PNAS
www.pnas.org

Supplementary Information for

Addressing Cellular Heterogeneity in Tumor and Circulation for Refined Prognostication

Su Bin Lim, Trifanny Yeo, Wen Di Lee, Ali Asgar Bhagat, Swee Jin Tan, Daniel S. W. Tan, Wan-Teck Lim, Chwee Teck Lim

Corresponding author: Chwee Teck Lim
Email: ctlim@nus.edu.sg

This PDF file includes:

SI Materials and Methods
Figures S1 to S22
Tables S1 to S8
SI References

www.pnas.org/cgi/doi/10.1073/pnas.1907904116

SI Materials and Methods

Cell culture

Both metastatic (H1650, H2009, and H1792) and non-metastatic (A549, H1975, H2228 and H2291) lung adenocarcinoma cancer cell lines were kept at 37°C conditions with 5% CO2, and allowed to achieve 80% confluency in T25 flasks before passaging. Cell lines were cultured in RPMI-1640 medium (Gibco™, ThermoFisher) except for H2009, which was maintained using DMEM:F12 basal medium (Gibco™, ThermoFisher). 0.25% Trypsin was used to detach the cells and centrifuge speed was set at 200 x g for 5 minutes during sub-culturing. Cells were resuspended in 1x PBS and diluted to achieve 80,000 cells/mL.

Primer design and multiplex gene panel

Primer pairs for preamplification and quantification were previously validated, except for few genes which were designed using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast), and *in-silico* validated with UCSC In-Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr). Forward and reverse sequences, reference, and length of each primer (bp) are listed in SI Appendix, Table S5. All primers were purchased from Integrated DNA Technologies unmodified with standard desalting. PCR-based multiplex preamplification protocol was developed using A549, MCF7, and MDA-MB-231 cancer cell lines. Validation of PCR performance and primer specificity of the multiplex PCR was done through melting curve analysis and/or gel electrophoresis. Amplified PCR products were run on a 2% agarose gel in 1x TAE buffer (all Life Technologies) stained with SYBR Safe (Life Technologies). The size of each amplicons was visualized and confirmed with GeneRuler 100bp DNA Ladder (Thermo Scientific).

ROC, AUC, and survival analyses

The area under the receiver operating characteristic (ROC) curve (AUC), sensitivity and specificity based on the best threshold were computed using the *pROC* package (1) in R. Univariate and multivariate Cox regression survival analyses were performed using the *survival* package in R. Hazard ratios (HRs) were adjusted for available clinical features in TCGA dataset (LUAD), including age at initial pathologic diagnosis, smoking status, AJCC pathologic stage, and gender.

Batch-effect correction and PCA

Batch-effect of CTC qPCR data was adjusted using the *ComBat* function from the *sva* package in R (2). PCA was performed using the *prcomp* function from the default *stats* package. The first two PCs were visualized with the *ggbiplot* package (https://github.com/vqv/ggbiplot).

Supplementary Figures

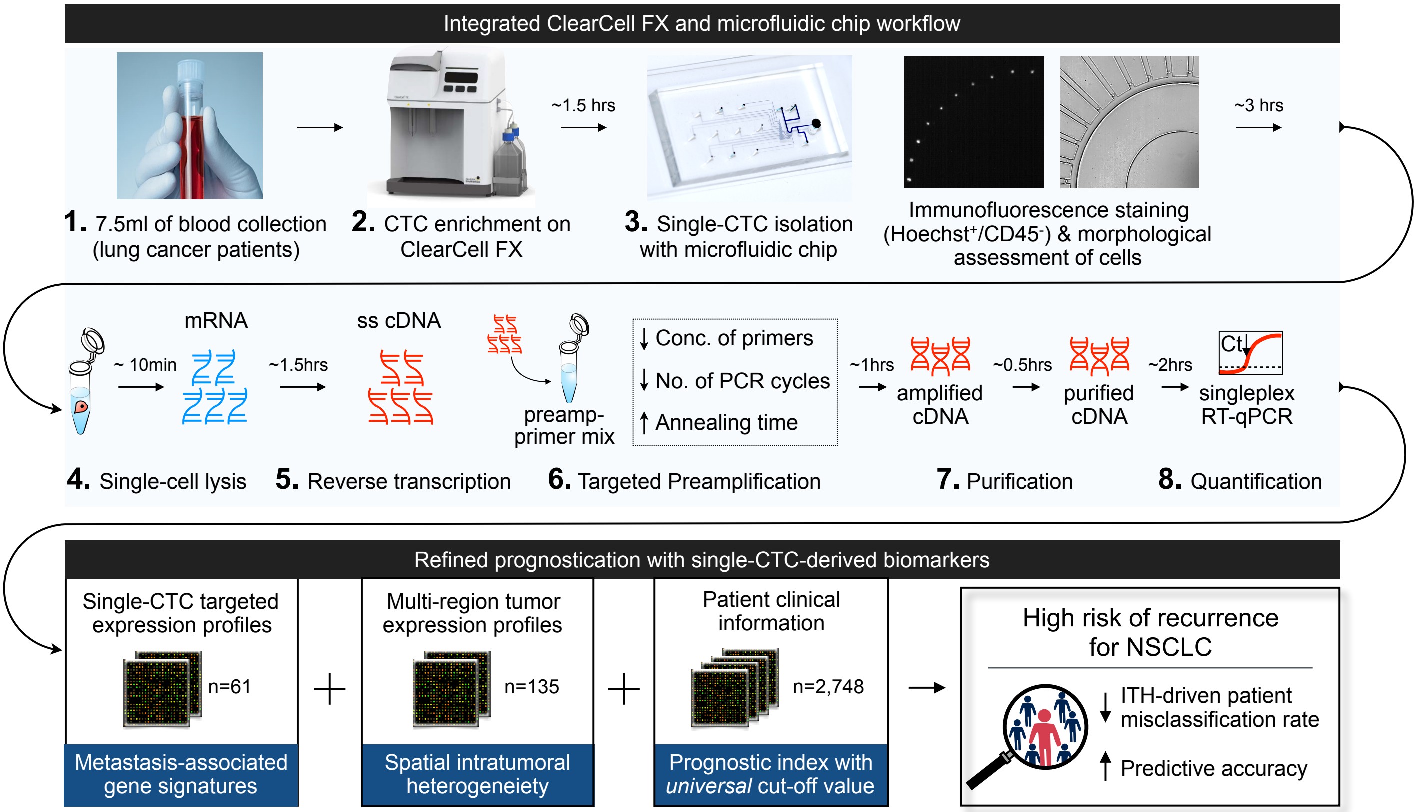


Fig. S1. Schematic of integrated workflow and study design. Top: integrated ClearCell FX and microfluidic chip workflow for single-cell analysis of CTCs. Bottom: big data approaches for refining prognostication with single-CTC-derived gene signatures and spatial intratumoral heterogeneity (ITH).

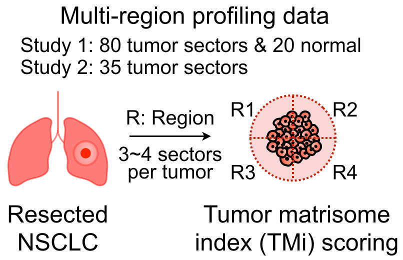
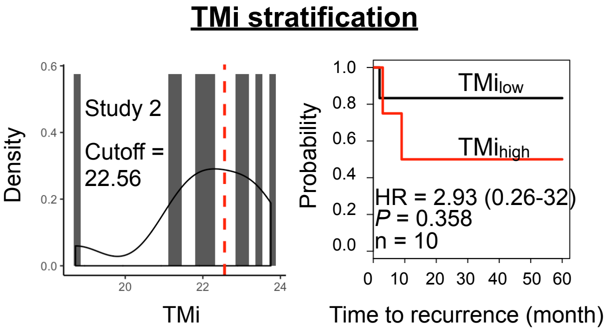


Fig. S2. Schematic outline of multi-region profiling data analysis.



Group	n	Events (Recurrence)	median	0.95 CL (lower)	0.95 CL (upper)
TMi$_{low}$	6	1	NA	NA	NA
TMi$_{high}$	4	2	9	3	NA

Fig. S3. Patient-level prognostic accuracy of TMi in Study 2. Gaussian kernel density distribution (left) and Kaplan-Meier survival curves (right) are drawn. The optimal cutoff value was used for patient stratification. Hazard ratio (HR), log-rank *P*-value, and the number of patients successfully stratified (n) are stated. The resulting summary statistics is shown below.

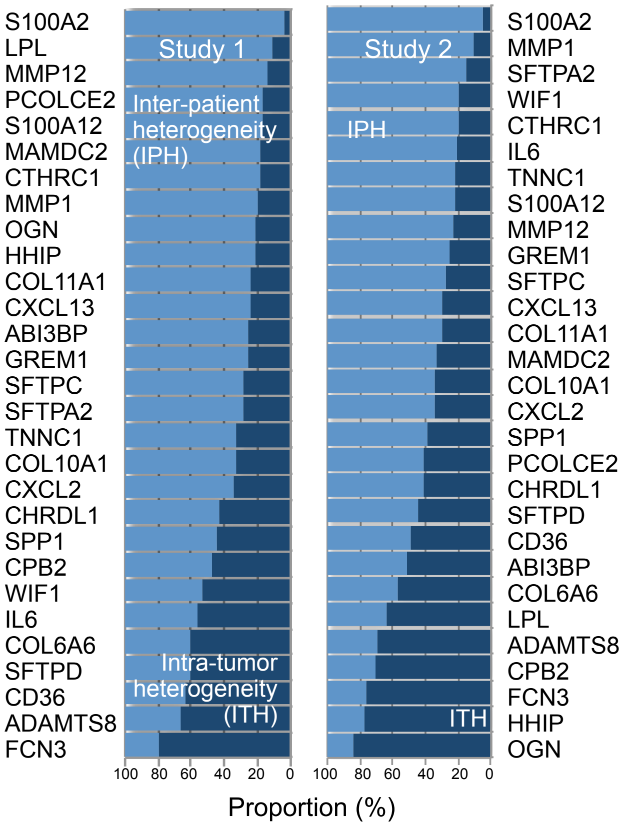


Fig. S4. ITH and IPH of each matrisome gene. Genes are ordered by increasing ITH.

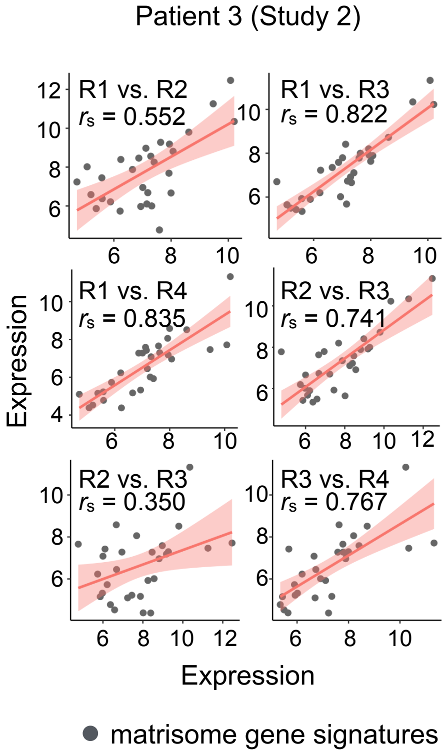


Fig. S5. All pairwise matrisome gene expression correlations for Patient 3 from Study 2. Spearman's rank correlation rho (r_s) values and linear regression lines with 95% CI are shown.

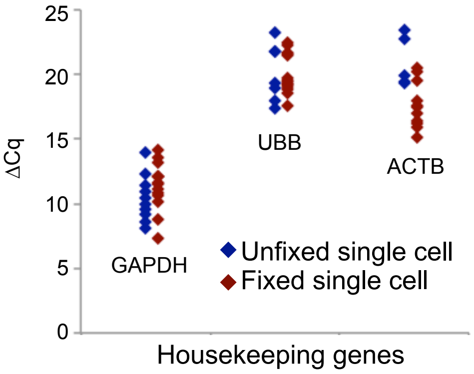


Fig. S6. Impact of 1% PFA on single-cell mRNA expression. Cq values of three housekeeping genes are shown for unfixed and fixed single A549 cells. The putative effects of 1% PFA fixation on mRNA expression was assessed with three housekeeping genes (GAPDH, UBB, and ACTB). A549 lung cancer cells were subjected to the same optimized workflow – from microfluidic enrichment of single cells to preamp-purification-quantification of the extracted RNAs. Three independent experiments were performed in duplicates.

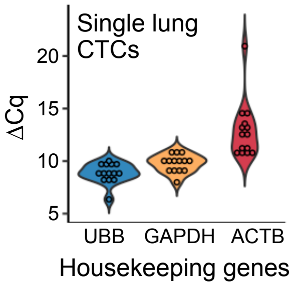


Fig. S7. Three housekeeping gene expression from 14 single patient-derived CTCs.

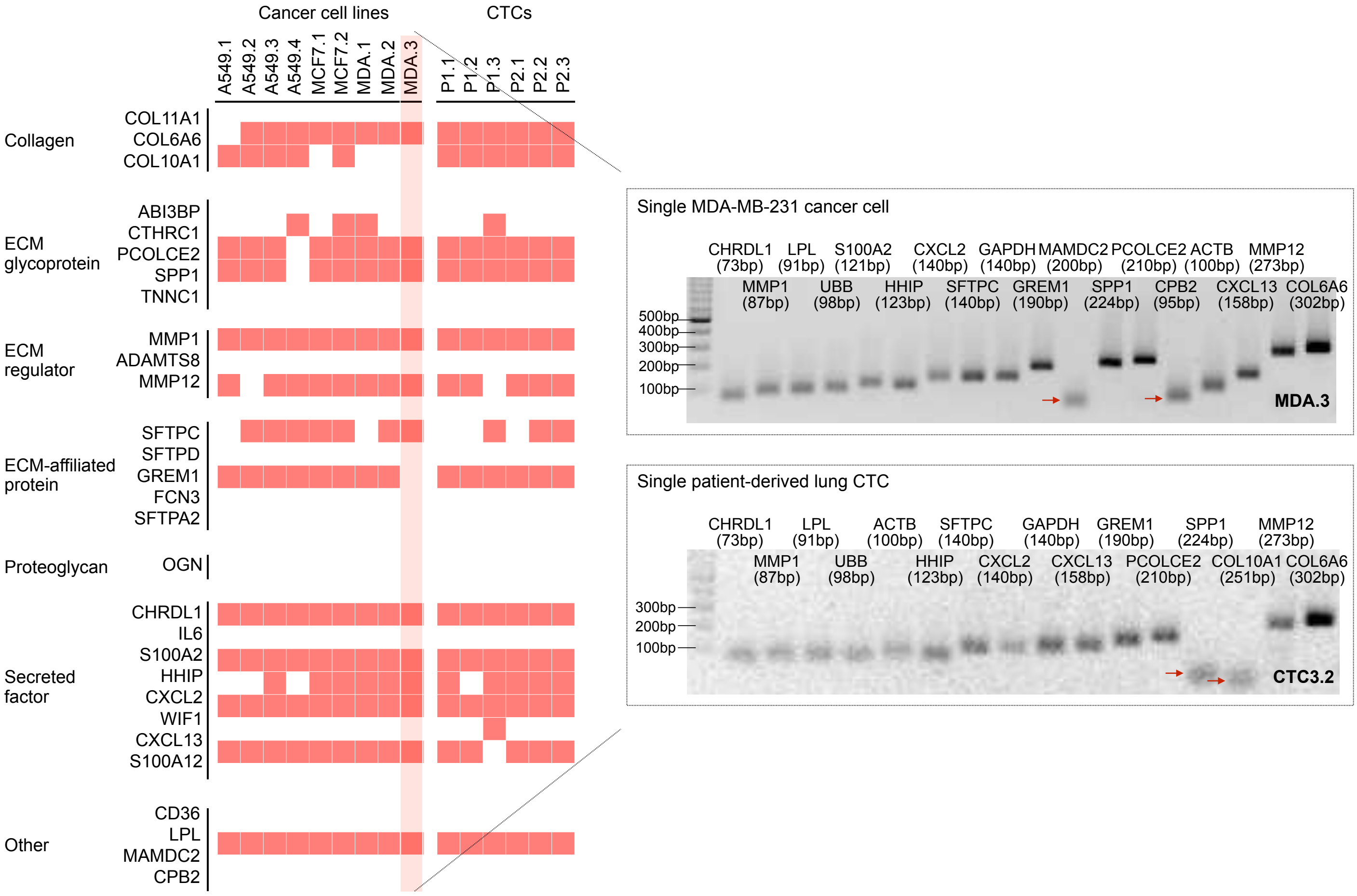


Fig. S8. Initial screening of 29 matrisome genes using single A549, MCF7, and MDA-MB-231 cancer cell lines. Exemplary gel electrophoresis images of PCR products preamplified with multi-gene PCR assays from single MDA-MB-231 cell and patient-derived CTC are shown.

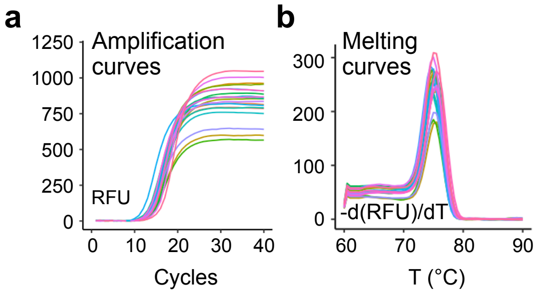


Fig. S9. (a) Amplification curves and (b) melting curves of amplicons obtained with CHRDL1 gene primer from 20 CTCs.

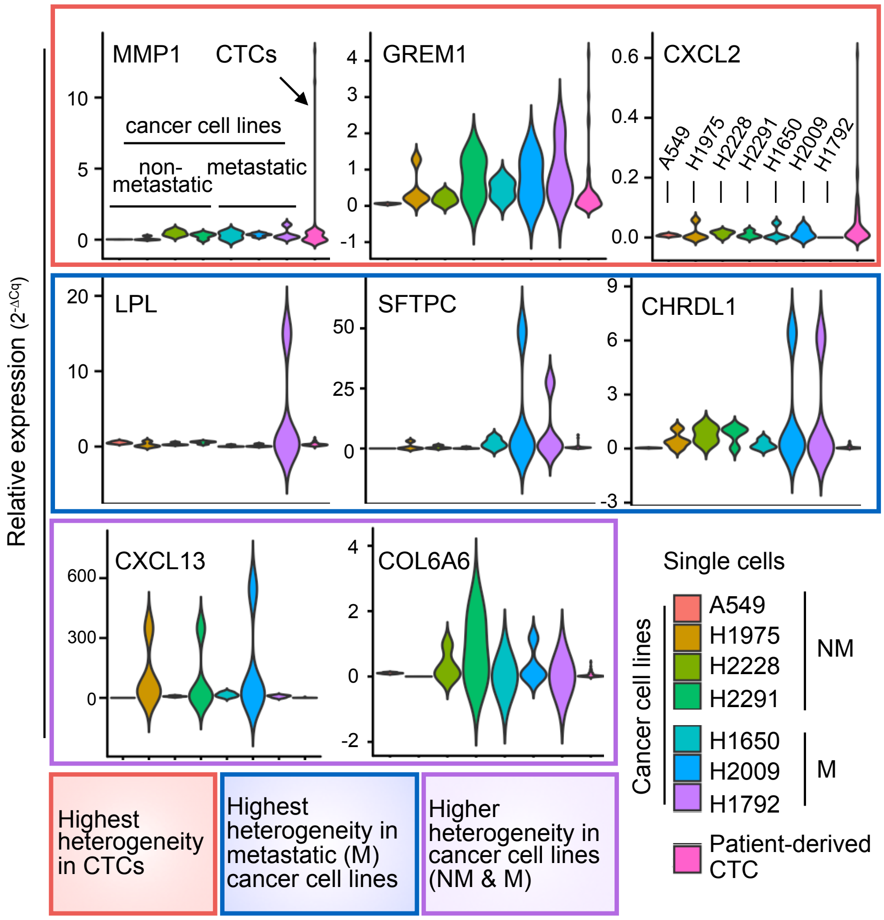


Fig. S10. Cell-to-cell heterogeneity in matrisome gene expression from six different lung cancer cell lines and patient-derived CTCs. GAPDH- and UBB-normalized gene expression values are shown. Single cells from metastatic (H1650, H2009, and H1792) and non-metastatic (A549, H1975, H2228 and H2291) lung cancer cell lines were cultured under the same conditions. Each experiment was performed in duplicate.

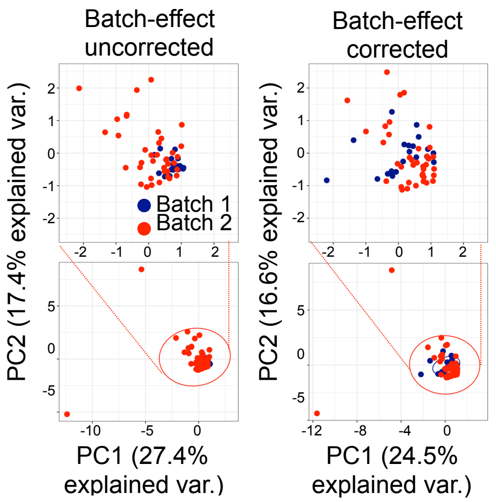


Fig. S11. PCA plots of targeted transcriptome data from 61 patient-derived CTCs obtained from two different laboratories demonstrating batch-effect correction.

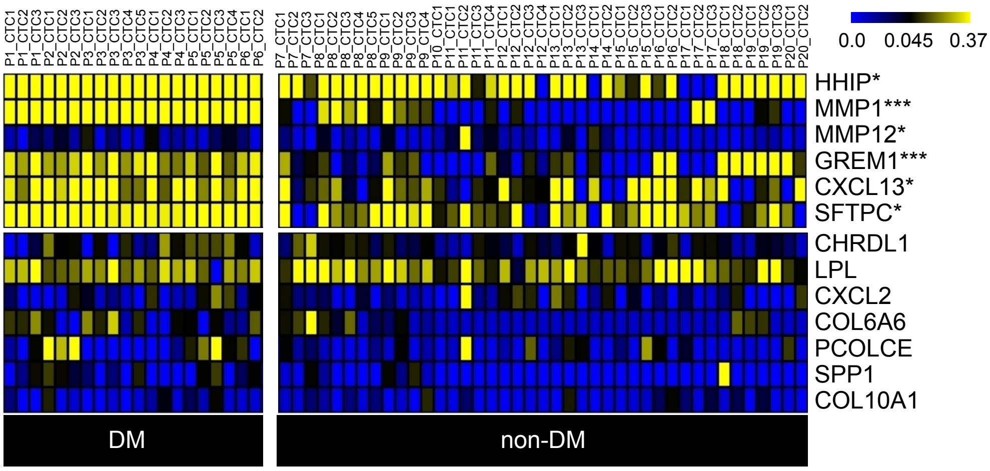


Fig. S12. Heatmap comparing expression profiles between DM and non-DM patient groups (*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$; Wilcoxon rank sum test).

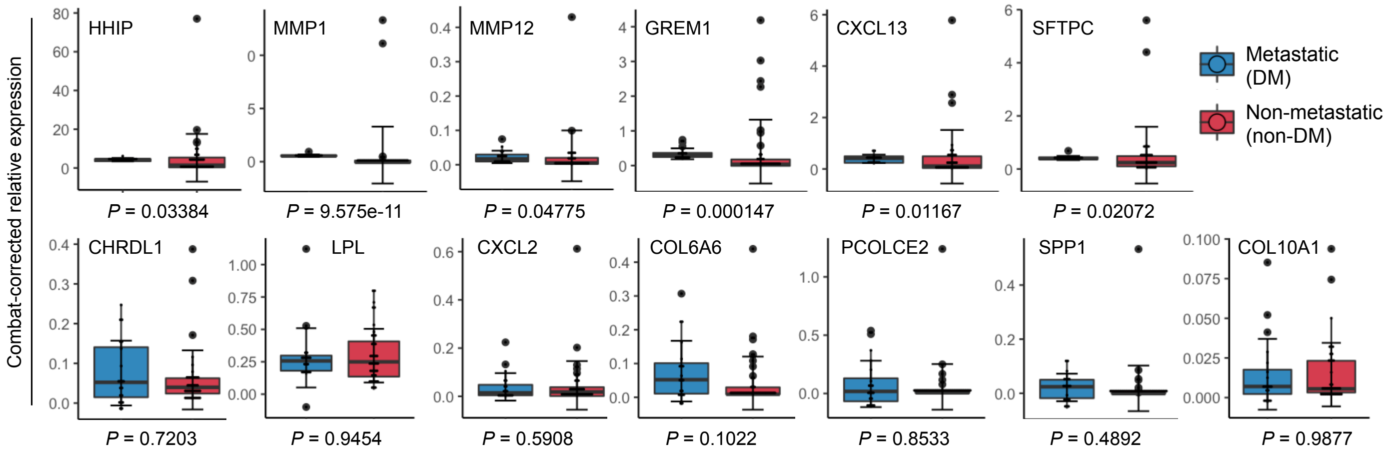
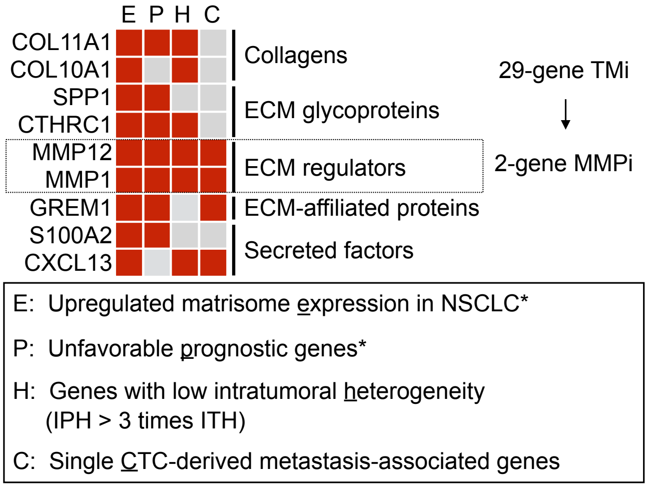


Fig. S13. Relative gene expression of matrisome genes in single CTCs comparing non-metastatic (non-DM) and metastatic (DM) patient group. GAPDH- and UBB-normalized gene expression values were subjected to Wilcoxon rank sum test.



	E	P	H	C	
COL11A1	🟥	🟥	🟥	⬜	Collagens
COL10A1	🟥	⬜	🟥	⬜	
SPP1	🟥	🟥	⬜	⬜	ECM glycoproteins
CTHRC1	🟥	🟥	🟥	⬜	
MMP12	🟥	🟥	🟥	🟥	ECM regulators
MMP1	🟥	🟥	🟥	🟥	
GREM1	🟥	🟥	⬜	🟥	ECM-affiliated proteins
S100A2	🟥	🟥	⬜	⬜	Secreted factors
CXCL13	🟥	⬜	🟥	🟥	

E: Upregulated matrisome expression in NSCLC*

P: Unfavorable prognostic genes*

H: Genes with low intratumoral heterogeneity
(IPH > 3 times ITH)

C: Single CTC-derived metastasis-associated genes

Fig. S14. Summary schematic of the refined prognostic gene signature tailored to the level of ITH and single-CTC-derived expression data. The asterisk denotes that data are based on our previous findings.

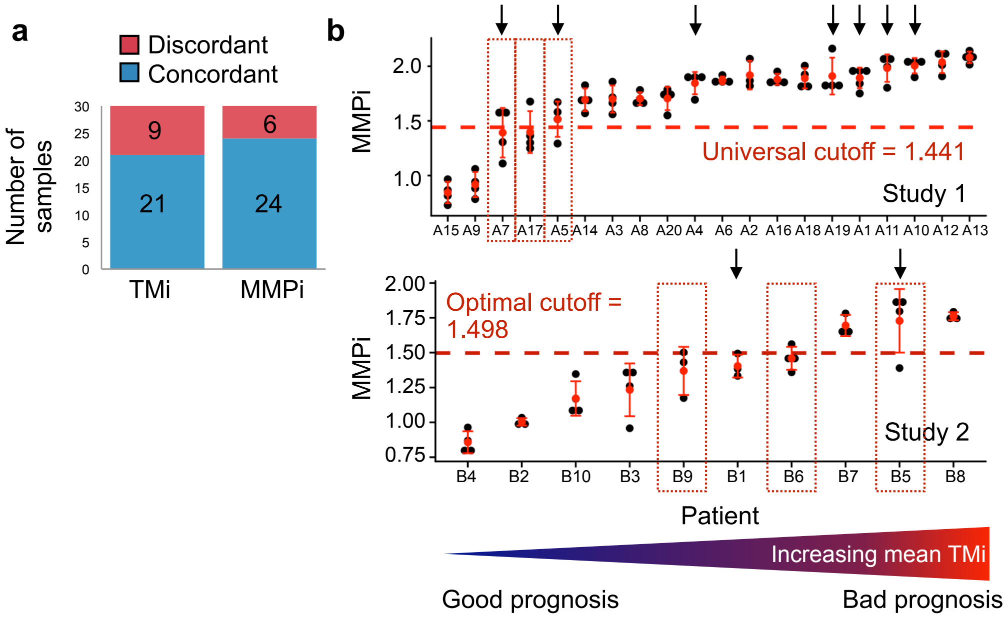
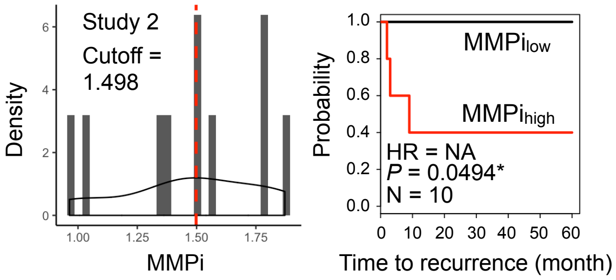


Fig. S15. Reduced number of discordant samples. (a) Number of ITH-driven misclassified patients in both studies. (b) MMPi distribution in spatially distinct tumor sectors. The universal cutoff value and the optimal cutoff value were used for patient stratification in Study 1 (top) and 2 (bottom), respectively. Dotted red boxes represent discordant samples having both MMPi_{low} and $\text{MMPi}_{\text{high}}$ tumor sectors. Arrows indicate TMi-defined discordant samples. Patients are ordered by increasing mean MMPi.

MMPi stratification



Group	n	Events (Recurrence)	median	0.95 CL (lower)	0.95 CL (upper)
MMPi_{low}	5	0	NA	NA	NA
$\text{MMPi}_{\text{high}}$	5	3	9	3	NA

Fig. S16. Patient-level prognostic accuracy of MMPi in Study 2. Gaussian kernel density distribution (left) and Kaplan-Meier survival curves (right) are drawn. The optimal cutoff value was used for patient stratification. Hazard ratio (HR), log-rank *P*-value, and the number of patients successfully stratified (n) are stated. The resulting summary statistics is shown below.

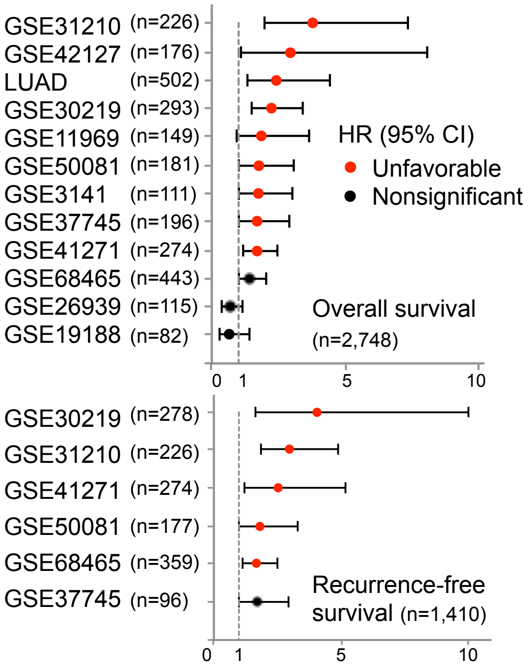
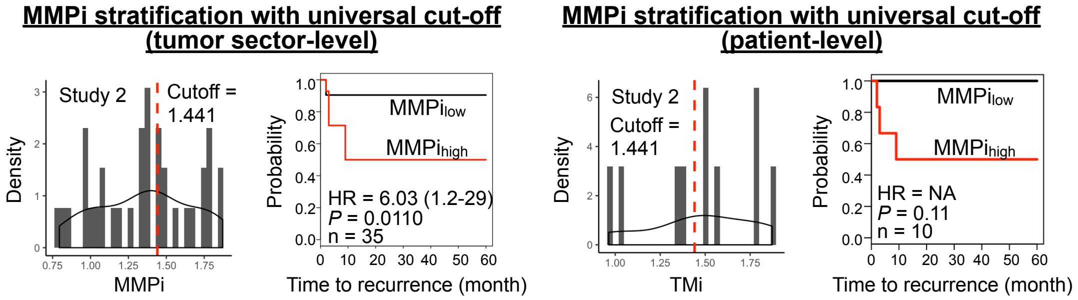


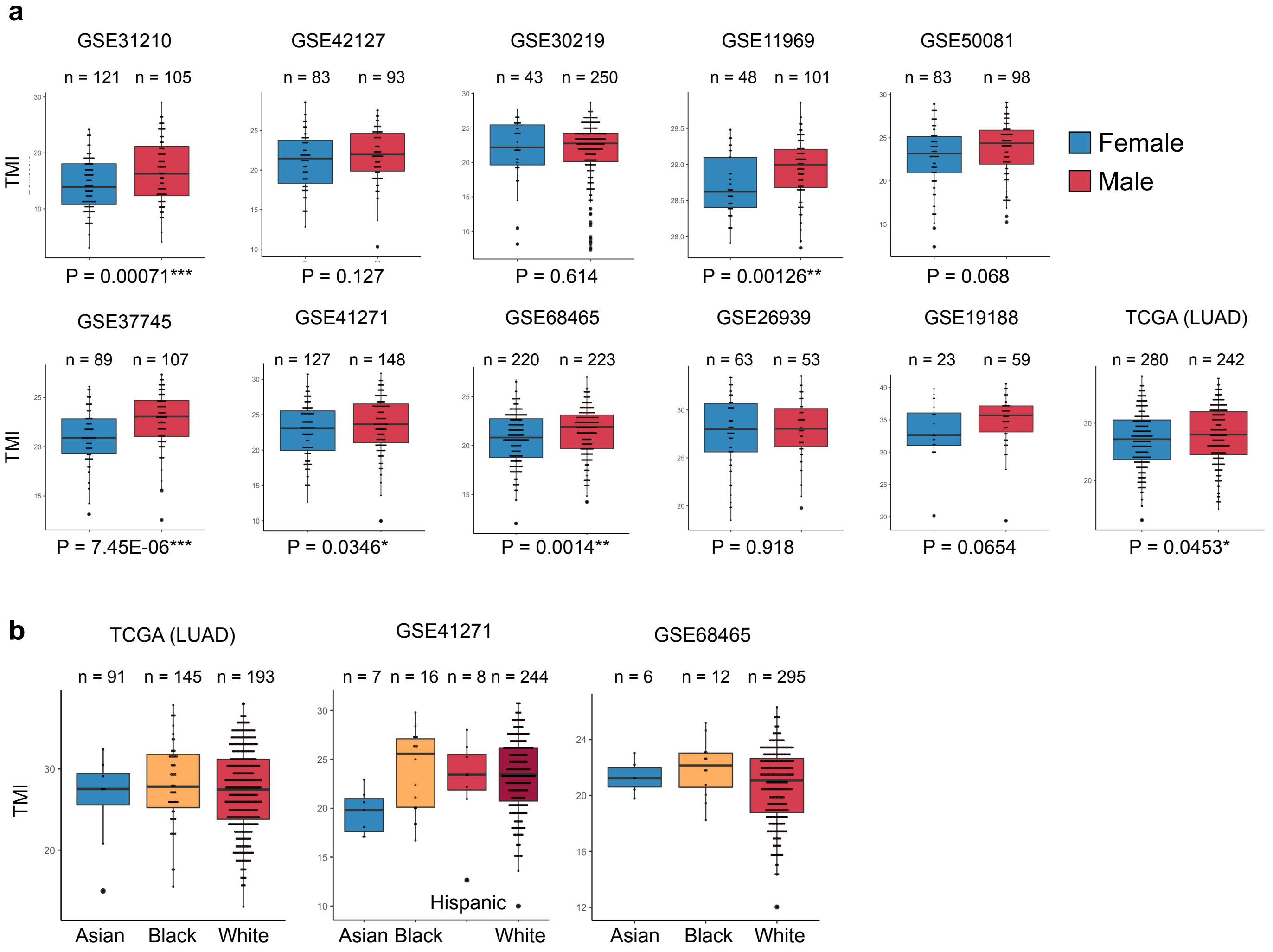
Fig. S17. Forest plot depicting the association between MMPi and overall survival (top) and recurrence-free survival (bottom) of lung cancer.



Group	n	Events (Recurrence)	median	0.95 CL (lower)	0.95 CL (upper)
MMPi_{low}	21	2	NA	NA	NA
$\text{MMPi}_{\text{high}}$	14	7	9	9	NA

Group	n	Events (Recurrence)	median	0.95 CL (lower)	0.95 CL (upper)
MMPi_{low}	4	0	NA	NA	NA
$\text{MMPi}_{\text{high}}$	6	3	9	3	NA

Fig. S18. Tumor sector- and patient-level prognostic accuracy of MMPi in Study 2. Predefined universal cutoff was applied for patient stratification. Gaussian kernel density distribution (left) and Kaplan-Meier survival curves (right) are drawn. The optimal cutoff value was used for patient stratification. Hazard ratio (HR), log-rank *P*-value, and the number of patients successfully stratified (n) are stated. The resulting summary statistics is shown below.



Wilcox rank sum test *P*	
Asian vs. Black	0.595
Asian vs. White	0.785
Black vs. White	0.260

Kruskal-Wallis test *P*
0.510

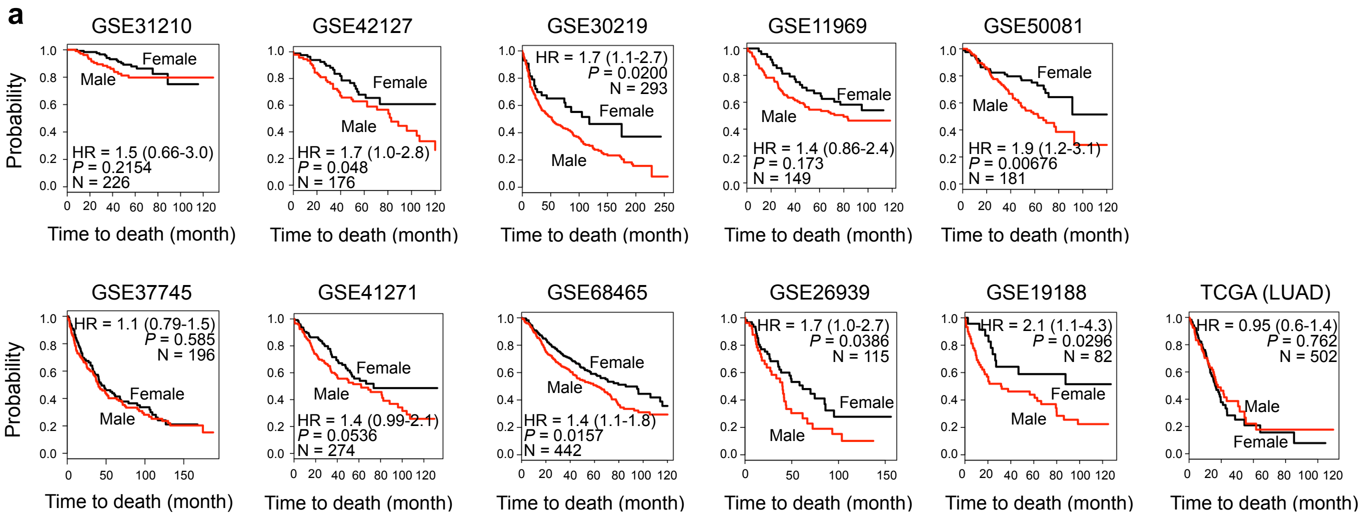
Wilcox rank sum test *P*	
Asian vs. Black	0.0329
Asian vs. Hispanic	0.0401
Asian vs. White	0.0054
Black vs. Hispanic	0.697
Black vs. White	0.53
Hispanic vs. White	0.937

Kruskal-Wallis test *P*
0.0442

Wilcox rank sum test *P*	
Asian vs. Black	0.494
Asian vs. White	0.698
Black vs. White	0.107

Kruskal-Wallis test *P*
0.256

Fig. S19. TMi by gender and ethnicity. (a) Gender differences in TMi. The accession codes of validation datasets, number of samples (n), and Wilcoxon rank sum test P values are stated. (*** *P* < 0.001, ** *P* < 0.01, * *P* < 0.05; Wilcoxon rank sum test). (b) Racial differences in TMi. The accession codes of validation datasets, number of samples (n), Wilcoxon rank sum test P values, and Kruskal-Wallis test *P* values are stated.



b

GSE41271	HR	95% Lower	95% Upper	Log-rank P
Asian	1			
Black	3.10	0.373	25.77	0.295
Hispanic	2.53	0.262	24.31	0.423
White	3.11	0.433	22.26	0.260
All	0.676			

GSE68465	HR	95% Lower	95% Upper	Log-rank P
Asian	1			
Black	1.88	0.378	9.318	0.441
White	2.26	0.559	9.159	0.252
All	0.458			

Fig. S20. Prognostic performance of gender and ethnicity. (a) Kaplan-Meier survival curves using previously analyzed validation datasets. Hazard ratio (HR), log-rank P-value (*P*), and the number of patients successfully stratified (N) determined from univariate Cox regression analyses are shown on each survival KM curve. Black and red KM curves represent female and male patients, respectively. (b) Estimates derived from univariate Cox regression analyses are shown.

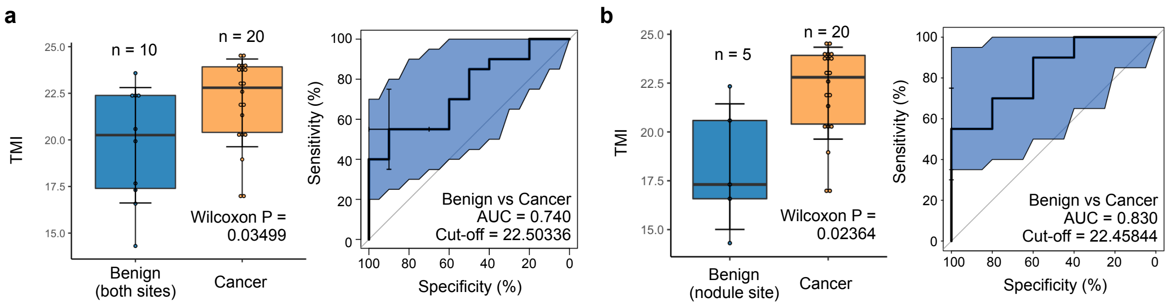
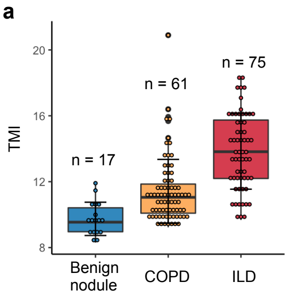


Fig. S21. TMi performance in distinguishing benign nodules from cancer using epithelial-lining fluid (ELF) samples, which were collected from subsegmental bronchi close to the pulmonary nodule and from the contralateral lung from patients with malignant or benign diagnosis (GSE27489). (a) Comparing TMi of benign ELF samples collected at both sites versus cancer ELF samples. n denotes the number of samples analyzed. No data were excluded from the analysis. (b) Comparing TMi of benign ELF samples collected at nodule site versus cancer ELF samples. The *optimal* TMi cut-off values were determined using the Cutoff Finder algorithm, as described in Methods. The receiver-operating characteristics (ROC) curves are drawn with 95% confidence intervals. Raw data was obtained and RMA-normalized using GEOquery and affy R/Bioconductor package, respectively. OGN expression data was missing in the dataset.



Wilcox rank sum test	
Benign nodule vs. COPD	6.64E-05
Benign nodule vs. ILD	7.51E-09
COPD vs. ILD	6.05E-10

Kruskal-Wallis test
1.03E-13

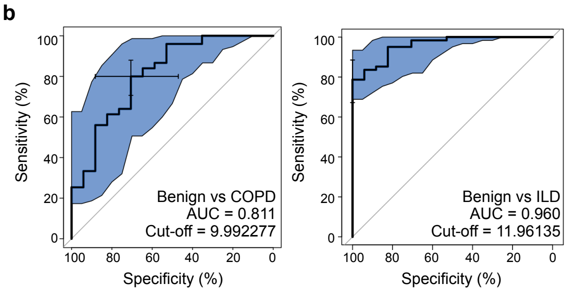


Fig. S22. TMi performance in distinguishing benign nodules from other types of lung disease (e.g., chronic obstructive pulmonary disease (COPD) and interstitial lung disease (ILD)) using tissue samples (GSE47460 probed with GPL6480 profiling platform). (a) Comparing TMi of benign nodules, COPD and ILD (left). n denotes the number of samples analyzed. No data were excluded from the analysis. Wilcox rank sum test and Kruskal-Wallis test *P*-values are stated (right). (b) ROC curves are drawn with 95% confidence intervals. The *optimal* TMi cut-off values were determined using the Cutoff Finder algorithm, as described in Methods.

Supplementary Tables

Table S1. Patient information (Study 1).

	Tumor	TMi	MMPi	Sex	Age	Histology	Stage
GSM835268_02_A_TK.CEL	Tumor	19.48971767	1.945513901	Female	44	adeno	2B
GSM835269_02_B_TK.CEL	Tumor	24.18669724	1.957698704	Female	44	adeno	2B
GSM835270_02_C_TK.CEL	Tumor	19.13931036	1.841950955	Female	44	adeno	2B
GSM835271_02_D_TK.CEL	Tumor	18.94933118	1.968295019	Female	44	adeno	2B
GSM835272_02_N_TK.CEL	Normal	8.388437177	0.935072211	Female	44	normal lung	
GSM835273_07_A_TK.CEL	Tumor	23.90353018	1.751130059	Male	65	adeno	2B
GSM835274_07_B_TK.CEL	Tumor	23.12958417	1.819835094	Male	65	adeno	2B
GSM835275_07_C_TK.CEL	Tumor	24.35945625	1.867754678	Male	65	adeno	2B
GSM835276_07_D_TK.CEL	Tumor	24.81849628	2.06761209	Male	65	adeno	2B
GSM835277_07_N_TK.CEL	Normal	9.727908671	0.71542968	Male	65	normal lung	
GSM835278_08_A_TK.CEL	Tumor	23.00907989	1.715346185	Female	54	mixed	1A
GSM835279_08_B_TK.CEL	Tumor	21.04990173	1.560188541	Female	54	mixed	1A
GSM835280_08_C_TK.CEL	Tumor	24.19324283	1.858029441	Female	54	mixed	1A
GSM835281_08_D_TK.CEL	Tumor	22.09079374	1.664151952	Female	54	mixed	1A
GSM835282_08_N_TK.CEL	Normal	8.370014181	0.846890731	Female	54	normal lung	
GSM835283_10_A_TK.CEL	Tumor	24.51049987	1.916142244	Male	67	mixed	1B
GSM835284_10_B_TK.CEL	Tumor	20.74919063	1.883030269	Male	67	mixed	1B
GSM835285_10_C_TK.CEL	Tumor	25.05444035	1.88754711	Male	67	mixed	1B
GSM835286_10_D_TK.CEL	Tumor	18.3045224	1.693422545	Male	67	mixed	1B
GSM835287_10_N_TK.CEL	Normal	7.168335446	0.551313247	Male	67	normal lung	
GSM835288_11_A_TK.CEL	Tumor	24.52852228	1.580925016	Male	65	mixed	1A
GSM835289_11_B_TK.CEL	Tumor	23.07041406	1.671643351	Male	65	mixed	1A
GSM835290_11_C_TK.CEL	Tumor	23.78853085	1.515724485	Male	65	mixed	1A
GSM835291_11_D_TK.CEL	Tumor	16.98162218	1.292276889	Male	65	mixed	1A
GSM835292_11_N_TK.CEL	Normal	8.821429566	1.163021843	Male	65	normal lung	
GSM835293_17_A_TK.CEL	Tumor	25.88265615	1.918466354	Male	54	adeno	2B
GSM835294_17_B_TK.CEL	Tumor	25.08586988	1.843655982	Male	54	adeno	2B
GSM835295_17_C_TK.CEL	Tumor	23.29688615	1.873410102	Male	54	adeno	2B
GSM835296_17_D_TK.CEL	Tumor	24.6069508	1.844512128	Male	54	adeno	2B
GSM835297_17_N_TK.CEL	Normal	8.64513943	0.836985511	Male	54	normal lung	
GSM835298_22_A_TK.CEL	Tumor	23.67359443	1.581452547	Male	59	adeno	2A
GSM835299_22_B_TK.CEL	Tumor	20.87986188	1.569179445	Male	59	adeno	2A
GSM835300_22_C_TK.CEL	Tumor	17.72216439	1.110041052	Male	59	adeno	2A
GSM835301_22_D_TK.CEL	Tumor	21.69870499	1.30578094	Male	59	adeno	2A
GSM835302_22_N_TK.CEL	Normal	8.787120919	0.767217609	Male	59	normal lung	
GSM835303_27_A_TK.CEL	Tumor	24.51309961	1.685494245	Female	65	adeno	1B
GSM835304_27_B_TK.CEL	Tumor	24.70293796	1.701697784	Female	65	adeno	1B
GSM835305_27_C_TK.CEL	Tumor	25.20381171	1.78177728	Female	65	adeno	1B
GSM835306_27_D_TK.CEL	Tumor	25.06414792	1.642298613	Female	65	adeno	1B
GSM835307_27_N_TK.CEL	Normal	9.421325576	0.850582851	Female	65	normal lung	
GSM835308_28_A_TK.CEL	Tumor	14.24464125	0.942084702	Male	61	adeno	1B
GSM835309_28_B_TK.CEL	Tumor	15.7745397	0.784777257	Male	61	adeno	1B
GSM835310_28_C_TK.CEL	Tumor	14.23122417	0.882796497	Male	61	adeno	1B
GSM835311_28_D_TK.CEL	Tumor	17.18289346	1.059977972	Male	61	adeno	1B
GSM835312_28_N_TK.CEL	Normal	6.659556999	0.559125972	Male	61	normal lung	
GSM835313_34_A_TK.CEL	Tumor	24.20591447	2.036795337	Male	58	squamous	1B
GSM835314_34_B_TK.CEL	Tumor	18.19810956	1.899685549	Male	58	squamous	1B
GSM835315_34_C_TK.CEL	Tumor	27.19756689	2.038919048	Male	58	squamous	1B
GSM835316_34_D_TK.CEL	Tumor	26.76463645	2.044136094	Male	58	squamous	1B
GSM835317_34_N_TK.CEL	Normal	8.441697328	0.719920682	Male	58	normal lung	
GSM835318_41_A_TK.CEL	Tumor	21.62777216	1.994425954	Male	65	adeno	1B
GSM835319_41_B_TK.CEL	Tumor	22.57742987	2.086207261	Male	65	adeno	1B

GSM835320_41_C_TK.CEL	Tumor	23.80694814	2.045588247	Male	65	adeno	1B
GSM835321_41_D_TK.CEL	Tumor	18.75460022	1.80296284	Male	65	adeno	1B
GSM835322_41_N_TK.CEL	Normal	7.806526006	0.819809632	Male	65	normal lung	
GSM835323_42_A_TK.CEL	Tumor	25.25557138	2.099046083	Male	62	squamous	1A
GSM835324_42_B_TK.CEL	Tumor	21.31813957	1.908175782	Male	62	squamous	1A
GSM835325_42_C_TK.CEL	Tumor	25.93059288	2.127901742	Male	62	squamous	1A
GSM835326_42_D_TK.CEL	Tumor	23.45166617	2.011141128	Male	62	squamous	1A
GSM835327_42_N_TK.CEL	Normal	6.71881133	0.529976911	Male	62	normal lung	
GSM835328_43_A_TK.CEL	Tumor	25.69601914	2.01497391	Male	61	mixed	1B
GSM835329_43_B_TK.CEL	Tumor	26.08874519	2.094855663	Male	61	mixed	1B
GSM835330_43_C_TK.CEL	Tumor	26.52775797	2.141321668	Male	61	mixed	1B
GSM835331_43_D_TK.CEL	Tumor	26.47414148	2.07300947	Male	61	mixed	1B
GSM835332_43_N_TK.CEL	Normal	12.31222553	1.366163649	Male	61	normal lung	
GSM835333_57_A_TK.CEL	Tumor	24.45139248	1.570066983	Male	69	adeno	1B
GSM835334_57_B_TK.CEL	Tumor	24.18599349	1.689656086	Male	69	adeno	1B
GSM835335_57_C_TK.CEL	Tumor	23.87251052	1.688045144	Male	69	adeno	1B
GSM835336_57_D_TK.CEL	Tumor	24.51992115	1.825527728	Male	69	adeno	1B
GSM835337_57_N_TK.CEL	Normal	6.554569172	0.581857174	Male	69	normal lung	
GSM835338_58_A_TK.CEL	Tumor	17.76759386	0.965247838	Female	38	adeno	1A
GSM835339_58_B_TK.CEL	Tumor	16.61872984	0.729615195	Female	38	adeno	1A
GSM835340_58_C_TK.CEL	Tumor	18.97458533	0.866156109	Female	38	adeno	1A
GSM835341_58_D_TK.CEL	Tumor	15.994222	0.815564553	Female	38	adeno	1A
GSM835342_58_N_TK.CEL	Normal	6.879915607	0.520434022	Female	38	normal lung	
GSM835343_59_A_TK.CEL	Tumor	20.84738228	1.874087724	Male	65	mixed	1B
GSM835344_59_B_TK.CEL	Tumor	23.45503325	1.950814725	Male	65	mixed	1B
GSM835345_59_C_TK.CEL	Tumor	24.50455309	1.831926545	Male	65	mixed	1B
GSM835346_59_D_TK.CEL	Tumor	23.68791919	1.861657643	Male	65	mixed	1B
GSM835347_59_N_TK.CEL	Normal	8.436073128	0.821111901	Male	65	normal lung	
GSM835348_62_A_TK.CEL	Tumor	18.23210735	1.676468481	Male	65	adeno	1B
GSM835349_62_B_TK.CEL	Tumor	14.96083824	1.300061003	Male	65	adeno	1B
GSM835350_62_C_TK.CEL	Tumor	16.5382644	1.248981281	Male	65	adeno	1B
GSM835351_62_D_TK.CEL	Tumor	18.55073652	1.363625909	Male	65	adeno	1B
GSM835352_62_N_TK.CEL	Normal	7.525615701	0.76230851	Male	65	normal lung	
GSM835353_85_A_TK.CEL	Tumor	25.16421012	1.923851162	Male	62	squamous	2B
GSM835354_85_B_TK.CEL	Tumor	23.57052677	1.800175571	Male	62	squamous	2B
GSM835355_85_C_TK.CEL	Tumor	24.81184367	1.830494442	Male	62	squamous	2B
GSM835356_85_D_TK.CEL	Tumor	26.16514466	2.004432342	Male	62	squamous	2B
GSM835357_85_N_TK.CEL	Normal	7.520300682	0.976481958	Male	62	normal lung	
GSM835358_93_A_TK.CEL	Tumor	22.24508544	1.8036776	Male	64	squamous	2B
GSM835359_93_B_TK.CEL	Tumor	20.49354309	1.829528038	Male	64	squamous	2B
GSM835360_93_C_TK.CEL	Tumor	22.84187389	1.844231482	Male	64	squamous	2B
GSM835361_93_D_TK.CEL	Tumor	26.48924072	2.161256938	Male	64	squamous	2B
GSM835362_93_N_TK.CEL	Normal	8.358020853	0.962153834	Male	64	normal lung	
GSM835363_98_A_TK.CEL	Tumor	23.95144029	1.717244658	Male	69	mixed	1B
GSM835364_98_B_TK.CEL	Tumor	22.40547004	1.787165347	Male	69	mixed	1B
GSM835365_98_C_TK.CEL	Tumor	23.32451814	1.763652674	Male	69	mixed	1B
GSM835366_98_D_TK.CEL	Tumor	24.33820635	1.5505964	Male	69	mixed	1B
GSM835367_98_N_TK.CEL	Normal	7.86715623	0.786301705	Male	69	normal lung	

Table S2. Patient information (Study 2). ADC=Adenocarcinoma; SCC=Squamous cell carcinoma; LCC=Large-cell carcinoma; PC=Pleomorphic carcinoma; RFS=recurrence-free survival.

Patient	MMPi	TMi	RFS	RFS_time	OS	Age	Sex	Ethnicity	Smoking
Pa1.R1	1.333	22.351	0	60	63	59	2_M	1_Asian	1_Never
Pa1.R2	1.384	22.783	0	60	63	59	2_M	1_Asian	1_Never
Pa1.R3	1.494	23.164	0	60	63	59	2_M	1_Asian	1_Never
Pa2.R1	0.984	16.629	0	60	62	62	1_F	2_Black	1_Never
Pa2.R2	0.992	18.717	0	60	62	62	1_F	2_Black	1_Never
Pa2.R3	1.033	17.139	0	60	62	62	1_F	2_Black	1_Never
Pa3.R1	1.344	20.092	0	52	53	63	2_M	2_Black	2_Current
Pa3.R2	0.958	15.809	0	52	53	63	2_M	2_Black	2_Current
Pa3.R3	1.261	19.310	0	52	53	63	2_M	2_Black	2_Current
Pa3.R4	1.371	22.261	0	52	53	63	2_M	2_Black	2_Current
Pa4.R1	0.870	20.587	0	60	60	70	2_M	3_Caucasian	3_Former
Pa4.R2	0.964	21.169	0	60	60	70	2_M	3_Caucasian	3_Former
Pa4.R3	0.794	20.936	0	60	60	70	2_M	3_Caucasian	3_Former
Pa4.R4	0.802	20.339	0	60	60	70	2_M	3_Caucasian	3_Former
Pa5.R1	1.390	18.411	0	60	62	47	2_M	3_Caucasian	3_Former
Pa5.R2	1.872	22.887	0	60	62	47	2_M	3_Caucasian	3_Former
Pa5.R3	1.797	22.646	0	60	62	47	2_M	3_Caucasian	3_Former
Pa5.R4	1.854	22.471	0	60	62	47	2_M	3_Caucasian	3_Former
Pa6.R1	1.359	20.606	0	36		73	1_F	3_Caucasian	2_Current
Pa6.R2	1.462	20.863	0	36		73	1_F	3_Caucasian	2_Current
Pa6.R3	1.561	21.196	0	36		73	1_F	3_Caucasian	2_Current
Pa6.R4	1.455	21.954	0	36		73	1_F	3_Caucasian	2_Current
Pa7.R1	1.782	22.898	1	9	12	64	1_F	3_Caucasian	3_Former
Pa7.R2	1.639	23.439	1	9	12	64	1_F	3_Caucasian	3_Former
Pa7.R3	1.663	23.064	1	9	12	64	1_F	3_Caucasian	3_Former
Pa8.R1	1.739	23.754	1	3	4	53	2_M	3_Caucasian	2_Current
Pa8.R2	1.793	23.605	1	3	4	53	2_M	3_Caucasian	2_Current
Pa8.R3	1.752	23.716	1	3	4	53	2_M	3_Caucasian	2_Current
Pa9.R1	1.174	20.104	1	2	7	47	1_F	3_Caucasian	2_Current
Pa9.R2	1.431	21.364	1	2	7	47	1_F	3_Caucasian	2_Current
Pa9.R3	1.502	20.750	1	2	7	47	1_F	3_Caucasian	2_Current
Pa10.R1	1.168	21.534	0	55	57	60	1_F	3_Caucasian	2_Current
Pa10.R2	1.347	21.682	0	55	57	60	1_F	3_Caucasian	2_Current
Pa10.R3	1.085	21.331	0	55	57	60	1_F	3_Caucasian	2_Current
Pa10.R4	1.086	22.068	0	55	57	60	1_F	3_Caucasian	2_Current
Patient	Histology	Tumor Size (cm)	Plural Invasion (+Y or -N)	Lymph Node (+Y or -N)	Path T	Path N	Path M	Final Path	Stage
Pa1.R1	ADC	7.2; 0.15	N	N	T3	N0	M0	IIB	II
Pa1.R2	ADC	7.2; 0.15	N	N	T3	N0	M0	IIB	II
Pa1.R3	ADC	7.2; 0.15	N	N	T3	N0	M0	IIB	II
Pa2.R1	ADC	2x1.8x1.5	N	N	T2a	N0	M0	IB	I
Pa2.R2	ADC	2x1.8x1.5	N	N	T2a	N0	M0	IB	I
Pa2.R3	ADC	2x1.8x1.5	N	N	T2a	N0	M0	IB	I
Pa3.R1	ADC	2.5 ; 1.2	Y	N	T3	N0	M0	IIB	II
Pa3.R2	ADC	2.5 ; 1.2	Y	N	T3	N0	M0	IIB	II
Pa3.R3	ADC	2.5 ; 1.2	Y	N	T3	N0	M0	IIB	II
Pa3.R4	ADC	2.5 ; 1.2	Y	N	T3	N0	M0	IIB	II
Pa4.R1	ADC	4.400	Y	Y	T2b	N1	M0	IIB	II
Pa4.R2	ADC	4.400	Y	Y	T2b	N1	M0	IIB	II
Pa4.R3	ADC	4.400	Y	Y	T2b	N1	M0	IIB	II
Pa4.R4	ADC	4.400	Y	Y	T2b	N1	M0	IIB	II
Pa5.R1	ADC	4.000	Y	N	T2a	N0	M0	NA	NA
Pa5.R2	ADC	4.000	Y	N	T2a	N0	M0	NA	NA
Pa5.R3	ADC	4.000	Y	N	T2a	N0	M0	NA	NA

Pa5.R4	ADC	4.000	Y	N	T2a	N0	M0	NA	NA
Pa6.R1	ADC	2.500	N	Y	T1b	N1	M0	NA	NA
Pa6.R2	ADC	2.500	N	Y	T1b	N1	M0	NA	NA
Pa6.R3	ADC	2.500	N	Y	T1b	N1	M0	NA	NA
Pa6.R4	ADC	2.500	N	Y	T1b	N1	M0	NA	NA
Pa7.R1	SCC	2.000	N	N	T1a	N0	M0	IA	I
Pa7.R2	SCC	2.000	N	N	T1a	N0	M0	IA	I
Pa7.R3	SCC	2.000	N	N	T1a	N0	M0	IA	I
Pa8.R1	SCC	5.5 X 5.5 X 5.0	Y	N	T2b	N0	M0	NA	NA
Pa8.R2	SCC	5.5 X 5.5 X 5.0	Y	N	T2b	N0	M0	NA	NA
Pa8.R3	SCC	5.5 X 5.5 X 5.0	Y	N	T2b	N0	M0	NA	NA
Pa9.R1	LCC	6.500	Y	Y	T2b	N2	M0	IIIA	III
Pa9.R2	LCC	6.500	Y	Y	T2b	N2	M0	IIIA	III
Pa9.R3	LCC	6.500	Y	Y	T2b	N2	M0	IIIA	III
Pa10.R1	PC	5.000	N	Y	T2a	N1	NA	NA	NA
Pa10.R2	PC	5.000	N	Y	T2a	N1	NA	NA	NA
Pa10.R3	PC	5.000	N	Y	T2a	N1	NA	NA	NA
Pa10.R4	PC	5.000	N	Y	T2a	N1	NA	NA	NA

Table S3. ITH and IPH of each matrisome gene. ITH=Intra-tumor heterogeneity; IPH=Inter-patient heterogeneity.

	Study 1		Study 2	
Gene	ITH	IPH	ITH	IPH
FCN3	0.796538301	0.203461699	0.761243739	0.238756261
ADAMTS8	0.671307954	0.328692046	0.690965167	0.309034833
CD36	0.640306122	0.359693878	0.487628442	0.512371558
SFTPD	0.608910173	0.391089827	0.447855874	0.552144126
COL6A6	0.604475942	0.395524058	0.570692099	0.429307901
IL6	0.557276679	0.442723321	0.207899716	0.792100284
WIF1	0.540786536	0.459213464	0.19857888	0.80142112
CPB2	0.474528442	0.525471558	0.710846954	0.289153046
SPP1	0.442066269	0.557933731	0.395235656	0.604764344
CHRDL1	0.434849653	0.565150347	0.414153763	0.585846237
CXCL2	0.34894564	0.65105436	0.346606335	0.653393665
COL10A1	0.337178989	0.662821011	0.339730429	0.660269571
TNNC1	0.33030078	0.66969922	0.215587393	0.784412607
SFTPC	0.290379747	0.709620253	0.27408	0.72592
SFTPA2	0.290379747	0.709620253	0.152589467	0.847410533
GREM1	0.256309362	0.743690638	0.252070741	0.747929259
ABI3BP	0.25399081	0.74600919	0.511661631	0.488338369
CXCL13	0.247111606	0.752888394	0.294946293	0.705053707
COL11A1	0.243488901	0.756511099	0.301611171	0.698388829
HHIP	0.219792639	0.780207361	0.775099639	0.224900361
OGN	0.217694732	0.782305268	0.843301286	0.156698714
MMP1	0.192482046	0.807517954	0.107362382	0.892637618
CTHRC1	0.190945601	0.809054399	0.198772383	0.801227617
MAMDC2	0.183685893	0.816314107	0.330655667	0.669344333
S100A12	0.169620253	0.830379747	0.220413805	0.779586195
PCOLCE2	0.165827036	0.834172964	0.41018351	0.58981649
MMP12	0.138695039	0.861304961	0.23283474	0.76716526
LPL	0.116966725	0.883033275	0.639720714	0.360279286
S100A2	0.033770015	0.966229985	0.056392523	0.943607477

Table S4. Patient information (liquid biopsies). Pt=Patient; Hist=Histology; ADC=Adenocarcinoma; NT=Neuroendocrine tumor (LCNEC/carcinoid); NS=Never-smoker; ES=Ex-smoker; CS=Current smoker

Pt	CTC counts	Group _CTC	RFS	RFS _time	pT	pN	pM	Gender	Age	Hist	Stage	OS	OS_ time	Smoking
1	3	G1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	3	G1	NA	NA	NA	NA	NA	Female	38	ADC	1	0	0	NA
3	5	G2	0	NA	NA	NA	NA	Female	56	ADC	4	0	15.6	NA
4	3	G1	0	NA	NA	NA	NA	Male	72	ADC	4	0	15.9	NA
5	4	G2	0	NA	NA	NA	NA	Male	56	ADC	3	0	14.7	NA
6	2	G1	1	NA	NA	NA	NA	Male	64	ADC	4	1	6.3	NA
7	3	G1	0	0	T4	N2	M0	Male	55	ADC	3	0	48.2	NS
8	5	G2	0	0	T1a	N0	M0	Male	71	ADC	1	0	22.6	ES
9	4	G2	0	0	T1b	N2	M0	Male	46	ADC	3	0	14.3	ES
10	1	G1	0	0	T1b	N0	M0	Female	60	ADC	1	0	48.4	NS
11	4	G2	0	0	T2a	N0	M0	Female	58	ADC	1	0	50.5	NS
12	4	G2	0	0	T1a	N0	M0	Female	56	ADC	1	0	28.3	CS
13	3	G1	0	0	T1a	N0	M0	Female	31	NT	1	0	5.6	CS
14	2	G1	1	9.1	T1b	N2	M0	Female	69	ADC	3	0	23.7	ES
15	3	G1	NA	0	T1b	N3	M0	Male	57	ADC	3	0	55.1	CS
16	2	G1	0	0	T1a	N0	M0	Male	62	ADC	1	0	14.7	ES
17	3	G1	1	25	T1a	N0	M0	Male	55	ADC	1	0	32.9	CS
18	2	G1	0	0	T2a	N1	M0	Female	70	ADC	2	0	14.1	NS
19	3	G1	0	0	T1b	N0	M0	Male	65	ADC	1	0	31.8	NS
20	2	G1	0	0	T1b	N1	M0	Male	67	ADC	2	0	37.1	ES

Table S5. 32-gene assay.

Gene	Forward	Reverse	Bp
ABI3BP	TTGCCAAGTCACTGTCCCAAT	GTGGCTGGACAGATTTGAAAAAT	95
ACTB	GGCACCCAGCACAATGAAG	GCCGATCCACACGGAGTACT	70
ADAMTS8	AACAAAAGCTGCTCCGTGAT	TCTGGTTCAGGTGGACGAAC	175
CD36	TGGGAAAGTCACTGCGACAT	TGCAATACCTGGCTTTTCTCA	80
CHRDL1	TGCGAGTACAATGGGACAAC	GTTGCCGATTCTGAAAGAGC	73
COL10A1	CAGGCAACAGCATTATGACC	CATTTGACTCGGCATTGG	251
COL11A1	TCGCATTGACCTTCCTCTTC	TCCCGTTGTTTTTGATATTC	113
COL6A6	CATCGGTGCTGCACTCAG	CTGTGGTACAGATGTTGCG	302
CPB2	GATATGCTTACAAAAATCCACATTGG	GCATTTTTGGCTGCTTGTTCT	95
CTHRC1	TTGTTCAGTGGCTCACTTCG	TTCAATGGGAAGAGGTCCTG	102
CXCL13	GAGGCAGATGGAACTTGAGC	CTGGGGATCTTCGAATGCTA	158
CXCL2	AAACCCAAGTTAGTTCAATCCTG	TCTTGGATTCCTCAGCCTCTAT	140
FCN3	GTGAGCCAGGAGATCCAGTG	GGGCGAAAGTACGGTTACCA	357
GAPDH	CAAGCTCATTTCCTGGTATGAC	CAGTGAGGGTCTCTCTCTTCCT	142
GREM1	TTAAGCAGACCATCCACGA	TGTAGTTCAGGGCAGTTGAGT	190
HHIP	ATGGTGGGTTGTGCTTTCC	AGTTGTGTTTGTGCTTTCTGCT	123
IL6	GGTACATCCTCGACGGCATCT	GTGCCTCTTTGCTGCTTTCAC	81
LPL	CGAGTCGTCTTTCTCCTGATGAT	TTCTGGATTCCAATGCTTCGA	91
MAMDC2	TCTTCAGAGAAACTTCCACCTC	CCATACACCATATGGGAGGC	200
MMP1	TGTGGACCATGCCATTGAGA	TCTGCTTGACCCTCAGAGACC	87
MMP12	CGGGCAACTGGACACATCTACC	GCTCCACGGGCAAAAACCAC	273
OGN	GTTGACATTGATGCTGTACCACCC	GCTTGGGAGGAAGAACTGGA	241
PCOLCE2	TGCTGTCAGAGGAGCAGCTA	GCAAAGGCTTGAAGGTGAAA	210
S100A12	CACATTCCTGTGCATTGAGG	TGCAAGCTCCTTTGTAAGCA	167
S100A2	AGCTTTGTGGGGGAGAAAGT	CAGTGATGAGTGCCAGGAAA	121
SFTPA2	GCCTGAAAAGAAGGAGCAGC	GGCTTCCAACACAAACGTCC	124
SFTPC	AGCAAAGAGGTCCTGATGGA	ACAATCACCACGACGATGAG	140
SFTPD	AAGTGGGCTTCCAGATGTTG	CTGTGCCTCCGTAAATGGTT	181
SPP1	ACTCGAACGACTCTGATGATGT	GTCAGGTCTGCGAAACTTCTTA	224
TNNC1	CAGCAAAGGGAAATCTGAGG	TGATGGTCTCGCCTGTAGC	122
UBB	GCTTTGTTGGGTGAGCTTGT	CGAAGATCTGCATTTTGACCT	98
WIF1	GAAGGTTGGCATGGAAGA	TGTAATTGGATTCAGGTGGA	145

Table S6. Univariate survival analyses (OS).

Dataset	Country	Sample size	Cutoff	HR	Lower (95% CI)	Upper (95% CI)	Log-rank P
GSE31210	Japan	226	2.064	3.78	1.95	7.35	0.000025
GSE42127	USA	176	0.9113	2.94	1.067	8.093	0.02871
LUAD	USA	502	0.3566	2.43	1.33	4.42	0.0028
GSE30219	France	293	1.071	2.23	1.464	3.403	0.0001287
GSE11969	Japan	149	0.06185	1.87	0.9592	3.641	0.06181
GSE50081	Canada	181	1.28	1.76	1.015	3.066	0.04138
GSE3141	USA	111	2.103	1.74	0.9947	3.041	0.04933
GSE37745	Sweden	196	2.181	1.72	1.021	2.899	0.03904
GSE41271	USA	274	1.237	1.71	1.174	2.488	0.004628
GSE68465	USA	443	2.047	1.42	0.9886	2.033	0.05651
GSE26939	USA	115	0.1165	0.67	0.391	1.146	0.1411
GSE19188	Netherlands	82	0.6477	0.66	0.3092	1.407	0.2781

Table S7. Univariate survival analyses (RFS).

Dataset	Country	Sample size	Cutoff	HR	Lower (95% CI)	Upper (95% CI)	Log-rank P
GSE30219	France	278	0.9589	4.08	1.66	10.01	0.00088
GSE31210	Japan	226	2.064	2.98	1.82	4.89	0.0000057
GSE41271	USA	274	1.848	2.53	1.23	5.19	0.0088
GSE50081	Canada	177	1.441	1.81	0.99	3.31	0.049
GSE68465	USA	359	2.038	1.7	1.15	2.5	0.0069
GSE37745	Sweden	96	1.908	1.69	0.96	2.97	0.064

Table S8. Multivariate survival analyses (TCGA).

TCGA	N=481	HR	Lower (95% CI)	Upper (95% CI)	Log-rank P
MMP Group (Cutoff = 0.3566)	G1	1			
	G2	1.9335	1.015	3.683	0.0449
Age at initial pathologic diagnosis		1.0186	0.9991	1.038	0.0623
Smoking Status	Lifelong non-smoker	1			
	Current smoker	0.865	0.4184	1.788	0.6955
	Current reformed smoker	1.465	0.7743	2.772	0.2405
pStage	Stage1	1			
	Stage2	2.8792	1.7525	4.73	0.0000298
	Stage3	4.2344	2.5937	6.913	7.88E-09
	Stage4	4.0642	2.01	8.218	0.0000949
Gender	Female	1			
	Male	0.8011	0.5341	1.202	0.2836

SI References

1. Robin X*, et al.* (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77.
2. Leek JT, Johnson WE, Parker HS, Jaffe AE, & Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28(6):882-883.