The Journal of Pathology: Clinical Research
J Path: Clin Res July 2015; **1**: 160–172
Published online 10 March 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cjp2.17

Original Article

A 12-gene signature to distinguish colon cancer patients with better clinical outcome following treatment with 5-fluorouracil or FOLFIRI

Eric R. Paquet,[1,2] Jing Cui,[3] David Davidson,[4] Natalia Pietrosemoli,[3] Houssein Hajj Hassan,[5] Serges P. Tsofack,[1] Annie Maltais,[1] Michael T. Hallett,[2] Mauro Delorenzi,[3,6] Gerald Batist,[4] Raquel Aloyz[4] and Michel Lebel[7]*

[1] *Molecular and Cellular Biology division, Centre de Recherche sur le Cancer de l'Université Laval, Quebec City, Québec, Canada*
[2] *Department of Biochemistry, Centre for Bioinformatics, McGill University, The Rosalind and Morris Goodman Cancer Research Centre, Montreal, Quebec, Canada*
[3] *Swiss Institute of Bioinformatics, Bioinformatics Core Facility, Lausanne, Switzerland*
[4] *Division of Experimental Medicine & Department of Oncology, McGill University, Segal Cancer Centre, Jewish General Hospital, Montreal, Quebec, Canada*
[5] *Biological Sciences Department, School of Arts and Sciences, Lebanese International University, Beirut, Lebanon*
[6] *Department of Oncology, University of Lausanne, Ludwig Center for Cancer Research, Lausanne, Switzerland*
[7] *Department of molecular biology, medical biochemistry and pathology, Laval University, Centre de Recherche du CHU de Québec, Centre Hospitalier de l'Université Laval, Québec, Canada*

**Correspondence to: Michel Lebel, Centre de Recherche du CHU de Québec, Centre Hospitalier de l'Université Laval (CHUL), 2705, boul. Laurier, Local T-2- 04, Québec, Québec G1V 4G2 Canada. e-mail: michel.lebel@crchudequebec.ulaval.ca*

Abstract

Currently, there is no marker in use in the clinical management of colon cancer to predict which patients will respond efficiently to 5-fluorouracil (5-FU), a common component of all cytotoxic therapies. Our aim was to develop and validate a multigene signature associated with clinical outcome from 5-FU therapy and to determine if it could be used to identify patients who might respond better to alternate treatments. Using a panel of 5-FU resistant and sensitive colon cancer cell lines, we identified 103 differentially expressed genes providing us with a 5-FU response signature. We refined this signature using a clinically relevant DNA microarray-based dataset of 359 formalin-fixed and paraffin-embedded (FFPE) colon cancer samples. We then validated the final signature in an external independent DNA microarray-based dataset of 316 stage III FFPE samples from the PETACC-3 (Pan-European Trails in Alimentary Tract Cancers) clinical trial. Finally, using a drug sensitivity database of 658 cell lines, we generated a list of drugs that could sensitize 5-FU resistant patients using our signature. We confirmed using the PETACC-3 dataset that the overall survival of subjects responding well to 5-FU did not improve with the addition of irinotecan (FOLFIRI; two-sided log-rank test $p = 0.795$). Conversely, patients who responded poorly to 5-FU based on our 12-gene signature were associated with better survival on FOLFIRI therapy (one-sided log-rank test $p = 0.039$). This new multigene signature is readily applicable to FFPE samples and provides a new tool to help manage treatment in stage III colon cancer. It also provides the first evidence that a subgroup of colon cancer patients can respond better to FOLFIRI than 5-FU treatment alone.

Keywords: chemotherapy; adjuvant; colon neoplasms/drug therapy; colon neoplasms/genetics; colon neoplasms/mortality; gene expression profiling; gene expression regulation; neoplastic/genetics; prognosis; treatment outcome

Received 7 November 2014; accepted 3 February 2015

†The Quebec-Clinical Research Organization in Cancer consortium Q-CROC (www.qcroc.ca), financed by the Pfizer-FRSQ Innovation Award, supported this work. E.R.P. is supported by a Doctoral scholarship from the Canadian Institutes of Health Research. S.P.T. is a scholar of the Q-CROC. M.D. and J.C. were supported in part by the Swiss National Science Foundation (SNF; http://www.snf.ch/E/; Grant No. 320030_135421); Krebsforschung Schweiz (KFS; http://www.krebsforschung.ch/; Grant No. 02697-08-2010); and by the Fondation Medic.

Abbreviations: 5-FU, 5-fluorouracil; DNA, Deoxyribonucleic acid; ECOG, Eastern Cooperative Oncology Group; FOLFIRI, 5-FU and irinotecan; HR, Hazard ratio; FDR, False discovery rate; FFPE, formalin-fixed and paraffin-embedded; IC_{50}, 50% inhibition values; MSI, microsatellite instability; PETACC-3, Pan-European Trails in Alimentary Tract Cancers; RFS, Recurrence-free survival; ROC, Receiver operating characteristics

Introduction

Colon cancer is the third leading cause of cancer-related death in the Western world (National Cancer Institute; 2009). Depending on the tumour stage at diagnosis, 5-year survival ranges from 93% for stage I patients to 44% for stage IIIC patients [1]. For stage II (node-negative) and III (node-positive) colon cancers, the treatment consists of resection of the primary tumour, which may be followed by adjuvant chemotherapy consisting of a 5-fluorouracil (5-FU) plus leucovorin, a common component of all cytotoxic therapies for colon cancer [2–4]. Unfortunately, around 20% of stage II and 30% of stage III patients will develop recurrent disease within 5 years of treatment, suggesting that 5-FU/leucovorin is not sufficient to eradicate all tumour cells within micrometastases [2–5]. Therefore, the aim of this study was to find pharmacogenomic markers, integrated in the form of a multigene signature that will identify patients associated with a better clinical outcome on 5-FU treatment. This multigene signature should be appropriate for testing in tumour specimens arising from surgical resections that are traditionally formalin-fixed and paraffin-embedded (FFPE) [6] and should enable clinicians to propose more effective treatment options for patients.

Several studies have already reported markers, in the form of multigene signatures, associated with response to 5-FU treatments [7,8]. However, none of these signatures have been identified or validated in a large cohort of clinical samples, limiting their use in the clinic. Other markers, including microsatellite instability (MSI) or DNA mismatch repair (MMR), 18q allele retention, thymidylate synthetase (TYMS), dihydropyrimidine dehydrogenase (DPYD) and methylenetetrahydrofolate reductase (MTHFR), have been suggested to predict response to 5-FU in stage II and III disease but have not been unambiguously proven to be associated with better clinical outcome from 5-FU treatment in a way to alter patient management [9–14]. We present here the results from an integrative approach, combining DNA microarray data from colon cancer cell lines with publically available datasets from clinical FFPE samples, that identified a 12-gene signature associated with better clinical outcome from 5-FU treatment in colon cancer.

Table 1. Calculated IC_{50} values for six colon cancer cell lines treated with 5-FU

Cell line	IC_{50} (μM)
Colo201	1.6 ± 0.2
HT29	2.2 ± 0.1
T84	1.8 ± 0.1
LoVo	4.1 ± 0.5
HTB39	6.6 ± 1.2
SW620	29.0 ± 4.2

Methods

Development of the 12-gene signature

We performed DNA microarray experiments using 5-FU resistant and sensitive colon cancer cell lines (Supporting Information Methods). From the DNA microarray data, we identified a list of 103 genes differentially expressed (limma, twofold change, FDR < 0.01, see Supporting Information Table 1). This list of genes provided a preliminary 5-FU response signature. As the list of genes exhibiting expression changes with 5-FU response contains some artifacts due to the cell culture based model used, we refined the list of 103 genes using a large gene expression dataset of 359 stage II colon cancer patients obtained from FFPE samples generated by Almac Diagnostics [6] to obtain a more clinically relevant signature applicable to patient tumour specimens. The Supporting Information Table 2

Table 2. Characterization of the 12 genes in the signature

Symbol	Description	Centroid	Biological processes
Down-regulated in resistant cell lines			
IRF7	Interferon regulatory factor 7	−0.84980259	Regulation of immune response
B2M	Beta-2 microglobulin	−0.837457154	Regulation of immune response
STAT1	Signal transducer and activator of transcription 1, 91 kDa	−0.850830843	Regulation of immune response
PARP14	Poly (ADP-ribose) polymerase family, member 14	−0.801034398	DNA damage response
PHF15	PHD finger protein 15	−0.819475008	Chromatin remodeling
COTL1	Coactosin-like F-actin binding protein 1	−0.798936749	Regulation of immune response
Up-regulated in resistant cell lines			
C17orf76-AS1	Antisense RNA	0.872447873	Regulation of apopotosis
XPC	Xeroderma pigmentosum, complementation group C	0.755607674	Nucleptode excision repair
UPF3A	UPF3 regulator of nonsense transcripts homolog A (yeast)	0.830744235	Nonsense-mediated mRNA decay
CDC16	Cell division cycle 16	0.800990646	Cell cycle regulation
CCNB1IP1	Cyclin B1 interacting protein 1, E3 ubiquitin protein ligase	0.842978607	Cell cycle regulation
TACC1	Transforming, acidic coiled-coil containing protein 1	0.802256744	Cell cycle regulation

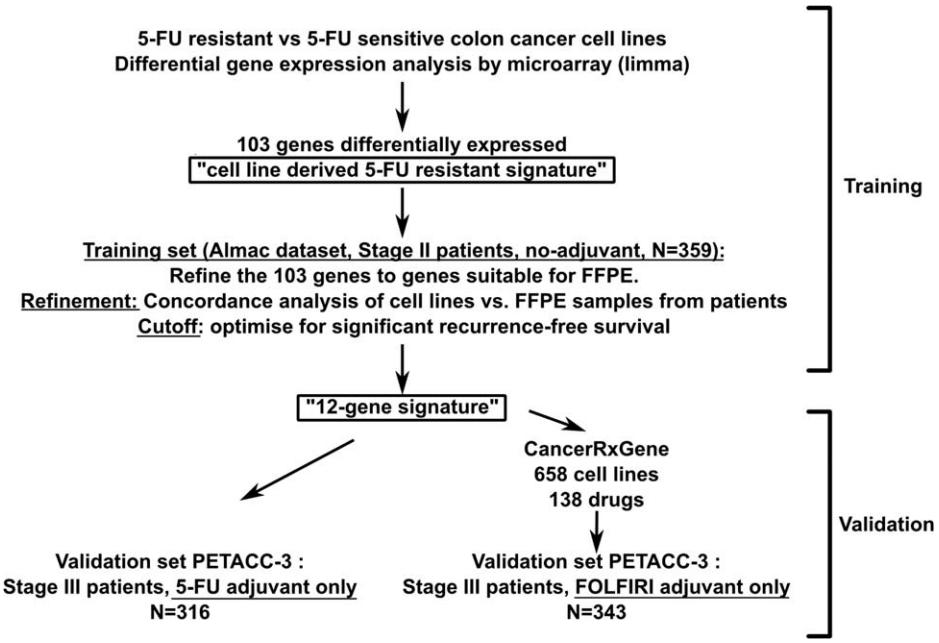

Figure 1. Schematic presentation of all the steps require for the definition of the final 12-gene signature.

provides characteristics of the Almac cohort. Importantly, no patient in this cohort received adjuvant chemotherapy. The refinement procedure consisted of keeping genes that behaved consistently in both the cell lines and FFPE datasets. We define consistency as the agreement between the directions of the modulated genes (over or under expressed) in our cell line derived signature and in the DNA microarray data from patients in a clinical data set. More precisely, we selected the genes in the signature so that genes simultaneously over and under expressed in the cell lines were also simultaneously over and under expressed in the patients DNA microarray data. A similar approach was recently used to refine a mouse-derived metastatic gene profile signature using a colon cancer dataset [15]. At the end of this procedure, we were left with 12 genes. Additional information regarding the generation of the 12-gene signature is available in the Supporting Information Methods.

To assign new colon cancer FFPE samples to the resistant or sensitive groups, we used a nearest centroid-based approach similar to what was used in several other multigene classifiers (e.g. ColoPrint [16]). Briefly, we used a defined 12-gene centroid using the mean of scaled values of the 5-FU resistant cancer cell lines. For every sample in a FFPE dataset, we computed Pearson's correlation to the 12-gene centroid to obtain a score ranging from -1 to 1. Before obtaining the optimal cutoff to split resistant and sensitive patients in the Almac dataset, we certified that the 12-gene signature score alone was associated with recurrence-free survival (RFS) and overall survival by univariate and multivariate Cox proportional hazard analysis. We then defined the cutoff score leading to the minimal univariate Cox p-value on RFS in the Almac dataset. Samples with a score <0.108 are considered sensitive to 5-FU and samples ≥ 0.108 are considered resistant. We confirmed that this final signature correctly classified 100% of the cell lines used in the first step of development of this signature in our original gene expression dataset as well as in another independent dataset from the CancerRxGene project (http://www.cancerrxgene.org/downloads/). A schematic presentation of all the steps required for the definition of the 12-gene signature is presented in Figure 1.

Analysis of the Almac dataset

We downloaded the DNA microarray data for the Almac dataset of tumour samples from 359 stage II patients with colon cancer who did not receive adjuvant chemotherapy directly from ArrayExpress using the accessions E-MTAB-863 and E-MTAB-864. We used the processed data provided by the Almac group in our analyses. This dataset is composed of 5014 probes and 359 samples generated on the Almac Affymetrix custom array. Supporting Information Table 2 provides characteristics of the Almac cohort. The expression values for the individual genes were

Table 3. Cox proportional hazards analysis for recurrence-free survival (RFS) in the Almac dataset using the 12-gene signature score (stage II, no-adjuvant, all lymph node negative)

	Univariate			Multivariate		
Parameter	HR	95% CI	*p*	HR	95% CI	*p*
Age	1.03	1.01 to 1.04	0.0050	1.03	1.01 to 1.05	0.0025
Grade						
2 vs 1	1.08	0.52 to 2.21	0.8409	1.07	0.52 to 2.24	0.8472
3 vs 1	1.66	0.74 to 3.69	0.2171	1.92	0.85 to 4.35	0.1167
Site: Proximal vs Distal	0.88	0.62 to 1.25	0.4635	0.84	0.58 to 1.23	0.3777
T-stage: T4 vs T3	1.82	1.21 to 2.73	0.0038	1.87	1.24 to 2.82	0.0028
Sex: Male vs Female	1.08	0.77 to 1.51	0.6638	1.08	0.77 to 1.53	0.6492
Number of examined lymph nodes	1.00	0.97 to 1.02	0.7005	0.99	0.97 to 1.02	0.5474
Mucinous subtype	0.91	0.59 to 1.41	0.6806	0.87	0.55 to 1.37	0.5413
12-gene signature score	1.99	1.35 to 2.94	0.0005	2.14	1.41 to 3.25	0.0003

first scaled and the average expression of duplicated genes was taken to obtain one vector of gene expression per gene symbol. This resulted in a final expression matrix containing 3050 genes and 359 samples. This final matrix was used for both the refinement procedure and to assign patients to the 5-FU resistant or sensitive groups using the 12-gene signature.

Analysis of the PETACC-3 dataset

We used data from stage III colon cancer patients from the Pan-European Trails in Alimentary Tract Cancers (PETACC-3) trial that were randomly allocated to one of two arms: (1) 5-FU only based chemotherapy (n = 316) or (2) FOLFIRI (5-FU + irinotecan) chemotherapy (n = 343; more information about this trial is available in Van Cutsem *et al*. [16]). The clinical information for patients in the PETACC-3 dataset is not publically available. Investigators of the PETACC-3 consequently performed the PETACC-3 portion of the analyses independently. Briefly, the PETACC-3 dataset (ArrayExpress E-MTAB-990), generated on the Almac Affymetrix custom chip (which is exactly the same chip used in the Almac dataset described above), was normalized using default rmaPLM function within the Bioconductor package affyPLM [17]. For each gene, the probeset with highest median absolute deviation was selected. The resulting gene expression matrix was scaled so that every single gene had a comparable level of expression. The final complete matrix composed of 659 patients was used to assign patients to the resistant and sensitive groups using the 12-gene signature (5-FU treated patients (n = 316) characteristics are in the Supporting Information Table 3; FOLFIRI treated patients (n = 343) characteristics are in the Supporting Information Table 4). Although the Almac and PETACC-3 datasets were generated using two different preprocessing methods, we tested whether this could have an impact on assignments made by our signature within the Almac dataset. We observed a 94% agreement when comparing the two different approaches. This observation suggests that the assignments made by our 12-gene signature is not dependent on the preprocessing methods. KRAS and BRAF mutations were detected using allele-specific real-time polymerase chain reaction [18]. The mutation status for all samples was confirmed using a second alternative technology (Sequenom) [19]. MSI status was determined using a panel of 10 mononucleotide and dinucleotide microsatellite loci by polymerase

Table 4. Cox proportional hazards analysis for overall survival in the Almac dataset using the 12-gene signature score (stage II, no-adjuvant, all lymph node negative)

	Univariate			Multivariate		
Parameter	HR	95% CI	*p*	HR	95% CI	*p*
Age	1.03	1.01 to 1.05	0.0014	1.04	1.01 to 1.06	0.0014
Grade						
2 vs 1	0.95	0.44 to 2.05	0.8921	0.85	0.39 to 1.87	0.6904
3 vs 1	1.65	0.70 to 3.90	0.2545	1.71	0.71 to 4.13	0.2349
Site: Proximal vs Distal	1.05	0.70 to 1.57	0.8191	1	0.65 to 1.55	0.9893
T-stage: T4 vs T3	1.7	1.06 to 2.71	0.0271	1.71	1.07 to 2.75	0.0263
Sex: Male vs Female	0.92	0.63 to 1.33	0.6449	0.91	0.62 to 1.35	0.6548
Number of examined lymph nodes	1	0.98 to 1.02	0.959	0.99	0.97 to 1.02	0.6137
Mucinous subtype	0.92	0.56 to 1.49	0.7235	0.93	0.56 to 1.55	0.783
12-gene signature score	2.16	1.39 to 3.34	0.0006	2.59	1.61 to 4.17	0.0001

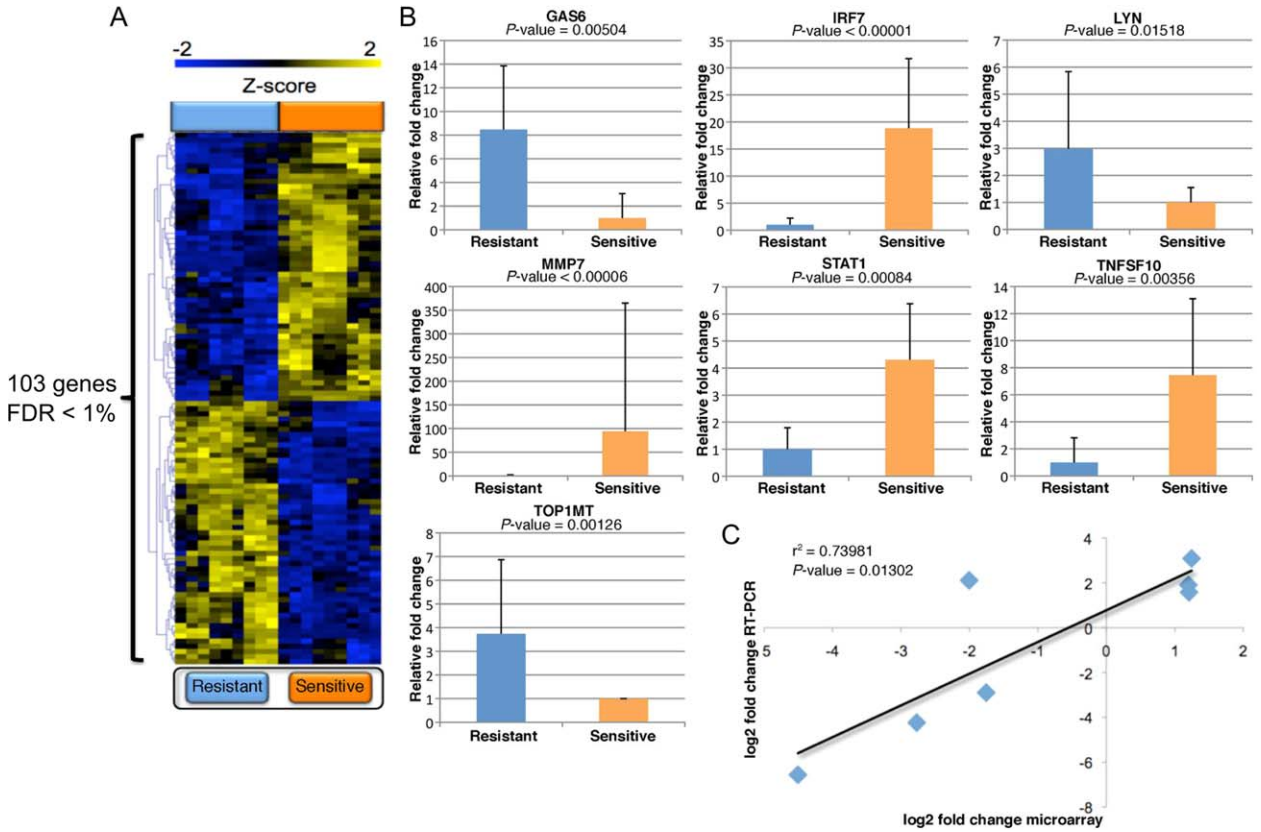

Figure 2. A: Supervised hierarchical clustering displaying the genes significantly modulated between the resistant and sensitive cell lines (52 under- and 51 over-expressed in resistant cell lines) using a cutoff of twofold change on expression and FDR < 0.01. B: Fold difference in gene expression between sensitive and resistant cell lines detected by qRT-PCR analyses. Seven genes were selected from the list of altered genes for this analysis. The results from sensitive or resistant cell lines were pooled to obtain an estimated fold change. C: Correlation between the qRT-PCR and the microarray expression data. The fold difference detected by qRT-PCR between sensitive and resistant cell lines correlated significantly confirming the microarray results (Pearson's correlation or $\rho = 0.739$, p-value $= 0.013$).

chain reaction amplification of normal/tumour DNA pairs. MSI was graded as high (MSI-H), with ≥ 3 unstable markers, low (MSI-L) with one to two unstable markers, and stable otherwise (MSS) [18].

Drug sensitizing analysis for patients resistant to 5-FU

Data from the CancerRxGene project version 2.0 were downloaded from the project website (http://www.cancerrxgene.org/downloads/). We used the preprocessed Affymetrix HG-U133A gene expression data for the 658 cell lines to assign the 5-FU resistant and sensitive cell lines. Briefly, single probe expressions were scaled for all the probes on the array. For multiple probes mapping to the same gene symbol, we took the mean of all the probes as the expression value for the gene symbol. We used the 12-gene signature to define the resistant and sensitive cell lines. To identify the drugs sensitizing predicted 5-FU resistant cell lines, we performed a Student's t-test on the $\log_{10}$ transformed IC_{50} for the individual drugs ($n = 130$) and compared the distribution of the sensitive and resistant $\log_{10}(IC_{50})$. We used the Benjamini–Hochberg corrected p-value from the Student's t-test to identify drugs that are significantly more toxic for resistant than for sensitive cells.

Statistical analysis

All Student's t, log-rank, univariate and multivariate Cox regression tests were performed in R version 2.14.0. All log-rank tests were performed using two-sided tests unless otherwise indicated in the figure legends. We used Wald's test for univariate and multivariate Cox regression analysis.

Table 5. Cox proportional hazards analysis for recurrence-free survival (RFS) in the PETACC-3 dataset (stage III, 5-FU adjuvant only, all lymph node positive)

	Univariate			Multivariate		
Parameter	HR	95% CI	*p*	HR	95% CI	*p*
Age	1.08	0.92 to 1.27	0.358	1.08	0.9 to 1.29	0.407
Grade: G-34 vs G-12	1.32	0.79 to 2.19	0.293	1.49	0.81 to 2.75	0.201
MSI: MSI-H vs MSS	0.43	0.19 to 0.97	0.043	0.48	0.2 to 1.17	0.108
N-stage: N2 vs N1	2.02	1.42 to 2.87	<0.001	2.08	1.42 to 3.05	<0.001
Sex: Female vs Male	0.8	0.56 to 1.17	0.251	0.85	0.57 to 1.26	0.417
Site: Right vs Left	0.92	0.64 to 1.33	0.666	1.35	0.89 to 2.07	0.161
T-stage						
T12 vs T3	0.49	0.22 to 1.12	0.092	0.49	0.21 to 1.16	0.104
T4 vs T3	1.64	1.05 to 2.56	0.031	1.7	1.06 to 2.73	0.028
BRAF: wt vs mut	1.16	0.54 to 2.5	0.697	1.14	0.48 to 2.69	0.763
KRAS: wt vs mut	0.79	0.55 to 1.14	0.208	0.76	0.51 to 1.13	0.181
Number of examined lymph nodes	0.81	0.65 to 1.01	0.058	0.75	0.59 to 0.95	0.017
12-gene signature: resistant vs sensitive	1.64	1.16 to 2.33	0.005	1.67	1.15 to 2.43	0.008

Results

Identification of a gene expression profile associated with resistance to 5-FU

To define a signature of resistance to 5-FU, we first performed DNA microarray analysis using mRNA extracted from colon cancer cell lines with varying levels of resistance to 5-FU. Dose response curves and a summary of the 50% inhibition values (IC_{50}) are presented in Table 1 and Supporting Information Figure 1. The difference in survival between resistant and sensitive cell lines was in the order of sevenfold. Bioinformatics analyses revealed that 103 genes (52 genes over and 51 under expressed) are significantly modulated in resistant cells (limma [20], > twofold change and false discovery rate (FDR) < 0.01, see Supporting Information Table 1). Hierarchical clustering of the genes significantly modulated between the resistant and sensitive cell lines is presented in Figure 2A. We validated the differential expression between resistant and sensitive cell lines for a selected panel of genes by quantitative real-time PCR and observed a good correlation between the fold change quantification and DNA microarray data with a Pearson's correlation or $\rho = 0.739$, p-value = 0.013 (Figure 2B,C, Supporting Information Figure 2, Supporting Information Table 5 for the primers used). These results confirmed the high quality of our microarray data.

Refinement to a 12-gene signature applicable on FFPE samples

Using the 103 genes associated with 5-FU resistance in the panel of colon cancer cell lines, we performed a refinement step to filter out genes not suitable for FFPE clinical specimens. Hence, we used a publically available gene expression DNA microarray dataset generated by Almac Diagnostics on 359 FFPE tissue specimens from stage II colon cancer that did not receive adjuvant chemotherapy. The Supporting Information Table 2 provides characteristics of the Almac cohort [6]. After refinement, a panel of 12 genes suitable for detection in clinical samples remained (Table 2). We then defined a 12-gene nearest centroid-based signature to assign a 12-gene signature score to every patient in the Almac dataset (see the centroid column in Table 2). We performed univariate and multivariate Cox regression analyses of the 12-gene signature score when adjusting for all available clinical variables: age, stage, grade, tumour location, sex, mucinous subtype and number of retrieved lymph nodes. We found that the 12-gene signature score is significantly associated with RFS in univariate analysis with a HR of 1.99 ($p = 0.0005$; Table 3) and in multivariate analysis with a HR of 2.14 ($p = 0.0003$; Table 3). The signature is also significantly associated with overall survival in univariate HR 2.6 ($p = 0.006$; Table 4) and multivariate HR 2.59 ($p = 0.0001$; Table 4). Thus, our statistical analyses of the Almac dataset indicate that our 12-gene signature score is associated with RFS and overall survival, independently of known prognostic factors and could be used as a prognostic tool in stage II colon cancer (Tables 3 and 4). Given that we were able to confirm that the 12-gene signature score alone is significantly associated with RFS and OS, we used this cohort to define a cutoff on the 12-gene signature score to classify patients as 5-FU resistant or 5-FU sensitive (see Supporting Information Materials and Methods).

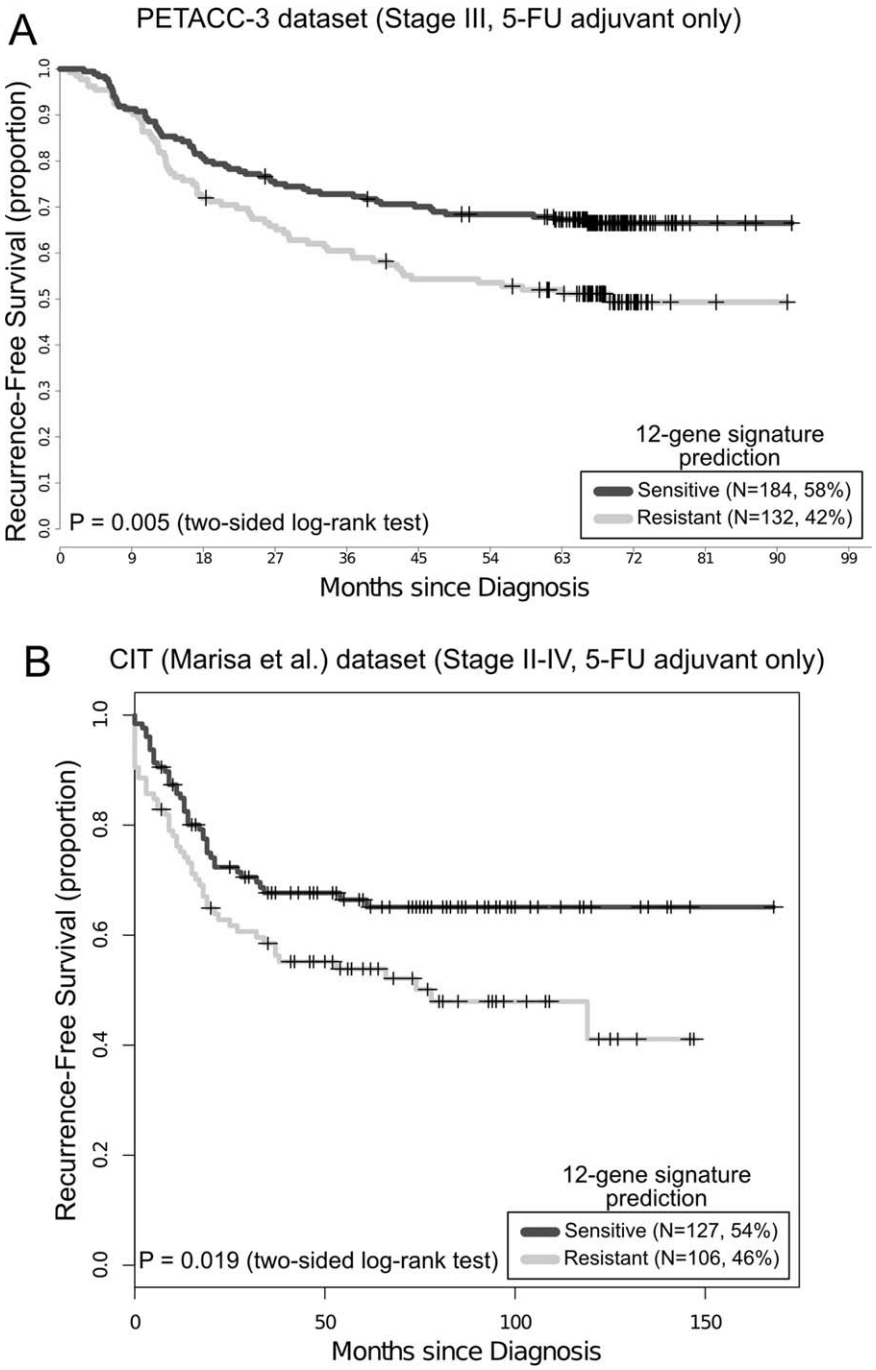

Figure 3. A: Kaplan–Meier curves displaying the difference in recurrence-free survival for patients predicted to be resistant or sensitive. All patients in this PETACC-3 dataset were treated with 5-FU. B: Kaplan–Meier curves displaying the difference in recurrence-free survival for stage II–IV patients predicted to be resistant or sensitive in the French national Cartes d'Identité des Tumeurs (CIT). All patients in the CIT dataset were treated with 5-FU.

Pathway analysis of the 12-gene signature

To identify the biological pathways captured by the 12-gene signature, we subjected the list of genes to gene ontology. Genes that are down regulated in the resistant patients are enriched for regulation of immune response while genes that are up regulated are enriched for cell cycle regulation (Table 2).

Table 6. Cox proportional hazards analysis for recurrence-free survival (RFS) in the CIT (Marisa *et al.*, reference [22]) dataset (stage II–IV, 5-FU adjuvant only)

	Univariate			Multivariate		
Parameter	HR	95% CI	*p*	HR	95% CI	*p*
Age	1.01	0.99 to 1.02	0.496	1.01	0.99 to 1.03	0.378
Stage						
3 vs 2	1.27	0.76 to 2.13	0.367	1.3	0.77 to 2.21	0.332
4 vs 2	3.35	1.65 to 6.79	0.001	3.37	1.66 to 6.84	0.001
Sex: Male vs Female	1.01	0.67 to 1.53	0.959	0.96	0.63 to 1.46	0.845
Site: Proximal vs Distal	0.68	0.43 to 1.08	0.104	0.7	0.43 to 1.13	0.143
12-gene signature: resistant vs sensitive	1.64	1.08 to 2.47	0.020	1.54	1.01 to 2.34	0.043

Association of the 12-gene signature with a better outcome from 5-FU in an independent cohort of colon cancer patients

We next assessed the predictive value of our signature in an independent cohort of 316 FFPE samples from stage III colon cancer patients treated only with 5-FU. We used the gene expression DNA microarray dataset generated by the PETACC (Pan European Trial Adjuvant Colon Cancer)-3 trial to test the ability of our signature to discriminate patients that have a better RFS following adjuvant treatment with 5-FU. The Supporting Information Table 3 provides characteristics of this PETACC-3 cohort [21]. Importantly, the bioinformatician in charge of developing the 12-gene signature using the Almac cohort (training set) did not have direct access to the PETACC-3 dataset (validation set). Consequently, investigators from the PETACC-3 group independently validated the 12-gene signature. All patients of the PETACC-3 cohort were assigned to either the resistant ($N = 132$) or sensitive ($N = 184$) groups and a significant difference in RFS between the two groups of patients was observed (Figure 3; two-sided log-rank test $p = 0.005$). By performing a multivariate analysis of the 12-gene signature stratification while adjusting for other clinical variables including MSI, BRAF and KRAS status, we found a significant association between the 12-gene signature and RFS in multivariate analysis with a HR of 1.67 ($p = 0.008$ in Table 5). A trend was similarly observed for overall survival (univariate HR = 1.4; $p = 0.087$, multivariate HR = 1.5; $p = 0.063$, Supporting Information Table 6). We validated these results in another independent dataset generated with RNA derived from fresh-frozen stage II, III and IV colon cancer tumours from patients enrolled in The French national Cartes d'Identité des Tumeurs (CIT) program [22]. All these patients were treated with 5-FU. We observed a significant difference between the resistant ($N = 106$) and sensitive ($N = 127$) defined patients using our 12-gene signature in the CIT dataset (Figure 3B; two-sided log-rank test $p = 0.019$). This result was also validated in a multivariate analysis with a HR of 1.54 ($p = 0.043$ in Table 6). This analysis, performed in the CIT and PETACC-3 dataset, confirmed the significant association between the 12-gene signature and better clinical outcome following 5-FU treatment.

A subgroup of patients with tumours resistant to 5-FU were associated with a better clinical outcome from chemotherapy containing a topoisomerase I inhibitor

To determine whether the 12-gene signature could identify patients with increased sensitivity to chemotherapies other than 5-FU, a recently generated dataset comprising the gene expression profiles of 658 cell lines and their IC_{50} values for 130 chemotherapeutic agents was analyzed [23]. First, using the 12-gene signature, the 658 cell lines were stratified into 5-FU resistant ($n = 327$) and sensitive cell lines ($n = 331$). By comparing the distribution of the IC_{50} values for the resistant and sensitive cell lines for all the 130 drugs, we identified a list of 18 drugs that are significantly more toxic for the 5-FU resistant cell lines (FDR < 0.05; Table 7). The list of drugs predicted to sensitize cell lines resistant to 5-FU contained the topoisomerase I inhibitor camptothecin, an analog of irinotecan, also used in a second arm of the PETACC-3 to treat a randomly selected group of patients in combination with 5-FU [24]. We, therefore, asked whether our 12-gene signature could distinguish between patients likely to exhibit a longer RFS on treatment with 5-FU and irinotecan (FOLFIRI) versus patients treated with 5-FU alone in this cohort. In the subgroup of patients, classified as 5-FU sensitive with our signature, we observed no significant difference in RFS when patients were treated with 5-FU alone or in combination with irinotecan (Figure 4A; log-rank test $p = 0.795$). In contrast, and perhaps more importantly, patients classified as

Table 7. Results from the cell line drug sensitivity analysis listing the drugs significantly sensitizing the cell lines predicted as resistant using our 12-gene signature compare to the cell lines predicted as sensitive to 5-FU

Drug name	Targets	Differential sensitization score	False discovery rate FDR (%)
Nutlin-3a	MDM2	0.81	0.00
Methotrexate	Dihydrofolate reductase (DHFR)	0.70	0.44
Axitinib	PDGFR, KIT, VEGFR	0.62	0.10
Vorinostat	HDAC inhibitor Class I, IIa, IIb, IV	0.55	0.00
OSI-906	IFG1R	0.53	1.14
CEP-701	FLT3, JAK2, NTRK1, RET	0.53	0.99
AZD-2281	PARP1/2	0.53	0.10
PD-173074	FGFR1/3	0.50	0.10
Camptothecin	TOP1	0.49	4.57
SB590885	BRAF	0.46	2.70
Cytarabine	DNA synthesis	0.45	4.42
681640	WEE1, CHK1	0.45	1.63
BAY-61-3606	SYK	0.42	3.84
Vinorelbine	Microtubules	0.42	4.57
BI-D1870	RSK1/2/3/5, PLK1, AURKB	0.39	2.70
TW-37	BCL-2, BCL-XL	0.39	1.63
Nilotinib	ABL	0.36	4.92
EHT-1864	Rac GTPases	0.31	4.57

resistant to 5-FU have a better RFS when treated with FOLFIRI than when treated with 5-FU alone (Figure 4B; one-sided log-rank test $p = 0.039$). A trend similarly observed for overall survival (Supporting Information Figure 3; resistant patients one-sided log-rank test $p = 0.063$). We also detect a statistical trend for the interaction between the chemotherapy regimen and our 12-gene signature using multivariate Cox regression (multivariate HR = 0.65; $p = 0.098$, Supporting Information Table 7). Together, these data indicate that patients predicted to be resistant to 5-FU were associated with a better clinical outcome from FOLFIRI therapy.

The fact that our signature is prognostic in the Almac dataset and associated with better outcome for resistant patients treated with FOLFIRI might suggest that the resistant patients represent a group of patients with deleterious tumour biology. To test this hypothesis, we compare our 12-gene signature with assignments obtained from the Almac signature in the PETACC-3 dataset. We were unable to find any association between the low-risk and high-risk assignments from Almac when compared with the assignments obtained from our signature (Fisher's exact test $p > 0.05$). This observation suggests that our signature captured a signal that is not associated with high-risk patients. In addition, we also examined whether our signature is associated with a worse prognosis [25]. We compared our 12-gene signature with well-known markers of proliferation like MKI67, AURKA and the proliferation score of Parker *et al*. [26] (Supporting Information Figure 6). We were unable to find an association of our signature with markers of proliferation (all Wilcoxon's test $p > 0.05$ and correlation $p > 0.05$). Taken together, these analyses suggest that our signature captures a signal that is not associated with the Almac signature or markers of proliferation.

Discussion

5-FU is the most widely used adjuvant anticancer drug in colon cancers, however, some patients will suffer from early recurrence, suggesting that some tumour cells were resistant to treatment (intrinsically or acquired early on during the treatment). There is a lack of knowledge regarding the underlying mechanism of 5-FU clinical response in colon adenocarcinoma. Thus, there is an immediate need for the identification of biomarkers to identify patients associated with a better clinical outcome on 5-FU treatment. In this study, we identified a cell line derived 5-FU response signature that was applicable on FFPE clinical samples. This 12-gene signature was significantly associated with RFS in early-stage colon cancer patients of the Almac group. More importantly, we found the same signature to be associated with RFS for patients treated with 5-FU in the PETACC-3 trial. These findings suggest that our 12-gene signature is associated with a better outcome from 5-FU in early-stage colon cancer patients. Finally, we demonstrated that a subset of patients classified as resistant to 5-FU using our 12-gene signature were associated with improved survival from the addition of irinotecan to their adjuvant treatment, thereby identifying a sub-population of patients more responsive to this treatment combination. This is a novel finding as a

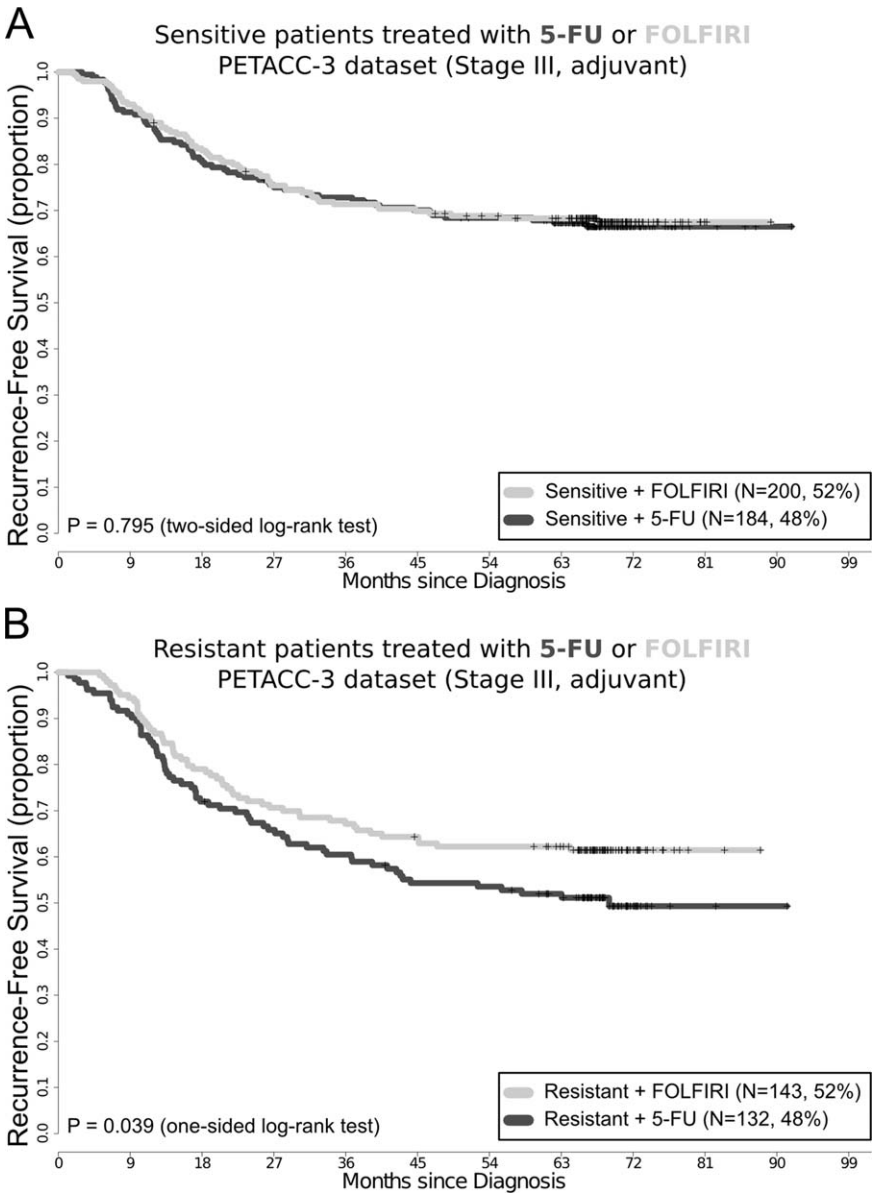

Figure 4. A: Kaplan–Meier curves displaying the difference in recurrence-free survival between patients predicted sensitive to 5-FU treated with 5-FU or FOLFIRI in the PETACC-3 dataset. B: Kaplan–Meier curves displaying the difference in recurrence-free survival between patients predicted resistant to 5-FU treated with 5-FU or FOLFIRI in the PETACC-3 dataset.

previous study from the PETACC-3 indicated that FOLFIRI did not confer a statistically significantly improvement in disease-free or overall survival compared to 5-FU alone [21]. However, it was suggested that a subgroup of patients could still benefit more from FOLFIRI and the application of our 12-gene signature supports this model [27]. Interestingly, our signature also indentifies a group of patients more sensitive to 5-FU that could be spared irinotecan treatment.

Several predictive biomarkers of response to 5-FU were proposed in the context of early-stage colon cancer. For example, the predictive value of MSI or DNA MMR for 5-FU treatment was studied extensively, but it is still a controversial issue and it is not clear how it could alter patient management particularly in stage III colon cancer [11,14]. Some studies suggested that MSI-High patients benefit from 5-FU [28] while other studies indicate that they do not benefit or that treatment with 5-FU is even detrimental [29,30]. Given the importance of MSI in colon cancer, we studied the interaction between our 12-gene signature and MSI in the PETACC-3 using the multivariate analysis presented in Table 5. We found that

our 12-gene signature has prognostic capability that is independent of MSI status as the hazard ratio of the 12-gene signature is still significant when we adjusted for MSI or for any other prognostic clinical variables. This suggests that the 12-gene signature is an independent marker allowing us to identify patients associated with a better outcome from 5-FU treatment.

There are some limitations to our results. First, all the analyses presented herein are derived from clinical specimens that have been obtained retrospectively. Clearly, a prospective study is needed to confirm the value of our 12-gene signature, its analytical validity, and reproducibility. Second, we were limited in our ability to analyze our hypothesis of putative sensitizing drugs for 5-FU resistant patients. Only camptothecin was selected because we had access to a dataset to test the hypothesis (i.e. the FOLFIRI arm of the PETACC-3 trial). Other drugs like methotrexate have the potential to also be effective for treating 5-FU resistant patients. Interestingly, one meta-analysis study indicated an improvement of response rate in colon cancer patients when 5-FU was combined with methotrexate [31]. We suspect that our 12-gene signature could identify patients showing significant clinical improvement from such a combination therapy. Third, regarding the statistical trend observed for the overall survival within the PETACC-3 cohort ($p = 0.063$, in Supporting Information Figure 3A,B), we believe this could possibly be explained by the fact that in the PETACC-3 trial, the treatment after relapse was not recorded and standardized which can be a confounding parameter for overall survival. Nevertheless, in view of these interesting and concordant results, a prospective study is commendable to confirm the value of our 12-gene signature. Fourth, with the recent publication of several new gene signatures in colon cancer, our 12-gene signature will need to be evaluated in comparison with those new signatures. Interestingly, recurrent biological pathways can be found in these different signatures such as changes in genes of the immune response and cell cycle progression [6,16,22,32–35] and we believe our 12-gene signature might capture similar signals. Fifth, the application of our 12-gene signature currently requires an entire dataset to estimate the scaling factors. Thus, our signature is not readily applicable on single samples from patients. However, our group has recently developed an approach enabling the translation of a dataset based signature to a signature applicable on raw gene expression profile from one patient [36]. We are in the process of performing the adaptation of our 12-gene signature using this approach. Finally, this study was not designed to answer if the predicted resistant or sensitive patients would improve with a treatment containing a combination of oxaliplatin and 5-FU (FOLFOX). This is due to the fact that oxaliplatin was not included in the publically available dataset that we used in our analysis (Almac, PETACC-3, and CancerRxGene) and a sufficiently large study including both DNA microarray data and clinical outcome is not available for FOLFOX treated patients. For example, large clinical trials like NSABP and MOSAIC does not usually provide DNA microarray gene expression data that are necessary to validate our 12-gene signature in those cohorts. Given FOLFOX is widely use for the treatment of early-stage colon cancer [37,38], this question will need to be addressed in a future study.

In this current era of personalized and precision medicine, biomarkers enabling the selection of appropriate treatment for patients are greatly needed. The 12-gene signature presented in this work identified a group of patients responding better to FOLFIRI than 5-FU alone. Given that this signature was validated in FFPE tumour specimens and contains only 12 genes, it has the potential to be rapidly translated into the clinic.

Acknowledgements

We thank J.-Y. Masson (Centre de Recherche en Cancérologie de l'Université Laval, Quebec City Canada) for technical advice and all the clinicians who enrolled patients and participated in the PETACC-3 trial (see Appendix at: http://jco.ascopubs.org/content/27/19/3117.long), in particular the coordinators D. Cunningham, R. Labianca and E. Van Cutsem. We are grateful to C. Rousseau for editing the manuscript. The Quebec-Clinical Research Organization in Cancer consortium Q-CROC (www.qcroc.ca), financed by the Pfizer-FRSQ Innovation Award, supported this work. E.R.P. is supported by a Doctoral scholarship from the Canadian Institutes of Health Research. S.P.T. is a scholar of the Q-CROC. M.D. and J.C. were supported in part by the Swiss National Science Foundation (SNF; http://www.snf.ch/E/; Grant No. 320030_135421); Krebsforschung Schweiz (KFS; http://www.krebsforschung.ch/; Grant No. 02697-08-2010) and by the Fondation Medic.

Author contributions

ERP, ML, RA and GB participated in the conception and design of the study. ML and RA provided the

study materials. ERP, JC, MD, DD, HHH, SPT, AM, RA and ML participated in the collection and assembly of the data. ERP, ML, JC, NP, MD, GB, RA and MTH participated in the analysis and interpretation of the data. The manuscript was drafted by ERP and ML and reviewed by all the authors. All the authors read and approved the final manuscript.

References

1. 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**: 1420–1425.
2. Wolmark N, Fisher B, Rockette H, *et al*. Postoperative adjuvant chemotherapy or BCG for colon cancer: results from NSABP protocol C-01. *J Natl Cancer Inst* 1988; **80**: 30–36.
3. Smith RE, Colangelo L, Wieand HS, *et al*. Randomized trial of adjuvant therapy in colon carcinoma: 10-year results of NSABP protocol C-01. *J Natl Cancer Inst* 2004; **96**: 1128–1132.
4. Wolmark N, Rockette H, Fisher B, *et al*. The benefit of leucovorin-modulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: results from National Surgical Adjuvant Breast and Bowel Project protocol C-03. *J Clin Oncol* 1993; **11**: 1879–1887.
5. Andre T, Boni C, Navarro M, *et al*. Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol* 2009; **27**: 3109–3116.
6. Kennedy RD, Bylesjo M, Kerr P, *et al*. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *J Clin Oncol* 2011; **29**: 4620–4626.
7. Mariadason JM, Arango D, Shi Q, *et al*. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. *Cancer Res* 2003; **63**: 8791–8812.
8. Boyer J, Allen WL, McLean EG, *et al*. Pharmacogenomic identification of novel determinants of response to chemotherapy in colon cancer. *Cancer Res* 2006; **66**: 2765–2777.
9. Watanabe T, Wu TT, Catalano PJ, *et al*. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med* 2001; **344**: 1196–1206.
10. Walther A, Johnstone E, Swanton C, *et al*. Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 2009; **9**: 489–499.
11. Vilar E, Tabernero J. Molecular dissection of microsatellite instable colorectal cancer. *Cancer Discov* 2013; **3**: 502–511.
12. Ezzeldin HH, Diasio RB. Predicting fluorouracil toxicity: can we finally do it? *J Clin Oncol* 2008; **26**: 2080–2082.
13. Tejpar S, Bertagnolli M, Bosman F, *et al*. Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery. *Oncologist* 2010; **15**: 390–404.
14. Sargent DJ, Shi Q, Yothers G, *et al*. Prognostic impact of deficient mismatch repair (dMMR) in 7,803 stage II/III colon cancer (CC) patients (pts): a pooled individual pt data analysis of 17 adjuvant trials in the ACCENT database. *ASCO Meeting Abstracts* 2014; **32**: 3507.
15. Smith JJ, Deane NG, Wu F, *et al*. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 2010; **138**: 958–968.
16. Salazar R, Roepman P, Capella G, *et al*. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *J Clin Oncol* 2011; **29**: 17–24.
17. Irizarry RA, Bolstad BM, Collin F, *et al*. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003; **31**: e15.
18. Roth AD, Tejpar S, Delorenzi M, *et al*. Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J Clin Oncol* 2010; **28**: 466–474.
19. De Roock W, Claes B, Bernasconi D, *et al*. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol* 2010; **11**: 753–762.
20. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; **3**: Article3.
21. Van Cutsem E, Labianca R, Bodoky G, *et al*. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol* 2009; **27**: 3117–3125.
22. Marisa L, de Reynies A, Duval A, *et al*. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013; **10**: e1001453.
23. Garnett MJ, Edelman EJ, Heidorn SJ, *et al*. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012; **483**: 570–575.
24. Masuda N, Fukuoka M, Kusunoki Y, *et al*. CPT-11: a new derivative of camptothecin for the treatment of refractory or relapsed small-cell lung cancer. *J Clin Oncol* 1992; **10**: 1225–1229.
25. Brown DC, Gatter KC. Ki67 protein: the immaculate deception? *Histopathology* 2002; **40**: 2–11.
26. Parker JS, Mullins M, Cheang MCU, *et al*. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009; **27**: 1160–1167.
27. Gasparini G, D'Andrea MR, Toffoli G. Irinotecan in the adjuvant treatment of colon cancer: is the story finished or does personalized therapy open new opportunities? *J Clin Oncol* 2010; **28**: e199.
28. Elsaleh H, Joseph D, Grieu F, *et al*. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet* 2000; **355**: 1745–1750.
29. Benatti P, Gafa R, Barana D, *et al*. Microsatellite instability and colorectal cancer prognosis. *Clin Cancer Res* 2005; **11**: 8332–8340.
30. Sargent DJ, Marsoni S, Monges G, *et al*. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J Clin Oncol* 2010; **28**: 3219–3226.

31. Meta-analysis of randomized trials testing the biochemical modulation of fluorouracil by methotrexate in metastatic colorectal cancer. Advanced Colorectal Cancer Meta-Analysis Project. *J Clin Oncol* 1994; **12**: 960–969.
32. Budinska E, Popovici V, Tejpar S, *et al*. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* 2013; **231**: 63–76.
33. Oh SC, Park YY, Park ES, *et al*. Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. *Gut* 2012; **61**: 1291–1298.
34. Sadanandam A, Lyssiotis CA, Homicsko K, *et al*. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* 2013; **19**: 619–625.
35. O'Connell MJ, Lavery I, Yothers G, *et al*. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. *J Clin Oncol* 2010; **28**: 3937–3944.
36. Paquet ER, Hallett MT. Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst* 2015; **107**: 357.
37. Andre T, Boni C, Mounedji-Boudiaf L, *et al*. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* 2004; **350**: 2343–2351.
38. Yothers G, O'Connell MJ, Allegra CJ, *et al*. Oxaliplatin as adjuvant therapy for colon cancer: updated results of NSABP C-07 trial, including survival and subset analyses. *J Clin Oncol* 2011; **29**: 3768–3774.

SUPPLEMENTARY MATERIAL ON THE INTERNET

Additional Supporting Information may be found in the online version of this article.

Table A1. Differential expression profile between resistant (LoVo, HTB39, SW620) and sensitive (Colo201, HT29, T84) cell lines.

Supplementary Table 2. Characteristics of the Almac Diagnostics dataset (Stage II, no adjuvant treatment).

Supplementary Table 3. Characteristics of the PETACC-3 dataset for Stage III patients that received 5-FU.

Supplementary Table 4. Characteristics of the PETACC-3 dataset for Stage III patients that received FOLFIRI.

Supplementary Table 5. Primers used for qRT-PCR.

Supplementary Table 6. Cox proportional hazards analysis for overall survival in the PETACC-3 dataset (Stage III, 5-FU adjuvant only, all lymph node positive).

Supplementary Table 7. Cox proportional hazards analysis for recurrence free survival reporting the interaction test between our 12-gene signature and treatment (Treatment x 12-gene signature) in the PETACC-3 dataset (Stage III, 5-FU adjuvant only, all lymph node positive).

