Supplementary Materials and Methods

Gene Expression Profiling

Total RNA was isolated from fresh-frozen tissues and cell lines, and the RNA integrity (RIN $>$ 5.5) was estimated using a 2100 Bioanalyzer (Agilent Technologie, Santa Cruz, CA). The concentration and purity were measured by NanoDrop 1000 (Thermo Scientific, Wilmington, DE) (ratio of the absorbance at 260 and 280 nm [A_{260}/A_{280}] $>$2). A total of 200 ng total RNA was amplified and incubated for 16 hours at 37°C. Array hybridization, washing, Cy3-streptavidin labeling, and scanning were performed on an iScan using reagents and protocols supplied by the manufacturer (Illumina, San Diego, CA). A total of 750 ng biotinylated cRNA were hybridized to humanRef-8v2 BeadChips (Illumina) for 18 hours at 58°C. Image analysis and data extraction were automated using iScan control software. Microarray data are available at www.ncbi.nlm.nih.gov/geo/ (accession number GSE26566).

Statistical Analysis

Data collection, preprocessing, and quantile normalization were performed with GenomeStudio v2010 as described in Andersen et al.[1] Genes were excluded from analysis if (1) expression levels less than 2-fold change as compared with the median expression values were found in more than 20% of samples, (2) variation in the $\log_2$ ratio was less than the 75th percentile, and (3) more than 20% of data were missing (BRB-ArrayTools v3.8.1, National Institutes of Health).

The signature value was assessed on the basis of overall survival. Survival and time to recurrence (TTR) analyses were performed by the Kaplan–Meier method with use of log-rank statistics (Prism 5; GraphPad, La Jolla, CA). Outcome class association was performed by class random variance method to build a prediction model using 7 different prediction algorithms, including compound covariate predictor (CCP), diagonal linear discriminator (DLD), 1-nearest neighbor (1NN), 3-nearest neighbors (3NN), nearest centroid (NC), support vector machine (SVM), and Bayesian compound covariate (BCC). To estimate accuracy of the prediction model, random permutations during leave-one-out cross-validation (LOOCV) were repeated 1000 times. Class comparison was performed by means of univariate tests with random variance modeling (10,000 random permutations; $P < .001$), depending on the number of classes. A global posttest was applied to test the significance of the gene set ($P < .05$). Cox proportional hazards model and Wald statistics were used to identify genes significantly associated with survival ($P < .01$). To estimate the accuracy, univariate permutation tests were repeated 10,000 times (BRB-ArrayTools v3.8.1). Functional annotation and pathways/networks analysis were performed by GSEA (MIT) and Ingenuity Pathway Analysis (IPA 8.7). The significance of each network and their connectivity was estimated in IPA. Analyses of equal variance and normality were performed by Bartlett's test or D'Agostino & Pearson omnibus test, respectively (α level at .05).

The differences between tumor epithelial and stromal compartments were assessed in 23 cases, and the significant genes were computed by means of paired bootstrap *t* test at $P < .001$ with 5000 repetitions. Class prediction was performed using the random Forest method and random variance modeling. In quantitative trait analysis, we applied Pearson correlation tests. The *P* values for significant genes were calculated based on 10,000 random permutations. The nominal significance level of each univariate test was at $P < .001$ (BRB-ArrayTools v3.8.1). The univariate and multivariate analyses of the clinical and pathological variables were performed using Fisher exact test and multivariate analysis (α level at .01), respectively (JMP8). Within the multivariate analysis, all variables tested under the univariate analysis were analyzed; the significant pairwise partial correlations (*r*) are given in Table 1. Integration of the data sets was performed following standardization using z-transformation. The cholangiocyte cell line H69 was used as a reference in generation of the cell line data set. Meta-analysis of the gene signature with publicly available data sets was performed in Oncomine premium research edition v4.4 (Compenda Bioscience, Ann Arbor, MI).

Laser Capture Microdissection

Tumor epithelial and stromal samples were microdissected from 23 patients using laser microdissection microscope LMD6000 (Leica, Buffalo Grove, IL). Frozen 10-μm-thick serial tissue sections were attached to 2-μm PEN-membrane slides (Leica). The first, middle, and last sections from each specimen were stained with H&E to confirm the presence of tumor tissue before dissection. Microdissection was followed by cresyl violet staining according to the manufacturer. Dissected material was collected directly into RNA extraction buffer (PicoPure RNA Isolation Kit; Arcturus Applied Biosystems, Foster City, CA) and stored at −80°C. The total area of epithelial and stromal tissue was analyzed for each tumor dissected. Handling time per slide did not exceed 20 minutes. The selection criteria (RIN $>$ 8) were determined based on the RNA quality. A total of 100 ng total RNA was amplified before array hybridization.

Mutational Analysis

Genomic DNA isolated from tumor and surrounding noncancerous liver tissue was analyzed for mutations of *KRAS*, *BRAF*, and *EGFR* by quantitative polymerase chain reaction–based assays approved for in vitro diagnostics (CE-IVD). The mutation test kits, which use allele-specific probes to identify the presence of 11 mutations in *KRAS* codons 12, 13, and 61, one mutation in *BRAF* (V600E) (EntroGen Inc, Torzana, CA), and 28 mu-

tations in *EGFR* (Qiagen, Valencia, CA), are designed to detect a low percentage of mutant DNA in a background of wild-type genomic DNA. Genomic DNA (50 ng) was assayed using a control fluorophore to assess the total DNA in the samples. The analysis was performed on an iQ5 real-time polymerase chain reaction system (BioRad, Hercules, CA) and performed according to the manufacturer's specifications.

Cell Culture and Treatment

Seven CCA cell lines were grown in Dulbecco's modified Eagle medium/F12 supplemented with 2 mmol/L L-glutamine, 1 U/mL penicillin/streptomycin, and 10% fetal calf serum. The normal cholangiocyte cell line (H69) was grown in BEGM (Lonza, Allendale, NJ) supplemented with a BEGM bullet kit, 2 mmol/L L-glutamine, 1 U/mL penicillin/streptomycin, and 10% fetal calf serum. Twenty-four hours before analysis, the culture medium was changed to Dulbecco's modified Eagle medium/F12. H69, KMCH, WITT, and KMBC were a gift from Dr G. J. Gores (Mayo Clinic, Rochester, MN). The remaining cholangiocarcinoma cell lines (YSCCC, HuCCT1, RBE, and SSP-25) were purchased from Riken BRC Cell Bank (Koyadai, Japan). Lapatinib (Tykerb; GlaxoSmithKline, Brentford, United Kingdom) was dissolved in dimethyl sulfoxide as a stock solution at 10 mmol/L and stored at −20°C. Trastuzumab (Herceptin; Genentech, San Francisco, CA) was dissolved in sterile apyrogen water and stored at 4°C. Each cell line was treated with lapatinib at the 50% lethal dose (LD_{50}) or 500 μg/mL trastuzumab for 7 days. The assay was repeated 8 times. Media and reagents were changed every 3 days.

Cytotoxicity Assay

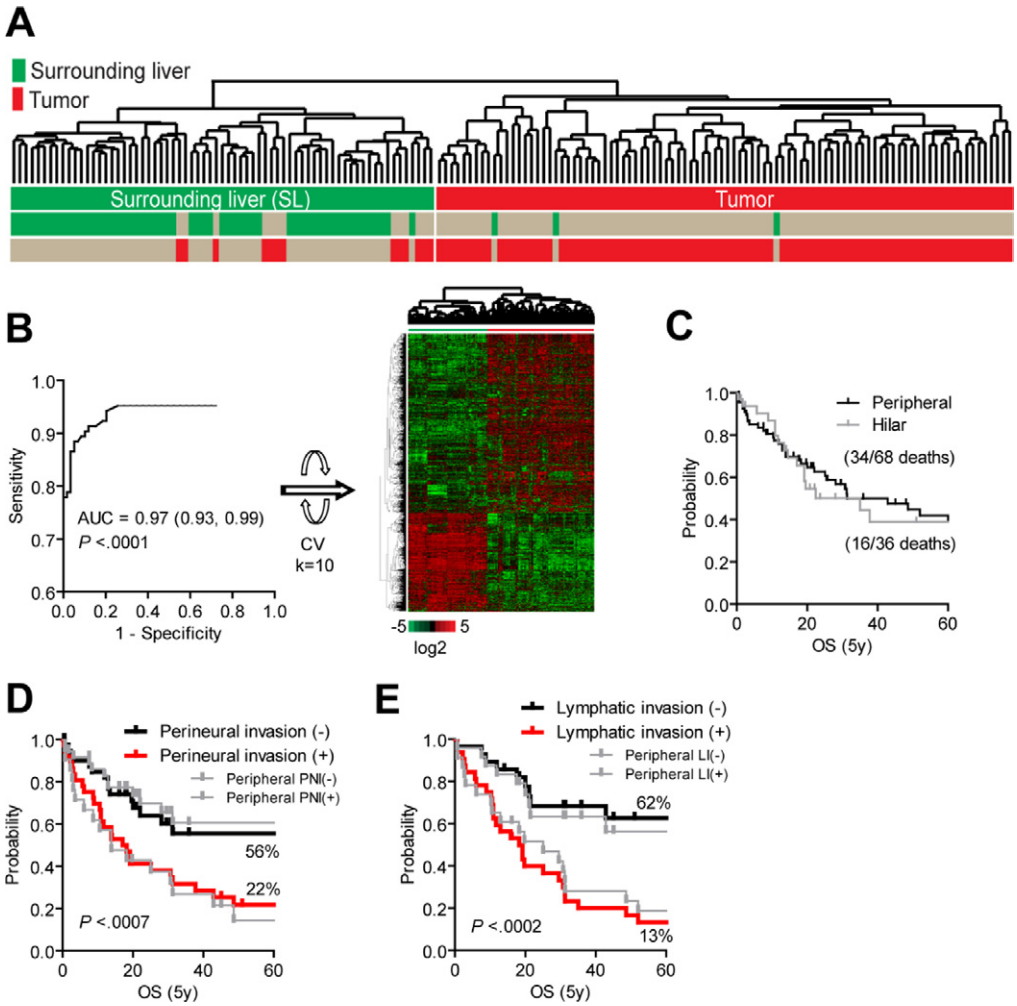
The sensitivity of each cell line to either lapatinib or trastuzumab was measured using WST-1 colorimetric assay according to the manufacturer's protocol (Roche Applied Sciences, Indianapolis, IN). The cells were plated at a density of 5×10^4 in 100 μL Dulbecco's modified Eagle medium per well of a 96-well plate. After 24 hours, the cells were treated with 0−100 μmol/L lapatinib or 0−500 μg/mL trastuzumab and incubated for 72 hours (Supplementary Figure 5). Absorbance of WST-1 was measured at 450 nm for 0.5−4 hours (M5; Molecular Devices, Sunnyvale, CA). Cell viability was defined as the absorbance of the treatment group compared with the control group and presented as percent change ± SD. The experiment was repeated (n = 8). The LD_{50} was estimated for each cell line using a nonlinear regression as log (inhibitor) versus variable normalized drug response ± 95% CI.

Immunohistochemical and Western Blot Analysis

Serial 8-μm-thick frozen sections were stained for MKI67, EGFR, MET, HER2, and pRPS6 (s235 and s244). Imaging was performed using a Zeiss 510 NLO confocal microscope (Carl Zeiss, Inc., Thornwood, NY) with a 20× objective lens. Image brightness parameters were adjusted across all comparison groups. For each tumor, the *H*-score was calculated by multiplying the staining intensity score (0, none; 1, weak; 2, moderate; 3, intense) by the percentage of cells stained at a given intensity (0−100) by an observer blinded to the sample identity. Staining intensity was determined with Zen software (Zeiss) from triplicate random images. The total composition of epithelial and stromal area in each tumor was evaluated by CK19/4′,6-diamidino-2-phenylindole (DAPI) stain and H&E sections and quantified by ImageJ v1.42 software. Western blotting was performed using antibodies against TMPRSS4, ITGA2, CEACAM6 (Abcam, Cambridge, MA), MET (Santa Cruz Biotechnology, Santa Cruz, CA), AKT, $pAKT^{S473}$, EGFR, $pEGFR^{Y1068}$, HER2, and $pHER2^{Y877}$ (Cell Signaling Technology, Darnes, MA) according to the manufacturer's recommendations.

Supplementary References

31. Andersen JB, Factor VM, Marquardt JU, et al. An integrated genomic and epigenomic approach predicts therapeutic response to zebularine in human liver cancer. Sci Transl Med 2010;2:54ra77.
32. Woo HG, Lee JH, Yoon JH, et al. Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. Cancer Res 2010;70:3034–3041.



Supplementary Figure 1. Hierarchical cluster analysis and evaluation of the patient cohort. (*A*) Unsupervised hierarchical clustering of 104 fresh-frozen CCAs (n = 104) and surrounding livers (SL, n = 59) separates tissues into 2 groups representing SL and tumors. (*B*) Sensitivity of the gene list to correctly predict the classification. Threshold of predicted probability for a sample to belong to a group is 0.8 based on the Bayesian compound covariate predictor. The sample is predicted as SL if the likelihood is greater than the threshold. The specificity is represented by the area under the receiver and operating characteristic (ROC) curve (AUC, 95% CI) and is calculated following 10,000 random permutations during k-fold (k = 10) cross-validation (CV) ($P < .001$). (*C*) Survival analysis of patients according to anatomic location of the tumor. (*D* and *E*) Analysis of overall survival based on (*D*) positive (PNI+) and negative (PNI-) perineural invasion (PNI) and (*E*) positive (LI+) and negative (LI-) lymphatic invasion (LI) as independent prognostic predictors. Kaplan–Meier and log-rank statistics were used to determine levels of significance for 2 groups. The analysis for peripheral tumor only is presented in *gray*.

Supplementary Table 1. List of Classifier Genes

	Gene symbol	Parametric *P* value	CV support (%)	Geometric mean (log_2 ratio in class 1)	Geometric mean (log_2 ratio in class 2)	Ratio	Chromosomal location
1	TMPRSS4	<1e-07	100	3.69	63.16	0.058	11q23.3
2	BIK	<1e-07	100	2.13	8.83	0.24	22q13.31
3	SERPINB5	<1e-07	100	4.39	55.51	0.079	18q21.3
4	KRT19	<1e-07	100	5	16.39	0.3	17q21.2
5	TRIM31	<1e-07	100	5.9	40.38	0.15	6p21.3
6	C13orf34	<1e-07	100	2.25	4.79	0.47	13q22.1
7	CDH3	<1e-07	100	14.07	158.43	0.089	16q22.1
8	CAPG	<1e-07	100	3.93	9.48	0.42	2p11.2
9	CEACAM6	<1e-07	100	2.65	12.57	0.21	19q13.2
10	IGFL2	<1e-07	100	11.88	106.2	0.11	19q13.32
11	HK2	<1e-07	100	4.44	15.02	0.3	2p13
12	STYK1	<1e-07	100	2.84	9.74	0.29	12p13.2
13	CEACAM5	<1e-07	100	3.91	61.27	0.064	19q13.1-q13.2
14	QPCT	<1e-07	100	1.61	4.15	0.39	2p22.2
15	LAMB3	<1e-07	100	8.06	32.58	0.25	1q32
16	TSPAN1	<1e-07	100	2.9	12.47	0.23	1p34.1
17	SLC6A14	<1e-07	100	24.01	343.86	0.07	Xq23-q24
18	RHBDL2	<1e-07	100	1.84	6.07	0.3	1p34.3
19	MXRA5	<1e-07	100	2.61	7.38	0.35	Xp22.33
20	GPR92	<1e-07	100	1.62	3.94	0.41	12p13.31
21	CTSE	<1e-07	100	3.4	27.19	0.13	1q31
22	ADAM8	<1e-07	100	2.92	8.33	0.35	10q26.3
23	DTX2	<1e-07	100	3.42	6.78	0.5	7q11.23
24	ASPHD2	<1e-07	100	2.53	7.56	0.33	22q12.1
25	FLJ22662	<1e-07	100	1.82	5.12	0.36	12p13.1
26	CYP2S1	<1e-07	100	1.21	8.55	0.14	19q13.1
27	C20orf42	<1e-07	100	10.38	45.97	0.23	20p12.3
28	ANKRD22	<1e-07	100	3.63	10.52	0.34	10q23.31
29	C19orf21	<1e-07	100	5.76	20.01	0.29	19p13.3
30	COL11A1	<1e-07	100	3.91	15.88	0.25	1p21
31	PSCA	<1e-07	100	2.53	29.37	0.086	8q24.2
32	C19orf33	<1e-07	100	4.85	23.33	0.21	19q13.2
33	GPR110	2.00E-07	100	1.81	5.37	0.34	6p12.3
34	ZBED2	1.00E-07	100	2.7	10.71	0.25	3q13.13
35	KRT17	<1e-07	100	6.37	48.71	0.13	17q12-q21
36	CD55	3.00E-07	100	1.37	3.48	0.39	1q32
37	ANLN	1.00E-07	100	3.83	9.28	0.41	7p15-p14
38	KCNK1	2.00E-07	100	2.16	4.27	0.51	1q42-q43
39	CORO2A	2.00E-07	100	2.25	4.7	0.48	9q22.3
40	ABHD9	4.00E-07	100	2.07	6.37	0.33	19p13.12
41	ITGB6	5.00E-07	100	4.54	16.04	0.28	2q24.2
42	KLK7	9.00E-07	100	3	22	0.14	19q13.33
43	PPP1R1B	1.00E-06	100	1.86	5.5	0.34	17q12
44	NRP2	1.30E-06	100	1.74	3.78	0.46	2q33.3
45	PLEK2	7.00E-07	100	4.89	11.32	0.43	14q23.3
46	C15orf48	1.20E-06	100	3.62	11.04	0.33	15q21.1
47	GJB5	2.80E-06	100	7.68	50.71	0.15	1p35.1
48	KIAA1199	1.00E-06	100	5.81	20.24	0.29	15q24
49	FA2H	1.50E-06	100	2.76	7.01	0.39	16q23
50	VDR	2.00E-06	100	2.88	6.61	0.44	12q13.11
51	AP1S3	1.60E-06	100	4.47	10.52	0.42	2q36.1
52	PTPRR	2.20E-06	100	2.63	10.02	0.26	12q15
53	SLC4A11	1.50E-06	100	2.51	5.5	0.46	20p12
54	DAPP1	3.30E-06	100	2.65	6.76	0.39	4q25-q27
55	MYEOV	2.10E-06	100	8.51	38.82	0.22	11q13
56	USP54	2.40E-06	100	4.7	10.37	0.45	10q22.2
57	SLC16A3	2.10E-06	100	2.75	5.56	0.49	17q25
58	GPR160	2.70E-06	100	2.36	4.84	0.49	3q26.2-q27
59	TJP3	2.90E-06	100	2.21	4.49	0.49	19p13.3
60	MYH14	2.90E-06	100	1.93	3.66	0.53	19q13.33
61	SCIN	3.00E-06	100	4.59	13.85	0.33	7p21.3

Supplementary Table 1. Continued

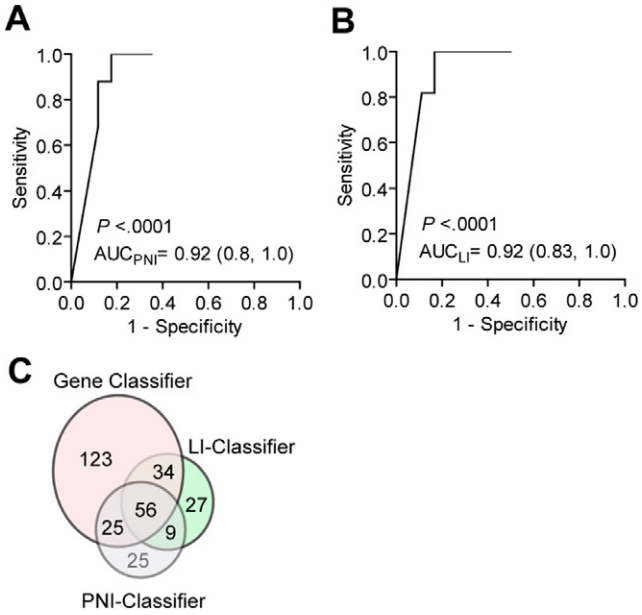
	Gene symbol	Parametric *P* value	CV support (%)	Geometric mean (log_2 ratio in class 1)	Geometric mean (log_2 ratio in class 2)	Ratio	Chromosomal location
62	S100P	4.60E-06	100	4.25	23.06	0.18	4p16
63	CDCP1	3.60E-06	100	3.72	7.39	0.5	3p21.31
64	FXYD3	5.80E-06	100	3.14	9.95	0.32	19q13.11-q13.12
65	BHLHB3	4.90E-06	100	2.39	4.75	0.5	12p11.23-p12.1
66	C1orf135	4.80E-06	100	8.02	21.8	0.37	1p36.11
67	PPP1R14D	6.80E-06	100	4.4	16.97	0.26	15q15.1
68	SFN	5.10E-06	100	14.15	44.98	0.31	1p36.11
69	MSLN	9.20E-06	100	7.61	60.56	0.13	16p13.3
70	ALDH3B1	6.90E-06	100	2.04	4.06	0.5	11q13
71	HSH2D	6.10E-06	100	5.06	10.7	0.47	19p13.11
72	CLIC3	9.70E-06	100	2.2	8.48	0.26	9q34.3
73	BCMO1	7.50E-06	100	3.05	9.38	0.32	16q21-q23
74	CDH17	8.40E-06	100	4.64	28.95	0.16	8q22.1
75	REG4	1.05E-05	100	2.99	23.16	0.13	1p13.1-p12
76	LIPH	1.00E-05	100	1.84	3.35	0.55	3q27
77	MYO1A	1.72E-05	100	1.84	10.36	0.18	12q13-q14
78	MTMR11	1.44E-05	100	1.84	3.61	0.51	1q12-q21
79	DHDH	1.41E-05	100	2.4	5.19	0.46	19q13.3
80	SLC6A20	2.45E-05	100	1.98	4.73	0.42	3p21.3
81	FAM83D	1.70E-05	100	2.75	5.19	0.53	20q11.22-q12
82	GPA33	2.19E-05	100	1.78	7.87	0.23	1q24.1
83	NOXO1	2.45E-05	100	2.56	7.26	0.35	16p13.3
84	NMU	2.84E-05	100	15.65	136.18	0.11	4q12
85	ST6GALNAC1	3.70E-05	100	2.2	10.06	0.22	17q25.1
86	APOBEC1	3.12E-05	100	3.53	15.58	0.23	12p13.1
87	ITGA2	1.99E-05	100	3.72	7.17	0.52	5q23-q31
88	GPR87	2.50E-05	100	3.99	12.77	0.31	3q24
89	CEP55	2.09E-05	100	19.54	42.94	0.45	10q23.33
90	HSPA1B	2.35E-05	100	2.15	3.59	0.6	6p21.3
91	C1orf178	3.85E-05	100	4.22	17.6	0.24	1p13.2
92	ST14	2.32E-05	100	4.98	9.49	0.52	11q24-q25
93	GJB3	2.90E-05	100	20.22	55.04	0.37	1p34
94	SIX4	4.16E-05	100	5.87	12.84	0.46	14q23
95	OLR1	4.35E-05	100	3.87	8.46	0.46	12p13.2-p12.3
96	BUB1	3.99E-05	100	5.62	10.45	0.54	2q14
97	FGFBP1	5.56E-05	100	5.52	30.28	0.18	4p16-p15
98	VCAN	4.69E-05	100	2.8	5.22	0.54	5q14.3
99	COL1A2	5.70E-05	100	2.11	3.9	0.54	7q22.1
100	C1GALT1	5.57E-05	100	2.18	3.81	0.57	7p14-p13
101	CENPE	5.22E-05	100	3.11	5.32	0.59	4q24-q25
102	PTK6	5.89E-05	100	4.01	10.13	0.4	20q13.3
103	DLG7	6.40E-05	100	19.86	45.71	0.43	14q22.3
104	GPRC5A	9.33E-05	100	3.51	11.7	0.3	12p13-p12.3
105	KLK6	7.94E-05	100	7.87	41.24	0.19	19q13.3
106	ELL3	8.43E-05	100	2.38	4.43	0.54	15q15.3
107	MMP12	7.74E-05	100	4.28	13.38	0.32	11q22.3
108	TPX2	7.53E-05	100	3.11	5.44	0.57	20q11.2
109	SMPDL3B	9.40E-05	100	2.59	4.9	0.53	1p35.3
110	CBLC	8.37E-05	100	2.58	4.15	0.62	19q13.2
111	SPOCD1	0.0001007	100	2.51	4.86	0.52	1p35.2
112	CKMT1B	0.0001279	100	3.81	13.61	0.28	15q15
113	KLF5	0.0001351	100	2.31	5.24	0.44	13q22.1
114	C9orf140	0.000112	100	4.99	9.2	0.54	9q34.3
115	LGALS3	0.0001273	100	2.02	3.54	0.57	14q21-q22
116	OLFM4	0.0001666	100	4.54	28.16	0.16	13q21.1
117	NFE2L3	0.0001243	100	3.58	6.55	0.55	7p15-p14
118	LGALS4	0.0001549	100	2.38	4.14	0.58	19q13.2
119	MUC17	0.0002118	100	3.01	37.54	0.08	7q22.1
120	KLK10	0.0001582	100	3.08	7.41	0.42	19q13.3-q13.4

Supplementary Table 1. Continued

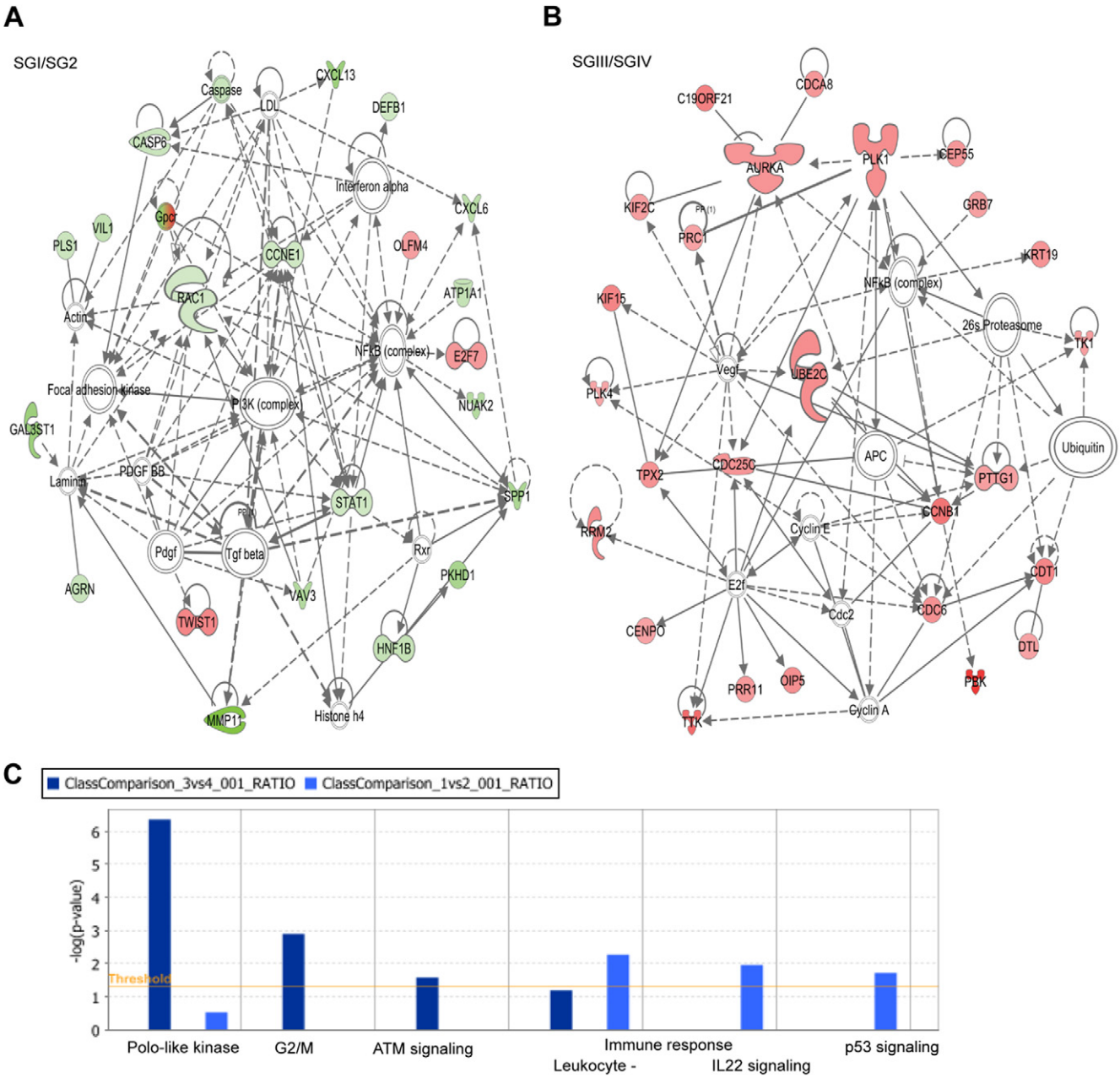
	Gene symbol	Parametric *P* value	CV support (%)	Geometric mean (log_2 ratio in class 1)	Geometric mean (log_2 ratio in class 2)	Ratio	Chromosomal location
121	UNC93B1	0.0001401	100	3.61	6.39	0.57	11q13
122	TFCP2L1	0.0001655	100	5.74	13.62	0.42	2q14
123	KRT15	0.0001798	100	2.69	6.46	0.42	17q21.2
124	SERPINB2	0.0002116	100	2.02	6.04	0.33	18q21.3
125	TMC7	0.0001995	100	3.96	7.89	0.5	16p12.3
126	KRT6B	0.0002213	100	2.99	8.65	0.35	12q12-q13
127	UBE2C	0.0001866	100	6.52	12.5	0.52	20q13.12
128	C20orf102	0.0002816	100	4.13	14.51	0.28	20q11.23
129	VILL	0.0002957	100	1.55	3.94	0.39	3p21.3
130	MYB	0.0002687	100	1.7	5.59	0.3	6q22-q23
131	SPIRE2	0.0002104	100	2.55	4.82	0.53	16q24
132	PKP3	0.0002312	100	33.57	94.45	0.36	11p15
133	SOX21	0.0002524	100	9.04	33.42	0.27	13q31-q32
134	CDC20	0.0002162	100	4.24	7.95	0.53	1p34.1
135	CDX2	0.0002672	100	4.83	19.19	0.25	13q12.3
136	PMAIP1	0.000233	100	4.98	9.85	0.51	18q21.32
137	LOC201175	0.0002307	100	4.46	9	0.5	17q21.31
138	PLAC8	0.0003352	100	2.3	6.47	0.36	4q21.22
139	ANXA3	0.0002974	100	3.34	6.06	0.55	4q13-q22
140	CDCA3	0.0002894	100	8.3	16.65	0.5	12p13
141	GPR35	0.0003212	100	3.26	6.23	0.52	2q37.3
142	PLA2G10	0.0003878	100	4.93	18.95	0.26	16p13.1-p12
143	GALNT12	0.0004242	96	1.9	4.23	0.45	9q22.33
144	WNK4	0.0004386	100	2.48	6.69	0.37	17q21-q22
145	EVI1	0.0004515	98	2.1	4.01	0.52	3q24-q28
146	LEMD1	0.0004187	100	28.77	105.72	0.27	1q32.1
147	PLAUR	0.0004512	100	1.89	3.49	0.54	19q13
148	CTHRC1	0.0004123	100	4.58	8.17	0.56	8q22.3
149	COL12A1	0.000436	100	2.27	4.38	0.52	6q12-q13
150	TOP2A	0.0004306	100	9.44	17.22	0.55	17q21-q22
151	KCNN4	0.0006333	93	3.94	12.15	0.32	19q13.2
152	TUFT1	0.0005077	100	2.91	4.52	0.64	1q21
153	HOXB7	0.0005198	100	7.14	18.81	0.38	17q21.3
154	BCL11B	0.0005385	100	1.96	3.14	0.62	14q32.2
155	MYOM3	0.0007207	82	2.66	5.14	0.52	1p36.11
156	CST6	0.0006506	87	2.85	8.78	0.32	11q13
157	KRT20	0.0006358	91	2.44	8.25	0.3	17q21.2
158	CLDN4	0.0006104	94	7.36	14.94	0.49	7q11.23
159	C1orf116	0.0006471	89	4.24	7.12	0.6	1q32.1
160	TTK	0.0006426	90	22.57	45.22	0.5	6q13-q21
161	AURKA	0.0006382	92	3.25	5.19	0.63	20q13.2-q13.3
162	SCEL	0.0006848	83	2.62	4.86	0.54	13q22
163	EPS8L1	0.0006661	84	4.53	8.73	0.52	19q13.42
164	KIF11	0.0006852	78	7.69	13.4	0.57	10q24.1
165	KIF15	0.0007712	76	11.81	24.78	0.48	3p21.31
166	GCNT3	0.0008445	65	4.12	8.83	0.47	15q21.3
167	NOX4	0.0007878	66	2.44	4.1	0.59	11q14.2-q21
168	SLC39A4	0.0007941	64	2.94	5.31	0.55	8q24.3
169	GALNT6	0.0008233	59	2	3.27	0.61	12q13
170	CCNB2	0.0007994	66	8.27	14.4	0.57	15q22.2
171	COL6A3	0.0008961	58	2.37	3.91	0.6	2q37
172	LAMC2	0.0008651	52	14.06	30.29	0.46	1q25-q31
173	DKK1	0.0009629	47	9.48	31.23	0.3	10q11.2
174	CTSL2	0.0008904	50	7.08	15.42	0.46	9q22.2
175	GJB4	0.0009991	44	3.45	7.37	0.47	1p34.3
176	HIST1H1C	0.0008892	62	3.26	1.92	1.7	6p21.3
177	SLC5A9	0.0008602	71	4.6	2.29	2	1p33
178	CRYBB2	0.0007377	75	3.56	2.11	1.69	22q11.2-q12.1|22q11.23
179	CLDN9	0.0007351	77	5.69	2.56	2.22	16p13.3
180	SMA4	0.0006846	81	3.3	1.96	1.69	5q13

Supplementary Table 1. Continued

	Gene symbol	Parametric *P* value	CV support (%)	Geometric mean (log_2 ratio in class 1)	Geometric mean (log_2 ratio in class 2)	Ratio	Chromosomal location
181	UBD	0.0005445	100	5.18	2.88	1.8	6p21.3
182	VN1R1	0.0005996	89	13.42	6.13	2.19	19q13.4
183	PGBD5	0.0004749	100	4.68	2.66	1.76	1q42.13
184	PTPN14	0.0003083	100	5.49	3.29	1.67	1q32.2
185	OXTR	0.0002728	100	12.64	5.24	2.41	3p25
186	PTHLH	0.00028	100	4.1	1.93	2.13	12p12.1-p11.2
187	TCF2	0.0002788	100	3.89	2.21	1.76	17cen-q21.3
188	ZDHHC1	0.0002288	100	3.64	2.25	1.62	16q22.1
189	HIST1H4H	0.0002183	100	5.51	2.64	2.09	6p21.3
190	ABHD3	0.0002191	100	3.21	1.97	1.63	18q11.2
191	CRYM	0.0002023	100	3.47	1.84	1.88	16p13.11-p12.3
192	FAM59A	0.0001592	100	3.37	2.16	1.57	18q12.1
193	C1QL4	2.00E-04	100	8.19	2.77	2.96	12q13.12
194	TNFRSF12A	0.0001169	100	4.37	2.76	1.58	16p13.3
195	GGT6	0.0001818	100	8.53	3.56	2.39	17p13.2
196	ATP1A1	8.92E-05	100	3.7	2.24	1.65	1p21
197	SYT12	0.0001143	100	4.47	1.94	2.31	11q13.1
198	ANKRD1	8.31E-05	100	17.62	5.64	3.12	10q23.31
199	PAQR5	7.51E-05	100	17.43	7.12	2.45	15q23
200	PSPH	7.12E-05	100	4.3	2.39	1.79	7p15.2-p15.1
201	CTXN1	6.28E-05	100	4.21	2.33	1.81	19p13.2
202	SCTR	7.12E-05	100	7.5	2.66	2.82	2q14.1
203	FAM38B	4.75E-05	100	3.79	2.21	1.72	18p11.22
204	C9orf150	4.63E-05	100	3.6	2.04	1.77	9p23
205	PATZ1	4.90E-05	100	6.01	3.05	1.97	22q12.2
206	FGFR2	5.80E-05	100	4.41	1.88	2.35	10q26
207	CDH6	5.20E-05	100	4.58	1.73	2.65	5p15.1-p14
208	PKHD1	3.22E-05	100	6.99	2.54	2.75	6p12.2
209	DEFB1	2.78E-05	100	3.87	1.97	1.96	8p23.2-p23.1
210	MLH3	1.64E-05	100	4.92	2.84	1.73	14q24.3
211	KIAA0802	1.03E-05	100	3.77	2.1	1.79	18p11.22
212	CXCL6	1.07E-05	100	8.67	3.21	2.71	4q21
213	RGS4	4.90E-06	100	8.06	2.62	3.07	1q23.3
214	DBNDD1	4.10E-06	100	6.18	3.33	1.85	16q24.3
215	KCNF1	4.50E-06	100	7.97	2.97	2.68	2p25
216	VTCN1	6.90E-06	100	6.38	1.89	3.38	1p13.1
217	KRT23	3.50E-06	100	52.84	12.31	4.29	17q21.2
218	NME5	6.20E-06	100	5.68	2.01	2.83	5q31
219	HKDC1	2.30E-06	100	10.1	4.86	2.08	10q21.3
220	MPP6	2.40E-06	100	3.58	1.69	2.12	7p15
221	SLC4A3	1.50E-06	100	16	4.56	3.51	2q36
222	C10orf132	1.00E-06	100	13.47	4.46	3.02	10q24.2
223	TMEM156	8.00E-07	100	9.67	3.15	3.07	4p14
224	RGS14	4.00E-07	100	3.89	2.09	1.86	5q35.3
225	VEPH1	6.00E-07	100	10.42	3.01	3.46	3q24-q25
226	GMNN	2.00E-07	100	6.96	2.59	2.69	6p22.2
227	SPP1	2.00E-07	100	6.72	2.94	2.29	4q21-q25
228	FLJ23861	3.00E-07	100	3.79	1.62	2.34	2q34
229	LOXL4	<1e-07	100	5.94	2.37	2.5	10q24
230	CSPP1	<1e-07	100	3.99	1.48	2.7	8q13.2
231	WNK2	2.00E-07	100	6.01	2.12	2.83	9q22.3
232	NUAK2	<1e-07	100	7.77	2.71	2.87	1q32.1
233	ITIH5	<1e-07	100	9.26	1.82	5.08	10p14
234	MCCC2	<1e-07	100	4.53	2.21	2.05	5q12-q13
235	GAL3ST1	<1e-07	100	12.77	2.65	4.82	22q12.2
236	SMYD2	<1e-07	100	3.81	1.66	2.3	1q41
237	DCDC2	<1e-07	100	19.64	3.06	6.42	6p22.1
238	LPGAT1	<1e-07	100	3.57	1.5	2.37	1q32



Supplementary Figure 2. Perineural invasion (PNI) and lymphatic invasion (LI) as predictors of overall survival. (*A* and *B*) Class prediction using (*A*) PNI and (*B*) LI as predictors of survival, respectively. The specificity is represented by the area under the ROC curve (AUC, 95% CI) using Bayesian compound covariate predictor modeling. (*C*) Venn diagram analysis of the overlap between our 238-gene classifier and PNI and LI predictors.



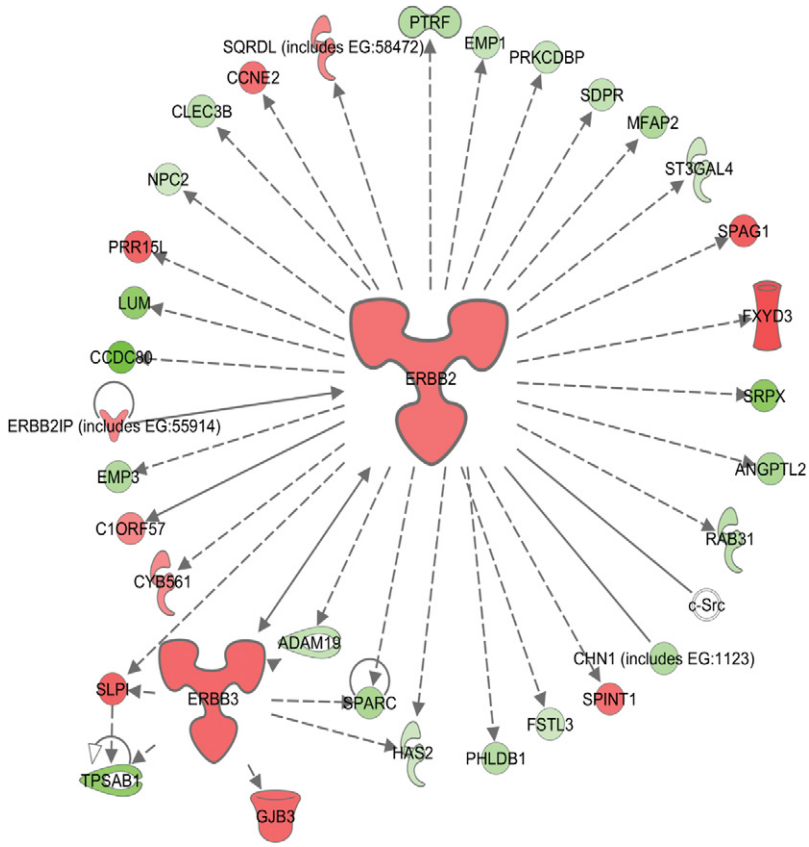
Supplementary Figure 3. Functional network analysis. (*A* and *B*) The top networks differentially expressed between subgroups (*A*) SGI/SGII and (*B*) SGIII/SGIV. Up-regulated and down-regulated genes are shown in *red* and *green*, respectively. *Dashed lines* represent indirect connectivity. (*C*) The significant canonical pathways differentiating SGI/SGII (*light blue*) versus SGIII/SGIV (*dark blue*). $P < .05$ by Fisher exact test.

Supplementary Table 2. List of Survival Genes and Network Association

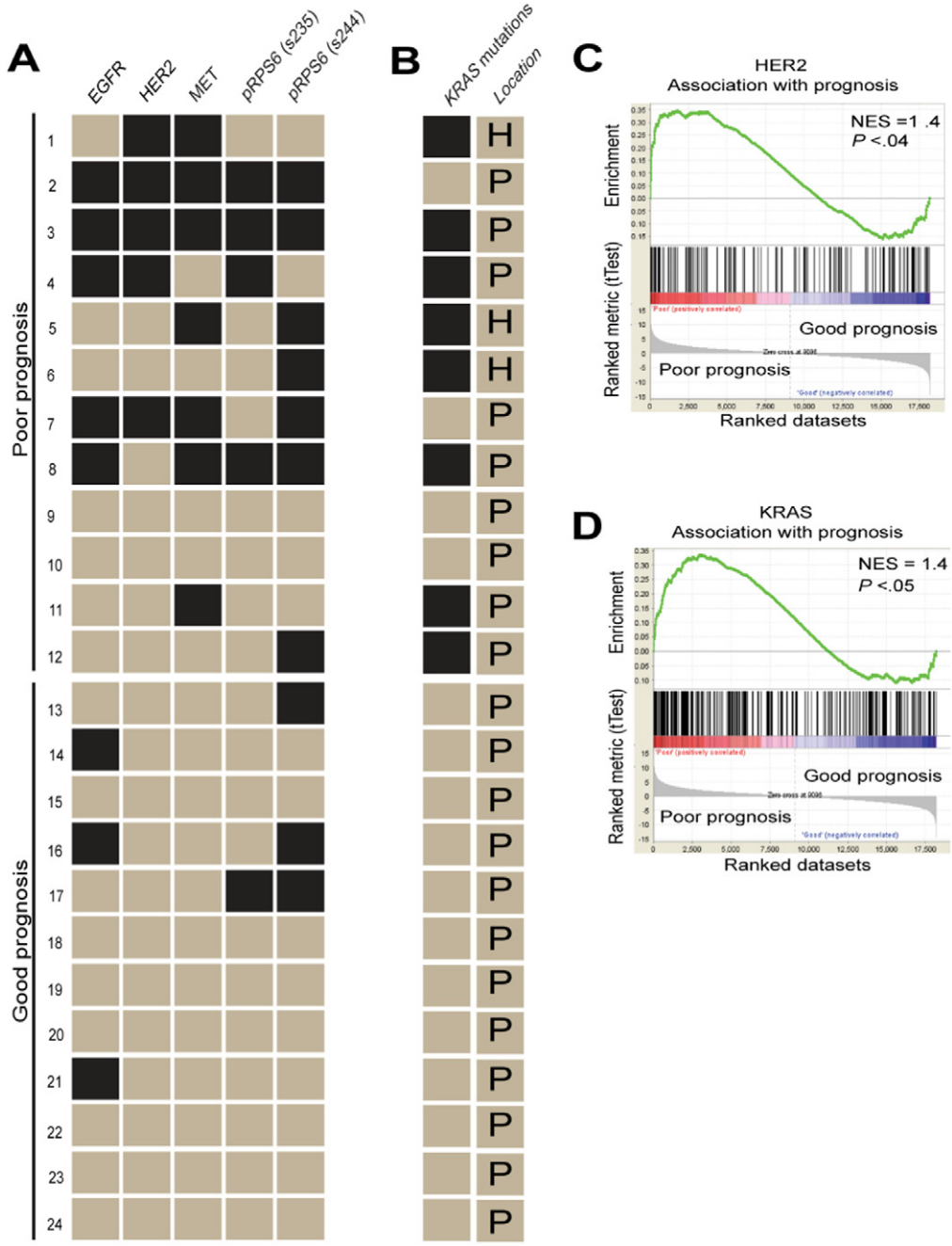
	Gene symbol	Parametric *P* value	Permutation *P* value	Hazard ratio	Network
1	CDH3	.000144	3.00E-04	1.348	CTNNB1/MYC
2	SCEL	.000627	8.00E-04	1.551	CTNNB1/MYC
3	CORO2A	.00077	7.00E-04	1.739	CTNNB1/MYC
4	QPCT	.000962	7.00E-04	1.67	CTNNB1/MYC
5	ANLN	.001047	.0012	1.534	CTNNB1/MYC
6	VTCN1	.001535	9.00E-04	0.557	CTNNB1/MYC
7	CDCP1	.002632	.003	1.642	CTNNB1/MYC
8	KRT17	.00328	.0031	1.182	CTNNB1/MYC
9	SYT12	.003786	.0034	0.531	CTNNB1/MYC
10	SIX4	.004195	.0046	1.413	CTNNB1/MYC
11	CEACAM6	.00917	.0064	1.366	CTNNB1/MYC
12	ADAM8	.000166	1.00E-04	1.581	TNF
13	GPR110	.000208	1.00E-07	1.744	TNF
14	SPOCD1	.002228	.0019	1.638	TNF
15	NUAK2	.007836	.008	0.675	TNF
16	ITGA2	1.59E-05	1.00E-07	2.015	VEGF/ERRB
17	PLEK2	.000319	1.00E-07	1.647	VEGF/ERRB
18	KRT19	.000516	4.00E-04	1.714	VEGF/ERRB
19	CST6	.001502	.0018	1.318	VEGF/ERRB
20	TMPRSS4	.002066	.0014	1.313	VEGF/ERRB
21	GJB3	.00284	.0025	1.389	VEGF/ERRB
22	UBE2C	.003421	.0027	1.468	VEGF/ERRB
23	CDC20	.003436	.0031	1.47	VEGF/ERRB
24	NRP2	.003883	.0038	1.596	VEGF/ERRB
25	MTMR11	.007299	.0053	1.617	VEGF/ERRB
26	TPX2	.009626	.0081	1.397	VEGF/ERRB
27	FXYD3	.009634	.0082	1.374	VEGF/ERRB
28	HK2	.009794	.0094	1.342	VEGF/ERRB
29	TTK	.009948	.0095	1.346	VEGF/ERRB
30	LOC201175	.000153	1.00E-07	1.829	Unknown association
31	C15orf48	.004142	.0035	1.422	Unknown association
32	C19orf33	.004256	.004	1.395	Unknown association
33	ASPHD2	.004468	.0029	1.532	Unknown association
34	GALNT6	.005952	.0049	1.461	Unknown association
35	C9orf140	.008904	.0083	1.411	Unknown association
36	CLIC3	.009536	.0073	1.309	Unknown association

Supplementary Table 3. Quantitative Trait Analysis (Molecular Functions)

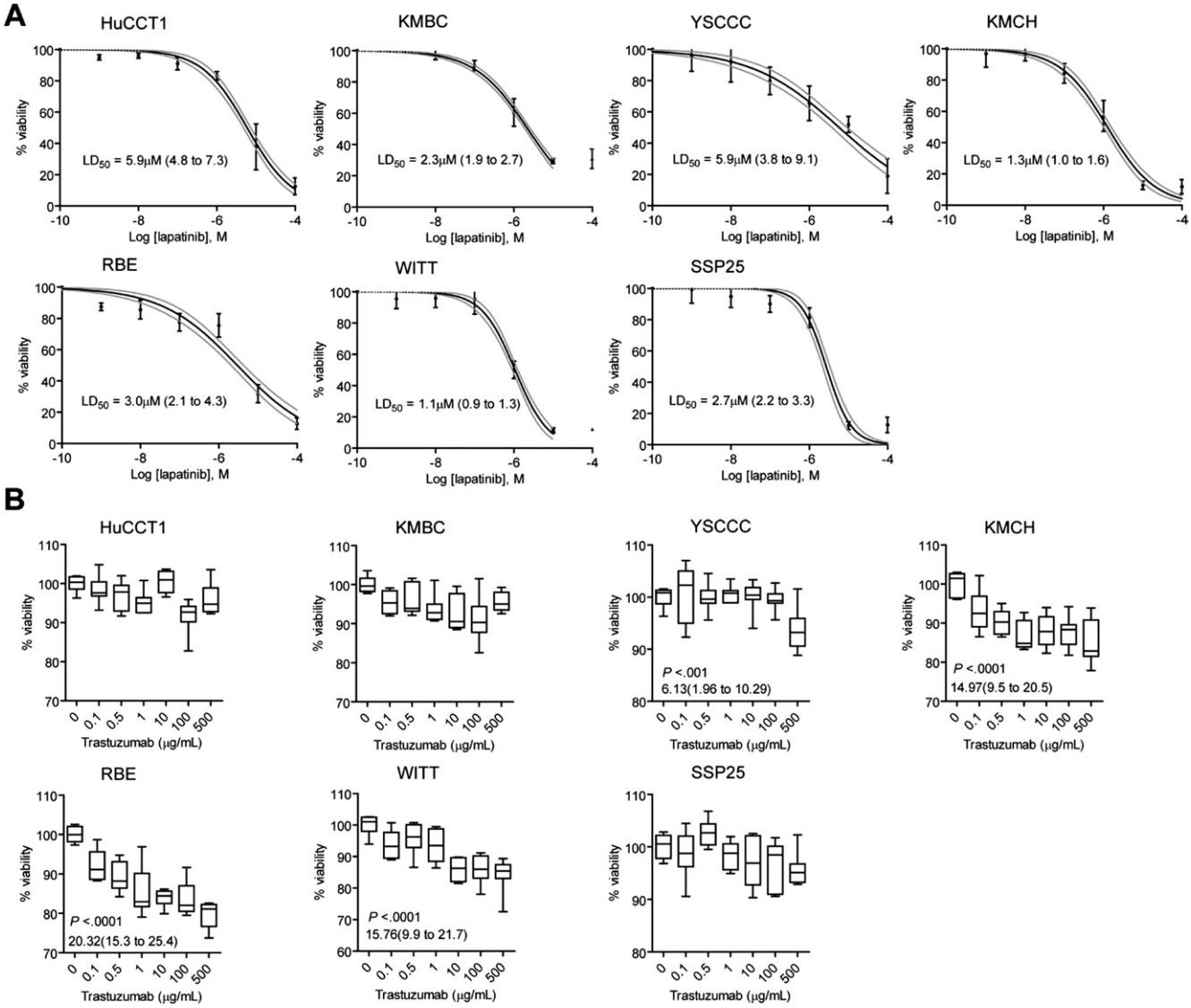
GO ID	GO term	Observed	Expected	Observed/expected
GO:0050431	TGF-β binding	5	1.23	4.05
GO:0032813	TNFR superfamily binding	5	1.44	3.47
GO:0004181	Metallocarboxypeptidase activity	7	2.06	3.4
GO:0008235	Metalloexopeptidase activity	8	2.47	3.24
GO:0005044	Scavenger receptor activity	8	2.47	3.24
GO:0004715	Protein tyrosine kinase activity	6	1.85	3.24
GO:0005507	Copper ion binding	7	2.26	3.09
GO:0005520	IGF binding	6	2.06	2.92
GO:0001948	Glycoprotein binding	6	2.06	2.92
GO:0001653	Peptide receptor activity	6	2.06	2.92
GO:0008528	Peptide receptor activity, G-protein coupled	5	1.85	2.7
GO:0005100	Rho GTPase activator activity	5	1.85	2.7
GO:0004180	Carboxypeptidase activity	8	3.09	2.59
GO:0051015	Actin filament binding	9	3.91	2.3
GO:0005099	Ras GTPase activator activity	6	2.88	2.08
GO:0004888	Transmembrane receptor activity	73	36.4	2.01
GO:0008081	Phosphoric diester hydrolase activity	7	3.5	2



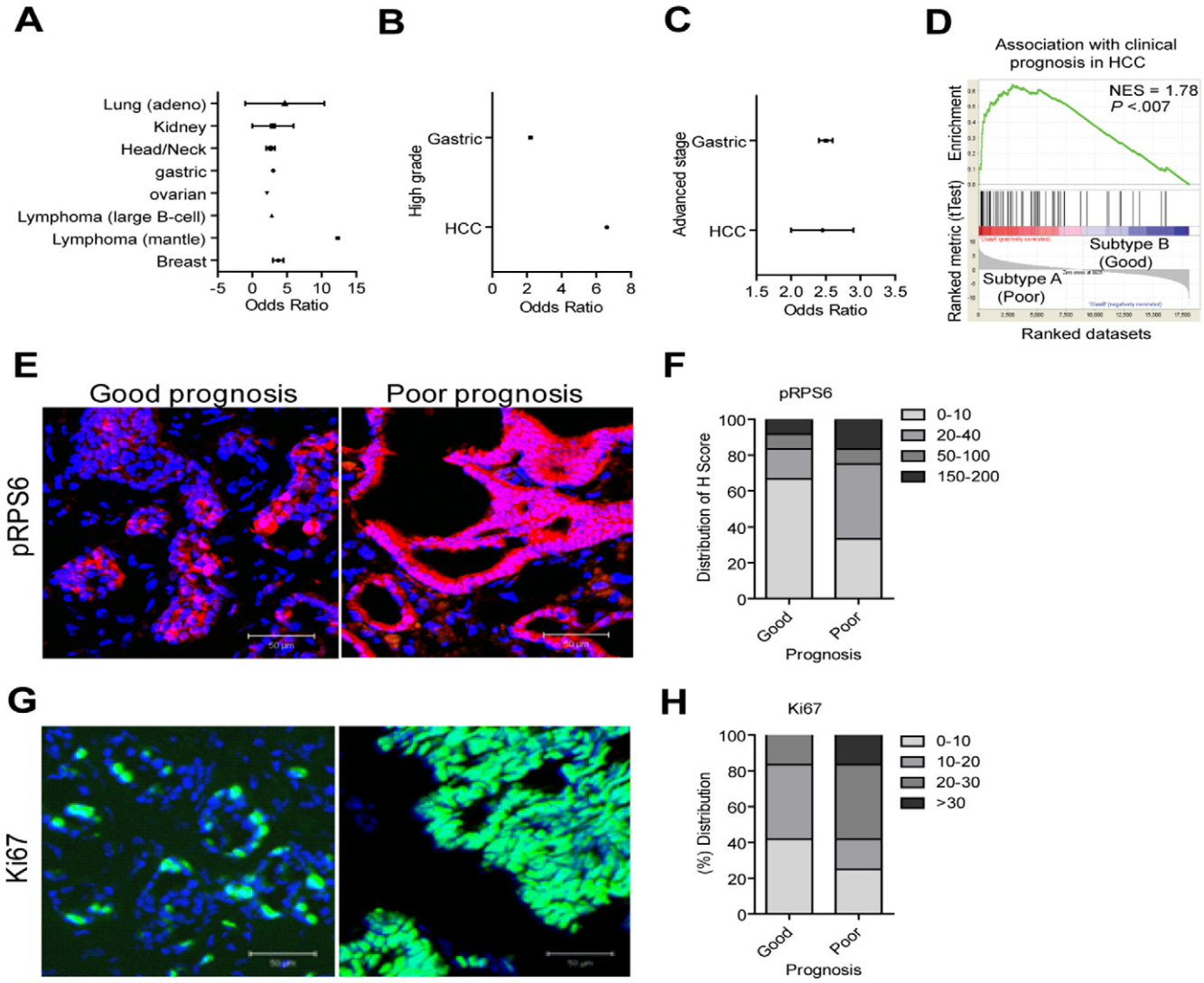
Supplementary Figure 4. Network analysis of the tumor microenvironment. Ingenuity Pathway Analysis tools were used to analyze the 1442 genes differentially expressed genes between the epithelial and stromal cell compartments. Up-regulated and down-regulated genes in tumor epithelium are shown in *red* and *green*, respectively. *Dashed lines* represent indirect connectivity.



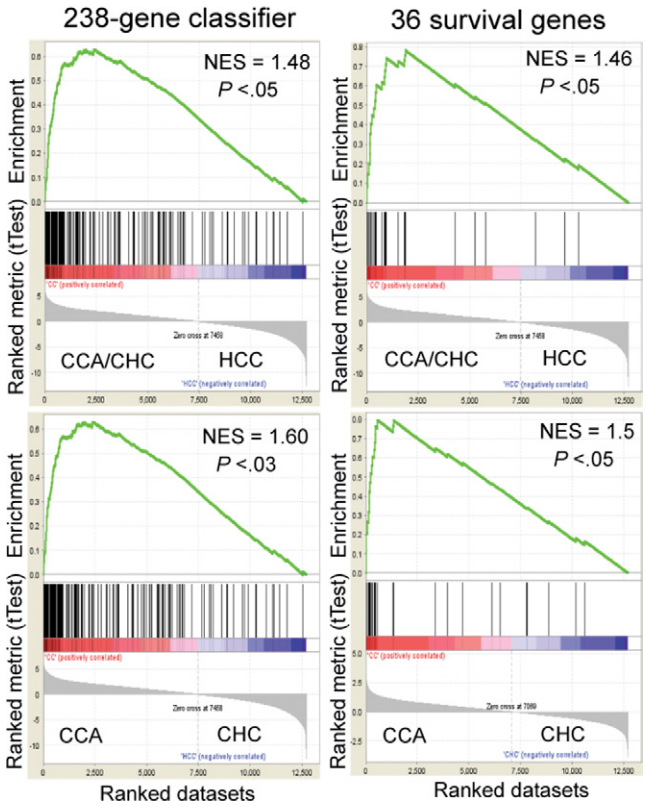
Supplementary Figure 5. Cross-activation of oncogenic pathways linked to prognosis. (*A*) Frozen serial tumor sections were stained with EGFR, MET, HER2, and pRPS6 (s235 and s244) antibodies. Expression levels were evaluated by *H*-score calculated by multiplying the staining intensity score (0, none; 1, weak; 2, moderate; 3, intense) by the percentage of cells stained at a given intensity (0−100). Image brightness parameters were adjusted across all comparison groups, and the median *H*-score in the poor prognosis group was used as the cutoff criterion for each staining. *Black squares* indicate a subject where the particular stain was above the cutoff. Numbers in the left margin denote individual patients. (*B*) The *KRAS* mutational status and location of tumors analyzed (P, peripheral; H, hilar). *Black squares* indicate subjects with mutated *KRAS*. (*C* and *D*) Gene set enrichment analysis showed enrichment and positive association of (*C*) HER2 and (*D*) KRAS with poor prognosis. The gene signatures were obtained from the Molecular Signatures Database.



Supplementary Figure 6. Estimate of LD_{50} for lapatinib and trastuzumab. Seven human CCA cell lines were treated with TKIs for 3 days. Growth inhibition was evaluated by WST viability assay. The experiments were repeated (n = 8). (*A*) For lapatinib, LD_{50} was calculated as log (inhibitor) versus variable normalized drug-response ± 95% CI using a nonlinear regression (dose range of 0, 10 nmol/L, 100 nmol/L, 1 μmol/L, 10 μmol/L, and 100 μmol/L). (*B*) The inhibitory effect of trastuzumab did not exceed 20% in any of the examined cell lines even at the maximum dose. The effect of trastuzumab at each concentration (0.1, 0.5, 1, 10, 100, and 500 μg/mL) is shown as *box plots*; *whiskers* are given for the 5th to 95th percentiles. The statistical significance was determined by one-way analysis of variance with Dunnett's multiple comparison tests and presented as mean difference ± 95% CI at the maximum upper limit dose used throughout subsequent experiments.



Supplementary Figure 7. Meta-analysis using the gene classifier. (*A*) The classifier was significantly associated with a study showing prognostic outcome if the odds ratio (OR) > 2 ($P < .0001$). Forest plot represents the OR $\pm$ 95% CI. (*B* and *C*) Association of the classifier with (*B*) high-grade and (*C*) advanced-stage HCC and gastric cancers, respectively. (*D*) GSEA. The classifier showed a significant enrichment and positive correlation with poor prognosis HCC (subtype A). (*E* and *F*) Because meta-analysis of the classifier showed a significant association with pRPS6-positive HCC, immunohistochemical analyses of (*E*) the mTOR pathway and (*G*) proliferation were performed in representative tumor samples from each prognostic subgroup (n = 12). (*E* and *F*) Tumors were stained for pRPS6 *(E)* and intensity of immunostaining was evaluated by *H*-score (distribution of bin values is shown in *F*). (*G* and *H*) Tumors in the poor prognostic group showed a higher rate of proliferation than tumors from patients with good prognosis. (*G*) Tumor samples from each prognostic subgroup (n = 12) were stained for Ki67. (*H*) Distribution of tumors with different percentage of Ki67-positive cells (shown on the *right*). *Scale bar* = 50 μm.



Supplementary Figure 8. Association of the gene classifier with CCA, hepatocellular CCA (CHC), and HCC. The 238-gene classifier and 36 survival genes showed significant enrichment and positive correlation with CCA compared with CHC and HCC, respectively. Data obtained from Woo et al.[2]