Translation and personalised medicine: Genome medicine meets reality

Winston Hide hidelab.wordpress.com @winhide
Sheffield Institute for Translational Neurosciences

http://dx.doi.org/10.6084/m9.figshare.1512438

Pre

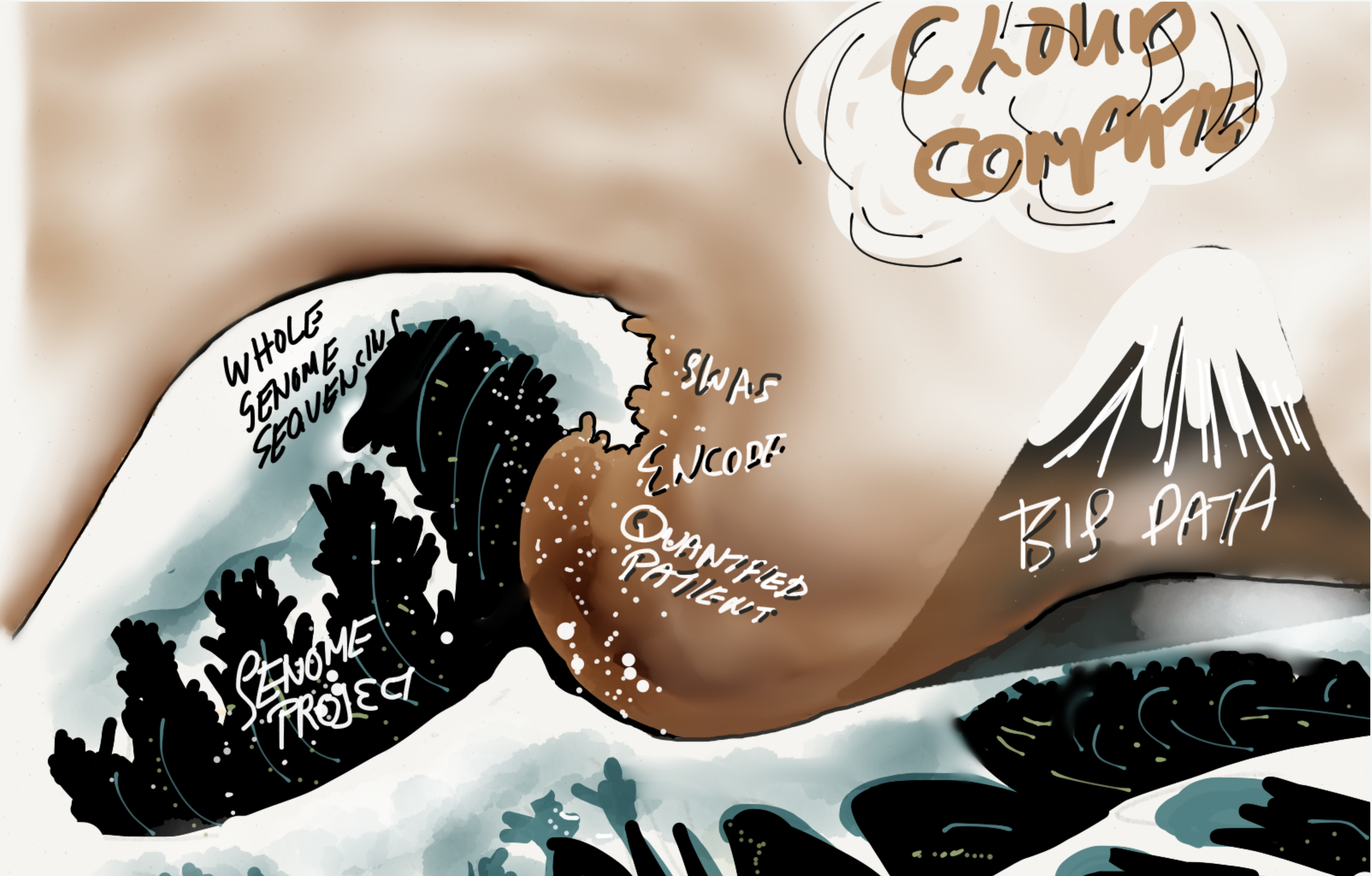
- I consult for Biogen Inc
- Russ Altman's **top translational informatics 2015** papers review has provided material throughout

What is Translational Medicine?

- Translation from **biological**
  - proteins, genes, small molecules
  - model organisms
  - cellular assays
- To **clinical**
  - diseases
  - symptoms
  - drugs

- **Scientific Discovery**
- **Health Improvement**

Brilliant successes with lots of failures


CLOUD COMPUTING
WHOLE GENOME SEQUENCING
GWAS
ENCODE
QUANTIFIED PATIENT
GENOME PROJECT
BIG DATA

Biomedical Discovery is becoming a data science

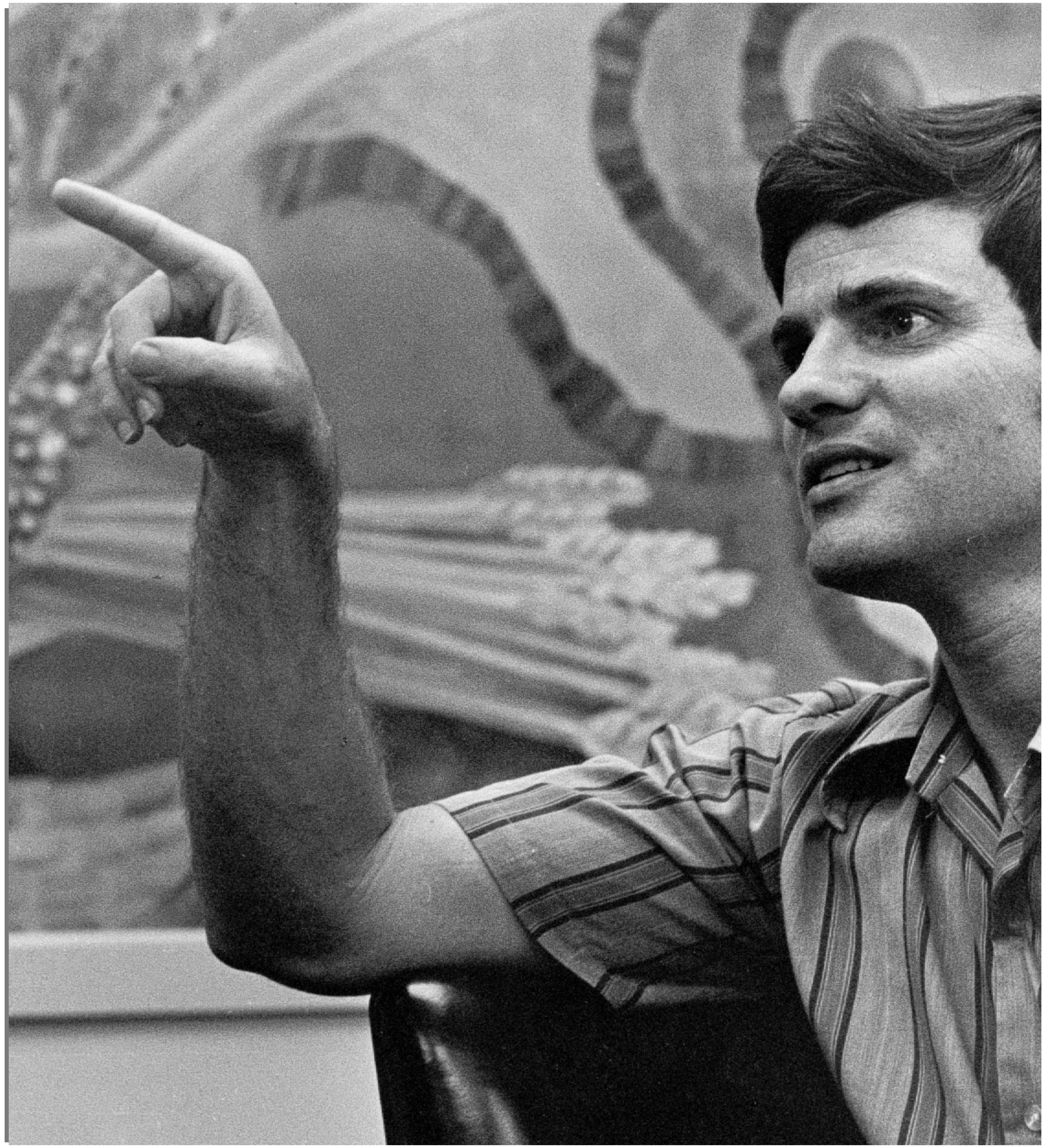
Lee Hood

(Institute for Systems Biology)

1993

Smith Kline Hiring spree

ISMB 1996 June 12-15, 1996, St. Louis, Missouri, USA



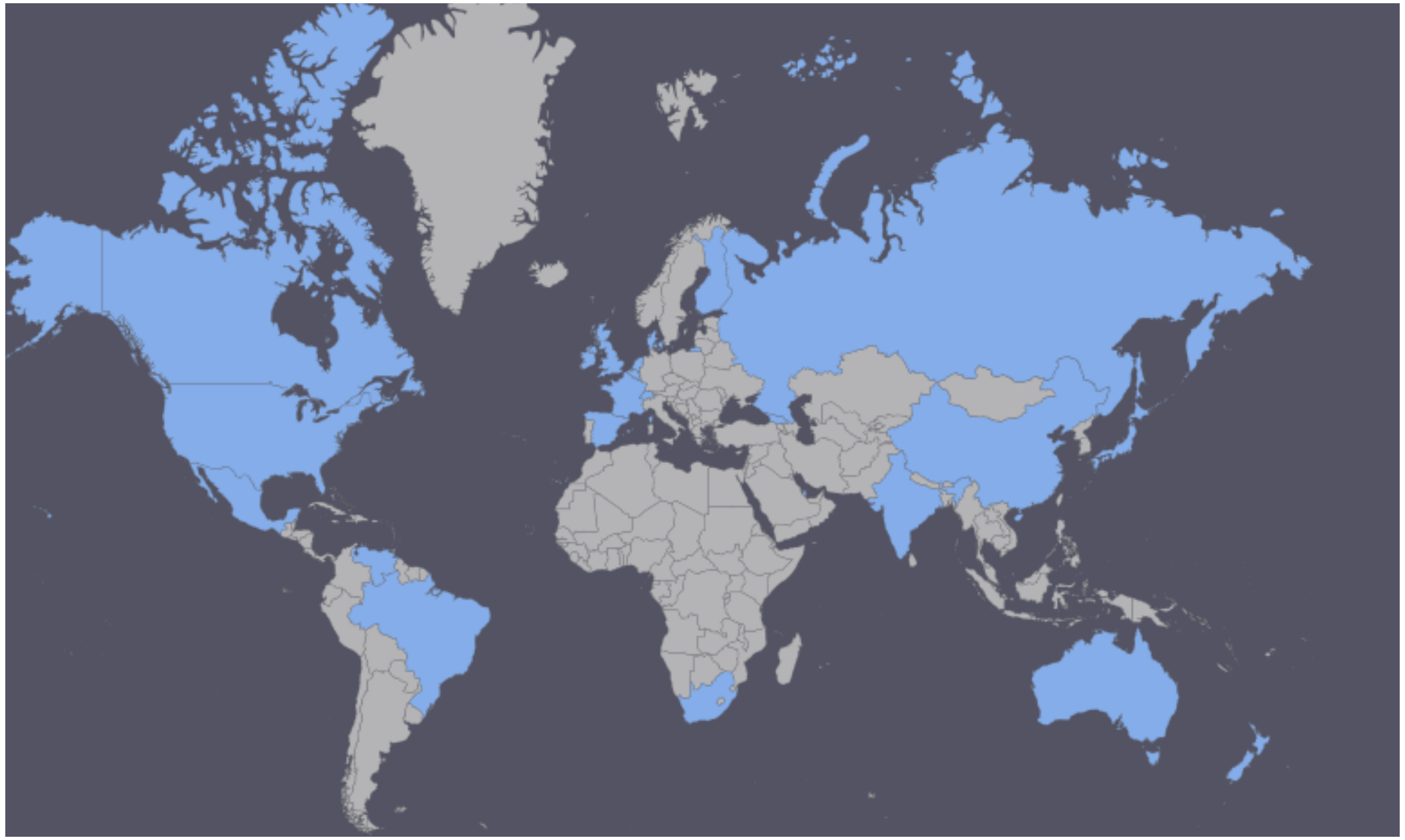
Clinicians are learning to type

Electronic **H**ealth **R**ecords

Hospital IT

Innovation and industry

Space and capacity



From population to patients

1M by 2016

Diagnostics

Genomepatri™ -
The new scientific way to map your health risks using DNA technology

Genetic counseling helps individuals take actionable steps towards a healthier 'you'

Apps + GeneticData
Real-Time Access

Turn genetic data into useful information.

Free, unlimited and secure storage of genetic data.

Get to know yourself with Insight.

The advanced DNA genetic test that offers you clear and actionable findings for healthier life and brighter future.

Companies offering analysis

1 Week

Craig Venter to offer DNA sequencing to consumers

Share this article

83 46 6 13

22 September 2015

Health Insurer

App ecosystem

netic
ofile

4 Million patients

Whole Exome

$250

Illumina-Helix

For $25,000 individuals can enjoy an “executive physical on steroids,”

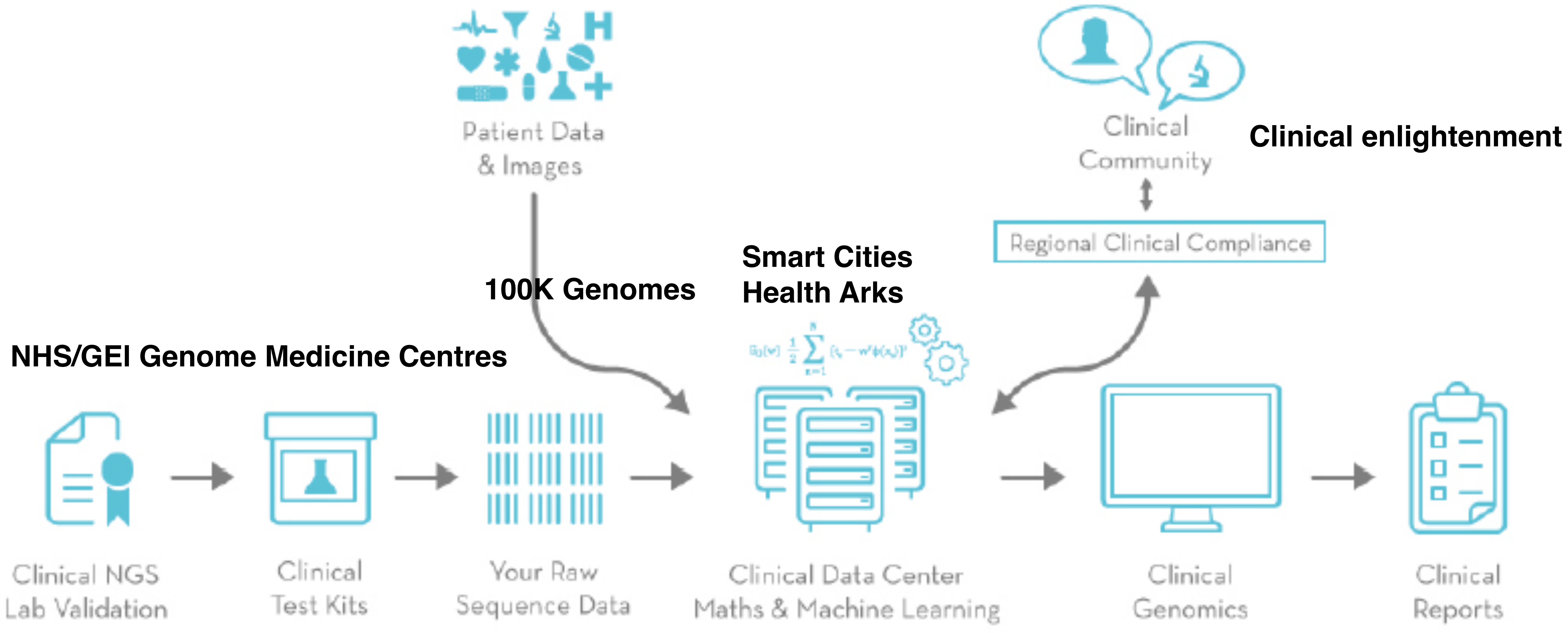
The company is the database

HLI complete omics workover

Converting human variation into clinical action?

- Guidelines for reporting disease-causing variants
- Evidence for candidate genes
  - Pathogenic variants
  - Supporting publications and reporting
- **Challenge of NGS clinical diagnostic sequencing interpretation**

mc Arthur, Nature 2015



Clinical enlightenment

100K Genomes

Smart Cities
Health Arks

NHS/GEI Genome Medicine Centres

The Genome-medicine_dev UK

Graphics from Sophia Genetics

Variant interpretation

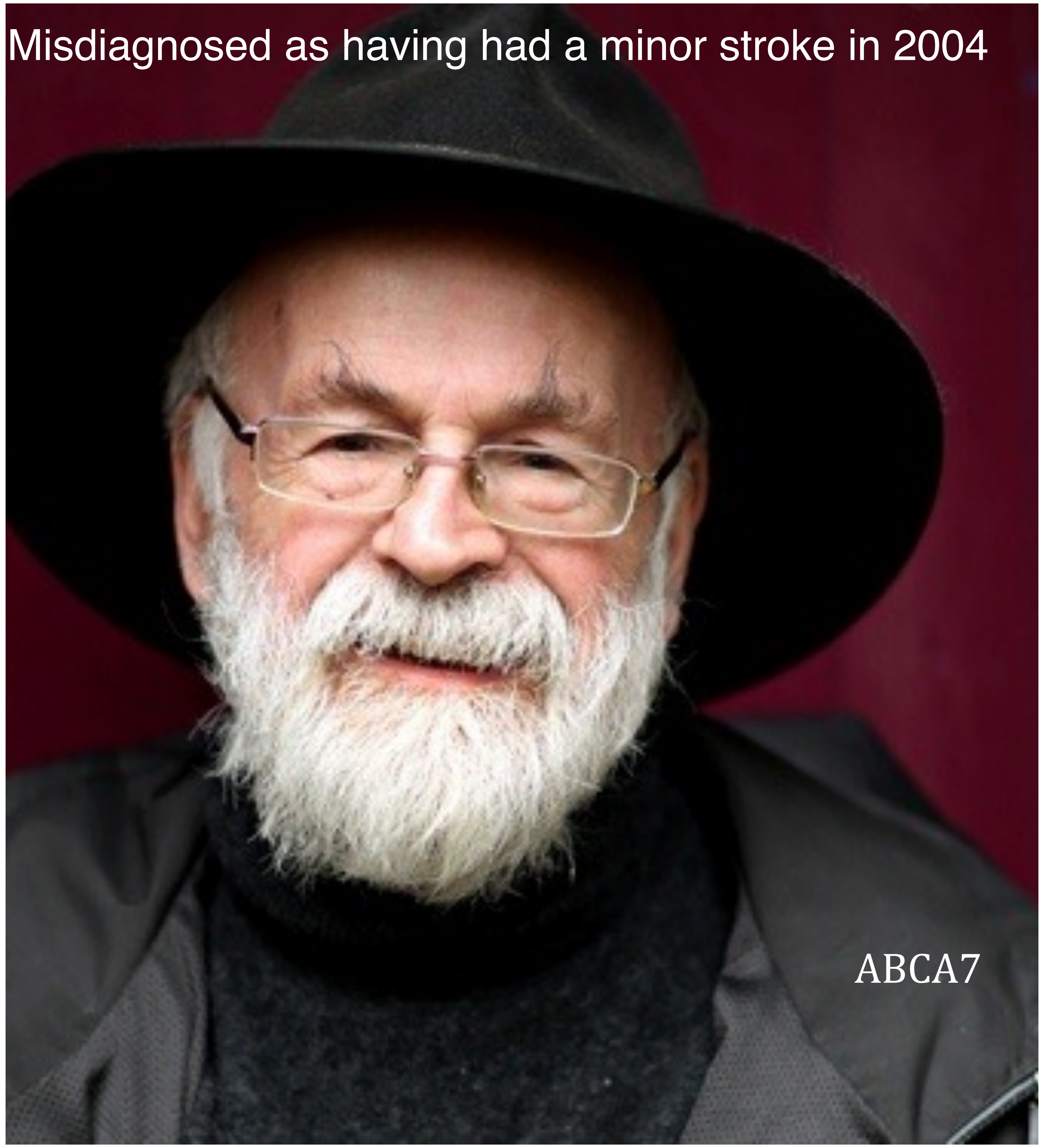
- Most **GWA signals** are non-coding
- Two alleles/genome - **allele specific** expression
- The genome is **3 dimensional** (HiC)
- **Functional interactions** occur after transcription and translation
- "orchestra of regulatory elements, variants at critical positions define what kind of music you are going to play” ***Albin Sandelin Karolinska***
- SNP linked to **specific expression states** eQTL
- **Genetic stratification** easier than functional interpretation

Sequence 1500 Alzheimer's genomes

437 Families - Sequence Whole Genomes

Discovered new variants that cause AD

Misdiagnosed as having had a minor stroke in 2004



ABCA7

Rudy Tanzi, Harvard Medical School

Motor Neurone Disease

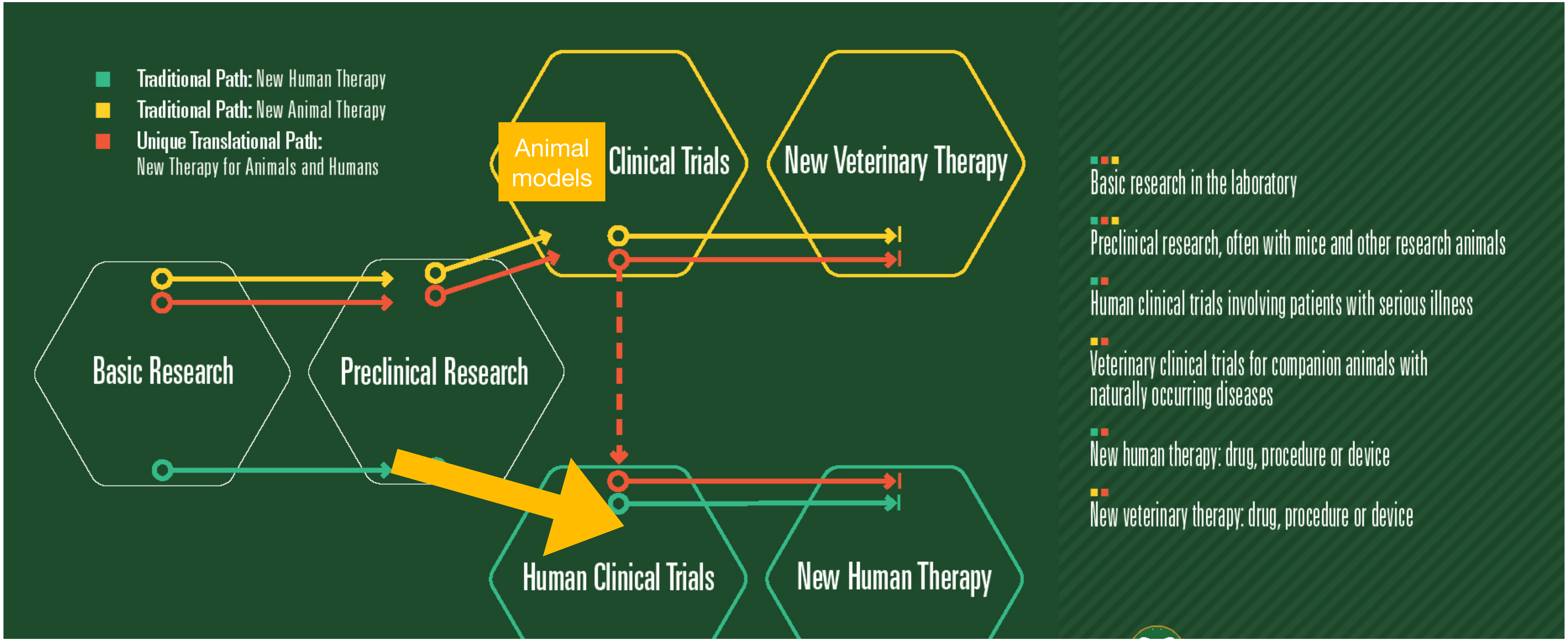
Progressive weakness of many of the muscles in the body

‘Handful’ of known genetic causes for about 10% of patients

Scale up to 15000 Amyotrophic Lateral Sclerosis Patients

15000 people with ALS compared to **7500** healthy individuals.

Therapy



- 🟩 **Traditional Path:** New Human Therapy
- 🟨 **Traditional Path:** New Animal Therapy
- 🟥 **Unique Translational Path:** New Therapy for Animals and Humans

🟩🟥🟨 Basic research in the laboratory

🟩🟥🟨 Preclinical research, often with mice and other research animals

🟩🟥 Human clinical trials involving patients with serious illness

🟨🟥 Veterinary clinical trials for companion animals with naturally occurring diseases

🟩🟥 New human therapy: drug, procedure or device

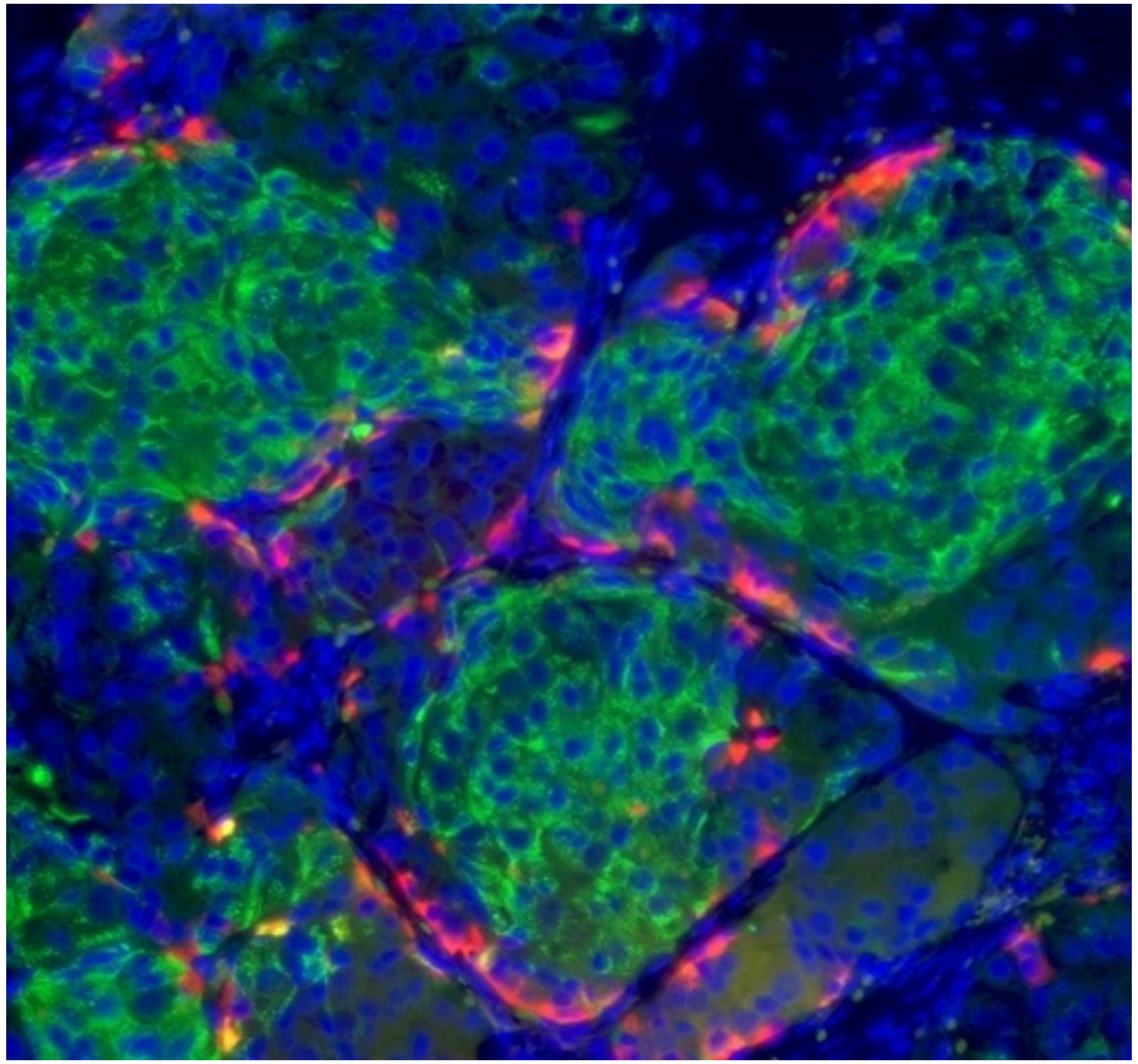
🟨🟥 New veterinary therapy: drug, procedure or device

Pathway to therapy is becoming humanised

Colorado State University Website

Translation of Stem cell biology

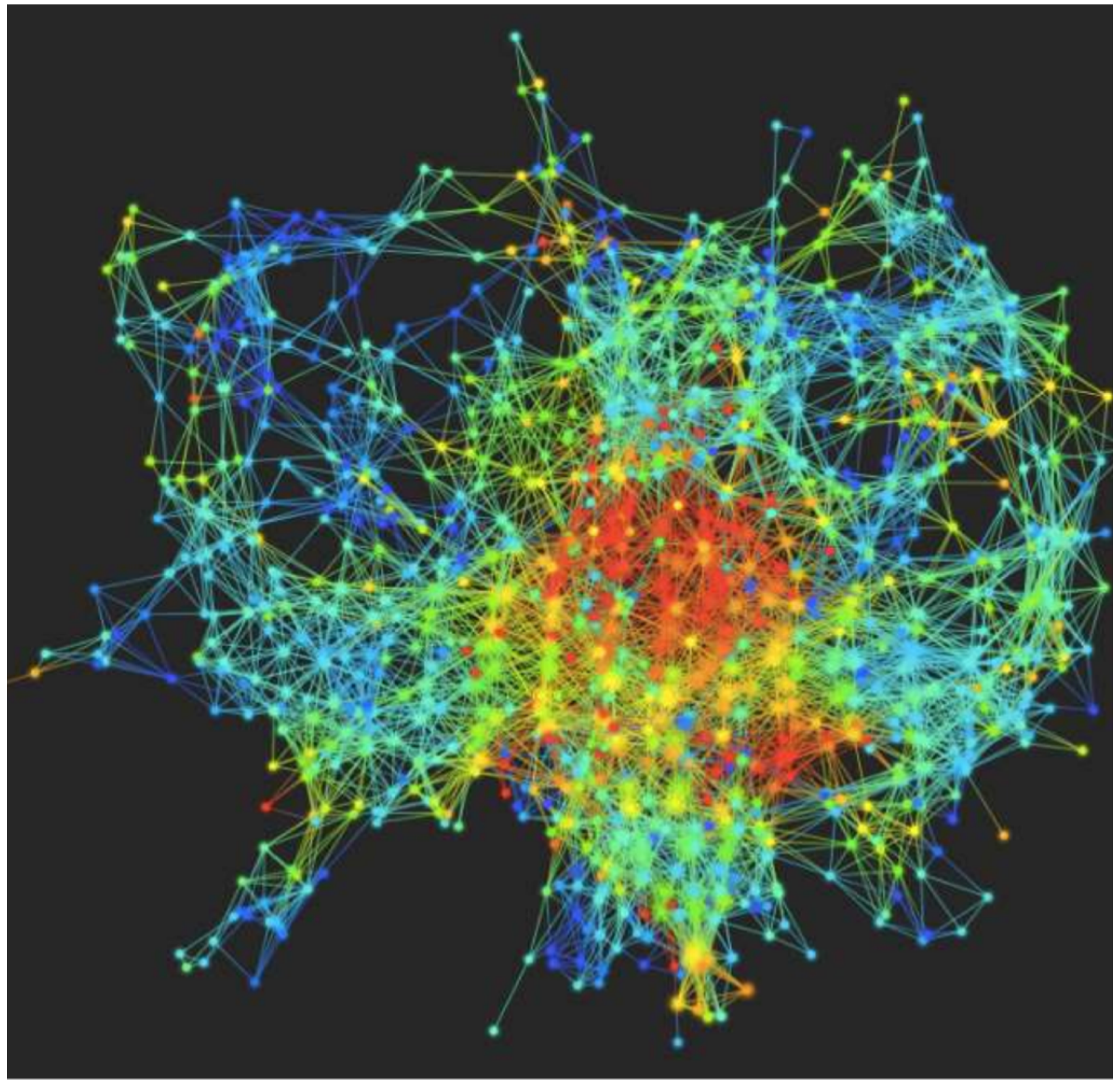
- **Transdifferentiated cell models**
  - Eggan and Woolf, first human transdifferentiated cell model: Amytrophic Lateral Sclerosis 'rescued' ALS phenotype
- **Regeneration**
  - Melton: SC-β cell transplantation ameliorates hyperglycemia in mice
  - Provide a platform for therapeutic development and disease modeling



"Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons.,"Woolf
"Generation of Functional Human Pancreatic β Cells In Vitro," Melton

Genes to networks to drugs

Interacting genes drive disease

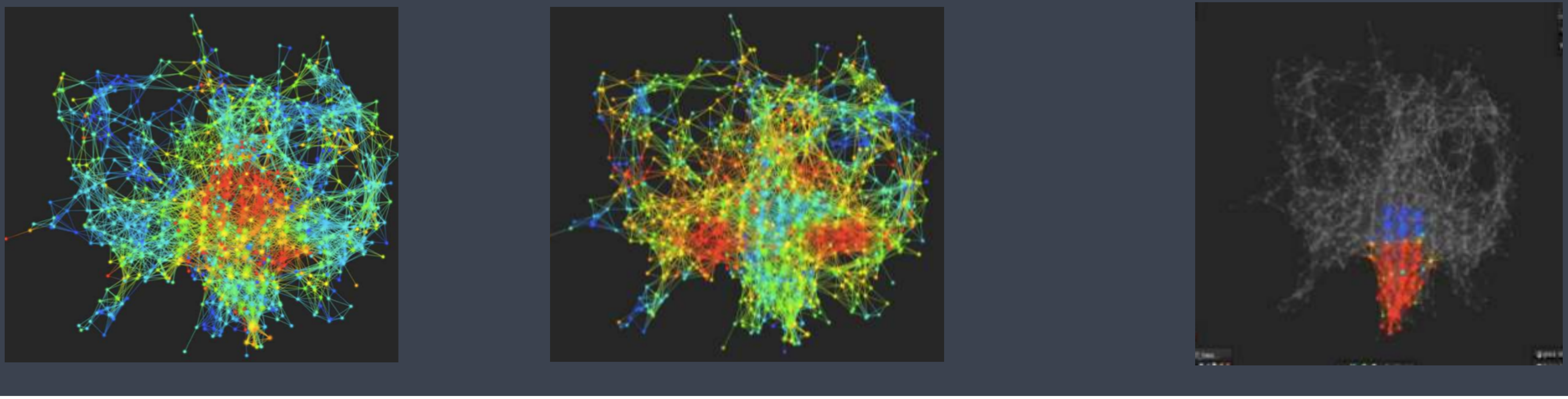


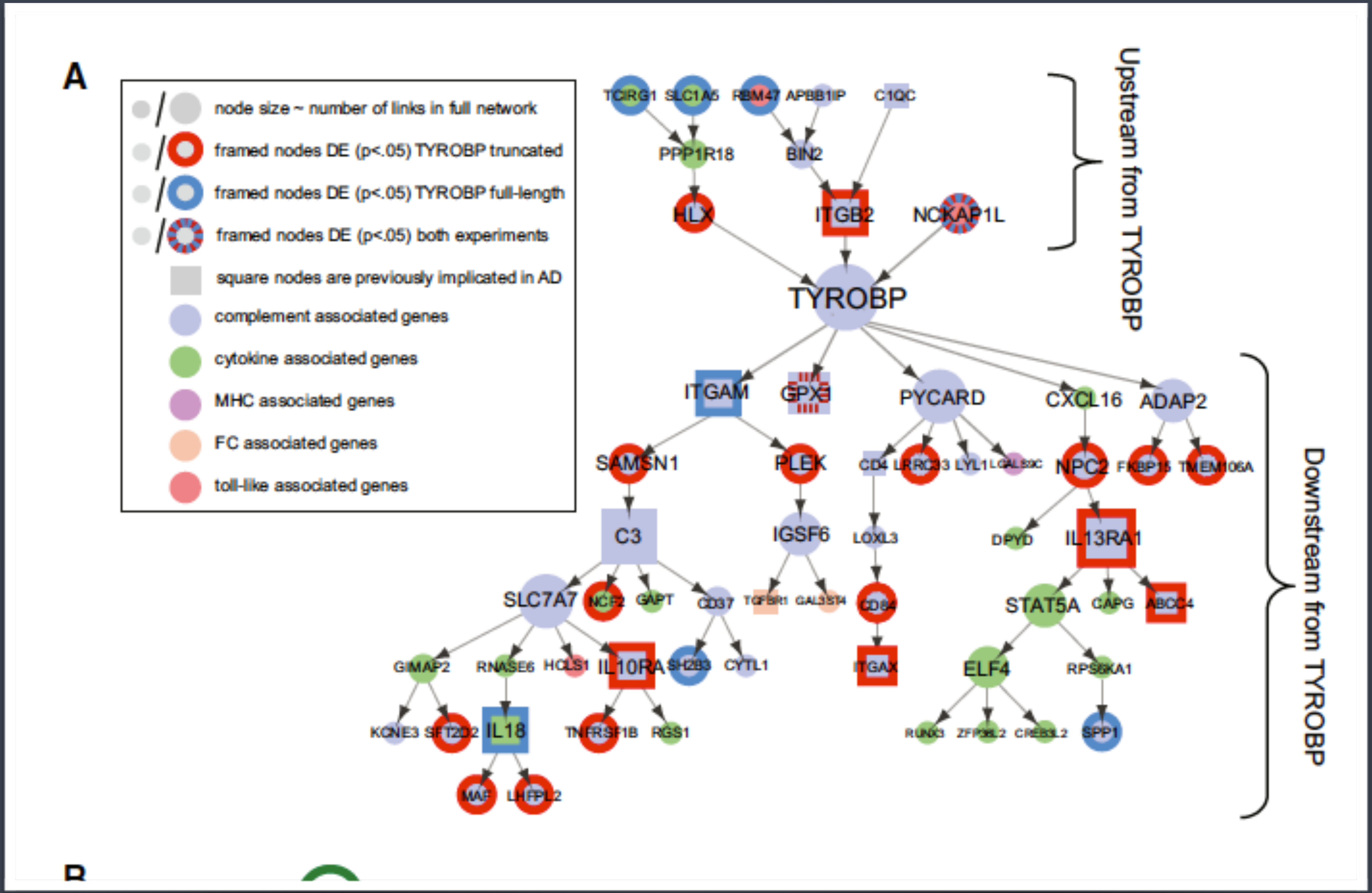
Models

- Fly
- Mouse
- Human
- Cohorts

- Genetics
- Mutant physiology
- iPSC models
- Genome variants

- Expression pathways
- Pathways/eQTL
- Functional assays/ expression
- Expression/omics




A
node size ~ number of links in full network
framed nodes DE (p<.05) TYROBP truncated
framed nodes DE (p<.05) TYROBP full-length
framed nodes DE (p<.05) both experiments
square nodes are previously implicated in AD
complement associated genes
cytokine associated genes
MHC associated genes
FC associated genes
toll-like associated genes
TCIRG1
SLC1A5
RBM47
APBB1IP
C1QC
PPP1R18
BIN2
HLX
ITGB2
NCKAP1L
Upstream from TYROBP
TYROBP
ITGAM
GPX3
PYCARD
CXCL16
ADAP2
SAMSN1
PLEK
CD4
LRRC33
LYL1
LGALS9C
NPC2
FKBP15
TMEM106A
C3
IGSF6
LOXL3
DPYD
IL13RA1
SLC7A7
NCF2
GAPT
CD37
TGFBR1
GAL3ST4
CD84
STAT5A
CAPG
ABCC4
GIMAP2
RNASE6
HCLS1
IL10RA
SH2B3
CYTL1
ITGAX
ELF4
RPS6KA1
KCNE3
SFT2D2
IL18
TNFRSF1B
RGS1
RUNX3
ZFP36L2
CREB3L2
SPP1
MAF
LHFPL2
Downstream from TYROBP
B

Pre-clinical validation

- **AMD NIH Partnership**
  - **Network sharing consortium**
  - **Preclinical consortia**
  - **Takeda - we will validate your model openly**

Alzheimer's Disease

The National Institutes of Health (NIH), 10 biopharmaceutical companies, and several non-profit organizations have designed an unprecedented new partnership. Managed through the Foundation for the NIH (FNIH), the Accelerating Medicines Partnership (AMP) brings high-level government, industry, and non-profit organization partners together to identify and validate the most promising biological targets of disease for new diagnostic and drug development. The partners have designed a bold milestone-driven research plan to tackle this challenge for **Alzheimer's disease**, as well as for type 2 diabetes and the autoimmune disorders of rheumatoid arthritis and systemic lupus erythematosus (lupus). Importantly, the AMP data and analyses will be made publicly available to the broad biomedical community. This fact sheet addresses the AMP research plan for Alzheimer's disease.

AMP AD

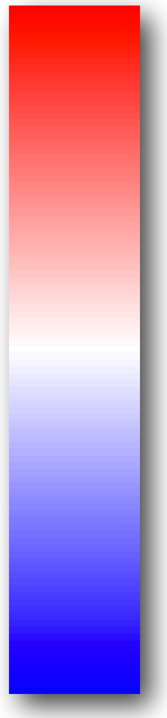
- Diverse datasets from multiple labs
- **Omics datasets** – RNAseq, Whole Exome, miRNA, Methylation, iPSC and MS
- **Mouse models** – APPPS1, TgCRND8, P301L tau (MAPT) mice and (JNPL3 strain), rTg4510
- The **network** is the data
- Target leader board

Quantitating function

Pathway Fingerprinting

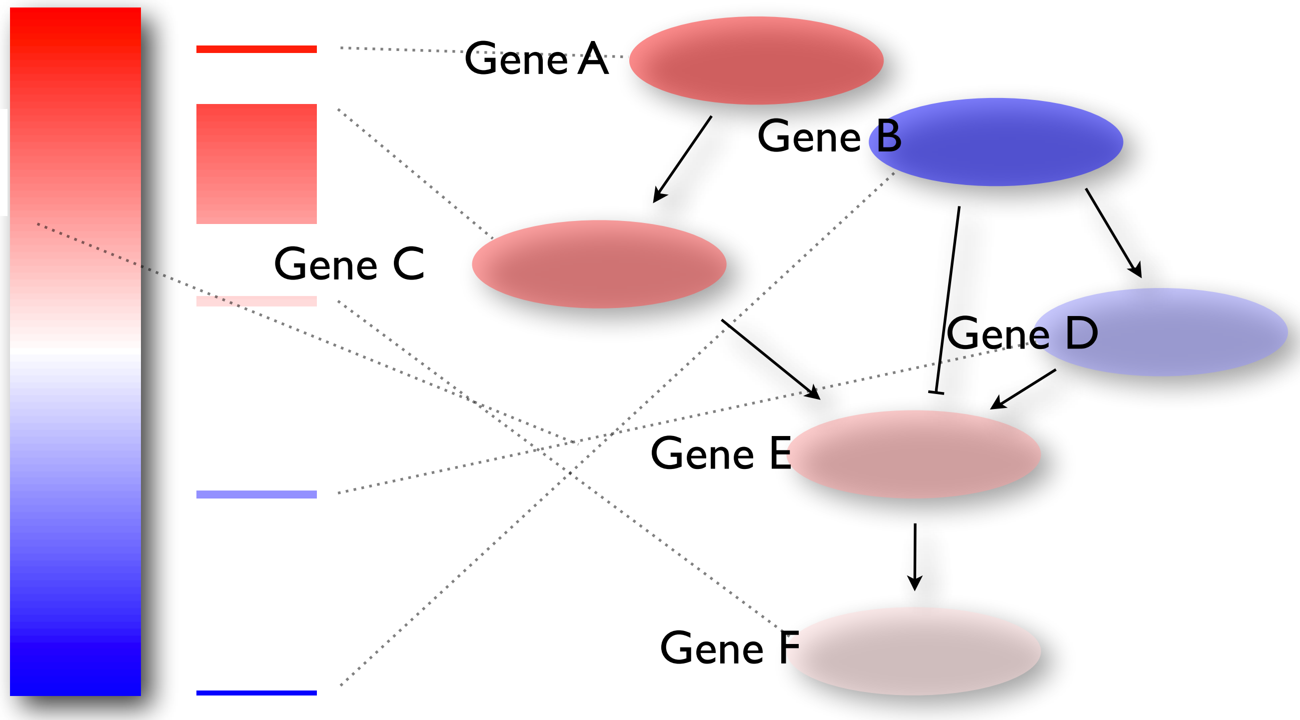
Pathway fingerprinting
A tool for universal comparison of functional states

Gene expression



Gene expression

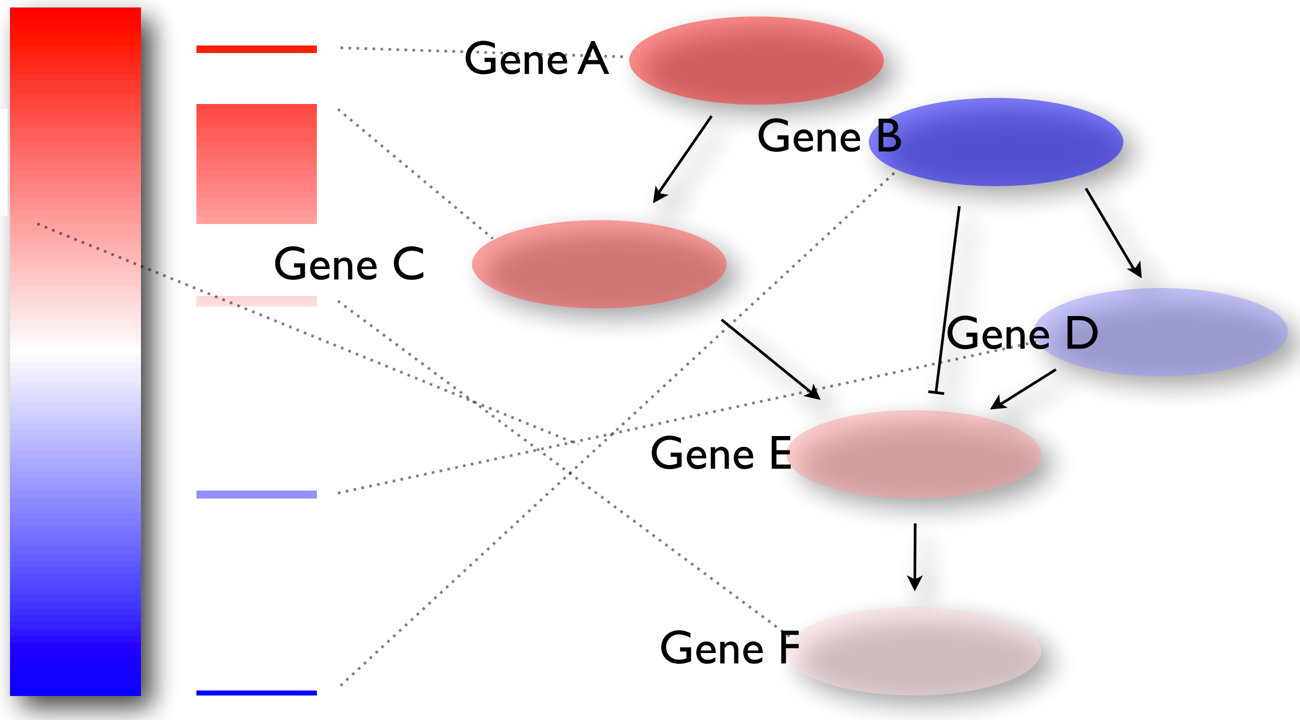
Pathway Expression



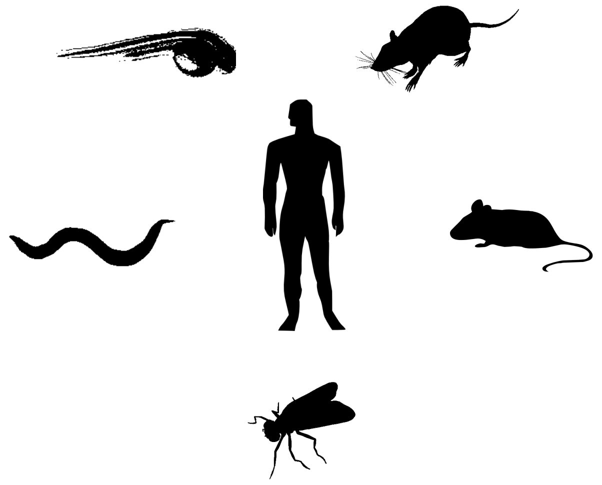
Gene expression

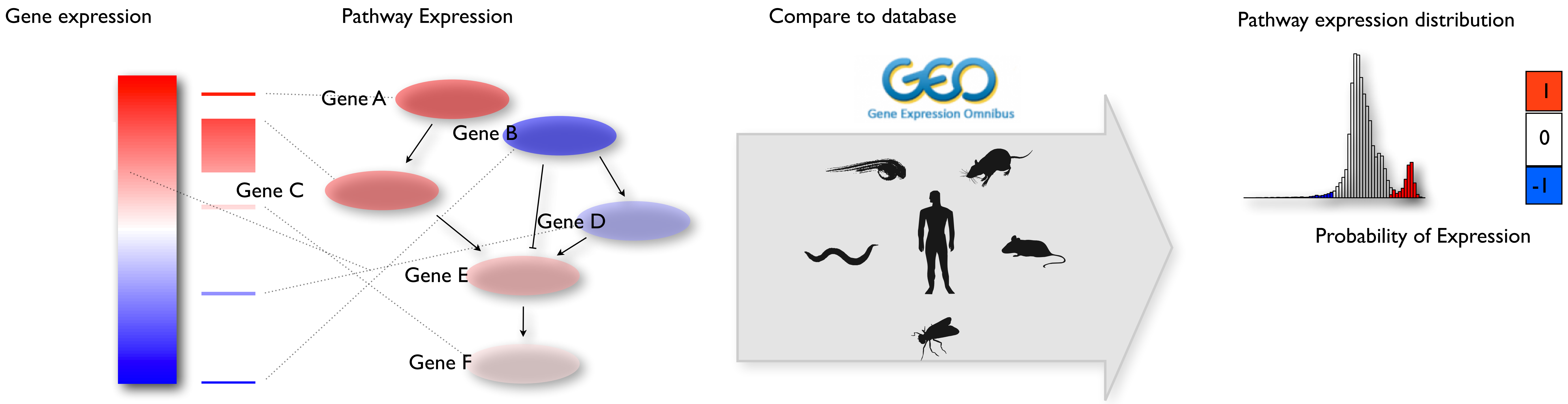
Pathway Expression

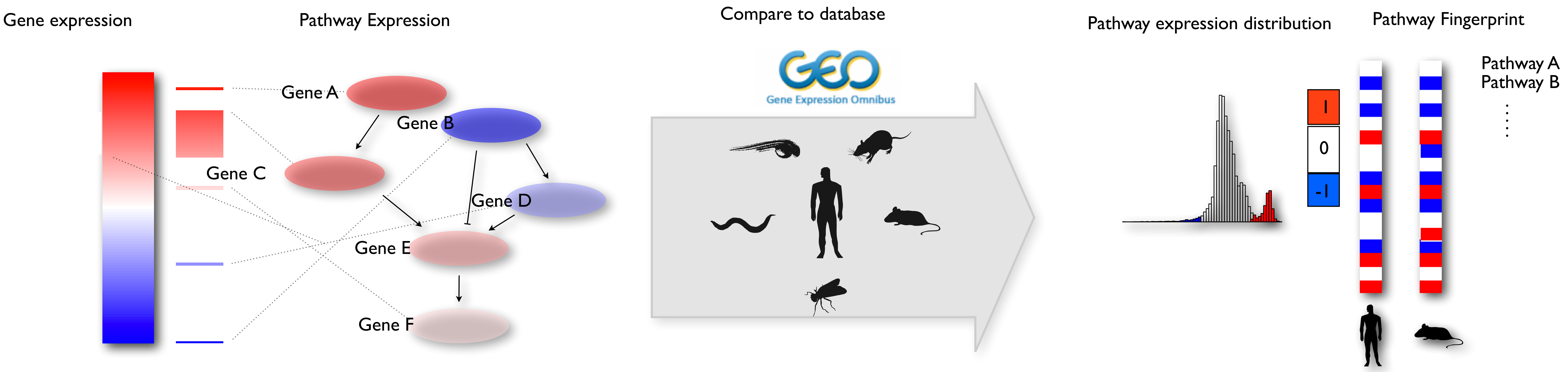
Compare to database



GEO
Gene Expression Omnibus

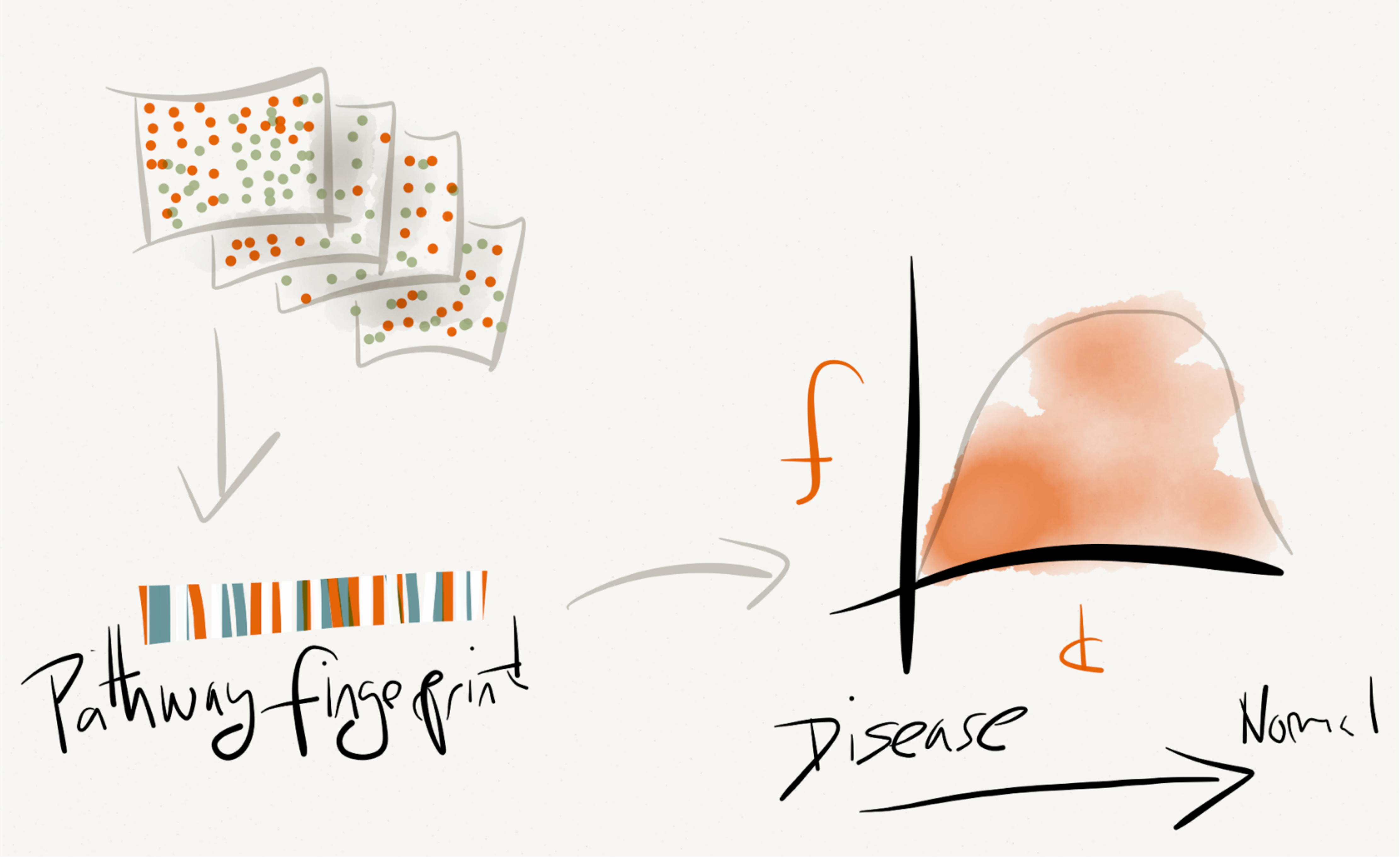



Gene expression
Pathway Expression
Compare to database
Pathway expression distribution
Gene A
Gene B
Gene C
Gene D
Gene E
Gene F
GEO
Gene Expression Omnibus
1
0
-1
Probability of Expression

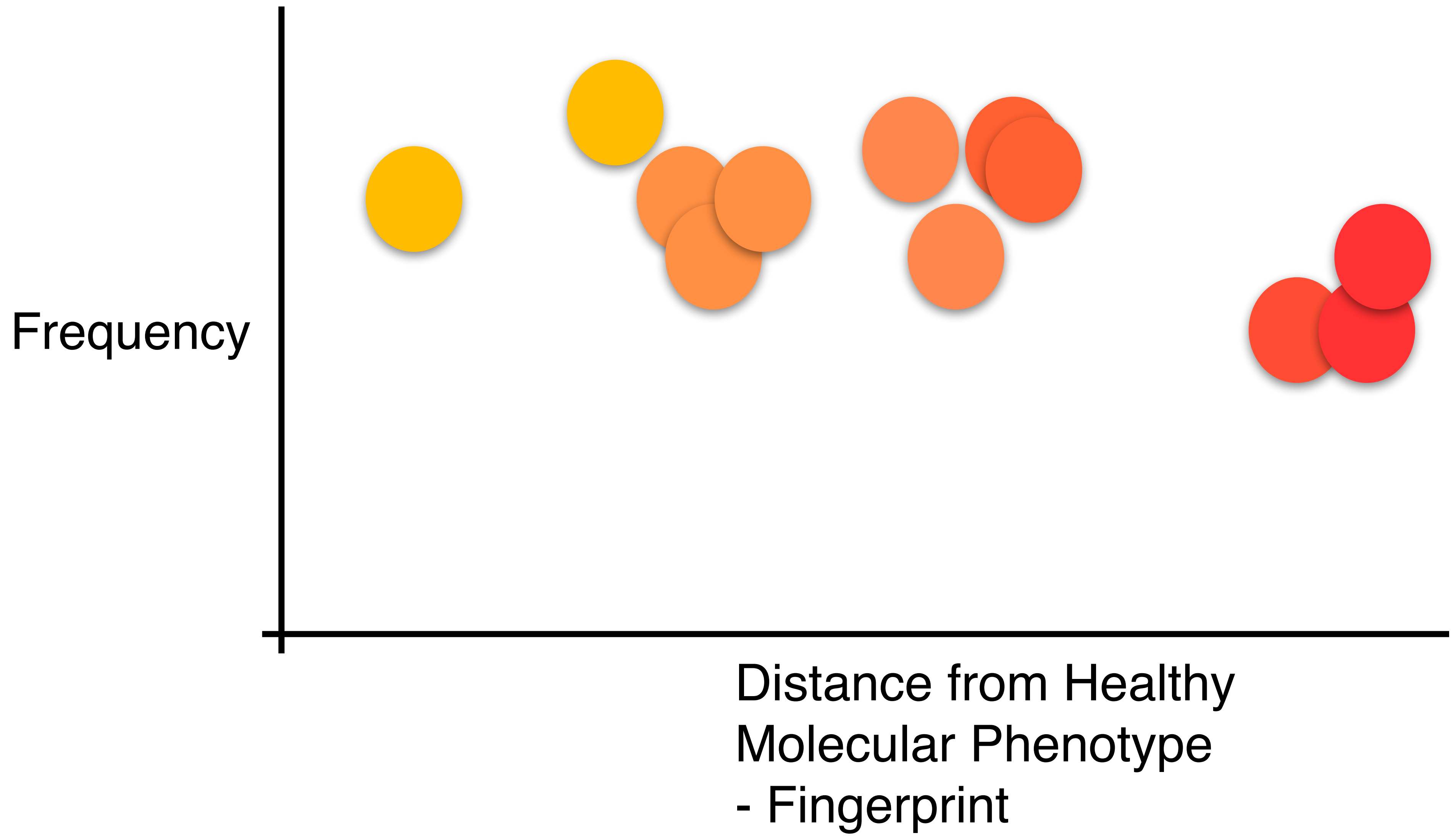


fit a two-component uniform normal mixture to the distribution of expression

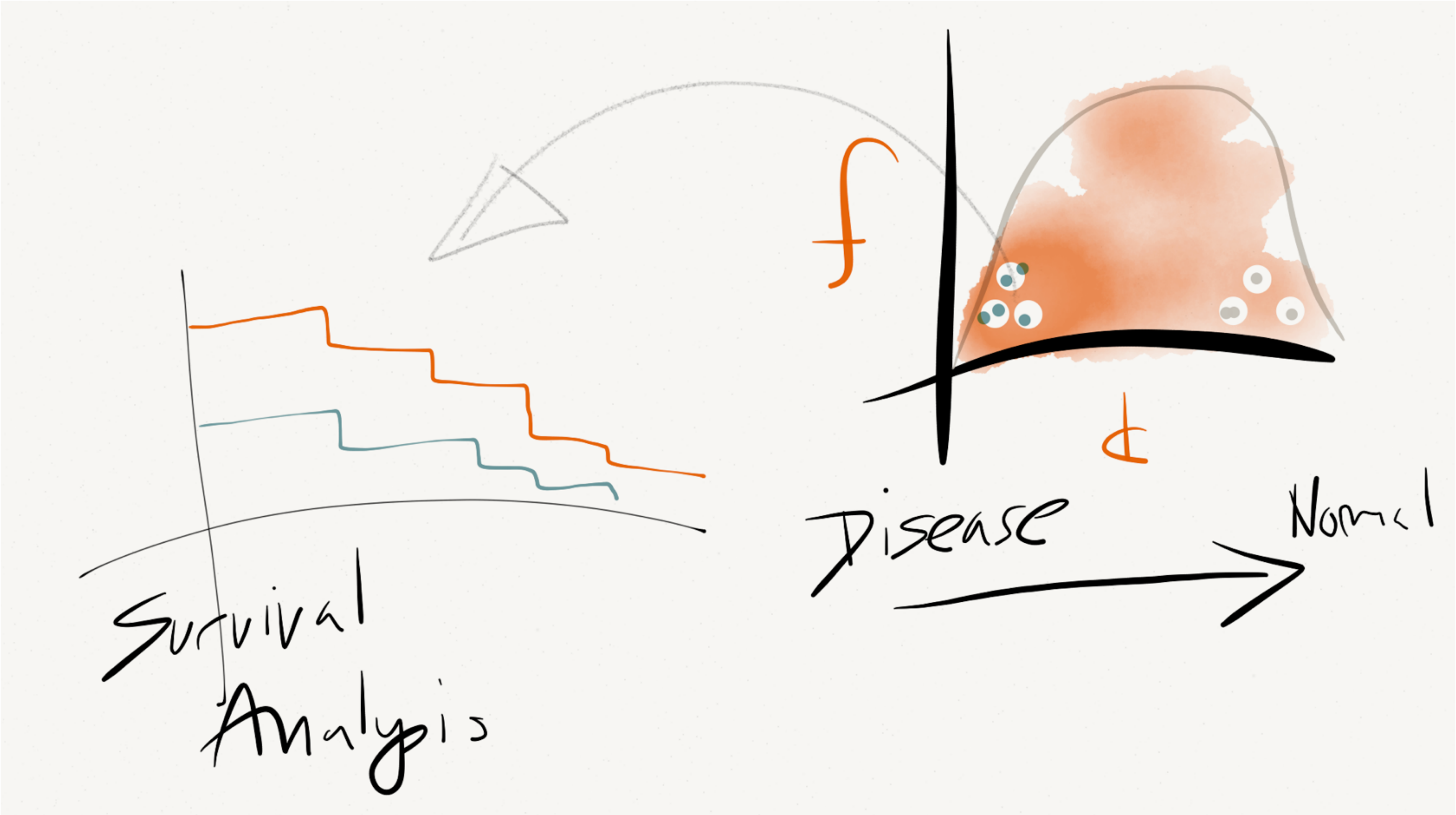
assumes that the expression scores for each gene set are normally distributed with outliers



Broad application across diseases

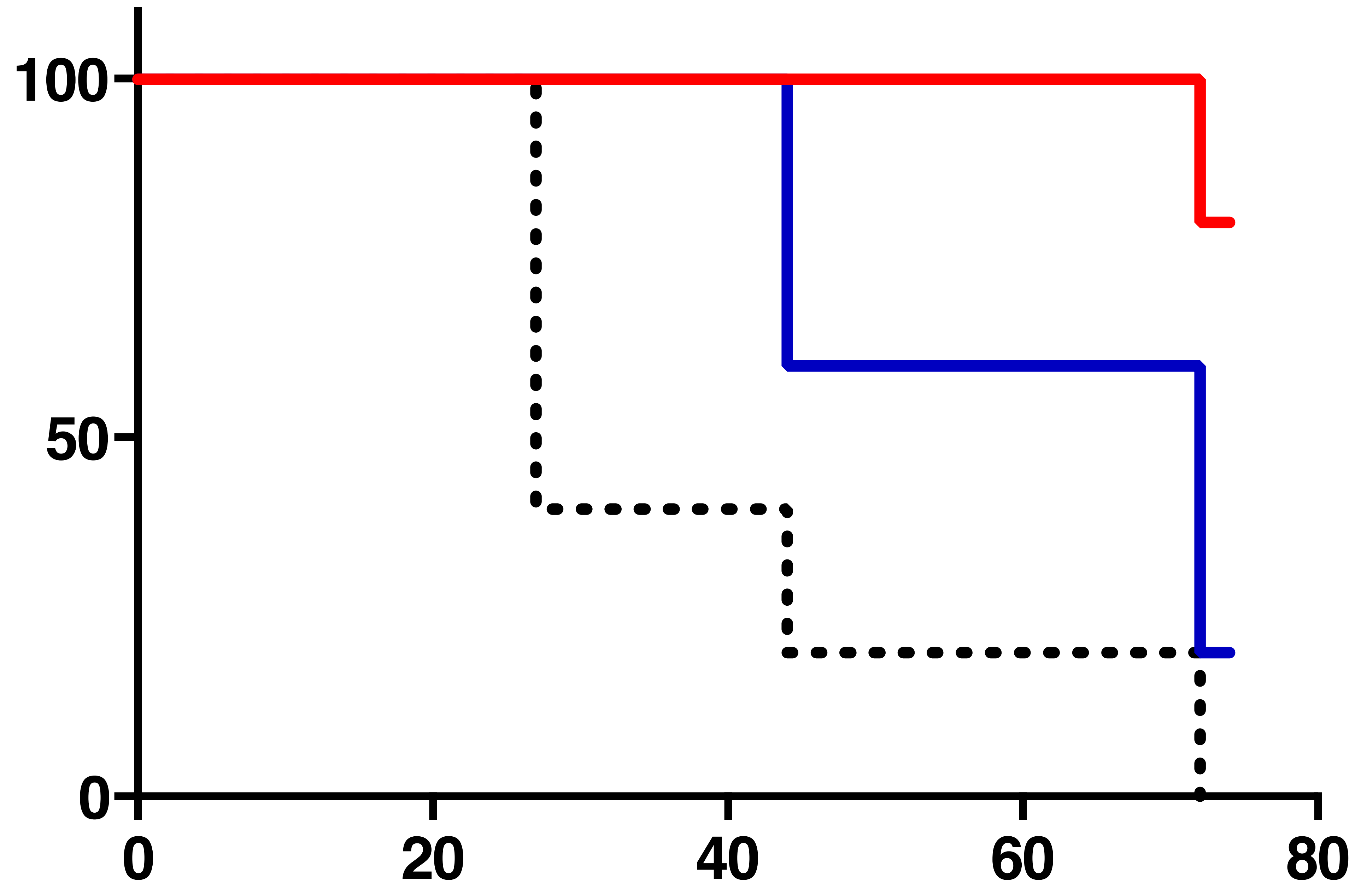


Patients and single cells as samples

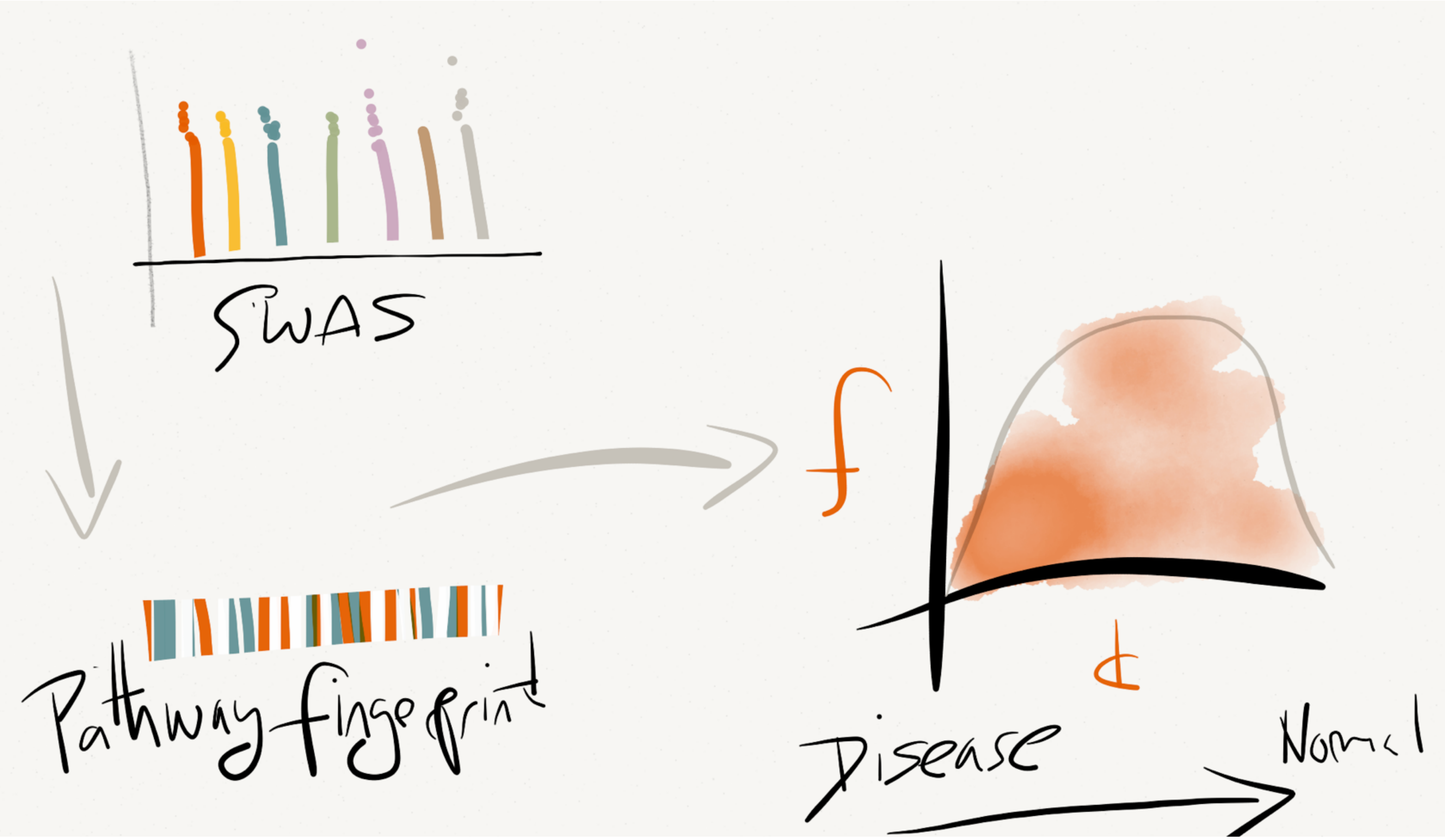


Survival outcome references

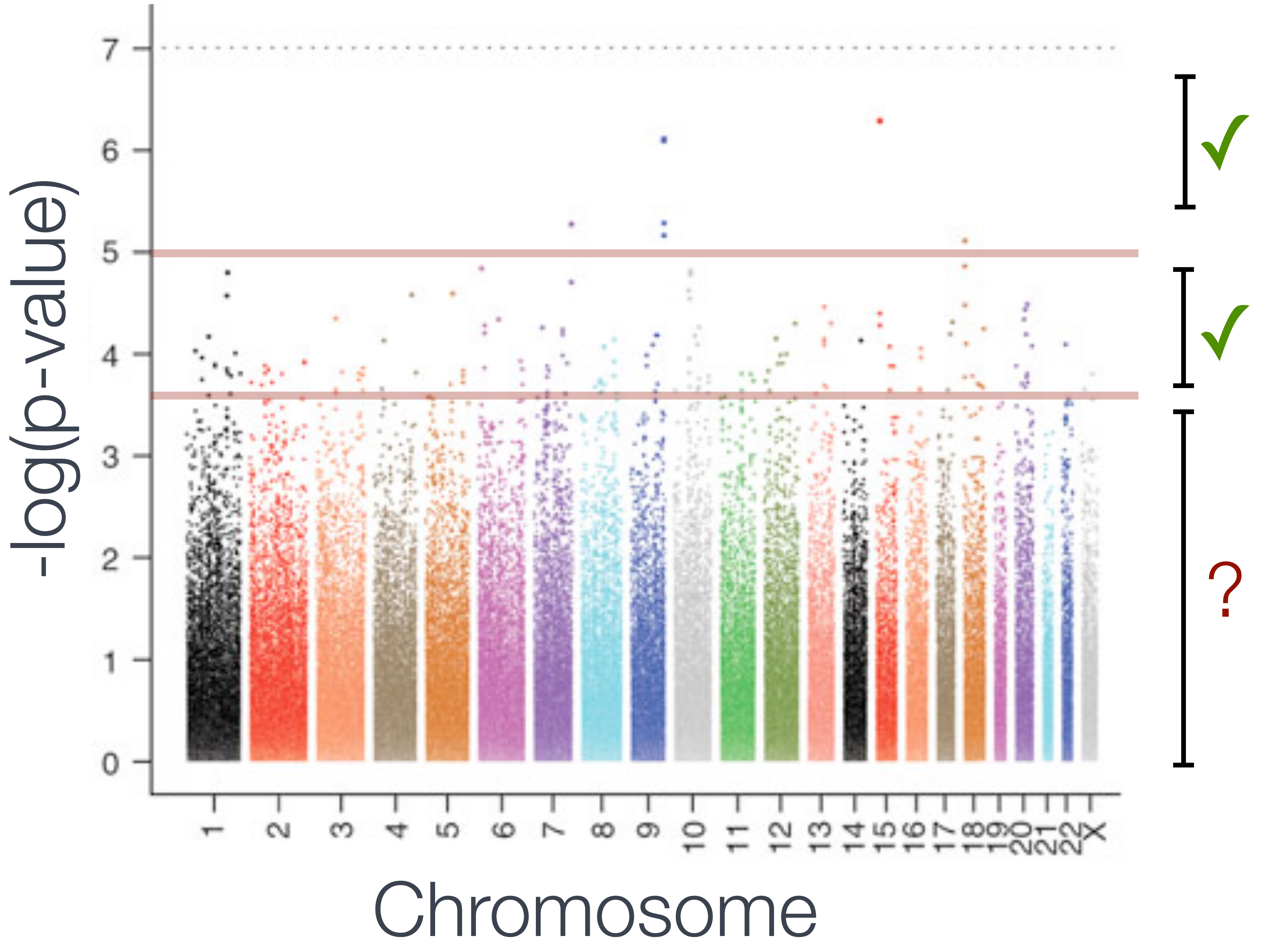
Improving Survival in Endotoxin Sepsis Model



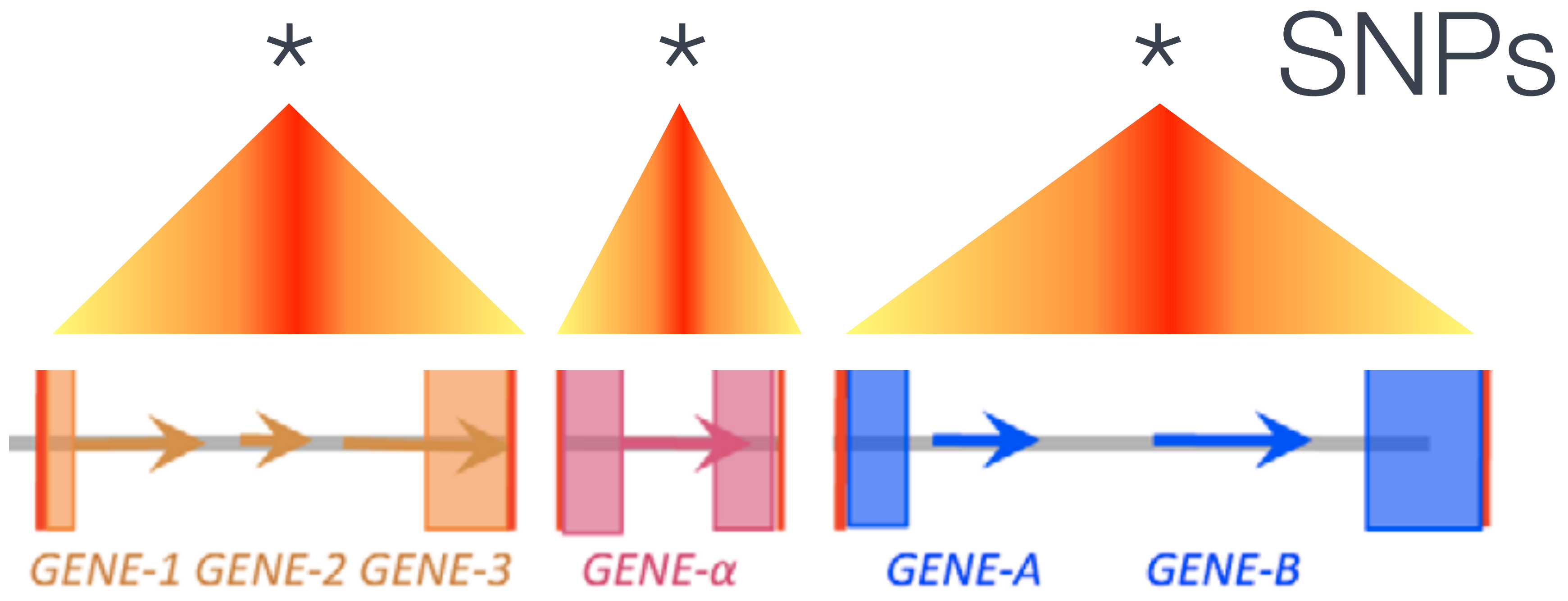
C57bl/6 female mice; LPS i.p. 23 mg/kg; Amitriptyline 10 mg/kg; n=5/group



Incorporate genetic variation



Functional approach

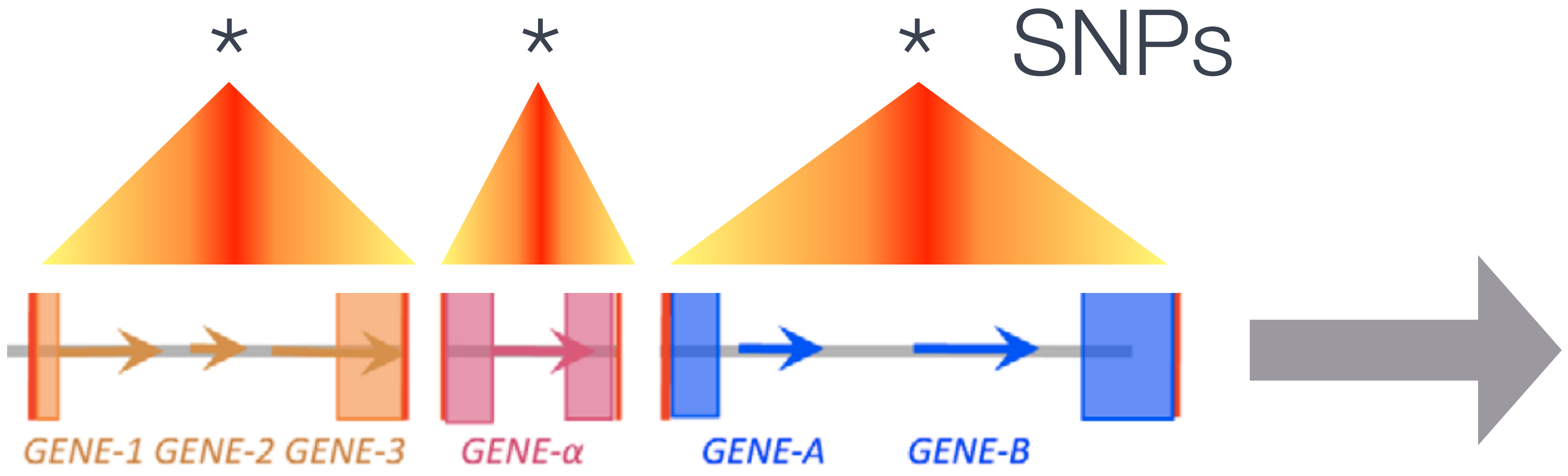


Adjust gene significance

Adapting fingerprints for GWA

Lung cancer survival study. Data courtesy of David Christiani and Mike Wang

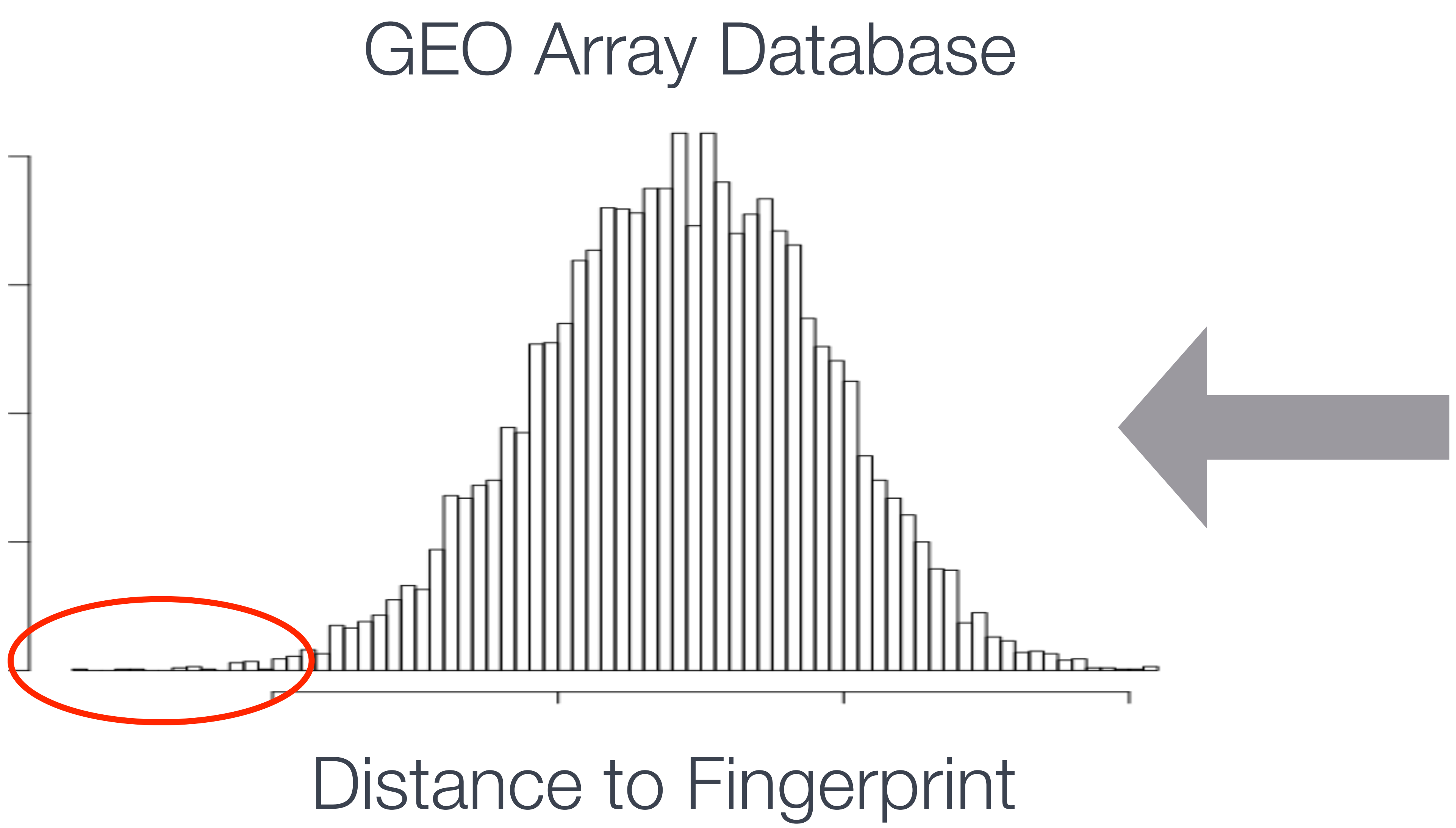
$p \leq 0.02$



- +1 CHRNA1 FI network
- +1 Cytokines, inflammatory response (Wiki)
- +1 IL-5 down regulated genes (Netpath)
- +1 MLH1 FI network
- +1 NAP1L1 FI network
- +1 EPRS FI network
- +1 VHL FI network
- +1 Proteasome degradation (Wiki)
- +1 RPS27A FI network

Adapting fingerprints for GWA

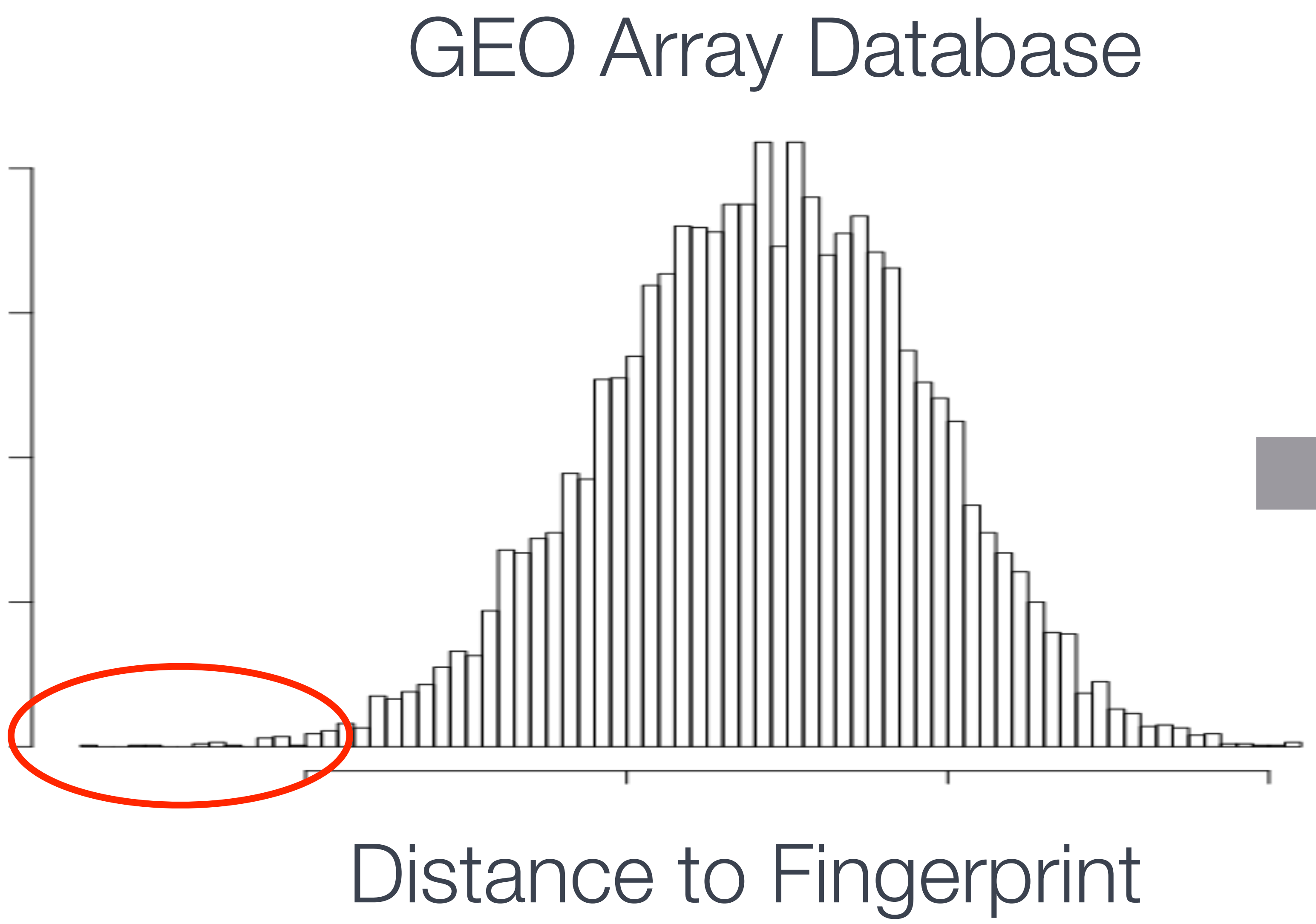
Lung cancer survival study. Data courtesy of David Christiani and Mike Wang



- +1 CHRNA1 FI network
- +1 Cytokines, inflammatory response (Wiki)
- +1 IL-5 down regulated genes (Netpath)
- +1 MLH1 FI network
- +1 NAP1L1 FI network
- +1 EPRS FI network
- +1 VHL FI network
- +1 Proteasome degradation (Wiki)
- +1 RPS27A FI network

Search for expression arrays matching GWA

Lung cancer survival study. Data courtesy of David Christiani and Mike Wang



Annotation	GEO ID
lung adenocarcinoma	GSE17475
early stage NSCLC	GSE19188
human cancer cell lines	GSE5823
lymph node biopsies	GSE6338
human pre-frontal cortex, liver, and colon tissues and colon tumors	GSE13471
...	

Search for expression arrays matching GWA

Lung cancer survival study. Data courtesy of David Christiani and Mike Wang

- +1 CHRNA1 FI network
- +1 Cytokines, inflammatory response (Wiki)
- +1 IL-5 down regulated genes (Netpath)
- +1 MLH1 FI network
- +1 NAP1L1 FI network
- +1 EPRS FI network
- +1 VHL FI network
- +1 Proteasome degradation (Wiki)
- +1 RPS27A FI network

Annotation	GEO ID
lung adenocarcinoma	**GSE17475**
early stage NSCLC	**GSE19188**
human cancer cell lines	**GSE5823**
lymph node biopsies	**GSE6338**
human pre-frontal cortex, liver, and colon tissues and colon tumors	**GSE13471**
...	

Organize arrays by GWA fingerprint

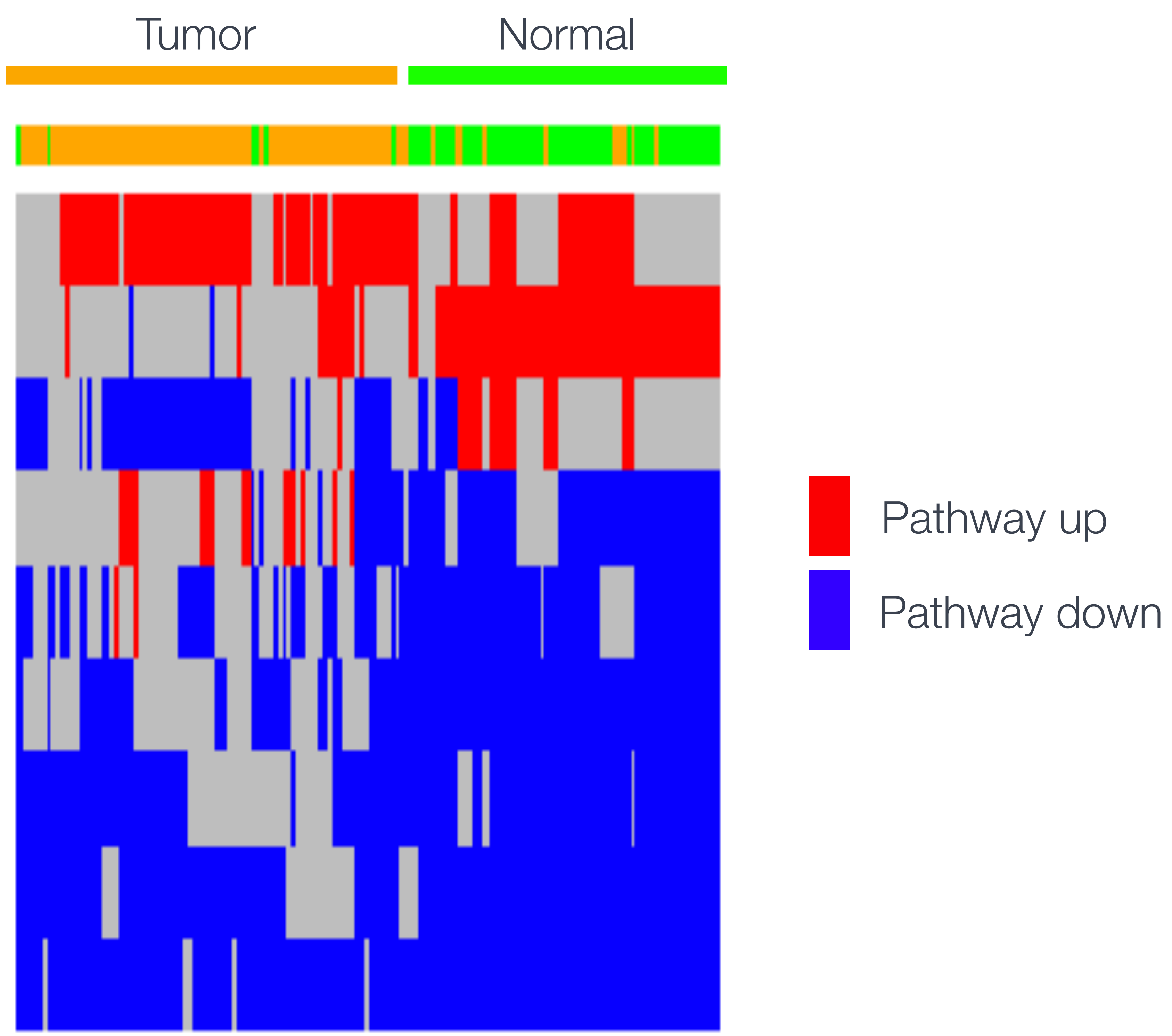
Lung cancer survival study. Data courtesy of David Christiani and Mike Wang

CHRNA1 FI network	+1
Cytokines, inflammatory response (Wiki)	+1
IL-5 down regulated genes (Netpath)	+1
MLH1 FI network	+1
NAP1L1 FI network	+1
EPRS FI network	+1
VHL FI network	+1
Proteasome degradation (Wiki)	+1
RPS27A FI network	+1

GWA Fingerprint

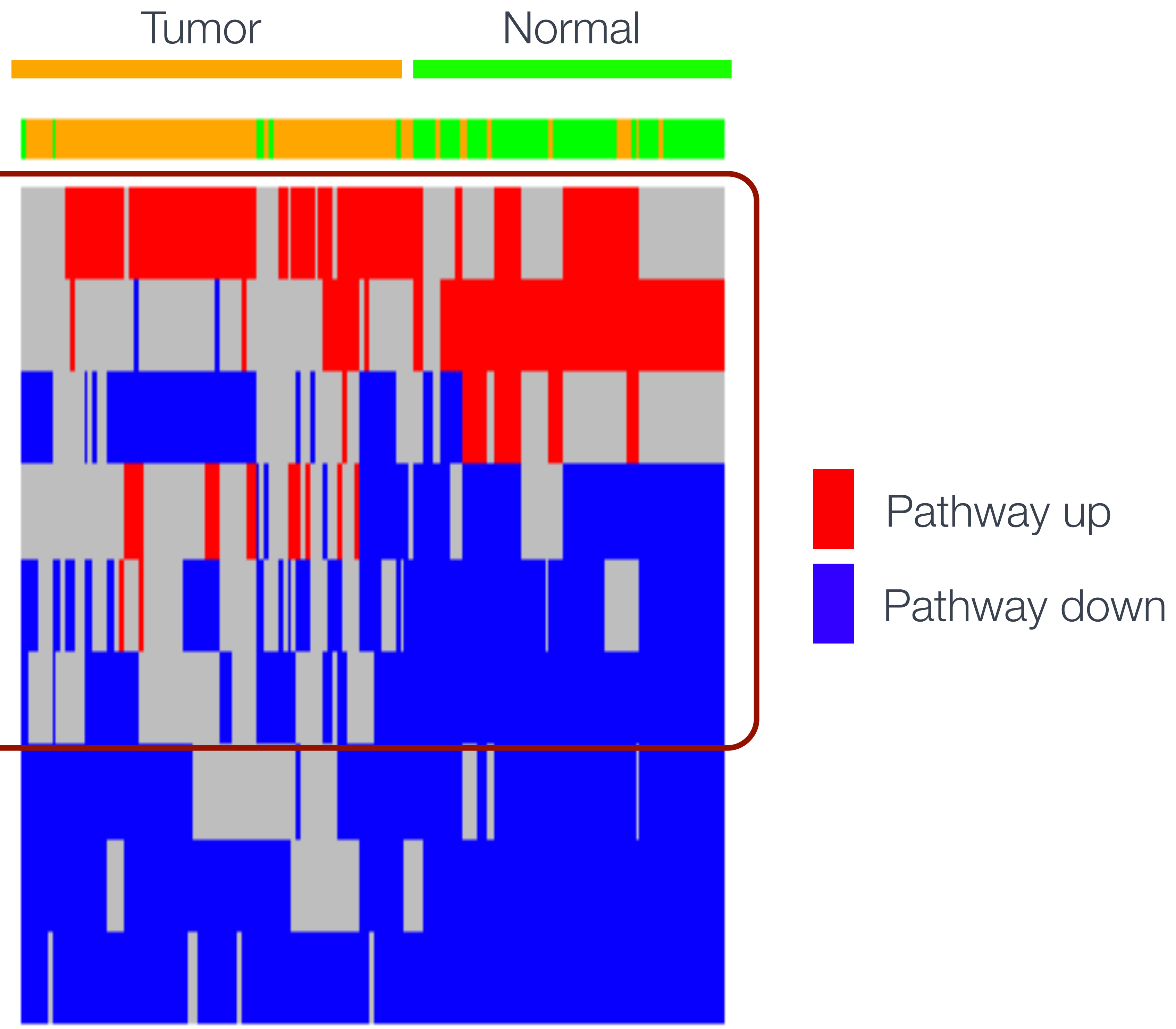
CHRNA1 FI network	+1
Cytokines, inflammatory response (Wiki)	+1
IL-5 down regulated genes (Netpath)	+1
MLH1 FI network	+1
NAP1L1 FI network	+1
EPRS FI network	+1
VHL FI network	+1
Proteasome degradation (Wiki)	+1
RPS27A FI network	+1

GWA Fingerprint



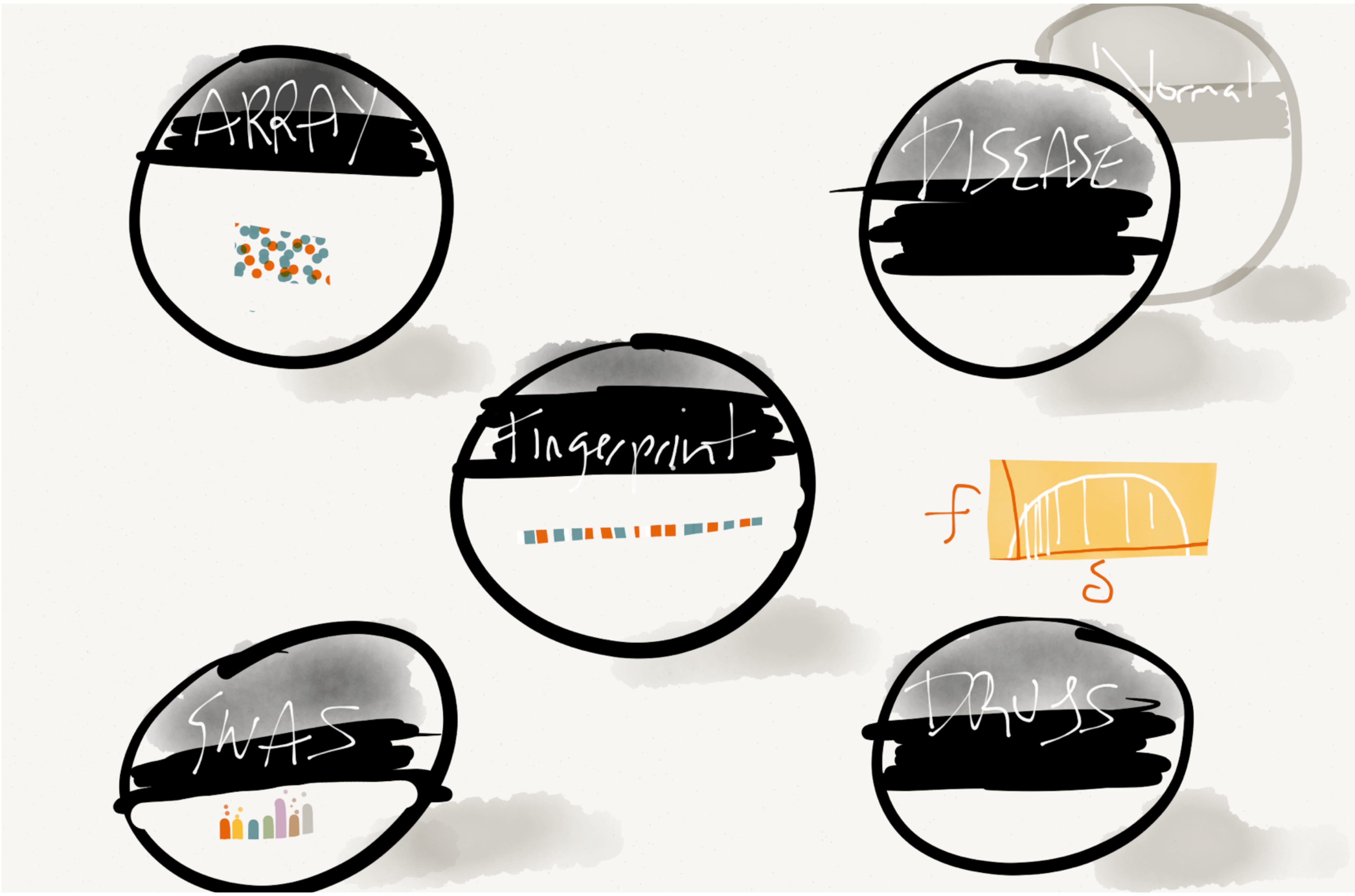
NSCLC arrays

CHRNA1 FI network	+1
Cytokines, inflammatory response (Wiki)	+1
IL-5 down regulated genes (Netpath)	+1
MLH1 FI network	+1
NAP1L1 FI network	+1
EPRS FI network	+1
VHL FI network	+1
Proteasome degradation (Wiki)	+1
RPS27A FI network	+1



GWA Fingerprint

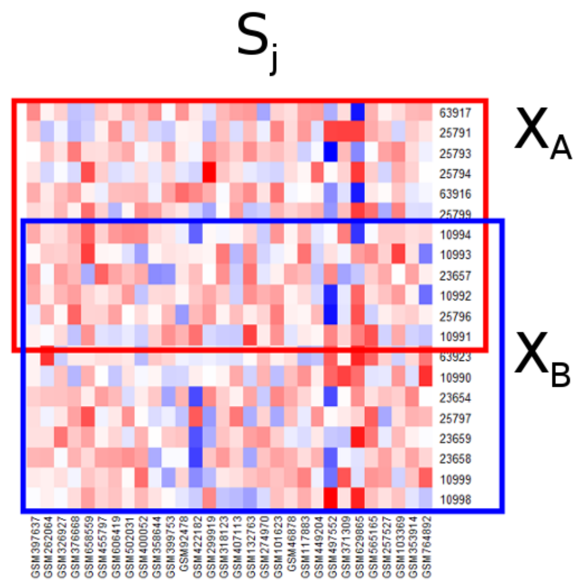
NSCLC arrays



Integrate

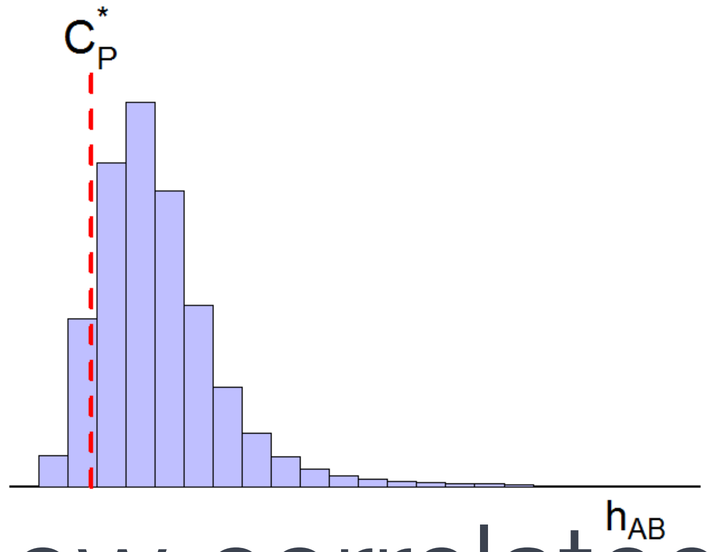
How do networks and pathways interact?

Pathway Expression Matrices



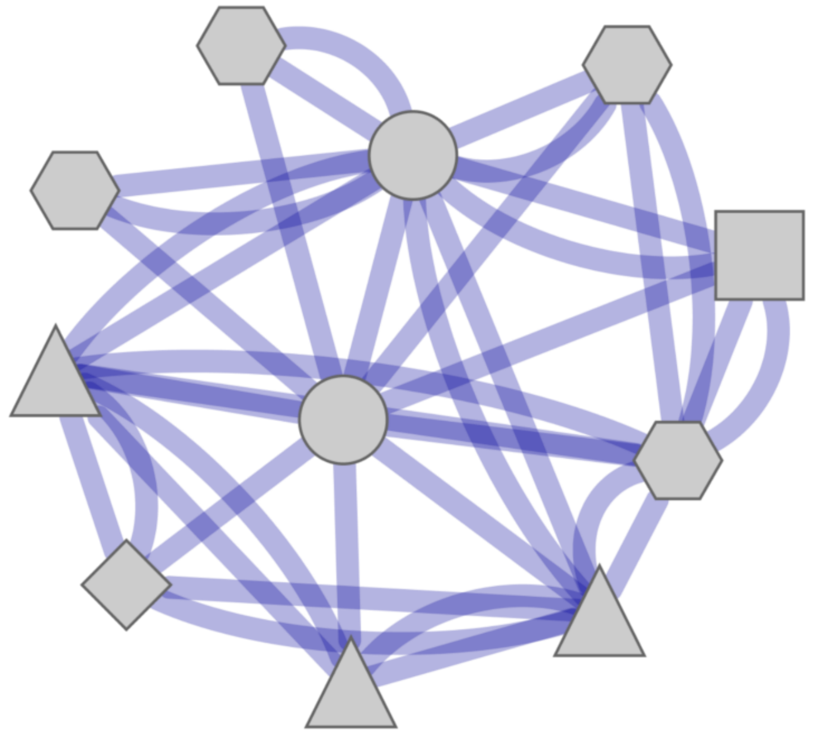
Internal Pathway Correlation 'coherence'

Pathway Agreement Threshold



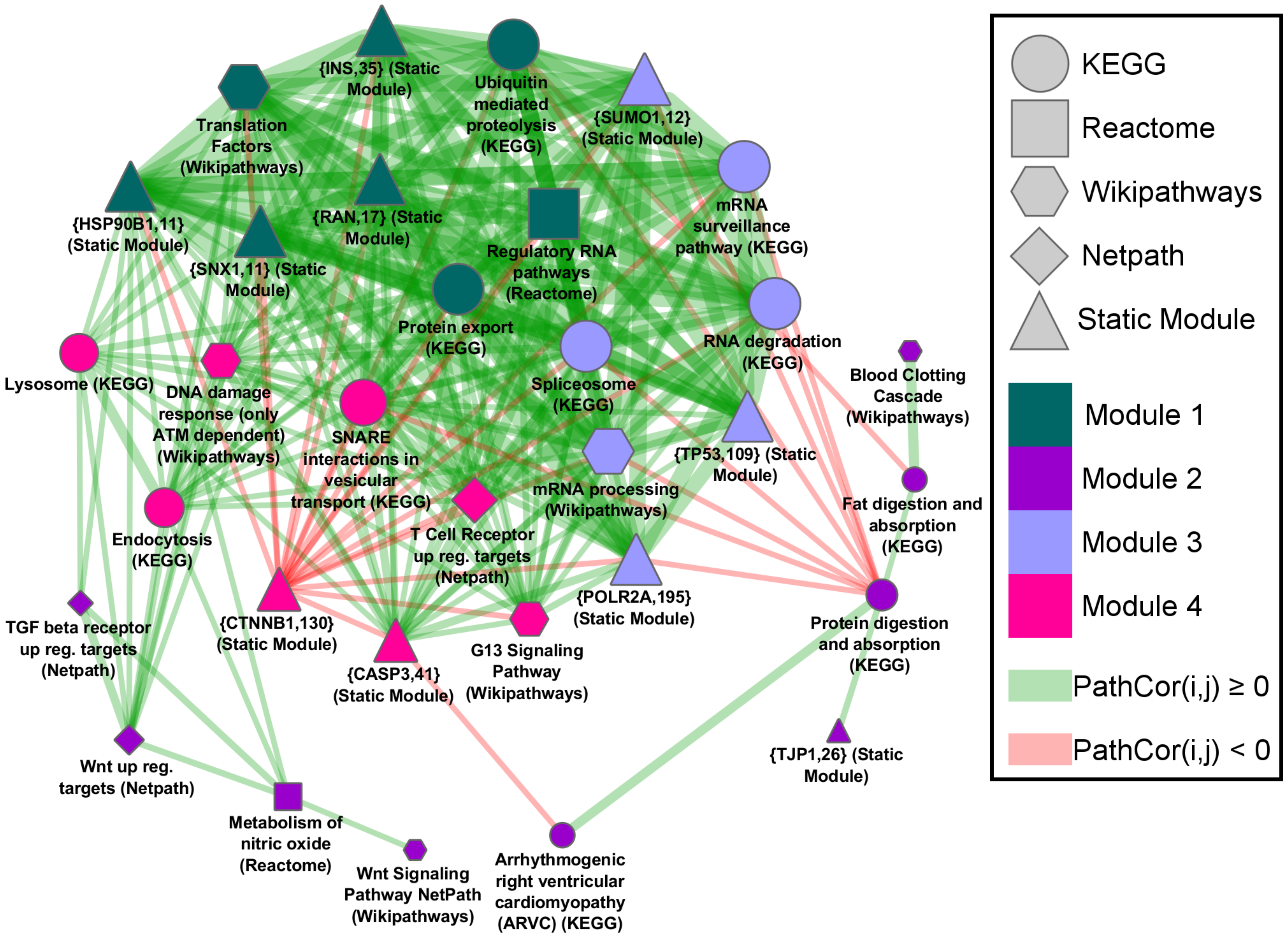
How correlated with the "pathway coherence" are external genes within another pathway?

Pathway Coexpression Network



Build network of correlations

Pathway Coexpression Network (PCxN): Clustering



Yered Pita-Juarez

Pathway Coexpression Network (PCxN): Clustering

Module 1 (165)	**Centrality ($c_e(i)$)**
{*SNX1*,11} (Static Module)	0.9783
{*RAN*,17} (Static Module)	0.9733
Protein export (KEGG)	0.9717

Module 2 (175)	**Centrality ($c_e(i)$)**
Protein digestion and absorption (KEGG)	0.4039
Wnt up reg. targets (Netpath)	0.3517
Metabolism of nitric oxide (Reactome)	0.2888

Module 3 (126)	**Centrality ($c_e(i)$)**
{*SUMO1*,12} (Static Module)	1.0000
mRNA surveillance pathway (KEGG)	0.9863
RNA degradation (KEGG)	0.9829

Module 4 (134)	**Centrality ($c_e(i)$)**
SNARE interactions in vesicular transport (KEGG)	0.8329
T Cell Receptor up reg. targets (Netpath)	0.7939
{*CASP3*,41} (Static Module)	0.7438

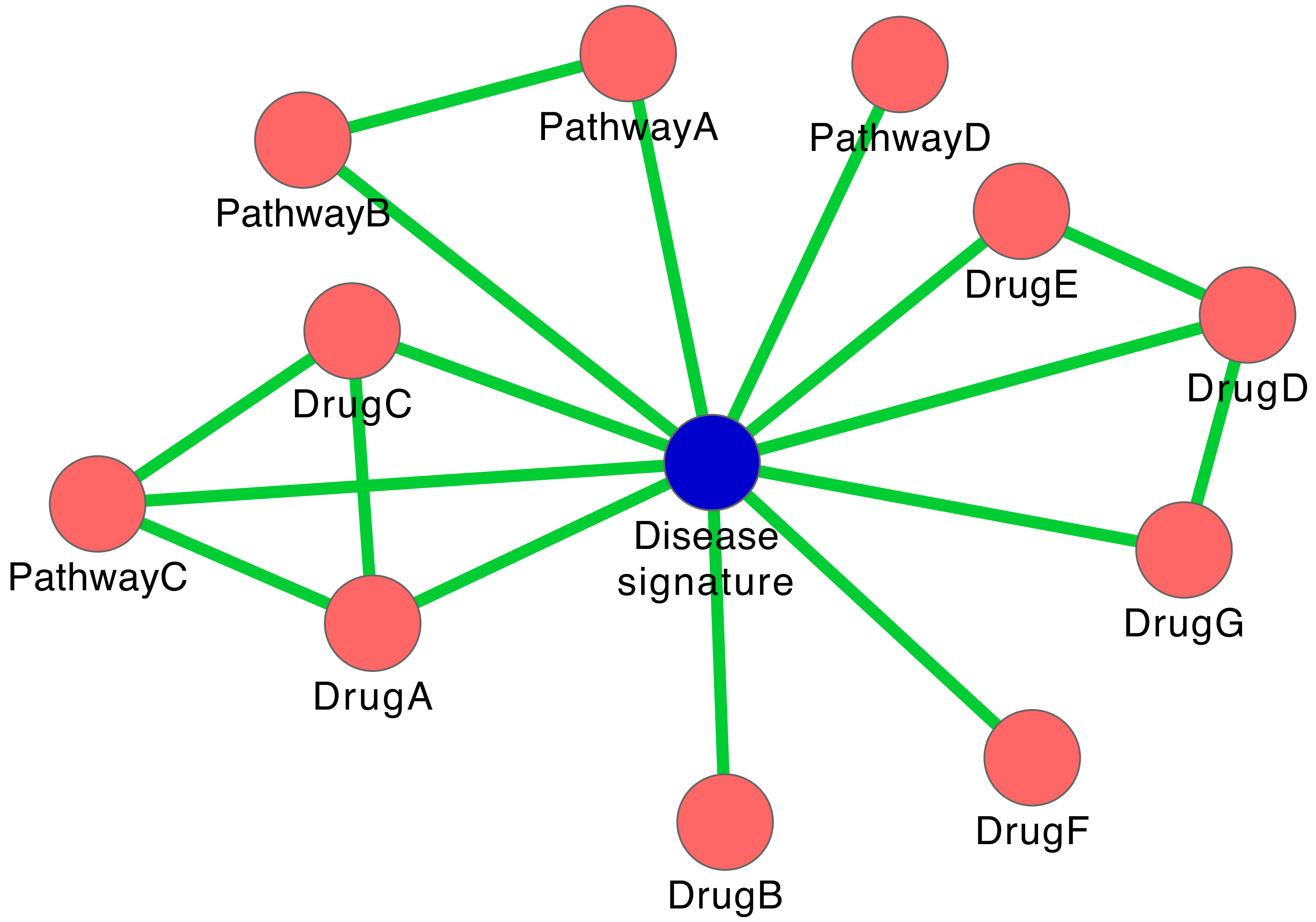
Axonogenesis

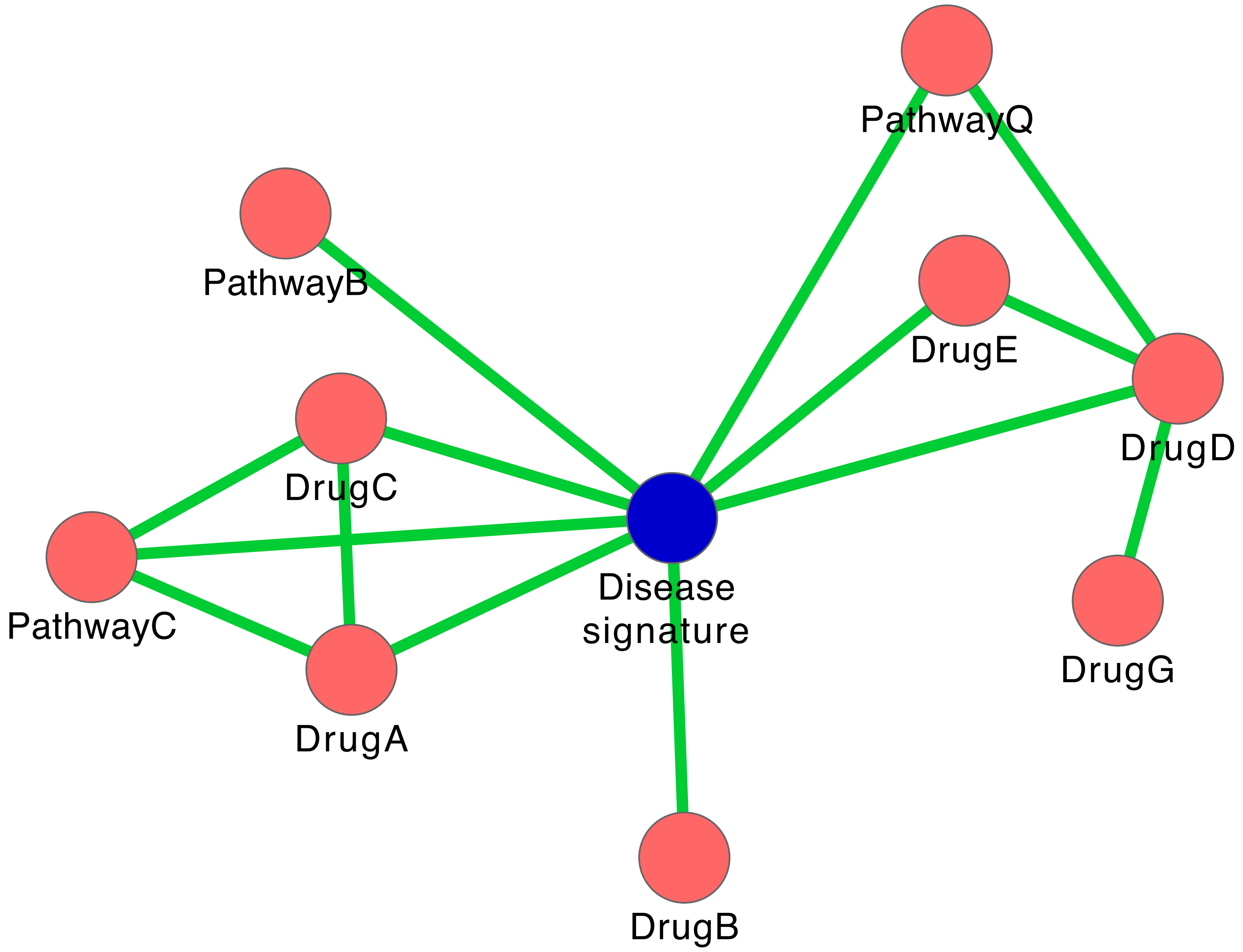
TRANSMISSION_OF_NERVE_IMPULSE	0.1676
SYNAPTIC_TRANSMISSION	0.1666
NEUROLOGICAL_SYSTEM_PROCESS	0.1501
SYSTEM_PROCESS	0.1496
CELL_CYCLE_GO_0007049	-0.1428
CELL_CYCLE_CHECKPOINT_GO_0000075	-0.1426
M_PHASE_OF_MITOTIC_CELL_CYCLE	-0.1419
DNA_REPLICATION	-0.1406
DNA_METABOLIC_PROCESS	-0.1403
MONOVALENT_INORGANIC_CATION_TRANSPORT	0.1383
METAL_ION_TRANSPORT	0.1382
M_PHASE	-0.1376
MITOSIS	-0.1368
ION_TRANSPORT	0.1357
CELL_CELL_SIGNALING	0.1349

Drugs, chemicals

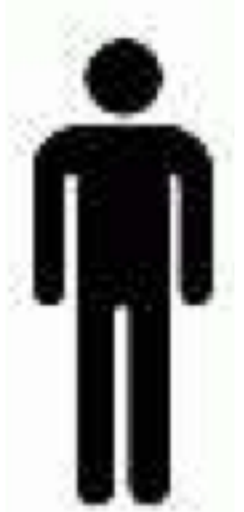
- **Comparative Toxicogenomics Database**
  - 2452 chemical genesets from ChemIDplus, DrugBank
- **PharmGKB**
  - 77 drug associated gene sets
- **Connectivity Map**
  - 6100 up regulated gene signatures (1% most up regulated genes)
  - 6100 down regulated gene signatures (1% most down regulated genes)

Network-based drug discovery pipeline



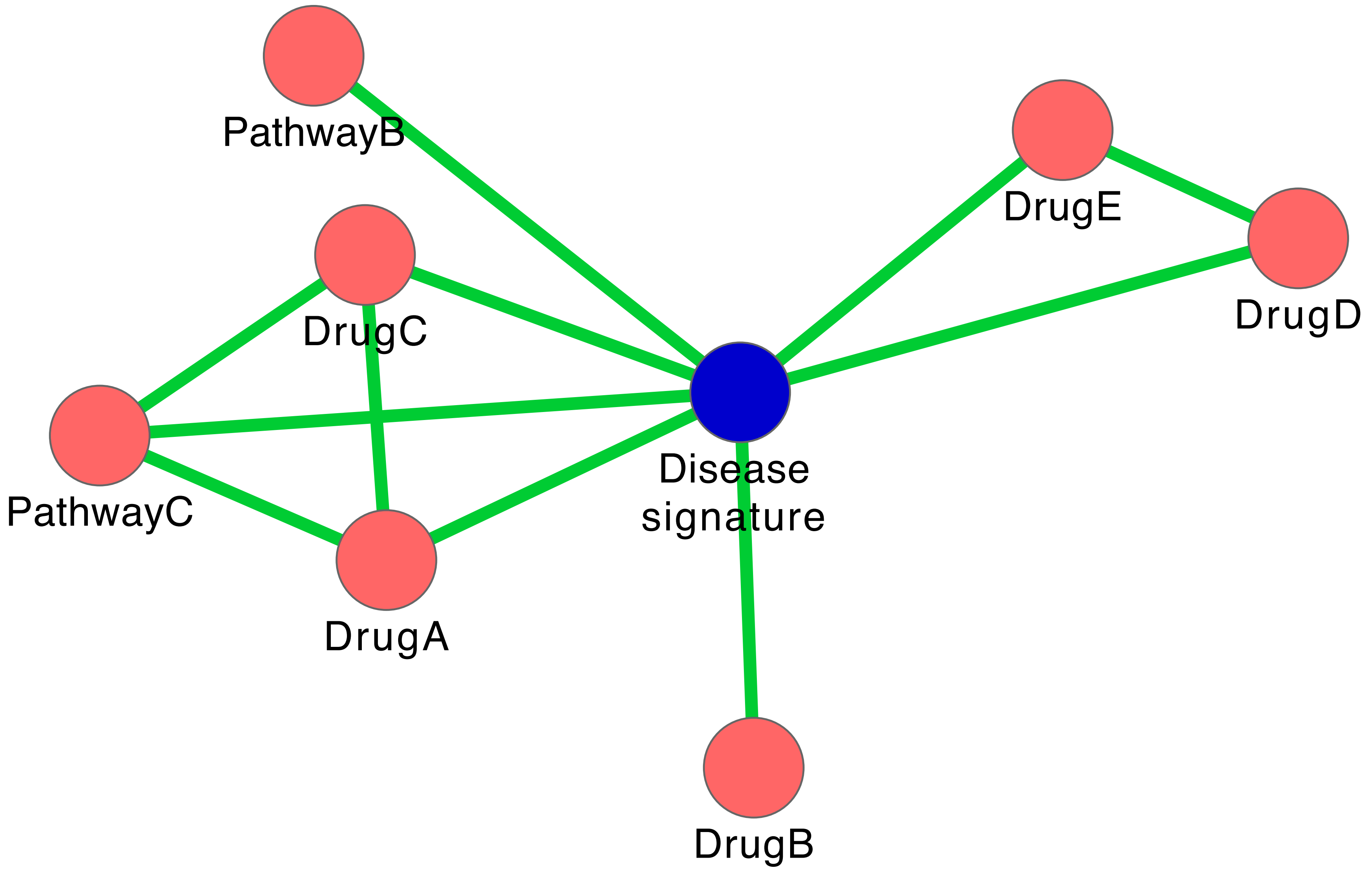


Network based on human gene expression correlation



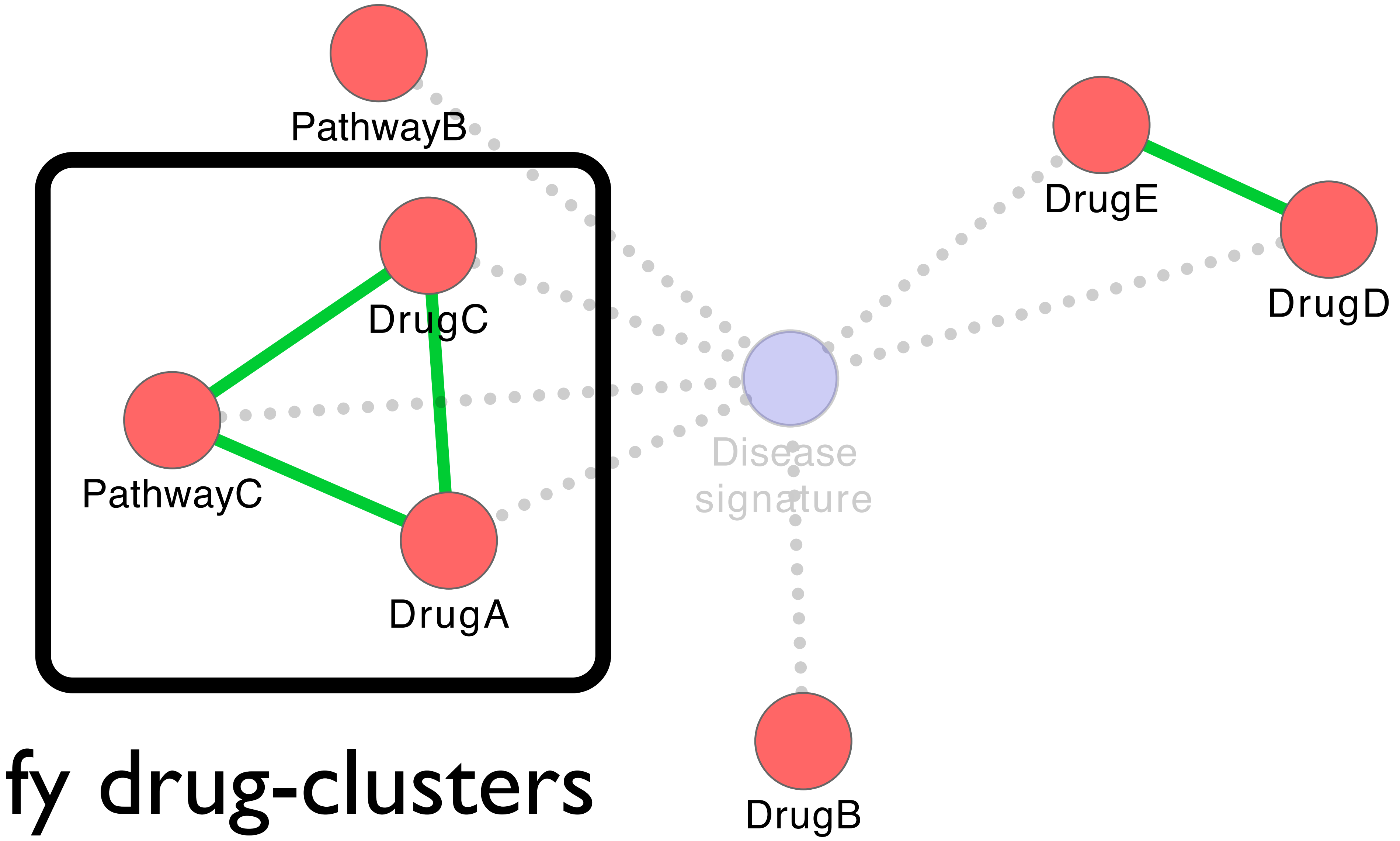
Network based on mouse gene expression correlation

Common network



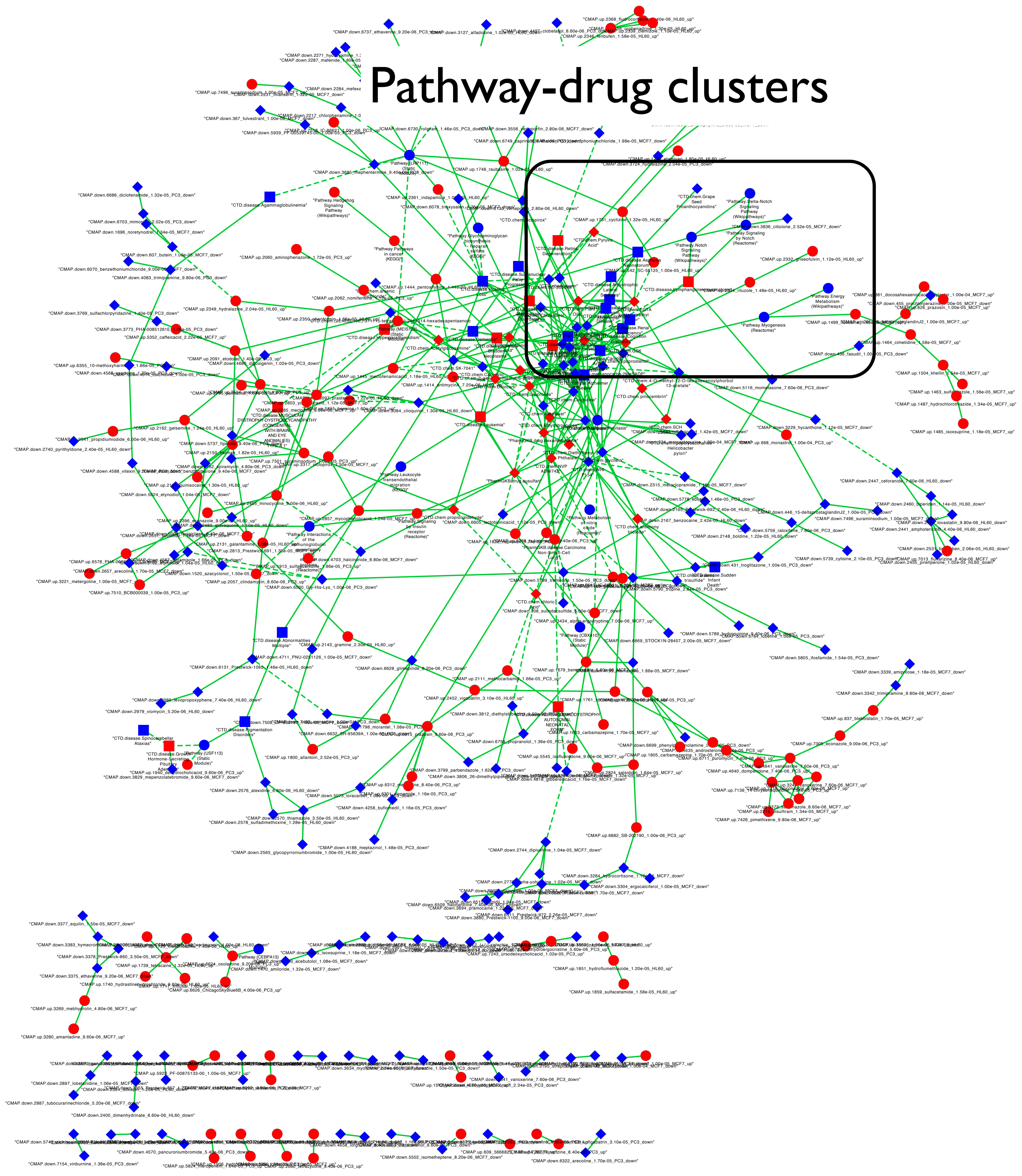
Define the consistent functional interactions

Topology Analysis



Identify drug-clusters

AD interaction network



🔵 Positively correlated to AD

🔴 Negative correlation to AD

— Positive correlation between nodes

- - - Negative correlation between nodes

◇ Chemical/Drug signature

□ Disease signature

○ Pathway signature

Top 100 Genetic candidates correlated pathways

Pathway.{SPTAN1,10} **Neuro-cyoskeletal**	2.22E-16
Pathway.{NAP1L1,23} **Nucleosome assembly**	2.22E-16
Pathway.{SIX3,11} **neurodevelopment**	2.22E-16
Pathway.{ETS1,13}	2.22E-16
Pathway.Cyanoamino acid metabolism (KEGG)	6.37E-12
Pathway.{HMGB1,14}	6.45E-12
Pathway.Acetylcholine Synthesis (Wikipathways)	1.36E-10
PharmGKB.drug.corticosteroids	1.86E-10
Pathway.Collecting duct acid secretion (KEGG)	1.76E-09
PharmGKB.drug.rosuvastatin	1.94E-09
PharmGKB.disease.Thromboembolism	1.76E-08
Pathway.{AKT1,48} (Static Module)	3.43E-08
Pathway.Endocytosis (KEGG)	1.73E-07
Pathway.{PLXNB1,15} **microtubule dynamics**	7.45E-07
Pathway.{AP1G1,16} **Adaptor Protein**	2.81E-06
Pathway.{RB1,11} **down regulated in Pre-ALS**	6.67E-06
Pathway.Carbohydrate digestion and absorption (KEGG)	2.19E-05
Pathway.{HTT,61} Huntington network is co-expressed	2.42E-05
Pathway.Glycerolipid metabolism (KEGG)	3.38E-05
Pathway.Biogenic Amine Synthesis (Wikipathways)	6.63E-05
Pathway.Glycosphingolipid biosynthesis - lacto and	6.93E-05

280 000 experiments
multiple platforms

Co-expressed modules

Roadblocks & Challenges

Provenance

Reproducibility

Data Sharing paradigms

Pre-competitive approaches

Create a cloud commons

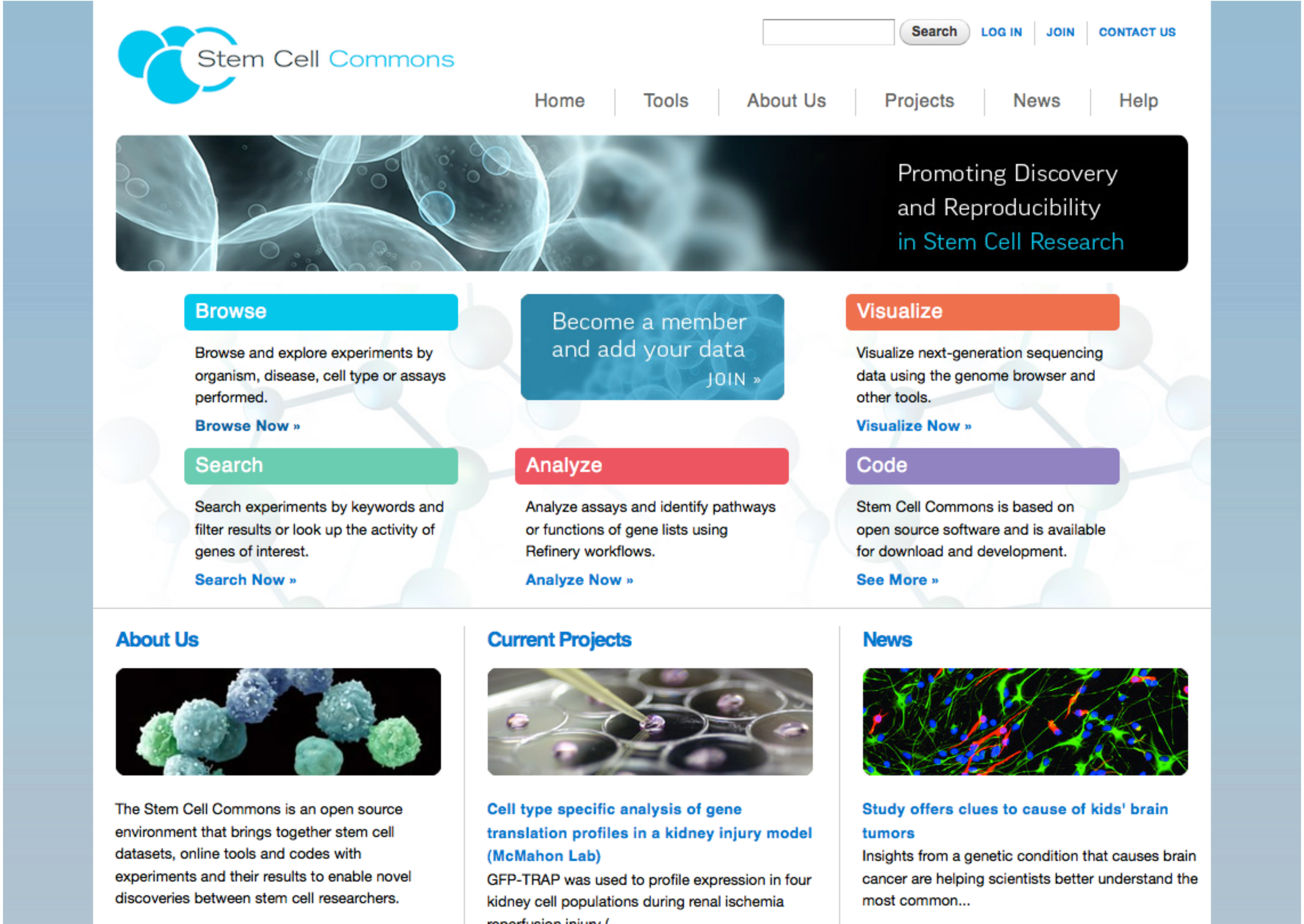
Nature Comment July 2015
Stein et al

Cloudy with a chance of genomes

- “every major genomic data set be uploaded into the most popular academic and commercial clouds available, and to pay for the long-term storage of the data in the clouds”

Provenance

- **Sage BioNetworks**
  - Nano Publication of Dataset/Workflow - DOI
- **HSCI Stem Cell Commons**
  - Publication of dataset to Data Publication
  - Workflows next to data



Promoting Discovery
and Reproducibility
in Stem Cell Research

Browse

Browse and explore experiments by organism, disease, cell type or assays performed.

Browse Now »

Become a member and add your data

JOIN »

Visualize

Visualize next-generation sequencing data using the genome browser and other tools.

Visualize Now »

Search

Search experiments by keywords and filter results or look up the activity of genes of interest.

Search Now »

Analyze

Analyze assays and identify pathways or functions of gene lists using Refinery workflows.

Analyze Now »

Code

Stem Cell Commons is based on open source software and is available for download and development.

See More »

About Us

The Stem Cell Commons is an open source environment that brings together stem cell datasets, online tools and codes with experiments and their results to enable novel discoveries between stem cell researchers.

Current Projects

Cell type specific analysis of gene translation profiles in a kidney injury model (McMahon Lab)

GFP-TRAP was used to profile expression in four kidney cell populations during renal ischemia reperfusion injury (

News

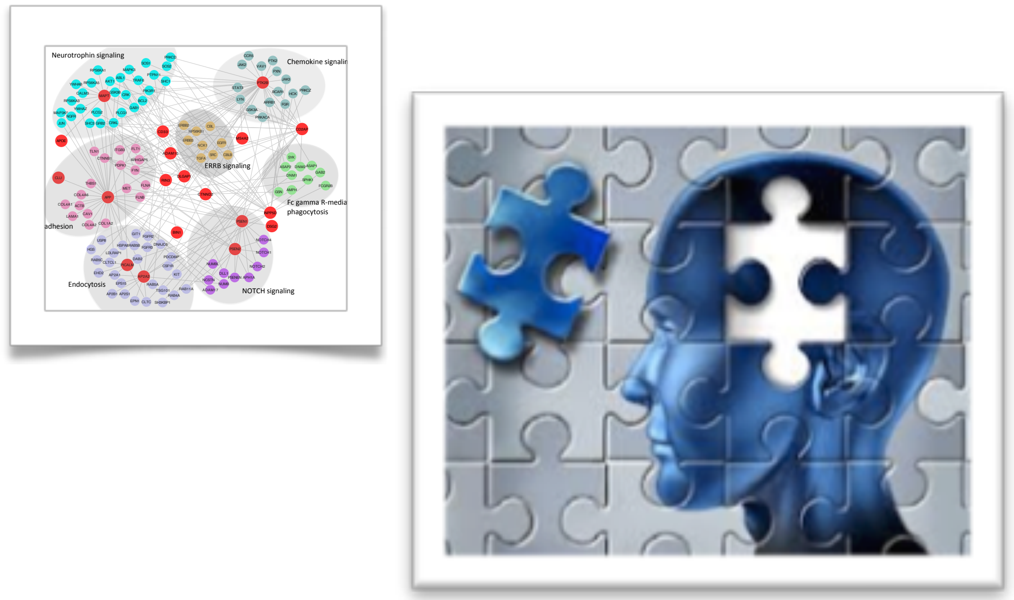
Study offers clues to cause of kids' brain tumors

Insights from a genetic condition that causes brain cancer are helping scientists better understand the most common...

SITraN
Sheffield Institute for Translational Neuroscience

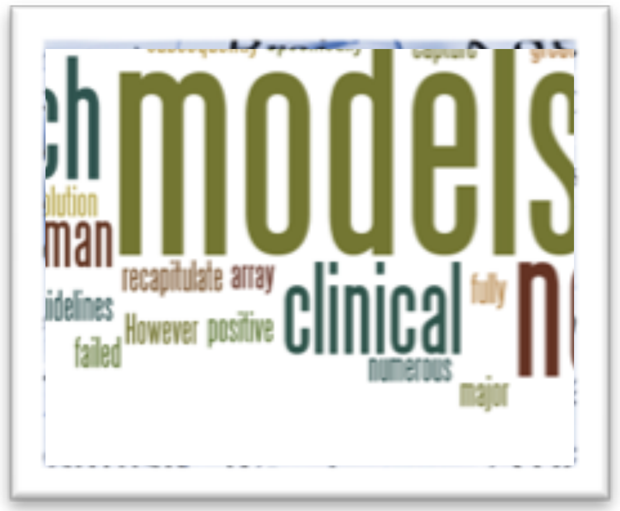
Centre for Genome Translation

Diseases



Diseases are described with phenotypes

Phenotypes

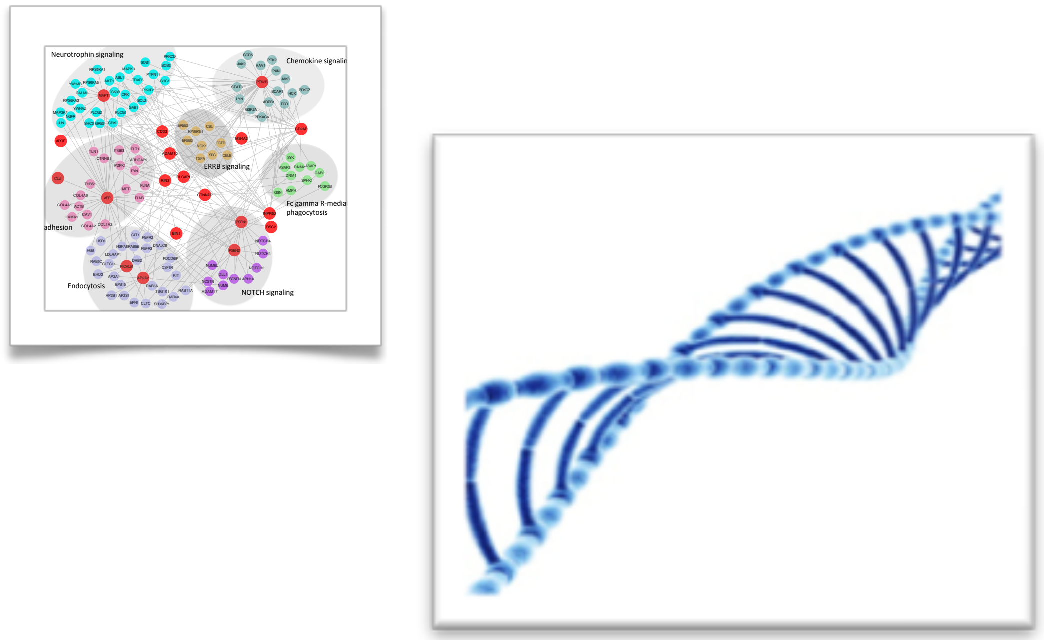


Diseases are associated with genetic variation

Cause
Target
Prediction
Repurpose

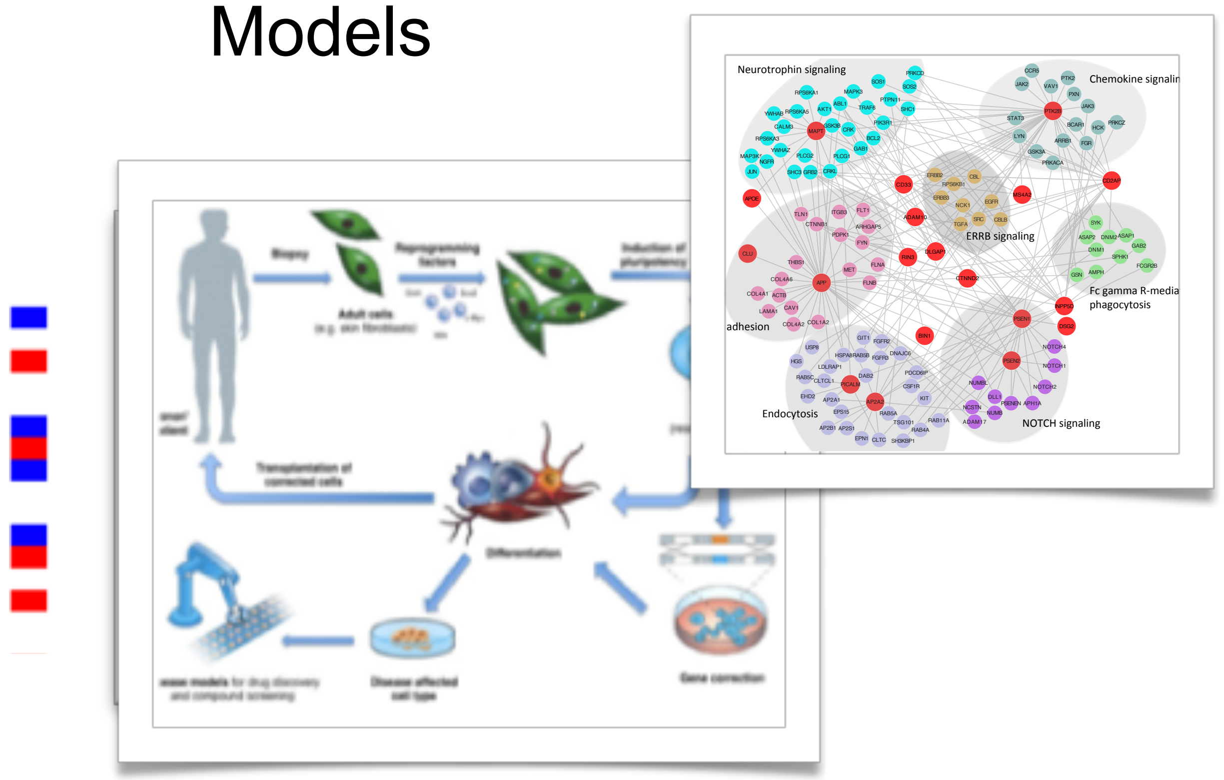
Phenotype comparisons suggest disease models

Genes



Models suggest candidate genes

Models



Physician scientists

Genomics
Prioritisation

Quantitative Functional Readouts

Deep Phenotyping

Biorepositories

Access to anonymised EHR

Animal models and validation

Computational Biology

Machine learning

Disease
Network models

Translation Commons

Target prioritisation

In silico validation

Context engine

Infrastructure - Models -**Target Prioritisation** - Repurposing

Take homes

- **Genomics is a growing driver** in target provision and stratification
- **Parochial knowledge** is king
- **Scaling is becoming overwhelming** and bottlenecks validation
- Computational infrastructure needed to address **translation between data driven and domain knowledge** approaches
- **Systems and network approaches** are likely to **yield** a growing cadre of **new** ranches of **targets**
- **Target prioritisation** goes hand in hand with **systematic functional quantitation**

Futurewatch

- • The data patient - realtime trials
- ○ Predicted vs treated patients - hospital is the home
- ○ Trained systems (Watson)
- ○ Genetics and environment
- ■ Facespace

The University Of Sheffield.

HSCI
HARVARD STEM CELL INSTITUTE®

Centre for Stem Cell Bioinformatics

- Win Hide
- **Shannan Ho Sui**
- **Yered Hammurabi Pita-Juarez**
- **Oliver Hofmann**

Centre for Genome Translation

- Win Hide
- **Gabriel Altschüler**
- David Jones
- WenBin Wei
- Sandeep Amberkar
- Claire Green

We are HIRING

hidelab.wordpress.com