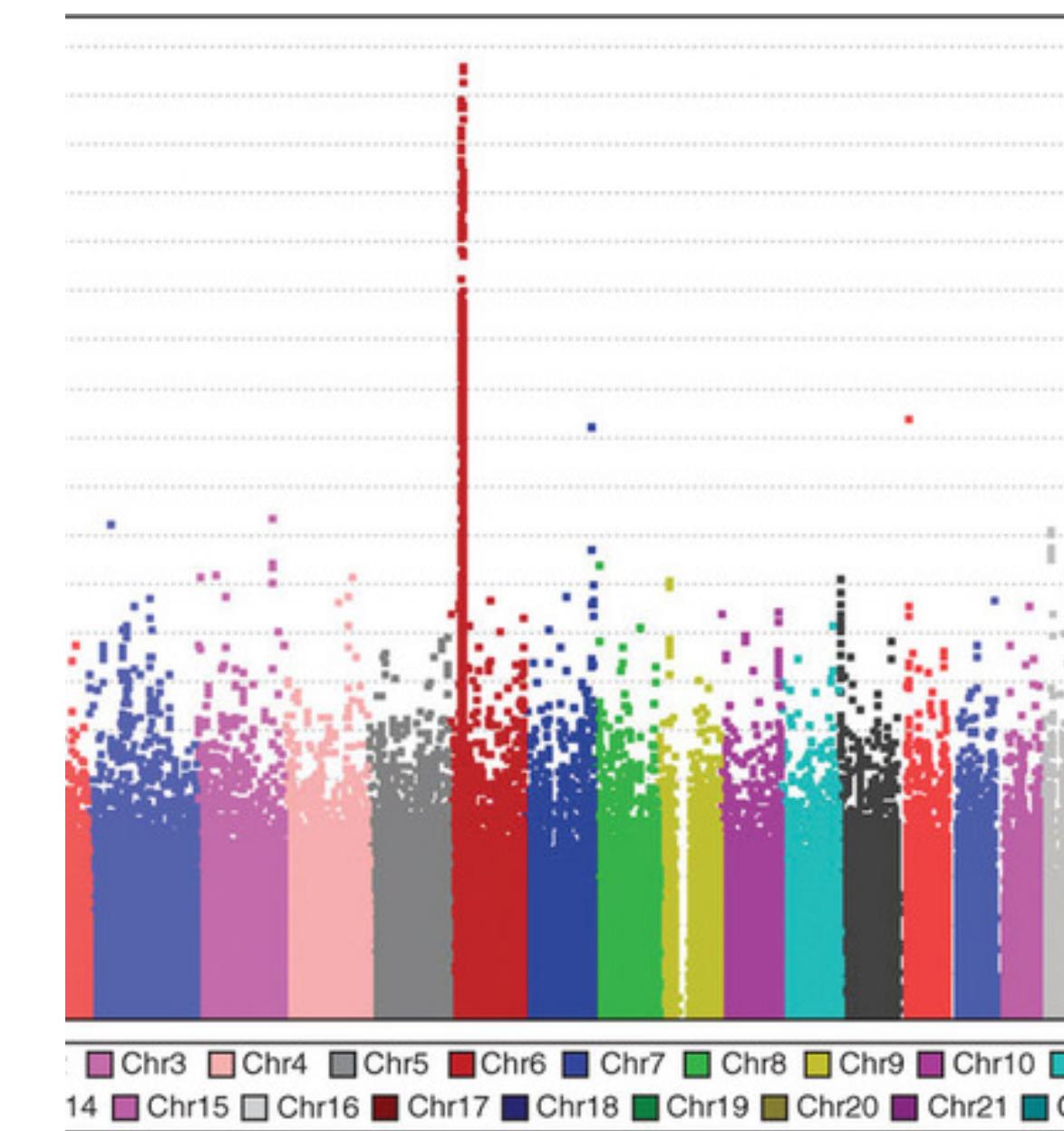


Molecular signatures and variation

Genome variants



Predicting health

- ▶ source of variation in the *estimates of average risks*
- ▶ *Sampling Error*
- ▶ SNPs vary between each platform
- ▶ Tiny fraction of all variants
- ▶ Role of most genes currently unknown in disease

Its a large cohort problem

Schizophrenia - nothing with 3000 cases

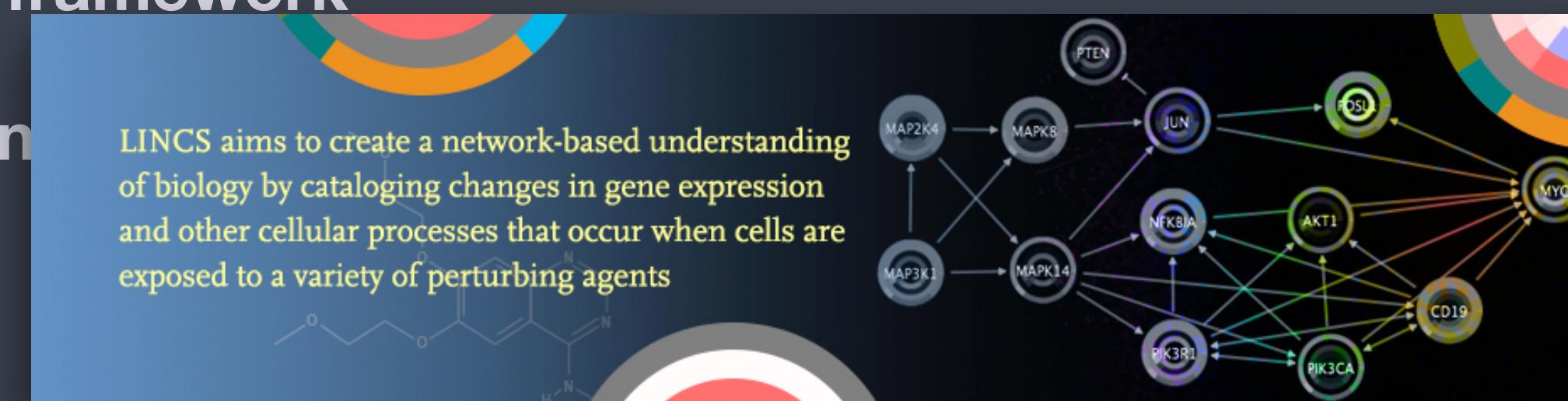
- 2 genes 10 000 cases
- 25 000 cases yield 62 - 93 significance loci -
association with post synaptic density
- 80 000 exome sequences yield **120+ loci**

Understanding function

- ▶ **Understanding cellular circuitry**
- ▶ **Common analytical framework**
- ▶ **Signatures of response**

Understanding function

- ▶ Understanding cellular circuitry
- ▶ Common analytical framework
- ▶ Signatures of response



Cohort/Samples for understanding disease

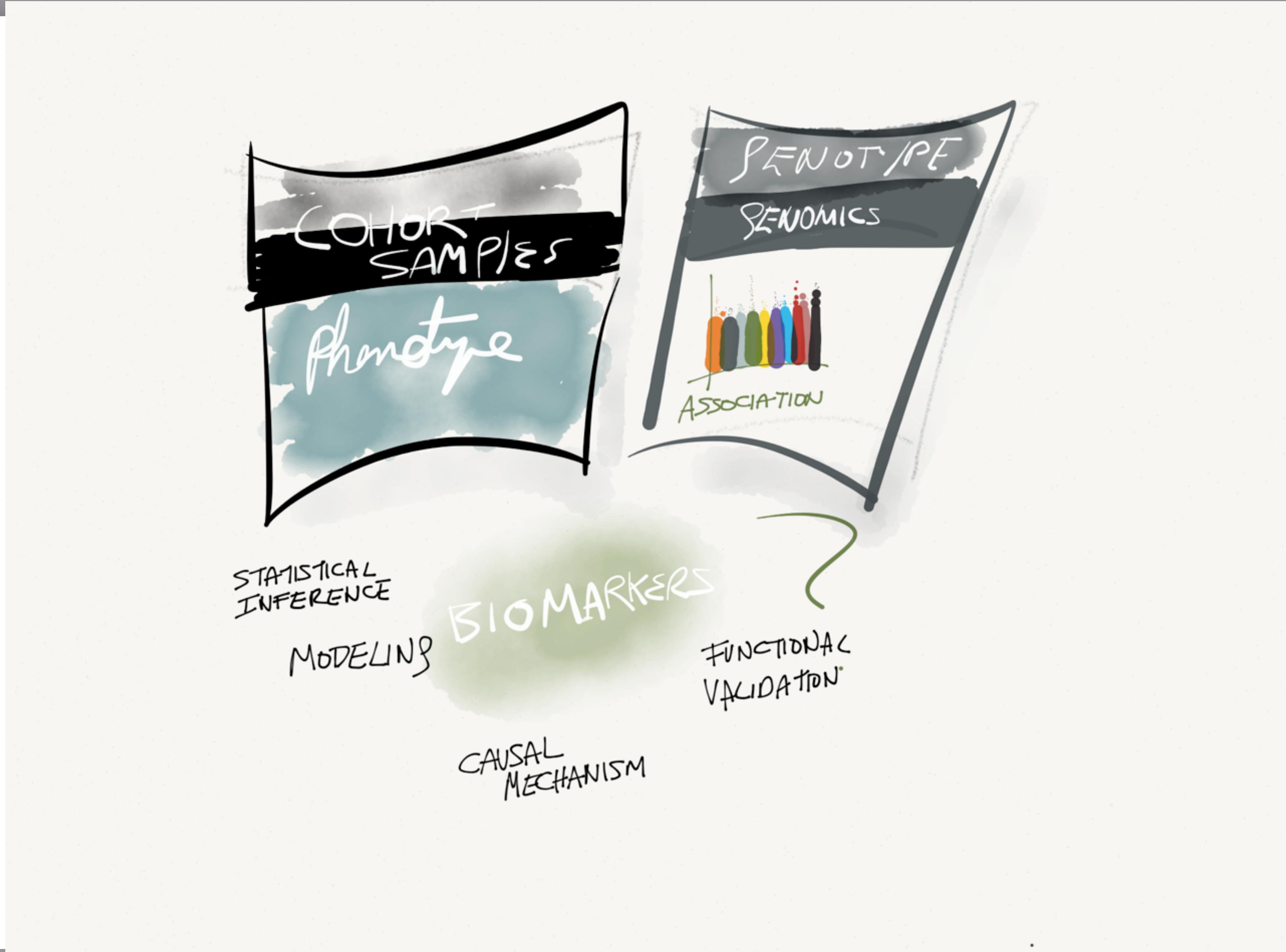
- ▶ “Classic” Phenotype
- ▶ Clinical/Physiological/Physical/Gene markers
- ▶ Molecular



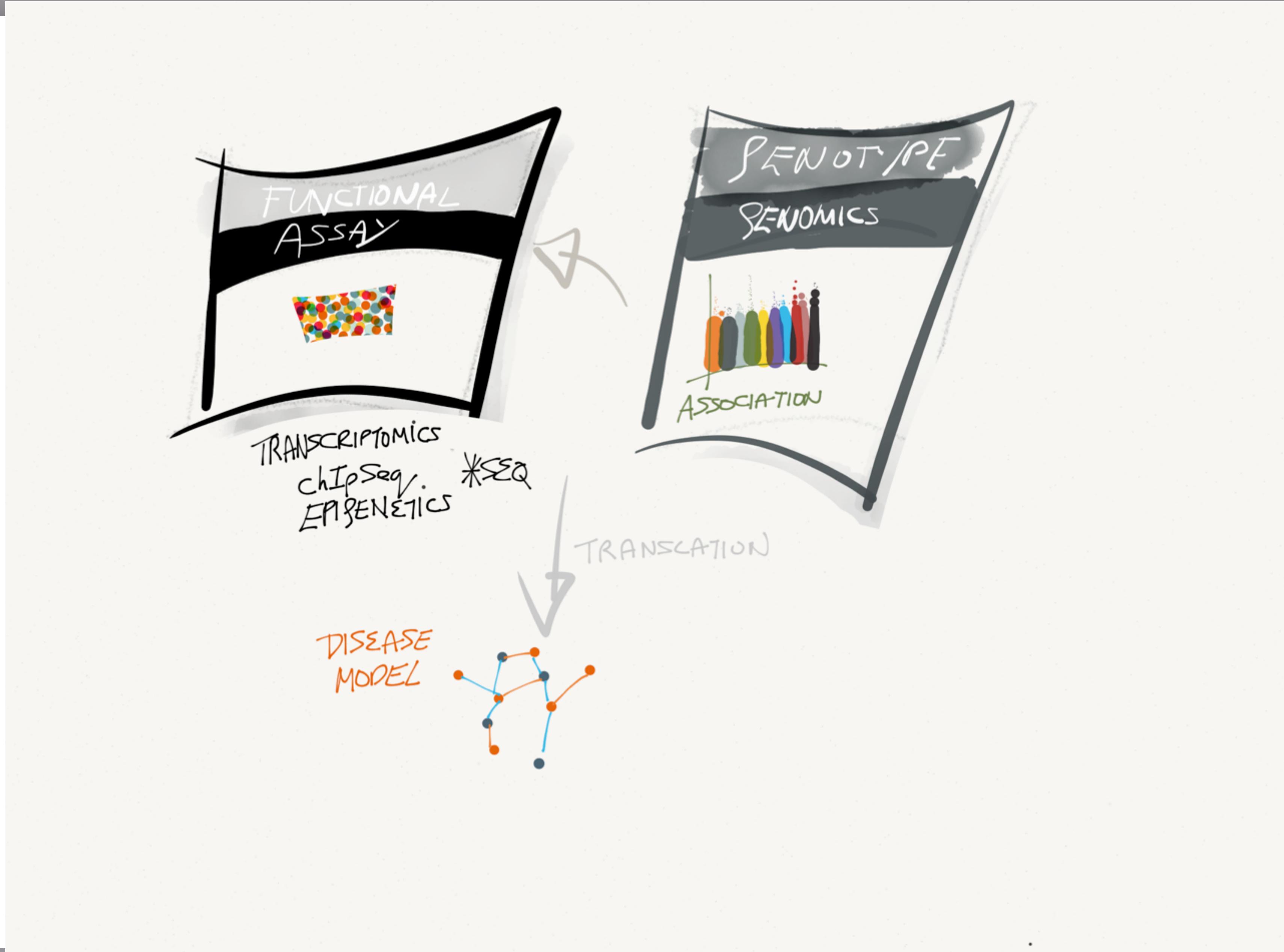


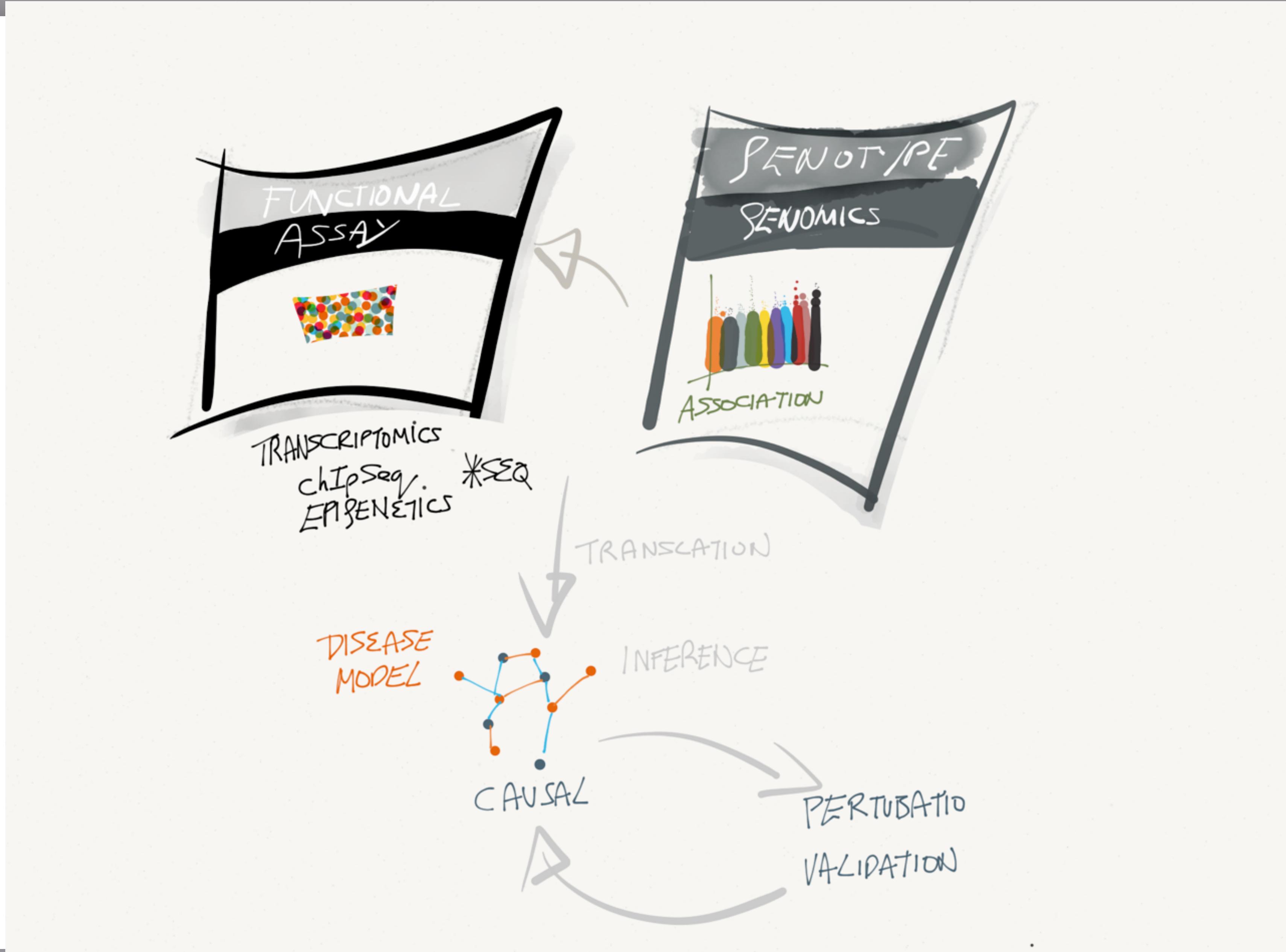












How to simplify and thus understand complexity?

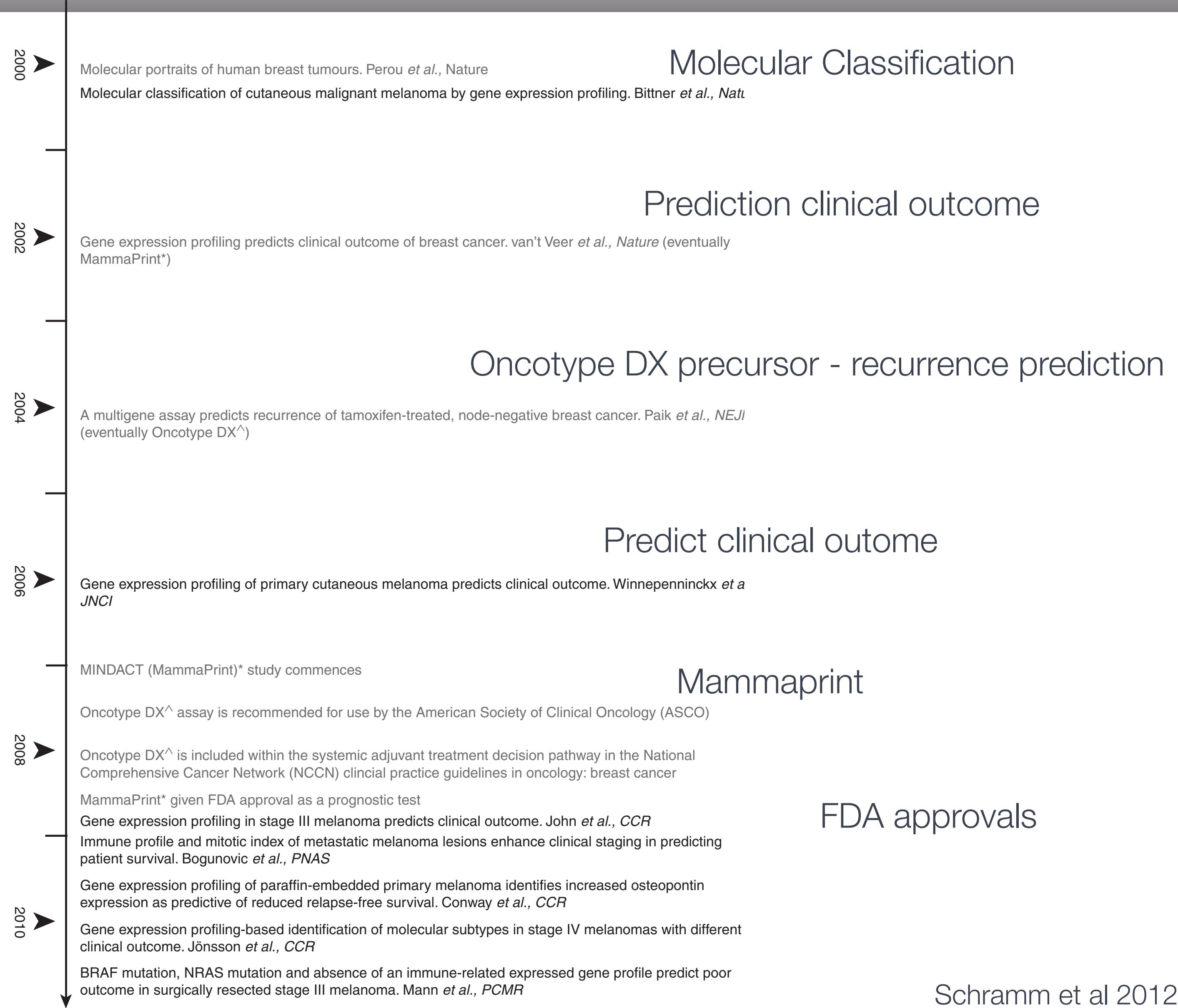
- ▶ Can we make consistent signatures?

Signatures



Reports on signatures

- ▶ 3752 publications for ‘disease signature’ in PubMed
- ▶ 468 ‘prognostic disease signature’
- ▶ 969 in breast cancer



Clinical prognostic value of gene expression signatures in colorectal cancers

Sanz-Pamplona et al PLoS One 2012

Clinical prognostic value of gene expression signatures in colorectal cancers

- ▶ 31 gene expression signatures
- ▶ 11 with outcome information
- ▶ 5 with significant association with prognosis
- ▶ all with **low reproducibility in independent data**
- ▶ “*better strategies for signature validation are needed*”

Sanz-Pamplona et al PLoS One 2012

Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome

David Venet¹, Jacques E. Dumont², Vincent Detours^{2,3*}

1 IRIDIA-CoDE, Université Libre de Bruxelles (U.L.B.), Brussels, Belgium, **2**IRIBHM, Université Libre de Bruxelles (U.L.B.), Campus Erasme, Brussels, Belgium, **3** WELBIO, Université Libre de Bruxelles (U.L.B.), Campus Erasme, Brussels, Belgium

Abstract

Bridging the gap between animal or *in vitro* models and human disease is essential in medical research. Researchers often suggest that a biological mechanism is relevant to human cancer from the statistical association of a gene expression marker (a signature) of this mechanism, that was discovered in an experimental system, with disease outcome in humans. We examined this argument for breast cancer. Surprisingly we found that gene expression signatures unrelated to cancer often affect cancer outcome.

Repeatability of published microarray gene expression analyses

John P A Ioannidis^{1–3}, David B Allison⁴, Catherine A Ball⁵, Issa Coulibaly⁴, Xiangqin Cui⁴, Aedín C Culhane^{6,7}, Mario Falchi^{8,9}, Cesare Furlanello¹⁰, Laurence Game¹¹, Giuseppe Jurman¹⁰, Jon Mangion¹¹, Tapan Mehta⁴, Michael Nitzberg⁵, Grier P Page^{4,12}, Enrico Petretto^{11,13} & Vera van Noort¹⁴

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Why Most Published Research Findings
Are False Ioannidis JPA (2005) Why most published research findings are false.
PLoS
Med 2: e124. doi:10.1371/journal.pmed.0020124

Viable, Useful Prognostic Signatures?

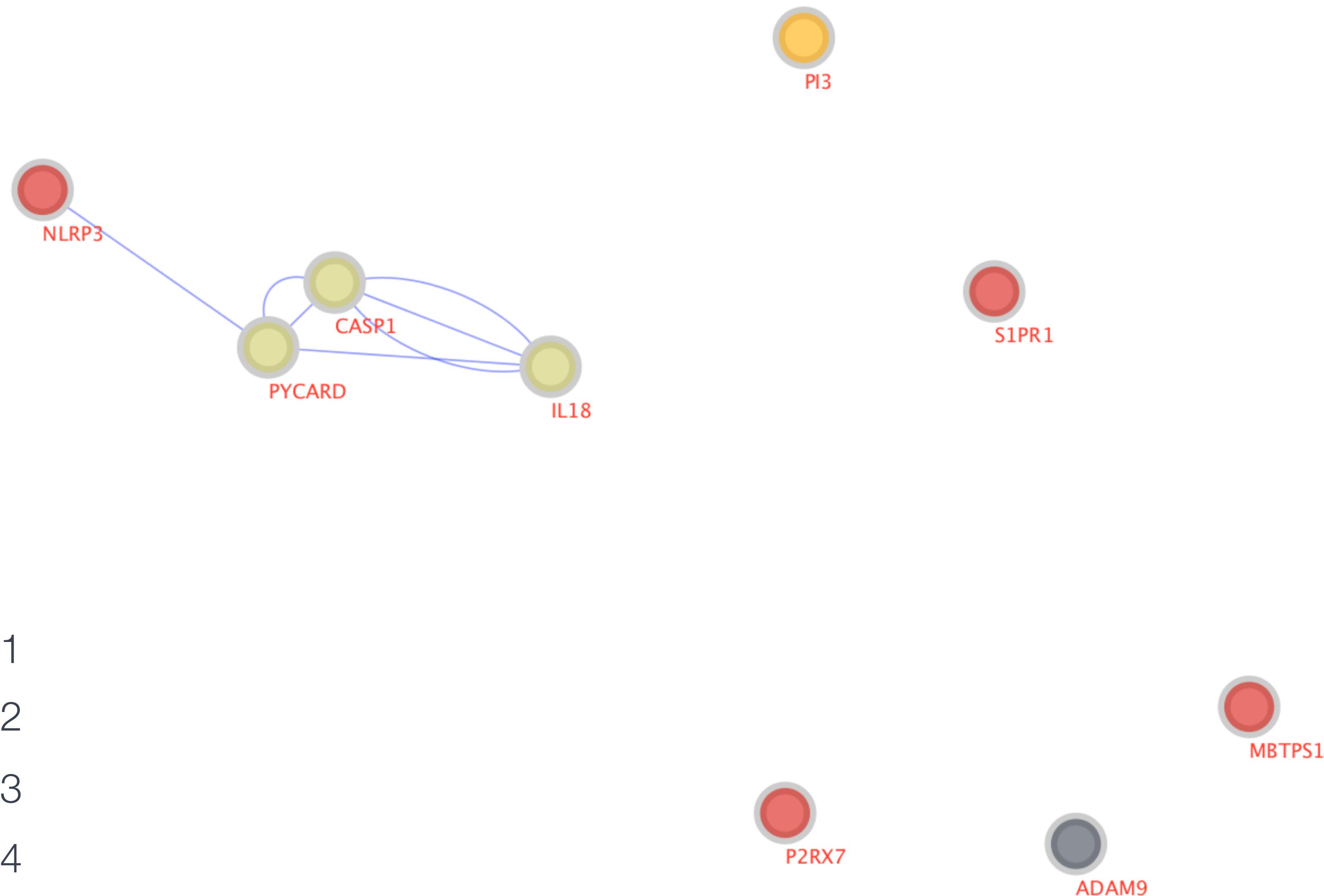
Take home

- ▶ Gene signatures can, at best, **complement traditional pathology**
- ▶ **Predictive signatures exist**, but lack broad efficacy and have poor reproducibility
- ▶ Signatures are valuable for helping understand correlation with molecular signatures data and GWAS data **variation, heterogeneity and aggregation** of physiological processes causal to a disease

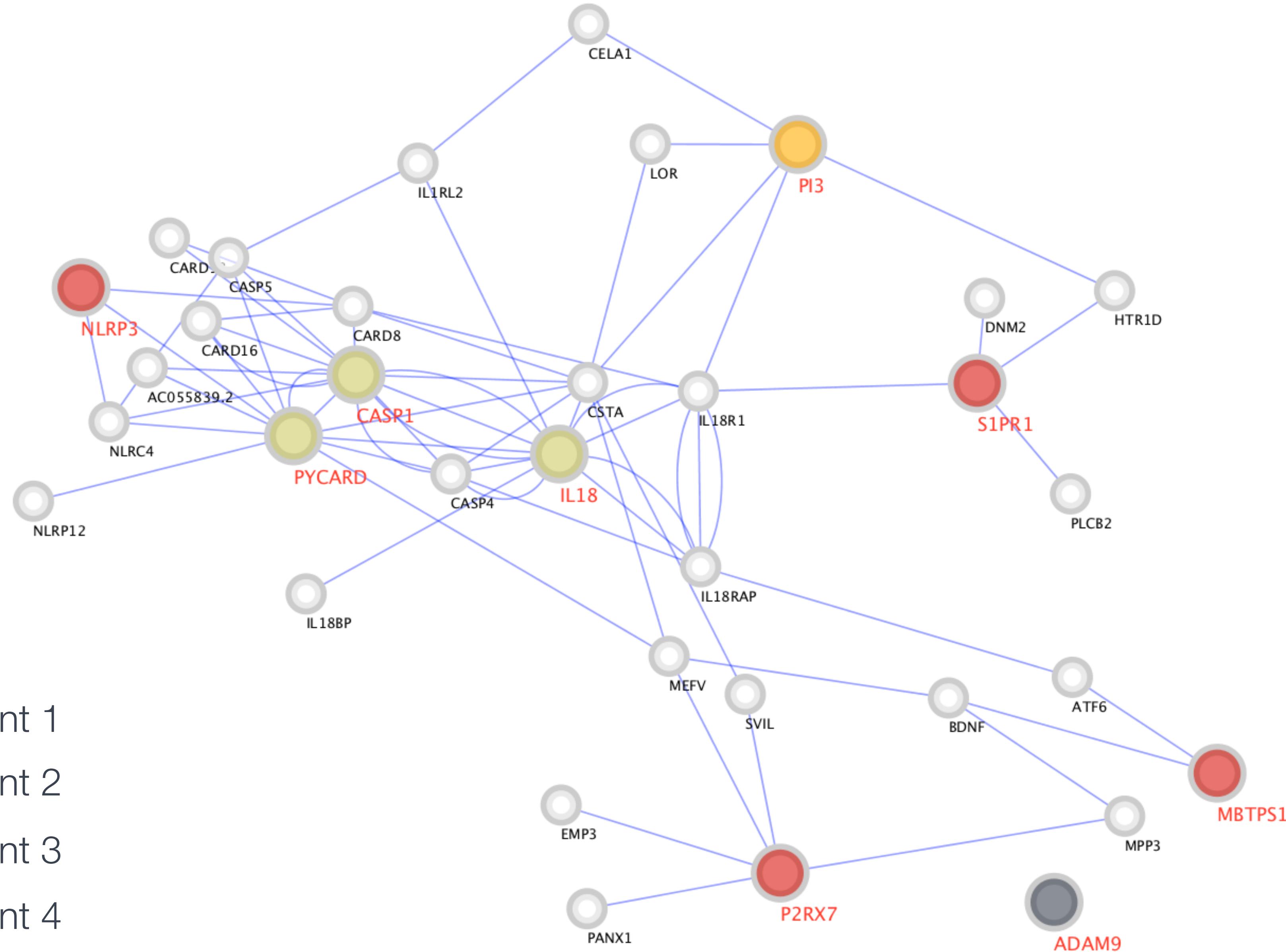
Defining the next level of cellular organisation

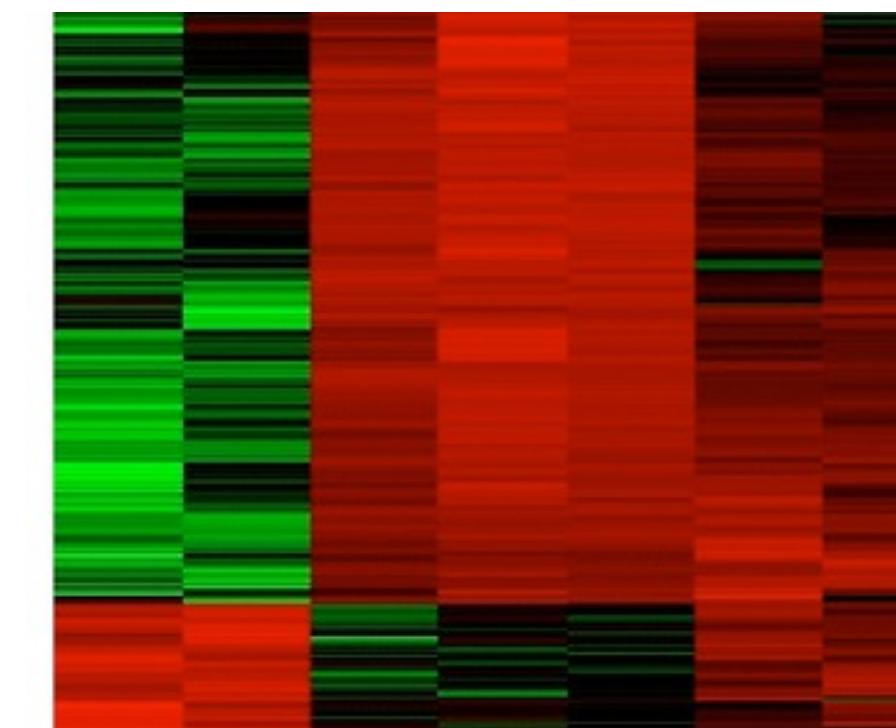
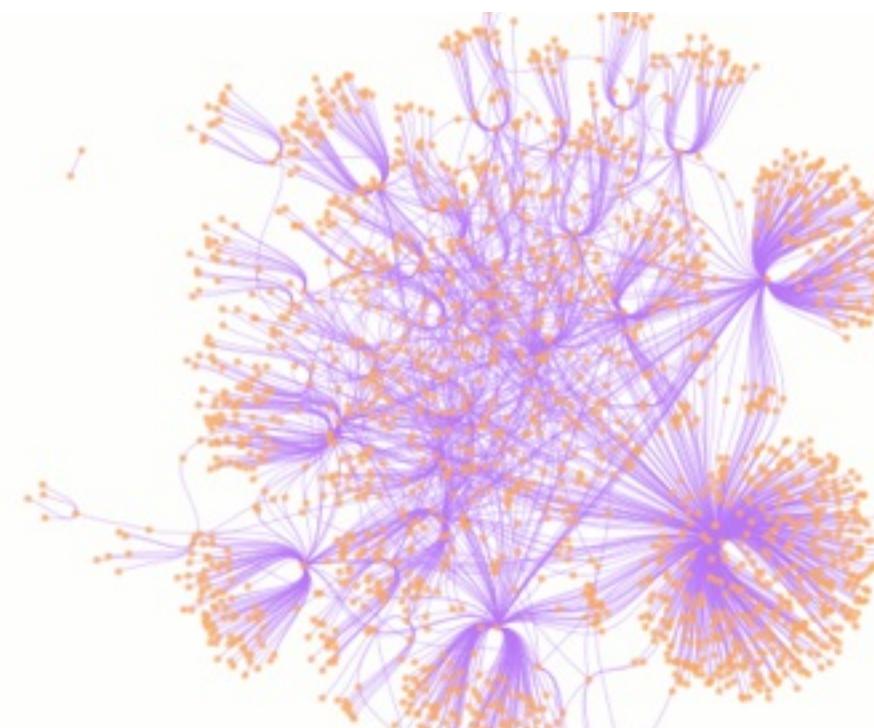
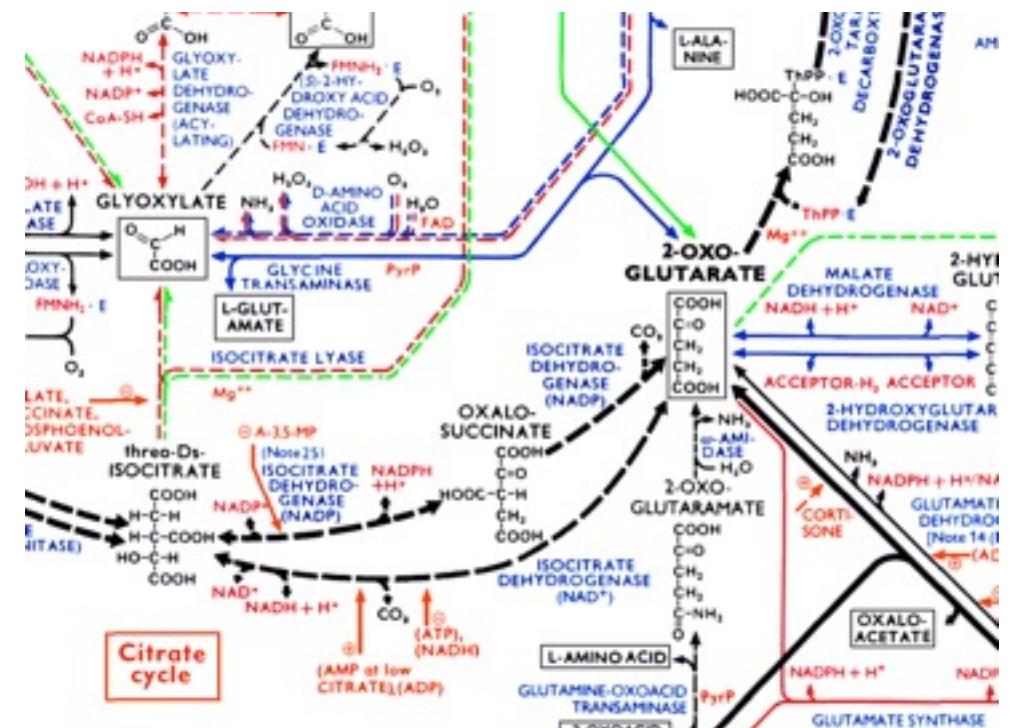
- ▶ Gene
- ▶ Gene-Gene
- ▶ Network
- ▶ Network and pathway interactions
- ▶ **Functional modules of the cell**

ARDS



- Patient 1
- Patient 2
- Patient 3
- Patient 4



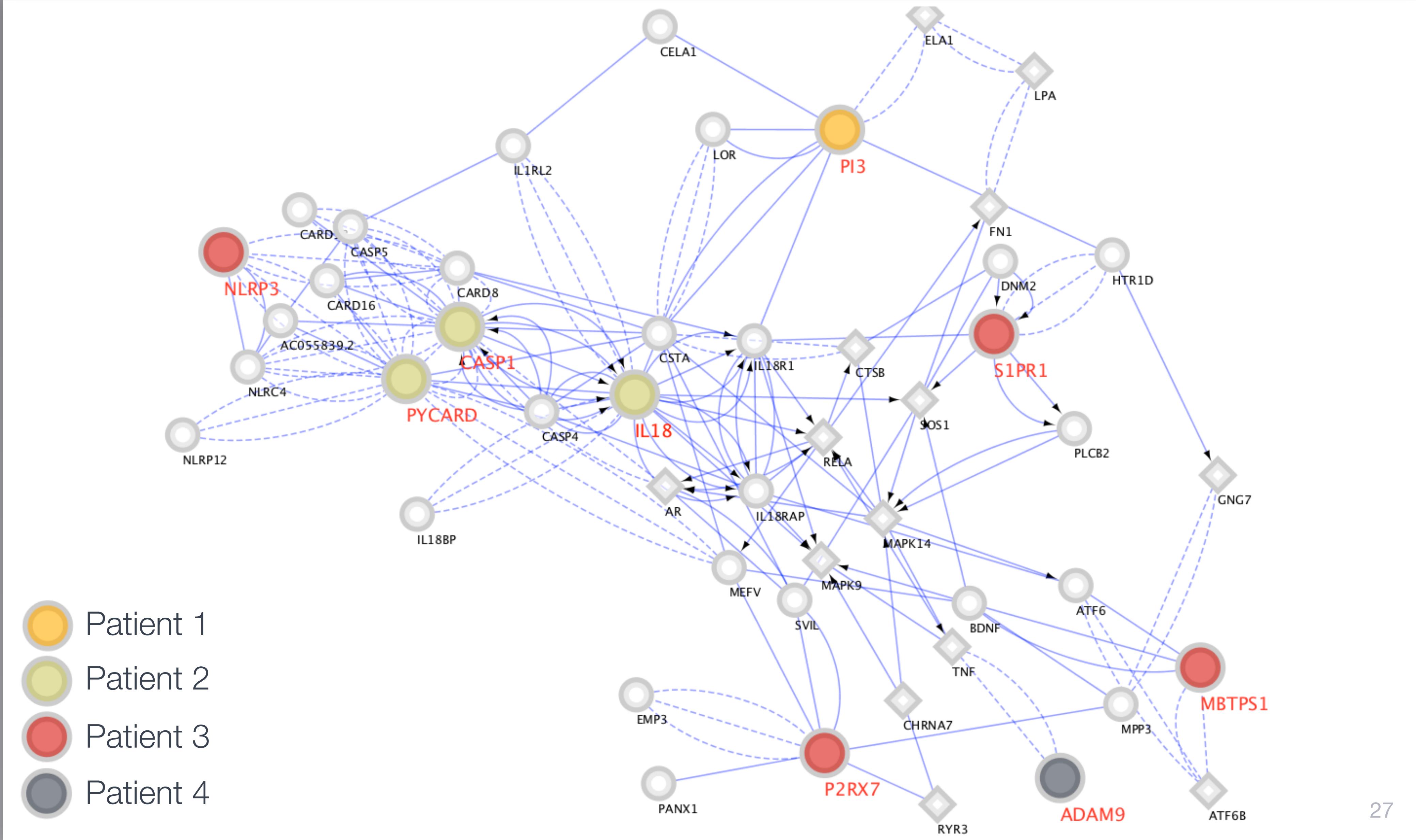


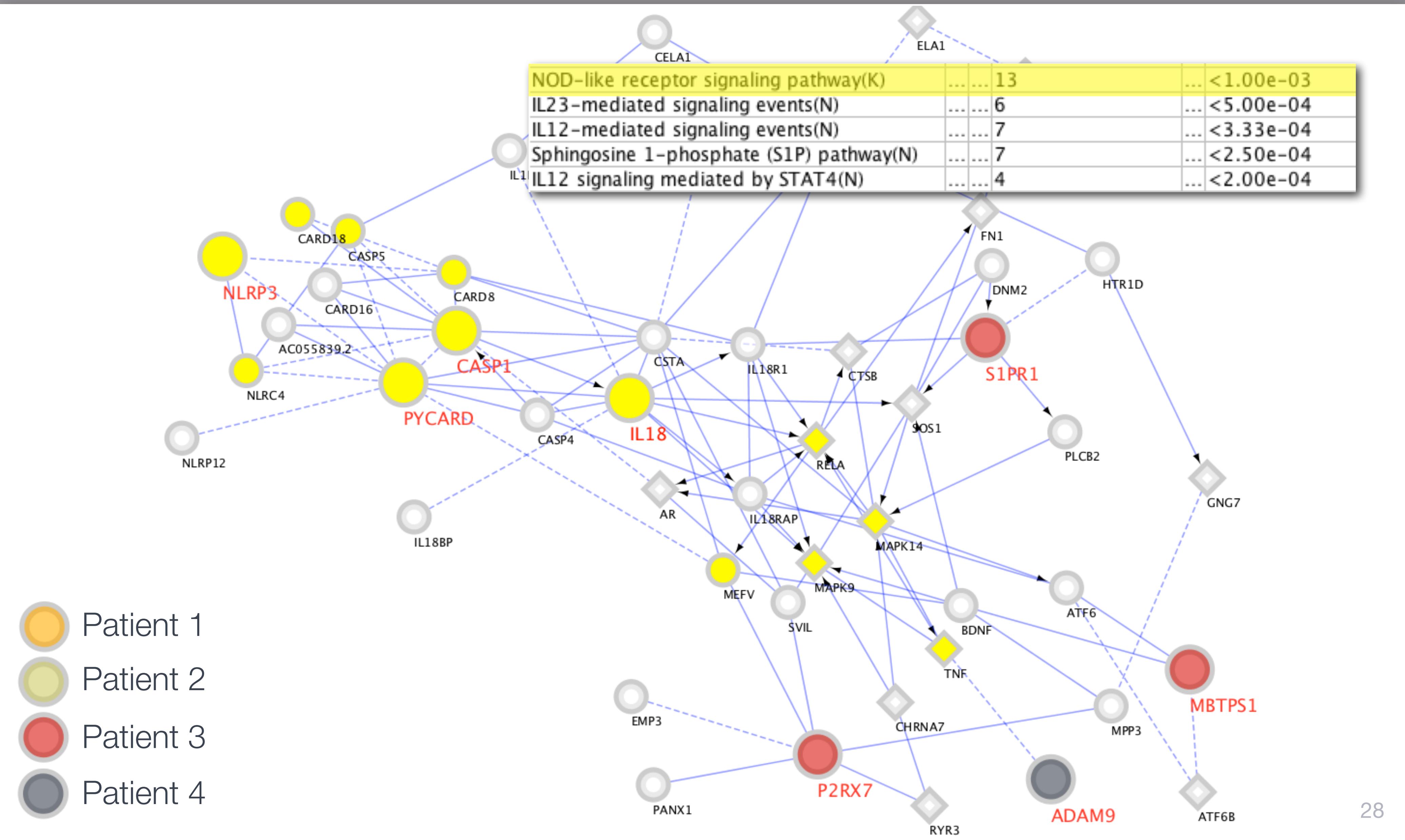
Known interactions

Protein-protein
binding

Consistent co-
expression

Transcriptional
regulation





Genes interact

Genes interact

- Lists

Genes interact

- Lists
- Modules (networks)

Genes interact

- Lists
- Modules (networks)
- Pathways (usually something is known)

Genes interact

- Lists
- Modules (networks)
- Pathways (usually something is known)
- Processes (open set of pathways or modules)

Genes interact

- Lists
- Modules (networks)
- Pathways (usually something is known)
- Processes (open set of pathways or modules)
- Biological function (process within a system)

Key Pathways

“Give me the key pathways that drive [PHENOTYPE]”

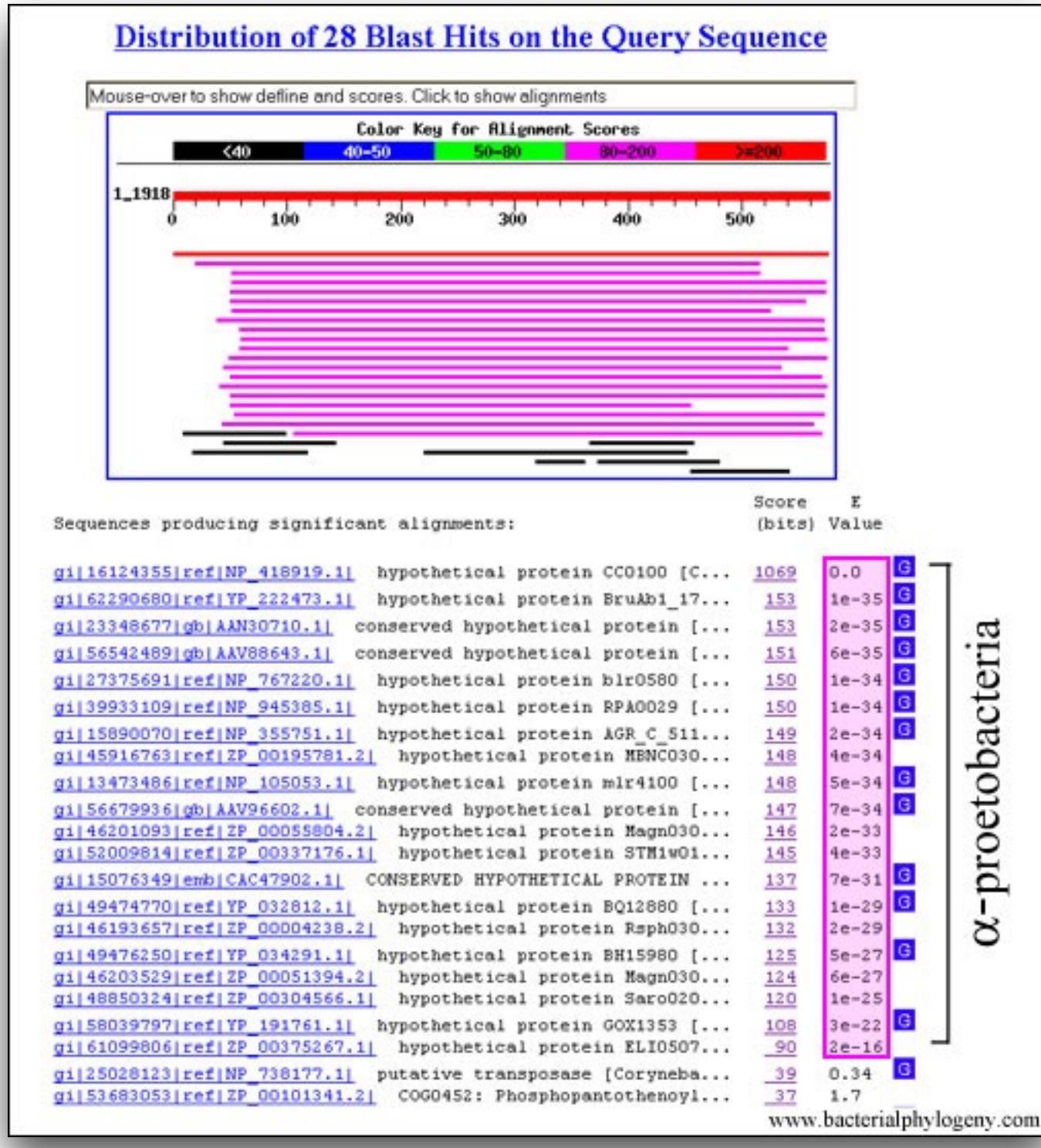
Pathway context

What does this pathway do?

“I’m not interested in this as I don’t know [GENESET NAME]

How good a target/validation will this be?

Context search



- Pathways do not act in isolation
- Pathway context > mechanism

BLAST - contextual list of gene names that provide understanding of your gene

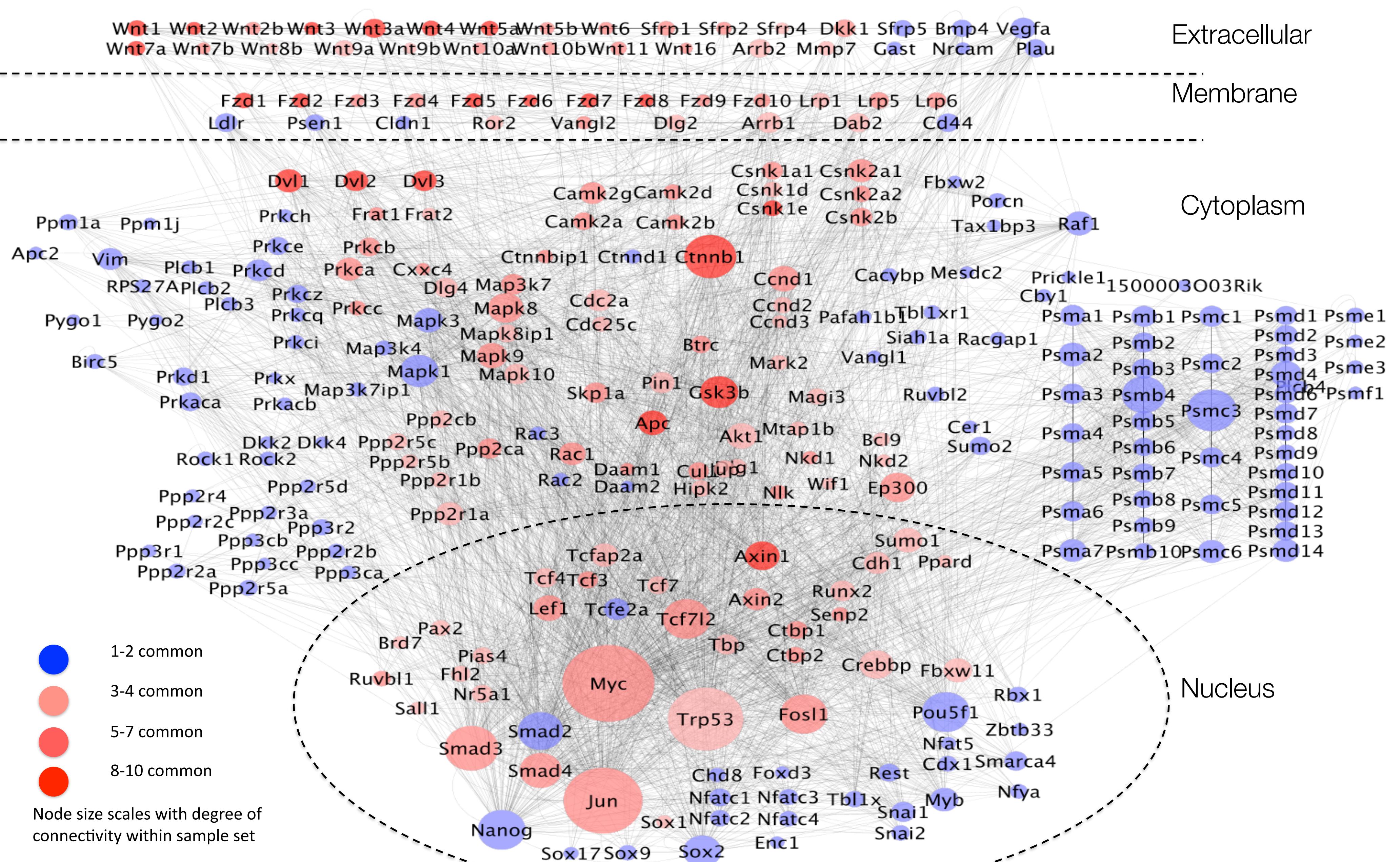
Function is not systematically described

Curated pathways, incomplete, drive understanding

coverage ± 6000 genes

Many different descriptions of the same pathways

Wnt consensus pathway



Function consists of interacting genesets

- Curated knowledge: validated, incomplete, inconsistent, *drives understanding*
- Data driven: judged by association with known function (enrichment with annotated terms), *validation dependent*

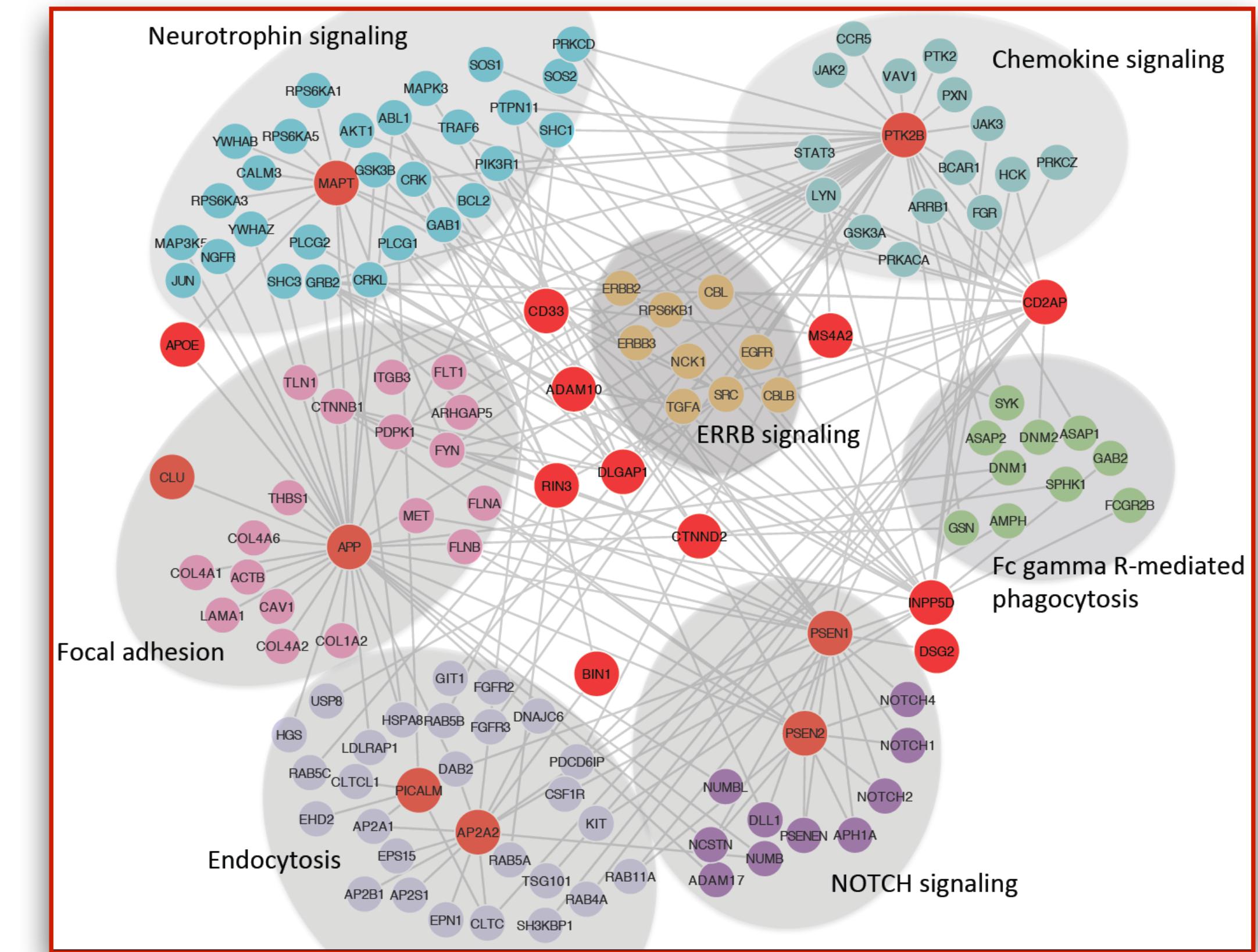
How do pathways relate to each other?

- How does my geneset relate to both known and data driven genesets and pathways?
 - *Where am I?*
 - What are the core pathways driving a phenotype?
 - *What is the mechanistic basis of my phenotype?*
 - What is the relationship between genetic/genomic underpinning and the functional phenotype?
 - *How to systematically translate genome variation to function?*

The functional landscape

Data driven: Network maps of the cell

- Sets of gene clusters tested for enrichment of existing GO terms



Simplified strategy

Simplified strategy

Gene-
Variants

GWA signals

Simplified strategy

Gene-
Variants

score

GWA signals

Map to genes

Simplified strategy

Gene-
Variants

GWA signals

score

Map to genes

Enrichment
systems

mSIGdb
DAVID

Simplified strategy

Gene-
Variants

GWA signals

score

Map to genes

Enrichment
systems

mSIGdb
DAVID

Phenotype
association

Simplified strategy

Molecular basis > Clinical interpretation

Gene-
Variants

score

GWA signals

Map to genes

Enrichment
systems

mSIGdb
DAVID

Phenotype
association