npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-024-00749-w

Integrating histopathology and transcriptomics for spatial tumor microenvironment profiling in a melanoma case study

Óscar Lapuente-Santana [1,2,3,4,6], Joan Kant [1,5,6] & Federica Eduati [1,2,3] ✉

Local structures formed by cells in the tumor microenvironment (TME) play an important role in tumor development and treatment response. This study introduces SPoTLIghT, a computational framework providing a quantitative description of the tumor architecture from hematoxylin and eosin (H&E) slides. We trained a weakly supervised machine learning model on melanoma patients linking tile-level imaging features extracted from H&E slides to sample-level cell type quantifications derived from RNA-sequencing data. Using this model, SPoTLIghT provides spatial cellular maps for any H&E image, and converts them in graphs to derive 96 interpretable features capturing TME cellular organization. We show how SPoTLIghT's spatial features can distinguish microenvironment subtypes and reveal nuanced immune infiltration structures not apparent in molecular data alone. Finally, we use SPoTLIghT to effectively predict patients' prognosis in an independent melanoma cohort. SPoTLIghT enhances computational histopathology providing a quantitative and interpretable characterization of the spatial contexture of tumors.

Pathology images of surgically removed or biopsied tumor tissue sections play a crucial role in modern clinical practice aiding in cancer diagnoses, subtype classification and treatment planning[1]. In the clinics, these tasks are predominantly performed via manual microscopic inspection of hematoxylin and eosin (H&E) stained images by expert pathologists for morphologic evaluation of tumors. Visual inspection is often complemented with molecular characterization of the tissue, which further enhances our understanding of tumor progression and helps identify important biomarkers essential for precision oncology[1,2].

The emergence of machine learning (ML) is revolutionizing the field of pathology by automating these tasks through computational pathology[3]. Moreover, ML algorithms have demonstrated the capability to predict various molecular features such as microsatellite instability[4], gene mutations[5–7] and chromosomal instability[8]. However, most existing algorithms focus on end-to-end predictions at the patient level[9], lacking insight into the spatial architecture of tumors.

Growing evidence highlights the critical role of spatial context and cellular composition in the tumor microenvironment on cancer prognosis and treatment outcomes[10,11]. To better understand how cellular organization affects tumor progression and treatment response, there is a need for more interpretable models that not only make accurate predictions but also provide clear insights into the reasoning behind those predictions (e.g., using attention-based learning)[12,13]. Incorporating spatial information into ML algorithms can enhance both explainability and predictive accuracy[14].

An alternative approach to digital pathology is computational staining, where algorithms are trained to identify distinctive features at the tissue, cellular or intracellular level from standard H&E images, mirroring the pathologist workflow[15–18]. This automated analysis can be used to derive quantitative descriptors of the spatial arrangement of different cell types from tissue-imaging slides[15,16,19]. However, computational staining relies on labor-intensive manual pathology annotations to train the models. To address this limitation, several approaches have been developed. For instance, Corredor et al. proposed an automated method that leverages

[1]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. [2]Eindhoven Artificial Intelligence Systems Institute, Eindhoven University of Technology, Eindhoven, The Netherlands. [3]Institute for Complex Molecular Systems, Eindhoven University of Technology, Eindhoven, The Netherlands. [4]Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [5]Princess Margaret Cancer Research Centre, University Health Network, Toronto, Canada. [6]These authors contributed equally: Óscar Lapuente-Santana, Joan Kant. ✉e-mail: f.eduati@tue.nl

knowledge of TIL morphology compared to cancerous cells (e.g., smaller nuclei, more rounded, and with darker, more homogeneous staining)[20], while Perera et al. introduced a strategy that requires only slide-level labels, using attention-based mechanisms to identify tumor-infiltrating lymphocytes (TILs)[21]. While promising, these methods are largely based on recognizing specific cellular morphologies, limiting their applicability to a narrow range of morphologically distinct cell types, such as TILs.

High-throughput molecular data, such as genomics and transcriptomics, are a rich resource of information on the TME and its response to therapy[22]. They have been used to study the impact of cellular composition[23,24] and functional state[25], and to characterize immune subtypes that are related to response to immune checkpoint blockers[26–28]. Yet, these data lack spatial information, making them complementary to imaging techniques, which accurately portray the spatial localization of immune cells relative to the tumor or its stroma. The availability of matching H&E slides and molecular data have paved the way for the identification of spatially-resolved biomarkers[6,7,29,30]. Combining H&E pathology with RNA sequencing (RNA-seq) shows promise also in spatially mapping TME cells without manual annotation. A pioneer study by Fu and colleagues have shown that weakly supervised learning can be leveraged to enable local mapping of transcriptionally defined lymphocytes[7].

In this study, we develop a pipeline to automatically characterize the spatial distribution of multiple cell types in the melanoma TME from H&E slides. Using data from 379 melanoma patients in The Cancer Genome Atlas (TCGA)[31], we train our model by integrating matching H&E slides and RNA-seq data. Using the deep-learning method from Fu et al.[7], we derive histopathological features on tiled H&E images. We then apply multi-task transfer learning to investigate the association between these features and transcriptome-based signatures of four key TME cell types: tumor cells, T cells, cancer-associated fibroblasts (CAFs), and endothelial cells.

Our approach generates spatial maps encoding the probability of different cell types in each tile of the slide, providing insights into their spatial distribution and co-occurrence. We then transform these maps into spatial graphs and utilize graph theory to extract interpretable features capturing cellular architecture within the TME. These spatial graph-based features not only distinguish between different microenvironment subtypes but also uncover local patterns of immune infiltration that are not detectable from molecular data alone. Lastly, we use these spatial features to build prognostic models for early-stage melanoma patients, and validate that these features can be reliably derived from new images, retaining their predictive power in forecasting patient outcomes in an independent validation cohort.

This proposed computational framework is provided as a tool called Spatial Profiling of Tumors by Leveraging Imaging and Transcriptomics (SPoTLIghT; https://github.com/SysBioOncology/SPoTLIghT) that can be applied on pathology slides to characterize the spatial biological contexture of tumors.

Results

Leveraging H&E imaging and bulk RNA-seq to portray the spatial architecture of the melanoma microenvironment

We used matching tumor tissue-imaging H&E slides and bulk RNA-seq data from 379 TCGA melanoma patients to spatially resolve cellular composition in the melanoma TME (Fig. 1a).

First, we used the PC-CHiP deep learning model from Fu and colleagues[7] to extract 1536 imaging features for thousands of tiles (with digital resolution of 512 pixels x 512 pixels, at 20x magnification) of the fresh-frozen (FF) H&E slides (Fig. 1a). These features consist of the last neurons of a convolutional neural network (CNN) trained to distinguish tumor from normal tissue[7].

Additionally, we quantified, for each melanoma slide, tumor cells, T cells, CAFs and endothelial cells based on molecular data (Fig. 1a). For tumor cells we derived three types of quantifications, one based on somatic DNA copy number[32] and two based on gene expression[33]. For T cells, CAFs and endothelial cells we selected multiple published transcriptomic-based signatures resulting in four, three and three quantification scores respectively (see Methods for a more detailed description of the approaches used for each cell type). We verified that, for each cell type, there was a good agreement between the quantifications obtained using the different approaches (Pearson correlation >0.7, $p < 0.05$).

In order to derive a spatial distribution of the four cell types in the whole slide, we formulated a weakly supervised multi-task transfer learning problem where we considered the multiple cell type-specific signatures as output variables (tasks). We trained four regularized multi-task linear regression (RMTLR) algorithms, i.e., one for each cell type, through cross-validation using the 1536 histopathological features as input (Methods, cross-validation performances in Supplementary Fig. 1). The advantage of multi-task learning is that it enables simultaneous learning from different cell type quantification methods. This is important as each method has its own specific capabilities and limitations[34]. By integrating information from all methods, the model becomes more robust and generalizable, effectively compensating for the individual shortcomings of each approach.

Finally, using our cell type-specific models, we computed per-tile predictions of the quantification of each cell type (Fig. 1a). We converted this information to cell type-specific probability maps capturing the spatial distribution of each cell type across the slide.

To assess the accuracy of our SPoTLIghT-derived spatial maps, we systematically compared them with spatial TIL maps generated using the computational staining method from Saltz et al.[16] across 334 melanoma patients from the TCGA cohort. For this comparison we only used formalin-fixed paraffin-embedded (FFPE) slides which were not used for model training, that was based on FF slides. We observed a good agreement between the maps reconstructed using the two methods (Fig. 1b), quantitatively confirmed by an average Dice score of 0.92 and an average Jaccard index of 0.74 (Fig. 1c).

We additionally validate the predictions of SPoTLIghT using two publicly available skin melanoma Xenium datasets from 10x Genomics. We processed and manually annotated the cells from the Xenium data using known cell type markers (see Methods, Supplementary Figs. 2a,b and 3a,b). Then SPoTLIghT was used to generate the cell type maps for the corresponding H&E images. By using a binning approach, we compared the annotated Xenium cell maps with the SPoTLIghT tile predictions (see Methods).

From a visual inspection of the Xenium slide with the add-on panel, we could observe that the SPoTLIghT-derived spatial maps were able to capture some global patterns that were seen in the Xenium maps (Fig. 1d; Supplementary Fig. 3c). For example, in the CAFs Xenium map, there was an oval-shaped area without CAFs, for which the tile predictions tended to be lower. In the T cells map, a distinct dark area in the bottom-left of the Xenium map was observed. This was in agreement with the SPoTLIghT-derived spatial map. The Spearman correlation coefficient was used to assess the relationship between the mean predicted probability and the number of cells per bin. Significant relationships were found across all cell types ($p < 0.05$) with a weak to moderate strength for the same slide (Fig. 1e). More specifically, the CAFs and T cells were found to show the strongest association with a coefficient of $R = 0.52$ and $R = 0.62$, respectively.

For the Xenium slide without the add-on panel, similar observations were made, namely that the SPoTLIghT-derived spatial map can capture global patterns better, while smaller resolution patterns may be lost (Supplementary Fig. 2c). In this case, significant associations were observed for T cells and tumor cells ($R = 0.33$, $p = 4.4 \times 10^{-8}$ and $R = 0.34$, $p = 1.8 \times 10^{-9}$) (Supplementary Fig. 2d).

Overall, these results highlight a good agreement between our predicted cell-type maps and those inferred from this established computational method or derived from single cell spatial transcriptomics data.

Cell types global quantification and tile-level co-localization correlate with microenvironment subtypes

From a first qualitative analysis we could observe that these cell type maps could inform on differences between four microenvironment subtypes as defined by Bagaev and colleagues[28] (Fig. 2a). Both immune-enriched (IE)

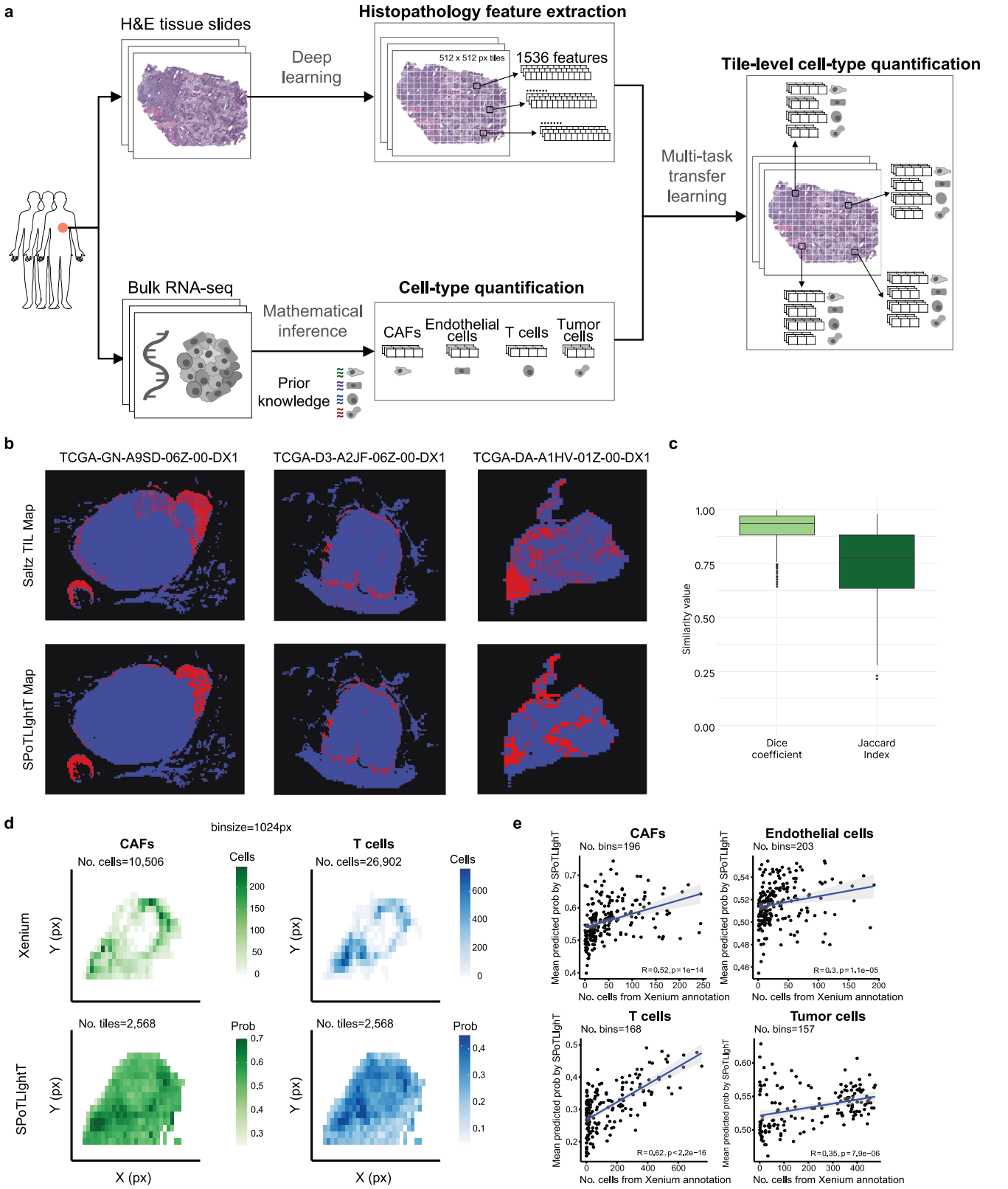

Fig. 1 | Inference of spatial cellular maps using SPoTLIghT. a Pipeline for prediction of tile-level cell type quantification. **b** Three examples of qualitative comparison of Saltz and SPoTLIghT tumor-infiltrating lymphocytes (TIL) maps. **c** Quantitative comparison of Saltz and SPoTLIghT TIL maps in terms of Dice and Jaccard similarity coefficients across 334 FFPE melanoma slides from TCGA. **d** Xenium (top) and SPoTLIghT (bottom) spatial CAFs and T cells maps based on bin-level cell-type annotations and on tile-level cell-type predictions, respectively. Probabilities were limited to 0.05 and 0.95 quantiles to enhance contrast, values outside of limits were assigned to lower or upper limits. **e** For each cell type, the correlation between Xenium bin-level predictions and SPoTLIghT tile-level predictions is shown.

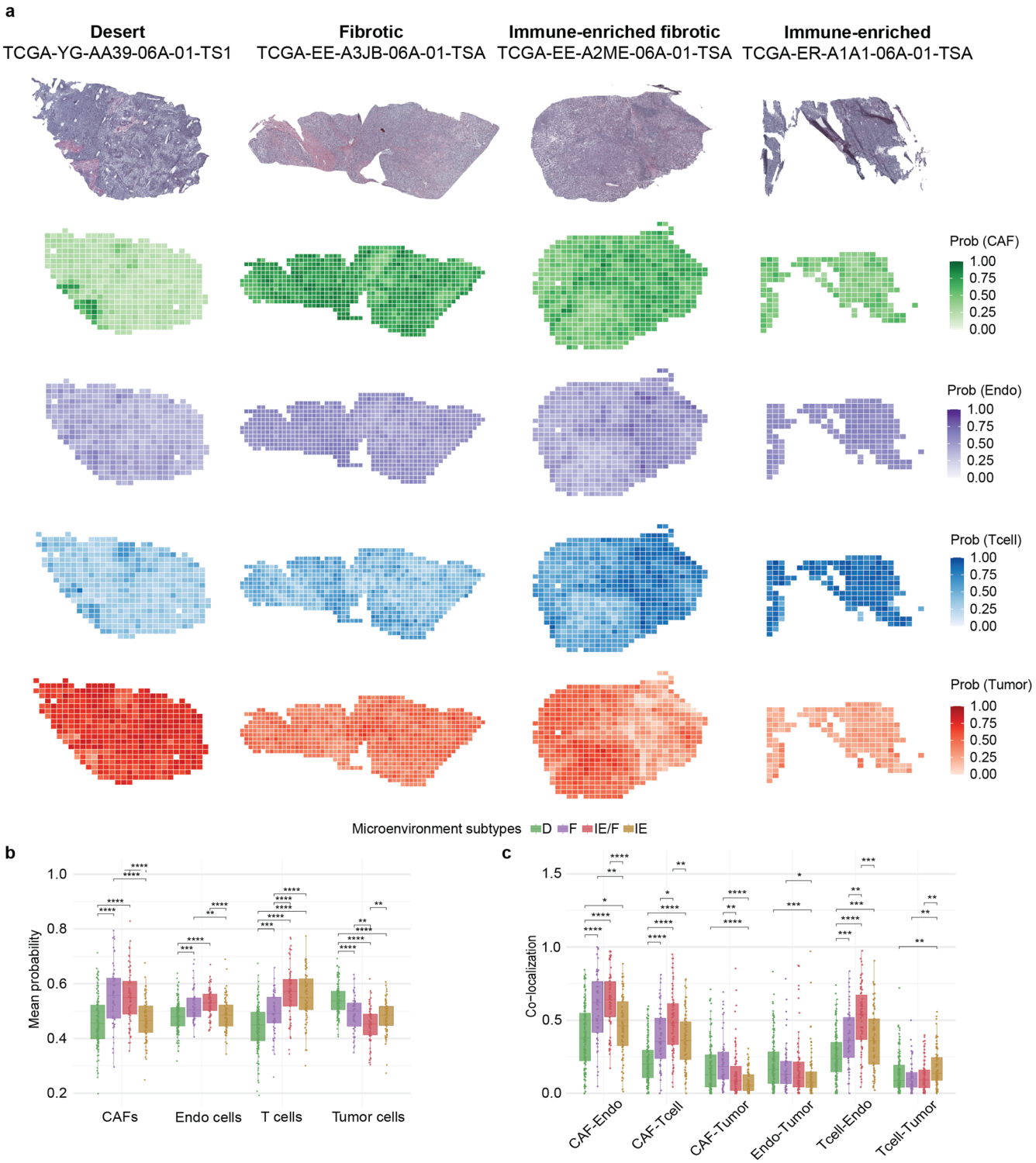

Fig. 2 | SPoTLIghT reveals distinct spatial patterns across microenvironment subtypes. a Cell type-specific probability maps for four examples of melanoma slides from distinct microenvironments: immune-depleted (D), fibrotic (F), immune-enriched fibrotic (IE/F) and immune-enriched (IE) subtypes. Note that each slide included two or three slices of tissue but only one slice is shown for the sake of visualization. **b** Slide-level quantification of each cell type and **c** co-localization assessment of cell type pairs across the distinct microenvironment subtypes (Wilcoxon rank-sum test; $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (**) and $p < 0.05$ (*)). Non-significant comparisons are not shown. The number of slides for the D, F, IE/F and IE microenvironment subtypes were 145, 57, 74 and 85 respectively.

and immune-enriched fibrotic (IE/F) slides showed high numbers of T cells. However, while T cells were distributed uniformly across the slide in the IE case, suggesting a good infiltration in the tumor tissue, they tended to be more clustered in the IE/F slides (e.g., T cells localized on the right part of the slide in the IE/F example in Fig. 2a) and co-localized with CAFs and endothelial cells. This suggests that, in the IE/F slide, T cells are not infiltrated in the area of tumor cells (e.g. in the left side of the slide in the IE/F slide in Fig. 2a), in agreement with the worst prognosis that characterizes IE/F tumors[28]. In F slides there is a general scarcity of T cells, which are absent from the main tumor cells' enriched area (right part of the slide in Fig. 2a). On the contrary, fibroblasts are spread over all the slides, including the tumor area. These considerations, albeit qualitative, were consistently

observed across several slides and provide the first evidence of the impressive potential that weakly labeled multi-task learning approach (i.e., cell type quantification provided at the patient level) has to provide spatial insights on the relative cellular distribution on H&E slides. This is particularly relevant given that spatial distribution is a complementary hallmark (distinguishing inflamed and excluded tumors) of patients' immune response and response to immunotherapy[35].

To expand on these qualitative observations, we started deriving quantitative features from the cell type-specific probability maps. First, we calculated, for all melanoma slides, the overall content of each cell type by averaging our per-tile predictions. As expected, cell type quantification is reflected in the four microenvironment subtypes (Fig. 2b): high proportions of stromal cells (CAFs and endothelial cells) were found in both fibrotic subtypes (F and IE/F), while enrichment of T cells was present in the immune-enriched subtypes (IE and IE/F). Tumor cells were most abundant in the D subtype.

To start diving into local spatial patterns, we investigated co-localization of pairs of cell types at the tile level (Fig. 2c), assuming that high co-localization indicates that two cell types are in physical proximity which might be indicative of cell-cell interaction. T cells were colocalized with tumor cells specifically in the IE subtype ($p < 0.01$ compared with all the other subtypes), suggesting a higher infiltration of T cells in the tumor. Instead, T cells were more co-localized with CAFs in the IE/F tissues ($p < 0.05$), suggesting potentially an immuno-inhibitory effect. While the co-occurrence of endothelial cells and T cells allowed to clearly distinguish the IE/F from the F slides ($p < 0.01$), the high co-localization of endothelial cells with CAFs stratified these fibrotic subtypes from the D and IE slides ($p < 0.001$). The latter observation is in line with the characteristic conversion of endothelial cells into CAF-like phenotype cells in fibrotic tumors[36].

Overall, our results suggest that even a simple quantitative feature of local cellular patterns such as co-localization of cell types can stratify different microenvironment subtypes. While the abundance of cell types in the sample affects their a priori probability of co-localization, this is clearly not the only factor, as most of the observations above could not be derived by looking only at patient/slide-level cell type quantification results.

Constructing spatial graphs and derivation of features

In the previous section, we showed that analyzing basic features reflecting the spatial convergence of various cell types offers valuable insights into distinct microenvironment subtypes. However, previous assessments of cell type co-localization were limited to isolated regions (tiles), without considering if this co-localization is consistent across the entire slide or concentrated in specific areas. To address this gap, we created spatial graphs from our tile-level quantifications, treating tiles as nodes (labeled as one or more cell types) connected by undirected edges to their direct spatial neighbors (Fig. 3a). This approach allowed us to extract graph-based features capturing node degree (ND), number of shortest paths (SP), size of the largest connected component (LCC), fraction of cell type specific clusters and proximity between the different classified clusters. Feature sets are illustrated in Fig. 3b–f and summarized in Table 1.

Graph-based features quantifying connectivity and proximity of cell types correlate with microenvironment subtypes

Our graph-based features (Table 1) provide a quantification of interpretable spatial structures from our cell type maps. To validate their value, we systematically investigated their association with the four above mentioned microenvironment subtypes.

We investigated the local proximity of different cell types using two different types of features: the average ND (i.e., the number of direct neighbors) for each cell type pair across the slide; and the number of SP between two cell types that have a maximum length of two (SP2). While ND is strongly affected by the global quantification of the different cell types (e.g., CAFs are always more likely to be direct neighbors of any other cell type in the fibrotic tissues, Supplementary Fig. 4), SP2 provides complementary information on microenvironment subtypes (Fig. 4a). An interesting example is the SP2 computed between T cells and tumor cells. This feature can clearly stratify IE slides from all the other subtypes, including the IE/F subtypes ($p < 0.001$), suggesting that a shorter distance between T cells and tumor cells specifically characterizes the IE subtype. To correct for the cell type quantification influence on the ND, we compared the calculated ND with the ones obtained when the cell types are randomly distributed across the slide and computed the effect size (ND_effsize, Supplementary Fig. 5). Besides accounting for the bias towards highly abundant cell types (e.g. CAFs do not prevail in the neighborhood of other cell types in fibrotic slides), this feature allowed us to capture actual spatial structure. For example, both the D and IE subtypes could be discerned based on the ND_effsize(Tumor,CAF) establishing that tumor cells were more distant from CAFs than if distributed at random in the IE subtype. Note that since this feature focuses more on spatial patterns, less significant results were found between distinct microenvironment subtypes, which are defined from molecular data.

Next, we used the LCC to investigate more global network properties (Fig. 4b). We observed that CAFs were more connected in the fibrotic subtypes than in the non-fibrotic subtypes ($p < 0.05$ in all pairwise comparisons of F and IE/F with IE and D). It is known that the CAFs are overall more abundant in these subtypes, but this confirms that they also form larger clusters. Similar observations can be reported for the T cells, which are more connected in the immune-enriched subtypes ($p < 0.001$ comparing IE and IE/F with F and D). Interestingly, endothelial cells were significantly more connected in the IE/F subtype than in all other subtypes ($p < 0.01$).

Cell type-specific spatial clusters are associated with microenvironment subtypes

Motivated by the observed connectivity in individual cell types, we used SCHC to identify clustering patterns of connectivity of nodes with similar predicted cell type probabilities that are also spatially linked. We built cell type-specific clusters based on the predictions of each individual cell type (more details in Methods and in Table 1). This means that nodes (i.e. tiles) were assigned to the same cluster if they were in geometrical proximity and they had a similar abundance of the specific cell type. Each cluster was then considered as enriched ('high') or depleted ('low') of a certain cell type based on the comparison with all clusters in all slides.

The calculated fractions of 'high' cell type clusters stratified the microenvironment subtypes similarly to the previous features (Supplementary Fig. 6). F and IE/F patients had a higher fraction of 'high' CAF clusters ($p < 0.0001$). In addition, the IE/F subtype also displayed a large number of clusters with high abundance of endothelial cells ($p < 0.05$). 'High' T cell clusters were mainly characteristic of IE/F and IE slides, whereas D slides displayed a higher fraction of 'high' tumor clusters ($p < 0.001$).

Since the total number of clusters per slide is fixed, a high fraction of clusters in general indicates that a given cell type is not only highly abundant but also more evenly distributed across the slide. However, when there is only a small fraction of 'high' clusters of a certain cell type in the slide, these could be located nearby or farther apart, with a different impact on the TME organization. Therefore, as additional features, we quantified the relative distance between clusters in terms of proximity. Higher values of proximity correspond to smaller distances between clusters and vice versa (see Methods for the formal definition of proximity) as shown in the results for cell type-specific clusters in Fig. 4c. In IE/F subtype compared to IE subtype, clusters enriched in T cells were in closer proximity to those enriched in endothelial cells (average prox_clust(Tcell,Endo (H-H)) = 0.74 (IE/F) vs 0.62 (IE), $p < 0.01$) and more distant from clusters depleted of endothelial cells (average prox_clust(T cell,Endo (H-L)) = 0.57 (IE/F) vs 0.65 (IE), $p < 0.05$). These findings could result from the formation of high endothelial venules which are major sites of inflammatory cells entry into tumors, promoting the development of fibrosis[37]. Clusters of CAFs and endothelial cells also showed to be in closer proximity in the IE/F subtype compared to the IE subtype (average prox_clust(CAF-Endo (H-H)) = 0.76 (IE/F) vs 0.63 (IE), $p > 0.001$). The close spatial interaction between these cell types could be explained by the role of CAFs in the modulation of tumor vasculature, for example, by increasing endothelial cell motility and permeability[38].

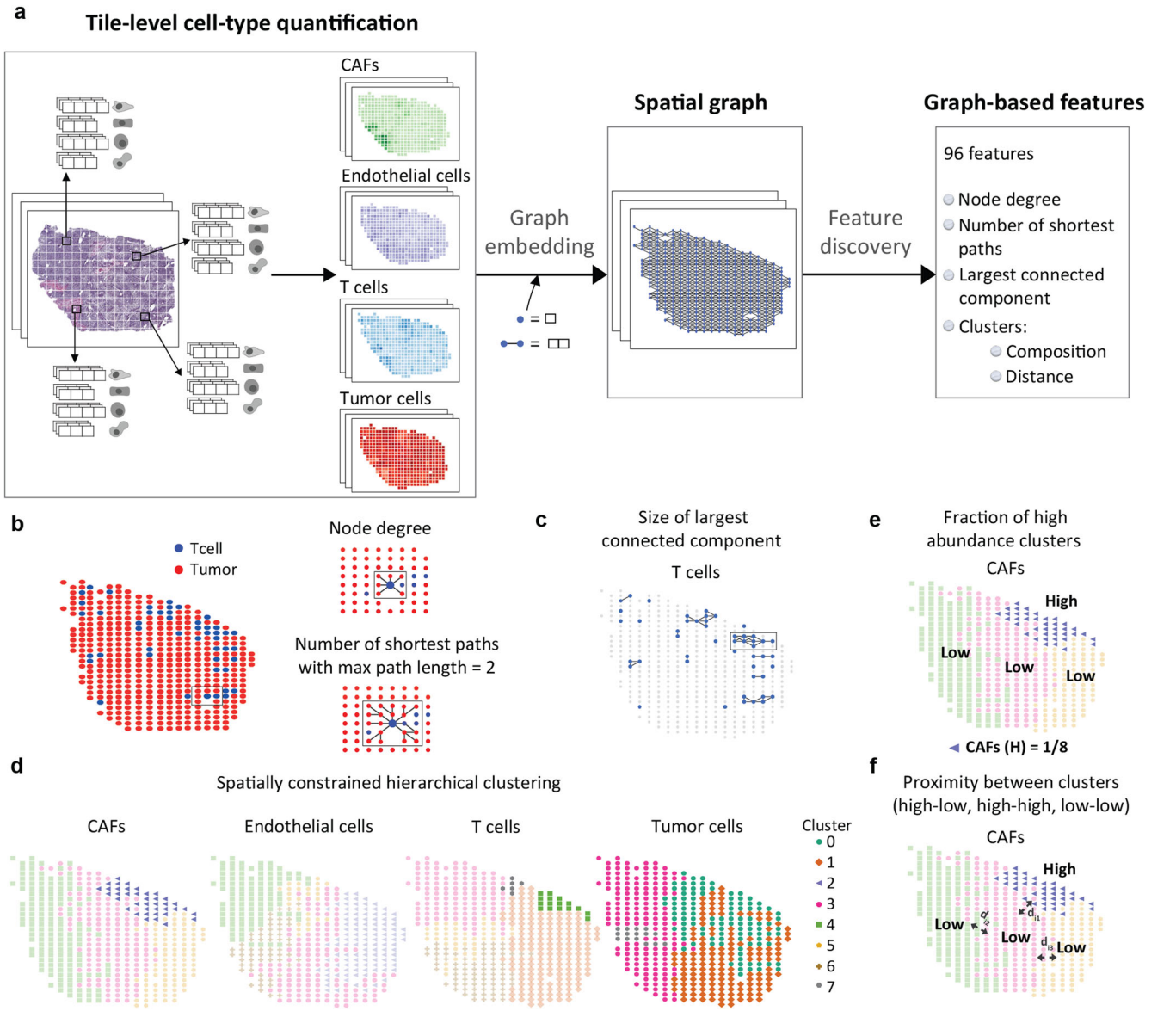

Fig. 3 | Derivation of quantitative and interpretable spatial features using SPoTLIghT. a Approach taken to extract graph-based features from our spatial cell type quantifications. **b** Illustration of local features derived from the spatial graphs for the T cells and tumor cells. Zoom-in in the highlighted area (rectangle) to illustrate: the average node degree and the number of shortest paths with a maximum length of two between T cells and tumor cells. **c** The largest connected component illustrated for T cells. **d** Cell type-specific clusters that were labeled as 'high' (darker colors) or 'low' (lighter colors) abundance depending on whether the average quantification in the cluster was higher or lower than the average quantification across all clusters and slides. **e** the fraction of high/low abundant clusters for the CAFs. **f** Concept of relative proximity between these clusters, where d represents the definition of distance between clusters. Please note that the example H&E slide included two separate pieces of tissue, but for the sake of visualization we are only showing the left piece. This is also the reason why not all eight clusters are visible.

Overall, we have shown that different sets of graph-based features are informative of the structure of the microenvironment and can explain differences in microenvironment subtypes.

Spatial graph-based interpretable features unmask local patterns not visible from molecular data

Our previous results demonstrated the ability of our derived features to describe characteristics associated with different microenvironment subtypes. Although microenvironment subtypes have been extensively validated and shown to be associated to some degree with immunotherapy response[28], they are intrinsically limited as they arise exclusively from molecular data which does not take spatial context into account. We, therefore, investigated the potential of our spatial graph-based features to uncover local patterns that could be missed by molecular subtypes.

In the previous analysis, the computed values for graph-based features exhibit considerable variation across slides within each molecular subtype, as evident from the broad distribution of these features. This variability suggests diverse spatial organization patterns within each molecular subtype. To exemplify, we investigated specific instances where slides, despite being assigned to the same molecular subtype and having similar estimated cell type proportions, displayed markedly distinct predicted spatial distributions.

We first visualized the cell type-specific probability maps of two slides that were classified as IE subtypes (Fig. 5a) and had comparable cell type quantifications (P(T cell) = 0.51 [left] vs 0.48 [right], P(Tumor) = 0.47 vs 0.51, P(CAF) = 0.48 vs 0.46, P(Endo) = 0.46 vs 0.44; see Supplementary Table 1). We then focused on the interpretation of how the graph-based features quantified observed differences in the cell type maps that have potential clinical implications. In both examples, we found that CAFs and

Table 1 | Description of the derived graph-based features

Feature	Description	Intuitive meaning
mean_ND(A,B) 16 features	Average number of neighbor nodes of cell type B surrounding nodes of cell type A. Nodes are assigned a cell type label if the predicted probability of that node (tile) for the given cell type is higher than 0.5.	Higher values of these features indicate higher connectivity. If cells are well-connected, they are more likely to interact.
ND_effsize(A,B) 16 features	Cohen's d measure of effect size computed comparing the mean_ND(A,B) with the null distribution obtained by recomputing the mean_ND randomly assigning the A or B cell type label to each node preserving the total number of cell type A and B nodes in the network. For a negative effect size, the true average mean_ND(A,B) is larger than the simulated average mean_ND(A,B) meaning that the two cell types in the actual slide are closer together compared to a random distribution of these two cell types. Vice versa for a positive effect size. Nodes are assigned a cell type label as described above.	
SP2(A,B) 8 features	Number of shortest paths (SP) with maximum length of two between nodes of cell type A and nodes of cell type B. Nodes are assigned a cell type label as described above.	
LCC(A) 4 features	Normalized size of the largest connected component (LCC) for cell type A. This is defined as the largest set of nodes of cell type A connected with at least one path between every pair of nodes, divided by the total number of nodes of cell type A. Nodes are assigned a cell type label as described above.	
frac_clust(A) 4 features	Fraction of clusters enriched in cell type A. This is defined using SCHC to cluster node (tile) quantifications (predicted probabilities) of cell type A in eight clusters. We used a fixed number of clusters to allow comparison of results between slides, and the number eight was empirically determined observing that results were robust to the number of selected clusters. Each cluster was defined as enriched ('high') or depleted ('low') in cell type A if the mean of cell type A computed across the nodes in the clusters is respectively higher or lower than the overall mean computed across all clusters of all slides for cell type A. Finally, the fraction of 'high'/'low' clusters was calculated by dividing the amount of 'high'/'low' clusters by the total number of clusters (8).	A higher fraction of these clusters reveals not only higher abundance but also more dispersed across the slide. Instead, lower fractions just suggest lower abundance.
prox_clust(A,B (H/L-H/L)) 36 features	Proximity between clusters enriched (H) or depleted (L) in cell type A and B. This is computed as follows: 1. For all pairs of clusters in each slide: a. Compute all minimum distances D between the nodes of two clusters c_1 and c_2. b. Count the number of distances with a maximum length: $S = \sum_{i=1}^{N} D_i \leq l_{\max}$, where D_i is the i-th minimum distance, $l_{\max} = \sqrt{(2xl_{tile})^2 + (2xl_{tile})^2}$, $l_{tile} = 462$. c. Calculate proximity as $proximity = \frac{S}{\max(size[c_1].size[c_2])}$ 2. Average the top three proximities.	Higher proximity values allude to an increased likelihood of interaction between clusters.

endothelial cells were mostly confined in the top-right and bottom-right respectively as indicated by their high mean_ND values (mean_ND(CAF, Endo) = 6.03 vs 4.82, mean_ND(Endo, CAF) = 5.13 vs 7.23). In the left slide, we saw areas of high T cells which are surrounded by CAFs and are kept confined from the areas of tumor cells. This is reflected by a low value in features related to the proximity between tumor and T cells (mean_ND(T cell, Tumor) = 2.5, mean_ND(Tumor, T cell) = 2.45; SP2(Tumor, T cell) = 1632). In contrast, in the right slide T cells are localized in close proximity to tumor cells and away from stromal areas. This is reflected by higher value in features related to proximity between tumor and T cells (mean_ND(T cell, Tumor) = 6, mean_ND(Tumor, T cell) = 3.64, SP2 = 2482) and lower proximity values between CAFs and T cells (mean_ND(T cell, CAF) = 1.31, mean_ND(CAF, T cell) = 2.02, SP2 = 651).

In the above example, the left slide showed less T cell infiltration into the tumor than the right slide even though both were considered to have an IE microenvironment. This qualitative observation can be quantitatively confirmed by looking at the spatial connectivity of T cells, as implied by the difference in the size of the LCC which was larger in the right slide as T cells had a more uniform distribution across the tumor area (LCC (left) = 0.44, LCC (right) = 0.58). Our clustering analysis revealed instead a low fraction of T cell and high fraction of tumor cell clusters in the right slide (frac_clust(T cell) = 0.25 and frac_clust(Tumor) = 0.62). This did not match our speculations about the association of T cell infiltration with the abundance of T cells clusters. We, therefore, examined the relative positioning of T cell and tumor clusters and we found that T cell clusters were far from 'low' tumor areas (prox_clust(Tcell,Tumor (H-L)) = 0.31) but close to 'high' tumor clusters (prox_clust(Tcell, Tumor (H-H)) = 0.48), confirming the observation of a more inflamed phenotype.

We then conducted a similar analysis but now studying two F slides (Fig. 5b), also with similar cell type quantifications (P(Tcell) = 0.53 [left] vs 0.55 [right], P(Tumor) = 0.53 vs 0.50, P(CAF) = 0.53 vs 0.57, P(Endo) = 0.47 vs 0.54; see Supplementary Table 1). The expected picture was now two slides with high stromal content but exhibiting completely different spatial distributions. The left F slide showed that CAFs were restricted to an adjacent area of the tumor core denoting less interaction between CAFs and tumor cells (mean_ND(CAF,Tumor) = 2.01, mean_ND(Tumor,CAF) = 1.39; prox_clust(CAF,Tumor (H-H)) = 0.08). On the contrary, the right F slide manifested a more uniform layout of CAFs bordering the tumor suggesting a higher degree of communication between these cell types (mean_ND(CAF,Tumor) = 4.75, mean_ND(Tumor,CAF) = 3.8; prox_clust(CAF,Tumor (H-H)) = 0.43). It is worth noticing the greater overlap between T cells and CAFs in the right case which kept T cells in the periphery of the tumor (prox_clust(CAF,Tcell (H-H)) = 0.69 [left] and 0.88 [right], prox_clust(Tcell,Endo (H-H)) = 0.67 and 0.81). Although peripheral T cells were also found in the left slide, these were clearly recruited to the tumor site in contrast to the right scenario (SP2(Tcell,Tumor) = 4030 [left] vs 1555 [right]).

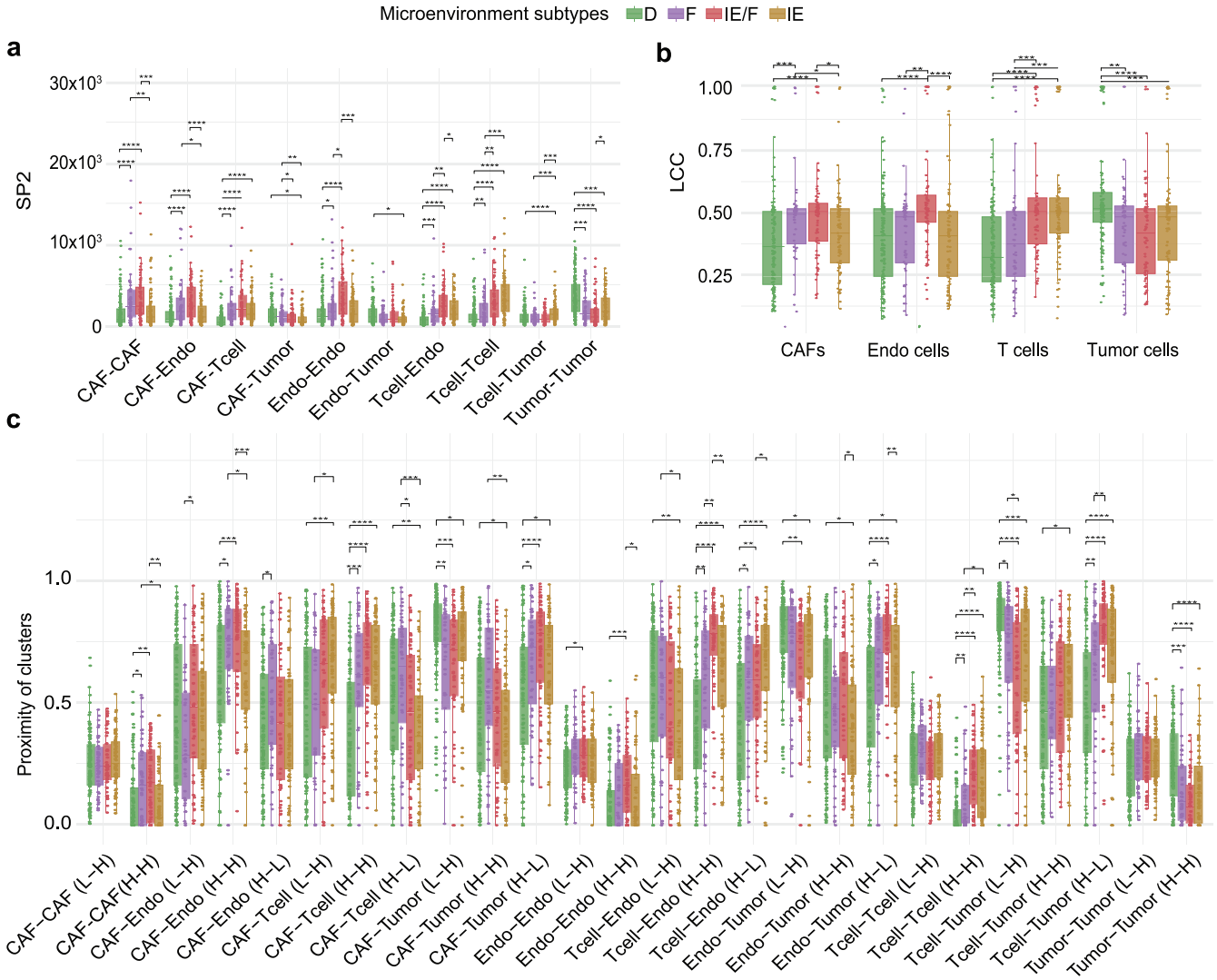

Fig. 4 | SPoTLIghT's spatial features distinguish microenvironment subtypes. **a** Number of shortest paths with maximum path length equal to two (SP2). **b** Size of the largest connected component (LCC). **c** Proximity between cell type-specific clusters classified as highly or lowly abundant. The distance between low clusters of the same cell type were not included. Wilcoxon rank-sum test; $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (**) and $p < 0.05$ (*). Non-significant comparisons are not shown.

Overall, these results support the notion that local spatial structures (e.g., T cell presence at the center or periphery of the tumor mass) cannot be directly captured from molecular data while they can be described by our derived spatial graph-based features.

Spatial graph-based interpretable features serve as prognostic biomarkers

We performed survival analysis to evaluate the association between the spatial features and overall survival adjusting for tumor staging. Using the Cox regression model, we observed that 30 out of 96 spatial features exhibited statistically significant association with overall survival in TCGA melanoma patients (Fig. 6a, Supplementary Table 2). Features related to the abundance of T cells (HR of mean_P(Tcell) = 0.07), their spread across the slide (HR of prox_clust(Tcell,Tcell (H-H)), frac_clust(Tcell) equal to 0.12, 0.44 respectively), their high connectivity (HR of LCC(Tcell) = 0.36) and their proximity with low tumor regions (HR of prox_clust(Tcell,Tumor (H-L) = 0.51) were positively associated with survival. Also the colocalization of CAFs and endothelial cells (HR of Coloc(CAF,Endo) = 0.49) and, perhaps surprisingly, the proximity of T cells and CAFs enriched clusters (HR of prox_clust(CAF, Tcell (H-H)) = 0.54) were associated with improved survival. On the other hand, presence and spread of tumor cells (HR of mean_P(Tumor), prox_clust(Tumor,Tumor (H,H)), frac_clust(Tumor), LCC(Tumor) equal to 173, 4.3, 2.9, 2.8 respectively), colocalization with endothelial cells (HR of Coloc(Endo, Tumor) = 3.38) and lack of T cells in tumor proximity (HR of prox_clust(Tcell,Tumor (L-H)) = 2.29) were associated with decreased survival.

Next, we wanted to validate the generalizability of SPoTLIghT's spatial features extraction pipeline by assessing if features identified as predictive of patient outcomes in the melanoma TCGA dataset, where these features were derived through transfer learning based on matched transcriptomics data, retain their prognostic ability in an independent melanoma cohort. The only publicly available melanoma cohort with H&E slides that we could retrieve was the Proteomic Tumor Analysis Consortium Cutaneous Melanoma (CPTAC-CM) cohort[39], consisting of 36 patients with paired histopathology images and information of 1-year survival (4 deceased, 32 alive) after surgery.

To focus specifically on survival prediction in early-stage patients, we trained a regularized (Elastic-net) logistic regression model on a subset of the TCGA melanoma cohort (see Methods), limiting the training data to stage I-III patients (15 deceased, 265 alive), as the CPTAC-CM cohort only contains early-stage patients. This allowed us to build a model tailored for

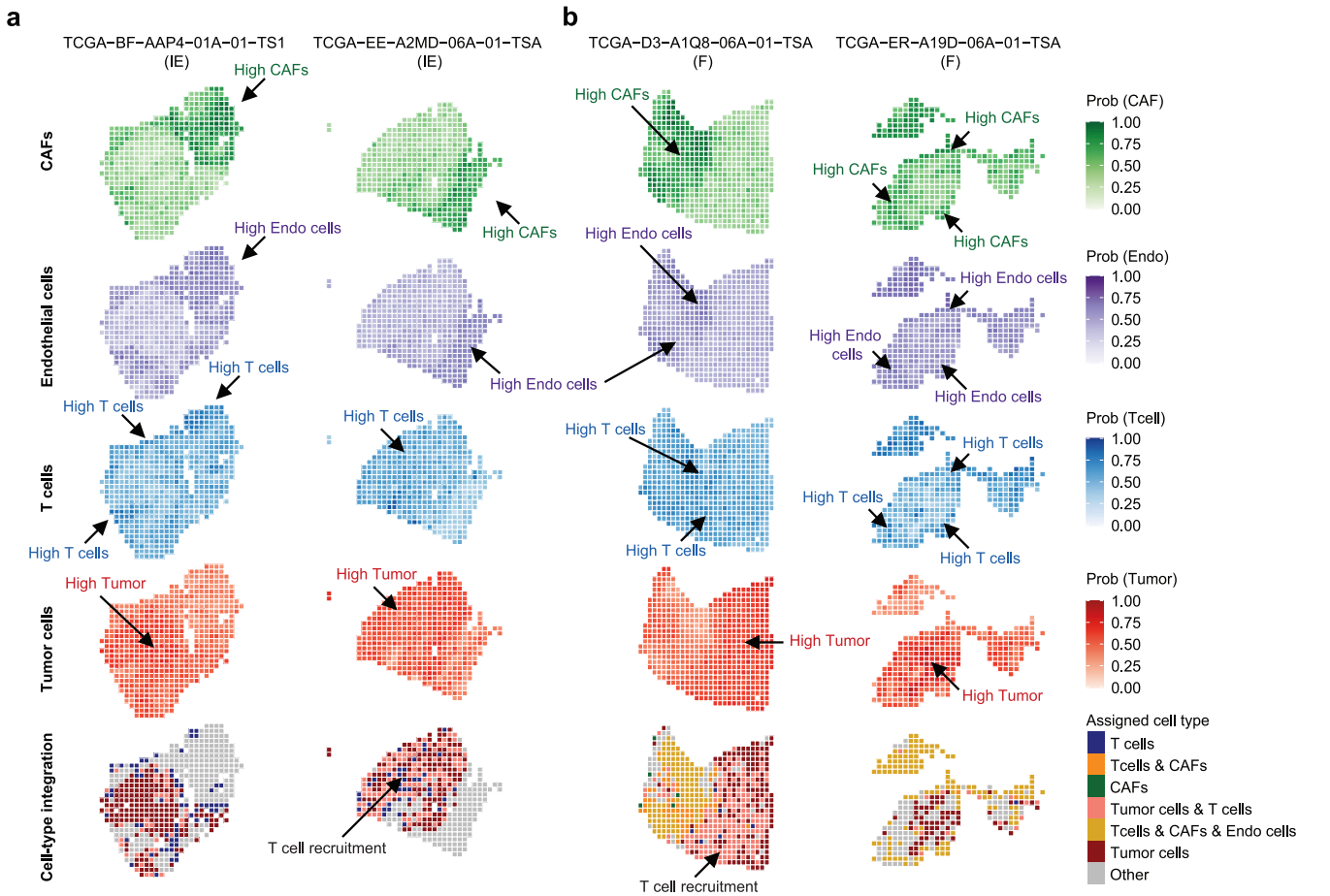

Fig. 5 | SPoTLIghT provides insights into local spatial structures. a Two slides classified as IE subtypes and **b** two slides classified as F subtypes which were predicted to have similar cell type abundance. In the left slide from **a** T cells can be seen surrounding the tumor whereas T cells are scattered around the tumor area in the right slide. In the left slide from **b** T cells are shown to be infiltrated into the tumor, while in the right slide T cells are surrounding the tumor and in close proximity with CAFs and endothelial cells.

predicting outcomes in early-stage patients, a distinct task from predicting survival for more advanced-stage cases. Deceased patients were assigned the positive class in order to estimate the probability of non-response to treatment.

The trained model was then applied out-of-the-box to spatial features extracted from the H&E slides of the validation CPTAC-CM cohort, for which no transcriptomics data were available. The model proved to be effective in predicting 1-year survival outcome in the validation cohort with area under the ROC curve (AUC) of 0.88 (Fig. 6b), accuracy of 72%, sensitivity of 75% and specificity of 72%. Additionally, we tested the model's performance when trained on all stages of patients from the TCGA dataset. The performance metrics remained comparable (see Supplementary Figure 7), demonstrating robustness.

To understand the rationale of the predictions, we analyzed the feature importance inherent in the model parameters identifying 13 robust features (using bootstrap, see Methods) that can be interpreted as biomarkers of 1-year survival status (Fig. 6c). The two most predictive positive biomarkers (associated with early death events) involve endothelial cells clusters, in particular the proximity of clusters enriched in endothelial and tumor cells (prox_clust(Endo,Tumor (H-H)) and the proximity of cluster with low and high abundance of endothelial cells (prox_clust(Endo,Endo (L-H)). In contrast, proximity of clusters enriched in endothelial cells (prox_clust(Endo,Endo (H-H)) and in T cells (prox_clust(Tcell,Tcell (H-H)) showed association with 1-year increased survival.

Finally, using survival curves we illustrated the individual prognostic effects of the five biomarkers which were both relevant for overall survival and selected in the 1-year predictive model (Fig. 6d). Overall, we can conclude that interconnection of T cells (LCC(Tcell)) and proximity of clusters abundant in T cells (prox_clust(Tcell,Tcell (H-H))) are strongly associated with good prognosis. Instead, communication between endothelial and tumor cells (mean_ND(Endo,Tumor)), proximity of tumor enriched and T cell depleted clusters (prox_clust(Tcell,Tumor (L-H))) and proximity of low T cells clusters (prox_clust(Tcell,Tcell (L-L))) are associated with decreased survival. Overall, these results revealed that the probability of survival increases with the presence of well-structured clusters of T cells, while it declines with T cells and endothelial cells constrained to lie distant from the tumor tissue.

Discussion

The nature of the TME involves multiple cell type populations of different functional states with distinct spatial organizations, altogether determining the progression of tumors and their response to therapy[40]. Such information could be derived from medical images although it might not be directly visible even to clinical experts. ML can support us in this task providing automatic approaches to extract latent knowledge from digital images of routinely collected histopathology slides.

In this study we employ ML techniques to explicitly enable the spatial characterization of the tumors and their microenvironment from H&E slides. Towards this goal, we derived spatial graph-based interpretable features by considering four relevant cell types of the TME and provided a new approach to quantitatively describe the spatial architecture of the melanoma

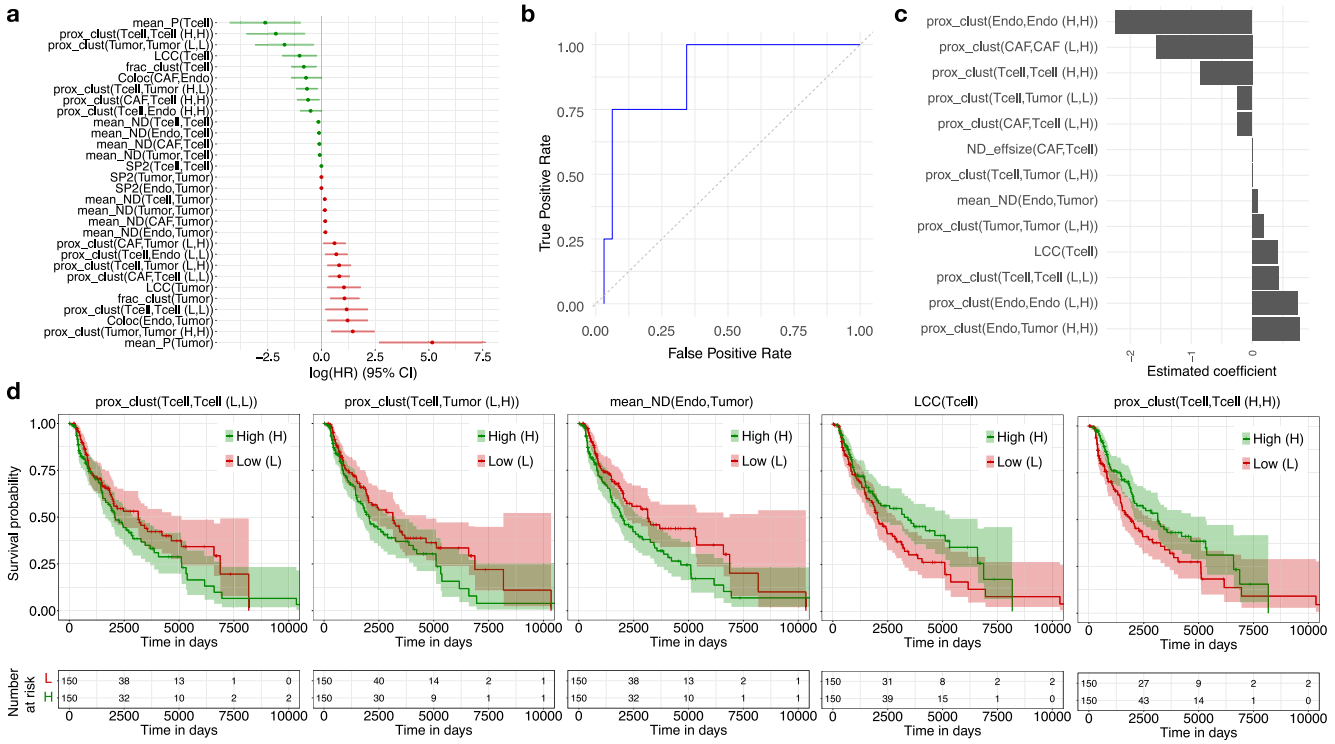

Fig. 6 | Spatial features predict survival outcomes in melanoma. a Hazard ratio (HR) estimate calculated based on univariate cox regression analysis of TCGA overall survival ($n = 320$; deceased = 153) showing the estimated HR (dots) and the 95% confidence interval (bars). Positive (red) and negative (green) HR means that the feature is associated with decreased and increased survival respectively. **b** Corresponding receiver operating characteristic (ROC) curve for prediction of 1-year vital status in the CPTAC melanoma cohort ($n = 36$ patients: deceased = 4, living = 32). **c** Bar plots showing the elastic-net logistic regression model coefficients for 13 recurrent features (non-zero in >50% of the bootstrap runs). **d** Kaplan–Meier (KM) curves displaying the estimated survival probability for two different groups (low and high) defined according to the values of corresponding spatial features.

microenvironment. This approach is implemented in a tool called SPoTLIghT, which has the potential to provide a better understanding of how local distribution of different cell types can affect survival and treatment outcomes.

We used transfer learning to quantify associations between information extracted from standard H&E pathology images and prior knowledge of cell type-specific expression signatures and copy-number profiles derived from molecular data. Employing multi-task transfer learning to learn simultaneously from multiple cell type signatures, helped us to reach more accurate cell type-specific models. In brief, our approach enabled us to gain a tile-level resolution of the quantification of multiple cell types and to capture cell type spatial local interactions which ultimately depicted patients' TME spatial architecture.

We validated our spatial maps by comparing them with tumor-infiltrating lymphocytes (TIL) maps derived from computational staining and with cell type maps derived from single cell spatial transcriptomics data. While constructing perfect spatial maps is beyond the scope of this paper—given the challenges of using unannotated H&E slides and bulk transcriptomics—our primary objective is to quantify spatial features that capture significant cellular patterns. These features have been validated through their expected associations with microenvironment subtypes: for instance, the co-localization of T cells with tumor cells in the immune-enriched (IE) subtype reflects established immune infiltration patterns, while the proximity of T cells to CAFs in the IE/F subtype indicates a potentially immunosuppressive environment[28]. Furthermore, the co-localization of endothelial cells with CAFs in fibrotic subtypes aligns with known roles of endothelial-to-mesenchymal transition in fibrosis[36]. Collectively, these findings emphasize the biological relevance of our spatial features and their ability to yield valuable insights into tumor microenvironments.

An advantage of our approach is that we do not rely on expensive imaging techniques or time-consuming manual curation for cell-type quantification. We can instead derive such information from pathology images and bulk RNA-seq which are routinely available in clinical settings. Importantly, while RNA-seq data were used for model training, our models can then be applied to derive our spatial features directly from the H&E slides also when no transcriptomics data is available. This is showcased in the paper by applying SPoTLIghT to derive spatial features from the H&E images of patients from the CPTAC-CM[39] public dataset. We showed that the derived spatial features could be used to build an accurate prognostic model of 1-year survival status (AUC = 0.88), identifying predictive positive and negative biomarkers. Interestingly, features capturing complex spatial patterns, such as proximity of cell type-specific clusters, proved more predictive than simpler features based on global quantification or simple co-localization. This confirmed the value of the graph embedding of the cell maps to derive graph properties.

A word of caution, however, should be applied when interpreting the prediction performance, as the test dataset was limited to only four positive cases. While the sensitivity of 75% suggests reasonable ability to identify true positives, further testing on larger datasets is necessary to confirm the robustness of these predictions.

As shown in our results, the derived features are able to describe the four immune/fibrotic microenvironment subtypes that are known to be informative of immunotherapy response[28]. For instance, the high abundance of endothelial cell clusters ($p < 0.05$) and their close proximity to 'high' T cell clusters show potential to discriminate between IE/F and IE subtypes (prox_clust(Tcell,Endo (H-H)) = 0.74 (IE/F) vs 0.62 (IE), $p < 0.01$). This is an interesting distinction because, while both subtypes are typically classified as immune-enriched tumors due to the high abundance of cytotoxic T cells, IE/F tumors are less prone to have favorable responses to immunotherapy as T cells are potentially kept far from tumor cells due to the presence of fibrotic tissue[35].

In this study, we validated our approach based on a TME classification that despite showing agreement with imaging features, it still emerges from molecular data which have inherent limitations on its ability to describe spatial relationships. To demonstrate that our approach overcomes this, we considered four example slides that were estimated with similar cell-type quantifications but that some of our features suggested having distinct spatial organizations. This was evident when looking at their spatial maps, supporting the idea that actual local spatial structural information can be captured by our graph-based features. This is an example of how our derived spatial features can improve patient stratification even in more challenging settings. It is also worth noticing that even though there was an imbalanced number of available melanoma patients across the TME subtypes (D: $n = 145$, F: $n = 57$, IE/F: $n = 74$, IE: $n = 85$), our features showed statistically significant differences between the subtypes involving a smaller number of patients.

A possible future application of SPoTLIghT will be prediction of immunotherapy response based on the spatial features. However, in clinical practice, H&E slides are often derived from formalin-fixed paraffin-embedded (FFPE) tissues, as FFPE better preserves morphology. Our approach is based on H&E slides from fresh-frozen (FF) tissues as these were used to extract RNA for molecular analysis[41]. In this paper, we successfully applied SPoTLIghT also to FFPE slides from both the TCGA and CPTAC cohorts, which were used for validation of the TIL maps and of the predictive features, respectively. These results demonstrate the potential generalizability of SPoTLIghT to FFPE slides. However, a systematic comparison between the features derived from FF and FFPE tissues is needed to ensure consistency in a clinical setting.

In this study we focused on the spatial relationships between four main cell types in the TME however the approach could be easily expanded to multiple cell types for which clear molecular signatures are available. Notably, this method does not rely on visually distinct cellular morphologies, paving the way for depicting a more complete landscape of tumor immune infiltration.

We envision that computational modeling of H&E slides providing such an understanding of the spatial relationships between multiple cell types in the TME can add considerable value to the pathology examination of tumor tissues. In fact, this knowledge can automatically lead to quantitative spatial biomarkers of tumors' immune phenotype from just standard pathology slides, opening new avenues for refining immuno-oncology treatment plans on a routine clinical basis.

Methods

TCGA tissue slides

FF and FFPE pathology images were downloaded for skin cutaneous melanoma (SKCM), from the GDC Data Portal (https://portal.gdc.cancer.gov/). Corresponding clinical data were downloaded from the supplementary files of the TCGA Pan-Cancer Clinical Data Resource[42]. Tissue slides were available for 476 patients. These patients sometimes had multiple samples and slides. Based on the available clinical data, a selection of slides was discarded. Primary tumor and metastatic samples were considered. Slides that missed information of expert pathology assessment of tumor purity were removed. Finally, all slides with a tumor purity < 80% were excluded, consistent with the threshold used for learning models of histopathological patterns[7]. This resulted in a final selection of 379 patients and 383 FF slides.

TCGA bulk RNA-sequencing data

Bulk RNA-seq data for 379 SKCM patients was downloaded via the Firehose tool from the BROAD Institute (https://gdac.broadinstitute.org), released January 28, 2016. The selection of patients was done according to the available TCGA tissue slides.

We extracted the gene expression data from "illuminahiseq_rnaseqv2-RSEM_genes" files. From these data, we used "raw_count" values as counts and we calculated transcripts per million (TPM) from "scaled_estimate" values multiplied by 1,000,000. We first removed those genes with a non-valid HGNC symbol and then we averaged the expression of those genes with identical HGNC symbols.

CPTAC melanoma cohort

The CPTAC melanoma cohort[39], comprising 79 FFPE slides from 36 patients with available information about 1-year survival after surgery, was used as an independent validation cohort. H&E images and clinical data were downloaded from The Cancer Imaging Archive (TCIA)[43].

Xenium datasets

Two publicly available FFPE melanoma Xenium datasets were downloaded from 10x Genomics (https://www.10xgenomics.com/datasets/human-skin-preview-data-xenium-human-skin-gene-expression-panel-add-on-1-standard and https://www.10xgenomics.com/datasets/human-skin-preview-data-xenium-human-skin-gene-expression-panel-1-standard). Both datasets consist of a single sample, where the first dataset was sequenced for a 282 gene panel (Xenium Human Skin Gene Expression Panel), while the other dataset also had an add-on panel of 100 genes. In addition, for both datasets the FFPE H&E slides, that were obtained after the Xenium run, were available.

Seurat was used to process the Xenium Ranger outputs. First, the cell-by-gene matrices were generated, followed by the removal of cells with less than 10 transcripts. The filtered matrices were normalized by Seurat's SCTransform[44]. After normalization, BANKSY[45] was used to identify clusters, and differential gene expression analysis was performed to obtain markers for each cluster. The average expression profiles and the differentially gene expression, together with known cell type markers were used to manually annotate the clusters. Clusters were only annotated with one of the cell types of interest (CAFs, T cells, endothelial cells and tumor cells), otherwise 'Undetermined' was used. Coordinates of the centroids for the annotated cells were converted to pixels by dividing it with the image scale factor 0.2125 (i.e. the pixel size).

TME immune phenotyping from molecular data

We used a previously defined classification of the TME to assign patients into different microenvironment subtypes: immune-enriched fibrotic (IE/F), immune-enriched non-fibrotic, fibrotic (F) and immune-depleted (D)[28]. This classification was made available for TCGA SKCM patients by the original work. Out of the 379 patients considered in our analysis, 361 patients had a corresponding TME classification.

Cell type quantification

We mathematically inferred the quantification of four cell types using cell type-specific prior knowledge of gene expression signatures and somatic DNA alterations as detailed below.

We quantified tumor cells based on three different scores. Tumor purity is inferred using ABSOLUTE[32], a method based on somatic DNA copy number data. Tumor purity is also calculated through ESTIMATE[33] which uses a fitted model of the sum of two gene signatures scores of stromal tissue and immune infiltration and the tumor purity quantification from ABSOLUTE. Tumor cell fractions were calculated using EPIC[46], which are relative to the overall composition of the tumor sample and are defined as 1 minus the total fraction of non-malignant cell types which include immune, stromal and endothelial cells.

To quantify T cells, we employ four different scores. Cytotoxic $CD8^+$ T cell fractions were computed using quanTIseq[47]. Effector cells signature scores were calculated from single-sample gene enrichment analysis (ssGSEA)[48] of genes related to effector T cells and NK cells[28]. Cytotoxic cells signature scores were computed using ssGSEA[26]. The immune score provided by ESTIMATE which is computed using ssGSEA of an immune signature involving 141 genes[33].

Four scores were used to estimate cancer associated fibroblasts (CAFs) abundance. The stromal score from ESTIMATE, which is defined by using

ssGSEA on 141 stromal genes[33]. A CAFs signature score calculated using ssGEA which is based on genes that are characteristic for CAFs and pericytes[28]. CAFs fractions were estimated using both MCP-counter[49] and EPIC[33].

We calculated three scores for endothelial cells quantification. Endothelial cell fractions were estimated first with EPIC[33] and then with xCell[50]. An endothelium signature score was calculated based on genes that were typical for endothelial cells[28].

EPIC, quanTIseq, MCP-counter and xCell perform in silico deconvolution of bulk gene expression based on cell type-specific gene signatures to estimate cell fractions. These methods were accessible through the immunedeconv[34] R package version 2.1.0. All mentioned gene signature scores were provided by the original publications.

Histopathological feature extraction

We used the method PC-CHiP[7] to quantify histopathological patterns of cancerous and normal tissues. Images were pre-processed as described in the Methods section of the original work. We used the retrained models that were made available by the authors (https://github.com/yufu2015/PC-CHiP) to extract 1536 features from the last hidden layer of their trained network.

Transfer learning: regularized multi-task linear regression (RMTLR)

RMTLR is defined, for N observations, by the following objective function:

$$\frac{1}{2N}\sum_{i=1}^{N} \| y_i - \beta_0 - x_i\beta\|_2^2 + \lambda \sum_{j=1}^{p}\left(\| \beta_j\|_2\right) \quad (1)$$

Where, for each observation i, y_i represents a q-dimensional row vector where each entry corresponds to a task (i.e. cell type score), and x_i is a row vector of size p, where each entry represents an observed feature. RMTLR aims to estimate a matrix β with p rows and q columns, whose rows represent the relation between one feature and all the tasks for a given cell type, and a vector β_0 of offsets (one for each task). The regularization term of RMTLR is a grouped version of lasso that aims at enforcing sparsity to entire rows of β[51,52]. The regularization term entails making two operations. First, the L2-norm of each row is taken, resulting in a vector of positive elements (the norm of each row). Then, the sum of the elements of this vector is made, and this latter operation corresponds to applying an L1-type regularization which promotes sparsity on the elements. Recalling that these elements are the norms of the rows of the original matrix, the result is to enforce group row-wise sparsity on β. In this way, the features corresponding to those rows of β that are set to zero do not contribute to the model. The strength of the regularization effect is tuned via the hyperparameter λ using 10-fold cross-validation.

We used RMTLR which was implemented as "linear_model.MultiTaskLasso" in the scikit-learn Python package version 1.2.0[53].

Multi-task transfer learning: model training based on TCGA data

Models were learned separately for each cell type. We trained RMTLR models using nested cross-validation (nCV) for model selection (5-fold outer loop) and for hyperparameter tuning (10-fold inner loop). Regularization and nCV allows better model generalization (avoiding overfitting on the original data). We randomly selected 50 tiles for each slide, and we split the folds at the patient-level. This ensured that all 50 tiles for a corresponding slide of a patient, or multiple slides of a patient, were found either all in the training or all in the test set and not split between the two. For each test set (outer fold), we first standardized the training set, and then we standardized the test set based on the mean and standard deviation of the training set. Model performances were evaluated using Spearman correlation.

Tile-level cell type quantification

For each slide, per-tile predictions were computed using the model learned when the slide was in the held-back validation set during nCV to prevent bias. Cell type-specific model predictions were computed by averaging across cell type-specific tasks. As the predicted values were z-scores, we transformed these into their associated cumulative probabilities. Probabilities were used as continuous values or binarized to assign each tile to one or more cell types if $P(i, c)>0.5$, where $P(i, c)$ was the predicted probability for a given tile i and cell type c.

Quantitative validation of T cell maps

We computed the T cells maps using SPoTLIghT for 334 FFPE TCGA-SKCM H&E slides for which Tumor-Infiltrating Lymphocytes (TIL) maps were available from Saltz et al.[16]. TIL assignments were provided as supplementary material from the original paper. We resized the SPoTLIghT T cell maps to match the resolution of Saltz TIL maps (100 pixels x 100 pixels, at 20x magnification) using nearest-neighbor interpolation. Finally, we computed both Jaccard and Dice similarity metrics to quantify the spatial overlap of SPoTLIghT and Saltz TIL maps. Note that both computational pipelines filter uniformative tiles differently, we only considered the tiles that were included in SPoTLIghT spatial maps.

Quantitative evaluation using Xenium in situ gene expression data

The two H&E slides from the Xenium datasets were used as inputs for SPoTLIghT to obtain the four cell type maps. For both slides, the gradient magnitude filter in the pipeline was relaxed to 10 to reduce the number of tiles that are discarded. The cell type maps were aligned with the image from the Xenium run using the provided 3×3 transformation matrix. After alignment, a grid was created with a bin size of 1024px, followed by assigning all tiles in the cell type maps to one of the bins. Then for each cell type, the mean predicted probability per bin was computed. Similarly, the annotated Xenium cells with their centroid coordinates were also binned using the same grid. Bins without any tiles and without any cells were removed. The Spearman correlation coefficient was then used to assess the relationship between the tile predictions and annotated cells. We computed this for the mean probability (SPoTLIghT) and number of cells per bin (Xenium).

Slide-level cell type quantification and tile-level cell type co-localization

For the slide-level quantification of cell types, we averaged the tile-level cell type predictions across all the tiles of each slide.

For tile-level quantification, the colocalization of each cell type was computed as the fraction of tiles that were labeled with two cell types based on the predicted probabilities. Note that tiles could be assigned to more than one label. For two cell types c_1 and c_2, the colocalization was calculated for each tile i across the N tiles of a given slide:

$$Colocalization = \frac{\sum_{i=1}^{N} P(i, c_1) \geq 0.5 \ \wedge \ P(i, c_2) \geq 0.5}{\sum_{i=1}^{N} P(i, c_1) \geq 0.5 \ \vee \ P(i, c_2) \geq 0.5} \quad (2)$$

Graph embedding

We embedded tile-level cell type quantifications into spatial graphs, where each node is represented by each tile of the image and its neighbors were assigned based on the surrounding tiles (up to 8 tiles per node). For this, we have used the Networkx Python package version 3.0[54].

Spatially constrained hierarchical clustering (SCHC)

To identify cell type clusters in our spatial graphs, we employed Spatially Constrained Hierarchical Clustering (SCHC) which considers both the predicted cell type abundances and the spatial position of the tiles.

Essentially, SCHC is a hierarchical clustering method but uses a connectivity matrix to ensure that only adjacent clusters are merged, resulting in clusters where its members are connected. The connectivity matrix, the cell type predictions and the spatial coordinates of the tiles were used to apply agglomerative clustering as implemented in the scikit-learn Python package version 1.2.0[53]. This form of clustering uses a bottom-up approach, starting with single samples as clusters and through a linkage criterion and applying the connectivity constraints, merges clusters. The Ward linkage criterion was used to minimize the within-cluster variance of the clusters that are merged. Clusters were classified as 'high' or 'low' abundance cell type (cell type-specific approach). Distances between clusters were computed using the Euclidean distance.

We also computed pan-cell type clusters considering the predictions of all cell types simultaneously. This means that nodes were assigned to the same cluster if they were in geographical proximity, and they had a similar distribution of cell types. Clusters were classified directly as belonging to one or more 'cell type' labels when the average prediction for that cell type(s) in the cluster was higher than the average computed across all clusters of all slides. The fraction of 'cell type' clusters and the proximity between 'cell type' clusters was computed analogously to the cell type-specific approach (see Table 1).

The results obtained were consistent with those from the cell type-specific approach (Supplementary Fig. 8) therefore we refer only to the latter one in the main text.

Spatial graph-based interpretable features

We reported the details of the derivation and meaning of these features in Table 1.

Predictive model

We trained a regularized logistic regression model on the TCGA data using the spatial features as predictors and the 1-year survival outcome as response variable. Since the CPTAC test dataset does not include stage IV patients, we only included stage I-III patients in the TCGA training dataset, this resulted in 265 alive and 15 deceased patients. Patients were considered 'alive' at 1 year when death_days_to ≥ 365 or when last_contact_days_to ≥ 365 and 'deceased' when death_days_to< 365. Aiming to predict the probability of non-response to treatment, we considered the deceased patients as positive class. The model was trained using Elastic net regression using the *glmnet* R package[55], where the hyperparameter alpha (representing the balance between L1 and L2 regularization) was set to 0.5 and the hyperparameter lambda (determining the balance between fit to the data and model complexity) was tuned using 5-fold cross-validation. Lambda was set to the value resulting in the lowest misclassification error in the cross-validation and used to retrain the logistic regression model on the training set. To compensate for the imbalanced dataset, model training was performed assigning different weights to the observation belonging to the positive and negative class, where the class weight was assigned as 1-(# of class members)/(# total observations). The model was retrained 100 times on bootstrapped data (i.e., resampling with replacement). The 100 bootstrapped models were then applied to predict 1-year survival in the independent CPTAC cohort (4 alive, 32 deceased) using as input only the spatial features computed on the CPTAC images. Final model predictions were derived as the median of the 100 predictions for each patient and used to compute the evaluation metrics. AUC was computed using the *pROC* R package[56]. Sensitivity, specificity and (balanced) accuracy were computed using the *caret* R package[57] using the default threshold at 0.5. The robust set of predictive features was derived as the features with non-zero coefficients in at least 50% of the bootstrap runs and their final coefficient values were computed as the median across all bootstrap runs.

Statistical analysis

We used Wilcoxon rank-sum test to compare the spatial features distribution across microenvironment subtypes. The effect size of the ND between the true and the simulated slide was calculated using Cohen's d.

Survival analysis

Univariate Cox's proportional hazard regression models were fitted including overall survival as dependent variable, the spatial features as fixed factors and using the tumor stage as covariate due to its potential confounding role. Patients were split into low and high at the median. We considered features as significant only if the 95% confidence intervals around the hazard ratio (HR) do not include the one. Survival curves were estimated using the Kaplan–Meier estimator. The univariate Cox regression analysis was performed using the survival R package[58], and survival curves were plotted using the survminer R package.

Data availability

All data used were publicly available for research use. The FF and FFPE slides of the TCGA-SKCM cohort were downloaded from https://portal.gdc.cancer.gov/. The CPTAC melanoma FFPE slides were downloaded from the NIH cancer imaging archive (https://www.cancerimagingarchive.net/collection/cptac-cm/). The FFPE melanoma Xenium datasets were downloaded from 10x Genomics (https://www.10xgenomics.com/datasets/human-skin-preview-data-xenium-human-skin-gene-expression-panel-add-on-1-standard and https://www.10xgenomics.com/datasets/human-skin-preview-data-xenium-human-skin-gene-expression-panel-1-standard).

Code availability

The source code required to reproduce the analyzes is available from GitHub: https://github.com/SysBioOncology/spatial_features_manuscript. SPoTLIghT has been implemented as a Nextflow pipeline to provide an automated workflow to derive the spatial graph-based interpretable features directly from H&E tissue images. This is available from Github: https://github.com/SysBioOncology/SPoTLIghT.

Received: 19 April 2024; Accepted: 28 October 2024;
Published online: 07 November 2024

References

1. D'Abbronzo, G. & Franco, R. The changing role of the pathologist in the era of targeted therapy in personalized medicine. *Expert Rev. Precis. Med. Drug Dev.* **6**, 295–297 (2021).
2. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
3. Cifci, D., Veldhuizen, G. P., Foersch, S. & Kather, J. N. AI in computational pathology of cancer: improving diagnostic workflows and clinical outcomes? *Annu. Rev. Cancer Biol.* **7**, 57–71 (2023).
4. Kather, J. N. et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
5. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
6. Kather, J. N. et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* **1**, 789–799 (2020).
7. Fu, Y. et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* **1**, 800–810 (2020).
8. Xu, Z. et al. Deep learning predicts chromosomal instability from histopathology images. *iScience* **24**, 102394 (2021).
9. Ghaffari Laleh, N. et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Med. Image Anal.* **79**, 102474 (2022).
10. Fridman, W. H., Pagès, F., Sautès-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306 (2012).
11. Fridman, W. H., Zitvogel, L., Sautès-Fridman, C. & Kroemer, G. The immune contexture in cancer prognosis and treatment. *Nat. Rev. Clin. Oncol.* **14**, 717–734 (2017).

12. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
13. Duanmu, H. et al. A spatial attention guided deep learning system for prediction of pathological complete response using breast cancer histopathology images. *Bioinformatics* **38**, 4605–4612 (2022).
14. Lu, C. et al. Feature-driven local cell graph (FLocK): New computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers. *Med. Image Anal.* **68**, 101903 (2021).
15. Diao, J. A. et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat. Commun.* **12**, 1613 (2021).
16. Saltz, J. et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**, 181–193.e7 (2018).
17. Wang, S. et al. Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer Res* **80**, 2056–2066 (2020).
18. Kather, J. N. et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* **16**, e1002730 (2019).
19. Failmezger, H. et al. Topological tumor graphs: a graph-based spatial model to infer stromal recruitment for immunosuppression in melanoma histology. *Cancer Res* **80**, 1199–1209 (2020).
20. Corredor, G. 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.* **25**, 1526–1534 (2019).
21. Perera, R. et al. Annotation-efficient deep learning for breast cancer whole-slide image classification using tumour infiltrating lymphocytes and slide-level labels. *Commun Eng* **3** (2024).
22. Sarhadi, V. K. & Armengol, G. Molecular Biomarkers in Cancer. *Biomolecules* **12** (2022).
23. Sturm, G., Finotello, F. & List, M. In silico cell-type deconvolution methods in cancer immunotherapy. *Methods Mol. Biol.* **2120**, 213–222 (2020).
24. Tran, K. A. et al. Performance of tumour microenvironment deconvolution methods in breast cancer using single-cell simulated bulk mixtures. *Nat. Commun.* **14**, 5758 (2023).
25. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
26. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
27. Tamborero, D. et al. A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin. Cancer Res.* **24**, 3717–3728 (2018).
28. Bagaev, A. et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39**, 845–865.e7 (2021).
29. Levy-Jurgenson, A., Tekpli, X., Kristensen, V. N. & Yakhini, Z. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. *Sci. Rep.* **10**, 18802 (2020).
30. El Nahhas, O. S. M. et al. Regression-based deep-learning predicts molecular biomarkers from pathology slides. *Nat. Commun.* **15**, 1253 (2024).
31. Cancer Genome Atlas Research Network. et al. The cancer genome atlas Pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
32. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
33. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
34. Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, i436–i445 (2019).
35. Galluzzi, L., Chan, T. A., Kroemer, G., Wolchok, J. D. & López-Soto, A. The hallmarks of successful anticancer immunotherapy. *Sci. Transl. Med*. **10** (2018).
36. Choi, K. J., Nam, J.-K., Kim, J.-H., Choi, S.-H. & Lee, Y.-J. Endothelial-to-mesenchymal transition in anticancer therapy and normal tissue damage. *Exp. Mol. Med.* **52**, 781–792 (2020).
37. Asrir, A. et al. Tumor-associated high endothelial venules mediate lymphocyte entry into tumors and predict response to PD-1 plus CTLA-4 combination immunotherapy. *Cancer Cell* **40**, 318–334.e9 (2022).
38. Leung, C. S. et al. Cancer-associated fibroblasts regulate endothelial adhesion protein LPP to promote ovarian cancer chemoresistance. *J. Clin. Invest.* **128**, 589–606 (2018).
39. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Cutaneous Melanoma collection (CPTAC-CM). The Cancer Imaging Archive https://doi.org/10.7937/K9/TCIA.2018.ODU24GZE (2018).
40. Lapuente-Santana, Ó. & Eduati, F. Toward systems biomarkers of response to immune checkpoint blockers. *Front. Oncol.* **10**, 1027 (2020).
41. Cooper, L. A. et al. PanCancer insights from the cancer genome atlas: the pathologist's perspective. *J. Pathol.* **244**, 512–524 (2018).
42. Liu, J. et al. An integrated TCGA Pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416.e11 (2018).
43. Clark, K. et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* **26**, 1045–1057 (2013).
44. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
45. Singhal, V. et al. BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat. Genet.* **56**, 431–441 (2024).
46. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* **6** (2017).
47. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* **11**, 34 (2019).
48. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
49. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
50. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
51. Yuan, M. & Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 49–67 (2006).
52. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
53. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
54. *Exploring Network Structure, Dynamics, and Function Using Networkx* (2008).
55. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
56. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* **12**, 1–8 (2011).
57. Kuhn, M. Building Predictive Models inRUsing thecaretPackage. *J. Stat. Softw*. **28** (2008).

58. Therneau, T. M. Survival Analysis [R package survival version 3.5-8] (2024).

Acknowledgements
The results shown here are in part based upon data generated by TCGA Research Network (https://www.cancer.gov/tcga) and by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC; https://proteomics.cancer.gov/programs/cptac). This work was supported by the Netherlands Organization for Scientific Research (NWO) Gravitation programme IMAGINE! (project number 24.005.009).

Author contributions
F.E. Conceptualization, supervision, data analysis and interpretation. O.L.S Conceptualization, algorithm development, data analysis and interpretation, validation analysis. J.K. Algorithm development, data analysis and interpretation, validation analysis. All authors contributed to writing the original draft and to review the manuscript. All authors have read and approved the manuscript. O.L.S. and J.K. contributed equally to this work.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-024-00749-w.

Correspondence and requests for materials should be addressed to Federica Eduati.