Carcinogenesis, 2020, Vol. 41, No. 12, 1671–1681

doi:10.1093/carcin/bgaa105
Advance Access publication October 9, 2020
Original Article

ORIGINAL ARTICLE

Digitally quantified CD8+ cells: the best candidate marker for an immune cell score in non-small cell lung cancer?

Thomas K.Kilvaer[1,2,*], Erna-Elise Paulsen[2,3], Sigve Andersen[1,2], Mehrdad Rakaee[2,4], Roy M.Bremnes[1,2], Lill-Tove Rasmussen Busund[4,5] and Tom Donnem[1,2]

[1]Department of Oncology, University Hospital of North Norway, Tromso, Norway, [2]Department of Clinical Medicine, UiT The Arctic University of Norway, Tromso, Norway, [3]Department of Pulmonary Medicine, University Hospital of North Norway, Tromso, Norway, [4]Department of Medical Biology, UiT The Arctic University of Norway, Tromso, Norway and [5]Department of Clinical Pathology, University Hospital of North Norway, Tromso, Norway

*To whom correspondence should be addressed. Department of Oncology, University Hospital of North Norway, 9038 Tromso, Norway. Tel: +47 776 26 765/+47 905 24 635; Fax: +47 77626779; Email: kilvaer@gmail.com

Abstract

The TNM classification is well established as a state-of-the-art prognostic and treatment-decision-making tool for non-small cell lung cancer (NSCLC) patients. However, incorporation of biological data may hone the TNM system. This article focuses on choosing and incorporating subsets of tissue-infiltrating lymphocyte (TIL), detected by specific immunohistochemistry and automatically quantified by open source software, into a TNM-Immune cell score (TNM-I) for NSCLC. We use common markers (CD3, CD4, CD8, CD20 and CD45RO) of TILs to identify TIL subsets in tissue micro-arrays comprising tumor tissue from 553 patients resected for primary NSCLC. The number of TILs is automatically quantified using open source software (QuPath). Their prognostic efficacy, alone and within a TNM-I model, is evaluated in all patients and histological subgroups. Compared with previous manual semi-quantitative scoring of TILs in the same cohort, the present digital quantification proved superior. As a proof-of-concept, we construct a TNM-I, using TNM categories and the CD8+ TIL density. The TNM-I is an independent prognosticator of favorable diagnosis in both the overall cohort and in the main histological subgroups. In conclusion, CD8+ TIL density is the most promising candidate marker for a TNM-I in NSCLC. The prognostic efficacy of the CD8+ TIL density is strongest in lung squamous cell carcinomas, whereas both CD8+ TILs and CD20+ TILs, or a combination of these, may be candidates for a TNM-I in lung adenocarcinoma. Furthermore, based on the presented results, digital quantification is the preferred method for scoring TILs in the future.

Introduction

Lung cancer, with a global estimate of 1.7 million patients who will succumb to the disease in 2018, remains the second and most common cause of cancer-related deaths, in both women and men (1). Lung cancer is mainly divided into two entities, small cell lung cancer and non-small cell lung cancer (NSCLC); of which, the latter comprises nearly 85% of all cases (2). Currently, patient prognoses are best predicted by the TNM classification system, describing the extent of the primary tumor (T), involved lymph nodes (N) and metastatic disease (M). Patients are categorized from stages I (localized cancer without lymph node involvement) to IV (metastatic disease). According to the current ESMO guidelines for early and locally advanced NSCLC, important clinical cutoffs are as follows: stage IB/IIA and stage IIIA/IIIB, representing the thresholds for adjuvant chemotherapy and treatment with curable versus palliative intent, respectively (3).

Although the TNM system incorporates essential information for clinical decision making, prognostication is further improved by

Received: July 18, 2020; **Revised:** October 1, 2020; **Accepted:** October 5, 2020

Abbreviations

CI	confidence interval
CRC	colorectal cancer
DSS	disease-specific survival
HR	hazard ratio
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
NSCLC	non-small cell lung cancer
TILs	tumor infiltrating lymphocytes
TMAs	tissue micro-arrays
TNM-I	TNM-Immune cell score

incorporating biomarkers directly related to cancer biology, development and progression (2). In the last two decades, tremendous effort has been put into this research area, and a plethora of potential biomarkers have been explored (2). For NSCLC patients, this pivotal work has provided several molecular targets/biomarkers implemented in clinical practice for predicting efficacy of treatment with targeted therapies and immunotherapy, though predominantly for subgroups of patients receiving palliative treatment (2,4,5).

In the last decade, clinical cancer research has shifted its focus, from mainly targeting examining tumor cell features, to exploring the multitude of other players in the tumor microenvironment (6). Especially, the increasing understanding of tumor–immune–system interactions has prompted new treatment strategies, leading to better treatment options and, in the case of melanoma, the potential cure of a small subset of patients (7–9). Checkpoint inhibitors constitute an elegant concept wherein the drugs unleash the body's own immune cells primed against neoplastic cells (7). Although research in melanoma patients has paved the road for immunotherapy in general, NSCLC trials have been in the frontline exploring biomarkers predictive of immunotherapy efficacy (4,5,7–9). Programmed death receptor 1 ligand (PD-L1) has hitherto proven to be the most reliable biomarker for this purpose (4,5). With this background, incorporation of immune information seems a logical step to further improve NSCLC prognostication. For colorectal cancer (CRC), Galon *et al*. have found the immune contexture to provide prognostic information equal or superior to the established TNM (10). More specifically, they found that an index composed of the presence or absence of tumor infiltrating lymphocytes (TILs), expressing the CD3⁺ pan-T-cell marker and/or the CD8⁺ cytotoxic T-cell marker in the tumor center and periphery, provides highly specific and sensitive prognostic information (10). As a consequence, they introduced the Immunoscore® (Immunoscore® is a registered trademark owned by INSERM) with the goal of supplementing TNM for CRC (10). A similar initiative is in place for breast cancer (11). In NSCLC, several large studies have established different immune-cell subsets as prognostic indicators, and, as in CRC, the most promising candidate markers up to date are CD3⁺, CD8⁺ and CD45RO⁺ expressing TILs (12–16).

We are currently working toward an implementation of a combined TNM-Immune cell score, a TNM-I, for NSCLC. In this study, we explore strategies for implementing a TNM-I for NSCLC, leveraging previous semi-quantitative as well as current digital, evaluation of CD3, CD4, CD8, CD20 and CD45RO in tissue micro-arrays (TMAs) of a cohort of 553 retrospectively collected primary tumors from NSCLC patients.

Materials and methods

Patients and clinical samples

The study population consisted of 553 consecutive NSCLC patients resected in 1990–2010 at the University Hospital of North-Norway (N = 295) and the Nordland Central Hospital (N = 258). Detailed information on the study population is previously published (17).

Tissue micro-array construction and immunohistochemistry

All tissue samples were reviewed by an experienced pathologist. The most representative areas containing tumor tissue was marked on the hematoxylin and eosin slide and sampled for TMA blocks. The TMAs were assembled using a tissue-arraying instrument (Beecher Instruments, Silver Springs, MD). The methodology is well documented (12,14,18). Briefly, a total of two stroma and tumor cores were sampled for each patient, respectively. The staining methods for immunohistochemistry utilized in this study are all validated and previously published (12,14).

Semi-quantitative scoring of immunohistochemistry

Representative and viable tissue sections were reviewed using a light microscope. The stromal and tumor compartments were scored for CD3⁺, CD4⁺, CD8⁺, CD20⁺ and CD45RO⁺ separately. For scoring TILs in the tumor compartment, a five-category scale with the following levels according to the percentage of positive cells was used: 0: <1%, 1: 1%–5%, 2: 6%–25%, 3: 26%–50%, 4: >50%. For scoring TILs in the stromal compartment, a four-category scale with the following levels according to the percentage of positive cells was used for all markers: 0: 0%–5%, 1: 6%–25%, 2: 26%–50%, 3: 50%. An exception was CD20 in the stromal compartment, for which the previously described five-category scale was used. When assessing a given core, the observers were blinded to each other's scores, clinical variables and patient's outcome.

Digital scoring of immunohistochemistry

TMA slides were digitized using a Pannoramic 250 Flash III (3DHistech, Budapest, Hungary) slide scanner and processed in QuPath v.0.1.3 (Queen's University, Belfast, Northern Ireland). TMA slides were dearrayed and preprocessed according to Bankhead *et al*. (19). After dearraying, TMAs were manually curated. Cell detection was conducted using QuPath's built-in 'Positive cell detection' (19). Figure 1 shows example cores for each marker, with and without overlay. Parameters for 'Positive cell detection' were tuned separately for each marker. Batch scripts are available in supplementary materials and methods. For each core, the total number of positive cells, irrespective of localization, was divided by the tissue area yielding positive cells per mm^2.

Statistical methods

All statistical analyses and visual representations were created using RStudio 1.2.5019 with R version 3.6.3 (20) and libraries 'survival' (21), 'plyr' (22), 'ggplot2' (23), 'grid', ' gridExtra' (24), 'irr' (25), 'cowplot', 'ggsignif' and 'reshape2' (26).

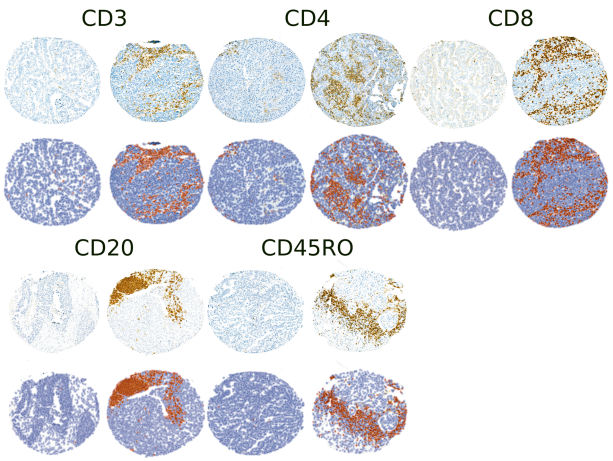

Figure 1. Low and high scores, with and without overlays for positive/negative cells, for CD3, CD4, CD8, CD20 and CD45RO.

Interobserver reliability was calculated using a two-way random-effects model with an absolute agreement definition and Cohen's kappa coefficient with equal weighting. Correlations were calculated using Pearson and Spearman methods for continuous and discrete variables, respectively. *T*-tests were utilized to examine the differences of continuous variables across groups. The Mann–Whitney *U* test was used to examine the association between distribution of different lymphocyte subsets across pathological stages.

Univariable survival analyses were conducted using the Kaplan–Meier method. Statistical difference between survival curves was assessed by the log-rank test. Disease-specific survival (DSS) was defined as the time from surgery to cancer-related death. Multivariable analyses, using the Cox proportional hazards model, were carried out to assess the independent value of pretreatment variables in the presence of other variables. Only variables with $P < 0.25$ from the univariable analyses, or deemed likely to influence outcome despite *P*-value, were explored in multivariable analyses. The significance level used was $P < 0.05$.

Ethical considerations

This study was approved by the Regional Committee for Medical and Health Research Ethics (Northern Norway, UNN: protocol ID: 2011/2503) and the need for patient consent waived. The National Data Inspection Board approved the collection and storing of the clinical database. The reporting of clinicopathological variables, survival data and biomarker expressions was conducted in accordance with the REMARK guidelines (27).

Results

Clinicopathological variables

Detailed information of the clinicopathological variables for the 553 patients in the population is already published (17). Briefly, the median age at time of diagnosis was 67 years and the median survival of survivors was 86 months. The population consisted of 32% females and the histological distribution was 307 lung squamous cell carcinomas (LUSC), 239 lung adenocarcinomas (LUAD) and 7 other. One hundred and sixty-eight patients received either adjuvant radiotherapy or chemotherapy or both.

Expression of CD3⁺, CD4⁺, CD8⁺, CD20⁺ and CD45RO⁺ TILs, their interobserver variability and correlations between digital and semi-quantitative scores

CD3, CD4, CD8. CD20 and CD45RO were expressed on the surface and in the cytoplasm of immune cells (Figure 1). All antibodies performed to expectation (Figure 1). Interestingly, CD4, which was considered to perform poorly during visual scoring, was easy to interpret in the digital domain.

Between-scorer agreements for semi-quantitative scores are summarized in Supplementary Table S1A. The ICC scores were generally higher for stromal (0.87–0.94) versus tumor scores (0.64–0.81). This observation was also reflected in the between-scorer Kappa values (0.62–0.77 stromal scores; 0.40–0.54 tumor scores).

Spearman's rank correlations between digital and semi-quantitative scores are summarized in Supplementary Table S1B. Since the semi-quantitative scoring was conducted in a microscope without annotation of the core of origin, only the final score for each patient, and not the score for each individual core, could be compared. Correlation coefficients were higher for digital and stromal scores (0.68–0.79) compared with digital and tumor scores (0.40–0.61).

Cutoff exploration and selection

All patients received a total of four semi-quantitative scores. Seventeen and thirteen possible dichotomized cutoffs could be obtained for each marker in the tumor (five-tiered scale) and the stromal (four-tiered scale) compartments, respectively. To visualize the prognostic impact of the investigated markers in relation to DSS, all potential dichotomized cutoffs from the semi-quantitative and digital analyses were plotted against *P*-values (Supplementary Figure S1). Due to superior prognostic information (Supplementary Figure S1), absolute reproducibility and ease of use, we decided to use digital counts in any subsequent analyses.

All potential cutoffs for digital TIL counts in the overall cohort, and in the LUSC and LUAD subgroups, are visualized in Figure 2. For simplicity of interpretation, dichotomization of the digital TIL counts was based on the cutoff yielding the lowest *P*-value in the overall cohort and then rounded to the nearest number divisible by 50. The final cutoffs, in cells/mm^2, were as follows: CD3 ≤ 1000; CD4 ≤ 550; CD8 ≤ 500; CD20 ≤ 400; CD45RO ≤ 250. The most discriminating prognostic information was found for LUSC patients, which translated into the prognostic differences we see for most markers in the overall cohort. For LUAD, there was hardly any significant values for any of the examined markers.

CD3⁺, CD4⁺, CD8⁺, CD20⁺ and CD45RO⁺ TILs and their correlations with clinicopathological variables

The correlations between clinicopathological variables and the dichotomized markers are presented in Table 1. To summarize, only a few reliable and significant correlations were observed. Female gender was associated with high CD3 ($P = 0.004$), increasing age was associated with high CD4 ($P = 0.029$) and increasing pStage was associated with CD4 ($P = 0.001$), CD8 ($P = 0.015$) and CD45RO ($P = 0.029$).

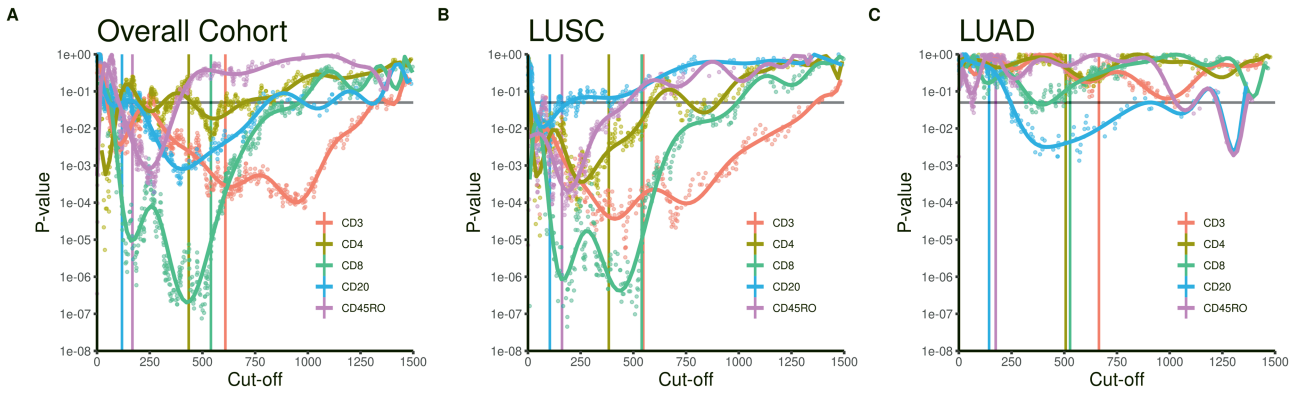

Figure 2. All possible dichotomized cutoffs for CD3, CD4, CD8, CD20 and CD45RO plotted against *P*-values indicating significance of DSS for all patient (**A**) and in the LUSC and LUAD histological subgroups (**B** and **C**). Vertical lines and numbers printed on the plot represents median and optimal cutoff values for each marker, respectively. For ease of interpretation, curves are slightly smoothened whenever possible (small dots represent the actual data points). DS, digital score; SQ, semi-quantitative.

Table 1. Digitally quantified CD3, CD4, CD8, CD20 and CD45RO density (cells/mm²) and their distribution and correlation with clinicopathological variables (chi-square and Fisher's exact tests)

	CD3			CD4			CD8			CD20			CD45RO		
	≤1000	>1000	*P*	≤550	>550	*P*	≤500	>500	*P*	≤400	>400	*P*	≤250	>250	*P*
Age			0.624			**0.029**			0.856			0.573			0.263
≤65	159	62		146	74		101	121		178	45		149	72	
>65	225	78		173	133		141	161		253	55		190	115	
Gender			**0.004**			0.788			0.783			0.122			0.946
Female	112	60		103	70		77	94		135	40		110	62	
Male	272	80		216	137		165	188		296	60		229	125	
Weight loss			0.791			0.328			0.082			0.936			0.699
<10%	346	127		283	189		211	259		388	90		303	169	
>10%	38	12		36	17		31	22		43	9		36	17	
Smoking			0.589			0.635			0.096			**0.002**			0.882
Never	11	6		9	9		10	8		15	3		12	6	
Present	242	91		204	129		142	191		260	78		212	121	
Previous	131	43		106	69		90	83		156	19		115	60	
ECOG			**0.029**			0.259			0.166			0.106			0.012
0	215	96		181	131		133	178		244	68		186	125	
1	139	38		116	61		91	87		154	28		124	55	
2	30	6		22	15		18	17		33	4		29	7	
Histology			**0.040**			0.118			0.574			0.573			0.109
LUSC	221	72		191	104		136	155		245	52		188	106	
LUAD	160	64		124	101		105	121		180	47		147	78	
LULCC	0	3		1	1		0	3		2	1		0	3	
LUADSC	2	1		2	1		1	2		3	0		3	0	
NOS	1	0		1	0		0	1		1	0		1	0	
tStage			0.653			**0.031**			0.270			0.270			0.342
T1a	8	3		6	7		7	4		9	4		9	3	
T1b	44	23		39	29		30	38		53	16		46	22	
T1c	67	25		49	43		41	50		69	23		58	34	
T2a	89	34		74	48		53	71		103	22		73	48	
T2b	53	19		39	33		28	43		59	13		40	32	
T3	75	25		67	35		48	52		86	15		73	29	
T4	48	11		45	12		35	24		52	7		40	19	
nStage			0.397			0.354			0.270			0.115			0.115
N0	258	102		216	146		163	197		288	77		229	133	
N1	86	24		66	45		49	62		98	14		69	42	
N2	40	14		37	16		30	23		45	9		41	12	
pStage			0.118			**0.001**			**0.015**			0.067			**0.029**
I	150	66		124	94		95	121		168	51		138	78	
II	129	47		99	80		73	105		150	31		105	74	
III	105	27		96	33		74	56		113	18		96	35	
Differentation			0.071			0.907			0.715			0.676			0.235
Well	149	70		128	87		97	122		182	39		132	87	
Intermediate	179	53		144	91		108	122		186	48		155	76	
Poor	56	17		47	29		37	38		63	13		52	24	
Vasc+			1.000			0.464			0.238			0.249			1.000
No	315	113		257	173		192	237		349	86		278	153	
Yes	68	25		60	33		48	44		80	13		59	33	

P-values below 0.05 are highlighted in bold. CD, cluster of differentiation; ECOG, eastern cooperative oncology group; NOS, no otherwise specified.

Table 2. Digitally quantified CD3, CD4, CD8, CD20 and CD45RO density (cells/mm^2) and a proposed TNM-Immune cell score (a combination of CD8 TIL density and pStage) as predictors of disease-specific survival in resected NSCLC patients and in subgroups with squamous cell carcinoma and adenocarcinoma (univariable analyses, log-rank test, ***n*** = 553, 307 and 239, respectively)

	All patients					LUSC					LUAD				
	N (%)	5 years	Median	HR (95% CI)	*P*	*N* (%)	5 years	Median	HR (95% CI)	*P*	*N* (%)	5 years	Median	HR (95% CI)	*P*
CD3					**<0.001**					**0.002**					0.071
≤1000	384 (69)	53	83	1.000		221 (72)	59	NA	1.000		160 (67)	47	57	1.000	
>1000	140 (25)	72	235	0.53 (0.4–0.71)		72 (23)	83	235	0.44 (0.29–0.67)		64 (27)	60	NA	0.66 (0.44–1)	
Missing	29 (5)					14 (5)					15 (6)				
CD4					**0.007**					**0.009**					0.206
≤550	319 (58)	54	98	1.000		191 (62)	58	235	1.000		124 (52)	49	57	1.000	
>550	207 (37)	66	NA	0.68 (0.51–0.89)		104 (34)	78	NA	0.56 (0.37–0.83)		101 (42)	55	NA	0.78 (0.53–1.14)	
Missing	27 (5)					12 (4)					14 (6)				
CD8					**<0.001**					**<0.001**					**0.04**
≤500	242 (44)	46	47	1.000		136 (44)	47	41	1.000		105 (44)	45	51	1.000	
>500	282 (51)	70	235	0.48 (0.36–0.63)		155 (50)	81	235	0.35 (0.24–0.53)		121 (51)	56	NA	0.68 (0.46–0.99)	
Missing	29 (5)					16 (5)					13 (5)				
CD20					**<0.001**					**0.046**					**0.003**
≤400	431(78)	55	98	1.000		245 (80)	62	235	1.000		180 (75)	45	54	1.000	
>400	100 (18)	76	NA	0.48 (0.34–0.66)		52 (17)	77	NA	0.55 (0.34–0.89)		47 (20)	74	NA	0.41 (0.26–0.64)	
Missing	22 (4)					10 (3)					12 (5)				
CD45RO					**<0.001**					**0.002**					0.173
≤250	339 (61)	54	83	1.000		188 (61)	58	127	1.000		147 (62)	49	57	1.000	
>250	187 (34)	67	NA	0.6 (0.46–0.79)		106 (35)	77	NA	0.51(0.34–0.75)		78 (33)	54	NA	0.76 (0.51–1.11)	
Missing	27 (5)					13 (4)					14 (6)				
TNM-I					**<0.001**					**<0.001**					**<0.001**
Fav	121 (22)	83	235	1.000		62 (20)	89	235	1.000		55 (23)	76	NA	1.000	
Int-Fav	166 (30)	69	190	2.03 (1.44–2.84)		95 (31)	76	NA	2.24 (1.36–3.68)		70 (29)	61	190	1.78 (1.11–2.85)	
Int-Poor	155 (28)	47	57	3.39 (2.35–4.89)		93 (30)	58	NA	3.52 (2.08–5.96)		61 (26)	33	44	3.38 (2.01–5.69)	
Poor	82 (15)	21	16	7.79 (4.7–12.92)		41 (13)	18	14	12.57 (5.68–27.82)		40 (17)	21	36	4.73 (2.46–9.1)	
Missing	29 (5)					16 (5)					13 (5)				

P-values below 0.05 are highlighted in bold. CD, cluster of differentiation; NA, not applicable.

Table 3. Multivariable models for the CD8 and CD20 TIL density (cells/mm^2) and TNM-I (a combination of CD8 TIL density and pStage) in the overall cohort and in the LUSC and LUAD subgroups (Cox proportional hazards test, n = 553, 307 and 239, respectively)

	CD8, overall		CD8, LUSC		CD8, LUAD		CD20, overall		CD20, LUSC		CD20, LUAD	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Gender												
Female	1		1		1		1		1		1	
Male	1.4 (1.03–1.89)	**0.028**	1.54 (0.93–2.56)	0.095	1.61 (1.08–2.41)	**0.019**	1.49 (1.1–2.02)	**0.011**	1.59 (0.95–2.66)	0.079	1.68 (1.12–2.52)	**0.012**
ECOG												
0	1		1				1		1			
1	1.43 (1.07–1.9)	**0.015**	1.57 (1.02–2.41)	**0.041**			1.45 (1.09–1.93)	**0.011**	1.98 (1.27–3.07)	**0.002**		
2	1.36 (0.73–2.54)	0.329	1.01 (0.4–2.59)	0.976			1.35 (0.74–2.47)	0.322	1.4 (0.58–3.36)	0.456		
Differentiation												
Poor	1		1						1			
Moderate	0.77 (0.58–1.04)	0.092	0.65 (0.42–1.01)	0.054					0.65 (0.42–1.01)	0.055		
Well	0.56 (0.35–0.92)	**0.022**	0.56 (0.25–1.26)	0.158					0.74 (0.33–1.69)	0.475		
Vascular infiltration												
No	1		1		1		1		1		1	
Yes	1.72 (1.23–2.41)	**0.002**	1.95 (1.19–3.19)	**0.008**	1.73 (1.06–2.82)	**0.027**	1.82 (1.29–2.56)	**<0.001**	2 (1.19–3.37)	**0.009**	1.65 (1.01–2.7)	**0.047**
pStage												
I	1		1		1		1		1		1	
II	1.55 (1.09–2.2)	**0.014**	1.42 (0.84–2.39)	0.190	1.93 (1.19–3.13)	**0.007**	1.6 (1.13–2.27)	**0.009**	1.49 (0.88–2.54)	0.138	1.79 (1.1–2.92)	**0.018**
III	3.36 (2.37–4.78)	**<0.001**	3.9 (2.29–6.64)	**<0.001**	3.28 (2.01–5.34)	**<0.001**	3.8 (2.69–5.38)	**<0.001**	4.53 (2.67–7.7)	**<0.001**	3.3 (2.02–5.38)	**<0.001**
CD8												
≤500	1		1		1							
>500	0.52 (0.39–0.69)	**<0.001**	0.41 (0.26–0.62)	**<0.001**	0.72 (0.49–1.07)	0.104						
CD20												
≤400							1		1		1	
>400							0.61 (0.4–0.94)	**0.025**	0.73 (0.39–1.38)	0.334	0.56 (0.3–1.03)	0.061

	CD8 TNM, overall		CD8 TNM, LUSC		CD8 TNM, LUAD	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Gender						
Female	1		1		1	
Male	1.42 (1.05–1.92)	**0.024**	1.72 (1.03–2.86)	**0.039**	1.56 (1.05–2.33)	**0.027**
ECOG						
0	1					
1	1.42 (1.06–1.89)	**0.017**				
2	1.3 (0.7–2.42)	0.412				

Table 3. Continued

	CD8 TNM, overall		CD8 TNM, LUSC		CD8 TNM, LUAD	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Differentiation						
Poor	1					
Moderate	0.8 (0.59–1.07)	0.135				
Well	0.58 (0.36–0.95)	**0.029**				
Vascular infiltration						
No	1		1		1	
Yes	1.76 (1.26–2.47)	**0.001**	2.2 (1.35–3.57)	**0.002**	1.75 (1.07–2.84)	**0.024**
CD8 TNM						
Fav	1		1		1	
Int-Fav	1.94 (1.19–3.16)	**0.007**	2.45 (1.14–5.25)	**0.022**	1.7 (0.9–3.21)	**0.102**
Int-Poor	3 (1.87–4.83)	**<0.001**	3.4 (1.61–7.19)	**0.001**	3.25 (1.74–6.05)	**<0.001**
Poor	7.41 (4.52–12.16)	**<0.001**	14.64 (6.89–31.14)	**<0.001**	4.67 (2.39–9.12)	**<0.001**

P-values below 0.05 are highlighted in bold. CD, cluster of differentiation; ECOG, eastern cooperative oncology group.

All markers were plotted against pStage to further explore the correlations between pStage and marker expression (Supplementary Figure S2). Overall, the number of positive immune cells decreased with increasing pStage. This observation was most prominent for CD4 and CD8 and between pStages I/II and III.

The differences in the total number of detected cells, and the number of positive cells for each marker, were compared in the LUSC versus LUAD subgroups. No differences in the total number of cells ($P = 0.36$) or the number of CD8 ($P = 0.56$), CD20 ($P = 0.57$) and CD45RO ($P = 0.87$) positive cells were observed. A tendency toward more positive cells in the LUAD subgroup was observed for CD3 ($P = 0.06$) and CD4 ($P = 0.04$).

Survival analyses

Univariable survival analyses of the investigated markers are summarized in Figure 2 and Table 2. Briefly, CD3, CD4, CD8, CD20 and CD45RO were all significant positive prognosticators of DSS in the overall cohort (P-values = <0.001, 0.007, <0.001, <0.001 and <0.001, respectively) and in the LUSC subgroup (P-values = 0.002, 0.009, <0.001, 0.046 and 0.002, respectively), whereas only CD8 and CD20 (P-values = 0.04 and 0.003) were significant positive prognosticators of DSS in the LUAD subgroup.

Based on the results from the cutoff explorations and univariable survival analyses, CD8 and CD20, both with excellent immunohistochemistry antibodies (Figure 1), excellent prognostication in the overall cohort and in the LUSC subgroup and reasonable prognostication in the LUAD subgroup (Table 2; Figure 2) and within most pStages (Supplementary Figures S3 and S4), were selected for multivariable analyses alongside key clinicopathological variables. Multivariable survival models are presented in Table 3.

High number of CD8 TILs was an independent significant prognosticator of favorable prognosis in the overall cohort [hazard ratio (HR) 0.52, 95% confidence interval [CI] 0.39–0.69, $P < 0.001$] and in the LUSC subgroup (HR 0.41, 95% CI 0.26–0.62) $P < 0.001$), but not in the LUAD subgroup (HR 0.72, 95% CI 0.49–1.07, $P = 0.104$).

High number of CD20 TILs was an independent prognosticator of favorable prognosis in the overall cohort (HR 0.61, 95% CI 0.40–0.94, $P = 0.025$), but not in the LUSC subgroup (HR 0.73, 95% CI 0.39–1.38, $P = 0.334$). In the LUAD subgroup (HR 0.56, 95% CI 0.30–1.03, $P = 0.06$), a tendency toward significance was observed.

TNM-Immune cell score for NSCLC

Building on the results from the multivariable analyses, the integration of pStage and CD8 TIL level was explored. Figure 4A, D and G shows 5-year survival of patients, in the overall cohort and the LUSC and LUAD subgroups, stratified by pStage and $CD8^+$ TILs. To build an integrated CD8 TNM, or TNM-I, groups with ≥80%, 60%–79%, 40%–59% and <40% five-year survival in the overall cohort were considered favorable (Fav), intermediate-favorable (Int-Fav), intermediate poor (Int-Poor) and poor (Poor), respectively. Results are presented in Table 2 (univariable) and Table 3 (multivariable). The TNM-I was an independent significant prognosticator for favorable prognosis in the overall cohort and in both the LUSC and LUAD subgroups.

Discussion

The immune contexture as a prognostic and/or predictive marker in NSCLC

In this article, we have explored the immune contexture in surgical specimens from NSCLC patients. Based on a digital,

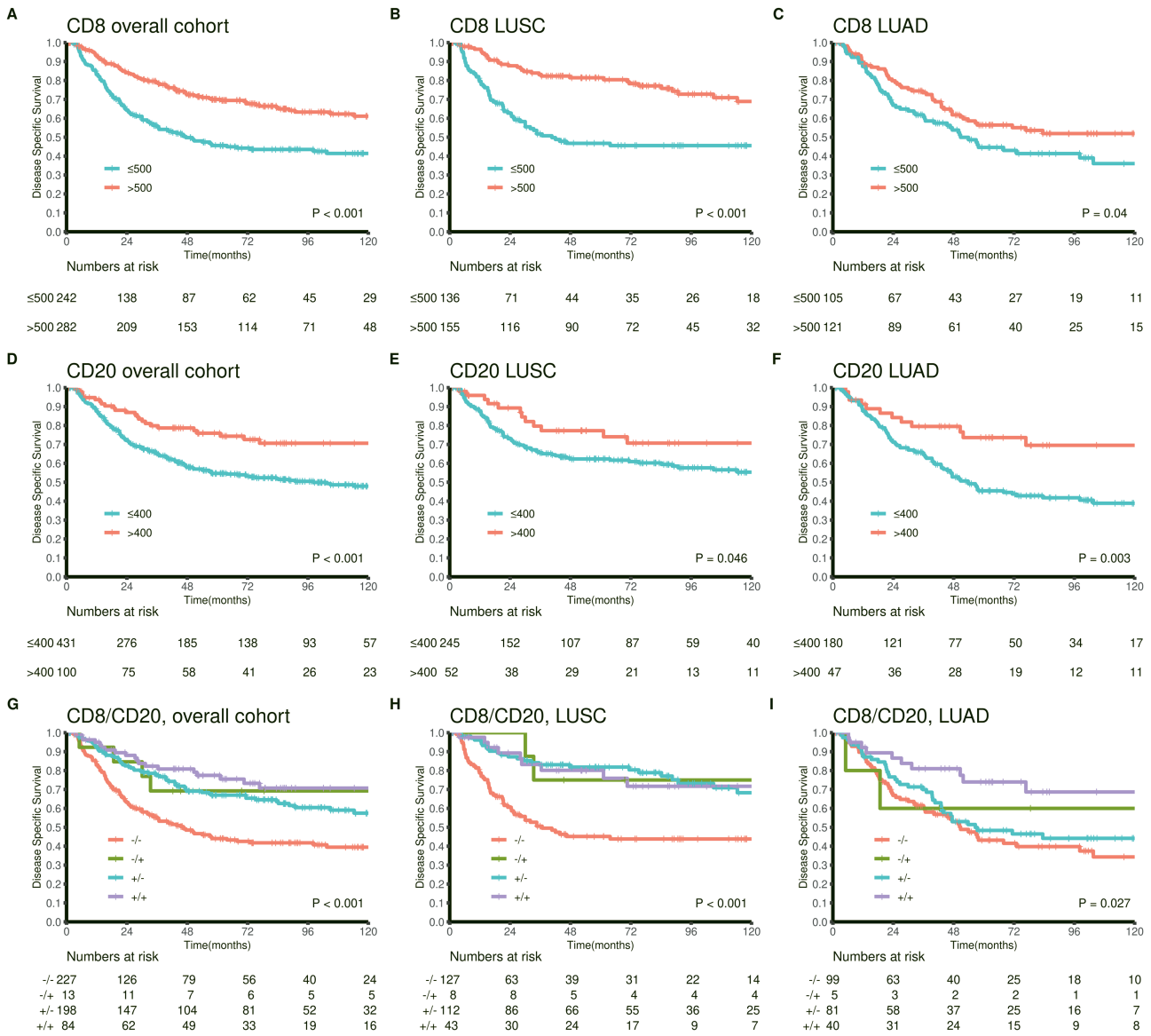

Figure 3. Disease-specific survival curves for ≤500 versus >500 CD8 TILs per mm² and ≤400 versus >400 CD20 TILs per mm² and their combination in (**A**, **D** and **G**) all patients; (**B**, **E** and **H**) the LUSC subgroup; (**C**, **F** and **I**) the LUAD subgroup.

quantitative evaluation of the main subsets of tumor infiltrating lymphocytes, in combination with pStage, we have shown that a TNM-Immune cell score for NSCLC patients is feasible. The resulting score yields increased prognostic accuracy at the cost of a fairly simple diagnostic test. As in our previous studies, the prognostic impact of the TNM-I is considerably stronger in LUSC compared with LUAD patients.

Automated digital versus manual quantification of TILs

In previous studies, we have utilized a semi-quantitative scoring approach (12,14). In these studies, immune cells in different compartments (intraepithelial versus stroma) were considered separately. Using this approach, immune cells in intratumoral stroma was found to be the superior prognosticator. In the present study, a digital scoring approach was applied and compared with semi-quantitative scores. As expected, the digital and semi-quantitative scores were significantly correlated (Supplementary Table S1). However, the continuous variables provided by digital quantification allowed fine tuned cutoff selection, which translated into additional prognostic information (Supplementary Figure S1) and superior multivariable models. Surprisingly, no difference in the prognostic impact of TILs was observed for the intraepithelial versus stromal compartments (data not shown). Furthermore, once established, digital quantification has better reproducibility and allows higher throughput. Hence, digital scores were considered superior to semi-quantitative scores.

Choosing candidate markers for an Immune cell score for NSCLC patients

Given an optimized pre-analytic workflow, candidate markers for a clinically used test utilizing immunohistochemistry should be subject to careful antibody selection, and optimization on an automated platform for ease of use (28). The markers under consideration should show significant discrimination between study groups and exhibit excellent reproducibility (28). CD3,

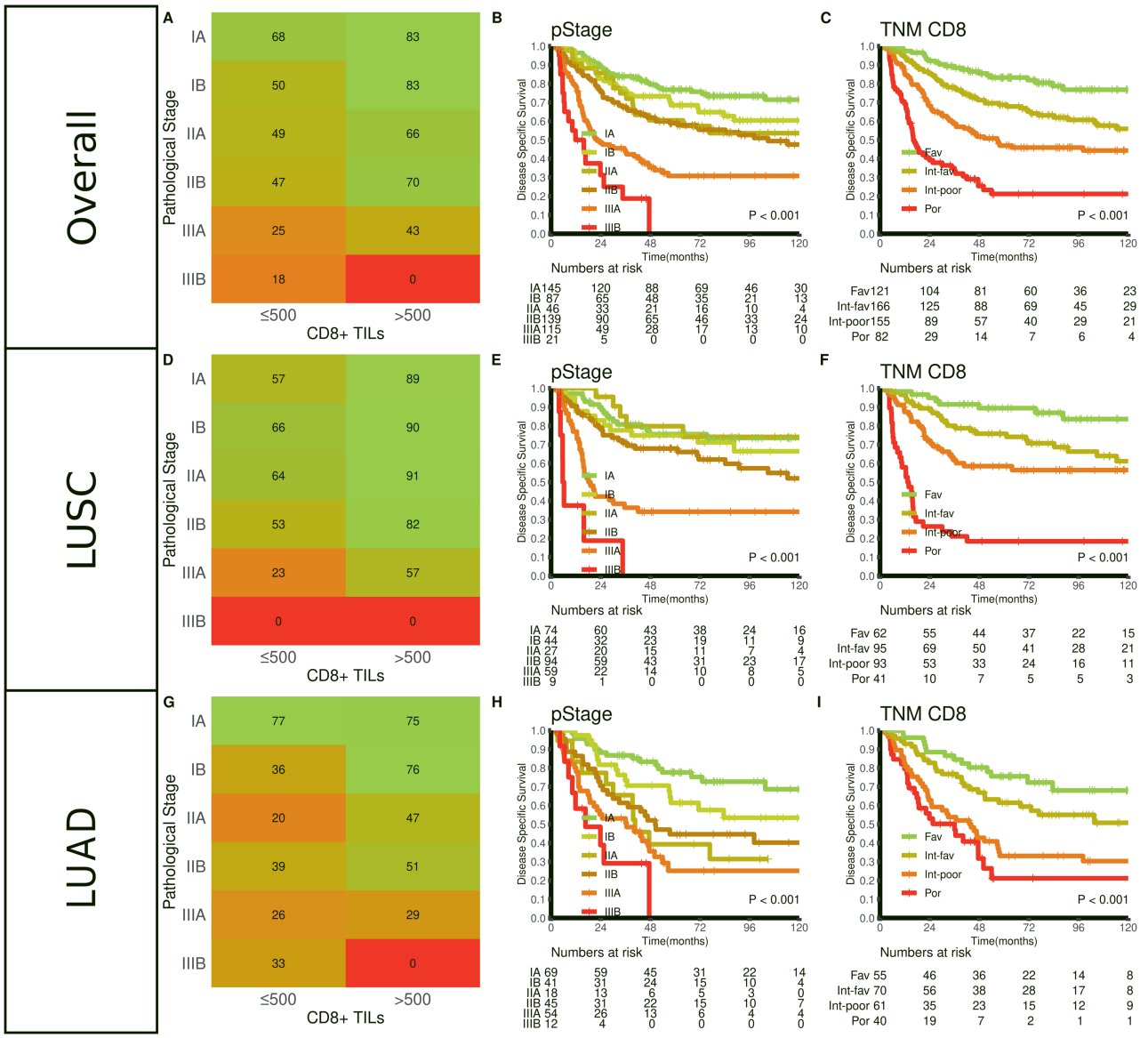

Figure 4. Five-year disease-specific survival in each pStage stratified by CD8⁺ TILs and color coded into distinct subgroups with similar survival and disease-specific survival curves according to pStage and to a proposed TNM-Immune cell score for all patients (**A–C**), the LUSC subgroup (**D–F**) and the LUAD subgroup (**G–I**).

CD4, CD8, CD20 and CD45RO are all well validated and in routine clinical use. All markers exhibited stellar staining (Figure 1).

Comparing DSS data (Table 2; Supplementary Figures S1 and S3), CD3⁺, CD8⁺ and CD45RO⁺ TILs are potential candidate markers for a TNM-I in the overall cohort, and in the LUSC subgroup, while CD8⁺ and CD20⁺ TILs are most promising in the LUAD subgroup. Moreover, CD8/CD20 combinations were explored using survival curves (Figure 3G–I; Supplementary Figure S4J–L). From these figures, it is evident that no additional benefit is gained from combining CD8/CD20 versus CD8 alone. For LUSC and LUAD, the observed survival differences rely almost solely on CD8⁺ and CD20⁺ TILs, respectively. Similar observations were made when exploring other possible combinations (data not shown). In multivariable analyses, CD8⁺ TILs was an independent positive prognosticator in the overall cohort and in the LUSC subgroup, whereas CD20 was near significant in the LUAD subgroup. CD8⁺ TILs is the most promising candidate for at TNM-I from our previous studies and corroborated by the Immunoscore® for CRC (10,13,16,29). In addition, the optimal cutoff for CD8⁺ TILs was near median, yielding a sufficient number of patients in the high and low groups for the combination with pStage. Hence, CD8⁺ TILs were considered the best candidate marker for a TNM-I in NSCLC.

The clinical utilization of a TNM-Immune cell score for NSCLC patients

This study demonstrates, as a proof-of-concept, that digitally quantified CD8⁺ TIL density easily incorporates with pStage to form an TNM-Immune cell score for NSCLC. As a supplement to the established prognostic toolbox, the TNM-I could help identify patients in need of adjuvant treatment or, more importantly, those who are not. As expected, the group with a favorable TNM-I (pStages IA and IB with a high CD8⁺ TIL density) exhibit excellent 5- and even 10-year disease-specific survival in our cohort (Figure 4; Tables 2 and 3; and Supplementary Figure

S4). More interesting, are the results for patients classified as TNM-I intermediate-favorable: no difference in DSS is observed between patients in pStage IA with low CD8+ TIL density and pStage IIA and B with high CD8+ TIL density. As pStage II is the recommended threshold for adjuvant chemotherapy (30), the TNM-I may force us to rethink this practice, either by considering adjuvant therapy to all TNM-I intermediate-favorable patients, thus including pStage IA/CD8+ low, or by moving the threshold to TNM-I intermediate-low (Figure 4; Supplementary Figure S4). Applying the TNM-I approach could potentially reduce the number needed to treat and increase adjuvant treatment efficacy. Additionally, as immunotherapy is gradually implemented in earlier stages of treatment (neo/adjuvant), it will be of great interest to see whether the TNM-I can add valuable prognostic/predictive information in this aspect as well.

Histology-specific adaptations of the TNM-I in NSCLC

The impact of CD8+ TIL density, while showing an impact in LUAD patients, is mainly found in LUSC patients, who are the main drivers of the results in the overall cohort. In addition, based on Figure 4, it is likely that a TNM-I for LUSC patients will be further improved by moving pStage II high CD8+ TIL density patients to the TNM-I favorable group. Interestingly, the mean density of some (CD3+ and CD4+), but not all (CD8+, CD20+ and CD45RO+) TILs were higher in LUAD compared with LUSC patients. This finding may indicate different immune responses in these two histological entities and warrant further investigation. For LUAD patients, although the CD8+ TIL-based TNM-I yields superior prognostication when compared with the conventional TNM, the results for CD20+ TILs are especially interesting. The observed prognostic impact of CD20+ TILs indicates that this cell type is involved in immune-activation in LUAD patients. CD20+ TILs are known to cluster in tertiary lymphoid structures. Based on recent publications in other cancer forms (31,32), it is pertinent to speculate if this result indicates similar processes in LUAD patients. However, CD20+ TILs were not significant in multivariable analysis (Table 3). A plausible explanation is that, because of their relative scarcity, the number of CD20+ TILs is not reliably estimated in TMAs. We are currently exploring the prognostic impact of tertiary lymphoid structures in NSCLC using whole slide images (to be published).

Future perspectives

The applicability of the TNM-I, presented as a result from this exploratory study, needs to be confirmed. Our research group has initiated a large multi-center study with an aim of implementing a TNM-I for NSCLC (NCT03299478). The presented TNM-I will be tested in this cohort upon study completion. If rapid implementation is warranted, confirmation in large, recent retrospective cohorts may be conducted. As suggested in the previous section, future studies should include sufficient numbers to explore histology-specific adaptations of the TNM-I.

Furthermore, in light of the recent advances in therapies directly influencing the activation and prolongation the tumor immune response, the concept of an Immune cell score gives promise beyond being a mere addition to a prognostic score. Active *in situ* immune-activation, measured using the CD8+ TIL density, in addition to or in place of PD-L1, may identify patients for whom immune-related therapy is likely to be beneficial. This could prove a pivotal clinical aspect of the TNM-I, which would benefit patients not only with regard to outcomes, but also alleviate the increasing financial burden of treatment, as recognized by ASCO (33).

Conclusions

Recently published evidence suggests that the introduction of a TNM-I for NSCLC will add meaningful prognostic stratification on the already excellent TNM system (12,13,16). We present data confirming that CD8+ TIL density is the most robust candidate for a TNM-I in NSCLC, as its incorporation improves upon the current TNM systems ability to predict disease-specific survival for NSCLC patients. Corroborating previous studies, the most pronounced prognostic effect of CD8+ TILs and the proposed TNM-I is observed in LUSC patients. However, contrary to prior studies, a non-negligible effect is also observed in LUAD patients. Furthermore, we show that CD8+ TILs may be automatically detected in digital TMAs and subsequently interpreted by freely available open source software, thus avoiding common hurdles for biomarker implementation such as additional cost and added complexity.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Figure S1. All possible dichotomized cutoffs for CD3, CD4, CD8, CD20 and CD45RO plotted against *P*-values indicating significance of disease-specific survival for digital scores (A) and semi-quantitative scores in stroma (B) and tumor (C). Vertical lines and numbers printed on the plot represents median and optimal cutoff values for each marker, respectively. For ease of interpretation, curves are slightly smoothed whenever possible (small dots represent the actual data points). Abbreviations: DS, digital score; SQ, semi-quantitative; DSS, disease-specific survival; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma

Figure S2. The median number of CD3, CD4, CD8, CD20 and CD45RO positive cells stratified by pStage (Two sided Wilcoxon rank-sum test). $^{**}P < 0.01$, $^{*}P < 0.05$.

Figure S3. All possible dichotomized cutoffs for CD3, CD4, CD8, CD20 and CD45RO plotted against *P*-values indicating significance of disease-specific survival in pStage I (A), II (B) and III (C). Dashed vertical lines represent median values for each marker, respectively. For easier interpretation all curves are slightly smoothened (small dots represent the actual data points). Abbreviations: DS, digital score; SQ, semi-quantitative; OS, overall survival; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma

Figure S4. Disease-specific survival curves for pStage I-III (panels A–C), and for ≤500 vs >500 CD8 TILs per mm^2 and ≤ 400 vs > 400 CD20 TILs per mm^2 and their combinations in pStage I-III, panels D-F, G-I and J-L, respectively. Abbreviations: TILs, tumor infiltrating lymphocytes; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma.

Acknowledgements

Thanks to Magnus L. Persson for making the TMA blocks and to Elin Richardsen and Samer Al-Saad for reclassification of the included tumor samples.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Conflict of Interest Statement: None declared.

References

1. Bray, F. *et al.* (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.*, 68, 394–424.
2. Kerr, K.M. *et al.*; Panel Members (2014) Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer. *Ann. Oncol.*, 25, 1681–1690.
3. Vansteenkiste, J. *et al.* (2013) Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.*, 24(Suppl. 6), vi89–98.
4. Garon, E.B. *et al.*; KEYNOTE-001 Investigators (2015) Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.*, 372, 2018–2028.
5. Borghaei, H. *et al.* (2015) Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *N. Engl. J. Med.*, 373(17):1627–1639.
6. Fridman, W.H. *et al.* (2012) The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer.*, 12, 298–306.
7. Khalil, D.N. *et al.* (2016) The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat. Rev. Clin. Oncol.*, 13, 394.
8. Larkin, J. *et al.* (2019) Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. *N. Engl. J. Med.*, 381, 1535–1546.
9. Hamid, O. *et al.* (2019) Five-year survival outcomes for patients with advanced melanoma treated with pembrolizumab in KEYNOTE-001. *Ann. Oncol.*, 30, 582–588.
10. Galon, J. *et al.* (2014) Towards the introduction of the "Immunoscore" in the classification of malignant tumours. *J. Pathol.*, 232(2):199–209.
11. Salgado, R. *et al.*; International TILs Working Group 2014 (2015) The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann. Oncol.*, 26, 259–271.
12. Paulsen, E.-E. *et al.* (2015) CD45RO$^+$ memory T lymphocytes – a candidate marker for TNM-Immunoscore in squamous non–small cell lung cancer. *Neoplasia*, 17(11):839–848.
13. Donnem, T. *et al.* (2016) Strategies for clinical implementation of TNM-Immunoscore in resected nonsmall-cell lung cancer. *Ann. Oncol.*, 27, 225–232.
14. Kilvaer, T.K. *et al.* (2016) The presence of intraepithelial CD45RO$^+$ cells in resected lymph nodes with metastases from NSCLC patients is an independent predictor of disease-specific survival. *Br. J. Cancer.*, 114, 1145–1151.
15. Hald, S.M. *et al.* (2013) CD4/CD8 co-expression shows independent prognostic impact in resected non-small cell lung cancer patients treated with adjuvant radiotherapy. *Lung Cancer.*, 80, 209–215.
16. Donnem, T. *et al.* (2015) Stromal CD8$^+$ T-cell density – a promising supplement to TNM staging in non–small cell lung cancer. *Clin. Cancer Res.*, 21(11):2635–2643.
17. Hald, S.M. *et al.* (2018) LAG-3 in non–small-cell lung cancer: expression in primary tumors and metastatic lymph nodes is associated with improved survival. *Clin. Lung Cancer.*, 19(3):249–259.e2.
18. Bremnes, R.M. *et al.* (2002) High-throughput tissue microarray analysis used to evaluate biology and prognostic significance of the E-cadherin pathway in non-small-cell lung cancer. *J. Clin. Oncol.*, 20, 2417–2428.
19. Bankhead, P. *et al.* (2017) QuPath: open source software for digital pathology image analysis. *Sci. Rep.*, 7, 16878.
20. R Core Team (2014) R: a language and environment for statistical computing. http://www.r-project.org/.
21. Therneau, T.M. *et al.* (2000) *Modeling survival data: extending the Cox model*. New York, NY: Springer.
22. Wickham, H. (2011) The split-apply-combine strategy for data analysis. *J. Stat. Softw*. 40(1):1–29.
23. Wickham, H. (2009) *Ggplot2: elegant graphics for data analysis*. New York, NY: Springer. http://had.co.nz/ggplot2/book.
24. Auguie, B. (2012) gridExtra: functions in Grid graphics. http://cran.r-project.org/package=gridExtra.
25. Gamer, M. *et al.* (2012) irr: Various Coefficients of Interrater Reliability and Agreement. http://cran.r-project.org/package=irr.
26. Wickham, H. (2007) Reshaping data with the reshape package. *J. Stat. Softw.*, 21(12):6–17.
27. McShane, L.M. *et al.*; Statistics Subcommittee of NCI-EORTC Working Group on Cancer Diagnostics (2006) Reporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast Cancer Res. Treat.*, 100, 229–235.
28. Lin, F. *et al.* (2014) Standardization of diagnostic immunohistochemistry: literature review and geisinger experience. *Arch. Pathol. Lab. Med.*, 138, 1564–1577.
29. Angell, H. *et al.* (2013) From the immune contexture to the immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr. Opin. Immunol.*, 25, 261–267.
30. Postmus, P.E. *et al.* (2017) Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.*, 28(Suppl. 4):iv1–iv21.
31. Helmink, B.A. *et al.* (2020) B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*, 577, 549–555.
32. Cabrita, R. *et al.* (2020) Author correction: tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*, 580, E1.
33. Schnipper, L.E. *et al.* (2016) Updating the American Society of Clinical Oncology Value Framework: revisions and reflections in response to comments received. *J Clin Oncol.*, 34(24), 2925–2934.