ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ANNALS OF ONCOLOGY driving innovation in oncology

ORIGINAL ARTICLE

Machine learning-based immune phenotypes correlate with STK11/KEAP1 co-mutations and prognosis in resectable NSCLC: a sub-study of the TNM-I trial

M. Rakaee[1,2,3*], S. Andersen[3,4], K. Giannikou[1,5], E.-E. Paulsen[3,6], T. K. Kilvaer[3,4], L.-T. R. Busund[2,7], T. Berg[2,7], E. Richardsen[2,7], A. P. Lombardi[7], E. Adib[1,8], M. I. Pedersen[3], M. Tafavvoghi[9], S. G. F. Wahl[10,11], R. H. Petersen[12,13], A. L. Bondgaard[14], C. W. Yde[15], C. Baudet[15], P. Licht[16], M. Lund-Iversen[17], B. H. Grønberg[10,11], L. Fjellbirkeland[18], Å. Helland[19,20,21], M. Pøhl[22], D. J. Kwiatkowski[1,23†] & T. Donnem[3,4†]

[1]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA; [2]Department of Clinical Pathology, University Hospital of North Norway, Tromso; [3]Department of Clinical Medicine, UiT The Arctic University of Norway, Tromso; [4]Department of Oncology, University Hospital of North Norway, Tromso, Norway; [5]Division of Hematology and Oncology, Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Los Angeles, USA; [6]Department of Pulmonology, University Hospital of North Norway, Tromso; [7]Department of Medical Biology, UiT The Arctic University of Norway, Tromso, Norway; [8]Lank Center for Genitourinary Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA; [9]Department of Community Medicine, UiT The Arctic University of Norway, Tromso; [10]Department of Oncology, St. Olav's Hospital, Trondheim University Hospital, Trondheim; [11]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway; [12]Department of Cardiothoracic Surgery, Copenhagen University Hospital, Rigshospitalet, Copenhagen; [13]Department of Clinical Medicine, University of Copenhagen, Copenhagen; [14]Department of Pathology, Copenhagen University Hospital, Rigshospitalet, Copenhagen; [15]Center for Genomic Medicine, Copenhagen University Hospital, Rigshospitalet, Copenhagen; [16]Department of Cardiothoracic Surgery, Odense University Hospital, Odense, Denmark; [17]Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo; [18]Department of Respiratory Medicine, Oslo University Hospital, University of Oslo, Oslo; [19]Department of Cancer Genetics, Institute for Cancer Research, Norwegian Radium Hospital, Oslo University Hospital, Oslo; [20]Department of Oncology, Oslo University Hospital, Oslo; [21]Department of Clinical Medicine, University of Oslo, Oslo, Norway; [22]Department of Oncology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [23]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, USA

Available online 24 April 2023

Background: We aim to implement an immune cell score model in routine clinical practice for resected non-small-cell lung cancer (NSCLC) patients (NCT03299478). Molecular and genomic features associated with immune phenotypes in NSCLC have not been explored in detail.

Patients and methods: We developed a machine learning (ML)-based model to classify tumors into one of three categories: inflamed, altered, and desert, based on the spatial distribution of CD8+ T cells in two prospective (n = 453; TNM-I trial) and retrospective (n = 481) stage I-IIIA NSCLC surgical cohorts. NanoString assays and targeted gene panel sequencing were used to evaluate the association of gene expression and mutations with immune phenotypes.

Results: Among the total of 934 patients, 24.4% of tumors were classified as inflamed, 51.3% as altered, and 24.3% as desert. There were significant associations between ML-derived immune phenotypes and adaptive immunity gene expression signatures. We identified a strong association of the nuclear factor-κB pathway and CD8+ T-cell exclusion through a positive enrichment in the desert phenotype. *KEAP1* [odds ratio (OR) 0.27, Q = 0.02] and *STK11* (OR 0.39, Q = 0.04) were significantly co-mutated in non-inflamed lung adenocarcinoma (LUAD) compared to the inflamed phenotype. In the retrospective cohort, the inflamed phenotype was an independent prognostic factor for prolonged disease-specific survival and time to recurrence (hazard ratio 0.61, P = 0.01 and 0.65, P = 0.02, respectively).

Conclusions: ML-based immune phenotyping by spatial distribution of T cells in resected NSCLC is able to identify patients at greater risk of disease recurrence after surgical resection. LUADs with concurrent *KEAP1* and *STK11* mutations are enriched for altered and desert immune phenotypes.

Key words: NSCLC, machine learning, immune phenotypes, STK11, KEAP1, NF-κB

*Correspondence to: Dr Mehrdad Rakaee, Department of Clinical Pathology, University Hospital of North Norway, 9019 Tromso, Norway. Tel: +47 77625237
E-mail: mehrdad.rakaee@uit.no (M. Rakaee).

†These authors contributed equally.

INTRODUCTION

The characteristics of the tumor microenvironment, notably immune cell subsets, have attracted the attention of researchers in the field of immuno-oncology.[1] Identifying immunotherapy biomarkers will become easier as knowledge increases about tumor and host immune-specific traits that affect prognosis and treatment responses. Wide differences in patient responses to treatment underscore the importance of predictive biomarkers for selecting optimal treatment strategies.[2] However, the current spectrum of biomarkers is extremely limited, particularly those that are studied in large prospective clinical trials.

There is strong evidence that CD8+ cytotoxic T lymphocytes (CTLs) play an important role in immune checkpoint treatment response.[3] CTLs interact with tumor-associated antigens on the target cells through the major histocompatibility complex class-I and T-cell receptor complex and can initiate apoptosis in malignant cells.[4] Effective CTL-mediated cytotoxicity requires direct contact between CTLs and tumor cells, thus it is increasingly recognized that the spatial distribution of CTLs in the tumor microenvironment may indicate, or correlate with, divergent responses to immunotherapy.[5,6] Three main phenotypes have been identified for tumor-immune classification, namely: (i) inflamed (or hot), with high levels of CTLs in the intratumoral compartment; (ii) altered, with moderate overall CTL infiltration; and (iii) desert (or cold), with none or scarce CTLs near tumor cells.[7] The association of clinical outcomes based on immune phenotypes has been documented,[8] but there is still no standardized method for defining the spatial distribution of CTL infiltrates. Furthermore, the molecular characteristics and mechanisms that shape the geographical distribution of CTLs are uncertain.

The TNM (tumor—node—metastasis) system is the gold standard for non-small-cell lung cancer (NSCLC) staging, providing guidance for the assessment of prognosis and the optimal treatment approach. However, there is significant heterogeneity in outcome within the same TNM stage.[9] Given the potential importance of spontaneous tumor immunity, as well as that induced by immune checkpoint inhibition treatment, incorporating immunological information into lung cancer staging may be an important advance in more precise cancer staging and prognostication.[10] For instance, immune scoring based on the assessment of CD8+ and CD3+ T lymphocytes in tumor subregions (intratumoral, and the tumor—stromal interface) was recently recommended for inclusion into the European Society for Medical Oncology (ESMO) Clinical Practice Guidelines for gastrointestinal cancers and the World Health Organization's (WHO) classification of digestive system tumors.[11] Moreover, there is a comparable clinical application of immune status for breast cancer.[12] We have previously reported on the utility of *in situ* immune components to predict outcomes after tumor resection in patients with early-stage NSCLC tumors,[13,14] and on responses to immunotherapy in advanced-stage tumors.[15,16]

Therefore, our intent was to develop a tumor-immune phenotype classifier and to test it in a prospective Scandinavian trial (NCT03299478; TNM-I). Recognizing the limitations of manual assessment of tumor-immune phenotype, including time requirement and large interobserver variability,[17] we pursued a machine learning (ML)-based approach. This ML method was used for categorizing tumor-immune phenotypes in patients with stage I-IIIA NSCLC in the TNM-I trial, similar to, but extending our previous work.[15] In addition, we examined the potential association of tumor-immune phenotype with genomic alterations and prognosis.

PATIENTS AND METHODS

The current study comprises 934 patients with resected stage I-IIIA NSCLC. The patients were from two independent cohorts: a multi-institutional Scandinavian clinical trial (NCT03299478; TNM-I, $n = 453$; Copenhagen and Odense in Denmark and Tromso and Trondheim in Norway), and a retrospective collection from centers in the Northern Norway health region (UNN, $n = 481$).

In the TNM-I cohort, patients were prospectively enrolled between August 2016 and February 2022. A web-based case report form (REDCap) was used for collecting histopathological and clinical data. In the UNN cohort, patients were operated on between 1990 and 2010.[13] In the TNM-I cohort, tumor staging and classification was in accordance with the latest guidelines [TNM American Joint Committee on Cancer (AJCC) eighth edition; WHO 2015]. Patients from the UNN cohort were restaged and reclassified according to this newer staging system.

DNA ($n = 215$) and RNA ($n = 132$) analyses were carried out using the TruSight Oncology 500 HT (TS0500, Illumina, San Diego, CA; Supplementary Figure S1, available at https://doi.org/10.1016/j.annonc.2023.04.005) and Pan-Cancer Immuno-Oncology 360 (IO360, NanoString, Seattle, WA) panels, respectively, in the TNM-I dataset.

The immune phenotyping of CD8/pan-cytokeratin (pCK) stained whole-tissue images was based on supervised ML models using QuPath (v.0.2.3)[18] on developmental (TNM-I) and validation (UNN) cohorts. Figure 1A summarizes the workflow used in the development of the algorithm. Training and evaluation of the ML model was conducted on 453 images of the TNM-I clinical trial, and tested further on the retrospective UNN cohort. The UNN cohort employed the same algorithms developed in the TNM-I set, with no modifications. For phenotyping, the cut-off was based on the lower (88 cells/mm^2) and upper (814 cells/mm^2) quartile values of CD8+ cell density in intratumoral and stromal compartments, respectively, in the entire material.[8] Detailed methods are reported in the Supplementary Material, available at https://doi.org/10.1016/j.annonc.2023.04.005.

RESULTS

Tissue classifier model development and immune phenotyping

A total of 934 stage I-IIIA NSCLC patients who had surgical resection as primary therapy were studied from one

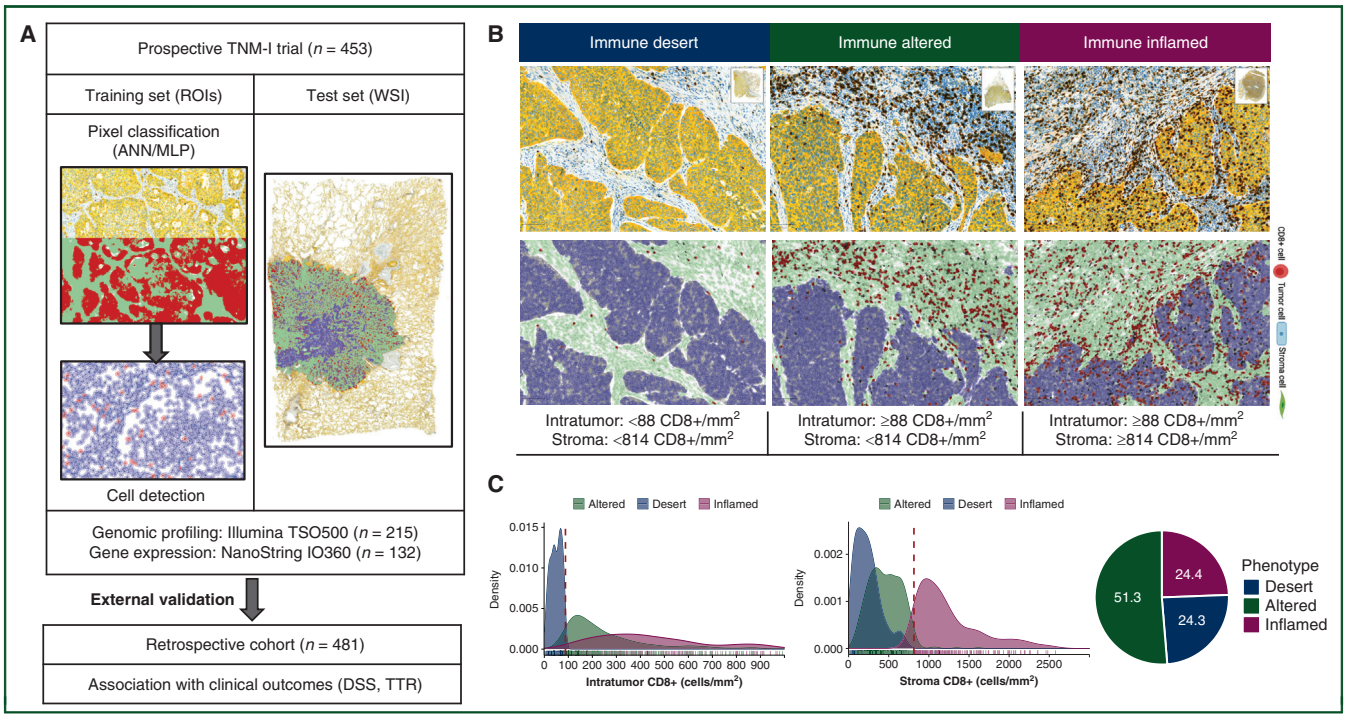

Figure 1. Machine learning (ML)-based immune phenotyping of whole-slide images (WSI) in non-small-cell lung cancer (NSCLC).
(A) The workflow of the developed ML model in the TNM-I cohort ($n = 453$) and further validation in the UNN ($n = 481$) dataset. (B) Example images of representative desert, altered, and inflamed *in situ* immune phenotypes are shown based on the spatial distribution of CD8+ in the stromal and intratumoral compartments. The images in the upper row show immunohistochemistry (IHC) co-staining for CD8/pan-cytokeratin (brown/yellow). Images of classified tissue with ML overlay are shown in the lower row. The stromal area/cells are represented by the green color. The tumor epithelial area/cells are represented by the blue color. The CD8+ cytotoxic T lymphocytes (CTL) are represented by the red color. (C) Range of CD8+ density in different subregions [intratumoral (red dashed line = Q1, 88 CD8+/mm^2) and stromal (red dashed line = Q3, 814 CD8+/mm^2)]. The distribution of immune phenotypes across the entire cohort ($n = 934$) is shown in a pie chart.
ANN, artificial neural networks; MLP, multilayer perceptron; ROIs, regions of interest.

prospective (TNM-I; 2016-2022) and one retrospective (UNN; 1990-2010) cohort. Baseline patient and tumor characteristics of the TNM-I and UNN cohorts were different, due to major changes in patient demographics and behaviors during collection (Table 1). For example, the contemporary TNM-I cohort had a higher proportion of patients with stage I NSCLC (54% versus 43%), former/never filtered cigarette/loose tobacco smokers (69% versus 37%), adenocarcinoma histology (68% versus 44%), and females (56% versus 34%), compared to the historical UNN cohort, as well as a higher median age at diagnosis (77 versus 68 years).

An ML-based pixel classifier was designed and trained to segregate tissue into two distinct compartments: stromal and intratumoral. To confirm the accuracy of the pixel classifier, a comparison was carried out with manual region annotation by a pathologist (ER) on a randomly selected subset ($n = 100$, excluding the training set). The intraclass correlation coefficients (ICCs) of CD8+ cell counts between automated and manual tissue segments were 0.99 for stromal subregions and 0.98 for intratumoral subregions (Supplementary Figure S2, available at https://doi.org/10.1016/j.annonc.2023.04.005).

In both cohorts ($n = 934$), and in intratumoral and stromal subregions, the median [interquartile range (IQR)] number of CD8+ cells was 186 (88-382) and 485 (266-814) per mm^2, respectively (Supplementary Figure S3, available at https://doi.org/10.1016/j.annonc.2023.04.005). Samples were classified into three immune phenotypes by spatial distribution of CD8+ cells (Figure 1B). The fraction of samples classified as inflamed, altered, and desert phenotypes was: 15%, 58.1%, and 26.9% for the TNM-I cohort; 33.5%, 44.7%, and 21.8% for the UNN cohort (Supplementary Figure S3, available at https://doi.org/10.1016/j.annonc.2023.04.005); and 24.4%, 51.3%, and 24.3% for the combined cohort (Figure 1C).

Immune phenotyping characterization by gene expression

ML-derived immune phenotypes were compared with immune gene expression profiles for 132 patients from the TNM-I cohort using whole-tissue sections. The profiles were assessed using a NanoString panel (IO360) comprising 770 immune-specific genes and 39 signatures (Supplementary Figure S4, available at https://doi.org/10.1016/j.annonc.2023.04.005). The adaptive cell subset signature scores were significantly increased stepwise from desert to inflamed subtypes (Figure 2A). For innate immunity signatures, macrophage scores ($P < 0.01$) and dendritic cell scores ($P < 0.001$) showed an association with immune phenotypes, while mast cell, neutrophil, and natural killer cell scores did not (Supplementary Figure S5, available at https://doi.org/10.1016/j.annonc.2023.04.005). In addition, macrophage/tumor-infiltrating lymphocyte (TIL) and neutrophil/TIL score ratios were highest in desert tumors and lowest in inflamed subgroups (Supplementary

Table 1. Baseline characteristics of the TNM-I and UNN cohorts for 934 resected stage I-IIIA NSCLC patients

	TNM-I cohort[a] $n = 453$ (%)	UNN cohort[b] $n = 481$ (%)	P[c]	Total $n = 934$ (%)
Cohort type	Prospective	Retrospective		
Inclusion time	2016-2022	1990-2010		
pStage			**0.004**	
Stage I	243 (54)	207 (43)		450 (48)
Stage II	127 (28)	157 (33)		284 (30)
Stage IIIA	83 (18)	117 (24)		200 (22)
pN status			0.06	
N0	342 (76)	341 (71)		683 (73)
N1 or N2	105 (23)	140 (29)		245 (26)
Nx	6 (1)			6 (1)
pT status			**0.002**	
T1	191 (42)	181 (38)		372 (40)
T2	171 (38)	167 (35)		338 (36)
T3	62 (14)	83 (17)		145 (15)
T4	29 (6)	50 (10)		79 (8)
Age, median (range)	77 (40-86)	68 (39-85)	0.2	68 (39-86)
Gender			**<0.001**	
Female	247 (56)	163 (34)		410 (44)
Male	206 (45)	318 (66)		524 (56)
Histology findings			**<0.001**	
LUAD	308 (68)	213 (44)		521 (56)
LUSC	126 (28)	261 (54)		387 (41)
Other[d]	19 (4)	7 (2)		26 (3)
LUAD subtype			**0.002**	
Solid	83 (27)	85 (40)		168 (32)
Acinar	141 (46)	74 (35)		215 (41)
Papillary	30 (10)	36 (17)		66 (13)
Micropapillary	10 (3)	15 (7)		25 (5)
Lepidic	22 (7)	3 (1)		25 (5)
Other[e]	22 (7)			22 (4)
Smoking			**<0.001**	
Current	139 (31)	304 (63)		443 (47)
Former/never	314 (69)	177 (37)		491 (53)
ECOG			**0.02**	
0	332 (73)	279 (58)		611 (65)
1	99 (22)	169 (35)		268 (29)
2	18 (4)	33 (7)		51 (5)
Unknown	4 (1)			4 (1)
Surgery type			**<0.001**	
Pulmonectomy	6 (1)	113 (23)		119 (13)
Lobectomy	436 (96)	346 (72)		782 (83)
Wedge	11 (3)	22 (5)		33 (4)
PD-L1 (TPS %)		NA		
<1	113 (25)			113 (25)
1-49	152 (34)			152 (34)
≥50	138 (30)			138 (30)
Unknown	50 (11)			50 (11)

Significant P-values in bold ($P < 0.05$).
ECOG, Eastern Cooperative Oncology Group; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; NA, not assessed; PD-L1, programmed death-ligand 1; pN, pathological node; pStage, pathological staging; pT, pathological tumor; TPS, tumor proportion score.
[a]Including samples from the following centers: Tromso, Trondheim, Odense, and Copenhagen.
[b]Including samples from the following centers: Tromso and Bodø.
[c]P values were based on χ^2 test.
[d]Including adenosquamous carcinoma, large cell carcinoma, and sarcomatoid.
[e]Including mucinous, fetal, and colloid adenocarcinoma.

Figure S5, available at https://doi.org/10.1016/j.annonc.2023.04.005).

We found no association between immune phenotypes and immune checkpoint gene expression (including *PDCD1*, *LAG3*, *TIGIT*, *CTLA4*, *CD274*, *PDCD1LG2*, *HAVCR2*, *IDO1*, and *VSIR*; Supplementary Figure S6, available at https://doi.org/10.1016/j.annonc.2023.04.005). In contrast, immunohistochemistry (IHC) assessment of programmed death-ligand 1 (PD-L1) expression showed that high (≥50%) PD-L1 tumors had a far greater frequency of the inflamed phenotype (25%) than tumors with either no PD-L1 expression (8%) or PD-L1 1%-49% (10%; χ^2 $P < 0.001$; Figure 2B).

We then carried out NanoString pathway analysis to identify biological processes that were substantially changed across immune phenotypes.[19] As expected, overall immune-related gene sets were highly up-regulated in inflamed phenotypes, while oncogenic pathways were enriched in desert phenotypes. Importantly, in addition to apoptosis and epigenetic regulatory metagenes, the nuclear factor (NF)-κB pathway was among the most attenuated gene sets in altered versus desert phenotypes (Figure 2C). This observation was further confirmed by gene set enrichment analysis using Reactome dataset, which showed that non-canonical NF-κB ($Q = 0.02$), cytokines ($Q = 0.1$), and T regulatory ($Q = 0.2$) pathway signaling had a significant negative enrichment score in inflamed versus altered phenotypes (Figure 2D, Supplementary Figure S7, available at https://doi.org/10.1016/j.annonc.2023.04.005).

Immune phenotypes and genomic changes

Next, we examined whether these immune phenotypes had different mutational profiles, in order to determine whether specific gene mutations have an effect on tumor immunogenicity. The DNA was analyzed using next-generation sequencing with the TSO500 panel, which detects 523 cancer-associated gene alterations. The variant classification of samples ($n = 215$; TNM-I cohort) is shown in Supplementary Figure S8, available at https://doi.org/10.1016/j.annonc.2023.04.005. In brief, the median (range) number of non-synonymous variants per sample was 12 (1-89), with the predominant mutation type of missense single nucleotide variants and the most common nucleotide change of C > A.

Since lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) have different genomic profiles and mutation patterns, we analyzed these tumors separately. The distributions of the most common mutations in each histological subgroup are shown in Figure 3A and B. In LUAD ($n = 137$), *TP53* (42%), *KRAS* (31%), and *STK11* (21%) were among the most commonly mutated genes, whereas in LUSC ($n = 68$), *TP53* (82%), *LRP1B* (40%), *ROS1* (18%), and *PIK3CA* (15%) were among the most commonly mutated genes. Mutation frequencies of the top 20 genes in the TNM-I cohort were similar to those seen in The Cancer Genome Atlas (TCGA) and Genomics Evidence Neoplasia Information Exchange (GENIE) (BPC NSCLC v2.0) lung (LUAD and LUSC) datasets (Supplementary Figures S9 and S10, available at https://doi.org/10.1016/j.annonc.2023.04.005).[20,21]

After stratification of the analysis, based on immune phenotypes across histology, compared to inflamed LUAD ($n = 72$), non-inflamed (altered and desert, $n = 65$) cases were significantly enriched for mutations in *KEAP1* [odds ratio (OR) 0.27, $Q = 0.02$] and *STK11* (OR 0.39,

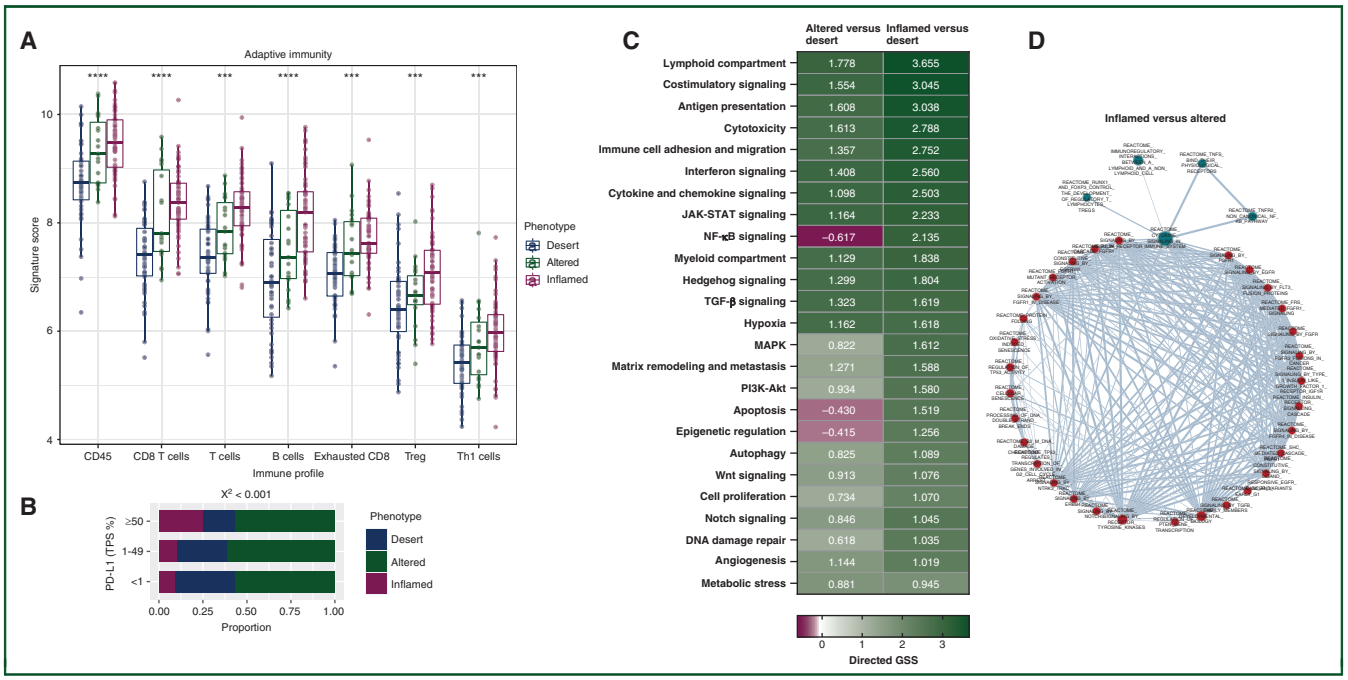

Figure 2. Molecular characteristics of immune phenotypes.
(A) Box plots of immune cell scores from the NanoString IO360 panel stratified by the machine learning (ML)-driven immune subclass. ***$P < 0.001$, ****$P < 0.0001$ by *U* test. (B) Distribution of immune phenotypes across programmed death-ligand 1 (PD-L1) immunohistochemistry (IHC) expression subgroups in the TNM-I cohort. (C) NanoString directed global significance scores (directed GSS) comparing the variation on gene expression across gene sets in altered versus desert and inflamed versus desert immune phenotypes. (D) Gene set enrichment analysis (GSEA) of the inflamed versus altered phenotypes using Reactome dataset and visualized with Enrichment Map (Cytoscape). The negative and positive enrichment scores are represented by blue and red colors, respectively.
NF-κB, nuclear factor-κB; TGF-β, transforming growth factor-β; TPS, tumor proportion score.

$Q = 0.04$). Likewise, *ATM* (OR 4.1, $Q = 0.03$) and *NF1* (OR 7.5, $Q = 0.03$) mutations were enriched among inflamed tumors (Figure 3C, Supplementary Figure S11, available at https://doi.org/10.1016/j.annonc.2023.04.005). Lollipop plots were constructed to demonstrate the distribution of all variants on specific functional domains of these genes. In the non-inflamed subgroup, the majority of somatic variations were distributed along the functional domains of *STK11* and *KEAP1*; but in the inflamed subgroup, *ATM* and *NF1* mutations were mainly located on the non-functional domain of *ATM* and *NF1* genes (Supplementary Figure S12, available at https://doi.org/10.1016/j.annonc.2023.04.005). Among the LUSC cases, no association with mutations was observed across the immune phenotypes.

Co-mutation analysis of the non-inflamed LUAD subset showed that both *TP53* and *KEAP1*, and *STK11* and *KEAP1* mutations often occurred together ($P < 0.001$; Supplementary Figure S13, available at https://doi.org/10.1016/j.annonc.2023.04.005). In the inflamed LUAD subset, *PTPRT* with *GRM3* and *ATRX*, *SETBP1* with *PIK3CG* and *SPTA1*, *NF1* and *SPTA1*, and *PTPRD* and *KRAS* mutations highly co-occurred, while none were mutually exclusive ($P < 0.001$; Supplementary Figure S13, available at https://doi.org/10.1016/j.annonc.2023.04.005). We then examined the impact of different immune phenotypes on the likelihood of somatic mutations in 10 cancer-related pathways.[22] The inflamed versus non-inflamed phenotypes were significantly associated with mutations in Wnt ($P = 0.004$) and cell cycle ($P < 0.001$) pathways (Supplementary Figure S14, available at https://doi.org/10.1016/j.annonc.2023.04.005).

We also compared tumor mutational burden (TMB) distribution across different immune phenotypes. The median (range) TMB was 8.6 mut/Mb (0.6-41 mut/Mb), and there was no difference in median TMB for the different immune subtypes. Further, there was no difference in median CD8+ density (cells/mm^2) in different tissue subregions based on low (<10) versus high (≥10) TMB (mut/Mb) levels (Figure 3D). Consistently, there was no significant difference in transversion events between inflamed and non-inflamed phenotypes (Supplementary Figure S15, available at https://doi.org/10.1016/j.annonc.2023.04.005).

Immune phenotype associations with disease stage and histology

The distribution of immune phenotypes according to baseline clinical attributes is shown in Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2023.04.005. No association was observed between the main histological subgroups or disease stages and immune phenotypes (Figure 4A). However, in LUAD, micropapillary and solid subtypes tended to have a higher proportion of the inflamed [40% (10/25), 27% (45/168), respectively] subgroup compared to lepidic (12%, 3/25) and papillary (21%, 14/66) subgroups (χ^2 $P = 0.24$, Figure 4A, Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2023.04.005).

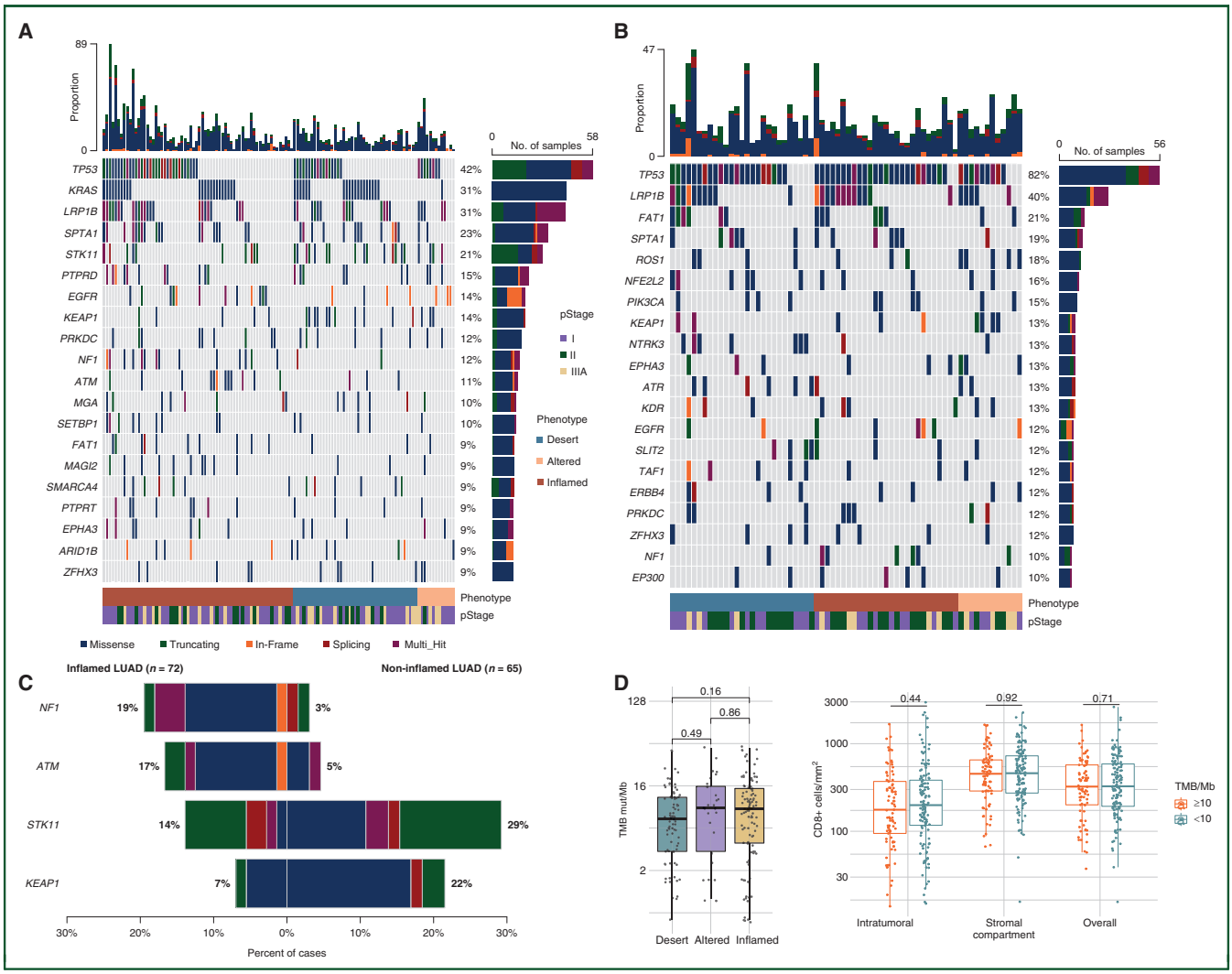

Figure 3. Mutation profiling of stage I-IIIA non-small-cell lung cancer (NSCLC) (TNM-I trial).
OncoPrint plot shows the top 20 mutated genes in lung adenocarcinoma (A; $n = 137$) and squamous cell carcinoma (B; $n = 68$) across immune phenotypes in the TNM-I genomic cohort. (C) Co-bar plot of statistically significant ($Q < 0.01$) differentially mutated genes in immune-inflamed versus non-inflamed. Color codes for variants are the same as for A. (D) Distribution of tumor mutational burden (TMB) (mut/Mb) across different immune phenotypes (left plot). Density (cells/mm^2) of tumoral, stromal, and total CD8+ cytotoxic T lymphocytes (CTLs) among NSCLC samples with low (<10 mut/Mb; $n = 130$) and high (≥10 mut/Mb; $n = 85$) TMB in the TNM-I cohort (right plot).

In contrast to immune phenotypes, 21 immune-related genes were significantly up-regulated when comparing LUAD versus LUSC as baseline (fold change >1.5; $Q < 0.05$; Figure 4B). This was further confirmed by pathway analysis in which immune-related gene sets were enriched in LUAD compared to LUSC, including antigen-presentation machinery, interferon signaling, and lymphoid compartments (Figure 4C). Similarly, in addition to a high intratumoral density of CD8+ (cells/mm^2), CD45, T, T helper, macrophages, dendritic cell, and mast cell signatures were significantly higher in LUAD compared to LUSC (Supplementary Figure S16, available at https://doi.org/10.1016/j.annonc.2023.04.005). In LUSC, only a few genes (RORC, DPP4, AREG) were up-regulated, despite the fact that many carcinogenic pathways were enriched, including Wnt and Hedgehog, compared to LUAD (Figure 4C).

Immune phenotypes and clinical outcome

The median (IQR) follow-up of the patients was 25 (17-36) and 83 (45-122) months for the TNM-I and UNN cohorts, respectively. Due to the short follow-up of the TNM-I cohort, the association of immune phenotypes with outcome data was determined only in the UNN cohort. In the UNN cohort, the median (95% confidence interval) time to recurrence (TTR) was 73 (52-not reached) months. We found that immune phenotypes (inflamed, altered, and desert) were significantly associated with 5-year disease-specific survival (DSS; 71% versus 57% versus 46%; $P = 0.002$) and TTR (64% versus 51% versus 41%; $P = 0.003$) (Figure 5A). Regarding the histology subgroups, immune inflamed was a positive prognostic factor for DSS ($P = 0.01$) and TTR ($P = 0.01$) in the LUSC subgroup. However, no significant association with survival was found for patients

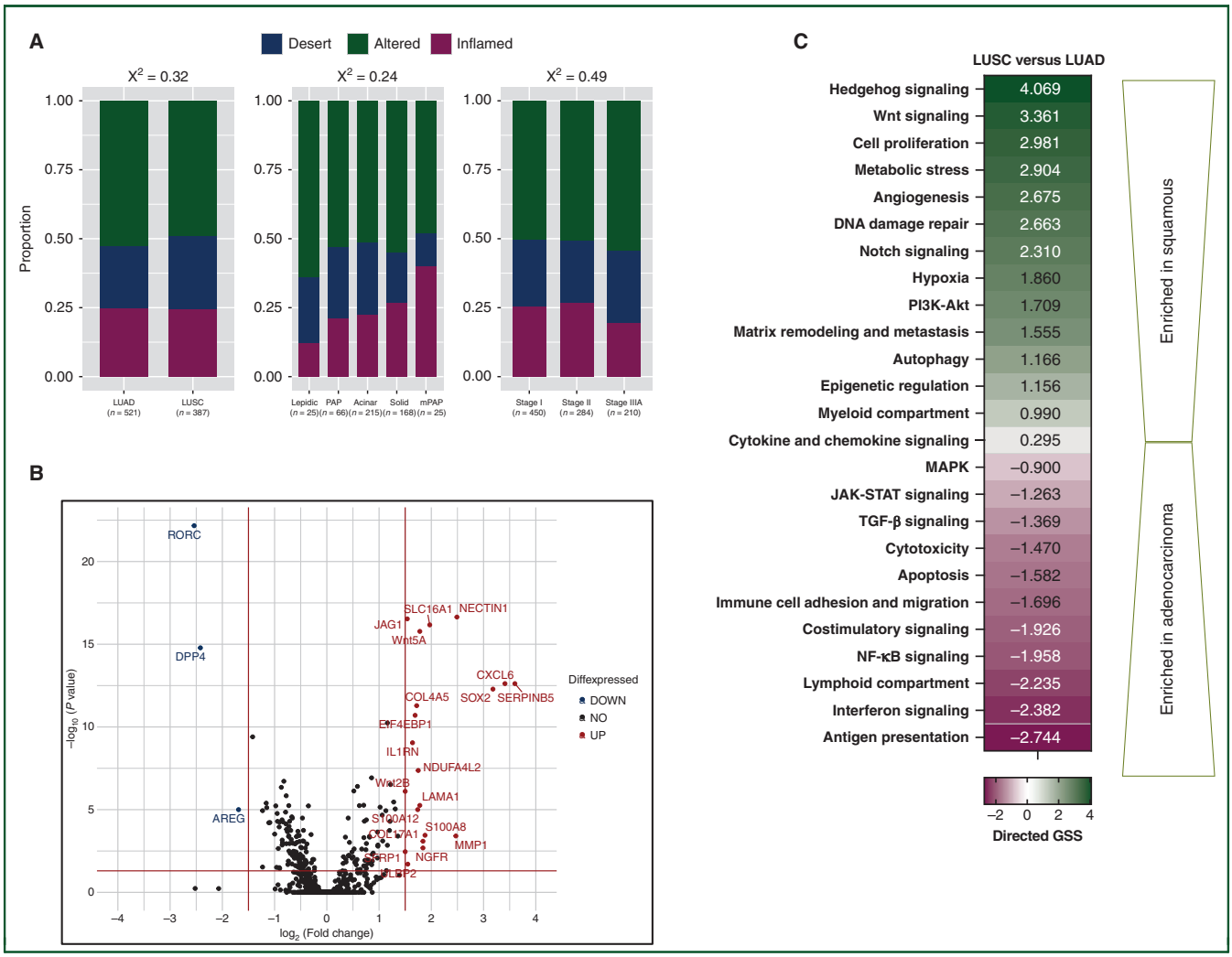

Figure 4. Non-small-cell lung cancer (NSCLC) histology and immune phenotypes.
(A) Distribution of immune phenotypes across main histological [lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)], LUAD subtypes, and pathological TNM (tumor—node—metastasis) stages of the entire NSCLC cohort ($n = 934$). (B) Differential expression of 770 immune-related genes between LUAD and baseline (LUSC) in the TNM-I cohort ($n = 132$). The *x*-axis shows the $\log_2$ fold change (FC) and the *y*-axis shows $-\log_{10}$ *P* values. Genes that have greater statistical significance ($-1.5 < FC > 1.5$; $Q < 0.05$) produced points with blue and red colors. (C) NanoString directed global significance scores (directed GSS) comparing the variation on biological gene set expressions between LUSC and LUAD.
mPAP, micropapillary; PAP, papillary; TGF-β, transforming growth factor-β.

with LUAD (Supplementary Figure S17, available at https://doi.org/10.1016/j.annonc.2023.04.005).

In multivariable analysis, after adjusting for pathological stage, histology, Eastern Cooperative Oncology Group performance score, and gender, the immune-inflamed phenotype was found to be an independent positive predictive factor for both DSS (inflamed versus desert, hazard ratio (HR) 0.61, $P = 0.016$; altered versus desert, HR 0.73, $P = 0.079$; Figure 5B) and TTR (Supplementary Figure S18, available at https://doi.org/10.1016/j.annonc.2023.04.005).

To explore the relative contribution of different prognostic biomarkers in classifying 5-year DSS and TTR, receiver operating characteristic (ROC) analysis was carried out including total, stromal, intratumoral CD8+ (cells/mm^2) density as continuous variable, immune phenotypes, and TNM granular staging. Followed by TNM staging [DSS area under the ROC curve (AUC) 0.66, $P < 0.001$; TTR AUC 0.63, $P < 0.001$], immune phenotypes had the highest AUC (DSS 0.58, $P = 0.001$; TTR 0.56, $P = 0.001$), while CD8 score in various tissue compartments (stromal, intratumoral, and overall) had lower AUC values for predicting risk of recurrence and disease survival (Figure 5C, Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2023.04.005).

DISCUSSION

This is the largest comprehensive study applying an automated procedure (ML) to analyze immune phenotypes in early-stage NSCLC tumor tissue. Our findings indicate that the inflamed phenotype is associated with a significantly lower risk of tumor recurrence and disease mortality in patients with NSCLC, independent of other clinical or molecular biomarkers. However, in subgroup analysis, this association was limited to the squamous cell carcinoma histology. We also found that non-inflamed immune

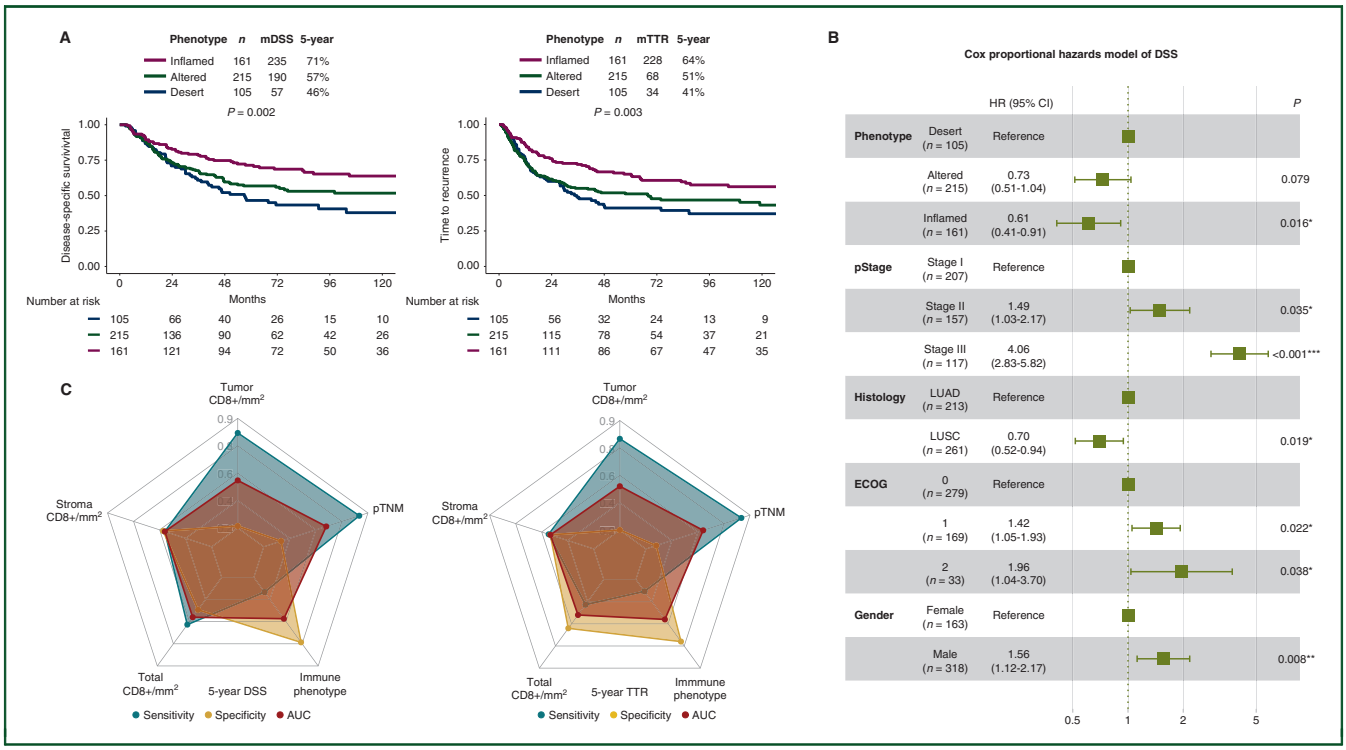

Figure 5. Immune phenotypes and clinical outcome.
(A) Kaplan–Meier curves of disease-specific survival (DSS) and time to recurrence (TTR) based on immune phenotypes in the UNN cohort ($n = 481$). (B) Multivariable analysis: forest plot of hazard ratio (HR) and 95% confidence interval (CI) for DSS according to known covariates in the UNN cohort. (C) Power of prognostic biomarkers to predict 5-year DSS (left) and TTR (right). Numbers in the plot represent the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.
ECOG, Eastern Cooperative Oncology Group; mDSS, median DSS in months; mTTR, median TTR in months; pTNM, pathological TNM (tumor–node–metastasis).

phenotypes are characterized by positive enrichment of a distinct proinflammatory pathway (NF-κB) and tend to have either or both *STK11* and *KEAP1* mutations.

ML-based immune phenotyping on whole-slide images

We used ML-based image analysis for immunological phenotyping of NSCLC tissues. Similar algorithms have been used in recent years to analyze histopathological images.[23] However, many have used unsupervised deep-learning approaches, whereas our method was pathologist-supervised, with specific steps for tissue segmentation based on pixel classification, and cell recognition based on hematoxylin and DAB optical density in nuclei. We created and tested this classification system on two independent lung cancer datasets. We used the open-source software QuPath, which has been adapted and validated for digital pathology image analysis.[15] In addition to association of ML-powered immune phenotypes with gene expression immune classifications (Figure 2A), our artificial neural network tissue classification algorithm obtained excellent accuracy compared to rigorous manual stromal and tumoral annotation by pathologists (ICC 0.99 and 0.98, respectively). Further, it is still not known whether analyses of region of interest (including tissue microarrays, hotspots, and selected regions) or whole-slide images are better for evaluating tissue biomarkers,[24] but we consider one of the strengths of our study to be the use of whole-slide images for immunological phenotyping.

Regarding the immune phenotyping cut-offs, Park et al. used the lower intratumoral (106 cells/mm^2) and upper stromal (357 cells/mm^2) 25th percentile (1 mm^2 patch size) for hematoxylin and eosin-based TIL analysis in 'advanced' NSCLC.[8] To our knowledge, this is the first study implementing ML models to immune phenotype of whole-slide images of 'early-stage' NSCLC patients based on CD8+ CTL density. On the basis of lower and upper 25th percentiles in our entire material, we used intratumoral 88 CD8+/mm^2 and stromal 814 CD8+/mm^2. Even though this classification was supported by gene expression immune signatures, the proposed cut-offs should be regarded as hypothesis-generating which should be confirmed in the TNM-I trial full dataset ($n = 1000$).

Genomic and immune phenotype

The *STK11* gene encodes liver kinase B1, which is a serine/threonine kinase that regulates and activates multiple downstream kinases that control metabolism and growth.[25] In our cohort, *STK11* deficiency affected ~21% of LUAD patients, similar to previous reports (Supplementary Figure S9, available at https://doi.org/10.1016/j.annonc.2023.04.005). *KEAP1* encodes an adaptor subunit of cullin 3-based E3 ubiquitin ligase, which serves a critical role in the regulation of the Nuclear factor erythroid 2-related

factor 2 transcription factor, which is a master regulator of detoxification and antioxidant genes. Mutations in *KEAP1* lead to activation of NRF2, which leads to resistance to multiple therapeutic interventions including immunotherapy.[26,27] In our cohort, *KEAP1* mutations were found in 14% of LUAD patients. We found that both *STK11* and *KEAP1* mutations were associated with a non-inflamed phenotype, and this was especially strong for tumors with mutations in both genes, which were much more common than predicted by chance. However, further investigation is warranted into the cross-communication between local immune reactions and *STK11*/*KEAP1* alterations. We also identified that there is no link between immunological phenotype and TMB in early-stage (stage I-IIIA) tumors, which is consistent with our previous findings in a large cohort of advanced NSCLC patients.[15]

NF-κB pathway

NF-κB has a key role in the development and function of regulatory T cells (T_{reg}), which have an immunosuppressive function that sustains self-tolerance and immune homeostasis.[28] T_{reg} are also known to suppress antitumor immune responses by influencing the function of TILs and macrophages. NF-κB signaling is also involved in the regulation of immune checkpoint expression in tumor cells, by inducing PD-L1 expression and promoting T-cell suppression, and thus tumor progression.[29] In this study, we found that NF-κB and T_{reg} pathways are highly enriched in non-inflamed compared to inflamed tumors. The desert phenotype was characterized by none or very-low-density CD8+ CTLs, and was either associated with alteration on oncogenic pathways gene sets (e.g. Wnt signaling; Supplementary Figures S14 and S19, available at https://doi.org/10.1016/j.annonc.2023.04.005) or with higher macrophage/TIL and neutrophil/TIL signature scores (Supplementary Figure S5, available at https://doi.org/10.1016/j.annonc.2023.04.005). Previous reports show that activation of Wnt signaling stimulates CTL exclusion in bladder cancer, melanoma, and NSCLC.[30-32] Overall, NF-κB is a promising therapeutic target promoting lymphocyte infiltration into tumors, turning the tumor microenvironment into a more inflamed state, and making it more susceptible to PD-L1 checkpoint blockade.[33]

Immune phenotypes and clinical outcome

Patients with advanced-stage non-inflamed NSCLC, including desert and excluded phenotypes, are less likely to respond to immune checkpoint inhibitors.[8] Similar findings have been reported in patients with advanced renal cell carcinomas and breast cancer with inflamed immune classifications, who have an improved response rate and progression-free survival to immunotherapy.[34,35] We previously reported, in a semi-quantitative manual approach to tissue microarrays, that stromal CD8+ CTLs had a better prognostic impact than intratumoral CD8+ CTLs in resected and early-stage NSCLC.[14] The significance of the inflamed phenotype for prognosis in resected NSCLC is a novel finding in the current study which uses fully quantitative and operator-independent methods.

However, we observed no significant prognostic value of immune phenotypes in LUAD histology, regardless of the clinical endpoint and found that more LUAD patients with solid or micropapillary-predominant subtypes are of the inflamed class than those with other LUAD subtypes. This unequal distribution of immune phenotypes across LUAD subtypes may explain why the prognostic significance of immune phenotypes varies across main histological subgroups. Nevertheless, we plan to extend our study to the full prospective TNM-I trial, which includes a larger number of LUAD NSCLC patients. Lastly, the results from our comparison of the different models also suggest that immune phenotypes are associated with higher performance power (AUC) and greater specificity compared to CD8+ subregion scores (cells per mm^2) for predicting the risk of recurrence and disease mortality.

Study limitations

We developed an automated model for assessing immune phenotypes and carried out comprehensive investigations of clinical and genomic features that could confound the immune cell score model. There were several limitations in our analysis: there was a relatively short follow-up of the TNM-I cohort (median 25 months), preventing analyses of the relationship between immunological phenotypes and clinical outcomes, and the analysis did not include copy number variants, which could be of significant importance in immune-related gene expression profiles of cancer patients receiving immunotherapy.[36]

The use of immune checkpoint inhibitors has recently been introduced in clinical practice both following NSCLC resection and in the neoadjuvant setting, whereas sequencing and optimal combinations with, for example, chemotherapy are being explored. Further, there are several ongoing trials on the impact of combining immune checkpoint inhibitors with adjuvant chemotherapy after NSCLC resection.[37] It will be intriguing to explore whether immunological phenotypes have the potential to provide predictive information in this area.

Conclusion

This interim analysis of data from a TNM-I clinical trial led to the development and validation of an ML model that accurately classifies spatial immune phenotypes in full-faced images on slide of resected NSCLC, using IHC for CD8+ and pCK. These immune phenotypes were associated with prognosis in multivariable analysis. Therefore, the ML-driven immune phenotypes could be a good candidate marker for establishing a TNM immune cell score in NSCLC.[10] In addition, mutation analysis of immune phenotypes revealed an association between *STK11* and *KEAP1* individual and co-mutations and poor local immune infiltration in LUAD, which could be a potential resistance mechanism to immune checkpoint blockade due to lack of tumor immunogenicity.[38]

ACKNOWLEDGEMENTS

We thank Dr Karen Ege Olsen at Department of Pathology, Odense University Hospital for providing the samples; Bonnie Svendsen at Department of Pathology, Rigshospitalet for helping with research data management and laboratory specimen management; and Marte Berglund at Molecular Pathology Lab, University Hospital of North Norway for NanoString analysis assistance.

FUNDING

MR was supported by The Norwegian Cancer Society [grant number 197879]. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DISCLOSURE

ÅH reports grants from Roche, AstraZeneca, BMS, Ultimovacs, Incyte, EliLilly, Novartis, and advisory role from AstraZeneca, Roche, Janssen, Bayer, Novartis, Takeda, Pfizer, EliLilly, MSD, BMS, Abbvie, Merck, and Sanofi outside the submitted work. MP reports advisory role from AstraZeneca, BMS, MSD, Pfizer, and Roche outside the submitted work. TB reports honoraria for lectures and consultancy from AstraZeneca, MSD, Novartis, Roche Diagnostics, and Illumina outside the submitted work. RHP reports advisory role from AstraZeneca, Roche, MSD and BMS and speaker fee from Medtronic, AMBU, Medela and AstraZeneca outside the submitted work. All other authors have declared no conflicts of interest.

REFERENCES

1. Lu S, Stein JE, Rimm DL, et al. Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: a systematic review and meta-analysis. *JAMA Oncol*. 2019;5(8):1195-1204.
2. McKean WB, Moser JC, Rimm D, Hu-Lieskovan S. Biomarkers in precision cancer immunotherapy: promise and challenges. *Am Soc Clin Oncol Educ B*. 2020;(40):e275-e291.
3. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133-150.
4. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. *Br J Cancer*. 2021;124(2):359-367.
5. Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. *Science*. 2015;348(6230):74-80.
6. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017;541(7637):321-330.
7. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov*. 2019;18(3):197-218.
8. Park S, Ock C-Y, Kim H, et al. Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non-small-cell lung cancer. *J Clin Oncol*. 2022;40(17):1916-1928.
9. Goldstraw P, Chansky K, Crowley J, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11(1):39-51.
10. Donnem T, Kilvaer TK, Andersen S, et al. Strategies for clinical implementation of TNM-Immunoscore in resected nonsmall-cell lung cancer. *Ann Oncol*. 2016;27(2):225-232.
11. Argilés G, Tabernero J, Labianca R, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(10):1291-1305.
12. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2019;30(8):1194-1220.
13. Rakaee M, Kilvaer TK, Jamaly S, et al. Tertiary lymphoid structure score: a promising approach to refine the TNM staging in resected non-small cell lung cancer. *Br J Cancer*. 2021;124(10):1680-1689.
14. Donnem T, Hald SM, Paulsen E-E, et al. Stromal CD8+ T-cell density—a promising supplement to TNM staging in non-small cell lung cancer. *Clin Cancer Res*. 2015;21(11):2635-2643.
15. 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*. 2023;9(1):51-60.
16. Rakaee M, Adib E, Ricciuti B, et al. Artificial intelligence in digital pathology approach identifies the predictive impact of tertiary lymphoid structures with immune-checkpoints therapy in NSCLC. *J Clin Oncol*. 2022;40(suppl 16):9065.
17. Dano H, Altinay S, Arnould L, et al. Interobserver variability in upfront dichotomous histopathological assessment of ductal carcinoma in situ of the breast: the DCISion study. *Mod Pathol*. 2020;33(3):354-366.
18. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):1-7.
19. White MG, Szczepaniak Sloane R, Witt RG, et al. Short-term treatment with multi-drug regimens combining BRAF/MEK-targeted therapy and immunotherapy results in durable responses in Braf-mutated melanoma. *Oncoimmunology*. 2021;10(1):1992880.
20. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;173(2):371-385.e18.
21. AACR Project GENIE Consortium. AACR project GENIE: powering precision medicine through an international consortium. *Cancer Discov*. 2017;7(8):818-831.
22. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell*. 2018;173(2):321-337.e10.
23. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*. 2019;16(11):703-715.
24. Patil SM, Tong L, Wang MD. Generating region of interests for invasive breast cancer in histopathological whole-slide-image. Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020 Institute of Electrical and Electronics Engineers Inc. Washington DC: IEEE Computer Society. 2020:723-728.
25. Hollstein PE, Eichner LJ, Brun SN, et al. The AMPK-related kinases SIK1 and SIK3 mediate key tumor-suppressive effects of LKB1 in NSCLC. *Cancer Discov*. 2019;9(11):1606-1627.
26. Ricciuti B, Arbour KC, Lin JJ, et al. Diminished efficacy of programmed death-(ligand)1 inhibition in STK11- and KEAP1-mutant lung adenocarcinoma is affected by KRAS mutation status. *J Thorac Oncol*. 2022;17(3):399-410.
27. Jaramillo MC, Zhang DD. The emerging role of the Nrf2-Keap1 signaling pathway in cancer. *Genes Dev*. 2013;27(20):2179-2191.
28. Perkins ND. The diverse and complex roles of NF-κB subunits in cancer. *Nat Rev Cancer*. 2012;12(2):121-132.
29. Peng J, Hamanishi J, Matsumura N, et al. Chemotherapy induces programmed cell death-ligand 1 overexpression via the nuclear factor-κB to foster an immunosuppressive tumor microenvironment in ovarian cancer. *Cancer Res*. 2015;75(23):5034-5045.
30. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. *Nature*. 2015;523(7559):231-235.
31. Sweis RF, Spranger S, Bao R, et al. Molecular drivers of the non- T-cell-inflamed tumor microenvironment in urothelial bladder cancer. *Cancer Immunol Res*. 2016;4(7):563-568.

32. Takeuchi Y, Tanegashima T, Sato E, et al. Highly immunogenic cancer cells require activation of the WNT pathway for immunological escape. *Sci Immunol*. 2021;6(65):eabc6424.
33. Yu H, Lin L, Zhang Z, Zhang H, Hu H. Targeting NF-κB pathway for the therapy of diseases: mechanism and clinical study. *Signal Transduct Target Ther*. 2020;5(1):209.
34. Braun DA, Hou Y, Bakouny Z, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med*. 2020;26(6):909-918.
35. Hammerl D, Martens JWM, Timmermans M, et al. Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nat Commun*. 2021;12(1):5668.
36. Lu Z, Chen H, Li S, et al. Tumor copy-number alterations predict response to immune-checkpoint-blockade in gastrointestinal cancer. *J Immunother Cancer*. 2020;8(2):e000374.
37. Provencio M, Calvo V, Romero A, Spicer JD, Cruz-Bermúdez A. Treatment sequencing in resectable lung cancer: the good and the bad of adjuvant versus neoadjuvant therapy. *Am Soc Clin Oncol Educ B*. 2022;(42):711-728.
38. Keenan TE, Burke KP, Van Allen EM. Genomic correlates of response to immune checkpoint blockade. *Nat Med*. 2019;25(3):389-402.