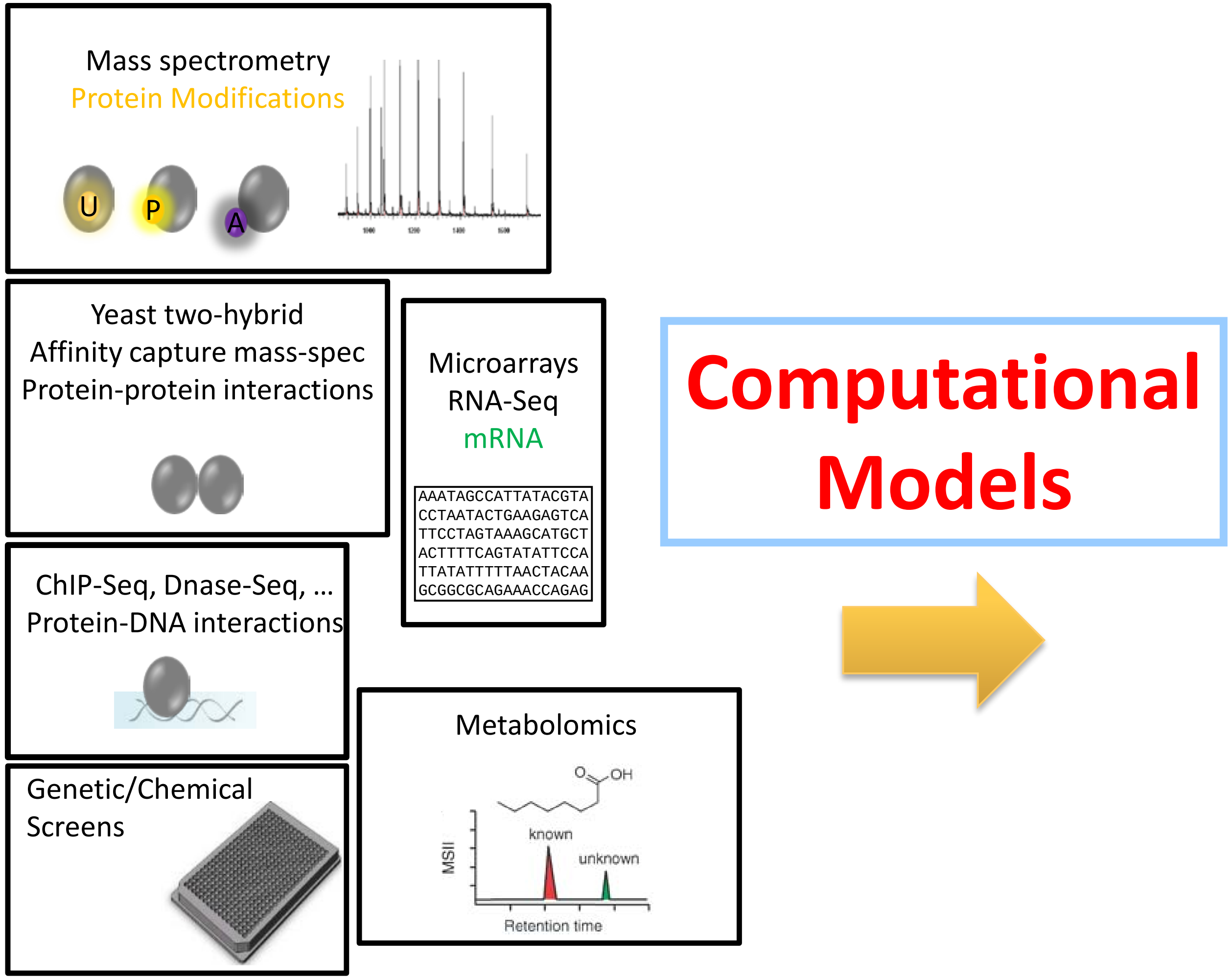
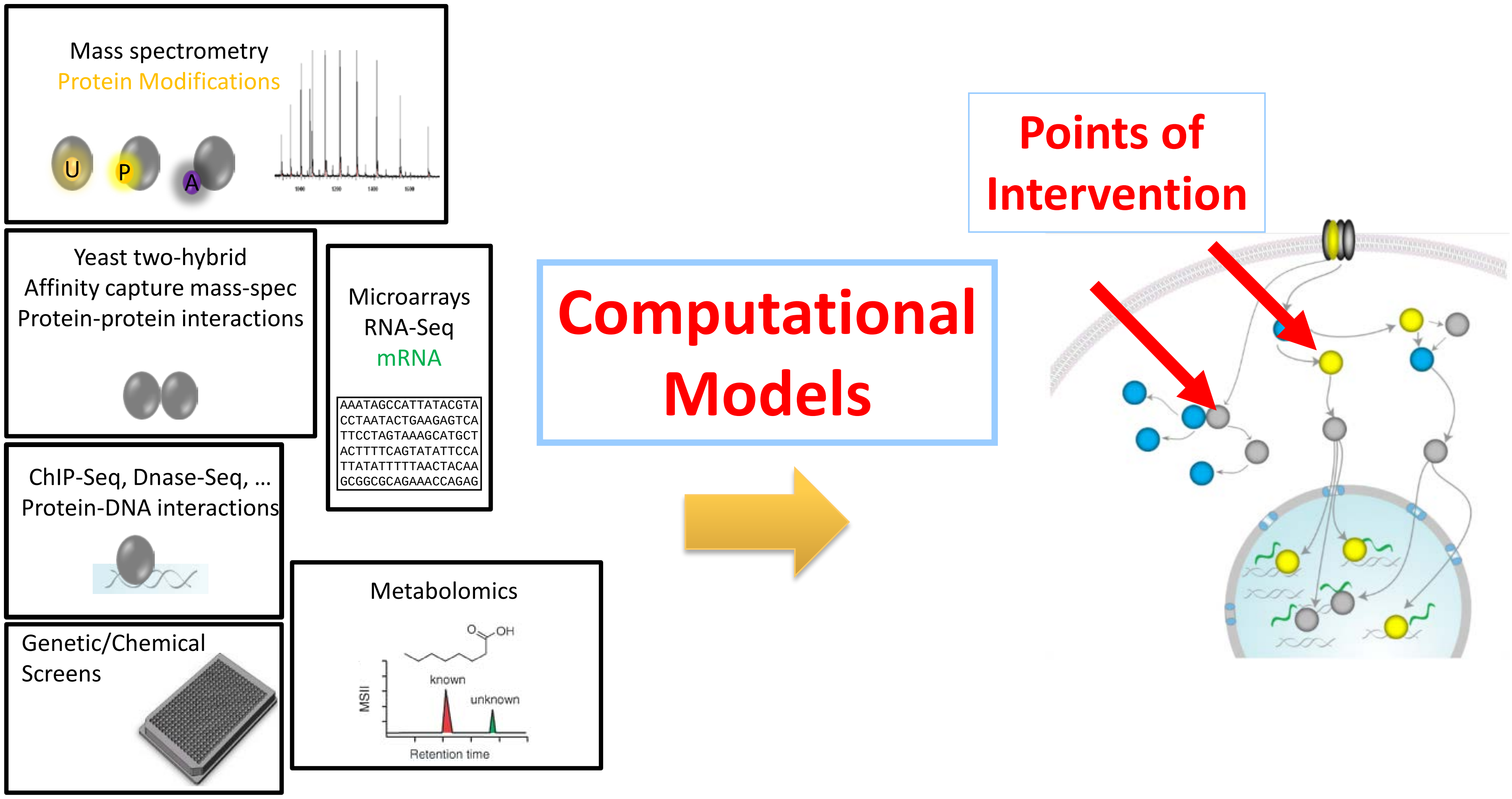
Integrating Multi-Omic Data to Understand Neurodegenerative Diseases

Ernest Fraenkel
Department of Biological Engineering
Massachusetts Institute of Technology



- Basic biology
  - transcription
  - signaling
  - microbiome

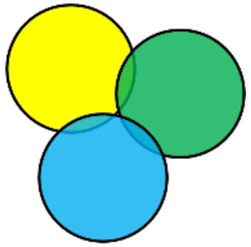
  …
- Tumor classification



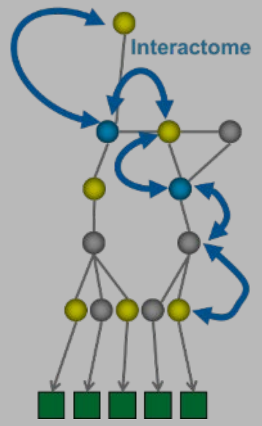
Mass spectrometry
Protein Modifications
U
P
A
Yeast two-hybrid
Affinity capture mass-spec
Protein-protein interactions
Microarrays
RNA-Seq
mRNA
AAATAGCCATTATACGTA
CCTAATACTGAAGAGTCA
TTCCTAGTAAAGCATGCT
ACTTTTCAGTATATTCCA
TTATATTTTTAACTACAA
GCGGCGCAGAAACCAGAG
ChIP-Seq, Dnase-Seq, …
Protein-DNA interactions
Genetic/Chemical
Screens
Metabolomics
O
OH
known
unknown
Retention time
Computational
Models
Points of
Intervention

Outline

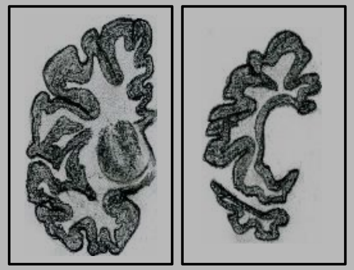
Why Data Integration is Hard



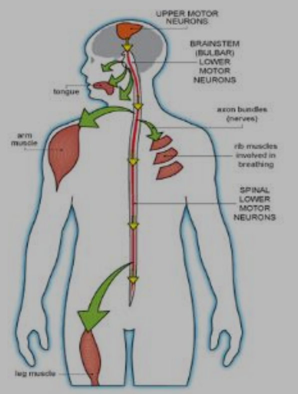
Networks Link the Data



Huntington's Disease



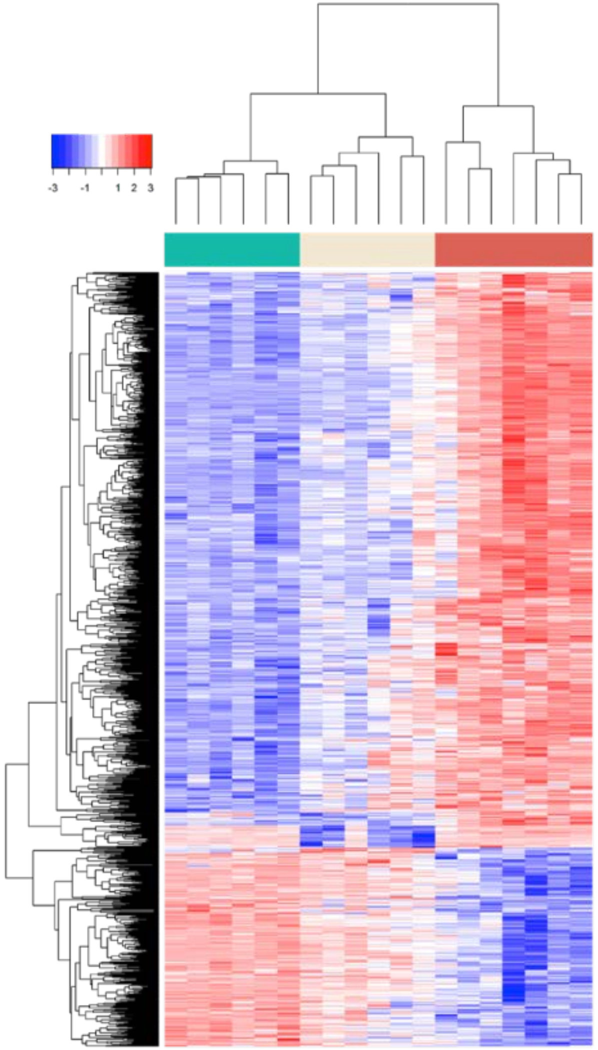
ALS



Standard ways to make sense of "omic" data

- Look for correlations

Cannot distinguish direct from indirect effects

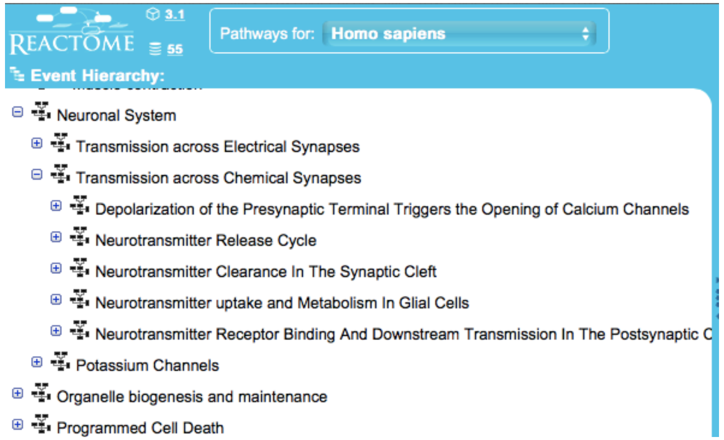


Standard ways to make sense of "omic" data

- Look for correlations

- Compare to known pathways

INGENUITY PATHWAY ANALYSIS



GO the Gene Ontology

MSigDB Molecular Signatures Database

Standard ways to make sense of “omic” data

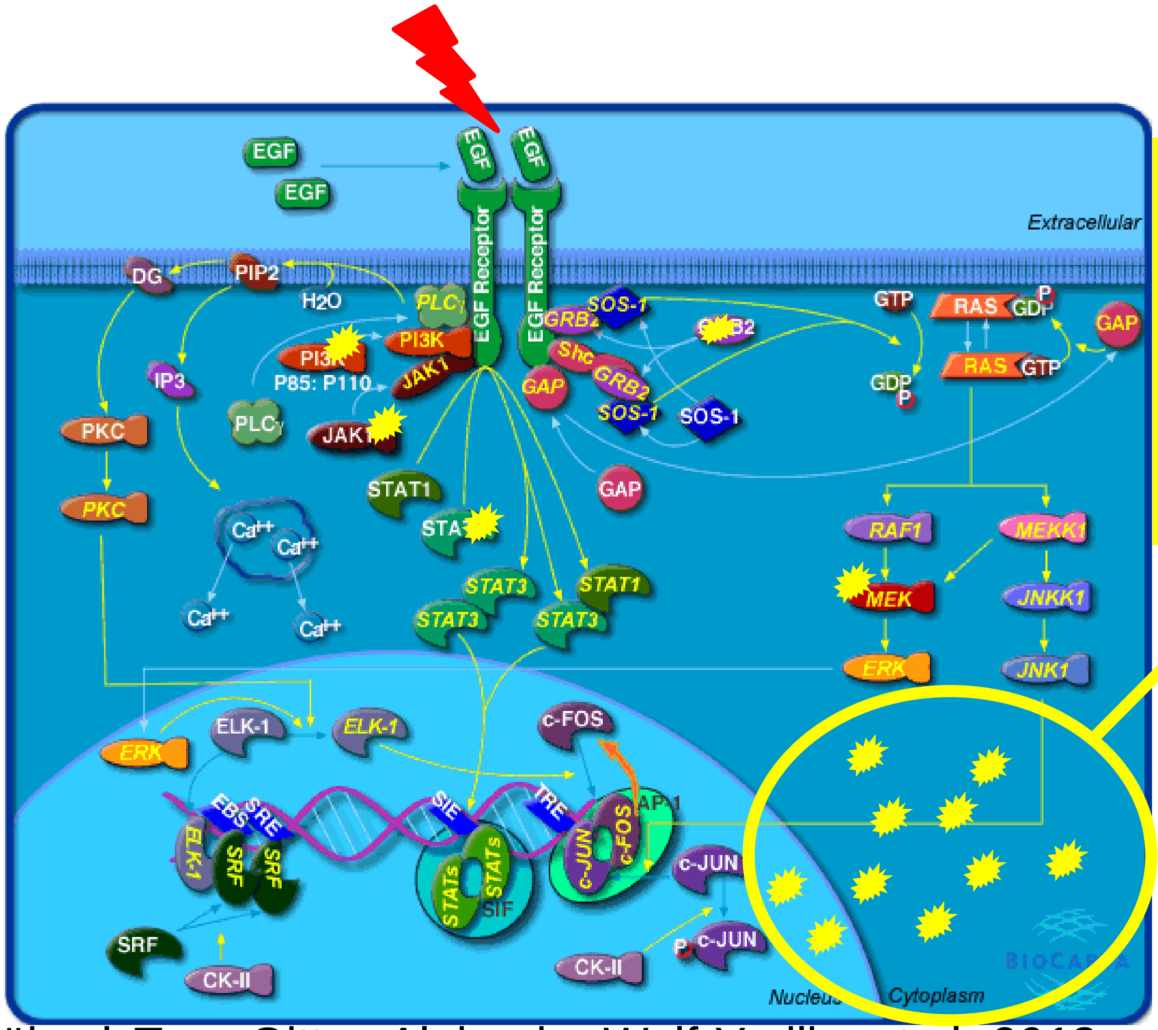
- Look for correlations

Cannot distinguish direct from indirect effects

- Compare to known pathways

Even best-studied ones are mostly unannotated

Most 'Omic Hits Don't Lie in Known Pathways

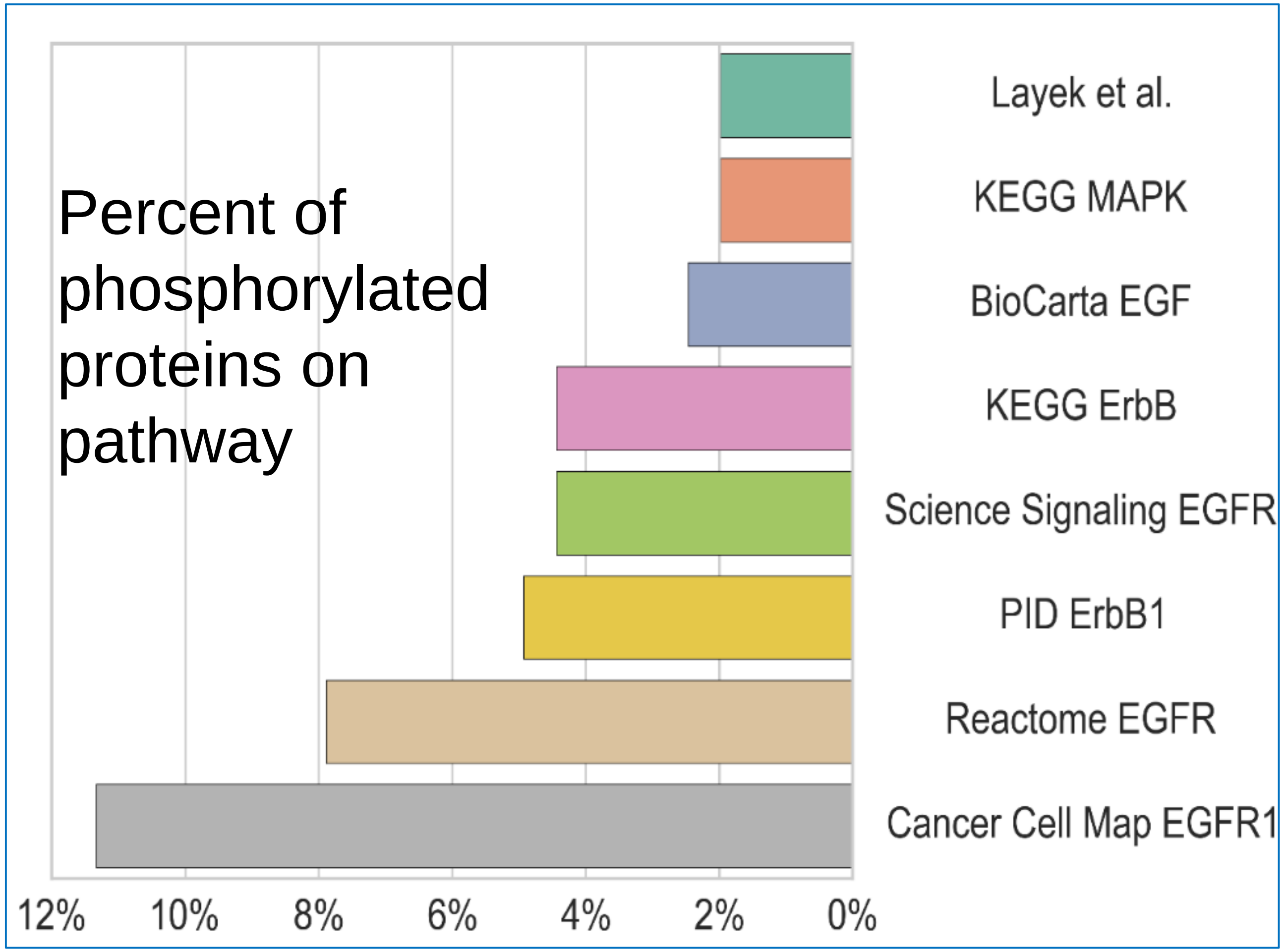


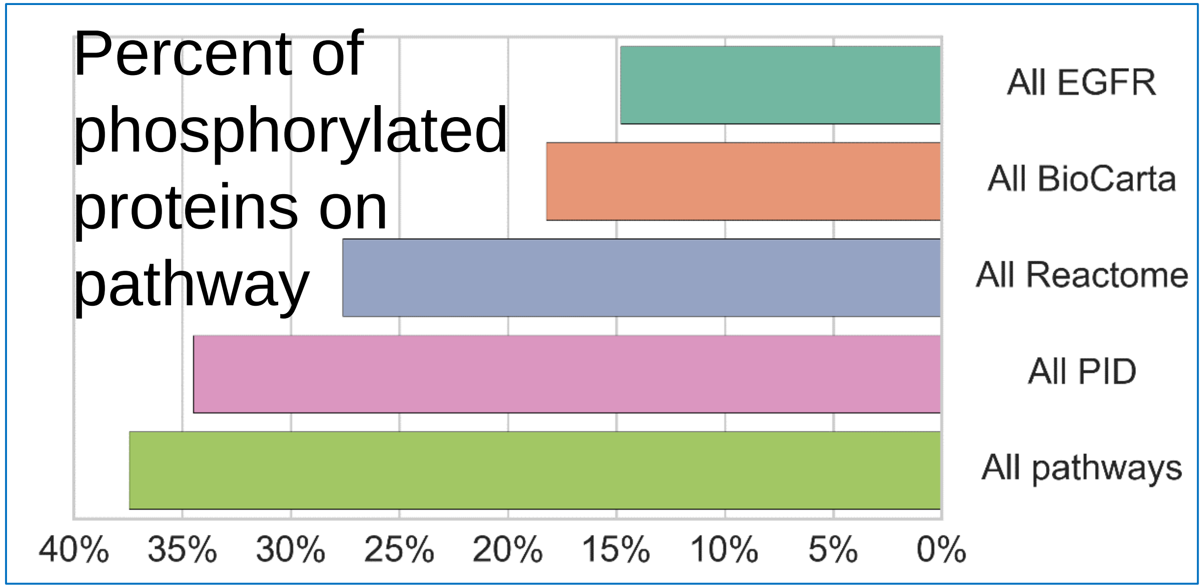
What % of activated proteins are off the pathway?

[Biocarta EGF signaling pathway]

Ali Sinan Köksal, Tony Gitter, Alejandro Wolf-Yadlin, et al. *2018.*

Most responding proteins are not on known pathways

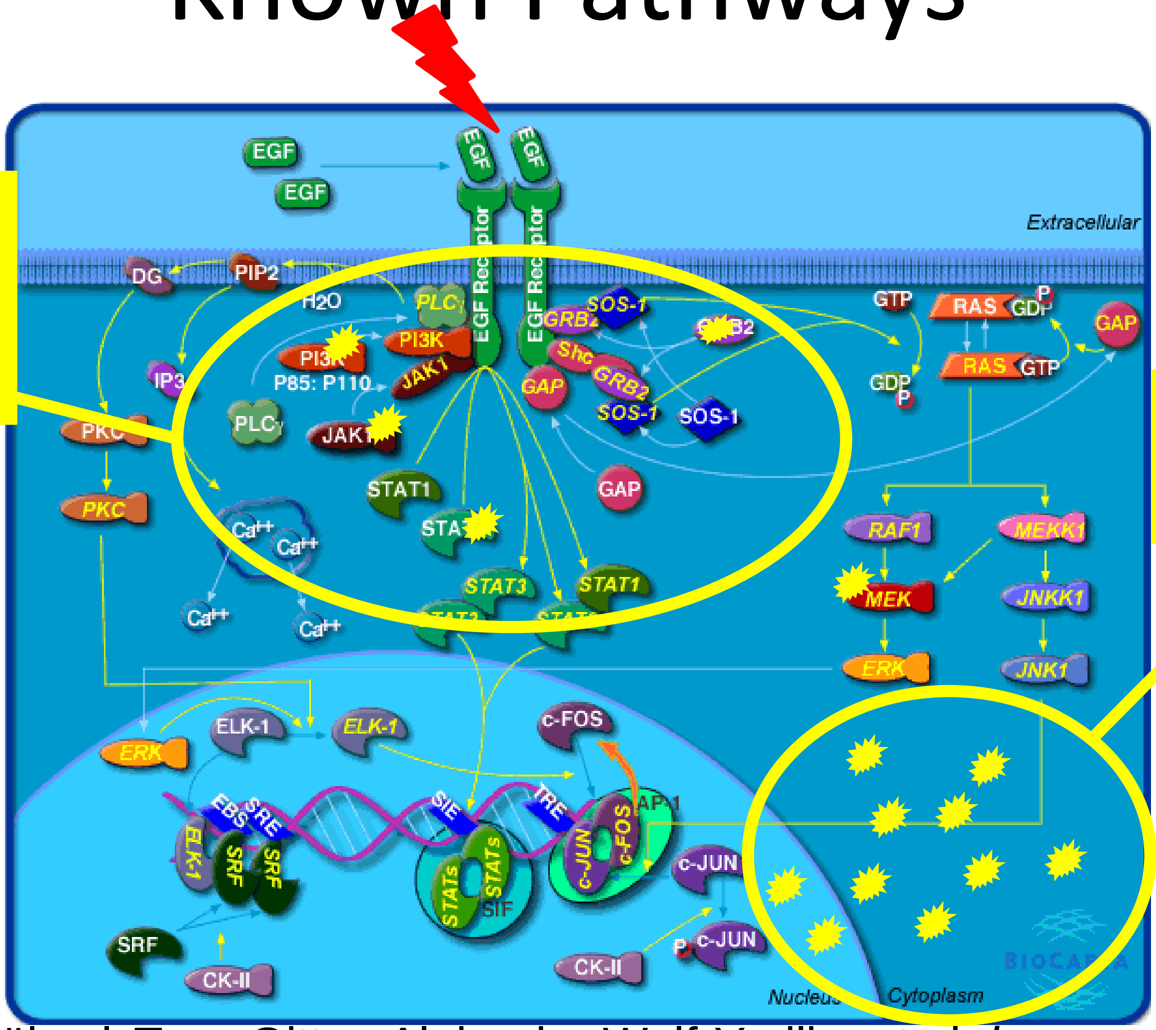




Ali Sinan Köksal, Tony Gitter, Alejandro Wolf-Yadlin, et al. *2018*

Challenge 1:
Most ‘Omic Hits Don’t Lie in Known Pathways

15% on the expected pathway



~85% off pathway

[Biocarta EGF signaling pathway]

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

Standard ways to make sense of "omic" data

- Look for correlations

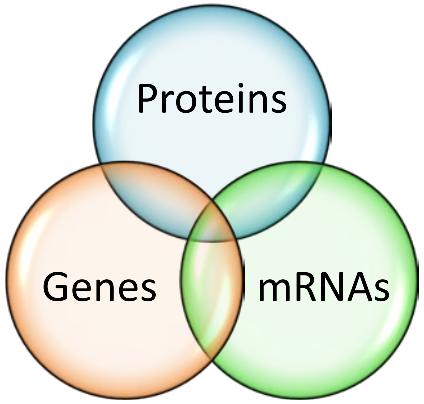
Cannot distinguish direct from indirect effects

- Compare to known pathways

Even best-studied ones are mostly unannotated

- Find overlap among different data types

Overlap is often *less* than expected at random



Esti Yeger-Lotem

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

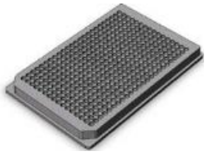
Laura Riva

Team Leader
Center for Genomic Science
Istituto Italiano di Tecnologia
Italy

nature genetics

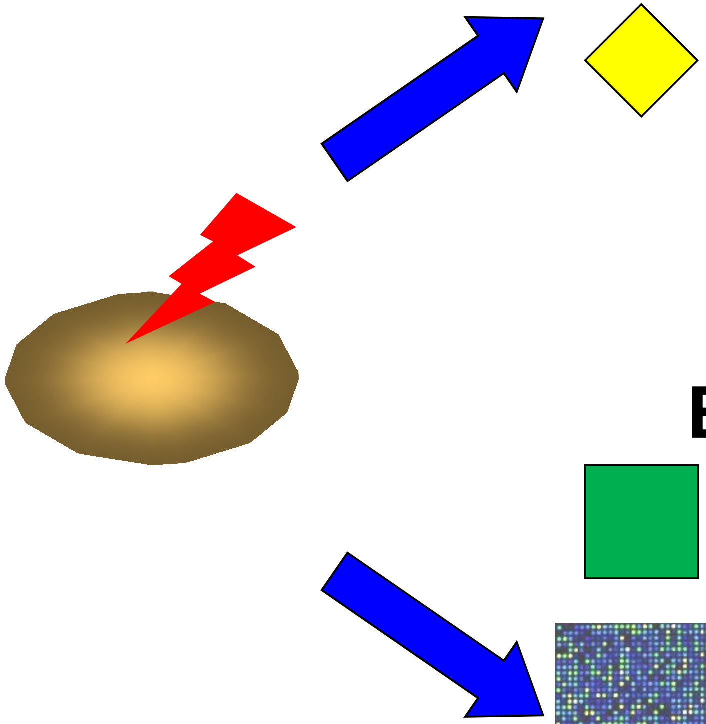
41(3):316-23
doi: 10.1038/ng.337

For 156 perturbations:



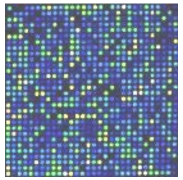
Genetic Data Enriched for:

- Transcriptional regulation
- Signal transduction

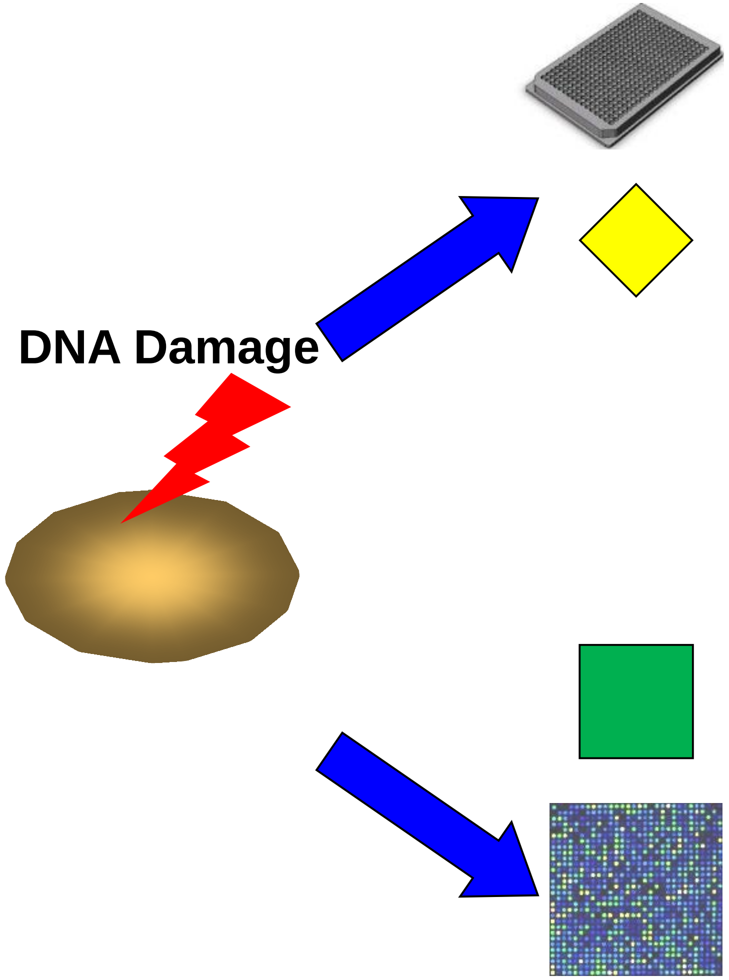


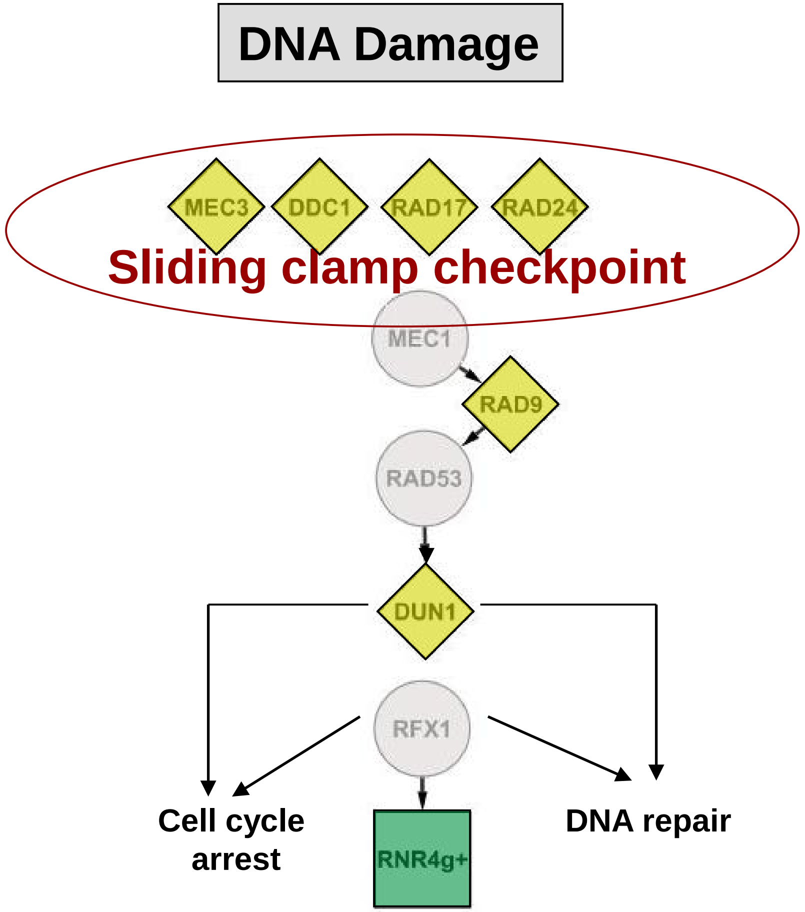
Expression Data Enriched for:

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



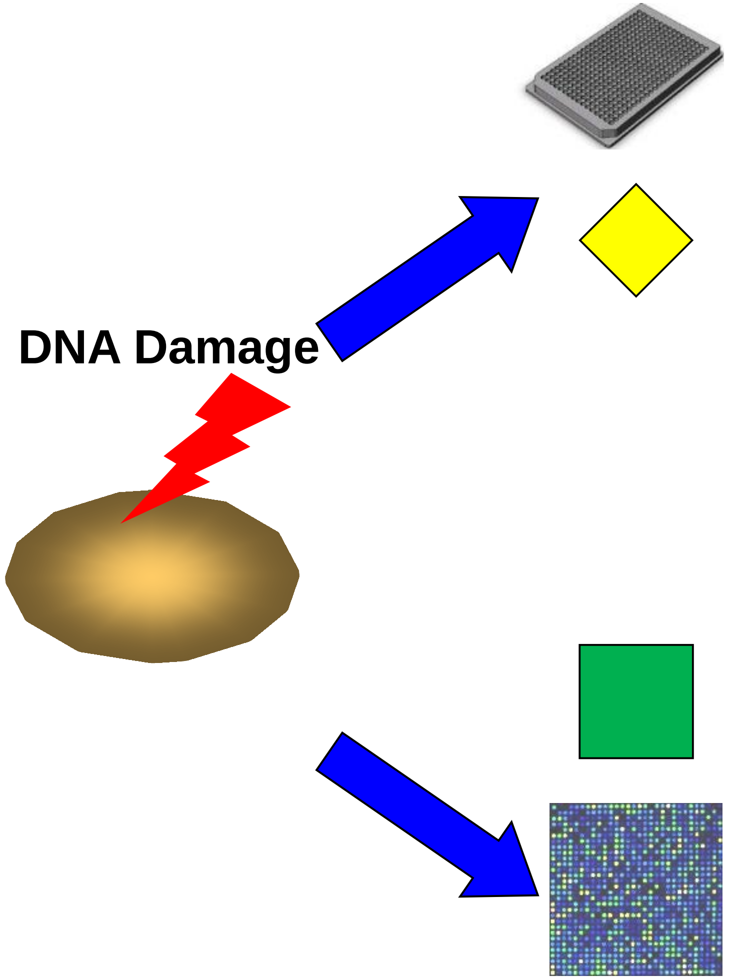
Genetic Data

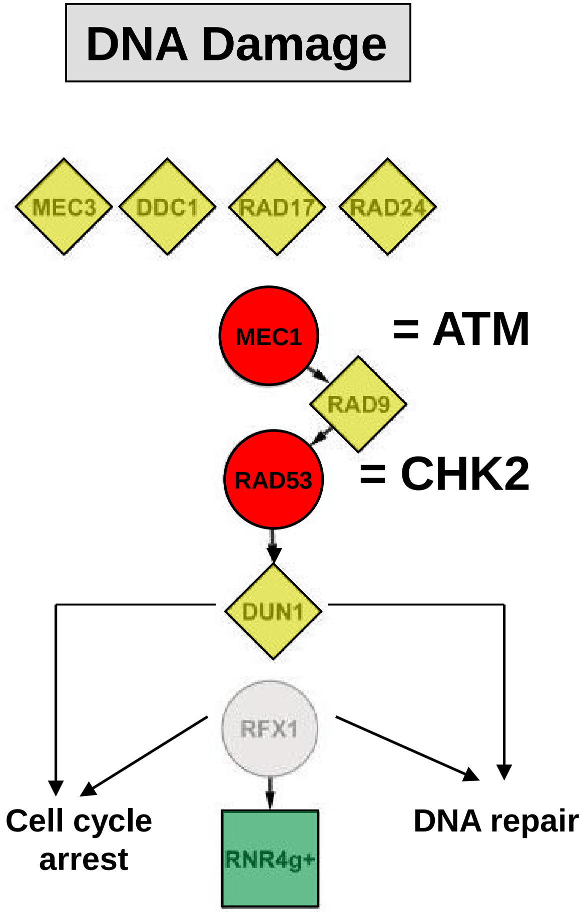




Expression Data

Genetic Data

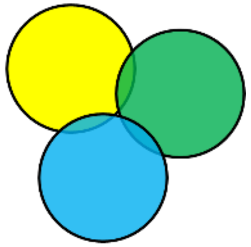




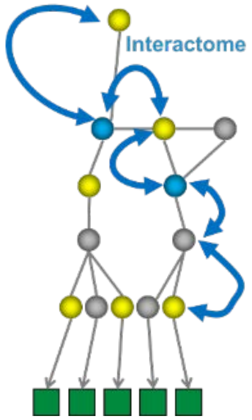
Expression Data

Outline

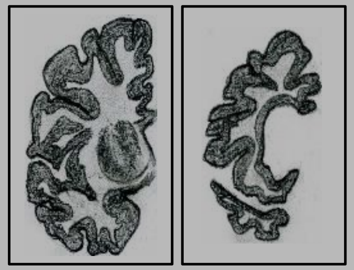
Why Data Integration is Hard



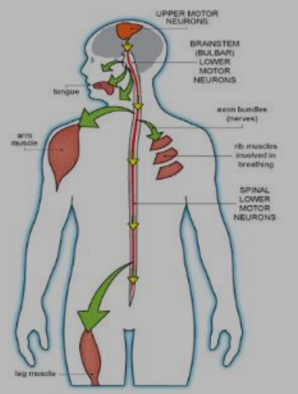
Networks Link the Data

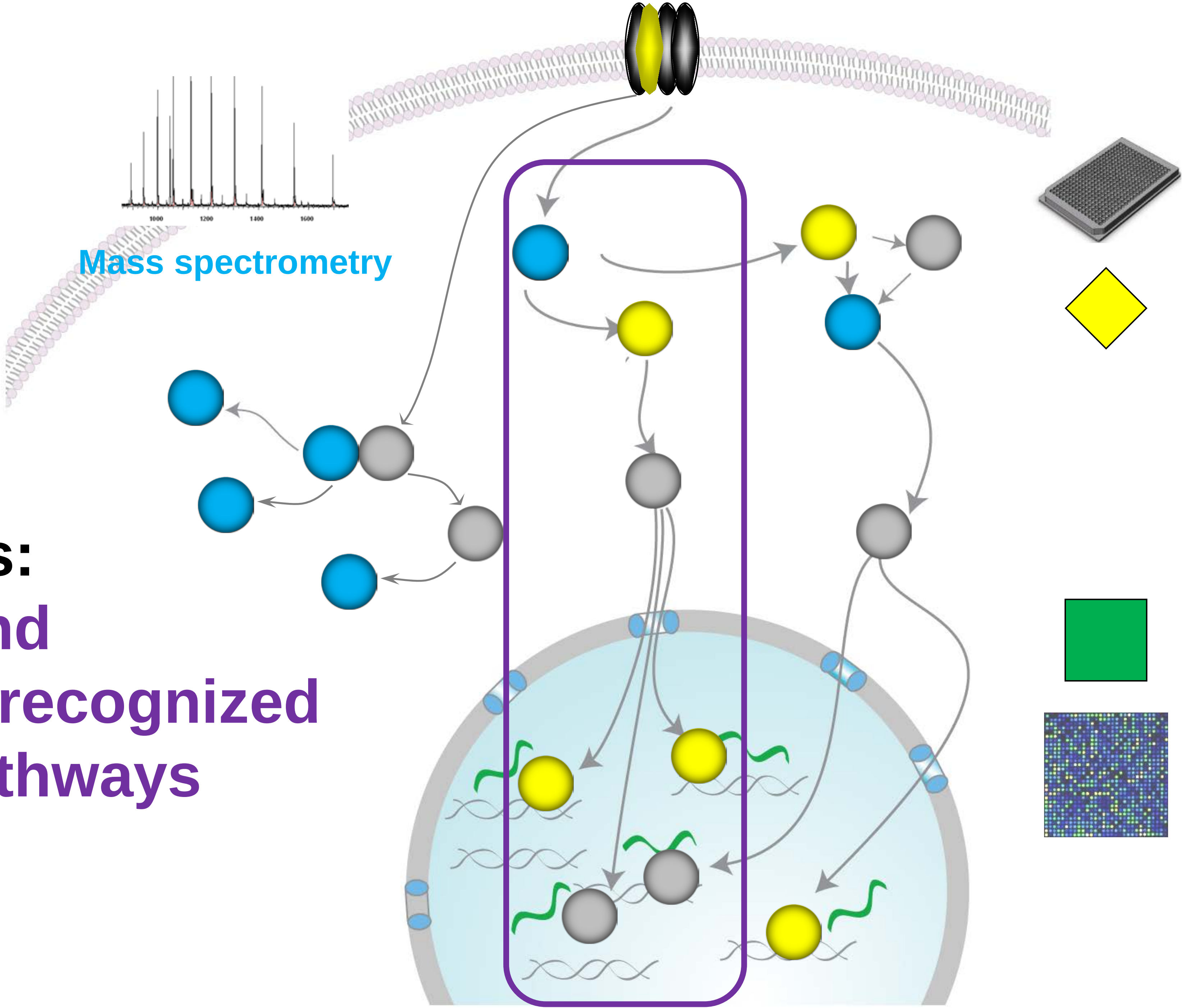


Huntington's Disease



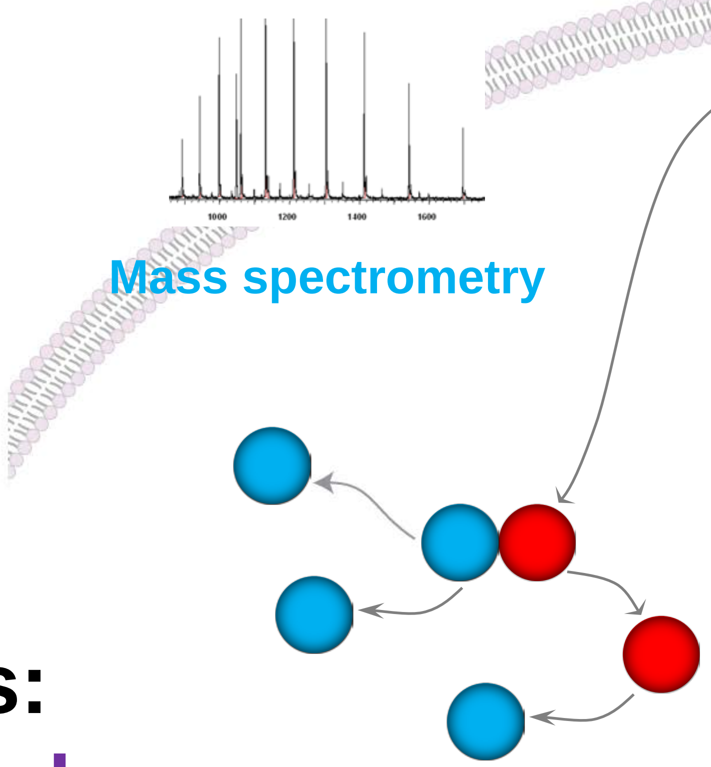
ALS

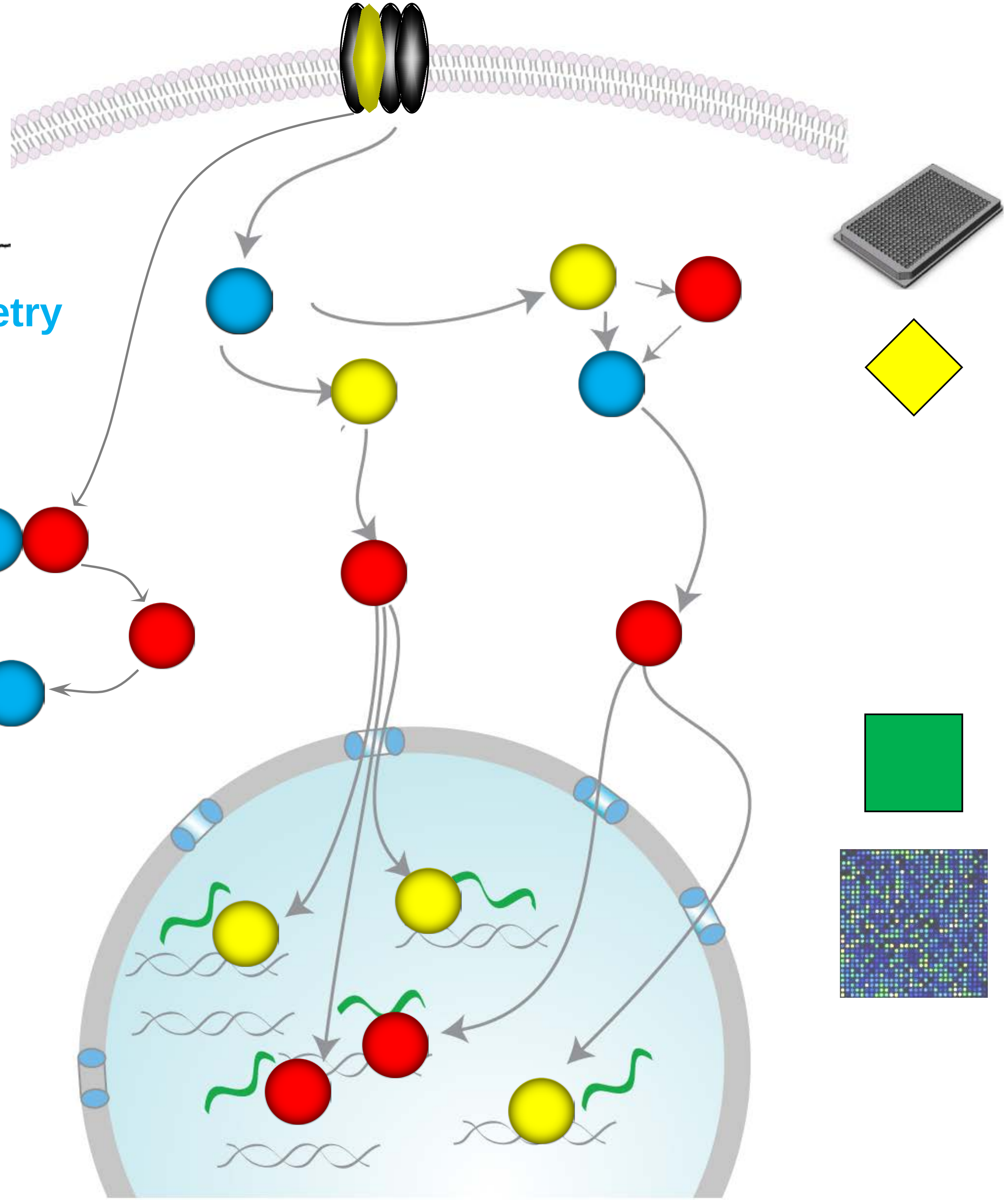




Goals:

1. Find unrecognized pathways



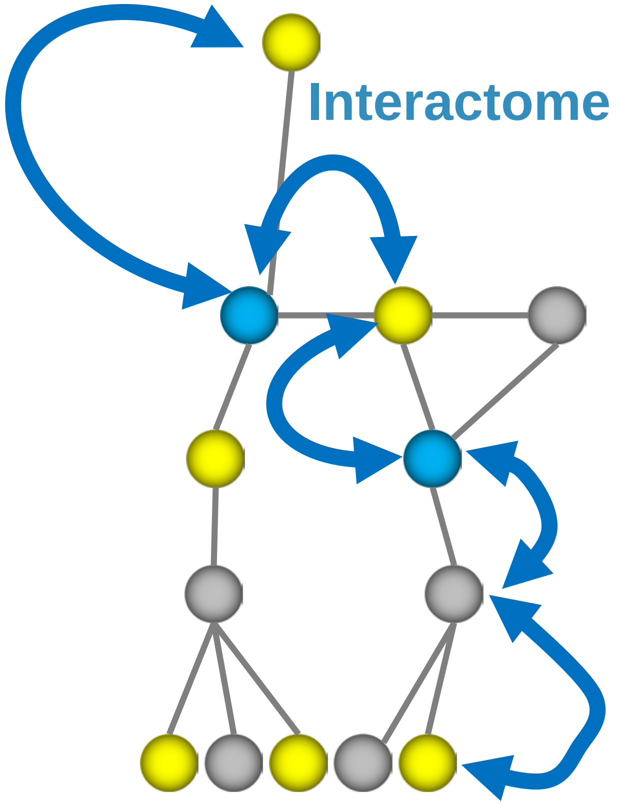


Goals:

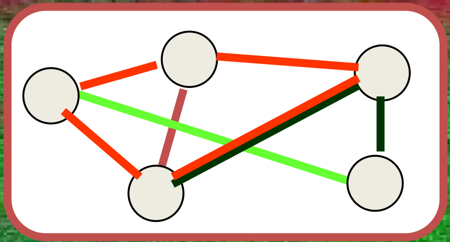
1. Find unrecognized pathways
2. Find “hidden” proteins

Approach

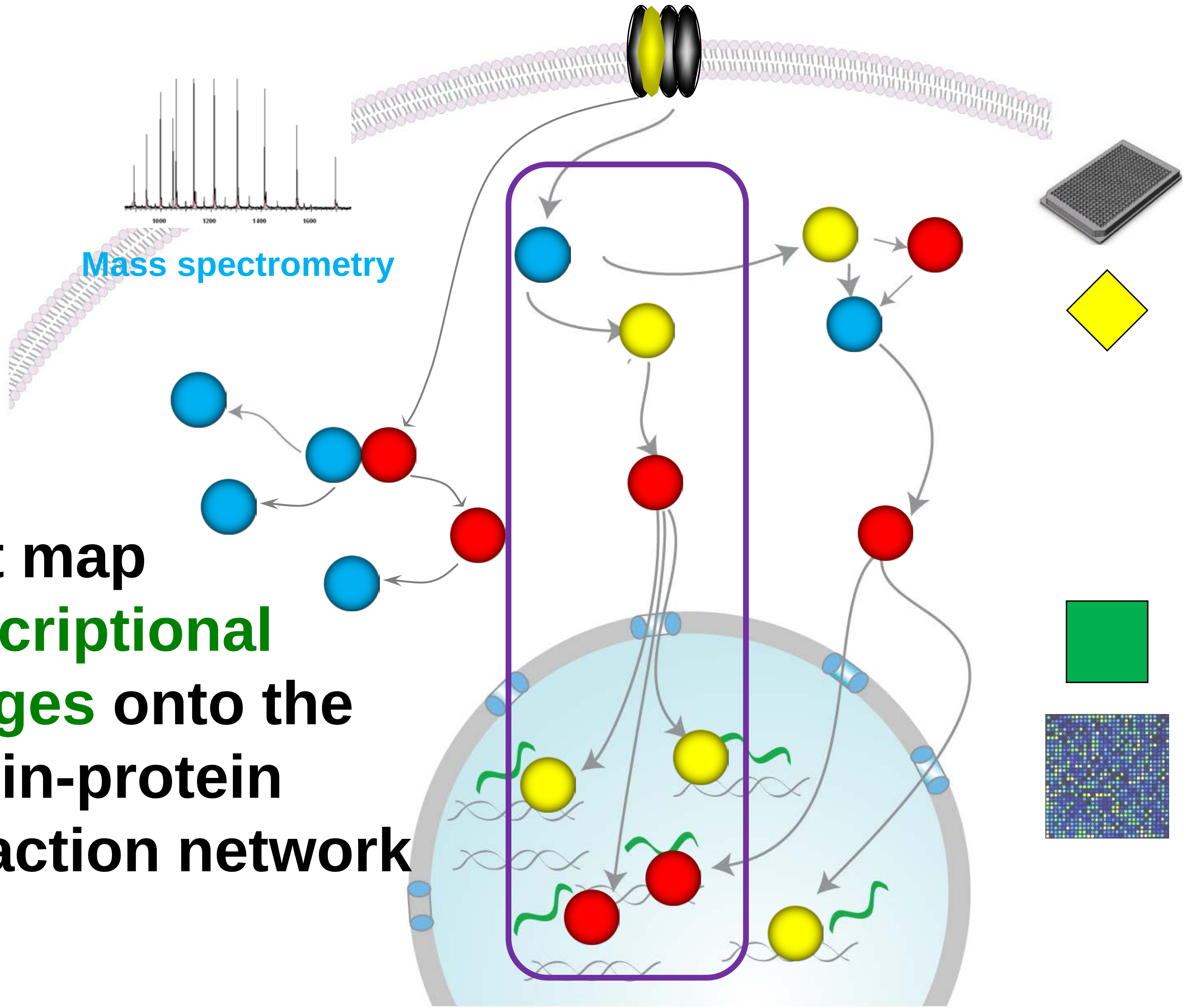
Map data onto a network of known interactions.



Interactome:
Known ***physical***
protein-protein interactions



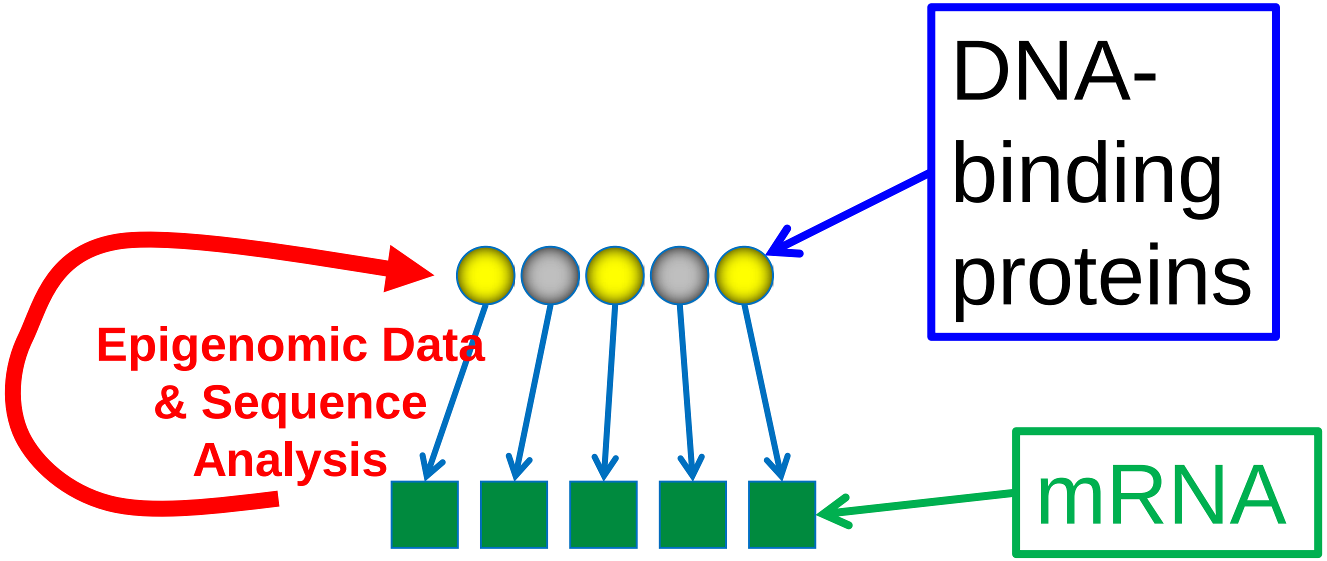
17,457 nodes
181,499 edges



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

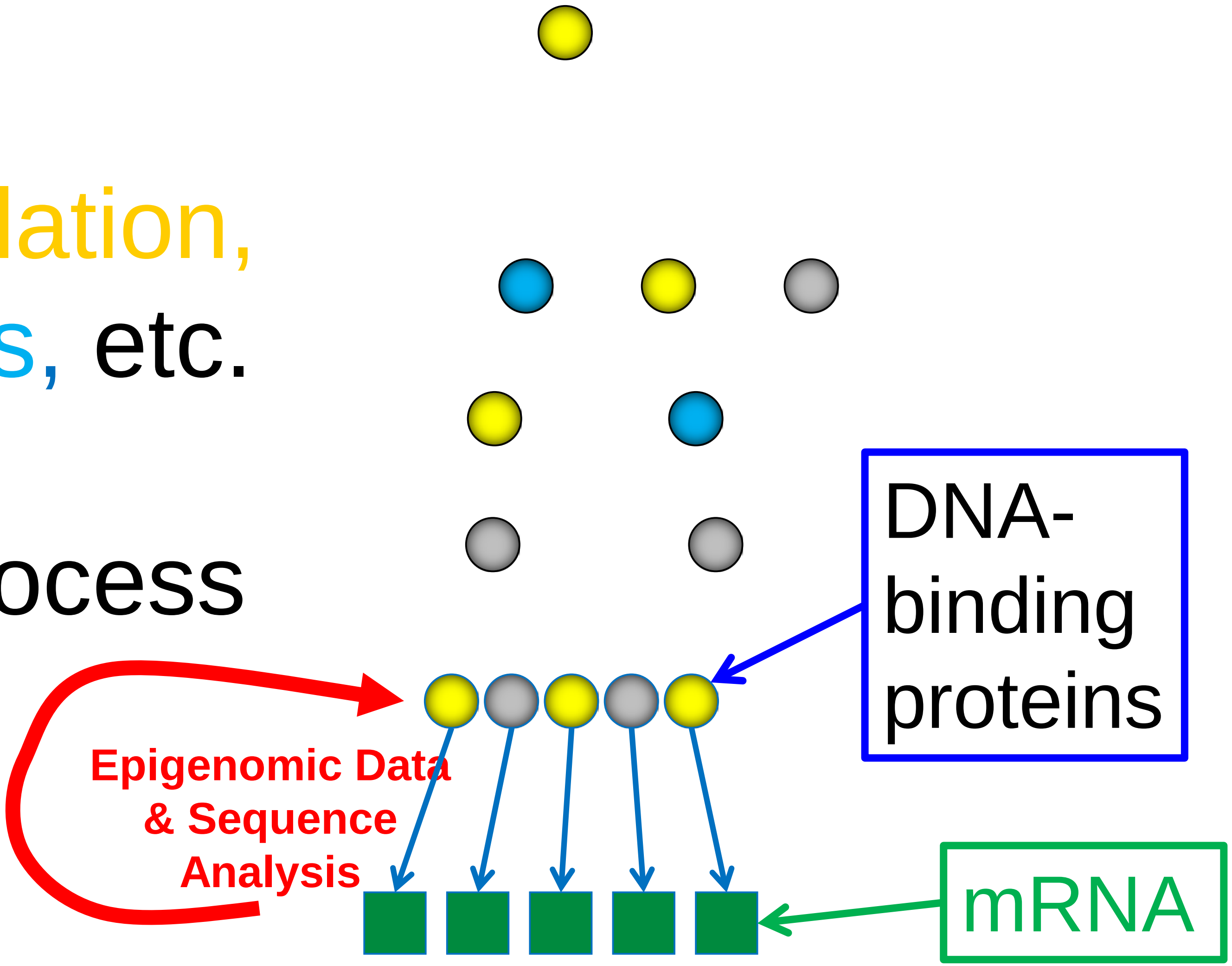
Step 1

Use expression data to find upstream signaling changes



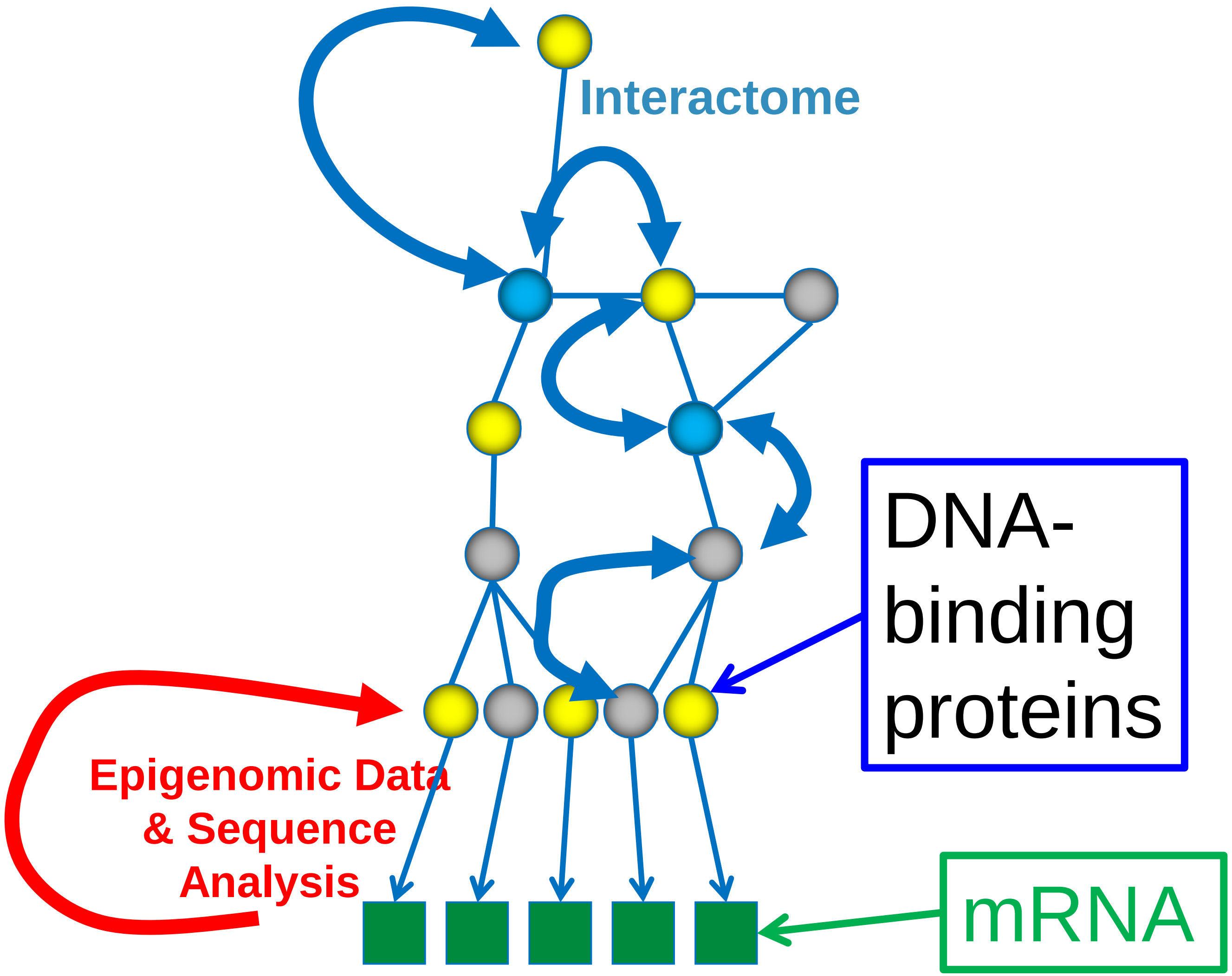
Step 2

Identify phosphorylation, metabolites, etc. relevant to disease process



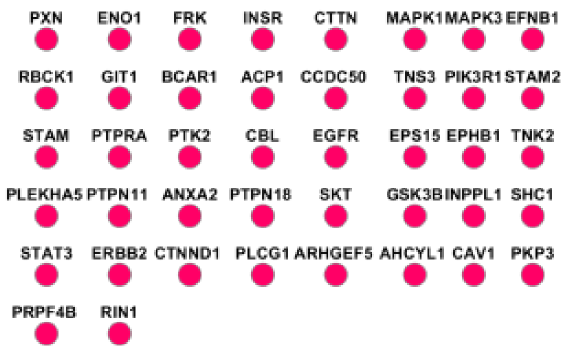
Step 3

Network integration

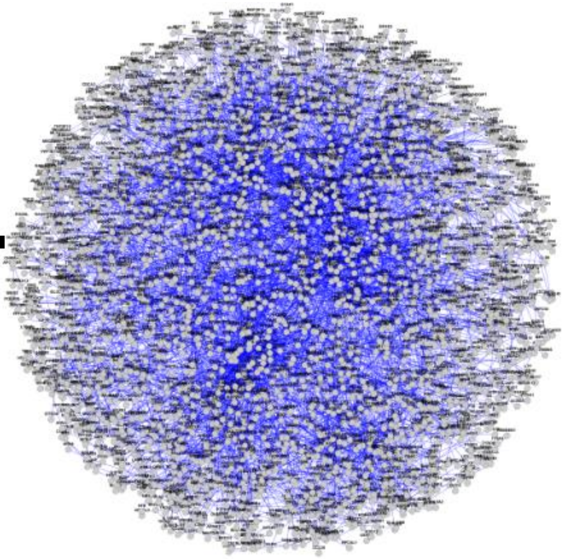


Interactome

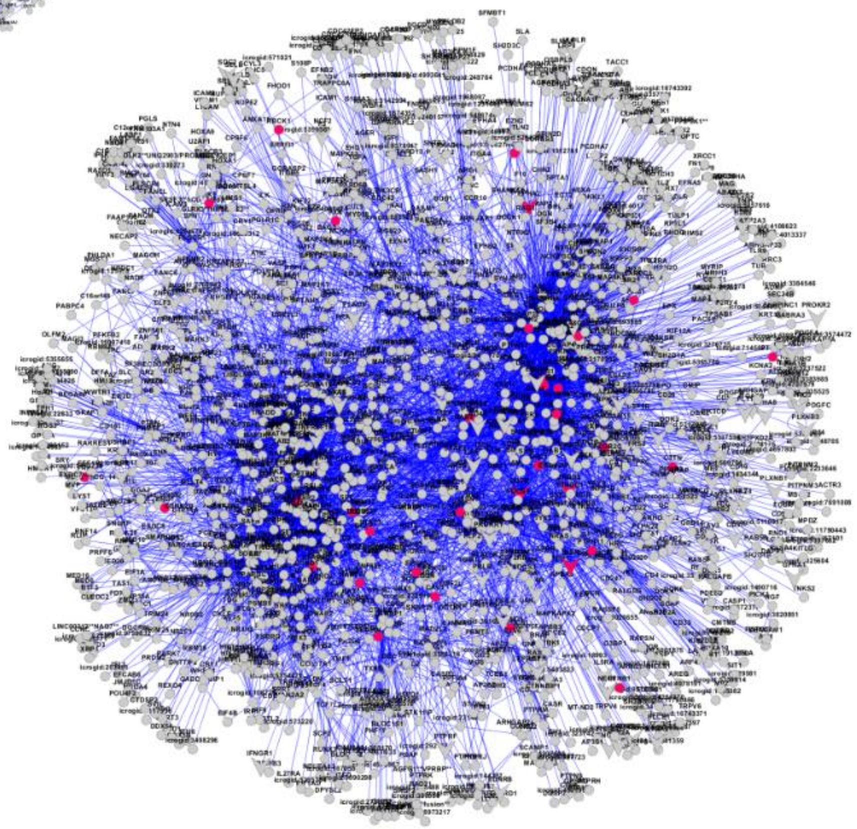
Experimental hits



+



Naïve methods



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

Approach

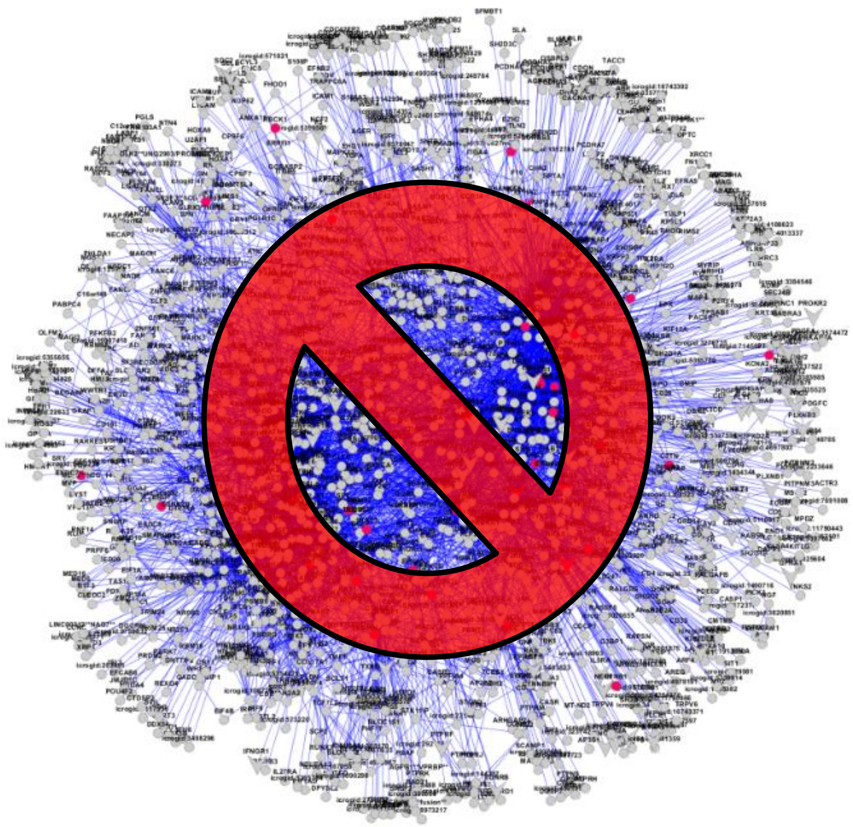
Lesson 1:

Network models make sense of diverse data.

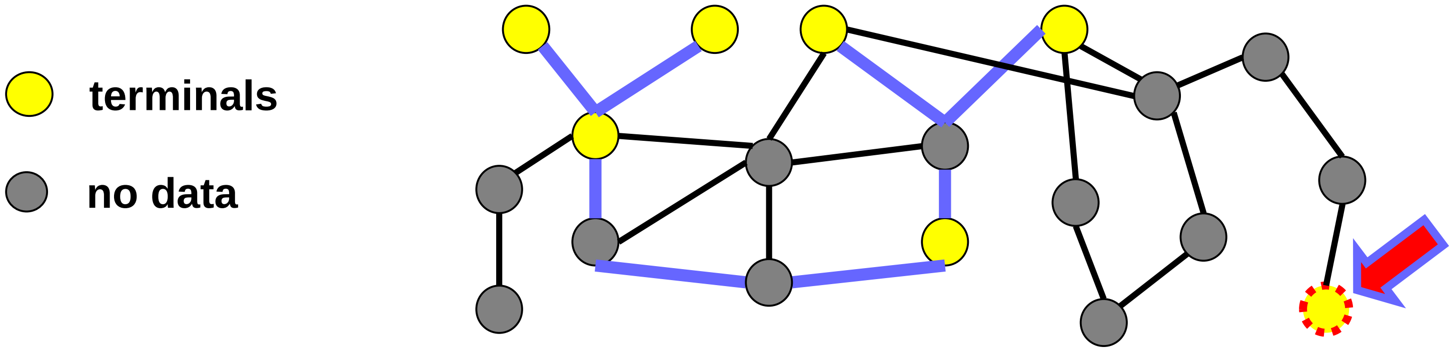
Lesson 2:

Hairballs do not!

Advanced network algorithms needed.



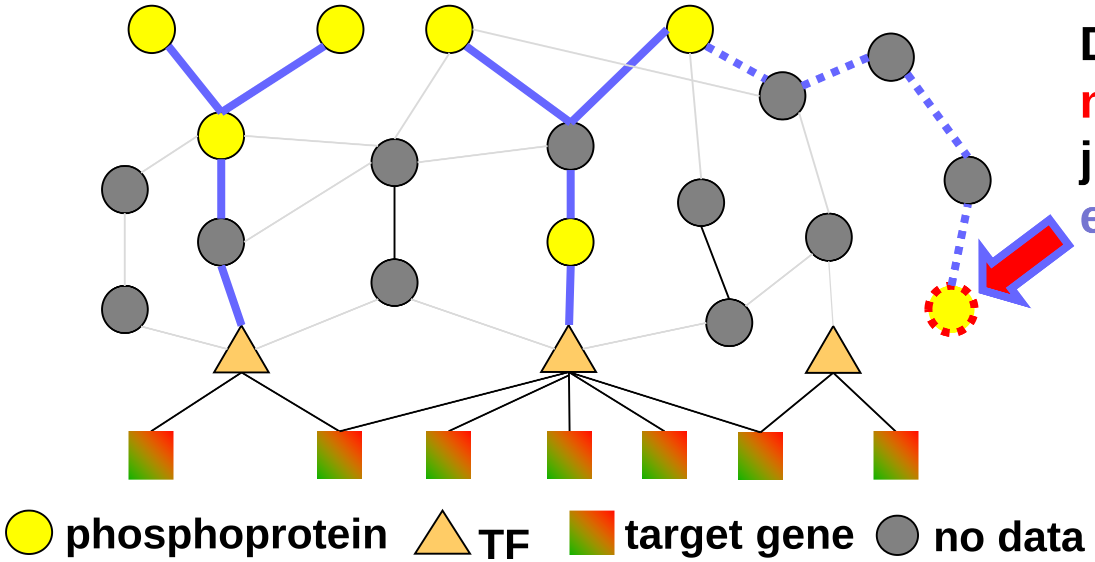
Avoiding False Positives



Trying to connect all the data is not necessarily the right approach

Prize-collecting Steiner Tree

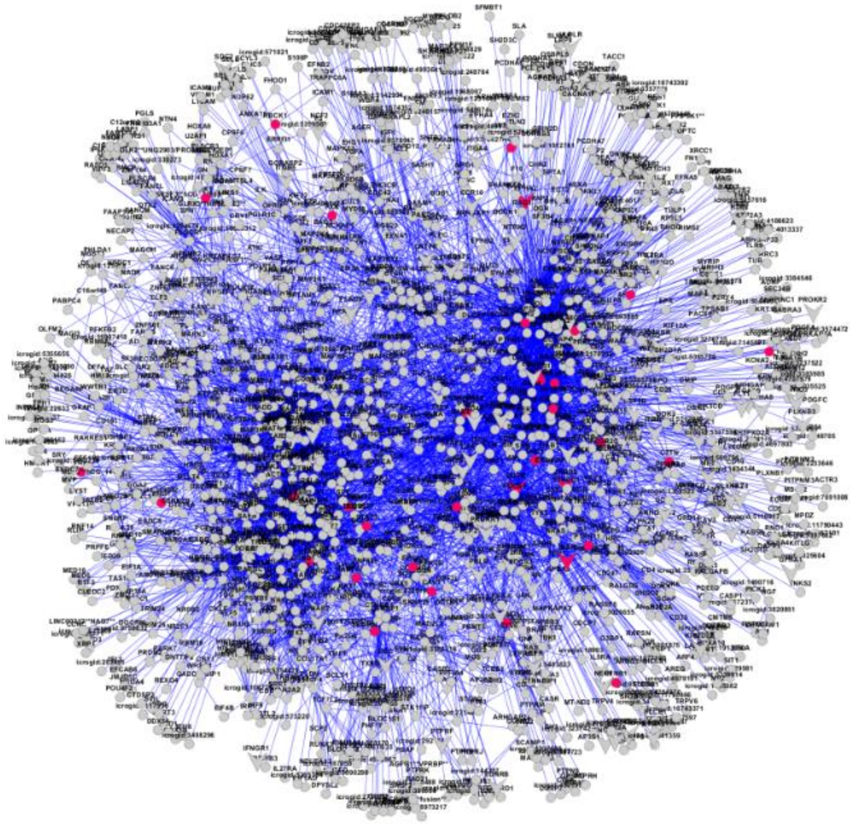
Method accounts for variable reliability of interactions and omic data



Does the node penalty justify the edge costs?

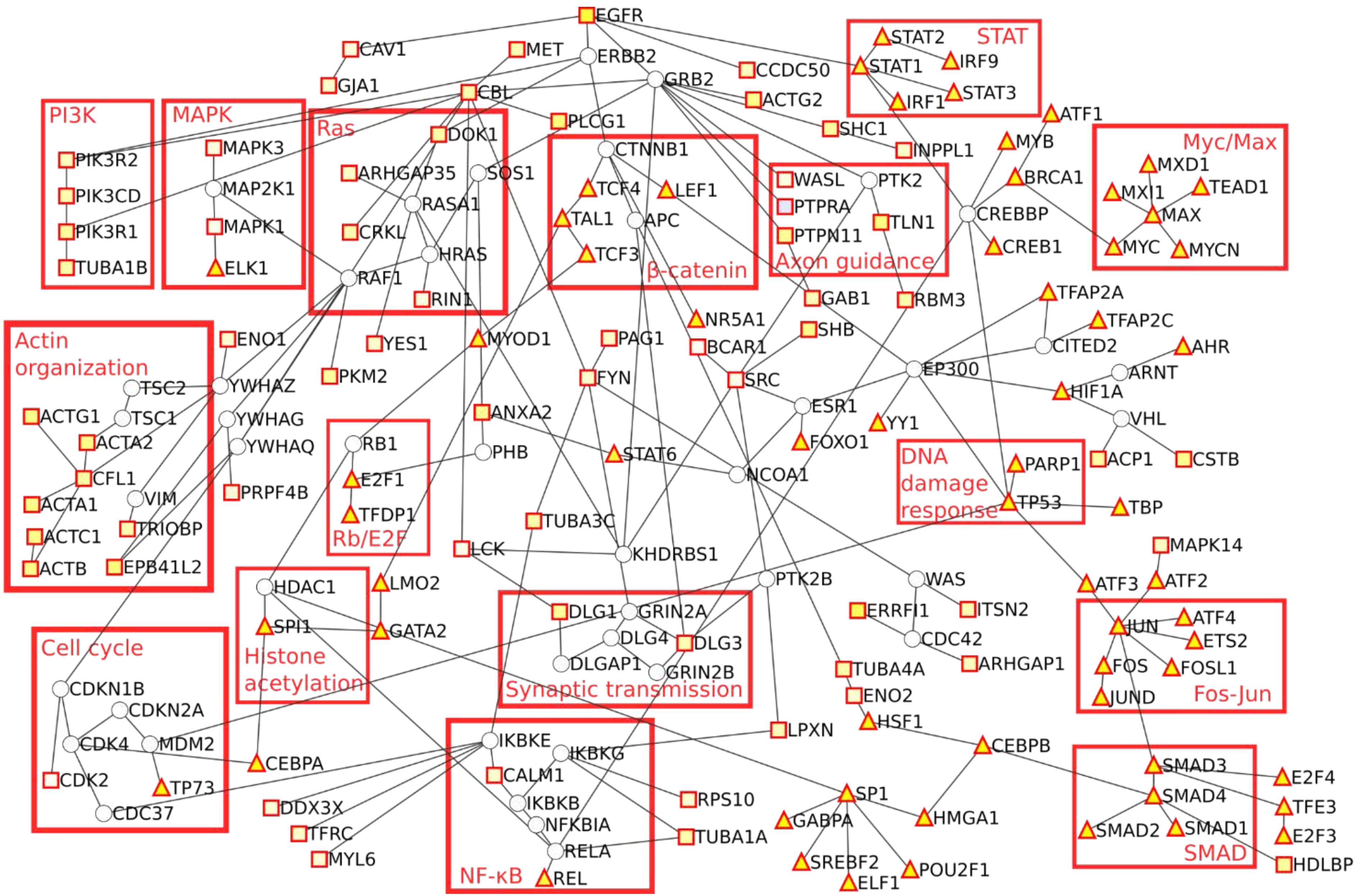
$$\boxed{\sum_{v \text{ not in } T} \beta \, \text{penalty}(v)} + \boxed{\sum_{e \text{ in } T} \text{cost}(e)}$$

Naïve Methods



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

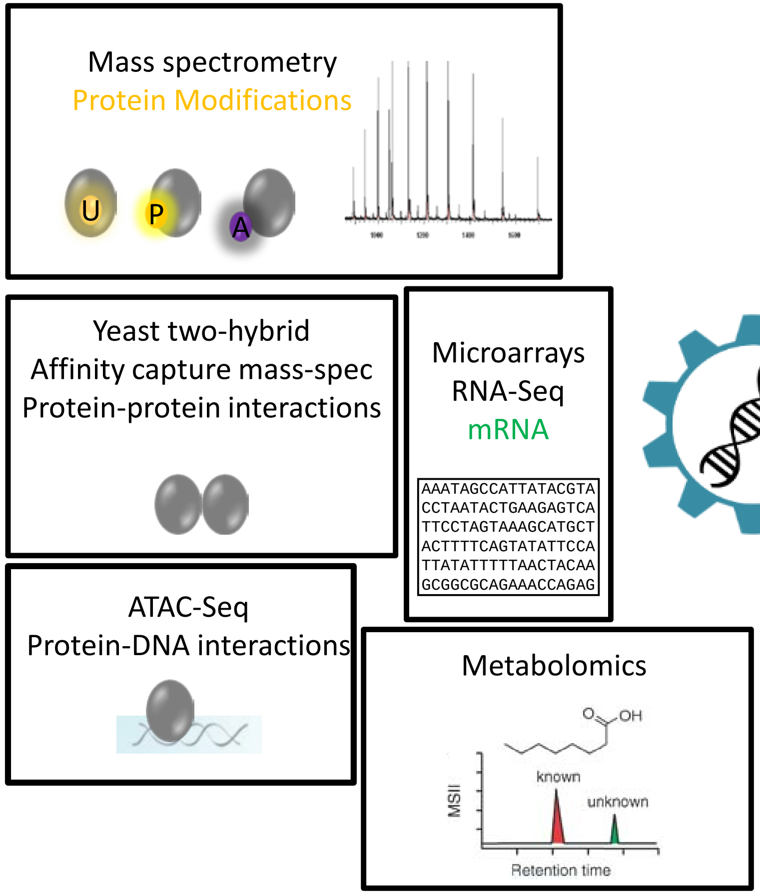
Linking Proteomic and Transcriptional Data through the Interactome and Epigenome Reveals a Map of Oncogene-induced Signaling
PLoS Comput Biol 9(2): e1002887. doi:10.1371/journal.pcbi.1002887

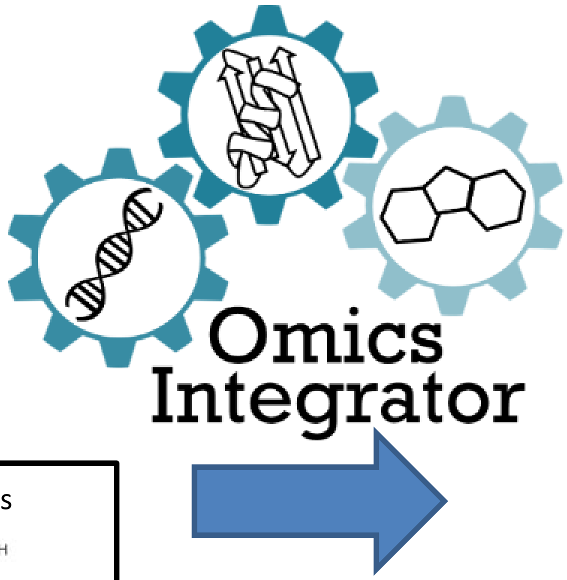


Linking Proteomic and Transcriptional Data through the Interactome and Epigenome Reveals a Map of Oncogene-induced Signaling

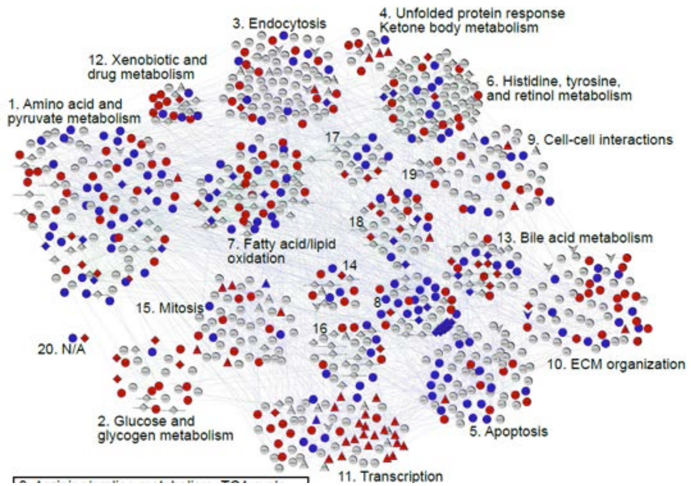
PLoS Comput Biol 9(2): e1002887. doi:10.1371/journal.pcbi.1002887

Interactome Models





- **Mechanistic**
- **Flexible**
- **Useful for Small N**
- **Patient Specific**



Nat Methods. doi: 10.1038/nmeth.3940.
Revealing disease-associated pathways by network integration of untargeted metabolomics.
PLoS Comput Biol. 2016 doi: 10.1371/journal.pcbi.1004879.
Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package.

PIÙMet

Revealing disease-associated pathways by network integration of untargeted metabolomics

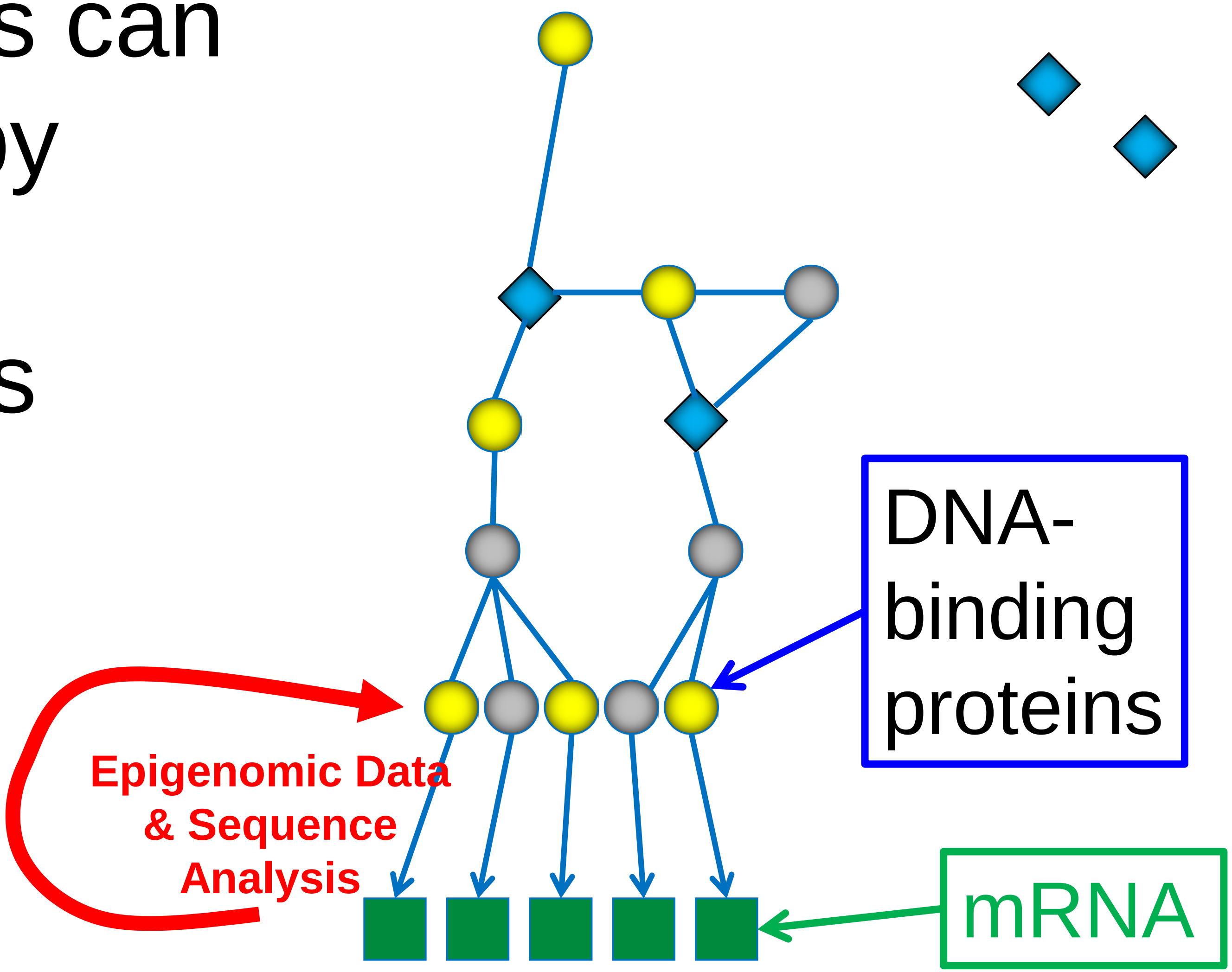
Leila Pirhaji[1], Pamela Milani[1], Mathias Leidl[2], Timothy Curran[1,3], Julian Avila-Pacheco[4], Clary B Clish[4], Forest M White[1,3], Alan Saghatelian[2,5] & Ernest Fraenkel[1,4]

NATURE METHODS | ADVANCE ONLINE PUBLICATION

PUBLISHED ONLINE 1 AUGUST 2016; DOI:10.1038/NMETH.3940

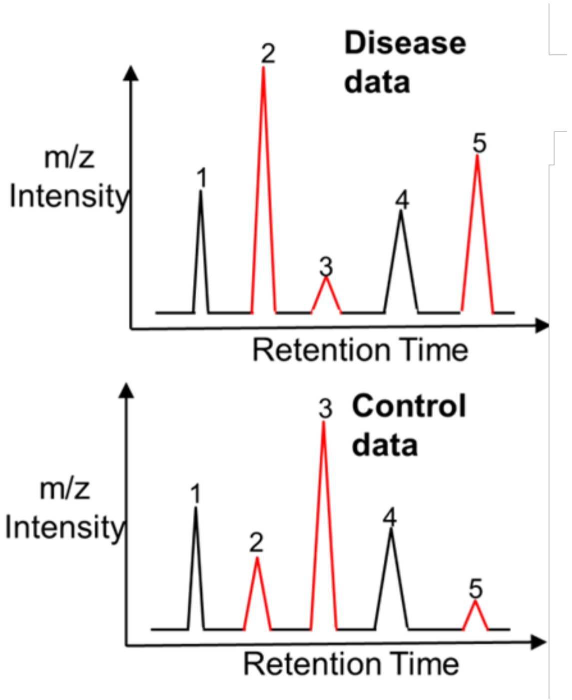
Metabolomics

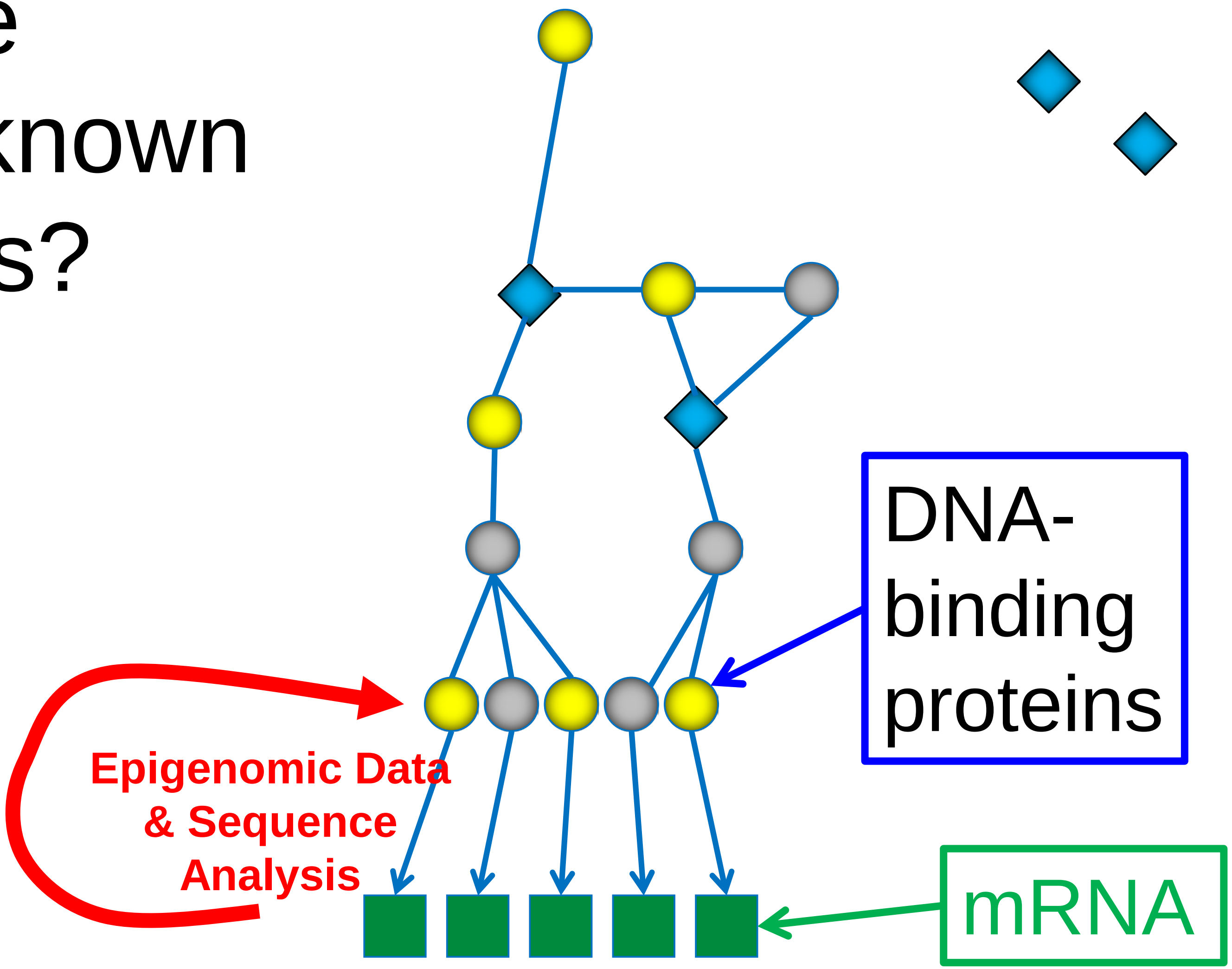
Metabolites can be linked by physical interactions



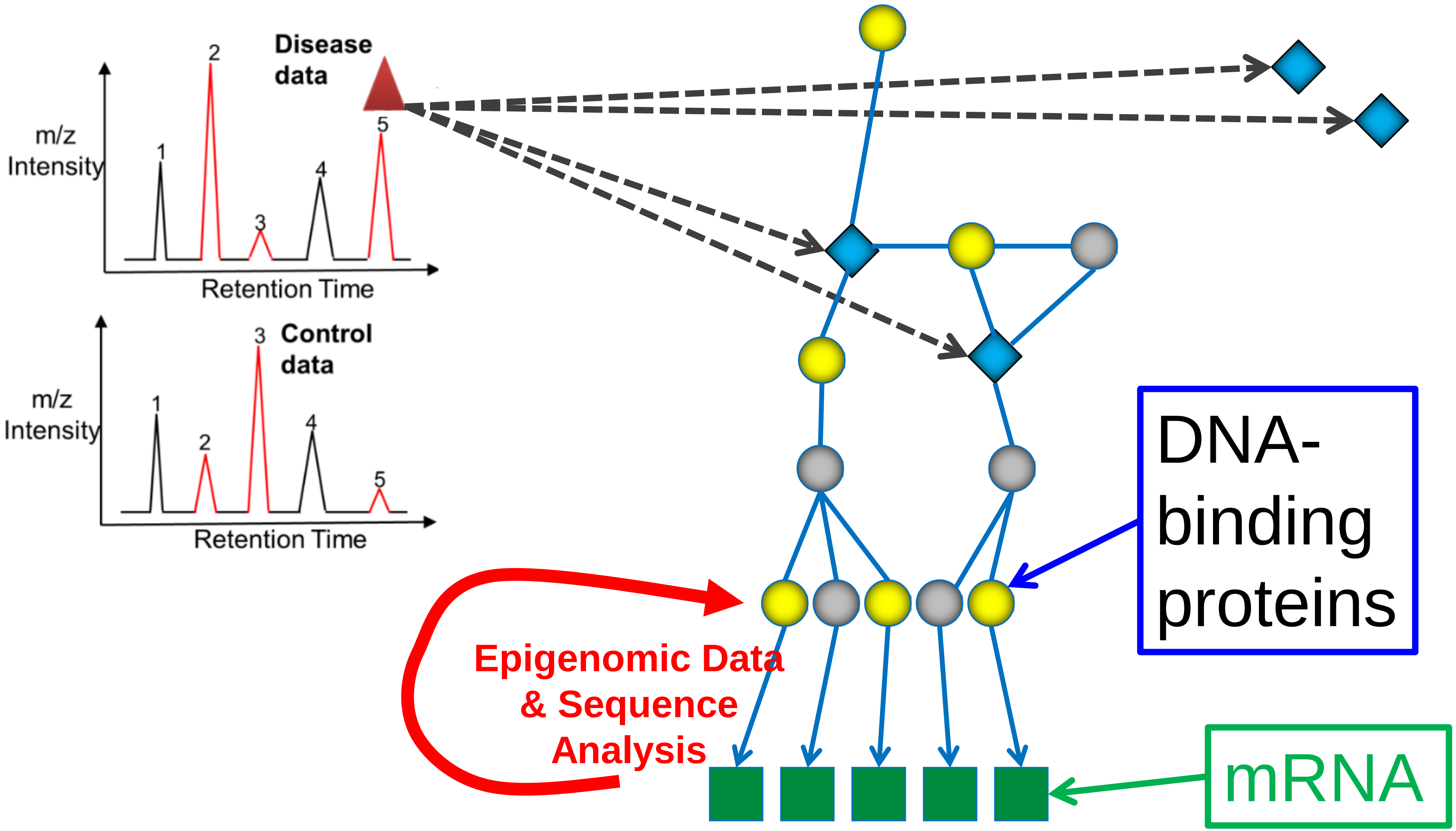
Metabolomics

How do we handle unknown metabolites?

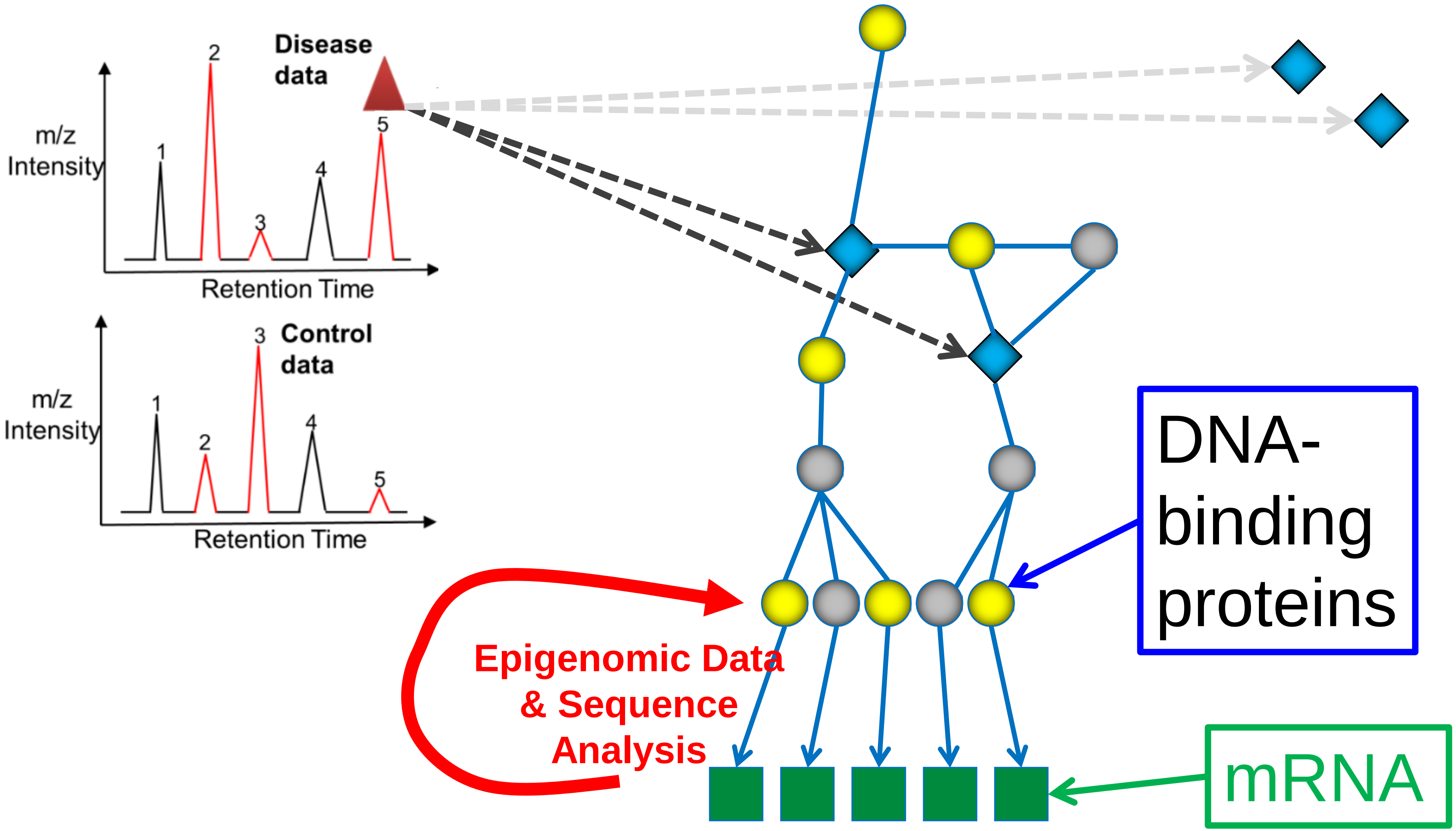




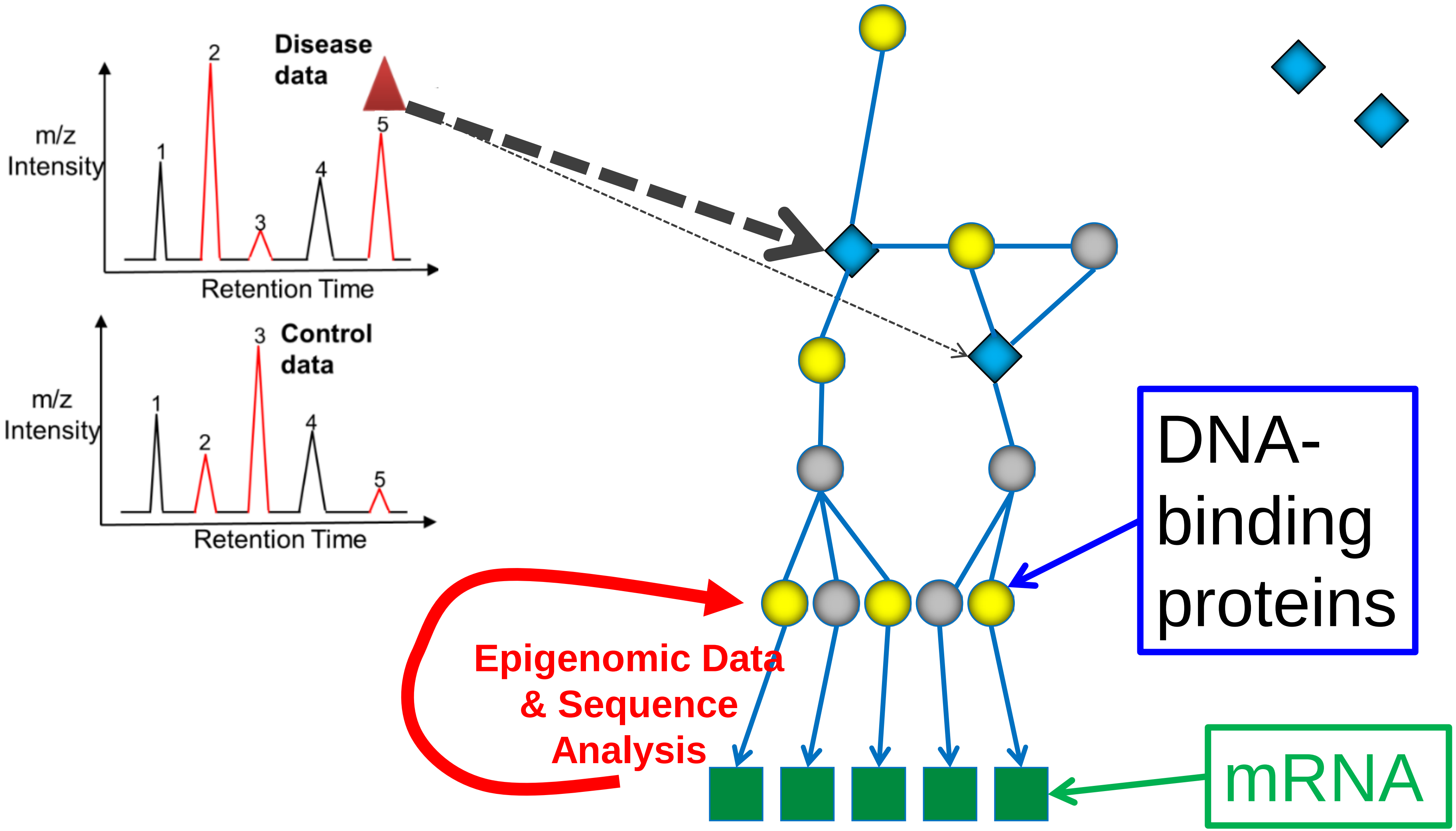
Initial Assignments Based on Mass



Connectivity Supports Some Assignments

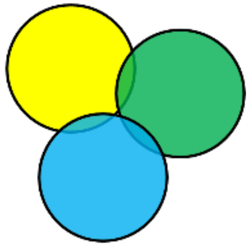


Robustness
Determines Weighted Assignments

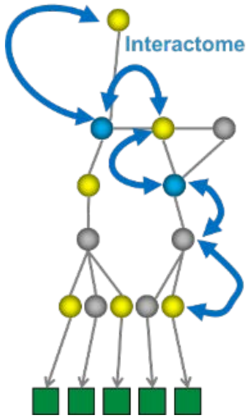


Outline

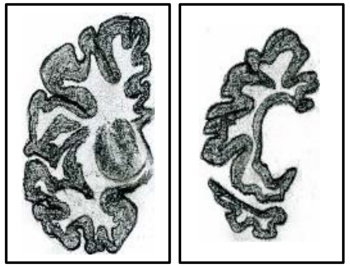
Why Data Integration is Hard



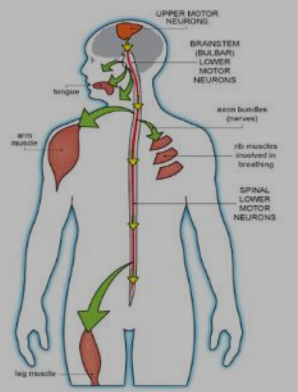
Networks Link the Data



Huntington's Disease

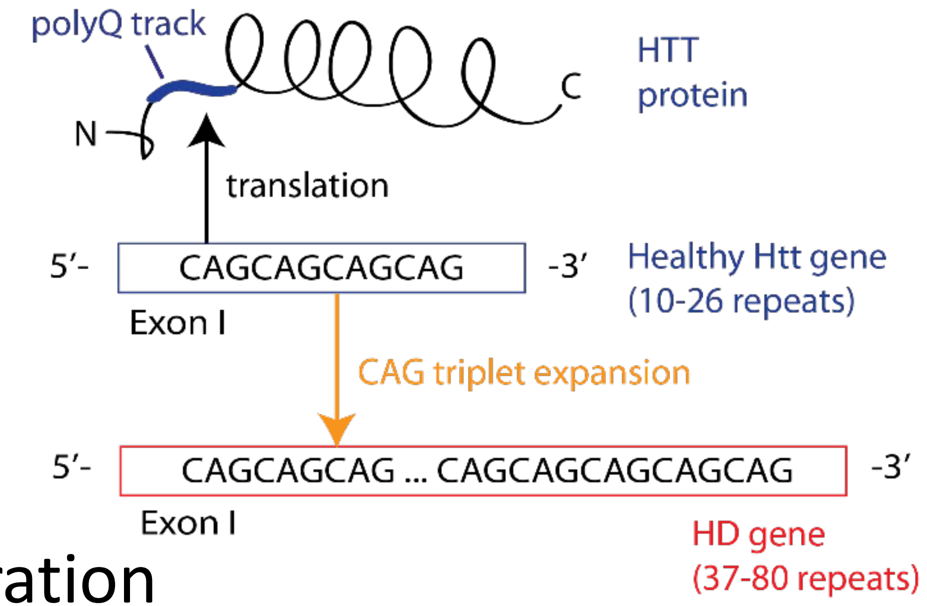


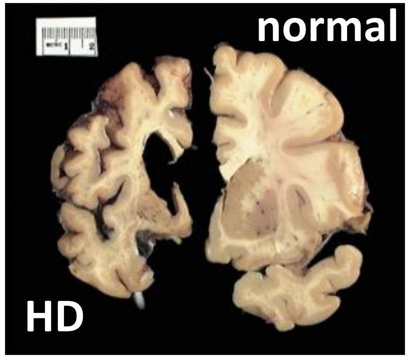
ALS



Huntington's Disease

Expanded CAG repeat
in gene for Huntingtin





Harvard Brain Tissue
Resource Center

Neurodegeneration

cognitive decline, psychiatric disturbance, chorea, dystonia

Pamela
Milani

Leila
Pirhaji

Amanda
Kedaigle

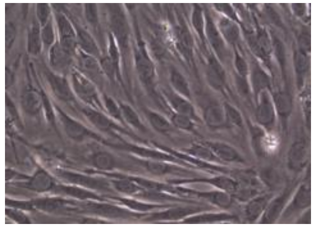
Brooke
Wassie

Simona
Dalin

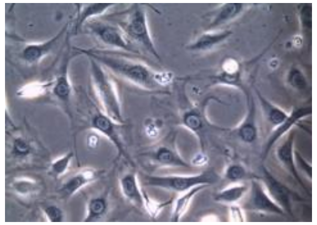
Lipidomics and Phosphoproteomics

Wild type: 7 repeats of (CAG)

STHdhQ7



STHdhQ111



Cell-line model of HD

Mutant: 111 repeats of (CAG)

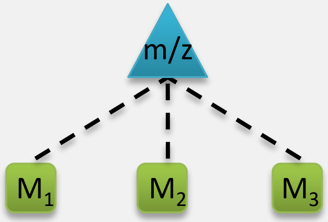
Affinity purify lipids

LC/MS

Map to network

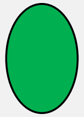
Affinity purify phosphotyrosine containing proteins

37 peaks

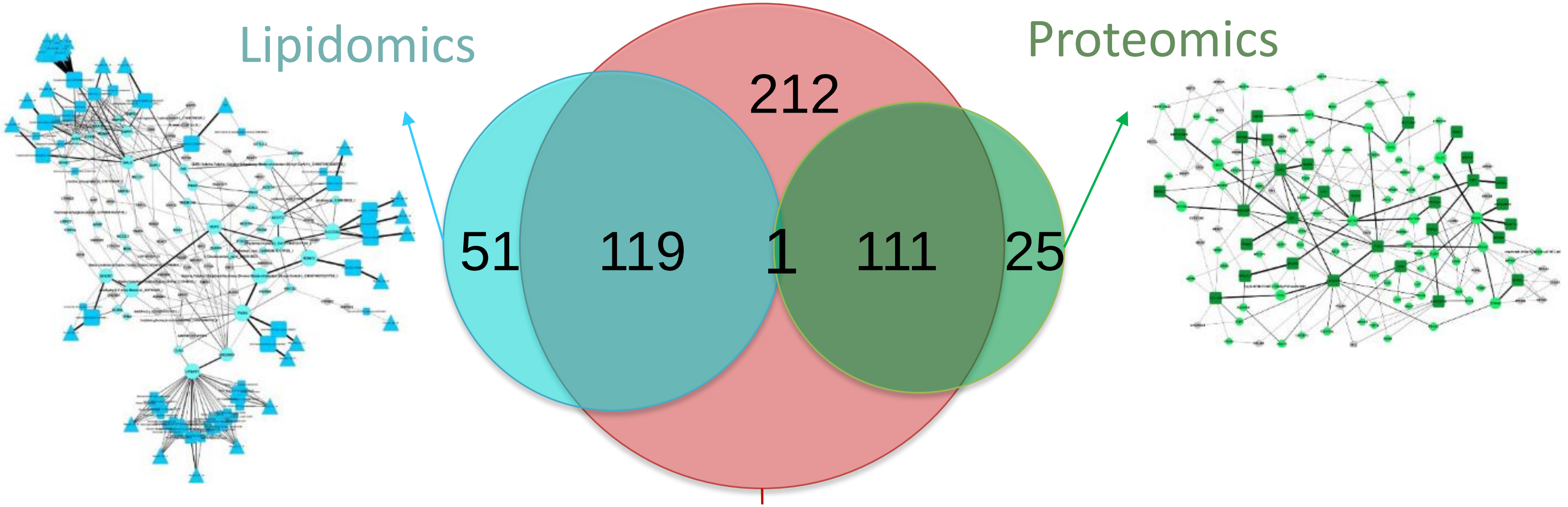


296 metabolites

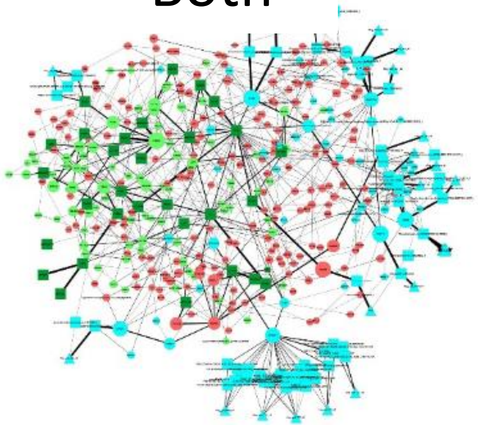
31 pY proteins



Value of multi-omics

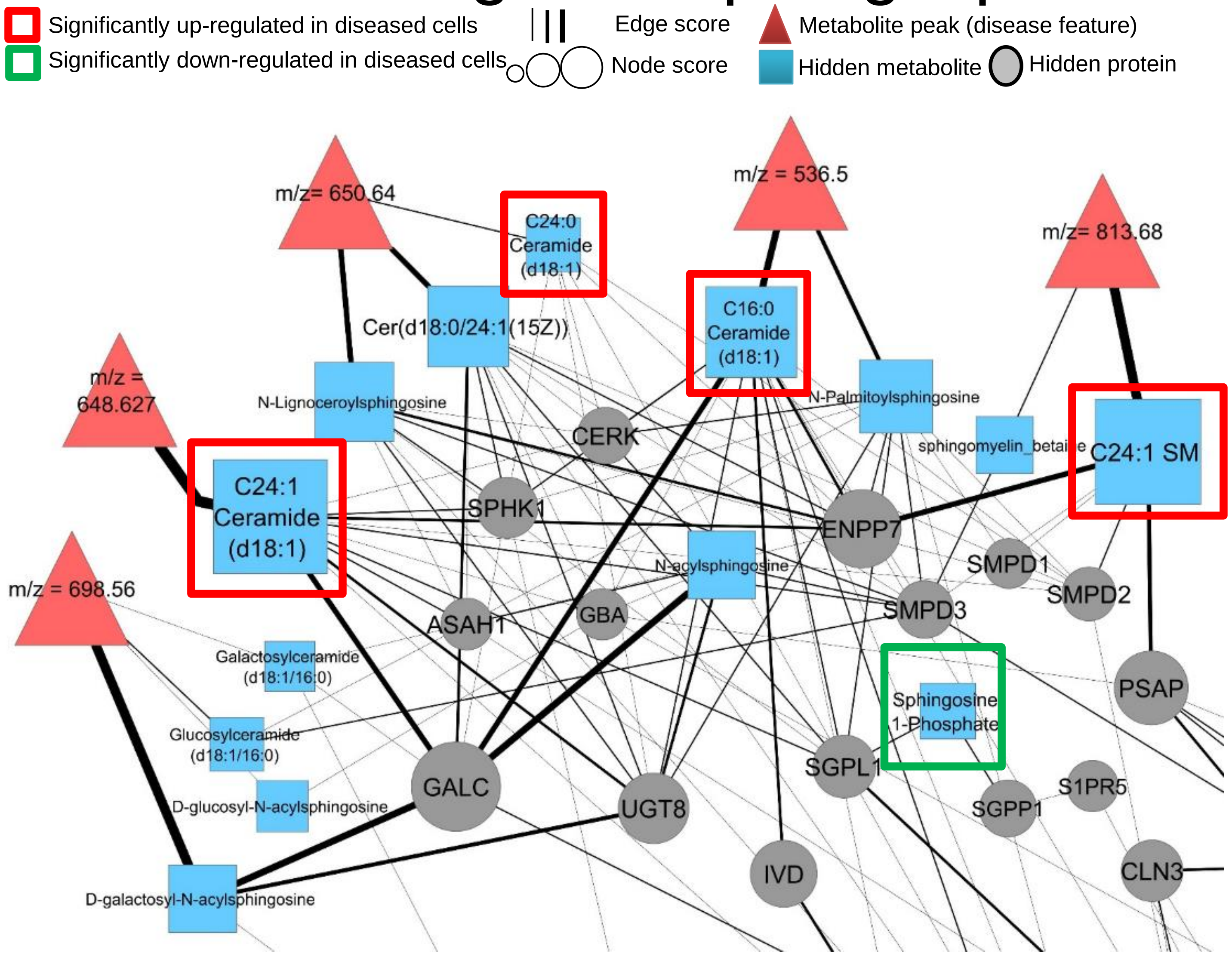


Both

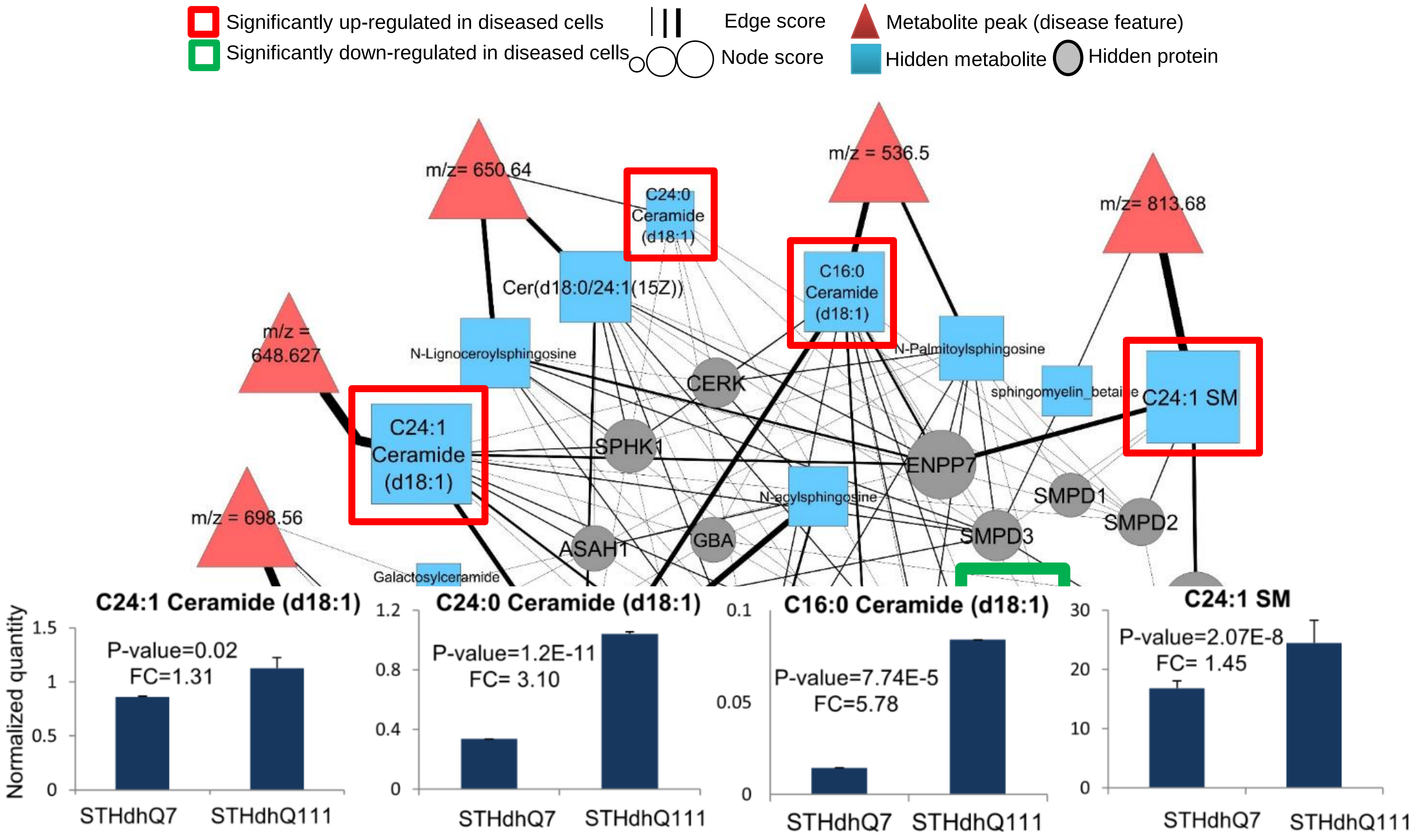


number of hidden nodes
found
using indicated data

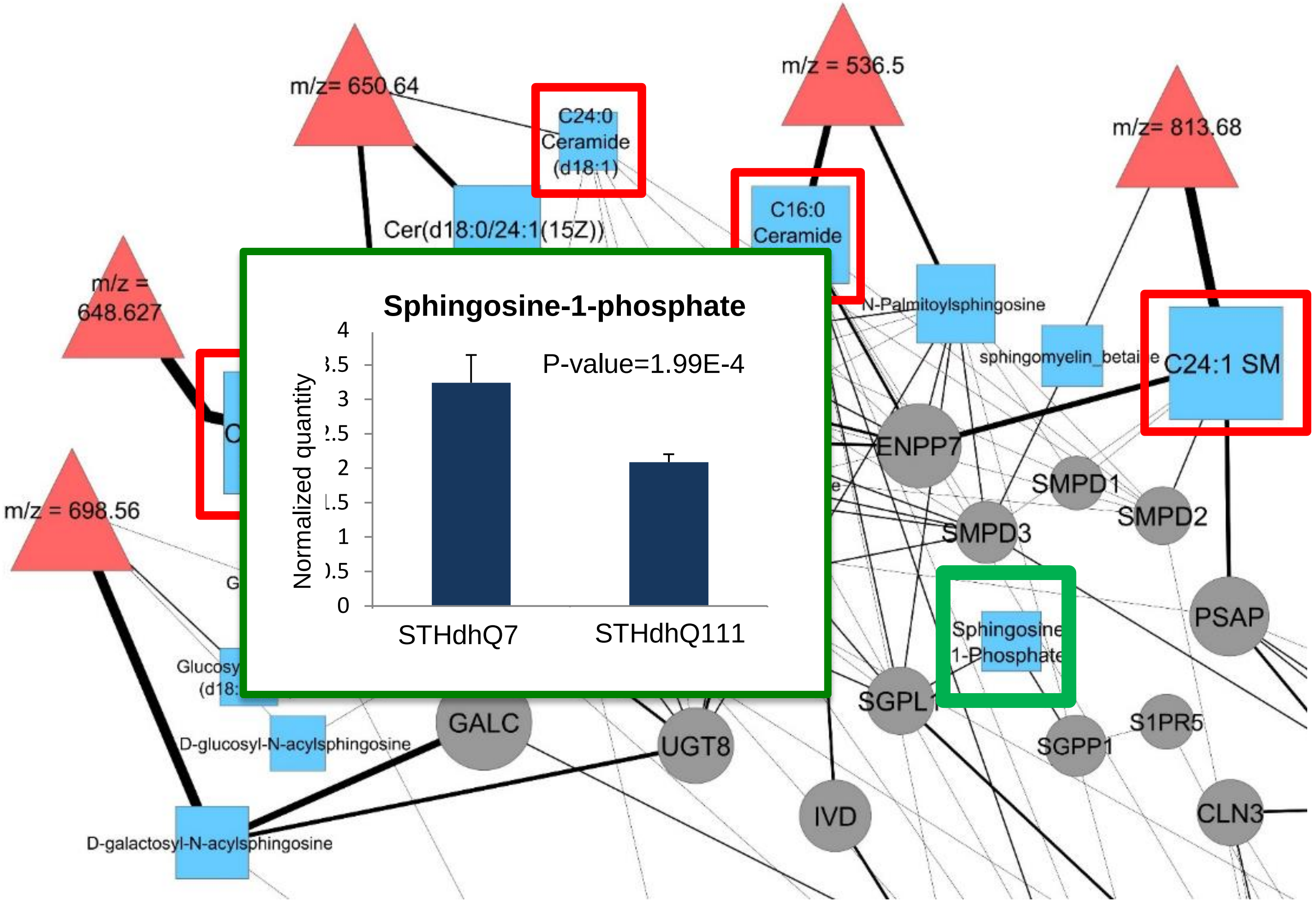
PIUMet identifies changes in sphingolipid metabolism



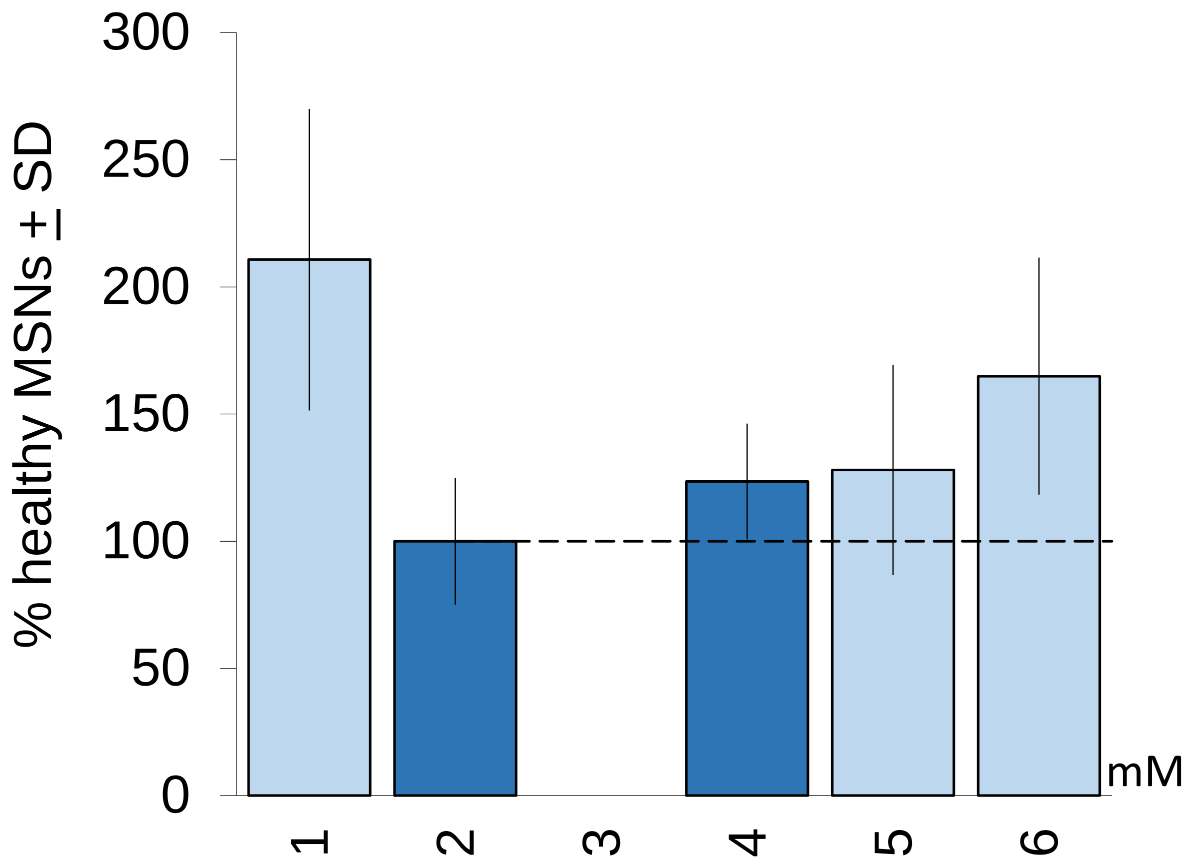
Sphingolipid changes experimentally verified



PIUMet identifies a potential therapeutic strategy



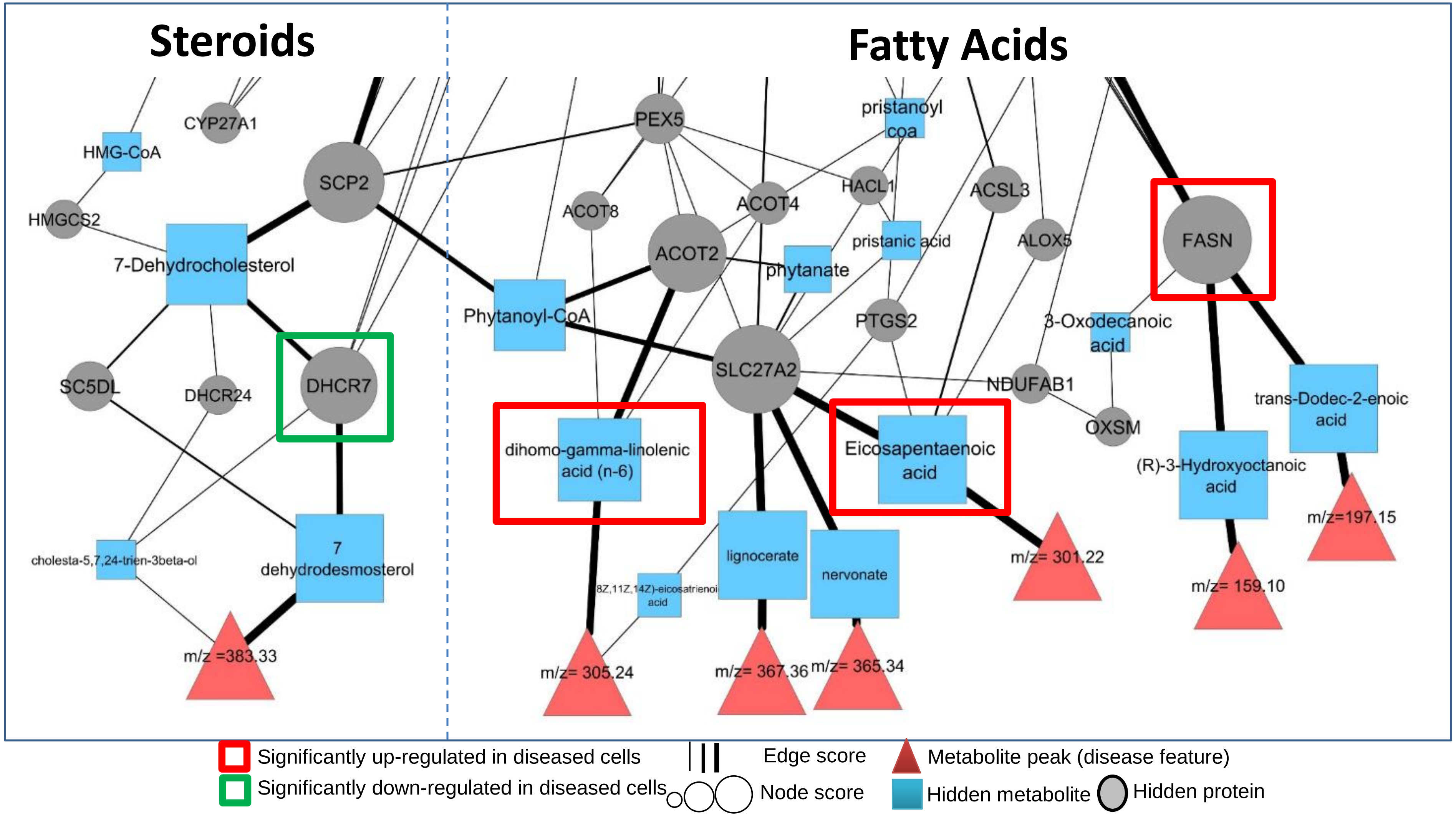
Inhibiting SPL enzyme protects neurons in rat brain slice culture



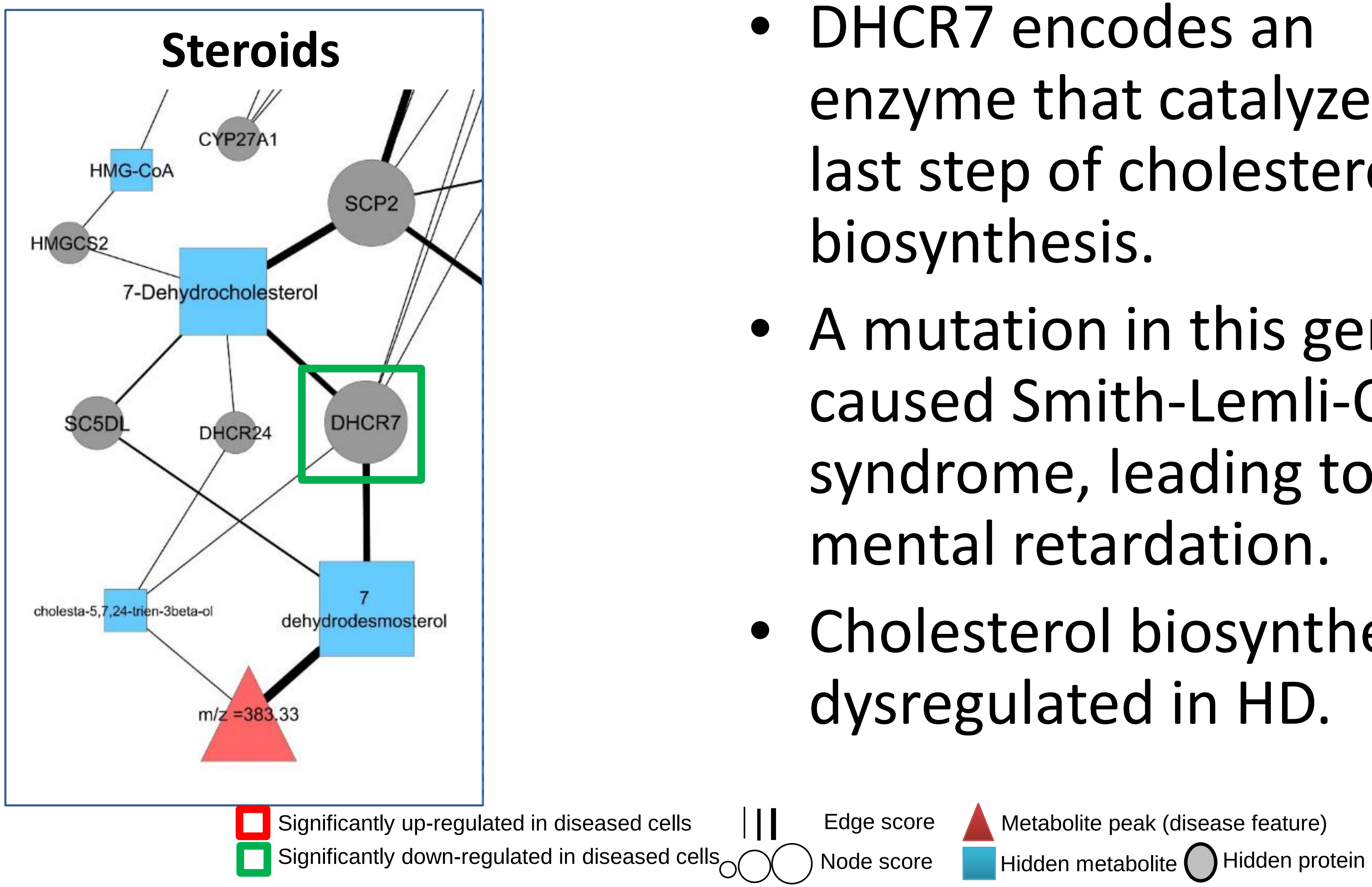

Denise Dunn


Don Lo

Novel Pathways

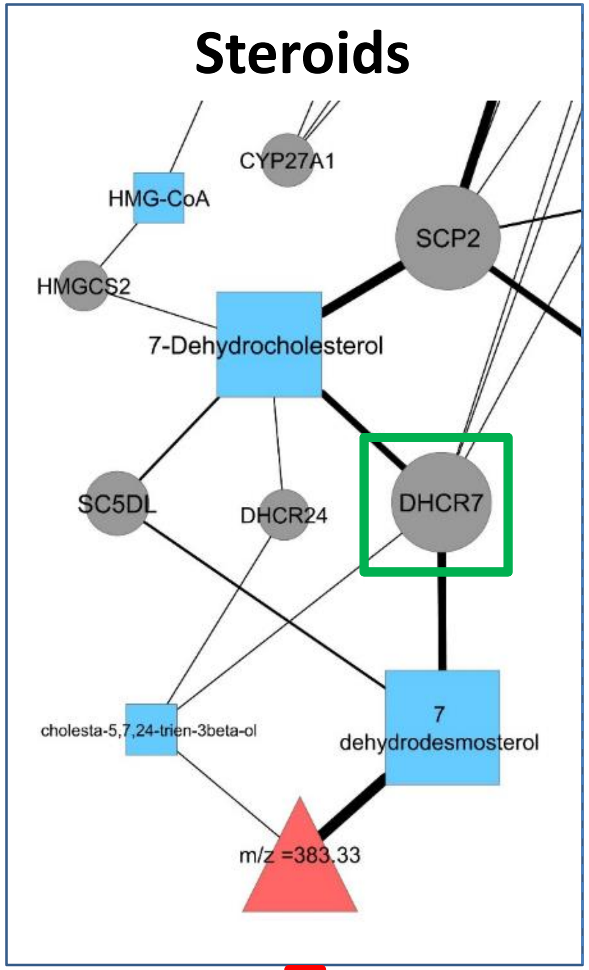


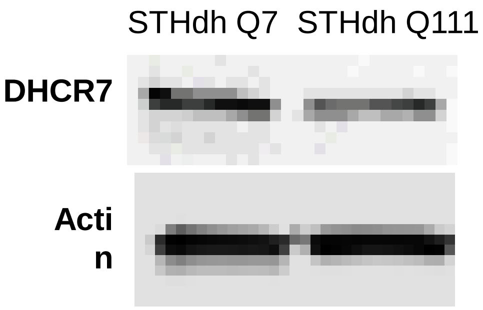
Novel Pathways

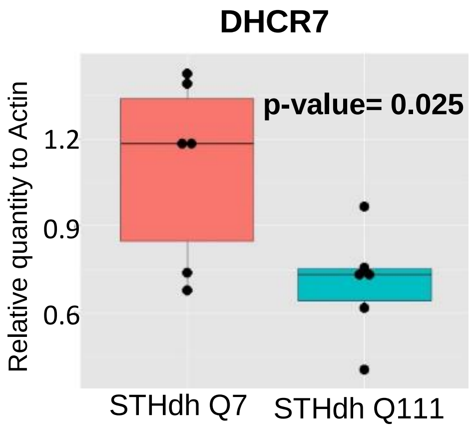


- DHCR7 encodes an enzyme that catalyzes the last step of cholesterol biosynthesis.
- A mutation in this gene caused Smith-Lemli-Opitz syndrome, leading to mental retardation.
- Cholesterol biosynthesis is dysregulated in HD.

Novel Pathways

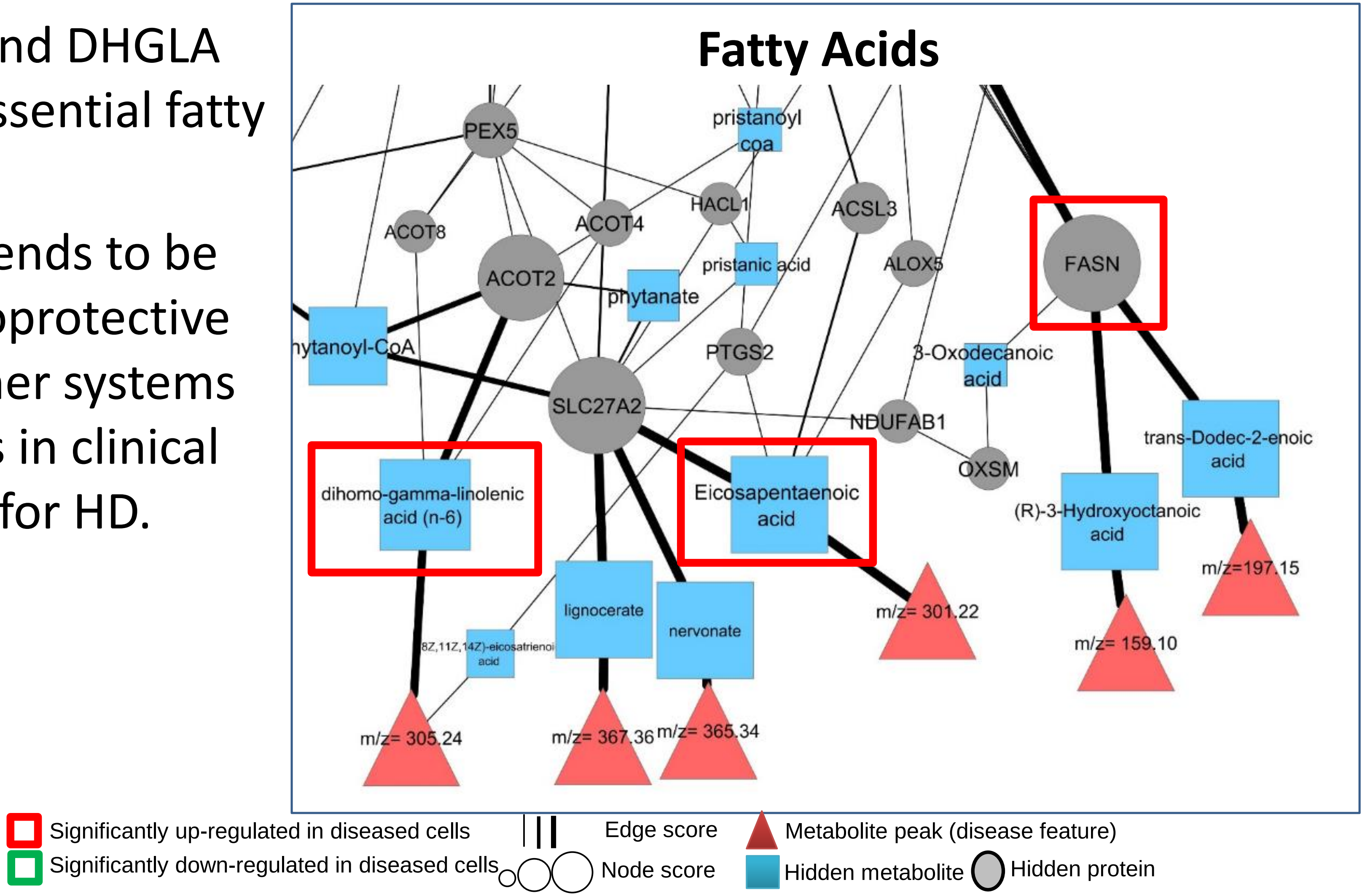






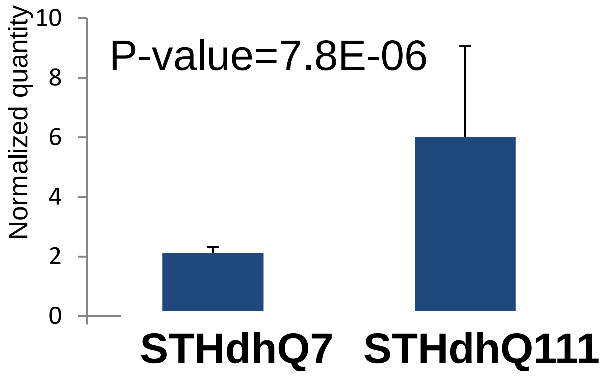
Novel Pathways

- EPA and DHGLA are essential fatty acids
- EPA tends to be neuroprotective in other systems and is in clinical trials for HD.

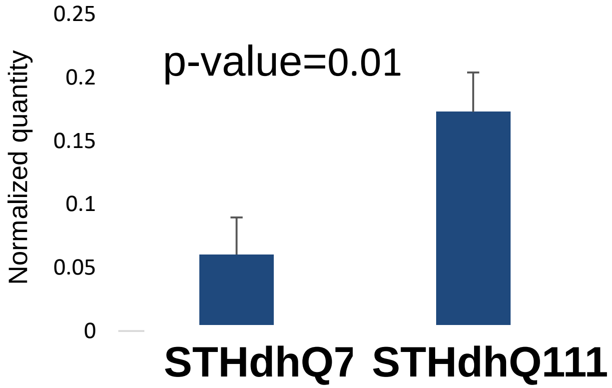


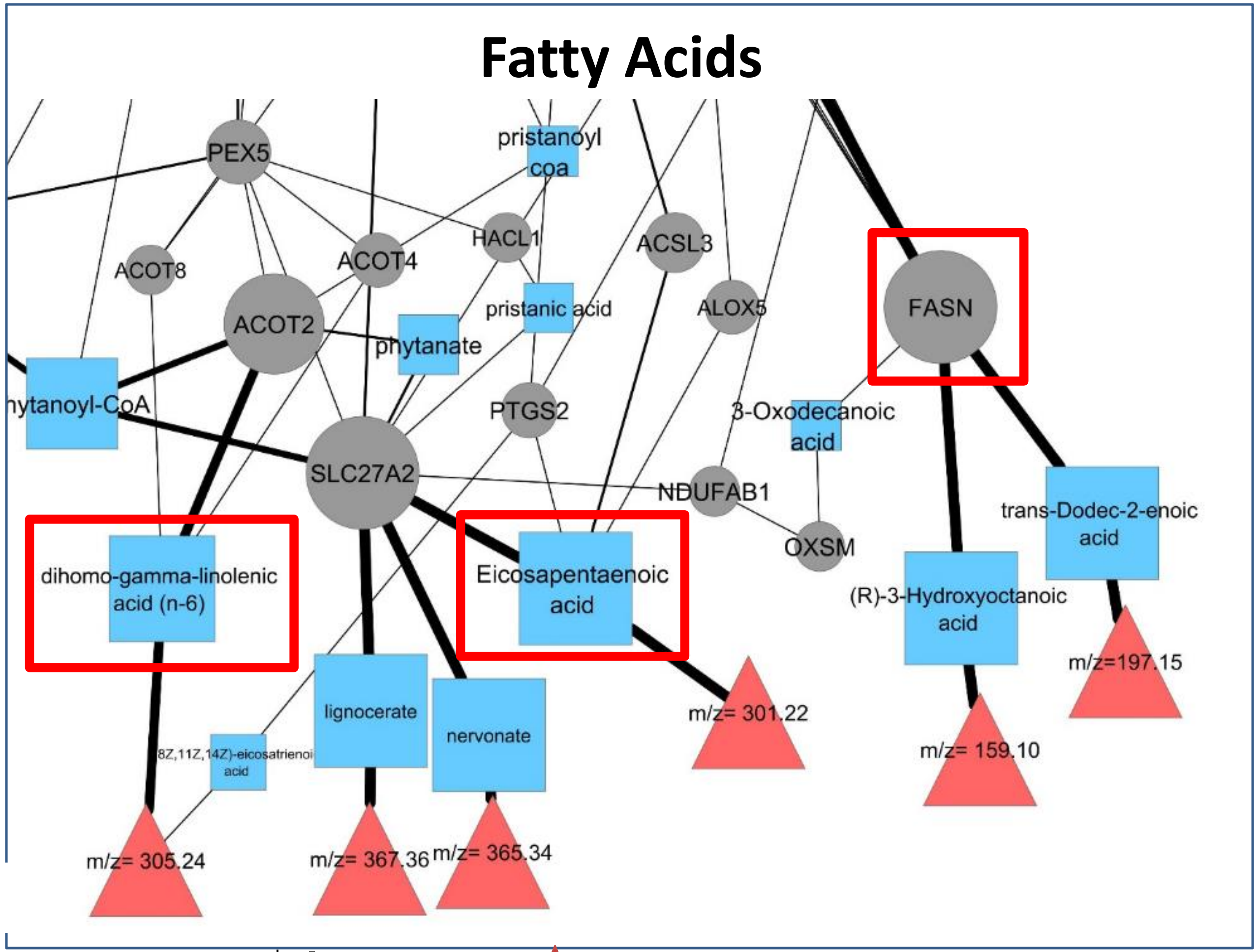
Novel Pathways

Eicosapentaenoic acid



Dihomo-gamma-linolenic acid

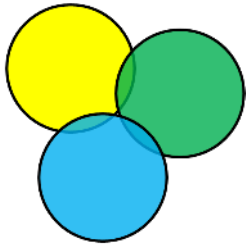




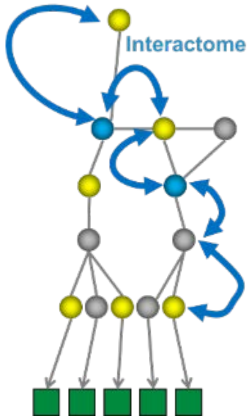
Significantly up-regulated in diseased cells
Significantly down-regulated in diseased cells
Edge score
Node score
Metabolite peak (disease feature)
Hidden metabolite
Hidden protein

Outline

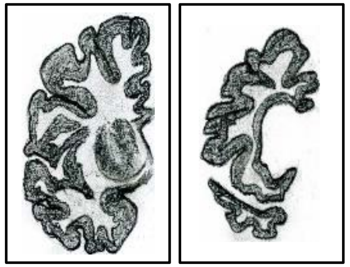
Why Data Integration is Hard



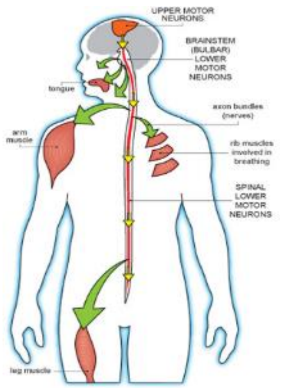
Networks Link the Data



Huntington's Disease

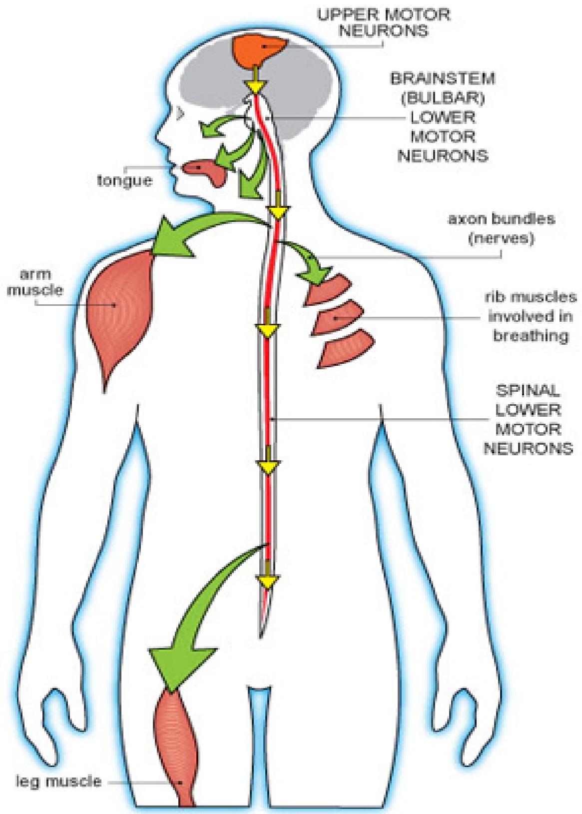


ALS



ALS: Fatal and poorly understood

- **Progressive**
  - Normally begins with mild symptoms and gradually affects most skeletal muscle
- **Fatal**
  - Patients lose the ability to perform vital functions, such as eating and breathing, resulting in death



Leslie Thompson

UCIrvine
University of California, Irvine

Jeff Rothstein

JOHNS HOPKINS
MEDICINE

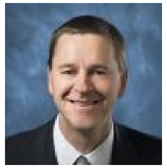
Steve Finkbeiner

UCSF
University of California
San Francisco

Jenny Van Eyk

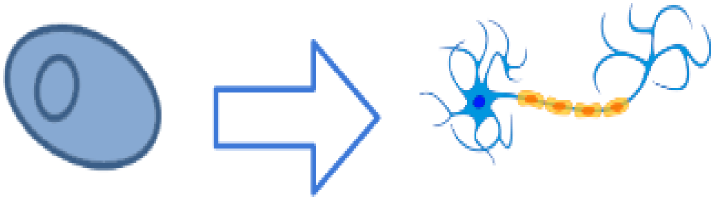
CEDARS-SINAI

Clive Svendsen



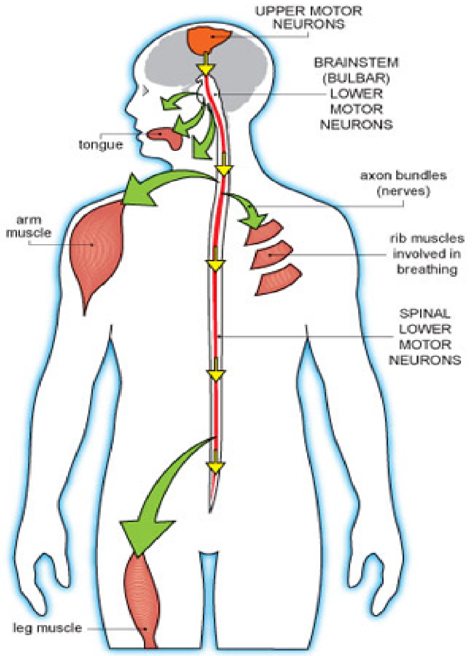
NEUROLINCS

Induced pluripotent stem cells provide personal models of disease



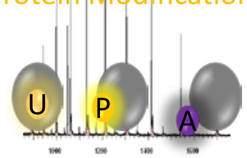
Induced Pluripotent Stem Cell (iPSC)

Motor Neuron



Mass spectrometry

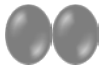
Protein Modifications



Yeast two-hybrid

Affinity capture mass-spec

Protein-protein interactions



RNA-Seq

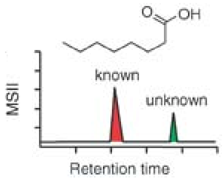
mRNA

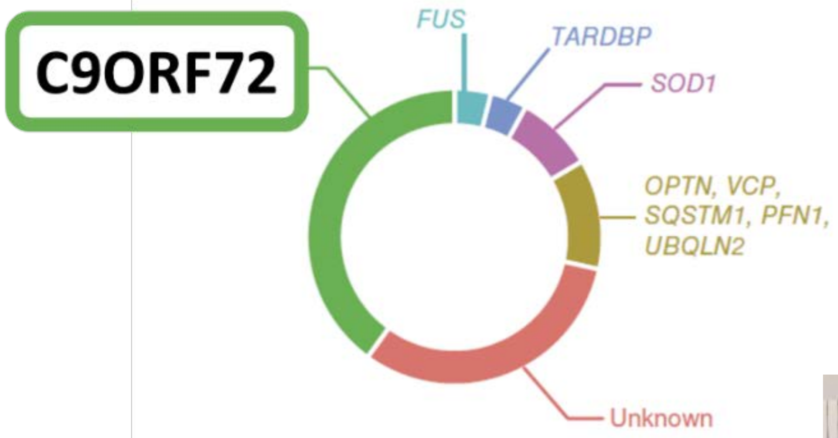
AAATAGCCATTATACGTA
CCTAATACTGAAGAGTCA
TTCCTAGTAAAGCATGCT
ACTTTTCAGTATATTCCA
TTATATTTTTAACTACAA
GCGGCGCAGAAACCAGAG

ATAC-Seq

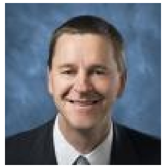
Protein-DNA interactions

Metabolomics

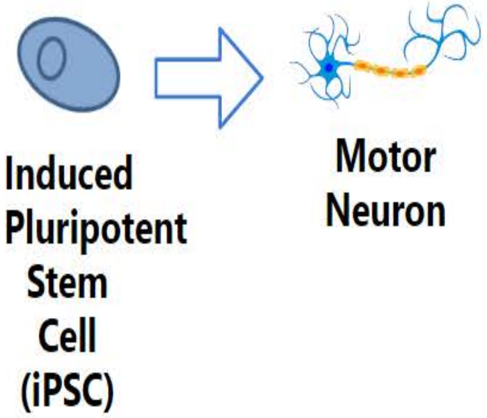




Clive Svendsen

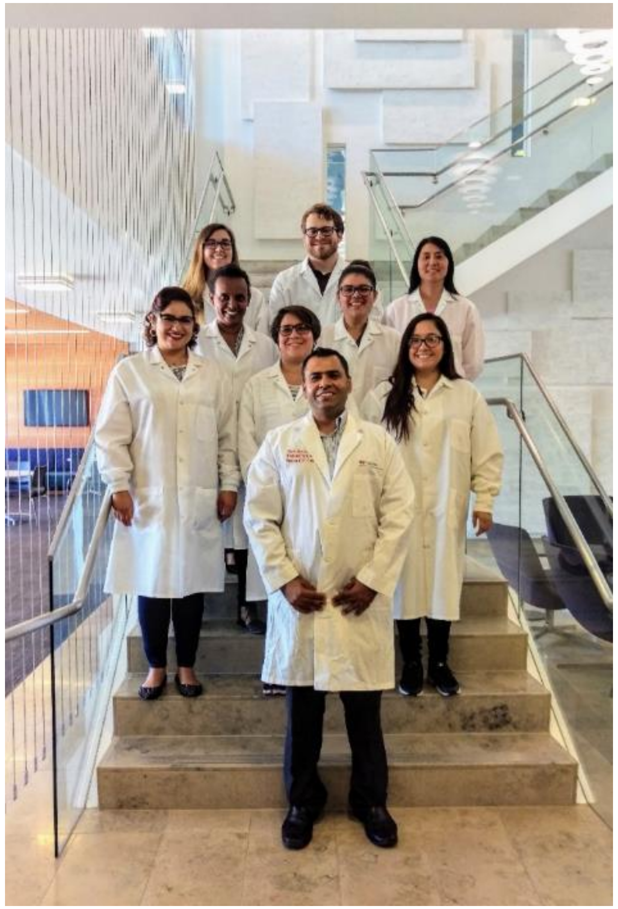


CEDARS-SINAI



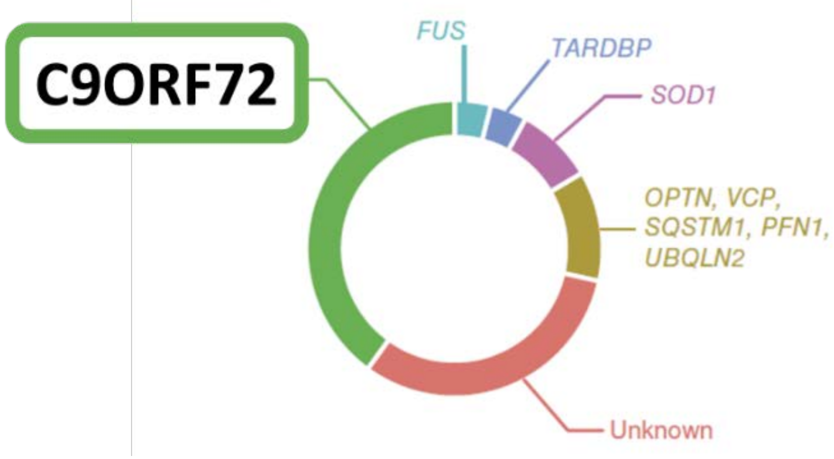
NEUROLINCS

Neurolincs.org



iPSC team

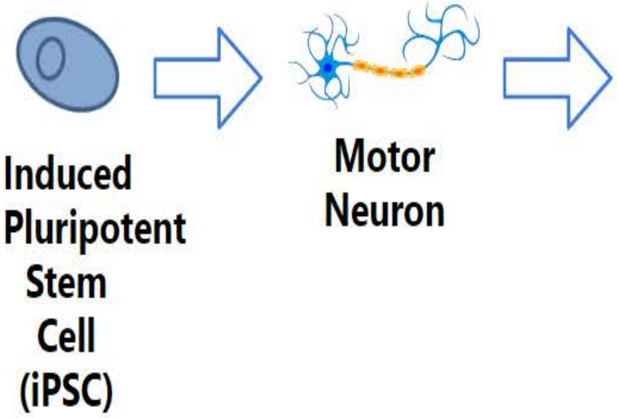
Differentiation team



Leslie
Thompson

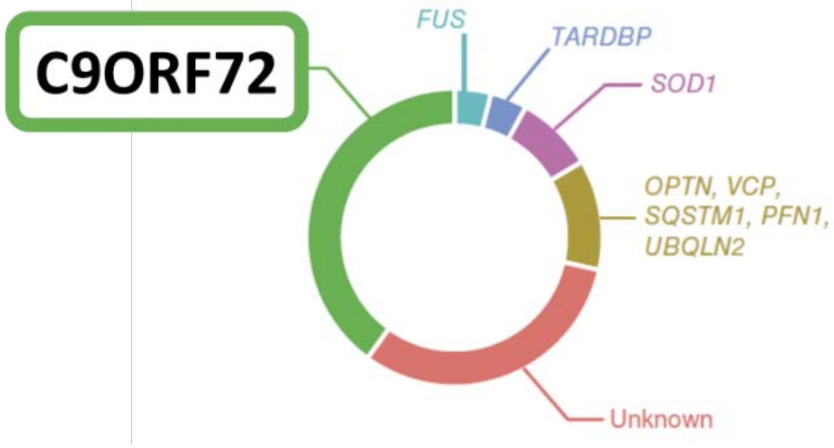
UCIrvine
University of California, Irvine

Total
RNA-Seq



NEUROLINCS

Neurolincs.org

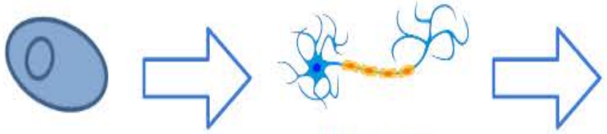


Leslie Thompson

UCIrvine
University of California, Irvine

Jenny Van Eyk

CEDARS-SINAI

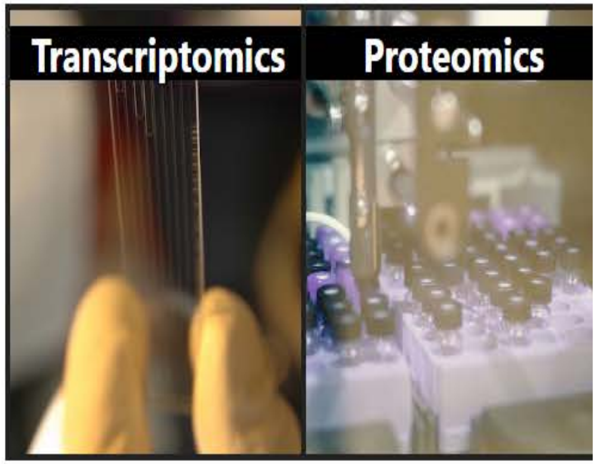


Induced Pluripotent Stem Cell (iPSC)

Motor Neuron

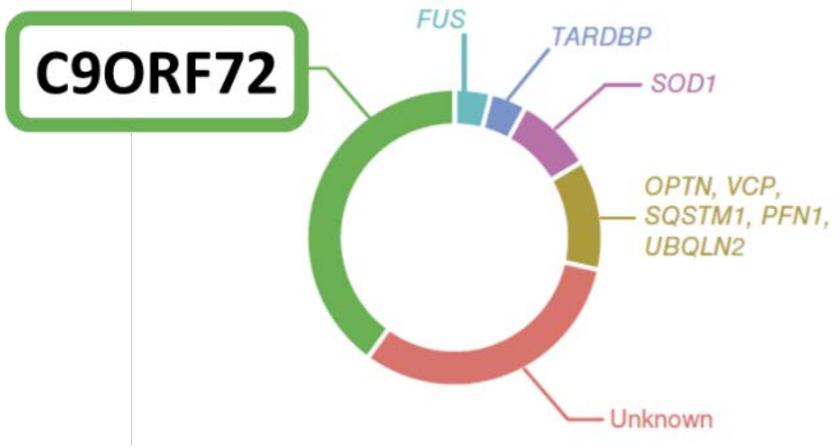
Total RNA-Seq

SWATH MS-based Proteomics



NEUROLINCS

Neurolincs.org

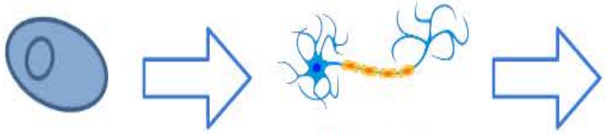


Leslie Thompson

UCIrvine
University of California, Irvine

Jenny Van Eyk

CEDARS-SINAI



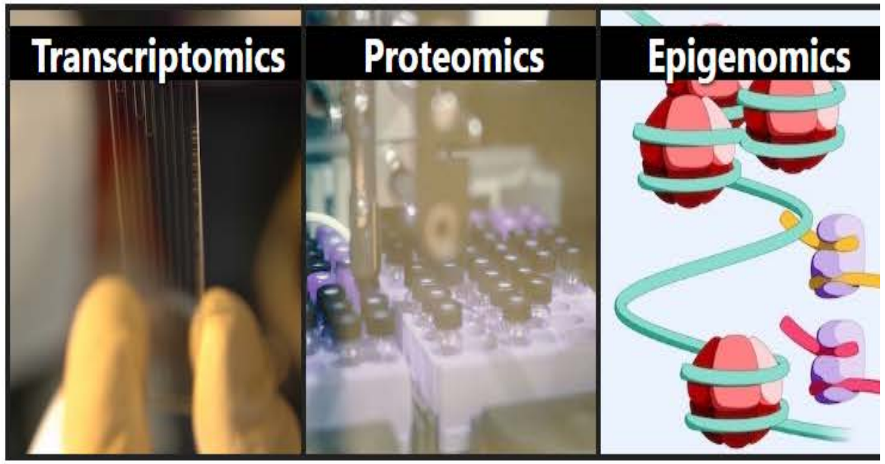
Induced Pluripotent Stem Cell (iPSC)

Motor Neuron

Total RNA-Seq

SWATH MS-based Proteomics

ATAC-Seq



Pamela Milani

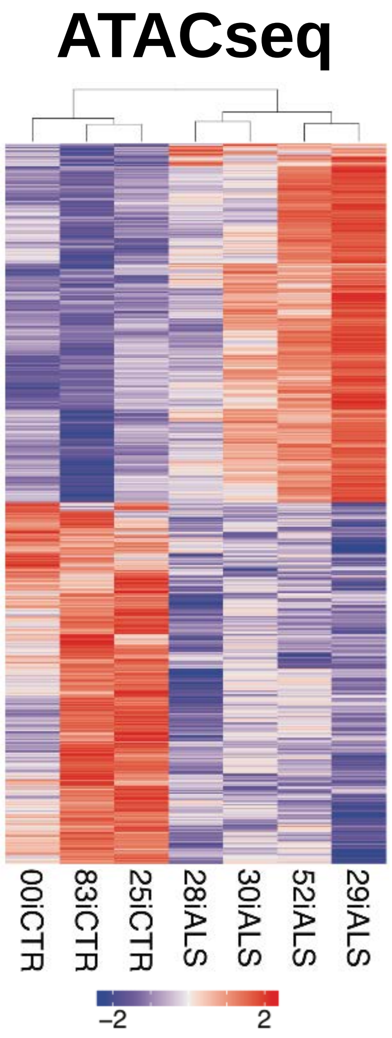
Miriam Adam

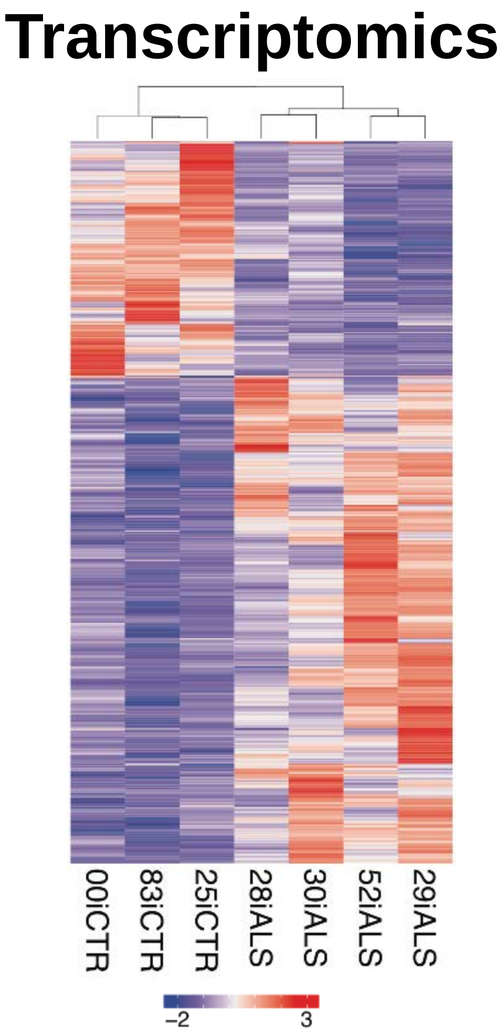
Brook Wassie

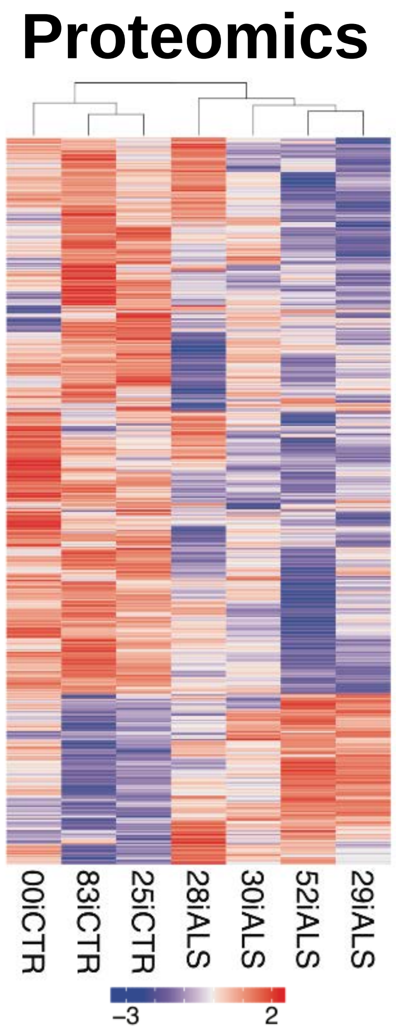
NEUROLINCS

Neurolincs.org

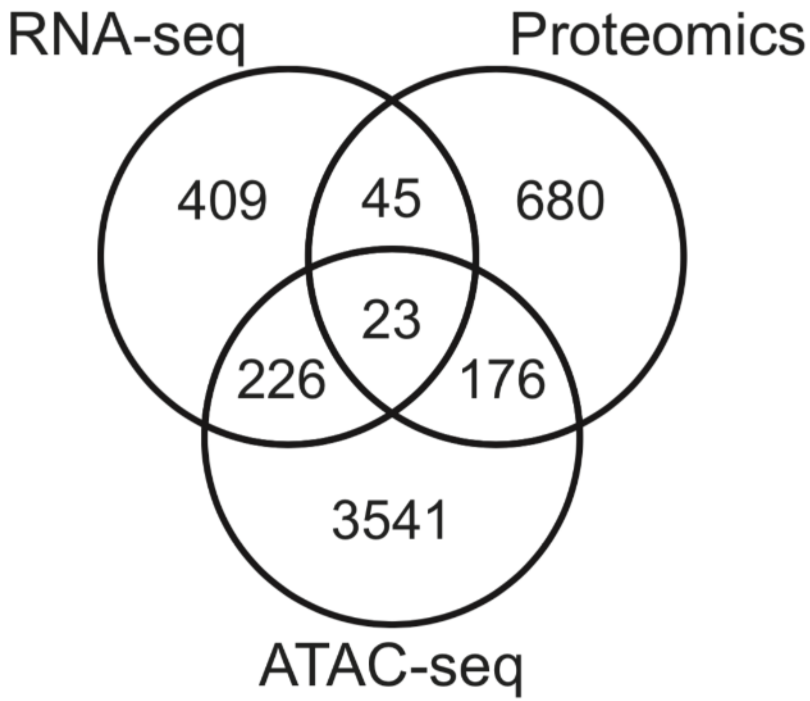
Mult-omic differences between c9ORF72 ALS and Control iMNs

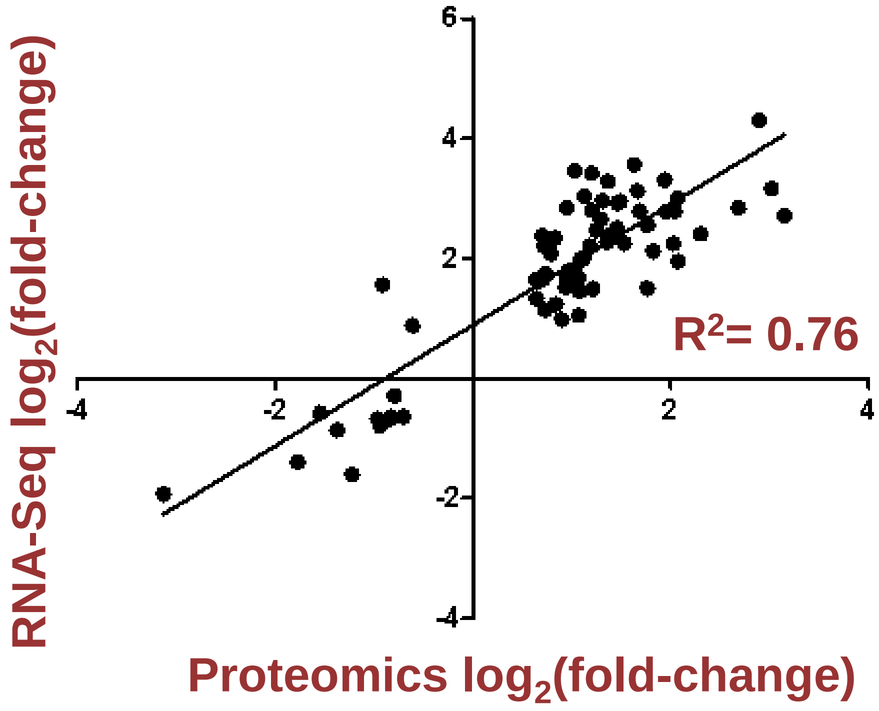




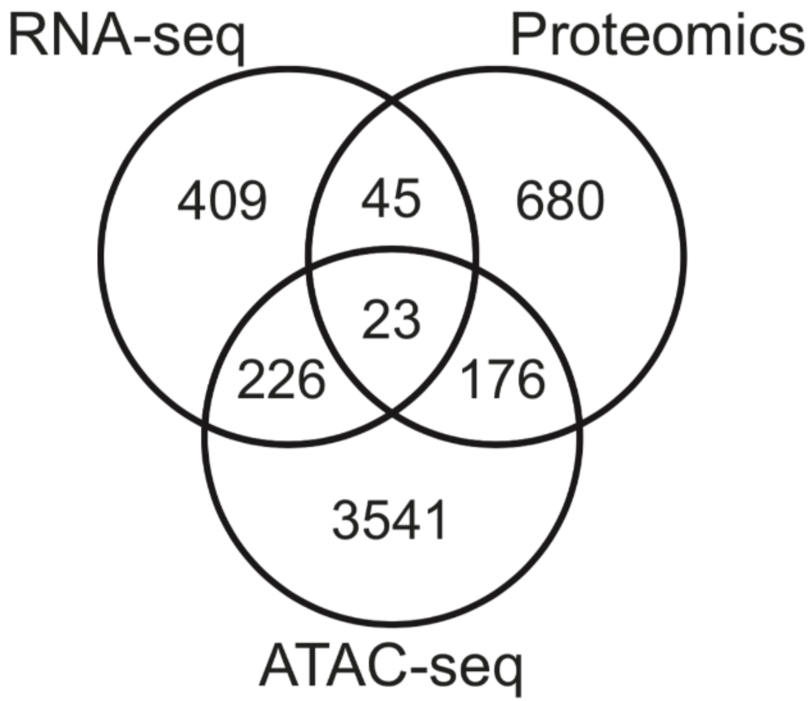


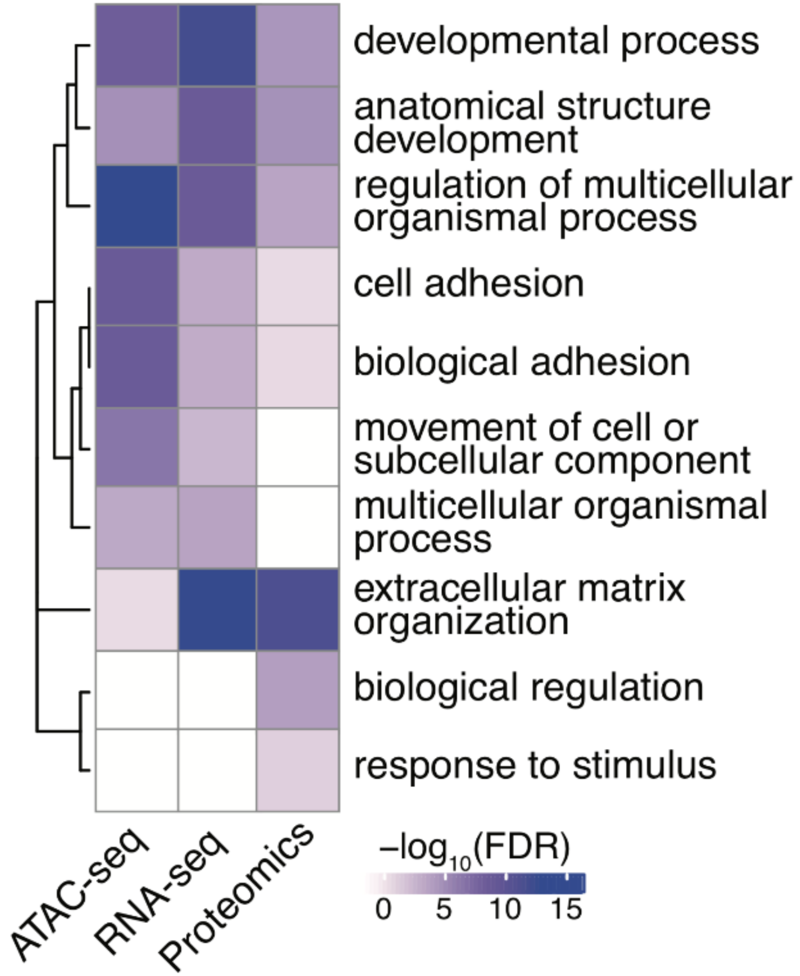
Low Overlap in Our Data, as Expected





Low Overlap, but Common Pathways



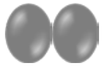


Network Integration

Johnny Li

Prior interactome network

Yeast two-hybrid
Affinity capture mass-spec
Protein-protein interactions



Omics Integrator

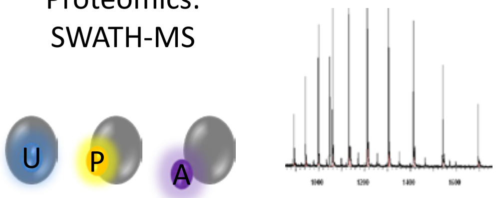
Omics from patient-derived iPSCs

ATAC-Seq
Protein-DNA interactions

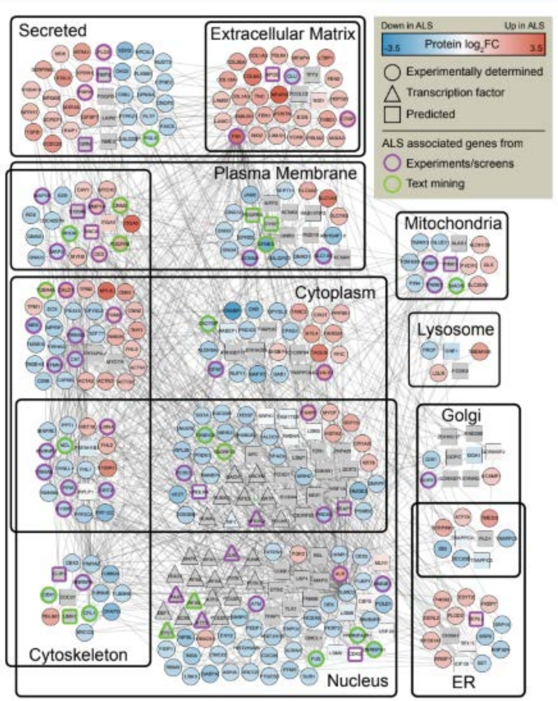
RNA-Seq
mRNA

AAATAGCCATTATACGTA
CCTAATACTGAAGAGTCA
TTCCTAGTAAAGCATGCT
ACTTTTCAGTATATTCCA
TTATATTTTTAACTACAA
GCGGCGCAGAAACCAGAG

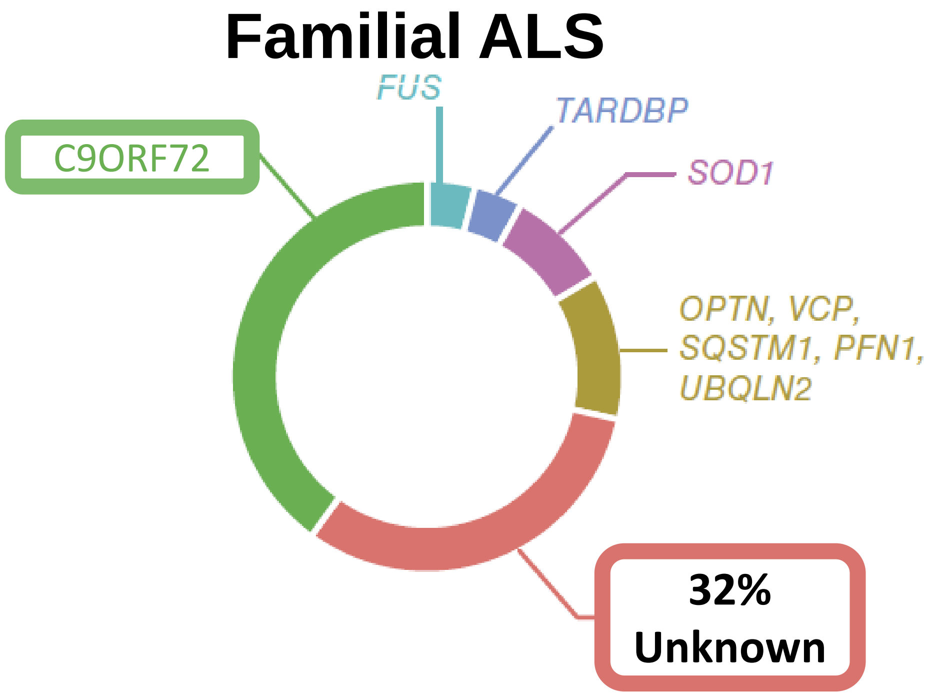
Proteomics:
SWATH-MS

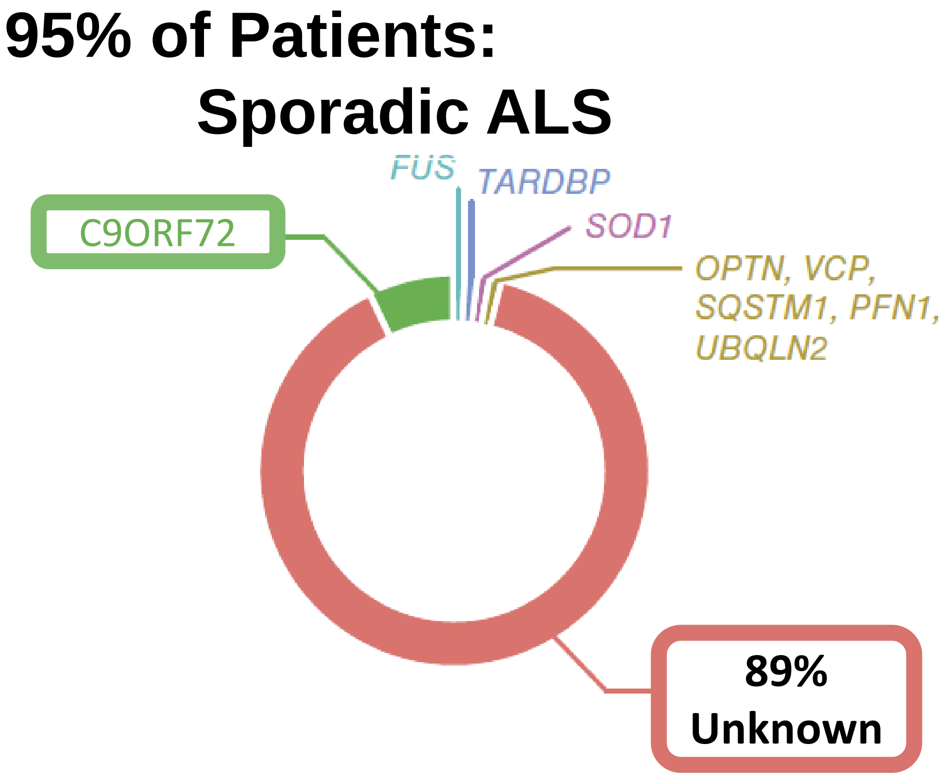


Test for robust and significant networks



Causes for 85% of ALS still not known

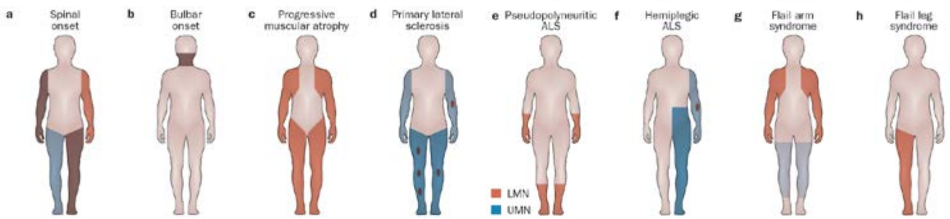


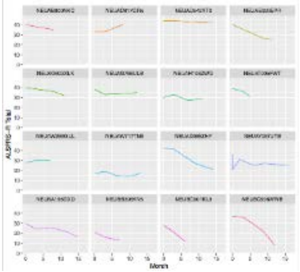


[State of play in amyotrophic lateral sclerosis genetics]
Nature Neuroscience 17,17–23 (2014)
doi:10.1038/nn.3584

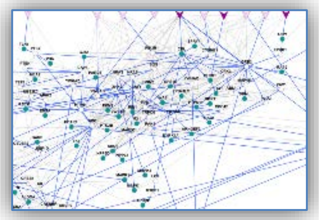
How many types of ALS are there?

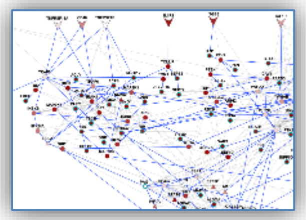
CLINICAL VARIANTS

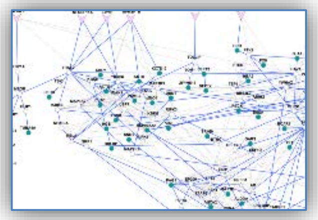


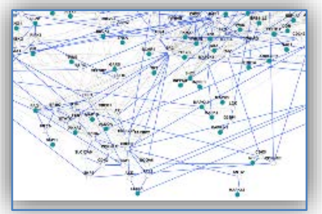


MOLECULAR VARIANTS

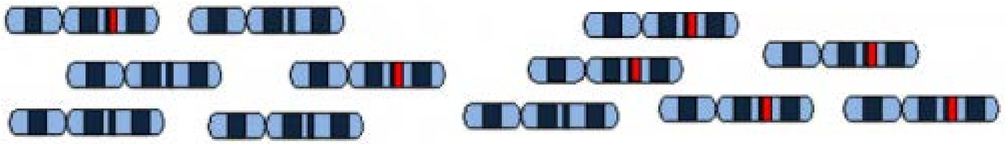


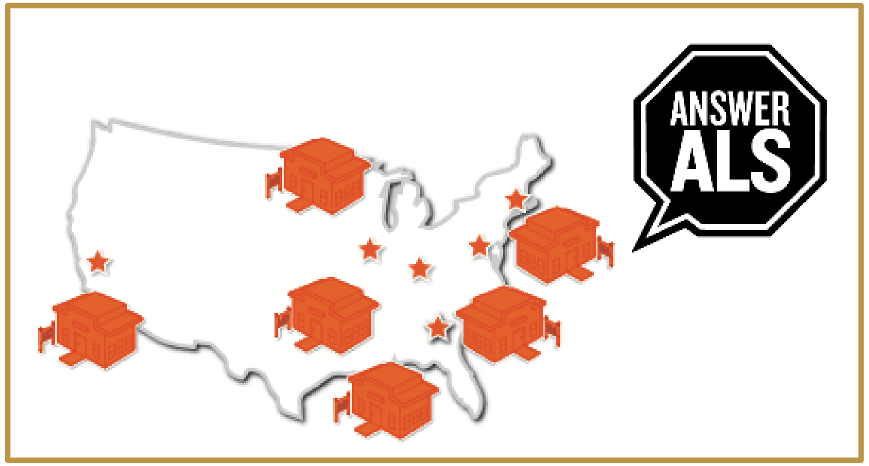






GENOMIC VARIANTS



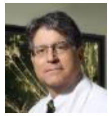


Clinical

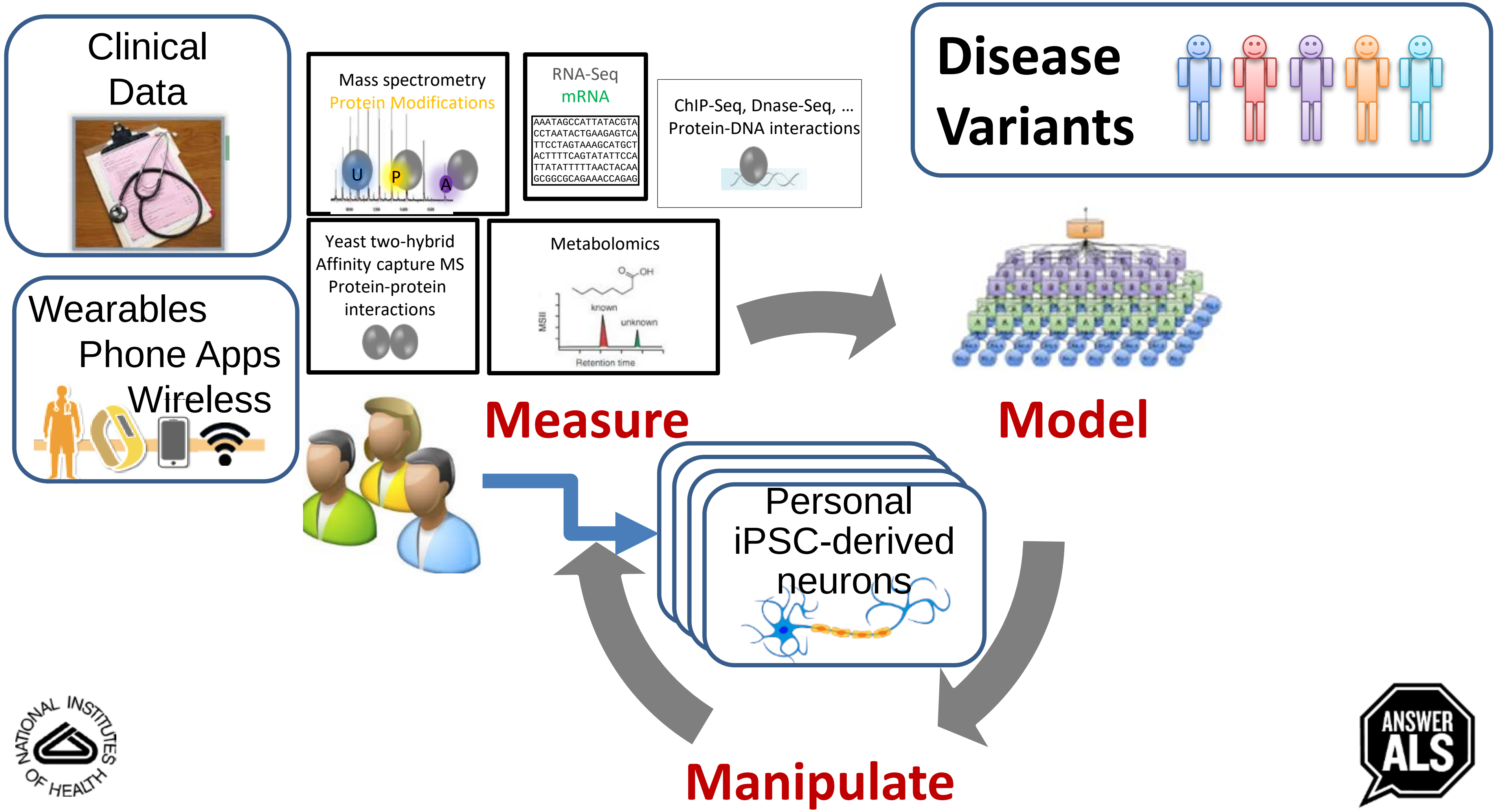
MGH 1811 Massachusetts General Hospital		James Berry
Johns Hopkins Medicine		Nicholas Maragakis
Emory University		Jonathan Glass
The Ohio State University		Stephen Kolb
Washington University in St. Louis		Timothy Miller
Cedars-Sinai		Bob Baloh

New York Genome Center

Research

Leslie Thompson		UC Irvine University of California, Irvine
Jeff Rothstein		Johns Hopkins Medicine
Steve Finkbeiner		UCSF University of California San Francisco
Ernest Fraenkel		MIT Massachusetts Institute of Technology
Jenny Van Eyk		Cedars-Sinai
Clive Svendsen		

The engineering design cycle could discover cures for ALS

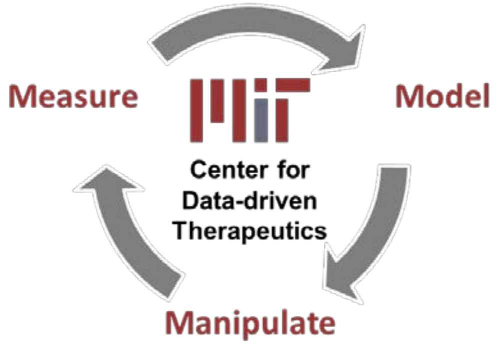


PGA NFL alsFINDINGaCURE Jay and Randy Fishman Family Foundation TEAM GLEASON ALS ASSOCIATION

The future

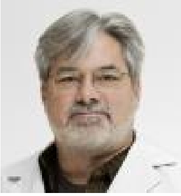
Rapid discovery of therapeutics through closed-loop design

Integration of clinical research, biological engineering and AI

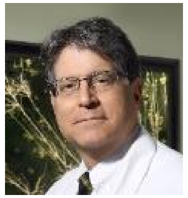


Leslie Thompson

Steve Finkbeiner



Jeff Rothstein



Jenny Van Eyk

Clive Svendsen