Practical and Robust Identification of Molecular Subtypes in Colorectal Cancer by Immunohistochemistry

Anne Trinh1,2, Kari Trumpi3, Felipe De Sousa E Melo4,5, Xin Wang1,6, Joan H. de Jong4, Evelyn Fessler4, Peter J.K. Kuppen4, Marlies S. Reimers6, Marloes Swets7, Miriam Koopman3, Iris D. Nagtegaal8, Marnix Jansen9,10, Gerrit K.J. Hooijer9, George J.A. Offerhaus3, Onno Kranenburg3, Cornelis J. Punt11, Jan Paul Medema4, Florian Markowitz7, and Louis Vermeulen4

Abstract

Purpose: Recent transcriptomic analyses have identified four distinct molecular subtypes of colorectal cancer with evident clinical relevance. However, the requirement for sufficient quantities of bulk tumor and difficulties in obtaining high-quality genome-wide transcriptome data from formalin-fixed paraffin-embedded tissue are obstacles toward widespread adoption of this taxonomy. Here, we develop an immunohistochemistry-based classifier to validate the prognostic and predictive value of molecular colorectal cancer subtyping in a multicenter study.

Experimental Design: Tissue microarrays from 1,076 patients with colorectal cancer from four different cohorts were stained with colorectal cancer from four different cohorts were stained for five markers (CDX2, FRMD6, HTR2B, ZEB1, and KER) by immunohistochemistry and assessed for microsatellite instability. An automated classification system was trained on one cohort using quantitative image analysis or semiquantitative pathologist scoring of the cores as input and applied to three independent clinical cohorts.

Results: This classifier demonstrated 87% concordance with the gold-standard transcriptome-based classification. Application to three validation datasets confirmed the poor prognosis of the mesenchymal-like molecular colorectal cancer subtype. In addition, retrospective analysis demonstrated the benefit of adding cetuximab to bevacizumab and chemotherapy in patients with RAS wild-type metastatic cancers of the canonical epithelial-like subtypes.

Conclusions: This study shows that a practical and robust immunohistochemical assay can be employed to identify molecular colorectal cancer subtypes and uncover subtype-specific therapeutic benefit. Finally, the described tool is available online for rapid classification of colorectal cancer samples, both in the format of an automated image analysis pipeline to score tumor core staining, and as a classifier based on semiquantitative pathology scoring.

Introduction

Colorectal cancer is a heterogeneous disease with an overall 5-year survival of below 60% (1). There is an urgent need to improve selection of early-stage patients who may benefit from adjuvant therapy, or to identify patients with metastasis who may profit from a specific targeted therapy. To facilitate this, stratification methods based on histopathologic characteristics are extensively implemented. For example, only patients with colorectal cancer with high-risk features such as high-grade and poorly differentiated morphology are believed to benefit from adjuvant chemotherapy (2). Although histopathologic classification is difficult to implement uniformly, associations with molecular characteristics have been noted, such as microsatellite instability (MSI) in serrated tumors (3). This provides a more robust/objective means of determining the suitability of a patient for a given therapy: For example, mutation in the KRAS/BRAF axis is a well-characterized determinant of resistance to anti-EGFR therapy in metastatic disease (4, 5). However, current mutational profiling provides only limited biomolecular understanding of the disease, particularly in chromosomal unstable disease where the large heterogeneity in patient response to
**Translational Relevance**

The recent stratification of colorectal cancer transcriptional profiles into subtypes with prognostic and predictive differences has an immediate clinical implication in personalization of therapies. However, a subsequent challenge is the adaptation of this classification system for diagnostic purposes given the requirement for a rapid scoring system, which uses minimal quantities of tumor material. This study resolves these issues, reporting an immunohistochemistry-based classifier, which uses either pathologist scoring or automated image analysis of tissue microarrays as inputs. Not only does this approach improve clinical utility of the current molecular taxonomy, it also allows retrospective access to large clinical cohorts for which only formalin-fixed paraffin-embedded material is available. Our retrospective analysis of four cohorts has validated the prognostic value of colorectal cancer subtyping. Furthermore, we have identified a subset of patients (approximately 30%) who benefitted from anti-EGFR therapy, potentially improving the efficacy of this class of drugs.

**Materials and Methods**

**Human colorectal cancer tissue specimens**

Four independent patient cohorts were used in this study: the AMC-AJCCII-90 (6), LIIMC (16), CAIRO (Trial Registration ID: NCT00312000; ref. 14), CAIRO (Trial Registration ID: NCT00208546; ref. 15) series, for which clinicopathologic characteristics are described in Table 1. The training set (AMC-AJCCII-90; ref. 6) comprised 90 stage II patients, for which 75 had adequate quantities of bulk tumor to perform IHC staining. Three 0.6-mm biopsies, which were stained for the five biomarkers of interest, were obtained from each patient. Exclusion criteria included damage to TMA cores and incomplete sets of cores, leaving a training set of 70 patients (Supplementary Fig. S1A).

The three validation sets on which CMS-IHC classification was performed included the CAIRO cohort (14), CAIRO cohort (15), and LIIMC cohort (16). A single 2-mm × 4-µm TMA core was available for each patient from the CAIRO and CAIRO series (17), whereas 3- × 0.6-mm × 4-µm cores were available for the LIIMC series (16). Patient material from all cohorts were fixed in formalin and embedded in paraffin. Exclusion criteria included insufficient primary material, unknown MSI status, and an incomplete set of IHC-stained cores following quality control. Of the 353 (LIIMC), 803 (CAIRO), and 559 (CAIRO) patients with clinical information, 240, 426, and 340 patients, respectively, were used in this analysis (Supplementary Fig. S1A). Comparison of clinical covariates of all cohorts prior to and after exclusion did not show any selection bias in terms of age, sex, or mutational status (6, 14–16).

**IHC staining, MSI status, and image acquisition**

Four markers were selected from previous transcriptomic analysis (6) for IHC staining in this study: (i) CDX2, a marker for differentiation which is expected to be highly expressed in epithelial-like tumors, (ii) HTR2B, which was shown to have high expression in mesenchymal-like tumors, (iii) FRMD6, a marker for goblet cells expressed in mesenchymal-like tumors, and (iv) ZEB1, a marker for EMT. In addition, pan-cytokeratin was selected to normalize the other markers for tumor content, which itself is expected to be higher in epithelial-like tumors.
Table 1. Summary of patient information from each patient cohort used in this study

<table>
<thead>
<tr>
<th>Clinical Information</th>
<th>AMC-AJCCII-90</th>
<th>LUMC</th>
<th>CAIRO</th>
<th>CAIRO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>90</td>
<td>353</td>
<td>820</td>
<td>559</td>
</tr>
<tr>
<td>Median age at surgery</td>
<td>73.4 (34.6–95.1)</td>
<td>68 (35.2–85.6)</td>
<td>63 (41–78)</td>
<td>62 (41–76)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>48 (53%)</td>
<td>179 (57%)</td>
<td>300 (37%)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>42 (47%)</td>
<td>174 (49%)</td>
<td>596 (63%)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td></td>
<td>4 (0.5%)</td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td>I</td>
<td>68 (19%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>133 (38%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>92 (26%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>55 (15%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>7 (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiation</td>
<td>G1 well</td>
<td>67 (19%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2 moderate</td>
<td>183 (58%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3 poor</td>
<td>27 (7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>76 (22%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment arm (CAIRO1/2)a</td>
<td>Arm A</td>
<td>410 (50%)</td>
<td>282 (50%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arm B</td>
<td>410 (50%)</td>
<td>277 (50%)</td>
<td></td>
</tr>
<tr>
<td>Vital statusb</td>
<td>Alive</td>
<td>124 (35%)</td>
<td>80 (30%)</td>
<td>124 (22%)</td>
</tr>
<tr>
<td></td>
<td>Death</td>
<td>229 (65%)</td>
<td>740 (90%)</td>
<td>435 (78%)</td>
</tr>
<tr>
<td>Median survival (months)c</td>
<td>39 (1.9 - 113.7)</td>
<td>88 (9 - 295.2)</td>
<td>15.2 (13 - 45.2)</td>
<td>20.4 (14.5 - 54.7)</td>
</tr>
<tr>
<td>Mutation information</td>
<td>MMR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSS</td>
<td>65 (72%)</td>
<td>215 (61%)</td>
<td>504 (61%)</td>
</tr>
<tr>
<td></td>
<td>MSI</td>
<td>25 (28%)</td>
<td>35 (10%)</td>
<td>19 (2%)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>103 (29%)</td>
<td>297 (36%)</td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td>Wild-type</td>
<td>70 (78%)</td>
<td>189 (23%)</td>
<td>321 (57%)</td>
</tr>
<tr>
<td></td>
<td>Mutation</td>
<td>20 (22%)</td>
<td>174 (21%)</td>
<td>208 (37%)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>457 (56%)</td>
<td>30 (5%)</td>
<td></td>
</tr>
<tr>
<td>BRAF</td>
<td>Wild-type</td>
<td>73 (81%)</td>
<td>332 (40%)</td>
<td>473 (85%)</td>
</tr>
<tr>
<td></td>
<td>Mutation</td>
<td>17 (19%)</td>
<td>23 (3%)</td>
<td>45 (8%)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>460 (56%)</td>
<td>41 (7%)</td>
<td></td>
</tr>
<tr>
<td>Classification</td>
<td>Number of patients excluded due to missing MSI data</td>
<td>0</td>
<td>103 (29%)</td>
<td>297 (36%)</td>
</tr>
<tr>
<td></td>
<td>Number of CM1 patients</td>
<td>25 (28%)</td>
<td>35 (10%)</td>
<td>19 (2%)</td>
</tr>
<tr>
<td></td>
<td>Number of patients with TMA cores for epithelial vs. mesenchymal classification</td>
<td>52 (58%)</td>
<td>256 (58%)</td>
<td>458 (56%)</td>
</tr>
<tr>
<td></td>
<td>Remaining patients with TMA cores following QC</td>
<td>49 (54%)</td>
<td>205 (58%)</td>
<td>407 (50%)</td>
</tr>
<tr>
<td></td>
<td>Number of cores used in study</td>
<td>121</td>
<td>613</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>Mean number of cores per patient</td>
<td>2.5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Number of classified patients</td>
<td>70 (78%)</td>
<td>240 (68%)</td>
<td>425 (52%)</td>
</tr>
</tbody>
</table>

aTreatment arm applicable to CAIRO2 Studies: Arm A is sequential therapy in CAIRO1. Arm B is combination therapy in CAIRO1. CAIRO2: Arm B is the additional cetuximab treatment.
bDFS used in AMC and LUMC cohorts. OS used in CAIRO cohorts.

cAMC and LUMC sets: right censored at 60 months to take into account deaths due to natural causes.

TMA slides of the various cohorts were stained with anti-HTR2B (1:75; Sigma; HPA012867), anti-FRMD6 (1:1500; Sigma; HPA027524), anti-CDX2 (1:200; Novus Biologicals; NB100-2136), anti-ZEB1 (1:500; Sigma; HPA027524), or anti-cytokeratin (AE1/AE3; 1:500; Thermo Scientific). After a secondary incubation with anti-rabbit-HRP or anti-mouse-HRP (Power- vision), staining was developed using DAB–rabbit-HRP or anti-mouse-HRP (PowerVision), counterstained with hematoxylin. Individual cores were scored by three trained observers (K. Trumpi, M. Jansen, and G.J.A. Offerhaus) for CDX2, FRMD6, HTR2B, KER intensity and content, and ZEB1 nuclear content in epithelial cells blinded for CMS subtype. Digital images of TMA slides from the AMC-AJCCII-90 and CAIRO2 series were acquired using the Olympus dotSlide system (Olympus). For the LUMC and CAIRO cohorts, an Aperio scanscope XT system (Leica Biosystems) was used. MSI status from the AMC and LUMC cohorts was identified using MSI analysis system (Promega; refs. 6, 16). In both CAIRO cohorts, MSI status was identified by IHC with antibodies against hMLH1, hMSH2, hMSH6, and hPMS2 as previously described (17, 18).

Automated image-analysis based classification

A classifier designed to distinguish colorectal cancer subtypes was trained using staining information from the AMC-AJCCII-90 cohort and then applied to three different patient cohorts (Supplementary Fig. S2A). This was first conducted using an automated image analysis approach, whereby digital images of TMA cores were segmented and quantitated as previously described (6) and summarized in Supplementary Fig. S2B. Brieﬂy, all segmented cores underwent intensity adjustment by rescaling each color channel to a [0, 255] range to ensure the brightness is consistent across samples. An entropy ﬁlter was applied to determine the main TMA area. A complete TMA core would typically occupy 40% to 60% of an image, and broken cores occupying less than 10% of the image were discarded. The image was deconvolved into hematoxylin-DAB color space using Ruifrok’s method (19). Stained regions were then determined by Otsu thresholding of the DAB channel, with a minimum threshold value of 0.3 employed to ensure that regions with low staining were not falsely considered as positive (20).
Pixel-level features were extracted from the image analysis pipeline instead of cell-level features due to difficulties in accurate segmentation. Prior to standardization with pan-cytokeratin, four parameters were extracted for each stain: (i) the total area occupied by positively stained cells, (ii) the fraction of the TMA core occupied by positively stained pixels, (iii) the average intensity of these pixels, and (iv) a combined area-intensity metric computed by multiplying the intensity with the area fraction. In addition, two "keratin-standardized" values were computed to take into account epithelial content: (v) a standardized TMA fraction (i.e., feature (ii) divided by the corresponding keratin TMA area fraction) and (vi) a standardized stain-content score (SCnorm) defined as:

$$SC_{\text{norm}} = \frac{\text{Intensity}_X \times \text{Area}_X \times \text{CoreArea}_{\text{KER}}}{\text{Intensity}_{\text{KER}} \times \text{Area}_{\text{KER}} \times \text{CoreArea}_X}$$

Thus, a total of 28 inputs were used for the classifier (six from each primary stain: CDX2, FRMD6, HTR2B, ZEB1 and four from pan-cytokeratin), which were representative of both tumor staining and total TMA staining (Supplementary Table S1).

The stains were selected based on differential expression between different subtypes (i.e., strongly positive staining or absent staining) and do not follow a normal distribution. Thus, to adjust for differences in sample preparation, core size, staining, and image capture between different cohorts, all features were rescaled to ensure 95% of values lay within a [0, 1] range (Supplementary Fig. S2C). The keratin-normalized features may have values outside of this range due to heterogeneity between neighboring sections; thus, values outside this range were right truncated to 1.5.

To classify patients into their colorectal cancer subtype, MSI status was first used to define patients which belong to the CMS1 subtype. The remaining patients were classified into "epithelial" (CMS2/3) or "mesenchymal" (CMS4) subtypes using a random forest classifier (21) trained on an AMC-AJCCII-90 set of 49 patients and 121 cores using all features extracted from image analysis. True labels were attained from gene-expression data (ref. 6; Supplementary Fig. S2D).

Briefly, a random forest classifier subsamples the dataset (two-thirds of patient samples) to construct a decision tree to separate CMS2/3 and CMS4 samples. This is repeated 1,000 times to construct a classification “forest.” The prediction probability for each patient sample is thus the frequency of classification of CMS2/3 compared with CMS4 in this “forest.” Cores with a random forest probability of 60% were scored as “mesenchymal.” Patient subtypes were determined using majority consensus (Supplementary Fig. S2E). Survival information for the cohorts was supplied after CMS classification was conducted. The required code to produce the pipeline is supplied as Supplementary Information.

**Semiquantitative classification system**

Decision rules implemented in the automated random forest classifier were extracted. The most common rules from 1000 decision trees were retained to construct a classifier, which requires semiquantitative pathologist scoring as input. Numeric thresholds in the AMC-AJCCII-90 set were compared with pathologist scoring to determine the appropriate equivalent semiquantitative thresholds. A random forest classifier was then constructed and applied to pathologist scoring in the CAIRO2 cohort. Detailed scoring guidelines are presented in the online classification tool (Supplementary Fig. S3; crcclassifier.shinyapps.io/appTesting).

**Comparison to the serrated pathway classification system**

Patient samples were classified based on mutational information into one of four pathways as previously described by Leggett and Whitehall (3): the “traditional” class is CIMP+, MSS, KRAS wt, and BRAF wt. The "alternate" pathway is characterized by MSS, KRAS mut, BRAF mut, and CIMP+. Tumors in the "serrated" pathway are BRAF mut, CIMP-, KRAS wt independent of MSI status. All remaining patients were grouped into an "unknown" category.

**Statistical and survival analysis**

All statistical analyses were conducted in R (22), and documented code to fully reproduce the study is supplied as Supplementary Information. Concordance between semiquantitative pathologist scoring and automated image analysis was conducted using a two-tailed Jonckheere Terpstra test for trend with significance assessed using 1,000 permutations of the data. Consistency in pathologist scoring was computed using intraclass correlation coefficients using a two-way mixed effects model.

Visualization of the directionality of markers in relation to the CMS2/3 and CMS4 subtypes was conducted using principal component analysis of all features. The directionality of each stain was illustrated by the keratin-normalized scores (SCnorm). Overall survival (OS) and disease-free survival (DFS) analyses were performed using multivariate Cox proportional hazards models accounting for age, stage, and sex, with colorectal cancer-specific 5-year follow-up, after which samples were right censored. Differences in survival were expressed as HRs with 95% confidence intervals and median survival time. Significance was tested using the log-rank test. Survival curves were calculated using the Kaplan–Meier method.

**Results**

**Developing the CMS-IHC classifier**

We have previously developed an IHC assay to identify molecular subtypes of colorectal cancer in the AMC-AJCCII-90 colorectal cancer patient series (6), and here we aim to optimize it for application to three external datasets: the LIJMC (16), CAIRO (14), and CAIRO2 (15) datasets (Supplementary Fig. S2). In developing a classification system, CMS1 patients were first separated using MSI status as it was observed to be almost ubiquitous in these patients. This leaves a two-class classification problem in separating mesenchymal-like CMS4 patients from epithelial-like CMS2/CMS3 patients. To distinguish between these two subtypes, a panel of four IHC stains (CDX2, FRMD6, HTR2B, and ZEB1) was selected based on differential gene expression. In addition, pan-cytokeratin (KER) was used to normalize for epithelial content.

The IHC staining procedures were first assessed by comparing transcriptome-based subtyping to semiquantitative scoring by pathologists in each patient (Fig. 1A). Epithelial-like TMA cores displayed increased CDX2 expression reflecting a higher degree of differentiation (23) and lower ZEB1 expression (24). CDX2 displayed a strong association with pathologist scoring of cellular differentiation ($\chi^2$ test, $P < 0.001$), and epithelial ZEB1 expression was more common in poorly differentiated tumors albeit not significantly (Supplementary Fig. S4A), supporting their use as...
markers in differentiating epithelial from mesenchymal tumors. HTR2B, a vasoactive neurotransmitter previously associated with hepatocellular tumor growth (25), is expressed in both epithelial-like and mesenchymal-like patients, but a higher intensity is observed in mesenchymal-like cases. FRMD6, which is expressed in goblet cells, also has higher expression in mesenchymal-like patient samples. These expression patterns are in line with the gene-expression levels of the corresponding genes as previously demonstrated (6).

To develop an automated classifier, quantitative measurements were first benchmarked against pathologist scoring to ensure that the same increasing trend observed in semiquantitative pathologist scoring could be captured in the quantitative data (Fig. 1B). All markers illustrated agreement between the two metrics (Jonckheere Terpstra test, \( P = 0.002 \)) with the exception of ZEB1, possibly due to expression in both the stroma and epithelium.

Similar results were observed using stain area or stain intensity alone (Supplementary Information).

No single stain demonstrated a clear distinction between the subtypes, and some interobserver variation in stain scoring was noted (Supplementary Fig. S4B and S4C), motivating the use of a combination of stains using an automated pipeline. Over 100 permutations of threefold cross-validation, individual features had higher error rates and variance in prediction accuracy compared with a classifier that encompasses all features (Supplementary Fig. S4D), motivating the inclusion of all stains into a final classifier. The five stain-based CMS-IHC classifier trained on the AMC-AJCCII-90 series demonstrated an out-of-bag error rate of 20% on an individual core level and 87% concordance between the gold-standard transcriptome-derived classifier and CMS-IHC classifier on a patient level (Fig. 1C). Furthermore, we successfully validated the association of patients classified as mesenchymal-like...
like with a dismal prognosis in stage II colorectal cancer. These patients had a median DFS of 14.5 months (Fig. 1D; HR, 6.73; 95% CI, 1.86–24.29; P = 0.007; log-rank test).

Finally, we compared the CMS classification system with another classification of colorectal cancer into “traditional,” “serrated,” and “alternative” subtypes using genomic information including BRAF, KRAS, MSI, and CIMP status (3). The CMS2/3 subtype showed a strong association with the “traditional” subtype characterized by the lack of mutations, and the CMS1 subtype showed interrelation with the “serrated” subtype (Supplementary Fig. S5A). However, this combination of four mutations cannot effectively identify mesenchymal CMS4-like patients, and the same prognostic differences observed using the CMS system could not be recapitulated (Supplementary Fig. S5B), motivating the use of information derived from transcriptomic data to distinguish CMS4 from CMS2/3.

**Classification of LUMC, CAIRO, and CAIRO2 cohorts by CMS-IHC**

TMAs from the LUMC, CAIRO, and CAIRO2 datasets were stained and classified using the AMC-AICCI-90-trained CMS-IHC classifier. Comparison of CMS-IHC classified cores in our validation sets demonstrated the same subtype distinctions in staining as the AMC-AICCI-90 series (Fig. 2A). To further support the notion that epithelial-like and mesenchymal-like cancers represent biologically distinct entities, principal component analysis using all extracted features highlighted that the principal direction of CDX2 expression was in line with the epithelial-like subtype (Fig. 2B). The directions of the HTR2B, FRMD6, and ZEB1 expression vectors were in line with the mesenchymal-like subtype, concordant with the differentiation observed by visual inspection. In addition, tumor budding was seen in some samples, and low CDX2 expression was noted in the corresponding serial sections (Supplementary Fig. S4D). This feature has previously been associated with poor prognosis (26), and is consistent with a mesenchymal-like phenotype.

**Prognostic value of CMS-IHC classification in three independent cohorts**

The training cohort of stage II patients demonstrated a CMS2/3:CMS4 ratio of approximately 2:1 using the “gold-standard” transcriptomic-based classification. Compared with this distribution, our validation datasets had an increased proportion of mesenchymal-like patients, consistent with these cohorts containing late stage patients (ref. 11; Fig. 2C). The LUMC series, comprising a roughly even distribution of stage I–IV patients, demonstrated an equal proportion of epithelial-like to mesenchymal-like patients (43% each). A reduced proportion of MSI+ patients (CAIRO 4%, CAIRO2 2%) and higher mesenchymal-like ratio (CAIRO 36%, CAIRO2 47%) was observed in the CAIRO cohorts, probably reflecting the fact that MSI+ tumors generally have a good prognosis and rarely metastasize (18).

In all cohorts, the mesenchymal-like subtype displayed significantly worse survival in a Cox proportional hazards model accounting for confounding variables including age, sex, stage, and treatment arm (Table 2). The mesenchymal-like arm in the LUMC cohort displayed a median DFS of 24 months (HR, 1.77; 95% CI, 1.20–2.62; P < 0.001; log-rank test; Fig. 2D). Both CAIRO sets, comprising stage IV patients, had lower mesenchymal-like median OS times compared with the epithelial-like subtype. In the CAIRO set, this value increased from 13.8 (95% CI, 12.5–16.4) to 19 (95% CI, 18.0–21.6) months. Similarly, the CAIRO2 cohort demonstrated an improvement from 20 months (95% CI, 15.5–22.9) to 23.8 months (95% CI, 21.7–27.4; Fig. 2E and F; Table 2).

In combination, these results confirm the notion that mesenchymal-like patients present with more advanced disease stage and worse disease outcome. This in turn demonstrates the utility of the CMS-IHC classifier in identifying a distinct molecular subtype of patients with colorectal cancer for whom dismal prognosis is a salient clinical feature.

**Predictive value of anti-EGFR therapy for "epithelial-like" cancers**

Following subtype classification, the CAIRO and CAIRO2 cohorts were revisited to determine whether there was a subtype-specific benefit in a specific treatment regimen. Our previous research suggests that mesenchymal-like tumors resist anti-EGFR therapy independent of KRAS/BRAF mutations (6), and we sought to validate this observation in an additional patient cohort: the CAIRO2 patient series, a clinical study to determine the efficacy of adding cetuximab, an anti-EGFR antibody, to a standard regimen of capecitabine, oxaliplatin, and bevacizumab in patients with advanced colorectal cancer (15).

KRAS/BRAF wild-type epithelial-like tumors illustrated an improved response to cetuximab therapy compared with mesenchymal-like tumors. Seventy-three percent of patients demonstrated partial or complete response to therapy, compared with only 50% in the mesenchymal-like case (χ² test; P = 0.1, Fig. 3A).

To determine whether this response to therapy translates to a long-term patient benefit, we evaluated patient survival with respect to treatment arm, subtype, and KRAS/BRAF mutation (Fig. 3B; Supplementary Table S2). OS of KRAS/BRAF-mutant patients was not significantly affected by cetuximab in both epithelial-like and mesenchymal-like subtypes, although a trend toward a detrimental effect was observed in both subtypes. Analysis of KRAS/BRAF wild-type patients demonstrated a significant beneficial effect of cetuximab in epithelial-like patients with median OS improving from 23 months (95% CI, 16.5–27.4) to 33.8 months (95% CI, 25.2–55.1). The HR for cetuximab-treated versus untreated cohort was 0.52 (95% CI, 0.31–0.87; P = 0.05, log-rank test). In contrast, no difference in survival of mesenchymal-like patients was observed in KRAS/BRAF wild-type patients (HR, 1.56; 95% CI, 0.91–2.65; P = 0.11; log-rank test), and a detrimental effect on survival was observed in mesenchymal-like patients harboring mutations in the KRAS/BRAF axis (HR, 1.75; 95% CI, 1.08–2.84; P = 0.06; log-rank test). This illustrates the utility of the CMS taxonomy to predict the efficacy of anti-EGFR therapy. More specifically, our analysis reveals a substantial group of patients (approximately 40%, comprising of mesenchymal-like) who despite lacking mutations in the KRAS/BRAF axis do not benefit from cetuximab therapy and in fact displayed a trend toward reduced OS (P = 0.12, log-rank test; Fig. 3B).

We performed a similar analysis on the CAIRO cohort to investigate if the combined or sequential administration of capcitabine, irinotecan, and oxaliplatin provides any subtype-specific responses but did not detect any significant differences (Supplementary Fig. S6).
Figure 2. Molecular and survival features of each subtype in validation cohorts. 

A, Staining of representative epithelial-like or mesenchymal-like patients in each cohort. Scale bar, 100 μm. 

B, Contribution of each stain to CMS separation illustrated on a PCA plot. 

C, Proportion of patients in each subtype for each data set. 

D, LUMC set, HR = 1.77 (95% CI, 1.20–2.62), P < 0.001. 

E, CAIRO cohort, HR = 1.39 (95% CI, 1.12–1.72), P = 0.03. 

F, CAIRO2 cohort, HR = 1.24 (95% CI, 0.96–1.59), P = 0.03. All P values calculated using the log-rank test.
Table 2. Contribution of each variable to multivariate cox proportional hazards models

<table>
<thead>
<tr>
<th>Clinical factor</th>
<th>CAIRO N (n event)</th>
<th>CAIRO2 N (n event)</th>
<th>LUMC N (n event)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–50</td>
<td>47 (40)</td>
<td>41 (31)</td>
<td>28 (12)</td>
</tr>
<tr>
<td>50–60</td>
<td>117 (107)</td>
<td>110 (83)</td>
<td>45 (17)</td>
</tr>
<tr>
<td>60–70</td>
<td>167 (144)</td>
<td>123 (102)</td>
<td>62 (43)</td>
</tr>
<tr>
<td>70+</td>
<td>95 (86)</td>
<td>54 (46)</td>
<td>105 (88)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>150 (131)</td>
<td>137 (107)</td>
<td>120 (76)</td>
</tr>
<tr>
<td>M</td>
<td>275 (246)</td>
<td>191 (155)</td>
<td>120 (84)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>38 (18)</td>
<td>38 (18)</td>
<td>38 (18)</td>
</tr>
<tr>
<td>II</td>
<td>92 (55)</td>
<td>92 (55)</td>
<td>92 (55)</td>
</tr>
<tr>
<td>III</td>
<td>70 (49)</td>
<td>70 (49)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Treatment arm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>210 (194)</td>
<td>161 (124)</td>
<td>120 (76)</td>
</tr>
<tr>
<td>B</td>
<td>216 (183)</td>
<td>167 (138)</td>
<td>120 (84)</td>
</tr>
<tr>
<td>CMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMS1</td>
<td>19 (16)</td>
<td>7 (6)</td>
<td>3.16 (1.36–7.34)</td>
</tr>
<tr>
<td>CMS2/3</td>
<td>253 (220)</td>
<td>167 (137)</td>
<td>3.16 (1.36–7.34)</td>
</tr>
<tr>
<td>CMS4</td>
<td>154 (141)</td>
<td>154 (119)</td>
<td>3.16 (1.36–7.34)</td>
</tr>
</tbody>
</table>
| Notes: CAIRO and CAIRO2 cohorts were trials for patients with metastasis.

Development of a pathologist-based classifier

In order to improve the portability of our classifier for use on an individual case basis, the features that most commonly appeared in decision rules were extracted and adapted to semiquantitative scores commonly assigned by pathologist scoring (Fig. 4A). This was trained on the AMC-AJCII-90 series and applied to the CAIRO2 cohort.

Comparison between the automated and semiquantitative classifiers highlighted a concordance of 78% in the CAIRO2 cohort, highlighting that the simplified classifier is capable of assigning subtype classes in the absence of continuous quantitative information (Fig. 4B).

To further assess the validity of this approach, survival analysis using this new classification system was conducted. The poor prognosis of mesenchymal-like subtype was illustrated using the pathologist-based classifier (HR, 1.34; 95% CI, 1.03–1.72; P = 0.003; log-rank test; Fig. 4C). In conjunction, across all KRAS/BRAF wild-type patients, the predictive value of the epithelial-like subtype for adjuvant cetuximab therapy compared with all other patient arms was highlighted (HR, 0.58; 95% CI, 0.40–0.84; P = 0.02; log-rank test; Fig. 4D). These results support the successful development of a portable version of our classifier compatible with pathologist-based semiquantitative scoring. We have made our classifier available as a free-to-use online resource, as shown in Supplementary Fig. S3 (https://crrclassifier.shinyapps.io/app-Testing/).

Discussion

Over the past decade, the stratification of patients into distinct molecular subtypes has been achieved in a number of cancers (27) including previous work on colorectal cancer (6–12). Such discrimination has been driven primarily by gene-expression profiling, where the requirements for sufficient bulk tumor, cost, and time impede widespread diagnostic adoption. Thus, there is an urgent need for rapid and cost-effective surrogates for gene-expression profiling in both the clinical setting and in translational research. Such tools are essential for the identification of patient subtyping to guide subsequent treatment and to access large patient cohorts from previous clinical trials for which only FFPE tissue available, but remain invaluable resources for biomarker and validation studies.

We have demonstrated as a proof of principle, the adaptation of a 146 gene-expression signature into a panel of five biomarkers as a potential diagnostic tool for the classification of patients in colorectal cancer. This method utilizes IHC staining on more readily available TMAs, and automated image analysis and classification to deliver objective and accurate scoring in our training set. The decision rules used within this classifier were simplified to permit pathologist scoring as an input, allowing for the identification of mesenchymal-like patients on a case-by-case basis. The portability of the method was highlighted in the successful classification of patients into distinct molecular subtypes in three independent cohorts, each validating the poor prognosis of the mesenchymal-like subtype independent of age, sex, and stage. In addition, we have demonstrated the predictive value of epithelial-like subtyping for anti-EGFR therapy in combination with anti-VEGF therapy and chemotherapy.

A current major challenge is the selection of patients that benefit from adding anti-EGFR agents to combination therapies as a first-line treatment for metastatic colorectal cancer. Although it has been established that patients displaying (K)RAS or BRAF mutations do not benefit from anti-EGFR therapy, and even may have worse disease outcome, a large proportion of patients who are wild type for these genes also do not benefit (15, 28). Our retrospective analysis of the CAIRO2 clinical trial assessing the efficacy of combined capecitabine and bevacizumab therapy with cetuximab (15) demonstrated a therapeutic benefit only in KRAS/BRAF wild-type epithelial-like patients, but not in KRAS/BRAF wild-type mesenchymal-like cancers. This finding has the possibility to further reduce the patients eligible for anti-EGFR therapy by approximately 40%, and thereby, importantly, increasing the efficacy of this class of drugs. Further randomized clinical trials are necessary to validate this effect. In contrast, no significant benefit
in combination compared with sequential therapy was observed in the CAIRO cohort.

Of note, our study aimed at generating a practical tool to classify patients with colorectal cancer into distinct molecular disease subtypes, rather than developing a prognostic or predictive biomarker assay *per se*. Our focus in particular was to separate patients based on differences in prognosis in a step toward determining suitable treatment options for each subtype. Currently, our classifier does not distinguish between different epithelial-like subtypes which have similar prognosis (i.e., the newly...
characterized canonical Wnt signaling CMS2 and metabolic CMS3 subtypes (ref. 11); however, we plan to extend our classifier to include suitable metabolic markers such as GLUT1 to assist in this discrimination. In addition, while no single marker demonstrated definitive separation between subtypes, the use of a trained algorithm allowed us to stratify large patient cohorts. Given that markers were selected based on transcriptome-based profiling rather than protein-based profiling (29), the accuracy of the classifier could be further improved using proteo-genomic approaches to select for optimal markers (10).

Additional work in developing standardized guidelines will be required to realize the clinical potential of this assay. First, our training set consisted of a small cohort of stage II tumors, and expansion of the training set to include patients with different clinicopathologic features from multiple centers will be required to minimize biases associated with such covariates. However, recent consensus subtyping has suggested that this classification system is largely independent of existing clinical parameters including stage and grade. A slight enrichment of late-stage patients in mesenchymal-like tumors was reported, consistent with our results (11). Second, the use of automated image-based classification requires standardization of protocols for biopsy sampling, TMA staining, and image analysis in order to minimize batch effects to ensure accurate implementation (30). Although scoring by a pathologist can take into account these differences, there is also the caveat of intertumor and intrascorer subjectivity, which has previously shown to vary results by up to 30% (31). In our pilot cohort, moderate consistency measured by ICC supports the use of the selected stains in a pathologist-based classifier. However, consistency could be further improved by standardizing processing guidelines, developing scoring criteria, and assessing the robustness of biomarkers to pathologist scoring. Nonetheless, a consistency of 78% in patient classification was attained by pathologist-based and an automated method, highlighting the promise of the system. Third, our current IHC assay, as is generally the case in pathology, is further complicated by the issue of intratumor heterogeneity in tissue specimens from the same patient, which affects our downstream subtype calling. We have used majority consensus in the presence of conflicting cores; however, the implications of tumor heterogeneity will need to be formally assessed to improve clinical utility of this classification method. Nonetheless, the development of a rapid IHC-based screening tool as a surrogate for gene-expression profiling...
Colorectal Cancer Subtypes by Immunohistochemistry

is a major step forward, demonstrating both clinical and research utility in allowing access to previous clinical trials to develop effective subtype-specific treatments for colorectal cancer.

Disclosure of Potential Conflicts of Interest
F. Markowetz reports receiving speakers bureau honoraria from Bayer Berlin. No potential conflicts of interest were disclosed by the other authors.

Data and Materials Availability
Code for the image analysis pipeline and development of the CMS-IHC classifier is available at: github.com/trinharn/crc-ihc-classification. The pathologist-based classifier is hosted at ccreclassifier.shinyapps.io/app/testing

Authors’ Contributions
Conception and design: A. Trinh, P.J.K. Kuppen, J.P. Medema, L. Vermeulen
Development of methodology: A. Trinh, K. Trumpi, J.H. de Jong, L. Vermeulen
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. Trumpi, F. De Sousa E Melo, E. Fessler, M.S. Reimers, M. Swets, M. Koopman, J.D. Nagtegaal, M. Jansen, G.J.A. Offerhaus, C.J. Punt
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Trinh, K. Trumpi, X. Wang, M.S. Reimers, C.J. Punt, F. Markowetz, L. Vermeulen

Acknowledgments
We would like to thank Rosemary Tate for statistical advice.

Grant Support
L. Vermeulen is supported by KWF grants (UVA2011-4969 and UVA2014-7245), Worldwide Cancer Research (14-1164), the Maag Lever Darm Stichting (MLDS-CDG-14-03), and the European Research Council (ERC-StG 638193). A. Trinh and F. Markowetz acknowledge the support of The University of Cambridge, Cancer Research UK, and Hutchison Whampoa Limited. Parts of this work were funded by CRUK core grant C14303/A17197 and A19274.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 14, 2016; revised June 23, 2016, accepted July 15, 2016, published OnlineFirst July 26, 2016.

References


Clinical Cancer Research

Practical and Robust Identification of Molecular Subtypes in Colorectal Cancer by Immunohistochemistry

Anne Trinh, Kari Trumpi, Felipe De Sousa E Melo, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-16-0680

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2016/09/03/1078-0432.CCR-16-0680.DC1
http://clincancerres.aacrjournals.org/content/suppl/2016/09/09/1078-0432.CCR-16-0680.DC2

Cited articles
This article cites 30 articles, 4 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/23/2/387.full#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/23/2/387.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/23/2/387.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.