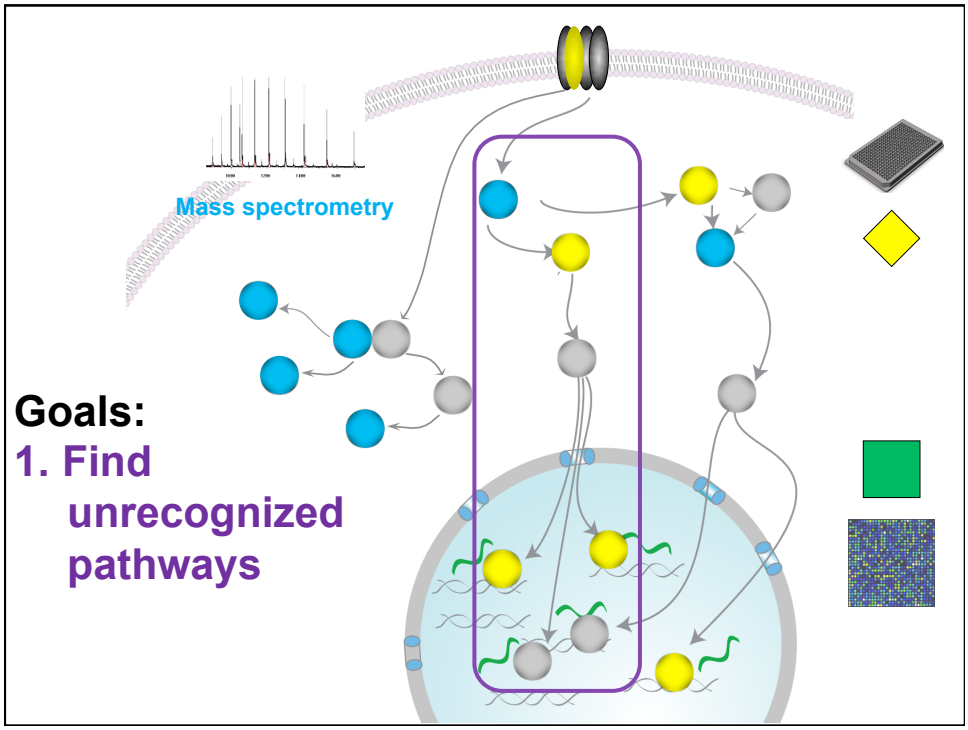
- **Transcriptional regulation**
- **Signal transduction**

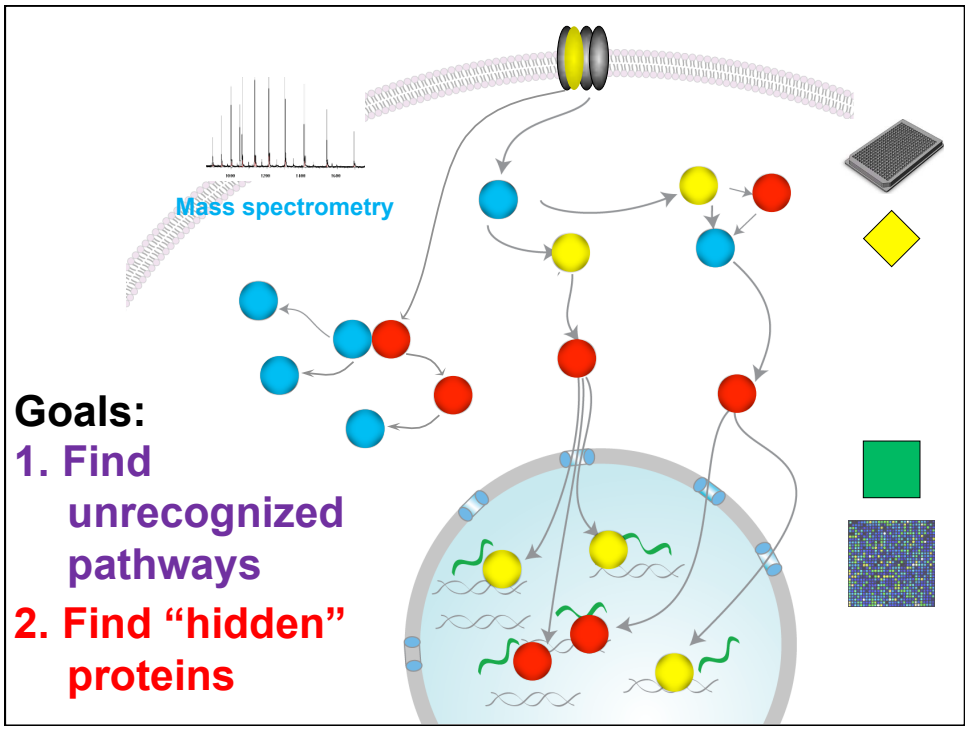
Expression Data Enriched for:

Metabolic Processes e.g., organic acid metabolic process, oxidoreductase activities

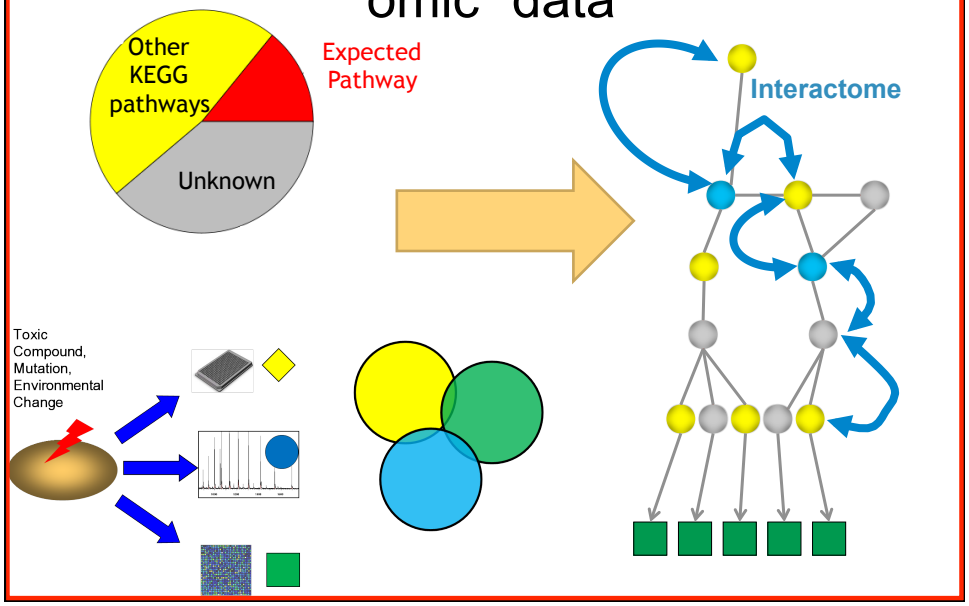

Genetic Data
DNA Damage
DNA Damage
MEC3
DDC1
RAD17
RAD24
Sliding clamp checkpoint
MEC1
RAD9
RAD53
DUN1
RFX1
Cell cycle arrest
RNR4g+
DNA repair
Expression Data


Genetic Data
DNA Damage
DNA Damage
MEC3
DDC1
RAD17
RAD24
MEC1
= ATM
RAD9
RAD53
= CHK2
DUN1
RFX1
Cell cycle arrest
RNR4g+
DNA repair
Expression Data

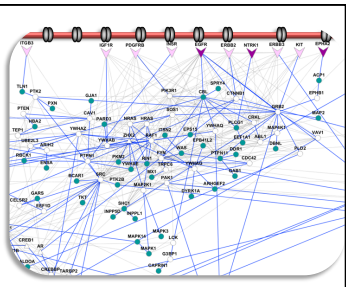

Mass spectrometry
Goals:
1. Find
unrecognized
pathways


Mass spectrometry
Goals:
1. Find
unrecognized
pathways
2. Find "hidden"
proteins

Changing how we think about "omic" data



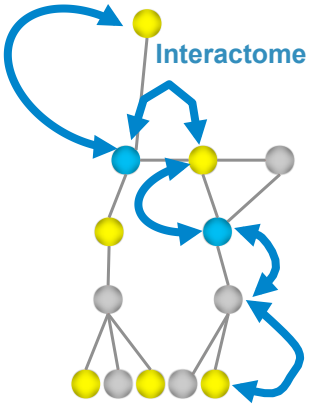
Outline



- Network Modeling
  - Approach
  - Predicting in vitro targets using RNA-Seq, DNase-Seq, and phospho-proteomics

Approach

Map data onto a network of known interactions.



Approach

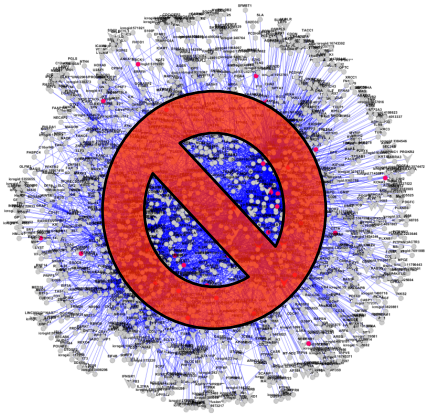
Lesson 1:

Network models make sense of diverse data.

Lesson 2:

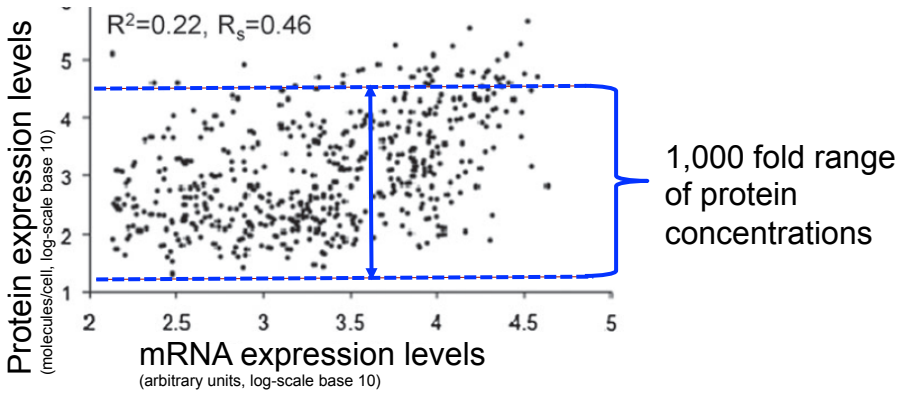
Hairballs do not!

Advanced network algorithms needed.



Approach

Hard to predict protein levels from mRNA

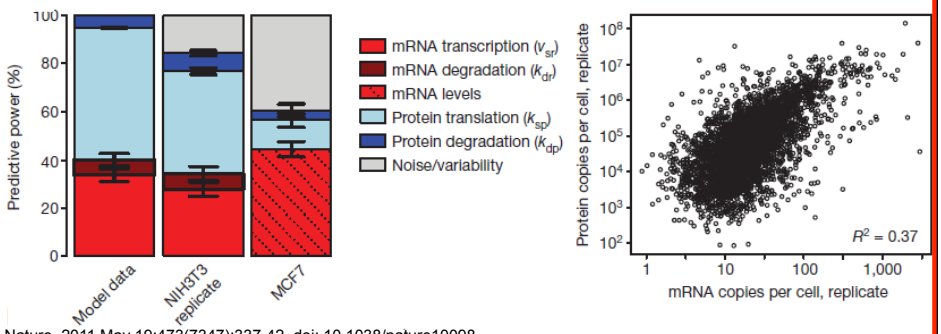


Raquel de Sousa Abreu, Luiz Penalva, Edward Marcotte and Christine Vogel, *Mol. BioSyst.*, 2009 DOI: 10.1039/b908315d

	SpectrumMill	msInspect	msBID	NSAF	RPKM	Microarray
SpectrumMill	-	0.91 (0.92)	0.91 (0.91)	0.90 (0.90)	0.49 (0.51)	0.36 (0.40)
msInspect	0.91 (0.92)	-	0.89 (0.91)	0.87 (0.88)	0.51 (0.53)	0.40 (0.44)
msBID	0.91 (0.91)	0.89 (0.91)	-	0.84 (0.89)	0.54 (0.54)	0.41 (0.42)
NSAF	0.90 (0.90)	0.87 (0.88)	0.84 (0.89)	-	0.51 (0.53)	0.42 (0.44)

Kang Ning, Damian Fermin, and Alexey I. Nesvizhskii J Proteome Res. 2012 April 6; 11(4): 2261–2271.

Hard to predict protein levels from mRNA



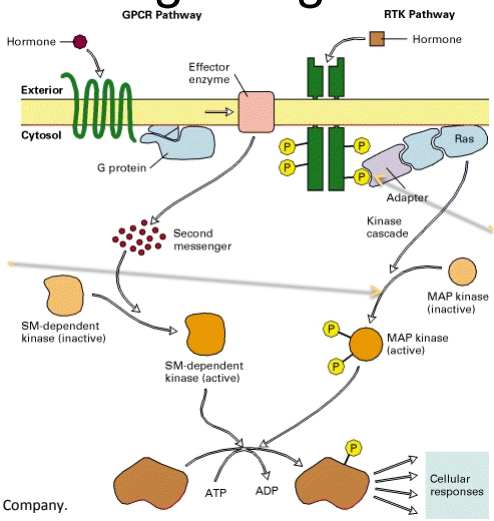
Nature. 2011 May 19;473(7347):337-42. doi: 10.1038/nature10098.
Global quantification of mammalian gene expression control.
Schwanhäusser B1, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M.

Statistics requantitates the central dogma

Transcription, not translation, chieflv determines protein abundance in mammals

SCIENCE 6 MARCH 2015 • VOL 347 ISSUE 6226

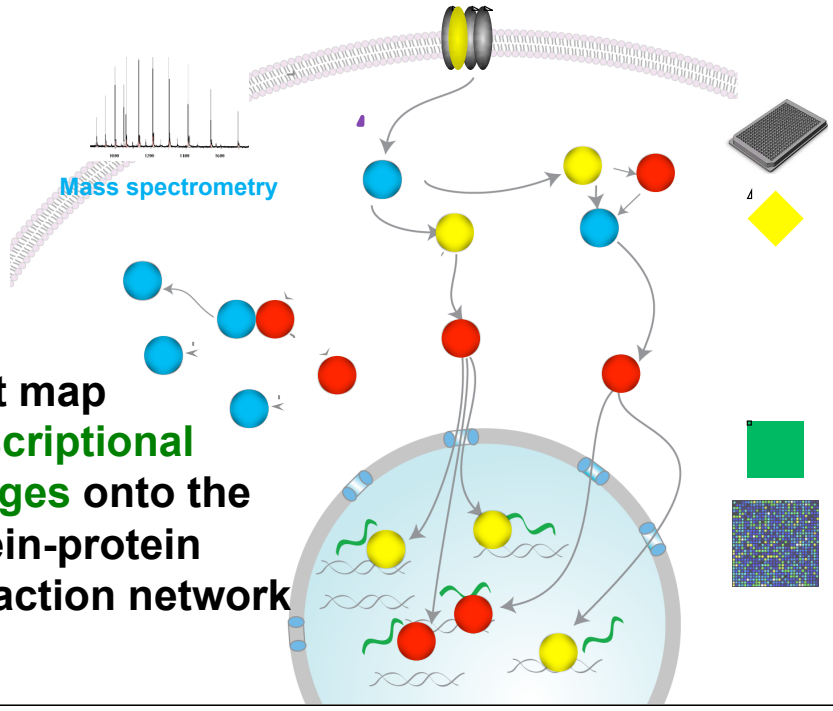
.. and mRNA levels do not reveal signaling



Phosphorylation
- **kinase**
- **phosphatase**

Protein-protein interactions

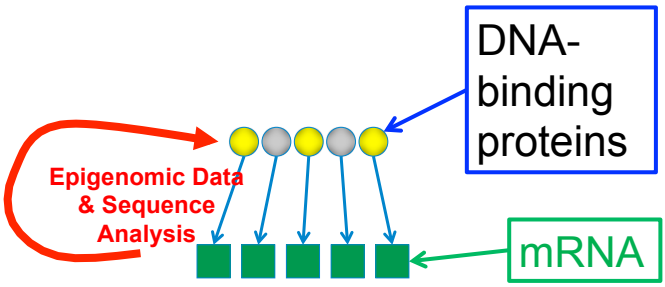
Enzymes
Microtubules
Histones
Transcription factors



Don't map transcriptional changes onto the protein-protein interaction network

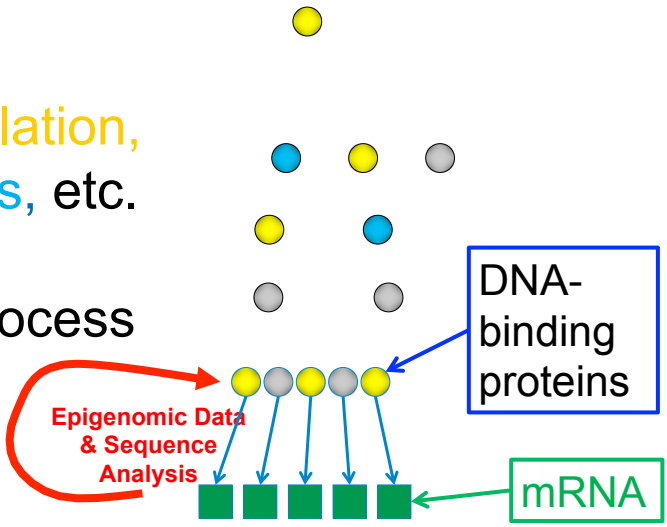
Step 1

Use expression data to find upstream signaling changes



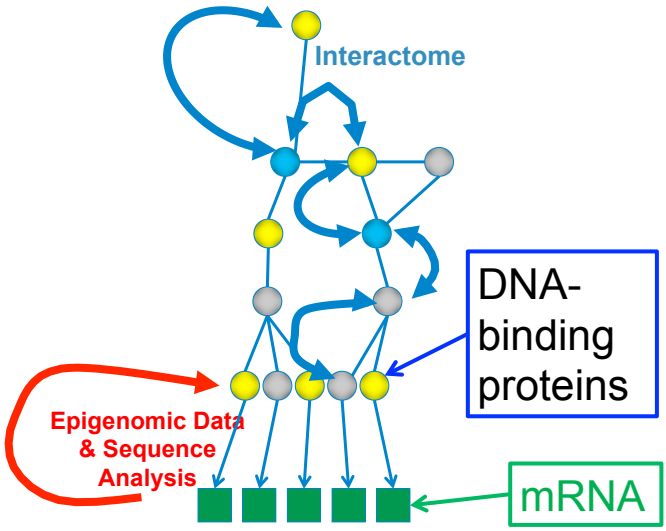
Step 2

Identify phosphorylation, metabolites, etc. relevant to disease process

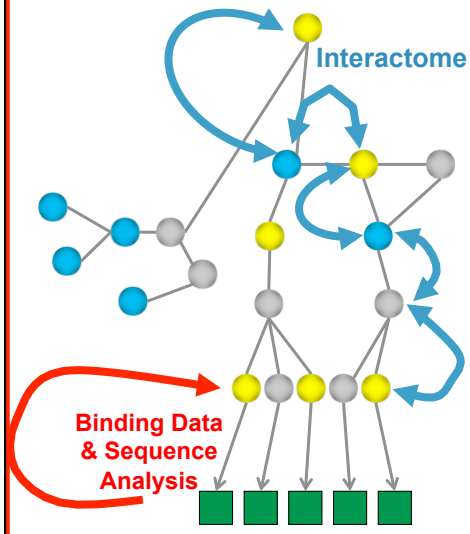


Step 3

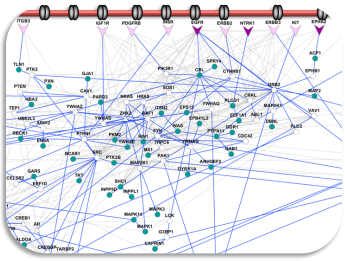
Network integration



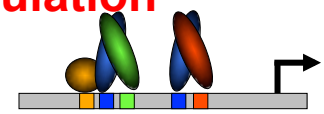
Overall Approach



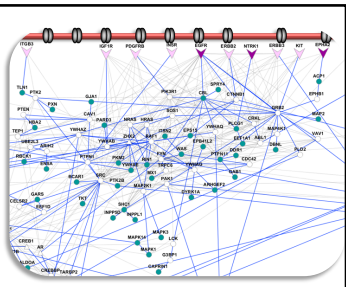
Model Networks



Map Transcriptional Regulation



Outline

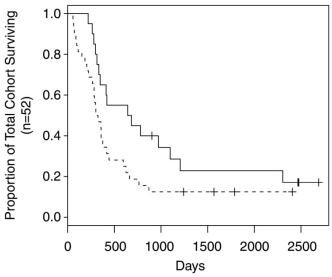


- Network Modeling
  - Approach
  - Predicting in vitro targets using RNA-Seq, DNase-Seq, and phospho-proteomics

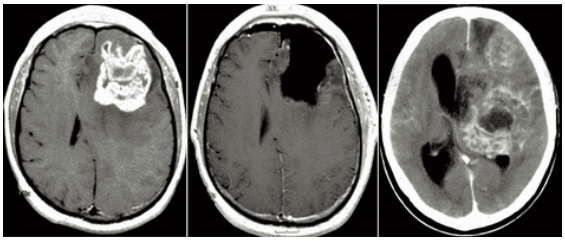
Carol Huang

Post-doctoral Associate
Salk Institute

Linking Proteomic and Transcriptional Data through the Interactome and Epigenome Reveals a Map of Oncogene-induced Signaling
PLOS Computational Biology, 2013



Pope W B et al. Radiology 2008;249:268-277

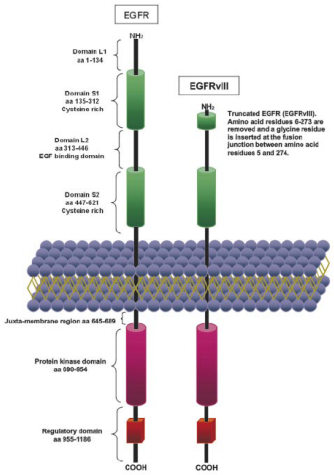


Presentation Post-op Recurrence

Weil RJ (2006) PLoS Med 3(1): e31.

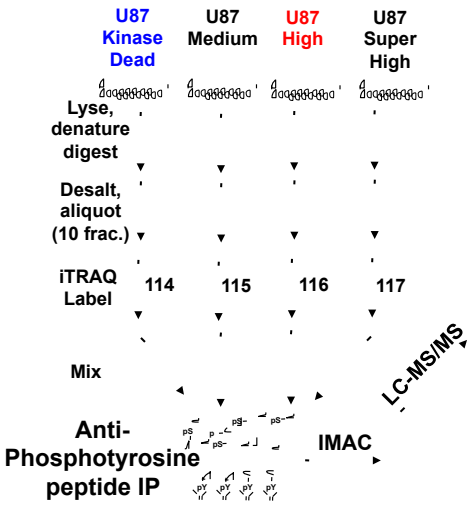
EGFR variant III mutation

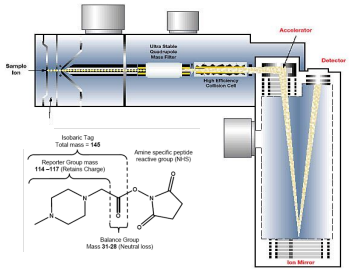
- Most common EGFR mutation in human cancer
- 60% of GBMs and 20% of anaplastic astrocytomas, also in lung, breast and prostate tumors
- EGFRvIII expression correlates with shorter life expectancies

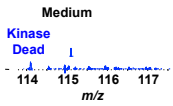


Pedersen et al., *Ann Oncol.*, 2001

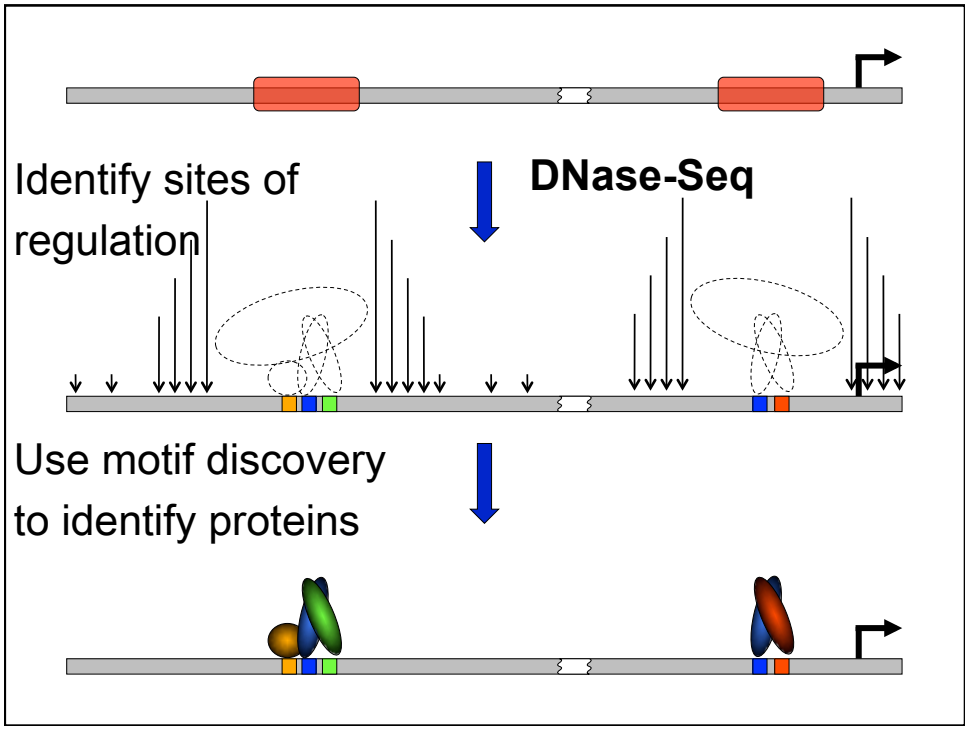
Compare Tyrosine Phosphorylation by Mass Spec White Lab, MIT

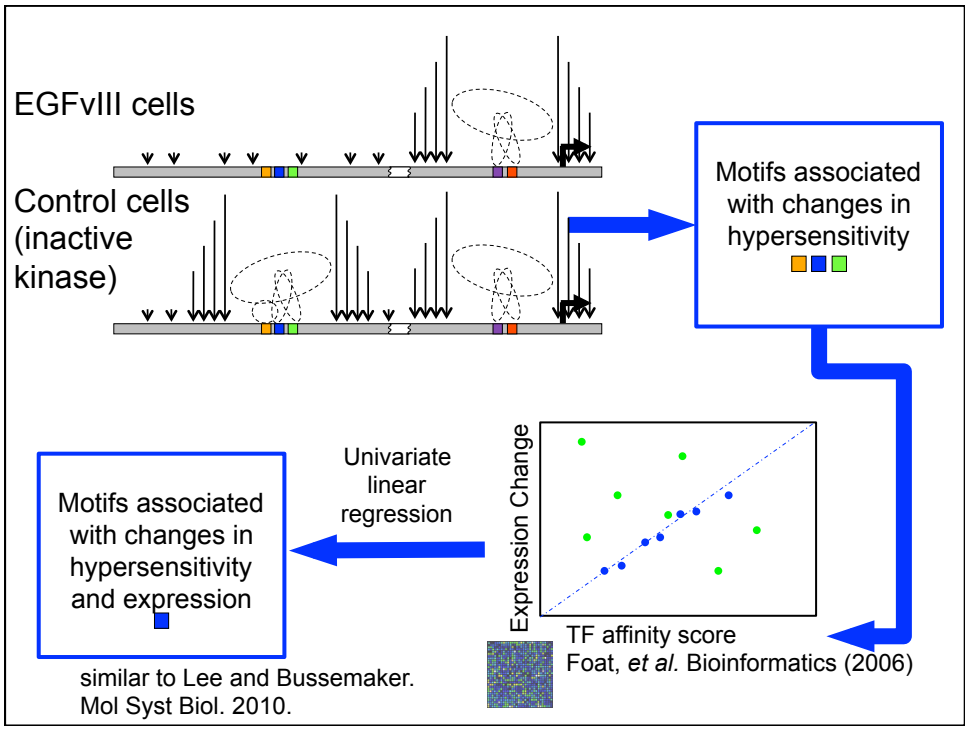




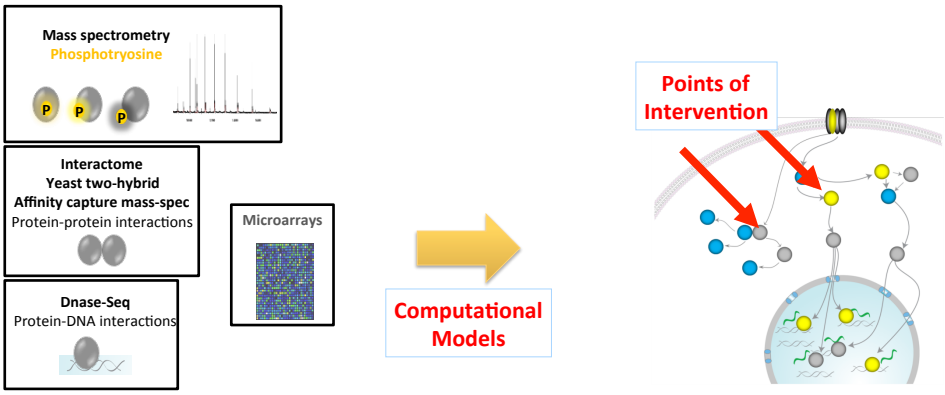


Huang *et al.*, *PNAS* 2007

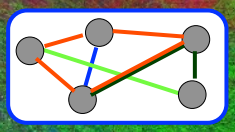

Identify sites of regulation
DNase-Seq
Use motif discovery to identify proteins


EGFvIII cells
Control cells (inactive kinase)
Motifs associated with changes in hypersensitivity
Univariate linear regression
Motifs associated with changes in hypersensitivity and expression
Expression Change
TF affinity score
Foat, *et al.* Bioinformatics (2006)
similar to Lee and Bussemaker. Mol Syst Biol. 2010.

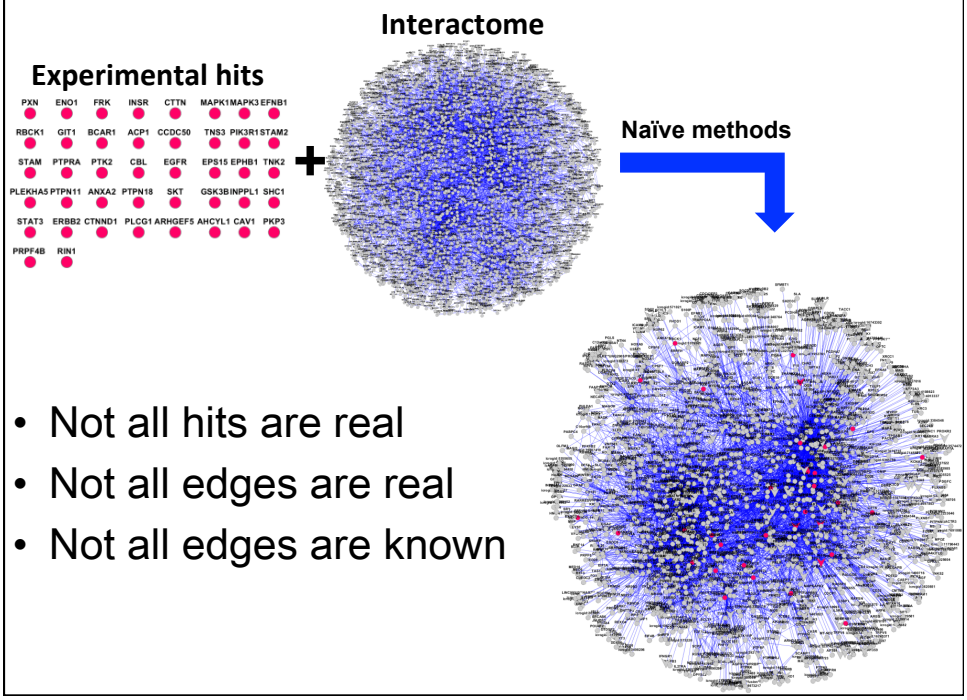
Assembling a EGFR vIII signaling network



Interactome: Known *physical* protein-protein interactions



24,039 nodes
234,455 edges



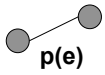
- Not all hits are real
- Not all edges are real
- Not all edges are known

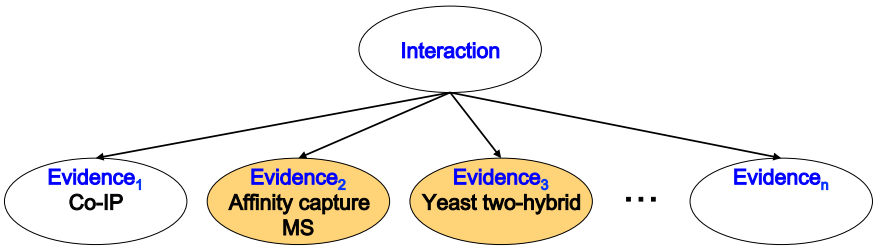
A Better Approach

- Find most probable connections.
- Only connect some of the data to each other:
  - False positives in data
  - Missing interactions in interactome

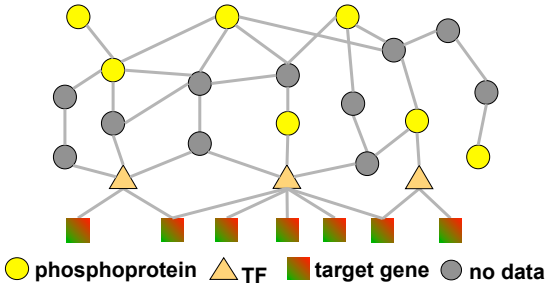
Probabilities Reflect Underlying Data

- Naïve Bayes approach integrates multiple types of evidence
  (Jansen *et al.* Science 2003; Myers et al. Genome Biology 2005.)
- Reliability determined by:
  - higher confidence experimental techniques
  - multiple experimental techniques

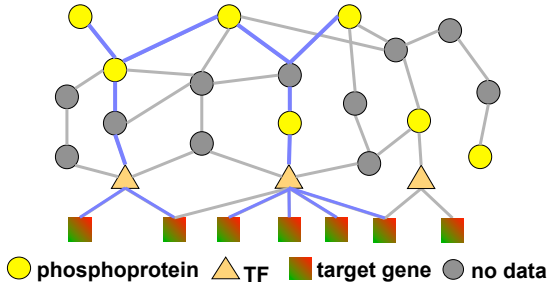




Prize-collecting Steiner Tree

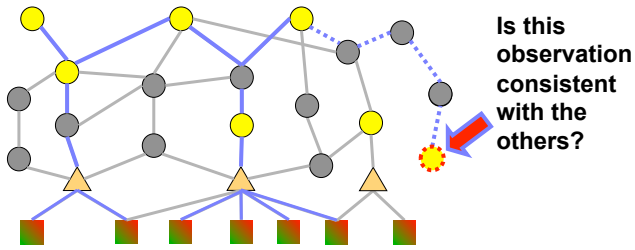


Prize-collecting Steiner Tree



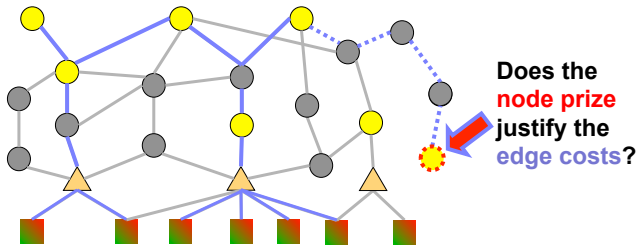
1. **Don't Include All Data**
2. **Avoid Unlikely Interactions**

Prize-collecting Steiner Tree



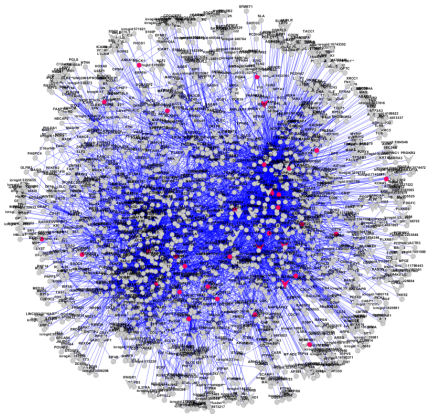
1. **Don't Include All Data**
2. **Avoid Unlikely Interactions**

Prize-collecting Steiner Tree

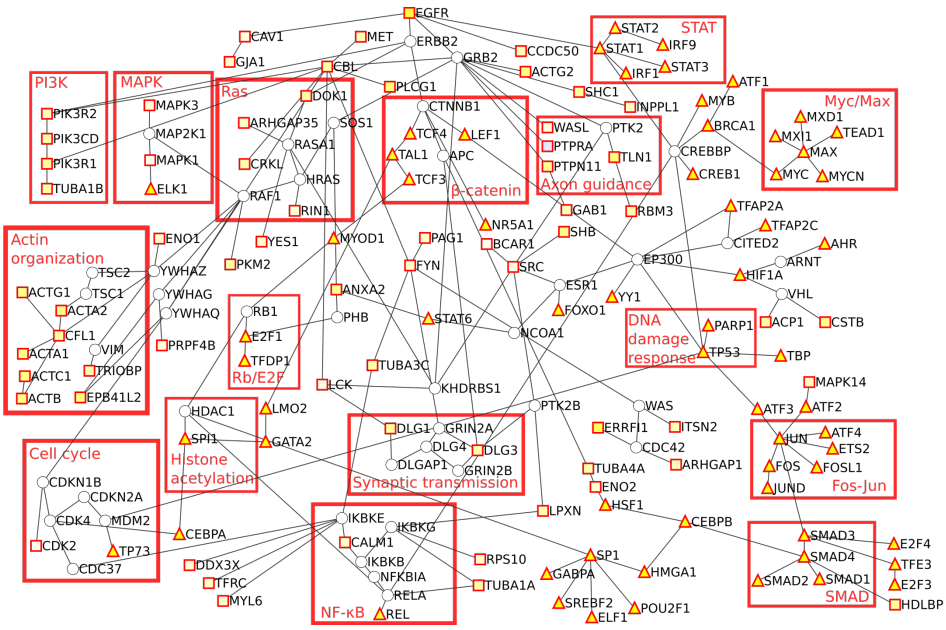


$$\sum_{v \text{ not in } T} \beta \text{prize}(v) + \sum_{e \text{ in } T} \text{cost}(e)$$

Naïve Methods

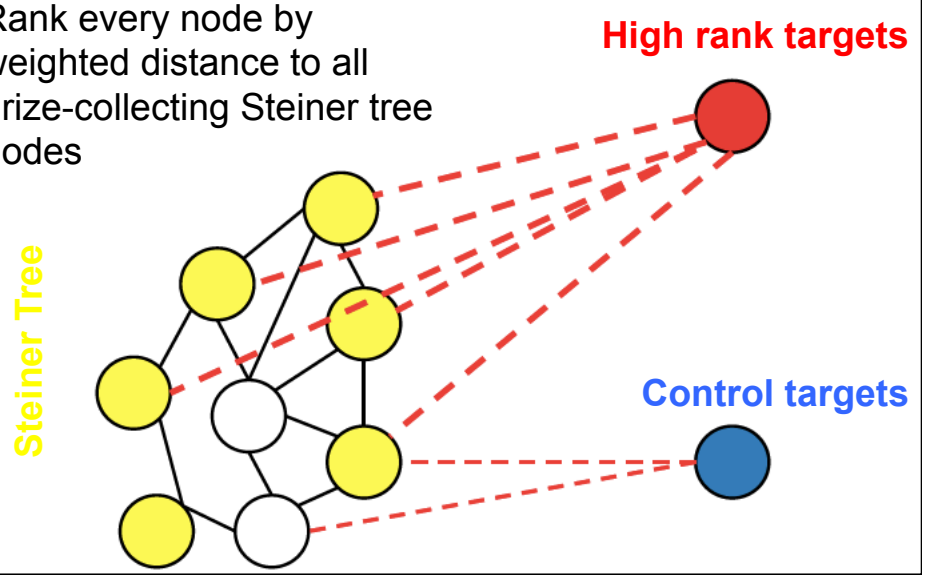


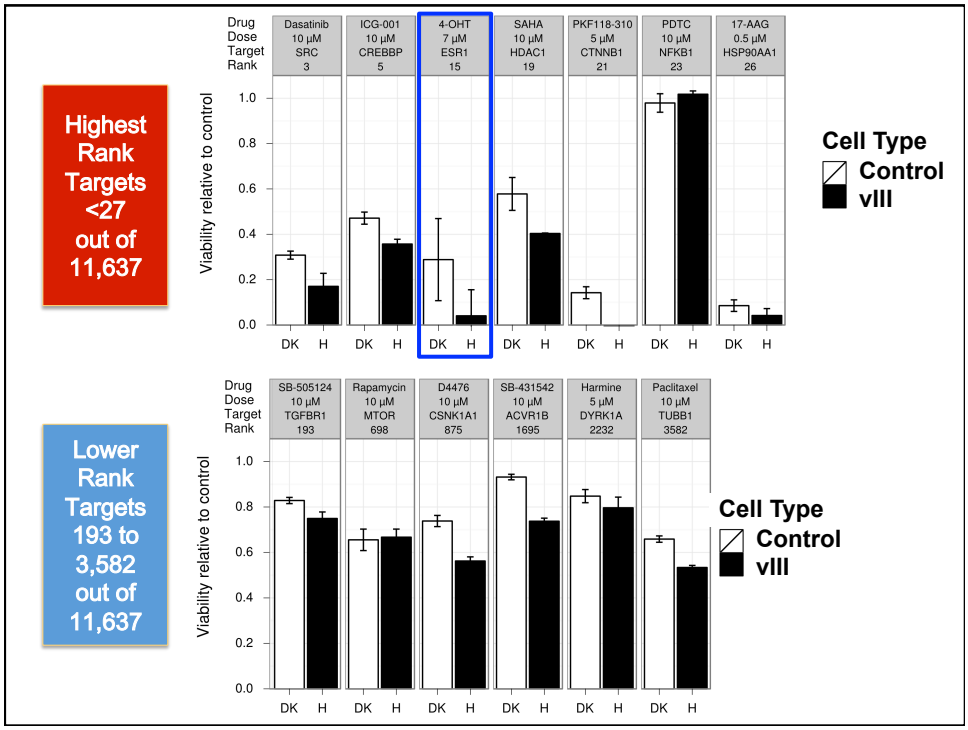
- >2,500 nearest neighbors of phosphoproteins
- >4,500 nearest neighbors of phosphoproteins +transcription factors

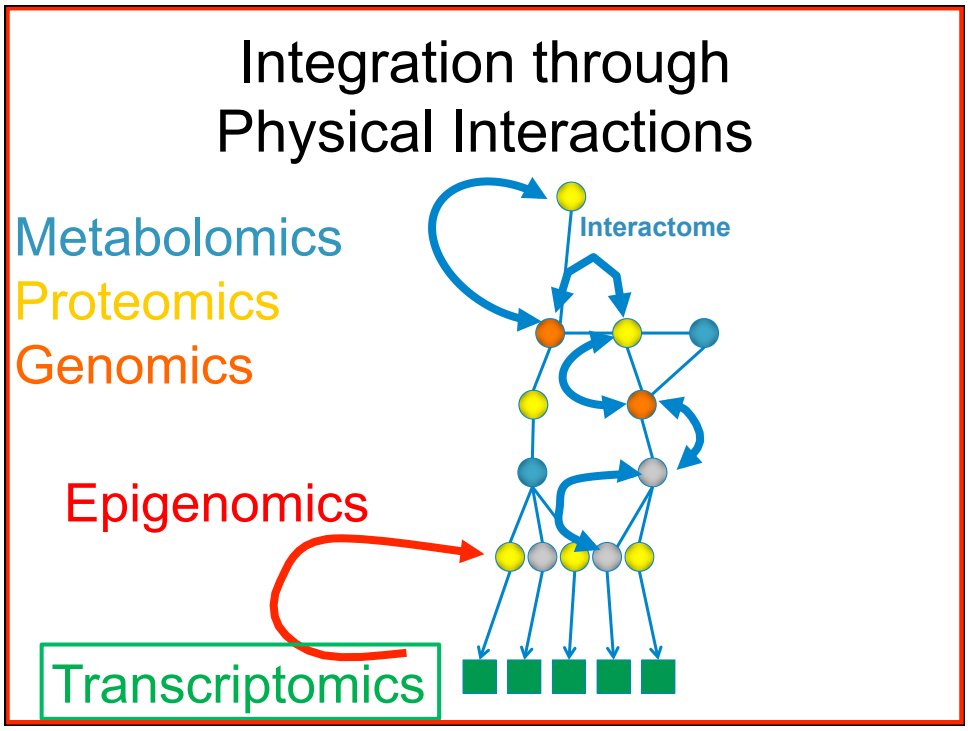


Can we find drug targets?

Rank every node by weighted distance to all prize-collecting Steiner tree nodes




Highest Rank Targets <27 out of 11,637
Lower Rank Targets 193 to 3,582 out of 11,637
Drug
Dose
Target
Rank
Dasatinib 10 μM SRC 3
ICG-001 10 μM CREBBP 5
4-OHT 7 μM ESR1 15
SAHA 10 μM HDAC1 19
PKF118-310 5 μM CTNNB1 21
PDTC 10 μM NFKB1 23
17-AAG 0.5 μM HSP90AA1 26
SB-505124 10 μM TGFBR1 193
Rapamycin 10 μM MTOR 698
D4476 10 μM CSNK1A1 875
SB-431542 10 μM ACVR1B 1695
Harmine 5 μM DYRK1A 2232
Paclitaxel 10 μM TUBB1 3582
Viability relative to control
1.0
0.8
0.6
0.4
0.2
0.0
DK
H
Cell Type
Control
vIII


Integration through Physical Interactions
Metabolomics
Proteomics
Genomics
Epigenomics
Transcriptomics
Interactome

Known Components

Genome/ Proteome-wide

Physical Relationships

Statistical Relationships

Differential equations

Interactome Models

Boolean logic, decision trees

Bayesian networks

mutual information

regression, clustering

Known Components

Genome/ Proteome-wide

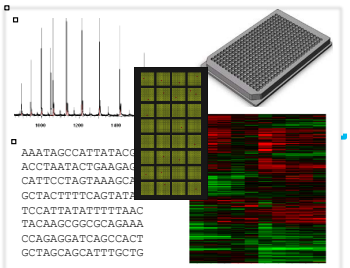
Physical Relationships

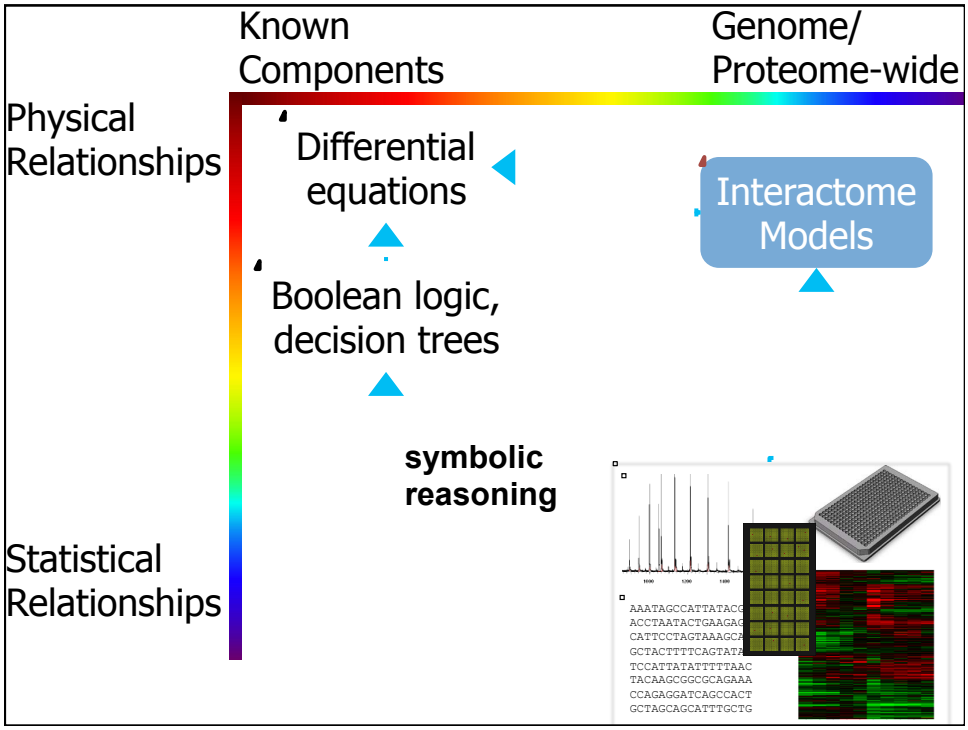
Statistical Relationships

Differential equations

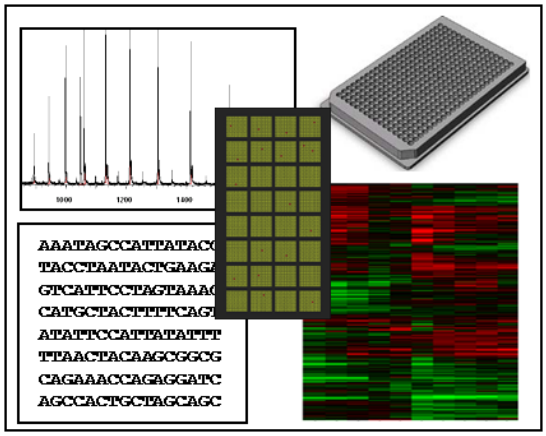
Interactome Models

Boolean logic, decision trees

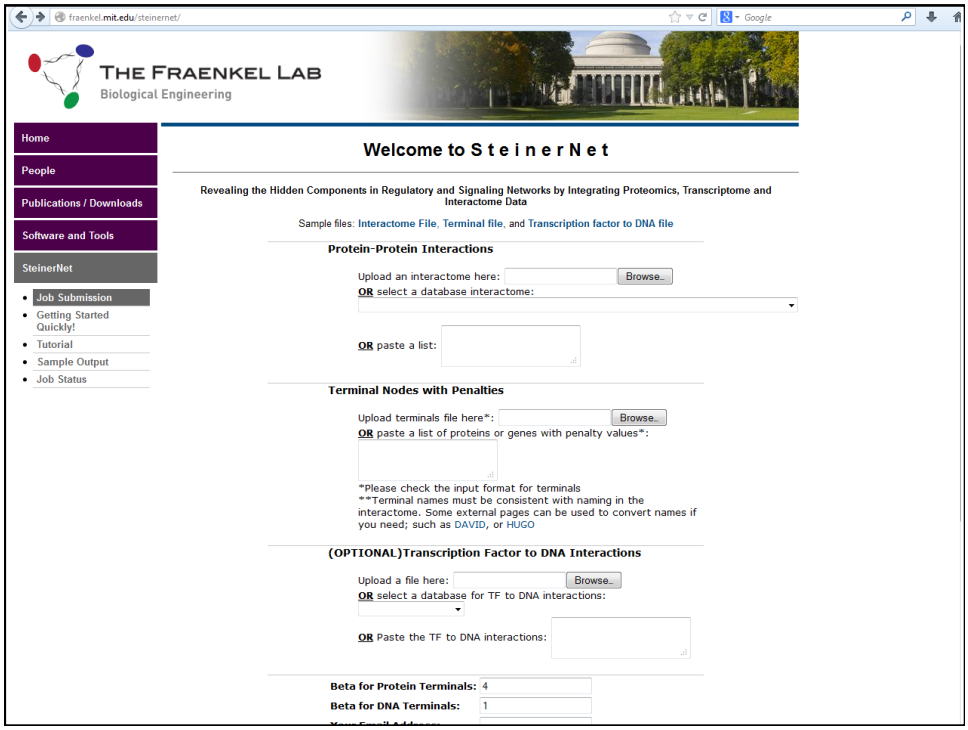


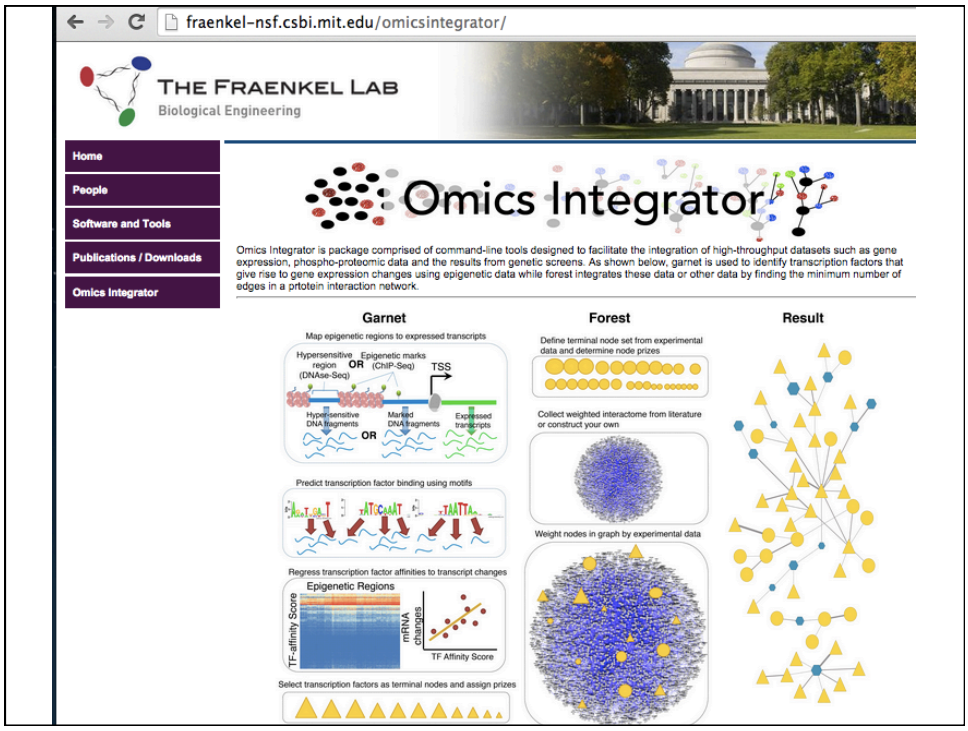


Embracing the change



108


fraenkel.mit.edu/steinernet/
THE FRAENKEL LAB
Biological Engineering
Home
People
Publications / Downloads
Software and Tools
SteinerNet
Job Submission
Getting Started Quickly!
Tutorial
Sample Output
Job Status
Welcome to S t e i n e r N e t
Revealing the Hidden Components in Regulatory and Signaling Networks by Integrating Proteomics, Transcriptome and Interactome Data
Sample files: Interactome File, Terminal file, and Transcription factor to DNA file
Protein-Protein Interactions
Upload an interactome here:
Browse...
OR select a database interactome:
OR paste a list:
Terminal Nodes with Penalties
Upload terminals file here*:
OR paste a list of proteins or genes with penalty values*:
*Please check the input format for terminals
**Terminal names must be consistent with naming in the interactome. Some external pages can be used to convert names if you need; such as DAVID, or HUGO
(OPTIONAL)Transcription Factor to DNA Interactions
Upload a file here:
OR select a database for TF to DNA interactions:
OR Paste the TF to DNA interactions:
Beta for Protein Terminals: 4
Beta for DNA Terminals: 1


fraenkel-nsf.csbi.mit.edu/omicsintegrator/
THE FRAENKEL LAB
Biological Engineering
Home
People
Software and Tools
Publications / Downloads
Omics Integrator
Omics Integrator
Omics Integrator is package comprised of command-line tools designed to facilitate the integration of high-throughput datasets such as gene expression, phospho-proteomic data and the results from genetic screens. As shown below, garnet is used to identify transcription factors that give rise to gene expression changes using epigenetic data while forest integrates these data or other data by finding the minimum number of edges in a prtotein interaction network.
Garnet
Forest
Result
Map epigenetic regions to expressed transcripts
Hypersensitive region (DNAse-Seq)
OR
Epigenetic marks (ChIP-Seq)
TSS
Hypersensitive DNA fragments
Marked DNA fragments
Expressed transcripts
Predict transcription factor binding using motifs
Regress transcription factor affinities to transcript changes
Epigenetic Regions
TF-affinity Score
mRNA changes
TF Affinity Score
Select transcription factors as terminal nodes and assign prizes
Define terminal node set from experimental data and determine node prizes
Collect weighted interactome from literature or construct your own
Weight nodes in graph by experimental data

Leslie Thompson

Forest White

David Housman

Jennifer Chayes

Christian Borgs

Riccardo Zecchina