OPEN ACCESS

ORIGINAL RESEARCH

Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage III colon cancer outcomes in PETACC08 study

Cynthia Reichling,[1] Julien Taieb,[2] Valentin Derangere,[3] Quentin Klopfenstein,[3] Karine Le Malicot,[4] Jean-Marc Gornet,[5] Hakim Becheur,[6] Francis Fein,[7] Oana Cojocarasu,[8] Marie Christine Kaminsky,[9] Jean Paul Lagasse,[10] Dominique Luet,[11] Suzanne Nguyen,[12] Pierre-Luc Etienne,[13] Mohamed Gasmi,[14] Andre Vanoli,[15] Hervé Perrier,[16] Pierre-Laurent Puig,[17] Jean-François Emile,[18] Come Lepage,[1] François Ghiringhelli [19]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2019-319292).

For numbered affiliations see end of article.

Correspondence to
Professor François Ghiringhelli, Département d'oncologie médicale, Georges-Francois Leclerc Centre, Dijon 21000, France; fghiringhelli@cgfl.fr

Received 14 June 2019
Revised 11 November 2019
Accepted 13 November 2019
Published Online First 28 November 2019

To cite: Reichling C, Taieb J, Derangere V, *et al*. *Gut* 2020;**69**:681–690.

ABSTRACT

Objective Diagnostic tests, such as Immunoscore, predict prognosis in patients with colon cancer. However, additional prognostic markers could be detected on pathological slides using artificial intelligence tools.

Design We have developed a software to detect colon tumour, healthy mucosa, stroma and immune cells on CD3 and CD8 stained slides. The lymphocyte density and surface area were quantified automatically in the tumour core (TC) and invasive margin (IM). Using a LASSO algorithm, DGMate (DiGital tuMor pArameTErs), we detected digital parameters within the tumour cells related to patient outcomes.

Results Within the dataset of 1018 patients, we observed that a poorer relapse-free survival (RFS) was associated with high IM stromal area (HR 5.65; 95% CI 2.34 to 13.67; p<0.0001) and high DGMate (HR 2.72; 95% CI 1.92 to 3.85; p<0.001). Higher CD3+ TC, CD3+ IM and CD8+ TC densities were significantly associated with a longer RFS. Analysis of variance showed that CD3+ TC yielded a similar prognostic value to the classical CD3/CD8 Immunoscore (p=0.44). A combination of the IM stromal area, DGMate and CD3, designated 'DGMuneS', outperformed Immunoscore when used in estimating patients' prognosis (C-index=0.601 vs 0.578, p=0.04) and was independently associated with patient outcomes following Cox multivariate analysis. A predictive nomogram based on DGMuneS and clinical variables identified a group of patients with less than 10% relapse risk and another group with a 50% relapse risk.

Conclusion These findings suggest that artificial intelligence can potentially improve patient care by assisting pathologists in better defining stage III colon cancer patients' prognosis.

Significance of this study

What is already known on this subject?

- ► CD3 and CD8 infiltrates are associated with prognostics for localised colorectal cancer.
- ► Immune infiltrate quantification outperforms tumour intrinsic prognostic variables.
- ► A standardised method called Immunoscore could be used to study these immune infiltrates using a centralised industrial platform.

What are the new findings?

- ► Based on pathological slides obtained from a large prospective study, we generated a new artificial intelligence software for studying both tumour intrinsic prognostic variables and CD3 and CD8 immune infiltrates in stage III colorectal cancer using an automatic procedure.
- ► We observed that CD8 does not provide added prognostic value in comparison to CD3 analysis alone; therefore, a single CD3 slide is sufficient for determining patient prognosis.
- ► The density of tumour stroma and tumour cell intrinsic variables are prognostic.
- ► Determining the density of tumour stroma and tumour cell intrinsic variables combined with CD3 could outperform the CD3/CD8 Immunoscore-like scoring in tumour prognosis.

How might it impact on clinical practice in the foreseeable future?

- ► Our findings show that both tumour stroma and tumour cell intrinsic variables, in association with immune cell infiltrates, should be taken into account during colorectal cancer prognosis.
- ► Here, we provide a novel, freely available and improved alternative to colorectal cancer prognosis using a single standard CD3 pathological slide.
- ► Validation by other studies involving stage II patients or patients treated with 5-fluorouracil alone are warranted to extend the clinical importance of this observation.

INTRODUCTION

Management of metastatic colon cancer (CC) has evolved considerably in recent years due to the availability of new anatomy, molecular biology or immunology data.[1] However, in localised tumours, adjuvant therapy only depends on the pathological

stage defined according to the tumour, node and metastases (TNM) classification.[2] This classification has some limits since prognosis can significantly vary among patients in the same stage. Indeed, in stage III patients, 5-year relapse-free survival (RFS) ranges from 44% to 83%.[3]

The in situ immune environment has also become important for determining patient prognosis and it appears that in most solid tumours, a high T-cell infiltration is associated with a decreased risk of tumour dissemination and improved survival. This correlation is well documented in CC, but also in melanoma, ovarian, breast, prostate and lung cancers.[4 5] In CC, Jérôme Galon *et al* proposed the Immunoscore concept, which studies CD3 and CD8 tumour infiltration in the tumour core (TC) and invasive margin (IM). Immunoscore allows more precise definition of patient prognosis than the TNM stage.[6] This infiltrate is associated with a lower risk of tumour dissemination and improved survival in CC.[7] Recently, an Immunoscore analysis, using a centralised method in a large multicentric prospective study including patients with stage I–III CC was able to distinguish three categories of patients with high, intermediate and low immunoscores and 8%, 19% and 32% recurrence within 5 years, respectively.[8] In addition to the Immunoscore analysis, many studies have underlined the prognostic role of tumour infiltrate lymphocytes in colorectal cancer (CRC).[9–15] However, Luigi Laghi *et al* demonstrated that CD3 infiltrates are only prognostic in stage II tumours.[16 17] Concerning stage III CRC, contrasting data are available in the literature. While Laghi *et al* showed that CD3 infiltrates cannot be independently used to predict patient clinical outcomes,[17] Sinicrope and Pagès, in two different clinical trials addressing FOLFOX-based adjuvant chemotherapy, demonstrated that CD3+ densities can independently predict patient outcomes in stage III CC.[18 19]

Additionally, non-immune factors are associated with outcomes in localised CC. For example, tumour localisation is known to have a prognostic impact, since patients with right CC have a poorer prognosis in metastatic settings.[20–22] Tumour molecular characteristics, such as RAS status, mismatch repair status and consensus molecular subtypes, could also be used to determine prognosis.[23–25] Further, artificial intelligence (AI) could be used to analyse virtual microscopic images and determine, with good accuracy, prognostic and tumour molecular characteristics.[26]

Here, we hypothesised that an AI software could be developed to analyse, in a single procedure on a tumour slide, both immune infiltration and tumour-related prognostic parameters. We further hypothesised that analyses of tumour-related variables generated by AI or combination of tumour-related and immune variables could outperform Immunoscore analyses.

METHODS

Patients

Studied patients belonged to the PETACC8 cohort,[27] an European phase III trial which studied in stage III CC adjuvant treatment with 12 cycles of FOLFOX-4 or a combination of cetuximab and FOLFOX-4. All 2559 patients were originally included between 22 December 2005 and 5 November 2009. Microsatellite stable (MSS) status, K-RAS, N-RAS and BRAF mutational statuses were determined as previously described.[25 28] Enrolled patients had signed an informed consent for translational research. Only 1018 patients of PETACC08 were included in this study, due to slide unavailability from the local pathologist or the absence of written informed consent for ancillary studies. Patients with slides without tumours were also excluded from the study.

CD3 AND CD8 STAINING

CD3 staining of the PETACC08 samples was carried out in Pr. Emile's lab. Slides were stained as previously described,[29] using Bond-Max Fr4.0 (Leica Biosystem) with CD3 primary antibodies (clone F7.2.38, Agilent). For CD8 staining, formalin-fixed paraffin-embedded slides were obtained from Fédération Francophone de Cancérologie Digestive. Slides were stained using anti-CD8 primary antibody (clone (C8/144B), Agilent) and a Bond III apparatus (Leica Biosystem). Once counterstained and permanently mounted, slides were digitised with a Nanozoomer HT2.0 (Hammamatsu) at ×20 magnification to generate a whole slide imaging (WSI) file in ndpi format.

GENERATION OF AI SOFTWARE

Tissue library generation step

All WSI files were automatically segmented by script with the QuPath software[30] using a super pixel strategy. This method tiled the tissue into thousands of parts. Then, 127 parameters (colour, saturation, brightness, texture, etc) were automatically calculated and extracted from each tile. The coordinates of each tile were exported to determine localisation subsequently.

Next, two pathologists hand-annotated the WSIs into different classes, that is, healthy (mucosa), tumour, stroma, immune cells, necrosis and empty space. By definition, tumour stroma consists of the basement membrane, fibroblasts, extracellular matrix, immune cells and vasculature,[31] but we requested that the pathologists exclude immune cells and designated the zone as 'stromal area'. Pathologists were also asked to select stromal areas rich in lymphocytes, which we designated as 'immune areas'. Necrotic tissue and tiles without tissue were grouped as 'other' for further analysis. This work was performed on 80 slides of different histological types. Discrepancies observed between the pathologists were reassessed by both pathologists in a joint meeting.

Classification model set-up step

From this training tissue library, we built a random forest[32] classification model. For histology differentiation, we selected the variables that most discriminated the different tissue classes described above using the VSURF algorithm.[33] A training model was then built for each histology differentiation group using the variables selected. For patients with unknown histological differentiation, we used meta training, regrouping all the training data available. The model used to classify WSI tiles was selected based on the data available from histology differentiation in the PETACC8 database. This model was called ColoClass.

TC and invasion margin estimation

The TC was obtained by merging adjacent tiles that were classified as tumour cells by the classification model. Once TC was estimated by tumour tiles' clustering, a 300 µm extra-boundary was automatically plotted. The area between this extra-boundary and TC is the IM, selected as tissue distance >500 µm from the tumour border in previous studies.[34 35] We tested different IM distances (200, 300, 400 and 500 µm) for CD3 and found that 300, 400 and 500 µm yielded similar results with strongly correlated variables and similar prognostic value, then we decided to select sample sets with 300 µm IM to reduce the duration for calculation.

CD3 and CD8 detection step

After measuring and exporting data from all WSI tiles, a script was run to detect any cell on WSI and export the coordinates. Thus, using QuPath detection scripting, positive cells for each

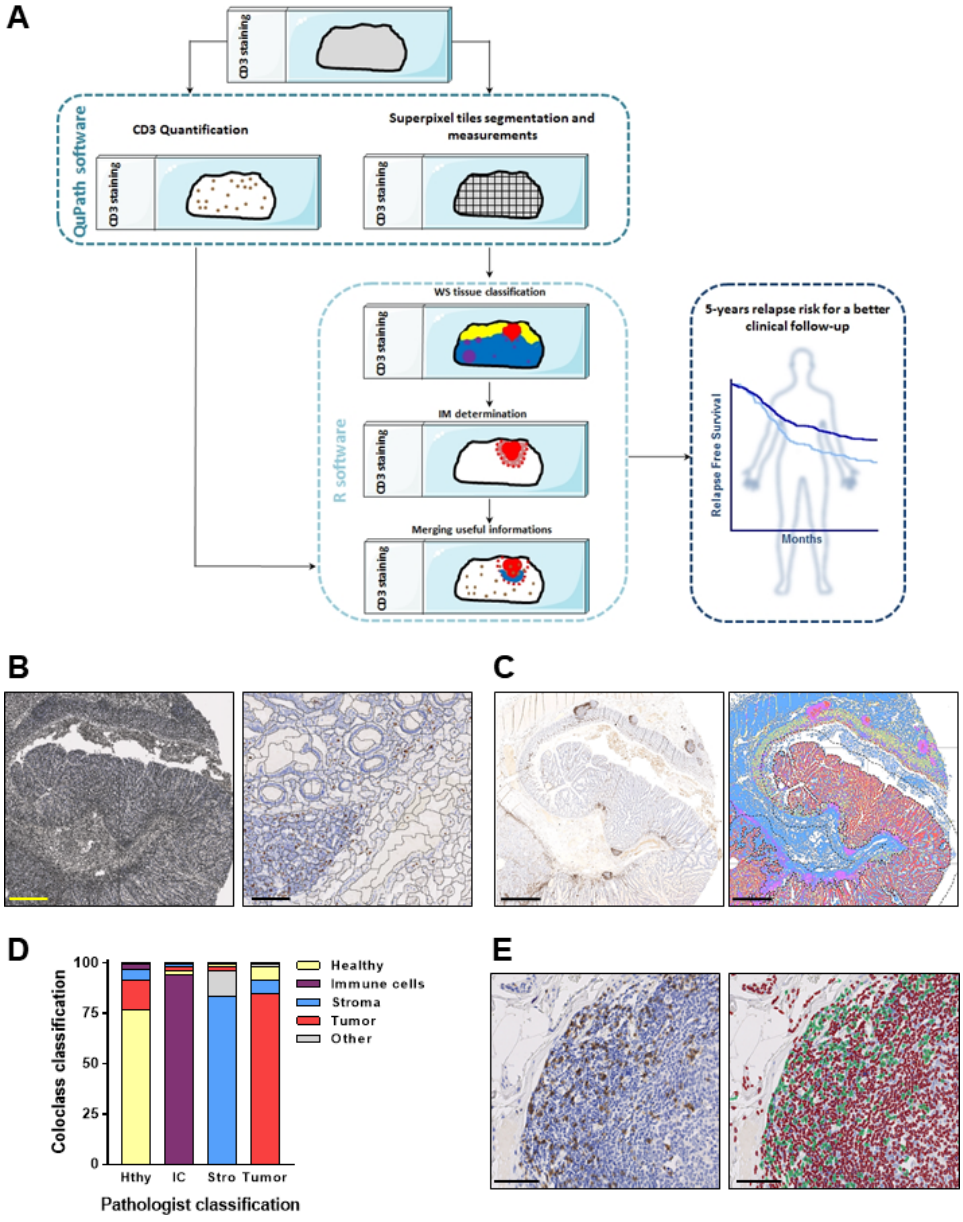

Figure 1 Tissue classification methodology and immune cell quantification. (A) Slides are segmented in thousands tiles using QuPath. Each tile is then classified with ColoClass R software. CD3 or CD8 staining is simultaneously evaluated with QuPath. All information are gathered to predict colon cancer relapse. (B) Representative pictures of a tiled slide at low magnification (left panel, scale bar 1 mm) and at high magnification (right panel, scale bar 250 µm). (C) Representative pictures of tissue classification from native slide (left panel) to ColoClass (right panel). Healthy mucosa is displayed in yellow, tumour in red, stroma in blue and immune cells in purple. The dotted line represents ColoClass IM estimation (scale bar 1 mm). (D) Validation of ColoClass versus pathologists. (E) Detection of positive cells on native slide (left panel) and using QuPath (right panel). Positive cells are displayed in green and negative cells in red (scale bar 100 µm). Hthy, healthy mucosa; IC, immune cells; IM, invasivemargin; Other, gathers white spaces and necrosis; Stro, stroma; WS, whole slide.

marker (ie, CD3 or CD8) were differentiated from negative ones. By gathering cell and tile coordinates, we were able to determine the accurate position of each cell and the class to which it belonged (ie, healthy, tumour, immune or stroma). The replicability of the method was tested by scanning several CD3 slides thrice, thus generating several .ndpi files. An independent bioinformatician then processed each file within our QuPath and R scripts, and checked DGMunes (DGMate (DiGital tuMor pArameTErs) associated with immune and stroma information) and subsequent scores. DGMuneS variance was ~5%. The concordance between semiquantitative evaluation of CD3 by two pathologists (~90% and 89%) was determined and the QuPath detection scripting was performed.

Classification model validation step

To validate our classification model, we randomly selected 53 slides which did not belong to the training dataset. These slides were identically processed (ie, segmentation, digital parameters measurement, coordinate extraction). Two pathologists were asked to classify some tiles on WSI using QuPath and discrepancies between them were reassessed by both pathologists in a joint

meeting. The annotated tiles were then exported and processed in ColoClass.

All data, the R code and Groovy script for QuPath are available on GitHub (https://github.com/Klopfe/PETACC8). A tutorial is supplied as online supplementary file.

STATISTICAL ANALYSIS

Survival analysis

The prognostic value of the different variables was tested through Cox proportional hazard models for RFS, which was defined as time to the first relapse or death from any cause. Survival probabilities were estimated using the Kaplan-Meier method, and survival curves were evaluated using the log-rank test. Patients with RFS periods longer than 5 years were censored.

DGMate score construction

QuPath[30] was used to measure 127 parameters in each software segmented tile. We had computed the mean of each tumour tile parameter for each slide, yielding 127 parameters per slide. Then a LASSO[36] algorithm was performed to select the variables that were related to the RFS using the glmnet R package.[37 38] The DGMate score is the linear predictor of the Cox model built on the discovery cohort with selected variables via the LASSO procedure.

Discovery and validation cohorts

To validate the DGMate score as a prognostic variable, we split the cohort in two different groups by random sampling, placing 70% of the patients in the discovery cohort and 30% in the validation cohort. Both groups were comparable for each clinical variable.

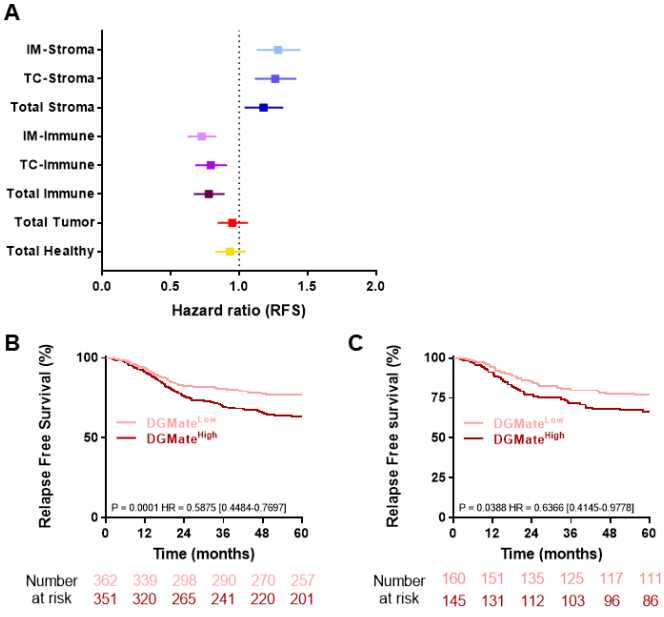

Figure 2 Predictive value of interest areas and digital features. (A) Forest plot representing the predictive value of stroma area (IM, TC and WSI), immune area (IM, TC and WSI), total tumour area and total healthy area on RFS. (B) Kaplan-Meier survival curve on discovery dataset (n=713) using DGMate split at median. (C) Kaplan-Meier survival curve on validation dataset (n=305) using DGMate split at median. IM, invasive margin; RFS, relapse-free survival; TC, tumour core; WSI, whole slide imaging.

Replication of Immunoscore

Immunoscore-like scores were generated as Immunoscore and assessed as previously described.[8] We computed the percentiles for CD3 IM, CD3 TC, CD8 IM and CD8 TC variables, from which the average percentile of the four variables was calculated for each patient. A three-category Immunoscore system was designed, and patients with scores ranging from 0 to 0.25, 0.25 to 0.7 and >0.7 were classed as having low, intermediate and high Immunoscores, respectively. A two-category Immunoscore system was also designed, and patients with scores ranging from 0 to 0.25 and >0.25 were classed as having low and high Immunoscores, respectively.

Predictive accuracy of Cox models

To evaluate the predictive accuracy of different models and to be able to compare their performance, we used 1000 bootstrap resampling and computed the predictive accuracy (AUC) for each bootstrap sampling. Model performances were compared using likelihood ratio tests, when the models were nested.

Nomogram construction

We used a nomogram representation of the multivariate Cox models combining the DGMuneS score, N stage, T stage, differentiation and RAS status to build a score. This score was then used to classify patients into three different categories. The score cut-offs were as follows: 20% of patients with the highest scores were classified as 'high', 20% with the lowest score as 'low' and the rest as 'intermediate'. This choice was driven by our decision to establish an intermediate group with a survival pattern similar to the global population.

Software and available data

R v3.3.3 was used for statistical analysis. Figures were performed using GraphPad 7.03.

RESULTS

Generation of an AI software to classify tissue structure in a CC pathological slide

In a haematoxylin CC tumour slide, six tissue structures were detected by a pathologist: the TC, the immune and the stromal tissues, necrotic areas, normal colon mucosae and areas without tissue. The IM was arbitrarily defined as an area 300 µm distant from the TC. Using the open source QuPath software, we performed tissue segmentation based on megapixel strategy, which regrouped pixels (called tiles) based on their similarity (figure 1A, methodological workflow and figure 1B), using a training set of 80 slides from different histological types. In total, 27 466 tiles were used to set up this tissue library. From this training tissue library, we built a random forest classification model (figure 1C) called ColoClass. To validate our classification model, we randomly selected 54 additional slides, which were classified by two pathologists. Annotated tiles were then exported and processed in ColoClass. In total, 26 659 tiles were processed and ~85% concordance (22 652 tiles) was found between the pathologist and ColoClass classifications (figure 1D). For a comparable dataset, the concordance between both pathologists was 87%. Similar results were observed independent of the histological differentiation type (online supplementary table S1). Positive T-cell (CD3 or CD8) detection was performed with QuPath and automatically attributed to the classified area. Pooling these information, the software is able to automatically determine the area corresponding to a particular

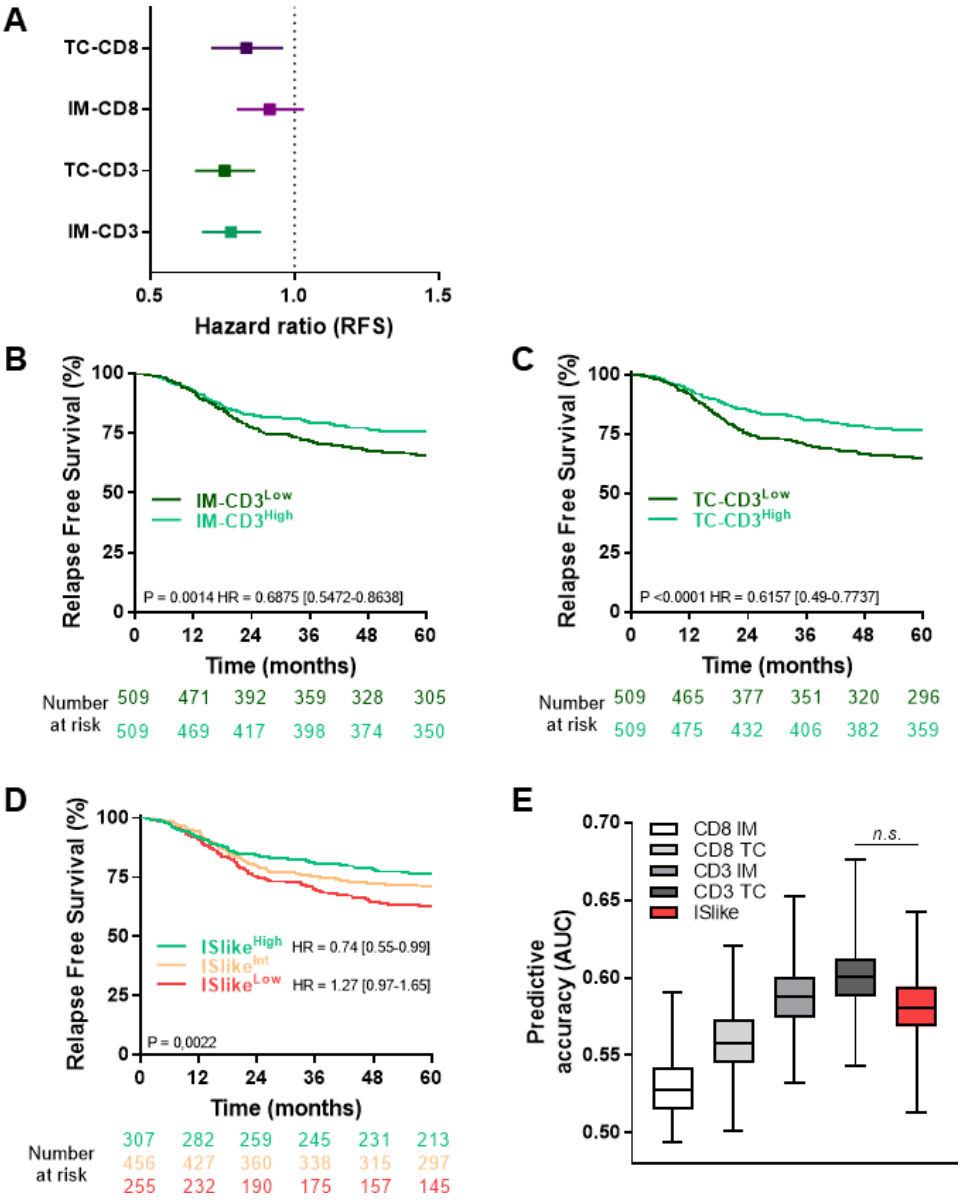

Figure 3 Prognostic value of immune T-cell analysis. (A) Forest plot representing the predictive value of CD3 and CD8 in IM and TC on RFS. (B) Kaplan-Meier RFS curved using CD3+ IM split at median. (C) Kaplan-Meier RFS curved using TC CD3 split at median. (D) Kaplan-Meier RFS curve using Immunoscore split at three risk groups (low 20%, intermediate 60% and high 20%). (E) Immunoscore predictive accuracy compared with CD3+ IM, CD3+ TC, CD8+ IM, CD8+ TC alone using a 1000× bootstrap strategy. AUC, area under the receiver operating characteristic curve; IM, invasive margin; ISlike, Immunoscore like; RFS, relapse-free survival; TC, tumour core; TC CD3, CD3 tumour-infiltrating lymphocytes present in the TC.

tissue on each slide, as well as the CD3 and CD8 cell infiltration in each tissue (figure 1E).

Prognostic role of tissue analysis

We tested the prognostic role of each variable on RFS in the PETACC8 cohort, including 1220 patients (flow chart, online supplementary figure 1). Some patients (n=202; 16.5%) were excluded after quality control, mostly due to the lack of tumour detection on the slide. Patient characteristics are presented in online supplementary table S2. We tested the relationship between each area determined by the software, used as a continuous variable, and RFS (figure 2A). High stromal and immune areas were, respectively, associated with poor and good outcomes. Data were also represented using Kaplan-Meier curves, and groups were separated using the median as the cut-off (online supplementary figure 2). Healthy and tumour areas were not associated with RFS. The stromal area was weakly anticorrelated with the immune area or CD3-TC (r=0.4015 or 0.3676, respectively; p<0.001). Stromal areas in IM and TC were strongly correlated (online supplementary figure 3). For further analysis, we decided to focus only on the stromal area in IM, due to a higher difference in HR and a more significant p-value overall. Similarly, immune areas were strongly correlated with CD3 infiltrate (online supplementary figure 3). Stromal area increased with T stage, but remained unaffected by N stage, sidedness, deficient mismatch repair (dMMR) status and RAS status (online supplementary table S3). The software analysed 127 parameters per tile, and we used a LASSO algorithm to select variables associated with outcomes, thereby deriving the DGMate score. In a training set of 713 patients, we selected eight variables (online supplementary table S4) which were associated with outcomes using the LASSO procedure (HR=2.718;

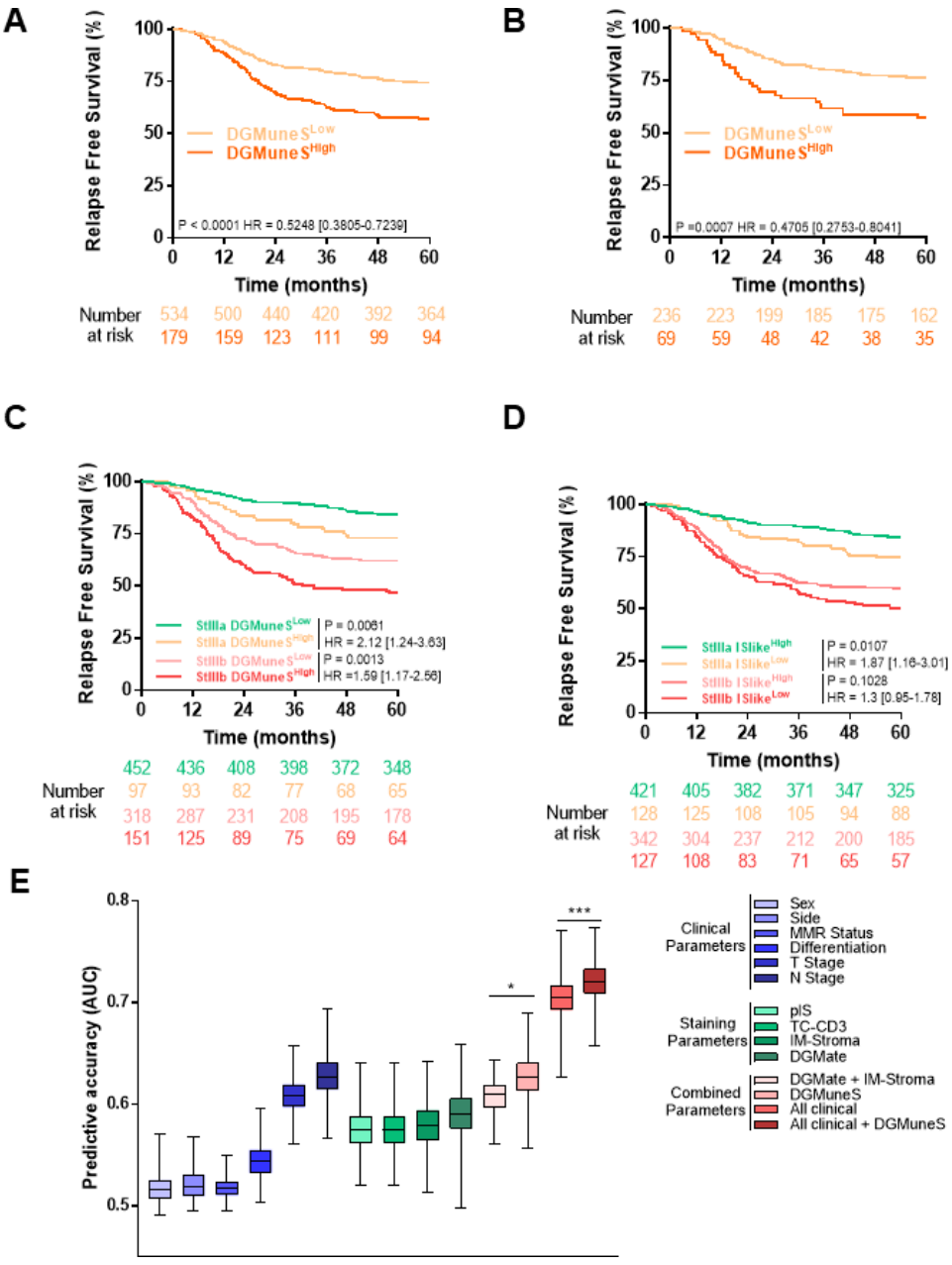

Figure 4 Composite variables improve prognosis prediction. (A) Kaplan-Meier relapse-free survival curve on discovery dataset (n=713) using DGMuneS split at third quartile. (B) Kaplan-Meier relapse-free survival curve on validation dataset (n=305) using DGMuneS split at third quartile. (C) Kaplan-Meier relapse-free survival curve on low-risk clinical stage (T1–3, (N1) (n=549) and high-risk clinical stage (T4 or N2) patients (n=469), both split at median, using DGMuneS. (D) Kaplan-Meier relapse-free survival curve on low-risk clinical stage (T1–3, (N1) (n=549) and high-risk clinical stage (T4 or N2) (n=469), using dichotomic ISlike score. (E) Predictive accuracy on patients' relapse depending on clinical parameters (blue shading), staining parameters (green shading) or combined parameters (red shading) based on 1018 patients from PETACC08 study using a 1000× bootstrap strategy. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. AUC, area under the receiver operating characteristic curve; IM, invasive margin; ISlike, Immunoscore like; MMR, mismatch repair; N stage, node stage; T stage, tumour stage; TC, tumour core; TC CD3, CD3 tumour-infiltrating lymphocytes present in the TC.

95% CI 1.853 to 3.988; p=3.1e-07 for continuous variable). We confirmed the prognostic role of this tumour signature in the validation dataset n=305 (HR=2.128; 95% CI 1.162 to 3.898); p=0.01 for continuous variable) (figure 2B,C). The DGMate score increased with T and N stage, sidedness, RAS status and dMMR status (online supplementary table S5).

Prognostic role of immune T-cell infiltration analysis

Based on the Immunoscore rationale, we evaluated CD3+ and CD8+ cells in both IM and TC, and studied CD3 and CD8 infiltration as a function of classical prognostic variables (online supplementary table S6). A higher T stage was associated with less CD3 infiltration, in both IM and TC. Similarly, a lower CD3 infiltration was observed in IM in N2 stage patients. Higher infiltration of CD3 and CD8 in IM and TC was observed in right-sided tumours. Higher infiltration of CD3 and CD8 in the TC, but not in IM, was observed with dMMR tumours; in contrast, RAS/BRAF mutated status did not impact immune infiltrates. We tested the correlation between the four immune variables (CD3+ IM, CD3+ TC, CD8+ IM, CD8+ TC) and

Table 1 Cox multivariate analysis of the relationship between clinical and image variables and RFS

Multivariate Cox models			
	Discovery cohort (n=713)		
	HR	95% CI	P value
Histopathology grading			
Poorly/undifferentiated	1		
Well/moderately	0.811	0.585 to 1.125	0.21
Unknown			
T			
T3	1		
T1/T2	0.516	0.240 to 1.111	0.09
T4	1.696	1.255 to 2.292	0.0006
N			
N1	1		
N2	2.181	1.655 to 2.875	3.10E-08
KRAS			
Mutated	1		
Wild type	0.644	0.488 to 0.848	0.002
DGMuneS	1.989	1.373 to 2.881	0.0003
C-index		0.69	
AIC		2567.345	

AIC, akaike information criterion; KRAS, kirsten rat sarcoma viral oncogene homolog; RFS, relapse-free survival.

observed a strong correlation between each variable (p<0.0001 with an R from 0.42 to 0.81) (online supplementary figure S3). Using the four immune variables as continuous variables, we tested their prognostic role in RFS (figure 3A). Using the median as the cut-off, high CD3+ IM, CD3+ TC and CD8+ TC were significantly associated with a better outcome, while CD8+ IM was close to significance (figure 3B,C and online supplementary figure S4). By combining these four variables in an 'Immunoscore like' (ISlike) manner as described by Pagès *et al*,[8] we observed a better prognosis for high ISlike patients (figure 3D). The ability of ISlike to predict RFS was compared with CD3+ IM, CD3+ TC, CD8+ IM or CD8+ TC used alone by testing the predictive accuracy for RFS, based on the time dependent area under the receiver operating characteristic curve (AUC) with 1000× bootstrap. ISlike did not significantly outperformed the TC-CD3 unique variable (likelihood ratio p=0.15) (figure 3E). Hence, for further analysis, we decided to use only CD3 in TC to assess immune variables, since this variable was more significantly associated with RFS.

Composite variable, including immune infiltrate, stromal area and DGMate, improved patient prognosis estimation

CD3+ TC, stromal area in IM and DGMate were weakly correlated (online supplementary figure S5). By combining CD3+ TC, stromal area in IM and DGMate, based on the discovery set, we generated a DGMuneS score, which was strongly associated with RFS in the discovery dataset and similar results were observed in the validation dataset (figure 4A,B). A multivariate Cox analysis for RFS, including all available clinical parameters and the DGMuneS composite score, revealed that the composite score remained independently associated with the outcome in both training and validation cohorts (table 1, online supplementary tables S2 and S7). Similar prognostic results were observed in both FOLFOX and FOLFOX–cetuximab groups (online supplementary figure S6). Subgroup analysis showed that the DGMuneS variable is significantly associated with prognosis in either T3N1 or T3N2 tumour stage groups (online supplementary table S8) and could significantly identify a group of patients with very good or poor outcomes within clinically low-risk (T1-3, N1) or high-risk (T4 or N2) groups, respectively (figure 4C). Contrastingly, the ISlike dichotomous variable was unable to significantly discriminate prognostic groups among high-risk clinical stage (T4 or N2) patients (figure 4D). No significant interaction was found between the digital and classical clinical prognosis variables with the exception of an interaction between DGMate and the histological grade (p=0.01). Using logistic regression, we tested the capacity of digital variables to predict classical prognostic variables, such as T stage, N stage, differentiation status and RAS/BRAF mutation. No variable was able to predict the differentiation status. Only DGMate could significantly predict the N stage and RAS/BRAF mutation status. All variables, TC-CD3, DGMate and stromal area, could be used to predict T stage (online supplementary table S9). The predictive accuracy of the DGMuneS score was evaluated by determining the time-dependent AUC; was found to be superior to tumour grade, RAS status, MSI status, sidedness or Immunoscore; and had a similar time-dependent AUC values as T stage and N stage. Furthermore, adding DGMuneS score to a model that combined all clinical variables (sex, side, MMR status, differentiation, T stage, N stage) significantly improved RFS prediction (likelihood ratio p=0.0007; figure 4E). Both tumour-related (stromal area and DGMate) and CD3 immune variables in TC are required to optimise the determination of patient prognosis (likelihood ratio p=0.04; figure 4E).

To address patient prognoses, we generated a nomogram tool based on variables retained in the multivariate model (ie, DGMuneS, T stage, N stage, tumour differentiation and RAS status). The prognostic score was based on the total number of points obtained on the nomogram (figure 5A). Patients were categorised into three risk groups, representing 20% lower, 60% intermediate and 20% higher scores. This separation was arbitrarily selected in such a way that the intermediate group had a similar RFS pattern to the entire cohort. A 5-year risk of relapse was 12% in the lower group (high vs low; HR 0.167; 95% CI 0.099 to 0.284; p=5e-13), 28% in the intermediate group (high vs intermediate; HR 0.428; 95% CI 0.319 to 0.573; p=5.6e-09) and 52% in the higher group (figure 5B). Using the same cut-off limits than in the training set, we identified the same risk groups in the validation dataset, with RFS represented by Kaplan-Meier curves (all p<0.001) (figure 5C).

DISCUSSION

Optimisation of adjuvant strategies for localised CC remains an important issue. On recent international guidelines based on the IDEA study,[39] it is suggested that patients with low-risk clinical stage CC only require 3 months of FOLFOX or XELOX regimen, while 6 months of oxaliplatin-based chemotherapy is recommended for high-risk clinical stage patients. With such treatment, 3-year RFS of 82% and 62% are attained for low-risk and high-risk clinical stage patients, respectively. Additional prognostic markers are necessary to better determine patient prognosis. Recent data in stage III CC patients show that BRAF or RAS mutations are independently associated with a shorter time to recurrence and overall survival in patients with MSS, but not with MSI status.[24] Moreover, immune infiltrate was also associated with tumour prognosis and Immunoscore was shown to predict outcomes in a large cohort of patients with stage I, II and III CC.[8]

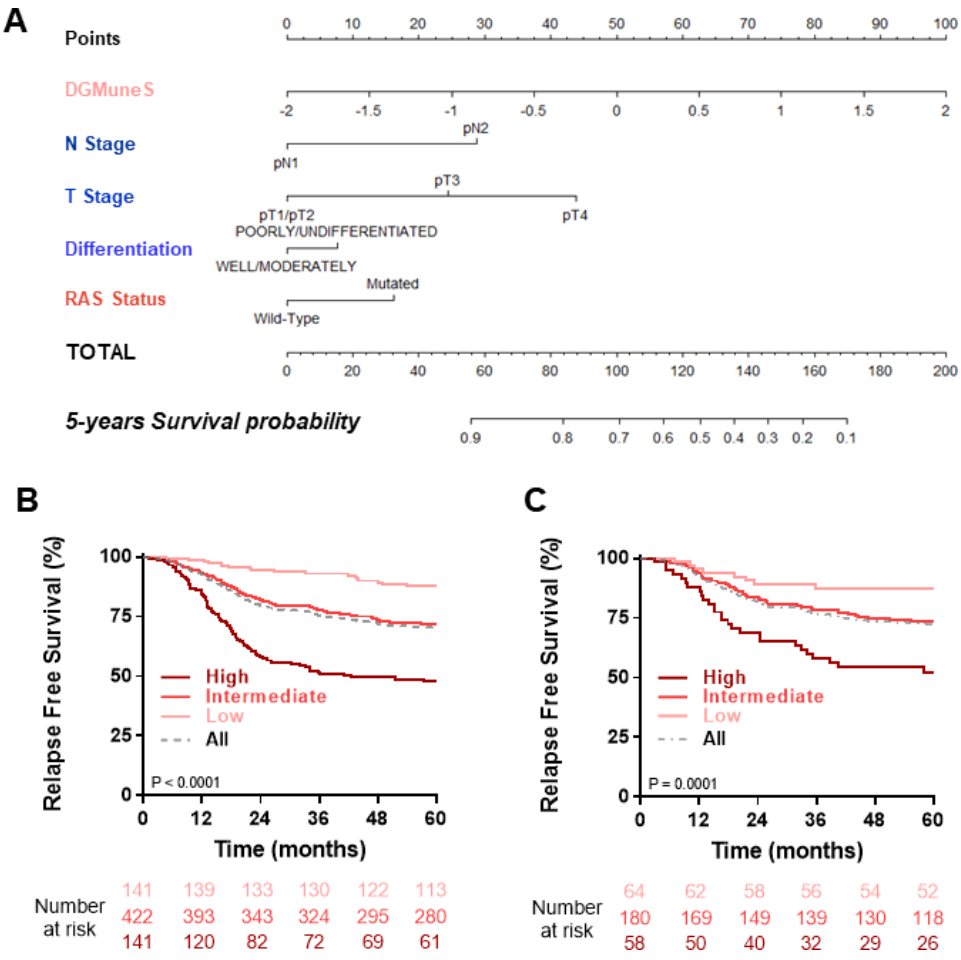

Figure 5 Nomogram tool based on variable retained in the multivariate model and relapse-free survival according to total score. (A) Nomogram representation of the multivariate model. Each parameter gives a number of points indicated on the upper line. The sum is indicated on total line and provides a 5-year survival probability. (B) Kaplan-Meier survival curve on discovery dataset (n=713) when nomogram score is split depending on relapse risk (low, light red line; high, dark red line or intermediate, red line), grey dotted line displays survival of whole discovery dataset. (C) Kaplan-Meier survival curve on validation dataset (n=305) when nomogram score is split depending on relapse risk (low, light red line; high, dark red line or intermediate, red line), grey dotted line displays survival of whole validation dataset. N stage, node stage; T stage, tumour stage.

We postulated that the analysis of tissue structure, tumour cell characteristics and immune infiltrate could be combined to predict CC outcomes and could outperform Immunoscore. Using random forest classifiers, we generated a software which could characterise tumour cells, normal and stromal areas. The software detected IM and enumerated immune infiltrates in each area. We observed that such tissue analysis was interesting from a prognosis point of view, since a large stromal area was linked to a poorer prognosis. Such data were reminiscent of previous data obtained in stage II and III CC, which underlined a poorer prognosis of high intratumorous stromal tissue patients.[40] Using a LASSO algorithm, we also isolated eight tumour cell parameters associated with patient outcomes, called the DGMate score. Surprisingly, this parameter was not associated with RAS, MSI status or the tumour differentiation status, suggesting that digital pathology combined with machine learning isolated independent prognostic phenotypic features that could not be recognised by human analysis.

Our analysis of CD3 and CD8 variables generated an ISlike score with time-dependent discrimination properties (AUC of 0.56) similar to the recent publication on the international validation of Immunoscore (AUC of 0.57).[8] Importantly, CD3 and CD8 variables were strongly correlated, and the ISlike score and CD3 tumour-infiltrating lymphocytes (TILs) present in the TC (TC CD3) yielded similar AUCs. So, in stage III CC, we do not believe that Immunoscore provides any additional value to a simple TC CD3 accumulation analysis. Most TC-CD3 are located in stromal tissue around tumour cell islets which are poorly invaded in most cases. Analysis of the prognostic role of CD3 TILs located in stromal areas present in the TC or TILs present in tumour islets did not outperform global TC CD3 analysis, so we have conserved this variable for further analyses (not shown). CD8 staining did not provide an added value in our study, although one limitation is that CD8 labelling was performed on older slides than used for CD3 labelling, raising the possibility that the staining was less efficient. Our study showed that a composite variable, including stromal area, CD3 and DGMate, was highly predictive of outcomes and had superior discriminatory properties when compared with ISlike score or clinical variables. A limitation of this observation could be our use of an adapted algorithm based on Immunoscore methodology. Indeed, this model provided results similar to Immunoscore when we compared time-dependent AUC of ISlike in our series and Immunoscore time-dependent AUC in the recent international validation of Immunoscore.[8] Additional studies to directly compare Immunoscore and DGMuneS should be

performed. To confirm our score, cross-validations will be necessary, and the use of different slide scanners should be addressed. The Immunoscore group proposed that T and N stages are not essential prognostic biomarkers and could be replaced by an immune variable. Our study demonstrates the opposite and clearly underlined that immune or digital variables are an added value, but could not replace clinical staging. Our data corroborate previous observations by Laghi *et al* which showed a dependency between N stage and CD3 infiltrates.[17] While our Immunoscore-like variable could not predict the prognosis in high-risk stage III patients, DGMuneS remained prognostic, suggesting that adding digital variables to an assessment of immune infiltrates improves tumour prognostic prediction. We have designed a nomogram to implement the scoring system. Currently, no prognostic model is available for estimating RFS in patients with stage III CC. Using this strategy, we have identified patients with reduced 5-year RFS (50%) rates. Such patients might probably require intensive follow-up and should be included in clinical trials testing intensive adjuvant therapy, such as the IROCAS study (NCT02967289). We have also pinpointed patients with a very good prognosis (90% 5-year RFS). In this context, the relative risk/benefit of adjuvant therapy should be discussed and clinical trials addressing adjuvant therapy minimisation should be initiated for such patients. We have provided here a nomogram and the required software, together with a tutorial. Using this open software, CD3 analysis could be performed by every pathologist.

Although this study involves a large and homogeneous cohort of stage III CC, our work also has limitations. The post hoc design of the analysis and the limited number of patients in some subgroups may restrict our conclusions. Training and validation cohorts for this work were performed by sampling the global trial population. Despite internal validation, external validation in prospective trials is warranted to validate the reproducibility of our software using different datasets. Similar studies should also be performed for stage II tumours, to test if our data can be validated in the context of patients who do not receive chemotherapy frequently.

Author affiliations
[1]Département d'hépato-gastroentérologie et en oncologie digestive, Hôpital du Bocage, Dijon, Bourgogne-Franche-Comté, France
[2]Service d'hépato-gastroentérologie, Hopital Europeen Georges Pompidou, Paris, France
[3]Plateforme de recherche biologique en oncologie, Georges-Francois Leclerc Centre, Dijon, Bourgogne-Franche-Comté, France
[4]Fédération Francophone de Cancérologie Digestive, Hôpital du Bocage, Dijon, Bourgogne-Franche-Comté, France
[5]Département d'hépato-gastroentérologie, Hospital Saint-Louis, Paris, Île-de-France, France
[6]Département d'hépato-gastroentérologie, Hôpital Bichat Claude-Bernard, Paris, Île-de-France, France
[7]Département d'hépato-gastroentérologie, CHU Besancon, Besancon, France
[8]Département d'onco-hématologie, Le Mans Universite, Le Mans, Pays de la Loire, France
[9]Département d'oncologie médicale, Institut de Cancérologie de Lorraine, Vandoeuvre-les-Nancy, Lorraine, France
[10]Département d'hépato-gastroentérologie et en oncologie digestive, Orleans University, Orleans, France
[11]Département d'hépato-gastroentérologie et en oncologie digestive, CHU Angers, Angers, Pays de la Loire, France
[12]Service d'Oncologie Médicale, CH Pau, Pau, Aquitaine-Limousin-Poitou, France
[13]Service d'Oncologie Médicale, Hospital Centre Saint Brieuc, Saint Brieuc, Bretagne, France
[14]Département d'hépato-gastroentérologie, Assistance Publique Hopitaux de Marseille, Marseille, Provence-Alpes-Côte d'Azu, France
[15]Département d'oncologie médicale, Clinique Sainte Marthe, Dijon, Bourgogne, France
[16]service d'oncologie, Hopital Saint Joseph, Marseille, Provence-Alpes-Côte d'Azu, France
[17]pole biologie, Hospital European George Pompidou, Paris, Île-de-France, France
[18]EA4340, Ambroise Pare Hospital, Beuvry, Hauts-de-France, France
[19]Département d'oncologie médicale, Georges-Francois Leclerc Centre, Dijon, Bourgogne-Franche-Comté, France

Acknowledgements The authors thank Elsevier Editing Service for English editing. The authors also thank Caroline Truntzer for performing the replicability testing.

Contributors FG designed the study, interpreted data and wrote the manuscript. CR conducted the majority of the experiments. QK generates the R software and perform statistical analysis. VD generates the groovy software. VD and CR perform histological slide analysis. JT and CL are clinical researchers who coordinate PETACC08 study. JFE collect and store all tissue sample, provide unstained slide for CD8 labelling and perform CD3 labelling. P-LP makes all molecular analysis (RAS, BRAF, MMR determination), KLM performed statistical analysis of PETACC08 and provided clinical database. J-MG, HB, FF, OC, MCK, JPL, DL, SN, P-LE, MG, AV, HP, JT, CL and FG are main clinical investigators of this study.

Funding This work was supported by Ligue National Contre le Cancer (Labelisation F. Ghiringhellli).

Competing interests JPL served on external advisory board or Sanofi Avantis France; received fee for travel from Ipsen, Novartis, Amgen, Roche; received fee for communication from Novartis and funding for research was provided by Merck Serono, Roche and MSD. DL received fee for travel from Merck Serono and Amgen. CL receives speakers bureau honoraria from Amgen, Novartis and Bayer and is a consultant/advisory board member for Novartis and Halio-DX. P-LP is a consultant/advisory board member for Merck Serono, Amgen, Boerhinger Ingelheim, Biocartis, Roche, Bristol-Myers Squibb and MSD. JT has received honoraria for speaker or advisory role from Sanofi, Roche, Merck, Amgen, Sirtex, Servier, Lilly, Celgene and MSD. FG served on external advisory boards for Roche. Research funding received from Roche, Genentech, Amgen, Enterome and Servier. Received funding for clinical trial from Astra Zeneca; received fee for communication from Amgen, Astra Zeneca, BMS, Sanofi, Merck-Serono and Servier and received fee for travel from Roche and Servier.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD
François Ghiringhelli http://orcid.org/0000-0002-5465-8305

REFERENCES

1 Van Cutsem E, Cervantes A, Adam R, *et al*. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol* 2016;27:1386–422.
2 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.
3 O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American joint Committee on cancer sixth edition staging. *J Natl Cancer Inst* 2004;96:1420–5.
4 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.
5 Senovilla L, Vacchelli E, Galon J, *et al*. Trial Watch: prognostic and predictive value of the immune infiltrate in cancer. *Oncoimmunology* 2012;1:1323–43.
6 Galon J, Pagès F, Marincola FM, *et al*. The immune score as a new possible approach for the classification of cancer. *J Transl Med* 2012;10:1.
7 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.
8 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.
9 Ropponen KM, Eskelinen MJ, Lipponen PK, *et al*. Prognostic value of tumour-infiltrating lymphocytes (TILs) in colorectal cancer. *J Pathol* 1997;182:318–24.
10 Morris M, Platell C, Iacopetta B. Tumor-Infiltrating lymphocytes and perforation in colon cancer predict positive response to 5-fluorouracil chemotherapy. *Clin Cancer Res* 2008;14:1413–7.
11 Huh JW, Lee JH, Kim HR. Prognostic significance of tumor-infiltrating lymphocytes for patients with colorectal cancer. *Arch Surg* 2012;147:366–72.

12 Di Caro G, Bergomas F, Grizzi F, *et al*. Occurrence of tertiary lymphoid tissue is associated with T-cell infiltration and predicts better prognosis in early-stage colorectal cancers. *Clin Cancer Res* 2014;20:2147–58.
13 Rozek LS, Schmit SL, Greenson JK, *et al*. Tumor-Infiltrating lymphocytes, Crohn's-Like lymphoid reaction, and survival from colorectal cancer. *J Natl Cancer Inst* 2016;108. doi:10.1093/jnci/djw027. [Epub ahead of print: 12 May 2016].
14 Williams DS, Mouradov D, Jorissen RN, *et al*. Lymphocytic response to tumour and deficient DNA mismatch repair identify subtypes of stage II/III colorectal cancer associated with patient outcomes. *Gut* 2019;68:465–74.
15 Kather JN, Krisam J, Charoentong P, *et al*. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* 2019;16:e1002730.
16 Cavalleri T, Bianchi P, Basso G, *et al*. Combined Low Densities of FoxP3$^+$ and CD3$^+$ Tumor-Infiltrating Lymphocytes Identify Stage II Colorectal Cancer at High Risk of Progression. *Cancer Immunol Res* 2019;7:751–8.
17 Laghi L, Bianchi P, Miranda E, *et al*. Cd3+ cells at the invasive margin of deeply invading (pT3-T4) colorectal cancer and risk of post-surgical metastasis: a longitudinal study. *Lancet Oncol* 2009;10:877–84.
18 Yoon HH, Shi Q, Heying EN, *et al*. Intertumoral Heterogeneity of CD3$^+$ and CD8$^+$ T-Cell Densities in the Microenvironment of DNA Mismatch-Repair-Deficient Colon Cancers: Implications for Prognosis. *Clin Cancer Res* 2019;25:125–33.
19 Pages F, Andre T, Taieb J, *et al*. Validation of the immunoscore prognostic value in stage III colon cancer patients treated with oxaliplatin in the prospective idea France cohort study (PRODIGE-GERCOR). *J Clin Oncol* 2019;37:3513.
20 Venook AP, Niedzwiecki D, Innocenti F, *et al*. Impact of primary (1º) tumor location on overall survival (OS) and progression-free survival (PFS) in patients (pts) with metastatic colorectal cancer (mCRC): analysis of CALGB/SWOG 80405 (Alliance). *J Clin Oncol* 2016;34:3504.
21 Brulé SY, Jonker DJ, Karapetis CS, *et al*. Location of colon cancer (right-sided versus left-sided) as a prognostic factor and a predictor of benefit from cetuximab in NCIC CO.17. *Eur J Cancer* 2015;51:1405–14.
22 Petrelli F, Tomasello G, Borgonovo K, *et al*. Prognostic survival associated with left-sided vs right-sided colon cancer: a systematic review and meta-analysis. *JAMA Oncol* 2017;3:211–9.
23 Guinney J, Dienstmann R, Wang X, *et al*. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;21:1350–6.
24 Taieb J, Le Malicot K, Shi Q, *et al*. Prognostic Value of *BRAF* and *KRAS* Mutations in MSI and MSS Stage III Colon Cancer. *J Natl Cancer Inst* 2017;109:djw272.
25 Taieb J, Zaanan A, Le Malicot K, *et al*. Prognostic effect of BRAF and KRAS mutations in patients with stage III colon cancer treated with leucovorin, fluorouracil, and oxaliplatin with or without cetuximab: a post hoc analysis of the PETACC-8 trial. *JAMA Oncol* 2016;2:643–53.
26 Coudray N, Ocampo PS, Sakellaropoulos T, *et al*. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559–67.
27 Taieb J, Tabernero J, Mini E, *et al*. Oxaliplatin, fluorouracil, and leucovorin with or without cetuximab in patients with resected stage III colon cancer (PETACC-8): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15:862–73.
28 Taieb J, Balogoun R, Le Malicot K, *et al*. Adjuvant FOLFOX +/- cetuximab in full RAS and BRAF wildtype stage III colon cancer patients. *Ann Oncol* 2017;28:824–30.
29 Emile J-F, Julié C, Le Malicot K, *et al*. Prospective validation of a lymphocyte infiltration prognostic test in stage III colon cancer patients treated with adjuvant FOLFOX. *Eur J Cancer* 2017;82:16–24.
30 Bankhead P, Loughrey MB, Fernández JA, *et al*. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017;7:16878.
31 Bremnes RM, Dønnem T, Al-Saad S, *et al*. The role of tumor stroma in cancer progression and prognosis: emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. *J Thorac Oncol* 2011;6:209–17.
32 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
33 Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. *R J* 2015;7:19–33.
34 Mlecnik B, Van den Eynde M, Bindea G, *et al*. Comprehensive Intrametastatic immune quantification and major impact of immunoscore on survival. *J Natl Cancer Inst* 2018;110:97–108.
35 Kather JN, Suarez-Carmona M, Charoentong P, *et al*. Topography of cancer-associated immune cells in human solid tumors. *Elife* 2018;7. doi:10.7554/eLife.36967. [Epub ahead of print: 04 Sep 2018].
36 Tibshirani R. Regression shrinkage and selection via the LASSO. *J R Stat Soc Series B Stat Methodol* 1996;58:267–88.
37 Simon N, Friedman J, Hastie T, *et al*. Regularization paths for COX's proportional hazards model via coordinate descent. *J Stat Softw* 2011;39:1–13.
38 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1–22.
39 Grothey A, Sobrero AF, Shields AF, *et al*. Duration of adjuvant chemotherapy for stage III colon cancer. *N Engl J Med* 2018;378:1177–88.
40 Huijbers A, Tollenaar RAEM, v Pelt GW, *et al*. The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the Victor trial. *Ann Oncol* 2013;24:179–85.