npj | precision oncology

Review article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-024-00671-1

The current landscape of spatial biomarkers for prediction of response to immune checkpoint inhibition

Hannah L. Williams[1,7] ✉, Ana Leni Frei[1,2,7], Thibaud Koessler[3,4,5], Martin D. Berger[6], Heather Dawson[1], Olivier Michielin[3,4,5] & Inti Zlobec[1]

Enabling the examination of cell-cell relationships in tissue, spatially resolved omics technologies have revolutionised our perspectives on cancer biology. Clinically, the development of immune checkpoint inhibitors (ICI) has advanced cancer therapeutics. However, a major challenge of effective implementation is the identification of predictive biomarkers of response. In this review we examine the potential added predictive value of spatial biomarkers of response to ICI beyond current clinical benchmarks.

ICI has demonstrated moderate efficacy across a variety of cancer types. This was highlighted in a multi-cancer study of ICI response across 27 cancer types. An objective response rate of >30% was observed in four cancer types (cutaneous squamous cell carcinoma, non-colorectal mismatch repair deficient (dMMR) tumours, melanoma, and Merkel-cell carcinoma). Despite this clinical success, approx. 37% of cancer types demonstrated an overall Response Rate (ORR) ≤ 10%[1,2]. Additionally, ICI-associated immune-related toxicities (immune-related adverse events (irAEs)) are not uncommon[3–5]. While their presence is thought to reflect an engaged immune response and is commonly associated with an observed anti-tumour response, clinical manifestations can impact a broad range of organs and in some cases, prove to be fatal (up to 1.3% of treated patients)[6]. Identification of a predictive biomarker for ICI response is of great importance to preserve patients from treatment-associated toxicities and relieve the financial burden on health systems from the use of inefficient therapies.

The current landscape of predictive biomarkers for ICI

The most prominently examined predictive biomarkers for response to ICI include PD-L1 expression (by immunohistochemistry), tumour mutational burden (TMB, the median number of mutations per megabase) and microsatellite status. TMB testing is FDA-approved in patients with unresectable or metastatic solid tumours for treatment with pembrolizumab following progression after prior treatment[7] while in patients with locally advanced or metastatic colorectal cancer, microsatellite instability (MSI) serves as a predictive biomarker and is associated with a favourable clinical outcome[8]. PD-L1 expression by immunohistochemistry is a widely implemented predictive biomarker for the prediction of response to anti-PD-1 and anti-PD-L1-based therapies for numerous tumour types[9]. For example, for non-small cell lung cancer (NSCLC) and gastric cancer, PD-L1 expression is FDA-approved as a companion diagnostic for treatment with pembrolizumab (PD-L1 IHC 22C3 pharmDx (Dako)[10].

PD-L1 expression can be quantified using the tumour proportion score (TPS), which accounts for PD-L1 expression in tumour cells, or the combined positive score (CPS) which accounts for PD-L1 staining in both tumour cells and surrounding immune cells[11]. The TPS is considered a reliable method for NSCLC; however, less robust for other solid cancers where the CPS is more commonly applied. Despite the common evaluation of PD-L1 expression, durable benefit from ICI has been observed in patients with low or non-detectable PD-L1 expression[12] questioning the sensitivity of the assay. Equally, PD-L1-positive patients with no response to ICI treatment have also been observed[13]. There is a clear need for more reliable predictive biomarkers to ICI and scope for a potential paradigm change in assessing patient outcomes for such in clinical trials[10]. Spatially resolved quantification of biomarkers such as PD-L1 or the discrete localisation of cell-cell interactions holds potential value for the identification of a new generation of "spatial biomarkers" of response to therapies such as ICI.

Deriving spatial context from tissue

We define spatial biomarkers as those derived from the quantitation of the spatial relationships of cellular components or target expression within tissue e.g., distance between cells, geographical distribution of target expression and/or cellular biomarkers across tissue. Enabling this quantitation requires the use of technical modalities which measure RNA or protein expression while retaining tissue architecture and cellular

[1]Institute for Tissue Medicine and Pathology, University of Bern, Bern, Switzerland. [2]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland. [3]Medical Oncology Department, Geneva University Hospitals, 4 rue Gabrielle-Perret-Gentil, 1205 Geneva, Switzerland. [4]Swiss Cancer Centre Léman, Lausanne, Geneva, Switzerland. [5]University of Geneva, Faculty of Medicine, Geneva, Switzerland. [6]Department of Medical Oncology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. [7]These authors contributed equally: Hannah L. Williams, Ana Leni Frei. ✉e-mail: hannah.williams@unibe.ch

organisation. The past decade has seen a rapid expansion in the development and commercialisation of spatially resolved omics technologies. Such technologies enable the visualisation and quantitation of numerous molecular targets and cellular components within tissue while retaining tissue architecture. A variety of spatially resolved omics technologies exist and have been comprehensively reviewed[14–17]. The studies featured in this review comprise largely of image-based technologies including imaging mass cytometry (IMC), co-detection by indexing (CODEX / Phenocycler (Akoya Biosciences), multiplex immunofluorescence (sequential, cyclic (CycIF) and MILAN), immunohistochemistry and one sequencing-based technology, the Digitial Spatial Profiler (DSP, Nanostring) (Fig. 1). The ability of such technologies to examine a high number of molecular targets means that numerous cell types can be examined within the same tissue section. The resolution of technologies ranges from user-selected regions of interest (ROIs) to single cells (cell segmentation). The retention of the spatial architecture of such features enables the extraction of spatial (x and y) coordinates and subsequent quantitation of their spatial distribution within the tissue which requires the implementation of a variety of computational methods. The most common spatial analysis methods described in this review include cell-cell distance calculation, nearest neighbour distance quantitation, radius-based cell density/frequency quantitation or target interaction/co-expression through image masking (Fig. 2), additional methods not featured have been reviewed by others[18–20]. Despite the appeal of newer spatial omics technologies, it is important to consider the ability to extract comparable spatial information (albeit at lower numbers of targets) from more traditional methods within pathology such as haematoxylin and eosin (H&E) or immunohistochemically stained sections. This type of spatial analysis often requires the implementation of more advanced computational modalities, such as artificial intelligence (AI) to see patterns beyond the naked eye.

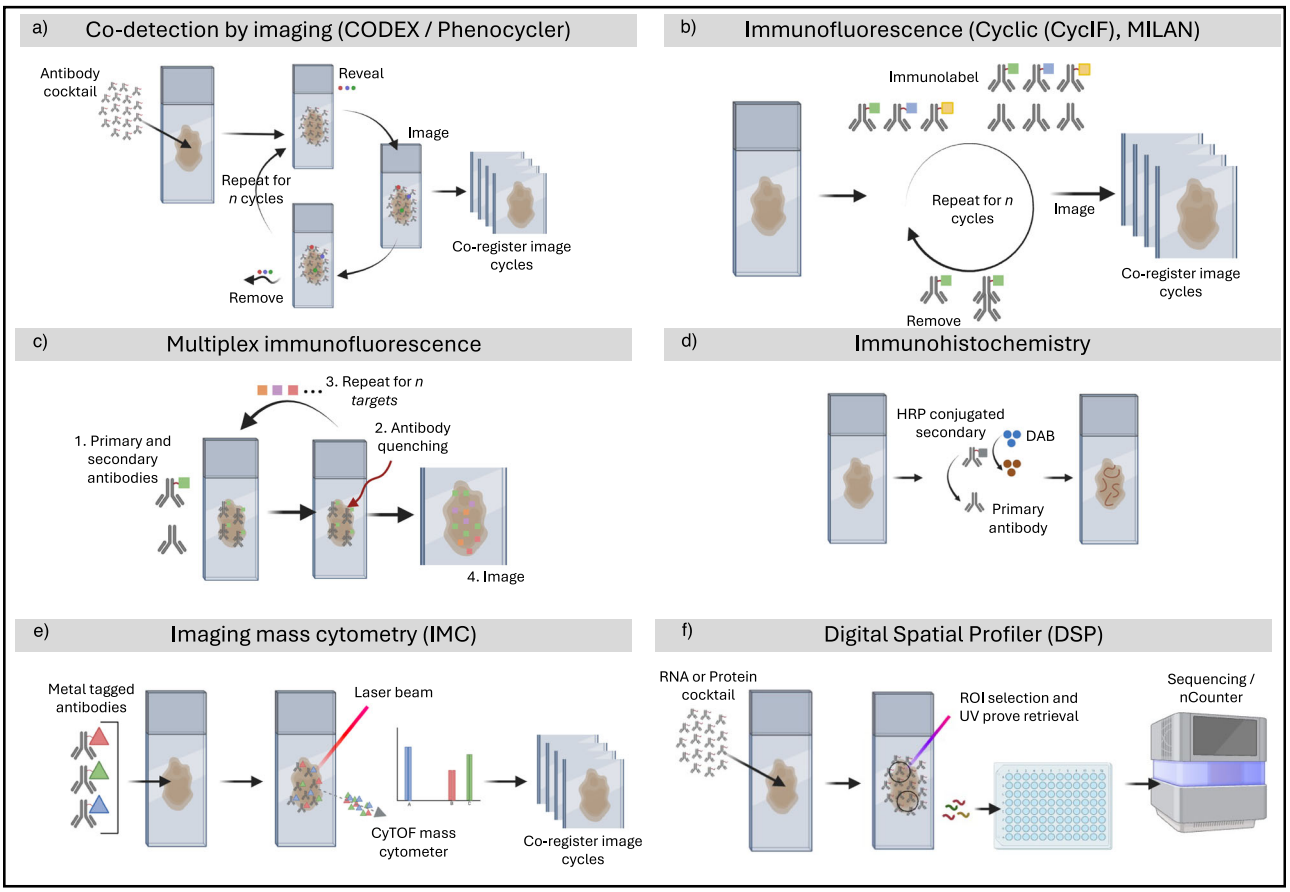

Fig. 1 | Technology and methods to acquire spatially resolved data and quantify spatial relationships. a Co-detection by imaging (CODEX/formerly Phenocycler) (Image-based, IB). Following labelling using an antibody cocktail conjugated with unique DNA barcodes a cyclical process of tissue imaging and reporter probe removal. Image co-registration of cycles enables downstream image analysis. **b** *Immunofluorescence (Cyclic (CycIF), multiple iterative labelling by antibody neo-deposition (MILAN) (IB)*: Iterative cycles of antibody incubation, CycIF: groups of 3 antibodies labelled with different fluorescent dyes for antibody incubation. MILAN: unconjugated primary antibodies followed by fluorescent labelled secondary antibodies, target imaging and removal. Images from each visualisation cycle are co-registered, and downstream image analysis is applied. **c** *Multiplex immunofluorescence (mIF) (IB)*: Iterative process of single antibody incubation, fluorescent visualisation with tyramide signal amplification and antibody removal. One imaging step acquires all fluorescent channels to visualise antibodies in the one tissue section to which downstream image analysis is applied. **d** *Immunohistochemistry (IHC) (IB)*: Visualisation of a single protein target, can also be performed for the co-visualisation of multiple targets. Primary antibody binding, secondary antibody conjugated to a chromogenic detection molecule, e.g., horseradish peroxidase (HRP) and visualisation through conversion of a chromogenic substrate, e.g., 3, 3'-diaminobenzidine (DAB) by HRP into a coloured signal to be visualised under a light microscope or brightfield scanner. **e** *Imaging mass cytometry (IMC) (IB)*: Following labelling with a cocktail of metal-tagged antibodies, labelled tissue undergoes laser ablation which vaporises the tissue (with bound metal-tagged antibodies) in sequential small spots. The ionised tissue is put through a mass cytometer, which detects and quantifies the metal isotopes and correlates them to specific antibodies. The intensity of each ablated spot is collated for each metal isotope and the spatial information on each spot is used to reconstruct images for downstream image analysis. **f** *Digital Spatial Profiler (DSP) (Sequencing-based)*: Applicable for the detection of RNA or protein molecules. For protein detection: following incubation with a cocktail of barcoded antibodies, fluorescently tagged primary antibodies are applied to aid region of interest (ROI) selection. For each ROI UV cleavage and aspiration of the antibody barcodes will proceed. Each aspirate from a unique ROI will be deposited into a 96-well plate for transcript counting via nCounter or NGS.

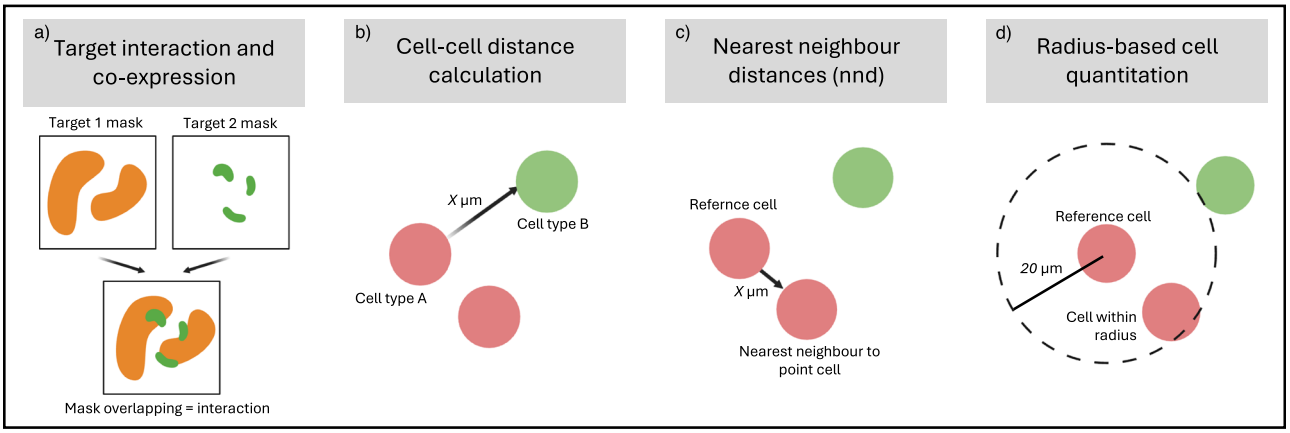

Fig. 2 | Data analysis methods to quantify spatial relationships from spatial omics technologies. a *Target interaction and co-expression*: The co-localisation of a target expression can be used to infer target interaction. This is achieved through the generation of masks per target of interest. Images with the corresponding masks are co-registered and regions of mask overlap are determined as interaction of targets. A similar approach can be utilised to assess target co-expression. For this, normally, cells are segmented from digital images and assessed for the co-expression of the target of interest. Once identified, the frequency or spatial organisation of co-expressing or interacting cells/targets can be further quantified. **b** *Cell–cell distance calculation*: The calculation of cell-cell distances involves the assignment of a reference cell phenotype (cell A) upon which distances are calculated to a second cell phenotype (Cell B). This is performed for each individual cell of the reference phenotype, and distance calculations are averaged to provide a reflection of the overall spatial distribution of cell B to cell A. **c** *Nearest neighbour distances*: This approach quantifies the distance between a reference Cell A and its nearest neighbouring cell. This calculation can be performed iteratively for each cell of interest, and distances averaged to provide a global overview of the proximity of neighbouring cells to a specific reference cell of interest. **d** *Radius-based cell quantitation*: For radius-based cell quantitation a reference cell A is defined upon which cell numbers and cell phenotypes within a radius of varying size are quantified.

Whether high plex image-based technologies or through the application of AI methods on H&E, the spatially resolved quantification of biomarkers such as PD-L1 or the discrete localisation of cell–cell interactions hold the potential to identify a new generation of "spatial biomarkers" of response to therapies such as ICI. Reflecting upon the title of this review, "how soon is now," we evaluated current literature examining spatial biomarkers of response to anti-PD-1 and or anti-PD-L1 therapy to answer: What is the current landscape of spatial biomarkers in the field of ICI response? Do spatial biomarkers provide added predictive value compared to current clinical benchmarks? How close to the clinical application are such spatial biomarkers?

Search criteria

We performed a literature search on PubMed in November 2023 for clinical trials and clinical cohorts examining image-based spatial biomarkers using validated search terms[21]. Inclusion criteria required the investigation of spatially resolved image-based biomarkers of response to immunotherapy on tissue which yielded 20 studies in total. This review focuses on tissue-based modalities for the assessment of spatial biomarkers from routine pathology specimens. Therefore, studies assessing imaging biomarkers from radiology, positive emission tomography, magnetic resonance imaging and computed tomography while holding the potential to obtain spatial biomarkers have been excluded from this review. Multiple studies in melanoma and NSCLC were identified enabling comparisons to be made between the types of spatial biomarkers identified in each disease. Additionally, we identified 6 studies assessing the use of AI for the identification of spatially resolved morphological biomarkers of response from H&E and IHC as well as automation of the quantitation of current benchmarks such as PD-L1 expression.

Multiplexed image-based spatial biomarkers of response

Melanoma

We identified seven studies assessing spatial biomarkers of response to ICI in melanoma. All studies included single-cell methodologies assessing protein expression, such as imaging mass cytometry or multiplex immunofluorescence (mIF) and one study applied digital spatial profiling (Nanostring) to assess compartment-level expression. From the studies summarised (Fig. 3, Table. 1), two spatially resolved features were frequently identified as predictors of response to ICI in metastatic melanoma patients: 1. PD-L1$^+$ macrophages, and 2. T-cells. Toki et al.[22] employed a compartment-based strategy in combination with DSP to examine the bulk expression of a panel of 44 protein markers as biomarkers of response in 60 pre-treatment melanoma formalin-fixed, paraffin-embedded (FFPE) tissue specimens comprised in a tissue microarray from stages I to IV. The three compartments include (1) CD68$^+$ macrophages, (2) CD45+ leucocytes and (3) S100 + HMB45$^+$ melanocytes. In univariate analysis assessment of target counts in each compartment revealed that high CD8+ counts in the macrophage compartment were significantly associated with response (CR/PR) ($P = 0.014$), progression-free survival (PFS) ($P = 0.0082$; HR, 0.42; 95% CI, 0.22–0.83) and prolonged overall survival (OS) ($P = 0.0119$; HR, 0.33; 95% CI, 0.14–0.78), indicating that a potential interaction between CD8+ T-cells and macrophages underly response mechanisms. In multivariate analysis, high PD-L1+ expression in the macrophage compartment was significant for response (CR/PR), PFS ($P = 0.0072$; HR, 0.36; 95% CI, 0.18–0.69) and OS ($P = 0.0032$; HR, 0.15; 95% CI, 0.065–0.35). This feature distinguished responders from non-responders irrespective of bulk PD-L1 expression in the melanocyte or the leucocyte compartments.

Using a similar compartment-based analysis, Lu et al.[23] identified in 11/23 high-risk resectable melanoma patients treated with either nivolumab only or a combination of ipilumumab/nivolumab high levels of ß2M, Beta-catenin, CD19 and CD8A can distinguish responders and non-responders (AUC = 0.998). ß2M, CD19 and CD8A are correlated with response, while Beta-catenin is correlated with non-response. Quantitation of the number of adjacent immune cells surrounding a given immune cell revealed that a higher immune neighbour number was observed in responders compared to non-responders for baseline samples ($P = 0.061$).

Similar features were identified in a study by Antoranz et al.[24]. Using Multiple iterative labelling by antibody neodeposition (MILAN), the authors of this study identified 18 cell phenotypes from a panel of 37 proteins in a cohort of 16 pre-treatment metastatic melanoma samples. Bulk expression measurements, cellular density and cell-cell distance measurements were made for each cell phenotype. The authors observed a significant spatial difference in the expression of PD-L1$^+$ macrophages between responders and non-responders. PD-L1 expression in macrophages close to the tumour edge and close to cytotoxic T-cells was found to

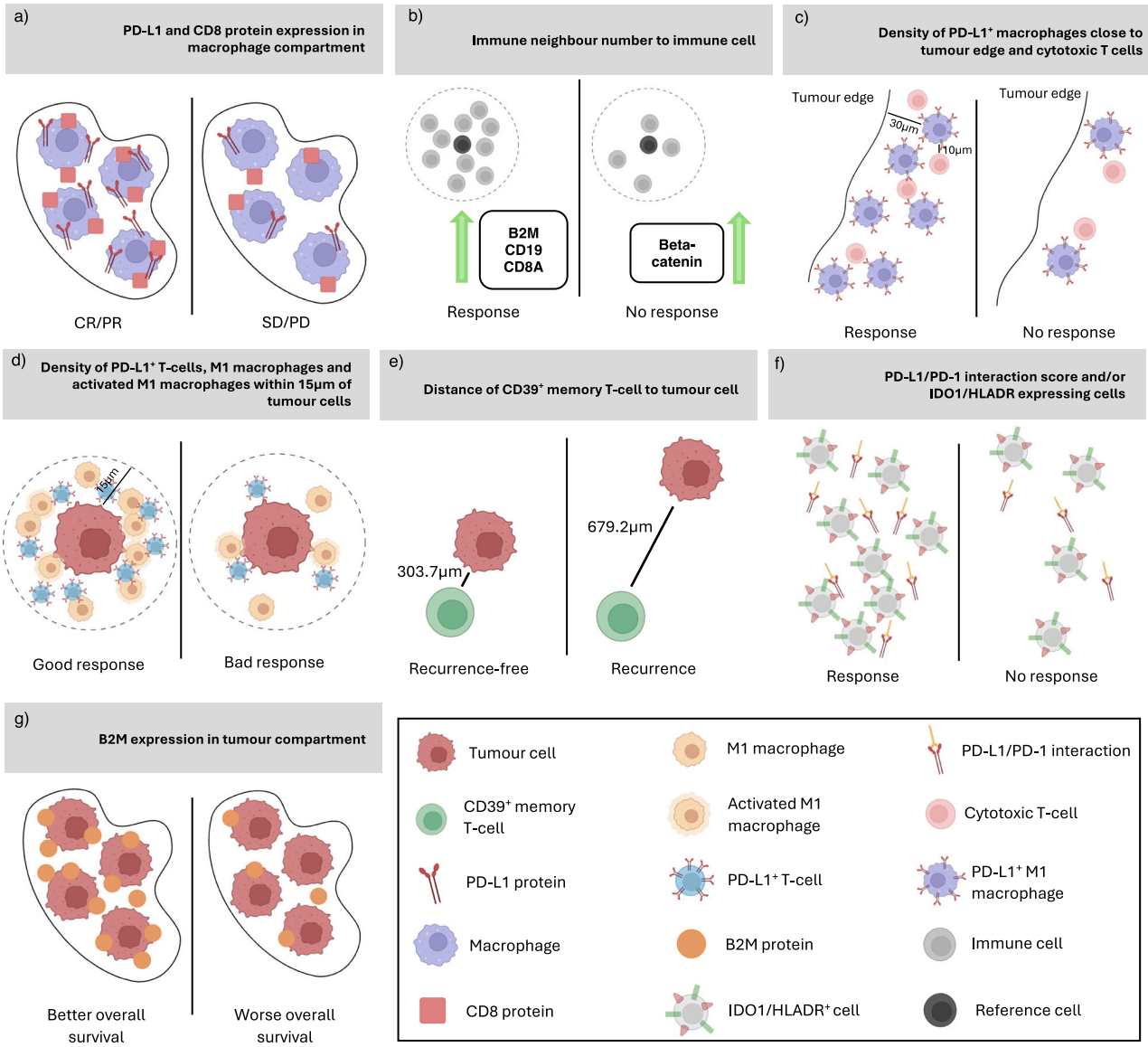

Fig. 3 | Spatial biomarkers for response to immune checkpoint inhibition in melanoma. a Toki et al.[22]: High PD-L1 and CD8 counts in macrophage compartment associated with CR/PR, prolonged PFS and OS. **b** Lu et al.[23]: High protein counts of B2M, CD19 and CD8A, and higher immune neighbour number associated with response. High protein counts of beta-catenin are associated with no response. **c** Antoranz et al.[24]: High density of PD-L1+ macrophages close to tumour edge (within 30 µm) and close to cytotoxic T-cells (within 10 µm) associated with improved response. **d** Kim et al.[25]: High density of PD-L1+ macrophages, M1 and activated M1 macrophages close to tumour cells (within 15 µm) associated with response. **e** Attrill et al.[26]: Shorter distance of CD39+ memory T-cells to tumour cells in recurrence-free patients. **f** Johnson et al.[27]: High PD-L1/PD-1 interaction score and/or high IDO1/HLADR+ co-expressing cells associated with highest likelihood of response. **g** Martinez-Morilla et al.[28]: High B2M protein expression associated with better overall survival.

have a superior predictive value (AUC 0.98) in comparison to bulk PD-L1 expression (AUC 0.68). More specifically and confirmed in a separate validation cohort, a high density of PD-L1$^+$ M1-polarised macrophages close to the tumour was significantly associated with response to ICI (AUC 0.98). Using similar visualisation methods, Kim et al.[25] employed a variety of spatial measures including direct cell-cell distances, nearest neighbour distances between cells near one another and radius-based measurements identifying the percentages of cell phenotypes within 15 mm of a specified cell type on a cohort of 23 stage IV melanoma patients. Through this analysis, they[25] identified that a high spatial density of PD-L1+ T-cells (OR, 15.0; 95% CI, 1.031–218.3; $P = 0.048$), M1 macrophages (OR, 1.226; 95% CI 1.002–1.501; $P = 0.048$) and activated M1 macrophages (OR, 15.0; 95% CI, 1.031–218.3; $P = 0.048$) close to tumour cells (within 15 µm) correlated with a favourable objective response to therapy. Through comparisons of immune cell densities in peri-tumoral and intra-tumoral regions the authors observed that densities of PD-L1+ T-cells tended to be higher in both peritumoral (P value not stated) and intra-tumoral regions ($P = 0.017$) of tumours in good responders while poor responders demonstrated a notably lower density of PD-L1+ T-cells in the peritumoral region compared to intra-tumoral regions.

Similarly, Attrill et al.[26] used multiplex immunofluorescence on a large cohort of 103 pre-treatment FFPE samples from stage III melanoma patients with in-transit or nodal metastasis. The authors focused on assessing intra-tumoral CD8$^+$ T-cell subtypes, namely, CD39$^+$CD103$^+$PD-1$^+$CD8$^+$ (resident memory (Trm)) T-cells, CD39$^-$CD103$^+$PD-1$^+$CD8$^+$ T-cells and CD39$^-$CD103$^-$PD-1$^-$CD8$^+$ T-cells, in association with therapy response. Spatial measures, including the average distance of each CD8$^+$ T-cell subtype to tumour cells and the percentage of each CD8$^+$ T-cell subtype within a 20 µm radius of tumour cells, were made. In the discovery cohort three CD8$^+$ T-cell phenotypes were found to be associated with

https://doi.org/10.1038/s41698-024-00671-1

Table 1 | Summary spatial biomarkers from derived from spatial omics methodologies

Ref	Method	Type of analyte	N targets	Cancer type	N samples	Spatial measure	Endpoint for response	Benchmark assessed?	Benchmark predictive result	Benchmark predictive of response?	Key feature of spatial biomarker
Toki et al.[22]	DSP	P	44	Melanoma	60	Compartment	RECIST criteria, PFS, OS	Yes	PDL1 counts in tumour did not associate with response.	Not in tumour compartment	**High PDL1 counts macrophage compartment. High CD8 counts in macrophage compartment.** High PD-L1 counts in macrophage compartment = CR/PR ($P = 0.0011$), prolonged PFS ($P = 0.0072$; HR, 0.36; 95% CI, 0.18–0.69) and OS ($P = 0.0032$; HR, 0.15; 95% CI, 0.065–0.35). High CD8 counts in macrophage compartm ent = CR/PR ($P = 0.014$), prolonged OS ($P = 0.0119$; HR, 0.33; 95% CI, 0.14–0.78) and PFS ($P = 0.0082$; HR, 0.42; 95% CI, 0.22–0.83).
Lu et al.[23]	DSP	P	29	Melanoma	23	Compartment	RECIST criteria	No	–	–	**High B2M, CD19, CD8A protein counts = R, High Beta-catenin = NR** Baseline samples: Higher immune neighbour number = R ($P = 0.061$) High B2M, CD19 and CD8A protein counts = R, High Beta-Catenin = NR (AUC = 0.998).
	mIF	P	2			Radial quantitation immune number in given distance of immune cell					
Antoranz et al.[24]	MILAN	P	37	Melanoma	16	Cell phenotype distance to edge	RECIST criteria	Yes	Bulk PDL1 not beneficial AUC 0.68	No	**High density PDL1+ macrophages close to tumour edge and cytotoxic T-cells = R (AUC = 0.998)**
						Radial quantitation of macrophage density within 30 µm tumour edge, 10 µm of cytotoxic T-cells					
Kim et al.[25]	mIF	P	7 and 6	Melanoma	23	Cell–cell distances	RECIST criteria	No	–	–	**High density PDL1 + T-cells, M1 macrophages, activated M1 macrophages within 15um of tumour cell = good response.** High density of PDL1+ Tcells (OR, 15.0; 95% CI, 1.031–218.3; *P*-value = 0.048), M1 macrophages (OR, 1.226; 95% CI 1.002–1.501; *P*-value = 0.048) and activated M1 macrophages (OR, 15.0; 95% CI, 1.031–218.3; *P*-value = 0.048) within 15 µm of tumour cells = good response.
						Nearest neighbour distance					
Attrill et al.[26]	mIF	P	6	Melanoma	103	Cell phenotype to tumour cell distance	Recurrence free survival	Yes	PDL1 status was not associated with recurrence irrespective of cut off. PD-L1 status >1%. Positive (HR 0.81, 95% CI 0.41–1.63, $P = 0.5560$)	No	**Tumour cells closer to CD39+ memory T-cells (Trm) in recurrence free patients.** Melanoma cells were significantly closer to CD39+ Trm/P1 in recurrence free (RF) compared to recurrence (Rec) patients: CD39+ Trm/P1 Rec mean = 679.2 µm, CD39+ Trm/P1 RF mean = 303.7 µm, $P = 0.0011$. Predictive model for RFS: High proportion of CD39+ memory Tcells (median = 8.226%, $P = 0.0022$) in addition to complete lymph node dissection and stage = longer RFS. (AUC 75.9%, 95% CI 65.3–86.5%)
						Radial quantiation of cell types within 20 µm of tumour cell					

https://doi.org/10.1038/s41698-024-00671-1

Table 1 (continued) | Summary spatial biomarkers from derived from spatial omics methodologies

Ref	Method	Type of analyte	N targets	Cancer type	N samples	Spatial measure	Endpoint for response	Benchmark assessed?	Benchmark predictive result	Benchmark predictive of response?	Key feature of spatial biomarker
Johnson et al.[27]	mIF	P	3,3, 3	Melanoma	166	Interaction score Co-expression	RECIST criteria/ irRECIST if prior not available	Yes	PDL1 expression alone did not discriminate between R and NR ($P = 0.67$, sensitivity = 0.58, specificity = 0.56, AUC = 0.57)	No	**High PDL1/PD1 interaction score and/or high IDO1/HLADR expressing cells = highest likelihood of response.** PD1/PDL1 interaction score sig. differentiated between R and NR = sensitivity 0.69 and specificity 0.82. Responders who had high PD1/PDL1 interaction score and high percentage of IDO1/HLADR+ cells demonstrated the highest likelihood of response (80%, $P \leqq 0.001$, sensitivity = 0.63, specificity = 0.81, AUC = 0.72)
Martinez-Morilla et al.[28]	IMC	P	21	Melanoma	60	Compartment	RECIST criteria	No	–	–	**High B2M expression = better response.** High levels of tumour B2M protein expression were associated with better response to ICI in discovery cohort: OS: HR (95% CI), 0.20 (0.05–0.74); $P = 0.015$). Trend in validation cohort OS: HR (95% CI), 0.60 (0.35–1.02); $P = 0.058$).
Ghiringhelli et al.[29]	mIF	P	2	NSCLC	265	Cell–cell distances Radial quantitation of cell types within 20 µm of tumour cell	Progression-free survival and overall survival	Yes (TPS)	Not significant in either training or validation set for OS: training set OS PDL1 1% cut-off: (HR = 0.102, CI = 0.63–1.65, $P = 0.93$), validation set OS PDL1 1% cut-off: (HR = 0.71, CI = 0.47–1.07, $P = 0.1$), training set OS PDL1 50% cut-off: (HR = 0.85, CI = 0.53–1.37, $P = 0.51$), validation set OS PDL1 50% cut-off: (HR = 0.71, CI = 0.47–1.07, $P = 0.1$). Significant only for PFS in validation cohort with PDL1 1% TPS cut-off: HR = 0.65, CI = 0.44–0.95, $P = 0.026$.	Yes, only PFS in validation cohort with 1% TPS cut-off	**High Immunoscore-IC score (CD8+ free of PD-L1+ cells, CD8 clusters, CD8 density and proximity of CD8 and PD-L1+ cells) = R.** Two-category (high-low immunoscore-IC)— Training set: Low score = high risk of recurrence. 24 months PFS: (unadjusted [HR] = 0.39, 95% CI (0.26–0.59), $P < 0.0001$). OS: (unadjusted [HR] = 0.42, 95% CI, $P < 0.0001$). Validation set: PFS: (unadjusted [HR] = 0.56, 95% CI (0.37–0.84), $P < 0.0054$). OS = (unadjusted [HR] = 0.43, 95% CI (0.28–0.66), $P < 0.0001$).
Qin et al.[30]	mIF	P	6	NSCLC	52	Nearest neighbour distance Radial quantiation of cell types within 15 µm of each T-cell and 40 µm of each tumour cell	RECIST criteria	Yes (TPS)	NR patients had significantly lower PD-L1 TPS than R (5.6% vs 20.6%; $P = 0.0163$). NR more likely to be PD-L1 negative (54%) than R (24%). 4/14 patients with no PD-L1 expression responded to ICI therapy.	Yes	**Higher % of helper T-cells (HTL) enaged with cytotoxic T-cells (CTL) and more engagment of cytotoxic T-cells with tumour cells in responders.** Lower % of HTLs engaged with CTL in NR v R (42.5% and 62.7%, respectively, $P = 0.0006$). Less engagement of CTLs with EC in NR vs R (33.8% vs 54.1%, $P = 0.0026$) NR had significantly lower spatial immune score compared to R. All patients scoring 0 did not respond to therapy. Universal responders had high CTL-EC engagement and high PD-L1 expression ($P = 0.0027$).

Table 1 (continued) | Summary spatial biomarkers from derived from spatial omics methodologies

Ref	Method	Type of analyte	N targets	Cancer type	N samples	Spatial measure	Endpoint for response	Benchmark assessed?	Benchmark predictive result	Benchmark predictive of response?	Key feature of spatial biomarker
Song et al.[31]	DSP	RNA	1812	NSCLC	18	Compartment	RECIST criteria	Yes, PDL1 and TMB	PDL1 expression (AUC = 0.778). TMB (AUC = 0.533)	Yes, PDL1 and TMB	**High stromal signature of 18 proteins including PDL1 = R.** High stromal signature of 18 proteins including PD-L1 (AUC = 0.838) = R. Patients with high stroma score has longer median OS ($P = 0.039$). Stromal immune composition: CD4 T-cells ($P = 0.014$), DCs ($P = 0.021$), macrophages ($P = 0.031$) and monocytes ($P = 0.029$) significantly higher in PR groups.
Moutafi et al.[32]	DSP	P	71	NSCLC	56	Compartment	RECIST criteria	Yes	PDL1 expression in tumour (HR = 0.67, $P = 0.017$) and immune compartment (HR = 0.52, $P = 0.008$) assoc. with longer PFS.	Yes, tumour and leucocyte compartment	**High CD44 expression in tumour compartment = longer PFS. Remained predictive when adjusting for PDL1 expression.** High CD44 expression in tumour compartment associated with longer PFS in UV: (HR = 0.76, 95% CI 0.61–0.96; $P = 0.024$) and MV: (HR = 0.68, 95% CI 0.46–99; $P = 0.043$). This remained predictive following correction for PDL1 TPS: PD-L1 ≥ 1% TPS cutpoint HR = 0.25, $P = 0.003$ in the discovery set; HR = 0.59, $P = 0.015$ in the validation set), or the PD-L1 ≥ 50% TPS cutpoint
	mIF	P	3								
Moutafi et al.[33]	DSP	P	71	NSCLC	56	Compartment	Progression-free survival and overall survival	Yes	High PDL1 expression in tumour cells ([OS; HR, 0.75; 95% confidence intervals (CI) 0.54–1.05; $P = 0.092$, PFS; HR, 0.67; 95% CI, 0.48–0.94; $P = 0.017$) and leucocytes ((CD45+/CD68−) [OS; HR, 0.41; 95% CI 0.24–0.69; $P < 0.001$, PFS; HR 0.52; 95% CI 0.32–0.84; $P = 0.008$) predictive of longer survival.	Yes, tumour and leucocyte compartment	**High CD66b expression in immune stroma compartment = shorter OS. Validated by mIF.** DSP P: High CD66b expression in immune stroma compartment associated with shorter OS (HR, 1.31; 95% CI 1.06–1.60; $P = 0.016$) and PFS (HR, 1.24; 95% CI 1.02–1.51; $P = 0.04$).
	mIF	P	3,3								
Zugazagoitia et al.[34]	DSP	P	44	NSCLC	53	Compartment	RECIST criteria	Yes	High PDL1 expression in leucocyte or macrophage compartment was not sig. in MV testing: (CD45 compartment—HR: 0.45, 95% CI: 0.15–1.23, $P = 0.11$, CD68 compartment – HR: 0.55, 95% CI: 0.22–1.39, $P = 0.17$)	No	**High CD56 in CD45 compartment = durable clinical benefit. High VISTA expression in tumour compartment = non-durable clincial benefit.** DSP: High levels of CD56 (OR: 6.7, 95% CI: 1.46–30.7, $P = 0.014$) and CD4 (OR: 8.55, 95% CI: 1.54–47.4, $P = 0.014$) in the CD45 compartment associated with durable CB. High levels of VISTA (OR: 0.09, 95% CI: 0.01–0.8, p = 0.031) and CD127 (OR: 0.08, 95% CI: 0.01–0.71, $P = 0.024$) in tumour compartment sig. predicted NCB. mIF: Sig. higher % of CD56+/CKPAN- cells in stroma in CB vs NCB (20.5% vs 9.7%, $P = 0.027$).
	mIF	P	3								

Table 1 (continued) | Summary spatial biomarkers from derived from spatial omics methodologies

Ref	Method	Type of analyte	N targets	Cancer type	N samples	Spatial measure	Endpoint for response	Benchmark assessed?	Benchmark predictive result	Benchmark predictive of response?	Key feature of spatial biomarker
Bortolomeazzi et al.[35]	IMC	P	42	CRC	29	Cell–cell distances	RECIST criteria	Yes	Durable benefit CRC has significantly lower TMB than non durable benefit CRCs $P = 6 \times 10^{-3}$	No	**$CD8^+GZMB^+PD\text{-}1^+$ and $CD8^+Ki67^+PD\text{-}1^+$ T cells were frequently closer to $CD68^+CD74^+PD\text{-}L1^+$ cells than to other cell types in those with DB-CRC**
Phillips et al.[36]	CODEX	P	55	CTCL	14	10 nearest neighbour distances	RECIST criteria	No	–	–	**High SpatialScore = NR, Low SpatialScore = R.** The SpatialScore was shown to be significnalty associated with response in pre-treatment specimens (p = 0.0182). A high SpatialScore represents increased T cell suppressive activity and is associated with non-response while a low SpatialScore represents increased T cell effector activity and is associated with response.
	mIF	P	7			Clustering of 10 nnd = cellular neighbourhoods					
Jia et al.[37]	mIF	P	15	Gastric cancer	60	Radial quantitation within 20 µm of tumour cell—Effective score: number of tumour-immune cell pairs/number of tumour cells. Effective percentage: number of paired cells/all tumour cells	irRECIST criteria	Yes	Patients with higher density ($P = 0.005$) and proportion ($P = 0.007$) of PDL1+ cells showed longer irPFS	Yes	**High risk score = shorter irPFS and irOS.** Risk score = $PDL1^+$ and $CD8^+PD\text{-}1^-$ proportion, Density ratios of the following: $CD8^+LAG\text{-}3^+PD\text{-}1^-/CD8^+$ cells, $PDL1^+/CD4^+FoxP3^+CTLA\text{-}4^+PDL1^+$ cells, $CD8^+LAG\text{-}3^+PD\text{-}1^-TIM\text{-}3^-/CD8^+LAG\text{-}3^+TIM\text{-}3^-$ cells. Higher risk score = shorter irPFS ([HR]: 3.19; $P < 0.001$; median irPFS: 4.87 vs 19.87 months) & shorter irOS (HR: 3.10; $P = 0.001$; median irPFS: 10.03 vs 24.87 months, respectively) in comparison to lower risk score.
Ma et al.[38]	mIF	P	4	Oesophageal cancer	36	Compartment	Progression-free survival and overall survival	Yes, TMB	Not assessed for response prediction	–	**Shorter distances from PD-L1+ tumour cells to PD-L1- dendritic cells and to PD-L1- macrophages = worse PFS and OS.** Shorter distances from PDL1+ tumours cells to PDL1− DCS and to PDL1− macrophages were associated with worse PFS ($P = 0.034$) and OS ($P = 0.003$). Higher TMB correlated with shorter distance between DCs and PDL1− tumour cells ($P = 0.029$) and DCs and PDL1+ tumour cells ($P = 0.049$).
Färkkilä et al.[39]	CycIF	P	36	Ovarian cancer	62	10 nearest neighbour distances	RECIST criteria	Yes, TMB	–	No	**Extreme responder 1**: High proportion of CD8+T− cells, PD-L1+ macrophages and dendritic cells. **Extreme responder 2:** Increased proximity of CD8+ T-cells to PD-L1− tumour cells, exhausted CD8+ T-cells closer to PD-L1+ tumour cells.
						Clustering of 10 nnd = cellular neighbourhoods					

Table 1 (continued) | Summary spatial biomarkers from derived from spatial omics methodologies

Ref	Method	Type of analyte	N targets	Cancer type	N samples	Spatial measure	Endpoint for response	Benchmark assessed?	Benchmark predictive result	Benchmark predictive of response?	Key feature of spatial biomarker
Hammerl et al.[40]	mIF	P	7,7	TNBC	53	Mean cell type densities at centre and tumour border	RECIST criteria	Yes	AUC of 0.66 (CI: 0.51–0.82)	Yes, immune cells only	**Enrichment of excluded and ignored phenotypes in NR, enrichment of inflamed phenotype. Spatial classifier derived from genesets corresponding to ICH phenotypes. Excluded geneset higher in NR, inflamed geneset higher in R.** Expression of the excluded gene-set was significantly higher in nonresponding (progressive disease (PD) patients) patients (odds-ratio (OR): 3.5; CI: 1.2–11.9), whereas expression of the inflamed gene-set was significantly higher. predictive value of the excluded, inflamed, or combined gene-sets, we observed areas under the curve (AUC) of 0.72 (CI: 0.52–0.89), 0.73 (CI: 0.54–0.94), and 0.75 (CI: 0.55–0.95), respectively. in responding patients (complete response (CR) +partial response. (PR) + stable disease (SD) for >24 weeks according to iRECIST criteria33) (OR: 0.4; CI: 0.18–0.92).
Wang et al.[41]	IMC	P	43	TNBC	243	Cell–cell interactions through whole cell mask contact	RECIST criteria	No	–	–	**High fraction of MHCII+ epithelial cells interacting with epithelial cells and high fraction of proliferating CD8+ TCF1+ T-cells interacting with TME cells = better response.** Identified cells in activation states. Fraction of MHCII+ epithelial cells interacting with epithelial cells ($P_{interaction} = 0.004$, FDR = 0.004, OR: 1.36, CI: 1.19–1.56) and fraction of proliferating CD8+ TCF1+ T-cells interacting with TME cells ($P_{interaction} = 8 \times 10^{-5}$, FDR = 0.003, OR: 1.72, CI: 1.37–2.17) were strongest predictors of response.

Abbreviations: *P* protein, *PFS* progression-free survival, *OS* overall survival, *RFS* regression-free survival, *CR* complete response, *PR* partial response, *SD* stable disease, *PD* progressive disease, *HR* hazard ratio, *CI* 95% confidence interval, *OR* objective response, *R* responder, *NR* non-responder, *Rec* recurrence, *RF* recurrence-free, *NSCLC* non-small cell lung cancer, *CRC* colorectal cancer, *CTCL* cutaneous T-cell lymphoma, *TNBC* triple-negative breast cancer, *DSP* digital spatial profiler, *mIF* multiplexed immunofluorescence, *MILAN* multiple iterative labelling by antibody neodisposition, *IMC* imaging mass cytometry, *CycIF* cyclic immunofluorescence.

recurrence: (1) in recurrence-free patients a high proportion of $CD39^+$ Trm ($CD39^+CD103^+PD\text{-}1^+CD8^+$) cells and a high proportion of CD39−CD103 + PD-1-CD8+ cells were identified and 2. in patients with recurrence a high proportion of $CD39^-CD103^-PD\text{-}1^-CD8+$ (bystander) cells was apparent. Incorporating spatial measures, in the validation cohort, a significantly higher percentage of tumour cells closer to CD39+ Trm cells was identified in recurrence-free patients (CD39+ Trm/P1 Recurrence: mean = 679.2 µm, CD39+ Trm/P1 Recurrence-free: mean = 303.7 µm, $P = 0.0011$). Bulk PD-L1 expression was not significantly associated with recurrence, irrespective of the cut-off used (PD-L1 status >1%. Positive (HR 0.81, 95% CI 0.41–1.63, $P = 0.5560$).

Conferring a slightly different perspective Johnson et al.[27] examined PD-1/PD-L1 interaction and IDO1/HLA-DR co-expression on a large multi-institution cohort of 166 stage IV melanoma specimens. To infer PD-1/PD-L1 interaction, the authors employed a pixel-based mask approach through the identification of $PD\text{-}L1^+$ pixels and dilation of these regions to infer interaction with neighbouring $PD\text{-}1^+$ pixels. A separate mask of pixels with IDO1/HLA-DR co-expression was also defined. This analysis demonstrated that those with the highest likelihood of response (80%, $P \leqq 0.001$, sensitivity = 0.63, specificity = 0.81, AUC = 0.72) had a significantly high PD-1/PD-L1 interaction score and/or high IDO1/HLA-DR co-expression in comparison to non-responders. Bulk PD-L1 or PD-1 expression alone, did not discriminate between response groups ($P = 0.67$, sensitivity = 0.58, specificity = 0.56, AUC = 0.57).

A final study from Martinez-Morilla et al.[28] demonstrated the application of compartment-based analysis using imaging mass cytometry (IMC) to screen for potential targets for further investigation. In this study, a 26-target IMC panel was applied to 60 specimens of stage III/IV metastatic melanoma patients treated with ICI. Assessment of target expression in 6 compartments (all cells, tumour cells (HMB45/S100), stroma (tumour cells subtracted from all cells), T-cells (CD3), B-cells (CD20) and macrophage (CD68)) identified a subset of markers for further validation. Multivariate analysis highlighted B2M as a potential target associated with ICI response (discovery cohort: HR (95% CI), 0.20 (0.05–0.74); $P = 0.015$, validation cohort: HR (95% CI), 0.60 (0.35–1.02); $P = 0.058$) however its specific clinical utility was inconclusive.

Non-small cell lung cancer

We identified six studies examining spatial biomarkers of response to ICI in NSCLC (2/6 on metastatic NSCLC) (Fig. 4, Table. 1). Two studies focused primarily on single cell analysis using either multiplex-immunofluorescence or IHC for response prediction in metastatic disease and both studies identified comparable features as predictors of response. Ghiringhelli et al.[29] examined spatial measures between PD-L1+ cells and CD8+ cells using multiplex IHC in a large cohort of 265 metastatic NSCLC patients treated with ICI. An Immunoscore-IC score was derived from the measurement of the proximity between PD-L1+ and CD8+ cells and quantification of the fraction of cell types within a 20 mm radius of a specified cell type. Patients with a low score (two-category Immunoscore-IC) demonstrated a higher risk of recurrence (endpoints: PFS and OS) in both training (24 months PFS: (unadjusted [HR] = 0.39, 95% CI (0.26–0.59), $P < 0.0001$). OS: (unadjusted [HR] = 0.42, 95% CI, $P < 0.0001$) and validation data sets (PFS: (unadjusted [HR] = 0.56, 95% CI (0.37–0.84), $P < 0.0054$). OS = (unadjusted [HR] = 0.43, 95% CI (0.28–0.66), $P < 0.0001$). A high Immunoscore-IC was associated with responders.

In a similar analysis using a low number of targets, Qin et al.[30] applied a 6-plex immunofluorescence assay to a cohort of 52 metastatic NSCLC patients to examine the cellular distribution and interaction of cells in the tumour microenvironment. The authors performed a variety of spatial measures between 6 major cell types (cytotoxic T-cells, helper T-cells, regulatory T-cells, epithelial tumour cell and "other" cells) including nearest neighbour and cell-cell engagement through the measurement of cell type frequency within a radius of between 15 and 40 µm from a specified cell of interest. Through this analysis, a significantly lower degree of engagement of cytotoxic T-cells with tumour cells (33.8% vs 54.1%, $P = 0.0026$) and/or helper T-cells (42.5% and 62.7%, respectively, $P = 0.0006$) was identified in non-responders compared to responders. These spatial measures were additive in predictive value to the percentage of PD-L1 tumour expression. Universal responders had high CTL-EC engagement and high PD-L1 expression ($P = 0.0027$).

In contrast, four studies employed compartment-based investigations, examining bulk expression of proteins from DSP (Nanostring) which was coupled with validation of promising targets by multiplex immunofluorescence. The most promising study from this group by Song et al.[31] for the identification of compartment-based spatial biomarkers of response focused on bulk RNA and protein expression in tumour and stromal compartments. To reduce the dimensionality and complexity of targets, the authors developed a signature score for each compartment. The stromal signature score comprising 18 protein targets demonstrated the strongest clinical value and greater predictive power (AUC 0.84) than bulk PD-L1 expression (AUC 0.78) and TMB (AUC 0.53). Interestingly, concordant with features identified in single-cell studies, many of the 11 targets were T-cell-associated. Furthermore, the stromal signature contained PD-L1 as one of the 11 targets.

The last three studies[32–34] using DSP employ the same compartment methodology i.e., assessing bulk protein expression in four compartments (tumour, leucocytes, macrophages, and immune stroma). The authors of the studies use univariate analysis to identify targets for further validation via multiplex immunofluorescence. The three studies identify potential targets associated with response i.e., high CD66b in the immune compartment[33] associated with shorter OS (HR, 1.31; 95% CI 1.06–1.60; $P = 0.016$) and PFS (HR, 1.24; 95% CI 1.02–1.51; $P = 0.04$), high CD44 expression in the tumour compartment[32] associated with predictive clinical benefit in test cohort (OR = 1.22, 95% CI 1.03–1.45; $P = 0.018$) and high CD56 expression in the leucocyte compartment were associated with durable clinical benefit (OR: 6.7, 95% CI: 1.46–30.7, $P = 0.014$) which may warrant further investigation.

Additional cancer types

In addition to the studies described, we identified seven studies across a variety of cancer types (colorectal cancer[35] ($n = 1$), cutaneous T-cell lymphoma[36] ($n = 1$), gastric cancer[37] ($n = 1$), oesophageal cancer[38] ($n = 1$), ovarian cancer[39] ($n = 1$) and triple-negative breast cancer[40,41] ($n = 2$)) (Table. 1). Many of these studies examined the spatial relationship of specific immune cell subtypes in relation to response.

Wang et al.[41] investigated the spatial relationship of cell types in different activation states in a cohort of 243 pre-treatment triple-negative breast cancer samples using IMC. The authors of this study quantified the frequency of cell-cell interactions of specific activated cell types. This was achieved through the generation of cell masks and the identification of neighbouring masks in contact with one another. The fraction of MHCII+ epithelial cells interacting with neighbouring epithelial cells (P interaction = 0.004, FDR = 0.004, OR: 1.36, CI: 1.19–1.56) and the fraction of proliferating CD8+ TCF1+ T-cells interacting with neighbouring tumour microenvironment cells (P interaction = 8.10-5, FDR = 0.003, OR: 1.72, CI: 1.37, 2.17) were strongest predictors of response. Similarly, Jia et al.[37] assessed the spatial location and functional status of multiple cell types in a cohort of 60 gastric cancer patients in relation to ICI response. In this study, the authors developed an effective score (number of tumour-immune cell pairs/number of tumour cells) and an effective percentage (number of pair cells/all tumour cells) within 20 µm of a tumour cell to reflect the cell-cell interactions occurring in the cohort samples. Through this, a risk score was generated combining multiple quantitative measures: PD-L1+ and CD8 + PD-1− proportion, density ratios of the following: CD8+ LAG-3− PD-1−/CD8+ cells, PD-L1+/CD4+ FoxP3+ CTLA-4+ PD-L1+ cells, CD8+ LAG-3+ PD-1− TIM-3−/CD8+ LAG-3+ TIM-3− cells. A higher risk score was associated with shorter immunotherapy-related (ir)PFS (HR: 3.19; $P < 0.001$; median irPFS: 4.87 vs 19.87 months and irOS (HR: 3.10; $P = 0.001$; median irPFS: 10.03 vs 24.87 months, respectively). Continuing the focus on the role of CD8+ T-cells for ICI response, Bortolomeazzi et al.[35]

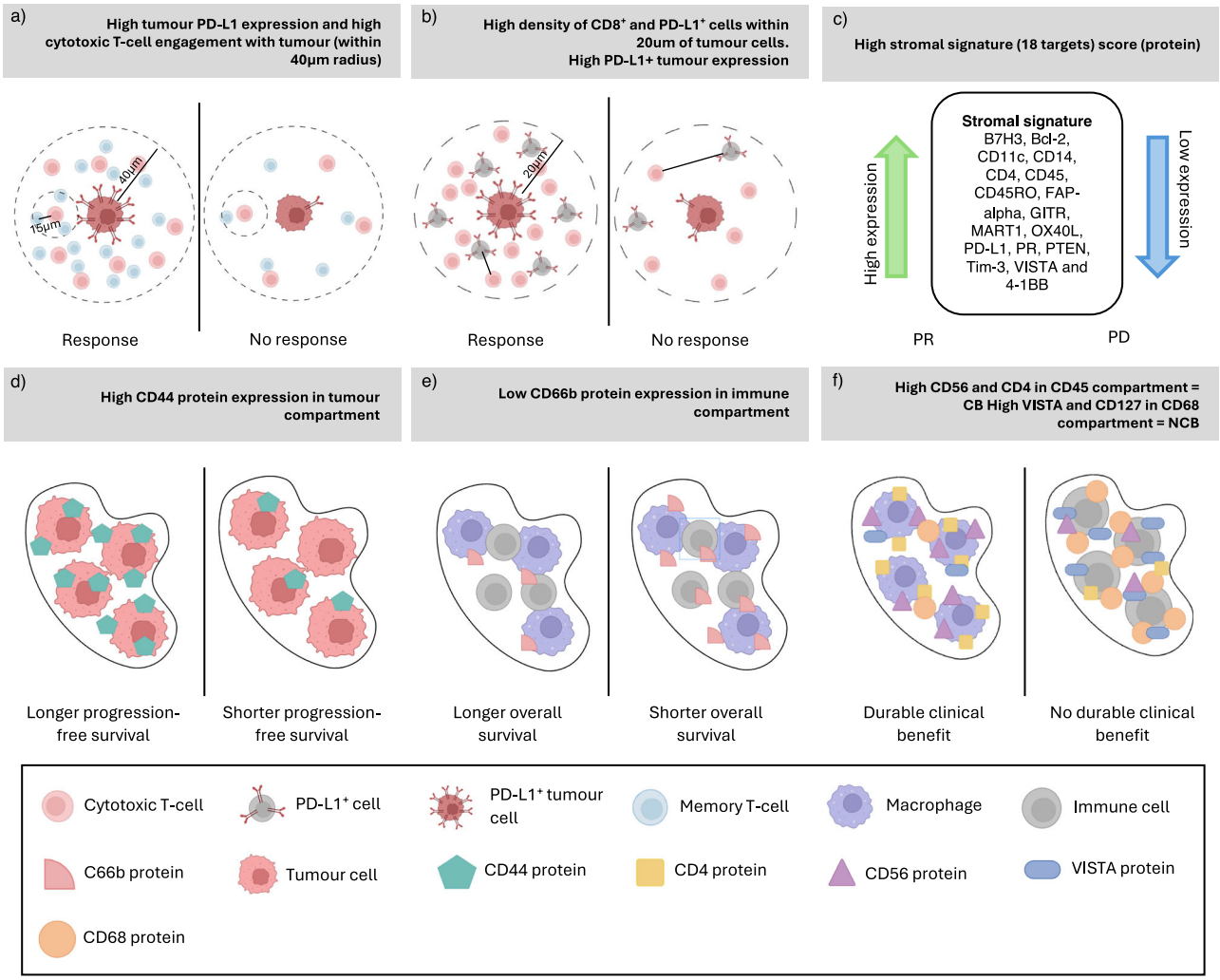

Fig. 4 | Spatial biomarkers for response to immune checkpoint inhibition in non-small cell lung adenocarcinoma. a Ghiringhelli et al.[29]: High Immunoscore-IC score (CD8+ free of PD-L1+ cells, CD8 clusters, CD8 density and proximity of CD8 and PD-L1+ cells) associated with response. **b** Qin et al.[30]: A higher percentage of helper T-cells engaged with cytotoxic T-cells (within 15 μm) and more engagement of cytotoxic T-cells with tumour cells (within 40 μm) in responders. **c** Song et al.[31]: High stromal signature comprised of 18 proteins including PD-L1 associated with response, longer median OS. **d** Moutafi et al.[32]: High CD44 expression in tumour compartment associated with longer PFS. **e** Moutafi et al.[33]: Low CD66b expression in immune stroma compartment associated with longer OS. **f** Zugazogoitia et al.[34]: High CD56 and CD4 in CD45 compartment associated with durable clinical benefit. High VISTA and CD127 in the CD68 compartment are associated with no durable clinical benefit.

examined the spatial relationship of cells of the TME and their predictive utility in comparison to TMB. In a cohort of 29 colorectal cancer patients, the immune landscape of hypermutated durable-benefit tumours was first characterised through RNA sequencing. These tumours demonstrated significant numbers of immune cells (immune hot) and specifically high levels of CD74+ macrophages and proliferating T-cells. To examine the spatial interactions of these cell types, the authors utilised IMC and measured distances between CD8+ GZMB+ PD-1+, CD8+ Ki67+ PD-1+ T cells and CD68+ CD74+ PD-L1+ cells and other cell types. Through this, it was demonstrated that CD8+ GZMB+ PD-1+ and CD8+ Ki67+ PD-1+ T cells were more frequently closer to CD68+ CD74+ PD-L1+ cells than other cell types in patients with durable-benefit CRC.

Phillips et al.[36] examined the topography of TME components in a cohort of 14 advanced cutaneous T-cell lymphoma patients using CODEX. Initial classification of cellular densities and their association with treatment response yielded negative results. Using a cellular neighbourhood (CN) method to identify coordinated multi-cellular communities, specific CNs were associated with response status, namely, tumour and CD4+ cell enriched CNs being associated with response, and conversely, Treg enriched CNs associated with non-response. To enable clinical translation of these observations the authors developed a SpatialScore derived from enrichment cell types of specific CNs associated with responder vs non-responder groups. The SpatialScore is quantified as a ratio of the average distance of CD4+ T cells to the nearest tumour cells and the average distance of Treg cells to the nearest tumour cells, representing a ratio of immune activity. The SpatialScore was shown to be significantly associated with response in pre-treatment specimens ($P = 0.0182$). A high SpatialScore represents increased T cell suppressive activity and is associated with non-response while a low SpatialScore represents increased T cell effector activity and is associated with response. To further simplify the potential implementation of the SpatialScore into a routine pathology department where access to technology such as CODEX would be a limiting factor, an 8-plex immunofluorescence panel readily applicable to imaging platforms already widely deployed in many clinical settings was developed. Validation of the predictive ability of the SpatialScore using this method (which requires a low number of targets) demonstrated a fivefold lower SpatialScore in responders (mean score 0.31) compared to non-responders (mean score 1.52) suggesting the assay had a higher sensitivity compared with initial methods. This study highlights the potential translation of spatial biomarkers from discovery to a more suited technology for clinical translation.

Artificial intelligence for image-based prediction of ICI response

In recent years, AI modalities have revolutionised the field of computational pathology, enabling the development of algorithms to perform a variety of functions i.e., automatic detection of histological features, prediction of patient outcome and novel biomarker discovery. Deep learning (DL) is a frequently implemented AI modality for computational pathology and requires several steps for its implementation (Fig. 5A). As a starting point, images are input for feature detection, most commonly from H&E-stained slides. Because of their large size, whole slide images (WSIs) cannot be processed at once and need to be divided into smaller image regions, commonly named tiles. These tiles are used to train DL models either in a supervised manner, where all tiles have a corresponding ground truth label, or semi-supervised where the ground truth is defined at the WSI level. After training, the model is fixed and can be applied to new unseen images. During inference, the same process is applied: the WSI is cut into tiles of predefined size and fed into the model. Tile-based predictions are finally aggregated to reconstruct the WSI-level prediction.

In the context of ICI response prediction, multiple groups have been using DL models to measure already known biomarkers such as PD-L1 positivity from pathology images and correlate it with patient response (Table. 2). Ligero et al.[42] used a semi-supervised approach to predict PD-L1 status from 233 IHC WSIs from an NSCLC cohort. They compared the performance of TPS and CPS. Reaching an AUC of 0.80 on a pan-cancer test cohort of 108 WSIs, they observed an improved association between PD-L1 predictions and response to ICI, compared to TPS and CPS. Also, patients classified as high PD-L1 had significantly longer median PFS compared to patients classified as low PD-L1. Along the same line, Shamai et al.[43] used breast cancer H&E TMA cores from 3,376 patients to predict PD-L1 status (Fig. 5B). Each H&E TMA was labelled as PD-L1 positive or negative by a pathologist, using the corresponding IHC-stained TMA as ground truth. A convolutional neural network (CNN) with residual connections was trained using resized H&E TMAs to predict PD-L1 positivity, reaching an AUC of 0.91–0.93. The inter-observer variability of pathologists assessing PD-L1 positivity showed a kappa agreement value of 0.77, meaning that independent pathologists might interpret PD-L1 differently and highlight the benefit of support systems to improve diagnosis. Interestingly, the authors also investigated morphological features associated with high and low prediction scores of PD-L1 status. Using image features, they mapped patient images to points in space (t-SNE). They found that a low prediction score was associated with dense desmoplastic stroma surrounding tumour ducts, various lumen sizes, low tumour-stroma ratio, and low immune cells. In contrast, high prediction scores correlated with crowded solid tumour nests, hyperchromatic nuclei, absence of lumens, small stromal area and presence (sometimes in abundance) of tumour-associated immune cells. Using a different approach, Park et al.[44] highlights the ability to derive spatial biomarkers with clinical utility from haematoxylin and eosin (H&E) stained slides (Fig. 5B). The authors established an AI model with Lunit Scope to segment tumour epithelium, tumour stroma and detect Tumour Infiltrating Lymphocytes (TILs) using a cohort of 3116 WSI from 25 cancer types. The results were validated on two NSCLC cohorts totalling 518 patients. Based on the amount of TILs in each segmented component (epithelium and stroma) they defined three immune phenotypes (IP): inflamed, immune-excluded and immune desert. They observed that inflamed IP was associated with response to ICI, significantly correlated with PD-L1 TPS $\geqq$ 50%, and had improved OS, TTR and longer PFS.

Instead of evaluating known biomarkers, some groups directly predicted response and patient outcomes from the digital H&E slides. Hu et al.[45] extracted tumour tiles from 190 H&E melanoma digital slides to train a DL model to predict responders vs non-responders. The model was tested on two test cohorts: 54 H&E melanoma slides with an AUC of 0.778 and 55 H&E lung cancer slides with an AUC of 0.645. The results were compared with a second DL model trained to identify TILs, reaching an AUC of 0.58 in the melanoma dataset, emphasising the benefit of using DL on images to leverage unknown features existing in the tissue slides. Shibaki et al.[46] trained a model to predict 1-year PFS through examination of the tumour immune microenvironment (TIME) using three different stainings: H&E, PD-L1 and double CD8 and FoxP3. All three stains were registered and tiles from the same regions were extracted (100,544 tiles from 78 patients). First, they extracted image features for each stained tile (H&E, PD-L+ and CD8/FoxP3) using an ImageNet pretrained EfficientNet. All three features were aggregated and passed through a second CNN for tile-based prediction. To compare image-based classification with patient information, they also trained a model based on clinical data. Using a combination of both image-based and patient-based models, they obtained a third model. Their combined model showed the highest performance with an AUC of 0.868, a patient-based model AUC of 0.789 (0.571–0.982), and an image-based model of 0.782. The PFS was longer in the high efficacy group than in the low efficacy group in all three models (patient information model, HR 0.468; pathological image model, HR 0.334; combined model, HR 0.353, 95% CI 0.195–0.37). The authors also retrieved the features that were most important for the classification decision and found that the importance of the pathological image predictions was higher than that of the patient information, highlighting the importance of the TIME in predicting patient response. The authors also showed that ML methods improved upon human count evaluations—AUC of human count, CD8+ lymphocyte: 0.681, FoxP3+ lymphocytes: 0.626, PD-L1 score: 0.567. Moreover, the PD-L1 CPS was comparable between patients who achieved and those who did not achieve 1-year PFS. Finally, Liu et al.[47] created an Ensemble model (ICIsNet) composed of two CNNs (DenseNet, an EfficientNet) and one Vision Transformer model (SwinVit v2) to predict good responders vs bad responders (Fig. 5B). Their cohort was composed of 313 H&E slides from 264 patients with advanced gastric cancer from 4 different centres. ICIsNet showed great performance on three validation data sets with AUC scores of 0.952, 0.920, and 0.962 for WSI-level predictions. Furthermore, overlaying heatmap predictions over the WSI, they could identify morphologic features associated with lack of response, such as poorer differentiation, diffuse tumour tissue, signet ring cells, cellular mucin pools and reduced lymphocytic infiltration. On the other hand, good responders showed tumour cells resembling normal tissue and increased lymphocytic infiltration.

These studies show the potential of using DL models to predict patient response to ICI from histopathology images such as H&E to high-plex imaging modalities. Half of the studies examined identified morphological features of response. These methods are also applicable to identify further features of response from more complex datasets such as those generated from spatial omics imaging-based platforms.

Discussion

The clinical realisation of ICI has significantly advanced cancer therapeutics. However, current predictive biomarkers for ICI response are sub-optimal. In this review, we sought to evaluate whether the assessment of the spatial distribution and interactions of cellular components in clinical samples could identify superior spatial biomarkers of response to ICI.

Tissue-based single-cell methodologies such as multiplex immunofluorescence dominated the types of spatial platforms utilised. All studies examined protein-based measures of target expression. While a variety of methods were applied for spatially resolved quantitation of cell–cell interactions and cellular distribution across the tissue samples, most commonly, radial quantitation and cell-to-cell distances were implemented. In the instance of studies utilising the DSP platform, spatial biomarkers were limited to compartment-based quantitation e.g., stromal or tumour, tissue compartment.

Many studies applied an initial discovery approach, examining, on average, 20 protein targets. It is unlikely that routine pathology departments would have the methods required to assess such a high number of targets from a single tissue section, proving clinical translation of such spatial biomarkers challenging. However, a subset of studies reduced their findings to a handful of markers in consideration of potential clinical implementation. Studies such as Phillips et al. demonstrate a potential paradigm for the transition from discovery to clinical translation. This dimensionality

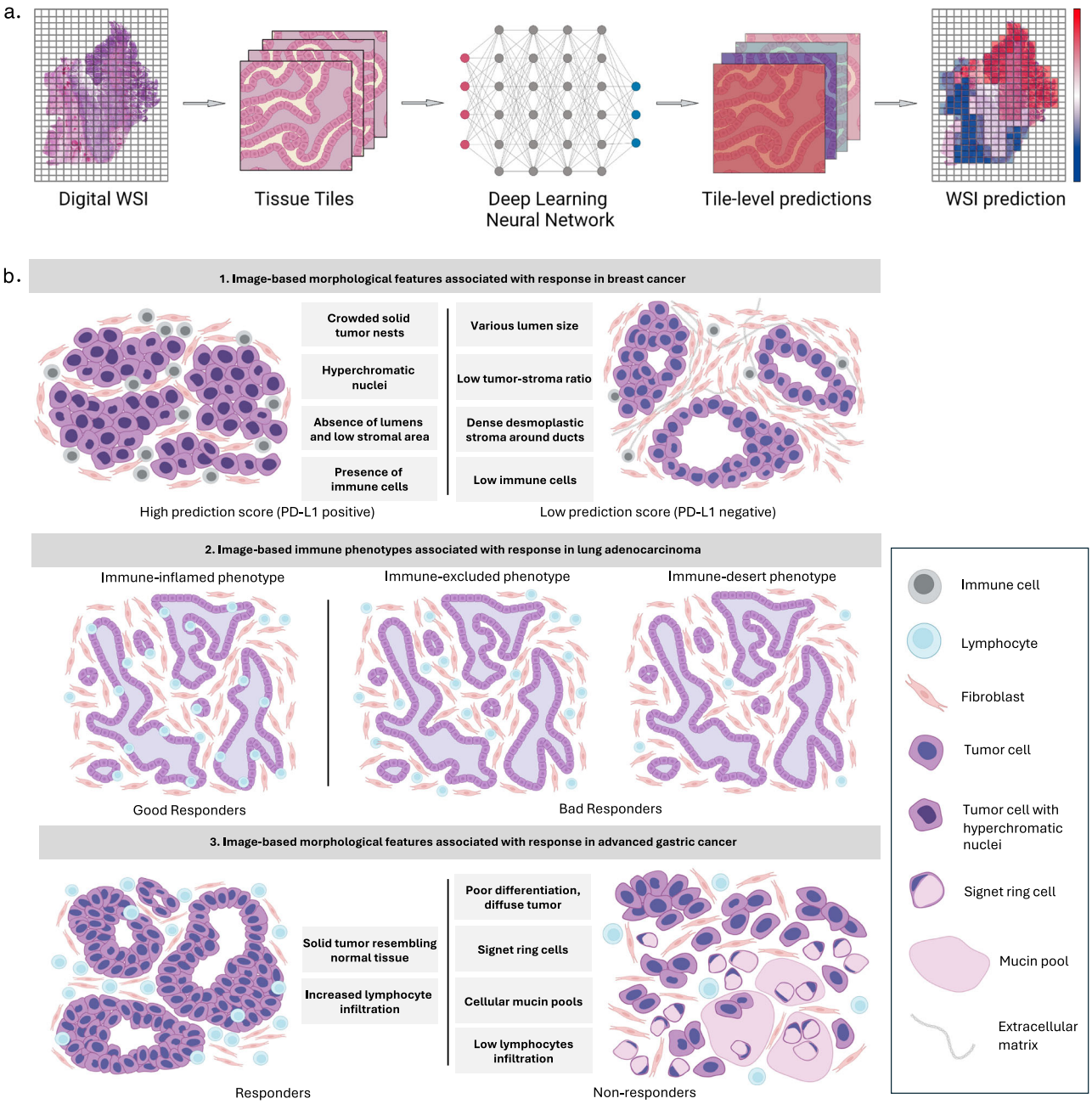

Fig. 5 | Artificial intelligence for image-based prediction of ICI response. **a** Workflow of deep learning predictions applied to histopathology whole slide images (WSIs). Tiles of predefined size are extracted from the digital WSI and fed into the pretrained deep learning model, resulting in tile-level predictions. Tiles are stitched back together to reconstruct the WSI and tile-level predictions are aggregated to generate a WSI-level prediction. Results can be visualised by overlaying a tile-based heatmap over the WSI. **b** Image-based derived morphological features associated with response to immune checkpoint inhibitors (ICI). Three studies used heatmaps to identify morphological features associated with response to ICI. 1. In breast cancer (Shamai et al.[43]), crowded solid tumour nests, hyperchromatic nuclei, absence of lumen, low stromal area and presence of immune cells were found in predicted PD-L1 positive patients, while negative patients were associated with the presence of various lumen size, low tumour to stroma ration, dense desmoplastic stroma around ducts and low immune cells presence. 2. In lung adenocarcinomas (Park et al.[44]), the presence of intra-epithelial lymphocytes (immune-inflamed) was found in responders, while non-responders were found to have various number of lymphocytes in the stroma (immune-excluded or immune-desert). 3. Advanced gastric cancer (Liu et al.[47]) responders showed solid tumours resembling normal tissue with increased lymphocyte infiltration whereas non-responders showed diffuse and poorly differentiated tumours with the presence of signet ring cells, cellular mucin pools and low lymphocyte infiltration.

reduction will be crucial for the clinical translation of spatial biomarkers both in terms of affordability and utilising methodologies such as multiplex immunohistochemistry or even H&E already routinely available in most pathology departments. Such studies demonstrate how 'spatial' does not necessarily equate with 'complex' and highlight the potential ease of integration of spatial measures into clinical trial protocols as a routine occurrence. The only example of successful clinical translational from the included studies is shown in the implementation of the IC-Immunoscore as developed by Ghiringhelli et al.[29] as a CE-marked in vitro diagnostic assay (Veracyte).

The bulk of studies that we identified focused on spatial biomarkers for response in melanoma and NSCLC. This enabled comparisons between the types of spatial biomarkers identified to be made within each disease type. In the context of melanoma, we observed trends in the types of spatial

https://doi.org/10.1038/s41698-024-00671-1

Table 2 | Summary of AI image-based studies for identification of spatial biomarkers

Ref	Image modality	Cohort details	Benchmark	Model prediction	Model Training	Model performance	Results	Derived morphological biomarkers
Ligero et al.[42]	IHC (PD-L1)	233 WSI from the NSCLC-Memorial Sloan Kettering cohort (TPS < 1%: 73; TPS ≥ 1%: 163) External vallidation cohort: Pan-cancer Val d'Hebron Institute of Oncology of 108 WSI (TPS < 1%: 59; TPS ≥ 1%: 49)	TPS and CPS determined by a trained pathologist. Slides with less than 100 tumour cells were excluded	Classify tiles as high PD-L1 (TPS ≥ 1%) vs low PD-L1 (TPS < 1%)	224 × 224px tiles extracted from ROIs. Feature extraction was performed (RetCCL and ResNet50) and used to train (5-fold cross-validaton) weakly supervised MIL model	AUC of 0.80 ± 0.03, Accuracy = 77%, True Positive = 37%, False Negative = 10%, False positive = 13%, True Negative = 40%, Sensitivity = 0.79, Specificity = 0.75, on the test cohort.	Improved association between PD-L1 prediction and response to ICI [HR: 1.5 (95% CI: 1–2.3), $P = 0.049$] compared to CPS [HR: 1.4 (95% CI: 0.96–2.2), $P = 0.082$] and TPS [HR: 1.2 (95% CI: 0.79–1.9), $P = 0.386$] Patients classified as high PD-L1 had significant longer median PFS	Not assessed
Shamai et al.[43]	H&E TMAs	BCCA TMA cores from 3376 breast cancer patients (PD-L1 negative: 2819, PD-L1 positive: 1671) BCCA divided into training: 2516 (74%) and testing 860 (25.5%)) External validation cohort: MA31 cohort 652 patients (PD-L1 negative: 252, PD-L1 positive: 23)	Pathologist PD-L1 scoring. MA31 cohort scored by two pathologists, kappa = 0.772	Classify TMA cores as PD-L1 positive vs negative	TMA cores resized to 512 × 512px images. 5-fold cross-validation CNN with residual connections inspired by ResNet. Ground truth labels were provided by pathologists comparing H&E cores with corresponding IHC cores	AUC 0.911 (95% CI: 0.891–0.925), 0.915 (95% CI: 0.883–0.937) and 0.854 (95% CI: 0.771–0.908) for internal cross-validation, BCCA test set and MA31 respectively. Negative predictive value (NPV) of 0.983 and sensitivity of 0.943 on BCCA test set and NPV of 0.993 and sensitivity of 0.957 for MA31 test set, compared with pathologist annotations	System able to detect patients likely to be PD-L1 negative and could be used as decision support in clinical practice. H&E breast cancer tissue structure contains information indicative of PD-L1 expression. PD-L1 expression correlated with presence of immune cells and more aggressive tumours.	Image-based morphological features associated with response: Non-responders: dense desmoplastic stroma around tumour ducts, various lumen size, low tumour to stroma ratio and low immune cells Responders: crowded solid tumour nests, hyperchromatic nuclei, absence of lumens, small stromal area and presence of immune cells.
Park et al.[44]	H&E	3116 WSI from 25 cancer types, including NSCLC External validation cohort composed of LUAD ($N = 461$), LUSC ($N = 462$) images from TCGA, SMC ($N = 1205$) and SNUBH ($N = 261$) Retrospective cohort of NSCLC patients treated with ICI to validate biomarker: SMC ($N = 299$) and SNUBH ($N = 219$)	TPS No correlation between immune phenotypes (IP) and baseline clinical attributes (smoking status, liver metastasis, lung immune prognostic score index)	Detect TILs, segment cancer epithelium (CE) and cancer stroma (CS)	AI model with Lunit Scope WSI were divided into 1 mm^2 tiles on which three immune phenotypes (IP) were defined: inflamed (TIL in cancer epithelium >106/mm^2), immune-excluded (TIL in cancer epithelium >106/mm^2 and TIL in cancer stroma >357/mm^2) and immune-desert TILs below the threshold in epithelium and stroma compartments.	AUC of 0.9539, 0.9871 and 0.9591 for CE, CS and TIL respectively on the external validation.	Significant higher response rate to ICI in patients with inflamed IP ($p < 0.01$) Significant correlation between inflamed IP and PD-L1 TPS ≥ 50% Significant higher PFS in inflamed IP patients compared to immune-excluded and immune-desert ($P < 0.001$) Significantly improved OS after ICI in patients with inflamed IP (HR 5 1.38, P, 05 for immune-excluded inflamed IP and 1.67, P, 0.05 for immune-desert IP inflamed IP)	Three immune phenotypes (IP) defined: Inflamed, immune-excluded and immune-desert Significant higher response rate to ICI in patients with inflamed IP ($p < 0.01$)

Table 2 (continued) | Summary of AI image-based studies for identification of spatial biomarkers

Ref	Image modality	Cohort details	Benchmark	Model prediction	Model Training	Model performance	Results	Derived morphological biomarkers
Hu et al.[45]	H&E	190 melanoma WSI from TCGA Two test cohorts: 54 melanoma WSI (responders: 15, non-responders: 23) and 55 lung cancer WSI	TILs image-based model performed better than the TIL model, highlighting the benefit of image-based DL to leverage unknown pathological features	classification of responders vs non-responders	256 × 256 tiles, features extracted using pretrained Xception model, PCA was applied to keep 20 most informative components that were used as input to SVM to predict responders vs non-responders. In parallel trained VGG-16 for tile-based TIL or non-TIL classification.	AUC of 0.778 (95% CI: 63.8–90.5%) and 0.645 (95% CI: 49.4–78.4%) for melanoma and lung test sets. TIL model AUC of 0.58 in melanoma test set.	Predicted responders had longer PFS ($P = 0.06$) image-based model performed better than the TIL model	Not assessed
Shibaki et al.[46]	H&E, IHC for PD-L1 and double CD8/FoxP3	78 ES-SCLC patients (8 patients ≥ 1-year PFS, 70 patients < 1-year PFS)	2 pathologists assessed: the number of each type of infiltrating immune cell (up to four 1 mm^2 regions), CPS	Prediction of 1-year PFS	128 × 128px tiles, 5-fold cross-validation, ImageNet pretrained EfficientNet-B0 for feature extraction from all three image modalities. Aggregated features were passed through a second CNN. Additional model based on clinical data. Final model used a combination of image-based and clinical-based models.	AUC of 0.868 [95% CI: 0.818–0.910], 0.789 [95% CI: 0.671–0.922] and 0.782 [95% CI: 0.735–0.844] for combined model, clinical model and image-based model respectively.	PD-L1 CPS and densities of CD8+ TILs and FoxP3+ TILS were comparable between patients who achieved and those who did not achieve 1-year PFS Image-based DL model better predicted patient PFS than manual cell counts, showing the importance of image-based features to predict patient outcome.	Not assessed
Liu et al.[47]	H&E	313 WSI from 264 advanced gastric cancer patients from 4 different centres Four test datasets FAH-SYSU ($N = 139$), FAH-NCU ($N = 90$), SAH-SYSU ($N = 25$), ACH-GZMU ($N = 10$)	Treatment response evaluated according to iRECIST1.1	Prediction of responders vs non-responders	1024 × 1024px tiles resized to 256 × 256px. Ensemble model (ICIsNet) composed of DenseNet, EfficientNet and SwinVit v2	Over the four test cohorts respectively: AUC of 0.952, 0.920, 0.962 and 1.0, Accuracy of 0.848, 0.870, 0.932, 0.900, Sensitivity of 1.0, 0.876, 0.962, 0.750, Specificity of 0.706, 0.863, 0.889, 1.0	The Ensemble model (ICIsNet) performed better than individual models. ICIsNet predictions are based on morphologic features that were identified using heatmaps overlays.	Overlaid heatmap to identify morphologic features: Non-responders: poor differentiation, diffuse tumour tissue, signet ring cells, cellular mucin pools and low lymphocytes infiltration Responders: solid tumour resembling normal tissue, increased lymphocyte infiltration

IHC immunohistochemistry, *H&E* haematoxylin and eosin, *TMA* tissue microarray, *TPS* tumour proportion score, *CPS* combined positive score, *HR* hazard ratio, *CI* 95% confidence interval, *PFS* progression-free survival, *OS* overall survival, *CNN* convolution neural network, *WSI* whole slide image, *NSCLC* non-small cell lung cancer, *TIL* tumour infiltrating lymphocytes.

biomarkers proving to be superior to bulk PD-L1 expression. Three studies identified macrophages (their proximity to tumour cells and PD-L1 expression) as an important cell type for response prediction and three studies highlighted the contribution CD8+ T-cells and their proximity to tumour cells has on improving prediction of response. This was also a feature of response in NSCLC where the potential importance of the immune contexture and its spatial relationship to tumour cells added predictive value when integrated with the clinical benchmark of tumour PD-L1 expression (TPS). Despite these trends within melanoma and NSCLC, no single common spatial biomarker of response was identified within or between disease types across the 26 studies examined. While this is likely to reflect differing biological modalities of response across disease types it may also be reflective in multiple points of heterogeneity between studies.

This became an important factor when assessing the added predictive value of spatial biomarkers. Of the 20 multiplexed image-based studies assessed, 14 studies compared their derived spatial biomarkers of response to benchmarks namely, PD-L1 expression or TMB. However, not all studies assessed PD-L1 status comparative to current clinical standards, e.g., PD-L1 immunohistochemistry TPS/CPS, making a direct evaluation of the predictive utility of each spatial biomarker challenging. Studies which did not evaluate clinical benchmarks for comparison often used multiplex immunofluorescence modalities for data acquisition rendering immunohistochemical assessment on the same tissue specimen challenging or impossible. Additionally, many study cohorts are derived from clinical trials where access to tissue is often limited which may render cutting of additional sections a limitation to further benchmarking experiments. Seven of the 14 studies identified a lack of predictive utility for their respective benchmark measure of PD-L1 expression, excluding those from NSCLC, where 5/6 of the studies confirmed the predictive utility of tumour PD-L1 expression (TPS). The utility of this is also reflected in the clinical application of TPS and not CPS for determining PD-L1 expression in this specific cancer context. Additionally, the same group of studies utilised differing predictive endpoints to determine the utility of their spatial biomarkers, ranging from RECIST criteria, heterogeneous classifications of responders and non-responders from RECIST criteria, progression-free survival, and overall survival. To enable direct comparisons between the utility of spatial biomarkers within cancer types, the field would benefit greatly from consortium-based efforts to compile large multi-disease cohorts on which data acquisition modalities, data analysis methods, clinical benchmarking and predictive endpoints could be homogenised.

The three AI-based studies investigating spatial biomarkers of response demonstrate how AI could be used to identify morphological features predictive of response to ICI from routine H&E slides. With H&E being part of routine clinical pathology, accessibility to a large cohort of slides is readily available. When coupled with the increasing digitisation of slide archives in pathology departments, the potential application of AI for the discovery of spatial biomarkers represents a powerful analysis modality with the potential to democratise spatial biomarker discovery beyond centres with the financial capability to perform more costly multiplexed image-based methods. Additionally, AI modalities could be a powerful analysis tool when applied to multiplexed image-based studies; however, we did not observe this approach to spatial biomarker discovery in the studies assessed in this review. Taken together, evidence suggests that there is a high potential for merging both modalities to accelerate the discovery and clinical translation of new predictive biomarkers for ICI response.

One important factor when considering the utility of predictive biomarkers is the distribution of expression throughout the tumour. Three studies (Antoranz et al., Kim et al., and Bortolomeazzi et al. assessed this and identified significant levels of heterogeneity for their respective biomarkers, highlighting the potential advantage of methodologies which retain the architecture of biomarker expression to facilitate accurate quantitation of tumour heterogeneity. Interestingly, Antoranz et al. identified that PD-L1+ macrophage expression was lower in the tumour area compared to the tumour stroma interface, with expression peaking at the tumour edge. Kim et al. study demonstrated that the spatial distribution of T-cell subsets and innate immune cells differed between responders and non-responders. This suggests that in some cases it is not merely the presence or absence of a spatial interaction but its distribution throughout the tumour which has an impact on response.

Taken together, the assessment and comparison of the above studies demonstrate the importance of assessing the immune contexture and cell-specific PD-L1 expression in a spatially resolved context. This paradigm yielded superior predictors of response to ICI in numerous studies. Realisation of this on a clinical scale will require large multi-disease studies utilising homogeneous modes of analysis and routine integration of spatial measures in clinical trial protocols. As demonstrated by the application of artificial intelligence modalities in this review, the analysis of standard H&E images represents a cost-effective and widely available resource upon which to examine spatial biomarkers of response. Keeping spatial simple may hold the key, in part, to bringing spatial biomarkers into routine clinical practice.

Received: 6 March 2024; Accepted: 5 August 2024;
Published online: 13 August 2024

References

1. Chen, J. et al. ICI efficacy information portal: a knowledgebase for responder prediction to immune checkpoint inhibitors. *NAR Cancer* **5**, zcad012 (2023).
2. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* **377**, 2500–2501 (2017).
3. Conroy, M. & Naidoo, J. Immune-related adverse events and the balancing act of immunotherapy. *Nat. Commun.* **13**, 392 (2022).
4. Choi, J. & Lee, S. Y. Clinical characteristics and treatment of immune-related adverse events of immune checkpoint inhibitors. *Immune Netw.* **20**, e9 (2020).
5. Yang, F. et al. Patterns of toxicity burden for FDA-approved immune checkpoint inhibitors in the United States. *J. Exp. Clin. Cancer Res.* **42**, 4 (2023).
6. Wang, D. Y. et al. Fatal toxic effects associated with immune checkpoint inhibitors: a systematic review and meta-analysis. *JAMA Oncol.* **4**, 1721–1728 (2018).
7. Marcus, L. et al. FDA approval summary: pembrolizumab for the treatment of tumor mutational burden-high solid tumors. *Clin. Cancer Res.* **27**, 4685–4689 (2021).
8. Puccini, A., Seeber, A. & Berger, M. D. Biomarkers in metastatic colorectal cancer: status quo and future perspective. *Cancers (Basel)* **14**, 4828 (2022).
9. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
10. Arkenau, H. T. PD-L1 in Cancer: ESMO Biomarker Factsheet. ESMO. https://oncologypro.esmo.org/education-library/factsheets-on-biomarkers/pd-l1-in-cancer (Published 2017. Accessed May 23, 2023).
11. Akhtar, M., Rashid, S. & Al-Bozom, I. A. PD−L1 immunostaining: what pathologists need to know. *Diagn. Pathol.* **16**, 94 (2021).
12. Sunshine, J. & Taube, J. M. PD-1/PD-L1 inhibitors. *Curr. Opin. Pharm.* **23**, 32–38 (2015).
13. Yang, F., Wang, J. F., Wang, Y., Liu, B. & Molina, J. R. Comparative analysis of predictive biomarkers for PD-1/PD-L1 inhibitors in cancers: developments and challenges. *Cancers (Basel)* **14**, 109 (2021).
14. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet*. https://doi.org/10.1038/s41576-023-00580-2 (Published online 2023).
15. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
16. Park, J. et al. Spatial omics technologies at multimodal and single cell/subcellular level. *Genome Biol.* **23**, 256 (2022).
17. Hu, B., Sajid, M., Lv, R., Liu, L. & Sun, C. A review of spatial profiling technologies for characterizing the tumor microenvironment in immuno-oncology. *Front Immunol.* **13**, 996721 (2022).

18. Summers, H. D., Wills, J. W. & Rees, P. Spatial statistics is a comprehensive tool for quantifying cell neighbor relationships and biological processes via tissue image analysis. *Cell Rep. Methods* **2**, 100348 (2022).
19. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nat. Biotechnol.* **40**, 308–318 (2022).
20. Parra, E. R. Methods to determine and analyze the cellular spatial distribution extracted from multiplex immunofluorescence data to understand the tumor microenvironment. *Front. Mol. Biosci.* **8**, 668340 (2021).
21. Yousefi-Nooraie, R., Irani, S., Mortaz-Hedjri, S. & Shakiba, B. Comparison of the efficacy of three PubMed search filters in finding randomized controlled trials to answer clinical questions. *J. Eval. Clin. Pract.* **19**, 723–726 (2013).
22. Toki, M. I. et al. High-plex predictive marker discovery for melanoma immunotherapy-treated patients using digital spatial profiling. *Clin. Cancer Res.* **25**, 5503–5512 (2019).
23. Lu, Y. et al. Resolution of tissue signatures of therapy response in patients with recurrent GBM treated with neoadjuvant anti-PD-1. *Nat. Commun.* **12**, 4031 (2021).
24. Antoranz, A. et al. Mapping the immune landscape in metastatic melanoma reveals localized cell–cell interactions that predict immunotherapy response. *Cancer Res.* **82**, 3275–3290 (2022).
25. Kim, S. et al. Differential RNA expression of immune-related genes and tumor cell proximity from intratumoral M1 macrophages in acral lentiginous melanomas treated with PD-1 blockade. *Biochim. Biophys. Acta Mol. Basis Dis.* **1868**, 166516 (2022).
26. Attrill, G. H. et al. Higher proportions of CD39+ tumor-resident cytotoxic T cells predict recurrence-free survival in patients with stage III melanoma treated with adjuvant immunotherapy. *J. Immunother. Cancer* **10**, e004771 (2022).
27. Johnson, D. B. et al. Quantitative spatial profiling of PD-1/PD-L1 interaction and HLA-DR/IDO-1 predicts improved outcomes of anti-PD-1 therapies in metastatic melanoma. *Clin. Cancer Res.* **24**, 5250–5260 (2018).
28. Martinez-Morilla, S. et al. Biomarker discovery in patients with immunotherapy-treated melanoma with imaging mass cytometry. *Clin. Cancer Res.* **27**, 1987–1996 (2021).
29. Ghiringhelli, F. et al. Immunoscore immune checkpoint using spatial quantitative analysis of CD8 and PD-L1 markers is predictive of the efficacy of anti- PD-1/PD-L1 immunotherapy in non-small cell lung cancer. *EBioMedicine* **92**, 104633 (2023).
30. Qin, A. et al. Cellular engagement and interaction in the tumor microenvironment predict non-response to PD-1/PD-L1 inhibitors in metastatic non-small cell lung cancer. *Sci. Rep.* **12**, 9054 (2022).
31. Song, X. et al. Spatial multi-omics revealed the impact of tumor ecosystem heterogeneity on immunotherapy efficacy in patients with advanced non-small cell lung cancer treated with bispecific antibody. *J. Immunother. Cancer* **11**, e006234 (2023).
32. Moutafi, M. K. et al. Spatially resolved proteomic profiling identifies tumor cell CD44 as a biomarker associated with sensitivity to PD-1 axis blockade in advanced non-small-cell lung cancer. *J. Immunother. Cancer* **10**, e004757 (2022).
33. Moutafi, M. et al. Discovery of biomarkers of resistance to immune checkpoint blockade in NSCLC using high-plex digital spatial profiling. *J. Thorac. Oncol.* **17**, 991–1001 (2022).
34. Zugazagoitia, J. et al. Biomarkers associated with beneficial PD-1 checkpoint blockade in non–small cell lung cancer (NSCLC) identified using high-plex digital spatial profiling. *Clin. Cancer Res.* **26**, 4360–4368 (2020).
35. Bortolomeazzi, M. et al. Immunogenomics of colorectal cancer response to checkpoint blockade: analysis of the KEYNOTE 177 trial and validation cohorts. *Gastroenterology* **161**, 1179–1193 (2021).
36. Phillips, D. et al. Immune cell topography predicts response to PD-1 blockade in cutaneous T cell lymphoma. *Nat. Commun.* **12**, 6726 (2021).
37. Jia, K. et al. Multidimensional immune profiling in Gastric Cancer Multiplex Immunohistochemistry Atlas from Peking University Cancer Hospital project informs PD-1/PD-L1 blockade efficacy. *Eur. J. Cancer* **189**, 112931 (2023).
38. Ma, X. et al. Spatial distribution and predictive significance of dendritic cells and macrophages in esophageal cancer treated with combined chemoradiotherapy and PD-1 blockade. *Front. Immunol.* **12**, 786429 (2022).
39. Färkkilä, A. et al. Immunogenomic profiling determines responses to combined PARP and PD-1 inhibition in ovarian cancer. *Nat. Commun.* **11**, 1459 (2020).
40. Hammerl, D. et al. Spatial immunophenotypes predict response to anti-PD-1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nat. Commun.* **12**, 5668 (2021).
41. Wang, X. Q. et al. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature* **621**, 868–876 (2023).
42. Ligero, M. et al. Weakly supervised deep learning predicts immunotherapy response in solid tumors based on PD-L1 expression. *Cancer Res. Commun.* **4**, 92–102 (2024).
43. Shamai, G. et al. Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. *Nat. Commun.* **13**, 6753 (2022).
44. Park, S. 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.* **40**, 1916–1928 (2022).
45. Hu, J. et al. Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images. *Transl. Oncol.* **14**, 100921 (2021).
46. Shibaki, R. et al. Machine learning analysis of pathological images to predict 1-year progression-free survival of immunotherapy in patients with small-cell lung cancer. *J. Immunother. Cancer* **12**, e007987 (2024).
47. Liu, Y. et al. Deep learning based digital pathology for predicting treatment response to first-line PD-1 blockade in advanced gastric cancer. *J. Transl. Med.* **22**, 438 (2024).

Acknowledgements

The authors would like to acknowledge the support of the Swiss National Science Foundation (SNSF) Ref: CRII5_193832. Figures created with BioRender.com.

Author contributions

HLW: concept, publication identification, paper writing. ALF: concept, paper writing. TK: concept, paper writing. MDB: concept, paper writing. HD: concept, paper writing. OM: concept. IZ: concept, paper writing. All authors read and approved the final draft of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Hannah L. Williams.