

1 **A 5 PROTEIN-BASED SIGNATURE FOR RESECTABLE LUNG SQUAMOUS-**
2 **CELL CARCINOMA IMPROVES THE PROGNOSTIC PERFORMANCE OF**
3 **THE TNM STAGING**

4 Elena Martínez-Terroba^{1,2}, Carmen Behrens³, Jackeline Agorreta^{1,2,4,5,‡}, Eduard Monsó⁶,
5 Laura Millares⁶, Enriqueta Felip⁷, Rafael Rosell⁸, José Luis Ramirez⁸, Ana Ramirez¹,
6 Wenceslao Torre^{4,9}, Ignacio Gil-Bazo^{1,4,5,10}, Miguel A. Idoate^{2,4,5,11}, Juan P. de-
7 Torres^{4,5,12}, Ruben Pio^{1,4,5,13}, Ignacio I. Wistuba^{3,14}, María J. Pajares^{1,2,4,5*} and Luis M.
8 Montuenga^{1,2,4,5,* ‡}

9 *These authors share senior authorship.

10

11 ¹Program in Solid Tumors, CIMA, Pamplona, Spain.

12 ²Department of Pathology, Anatomy and Physiology, School of Medicine, University of
13 Navarra, Pamplona, Spain.

14 ³Department of Thoracic/Head and Neck Medical Oncology, The University of Texas
15 M.D. Anderson Cancer Center, Houston, TX, USA.

16 ⁴Navarra Health Research Institute (IDISNA), Pamplona, Spain.

17 ⁵Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain.

18 ⁶Neumology Service, Parc Taulí University Hospital, Sabadell, Spain; CIBER de
19 Enfermedades Respiratorias-CIBERES, Instituto de Salud Carlos III, Madrid, Spain.

20 ⁷Vall d'Hebron University Hospital, Institute of Oncology, Barcelona, Spain.

21 ⁸Catalan Institute of Oncology, Hospital Germans Trias i Pujol, Badalona, Spain.

22 ⁹Department of Thoracic Surgery, Clínica Universidad de Navarra, Pamplona, Spain

23 ¹⁰Department of Oncology, Clínica Universidad de Navarra, Pamplona, Spain.

24 ¹¹Department of Pathology, Clínica Universidad de Navarra, Pamplona, Spain.

25 ¹²Department of Neumology, Clínica Universidad de Navarra, Pamplona, Spain.

26 ¹³Department of Biochemistry and Genetics, School of Sciences, University of Navarra,
27 Pamplona, Spain.

28 ¹⁴Department of Translational Molecular Pathology, The University of Texas M.D.
29 Anderson Cancer Center, Houston.

30

31 **‡Corresponding author:** Program in Solid Tumors, Center for Applied Medical
32 Research (CIMA), Pio XII Avenue 55, 31008 Pamplona, Spain. Tel.: +34 948194700.
33 Fax: +34 948194714. E-mail addresses: jagorreta@unav.es and lmontuenga@unav.es

34

35

36

37

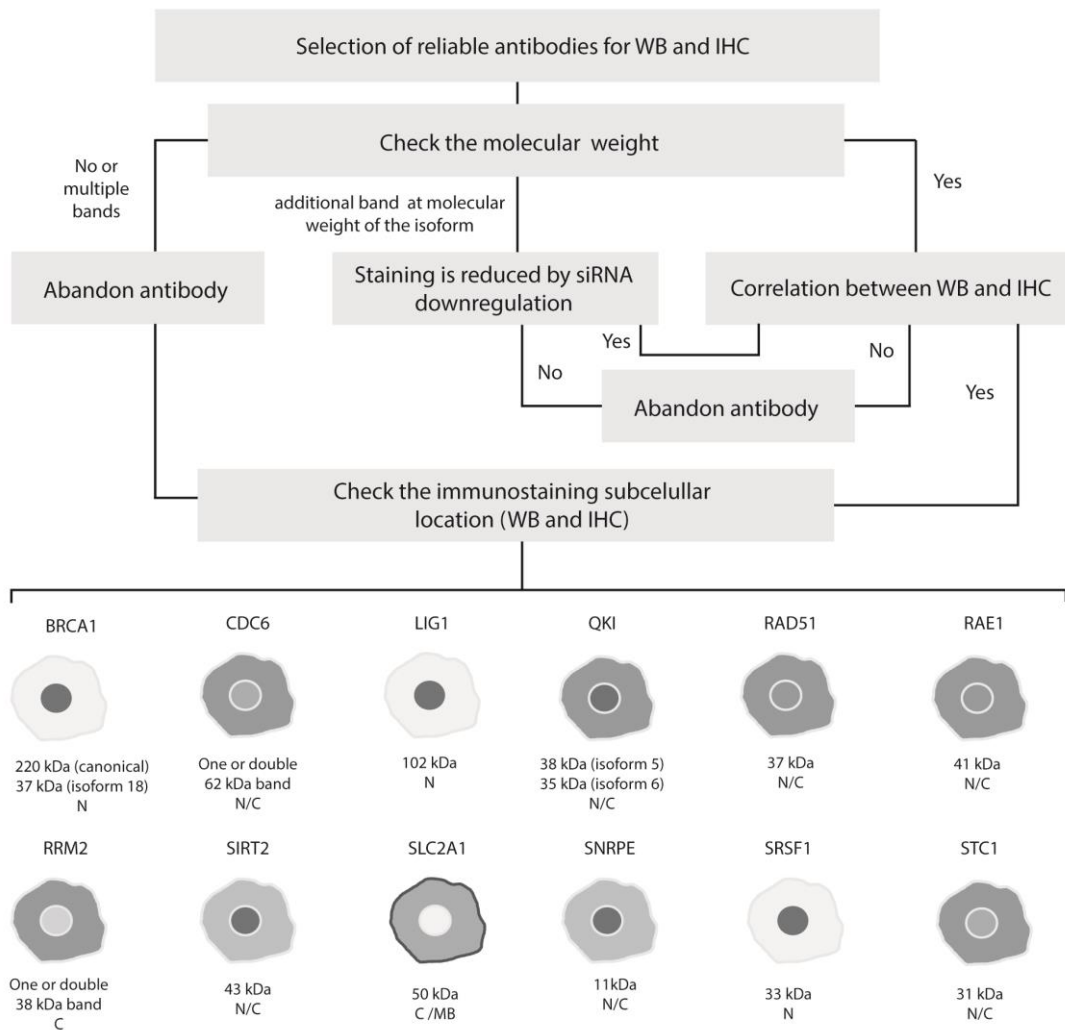
38

39

40

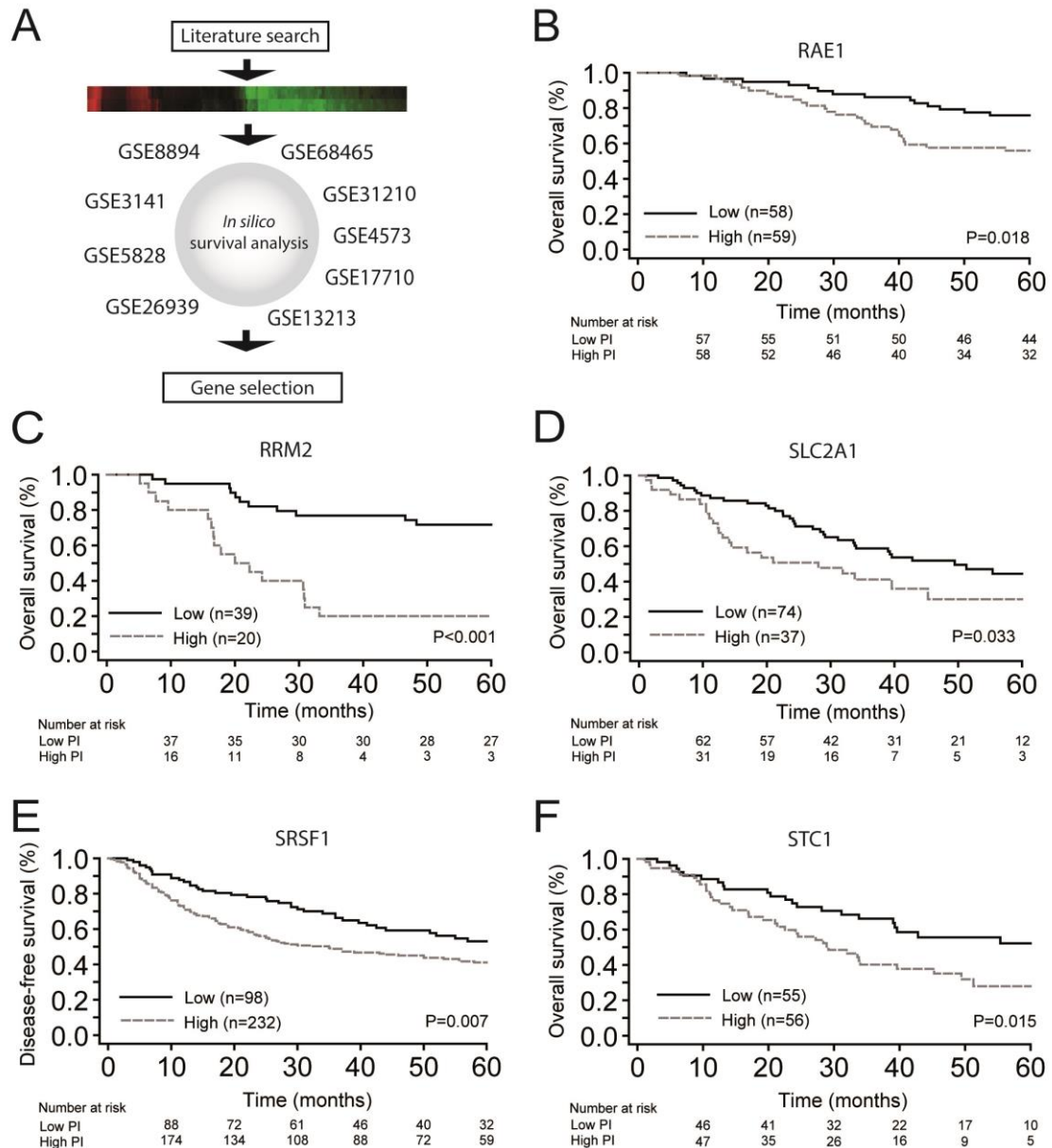
41

Supplementary Figure 1



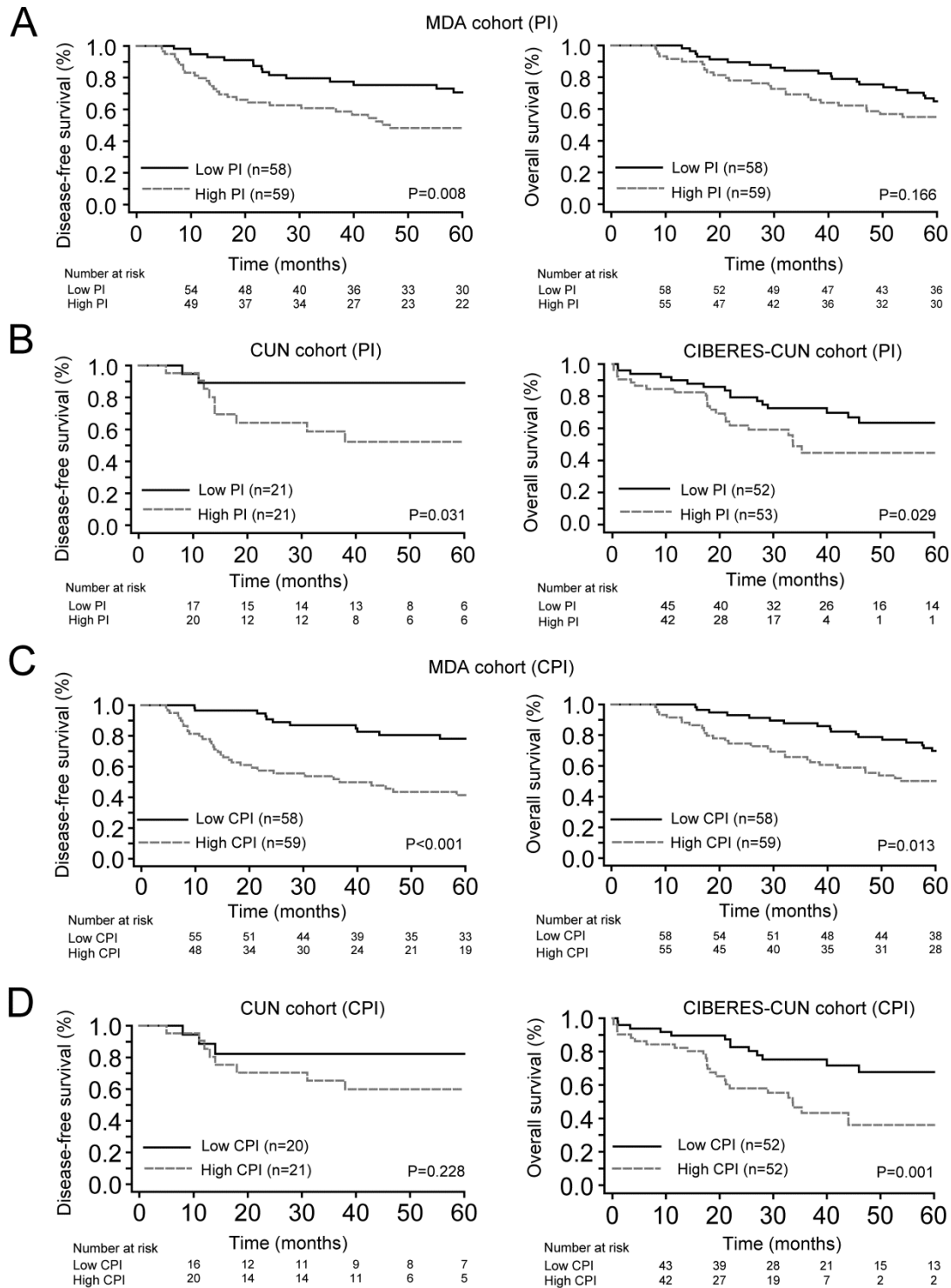
Supplementary Figure 1. Pipeline for antibody validation. We show a schematic draw of the IHC staining (grey gradient correlates with intensity) and the molecular weight observed by WB (N: nuclear; C: cytoplasmic and MB: membrane).

Supplementary Figure 2



Supplementary Figure 2. Survival analysis in microarray datasets. **A.** Outline showing the microarray datasets used in the study to select the candidate genes. **B.** Representative Kaplan-Meier survival curve for RAE1 expression in GSE13213 dataset. Patients were stratified into two groups according to the median. **C.** Representative Kaplan-Meier survival curve for RRM2 expression in GSE5828 dataset. Patients were stratified into two groups according to the upper tercile. **D.** Representative Kaplan-Meier survival curve for SLC2A1 expression in GSE3141 dataset. Patients were stratified into two groups according to the upper tercile. **E.** Representative Kaplan-Meier survival curve for SRSF1 expression in GSE68465 dataset. Patients were stratified into two groups according to the lower tercile. **F.** Representative Kaplan-Meier survival curve for STC1 expression in GSE3141 dataset. Patients were stratified into two groups according to the median. Differences between groups were evaluated using the log-rank test.

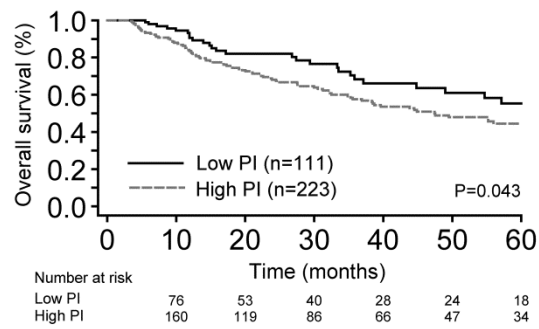
Supplementary Figure 3



Supplementary Figure 3. Association between PI or CPI and survival in SCC patients. **A.** Kaplan-Meier survival curves for high and low prognostic index (PI) groups in MDA patients for DFS (left panel) and OS (right panel). **B.** Kaplan-Meier survival curves for DFS in CUN (left panel) and for OS in CIBERES-CUN cohort (right panel). The risk groups were stratified according to the median of the PI. **C.** Association between CPI and survival in patients from MDA cohort. Kaplan-Meier survival curves

were generated for DFS (left panel) and OS (right panel). **D.** Kaplan-Meier survival curves for DFS in CUN (left panel) and for OS in CIBERES-CUN cohort (right panel). Patients were stratified into two risk groups according to the median of the CPI. Differences between groups were evaluated using the log-rank test.

Supplementary Figure 4



Supplementary Figure 4. Survival analysis of PI in SCC patients from TCGA at mRNA level. A. Kaplan-Meier survival curve for high and low prognostic index (PI) groups. Patients were stratified into two risk groups according to the first tercile of the PI. Differences between groups were evaluated using the log-rank test.

Supplementary Table 1. Patient clinical and pathological characteristics

	Training cohort MDA	Validation Cohort NATCH	Validation cohort CIBERES-CUN	
	MDA ¹ n=117	NATCH ² n=96	CIBERES ³ n=63	CUN ⁴ n=42
Stage*				
IA	27 (23.1%)	13 (13.5%)	16 (25.4%)	15 (35.7%)
IB	23 (19.7%)	44 (45.8%)	18 (28.6%)	5 (11.9%)
IIA	5 (4.3%)	2 (2.1%)	4 (6.3%)	4 (9.5%)
IIB	32 (27.4%)	17 (17.7%)	17 (27.3%)	7 (16.7%)
IIIA	24 (20.5%)	16 (16.7%)	8 (12.7%)	9 (21.4%)
IIIB	6 (5.1%)	4 (4.2%)	0 (0%)	1 (2.4%)
IV	0 (0%)	0 (0%)	0 (0%)	1 (2.4%)
Gender				
Male	71 (60.7%)	93 (96.9%)	59 (93.7%)	40 (95.2%)
Female	46 (39.3%)	3 (3.1%)	4 (6.3%)	2 (4.8%)
Age (years)				
<65	8 (6.8%)	46 (47.9%)	24 (38.1%)	21 (50.0%)
≥65	109 (93.2%)	50 (52.1%)	39 (61.9%)	21 (50.0%)
Adjuvant treatment				
No	67 (57.3%)	44 (45.8%)		27 (64.3%)
Yes	45 (38.5%)	52 (54.2%)		15 (35.7%)
Unknown	5 (4.2%)	0 (0%)	63 (100%)	0 (0%)
Tobacco				
Never	0 (0%)	34 (35.4%)		0 (0%)
Former	60 (51.3%)	18 (18.8%)		29 (69.0%)
Current	57 (48.7%)	44 (45.8%)		13 (31.0%)
Unknown	0 (0%)	0 (0%)	63 (100%)	0 (0%)

Data are n (%). ¹MDA = M.D. Anderson Cancer Center; ²NATCH = NATCH trial; ³CIBERES Pulmonary Biobank Platform; ⁴CUN = Clínica Universidad de Navarra. *8th TNM edition was used for tumor classification in MDA and CIBERES-CUN cohorts, and 6th TNM edition was used in NATCH cohort.

Supplementary Table 2. Conditions for protein detection by IHC and WB in cell extracts and FFPE samples.

Protein	Reference	Host ¹	Class ²	Pre-treatment ³	IHC ⁴	IHC detection method	WB ⁵
BRCA1	Genetex, Irvine (CA) (GTX70113)	M	M	PT-CIT pH6 (20 min 95 °C)	1:500	EnVision M	1:1000
CDC6	Santa Cruz, Dallas (TX) (sc-56273)	M	M	PT-EDTA pH9 (20 min 95 °C)	1:20	EnVision M	1:1000
LIG1	Areta International, Gerenzano (VA)	M	M	PT-EDTA pH9 (20 min 95 °C)	1:500	EnVision M	1:2000
QKI	Sigma, St. Louis (MO) (HPA019123)	R	P	PT-EDTA pH9 (20 min 95 °C)	1:500	EnVision R	1:2000
RAD51	Calbiochem, San Diego (CA) (PC130)	R	P	PT-EDTA pH9 (20 min 95 °C)	1:100	EnVision R	1:8000
RAE1	Genetex, Irvine (CA) (EPR6923)	R	M	PT-CIT pH6 (20 min 95 °C)	1:250	EnVision R	1:1000
RRM2	Sigma, St. Louis (MO) (WH0006241M1)	M	M	PT-CIT pH6 (20 min 95 °C)	1:3000	EnVision M	1:1000
SIRT2	Santa Cruz, Dallas (TX) (sc-28298)	M	M	PT-CIT pH6 (20 min 95 °C)	1:50	EnVision M	1:500
SLC2A1	Proteintech, Rosemont (IL) (21829-1-AP)	R	P	PT-CIT pH6 (20 min 95 °C)	1:1000	EnVision R	1:1000
SNRPE	Proteintech, Rosemont (IL) (20407-1-AP)	R	P	PT-CIT pH6 (20 min 95 °C)	1:200	EnVision R	1:500
SRSF1	Zymed, San Francisco (CA) (32-4600)	M	M	PT-EDTA pH9 (20 min 95 °C)	1:200	EnVision M	1:10000
STC1	Santa Cruz, Dallas (TX) (sc-14346)	G	P	PT-CIT pH6 (20 min 95 °C)	1:100	Secondary antibody/ Streptavidin HRP (1:200)	1:1000
β-Actin	Sigma, St. Louis (MO) (A1978)	M	M				1:4000
GAPDH	AbD Serotec (Bio-Rad), Hercules (CA) (MCA2427)	M	M				1:10000
GRB2	BD Biosciences, San Jose (CA) (610111)	M	M				1:1000
lamin A/C	Cell Signaling, Danvers (MA) (2032)	R	P				1:5000

¹M: Mouse, **R**: Rabbit, **G**: Goat; ²M: Monoclonal, **P**: Polyclonal; ³PT-CIT: PT-linker buffer citrate (ThermoScientific, Waltham, MA), **PT-EDTA**: PT-linker buffer Tris EDTA (ThermoScientific, Waltham, MA); ⁴Dilution for primary antibody in IHC; ⁵Dilution for primary antibody in WB.

Supplementary Table 3. Univariate and multivariate Cox proportional hazards analysis of PI and clinico-pathological parameters for DFS and OS in the MDA cohort

Training cohort	MDA n (%)	DFS ¹				OS ²			
		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value
PI³		2.55 (1.60-4.05)	<0.001	2.78 (1.72-4.50)	<0.001	1.77 (1.14-2.76)	0.012	1.82 (1.16-2.83)	0.008
Stage			<0.001		<0.001		0.023		0.046
I	50 (42.7%)	1		1		1		1	
II	37 (31.6%)	1.25 (0.56-2.80)	0.577	1.50 (0.65-3.45)	0.339	0.89 (0.42-1.86)	0.753	0.97 (0.46-2.05)	0.936
III	30 (25.7%)	3.83 (1.87-7.86)	<0.001	4.27 (2.03-8.99)	<0.001	2.19 (1.12-4.30)	0.022	2.14 (1.08-4.25)	0.030
Gender			0.232		0.371		0.562		
Male	71 (60.7%)	1		1		1			
Female	46 (39.3%)	0.68 (0.36-1.28)		0.74 (0.39-1.43)		0.84 (0.46-1.52)			
Age (years)			0.323				0.378		
<65	8 (6.8%)	1				1			
≥65	109 (93.2%)	0.60 (0.21-1.67)				0.63 (0.23-1.76)			
Adjuvant treatment			0.623				0.751		
No	67 (57.3%)	1				1			
Yes	45 (38.5%)	1.35 (0.74-2.48)				1.25 (0.69-2.31)			
Unknown	5 (4.2%)								
Tobacco			0.715				0.168		0.212
Former	60 (51.3%)	1				1		1	
Current	57 (48.7%)	1.12 (0.62-2.02)				1.51 (0.84-2.70)		1.46 (0.81-2.65)	

¹Events/n = 44/117. ²Events/n = 46/117. ³PI: prognostic index.

Supplementary Table 4. Univariate and multivariate Cox proportional hazards analysis of PI and clinico-pathological parameters for DFS and OS in the NATCH cohort.

Training cohort	MDA n (%)	DFS ¹				OS ²			
		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value
PI³			0.041		0.042		0.024		0.031
Low	64 (66.7%)	1		1		1		1	
High	32 (33.3)	1.97 (1.03-3.76)		2.01 (1.03-3.94)		2.29 (1.12-4.69)		2.29 (1.08-4.86)	
Stage			0.002		0.003		<0.001		0.001
I	57 (59.4%)	1		1		1		1	
II	19 (19.8%)	0.69 (0.24-2.04)	0.505	0.75 (0.25-2.23)	0.608	0.70 (0.20-2.42)	0.568	0.82 (0.23-2.90)	0.649
III	20 (20.8%)	3.10 (1.54-6.26)	0.002	3.10 (1.52-6.33)	0.002	4.00 (1.86-8.60)	<0.001	4.12 (1.88-9.27)	0.002
Gender			0.057		0.022		0.038		0.006
Male	93 (96.9%)	1		1		1		1	
Female	3 (3.1%)	0.25 (0.06-1.04)		5.81 (1.29-26.14)		4.57 (1.09-19.27)		8.80 (1.89-41.04)	
Age (years)			0.911				0.526		
<65	46 (47.9%)	1				1			
≥65	50 (52.1%)	1.03 (0.54-1.99)				0.79 (0.39-1.63)			
Adjuvant treatment			0.660				0.674		
No	44 (45.8%)	1				1			
Yes	52 (54.2%)	0.87 (0.45-1.65)				0.86 (0.42-1.76)			
Tobacco			0.989				0.921		
Never	34 (35.4%)	1				1			
Former	18 (18.8%)	0.98 (0.38-2.48)	0.961			0.82 (0.28-2.40)	0.715		
Current	44 (45.8%)	1.04 (0.50-2.16)	0.918			1.00 (0.45-2.22)	0.995		

¹Events/n = 37/96. ²Events/n = 30/96. ³PI: prognostic index.

Supplementary Table 5. Univariate and multivariate Cox proportional hazards analysis of PI and clinicopathological parameters for DFS and OS in CIBERES-CUN cohort.

Validation cohort	CUN n (%)	DFS ^{1,2}				CIBERES -CUN n (%)	OS ³			
		Univariate analysis		Multivariate analysis			Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value		HR (95% CI)	P value	HR (95% CI)	P value
PI⁴			0.020		0.050			0.055		0.074
Low	28 (66.6%)	1		1		70 (66.7%)	1		1	
High	14 (33.3%)	4.32 (1.25-14.86)		3.55 (0.98-12.87)		35 (33.3%)	1.88 (0.99-3.57)		1.84 (0.94-3.60)	
Stage			0.274		0.289			0.040		0.042
I	20 (47.6%)	1				54 (51.4%)	1		1	
II	11 (26.2%)	4.61 (0.84-25.19)	0.078	4.67 (0.84-25.95)	0.078	32 (30.5%)	2.73 (1.10-5.74)	0.008	2.51 (1.19-5.30)	0.016
III	10 (23.8%)	4.73 (0.92-24.38)	0.063	4.49 (0.86-23.38)	0.075	18 (17.1%)	2.72 (1.17-6.30)	0.020	3.00 (1.28-7.02)	0.011
IV	1 (2.4%)	0.99 (0.00-0.00)	0.988	0.00 (0.00-0.00)	0.991	1 (1.0%)	0.00 (0.00-0.00)	0.975	0.00 (0.00-0.00)	0.976
Gender			0.571					0.328		
Male	40 (95.2%)	1				99 (94.3%)	1			
Female	2 (4.8%)	0.05 (0.00-2044)				6 (57%)	0.37 (0.05-2.70)			
Age (years)			0.153		0.399			0.854		
<65	21 (50%)	1		1		45 (42.9%)	1			
≥65	21 (50%)	2.64 (0.70-9.96)		1.84 (0.45-7.57)		60 (57.1%)	1.06 (0.56-2.00)			
Adjuvant treatment⁵			0.591							
No	27 (64.3%)	1								
Yes	15 (35.7%)	1.38 (0.42-4.54)								
Tobacco⁵			0.430							
Former	29 (69.0%)	1								
Current	13 (31.0%)	0.54 (0.12-2.50)								

¹Events/n = 11/42. ²Complete data available for CUN patients. ³Events/n = 35/105. ⁴PI: prognostic index. ⁵Adjuvant treatment and tobacco history were not collected in the CIBERES cohort of patients

Supplementary Table 6. Prognostic value of PI and stage in univariate and bivariate models for MDA and CIBERES-CUN cohorts.

	MDA cohort				CIBERES-CUN cohort			
	DFS (n=117)		OS (n=117)		DFS (n=42)		OS (n=105)	
	LR	P value	LR	P value	LR	P value	LR	P value
Stage	18.879	<0.001	7.950	0.019	5.238	0.155	9.776	0.021
Stage & PI	37.364	<0.001	14.535	0.002	9.950	0.041	15.592	0.004
LR-Chi ² improvement ¹	18.710	<0.001	6.436	0.011	5.817	0.016	5.945	0.015

¹LR-Chi²: Likelihood ratio/Chi-squared test statistic

Supplementary Table 7. Univariate and multivariate Cox proportional hazards analysis of CPI and clinicopathological parameters for DFS and OS in MDA cohort.

Training cohort	MDA n (%)	DFS ¹				OS ²			
		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value
CPI³		2.72 (1.94-3.81)	<0.001	2.69 (1.92-3.76)	<0.001	1.79 (1.30-2.45)	<0.001	1.77 (1.29-2.44)	<0.001
Gender			0.232		0.385		0.562		
Male	71 (60.7%)	1		1		1			
Female	46 (39.3%)	0.68 (0.36-1.28)		0.75 (0.40-1.43)		0.84 (0.46-1.52)			
Age (years)			0.323				0.378		
<65	8 (6.8%)	1				1			
≥65	109 (93.2%)	0.60 (0.21-1.67)				0.63 (0.23-1.76)			
Adjuvant treatment			0.623				0.751		
No	67 (57.3%)	1				1			
Yes	45 (38.5%)	1.35 (0.74-2.48)				1.25 (0.69-2.31)			
Unknown	5 (4.2%)								
Tobacco			0.715				0.168		0.213
Former	60 (51.3%)	1				1		1	
Current	57 (48.7%)	1.12 (0.62-2.02)				1.51 (0.84-2.70)		1.45 (0.81-2.60)	

¹Events/n = 44/117. ²Events/n = 46/117. ³CPI: combined prognostic index.

Supplementary Table 8. Univariate and multivariate Cox proportional hazards analysis of CPI and clinicopathological parameters for OS in CIBERES-CUN cohort.

Validation cohort	CUN n (%)	DFS ^{1,2}				CIBERES-CUN n (%)	OS ³			
		Univariate analysis		Multivariate analysis			Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value		HR (95% CI)	P value	HR (95% CI)	P value
CPI⁴			0.011		0.012			0.001		0.001
Low	27 (65.8%)	1		1		69 (66.3%)	1		1	
High	14 (34.1%)	5.66 (1.50-21.38)		5.52 (1.46-20.89)		35 (33.7%)	2.91 (1.53-5.54)		2.88 (1.51-5.49)	
Gender			0.564					0.311		0.351
Male	39 (95.1%)	1				97 (94.2%)	1		1	
Female	2 (4.9%)	0.05 (0.00-1718)				6 (5.8%)	0.36 (0.05-2.61)		0.39 (0.05-2.84)	
Age (years)			0.182		0.209			0.860		
<65	20 (48.8%)	1		1		44 (42.7%)	1			
≥65	21 (51.2%)	2.47 (0.65-9.32)		2.35 (0.62-8.88)		59 (57.3%)	1.06 (0.56-2.00)			
Adjuvant treatment			0.495							
No	27 (65.9%)	1								
Yes	14 (34.1%)	1.51 (0.46-4.96)								
Tobacco			0.397							
Former	28 (68.3%)	1								
Current	13 (31.7%)	0.51 (0.11-2.39)								

¹Events/n = 11/41. ²Complete data available for CUN patients. ³Events/n = 39/104. ⁴CPI: combined prognostic index.

Supplementary Table 9. Univariate and multivariate Cox proportional hazards analysis of CPI and clinicopathological parameters for OS in CIBERES-CUN cohort.

Validation cohort	CUN n (%)	DFS ^{1,2}				CIBERES-CUN n (%)	OS ³			
		Univariate analysis		Multivariate analysis			Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value		HR (95% CI)	P value	HR (95% CI)	P value
CPI⁴		2.72 (1.34-5.55)	0.006	2.54 (1.27-5.09)	0.009		1.56 (1.20-2.02)	0.001	1.56 (1.20-2.02)	0.001
Gender			0.564					0.311		0.319
Male	39 (95.1%)	1				97 (94.2%)	1		1	
Female	2 (4.9%)	0.05 (0.00-1718)				6 (5.8%)	0.36 (0.05-2.61)		0.36 (0.05-2.66)	
Age (years)			0.182		0.306			0.860		
<65	20 (48.8%)	1		1		44 (42.7%)	1			
≥65	21 (51.2%)	2.47 (0.65-9.32)		2.02 (0.52-7.80)		59 (57.3%)	1.06 (0.56-2.00)			
Adjuvant treatment			0.495							
No	27 (65.9%)	1								
Yes	14 (34.1%)	1.51 (0.46-4.96)								
Tobacco			0.397							
Former	28 (68.3%)	1								
Current	13 (31.7%)	0.51 (0.11-2.39)								

¹Events/n = 11/41. ²Complete data available for CUN patients. ³Events/n = 39/104. ⁴CPI: combined prognostic index