Network-based stratification of tumor mutations

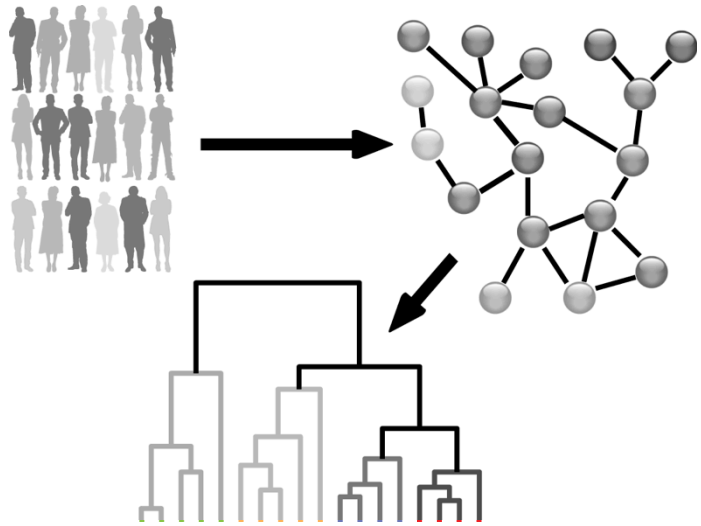
Matan Hofree, John P. Shen, Trey Ideker

- TCGA annual symposium 2012 -

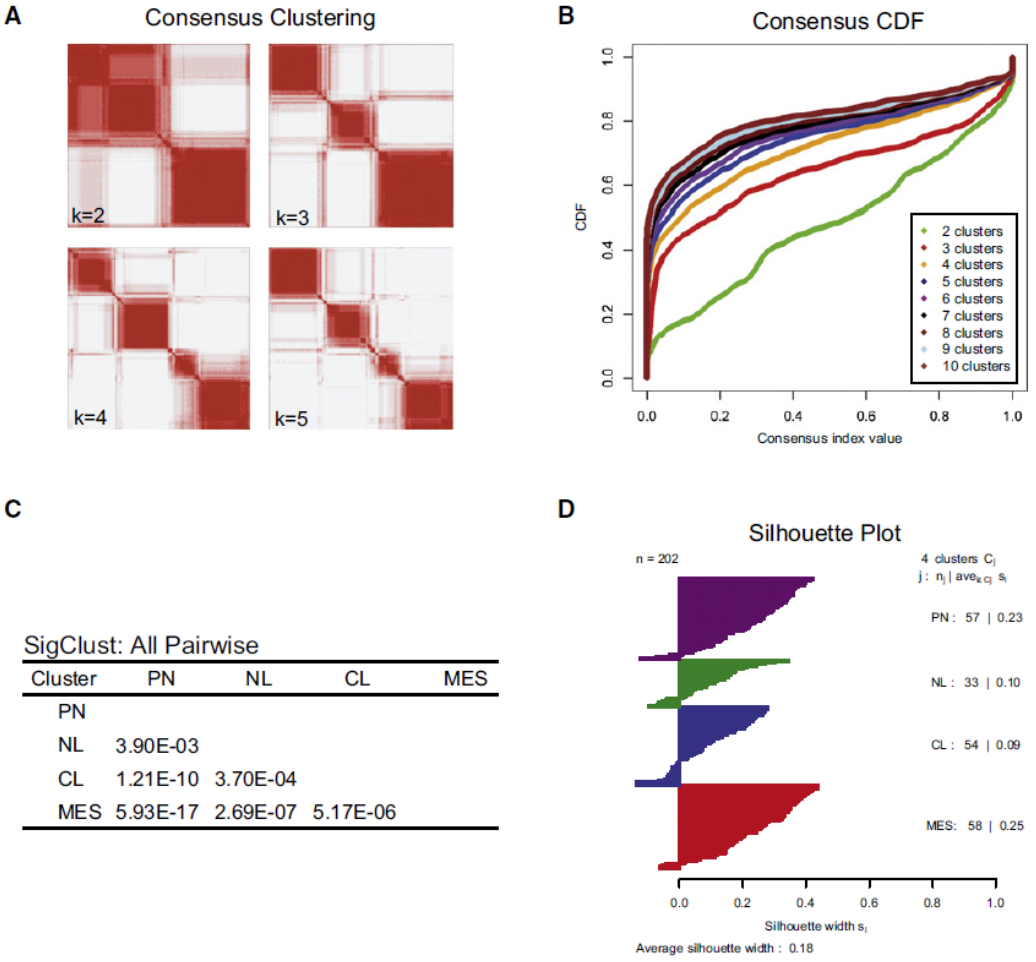
Stratification: dividing cancer into subtypes

Why stratify?

- Better patient prognostics
- A better understanding of tumor biology
- New subtype specific drug targets
- Better patient tailored treatment

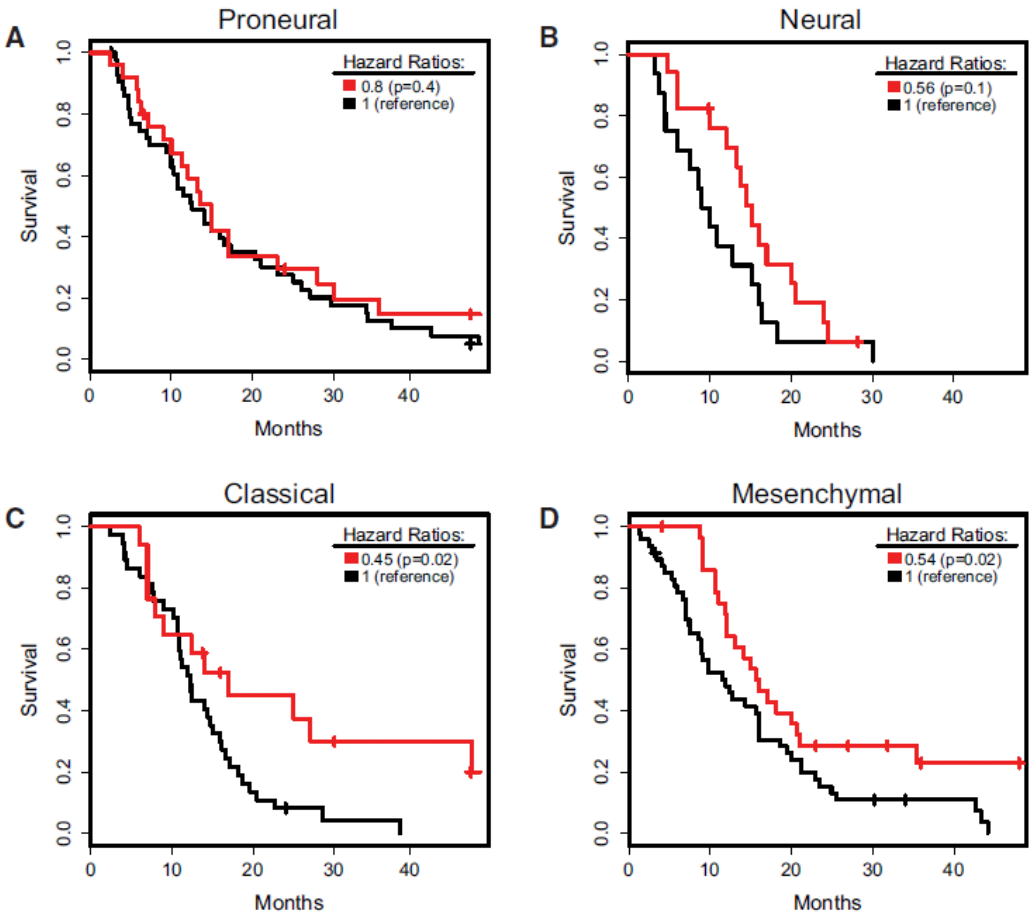


Efforts to stratify using gene expression



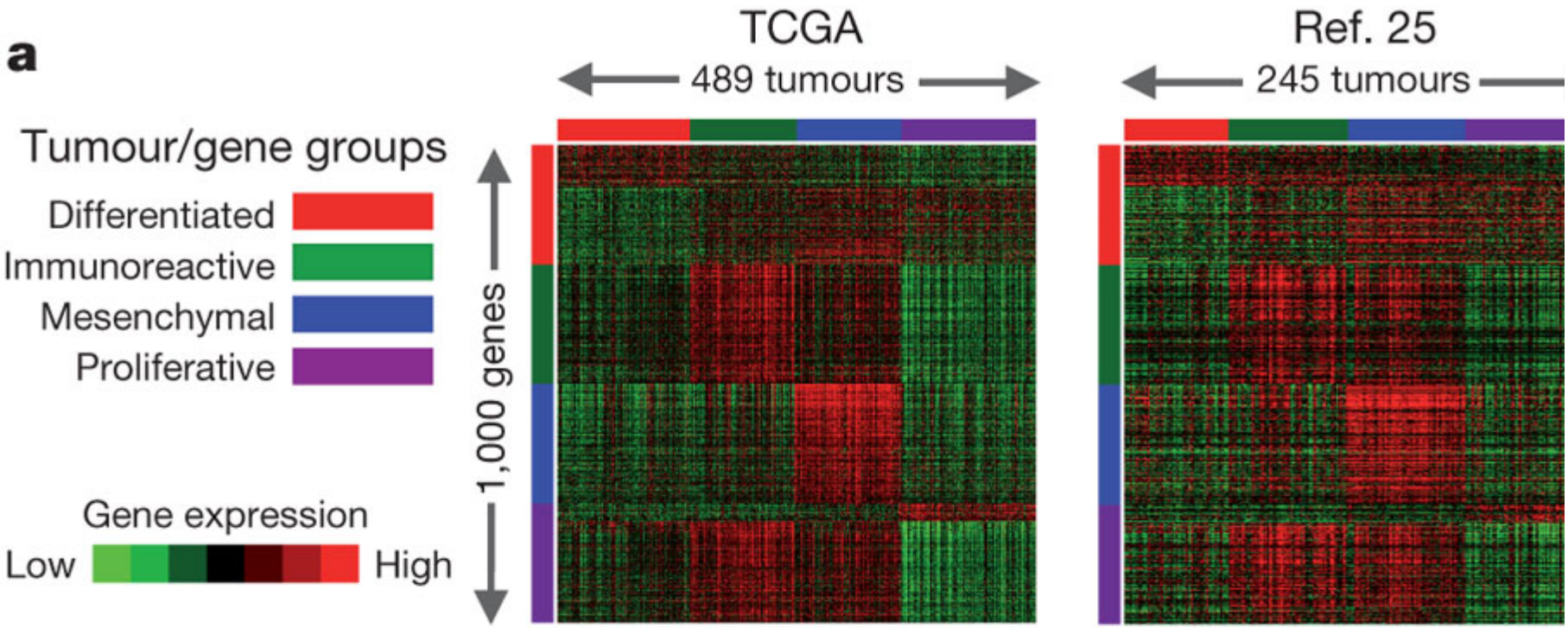
Verhaak, R.G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98-110 (2010).

Four GBM subtypes associated with different survival odds



Verhaak, R.G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98-110 (2010).

Clustering of gene expression in ovarian cancer (OV)



No association to a clinical phenotype was reported (for these subgroups).

T.C.G.A.R.N. (TCGA), Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-15 (2011).

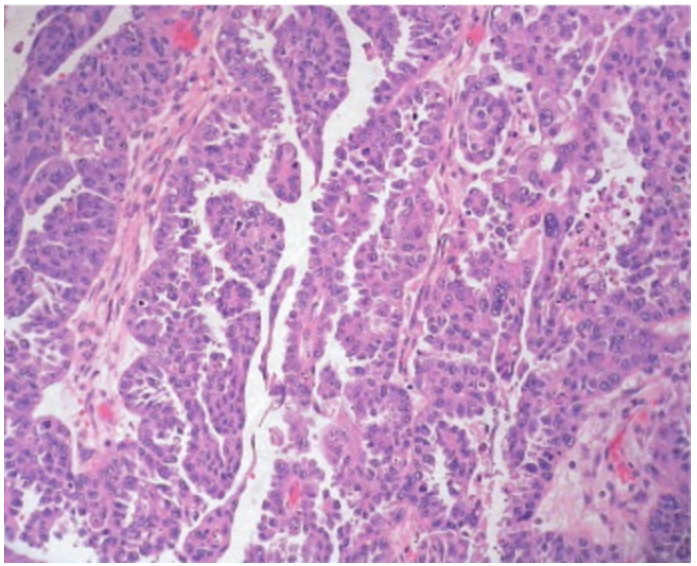
The cancer genome atlas (TCGA)

- +20 Cancer cohorts with 50-800 individuals
- Patient matched samples of different measurement types including:
  - mRNA expression
  - Copy number variations
  - Single nucleotide polymorphisms
  - Methylation
  - miRNA
  - Protein expression
  - Patient genomes (somatic mutations)

The Cancer Genome Atlas

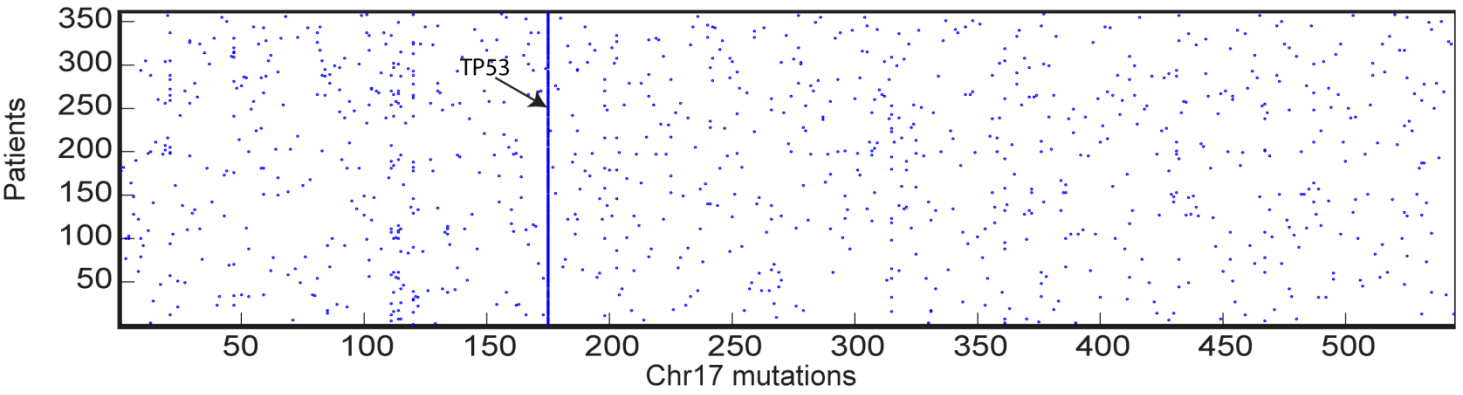
Somatic mutations in high grade serous ovarian cancer

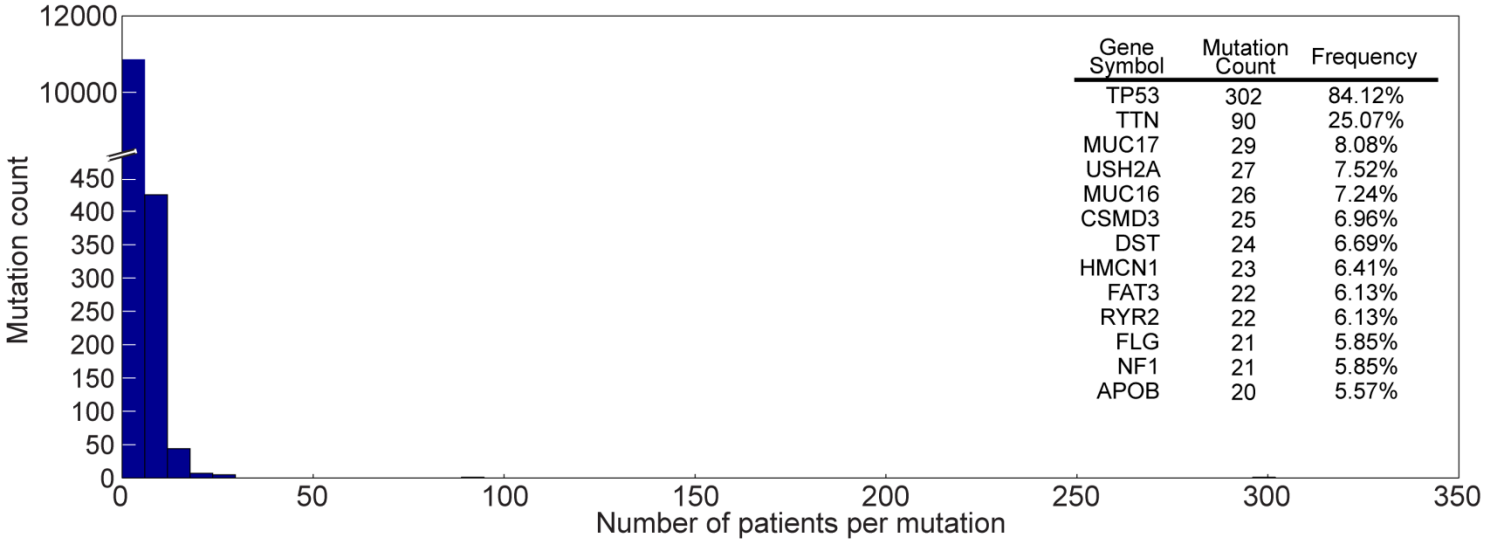
- 359 matched patient/tumor exome sequenced with Illumina GAIIx
- 11,231 somatic mutations



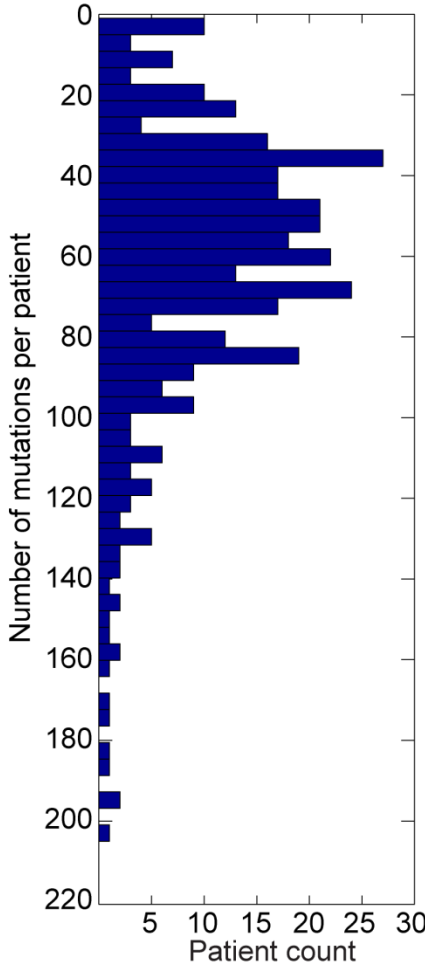
The Cancer Genome Atlas

Why is it hard to cluster somatic mutation genotypes?

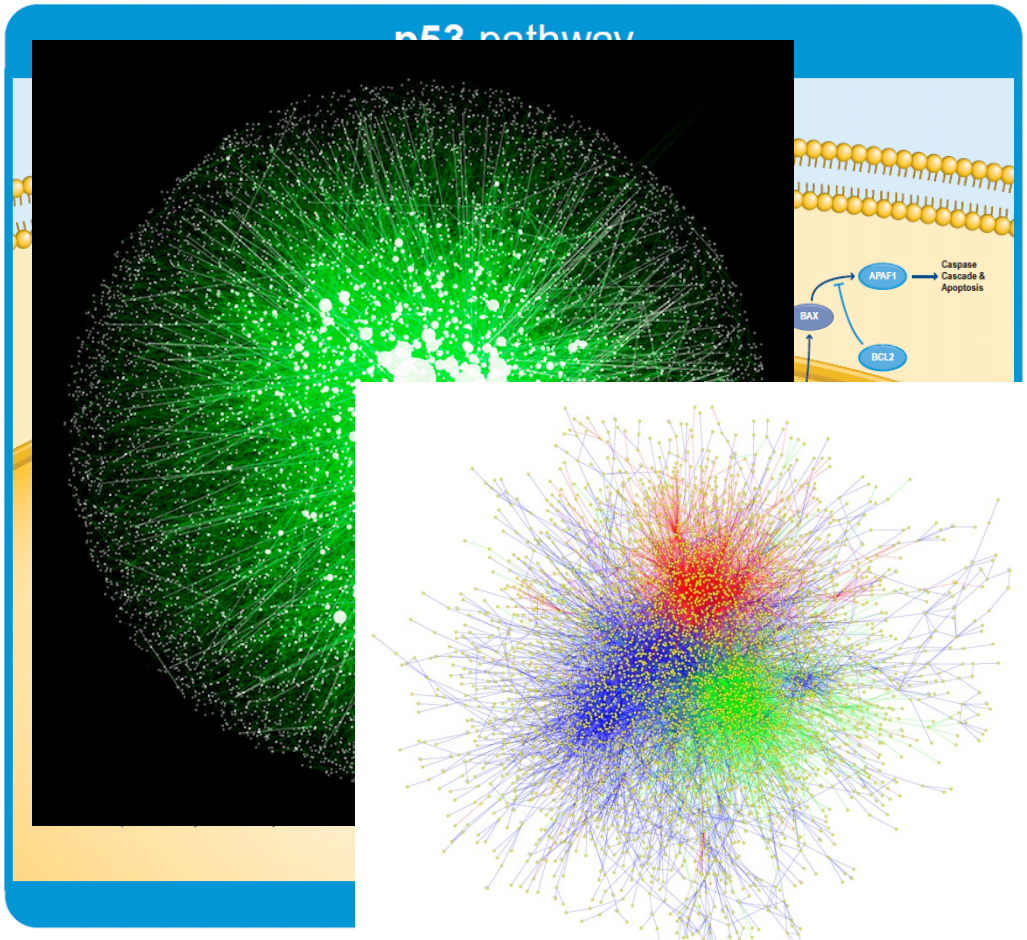




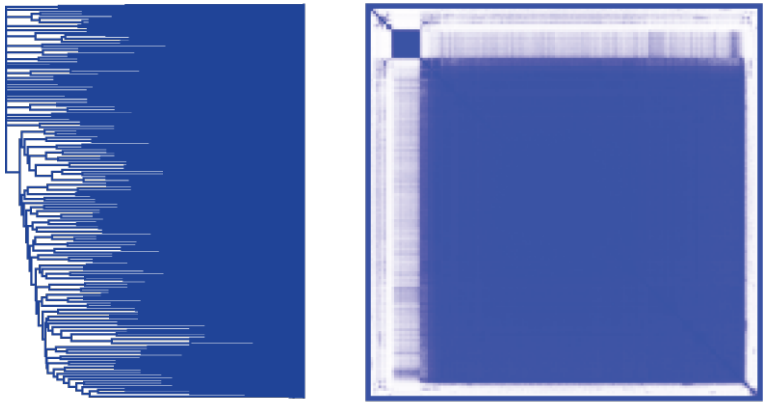
Gene Symbol	Mutation Count	Frequency
TP53	302	84.12%
TTN	90	25.07%
MUC17	29	8.08%
USH2A	27	7.52%
MUC16	26	7.24%
CSMD3	25	6.96%
DST	24	6.69%
HMCN1	23	6.41%
FAT3	22	6.13%
RYR2	22	6.13%
FLG	21	5.85%
NF1	21	5.85%
APOB	20	5.57%



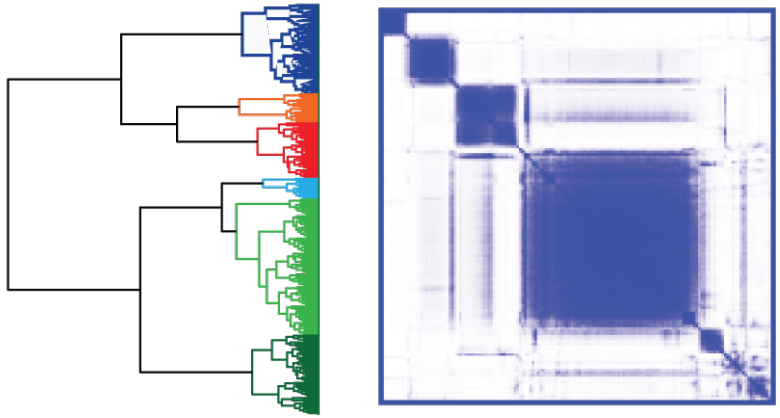
Improving stratification with networks



Regular Consensus Clustering NMF



Network based stratification



Network-based stratification:

Draw a bootstrap sample of genes from *G*(*patients x genes)* matrix.

↓

Network smoothing:
For each patient project mutations onto a network (*A*) and propagate.

↓

Network clustering:
Cluster smoothed *F*(*patients x genes*) matrix using Network NMF

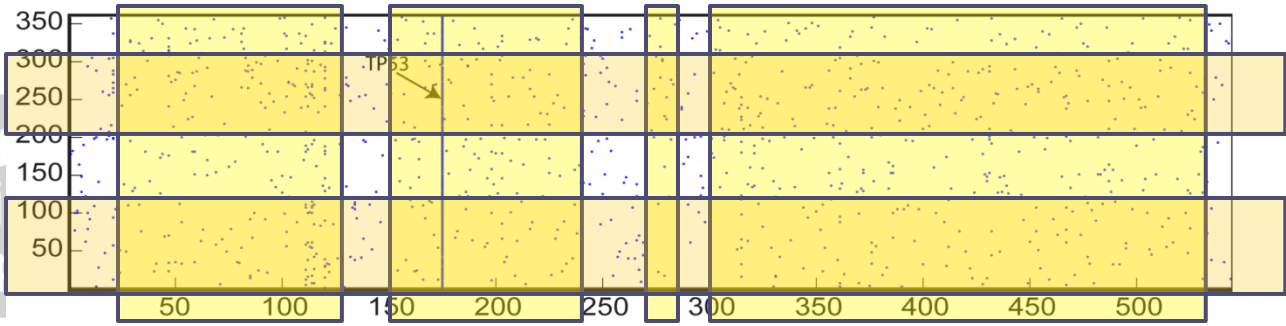
↓

Repeat *N* times and aggregate into a (*patients x patients*) consensus matrix

Network-based stratification:

Draw a bootstrap sample of genes from *G*(*patients x genes)* matrix.

Network smoothing
For each patient proj
onto a network (*A*) and



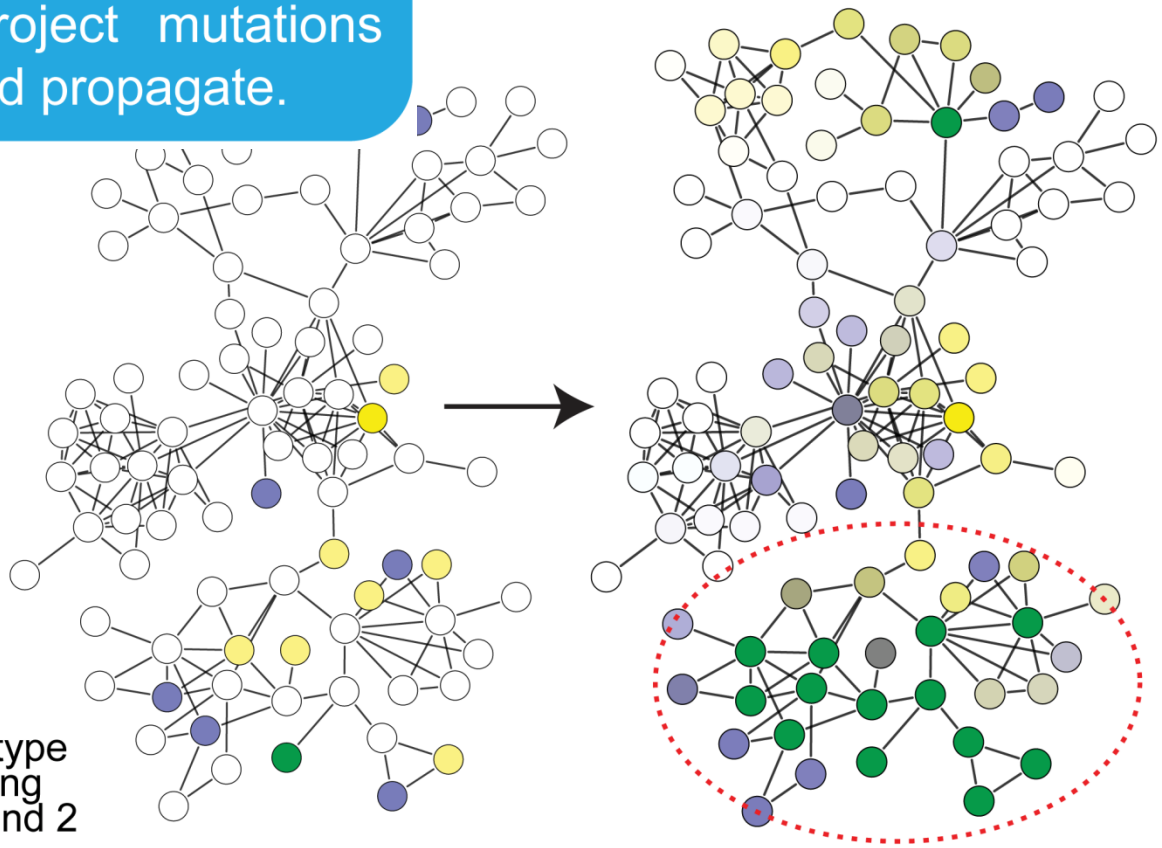
Network clustering:
Cluster smoothed *F*(*patients x genes*) matrix using Network NMF

Repeat *N* times and aggregate into a (*patients x patients*) consensus matrix

Network-based stratification:

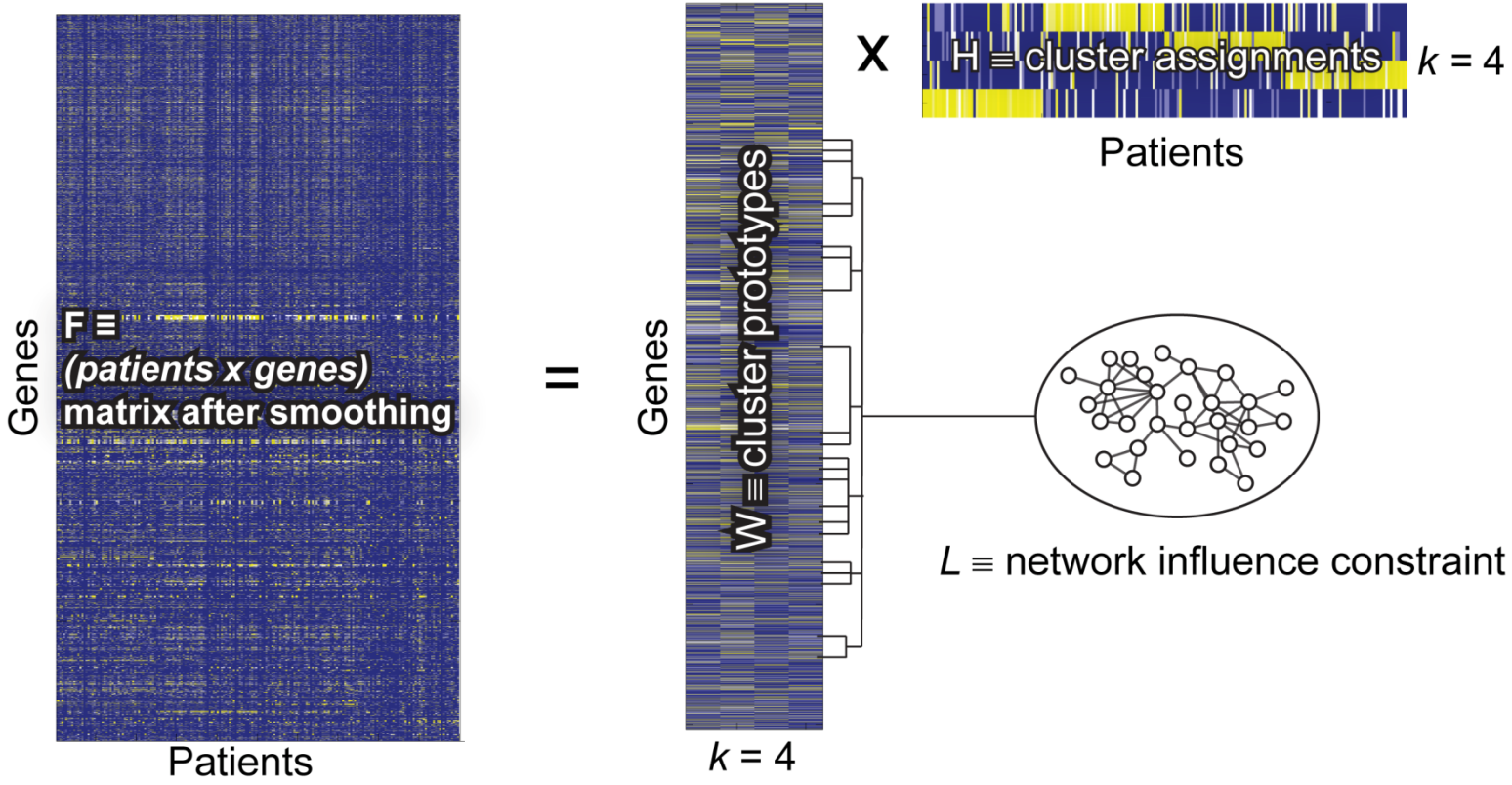
Draw a bootstrap sample of genes from *G(patients x genes)* matrix.

Network smoothing:
For each patient project mutations onto a network (*A*) and propagate.



Network-based stratification:

Draw a bootstrap sample of genes from *G(patients x genes)* matrix.



Consensus clustering NMF of somatic mutations

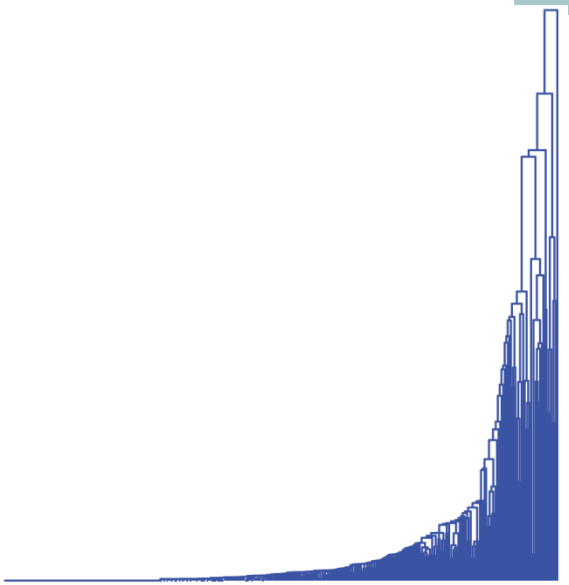
Network-based stratification

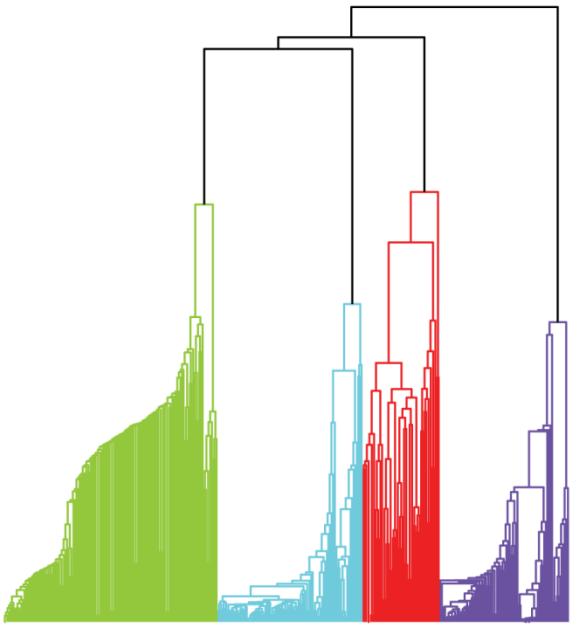
Draw a bootstrap sa
from *G*(*patients x gen*

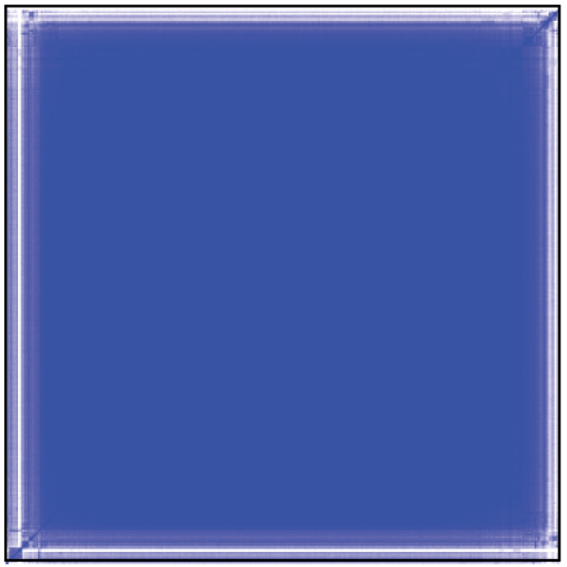
Network smoothi
For each patient p
onto a network (*A*) ar

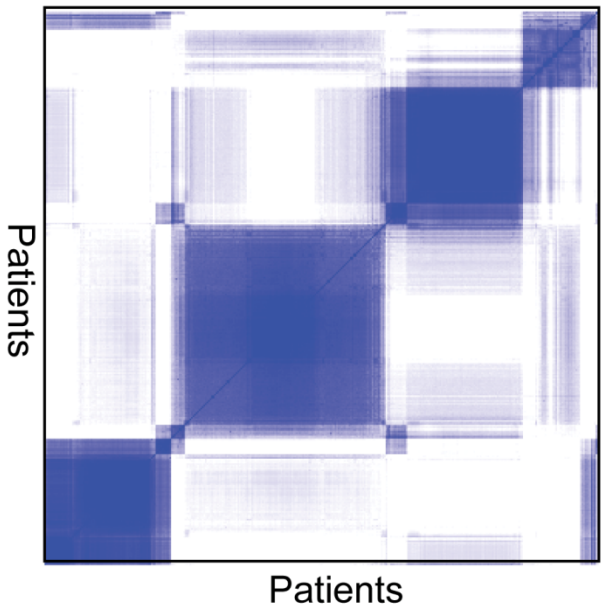
Network clusterir
Cluster smoothed *F*(*p*
matrix using Network

Repeat *N* times and
(*patients x patients*) consensus matrix

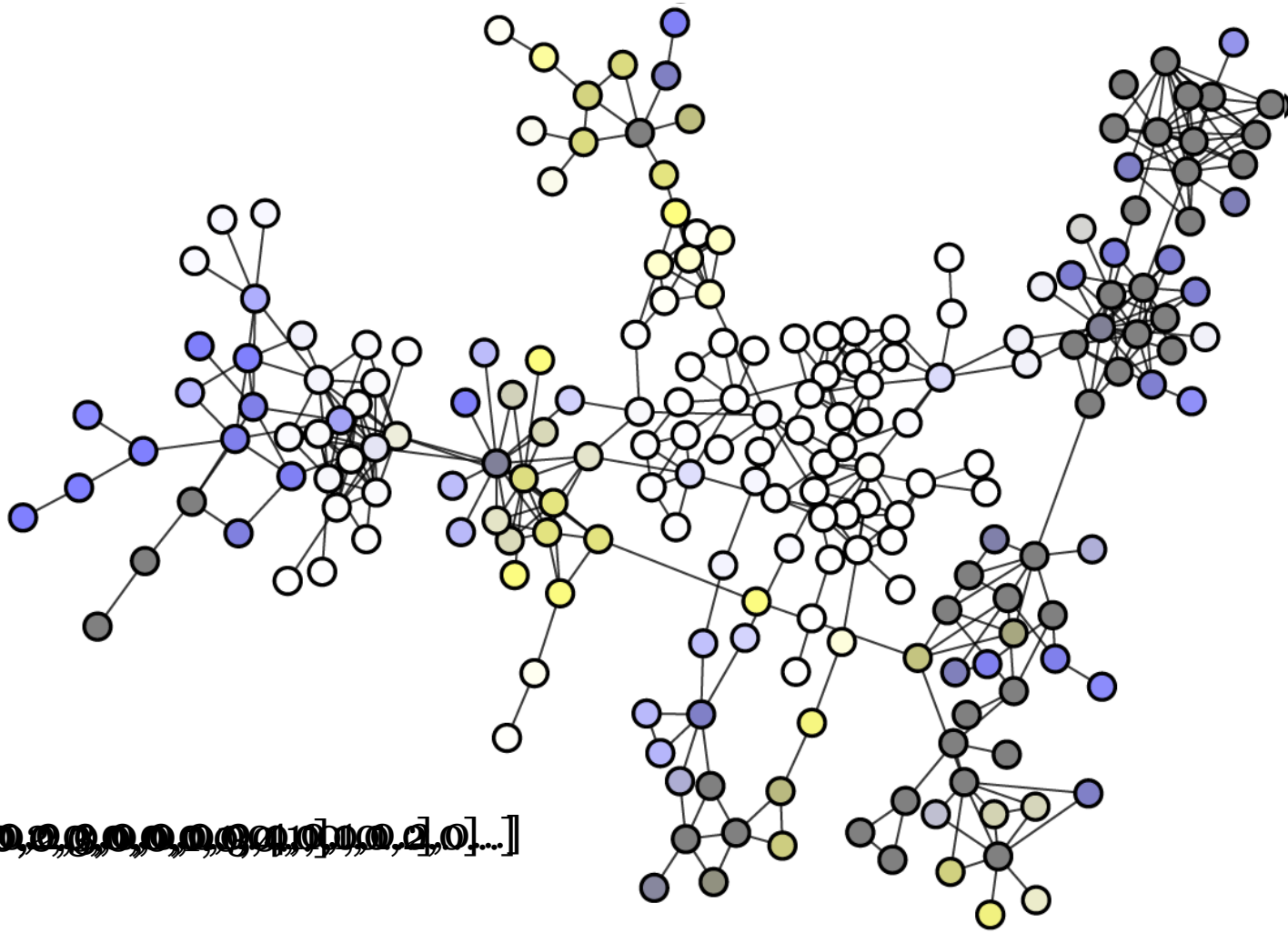








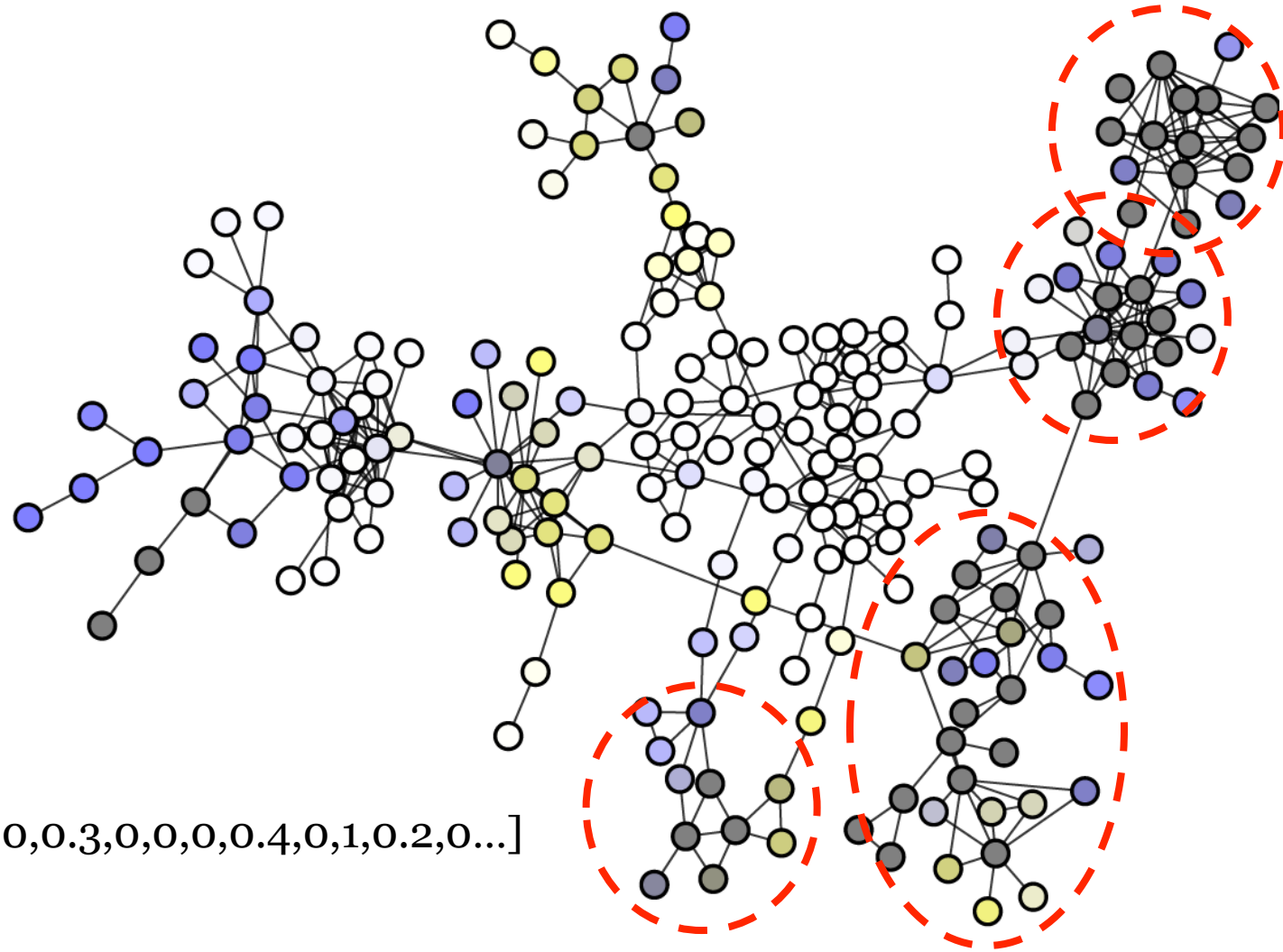
An intuition for network smoothing



Genotype A: [...,
[0,0,3,1,0,2,0,1,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,2,0,...]

Genotype B: [[...,
[0,0,3,0,0,2,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,2,0,...]]

An intuition for network smoothing



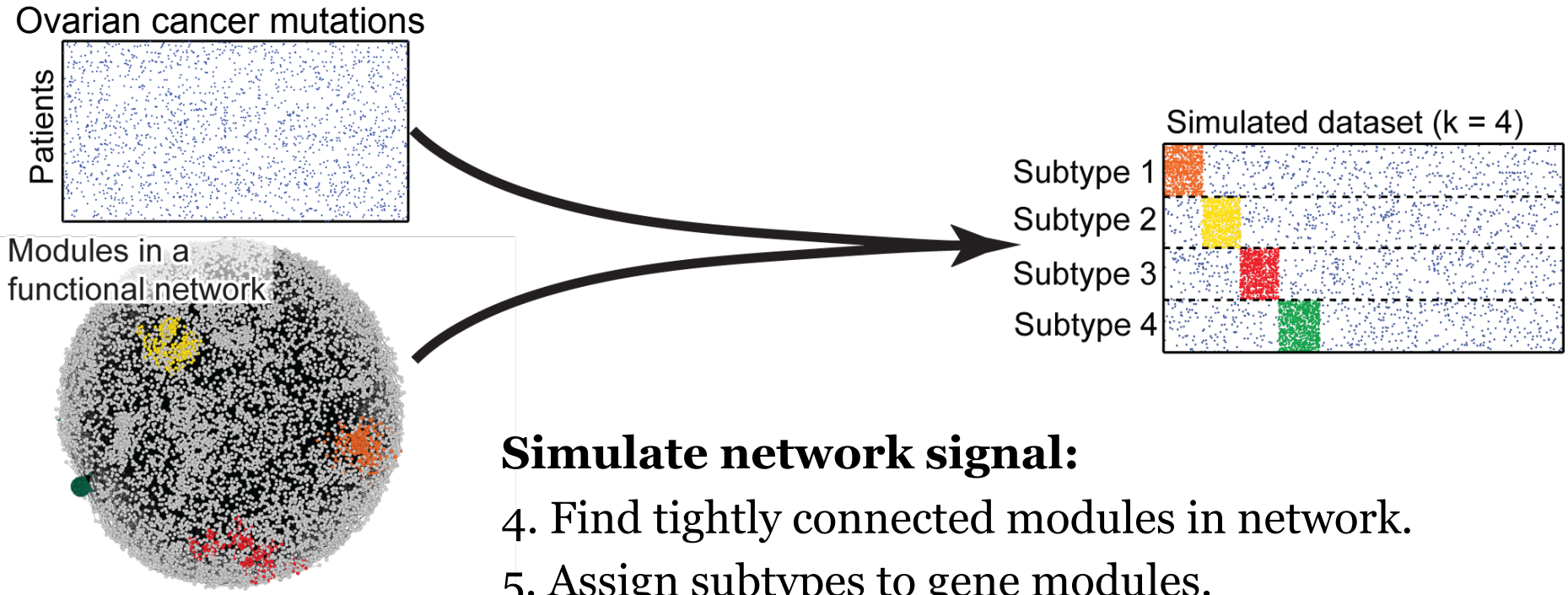
Genotype A: [...,
0,0.3,1,0.4,0.4,1,0.2,0,0,0.3,0,0,0,0.4,0,1,0.2,0...]

Genotype B: [...,
0,0.3,1,0.2,0.4,0.2,0,0,0.4,1,0,0,0,0,0.2,0.4,0.5,0...]

Simulating ‘network’ data

Simulate background mutations:

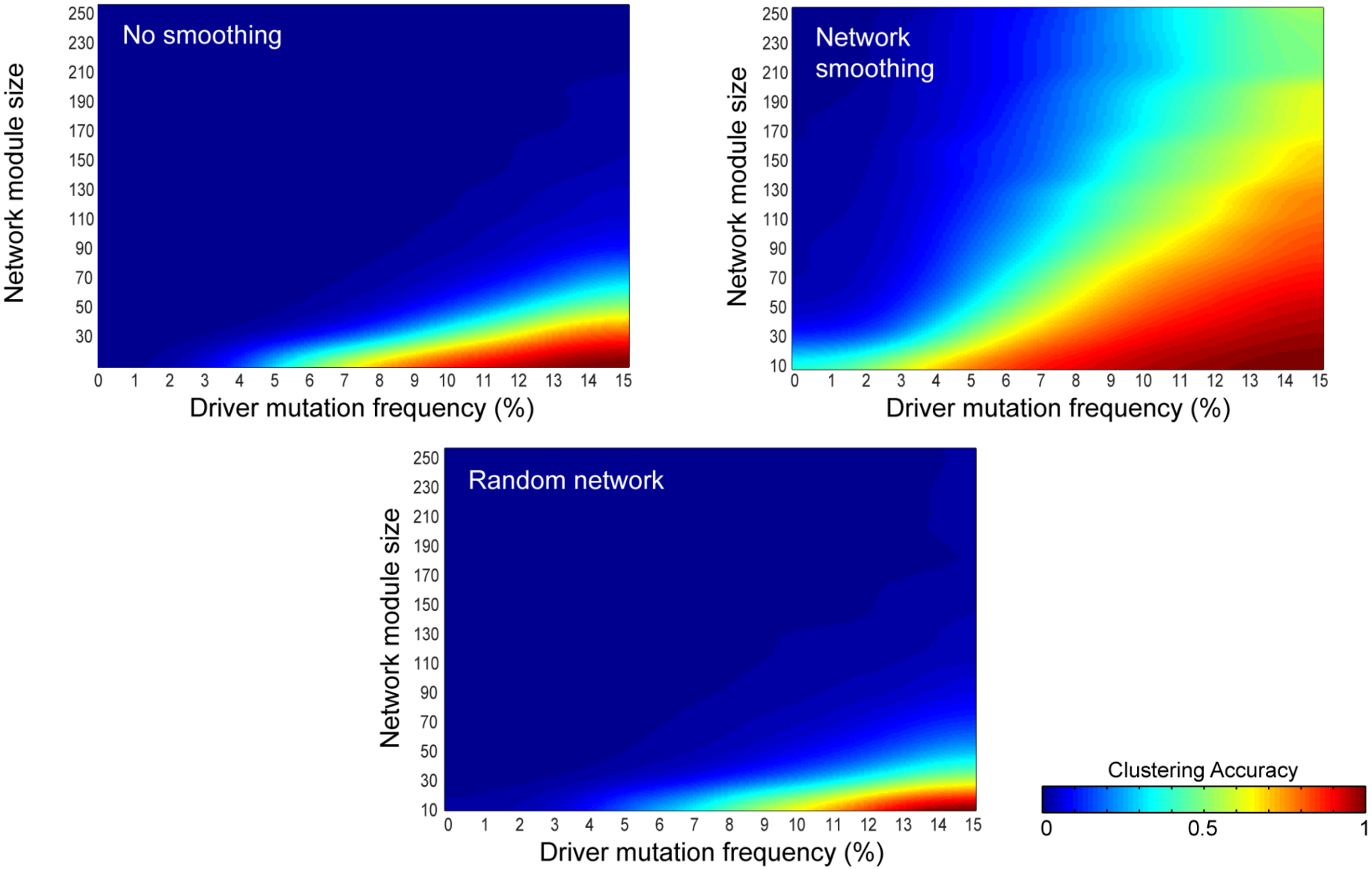
1. Sample patients from dataset.
2. Permute mutated genes.
3. Divide patients into k subtypes.



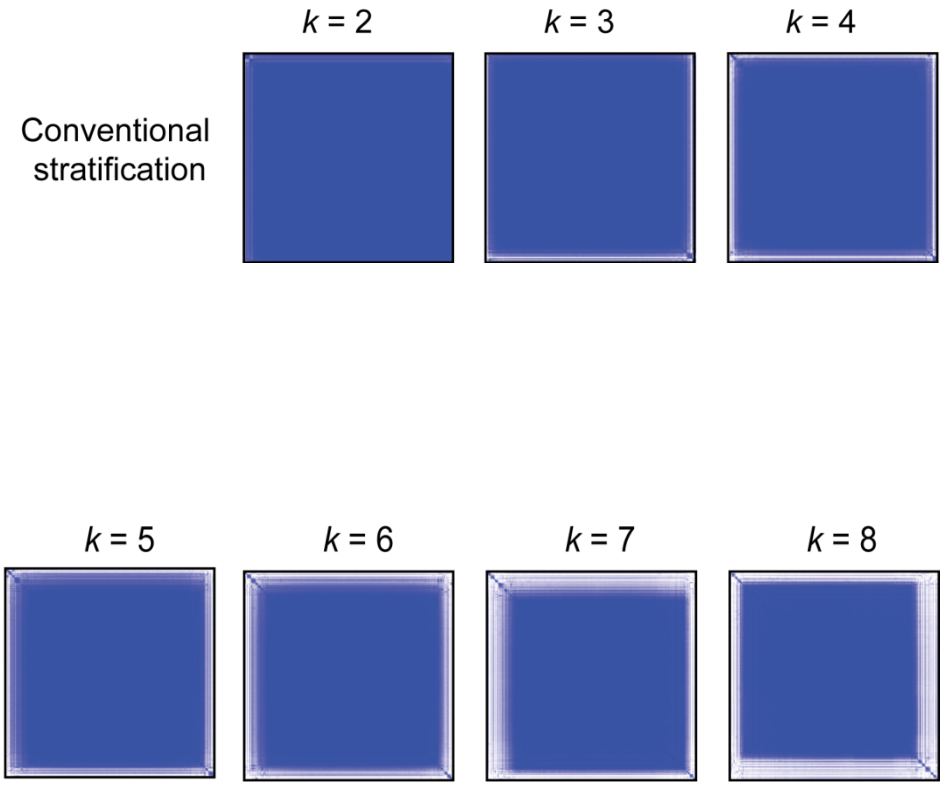
Simulate network signal:

4. Find tightly connected modules in network.
5. Assign subtypes to gene modules.
6. For each patient move m% of mutated genes to modules in the patient’s subtype.

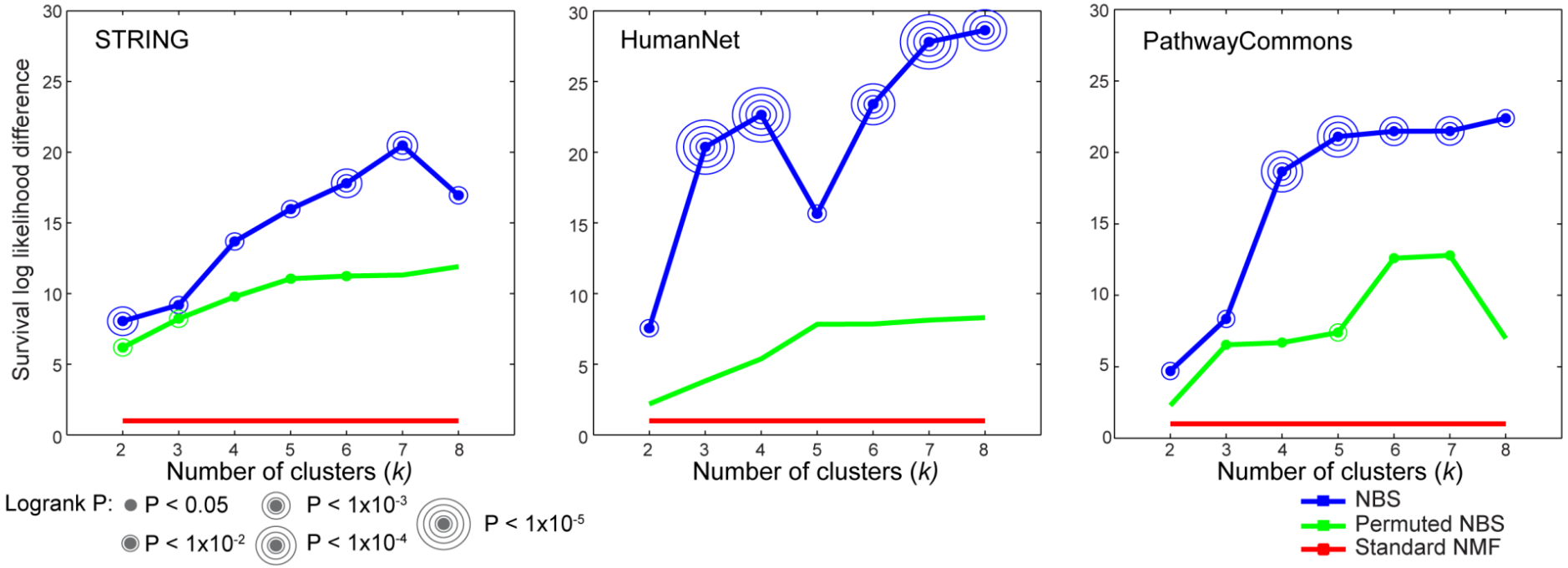
Simulation - a different landscape



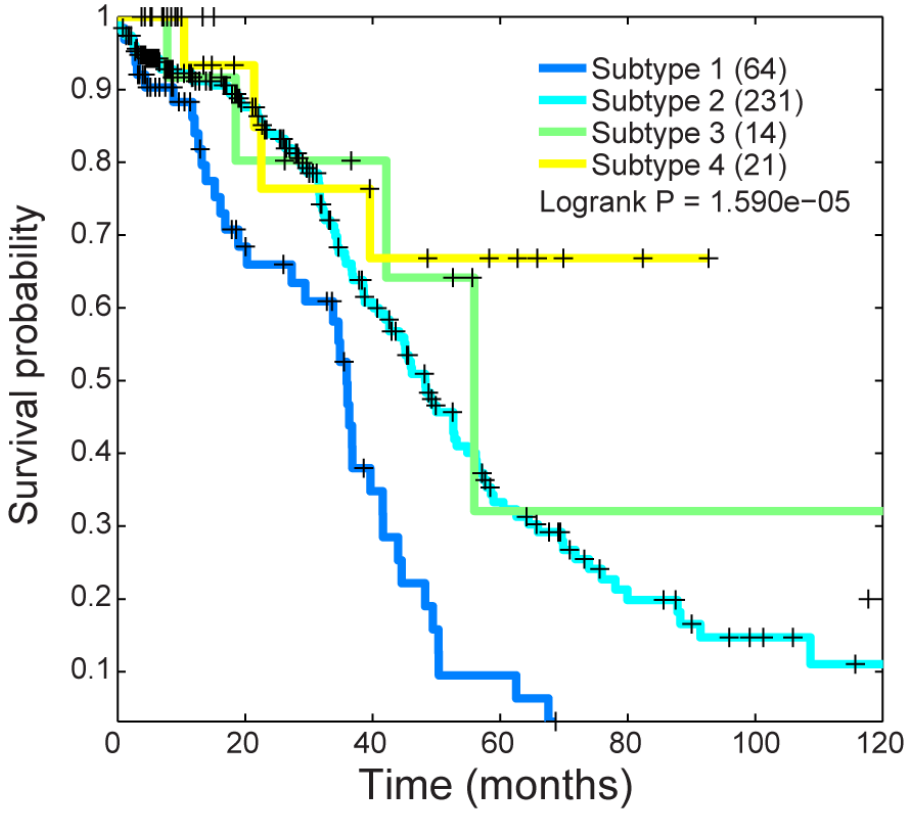
Network-based stratification on somatic mutations from TCGA ovarian cancer

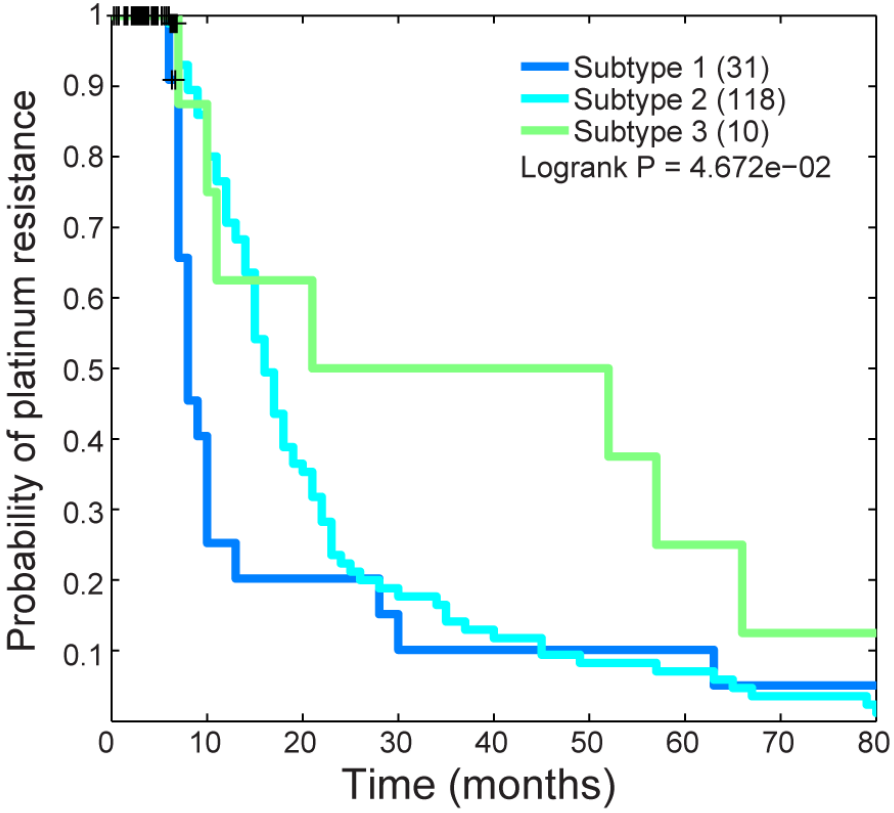


Association with patient survival

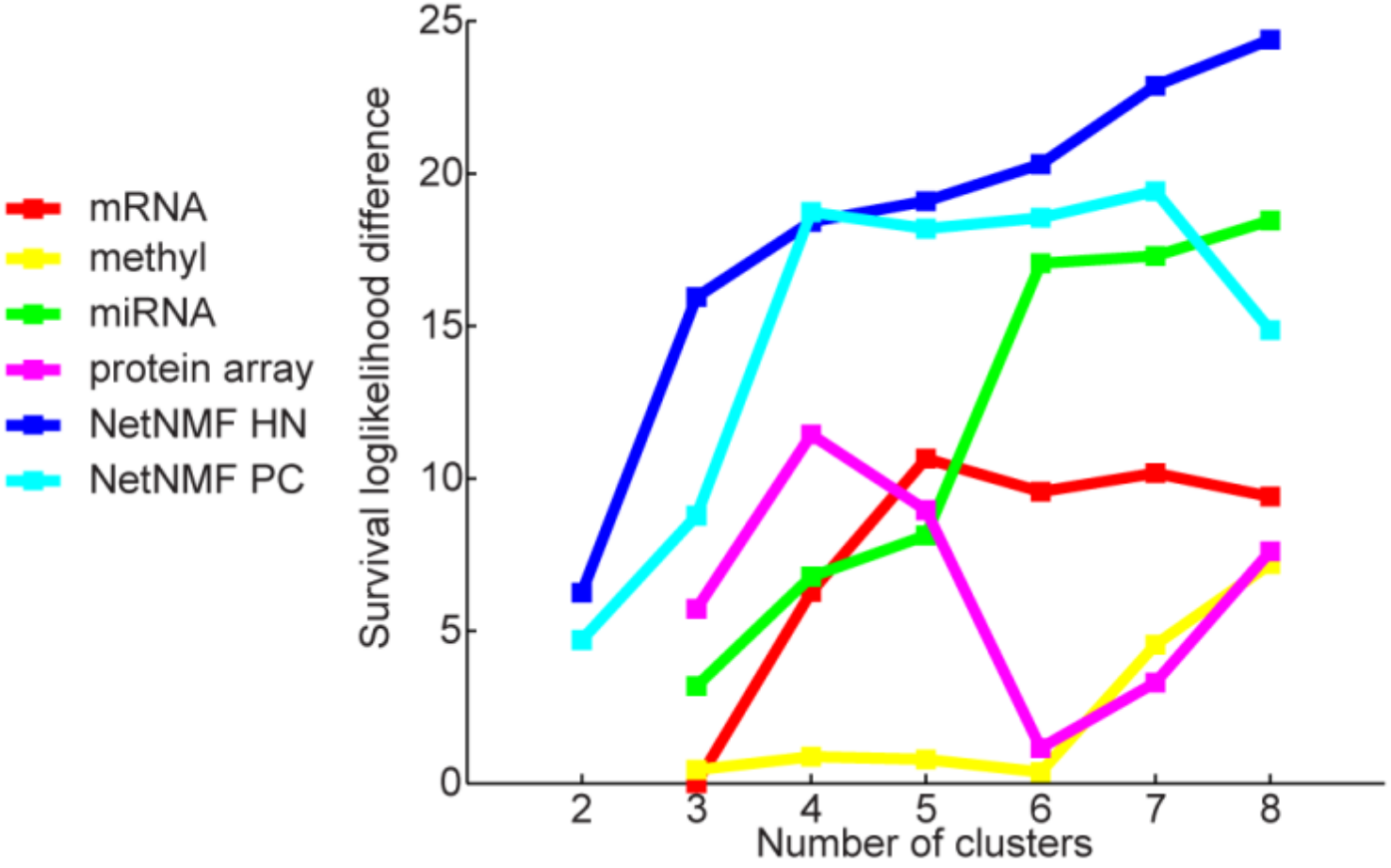


Association with patient survival





Comparing to other data types



FIREHOSE
Broad GDAC

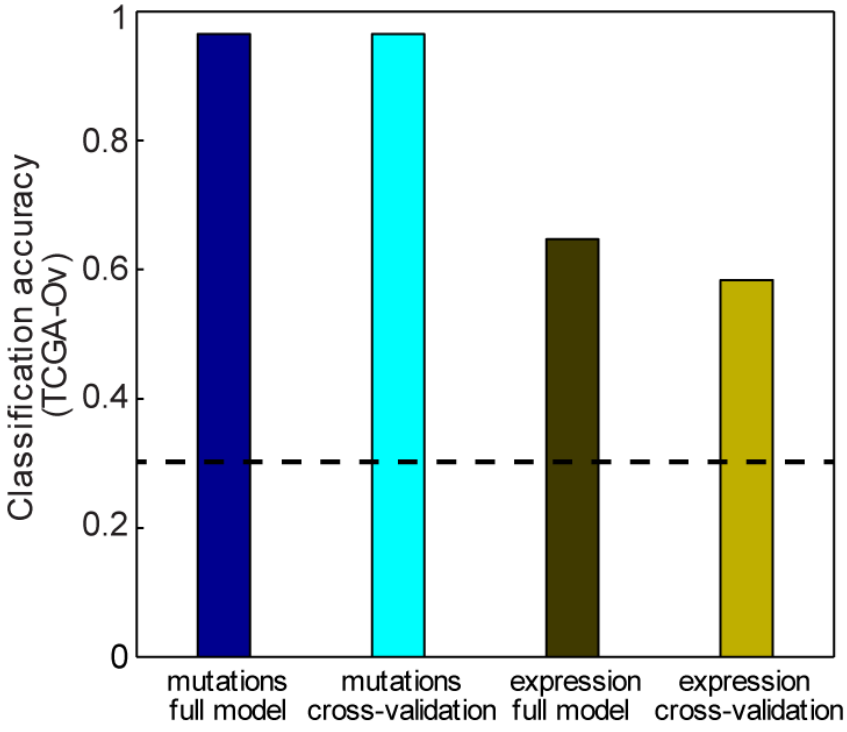
Clinical translation of subtypes using expression signature

Measuring the expression of a gene set is easier than sequencing a genome.

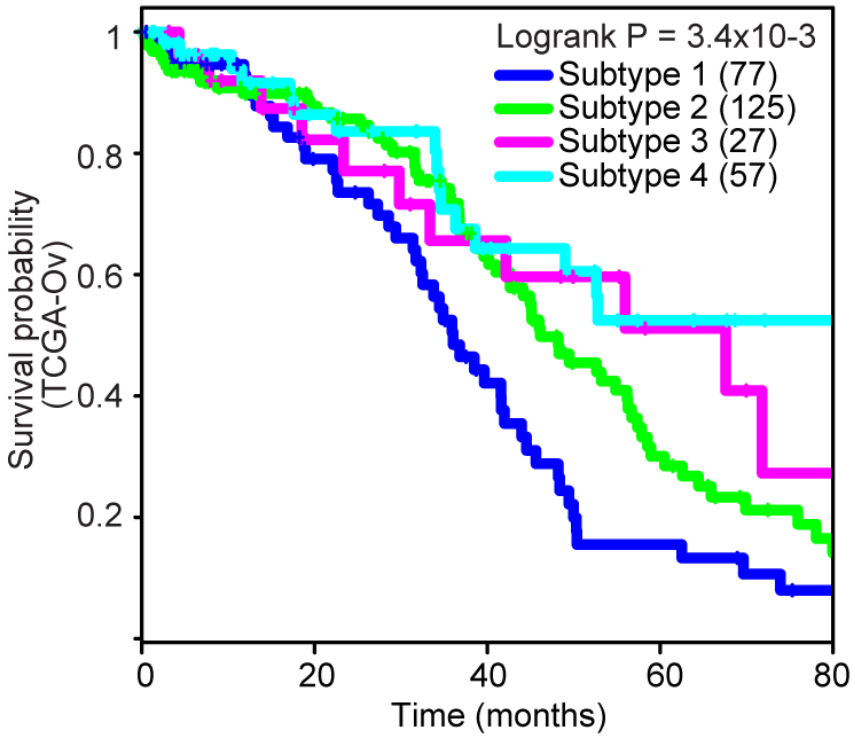
1. **Define subtypes** using somatic mutations which predict a clinical phenotype (survival, drug response).
2. **Train a model on matched gene expression** to predict subtypes on the same set of patients.
3. **Predict subtypes** using expression on new patients.

Clinical translation of TCGA-OV subtypes

Classification accuracy recovering NBS subtypes

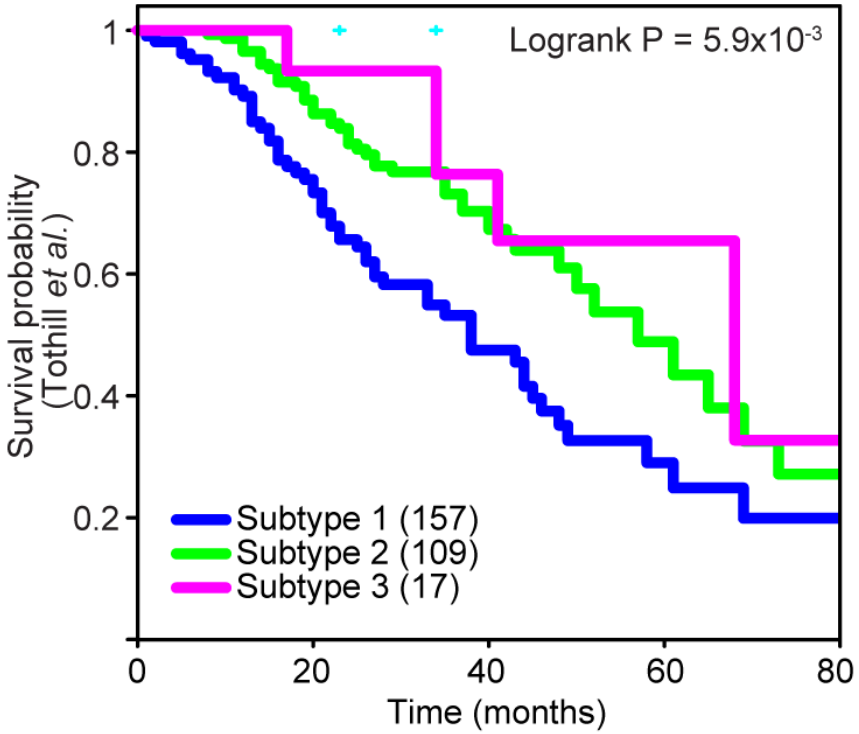


Overall survival with expression recovered subtypes

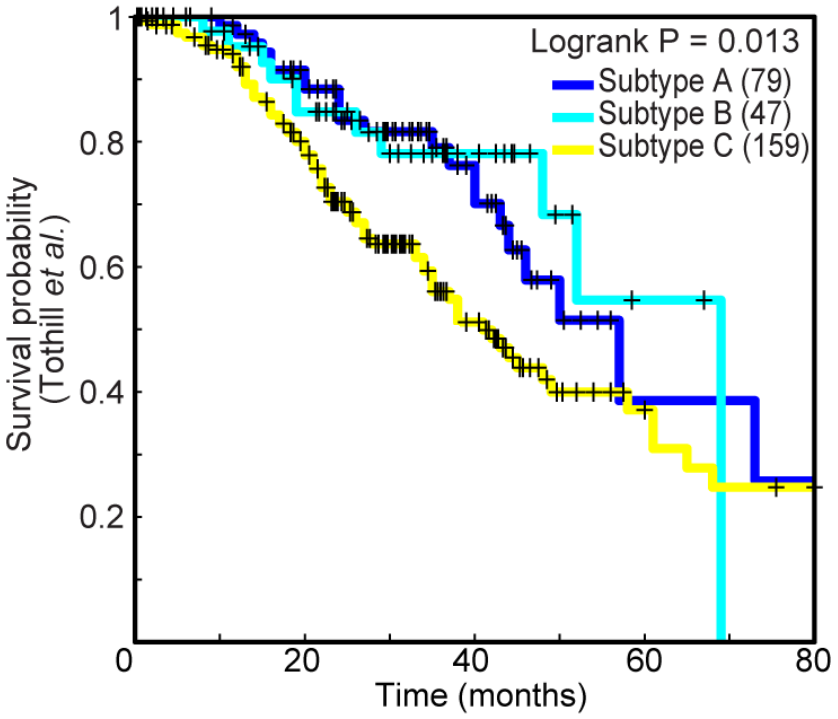


Clinical translation of TCGA-OV subtypes

Somatic Mutation subtypes recovered in Tothill[1] gene expression cohort

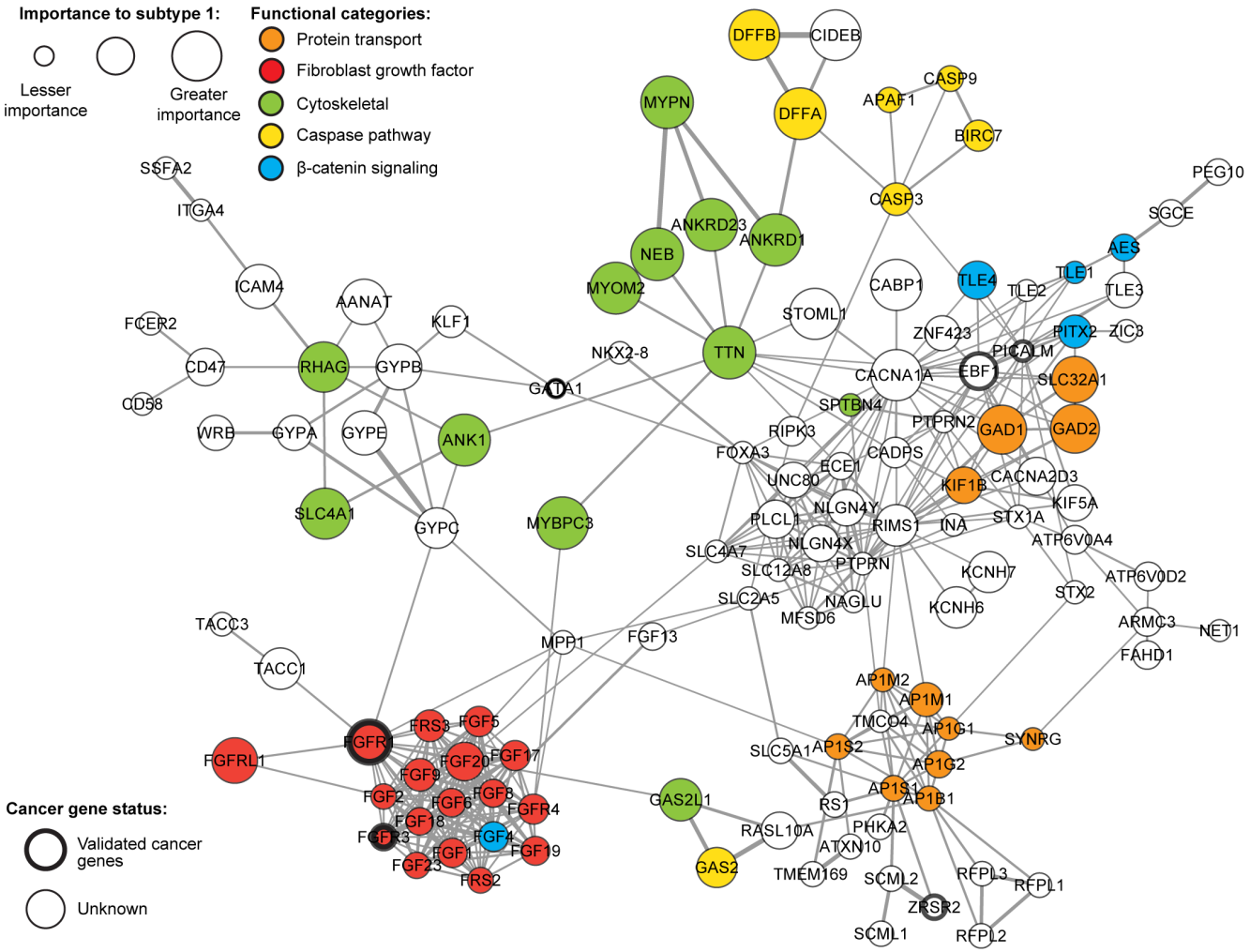


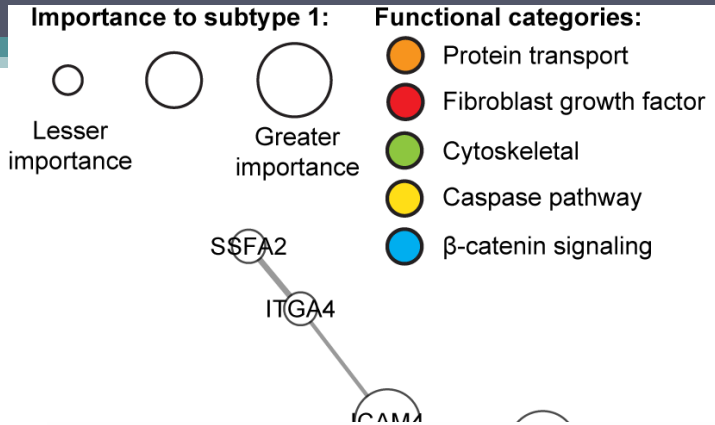
Standard consensus clustering NMF

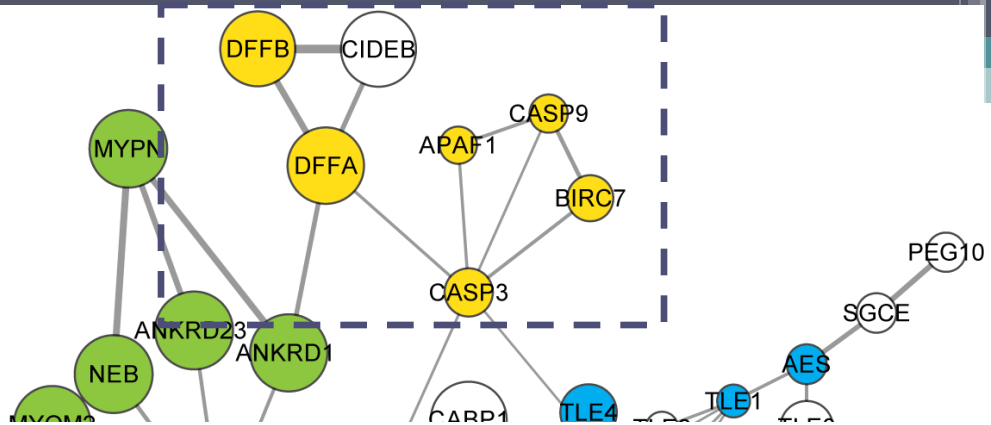


1 Tothill RW, *et al.*, Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* 2008, **14:**5198-5208.

A characteristic network for subtype 1







[CANCER RESEARCH 59, 3077–3083, July 1, 1999]

Cisplatin-induced Apoptosis Proceeds by Caspase-3-dependent and -independent Pathways in Cisplatin-resistant and -sensitive Human Ovarian Cancer Cell Lines[1]

Karen M. Henkels and John J. Turchi[2]

Department of B

ABSTRACT

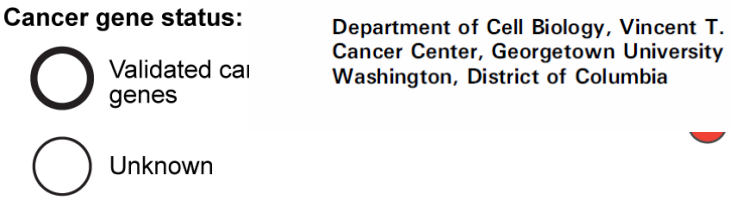
We have as
cell death in
and two drug

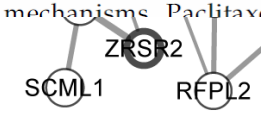
Src inhibition enhances paclitaxel cytotoxicity in ovarian cancer cells by caspase-9-independent activation of caspase-3

Ting Chen, Yolande Pengetnze, and Christopher C. Taylor

Department of Cell Biology, Vincent T. Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, District of Columbia

largely the result of late presentation of patients due to clinically silent symptoms until disseminated and metastatic disease has been well established. Following tumor debulking, patients generally receive chemotherapy treatment with paclitaxel, platinum-based agents, or a combination of both (2, 3). These agents act via different mechanisms. Paclitaxel binds tubulin





COMMENTARY

Cancer Biology & Therapy 10:5, 505-508; September 1, 2010; © 2010 Landes Bioscience

RESEARCH PAPER

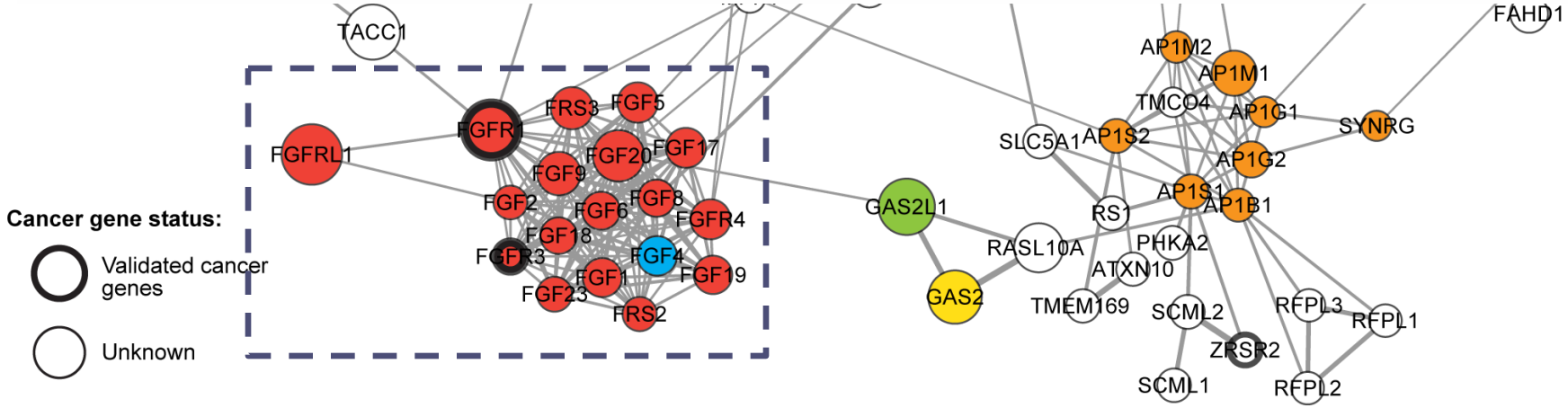
Cancer Biology & Therapy 10:5, 495-504; September 1, 2010; © 2010 Landes Bioscience

Inhibition of FGFR2 and FGFR1 increases cisplatin sensitivity in ovarian cancer

Claire Cole,[1] Sin Lau,[2] Alison Backen,[1] Andrew Clamp,[1] Graham Rushton,[1] Caroline Dive,[3] Cassandra Hodgkinson,[3] Rhona McVey,[4] Henry Kitchener[5] and Gordon C. Jayson[1]

[1]Cancer Research UK and University of Manchester Dept. Translational Angiogenesis; Paterson Institute; Withington, Manchester UK; [2]Department of Oncology; Blackpool Victoria Hospital; Blackpool, UK; [3]Cancer Research UK and University of Manchester Clinical and Experimental Pharmacology Group; Paterson Institute; Withington, Manchester UK; [4]Department Gynaecological Histopathology; and [5]Dept. Gynaecological Oncology; St. Marys Hospital; Manchester, UK

Key words: ovarian cancer, fibroblast growth factor, fibroblast growth factor receptor, shRNAi, cisplatin

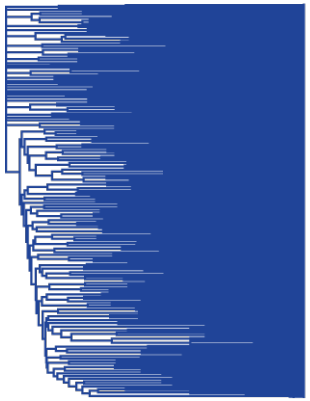


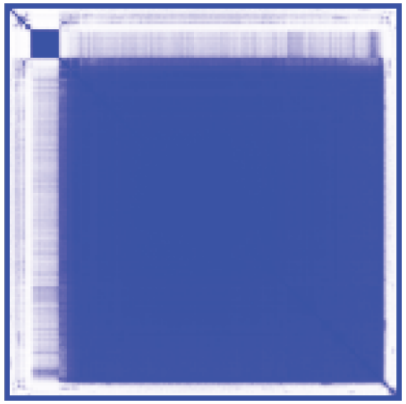
Conclusions

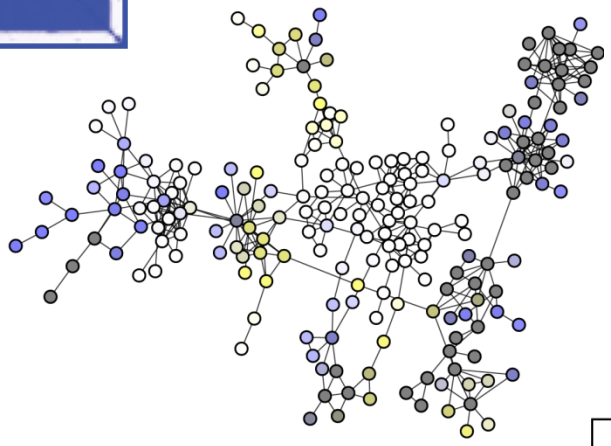
- Network-based stratification recovers biologically relevant subtypes of ovarian cancer.
- Somatic mutation subtypes are different from those recovered from other molecular profiles.
- These subtypes can be recapitulated using gene expression.
- Each subtype seems to have specific effected subnetworks.

One slide summary

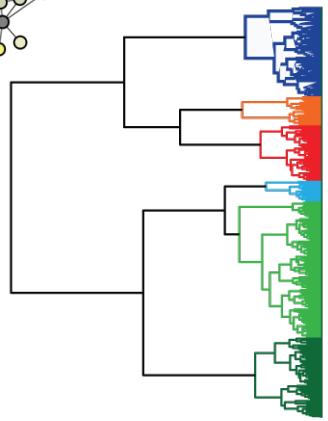
Regular Consensus Clustering NMF

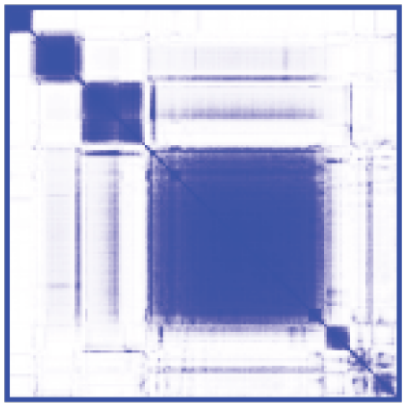






Network based stratification





Acknowledgements

Ideker Lab:

- John P. Shen

- Janusz Dutkwoski

- Andy Gross
- Rohith Srivas
- Gordon Bean

Moores cancer center:

- Stephen Howell
- Kelly Frazer
- Jean Wang

UCSD Computer Science:

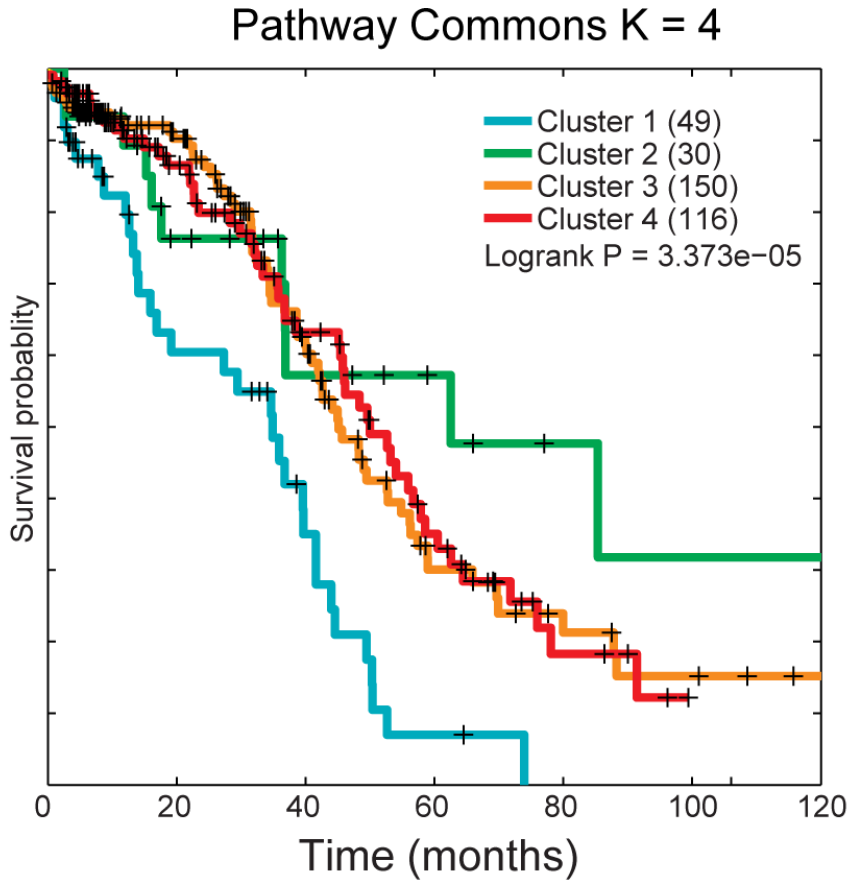
- Lawrence Saul

Trey Ideker

UCSD's Geisel Library

Thank you for listening…

A similar clustering from different networks

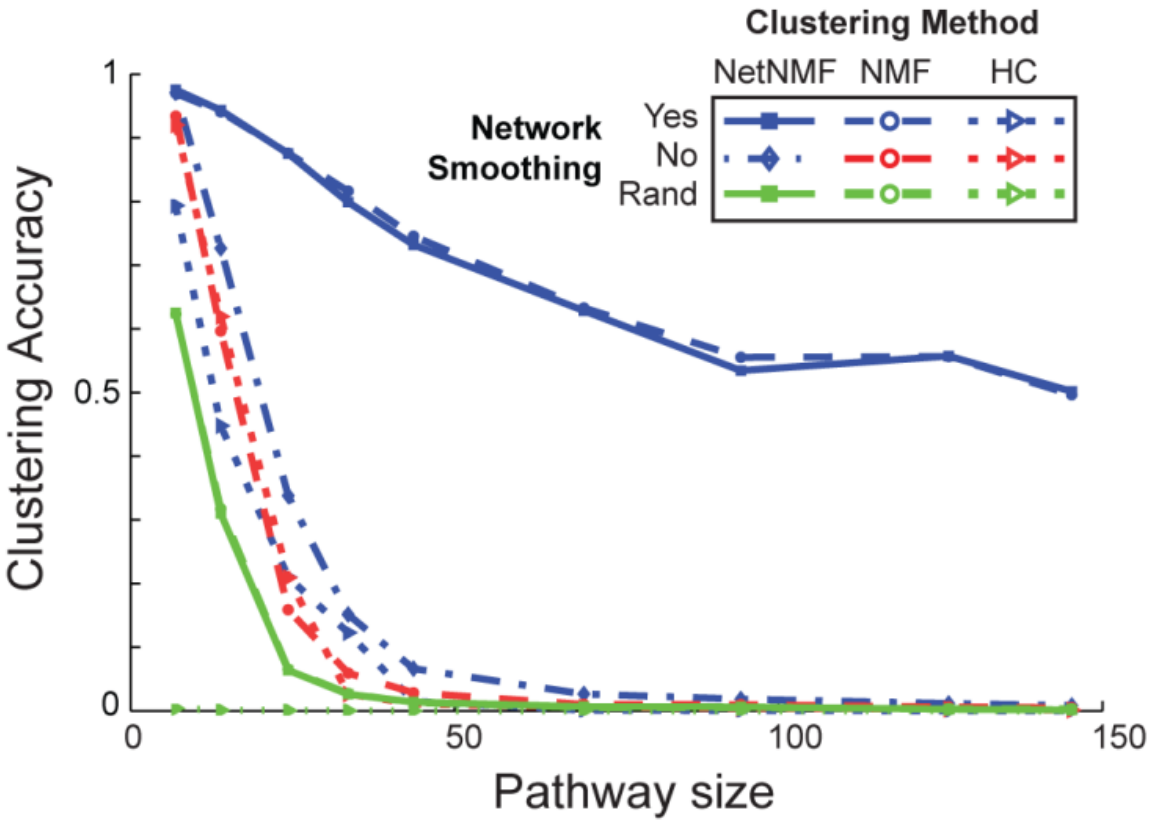


Pathway Commons

HumanNet	1	2	3	4
1	**35**	4	13	5
2	5	22	**104**	**95**
3	0	3	4	6
4	1	0	12	5

p-value (χ^2) =3.98×10^{-27}

Simulation results



Network regularized NMF

- NMF has been 'augmented' with many forms of regularizations:

$$\min_{W,H>0} \|X - WH\|_F^2 + \alpha\|W\| + \beta\|H\|$$

- We suggest adding a term for 'network sparsity' of W. Let $\boldsymbol{K}$ be the graph laplacian of a nearest neighbors graph induced by given network.

$$\min \|X - WH\|_F^2 + \rho \cdot trace(W^T K W)$$

Potential clinical covarites