Network-based stratification of tumor mutations

Matan Hofree

Goal

- Tumor stratification: to divide a heterogeneous population into clinically and biologically meaningful subtypes based on molecular profiles

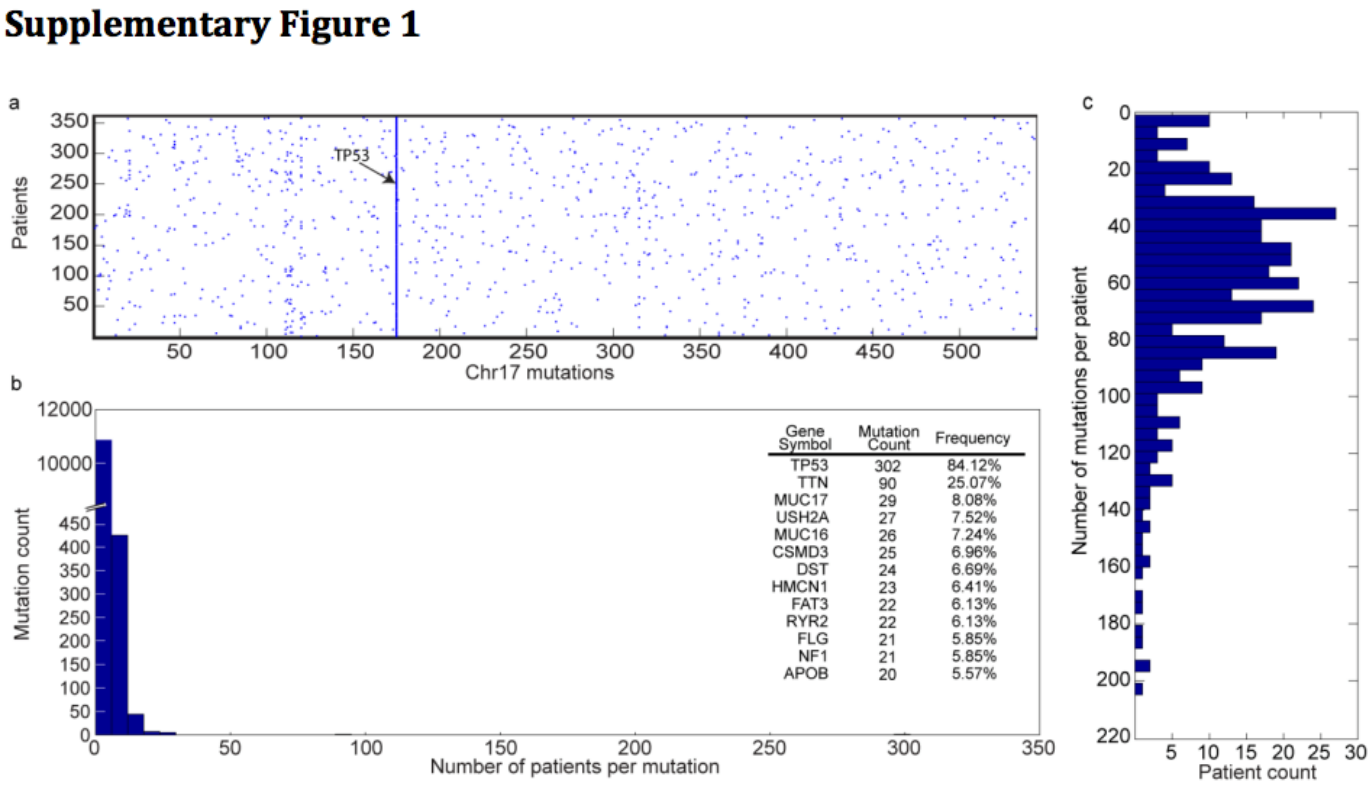
Previous attempts

- Glioblastma and breast cancer – mRNA expression data

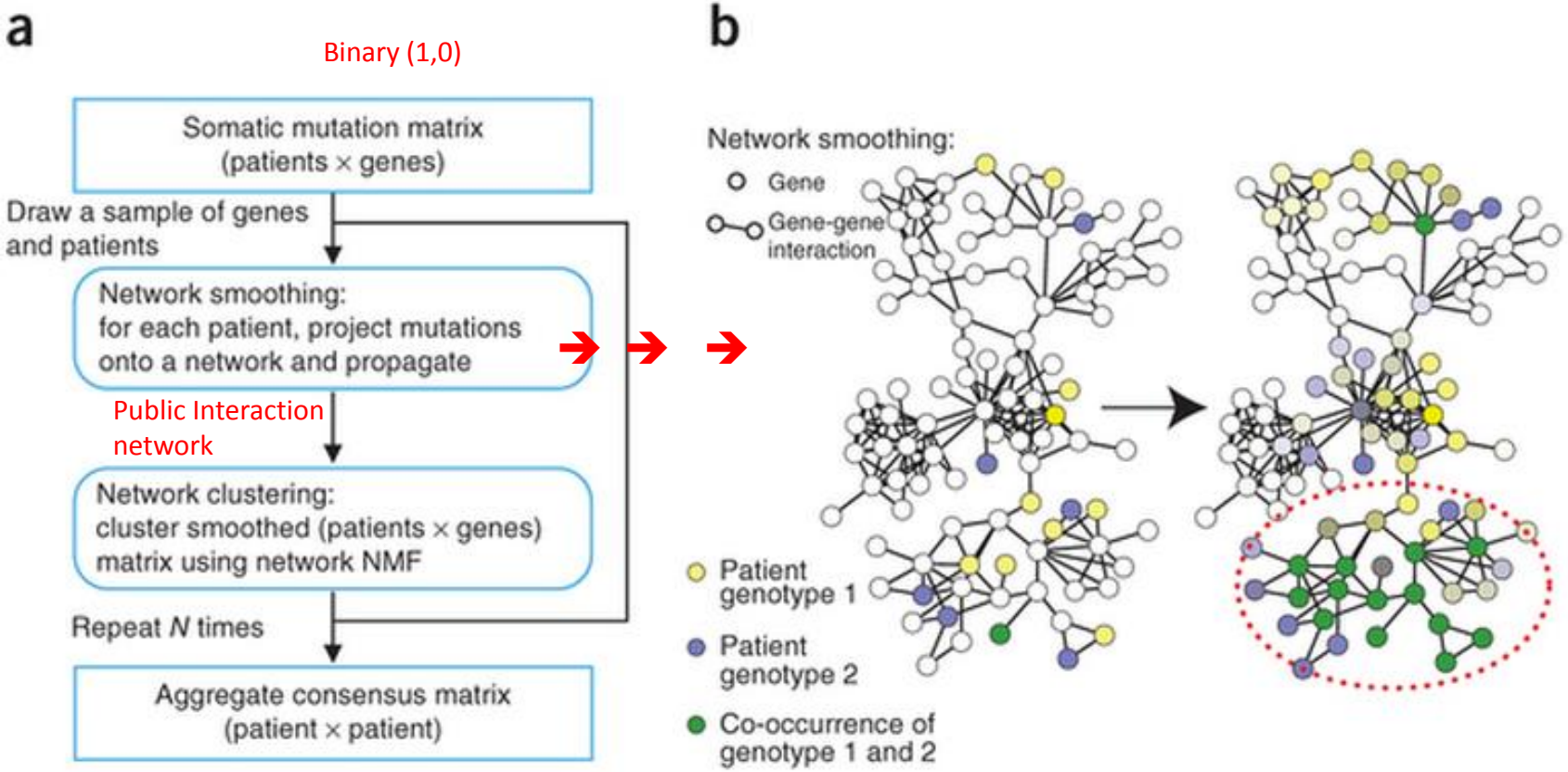
- Colorectal adenocarcinoma and small-cell lung cancer – expression data **not** correlate with clinical phenotype

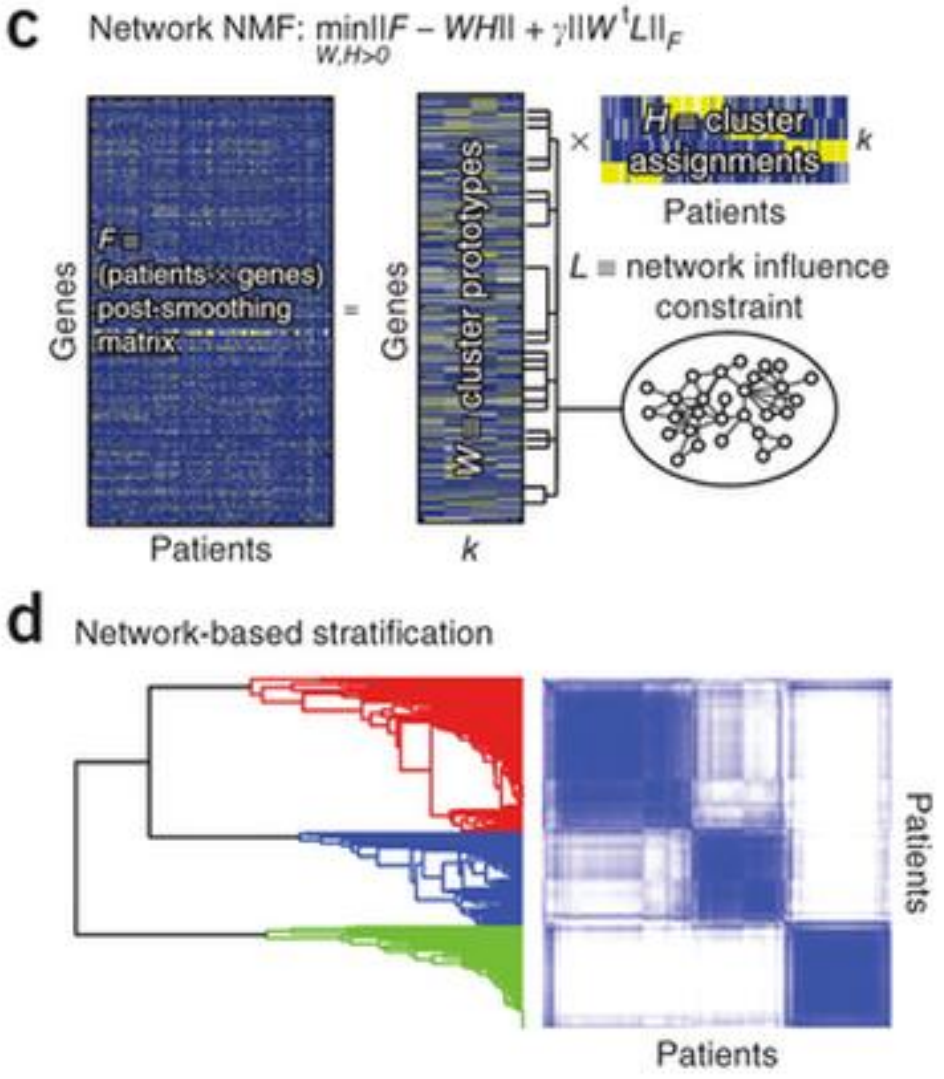
Somatic mutation profile

- Compare the genome or exome of a patient's tumor to that of the germ line
- Sparse



Overview of network-based stratification




c
Network NMF: $\min_{W,H>0} \|F - WH\| + \gamma \|W^T L\|_F$
F ≡ (patients × genes) post-smoothing matrix
Genes
Patients
=
W ≡ cluster prototypes
Genes
k
×
H ≡ cluster assignments
k
Patients
L ≡ network influence constraint
d
Network-based stratification
Patients
Patients

Network smoothing

- $F_{t+1} = \alpha F_t A + (1-\alpha)F_0$

F_0: patients * genes matrix

A: adjacency matrix of the gene interaction network (STRING, HumanNet and PathwayCommons)

α: tuning factor that determines how far a mutation signal can diffuse

Network-regularized NMF

- Min $|| F - WH ||^2 + \text{trace}(W^tKW)$

Patient * gene matrix

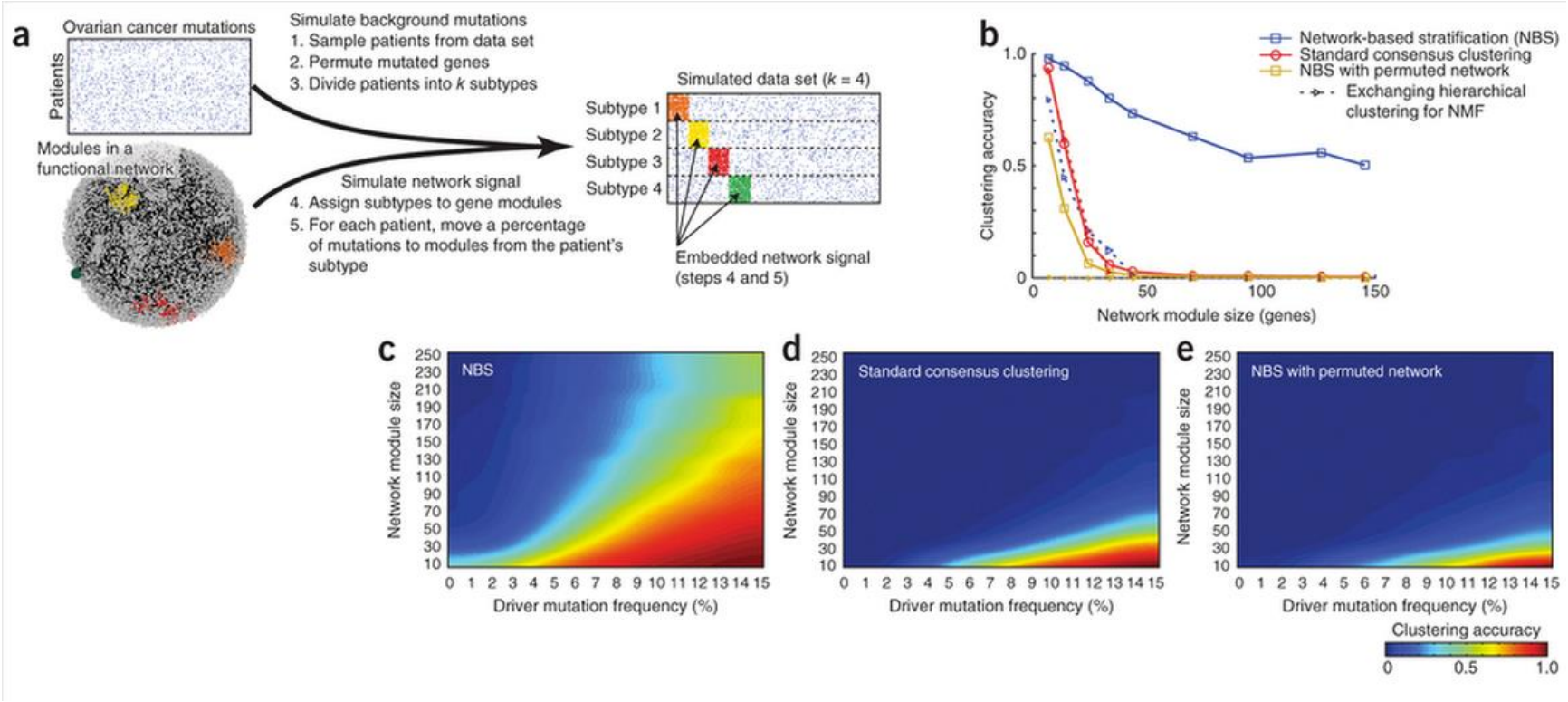
W: a collection of basis vectors, “metagenes”
H: the basis of vector loading
Trace(W^tKW): constrain the basis vectors(W) to respect local network neighborhoods
K: derived from the original network

Simulation Assessment

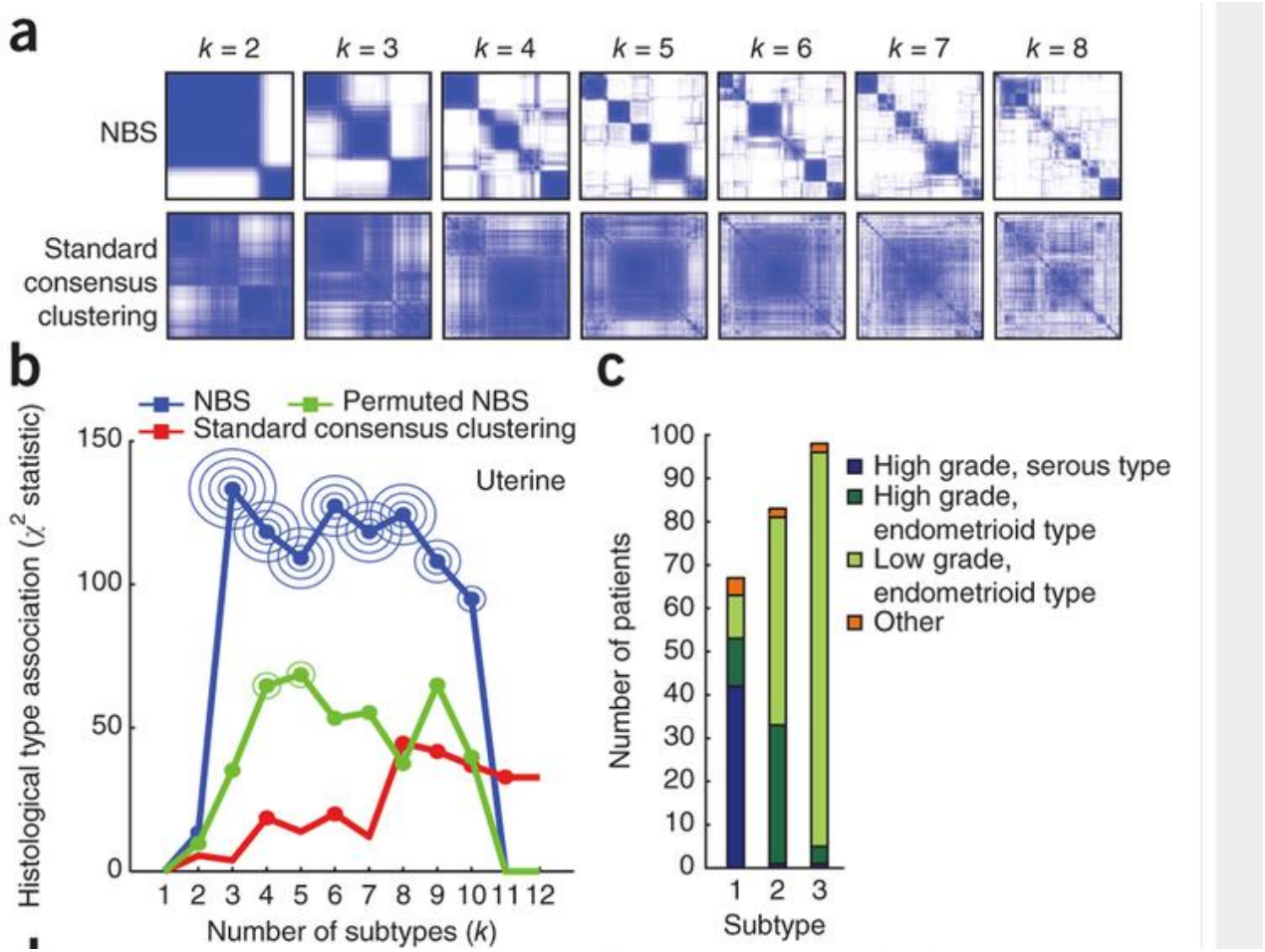
K=4

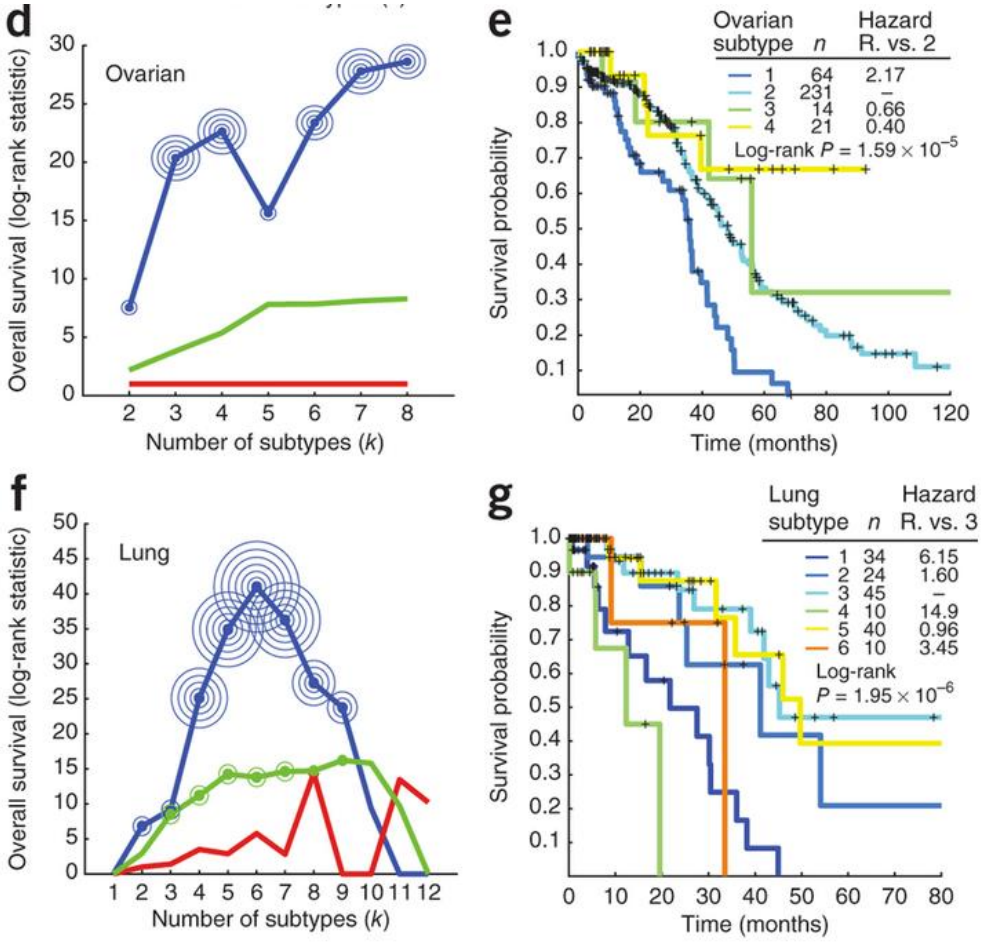
Driver mutation f: 0% to 15%

The size of network modules: 10-250

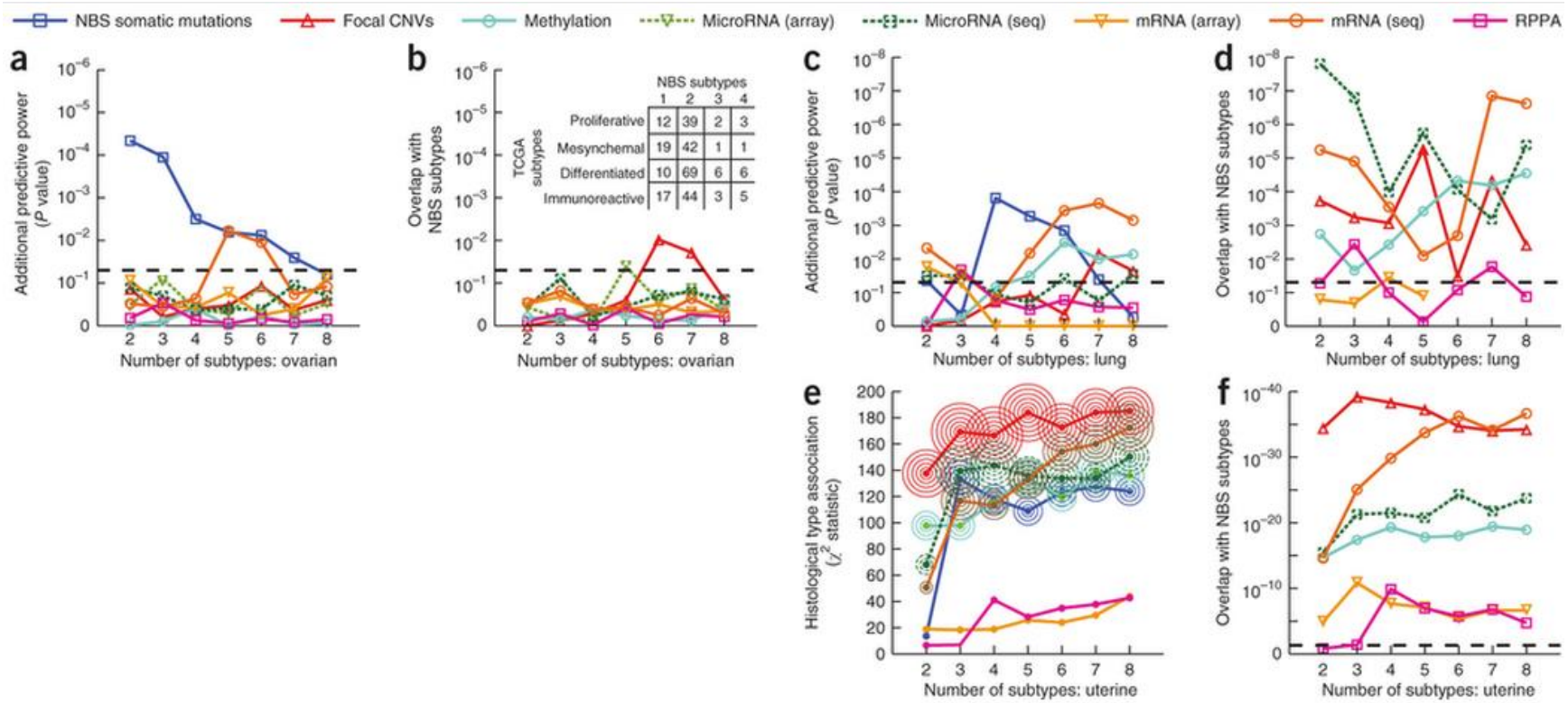


Results- NBS of somatic tumor mutations

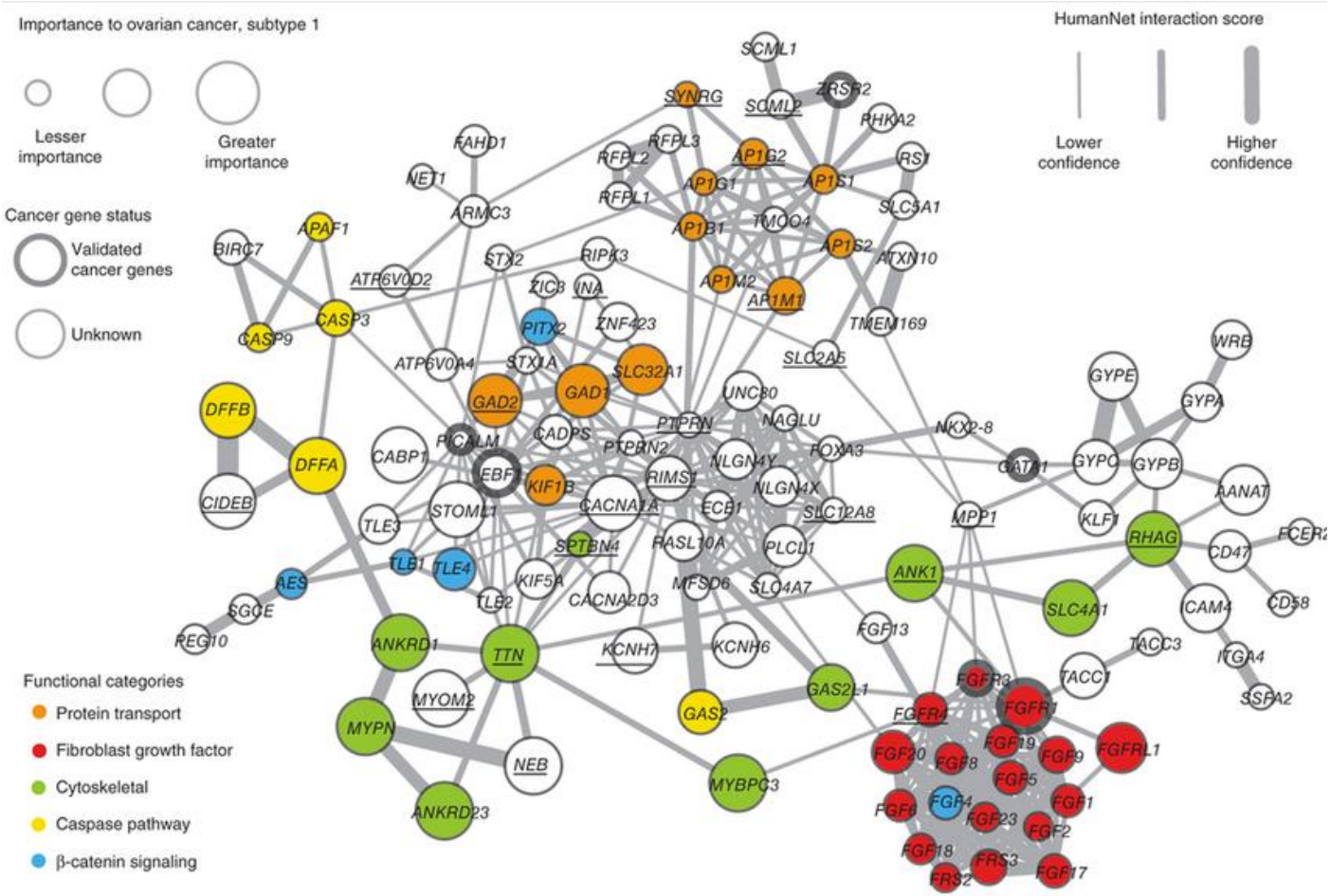



d
Ovarian
Overall survival (log-rank statistic)
Number of subtypes (k)
e
Ovarian subtype n Hazard R. vs. 2
1 64 2.17
2 231 –
3 14 0.66
4 21 0.40
Log-rank P = 1.59 × 10−5
Survival probability
Time (months)
f
Lung
Overall survival (log-rank statistic)
Number of subtypes (k)
g
Lung subtype n Hazard R. vs. 3
1 34 6.15
2 24 1.60
3 45 –
4 10 14.9
5 40 0.96
6 10 3.45
Log-rank
P = 1.95 × 10−6
Survival probability
Time (months)

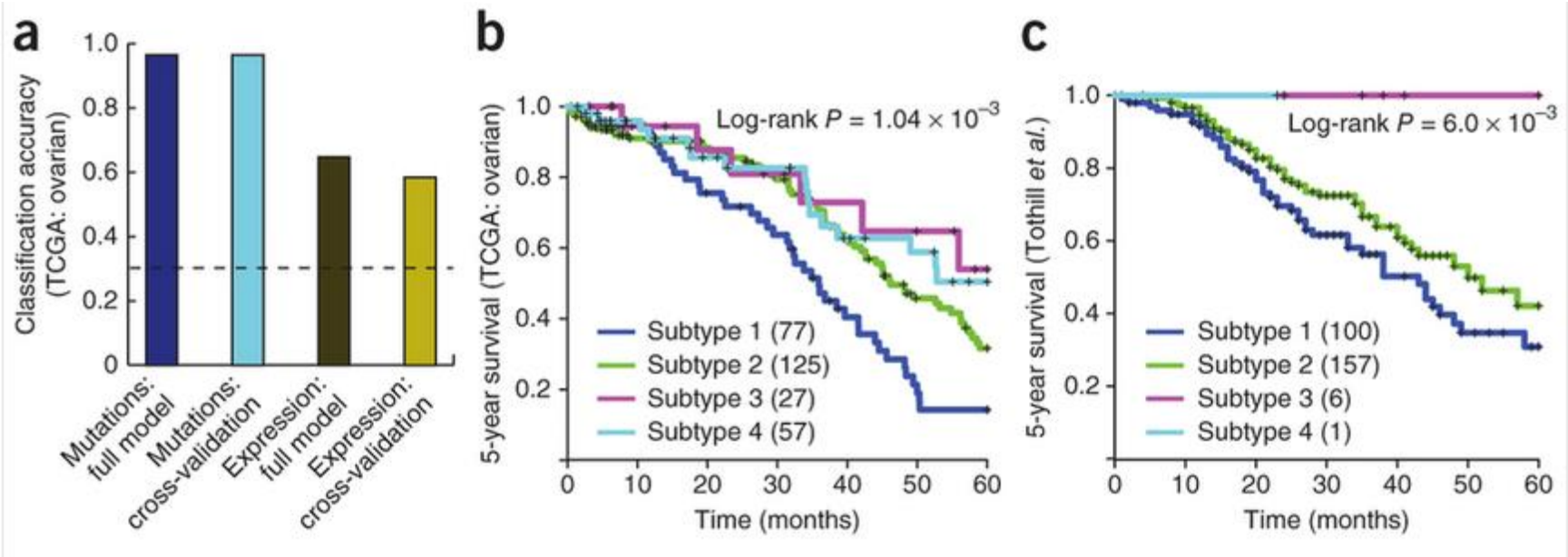
Results-Predictive power and overlap of subtypes derived from different TCGA datasets



Network view of genes with high network-smoothed mutation scores in HumanNet ovarian cancer type 1



From mutation-derived subtypes to expression signatures



Effects of different types of mutations on stratification