Supplemental material

Table S1. Basic information of gene sets included in the study.

Gene sets	Gene components
SCCup	CDCA5, DSC3, DSG3, EPT1, KRT14, KRT5, OSGEP, PARD3B, PSMA8, PVRL1, SFN, SLC2A1, TOP2A, TPX2
SCCdown	ABCA8, ADAMTS8, ADAMTSL3, ADARB1, AGER, AKAP2, CAV1, NR2F1, PGM5, PHACTR2, RASL12, SRCRB4D, ST6GALNAC5, VWF
maleaffy	USP9Y, ZFY, CYorf14, EIF1AY, CYorf15B, DDX3Y, JARID1D, UTY, CYorf15A, PRKY, CD24L4
femaleaffy	XIST, AZI2, EREG, GPM6B
KEGG_MISMATCH_REPAIR	MLH3, POLD1, MLH1, POLD2, RFC1, MSH2, RFC3, RFC2, MSH3, POLD4, PMS2, RFC4, LIG1, RFC5, RPA1, MSH6, RPA3, POLD3, RPA2, PCNA, SSBP1, RPA4, EXO1
MITOSIS	KIF22, KIF25, NBN, KNTC1, PKMYT1, TTK, AURKA, CDC16, TTN, TGFB1, KIF2C, TRIAP1, DDX11, CDKN2B, PCBP4, TARDBP, PRMT5, TGFA, CCNA2, CDCA5, ZW10, KIF11, ANAPC5, RAN, KIF15, RINT1, ANAPC4, TPX2, NUSAP1, ESPL1, UBE2C, DCTN3, DCTN2, CHMP1A, MAD2L1, EREG, ZWINT, BUB1B, CLIP1, AKAP8, MAD2L2, CHFR, NEK6, RAD17, PPP5C, PAM, NEK2, PML, CETN1, ANLN, ANAPC10, ANAPC11, BRSK1, RCC1, PIN1, NUMA1, NCAPH, NPM2, BUB1, PBRM1, CDK13, EGF, GML, CD28, SSSCA1, PDS5B, CDC23, CENPE, NDC80, BIRC5, SUGT1, CDC25C, CDC27, SMC3, ATM, CDC25B, SMC4, NOLC1, PLK1, EPGN, SMC1A, CIT
GO_INNER_CELL_MASS_CELL_PROLIFERATION	BRCA2, GINS1, GINS4, NCAPG2, NDEL1, PALB2, PRPF19, SALL4, SBDS, TAF8, ZNF259
INTERFERON_GAMMA_PRODUCTION	IL27, CEBPG, IL18, CD276, TLR3, INHA, FOXP3, TLR7, TLR8, TLR9, INHBA, IL12A, IL12B, EBI3
CD8+ T	TRBC1, TRBV19, TRBV5-4, TRBV21-1, CCL5, CD3D, TRBV3-1, LOC647353, LCK, CD8A, IL7R, PRKCH, KLRK1

Table S2. Clinicopathological information of included cases

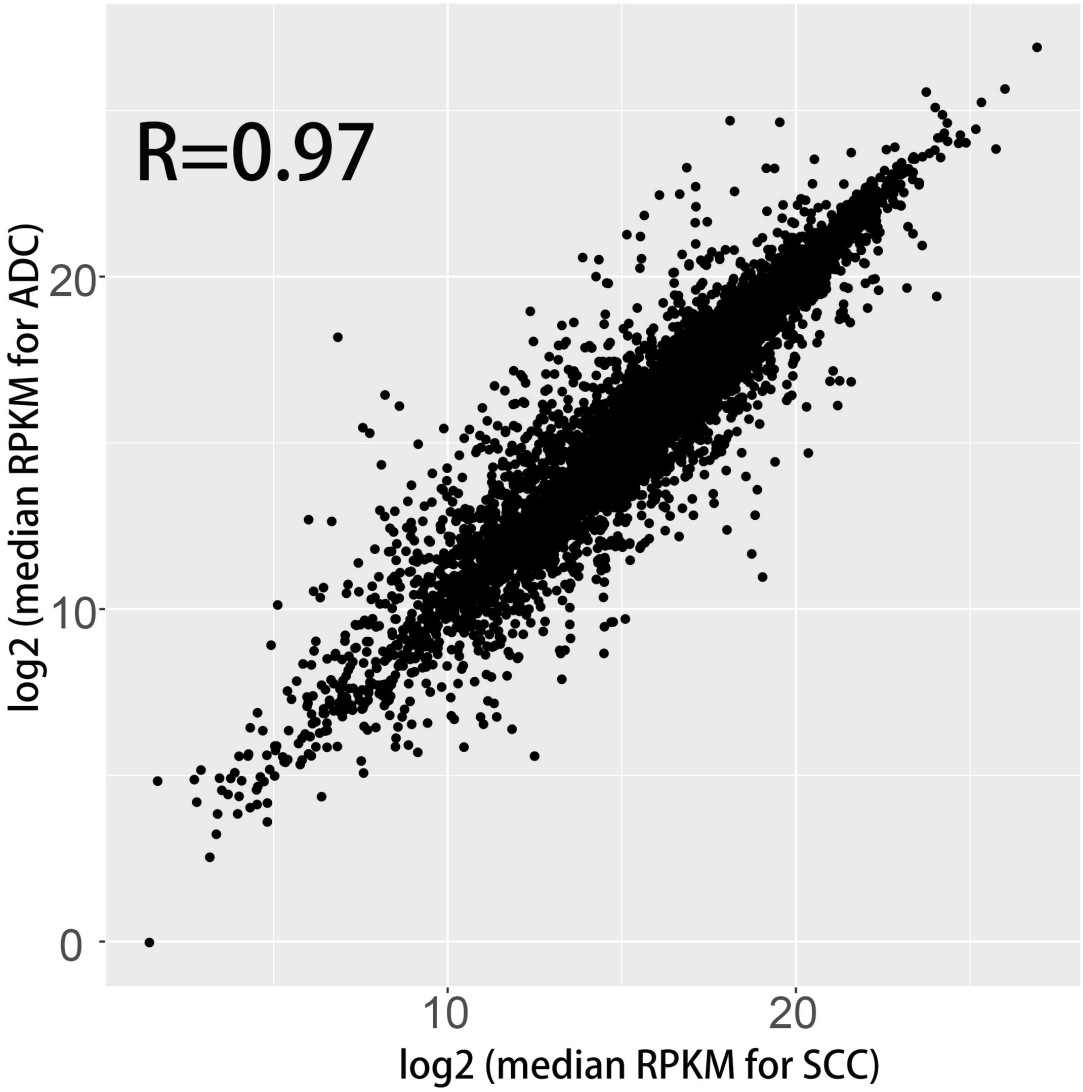
	Subgroups	No.
Age	Median(sd)	67 (9.4)
	<=65 years	427
	>65 years	558
	NA	28
Gender	Male	607
	Female	406
Histology	ADC	513
	SCC	500
TNM	Ⅰ	517
	Ⅱ	283
	Ⅲ	168
	Ⅳ	33
	NA	12
EGFR	Driver mutant	69
	Wild type	886
	Passenger mutant	20
KRAS	Driver mutant	171
	Wild type	801
	Passenger mutant	3
Smoking	Smokers	774
	Non-smokers	239
	Unknown	3
Median follow-up time (months)		28.633
status	Alive	653
	Dead	340
	NA	20

Abbreviations： ADC, adenocarcinoma; SCC, squamous cell carcinoma; EGFR, epidermal growth factor receptor; KRAS, Kirsten rat sarcoma viral oncogene; sd, standard deviation; NA, not available.

Table S3. Clinicopathological information of included cases

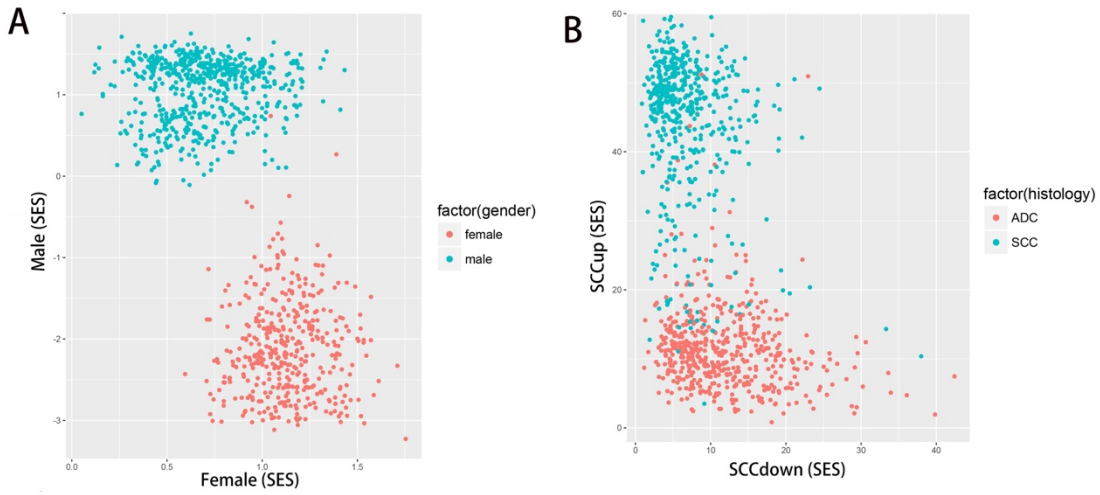
	Subgroups	No.
Age	Median(sd)	64.56(10.0)
	<=65 years	139
	>65 years	136
	NA	0
Gender	Male	148
	Female	127
	NA	0
Histology	ADC	183
	SCC	80
	others	12
	NA	0
TNM	Ⅰ	133
	Ⅱ	50
	Ⅲ	86
	Ⅳ	6
	NA	0
Smoking	Smokers	272
	Non-smokers	3
	Unknown	0
Median follow-up time (months)		36.2
status	Alive	151
	Dead	124
	NA	0

Abbreviations： ADC, adenocarcinoma; SCC, squamous cell carcinoma; sd, standard deviation; NA, not available.



Supplementary Figure 1. compatibility for merging LUAD and LUSC datasets.

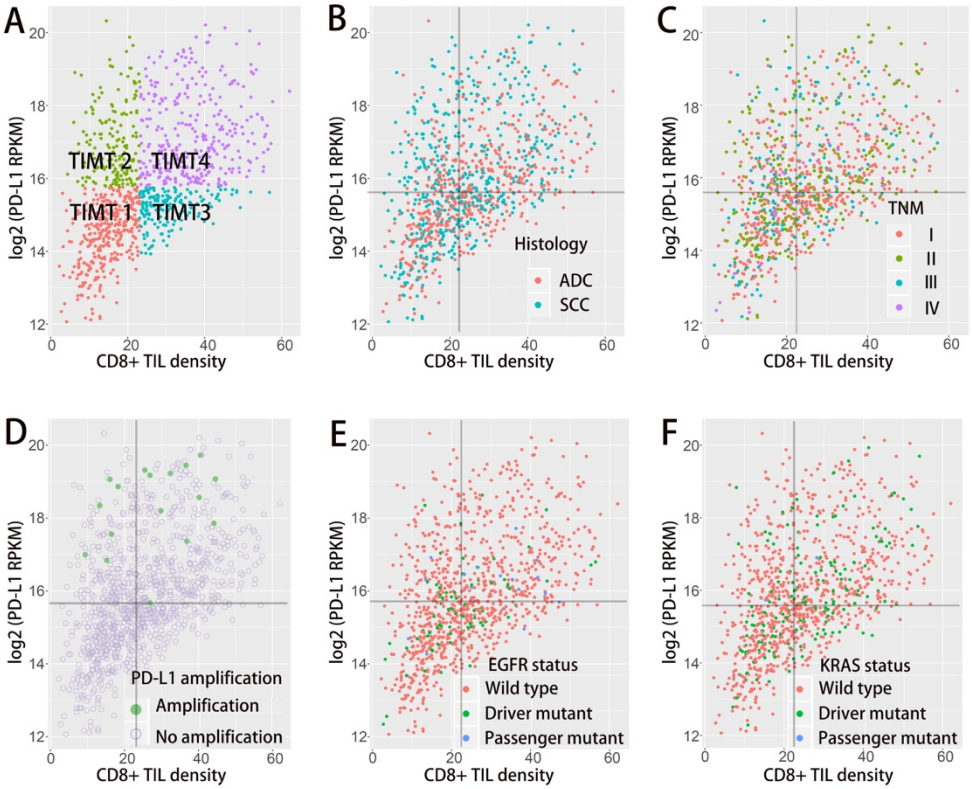
The median RPKM value of each of the 24991 gene symbol was computed for LUAD dataset and LUSC dataset respectively. Each spot on the scatterplot stands for a single gene, with x-axis value and y-axis stand for its log2 transferred median RPKM value in LUSC and LUAD respectively. Correlation of the overall gene expression in the two datasets was computed through Pearson correlation, with the coefficient R=0.97, which indicates that the two datasets are homogeneous and can be merged for further analysis.



Supplementary Figure 2. Validation of SES by distribution analysis.

A Plots of SES for the female gene set vs male gene set (A) and SCC down expressing gene set vs SCC up expressing gene set (B) in annotated subtypes of NSCLC samples: female (red dots), male (blue dots) and ADC (red dots), SCC (blue dots).

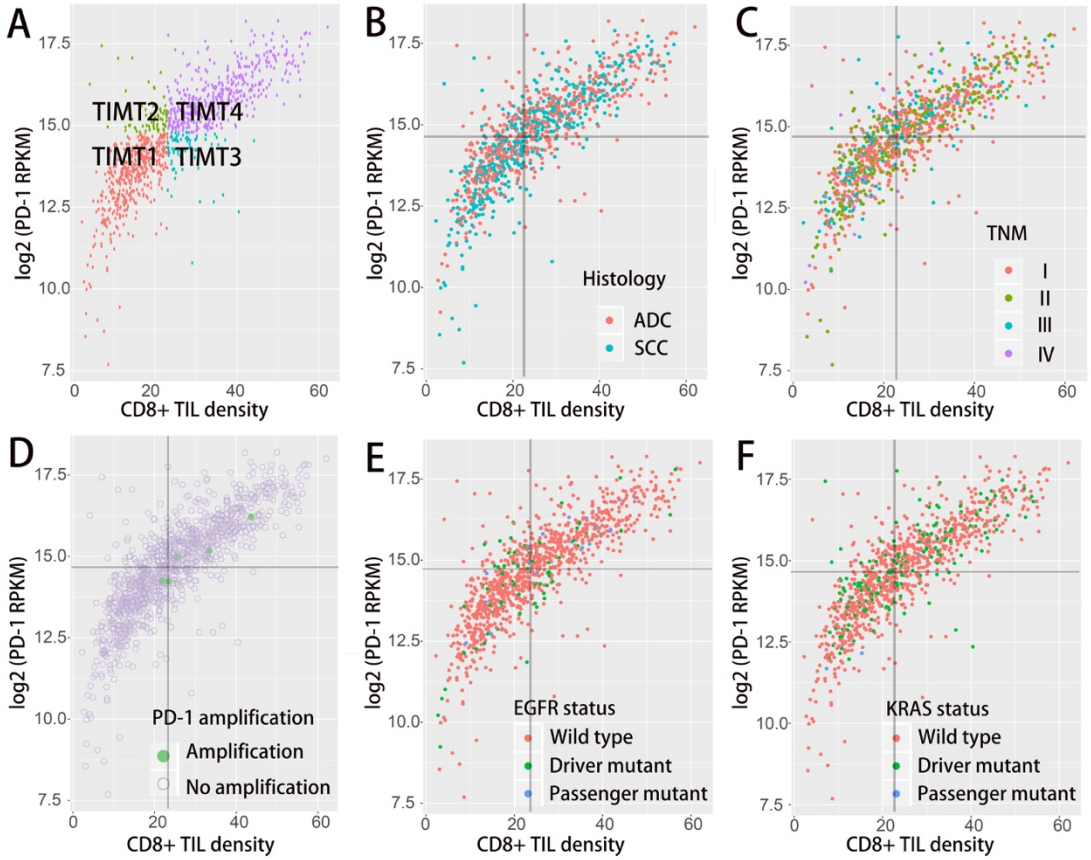
Abbreviations: ADC, adenocarcinoma; SCC, squamous cell carcinoma.



Supplementary Figure 3. Distribution of PD-L1 and CD8+ TIL density expression in NSCLC subgroups.

Tumor microenvironment immune type (TMITs) determined by median levels of PD-L1 expression and CD8+ TIL density.

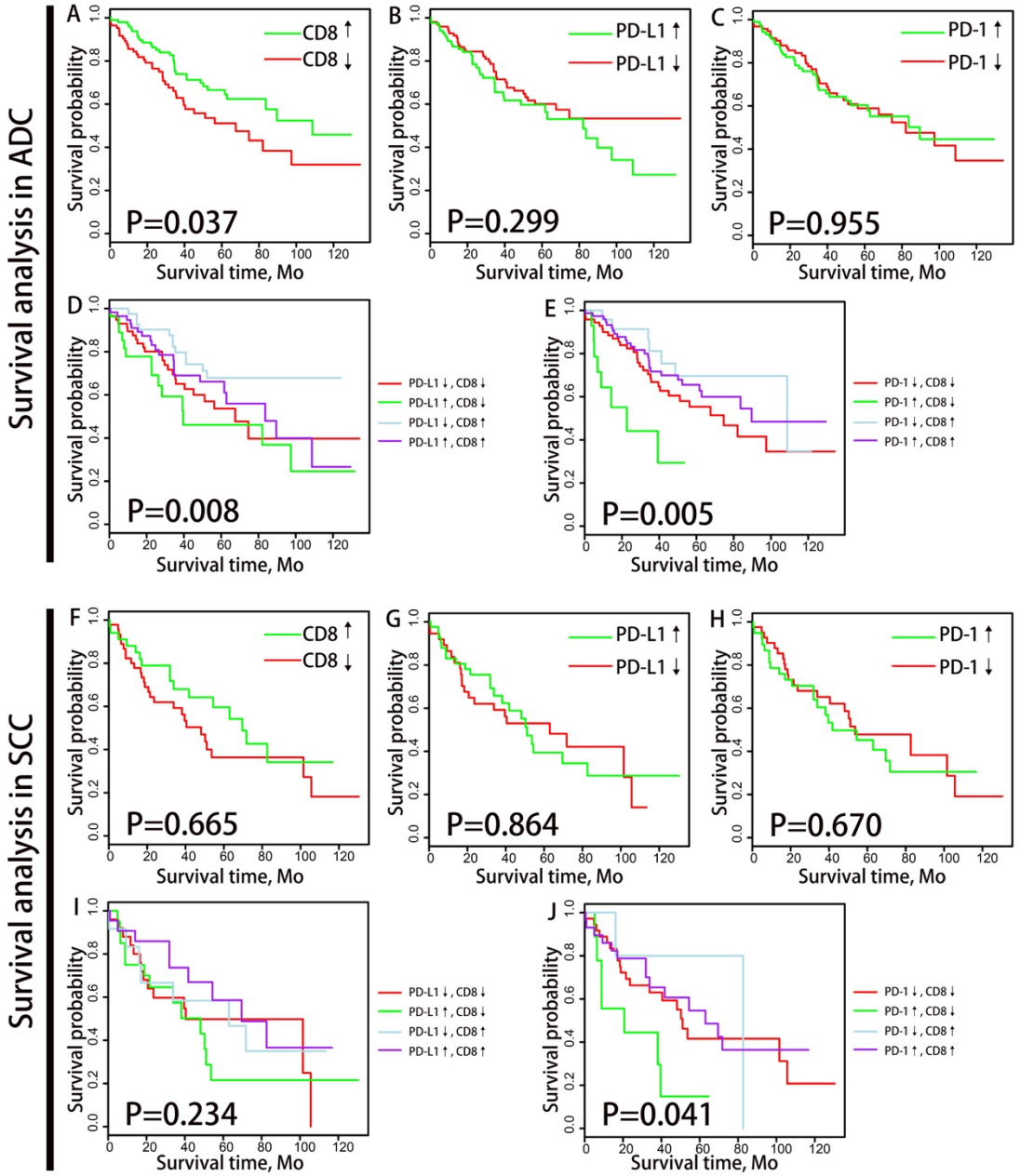
Abbreviations: PD-1, programmed cell death ligand-1; ADC, adenocarcinoma; SCC, squamous cell carcinoma; EGFR, epidermal growth factor receptor; KRAS, Kirsten rat sarcoma viral oncogene; RPKM, reads per kilobase of transcript per million reads mapped.



Supplementary Figure 4. Distribution of PD-1 and CD8+ TIL density expression in NSCLC subgroups.

Tumor microenvironment immune type (TMITs) determined by median levels of PD-L1 expression and CD8+ TIL density.

Abbreviations: PD-1, programmed cell death-1; ADC, adenocarcinoma; SCC, squamous cell carcinoma; EGFR, epidermal growth factor receptor; KRAS, Kirsten rat sarcoma viral oncogene; RPKM, reads per kilobase of transcript per million reads mapped.



Supplementary Figure 4. Survival analysis in validation cohort according to histologic subtypes.

Survival analysis in validation cohort were performed for ADC (A-E) and SCC (F-J) respectively. Kaplan-Meier plots of overall survival according to CD8+ TILs density and PD-1/PD-L1 expression level are graphed (A-C, F-H). PD-1/PD-L1↓ and PD-1/PD-L1↑ refer to tumors with PD-1/PD-L1 mRNA expression value less or more than the median respectively. CD8↓ and CD8↑ refer to tumors with CD8+ TILs enrichment score more or less than the median respectively. Tumors were classified into four TIMTs based on the combination of CD8+ TIL and PD-1/PD-L1, with median values as cutoff points. Kaplan-Meier plots of overall survival according to TIMTs are graphed (D, E, I, J).