Automated analysis of lymphocytic infiltration, tumor budding, and their spatial relationship improves prognostic accuracy in colorectal cancer

Ines P. Nearchou[1], Kate Lillard[2], Christos G. Gavriel[1], Hideki Ueno[3], David J. Harrison[1], Peter D. Caie[1]

[1]Quantitative and Digital Pathology, School of Medicine, University of St Andrews, St Andrews, KY16 9TF, UK

[2]Indica Labs, Inc, 2469 Corrales Rd Bldg A-3, Corrales, NM 87048 USA

[3]Department of Surgery, National Defense Medical College, 3-2 Namiki, Tokorozawa, Saitama 359-8513, Japan

Running Title: Automated image analysis in colorectal cancer prognosis

Keywords: digital image analysis, digital pathology, tumor budding, immunoscore, prognosis

Additional Information:

Corresponding Author: Ines P. Nearchou
School of Medicine,
University of St Andrews
St Andrews, KY16 9TF, UK
Tel: +44 (0)1334 463603
Fax: +44 (0)1334 467470
Email: i.p.nearchou@gmail.com

Disclosure/conflict of interest: K. Lillard is a full-time employee of Indica Labs, Inc.

Word Count: 3703

Total number of figures and tables: 6

Abstract:

Both immune profiling and tumor budding significantly correlate with colorectal cancer (CRC) patient outcome,but are traditionally reported independently. This study evaluated the association and interaction between lymphocytic infiltration and tumor budding, coregistered on a single slide, in order to determine a more precise prognostic algorithm for patients with stage II CRC. Multiplexed immunofluorescence and automated image analysis were used for the quantification of $CD3^+CD8^+$ T cells, and tumor buds (TBs), across whole slide images of three independent cohorts (training cohort: n = 114, validation cohort 1: n = 56, validation cohort 2: n = 62). Machine learning algorithms were used for feature selection and prognostic risk model development. High numbers of TBs (HR = 5.899, 95% CI, 1.875 - 18.55), low $CD3^+$ T cell density (HR = 9.964, 95% CI 3.156 - 31.46), and low mean number of $CD3^+CD8^+$ T cells within 50 μm of TBs (HR = 8.907, 95% CI 2.834 - 28.0) were associated with reduced disease-specific survival. A prognostic signature, derived from integrating TBs, lymphocyte infiltration, and their spatial relationship, reported a more significant cohort stratification (HR = 18.75, 95% CI 6.46–54.43), than TBs, Immunoscore, or pT stage. This was confirmed in two independent validation cohorts (HR = 12.27, 95% CI 3.524–42.73, HR = 15.61, 95% CI 4.692-51.91). The investigation of the spatial relationship between lymphocytes and TBs within the tumor microenvironment improves accuracy of prognosis of patients with stage II CRC through an automated image analysis and machine learning workflow.

Introduction:

Tumor Node Metastasis (TNM) staging remains the gold standard for stratification of colorectal cancer (CRC) patients into prognostic subgroups, and is fundamental for treatment selection (1,2). Surgical resection without the use of adjuvant treatment results in long-term disease-free survival in most stage II CRC cases. However, 20% of patients experience disease-specific death (3). In addition, some stage III patients experience better outcomes than some stage II patients. This is in spite of revisions to the TNM guidelines, which, for example, subdivide pT stage II into three categories (4). Alongside pT stage, tumor differentiation aids the pathologist in sub-stratifying patients into high or low risk of tumor progression; however grade is highly subjective (5). Tumor budding, which morphologically resembles small, poorly-differentiated clusters of cells (1-4 cells) disseminating at the invasive margin, has consistently been correlated with poor CRC prognosis (6–8).

Both TNM staging and tumor budding concentrate solely on the tumor, but it is now well established that the host interaction, such as immune infiltrate, within the tumor microenvironment (TME) plays a crucial role in tumor progression (9,10). Specifically, in CRC, the pioneering studies of Galon and colleagues demonstrated that $CD3^+$ and $CD8^+$ T-cell infiltration was an independent prognostic factor predicting patient survival (9,11). Tumor buds (TBs) are hypothesized to represent a highly invasive tumor subpopulation, especially if they also have the ability to evade the host immune response (6–8). The interaction between immune response and TBs may therefore hold vital information pertaining to the tumor's potential to metastasize.

Previous research has demonstrated the ability to automatically quantify $CD3^+$ and $CD8^+$ T cells within the TME, thereby achieving a greater level of standardization than with subjective, manual reporting (9). Similarly we have previously shown that tumor budding can be standardized through automated image analysis (12,13). Although the automatic quantification of both immune infiltrate and tumor budding has been associated with patient prognosis, they have not been studied together. Furthermore, by coregistration on the same tissue section it is possible to accurately study the spatial interaction with each other. We apply such an approach in this study to investigate if quantifying both features alongside their spatial interaction allows for a more precise and accurate prognosis for stage II CRC patients.

Materials and methods

Patient Material and Clinical Data

This study included three independent cohorts of stage II CRC patients who underwent surgical resection in Edinburgh, UK, hospitals over the years 2002–2004 (n = 170) and the National Defense Medical College Hospital (NDMCH), Japan, over the years 2006–2011(n = 62). The formalin-fixed paraffin-embedded (FFPE) block that contained the deepest cancer invasion for each patient was selected after the review of hematoxylin and eosin (H&E) slides. Cancer blocks from Edinburgh hospitals comprised specimens of cross-sectional cut whereas the ones from the NDMCH were of longitudinal cut. Associated clinical data included features from the original pathology report, such as TNM staging and differentiation as well as detailed follow-up information including disease specific survival (DSS); up to 11.5 years for the Edinburgh cohort and 8.6 years for the Japanese cohort. 114 consecutive patients who underwent surgical resection in Edinburgh hospitals over the years 2002–2003 were used as a training cohort for this study. All available patients who underwent surgical resection in Edinburgh hospitals in the following year (2004) were used as a first validation cohort (n = 56). We then used a similar number of patients who underwent surgical resection in the NDMCH, Japan over the years 2006–2011 as an international validation set (n = 62). The clinicopathological data of all the cohorts are shown in Table 1. This study was conducted in accordance with the declaration of Helsinki. Ethical approval was obtained after review by the NHS Lothian NRS BioResource, REC-approved Research Tissue Bank (REC approval ref: 13/ES/0126), granted by East of Scotland Research Ethics Service and by the Ethics Committee of the National Defense Medical College (approval ref: No.2992). The ethical approval permitted us to use archival diagnostic samples once they were de-identified for the purposes of the research.

Immunofluorescence and Image Capture

Immunohistochemistry antibody optimization

We selected externally validated specific primary antibodies: primary antibody against $CD8^+$ cytotoxic T cells (monoclonal mouse anti-human CD8, M7103, Dako, Glostrup, DK) was previously used for the international validation of the consensus Immunoscore by Pages et al.(14), primary antibody against pan T cells (Polyclonal Rabbit Anti-Human CD3, A045201-2, Dako, Glostrup, DK) used by Harter et al. (15), in assessment of tumor-infiltrating lymphocytes in human brain metastases and for pan-cytokeratin (PanCK) (mouse monoclonal anti-human cytokeratin, M351501-2, Dako, Glostrup, DK) previously used by Koelzer et al. (16), for the assessment of TBs in CRC.

In order to assess the in-house antibody performance, each antibody was individually assessed by brightfield immunohistochemistry (IHC). The primary antibodies against CD3 and CD8 T cells were each tested individually on 3μm thick serial sections of normal tonsil FFPE tissue, known to have high lymphocytic cell concentration. For assessment of PanCK primary antibody, a CRC optimizing tissue microarray, consisting of 100 cores taken from primary tissue blocks, was used. Expression of all cell markers was detected using 3,3'-Diaminobenzidine (DAB) chromogen (K3468, Dako, Glostrup, DK). Nuclei were counterstained with hematoxylin (RBA-4213-00A, CellPath, Newton, UK). In parallel, H&E

(RBB-0100-00A, CellPath, Newton, UK) staining was performed on serial sections and specificity of antibodies from the IHC was assessed and compared visually by an experienced pathologist (DJH) and research scientists (IPN and PDC). No CD8 positive but CD3 negative cells were present, giving confidence to the antibodies' specificity. Positive and negative controls were used in each immunohistochemistry staining; human tonsil FFPE tissue without primary antibody for CD3 and CD8 T cells and a CRC optimizing tissue microarray without primary antibody for PanCK. Western blot analysis was also used to assess the specificity of the PanCK antibody using CRC cell lines.

Immunofluorescence staining optimisation

Once specificity of antibodies was confirmed in brightfield, each target was assessed by a uniplex immunofluorescence. Uniplex immunofluorescence was performed on serial sections of the CRC optimising tissue microarray (specified above) with the use of individual tyramide signal amplification (TSA) fluorescence kits. After deparaffinization, slides were placed in a microwaveable pressure cooker containing boiling antigen retrieval (AR) buffer (0.1M Sodium Citrate pH6) and microwaved for 5 min at 75 °C. Slides were allowed to cool down in running tap water for 20min at room temperature and were then rinsed with 0.1% Phosphate buffered saline-Tween 20 (PBS-T). PBS-T was used for all washing steps. Tissues were then blocked with 3% hydrogen peroxide for 5 minutes and Dako serum free protein block (X090930-2, Dako, Glostrup, DK) was then applied for 10 minutes. Slides were then incubated for 30 min with the same antibodies as for immunohistochemistry at these dilutions: CD3 (dilution 1:400), CD8 (dilution 1:200) and PanCK (dilution 1:100). Slides were then washed (3 x 5min). The slides that had previously been incubated with CD3 or CD8 antibodies were then incubated in pre-diluted anti mouse or anti rabbit HRP conjugated secondary antibodies (K400111-2 and K400311-2 respectively, Dako, Glostrup, Denmark) for 30 min. Following 3 more washes of 5 min, slides were incubated for 10 min with either TSA® fluorescein (FITC) for CD3 antibody visualisation or TSA® Cyanine 5 for CD8 antibody visualisation (NEL741B001KT and NEL745B001KT respectively, Perkin Elmer, MA, USA) at 1:100 dilution. The slide that had been previously incubated with PanCK antibody was then incubated in anti-mouse AlexaFluor 555 (A21422, ThermoFisher, MA, USA) in antibody diluent (S080983-2, Dako, Glostrup, DK) at 1:100 dilution for 30 min. Finally, following a washing step, the slides were counterstained for 10 min with Hoechst 33342 (H3570, ThermoFisher, MA, USA) at 1:20 dilution in de-ionised water. Sections were then dehydrated in 80% ethanol and mounted with ProLong Gold Antifade Reagent (P36930, Thermo Fisher Scientific Inc., MA, USA). Optimal antibody dilutions in the uniplex immunofluorescence slides were chosen to ensure best signal to noise and specific staining pattern when digitised on the Zeiss Axioscan.z1 whole slide scanner (Zeiss, Oberkochen, DE). Fluorescence slides were compared with serial section of brightfield DAB slides using the same antibodies. Positive and negative controls were included in each staining run. The positive controls for the CD3 and CD8 antibodies were the tonsil tissue as specified above.

Once the appropriate dilution of primary antibodies, secondary antibodies and fluorophores were found under uniplex immunofluorescence, we then assessed them multi-plexed on a single tissue section. In order to eliminate species cross reactivity, microwave stripping was performed between the CD8 and PanCK staining. Microwave stripping was performed as follows. AR buffer was heated in a pressure cooker in the microwave for 12 min prior to adding the slides in the solution. Slides were then microwaved for 17 min using the auto defrost function. Sections were then allowed to cool down in running tap water for 20 min and washed for 5 min. The immunofluorescence protocol utilised to label the study samples was performed as above using a DAKO Link48 Autostainer (Dako, Glostrup, DK). The first

run of labelling included the primary antibody against CD8 prior to antibody stripping, and labelling with a primary antibody cocktail of CD3 and PanCK prior to counterstaining with Hoechst. The multi-plexed section was compared to the uniplex labelled serial sections and we confirmed that there was no increase or decrease in the level of specific signals or nonspecific background signals.

Whole slide fluorescence images were captured with a 20x objective using an Axioscan.Z1 (Zeiss, Oberkochen, DE). Exposure times for Hoechst, Cy3, Cy5 and FITC channels were 12, 120, 8 and 25ms, respectively (see supplementary methods for detailed protocol of antibody optimization and multiplexed immunofluorescence).

Digital Image Analysis

Digital whole slide fluorescence images, in Zeiss' .czi file format, were uploaded into HALO Next Generation Image Analysis software (version 2.1.1637.18) (IndicaLabs, Corrales, NM) for image analysis. Full algorithm workflow and settings are shown below:

Nuclear segmentation:

All nuclei in the whole slide image were automatically segmented using the commercially available High-Plex FL (version 2.0) module within the HALO Next Generation image analysis software. Optimized module parameters for nuclei detection included dye weights (Hoechst nucleus weight = 4, Cy5 nucleus weight = 0.232, FITC nucleus weight = 0.232), nuclear contrast threshold (0.5), minimum nuclear intensity (0.079), nuclear segmentation aggressiveness (0.65) and using the default nuclear size setting (1-500μm^2). All samples were analyzed using these specific settings.

Lymphocytic Infiltration Analysis

Within the same software module in which we segmented the nuclei, (High-Plex FL), cells were then classified as CD3 (FITC) or CD8 (Cy5) positive based on dye nucleus positive threshold (Cy5 = 0.132, FITC = 0.15), cytoplasm positive threshold (Cy5 = 0.075, FITC = 0.5) and membrane positive threshold (Cy5 = 0.075, FITC = 0.5). The algorithm settings were kept constant across all patients. Next, the algorithm automatically quantified the number of each lymphocyte classification across the whole slide image. The flood tool was used to identify the invasive front. An invasive margin (IM) was created which consisted of a width span of 500µm in and out from the invasive front. The tumor core area (CT) was set to be the whole tumor area excluding the 500µm before the invasive front (Fig. 1A). Lymphocytes ($CD3^+$ and $CD8^+$ T cells) infiltrating both the CT and the IM of the tumor were automatically quantified (Fig. 1B) and their densities (cells/ mm^2) were exported to assess their prognostic significance.

Immunoscore Evaluation

Immunoscore was performed based on the methodology described previously by Galon et al(9). Using the lymphocytic infiltration analysis results, $CD3^+$ and $CD8^+$ lymphocytes were counted in both the IM and CT resulting in the reporting of the following features: $CD3^+$ T cells in the CT, $CD3^+$ T cells in the IM, $CD8^+$ cells in the CT and $CD8^+$ T cells in the IM. Optimal cutoff points for the four features were calculated based on survival data of the

training cohort (Supplementary Table S1). Values above the optimal cutoff points for each feature were scored 1 whereas values below were scored 0. An immunoscore was then calculated by the sum of all scores for each patient sample. Finally, similarly to Liu et al. (17), patients were stratified into two groups; High Immunoscore > 2 and Low Immunoscore ≤ 2.

Tumor Budding Analysis

The PanCK intensity within each whole slide image was initially assessed using the Area Quantification FL v1.2 module without any changes to the commercially available module thresholds. Patient samples were grouped into two categories based on their resulting PanCK intensity; low (average dye intensity < 2.16×10^{-2}) and high (average dye intensity > 2.16×10^{-2}). Using samples from within each PanCK group (high or low), machine learning classifiers were trained to segment tumor from non-tumor region. Two High Plex FL (version 2.0) modules were next created; one using the classifier trained from low PanCK patient samples and one from the high PanCK patient samples. Patients with an average PanCK intensity below 2.16-02 were analyzed to segment tumor from non-tumor using the classifier termed 'low PanCK' and the other with the classifier termed 'high PanCK'. These modules were run within a 1000μm border inwards from the invasive front; termed the tumor budding region of interest (TBROI) (Fig. 1C). Within the tumor regions detected, nuclei were detected using the methodology described above and cancer cells were classified based on PanCK positivity in nucleus (0.1), cytoplasm (0.08) and membrane (0.05) (Fig. 1D). These cancer cell classification thresholds were kept constant across the entire cohort of images. TBs were classified as tumor clusters containing up to 4 $PanCK^+$ cells. TB density (TBs / mm^2) and number were then exported from the algorithm. Mucin pools and areas of necrosis were excluded.

Spatial Analysis of Lymphocytes and Tumor Buds

Spatial coordinates of the lymphocytes and TBs were imported into a spatial plot within the HALO software. From this plot, the Spatial Analysis algorithm was utilized to calculate the number of $CD3^+$ and $CD8^+$ T cells within a range between 0-100 μm radii of TBs (Fig. 1E) in consecutively increasing 10 μm steps therefore creating 11 classes; 0-10 μm, 0-20 μm, 0-30 μm and so on. All classes of radius size were significantly associated with survival, however after assessing the prognostic significance of iterative groupings of varying radius sizes, there was a statistical difference observed between the 0-50 μm and the 0-100 μm groupings. The classes were therefore merged into two categories: 0-50 μm and 0-100 μm radii of TBs. Finally, the number and densities of $CD3^+$ and $CD8^+$ T cells within these distances were established.

All features exported from the image analysis are listed in Supplementary Table S2.

Statistical Analysis

All statistical analysis, unless otherwise stated, was performed in RStudio 1.1.419 running R 3.4.3 (18,19). Continuous image analysis data, categorical clinical data and survival data were loaded into R. The relationship between lymphocyte distribution patterns, TBs, and

patient characteristics was assessed using the Chi-Square test. For comparison with categorical patient data, image analysis features were divided into low and high groups according to optimal cut-off points based on survival data. To assess the relationship between lymphocytes and TBs, Spearman correlation coefficients (r) were applied on continuous data. All $r = 0$ values indicated no association between features, $r > 0$ indicated a positive association where as one variable increases, so does the other and $r < 0$ indicated a negative association where as one variable increases the other decreases. Associations between features were assessed on the training cohort.

Univariate Cox regression was performed on both the categorical clinical and continuous image analysis data. All $P < 0.05$ values were considered statistically significant.

The least absolute shrinkage and selection operator (LASSO) penalized Cox proportional hazard regression was performed for identification of significant features (20), using the glmnet package (21). The significant features identified by LASSO were then inputted into a random forest (RF) ($n = 500$) decision tree model using the randomForest package (22) and were ranked according to their mean decrease Gini coefficient. Features with a mean decrease Gini of above 4 were taken forward for further analysis. Optimal cutoff points for these 4 features were calculated and iterative combinations of these features were tested. The model with most predictive value was then selected; termed the "tumor bud-immuno spatial index" (TBISI). In order to avoid overfitting during the development of the model, validation was performed for both LASSO Cox regression model (10-fold cross validation) and random forest (out of the bag). Univariate Cox regression was performed for; the TBISI, pT stage, TB number and Immunoscore, and a bootstrap resampling technique was applied (number of samples = 1000) in SPSS 24 (23). Multivariate Cox Regression with a forward stepwise method and the Receiver Operating Characteristic (ROC) curve were also applied on to compare the TBISI, pT stage TB number and Immunoscore in SPSS 24. The prognostic value of TBISI was then assessed on two independent validation cohorts. All validations were performed using the fixed optimal cut-off points of the training cohort. All optimal cutoff points for the categorization of image analysis features including tumor budding, the immunoscore and for the development of the TBISI were based on disease specific survival and were calculated using the survminer package (24). All Kaplan Meier (KM) curves were plotted using the survival package (25). The P values from KM curves were corrected for false discovery rate (FDR) using the Benjamini-Hochberg procedure to account for multiple hypothesis testing (26).

Study workflow is illustrated in Supplementary Fig. S1.

Results

Patient Characteristics

The clinicopathological characteristics of all three cohorts of stage II CRC patients are shown in Table 1. This study cohort comprised a training cohort of 114 stage II CRC patients, of which 57 were female and 57 were male. The validation cohorts included 34 male and 22 female patients in validation cohort 1 and 42 male and 20 female patients in validation cohort 2. The range of patients' age was 37–96 years old. There were 43 right-sided colon carcinomas, 38 left-sided and 33 rectal cancers in the training cohort and 29 and 14 right-sided colon carcinomas, 9 and 22 left-sided and 18 and 26 rectal cancers in validation cohort 1 and 2, respectively.

Correlations of Patient Characteristics with Lymphocyte Density and TB Number

Automated image analysis was performed to quantify TBs (within TBROI), $CD3^+$ cell, and $CD8^+$ T-cell densities (within IM and CT). Optimal cutoff points based on survival data were used to categorize image analysis features of the training cohort into low and high groups (Supplementary Table S1). TB number and density were highly correlated with each other (correlation coefficient = 0.7466; $P < 2.2 \times 10^{-16}$; therefore, TB number was selected to be used for correlation assessments. The relationship between patient characteristics and both lymphocyte distribution patterns and TB number was then assessed. A high density of $CD3^+$ T cells at the IM was associated with no extramural lymphovascular invasion (EMLVI; P = 0.049), whereas in the CT it was correlated with moderate differentiation status and non-mucinous histological type (P = 0.007 and P = 0.023, respectively). A high density of $CD8^+$ cells at the IM and across the whole tumor section (WTS) was significantly associated with female gender (P = 0.016 and P = 0.030 respectively). Finally, advanced pT stage was significantly associated with a high TB number at the TBROI (P = 0.031) (Table 2).

Relationship between Tumor Budding and Lymphocyte Density

TB number was tested for association with lymphocyte densities of the training cohort, using the Spearman correlation coefficient (Fig. 2). This indicated a weak negative relationship between TB number and all lymphocytic cell densities ($r < 0$) regardless of distribution pattern (ie. IM, CT, WTS) or lymphocytic cell subpopulations ($CD3^+$ or $CD8^+$). However, a higher density of $CD3^+$ T cells at the IM, $CD8^+$ T cells at the IM, and WTS were significantly correlated with a lower TB number (P = 0.016, P = 0.037, and P = 0.041 respectively) (Supplementary Table S3).

Survival Analysis - Clinicopathological Data

To assess the prognostic significance of the categorical clinicopathological data of the training cohort, univariate Cox-regression was applied. KM survival analysis was used to further assess any significant prognostic features from the clinicopathological data of the training cohort (Fig. 3). Univariate Cox-regression showed that only pT stage (HR =4.143; 95% CI, 1.480-11.570; P = 0.006) was significantly associated with disease specific survival (Table 1). KM survival function was then applied and showed that patients of T4 stage have poor DSS (73% survived) compared to T3 stage patients (91% survived) (P = 0.003) (Fig. 3A).

Survival Analysis - Image Analysis Features

To explore the prognostic effects of the quantified lymphocytic distribution patterns, TB number and density within the training cohort, we performed univariate Cox-regression analysis. This analysis was also performed for the spatial relationships between lymphocytes and TBs. The following specific continuous features were calculated and exported from the image analysis algorithm: a) in IM, CT and WTS; densities of $CD3^+$ and $CD8^+$ T cells and ratio of $CD3^+$ to $CD8^+$ T cells; b) TB number and density at the TBROI, c) within 0–50 μm and 0–100 μm radius of TBs; number and mean number of $CD3^+$ and $CD8^+$ T cells. The densities of both $CD3^+$ and $CD8^+$ T cells were significantly associated with disease specific survival ($P < 0.05$). TB number and density conferred prognostic significance (HR =1.001; 95% CI, 1.000–1.001; $P = 0.0004$ and HR =1.030; 95% CI, 1.000–1.060; $P = 0.0446$ respectively). KM analysis showed TB number to be a significant prognostic factor; 95% of patients in the low tumor budding group survived compared to the high tumor budding group, in which 73% of patients survived ($P = 0.0006$) (Fig. 3D). Furthermore, a high mean number of lymphocytes near TBs was significant in predicting disease specific survival for stage II CRC patients ($P < 0.05$) (Supplementary Table S2).

Significant Feature Selection

All clinicopathological and image analysis data of the training cohort (listed in Supplementary Table S2) were input for a LASSO penalized Cox proportional hazard regression. This was performed to identify the features that add significant value to the prediction of disease specific survival over time and therefore to reduce the data dimensionality. The analysis reported that there were 8 significant candidate features ($CD3^+$ T-cell density at WTS, mean $CD3^+CD8^+$ T-cell number within 0–50 μm of TBs, TB Number, $CD8^+$ T-cell density in CT, pT stage, EMLVI, age, and differentiation). The 8 features which added value to the model were used as further input into a RF decision tree model for optimal prognostic feature selection. The RF model reported 4 features with a mean decrease Gini of above 4 ($CD3^+$ T-cell density in WTS, mean $CD3^+CD8^+$ T-cell number within 0–50 μm of TBs, TB number, $CD8^+$ T-cell density in CT) (Table 3).

Development of Tumor Bud Immuno-Spatial Index

Optimal cutoff points for these 4 features ($CD3^+$ T-cell density in WTS, mean $CD3^+CD8^+$ T-cell number within 0–50 μm of TBs, TB number, $CD8^+$ T-cell density in CT) were calculated for the training cohort using the patient survival data (389.6 cells/mm^2, 4.1, 1104.0, and 114.7 cells/mm^2, respectively). The least significant feature was then removed in an iterative process until the removal of a feature decreased the prognostic value of the model. This led to the creation of the TBISI consisting of 3 features; $CD3^+$ T-cell density in WTS, mean $CD3^+CD8^+$ T-cell number within 0–50 μm of TBs and TB number. Patients with $CD3^+$ T-cell density in WTS and mean $CD3^+CD8^+$ T-cell number within 0–50 μm of TBs below the cutoff point and TB number above the cutoff point were classified as a “high risk” of disease specific death group; the remainder of the patients were classified as a “low-risk” group. KM curves and bootstrap univariate cox regression p values were used to determine the significance of this index. The created TBISI was highly significant in the prediction of disease specific death from both KM analysis and cox regression (HR = 18.75, 95%CI 6.46-54.43; $p = 1.202 \times 10^{-13}$) (Fig. 3J). KM and univariate cox-regression analyses were also applied to pT stage (HR = 4.143, 95% CI 1.48-11.57; $P = 0.0033$) (Fig. 3A), TB number (HR

= 5.899, 95% CI 1.875–18.55; *P* = 0.0006) (Fig. 3D) and the studies' calculated Immunoscore (HR = 7.02, 95% CI 2.49–19.84; *P* = 1.922 x 10^{-05}) (Fig. 3G). Multivariate Cox regression model (forward stepwise) revealed that this TBISI was the only feature that added value to the model when pT stage, TB number, Immunoscore, and TBISI were included (Supplementary Table S4). Finally, the ROC curve showed that the TBISI had the greatest area under the curve (Area = 0.785, 95% CI 0.629–0.941) when compared to pT stage or each of the model's component individually (Supplementary Fig. S2).

Validation of Tumor Budding, Immunoscore and Tumor Bud-Immuno Spatial Index

The TBISI developed from the training set, alongside the Immunoscore and TBs, were evaluated on two independent validation cohorts; Validation Cohort 1 patients from Edinburgh hospitals (*n* = 56) and International Validation Cohort 2 patients from the NDMCH, Japan (*n* = 62). Cutoff points calculated from the training set data were directly applied to the two validation cohorts. Univariate Cox regression and KM analyses were calculated using the patient data from the two validation cohorts and used to assess the clinicopathological data, TB number, immunoscore, and the model TBISI (Table 1). pT stage was significantly associated with patient survival in validation cohort 1 (HR = 5.00, 95% CI 1.52–16.46; *P* = 0.008) (Fig. 3B), but not in the Japanese validation cohort 2 (Fig. 3C). TB number was significant in both validation cohort 1 and 2 (HR = 5.29, 95% CI 1.14–24.50; *P* = 0.033 and HR = 8.816, 95% CI 2.70–28.82; *P* = 0.0003) (Fig. 3E and 3F respectively), whereas the immunoscore was only significant in the validation cohort 1 (HR = 4.72, 95% CI 1.02–21.87; *P* = 0.047) (Fig. 3H), although trended to significance in validation cohort 2 (Fig. 3I).

TBISI was highly significant on both validation cohorts; Validation Cohort 1 with HR = 12.27, 95% CI 3.52–42.73; p = 8.2 x 10^{-05} and in Validation Cohort 2 with HR = 15.61, 95% CI 4.69–51.91; p = 7.42 x 10^{-06} (Fig. 3K and 3L, respectively). The TBISI demonstrated an increase in hazard ratio and significance in both validation cohorts than when analyzing the sum of its parts independently.

Discussion

Risk assessment and thus treatment choices for stage II CRC patients are in need of improvement. A growing awareness exists that the TME contributes to cancer differentiation, progression, invasion, and metastasis (27–29). In CRC, the two most promising candidate prognostic features involve two distinct aspects of the TME: tumor budding (2,30–32) and the Immunoscore (9,10,14,33). Tumor budding represents small clusters of infiltrating cancer cells within the IM, whereas the Immunoscore quantifies tumor infiltrating lymphocytes (TILs) within both the IM and the CT. However, these features are predominantly assessed independently of each other. The question, therefore, arises as to whether these two features combined augment their individual prognostic significance and whether their association and spatial relationship adds further prognostic value for stage II CRC patients. The work presented here utilizes automated image analysis to quantify and compare TBs and TILs coregistered to the same tissue section.

We report four major findings: first, we have shown significant associations between clinicopathological factors, $CD3^+$ $CD8^+$ T-cell densities, and TB number; second, we have confirmed that Immunoscore and tumor budding are strong independent prognostic factors for disease specific survival in stage II CRC; third, we have demonstrated that the spatial interaction between lymphocytes and TBs holds additional prognostic significance; and finally, a standardized prognostic index was created combining all the above features to achieve a higher significant cohort stratification than any individual prognostic feature tested.

Emerging technologies in tissue analysis have allowed further understanding of the role of the TME in disease progression. Automated image analysis has been shown to have a high potential for unbiased and reproducible assessment of histopathological features in CRC, such as for TBs, lymphatic vessel invasion (12), and TILs (14), as well as in other disease types (34,35). To an extent, this is assisted by the advancement of acquiring multiparametric data from a single tissue section, which can provide a more complete understanding of the underlying mechanisms within the TME. Even though some multiplex immunohistochemistry or immunofluorescence methodologies have shown the ability of high-level multiplexing on a single FFPE tissue sample, the analysis may be restricted to manually selected regions of interest or through their applications on tissue microarrays (36). Such methodologies can also be associated with long acquisition times for whole tissue sections and sequential stitching of image tiles into one large image (37). In this study we quantified TILs and TBs coregistered to a single whole slide section through the use of four-plexed immunofluorescence, whole slide scanning, and automated image analysis. This methodology permitted the colocalization of the four tissue markers, required to quantify the features measured in this study, across the whole tissue section and at single cell resolution. Spatial coordinates of each cell type were exported from the same tissue section ensuring more accurate analysis than is achieved from coregistering serial sections. This methodology therefore conserves precious tissue, reduces scanning time and cost, while also allowing the objective analysis of the entire TME and its cellular interactions.

Our results showed that the presence of high $CD3^+$ T-cell density correlated with favorable clinicopathological features, at the IM with the absence of EMLVI and at the CT with moderate differentiation status and non-mucinous histological type. These data concur well with prognostic analyses on the propitious impact of overall TIL densities in CRC patients

(10,38). We also found that a high CD8 T-cell density was correlated with the female gender, a finding that has previously been shown in invasive CRC by Jang (39). However, the mechanisms underlying this observation remain unknown. An elevated number of TBs was significantly associated with advanced pT stage, correlating favorably with previous literature showing TBs being associated with adverse prognostic features (40–42). These findings highlight the impairing effect T cells have on tumor progression, while suggesting that high tumor budding cancers harbor a more aggressive phenotype.

Our data indicate a significant inverse relationship between the densities of TILs and TB number, which substantiates the work of Zlobec et al. (43), who found the absence of TILs to be associated with presence of tumor budding. This may suggest that during the transition of TBs from epithelial to mesenchymal phenotype, they might occupy a niche that may not illicit an immune response. Indeed, it has been suggested that TBs lose MHC expression during this transition (7), and others have shown that TB positivity is significantly associated with PD-L1 positivity (44). The data presented here adds further support to the concept of TBs having the ability to evade the antitumoral host response and thereby embark on the first steps of the metastatic cascade.

Despite repeated evidence of tumor budding conferring prognostic significance in CRC (8,12,42,45), its routine clinical assessment has not yet been established due to the lack of standardized methodology of assessment. Most studies investigating tumor budding rely upon the use of manual quantification, often based on subjective selection of high magnification regions of interest (2). The use of cytokeratin immunohistochemistry has previously shown to improve the detection of TBs as well as increasing inter-observer reproducibility (30). Furthermore, cytokeratin immunohistochemistry has been recommended by the International Tumor Bud Consensus Conference to help TB quantification in challenging cases such as tumors regions with strong inflammation or ruptured glands (2). In the current study, tumor budding was identified by the use of cytokeratin immunofluorescence and quantified by automated and standardized image analysis across the entire TBROI. Previous studies have highlighted the difficulty in the qualitative assessment of TBs in circumstances of gland fragmentation, stromal cells mimicking TBs, inflammation (46), microvesicles, and membrane fragments that might be cytokeratin-positive (30). In order to address these concerns, we ensured the presence of whole nuclei (n = 1–4) for the TB identification and only counted the isolated clusters of cells as buds. Quantifying TBs at the whole TBROI has the advantage of being representative of the whole invasive margin rather than a manually selected region of interest. However, our methodology has the potential to "dilute" the significance of TBs represented by hotspots due to the tumor's innate heterogeneity. Like previous semi-quantitative studies (6,42) our quantification methodology reported TB number and density at the TBROI to be significantly correlated to poor patient survival.

The Immunoscore, proposed by Galon et al. (38,47,48), has been validated in a large international colon cancer study (14). It is not only a promising prognostic index (11,38), but it has been shown to be superior to TNM staging when assessing CRC prognosis (47,49). This score is based on the quantification of the *in situ* immune infiltration by the chromogenic labelling of two sequential tissue sections; one with $CD3^+$ T cells and one with $CD8^+$ T cells and quantified in two regions (IM and CT). By applying our automated image analysis workflow across a single tissue section, our Immunoscore results compared

favorably with those of Galon et al. (9), showing that a high density of TILs was correlated with better survival outcomes in stage II CRC.

In order to reflect the dynamic processes present at the tumor front, T lymphocytes in close (50 and 100 μm) proximity of TBs in a 2-dimensional plot were quantified. Our results revealed that a high number of T lymphocytes near TBs was associated with better survival. This supports the theory that TBs represent a heterogeneous subpopulation of cells and, although hypothesized to be more invasive than well differentiated tumor glands, retain the potential to be recognized by immune-surveillance mechanisms. This process has previously been described as "nipping in the bud" (43). This theory was also supported by a study on the microenvironment of TBs by Lugli et al. (50), reporting high lymphocyte to TB ratio as a good prognostic factor in patients with CRC and by Lang-Schwarz et al. (51), reporting the combination of both TILs and tumor budding predicting survival in CRC. Due to the high complexity and heterogeneity within the whole tissue slide, spatial mapping of specific features by eye is difficult, subjective, and time consuming. Selected regions of interest to study this interplay between different features and cells might not represent the spatial relationships present in the whole tissue. Here we propose a rapid and completely unbiased method capturing every single event at the single cell resolution and therefore better reporting on the interactions among TBs and TILs seen at the whole IM. Furthermore, such methodology can reveal specific patterns that may advance our understanding of the interactions occurring between cancer cells and their microenvironment.

Multiple features reported from the clinicopathological report and by automated image analysis demonstrated prognostic significance. In order to reduce the dimensionality of the data and identify the independently significant features, a Cox regression model with LASSO regularization was applied to all 32 reported features in this study. This reported a reduced and independently significant set of 8 features: 4 from the clinicopathological report and 4 from image analysis. This reduced feature set was input for a RF machine learning model, which consisted of 500 trees. The advantage of feature selection through RF is the thorough testing of all significant features across multiple models while building in validation. This machine learning application reported only 4 features with a high mean decrease Gini coefficient, all of which were exported from image analysis. In order to avoid overfitting of our model on the training cohort, we included multiple validation steps within the workflow. Through this, we demonstrate that a model in which patients with a combined low $CD3^+$ T-cell density in WTS, low mean number of TILs within 50 μm proximity to TBs, and high TB number conferred a survival disadvantage compared to any other combination. When a multivariate forward stepwise Cox regression including pT stage, TB number, Immunoscore, and TBISI was applied, the only feature reported to significantly add value to the model was the TBISI, whereas pT stage, TB number, and Immunoscore were not included in the equation. The new prognostic model of TBISI was then successfully validated across two independent cohorts, the second of which included patients with different genetic background, tumor genetic features, and gut microbiota, compared with the first cohort. This model and its cutpoints did not need to be altered for the Japanese validation cohort despite the fact that their tissue was sectioned longitudinally compared to the other cohorts that were cut at a cross-section. pT stage in the Japanese cohort, unlike both Edinburgh ones, was not significantly correlated with DSS in univariate analysis. This reflects the heterogeneous nature of stage II CRC disease and further highlights the need for additional methodologies,

such as the TBISI reported here, to more accurately stratify these patients for both prognostic reasons and potentially identifying patients that may benefit from adjuvant therapy.

In conclusion, this study reports an automated image analysis–based workflow with the ability to quantify the tumor-infiltrating immune cells; total T cells ($CD3^+$) and cytotoxic T cells ($CD8^+$), and TBs at the invasive front, across a single tissue section. Additionally, through a machine learning approach, our results show evidence that the spatial association of lymphocytes and TBs has a high prognostic significance in stage II CRC patients. This combinational prognostic model demonstrated augmented accuracy and precision over that gained by either characterizing the lymphocytic infiltration or tumor budding in isolation. Furthermore, the final model contained no clinical parameters, thus demonstrating the benefit of automated profiling of the TME to provide a more precise and accurate prognosis. The methodology completely removes human subjectivity and negates the need for experienced staff to spend valuable time quantifying complex features across large areas of a whole slide section. This research can serve as a base for future studies on the prognostic significance of the interplay between different cell types within the patient's heterogeneous and heterotypic TME.

Authors' Contributions

Conception and design: I.P. Nearchou, P.D. Caie
Development of methodology: I.P. Nearchou and P.D. Caie
Acquisition of data: I.P. Nearchou, C.G. Gavriel, P.D. Caie
Analysis and interpretation of data: I.P. Nearchou, P.D. Caie
Writing, review and/or revision of the manuscript: I.P. Nearchou, D.J. Harrison, P.D. Caie
Administrative, technical, or material support: I.P. Nearchou, K. Lillard, H.Ueno, D.J. Harrison, P.D. Caie
Study supervision: K. Lillard, D.J. Harrison, P.D. Caie

Acknowledgments

The authors would like to acknowledge Frances Rae from Tissue Governance NHS Lothian and Dr Yoshiki Kajiwara from the National Defense Medical College for securing ethics and patient follow-up data as well as In Hwa Um and Mustafa Elshani from the School of Medicine, University of St Andrews for cutting the tissue sections. Funding for the study was provided by Medical Research Scotland and Indica Labs, Inc. provided in-kind resource.

References:

1. Li J, Yi C-H, Hu Y-T, Li J-S, Yuan Y, Zhang S-Z, et al. TNM Staging of Colorectal Cancer Should be Reconsidered According to Weighting of the T Stage: Verification Based on a 25-Year Follow-Up. Medicine (Baltimore). Wolters Kluwer Health; 2016 Feb;95(6):e2711.

2. Lugli A, Kirsch R, Ajioka Y, Bosman F, Cathomas G, Dawson H, et al. Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. Mod Pathol. 2017 Sep 26;30(9):1299–311.

3. Nitsche U, Maak M, Schuster T, Künzli B, Langer R, Slotta-Huspenina J, et al. Prediction of Prognosis Is Not Improved by the Seventh and Latest Edition of the TNM Classification for Colorectal Cancer in a Single-Center Collective. Ann Surg. 2011 Nov;254(5):793–801.

4. Ajcc. AJCC Cancer Staging. 7th ed. Edge S, Byrd DR, Compton CC, Fritz AG, Greene F, Trotti A, editors. Springer; 2010. 648 p.

5. Fleming M, Ravula S, Tatishchev SF, Wang HL. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol. AME Publications; 2012 Sep;3(3):153–73.

6. Lugli A, Karamitopoulou E, Zlobec I. Tumour budding: a promising parameter in colorectal cancer. Br J Cancer. 2012;106.

7. Koelzer VH, Herrmann P, Zlobec I, Karamitopoulou E, Lugli A, Stein U. Heterogeneity analysis of Metastasis Associated in Colon Cancer 1 (MACC1) for survival prognosis of colorectal cancer patients: a retrospective cohort study. BMC Cancer. BioMed Central; 2015 Mar 21;15:160.

8. Koelzer VH, Canonica K, Dawson H, Sokol L, Karamitopoulou-Diamantis E, Lugli A, et al. Phenotyping of tumor-associated macrophages in colorectal cancer: Impact on single cell invasion (tumor budding) and clinicopathological outcome. Oncoimmunology. Taylor & Francis; 2016 Apr;5(4):e1106677.

9. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. J Pathol. 2014 Jan;232(2):199–209.

10. Pagès F, Kirilovsky A, Mlecnik B, Asslaber M, Tosolini M, Bindea G, et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. J Clin Oncol. 2009 Dec 10;27(35):5944–51.

11. Kwak Y, Koh J, Kim D-W, Kang S-B, Kim WH, Lee HS. Immunoscore encompassing CD3+ and CD8+ T cell densities in distant metastasis is a robust prognostic marker for advanced colorectal cancer. Oncotarget. Impact Journals, LLC; 2016 Dec 6;7(49):81778–90.

12. Caie PD, Turnbull AK, Farrington SM, Oniscu A, Harrison DJ. Quantification of tumour budding, lymphatic vessel density and invasion through image analysis in colorectal cancer. J Transl Med. 2014 Jun 1;12(1):156.

13. Caie PD, Zhou Y, Turnbull AK, Oniscu A, Harrison DJ. Novel histopathologic feature identified through image analysis augments stage II colorectal cancer clinical

reporting. Oncotarget. 2016;7(28).

14. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou F-S, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet. 2018 May 10;391(10135):2128–39.

15. Harter PN, Bernatz S, Scholz A, Zeiner PS, Zinke J, Kiyose M, et al. Distribution and prognostic relevance of tumor-infiltrating lymphocytes (TILs) and PD-1/PD-L1 immune checkpoints in human brain metastases. Oncotarget. Impact Journals, LLC; 2015 Dec 1;6(38):40836.

16. Koelzer VH, Assarzadegan N, Dawson H, Mitrovic B, Grin A, Messenger DE, et al. Cytokeratin-based assessment of tumour budding in colorectal cancer: analysis in stage II patients and prospective diagnostic experience. J Pathol Clin Res. Wiley-Blackwell; 2017 Jul;3(3):171–8.

17. Liu R, Peng K, Yu Y, Liang L, Xu X, Li W, et al. Prognostic Value of Immunoscore and PD-L1 Expression in Metastatic Colorectal Cancer Patients with Different RAS Status after Palliative Operation. Biomed Res Int. Hindawi; 2018 Jan 31;2018:1–8.

18. RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA [Internet]. 2015. Available from: http://www.rstudio.com/.

19. R: The R Project for Statistical Computing [Internet]. 2018. Available from: https://www.r-project.org/

20. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Vol. 58, Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1996. p. 267–88.

21. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010 Feb 2;33(1):1–22.

22. Andy Liaw and Matthew Wiener. Classification and Regression by randomForest. R News. 2002;2(3):18–22.

23. IBM Corp. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp. 2016.

24. Alboukadel Kassambara, Marcin Kosinski , Przemyslaw Biecek SF. Drawing Survival Curves using “ggplot2” [R package survminer version 0.4.2]. Comprehensive R Archive Network (CRAN); 2018;

25. Therneau TM, Grambsch PM. Modeling survival data : extending the Cox model. Springer; 2000. 350 p.

26. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Vol. 57, Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1995. p. 289–300.

27. Fridman WH, Zitvogel L, Sautès–Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol. 2017 Jul 25;14(12):717–34.

28. Barnes TA, Amir E. HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer SUBTYPES OF IMMUNE CELL INFILTRATE. 2017;117.

29. Soncin I, Sheng J, Chen Q, Foo S, Duan K, Lum J, et al. The tumour

microenvironment creates a niche for the self-renewal of tumour-promoting macrophages in colon adenoma.

30. Koelzer VH, Zlobec I, Berger MD, Cathomas G, Dawson H, Dirschmid K, et al. Tumor budding in colorectal cancer revisited: results of a multicenter interobserver study. Virchows Arch. 2015 May 21;466(5):485–93.

31. van Wyk HC, Park J, Roxburgh C, Horgan P, Foulis A, McMillan DC. The role of tumour budding in predicting survival in patients with primary operable colorectal cancer: A systematic review. Cancer Treat Rev. 2015 Feb;41(2):151–9.

32. Zlobec I, Lugli A. Tumour budding in colorectal cancer: molecular rationale for clinical translation. Nat Rev Cancer. 2018 Jan 29;18(4):203–4.

33. Pagès F, Galon J, Dieu-Nosjean M-C, Tartour E, Sautès-Fridman C, Fridman W-H. Immune infiltration in human tumors: a prognostic factor that should not be ignored. Oncogene. 2010 Feb;29(8):1093–102.

34. Brown JR, Wimberly H, Lannin DR, Nixon C, Rimm DL, Bossuyt V. Multiplexed Quantitative Analysis of CD3, CD8, and CD20 Predicts Response to Neoadjuvant Chemotherapy in Breast Cancer. Clin Cancer Res. 2014 Dec 1;20(23):5995–6005.

35. Donovan MJ, Scott R, Khan F m, Zeineh J, Fernandez G. The application of artificial intelligence and machine learning to automate Gleason grading: Novel tools to develop next generation risk assessment assays. J Clin Oncol. 2018 Feb 20;36(6_suppl):170–170.

36. Tsujikawa T, Kumar S, Borkar RN, Azimi V, Thibault G, Chang YH, et al. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. Cell Rep. NIH Public Access; 2017;19(1):203–17.

37. Lin J-R, Izar B, Wang S, Yapp C, Mei S, Shah PM, et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. Elife. 2018 Jul 11;7.

38. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the “Immunoscore” in the classification of malignant tumours. J Pathol. Wiley-Blackwell; 2014 Jan;232(2):199–209.

39. Jang TJ. Progressive Increase of Regulatory T Cells and Decrease of CD8+ T Cells and CD8+ T Cells/Regulatory T Cells Ratio during Colorectal Cancer Development. Korean J Pathol. Korean Society of Pathologists and Korean Society for Cytopathology; 2013 Oct;47(5):443–51.

40. Lohneis P, Sinn M, Klein F, Bischoff S, Striefler JK, Wislocka L, et al. Tumour buds determine prognosis in resected pancreatic ductal adenocarcinoma. Br J Cancer. Nature Publishing Group; 2018 May 14;1.

41. Gujam FJA, McMillan DC, Mohammed ZMA, Edwards J, Going JJ. The relationship between tumour budding, the tumour microenvironment and survival in patients with invasive ductal breast cancer. Br J Cancer. Nature Publishing Group; 2015 Sep 11;113(7):1066–74.

42. Ueno H, Murphy J, Jass JR, Mochizuki H, Talbot IC. Tumour “budding” as an index to estimate the potential of aggressiveness in rectal cancer. Histopathology. 2002

Feb;40(2):127–32.

43. Zlobec I, Lugli A, Baker K, Roth S, Minoo P, Hayashi S, et al. Role of APAF-1, E-cadherin and peritumoural lymphocytic infiltration in tumour budding in colorectal cancer. J Pathol. 2007 Jul;212(3):260–8.

44. Kim K-J, Wen X-Y, Yang HK, Kim WH, Kang GH. Prognostic Implication of M2 Macrophages Are Determined by the Proportional Balance of Tumor Associated Macrophages and Tumor Infiltrating Lymphocytes in Microsatellite-Unstable Gastric Carcinoma. Singh PK, editor. PLoS One. 2015 Dec 29;10(12):e0144192.

45. Tanaka M, Hashiguchi Y, Ueno H, Hase K, Mochizuki H. Tumor budding at the invasive margin can predict patients at high risk of recurrence after curative surgery for stage II, T3 colon cancer. Dis Colon Rectum. 2003 Aug;46(8):1054–9.

46. Mitrovic B, Schaeffer DF, Riddell RH, Kirsch R. Tumor budding in colorectal carcinoma: time to take notice. Mod Pathol. 2012 Oct 13;25(10):1315–25.

47. Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. Nature Publishing Group; 2012 Apr 15;12(4):298–306.

48. Galon J, Pagès F, Marincola FM, Angell HK, Thurin M, Lugli A, et al. Cancer classification using the Immunoscore: a worldwide task force. J Transl Med. 2012 Oct 3;10(1):205.

49. Lea D, Håland S, Hagland HR, Søreide K. Accuracy of TNM staging in colorectal cancer: a review of current culprits, the modern role of morphology and stepping-stones for improvements in the molecular era. Scand J Gastroenterol. Taylor & Francis; 2014 Oct 21;49(10):1153–63.

50. Lugli A, Karamitopoulou E, Panayiotides I, Karakitsos P, Rallis G, Peros G, et al. CD8+ lymphocytes/ tumour-budding index: an independent prognostic factor representing a “pro-/anti-tumour” approach to tumour host interaction in colorectal cancer. Br J Cancer. Nature Publishing Group; 2009 Oct 20;101(8):1382–92.

51. Lang-Schwarz C, Melcher B, Haumaier F, Lang-Schwarz K, Rupprecht T, Vieth M, et al. Budding and tumor infiltrating lymphocytes – combination of both parameters predicts survival in colorectal cancer and leads to new prognostic subgroups. Hum Pathol. W.B. Saunders; 2018 May 19;

Features	Training Cohort (*n* = 114)			Validation Cohort 1 (*n* = 56)			Validation Cohort 2 (*n* = 62)		
	Freq(%)	HR (95%CI)	*P*	Freq(%)	HR (95%CI)	*P*	Freq(%)	HR (95%CI)	*P*
Clinicopathological									
Age		1.466 (0.782-2.747)	0.233		1.603 (0.788-3.260)	0.192		2.177 (0.964-4.916)	0.061
≤70	46(40.4)			24(42.9)			40(64.5)		
71-79	32(28.0)			14(25.0)			18(29.0)		
≥80	36(31.6)			18(32.1)			4(6.5)		
Gender		0.851 (0.308-2.347)	0.755		0.956 (0.279-3.272)	0.943		0.415 (0.092-1.876)	0.253
Male	57(50.0)			34(60.7)			42(67.7)		
Female	57(50.0)			22(39.3)			20(32.3)		
pT stage		**4.143 (1.48-11.57)**	**0.006**		**5.003 (1.521-16.460)**	**0.008**		2.313 (0.497-10.770)	0.285
pT3	88(77.2)			46(82.1)			56(90.3)		
pT4	26(22.8)			10(17.9)			6(9.7)		
Tumour site		0.663 (0.337-1.305)	0.234		0.620 (0.255-1.510)	0.293		2.590 (1.062-6.319)	0.037
Left	38(33.3)			9(16.1)			22(35.5)		
Right	43(37.7)			29(51.8)			14(22.6)		
Rectal	33(28.9)			18(32.1)			26(41.9)		
Differentiation		2.025 (0.968-4.238)	0.061		1.251 (0.252-6.204)	0.784		0.694 (0.378-1.276)	0.240
Moderate	91(79.8)			14(25.0)			20(32.3)		
Poor	20(17.5)			8(14.3)			7(11.3)		
Well	3(2.6)			34(60.7)			34(54.8)		
N/A	0			0			1(1.61)		
Max.tum.diam		1.224 (0.411-3.646)	0.716		1.590 (0.448-5.637)	0.473		N/A	
<5	58(50.9)			25(44.6)					
≥5	40(35.1)			26(46.4)					
N/A	16(14.0)			5(8.9)					
Ap.node		0.366 (0.082-1.641)	0.189		**0.240 (0.636-0.910)**	**0.035**		N/A	
None	8(7.0)			6(10.7)					
≥1	101(88.6)			50(89.3)					
N/A	5(4.4)								
EMLVI		0.342 (0.117-1.000)	0.050		**0.113 (0.020-0.636)**	**0.013**		N/A	
Yes	18(15.8)			3(5.36)					
No	96(84.2)			34(60.7)					
N/A	0			19(33.9)					
Tumour type		<0.001 (0-Inf)	0.998		1.858 (0.839-4.110)	0.126		1.395 (0.178-10.950)	0.751
Adenocarcinoma	105(92.1)			45(80.4)			56(90.3)		
Mucinous	5(4.4)			8(14.3)			5(8.1)		
Mixed	4(3.5)			3(5.4)			0		
N/A	0			0			1(1.6)		
ImageAnalysis									
Immunoscore		**7.020 (2.485-19.840)**	**<0.001**		**4.720 (1.019-21.870)**	**0.047**		3.314 (0.968-11.350)	0.057
Low	22 (19.3)			27 (48.2)			20 (32.3)		
High	92 (80.7)			29 (51.8)			42 (67.7)		
TBNumber		**5.899 (1.875-18.55)**	**<0.001**		**5.289 (1.142-24.500)**	**0.033**		**8.816 (2.697-28.820)**	**<0.001**
Low	73 (64.0)			29 (51.8)			48 (77.4)		
High	41 (36.0)			27 (48.2)			14 (22.6)		
TBISI		**18.75 (6.460-54.430)**	**<0.001**		**12.270 (3.524-42.730)**	**<0.001**		**15.610 (4.692-51.910)**	**<0.001**
Low	102 (89.5)			45 (80.4)			56 (90.3)		
High	12 (10.5)			11 (19.6)			6 (9.7)		

Table 1. Univariate Cox Regression analysis for clinicopathological data, Immunoscore, Tumor Bud (TB) Number and Tumor Bud-Immuno Spatial Index (TBISI) for all three cohorts. Significant features ($P < 0.05$) are shown in bold. Max.tum.diam = Maximum tumor diameter, Ap.node. = Apical node, EMLVI = Extramural lymphovascular invasion.

Feature	Total Freq(%)	CD3⁺ density (cells/mm²)						CD8⁺ density (cells/mm²)						TB Number	
		IM		CT		WTS		IM		CT		WTS			
		Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
		n = 11	*n* = 103	*n* = 35	*n* = 79	*n* = 29	*n* = 85	*n* = 36	*n* = 78	*n* = 16	*n* = 98	*n* = 21	*n* = 93	*n* = 73	*n* = 41
Age		*P* = 0.521		*P* = 0.536		*P* = 0.532		*P* = 0.151		*P* = 0.267		*P* = 0.164		*P* = 0.681	
≤70	46(40.4)	6(5.3)	40(35.1)	14(12.3)	32(28.0)	12(10.5)	34(29.8)	18(15.8)	28(24.6)	6(5.3)	40(35.1)	11(9.7)	35(00.0)	31(27.1)	15(13.2)
71-79	32(28.0)	3(2.6)	59(25.4)	12(10.5)	20(17.5)	10(8.8)	22(19.3)	11(9.7)	21(18.4)	7(6.1)	25(21.9)	7(6.1)	25(21.9)	21(18.4)	11(9.7)
≥80	36(31.6)	2(1.8)	34(29.8)	9(7.9)	27(23.7)	7(6.1)	29(25.4)	7(6.1)	29(25.4)	3(2.6)	33(28.9)	3(2.6)	33(28.9)	21(18.4)	15(13.2)
Gender		p = 0.341		p = 0.839		p = 0.830		p = 0.016*		p = 0.590		p = 0.030*		p = 0.845	
Male	57(50)	7(6.1)	50(43.9)	17(14.9)	40(35.1)	15(13.2)	42(36.8)	24(21.1)	33(28.9)	9(7.9)	48(42.1)	15(13.2)	42(36.8)	36(31.6)	21(18.4)
Female	57(50)	4(3.5)	53(46.5)	18(15.8)	39(34.2)	14(12.3)	43(37.7)	12(10.5)	45(39.5)	7(6.1)	50(43.9)	6(5.3)	51(44.7)	37(32.5)	20(17.5)
pT Stage		p = 0.254		p = 0.052		p = 0.139		p = 0.390		p = 0.385		p = 0.904		p = 0.031*	
pT3	88(77.2)	10(8.8)	78(68.4)	23(20.2)	65(57.0)	19(16.7)	69(60.5)	26(22.8)	62(54.4)	11(9.7)	77(67.5)	16(14.0)	72(63.2)	61(53.5)	27(23.7)
pT4	26(22.8)	1(0.9)	25(21.9)	12(10.5)	14(12.3)	10(8.8)	16(14.0)	10(8.8)	16(14.0)	5(4.4)	21(18.4)	5(4.4)	21(18.4)	12(10.5)	14(12.3)
Tumour site		p = 0.735		p = 0.250		p = 0.827		p = 0.283		p = 0.971		p = 0.857		p = 0.467	
Left	38(33.3)	4(3.5)	34(29.8)	15(13.2)	23(20.2)	11(9.7)	27(23.7)	15(13.2)	23(20.2)	5(4.4)	33(28.9)	8(7.0)	30(26.3)	23(20.2)	15(13.2)
Right	43(37.7)	3(2.6)	40(35.1)	13(11.4)	30(26.3)	10(8.8)	33(28.9)	10(8.8)	33(28.9)	6(5.3)	37(32.5)	7(6.1)	36(31.6)	26(22.8)	17(14.9)
Rectal	33(28.9)	4(3.5)	29(25.4)	7(6.1)	26(22.8)	8(7.0)	25(21.9)	11(9.7)	22(19.3)	5(4.4)	28(24.6)	6(5.3)	27(23.7)	24(21.1)	9(7.9)
Differentiation		p = 0.366		p = 0.007*		p = 0.233		p = 0.932		p = 0.549		p = 0.739		p = 0.354	
Moderate	91(79.8)	8(7.0)	83(72.8)	22(19.3)	69(60.5)	20(17.5)	71(62.3)	28(24.6)	63(55.3)	13(11.4)	78(68.4)	20(17.5)	71(62.3)	61(53.5)	30(26.3)
Poor	20(17.5)	2(1.8)	18(15.8)	12(10.5)	8(7.0)	8(7.0)	12(10.5)	7(6.1)	13(11.4)	2(1.8)	18(15.8)	8(7.0)	12(10.5)	10(8.8)	10(8.8)
Well	3(2.6)	1(0.9)	2(1.8)	1(0.9)	2(1.8)	1(0.9)	2(1.8)	1(0.9)	2(1.8)	1(0.9)	2(1.8)	1(0.9)	2(1.8)	2(1.8)	1(0.9)
Max.tum.diam		p = 0.816		p = 0.055		p = 0.565		p = 0.579		p = 0.305		p = 0.854		p = 0.818	
<5	58(50.9)	5(4.4)	56(49.1)	14(12.3)	44(38.6)	13(11.4)	45(39.5)	19(16.7)	39(34.2)	6(5.3)	52(45.6)	11(9.7)	47(41.2)	39(34.2)	19(16.7)
≥5	40(35.1)	4(3.5)	36(31.6)	17(14.9)	23(20.2)	11(9.7)	29(25.4)	11(9.7)	29(25.4)	7(6.1)	33(28.9)	7(6.1)	33(28.9)	26(22.8)	14(12.3)
N/A	16(14.0)	2(1.8)	14(12.3)	4(3.5)	12(10.5)	4(3.5)	12(10.5)	6(5.3)	10(8.8)	3(2.6)	13(11.4)	3(2.6)	13(11.4)	8(7.0)	8(7.0)
Ap.node.exam		p = 0.814		p = 0.736		p = 0.386		p = 0.233		p = 0.338		p = 0.146		p = 0.417	
None	8(7.0)	1(0.9)	7(6.1)	2(1.8)	6(5.3)	3(2.6)	5(4.4)	4(3.5)	4(3.5)	2(1.8)	6(5.3)	3(2.6)	5(4.4)	4(3.5)	4(3.5)
≥1	101(88.6)	10(8.8)	91(79.8)	31(27.2)	70(61.4)	24(21.1)	77(67.5)	30(26.3)	71(62.3)	13(11.4)	88(77.2)	17(14.9)	84(73.7)	65(57.0)	36(31.6)
N/A	5(4.4)	0(0)	5(4.4)	2(1.8)	3(2.6)	2(1.8)	3(2.6)	2(1.8)	3(2.6)	1(0.9)	4(3.5)	1(0.9)	4(3.5)	4(3.5)	1(0.9)
EMLVI		p = 0.049*		p = 0.053		p = 0.153		p = 0.201		p = 0.697		p = 0.650		p = 0.800	
Yes	18(15.8)	4(3.5)	14(12.3)	9(7.9)	9(7.9)	7(6.1)	11(9.7)	8(7.0)	10(8.8)	2(1.8)	16(14.0)	4(3.5)	14(12.3)	12(10.5)	6(5.3)
No	96(84.2)	7(6.1)	89(78.1)	26(22.8)	70(61.4)	22(19.3)	74(64.9)	28(24.6)	68(59.6)	14(12.3)	82(71.9)	17(14.9)	79(69.3)	61(53.5)	35(30.7)
Tumour type		p = 0.054		p = 0.023*		p = 0.103		p = 0.316		p = 0.175		p = 0.625		p = 0.662	
Adenocarcinoma	105(92.1)	9(7.9)	96(84.2)	31(27.2)	74(64.9)	26(22.8)	79(69.3)	35(30.7)	70(61.4)	14(12.3)	91(79.8)	20(17.5)	85(65.8)	66(57.9)	39(34.2)
Mucinous	5(4.4)	2(1.8)	3(2.6)	4(3.5)	1(0.9)	3(2.6)	2(1.8)	1(0.9)	4(3.5)	2(1.8)	3(2.6)	1(0.9)	4(3.5)	4(3.5)	1(0.9)
Mixed	4(3.5)	0(0)	4(3.5)	0(0)	4(3.5)	0(0)	4(3.5)	0(0)	4(3.5)	0(0)	4(3.5)	0(0)	4(3.5)	3(2.6)	1(0.9)

Table 2. Lymphocyte distribution patterns and tumor bud number in relation to clinicopathological characteristics. IM = invasive margin, CT = tumor core, WTS = whole tumor section, TB = tumor bud, Max.Tum.Diam = Maximum Tumor Diameter, Ap.Node.Exam = Apical Node Exam, EMLVI = Extramural lymphovascular invasion. * Statistical significance, Chi-Square.

Features	Coefficients	
	LASSO	Mean Decrease Gini
$CD3^+$ density in WTS	**-0.0006**	**5.7932**
Mean $CD3^+CD8^+$ number within 0-50μm of TB	**-0.8951**	**5.4673**
TB Number	**0.0004**	**4.7092**
$CD8^+$ density in CT	**-0.0004**	**4.4586**
pT	0.8394	1.3724
EMLVI	-0.8623	1.1499
Age	0.5791	1.1071
Differentiation	0.4967	0.9728

Table 3. LASSO penalized Cox proportional hazard regression and Random Forest Gini coefficients for the significant features. WTS = whole tumor section, CT = tumor core, TB = tumor bud, EMLVI = Extramural lymphovascular invasion.

Figure Legends:

Figure 1. Digital image analysis method. A) Full tissue area (yellow line) and invasive front (green line) are outlined, pancytokeratin (PanCK), $CD3^+$, and $CD8^+$ cells annotated in green, yellow and red respectively; invasive margin region is highlighted in green (IM) and the tumor core region in blue (CT), **B)** detection and classification of lymphocyte cell type ($CD3^+$ cells in yellow, $CD8^+$ cells in red and their colocalization in orange mask), **C)** Full tissue area (yellow line) and Invasive front (green line) are outlined, PanCK, $CD3^+$, and $CD8^+$ cells annotated in green, yellow, and red, respectively; area of tumor bud quantification is highlighted in green, **D)** tumor to stroma segmentation and PanCK cell quantification within the tumor areas, **E)** Proximity Analysis of lymphocytes to tumor buds (TBs). $CD3^+$ cells are shown in blue, $CD8^+$ cells in orange, and TBs in gray**.** Proximity line series is shown for lymphocytes within 50 μm of TBs.

Figure 2. Spearman correlation matrix for lymphocyte densities and tumor bud number. Spearman correlation coefficient is shown for all relationships. A coefficient with either +1 (blue), 0 (white), or -1 (red) value indicates a perfect association, no association, and a perfect negative association of ranks, respectively.

Figure 3. Kaplan Meier survival analysis for pT stage, tumor bud (TB) number, immunoscore, and Tumor Bud-Immuno spatial index (TBISI) for training cohort, validation cohort 1, and validation cohort 2. (A) pT stage for training cohort, **(B)** pT stage for validation cohort 1, and **(C)** pT stage for validation cohort 2. **(D)** TB number for training cohort, **(E)** TB number for validation cohort 1, and **(F)** TB number for validation cohort 2; High represents the group of patients with TB number above the optimal cutoff point (1104.0). Low represents the group of patients with TB Number below the cutoff point. **(G)** Immunoscore for training cohort, **(H)** Immunoscore for validation cohort 1, and **(I)** Immunoscore for validation cohort 2. High Immunoscore represents the group of patients with an Immunoscore > 2, and Low Immunoscore represents the group of patients with an Immunoscore ≤ 2. **(J)** TBISI for training cohort, **(K)** TBISI for validation cohort 1, and **(L)** TBISI for validation cohort 2. High risk represents the group of patients who have $CD3^+$ density in whole tumor section and mean $CD3^+CD8^+$ cell number within 0–50 μm of TBs below the cutoff point (389.6 cells / mm^2 and 4.1, respectively) and TB number above the cutoff point (1104.0). Low risk represents all other patients.

Figure 1

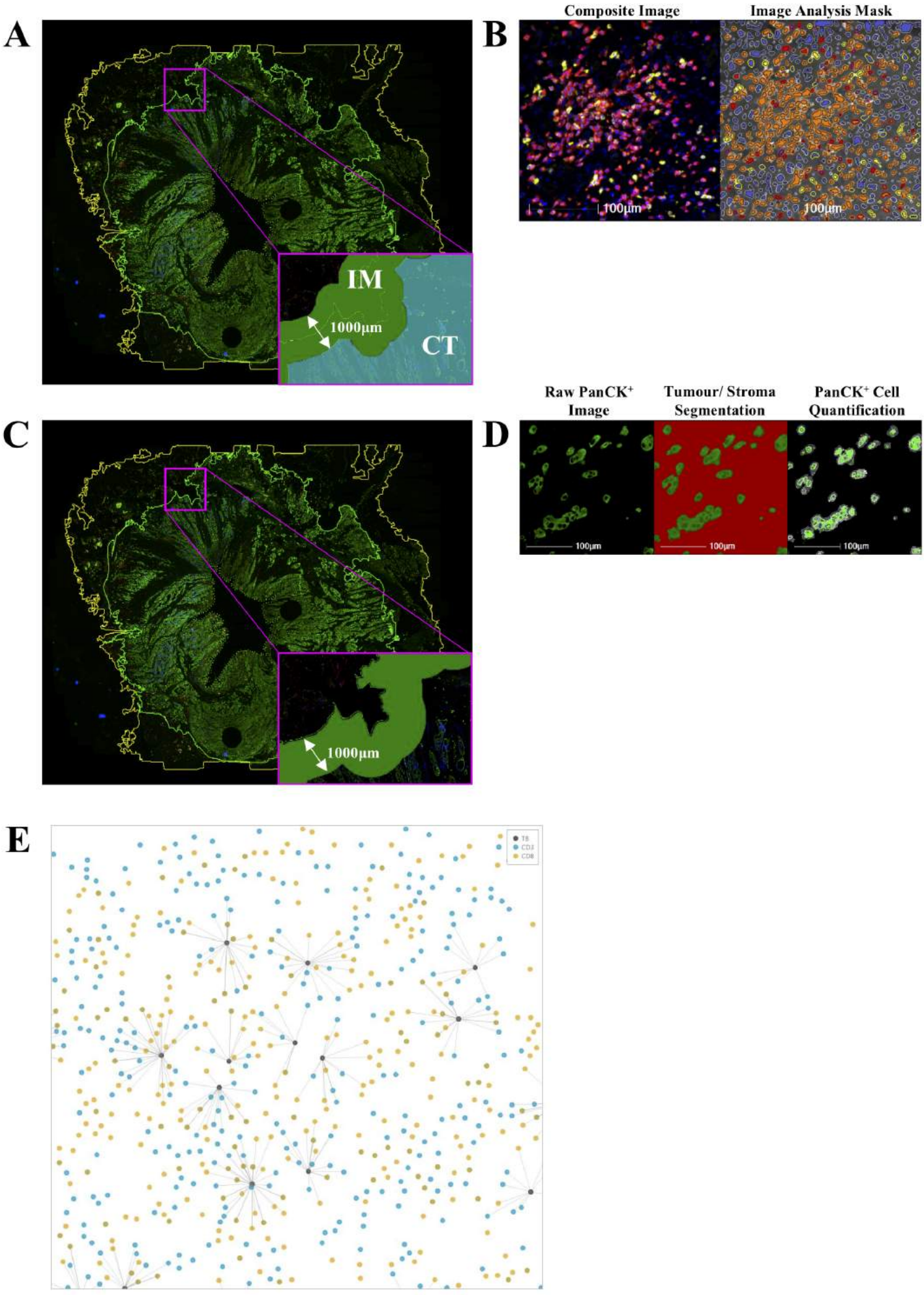
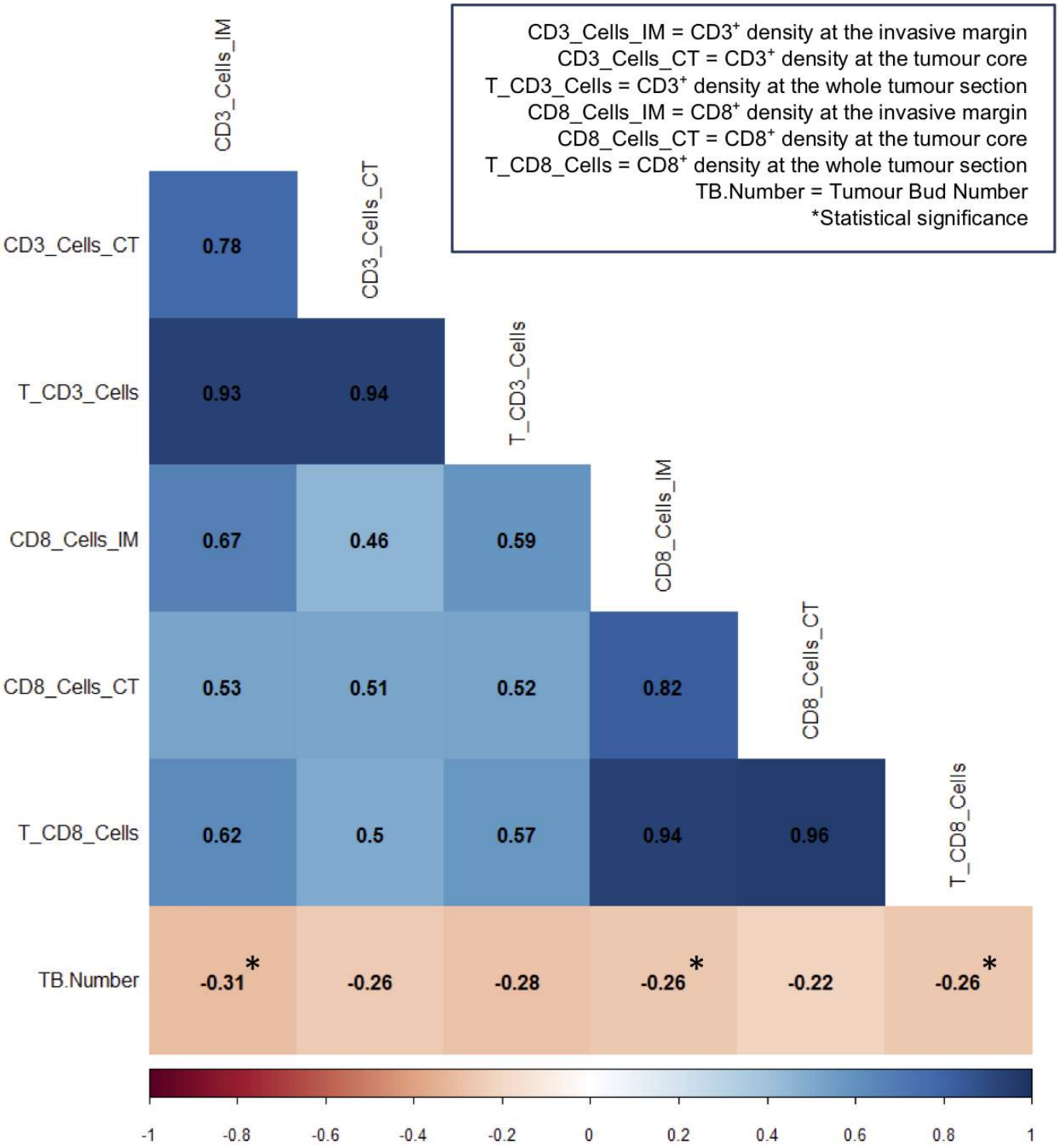
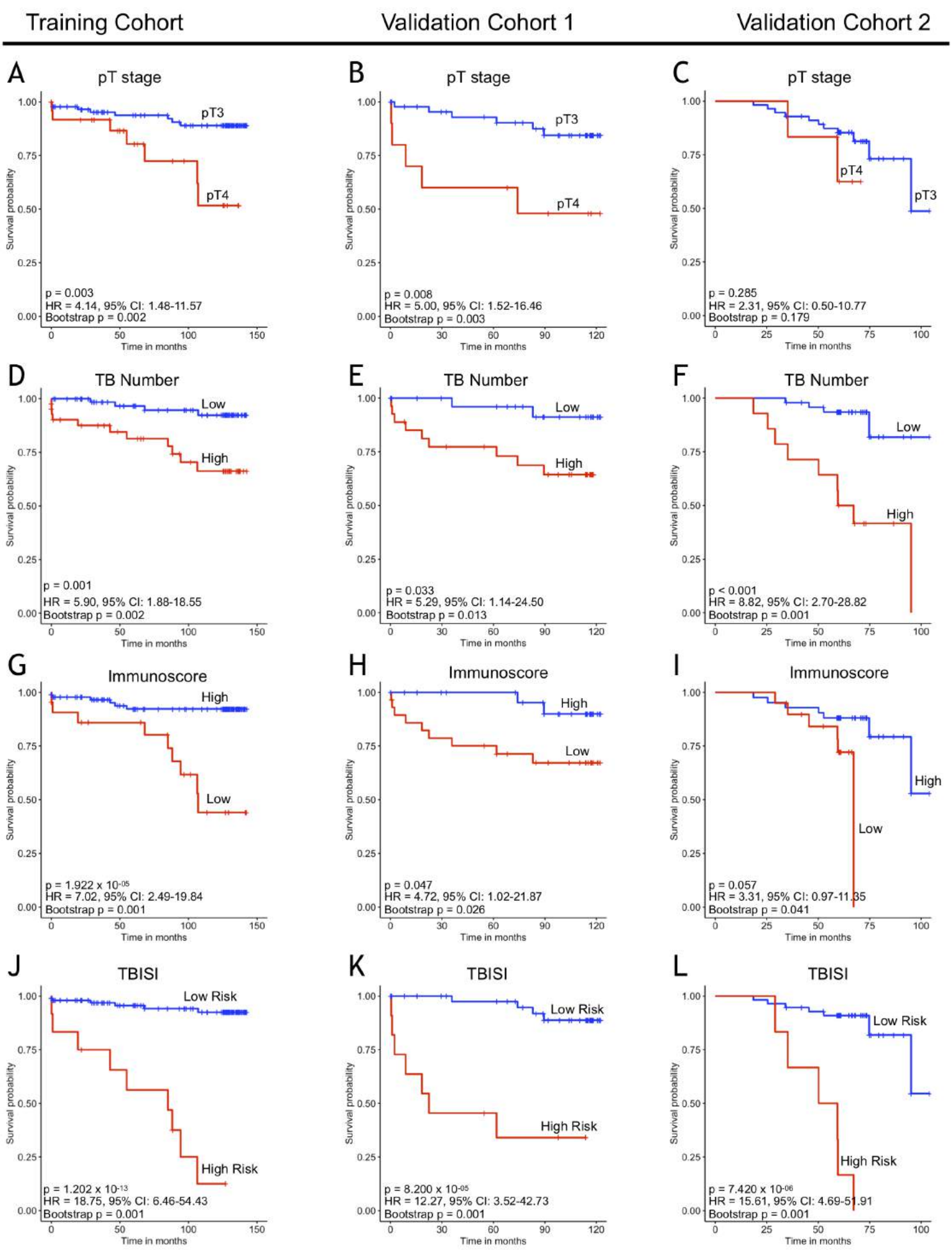

A
IM
1000µm
CT
B
Composite Image
Image Analysis Mask
100µm
100µm
C
1000µm
D
Raw PanCK+ Image
Tumour/ Stroma Segmentation
PanCK+ Cell Quantification
100µm
100µm
100µm
E
TB
CD4
CD8

Figure 2


CD3_Cells_IM = CD3+ density at the invasive margin
CD3_Cells_CT = CD3+ density at the tumour core
T_CD3_Cells = CD3+ density at the whole tumour section
CD8_Cells_IM = CD8+ density at the invasive margin
CD8_Cells_CT = CD8+ density at the tumour core
T_CD8_Cells = CD8+ density at the whole tumour section
TB.Number = Tumour Bud Number
*Statistical significance
CD3_Cells_IM
CD3_Cells_CT
T_CD3_Cells
CD8_Cells_IM
CD8_Cells_CT
T_CD8_Cells
CD3_Cells_CT
T_CD3_Cells
CD8_Cells_IM
CD8_Cells_CT
T_CD8_Cells
TB.Number
0.78
0.93
0.94
0.67
0.46
0.59
0.53
0.51
0.52
0.82
0.62
0.5
0.57
0.94
0.96
-0.31*
-0.26
-0.28
-0.26*
-0.22
-0.26*
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1


Training Cohort
Validation Cohort 1
Validation Cohort 2
A
pT stage
pT3
pT4
p = 0.003
HR = 4.14, 95% CI: 1.48-11.57
Bootstrap p = 0.002
B
pT stage
pT3
pT4
p = 0.008
HR = 5.00, 95% CI: 1.52-16.46
Bootstrap p = 0.003
C
pT stage
pT4
pT3
p = 0.285
HR = 2.31, 95% CI: 0.50-10.77
Bootstrap p = 0.179
D
TB Number
Low
High
p = 0.001
HR = 5.90, 95% CI: 1.88-18.55
Bootstrap p = 0.002
E
TB Number
Low
High
p = 0.033
HR = 5.29, 95% CI: 1.14-24.50
Bootstrap p = 0.013
F
TB Number
Low
High
p < 0.001
HR = 8.82, 95% CI: 2.70-28.82
Bootstrap p = 0.001
G
Immunoscore
High
Low
p = 1.922 x 10-05
HR = 7.02, 95% CI: 2.49-19.84
Bootstrap p = 0.001
H
Immunoscore
High
Low
p = 0.047
HR = 4.72, 95% CI: 1.02-21.87
Bootstrap p = 0.026
I
Immunoscore
High
Low
p = 0.057
HR = 3.31, 95% CI: 0.97-11.35
Bootstrap p = 0.041
J
TBISI
Low Risk
High Risk
p = 1.202 x 10-13
HR = 18.75, 95% CI: 6.46-54.43
Bootstrap p = 0.001
K
TBISI
Low Risk
High Risk
p = 8.200 x 10-05
HR = 12.27, 95% CI: 3.52-42.73
Bootstrap p = 0.001
L
TBISI
Low Risk
High Risk
p = 7.420 x 10-06
HR = 15.61, 95% CI: 4.69-51.91
Bootstrap p = 0.001
Survival probability
Time in months