

## Article

# EMT-Related Genes Have No Prognostic Relevance in Metastatic Colorectal Cancer as Opposed to Stage II/III: Analysis of the Randomised, Phase III Trial FIRE-3 (AIO KRK 0306; FIRE-3)

Elise Pretzsch <sup>1,2</sup>, Volker Heinemann <sup>2,3,4</sup>, Sebastian Stintzing <sup>5</sup> , Andreas Bender <sup>6,7</sup> , Shuo Chen <sup>7</sup>, Julian Walter Holch <sup>3</sup>, Felix Oliver Hofmann <sup>1,2</sup> , Haoyu Ren <sup>1</sup>, Florian Bösch <sup>8</sup>, Helmut Küchenhoff <sup>6</sup>, Jens Werner <sup>1</sup> and Martin Konrad Angele <sup>1,\*</sup>

- <sup>1</sup> Department of General, Visceral, and Transplant Surgery, Ludwig-Maximilians-University Munich, 80539 Munich, Germany
  - <sup>2</sup> German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, partner site 80336 Munich, Germany
  - <sup>3</sup> Department of Hematology/Oncology and Comprehensive Cancer Center Munich, LMU University Hospital Munich, Ludwig-Maximilians-University Munich, 81377 Munich, Germany
  - <sup>4</sup> Comprehensive Cancer Center (CCC Munich LMU), LMU University Hospital Munich, 81377 Munich, Germany
  - <sup>5</sup> Department of Hematology, Oncology, and Cancer Immunology (CCM), Charité-Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, 10117 Berlin, Germany
  - <sup>6</sup> Statistical Consulting Unit, StaBLab, Department of Statistics, LMU Munich, 80539 Munich, Germany
  - <sup>7</sup> Munich Center for Machine Learning, LMU Munich, 81377 Munich, Germany
  - <sup>8</sup> Department of General, Visceral, and Pediatric Surgery, University Medical Center Goettingen, 37075 Göttingen, Germany
- \* Correspondence: martin.angele@med.uni-muenchen.de; Tel.: +49-89-44000; Fax: +49-89-440075474



**Citation:** Pretzsch, E.; Heinemann, V.; Stintzing, S.; Bender, A.; Chen, S.; Holch, J.W.; Hofmann, F.O.; Ren, H.; Bösch, F.; Küchenhoff, H.; et al. EMT-Related Genes Have No Prognostic Relevance in Metastatic Colorectal Cancer as Opposed to Stage II/III: Analysis of the Randomised, Phase III Trial FIRE-3 (AIO KRK 0306; FIRE-3). *Cancers* **2022**, *14*, 5596. <https://doi.org/10.3390/cancers14225596>

Academic Editor: Roberto Benelli

Received: 13 October 2022

Accepted: 9 November 2022

Published: 14 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Despite huge advances in local and systemic therapies, the 5-year relative survival rate for patients with metastatic CRC is still low. To avoid over- or undertreatment, proper risk stratification with regard to treatment strategy is highly needed. As EMT (epithelial-mesenchymal transition) is a major step in metastatic spread, this study analysed the prognostic effect of EMT-related genes in stage IV colorectal cancer patients using the study cohort of the FIRE-3 trial, an open-label multi-centre randomised controlled phase III trial of stage IV colorectal cancer patients. Overall, the prognostic relevance of EMT-related genes seems stage-dependent. EMT-related genes have no prognostic relevance in stage IV CRC as opposed to stage II/III.

**Abstract:** Introduction: There is no standard treatment after resection of colorectal liver metastases and the role of systemic therapy remains controversial. To avoid over- or undertreatment, proper risk stratification with regard to postoperative treatment strategy is highly needed. We recently demonstrated the prognostic relevance of EMT-related (epithelial-mesenchymal transition) genes in stage II/III CRC. As EMT is a major step in CRC progression, we now aimed to analyse the prognostic relevance of EMT-related genes in stage IV CRC using the study cohort of the FIRE-3 trial, an open-label multi-centre randomised controlled phase III trial of patients with metastatic CRC. Methods: Overall and progression free survival were considered as endpoints ( $n = 350$ ). To investigate the prognostic relevance of EMT-related genes on either endpoint, we compared predictive performance of different models using clinical data only to models using gene data in addition to clinical data, expecting better predictive performance if EMT-related genes have prognostic value. In addition to baseline models (Kaplan Meier (KM), (regularised) Cox), Random Survival Forest (RSF), and gradient boosted trees (GBT) were fit to the data. Repeated, nested five-fold cross-validation was used for hyperparameter optimisation and performance evaluation. Predictive performance was measured by the integrated Brier score (IBS). Results: The baseline KM model showed the best performance (OS: 0.250, PFS: 0.251). None of the other models were able to outperform the KM when using clinical data only according to the IBS scores (OS: 0.253 (Cox), 0.256 (RSF), 0.284 (GBT); PFS: 0.254 (Cox), 0.256 (RSF), 0.276 (GBT)). When adding gene

data, performance of GBT improved slightly (OS: 0.262 vs. 0.284; PFS: 0.268 vs. 0.276), however, none of the models performed better than the KM baseline. Conclusion: Overall, the results suggest that the prognostic relevance of EMT-related genes may be stage-dependent and that EMT-related genes have no prognostic relevance in stage IV CRC.

**Keywords:** colorectal cancer; metastasis; EMT; EMT-related genes

## 1. Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide, being the second leading cause of cancer-related deaths [1] with more than 1.9 million new cases each year. About 20% of patients present with synchronous liver metastases and up to 50% of patients develop distant metastases during their disease with the liver being the most frequent site of metachronous spread [1–5]. Despite huge advances in local and systemic therapies, the 5-year relative survival rate for patients with metastatic CRC (mCRC) still ranges between 14–17% [6]. To date, the personalised approach to treat mCRC as recommended by national (German S3-Leitlinie, NCCN) and international (ESMO, ESMO-Asia) guidelines is limited to the analysis of microsatellite (MSI) status and mutational analysis of RAS (rat sarcoma oncogene) and B-RAF [7–10]. However, extended molecular testing has the potential to identify druggable targets beyond standardised treatment options, establish biomarkers that allow better and more precisely risk stratification, predict prognoses, and improve clinical decision-making in a precision medicine approach. There is still no standard treatment after the resection of colorectal liver metastases (CRLM) and the role of systemic therapy remains controversial. Whereas adjuvant treatment is recommended after surgery in stage III CRC with nodal spread, there is no standard recommendation for systemic treatment after surgery in stage IV CRC with distant spread, a rationale that might not seem conclusive [9]. To avoid over- or undertreatment, a proper risk stratification regarding the postoperative treatment strategy is highly needed (precision oncology). Successful stratification of risk groups based on tumour biology reflected in longer disease-free survival in high-risk groups receiving additive systemic treatment and avoidance of unnecessary adverse effects of systemic treatment in low-risk groups could lead to a paradigm shift in the treatment strategy of stage IV CRC.

In this respect, we recently demonstrated the prognostic relevance of EMT-related (epithelial-mesenchymal transition) genes in stage II/III CRC. Further, we proposed an EMT-related gene signature that identified patients at risk of relapse in multiple CRC cohorts. This EMT-related gene signature was a strong predictive indicator for recurrence in stage II/III CRC patients and associated with overall survival [11]. With the aim to optimise patient outcome, the respective EMT signature might help to stratify patients according to their tumour biology, and contribute to personalised treatment in the future.

EMT is a key program that enables stationary epithelial cells to lose their cell-cell adherence and acquire mesenchymal properties, including enhanced mobility, invasiveness, increased resistance to apoptosis, and degradation and production of extracellular matrix components, that are all essential for invasion and metastasis. In this respect, EMT is associated with an aggressive phenotype, pivotal for tumour progression and the prerequisite for metastatic spread [12,13]. EMT is regulated at different molecular levels that lead to the loss of E-Cadherin as the critical event with a subsequent activation of all major cancer cell intrinsic signaling pathways. Whereas EMT and EMT-related gene signatures have been demonstrated to be associated with prognosis and therapeutic resistance in non-metastatic stages of CRC and various other tumour entities [11,14–17], the prognostic role of EMT-related genes and our previously proposed EMT-related gene signature in mCRC remains uncertain.

As EMT is a major step in CRC progression and metastatic spread, it was the aim of this study to analyse the prognostic relevance of EMT-related genes in stage IV CRC patients

using the study cohort of the FIRE-3 trial. The FIRE-3 trial was an open-label multi-centre randomised controlled phase III trial for first-line treatment of patients with RAS wild-type (wt) mCRC patients [18]. In this respect, we aimed to assess whether the prognostic value of the previously identified EMT-related genes and our proposed EMT-related gene signature in stage II/III CRC can be validated in the metastatic setting of CRC and potentially be used for risk stratification and guidance of systemic treatment in an individualised therapy approach in stage IV CRC.

## 2. Materials and Methods

### 2.1. Study Design

FIRE-3 was designed as an open-label, multi-centre, randomised phase III trial that evaluated the combination of FOLFIRI plus cetuximab or bevacizumab as first-line regimen in irresectable RASwt mCRC. Treatment protocol, regulatory aspects of trial conduct, safety and efficacy, outcome, molecular subgroups, and next-generation sequencing results were published in the studies by Heinemann et al., Stintzing et al., Stahler et al., and Modest et al. [14,18–22].

### 2.2. Patients

Clinical data was available from 752 patients. In 416 cases, genetic information was also available. After removing missing values and a single subject with primary tumour location on both sides (rather than left or right), 350 patients with complete data remained for the analysis, of which 237 were male and 113 were female. For each patient, the corresponding treatment, tumour location, metastatic status (solitary liver metastasis (nonresectable metastases confined to the liver at time of diagnosis) versus distant metastasis in more than one organ), BRAF<sup>V600E</sup> status (wild-type (wt) versus mutated (mut)), survival parameters (overall survival (OS), progression-free survival (PFS)) along with 191 variables containing gene expression data were collected (Table 1).

**Table 1.** Patient characteristics.

	Sub Category	Frequency	Percent (%)
Total valid records		350	100
Gender	Female	113	32.28
	Male	237	67.71
Type of treatment	Cetuximab	165	47.14
	Bevacizumab	185	52.86
Tumour location	Left	273	78.00
	Right	77	22.00
Solitary liver metastasis	Yes	116	33.14
	No	234	66.86
BRAF <sup>V600E</sup>	Wt	265	75.71
	Mut	20	5.71
	Not tested	65	18.57
Overall survival (OS)	Censored	37	10.57
	Dead	313	89.43
Progression-free survival (PFS)	Censored	21	6.00
	Progression or dead	329	94.00

### 2.3. Gene-Expression Analysis

Using formalin-fixed paraffin-embedded (FFPE) samples of primary tumour tissue, gene expression analysis was carried out using ALMAC's Xcel<sup>TM</sup> gene-expression array at ALMACs laboratories [23]. All analyses were approved by the ethics committee of the Ludwig-Maximilians-University, Munich (#186-15).

#### 2.4. EMT-Related Dataset

The EMT-related dataset investigated in this study was derived from public databases as previously described [11]. In this respect, transcriptome profiles and clinical information of 1780 stage II/III CRC patients from 15 public datasets were investigated. Coefficient variant analysis was used to select reference genes for normalising gene expression levels. Univariate, LASSO, and multivariate Cox regression analyses were combined to develop the originally studied EMT related dataset [11].

#### 2.5. Outcomes

Endpoints investigated in this study included progression-free survival (PFS) (time from randomisation to disease progression or death from any cause) and overall survival (OS) (time from randomisation to death from any cause).

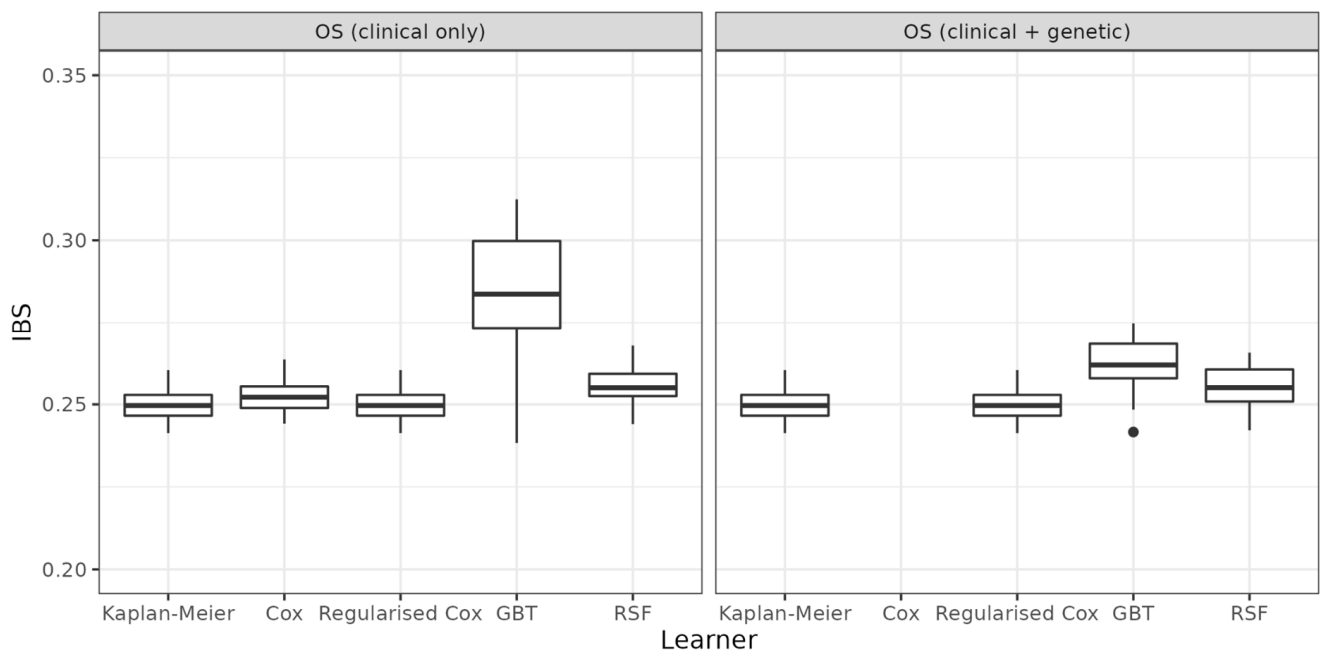
#### 2.6. Statistical Analysis

For each endpoint (OS and PFS), we trained different models and investigated their predictive performance. In order to investigate the prognostic relevance of gene data, we compared the predictive performance of the models using clinical data only to model using gene data in addition to clinical data, expecting better predictive performance if gene data has prognostic value. The models used for comparison were Kaplan-Meier (KM) [24], (regularised) Cox Regression (Cox) [25], Random Survival Forest (RSF) [26,27], and Gradient Boosted Trees with Cox Loss (GBT) [28]. KM served as a baseline for the prediction without taking any covariate information into account, the Cox models served as a baseline for a model with covariates but without non-linear effects and interactions. The predictive performance of the models was measured by the integrated Brier Score (IBS) evaluated at the median survival time [29].

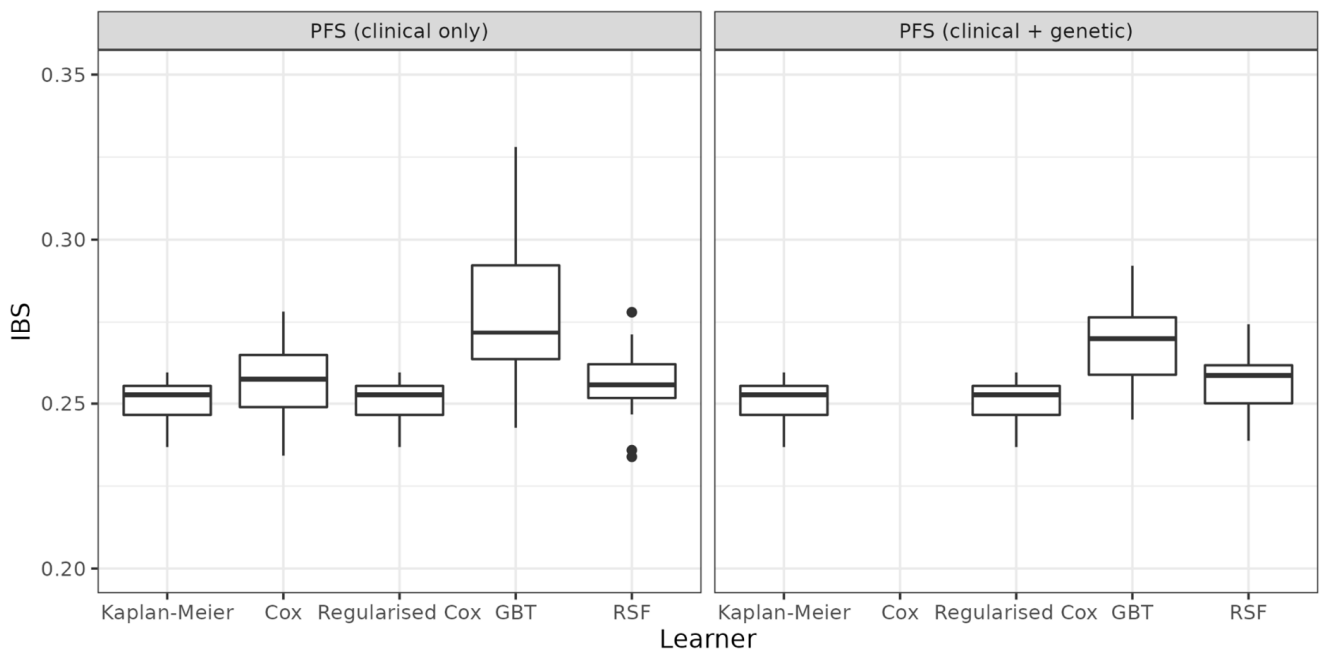
For regularised Cox, RSF, and GBT, hyperparameter optimisation (HPO) was performed via random search [30] and three-fold cross validation (CV) on the respective training data (see Appendix A for details). Performance evaluation was based on a repeated five-fold CV with five repetitions. All analyses including training the models were performed in the R programming environment (version 4.1.3). For the setup of survival tasks, model training, HPO, and performance evaluation, we used *mlr3proba* [31] with the *mlr3* [32] ecosystem. All code used for the analysis is available from GitHub: <https://github.com/adibender/EMT-gene-fire3-prognostic-relevance> (accessed on 1 January 2022).

### 3. Results

The results of the benchmark experiments are given in Figure 1 (OS), Figure 2 (PFS), and Table 2. The boxplots indicate the distribution of the 25 IBS values calculated on the test data from the respective iteration of the repeated cross-validation. Lower values of the IBS indicate better predictive performance. The results indicate that predictive performance cannot be improved when using gene data in addition to clinical data (comparison of left and right panels, respectively). While mean values for GBT decrease slightly when comparing performance with and without genetic data (OS: 0.262 vs. 0.284; PFS: 0.268 vs. 0.276), the performance is still worse than the mean performance of RSF or unregularised Cox with clinical data only. Furthermore, none of the models using additional information (clinical and/or genetic) can outperform the KM baseline (OS: 0.250, PFS: 0.251). Performance of regularised Cox regression is identical to KM, as all coefficients are penalised to zero.



**Figure 1.** Benchmark experiments with respect to overall survival. Predictive performance of five learners based on 25 values of the IBS. Lower values indicate better performance. Performance based on clinical data only (**left**) is compared to performance based on clinical and gene data (**right**). Performance of the unregularised Cox model is omitted for the high-dimensional setting (data with genes).



**Figure 2.** Benchmark experiments with respect to progression-free survival. Predictive performance of five learners based on 25 values of the IBS. Lower values indicate better performance. Performance based on clinical data only (**left**) is compared to performance based on clinical and gene data (**right**). Performance of the unregularised Cox model is omitted for the high-dimensional setting (data with genes).

**Table 2.** Benchmark experiments with respect to overall survival and progression-free survival. Aggregated predictive performance of five learners based on 25 values of the IBS. Lower values indicate better performance. Performance of the unregularised Cox model is omitted for the high-dimensional setting (data with genes). Performance of KM is identical between the different settings (clinical only vs. clinical + genetic).

Task	Learner	Mean (sd)	Median
OS (clinical only)	Kaplan-Meier	0.25 (0.005)	0.2496
	Cox	0.253 (0.006)	0.2522
	Regularised Cox	0.25 (0.005)	0.2496
	GBT	0.284 (0.019)	0.2837
	RSF	0.256 (0.007)	0.2550
OS (clinical + genetic)	Kaplan-Meier	0.250 (0.005)	0.2496
	Cox	—	—
	Regularised Cox	0.25 (0.005)	0.2496
	GBT	0.262 (0.009)	0.2619
	RSF	0.255 (0.007)	0.2551
PFS (clinical only)	Kaplan-Meier	0.251 (0.007)	0.2527
	Cox	0.257 (0.013)	0.2574
	Regularised Cox	0.251 (0.007)	0.2527
	GBT	0.276 (0.021)	0.2717
	RSF	0.256 (0.011)	0.2557
PFS (clinical + genetic)	Kaplan-Meier	0.251 (0.007)	0.2527
	Cox	—	—
	Regularised Cox	0.251 (0.007)	0.2527
	GBT	0.268 (0.013)	0.2697
	RSF	0.258 (0.009)	0.2585

#### 4. Discussion

Advances in genomic and transcriptomic analyses have shifted cancer therapy towards a precision medicine approach and allowed better understanding of the molecular alterations of CRC with regard to tumour initiation, progression, and resistance [33]. While the TNM staging system in combination with molecular markers (RAS, BRAF, MSI) is the backbone of therapeutic decisions and used as a guideline for survival estimates, there is a wide variation in prognosis among CRC patients with the same TNM stage and a survival paradox of stage II/III CRC patients on account of the inherent heterogeneity that traditional clinicopathological and molecular features fail to explain. In this respect, identification of innovative markers and risk factors based on tumour biology that can guide the administration of systemic treatment (targeted therapies, postoperative additive treatments) in CRC need to be introduced into the clinical arena.

In this respect, we recently demonstrated that EMT-related genes have prognostic relevance in stage II/III CRC. Further, we developed an innovative prognostic model based on our proposed EMT-related gene signature predicting recurrence of stage II/III CRC patients, offering a potential explanation with regard to tumour biology beyond traditional clinicopathological characteristics for the mechanisms underlying the observed survival paradox [11]. We now aimed to assess the prognostic relevance of EMT-related genes in a metastasized CRC setting using the study cohort of the FIRE-3 trial, a multi-centre randomised controlled phase III trial of mCRC patients [18].

Analyses using (regularised) Cox and RSF showed no improvement in predictive performance according to IBS when using gene data in addition to clinical data (see Figures 1 and 2; Table 2), and therefore no prognostic effect of EMT-related genes in stage IV CRC. GBT performed slightly better when using gene data in addition to clinical data, but still performed worse than the RSF or unregularised Cox using clinical data only. Furthermore, according to our results, none of the models using covariate information (clinical and/or gene data) could outperform the KM with respect to predictive performance.

We have previously shown that FOLFIRI plus cetuximab was associated with improved OS in patients with RASwt mCRC relative to those treated with FOLFIRI plus bevacizumab [18]. This FOLFIRI plus cetuximab conferred OS benefit, however, was in the absence of differences in investigator-assessed objective response or PFS. In an attempt to elucidate the underlying relationship for these unexplained results, metrics of tumour dynamics were assessed, and centralised radiological review revealed that FOLFIRI plus cetuximab induced superior objective response, frequency of early tumour shrinkage and depth of response compared with FOLFIRI plus bevacizumab. In this respect, early tumour shrinkage and depth of response were associated with OS in both treatment groups [19]. These results highlight the importance of new innovative metrics that reflect tumour biology to predict therapeutic response and outcome. Indeed, this is underscored by the increasing evidence that evaluation of response according to RECIST criteria may not adequately capture the quality and quantity of response to targeted therapies in mCRC [24]. As depth of response is an on-treatment parameter occurring approximately 3.5 months after the beginning of treatment, this parameter might rather be used for retrospective analyses than initial clinical decision making. On the other hand, early tumour shrinkage is a useful parameter to guide decision making in the early phase of treatment, however, the parameter by itself does not consider the subgroup of patients that show no early shrinkage, though they are slow responders who could still benefit from continuation of treatment. In this respect, there is still a need for further metrics to supplement the RECIST criteria that are easy to obtain, feasible in a clinical setting, and can predict therapeutic efficacy not only with improved precision, but also at an early stage. To this aim, we assessed the relevance of our proposed EMT-related signature that fulfilled the above-described requirements and was able to predict recurrence and survival in stage II/III CRC patients, in a metastatic setting.

To our knowledge, we are first to assess the prognostic relevance of EMT-related genes in stage IV CRC. The results of this study indicate that in an advanced stage, EMT-related characteristics may give insights into the underlying tumour biology and mechanisms that have preceded metastatic spread but do not add further value to the determination of prognosis.

In fact, the role of EMT in metastatic tumours is still under debate. EMT is a critical process for tumour progression in which epithelial cells lose their epithelial features and acquire mesenchymal characteristics, such as invasion and motility. During EMT, cancer cells are considered to gain a more aggressive phenotype and are more prone to develop metastatic spread. Studies suggest that induction of EMT is critical for the initial steps of metastatic spread but not for metastatic seeding and outgrowth at the distant site. Other studies point out that the mere presence of tumour cells displaying EMT characteristics in the primary tumour does not prove that EMT is even absolutely required for metastatic spread. In line with this, not all cells that have undergone EMT will successfully metastasize. Further studies suggest that EMT may be a temporal function of metastatic progression that may play a pivotal role and predict prognosis in earlier stages of CRC, as also described by our study groups, whereas successful spread may be dependent on various factors including transcription factors, miRNAs, and noncoding RNAs, epigenetic regulators, environmental factors, and multiple other signaling molecules [34].

We have previously assessed the relevance of consensus-molecular subgroups (CMS), grouping CRC according to their gene-signature in four different subtypes, and found that OS in CMS4 (defined by EMT and an activated tissue growth factor (TGF)- $\beta$  path-

way making this subgroup more chemo-resistant) favored FOLFIRI plus cetuximab over FOLFIRI plus bevacizumab. However, similar to this study and from a clinical standpoint, CMS classification appeared not to be of superior value with regard to patient selection and optimal treatment [14]. In summary, evaluation of EMT-related genes has prognostic relevance in non-metastatic CRC, however, after the tumour has spread, they do not appear to add further value.

There are also limitations of this study. The material that the gene-expression analysis in this investigation was based on, was predominantly derived from primary tumours, as liver tissue was not available in the majority of cases, due to irresectability. In this respect, EMT-related characteristics in the primary tumour may not have prognostic relevance after the primary tumour has spread, however, although it has been shown that CMS classification changes from primary tumour to metastases, it remains unclear whether the expression of EMT related genes changes within metastases, which would be an interesting research question for future studies [35]. We used a range of different methods to be able to capture different types of effects that genes could exhibit on OS or PFS (linear or non-linear effects, interactions, deviation from proportional hazards). The results indicate no prognostic effect of gene data, as defined by predictive performance measured by the IBS, however, given the usually small effect sizes associated with individual genes, the sample size for this study might have been too low in order to detect them, if present. Performance of GBT could probably be improved through additional tuning, however, given the results obtained from other learners, it is doubtful that performance of the KM could be improved upon.

Overall, EMT-related genes did not show prognostic relevance in stage IV CRC and are not of additional value compared to common parameters used regarding patient selection and clinical decision making. In this respect, further studies that investigate novel metrics that reflect tumour biology to predict therapeutic response and outcome in mCRC and that can supplement the currently limited decision-making tools are warranted.

## 5. Conclusions

We have assessed the prognostic relevance of EMT-related genes in stage IV CRC. The results of this study indicate that in an advanced stage, EMT-related characteristics may give insights into the underlying tumour biology and mechanisms that have preceded metastatic spread but do not add further value to the determination of prognosis compared to common parameters used regarding patient selection and clinical decision making.

**Author Contributions:** Conceptualization, E.P., V.H., S.S. and M.K.A.; methodology, E.P., J.W.H. and M.K.A.; formal analysis, A.B., S.C. and H.K.; investigation, E.P., A.B., S.C. and H.K.; resources, E.P., V.H., S.S., H.R., F.B. and J.W.H.; data curation, V.H., S.S., H.R. and F.B.; writing—original draft preparation, E.P.; writing—review and editing, V.H., S.S., F.O.H., F.B., J.W. and M.K.A.; visualization, E.P. and A.B.; supervision, M.K.A.; funding acquisition, V.H. and S.S.; All authors have read and agreed to the published version of the manuscript.

**Funding:** The legal funder (sponsor) of the trial is University Hospital, LMU Munich. This work was supported by grants from Pfizer GmbH, Germany, and Merck KGaA, Darmstadt, Germany. The funding sources had a role in the design and conduct of the study, and in the collection, management, analysis and interpretation of the data. Prior to 2009, cetuximab was supplied by Merck Serono GmbH, an affiliate of Merck KGaA, Darmstadt, Germany. Merck KGaA reviewed the paper for medical accuracy only and had no role in the decision to submit the paper for publication. The authors are fully responsible for the content of this paper, and the views and opinions described in the publication solely reflect those of the authors. Open Access funding enabled and organised by Projekt DEAL.

**Institutional Review Board Statement:** The protocol and informed consent forms were approved by the ethic committee of the Medical Faculty of the Ludwig-Maximilians-University (reference number: 370-06). The study was conducted in accordance with the Declaration of Helsinki.



**Informed Consent Statement:** Informed consent was obtained from all subjects prior to participating in the study.

**Data Availability Statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Conflicts of Interest:** VH: Honoraria: Merck, Amgen, Roche, Sanofi, SIRTEX, Servier, Pfizer, Pierre-Fabre, Astra-Zeneca. Consulting: Merck, Amgen, Roche, Sanofi, SIRTEX, BMS; MSD, Novartis, Boehringer Ingelheim, Servier, Pierre-Fabre, Celgene, Terumo. Research funding (for the institution): Merck, Amgen, Roche, Sanofi, Pfizer, BoehringerIngelheim, Sirtex, Bayer, Servier. Travel accommodation expenses: Merck, Roche, Amgen, SIRTEX, Bayer, Servier. SS: Honoraria: AMGEN, Bayer, BMS, ESAI, Lilly, Merck KGaA, MSD, Pierre-Fabre, Roche, Sanofi, Servier, Taiho, Takeda, Financial relationships: Merck KGaA, Pierre-Fabre, Servier, Roche. Advisory role: AMGEN, Bayer, BMS, ESAI, Lilly, Merck KGaA, MSD, Pierre-Fabre, Roche, Sanofi, Servier, Taiho, Takeda. JH: Advisory role: Roche Honoraria: Roche, Travel support: Novartis.

## Appendix A

Model	Hyperparameter	Searching Range
Random Survival Forest	Minimum Node Size	(1, 50)
	Number of variables to possibly split at in each node	(0, 1)
	Fraction of observations to sample	(0.1, 1)
	Sample with replacement	(TRUE, FALSE)
	Splitting rule	(maxstat, logrank)
Cox (regularised)	Elastic net mixing parameter alpha; penalty parameter is trained internally via 10-fold CV	(0, 1)
Gradient Boosting Tree	Maximum depth of a tree	(1, 20)
	Subsample ratio of the training instances	(0.1, 1.0)
	Number of rounds	(10, 5000)
	Step size shrinkage	(0.01, 1.0)
	Subsample ratio of columns when constructing each tree	(0.01, 1.0)
	Minimum loss reduction required to make a further partition on a leaf node of the tree	(0.01, 3)
	The way new nodes are added to the tree	(Depth-wise, Loss-guide)

In order to assess predictive performance of different algorithms and data modalities (clinical and genetic), we set up a benchmark experiment with a nested cross-validation, 5 times repeated 5-fold CV in the outer loop (for performance evaluation) and a 3-fold CV in the inner loop for hyperparameter optimisation (HPO). Random searching mechanism is deployed during the optimisation, where the number of search iterations depends on the dimension of the search space. Concretely, the number of iterations is determined by the size of the search-space based on a linear growth strategy  $B = b + k * D$  in which  $B$  is the number of evaluations,  $D$  is the dimension of the search space,  $b$  is 25 and  $k$  is 60 in our experiments.

## References

1. Brenner, H.; Kloor, M.; Pox, C.P. Colorectal cancer. *Lancet* **2014**, *383*, 1490–1502. [[CrossRef](#)]
2. Manfredi, S.; Lepage, C.; Hatem, C.; Coatmeur, O.; Faivre, J.; Bouvier, A.M. Epidemiology and management of liver metastases from colorectal cancer. *Ann. Surg.* **2006**, *244*, 254–259. [[CrossRef](#)] [[PubMed](#)]
3. Manfredi, S.; Bouvier, A.M.; Lepage, C.; Hatem, C.; Dancourt, V.; Faivre, J. Incidence and patterns of recurrence after resection for cure of colonic cancer in a well defined population. *Br. J. Surg.* **2006**, *93*, 1115–1122. [[CrossRef](#)] [[PubMed](#)]
4. Elferink, M.A.G.; de Jong, K.P.; Klaase, J.M.; Siemerink, E.J.; de Wilt, J.H.W. Metachronous metastases from colorectal cancer: A population-based study in North-East Netherlands. *Int. J. Colorectal Dis.* **2015**, *30*, 205–212. [[CrossRef](#)]
5. Brandi, G.; de Lorenzo, S.; Nannini, M.; Curti, S.; Ottone, M.; Dall’Olio, F.G.; Barbera, M.; Pantaleo, M.A.; Biasco, G. Adjuvant chemotherapy for resected colorectal cancer metastases: Literature review and meta-analysis. *World J. Gastroenterol.* **2016**, *22*, 519–533. [[CrossRef](#)]
6. *Colorectal Cancer Early Detection, Diagnosis, and Staging*; American Cancer Society: Atlanta, GA, USA, 2022.
7. Van Cutsem, E.; Cervantes, A.; Adam, R.; Sobrero, A.; van Krieken, J.H.; Aderka, D.; Aguilar, E.A.; Bardelli, A.; Benson, A.; Bodoky, G.; et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* **2016**, *27*, 1386–1422. [[CrossRef](#)]
8. Benson, B.; Venook, A.P.; Cederquist, L.; Chan, E.M.; Chen, Y.; Cooper, H.S.; Fichera, A.; Grem, J.L.; Grothey, A.; Hochster, H.S.; et al. Journal of the national comprehensive cancer network, colon cancer, version 1.2021, NCCN clinical practice guidelines in oncology. *NCCN Clin. Pract. Guidel. Oncol.* **2021**, *19*, 329–359.
9. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe and L. Onkologie). *AWMF Regist. 021/007OL; S3-Leitlinie Kolorektales Karzinom*; Langversion 2.1. 2019; pp. 1–328.
10. Yoshino, T.; Arnold, D.; Taniguchi, H.; Pentheroudakis, G.; Yamazaki, K.; Xu, R.H.; Kim, T.W.; Ismail, F.; Tan, I.B.; Yeh, K.H.; et al. Pan-Asian adapted ESMO consensus guidelines for the management of patients with metastatic colorectal cancer: A JSMO-ESMO initiative endorsed by CSCO, KACO, MOS, SSO and TOS. *Ann. Oncol.* **2018**, *29*, 44–70. [[CrossRef](#)]
11. Ren, H.; Bösch, F.; Pretzsch, E.; Jacob, S.; Westphalen, C.B.; Holch, J.W.; Werner, J.; Angele, M.K. Identification of an EMT-related gene signature predicting recurrence in stage II/III colorectal cancer—A retrospective study in 1780 patients. *Ann. Surg.* **2022**, *276*, 1–25. [[CrossRef](#)]
12. Cao, M.L.H.; Xu, E.; Liu, H.; Wan, L. Epithelial-mesenchymal transition in colorectal cancer metastasis: A system review. *Pathol. Resarch Pract.* **2015**, *211*, 557–569. [[CrossRef](#)]
13. Pretzsch, E.; Bösch, F.; Neumann, J.; Ganschow, P.; Bazhin, A.; Guba, M.; Werner, J.; Angele, M.K. Mechanisms of metastasis in colorectal cancer and metastatic organotropism: Hematogenous versus peritoneal spread. *J. Oncol.* **2019**, *2019*, 7407190. [[CrossRef](#)] [[PubMed](#)]
14. Stintzing, S.; Wirapati, P.; Lenz, H.-J.; Neureiter, D.; von Weikersthal, L.F.; Decker, T.; Kiani, A.; Kaiser, F.; Al-Batran, S.; Heintges, T.; et al. Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial. *Ann. Oncol.* **2019**, *30*, 1796–1803. [[CrossRef](#)]
15. Findlay, V.J.; Wang, C.; Watson, D.K.; Camp, E.R. Epithelial-to-mesenchymal transition and the cancer stem cell phenotype: Insights from cancer biology with therapeutic implications for colorectal cancer. *Cancer Gene Ther.* **2014**, *21*, 181–187. [[CrossRef](#)] [[PubMed](#)]
16. Chen, S.; Zhao, E. Development and validation of a robust epithelial-mesenchymal transition (EMT)-related prognostic signature for hepatocellular carcinoma. *Clin. Res. Hepatol. Gastroenterol.* **2021**, *45*, 101587. [[CrossRef](#)] [[PubMed](#)]
17. Dai, W.; Xiao, Y.; Li, J.; Hong, L.; Zhang, J.; Pei, M.; Lin, J.; Liu, S.; Wu, X.; Xiang, L.; et al. Identification of an EMT-related gene signature for predicting overall survival in gastric cancer. *Front. Genet.* **2021**, *12*, 1–18. [[CrossRef](#)] [[PubMed](#)]
18. Heinemann, V.; von Weikersthal, L.F.; Decker, T.; Kiani, A.; Vehling-Kaiser, U.; Al-Batran, S.-E.; Heintges, T.; Lerchenmüller, C.; Kahl, C.; Seipelt, G.; et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): A randomised, open-label, phase 3 trial. *Lancet Oncol.* **2014**, *15*, 1065–1075. [[CrossRef](#)]
19. Stintzing, S.; Modest, D.; Rossius, L.; Lerch, M.; von Weikersthal, L.F.; Decker, T.; Kiani, A.; Vehling-Kaiser, U.; Al-Batran, S.; Heintges, T.; et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab for metastatic colorectal cancer (FIRE-3): A post-hoc analysis of tumour dynamics in the final RAS wild-type subgroup of this randomised open-label phase 3 trial. *Lancet Oncol.* **2016**, *17*, 1426–1434. [[CrossRef](#)]
20. Stintzing, S.; Miller-Phillips, L.; Modest, D.; von Weikersthal, L.F.; Decker, T.; Kiani, A.; Vehling-Kaiser, U.; Al-Batran, S.; Heintges, T. Impact of BRAF and RAS mutations on first-line efficacy of FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab: Analysis of the FIRE-3 (AIO KRK-0306) study. *Eur. J. Cancer* **2017**, *79*, 50–60. [[CrossRef](#)]
21. Stahler, A.; Stintzing, S.; von Einem, J.C.; Westphalen, C.B.; Heinrich, K.; Krämer, N.; Michl, M.; Modest, D.P.; von Weikersthal, L.F.; Decker, T.; et al. Single-nucleotide variants, tumour mutational burden and microsatellite instability in patients with metastatic colorectal cancer: Next-generation sequencing results of the FIRE-3 trial. *Eur. J. Cancer* **2020**, *137*, 250–259. [[CrossRef](#)]
22. Modest, D.P.; Heinemann, V.; Folprecht, G.; Denecke, T.; Pratschke, J.; Lang, H.; Bemelmans, M.; Becker, T.; Rentsch, M.; Seehofer, D.; et al. Factors that influence conversion to resectability and survival after resection of metastases in RAS WT metastatic colorectal cancer (mCRC): Analysis of FIRE-3- AIOKRK0306. *Ann. Surg. Oncol.* **2020**, *27*, 2389–2401. [[CrossRef](#)]

23. Stahler, A.; Heinemann, V.; Holch, J.W.; Einem, J.C.; Westphalen, C.B.; Heinrich, K.; Schlieker, L.; Jelas, I.; Alig, A.H.S.; Fischer, L.E.; et al. Mutational profiles of metastatic colorectal cancer treated with FOLFIRI plus cetuximab or bevacizumab before and after secondary resection (AIO KRK 0306; FIRE-3). *Int. J. Cancer* **2021**, *149*, 1935–1943. [[CrossRef](#)] [[PubMed](#)]
24. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete samples. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [[CrossRef](#)]
25. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)] [[PubMed](#)]
26. MWright, N.; Ziegler, A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [[CrossRef](#)]
27. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
28. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
29. Graf, E.; Schmoor, C.; Sauerbrei, W.; Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. *Stat. Med.* **1999**, *18*, 2529–2545. [[CrossRef](#)]
30. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
31. Sonabend, R.; Király, F.J.; Bender, A.; Bischl, B.; Lang, M. mlr3proba: An R package for machine learning in survival analysis. *Bioinformatics* **2021**, *37*, 2789–2791. [[CrossRef](#)]
32. Lang, M.; Binder, M.L.M.; Richter, J.; Schratz, P.; Pfisterer, F.; Coores, S.; Au, Q.; Casalicchio, G.; Kotthoff, L.; Bischl, B.; et al. mlr3: A modern object-oriented machine learning framework in R. *J. Open Source Softw.* **2019**, *4*, 1903. [[CrossRef](#)]
33. Dienstmann, R.; Vermeulen, L.; Guinney, J.; Kopetz, S.; Tejpar, S.; Tabernero, J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **2017**, *17*, 79–92. [[CrossRef](#)]
34. Mittal, V. Epithelial mesenchymal transition in tumor metastasis. *Annu. Rev. Pathol. Mech. Dis.* **2018**, *13*, 395–412. [[CrossRef](#)] [[PubMed](#)]
35. Schlicker, A.; Ellappalayam, A.; Beumer, I.J.; Snel, M.H.J.; Mittempergher, L.; Diosdado, B.; Dreezen, C.; Tian, S.; Salazar, R.; Loupakis, F.; et al. Investigating the concordance in molecular subtypes of primary colorectal tumors and their matched synchronous liver metastasis. *Int. J. Cancer* **2020**, *147*, 2303–2315. [[CrossRef](#)] [[PubMed](#)]