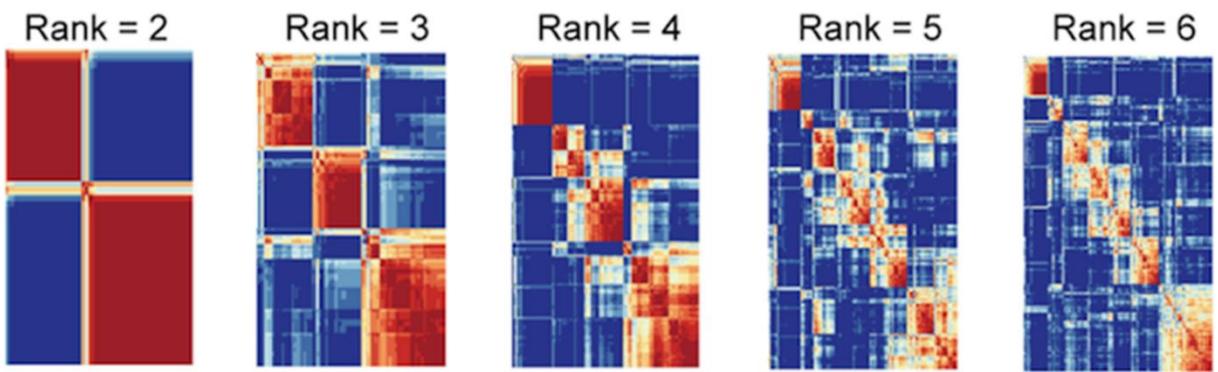


## SUPPLEMENTARY FIGURES



Supplementary Figure 1. Heatmap representation of NMF clustering.

### A GSE72094

Variable	N	Hazard ratio	P
Age	393	1.01 (0.99, 1.03)	0.448
Gender			
Female	219	Reference	
Male	174	1.80 (1.22, 2.67)	0.003
Stage			
I	254	Reference	
II	67	2.05 (1.27, 3.31)	0.003
III	57	3.09 (1.90, 5.03)	< 0.001
IV	15	3.38 (1.53, 7.47)	0.003
Smoking status			
Ever	298	Reference	
Never	30	1.18 (0.50, 2.79)	0.712
Unknown	65	1.47 (0.69, 2.43)	0.128
EGFR status			
MUT	41	Reference	
WT	352	2.70 (0.95, 7.62)	0.061
KRAS status			
MUT	138	Reference	
WT	255	0.84 (0.56, 1.25)	0.389
TP53 status			
MUT	95	Reference	
WT	298	1.23 (0.78, 1.94)	0.372
STK11 status			
MUT	64	Reference	
WT	329	1.27 (0.75, 2.16)	0.368
subtype			
Low risk	202	Reference	
High risk	191	2.08 (1.38, 3.15)	< 0.001

### B GSE68465

Variable	N	Hazard ratio	P
Age	433	1.03 (1.01, 1.04)	<0.001
Gender			
Female	214	Reference	
Male	219	1.32 (0.97, 1.79)	0.08
Grade			
G1	60	Reference	
G2	208	0.79 (0.47, 1.33)	0.37
G3	165	0.85 (0.48, 1.48)	0.56
Smoking status			
Currently smoking	31	Reference	
Never smoked	48	0.70 (0.35, 1.42)	0.32
Smoked in the past	264	0.73 (0.44, 1.23)	0.24
Unknown	90	0.63 (0.35, 1.12)	0.12
subtype			
Low risk	195	Reference	
High risk	238	2.45 (1.70, 3.54)	< 0.001

### C GSE50081

Variable	N	Hazard ratio	P
Age	127	1.02 (0.99, 1.05)	0.260
Gender			
Female	62	Reference	
Male	65	1.39 (0.78, 2.48)	0.259
Stage			
I	92	Reference	
II	35	2.44 (1.37, 4.34)	0.002
Smoking status			
Current	36	Reference	
Ever	56	0.81 (0.42, 1.57)	0.527
Never	23	0.78 (0.31, 1.95)	0.594
Unknown	12	1.14 (0.41, 3.21)	0.802
Subtype			
Low risk	77	Reference	
High risk	50	2.09 (1.19, 3.67)	0.010

### D GSE42127

Variable	N	Hazard ratio	P
Age	132	1.02 (0.99, 1.06)	0.243
Gender			
Female	65	Reference	
Male	67	1.17 (0.59, 2.34)	0.656
Stage			
I	89	Reference	
II	22	1.53 (0.73, 3.23)	0.260
III	20	1.56 (0.66, 3.67)	0.312
IV	1	26.78 (2.50, 286.87)	0.007
subtype			
Low risk	79	Reference	
High risk	53	2.44 (1.25, 4.74)	0.009

### E GSE41271

Variable	N	Hazard ratio	P
Age	181	1.01 (0.99, 1.04)	0.349
Gender			
Female	90	Reference	
Male	91	1.48 (0.86, 2.52)	0.155
Stage			
I	95	Reference	
II	27	1.32 (0.65, 2.65)	0.439
III	43	2.50 (1.42, 4.41)	0.002
IV	16	3.26 (1.39, 7.62)	0.006
Smoking Status			
Ever	156	Reference	
Never	25	1.17 (0.49, 2.78)	0.719
subtype			
Low risk	101	Reference	
High risk	80	1.79 (1.06, 3.03)	0.028

### F GSE31210

Variable	N	Hazard ratio	P
Age	226	1.04 (0.99, 1.09)	0.148
Gender			
Female	121	Reference	
Male	105	0.92 (0.34, 2.54)	0.878
Stage			
I	168	Reference	
II	58	3.26 (1.62, 6.54)	< 0.001
Smoking status			
Ever	111	Reference	
Never	115	0.95 (0.34, 2.67)	0.922
Subtype			
Low risk	149	Reference	
High risk	77	3.44 (1.62, 7.30)	0.001

### G GSE30219

Variable	N	Hazard ratio	P
Age	85	1.04 (1.01, 1.09)	0.025
Gender			
Male	66	Reference	
Female	19	1.44 (0.64, 3.23)	0.380
Tumor size			
T1	71	Reference	
T2	12	1.45 (0.64, 3.28)	0.376
T3	2	0.55 (0.05, 5.86)	0.622
Node status			
N0	82	Reference	
N1	3	1.33 (0.23, 7.52)	0.749
subtype			
Low risk	41	Reference	
High risk	44	2.82 (1.31, 6.05)	0.008

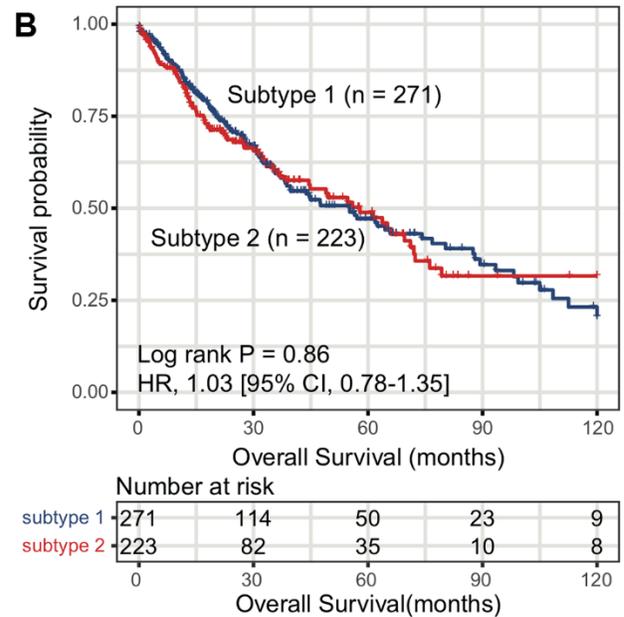
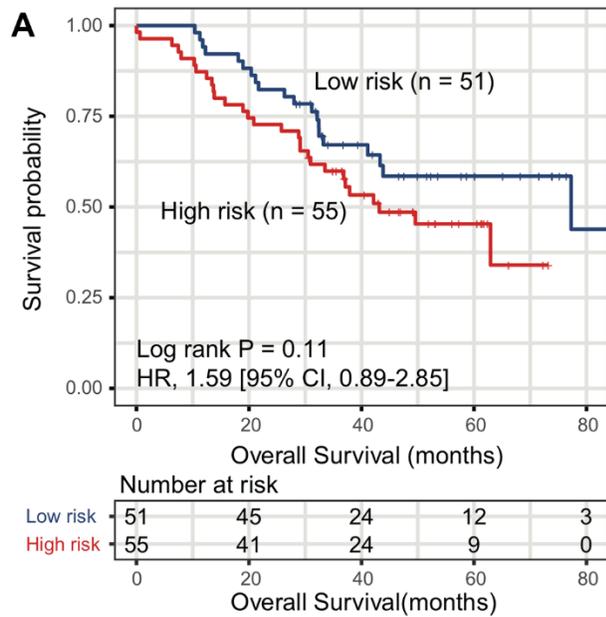
### H GSE13213

Variable	N	Hazard ratio	P
Age	116	1.01 (0.98, 1.04)	0.47
Gender			
Male	60	Reference	
Female	56	0.59 (0.25, 1.37)	0.22
Stage			
I	78	Reference	
II	13	1.35 (0.54, 3.41)	0.52
III	25	3.01 (1.59, 5.68)	< 0.001
Smoking Status			
Ever	61	Reference	
Never	55	1.31 (0.57, 3.04)	0.53
EGFR status			
MUT	45	Reference	
WT	71	0.67 (0.34, 1.31)	0.24
KRAS Status			
MUT	15	Reference	
WT	101	0.83 (0.35, 1.94)	0.66
TP53 Status			
MUT	38	Reference	
WT	78	0.97 (0.51, 1.81)	0.92
subtype			
Low risk	64	Reference	
High risk	52	2.00 (1.02, 3.89)	0.04

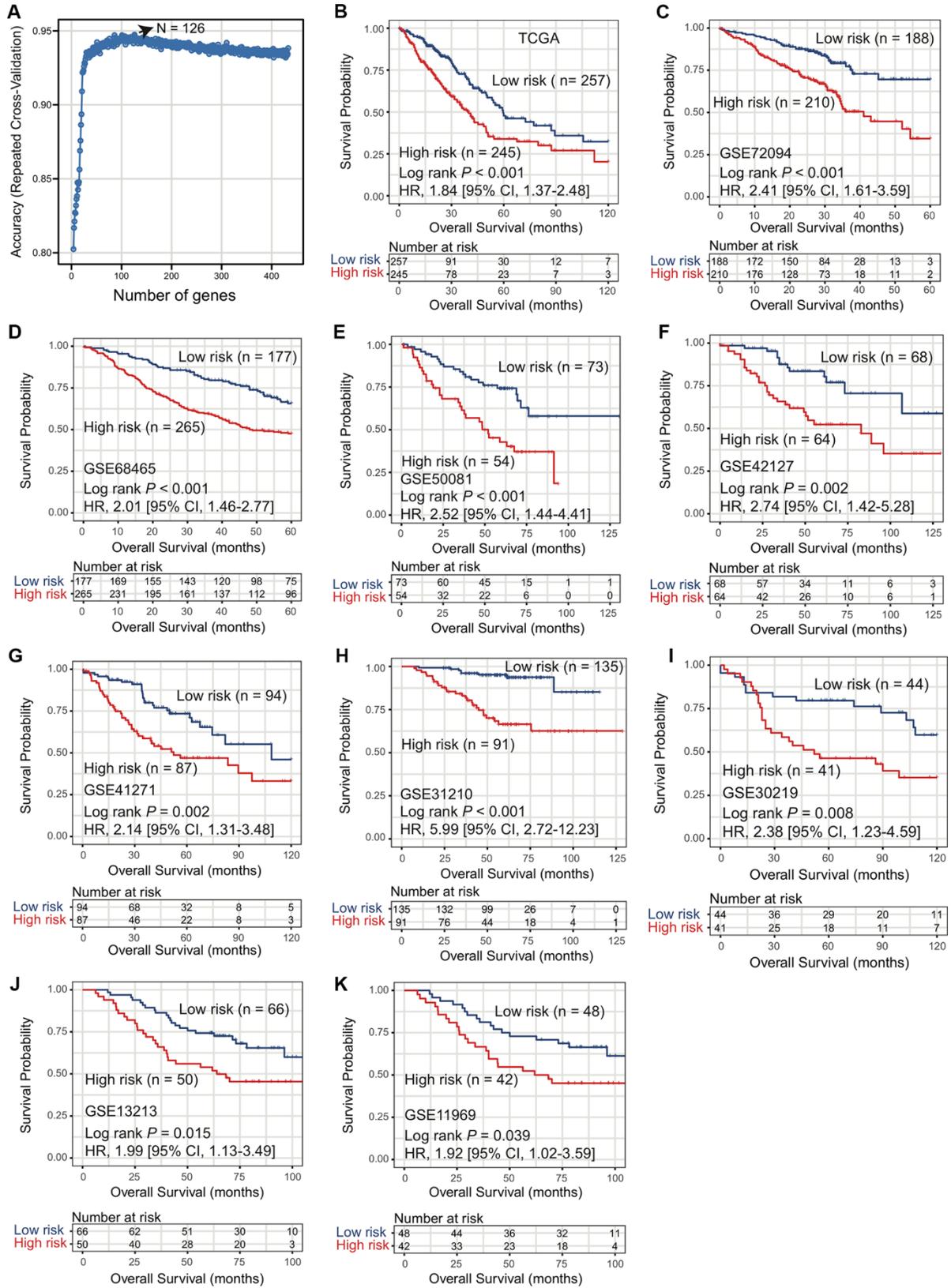
### I GSE11969

Variable	N	Hazard ratio	P
Age	90	1.01 (0.97, 1.04)	0.758
Gender			
Female	43	Reference	
Male	47	1.44 (0.57, 3.65)	0.446
Stage			
I	52	Reference	
II	13	1.50 (0.55, 4.08)	0.424
III	25	3.35 (1.61, 6.97)	0.001
Grade			
G1	25	Reference	
G2	31	0.89 (0.34, 2.33)	0.816
G3	34	1.69 (0.62, 4.57)	0.303
Smoking status			
Ever	45	Reference	
Never	45	1.15 (0.46, 2.86)	0.766
EGFR status			
MUT	32	Reference	
WT	58	0.53 (0.24, 1.15)	0.109
KRAS Status			
MUT	10	Reference	
WT	80	0.61 (0.22, 1.67)	0.338
TP53 Status			
MUT	29	Reference	
WT	61	0.99 (0.46, 2.14)	0.976
subtype			
Low risk	50	Reference	
High risk	40	1.62 (0.74, 3.55)	0.228

**Supplementary Figure 2.** Prognostic significance of high-risk versus low-risk subtypes using multivariate Cox model in 9 validation cohorts of (A) GSE72094, (B) GSE68465, (C) GSE50081, (D) GSE42127, (E) GSE41271, (F) GSE31210, (G) GSE30219, (H) GSE13213, and (I) GSE11969

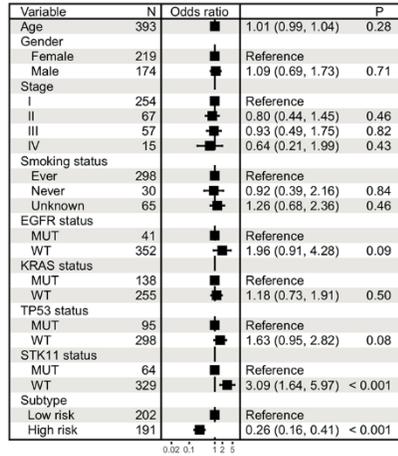


**Supplementary Figure 3.** (A) Kaplan-Meier plot of identified 2 subtypes in GSE81089 dataset. (B) Kaplan-Meier plot of identified TCGA LUSC subtypes.

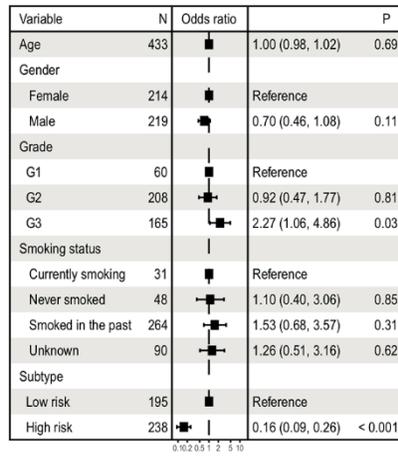


**Supplementary Figure 4.** (A) Relation between classification accuracy and selected genes via recursive feature elimination algorithm. (B–K) Kaplan-Meier plot of identified subtypes using 126 genes in TCGA and 9 validation cohorts.

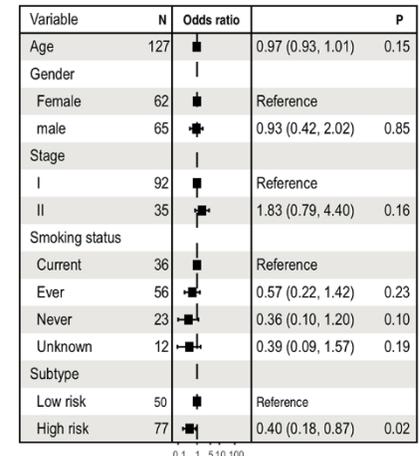
### A GSE72094



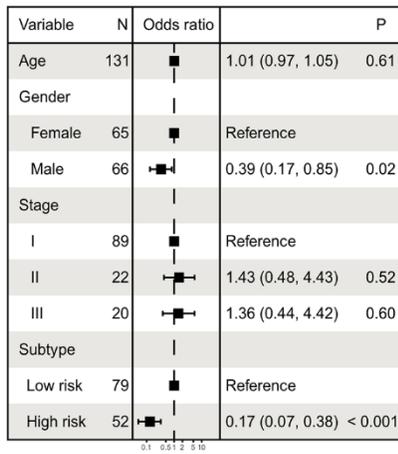
### B GSE68465



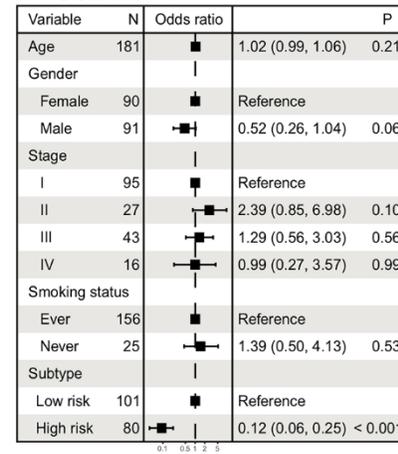
### C GSE50081



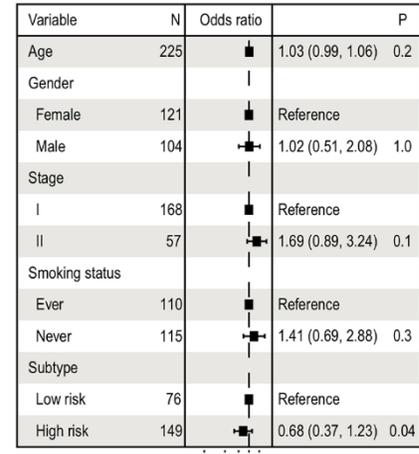
### D GSE42127



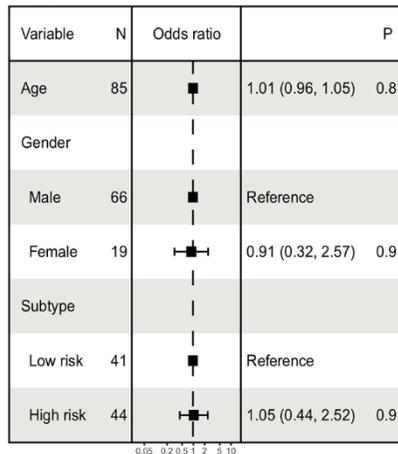
### E GSE41271



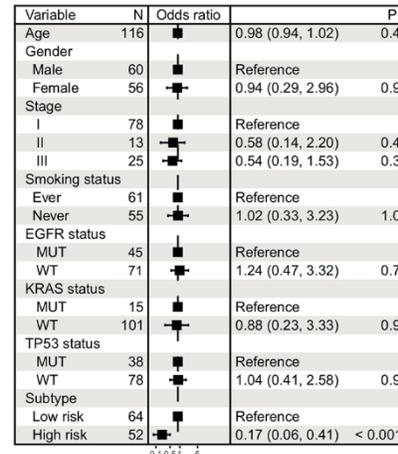
### F GSE31210



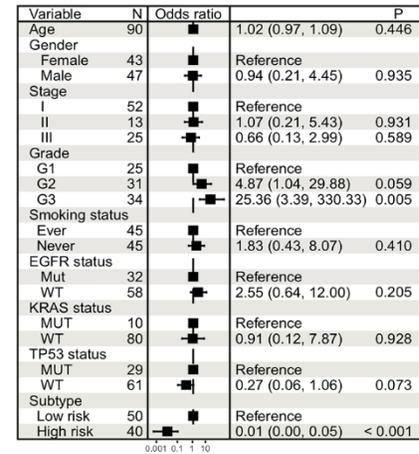
### G GSE30219



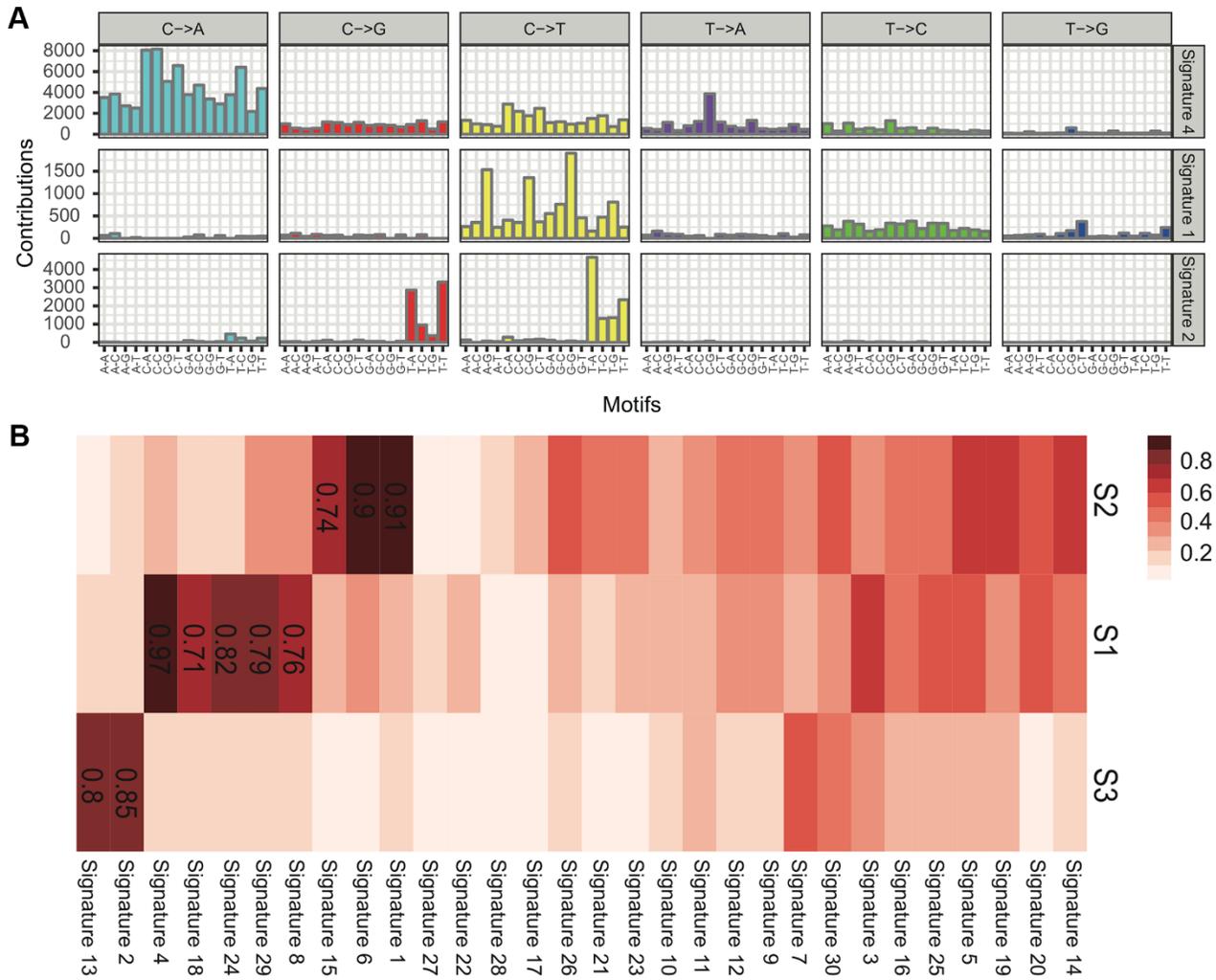
### H GSE13213



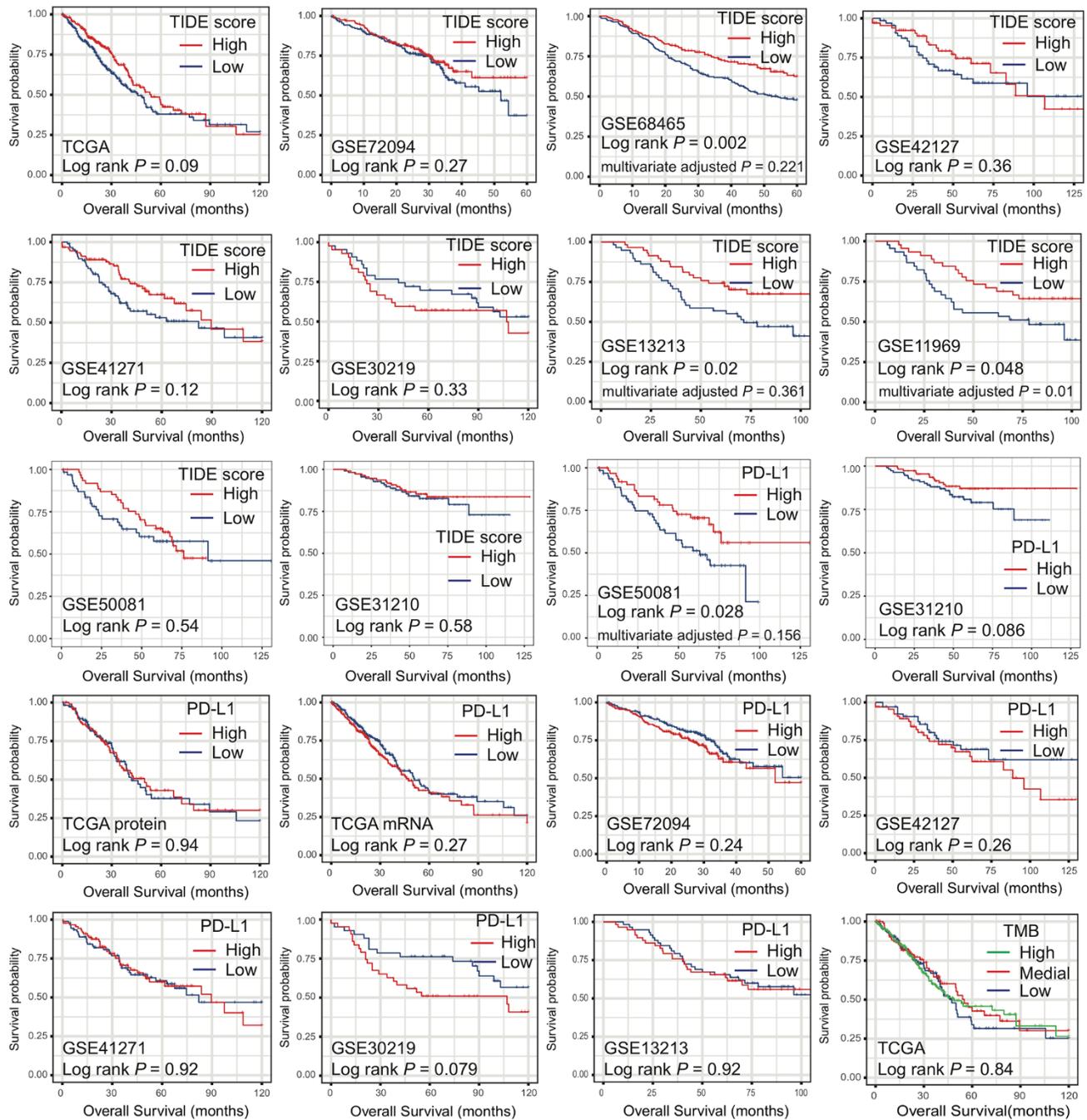
### I GSE11969



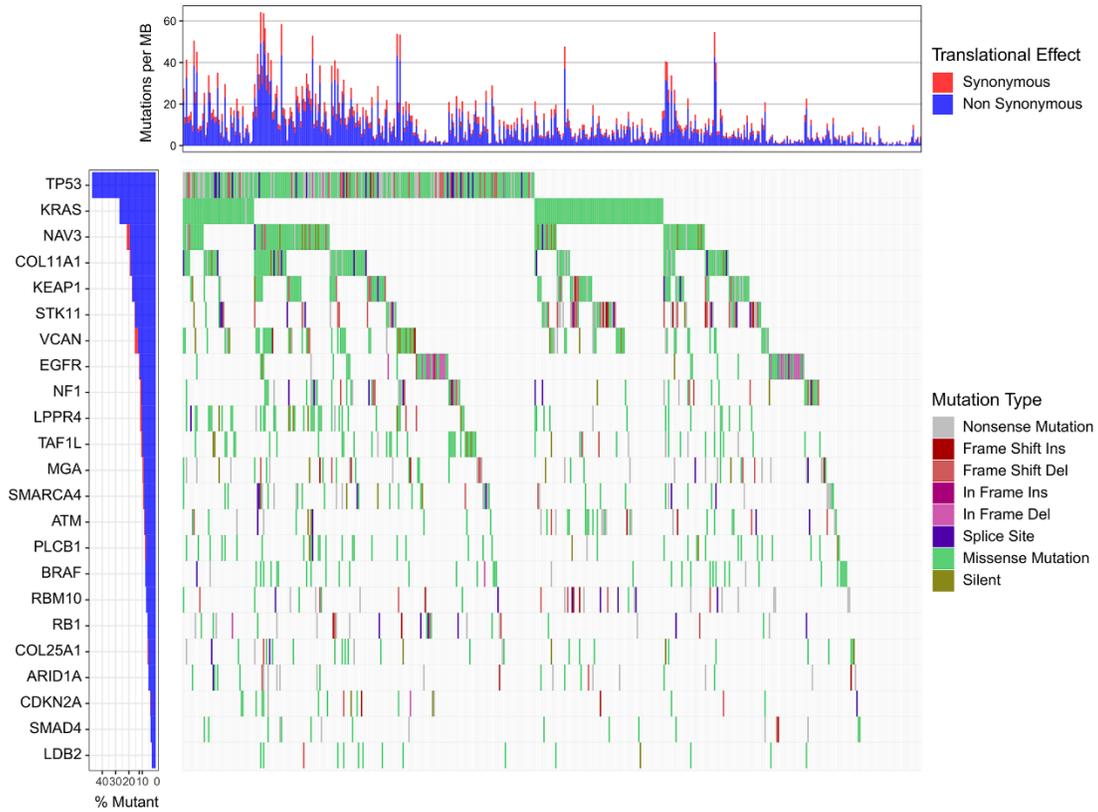
**Supplementary Figure 5.** Association of identified LUAD subtypes with TIDE score using multivariate logistic analysis in 9 validation cohorts of (A) GSE72094, (B) GSE68465, (C) GSE50081, (D) GSE42127, (E) GSE41271, (F) GSE31210, (G) GSE30219, (H) GSE13213, and (I) GSE11969.



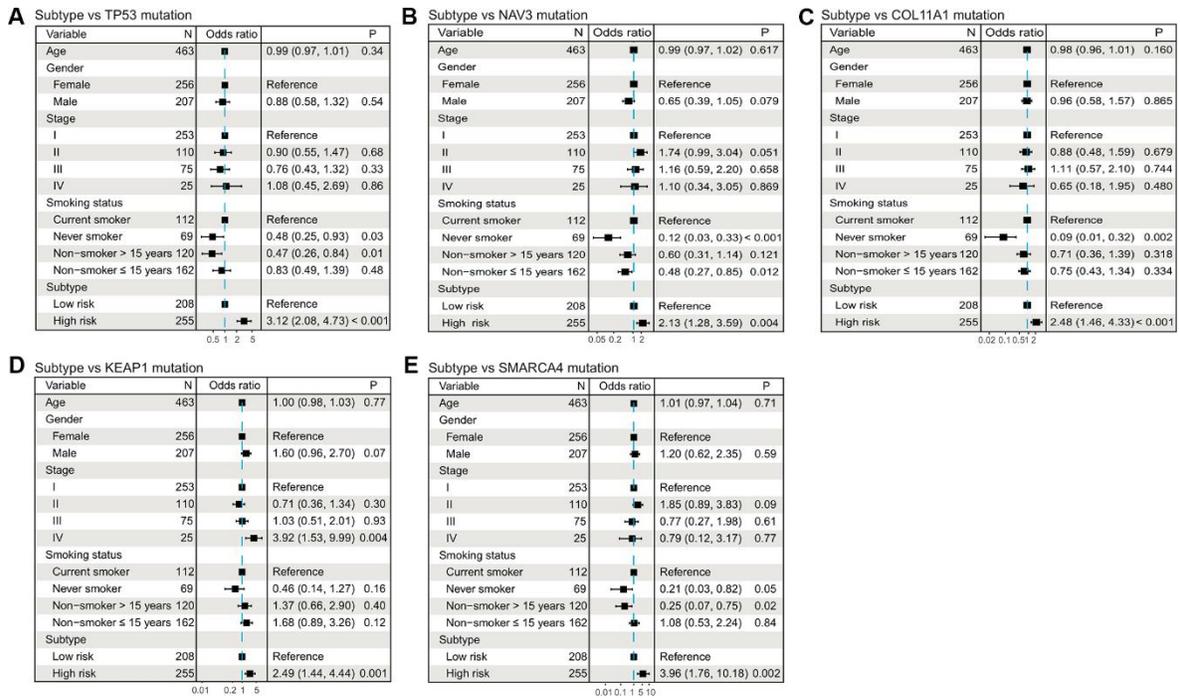
**Supplementary Figure 6.** (A) Mutational signatures extracted from TCGA LUAD cohort and (B) their cosine similarity with COSMIC mutational signatures.



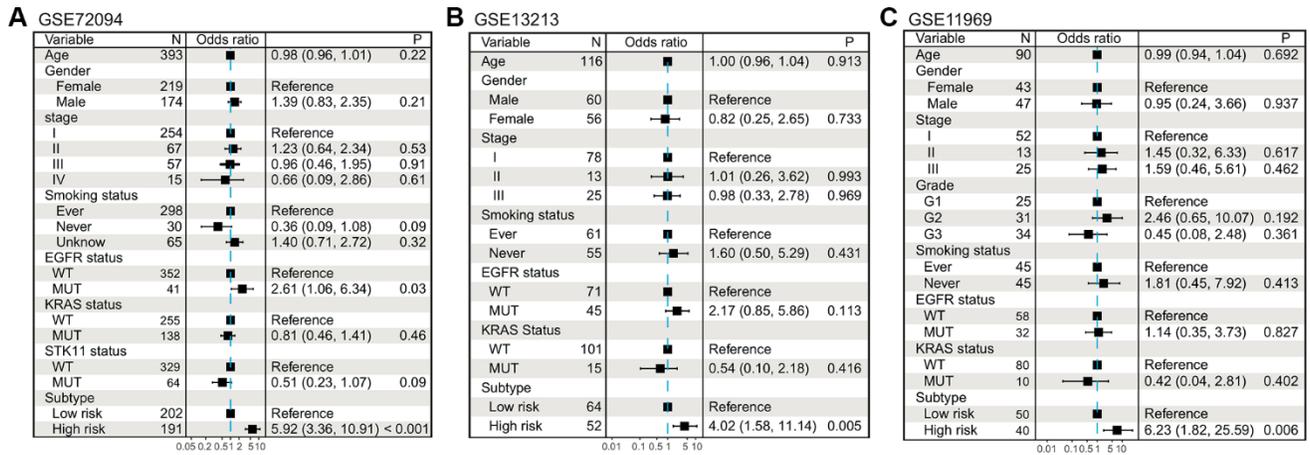
Supplementary Figure 7. Kaplan-Meier plots with respect to TIDE score, *PD-L1* expression, and TMB in TCGA and 9 validation cohorts.



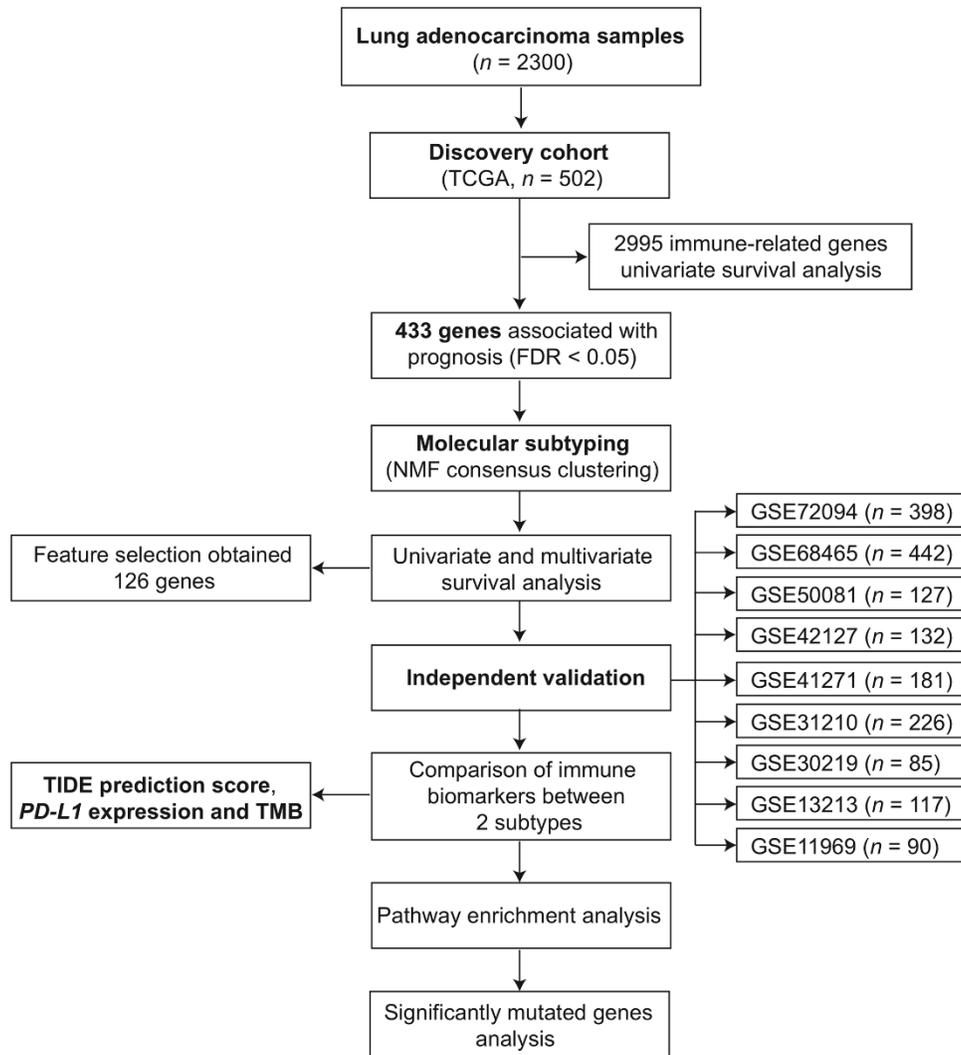
Supplementary Figure 8. Mutational landscape of SMGs in TCGA LUAD cohort.



Supplementary Figure 9. Associations between mutations in *TP53*, *NAV3*, *CLO11A1*, *KEAP1* and *SMARCA4* and identified 2 LUAD subtypes using multivariate logistic analysis.



**Supplementary Figure 10.** Associations of *TP53* mutation with high-risk subtype in (A) GSE72094, (B) GSE13213 and (C) GSE11969.



**Supplementary Figure 11.** Flow chart of our study. TCGA and 9 public LUAD cohorts containing 2300 samples were included to perform relevant analyses.