Original Research

Prospective validation of a lymphocyte infiltration prognostic test in stage III colon cancer patients treated with adjuvant FOLFOX*

Jean-François Emile a,b,*, Catherine Julié a,b, Karine Le Malicot c, Come Lepage c,d,e, Josep Tabernerof, Enrico Mini g, Gunnar Folprecht h, Jean-Luc Van Laethem i, Stéphanie Dimet j, Camille Bouagnon-Rombi k, Marc-Antoine Allard a, Frédérique Penault-Llorca l, Jaafar Bennounam m, Pierre Laurent-Puig n,o,p, Julien Taieb n,q, the PETACC8 Study Investigators1

a EA4340, Versailles University, Paris-Saclay University, Boulogne, France
b Assistance Publique Hôpitaux de Paris, Hopital Ambroise Paré, Boulogne, France
c Fédération Francophone de Cancérologie Digestive (FFCD), Dijon, France
d Department of Hepato-Gastroenterology, Dijon University Hospital, Dijon, France
e UMR 866, INSERM, Dijon, France
f Vall D’Hebron University Hospital and Institute of Oncology (VHIO), Universitat Autònoma de Barcelona, Barcelona, Spain
g Section of Internal Medicine, Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy
h Medical Department I, University Hospital Carl Gustav Carus, Dresden, Germany
i Department of Gastroenterology, Erasme Hospital University, Brussels, Belgium
j Department of Pathology, Hospital of La Roche sur Yon, La Roche sur Yon, France
k Department of Pathology, University Hospital of Reims, Reims, France
l Department of Pathology, Centre Jean Perrin, Clermont-Ferrand, France
m Department of Gastroenterology, Centre Rene Gauducheau, Saint-Herblain, France
n Université Paris Descartes, Sorbonne Paris Cité, Paris, France
o Assistance Publique Hôpitaux de Paris, Department of Biology, Hôpital Européen Georges Pompidou, Paris, France
p INSERM UMR-S1147 Centre Universitaire des Saints Pères, Paris, France
q Paris Descartes University, Paris Sorbonne cité, Department of Digestive Oncology, HEGP, Paris, France

Received 27 January 2017; received in revised form 25 April 2017; accepted 29 April 2017
Available online 23 June 2017

* This study was presented in part at the Gastrointestinal (GI) meeting of the American Society of Clinical oncology (ASCO) in January 2016 (San Francisco, CA, USA), and at the Journées Francophones d’Hépatologie et Oncologie Digestive (JFHOD) in March 2016 (Paris, France).

* Corresponding author: EA4340 & Pathology Department, Ambroise Paré Hospital, 9 Av. Ch. de Gaulle, 92104 Boulogne, France.
E-mail address: jean-francois.emile@uvsq.fr (J.-F. Emile).

1 Please see the Supplementary Appendix for a list of the Investigators/Collaborators.
KEYWORDS
Colorectal carcinoma; Immune response; Biomarker; Prospective cohort study

Abstract  Background: The prognostic value of lymphocyte infiltration (LI) of colorectal carcinoma (CC) has been demonstrated by several groups. However, no validated test is currently available for clinical practice. We previously described an automated and reproducible method for testing LI and aimed to validate it for clinical use.

Patients and methods: According to National Institutes of Health criteria, we designed a prospective validation of this biomarker in patients included in the PETACC8 phase III study. Primary objective was to compare percentage of patients alive and without recurrence at 2 years in patients with high versus low LI (#NCT02364024). Associations of LI with patient recurrence and survival were analysed, and multivariable models were adjusted for treatment and relevant factors. Automated testing of LI was performed on virtual slides without access to clinical data.

Results: Among the 1220 CC patients enrolled, LI was high, low and not evaluable in 241 (19.8%), 790 (64.8%) and 189 (15.5%), respectively. Primary objective was met with a 2-year recurrence rate of 14.4% versus 21.1% in patients with high and low LI, respectively ($p = 0.02$). Patients with high LI also had better disease free survival (DFS) and overall survival (OS). Tumour stage, grade, RAS status and BRAF status were with LI the only prognostic markers in multivariable analysis for OS. Subgroup analyses revealed that high LI had better DFS and OS in mismatch repair (MMR) proficient patients, and in patients without RAS mutation, but not in MMR deficient and RAS mutated patients.

Conclusion: Although this is the first validation with high level of evidence (IIB) of the prognostic value of a LI test in colon cancers, it still needs to be confirmed in independent series of colon cancer patients.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Colorectal carcinoma (CC) is the 4th cause of cancer death worldwide [1]. After surgical resection of stage III colon carcinoma, up to 50% of patients develop recurrence and die from metastatic disease [2]. In 2004, adjuvant treatment with fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) was shown to be better than fluorouracil and leucovorin alone [3], and oxaliplatin benefit was confirmed in other trials [4,5]. FOLFOX4 is thus currently the standard adjuvant treatment for stage III patients and probably improves long term disease free survival (DFS) by 10–15% as compared to surgery alone [5,6]. The most recent clinical trials for stage III CC patients, investigated either intensifying adjuvant treatment by adding targeted therapies to FOLFOX [7–9] or shortening treatment duration to decrease adjuvant treatment toxicities (clinicaltrials.gov #NCT00958737). Obviously a better prognostic assessment of these patients using validated biomarkers, as well as biomarkers predictive for efficacy of a therapy, would help to select patients for either more or less intensive adjuvant therapy.

In 1986, lymphocyte infiltration (LI) of rectal carcinoma was reported to be associated with a better prognosis [10]. More than 300 publications investigating CC have confirmed this seminal publication. Lymphocytes were counted on haematoxylin and eosin or immunohistochemistry stained slides, manually or with image analysis software, on whole slides or on tissue arrays [11–14]. However despite all these publications and the need of biomarkers, no test has been validated today for clinical use. The lack of such test is due to a technical reason: immunostaining of lymphocytes is highly dependent on preanalytic conditions. It is also due to a biological reason: the density of lymphocytes within a given CC may vary from 1 to 50, depending on a biological reason: the density of lymphocytes within a given CC may vary from 1 to 50, depending on a biological reason: the density of lymphocytes within a given CC may vary from 1 to 50, depending on a biological reason: the density of lymphocytes within a given CC may vary from 1 to 50, depending on the counted areas [15]. Thus in all the previously published series, the cutoff of lymphocyte density to conclude the result as high or low was determined after analysing the whole retrospective series of patients.

We previously described a robust and reproducible test evaluating LI in CC [15]. This test is based on automated counting of CD3 lymphocytes within hundreds of small areas localised in both parts of the tumour margin. Each lymphocyte density is associated with the distance of the analysed area from the tumour margin, and used to generate a curve of variation of lymphocyte densities according to the distance from tumour margin. The local immune response (IR) to a given tumour is then interpreted on the curve, and does not depend on the preanalytic conditions.

The present work was designed to validate this test with a high level of evidence in a large pan-European prospective study of patients with stage III CC treated with adjuvant FOLFOX4 plus or minus cetuximab.

2. Materials and methods

2.1. Patient

Study design has been published on clinicaltrials.gov web site in February 2015 (#NCT02364024). Patients from...
the PETACC8 clinical trial that had signed an informed consent for translational research were enrolled. The PETACC8 trial tested FOLFOX ± cetuximab for 6 months versus FOLFOX [9].

Mismatched repair (MMR) status, mutations within exon 2, 3 and 4 of KRAS and NRAS, and mutations within exon 15 of BRAF of tumours were determined as previously described [16,17].

2.2. LI status determination

Assessment of LI status (i.e. high versus low) was performed as previously described [15]. For each patient, one tumour slide was stained using Bond-Max Fr4.0 (Leica Biosystem, Newcastle, UK) with CD3 primary antibody (clone F7.2.38, Dako France). Virtual slides of format either mrx or ndpi were then analysed using the Visilog 7.0 software (FIE, Saint-Aubin, France). The areas analysed were rectangles (1 × 4 mm). Lymphocyte cell densities were determined every 5 μm. For each patient, one to nine rectangles were analysed, corresponding to 800–7200 counts of lymphocyte densities; each count being associated with its position to the tumour invasive front. Data obtained from each 4 mm [2] rectangle were used to generate a curve of the variation of LI densities with distance from tumour margins. Tumours were then classified according to the pattern of these curves as previously described [15], in four groups (Supplementary Fig. 1). High and low IR were defined prior initiation of the study.

2.3. Statistical analyses

The primary end-point was the percentage of patients alive and without any recurrence at 2 years. This time to recurrence (TTR) was defined as the time between randomisation and the occurrence of local or metastatic recurrence or death linked to disease progression (whichever occurred first). Patients without any event were censored at the date of last news. Overall survival (OS) and DFS were defined as the time between randomisation and event. Patients without event were censored at the date last known to be alive. As DFS and OS were not different in both arms of the PETACC8 trial [9], patients were analysed together for the prognostic value of LI. We tested the interaction between the LI score and treatment for the end-points, and found no statistical interaction.

As no pertinent data were available to construct the primary hypotheses of the present prospective study, we randomly selected a subset of 300 patients from PETACC8 without knowledge of clinical data. Based on the results of this sample, we expected a difference of 13% in 2-year TTR between the 2 groups of patients in favour of high IR patients (76% versus 89%) in low versus high IR patients, respectively. According to the method of comparison of two proportions by a binomial distribution, with a power of 85% and an alpha risk of 5%, the number of patients to be included was 167 in each arm. The frequency of patients with high IR was 19.5% in the exploration set of patients. Thus at least 856 patients should be included in the study. In order to study outcome regarding IR in molecular subgroups (KRAS, BRAF or MMR status) previously analysed in PETACC8 series [16,18], we finally decided to include 1220 patients.

For comparisons of baseline characteristics, categorical outcomes were analysed with χ^2 tests, and the primary outcome and continuous outcomes were compared with standard parametric or non-parametric tests. Continuous variables are presented as the mean and median inter-quartile range.

TTR, DFS and OS curves were estimated with the Kaplan–Meier method and Cox models was used to estimate the hazard ratio (HR) and corresponding 95% confidence intervals (CIs). Factors included in the multivariable analyses were the treatment group, imbalanced baseline variables and prognostic factors identified in unadjusted (or univariate) analyses (p < 0.05) or already known to be clinically relevant.

Analyses were carried out according with a two-sided significance level of 5%. Results were uncorrected for multiple comparisons. All statistical analyses were done with the SAS statistical software package (version 9.4).

3. Results

Lymphocyte infiltration was quantified in 1220 patients corresponding to 2/3 of the patients of the PETACC8 study that were potentially available for the present study (Supplementary Fig. 2). The selected population was representative of the whole PETACC8 population (Supplementary Table 1).

3.1. Assessment of lymphocyte infiltration

As already shown in a retrospective series [15], the density of LI in a single tumour was highly variable, and the main parameter of variation was the distance from tumour invasive front. As decided before starting the study (clinicaltrials.gov #NCT02364024), patients with pattern 1 and 2 were considered as low IR, while those with pattern 3 and 4 as high IR.

One hundred eighty-nine patients (15%) could not be classified, either because the curve of density did not fit with the defined patterns (n = 123, 10%), or for technical failure (n = 66, 5%) such as absence of tumour/non-tumour interface or bad fixative conditions. Unclassified patients did not differ from other patients, except for pT status (Supplementary Table 2).

Among the 1031 classified patients, 241 (23%) had high and 790 (77%) had low LI. Patients with high and low IR had similar characteristics at time of surgery
(Supplementary Table 3), except for MMR status. As expected [19] deficient MMR (dMMR) tumours had a higher IR than those with proficient MMR (pMMR). Indeed 31 of the 98 (32%) dMMR cases had a high LI, while only 210 of the 933 (22%) pMMR cases had a high LI ($p = 0.04$). Patients also differed for treatment arm randomised after surgery; with high IR in 104/506 (20.6%) versus 137/525 (26.1%) for arm with or without cetuximab, respectively ($p = 0.036$).

3.2. Immune response and outcome

Primary objective was met with a 2-year TTR of 14.4% versus 21.1% in patients with high and low IR, respectively ($p = 0.02$).

Moreover, patients with high IR had longer DFS than those with low IR (Fig. 1), with a 3-year DFS at 81% and 72%, respectively (HR 0.69, 95% CI [0.51–0.92], $p = 0.01$). High IR was also associated with a better 5-year OS (89% versus 80%, HR = 0.58, 95% CI [0.40–0.85], $p = 0.0048$).

3.3. Subgroup analysis

In the subgroup of pMMR, high IR was also associated with a better DFS (3-year DFS 80.3% versus 71.7%, HR 0.72%, respectively (HR 0.69, 95% CI [0.52–0.95], $p = 0.022$) and OS (5 years OS 88.7% versus 79.2%, HR 0.58, 95% CI [0.39–0.87], $p = 0.008$) (Fig. 1C and D). In dMMR subgroup patients with high IR also seemed to have a better DFS (HR 0.61 [0.23–1.64] $p = 0.32$) and OS (HR 0.61 [0.17–2.20] $p = 0.45$). We also checked the prognostic value of IR status in the subgroup of patients with and without RAS mutations. High IR was associated with a better 3-year DFS (86.2% versus 77.8%, HR 0.58, 95% CI [0.35, 0.95], $p = 0.027$) and 5-year OS (91.9% versus 84.4%, HR 0.51, 95% CI [0.27, 0.97], $p = 0.038$) in RAS WT patients (Fig. 2A and B). In patients with RAS mutated tumours, a trend for a better DFS was observed in patients with high IR regarding 3-year (76.6% versus 68.0%) and 5-year DFS (85.1% versus 75.0%).

Concerning BRAF status, high IR was prognostic in BRAF WT patients for both DFS (HR 0.73 (0.54–1.00), $p = 0.046$) and OS (HR 0.58 (0.38–0.88), $p = 0.009$) (Fig. 2C–D).

3.4. Multivariable analysis for OS

Prognostic markers detected by unadjusted (or univariate) analyses were similar to those previously published in a translational ancillary study of the PETACC8 population [20], and these markers were included in the multivariable analysis for OS (Table 1). Tumour stage, grade RAS status and BRAF status were with IR the only independent prognostic markers found in multivariable analysis. The HR (95% CI) and $p$ values were: $pT3$ versus $pT2$ 4.28 (1.1–17.5), $p = 0.04$, $pT4$ versus $pT2$ 8.31 (2.0–34.4), $p = 0.003$, $pN$ 2.55 (1.8–3.6), $p < 0.0001$, grade 1.67 (1.1–2.4), $p = 0.008$, RAS 2.17 (1.5–3.2), $p < 0.0001$, BRAF 2.02 (1.1–3.7) $p = 0.02$, and LI 0.6 (0.4–0.98), $p = 0.04$.

4. Discussion

The aim of this study was to prospectively validate a test evaluating the LI of colon cancer. The primary objective of percentage of patients alive and without any recurrence at 2 years (TTR) is achieved, with a lower recurrence at 2 years of patients with high immune response (IR). Secondary objectives, which consisted in demonstrating the prognostic value of this test, are also achieved with a better DFS and OS in patients with high IR.

To validate our test of LI with the high level of evidence required for clinical use (IIB), we design a prospective study according to international recommendations [20,21] and published the objectives, hypothesis and statistical methods before its initiation (clinicaltrials.com #NCT02364024). To our knowledge, despite the high number of publications on LI in solid tumours, this is the first validation with such level of evidence. The only divergence with the planned statistics was the inclusion of a higher number of patients than required, which was decided in order to give higher power to subgroup analyses.

The test used for our study is based on hundreds of automated counts of lymphocyte by image analysis on virtual whole slide stained by standard immunohistochemistry. Interpretation of all lymphocyte densities obtained for each tumour is based on the patterns of the curves of densities from 2 mm outside to 2 mm inside the tumour margin. For these reasons, the results are not dependent of the preanalytic conditions, and notably of tissue fixation. Indeed interpretation of the lymphocyte densities was not based on absolute numbers, but on variations within the slides (a slide is its own control). Furthermore we previously showed that the variations of lymphocyte densities within a tumour are mainly dependent on the distance from the tumour margin. Thus our expression of densities according to this distance attenuates this major disturbing phenomenon. Using this method, we were able to prospectively classify 84.5% of the tumours obtained from 237 different centres localised in nine countries. By contrast the Immunoscores presented by Galon failed to classify 31% of the patients, although performed by selected and specifically formed centres [22]. Most of the failures in our series were due to the absence of instructions provided to the pathologist during sample collection, and will easily be corrected in the future.

The work of the group of Galon and Pagès provided major insights for understanding the IR against CC. However, it is still not clear whether adding other lymphocyte marker to CD3 could improve the value of the immune scoring of CC. Their original immune score
Fig. 1. Survivals according to immune response score. **Whole population**; (A) disease free survival, (B) overall survival. **pMMR subgroup**; (C) disease free survival, (D) overall survival.
Fig. 2. Overall survivals according to immune response score in the RAS and BRAF subgroups. (A) RAS wild type, (B) RAS mutant, (C) BRAF wild type, (D) BRAF mutant.
combined CD3 and CD45RO that were quantified within tissue micro-arrays [23]. In 2009, they combined CD45R0 and CD8 [24], and more recently presented another Immunoscore combining CD3 and CD8 on whole slides [22]. Interestingly Salama et al. used simultaneously CD45RO, CD8 and FoxP3, and found that FoxP3, but not CD8 neither CD45RO, was an independent prognostic marker [14]. Finally Laghi et al. obtained convincing results with CD3 only [13].

The better prognosis of patient with stage III dMMR tumour treated with adjuvant FOLFOX has been reported in some series [6,25], but was not confirmed in recent studies [26,27]. As already reported by other groups [19,28], we detected a high LI in dMMR than in pMMR CC. However 22% of pMMR CC also had a high LI. This may have major therapeutic implications. Indeed, clinical benefit of treatments with checkpoint inhibitors seems to be restricted to patients whose tumour has a high LI, with clonal neoantigens [28]. For patients with CC, successful treatment with checkpoint inhibitors was recently reported in dMMR patients [29], and most ongoing phase II or III trials are limited to dMMR patients. We suspect that some patients with pMMR CC but high IR may also benefit from immunotherapies.

In patients with CC undergoing surgical resection, prognosis and management are usually based entirely on the tumour-node-metastasis classification, and validated prognostic biomarkers are needed to improve adjuvant strategies. A large number of retrospective studies have been published during the last 20 years, however very few biomarkers have been validated with a good level of evidence. Tumour budding is a promising biomarker [30] and might be related with lymphocyte infiltrate, but its prognostic value still remained to be validated. We validated in the present multicentric prospective study of stage III colon CC treated with adjuvant FOLFOX4 the prognostic value of a test evaluating LI. We also showed by multivariable analysis that this test was independent of already known prognostic factors, such as tumour stage (T and N) and grade, and \( K_{NRA S} \) and \( B R A F \) status. The growth of cancers does not only depend on intrinsic characteristics of the tumours (mainly depending on genetic and epigenetic alterations), but also on immune host response. For oncology daily practice, it would obviously be more efficient to use a score combining markers of both host response and tumour molecular alterations. The optimal combination of biomarkers to determine the prognosis of localised CC treated by adjuvant FOLFOX

Table 1
Multivariable analysis for OS.

<table>
<thead>
<tr>
<th>N = 744 (deaths = 151)</th>
<th>HR</th>
<th>Intervalle de Confiance à 95% (95% confidence interval)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immune response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High IR versus low IR</td>
<td>0.62</td>
<td>0.390</td>
<td>0.976</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOLFOX + cetuximab versus FOLFOX</td>
<td>1.288</td>
<td>0.930</td>
<td>1.784</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female versus male</td>
<td>0.884</td>
<td>0.635</td>
<td>1.232</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;70 years versus age ≤ 70 years</td>
<td>0.948</td>
<td>0.564</td>
<td>1.594</td>
</tr>
<tr>
<td>Histopathological grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3–G4 versus G1–G2</td>
<td>1.668</td>
<td>1.144</td>
<td>2.432</td>
</tr>
<tr>
<td>Tumour localisation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right localisation versus left localisation</td>
<td>0.982</td>
<td>0.685</td>
<td>1.408</td>
</tr>
<tr>
<td>pT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT3 versus pT2</td>
<td>4.285</td>
<td>1.051</td>
<td>17.462</td>
</tr>
<tr>
<td>pT4 versus pT2</td>
<td>8.310</td>
<td>2.006</td>
<td>34.430</td>
</tr>
<tr>
<td>pN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pN2 versus pN1</td>
<td>2.548</td>
<td>1.807</td>
<td>3.592</td>
</tr>
<tr>
<td>Mutation RAS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutated versus wild type</td>
<td>2.165</td>
<td>1.469</td>
<td>3.192</td>
</tr>
<tr>
<td>Mutation BRAF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutated versus wild type</td>
<td>2.016</td>
<td>1.098</td>
<td>3.702</td>
</tr>
<tr>
<td>WHO performance status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–2 versus 0</td>
<td>1.367</td>
<td>0.937</td>
<td>1.996</td>
</tr>
<tr>
<td>Bowel obstruction and perforation</td>
<td>0.828</td>
<td>0.560</td>
<td>1.224</td>
</tr>
<tr>
<td>Vascular/lymphatic infiltration</td>
<td>0.948</td>
<td>0.657</td>
<td>1.367</td>
</tr>
<tr>
<td>MMR status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dMMR versus pMMR</td>
<td>0.645</td>
<td>0.331</td>
<td>1.258</td>
</tr>
</tbody>
</table>

dMMR, deficient MMR; pMMR, proficient MMR.
is currently under evaluation. This validated immune test should also be used for stratifying patients in the future CC adjuvant trials.

**Funding**

This study was supported by Région Ile de France, Fédération Francophone de Cancérologie Digestive (FFCD) and Association pour la Recherche et l’Enseignement en Pathologie (AREP). The PETACC8 study was sponsored by the FFCD that was responsible for the study management. Merck KGaA and Sanofi-Aventis supported the PETACC8 study: Merck provided the study cetuximab and financial support for study management; Sanofi-Aventis provided financial support for the provision of oxaliplatin to Belgian sites when necessary.

**Conflict of interest statement**

J.F. Emile received honoraria from Amgen and Merck Serono. J. Tabernero received honoraria Consultant/Advisory role from Amgen, Bayer, Boehringer Ingelheim, Celgene, Chugai, Lilly, MSD, Merck Serono, Novartis, Roche, Sanofi, Symphogen, Takeda and Taiho. G. Folprecht declared research funding from Merck KGaA and honoraria from Merck KGaA, Roche, Lilly, BMS and Amgen; P. Laurent-Puig declared providing advisory roles and lectures for Sanofi, Merck Serono, Amgen, Roche, Genomic Health, Myriad Genetics, Integragen and Pfizer. J. Taieb received honoraria from Sanofi and Merck KGaA. All remaining authors have declared no conflicts of interest.

**Acknowledgements**

The authors thank M Bakari, G Faucher, C Le Gall, S Surel and N Terrones for excellent technical pathology work, and M Schneider for her managing task in the PETACC8 trial. The authors also want to thank all the investigators of PETACC8 (list is provided in supplementary data). Finally, the authors thank all the patients who participated to the PETACC8 study and consented for the translational studies.

**Appendix A. Supplementary data**

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejca.2017.04.025.

**References**


