Abstract
The prognostic potential of individual clinical and molecular parameters in stage II/III colon cancer has been investigated, but a thorough multivariable assessment of their relative impact is missing.

Reference

DOI: 10.1093/jnci/djs427
PMID: 23104212

Available at:
http://archive-ouverte.unige.ch/unige:32264

Disclaimer: layout of this document may differ from the published version.
Integrated Analysis of Molecular and Clinical Prognostic Factors in Stage II/III Colon Cancer

Arnaud D. Roth, * Mauro Delorenzi, * Sabine Teijpar, * Pu Yan, Dirk Klingbiel, Roberto Fiocca, Giovanni d’Ario, Laura Cisar, Roberto Labianca, David Cunningham, Bernard Nordlinger, Fred Bosman, Eric Van Cutsem

Manuscript received December 13, 2011; revised September 5, 2012; accepted September 7, 2012.

*These authors contributed equally to this work.

Correspondence to: Arnaud Roth, MD, Oncosurgery Unit, Geneva University Hospital, Geneva, Switzerland (e-mail: arnaud.roth@hcuge.ch).

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please e-mail: journals.permissions@oup.com.

Integrated Analysis of Molecular and Clinical Prognostic Factors in Stage II/III Colon Cancer

Arnaud D. Roth, * Mauro Delorenzi, * Sabine Teijpar, * Pu Yan, Dirk Klingbiel, Roberto Fiocca, Giovanni d’Ario, Laura Cisar, Roberto Labianca, David Cunningham, Bernard Nordlinger, Fred Bosman, Eric Van Cutsem

Manuscript received December 13, 2011; revised September 5, 2012; accepted September 7, 2012.

*These authors contributed equally to this work.

Correspondence to: Arnaud Roth, MD, Oncosurgery Unit, Geneva University Hospital, Geneva, Switzerland (e-mail: arnaud.roth@hcuge.ch).

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please e-mail: journals.permissions@oup.com.

Colon cancer is heterogeneous in clinical behavior and in the molecular mechanisms underlying its pathogenesis. Markers that identify these heterogeneous groups are needed to implement tailored therapeutic strategies. In the adjuvant setting, TNM classification remains the only validated prognostic tool. Since the seminal paper by Watanabe et al. (1), a myriad of candidate molecular biomarkers has been reported, but none has been rigorously assessed in a large patient dataset by multivariable analysis and subsequently implemented into clinical practice (2–4). Reasons for this include the limited number of patients studied, retrospectively collected heterogeneous patient series, lack of validation series, and lack of multivariable assessment to identify individual contributions (5, 6). Currently stage II colon cancer, a group where treatment decisions depend largely on prognostic prediction, is assessed by a set of high-risk markers (ASCO markers), whereas stage III cancers are undifferentiated and all are treated. The aim of this study was to identify prognostic subgroups in stage II and III colon cancer by careful assessment of clinical and molecular markers (7, 8).

We addressed this in a large randomized phase III trial (PETACC3, ClinicalTrials.gov NCT00026273) in which 3278 patients were accrued to assess the influence of adding irinotecan to 5-fluorouracil/leucovorin (5-FU/LV) in the adjuvant treatment of stage II/III disease (8). The trial did not meet its primary endpoint, and after a median follow-up of 66.3 months, no difference in disease-free survival (DFS) or overall survival (OS) was observed. During accrual, formalin-fixed paraffin-embedded (FFPE) tissue samples of the colectomy specimens from 1564 patients were prospectively collected, and the status of tumor molecular markers previously identified as promising for outcome in this setting was assessed. These included KRAS and BRAF mutations, loss of heterozygosity at chromosome 18q21 (18qLOH), microsatellite instability (MSI), and SMAD4 protein expression (7). We have performed multivariable analyses to establish which among the five promising markers have the most clinically significant prognostic value. In addition, we explored the possibility of constructing classification algorithms using the strongest variables to identify novel prognostic subgroups after adjuvant treatment but currently not addressed by TNM staging alone.

Patients and Methods

Patients of age between 18 and 76 years, with completely resected, histologically proven stage II and III adenocarcinoma of the colon,
were eligible for the clinical study. After providing written informed consent in which the planned translational study program was clearly specified, patients were randomized to receive 6 months of either 5-FU/LV alone or in combination with irinotecan on one of two different schedules and were then regularly followed up (8).

On the first schedule, patients received irinotecan (180 mg/m\(^2\) as a 30- to 90-minute infusion, day 1) and 5-FU/LV as the LV5FU2 (de Gramont) regimen (LV 200 mg/m\(^2\) as a 2-hour infusion, followed by 5-FU as a 400 mg/m\(^2\) bolus and then a 600 mg/m\(^2\) continuous infusion over 22 hours, days 1 and 2) every 2 weeks for 12 cycles or LV5FU2 alone on the same schedule. On the second schedule, patients were randomized to receive irinotecan (80 mg/m\(^2\) as a 30- to 90-minute infusion, day 1, every week for 6 weeks, followed by 2 weeks’ rest, for up to four cycles) and high-dose 5-FU on an Arbeitsgemeinschaft Internationale Onkologie (AIO) schedule (LV 500 mg/m\(^2\) as a 2-hour infusion, followed by 5-FU 2000 mg/m\(^2\) as a 24-hour infusion, on the same days as irinotecan) or the AIO regimen alone. The primary study endpoint was DFS, defined as the time from the date of randomization to the first date of local, regional, or distant relapse, second primary malignancy (colon or other) or death. The secondary endpoint endpoints were relapse-free survival (RFS), defined as the time from the date of random allocation to the first date of local, regional, or distant relapse, the occurrence of a second primary colon cancer or death, and overall survival (OS), defined as the time from random allocation until death (8).

Tumor biomarkers were analyzed in a subpopulation of these patients called the marker analysis population, and RFS was taken as the primary endpoint instead of DFS in order to avoid interference due to the occurrence of second primary malignancies other than colon by censoring them. Patient characteristics in the marker analysis population are detailed in Supplementary Tables 1 and 2 (available online).

**KRAS and BRAF Mutation Detection**

FFPE tumor and normal tissue was macroscopically dissected and genomic DNA was extracted using a standard phenol/chloroform extraction protocol (7). **KRAS** (codons 12 and 13) and **BRAF** (V600E) mutations were determined by an allelic discrimination assay. Briefly, for the first 600 cases, the presence of **KRAS** mutations were determined by direct sequencing of a 286-base pair fragment of **KRAS** exon 2 (PCR primers; Supplementary Table 3, available online). PCR amplification was carried out as previously described (9). One microliter of digested DNA was added to a 50 µL reaction containing 0.2 µM of each primer, 0.2 mM of each dNTP, 2 mM MgCl\(_2\), and 1.5 U of a Taq DNA polymerase (Roche Diagnostics, Vilvoorde, Belgium). PCR reactions were performed in a GeneAmp PCR system 2700 thermal cycler (Applied Biosystems, Foster City, CA) under the following conditions: an initial incubation at 95°C for 5 minutes, 40 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute, followed by a final extension at 72°C for 7 minutes. The presence of an appropriately sized PCR product was confirmed by resolving on a 2% agarose gel. PCR products were purified using ExoSAP-IT (USB Corporation, High Wycombe, UK) and sequenced using fluorescent BigDye-terminator cycle sequencing kit (Applied Biosystems). Sequencing reaction products were resolved on an ABI 3100 sequencing instrument (Applied Biosystems). Mutations were identified by visual analysis of the sequence chromatograms using SeqMan (DNASTar) or SeqScape (Applied Biosystems).

Because of the low success rate of this technique in FFPE samples, we subsequently switched to and analyzed all samples with an allelic discrimination assay on a 7500 Real Time PCR System (Applied Biosystems) (9,10). Seven **KRAS** exon 2 mutations were screened for: G12V, G12D, G13D, G12C, G12A, G12S, and G12R located within the codon 12, and G13D in codon 13. This technique was also used to screen samples for the **BRAF** mutation at codon 600 and is based on that initially published by Lievre et al. (9). Briefly, the presence of mutations in KRAS and BRAF was determined by an allelic discrimination assay on an ABI 7500 Sequence Detection System (Applied Biosystems). Briefly, reactions were performed in 25 µL comprising 10 ng of DNA, 1X of specific primers and probes, and 1X TaqMan Universal PCR Master Mix (Applied Biosystems). DNA was then submitted to the following cycle conditions: 95°C for 10 minutes; 50 cycles, 92°C for 15 seconds; and 60°C for 1 minute. Data were analyzed with SDS v1.3 software (Applied Biosystems).

**Loss of Heterozygosity of Loci at Chromosome 18q21**

Loss of heterozygosity (LOH) analysis of loci at chromosome 18q21 was performed in DNA samples by genotyping seven single nucleotide polymorphisms (SNPs) in the selected region by pyrosequencing on normal and tumor DNA. For each of the 7 SNPs, the rs number/localization/position/allele frequencies are given: rs965/18q11.2/30881486/T = 0.54/C = 0.45; rs2456/18q12.1/39103191/G = 0.69/A = 0.30; rs736839/18q16/44782063/C = 0.76/T = 0.33; rs140686-18q16/46057352/A = 0.56/G = 0; rs1025854/18q20/49347390/A = 0.56/G = 0.44; rs2226/18q21.5/56435992/A = 0.24/G = 0.75; rs2872/18q22/69725387/G = 0.72/C = 0.27. Pyrosequencing cutoffs for heterozygosity and LOH determination were established by first analyzing the ranges of the first allele (C) for the different SNPs on known heterozygous control samples. These ranges reflect the precision of the pyrosequencing in detecting the heterozygous state. Ranges are different for individual SNPs as they are sequence- and context-dependent. Expected signal for C is 50, actual ranges are given below: rs965: 46.3–51.1; rs2456: 44.5–51.7; rs736839: 50.5–54.7; rs140686: 43.1–52.5; rs1025854: 42–49.8; rs2226: 43.9–52.2; rs2872: 43.3–49.4. A scoring system for the tumor samples was established based on these ranges: no LOH if the frequency of the first allele signal was within the range of the controls, expanded with 5 on each side; LOH if the frequency of the first allele signal was within the range of the controls, expanded with 5 on each side; LOH if the frequency of the first allele signal was outside the range of the controls, expanded with 10 on each side; LOH unknown if the frequency of the first allele signal was between LOH and no LOH. SNPs were scored noninformative if they were homozygous on the corresponding normal DNA sample, and “No Result” was entered if the pyrosequencing failed on either the normal or tumor DNA sample. Only patients with at least two informative SNPs were considered for the analysis (N = 831 [59%]; Supplementary Figure 1, available online).

**Microsatellite Instability**

MSI was analyzed using a panel of 10 different microsatellite loci containing mono- or dinucleotide repeated sequences, as recommended by the international guidelines for evaluation of MSI in...
colorectal cancer (11). The panel consisted of the five markers from the Bethesda reference panel, with the addition of five markers that were also suggested during the International Workshop on IINPCC in 1997. Eight of these markers were organized into two multiplexes: multiplex-A, with BAT-25, BAT-26, D2S123, and D5S346; and multiplex-B with TGFBR2, BAT-40, D17S787, and D18S69. D17S250 and D18S58 were run in separate PCR reactions. PCR fragments were labeled with 6-FAM, HEX, or TET.

Amplicons from paired tumor and normal DNA from the same patient were analyzed on an ABI 3100 Genetic Analyzer (Applied Biosystems). Primers (Applied Biosystems) were diluted in sterile water to a final concentration of 100 µM and stored at −20°C. After optimizing procedures, optimal PCR mixes were defined. Each reaction volume of 25 µL consisted of 10–20 ng template, primers (concentrations ranged 0.125–5 µM), PCR Master Mix containing Taq DNA polymerase, MgCl2, dNTPs, and reaction buffer (Promega Benelux, Leiden, the Netherlands), in sterile water. Reactions were performed on a GeneAmp 9700 Thermal Cycler (Applied Biosystems) and, after an initial denaturation step at 95°C for 2 minutes, were cycled 35 times at a denaturation of 95°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 30 seconds, after which a final extension step at 72°C for 7 minutes was performed. PCR products were analyzed using the automated ABI Prism Sequencer Model 3100 Genetic Analyzer (Applied Biosystems). In brief, deionized formamide was combined with GeneScan 120 LIZ size standard (Applied Biosystems) and 1 µL of PCR product in a genetic analyzer sample tube using Multiprobe II Automated Liquid Handling System (Perkin-Elmer Life Sciences, Wellesley, MA). Samples were then denatured in a heat block for 3 minutes at 95°C, chilled on ice, and stored at −20°C prior to loading on the ABI Prism 3100 Genetic Analyzer. The run was carried out in accordance with the supplier’s protocol. Loci were scored for MSI using GeneScan software (Applied Biosystems). Only those fragments in which both tumor and normal DNA showed signals above background levels were considered assessable. A locus was scored unstable if unequivocal instabilities were seen in the tumor sample in comparison to the paired normal DNA. These instabilities could be either the appearance of additional peaks or shifts in fragment lengths. MSI was graded as high (MSI-H) when 30% or more of markers were unstable, low when 10%–20% of markers were unstable, and stable when all markers displayed no instability. For the statistical analysis, MSI low (MSI-L) and MS stable (MS-S) populations were pooled and identified as MS-L/S (12).

Immunohistochemistry

The expression of SMAD4 was determined by immunohistochemistry (IHC), following a standard streptavidin/peroxidase procedure. Deparaffinization and rehydration of the sections was followed by heat-induced epitope retrieval, and sections were stained with Smad4 (B-8 clone) mouse monoclonal antibody (0.2 mg/mL; Santa Cruz Biotechnology, CA), which recognizes the full-length human SMAD4 molecule (amino acid 1–552).

Semiquantitative scoring of the IHC staining was performed by two independent observers per marker. Scoring was based upon the percentage of tumor cells stained in an overall assessment of all tumor tissue available. Occurrence of clusters of unstained cells was considered as loss, regardless of the percentage of unstained cells. A representative example of loss of focal expression of SMAD4 is shown in Supplementary Figure 2 (available online).

Statistical Analysis

Relationships between the occurrence frequencies of variables were evaluated by Fisher’s exact test. Prognostic importance of variables was assessed using Cox’s proportional hazard regression models; data were summarized with hazard ratios (HR) and their 95% confidence intervals (CI); and P-values were estimated by Wald tests or likelihood ratio tests (LRT). Graphical assessment of proportional hazards based on Schoenfeld residuals showed no important departure from proportionality (no association with time) for any of the variables.

The significance of an interaction term with stage and treatment, respectively, was also tested by a Wald test in presence of both main effects. The Kaplan–Meier method was used to compute survival curves, median survival, and estimated survival rates at fixed time points, and differences between curves were assessed using the log rank test. Recursive partitioning trees for survival were constructed using the algorithm in the party (13) add-on package to the R software system, which implements a method that separates the variable selection and splitting steps. The method is based on locally optimal decisions and, like recursive partitioning methods in general, does not additionally search for a globally optimal solution. At each node, rejection of the global null hypothesis of independence between the response and each covariate at a specified level of significance (.05) was tested. If rejected, the node was split using the explanatory variable most strongly associated with survival. The binary partition is then performed taking the partition that maximizes a two-sample statistic (13). All variables we considered were categorical, with prespecified level two for BRAF mutation, SMAD4 expression, and MSI status, and three for nodal status and T stage.

Patients with missing values for the variables used were retained in the multivariable regression models, introducing a separate level for the missing values, but were excluded in the tree and the follow-up analyses about prognostic subgroups.

All reported P-values were two-sided. The reported P-values were not adjusted for multiple testing, but the cutoff for statistical significance was set at .01, corresponding to the usual cutoff of .05 adjusted with a Bonferroni correction for five tests. The main aim of the study was to make decisions regarding the prognostic role of the five molecular markers included in the study. Further correction due to the multiplicity of models (OS, RFS, univariate, and multivariable models) was not performed, considering that the results of these four models are positively correlated to each other. The same cutoff was used for testing significance of interaction terms with stage and treatment (OS and RFS, univariate only for each of the five molecular markers). In this way, good statistical power could be maintained without inflating the risk for type II errors (false negatives). Analyses were performed using the R software system for statistical computing (14).

Results

The pathology material from each enrolled patient was prospectively requested from participating institutions, and samples from 1564 patients were successfully collected during study accrual. The
Table 1. Marker alteration rates reported in the literature and in the PETACC3 marker analysis population (n = 1404)*

<table>
<thead>
<tr>
<th>Marker</th>
<th>Reported alteration rate</th>
<th>PETACC3</th>
<th>Stage II (n = 420)</th>
<th>Stage III (n = 984)</th>
<th>P†</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSI-H</td>
<td>10%–17%</td>
<td>15%</td>
<td>22%</td>
<td>12%</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SMAD4 expression†</td>
<td>13%–63%</td>
<td>21%</td>
<td>18%</td>
<td>23%</td>
<td>.03</td>
</tr>
<tr>
<td>18qLOH</td>
<td>70%</td>
<td>88%</td>
<td>87%</td>
<td>8%</td>
<td>1</td>
</tr>
<tr>
<td>KRAS mutation</td>
<td>32%–42%</td>
<td>37%</td>
<td>36%</td>
<td>37%</td>
<td>.80</td>
</tr>
<tr>
<td>BRAF mutation</td>
<td>10%</td>
<td>8%</td>
<td>8%</td>
<td>8%</td>
<td>1</td>
</tr>
</tbody>
</table>

* Number of patients screened from the total population for each marker and the success rate for determining individual marker status has been reported previously (7). MSI-H = microsatellite instability high; 18qLOH = loss of heterozygosity on chromosome 18q21.  
† Fisher’s exact test for stage II vs stage III disease.  
‡ Any loss of expression.

We next performed multivariable analysis to identify independent predictors of outcome. Multivariable modeling identified T stage, N stage, number of examined lymph nodes, age, MSI status (HR 0.54, 95% CI = 0.37 to 0.81, P = .003), and SMAD4 expression (HR 1.47, 95% CI = 1.19 to 1.81, P < .001) as the most important independent determinants of RFS (Table 3). In particular, MSI and SMAD4, despite being associated with each other and with other prognostic variables (Supplementary Tables 6 and 7, available online), both contributed additional prognostic information in a model with all other factors included. Notably, 18qLOH status lost all prognostic value for RFS when analyzed in a multivariable model (HR 0.93, 95% CI = 0.70 to 1.24, P = .64).

In the multivariable analysis for factors predicting OS (MSI-H status; HR of death = 0.43, 95% CI = 0.27 to 0.70, P = .001; SMAD4 loss: HR = 1.58, 95% CI = 1.23 to 2.01, P < .001), results were comparable with those described for RFS (Table 3). A notable exception was again that BRAF mutation status was prognostic for OS (HR 1.56, 95% CI = 1.02 to 2.39, P = .04) although not for RFS, in line with our previous report (10).

We examined the data to identify if any of the prognostic effects found with the molecular markers were different between stage II and stage III disease by testing interaction terms in survival regression models with stage and one marker at a time (Supplementary Table 8, available online). No statistically significant differences were found (all P > .01). Similarly, no strong evidence was found for differences of effects among the two treatment arms (Supplementary Table 9, available online). Thus, the results presented above in the full cohort are meaningful average effects across the two stages and are independent of the addition of irinotecan to 5-FU in the adjuvant treatment. However, the power to detect such differences is not high, and the possibility of moderately different effects by stage or treatment cannot be excluded as discussed below.

In order to confirm that the strongest molecular predictors of survival could be integrated with the dominating role of T and N stage in our survival data, we applied recursive partitioning prognostic models. Analysis for RFS (Figure 1) with only T and N variables gave a prognostic tree model that resulted in six prognostic subgroups (Figure 1, A): N2, T4N1, T3N1, T(1,2)N1, T4N0, and T3N0. This model agrees remarkably well with the recent findings of the AJCC (20,21) and supports the efficacy of this approach. If a less stringent cutoff level was used, the model was refined to include further subdivision of the N2 group by T stage.

When the molecular variables, SMAD4 expression, MSI status, and BRAF mutation status, were integrated into the model, SMAD4 expression and MSI status but not BRAF mutation status
were included in prognostic trees for both RFS (Figure 1, B) and OS (Figure 2, B). Under patient resampling, the detailed structure of the tree was not stable, but MSI and SMAD4 were almost always used (MSI 98%, SMAD4 95% of the RFS trees). When analyzing patient subsets, the most important prognostic contributions of the molecular markers evidenced by the trees were the prognostic stratification by SMAD4 expression status in both the stage III subgroups—the N1 and N2 groups—and the stratification by MSI in the N0 and N1 groups, that is in stage II but also in the T3N1 subgroup of stage III.

To better identify the potential contributions of the molecular markers in the clinical setting, we further tested the potential refinement of risk estimation in N0 groups stratified by MSI status and in the T3N1 group (the largest TNM subgroup in our data) by concomitant use of MSI and SMAD4 status.

Patients with T3N0/MSI-H tumors (66/318, 21% of T3N0) had estimated 5-year RFS and OS rates of 97.0% (95% CI = 92.9 to 100) and 98.5% (95% CI = 95.6 to 100), respectively (Figure 3, A and B). Patients with T4N0/MSI-H tumors (20/77, 26% of T4N0) had 5-year RFS and OS rates of 85% (95% CI = 71 to 100) and 95% (95% CI = 86 to 100), respectively (Figure 3, A and B). In patients with N0 tumors with MSI status (N = 395), the improvement obtained in prediction of long-term outcome with the use of the MSS/MSI-H split was, judged by HR and LRT, no less important than that obtained with the use of the T3-T4 distinction. Thus, for predicting RFS and OS, respectively, for MSI (MSS vs MSI-H), HR = 0.26, P < .001, and HR = 0.16, P < .001, and for T stage (T3 vs T4), HR = 2.4, P = .001, and HR = 2.3, P = .015.

The analysis was not powered to address the question whether the predictive role of MSI status for RFS and OS, respectively, was of the same magnitude in T3N0 tumors, HR = 0.16, P < .001, and HR = 0.12, P = .004, compared with the smaller group of T4N0 tumors, HR = 0.38, P = .077, and HR = 0.18, P = .033. These results confirmed the evidence for the strong prognostic relevance of MSI status in stage II colon cancer and the considerable importance for its use in treatment decisions in association with T stage.

We similarly explored the inclusion of MSI and SMAD4 expression status for risk stratification of patients in the T3N1 subgroup (n = 408 after exclusion of MSI and/or SMAD4 cases with missing values). The addition of either factor was statistically significant (LRT: P < .03 for MSI, P = .01 for SMAD4). The RFS (Figure 4, A) and OS (Figure 4, B) curves for the T3N1 populations are shown. In patients with T3N1 tumors, the 5-year RFS was 72%
Overall survival
Relapse-free survival
Outcome parameter prognostic variable
HR (95% CI) | P | Interpretation on outcome
---|---|---
Age (decades) | 1.11 (1.01 to 1.22) | .03 | —
Sex (male vs female) | 1.12 (0.92 to 1.36) | .26 | —
Grade (G-3/4 vs G-1/2) | 1.32 (0.96 to 1.79) | .07 | —
T stage (T3 vs T1/2) | 2.08 (1.20 to 3.57) | .009 | T3 worse than T1/2
T stage (T4 vs T3) | 1.73 (1.38 to 2.17) | <.001 | T4 worse than T3
N stage (N1 vs N0) | 1.73 (1.31 to 2.28) | <.001 | N1 worse than N0
N stage (N2 vs N0) | 3.97 (2.01 to 5.24) | <.001 | N2 worse than N0
Number of examined lymph nodes (increments of 10) | 0.79 (0.69 to 0.90) | <.001 | More examined lymph nodes better
Site (right vs left) | 1.01 (0.82 to 1.25) | .91 | —
Treatment group (FOLFIRI vs 5-FU/LV) | 0.91 (0.75 to 1.09) | .31 | —
MSI status (MSI-H vs MS-L/S) | 0.54 (0.37 to 0.81) | .003 | MSI-H better than MS-L/S
SMAD4 expression (any loss vs no loss) | 1.47 (1.19 to 1.81) | <.001 | Any loss of expression worse than no loss of expression
18qLOH status (LOH vs no LOH) | 0.93 (0.70 to 1.24) | .64 | —
BRAF mutation status (mutation vs wild-type) | 1.17 (0.79 to 1.73) | .44 | —
KRAS mutation status (mutation vs wild-type) | 1.06 (0.85 to 1.30) | .65 | —
Overall survival
Age (decades) | 1.17 (1.04 to 1.31) | .008 | Increasing age worse
Sex (male vs female) | 1.23 (0.98 to 1.55) | .07 | —
Grade (G-3/4 vs G-1/2) | 1.40 (0.99 to 1.99) | .06 | —
T stage (T3 vs T1/2) | 2.22 (1.14 to 4.35) | .02 | —
T stage (T4 vs T3) | 1.94 (1.50 to 2.52) | <.001 | T4 worse than T3
N stage (N1 vs N0) | 1.83 (1.30 to 2.57) | .001 | N1 worse than N0
N stage (N2 vs N0) | 4.53 (3.24 to 6.33) | <.001 | N2 worse than N0
Number of examined lymph nodes (increments of 10) | 0.73 (0.63 to 0.86) | <.001 | More examined lymph nodes better
Site (right vs left) | 1.09 (1.01 to 1.65) | .04 | —
Treatment group (FOLFIRI vs 5-FU/LV) | 0.94 (0.75 to 1.17) | .58 | —
MSI status (MSI-H vs MS-L/S) | 0.43 (0.27 to 0.70) | .001 | MSI-H better than MS-L/S
SMAD4 expression (any loss vs no loss) | 1.58 (1.23 to 2.01) | <.001 | Any loss of expression worse than no loss of expression
18qLOH status (LOH vs no LOH) | 0.78 (0.56 to 1.09) | .14 | —
BRAF mutation status (mutation vs wild-type) | 1.56 (1.02 to 2.39) | .04 | —
KRAS mutation status (mutation vs wild-type) | 1.10 (0.86 to 1.42) | .45 | —

* P-values are from multivariable Cox proportional hazards survival regression statistics (Wald test). 5-FU = 5-fluorouracil; LV = leucovorin; FOLFIRI = 5-fluorouracil, leucovorin, irinotecan; MSI = microsatellite instability; MSI-H = microsatellite instability high; MS-L/S = microsatellite instability low; microsatellite stable; LOH = loss of heterozygosity.

(95% CI = 68 to 77) and 5-year OS was 83% (95% CI = 80 to 87). Patients with MSI-H tumors with no loss of SMAD4 expression comprised 11% (44/408) of T3N1 patients and displayed comparatively good prognosis, 5-year RFS of 86.3% (95% CI = 77 to 97), and 5-year OS of 90.8% (95% CI = 83 to 100). In contrast, patients with MS-L/S tumors with any loss of SMAD4 expression accounted for 20% (82/408) of T3N1 patients and demonstrated worse prognosis, 5-year RFS of 62% (95% CI = 52 to 73), and 5-year OS of 72% (95% CI = 63 to 83).

**Discussion**

In colon cancer, clinical practice is still entirely based on T and N stage. A variety of markers have been proposed to refine T- and N-based groups, but their integration into the clinical setting requires extensive validation of their relative value and optimal use (2,6). In this study, we have provided an integrated analysis of both clinical and molecular markers to assess their relative value and optimal positioning in the diagnostic workup. For this, univariate and multivariable analysis tools and recursive partitioning were used. Considering the heterogeneity of colon cancer and the complex interactions between markers, the estimation of prognostic effects depends on which variables are included, perhaps explaining some of the inconsistencies reported between studies. The most useful data are those from multivariable models attempting to include all relevant factors. The design of the PETACC3 trial allowed the inclusion of multiple clinical, pathological, and molecular variables in a multivariable model, without the issues of validity that accompany large studies that collect data from multiple sources (22). The aim of the study was to make decisions on the evidence for a prognostic role of five molecular markers in the presence of a full set of important adjustment variables.

In terms of validation of previously proposed biomarkers, our dataset suggests, in agreement with others (5), that 18qLOH is less important than previously assumed. Since Jen et al. (23) published the prognostic significance of 18qLOH in stage II colorectal cancer, later confirmed in stage III disease (1), interest in this marker has remained, although subsequent studies failed to confirm these findings (24,25). A meta-analysis performed on 17 publications concluded that prospective studies using consistent methodology...
are needed to quantify the effect of 18qLOH and its role in patients with stage II-III disease (15).

SMAD4 is an important transcriptional mediator in the TGF-β signalling pathway and has been implicated in colon cancer development in preclinical studies (26,27). Loss or low expression of SMAD4 has been associated with poor prognosis in colorectal cancer patients in small series (28,29). Ours are the only data so far based on a large homogeneous series to further validate SMAD4 expression status as an independent prognostic variable. We conclude, therefore, that even focal loss of expression of SMAD4 is a good candidate for patient stratification in stage III disease. Furthermore, BRAF mutation status was found to be a predictor of OS despite not having prognostic significance for RFS, suggesting that its effect is primarily on survival after relapse, as we reported previously (30). SMAD4 expression status, however, was predictive of both RFS and OS and may be a more important tool to predict relapse of stage III disease after adjuvant treatment.

We confirmed and expanded on earlier findings that MSI status constitutes a remarkable prognostic factor in stage II/III disease (31,32). The major effects on both RFS and OS, recently reported in untreated populations (33,34), were reproduced in treated patients in PETACC3. Consistent with the report by Kerr et al. (33), we found the effect of MSI status was strong and independent.

**Figure 1.** Conditional inference trees incorporating molecular variables with TNM staging into prognostic models for RFS: A) based on T and N stages; B) MSI status, SMAD4 expression status, and BRAF mutation status were introduced to test how they might improve the prognostic value of TNM staging. *P*-values are from log-rank statistics. LOE = loss of expression; MSI = microsatellite instability; MSI-H = microsatellite instability high; MS-L/S = microsatellite instability low and microsatellite stable; RFS = relapse-free survival.
of clinical variables, including T stage, sex, and grade. Interaction tests of each of the five markers with stage and treatment respectively did not reach statistical significance. Nonetheless, we observed borderline significance for interaction of MSI with stage and of SMAD4 with irinotecan treatment (Supplementary Tables 8 and 9, available online).

One observed trend implies the possibility of a stronger prognostic effect of MSI in stage II compared with stage III disease, which was also found for 18qLOH (Supplementary Table 8, available online). A more surprising trend was found with SMAD4 expression, which might imply that the addition of irinotecan would be harmful in the high-risk subgroup that has partial loss of SMAD4 expression and conversely might be of higher benefit in the population that has no loss of expression (Supplementary Tables 9 and 10, available online). Although this test reached borderline statistical significance ($P = .02$), its implications seem implausible, but the association should be reanalyzed in future data.

Finally, our analyses suggest that clinically relevant refinement of prognostic assessment with TNM classification might be achieved with the inclusion of information on molecular markers. Although used as a pretherapy staging system, TNM is based on the study of cohorts of treated and untreated patients (21). As such, it can be used as a basis for the exploration of new markers arising from the study of patient populations followed from the time of diagnosis, which might have the potential to refine it. MSI-H,
loss of SMAD4 expression, and \( \text{BRAF} \) mutated tumors constitute patient subgroups that are relatively small, limiting the variation in survival that can be explained by these variables. Nonetheless, this study was large enough for a fairly precise estimation of the differences. With hazard ratios of similar sizes as those from the \( T \) and \( N \) subclasses, the prediction of patient survival for these groups...
Figure 4. Kaplan–Meier curves incorporating patient tumor MSI status and SMAD4 expression status in selected TNM subgroups. A) RFS and B) OS, respectively, in patients with T3N1 tumors with 95% confidence bands as shown as intermitted lines (censoring marks suppressed) and with T3N1 tumors subdivided by MSI and SMAD4 status as indicated (no curve is shown for the group of six patients who were MSI-H with loss of SMAD4 expression). Curves for T3N0 and T3N2 are shown for reference (black as indicated, censoring marks suppressed). LOE = loss of expression; MSI = microsatellite instability; MSI-H = microsatellite instability high; MS-L/S = microsatellite instability low and microsatellite stable; RFS = relapse-free survival; OS = overall survival.
could be substantially improved clinically by the inclusion of the biomarker information.

This study also had some limitations. A trend to a possible interaction between irinotecan and SMAD4 needs to be further investigated in a larger cohort. All patients in our cohort received adjuvant treatment, and therefore these results would be strengthened if replicated retrospectively in an untreated patient population to ensure the absence of treatment interaction. Additionally, as for any work of this kind, a validation in other series would be appropriate.

How to incorporate these variables into current clinical practice requires further study. The exploratory statistical models from our data suggest several relevant possibilities. Thus, in the T3N1 group, patients with MSI-H tumors with no loss of SMAD4 expression (11%) had survival similar to those with T3N0 tumors (stage II A disease), whereas those with tumors displaying MS-L/S and loss of SMAD4 expression (20%) had statistically significantly worse survival, approaching that of the T3N2 population (Figure 4). If replicated in untreated patients, this result could imply that 11% of the T3N1 population could be spared adjuvant chemotherapy, allowing cost containment and avoiding unnecessary toxicity. The same reasoning might hold true for patients with T4N0M0, MSI-H tumors (26% of the T4N0M0 population) whose survival was comparable with that of patients with T3N0M0 tumors (Figure 3) but was higher than that of the patients with T3N0M0 and MS-L/S tumors. We emphasize that all patients in PETACC3 received adjuvant therapy and therefore these results cannot be translated directly into clinical practice. However, a recent report on the molecular risk score performed on stage II patients in the QUASAR study, which includes a no adjuvant treatment arm, suggested that the molecularly assessed risk or recurrence was not substantially affected by the treatment and that it had no important relationship with adjuvant treatment efficacy (35). Therefore, once they are validated in other series, our results should serve as a basis for the refinement of future clinical investigations into the adjuvant treatment of colon cancer.

In conclusion, our data indicate that after further validation in large multivariable analyses, the integration of molecular markers might improve the risk assessment of stage II and III colon cancer. Given that informative molecular markers will continue to be discovered in the postgenomic era, the challenge will be to develop and validate combinations of variables with independent contributions to prognosis in multifactorial models. This will facilitate the development of more personalized therapeutic strategies for the benefit of the individual patient.

References

Funding
The PETACC3 trial was initiated by Rhône-Poulenc, which became Aventis after merging with Hoechst AG. The rights of CPT11 and the trial were then bought by Pfizer at the end of 2005. The industry sponsored the PETACC3 trial and funded the molecular laboratory analysis and supported in part the statistical analysis in the translational program described in the manuscript.

The statistical analysis of this work was also supported in part by a grant of the Swiss National Science Foundation (SNF 320030_135421) to AR and MD, the Medic Foundation, and a grant of the Krebsforschung Schweiz (KFS 02697-08-2010) to AR and MD.

Notes
The PETACC3 trial and its translational research program were conducted in collaboration with the European Organization for Research and Treatment of Cancer (EORTC #40993), the Swiss Group for Clinical Cancer Research (SAKK 60/00), and The Swiss Institute of Bioinformatics (SIB). Pfizer supported medical writing services provided by Paul Hoban from Cancer Communications and Consultancy Ltd (Knutsford, UK). EVC and ST are senior investigators of the fund for Scientific Research Flanders, Belgium (Fonds Wetenschappelijk Onderzoek-Vlaanderen), Bart Biesmans (Digestive Oncology Unit, University Hospital of Gasthuisberg), Monica Carto (Department of Surgical and Morphological Science, University of Genova), and Astrid Bachman (Department of Pathology, Lausanne University) provided technical assistance, and Daniel Dietrich (Swiss Group for Clinical Cancer Research [SAKK]) provided statistical support. They did not receive compensation for their work.

Affiliations of authors: Oncosurgery Unit, Geneva University Hospital, Geneva, Switzerland (ADR); Département de formation et recherche, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland (MD); The Swiss Group for Clinical Cancer Research (SAKK) (DK), Swiss Institute of Bioinformatics (GdA, MD, DK), Lausanne, Switzerland; Department of Pathology, Lausanne University, Lausanne, Switzerland (FB, PY); Digestive Oncology Unit, University Hospital Gasthuisberg, Leuven, Belgium (ST, EVC); Department of Surgical and Morphological Science, University of Genova, Italy (RF); Unit of Medical Oncology, Ospedali Riuniti, Bergamo, Italy (RL); Pfizer Inc, New York, NY (LC); Department of Medical Oncology, The Royal Marsden Hospital, Sutton, UK (DC); Hôpital Ambroise Paré, CHU Paris Ouest, Boulogne, France (BN).