



OPEN ACCESS

EDITED BY

Sharon R. Pine,
University of Colorado Anschutz Medical
Campus, United States

REVIEWED BY

Marahaini Musa,
University of Science Malaysia (USM), Malaysia
Marco Tonello,
Veneto Institute of Oncology (IRCCS), Italy
Chuanwen Fan,
Linköping University, Sweden
Sergio Facchini,
University of Campania Luigi Vanvitelli, Italy

*CORRESPONDENCE

Vlad Popovici

✉ vlad.popovici@recetox.muni.cz

RECEIVED 08 January 2024

ACCEPTED 18 March 2024

PUBLISHED 19 April 2024

CITATION

Budinská E, Čarnogurská M, Ivković TC,
Macháčková T, Boudná M, Pífková L, Slabý O,
Bencsiková B and Popovici V (2024) An
invasion front gene expression signature
for higher-risk patient selection in stage
IIA MSS colon cancer.
Front. Oncol. 14:1367231.
doi: 10.3389/fonc.2024.1367231

COPYRIGHT

© 2024 Budinská, Čarnogurská, Ivković,
Macháčková, Boudná, Pífková, Slabý,
Bencsiková and Popovici. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

An invasion front gene expression signature for higher-risk patient selection in stage IIA MSS colon cancer

Eva Budinská¹, Martina Čarnogurská¹, Tina Catela Ivković²,
Táňa Macháčková^{2,3}, Marie Boudná^{2,3}, Lucie Pífková^{2,3},
Ondřej Slabý^{2,3}, Beatrix Bencsiková⁴ and Vlad Popovici^{1*}

¹RECETOX, Faculty of Science, Masaryk University, Brno, Czechia, ²Central European Institute of Technology, Masaryk University, Brno, Czechia, ³Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czechia, ⁴Department of Comprehensive Cancer Care, Masaryk Memorial Cancer Institute, Brno, Czechia

Stage II colon cancer (CC) encompasses a heterogeneous group of patients with diverse survival experiences: 87% to 58% 5-year relative survival rates for stages IIA and IIC, respectively. While stage IIA patients are usually spared the adjuvant chemotherapy, some of them relapse and may benefit from it; thus, their timely identification is crucial. Current gene expression signatures did not specifically target this group nor did they find their place in clinical practice. Since processes at invasion front have also been linked to tumor progression, we hypothesize that aside from bulk tumor features, focusing on the invasion front may provide additional clues for this stratification. A retrospective matched case-control collection of 39 stage IIA microsatellite-stable (MSS) untreated CCs was analyzed to identify prognostic gene expression-based signatures. The endpoint was defined as relapse within 5 years vs. no relapse for at least 6 years. From the same tumors, three different classifiers (bulk tumor, invasion front, and constrained baseline on bulk tumor) were developed and their performance estimated. The baseline classifier, while the weakest, was validated in two independent data sets. The best performing signature was based on invasion front profiles [area under the receiver operating curve (AUC) = 0.931 (0.815–1.0)] and contained genes associated with KRAS pathway activation, apical junction complex, and heme metabolism. Its combination with bulk tumor classifier further improved the accuracy of the predictions.

KEYWORDS

colon cancer, invasion front, early stage, prognostic signature, stage II/MSS

1 Introduction

Despite important progress made in early detection and treatment over the last decades, colon cancer (CC) is still one of the major causes of death among all solid tumor cancers accounting for more than 600,000 deaths yearly (1). The TNM (tumor–node–metastasis) staging remains the cornerstone of patient management and outcome prediction, even though several other predictors have been proposed, including commercially available gene signatures, such as Oncotype Dx Colon (2), ColoPrint (3), and ColDX (4), or immune system scoring, such as Immunoscore Colon (5). Globally, stage II CC, accounting for 35%–40% of newly diagnosed cases (SEER Cancer Stat Facts: Colorectal Cancer; <https://seer.cancer.gov/statfacts/html/colorect.html>), has a good prognosis, with 5-year relative survival rates of 58%–87% (6). However, compared to other stages, it is more heterogeneous with low, intermediate, and high risk for metastatic dissemination subgroups, as recognized in the revised categorization (6). Microsatellite instability (MSI) or deficiency in DNA mismatch repair (dMMR) are characteristics of a low-risk group, with more than 90% 5-year overall survival (7). The high-risk (pT4bN0, stage IIC) or intermediate-risk microsatellite-stable (pT4aN0/MSS, stage IIB/MSS) patients are generally treated with adjuvant chemotherapy after curative surgical resection (8). The benefits from adjuvant therapy are not clear in these patients probably because direct evidence from clinical trials is still insufficient (9). However, the low-risk patients (pT3N0, stage IIA) are usually spared the adjuvant treatment, but still, approximately 13% of them will die within 5 years (6). Therefore, it is of utmost importance to develop better prognostic tools, eventually integrated with the TNM staging, targeting the earlier stage where the benefit from adjuvant treatment may potentially be significant.

All the transcriptomic signatures proposed so far considered whole-tumor sampling for RNA extraction. Still, mounting evidence suggests that processes taking place at the invasion front would be equally prognostic, if not even more. The activation of epithelial-to-mesenchymal transition (EMT) at aberrant expression of nuclear β -catenin as invasion front markers of tumor progression has been recognized previously (10, 11). Also, the infiltrative configuration of the invasion front and the presence of tumor budding have been recognized as additional prognostic parameters (12, 13). It has been proposed that the balance of pro- and anti-tumor factors at the invasion front may be decisive for tumor progression (14) and overexpression of *ZEB2* (an epithelial-to-mesenchymal transition-associated gene) as the invasion front has been identified as an independent prognostic factor in a general CC patient population (15). Additionally, the immune reaction scored along the invasion front could be used to stratify the CC patients into three distinct risk groups (5). In addition, the histopathologic characteristics of the reactive stroma at the invasion front have been shown to bear prognostic potential (16). Thus, it is of interest whether transcriptomics of the invasion front may bring novel discriminative markers that could improve patient stratification.

The goal of the present pilot study is twofold: to assess the prognostic utility of invasion front gene expression and develop a

predictor of early relapse within the low-risk stage IIA/MSS colon cancers. From the same group of patients, we develop gene signatures from both bulk tumor (traditional tumor sampling) and tumor invasion front predicting the risk of relapse, and we compare their performance. As the study has a limited sample size, we opted for increasing the contrasts between the groups by selecting patients with relapse within 5 years vs. patients with no relapse for at least 6 years.

2 Materials and methods

2.1 Samples

This retrospective matched case-control study used tumor samples from patients with CC who underwent surgery at Masaryk Memorial Cancer Institute, Brno, Czech Republic, in the years 1998–2018. Inclusion criteria for this study were as follows: age >18 years, clinically and histopathologically confirmed diagnosis of primary CC, stage IIA (pathology T-stage 3, N0), microsatellite-stable primary tumors, and no adjuvant chemotherapy. Standard clinical and histopathological variables (TNM, grade, etc.) were retrieved for all patients. The “early relapse” group was defined as those patients experiencing a relapse within 5 years from the date of diagnosis, while the “no relapse” group consisted of patients who did not experience a relapse for at least 6 years. The relapse was defined as any disease recurrence or disease-related death except for any second primary cancers. To the extent possible, the two groups were further matched in terms of gender, age, and grade distribution. Failure of laboratory analyses (problematic sample preparation, low quality and/or quantity of isolated RNA, and low quality of expression data) was a reason for excluding these samples from the study.

From each tumor block, two different regions were sampled in adjacent sections: one representing the bulk tumor and one only the invasion front (Supplementary Figure 1). Each sample was profiled independently.

2.2 Expression profiling

The RNA extraction was performed from formalin-fixed paraffin-embedded histopathological slides using AllPrep DNA/RNA Kits (Qiagen, Hilden, Germany) according to manufacturer's instructions. The extracted RNA served as input for a GeneChip WT Pico Reagent Kit (Thermo Fisher Scientific, Waltham, MA, USA) for analysis of the transcriptome on whole-transcriptome arrays. Total RNA from HeLa cells provided in the kit was used as a positive control together with high-quality low-concentration RNA isolated from a serum as a low-input control. Clariom D Array for human samples (Thermo Fisher Scientific, Waltham, MA, USA) was used for target hybridization to capture both coding and multiple forms of non-coding RNA. Finally, the arrays were scanned using Affymetrix GeneChip Scanner 3000 7 G (Thermo Fisher Scientific, Waltham, MA, USA). All the samples complied with the quality control

requirements, and none of the samples were excluded from the analysis.

2.3 Bioinformatics analyses

All resulting CEL files were processed using Bioconductor (17) (v.3.15) packages oligo (18) (v.1.60), affycoretools (v1.68), and, for Clariom D chip annotation, pd.clariom.d.human (v.3.14). For the quality control, we used AffyPLM (v.147) and imposed a maximal median Normalized Unscaled Standard Error (NUSE) of 1.12. All chips passing the quality control steps were normalized together using RMA (oligo) with core-probeset summarization. Further, the array data were summarized at gene level by selecting the most variable probeset per unique EntrezID, and entries corresponding to missing HUGO symbols, speculative transcripts, microRNA, and short non-coding RNA were discarded resulting in a reduced list of 28,663 unique genes.

For the identification of differentially expressed genes, we used linear models (limma package v.3.52.2) with a cut-off for false discovery rate (FDR) of 0.1. The pathways were scored in terms of enrichment in specific signatures using gene set enrichment analysis (GSEA) (19) as implemented in fgsea package (v.1.22.0). MSigDB (hallmark gene sets collection “H” v.7.4.1) (20) was used as the main source for gene sets and pathways. The gene expression classifiers were based on ElasticNet model as implemented in the R package caret (v.6.0). All data analyses were performed in R 4.3 (R Development Core Team, 2022).

The development of the predictive models required the following two major steps: feature generation and classifier training. These two steps were embedded in an external leave-one-out loop for estimating the performance. The main performance parameter of the model was the area under the receiver operating curve (AUC) with sensitivity and specificity also estimated and reported. For the feature generation step, we first selected the most predictive (in terms of AUC) and stable genes and grouped them into modules according to gene signatures from MSigDB (H-section). For estimating the stability of each gene, we generated $b = 50$ bootstraps of the current training set (at each iteration of the leave-one-out procedure) and recorded the AUC and direction of the association of the gene with the outcome. We defined the direction of a gene g as $d_g = +1$ if the average expression of the gene in the “early relapse” group was higher than in the “no relapse” group; otherwise, $d_g = -1$. The AUC for a gene was the average AUC from bootstrapping procedure, and the gene was considered stable if the direction of the association with the outcome was constant (over the b bootstraps). The gene modules were generated from MSigDB gene signatures by selecting the top five (in terms of AUC) subsets of n_g genes from each signature. The value of a module was defined as $n_g^{-1} \sum d_g x_g^g$, where x_g is the expression of gene g in the module. By extension, the names of the gene modules were taken from the names of the corresponding signatures even though they no longer represented their de-/activation status. Then, an ElasticNet model was fitted on the top n_f gene modules. To minimize the chances of overfitting, the tested

domain for n_g and n_f was limited to values 3, 4, and 5. No constraint was imposed on the number of times a gene could be selected in different modules (the signatures from MSigDB overlapped) nor on selecting only one module per gene signature. While this choice introduces potential redundancy in the model, it also improves its robustness.

To validate the modeling approach, we used two independent data sets (21) compatible with our experimental design (with the exception of unknown MSI status) publicly available from ArrayExpress under accession numbers E-MTAB-863 and E-MTAB-864, respectively. We further limited the set of genes to the intersection of the two platforms (Clariom D for our study and Affymetrix customized Almac array for the independent sets) resulting in 13,274 common symbols. Also, in the validation sets (denoted KEN1 and KEN2), we considered only the patients in our target group (pT3/pN0/pM0); the rest of the expression profiles were used for mitigating the differences between the two microarray platforms. The model built (and validated) on the restricted set of genes was considered as a baseline model. Additionally, as the two external data sets contained survival data as well, we estimated the probability of survival in the two predicted groups using the Kaplan–Meier estimator and tested for significant difference between the curves using the log-rank test.

The main analysis considered the full set of genes available on our platform (Clariom D) and concerned the two sampling regions as follows: bulk tumor and invasion front, respectively.

3 Results

In total, $n = 39$ patients were identified fitting the selection criteria [19 cases of early relapse (12 men) vs. 20 cases of no relapse (11 men)] resulting in 39 bulk tumor profiles. For the same patients, $n = 35$ [17 early relapse (11 men) and 18 no relapse (10 men)] good quality invasion front profiles were also generated. No statistically significant differences were found between groups regarding age, tumor location, or grade (Table 1).

3.1 Differentially expressed genes and pathways

The differential expression analyses of both bulk tumor and invasion front samples revealed no genes with significantly different expressions between early and no relapse groups after adjusting for multiple testing. Nevertheless, 204 and 333 genes had a significant (un-adjusted) p-value (≤ 0.01) within the bulk tumor and invasion front samples, respectively. Using the t -statistics estimated by limma as input for ordering the whole set of genes for GSEA, we identified a number of pathways/gene sets differentially activated between the early relapse and no relapse groups (Figure 1). The full list of significant (un-adjusted p-value) genes (p-value ≤ 0.01) is given in Supplementary Table 1 and the GSEA results in Supplementary Table 2.

TABLE 1 Basic patient population demographics for the training set.

	Early relapse (within 5 years)	No relapse (for at least 6 years)	p-Value	Test
N	19	20		
Age [mean (SD)]	69.5 (9.22)	68.9 (9.56)	0.849	Student's t-test
Gender				
Female	7	9	0.747	Fisher's exact test
Men	12	11		
Grade				
G2	18	20	0.487	Fisher's exact test
G3	1	0		
Tumor site				
Right (including transverse colon)	14	12	0.501	Fisher's exact test
Left	5	8		

All patients were stage II/A, microsatellite stable.

3.2 Early relapse predictors

To validate the approach, we developed a baseline predictor of early relapse cases using a restricted set of genes common to the two platforms (Clariom D and Almac) and based on bulk tumor profiles. The optimal model used $n_f = 5$ gene modules each with $n_g = 4$ genes

(see Table 2). Its estimated leave-one-out performance was $AUC_0 = 0.795(95\%CI = 0.625 - 0.964)$ (Figure 2A). The binary classification performance (for the default cut-off of 0.5) was sensitivity $Se = 0.737(95\%CI = 0.488 - 0.908)$ and specificity $Sp = 0.8(95\%CI = 0.563 - 0.943)$. At the same time, the observed performance on the validation sets was $AUC_{KEN1} = 0.731(95\%CI = 0.636 - 0.827)$ and $AUC_{KEN2} = 0.768(95\%CI = 0.612 - 0.874)$ being superior to the one reported elsewhere (21) (Supplementary Figure 2). The Kaplan–Meier curves for predicted groups (“no relapse” and “early relapse”) were significantly different ($p < 0.001$) (Supplementary Figure 3).

For the genes in the modules, a positive sign (explicit or implicit) indicates its higher expression in the “early relapse” group, while the negative sign indicates the reverse situation.

With the modeling approach validated, we studied the predictive power of the profiles derived from bulk tumor and invasion front regions. First, we compared the univariate (per-gene) AUCs for bulk and invasion front profiles (Figure 2B, Supplementary Table 3) estimated using all samples. It was apparent that the invasion front expression profiles were more predictive with the top ranking genes having consistently higher univariate AUC (2%–5%). Also, there were almost twice as many genes from the invasion front with $AUC > 0.7$ than from bulk tumor profiles (Supplementary Table 3).

The predictors built from the bulk and invasion front profiles confirmed this tendency (Figure 2A): the leave-one-out estimated performance for invasion front was $AUC_i = 0.931(95\%CI = 0.815 - 1.0)(Se = 0.882, Sp = 0.833)$, superior to the bulk tumor performance: $AUC_b = 0.887(95\%CI = 0.750 - 1.0)(Se = 0.895, Sp = 0.75)$. The two models are given in Table 2 and further gene annotations in Supplementary Table 4.

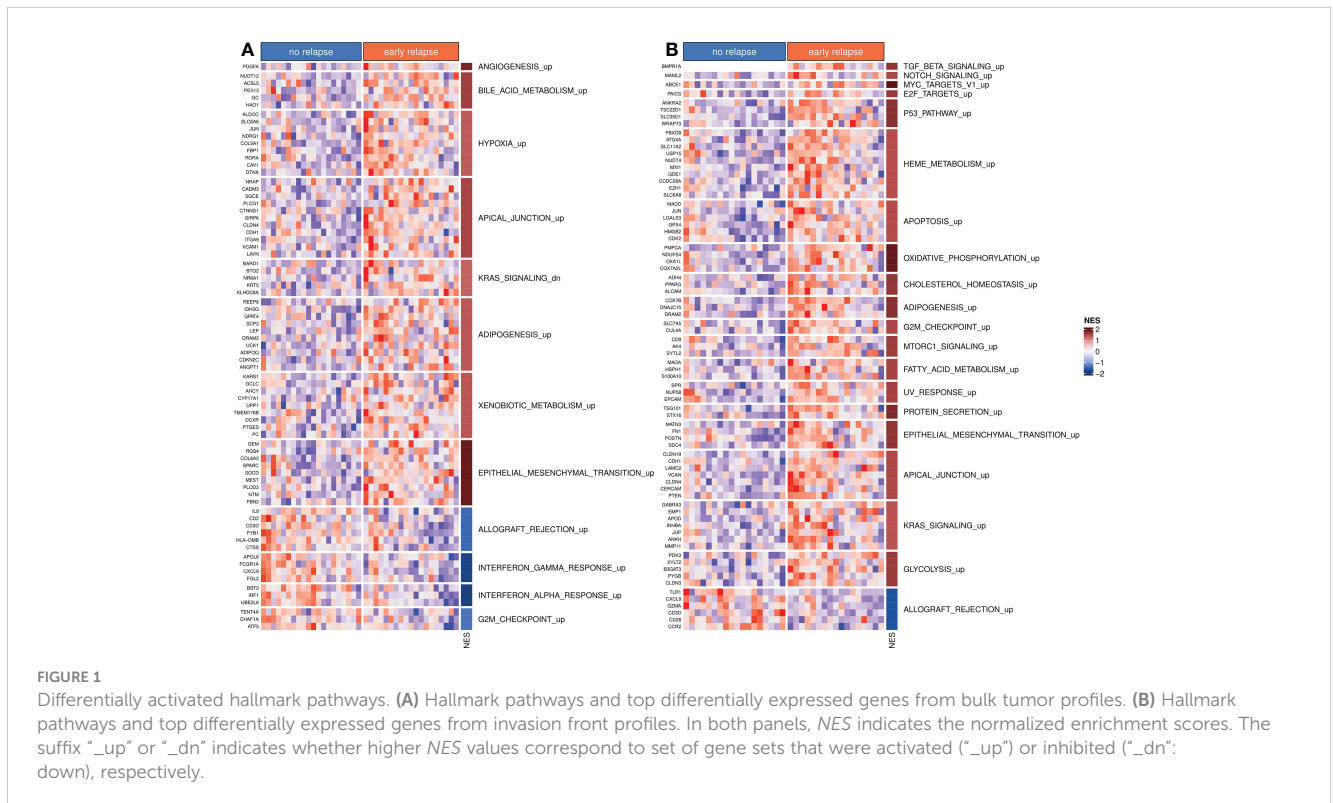


TABLE 2 Predictive models and their performance.

Model	Modules and coefficients	Module coefficient	Genes in modules	Leave-one-out performance estimates (with 95% confidence intervals)
Baseline model	INTERFERON_GAMMA_RESPONSE_up1 INTERFERON_GAMMA_RESPONSE_up2 INTERFERON_GAMMA_RESPONSE_up3 TNFA_SIGNALING_VIA_NFKB_up1 TNFA_SIGNALING_VIA_NFKB_up2	1.0545 -0.7575 2.0305 1.8225 -0.9185	LATS2 - IRF1 - TRIM14 - APOL6 LATS2 - CXCL9 - TRIM14 - APOL6 LATS2 - IRF1 - TRIM14 - CXCL9 DUSP1 + LAMB3 - IRF1 - SLC2A6 DUSP1 + JUN - IRF1 - SLC2A6	AUC = 0.795 (0.625–0.964) Se = 0.737 (0.488–0.908) Sp = 0.8 (0.563–0.943)
Bulk tumor model	INFLAMMATORY_RESPONSE_up IL6_JAK_STAT3_SIGNALING_up APICAL_JUNCTION_up	1.1161 -0.1483 0.6747	EBI3 + KCNMB3 + TLR2 - IRF1 - TACR3 EBI3 + HAX1 + TLR2 - IRF1 - CXCL9 CLDN4 + LAYN + ITGA9 + NRAP + CADM3	AUC = 0.887 (0.75–1.0) Se = 0.895 (0.669–0.987) Sp = 0.75 (0.509–0.913)
Invasion front model	APICAL_JUNCTION_up KRAS_SIGNALING_up HEME_METABOLISM_up	0.1652 0.1527 0.0915	VCAN + CLDN19 + PTEN + CDH1 GABRA3 + APOD + JUP - TMEM100 EZH1 + CCDC28A + FBXO9 + SLC6A8	AUC = 0.931 (0.815–1.0) Se = 0.882 (0.636–0.985) Sp = 0.833 (0.586–0.964)

3.3 Combining predictors

We also compared the scores (posterior probabilities from ElasticNet models) produced by the two models (Figure 2C). The correlations (Pearson correlation: 0.564, Spearman correlation: 0.582) between the scores were modest, as was Cohen's kappa coefficient ($\kappa = 0.484$ between the class assignments based on these scores). This indicated a certain degree of complementarity between the two models, and we speculatively created an average score (from leave-one-out scores of matched tumor bulk and invasion front samples) and used it for predicting the groups. The new score indeed improved on all previous predictions— $AUC = 0.977$ ($95\%CI = 0.907 - 1.0$), $Se = 0.941$, $Sp = 0.889$).

4 Discussion

The intermediate-risk group of patients with stage II colon cancer is heterogeneous in terms of survival experience: while most of the patients fare well without any adjuvant chemotherapy, others relapse much sooner. Reliably identifying the patients at risk for early relapse is, therefore, fundamental.

Our pilot study addressed two problems: First, developing a gene-based predictor for the stage IIA colon cancer patients who, despite being considered as low risk of relapse by current guidelines, are relapsing within 5 years. The second problem addressed aimed at investigating whether the invasion front is more predictive for the early relapse. Benefitting from a matched data set on which both bulk tumor and invasion front were profiled, we developed two predictive models. In our data, the invasion front model proved to be significantly superior to the bulk tumor model. This suggests that the dynamic changes happening on the contact border between the tumor and the normal tissue of the host may bear more information about the invasiveness potential of the tumor.

The targeted patient population appears to be rather homogeneous from the perspective of transcriptomics, with no gene significantly differentially expressed between “no relapse” and “early relapse” groups, after adjustment for multiple hypotheses testing. Nevertheless, several genes reached statistical significance

when considered individually with more genes in the case of invasion front samples. Using the results from the differential expression analysis as input for gene set enrichment analyses, several significantly deregulated pathways/gene sets were identified. Some of them were common between bulk tumor and invasion front samples, most notably the epithelial-to-mesenchymal transition pathway, which was strongly up regulated in early relapse cases. Interestingly, the KRAS activation appeared in contrasting instances between the following two types of samples: in bulk tumors, the KRAS-down gene set was activated in the “early relapse” group, while in invasion front samples, the KRAS-up gene set was activated in the same group of patients, indicating a differential activation of KRAS between bulk tumor and invasion front regions within early CC.

The first predictor for early relapse established a baseline model and performance and validated the modeling approach. However, it was limited in the number of genes covered, as the two independent validation sets originated from an older microarray platform. Nevertheless, we were able to construct and validate a relatively strong classifier from bulk tumor profiles. The validation sets (21) were not selected for MSS, as this was not reported, but the baseline model performed close to the estimated performance. While the baseline classifier relied on five gene modules, the features selected by the algorithm referred to only two of the following MSigDB's pathways: interferon-gamma ($INF-\gamma$) and tumor necrosis factor-alpha ($TNF-\alpha$) via nuclear factor- $\kappa\beta$, related to antitumor immunity and inflammatory processes, respectively. More interestingly, one gene—*IRF1* (interferon regulatory factor 1)—was common to both pathways (and to both bulk tumor models) and selected in four out of five modules being downregulated in the early relapse group. Upregulation of this gene was shown to be related to better survival and tumor radiosensitivity (22). We also note that the model could be further simplified to a model with only two modules ($INF-\gamma$ and $TNF-\alpha$) each of five genes; however, this combination was not foreseen when training the models (we imposed $n_f = 3,4,5$).

The same modeling approach was applied on tumor bulk and invasion front profiles considering all the genes present on our platform (still limited to the hallmark pathways of MSigDB). This led to the development of two models of which the invasion front

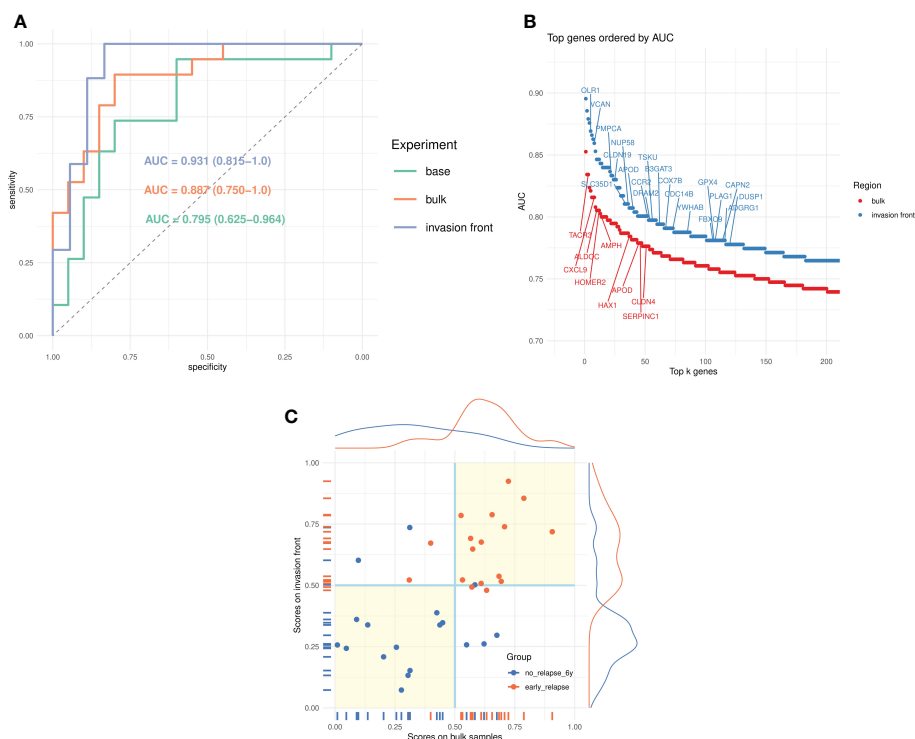


FIGURE 2

Prediction of early relapse. (A) Receiver operating characteristics (ROC) curves for the three models (baseline, bulk tumor, and invasion front) and the corresponding AUCs. (B) Univariate AUC, based on all samples, for top $k = 200$ genes from bulk tumor and invasion front expression profiles. The top genes (AUC > 0.775) from MSigDB hallmark signatures are marked. (C) Scatter plots of scores from bulk tumor and invasion front (35 samples) and their marginal distributions. The points are colored according to their true category, and the quadrants marked (light yellow background) indicate regions of agreement for the two classifiers.

signature had the best performance while both being superior to the baseline model. As the models were derived from tumor samples originating from the same patients, comparing the two allowed us to gain more insights into the predictive power of the invasion front. We first investigated the predictive utility (in terms of AUC) of each gene and found more genes from the invasion front having higher AUCs than from bulk tumors (see also [Supplementary Table 3](#)). While these results hinted toward more prognostic value of the invasion front signatures, it was the multivariable models (ElasticNets) that showed this being true in practice. Both models comprised of three gene modules with apical junction being a common term. However, the genes selected in the two “apical junction” modules were different with those from the invasion front pointing also toward EMT (*VCAN*) and estrogen receptor (*CDH1*). Also, we note the *KRAS*-related module present in the invasion front signature, which, corroborated with the results of GSEA ([Figure 1](#); [Supplementary Table 2](#)), points toward a stronger *KRAS* pathway activation in early relapse patients. While specific mutations of the *KRAS* oncogene were shown to be predictive for overall survival in some studies (23, 24), they appeared not to be predictive for relapse-free survival (25). A more detailed annotation of all genes, with further references, is given in [Supplementary Table 4](#). We also noted that the proposed marker gene for invasion front (15), *ZEB2*, was prognostic in our data as well, but with lower performance [$AUC_{ZEB2} = 0.716$ (0.521–0.910); [Supplementary Table 3](#)].

Our pilot study has some limitations as well: the invasion front signature could not be validated on external independent data because no similar data collections exist. We make our data publicly available to begin filling this gap. Second, the sample size did not allow for more analyses. For example, the observation that combining invasion front and bulk tumor signatures into a stronger predictor was made *post hoc*, and it would require another data set for its statistical assessment.

Another aspect pertains to the definition/delineation of the invasion front. We expect a relatively significant inter-observer variability. Thus, for the future results to be validated independently, a consensus must be reached between pathologists to stabilize the sampling regions.

In conclusion, our study proposes a novel invasion front-derived gene signature for predicting high-risk patients within the stage IIA colon cancer group. Its combination with bulk tumor signature further improved the prediction suggesting that a combined, dual sampling of core and border of the tumor may lead to a practical and precise predictor.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found below: <https://www.ebi.ac.uk/arrayexpress/>, E-MTAB-13695.

European Commission is not responsible for any use that may be made of the information it contains.

Ethics statement

The studies involving humans were approved by Research Ethics committee of Masaryk University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

EB: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft. MČ: Methodology, Writing – original draft. TI: Investigation, Methodology, Writing – original draft. TM: Data curation, Writing – original draft. MB: Data curation, Writing – original draft. LP: Data curation, Writing – original draft. OS: Conceptualization, Methodology, Writing – original draft. BB: Data curation, Funding acquisition, Investigation, Project administration, Writing – original draft. VP: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Supported by the Ministry of Health of the Czech Republic, grant nr. 19-03-00298. All rights reserved. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825410 (ONCOBIOME project). This publication reflects only the author's view, and the

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Gray RG, Quirke P, Handley K, Lopatin M, Magill L, Baehner FL, et al. Validation study of a quantitative multigene reverse transcriptase–polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. *JCO.* (2011) 29:4611–9. doi: 10.1200/JCO.2010.32.8732
3. Kopetz S, Tabernero J, Rosenberg R, Jiang Z-Q, Moreno V, Bachleitner-Hofmann T, et al. Genomic classifier coloPrint predicts recurrence in stage II colorectal cancer patients more accurately than clinical factors. *Oncol.* (2015) 20:127–33. doi: 10.1634/theoncologist.2014-0325
4. Niedzwiecki D, Frankel WL, Venook AP, Ye X, Friedman PN, Goldberg RM, et al. Association between results of a gene expression signature assay and recurrence-free interval in patients with stage II colon cancer in cancer and leukemia group B 9581 (Alliance). *JCO.* (2016) 34:3047–53. doi: 10.1200/JCO.2015.65.4699
5. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou F-S, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a

Acknowledgments

The authors thank the RECETOX Research Infrastructure (No LM2023069) financed by the Ministry of Education, Youth and Sports for supportive background. Supported by Ministry of Health of the Czech Republic, grant nr. 19-03-00298. All rights reserved. This work was supported from the European Union's Horizon 2020 research and innovation program under grant agreement No 857560.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1367231/full#supplementary-material>

- prognostic and accuracy study. *Lancet.* (2018) 391:2128–39. doi: 10.1016/S0140-6736(18)30789-X
6. Gunderson LL, Jessup JM, Sargent DJ, Greene FL, Stewart AK. Revised TN categorization for colon cancer based on national survival outcomes data. *JCO.* (2010) 28:264–71. doi: 10.1200/JCO.2009.24.0952
7. Dienstmann R, Mason MJ, Sinicrope FA, Phipps AI, Tejpar S, Nesbakken A, et al. Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. *Ann Oncol.* (2017) 28:1023–31. doi: 10.1093/annonc/mdx052
8. Taieb J, Karoui M, Basile D. How I treat stage II colon cancer patients. *ESMO Open.* (2021) 6(4):100184. doi: 10.1016/j.esmoop.2021.100184
9. Argilés G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* (2020) 31:1291–305. doi: 10.1016/j.annonc.2020.06.022
10. Brabletz T, Hlubek F, Spaderna S, Schmalhofer O, Hiendlmeyer E, Jung A, et al. Invasion and metastasis in colorectal cancer: epithelial-mesenchymal transition, mesenchymal-epithelial transition, stem cells and β -catenin. *Cells Tissues Organs.* (2005) 179:56–65. doi: 10.1159/000084509

11. Bronsert P, Enderle-Ammour K, Bader M, Timme S, Kuehs M, Csanadi A, et al. Cancer cell invasion and EMT marker expression: a three-dimensional study of the human cancer-host interface: 3D cancer-host interface. *J Pathol.* (2014) 234:410–22. doi: 10.1002/path.4416
12. Compton CC, Fielding LP, Burgart LJ, Conley B, Cooper HS, Hamilton SR, et al. Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med.* (2000) 124:979–94. doi: 10.5858/2000-124-0979-PFICC
13. Graham RP, Vierkant RA, Tillmans LS, Wang AH, Laird PW, Weisenberger DJ, et al. Tumor budding in colorectal carcinoma: confirmation of prognostic significance and histologic cutoff in a population-based cohort. *Am J Surg Pathol.* (2015) 39:1340–6. doi: 10.1097/PAS.0000000000000504
14. Zlobec I, Lugli A. Invasive front of colorectal cancer: dynamic interface of pro-/anti-tumor factors. *World J Gastroenterol.* (2009) 15:5898–906. doi: 10.3748/wjg.15.5898
15. Kahlert C, Lahes S, Radhakrishnan P, Dutta S, Mogler C, Herpel E, et al. Overexpression of ZEB2 at the invasion front of colorectal cancer is an independent prognostic marker and regulates tumor invasion *in vitro*. *Clin Cancer Res.* (2011) 17:7654–63. doi: 10.1158/1078-0432.CCR-10-2816
16. Martin B, Grosser B, Kempkens L, Miller S, Bauer S, Dhillon C, et al. Stroma AReactive invasion front areas (SARIFA)-A new easily to determine biomarker in colon cancer-results of a retrospective study. *Cancers.* (2021) 13(19):4880. doi: 10.4324/9781003101857
17. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods.* (2015) 12:115–21. doi: 10.1038/nmeth.3252
18. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics.* (2010) 26:2363–7. doi: 10.1093/bioinformatics/btq431
19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
20. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst.* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004
21. Kennedy RD, Bylesjo M, Kerr P, Davison T, Black JM, Kay EW, et al. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *JCO.* (2011) 29:4620–6. doi: 10.1200/jco.2011.35.4498
22. Xu X, Wu Y, Yi K, Hu Y, Ding W, Xing C. IRF1 regulates the progression of colorectal cancer via interferon-induced proteins. *Int J Mol Med.* (2021) 47:104. doi: 10.3892/ijmm.2021.4937
23. Imamura Y, Morikawa T, Liao X, Lochhead P, Kuchiba A, Yamauchi M, et al. Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers. *Clin Cancer Res.* (2012) 18:4753–63. doi: 10.1158/1078-0432.CCR-11-3210
24. Jones RP, Sutton PA, Evans JP, Clifford R, McAvoy A, Lewis J, et al. Specific mutations in KRAS codon 12 are associated with worse overall survival in patients with advanced and recurrent colorectal cancer. *Br J Cancer.* (2017) 116:923–9. doi: 10.1038/bjc.2017.37
25. Popovici V, Budinska E, Bosman FT, Tejpar S, Roth AD, Delorenzi M. Context-dependent interpretation of the prognostic value of BRAF and KRAS mutations in colorectal cancer. *BMC Cancer.* (2013) 13:439. doi: 10.1186/1471-2407-13-439