





OPEN

MOSBY enables multi-omic inference and spatial biomarker discovery from whole slide images

Yasin Şenbabaoğlu , Vignesh Prabhakar, Aminollah Khormali, Jeff Eastham, Evan Liu, Elisa Warner, Barzin Nabet, Minu Srivastava, Marcus Ballinger & Kai Liu 

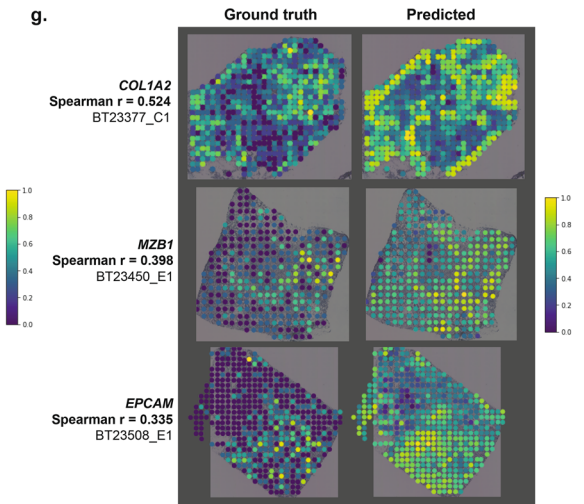
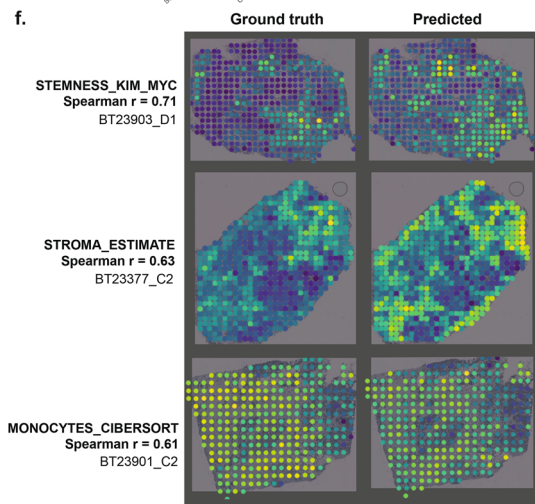
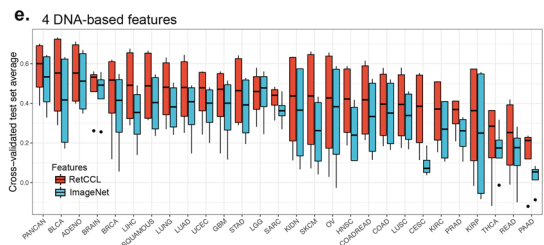
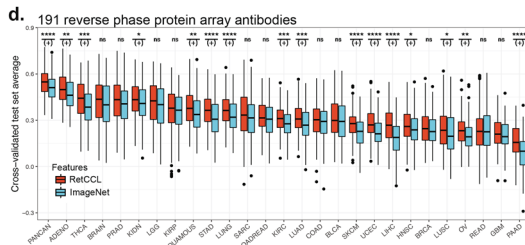
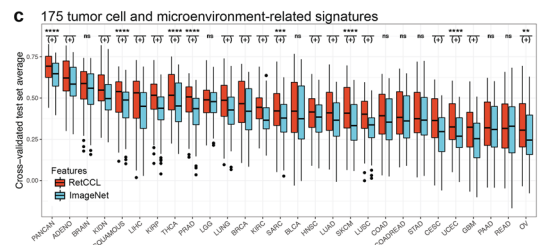
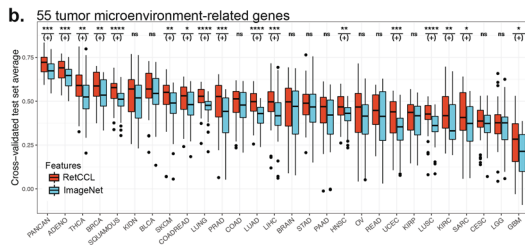
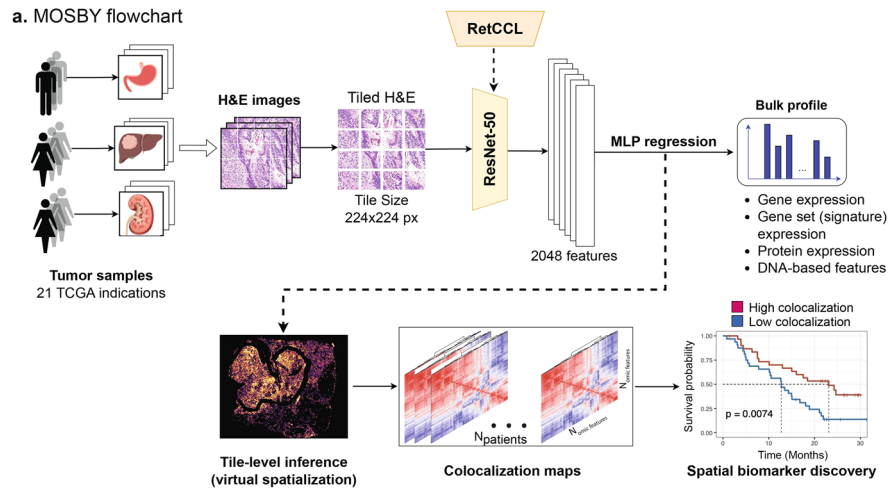
The utility of deep neural nets has been demonstrated for mapping hematoxylin-and-eosin (H&E) stained image features to expression of individual genes. However, these models have not been employed to discover clinically relevant spatial biomarkers. Here we develop MOSBY (Multi-Omic translation of whole slide images for Spatial Biomarker discovery) that leverages contrastive self-supervised pretraining to extract improved H&E whole slide images features, learns a mapping between image and bulk omic profiles (RNA, DNA, and protein), and utilizes tile-level information to discover spatial biomarkers. We validate MOSBY gene and gene set predictions with spatial transcriptomic and serially-sectioned CD8 IHC image data. We demonstrate that MOSBY-inferred colocalization features have survival-predictive power orthogonal to gene expression, and enable concordance indices highly competitive with survival-trained multimodal networks. We identify and validate (1) an ER stress-associated colocalization feature as a chemotherapy-specific risk factor in lung adenocarcinoma, and (2) the colocalization of T effector cell vs cysteine signatures as a negative prognostic factor in multiple cancer indications. The discovery of clinically relevant biologically interpretable spatial biomarkers showcases the utility of the model in unraveling novel insights in cancer biology as well as informing clinical decision-making.

The developments in high-throughput RNA, DNA, and protein assays as well as in digital pathology have enabled multiple different high-resolution perspectives on a given tissue. Sampled from the same tissue, these data modalities are inherently linked, due to arising from highly similar cell populations and potentially the same biological state. Successfully inferring this common biology is key for relating different data modalities, and predicting one modality from another. Deep neural nets have emerged as an effective and flexible framework to represent the common biology underlying different readouts, opening the way to ‘translate’ any data modality into another^{1–5}. A high utility use case for this translation task involves digital pathology where relatively easy-to-acquire hematoxylin and eosin (H&E) stained whole slide images (WSI) can be used to infer high-throughput molecular data that is time consuming or expensive to acquire.

The success of inferring the joint biology (i.e. underlying latent distribution) between WSIs and molecular data relies heavily on extracting informative features from WSIs. Contrastive self-supervised learning has made a breakthrough in computer vision by learning high quality image representations in no-annotation settings as shown in curated benchmark datasets such as ImageNet^{6,7}. Contrastive learning aims to discriminate between positive and negative images where positive pairs are obtained from augmentations from the same image and negative pairs are augmentations from different images⁸. Contrastive self-supervised learning has also shown promise in identifying cancer-specific morphological features in large-scale and heterogeneous histology datasets and training feature extractor models that outperform ImageNet-pretrained networks^{9–11}. However, the superior performance of self-supervised features has largely been demonstrated for classification tasks, with limited attention on regression tasks. For instance, Wang et al.⁸ developed an oncology-focused feature extractor RetCCL that was trained in more than 34,000 histology images and achieved state-of-the-art performance in cancer subtype classification. Regression tasks, such as inferring gene expression levels from WSI features, have been addressed with multi-output regression networks^{1,12}; yet, self-supervised learning-based image features remain largely underexplored in this context¹³.

Among molecular data modalities, the inference of bulk and single-cell transcriptomics from H&E WSI features has received considerable attention^{1,2,12}; however, an underexplored extension of this question is whether mapping the same image features to proteomic data holds greater promise due to proteins being more proximally associated with phenotype. The full power and limitations of image to molecular data ‘translations’ will become

Genentech, Inc., South San Francisco, CA 94080, USA. ✉email: shenbaba@gmail.com; vincentliuk@gmail.com



more evident as these deep learning models are expanded to learn multiple -omic data modalities such as transcriptomics, proteomics, metabolomics, and whole exome or SNP array-based DNA measurements.

Here, we present the MOSBY (Multi-Omic translation of whole slide images for Spatial Biomarker discoveryY) model that employs contrastive self-supervised learning-based (RetCCL) features from H&E images to infer high-throughput molecular data such as transcriptomics (single genes or gene signatures), proteomics, and selected DNA-based features (e.g. cancer DNA fraction, average DNA methylation). MOSBY adopts a multiple instance learning approach where it partitions gigapixel WSIs to small tiles, makes tile-level predictions as an intermediate step, and then aggregates those to obtain slide-level predictions. At the inference stage, MOSBY enables in silico spatialization of bulk omic profiles by reconstructing tiles from a given WSI. Correlation between two omic features across all tiles of a slide reveals patient-specific colocalization or spatial exclusion patterns.

Figure 1. (a) MOSBY flowchart: H&E-stained whole slide images were downloaded from TCGA and partitioned into tiles. Image features were extracted using RetCCL-pretrained weights and the ResNet-50 architecture. A 2-layer perceptron was trained to learn the mapping from image features to bulk omic profiles. A by-product of the process, tile level predictions were utilized to achieve virtual spatialization of omic profiles. Pairwise correlation of omic features resulted in a colocalization map for each patient. Colocalization maps were then flattened, and used as covariates in survival analysis with the goal of spatial biomarker discovery. (b,c,d,e) Test set Spearman correlation between omic feature predictions and ground truth, averaged over 5 cross-validation folds. Each data point is an omic feature, and the x-axis shows different single- and multi-indication runs. Red and turquoise boxes indicate results from RetCCL vs. ImageNet-pretrained ResNet-50 architecture respectively. Results are shown for (b) 55 TME-related genes, (c) 175 tumor cell and microenvironment-related signatures, (d) 191 reverse phase protein array antibodies, and (e) 4 DNA-based features. Mann–Whitney hypothesis tests for (b), (c), (d) were implemented with the `ggpubr compare_means` function to compare RetCCL and ImageNet based results (Hypothesis tests were not performed for (e) due to the small number of features, $N=4$). Significance levels are provided in the figure, original P -values are provided in Supplementary Table 5. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns: $p > 0.05$. Lower and upper hinges in box plots correspond to first and third quartiles, while whiskers extend to 1.5 times interquartile range. (f–g) Concordance between ST ground truth and MOSBY-predicted levels of select (f) signatures, (g) genes. Normalized expression levels in ground truth and MOSBY-predicted levels were both mapped to the [0–1] interval to enable comparison.

Expanded to the cohort, these spatial features allow identification of clinically relevant spatial biomarkers. We demonstrate the performance of MOSBY in pan-cancer TCGA data, and spatially validate results with (1) spatially resolved transcriptomic data in breast cancer, and (2) serially-sectioned CD8 IHC images in urothelial cancer. The selected features for signature models aimed to encompass a broad spectrum of biology important for the tumor microenvironment (TME) such as immune and stroma cell, oncogenic pathway, and metabolic signatures. For gene models, the feature set focused on TME cell type marker genes as cell types are more likely to have morphological differences that can be captured with computer vision models. We finally utilize signature models to showcase the derivation of biologically interpretable colocalization features consistently associated with risk and disease state in human cancers.

Results

We developed the MOSBY model that learned a mapping from RetCCL (contrastive self-supervised learning)-based WSI features to bulk transcriptome, proteome, whole exome, SNP-array, and DNA methylome based profiles (Fig. 1a). We limited our study to 21 TCGA indications¹⁴ with at least 200 paired RNA-seq and H&E whole slide images, resulting in a total of 12,592, 10,192, and 12,090 images matching with transcriptome, proteome and DNA-based data respectively (breakdown by indication in Supplementary Table 1). Analyzed transcriptomic features consisted of 55 TME-related genes^{15,16} (Supplementary Table 2) and 175 gene sets that covered tumor-related processes^{17–32}, metabolic pathways³³, and TME cell type or process signatures^{34–37} (Supplementary Table 3). The proteome model involved 191 total and phosphoprotein antibodies from the TCGA reverse-phase protein array (RPPA) panel that focused on tumor-intrinsic and oncogenic processes³⁸. Tested DNA-based features were limited to tumor purity, cancer DNA fraction, subclonal genome fraction, and average DNA methylation (all continuous measures bound to the interval [0,1]).

In addition to single indication models, we also trained MOSBY in pan-tissue, pan-cancer, pan-squamous and pan-adenocarcinoma settings. Pan-tissue models consisted of LUNG (lung adenocarcinoma and squamous cell carcinoma), KIDNEY (clear cell and papillary renal cell carcinoma), BRAIN (glioblastoma and low grade glioma), and COADREAD (colon and rectal adenocarcinoma). The pan-adenocarcinoma (ADENO) model consisted of pancreatic, lung, stomach, colon, rectal, prostate, and ovarian adenocarcinomas; while the pan-SQUAMOUS model included squamous cell carcinomas of the lung, cervix, and head and neck. The ADENO and SQUAMOUS models allowed us to investigate whether histological similarities beyond tissue architecture enabled MOSBY to learn a better mapping from WSI features to multi-omic data.

Similar to the HE2RNA¹² model, a 2-layer perceptron was adopted as the multi-output regression network that mapped image features to omic variables, and a maximum of 8000 image tiles per slide were used. Following deviations from the HE2RNA model were implemented in TCGA: (1) Separate models were trained for gene, signature, protein, and DNA variables; (2) image tiles were randomly selected from each WSI to capture an unbiased representation of the entire slide, and were all used in training without clustering; (3) batch-normalized transcriptomic and proteomic data were used as ground truth to enable across-indication comparisons with resulting MOSBY predictions; (4) tile-level predictions of omic features were aggregated by averaging all available tiles to obtain slide-level predictions.

Multiple model architectures (i.e. number of perceptron layers) and average pooling approaches (varying number of averaged tiles in {100,500,1000,all}) were explored in independent urothelial carcinoma datasets IMvigor210³⁹ and IMvigor211⁴⁰. Results showed that average pooling across all available tiles led to more accurate test set predictions in this setting (Supplementary Fig. 1a, Supplementary Table 4a), however, a 1-layer perceptron can in some cases be preferable to 2-layer perceptron in MOSBY (Supplementary Fig. 1b, Supplementary Table 4b).

Contrastive self-supervised pretraining benefits prediction of omic data from H&E whole slide images

The RetCCL feature extractor utilized TCGA H&E images during contrastive training, however it was not supervised by gene expression information. Hence, there are no a priori guarantees for RetCCL-based image features to predict gene expression with higher accuracy than ImageNet-based features. Thus, we investigated the performance differences between MOSBY models trained with RetCCL- or ImageNet-based image features. Models were trained with fivefold cross-validation (80/20 percent training/test set split). A random one fifth of the training set was also allocated as validation set to determine an early stopping criterion using the Spearman correlation between slide-level model prediction and ground truth omic data. Spearman correlation was preferred to mitigate the effects of outlier samples in relatively smaller cohorts. Correlation coefficients from test sets were subsequently averaged across five folds to obtain the final performance score for a given feature.

RetCCL-based image features consistently led to higher cross-validated test set averages in all four omic data types compared to features extracted with ImageNet-pretrained ResNet-50 architecture (Fig. 1b–e, Supplementary Table 5). The pan-cancer model (PANCAN) with RetCCL-based features achieved the highest performance for all tested data types with median cross-validated Spearman correlation of 0.722 in single genes (0.673 with ImageNet-features) (Fig. 1b), 0.693 in signatures (0.647 with ImageNet-features) (Fig. 1c), 0.549 in proteomic data (0.512 with ImageNet-features) (Fig. 1d), and 0.6 in DNA features (0.533 with ImageNet-features) (Fig. 1e). In terms of single indication models, the thyroid cancer model (THCA) achieved best performance for single gene and protein expression data sets with RetCCL-based features. Median cross-validated Spearman correlations in these models were 0.59 vs 0.524 for single gene, and 0.443 vs 0.385 for protein expression data with RetCCL and ImageNet-based features respectively (Fig. 1b,d). The single indication models achieving best performance for signature and DNA data were the liver and bladder cancer models respectively (LIHC and BLCA) (Fig. 1c,e). Across tested signatures, the LIHC model showed a median cross-validated Spearman correlation of 0.532 with RetCCL-based vs 0.45 with ImageNet-based features. For tested DNA features, the BLCA model achieved a median cross-validated Spearman correlation of 0.553 with RetCCL-based vs 0.417 with ImageNet-based features.

To address the information leakage possibility in TCGA, the RetCCL vs ImageNet comparison was repeated in urothelial carcinoma datasets IMvigor210³⁹ and IMvigor211⁴⁰. Gene and signature models were trained in the same fashion as in TCGA, and cross-validated averages of test set correlations were compared between RetCCL- and ImageNet-based models. In both gene and signature models, MOSBY slide level predictions were highly significantly better with RetCCL-based features (Supplementary Fig. 1c, Supplementary Table 6a,b). Taken together, these results indicated that contrastive self-supervised pretraining in large-scale histology datasets benefits prediction of multiomic data from H&E-stained whole slide images.

Further inspection of multi-indication models in TCGA showed that cohort differences between indications may lead to exaggerated performance estimates, and a pan-cancer model has limited utility in a clinical deployment scenario (Supplementary Notes, Supplementary Fig. 2).

MOSBY tile level predictions are validated with spatially resolved ground truth

After showing the benefit of contrastive self-supervised training for ‘slide level’ predictions, we validated MOSBY ‘tile level’ predictions with spatially resolved transcriptomic data in breast cancer, and CD8 immunohistochemistry (IHC) whole slide image data in urothelial cancer. For both tasks, inference models were trained with data from 80% of patients, with the remaining 20% used as the validation set to determine an early stopping criterion.

The inference model for the former validation task was trained in TCGA (BRCA cohort, N = 1576 WSIs), and subsequently deployed on H&E image tiles from a publicly available spatially resolved transcriptomic breast cancer dataset⁴¹ (referred to as ST from here on, N = 68 WSIs). Image tiles were centered around ST spot coordinates to enable one-to-one comparison between spot level ground truth data and tile level model predictions (Methods). For individual slides, the MOSBY signature model predictions showed the highest concordance with ground truth for ‘poor differentiation’ (Stemness_Kim_Myc, Spearman $r = 0.71$) and stromal features (Stroma_Estimate, Spearman $r = 0.63$) (Fig. 1f, Supplementary Fig. 3a). Across 68 slides, concordance was highest for a monocyte signature with a median Spearman correlation of 0.238 (Supplementary Fig. 3a), and a maximum of 0.611 (Fig. 1f). This large variation across slides was observed for all tested signatures suggesting that the quality of spatially resolved data is critical in validating MOSBY predictions. Of note, CD8 T cell infiltration and proliferation-related model predictions also showed strong concordance with ground truth, demonstrating the variety in phenotypes captured successfully by the model (Supplementary Fig. 3b).

Spot level concordance for single gene expression values was overall lower than that for signatures (Supplementary Fig. 3d), potentially driven by the large degree of zero reads (i.e. dropouts) in ground truth data (Supplementary Fig. 3e). However, model predictions for genes associated with stroma, plasma cells, and epithelial features showed good performance for individual slides (COL1A2, MZB1, EPCAM respectively) (Fig. 1g). CD68, a myeloid marker, showed the highest median concordance in gene models (Spearman $r = 0.127$), which was low in magnitude but consistent with the high concordance of myeloid cells in signature models (Supplementary Fig. 3d).

The gene–gene and signature–signature correlation structure in spatial transcriptomics data also showed that MOSBY predictions had stronger concordance with signature level ground truth data, again highlighting that computing gene signature scores is an effective strategy to deal with the dropout issue in spatially resolved transcriptomic datasets (Supplementary Notes, Supplementary Fig. 3).

MOSBY tile-level predictions were further validated with CD8 antibody-stained IHC image data (Methods). Inference models were trained with paired H&E and RNA-seq data from urothelial carcinoma trials IMvigor210³⁹ and IMvigor211⁴⁰ (Methods, N = 1460). Comparing CD8-associated gene and signature model predictions with

IHC ground truth showed that MOSBY successfully learned a CD8 T cell-specific representation on H&E images (Figs. 2 and 3, Supplementary Notes).

MOSBY predicts stroma, immune, and proliferation features with highest accuracy

In TCGA indications, cross-validated test set performance in signature and protein models was used to determine biological features with the highest prediction accuracy, and those with 25 highest performances are highlighted in Fig. 4a,d respectively. The best-performing signatures ESTIMATE Stroma and ESTIMATE Immune had median correlations 0.627 and 0.616 across all tested indications (Fig. 4a). In randomly split test sets, the Spearman correlation of ESTIMATE Stroma and ESTIMATE Immune signatures reached as high as 0.787 and 0.781 in skin cutaneous melanoma and thyroid cancer respectively (Fig. 4b). Other best-performing signatures were again largely enriched in stroma and immune features as well as general mesenchymal characteristics (such as Hallmark EMT and Taube et al. mesenchymal signatures). (Fig. 4a). Of note, particular signatures known to play an important role in specific cancer indications also showed strong prediction accuracy in those pertinent settings. For instance, the Hallmark Angiogenesis signature had Spearman $r=0.746$ in the liver cancer model (LIHC), and a fatty acid elongation signature had Spearman $r=0.741$ in the low grade glioma setting (LGG) (Fig. 4b). After adjusting for multiple hypothesis testing with false discovery rate (adj. $p < 0.05$), 14 out of 21 tested indications showed significant performance (as measured by Spearman r) for more than 90% of tested signatures (total $N=175$ signatures) (Fig. 4c).

In contrast with our signature set that represented both tumor-intrinsic pathways and cell populations in the TME, the TCGA RPPA antibody set was heavily focused on tumor-intrinsic pathways. Therefore, the best-performing proteins had a diversity of representation from tumor-intrinsic characteristics, such as proliferation (Cyclin-B1, FOXM1), DNA repair (MSH6, PARP1), and apoptosis (cleaved Caspase7) (Fig. 4d). MSH6 and Cyclin B1 were the two best-performing proteins that were tested in at least 20 indications (Fig. 4d). These proteins had median Spearman $r=0.423$ and 0.417 respectively across tested indications, but correlations in individual indications were as high 0.676 in PRAD for Cyclin B1, and 0.593 in LGG for MSH6. Overall, lower prediction accuracy in protein models was expected due to the lower signal-to-noise ratio in the RPPA technology compared to RNA-seq. However, in specific settings such as the PRAD model, both total and phosphoproteins showed strong prediction accuracy in test sets. Progesterone receptor (PR) and cMET_pY1235 reached Spearman $r=0.614$ and 0.631 respectively in this indication (Fig. 4e). Also, the AKT/mTOR pathway showed evidence of strong prediction accuracy in the sarcoma setting where S6 and RICTOR_pT1135 antibodies showed Spearman $r=0.731$ and 0.695 respectively (Fig. 4e). Despite paucity of representation in the RPPA panel, stromal and immune features were also found among the best-performing proteins such as ECM-associated Collagen-VI, Fibronectin and T cell-associated Lck (Fig. 4d). Overall, 11 out of 20 tested indications showed significant performance (adj. $p < 0.1$) for more than 90% of tested antibodies (total $N=191$ antibodies) (Fig. 4f).

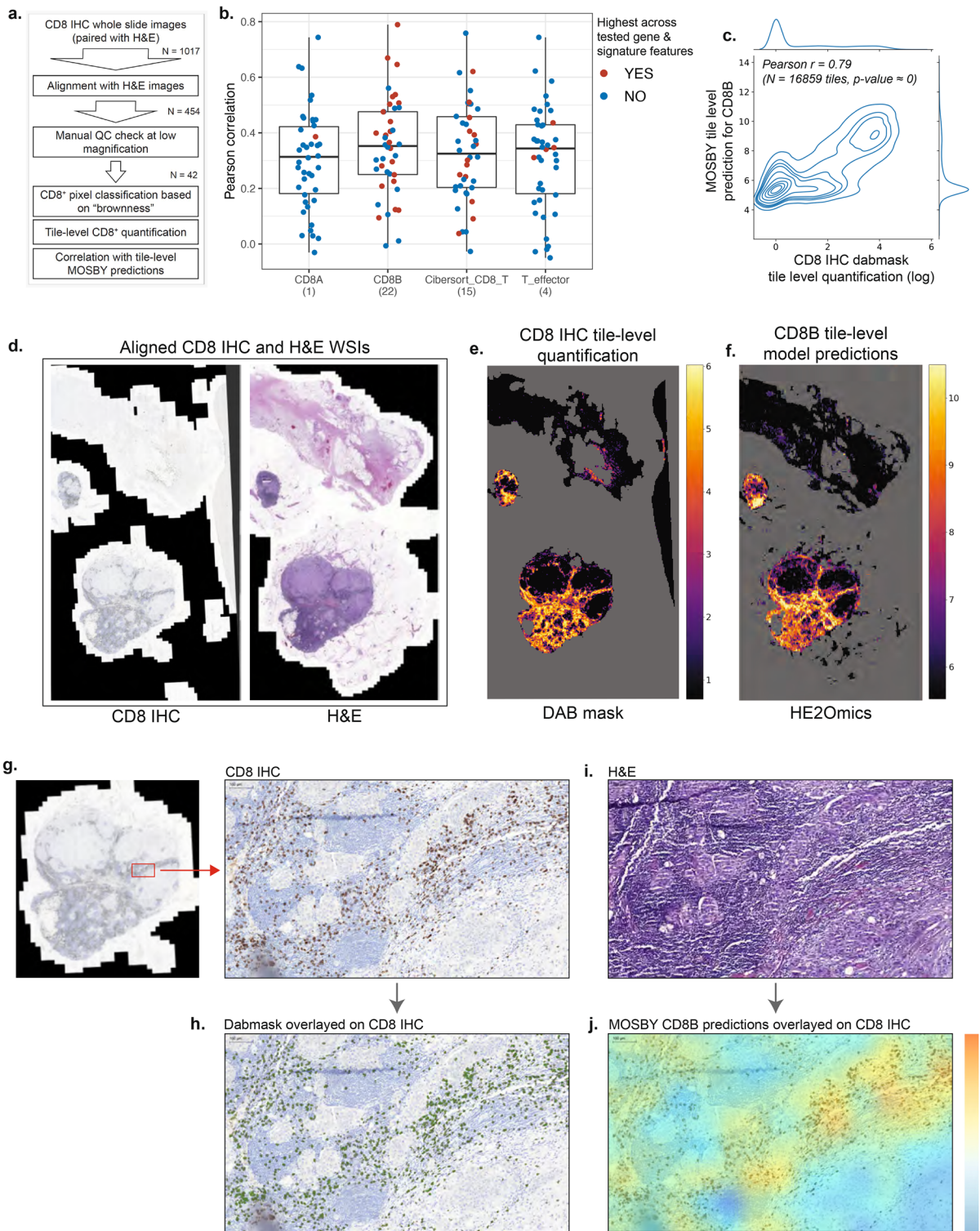
In terms of single gene models, best performing features confirmed the strong prediction accuracy associated with stromal features observed above; the highest-ranking genes were marker genes of fibroblasts (e.g. LUM, COL5A1) (Fig. 4g). Top-ranking genes also included markers of macrophages (e.g. CSF1R), and T cells (e.g. CD3E). In DNA models, tumor cellularity measures (tumor purity, cancer DNA fraction) achieved better performance than subclonal genome fraction, and average DNA methylation features (Fig. 4h).

Taken together, these results highlight the promise of MOSBY predictions for profiling cell populations in complex tissue architectures. Models trained in TCGA lung adenocarcinoma were further deployed in IMpower150⁴², an independent validation cohort with pathologist-annotated WSIs. Comparison of pathologist cancer epithelia annotations with model predictions indicated that MOSBY is effective in inferring intratumor heterogeneity (Supplementary Notes, Supplementary Fig. 4). Moreover, model predictions for commonly used epithelial markers revealed important biology, such as DNA-based tumor cellularity estimates being preferable to RNA-based epithelia signatures (e.g. Taube et al.²⁰) in demarcating cancer regions via bulk measurements (Supplementary Notes, Supplementary Fig. 5).

Spatial patterns inferred from tile-level MOSBY predictions increase survival predictive power of gene signature-based models

MOSBY tile-level predictions enable in silico spatialization for a tested omic feature, as well as assess spatial correlation between two tested features (positive and negative correlations indicating colocalization and spatial exclusion respectively). We define a 'colocalization feature' as the Pearson correlation between tile-level predictions of two omic features. As correlation coefficients were computed across all tiles on a slide, colocalization features represent slide-level as opposed to 'local' spatial patterns. We focused on our signature panel ($N=175$) as the omic features to derive colocalization features and investigate survival associations, as the signature panel covered both tumor and non-tumor TME components. Moreover, using signatures as opposed to single genes enable the discovery of 'biologically interpretable' spatial biomarkers as well-designed signatures capture pathway and cell type-related gene expression with higher fidelity. For a slide, correlations from all pairwise combinations of signatures ($N=15,225$) (i.e. the collection of all colocalization features) are referred to as the 'colocalization map' from here on (Fig. 1a).

A survival analysis was performed to ask whether slide-level patterns in colocalization maps harbored survival signals that could not be captured by the mere magnitude of signature expression. Three L1-regularized Cox proportional hazards regression models were fit to address this question, and the ability of the models to predict survival was computed with concordance indices (c-index) (Fig. 5a) (Methods). The first model only used flattened colocalization features (MOSBY predictions, $N=15,225$). The second Cox model only used signature expression levels ($N=175$) with the goal of assessing the survival predictive power of gene expression 'magnitude'. The third model was a joint model combining all signatures but only lasso-selected colocalization



features from the first model in order to prevent colocalization features from dominating the model. Each model was run across 10 cross-validation folds to optimize the shrinkage parameter and also obtain mean and standard error estimates for the relevant c-index (**Methods**).

This process was implemented separately for all tested TCGA indications, and c-index mean and standard error estimates were plotted (Fig. 5a). We observed that the joint (i.e. third) model had a higher c-index than the signature-only model in most indications. This finding revealed that slide-level spatial patterns, discovered with an inference engine such as MOSBY, have survival predictive power that could not be captured by gene expression alone. As MOSBY colocalization features are biologically interpretable, this opens the door to discovering potentially 'actionable' colocalization or spatial exclusion patterns that predict clinical outcomes. Moreover, the c-indices achieved in the joint model were highly competitive with or higher than those reported in the literature

◀ **Figure 2.** Validation of MOSBY in silico spatialization with CD8 IHC whole slide images in IMvigor211. (a) Workflow for computational alignment of serially sectioned CD8 IHC and H&E whole slide images, and correlation of CD8 IHC-based ground truth values with MOSBY tile-level predictions. The derivation of tile level ground truth values by applying DAB mask on CD8 IHC images is described in **Methods**. (b) Pearson correlation between tile-level ground truth (CD8 IHC) and MOSBY-predicted values for the 42 slides that satisfied quality control criteria in (a). The x-axis shows gene (*CD8A*, *CD8B*) and signature (Cibersort CD8 T cell, T effector cell) features compared with CD8 IHC. For a given slide, the feature with the highest CD8 IHC correlation is denoted with a red dot, while the other three features are denoted with blue dots. The number of slides where each feature had the highest CD8 IHC correlation is shown in parenthesis under the x-axis labels. Lower and upper hinges in box plots correspond to first and third quartiles, while whiskers extend to 1.5 times interquartile range. (c–j) Description of results for the slide where MOSBY predictions are the most concordant with CD8 IHC-based ground truth levels: (c) Density plot showing the correlation between tile-level ground truth data (DAB mask applied on CD8 IHC WSI) and MOSBY *CD8B* predictions (Pearson $r = 0.79$, $N = 16,859$ tiles, two-sided $p \approx 0$ using the exact beta distribution of r). (d) Computationally aligned CD8 IHC and H&E whole slide images. (e) Heat map showing CD8 IHC tile level quantification. DAB mask is applied on CD8 IHC images to call ‘brown’ pixels. Brown pixels are counted within each 100×100 px window. Count values are log-transformed, and then clipped at 50th and 98th percentiles for contrast. Background regions are denoted with gray. (f) Heat map showing *CD8B* tile-level model predictions. Predicted values are clipped at 50th and 98th percentiles for contrast. Background regions are denoted with gray. (g) High magnification example of a CD8 T cell-rich region on the CD8 IHC image. (h) DAB mask classification of ‘brown’ pixels (denoted with green) overlaid on CD8 IHC image. (i) Corresponding region on the H&E image as identified by slide-level computational alignment. (j) *CD8B* model predictions inferred from the H&E image in (i) overlaid on the CD8 IHC image in (g). Orange and blue show high and low values respectively.

from end-to-end survival-trained multimodal models utilizing genomic, transcriptomic, and image datasets⁴³. Of note, comparing colocalization-only and signature-only models, we also found that the total survival predictive power of slide-level spatial patterns was not as high as that of signature levels in most TCGA indications. Ovarian and rectal cancer results were an exception to this general pattern (Fig. 5a), suggesting spatial biomarkers discovered in these indications may have the greatest potential to lead to novel insights.

Colocalization maps enable discovery of biologically interpretable spatial biomarkers

We next investigated consistent spatial predictors of risk that were supported across multiple indications and also showed evidence of tumor specificity. In each indication, survival effects and tumor specificity of colocalization features were explored with two tests: (1) A univariate Cox regression model for survival, and (2) a Mann–Whitney test to compare tumor vs. normal levels of colocalization. A potential spatial biomarker of risk was defined as a colocalization feature that was significantly associated with poor overall survival ($p < 0.05$) and also had elevated levels of colocalization in the tumor (adj. $p < 0.05$). Given these criteria, four colocalization features had evidence in four different TCGA indications to be a spatial biomarker of risk (Fig. 5b). Of these four, the colocalization between an ER stress signature (XBP1s targets ER17²²) and a neurotransmitter signature was associated with poor survival and malignant state in colon adenocarcinoma, lung adenocarcinoma, liver hepatocellular and ovarian cancers (Fig. 5c,d). In an independent non-squamous lung cancer study involving both immune checkpoint blockade (atezolizumab) and chemotherapy arms (IMpower110), this colocalization feature was also found to be associated with poor survival in the chemotherapy, but not immunotherapy arm, suggesting higher relevance as a resistance factor in chemotherapy (Fig. 5e,f). Of note, the ER17 and neurotransmitter signatures were not individually found to be associated with risk in any of the four mentioned indications (Supplementary Fig. 6a,b). Visual inspection of WSIs indicated that the expression of ER17 and neurotransmitter signatures primarily came from the microenvironment (as opposed to tumor region) in the case of high colocalization (Supplementary Fig. 6c). In low colocalization cases, the neurotransmitter signature expression primarily came from the tumor region whereas the ER17 signature expression was again predominantly in the microenvironment (Supplementary Fig. 6d).

Focusing on colocalization features involving immune system signatures, the T effector cell vs cysteine colocalization was identified as the most consistent spatial biomarker of risk in TCGA. This colocalization feature was associated with poor survival and also showed significant tumor enrichment in breast, squamous lung, and ovarian cancers (Fig. 6a,b). T effector and cysteine signatures were not individually found to be associated with risk in any of these indications (Supplementary Fig. 7a,b). Visual inspection of WSIs indicated that the expression of T effector and cysteine signatures primarily came from the microenvironment (as opposed to tumor region) in the case of high colocalization (Supplementary Fig. 7c). As high colocalization is a risk factor, this expression pattern may be suggestive of a cysteine-associated immunosuppressive TME. In low colocalization cases, the T effector signature expression primarily came from the microenvironment whereas the cysteine signature expression was predominantly in the tumor region (Supplementary Fig. 7d).

The strongest survival effect for T effector vs cysteine colocalization was observed in breast cancer, where we investigated other immune cell types and found that immune vs cysteine colocalization was a general negative prognosis biomarker in this indication. Significant survival associations were observed for both lymphocytes/NK cells (Fig. 6c), and myeloid populations (Fig. 6d). The same immune vs cysteine colocalization features were also found to be negative prognostics in the atezolizumab arm of Impower110 non-squamous cohort (Fig. 6e,f). Of note, most of these features were not prognostic in the chemotherapy arm (Supplementary Fig. 7e,f), however

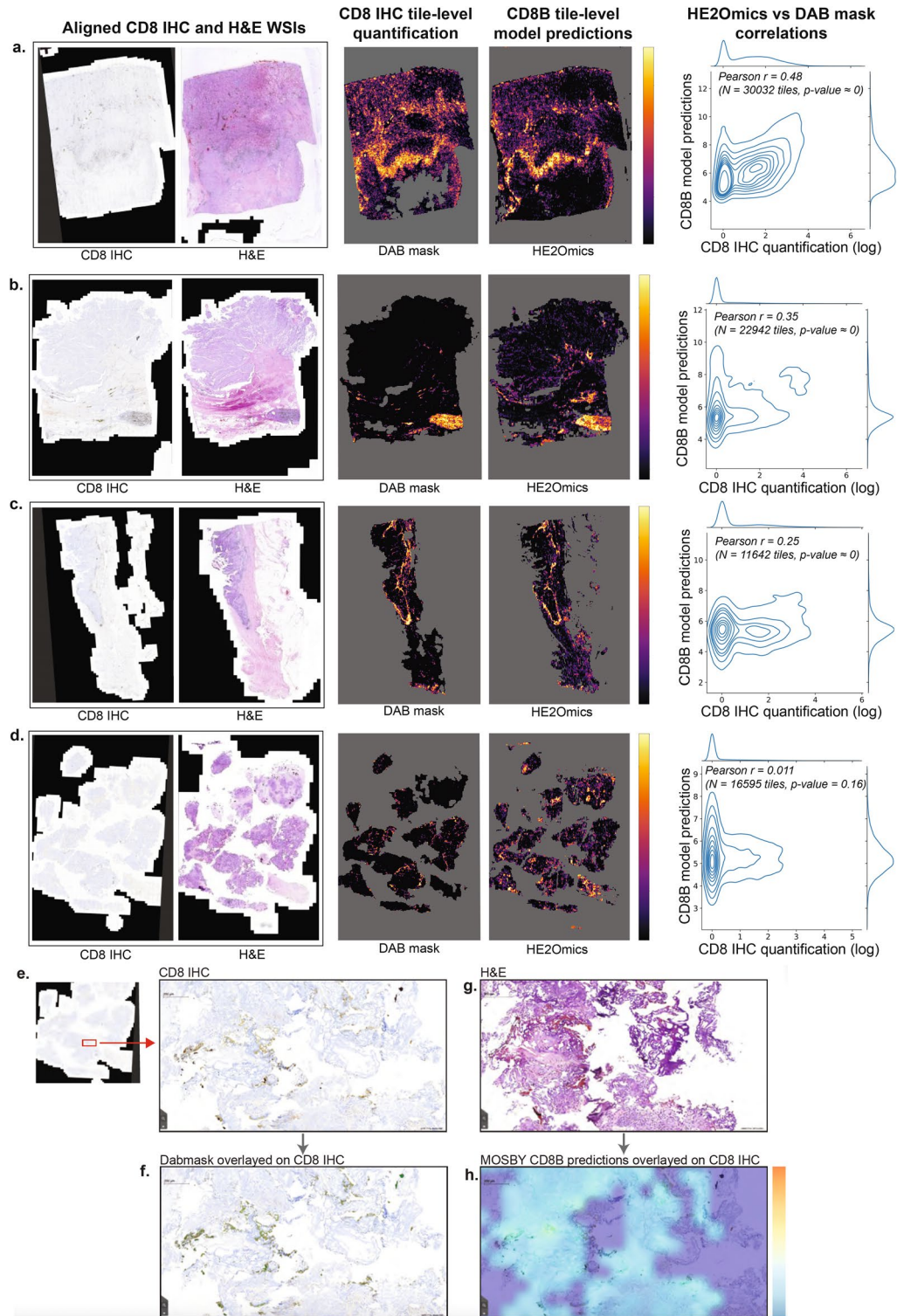


Figure 3. Representative slides showing the variability for CD8 IHC vs. MOSBY concordance in IMvigor211. **(a–d)** Slides from the 75th, 50th, 25th, and 0th percentiles of the Pearson correlation between *CD8B* model predictions and CD8 IHC ground truth quantification. In each figure, subpanels from left to right depict (1) computationally aligned CD8 IHC and H&E whole slide images, (2) heat map showing CD8 IHC tile level quantification (plotted values derived from 100×100 px windows, and clipped at 50th and 98th percentiles) with gray-colored background regions, (3) heat map showing *CD8B* tile level model predictions (clipped at 50th and 98th percentiles) with gray-colored background regions, (4) density plot showing the correlation between tile-level ground truth data (DAB mask applied on CD8 IHC WSI) and MOSBY *CD8B* predictions. **(a)** Slide from the 75th percentile with $r = 0.48$. **(b)** Slide from the 50th percentile with $r = 0.35$. **(c)** Slide from the 25th percentile with $r = 0.25$. **(d)** Slide from the 0th percentile with $r = 0.011$. **(e)** High magnification example of a CD8 T cell-poor region on the slide with $r = 0.011$ from **(d)**. This region shows the presence of brown-staining artifacts. **(f)** DAB mask classification of 'brown' pixels (denoted with green) overlaid on CD8 IHC image. Brown-staining artifacts are also captured with DAB mask. **(g)** Corresponding region on the H&E image as identified by slide-level computational alignment. **(h)** *CD8B* model predictions inferred from the H&E image in **(g)** overlaid on the CD8 IHC image in **(e)**. Orange and blue show high and low values respectively.

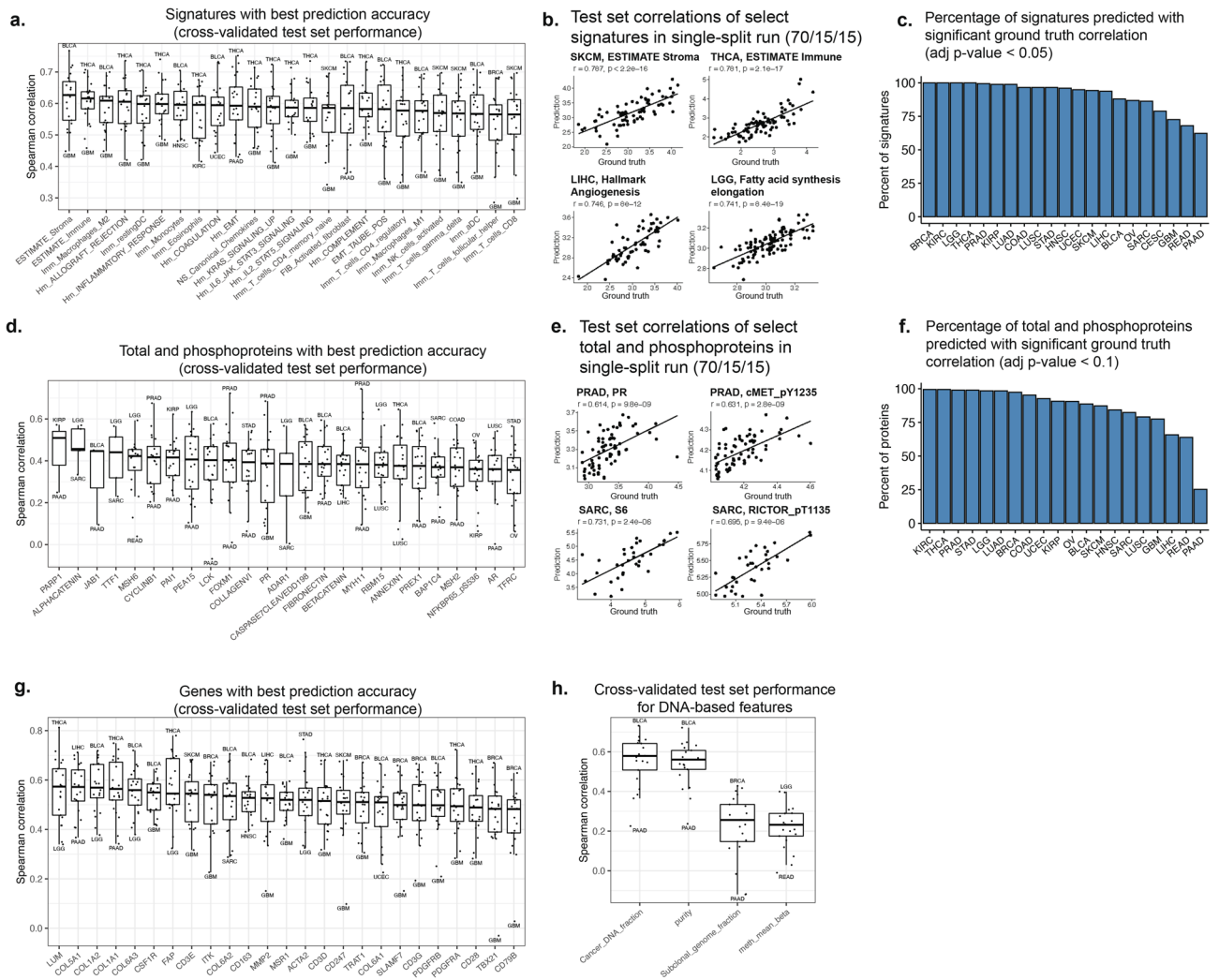


Figure 4. Best performing features in MOSBY signature, protein, gene, and DNA panels. **(a)** Top 25 signatures with best prediction accuracy according to cross-validated test set performance in single-indication models. **(b)** Test set correlations of select signatures in a single split run (70% training, 15% validation, and 15% test sets). **(c)** Percentage of signatures predicted with significant ground truth correlation (adjusted p -value < 0.05). **(d)** Top 25 proteins/phosphoproteins with best prediction accuracy according to cross-validated test set performance in single-indication models. **(e)** Test set correlations of select proteins/phosphoproteins in a single split run (70% training, 15% validation, and 15% test sets). **(f)** Percentage of proteins/phosphoproteins predicted with significant ground truth correlation (adjusted p -value < 0.1). **(g)** Top 25 genes with best prediction accuracy according to cross-validated test set performance in single-indication models. **(h)** Cross-validated test set performance of all tested DNA-based features. Lower and upper hinges in box plots correspond to first and third quartiles, while whiskers extend to 1.5 times interquartile range.

did not qualify as predictive biomarkers since the survival association differences between atezolizumab and chemotherapy arms were not significant.

Discussion

The MOSBY workflow achieves prediction of bulk omic profiles from H&E WSI features. Our results showed that, compared to ImageNet-based pretraining, self-supervised pretraining in large histological datasets allows creation of inference engines (e.g. RetCCL) that enable a more accurate mapping from image features to gene, signature, protein, and DNA-based measurements. We demonstrated that the most accurately predicted features by MOSBY involved processes such as proliferation, immune/stromal infiltration, differentiation, and epithelial-to-mesenchymal transition. This finding suggests that self-supervised features learned from pan-cancer histological datasets run the risk of accentuating biological processes and pathways that show the highest variation across different cancer indications. Training feature extractors on images from a single indication may be required to capture biological processes that play an important role in one or only a few cancer indications. RetCCL-based features showed promise by capturing angiogenesis in hepatocellular carcinoma, and fatty acid biology in low grade glioma. Yet, the accumulation of even larger histological datasets in the future have potential to allow

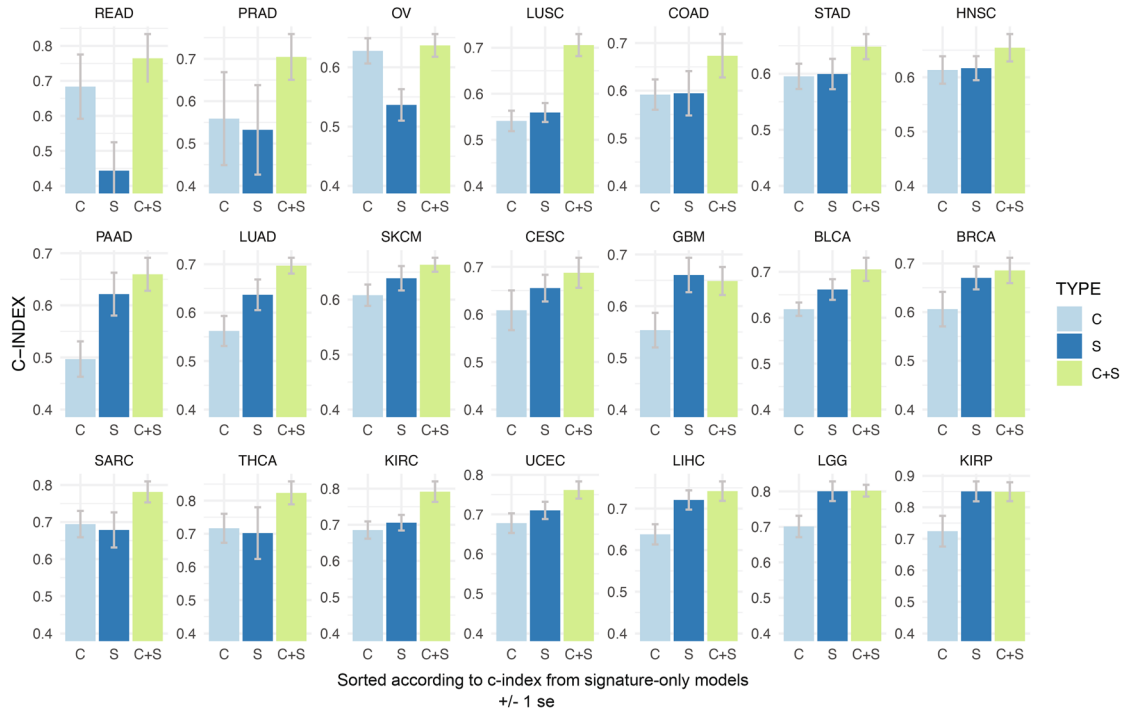
a. L1-regularized Cox regression models to compare the survival predictive power of signature expression and colocalization features

C: Colocalization, **S:** Signature-only, **C+S:** Colocalization + Signatures

C (model 1): Survival ~ Colocalization (N=15225)

S (model 2): Survival ~ Signature expression (N=175)

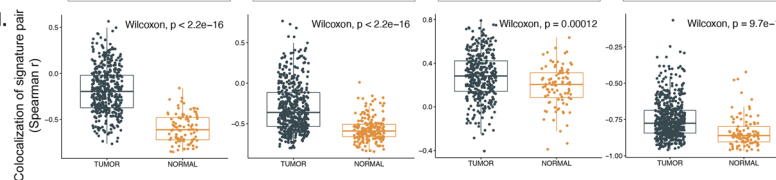
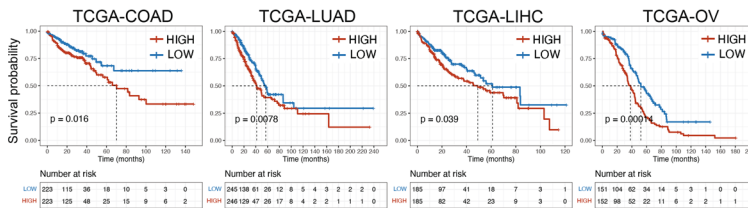
C+S (model 3): Survival ~ Signature expression + Lasso-selected colocalization features (N varies by indication)



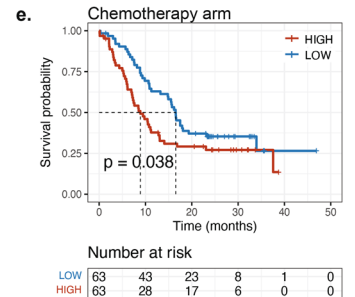
b.

Colocalization partner 1	Colocalization partner 2	Indications	Count
ER stress XBP1s targets ER17	Sabatini Neurotransmitter	COAD, LUAD, LIHC, OV	4
Hallmark Coagulation	Sabatini Vitamin A	COAD, LIHC, LUSC, OV	4
Sabatini Complex II	Sabatini Urea	BRCA, COAD, LUAD, STAD	4
Sabatini Custom Triglycerides	Sabatini Ketone Bodies	BLCA, COAD, LIHC, LUSC	4

c. ER stress XBP1s targets ER17 vs Sabatini Neurotransmitter colocalization



e. IMpower110 non-squamous cohort



f. Atezolizumab arm

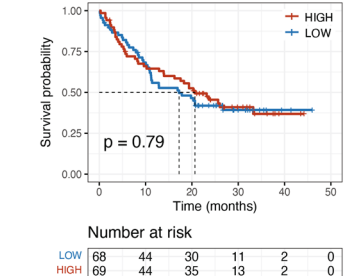


Figure 5. Survival predictive power comparison between gene signatures and colocalization features. **(a)** Concordance indices of survival models involving only colocalization features (denoted with C), only signatures (denoted with S), or signatures and lasso-selected colocalization features (denoted with C+S) in 21 TCGA indications. Indications are ordered according to increasing concordance index in signature-only models. Error bars denote ± 1 standard error. **(b)** Colocalization features most consistently associated with poor overall survival and elevated tumor levels in multiple TCGA indications. **(c,d)** Survival and differential tumor colocalization evidence for the ER17 vs. Neurotransmitter signature pair in TCGA colon, liver, lung, and ovarian cancer cohorts. **(c)** Median-split Kaplan–Meier plots and logrank test p-values, **(d)** tumor vs normal box plots and Wilcoxon test p-values. **(e,f)** Median-split Kaplan–Meier plots and logrank test p-values for the colocalization between ER17 and Neurotransmitter signatures in IMpower110 non-squamous cohort: **(e)** Chemotherapy arm, **(f)** Atezolizumab arm. Lower and upper hinges in box plots correspond to first and third quartiles, while whiskers extend to 1.5 times interquartile range. Red and blue color denote high and low levels respectively.

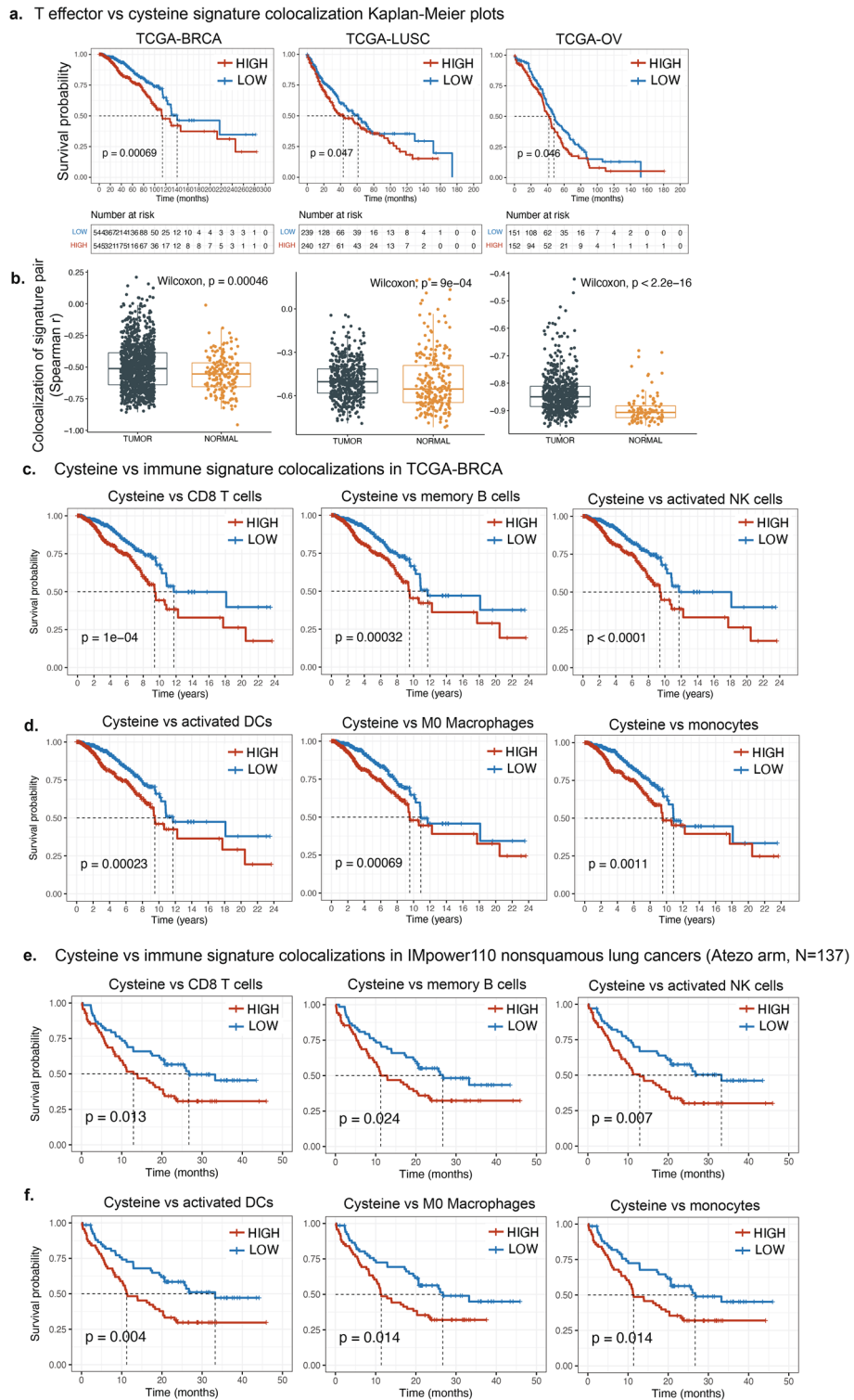


Figure 6. Immune system-related spatial biomarkers consistently associated with risk. **(a,b)** Survival and differential tumor colocalization evidence for T effector cell and Sabatini Cysteine signature pair in TCGA breast, lung squamous, and ovarian cancer cohorts: **(a)** Median-split Kaplan–Meier plots and logrank test p-values, **(b)** tumor vs normal box plots and Wilcoxon test p-value. Lower and upper hinges in box plots correspond to first and third quartiles, while whiskers extend to 1.5 times interquartile range. **(c,d,e,f)** Median-split Kaplan–Meier plots and logrank test p-values for Sabatini Cysteine colocalization with various immune cell populations: **(c,d)** TCGA breast cancer cohort: **(c)** Lymphocytes and NK cells, **(d)** myeloid cell populations. **(e–f)** Atezolizumab arm of IMpower110 non-squamous cohort (N = 137): **(e)** Lymphocytes and NK cells, **(f)** myeloid cell populations. Red and blue color denote high and low levels respectively.

refinement of image features relevant for indication-specific pathways, thus making possible the discovery of a greater number of clinically relevant spatial biomarkers.

MOSBY, as with many other deep learning models, adopts a weakly-supervised approach by first making tile-level predictions and then aggregating tiles to obtain a prediction at the WSI level^{4,44}. Although an intermediate output of the model, tile-level predictions enable *in silico* spatialization of whole slide-level annotations, opening the way to inferring intratumor heterogeneity for the slide-level information used as ground truth^{12,43}. Spatial intratumor heterogeneity patterns learned from a cohort of patients subsequently allow investigation of clinically relevant biomarkers. In MOSBY, spatial patterns are captured by pairwise colocalization features. For a given signature pair, the colocalization value on a slide is defined as the Pearson correlation across all tiles. Thus, MOSBY colocalization features capture *slide-level* but not *local* spatial patterns. Local processes such as tertiary lymphoid structures are known to affect patient survival and response to cancer immunotherapy^{45,46}. We demonstrated in this study that slide-level spatial patterns also carried survival signals, and increased predictive power of gene signature-based survival models in most TCGA indications. Moreover, we noted that the concordance indices of joint models (both colocalization features and gene signature levels) either surpassed or were comparable to those of multimodal deep learning models that employed the complete set of WSI and omic (RNA-seq, mutation status, copy number variation) data in TCGA⁴³. This finding indicated that the 175-signature panel we defined was sufficient to capture most biological processes important for clinical outcome.

End-to-end neural networks trained to predict survival have been lacking in terms of direct biological interpretation of image regions important for the model^{43,47}. These models may incorporate mechanisms such as attention heatmaps and spatial credit assignment to increase interpretability^{43,48}, yet still require pathologist efforts to examine important image regions whereby interpretation remains limited to phenotypes visible by human eye. In contrast, MOSBY spatial features are biologically interpretable by design, which is an advantage of this approach over end-to-end neural networks trained to predict survival^{43,47}. In this study, we showed that the colocalization of an endoplasmic reticulum (ER) stress-related signature and a neurotransmitter signature is both elevated in tumors and associated with poor overall survival in four TCGA indications. The poor survival association in lung adenocarcinoma was also validated in the chemotherapy arm of an independent NSCLC cohort (nonsquamous samples in Impower110), suggesting this colocalization may be a chemotherapy-specific risk factor. Moreover, we identified the T effector cell vs cysteine signature colocalization as a TME-related risk factor in multiple TCGA indications, as well as in the immunotherapy arm of Impower110 nonsquamous cohort. These results showcase the high utility of colocalization maps for discovering biologically interpretable clinically relevant spatial biomarkers.

A limitation of our study is that MOSBY colocalization maps are not able to capture local spatial patterns. Future work involves the investigation of a graph neural network-based model on tile-level MOSBY predictions where we can capture local as well as slide-level patterns. Moreover, transformer-based architectures may increase the expressiveness of our model to allow a more accurate mapping from image features to bulk omic profiles.

Methods

Data

TCGA: Batch-normalized RNA sequencing, RPPA datasets as well as clinical and DNA-based data were obtained from the PanCanAtlas publications page of the Genomic Data Commons website (<https://gdc.cancer.gov/about-data/publications/pancanatlas>). H&E-stained slide images (i.e. Tissue Slides) were downloaded from GDC Data Portal (<https://portal.gdc.cancer.gov/>).

Computation of signature scores from bulk RNA-seq data: In TCGA and IMvigor RNA-seq datasets, normalized gene expression values were log-transformed ($\log(x + 1)$), z-transformed across samples to have 0 mean, unit variance, and subsequently averaged across genes to arrive at a single signature score for each sample. In TCGA, signature scoring was performed on the batch-normalized pan-cancer RNA-seq dataset, and hence signature scores were comparable across cancer indications.

Spatially-resolved transcriptomic data: Publicly available breast cancer tissue slides and spatial transcriptomic assays processed by the Spatial Transcriptomics method^{41,49} were downloaded. The dataset contained 68 WSIs from a total of 23 patients along with spot coordinates and respective RNA-seq expression values from the spatial transcriptomics assay. WSIs were tiled into 224×224 patches as input for the MOSBY model. Each tile was centered around the pixel coordinate of an assay spot to represent the 100 μm region of the spot. Log-normalized gene expression values were used as ground truth. Signature scores were calculated by first creating an AnnData object using the `anndata` package then applying the `score_genes` function from `scanpy`⁵⁰. All analyses were conducted using Python 3.10.

MOSBY preprocessing

Image tiling: On whole slide images, foreground elements (tissue) and background (glass) were isolated through luminosity-based segmentation. The Python library OpenSlide was used as a backend for generating 224×224 px foreground tiles at 0.5mpp resolution. The same tiling protocol was leveraged for both training and inference, allowing tiles to be mapped back to the original WSI positions for visualization (Supplementary Fig. 8).

Feature Extraction: Contrastive self-supervised learning-based RetCCL⁸ was used to extract image features. RetCCL employs a ResNet-50 architecture to extract 2048 features for each image tile.

Model training

TCGA: A maximum of 8000 tiles were selected randomly for each slide to yield an unbiased representation of the slide, and all tiles were concurrently used for training. A 2-layer perceptron (512 and 256 nodes per layer) was used to map image features to omic variables. Number of epochs was set to a maximum of 300, with early

stopping allowed with a *patience* of 30 epochs. Ground truth RNA-seq data were log-transformed. Ground truth signature and protein levels had negative values, and thus were shifted to make all values nonnegative. The model was trained with MSE loss between prediction and ground truth levels, while Spearman correlation between these quantities was used as early termination criterion in the validation set. A batch size of 64, and AdamW optimizer with 1e-3 weight decay were used. Learning rate scheduler was implemented with step size 5 and gamma 0.9.

Fivefold cross-validation was performed (64% training, 16% validation, and 20% test set in each fold) to assess model performance. Information leak was prevented by assigning all WSIs from the same patient to the same partition. For inference, full models were trained with 80% of the data, with the remaining 20% used as validation set to check early stopping criterion.

IMvigor210 and IMvigor211: A full model was trained using both IMvigor210 and IMvigor211 images, and largely following the parameters used in TCGA training. Patients were split into 80% training and 20% validation set, stratified by trial to prevent information leak. Cross-validation models were trained using 64% training, 16% validation, and 20% test set (in each one of the 5 folds) for the RetCCL vs ImageNet comparison. The signature and gene model consisted of 175 signatures (Supplementary Table 3) and 73 genes (Supplementary Table 2) respectively. Different from TCGA runs, a maximum of 4000 randomly selected tiles per WSI were used during training, and WSIs having less than 200 tiles were filtered out. Ground truth RNA-seq data were log-transformed and standardized to have 0 mean and unit variance. Both gene and signature expression levels were shifted to make all values nonnegative.

Computational hardware and software

Programming languages Python (v3.10.8, <https://www.python.org/downloads/release/python-3108/>) and R (v4.1.1, <https://cran.r-project.org/src/base/R-4/R-4.1.1.tar.gz>) were used in this study. MOSBY was built with the PyTorch library (v1.11.0) in Python as a novel implementation of the HE2RNA¹² model. Python libraries used for data processing included NumPy (v1.23.5), Pandas (v1.2.4), Scikit-learn (v0.24.1), OpenSlide (v1.1.1), Zarr (v2.12.0), TiffFile (v2020.10.1), and OpenCV-cv2 (v4.5.5). Whole slide image tiling, RetCCL feature extraction and MOSBY model training were implemented in NVIDIA Tesla V100 Tensor Core GPUs (graphics processing units). Deep learning models were trained with NVIDIA Cuda compiler (v12.1.105). Data visualization in Python was implemented with Matplotlib (v3.3.4) and Seaborn (v0.11.1) libraries. Python statistical analyses such as Spearman and Pearson correlation were implemented with the SciPy library (v1.10.1).

Data processing in R was implemented with dplyr (v1.0.8), magrittr (v2.0.2), and reshape2 (v1.4.4) libraries. Data visualization in R was performed using ggplot2 (v3.3.5), ggpubr (v0.4.0), and ggsci (v2.9) libraries.

CD8 IHC and H&E whole slide image alignment

IMvigor210 and IMvigor211 slides were scanned by an external party (CellCarta, Montreal, QC) on a Panoramic 250 (3DHitech, Budapest, Hungary) with a 20× or 40× objective. Digital files were transferred to Genentech and converted to the Aperio SVS image format for all viewing and analysis. Hematoxylin and Eosin (HE) slides were aligned with a CD8 stained section from the same block (1017 total sample pairs). These sections were almost never serial, and displayed variable degrees of adjacency characteristic of the block being refaced between sections. Image alignment was performed in Matlab (R2022a, Mathworks, Natick, MA.) by downsampling to ~20 um per pixel and converting them to normalized grayscale images, before calculating an affine transformation using mutual information as the underlying matching metric. The transformation was then upsampled before being applied to the original CD8 high magnification image data bringing it into alignment with the HE image. Various alignment metrics were produced (intersection over union, normalized cross correlation) to select for correctly aligned images (454 sample pairs) before a final manual QC check at low magnification (at least 95% of tissue present on both slides, with the majority of visible structures in close proximity) that was not exhaustive and resulted in 42 sample pairs subjected to further analysis.

A binary mask was then produced from the aligned CD8 image using both HSV thresholding and a blue-normalized “brownness” algorithm⁵¹.

CD8 IHC tile-level quantification

Tile size used for H&E images was set to 224×224 pixels at 0.5mpp. In cases where the native resolution was different from 0.5mpp, tiling was performed with an adjusted tile size at the native resolution, and tiles were subsequently downsampled or upsampled to arrive at 224 pixels at 0.5mpp. In most IMvigor slides, the native resolution was 0.243mpp resulting in an adjusted tile size of approximately 460×460 pixels. CD8 IHC images were tiled using the ‘adjusted’ tile size to match with the corresponding H&E tiles. The CD8 IHC count for a tile was computed using a convolution approach: Each tile was split into 30×30 p subtiles, for which the number of 1 s (brown pixels) was counted. The counts across subtiles were averaged to obtain the value for the tile. This average value was log-transformed ($\log(x + 1)$) in comparisons with MOSBY model predictions.

Survival analysis

TCGA concordance index analysis: MOSBY signature models (full models trained with 80%-20% training/validation split) were used to make inference on all slides in a cancer indication. For a given slide, the colocalization for a signature pair was computed with a Pearson correlation across tile-level predictions of the two signatures. The collection of all pairwise correlation values for a slide formed the ‘slide colocalization matrix’ ($N = 175 \times 175$). For patients with multiple existing H&E slides, the patient-level colocalization matrix was computed as the average of all pertinent slide-level colocalization matrices. Three sets of cross-validation runs were implemented for L1-regularized Cox regression models. The inputs to these survival models consisted of: Model 1) Flattened patient-level colocalization maps ($N = 15,225$ features). Model 2) patient-level signature scores ($N = 175$

signatures). Model 3) all tested signatures (N = 175) and lasso-selected colocalization features from Model 1. The input features were regressed against overall survival in all models and for all tested indications.

In all three models, the sequence of possible values for lambda (shrinkage parameter) was internally determined in the `cv.glmnet` function from the `glmnet`⁵² R library (v4.1.3) prior to cross-validation runs. The lambda maximizing the mean Harrel's concordance measure across 10 cross-validation folds was chosen as optimal, and used to determine the concordance index estimate for the model. The standard error estimate for the model concordance index was calculated across the 10 cross-validation folds.

Kaplan–Meier plots: The `survminer` R library (v0.4.9) was used with a median cutoff to generate Kaplan–Meier plots. Log-rank test p-values were obtained internally in `survminer` using the `survdiff` function in the `survival`⁵³ library (v3.3.1).

Data sharing

For up to date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: https://go.roche.com/data_sharing.

Data availability

The accession URLs for publicly available data analyzed in this study (TCGA, spatial transcriptomics) are listed in the Data section of Methods. Datasets from clinical trials IMpower150 (H&E image data), IMpower110 (H&E image and clinical data), IMvigor210 (RNA-seq, H&E image and CD8 IHC image data), IMvigor211 (RNA-seq, H&E image and CD8 IHC image data) were also analyzed in the current study. IMvigor210 RNA-seq data is available at the European Genome-phenome archive (EGA) under the accession number EGAS00001002556, and was also published as an R package (<http://research-pub.gene.com/IMvigor210CoreBiologies/>). IMpower110 and IMpower150 datasets are not publicly available as data release is designated for the pending primary biomarker manuscripts. IMvigor210 and IMvigor211 H&E and CD8 IHC image data as well as IMvigor211 RNA-seq data that support the findings of this study are available from Roche, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Roche. For up to date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: https://go.roche.com/data_sharing.

Received: 8 February 2024; Accepted: 1 August 2024

Published online: 06 August 2024

References

- Comiter, C. *et al.* Inference of single cell profiles from histology stains with the Single-Cell omics from Histology Analysis Framework (SCHAF). <https://doi.org/10.1101/2023.03.21.533680> (2023).
- Alsaafin, A., Safarpour, A., Sikaroudi, M., Hipp, J. D. & Tizhoosh, H. R. Learning to predict RNA sequence expressions from whole slide images with applications for search and classification. *Commun. Biol.* **6**, 304 (2023).
- Yang, K. D. *et al.* Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nat. Commun.* **12**, 31 (2021).
- Tsai, P.-C. *et al.* Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nat. Commun.* **14**, 2102 (2023).
- Haviv, D., Gatie, M., Hadjantonakis, A. K., Nawy, T. & Peer, D. The covariance environment defines cellular niches for spatial inference. *Nat. Biotechnol.* <https://doi.org/10.1101/2023.04.18.537375> (2023).
- Chen, X., Fan, H., Girshick, R. & He, K. Improved baselines with momentum contrastive learning. Preprint at <http://arxiv.org/abs/2003.04297> (2020).
- Oquab, M. *et al.* DINOv2: Learning robust visual features without supervision. Preprint at <http://arxiv.org/abs/2304.07193> (2023).
- Wang, X. *et al.* RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval. *Med. Image Anal.* **83**, 102645 (2023).
- Fremont, S. *et al.* Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: A combined analysis of the PORTEC randomised trials and clinical cohorts. *Lancet Digit. Health* **5**, e71–e82 (2023).
- Schirris, Y., Gavves, E., Nederlof, I., Horlings, H. M. & Teuwen, J. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. *Med. Image Anal.* **79**, 102464 (2022).
- Ciga, O., Xu, T. & Martel, A. L. Self supervised contrastive learning for digital histopathology. *Mach. Learn. Appl.* **7**, 100198 (2022).
- Schmauch, B. *et al.* A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat. Commun.* **11**, 3877 (2020).
- Nahhas, O. S. M. E. *et al.* Regression-based Deep-Learning predicts molecular biomarkers from pathology slides. Preprint at <http://arxiv.org/abs/2304.05153> (2023).
- Hoadley, K. A. *et al.* Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018).
- Bagaev, A. *et al.* Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39**, 845–865.e7 (2021).
- Patil, N. S. *et al.* Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer. *Cancer Cell* **40**, 289–300.e4 (2022).
- Liberzon, A. *et al.* The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
- Miranda, A. *et al.* Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9020–9029 (2019).
- Şenbabaoglu, Y. *et al.* Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* **17**, 231 (2016).
- Taube, J. H. *et al.* Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15449–15454 (2010).

21. Masiero, M. *et al.* A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis. *Cancer Cell* **24**, 229–241 (2013).
22. Harnoss, J. M. *et al.* IRE1 α disruption in triple-negative breast cancer cooperates with antiangiogenic therapy by reversing ER stress adaptation and remodeling the tumor microenvironment. *Cancer Res.* **80**, 2368–2379 (2020).
23. Gene Ontology Consortium *et al.* The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
24. Hobert, O., Carrera, I. & Stefanakis, N. The molecular and gene regulatory signature of a neuron. *Trends Neurosci.* **33**, 435–445 (2010).
25. Robertson, A. G. *et al.* Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e25 (2017).
26. Tsai, H. K. *et al.* Gene expression signatures of neuroendocrine prostate cancer and primary small cell prostatic carcinoma. *BMC Cancer* **17**, 759 (2017).
27. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
28. Xu, Q. *et al.* Pan-cancer transcriptome analysis reveals a gene expression signature for the identification of tumor tissue origin. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc.* **29**, 546–556 (2016).
29. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40**, 499–507 (2008).
30. Bhattacharya, B., Puri, S. & Puri, R. K. A review of gene expression profiling of human embryonic stem cell lines and their differentiated progeny. *Curr. Stem Cell Res. Ther.* **4**, 98–106 (2009).
31. Shats, I. *et al.* Using a stem cell-based signature to guide therapeutic selection in cancer. *Cancer Res.* **71**, 1772–1780 (2011).
32. Kim, J. *et al.* A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–324 (2010).
33. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).
34. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
35. Mariathasan, S. *et al.* TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
36. Böttcher, J. P. & Reis e Sousa, C. The role of type 1 conventional dendritic cells in cancer immunity. *Trends Cancer* **4**, 784–792 (2018).
37. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
38. Akbani, R. *et al.* A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat. Commun.* **5**, 3887 (2014).
39. Balar, A. V. *et al.* Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: A single-arm, multicentre, phase 2 trial. *Lancet Lond. Engl.* **389**, 67–76 (2017).
40. Powles, T. *et al.* Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): A multicentre, open-label, phase 3 randomised controlled trial. *Lancet Lond. Engl.* **391**, 748–757 (2018).
41. Stenbeck, L., Bergenstråhle, L., Lundeberg, J. & Borg, Å. Human breast cancer in situ capturing transcriptomics. *Mendeley* <https://doi.org/10.17632/29ntw7sh4r.5> (2021).
42. Socinski, M. A. *et al.* Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N. Engl. J. Med.* **378**, 2288–2301 (2018).
43. Chen, R. J. *et al.* Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* **40**, 865–878.e6 (2022).
44. Lu, M. Y. *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
45. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577**, 549–555 (2020).
46. Trüb, M. & Zippelius, A. Tertiary lymphoid structures as a predictive biomarker of response to cancer immunotherapies. *Front. Immunol.* **12**, 674565 (2021).
47. Katzman, J. L. *et al.* DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **18**, 24 (2018).
48. Javed, S. A. *et al.* Additive MIL: intrinsically interpretable multiple instance learning for pathology. *Adv. Neural Inf. Process. Syst.* <https://doi.org/10.48550/ARXIV.2206.01794> (2022).
49. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
50. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
51. Brey, E. M. *et al.* Automated selection of DAB-labeled tissue for immunohistochemical quantification. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **51**, 575–584 (2003).
52. Simon, N., Friedman, J., Tibshirani, R. & Hastie, T. Regularization paths for cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39**, 1–13 (2011).
53. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).

Acknowledgements

We would like to thank all of the study participants and their families, and all of the site investigators, study coordinators, and staff. We also would like to thank Brandon Kayser, Robert Johnston, Aïcha Bentaieb, Dan Ruderman, Hector Corrada Bravo, Jason Hackney, and colleagues from the Oncology Reverse Translation team for providing critical feedback on the manuscript. This work was supported by Genentech, Inc.

Author contributions

K.L., A.K, Y.S. devised the study. Y.S., V.P, A.K, J.E, E.L, E.W., B.N., M.S., M.B. acquired, analyzed, and interpreted the data. Y.S. wrote the manuscript with input from the remaining authors. All authors reviewed the manuscript.

Funding

This work was supported by Genentech, Inc.

Competing interests

J.E., E.L., B.N., M.S., M.B. are employees of Genentech, Inc. and shareholders in F. Hoffmann La Roche, Ltd. Y.S. and K.L. were employees of Genentech, Inc. V.P. was an external partner at Genentech, Inc. A.K. was an intern at Genentech, Inc. E.W. was an intern at Genentech, Inc.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69198-6>.

Correspondence and requests for materials should be addressed to Y.Ş. or K.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024