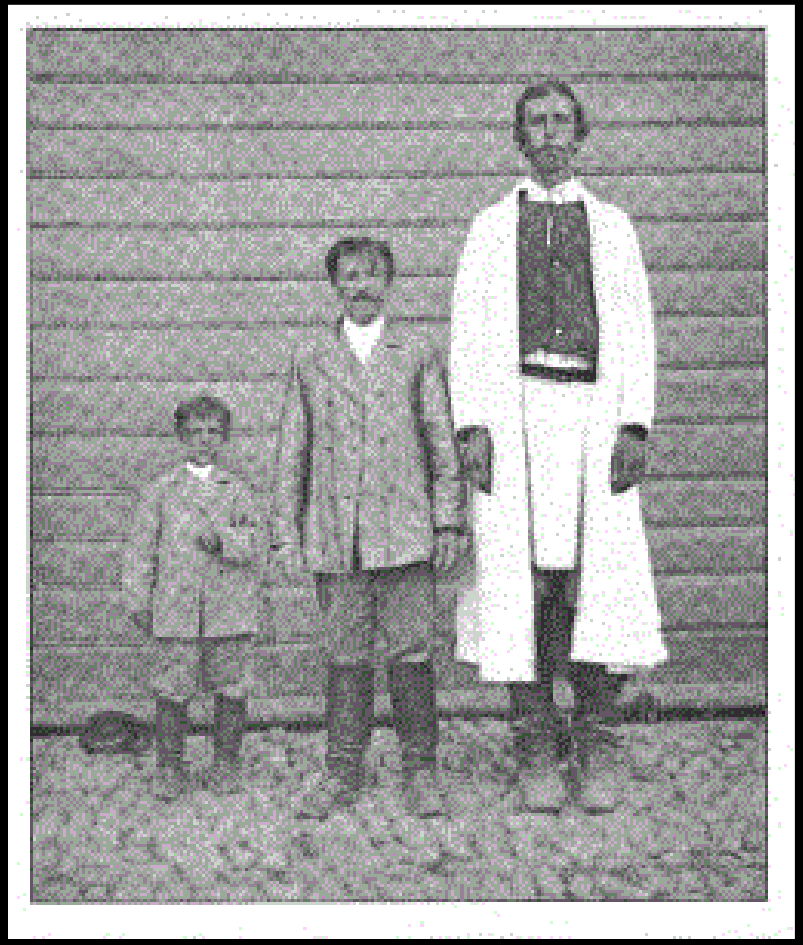
Understanding Genotype-Phenotype relations in Cancer via Network Approaches

Teresa Przytycka

NIH / NLM / NCBI

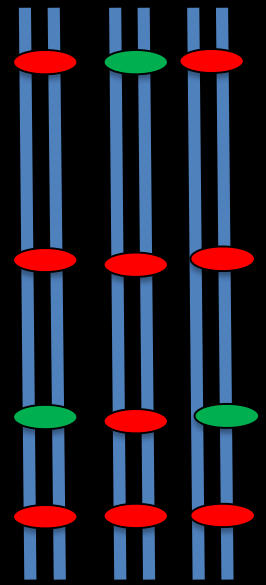
Phenotypes

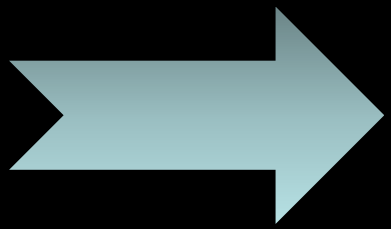


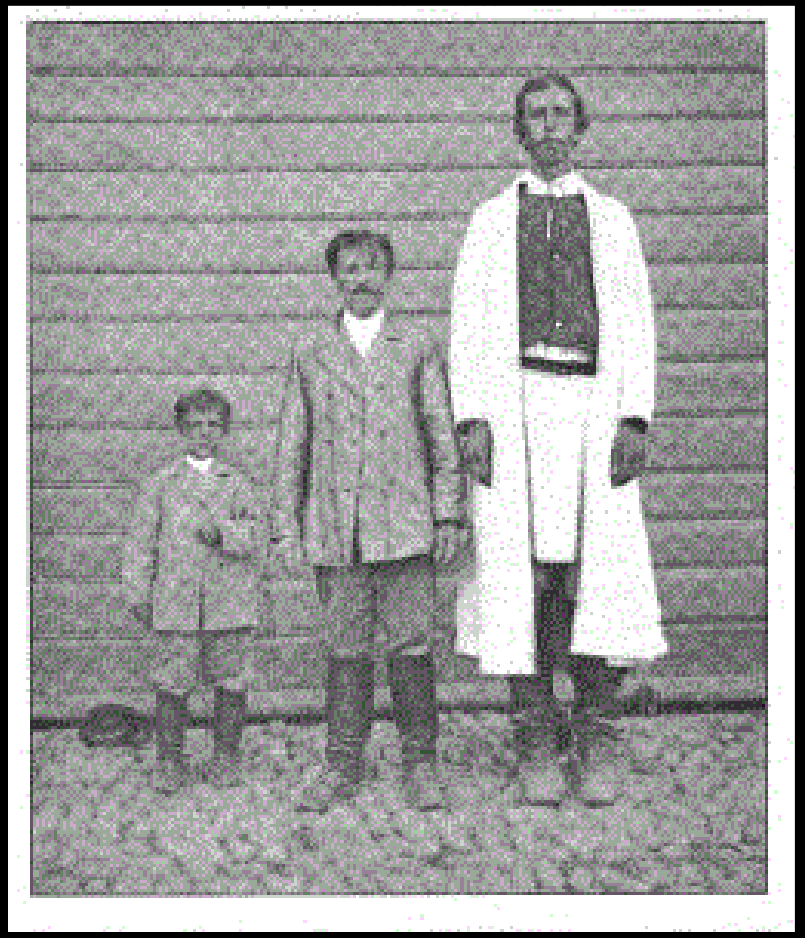
Journal “Wisla” (1902)

Genotypes/causes

Phenotypes







Journal "Wisla" (1902) Picture from a local fare in Lublin, Poland

Key challenges in cancer genotype-phenotype analysis

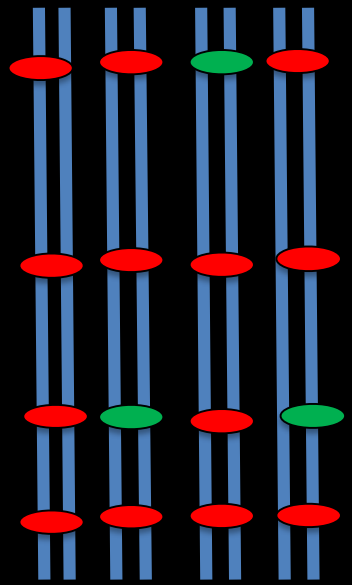
- Complexity: Multiple driver mutations are typically required for caner progression
    - Driver mutations /alterations– mutations contributing to cancer progression
    - Passenger mutations – neutral mutations accumulating during cancer progression
- Heterogeneity: Phenotypically similar cancer cases might be caused by different sets of driver mutations
- Some driver mutations are rare
- Epistasis – masking of the effect of one mutation by another mutation
- Cancer evolution

Network/Systems biology view

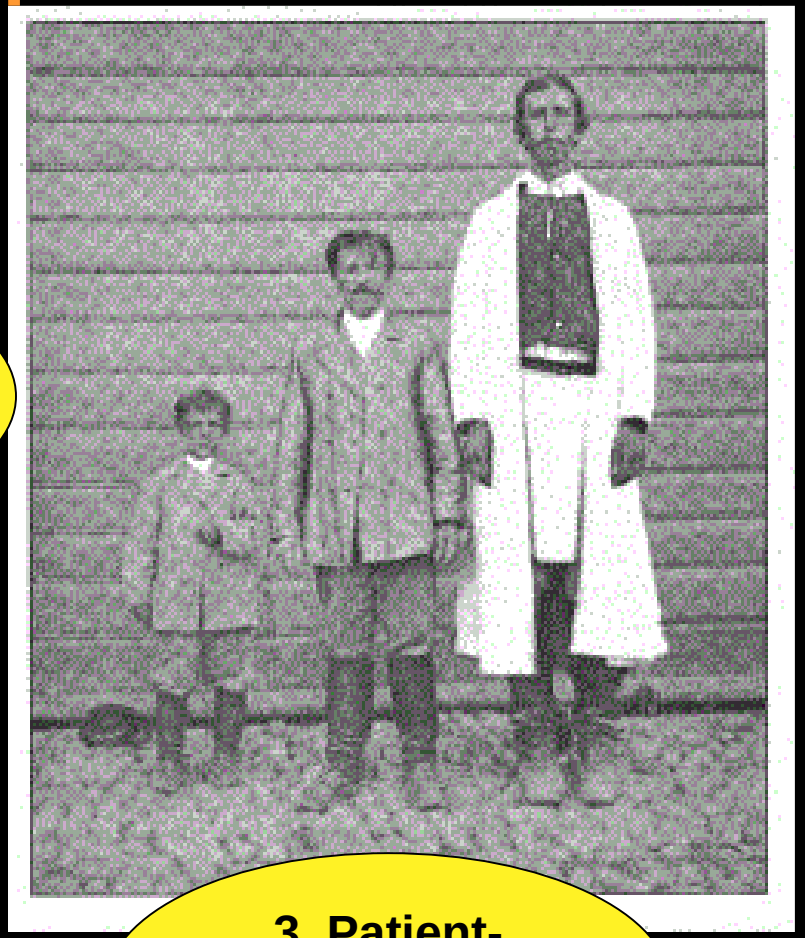
Motivation: Molecules function in the context of interaction networks :

- Effects of genetic alteration propagate through the interaction network affecting downstream genes
- Different driver mutations often dys-regulate common pathways

Utilizing Networks for Understanding Genotype-Phenotype effects



2. Network based signal propagation

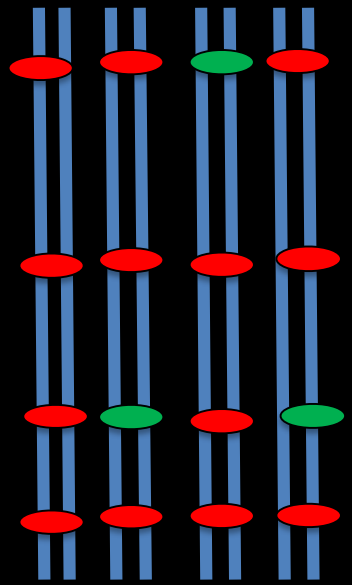


1. Dys- regulated Networks

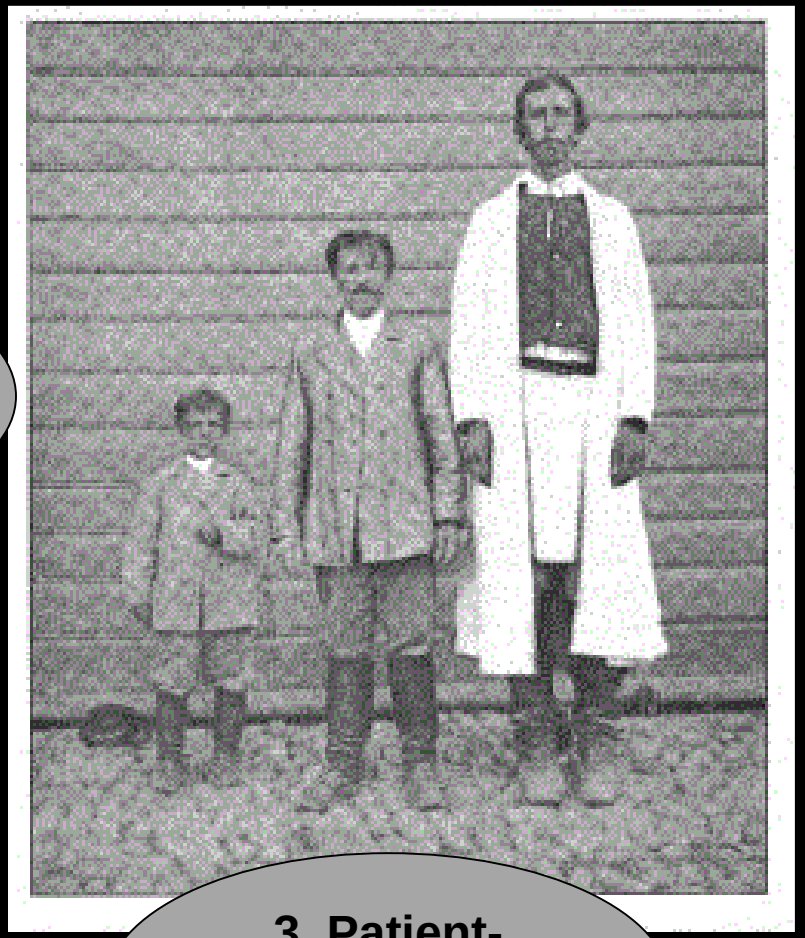
3. Patient-similarity Networks

Genotypes

Phenotypes



2. Network based signal propagation

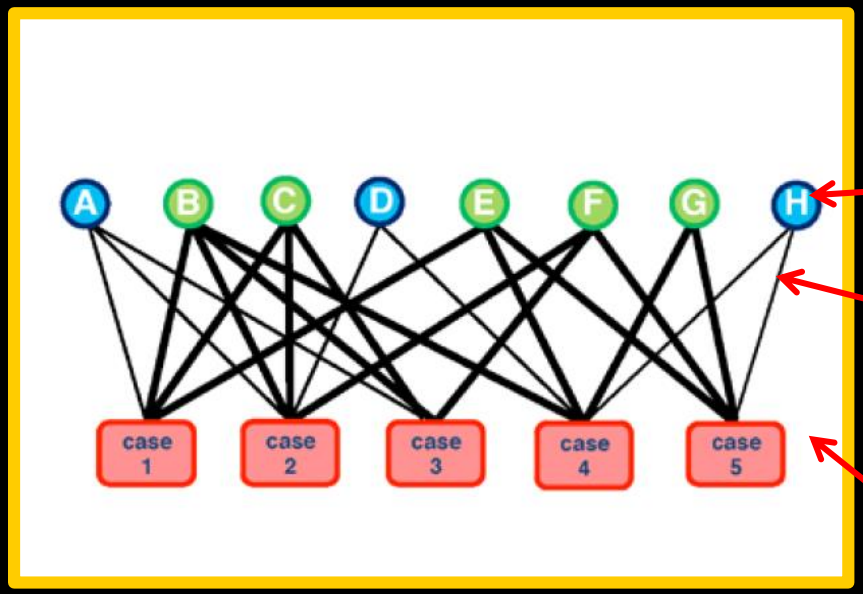


1. Dys- regulated Networks

3. Patient-similarity Networks

Set cover approach as a method to find cancer drivers/markers – parsimony approach

Goal: Given a set of dysregulated genes and disease cases, find a representative set of dysregulated genes



Genes

Gene "covers" the case (it is altered in this case)

Disease Cases

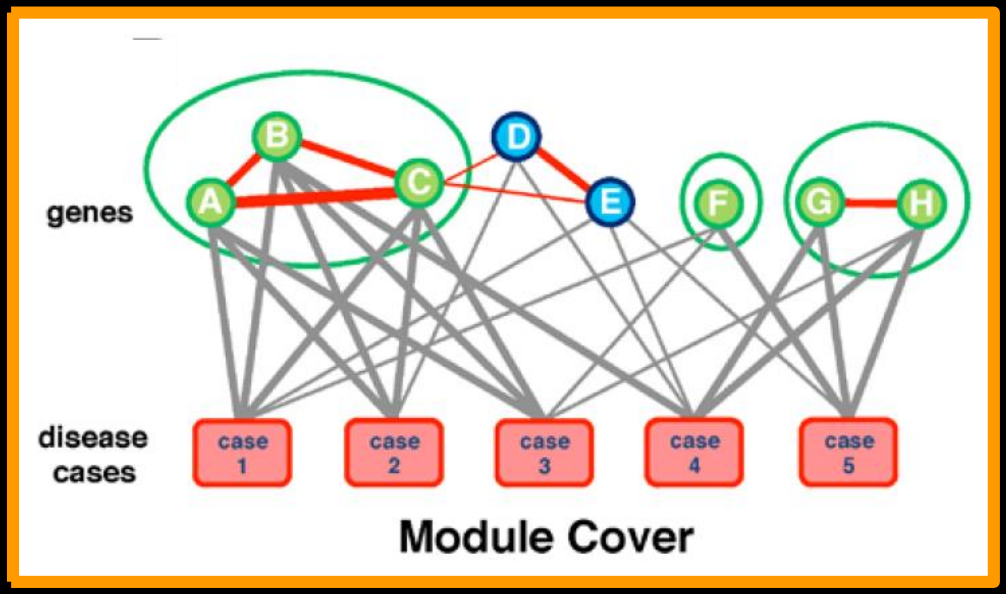
Module Cover Approach

Optimization problem:

Find smallest cost set of modules so that each disease case is covered at least k times

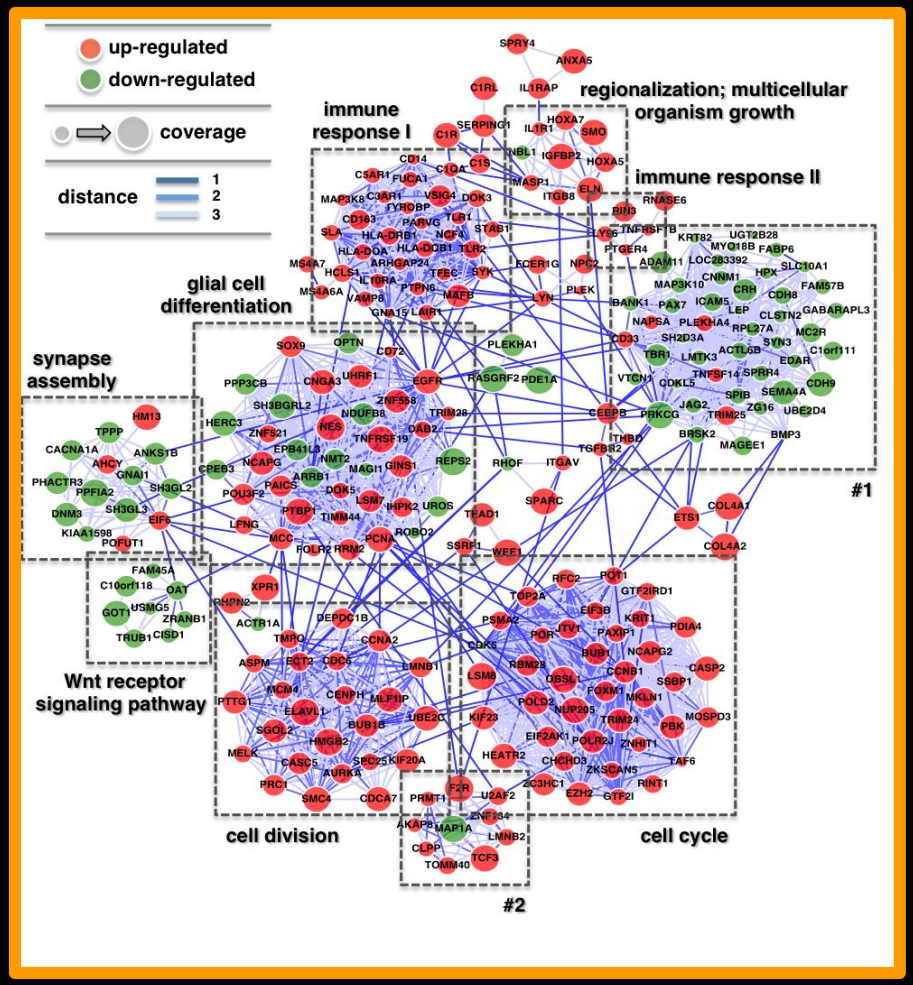
Cost is a function of:

- ↓ **distance in the network of genes in same module**
- ↓ **A similarity measure (application dependent)**
- ↑ **number of modules (parameterized penalty)**



Kim et al. PSB 2013

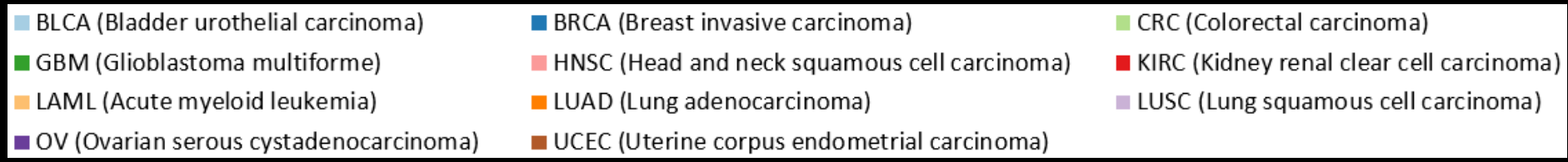
Module Cover: Glioblastoma Data



Signature modules from GBM Dataset (REMBRANDT)

Kim et al. PSB 2013

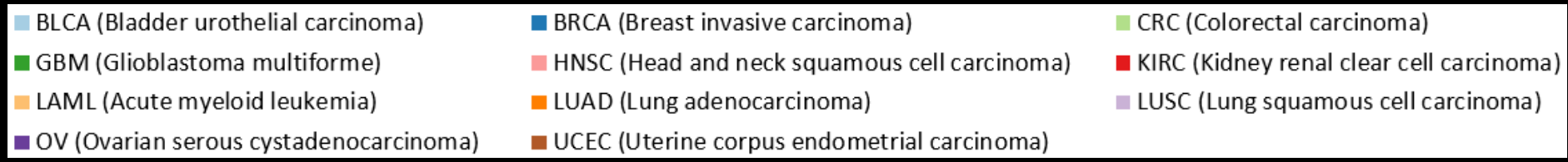
The Pan-Cancer initiative



- genetic and epigenetic aberrations in cancer samples from thousands of cancer patients over
- 12 cancer types

- Questions:
  - Differences
  - Similarities

The Pan-Cancer initiative



- genetic and epigenetic aberrations in cancer samples from thousands of cancer patients over
- 12 cancer types

- Questions:
  - Differences
  - Similarities

Network based approaches

Network based stratification (Ideker)

HotNet2 (Ben Raphael),
MEMCover (this presentation)

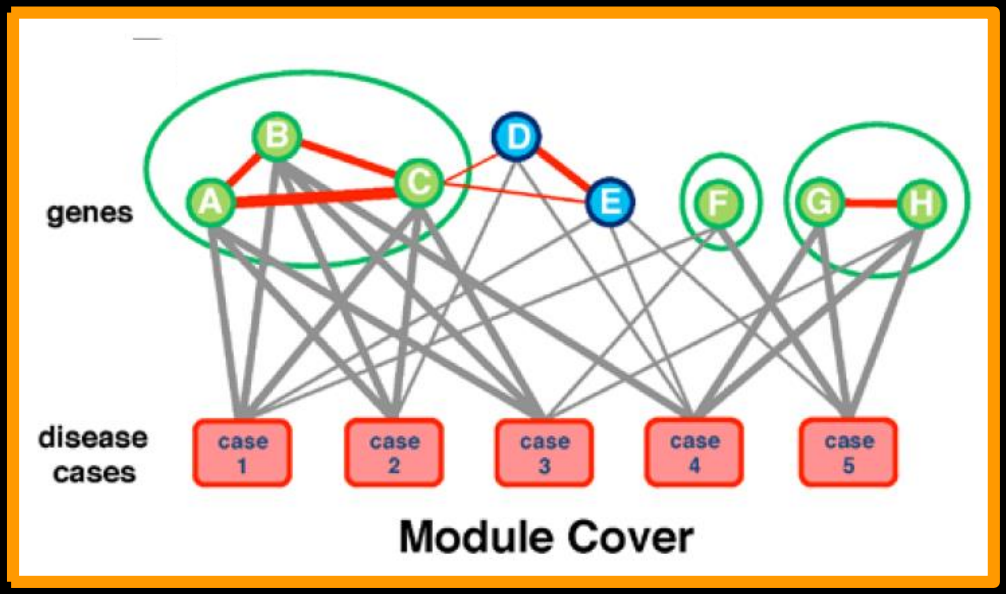
Module Cover Approach

Optimization problem:

Find smallest cost set of modules so that each disease case is covered at least k times

Cost is a function of:

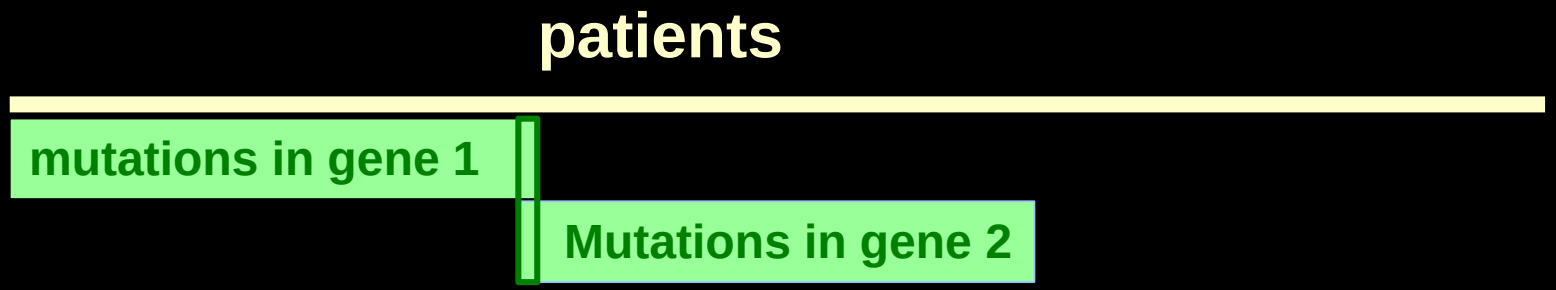
- ↓ **distance in the network of genes in same module**
- ↓ **A similarity measure (application dependent) ????**
- ↑ **number of modules (parameterized penalty)**



Kim et al. PSB 2013

In many cancer types cancer drivers are often mutually exclusive

Thomas et al 2007



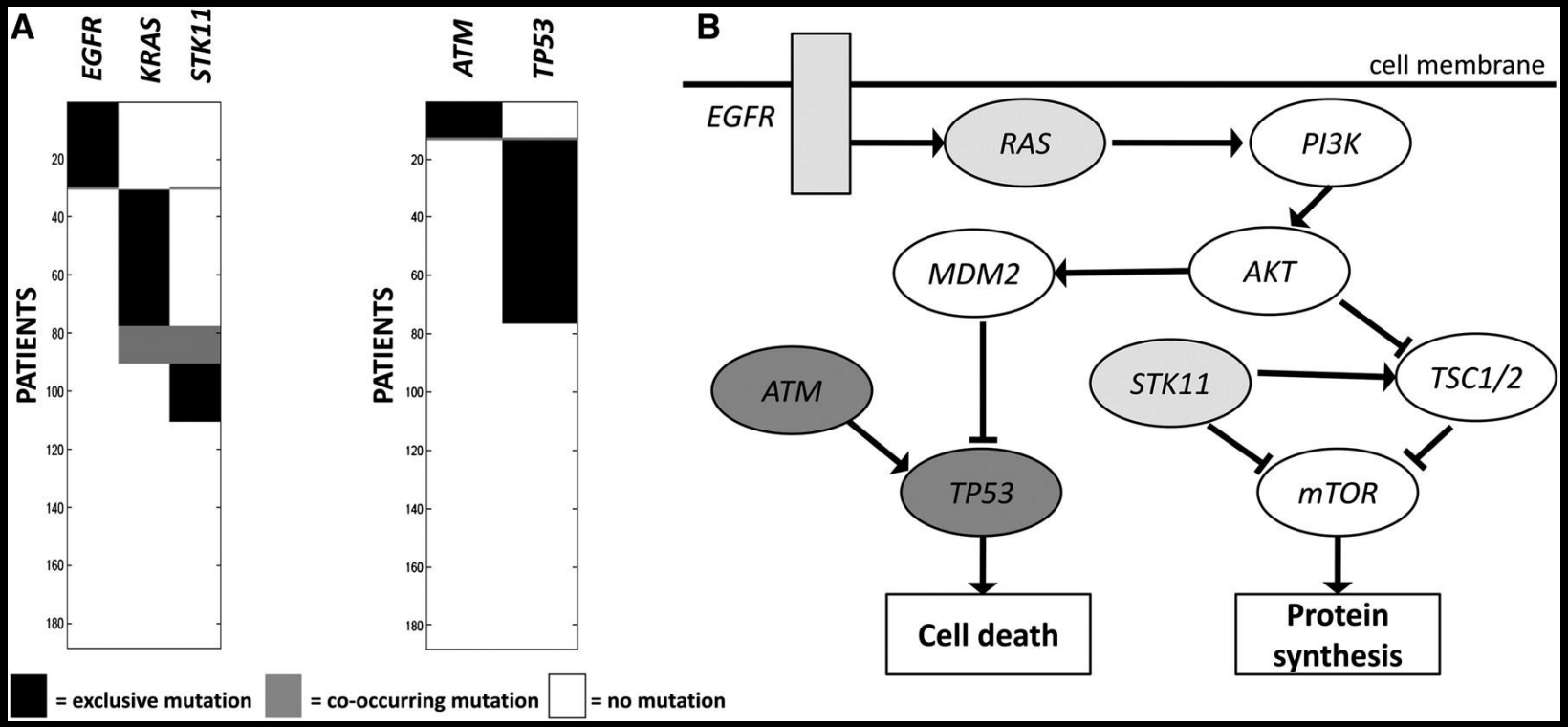
Possible explanations

- any of the two drivers alone gives sufficient growth advantage
- negative genetic interactions between drivers

Mutually exclusive pairs often act in the same pathway

Thomas et al 2007
Ciriello, et al., 2012;
Vandin, et al., 2012;
Leiserson, et al., 2013

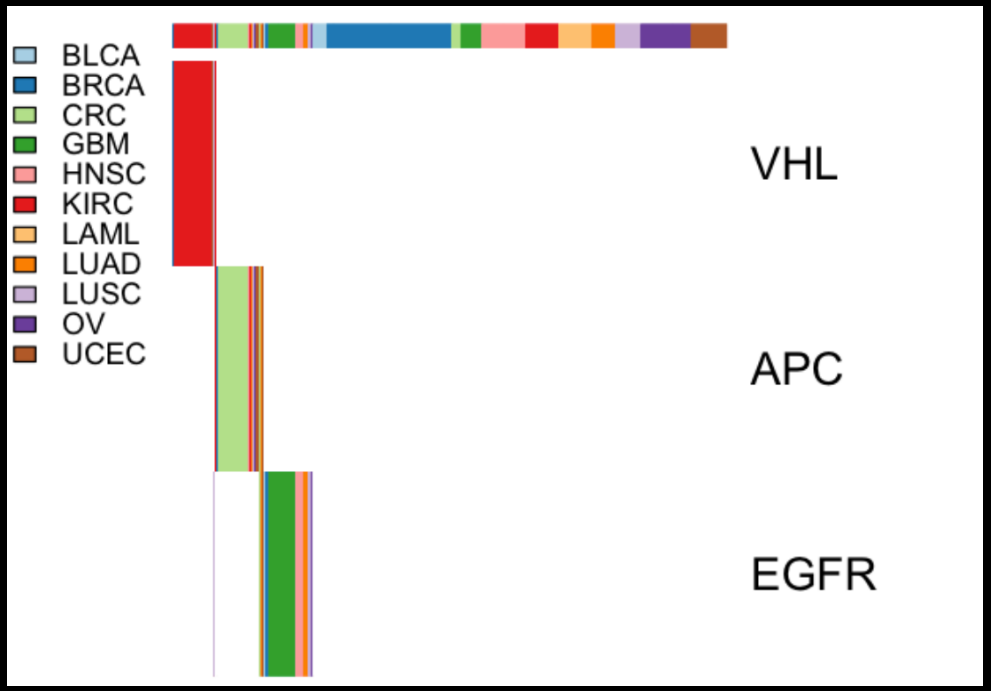
Example from Vandin et al.
(lung adenocarcinoma data)



Mutual Exclusivity and PanCancer TCGA

Can Mutual Exclusivity principle help identifying common pathways dysregulated across cancer types?

Cancer type specific mutations are mutually exclusive but not in necessarily in the same pathway



Introducing classification of mutual exclusivity

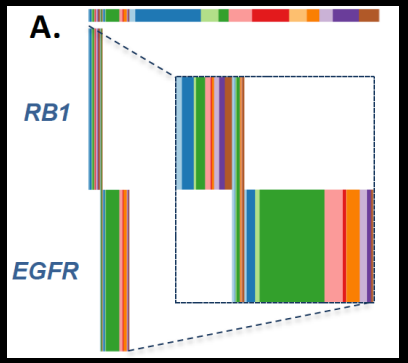
- **Within tissue exclusivity**
  WITHIN_ME

- **Across tissues exclusivity**
  ACROSS_ME

- **Between tissues exclusivity**
  BETWEEN_ME

Introducing classification of mutual exclusivity

- **Within tissue exclusivity**
  WITHIN_ME



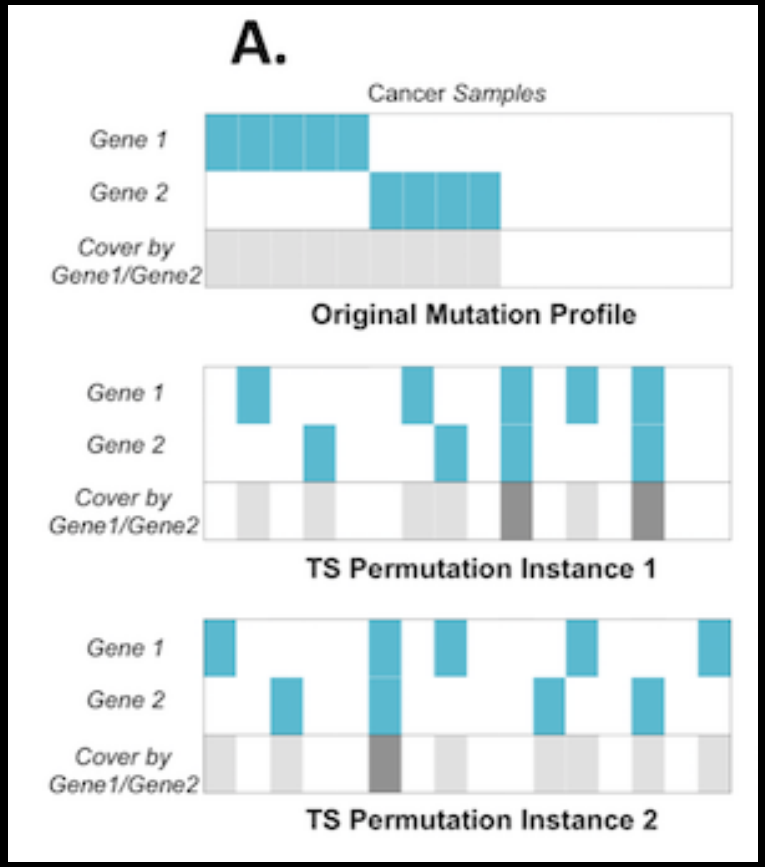
Traditional permutation test

- **Across tissues exclusivity**
  ACROSS_ME

- **Between tissues exclusivity**
  BETWEEN_ME

Permutation Test (within cancer type)

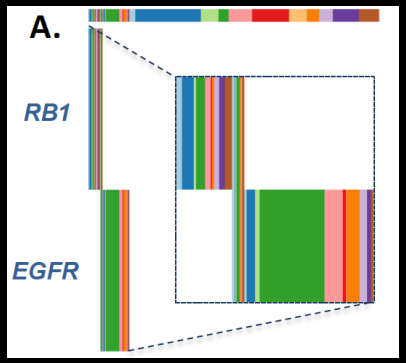
To preserve the mutation rates of each gene and each sample, in each iteration, two (gene, sample) pairs are randomly and swapped



Introducing classification of mutual exclusivity

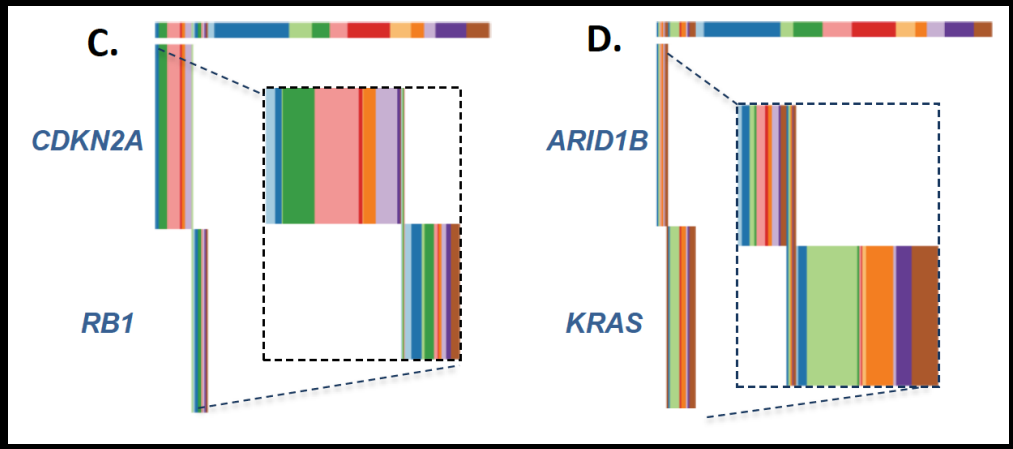
- **Within tissue exclusivity**
  WITHIN_ME

Traditional permutation test



- **Across tissues exclusivity**
  ACROSS_ME

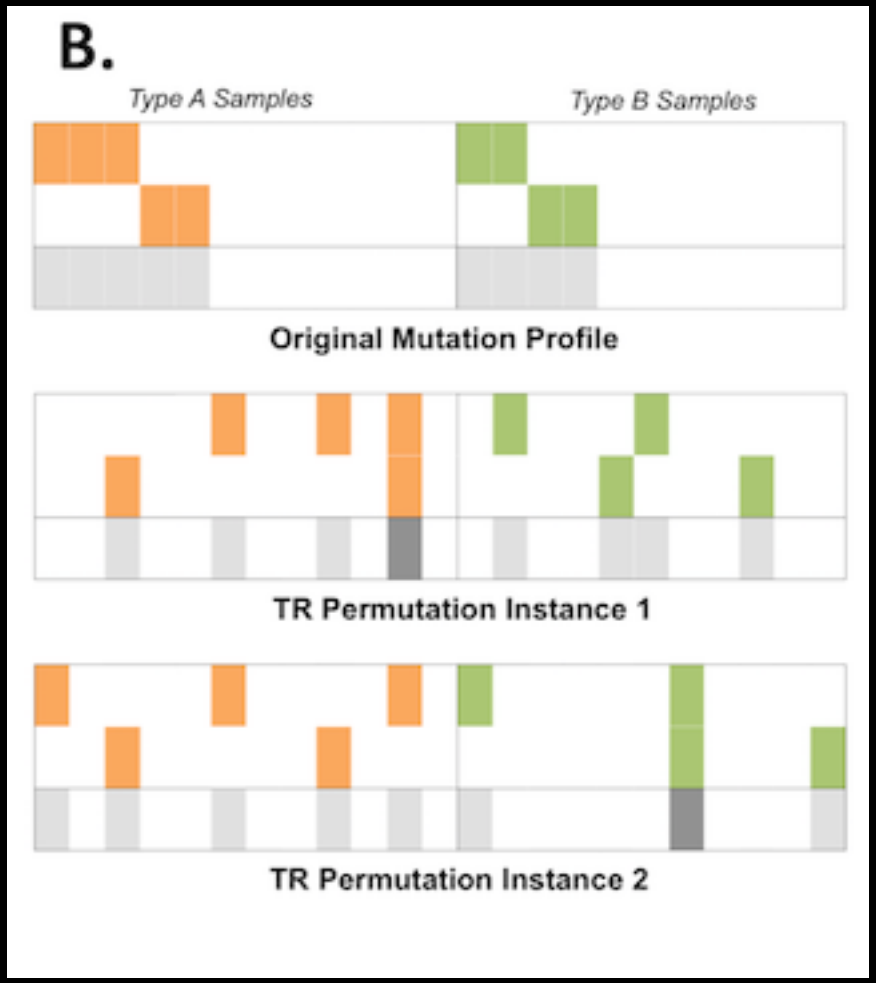
Type-restricted permutation test



- **Between tissues exclusivity**
  BETWEEN_ME

Permutation Test (across cancer type)

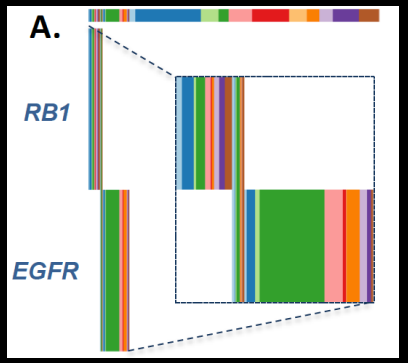
In each iteration, two (gene, sample) pairs are randomly chosen from the same cancer type and swapped



Introducing classification of mutual exclusivity

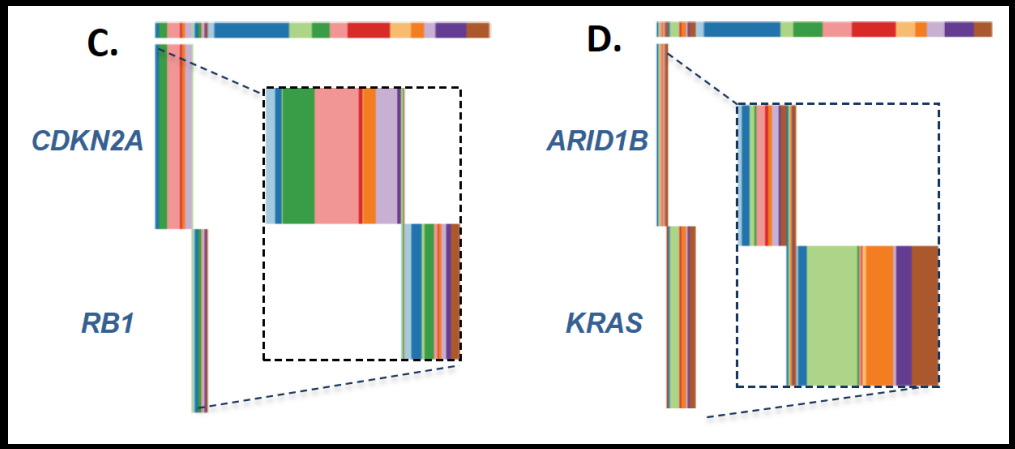
- **Within tissue exclusivity**
  WITHIN_ME

Traditional permutation test



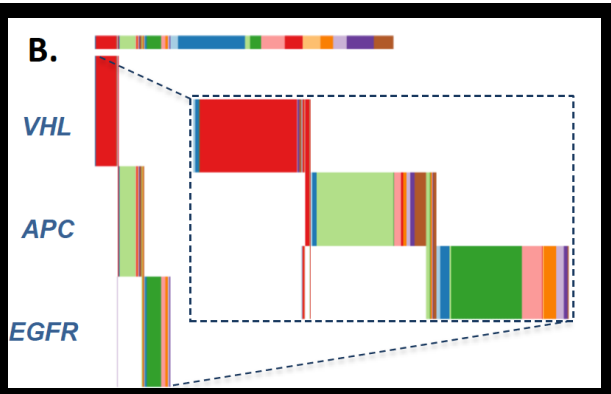
- **Across tissues exclusivity**
  ACROSS_ME

Type-restricted permutation test



- **Between tissues exclusivity**
  BETWEEN_ME

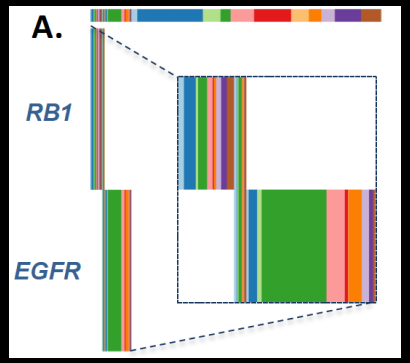
Type-oblivious permutation test



Introducing classification of mutual exclusivity

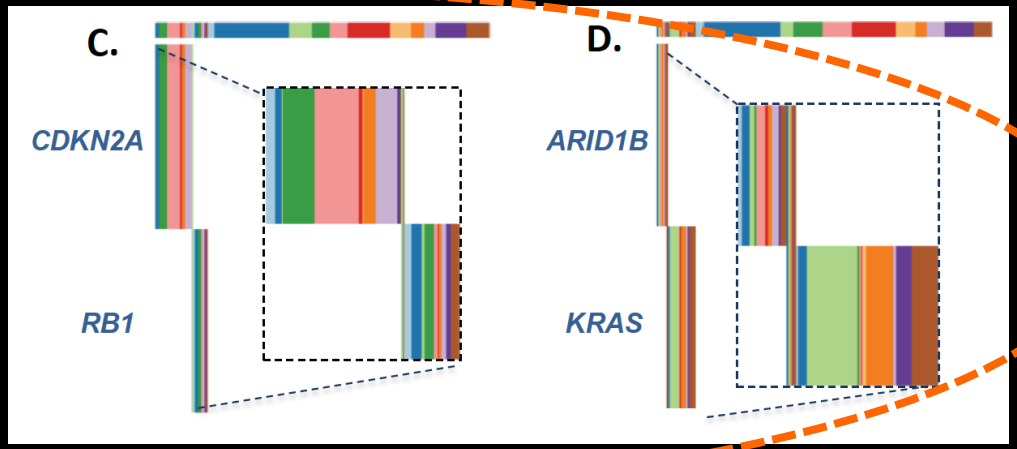
- **Within tissue exclusivity**
  WITHIN_ME

Traditional permutation test



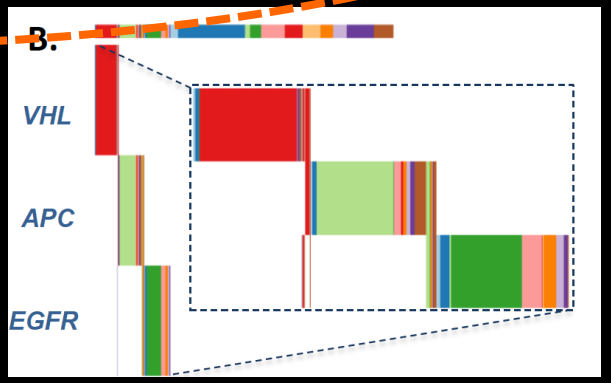
- **Across tissues exclusivity**
  ACROSS_ME

Type-restricted permutation test



- **Between tissues exclusivity**
  BETWEEN_ME

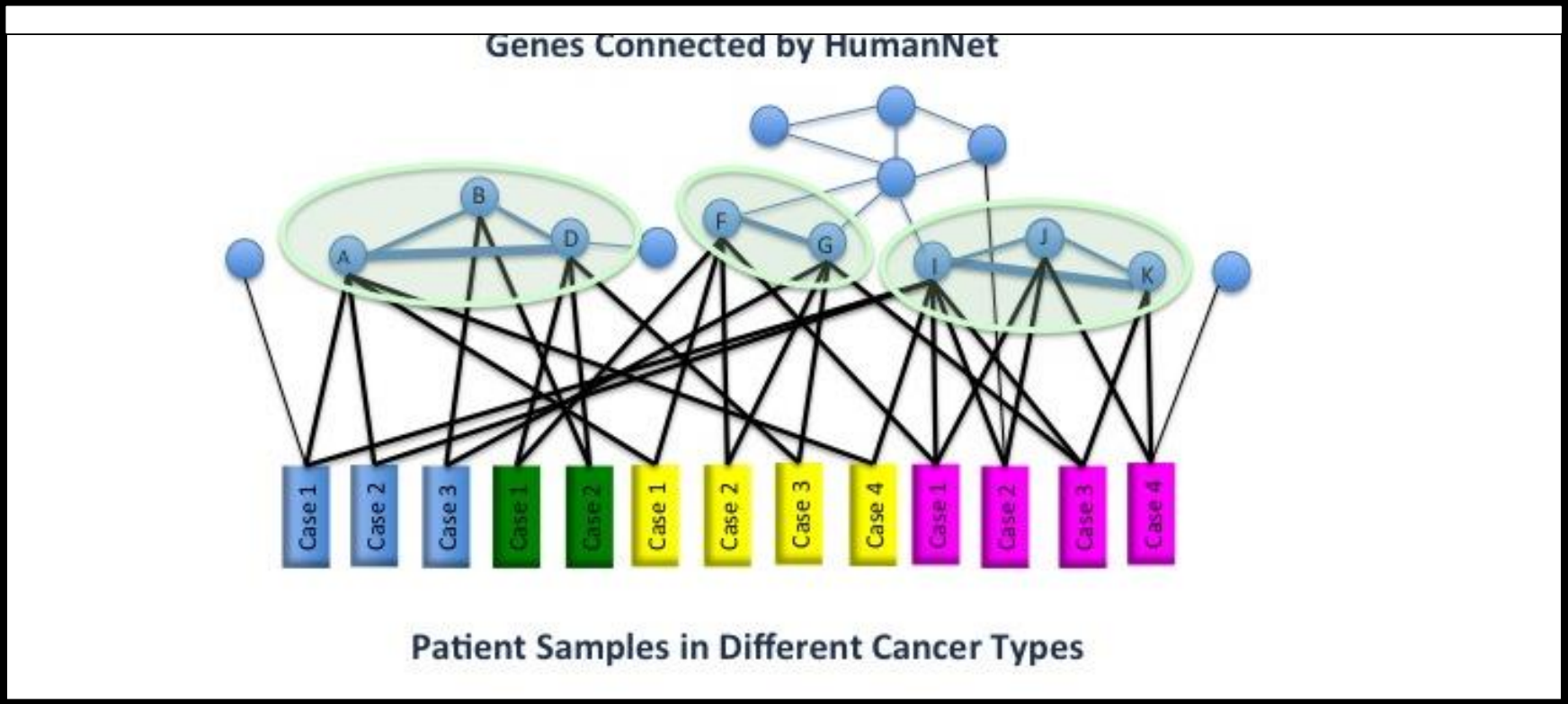
Type-oblivious permutation test



Finding cross-cancer dysregulated modules by combining interaction and ACROSS_ME

MEMCover – Mutual Exclusivity Module Cover

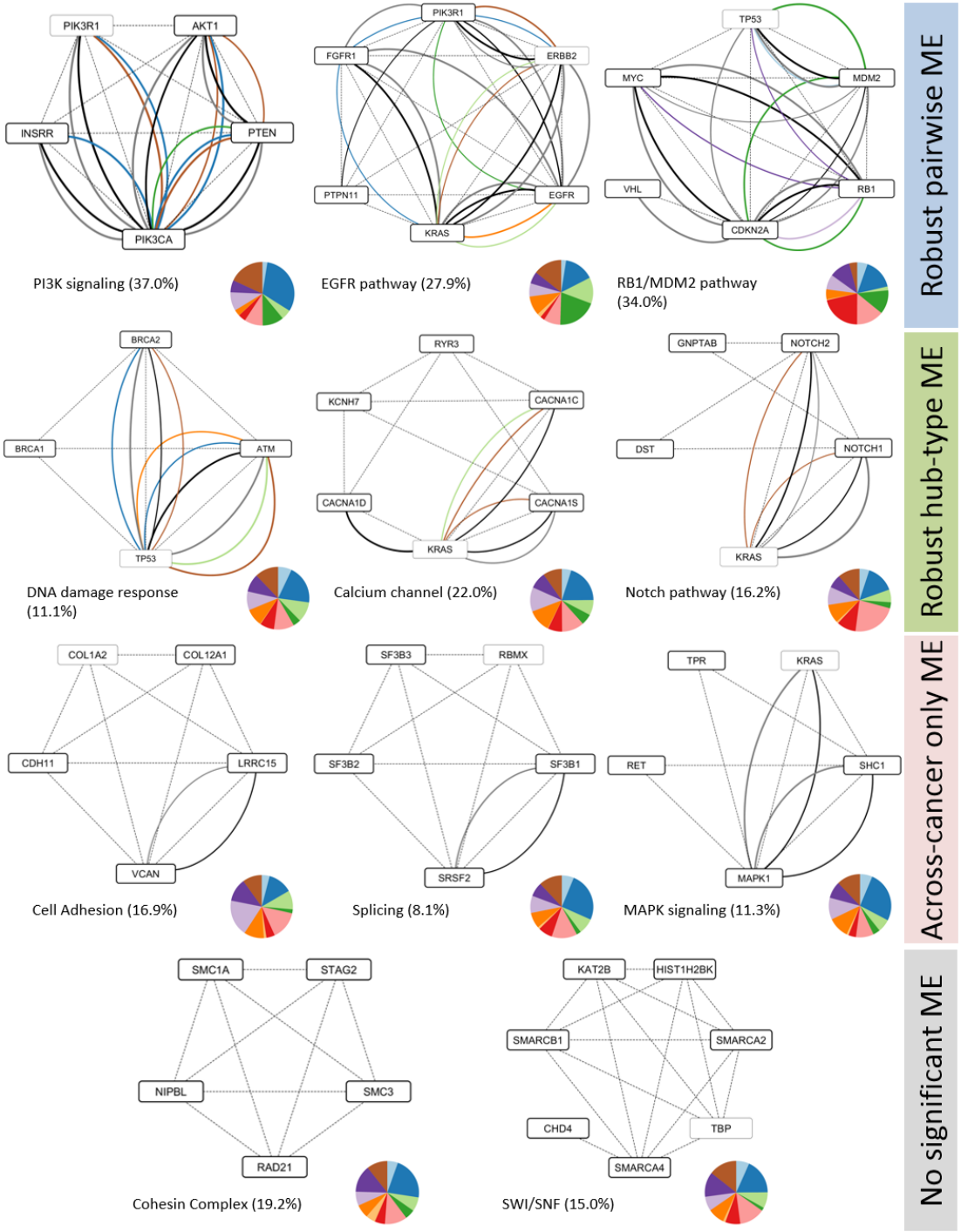
MEMCover Algorithm



Cost function considers:

- ↓ *edge confidence weights,*
- ↓ *ACROSS_ME scores,*
- ↑ *constant cost per module ,*
- ↓ *weight of covering edge.*

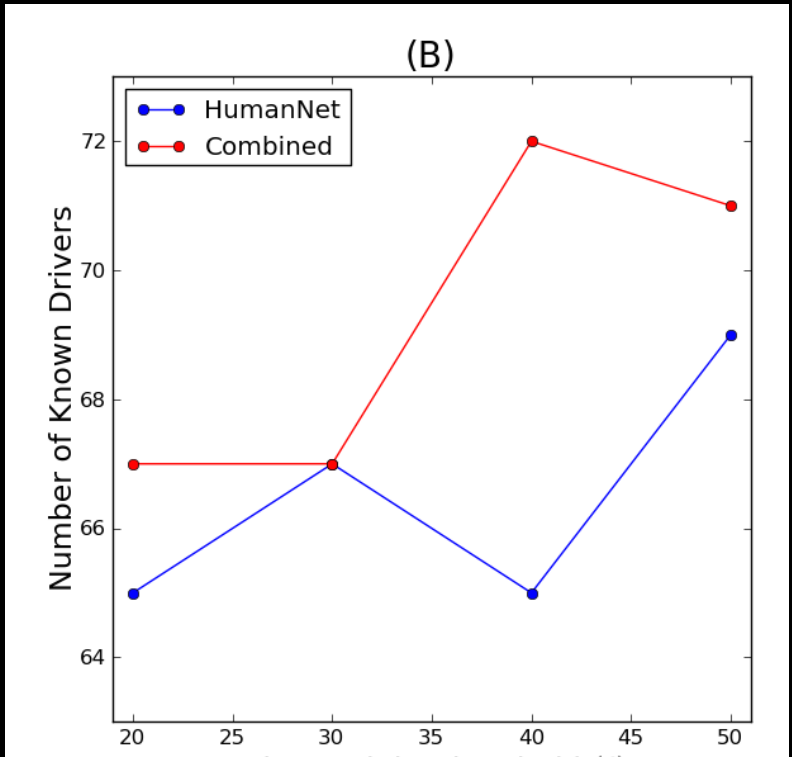
(to utilize scores given by some mutation calling programs)


PIK3R1
AKT1
INSRR
PTEN
PIK3CA
PI3K signaling (37.0%)
PIK3R1
FGFR1
ERBB2
PTPN11
EGFR
KRAS
EGFR pathway (27.9%)
TP53
MYC
MDM2
VHL
RB1
CDKN2A
RB1/MDM2 pathway (34.0%)
Robust pairwise ME
BRCA2
BRCA1
ATM
TP53
DNA damage response (11.1%)
RYR3
KCNH7
CACNA1C
CACNA1D
CACNA1S
KRAS
Calcium channel (22.0%)
GNPTAB
NOTCH2
DST
NOTCH1
KRAS
Notch pathway (16.2%)
Robust hub-type ME
COL1A2
COL12A1
CDH11
LRRC15
VCAN
Cell Adhesion (16.9%)
SF3B3
RBMX
SF3B2
SF3B1
SRSF2
Splicing (8.1%)
TPR
KRAS
RET
SHC1
MAPK1
MAPK signaling (11.3%)
Across-cancer only ME
SMC1A
STAG2
NIPBL
SMC3
RAD21
Cohesin Complex (19.2%)
KAT2B
HIST1H2BK
SMARCB1
SMARCA2
CHD4
TBP
SMARCA4
SWI/SNF (15.0%)
No significant ME

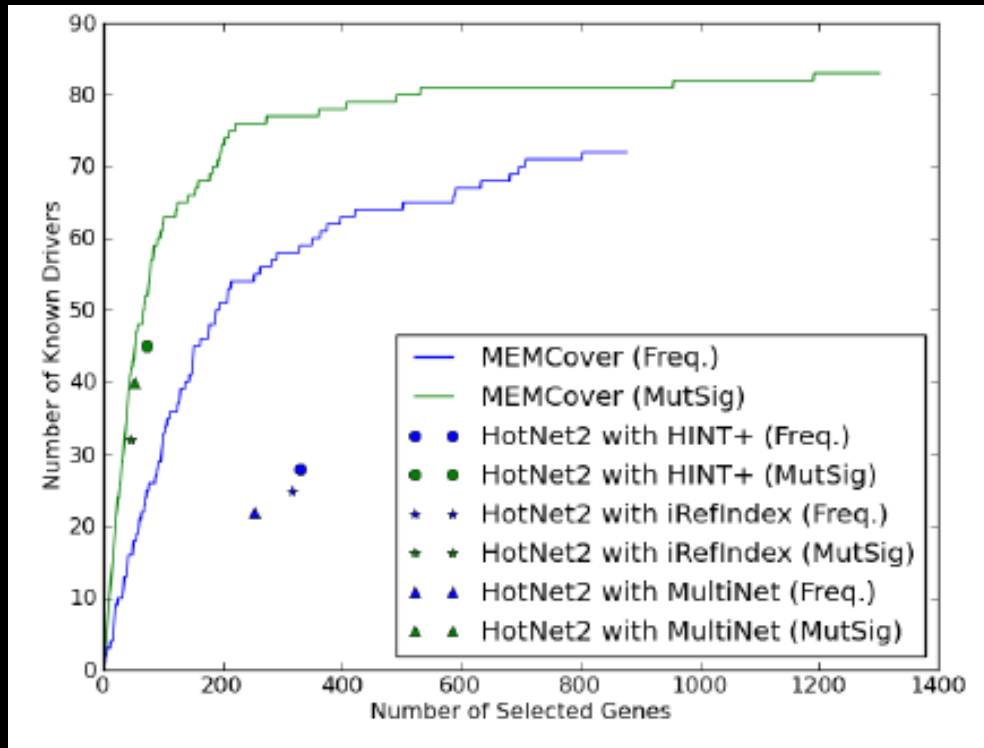
Does putting together ACROSS_ME and interaction data actually helps

MEMCover we find more cancer drivers

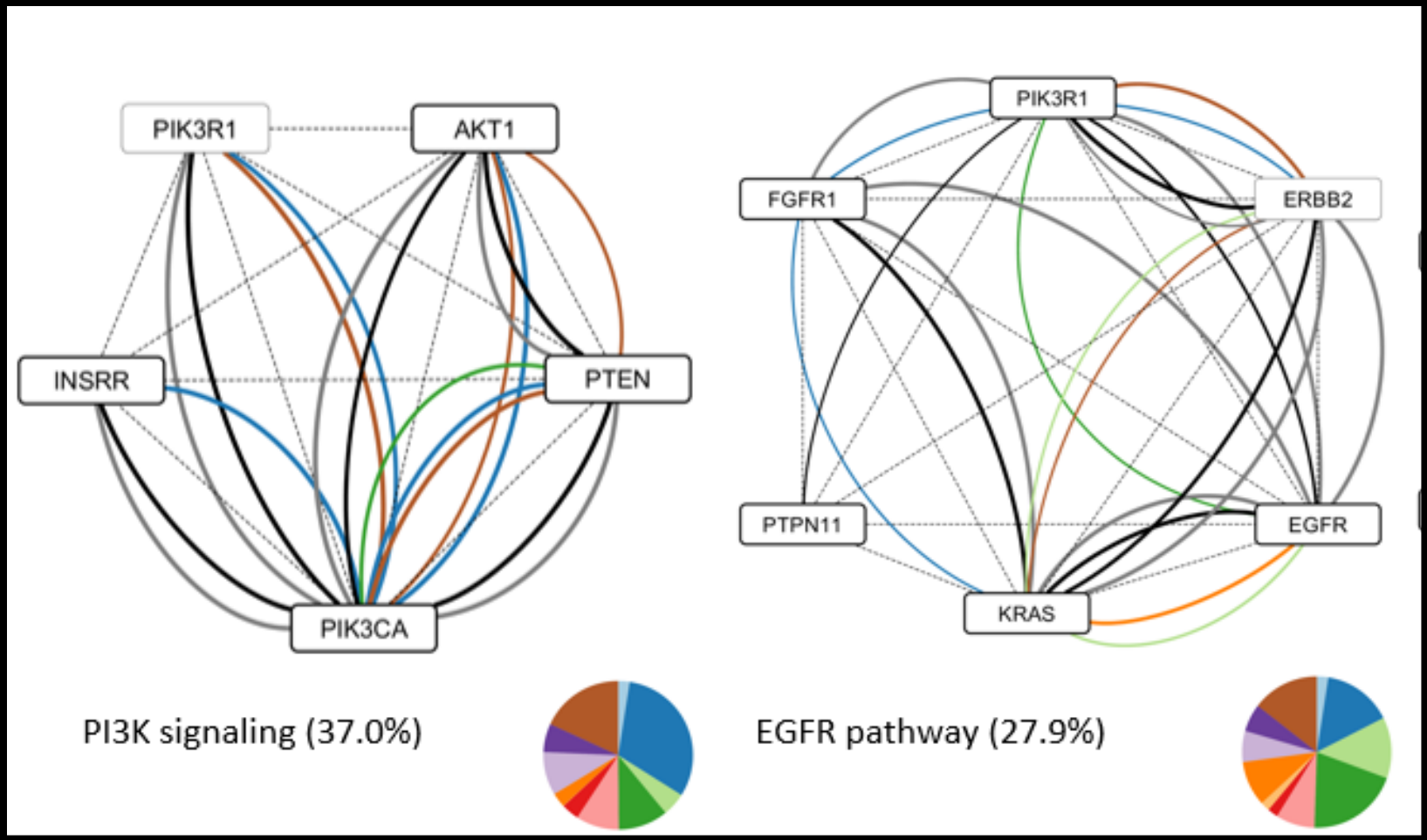
Compared to Module Cover



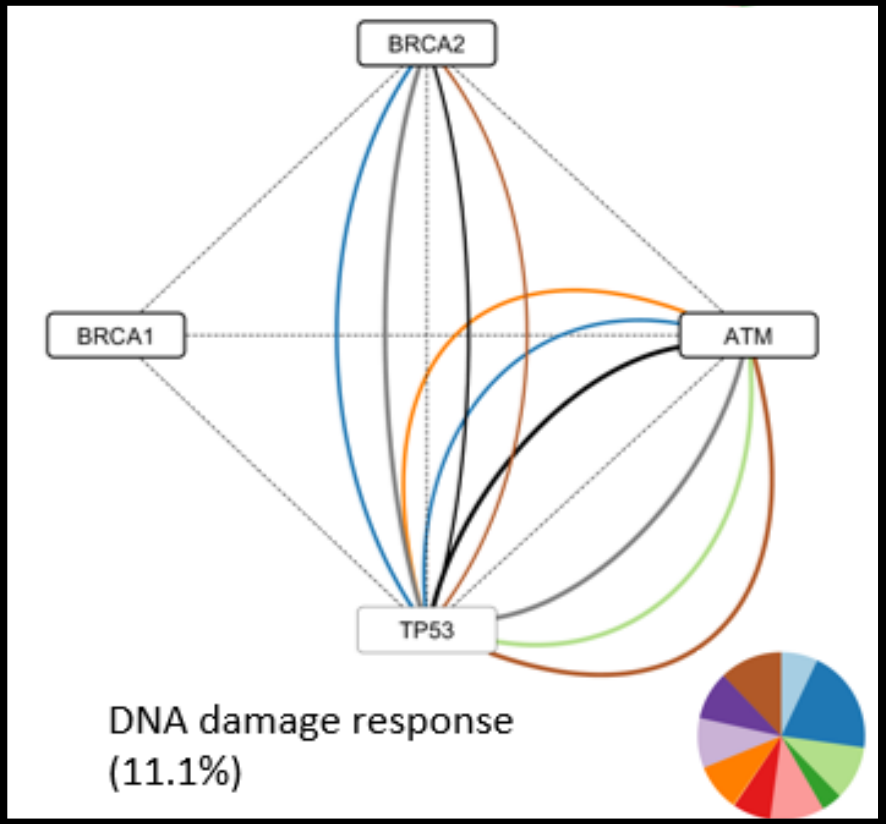
Compared to HotNet2



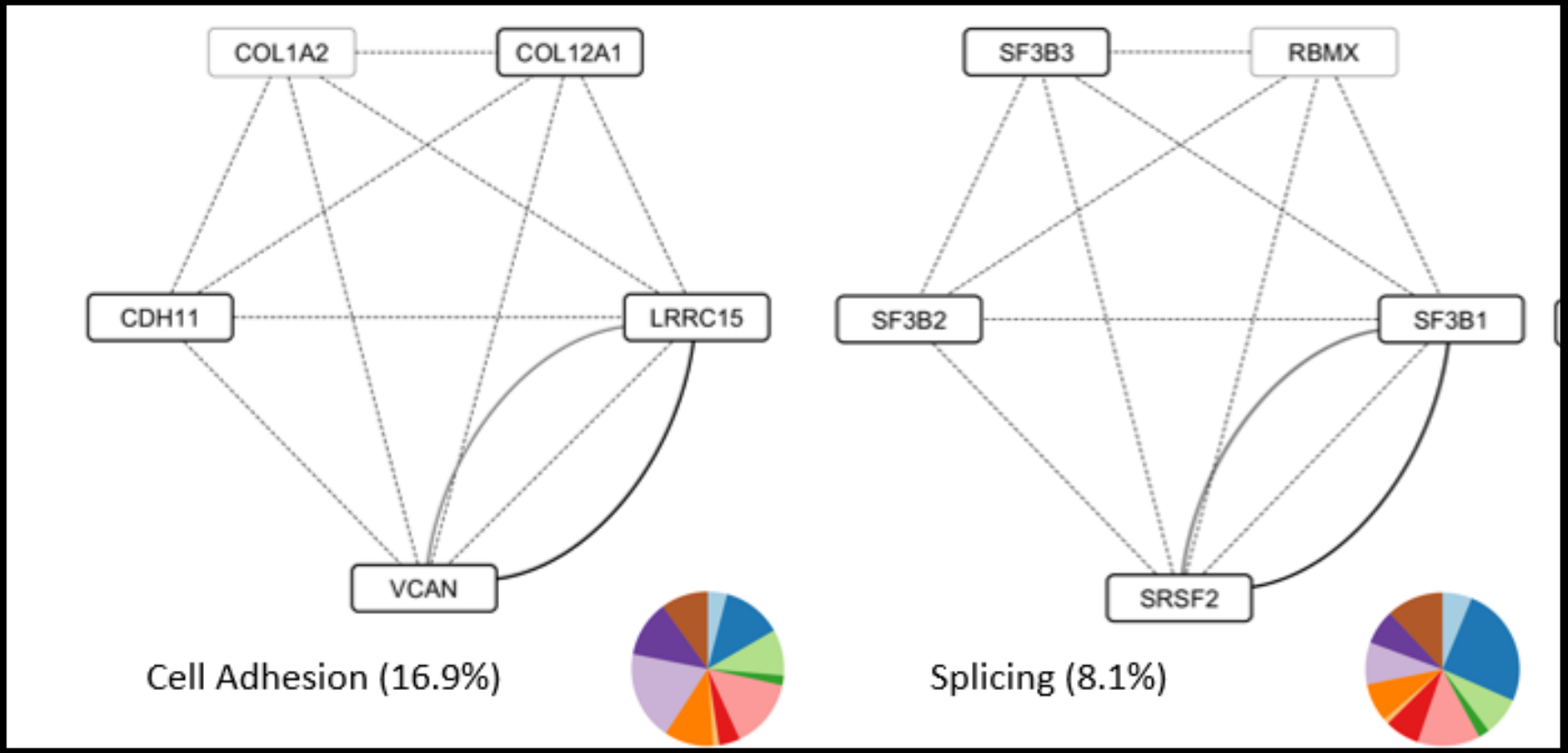
Robust mutual exclusivity within some modules



Hub-like ME within some modules



Across ME only within some modules



LETTER

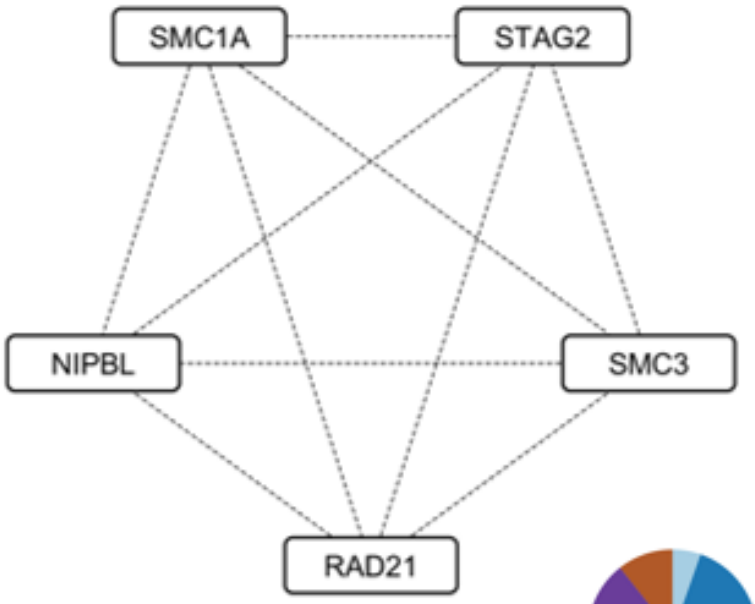
doi:10.1038/nature14985

The spliceosome is a therapeutic vulnerability in MYC-driven cancer

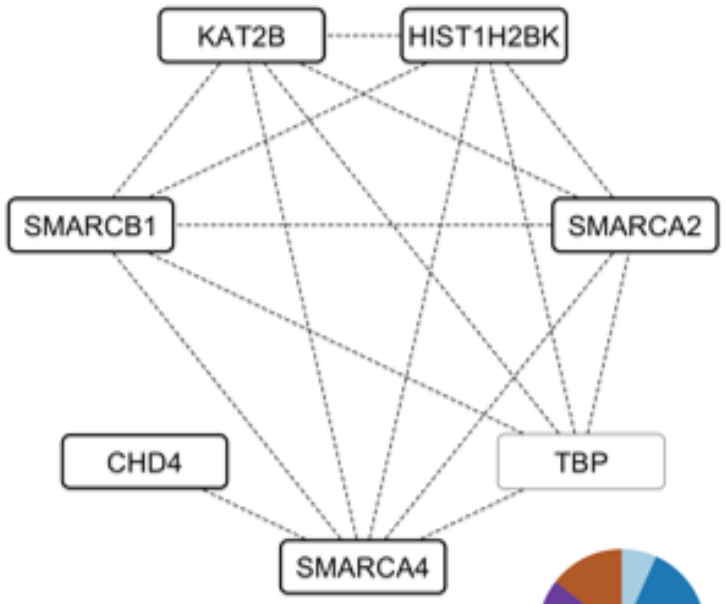
Tiffany Y.-T. Hsu[1,2,3,4], Lukas M. Simon[4], Nicholas J. Neill[1,4], Richard Marcotte[5], Azin Sayad[5], Christopher S. Bland[1,4], Gloria V. Echeverria[6,7,8], Tingting Sun[1,4], Sarah J. Kurley[1,4], Siddhartha Tyagi[1,4], Kristen L. Karlin[1,4], Rocio Dominguez-Vidaña[1,2,4], Jessica D. Hartman[4]†, Alexander Renwick[4], Kathleen Scorsone[9], Ronald J. Bernardi[9], Samuel O. Skinner[1,10], Antrix Jain[1], Mayra Orellana[1,4], Chandraiah Lagisetti[11], Ido Golding[1,10], Sung Y. Jung[1], Joel R. Neilson[2,6], Xiang H.-F. Zhang[12], Thomas A. Cooper[6,7,8], Thomas R. Webb[11], Benjamin G. Neel[5,13], Chad A. Shaw[4] & Thomas F. Westbrook[1,2,4]

- **But little ME between MYC and Spliceoseme**
- **Possible ME** between **Myc** and **SNRNP 200,** p-value < 0.03
- **ME** between **PIK3CA and SF3B4**, p-value 0

No ME within some modules



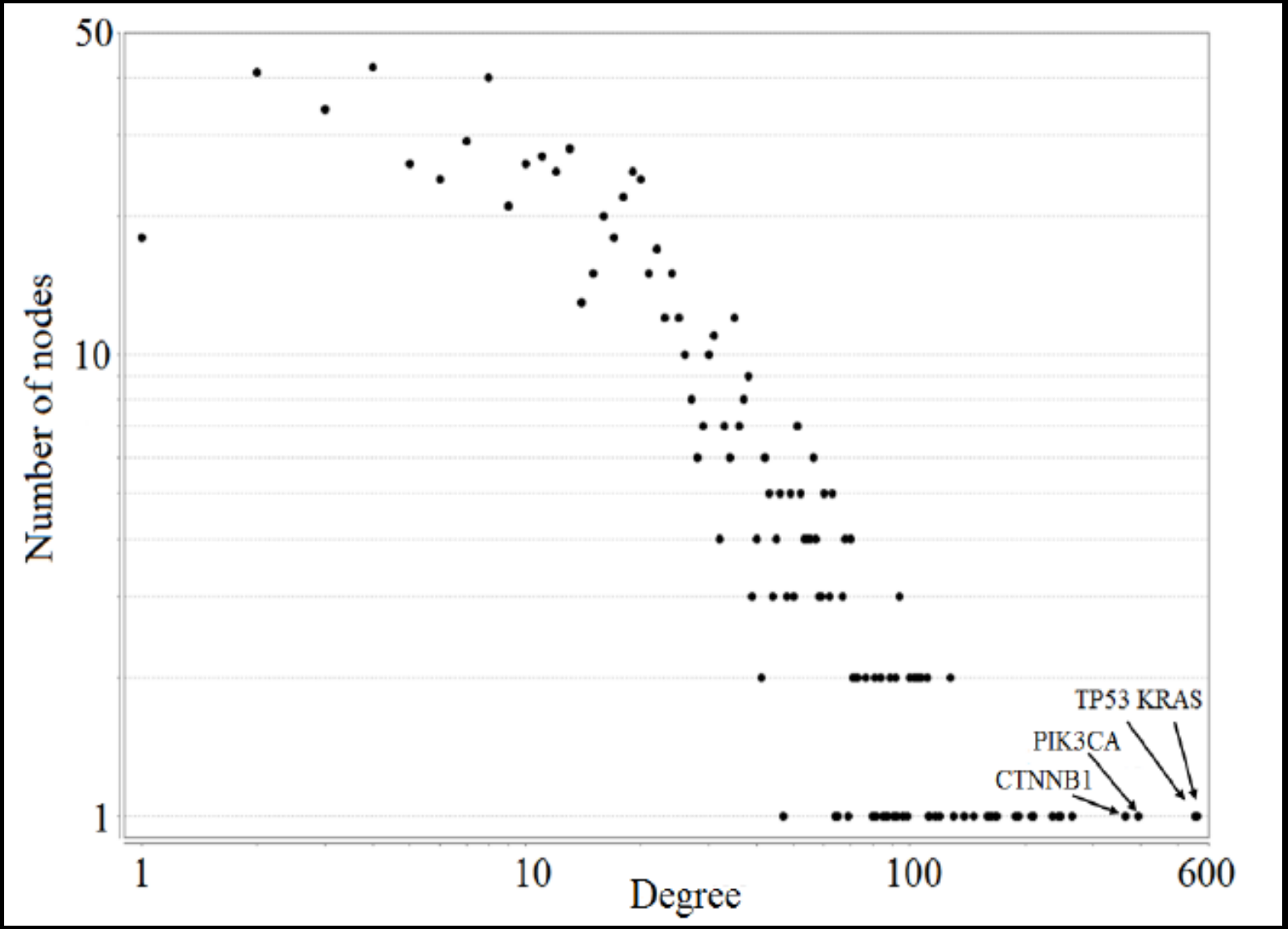
Cohesin Complex (19.2%)



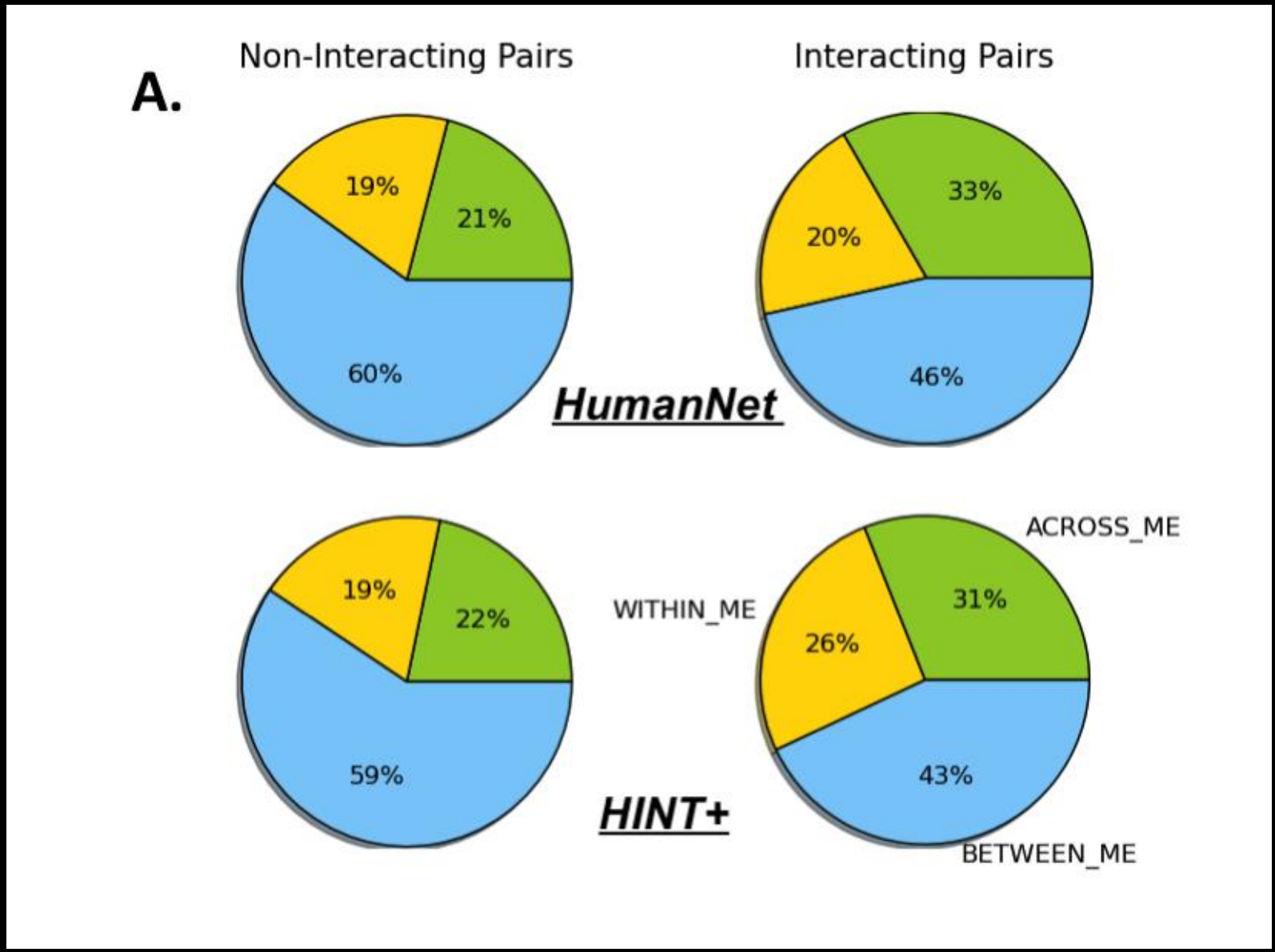
SWI/SNF (15.0%)

ME is not restricted to genes from same pathway

Mutual Exclusivity Hubs



ME relation and interactions

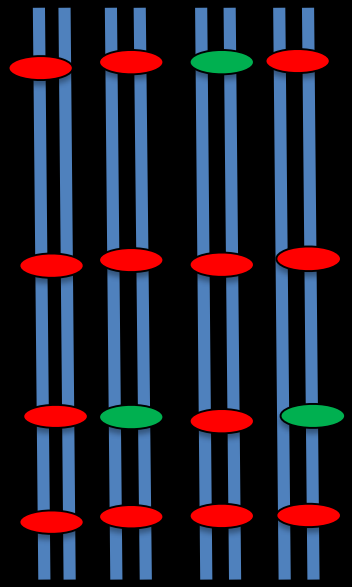


Summary

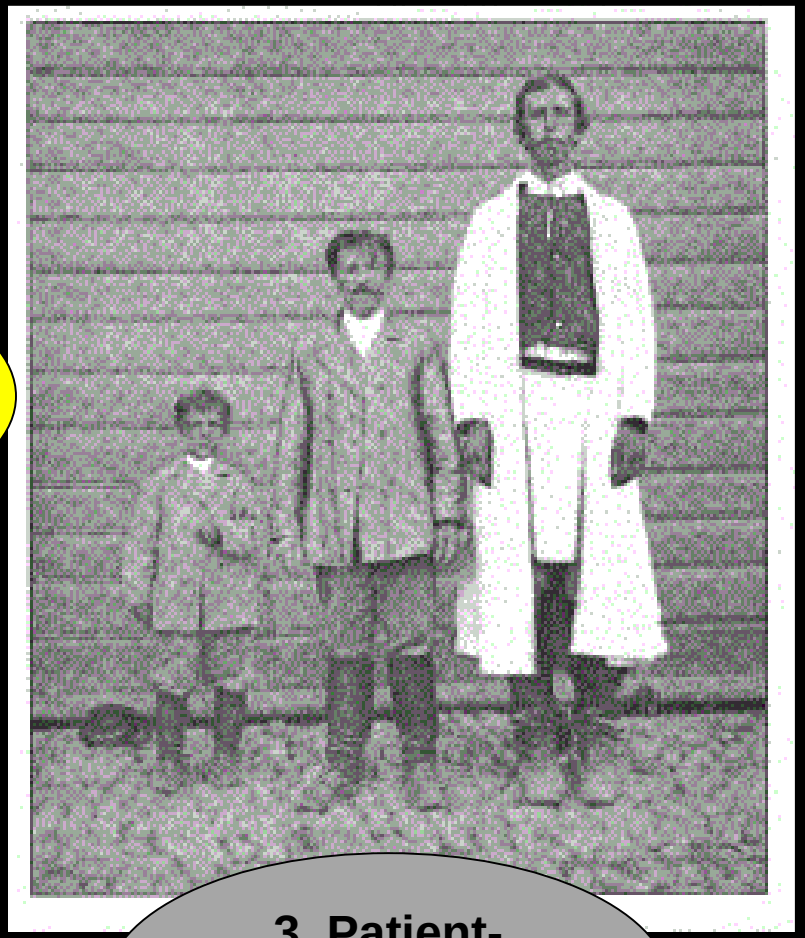
- *Combining ME with interaction network improves identification of PanCancer dysregulated modules*
- *While ME pairs are biased towards functionally interacting pairs but there is a lot of ME between non-interacting genes*
- Some dysregulated modules show no within module ME but show ME with genes from other pathways (inconsistent with Multi DENDRIX assumptions)
- *Mutual exclusivity hubs and are potent cancer drivers*

Genotypes

Phenotypes



2. Network based signal propagation

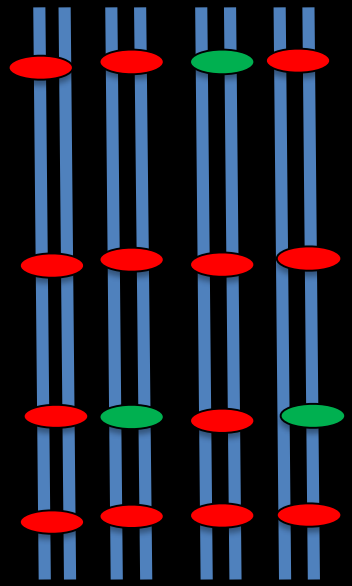


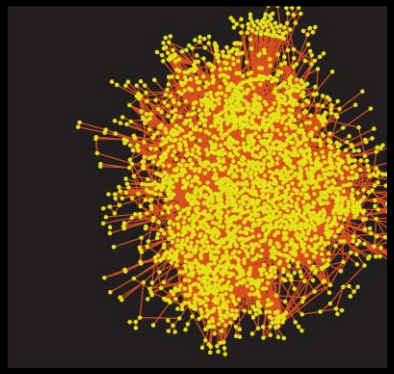
1. Dys- regulated Networks

3. Patient-similarity Networks

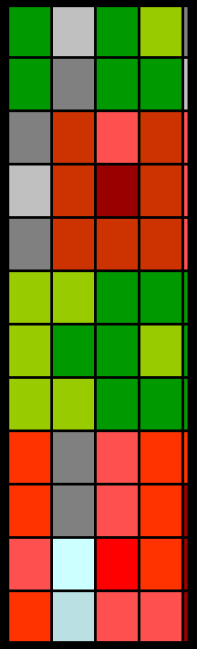
Information flow from genotypic changes to expression changes

Copy number aberrations or/and mutations





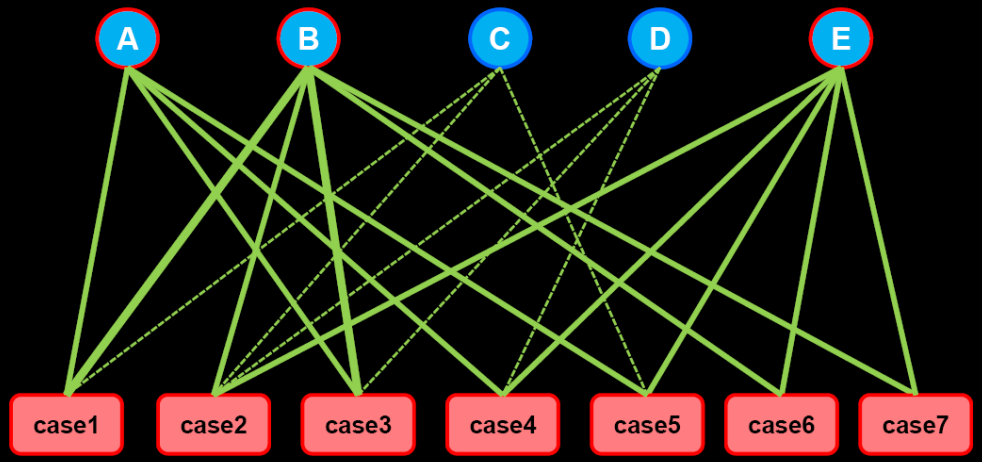
Gene expression



Kim et al. PolS CB 2011/RECOMB 2010

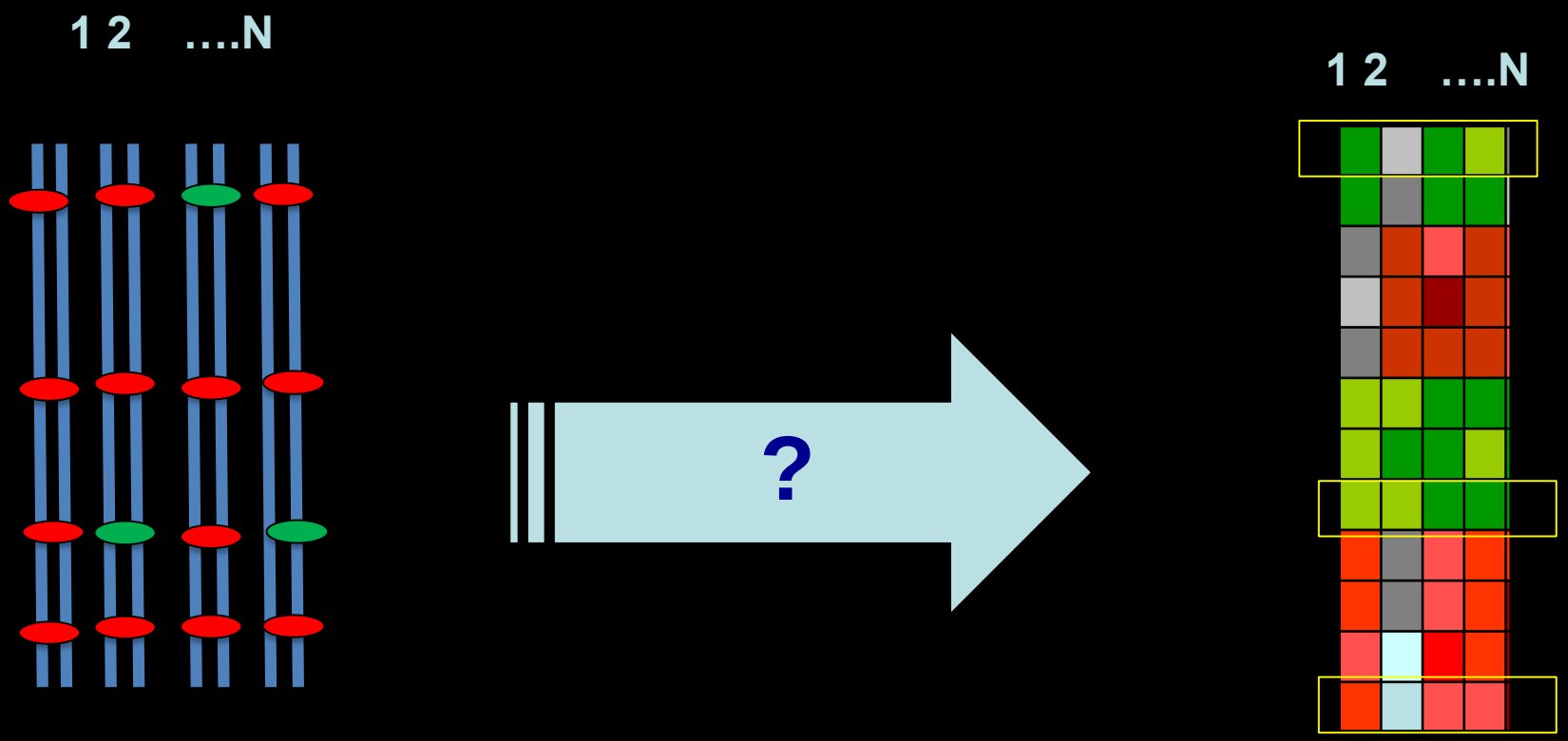
Selecting “signature” genes

Find smallest set of genes so that each case is “covered” (=over/under expressed)” at least specified number of times

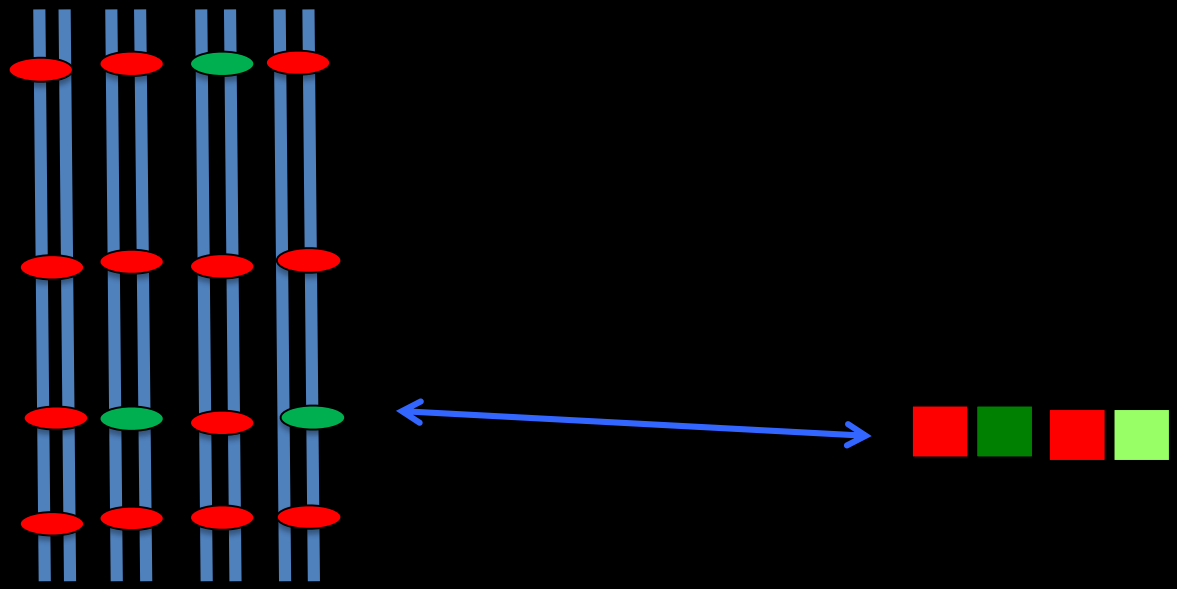


Kim et al. PolS CB 2011/RECOMB 2010

Explaining expression changes in the signature genes

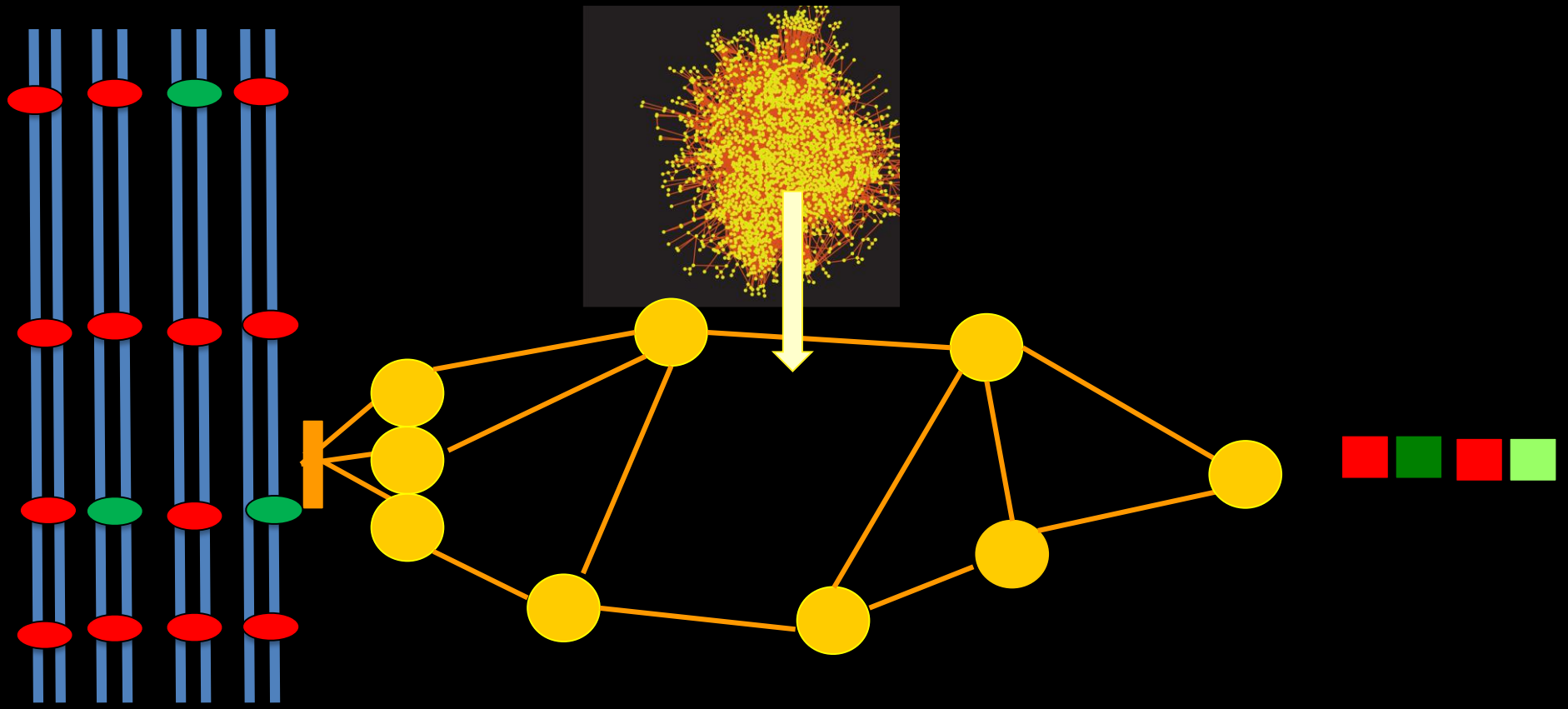


eQTL analysis links expression variability to genotypic variability



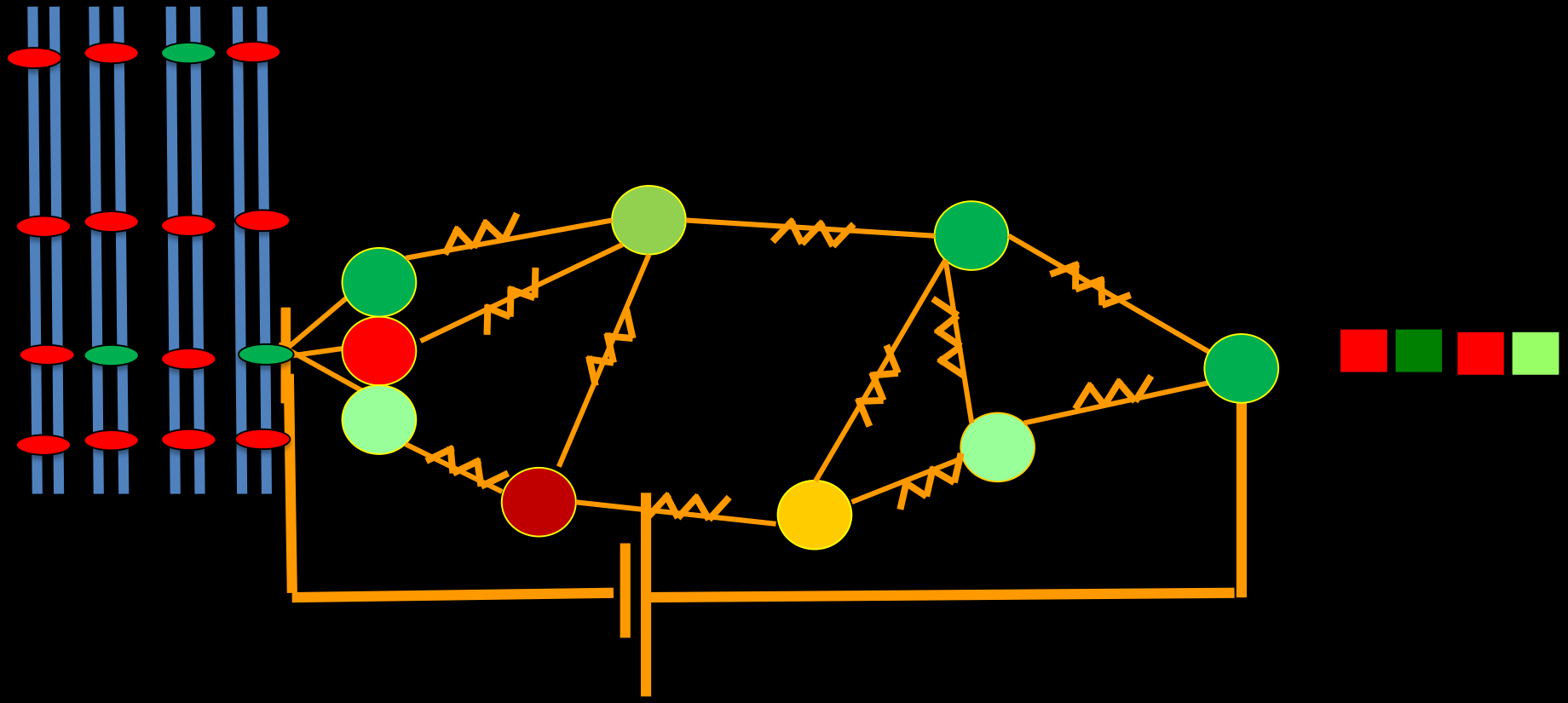
Tu *et al* Bioinfomatcis 2006
Suthram *et al* MSB 2008
Kim et al. PolS CB 2011/RECOMB 2010

Uncovering pathways of information flow between CNV and target gene



Tu *et al* Bioinfomatcis 2006
Suthram *et al* MSB 2008
Kim et al. PolS CB 2011/RECOMB 2010

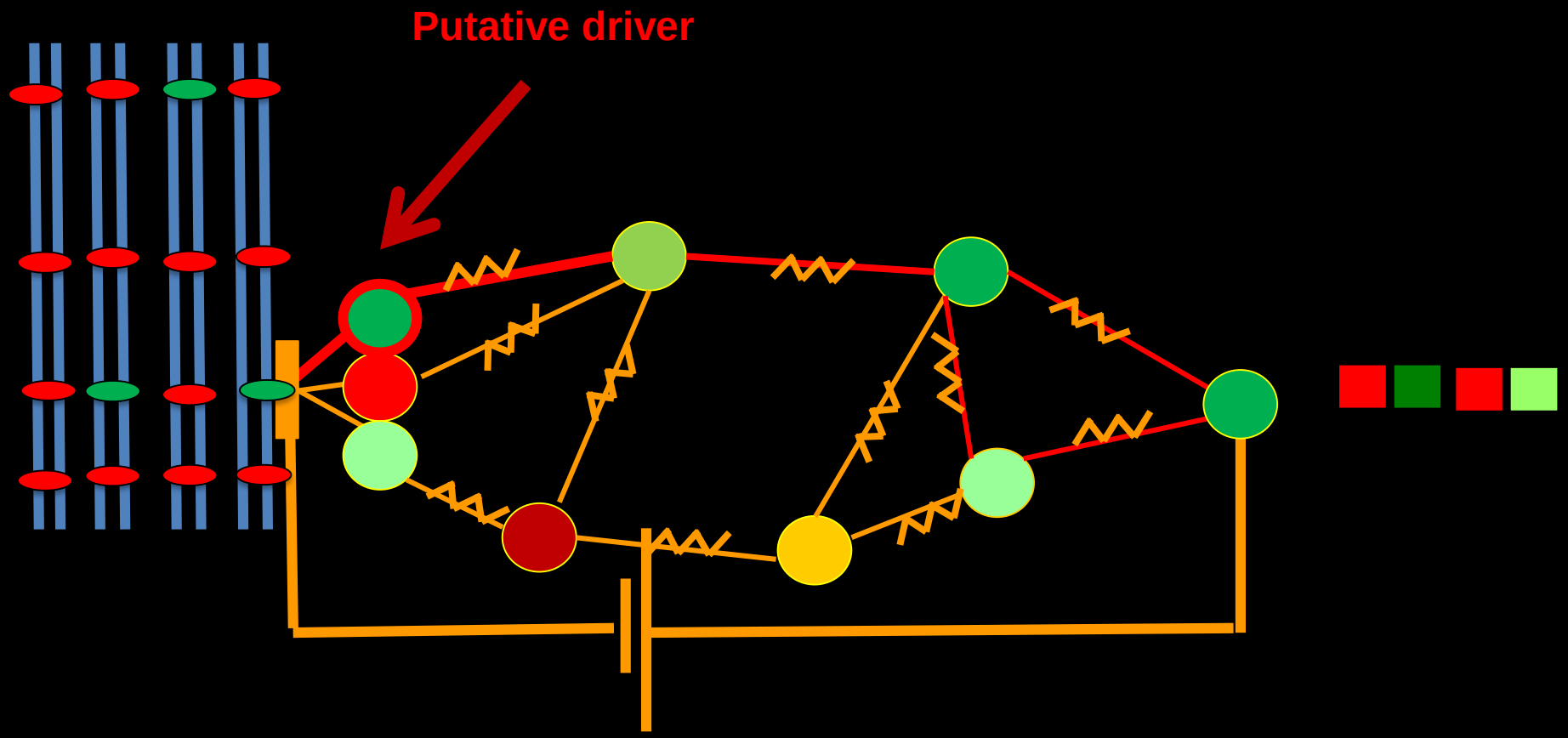
Adding resistances differentiate likelihoods of the edges



Resistance - set to favor most likely path -based on gene expression values

(reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)

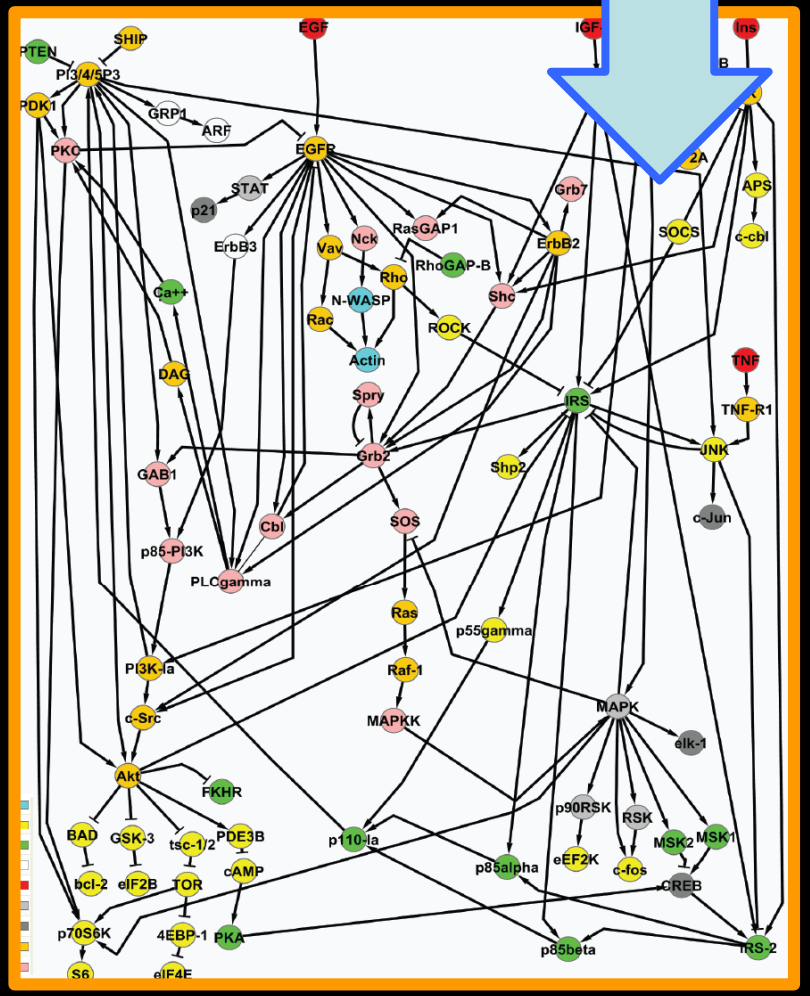
Finding subnetworks with significant current flow

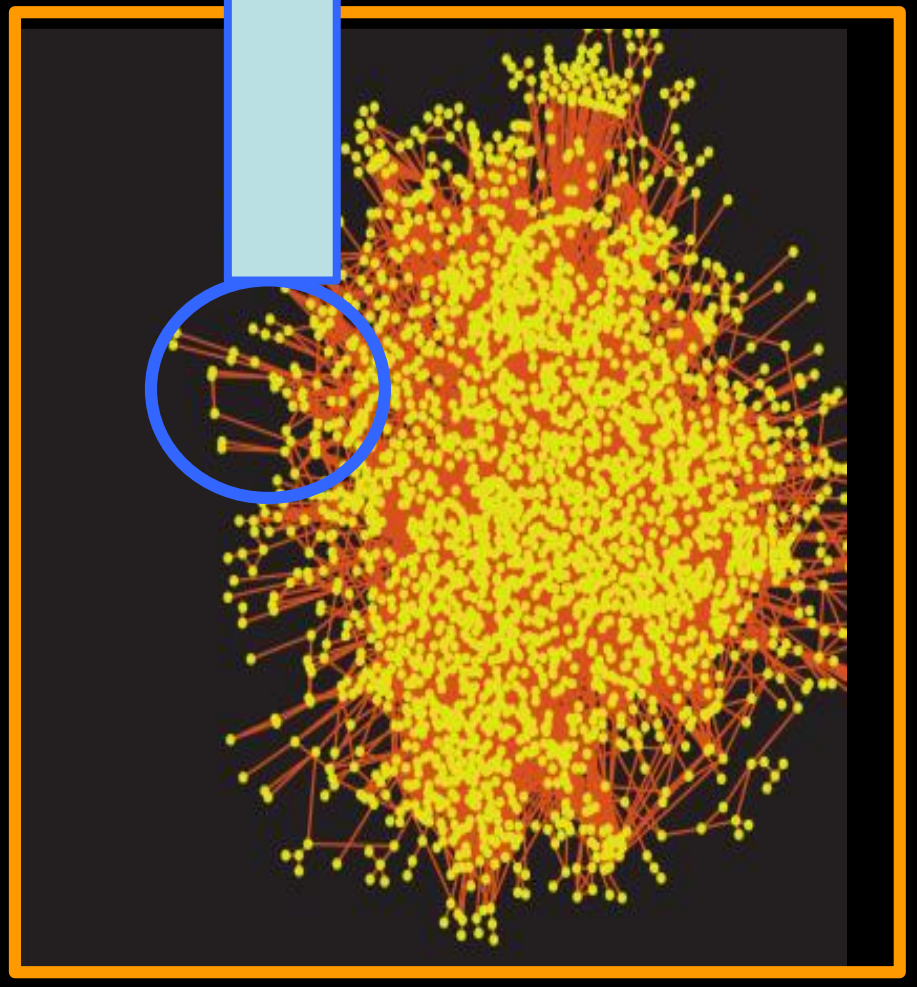


Resistance - set to favor most likely path -based on gene expression values
(reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)

Quest for interpretation

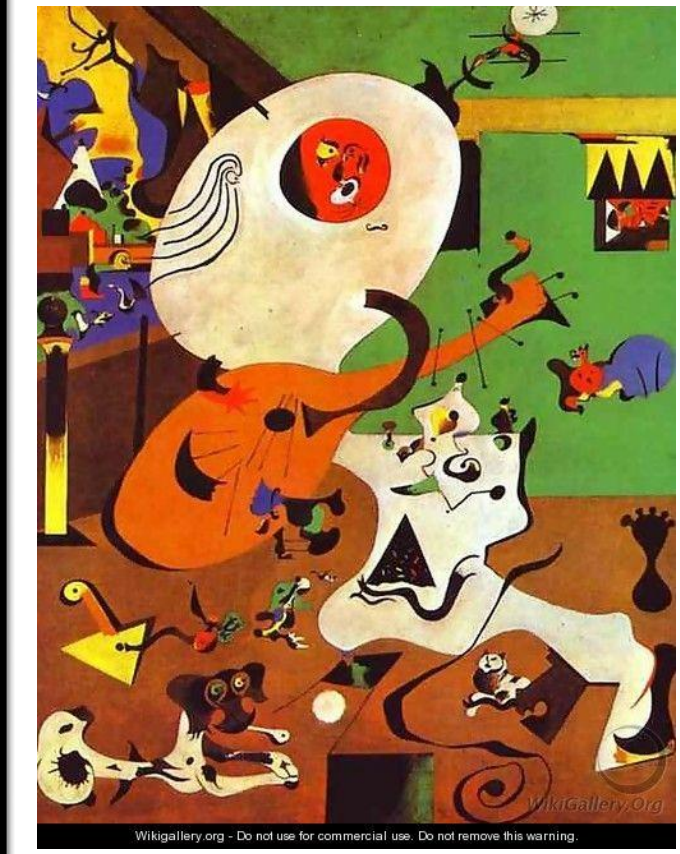
GO enrichment analysis



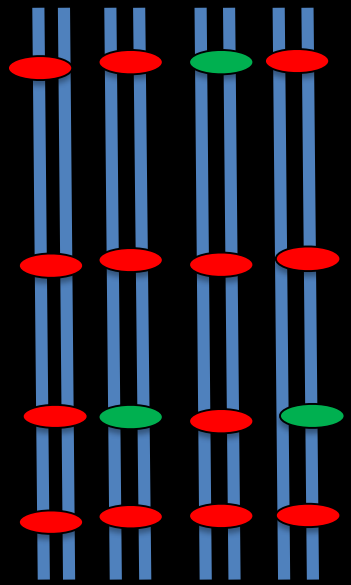


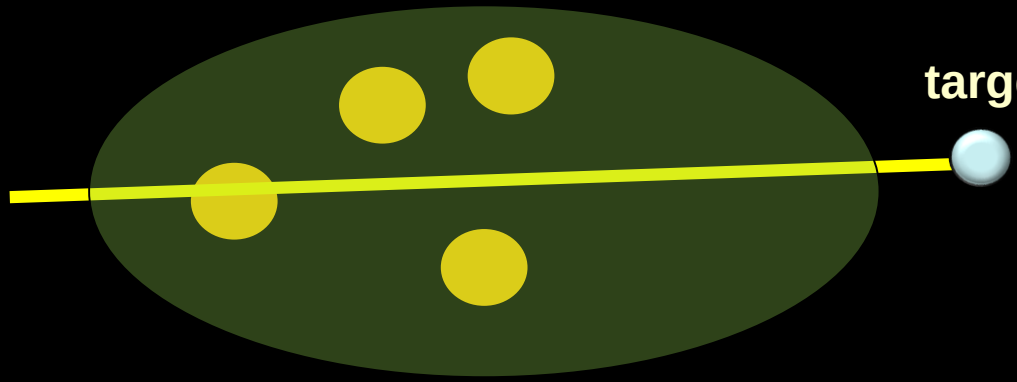
Quest for interpretation

The Lute Player, Hendrick Maertensz Sorgh (1610-1670),
Rijksmuseum, Amsterdam
(public domain)



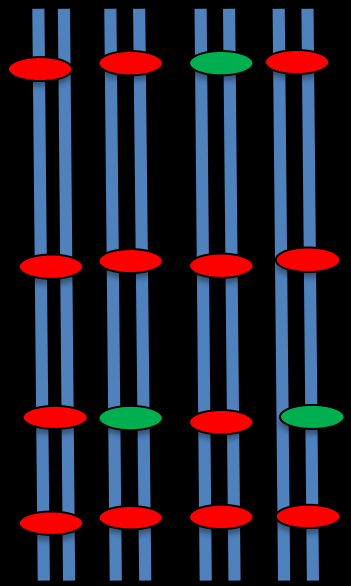
Dutch Interior 1, Joan Miro' (1893–1983)
Museum of Modern Art, New York

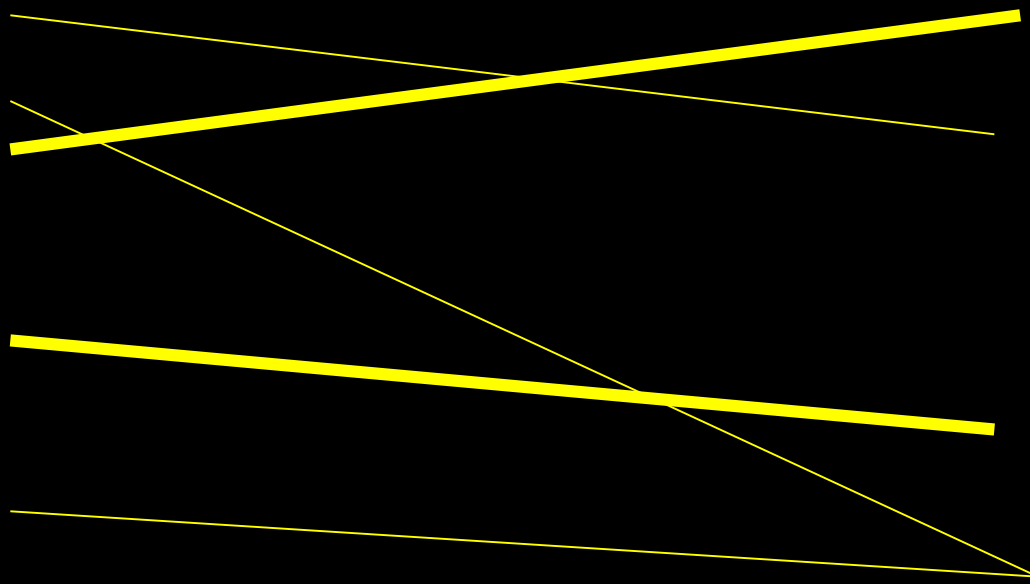


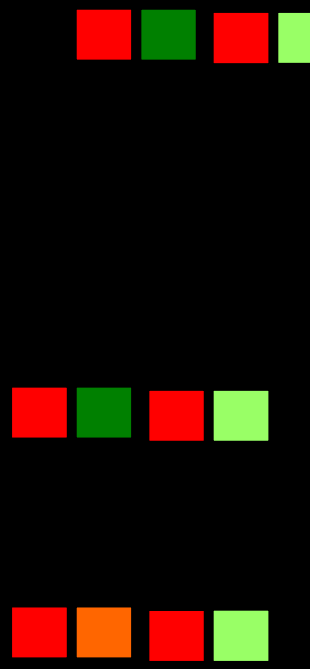


Kim et al. PolS CB 2011/RECOMB 2010

Repeat for other genes and significantly associated loci





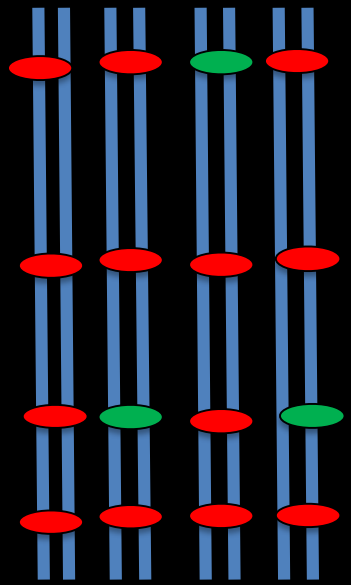


Cancer Cases
CNV data

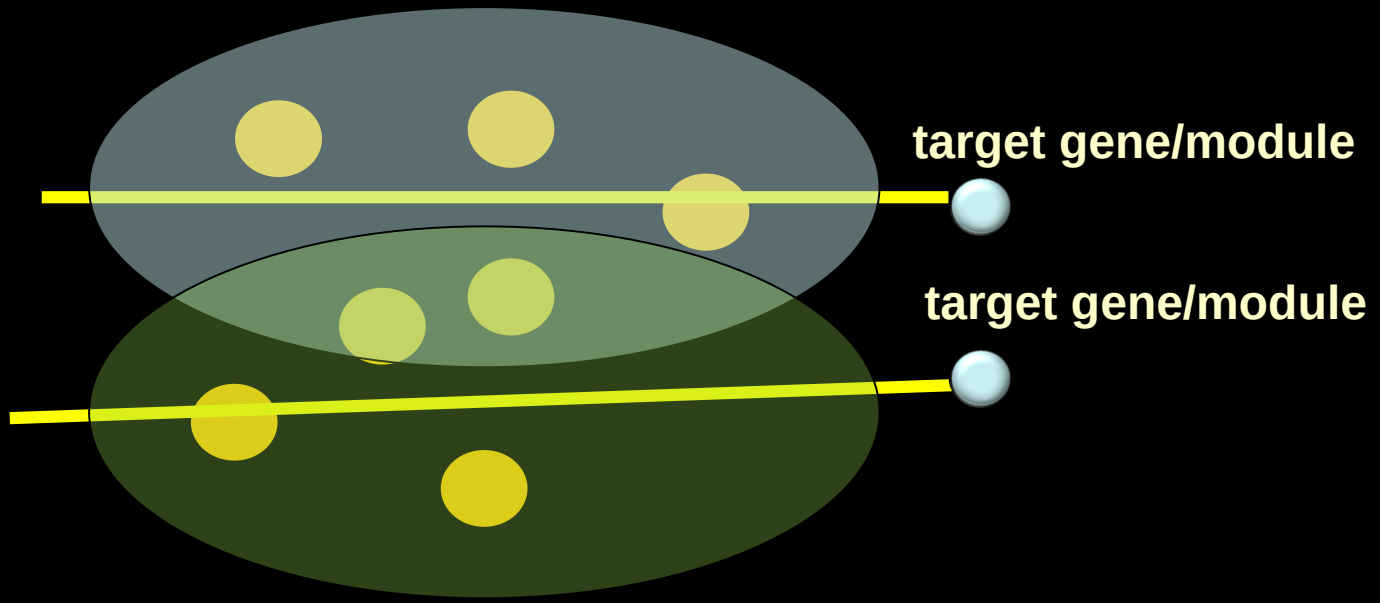
Cancer Cases
Gene expression data

Are there common functional pathways?

Cancer Cases
CNV data



Cancer Cases
Gene expression data

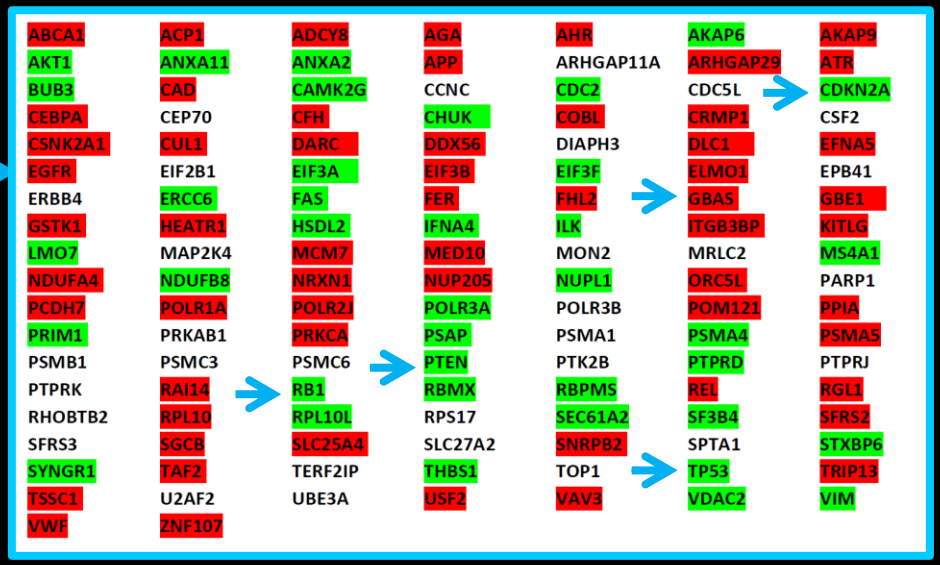


Kim et al. PolS CB 2011/RECOMB 2010

Gene Hubs

MYC(110)	E2F1(88)	E2F4(43)	CREBBP(34)	GRB2(27)	SP3(26)	ESR1(25)
TFAP2A(25)	NFKB1(23)	MYB(22)	JUN(22)	E2F2(22)	RELA(21)	AR(21)
SP1(20)	RPS27A(20)	MAPK3(19)	POU5F1(17)	HIF1A(16)	PPARA(15)	CDC42(15)
UBA52(13)	CDK7(13)	YBX1(13)	YWHAZ(12)	CEBPB(12)	POU2F1(12)	UBE2I(11)
SMAD3(11)	TAL1(11)					

Driving Copy number aberrations

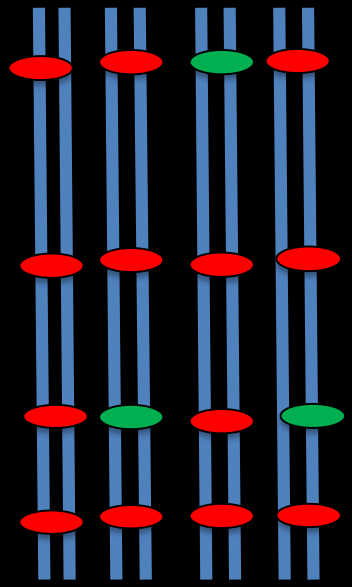


Pathway Hubs

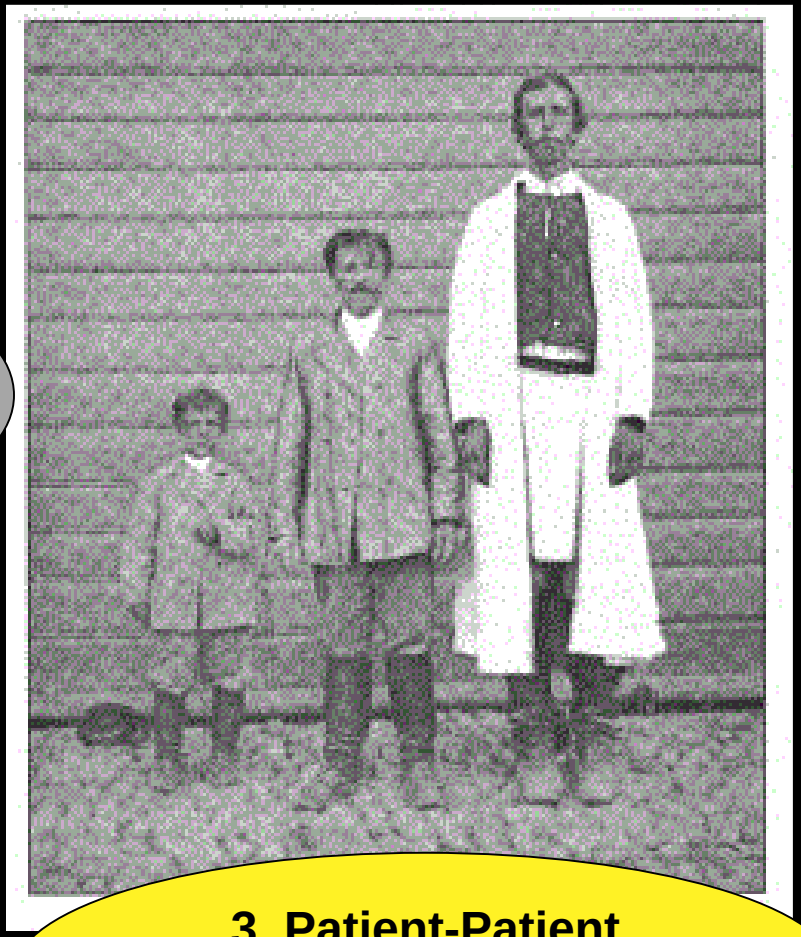
GO biological process	#
cell cycle arrest	10
epidermal growth factor receptor signaling pathway	9
negative regulation of cell growth	9
Ras protein signal transduction	9
regulation of sequestering of triglyceride	8
cell proliferation	7
nuclear mRNA splicing, via spliceosome	7
regulation of cholesterol storage	7
nucleotide-excision repair	7
RNA elongation from RNA polymerase II promoter	7
insulin receptor signaling pathway	6
transcription initiation from RNA polymerase II promoter	6
N-terminal peptidyl-lysine acetylation	5
phosphoinositide-mediated signaling	5
positive regulation of lipid storage	4
positive regulation of specific transcription from RNA polymerase II promoter	3
positive regulation of epithelial cell proliferation	3
base-excision repair	2
negative regulation of hydrolase activity	2
gland development	2
positive regulation of MAP kinase activity	2
regulation of nitric-oxide synthase activity	2
estrogen receptor signaling pathway	2
regulation of receptor biosynthetic process	2
response to organic substance	2
JAK-STAT cascade	2
regulation of transforming growth factor-beta2 production	2
G1/S transition of mitotic cell cycle	2
SMAD protein nuclear translocation	2

Genotypes

Phenotypes



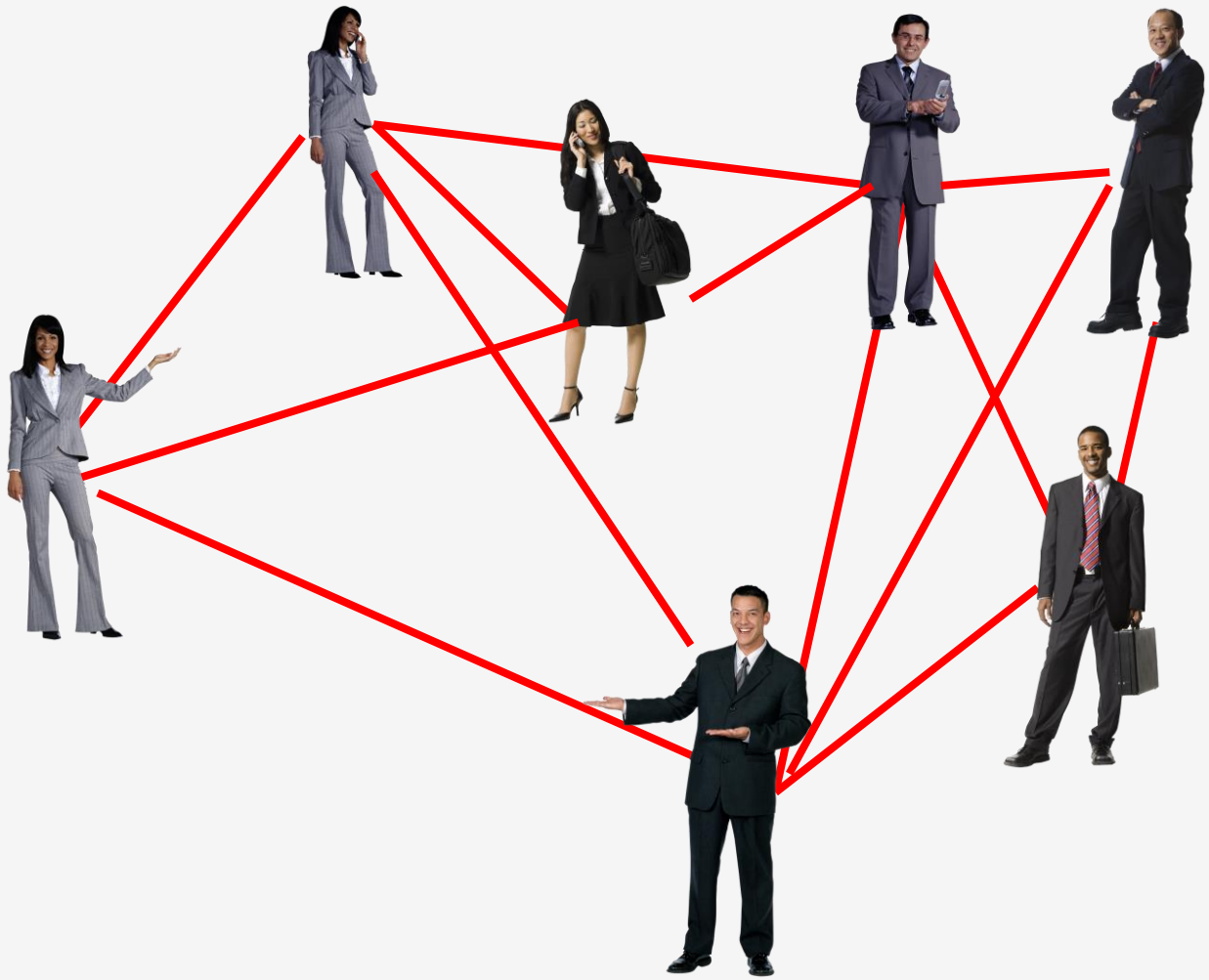
2. Network based signal propagation



1. Dys- regulated Networks

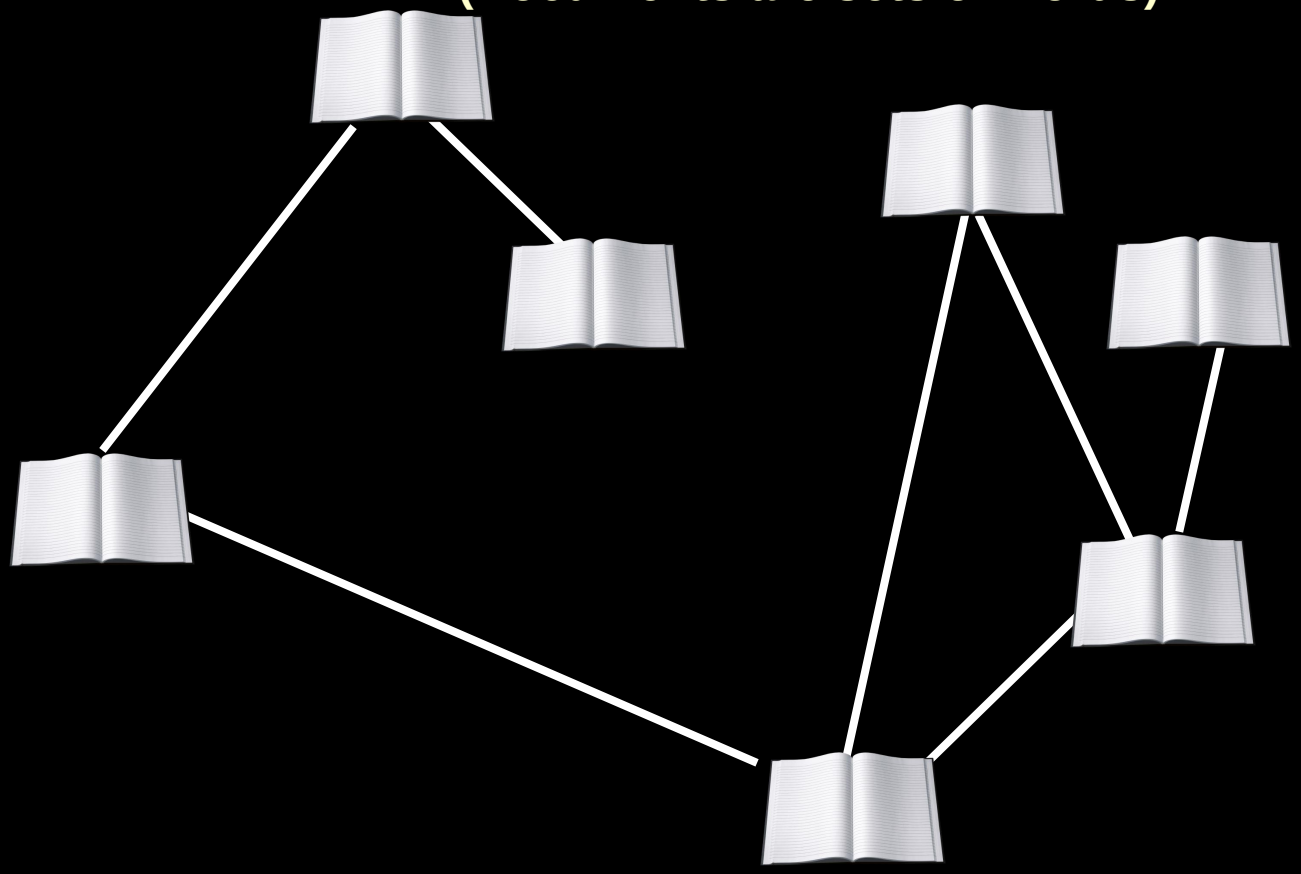
3. Patient-Patient similarity Networks

Phenotype similarity network



Document similarity network

(Documents are sets of words)



Chang J, Blei DM: Hierarchical Relational Models for Document Networks. Ann Appl Stat 2010, 4(1):124-150.

Topic Model to divide documents into topics

Topic I

president 0.45
senate 0.37
legislation 0.21
….

Topic II

police 0.51
Accident 0.29
search 0.17
….

Topics III

Competition 0.38
Track 0.22
Runner 0.14
….

Topic IV

Audience 0.27
Play 0.15
Actor ….

Chang J, Blei DM: Hierarchical Relational Models for Document Networks. Ann Appl Stat 2010, 4(1):124-150.

Phenotypic versus explanatory features

Phenotypic features (looks) :

Survival time
Response to drugs,…..
Gene expression profile

Explanatory features (words)

- mutations, CNV, micro RNA level;
- Epigenetic factors,
- Sex, age, environment ….

Key idea

neighbors in patient network should have similar explanatory features

Based on patient's features represent each patient as mixture of the subtypes

Subtype I

EGFR_A 0.45
NF1_M 0.37
PTEN_A 0.21
….

Subtype II

PDGFA_A 0.51
IDH1_M 0.29
M53_M 0.17
….

Subtype III

mirR218_H 0.38
ICDK2_D 0.22
SHC1_M 0.14
….

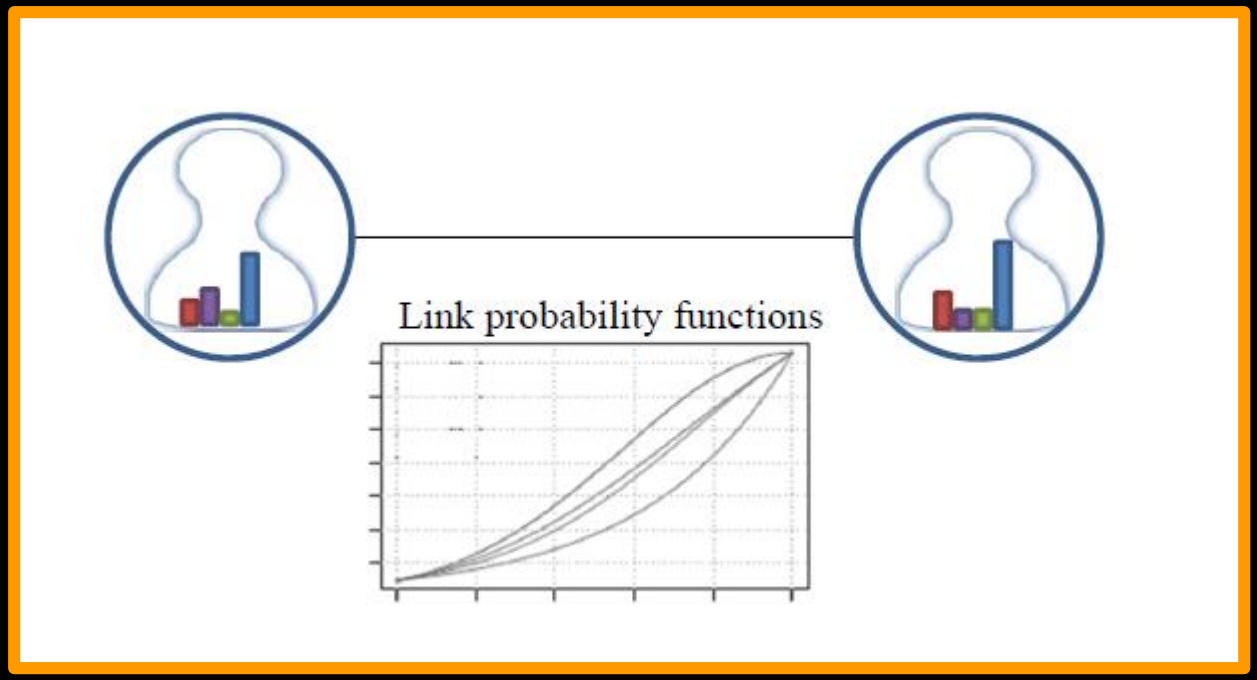
Features:
EGFR_A
NF1_M
CDKN2B_D
.
.
;

Subtype IV

CDK2B_D 0.37
EGFR_A 0.25
….

Cho et al. NAR 2013/RECOMB 2012

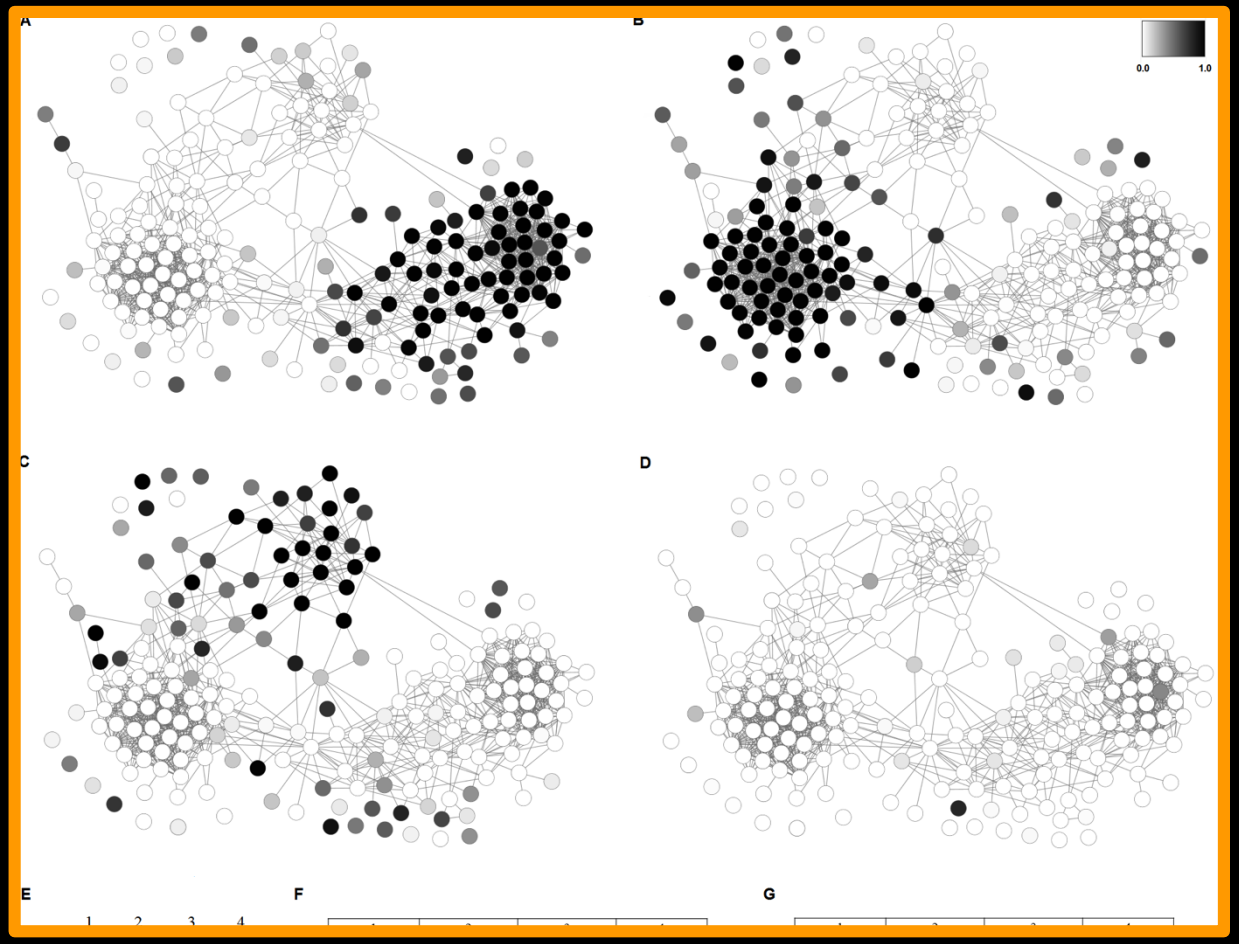
Generate edges based on similarity of subtype mixtures

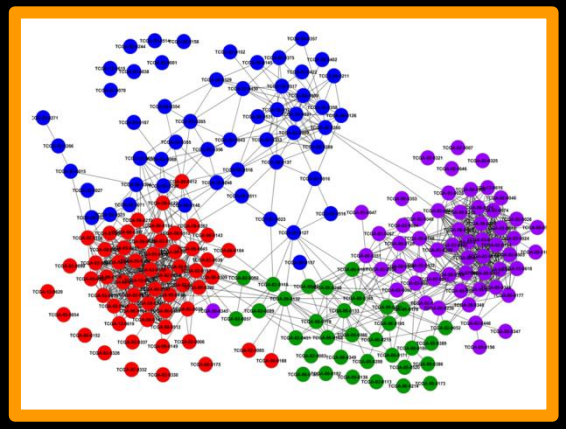


Optimize parameters to maximize likelihood of the patient -patient network

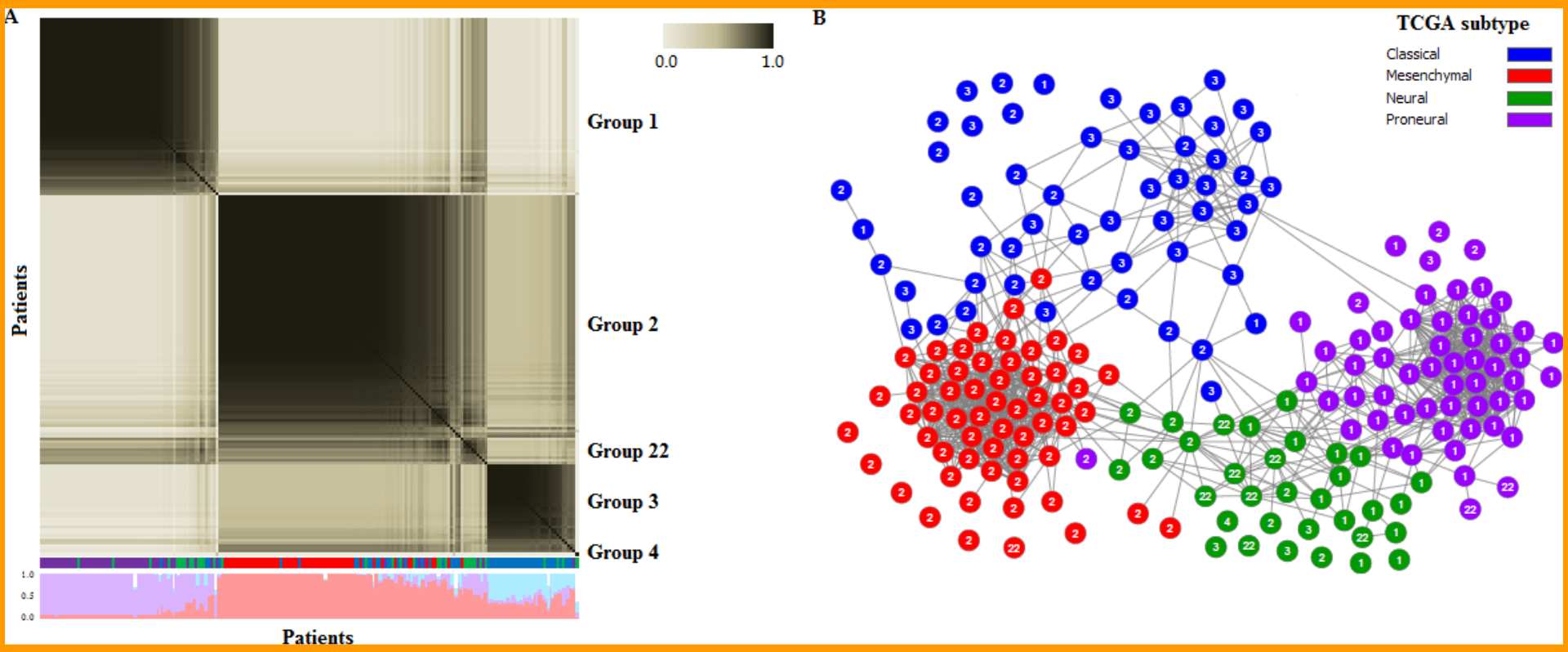
Cho et al. NAR 2013/RECOMB 2012

Visualization of subtypes distribution form a sample model





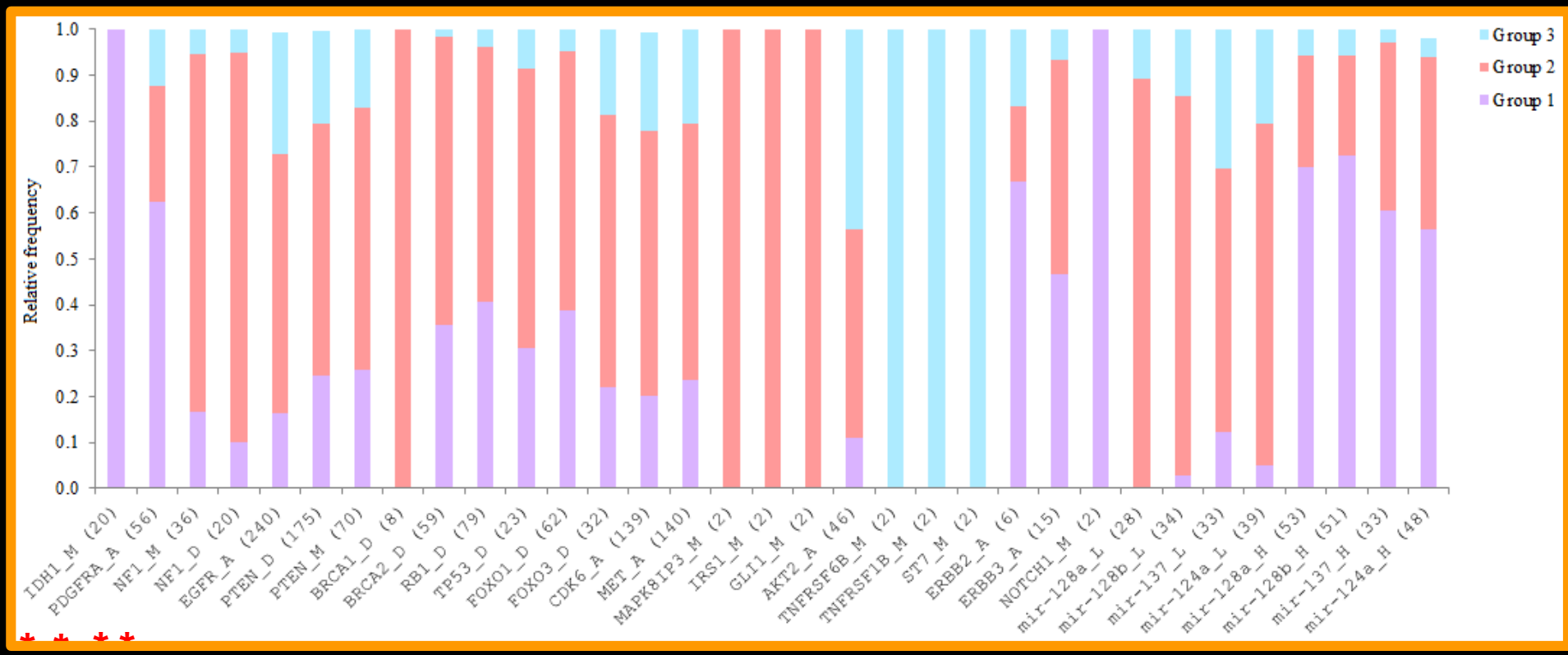
Patient-patient relationship based on1000 models



Observation: No separate Neural group

Cho et al. NAR 2013/RECOMB 2012

Selected cancer related features



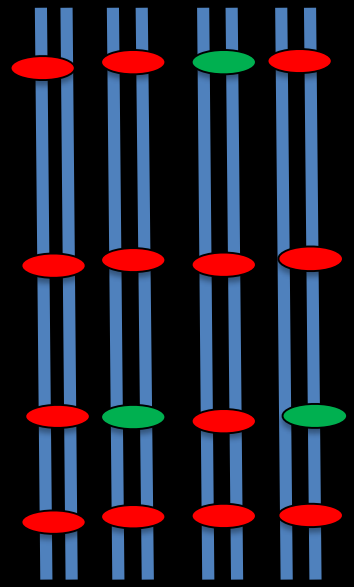
Observations: correctly recovered features form Varhaak et al. (TCGA)

AKT2 – most important defining feature of the Classical group

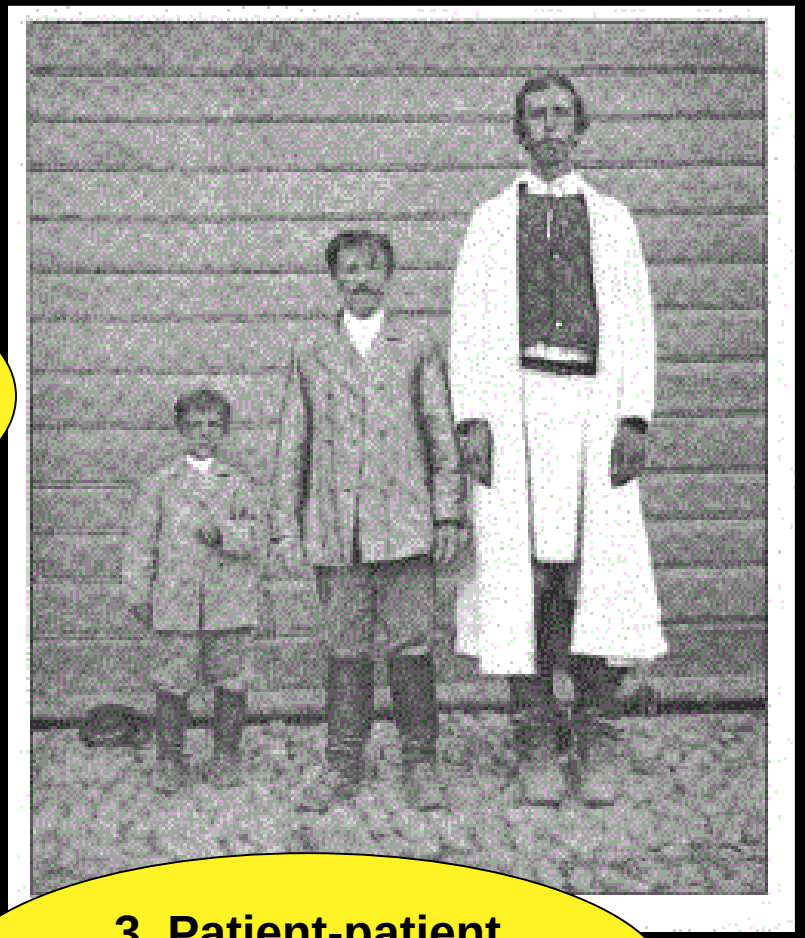
Potential benefits of using dys-regulated pathways as features

Genotypes

Phenotypes



2. Network based signal propagation



1. Dysregulated Networks

3. Patient-patient similarity Networks

oland

Acknowledgments

Przytycka's group

DongYeon Cho

Phuong Dao

Jan Hoinka

YooAh Kim

Yjije Wang

Damian Wojtowicz

Support: Intramural research program NLM / NIH

Using 1,000 models to infer:

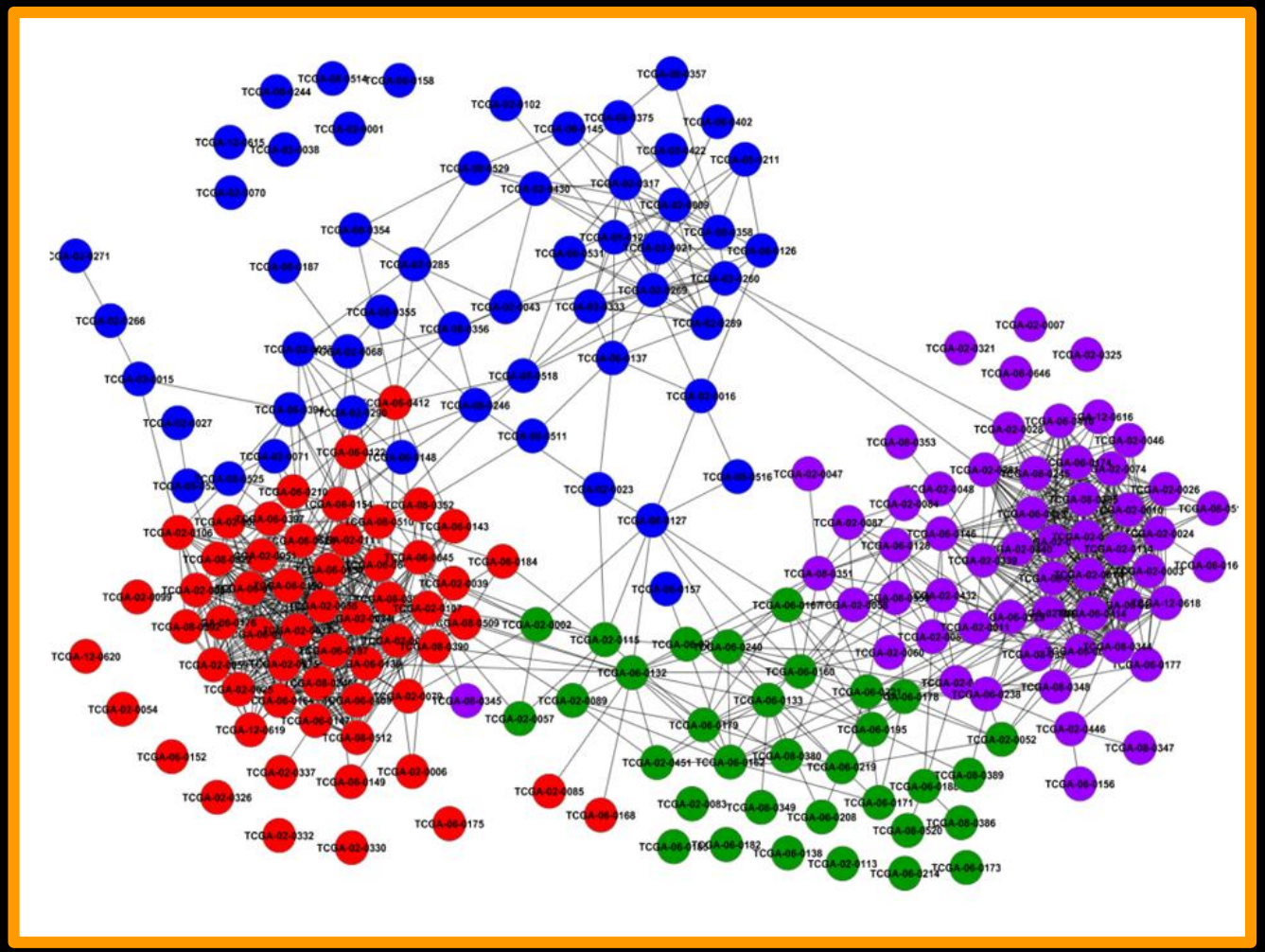
- Probabilistic relation between patients
- Probabilistic relation between features
- Probabilistic elation between features and patients

Case study of GBM (Glioblastoma Multiforme)

Varhaak et al. Classification

- Mesenchymal
- Classical
- Proneural
- Neural

patient network for GMB



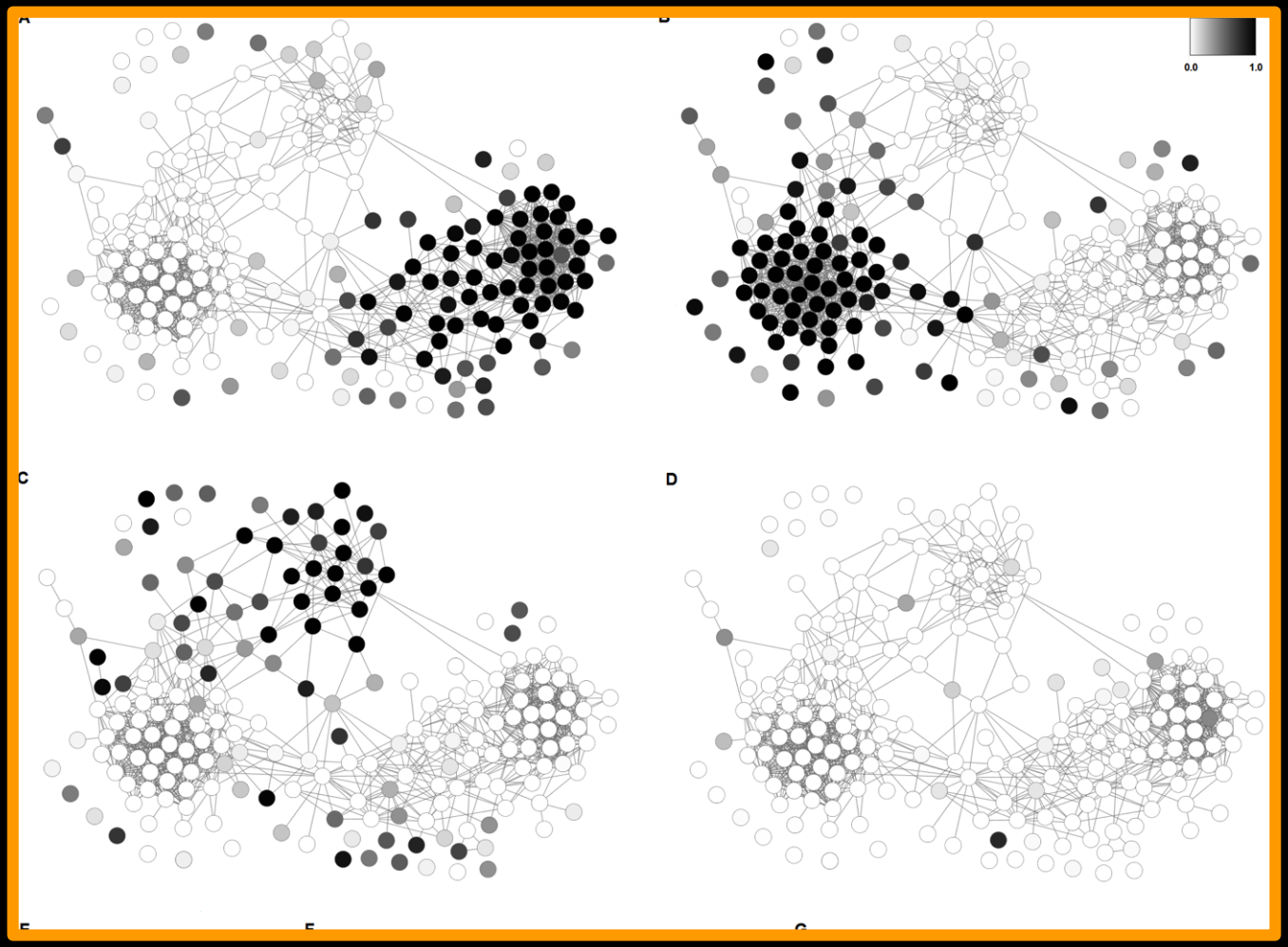
Simultaneous modeling of phenotypic and explanatory features

In each model we assume

- k subtypes
- each subtype is defined by probability distribution of (explanatory) features
- each patient is a mixture of these subtypes
- patients with similar phenotypic features have mixtures

Chang J, Blei DM: Hierarchical Relational Models for Document Networks. Ann Appl Stat 2010, 4(1):124-150.

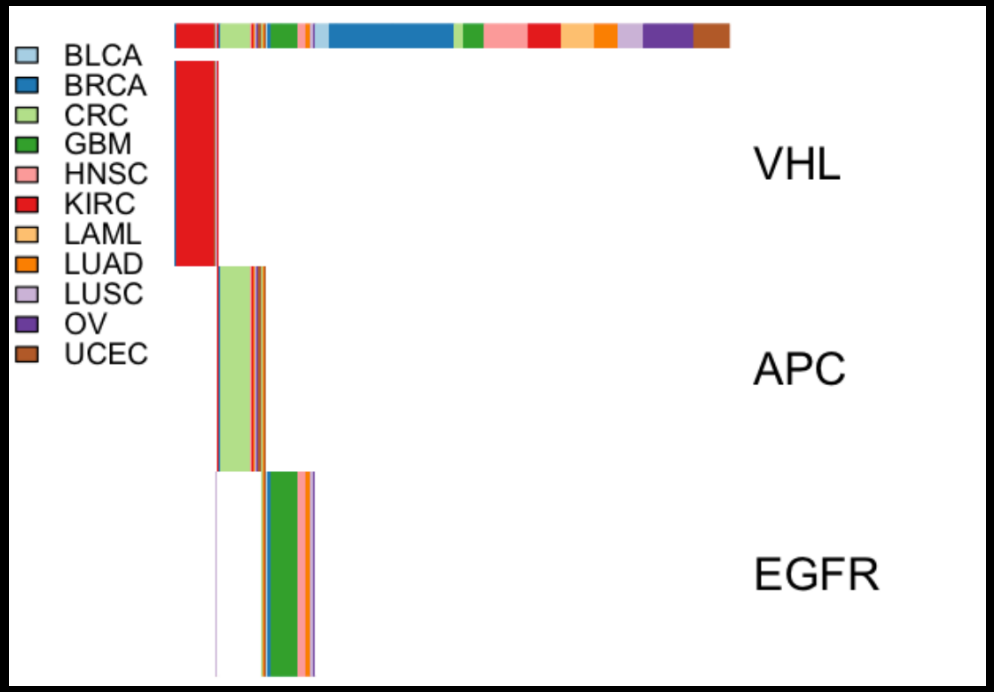
Visualization of subtypes distribution form a sample model



Cho et al. NAR 2013/RECOMB 2012

Mutual Exclusivity and PanCancer TCGA

Can Mutual Exclusivity principle help identifying common pathways dysregulated across cancer types?



Mutual exclusivity is between cancer type specific drivers (expected)
Genes are not in the same pathway (a general property?)

Interaction networks are elucidated by a variety of experimental techniques