Supplementary Table 1. Patient characteristics from Director's Challenge study (Shedden et al., 2008)

	UM & HLM (Training, *n*=256)	MSK (Test 1, *n*=104)	DFCI (Test 2, *n*=82)
Median follow-up (months)	48	44	50
Age (mean, s.d.)	65 (10)	65 (10)	61 (10)
Sex (% male)	55%	36%	55%
Tumor Stage			
Stage I	61% (157/256)	61% (63/104)	68% (56/82)
Stage II	19% (49/256)	19% (20/104)	32% (26/82)
Stage III	18% (47/256)	20% (21/104)	-
Stage IV	-	-	-
Unknown	1% (3/256)	-	-

Supplementary Table 2. Summary of prognostic gene signatures discovered using the proposed methodology.

Gene Selection Approach	Number of signaling hallmarks used in gene selection	Number of gene signatures generating significant stratification (log-rank $P < 0.05$; Kaplan-Meier analyses) in all datasets		
		with significant hazard ratio in all stages	with significant hazard ratio in all stages and stage I	with significant hazard ratio in all stages, stage I, and stage I not receiving chemotherapy
Network-based (Approach 1)	7	4	1	0
	6	9	5	4
Network + Random Forests (Approach 2)	7	4	4	1
	6	3	2	0
Network + Relief (Approach 3)	7	7	4	0
	6	47	26	16

Supplementary Table 3. Prognostic signatures identified with Approach 1 that generated significant stratifications in patients with all stages, stage I only, and stage I without receiving chemotherapy.

No.	Concurrently Coexpressed Signaling Hallmarks	Signature genes
S1	MET, EGF, KRAS, RB1, E2F1, E2F5	HSPA9, PRDX6, SUPT7L, LEPROT, MPI, QPCT, SLC39A8, ADH1B, MTX1, RAD17, HIPK1, ZFR, CLIC2, TFPI, HEXA, LYST, DYNLRB1, GCC1, CPEB1, ATP1A1, ABHD11
S2	EGF, EGFR, KRAS, TP53, E2F3, E2F4	TOMM34, RPS6KA1, ADD2, MPPED1, DNAJC4, IL12RB2, ICA1, THY1, LOC399491, FHL1, WDR43, LRRC23, MRPL13, ZC3H7A, GRHL2, APOA2, CPEB1, LOC100294391, ATP1A1
S3	EGF, KRAS, TP53, E2F1, E2F2, E2F4	EEF1B2, TOMM70A, TOMM34, IRF3, DDT, RPS6KA1, SC65, SMAD3, PPM1E, MOCS3, DNAJC4, DNAJA2, GRK6, ZNF592, THY1, FHL1, ACTA2, GRM8, GRHL2, APOA2, CPEB1, FBXO31, PDCD1LG2, HDLBP
S4	EGF, KRAS, TP53, E2F1, E2F2, E2F5	PRDX6, ANXA6, TOMM70A, TOMM34, IRF3, RPS6KA1, KATNA1, MPHOSPH9, CCDC9, ZNF141, SCNN1G, DNAJA2, ABCF2, HBS1L, APLP1, ITCH, MTX1, GRK6, NUP214, ANXA9, ELN, ZFR, ZNF592, ACTA2, GRM8, NRN1, APOA2, CPEB1, PDCD1LG2, MUM1, HDLBP, RING1

Supplementary Table 4. Prognostic signatures identified with Approach 2 that generated significant stratifications in patients with all stages, stage I only, and stage I without receiving chemotherapy.

No.	Concurrently Coexpressed Signaling Hallmarks	Signature genes
S5	MET, EGFR, E2F2, KRAS, TP53, E2F1, E2F3	CD86, LHX2, GBX1, HEMK1, CPEB1

Supplementary Table 5. Prognostic signatures identified with Approach 3 that generated significant stratifications in patients with all stages, stage I patients only, and stage I patients without receiving chemotherapy.

No.	Concurrently Coexpressed Signaling Hallmarks	Signature genes
S6	MET, EGF, EGFR, KRAS, TP53, E2F3	CD86, ICA1, RPAP3, CPEB1
S7	MET, EGF, EGFR, KRAS, E2F2, E2F3	ANXA6, SLC17A7, CD86, GAS7, TAF4, ARNT, CPEB1
S8	MET, EGF, KRAS, TP53, E2F1, E2F2	EEF1B2, SNRPD2, PRDX6, ANXA6, TOMM70A, NIPSNAP1, IL13RA1, IRF3, DDT, ABCC4, RPS6KA1, SMAD3, CD86, CCDC9, OPRL1, CLDN6, DNAJA2, CCL19, MTX1, MAPK9, ANXA9, ZFR, THY1, SFRS2B, IVD , MKRN2, GRHL2, CPEB1, FBXO31, PDCD1LG2, C20orf30, MUM1, OR1F1□
S9	MET, EGF, KRAS, E2F1, E2F3, E2F5	HSPA9, ANXA6, MPI, ACTL6A, RPS6KA1, RTCD1, SLC12A2, CCDC9, NDUFAF3, FLT3LG, ANXA9, ZFR, CLIC2, SOSTDC1, TRMU, TCF3 , DYNLRB1, CPEB1, C20orf46, LOC100294391, ATP1A1, MUM1, ABHD11
S10	EGF, EGFR, KRAS, TP53, RB1, E2F2	RPL18, VIPR2, MOCS3, DNAJC4, ADAMTSL3, WDR12, HDLBP
S11	EGF, EGFR, KRAS, TP53, E2F3, E2F4	TOMM34, RPS6KA1, ADD2, MPPED1, DNAJC4, IL12RB2, ICA1, THY1, LOC399491, FHL1, WDR43, LRRC23, MRPL13, ZC3H7A, GRHL2, APOA2, CPEB1, LOC100294391, ATP1A1
S12	EGF, EGFR, TP53, RB1, E2F1, E2F2	MOCS3, DNAJC4, CCBP2, THY1, SFRS2B, PUM2, HDLBP
S13	EGF, KRAS, TP53, RB1, E2F1, E2F2	PRDX6, MOCS3, OPRL1, HBS1L, MTX1, ZFR, SPIN1, CPEB1, OR1F1, HDLBP
S14	EGF, KRAS, TP53, RB1, E2F1, E2F4	DDT, MOCS3, MPPED1, DNAJC4, RGL1, CEP57, THY1, TFPI, LRRC23, MRPL13, CPEB1, FBXO31, ATP1A1, HDLBP, SFTPB
S15	EGF, KRAS, TP53, RB1, E2F1, E2F5	RPL30, PRDX6, SNX2, LEPROT, MPI, KATNA1, SLC39A8, HBS1L, MTX1, ELN, ZFR, ANGEL1 , TFPI, LRRC23, NRN1, SLC35F2, HMBOX1, CPEB1, ATP1A1, GINS2, HDLBP
S16	EGF, KRAS, TP53, E2F1, E2F2, E2F4	EEF1B2, TOMM70A, TOMM34, IRF3, DDT, RPS6KA1, SC65, SMAD3, PPM1E, MOCS3, DNAJC4, DNAJA2, GRK6, ZNF592, THY1, FHL1, ACTA2, GRM8, GRHL2, APOA2, CPEB1, FBXO31, PDCD1LG2, HDLBP
S17	EGF, KRAS, TP53, E2F1, E2F2, E2F5	PRDX6, ANXA6, TOMM70A, TOMM34, IRF3, RPS6KA1, KATNA1, MPHOSPH9, CCDC9, ZNF141, SCNN1G, DNAJA2, ABCF2, HBS1L, APLP1, ITCH, MTX1, GRK6, NUP214, ANXA9, ELN, ZFR, ZNF592, ACTA2, GRM8, NRN1, APOA2, CPEB1, PDCD1LG2, MUM1, HDLBP, RING1
S18	EGF, KRAS, TP53, E2F2, E2F3, E2F5	KIAA0040, KCNS3, KCNA4, COL14A1, CPEB1, RING1
S19	EGF, KRAS, RB1, E2F1, E2F3, E2F5	HSPA9, ABHD11, C9orf156
S20	EGFR, KRAS, RB1, TP53, E2F1, E2F2	TRAP1, PRMT2, MOCS3, DNAJC4, CCL8, TFCP2L1, LOH3CR2A, HDLBP, PKNOX2
S21	EGFR, KRAS, RB1, E2F5, TP53, E2F2	TRAP1, VIPR2, TCP10, TBX1, CCL8, LDLR, WDR12, PRR15L, HDLBP

Supplementary Table 6. Multivariate Cox proportional hazard analysis of the 10-gene risk score and major clinical covariates including gender, age, and tumor stage on the combined testing cohorts (MSK and DFCI)

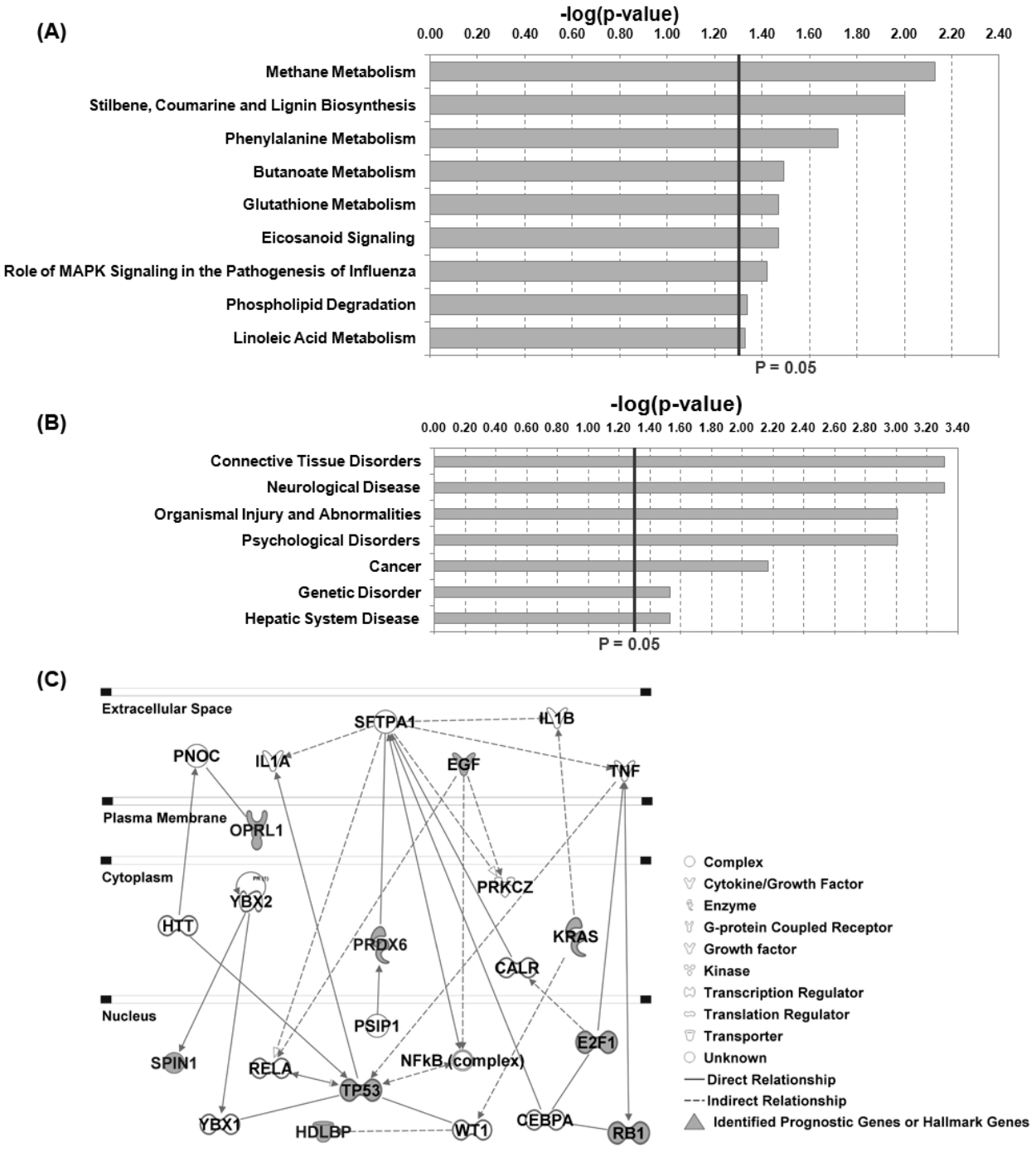
Variable*	***P*-value**	**Hazard Ratio (95% CI)** [Ψ]	
Analysis without 10-gene risk score			
Gender (Male)	0.22	1.34	(0.84,2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95,2.74)
Cancer Stage			
Stage II	6.25E-05	2.91	(1.72,4.91)
Stage III	1.09E-05	4.16	(2.20,7.85)
Analysis with 10-gene risk score			
Gender (Male)	0.28	1.30	(0.81, 2.09)
Age at diagnosis (> 60)	0.09	1.59	(0.93, 2.70)
Cancer Stage			
Stage II	1.62E-04	2.74	(1.62, 4.63)
Stage III	4.58E-06	4.45	(2.35, 8.43)
10-gene risk score	8.61E-04	3.63	(1.70, 7.77)

* Gender was a binary variable (0 for female and 1 for male); age at diagnosis was a binary variable (0 for < 60 years old and 1 otherwise); cancer stage was a categorical variable with 3 categories (Stage I [as the reference group], Stage II, and Stage III).
Ψ denotes confidence interval.

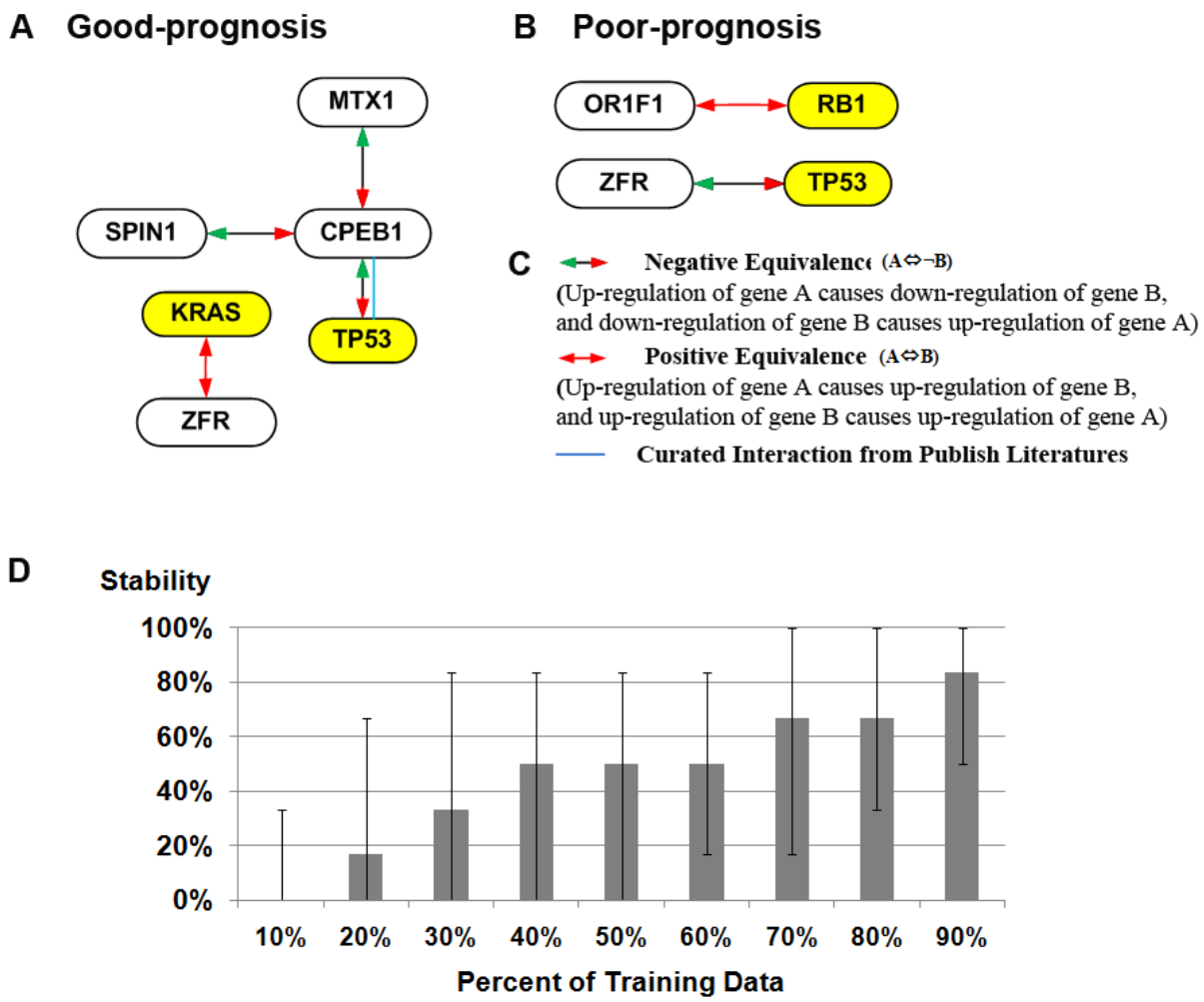
Functional pathway Analysis

Proprietary web-based software Ingenuity Pathway Analysis (IPA, Ingenuity Systems®, www.ingenuity.com) was used to derive curated molecular interactions, including both physical and functional interactions, as well as pathway relevance reported in the literature. The databases and software toolsets weigh and integrate information from numerous sources, including experimental repositories and text collections from published literature. Core analysis was used to identify significant biological processes and functions from the merged network related to the identified 10-gene signature in human tissues and cell lines.

The discovered biomarkers may reveal fundamental molecular mechanisms of this deadly disease, and enhance our understanding of why patients with certain molecular tumor characteristics have a poor clinical outcome and how their outcome could be improved. Functional pathway studies with IPA confirmed the interactions between the major NSCLC signaling pathways and the identified 10-gene signature. Nine canonical pathways were significantly (P<0.05; adjusted with BH tests) associated with the 10 prognostic genes. These pathways include methane metabolism and phenylalanine metabolism related to cell cycle, eicosanoid signaling that mediates inflammation and immunity, and MAPK signaling related to cell death, tissue morphology and inflammatory response (Supplementary Fig. 1A). Cancer is among the top 5 most significant disease and disorders (P<0.05; adjusted with BH tests) in the network related to the 10 prognostic genes (Supplementary Fig. 1B). Furthermore, 4 of the 10 prognostic genes were involved in interactions with major lung cancer signaling proteins, including *TP53*, *KRAS*, *EGF*, *E2F1*, and *RB1* as reported in the literature (Supplementary Fig. 1C). For instance, high density lipoprotein binding protein *HDLBP* is indirectly related to *TP53* through a tumor-suppressor gene *WT1*, as *HDLBP* is regulated by *WT1* (1), which in turn interacts with *TP53* (2). A prognostic gene, peroxiredoxin 6 (*PRDX6*), promotes invasion and metastasis of lung cancer cells (3). Direct interaction of pulmonary surfactant *SP-A* gene and *PRDX6* could be important in the regulation of lung surfactant phospholipid metabolism (4). The computationally derived networks delineated expression patterns among the 10 prognostic genes and the signaling hallmarks specifically associated with the good- and poor-prognosis groups (Supplementary Fig. 2). Specifically, the interaction between *CPEB1* and *TP53* was confirmed in a published study (5). These identified biomarkers associated with lung cancer progression and metastasis could be potential therapeutic targets in future clinical treatments.



Supplementary Figure 1. **Functional pathway analysis of the 10 prognostic genes.** Core analysis was performed with Ingenuity Pathway Analysis (IPA). Significant canonical pathways retrieved from IPA (A). Cancer was a significant biological function in the disease and disorders category (B). Curated interactions related to the 10 signature genes were also revealed from the literature (C).



Supplementary Figure 2. Disease-specific coexpression relations among the 10-gene prognostic signature and the 6 lung cancer signaling hallmarks. The disease-specific expression patterns, i.e., coexpression patterns specific for the good-prognosis group (A) and the poor-prognosis group (B) that were commonly present in both training (UM & HLM) and combined test cohorts (MSK & DFCI) were illustrated. The interpretation of the coexpression patterns is provided in (C). The stability of the networks in (A) and (B) was evaluated by using random subsets of the training samples in 100 iterations (D). The stability is defined as the portion of disease-specific coexpression relations obtained from the original data that are retrieved by using only a random subset of the training data and the full test data.

Disease-specific coexpression networks assessment

Gene co-expression networks derived from the implication networks algorithm was evaluated on precision and false discovery rate (*FDR*). Five gene set collections (positional, curated, motif, computational, and Gene Oncology) and canonical pathway databases from the MSigDB [1] were used to evaluate the biological relevance of computationally derived coexpression relations. A co-expression relation was considered a true positive (TP) if the pair of genes belongs to the same gene set or pathway in any investigated database. If a pair of genes does not share any gene set or pathway, the co-expression relation was considered a false positive (FP). A co-expression relation was labeled as non-discriminatory (ND) if at least one gene in the pair is not annotated in a database (6). Co-expression relations labeled as ND were excluded in the evaluation as they were not confirmed.

Precision and *q*-value of the disease-mediated coexpression networks are defined as:

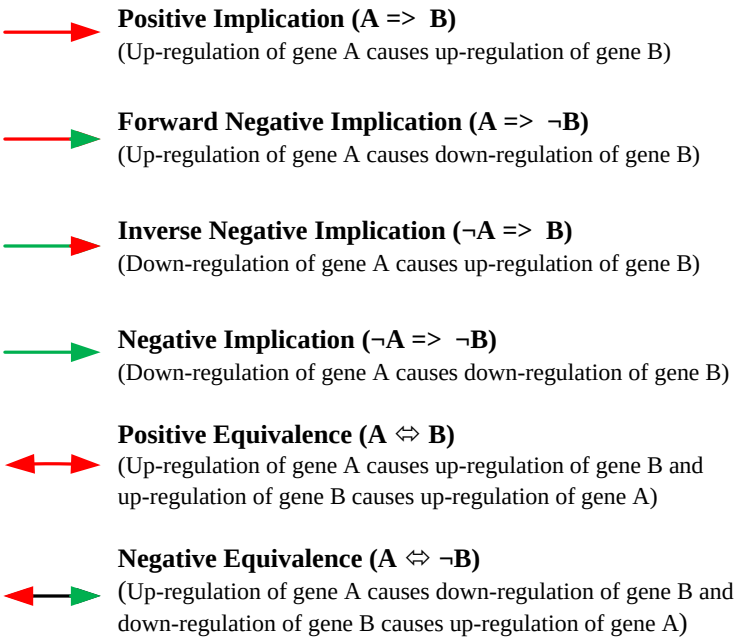
$$Precision = \frac{TP}{TP + FP}$$

$$q - value = \frac{FP}{TP + FP}$$

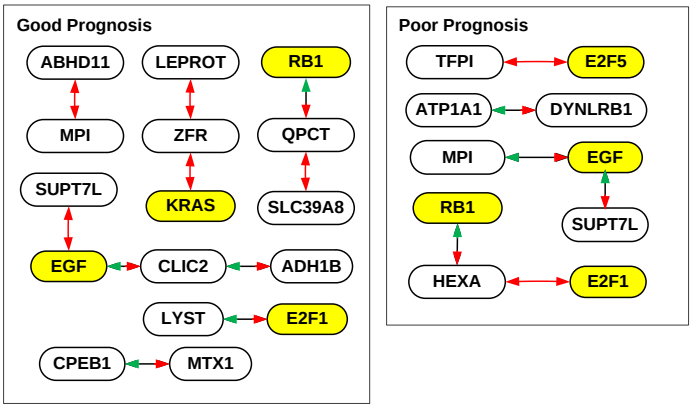
Null distributions of precisions and *q*-values were generated in 1,000 random permutations of the class labels in the test cohorts. From the null statistics, the *FDR* of the disease-mediated coexpression networks is the average of *q*-value from the null distribution.

For each of the 21 signatures and their respective concurrent co-expressed signaling hallmarks (Supplementary Table 3-5), the disease-specific co-expression networks commonly present in both training and test cohorts were retrieved and further validated for biological evaluation as presented in figures Supplementary Fig. 3-23.

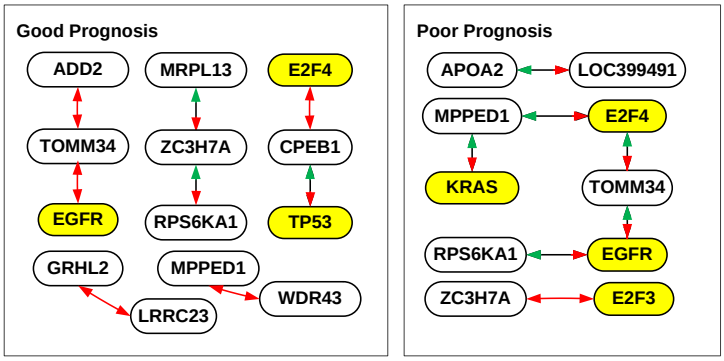
[1] http://www.broadinstitute.org/gsea/msigdb/collections.jsp



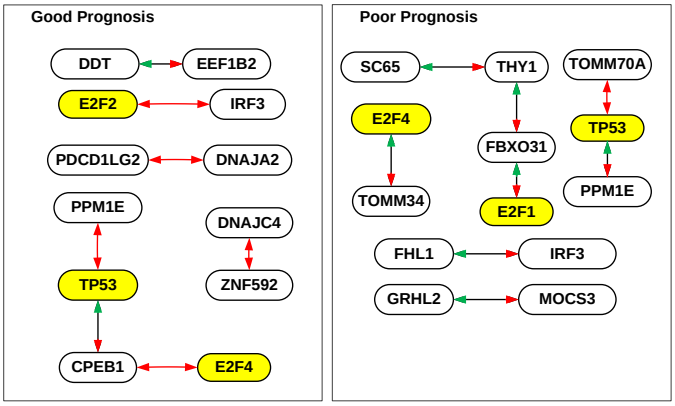
Supplementary Figure 3. Legend of expression relations of the disease-specific coexpression networks represented in the six implication rules.



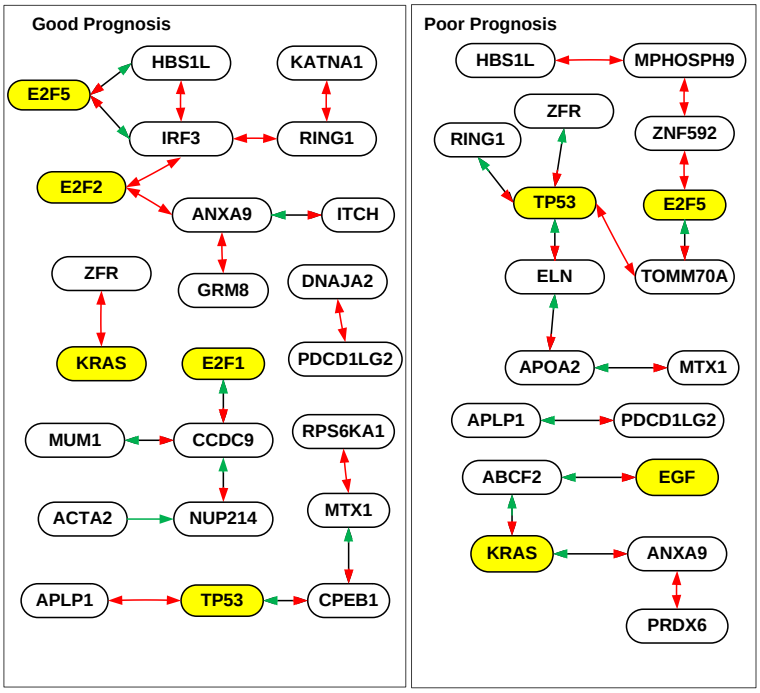
Supplementary Figure 4 . Disease-specific coexpression networks for signature S1 (precision = 0.71, *FDR* = 0.08)



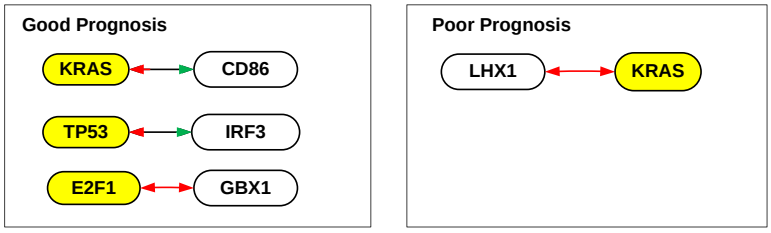
Supplementary Figure 5. Disease-specific coexpression networks for signature S2 (precision = 1, *FDR* = 0.10)



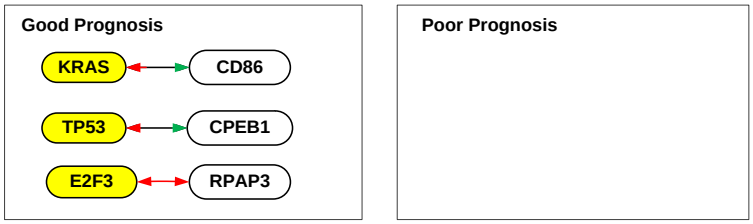
Supplementary Figure 6. Disease-specific coexpression networks for signature S3 (precision = 1, *FDR* = 0.03)



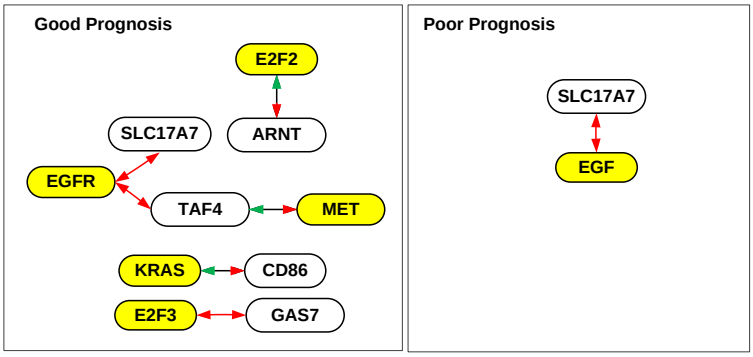
Supplementary Figure 7. Disease-specific coexpression networks for signature S4 (precision = 1, *FDR* = 0.01)



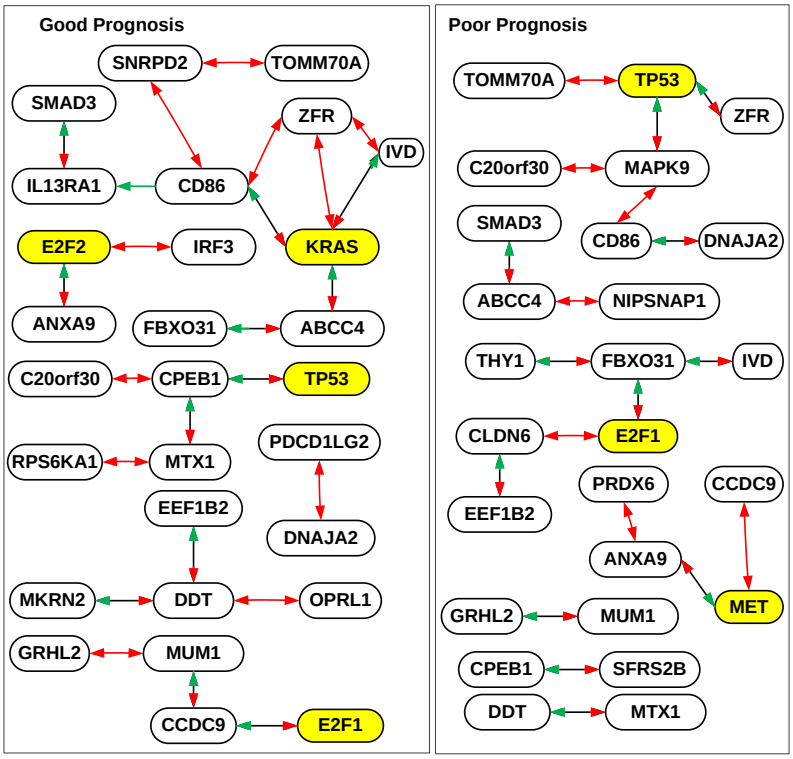
Supplementary Figure 8. Disease-specific coexpression networks for signature S5 (precision = 1, *FDR* = 0)



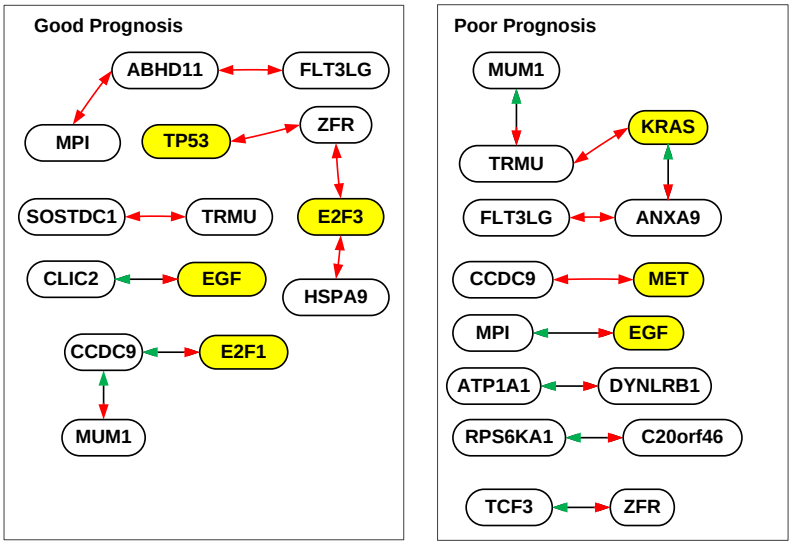
Supplementary Figure 9. Disease-specific coexpression networks for signature S6 (precision = 1, *FDR* = 0)



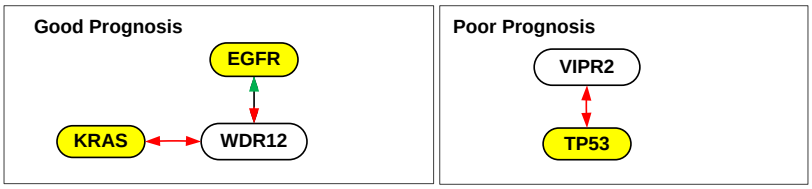
Supplementary Figure 10. Disease-specific coexpression networks for signature S7 (precision = 0.86, *FDR* = 0.10)



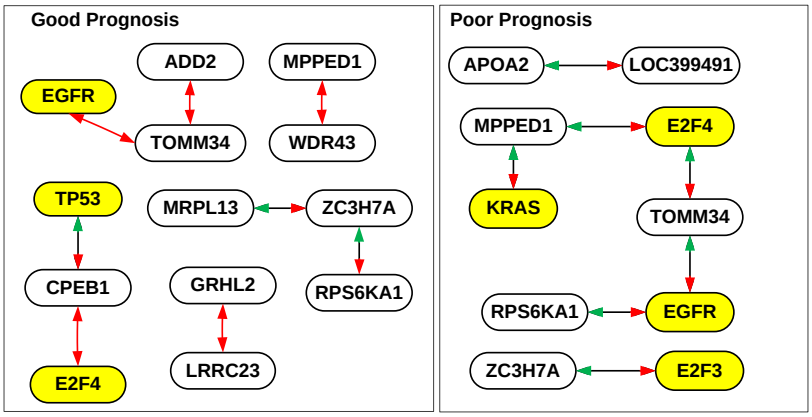
Supplementary Figure 11. Disease-specific coexpression networks for signature S8 (precision = 0.95, *FDR* = 0.05)



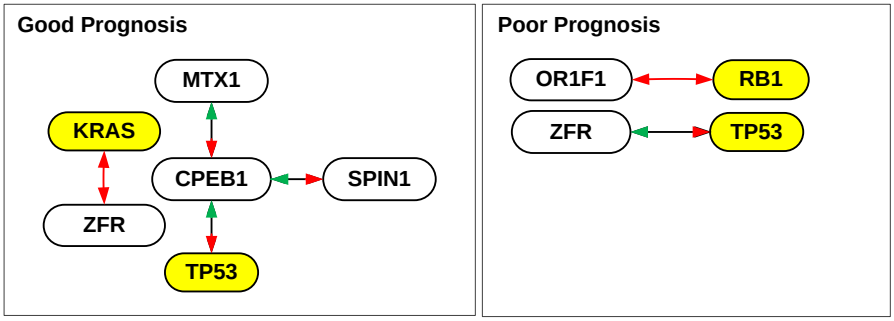
Supplementary Figure 12. Disease-specific coexpression networks for signature S9 (precision = 1, *FDR* = 0.02)



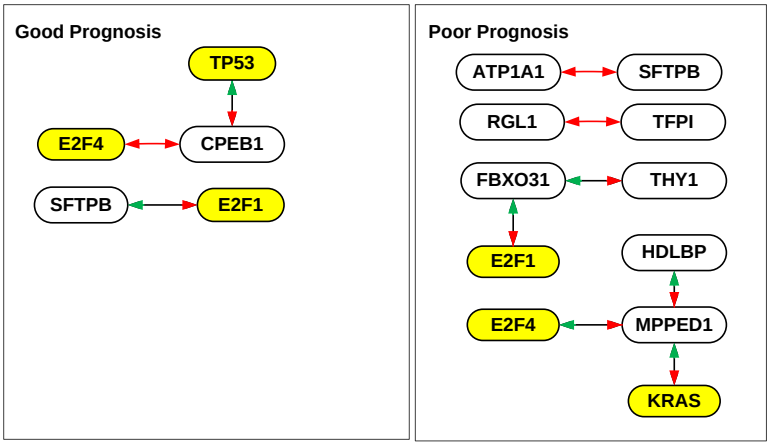
Supplementary Figure 13. Disease-specific coexpression networks for signature S10 (precision = 1, *FDR* = 0.08)



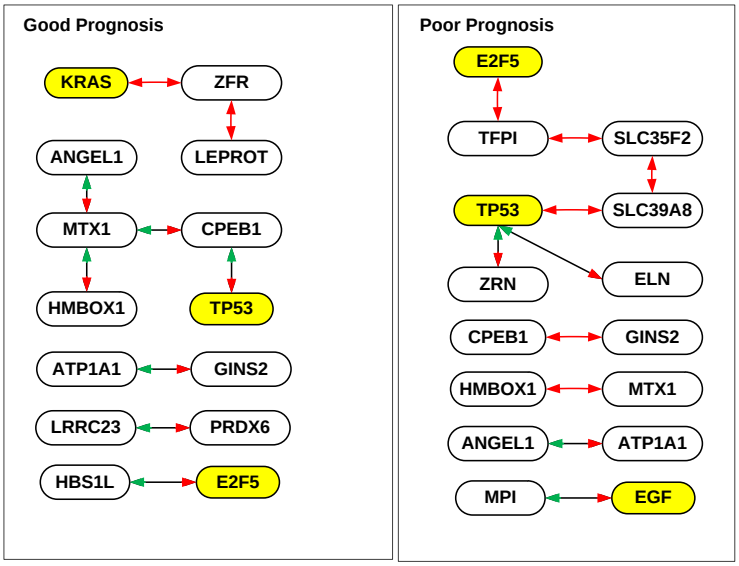
Supplementary Figure 14. Disease-specific coexpression networks for signature S11 (precision = 1, *FDR* = 0.06)



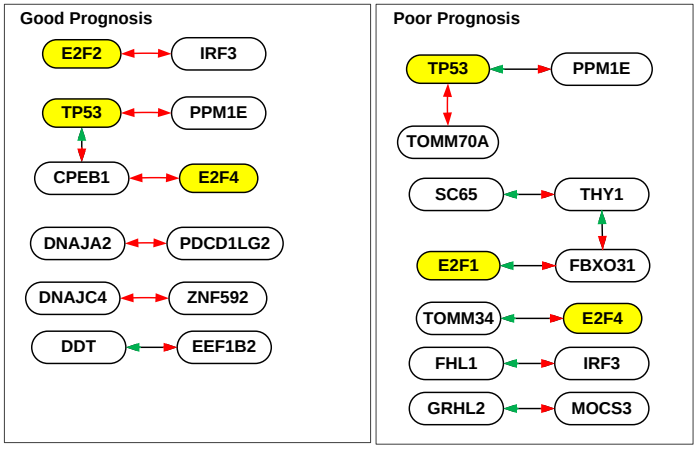
Supplementary Figure 15. Disease-specific coexpression networks for signature S13 (precision = 1, *FDR* = 0)



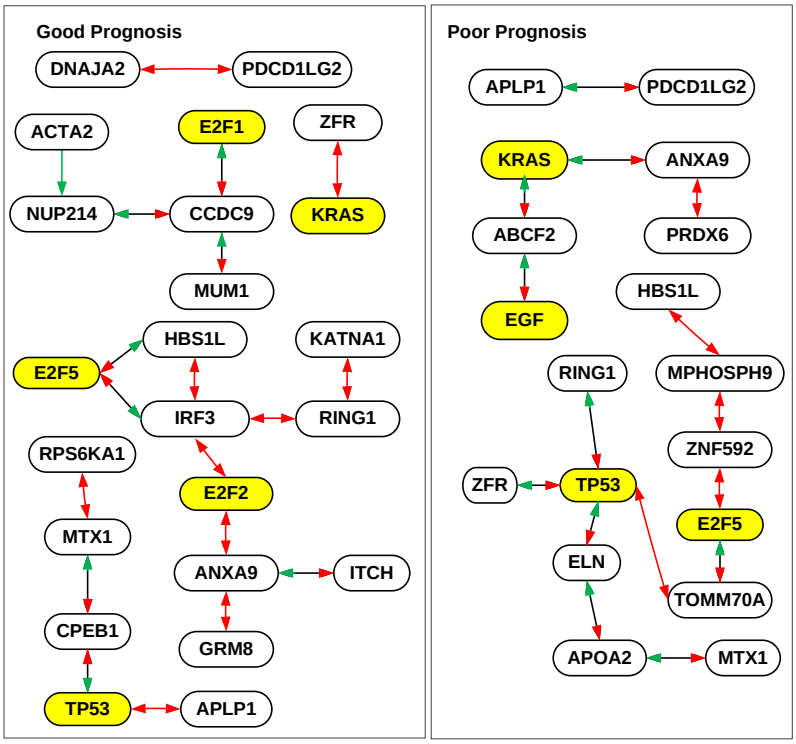
Supplementary Figure 16. Disease-specific coexpression networks for signature S14 (precision = 1, *FDR* = 0.01)



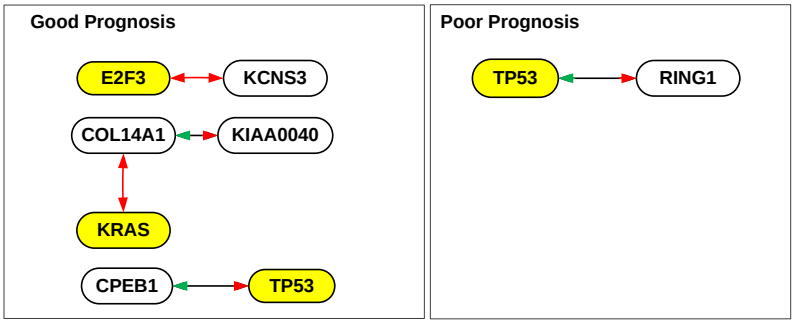
Supplementary Figure 17. Disease-specific coexpression networks for signature S15 (precision = 1, *FDR* = 0.05)



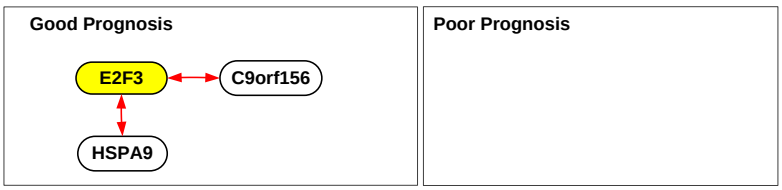
Supplementary Figure 18. Disease-specific coexpression networks for signature S16 (precision = 1, *FDR* = 0.03)



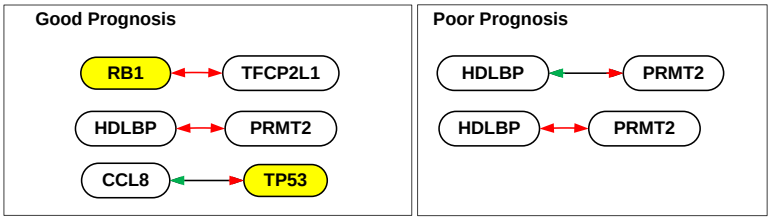
Supplementary Figure 19. Disease-specific coexpression networks for signature S17 (precision = 1, *FDR* = 0.02)



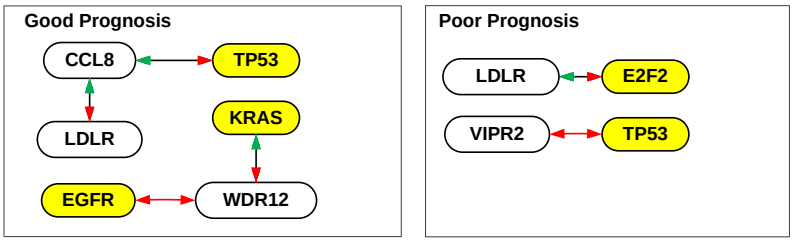
Supplementary Figure 20. Disease-specific coexpression networks for signature S18 (precision = 1, *FDR* = 0)



Supplementary Figure 21. Disease-specific coexpression networks for signature S19 (precision = 1, *FDR* = 0)



Supplementary Figure. 22. Disease-specific coexpression networks for signature S20 (precision = 1, *FDR* = 0.05)



Supplementary Figure 23. Disease-specific coexpression networks for signature S21 (precision = 1, *FDR* = 0.001)

Reference List

1. Rae FK, Martinez G, Gillinder KR, Smith A, Shooter G, Forrest AR, Grimmond SM, Little MH. Anlaysis of complementary expression profiles following WT1 induction versus repression reveals the cholesterol/fatty acid synthetic pathways as a possible major target of WT1. Oncogene. 2004 Apr 15;23(17):3067-79.

2. Maheswaran S, Park S, Bernard A, Morris JF, Rauscher FJ, III, Hill DE, Haber DA. Physical and functional interaction between WT1 and p53 proteins. Proc.Natl.Acad.Sci.U.S A. 1993 Jun 1;90(11):5100-4.

3. Ho JN, Lee SB, Lee SS, Yoon SH, Kang GY, Hwang SG, Um HD. Phospholipase A2 activity of peroxiredoxin 6 promotes invasion and metastasis of lung cancer cells. Mol.Cancer Ther. 2010 Apr;9(4):825-32.

4. Wu YZ, Manevich Y, Baldwin JL, Dodia C, Yu K, Feinstein SI, Fisher AB. Interaction of surfactant protein A with peroxiredoxin 6 regulates phospholipase A2 activity. J.Biol.Chem. 2006 Mar 17;281(11):7515-25.

5. Burns DM, Richter JD. CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. Genes Dev. 2008 Dec 15;22(24):3449-60.

6. Ucar D, Neuhaus I, Ross-MacDonald P, Tilford C, Parthasarathy S, Siemers N, Ji RR. Construction of a reference gene association network from multiple profiling data: application to data analysis. Bioinformatics. 2007 Oct 15;23(20):2716-24.