RESEARCH ARTICLE

Identification of a gene expression signature associated with brain metastasis in colorectal cancer

Marlies Michl[1,2] · Francesco Taverna[3] · Christine Woischke[3] · Pan Li[3] · Frederick Klauschen[2,3,4] · Thomas Kirchner[3,4] · Volker Heinemann[1,2,4] · Michael von Bergwelt-Baildon[1,2,4] · Arndt Stahler[5] · Tobias Marcus Herold[1,4] · Vindi Jurinovic[6] · Jutta Engel[7] · Jörg Kumbrink[3,4] · Jens Neumann[3,4]

Abstract

Purpose Brain metastasis (BM) in colorectal cancer (CRC) is a rare event with poor prognosis. Apart from (K)RAS status and lung and bone metastasis no biomarkers exist to identify patients at risk. This study aimed to identify a gene expression signature associated with colorectal BM.

Methods Three patient groups were formed: 1. CRC with brain metastasis (BRA), 2. exclusive liver metastasis (HEP) and, 3. non-metastatic disease (M0). RNA was extracted from primary tumors and mRNA expression was measured using a NanoString Panel (770 genes). Expression was confirmed by qPCR in a validation cohort. Statistical analyses including multivariate logistic regression followed by receiver operating characteristic (ROC) analysis were performed.

Results EMILIN3, MTA1, SV2B, TMPRSS6, ACVR1C, NFAT5 and SMC3 were differentially expressed in BRA and HEP/M0 groups. In the validation cohort, differential NFAT5, ACVR1C and SMC3 expressions were confirmed. BRA patients showed highest NFAT5 levels compared to HEP/M0 groups (global $p = 0.02$). High ACVR1C expression was observed more frequently in the BRA group (42.9%) than in HEP (0%) and M0 (7.1%) groups (global $p = 0.01$). High SMC3 expressions were only detectable in the BRA group (global $p = 0.003$). Only patients with BM showed a combined high expression of NFAT5, ACVR1C or SMC3 as well as of all three genes. ROC analysis revealed a good prediction of brain metastasis by the three genes (area under the curve (AUC) $= 0.78$).

Conclusions The NFAT5, ACVR1C and SMC3 gene expression signature is associated with colorectal BM. Future studies should further investigate the importance of this biomarker signature.

Keywords Colorectal cancer · Brain metastasis · Gene expression signature · Metastatic organotropism

Jörg Kumbrink and Jens Neumann have contributed equally to this work.

Christine Woischke and Jutta Engel: Deceased.

✉ Jörg Kumbrink
Joerg.Kumbrink@med.uni-muenchen.de

1 Department of Medicine III, University Hospital, Ludwig-Maximilian-University of Munich, Munich, Germany

2 Department of Haematology and Oncology, Comprehensive Cancer Center Munich, Ludwig-Maximilian-University of Munich, Munich, Germany

3 Institute of Pathology, Faculty of Medicine, Ludwig-Maximilian-University of Munich, Munich, Germany

4 German Cancer Consortium (DKTK), Partner Site Munich and German Cancer Research Centre (DKFZ), Heidelberg, Germany

5 Department of Hematology, Oncology, and Tumorimmunology, Corporate Member of Freie Universitaet Berlin and Humbolt-Universitaet zu Berlin, Charité – Universitaetsmedizin Berlin, Berlin, Germany

6 Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilian-University of Munich, Munich, Germany

7 Munich Cancer Registry (MCR), Ludwig-Maximilian-University of Munich, Munich, Germany

Introduction

Although significant progress has been made in cancer medicine understanding of organ specific metastasis development remains limited. However, knowledge on molecular mechanisms of organotropic metastasis is essential for biomarker-based prediction and prognosis, invention of innovative therapeutic strategies, and consequently improvement of patient outcomes [1]. The term `metastatic organotropism´ describes the distribution of distant metastases to certain organs in a non-random process which is regulated by multiple factors such as subtypes of cancer, molecular features of cancer cells, host immune microenvironment, as well as cross-talk and interactions with local cells [1].

To date, formation of brain metastasis (BM) from colorectal carcinoma (CRC) is virtually not understood. Compared to other solid tumors, colorectal BM is less common and a rare event. Incidence rates are reported between one and four percent, however, increasing numbers are observed in the last decades [2–4]. Thus far, only two independent factors exist to predict the development of BM in CRC, namely first, the presence of lung or bone metastasis and second the presence of a (K)RAS mutation [5–7]. Patients with KRAS mutated CRC carry a 3.7-fold higher risk to develop BM during their course of disease [8]. Primary tumor site was recently shown to have no predictive impact [4]. When BM is diagnosed prognosis is utterly devastating with survival times of only few months [9, 10]. Thus, intense efforts for a better understanding of pathogenetic mechanisms regarding the formation of colorectal BM as well as the characterization and identification of patients at risk is urgently needed [11].

Biological findings accumulate that corroborate the hypothesis of organotropism implicating that different molecularly codified organotropic CRC-types determine the metastatic pattern early in disease [12–15]. Thus, intensified investigations regarding pathogenetic mechanisms of colorectal BM development, molecular pathological profiles and debates on how best to identify CRC patients at risk are justified.

Earlier we showed that—in contrast to liver metastasis—other mechanisms than deregulation of Wnt/β-catenin-signaling and acquisition of cancer stemness are required for formation of BM [14]. In contrast, the hypothesis of stem cell driven brain metastatic genesis in CRC was strengthened by the detection of stem cell properties in human brain metastasis stem cell lines.

The present work aimed to identify a gene expression signature associated with BM in CRC. For this purpose, a case control study population was designed consisting of three CRC patient cohorts with different organotropic metastatic phenotypes. Gene expression profiling on mRNA level was performed in primary tumor tissue specimen and compared between groups. Validation of identified markers was carried out applying reverse transcriptase (RT) qPCR expression analysis.

Materials and methods

Patients

All patients involved in the present analysis had a histologically proven diagnosis of CRC with either brain metastasis (BM) or exclusive liver metastasis (HEP) or non-metastatic disease for at least five years (M0) and were diagnosed at the Institute of Pathology, Faculty of Medicine, Ludwig-Maximilians-University (LMU) Munich. Suitable participants for the BM group were selected from the previously built database containing 228 patients with CRC and BM [10, 16] by taking into account the availability of sufficient tumor tissue specimens for the planned expression analyses. The matching HEP and M0 groups were identified via a systematic database search in collaboration with the Munich Cancer Registry (MCR). The MCR covered an estimated population of 4.9 million inhabitants in southern Germany. Proceedings on patient selection for the present study are illustrated in Fig. 1 Proceedings on patient selection for matched-pair analysis were described by our group earlier [14, 15].

Study design

The present investigation consisted of a matched-pair analysis where patients from all three groups were matched according to gender, primary tumor site (colon versus rectum), sidedness of primary (right versus left colon), grading and pT-category where applicable. As suitable for a matched-pair analysis, all groups consisted of equal patient numbers. Availability of sufficient tumor tissue specimen limited patient numbers to 15 patients per group.

Histopathological samples

Histopathological diagnosis and classification were reviewed for every available tumor specimen at the accredited Institute of Pathology of the University of Munich (Germany). Histopathological grade was confirmed by an experienced pathologist.

RNA extraction from FFPE samples

Sections from formalin fixed paraffin embedded (FFPE) tissue samples were prepared followed by hematoxylin–eosin staining of one slide. Areas with a minimum percentage
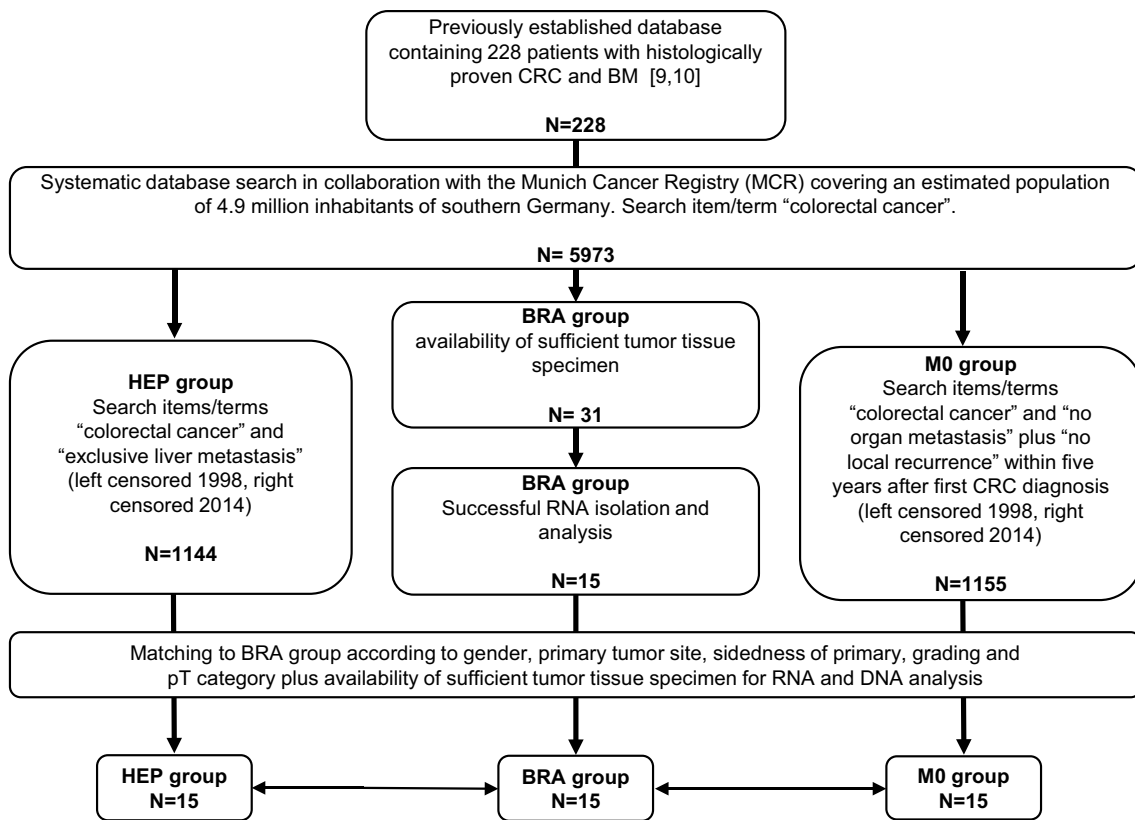
Fig. 1 Consort diagram. Illustration of patient selection for the present analysis

of 50% tumor cells were microdissected from subsequent unstained sections and used for RNA preparation. Total RNA was extracted from six to 12 sections of FFPE tissue sections using the RNeasy FFPE Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. RNA yield and purity were assessed using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Rockland, USA).

NanoString® nCounter expression analysis

mRNA expression was measured with the NanoString nCounter FLEX Analysis System (NanoString Technologies, Seattle, USA) using 100 ng of total RNA and the PanCancer Progression Panel (770 genes). The Nanostring expression analysis was performed as described previously [12, 17]. Briefly, the nCounter CodeSet was hybridized to 100 ng total RNA for 18 h at 65 °C. Quality control and normalization of the expression data was performed using the default nSolver v4.0 software settings by utilizing reference genes, positive/negative controls, total counts, and binding densities in each sample. Unsupervised hierarchical clustering and

principal component analysis (PCA) were performed with the ClustVis web tool [18]. Default settings were used except for heatmap clustering distance for rows and columns, which was set to Euclidean.

Reverse transcriptase (RT) qPCR expression analysis

Total RNA (25 ng/μl within the reverse transcription reaction) was transcribed into cDNA using Random Hexamer Primer and the RevertAid™ Reverse Transcriptase kit (both Thermo Fisher Scientific, Waltham, USA). NFAT5, ACVR1C and SMC3 expression was analyzed by qPCR using primers and UPL (universal probe library) probes (Roche, Basel, Switzerland) displayed in Table S1 and the LightCycler® 480 Probes Master mix (Roche). qPCRs were analyzed on a Bio-Rad® CFX Connect™ Real-Time PCR Detection System with the Bio-Rad® CFX Manager™ Software 3.1 (Bio-Rad Laboratories, Hercules, USA). GAPDH, YWHAZ and ACTB (β-Actin) were used for normalization of gene of interest (GOI) expression. Similar PCR efficiencies (> 95%) were achieved for all investigated genes. Any analysis with no Cq value or a Cq value above 40 was considered "undetectable" and expression was set to 0.

Relative GOI mRNA expressions were calculated by division of measured Cq values by the average Cq values of GAPDH, YWHAZ and ACTB. The threshold value for high gene expression was determined using ROC (receiver operating characteristic) analysis and Youden's index. The determined cut-offs for high expression were: $NFAT5 \geq 1.87 \times 10^{-2}$, $ACVR1C \geq 6.60 \times 10^{-3}$, and $SMC3 \geq 13.73 \times 10^{-2}$.

Statistical analysis

For comparison of patient and tumor characteristics between groups, a two-sided chi-squared test was used. The global testing of the relative mRNA expression was achieved using the Kruskal–Wallis test and for the head-to-head comparisons of the relative mRNA expression the Mann–Whitney-U test was performed. Outliers were selected with Grubbs's test and not considered in the calculations. The significance of correlations of high gene expression and biomarker combinations was calculated using a chi-squared test. Individual risk prediction for brain metastasis was computed by using multivariate logistic regression to obtain coefficients for each gene. Coefficients were multiplied with the continuous expression values for the corresponding gene and subsequently added. To determine how well the risk prediction model discriminates patients with and without brain metastasis ROC analysis was performed. For all statistical tests, SPSS V. 26.0 (IBM Inc., Armonk, NY) or GraphPad PRISM 8 (GraphPad Software, Inc., USA) were employed. A p-value lower than 5% ($p < 0.05$) was considered statistically significant. Global p-values define comparisons between all groups.

Results

Patient and tumor characteristics

The study population consisted of 45 patients (19 male [42%], 26 female [58%]) with histologically proven adenocarcinoma of the colorectum and brain metastasis (BRA; $N = 15$), exclusive liver metastasis (HEP; $N = 15$) or non-metastatic disease (M0; $N = 15$) as defined above. Baseline patient and tumor characteristics of the analyzed patient cohort are presented in Table 1.

Identification of a 7-gene expression signature for brain metastatic CRC

In a PILOT study (BRA; $N = 6$; M0 and HEP; $N = 12$ each), analysis of mRNA expression of 770 genes performed with the Nanostring PanCancer Progression Panel was conducted. Analyses revealed differential gene expressions in the investigated patient groups. Specifically, seven genes namely EMILIN3 (Elastin Microfibril Interfacer 3), MTA1 (Metastasis Associated 1), SV2B (Synaptic Vesicle Glycoprotein 2B), TMPRSS6 (Transmembrane Serine Protease 6), ACVR1C (Activin A Receptor Type 1C), NFAT5 (Nuclear Factor Of Activated T Cells 5) and SMC3 (Structural Maintenance Of Chromosomes 3) were significantly differentially expressed in patients of the BRA group (high expression) and patients from the HEP and M0 group (low expression) (Fig. 2a). Moreover, a perfect separation of the BRA group vs M0 and HEP groups was achieved with the seven gene expressions by unsupervised hierarchical clustering (Fig. 2b) and principal component analysis (PCA, Fig. 2c).

Validation of the identified gene expression profile with qPCR

To confirm the results of the pilot study, expression of the seven identified differentially expressed genes (DEG) was analyzed by qPCR in a larger validation cohort. NFAT5, ACVR1C and SMC3 expression levels were associated with BM, thus confirming the results of the pilot study. Relative mRNA expressions of NFAT5, ACVR1C and SMC3 are listed in Table 2 and illustrated in Fig. 2d. Patients with BM showed the highest mRNA expression of NFAT5 ($BRA = 5.08 \times 10^{-2}$ [95% CI 2.65×10^{-2}–10.03×10^{-2}] compared to the HEP group ($HEP = 3.00 \times 10^{-2}$ [95% CI 1.32×10^{-2}–4.18×10^{-2}]) and the M0 group ($M0 = 1.78 \times 10^{-2}$ [95% CI 0.81×10^{-2}–6.36×10^{-2} global $p = 0.02$). Expression in each case of the BRA group exceeded the threshold value for high NFAT5 expression (BRA: 100.0%; HEP: 60.0%; M0: 42.9%; global $p = 0.004$). Regarding ACVR1C, patients with BM had a stronger ACVR1C expression ($BRA = 2.69 \times 10^{-3}$ [95% CI 0–13.74×10^{-3}]; $HEP = 0.18 \times 10^{-3}$ [95% CI 0–1.71×10^{-3}]; $M0 = 0.33 \times 10^{-3}$ [95% CI 0–4.72×10^{-3}]) than patients in the HEP and M0 group nearly reaching the level of significance (global $p = 0.08$). High ACVR1C expression was observed more frequently in the BRA group (42.9%) than in the HEP (0%) and M0 (7.1%) group (global $p = 0.01$). Furthermore, the presence of BM was associated with stronger SMC3 expressions than observed in the HEP and M0 group ($BRA = 8.16 \times 10^{-2}$ [95% CI 4.65×10^{-2}–21.85×10^{-2}]; $HEP = 6.63 \times 10^{-2}$ [95% CI 4.13×10^{-2}–9.73×10^{-2}]; $M0 = 6.95 \times 10^{-2}$ [95% CI 3.68×10^{-2}–9.63×10^{-2}]; global $p = 0.34$). High SMC3 expression was only detectable in the BRA group (35.7%), whereas no case of the HEP and M0 group (global $p = 0.003$) displayed high SMC3 expression.

Table 1 Patient and tumor baseline characteristics of the entire study population (*N*=45)

Total *N*=45	BRA *N*=15 (%)	HEP *N*=15 (%)	M0 *N*=15 (%)	*P* Global	BRA vs. HEP
Sex					
Female	9 (60.0)	7 (46.7)	10 (66.7)	0.53	0.46
Male	6 (40.0)	8 (53.3)	5 (33.3)		
Age at first diagnosis of CRC					
Mean, years	64.1	62.5	73.3	0.004	0.66
≥ 70 years	4 (26.7)	2 (13.3)	11 (73.3)	0.002	0.36
≥ 65 years	7 (46.7)	7 (46.7)	13 (86.7)	0.04	1.00
< 65 years	8 (53.3)	8 (53.3)	2 (13.3)		
Sidedness of primary					
Right colon	6 (40.0)	6 (40.0)	6 (40.0)	1.00	1.00
Left colon	9 (60.0)	9 (60.0)	9 (60.0)		
Primary tumour site					
Colon	9 (60.0)	9 (60.0)	9 (60.0)		
Rectosigmoid	–	–	–	1.00	1.00
Rectum	6 (40.0)	6 (40.0)	6 (40.0)		
Grading					
Low grade (G1, G2)	10 (66.7)	10 (66.7)	9 (60.0)	0.91	1.00
High grade (G3)	5 (33.3)	5 (33.3)	6 (40.0)		
pT stage					
pT1	–	–	–		
pT2	–	–	4 (26.7)		
pT3	11 (73.3)	12 (80.0)	7 (46.7)	0.07	0.59
pT4	3 (20.0)	3 (20.0)	4 (26.7)		
Unknown	1 (6.7)	–	–		
pN status					
pN0	4 (26.7)	4 (26.7)	4 (26.7)		
pN1	3 (20.0)	3 (20.0)	8 (53.3)		
pN2	7 (46.7)	8 (53.3)	3 (20.0)	0.25	0.79
Unknown	1 (6.7)	–	–		

Different gene expression combinations and head-to-head comparisons of the investigated organotropic patient groups

Next, we tested for the association of combined expressions of the identified DEG with certain groups. Patients with BM were characterized by the simultaneous presence of high NFAT5, ACVR1C and/or SMC3 expression. A combined high expression of NFAT5 and ACVR1C (*N*=5 [38.5%]), NFAT5 and SMC3 (*N*=5 [35.7%]), ACVR1C and SMC3 (*N*=3 [23.1%]) as well as of all three genes (*N*=3 [23.1%]) was observed exclusively in the BRA group (Table 3). No case of the HEP and M0 group expressed one of these combinations, resulting in a significant difference between groups. Head-to-head- expression comparisons between groups (Table 4) showed a significant difference in NFAT5 expression between the BRA and the M0 group (*p*=0.004) and, in ACVR1C (BRA vs. HEP group; *p*=0.04). High NFAT5, ACVR1C and SMC3 expression levels differed significantly between the BRA and M0 group (NFAT5 *p*=0.001; ACVR1C *p*=0.03; SMC3 *p*=0.01) and the BRA and HEP group (NFAT5 *p*=0.01; ACVR1C *p*=0.01; SMC3 *p*=0.01), but not between the HEP and M0 group (NFAT5 *p*=0.36, ACVR1C *p*=0.33; SMC3 *p*=1.00). Analyses of combined biomarker expressions revealed only significant differences when comparing high combined NFAT5 and ACVR1C expression (BRA vs. M0, *p*=0.01; BRA vs. HEP, *p*=0.01) as well as high NFAT5 and SMC3 expression (BRA vs. M0, *p*=0.02; BRA vs. HEP, *p*=0.01).

To test whether the three DEG signature can be used to predict the risk of BM, a multivariate logistic regression and subsequent ROC analysis was applied. An area under the curve (AUC) of 0.78 was achieved, even though patient numbers were low (Supplementary Fig. 1), suggesting an acceptable discrimination between patients that will develop brain metastases and those who will not.
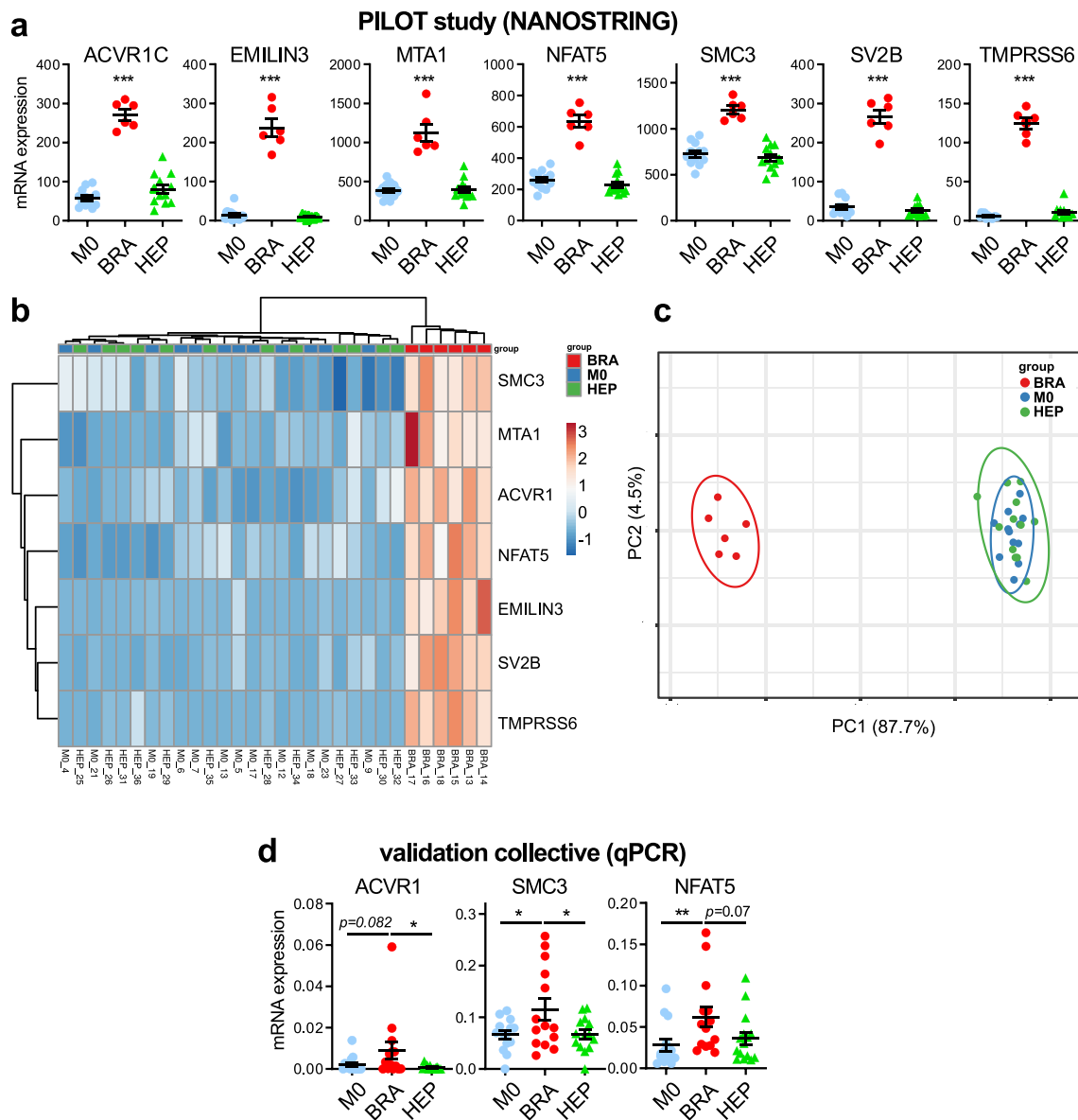
Fig. 2 Significantly differentially expressed genes (DEGs) in the M0, BRA and HEP groups. **a** Expression of BRA group associated genes as measured by nanostring analysis in the pilot study cohort (BRA, $n = 6$; M0 and HEP, $n = 12$). **b** Unsupervised hierarchical clustering using the seven DEG. **c** Principal component analysis (PCA) utilizing the seven DEG. **d** Expression of BRA group associated genes as measured in the validation collective ($n = 15$ for each group) by qPCR. **a–d** *BRA* patients with brain metastases; *HEP* patients with liver metastases; *M0* patients without metastases. *, $p < 0.05$. **$p < 0.01$. ***, $p < 0.001$

Discussion

The present work aimed to identify a gene expression signature predictive for the development of colorectal brain metastasis. Brain metastasis from CRC represents a rare event, but numbers are increasing. Still, prognosis for most CRC patients affected is outstandingly poor [10, 16] and thus, there is a high medical need to identify patients at risk. To date, the only known independent factors to predict the development of BM in CRC are the presence of lung or

bone metastasis as well as the presence of a (K)RAS mutation [5–8].

In this manuscript we report results from a case–control-analysis comparing gene expression profiles of three CRC patient cohorts with different organotropic metastatic phenotypes. The three study groups consisted of (1) CRC patients with brain metastasis (BRA group), (2) CRC patients with exclusive liver metastasis (HEP group) and CRC patients without metastatic disease within five years after CRC diagnosis (M0 group).

Table 2 Relative mRNA expression and frequency of high and low mRNA expression of NFAT5, ACVR1C and SMC3 comparing the three patient cohorts

Total $N=45$	BRA $N=15$ (95% CI)	HEP $N=15$ (95% CI)	M0 $N=15$ (95% CI)	Global P
Relative mRNA expression				
NFAT5	5.08×10^{-2} (2.65×10^{-2}; 10.03×10^{-2})	3.00×10^{-2} (1.32×10^{-2}; 4.18×10^{-2})	1.78×10^{-2} (0.81×10^{-2}; 6.36×10^{-2})	**0.02**
ACVR1C	2.69×10^{-3} (0; 13.74×10^{-3})	0.18×10^{-3} (0; 1.71×10^{-3})	0.33×10^{-3} (0; 4.72×10^{-3})	0.08
SMC3	8.16×10^{-2} (4.65×10^{-2}; 21.85×10^{-2})	6.63×10^{-2} (4.13×10^{-2}; 9.73×10^{-2})	6.95×10^{-2} (3.68×10^{-2}; 9.63×10^{-2})	0.34
Frequency of high and low mRNA expression				
NFAT5 expression	$N=14$ (%)	$N=15$ (%)	$N=14$ (%)	
Low	0 (0)	6 (40.0)	8 (57.1)	**0.004**
High	14 (100)	9 (60.0)	6 (42.9)	
ACVR1C expression	$N=14$ (%)	$N=13$ (%)	$N=14$ (%)	
Low	8 (57.1)	13 (100)	13 (92.9)	**0.01**
High	6 (42.9)	0 (0)	1 (7.1)	
SMC3 expression	$N=14$ (%)	$N=14$ (%)	$N=14$ (%)	
Low	9 (64.3)	14 (100)	14 (100)	**0.003**
High	5 (35.7)	0 (0)	0 (0)	

Significant p-values are printed in bold

Table 3 Frequency of biomarker combinations comparing the three patient cohorts. Patient numbers are indicated without outliers

Biomarker combinations	BRA		HEP		M0		Global P
	Positive	Negative	Positive	Negative	Positive	Negative	
	$N=13$ (%)		$N=13$ (%)		$N=13$ (%)		
NFAT5 high expression PLUS ACVR1C high expression	5 (38.5)	8 (61.5)	0 (0.0)	13 (100)	0 (0)	13 (100)	**0.003**
	$N=14$ (%)		$N=14$ (%)		$N=13$ (%)		
NFAT5 high expression PLUS SMC3 high expression	5 (35.7)	9 (64.3)	0 (0.0)	14 (100)	0 (0)	13 (100)	**0.004**
	$N=13$ (%)		$N=12$ (%)		$N=13$ (%)		
ACVR1C high expression PLUS SMC3 high expression	3 (23.1)	10 (76.9)	0 (0)	12 (100)	0 (0)	13 (100)	**0.04**
	$N=13$ (%)		$N=12$ (%)		$N=12$ (%)		
NFAT5 high expression PLUS ACVR1C high expression PLUS SMC3 high expression	3 (23.1)	10 (76.9)	0 (0)	12 (100)	0 (0)	12 (100)	**0.049**

Confidence intervals (CI) are shown in brackets

Significant p-values are printed in bold

For the present study, we deliberately chose an approach from the clinicians perspective and analyzed primary CRC tumor tissue facing the question whether primary CRC tissue provides information on metastatic organotropism with focus on colorectal BM. Therefore, we designed a reverse translational study "from bedside to bench" and formed the above described organotropic patient groups by incorporating a matched-pair technique to make groups as homogenous as possible. To our knowledge, no previous studies on this topic with a comparable study design have been published so far.

In the PILOT study, we identified the seven genes EMI-LIN3, MTA1, SV2B, TMPRSS6, ACVR1C, NFAT5 and SMC3, showing a significantly higher mRNA expression in

Table 4 Head-to-head comparisons between groups according to mRNA expression and biomarker combinations

	BRA↔M0	HEP↔M0	BRA↔HEP
Relative mRNA expression			
NFAT5	**0.004**	0.29	0.07
ACVR1C	0.08	0.91	**0.04**
SMC3	0.21	0.92	0.23
High mRNA expression			
NFAT5 high expression	**0.001**	0.36	**0.01**
ACVR1C high expression	**0.03**	0.33	**0.01**
SMC3 high expression	**0.01**	1.00	**0.01**
Biomarker combinations			
NFAT5 high expression PLUS *ACVR1C* high expression	**0.01**	1.00	**0.01**
NFAT5 high expression PLUS *SMC3* high expression	**0.02**	1.00	**0.01**
ACVR1C high expression PLUS *SMC3* high expression	0.07	1.00	0.08
NFAT5 high expression PLUS *ACVR1C* high expression PLUS *SMC3* high expression	0.08	1.00	0.08

Significant p-values are printed in bold

the BRA group compared to the HEP and M0 group. Results on high mRNA expression of NFAT5, ACVR1C and SMC3 in patients from the BRA group and low expression of the same three genes in the HEP and M0 group were confirmed in the validation cohort with a second independent analysis method. Moreover, we show that this three gene expression signature might have predictive impact for the formation of colorectal BM as an AUC of 0.78 was achieved even if the collective was small.

The Nuclear Factor of Activated T-cells 5 (NFAT5) was originally identified as tonicity regulated transcription factor and plays a central role in the adaptation of cells to osmotic stress [19]. NFAT5 is upregulated by hyperosmolarity caused by local inflammatory reaction e.g. induced by tumor growth [19]. Thus, strong NFAT5 expression could indicate an activated immune response that curtails tumor aggressiveness and consequently tapers down the tempo of metastatic spread. As known, mCRC patients with BM often show longer courses of disease compared to mCRC patients with liver and/or peritoneal metastasis suggesting less aggressive tumor biology or potentially a better immunologic tumor control [15, 20, 21].

The Activin A Receptor Type IC (ACVR1C) also known as ALK7 and its ligand nodal growth differentiation factor (NODAL), is a type I receptor serine/threonine kinase to

which TGF-β ligands bind [22]. By activating the subsequent signaling pathway cell proliferation is reduced [23, 24]. In a cancer stem cells enriched colorectal cancer spheroid cell model, ACVR1C was described as one of the six key molecules involved in signaling pathways for controlling various aspects of cancer stem cells [25]. Increased expression of ACVR1C seems to be associated with a less aggressive tumor growth and high ACVR1C expression was a positive prognostic factor in several tumor entities [26].

The Structural Maintenance of Chromosomes 3 (SMC3) is a member of the SMC protein family and a key regulator of DNA repair, chromosome condensation and chromosome segregation [27, 28]. Since SMC3 may also influence the activation of β-catenin [29] and this in turn can lead to EMT [30], SMC3 may directly activate metastatic growth.

Certainly, there are several limitations of the present investigation. First, low patient and tumor numbers limit meaningfulness of the presented results and merely grant this data a hypothesis-generating impact. Second, the retrospective and explorative study design implicates compromises in quality and completeness of available data on e.g. patients and tumor characteristics. However, we are convinced that for rare cancers which certainly include brain metastatic CRC acquisition of tumor tissue and data from prospective trials is not realistic.

In conclusion, the presented work identified the seven genes EMILIN3, MTA1, SV2B, TMPRSS6, ACVR1C, NFAT5 and SMC3 associated with the formation of colorectal BM during the course of disease but not with liver metastasis or non-metastatic disease. High mRNA expression of the three genes NFAT5, ACVR1C and SMC3 was confirmed with a second validation analysis technique. We suggest that primary colorectal tumors apparently contain gene expression markers which precede the formation of BM. Even if the present study cohort is small and prone to bias, considerations on carrying out such analyzes systematically are justified. Further, underlying mechanisms need to be validated in larger study cohorts and functional experiments.

tables and figures. TK, VH, MBB and JK supervised the project, had advisory function, reviewed the manuscript critically.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest Within the submitted work none of all authors has a conflict of interest to declare. However, apart from the submitted work MM received honoraria for talks and advisory boards by SIR-TeX, Roche, AstraZeneca, Novartis, Merck, Sanofi, Lilly and MSD, travel expenses by SIRTeX, Amgen, Sobi and Merck. AS received honoraria for Advisory Roles from BMS and NovoCure, honoraria for talks from Roche, Taiho Pharmaceutical, Servier and expenses for travel and accommodations from Roche, Merck KGaA, Amgen, Pfizer and Lilly Oncology. JN received honoraria for talks and advisory boards by Roche, Novartis, Pfizer, MSD, BMS, Falk and Pierre Fabre. VH received honoraria for talks, advisory boards and travel expenses by Merck, Amgen, Roche, Takeda, Servier, Pierre Fabre, Taiho, Lilly Oncology, Servier, Sanofi and Bayer Pharmaceuticals.

Ethical approval and consent to participate Ethics approval and content was obtained from the local ethics committee (Approval Number 18-105). The study was conducted in accordance with the Declaration of Helsinki.

Informed consent Informed consent was obtained from all subjects involved in the study.

References

1. Chen W, Hoffmann AD, Liu H, Liu X. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. NPJ Precis Oncol. 2018;2:4.

2. Nieder C, Spanne O, Mehta MP, Grosu AL, Geinitz H. Presentation, patterns of care, and survival in patients with brain metastases: what has changed in the last 20 years? Cancer. 2011;117:2505–12. https://doi.org/10.1002/cncr.25707.

3. Tabouret E, Chinot O, Metellus P, Tallet A, Viens P, Goncalves A. Recent trends in epidemiology of brain metastases: an overview. Anticancer Res. 2012;32:4655–62.

4. Franceschi W, Bliggenstorfer J, Sarode AL, Ginesi M, Steinhagen E, Stein SL. Primary Colorectal Tumor Location and Predictors for Metastasis to the Brain. Cureus. 2023;15: e39735. https://doi.org/10.7759/cureus.39735.

5. Lipsyc M, Yaeger R. Impact of somatic mutations on patterns of metastasis in colorectal cancer. J Gastrointest Oncol. 2015;6:645–9. https://doi.org/10.3978/j.issn.2078-6891.2015.045.

6. Qiu M, Hu J, Yang D, Cosgrove DP, Xu R. Pattern of distant metastases in colorectal cancer: a SEER based study. Oncotarget. 2015;2015(6):38658–66. https://doi.org/10.18632/oncotarget.6130.

7. Yaeger R, Cowell E, Chou JF, Gewirtz AN, Borsu L, Vakiani E, et al. RAS mutations affect pattern of metastatic spread and increase propensity for brain metastasis in colorectal cancer. Cancer. 2015;121:1195–203. https://doi.org/10.1002/cncr.29196.

8. Tie J, Lipton L, Desai J, Gibbs P, Jorissen RN, Christie M, et al. KRAS mutation is associated with lung metastasis in patients with curatively resected colorectal cancer. Clin Cancer Res. 2011;17:1122–30. https://doi.org/10.1158/1078-0432.CCR-10-1720.

9. Sperduto PW, Kased N, Roberge D, Xu Z, Shanley R, Luo X, et al. Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. J Clin Oncol. 2012;30:419–25. https://doi.org/10.1200/JCO.2011.38.0527.

10. Michl M, Thurmaier J, Schubert-Fritschle G, Wiedemann M, Laubender RP, Nussler NC, et al. Brain metastasis in colorectal cancer patients: survival and analysis of prognostic factors. Clin Colorectal Cancer. 2015;14:281–90. https://doi.org/10.1016/j.clcc.2015.05.009.

11. Desette A, Guichet PO, Emambux S, Masliantsev K, Cortes U, Ndiaye B, et al. Deciphering brain metastasis stem cell properties from colorectal cancer highlights specific stemness signature and shared molecular features. Cell Mol Gastroenterol Hepatol. 2023. https://doi.org/10.1016/j.jcmgh.2023.07.008.

12. Jacob S, Bosch F, Schoenberg MB, Pretzsch E, Lampert C, Haoyu R, et al. Expression of CIB1 correlates with colorectal liver metastases but not with peritoneal carcinomatosis. BMC Cancer. 2021;21:1243. https://doi.org/10.1186/s12885-021-08927-w.

13. Pretzsch E, Bosch F, Neumann J, Ganschow P, Bazhin A, Guba M, et al. Mechanisms of metastasis in colorectal cancer and metastatic organotropism: hematogenous versus peritoneal spread. J Oncol. 2019;2019:7407190. https://doi.org/10.1155/2019/7407190.

14. Michl M, Heinemann V, Jung A, Engel J, Kirchner T, Neumann J. Expression of cancer stem cell markers in metastatic colorectal cancer correlates with liver metastasis, but not with metastasis to the central nervous system. Pathol Res Pract. 2015;211:601–9. https://doi.org/10.1016/j.prp.2015.05.006.

15. Michl M, Taverna F, Kumbrink J, Schiergens TS, Heinemann V, Engel J, et al. Biomarker alterations associated with distinct patterns of metastatic spread in colorectal cancer. Virchows Arch. 2021;478:695–705. https://doi.org/10.1007/s00428-020-02983-6.

16. Thurmaier J, Heinemann V, Engel J, Schubert-Fritschle G, Wiedemann M, Nussler NC, et al. Patients with colorectal cancer and brain metastasis: The relevance of extracranial metastatic patterns predicting time intervals to first occurrence of intracranial metastasis and survival. Int J Cancer. 2021;148:1919–27. https://doi.org/10.1002/ijc.33364.

17. Ni H, Kumbrink J, Mayr D, Seiler A, Hagemann F, Degenhardt T, et al. Molecular prognostic factors for distant metastases in

premenopausal patients with HR+/HER2- early breast cancer. J Pers Med. 2021;11:835.

18. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res. 2015;43:W566-570. https://doi.org/10.1093/nar/gkv468.

19. Neuhofer W. Role of NFAT5 in inflammatory disorders associated with osmotic stress. Curr Genomics. 2010;11:584–90. https://doi.org/10.2174/138920210793360961.

20. Jauliac S, Lopez-Rodriguez C, Shaw LM, Brown LF, Rao A, Toker A. The role of NFAT transcription factors in integrin-mediated carcinoma invasion. Nat Cell Biol. 2002;4:540–4. https://doi.org/10.1038/ncb816.

21. Quan J, Ma C, Sun P, Wang S, Zhuang M, Liu Z, et al. Brain metastasis from colorectal cancer: clinical characteristics, timing, survival and prognostic factors. Scand J Gastroenterol. 2019;54:1370–5. https://doi.org/10.1080/00365521.2019.1686056.

22. Bondestam J, Huotari MA, Moren A, Ustinov J, Kaivo-Oja N, Kallio J, et al. cDNA cloning, expression studies and chromosome mapping of human type I serine/threonine kinase receptor ALK7 (ACVR1C). Cytogenet Cell Genet. 2001;95:157–62. https://doi.org/10.1159/000059339.

23. Li X, Ventura A. ALK7 erects a suppressive barrier to tumor progression and metastasis. Dev Cell. 2019;49:304–5. https://doi.org/10.1016/j.devcel.2019.04.025.

24. Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell. 2003;113:685–700. https://doi.org/10.1016/s0092-8674(03)00432-x.

25. Rengganaten V, Huang CJ, Tsai PH, Wang ML, Yang YP, Lan YT, et al. Mapping a circular RNA-microRNA-mRNA-signaling regulatory axis that modulates stemness properties of cancer stem cell populations in colorectal cancer spheroid cells. Int J Mol Sci. 2020;21:7864.

26. Loomans HA, Andl CD. Activin receptor-like kinases: a diverse family playing an important role in cancer. Am J Cancer Res. 2016;6:2431–47.

27. Barber TD, McManus K, Yuen KW, Reis M, Parmigiani G, Shen D, et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proc Natl Acad Sci U S A. 2008;105:3443–8. https://doi.org/10.1073/pnas.0712384105.

28. Ghiselli G, Iozzo RV. Overexpression of bamacan/SMC3 causes transformation. J Biol Chem. 2000;275:20235–8. https://doi.org/10.1074/jbc.C000213200.

29. Ghiselli G, Coffee N, Munnery CE, Koratkar R, Siracusa LD. The cohesin SMC3 is a target the for beta-catenin/TCF4 transactivation pathway. J Biol Chem. 2003;278:20259–67. https://doi.org/10.1074/jbc.M209511200.

30. Kim K, Lu Z, Hay ED. Direct evidence for a role of beta-catenin/LEF-1 signaling pathway in induction of EMT. Cell Biol Int. 2002;26:463–76.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.