Supplementary Appendix

Integrated genomic analysis of recurrence-associated small non-coding RNAs in esophageal cancer

Hee-Jin Jang, M.D.; Hyun-Sung Lee, M.D., Ph.D.; Bryan M. Burt, M.D.; Geon Kook Lee, M.D., Ph.D.; Kyong-Ah Yoon, Ph.D.; Yun-Yong Park, Ph.D.; Bo Hwa Sohn, Ph.D.; Sang Bae Kim, Ph.D.; Moon Soo Kim, M.D.; Jong Mog Lee, M.D.; Jungnam Joo, Ph.D.; Sang Cheol Kim; Ju Sik Yun, M.D.; Kook Joo Na, M.D., Ph.D.; Yoon-La Choi, M.D.; Jong-Lyul Park, Ph.D.; Seon-Young Kim, Ph.D.; Yong Sun Lee, Ph.D.; Leng Han, Ph.D.; Han Liang, Ph.D.; Duncan Mak; Jared K. Burks; Jae Ill Zo,M.D.,Ph.D.; David J. Sugarbaker, M.D.; Young Mog Shim, M.D., Ph.D.; Ju-Seog Lee, Ph.D.

Table of Contents

SUPPLEMENTARY METHODS

Exclusion criteria

We did not enroll patients who: (1) received perioperative chemotherapy and/or radiotherapy at any period except confirmation of recurrence, (2) had a prior cancer diagnosis, (3) had double primary synchronous cancer, (4) had cervical esophageal cancer, or (5) were not monitored completely.

Preoperative evaluation and surgical policy

Preoperative evaluations included routine blood examination, esophagogastroduodenoscopy with biopsy, endoscopic ultrasonography, chest and abdominal computed tomography (CT), pulmonary function testing, electrocardiography, positron emission tomography (PET)-CT, bronchoscopy, and, if necessary, cardiac function testing. Tumors were staged according to the 7th edition of the American Joint Committee on Cancer TNM staging system.[49]

For patients with middle and lower ESCC, a two-field lymph node dissection and an intrathoracic esophagogastrostomy were performed using the entire stomach as a conduit, and an anastomosis was performed with a 28-mm end-to-end anastomosis stapler (Autosuture, U.S. Surgical Corp., Norwalk, CT, USA). For patients with upper esophageal cancer, a three-field lymph node dissection was routinely performed. If an intrathoracic esophagogastrostomy was possible, an intrathoracic anastomosis was performed using the entire stomach. For all others, a cervical esophagogastrostomy with a gastric tube was performed. The gastric tube was made with 75-mm and 55-mm titanium linear cutter staplers (Ethicon Ltd., Somerville, NJ, USA), and the cervical anastomosis was performed in the left side of the neck with a 25-mm end-to-end anastomosis stapler. The stomach was positioned in the posterior mediastinum. A pyloroplasty was routinely performed in all cases using finger disruption of the pylorus; the pylorus was pinched between the index finger and thumb until the pylorus ring was broken off.

An esophagography was performed on postoperative day 7. Patients were allowed to take sips of water after the absence of an anastomosis leak was confirmed, and a full liquid diet was implemented on the following day. If the patient tolerated the liquid diet, the diet was then changed to soft foods. We encouraged patients to ambulate soon after food was tolerated.

Follow-up

Survival time was defined as the time elapsed between the operation and death or between the operation and the most recent follow-up visit. In the NCCK, regular follow-up was conducted by telephone or mail twice a year, in April and October. A chest CT was routinely performed during the first follow-up visit after discharge. In addition, all patients in NCCK and SMC, the high-volume centers, underwent regular evaluations including a routine blood examination, chest x-ray, and chest CT every 3 months during the first 2 years. Subsequently, all patients were monitored annually. PET-CT and esophagogastroduodenoscopy were performed annually or more frequently if necessitated according to clinical history and clinical examination findings. However, the follow-up period, methods, and modality varied in CNU, the low-volume center.

Sample preparation

Total RNA from fresh frozen tissues was extracted with TRIzol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's procedures. RNA quality was assessed with an Agilent 2100 bioanalyzer and the

RNA 6000 NanoChip kit (Agilent Technologies, Santa Clara, CA), and the quantity was determined with an ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

FFPE samples were sliced to 10 μm thickness, and 2 slices were put into a 1.5-ml tube. After deparaffinization, total RNA was extracted with a RecoverAll Total Nucleic Acid Isolation Kit (Ambion, Austin, TX, USA) following the manufacturer's instructions. RNA quality and quantity were assessed with the aforementioned methods.

Small non-coding RNA microarray hybridization and analysis

For sncRNA expression array analysis, we used a GeneChip miRNA 2.0 array (Affymetrix). We labeled 1 μg of total RNA with the FlashTag Biotin HSR RNA Labeling Kit (Genisphere LLC, Hatfield, PA, USA) following the manufacturer's recommendations. Then, labeled sncRNA was hybridized to the array and incubated as described in the manufacturer's protocol (Affymetrix). The chips were then washed and stained with GeneChip Fluidics Station 450 (Affymetrix) and scanned with a GeneChip scanner 3000 7G (Affymetrix). Feature extraction was performed with Affymetrix Command Console software. We analyzed all microarray data with the Robust MultiArray Average algorithm and implemented quantile normalization[50] with log2 transformation of gene expression intensities with BRB-Array Tools version 4.3.0 (Biometric Research Branch, National Cancer Institute, Bethesda, MD, USA)[51] and the R-script from the Bioconductor project (www.bioconductor.org). Then, we selected the human sncRNAs and adjusted data with mean values for genes and arrays, respectively. An unsupervised hierarchical clustering algorithm was applied using the uncentered correlation coefficient as the measure of similarity and the method of average linkage[52] (Cluster 3.0). Java Treeview 1.60 (Stanford University School of Medicine, Stanford, CA, USA) was used for tree visualization. The microarray data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus **(GSE55857).**

Northern blot

To determine the identity of miR-886 in ESCC cells, we performed Northern blot assays. Briefly, total RNAs (1 μg each) from cell lines were run on 15% denaturing polyacrylamide gels, and then transferred onto Genescreen Plus Hybridization Transfer Membranes (Perkin-Elmer, Waltham, MA, USA). Hybridization was done in Hybridization Buffer ULTRAhyb®-Oligo (Grand Island, NY, USA) containing an antisense oligonucleotide against miR-886-3p (whose sequence is 5'-AAAAGGGTCAGTAAGCACCCGCG-3') that was 5'-end labeled with γ-^{32}P-ATP. After hybridization in a 37°C oven overnight, the membrane was washed with 2XSSC/0.5% SDS, twice at room temperature for 5 min each, and then twice again at 37°C for 30 min.

Gene expression microarray hybridization and analysis

To identify the mRNA expression profile of ESCC cancer samples, we carried out mRNA microarray experiments with a HumanHT-12 v4 Expression Beadchip Kit (Illumina, San Diego, CA, USA). With a TotalPrep RNA Amplification Kit (Illumina), we labeled and hybridized 750 ng of total RNA according to the manufacturer's protocols. After beadchips were scanned with a BeadArray Reader (Illumina), microarray data were normalized via quantile normalization,[50] and the normalized values were transformed logarithmically on a base 2 scale with the R-script. The microarray data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus **(GSE55856).**

qRT-PCR

To verify the sncRNA expression profiles on microarrays of fresh frozen tissue, the abundance of miR-223, miR-1269a, and nc886 was measured with qRT-PCR. We also used qRT-PCR to evaluate the expression level of 3 sncRNA signatures in samples obtained from SMC and CNU. Reverse transcriptase reactions for two miRNAs (miR-223 and miR-1269a) were performed with a miScript II RT Kit (Qiagen, Valencia, CA, USA) and 500 ng of total RNA per reaction, as described in the manufacturer's manual. We performed qRT-PCR for miRNA with a miScript SYBR Green PCR Kit (Qiagen) and cDNA equivalent of 500 ng total RNA per reaction with Universal RT primers at 95°C for 15 minutes, followed by 40 cycles at 94°C for 15 seconds, at 55°C for 30 seconds, and at 70°C for 30 seconds. We used the mean value of two stable miRNAs (miR-132 and miR-652) as an endogenous control to normalize the data for the ΔΔCt method of relative quantification.

For nc886, cDNA was synthesized from 500 ng of total RNA with the amfiRivert Platinum cDNA synthesis mix (GenDepot, Barker, TX, USA) and amplified with SYBR Premix Ex Taq II (Tli TNaseH Plus) (TaKaRa, Mountain View, CA, USA), according to the manufacturer's instructions. The cycling conditions were 95°C for 30 seconds, followed by 40 cycles at 95°C for 15 seconds, at 64°C for 20 seconds, and at 72°C for 20 seconds. Relative amounts of mRNA were calculated from the threshold cycle number using the expression of PPIA (Qiagen) as an endogenous control.[14]

We performed qRT-PCR with the Eppendorf Mastercycler ep Gradient S Thermocycler (Eppendorf, Hamburg, Germany). Each sample was amplified in triplicate and the values were averaged. The primers used are described in **Supplementary Table 4**.

Cell culture

An immortalized esophageal epithelial cell line (Het-1A), a Barrett's esophageal cell line (BE-3), an esophageal metaplastic cell line (OE-33), an esophageal adenocarcinoma cell line (SK-4), and 4 ESCC cell lines (TE-1, TE-8, TE-12, and TT) were used in this study. All cell lines were provided by Drs. Xiaochun Xu and Julie J. Izzo at The University of Texas M.D. Anderson Cancer Center. Het-1A cells were maintained in keratinocyte serum-free medium containing 5 ng/ml human recombinant epidermal growth factor (Sigma-Aldrich, St Louis, MO, USA), 0.05 mg/ml bovine pituitary extract (Invitrogen), 0.005 mg/ml human recombinant insulin (Invitrogen), and 500 ng/ml hydrocortisone (Sigma-Aldrich) at 37°C in a humidified atmosphere of 5% CO_2, as recommended by ATCC. BE-3 and TT cell lines were maintained in Dulbecco's modified Eagle medium (Corning, Manassas, VA, USA), and the other cells were maintained in RPMI 1640 medium (Corning) supplemented with 10% fetal bovine serum (Life Technologies, Grand Island, NY, USA) and 1% antibiotic-antimycotic solution (Sigma-Aldrich) at 37°C in a humidified atmosphere of 5% CO_2. Total RNA was purified with the mirVana miRNA Isolation Kit (Ambion), and RNA quality and quantity were checked via the aforementioned method for further analysis. Contaminating genomic DNA was removed via treatment with DNase. Cell lines were validated by STR DNA fingerprinting using the AmpF_STR Identifiler kit (Applied Biosystems, Grand Island, NY) according to manufacturer's instructions. The STR profiles were compared to known ATCC fingerprints (http://www.atcc.org/) and to the Cell Line Integrated Molecular Authentication database (CLIMA, version 0.1.200808, http://bioinformatics.istge.it/clima/). The STR profiles matched known DNA fingerprints or were unique. Cell lines were tested for mycoplasma contamination

Apoptosis Assay for Drugs

All compounds (Panobinostat (LBH589; Catalog No. S1030), LAQ824 (Dacinostat; Catalog No. S1095), Vorinostat (SAHA, MK0683; Catalog No. S1047), BI2536 (Catalog No. S1109), BI6727(Volasertib; Catatlog No. S2235), BEZ235 (NVP-BEZ235, Dactolisib; Catalog No. S1009), and Tacrolimus (FK506; Catalog No. S5003),) were purchased from Selleckchem (www.selleckchem.com).

After overnight incubation with 10^6 cells from each cancer cell line (Het-1a (low risk), TE-1(intermediate risk), and TE-8 (high risk)), IC50 dose of each drug for high risk cell line was administrated on 6-well plates. Following drug addition, the plates were incubated for 24 hours at 37°C in a humidified

atmosphere of 5% CO_2. And then, the apoptotic assay was performed by using Annexin V-FITC and DAPI. All experiments were independently repeated in duplicate.

Single cell analysis with flow cytometry and mass cytometry

Single cell preparations from whole blood

Human fresh blood samples from healthy donors (Gulf Coast Regional Blood Center) were diluted 1:1 with PBS in a conical tube. The diluted samples were underlaid with a volume of Ficoll® that is equal to the original sample volume, and then centrifuged for 20 minutes (800 x g) with the brake OFF. The PBMCs located at the interface of the PBS and Ficoll® layers were harvested and transferred into a fresh tube filled with PBS to wash the cells. The cell suspension was centrifuged for 5 minutes (400 x g) at 20°C, and the supernatant discarded. The cell pellet was resuspended in freezing media and 1mL of cell suspension was aliquoted into each cryovial. The cryovials were placed in a Mr. Frosty and in a -80°C freezer for 24 hours. Immediately after this, the cryovials with PBMCs were placed into a liquid nitrogen tank for long term storage.

Generation of monocyte-derived dendritic cells (moDC)

To generate moDCs from freshly isolated PBMCs we used dendritic cell generation medium DXF(C-28052) from PromoCell. Freshly isolated cells were plated in an appropriate amount of PromoCell DC Generation Medium DXF without cytokines. Mononuclear cells (2-3 million/cm^2) were incubated for 1 hour at 5% CO_2 and 37°C in the incubator. Non-adherent cells were loosened by vigorously swirling the tissue culture vessel. The cells were then aspirated and an appropriate amount of PromoCell DC Generation Medium DXF supplemented with 1x Cytokine Pack moDC DXF (IL-4 and GM-CSF) was added. The cells were incubated for 3 days at 37°C and 5% CO_2. We changed the medium on day 3. On day 6, we harvested immature moDC. To complete the moDC maturation process, the entire volume with 1x of Component B of the Cytokine Pack moDC DXF (supplied at 100x) was supplemented on day 6 without changing the medium. Cells were then incubated at 37°C and 5% CO_2 for an additional 48 hours. After dislodging loosely attached cells by pipetting up and down several times, we transferred the medium containing the cells into a 50 ml tube, as mature moDCs are non-adherent cells and exhibit a unique morphology originating from their multiple long thread-like dendrites. We spun down harvested moDCs at 180 x g for 10 minutes and discarded the supernatant.

In vitro simulation of tumor microenvironment

Pooled IgG was isolated from plasma from ten healthy donors using NAb™ Protein G Spin Columns (Thermo Scientific). The human subject's protocols were approved by Baylor College of Medicine and St. Luke's Hospital's Institutional Review Board. For tumor–antibody complexes, Tumor cells were fixed in 2% paraformaldehyde for 20 min, washed extensively in PBS and coated for 30 min with allogeneic IgG, and incubated with 1–3 µg allogeneic IgG per 1 x 10^5 tumor cells, and were then washed to remove excess antibodies.[53]

We mixed tumor cells and immune cells by combining 1x10^5 cells from the PBMC preparation and 1x10^5 cells from the moDC preparation and adding the combination to each cancer cell line containing 5x10^5 cells on 6 wells.[54-57] imoDCs are activated with 1 µg/ml high molecular mass polyinosinic-polycytidylic acid (poly(I:C)) (InvivoGen, San Diego, California). After a 24 hour incubation at 37°C with 5% CO_2, the cells are harvested for flow cytometry or CyTOF. All *in vitro* activations of DCs were independently repeated in duplicate.

Flow cytometry

For cell surface staining, monoclonal antibodies conjugated to PE-Cy5, PE, PE-Cy7, and Brilliant Violet 421 specific for the following antigens were used: CD45(TU116), CD40 (5C3), CD86(2331), HLA-

DR(G46-6) from BD Biosciences (San Jose, California). Flow cytometry was performed on a LSR Fortessa (BD Biosciences) and data sets were analyzed using FlowJo software (Tree Star, Inc.).

Mass cytometry (CyTOF)

Single cell preparations were washed with staining buffer; then, antibodies and cells were mixed and kept at room temperature for 30 min. Antibodies were chosen to facilitate the identification of cancer cells and major immune cell types and to define the immune phenotype likely to be affected by immune stimulation. The 17 antibodies used are listed in **Supplementary Table 8**. Five minutes before staining with antibodies, cell ID™-cisplatin (MaxPar®) was be added to check cell viability and then, the samples were fixed with 1.6% paraformaldehyde. In addition, we used permeabilization and intercalation with Cell-ID™ Intercalator-Ir. Cell-ID Intercalator-Ir is a cationic nucleic acid intercalator that contains naturally abundant Iridium (191Ir and 193Ir) and is used for identifying nucleated cells in CyTOF® analysis. When cells are stained with Intercalator-Ir, it binds to cellular nucleic acids. Detection of both stable isotopes enabled the identification of nucleated cells. Intercalator-Ir is a live cell membrane-impermeable dye and therefore requires the cells to be fixed and/or permeabilized before staining. After overnight intercalation at 4°C, stained cells were analyzed on a mass cytometer (CyTOF2™ mass cytometer, DVS Sciences) at an event rate of 400 to 500 cells per second. All experiments were performed in duplicate. Data files for each sample were normalized with Normalizer v0.1 MCR and gated (**Supplementary Figure 14**).

References

49. Rice TW, Rusch VW, Apperson-Hansen C, et al. Worldwide esophageal cancer collaboration. Dis Esophagus 2009;22:1-8.
50. Bolstad BM, Irizarry RA, Åstrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185-93.
51. Simon R, Lam A, Li M-C, et al. Analysis of gene expression data using BRB-Array Tools. Cancer Inform 2007;3:11-7.
52. Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863-68.
53. Carmi Y, Spitzer MH, Linde IL, et al. Allogeneic IgG combined with dendritic cell stimuli induce antitumour T-cell immunity. Nat 2015;521:99-104.
54. Romagnoli GG, Zelante BB, Toniolo PA, et al. Dendritic cell-derived exosomes may be a tool for cancer immunotherapy by converting tumor cells into immunogenic targets. Front Immunol 2014;5:692.
55. Kuhn S, Ronchese F. Monocyte-derived dendritic cells: Emerging players in the antitumor immune response. Oncoimmunology 2013;2:e26443.
56. Palucka K, Banchereau J. Cancer immunotherapy via dendritic cells. Nat Rev Cancer 2012;12: 265-77.
57. Rauch C, Ibrahim H, Foster N. Membranes, molecules and biophysics: enhancing monocyte derived dendritic cell (MDDC) immunogenicity for improved anti-cancer therapy. J Cancer Ther Res 2013; 2:20,.

Supplementary Table 1. Clinico-pathologic characteristics of patients.

Variables		Discovery cohort	Validation cohort	P value	TCGA ESCA	TCGA LUSC
Number of patients		**108**	**214**		**151**	**147**
Sex	Men	102 (94.4 %)	206 (96.3 %)	0.564	130 (86.1%)	101 (68.7 %)
	Women	6 (5.6 %)	8 (3.7 %)		21 (13.9 %)	46 (31.3 %)
Age, median (range), y		66 (46-82)	66 (40-88)	0.631	62 (27-90)	71 (39-84)
Smoking		69 (63.9 %)	140 (65.4 %)	0.805	92 (60.9 %)	111 (75.6 %)
Pathologic T status	Tis	-	3 (1.4 %)	<0.001		
	1	5 (4.6 %)	56 (26.2 %)		25 (16.5 %)	31 (21.1%)
	2	12 (11.1 %)	41 (19.2 %)		33 (21.8 %)	98 (66.7%)
	3	84 (77.8 %)	100 (46.7 %)		75 (49.7 %)	11 (7.5%)
	4	7 (6.5 %)	14 (6.5 %)		4 (2.7 %)	7 (4.8%)
	NA				14 (9.3 %)	
Pathologic N status	0	38 (35.2%)	108 (50.5%)	0.083	64 (42.4 %)	88 (59.9%)
	1	43 (39.8%)	63 (29.4%)		55 (36.4 %)	45 (30.6%)
	2	18 (16.7%)	29 (13.6%)		11 (7.3 %)	13 (8/8%)
	3	9 (8.3%)	14 (6.5%)		6 (4.0 %)	
	X				2 (1.3 %)	1 (0.7%)
	NA				13 (8.6%)	1 (0.7%)
Histologic differentiation	Well	10 (9.3%)	59 (27.6%)	<0.001	15 (9.9 %)	
	Intermediate	81 (75%)	126 (58.9%)		59 (39.1 %)	-
	Poorly	17 (15.7%)	29 (13.5%)		40 (26.5 %)	
	NA				37 (24.5 %)	
Tumor location	Upper	22 (20.4%)	12 (5.6%)	0.011	4 (2.7 %)	
	Mid	54 (50%)	109(50.9%)		38 (25.2 %)	-
	Lower	32 (29.6%)	93 (43.5%)		108 (71.5 %)	
					1 (0.6 %)	
Pathologic TNM stage*	0	-	3 (1.4%)	<0.001		
	I	-	46 (21.5%)		18 (11.9 %)	80 (54.4%)
	II	39 (36.1%)	85 (39.7%)		62 (41.0 %)	42 (28.6%)
	III	69 (63.9%)	80 (37.4%)		49 (32.5 %)	23 (15.6%)
	IV				5 (3.3 %)	2 (1.4%)
	NA				17 (11.3 %)	
Postoperative complications		48 (44.4%)	97 (45.3%)	0.906	-	-
Follow-up, median, mo		21.6	35.4	0.583	16.2	50.3

* The pathologic TNM stage was determined according to 7th edition of the American Joint Committee on Cancer staging manual. The staging system in TCGA cohorts has been combined with that in the 6th edition.

ESCA denotes esophageal cancer, LUSC lung squamous cell carcinoma, and TCGA The Cancer Genome Atlas.

Supplementary Table 2. Summarized expression of mapped Affymetrix sncRNA type.

Category	Abbreviation	Number
Mature microRNA	miRNA	1251
Precursor microRNA	pre-miRNA	1121
Small nucleolar RNA	SnoRNA	1792
Small nucleolar RNA C/D box	SnoRNA C/D box	282
Small nucleolar RNA H/ACA box	SnoRNA H/ACA box	161
Small Cajal body-specific RNA	ScaRNA	32
Total		4639

Supplementary Table 3. Univariable Cox regression analysis of the differentially expressed small non-coding RNAs associated with recurrence-free survival in the discovery set of 108 patients with esophageal squamous cell carcinoma.

sncRNAs	Hazard Ratio* (95% CI)	Regression coefficient	Z score	P value	False Discovery Rate
hsa-miR-223	0.715 (0.568-0.9)	-0.336	-2.86	0.004	0.017
hsa-miR-193a-3p	0.738 (0.568-0.96)	-0.303	-2.268	0.023	0.033
hsa-miR-886-3p	0.82 (0.685-0.983)	-0.198	-2.143	0.032	0.039
tRNA (Lys3)	1.281 (1.021-1.608)	0.248	2.135	0.033	0.039
SNORA24	0.748 (0.565-0.989)	-0.29	-2.036	0.042	0.041
hsa-miR-1269a	1.163 (1.004-1.348)	0.151	2.008	0.045	0.041
SNORA40	0.736 (0.545-0.995)	-0.306	-1.994	0.046	0.041
SNORD126	0.822 (0.677-0.998)	-0.196	-1.984	0.047	0.041

* We calculated the hazard ratios and *p* values with the unadjusted Cox proportional hazards model in BRB-Array Tools 4.3.0.

CI denotes confidence interval.

Supplementary Table 4. Primers used for quantitative real time-polymerase chain reaction (qRT-PCR).

Gene	Primer Direction	Primer Sequence (5' → 3')
nc886	Forward	CGGGTCGGAGTTAGCTCAAGCGG
	Reverse	AAGGGTCAGTAAGCACCCGCG
RNU6	Forward	TGCTCGCTTCGGCAGCACATAT
	Reverse	TGGAACGCTTCACGAATTTGCG
β-actin	Forward	CAAGAGATGGCCACGGCTGCT
	Reverse	TCCTTCTGCATCCTGTCGGCA
miR-132		TAACAGTCTACAGCCATGGTCG
miR-223		TGTCAGTTTGTCAAATACCCCA
miR-652		AATGGCGCCACTAGGGTTGTG
miR-1269a		CTGGACTGAGCCGTGCTACTGG

Supplementary Table 5. Univariable Cox regression analysis of sncRNA signature expression and survival in the combined validation sets (n=214).

Variables	Recurrence-free survival		Overall survival	
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Sex (M *vs.* F)	2.45 (0.34-17.7)	0.375	1.37 (0.33-5.65)	0.659
Age (≥65 *vs.* <65 years)	2.89 (1.46-5.40)	0.001	2.33 (1.32-4.13)	0.004
Smoking (Yes *vs.* No)	2.10 (1.11-3.99)	0.023	2.03 (1.09-3.77)	0.025
TNM staging				
T (T3-4 *vs.* T1-2)	4.73 (2.44-9.17)	<0.001	3.40 (1.88-6.13)	<0.001
N (N1-3 *vs.* N0)	4.24 (2.32-7.75)	<0.001	4.12 (2.32-7.30)	<0.001
Grade (G2-3 *vs.* G1)	1.53 (0.82-2.86)	0.180	1.65 (0.90-3.02)	0.108
Tumor Location (upper & mid *vs.* lower)	1.07 (0.63-1.83)	0.796	1.25 (0.74-2.14)	0.407
TNM stage (stage III *vs.* I & II)	5.04 (2.87-8.86)	<0.001	4.29 (2.50-7.37)	<0.001
Postoperative complications (Yes *vs.* No)	2.03 (1.18-3.49)	0.010	1.95 (1.15-3.32)	0.013
sncRNAs				
nc886	0.78 (0.69-0.89)	<0.001	0.93 (0.83-1.04)	0.216
miR-223	0.77 (0.68-0.87)	<0.001	0.95 (0.84-1.06)	0.350
miR-1269a	1.39 (1.15-1.69)	0.001	1.39 (1.14-1.68)	0.001
Recurrence-Risk Assessment Score (high *vs.* intermediate & low risk)	3.28 (1.85-5.82)	<0.001	2.28 (1.34-3.86)	0.002

CI denotes confidence interval.

Supplementary Table 6. Multivariable Cox regression analysis* of sncRNA signature expression and survival in the validation set (n=214).

	Hazard ratio (95% CI)	P value
Recurrence-free survival		
Recurrence-Risk Assessment Score (high *vs.* intermediate and low risk)	2.27 (1.26-4.09)	0.007
TNM stage (stage III *vs.* I–II)	4.13 (2.31-7.39)	<0.001
Age (≥65 years *vs.* <65 years)	2.35 (1.26-4.41)	0.008
Smoking	2.24 (1.17-4.29)	0.014
Overall survival		
Recurrence-Risk Assessment Score (high *vs.* intermediate and low risk)	1.79 (1.04-3.06)	0.035
TNM stage (stage III *vs.* I–II)	3.47 (1.99-6.03)	<0.001
Age (≥65 years *vs.* <65 years)	1.95 (1.09-3.47)	0.023

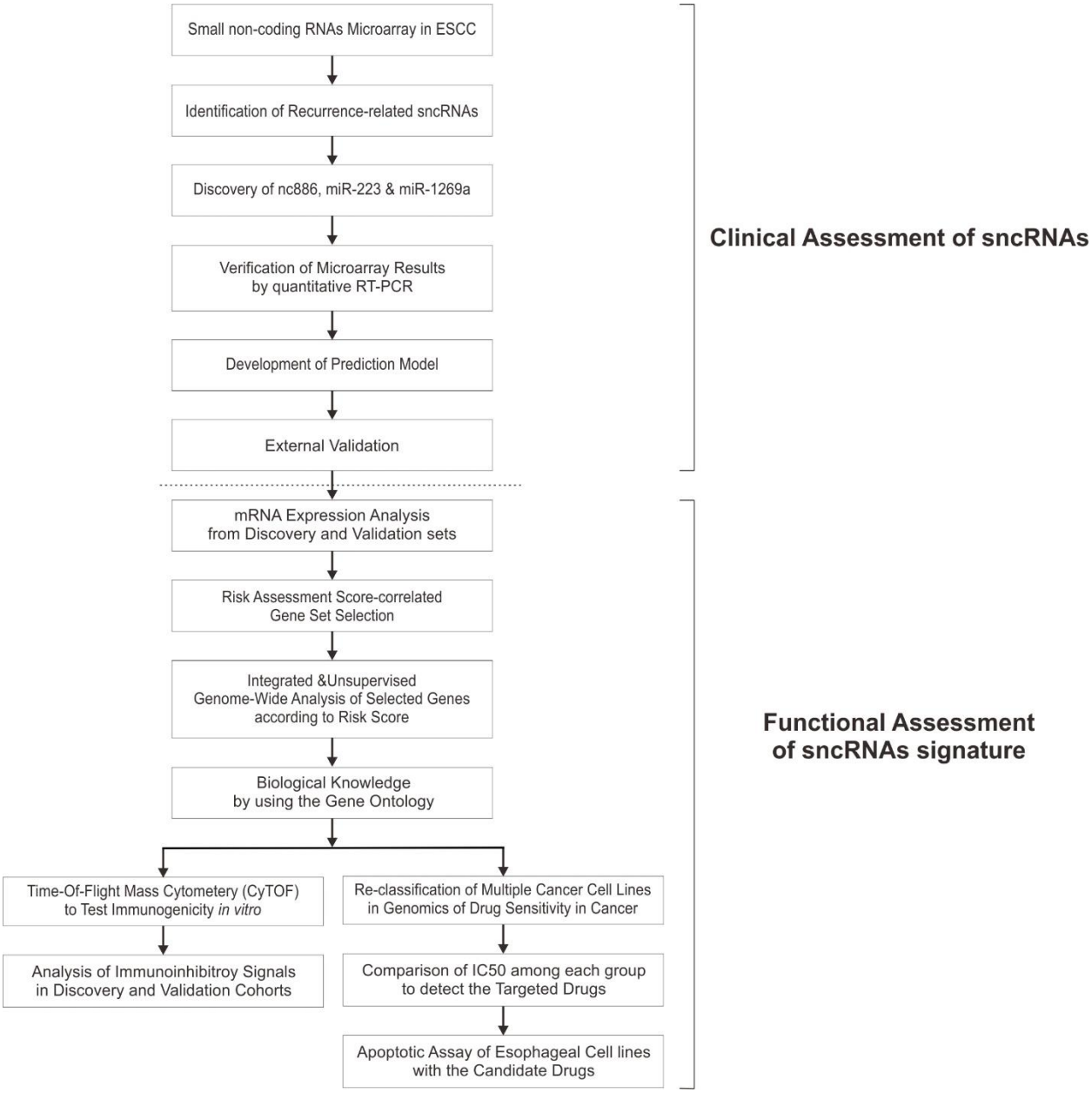
* We calculated hazard ratios and P values with an adjusted multivariate Cox proportional hazards regression model, including Recurrence-Risk Assessment Score (high risk *vs.* intermediate and low risk), sex, age (<65 years *vs.* ≥65 years), smoking, TNM stage (stage III *vs.* I–II), and postoperative complications as covariates. We selected variables with the backward stepwise approach. Only variables that were significantly associated with survival are presented (P<0.05). CI denotes confidence interval.

Supplementary Table 7. Common mRNA genes significantly correlated with RAS in the discovery, esophageal cancer, and lung squamous cell carcinoma cohorts from TCGA.

Positive correlation (507)											Negative correlation (106)		
ABCA13	BTRC	CLUAP1	EXTL2	HNRNPR	MBLAC2	PARP1	RB1CC1	SNAP47	UNG	ZNF658	ACOT11	ITGAX	TRIM54
ABCC5	C12orf26	CMAS	FADS1	HNRPDL	MCCC1	PARP2	RBBP4	SNW1	USF2	ZNF667	ADCY4	KALRN	TSPAN11
ABI2	C12orf29	CNBP	FAM117B	HPS4	MED30	PAX6	RBBP5	SP3	USP1	ZNF7	ALOX5AP	KCNK6	UNC13D
ACAP2	C12orf32	CNIH4	FAM172A	HRK	MEGF9	PBX1	RBBP8	SPAST	USP21	ZNF793	ALS2CL	KIF21B	VWA1
ACPL2	C12orf65	CNOT8	FAM179B	IER3IP1	METAP1	PCGF6	RBL1	SRGAP2	VANGL2	ZNHIT6	AMN	LILRA5	ZNF600
ACTL6A	C12orf76	COL4A5	FAM190B	IPO9	METTL14	PCMTD1	RBL2	SRP9	VPS13B	ZXDC	ANPEP	LILRB3	ZNF69
ACTR6	C14orf135	COMMD5	FAM20B	IQCB1	METTL3	PCMTD2	RBM23	SS18	VPS45	ZZZ3	APOL1	MAFF	
ACYP1	C16orf80	COPS8	FAM48A	IRF2BP2	MEX3C	PCNP	RBM4B	STAG1	WBP11		ARFGAP1	MAP2K3	
ACYP2	C16orf87	CRKL	FAM60A	IRX3	MKRN3	PDCD4	RBMX	STAU2	WBP4		ARRB1	MED8	
ADHFE1	C18orf54	CSRNP2	FAM92A1	IRX5	MLF1	PDIK1L	REV1	STMN1	WDR67		ATG4A	MEI1	
ADIPOR1	C1orf51	CSTF3	FBXO21	ITFG2	MLLT10	PDS5A	RFC1	SYT1	WDR72		BCL2L15	MICAL1	
ADNP	C1orf55	CTCF	FBXO3	JMY	MORN4	PFN2	RFC4	TAF2	WDSUB1		BIRC3	MOGAT3	
ADSS	C20orf72	CTHRC1	FBXO30	KCNMB2	MRFAP1L1	PGAP1	RFC5	TBCE	XRCC5		BTNL3	MUC17	
AES	C2orf68	DBT	FEM1A	KCNMB3	MRPL47	PGBD5	RFWD2	TBL1XR1	XRCC6BP1		C19orf33	MYO18A	
AGA	C3orf17	DCAF16	FTO	KCNS3	MRS2	PHF2	RHOT1	TCTEX1D2	YEATS2		C2CD4A	ORMDL2	
AGK	C3orf58	DCLRE1A	FUBP3	KCTD3	MSH6	PHF20L1	RIOK1	TCTN1	YPEL5		CAPN10	PCSK6	
AGL	C4orf27	DDX59	FXR1	KEAP1	MTBP	PHIP	RMI1	TDG	YWHAQ		CASP7	PFKFB3	
AHCTF1	C6orf162	DEAF1	FZD10	KHDC1	MTERFD1	PIAS2	RNF115	TDRD3	YY1		CCL15	PHLDA2	
ALDH9A1	C6orf52	DEGS1	FZD6	KHDRBS1	MTX2	PIAS3	RNF2	TEAD2	ZBED5		CCR2	PLEKHB2	
AMN1	C9orf78	DEK	FZD7	KIAA0907	MTX3	PIGC	RPRD2	TERF1	ZBTB22		CCRL1	PTAFR	
ANAPC10	C9orf9	DERA	GABARAPL1	KIAA1191	NAE1	PIGY	RRAGB	TESK2	ZBTB43		CD300LF	PTK6	
ANAPC7	CABYR	DHRS4L2	GABRE	KIAA1324L	NAP1L1	PIK3CA	RSBN1L	THAP10	ZC3H14		CD3D	PTP4A2	
ANGEL2	CBFA2T2	DHX36	GATAD2B	KIAA1328	NAPB	PITX2	RSRC2	THAP9	ZCCHC11		CLDN2	PTPRE	
ANKRD17	CBFB	DIRC2	GGPS1	KIAA1467	NARG2	PJA1	RYK	TIA1	ZCCHC8		CLMN	PTPRJ	
ANKRD46	CBR4	DMTF1	GINS3	KIAA1704	NCBP2	PKD2	SAV1	TIGD5	ZFP30		CPM	PVR	
AP3M2	CBX5	DNAJC19	GLI4	KIDINS220	NEK1	PLAC2	SBF2	TIGD7	ZFP37		CSF3	RASGRP4	
API5	CBY1	DNAJC27	GMNN	KIF11	NEK2	PLRG1	SCAND3	TIPRL	ZFP64		CSF3R	RASSF6	
ARHGAP19	CCDC102A	DNAL4	GNAS	KLF11	NGFRAP1	PMAIP1	SCARA3	TMEM194A	ZHX1		CTH	REPS2	
ARL2BP	CCDC115	DPY19L4	GNPDA2	KLHDC5	NIPSNAP1	POLE3	SCPEP1	TMEM209	ZMYM4		CTSE	RHBDL2	
ARL6IP6	CCDC117	DPY30	GOLGA7	KLHL12	NMNAT3	POT1	SEC24B	TMPO	ZNF134		CXCL1	RHOH	
ARV1	CCDC59	DR1	GPM6B	KLHL20	NPL	PPHLN1	SENP2	TOPORS	ZNF148		CXCL5	RNF213	
ASF1A	CCDC91	DRG1	GPNMB	KLHL24	NRF1	PPM1A	SENP5	TOR1AIP1	ZNF16		CXCL6	RORC	
ASH2L	CCNE2	DSC3	GPR19	KLHL7	NSMAF	PPP1CB	SENP7	TP63	ZNF184		DENND1C	RPS27L	
ATAD2	CD9	DSTYK	GRHL2	KNTC1	NUDCD2	PPP1R2	SEPHS1	TRAPPC6B	ZNF193		DMBT1	RTEL1	
ATP11B	CDC7	DUT	GTF2IP1	KTN1	NUDT21	PRPSAP2	SEPT3	TRIM23	ZNF207		ERGIC1	SH2D2A	
ATP6V1E2	CDCA4	DVL3	GTF3C3	LANCL1	NUF2	PSIP1	SERPINI1	TRIT1	ZNF236		FAM40B	SH3KBP1	
ATP6V1G1	CDK2AP1	DYM	GTPBP8	LANCL2	NUP133	PSMD10	SERTAD4	TRMT5	ZNF250		FLT4	SIL1	
AZIN1	CENPC1	EAPP	H2AFY2	LAPTM4B	NUP153	PTDSS1	SET	TROVE2	ZNF260		FPR2	SLAMF6	
BBS10	CENPF	EID2	HDAC2	LARP7	NUP54	PTK2	SFRP2	TSEN15	ZNF271		FRMD5	SLC19A3	
BBS2	CENPQ	EIF4A2	HDGFRP3	LEF1	NUPL2	PTPN13	SIAH1	TTC32	ZNF273		FUT6	SLC29A1	
BBS7	CEP63	EIF4ENIF1	HECA	LINC00909	OGFRL1	PUM2	SKA2	TTC35	ZNF280D		GALE	SLC4A4	
BBX	CEP68	ELF2	HELQ	LINC00938	OPA1	QKI	SLBP	TTC5	ZNF32		GALNT3	SLC5A1	
BCL2L13	CHD9	ELOVL5	HIVEP1	LINC01089	OSCP1	RAB2B	SLC39A6	TTC8	ZNF362		GIMAP8	SLC6A20	
BCL7A	CHEK2	ENAH	HLTF	LOC202781	OTUD6B	RABL2A	SLMO1	TUB	ZNF397		GRAP	SLCO4A1	
BEX2	CHODL	ENSA	HMGB2	LRCH3	OXR1	RABL2B	SMAD4	TUBA1A	ZNF436		HGD	SPNS2	
BEX4	CHRAC1	ENY2	HMGXB4	LRRC40	PABPC4L	RAD21	SMAD5	TXNL1	ZNF512		HNF4A	TCF7	
BOLA1	CIDEB	EP300	HNRNPA0	LYPLAL1	PAIP2	RAD23B	SMARCAD1	UBE2E3	ZNF532		ICA1	TM4SF5	
BRD7	CKAP2	ERMAP	HNRNPA1L2	MAGEF1	PAPSS1	RAD51C	SMARCD1	UBE2V2	ZNF559		IL1B	TMEM106A	
BRMS1L	CLDND1	ETNK2	HNRNPD	MANEA	PARD6G	RALGAPA1	SMC3	UBR5	ZNF606		IL4R	TNFSF15	
BTF3L4	CLIP4	EXOSC9	HNRNPH3	MATR3	PARL	RAP2A	SMC4	UBR7	ZNF644		IL8	TRIM36	

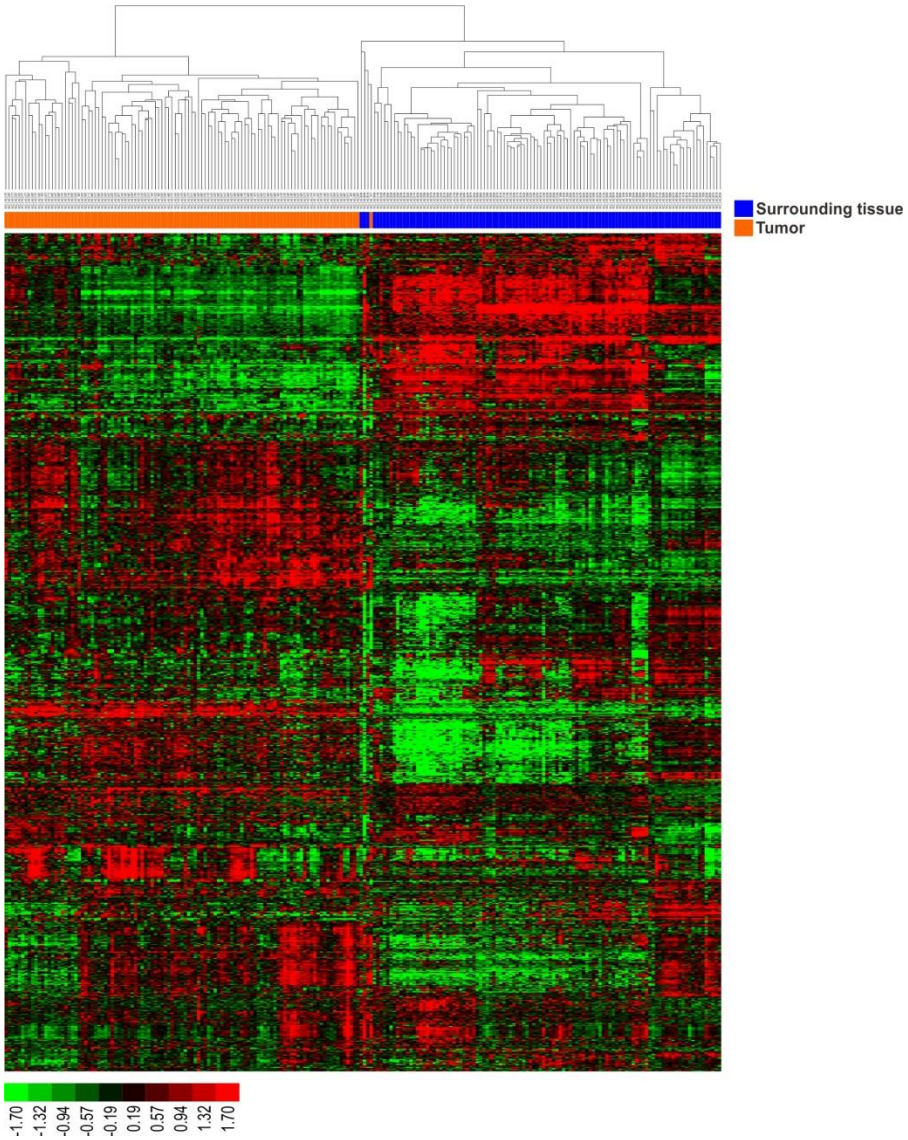
Supplementary Table 8. CyTOF staining panel design with 17 metal-conjugated antibodies.

Label	Target	Vendor_ID	Cat. No	Clone
142Nd	CD19	DVS-Fluidigm	3142001B	HIB19
145Nd	CD4	DVS-Fluidigm	3145001B	RPA-T4
146Nd	CD8a	DVS-Fluidigm	3146001B	RPA-T8
147Sm	CD278=ICOS	BioLegend	313502	C398.4A
150Nd	CD61	DVS-Fluidigm	3150001B	VI-PL2
151Eu	CD123	DVS-Fluidigm	3151001B	6H6
154Sm	CD45	DVS-Fluidigm	3154001B	HI30
160Gd	CD279=PD1	Miltenyi	130-096-168	PD1.3.1.3
161Dy	CD152=CTLA-4	DVS-Fluidigm	3161004B	14D3
162Dy	CD69	DVS-Fluidigm	3162001B	FN50
163Dy	CD56	DVS-Fluidigm	3163007B	NCAM16.2
169Tm	CD25	DVS-Fluidigm	3169003B	2A3
170Er	CD3	DVS-Fluidigm	3170001B	UCHT1
172Yb	CD274=PD-L1	BioLegend	329702	29E.2A3
173Yb	CD14	BioLegend	325602	HCD14
174Yb	HLA-DR	DVS-Fluidigm	3174001B	L243
176Yb	CD127=IL-7Ra	DVS-Fluidigm	3176004B	A019D5
191Ir	Intercalation	DVS-Fluidigm		
Viability	Cisplatin	DVS-Fluidigm		



Supplementary Figure 1. Schematic diagram of the strategy for developing a risk assessment score for recurrence (RAS) and uncovering the underlying biology associated with RAS in esophageal squamous cell carcinoma.

ESCC denotes esophageal squamous cell carcinoma, sncRNAs small non-coding RNAs, quantitative RT-PCR quantitative real time-polymerase chain reaction, and TCGA The Cancer Genome Atlas.

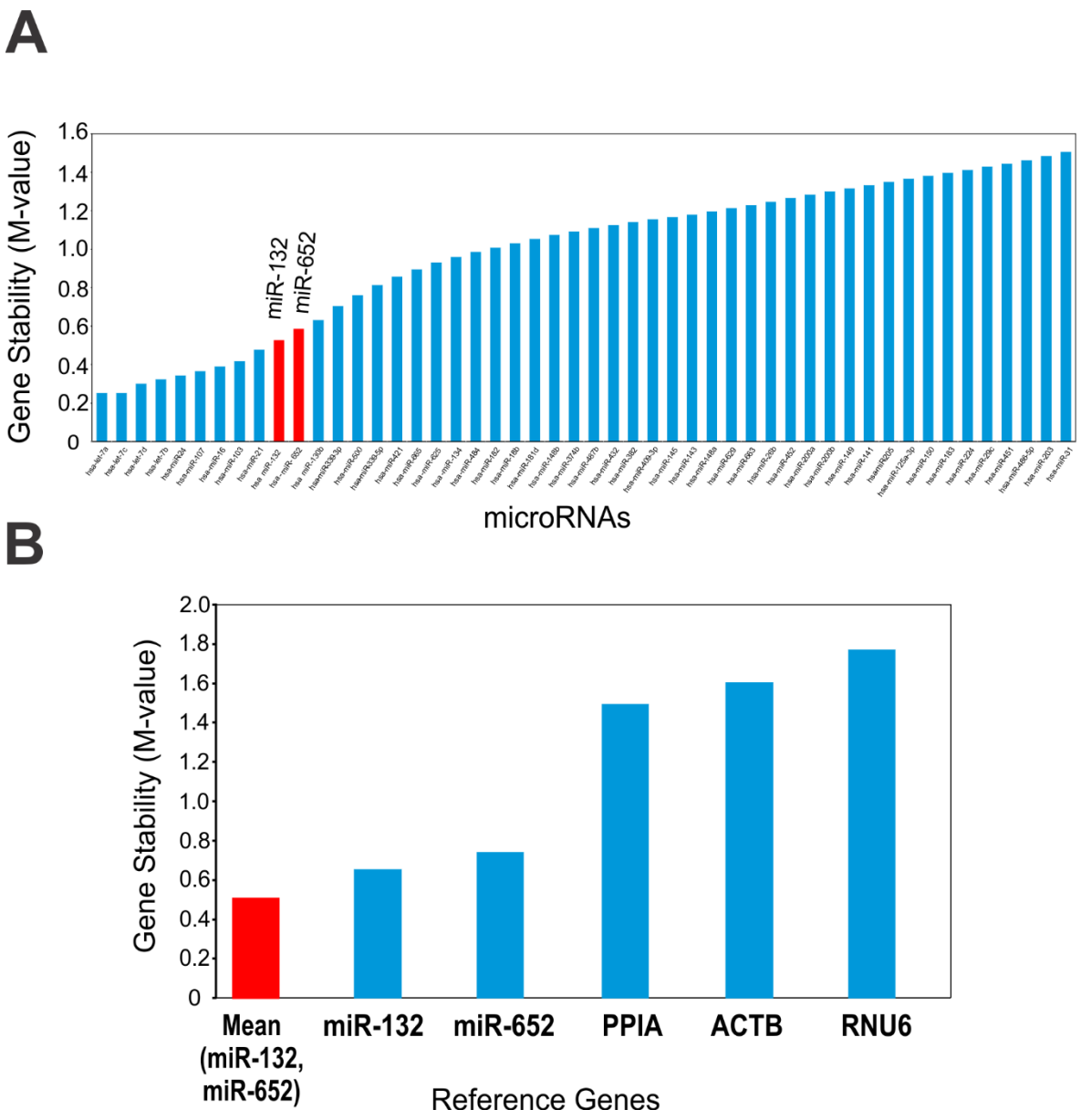


Supplementary Figure 2. Unsupervised hierarchical clustering analysis of sncRNA expression data from patients with ESCC (discovery set).

To evaluate the quality of ESCC tumor and adjacent normal tissues, unsupervised hierarchical clustering was applied to sncRNA expression data from 108 ESCC in the discovery set. Hierarchical clustering analysis was performed with selected sncRNAs (901 sncRNAs features) having an expression level that had at least 2-fold difference relative to median value across tissues in at least 20 tissues. The clustering clearly separated tumor (orange bar) and normal (blue bar) tissue samples. Only one sample was misclassified.

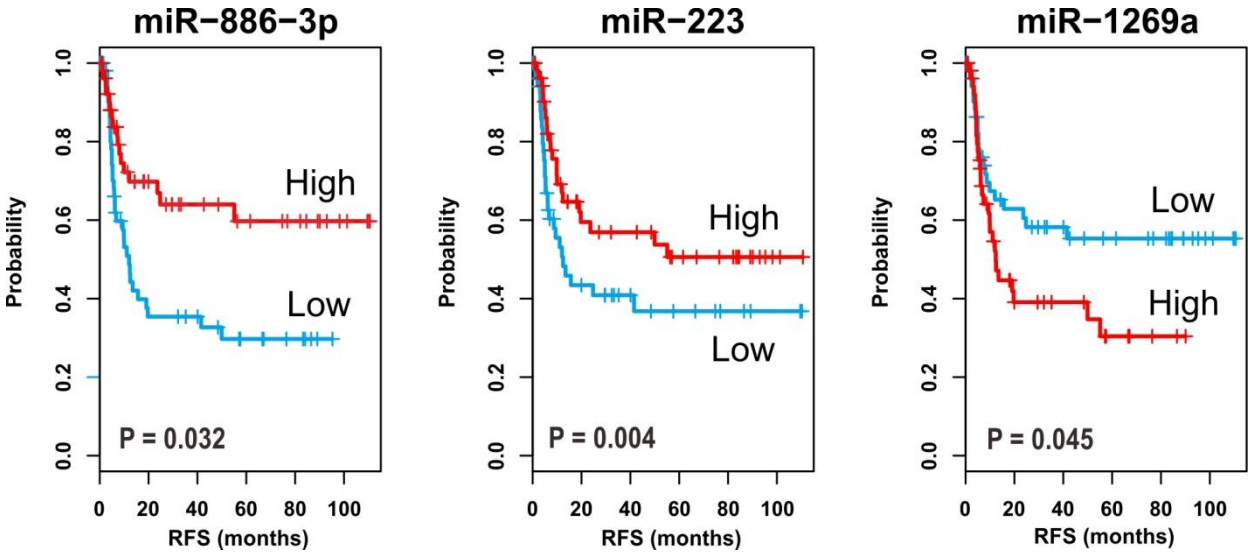
The data are presented in a matrix format in which rows represent individual sncRNA and columns represent individual tissue. Each cell in the matrix represents the expression level of an sncRNA feature in an individual tissue. The red and green colors in the cells reflect relative high and low expression levels, respectively, as indicated in the scale bar (log2 transformed scale).

sncRNA denotes small non-coding RNA.



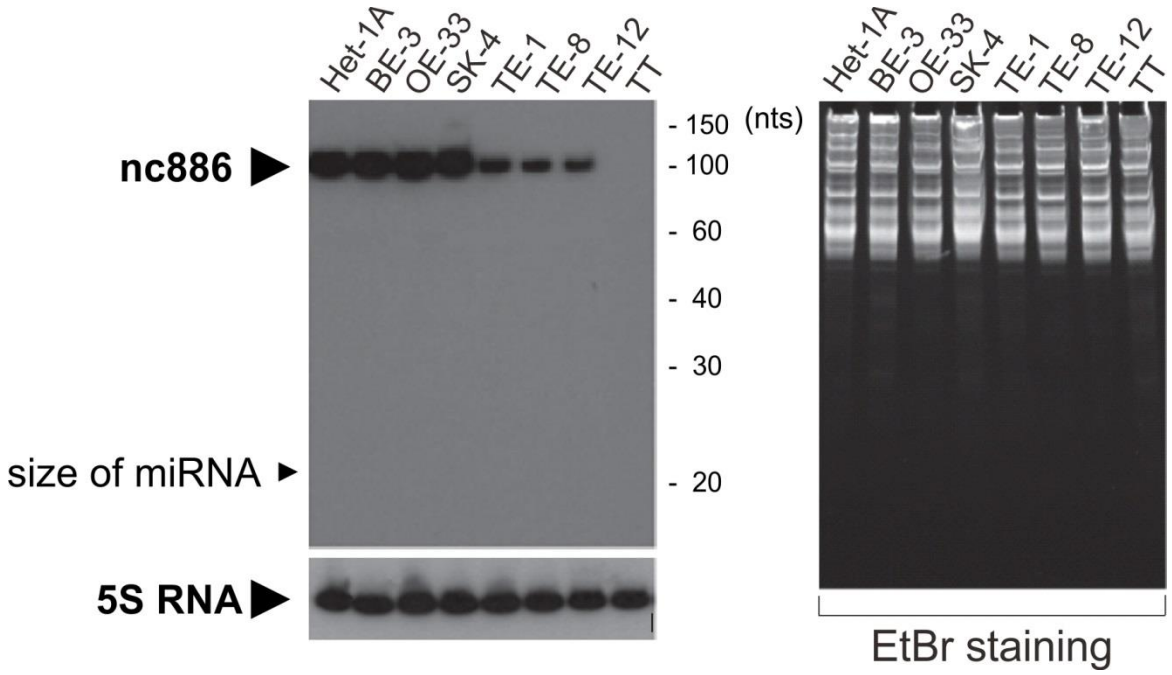
Supplementary Figure 3. geNorm expression stability plots for finding reference genes for qRT-PCR data normalization.

(a) Expression stability plot from sncRNA microarray data of tumor and normal tissues in the discovery set. To select the most stable microRNAs (miRNAs) for normalization in miRNA qRT-PCR data, we used geNorm to calculate the gene stability value (M value), which is the average variation of each gene with all other control genes. Genes with lower M values are considered more stable. The two miRNAs selected as reference among the most stable 50 miRNAs are highlighted in red. **(b)** Expression stability plot from qRT-PCR of reference genes. To check the stability of reference genes for qRT-PCR experiments, the M value was calculated for 3 generally used reference genes (PPIA, ACTB and RNU6) as well as two selected reference miRNAs (miR-132 and miR-652). PPIA denotes cyclophilin A, ACTB β-actin, RNU6 small nuclear RNA 6, and qRT-PCR quantitative real time-polymerase chain reaction.



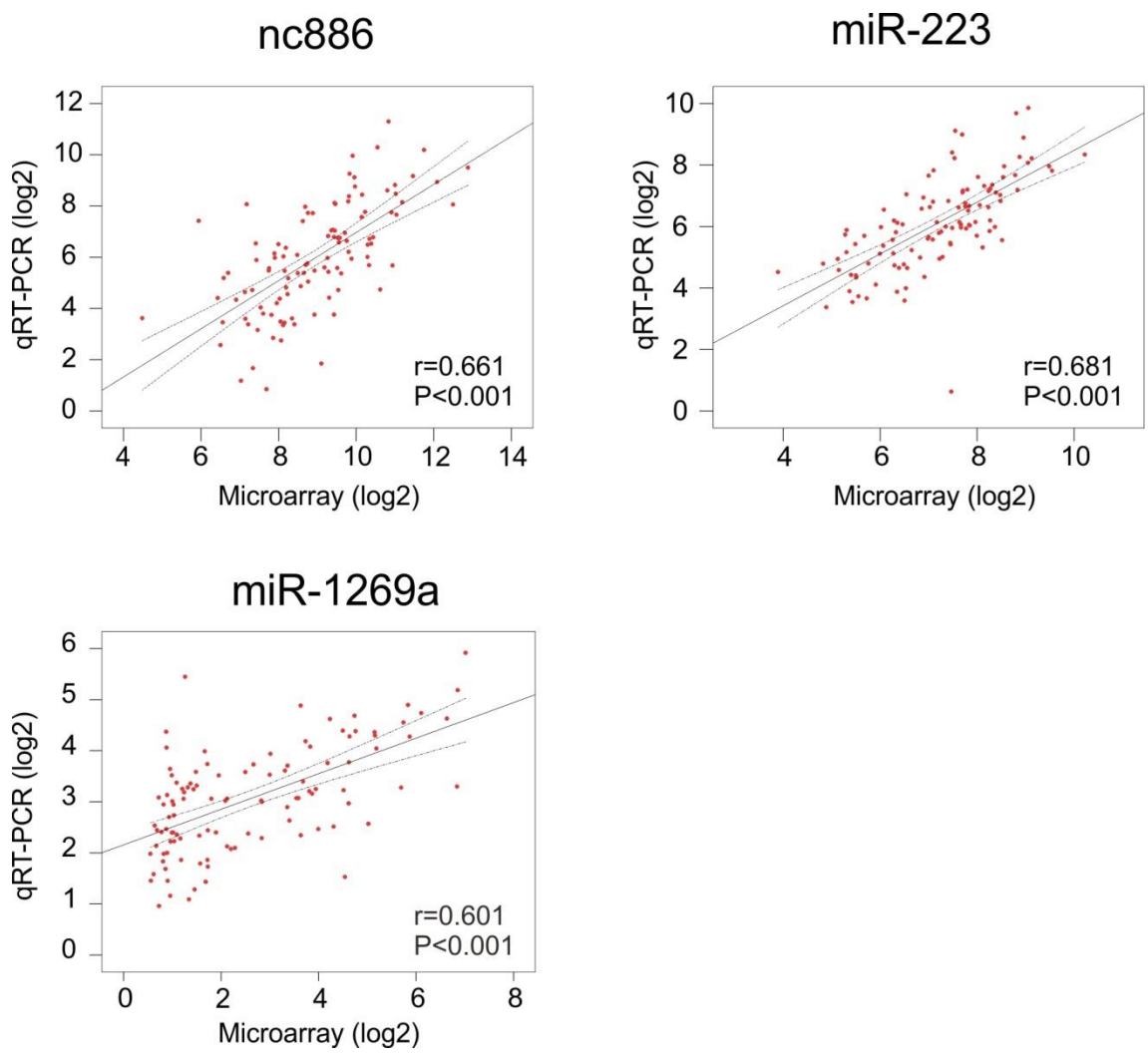
Supplementary Figure 4. Prognostic small non-coding RNAs in esophageal squamous cell carcinoma.

Kaplan-Meier plots of recurrence-free survival of two subgroups in patients with esophageal squamous cell carcinoma when patients were dichotomized by median expression values of each sncRNA. The + symbols in the panel indicate censored data. RFS denotes recurrence-free survival.



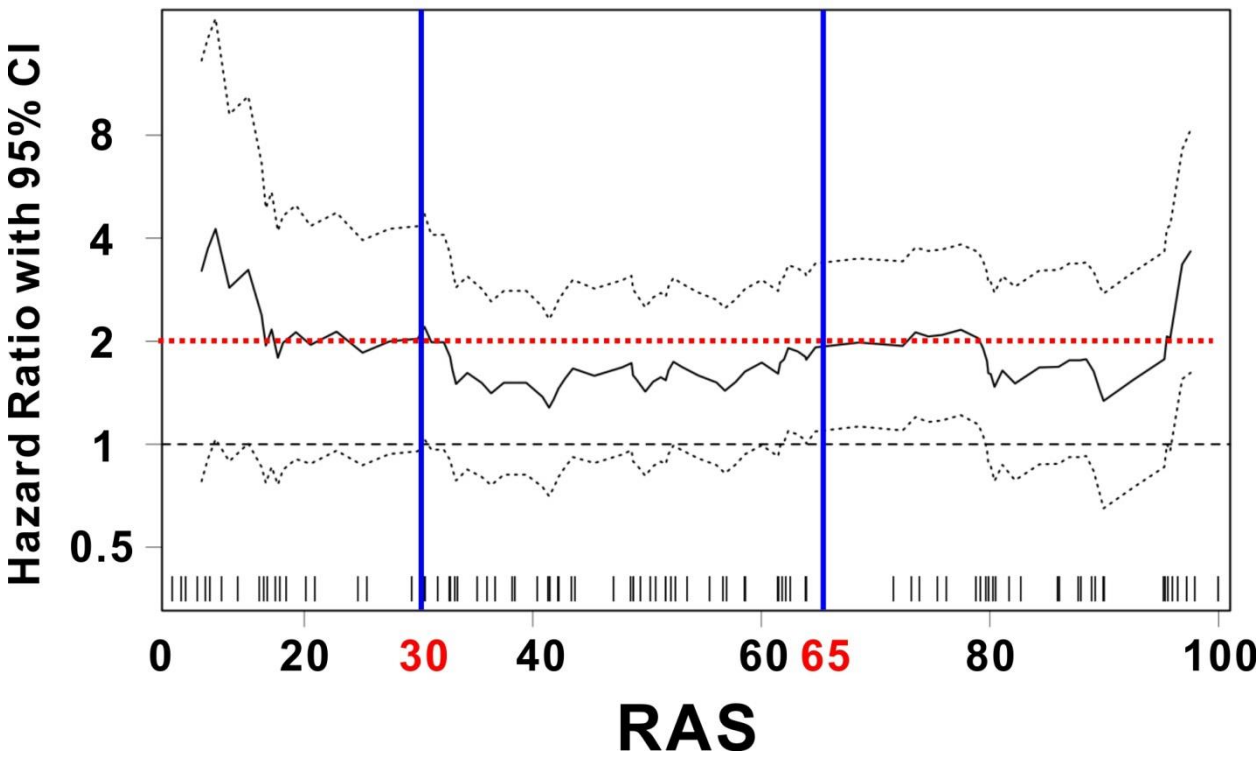
Supplementary Figure 5. Northern hybridization of miR-886.

Northern hybridization was carried out with RNA from ESCC cell lines (TE-1, TE-8, TE-12, and TT), immortalized esophagus squamous cell line (Het-1A), esophageal metaplastic cell line (BE-3) and esophageal adenocarcinoma cell lines (OE-33, and SK-4). EtBr staining is shown for equal loading. EtBr denotes ethidium bromide.



Supplementary Figure 6. Correlation between sncRNA expression data generated from microarray and qRT-PCR experiments.

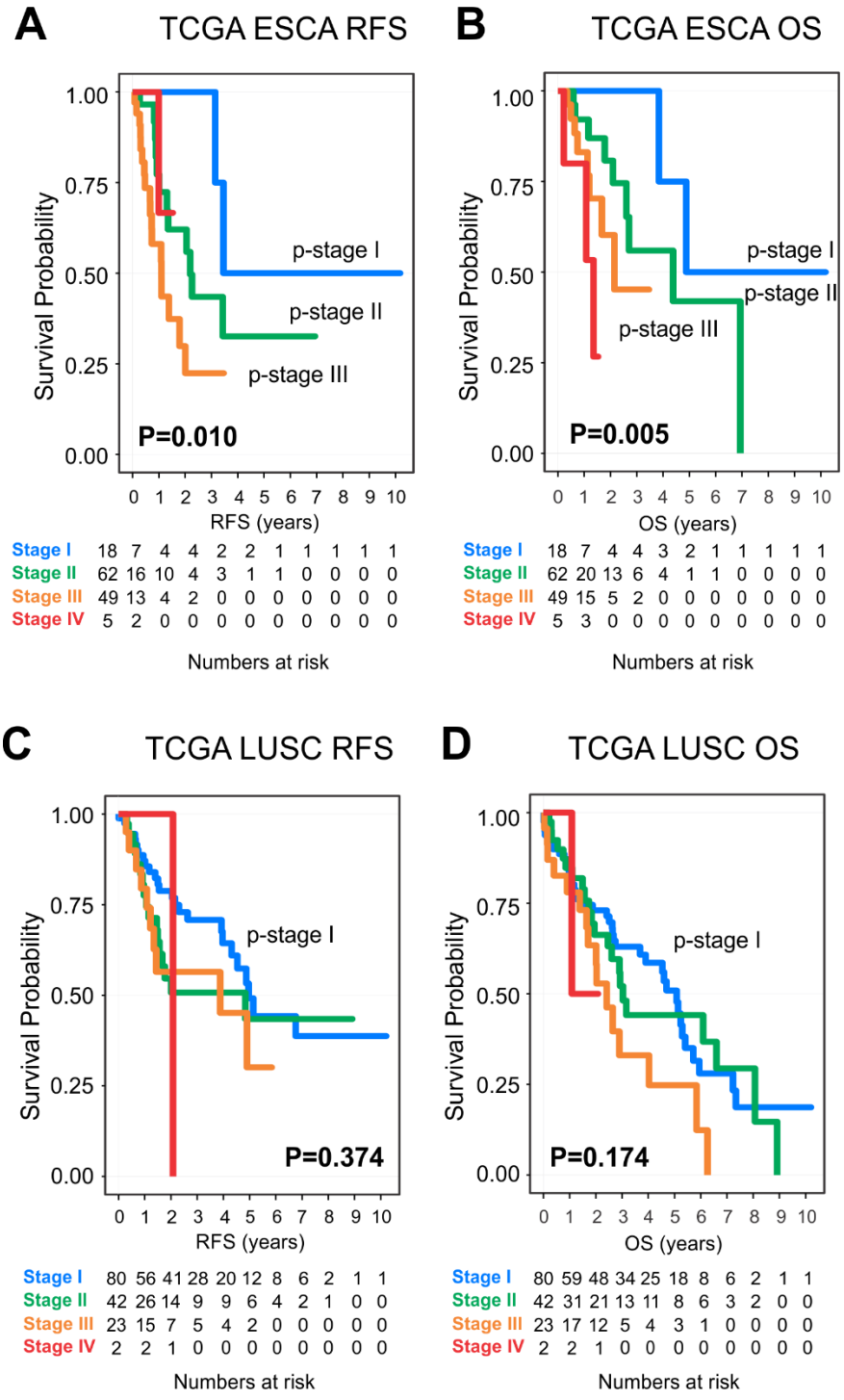
The significance of the correlation between sncRNA expression data generated from microarray and that of qRT-PCR experiments was estimated with the Spearman correlation method. qRT-PCR denotes quantitative real time-polymerase chain reaction.



Supplementary Figure 7. Hazard ratio plot to optimize the cutoff of RAS.

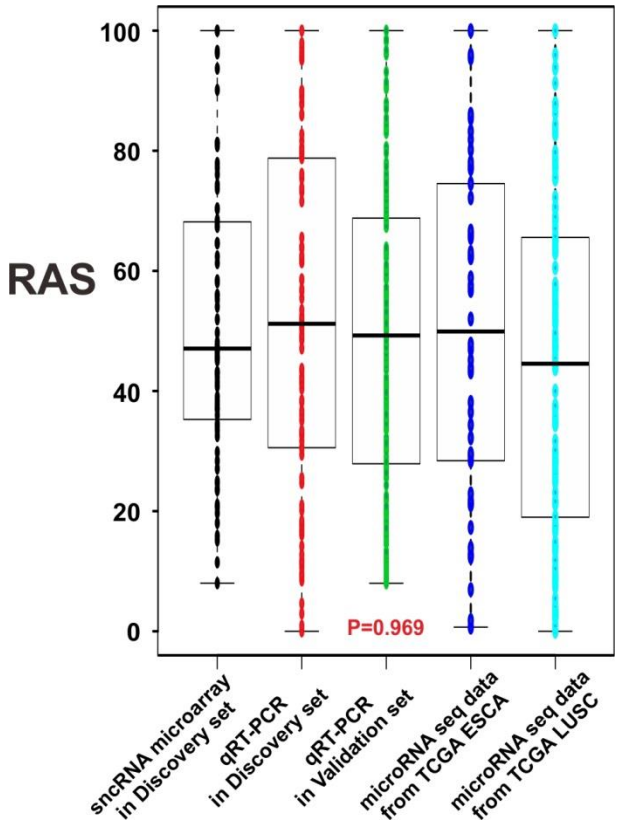
The optimal cutoff depends on the method that was chosen for cutoff optimization. Here, we optimized the cutoff by considering the significant alteration of hazard ratios (HR) and the largest interval of RAS.

The optimal cutoff is marked by a vertical line, and the horizontal dot line in red indicates HR 2.0. The HR with 95% confidential interval is shown in dependence of possible cutoff points. The distribution of the RAS is shown as rug plot at the bottom of the figure. For each cutoff point, a Cox analysis of RAS and the recurrence-free survival variable is executed. The fits are done using the functions coxph and survfit from the R package survival.



Supplementary Figure 8. Survival of TCGA cohorts based on pathologic staging, to check the quality of clinical data.

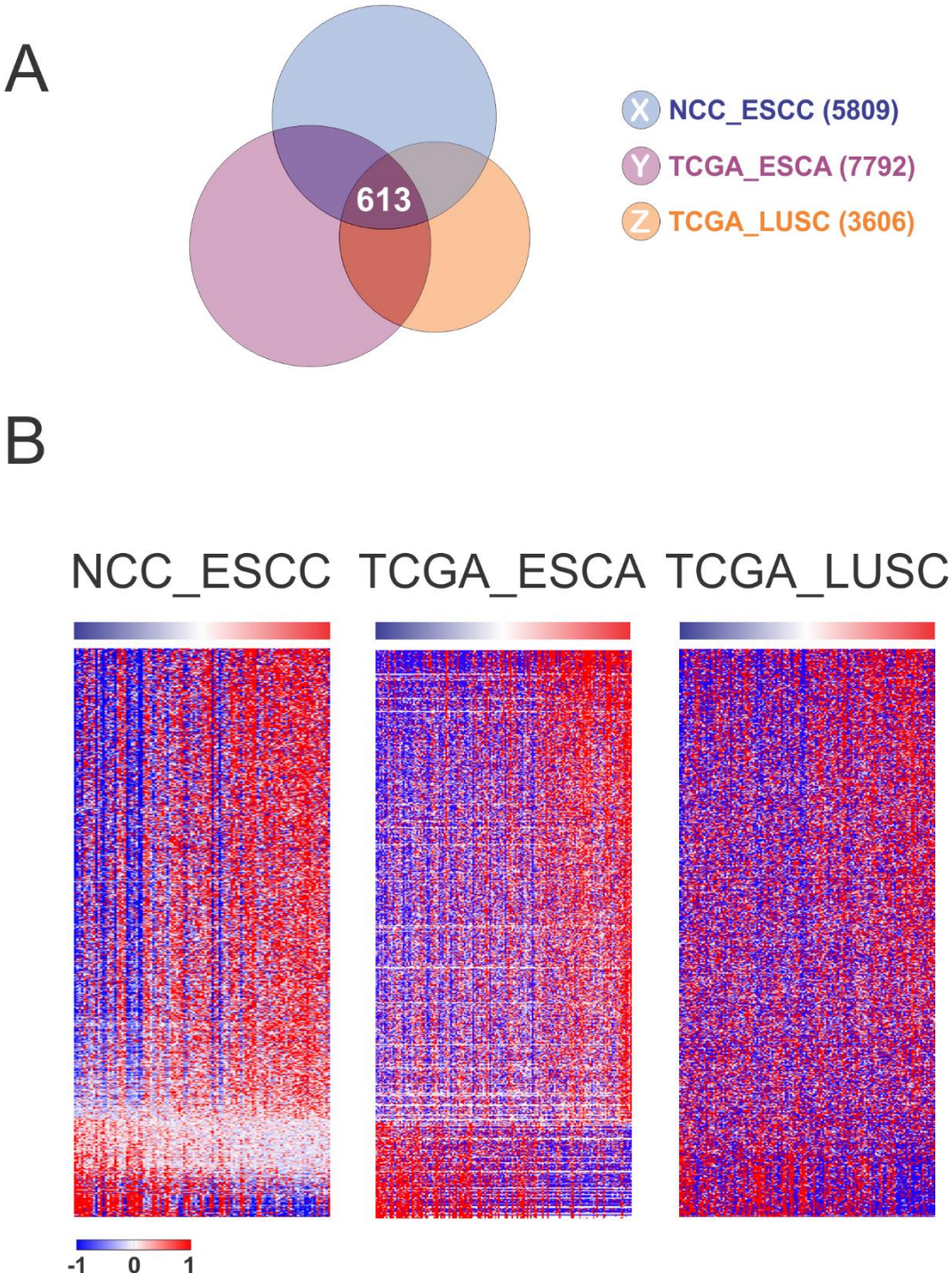
(a,b) Discrete RFS and OS curves based on pathologic staging of ESCA. **(c,d)** Tendency of different RFS and OS in pathologic staging of LUSC.
ESCA denotes esophageal cancer, LUSC lung squamous cell carcinoma, OS overall survival, RFS recurrence-free survival, and TCGA The Cancer Genome Atlas.



Supplementary Figure 9. Comparison of the range of risk assessment score for recurrence (RAS) based on various platforms.

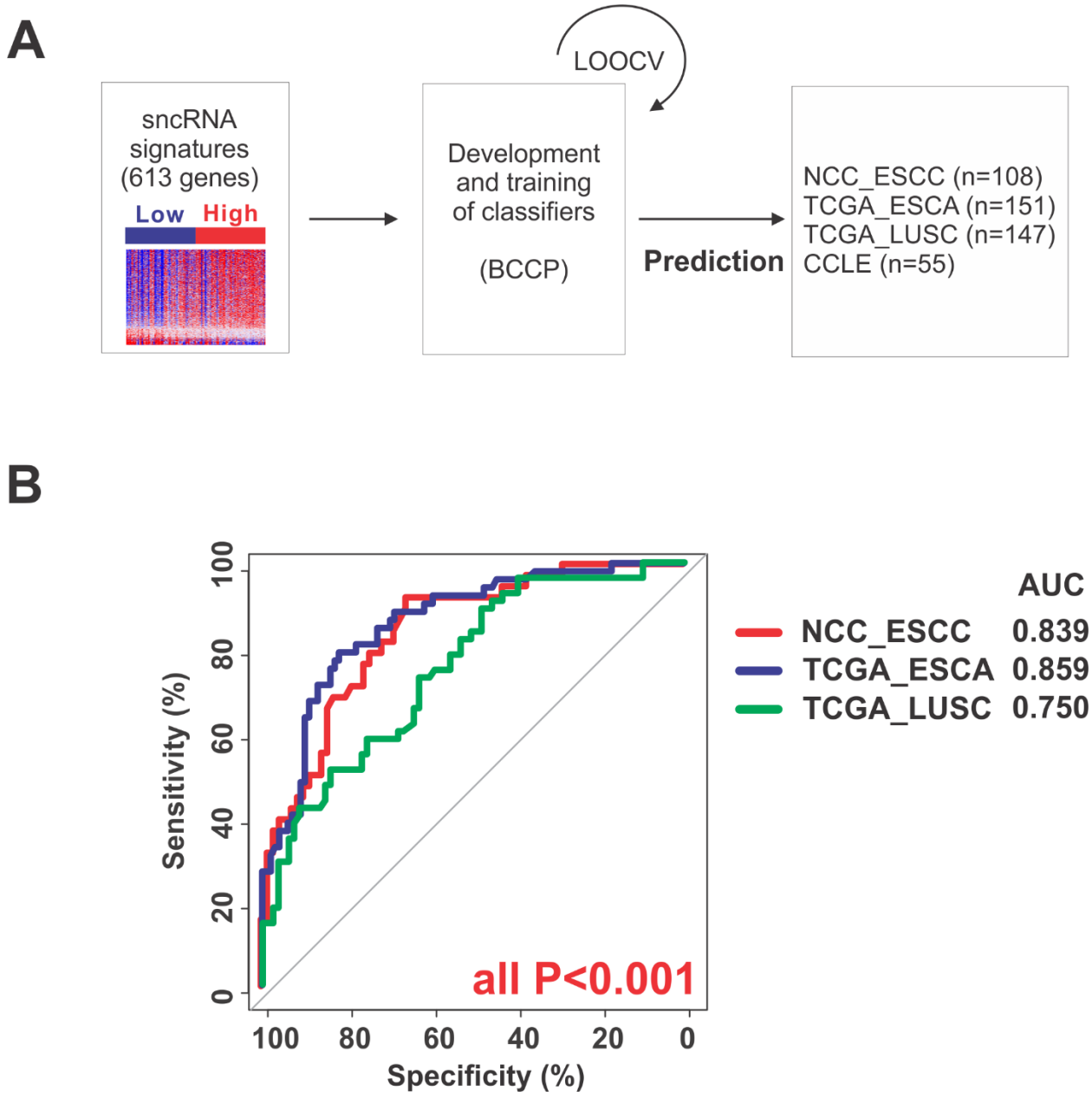
To prove that RAS from various platforms can be applied to multiple cohorts, we compared the mean value in each platform with the t-test (P=0.969) and applied the homogeneity test and Levine's test to prove equal variance (P=0.644).

RAS denotes risk assessment score for recurrence.



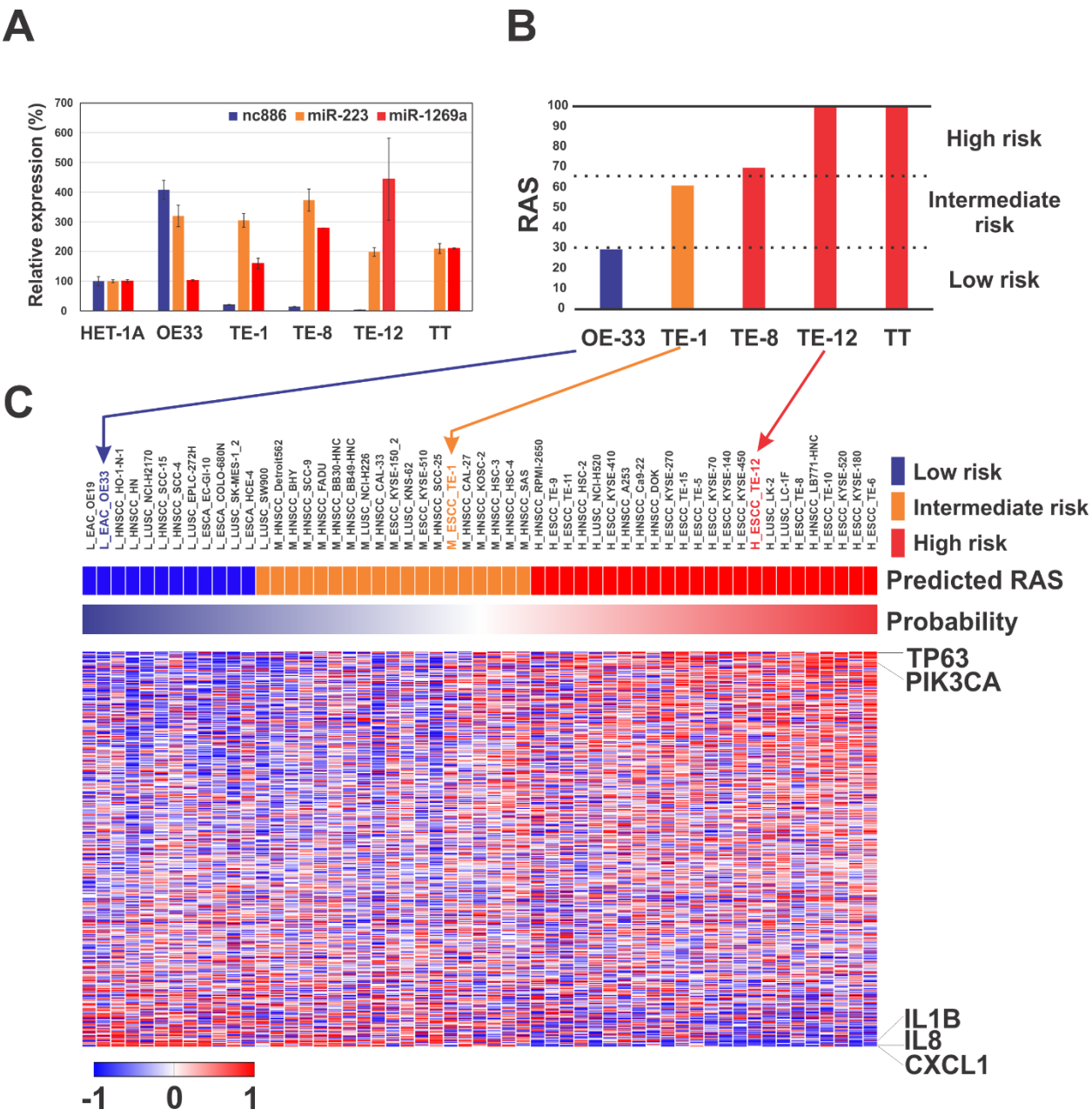
Supplementary Figure 10. RAS-correlated mRNA signature from human esophageal cancer and lung squamous cell carcinoma.

(a) Venn diagram of genes whose expression is significantly correlated with RAS (P<0.01). The 613 genes are commonly overlapped in three cohorts - discovery cohort (NCC_ESCC), esophageal cancer (TCGA_ESCA), and lung squamous cell carcinoma (TCGA_LUSC) from TCGA. **(b)** Expression of selected genes shared among three patient groups. Colored bars at the top of the heat map represent samples, as indicated.



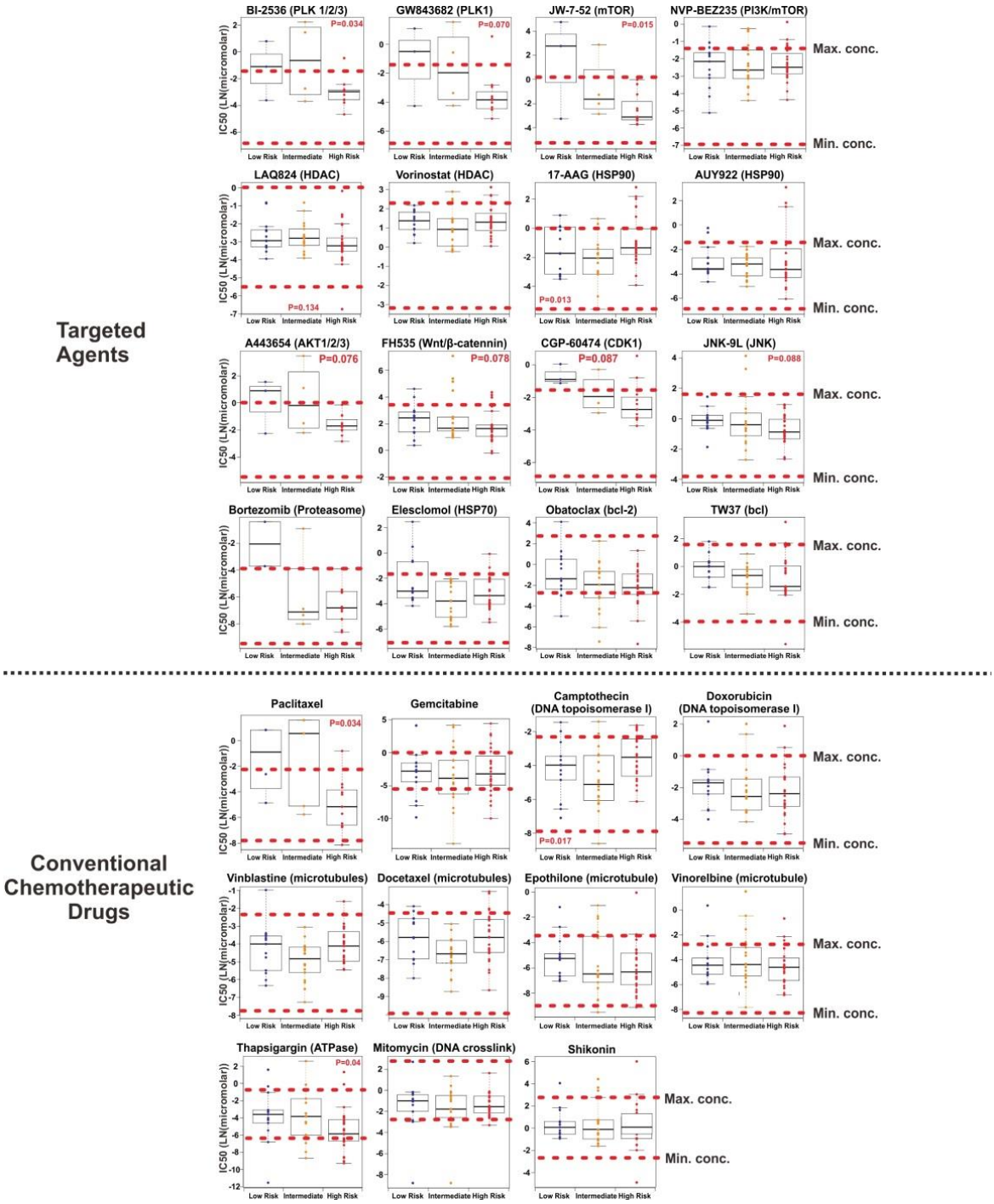
Supplementary Figure 11. Validation of RAS-correlated mRNA signature through multiple cohorts.

(a) Schematic overview of the strategy used for constructing prediction models and evaluating predicted outcomes according to gene expression signatures. **(b)** Receiver operating characteristic (ROC) curve between RAS and the probability of high-risk group predicted by mRNA signature.
BCCP denotes Bayesian compound covariate predictor, LOOCV leave-one-out cross validation.



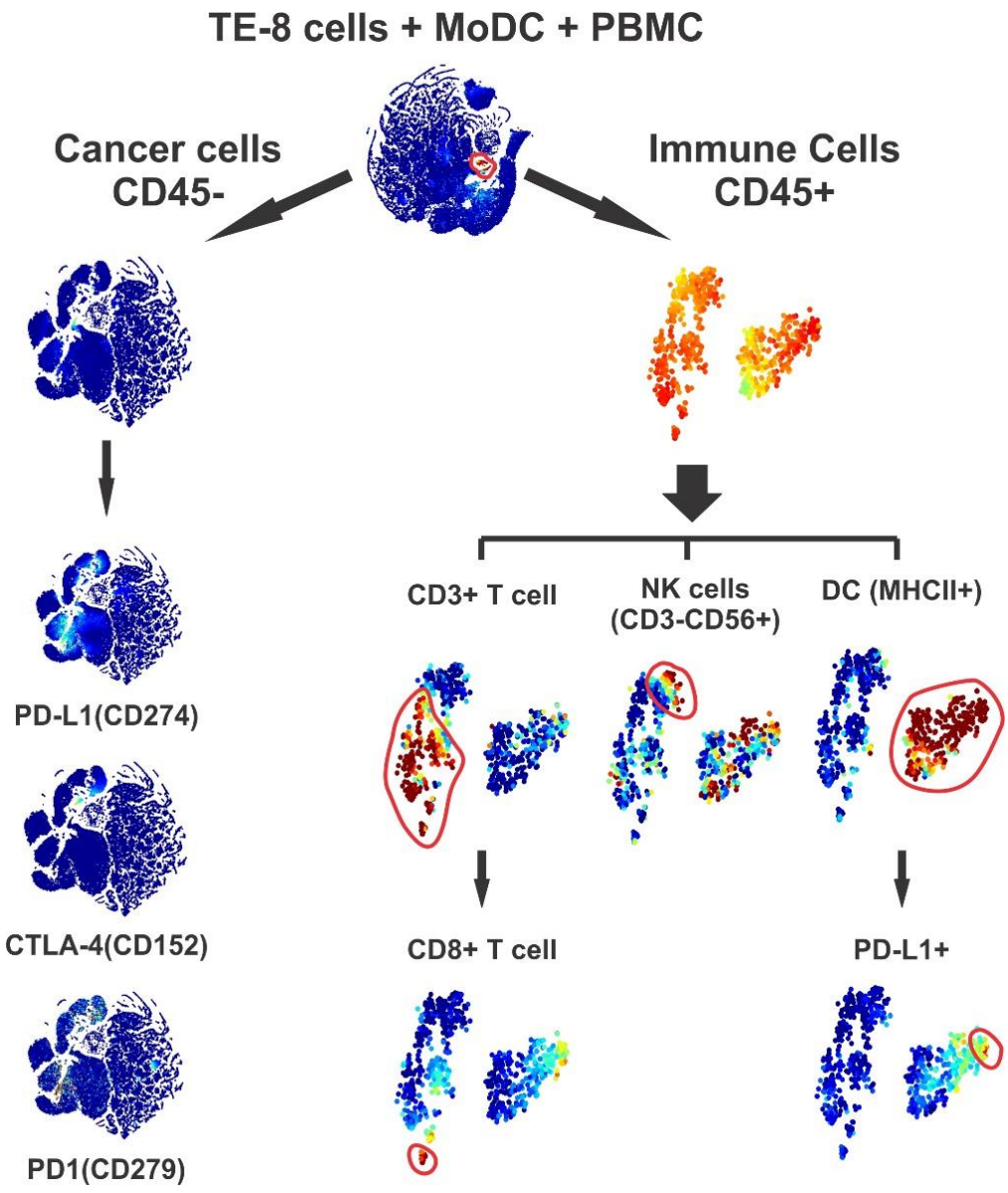
Supplementary Figure 12. Application of RAS-correlated mRNA signature to 55 cell lines of esophageal cancer, lung squamous cell carcinoma, and head and neck squamous cell carcinoma from genomics of drug sensitivity in cancer database.

(a) Relative expression of three sncRNAs in esophageal epithelial and cancer cell lines based on qRT-PCR data. **(b)** Each cell line was classified into three risk groups, according to our RAS scoring system. OE-33 was in the low-risk group, TE-1 in the intermediate, and TE-8, TE-12 and TT were in the high-risk groups. **(c)** The categorization based on RAS-correlated mRNA signatures showed the same result as the categorization based on qRT-PCR data of three sncRNAs.



Supplementary Figure 13. Drug rearrangement for risk-stratified esophageal cancer by leveraging the genomics of drug sensitivity in cancer database.

Fifty-five cell lines from esophageal cancer, lung squamous cell, and head and neck squamous cell carcinoma were selected and classified into three risk groups by using a prediction model derived from RAS-correlated 613 mRNA signatures. The IC_{50} values of about 15% of the drugs were within the acceptable range for cell lines with high risk.



Supplementary Figure 14. Immunophenotyping of high-risk TE-8 ESCC cell line after incubation with moDCs and PBMCs.

Manual gating strategy to characterize cancer cells and major immune cell types with CyTOF. Data were obtained from TE-8 cell lines incubated with moDCs and PBMCs. The proportion of immune cells was markedly decreased in tumor microenvironment that included the TE-8 cell line. There was no alteration of PD-L1 expression in both TE-8 cells and dendritic cells after incubation with moDCs and PBMCs. The viSNE plots were generated with CYT, which is an interactive visualization tool designed for the analysis of high-dimensional mass or flow cytometry data.

CyTOF denotes Time-Of-Flight Mass Cytometry, moDCs monocyte-derived dendritic cells, and PBMCs peripheral blood mononuclear cells.