


Inflamed immune phenotype predicts favorable clinical outcomes of immune checkpoint inhibitor therapy across multiple cancer types

Jeanne Shen ,^{1,2} Yoon-La Choi,^{3,4} Taebum Lee,⁵ Hyojin Kim,⁶ Young Kwang Chae,⁷ Ben W Dulken,¹ Stephanie Bogdan,² Maggie Huang,⁸ George A Fisher,⁹ Sehhoon Park ,¹⁰ Se-Hoon Lee,¹⁰ Jun-Eul Hwang,¹¹ Jin-Haeng Chung,⁶ Leeseul Kim,¹² Heon Song,¹³ Sergio Pereira,¹³ Seunghwan Shin,¹³ Yoojoo Lim,¹³ Chang Ho Ahn,¹³ Seulki Kim,¹³ Chiyoon Oum,¹³ Sukjun Kim,¹³ Gahee Park,¹³ Sanghoon Song,¹³ Wonkyung Jung,¹³ Seokhwi Kim,¹⁴ Yung-Jue Bang,¹⁵ Tony S K Mok,¹⁶ Siraj M. Ali,¹³ Chan-Young Ock¹³

To cite: Shen J, Choi Y-L, Lee T, *et al.* Inflamed immune phenotype predicts favorable clinical outcomes of immune checkpoint inhibitor therapy across multiple cancer types. *Journal for ImmunoTherapy of Cancer* 2024;**12**:e008339. doi:10.1136/jitc-2023-008339

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/jitc-2023-008339>).

JS, Y-LC, TL, HK and YKC contributed equally.

Preliminary results were presented in part at the ASCO Annual Meeting 2022 (June 3 – June 7, 2022), Chicago, Illinois, USA.

Accepted 27 January 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Jeanne Shen;
jeannes@stanford.edu

Dr Chan-Young Ock;
ock.chanyoung@lunit.io

ABSTRACT

Background The inflamed immune phenotype (IIP), defined by enrichment of tumor-infiltrating lymphocytes (TILs) within intratumoral areas, is a promising tumor-agnostic biomarker of response to immune checkpoint inhibitor (ICI) therapy. However, it is challenging to define the IIP in an objective and reproducible manner during manual histopathologic examination. Here, we investigate artificial intelligence (AI)-based immune phenotypes capable of predicting ICI clinical outcomes in multiple solid tumor types.

Methods Lunit SCOPE IO is a deep learning model which determines the immune phenotype of the tumor microenvironment based on TIL analysis. We evaluated the correlation between the IIP and ICI treatment outcomes in terms of objective response rates (ORR), progression-free survival (PFS), and overall survival (OS) in a cohort of 1,806 ICI-treated patients representing over 27 solid tumor types retrospectively collected from multiple institutions.

Results We observed an overall IIP prevalence of 35.2% and significantly more favorable ORRs (26.3% vs 15.8%), PFS (median 5.3 vs 3.1 months, HR 0.68, 95% CI 0.61 to 0.76), and OS (median 25.3 vs 13.6 months, HR 0.66, 95% CI 0.57 to 0.75) after ICI therapy in IIP compared with non-IIP patients, respectively ($p < 0.001$ for all comparisons). On subgroup analysis, the IIP was generally prognostic of favorable PFS across major patient subgroups, with the exception of the microsatellite unstable/mismatch repair deficient subgroup.

Conclusion The AI-based IIP may represent a practical, affordable, clinically actionable, and tumor-agnostic biomarker prognostic of ICI therapy response across diverse tumor types.

INTRODUCTION

Immune checkpoint inhibitors (ICI) have become a major part of the standard of care for various tumor types.¹ Several predictive and prognostic biomarkers, including

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Recent studies have demonstrated potential association between the distribution of tumor-infiltrating lymphocytes (TIL) within the tumor microenvironment and response to immune checkpoint inhibitor (ICI) therapies. However, manual evaluation of TILs can be time-consuming, labor intensive, and subject to interobserver variability.

WHAT THIS STUDY ADDS

⇒ This study demonstrates the ability of an artificial intelligence (AI) model that runs on routine H&E-stained pathology whole-slide images of pre-treatment tumor samples to predict ICI treatment outcomes in a real-world multicenter cohort of 1,806 ICI-treated patients representing over 27 different solid tumor types.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ An AI-based assessment of immune phenotypes is associated with better clinical outcomes after ICI therapy across diverse solid tumor types, suggesting its potential as a prognostic biomarker for ICI treatment planning.

programmed cell death ligand 1 (PD-L1) expression, high microsatellite instability/mismatch repair deficiency (hereafter referred to as MSI), and tumor mutational burden (TMB) have been approved for use to guide ICI treatment decisions, but concern exists regarding their technical and clinical limitations.² PD-L1 expression by immunohistochemistry (IHC) has been the most extensively investigated. However, pivotal studies supporting the Food and Drug Administration's (FDA) ICI drug

approvals have shown that PD-L1 IHC predicts ICI treatment response in only 28.9% of cases,³ limited to specific tumor types. Furthermore, PD-L1 IHC performance and interpretation strongly rely on factors such as specimen fixation method, the choice of antibody clone and staining platform, pathologist experience level, and the specific IHC scoring method used. Moreover, the thresholds for PD-L1 positivity vary by cancer type and treatment indication.⁴

MSI and TMB have recently been approved as tissue-agnostic biomarkers of ICI response, as MSI and TMB-high are associated with increased numbers of tumor neoantigens.^{5,6} However, the overall prevalence of MSI is less than 2% in pan-cancer studies,^{7,8} and estimation of TMB by next-generation-sequencing (NGS) has also been subject to technical and practical limitations. NGS testing has more stringent specimen requirements, a longer turnaround time, and higher costs. Although a cut-off of 10 or more mutations per megabase (Mb) is often used for defining TMB-high, as per a recent FDA approval, this threshold is limited to the few tumor types investigated in one study⁶ and likely needs to be adjusted for other tumor types.^{9–12}

Given the limitations of current biomarkers, there is a significant need for additional novel biomarkers that are more time and cost efficient, universally applicable, and able to capture a significant proportion of potential ICI responders. The inflamed immune phenotype (IIP) is a potential new biomarker which is directly associated with, or reflective of, the mechanism of action of ICIs, which impact the activity of immune cells within the tumor microenvironment (TME).

Clinical outcomes after ICI treatment have been associated with the spatial localization of tumor-infiltrating lymphocytes (TILs) within the TME across several tumor types.^{13,14} TIL evaluation on H&E-stained tumor pathology slides collected during routine clinical care could potentially serve as a novel tumor-agnostic biomarker for ICI response. Advances in artificial intelligence (AI), and specifically, deep learning, offer the possibility of more objective and reproducible automated computational TIL assessment.¹⁵ Building on our prior work in automated histopathologic cancer region segmentation and object detection,^{16,17} we recently demonstrated the ability of deep learning to perform TME immune phenotyping on H&E-stained whole-slide images (WSI) of non-small cell lung cancer (NSCLC) and nasopharyngeal carcinoma, showing that the IIP is correlated with survival and response to ICI treatment.^{18–20} In the current study, we extend this work in a pan-cancer analysis to assess whether the immune phenotype (IP), as determined by an automated deep learning model which performs TIL analysis on routine H&E-stained WSI, might serve as a novel, clinically-actionable “tumor-agnostic” biomarker for predicting ICI treatment outcomes in a large, real-world sample of patients representing over 27 different solid tumor types.

METHODS

Classification of the artificial intelligence-based immune phenotype

The Lunit SCOPE IO model (Lunit, Seoul, Republic of Korea)¹⁸ is a deep learning model that classifies the IP of the TME based on TIL distribution and density within H&E WSI. Model inference consists of two main stages: (1) tissue segmentation and cell detection, followed by (2) IP classification based on the outputs from the preceding stage.

Deep-learning-based tissue segmentation and cell detection stage

In this first stage, a convolutional neural network (CNN), hereafter referred to as the tissue segmentation model, performs semantic segmentation of cancer area (CA) and cancer stroma (CS) within a WSI (where CA refers to cancer epithelium or, in the case of non-epithelial tumors, the non-stromal tumor cells). In parallel, another CNN detects TILs using a cell detection model that identifies both tumor cells and lymphocytes. The data sets for training and tuning (optimizing) these CNNs were drawn from a pool of 17,296 H&E WSIs of over 24 different solid tumor types, including NSCLC and rarer cancer types (online supplemental tables 1a,b), collected from over nine different sources/institutions, scanned at 20× to 40× magnification (0.25 μm to 0.5 μm per pixel). As WSIs cannot be directly input into models due to their large size, representative 1,024×1,024 pixel patches were computationally extracted from pathologist-delineated tumor regions from each WSI, summing to a total effective area of 3.44×10¹⁰ μm² for training and 8.75×10⁹ μm² for tuning. The training and tuning sets for the tissue segmentation model consisted of 55,325 patches extracted from 13,962 WSIs and 13,962 patches extracted from 849 WSIs, respectively, manually annotated for CA and CS regions by board-certified pathologists. The training and tuning subsets for the cell detection model consisted of 5,698 patches extracted from 2,485 WSIs and 1,925 patches extracted from 849 WSIs, respectively, manually annotated by board-certified pathologists for tumor cells and lymphocytes by placing a point annotation within the nucleus of each cell. All annotations were independently verified by a second pathologist before being used as the ground truth for model development. A total of 104 board-certified pathologists participated in annotation.

The architecture of the tissue segmentation model was based on DeepLabV3+,²¹ with a Resnet-34 backbone as a feature extractor.²² This model takes as input patches of size 1,024×1,024 pixels, performing semantic segmentation on each patch (predicting the likelihood of each pixel belonging to the CA, CS, or background non-cancer tissue classes) and outputting a tissue class probability map of size 1,024×1,024 pixels. The model was trained using a Dice loss function²³ and optimized using the Adam optimizer²⁴ with a learning rate of 0.0001, achieving a performance of 0.82 and 0.67 on the Intersection-over-Union metric for CA and CS, respectively (please see online supplemental methods for additional details).

The cell detection model was also based on the DeepLabV3+ and Resnet-34 architectures.^{21–22} Cell nuclei were annotated as points, but this model casts the cell detection problem as a dense pixel prediction task. We therefore generated a circle of radius 0.95 μm centered around each point annotation during the training stage and trained the model using a Dice loss function²³ with Adam optimization²⁴ with a learning rate of 0.002. The inputs to the model were also patches of size 1,024 \times 1,024 pixels, with the outputs being probability maps of size 1,024 \times 1,024 pixels where each pixel represents the likelihood of a TIL existing in that location. A post-processing stage was applied to extract the location of the cells (in online supplemental methods). The F1-score for this model on lymphocyte detection was 0.69.

The performance of the tissue region segmentation and cell detection models was validated on a separate internal validation set consisting of 356 WSIs from over 17 different tumor types collected from over 8 different sources/institutions (online supplemental table 2), scanned at 20 \times to 40 \times magnification (0.25 μm to 0.5 μm per pixel), with resultant area under the receiver operating characteristic (AUROC) values for segmentation of CA and CS and TIL detection of above 0.95 (figure 1B). The outputs from the preceding tissue segmentation and cell detection models were used in the subsequent IP classification stage.

Immune phenotype classification stage

In this stage, we sought to classify the tumor present in each patient's WSI into one of three general IP.^{25–26} In the IIP, there is a high density of TILs present within the CA. In the immune excluded phenotype (IEP), TILs are abundant within the CS but excluded from the CA. In the immune desert phenotype (IDP), TILs are scarce within both the CA and CS.

To perform TIL analysis within WSI of various sizes, each WSI was divided into 1 mm^2 grids, with the IP of each grid classified as IIP, IEP, or IDP based on empirically determined TIL density criteria (see next paragraph below). The overall WSI-level Inflamed Score (IS), Immune-Excluded Score (IES), and Immune-Desert Score (IDS) were calculated by dividing the number of grids having that respective phenotype over the total number of grids analyzed within the WSI. If the WSI-level IS exceeded a prespecified threshold, that WSI was classified as having a WSI-level IIP. The overall workflow for IP classification is illustrated in figure 1A.

Determination of the TIL density cutoffs for grid-level IP classification and the cancer-type agnostic IS threshold for classification of an WSI as IIP was based on prior evidence demonstrating that a T-cell-inflamed gene expression profile (as characterized by the interferon-gamma responsive gene (*IFNG*) signature) is predictive of ICI clinical response.²⁷ We hypothesized that ICI responders would exhibit a stronger *IFNG* signature and comprise approximately 25% of patients with pan-carcinoma. Therefore, the TIL density cut-off and optimal IS threshold were

determined as those scores that predicted the upper 25% of *IFNG* signature levels in The Cancer Genome Atlas (TCGA) pan-carcinoma data set (n=7,454, online supplemental table 3)²⁸ with the highest AUROC and greatest sum of sensitivity and specificity.

We found that an intratumoral TIL density cut-off of 200/ mm^2 yielded the highest AUROC (0.7772) for predicting a 75th percentile or higher *IFNG* signature level. As the *IFNG* signature levels in the TCGA data set were derived from bulk sequencing data without distinction between cancer parenchyma and stroma, we set a consistent cut-off of 200/ mm^2 for stromal (CS) TILs to ensure a comparable distribution to that observed in cancer parenchyma (CA). The IP of each grid was therefore classified using the following criteria: grid-level IIP, if the TIL density within the total CA in the grid is $\geq 200/\text{mm}^2$; grid-level IEP, if the TIL density within the total CA is $< 200/\text{mm}^2$ and that within the total CS is $\geq 200/\text{mm}^2$; and grid-level IDP, if the TIL density is $< 200/\text{mm}^2$ in both the total CA and CS within the grid. At the 200/ mm^2 TIL cut-off, an optimal IS of 19.479% resulted in the maximum sum of sensitivity (65.0%) and specificity (78.2%), irrespective of TMB status (AUROC 0.76 for the TMB-high population and 0.77 for the TMB-low population). Therefore, we set 20.0% as the IS threshold for WSI-level IIP classification, regardless of cancer type (hereafter referred to as the *universal threshold*) (figure 1C, online supplemental table 3).

ICI-treated patient data set

The final optimized Lunit SCOPE IO model was applied to an independent real-world data set (N=1,806 patients) of H&E-stained WSIs scanned at 40 \times magnification (0.25 μm per pixel), derived from pre-ICI treatment formalin-fixed, paraffin-embedded (FFPE) surgical resection and biopsy specimens with accompanying clinical outcomes, including progression-free survival (PFS), overall survival (OS), and best overall response (BOR) after ICI monotherapy or ICI combination therapy, as assessed by Response Evaluation Criteria In Solid Tumors (RECIST) V.1.1.²⁹ The WSIs were collected from Stanford University Medical Center (Stanford, n=688), Samsung Medical Center (SMC, n=653), Seoul National University Bundang Hospital (SNUBH, n=269), Chonnam National University Hospital (CNUH, n=183), and Northwestern Memorial Hospital (Northwestern, n=13). ICI combination therapy regimens included at least one other anti-neoplastic drug, such as conventional chemotherapy. All work was conducted in accordance with the Declaration of Helsinki for biomedical research, after institutional review board approval at each participating institution.

PD-L1 immunohistochemistry and other ancillary biomarker testing

PD-L1 IHC was performed using the US FDA-approved Dako PD-L1 IHC 22C3 PharmDx kit (Agilent Technologies, Santa Clara, California, USA), with scoring of PD-L1 expression (%) determined using the Tumor Proportion

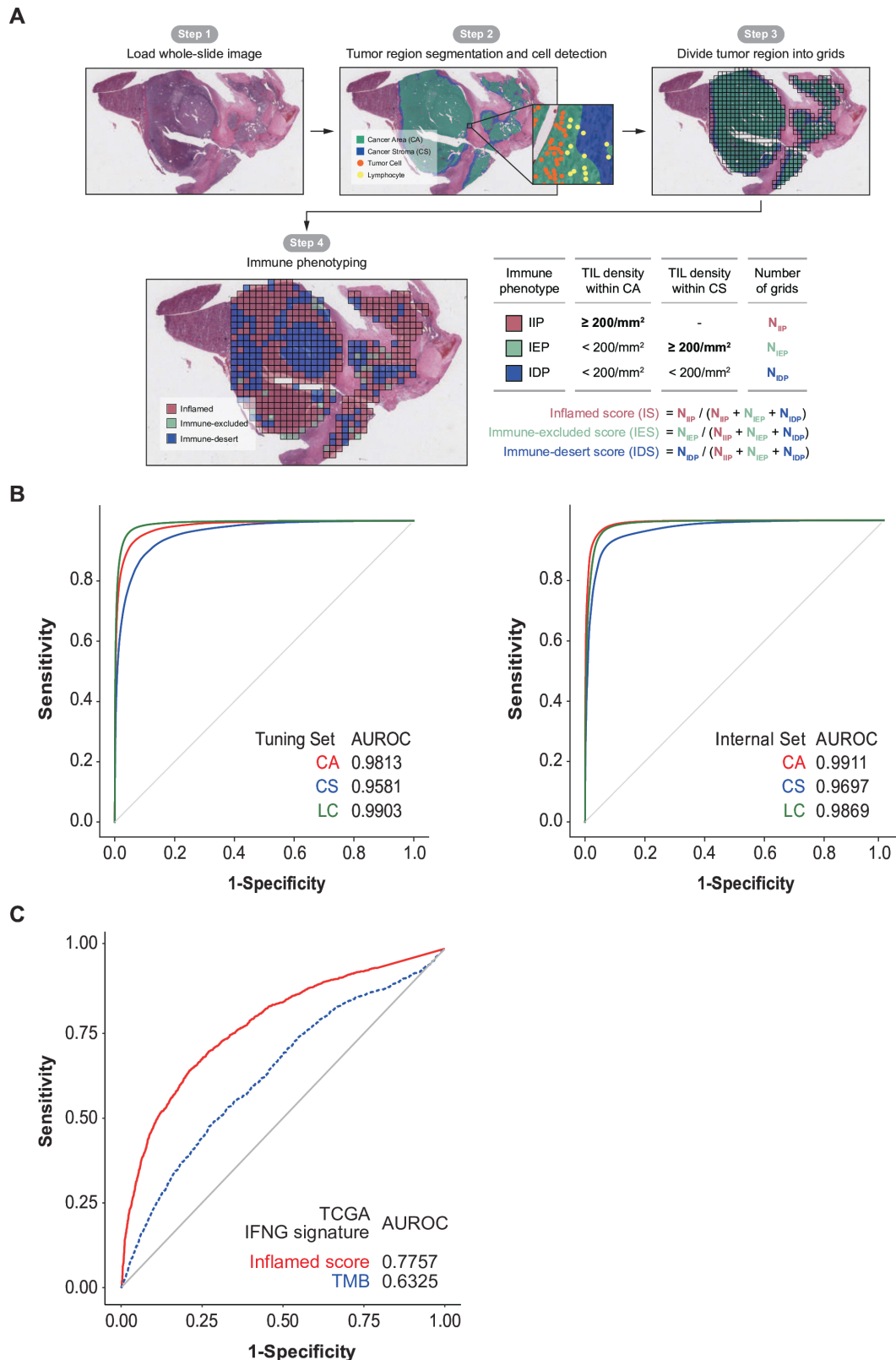


Figure 1 Pan-cancer performance validation of the Lunit SCOPE IO, AI-powered H&E-stained WSI analyzer, and classification of the immune phenotype. (A) Workflow for AI-powered TIL analysis and IP classification. (B) ROC curves for segmentation of CA and CS and TIL (LC) detection on the tuning (n=849) and internal validation sets (n=356). (C) ROC curves for determination of the optimal Inflamed Score threshold (*universal threshold*) for WSI-level IP classification (IIP vs non-IIP) in the TCGA data set (n=7,454). AI, artificial intelligence; AUROC, area under the ROC; CA, cancer area; CS, cancer stroma; IDP, immune-desert phenotype; IEP, immune-excluded phenotype; IFNG, interferon-gamma-responsive gene; IIP, inflamed immune phenotype; IP, immune phenotype; LC, lymphocyte; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas; TIL, tumor-infiltrating lymphocyte; TMB, tumor mutational burden; WSI, whole-slide image.

Score (TPS), representing the percentage of viable tumor cells showing partial or complete membranous staining for PD-L1 with 1+ to 3+ intensity.³⁰

Determination of microsatellite status at Stanford was done using DNA mismatch repair IHC and/or MSI PCR. For MSI PCR, standard multiplex PCR amplification of a panel of five microsatellites (BAT-25, BAT-26, MONO-27, NR-21, and NR-24) was performed with comparison of tumor and normal samples from the same patient by the Stanford Molecular Pathology laboratory (order code: TMSI (Tumor Microsatellite Instability), additional assay details available at: <https://stanfordlab.com/content/stanfordlab/en/test-details/t/TMSI.html>). If two of five microsatellite loci showed a difference in length between tumor and normal samples, the tumor was designated MSI high. If only one locus showed a difference in length, the tumor was considered MSI low. If no loci showed a difference in length, the tumor was designated microsatellite stable. IHC for DNA mismatch repair proteins was performed using standard protocols with monoclonal antisera reacting to MLH1 (clone G168-728, BD Biosciences), MSH2 (clone FE11, Calbiochem), MSH6 (clone 44, Cell Marque), and PMS2 (clone MRQ-28, Cell Marque). Normal expression was defined as nuclear staining within tumor cells, using the nuclei of stromal cells and infiltrating lymphocytes as positive internal controls.

The determination of TMB status on the Stanford samples was made using an NGS-based targeted gene panel, the Stanford Actionable Mutation Panel for Solid Tumors (order code: STAMPT, additional assay details available at: <https://stanfordlab.com/content/stanfordlab/en/test-details/s/STAMPT.html>) and/or the FoundationOne CDx panel (Foundation Medicine, Cambridge, Massachusetts, USA). Microsatellite and TMB status were determined by whole-exome sequencing, and MSI was quantified by using MSIsensor at SMC, as previously described.^{31 32}

Statistical analysis

Receiver operating characteristic curves and the AUROC were used to evaluate the performance of the AI models in this study. For PFS and OS estimation, the Kaplan-Meier method was used, and the log-rank test was used to assess differences in PFS and OS between groups. HRs and 95% CIs were computed using the Cox proportional hazards model. Between-group differences in categorical variables were compared using Fisher's exact test, and differences in means or medians for continuous variables were assessed using the non-parametric Mann-Whitney U test. All p values were two-tailed, with a significance threshold of $p < 0.05$.

RESULTS

Distribution of the inflamed immune phenotype in a large-scale ICI-treated cohort

We examined the prevalence of the H&E-based WSI-level IIP (hereafter simply referred to as the IIP) across multiple tumor types and explored its potential as a biomarker for guiding ICI treatment planning, using the retrospectively-collected FFPE pre-ICI treatment tumor WSI from our large, multicenter cohort of ICI-treated patients (N=1,806) representing over 27 different solid tumor types (online supplemental figure S1 and table 3). The clinical and histopathologic characteristics of these patients are summarized in table 1 and table 2, respectively. Most samples were collected from the primary tumor (62.0%), and the most prevalent tumor type was NSCLC (49.7%). 1,502 (83.2%) patients received ICI monotherapy (mono) and 304 (16.8%) received an ICI in combination with at least one other anti-neoplastic drug (ICI combo). Most patients received the ICI as part of their first (25.1%) or second line (43.5%) of treatment.

Of the 798 patients with available PD-L1 TPS results, the proportions with PD-L1 TPS < 1% and PD-L1 TPS ≥ 1% were 24.9% and 75.1%, respectively. Among the patients with both microsatellite and TMB status available (n=130), 67.7% had microsatellite stable/low microsatellite instability (MSS/MSI-L), TMB-low tumors, while 32.3% had MSI or TMB-high tumors (cut-off of 10 mutations per Mb).

By Lunit SCOPE IO analysis, 636 of the 1,806 patients (35.2%) were classified as IIP (online supplemental table 4). The IIP was highly enriched in patients with nasopharyngeal carcinoma (68.0%), melanoma (56.3%), renal cell carcinoma (52.9%), and NSCLC (33.7%) (figure 2). With regard to ICI treatment line, 39.4%, 34.2%, and 31.5% of patients receiving first-line, second-line, and ≥ third-line treatment were classified as IIP. The IIP proportion was 40.7% in TPS ≥ 1% patients and 21.6% in TPS < 1% patients, respectively. In the primary tumor, lymph node, and distant metastatic samples, the IIP proportions were 35.1%, 41.4%, and 32.2%, respectively. While 33.3% of tumors that were MSI and/or TMB-high (≥ 10 mutations/Mb) were IIP, a substantial proportion (26.1%) of tumors that were both MSS/MSI-L and TMB-low were also IIP (online supplemental figure S2A–E).

Association between the inflamed immune phenotype and ICI-treatment outcomes across multiple tumor types when applying a universal threshold

In the overall cohort (N=1,806), the objective response rate (ORR) was significantly higher in IIP than in non-IIP patients (26.3% vs 15.8%, $p < 0.001$, figure 3A) and there was a significant downward trend in the median IS with respect to RECIST response groups ($p < 0.001$, online supplemental figure S3A). Interestingly, only the IS was positively prognostic of response to ICI, but not the IES and IDS (online supplemental table 5 and figure S3B). In the subset of 798 patients with available PD-L1 TPS results, the AUROC for predicting the best overall ICI response

**Table 1** Clinical characteristics of the multicenter study cohort

| | All (N=1,806) | Stanford (n=688) | SMC (n=653) | SNUBH (n=269) | CNUH (n=183) | Northwestern (n=13) |
|--|------------------|---------------------|----------------|------------------|-----------------|------------------------|
| Gender, n (%) | | | | | | |
| Female | 577 (31.95) | 267 (38.81) | 153 (23.43) | 92 (34.20) | 59 (32.24) | 6 (46.15) |
| Male | 1229 (68.05) | 421 (61.19) | 500 (76.57) | 177 (65.80) | 124 (67.76) | 7 (53.85) |
| ECOG PS, n (%) | | | | | | |
| Patients evaluated | 1060 | 0 | 642 | 230 | 183 | 5 |
| 0 | 60 (5.66) | 0 (0.00) | 46 (7.17) | 10 (4.35) | 3 (1.64) | 1 (20.0) |
| 1 | 882 (83.21) | 0 (0.00) | 518 (80.69) | 185 (80.43) | 176 (96.17) | 3 (60.00) |
| 2 | 116 (10.94) | 0 (0.00) | 77 (11.99) | 34 (14.78) | 4 (2.19) | 1 (20.0) |
| 3 | 2 (0.19) | 0 (0.00) | 1 (0.15) | 1 (0.44) | 0 (0.00) | 0 (0.00) |
| ICI treatment, n (%) | | | | | | |
| Pembrolizumab | 914 (50.61) | 388 (56.39) | 326 (49.92) | 119 (44.24) | 71 (38.80) | 10 (76.92) |
| Nivolumab | 664 (36.77) | 226 (32.85) | 264 (40.43) | 97 (36.06) | 75 (40.98) | 2 (15.38) |
| Atezolizumab | 188 (10.41) | 53 (7.70) | 45 (6.89) | 53 (19.70) | 37 (20.22) | 0 (0.00) |
| Avelumab | 18 (1.00) | 3 (0.44) | 15 (2.30) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Durvalumab | 12 (0.66) | 9 (1.31) | 2 (0.31) | 0 (0.00) | 0 (0.00) | 1 (7.70) |
| Others (cemiplimab, tislelizumab) | 10 (0.55) | 9 (1.31) | 1 (0.15) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Treatment line, n (%) | | | | | | |
| Patients evaluated | 1376 | 258 | 653 | 269 | 183 | 13 |
| First-line | 345 (25.07) | 136 (52.70) | 126 (19.30) | 27 (10.04) | 46 (25.14) | 10 (76.92) |
| Second-line | 599 (43.53) | 103 (39.90) | 303 (46.40) | 119 (44.24) | 72 (39.34) | 2 (15.38) |
| ≥Third-line | 432 (31.40) | 19 (7.40) | 224 (34.30) | 123 (45.72) | 65 (35.52) | 1 (7.70) |
| Regimen, n (%) | | | | | | |
| ICI monotherapy | 1502 (83.17) | 430 (62.50) | 624 (95.56) | 261 (97.03) | 183 (100.00) | 4 (30.77) |
| ICI combination therapy | 304 (16.83) | 258 (37.50) | 29 (4.44) | 8 (2.97) | 0 (0.00) | 9 (69.23) |
| ICI combination therapy regimen, n (%) | | | | | | |
| Patients evaluated | 193 | 148 | 28 | 8 | 0 | 9 |
| IO+IO | 36 (18.65) | 35 (23.65) | 0 (0.00) | 0 (0.00) | 0 (0.00) | 1 (11.11) |
| IO+anti-VEGF | 27 (13.99) | 22 (14.87) | 1 (3.57) | 4 (50.00) | 0 (0.00) | 0 (0.00) |
| IO+chemotherapy | 95 (49.22) | 57 (38.51) | 27 (96.43) | 4 (50.00) | 0 (0.00) | 7 (77.78) |
| IO+PARP inhibitor | 7 (3.63) | 7 (4.73) | 0 (0.00) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| IO+RAF/MEK inhibitor | 5 (2.59) | 5 (3.38) | 0 (0.00) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| IO+others* | 23 (11.92) | 22 (14.86) | 0 (0.00) | 0 (0.00) | 0 (0.00) | 1 (11.11) |
| Specimen type, n (%) | | | | | | |
| Patients evaluated | 1780 | 663 | 653 | 269 | 183 | 12 |
| Surgery | 714 (40.11) | 368 (55.51) | 159 (24.35) | 97 (36.06) | 90 (49.18) | 0 (0.00) |
| Biopsy | 1066 (59.89) | 295 (44.49) | 494 (75.65) | 172 (63.94) | 93 (50.82) | 12 (100.00) |
| Tissue harvest site, n (%) | | | | | | |
| Patients evaluated | 1803 | 688 | 653 | 269 | 182 | 11 |
| Primary tumor | 1118 (62.01) | 362 (52.62) | 359 (54.98) | 250 (92.94) | 145 (79.67) | 2 (18.18) |
| Lymph node metastasis | 256 (14.20) | 73 (10.61) | 169 (25.88) | 8 (2.97) | 3 (1.65) | 3 (27.27) |
| Distant metastasis | 429 (23.79) | 253 (36.77) | 125 (19.14) | 11 (4.09) | 34 (18.68) | 6 (54.55) |

Continued

Table 1 Continued

| | All (N=1,806) | Stanford (n=688) | SMC (n=653) | SNUBH (n=269) | CNUH (n=183) | Northwestern (n=13) |
|--|------------------|---------------------|----------------|------------------|-----------------|------------------------|
| Time between tissue harvest date and start of ICI treatment, n (%) | | | | | | |
| Patients evaluated | 1796 | 682 | 653 | 266 | 182 | 13 |
| <365 days | 1193 (66.43) | 445 (65.25) | 450 (68.91) | 162 (60.90) | 123 (67.58) | 13 (100.00) |
| ≥365 days | 603 (33.57) | 237 (34.75) | 203 (31.09) | 104 (39.10) | 59 (32.42) | 0 (0.00) |

*Others include miscellaneous targeted therapy drugs and early phase immunotherapy drugs that are designed for targets other than PD-1/PD-L1 or CTLA-4.

CNUH, Chonnam National University Hospital ; CTLA-4, cytotoxic T-lymphocyte associated antigen 4; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ICI, immune checkpoint inhibitor; IO, immuno-oncology; MEK, mitogen-activated protein kinase; PARR, poly-ADP ribose polymerase; PD-1, programmed cell death protein 1; PD-L1, programmed cell death ligand 1 ; RAF, rapidly accelerated fibrosarcoma kinase; SMC, Samsung Medical Center; SNUBH, Seoul National University Bundang Hospital ; VEGF, vascular endothelial growth factor.

was 0.60 for the IS and 0.68 for the TPS (online supplemental table 6). However, in the subset of these patients without NSCLC, the corresponding AUROC was 0.70 for the IS and 0.64 for the TPS (online supplemental table 6). Median PFS was significantly longer in IIP compared with non-IIP patients (5.3 vs 3.1 months, HR 0.68, 95% CI 0.61 to 0.76, $p<0.001$) (figure 3A). A similar improvement in the median OS after ICI treatment was also observed in IIP compared with non-IIP patients (25.3 vs 13.6 months, HR 0.66, 95% CI 0.57 to 0.75, $p<0.001$). The same trends were observed in the subset of 909 patients without NSCLC (online supplemental figure S4), suggesting that the results for the overall cohort were not solely driven by effects in the NSCLC subset.

On subgroup analysis, the IIP was prognostic of favorable PFS, irrespective of ICI regimen (monotherapy HR 0.68, 95% CI 0.60 to 0.77, $p<0.001$; combo therapy HR 0.68, 95% CI 0.51 to 0.91, $p=0.008$, figure 3B) and PD-L1 TPS status (positive HR 0.67, 95% CI 0.56 to 0.81, $p<0.001$; negative HR 0.66, 95% CI 0.44 to 0.98, $p=0.038$, (online supplemental figure S5A).

Additionally, the IIP was consistently prognostic of favorable response to ICI across various subgroups, including: first-line and second-line ICI treatment; the timing of FFPE tissue collection relative to the start of ICI treatment (less than or ≥ 1 year before ICI treatment); MSS/MSI-L and TMB-low; histologic subtype; specimen type (surgical resection or biopsy); and tissue harvest site (primary, lymph node or distant metastasis). However, the IIP was not significantly associated with favorable response to ICIs in the MSI and TMB-high subgroups or in patients with squamous cell carcinomas (online supplemental figure S5B).

Application of individual (tumor type-specific) thresholds in defining the inflamed immune phenotype

Given the variability in the proportions of IIP patients within each tumor type when applying a universal threshold, we also performed a stratified analysis in which an individual threshold, defined as the IS which distinguished the top 20% of IS's from the remaining

80% within each tumor type, was used to define the IIP. Although these individual thresholds varied across the cancer types, the overall trends in improved clinical outcomes for IIP patients were similar to those observed when using the universal threshold. The improvements in ORR, PFS, and OS for the IIP compared with non-IIP patients remained significant after applying individual thresholds for each tumor type within both the overall cohort and the subset of patients with tumors other than NSCLC (online supplemental figure S6A and 7). The subgroup analyses showed similar trends, although the improvement in PFS for IIP versus non-IIP patients did not reach statistical significance within the subgroup of MSS/MSI-L and TMB-low patients (HR 0.43, 95% CI 0.17 to 1.07, $p=0.062$), or in the subgroup of patients receiving the ICI as first-line treatment (HR 0.74, 95% CI 0.53 to 1.04, $p=0.078$, online supplemental figure 6B).

Association between the other immune phenotypes and ICI treatment outcomes

We further analyzed clinical outcomes of ICI treatment with respect to the IES and IDS, using two thresholds: a 20% threshold, consistent with the IS threshold for the IIP, and a 33.3% threshold for ternary classification of the WSI-level IP. In the overall cohort (N=1,806), WSI-level non-IIP patients exhibited a significant increase in median PFS compared with IEP patients (5.0 vs 3.3 months, HR 1.26 with a 20% threshold; 4.9 vs 3.0 months, HR 1.29 with a 33.3% threshold; both $p<0.001$) (online supplemental table 7). Additionally, WSI-level non-IIP patients showed higher ORR at both thresholds—25.5% (20% threshold) and 24.7% (33.3% threshold)—compared with IEP patients (both $p<0.001$). However, there were no significant differences in OS between WSI-level non-IIP and IEP patients. With regard to the IDS, median OS was significantly increased in WSI-level non-IDP compared with IDP patients, only at the 33.3% threshold (18.0 vs 14.0 months, HR 1.18, $p=0.014$); no significant differences were observed for the other outcomes.

Table 2 Histopathologic characteristics of the multicenter study cohort

| | All (N=1806) | Stanford (n=688) | SMC (n=653) | SNUBH (n=269) | CNUH (n=183) | Northwestern (n=13) |
|---|-----------------|---------------------|----------------|------------------|-----------------|------------------------|
| Pathology, n (%) | | | | | | |
| Adenocarcinoma | 678 (37.54) | 155 (22.53) | 325 (49.77) | 129 (47.95) | 61 (33.33) | 8 (61.54) |
| Squamous cell carcinoma | 517 (28.63) | 168 (24.42) | 270 (41.35) | 61 (22.68) | 16 (8.74) | 2 (15.38) |
| Others | 611 (33.83) | 365 (53.05) | 58 (8.88) | 79 (29.37) | 106 (57.93) | 3 (23.08) |
| Primary origin, n (%) | | | | | | |
| NSCLC | 897 (49.67) | 117 (17.01) | 535 (81.93) | 230 (85.50) | 2 (1.09) | 13 (100.00) |
| HNSCC | 148 (8.19) | 101 (14.68) | 35 (5.36) | 0 (0.00) | 12 (6.56) | 0 (0.00) |
| Melanoma | 144 (7.97) | 103 (14.97) | 0 (0.00) | 0 (0.00) | 41 (22.40) | 0 (0.00) |
| Urothelial carcinoma | 109 (6.04) | 69 (10.03) | 0 (0.00) | 0 (0.00) | 40 (21.86) | 0 (0.00) |
| Renal cell carcinoma | 87 (4.82) | 73 (10.61) | 0 (0.00) | 0 (0.00) | 14 (7.65) | 0 (0.00) |
| Esophageal carcinoma | 78 (4.32) | 8 (1.16) | 66 (10.11) | 0 (0.00) | 4 (2.19) | 0 (0.00) |
| Biliary tract carcinoma | 42 (2.33) | 6 (0.87) | 0 (0.00) | 0 (0.00) | 36 (19.67) | 0 (0.00) |
| Ovarian carcinoma | 37 (2.05) | 15 (2.18) | 0 (0.00) | 22 (8.18) | 0 (0.00) | 0 (0.00) |
| Non-melanoma skin carcinoma | 32 (1.77) | 32 (4.65) | 0 (0.00) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Colorectal carcinoma | 28 (1.55) | 28 (4.07) | 0 (0.00) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Endometrial carcinoma | 25 (1.38) | 12 (1.74) | 0 (0.00) | 13 (4.83) | 0 (0.00) | 0 (0.00) |
| Nasopharyngeal carcinoma | 25 (1.38) | 13 (1.89) | 12 (1.84) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Breast carcinoma | 24 (1.33) | 24 (3.50) | 0 (0.0) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Hepatocellular carcinoma | 20 (1.11) | 13 (1.89) | 0 (0.0) | 0 (0.00) | 7 (3.83) | 0 (0.00) |
| Salivary gland carcinoma | 14 (0.77) | 10 (1.45) | 4 (0.61) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Other tumor types* | 96 (5.32) | 64 (9.30) | 1 (0.15) | 4 (1.49) | 27 (14.75) | 0 (0.00) |
| PD-L1 22C3 TPS, n (%) | | | | | | |
| Patients evaluated | 798 | 126 | 463 | 207 | 2 | 0 |
| TPS<1% | 199 (24.94) | 48 (38.10) | 100 (21.60) | 51 (24.64) | 0 (0.00) | 0 (0.00) |
| TPS≥1% | 599 (75.06) | 78 (61.90) | 363 (78.40) | 156 (75.36) | 2 (100.00) | 0 (0.00) |
| Microsatellite status and TMB status, n (%) | | | | | | |
| Patients evaluated | 130 | 67 | 63 | 0 | 0 | 0 |
| MSS/MSI-L and TMB-low | 88 (67.69) | 38 (56.72) | 50 (79.37) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| MSI or TMB-high | 42 (32.31) | 29 (43.28) | 13 (20.63) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| Microsatellite status†, n (%) | | | | | | |
| Patients evaluated | 208 | 131 | 77 | 0 | 0 | 0 |
| MSS/MSI-L | 192 (92.31) | 117 (89.31) | 75 (97.40) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| MSI | 16 (7.69) | 14 (10.69) | 2 (2.60) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| TMB status‡, n (%) | | | | | | |
| Patients evaluated | 141 | 62 | 79 | 0 | 0 | 0 |
| TMB-low (<10/Mb) | 113 (80.14) | 45 (72.58) | 68 (86.08) | 0 (0.00) | 0 (0.00) | 0 (0.00) |
| TMB-high (≥10/Mb) | 28 (19.86) | 17 (27.42) | 11 (13.92) | 0 (0.00) | 0 (0.00) | 0 (0.00) |

*Other tumor types included gastric carcinoma, carcinoma of unknown primary, uterine cervical carcinoma, pancreatic carcinoma, thyroid carcinoma, neuroendocrine carcinoma, anal carcinoma, germ cell tumor, adrenal cortical carcinoma, penile carcinoma, prostate carcinoma, primary peritoneal carcinoma, and small intestinal carcinoma.

†Microsatellite status was determined by DNA mismatch repair immunohistochemistry and/or MSI PCR at Stanford, and by whole-exome sequencing (WES) at SMC.

‡TMB status was determined by next-generation sequencing-based targeted gene panels (FoundationOne CDx and/or Stanford Actionable Mutation Panel for Solid Tumors) at Stanford, and by WES at SMC.

CNUH, Chonnam National University Hospital ; HNSCC, head and neck squamous cell carcinoma; MSI, high microsatellite instability/mismatch repair deficiency; MSI-L, low microsatellite instability; MSS, microsatellite stable; NSCLC, non-small cell lung cancer; PD-L1, programmed death-ligand 1; SMC, Samsung Medical Center; SNUH, Seoul National University Bundang Hospital ; TMB, tumor mutational burden; TPS, Tumor Proportion Score.

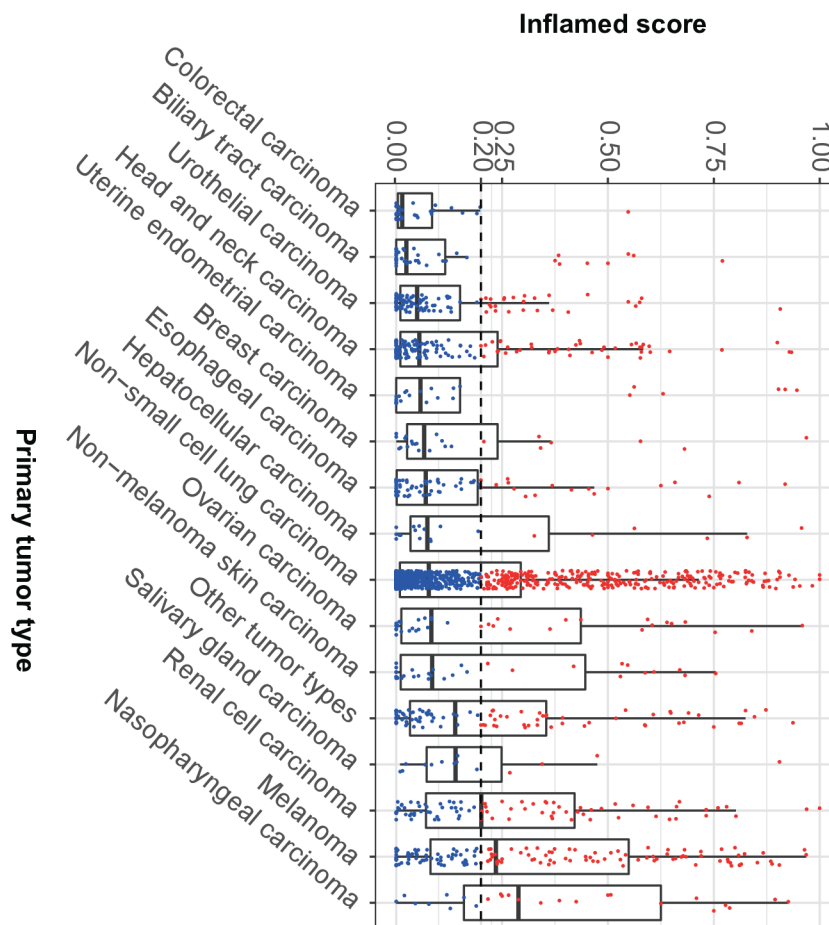


Figure 2 Distribution of Inflamed Scores by tumor type. Box plots depict the distribution of Inflamed Scores (IS), categorized into IIP (red dots) and non-IIP (blue dots) by primary tumor type in the combined cohort (N=1,806). The thick horizontal lines represent the median IS, with boxes delineating the first to third quartiles; the whiskers extend to the minimum and maximum points within 1.5 times the IQR. The dashed line represents the *universal threshold* for the IIP. IIP, inflamed immune phenotype.

DISCUSSION

Here, we present an automated deep learning model, Lunit SCOPE IO, which classifies the immune phenotype of the TME on H&E-stained WSI using TIL distribution and density analysis. We demonstrate, for the first time, the ability of an AI model which runs on routine H&E-stained WSI of pre-ICI treatment FFPE tumor samples to predict ICI clinical outcomes across a broad range of solid tumor types, using a real-world multicenter cohort of 1,806 ICI-treated patients. We show that the IIP, as defined by either a *universal threshold* or *tumor type-specific thresholds*, is significantly prognostic of favorable clinical outcomes after ICI treatment. Furthermore, the IIP appears to correlate with significantly prolonged PFS, regardless of ICI treatment regimen or PD-L1 expression level, and is prognostic of favorable PFS in MSS/MSI-L, TMB-low patients, a clinically important subgroup in whom biomarkers are urgently needed.

Although the utility of the TMB and MSI as universal biomarkers of potential ICI response has primarily been attributed to their association with increased tumor neoantigenicity and heterogeneity of T-cell receptor clones,^{33–35} some criticism has since been directed toward their reliability as predictive biomarkers.³⁶ Furthermore,

even in the TMB-high or MSI setting, which would be expected to increase TIL recruitment to CAs, stromal interference modulated by the transforming growth factor (TGF)-beta or other immunosuppressive pathways may result in the spatial exclusion of TILs from these CAs (as reflected in the IEP).²⁶ This is supported by the observation that, of the patients in our cohort for whom both MSI and TMB status were available, 57% of patients whose tumors were MSI and TMB-high were also of the IEP. In addition, immunoediting, whereby less immunogenic tumor cell clones are selected for, may result in a decreased antitumor immune response, even in tumors with a high mutational burden.³⁷

Immune phenotyping based on TIL analysis avoids many of the limitations of the TMB and MSI. Our TIL analysis directly assesses the degree of lymphocytic infiltration of both CA and CS regions, allowing for the detection of tumors which might be TMB-high or MSI but unresponsive to ICI therapy due to immunoediting or the activation of immunosuppressive pathways, as these tumors would be classified as IEP (eg, non-IIP) tumors by TIL analysis. Furthermore, the H&E-based IIP appears to reflect an active antitumor immune response, based on the observed correlation between the IS and high IFNG

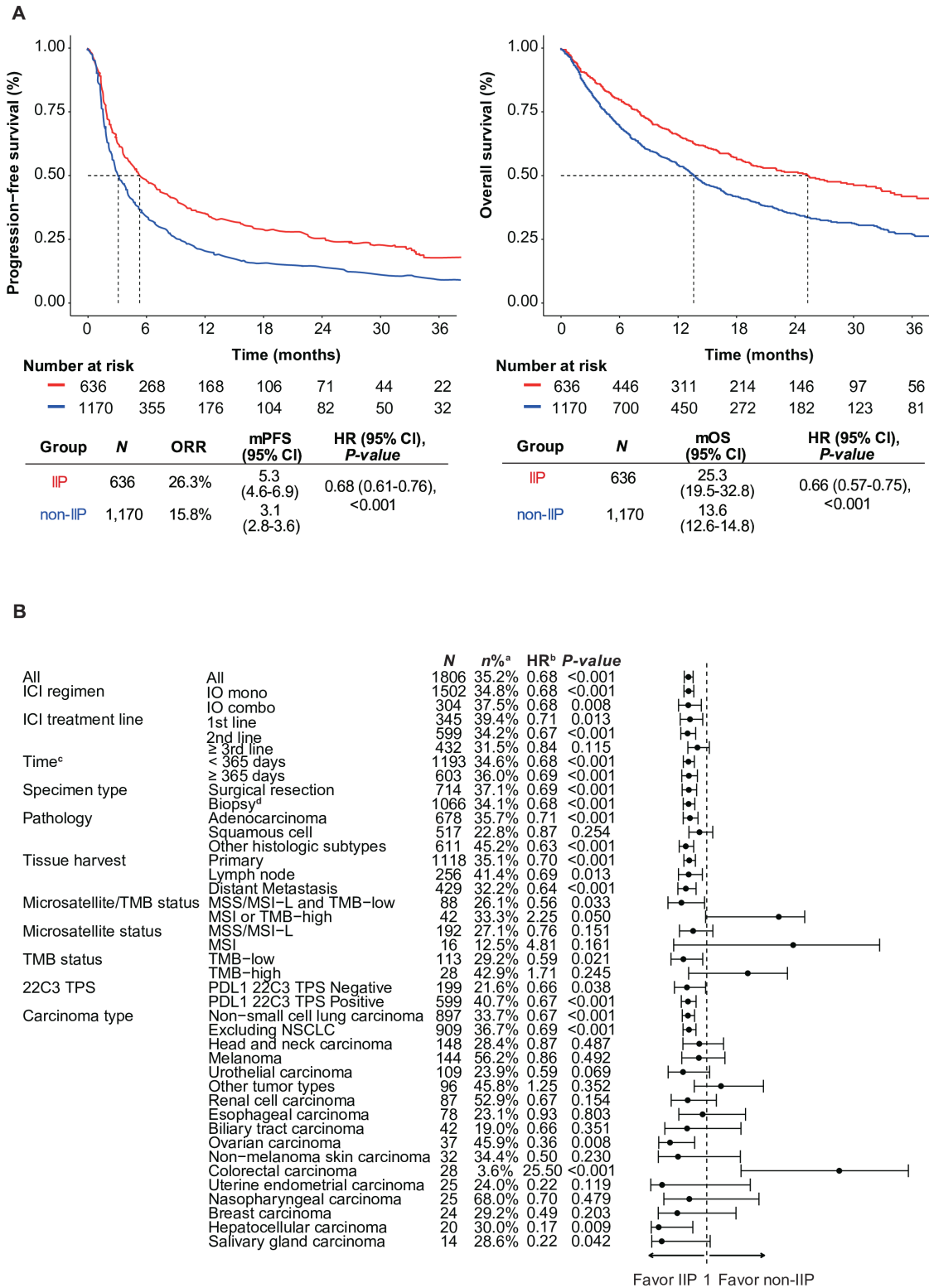


Figure 3 Correlation of clinical outcomes with the immune phenotype across multiple tumor types by the *universal threshold*. (A) Kaplan-Meier survival analysis of PFS (left) and OS (right) after ICI treatment, according to IP (IIP/non-IIP) as defined using the *universal threshold* in the combined cohort ($p < 0.001$). P values were calculated using a two-sided log-rank test. The Cox proportional hazards model was used for calculation of HRs and corresponding 95% CIs. (B) Forest plot of PFS after ICI treatment according to baseline patient characteristics, with comparison of PFS between IIP and non-IIP subgroups, as defined using the *universal threshold* in the combined cohort. Dots and whiskers represent HRs and 95% CIs, respectively. ICI, immune checkpoint inhibitor; IIP, inflamed immune phenotype; IP, immune phenotype; mPFS, median progression-free survival; mOS, median overall survival; MSI-L, low microsatellite instability; MSS, microsatellite stable; NSCLC, non-small cell lung cancer; ORR, objective response rate; OS, overall survival; PD-L1, programmed cell death ligand 1; PFS, progression-free survival; TMB, tumor mutational burden; TPS, Tumor Proportion Score.

pathway activation in a pan-cancer TCGA data set. H&E-based immune phenotyping uses pre-existing FFPE H&E slides collected during the course of routine clinical care, therefore requiring no additional tissue section procurement, and computational TIL assessment will enable more objective time-efficient and labor-efficient analysis at scale, avoiding interobserver variability and bias in interpretation.^{38–40}

The promise of computational TIL analysis of routine H&E-stained images has been demonstrated by earlier studies showing a correlation between the spatial architecture of TILs and patient prognosis (1) in non-ICI-treated patients with cancer^{41–43} and (2) in small cohorts of patients with NSCLC treated with immunotherapy.⁴⁴ For example, in a seminal study using over 5,000 H&E WSIs from 13 tumor types represented in the TCGA, Saltz and colleagues applied deep learning-based binary patch classification (wherein each tissue-containing patch in a WSI was classified as either positive or negative for TILs) to generate a patch-based WSI-level spatial TIL map, finding a significant association between various structural features derived from these TIL maps and OS across four TCGA tumor types (breast invasive adenocarcinoma, lung adenocarcinoma, prostate adenocarcinoma, and cutaneous melanoma).⁴³ Our study builds on all of these prior contributions by presenting the first comprehensive pan-cancer analysis examining the association between AI-enabled TIL-based immune phenotypes, as assessed on routine H&E-stained slides representing over 27 different solid tumor types, and ICI treatment outcomes. Furthermore, we apply a universal cut-off across multiple tumor types to evaluate the feasibility of using the TIL-based immunophenotype as a tumor-type agnostic biomarker of ICI response, which has not previously been done.

We acknowledge that this study was subject to limitations. Given its focus on ICI-treated patients, it is likely that our real-world data set was enriched for IIP patients, as many of the eligibility requirements for ICI treatment are tied to the relative “immunogenicity” of the tumor. Therefore, it is possible that our current model might not generalize as well when applied to less immunogenic tumor types. Due to the retrospective nature of the study, we were unable to systematically control for heterogeneity in patient treatment regimens and other potential confounders. For instance, although previous reports have not clearly shown a significant difference in efficacy between different ICI agents,^{45–47} we could not entirely exclude the possibility of confounding of ICI response rates by the specific ICI regimen used. In addition, the limited number of patients in the current data set for whom ancillary biomarker status was available (MSI, PD-L1 and TMB) precluded more sophisticated analyses of the relationship between these biomarkers and the AI-based immune phenotype. However, the results from our analysis of the TCGA data set showed that the IS was predictive of a high *IFNG* expression signature regardless of TMB status, suggesting that the AI-based immune phenotype contributes additional predictive and prognostic value independent of the TMB.

Also, as the current study was focused on the development and validation of a readily scalable H&E-based TIL analysis model, more granular analyses of individual TIL types and activation states based on immunohistochemical and gene expression profiling were not performed on the current data set. However, in prior analyses, we have found that H&E-based TIL distribution and density analysis indirectly reflects antitumor lymphocytic activity, as assessed by gene expression profile-based cytolytic activity scores and *IFNG* signatures.¹⁸ Lastly, our study was subject to sample size limitations for some cancer types. For example, we were unable to include mesenchymal tumors or other rarer tumor types, which will be important to include in future studies as larger ICI-treated cohorts become available. In addition, it should be noted that, in our study, the directionality of the association between the IIP and PFS for colorectal carcinoma (CRC) was contrary to what might typically be expected. We believe that this might have been due to the small sample size; among the 28 patients with CRC in the study, only one was classified as IIP when we applied the universal IS cut-off. This patient belonged to the MSI-high group and had the best overall response of stable disease with pembrolizumab monotherapy. In the remaining 24 non-IIP patients with BOR data available, four showed a partial or better response. When the IS cut-off for IIP was set to 9.1% (resulting in six IIP patients), the difference in PFS for the IIP versus non-IIP patients with CRC was not found to be statistically significant ($p=0.305$). In future studies, we plan to conduct more comprehensive analyses of the relationships between the IS, PD-L1, MSI status, and TMB in order to develop a more robust model for predicting BOR. We also believe that further investigations in individual tumor types with larger sample sizes are strongly warranted to optimize thresholds and to more definitively determine whether a universal or tumor type-specific threshold would be more appropriate.

Nonetheless, in this first effort to examine the correlation between clinical outcomes (including ORR, PFS, and OS) in ICI-treated patients and the H&E TIL-based immune phenotype, as assessed in a large multi-institutional cohort encompassing multiple diverse tumor types, we observed convincing results suggesting that the IIP may represent a practical, clinically actionable biomarker of favorable clinical outcomes, particularly in patients with PD-L1 negative, MSS/MSI-L, and TMB-low tumors, in whom biomarkers are urgently needed. Furthermore, we demonstrate that the application of deep learning to H&E-based immune phenotyping can provide an automated, readily scalable tool for guiding the selection of patients for ICI treatment across a wide range of different solid tumors. Further optimization and validation of the IIP thresholds used in this study in prospective clinical trials represents an important next step, which, if successful, might 1-day enable more precise selection of patients for ICI therapy.

Author affiliations

¹Department of Pathology, Stanford University School of Medicine, Stanford, California, USA

²Center for Artificial Intelligence in Medicine & Imaging, Stanford University, Stanford, California, USA

³Department of Pathology and Translational Genomics, Sungkyunkwan University School of Medicine, Suwon, Korea (the Republic of)

⁴Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Korea (the Republic of)

⁵Department of Pathology, Chonnam National University Medical School, Gwangju, Korea (the Republic of)

⁶Department of Pathology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Korea (the Republic of)

⁷Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

⁸UCLA Health, University of California, Los Angeles, Los Angeles, California, USA

⁹Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

¹⁰Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea (the Republic of)

¹¹Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Korea (the Republic of)

¹²AMITA Health Saint Francis Hospital Evanston, Evanston, Illinois, USA

¹³Lunit, Seoul, Korea (the Republic of)

¹⁴Department of Pathology, Ajou University School of Medicine, Suwon, Korea (the Republic of)

¹⁵Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea (the Republic of)

¹⁶Department of Clinical Oncology, The Chinese University of Hong Kong, New Territories, Hong Kong

Twitter Gahee Park @loki_phd

Acknowledgements The authors would like to thank Christopher King from the Stanford Department of Pathology and Curtis P. Langlotz and Johanna Kim from the Stanford Center for Artificial Intelligence in Medicine & Imaging for additional administrative support, and Sang Yong Song, Sangjoon Choi, Hyun Ae Jung, Jong Mu Sun, Jin Seok Ahn and Myung Ju Ahn from Samsung Medical Center; Yoo Duk Choi and Kyung Hwa Lee from Chonnam National University Medical School; Soeick Cho, Huijeong Kim, Senghui Seo, Jiwon Shin, Jisoo Shin, Seungje Lee, Eunji Baek, Seonwook Park, Mohammad Mostafavi, Jeongun Ryu, Minuk Ma, Donggeun Yoo, Sangheon Ahn and Kyunghyun Paeng from Lunit; and Koungh Jin Suh, Se Hyun Kim, Yu Jung Kim and Jong Seok Lee of Seoul National University Bundang Hospital for contributing to the development of Lunit SCOPE IO.

Contributors Conception and design: JS, C-YO, Y-JB, and TSKM. Financial support: JS and C-YO. Administrative support: JS, Y-LC, TL, HK, YKC, and S-HL. Provision of study material or patients: JS, Y-LC, TL, HK, YKC, BWD, SB, MH, GAF, SP, S-HL, J-EH, J-HC, and LK. Collection and assembly of data: JS, Y-LC, TL, HK, YKC, SSh, YL, CHO, Seulki Kim, CO, WJ. Data analysis and interpretation: HS, SP, CO, Sukjun Kim, GP, SSo, Seokhwi Kim, Y-JB, TSKM, SA, and C-YO. Manuscript writing: All authors. Final approval of manuscript: All authors. Accountable for all aspects of the work: All authors. Guarantor: C-YO.

Funding Funding for this study was provided by Lunit Inc., with additional infrastructural support from the Stanford Center for Artificial Intelligence in Medicine & Imaging (AIMI) and the Department of Pathology, Stanford University School of Medicine. JS additionally received support from the United States National Cancer Institute (NCI), National Institutes of Health (NIH) (R01 CA270437).

Competing interests JS and SB received institutional research funding from Lunit, Inc. HS, SP, SSh, YL, CHO, Seulki Kim, CO, Sukjun Kim, GP, SSo, WJ, SA and C-YO are employees of Lunit, Inc. Y-JB is a Consultant/Advisory Board member for Merck Sharp and Dohme (MSD), Merck Serono, Daiichi-Sankyo, Astellas, Alexo Oncology, Samyang Biopharm, Hanmi, Daewoong, and Amgen, and received institutional research grants for clinical trials from Genentech/Roche, MSD, Merck Serono, Daiichi Sankyo, Astellas, and Amgen in the past 3 years. Other authors declare no potential conflicts of interest.

Patient consent for publication Not applicable.

Ethics approval All work was conducted according to appropriate guidelines for the protection of human subjects in biomedical research, and with institutional review board (IRB) approval at Stanford (IRB no. 58610), SMC (IRB no. 2018-06-103

and 2021-02-011), SNUBH (IRB no. B-2006-619-307 and B-2209-780-303), CNUH (IRB no. CNUH-2021-023) and Northwestern University (IRB no. STU00207117). Obtaining informed consent from individual patients was waived, considering the retrospective nature of this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The TCGA data used in the current study are publicly available via the National Cancer Institute Genomic Data Commons portal (<https://portal.gdc.cancer.gov/>). The data in the multi-institutional ICI-patient external validation data set are not publicly available due to institutional restrictions governing human subject privacy protection. As these data are stored in controlled access repositories, the data, or a subset of the data, may be made available upon reasonable request following submission of a research protocol detailing intended use and institutional review board ethics review and approval; inquiries should be directed to JS (Stanford), Y-LC (SMC), TL (CNUH), HK (SNUBH), and YKC (Northwestern University). The remaining data are available within the Article and Supplementary materials.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Jeanne Shen <http://orcid.org/0000-0002-1519-0308>

Sehhoon Park <http://orcid.org/0000-0001-9467-461X>

REFERENCES

- Twomey JD, Zhang B. Cancer immunotherapy update: FDA-approved checkpoint inhibitors and companion diagnostics. *AAPS J* 2021;23:39-39..
- Wang Y, Tong Z, Zhang W, *et al.* FDA-approved and emerging next generation predictive biomarkers for immune checkpoint inhibitors in cancer patients. *Front Oncol*;11:2115.
- Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. *J Immunother Cancer* 2019;7:278.
- Paver EC, Cooper WA, Colebatch AJ, *et al.* Programmed death ligand-1 (PD-L1) as a predictive marker for immunotherapy in solid tumours: A guide to immunohistochemistry implementation and interpretation. *Pathology* 2021;53:141-56.
- Marabelle A, Le DT, Ascierto PA, *et al.* Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *JCO* 2020;38:1-10.
- Marabelle A, Fakih M, Lopez J, *et al.* Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol* 2020;21:1353-65.
- Cortes-Ciriano I, Lee S, Park WY, *et al.* A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* 2017;8:15180.
- Bonneville R, Krook MA, Kautto EA, *et al.* Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol* 2017;2017:PO.17.00073.
- Kim ES, Velcheti V, Mekhail T, *et al.* Blood-based tumor mutational burden as a biomarker for atezolizumab in non-small cell lung cancer: the phase 2 B-F1RST trial. *Nat Med* 2022;28:939-45.
- Samstein RM, Lee C-H, Shoushtari AN, *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* 2019;51:202-6.

- 11 Rizvi NA, Cho BC, Reinmuth N, *et al.* Durvalumab with or without tremelimumab vs standard chemotherapy in first-line treatment of metastatic non-small cell lung cancer: the MYSTIC phase 3 randomized clinical trial. *JAMA Oncol* 2020;6:661–74.
- 12 McGrail DJ, Pilié PG, Rashid NU, *et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol* 2021;32:661–72.
- 13 Alexandrov LB, Nik-Zainal S, Wedge DC, *et al.* Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
- 14 Pajjens ST, Vledder A, de Bruyn M, *et al.* Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell Mol Immunol* 2021;18:842–59.
- 15 Amgad M, Stovgaard ES, Balslev E, *et al.* Report on computational assessment of tumor infiltrating lymphocytes from the international immuno-oncology biomarker working group. *NPJ Breast Cancer* 2020;6:1–13.
- 16 Paeng K, Park S, Kim M, eds. *A Unified Framework for Tumor Proliferation Score Prediction in Breast Histopathology*. Cham: Springer International Publishing, 2017.
- 17 Pantanowitz L, Hartman D, Qi Y, *et al.* Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagn Pathol* 2020;15:1–10.
- 18 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;40:1916–28.
- 19 Jung HA, Park K-U, Cho S, *et al.* A phase II study of nivolumab plus gemcitabine in patients with recurrent or metastatic nasopharyngeal carcinoma (KCSG HN17-11). *Clin Cancer Res* 2022;28:4240–7.
- 20 Roh W, Geffen Y, Cha H, *et al.* High-resolution profiling of lung adenocarcinoma identifies expression subtypes with specific biomarkers and clinically relevant vulnerabilities. *Cancer Res* 2022;82:3917–31.
- 21 Chen L-C, Zhu Y, Papandreou G, *et al.* Encoder-Decoder with Atrous separable Convolution for semantic image Segmentation. proceedings of the European conference on computer vision (ECCV). 2018
- 22 He K, Zhang X, Ren S, eds. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
- 23 Milletari F, Navab N, Ahmadi S-A. V-Net: fully Convolutional neural networks for volumetric medical image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV); Stanford, CA, USA.
- 24 Kingma D, Adam JB. A method for stochastic optimization in: International conference on learning representations (ICLR). arXiv preprint arXiv:1412.6980; 2014.
- 25 Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017;541:321–30.
- 26 Mariathasan S, Turley SJ, Nickles D, *et al.* TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 2018;554:544–8.
- 27 Ayers M, Lunceford J, Nebozhyn M, *et al.* IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017;127:91190:2930–40.:
- 28 The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;45:1113–20.
- 29 Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- 30 Roach C, Zhang N, Corigliano E, *et al.* Development of a Companion Diagnostic PD-L1 Immunohistochemistry Assay for Pembrolizumab Therapy in Non-Small-cell Lung Cancer. *Applied Immunohistochemistry & Molecular Morphology* 2016;24:392–7.
- 31 Shim JH, Kim HS, Cha H, *et al.* HLA-corrected tumor mutation burden and homologous recombination deficiency for the prediction of response to PD-(L)1 blockade in advanced non-small-cell lung cancer patients. *Ann Oncol* 2020;31:S0923-7534(20)39295-4:902–11.:
- 32 Niu B, Ye K, Zhang Q, *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 2014;30:1015–6.
- 33 Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* 2015;348:69–74.
- 34 Le DT, Uram JN, Wang H, *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015;372:2509–20.
- 35 Timmermann B, Kerick M, Roehr C, *et al.* Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLOS ONE* 2010;5:e15661e15661.
- 36 McGranahan N, Rosenthal R, Hiley CT, *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 2017;171:S0092-8674(17)31185-6:1259–1271..
- 37 Schreiber RD, Old LJ, Smyth MJ. Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion. *Science* 2011;331:1565–70.
- 38 Denkert C, von Minckwitz G, Darb-Esfahani S, *et al.* Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol* 2018;19:40–50.
- 39 Brunyé TT, Mercan E, Weaver DL, *et al.* Accuracy is in the eyes of the pathologist: The visual interpretive process and diagnostic accuracy with digital whole slide images. *J Biomed Inform* 2017;66:171–9.
- 40 Brambilla E, Le Teuff G, Marguet S, *et al.* Prognostic effect of tumor lymphocytic infiltration in resectable non-small-cell lung cancer. *J Clin Oncol* 2016;34:1223–30.
- 41 Yuan Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. *J R Soc Interface* 2015;12:20141153:103.:
- 42 Corredor G, Wang X, Zhou Y, *et al.* Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non-small cell lung cancer. *Clin Cancer Res* 2019;25:1526–34.
- 43 Saltz J, Gupta R, Hou L, *et al.* Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep* 2018;23:181–93.
- 44 Barrera C, Corredor G, Viswanathan VS, *et al.* Deep computational image analysis of immune cell niches reveals treatment-specific outcome associations in lung cancer. *NPJ Precis Oncol* 2023;7:52.
- 45 Al-Showbaki L, Nadler MB, Desnoyers A, *et al.* Network meta-analysis comparing efficacy, safety and tolerability of Anti-PD-1/PD-L1 antibodies in solid cancers. *J Cancer* 2021;12:14:4372–8.:
- 46 Duan J, Cui L, Zhao X, *et al.* Use of immunotherapy with programmed cell death 1 vs programmed cell death ligand 1 inhibitors in patients with cancer: a systematic review and meta-analysis. *JAMA Oncol* 2020;6:375–84.
- 47 Jiang M, Liu C, Ding D, *et al.* Comparative efficacy and safety of Anti-PD-1/PD-L1 for the treatment of non-small cell lung cancer: a network meta-analysis of 13 randomized controlled studies. *Front Oncol* 2022;12:827050.