SUPPLEMENTAL FIGURES

Figure S1 Use of IGHV mutation status to stratify patient samples. Survival curves on SC→TX for the 17 European patients (A) or for the patient cohort in Friedman et al (2009) (B).

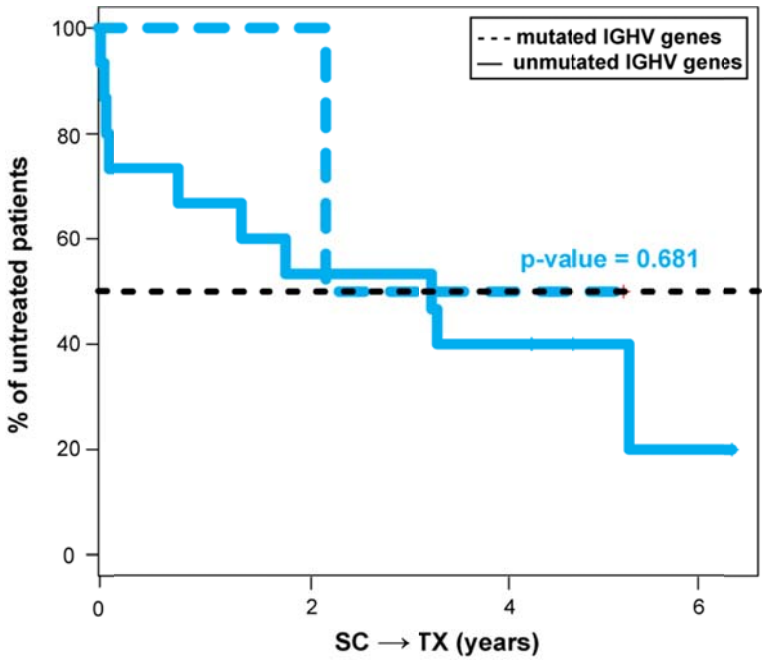
Figure S2 Detection of known disease genes with cancer susceptibility. The enrichment of disease genes is shown for the 38 subnetworks or top 230 genes selected from the 130 patients from UCSD, or for random subnetworks of the same size as the identified 38 subnetworks, but without regard to the expression profiles. Bars chart the percentage of disease genes among all genes covered in the markers. Numbers above the bars are the hypergeometric P-values of enrichment.

Figure S3 Schematic overview of subnetwork identification. Protein interaction networks are used to assign sets of genes to discrete subnetworks. Gene expression profiles of tissue samples are transformed into a "subnetwork activity matrix". For a given subnetwork M_k in the interaction network, the activity is a combined z-score derived from the expression of its individual genes. After overlaying the expression vector of each gene on its corresponding protein in the interaction network, subnetworks with discriminative activities are found via a greedy search. Significant subnetworks are selected based on null distributions estimated from permuted subnetworks. Subnetworks are then used to identify disease genes, and the subnetwork activity matrix is used to train a classifier for prognosis of newly diagnosed patients.

Figure S4 Enriched biological processes in the significantly predictive subnetowrks. The 38 subnetworks are enriched for proteins functioning in a common biological function as annotated by Gene Ontology database (hypergeometric test with a false discovery rate of 5%). Enriched terms from the Biological Process category, are depicted in the top.

Figure S1

A.


mutated IGHV genes
unmutated IGHV genes
p-value = 0.681
% of untreated patients
SC ⟶ TX (years)
100
80
60
40
20
0
0
2
4
6

B.

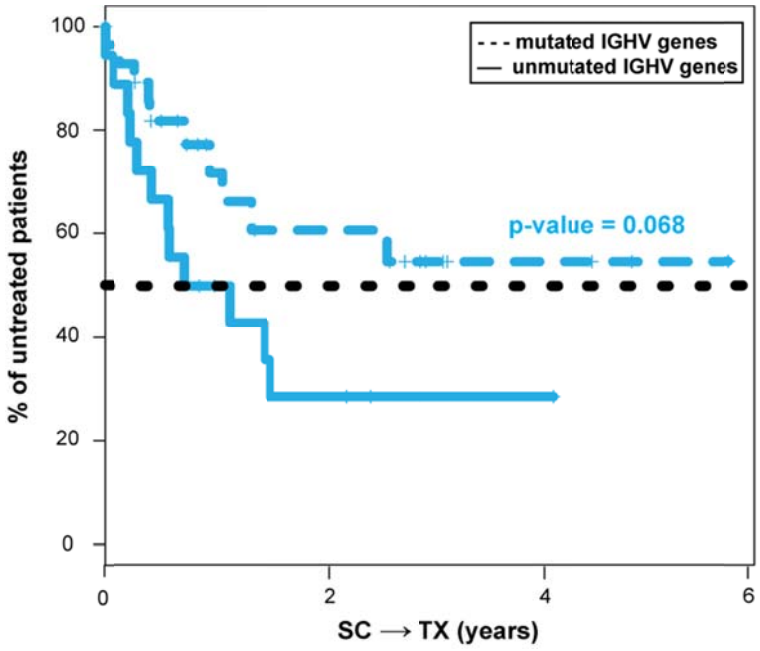

mutated IGHV genes
unmutated IGHV genes
p-value = 0.068
% of untreated patients
SC ⟶ TX (years)
100
80
60
40
20
0
0
2
4
6

Figure S2

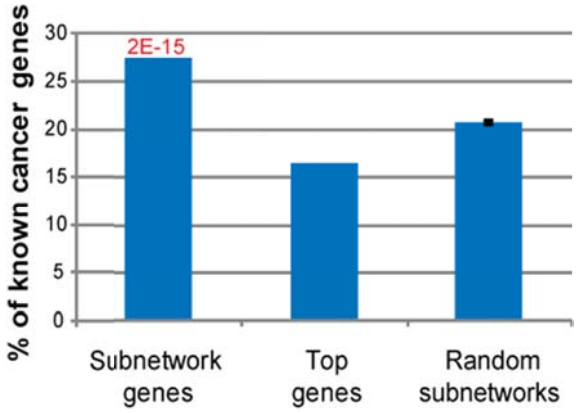

% of known cancer genes
30
25
20
15
10
5
0
2E-15
Subnetwork genes
Top genes
Random subnetworks

Figure S3

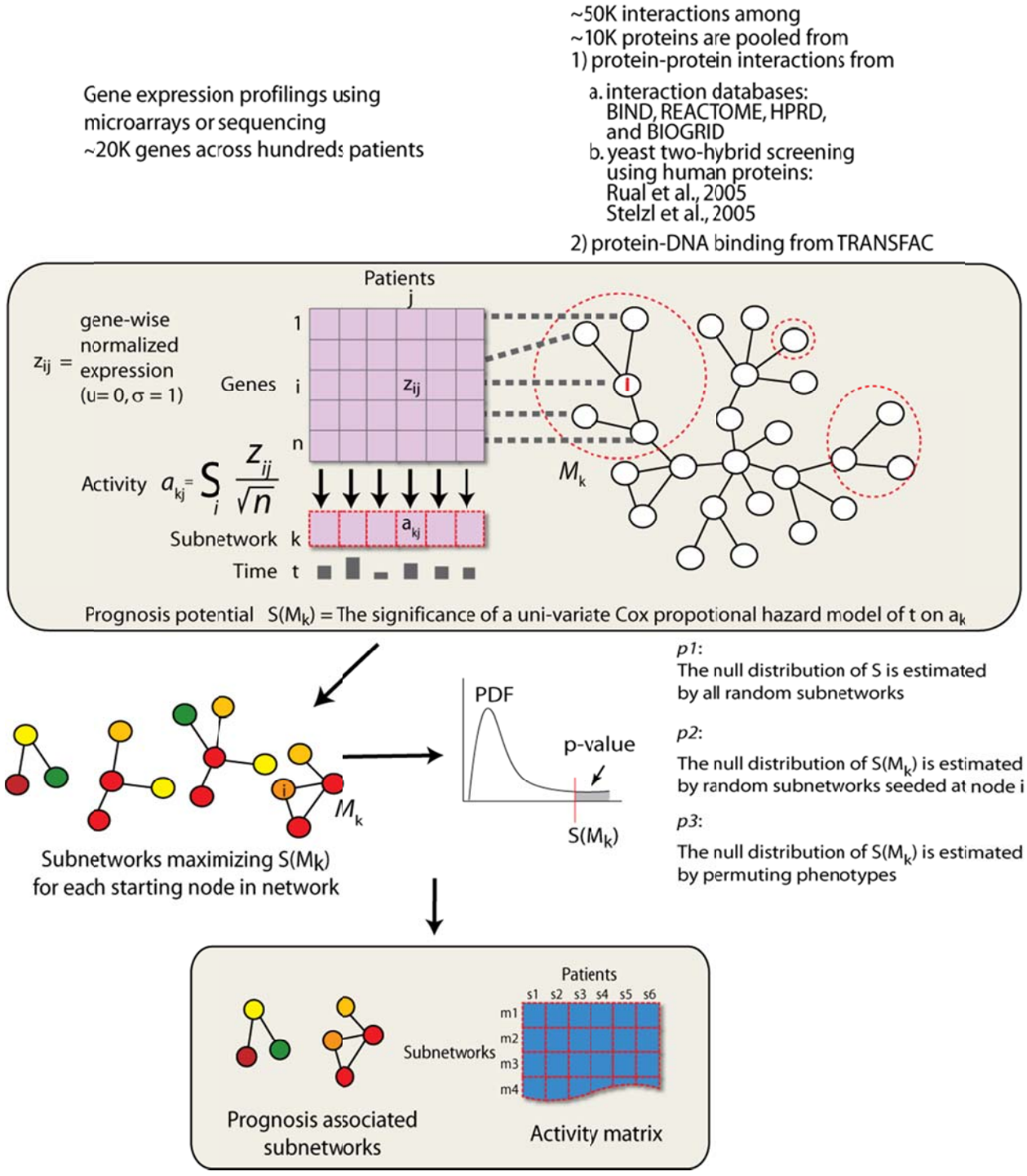
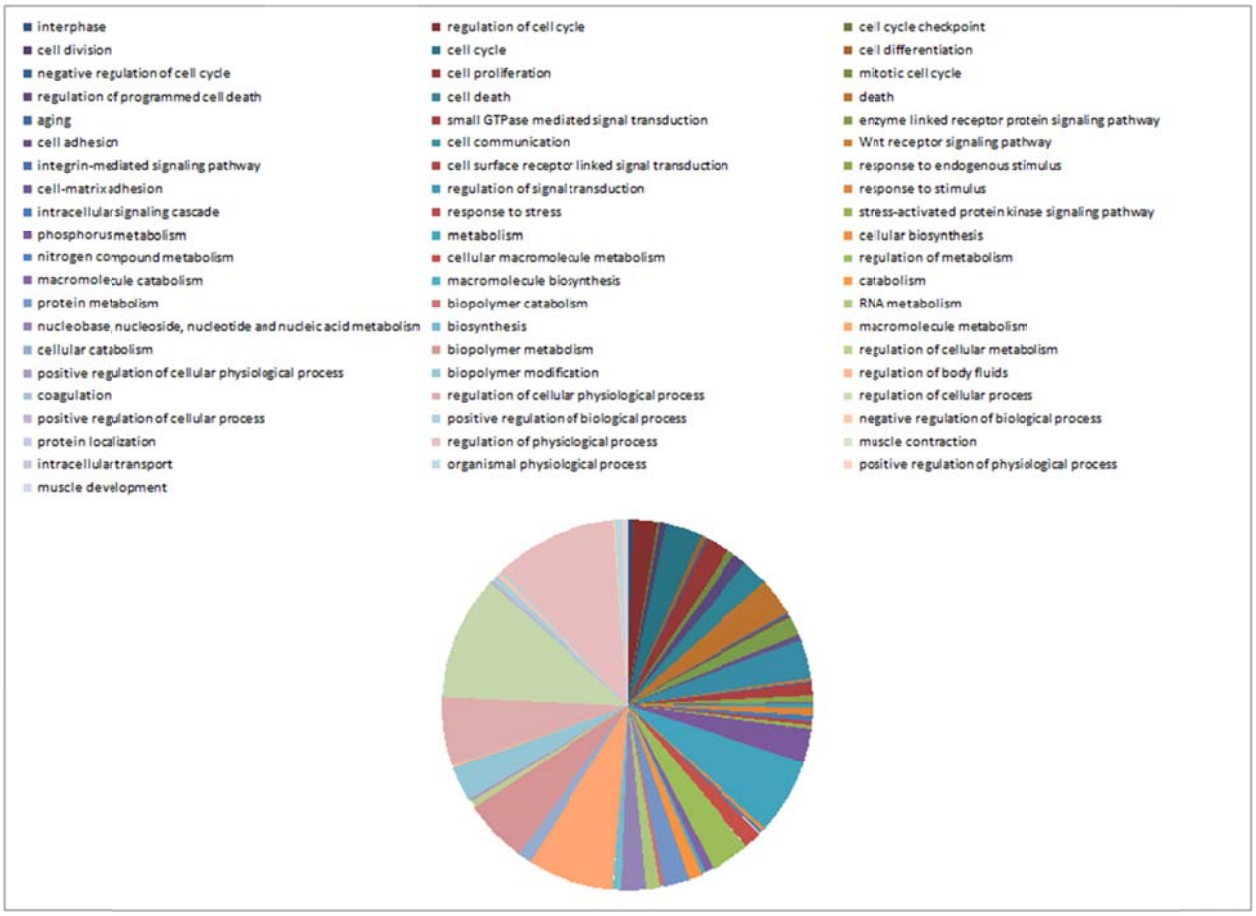

~50K interactions among
~10K proteins are pooled from
1) protein-protein interactions from
a. interaction databases:
BIND, REACTOME, HPRD,
and BIOGRID
b. yeast two-hybrid screening
using human proteins:
Rual et al., 2005
Stelzl et al., 2005
2) protein-DNA binding from TRANSFAC
Gene expression profilings using
microarrays or sequencing
~20K genes across hundreds patients
Patients
j
gene-wise
normalized
expression
(u= 0, σ = 1)
Genes
Activity
Subnetwork k
Time t
Prognosis potential S(Mk) = The significance of a uni-variate Cox propotional hazard model of t on ak
PDF
p-value
S(Mk)
Subnetworks maximizing S(Mk)
for each starting node in network
p1:
The null distribution of S is estimated
by all random subnetworks
p2:
The null distribution of S(Mk) is estimated
by random subnetworks seeded at node i
p3:
The null distribution of S(Mk) is estimated
by permuting phenotypes
Patients
s1 s2 s3 s4 s5 s6
m1
m2
m3
m4
Subnetworks
Prognosis associated
subnetworks
Activity matrix

Figure S4


interphase
cell division
negative regulation of cell cycle
regulation of programmed cell death
aging
cell adhesion
integrin-mediated signaling pathway
cell-matrix adhesion
intracellular signaling cascade
phosphorus metabolism
nitrogen compound metabolism
macromolecule catabolism
protein metabolism
nucleobase, nucleoside, nucleotide and nucleic acid metabolism
cellular catabolism
positive regulation of cellular physiological process
coagulation
positive regulation of cellular process
protein localization
intracellular transport
muscle development
regulation of cell cycle
cell cycle
cell proliferation
cell death
small GTPase mediated signal transduction
cell communication
cell surface receptor linked signal transduction
regulation of signal transduction
response to stress
metabolism
cellular macromolecule metabolism
macromolecule biosynthesis
biopolymer catabolism
biosynthesis
biopolymer metabolism
biopolymer modification
regulation of cellular physiological process
positive regulation of biological process
regulation of physiological process
organismal physiological process
cell cycle checkpoint
cell differentiation
mitotic cell cycle
death
enzyme linked receptor protein signaling pathway
Wnt receptor signaling pathway
response to endogenous stimulus
response to stimulus
stress-activated protein kinase signaling pathway
cellular biosynthesis
regulation of metabolism
catabolism
RNA metabolism
macromolecule metabolism
regulation of cellular metabolism
regulation of body fluids
regulation of cellular process
negative regulation of biological process
muscle contraction
positive regulation of physiological process

SUPPLEMENTAL METHODS

Selecting significant subnetworks. To assess the significance of the identified subnetworks, **M**s, three tests of significance are performed. For the first test, we perform the same search procedure over 100 random trials in which the expression vectors of individual genes are randomly permuted on the network. Such permutation disrupts the correlation between expression and interaction. The score **S**(**M**), –log *p*-value of a χ^2 test on a Cox proportional hazard model on *T*, of each real subnetwork is indexed on the 'global' null distribution of all random subnetwork scores. The second test indexes each real subnetwork score on a 'local' null distribution, estimated from the **S**(**M**) scores of 100 random subnetworks initialized from the same seed protein as the real subnetwork. Third, we test whether **S**(**M**) of a Cox proportional hazard model on real *T* is stronger than that obtained with random assignments of *T* to patients. For the random model, these assignments are permuted in 20,000 trials, yielding a null distribution of mutual information scores for each trial; the real score of each subnetwork is indexed on this null distribution. In this study, significant subnetworks are selected that satisfy all three tests with $P_1<0.05$, $P_2<0.05$, and $P_3<0.00005$, according to the three different null distributions of **S**(**M**).

DNA primers for Real time PCR in serial gene expression experiments

1. **TCEB3**

Forward 5'-	TTTGCCAGGGACCTAGTGG	-3'
Reverse 5'-	CGCTTTCGGGAATTGCTCT	-3'

2. **MED9**

Forward 5'-	CCTTTGGTTCACAACATCATCAA	-3'
Reverse 5'-	CTGGAACTTGCTTTTGAGGG	-3'

3. **CEBPA**

Forward 5'-	CCACGCCTGTCCTTAGAAAG	-3'
Reverse 5'-	CCCTCCACCTTCATGTAGAAC	-3'

4. **CEBPB**

Forward 5'-	GGCCCTGAGTAATCGCTTAAAG	-3'
Reverse 5'-	TCCCAAAATATACAGACGCCTC	-3'

5. **CSPG6**

Forward 5'-	CCCCAGGAAGCATTTGAAAAG	-3'
Reverse 5'-	CTGCTCGGAGAAATTTACAAACTG	-3'

6. **PFTK1**

Forward 5'-	TGGCCTGGAGTTCATTCTTTAC	-3'
Reverse 5'-	AACATTGTAGGAGCTTGGAGG	-3'

7. ACVR1

Forward 5'-	GAAGATATGAGGAAGGTAGTCTGTG	-3'
Reverse 5'-	AGTGCTGTGAGTCTTGCG	-3'

8. FKBP4

Forward 5'-	CAATATGTTTGAGAGGCTGGC	-3'
Reverse 5'-	CTATGCTTCTGTCTCCACCTG	-3'

9. DYNLL1

Forward 5'-	ACATAGAGAAGGACATTGCGG	-3'
Reverse 5'-	GCCCAGGTAGAAGTAGATGAAG	-3'

10. SMAD2

Forward 5'-	GCCGTCTATCAGCTAACTAGAATG	-3'
Reverse 5'-	TTTGTCCAACCACTGTAGAGG	-3'

11. IRAK2

Forward 5'-	AAGCGAGTGGACATCTTCAG	-3'
Reverse 5'-	CTGCTTGGAATATCACTGAGGA	-3'

12. SUPT3H

Forward 5'-	ATTTCGAGACTGGTTGGACTG	-3'
Reverse 5'-	GGTTACCATGTCTTGCCTCAC	-3'

13. CSNK2A1

Forward 5'-	TTCAGTGCCAACCCCTTC	-3'
Reverse 5'-	AGGCATCAGGAGACAGATAGG	-3'

14. SKP2

Forward 5'-	CCAACACCTATCACTCAGTCG	-3'
Reverse 5'-	TCTGTATGTTTGAGGGCATCC	-3'

15. CDC26

Forward 5'-	GACGGAAACCAACACGCCTA	-3'
Reverse 5'-	GCCTCCTACAACTTCCACATC	-3'

16. TNFRSF7

Forward 5'-	GCTCCGATTTTATTCGCATCC	-3'
Reverse 5'-	TGTAACGACAAGGCTCTGC	-3'

17. MCP

Forward 5'-	CCTCCATCTAGTACAAAACCTCC	-3'
Reverse 5'-	CACAGCAATGACCCAAACATC	-3'

18. DMD

Forward 5'-	AGAAATACCCCTGGAAAGCC	-3'
Reverse 5'-	TTCTGCTCCTTCTTCATCTGTC	-3'

19. CCT4

Forward 5'-	CAGAACTAAGAAACCGGCATG	-3'
Reverse 5'-	TCAGTTGCAAGAGTCAGAGC	-3'

20. CCT7

Forward 5'-	ATGCCCACACCAGTTATCCTA	-3'
Reverse 5'-	CAGGGTAGTTCTTACAGCCTCA	-3'

21. CREB3

Forward 5'-	CCTTGTACCTGCTATGTACTCC	-3'
Reverse 5'-	TCTTTCGGCACTTCTGACTG	-3'