



HAL
open science

Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer.

Florence Marliot, Xiaoyi Chen, Amos Kirilovsky, Thomas Sbarrato, Carine El Sissy, Luciana Batista, Marc van den Eynde, Nacilla Haicheur-Adjouri, Maria-Gabriela Anitei, Ana-Maria Musina, et al.

► To cite this version:

Florence Marliot, Xiaoyi Chen, Amos Kirilovsky, Thomas Sbarrato, Carine El Sissy, et al.. Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer.. Journal for Immunotherapy of Cancer, 2020, 8 (1), 10.1136/jitc-2019-000272 . hal-03023246

HAL Id: hal-03023246


<https://hal.sorbonne-universite.fr/hal-03023246v1>

Submitted on 25 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer

Florence Marliot,¹ Xiaoyi Chen,² Amos Kirilovsky,¹ Thomas Sbarrato ³,
Carine El Sissy,¹ Luciana Batista,⁴ Marc Van den Eynde,⁵ Nacilla Haicheur-Adjouri,⁶
Maria-Gabriela Anitei,⁷ Ana-Maria Musina,⁷ Viorel Scripcariu,⁷
Christine Lagorce-Pagès,¹ Fabienne Hermitte,³ Jérôme Galon,¹ Jacques Fieschi,³
Franck Pagès¹

To cite: Marliot F, Chen X, Kirilovsky A, et al. Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer. *Journal for ImmunoTherapy of Cancer* 2020;**8**:e000272. doi:10.1136/jitc-2019-000272

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jitc-2019-000272>).

Accepted 09 March 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Professor Franck Pagès;
franck.pages@aphp.fr

ABSTRACT

Background New and fully validated tests need to be brought into clinical practice to improve the estimation of recurrence risk in patients with colon cancer. The aim of this study was to assess the analytical performances of the Immunoscore (IS) and show its contribution to prognosis prediction.

Methods Immunohistochemical staining of CD3+ and CD8+ T cells on adjacent sections of colon cancer tissues were quantified in the core of the tumor and its invasive margin with dedicated IS modules integrated into digital pathology software. Staining intensity across samples collected between 1989 and 2016 (n=595) was measured. The accuracy of the IS workflow was established by comparing optical and automatic counts. Analytical precision of the IS was evaluated within individual tumor block on distant sections and between eligible blocks. The IS interlaboratory reproducibility (n=100) and overall assay precision were assessed (n=3). Contribution of the IS to prediction of recurrence based on clinical and molecular parameters was determined (n=538).

Results Optical and automatic counts for CD3+ or CD8+ were strongly correlated (r=0.94, p<0.001 and r=0.92, p<0.001, respectively). CD3 and CD8 staining intensities were not altered by the age of the tumor block over a period of 30 years. Neither the position of tested tissue sections within a tumor block nor the selection of the tissue blocks affected the IS. Reproducibility of the IS was not affected by multiple variables (eg, antibody lots, DAB revelation kits, immunohistochemistry automates and operators). Interassay repeatability of the IS was 100% and interlaboratory reproducibility between two testing centers was 93%. Finally, in a case series of patients with stage II–III colon cancer, the relative proportion of variance for time to recurrence was greatest for the IS (53% of prognostic variability) in a model that included IS, T-stage, microsatellite instability status and total number of lymph nodes.

Conclusion IS is a robust and validated clinical assay leveraging immune scoring to predict recurrence risk of patient with localized colon cancer. The strong and independent prognostic value of IS should pave the way for its use in clinical practice.

INTRODUCTION

Colorectal cancer represents one of the most commonly diagnosed cancers worldwide.¹ Estimation of prognosis and decision on the treatment of patients are based on features of tumor extension (the Union for International Cancer Control (UICC) tumor-node-metastasis (TNM) classification) and tumor-cells differentiation.^{2,3} However, clinical outcome can significantly vary among patients within the same tumor stage and the TNM classification alone does not predict response to therapy.⁴ Given these limitations, much effort has been devoted to the identification of tumor cell-based biomarkers, mutational status, molecular pathway and tumor gene expression-based stratification approaches to further refine risk estimation.^{4–7} However, only a few of these factors have been validated as robust independent prognostic markers for care of patients with colon cancer.^{5,6} Novel insights into the molecular biology of colon cancer have enabled new therapeutics to be developed, but new biomarkers of high prognostic or predictive value are still needed to select the most optimal adjuvant management strategies.⁸

There is evidence that anticancer immune responses contribute to the control of tumor growth and metastasis.^{9–11} The cytotoxic adaptive immune response has been almost consistently associated with a favorable prognostic in solid tumors.¹² We previously showed that in situ adaptive immune reaction in the tumor core (CT) and its invasive margin (IM) strongly correlates with time to recurrence (TTR), disease-free survival (DFS) and overall survival (OS),^{13–15} regardless of cancer stage.¹³ Digital pathology leveraging automated image

analysis has been suggested as a potential standardization method to improve the accuracy and reproducibility of the immune cells quantification.^{16 17} Consequently, we developed a prognostic immune scoring system, termed Immunoscore (IS), that quantitates both CD3+ lymphocytes and CD8+ cytotoxic T cells in the CT and IM. The prognostic performance of IS has been validated through an international pivotal validation study of 3539 patients with stage I–III colon cancer.¹⁸

However, additional sources of variability related to sample preparation or immunostaining technique must also be controlled, that is, lot-to-lot variability of critical reagents such as primary antibodies, the choice of tumor block among those potentially eligible, the reproducibility of the sectioning process or external factors such as age of the paraffin block and storage conditions.^{19 20} The demonstration of the analytical performance characteristics (IS robustness according to the parameters mentioned above and an intralaboratories and interlaboratories reproducibility) are necessary for implementation of the IS into clinical practice.

In this study, we describe the development of the IS, validate its analytical performance and examine its prognostic value compared with tumor biomarkers.

MATERIAL AND METHODS

Patients and specimens

The prognostic value of the IS was assessed in the French cohort of patients with stages I–III colon cancer included in the international validation study of the IS between 1989 and 2009 (n=538)¹⁸ and in a French cohort of patients with stage I–III colon cancer from 2 studies: the Immunoscore as a prognostic marker for patients with a colorectal cancer (ImmuCol) and personalizing medicine of colorectal cancer (ImmuCol2), between 2012 and 2016 (n=57; NCT02274753, NCT01688232). Samples were excluded if the patient did not meet the inclusion criteria (clinical quality control: rectal cancer, Tis/Tx tumor, metastatic cancer, neoadjuvant chemotherapy, missing mortality or recurrence status).¹⁸ Histopathological and clinical findings of these datasets were scored according to the eighth UICC-TNM staging system.²¹ Formalin-fixed paraffin-embedded (FFPE) surgical tissues from the colon cancers were collected and used for the IS analyses.

A second cohort of patients with stage I–III colon cancer (n=100) treated in the Department of Surgical Oncology of the Regional Institute of Oncology (Iasi, Romania) was used to assess the between-lab agreement for the immune cell densities and IS levels.

The analytical precision elements of this study were determined using 13 anonymous FFPE colon cancer blocks (Indivumed, Hamburg, Germany and ProteoGenex Culver City, California, USA).

Clinical outcomes

The primary outcome was TTR defined as time from surgery to disease recurrence. Secondary outcomes were

DFS defined as the time from surgery to the first observation of disease recurrent or death due to any cause, and OS defined as time from surgery to death due to any cause.

Immunohistochemistry

FFPE surgical tissue specimens were used for the immunohistochemical analyses. H&E staining confirmed the presence of both adenocarcinoma tissue and normal tissue and the evidence of immunity cells infiltration in the CT and its IM. Two adjacent tissue 4 µm sections were cut from the FFPE tissue blocks and processed for immunohistochemistry. Primary antibodies were directed against CD3+ (2GV6, Ventana, Tuscon, Arizona, USA) and CD8+ (C8/144B, Dako, Glostrup, Denmark) and were visualized by Ultraview Universal DAB IHC Detection Kit (Ventana) in center 1 (Laboratory of Immunology, AP-HP, European Hospital Georges Pompidou, Paris, France). In center 2 (HaliDx, Marseille, France), sections were immunostained with antihuman CD3+ (clone HDx2, HaliDx) and antihuman CD8+ (clone HDx1, HaliDx) monoclonal antibodies revealed according to the CE-IVD Immunoscore Kit Instructions (HaliDx).

Digitization of immunostained sections and quantification of immune cells

In center 1, slides were digitized on the NanoZoomer HT scanner (Hamamatsu Photonics, Japan) with the ×20 magnification and 0.45 µm/pixel resolution. A customized IS module integrated into Developer XD digital pathology software (Definiens, Munich, Germany) was used to quantify CD3+ and CD8+ T cells in the CT and IM, divided in tiles (720 µm per side). In center 2, the NanoZoomer XR scanner (Hamamatsu Photonics) was used, with the identical setting. An upgraded version of the IS module (Immunoscore Analyzer, HaliDx) was used to quantify CD3+ and CD8+ cells. In both centers, the mean and distribution of the intensity of stained cells were used as internal quality controls of each immunostaining. Automatic tissue detection (tumor, healthy non-epithelial tissue and epithelium) and IM defined as a region of 720 µm width; 360 µm on each side of the border between malignant cells and peritumoral stroma were generated automatically. If required, automated detection of the CT region was manually corrected and validated by a pathologist.

Immunoscore determination

For each case, CD3+ and CD8+ cell densities in the CT and its IM were compared with those obtained in the training and internal validation sets of the international validation study cohort¹⁸ and converted into percentiles. The mean of the 4 percentiles obtained for CD3+ and CD8+, either in the CT or IM, was then calculated and converted into an IS. The IS was categorized into three groups, with a mean percentile of 0%–25% (IS0–1), >25%–70% (IS2) and >70%–100% (IS3–4). IS were further categorized into IS Low (IS0-1) and IS High (IS2-3-4) (figure 1G).

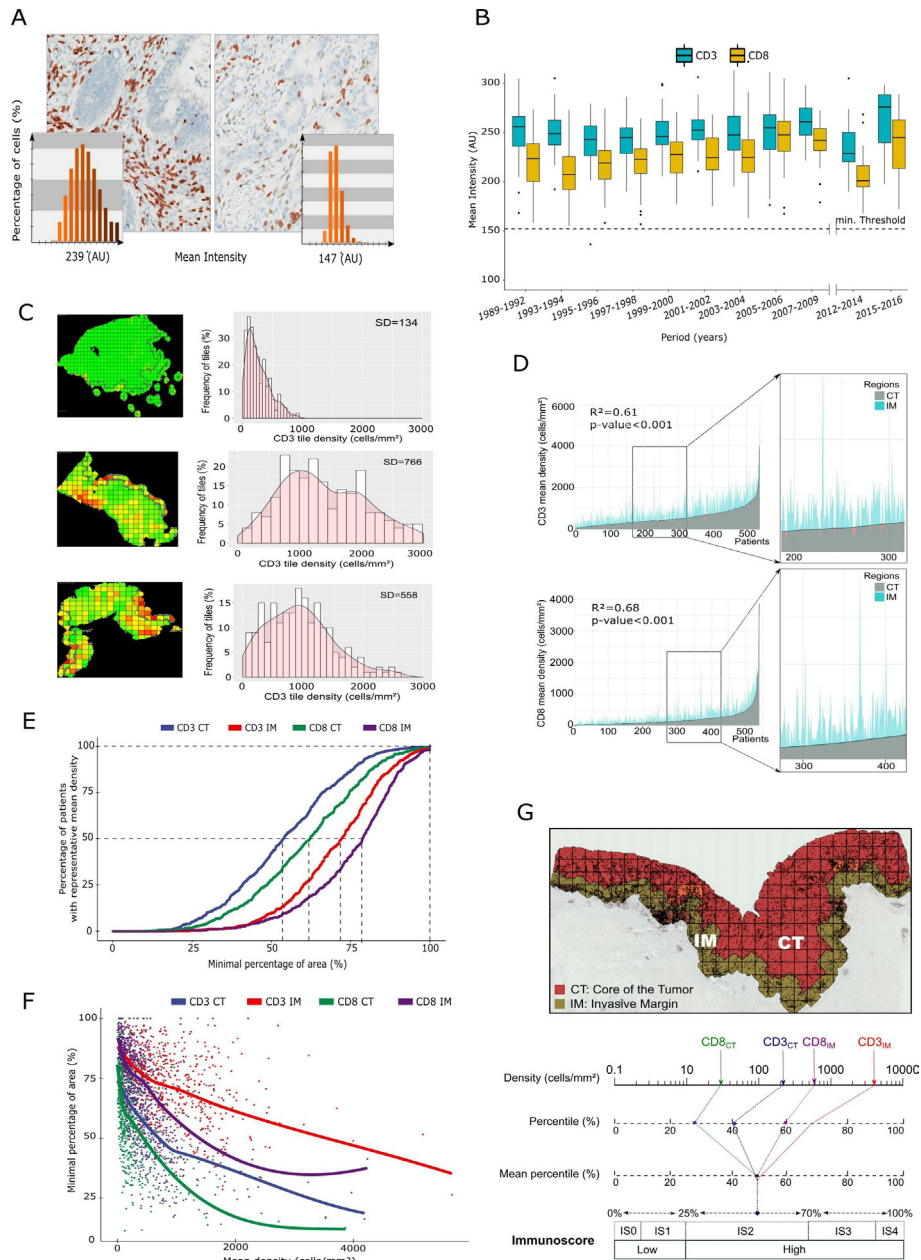


Figure 1 Immunoscoring (IS) determination. (A) Representative digital pathology image of strong CD3+ T cells immunostaining (left). Histogram of CD3+ cells intensities detected by the software (mean intensity of 239 AU; bottom left). Example of weak CD3+ immunostaining, which leads to underestimation of cell counts by the software as shown by the absence of red outline around some immunostained cells (right). Histogram of the CD3+ cells intensities detected by the software (mean intensity of 147 AU; bottom right). (B) The effect of sample storage time on immunostaining performance. Box plots showing the IQR of the CD3+ and CD8+ staining intensities in colon cancer samples (n=595) according to the age of tumor blocks (between 1989 and 2016). The dashed line represents the 152 AU mean intensity threshold. (C) Heterogeneity of CD3+ T-cell infiltration in three colon cancer tumor samples. The tumor core (CT) and invasive margin (IM) regions are divided into tiles with the density of CD3+ T cells in each tile figured from green (lowest density) to red (highest density) (left). Unimodal distribution of CD3+ T cells with a weak and homogeneous immune infiltration (right, upper graph). Bimodal distribution with a heterogeneous immune T-cells infiltration (middle graph). Multimodal distribution with an intermediate and heterogeneous immune cell infiltration (lower graph). SDs of the immune cells densities in tiles are indicated. (D) Comparison of the CD3+ or CD8+ T cells mean densities in the CT (gray) and in IM (blue) for each patient of the cohort (n=538). Patients are ordered by an increasing immune cells density in the CT region. Spearman's correlations with R^2 for CT vs IM are provided for each marker. (E) Minimum area (%) required for each marker (CD3+ and CD8+), in each region (CT and IM), which allows to estimate immune cells density equal to the whole region ($\pm 10\%$) in all patients (n=538). (F) Impact of immune cells density (CD3+ and CD8+) in each region of interest (CT and IM) on the minimum surface necessary to estimate immune cell density equal to the whole region ($\pm 10\%$). (G) Calculation of the clinical assay IS groupings. Chart illustrating the IS calculation method. Densities of CD3+ CT, CD3+ IM, CD8+ CT and CD8+ IM converted into percentile values, determined by the international validation study.¹⁸ The mean percentile of the four markers is calculated and translated into a five-category (IS0, IS1, IS2, IS3, IS4) or a two-category scoring system IS Low and IS High.

All analyses were performed in center 1 except experiments reported in [figure 2B,E](#) and supplementary tables, performed in center 2.

Statistics

The hierarchical clustering method was used to explore the relation between clinical (age, perforation), pathological (T stage, N stage, vascular, lymphatic or perinervous tumor emboli, number of lymph nodes) and molecular characteristics (*KRAS*, *APC*, *P53*, *PI3K*, *BRAF*, microsatellite instability (MSI) status) of the tumors and the IS.²² An age threshold of 70 years and a lymph nodes threshold of 12 were used. Data of each patient were scaled to mean 0 and an SD of 1. Manhattan distance was used to compute the dissimilarity for hierarchical clustering.

Based on the Clinical and Laboratory Standards Institute EP05-A3 guidelines, the analytical precision statistics were compared using analysis of variance (ANOVA) for the repeatability and an ANOVA test (Variance Component Analysis model) for the assay intermediate precision were used to assess the percentage contribution of source of variation (for each component and measurement) to the total variability. Similarly, intrablock repeatability of IS on all 10 samples was assessed using a mixed model (REML mixed model from R VCA package) in order to assess the variation expected in results when considering intrablock as part of the model. A bandwidth of $1.96 \times \text{error}$ was then calculated.

The relative proportion of explained variance (Cox & Snell pseudo R^2) in TTR, which was accounted by different categories of predictor covariates, was calculated. The full multivariate model was initiated and each time one predictor was removed from the model, the differences of pseudo R^2 were computed, then the proportion of each predictor covariate was obtained. Covariates were sequentially removed from the model by decreasing importance. For this analysis, molecular markers with missing values were excluded and the multivariate Cox model was built with eight parameters used to describe the severity of colon cancer. Figures and statistical analyses were performed using the R software (V.3.3.1). To assess the agreement between automatic and manual T-cells counts, intraclass correlation coefficient was calculated using the psych-package. The Bland-Altman plots were generated with their limits of agreement, defined as the mean of the differences ± 1.96 SD above and below the mean difference (known as the bias). Pearson's and Spearman's correlations were applied to calculate the r and R^2 , respectively. The log-rank T-test and log-rank test for trend were applied to calculate p values. The p values lower than 0.05 were considered statistically significant. The ggplot2 package and survival package were used to model the hierarchical clustering and calculate the relative proportion of explained variance (Cox & Snell pseudo R^2).

Kaplan-Meier method estimates of survival were used to illustrate the survival curves, and survival was measured

in months from the resection to either recurrence or the last review.

RESULTS

Analytical validation of the IS

To monitor the variability of the CD3+ and CD8+ T cells immunostaining intensity across samples, an internal control was established to ensure the validity of the counting by the software. Illustrations of an adequate staining intensity (mean intensity of 239 AU) for CD3+ T cells allowing a valid count of cells and a faint staining (mean intensity of 147 AU) leading to an underestimation of the number of stained cells are shown in [figure 1A](#). A minimum value of 152 AU for the mean intensity of all stained cells detected by the software, ensured a detection of at least 70% of the stained cells,¹⁸ and was defined as acceptance criterion.

The influence of the tumor block age on the intensity of staining was tested on tissue samples from a retrospective monocentric cohort of 595 patients with colon cancer treated between 1989 and 2016. Fixation time, concentration of the formalin and storage environment remained constant throughout this period of time. Sample stability for CD3 and CD8 detection was maintained over a period of 30 years as shown by the intensity of stained cells above a given intensity threshold ([figure 1B](#)).

Accuracy of an automated method for counting stained CD3+ and CD8+ T cells

The accuracy of the automated IS workflow was established by counting CD3+ and CD8+ T cells manually or with a dedicated software program in 50 tiles randomly selected from immunostaining (online supplementary figure S1). Optical and automatic counts of stained cells for CD3+ or CD8+ in all the tiles (ranging between 50 and 2000 cells per tile) were strongly correlated ($r=0.94$, $p<0.001$ and $r=0.92$, $p<0.001$, respectively). This result was confirmed with intraclass correlation coefficient of 0.90 for CD3+ (95% CI 0.81 to 0.94) and of 0.91 for CD8+ (95% CI 0.85 to 0.95). The Bland-Altman plots showed a strong agreement between the two counting methods. Differences observed above 800 cells/tile of 720 μm side were probably due to the difficulty of optical estimation at very high cell densities (online supplementary figure S1).

Assessment of the CD3+ and CD8+ T-cells distribution across tumors and definition of the IS

The distribution of CD3+ and CD8+ T cells in the CT and its IM was also assessed in tumor samples from the cohort of 538 patients with stage I–III colon cancer tumors. Unimodal, bimodal or multimodal distribution of immune cells densities were observed, as illustrated for the CD3+ densities ([figure 1C](#)). An average SD of 505 cells/ mm^2 (from 7 cells/ mm^2 to 1781 cells/ mm^2) for CD3+ T cells and 199 cells/ mm^2 (7 cells/ mm^2 to 1257 cells/ mm^2) for CD8+ T cells across the tiles defined in all tumor sections illustrated immune infiltrate heterogeneity (data

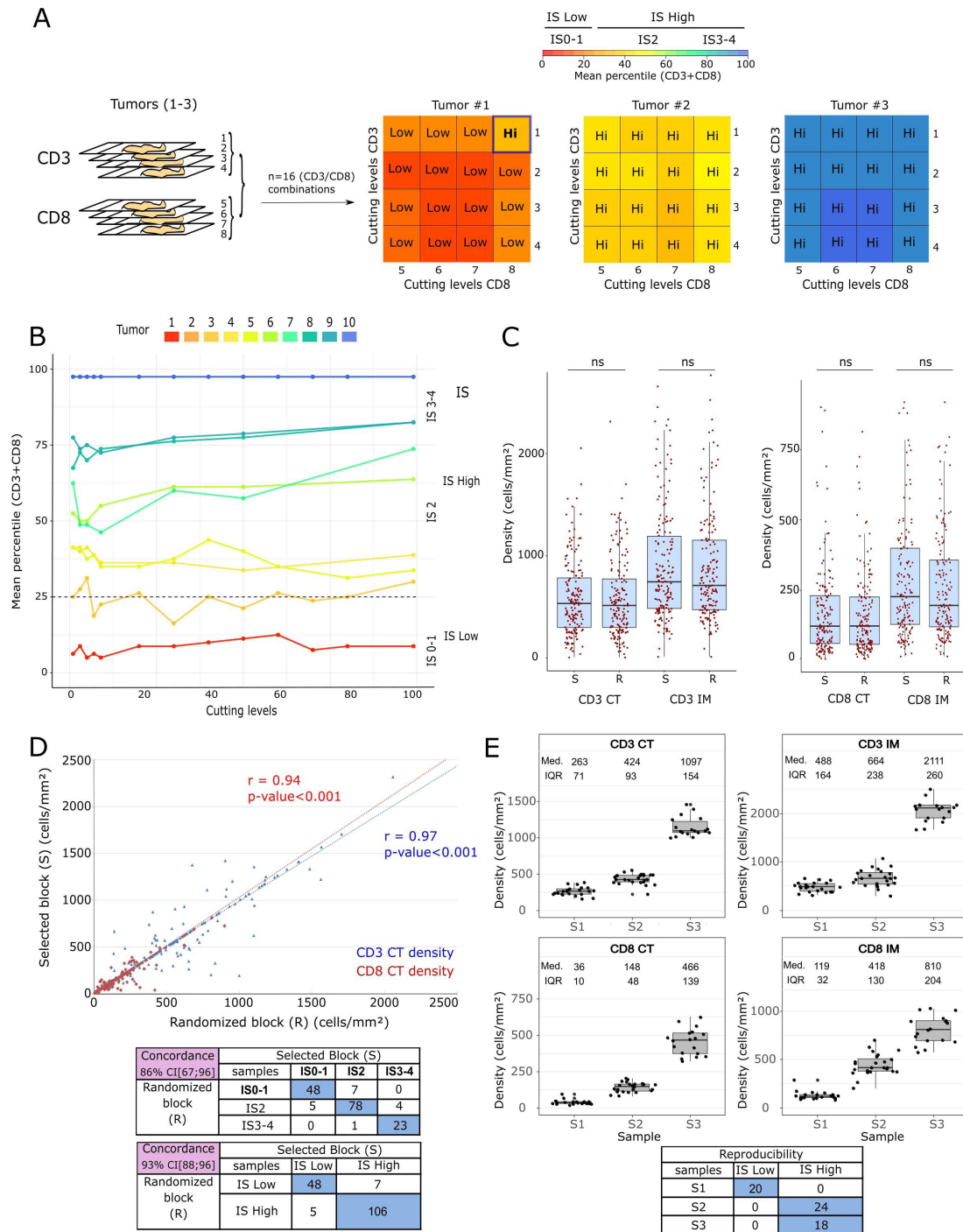


Figure 2 Analytical performance of the Immunoscore (IS). (A) Sample repeatability: assay stability in serial tumor sections. Impact of the selected sections on IS on three colon cancer samples according to the cutting level of the tumor block. For each tumor, four adjacent sections were immunostained for CD3+ followed by four adjacent sections immunostained for CD8+. All CD3+/CD8+ combination slides were tested per tumor. The concordance matrix shows the mean percentile (from red to blue; 0–100) of each combination (CD3+/CD8+) and the associated IS category. (B) Mean percentile (CD3+ and CD8+) and the IS variations according to the cutting level of the tumor block from top to bottom. A total of 13 levels (1, 3, 5, 7, 9, 20, 30, 40, 50, 60, 70, 80, 99) of cutting from 100 adjacent cuts of each tumor block (n=10) were investigated for IS. (C) Intermediate precision: impact of tumor blocks selection on the IS. Box plots show the IQR of CD3+ and CD8+ immune densities from the tumor block selected by the pathologist (S) and from a random block (R) for each patient of the cohort (n=166). Paired t-tests were calculated to assess differences between groups. (D) Pearson's correlation with r between densities in the randomized block and selected block for the CD3+ CT (blue) and CD8+ CT (red) (top). Contingency tables show the IS categories obtained for each case with the selected and random block (bottom). (E) IS assay precision and lot-to-lot reproducibility. Tumor cut from three colon cancers were assessed for CD3+ and CD8+ T cells densities using three different antibody lots, three DAB revelation kit lots, two Benchmark autostainers, three different runs and three operators. Contingency tables showing the IS classification concordance for each sample (bottom). CT, core of the tumor; IM, invasive margin.

not shown). In addition, for each marker, immune cells densities in the CT and its IM were only weakly correlated (figure 1D), where patients (n=538) were ordered by increasing cell density of CD3+ T cells in the CT (R^2 for CT vs IM of 0.61 for CD3+ and of 0.68 for CD8+, $p<0.001$).

Investigation of 55%–78% of the total area of each region was necessary to obtain a reliable estimation of CD3+ or CD8+ mean densities for 50% of the patient cohort (figure 1E). These percentages of total area presented a surface of 40.9 and 48.3 mm² in CT and 14.7 and 15.4 mm² in IM for CD3 and CD8 markers, respectively (online supplementary figure S2). The minimum area to be analyzed was larger when the immune infiltration for the marker of interest was low (figure 1F) and this effect was more pronounced in the CT than in the IM region.

Based on these results, the CD3+ and CD8+ immune T-cells densities were calculated on the entire regions (CT, IM) in order to build the IS scoring system (figure 1G). Densities of CD3+ and CD8+ cells in the CT and IM regions were converted into percentiles values according to a reference cohort of 700 patients with colon cancer from 13 countries, provided by the Society for Immunotherapy of Cancer validation study.¹⁸ For each patient, the mean of the four percentiles (CD3+ CT, CD3+ IM, CD8+ CT, CD8+ IM) was translated into five categories, which were further collapsed into two groups as described in the ‘Materials and methods’ section.

Intratumor and intertumor tissue block precision of the IS

Repeatability of the IS was explored between adjacent slides from three tumor tissue blocks with low or high immune infiltration (figure 2A). For each block, adjacent sections were stained for CD3+ T cells (four slides) and CD8+ T cells (four slides), which allowed for 16 CD3+/CD8+ IS combinations. Variation of the mean percentile for the 16 combinations was low, with an SD of 5.4%, 4.3% and 0.91% per tumor tissue block (data not shown). Overall concordance of the IS testing between the tumor tissue blocks was 98%. A shift from one to another IS category was not observed for any of the CD3+/CD8+ combinations for the high IS blocks. Only a one level IS change among 16 combinations was observed for tumor #1 (IS Low), whose mean percentile was close to the 25% cut-off between low and High IS.

Since two adjacent sections of 4 μm represented <1% of the total thickness of a typical tumor block, it was therefore important to ensure that the IS testing was robust to tumor heterogeneity. We examined the inter-assay repeatability by testing distant sections from the same tissue blocks. The experiment was conducted on 10 different tumors covering the analytical range of the IS and was inclusive of tissues near the IS cut-off. The IS was evaluated on 13 levels distributed from cut 1 to cut 100 (adjacent slides from cut levels 1, 3, 5, 7, 9, 20, 30, 40, 50, 60, 70, 80, 99) from each tissue block (figure 2B). A low variability of the mean percentile (CD3+ and CD8+) was observed across the sections for all tumors. The global

intrablock repeatability SD for the 10 samples studied on the set of 13 levels was of ±8 percentiles. Overall, precision remained high, with IS group concordance at 95%. Shifts from one to another IS category were observed for blocks with mean percentile very close to the cut point between two categories.

In clinical practice, there are sometimes multiple tumor blocks available for the same patient tumor(s). The differences in immune cells densities and IS between a block selected by a pathologist and a block randomly selected among all eligible blocks was determined for 166 patients with colon cancer included in the international validation cohort (figure 2C). No significant differences between tumor blocks (random vs selected) were observed neither immune marker nor tumor regions (CT; IM) as assessed by paired t-tests. The intertumor tissue block correlation based on Pearson’s correlation coefficient (r) was 0.94 for CD3+, and 0.97 for CD8+ T cells ($p<0.001$). The inter-assay repeatability between selected and randomized tumor tissue blocks was 93% (figure 2D).

Robustness of the IS

Further intermediate precision, lot-to-lot reproducibility and intra-assay repeatability were assessed in six immunostaining runs with two Benchmark XT autostainers, three CD3+ and three CD8+ antibody lots, three DAB revelation kit lots and three operators. The IQRs of CD3+ and CD8+ immune densities in the CT and IM are illustrated in figure 2E. The intra-assay repeatability was assessed for each sample from duplicates (two adjacent slides). For each duplicate, the difference of densities and mean percentile was calculated and is reported in online supplementary table S1. Overall, the SD of repeatability (S_r) was 90.66 cells/mm² for the densities and 3.92% for the mean percentile, thus highlighting the high level of intra-assay repeatability in the IS readout.

To resolve the total variability of the IS workflow into different components, the general and lot-to-lot reproducibility studies were performed and analyzed with a mixed effects ANOVA-Variance Component Analysis model on mean percentiles (online supplementary table S2). None of the individual components significantly contributed to IS variability. Furthermore, the categorical IS concordance was 100% among the three tested tumors (figure 2E).

Impact of interlaboratory reproducibility on the IS

The assessment of IS reproducibility was performed in two laboratories; each laboratory had its own IS workflow including staining, scanning and analysis. Non-consecutive cutting levels from the same tumor block were used to assess the IS of 100 cases from Iași Hospital. The interlaboratory mean percentiles (CD3+ and CD8+) correlation was 0.94 ($p<0.001$) and the categorical IS concordance between the two centers was 93% (online supplementary figure S3).

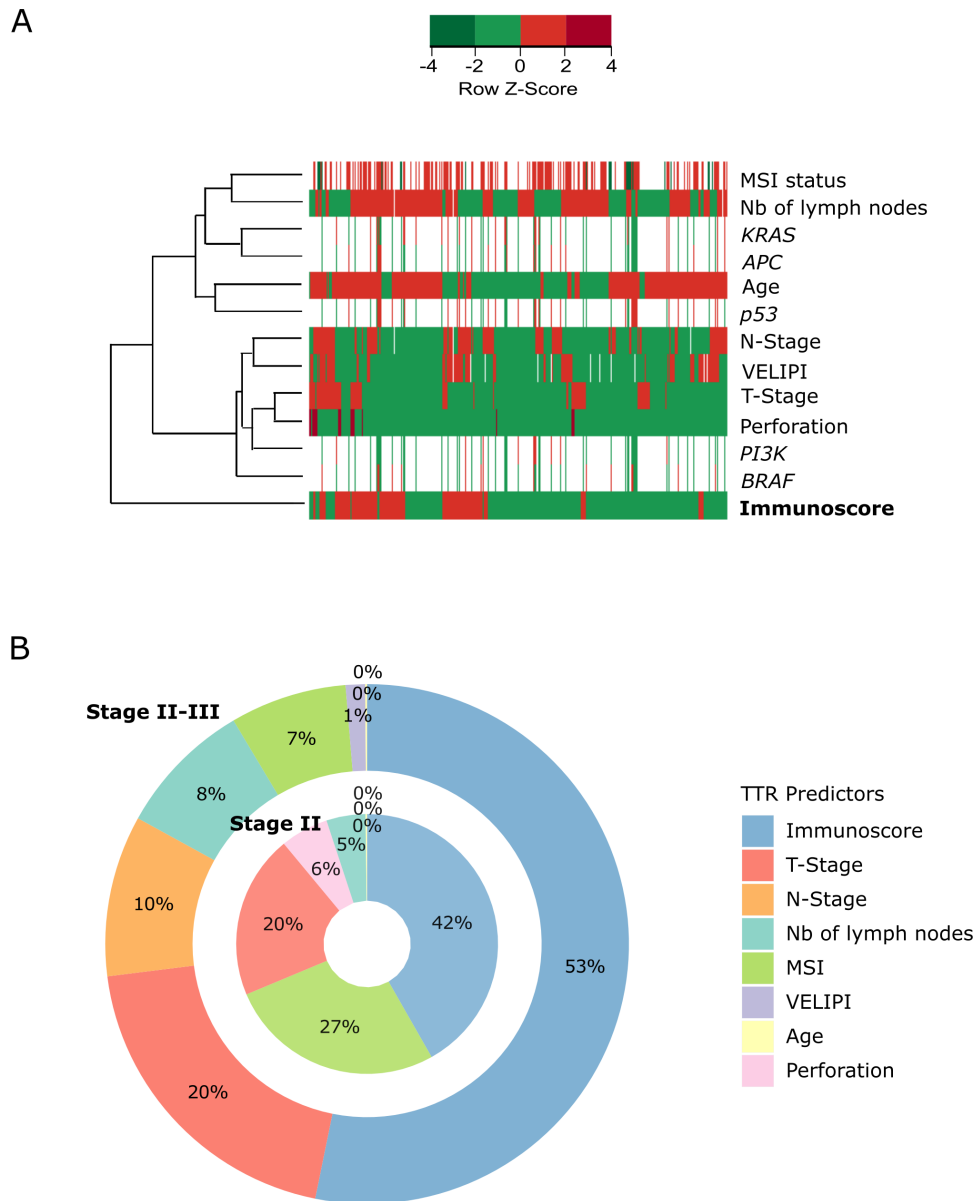


Figure 3 Relative prognostic contribution of clinicopathological and molecular factors. (A) Unsupervised hierarchical clustering testing the similarity between Immunoscore (IS) and clinical and biological parameters of severity ($n=12$) in patients with stage I–III colon cancer ($n=538$). Each parameter is in ordinal value and the normalized z-scores are shown in the heat map. Beneficial (green) and adverse (red) parameters are represented. Missing values are indicated (white). Manhattan distance was used to compute the similarity for the hierarchical clustering. (B) Ring charts illustrating the relative proportion of explained variance (Cox & Snell pseudo R^2) of each risk parameter to recurrence risk in patients with stage II–III ($n=229$) and stage II colon cancer ($n=141$). TTR, time to recurrence.

Prognostic validation of the IS

The singularity of the IS with respect to the clinical and molecular parameters of severity commonly used in clinical practice was first tested on the cohort of 538 patients with stage I–III colon cancer (figure 3A). Hierarchical clustering of the covariate relationships (Manhattan distance) showed that MSI status and the number of tumor-invaded lymph nodes were clustered together. *KRAS*, *APC* and *p53* were clustered together away from *PI3K* and *BRAF* genes. Interestingly, the IS was dissimilar to the other clusters in the hierarchical clustering. To examine the relative contribution of the IS to patient

recurrence risk, ring charts were generated in order to illustrate the relative proportion of explained variance of the IS with the other risk parameter to recurrence (TTR predictor) in the cohort of 229 patients with stage II–III colon cancer (figure 3B). The IS outperformed other clinical parameters of severity such as T stage, N stage and MSI status. The relative proportion of variance for TTR explained by IS, T stage, N stage and MSI status was 53%, 20%, 10%, 8% and 7%, respectively for patients with stage II–III colon cancer ($n=229$). In patients with localized stages II colon cancer ($n=141$), this observation persisted. Significant difference of TTR among patient's

groups, stratified by IS were observed (online supplementary figure S4A,B). HR forest plots for TTR, DFS and OS according to the IS in patients with stage II (n=292) and stage II–III (n=448) colon cancer are provided in online supplementary figure S4C. In patients with stage II or stage II–III colon cancer, an IS High was associated with the lowest risk of relapse (TTR; all log-rank tests $p < 0.05$). For example, 5 years TTR was 87.9%, 77.1% and 57.0% for IS3–4, IS2 and IS0–1, respectively in patients with stage II–III colon cancer.

DISCUSSION

Polymorphic and variable immune infiltrates composed of almost all types of immune cells are observed between solid tumors. The different immune cell subsets are associated with variable prognostic significance, sometimes with opposite direction.¹² The in situ immune ‘picture’ in the tumor is therefore complex. An integrative analysis of immune populations infiltrating colon cancer has recently shown the major influence of particular subsets, for example, Th1-oriented immune cells and cytotoxic T lymphocytes for long-term survival of patients.^{13 15} The positive impact of the tumor infiltrating lymphocytes on survival of patients with colorectal cancer, independent of criteria of tumor extension, was already observed in 1986 by Jass.²³ Thirty years after, a comprehensive immune assessment has not yet been integrated into our clinical practice probably due to the interobserver variability between pathologists and a lack of standardization in methodology.^{17 24}

We have recently implemented the consensus IS as a test to predict clinical outcome in patients with stage I–III colon cancer.¹⁸ The procedures and strategies of quantification of the IS (ie, CD3+ and CD8+ quantification in the CT and its IM) were defined by an international consortium.¹⁸ CD3+ and CD8+ markers were selected based on their associated staining quality, and respective non-redundant prognostic contributions.^{13 15 25} It was previously shown that the combined analysis of tumor regions (CT plus IM) improves the accuracy of prediction of survival for the different patient groups compared with single-region analysis.¹⁵ We have refined those observations here by demonstrating that the cell densities must be assessed on whole regions (CT and its IM) to more accurately measure the mean density of the total region of interest, as well as ensure reproducibility between laboratories. Therefore, the IS test is performed on whole regions (CT and IM). Several the FFPE tumor tissue blocks are often eligible for evaluation of the IS and a wide range of cutting levels can also be achieved within the same block. This raises the question whether the IS is impacted by tumor heterogeneity and FFPE block bias. Here, we showed that the CD3+ and CD8+ cell densities and the average percentiles remain remarkably constant across a wide range of cutting levels within the same block or between blocks. Moreover, the IS categorization into two classes further increases the concordance between

multiple measurements. Overall, the IS intra-assay and interassay agreement was 95% and 90%, respectively. In patients close to the cut point, results can be provided as IS category and also as mean percentile. A graphic showing for each mean percentile (ie, IS as a continuous variable) the survival prediction at 3 and 5 years could ultimately help clinician to have the most accurate prognostic evaluation to guide therapeutic decision.

None of the technical components of the IS assay, including antibodies, revelation DAB kit lots, instruments and operators significantly contributed to variability in CD3 and CD8 immunostaining nor IS classification, consistent with previous analyses. We had already observed a very good homogeneity of the IS categorization assessed by different observers evaluating the same CD3+ and CD8+ immunostainings.¹⁸ IS reproducibility was determined by comparison of 36 images of colon cancer sections stained for CD3+ and CD8+, which were re-analyzed by eight pathologists from different centers, and this procedure demonstrated a small variation (2.1%) in the mean percentile of CD3+ and CD8+ T-cell densities between observers.¹⁸ Comparison of the IS categorization on non-consecutive cutting levels in terms of homogeneity in our study showed again very high agreement (93%) between two centers performing the staining, scanning and analysis with their own instrument workflows. In a separate validation analysis, the IS emerged as an independent prognostic parameter in our study. As such, it can provide new information on the host’s defense against the tumor, which is the essential puzzle piece into immunotherapy success. This observation is in line with the recent results of the international validation study of the IS in 3539 patients with stage I–III colon cancer.¹⁸

CONCLUSION

This study details the analytical performance characteristics of the IS, measuring the immune response to cancer and improving the estimation of risk of recurrence for patients with colon cancer. The analyses showed that the IS scoring system is a robust, reproducible, quantitative and standardized immune assay, with a high prognostic performance, independent of the prognostic markers currently used in clinical practice. This paves the way for the use of the IS test in clinical practice in colonic tumors and probably in the near future in other solid tumors.

Author affiliations

¹Laboratory of Integrative Immunology and cancerology, INSERM, University of Paris, Cordeliers Research centre, Immunomonitoring Platform, Laboratory of Immunology, AP-HP (Assistance Publique-Hôpitaux de Paris) Hôpital Européen Georges Pompidou, Paris, France

²Laboratory of Information Sciences to Support Personalized Medicine, Cordeliers Research Centre, Paris, France

³HaliDx, Marseille, France

⁴Innate Pharma SA, Marseille, France

⁵Department of clinical and medical oncology, Cliniques universitaires Saint-Luc Institut Roi Albert II, Bruxelles, Belgium

⁶Immunomonitoring Platform, Laboratory of Immunology, AP-HP (Assistance Publique-Hôpitaux de Paris) Hôpital Européen Georges Pompidou, Paris, France
⁷Department of Surgery, Grigore T Popa University of Medicine and Pharmacy Faculty of Medicine, Iasi, Romania

Acknowledgements The authors would like to thank Magdalena Benetkiewicz (PhD) for reviewing/editing assistance.

Contributors FP, JG, JF had full access to the study data and take responsibility for the integrity of the data and accuracy of the data analysis. Study concept and design: FP, JG, FH, JF. Statistical analysis: AK, XC, LB, TS. Acquisition, analysis or interpretation of data: all authors. Critical revision of the manuscript for important intellectual content: all authors. Obtained funding: FP, JG. Administrative, technical or material support: FM, NH-A, MVdE, M-GA, A-MM, VS, CL-P. Drafting of the manuscript: FM, FP, JF.

Funding The work was supported by grants from National Institute of Health and Medical Research (INSERM), LabEx Immuno-oncology, Transcan ERANet European project, Association pour la Recherche contre le Cancer, CARPEM, AP-HP (IMMUCOL1 and IMMUCOL2), INCA translationnel, and Paris Descartes University.

Competing interests JG and FP have patents associated with the immune prognostic biomarkers. JG and FH are HaliDx co-founders and shareholders. All other authors declare no competing interests. The IS is a registered trademark owned by the National Institute of Health and Medical Research.

Patient consent for publication Not required.

Ethics approval The study was conducted in accordance with the protocol, good clinical practice standards and the Declaration of Helsinki and was approved by the local ethics committee (2009- A00707-50, 2015/13915, 2012/02NICB, 2014/50NICB).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Thomas Sbarrato <http://orcid.org/0000-0003-1283-5841>

REFERENCES

- Araghi M, Soerjomataram I, Jenkins M, *et al.* Global trends in colorectal cancer mortality: projections to the year 2035. *Int J Cancer* 2019;144:2992–3000.
- Weiser MR, Martin R. Ajcc 8th edition: colorectal cancer. *Ann Surg Oncol* 2018;25:1454–5.
- Weitz J, Koch M, Debus J, *et al.* Colorectal cancer. *Lancet* 2005;365:153–65.
- Nagtegaal ID, Quirke P, Schmoll H-J. Has the new TNM classification for colorectal cancer improved care? *Nat Rev Clin Oncol* 2011;9:119–23.
- Guinney J, Dienstmann R, Wang X, *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;21:1350–6.
- Sinicropo FA, Okamoto K, Kasi PM, *et al.* Molecular biomarkers in the personalized treatment of colorectal cancer. *Clin Gastroenterol Hepatol* 2016;14:651–8.
- Locker GY, Hamilton S, Harris J, *et al.* ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol* 2006;24:5313–27.
- Tumeh PC, Harview CL, Yearley JH, *et al.* Pd-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 2014;515:568–71.
- Mlecnik B, Bindea G, Pagès F, *et al.* Tumor immunosurveillance in human cancers. *Cancer Metastasis Rev* 2011;30:5–12.
- Koebel CM, Vermi W, Swann JB, *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* 2007;450:903–7.
- Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* 2011;331:1565–70.
- Fridman WH, Pagès F, Sautès-Fridman C, *et al.* The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* 2012;12:298–306.
- Galon J, Costes A, Sanchez-Cabo F, *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006;313:1960–4.
- Mlecnik B, Tosolini M, Kirilovsky A, *et al.* Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J Clin Oncol* 2011;29:610–8.
- Pagès F, Kirilovsky A, Mlecnik B, *et al.* In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol* 2009;27:5944–51.
- Plancoulaine B, Laurinaviciene A, Meskauskas R, *et al.* Digital immunohistochemistry wizard: image analysis-assisted stereology tool to produce reference data set for calibration and quality control. *Diagn Pathol* 2014;9(Suppl 1):S8.
- Gray A, Wright A, Jackson P, *et al.* Quantification of histochemical stains using whole slide imaging: development of a method and demonstration of its usefulness in laboratory quality control. *J Clin Pathol* 2015;68:192–9.
- Pagès F, Mlecnik B, Marliot F, *et al.* International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* 2018;391:2128–39.
- Thompson SM, Craven RA, Nirmalan NJ, *et al.* Impact of pre-analytical factors on the proteomic analysis of formalin-fixed paraffin-embedded tissue. *Proteomics Clin Appl* 2013;7:241–51.
- Craven RA, Cairns DA, Zougman A, *et al.* Proteomic analysis of formalin-fixed paraffin-embedded renal tissue samples by label-free MS: assessment of overall technical variability and the impact of block age. *Proteomics Clin Appl* 2013;7:273–82.
- Amin MB, Greene FL, Edge SB, *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 2017;67:93–9.
- Mlecnik B, Bindea G, Angell HK, *et al.* Integrative analyses of colorectal cancer show immunoscore is a stronger predictor of patient survival than microsatellite instability. *Immunity* 2016;44:698–711.
- Jass JR. Lymphocytic infiltration and survival in rectal cancer. *J Clin Pathol* 1986;39:585–9.
- Yagi Y, Gilbertson JR. A relationship between slide quality and image quality in whole slide imaging (WSI). *Diagn Pathol* 2008;3(Suppl 1):S12.
- Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu Rev Immunol* 1989;7:145–73.