Multi-modal digital pathology for colorectal cancer diagnosis by high-plex immunofluorescence imaging and traditional histology of the same tissue section

Jia-Ren Lin[1,2,*], Yu-An Chen[1,2,*], Daniel Campton[3,*], Jeremy Cooper[3], Shannon Coy[1,4], Clarence Yapp[1,2], Juliann B. Tefft[1,2], Erin McCarty[3], Keith L. Ligon[4], Scott J. Rodig, Steven Reese[3], Tad George[3], Sandro Santagata[1,2,4,±], Peter K. Sorger[1,2,±]

[*] These authors contributed equally

[±] These authors contributed equally

Human Tissue Atlas Center

[1]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, 02115, USA.

[2]Ludwig Center at Harvard, Harvard Medical School, Boston, MA 02115, USA.

[3]RareCyte, Inc., 2601 Fourth Ave., Seattle, WA, 98121, USA.

[4]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

Keywords

Precision medicine, machine learning, immune profiling, digital pathology, colorectal cancer, multiplex microscopy, diagnostics, multiplexed imaging, computational biology, spatial analysis, fluorescence microscopy, neoplasms, pathology, molecular pathology

Pre-publication correspondance

peter_sorger@hms.harvard.edu cc: lsp-papers@hms.harvard.edu

Cell Phone: 617-797-4928

ABSTRACT

Precision medicine is critically dependent on better methods for diagnosing and staging disease and predicting drug response. Histopathology using Hematoxylin and Eosin (H&E) stained tissue - not genomics – remains the primary diagnostic modality in cancer. Moreover, recently developed, highly multiplexed tissue imaging represents a means of enhancing histology workflows with single cell mechanisms. Here we describe an approach for collecting and analyzing H&E and high-plex immunofluorescence (IF) images from the same cells in a whole-slide format suitable for translational and clinical research and eventual deployment in diagnosis. Using data from 40 human colorectal cancer resections (60 million cells) we show that IF and H&E images provide human experts and machine learning algorithms with complementary information. We demonstrate the automated generation and ranking of computational models, based either on immune infiltration or tumor-intrinsic features, that are highly predictive of progression-free survival. When these models are combined, a hazard ratio of ~0.045 is achieved, demonstrating the ability of multi-modal digital pathology to generate high-performance and interpretable biomarkers.

INTRODUCTION

The microanatomy of fixed and stained tissues has been studied using light microscopy for over two centuries[1,2], and immunohistochemistry (IHC) has been in widespread use for 50 years[3]. Histopathology review of hematoxylin and eosin (H&E) stained tissue sections, complemented by IHC and exome sequencing, remains the primary approach for diagnosing and managing many diseases, particularly cancer[4]. More recently, a range of computational methods have been developed to automatically extract information from H&E images[5] and the use of machine learning and artificial intelligence approaches (ML/AI) is leading to rapid progress in computer-assisted diagnosis[6]. However, the images in current digital pathology systems – acquired from conventional histology and IHC methods – generally lack the molecular precision and depth of quantitative analysis needed to optimally predict outcomes, guide the selection of targeted therapies, and enable research into the molecular mechanisms of disease (see Wharton et al. for a thorough review)[7].

The transition from H&E-based histopathology to digital technologies[8] is occurring concurrently with the introduction of methods for obtaining 10-80-plex data from fixed tissue sections (e.g., MxIF, CyCIF, CODEX, 4i, mIHC, MIBI, IBEX, and IMC[9–15]). These high-plex imaging methods enable deep morphological and molecular analysis of normal and diseased tissues from humans and animal models[12,16–19] and generate spatially resolved information that is an ideal complement to other single cell methods, such as scRNA sequencing. Whereas some imaging methods require frozen samples, those that are compatible with formaldehyde-fixed and paraffin-embedded (FFPE) specimens – the type of specimens universally acquired for diagnostic purposes – make it possible to tap into large archives of human biopsy and resection specimens[20,21]. Moreover, whereas many high-plex imaging studies to date involve tissue microarrays (TMA; arrays of many 0.3 to 1 mm specimens on a single slide) or the small fields of view characteristic of mass-spectrometry based imaging[9,11], whole-slide imaging is required for clinical research and practice both to achieve sufficient statistical power[22] and as an FDA requirement[23].

Histopathology review of H&E images, a top-down approach, exploits prior knowledge about the cellular and acellular structures and morphologies associated with disease to analyze images[24]. In contrast, research using highly multiplexed imaging most commonly relies on a bottom-up approach in which cell types are enumerated and neighborhoods associated with disease are identified computationally[9,11]. A substantial opportunity exists to link these approaches in research and diagnostic settings, thereby combining standard clinical practice with single cell analysis of the tumor microenvironment. An ideal instrument for this purpose should have sufficient plex and resolution to distinguish tumor and immune cell types, enable efficient data acquisition with minimal human intervention, and, critically, allow the collection of same-cell high-quality H&E images for pathology review. A first-principles analysis suggests these requirements can be met with an instrument having 16-20 IF channels: 7-8 to subtype immune cells, 3-4 to detect and subtype tumor cells, 3-4 to identify relevant tissue structures, and 3-4 to examine specific tumor or therapeutic mechanisms (see **Extended Data Table 1** for example antibody panels) with the possibility of deeper analysis as needed in some cases.

The relative complexity of existing highly multiplexed imaging assays has prevented their wide adoption in the clinic; the current standard in clinical research is 5 to 6-plex imaging of tissue sections using a Perkin Elmer Vectra Polaris™ (now Akoya PhenoImager HT™) combined in some cases with H&E imaging of adjacent sections[25,26]. Achieving higher plex than this in a diagnostic setting will likely require parallel (one-shot) fluorescence acquisition rather than the sequential process developed by Gerdes et al.[10] and subsequently extended by our group[15] and others[27]. The unrealized possibility of visualizing the same cells with both H&E and >6 plex images would also facilitate analytical approaches that link molecular data to disease-associated histological features.

In this paper, we describe an approach to one-shot, whole-slide, 16 to 18-channel immunofluorescence (IF) imaging followed by H&E staining and imaging of the same tissue and then

explore its use in the generation of spatial biomarkers prognostic of tumor progression. Using FFPE specimens from multiple tumor types, we compare the performance of this multimodal "Orion™" method, and a commercial-grade instrument that implements it, to established IHC and cyclic data acquisition by CyCIF[28]. We show that joint analysis of H&E and IF same-section images substantially improves our ability to identify and interpret image features significantly associated with disease progression by facilitating the transfer of anatomical annotation from H&E images (e.g., distinguishing normal tissue from a tumor) while labeling H&E images using high-plex data. We also show that machine learning (ML) models generated from molecular analysis of high-plex IF images can be combined with ML of H&E images to aid in feature identification and interpretation (substantially extending previous data on joint analysis of molecular and H&E images)[29,30]. In combination, the top down and bottom up approaches generate potential biomarkers that are highly predictive of progression free survival (PFS) in a 40-patient cohort. Of note, our analysis involves a large amount of data by the standards of high-plex tissue imaging but the number of patients is too small for validation of a clinical test. Thus, the current work should be considered a proof-of-principles study; fortunately, the Orion method is scalable to the larger cohorts needed to test and validate biomarkers for clinical use as soon as these cohorts can be assembled.

RESULTS

Constructing and testing the Orion platform.

We investigated multiple approaches for achieving one-shot high-plex IF followed by H&E imaging of the same cells (i.e., from the same tissue section). Because eosin fluoresces strongly in the 530 - 620 nm range, it proved impractical to perform H&E staining prior to IF. However, board-certified pathologists confirmed that "clinical-grade" H&E images could be obtained after one or a small number of IF cycles when staining was performed using an industry-standard Ventana automated slide stainer

(or a similar machine from other vendors)[31]. An additional limitation of multiplexed fluorescence microscopy is that the overlap in excitation and emission spectra limits the number of fluorophores (typically to five to six) that can be accommodated within the bandpass useful for antibody labeling (~350 to 800 nm). This can be overcome using tuned emission and excitation filters and spectral deconvolution (typically of 6 - 10 channels)[32] or by dispersing emitted light using a diffraction grating and then performing linear unmixing[33,34]. However, unmixing of complex spectra (e.g., from an image stained with 10 or more fluorophores) has historically resulted in a substantial reduction in sensitivity and has not been widely implemented.

To develop the Orion platform, we tested >100 chemical fluorophores from different sources to identify 18 ArgoFluor™ dyes that were compatible with spectral extraction by discrete sampling based on the following properties: (i) emission in the 500 - 875 nm range; (ii) high quantum-efficiency; (iii) good photostability; and (iv) compatibility with each other in high-plex panels (**Extended Data Fig. 1a, Extended Data Table 2**). ArgoFluor dyes were covalently coupled to commercial antibodies directed against lineage markers of immune (e.g., CD4, CD8, CD68), epithelial (cytokeratin, E-cadherin), and endothelial (CD31) cells as well as immune checkpoint regulators (PD-1, PD-L1), and cell state markers (Ki-67), to generate panels suitable for studying the microenvironment and architecture of epithelial tumors and adjacent normal tissue (**Extended Data Fig. 1b**). An accelerated aging test demonstrated excellent reagent stability, estimated to be >5yr at -20ºC storage (**Extended Data Fig. 1c**).

With support from an NCI SBIR grant, a commercial-grade Orion instrument was developed. The instrument utilizes seven lasers (**Fig. 1a** and **Extended Data Fig. 1d**) to illuminate the sample and collects the emitted light with 4X to 40X objective lenses (0.2 NA to 0.95 NA). The system employs multiple tunable optical filters[35] that use a non-orthogonal angle of incidence on thin-film interference filters to shift the emission bandpass[36]. These filters have 90-95% transmission efficiency and enable collection of 10 - 15 nm bandpass channels with 1 nm center wavelength (CWL) tuning resolution over

a wide range of wavelengths (425 to 895 nm). Narrow bandpass emission channels improve specificity and the consequent reduction in signal strength is overcome by using excitation lasers that are ~10 times brighter than conventional LED illuminators and a sensitive scientific CMOS detector. Raw image files are processed computationally to correct for system aberrations such as geometric distortions and camera non-linearity[37], followed by spectral extraction to remove crosstalk, thereby isolating individual biomarker signals to one per imaging channel. The features of single-cells and regions of tissue are then computed using MCMICRO software[38]. The Orion instrument has an integrated brightfield mode, but the H&E images used in this study were also acquired using an Aperio GT450 microscope (Leica Biosystems), which is a gold standard for diagnostic applications[39] (**Fig. 1a**).

Validating high-plex one-Shot fluorescence imaging.

To test the Orion approach, we collected images from three sets of FFPE specimens: (i) human tonsils, a standard tissue for antibody qualification, (ii) 40 stage I-IV colorectal cancer (CRC) resections from the archives of the Brigham and Women's Hospital Pathology Department (key features of this cohort are described in **Extended Data Table 3**), and (iii) specimens of multiple tumor types available on TMA (**Extended Data Table 4**). We optimized the panel on tonsil and applied it successfully to this CRC cohort and to other tissue types represented on the TMA (see Methods). We included a dedicated autofluorescence channel (445 nm excitation / 485 nm emission, CWL) that provided valuable information on tissue morphology and components of connective tissue structures and blood vessels (**Fig. 1b**)[40]. This channel was also used to extract autofluorescence from the IF channels and improve biomarker signal to noise ratio (SNR). The images in this paper therefore represent 18-plex imaging (16 antibody channels, autofluorescence and nuclear stain) plus H&E. Inspection of extracted images revealed error-free whole-slide imaging of 1,000 or more adjacent tiles (**Fig. 1c**)[41] as well as bright in-focus staining of cellular and cellular substructures within each tile (**Fig. 1d**). To quantify the

effectiveness of spectral extraction, we imaged serial sections of human tonsil each stained with an individual antibody conjugated to a different ArgoFluor fluorophore and then recorded data in all channels. Prior to extraction, spectral cross talk between adjacent channels averaged ~35% and this was reduced 35-fold to <1% following spectral extraction (**Fig. 1e;** crosstalk among all channel pairs was reduced to <0.5%). When a tissue section was subjected to one-shot 16-plex antibody labeling, we observed "cross-talk" only for antibodies that stain targets co-localized on the same types of cells (e.g., co-staining of T-cell membranes by antibodies against CD3e and CD4 resulted in a high degree of pixel intensity correlation across these two channels; **Extended Data Fig. 1e).**

The staining patterns obtained by ArgoFluor-antibody conjugates were similar to those obtained by conventional IHC performed on the same specimen using the same antibody clones (see Du et al.[42] for details of approach; **Fig. 2a** and **Extended Data Fig. 2**). In addition, when adjacent tissue sections from CRC patients were imaged using Orion and the well-established method of cyclic immunofluorescence (CyCIF)[15] images looked similar and the fractions of cells scoring positive for identical markers were highly correlated (**Fig. 2b, 2c** and **Extended Data Fig. 2b** shows four examples with $\rho = 0.8$ to 0.9). Furthermore, projection of the high dimensional Orion data using t-SNE successfully resolved multiple immune and tumor cell types (**Fig. 2d** and **Extended Data Fig. 2c**). We conclude that the Orion method generates single cell data that are qualitatively and quantitatively similar to data generated by established methods such as IHC and CyCIF. Moreover, Orion is compatible with a range of tumor types including carcinomas, lymphomas, gliomas, and sarcomas as well as normal and non-neoplastic disease tissues (**Extended Data Fig. 3a**).

There are situations in which 16-20 fluorescent channels are likely to be insufficient. We therefore performed CyCIF after Orion imaging (but prior to H&E staining). We found that the standard CyCIF signal reduction ("bleaching") procedure[15] reduced ArgoFluor intensity by over 95% on average (**Extended Data Fig. 3b**), enabling the collection of multiple rounds of multiplexed imaging data after

the original Orion imaging round (**Fig. 2e**). As additional antibody panels become available, it will also be possible to perform sequential rounds of high-plex Orion imaging, although the quality of H&E images will decrease as the number of IF cycles increases.

Integrated analysis of IF and H&E images

The Orion method enables same-cell H&E and IF comparison (**Fig. 3a**), as opposed to existing methods that require use of adjacent tissue sections. We found that molecular labels obtained from IF enabled more complete enumeration of lymphocytes than inspection of H&E images by trained pathologists alone; for example, cells in CD4, CD8 T cell and B cell lineages are similar by H&E but clearly distinguishable by IF (arrows in **Fig. 3a**). Conversely, some cells and cell states were more readily defined in H&E images based on morphologic features than by immunofluorescence markers; this included eosinophils, neutrophils that could not be subtyped in IF images but whose morphology is highly characteristic in H&E data, as well as the prophase, metaphase, anaphase and telophase stages of mitosis (arrows and dashed lines in **Fig. 3b**). To quantify the amount of complementary information in H&E and IF images, we computed the number of cells (as identified by nuclear segmentation) in the 40-specimen CRC dataset that could not be assigned a clear identity using IF images and found that it varied from 6.5 to 42% of total nuclei (median = 16%) (**Fig. 3c**). We have previously observed a similar fraction of "unidentifiable" cells even with 40-60 plex CyCIF imaging[22] and surmised that these cells are either negative for all antibody markers included in the panel or have morphologies that are difficult to segment[43]. We therefore used a previously published ML model trained on H&E data[44] to identify those cells missing labels in Orion IF images (see Methods for details of this model and its performance) and found that >50% were predicted to be smooth muscle, stromal fibroblasts, or adipocytes (**Fig. 3d**). These predictions were confirmed by visual inspection of the H&E images by pathologists (**Fig. 3e**). We also identified specimens (e.g., from patient 26, **Fig. 3f** and **Extended Data Fig. 3c**) in which a region

of epithelium was weakly stained by pan-cytokeratin, E-cadherin, and immune markers making the cells difficult to identify by IF. Inspection of H&E images showed that these cells corresponded to a serrated adenoma whereas nearby invasive low-grade adenocarcinoma cells stained strongly for pan-cytokeratin and E-cadherin. Differential staining of cytokeratin isoforms in serrated adenoma and adenocarcinoma has been described previously[45] and in specimen C26 we speculate it may also reflect clonal heterogeneity. We conclude that the availability of H&E and IF images of the same set of cells substantially increases the fraction of cell types and states that can be identified as compared to either type of data alone. This is particularly true of cells for which specific molecular markers do not exist (e.g., stromal fibroblasts) or are lost due to tumor sub-clonality[22] as well as cells that are highly elongated or have multiple nuclei and are thus difficult to segment.

Identifying tumor features predictive of disease progression.

The classification of cancers for diagnostic purposes using American Joint Committee on Cancer (AJCC/UICC-TNM classification) criteria is based primarily on tumor-intrinsic characteristics (tumor, lymph node, and metastases, the TNM staging system)[46]. However, the extent and type of immune infiltration also plays a major role in therapeutic response and survival[47]. In colorectal cancer (CRC) this has given rise to a clinical test, the Immunoscore®[48], that is predictive of disease progression in multicenter cohort studies (as measured by progression-free survival, PFS, or overall survival, OS) and of time to recurrence in stage III cancers in a Phase 3 clinical trial[49]. The Immunoscore uses IHC to evaluate the number of CD3 and CD8-positive T cells at the tumor center (CT) and the invasive margin (IM; for Immunoscore this is defined as a region encompassing 360 μm on either side of the invasive boundary; in this work it is ± 100 μm from the boundary)[50,51]. The hazard ratio (HR; the difference in the rate of progression) between patients with tumors containing few immune cells in both the CT and the IM (Immunoscore = 0) and patients with tumors containing many cells in both compartments

(Immunoscore = 4) has been reported to be 0.20, (95% CI 0.10–0.38; $p < 10^{-4}$) in a Cox regression model, with increasing score correlating with longer survival[52]. This is a clinically significant difference that can be used to inform key treatment decision: for example, whether or not to deliver chemotherapy following surgery (i.e., adjuvant therapy)[53].

Using Orion data, we developed an automated method to recapitulate key aspects of the Immunoscore using PFS as measure of survival. First, we detected the tumor-stromal interface and generated masks that matched the criteria for CT and IM (± 100 μm around the tumor boundary; **Fig. 4a**). CD3 and CD8 positivity in single cells was determined by Gaussian Mixture Modeling[54] with the median positive fraction for each marker (CD3 or CD8) in each region (CT or IM) across all 40 CRC cases used as the cutoff for assigning a subscore of 0 or 1; the sum of the four subscores was used as the final score for Image Feature Model 1 (IFM1; **Fig. 4b**). The scoring method was intentionally simplified to avoid a need for tuning of adjustable parameters but nonetheless yielded a HR = 0.209 (95% CI 0.094-0.465; $p = 10^{-4}$) (**Fig. 4c**), similar to Immunoscore itself. Next, we used the underlying logic of Immunoscore to leverage multiple Orion channels. A total of 13 immune focused markers were used to generate ~15,000 marker combinations (IFMs), each composed of four markers within the CT and IM domains (**Fig. 4d**). Scores for each CRC case were binarized into high and low scores based on median intensities. When HRs were calculated we found that nearly 2,500 IFMs exceeded IFM1 in performance (**Extended Data Fig. 4a, 4b,** and **4c**). The optimal model (IFM2) exhibited an HR = 0.0785 (95% CI: 0.036-0.172, $p = 2 \times 10^{-06}$) (**Fig. 4d** and **4e**) and comprised the fractions of α-SMA^{+} cells in the CT, and CD45^{+}, PD-L1^{+} and CD4^{+} cells in the IM. Leave-one-out resampling showed that IFM2 was significantly better than IFM1 and demonstrated stable ranking with respect to HR ($p = 3.4 \times 10^{-14}$; adjusted p value based on the Benjamini-Hochberg Procedure $p_{adj} = 5.01 \times 10^{-9}$). 500-fold bootstrapping also confirmed a distribution of hazard ratios for IFM2 that was significantly lower than for IFM1 (**Fig. 4f**).

Histologic review of H&E images showed that IFM2-high tumors that exhibited slow progression (e.g., patients C34) had extensive lymphohistiocytic chronic inflammation including large lymphoid aggregates and tertiary lymphoid structures (TLS) at the tumor invasive margin (so-called "Crohn's-like lymphoid reaction")[55], whereas IFM2-low tumors had relatively few lymphoid aggregates and no TLS (e.g., patients C09) (**Fig. 4g** and **Extended Data Fig. 4d**). IFM2-low tumors were also more invasive than IFM2-high tumors but scoring was independent of histologic subtypes (e.g., conventional vs. mucinous morphology) and did not correlate with histologic grade (low vs. high grade carcinoma). Thus, IFM2 is likely to capture hyperactivity of the immune microenvironment around the invasive tumor margin and potential inactivation of tumor-associated fibroblasts. More generally, we conclude that Orion data can be used to automate previously described image-based biomarkers based on single-channel IHC and identify new marker combinations that significantly outperform them (see limitations sections for further discussion of the differences between the case number used in this paper and the number required for clinical validation of a biomarker).

Identifying new progression markers.

As an unbiased bottom-up means of identifying new progression models, we used Latent Dirichlet Allocation (LDA)[56], a probabilistic modeling method that reduces complex assemblies of intermixed entities into distinct component communities ("topics"). LDA is widely used in text mining and biodiversity studies and can detect recurrent arrangements of words or natural elements while accounting for uncertainty and missing data[57,58]. We separated CRC specimens into tumor and adjacent normal tissue using H&E data and an ML/AI model[44] and performed LDA at the level of individual IF markers on cells in the tumor region (**Fig. 5a**). This yielded 12 spatial features (topics) that recurred across the dataset (the number of topics was optimized by calculating the perplexity; see Methods for details) (**Extended Data Fig. 5a**). Visual inspection of images by a pathologist confirmed that marker

probabilities matched those computed for different topics and that the frequency distribution of each topic varied, sometimes substantially, among CRC samples (**Fig. 5b** and **Extended Data Fig. 5**). The strongest correlations between topics and PFS for the 40 CRC cohort were found to be -0.52 ($p < 0.001$) for Topic 7, comprising pan-cytokeratin and E-cadherin positivity, with little contribution from immune cells, and +0.57 ($p < 0.001$) for Topic 11, comprising CD20 positivity with minor contributions from CD3, CD4, and CD45 (**Fig. 5b-5f** and **Extended Data Fig. 5**). In contrast, topics involving the proliferation marker Ki-67$^+$ (Topic 6), PD-L1 positivity (Topic 9), or immune cells markers (CD45$^+$ or CD45RO$^+$; Topics 3 and 10) exhibited little or no correlation with survival (**Extended Data Fig. 5**).

Given the correlation of Topic 7 with PFS, we constructed a Kaplan-Meier curve for tumors having a proportion of Topic 7 below the 25th percentile versus those above this threshold (including all cells in the specimen) and observed HR = 0.24 (**Fig. 6a;** CI 95%: 0.10 – 0.54; $p < 10^{-3}$). Thus, LDA had discovered – via direct analysis of high-plex IF data – a tumor-intrinsic feature distinct from immune infiltration that was significantly associated with poor patient survival. One limitation of this, and many other models built using ML methods such as LDA is poor interpretability. In the case of Topic 7, the primary molecular features were pan-cytokeratin and E-cadherin positivity, but Topic 8 was similar in composition while exhibiting no correlation with PFS ($r = 0.01$; **Fig. 5c** and **Extended Data Fig. 5**). To identify the tumor histomorphology corresponding to these topics we transferred labels from IF to the same section H&E images, trained a convolutional neural network (CNN) on the H&E data, and inspected the highest scoring tumor regions (**Extended Data Fig. 6a**). In the case of Topic 7 these were readily identifiable as regions of poorly differentiated/high-grade tumor with stromal invasion (**Fig. 6a** and **6b**). In contrast, Topic 8 consisted predominantly of intestinal mucosa with a largely normal morphology and some areas of well-differentiated tumor (**Fig. 6b** and **Extended Data Fig. 6b).** When we inspected Orion and CyCIF images of specimens with a high proportion of Topic 7 (e.g., patient C06, **Extended Data Fig. 7**) we found that the E-cadherin to pan-cytokeratin ratios were low relative to

normal mucosa or Topic 8 (Na,K-ATPase expression was also low). These are features of cells undergoing an epithelial-mesenchymal transition (EMT), which is associated in CRC with progression and metastasis[59]. However, some features of EMT were not observed in Topic 7-positive cells: proliferation index was high (40-50% Ki67 and PCNA positivity) and staining for the EMT marker and transcription factor ZEB1 was low (when assessed using CyCIF data)[60]. Thus, even though the molecular and morphological features of Topic 7 were consistent with each other, H&E morphology was more readily interpretable with respect to long established features of CRC progression. It has been observed previously that interpretability increases confidence in a potential biomarker and substantially improves its chances of clinical translation[61].

Only about one-third of patients scored high for IFM1 and low for IFM3 (the combination correlated with the longest PFS; **Fig. 6d**), so we reasoned that it would be effective to combine the two models. Using the composite model (IFM4), we observed near perfect discrimination between progressing and non-progressing CRC patients with HR = 0.045; (95% CI = 0.021 to 0.098; $p = 1.4 \times 10^{-6}$) (**Fig. 6e**). This demonstrates that immunological and tumor-intrinsic features of cancers arising from top-down and bottom-up analysis can be effectively combined to generate prognostic models with high predictive value. Of note, no parameter tuning (e.g., setting thresholds for positivity) was involved in the generation of IFMs 1-3 or the highly performative IFM4 hybrid model. Experience with Immunoscore shows that parameter tuning using larger cohorts of patients can further boost performance.

DISCUSSION

In this paper, we describe an approach to multimodal tissue imaging that combines high-plex, subcellular resolution IF with conventional H&E imaging of the same cells. The approach required developing a new Orion platform whose staining and imaging workflow uniquely enables single-shot

high-plex IF data acquisition while preserving the sample for high-quality same-section H&E imaging. We show that multimodal tissue imaging has substantial benefits for human observers and machine-learned models; most obviously, it facilities the use of extensive historical knowledge about tissue microanatomy based on histopathological analysis of H&E images in the interpretation of molecular data derived from multiplexed molecular imaging. Moreover, human experts and ML algorithms can exploit H&E images to classify cell types and states that are not readily identifiable in multiplexed data. H&E and autofluorescence imaging are also effective at characterizing acellular structures that organize tissues at mesoscales (e.g., the elastic lamina of the vessel wall). Conversely, by overlaying molecular data on H&E images it is possible to discriminate cell types that have similar morphologies but different functions. The ability of molecular data to label cell types in H&E images should substantially improve supervised learning for ML/AI modeling[7,62] and the use of H&E data to analyze ML models trained on molecular data. The topic of "black box" versus interpretable AI is a major point of discussion in medicine in general[63], but in the case of pathology it is highly likely that interpretability will improve uptake, facilitate further research, and potentially improve reproducibility.

The Orion instrument currently supports up to 20-plex data acquisition (including DNA and one or more autofluorescence channels), but we find that 18-plex data collection is more robust – hence its use in this paper. It is nonetheless likely that several additional channels can be added to the method as we identify fluorophores more optimally matched to available lasers and optical elements. It is also possible to perform cyclic image acquisition (CyCIF) after Orion, increasing the number of molecular channels dramatically. However, H&E staining must be performed after all IF is complete, and H&E image quality goes down as the number of IF cycles increases. In the applications that we describe, implementing a performative image-based prognostic test required only a subset of the antibody channels and it is therefore likely that high-plex IF (possibly two cycles of Orion) will be most

important for exploratory and research studies and somewhat lower-plex imaging suitable for deployed image-based diagnostics, with attendant reductions in test complexity and cost.

It is not surprising that multiplexed molecular data from IF images adds information to H&E imaging. More surprising are the many cell types and structures that are difficult to identify in multiplexed images and readily identified in H&E images by histopathologists or the ML algorithms they train. This includes acellular structures, cell types for which good markers are not readily available, highly elongated and multi-nucleated cells that are difficult to segment with existing algorithms (e.g., muscle), and – most remarkably – tumor cells themselves. Many tumor types lack a definitive cell-type marker, and even when such markers are available, some cells in a tumor express these markers poorly likely due to sub-clonal heterogeneity. In contrast, pathologists are skilled at identifying dysplastic and transformed cells in H&E images. Therefore, H&E images are potentially more reliable than molecular images for the identification of some types of tumor cells. Conversely, many immune cell types cannot be reliably differentiated using H&E images, and their presence can also be difficult to discern when cells are crowded; the use of IF lineage markers provide critical new information in these cases.

The complementary strengths of H&E and IF imaging can be exploited by ML/AI algorithms that are increasingly used to process tissue images in clinical and research settings[62]. For example, we show that models trained to recognize disease-associated structures in H&E images, which is an area of intensive development in digital pathology[64], can improve the analysis and interpretation of multiplexed IF data. The converse is also true: IF images can be used to automatically label structures in H&E images (e.g., immune cell types) to assist in supervised learning on these images. This is a significant development because the labor associated with labeling of images – currently by human experts – is a major barrier to the development of better ML models. Combined H&E and IF images will be of immediate use in ML-assisted human-in-the loop environments that represent the state of the art in image interpretation in a research setting[65].

We find demonstrate the use of automated image processing on H&E and molecular data to identify image features prognostic of tumor progression[66]. For example, Immunoscore is a pathology-driven (top-down) clinical test that uses H&E and IHC data on the distribution of specific immune cell types at the tumor margin to predict outcome for patients with CRC. In this paper, we reproduced the logic of Immunoscore on a cohort of 40 CRC patients and using automated scripts show that it is possible to substantially improve upon it using additional immune markers (in terms of Hazard Ratios computed from PFS data; see limitations section below)[67]. In a distinct but complementary bottom-up approach, we used a spatially sensitive statistical model (LDA) of IF data to identify cell neighborhoods significantly associated with CRC progression. The top-performing feature in this case is tumor-cell intrinsic and is based on the distributions of cytokeratin and E-cadherin, two epithelial cell markers. Precisely why this is a prognostic feature is unclear from the IF data alone: other features involving similar markers are not predictive. However, inspection of corresponding H&E data (and training of an ML model) showed that LDA had identified local tumor morphologies typical of poorly differentiated/high-grade tumor with stromal invasion, increasing our confidence in the model. Because the features in the tumor-intrinsic model were distinct from and uncorrelated with the immune markers in Immunoscore, combining the two sets of features significantly improved the hazard ratio relative to either model used alone. We therefore anticipate that many opportunities will emerge for joint learning from H&E and IF data using adversarial, reinforcement, and other types of ML/AI modeling for research purposes, development of novel biomarkers, and analysis of clinical H&E data at scale[6]. The immediate availability of Orion as a commercial platform and our use of open-source software and OME (Open Microscopy Environment)[68] and MITI (Minimum Information about Tissue Imaging)[69] compliant data standards makes the approach we describe readily available to other investigators.

Limitations

The images in this paper represent one of the largest datasets collected to date using high-plex IF methods – 40 whole-slide CRC sections (representing 7.8 x 10^5 individual image tiles and ca. 6.2 x 10^7 segmented cells) and the only high-plex multimodal image collection currently available. However, the prognostic image feature models (IFMs) that we derive from these data cannot not be regarded as validated biomarkers or clinical tests[70]. Systematic metanalysis has identified a range of factors that negatively impact the reliability and value of prognostic biomarkers[71], particularly those based on new technology and multiplexed assays[72]. In the current work, specific limitations relative to REMARK recommendations[73] include a relatively small cohort size, the absence of pre-registeration[74], the acquisition of specimens from a single institution, and the use of leave-one out cross-validation rather than validation on an independent cohort. In particular, given the limited number of specimens analyzed in the current study as compared to conventional practice in histopathology-based biomarker studies (in which 500-1,000 cases is not uncommon) we are not able to fully control for all relevant covariates (e.g., depth of invasion, age, race, clinical stage etc.). These and other concerns will be addressable as we gain access to larger and more diverse collections of tissue blocks from which fresh sections can be cut and multi-modal imaging performed. With all of the advantages attendant to automated data acquisition and ML-based image analysis we anticipate that it will be feasible to progress in a few years to validated clinical tests that can be added to colorectal cancer treatment guidelines[53], substantially improving opportunities for personalized therapy.

ACKNOWLEDGEMENTS

This work was supported by NCI grants U54-CA225088 and U2C-CA233262 (PKS, SS), an NCI SBIR small business grant to RareCyte and PKS (R41-CA224503), and commercial investment from RareCyte; data processing software was developed with support from a Team Science Grant from the Gray Foundation and Ludwig Cancer Research (PKS, SS). SS is supported by the BWH President's

Scholars Award. We are grateful to all members of the HMS Laboratory of Systems Pharmacology (LSP) engaged in tissue imaging (see https://www.tissue-atlas.org/), to Joe Victor, and to members of the RareCyte software and hardware development teams.

AUTHOR CONTRIBUTIONS

J.R.L, Y.C., D.C., J.C., and E.M performed experiments and imaging. J.R.L., Y.C., D.C., J.C., S.C., C.Y., S.R., and T.G. performed data analysis. P.K.S., S.S., T.G., J.R.L., Y.C., and J.B.T. wrote the paper and all authors reviewed drafts and the final manuscript. J.B.T., J.R.L., and Y.C. prepared the figures. K.L.L., S.J.R., and S.S. supervised clinical research, and S.R., T.G., S.S., and P.K.S. supervised the overall research.

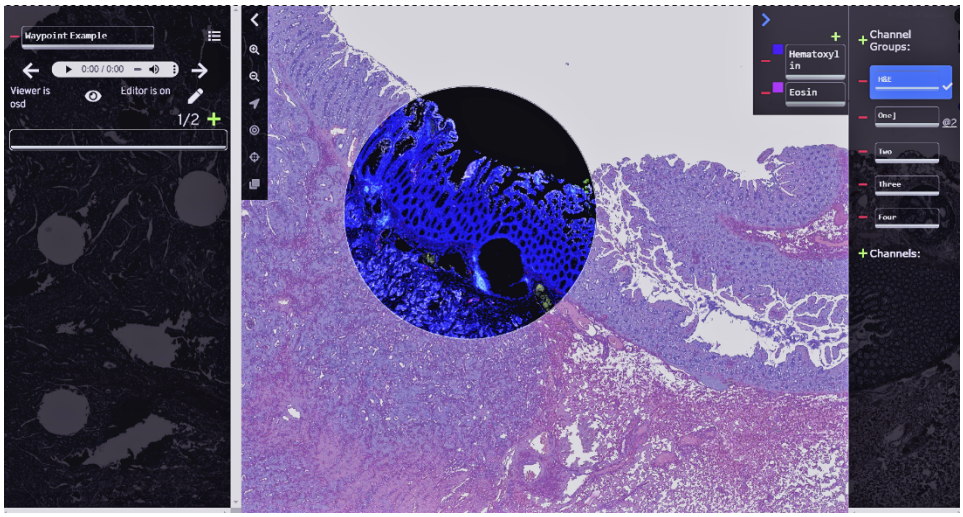
DECLARATION OF INTERESTS

PKS is a co-founder and member of the BOD of Glencoe Software, a member of the BOD for Applied Biomath, and a member of the SAB for RareCyte, NanoString, and Montai Health; he holds equity in Glencoe, Applied Biomath, and RareCyte. PKS is a consultant for Merck and the Sorger lab has received research funding from Novartis and Merck in the past five years. YC is a consultant for RareCyte. DC, JC, EM, SR, and TG are employees of RareCyte. The DFCI receives funding for KLL's research from the following entities: Amgen, Travera, and X4. DFCI and KLL have patents related to molecular diagnostics of cancer. SJR receives research support from Bristol-Myers-Squibb and KITE/Gilead. SJR is on the Scientific Advisory Board for Immunitas Therapeutics. The other authors declare no outside interests.

DATA AVAILABILITY (FOR REVIEWERS – TO BE UPDATED UPON PUBLICATION)

In keeping with the policies of the NCI Human Tumor Atlas Network (HTAN), all primary image and feature data described in this manuscript will be available via the HTAN data portal at https://htan-portal-nextjs.vercel.app/. However, it currently takes several months for data to appear on this portal; we

have therefore made all data available via S3 and GitHub using links found at https://labsyspharm.github.io/orion-crc/. Reviewers are directed in particular to images in the cloud-based MINERVA viewer with which it is possible to zoom and pan on H&E and high-plex data; we include a test implementation of an interactive lens that makes it possible to switch between H&E and Orion IF data (screenshot below).We expect this tool to be available for all of the data in the current paper by the time it goes to press.



Browser-based MINERVA image viewer with an H&E to IF interactive lens. See Rashid et al. for more information on the software (DOI: 10.1038/s41551-021-00789-8). The lensing feature is in development but will be available for all data in this manuscript, enabling “Google Maps” style interaction with the data without download or special software.

MATERIALS AND METHODS

Ethics and tissue cohort

Our research complies with all relevant ethical regulations and was reviewed and approved by the Institutional Review Boards (IRB) at Brigham and Women's Hospital (BWH), Harvard Medical School (HMS), and Dana Farber Cancer Institute (DFCI). Formaldehyde-fixed and paraffin-embedded (FFPE) tissue samples were used after diagnosis and informed written patient consent under Dana-Farber Cancer Institute IRB protocol 17-000. The study is compliant with all relevant ethical regulations regarding research involving human tissue specimens.

Tissue preparation.

Blocks of FFPE tonsil (AMSBIO, cat# 6022CS) and lung adenocarcinoma (AMSBIO, cat# 28004) and colorectal adenocarcinoma from the BWH Pathology Department archives were cut at 5 µm thickness using a rotary microtome and the sections were mounted onto Superfrost™ Plus microscope glass slides (Thermo Fisher, Catalog No.12-550-15). The slides were dried at 37°C overnight and baked at 59°C for one hour. Slides were stored at 4°C until use.

Fluorophores for Orion™ imaging.

The Orion™ instrument is designed to work with an optimized set of fluorophores from RareCyte, branded as ArgoFluor™ dyes whose emission peaks cover the spectrum from green to far-red (**Extended Data Table 2**). Although the instrument can also be used with other commercially available dyes, the ArgoFluor™ dyes have been strategically chosen based on a combination of properties that include resistance to photobleaching, narrow excitation and emission spectra, and high quantum efficiency. To date, the company has optimized 18 ArgoFluor™ dyes, with others in development.

Immunofluorescence antibodies.

Antibodies were obtained in carrier-free PBS and conjugated directly to either biotin for α-SMA, digoxygenin for pan-cytokeratin or to ArgoFluor™ dyes (RareCyte, Inc.) using amine conjugation chemistry. After determining labeling efficiency using absorbance spectroscopy, the conjugated antibodies were diluted in PBS-Antibody Stabilizer (CANDOR Bioscience GmbH, Catalog No. 130050) to a concentration of 200 µg/mL. Antibodies used in immunofluorescence studies are listed in the **Extended Data Table 2**.

Immunofluorescence staining.

Slides were de-paraffinized and subjected to antigen retrieval for 5 minutes at 95°C followed by 5 minutes at 107°C, using pH8.5 EZ-AR 2 Elegance buffer (BioGenex, Catalog No. HK547-XAK). To reduce tissue autofluorescence, slides were placed in a transparent reservoir containing 4.5% H_2O_2 and 24 mM NaOH in PBS and illuminated with white light for 60 minutes followed by 365 nm light for 30 minutes at room temperature as previously described[15]. Slides were rinsed with surfactant wash buffer (0.025% Triton X-100 in PBS), placed in a humidified stain tray, and incubated in Image-iT™ FX Signal Enhancer (Thermo Fisher, Catalog No. I36933) for 15 minutes at room temperature. After rinsing with surfactant wash buffer, the slides were placed in a humidity tray and stained with the panel of fluor- and hapten-labeled primary antibodies in PBS-Antibody Stabilizer (CANDOR Bioscience GmbH, Catalog No.130 050) containing 5% mouse serum and 5% rabbit serum for 2 hours at room temperature. Slides were then rinsed again with surfactant wash buffer and placed in a humidified stain tray and incubated with Hoechst 33342 (Thermo Fisher Catalog no. H3570), ArgoFluor™ 845 mouse-anti-DIG, and ArgoFluor™ 875-conjugated streptavidin in PBS-Antibody Stabilizer containing 10% goat serum for 30 minutes at room temperature. The slides were then rinsed a final time with surfactant wash buffer and PBS, coverslipped with ArgoFluor™ Mounting Media (RareCyte, Inc.) and dried overnight.

ArgoFluor™-antibody conjugate stability testing.

Antibody accelerated-aging studies were performed to determine ArgoFluor™-antibody conjugation stability. Reagent stability was measured using the ratio of quantitative metrics obtained with the accelerated condition (21.6ºC) to those obtained with the storage condition (-20ºC). Tissue validation (Orion IF): Single-cell mean fluorescence intensity (MFI) data obtained by imaging FFPE tonsil stained with the ArgoFluor™ conjugate was gated using a Gaussian mixture model to obtain the percent of positive cells and S:B values (S and B refer to the MFI of cells with values above (S, Signal) and below (B, Background) the gated threshold). Fluor stability (Orion IF): Single bead MFI data was obtained by imaging Ig-capture beads incubated with (S) or without (B) the ArgoFluor™ conjugate. Binding stability (Flow Cytometry): Intensity data from peripheral blood mononuclear cells (PBMC) stained with the ArgoFluor conjugated antibody was manually gated to obtain % Positive and S:B values (S and B refer to the MFI of cells with values above (S) and below (B) the gated threshold).

The Orion method and instrumentation.

The Orion instrument was designed with the following performance goals: (1) whole-slide imaging; (2) rapid single-pass data collection; (3) sub-cellular imaging resolution; (4) sufficient immunoprofiling depth; (5) bright-field imaging; (6) optical and mechanical stability for accurate image tile stitching; and (7) compatibility with established image data standards and formats. ArgoFluor™-conjugated antibodies along with Hoechst dye and tissue autofluorescence were excited by seven laser lines, ranging from 405 to 730 nm (**Extended Data Table 2**). To separate the overlapping emission spectra, images were captured through a set of nine bandpass filters, which can achieve a tunable narrow band detection window (10 - 15 nm) throughout the spectrum from 425 nm to 894 nm. For a specific sample, the detection bands were chosen to optimize color separation, implemented with RareCyte Inc.'s Artemis™ software. Tuning of these filters is based on the well-known fact that the spectrum of a thin-film interference filter shifts toward shorter wavelengths when the angle of incidence shifts away from 0

degrees (orthogonal to the filter surface). The filters were motorized such that any narrow band of 10 - 15 nm can be achieved across the entire fluorescence spectrum. Narrow bandpass emission channels improve specificity; the resulting lower signal is overcome by using high power excitation lasers, which yield power at the sample plane ranging from 270 to 600 mW, more than 10 times greater than a typical fluorescence microscope.

One-shot antibody IF imaging with the Orion instrument.

Whole slides were scanned using the Orion instrument using acquisition settings optimized for the specific antibody panels. Briefly, acquisition channel parameters were defined for each biomarker plus an additional channel dedicated to tissue autofluorescence, and included excitation laser, emission center wavelength (CWL), and exposure times. The nuclear channel was scanned at low resolution to identify tissue boundaries, followed by surface mapping at 20x to find the tissue in the z-axis. Whole tissue was acquired at 20x following the surface map within the specified tissue boundaries by collecting all channels for a single field of view (FOV) before proceeding to the next partially overlapping FOV. Raw image files were processed to correct for system aberrations, then signal from individual targets were isolated to separate channels using the Spectral Matrix obtained with control samples, followed by stitching of FOVs to generate a continuous open microscopy environment (OME) pyramid TIFF image.

Same Section H&E staining and imaging.

After Orion imaging was complete, slides were de-coverslipped by immersion in 1x PBS at 37°C until the coverslips fell away from the slide. Slides were rinsed in distilled water for 2 minutes, then stained by a routine regressive H&E protocol using Harris Hematoxylin (Leica, Catalog No. 3801575) and alcoholic eosin Y (Epredia, Catalog No. 71211). Coverslipping was performed with toluene-based mounting media (Thermo Scientific, Catalog No. 4112). After drying for 24 hours, slides were scanned on an Orion system in brightfield mode, using the same scan area used for IF image acquisition. H&E

images were also acquired using an Aperio GT450 microscope (Leica Biosystems), and the H&E images were registered to the IF images using ASHLAR[41] and PALOM software (https://github.com/Yu-AnChen/palom).

Pathology annotation of H&E images performed after Orion immunofluorescence imaging.

H&E images were annotated by a board-certified anatomic pathologist (SC and SS). The histologic features of each tissue section were defined and labeled in OMERO PathViewer software on whole slide images according to morphologic criteria[75] including normal mucosa, hyperplastic mucosa, adenomatous mucosa (tubular or serrated), invasive adenocarcinoma (tumor), lymphovascular invasion (LVI), peri-neural invasion (PNI), secondary lymphoid structures/Peyer's patches (SLS), tertiary lymphoid structures (TLS), lymphoid aggregates (without identifiable germinal center formation), lymph nodes. Tertiary lymphoid structures were morphologically defined by the presence of a lymphoid aggregate with germinal center formation and an anatomic distribution and appearance inconsistent with a secondary lymphoid structure (Peyer's patch or lymph node).

CyCIF imaging.

Tissue-based cyclic immunofluorescence (CyCIF) was performed as previously described[15] following the methods available in protocols.io (dx.doi.org/10.17504/protocols.io.bjiukkew). Data from specimens C1-C17 was acquired as previously reported[22] and computed cell counts were compared in this study with cell counts derived from Orion images of adjacent sections from the same specimens. A BOND RX Automated Slide Stainer was used to bake FFPE slides at 60°C for 30 minutes. Dewaxing was performed using Bond Dewax solution at 72°C, and antigen retrieval was performed using BOND Epitope Retrieval Solution 1 (Leica Biosystems) at 100°C for 20 minutes. Slides then underwent multiple cycles of antibody incubation, imaging, and fluorophore inactivation to perform the CyCIF process. All antibodies were incubated overnight at 4°C in the dark. Slides were stained with Hoechst

33342 for 10 minutes at room temperature in the dark following antibody incubation in every cycle. Coverslips were wet-mounted using 200µL of 10% Glycerol in PBS prior to imaging. Images were taken using a 20x objective (0.75 NA) on a CyteFinder™ slide scanning fluorescence instrument (RareCyte Inc. Seattle WA). Fluorophores were inactivated by incubating slides in a 4.5% H_2O_2, 24mM NaOH in PBS solution and placing under an LED light source for 1 hr. For CyCIF after Orion imaging, slides were immersed in 1x PBS at 37°C until the coverslips fell away from the slide. The standard CyCIF method was subsequently performed on these slides.

Immunohistochemistry.

FFPE sections were de-paraffinized, dehydrated, and endogenous peroxidase activity was blocked. Antigen retrieval was performed for 20 minutes at 100°C, at pH9, using BOND Epitope Retrieval Solution 2 (Leica Biosystems). Detection was achieved using a Bond Polymer Refine Detection® DAB chromogen kit and counterstained with hematoxylin. Slides were scanned using a RareCyte CyteFinder instrument. Primary antibodies used in immunohistochemistry are listed in **Extended Data Table 2**.

Orion image processing data quantification

Image stitching and segmentation. Image data processing was performed using MCMICRO modules[38]. Briefly, stitched, registered, illumination and geometric distortion corrected images were generated by the Orion platform. Single-cell segmentation was performed with UNMICST2 and cell masks were generated by 5-pixel dilation of the nucleus masks. Mean intensity of each channel and morphological features were quantified for each cell masks. Image and data analysis was performed using customized scripts in Python, ImageJ and MATLAB. All code is available on GitHub (https://github.com/labsyspharm/orion-crc).

Analysis of channel crosstalk

Single-plex tonsil images. Tonsil FFPE sections stained with single antibody-ArgoFluor underwent standard acquisition and extraction process using the Orion instrument. The pixel intensities of all 18 channels from 17 samples were used to quantify bleed through of a given antibody-ArgoFluor complex to the other channels before and after spectral extraction.

18-plex tonsil image. Pearson's correlation coefficients between all channel pairs were computed using pixel intensities in the 18-plex tonsil image before and after spectral extraction.

Computational analysis of Orion images and derivation of image feature models

IFM computation from Orion data. IFM1 was designed to replicate the logic of the Immunoscore method and was calculated in a semi-automated manner using Orion data. In brief, quantitative data of tumor and immune markers (pan-cytokeratin, CD3e, and CD8a) were gated for high and low cells using a Gaussian Mixture Model (GMM) and confirmed by inspection. After gating, the pan-cytokeratin$^+$ cells were then used to generate tumor masks using a K-Nearest Neighbor (KNN) model (kernel size = 25 cells). The tumor margins were derived from tumor masks by expanding 100 microns in either direction from the point of stroma-tumor contact. The CD3$^+$ and CD8$^+$ fraction, defined as marker positive cells divided by the total of all successfully segmented cells of all types in either the tumor center (TC) or invasive margin (IM). Tumor and margins were enumerated independently in each sample. The median values of all samples were used as a cutoff to defined a subscore as follows: below the median scored as 0 and above the median scored as 1. The final IFM1 value was calculated by adding together the subscores for CD3 and CD8 positive cells in the TC and IM regions (see **Fig. 4b** for a flow diagram). The IFM1 score therefore ranged from 0 (CD3$^+$ and CD8$^+$ low in both regions) to 4 (CD3$^+$ and CD8$^+$ high in both regions). Similar logic was used to generated other combinations of IFMs. 13 selected immune markers (CD3, CD8, CD45, CD45RO, CD68, CD163, CD4, CD20, α-SMA, FOXP3, PD-1, PD-L1) were gated as described above, and 26 parameters (each marker in the tumor or tumor/stromal

interface regions) were generated. The complete combination of 4 out 26 parameters was tested against PFS days for Hazard Ratio (HR). IFM2 was the 3rd best IFM among those combinations, excluding the 1st and 2nd best combinations which had some of the same markers as IFM1 (i.e., CD3 and CD8); the difference in performance between the top performing models was insignificant.

Leave-one-out (LOO) test and bootstrapping analysis for IFM2. In the LOO test, the ranks of IFM1 and IFM2 were recalculated with the 40 set of samples (n = 39); each set left out one sample from the original cohort. The collections of ranks from IFM1 and IFM2 were then tested with pairwise t-test. For bootstrapping, the 500 set of randomly selected samples were used to recalculate the hazard ratios of IFM1 and IFM2 as described above. The collections of hazard ratios from IFM1 and IFM2 were then tested with the pairwise t-test. To adjust for multiple hypotheses, the Benjamini-Hochberg Procedure was used with FDR = 0.1.

*Latent Dirichlet Allocation for IFM3 and IFM4***.** Latent Dirichlet Allocation (LDA) was used to compute spatial neighborhoods as described[22]. First, each sample was divided into "grids" of 200 microns^2, and marker frequency was calculated in each grid. The summarized probabilities of all samples were then used to generate the LDA model with 12 topics using collapsed Gibbs sampling in MATLAB. The optimal topic number was determined via varying numbers (between 8 to 16) of topics and evaluating the goodness-of-fit by calculating the perplexity of a held-out set. After fitting a global LDA model, the individual samples were then applied with the same models to assign topics at the single-cell level.

Convolutional Neural Network to identify IFM3 in H&E images

The transfer learning of a GoogLeNet model was done as follows. First, the patch images of 224 x 224 pixels^2 were generated from post-Orion H&E images, and the LDA topics were assigned to each patch using Orion data. To exclude low confidence training data, only patches with more than 20 cells and the percentage of the dominant topic over 60% were used. The selected patches were than separated into training, validation, and test sets as the ratio 0.6:0.2:0.2. The training was done with MATLAB (version

2019b) and the results are shown in **Extended data Fig. 6a**. Scripts and training data are available at https://github.com/labsyspharm/orion-crc.

A publicly available DenseNet161 model (https://doi.org/10.1101/2021.12.23.474029) trained with the 100K CRC H&E dataset (https://doi.org/10.5281/zenodo.1214456) was used to classify the post-Orion H&E image patches (112 μm^2) for all the CRC samples. WSI patch prediction was performed with TIAToolbox v1.1.0 (https://doi.org/10.1101/2021.12.23.474029) on a Windows PC with Nvidia GeForce GTX 1080 graphics card and using batch size = 32. Model performance was reported as F_1 = 0.992. As described in the training dataset, there are 9 output classes: adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), colorectal adenocarcinoma epithelium (TUM). Scripts for reproducing the inference results can be found at https://github.com/labsyspharm/orion-crc).

Outcome analysis

Outcome analysis was performed using Kaplan-Meyer estimation and log-rank test utilizing the MatSurv function in MATLAB[76]. Cutoffs for IFM1, IFM2, and IFM3 were selected at the median value of the entire cohort, and cutoff for IFM4 were selected based on IFM1 & IFM3 as described. Hazard ratios and confidence intervals were calculated with the log-rank approach: HR = (Oa/Ea)/(Ob/Eb), where Oa & Ob are the observed events in each group and Ea & Eb are the number of expected events[78].

DATA AVAILABILITY (AT PUBLICATION – SEE INFORMATION FOR REVIEWERS ABOVE)

Data used in the preparation of this manuscript are detailed in the Source Data file provided with the manuscript. All image and derived data are available without restriction via the NCI Human Tumor

Atlas Network (HTAN) Portal (https://htan-portal-nextjs.vercel.app/) in accordance with NCI Moonshot Policies.

CODE AVAILABILITY

Code used in this manuscript is available under an MIT open source license at the following repository: https://github.com/labsyspharm/orion-crc.

REFERENCES

1. Bock, O. A history of the development of histology up to the end of the nineteenth century. *Research* (2015).
2. Paget, S. THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. *The Lancet* **133**, 571–573 (1889).
3. Coons, A. H., Creech, H. J. & Jones, R. N. Immunological Properties of an Antibody Containing a Fluorescent Group. *Proceedings of the Society for Experimental Biology and Medicine* **47**, 200–202 (1941).
4. Robbins & Cotran Pathologic Basis of Disease - 9780323531139. *US Elsevier Health* https://www.us.elsevierhealth.com/robbins-cotran-pathologic-basis-of-disease-9780323531139.html.
5. Demir, C. & Yener, B. Automated cancer diagnosis based on histopathological images: A systematic survey. (2004).
6. Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. *Lab Invest* **101**, 412–422 (2021).
7. Wharton, K. A. *et al.* Tissue Multiplex Analyte Detection in Anatomic Pathology – Pathways to Clinical Implementation. *Frontiers in Molecular Biosciences* **8**, (2021).
8. Abels, E. *et al.* Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J. Pathol.* **249**, 286–294 (2019).
9. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).
10. Gerdes, M. J. *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11982–11987 (2013).

11. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).

12. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).

13. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, (2018).

14. Tsujikawa, T. *et al.* Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep* **19**, 203–217 (2017).

15. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* **7**, e31657 (2018).

16. Färkkilä, A. *et al.* Immunogenomic profiling determines responses to combined PARP and PD-1 inhibition in ovarian cancer. *Nat Commun* **11**, 1459 (2020).

17. Launonen, I.-M. *et al.* Single-cell tumor-immune microenvironment of BRCA1/2 mutated high-grade serous ovarian cancer. *Nat Commun* **13**, 835 (2022).

18. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

19. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).

20. Burger, M. L. *et al.* Antigen dominance hierarchies shape TCF1+ progenitor CD8 T cell phenotypes in tumors. *Cell* **184**, 4996-5014.e26 (2021).

21. Gaglia, G. *et al.* Temporal and spatial topography of cell proliferation in cancer. *Nat Cell Biol* **24**, 316–326 (2022).

22. Lin, J.-R. *et al.* Multiplexed 3D atlas of state transitions and immune interactions in colorectal cancer. *bioRxiv* 2021.03.31.437984 (2021) doi:10.1101/2021.03.31.437984.

23. Health, C. for D. and R. Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices. *U.S. Food and Drug Administration* http://www.fda.gov/regulatory-information/search-fda-guidance-documents/technical-performance-assessment-digital-pathology-whole-slide-imaging-devices (2019).
24. Weiser, M. R. AJCC 8th Edition: Colorectal Cancer. *Ann Surg Oncol* **25**, 1454–1455 (2018).
25. O'Meara, T. A. *et al.* Abstract P1-04-05: Multiplexed immunofluorescence staining of intra-tumoral immune cell populations and associations with immunohistochemical, clinical, and pathologic variables in breast cancer. *Cancer Research* **82**, P1-04–05 (2022).
26. Berry, S. *et al.* Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science* **372**, eaba2609 (2021).
27. Jones, J. A. *et al.* Oligonucleotide conjugated antibody strategies for cyclic immunostaining. *Sci Rep* **11**, 23844 (2021).
28. Lee, S. *et al.* Novel charged sodium and calcium channel inhibitor active against neurogenic inflammation. *Elife* **8**, e48118 (2019).
29. Burlingame, E. A., Margolin, A. A., Gray, J. W. & Chang, Y. H. SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks. *Proc SPIE Int Soc Opt Eng* **10581**, 1058105 (2018).
30. Prichard, J. W. *et al.* TissueCypher^{TM}: A systems biology approach to anatomic pathology. *J Pathol Inform* **6**, 48 (2015).
31. Hassell, L. A., Glass, C. F., Yip, C. & Eneff, P. A. The combined positive impact of Lean methodology and Ventana Symphony autostainer on histology lab workflow. *BMC Clin Pathol* **10**, 2 (2010).
32. McRae, T. D., Oleksyn, D., Miller, J. & Gao, Y.-R. Robust blind spectral unmixing for fluorescence microscopy using unsupervised learning. *PLOS ONE* **14**, e0225410 (2019).

33. Garini, Y., Young, I. T. & McNamara, G. Spectral imaging: principles and applications. *Cytometry A* **69**, 735–747 (2006).

34. Zimmermann, T. Spectral imaging and linear unmixing in light microscopy. *Adv Biochem Eng Biotechnol* **95**, 245–265 (2005).

35. Favreau, P. *et al.* Thin-film tunable filters for hyperspectral fluorescence microscopy. *J Biomed Opt* **19**, 011017 (2014).

36. Anderson, N., Beeson, R. & Erdogan, T. Angle-Tuned Thin-Film Interference Filters for Spectral Imaging. *Optics and Photonics News* **13**, 1–2 (2011).

37. Zeng, Z. *et al*. Computational methods in super-resolution microscopy. *Frontiers Inf Technol Electronic Eng* **18**, 1222–1235 (2017).

38. Schapiro, D. *et al.* MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nat Methods* **19**, 311–315 (2022).

39. Babawale, M. *et al.* Verification and Validation of Digital Pathology (Whole Slide Imaging) for Primary Histopathological Diagnosis: All Wales Experience. *J Pathol Inform* **12**, 4 (2021).

40. STEINER, K. Fluorescence Microscopy of Normal and Pathologic Keratin. *Archives of Dermatology* **82**, 352–361 (1960).

41. Muhlich, J., Chen, Y.-A., Russell, D. & Sorger, P. K. Stitching and registering highly multiplexed whole slide images of tissues and tumors using ASHLAR software. (2021) doi:10.1101/2021.04.20.440625.

42. Du, Z. *et al.* Qualifying antibodies for image-based immune profiling and multiplexed tissue imaging. *Nat Protoc* **14**, 2900–2930 (2019).

43. Yapp, C. *et al*. UnMICST: Deep learning with real augmentation for robust segmentation of highly multiplexed images of human tissues. 2021.04.02.438285 Preprint at https://doi.org/10.1101/2021.04.02.438285 (2022).

44. Kather, J. N. *et al.* Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOS Medicine* **16**, e1002730 (2019).

45. Tatsumi, N. *et al.* Expression of Cytokeratins 7 and 20 in Serrated Adenoma and Related Diseases. *Dig Dis Sci* **50**, 1741–1746 (2005).

46. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. *CA Cancer J Clin* **67**, 93–99 (2017).

47. Paijens, S. T., Vledder, A., de Bruyn, M. & Nijman, H. W. Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell Mol Immunol* **18**, 842–859 (2021).

48. Galon, J. *et al.* Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *The Journal of Pathology* **232**, 199–209 (2014).

49. Pagès, F., Taieb, J., Laurent-Puig, P. & Galon, J. The consensus Immunoscore in phase 3 clinical trials; potential impact on patient management decisions. *Oncoimmunology* **9**, 1812221.

50. Angell, H. K., Bruni, D., Barrett, J. C., Herbst, R. & Galon, J. The Immunoscore: Colon Cancer and Beyond. *Clin Cancer Res* **26**, 332–339 (2020).

51. Galon, J. *et al.* Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* **313**, 1960–1964 (2006).

52. Pagès, F. *et al.* International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet* **391**, 2128–2139 (2018).

53. Argilés, G. *et al.* Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* **31**, 1291–1305 (2020).

54. Pan, K., Kokaram, A., Hillebrand, J. & Ramaswami, M. Gaussian mixtures for intensity modeling of spots in microscopy. in 121–124 (2010). doi:10.1109/ISBI.2010.5490398.

55. Graham, D. M. & Appelman, H. D. Crohn's-like lymphoid reaction and colorectal carcinoma: a potential histologic prognosticator. *Mod Pathol* **3**, 332–335 (1990).

56. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).

57. Valle, D., Baiser, B., Woodall, C. W. & Chazdon, R. Decomposing biodiversity data using the Latent Dirichlet Allocation model, a probabilistic multivariate statistical method. *Ecology Letters* **17**, 1591–1601 (2014).

58. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).

59. Cao, H., Xu, E., Liu, H., Wan, L. & Lai, M. Epithelial-mesenchymal transition in colorectal cancer metastasis: A system review. *Pathol Res Pract* **211**, 557–569 (2015).

60. Zhang, P., Sun, Y. & Ma, L. ZEB1: at the crossroads of epithelial-mesenchymal transition, metastasis and therapy resistance. *Cell Cycle* **14**, 481–487 (2015).

61. Ludwig, J. A. & Weinstein, J. N. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* **5**, 845–856 (2005).

62. Granter, S. R., Beck, A. H. & Papke, D. J. AlphaGo, Deep Learning, and the Future of the Human Microscopist. *Arch. Pathol. Lab. Med.* **141**, 619–621 (2017).

63. Wang, F., Kaushal, R. & Khullar, D. Should Health Care Demand Interpretable Artificial Intelligence or Accept 'Black Box' Medicine? *Ann Intern Med* **172**, 59–60 (2020).

64. Bhinder, B., Gilvary, C., Madhukar, N. S. & Elemento, O. Artificial Intelligence in Cancer Research and Precision Medicine. *Cancer Discov* **11**, 900–915 (2021).

65. Evans, T. *et al.* The explainability paradox: Challenges for xAI in digital pathology. *Future Generation Computer Systems* **133**, 281–296 (2022).

66. Savadjiev, P. *et al.* Image-based biomarkers for solid tumor quantification. *Eur Radiol* **29**, 5431–5440 (2019).

67. Bruni, D., Angell, H. K. & Galon, J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer* **20**, 662–680 (2020).

68. Swedlow, J. R., Goldberg, I., Brauner, E. & Sorger, P. K. Informatics and quantitative analysis in biological imaging. *Science* **300**, 100–102 (2003).

69. Schapiro, D. *et al.* MITI minimum information guidelines for highly multiplexed tissue images. *Nat Methods* **19**, 262–267 (2022).

70. Research, C. for D. E. and. Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification Guidance for Industry. *U.S. Food and Drug Administration* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-histopathology-and-its-associated-methodologies-support-biomarker-qualification (2020).

71. Ntzani, E. E. & Ioannidis, J. P. A. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* **362**, 1439–1444 (2003).

72. Hemingway, H., Riley, R. D. & Altman, D. G. Ten steps towards improving prognosis research. *BMJ* **339**, b4184 (2009).

73. McShane, L. M. *et al.* Reporting Recommendations for Tumor Marker Prognostic Studies. *JCO* **23**, 9067–9072 (2005).

74. Warren, M. First analysis of ‘pre-registered’ studies shows sharp rise in null findings. *Nature* (2018) doi:10.1038/d41586-018-07118-1.

75. *Digestive System Tumours: WHO Classification of Tumours*. (World Health Organization, 2019).

76. Creed, J. H., Gerke, T. A. & Berglund, A. E. MatSurv: Survival analysis and visualization in MATLAB. *Journal of Open Source Software* **5**, 1830 (2020).

77. Schumacher, T. N. & Thommen, D. S. Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022).

Fig. 1 | Same-section immunofluorescence and H&E using the Orion™ Platform.

a, Schematic of one-shot 16 to 20-channel multiplexed immunofluorescence imaging with the Orion™ method followed by Hematoxylin and Eosin (H&E) staining of the same section using an automated slide stainer and scanning of the H&E-stained slide in transillumination (brightfield) mode. This method of discriminating the emission spectra of fluorophores is repeated using seven excitation lasers spaced across the spectrum (see **Extended Data Fig. 1a** and Methods section). Using polychroic mirrors and tunable optical filters, emission spectra are extracted to discriminate up 20 channels including signal from fluorophore-labelled antibodies (15-20 in most experiments), the nuclear stain Hoechst 33342, and tissue intrinsic autofluorescence. **b,** Left panels: Orion multiplexed immunofluorescence image showing CD31, α-SMA, Hoechst (DNA), and signal from the tissue autofluorescence channel (AF) from a colorectal cancer FFPE specimen (C04); this highlights an artery outside of the tumor region with red blood cells in the vessel lumen and elastic fibers in the internal and external elastic lamina of the vessel wall, numerous smaller vessels (arterioles), and stromal collagen fibers (inset displays arterioles). Right panels: images of the H&E staining from the same tissue section (histologic landmarks are indicated). Scalebars 50 µm. **c,** Orion multiplexed immunofluorescence image (showing CD45, pan-cytokeratin, CD31, and α-SMA) from a whole tissue FFPE section of a colorectal cancer (C04) and matched H&E from the same section. Holes in the images are regions of tissue ('cores') removed in the construction of TMAs. Scalebar 5 mm. **d,** Zoom-in views of the regions indicated by arrowheads in panel **c**; marker combinations indicated. Scalebar 20 µm. **e,** Intensities of fluorochromes (columns in heatmaps) in each Orion channel (rows in heatmaps) prior to (top) and after (bottom) spectral extraction. The extraction matrix was determined from control samples scanned using the same acquisition settings that were used for the full panel. The control samples included: unstained lung tissue (for the autofluorescence channel), tonsil tissue stained with Hoechst, and tonsil tissue stained in single-plex with ArgoFluor-

conjugates used in the panel (for the biomarker channels). The values in each column were normalized to the maximum value in the column.

Fig. 2 | Qualifying 16-plex single-shot Orion antibody panel.

a, Panels of images from FFPE tonsil sections showing single-antibody immunohistochemistry (IHC) for pan-cytokeratin, Ki-67, CD8a, CD163, and the matching channels extracted from 16-plex Orion immunofluorescence (IF) images (H&E stain was performed on the same section as the Orion imaging). Scalebars 50 µm. **b,** Orion IF images and cyclic immunofluorescence (CyCIF) images from neighboring sections of an FFPE colorectal adenocarcinoma; Scalebars 50 µm. The CyCIF images were done with 2x2 binning while Orion images were obtained with no binning. **c,** Plots of the fraction of cells positive for the indicated markers from whole slide Orion IF and CyCIF images acquired from neighboring sections from 16 FFPE colorectal cancer specimens. Pearson correlation coefficients are indicated. **d,** t-distributed stochastic neighbor embedding (t-SNE) plots of cells derived from CyCIF (left panels) and Orion IF images (right panels) of a FFPE colorectal cancer specimen (C01) with the fluorescence intensities of immune (CD45, pan-cytokeratin, CD8a, α-SMA) markers overlaid on the plots as heat maps. **e,** Same-slide Orion and CyCIF experiment. The tonsil samples were first processed with 16-panel Orion antibodies; PD-L1, CD4, CD8a, Ki-67, and α-SMA are shown. After imaging, fluorophores were inactivated by bleaching using the standard CyCIF protocol, then three-cycles of four-channel CyCIF staining and imaging were performed using the antibodies indicated.

Fig. 3 | Combined H&E and Orion to identify cell/tissue types.

a, Representative images of Orion IF and same-section H&E from an area of normal colon (from colorectal cancer resection specimen C02). Scalebars 50 µm. **b,** Cell types not specifically identified by markers in the Orion panel but readily recognized in H&E images including eosinophils, neutrophils,

and cells undergoing mitoses (selected cells of each type denoted by arrowheads and dashed lines). Scalebars 10 µm. **c,** Spatial maps of the positions of cells (~15% of total cells) that were not detected by the Orion IF panel in a colorectal cancer specimen overlaid onto the corresponding H&E image (specimen: C01); red dots denote cells with identifiable nucleus but not subtyped using the antibody panel. **d,** Upper panel: Spatial map of nine tissue classes determined from the H&E image using a convolutional neural network (CNN) model for various cell types as indicated[44]. Lower panel: Percent of total of "unidentifiable" cells assigned to a specific tissue class by the CNN applied to the H&E image. **e,** Example same-section Orion IF and H&E images from areas enriched for 'non-detected' cells; examples include areas predicted to be rich in stroma and smooth muscle; Scalebars 100 µm. **f,** Orion IF and H&E images from colorectal cancer resection specimen C26, showing an area of serrated adenoma with low pan-cytokeratin expression (markers as indicated). Whole slide image indicating the location of this region is shown in **Extended Data Fig. 3c.** Scalebars 300 µm.

Fig. 4 | Recapitulating and extending the Immunoscore tissue immune test using Orion images. **a,** Map of tumor center and invasive-margin compartments for specimen C04 overlaid on an H&E image with the density of $CD3^+$ cells shown as a contour map (yellow) and the positions of $CD8^+$ T cells as blue dots. The arrow indicates the zoom-in images shown below. Lower panel shows selected channels from a portion of the Orion image for C04 spanning the invasive boundary (denoted by green shading). **b,** Flow chart for the calculation of Image Feature Model 1 (IFM1) that recapitulates key features of the Immunoscore test. **c,** Upper panel: Box-and-whisker plots for progression-free survival (PFS) for 40 CRC patients based on actual IFM1 scores (midline = median, box limits = Q1 (25th percentile)/Q3 (75th percentile), whiskers = 1.5 inter-quartile range (IQR), dots = outliers (>1.5IQR) or scores stratified into two classes as follows, low: score ≤ 2, high: score = 3 or 4 (pairwise two-tailed t-test p = 0.002. Lower panel: Kaplan Meier plots computed using IFM1 binary classes (HR, hazards ratio; 95%

confidence interval; logrank p-value). **d,** Flow chart for calculation of additional models that use the underlying logic of Immunoscore but considering 13 markers. The image processing steps are the same as in panel *a*. The rank positions of IFM1 and IFM2 are shown relative to all other 14,950 combinations of parameters that were considered. **e,** (Left) Box-and-whisker plots for PFS for 40 CRC patients based on IFM2 scores. (Right) Kaplan Meier plots computed using IFM2 binary classes stratified into two classes as follows, low: score ≤ 2, high: score = 3 or 4 (HR, hazards ratio; 95% confidence interval; logrank p-value). **f.** Plots of leave-one-out cross-validation of ranks from IFM1 and IFM2 (unadjusted p = 3.35×10^{-14} and adjusted using the Benjamini-Hochberg Procedure; p=5.0×10^{-9}) and bootstrapping of hazard ratios from IFM1 and IFM2 (unadjusted p = 1.94×10^{-25} and adjusted p = 2.91×10^{-20}). Detailed analysis was described in the methods section and pairwise two-tailed t-test were used unless otherwise mentioned. **g,** Representative Orion IF images of cases with high IFM2 (score = 4 in specimen C34) and low IFM2 (score = 0 in specimen C09). IF images show DNA, pan-cytokeratin, α-SMA, CD45, and PD-L1; Scalebars 100 µm.

Fig. 5 | Bottom-up development of a tumor-intrinsic image feature model.

a, Positions in specimen C39 of three selected topics identified using Latent Dirichlet Allocation (LDA). Topic locations are overlaid on an H&E image; Scalebar 5 mm. **b,** Left: Markers making up selected LDA topics as shown with size of the text proportional to the frequency of the marker but with colored text scaled by 50% for clarity; Radar plot indicating the fraction of cells positive for each marker in Topic 7, 8, and 11 (data for all others topics show in **Extended Data Fig. 5**). **c,** Immunofluorescence images showing expression of pan-cytokeratin, α-SMA, CD45, and CD20 for the indicated LDA topics. The position of each image frame is denoted by the yellow boxes in panel a. Scalebars 100 µm. **d,** Pearson correlation plots of progression-free survival (PFS) and Fraction of Topic 7, 8 and 11 in 40 CRC patients. Topic 11 corresponded to TLS, whose presence is known to correlate with good

outcome[77]. **e,** Fraction of Topics 7, 8, and 11 in colorectal cancer specimens C1-C40. **f,** Box-and-whisker plots showing fractions of Topic 7, 8, and 11 positive cells for indicated markers; midline = median, box limits = Q1 (25th percentile)/Q3 (75th percentile), whiskers = 1.5 inter-quartile range (IQR), dots = outliers (>1.5IQR)). Pairwise t-test p values indicated.

Fig. 6 | LDA Topic 7 corresponds to aggressive tumor regions and is correlated with poor outcomes.

a, Kaplan Meier plots of PFS for 40 CRC patients based on the fraction of Topic 7 present in the tumor domain and stratified as follows: high class: over 11% of topic 7 cells over all cells; and low class: under 11% of topic 7 cells over all cells (HR, hazards ratio; 95% confidence interval; logrank p-value). **b,** Representative images of Topic 7 extracted from all specimens using a convolutional neural network (GoogLeNet) trained on LDA data. **c,** Spatial map of LDA Topic 7 and H&E image from colorectal cancer sample C02. **d,** Plot of fraction of Topic 7 (IFM3) versus IFM1 score for 40 CRC patients. **e,** Kaplan Meier plots of PFS for 40 CRC patients stratified using IFM4 which was binarized as follows: class 1: IFM1 high and Topic 7 (IFM3) low group; class 2: all other patients – i.e., either low IFM1 and/or high Topic 7 (IFM3) (HR, hazards ratio; 95% confidence interval; logrank p-value).

SUPPLEMENTARY INFORMATION

EXTENDED DATA FIGURE LEGENDS

Extended Data Fig. 1 | Features of the fluorophores, signal extraction, antibodies, and instrumentation used in the Orion™ Method.

a, Emission spectra of the ArgoFluor dyes used in this study with overlaid filter profiles. Each row shows fluorophores excited using the same laser (denoted by the colored vertical line). From left to right within each laser row: 405 laser (Hoechst 33342); 445 laser (Autofluorescence); 470 laser (ArgoFluor

515, ArgoFluor 555L); 520 laser (ArgoFluor 535, ArgoFluor 550); 555 laser (ArgoFluor 572, ArgoFluor 584, ArgoFluor 602, ArgoFluor 624, ArgoFluor 660L); 640 laser (ArgoFluor 662, ArgoFluor 686, ArgoFluor 706, ArgoFluor 730); 730 laser (ArgoFluor 760, ArgoFluor 795, ArgoFluor 845, ArgoFluor 875). For the 405-laser data collection, tonsil tissue stained with Hoechst 33342 was used as the sample. For the 445-laser data collection, unstained lung tissue was used. Single color Ig-capture beads generated by incubation with antibodies conjugated to the indicated ArgoFluor dye were used as the sample for all other collections. For each sample, data was collected into multiple Orion channels spanning a wide range of wavelengths (in 2 nm center wavelength increments). **b,** Channel images of FFPE tonsil section stained, imaged, and processed with Orion platform showing distinct spatial patterns with minimal channel crosstalk. **c,** Stability of fluorophore and of epitope recognition in solution and in tissues is shown for ArgoFluor 572 conjugated anti-CD4 antibody. Quantitative stability metrics were generated from three different assays to compare reagents stored at an accelerated aging condition (+21.6ºC) to reagents stored at the recommended condition (-20ºC) based on the Arrhenius equation (storage for 3.5 months at the accelerated aging condition is equivalent to 5 years at the recommended storage condition). *Fluorochrome stability*: The intensity of Ig-capture beads incubated with (signal) or without (background) antibody was measured from images scanned with the Orion system. The histogram overlay shows the intensity distribution for unlabeled beads (orange) and for beads incubated with antibody stored for 3.5 months at -20ºC (red) or +21.6ºC (blue). The mean fluorescence intensity (MFI) was obtained from these distributions, as well as the MFI signal-to-background (S:B) ratios. The dot plots show accelerated-to-real time CD4 MFI ratios (left plot) and S:B ratios (right plot) across 7 computed time points. *Antibody binding stability*: Human peripheral blood mononuclear cells (PBMC) were stained with accelerated-aged (blue) or real-time-aged (red) ArgoFluor 572 conjugated anti-CD4 antibody and analyzed using flow cytometry (3.5-month real-time / 5-year accelerated time point shown in histogram. The MFI was obtained for the positive (signal) and negative (background) populations,

allowing derivation of S:B ratios. The dot plots show accelerated-to-real time CD4 MFI ratios (left) and S:B ratios (right) across 7 time points. For tissue-based antibody stability testing, images of serial sections from FFPE tonsil stained with real-time aged (top)z and accelerated-aged (bottom) antibodies were obtained using the Orion system. Single cell segmentation and intensity measurements were obtained with QuPath software, and a Gaussian mixture model threshold was applied to determine positive cells (signal) from negative cells (background) to determine S:B for both conditions. These methods demonstrate equivalent performance for both storage conditions in the three assays (S:B ratio of 0.93 for the 3.5-month real-time / 5-year accelerated time point). **d,** Schematic of the Orion optical system. The Orion imaging system has fluorescence and brightfield imaging modes. *Fluorescence imaging*: The Orion system is a class 1 LASER product which uses 7-color LASER illumination (one at a time) to illuminate a sample on a microscope slide. The illumination beam emanates from the source in a fiber optic cable, then shaped with beam conditioning optics, and redirected via a beam splitter and path folding mirrors through an objective lens which focuses it onto the sample. Excitation light passing through the sample is stopped by a beam block preventing damage to the transmitted light source. Laser-excited fluorophores in the sample emit light that is collected by the objective lens. This light is redirected via the beam splitter and path folding mirrors through a tube lens for focusing, fixed and rotatable compensation elements for optical corrections, and a tunable emission filter prior to collection by a sCMOS camera. *Brightfield imaging*: The Orion system utilizes LED transillumination of the sample on the microscope slide. Chromogenic stains in the sample absorb a portion of the light, and the remainder is collected by the objective lens. The light follows the same path as the fluorescence emission described above, with the exception that a window is used instead of an emission filter. **e,** Validation of minimal channel crosstalk in 18-plex tonsil image after spectral extraction. Pearson's correlation coefficients between all channel pairs were calculated using the paired pixel intensities.

Square boxes with colored borders denote excitation lasers. High correlation coefficients were only found in channel pairs that contains target markers that are in close proximity.

Extended Data Fig. 2 | Qualifying 16-plex single-shot Orion antibody panel relative to immunohistochemistry and Cyclic Immunofluorescence (CyCIF).

a, Panels of images from FFPE tonsil sections showing single-antibody immunohistochemistry (IHC) for the indicated markers and matching channels extracted from the 16-plex Orion immunofluorescence (IF) images (H&E stain was performed on the same section as the Orion imaging). Scalebars 50 µm. **b,** Plots of the fraction of positive for the indicated markers (CD45, CD68, CD20, CD4) from whole slide Orion IF and CyCIF images acquired from neighboring sections from 16 FFPE colorectal cancer specimens. Pearson correlation coefficients are indicated. **c,** t-distributed stochastic neighbor embedding (t-SNE) plots of cells from Orion IF image (specimen: C01). Log transformed marker intensities (CD31, CD20, E-cadherin, Ki-67, pan-cytokeratin, α-SMA) were used to color the dots in each panel.

Extended Data Fig. 3 | Orion imaging of a different disease histologies and CyCIF following Orion imaging.

a, 16-plex (18 channel) Orion image from a tissue microarray (TMA) containing normal and diseased human tissues including inflammatory and neoplastic diseases (Examples highlighted are lung squamous cell carcinoma, prostate adenocarcinoma, ovarian cancer, and breast; DNA, pan-cytokeratin, KI-67, α-SMA, CD45 and CD31 are displayed. scalebars 2 mm and 400 µm, as indicated. **b,** Left panel: Orion image of normal colon showing E-cadherin, CD11b, CD45, CD163, Ki-67, and DNA (Sytox) signal. Right panel: same area of normal colon following inactivation of Orion fluorophores (see Methods). **c,** Orion IF image from colorectal cancer resection specimen C26, showing an area of serrated adenoma

with low pan-cytokeratin expression (markers as indicated). Higher magnification inset as indicated by the box is shown in **Fig. 3f.** Scalebar 3 mm.

Extended Data Fig. 4 | Assessment of individual markers to Image Feature Models of patient prognosis derived from Orion immunofluorescence images.

a, Upper: Ranking of 1/hazard ratio (HR) for each Image Feature Model (IFM1 to IFM14,950) calculated by determining the frequency of cells positive for one or more of 13 markers in Orion IF images lying within (tumor center: CT) or outside of a region 100 µm from the tumor invasive margin (IM) model (n = 40 patients). Ranking position of IFM1 is indicated. IFM2 showed an HR = 0.0785 (95% CI: 0.0358-0.172, p = 1.91 $x10^{-06}$). Lower: Full heat map showing the selected markers at the tumor or margin in each combination. A total 14,950 combinations were generated as the set of 4 out of 26 parameters (13 markers in 2 regions). **b,** Enrichment plots showing enrichment scores (ES) for positive cells denoted by the indicated markers (and their location in the tumor or at the tumor margin) based on the 16-plex Orion images, indicating whether the marker/location feature is enriched in the image feature models linked to the best hazard ratios. The green lines represent the running ES for a given marker/location as the analysis proceeds down the ranked list. The value at the peak is the final ES. **c,** Regression line scatter plot showing fraction of positive cells for indicated markers from the Orion 16-plex images vs. progression-free survival (PFS, days) for 40 patients with colorectal cancer. Each dot represents measurements from a single patient. R^2 for each plot is displayed. **d,** Representative Orion IF images of cases with high IFM2 (IS = 4 in specimen C34) and low IFM2 (IS = 0 in specimen C09). IF images show DNA, pan-cytokeratin, α-SMA, CD45, and PD-L1; Scalebars 200 and 300 µm as indicated. Higher magnification regions of interest are shown in **Fig. 4g**.

Extended Data Fig. 5 | Cellular neighborhoods in colorectal cancer resections.

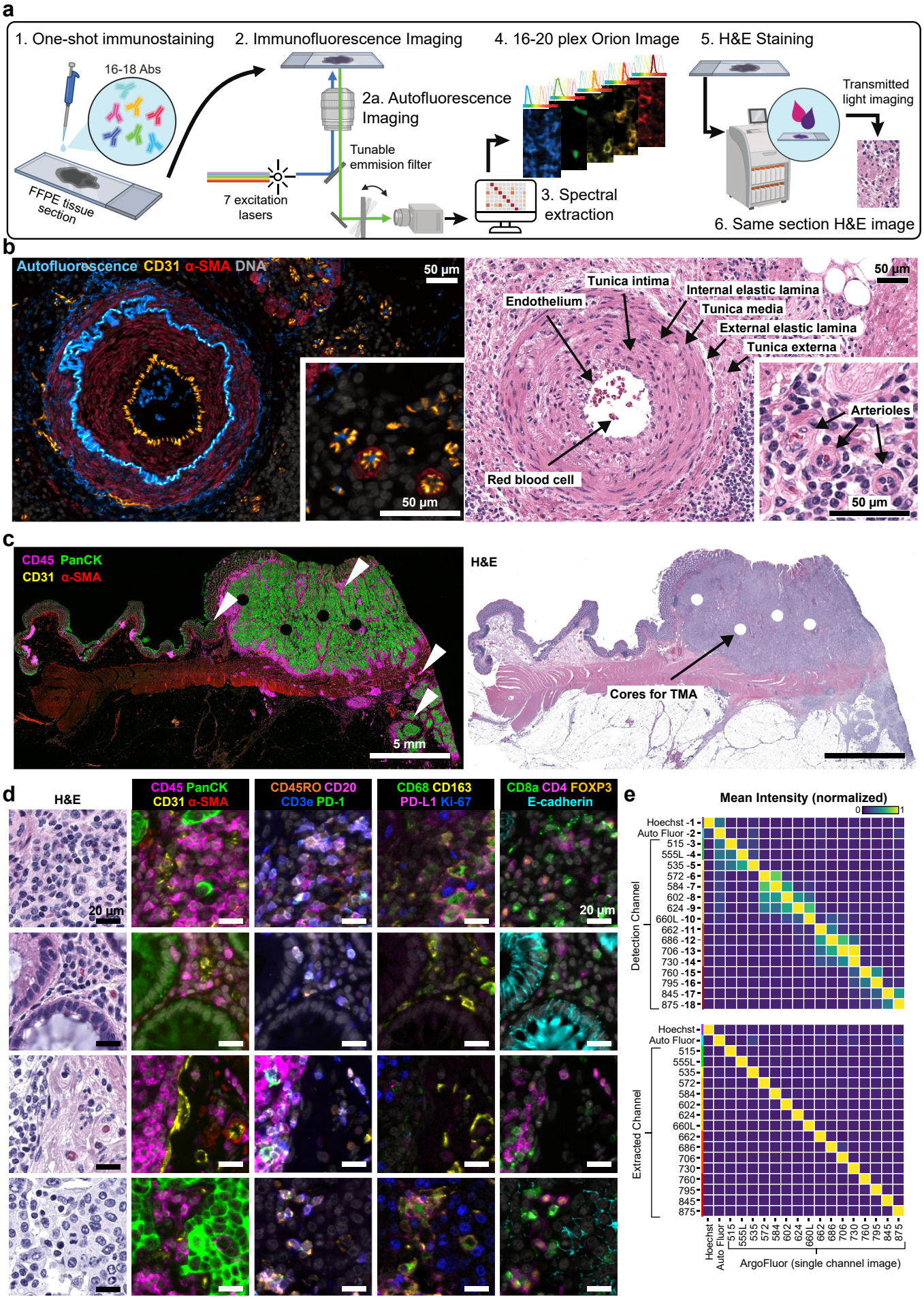
a, Latent Dirichlet Allocation (LDA) probabilistic modeling was used to analyze Orion immunofluorescence data from 40 colorectal cancer specimens to reduce cell populations into neighborhoods ("topics") defined by patterns of single-cell marker expression. The analysis identified 12 topics that recurred across the dataset. Within each box is the LDA plot for the indicated topic (top) and a regression line scatter plot indicating the fraction of each tumor composed of the indicated LDA topic and the relationship to progression-free survival (PFS, days). Each dot represents measurements from a single patient. R^2 for each plot is displayed. **b,** Bar plot depicting the proportional distribution of the LDA Topics in the 40 colorectal cancer specimens.

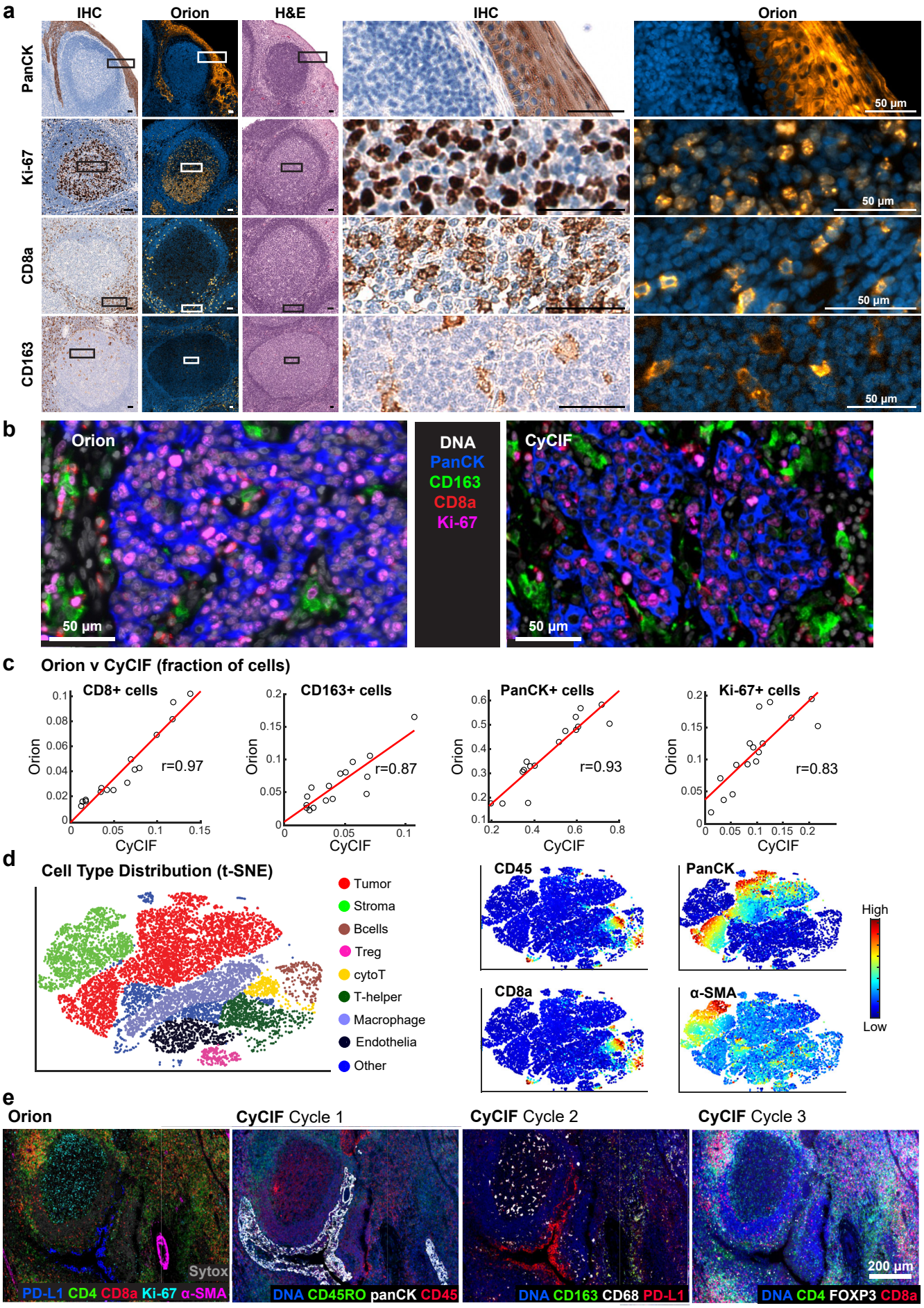
Extended Data Fig. 6 | Evaluation of the performance of a Convolution Neural Network used to identify cellular neighborhood Topic 7 from H&E images of colorectal cancer.

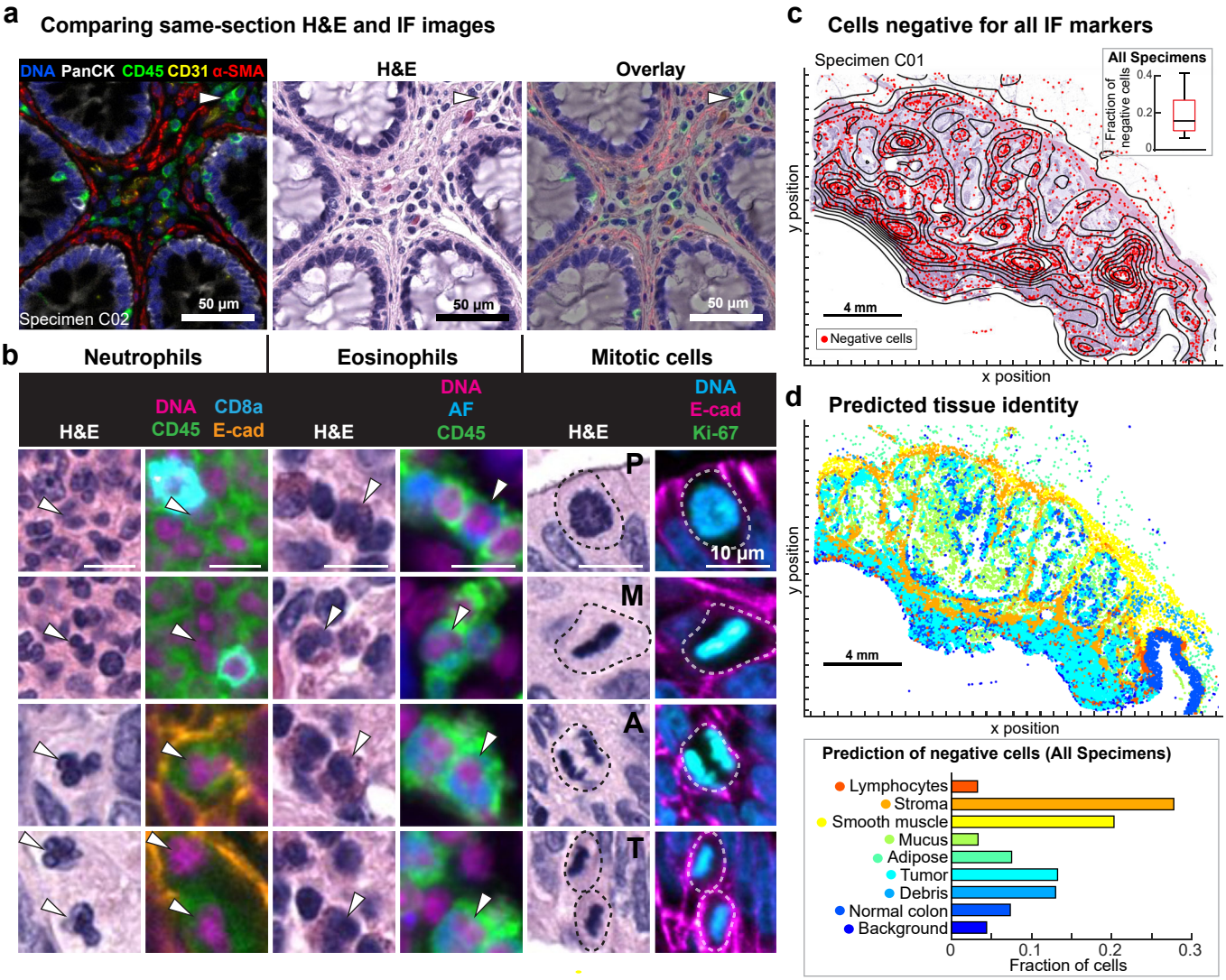
a, Confusion matrix table showing performance of GoogLeNet convolutional neural network (CNN) trained using H&E data from Latent Dirichlet Allocation (LDA) Topic 7 and its performance in identifying Topic 7 cells from H&E data. Topic 0 contains the rest of the topics (3, 5, 6, 9, 10, 11, 12). Target class (ground truth) was assigned from LDA analysis of Orion images and Output class (predicted) was assigned by the GoogLeNet CNN. **b**, Gallery of representative H&E images of true positives for topic 8; Scalebars 50 µm.

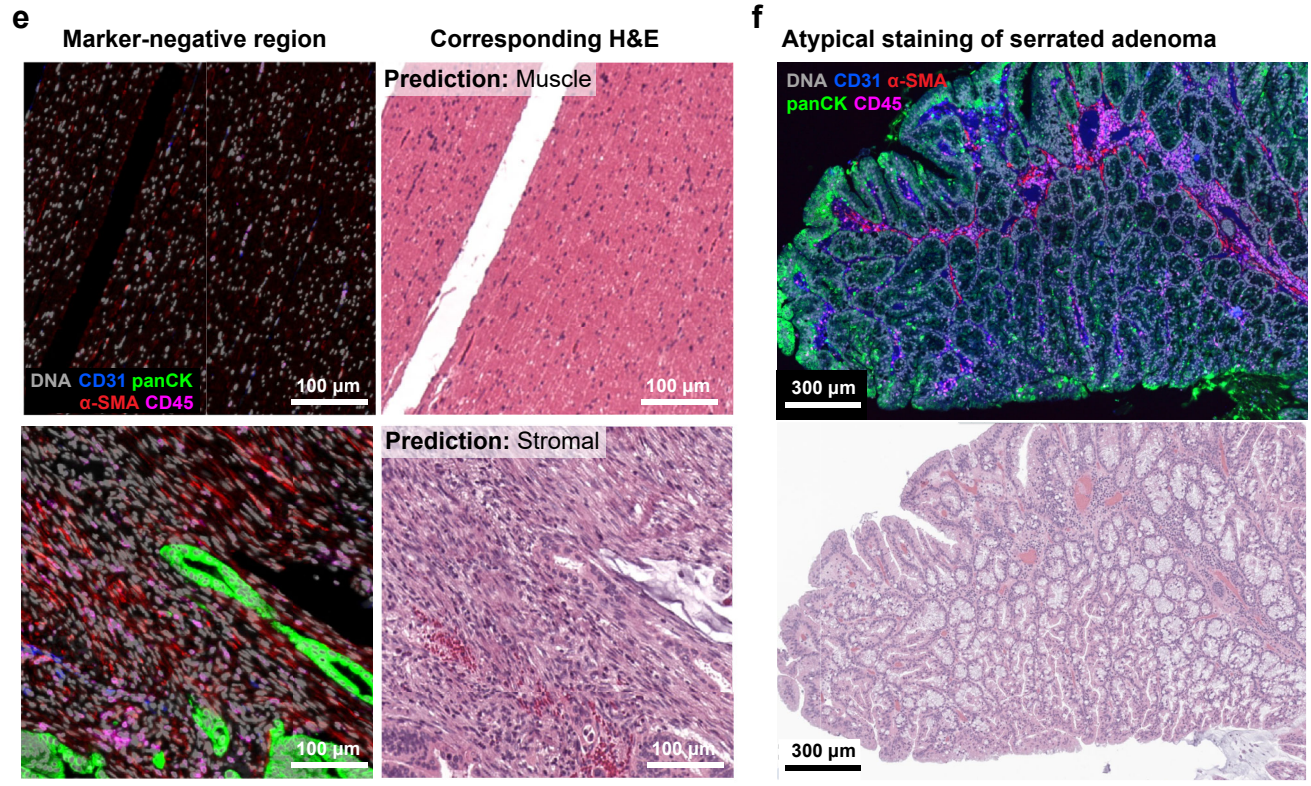
Extended Data Fig. 7 | Additional biomarkers of Topic 7 from CyCIF images.

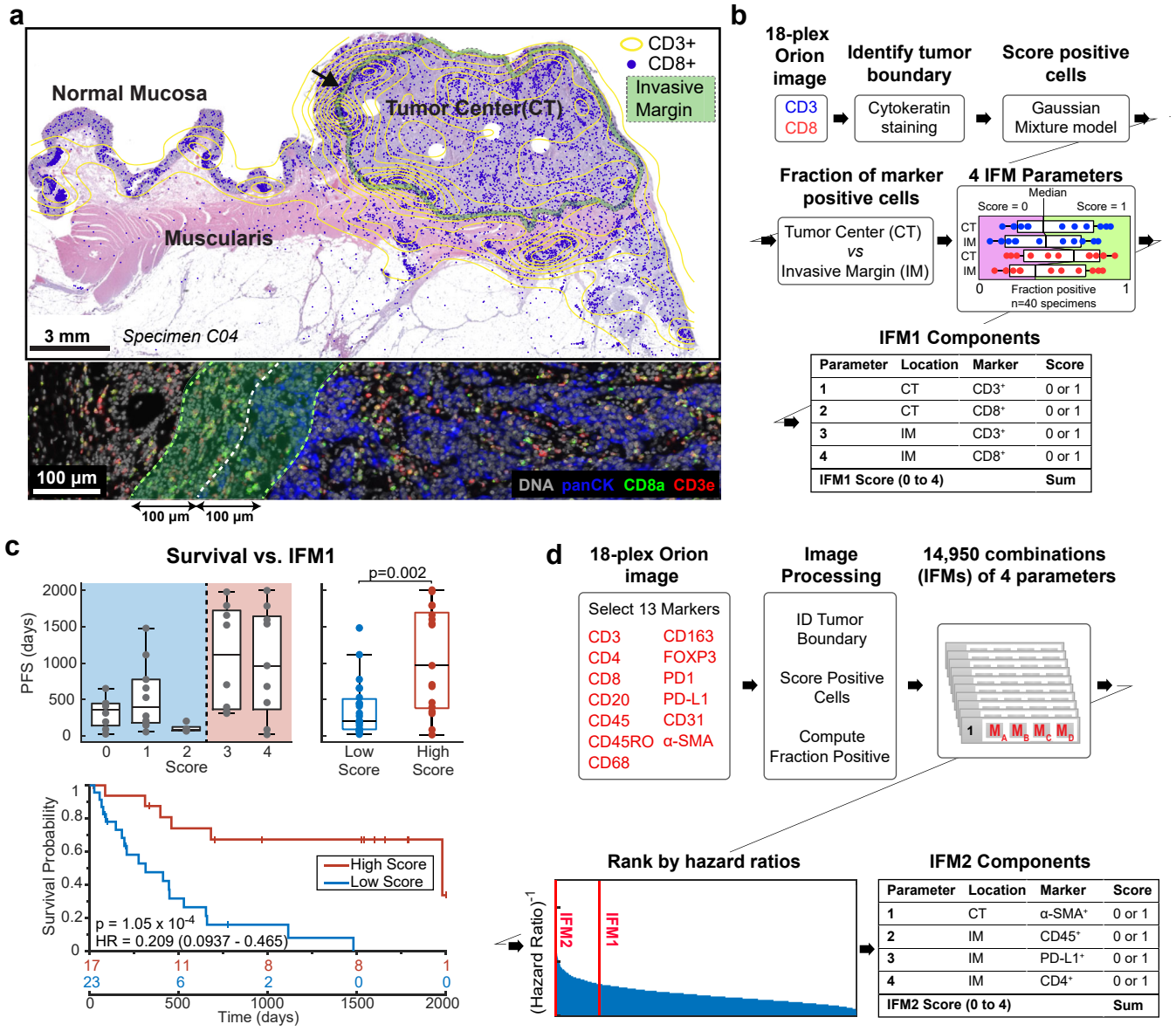
The sample C06 was selected based on the high fraction of Topic 7 cells from the Orion data. The CyCIF images obtained from the same specimen, but a different section are displayed.

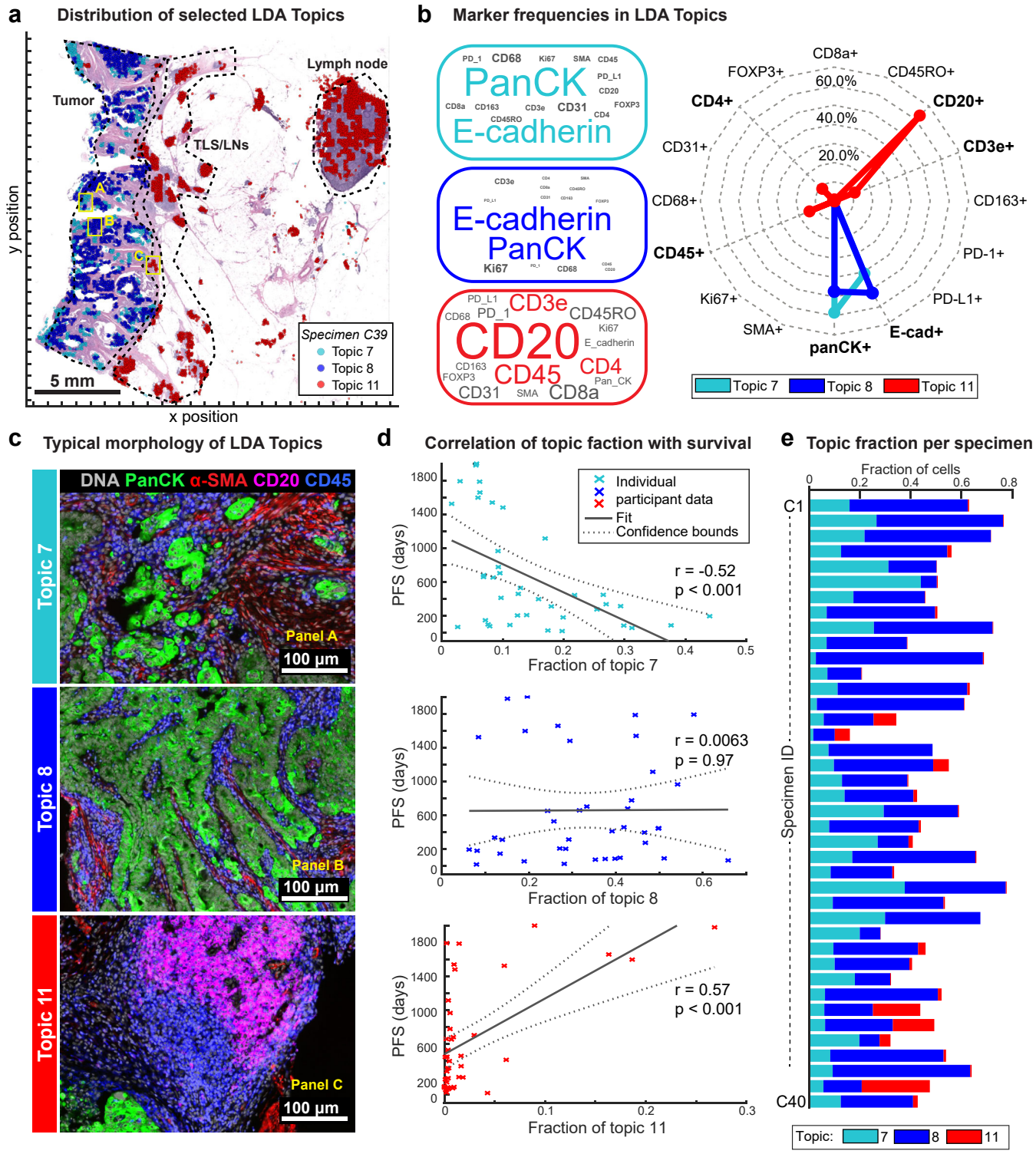

a
1. One-shot immunostaining
16-18 Abs
FFPE tissue section
2. Immunofluorescence Imaging
2a. Autofluorescence Imaging
Tunable emmision filter
7 excitation lasers
3. Spectral extraction
4. 16-20 plex Orion Image
5. H&E Staining
Transmitted light imaging
6. Same section H&E image
b
Autofluorescence CD31 α-SMA DNA
50 µm
Tunica intima
Internal elastic lamina
Endothelium
Tunica media
External elastic lamina
Tunica externa
Arterioles
Red blood cell
c
CD45 PanCK
CD31 α-SMA
5 mm
H&E
Cores for TMA
d
H&E
CD45 PanCK CD31 α-SMA
CD45RO CD20 CD3e PD-1
CD68 CD163 PD-L1 Ki-67
CD8a CD4 FOXP3 E-cadherin
20 µm
e
Mean Intensity (normalized)
Detection Channel
Extracted Channel
Hoechst -1
Auto Fluor -2
515 -3
555L -4
535 -5
572 -6
584 -7
602 -8
624 -9
660L -10
662 -11
686 -12
706 -13
730 -14
760 -15
795 -16
845 -17
875 -18
ArgoFluor (single channel image)

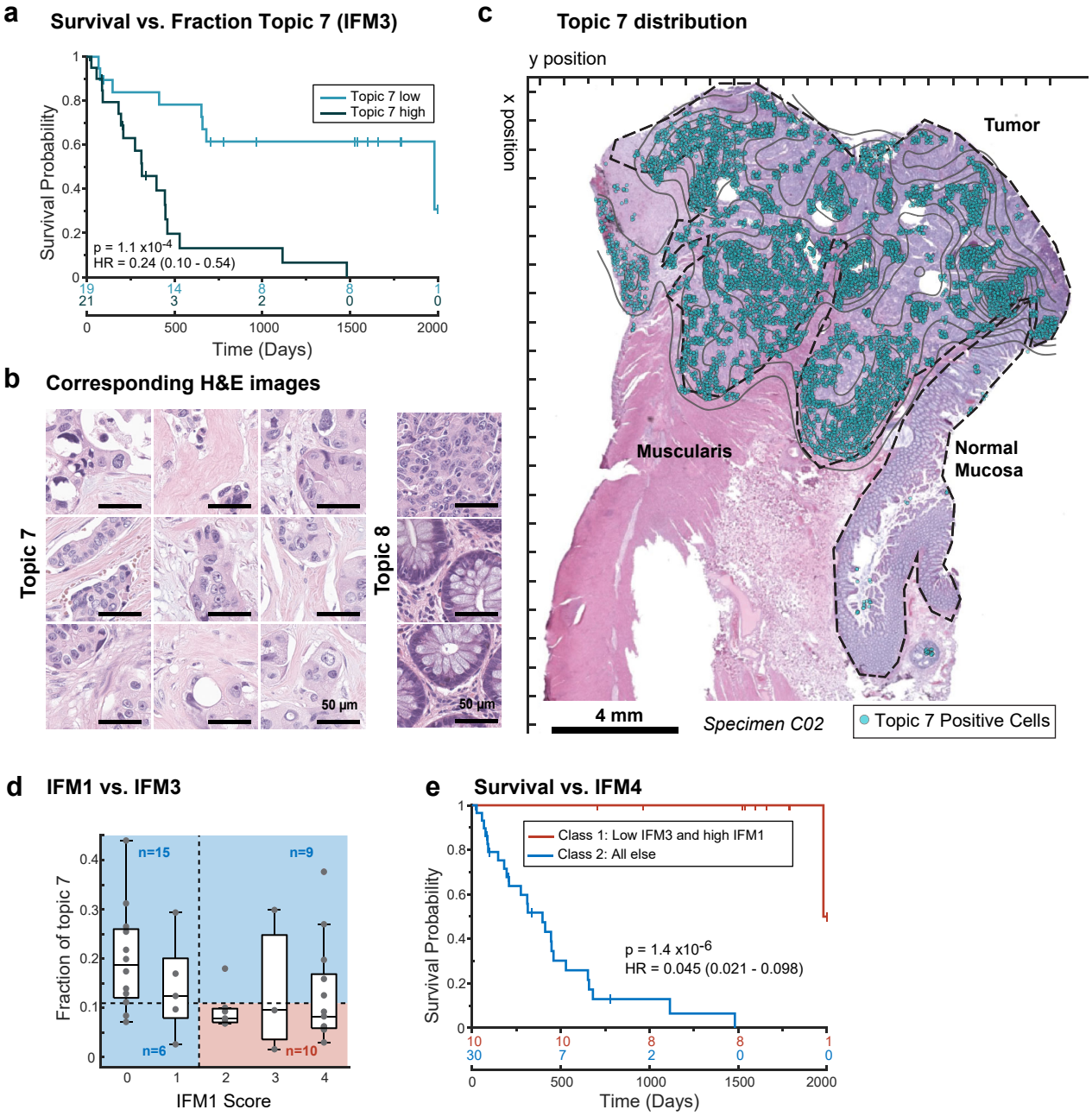

a
IHC
Orion
H&E
IHC
Orion
PanCK
Ki-67
CD8a
CD163
50 µm
b
Orion
50 µm
DNA
PanCK
CD163
CD8a
Ki-67
CyCIF
50 µm
c
Orion v CyCIF (fraction of cells)
CD8+ cells
r=0.97
CD163+ cells
r=0.87
PanCK+ cells
r=0.93
Ki-67+ cells
r=0.83
Orion
CyCIF
d
Cell Type Distribution (t-SNE)
Tumor
Stroma
Bcells
Treg
cytoT
T-helper
Macrophage
Endothelia
Other
CD45
PanCK
CD8a
α-SMA
High
Low
e
Orion
CyCIF Cycle 1
CyCIF Cycle 2
CyCIF Cycle 3
Sytox
PD-L1 CD4 CD8a Ki-67 α-SMA
DNA CD45RO panCK CD45
DNA CD163 CD68 PD-L1
200 µm
DNA CD4 FOXP3 CD8a


a
Comparing same-section H&E and IF images
DNA PanCK CD45 CD31 α-SMA
H&E
Overlay
Specimen C02
50 µm
b
Neutrophils
Eosinophils
Mitotic cells
H&E
DNA CD45
CD8a E-cad
DNA AF CD45
DNA E-cad Ki-67
P
M
A
T
10 µm
c
Cells negative for all IF markers
Specimen C01
All Specimens
Fraction of negative cells
y position
x position
4 mm
Negative cells
d
Predicted tissue identity
Prediction of negative cells (All Specimens)
Lymphocytes
Stroma
Smooth muscle
Mucus
Adipose
Tumor
Debris
Normal colon
Background
Fraction of cells

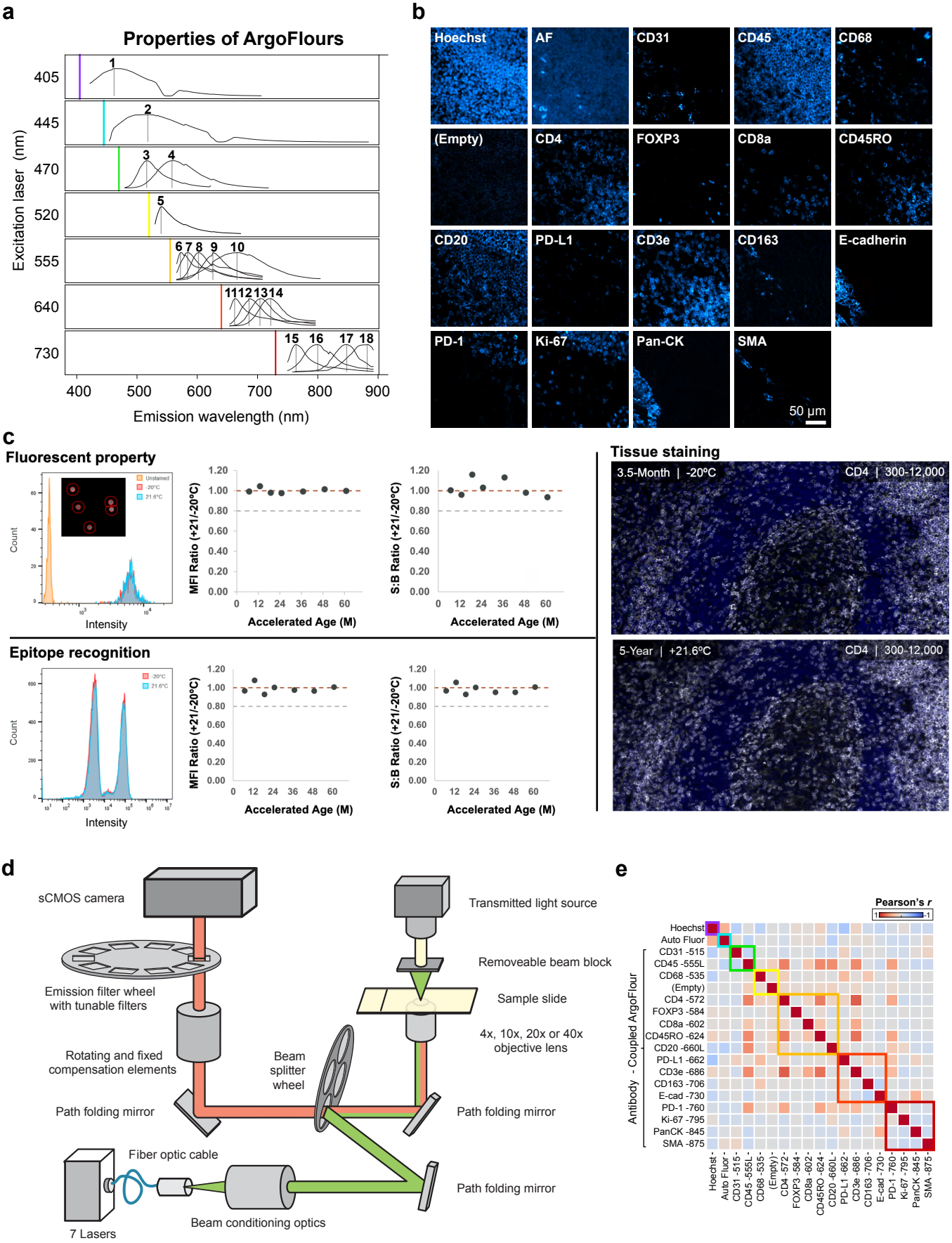

e
Marker-negative region
Corresponding H&E
DNA CD31 panCK
α-SMA CD45
100 µm
Prediction: Muscle
Prediction: Stromal
f
Atypical staining of serrated adenoma
DNA CD31 α-SMA
panCK CD45
300 µm


a
CD3+
CD8+
Invasive Margin
Normal Mucosa
Tumor Center(CT)
Muscularis
3 mm
Specimen C04
100 µm
DNA panCK CD8a CD3e
100 µm
100 µm
b
18-plex Orion image
CD3
CD8
Identify tumor boundary
Cytokeratin staining
Score positive cells
Gaussian Mixture model
Fraction of marker positive cells
Tumor Center (CT)
vs
Invasive Margin (IM)
4 IFM Parameters
Median
Score = 0
Score = 1
CT
IM
CT
IM
0
Fraction positive
1
n=40 specimens
IFM1 Components
Parameter Location Marker Score
1 CT CD3+ 0 or 1
2 CT CD8+ 0 or 1
3 IM CD3+ 0 or 1
4 IM CD8+ 0 or 1
IFM1 Score (0 to 4) Sum
c
Survival vs. IFM1
PFS (days)
Score
p=0.002
Low Score
High Score
Survival Probability
High Score
Low Score
p = 1.05 x 10-4
HR = 0.209 (0.0937 - 0.465)
Time (days)
d
18-plex Orion image
Select 13 Markers
CD3 CD163
CD4 FOXP3
CD8 PD1
CD20 PD-L1
CD45 CD31
CD45RO α-SMA
CD68
Image Processing
ID Tumor Boundary
Score Positive Cells
Compute Fraction Positive
14,950 combinations (IFMs) of 4 parameters
Rank by hazard ratios
(Hazard Ratio)-1
IFM2
IFM1
IFM2 Components
Parameter Location Marker Score
1 CT α-SMA+ 0 or 1
2 IM CD45+ 0 or 1
3 IM PD-L1+ 0 or 1
4 IM CD4+ 0 or 1
IFM2 Score (0 to 4) Sum
e
Survival vs. IFM2
Low Score
High Score
PFS (days)
Score
Survival Probability
High Score
Low Score
p = 1.91 x 10-6
HR=0.0785 (0.0358-0.172)
Time (Days)
f
Cross- validation
Rank, lower the better
Rank by Hazard Ratio
IFM1
IFM2
Bootstrap Sampling
HR, lower the better
Hazard ratio
IFM1
IFM2
g
IFM2 = 4 (C34)
DNA PD-L1 PanCK α-SMA CD4
100 µm
IFM2 = 0 (C09)
DNA PD-L1 PanCK α-SMA CD4

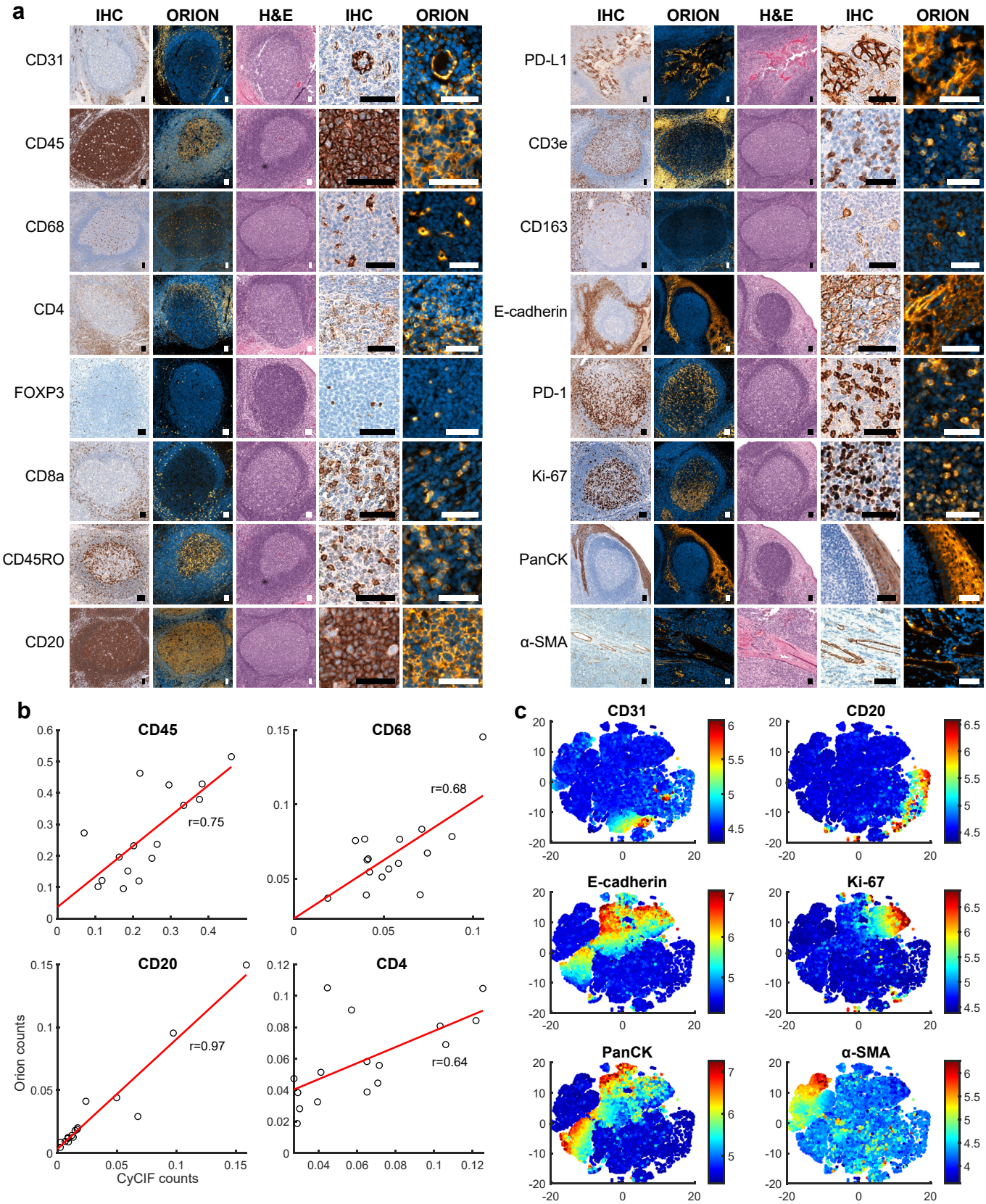

a
Distribution of selected LDA Topics
Tumor
Lymph node
TLS/LNs
A
B
C
Specimen C39
Topic 7
Topic 8
Topic 11
5 mm
y position
x position
b
Marker frequencies in LDA Topics
PanCK
E-cadherin
E-cadherin
PanCK
CD20
CD3e
CD45RO
CD45
CD4
CD8a
CD31
CD8a+
CD45RO+
CD20+
CD3e+
CD163+
PD-1+
PD-L1+
E-cad+
panCK+
SMA+
Ki67+
CD45+
CD68+
CD31+
CD4+
FOXP3+
60.0%
40.0%
20.0%
Topic 7
Topic 8
Topic 11
c
Typical morphology of LDA Topics
DNA PanCK α-SMA CD20 CD45
Topic 7
Topic 8
Topic 11
Panel A
Panel B
Panel C
100 µm
d
Correlation of topic faction with survival
Individual
participant data
Fit
Confidence bounds
PFS (days)
r = -0.52
p < 0.001
Fraction of topic 7
r = 0.0063
p = 0.97
Fraction of topic 8
r = 0.57
p < 0.001
Fraction of topic 11
e
Topic fraction per specimen
Fraction of cells
C1
C40
Specimen ID
Topic: 7 8 11
f
PanCK+
E-cadherin+
CD20+
CD45+
CD68+
Fraction positive
Topic
p<0.01
NS
p=0.01
p=0.08


a
Survival vs. Fraction Topic 7 (IFM3)
Survival Probability
Time (Days)
Topic 7 low
Topic 7 high
p = 1.1 x10-4
HR = 0.24 (0.10 - 0.54)
b
Corresponding H&E images
Topic 7
Topic 8
50 µm
c
Topic 7 distribution
y position
x position
Tumor
Muscularis
Normal Mucosa
4 mm
Specimen C02
Topic 7 Positive Cells
d
IFM1 vs. IFM3
Fraction of topic 7
IFM1 Score
n=15
n=6
n=9
n=10
e
Survival vs. IFM4
Survival Probability
Time (Days)
Class 1: Low IFM3 and high IFM1
Class 2: All else
p = 1.4 x10-6
HR = 0.045 (0.021 - 0.098)

Extended Data Figure 1

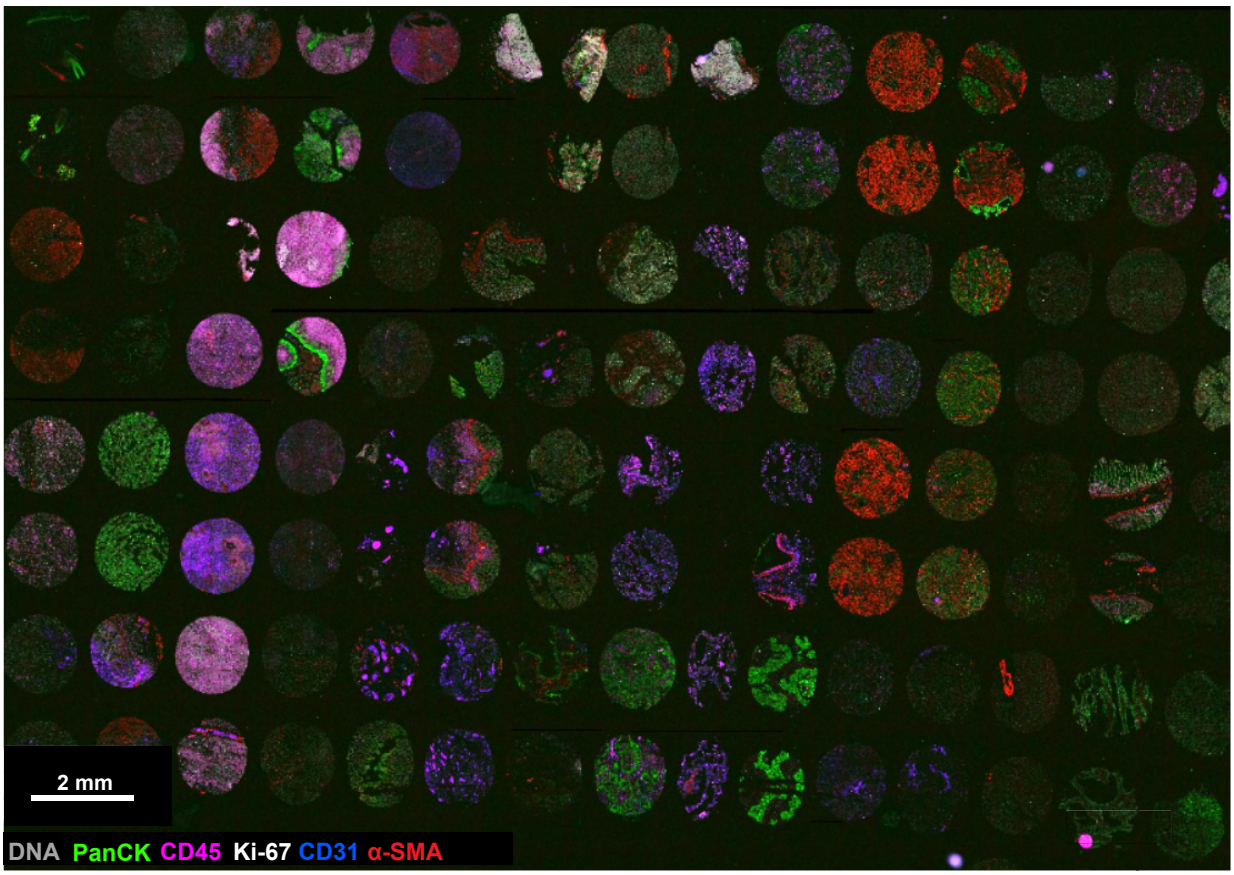


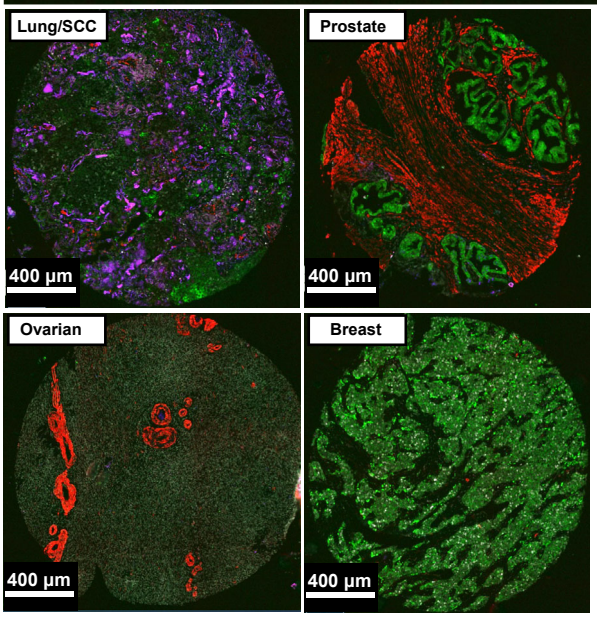
Extended Data Figure 2



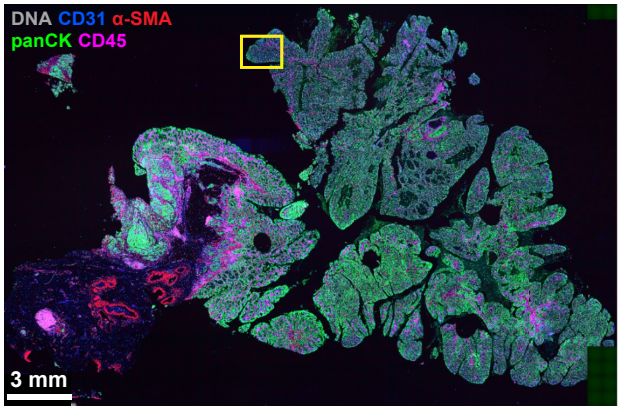
Extended Data Figure 3

a

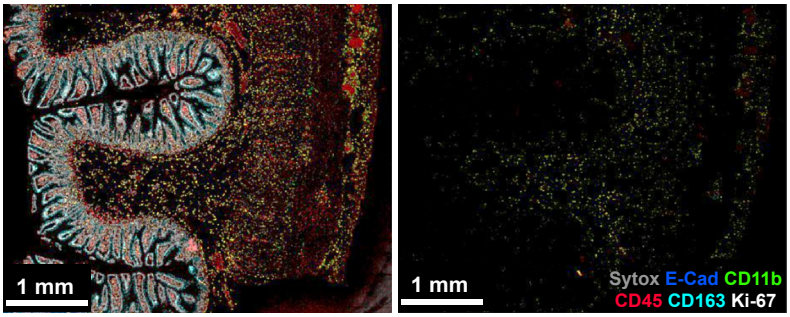




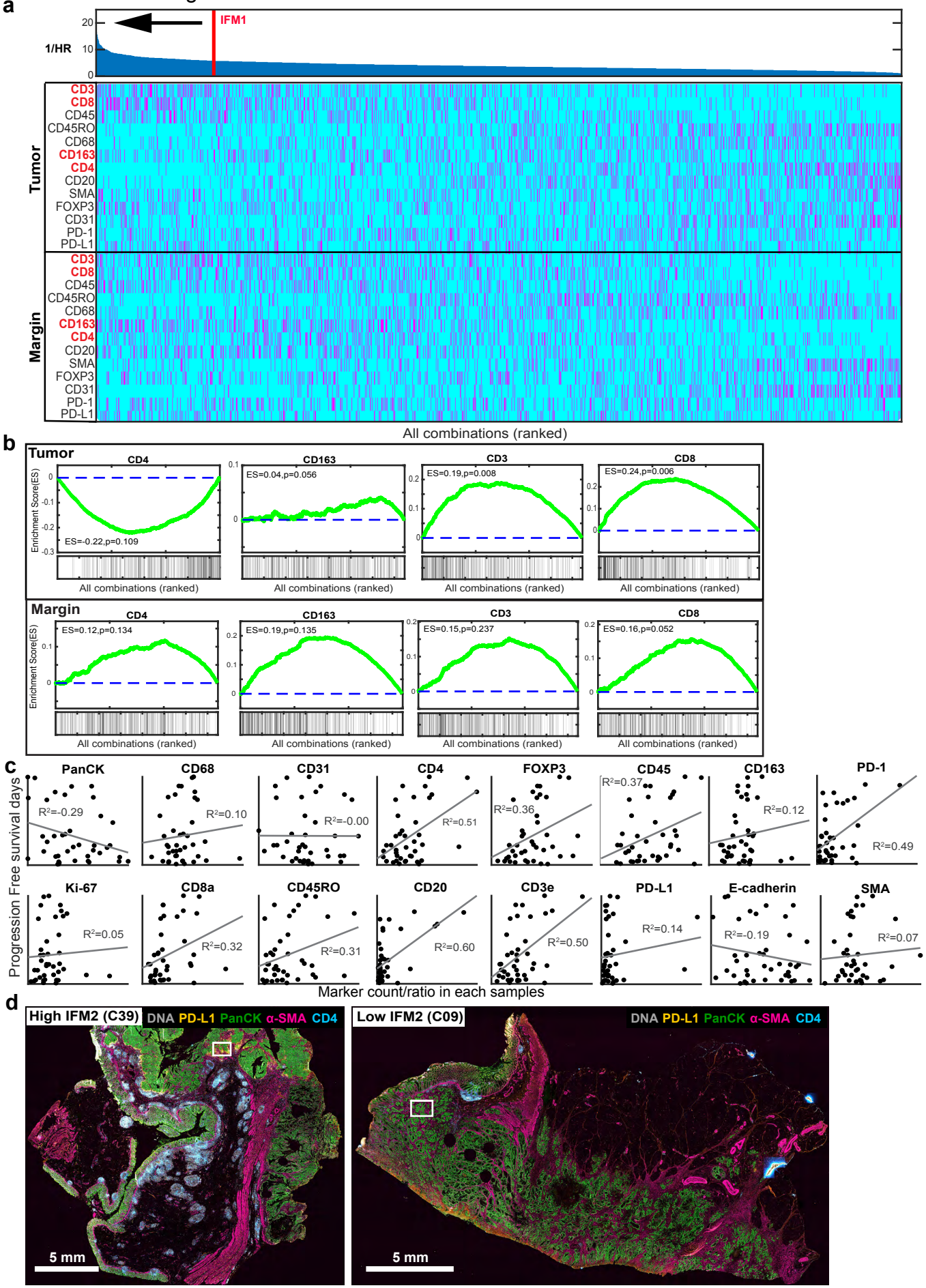
c



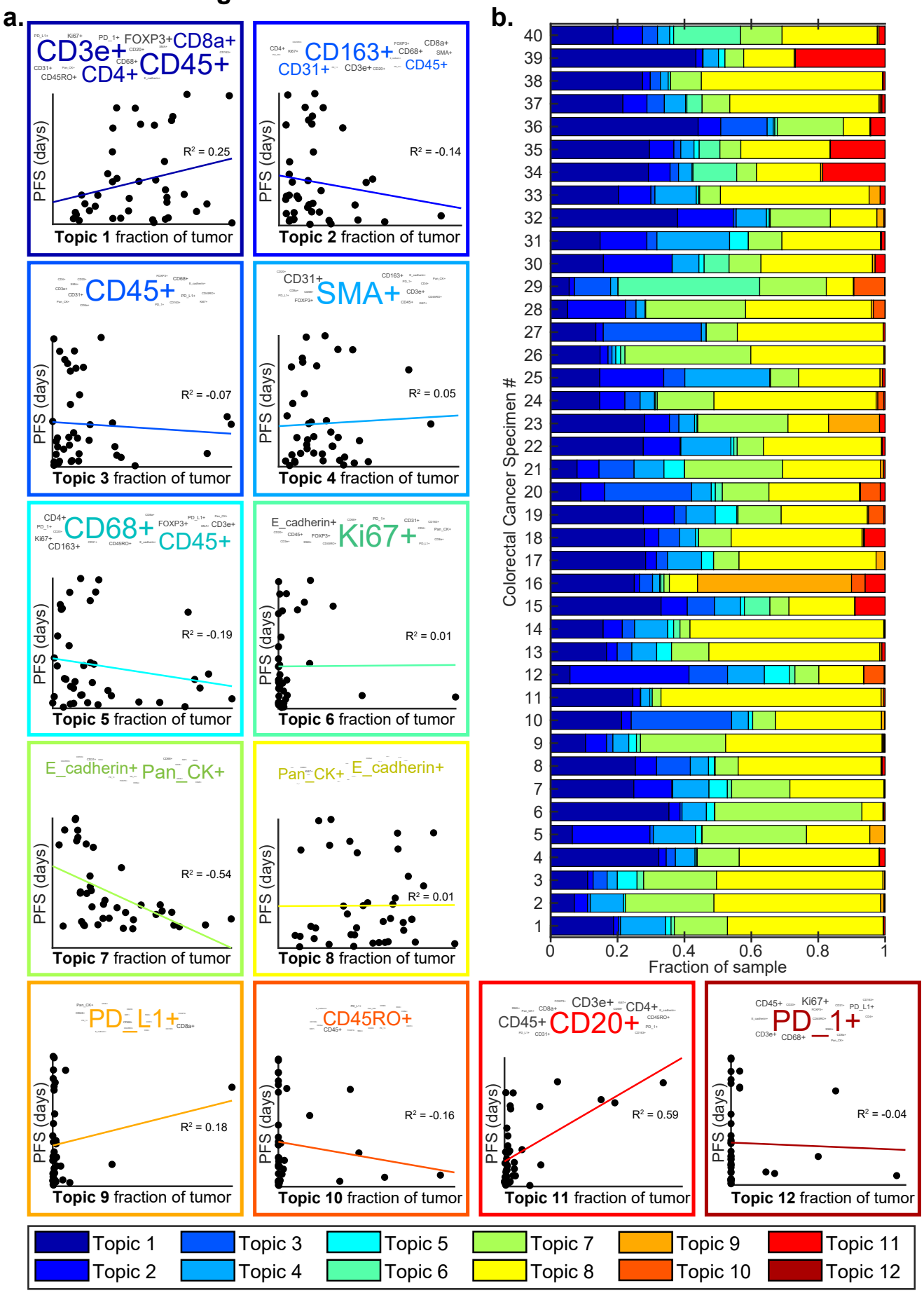
b



Extended Data Figure 4



Extended Data Figure 5



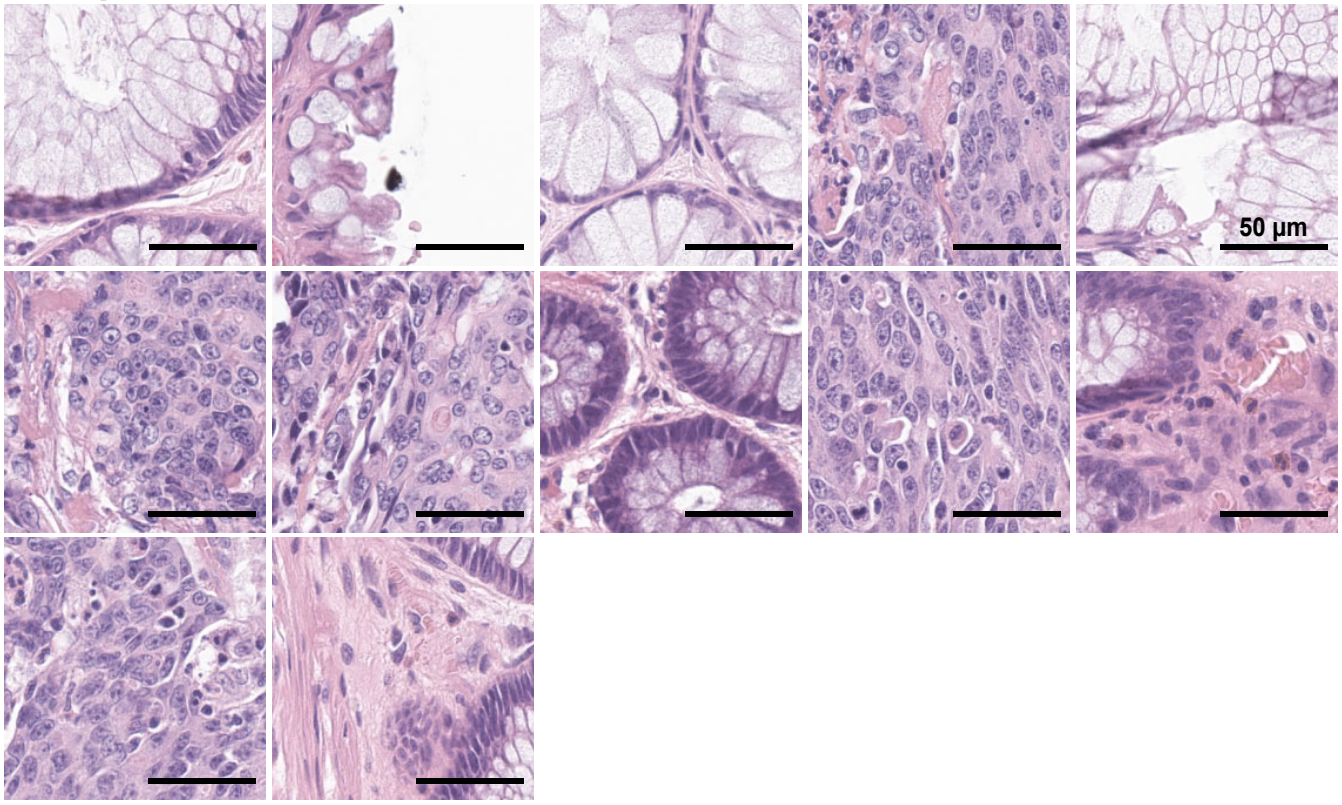
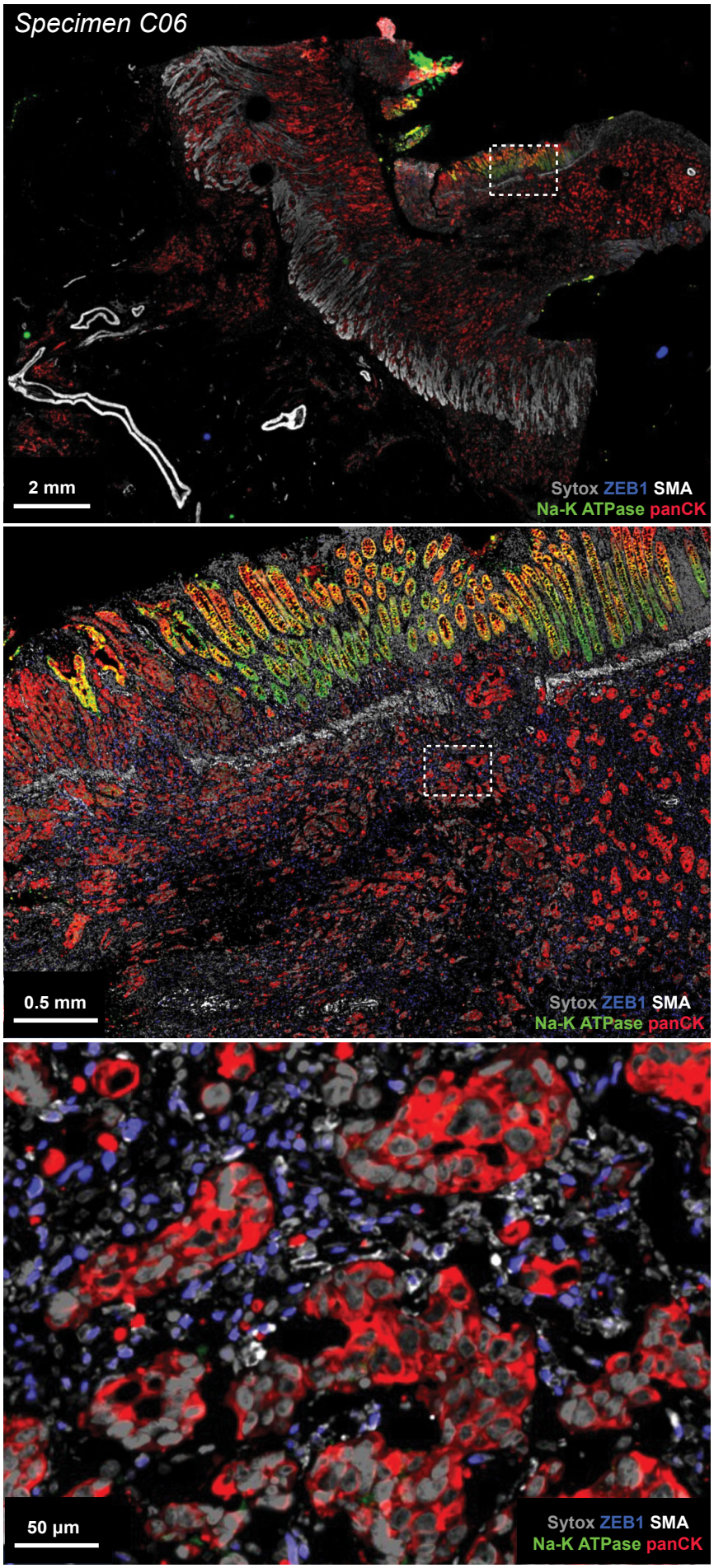
Extended Data Figure 6

a

Prediction \ Ground truth	Topic 0	Topic 1	Topic 2	Topic 4	Topic 7	Topic 8	Precision
Topic 0	**0** 0.0%	**33** 0.3%	**8** 0.1%	**6** 0.1%	**1** 0.0%	**20** 0.2%	0.0% 100%
Topic 1	**6** 0.1%	**569** 4.8%	**22** 0.2%	**33** 0.3%	**18** 0.2%	**252** 2.1%	63.2% 36.8%
Topic 2	**635** 5.4%	**815** 6.9%	**1051** 8.9%	**1163** 9.9%	**378** 3.2%	**294** 2.5%	24.2% 75.8%
Topic 4	**15** 0.1%	**46** 0.4%	**70** 0.6%	**313** 2.7%	**24** 0.2%	**119** 1.0%	53.3% 46.7%
Topic 7	**214** 1.8%	**271** 2.3%	**69** 0.6%	**226** 1.9%	**2082** 17.7%	**960** 8.1%	54.5% 45.5%
Topic 8	**0** 0.0%	**180** 1.5%	**1** 0.0%	**0** 0.0%	**3** 0.0%	**1887** 16.0%	91.1% 8.9%
Recall	0.0% 100%	29.7% 70.3%	86.1% 13.9%	18.0% 82.0%	83.1% 16.9%	53.4% 46.6%	**50.1%** **49.9%**

b

Topic 8



Extended Data Figure 7