

SCIENTIFIC REPORTS



OPEN

Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer

Korsuk Sirinukunwattana¹, David Snead², David Epstein³, Zia Aftab⁴, Imaad Mujeeb⁴, Yee Wah Tsang², Ian Cree⁵ & Nasir Rajpoot^{6,2,7}

Distant metastasis is the major cause of death in colorectal cancer (CRC). Patients at high risk of developing distant metastasis could benefit from appropriate adjuvant and follow-up treatments if stratified accurately at an early stage of the disease. Studies have increasingly recognized the role of diverse cellular components within the tumor microenvironment in the development and progression of CRC tumors. In this paper, we show that automated analysis of digitized images from locally advanced colorectal cancer tissue slides can provide estimate of risk of distant metastasis on the basis of novel tissue phenotypic signatures of the tumor microenvironment. Specifically, we determine what cell types are found in the vicinity of other cell types, and in what numbers, rather than concentrating exclusively on the cancerous cells. We then extract novel tissue phenotypic signatures using statistical measurements about tissue composition. Such signatures can underpin clinical decisions about the advisability of various types of adjuvant therapy.

Cell function and behavior cannot be fully understood without the context of their microenvironment. Communication between cells and their surroundings allows the functional organization of cells into tissues and organs. It also plays a vital role in maintaining tissue homeostasis by generating signals that suppress and revert malignant phenotypes¹. Experiments in animal and cell culture models have demonstrated that certain conditions of the microenvironment can cause potent cancerous cells to revert to an almost normal phenotype^{2,3}. Although the normal tissue microenvironment is known to be resilient to tumorigenesis, false signals in the microenvironment can disrupt tissue homeostasis and subsequently initiate tumors. The microenvironment in which tumor exists is both complex and heterogeneous, inhabited by a multitude of cellular and non-cellular components including tumor cells, extracellular matrix, tumor stroma, blood vessels, inflammatory cells, signaling molecules⁴⁻⁶. Studies over the last decade have increasingly recognized the role of these different components in the development and progression of tumors⁵. This paper adds to this evidence and shows how its quantification may be automated.

Metastasis is the major cause of morbidity and death in colorectal cancer (CRC). The 5-year survival rate in CRC patients with distant metastasis is approximately 10%, considerably smaller than 70% with regional metastasis and 90% without metastasis⁷. Patients at high risk of developing distant metastasis could benefit from appropriate adjuvant and follow-up treatments if stratified accurately. The literature reports several histopathological features carrying prognostic value for CRC progression. Each of the features reflects competing cellular stimuli that influence tumor progression or suppression within the microenvironment. Type, density, and relative locations of different tissue components in the tumor microenvironment are crucial in determining progression and patient survival in CRC. For instance, the number of cytotoxic and memory T cells in the tumor center and the invasive margin have been linked to an improved prognosis of CRC⁸. Similarly, numerous studies have reported cancer-associated fibroblasts (CAFs) and desmoplasia to be important histopathological features associated with an unfavorable prognosis for CRC and an increased mortality rate⁹⁻¹³. Analogous to a wound that never heals^{14,15}, tumors stimulate many associated responses, wherein normal fibroblasts have been reported to acquire a cancer-associated phenotype^{5,16}.

¹Department of Engineering Science, University of Oxford, Oxford, UK. ²Department of Pathology, University Hospitals Coventry and Warwickshire, Coventry, UK. ³Mathematics Institute, University of Warwick, Coventry, UK. ⁴Hamad Medical Corporation, Doha, Qatar. ⁵International Agency for Research on Cancer, Lyon, France. ⁶Department of Computer Science, University of Warwick, Coventry, UK. ⁷The Alan Turing Institute, London, UK. Correspondence and requests for materials should be addressed to K.S. (email: korsuk.sirinukunwattana@eng.ox.ac.uk) or N.R. (email: n.m.rajpoot@warwick.ac.uk)

Clinical feature	UHCW cohort	HGH cohort	Total
Number of cases	72	30	102
Age (year)			
Median	70.5	55	67
Range	32–90	34–79	32–90
Gender			
Female	34	9	43
Male	38	21	59
Tumor histological type			
Adenocarcinoma	63	26	89
Mucinous	9	3	12
Not available	0	1	1
Tumor differentiation			
Well differentiated	6	6	12
Moderately differentiated	37	20	57
Poorly differentiated	16	4	20
Not available	13	0	13
T stage			
pT3	56	26	82
pT4	16	4	20
5-year metastasis			
No	52	23	75
Yes	20	7	27
Median metastasis-free survival (year)			
With distant metastasis	1.148	Not available	1.148
Without distant metastasis	>5	Not available	>5

Table 1. A summary of clinicopathological data.

Furthermore, the extent of necrosis in CRC has been reported to correlate strongly with cancer progression and patient survival^{13,17,18}. The link between necrosis and tumor progression is possibly due to the hypoxic nature of tumors, which drives tumor infiltrating inflammatory cells, namely phagocytic macrophages and granulocytes, to secrete pro-inflammatory cytokines which in turn promote cell proliferation⁴.

In this study, we investigate the significance of tissue phenotypic and morphometric features, exploring in particular cellular heterogeneity in tumor microenvironments, in determining metastatic potential in CRC patients diagnosed with advanced primary tumors. Based on the AJUCC/UICC-TNM staging system¹⁹, this group of patients have a primary tumor that has grown through the outer lining of colon wall (T3/T4), have no lymph nodes that are affected by cancer cells (N0), and no clinical evidence of distant metastasis at the time of diagnosis (M0). Detailed quantitative analysis was performed on whole slide images (WSIs) of CRC histology slides, stained with routine Hematoxylin & Eosin (H&E) dyes in a fully quantitative manner, using bespoke image analysis methods to provide an objective and reproducible assessment. Quantitative analysis of various types of cell population reveals novel tissue phenotypic features, derived from both cell-cell connection frequencies and tissue appearance, with significant association with metastasis incidence and distant metastasis-free survival (DMFS) in the advanced primary CRC tumors.

Results

Quantifying tissue phenotypic signatures of CRC tumors. In this study, WSIs of Hematoxylin and Eosin (H&E)-stained histological sections from 102 patients with advanced node negative primary CRC tumors (T3/T4, N0, M0) were acquired from two independent cohorts from two different institutes: University Hospitals Coventry and Warwickshire (UHCW, 72 patients) and Hamad General Hospital (HGH, 30 patients). Summary details of the cohorts and clinical information are given in Table 1.

CRC, like other solid tumors, is a disease of substantial heterogeneity^{20,21}. Different parts of the same tumor can exhibit different features including cellular morphology, gene expression, metabolism, motility, angiogenic, proliferative, immunogenic and metastatic potential²². The tumor microenvironment is composed of diverse cell types; each plays a different role in tumor development and progression — some support and promote tumor progression while others play host protective roles⁵. The biological functions of cells are not only determined by their type but are also greatly influenced by their surrounding context. It follows that tissue morphometric signatures measuring tumor heterogeneity could be computed from the analysis of distributions and relative locations of cellular populations in the tumor microenvironment.

Here, we outline the quantification of digital tissue phenotypic signatures (see Methods for details). We divided each tumor histology image (i.e., each WSI) into small square regions or sub-images (Fig. 1a) and analyzed the small sub-images to obtain local characteristics that were then summarized to characterize the entire tumor section. We first applied our artificial intelligence (AI) based algorithm²³, which was recently shown to be the state-of-the-art in

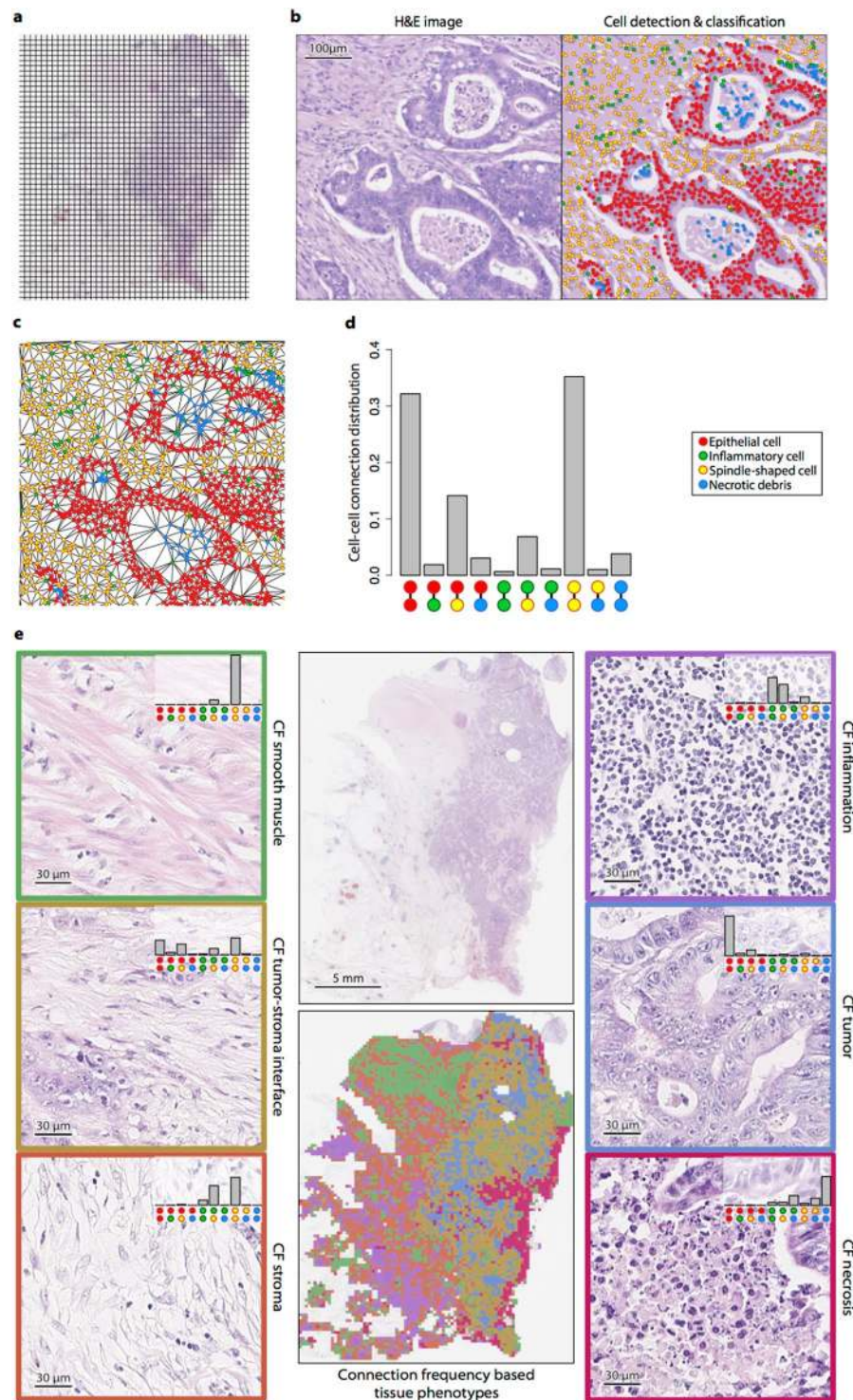


Figure 1. Profiling tissue morphometric phenotypes. A WSI was divided into small regions of size $200 \times 200 \mu\text{m}^2$ (a). Cellular components in the image were localized and classified into 4 different cell types, including malignant epithelial cell, inflammatory cell, spindle-shaped cell and necrotic debris, based on their nuclear morphology and surrounding tissue context (b). A cell network was subsequently constructed from the cell detection and classification results, in which nodes in the network represent cells and edges conceptualize relationships among them (c). A distribution of cell-cell connections was calculated for each small region (d). According to their distributions of cell-cell connections, tissue regions were profiled into 6 different phenotypes (e).

detecting and distinguishing between four types of cells based on their morphology and context, to each sub-image. The four types of cells were: malignant epithelial cells, spindle-shaped cells (normal fibroblasts, cancer-associated fibroblasts and smooth muscle cells), inflammatory cells (eosinophils, lymphocytes and neutrophils), and necrotic debris

Feature	Differentiation		Histological type		T stage	
	p-value	r^2	p-value	r^2	p-value	r^2
CF smooth muscle ratio	0.761	0.001	0.458	0.005	0.830	0.000
CF inflammation ratio	0.293	0.011	0.946	0.000	0.205	0.016
CF tumor-stroma interface ratio	0.502	0.004	0.537	0.004	0.005	0.079
CF tumor ratio	0.033	0.045	0.571	0.003	0.378	0.008
CF stroma ratio	0.755	0.001	0.897	0.000	0.628	0.002
CF necrosis ratio	0.194	0.017	0.306	0.010	0.746	0.001

Table 2. Association between the CF tissue phenotypic features and standard clinical features. Mann-Whitney test's p-value and coefficient of determination (r^2) are used to assess the association between features. The results with p-value less than 0.05 is considered statistically significant (bold).

(Fig. 1b). This allowed us to do quantification of tissue morphological characteristics associated with tumor, based on both distributions and relative spatial locations of diverse cell types. For each small tissue region (sub-image) in the large WSI, we then constructed a cell network (Fig. 1c). Each vertex of the network represents a cell of a certain type, and an edge denotes a cell-cell connection between immediately neighboring cells. Based on the distribution of cell-cell connections in the network (Fig. 1d), we then grouped the local tissue regions into different phenotypes using an unsupervised learning approach. The six resulting connection frequency (CF) based tissue phenotypes were visually discernible with each phenotype corresponding mainly to local areas of smooth muscle, inflammation, tumor-stroma interface, tumor, stroma, or necrosis (Fig. 1e). Finally, we used the ratio of the area of each CF tissue phenotype to the total tissue area to give digital tissue phenotypic signature of each tumor sample (Methods).

To further examine the extent to which the aforementioned automatically derived cell-cell CF tissue phenotypes correlate with known tissue types, we also quantified the tissue types by means of appearance based (AP) tissue segmentation. The tissue content of each WSI was automatically segmented into the following eight categories: tumor, stroma, loose connective tissue, normal/hyperplastic mucosa, smooth muscle, necrosis, fat, and inflammation (Fig. S1). We then investigated correlation between the CF and AP based tissue phenotypes. These are smooth muscle, inflammation, tumor, stroma, and necrosis. The Spearman correlation coefficients for individual pairs of CF and AP features range from 0.427 to 0.698 (Fig. S2), indicating moderate correspondence between the automatically-derived phenotypes and the underlying tissue types.

In addition to the phenotypic and standard clinical features, we considered the following automatically-derived features: Morisita index²⁴, stroma-tumor ratio^{9,11,12}, and necrosis-tumor ratio^{13,17,18}. These features have previously been identified as having prognostic significance for CRC or other malignancies. Morisita index measures the spatial coexistence of inflammatory cell and malignant epithelial cells²⁴. Stroma-tumor ratio is defined as the proportion of the total area of stroma to the total area of combined stroma and tumor in the tissue. Necrosis-tumor ratio is defined in the similar manner as that of stroma-tumor ratio (Methods). It is worth noting that in the above studies, the stroma-tumor ratio and necrosis-tumor ratio were semi-quantitatively assessed on manually selected small regions of histological slides. In contrast, we measured these quantities with greater precision and using all regions of our WSIs, thus avoiding subjective bias.

Association between phenotypic and clinical features. Here, we determined the strength of association between the CF tissue phenotypic features and standard clinical features normally used in routine prognostication of colon cancer (Table 2). The clinical features included tumor differentiation, tumor histological type, and primary tumor (T) stage. For example, to check whether there is association between the CF inflammation ratio and the T stage, we test if the distribution of CF inflammation ratio of the group of samples that are annotated as pT3 stage is significantly different from the distribution of samples that are annotated as pT4 stage using Mann-Whitney U test (also known as Wilcoxon rank-sum test).

We found statistically significant association between CF tumor-stroma interface ratio and T stage ($p = 0.027$). Nonetheless, the relatively small values of the coefficients of determination ($r^2 = 0.048$) indicate that CF tumor-stroma interface ratio are only weakly associated with T stage. There is no statistically significant association between other pairs of the CF phenotypic features and the standard clinical features. Altogether, these results suggest that the CF tissue phenotypic features are not strongly associated with standard clinical features and, therefore, are potentially new features whose prognostic significance is worth further investigation.

Logistic regression analysis. To assess the significance of each phenotypic feature in identifying a patient's risk of subsequent distant metastasis, we carried out logistic regression analysis. Odds ratio factor and 95% confidence interval (CI) estimates were obtained for each feature to quantify the risk of distant metastasis incidence associated with the phenotypic features (Methods). In the multivariate analysis, the effect of individual features was adjusted for the effect of standard clinical features, as well as, the cohort membership since samples from two cohorts (UHCW and HGH) were used in the analysis.

The results show that CF smooth muscle and inflammation ratios are statistically significant ($p < 0.05$) in univariate and multivariate analyses ($p < 0.05$, Table 3 and Table S1). The interquartile change in CF smooth muscle ratio increases the odds of distant metastasis by a factor of 1.889 (95% CI: 0.903–3.95) in univariate analysis and by 2.101 (95% CI: 0.919–4.801) in multivariate analysis. The interquartile change in CF inflammation ratio, on the other hand, decreases the odds by a factor of 0.3 (95% CI: 0.119–0.758) in the univariate analysis and 0.305

Feature	Univariate			Multivariate		
	Odds ratio factor	p-value	AUC	Odds ratio factor	p-value	AUC
Standard histological features						
Differentiation (MD → PD)	1.726 (0.591,5.042)	0.323	0.488			
Histological type (Adenocarcinoma → Mucinous)	0.941 (0.234,3.779)	0.886	0.477			
T stage (pT3 → pT4)	2.211 (0.788,6.2)	0.138	0.565			
Connection frequency (CF) based tissue phenotypic features						
CF smooth muscle ratio (0.161 → 0.368)	1.889 (0.903,3.95)	0.029	0.601	2.101 (0.919,4.801)	0.019	0.591
CF inflammation ratio (0.042 → 0.139)	0.3 (0.119,0.758)	0.027	0.641	0.305 (0.11,0.846)	0.04	0.572
CF tumor-stroma interface ratio (0.116 → 0.231)	1.122 (0.568,2.217)	0.928	0.482	0.974 (0.476,1.994)	0.842	0.539
CF tumor ratio (0.079 → 0.23)	0.455 (0.219,0.948)	0.082	0.626	0.487 (0.227,1.044)	0.117	0.581
CF stroma ratio (0.182 → 0.279)	0.711 (0.412,1.226)	0.469	0.536	0.711 (0.401,1.259)	0.472	0.513
CF necrosis ratio (0.023 → 0.054)	0.75 (0.363,1.552)	0.459	0.52	0.753 (0.356,1.589)	0.433	0.522
Appearance (AP) based tissue phenotypic features						
AP smooth muscle ratio (0.136 → 0.334)	1.625 (0.782,3.379)	0.37	0.542	2.434 (1.014,5.843)	0.104	0.532
AP inflammation ratio (0.025 → 0.072)	0.388 (0.174,0.866)	0.028	0.627	0.404 (0.176,0.926)	0.033	0.581
Other features						
Morisita index ²⁴ (0.356 → 0.533)	1.51 (0.756,3.017)	0.405	0.508	1.352 (0.658,2.777)	0.553	0.52
Stroma-tumor ratio ^{9,11,12} (0.403 → 0.622)	0.985 (0.546,1.777)	0.211	0.586	0.91 (0.481,1.721)	0.202	0.555
Necrosis-tumor ratio ^{13,17,18} (0.079 → 0.227)	0.6 (0.285,1.265)	0.409	0.527	0.633 (0.294,1.36)	0.502	0.511
Cohort (UHCW → HGH)	0.791 (0.294,2.131)	0.64	0.483			

Table 3. Prognostic values of different features according to the logistic regression analysis. Each morphological feature is adjusted by the standard histological features in the multivariate analysis. The statistical significance of each feature is assessed by the likelihood ratio test's p-value. An interquartile change for a continuous variable or categorical change for a categorical variable is noted by ($x \rightarrow y$). A 95% confidence interval of the estimate of odds ratio factor is noted by (x, y). A statistically significant result at the 0.05 is highlighted in bold. AUC in the multivariate analysis refers to the AUC of the multivariate model rather than an individual feature.

(95%CI: 0.11–0.846). Despite the fact that CF smooth muscle and inflammation ratios are separately shown to be statistically significant in both the univariate and multivariate analyses, when considered together in the multivariate model (Table S1), their joint contribution towards the prediction of metastasis development becomes less clear. This is likely due to a moderate degree of correlation ($\rho = -0.64$, Fig. S3) between the features. Thus, when one is used, the other should probably be disregarded.

We investigated if the above statistical results could be achieved by means of the AP smooth muscle ratio and AP inflammation ratio features. Only the AP inflammation ratio is shown to be statistically significant in the univariate and multivariate analyses ($p < 0.05$, Table 3).

Distant metastasis-free survival analysis. Next, we investigated the prognostic significance of various features, using DMFS as a criterion. The analysis was carried out on all cases from the UHCW cohort (72 cases), for which survival data were available. In our multivariate analysis, the effect of individual features was adjusted for the effect of standard clinical features.

The tissue CF smooth muscle ratio feature and the AP inflammation ratio feature were shown to be influencing features in determining the DMFS probability of the patients under Cox proportional hazards models ($p < 0.05$, Table 4 and Table S2). The effect of the interquartile change in CF smooth muscle ratio is to increase the hazard by 1.770 times (95% CI: 0.676–4.635) in the univariate analysis and by 2.106 times (95% CI: 0.793–5.595) in the multivariate analysis. The effect of interquartile change in AP inflammation ratio on the DMFS probability is to reduce the hazard by a factor of 0.376 (95% CI: 0.191–0.741) in the univariate analysis and by a factor of 0.389 (95% CI: 0.189–0.803) in the multivariate analysis. In addition, when CF smooth muscle and AP inflammation are compared together in the same multivariate model, the effect of each feature on the DMFS probability vanishes (Table S2). There is a statistically significant difference between the survival distributions of cases when stratified by AP inflammation ratio (log-rank $p < 0.05$, Table 4, Fig. 2). Stratification by other features does not yield statistically significant results (Fig. S4).

In summary, CF smooth muscle and AP inflammation ratios are shown to be important prognostic factors for DMFS in the univariate and multivariate Cox regression analyses. Nonetheless, they are not shown to be independent of each other and therefore when one is used, the other should probably be disregarded.

Discussion

The goal of this study was to investigate the prognostic significance of novel image-based quantitative morphometric features derived from diverse cellular populations that constitute the tumor microenvironment of CRC with advanced primary tumors (T3/T4, N0, M0).

Feature	Univariate				Multivariate		
	Hazard ratio factor	Score test p-value	Log-rank test p-value	AUC	Hazard ratio factor	Wald test p-value	AUC
Standard histological features							
Differentiation (MD → PD)	1.306 (0.457,3.729)	0.575	0.575	0.455			
Histological type (Adenocarcinoma → Mucinous)	0.756 (0.175,3.261)	0.707	0.707	0.488			
T stage (pT3 → pT4)	1.853 (0.711,4.830)	0.200	0.200	0.537			
Connection frequency (CF) based tissue phenotypic features							
CF smooth muscle ratio (0.179 → 0.379)	1.770 (0.676,4.635)	0.025	0.066	0.617	2.106 (0.793,5.595)	0.014	0.586
CF inflammation ratio (0.037 → 0.103)	0.418 (0.186,0.943)	0.088	0.108	0.583	0.415 (0.183,0.945)	0.099	0.558
CF tumor-stroma interface ratio (0.116 → 0.241)	0.860 (0.411,1.798)	0.842	0.737	0.434	0.802 (0.379,1.696)	0.740	0.526
CF tumor ratio (0.079 → 0.222)	0.496 (0.235,1.047)	0.070	0.427	0.637	0.502 (0.233,1.083)	0.119	0.583
CF stroma ratio (0.172 → 0.275)	0.531 (0.246,1.144)	0.228	0.129	0.582	0.513 (0.234,1.121)	0.229	0.586
CF necrosis ratio (0.023 → 0.053)	0.731 (0.329,1.622)	0.363	0.369	0.511	0.624 (0.273,1.425)	0.291	0.526
Appearance (AP) based tissue phenotypic features							
AP smooth muscle ratio (0.141 → 0.341)	2.055 (0.840,5.029)	0.207	0.364	0.573	2.952 (1.2,7.262)	0.052	0.573
AP inflammation ratio (0.027 → 0.076)	0.376 (0.191,0.741)	0.001	0.001	0.674	0.389 (0.189,0.803)	0.004	0.638
Other features							
Morisita index ²⁴ (0.356 → 0.537)	1.460 (0.663,3.215)	0.279	0.388	0.526	1.376 (0.612,3.094)	0.343	0.517
Stroma-tumor ratio ^{8,11,12} (0.383 → 0.598)	0.934 (0.541,1.613)	0.363	0.818	0.537	0.864 (0.482,1.547)	0.332	0.543
Necrosis-tumor ratio ^{13,17,18} (0.074 → 0.224)	0.671 (0.295,1.526)	0.563	0.529	0.510	0.657 (0.289,1.494)	0.465	0.515

Table 4. Prognostic values of different features according to the Cox proportional hazards regression analysis on the UHCW cohort. Each morphological feature is adjusted by the standard histological features in the multivariate analysis. An interquartile change for a continuous variable or categorical change for a categorical variable is noted by ($x \rightarrow y$). A 95% confidence interval of the estimate of hazard ratio factor is noted by (x, y). A statistically significant result at the 0.05 significance level is highlighted in bold. AUC in the multivariate analysis refers to the AUC of the multivariate model rather than an individual feature.

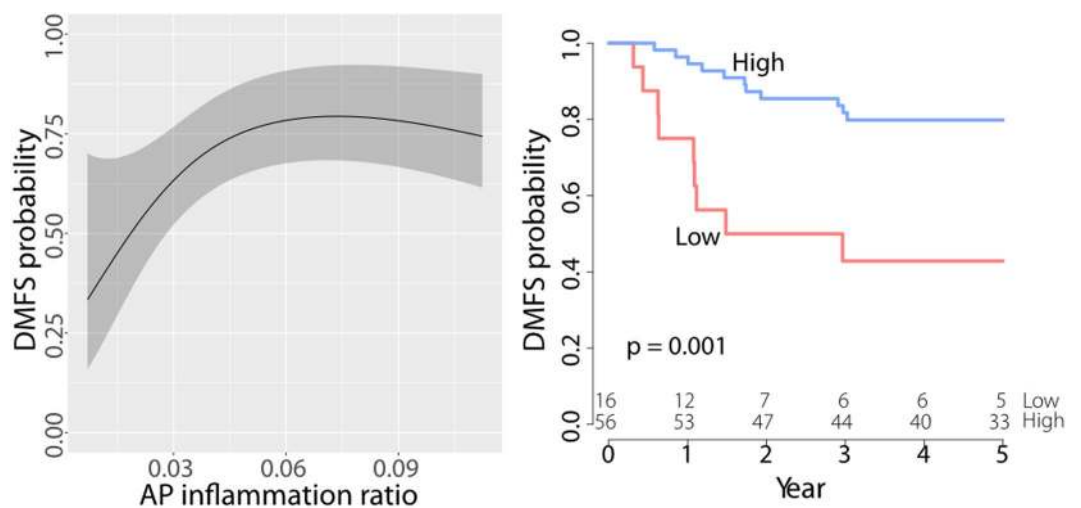


Figure 2. Prognostic values of the AP inflammation ratio in the univariate survival analysis. (Left) A 5-year DMFS estimate. The gray shaded regions indicate the 95% confidence intervals of the estimates. (Right) Kaplan-Meier curves. A log-rank p-value was computed for each pair of Kaplan-Meier estimates.

Digital phenotypic features vs histological features. To fully explore the rich microscopic level information available in a tissue section, we have developed an automated system to provide quantitative measurements and to avoid subjectivity from visual assessment. The analysis was conducted on WSIs of H&E-stained

formalin-fixed paraffin-embedded (FFPE) histological sections. Unlike previous works that identify diverse cellular components in a tumor section^{25,26}, our morphometric features are not limited to tumor cells, lymphocytes, and stromal cells, but also include other types of inflammatory cells, spindle-shaped cells, and necrotic debris. In addition, we explored the relationship between these cellular components through a cell-cell network in order to characterize the morphological and tissue phenotypic heterogeneity of tumor. Our system did not adopt a commonly used approach^{27,28} that calculates a large number of features followed by feature selection methods to select a handful of features suitable for the objectives of the analysis. Although such an exploratory approach has proved successful in some applications^{27,28}, the resulting features may not be easily interpretable in clinical terms. Moreover, if sufficiently many features are tried, it is likely that one of them will turn out to be “statistically significant” and so this approach requires follow-up tests of reproducibility. Instead, we investigated a small set of 8 features (6 CF + 2 AP phenotypic features), automatically found through unsupervised phenotyping and segmentation (see Methods for details). These features are visually meaningful as they correspond to distinct histological patterns of CRC tissue (Fig. 1).

Our systematic analysis shows that (a) the CF smooth muscle, CF inflammation ratios, and AP inflammation ratio are potentially independent markers predicting the occurrence of distant metastasis (binary logistic regression analysis) and (b) the CF smooth muscle and AP inflammation ratios are potential prognostic markers of 5-year DMFS for CRC patients diagnosed with advanced primary tumor (Cox proportional hazards regression analysis). CF smooth muscle ratio essentially measures the amount of the smooth muscle that is part of the colon wall. It quantifies the extent of spread and potential advancement of the tumors — the concept is related (but not similar) to other measures such as T stage, tumor-stroma ratio^{9,11,12}, and tumor border configuration^{29,30}. Low CF smooth muscle ratio is strongly associated with favorable prognosis. CF inflammation and AP inflammation ratios largely measure the amount of inflammation within the tumor tissue. High inflammation ratio is strongly associated with favorable prognosis, which supports the host-protective role of inflammatory cells in CRC that has been described by several studies^{8,31,32}. From this observation, one may hypothesize about the biological relevance of each of our automatically derived tissue phenotypes for tumor development and progression.

In logistic regression and survival analyses, the obtained AUCs for CF smooth muscle, CF inflammation and AP inflammation ratios are approximately within the range of 0.57–0.64. This indicates that the classifiers and survival predictors perform better than random but considered not satisfactory in general. This may imply that the metastasis risk of CRC cannot be rigorously assessed by only a single or a few variable(s). Leveraging other sources of information, such as molecular data, clinical record, and other imaging modalities, in conjunction with the proposed tissue phenotypic signature is one of the possibilities to address the performance issue.

The prognostic value of stroma-tumor ratio^{9,11,12} and necrosis-tumor ratio^{13,17,18} could not be confirmed in this study. It should be emphasized that, in those studies, both the ratios were semi-quantitatively measured in manually selected tumor-rich areas and were inevitably prone to observer bias. By contrast, our study measured these quantities in a fully automated and quantitative manner from all regions of the tumor section and therefore can be considered to be more objective and reproducible.

Uncertainty found in our analysis pertaining to the prognostic impact of standard clinical factors has also been confirmed in existing literature^{33–38}. Despite the fact that tumor differentiation has been consistently shown to be a prognostic feature independent of stage^{39–43}, the conventional grading process is subjective by its very nature and can exhibit a substantial degree of observer variability^{33,34}. It is also worth noting that according to the revised WHO criteria⁴⁴, only poorly differentiated tumor histology without mismatch repair protein deficiency is considered a high-risk factor. Presence of the mucinous histologic type in general is not an independent prognostic factor, given that available results are contradictory^{35,36}. Recent data have demonstrated the primary tumor extent (T4 stage) to be a likely prognostic factor for recurrence/metastasis^{45–47}. Nevertheless, like other semi-quantitative features, there have been reports of variability in assessment of the degree of tumor extent^{37,38}. Results from our analysis also indicate that T4 tumors have adverse DMFS outcome compared to T3 tumors, though the difference is not statistically significant.

The samples in this study come from patients diagnosed with stage II (Dukes stage B) CRC. This is characterized by advanced primary tumor with neither lymph node nor distant metastasis involvement (T3/T4, N0, M0). Stage II CRC consists of a heterogeneous population; some subgroups appear more likely to develop distant metastasis than others. Although adjuvant chemotherapy treatment is effective in other stages of the disease, there is a limited incremental benefit that stage II CRC patients could derive from this type of treatment in general^{48–50}. Due to the high financial cost and morbidity of the treatment coupled with uncertainty over which patients will relapse, there has long been a debate as to whether adjuvant chemotherapy treatment should be given to the patients, since a majority of the patients will already have been cured by surgical resection alone. In the absence of molecular or genetic predictive markers for chemotherapy response^{45,51,52}, improved prognostication accuracy seems to be the only key to better identify candidates who could potentially benefit the most from systemic therapies and thereby avoid unnecessary overtreatment as well as provide more efficient use of healthcare resources.

Even though several histological features have been demonstrated to be potential prognostic markers for recurrence or distant metastasis in stage II CRC, their prognostic significance is less clear and needs further validation. Primary tumor (T) stage and the number of lymph nodes examined have been recommended as risk factors by the National Comprehensive Cancer Network⁵³. Moreover, there is a controversy as to whether examining more lymph nodes can, in fact, reduce tumor staging error and in turn result in improved stage II patient survival^{54,55}. High-frequency microsatellite instability has been associated with improved disease-free survival in one study⁵² while in another study the effect was the opposite⁵¹. Gene expression profile is another factor that has shown promise for prediction of recurrence^{45,46}.

Study limitations. Based on the makeup of our dataset and the results from our analysis, we hypothesize that high CF smooth muscle ratio and low CF or AP inflammation ratios are potential risk factors for distant metastasis in stage II CRC. There are nevertheless some limitations of this study as described below.

Firstly, while it is possible that there may be subjectivity in the sample selection by the pathologist when selecting slides showing the deepest invasion into the bowel wall and/or the worst differentiated parts of the tumor, that was considered as part of the inclusion criteria in this study (Methods). Furthermore, most tumor sections used in this study were of $2\text{--}3\text{ cm}^2 \times 4\text{--}5\ \mu\text{m}$ across the face of the tumour in the horizontal plane. In volumetric terms, this is clearly a small proportion of the tumour and it is possible that there is sampling bias, as is the case for most such studies. It is also worth mentioning that the limited number of metastatic cases ($n = 27$) in this study may have rendered the analyses underpowered to detect prognostic effects of some features.

Secondly, although our cell detection and classification approach²³ was developed to be robust to a certain degree of variation of images arising from factors such as stain inconstancy, batch effects, failed autofocus, and artefacts in the tissue preparation process, it remains to be tested if the degree of variation is excessive. Good image quality is therefore critical if the system is to produce accurate results. This issue can be addressed by careful tissue preparation and slide scanning.

Thirdly, due to the nature of the H&E stain and cellular morphology, our system is capable of identifying only a limited number of cell categories that are somewhat coarse. IHC stains could provide an effective means of identifying more specific cell types, such as different types of immune cells and fibroblasts (normal fibroblasts or CAFs), at the additional costs of IHC slide preparation and associated antibodies.

Fourthly, the phenotyping proposed in this work was done on the basis of local cell-cell connection frequencies and also on the basis of appearance and other important contextual information such as tissue textures. This, on the one hand, can be seen as a limitation of the proposed quantitative tissue phenotyping approach, as it relies on local cell populations to generate global statistics. On the other hand, a number of studies have reported that normal cells of various types undergo transformation when coming into contact with tumor cells, thus resulting in some of the previously normal cells exhibiting new biological functions different from the original ones. The proposed approach focuses on cellular morphology and cellular context and avoids influences from other possibly misleading contextual information.

Finally, our analysis was based on a single dataset consisting of two independent cohorts from different institutes. To further confirm the reproducibility of the results and generalizability of our automated histologic quantification system, large-scale validation using independent cohorts from multiple institutes is required. To be translated into clinical practice, these limitations will need to be carefully addressed.

The outlook. With the increasing uptake of digital slide scanning technology in histopathology laboratories, digitized WSIs will gradually replace glass slides in routine pathology workflow⁵⁶. This presents an opportunity to advance image analytical techniques and computational algorithms for quantitative analysis of tissue morphology and consequently to provide an accurate and reproducible means for the diagnosis and prognostication of cancers. This is the first step towards effective treatment, decision-making, and personalized medicine with computational support. In this work, we have demonstrated the usefulness of such morphometric tools to reveal prognostic features in CRC. Our morphometric analysis is not restricted to images of FFPE CRC tissues but is also applicable to frozen tissue images as well as to images from different types of cancers. This morphometric approach was not designed to replace pathologists, but rather to provide additional information to assist in their diagnostic decision-making and risk stratification. Another potentially important direction would be to investigate potential associations between genomic alterations and digital tissue phenotypic signatures reflecting measurable aspects of in the tumor microenvironment.

Methods

Experimental design. The main objective of this study was to assess the significance of tissue phenotypic features for determining distant metastasis in advanced primary CRC. Specifically, we asked what quantitative tissue phenotypic features are biologically meaningful and important in predicting the subsequent development of distant metastasis and the distant-metastasis-free survival. Based on results from our statistical analyses, we have shown that digital tissue phenotypic features are independent prognostic factors for distant metastatic potential in CRC patients with advanced primary tumors (T3/T4, N0, M0). The sample size for logistic and Cox proportional hazards regression analyses was calculated based on the concept of events per variable^{57–59} which⁶⁰ indicates that a minimum of 30 metastatic subjects would be sufficient to control for a type I error rate at 7%, 95% CI coverage of 93%, and a relative bias of 7% of the estimate in the Wald test. We retrospectively recruited 130 CRC subjects (90 UHCW + 40 HGH) with advanced primary tumors. The enrollment was stopped when the calculated sample size was reached. We excluded cases without a 5-year distant metastasis status or with clinical evidence of metastasis at the time of diagnosis. We further excluded outlier cases whose tissue section had no tumor. In total, 28 cases (18 UHCW + 10 HGH) were excluded and there were 27 metastatic cases left. Our analyses were conducted on H&E-stained WSIs of tumor sections. In view of the limited number of cases, randomization was not used in any experiments.

Patient and clinical information. This study involved two independent cohorts of CRC patients from two institutes. The first cohort consisted of 72 patients initially admitted for CRC treatment during the years 2006 to 2010 at University Hospitals Coventry and Warwickshire (UHCW), Coventry, UK. The second cohort comprised 30 patients admitted during the years 2007 to 2012 at Hamad General Hospital (HGH), Doha, Qatar. For each case, clinical data included tumor histological type, differentiation, stage of the primary tumor (T), lymph node metastasis (N), and distant metastasis (M). The 5-year DMFS data were available only for UHCW cases. All CRC patients were diagnosed with locally advanced tumors (T3/T4) and negative lymph node (N0), and distant

metastasis free (M0). The TNM classification was reviewed and conducted according to the AJUCC/UICC-TNM staging system¹⁹. Summary details of the clinical information are given in Table 1.

The data used for this study including the WSIs and clinical information was provided after de-identification and informed patient consent was obtained from all subjects. Ethics approval for this study was obtained from the National Research Ethics Service North West (REC reference 15/NW/0843) and the Medical Research Center (RC/35213/2013) for the HGH cohort. All the experiments were carried out in accordance with approved guidelines and regulations.

Histological samples and imaging. For each case, tissue sections were prepared from an FFPE tumor tissue block and were then stained with H&E. Each tissue section was prepared in the pathology laboratory of the UHCW hospital. Histological slides were digitally scanned using the Omnyx VL120 Scanner (GE Omnyx, LLC) with an $\times 40$ setting (equivalent to $0.275 \mu\text{m}/\text{pixel}$). The scanned images were manually reviewed to control for failed autofocus. The tumor slides of all the cases were reviewed by the pathologists (DS, YT and IM) and the slides showing the deepest invasion into the bowel wall and/or the worst differentiated parts of the tumor, were selected for analysis. The reviewing pathologists agreed with the selection of slides in all cases.

Detection and classification of cells based on nuclear appearance. Two separate convolutional neural networks (CNNs) were trained, one for detection and another for classification of cells²³. A spatially-constrained CNN produced a probability map assigning to each pixel the probability of being the center of a cell. Subsequently, the locations of cells were estimated by the local maxima of the probability map. To classify a detected cell, multiple small sub-images in the neighborhood of the detected cell were extracted and then fed to the neighboring ensemble predictor (NEP). The NEP was trained to classify 4 cell types: malignant epithelial cells, inflammatory cells (including eosinophils, lymphocytes, and neutrophils), spindle-shaped cells (including normal fibroblasts, CAFs and smooth muscle cells), and necrotic debris.

The training and validation of the two algorithms were carried out on a dataset consisting of more than 20,000 cells, annotated by an experienced pathologist and a trained observer. The pixel resolution of images in the dataset was reduced to $0.55 \mu\text{m}$ (equivalent to using a $\times 20$ microscope objective). This dataset consisted of certain H&E-stained WSIs from cases that were initially excluded from the study. Based on a 2-fold cross-validation, the cell detection algorithm achieved an F1-score of 0.802 and the cell classification algorithm a multiclass AUC score⁶¹ of 0.917. For more details of the cell detection and classification method and the running time of the methods, see Sirinukunwattana *et al.*²³ and Table S3.

Quantifying local tissue characteristic. We first split a WSI into small non-overlapping image tiles of size $200 \times 200 \mu\text{m}^2$ (Fig. 1a), which was within the limit of effective intercellular communication distance⁶². For each image tile, a cell network (in computational terms, a graph) was constructed based on cell detection and classification results (Fig. 1b). The vertices of the network represent cells of different types. The network itself is the associated Delaunay triangulation (Fig. 1c), so that an edge represents a connection between a pair of neighboring cells. The edges connecting cells in one tile with cells in an adjacent tile were not considered. Since there are 4 cell classes, there are 10 possible pairs of cell-cell connections in the network. We then used the distribution of different cell-cell connection types (Fig. 1d) to characterize a given image tile.

Tissue phenotyping using cell-cell connection frequencies. In order to group image tiles into different phenotypes, we first calculate a feature vector based on cell-cell connection frequencies. We consider the 4-element set $A = \{M, I, S, N\}$, where M denotes the malignant epithelial type, I the inflammatory type, S the spindle-shaped type, and N the necrotic debris type. We also identify A with 1, 2, 3, 4 and define an indexing set $Q = \{(i, j) | i \geq j\}$. Let $h = [h_{(i,j)} | (i, j) \in Q] \in R^{10}$ be the ten-dimensional cell-cell connection frequency vector representing the frequencies of all cell-cell connections, where $h_{(i,j)} \in [0, 1]$ denotes the proportion of connection frequencies between cells of types i and j . We calculated this vector for every image tile extracted from every WSI in the dataset.

Next, we performed k -medoid clustering on all frequency vectors, calculated as above, for all tiles in all WSIs in the dataset in order to group image tiles into different phenotypes. This unsupervised algorithm (we used the k -medoid algorithm implemented in Matlab 2016b) automatically finds a set of medoids — representative frequency vectors for tile phenotypes within the data — and assigns a phenotype label to each tile according to its nearest medoid. We employed the Chi-squared distance between a frequency vector h and a medoid m given by:

$$d(h, m) = \sum_{k \in Q} \frac{(h_k - m_k)^2}{h_k + m_k}$$

We initialized the medoids randomly and ran the clustering algorithm 100 times for each trial. We then used the results from the replicate that yielded the smallest total sum of distances between the frequency vectors and their corresponding medoids. The criteria used to determine the number of phenotypes k were the similarity between the phenotypes and the correlation between tissue morphometric features derived from the phenotypes (described below). The similarity between a pair of phenotypes was measured in terms of the Chi-squared distance between the pair of medoids representing the phenotypes. Correlation between a pair of features was measured by the Spearman correlation coefficient. In order to find a suitable number of distinct phenotypes k , we chose the maximum number of phenotypes that produced relatively high values of Chi-squared distance and relatively low values of correlation between distinct features. A distance value less than 0.2 and a correlation coefficient value greater than 0.8 were considered undesirable. We found that $k = 6$ is the maximum number of phenotypes that satisfies both criteria (Fig. S5).

Examples of image tiles from different tissue phenotypes discovered using cell-cell connection frequencies are shown in Fig. 1e. As can be observed in Fig. 1e, the six connection frequency (CF) based phenotypes found automatically corresponded well with the following distinct tissue phenotypes: smooth muscle, inflammation, tumor-stroma interface, tumor, stroma, and necrosis.

Tissue phenotyping based on appearance. We also trained a deep learning based CNN for patch-based tissue phenotyping, in which the following 9 categories of image patches were explicitly considered: normal, non-tissue denotes the proportion of connection background, loose connective tissue (submucosa), fat (adipose), stroma (desmoplasia), inflammation, necrosis, smooth muscle, and tumor. Each image patch was of size 32×32 pixels with a pixel resolution of $2.2 \mu\text{m}/\text{pixels}$ ($\sim 5 \times$ objective). The architecture of the CNN was a simplified version of that proposed by Simonyan *et al.*⁶³.

In developing this appearance (AP) based approach to tissue phenotyping, we used a dataset consisting of 193 sub-images, each of size $1,346 \times 982$ pixels. These images were extracted from WSIs of cases that were initially excluded from the study. A trained observer (KS) annotated all images. We randomly split the images into three parts with 52.5% for training, 17.5% for validation, and 30% for testing. Each WSI contributed images to only one part of the split. For training and validation, we extracted multiple patches of size 32×32 pixels from the training and validation images. We selected the version of the algorithm that yielded the best performance on the validation part. In testing, for each test image, we extracted patches in a sliding-window fashion and classified each of them separately before merging the results together to obtain a segmentation result for the whole image. The correct classification accuracies for the 9 tissue phenotypes were as follows: normal 98.9%, non-tissue background 99.9%, loose connective tissue (submucosa) 98.4%, fat (adipose) 97.9%, stroma (desmoplasia) 90.4%, inflammation 99.3%, necrosis 98.2%, smooth muscle 97.5%, and tumor 96.0%.

We ran the trained segmentation algorithm on the 108 H&E-stained WSI images, used in the analyses. Examples of the segmentation results can be seen in Fig. S1. Furthermore, as a quality control, segmentation results of 10 images (out of 108 images) were randomly selected and then reviewed by expert pathologists (DS, IC).

Automatically-derived tissue phenotypic features. The CF and AP based tissue phenotypic features were calculated as follows:

$$\text{phenotype ratio} = \frac{\text{area of the tissue phenotype}}{\text{total tissue area}}$$

Here, the tissue area was computed from all tissue types excluding the normal and fat regions. The other tissue phenotypic features were quantified as follows:

$$\text{stroma-tumor ratio} = \frac{\text{stroma area}}{\text{stroma area} + \text{tumor area}}$$

$$\text{necrosis-tumor ratio} = \frac{\text{necrosis area}}{\text{necrosis area} + \text{tumor area}}$$

where stroma, tumor and necrosis areas were obtained from the AP based phenotyping results.

Statistical analyses. Our analysis did not distinguish well differentiated from moderately differentiated tumors—as recommended by Compton *et al.*^{64,65}, this helps to avoid contradictory labelling by two different observers, or even by a single observer, looking at the same sample on two different occasions. Missing data were filled in with 100 imputed values using the multiple imputation method implemented in the R ‘mice’ library⁶⁶. Analyses were performed on every imputed dataset and the results were combined to yield an overall estimate⁶⁷. The significance level was set to 0.05 for all the tests described below.

Association between the tissue phenotypic and standard clinical features was tested by the Mann-Whitney test and the strength of association was determined through coefficients of determination (r^2) of the test^{68,69}. The median p-value and r^2 were reported for a variable with multiple imputed values. We used the ‘rms’ library in R⁷⁰ to fit logistic regression models, to calculate the area under the receiver operating characteristic curve (AUC), and to perform survival and bootstrap analyses.

Logistic regression analysis was performed to assess the predictive power of each phenotypic feature in identifying patients with a propensity for distant metastasis development. Effects of the automatically-derived features were gauged after adjusting for the standard clinical variables and cohort indicator variable in multivariate logistic regression models. A total of 102 cases (72 UHCW and 30 HGH) were used in the analysis. The 5-year metastasis status was treated as a binary outcome and features were treated as predictors in regression models. Estimated odds ratio and its 95% CI were obtained for each feature to quantify the risk of distant metastasis development associated with the feature. We reported the factor of change in odds ratio when the value of a feature changes from the baseline value to the new value. For a continuous feature, the baseline and the changed values were set to the 1st and 3rd quartiles of the feature. Furthermore, likelihood ratio p-values were computed to assess goodness of fit of predictive models contributed by various features.

Survival analysis was performed to determine the prognostic value for DMFS associated with each feature. Univariate and multivariate Cox proportional hazards regression analyses were conducted on 72 cases from the UHCW cohort for which DMFS data were available. The former was used to evaluate the prognostic impact of each feature separately while the latter was used to assess the prognostic value of image-based tissue phenotypic features while adjusting for the effects of the clinical features. Rao’s score test and Wald test were employed in the univariate and multivariate analyses, respectively, to test whether the regression coefficient corresponding to a

particular feature in the Cox proportional hazards model was nonzero. Note that the score test is equivalent to the log-rank test when only a single categorical feature is considered in the model⁷¹. Hazard ratio and 95% CI estimates were obtained for each feature. To internally validate the performance of each fitted Cox proportional hazards model in predicting the survival probability, a bootstrap routine⁷² with 100 resampling replicates was employed to estimate the AUC. The statistical significance difference between survival stratifications was determined through the log-rank test using the R 'survival' library⁷³. The cutoff with minimum p-value was used for stratification, and the p-value was adjusted according to Altman's correction⁷⁴ in case of a continuous feature.

Data Availability

Extracted image features and codes to perform statistical analyses have been included in the Supplementary Information files. The datasets used and analyzed in the current study will be made available upon request.

References

- Bissell, M. J. & Hines, W. C. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. medicine* **17**, 320–329 (2011).
- Stoker, A. W., Hatier, C. & Bissell, M. J. The embryonic environment strongly attenuates v-src oncogenesis in mesenchymal and epithelial tissues, but not in endothelia. *The J. cell biology* **111**, 217–228 (1990).
- Weaver, V. M. *et al.* Reversion of the malignant phenotype of human breast cells in three-dimensional culture and *in vivo* by integrin blocking antibodies. *The J. cell biology* **137**, 231–245 (1997).
- Whiteside, T. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* **27**, 5904 (2008).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. medicine* **19**, 1423–1437 (2013).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA: a cancer journal for clinicians* **66**, 7–30 (2016).
- Galon, J. *et al.* Type, density and location of immune cells within human colorectal tumors predict clinical outcome. *Sci.* **313**, 1960–1964 (2006).
- Mesker, W. E. *et al.* The carcinoma–stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. *Anal. Cell. Pathol.* **29**, 387–398 (2007).
- Crispino, P. *et al.* Role of desmoplasia in recurrence of stage II colorectal cancer within five years after surgery and therapeutic implication. *Cancer investigation* **26**, 419–425 (2008).
- West, N. *et al.* The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients. *Br. journal cancer* **102**, 1519 (2010).
- Huijbers, A. *et al.* The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. *Annals. Oncol.* **24**, 179–185 (2012).
- Jayasinghe, C., Simiantonaki, N. & Kirkpatrick, C. J. Histopathological features predict metastatic potential in locally advanced colon carcinomas. *BMC cancer* **15**, 14 (2015).
- Dvorak, H. F. Tumors: wounds that do not heal. *New Engl. J. Medicine* **315**, 1650–1659 (1986).
- Schäfer, M. & Werner, S. Cancer as an overhealing wound: an old hypothesis revisited. *Nat. reviews Mol. cell biology* **9**, 628–638 (2008).
- Tommelein, J. *et al.* Cancer-associated fibroblasts connect metastasis-promoting communication in colorectal cancer. *Front. oncology* **5** (2015).
- Pollheimer, M. J. *et al.* Tumor necrosis is a new promising prognostic factor in colorectal cancer. *Hum. pathology* **41**, 1749–1757 (2010).
- Richards, C. *et al.* Prognostic value of tumour necrosis and host inflammatory responses in colorectal cancer. *Br. J. Surg.* **99**, 287–294 (2012).
- Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals surgical oncology* **17**, 1471–1474 (2010).
- Punt, C., Koopman, M. & Vermeulen, L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat. Rev. Clin. Oncol* **14**, 235–246 (2017).
- Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. biotechnology* **29**, 1120–1127 (2011).
- Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochimica et Biophys. Acta (BBA)-Reviews on Cancer* **1805**, 105–117 (2010).
- Sirinukunwattana, K. *et al.* Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging* **35**, 1196–1206 (2016).
- Maley, C. C., Koelble, K., Natrajan, R., Aktipis, A. & Yuan, Y. An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. *Breast Cancer Res.* **17**, 131 (2015).
- Yuan, Y. *et al.* Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. translational medicine* **4**, 157ra143–157ra143 (2012).
- Nawaz, S., Heindl, A., Koelble, K. & Yuan, Y. Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Mod. Pathol.* **28**, 766 (2015).
- Beck, A. H. *et al.* Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. translational medicine* **3**, 108ra113–108ra113 (2011).
- Yu, K.-H. *et al.* Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. communications* **7** (2016).
- Koelzer, V. H. & Lugli, A. The tumor border configuration of colorectal cancer as a histomorphological prognostic indicator. *Front. oncology* **4** (2014).
- Karamitopoulou, E. *et al.* Tumour border configuration in colorectal cancer: proposal for an alternative scoring system based on the percentage of infiltrating margin. *Histopathol.* **67**, 464–473 (2015).
- Pages, F. *et al.* Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene* **29**, 1093 (2010).
- Ohtani, H. Focus on TILs: prognostic significance of tumor infiltrating lymphocytes in human colorectal cancer. *Cancer Immun. Arch.* **7**, 4 (2007).
- Jass, J. *et al.* The grading of rectal cancer: historical perspectives and a multivariate analysis of 447 cases. *Histopathol.* **10**, 437–459 (1986).
- Compton, C. C. Pathology report in colon cancer: what is prognostically important? *Dig. Dis.* **17**, 67–79 (1999).
- Hyngstrom, J. R. *et al.* Clinicopathology and outcomes for mucinous and signet ring colorectal adenocarcinoma: analysis from the National Cancer Data Base. *Annals surgical oncology* **19**, 2814–2821 (2012).
- Kim, S. H. *et al.* Prognostic value of mucinous histology depends on microsatellite instability status in patients with stage III colon cancer treated with adjuvant FOLFOX chemotherapy: a retrospective cohort study. *Annals surgical oncology* **20**, 3407–3413 (2013).

37. Zeng, Z. *et al.* Serosal cytologic study to determine free mesothelial penetration of intraperitoneal colon cancer. *Cancer* **70**, 737–740 (1992).
38. Shepherd, N. A., Baxter, K. J. & Love, S. B. The prognostic importance of peritoneal involvement in colonic cancer: a prospective evaluation. *Gastroenterol.* **112**, 1096–1102 (1997).
39. Chapuis, P. *et al.* A multivariate analysis of clinical and pathological variables in prognosis after resection of large bowel cancer. *Br. journal surgery* **72**, 698–702 (1985).
40. Griffin, M. R., Bergstrahl, E. J., Coffey, R. J., Beart, R. W. & Melton, L. J. Predictors of survival after curative resection of carcinoma of the colon and rectum. *Cancer* **60**, 2318–2324 (1987).
41. Wiggers, T., Arends, J. W. & Volovics, A. Regression analysis of prognostic factors in colorectal cancer after curative resections. *Dis. colon & rectum* **31**, 33–41 (1988).
42. Newland, R. C., Dent, O. F., Lyttle, M. N., Chapuis, P. H. & Bokey, E. L. Pathologic determinants of survival associated with colorectal cancer with lymph node metastases. A multivariate analysis of 579 patients. *Cancer* **73**, 2076–2082 (1994).
43. Jessup, J. M., Stewart, A. K. & Menck, H. R. The National Cancer Data Base report on patterns of care for adenocarcinoma of the rectum, 1985–1995. *Cancer* **83**, 2408–2418 (1998).
44. Hamilton, S. R., *et al.* WHO classification of tumours. Pathology and genetics of tumours of the digestive system. Geneva: World health organization (2000).
45. Kerr, D. *et al.* A quantitative multigene RT-PCR assay for prediction of recurrence in stage II colon cancer: Selection of the genes in four large studies and results of the independent, prospectively designed QUASAR validation study. *J. Clin. Oncol.* **27**, 4000–4000 (2009).
46. Salazar, R. *et al.* Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *J. clinical oncology* **29**, 17–24 (2010).
47. Tsikitis, V. L., Larson, D. W., Huebner, M., Lohse, C. M. & Thompson, P. A. Predictors of recurrence free survival for patients with stage II and III colon cancer. *BMC cancer* **14**, 336 (2014).
48. Hartung, G. *et al.* Adjuvant therapy with edrecolomab versus observation in stage II colon cancer: a multicenter randomized phase III study. *Oncol. Res. Treat.* **28**, 347–350 (2005).
49. Quasar Collaborative Group. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomized study. *The Lancet* **370**, 2020–2029 (2007).
50. Schippinger, W. *et al.* A prospective randomised phase III trial of adjuvant chemotherapy with 5-fluorouracil and leucovorin in patients with stage II colon cancer. *Br. journal cancer* **97**, 1021 (2007).
51. Ribic, C. M. *et al.* Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *New Engl. J. Medicine* **349**, 247–257 (2003).
52. Kim, G. P. *et al.* Prognostic and predictive roles of high-degree microsatellite instability in colon cancer: a National Cancer Institute–National Surgical Adjuvant Breast and Bowel Project Collaborative Study. *J. Clin. Oncol.* **25**, 767–772 (2007).
53. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: colon cancer http://www.nccn.org/professionals/physician_gls/PDF/colon.pdf (2010).
54. Wong, S. L. *et al.* Hospital lymph node examination rates and survival after resection for colon cancer. *Jama* **298**, 2149–2154 (2007).
55. Moore, J., Hyman, N., Callas, P. & Littenberg, B. Staging error does not explain the relationship between the number of lymph nodes in a colon cancer specimen and survival. *Surg.* **147**, 358–365 (2010).
56. Snead, D. R. *et al.* Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathol.* **68**, 1063–1072 (2016).
57. Concato, J., Peduzzi, P., Holford, T. R. & Feinstein, A. R. Importance of events per independent variable in proportional hazards analysis I. Background, goals and general strategy. *J. clinical epidemiology* **48**, 1495–1501 (1995).
58. Peduzzi, P., Concato, J., Feinstein, A. R. & Holford, T. R. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *J. clinical epidemiology* **48**, 1503–1510 (1995).
59. Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. A simulation study of the number of events per variable in logistic regression analysis. *J. clinical epidemiology* **49**, 1373–1379 (1996).
60. Vittinghoff, E. & McCulloch, C. E. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology* **165**(6), 710–718 (2007).
61. Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. learning* **45**, 171–186 (2001).
62. Francis, K. & Palsson, B. O. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl. Acad. Sci.* **94**, 12258–12262 (1997).
63. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
64. Compton, C. C. *et al.* Prognostic factors in colorectal cancer: College of American Pathologists consensus statement 1999. *Arch. pathology & laboratory medicine* **124**, 979–994 (2000).
65. Compton, C. C. Updated protocol for the examination of specimens from patients with carcinomas of the colon and rectum, excluding carcinoid tumors, lymphomas, sarcomas and tumors of the vermiform appendix: a basis for checklists. *Arch. pathology & laboratory medicine* **124**, 1016–1025 (2000).
66. Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. statistical software* **45** (2011).
67. Rubin, D. B. *Multiple imputation for nonresponse in surveys*, vol. 81 (John Wiley & Sons, 2004).
68. Cohen, B. H. *Explaining psychological statistics* (John Wiley & Sons, 2008).
69. Fritz, C. O., Morris, P. E. & Richler, J. J. Effect size estimates: current use, calculations and interpretation. *J. experimental psychology: Gen.* **141**, 2 (2012).
70. Harrell Jr, F. rms: Regression Modeling Strategies. R package version 4.5-0, 2016 (2016).
71. Therneau, T. & Grambsch, P. *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health <https://books.google.co.th/books?id=9kY4XRuUMUsC> (Springer, 2000).
72. Harrell, F. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression and survival analysis* (Springer, 2015).
73. Therneau, T. A package for survival analysis in S. R package version 2.38, 2015 (2015).
74. Altman, D. G., Lausen, B., Sauerbrei, W. & Schumacher, M. Dangers of using optimal cutpoints in the evaluation of prognostic factors. *JNCI: J. Natl. Cancer Inst.* **86**, 829–835 (1994).

Acknowledgements

This paper was made possible by NPRP grant number NPRP5-1345-1-228 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors. The authors would like to acknowledge the contribution by Sean James, Kayleigh Patterson, Dr. Aisha Meskiri and Dr. Asha Rupani who involved in the preparation, staining and scanning of histology slides of the CRC samples used in this study.

Author Contributions

N.R., D.S., I.C. and K.S. designed the study. Z.A. and D.S. collected the clinical data. I.M. and D.S. reviewed and graded the histological samples. D.S., Y.T. and I.M. reviewed and selected slides for analysis. K.S. and N.R. developed the image analysis tools. Y.T. and K.S. generated ground truth data for training the nuclear detection and classification algorithm. Y.T., D.S. and I.C. reviewed the results of the algorithms. K.S. conducted all the statistical analyses. N.R., D.E. and I.C. supervised the statistical analyses and interpreted the results. K.S. drafted the manuscript. All authors were involved in discussion of the results and finalization of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-31799-3>.

Competing Interests: The authors declare no competing interests.

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



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

© The Author(s) 2018