Tissue Biomarker Development in a Multicentre Trial Context: a Feasibility Study on the PETACC3 Stage II and III Colon Cancer Adjuvant Treatment Trial

Fred T. Bosman, Pu Yan, Sabine Tejpar, Roberto Fiocca, Eric Van Cutsem, Richard D. Kennedy, Daniel Dietrich, and Arnaud Roth

Abstract

Purpose: We evaluated the feasibility of biomarker development in the context of multicenter clinical trials.

Experimental Design: Formalin-fixed, paraffin-embedded (FFPE) tissue samples were collected from a prospective adjuvant colon cancer trial (PETACC3). DNA was isolated from tumor as well as normal tissue and used for analysis of microsatellite instability, KRAS and BRAF genotyping, UGT1A1 genotyping, and loss of heterozygosity of 18 q loci. Immunohistochemistry was used to test expression of TERT, SMAD4, p53, and TYMS. Messenger RNA was retrieved and tested for use in expression profiling experiments.

Results: Of the 3,278 patients entered in the study, FFPE blocks were obtained from 1,564 patients coming from 368 different centers in 31 countries. In over 95% of the samples, genomic DNA tests yielded a reliable result. Of the immunohistochemical tests, p53 and SMAD4 staining did best with reliable results in over 85% of the cases. TERT was the most problematic test with 46% of failures, mostly due to insufficient tissue processing quality. Good quality mRNA was obtained, usable in expression profiling experiments.

Conclusions: Prospective clinical trials can be used as framework for biomarker development using routinely processed FFPE tissues. Our results support the notion that as a rule, translational studies based on FFPE should be included in prospective clinical trials. (Clin Cancer Res 2009;15(17):OF1–6)

In spite of remarkable progress in our understanding of the molecular mechanisms involved in the development of cancer, the impact of this knowledge on cancer care has fallen short of expectations. Several reasons to explain this slow progress can be proposed. First, adjuvant treatment, although rationally addressing the risk of recurrent disease, tends to be beneficial for a limited proportion of patients because it cannot be predicted who will recur or which recurring patient will respond to the adjuvant treatment. Biomarkers, capable of predicting recurrent disease as well as (non)response to adjuvant treatment modality, potentially provide a solution to this problem, as exemplified by the predictive value of c-Kit mutations for the response of GIST to the tyrosine-kinase inhibitor imatinib (1) or of KRAS mutations to epidermal growth factor receptor blocking agents (2, 3). Second, the long duration of clinical trials, which usually have outcome (disease free or overall survival) as endpoint, hampers rapid introduction of new treatment modalities (4). Biomarkers indicating who responds would constitute surrogate study end-points that could be reached more rapidly (2).

Few published putative biomarkers have made it into clinical practice because (a) most biomarker development studies were done on small patient numbers and significance in multivariate analysis was not studied or not confirmed; (b) often studies published are based on a learning set only, without independent validation on a test set; (c) most studies are retrospective and suffer from incomplete data or insufficient data quality; and (d) the tests proposed are often derived from experimental studies with insufficient validation in a clinical setting, taking factors such as tissue preservation, sampling adequacy, and test reproducibility into account. Biomarker development could profit immensely from large clinical trials (2, 4, 5). Sufficient patient numbers can be accrued, patient data are prospectively collected, and thus generally of high quality and often central...
Translational Relevance

Biomarker development for diagnostic, prognostic, or predictive purposes is a timely topic in clinical cancer medicine and cancer research. Biomarker development and validation depends on the availability of large well-documented case collections along with appropriate biosamples. In this study, we focus on how formalin-fixed paraﬃn-embedded tissue samples may be used for tissue biomarker development. The experience gained in the study exempliﬁes how in the context of clinical trials inclusion of well-planned translational studies on routinely collected formalin-ﬁxed paraﬃn-embedded can be used for a wide range of biomarker development platforms. We consider it important to draw attention to this approach as it would be preferable to, as a rule, include such studies in clinical trials.

review of histopathologic material is required, which provides a logical stepping stone for centralized collection of tissue samples. A limitation of this approach is that usually only formalin-ﬁxed paraﬃn-embedded (FFPE) tissue samples are available with inherent problems for detailed molecular testing (6). Nonetheless, several studies exploring the potential of this approach has been published (7–11). We embarked upon a search for new biomarkers in colon cancer with as starting point the PETACC-3 trial.7

It is the purpose of this article to communicate our experience with this approach, analyze successes and failures, and propose inclusion of this type of approach as a routine in clinical trial design.

Materials and Methods

The trial. The trial, a nonblinded multicenter randomized phase III study conducted within the Pan-European trial Adjuvant Colon Cancer network, was designed to study whether addition of irinotecan to infusional 5-Fluorouracil/FA would improve disease-free survival (DFS) when compared with 5-Fluorouracil/FA alone as adjuvant treatment in stage II and III colon cancer patients. Among the secondary objectives was assessment of the prognostic value of a set of gene and protein expression markers for DFS or overall survival. The trial was conducted according to the Declaration of Helsinki and its conduction was monitored by a steering committee and an independent data monitoring committee.

The tissue bank. The translational studies were included in the original study protocol. Oncology departments having submitted patients were requested to submit one FFPE tissue block containing a signiﬁcant quantity of tumor and normal tissue. A brief description of the project was provided to stimulate participation. When the trial was in full swing, a newsletter was semiannually circulated with statistics on the ﬂow of patient material and the progress made on the analyses undertaken, with the same goal.

From the blocks, 20 5-μm sections were cut, mounted on super frost slides, and stored at 4°C in light tight boxes. Later on, once the decision was taken to include molecular genetic analyses, ﬁve additional 20-μm sections were cut for DNA extraction. About halfway through the collection of the material, the steering committee agreed with the proposal to prepare a tissue microarray. This was done by taking ﬁve 0.6-mm samples from the tumor periphery and center, as well as three samples of normal tissue. A manual tissue arrayer was used according to procedures described previously (12).

Remotely remaining material was returned to the submitting pathology laboratories as a rule not later than 3 mo after reception of the blocks.

Histology. The context of the study, each tumor was reclassiﬁed by one expert pathologist (FR), and graded according to WHO criteria (13).

Immunohistochemical markers. The following markers were chosen as targets, based on literature considerations:

- Telomerase, reasoning that tumor stem cells might be characterized by telomerase expression and their proportion might be a parameter indicative of tumor behavior. This was achieved through immunohistochemical detection of TERT, the reverse transcriptase protein of the telomerase complex (14–16).
- P53, as this remains one of the most studied parameters in colorectal cancer (17).
- SMAD4 as a well-recognized prog nostic parameter in colorectal cancer (18).
- Thymidylate synthetase (TYMS), thymidylate phosphorylase (TP), and dihydropyrimidine dehydrogenase as genes important in pyrimidine metabolism, in the context of the 5-Fluorouracil adjuvant treatment (21–24).

For efﬁciency purposes, immunohistochemistry for different sets of markers was conducted in two different laboratories (RF, Department of Morphology, University of Genoa; p53 and SMAD4; FTB, Institute of Pathology, University of Lausanne Medical Center, Lausanne, CH; TYMS, thymidylate phosphorylase, dihydropyrimidine dehydrogenase, TERT). The two laboratories exchanged protocols and stained slides, to obtain comparable staining results. Staining was done on 5-μm parafﬁn sections that had been stored for various periods of time (up to 2 y) as described above.

Standard methodology was used with heat-induced epitope retrieval after deparafﬁnization and rehydration of the sections. The used antibodies and the applied staining conditions are provided as Supplementary Data. Semiquantitative scoring of the immunohistochemistry results was done by at least two independent observers per marker. Scoring was based on the percentage of tumor cells stained, in an overall assessment of all tumor tissue available. Scoring details differed per marker, depending on the characteristics of the immunoreactivity patterns obtained.

Molecular genetic markers. For molecular genetic analysis, tissue sections were retrieved from the stock as described above. Areas with normal tissue and tumor tissue were identiﬁed by microscopy of an H&E-stained slide. These were macroscopically disected by scraping with a scalpel blade. Tissue thus collected was deparafﬁnized and genomic DNA was extracted using a standard a standard phenol/chloroform extraction protocol.

DNA samples were dispatched at ambient temperature to the molecular genetics reference laboratory (ST, Centre for Human Genetics, University Hospital Gasthuisberg, Leuven, Belgium).

The presence of KRAS (codon 12) and BRAF mutations (codon 600) was determined by an allelic discrimination assay. Loss of heterozygosity (LOH) analysis of the 18q21 locus was done by genotyping seven single nucleotide polymorphisms in the selected regions on normal and tumor DNA by pyrosequencing. UGT1A1 genotyping was done by determining the status of the TA repeat (5, 6, 7, or 8) in the TATA box of the UGT1A1 promoter region on normal DNA by PCR amplification using the primer set previously published (25). For MSI determination, a panel of 10 mononucleotide and dinucleotide microsatellite loci was used. MSI was graded as high (MSI-H) when three or more


Published OnlineFirst August 18, 2009; DOI: 10.1158/1078-0432.CCR-09-0741
markers were positive, low (MSI-L) when one or two markers were positive, and stable when all markers were negative (26).

**Microarray analysis.** RNA was extracted from five stored (at 4°C in light tight boxes) or freshly cut (from the same block) 5-μm FFPE sections, amplified and labeled with biotin. The yield of total RNA and amplified cDNA was assessed and the amplified cDNA was then hybridized to a colorectal cancer disease specific microarray (Almac Diagnostics). Array scanning yielded data that was assessed on the basis of parameters obtained from bioinformatics and statistics toolboxes in Affymetrix GeneChip Operating System report files (Affymetrix) and Matlab.

**Data management and statistics.** A Web-based tool for the on-line submission of the results was developed, permitting direct downloading of data into the database. Data collection was monitored by the statistics unit of the Swiss Group for Clinical Cancer Research. The effects of covariates on survival outcomes, expressed as hazard ratios with confidence intervals, was determined through Cox regression. It was decided that for variables without established cut-points, an approach through receiver operator characteristics (ROC) curves would be applied. Associations between biomarkers, or biomarkers and established prognostic variables, were determined by χ² test. Kaplan-Meier curves and Cox regressions of disease-free survival (DFS), relapse-free survival (RFS), and overall survival were calculated for each biomarker, univariate and multivariate, and reported as P values, hazard ratios, and 95% confidence intervals.

### Results

As this article is intended as a feasibility study, details of the results are not presented. A summary of the patients accrued in the trial and entered in the translational study is provided in Table 1 and Fig. 1.

**Sample size, sample quality, time frame.** The translational study protocol required participating institutions to submit for each patient a FFPE tissue block. Considerable efforts in terms of informing participating centers and pathologists paid off: FFPE blocks were obtained of 1,564 cases of 3,278 (48%) from 368 sites in 31 countries, as is summarized in Table 1 and Fig. 1. When the available blocks did not allow identification of a sufficient quantity of normal and tumor tissue, the case was excluded from the study. As a consequence, overall, 1,401 cases (419 stage II and 982 stage III) could be included in the study.

The study accrued from January 2000 to April 2002 and paraffin blocks were collected until April 2004. Tests were run in 2005/6 but data analysis was possible only when the first follow-up data became available early in 2007 and final analysis only in 2008. A study of this scope will, as a consequence, easily spread over a 5- to 10-year period.

**Quality of the obtained results.** How the different tests fared in terms of number of cases with a usable result, and how the frequency of abnormalities found relates to what has been published is listed in Table 2. Immunohistochemistry results varied according to the marker studied. For p53 and SMAD4, robust results were obtained with few technical complications and reproducible semiquantitative assessment (98% success rate). This finding is confirmed by a recently published similar study on colorectal cancer (11). For IERT, however, major problems were specificity of the available antibodies and variations in tissue-processing quality impacting on the obtained results. This is reflected in a high number of cases judged of insufficient quality (success rate of 787 of 1,401 or 56%). Thymidylate phosphorolysase and dihydropyrimidine dehydrogenase results were totally irreproducible and not included in the study.

The molecular genetic tests were highly successful. The amount of tumor DNA obtained varied according to the amount of tumor tissue available per section, on the average 1,200 ng/cm² of tissue. The results of the DNA tests were highly satisfactory with a usable result in between 1,259 cases (90% for MSI) and 1,304 cases (93% for BRAF). Preliminary results show that the percentage of cases with MSI or with a KRAS or BRAF mutation is in close agreement with percentages published, confirming the reliability of the obtained results (Table 2).

**Transcriptional profiling pilot study.** The average concentration of RNA extracted for the stored sections was 140 ng/mL of extraction buffer and 193.6 ng/mL for the freshly cut sections. The associated 260/280 absorbance ratios were consistent between the paired samples and ranged between 1.8 and 2.4. Following amplification, all samples generated the recommended minimum cDNA concentration of 200 ng/μL required for hybridization. On a proprietary cDNA microarray platform (27) both the stored and the freshly cut samples yielded very good results (Table 2).

### Table 1. Number of patients and events

<table>
<thead>
<tr>
<th>No. of patients</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. patients</td>
<td>3,278</td>
</tr>
<tr>
<td>Patients eligible with DFS of &gt;1 d</td>
<td>3,128</td>
</tr>
<tr>
<td>Patients with a relapse</td>
<td>1,180</td>
</tr>
<tr>
<td>No. of events before 3 y</td>
<td>1,061</td>
</tr>
<tr>
<td>No. of events after 3 y</td>
<td>202</td>
</tr>
<tr>
<td>Patients with marker data</td>
<td>1,452</td>
</tr>
<tr>
<td>Patients excluded from analysis*</td>
<td>69</td>
</tr>
<tr>
<td>Patients available for analysis</td>
<td>1,404</td>
</tr>
</tbody>
</table>

*Reasons for exclusion: samples with missing patient numbers: 4, patients with double measurements: 15, patients not identified in survival data: 9, patients with a DFS of 1 d: 20, ineligible patients other than cardiovascular reasons and wrong stage: 19.
percentage present calls (ranging 24.9-42.2%) with on average 3% more calls with the freshly cut samples. For the purposes of molecular tumor classification or biological pathway analysis, we deem a present call of at least 20% as acceptable. Principle component analysis showed that stored and freshly cut samples from the same block cluster together tightly, indicating very similar transcriptional profiles.

**Statistical considerations.** Two strategies were followed for the data analysis. A set of standard approaches was formulated in an analysis plan and the analyses were conducted following these methods, with amendments imposed when data characteristics did not match the expectations. For example, most of the biomarker data turned out to have only a very small number of well-populated levels of measurement, so that systematic ROC analysis for optimal cutoffs and was not very informative. This included the classic univariate and subgroup analyses dictated by the two treatment arms and the distinction between stage II and stage III tumors.

Additional analyses were devised later, driven by hypothesis based on new literature and induced by intermediate results of multivariate model building. Data and results published in the meantime suggested for example the importance of distinguishing tumors depending on microsatellite stability into MSI and microsatellite stable classes with differing characteristics. Statistical analysis of survival data showed the expected strong relationship between tumor-node-metastasis stages and survival, in particular, the strong associations between node positivity and tumor size with earlier relapse. Thanks to the relatively large sample size of the study, confidence intervals were relatively small allowing fairly precise estimation of the hazard ratios, the estimates being in agreement with previous reports, and \( P \) values of tests of a null hypothesis of no effect were exceedingly small.

Survival effects of biomarkers were of weaker magnitude, so that multivariate model building could not be based uniquely on purely statistical considerations. An intense interplay between analysts and medical experts was essential to guide the model-building and interpretation process.

A major challenge in the design of the analyses and the interpretation of the results was the substantive need for stratified analyses for interpretation of results and for assessing medically relevant questions while avoiding inflation of false-positive conclusions. Avoiding to generate too many not well-supported hypothesis, while not missing new important insights by limiting the set of analyses too drastically was found difficult.

Other problems were the high number of missing or non-informative values for some biomarkers (such as the LOH markers), with the risk of biased estimation if the missing values were not missing at random, the need to consider more sophisticated methods of analysis in view of the increasingly complex data and difficulties in defining marker status coherently (such as for 18q LOH).

**Discussion**

Earlier biomarker development studies have used clinical trials as infrastructural basis for collection of patient data and biological samples, the two essential elements of this type of translational study (28). A similar approach was recently published, focusing on predictive biomarkers in metastatic colorectal cancer and in the context of a clinical trial exploring the efficiency of palliative chemotherapy (11). We report one of the largest of such studies as yet done in curatively treated stage II and III colon cancer with the involvement of a wide range of analytic techniques. Our experience shows that when the translational study is conceived along with the trial design, and sample procurement is included in the operating procedures established in the trial, a high percentage of samples can be obtained. General informed consent allowed the addition of new biomarkers to the study. Requests for the study of new markers or analytic procedures were submitted to the trial steering committee, allowing external oversight of the validity of new projects. Several groups were involved in the marker studies, which seemed very fortunate given the load of work and the expansion of scope of the projects over time, with the availability of new technology. Productive evolution of the study has depended largely on this approach.

Immunohistochemistry did not do better than DNA and RNA analysis, contrary to what we originally expected. Immunohistochemistry had a high success rate for well-established markers (for p53 and SMAD4 98%), but for newer markers for which staining and scoring procedures had to be developed, this was not the case (for TERT 56%). In contrast, our DNA analysis procedures were robust and highly satisfactory, with biomarker abnormality scores in the range of those published,

---

**Table 2. Number of cases available and frequency of alterations observed for each marker**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Patients</th>
<th>%</th>
<th>% abnormal</th>
<th>Published range %</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTERT (&gt;15% of the cells positive)</td>
<td>818</td>
<td>56</td>
<td>46</td>
<td>N.a.</td>
</tr>
<tr>
<td>TS (&gt;25% cells positive)</td>
<td>1,254</td>
<td>86</td>
<td>34</td>
<td>N.a.</td>
</tr>
<tr>
<td>SMAD4 (&gt;35% of cells)</td>
<td>1,428</td>
<td>98</td>
<td>12</td>
<td>13-63</td>
</tr>
<tr>
<td>PS3 (&gt;35% of cells positive)</td>
<td>1,432</td>
<td>99</td>
<td>35</td>
<td>25-76</td>
</tr>
<tr>
<td>MSI</td>
<td>1,296</td>
<td>89</td>
<td>15</td>
<td>10-17</td>
</tr>
<tr>
<td>KRAS</td>
<td>1,345</td>
<td>92</td>
<td>37</td>
<td>32-40</td>
</tr>
<tr>
<td>BRAF</td>
<td>1,353</td>
<td>93</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>LOH in 18q: at least one SNP informative</td>
<td>1,220</td>
<td>84</td>
<td>69</td>
<td>70</td>
</tr>
<tr>
<td>Histologic grading</td>
<td>1,439</td>
<td>99</td>
<td>N.ap.</td>
<td>N.ap.</td>
</tr>
<tr>
<td>UGT1A1</td>
<td>1,306</td>
<td>90</td>
<td>N.ap.</td>
<td>N.ap.</td>
</tr>
<tr>
<td>Total</td>
<td>1,452</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: N.a., no data available in the literature; SNP, single nucleotide polymorphism; n.ap., not applicable for this marker.
confirming the reliability of the obtained results. The results of the mRNA expression profiling pilot study were surprisingly positive. The amount of RNA that could be extracted from one 5-μm section was sufficient and the storage conditions were found less important than has been suggested in the literature (29). It has been suggested to dip cut sections in paraffin to reduce degradation of mRNA. We found, however, no difference between the quality of RNA obtained from freshly cut sections and that from sections that had been stored at 4°C for >2 years. Further experiments with actual profiling and comparing tissue samples between different institutions will clarify whether or not mRNA expression profiling constitutes an approach that can be reliably applied to FFPE samples collected in multicenter trials. How array-based (epi)genome-oriented studies will fare remains to be established, although our recent pilot experiments (data not shown) indicate that array genomic hybridization can be reliably done, in agreement with published data (30–32).

Finally, although we disposed of a large series of cases, with a multitude of variables, subgroups were occasionally too small to attain statistically significant results. As a consequence, we could not perform the ROC approach for cut-point definition as foreseen. In addition, for the same reason, validation of our results in separate learning and test sets was not attainable. This could not perform the ROC approach for cut-point definition as foreseen. In addition, for the same reason, validation of our results as foreseen. In addition, for the same reason, validation of our results in separate learning and test sets was not attainable. This will have to be done on separate but similar case collections. Furthermore, with the application of high throughput methods to large case collections, bioinformatics expertise becomes indispensable for adequate data analysis.

In conclusion, our experience shows that:

1. Translational studies in the context of multicenter trials constitute a promising tool for the development of clinically useful biomarkers; general informed consent is essential for optimal use of the collected biospecimens.
2. For optimal design, these studies need to be conceived along with the development of the trial and tissue collection should be an integral component of the trial.
3. The wealth of possibilities using this approach is such that these studies are best done in a multidisciplinary consortium approach rather than within the experience available in a single research group.
4. Given appropriately adapted protocols, FFPE can be used in high throughput analysis platforms including mRNA expression profiling.

5. Immunohistochemical markers might be less reliable than genomic markers, largely due to important heterogeneity in tissue processing techniques and problems in quantification; immunohistochemistry, however, offers the opportunity of detecting markers in situ.
6. Advanced bioinformatics expertise needs to be incorporated into the study group as of the initiation of the study.

Disclosure of Potential Conflicts of Interest

F. Bosmann and A. Roth have received a commercial research grant from Pfizer; A. Roth, S. Teijpar, and E. van Cutsem are consultants for Pfizer; R. Kennedy is employed by ALMAC.

Acknowledgments

We thank Dr. Mauro Delorenzi for providing essential bioinformatics support and critical reading of the manuscript.

References


Tissue Biomarker Development in a Multicentre Trial Context: a Feasibility Study on the PETACC3 Stage II and III Colon Cancer Adjuvant Treatment Trial

Fred T. Bosman, Pu Yan, Sabine Tejpar, et al.

Clin Cancer Res Published OnlineFirst August 18, 2009.

Updated version Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-09-0741