Supplemental Online Content

Rakaee M, Adib E, Ricciuti B, et al. Association of machine learning–based assessment of tumor-infiltrating lymphocytes on standard histologic images with outcomes of immunotherapy in patients with NSCLC. *JAMA Oncol.* Published online November 10, 2022. doi:10.1001/jamaoncol.2022.4933

responders versus non-responders after ICI treatment in overall discovery cohort and PD-L1 subsets. 95% CI in parentheses

This supplemental material has been provided by the authors to give readers additional information about their work.

eSupplementary method

Inclusion time

All NSCLC patients treated with ICI agents were included at each institution for the following time intervals: August 2014 to May 2019 at the Dana-Farber Cancer Institute (DFCI), February 2014 to August 2021 at the Imperial College of London (ICL), and January 2015 to September 2021 for the Amsterdam University Medical Center (UMC) cohort.

Ethical clearance

Clinicopathological data and clinical endpoints were retrieved from each patient's chart by experienced oncologists, who were blinded to assessment of TIL levels by ML. This study was approved by the institutional review board at each center (DFCI: DF/HCC #02-180; ICL: 17/WA/0161/R18009; UMC: U2017.003).

Tumor genomic profiling and TMB assessment

In the discovery cohort, DNA sequencing was performed using the OncoPanel massively parallel sequencing assay. Briefly, pathologists identify an area of interest on a H&E stained slide, that is relatively tumor rich and free of artifacts; this region is then collected from parallel unstained slides, and used for targeted capture sequencing. Three successive versions of OncoPanel have been used, covering 275, 300, and 447 cancer-related genes [1]. Tumor mutational burden (TMB) was defined as the number of somatic, non-synonymous, single nucleotide variants and small indel mutations per megabase of genome sequenced [2]. In the ICL and UMC validation cohorts, mutation analysis was performed using a custom-built QIAseq targeted DNA panel (TruSeq Amplicon - Cancer Panel; TSACP) based on multiplex PCR targeted enrichment technology comprising 207 amplicons of 50 oncogenes frequently mutated in solid tumors. The exons assessed for KRAS and EGFR were: 2-4; 3,7,15,18-21, respectively.

Tumor cellularity

In the DFCI cohort, tumor cellularity was visually estimated by a pathologist for the region of interest on H&E slides that guided the collection of DNA from parallel unstained slides. The tumor cellularity was defined as the percentage of tumor epithelial cells relative to other cells

including stromal, inflammatory and normal epithelial cells. For some cases the visual tumor cellularity estimate was adjusted based on the variant allele frequency of driver mutations in the tumor [3]. The tumor cellularity data used in this study was retrieved from Oncopanel sequencing reports.

PD-L1 expression

The PD-L1 tumor proportion score (TPS) was determined using immunohistochemistry as the percent of PD-L1 positive tumor cells relative to all tumor cells present on the slide in the area of interest, irrespective of staining intensity. At least 100 viable tumor cells were required for evaluation of PD-L1 expression. As part of routine clinical care, following in-house validated and IVD PD-L1 assays were used for each cohort: DFCI (E1L3N, Cell Signaling Technology; 22C3, Dako), ICL (22C3, Dako; SP263, Roche) and UMC (22C3, Dako; 28-8, Agilent). PD-L1 scoring was available for 527 of the 685 patient samples.

Quantification of TILs

H&E slides were digitalized using Aperio ScanScope AT (0.49 microns/pixel, Leica Biosystems, Germany) for the discovery cohort, and Pannoramic 250 Flash III (0.24 microns/pixel, 3DHistech, Hungary) for the validation cohort. Supervised machine learning algorithms (QuPath v.0.2.3, Queen's University, Belfast, Northern Ireland) were sequentially employed to build an automated TIL scoring model, based on the parameters we have previously described [4], and after adjustment for current cohort setting, in the following order: 1) Color deconvolution to estimate the stain vectors and to normalize the RGB channels per slide, as H&E intensity varied on different slides. 2) Watershed segmentation to identify cells based on size, shape, and optical density (OD) of nuclei in the hematoxylin layer (calculating 33 features for each cell) [5]. Set up parameters for hematoxylin OD: pixel size: 0.5 μm, background radius: 10 μm, median filter radius: 1 μm, sigma: 1.5 μm, minimum area: 7 μm2 , maximum area:500 μm2, intensity threshold: 0.1, background intensity: 2, cell Expansion: 2 μm, watershed post process + include nuclei + smooth boundaries + make measurements: true. 3) Adding intensity and smoothed object features, calculating Haralick texture features (haralick distance: 1, haralick bins: 32) and gaussian-weighted averages per cell. 4)Training of an object-based classifier. For this step, thoracic pathologist annotation-guided random decision forest cell classifiers were trained to

retain TILs, tumor and stroma cells. TILs were defined as mononuclear immune cells including lymphocytes and plasma cells [6,7]. Seven distinct cell classifiers were trained, one for each tissue type. The training set consisted of 1/3 of each tissue type population (lung = 75/224, lymph node = 20/60, pleura = 14/41, brain = 12/37, liver = 9/27, soft tissue = 9/27, other tissue =10/30). For building the training images for each tissue type, multiple (lung: 2-3; other sites: 3-5) random ROIs (0.25 mm2) were harvested from each case. After building the training images, as starting point, cell labeling process were initiated in the training images, and as more cells were annotated and curated, the model improved, resulting in a decrease in the number of misclassifications. This step continued until the classifiers achieved pathologist-level performance. The accuracy of the final locked classifiers were verified by second observer (pathologist). The final classifiers were deployed on the whole-slide images for each tissue set, followed by manual review of each image (training set) for classification accuracy. Later, the locked classifiers (per tissue sites) were run on the rest of the cohort. Quality control for cell detection and cell classification was performed by pathologists (J.V; E.R; W.S) on the training images and randomly in the full set of the cohort.

The validation cohort (ICL, UMC), used the same algorithms developed in the discovery set with minor adjustment. Briefly, in the validation set, the new data are fed into the trained model of discovery cohort to generate cell predictions. No changes was interpolated on the steps (from estimation of color vectors to cell detection) prior cell classification. For each tissue-based cell classifier, after few rounds of cell classification review and correction by pathologists (E.R; W.S), the re-trained classifier was implemented. In both cohorts, the entire image analysis to determine TIL levels was performed blinded to patient outcome data.

Endpoints & statistics

Objective response rate (ORR) and progression-free survival (PFS) were determined by blinded radiology review using Response Evaluation Criteria In Solid Tumors (RECIST) version 1.1. PFS was defined as the time from PD-(L)1 inhibitor start to progression or death, and for those without progression, censoring was done at the time of the last disease assessment scan showing no progression. Overall survival (OS) was calculated from the time of PD-(L)1 inhibitor start to death. Patients who were still alive at the time of data analysis were censored at the date of last contact.

TILs were quantified as the number of cells per mm2, and were analyzed for association with clinicopathological variables using either Fisher's exact or chi-square tests, as appropriate. For simplicity of interpretation, dichotomization of the TILs/mm2 counts was based on the cutoff yielding the lowest P-value (eFigure S6) in the discovery cohort and then rounded to the nearest number divisible by 50. PD-L1 TPS (<1, 1-49, ≥50 %) and TMB (<10 vs. ≥10 mu/Mb) cutoffs for subgroup analysis were in accordance with routine established and FDA recommendations [8,9]. Continuous variables were compared with Mann–Whitney U test for 2-group comparisons or the Kruskal-Wallis exact test for multiple comparisons. The Intraclass Correlation Coefficient (ICC) was calculated using a two way random effect with absolute agreement definition. Kaplan-Meier methodology was used to estimate event-time distributions. Log-rank tests were used to test for differences in event-time distributions, and Cox models were fitted to obtain estimates of hazard ratios in univariate and multivariable models. The proportional hazard assumption of the Cox models was tested using Schoenfeld residuals. Receiver operating characteristic curves were generated to assess the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of continuous variables with the area under the receiver operating characteristic curve (AUC). Survival and statistical analyses were performed using SPSS (v. 26) and R (v. 4.1.2) packages. P-values less than 0.05 were considered statistically significant.

Extended references:

1. Alessi J V, Ricciuti B, Spurr LF, et al. SMARCA4 and other SWI/SNF family genomic alterations in non-small cell lung cancer: Clinicopathological characteristics and outcomes to immune checkpoint inhibition. *J Thorac Oncol*. April 2021. doi:10.1016/j.jtho.2021.03.024

2. Ricciuti B, Wang X, Alessi J V., et al. Association of High Tumor Mutation Burden in Non–Small Cell Lung Cancers With Increased Immune Infiltration and Improved Clinical Outcomes of PD-L1 Blockade Across PD-L1 Expression Levels. *JAMA Oncol*. 2022:1-9. doi:10.1001/jamaoncol.2022.1981

3. Patel NM, Jo H, Eberhard DA, et al. Improved Tumor Purity Metrics in Next-generation Sequencing for Clinical Practice: The Integrated Interpretation of Neoplastic Cellularity and Sequencing Results (IINCaSe) Approach. *Appl Immunohistochem Mol Morphol*.

2019;27(10):764-772. doi:10.1097/PAI.0000000000000684

4.   Väyrynen JP, Lau MC, Haruki K, et al. Prognostic Significance of Immune Cell Populations Identified by Machine Learning in Colorectal Cancer Using Routine Hematoxylin and Eosin-Stained Sections. *Clin Cancer Res*. 2020;26(16):4326-4338. doi:10.1158/1078-0432.CCR-20-0071

5.   Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng*. 2010;57(4):841-852. doi:10.1109/TBME.2009.2035102

6.   Salgado R, Denkert C, Demaria S, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol Off J Eur Soc Med Oncol*. 2015;26(2):259-271. doi:10.1093/annonc/mdu450

7.   Bai Y, Cole K, Martinez-Morilla S, et al. An Open Source, Automated Tumor Infiltrating Lymphocyte Algorithm for Prognosis in Triple-Negative Breast Cancer. *Clin Cancer Res*. June 2021:clincanres.0325.2021. doi:10.1158/1078-0432.ccr-21-0325

8.   Lamberti G, Spurr LF, Li Y, et al. Clinicopathological and genomic correlates of programmed cell death ligand 1 (PD-L1) expression in nonsquamous non-small-cell lung cancer. *Ann Oncol Off J Eur Soc Med Oncol*. 2020;31(6):807-814. doi:10.1016/j.annonc.2020.02.017

9.   Subbiah V, Solit DB, Chan TA, Kurzrock R. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) ≥10: a decision centered on empowering patients and their physicians. *Ann Oncol*. 2020;31(9):1115-1118. doi:10.1016/J.ANNONC.2020.07.002
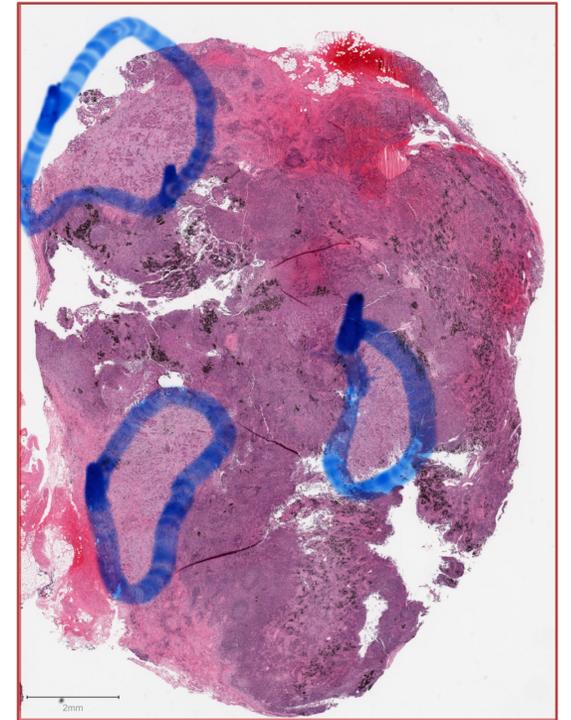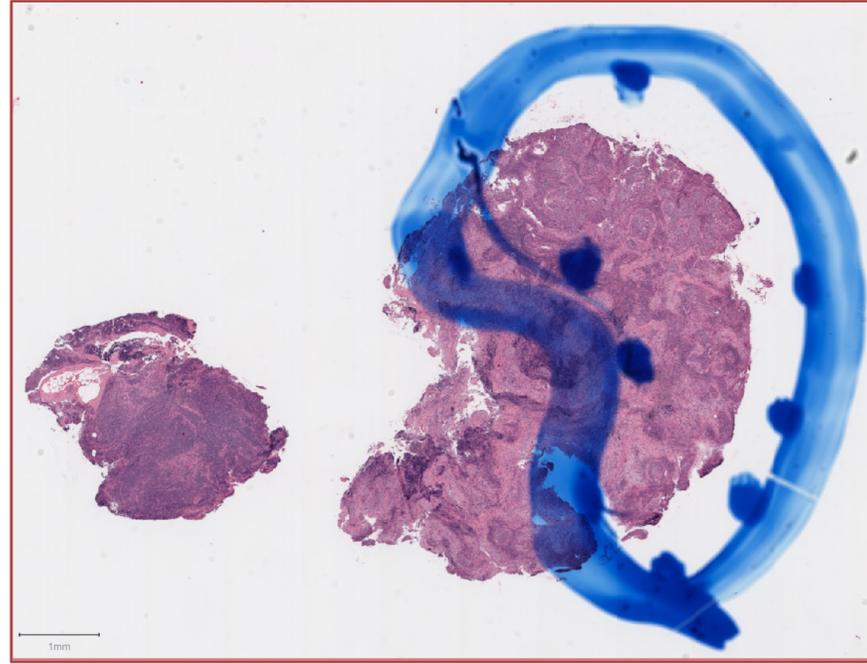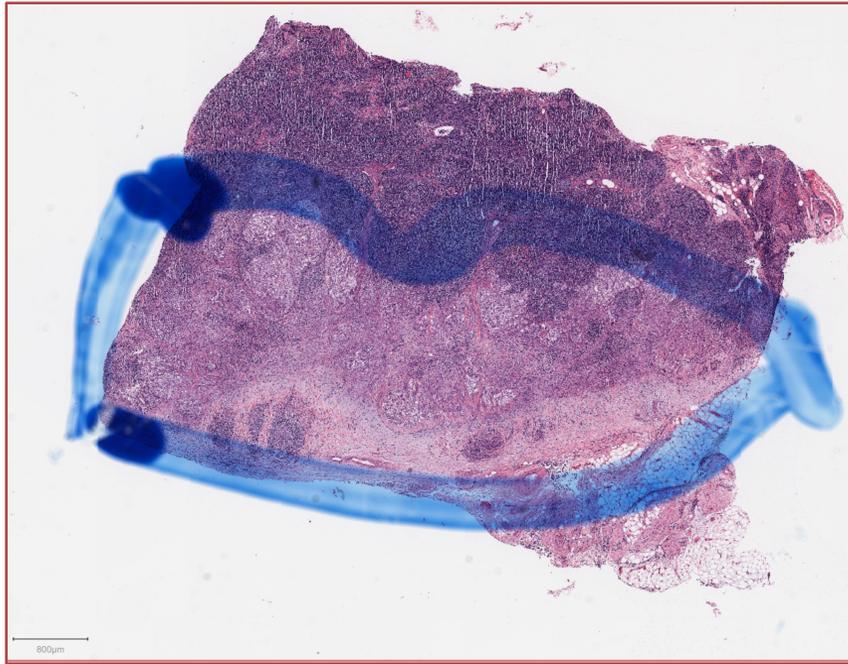
eFigure S1 : **Areas included for TILs assessment in LN tissues.** The area marked by pathologists and used for DNA sequencing on parallel slides was harvested for image analysis in LN tissues. As marked up, the pre-existing lymphoid stroma and noncontiguous lymph node structure

were excluded from the analysis. Representative H&E images from three different cases.
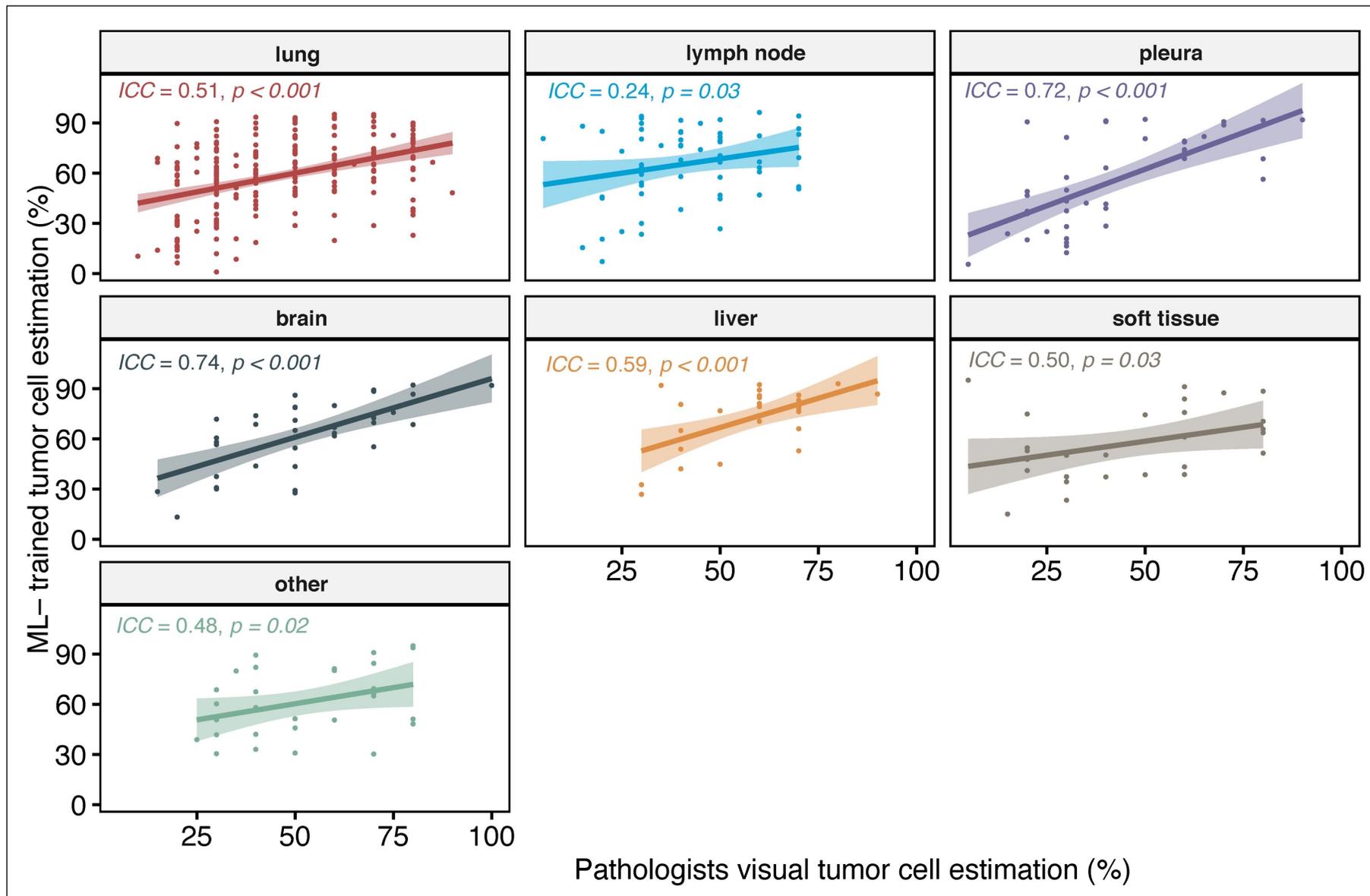
eFigure S2: ML-based tumor content versus manual estimation in DFCI cohort. The intraclass correlation coefficients (ICC) between ML-trained and visual (VAF adjusted, see eSupplementary method) tumor cell estimations (%) for different tissue sites.
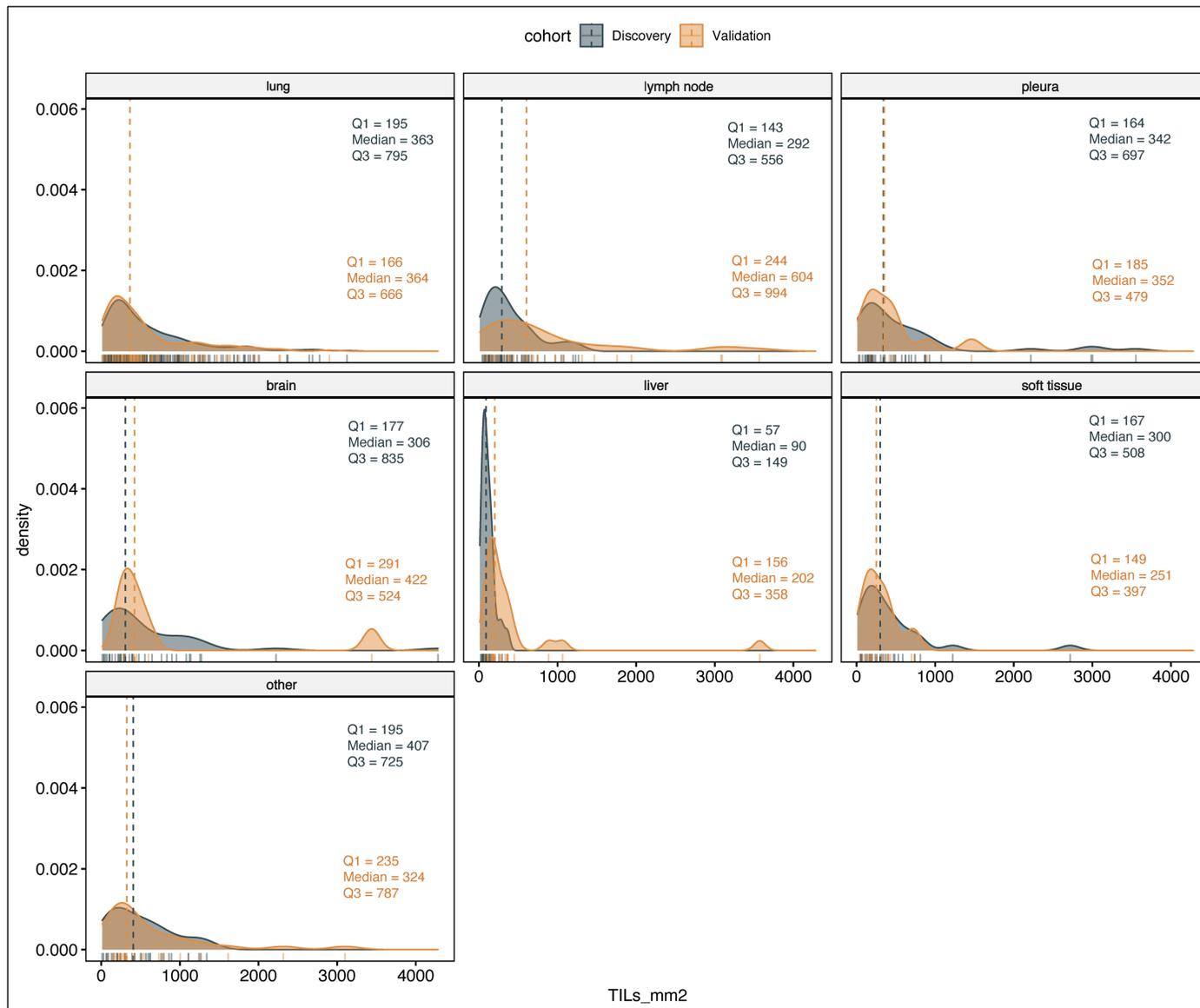
eFigure S3: Range of TIL density in different sample types. Density plot of TILs/mm2 level in different tissue sites in discovery (gray) and validation (yellow) cohorts. (dash line = median).
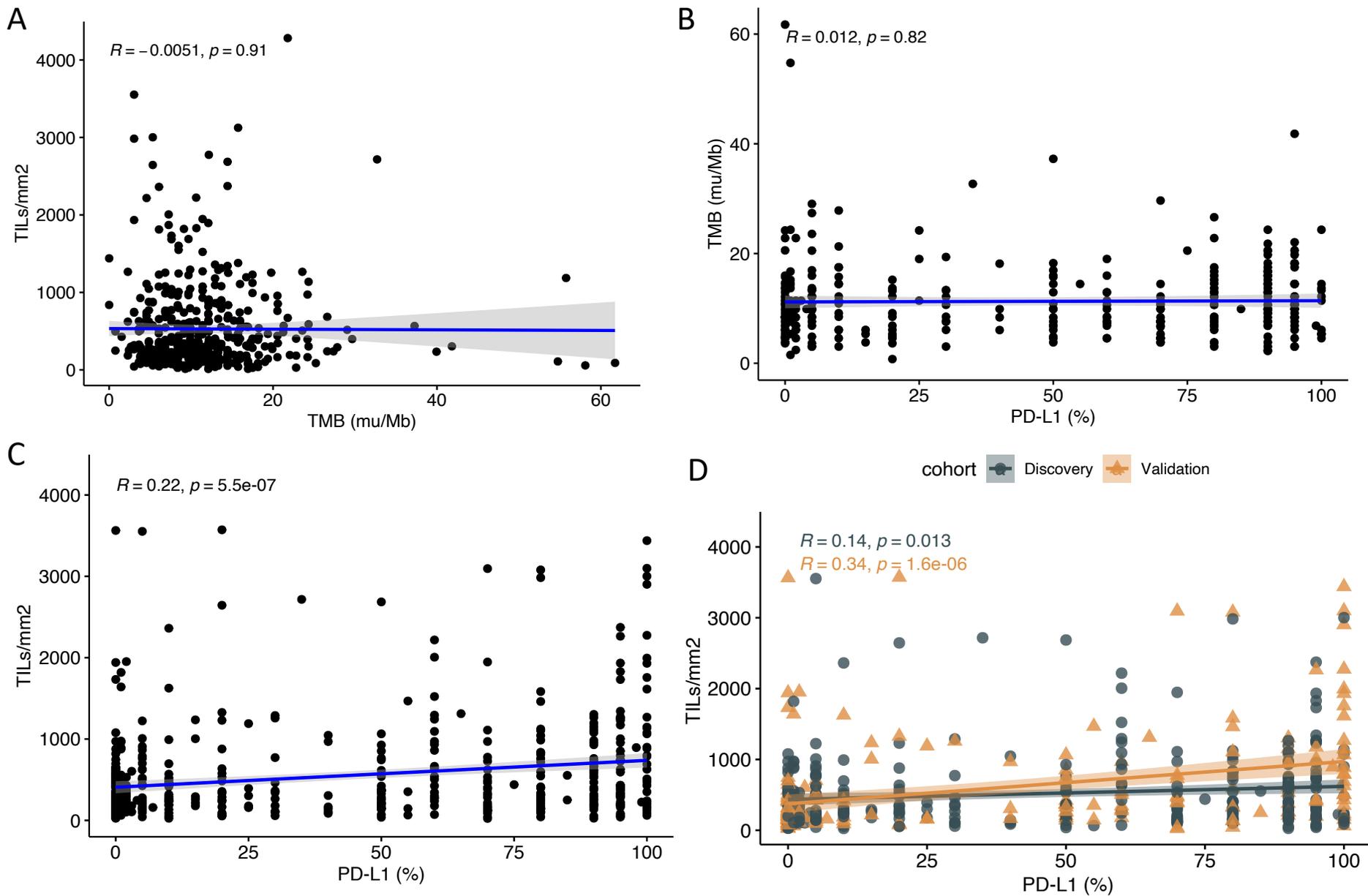
eFigure S4: Correlation between TIL levels and TMB/PD-L1. A lack of correlation is seen between **A)** TILs and TMB, **B)** TMB and PD-L1 in the DFCI cohort; Weak correlation is seen between **C)** TILs and PD-L1 in the entire cohort, and in each of the discovery and validation cohorts

KRAS subtypes in adenocarcinoma (n = 149)

	N	
52 %	77	KRAS G12C
15 %	22	KRAS G12V
8 %	12	KRAS G12D
7 %	11	KRAS G12A
5 %	7	KRAS Q61H
4 %	6	KRAS G13C
1 %	2	KRAS G12F
1 %	2	KRAS G12R
1 %	2	KRAS G13R
1 %	2	KRAS L19F
<1 %	1	KRAS A146V
<1 %	1	KRAS G12E
<1 %	1	KRAS G12S
<1 %	1	KRAS G13D
<1 %	1	KRAS G13V
<1 %	1	KRAS Q61L

EGFR subtypes in adenocarcinoma (n = 36)

	N	
39%	14	exon 19 deletion
39%	14	L858R
17%	6	exon 20 insertion
5%	2	G719
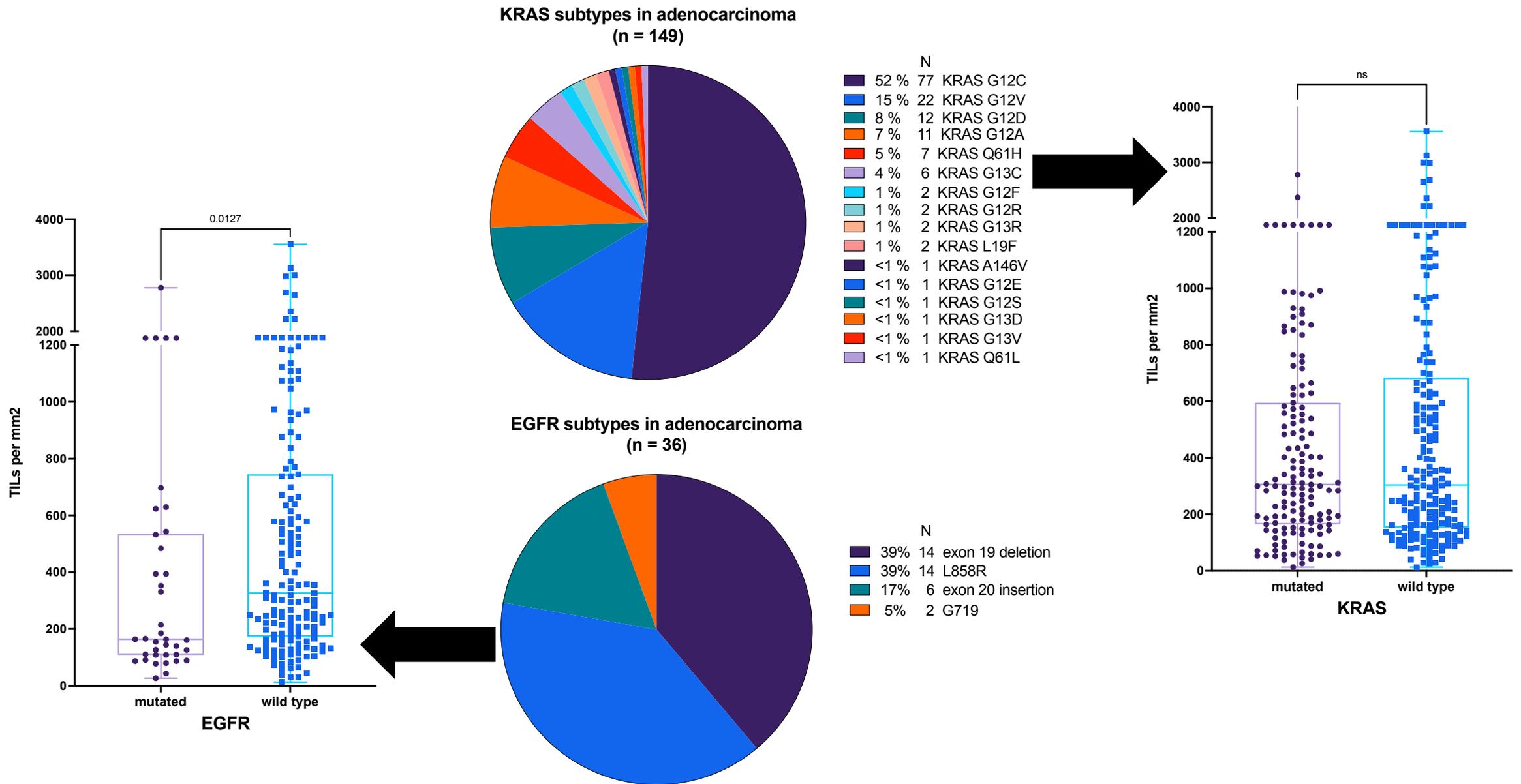
eFigure S5: TIL levels according to KRAS and EGFR mutation status in the discovery cohort. The pie charts show the proportions of KRAS (top) and EGFR (bottom) mutation subtypes seen in adenocarcinoma. Right and left box plots show the distribution of TIL levels in mutated vs wild type EGFR and KRAS cases, respectively.
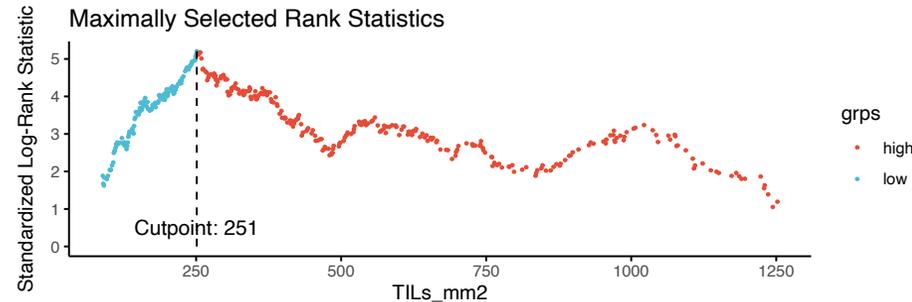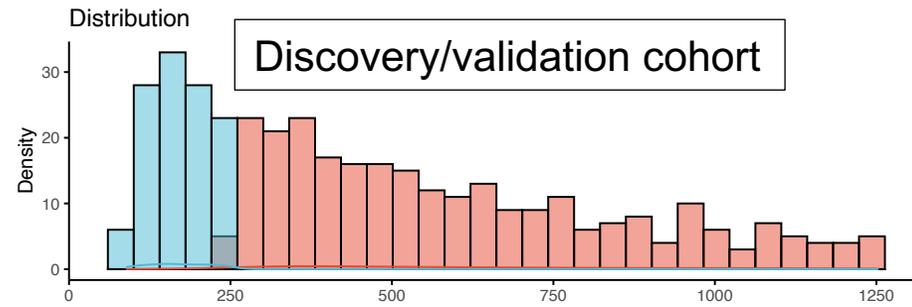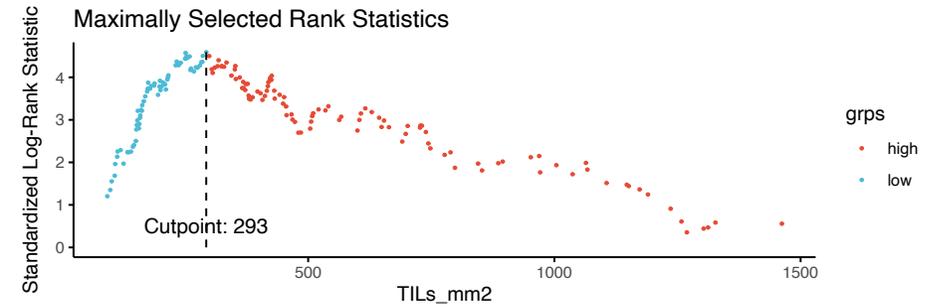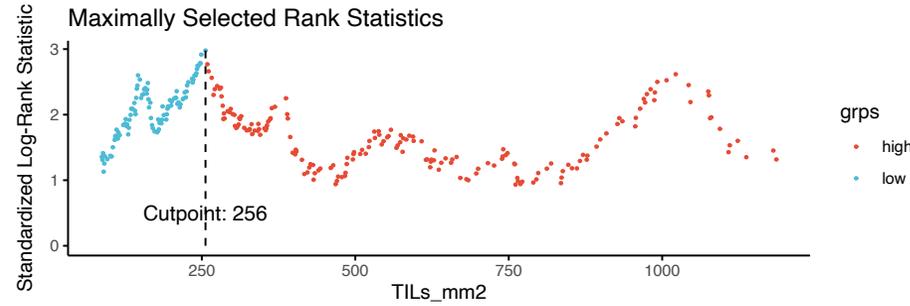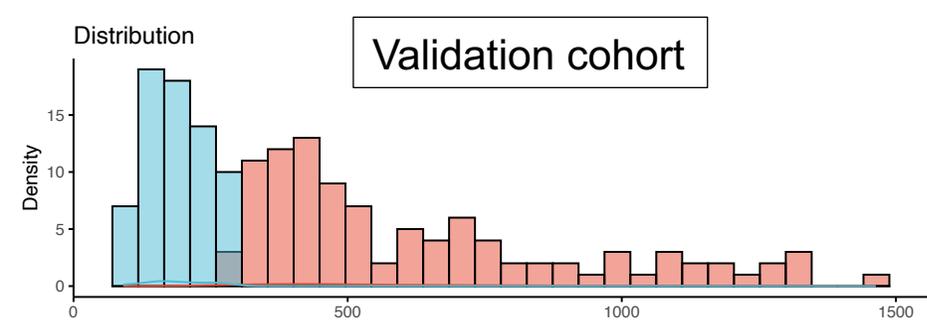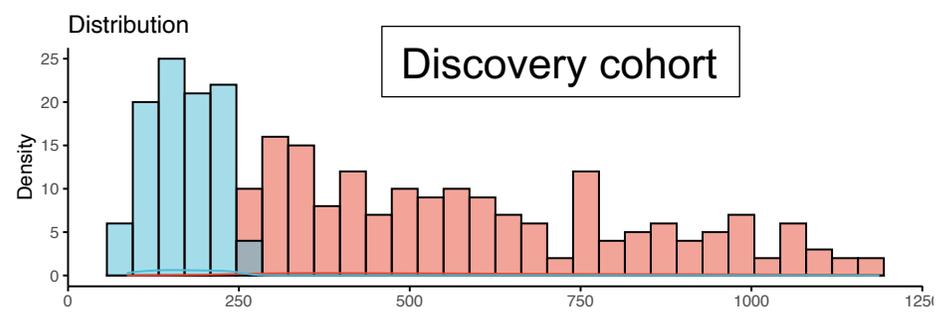
eFigure S6: Log-rank test cutoff identification. Maximally selected log-rank test on progression-free survival in the discovery (n = 446, upper left), validation (n= 239, upper right), and merged discovery and validation cohort (n = 685, lower).
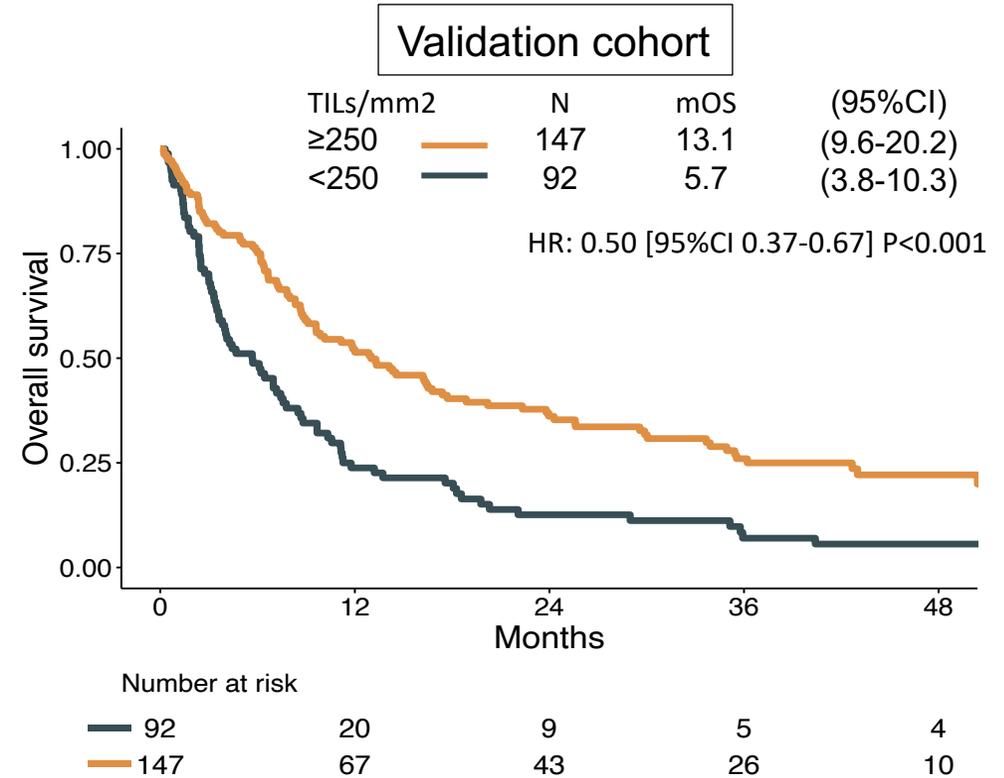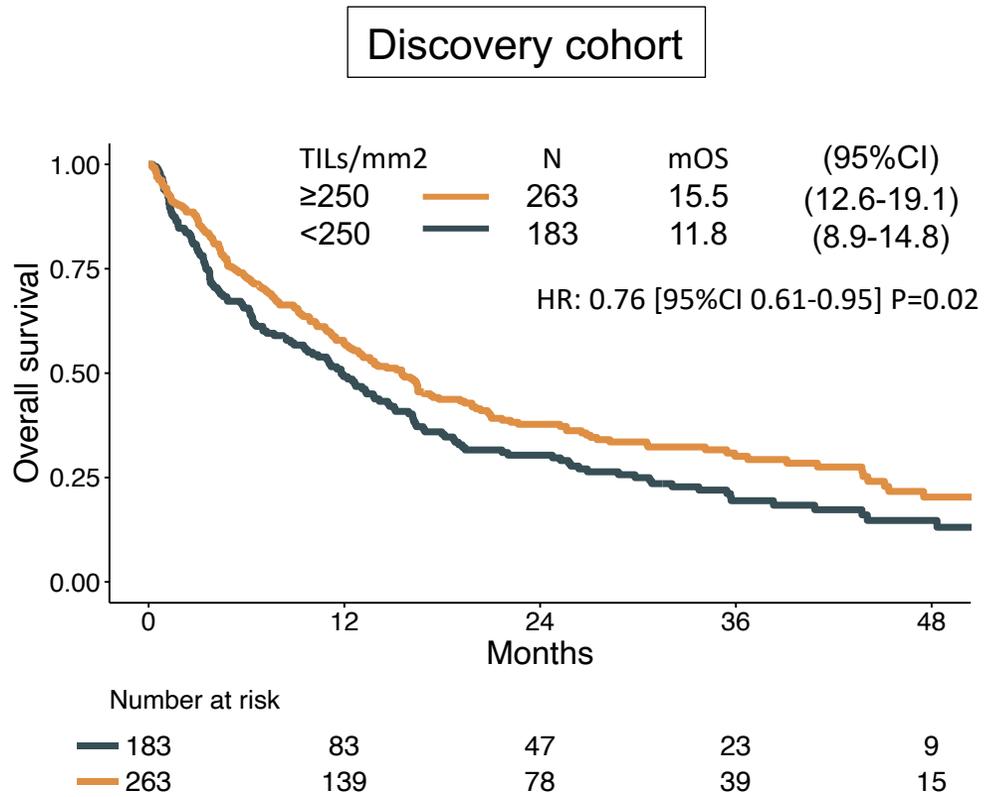
eFigure S7: TIL levels and overall survival to immunotherapy. Overall survival (OS) according to TIL levels (<250 vs. ≥250 cells/mm2) in the discovery and validation cohorts.

mOS, median OS in months; HR, hazard ratio; CI, confidence interval.

Cox proportional hazard model of OS in discovery cohort | Cox proportional hazard model of OS in validation cohort

	Discovery HR (95% CI)	Discovery P	Validation HR (95% CI)	Validation P
TILs/mm2 (≥ vs < 250)	0.74 (0.56 - 0.97)	0.028 *	0.75 (0.62 - 0.89)	0.001 **
PD-L1/TPS (≥ vs < 50)	0.67 (0.50 - 0.90)	0.009 **	0.70 (0.47 - 1.02)	0.064
TMB/Mb (≥ vs < 10)	0.76 (0.58 - 1.00)	0.046 *	N.A	
ECOG (≥ 2 vs 0-1)	2.51 (1.82 - 3.47)	<0.001 ***	2.85 (1.82 - 4.45)	<0.001 ***
Tumor type (metastatic vs primary)	1.04 (0.79 - 1.36)	0.796	0.65 (0.46 - 0.91)	0.013 *
Tissue type (resection vs biopsy)	0.77 (0.59 - 1.01)	0.058	1.14 (0.89 - 1.45)	0.302
ICI line (≥ 2 vs 1)	1.50 (1.12 - 2.02)	0.007 **	1.10 (0.74 - 1.64)	0.644
Smoking (ever vs never)	1.12 (0.75 - 1.67)	0.587	N.E	
Sex (male vs female)	N.E		0.83 (0.59 - 1.17)	0.282

Discovery axis: 0.5 1 1.5 2 2.5 3 3.5 4
Validation axis: 0.5 1 2 5

eFigure S8: Multivariable models. Forest plot based on the results of multivariate analysis of the significant independent predictive factors

associated with overall survival (OS) in the discovery (left) and validation (right) cohorts.
N.E, not entered; N.A, not available.
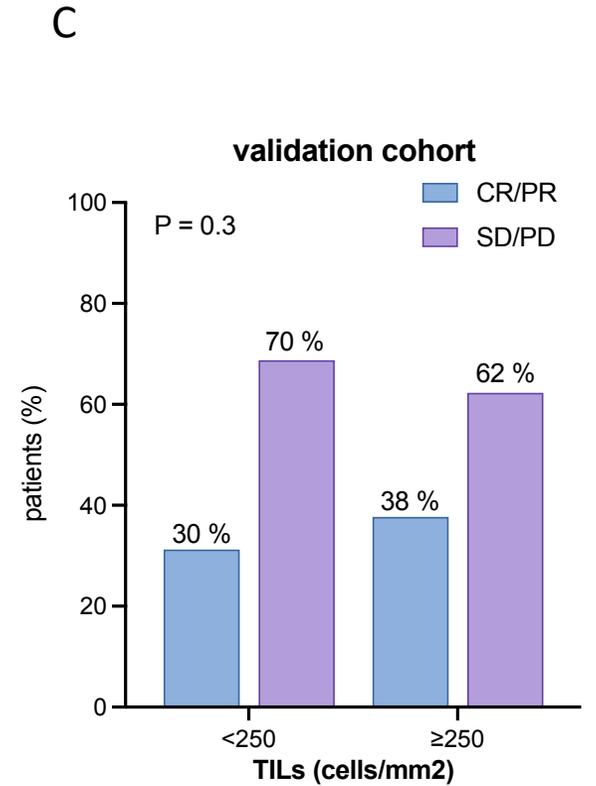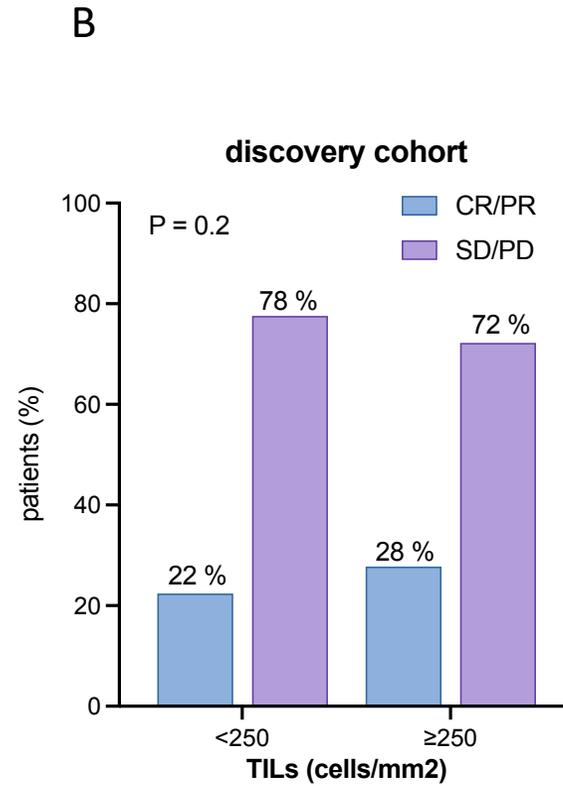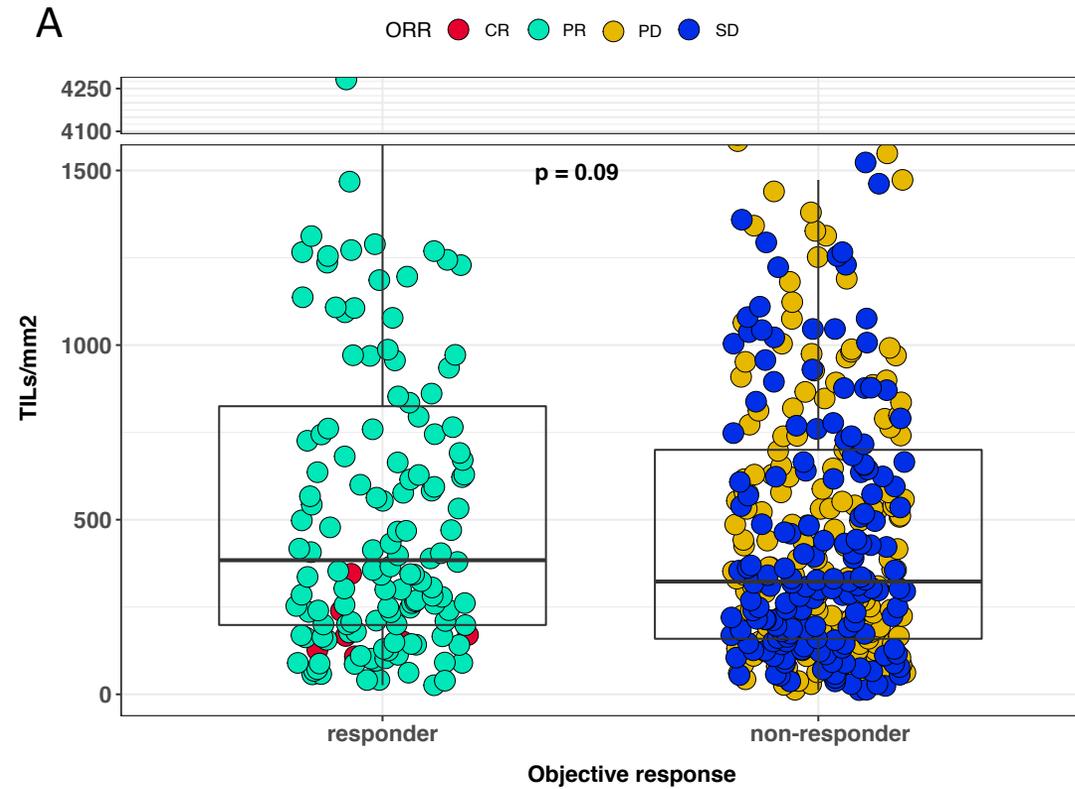
eFigure S9: TILs and objective response rate (ORR): A) TIL density in the combined cohort of patients who had complete response (CR)/ partial response (PR) or stable disease (SD)/progressive disease (PD). ORR was available for only 547 out of 685 patients. B-C) Proportion of patients with CR/PR and SD/PD in the TIL high versus TIL low groups in discovery and validation sets.

First-line ICI monotherapy

TILs/mm2	N	mPFS	(95%CI)	ORR*
≥250	183	7.1	(6.1-9.9)	38%
<250	87	3.7	(2.7-4.9)	28%

HR: 0.52 [95%CI 0.39-0.71] P<0.001

TILs/mm2	N	mOS	(95%CI)
≥250	183	24.3	(16.5-35.4)
<250	87	13.3	(9.6-18.1)

HR: 0.59 [95%CI 0.43-0.81] P<0.001

Number at risk (PFS): 87, 14, 6, 1, 0; 183, 59, 37, 16, 0

Number at risk (OS): 87, 42, 22, 7, 1; 183, 104, 69, 34, 6

Subsequent-line ICI monotherapy

TILs/mm2	N	mPFS	(95%CI)	ORR*
≥250	227	3.2	(2.7-3.9)	23%
<250	188	2.3	(1.8-2.7)	21%

HR: 0.76 [95%CI 0.62-0.93] P=0.008

TILs/mm2	N	mOS	(95%CI)
≥250	227	11.5	(9.6-13.6)
<250	188	6.9	(5.8-10.1)

HR: 0.77 [95%CI 0.62-0.95] P=0.01

Number at risk (PFS): 188, 26, 13, 7, 3; 227, 50, 18, 8, 7

Number at risk (OS): 188, 61, 34, 21, 12; 227, 102, 52, 31, 19
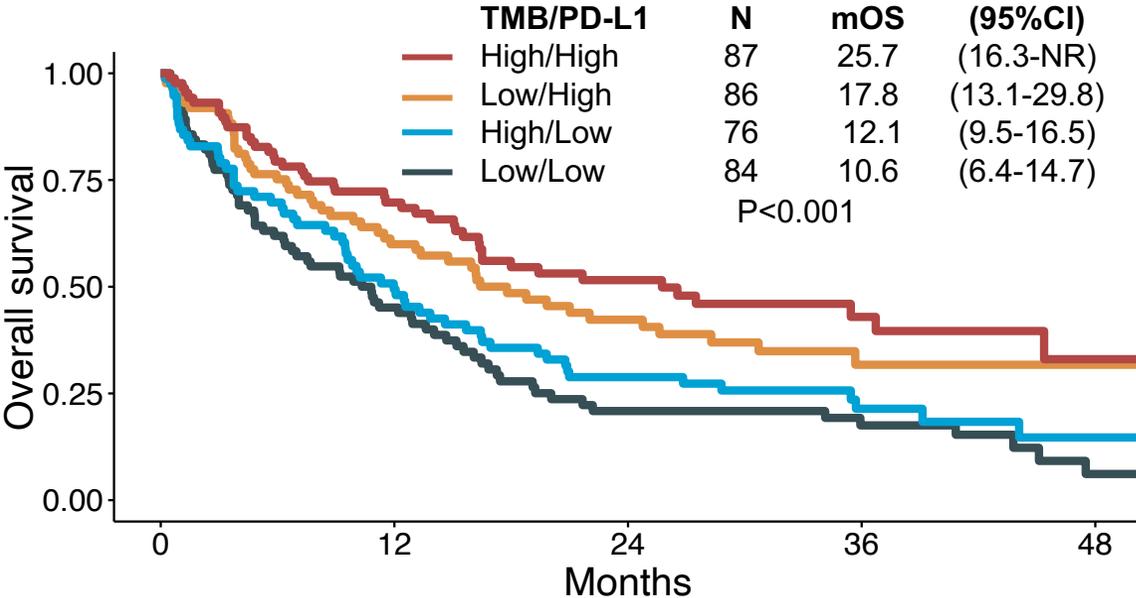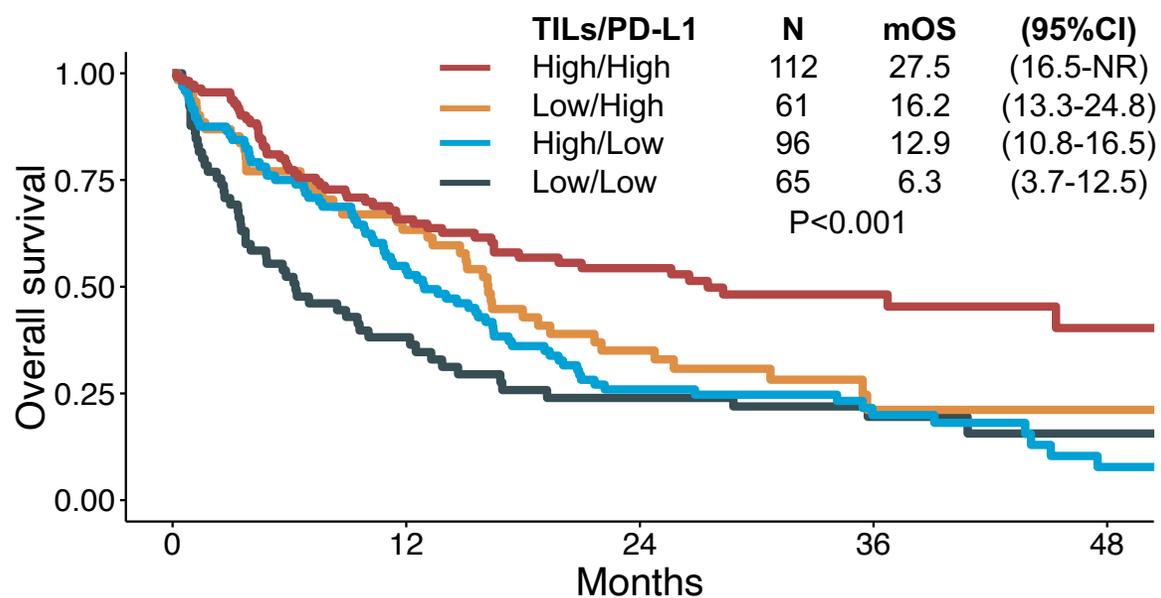
eFigure S10: Treatment line-TILs interaction. Progression-free and overall survival in treatment subgroups of the combined cohorts.
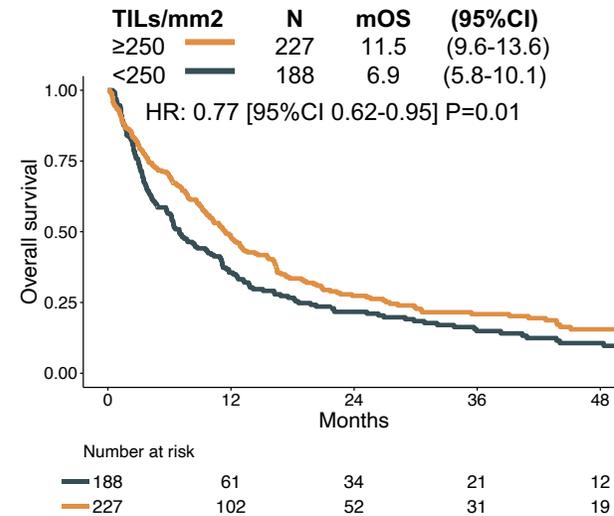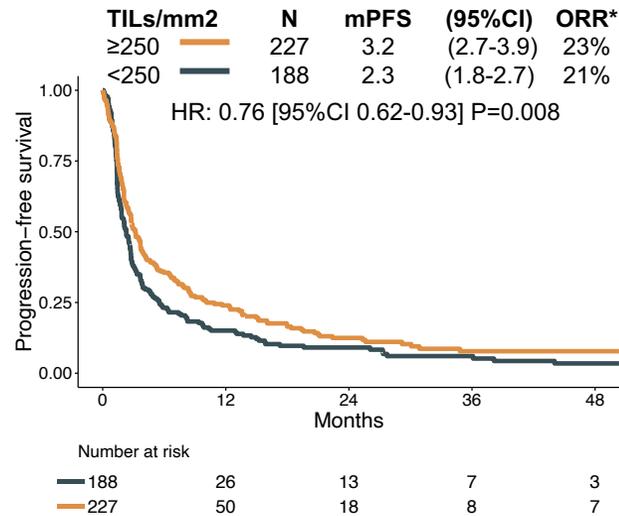
HR, hazard ratio; CI, confidence interval; mPFS, median PFS in months; mOS, median OS in months; ORR, objective response rate.
 * ORR was available for 547 of 685 of the patients in the combined cohort.

eFigure S11: Combined TILs/PD-L1, TMB/PD-L1 and immunotherapy outcome. Overall survival (OS) of combined TILs (<250 vs. ≥250 cells/mm2)/(<50 vs. ≥50%) and TMB(<10 vs. ≥10mu/Mb)/PD-L1(<50 vs. ≥50%) scores in the DFCI cohort.

HR, hazard ratio; CI, confidence interval; mOS, median OS in months; NR, not reached
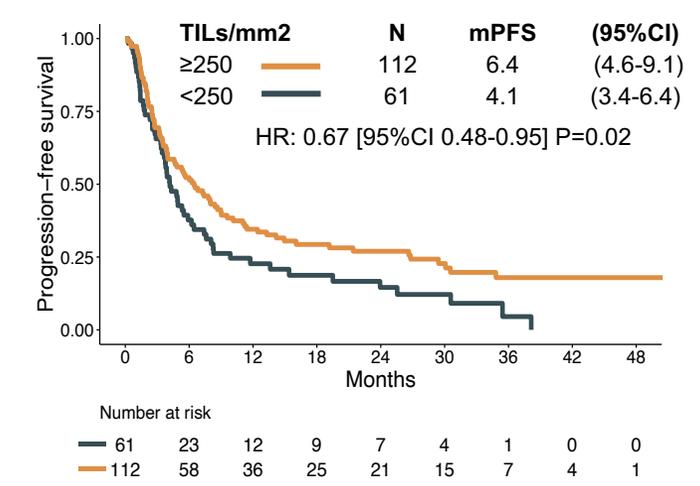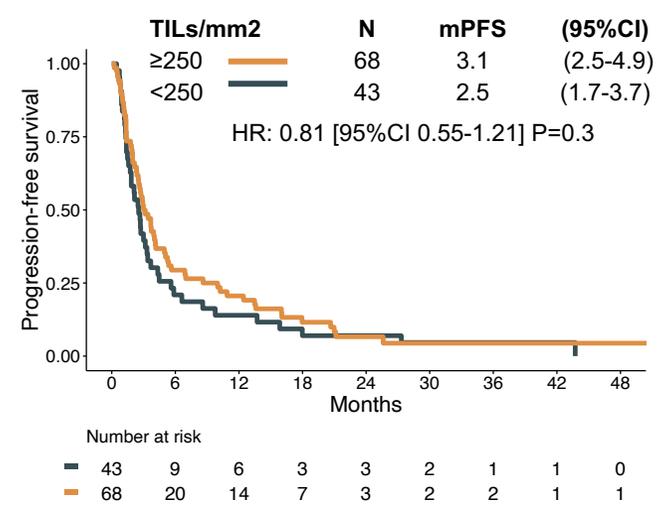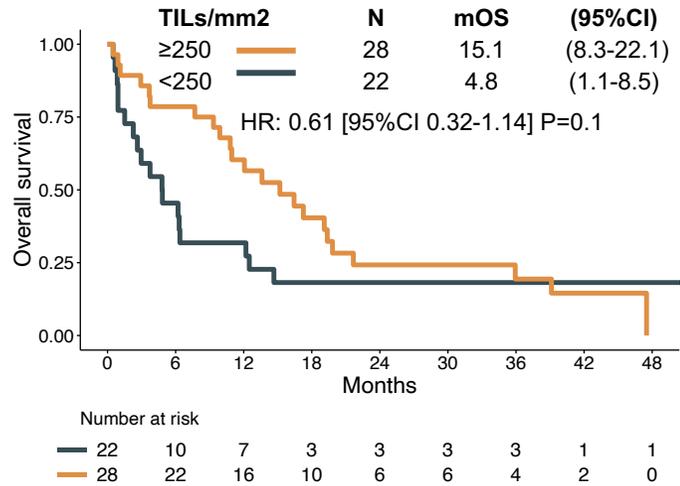
eFigure S12: TILs and immunotherapy outcome based on PD-L1 stratification. Overall survival (OS) and progression-free survival (PFS) according to TILs (<250 vs. ≥250 cells/mm2) in response to ICI therapy in the PD-L1 subgroups of the discovery (upper row) and validation (lower row) cohorts. HR, hazard ratio; CI, confidence interval; mPFS, median PFS in months; median OS in months; NR, not reached
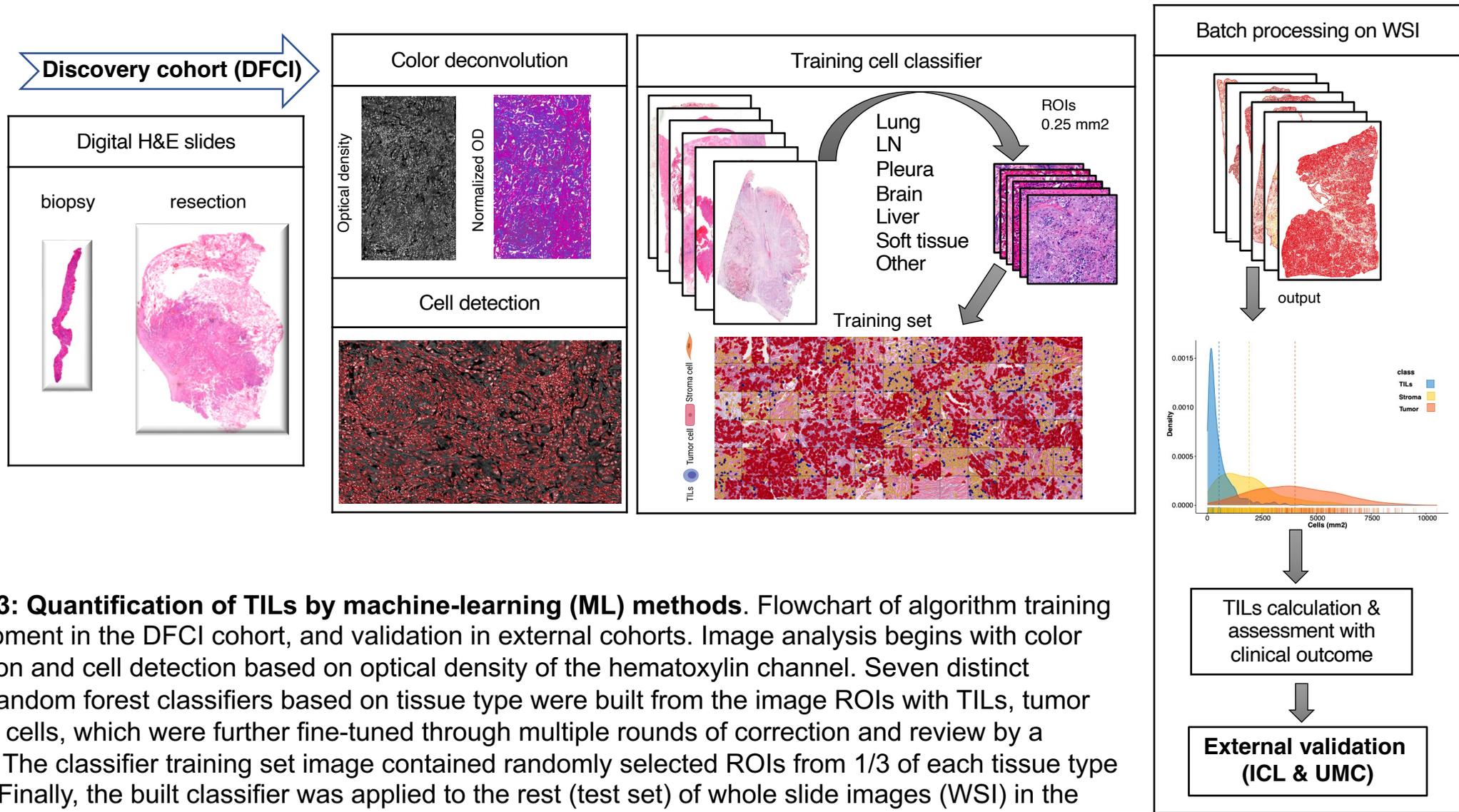
eFigure S13: Quantification of TILs by machine-learning (ML) methods. Flowchart of algorithm training and development in the DFCI cohort, and validation in external cohorts. Image analysis begins with color deconvolution and cell detection based on optical density of the hematoxylin channel. Seven distinct pretrained random forest classifiers based on tissue type were built from the image ROIs with TILs, tumor and stromal cells, which were further fine-tuned through multiple rounds of correction and review by a pathologist. The classifier training set image contained randomly selected ROIs from 1/3 of each tissue type population. Finally, the built classifier was applied to the rest (test set) of whole slide images (WSI) in the DFCI cohort. Application of trained classifier resulted in TIL, tumor and stroma cell measurements. The ML model was subsequently tested in validation cohorts from ICL and UMC. DFCI, Dana Farber Cancer Institute; OD, optical density; ROIs, region of interests; ICL, Imperial College of London; UMC, Amsterdam University Medical Center.

Tables

eTable S1: Patient characteristics according to TIL levels in the discovery and validation cohorts.

	Discovery cohort			Validation cohort		
	TILs low (<250 cell/mm2) N = 183 (%)	TILs high (≥250 cell/mm2) N =263 (%)	*P*	TILs low (<250 cell/mm2) N = 92 (%)	TILs high (≥250 cell/mm2) N =147 (%)	*P*
Treatment agent			0.1			**<0.001**
Pembrolizumab	95 (52)	151 (57)		27 (29)	90 (61)	
Nivolumab	67 (37)	97 (37)		56 (61)	50 (34)	
Atezolizumab	15 (8)	13 (5)		8 (8)	6 (4)	
Other [a]	6 (3)	2 (1)		1 (2)	1 (1)	
ICI line			0.3			**<0.001**
1	68 (37)	111 (42)		19 (21)	72 (49)	
≥2	115 (63)	152 (58)		73 (79)	75 (51)	
Age			0.3			0.1
<66	96 (52)	125 (48)		50 (54)	69 (47)	
≥66	87 (48)	138 (52)		42 (46)	78 (53)	
Gender			0.7			0.2
female	104 (57)	144 (55)		34 (37)	63 (43)	
male	79 (43)	119 (45)		58 (63)	84 (57)	
Histology			0.1			0.04
LUAD	150 (82)	202 (77)		52 (56)	106 (72)	
LUSC	24 (13)	40 (15)		33 (36)	22 (15)	
other	9 (5)	21 (8)		6 (8)	17 (13)	
Smoking			0.08			0.8
never	30 (17)	28 (11)		9 (9)	16 (11)	
ever	153 (83)	235 (89)		80 (87)	128 (87)	
unknown				3 (4)	3 (2)	
ECOG			0.5			0.4
0-1	146 (80)	216 (82)		74 (80)	123 (83)	
≥2	37 (20)	45 (17)		15 (16)	21 (14)	
unknown		2 (1)		3 (4)	3 (3)	
Specimen site			**<0.001**			0.1
lung	77 (42)	147 (56)		48 (52)	76 (52)	
lymph node	24 (13)	36 (14)		8 (9)	24 (16)	
pleura	19 (10)	22 (8)		5 (5)	6 (5)	
brain	15 (8)	22 (8)		1 (1)	6 (4)	
liver	24 (13)	3 (1)		13 (14)	10 (7)	
soft tissue	12 (7)	15 (6)		8 (9)	7 (4)	
other	12 (7)	18 (7)		9 (10)	18 (12)	
Tumor type			0.03			0.4
primary	72 (40)	130 (49)		44 (48)	73 (50)	
metastatic	108 (59)	125 (47)		48 (52)	74 (50)	
unknown	3 (1)	8 (4)				

Tissue type			0.1			0.8
biopsy	85 (46)	109 (41)		73 (79)	119 (81)	
resection	98 (54)	154 (59)		19 (21)	28 (19)	

[a] Including durvalumab and commercial immunotherapy agents.

Abbreviations: ICI, Immune checkpoint inhibitors; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

eTable S2: Machine-learning derived quantitative detail of the cell subsets from the histological H&E images in the entire cohort, including both the discovery and validation cohorts (n = 685).

	Mean	Min	Median	Max	IQR	Q1	Q3
TIL cells/mm2[a]	543	12	327	4282	517	164	681
Tumor cells/mm2[b]	4017	115	3874	10497	2822	2568	5390
Stroma cells/mm2[c]	1922	14	1686	9430	1695	885	2580
TIL (%)[d]	8	0.2	7	39.6	7.1	3	11
Tumor (%)[e]	62	1	65	99.5	32.5	47	80
Stroma (%)[f]	31	1	29	94	29	14	43
TSP (%)[g]	34	0.1	31	98.8	32	15	48

[a] calculated by: #lymphocytes/tissue size
[b] calculated by: #tumor cells/tissue size
[c] calculated by: #stroma cells/tissue size
[d] calculated by: (#lymphocytes/#detected cells) x 100 %
[e] calculated by: (#tumor cells/#detected cells) x 100 %
[f] calculated by: (#stroma cells/#detected cells) x 100 %
[g] calculated by: c/(c + b) x 100 %

Abbreviations: TSP, tumor-stroma proportion; IQR, interquartile range

eTable S3: Clinicopathologic variables, including TILs, in association with PFS and OS to ICIs in A) discovery and B) validation cohorts (Univariate analyses, Log-rank test, unadjusted Cox proportional hazard ratios). Variables with P < 0.25 (bold) were selected for multivariable analysis.

	Progression-free survival			Overall survival		
A) Discovery cohort	HR	95% CI	*P*	HR	95% CI	*P*
TILs/mm2(≥ vs<250)	0.74	(0.61- 0.90)	**0.003**	0.77	(0.61-0.95)	**0.02**
PD-L1 (≥ vs < 50%)	0.57	(0.45-0.72)	**<0.001**	0.57	(0.43-0.74)	**<0.001**
TMB/Mb (≥ vs < 10)	0.77	(0.64-0.94)	**0.01**	0.78	(0.62-0.97)	**0.02**
ECOG (≥ 2 vs 0-1)	1.65	(1.29-2.11)	**<0.001**	2.44	(1.87-3.18)	**<0.001**
ICI line (≥ 2 vs 1)	1.39	(1.14-1.70)	**<0.001**	1.64	(1.29-2.07)	**<0.001**
Smoking (ever vs never)	0.67	(0.50-0.90)	**0.01**	0.81	(0.59-1.12)	**0.20**
Tumor type (metastatic vs primary)	1.15	(0.95-1.40)	**0.16**	1.08	(0.87-1.34)	0.49
Tissue type (resection vs biopsy)	0.87	(0.71-1.06)	**0.17**	0.85	(0.68-1.06)	**0.15**
Age (≥ vs < 66)	0.96	(0.79-1.17)	0.72	1.06	(0.85-1.32)	0.60
Sex (male vs female)	0.95	(0.78-1.16)	0.60	1.01	(0.81-1.26)	0.95
B) Validation cohort						
TILs/mm2(≥ vs<250)	0.47	(0.35-0.63)	**<0.001**	0.50	(0.37-0.67)	**<0.001**
PD-L1 (≥ vs < 50%)	0.46	(0.33-0.65)	**<0.001**	0.62	(0.44-0.86)	**<0.001**
ECOG (≥ 2 vs 0-1)	1.76	(1.18-2.63)	**0.01**	2.34	(1.58-3.48)	**<0.001**
ICI line (≥ 2 vs 1)	1.91	(1.40-2.60)	**<0.001**	1.46	(1.08-1.97)	**0.02**
Smoking (ever vs never)	0.82	(0.52-1.31)	0.40	0.79	(0.49-1.25)	0.31
Tumor type (metastatic vs primary)	0.75	(0.56-1.00)	**0.05**	0.76	(0.57-1.01)	**0.06**
Tissue type (resection vs biopsy)	1.25	(1.03-1.51)	**0.02**	1.26	(1.04-1.52)	**0.02**
Age (≥ vs < 66)	1.12	(0.84-1.50)	0.43	1.08	(0.80-1.44)	0.62
Sex (male vs female)	0.80	(0.60-1.08)	**0.14**	0.83	(0.61-1.11)	**0.21**

eTable S4: Comparison of different single and multi-assays sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for responders versus non-responders after ICI treatment in overall discovery cohort and PD-L1 subsets. 95% CI in parentheses.

Discovery cohort (n=446)	AUC	*P*	Sensitivity	Specificity	PPV	NPV
PD-L1/TMB	0.70(0.63-0.76)	**<0.001**	0.62 (0.51-0.72)	0.70 (0.63-0.75)	0.43(0.36-0.54)	0.83(0.76-0.87)
PD-L1/TILs	0.68(0.63-0.75)	**<0.001**	0.79 (0.69-0.87)	0.56 (0.49-0.62)	0.40(0.33-0.53)	0.87(0.81-0.90)
PD-L1	0.68(0.62-0.75)	**<0.001**	0.78(0.69-0.86)	0.52(0.45-0.58)	0.37(0.31-0.51)	0.87(0.79-0.89)
TMB	0.59(0.53-0.65)	0.05	0.48(0.38-0.57)	0.69(0.63-0.74)	0.35(0.29-0.44)	0.79(0.72-0.83)
TILs	0.55(0.49-0.61)	0.08	0.70(0.61-0.78)	0.40(0.35-0.45)	0.29(0.24-0.38)	0.80(0.72-0.83)
PD-L1 negative subgroup (n=50)						
TILs	0.77(0.63-0.88)	**0.03**	0.67(0.22-0.95)	0.86(0.72-0.95)	0.40(0.22-0.88)	0.95(0.73-0.98)
TMB	0.65 (0.51-0.78)	0.3	0.66(0.23-0.96)	0.72(0.56-0.84)	0.25(0.14-0.78)	0.93(0.70-0.97)
PD-L1 intermediate subgroup (n=111)						
TILs	0.50(0.40-0.59)	0.9	N.A	N.A	N.A	N.A
TMB	0.49(0.39-0.58)	0.9	N.A	N.A	N.A	N.A
PD-L1 high subgroup (n=173)						
TILs	0.53(0.43-0.61)	0.6	0.55(0.42-0.67)	0.59(0.49-0.68)	0.45(0.35-0.58)	0.67(0.56-0.76)
TMB	0.58(0.49-0.66)	0.07	0.71(0.58-0.81)	0.45(0.36-0.55)	0.44(0.35-0.59)	0.72(0.59-0.79)

Abbreviation: N.A, not applicable