





Investigating the concordance in molecular subtypes of primary colorectal tumors and their matched synchronous liver metastasis

Andreas Schlicker¹  | Architha Ellappalayam² | Ines J. Beumer² | Mireille H. J. Snel² | Lorenza Mittempergher² | Begona Diosdado³ | Christa Dreezen² | Sun Tian² | Ramon Salazar⁴ | Fotios Loupakis⁵  | Filippo Pietrantonio^{6,7}  | Cristina Santos Vivas⁴ | Maria Mercedes Martinez-Villacampa⁴ | Alberto Villanueva⁸ | Xavier Sanjuán⁹ | Marta Schirripa⁵ | Matteo Fassan¹⁰ | Antonia Martinetti⁶ | Giovanni Fucà⁶  | Sara Lonardi⁵ | Ulrich Keilholz¹¹ | Annuska M. Glas² | René Bernards^{2,3} | Loredana Vecchione^{11,12}

¹Bayer AG, Berlin, Germany

²Agendia, Amsterdam, The Netherlands

³Division of Molecular Carcinogenesis, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands

⁴Medical Oncology Department, Catalan Institute of Oncology, ONCOBELL - IDIBELL, Bellvitge Biomedical Research Institute, L'Hospitalet de Llobregat, Catalonia, Spain

⁵Department of Oncology, Veneto Institute of Oncology IOV - IRCCS, Padova, Italy

⁶Medical Oncology Department, Fondazione IRCCS, Istituto Nazionale dei Tumori, Milan, Italy

⁷Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy

⁸Translational Research Laboratory, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Institut Català d'Oncologia, Hospitalet, Barcelona, Spain

⁹Department of Pathology, Bellvitge Hospital, L'Hospitalet de Llobregat, Catalonia, Spain

¹⁰Department of Medicine, Surgical Pathology Unit, University of Padua, Padua, Italy

¹¹Charité Comprehensive Cancer Center, Berlin, Germany

¹²Department of Hematology, Oncology and Tumor Immunology (CCM) Charité - Universitaetsmedizin Berlin, Berlin, Germany

Correspondence

René Bernards, Division of Molecular Carcinogenesis, Oncode Institute, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
Email: r.bernards@nki.nl

Abstract

To date, no systematic analyses are available assessing concordance of molecular classifications between primary tumors (PT) and matched liver metastases (LM) of metastatic colorectal cancer (mCRC). We investigated concordance between PT and

Abbreviations: BRAFm-like, BRAF mutant like; BRAF, v-RAF murine sarcoma viral oncogene homolog B; BRAFwt-like, BRAF wild type like; CMS, consensus molecular subtype; CMS1, consensus molecular subtype 1; CMS2, consensus molecular subtype 2; CMS3, consensus molecular subtype 3; CMS4, consensus molecular subtype 4; CRC, colorectal cancer; EB, Ethical Board; EMT, epithelial to mesenchymal transition; FFPE, formalin-fixed paraffin-embedded; HR, hazard ratio; ICF, informed consent form; ICO-IDIBELL, Catalan Institute of Oncology - Bellvitge Biomedical Research Institute; INT, Istituto Nazionale dei Tumori; IOV, Istituto Oncologico Veneto; KRAS, Kirsten Rat Sarcoma Viral Oncogene Homolog; KW-test, Kruskal-Wallis test; LM, liver metastasis; mCRC, metastatic colorectal cancer; mOS, median overall survival; MoTriColor, Molecularly guided trials strategies in patients with advanced newly molecular defined subtypes of colorectal cancer; MSI, microsatellite instable; MW-test, Mann-Whitney; OS, overall survival; PT, primary tumor; QC, quality control; SpCorr, Spearman's rank-order correlation; TGFB, transforming growth factor-beta 1; TGFβA-like, transforming growth factor-beta 1 activating-like; TGFβI-like, transforming growth factor-beta 1 inactivating-like.

Andreas Schlicker, Architha Ellappalayam, and Ines J. Beumer contributed equally to first authorship.

René Bernards and Loredana Vecchione contributed equally to senior authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC

Loredana Vecchione, Department of Hematology, Oncology and Tumor Immunology (CCM), Charité Comprehensive Cancer Center, Charité Universitätsmedizin Berlin, Charitéplatz 1/Hufelandweg 9, Ebene 4, Haus 2622, Raum 004, 10117 Berlin, Germany. Email: loredana_vecchione@hotmail.com, loredana.vecchione@charite.de

Funding information

INTRACOLOR grant funded under the programme ERA-NET on TRANSCAN 2

LM for four clinically relevant CRC gene signatures. Twenty-seven fresh and 55 formalin-fixed paraffin-embedded pairs of PT and synchronous LM of untreated mCRC patients were retrospectively collected and classified according to the MSI-like, BRAF-like, TGFβ-activated-like and the Consensus Molecular Subtypes (CMS) classification. We investigated classification concordance between PT and LM and association of TGFβ-like and CMS classification with overall survival. Fifty-one successfully profiled matched pairs were used for analyses. PT and matched LM were highly concordant in terms of BRAF-like and MSI-like signatures, (90.2% and 98% concordance, respectively). In contrast, 40% to 70% of PT that were classified as mesenchymal-like, based on the CMS and the TGFβ-like signature, respectively, lost this phenotype in their matched LM (60.8% and 76.5% concordance, respectively). This molecular switch was independent of the microenvironment composition. In addition, the significant change in subtypes was observed also by using methods developed to detect cancer cell-intrinsic subtypes. More importantly, the molecular switch did not influence the survival. PT classified as mesenchymal had worse survival as compared to nonmesenchymal PT (CMS4 vs CMS2, hazard ratio [HR] = 5.2, 95% CI = 1.5-18.5, $P = .0048$; TGFβ-like vs TGFβi-like, HR = 2.5, 95% CI = 1.1-5.6, $P = .028$). The same was not true for LM. Our study highlights that the origin of the tissue may have major consequences for precision medicine in mCRC.

KEYWORDS

colorectal cancer molecular classification, gene expression profile of primary and synchronous liver metastasis, molecular concordance between primary and liver metastasis

1 | INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers worldwide, with an estimated 1.2 million cases and over 600 000 deaths per year.¹ Due to its relatively asymptomatic progression, patients are frequently diagnosed with metastatic disease, which is associated with a five-year survival rate of around 10%.² Since biopsies and surgical tissue of metastatic lesions are difficult to obtain, treatment choice is mainly driven by the analysis of the archived primary tumor.

Coding mutations have been reported to be highly concordant between primary tumors (PT) and matched liver metastasis (LM).³ This is also the case for epigenetic and microbiome profiles.⁴⁻⁶ In contrast, copy number profiles are discordant^{7,8} possibly pointing at larger genomic differences between PT and LM.

CRC can also be classified into different molecular subtypes based on gene expression patterns.⁹⁻¹⁵ The different molecular subtypes are characterized by the activation of different biological processes, such as microsatellite instability (MSI) and immune infiltration signaling, canonical epithelial signaling activation, metabolic dysregulation and mesenchymal characteristics. Although these subgroups have different prognosis, their predictive value, especially regarding the efficacy of targeted agents, remains under investigation. In this context, the MoTriColor consortium is currently

What's new?

No systematic analyses have assessed concordance of molecular classifications between primary tumors and matched liver metastases in metastatic colorectal cancer (mCRC). Here, the authors show that 40% to 70% of primary colon tumors cease to exhibit an epithelial-to-mesenchymal transition phenotype (EMT) at the transcription level in their matched liver metastasis (LM). While EMT-positive PT show worse outcome compared to EMT-negative PT, this is not true for LM. The data argue in favor of using the primary tumor for molecular analysis rather than distant metastases. Overall, this study highlights that tissue origin may have major consequences for precision medicine in mCRC.

exploring the efficacy of specific treatment strategies in molecularly defined CRC subgroups.¹⁶ Published and validated transcriptomic signatures were mainly developed in stage II and stage III disease, while metastatic CRC (mCRC) was not systematically investigated. Moreover, most of the data were generated from archival primary tumors.

Importantly, no systematic studies have investigated the concordance of classification of PT and LM according to different gene expression signatures. Few studies have reported about similarity in transcriptomic profiles between PT and matched LM, but they were inconclusive because of their small size and inclusion of synchronous and metachronous tumors.^{17,18} Here, we aimed to systematically study PT and their matched LM and to assess if gene expression signatures currently investigated in the MoTriColor consortium¹⁶ as well as the Consensus Molecular Subtype (CMS) classification¹⁵ are concordant between matched pairs. Recently, Trumpi et al¹⁹ reported that chemotherapy can affect the molecular classification of CRC. In addition, Isella et al²⁰ showed that the features of the mesenchymal subtype can be ascribed to its stromal component. Therefore, to avoid any bias that could be related to systemic treatment and different metastatic locations, we only included untreated primary CRC and their matched synchronous LM. Moreover, we investigated if the classification of some tumors, especially the ones classified as belonging to mesenchymal subgroups, could be influenced by the tumor micro-environment. Such data could help to understand if the transcriptomic molecular profiling of PT is sufficient to inform treatment choice or if molecular profiling of matching metastases is required to guide clinicians for individualized treatment recommendations in mCRC.

2 | MATERIALS AND METHODS

2.1 | Patient samples

We collected retrospectively samples of PT and matching synchronous LM from three different academic institutions: Catalan Institute of Oncology (ICO)-Bellvitge Biomedical Research Institute (IDIBELL-Barcelona), Istituto Nazionale dei Tumori (INT-Milan) and Istituto Oncologico Veneto (IOV-Padua). Samples were collected from treatment-naïve cases with synchronous liver metastases at time of diagnosis and available clinical-pathological annotations. We restricted our study to these inclusion criteria to exclude potential effects of earlier treatments or different metastatic locations.¹⁹ Clinicopathological annotations included are reported in Table 1.

Based on these eligibility criteria, 82 matched mCRC pairs were collected. Of these, 24 fresh and 38 formalin-fixed paraffin-embedded (FFPE) tissue pairs were successfully processed and passed quality control. For 11 patients, we received both fresh and FFPE tissues of matched pairs from ICO-IDIBELL, which were used to investigate the influence of tissue preservation technique on gene expression (Figure 1).

Research was performed according to the principles of the Declaration of Helsinki. All patients were under clinical follow-up surveillance according to the Spanish or Italian National Guidelines. All patient samples and data were anonymized in accordance with national ethical guidelines²¹ and study samples had Institutional Review Board approvals for the anonymized use of archival tissues. In particular, the Institutional Review Board of the INT approved the study (study number 117/15) and all alive patients signed a written informed consent. The Ethical Board (EB) of the IOV approved the study (study

2017/70) and the local EB of ICO-IDIBELL approved the study (PR030/17; study 2017/70). For ICO-IDIBELL, none of the patients signed a written informed consent form (ICF) because patients were dead or lost during the follow-up. The Spanish law allows using tumor samples collected before 2006 without an ICF if it is not possible to have it.

2.2 | Microarray processing and quality control

Total RNA was isolated from fresh-frozen and FFPE tissues with at least 30% of tumor cells. If possible, tissue enrichment was performed for samples that did not meet these criteria. RNA isolation and microarray processing were performed as described previously.^{9,10,12,13} For fresh tissue, RNA was isolated using the RNeasy micro kit (Qiagen, Hilden, Germany). Quality was assessed using an RNA 6000 Nano total RNA-Chip (Agilent Technologies, Santa Clara, California). Only samples with RIN ≥ 6 were included in further analyses. Two hundred nanograms of total RNA were reverse transcribed, amplified and labeled with either Cy3 (sample) or Cy5 (reference sample) using the QuickAmp Labeling kit (Agilent Technologies), and subsequently purified using the Qiagen RNeasy mini kit. Cy3-labeled cDNA and Cy5-labeled cDNA were pooled (equimolar) and hybridized to the microarray.

For FFPE tissues, RNA was isolated using the RNeasy FFPE kit (Qiagen). Fifty nanograms of total RNA were reverse transcribed and amplified using the TransPLEX C-WTA whole-transcriptome amplification kit (Rubicon, Ann Arbor, Michigan) and labeled with Cy3 using the Genomic DNA Enzymatic Labeling Kit (Agilent Technologies). For the microarray processing, cDNA was hybridized to custom full genome arrays (array design based on Agilent Catalog #G2514F) and washed according to the Agilent standard hybridization protocol (Agilent Oligo Microarray Kit, Agilent Technologies). Arrays were scanned with a dual laser scanner (Agilent Technologies).

Probes that showed nonuniformity of the signal as identified by the feature extraction software were omitted from further analyses. Image analysis of the scanned arrays was performed to quantify fluorescent intensities using Feature Extraction software version 9.5 and 11.5.1.1 (Agilent Technologies), for fresh and FFPE tissues, respectively. The feature extraction process included within-array normalization, which was performed using the default method for within-array normalization of Agilent microarrays (Lowess correction method using a linear polynomial [locally weighted linear least square regression]). Background correction was not applied. The final data sets contained expression values for 32,164 unique probes for our entire cohort. Expression values were calculated as sample/reference ratios using within-array normalized signals ($\log_{10}[\text{Cy3}/\text{Cy5}]$) for fresh tissue and represented the gMeanSignal intensities for FFPE tissue.

2.3 | Data analysis

Analyses and visualization of transcriptome data were performed in R²² and RStudio. To investigate the contribution of the tissue

TABLE 1 Patients' characteristics for the successfully profiled matched pairs

	ICO		INT		IOV		P-value
	n	%	n	%	n	%	
Age at diagnosis							
Median-range	59.3	53.4-73.3	64.4	39.5-78.5	70.0	35.6-76.8	.015
Age at surgery							
Median-range	62.6	52.6-75.1	64.4	39.5-78.5	73.7	35.6-76.8	.022
Gender							
Female	7	28.0	2	15.4	4	30.8	.615
Male	18	72.0	11	84.6	9	69.2	
Tumor location							
Right	7	28.0	3	23.1	4	30.8	.055
Left	7	28.0	5	38.5	9	69.2	
Rectum	11	44.0	5	38.5	0	0.0	
Grade							
G1	20	87.0	0	0.0	0	0.0	<.001
G2	0	0.0	6	46.2	10	76.9	
G3	3	13.0	7	53.8	3	23.1	
Diameter							
Median-range	47.5	20-60	40	16-75	35	12-80	.605
MSI status							
MSS	2	8.0	12	92.3	9	69.2	<.001
MSI	0	0.0	1	7.7	1	7.7	
Missing	23	92.0	0	0.0	3	23.1	
BRAF							
Wild	0	0.0	10	76.9	12	92.3	<.001
Mutant	0	0.0	1	7.7	1	7.7	
Missing	25	100.0	2	15.4	0	0.0	
KRAS							
Wild	2	8.0	6	46.2	6	46.2	<.001
Mutant	0	0.0	6	46.2	7	53.8	
Missing	23	92.0	1	7.7	0	0.0	

Note: Twenty-five matched pairs were provided by ICO (Catalan Institute of Oncology, Barcelona, Spain) while 13 matched pairs were provided both by INT (Istituto Nazionale dei Tumori di Milano, Italy) and IOV (Istituto Oncologico Veneto, Padua, Italy), respectively. The following clinical variables were considered: age at diagnosis, age at surgery, gender, tumor location (right side colon, left side colon, rectum), tumor grade (G1, G2, G3), tumor diameter, MSI-status, BRAF status, KRAS status. P values are referring to differences in distribution of clinical variables across the three different centers. Abbreviation: NA, not applicable.

preservation on gene expression levels, 11 unique patients' pairs for which both fresh and FFPE tissues were provided, underwent further analyses. In particular, 22 fresh samples (primary and metastasis) and 22 FFPE samples (primary and metastasis) were first median centered separately to remove probe specific bias and then combined together. Next, samples were clustered using the 64 genes of the MSI-like signature, the 277 genes of the TGF β a-like signature and the 266 unique genes of the CMS classification. For these 11 patients, we finally included the data from the FFPE tissue pairs, thus giving a total of 51 matched pairs. Hierarchical unsupervised clustering of these 51 matched pairs was performed on genome-wide transcriptome level. All clusterings (R version 3.1.3) were performed using the "ward.D2"

method in the hclust function, and visualized using ggplot2 (version 3.0.01), dendextend (version 1.12.0), dplyr (version 0.8.3). Colors of the dendrogram bars were generated using ColorBrewer (version 2.0).

2.4 | Signature and classification readout

Molecular subtyping was performed on the microarray transcriptome data of the 51 matched pairs, using five patented signatures for molecular subtype classification in colon cancer (MSI64-gene signature for fresh tissues, MSI-like FFPE signature, BRAF58-gene signature for fresh samples, BRAF mutant-like FFPE signature, ABC

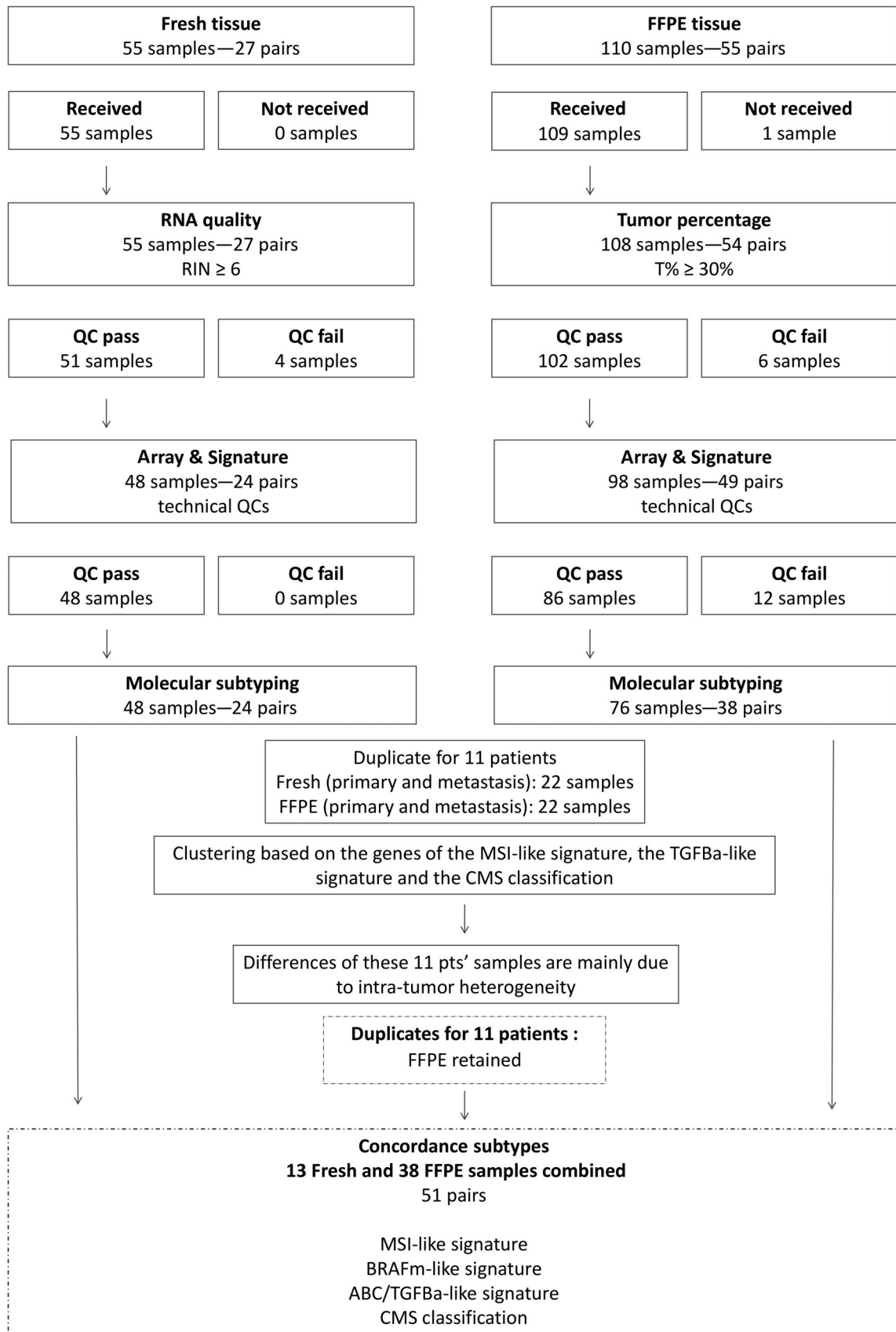


FIGURE 1 Legend on next page.

classification for fresh samples), one under patenting (TGFB activating-like signature) and the consensus classification (CMS classification). Additionally, we applied the CMScaller signatures,²³ which are based on cancer cell-intrinsic gene markers.

2.4.1 | Proprietary signatures

The MSI64-gene signature¹³ was developed using fresh tissues to identify patients with a gene expression pattern similar to patients that were MSI-high by clinical tests, and categorizes tumors as microsatellite stable (MSS)-like or unstable (MSI)-like. The BRAF58-gene signature^{10,12} was developed using fresh tissues to identify patients with a gene expression pattern similar to patients with BRAF V600E mutations, and categorizes tumors into BRAF wild-type (BRAFWt)-like and BRAF mutant (BRAFM)-like. Both signatures were also adapted for use in FFPE tissues. The ABC classification⁹ was specifically developed for fresh tissues and identifies tumors as A-type (DNA mismatch repair-deficient epithelial subtype), B-type (proliferative epithelial subtype) or C-type (mesenchymal subtype). The TGFB activated (TGFBa)-like signature was developed specifically for FFPE tissue. The signature categorizes tumors into TGFB inactivated (TGFBi)-like and activated (TGFBa)-like, with the TGFBa-like group resembling the C-type and the TGFBi-like the AB-type of the ABC classification, respectively. The TGFBa-like signature, even if not yet published, is under investigation in the frame of the MoTriColor consortium.¹⁶ Classification results were generated using proprietary software based on MATLAB (MathWorks Inc, Natick, Massachusetts).

2.4.2 | Publicly available classification

Probe sequences were aligned to the human transcriptome using NCBI-Blast to obtain the latest annotation information for generating the CMS calls. The CMS classification¹⁵ was performed in R (version 3.1.1²²) and RStudio (version 0.98.994) using the CMS calls specific for the Agilent-platform. Additionally, the CMScaller signatures²³ were performed in R (version 0.99.1) which has as input a matrix with gene expression data and a CMS template.

2.5 | Stroma percentage and microenvironment assessment

The stroma percentage of the FFPE tissue slides were visually scored in a blinded manner. The scoring percentages of the hematoxylin and

eosin (H&E) stained 5 μ M thick sections were scanned on an Aperio ScanScope XT (Leica Biosystems, Wetzlar, Germany) and uploaded to the Aperio eSlide Manager (Leica Biosystems). Pre-existing healthy tissue, necrotic and mucinous areas were excluded from the scoring. 1X amplification was used to determine the relative percentages corresponding to the desmoplastic stroma. The tumor epithelium areas were determined within the tumor field. Stromal percentage (surrogate) was defined as 100 minus the tumor percentage. Moreover, to estimate the composition of the tumor microenvironment, we utilized the Microenvironment Cell Populations-counter (MCP-counter) method²⁴ which allows a robust quantification from transcriptomic data of both immune and stromal cell populations in heterogeneous tissue.

2.6 | Statistical analysis

Data were analyzed using SPSS 22.0 for Windows (SPSS Inc. Chicago, Illinois). For all statistical analyses, a two-sided *P*-value of .05 or less was considered statistically significant.

Sample population homogeneity and normality of distribution were tested using, respectively, a Pearson chi-square statistic for categorical variables and an Independent Samples Kruskal-Wallis (KW) test for the continuous variables.

The overall concordance of molecular profile between PT and matched LM was estimated using categorical classifications for all gene expression signatures. For BRAF-like, MSI-like and ABC/TGFBa-like signatures, the switch between tumor types was calculated using generalized estimating equations to fit a repeated measures logistic regression.

Sankey plots were generated using R (version 3.6.1) and the package networkD3 (version 0.4).

The relationship between stroma percentage (S%) and tissue type (PT/LM) or molecular subtypes was investigated using a paired *t* test for paired PT and LM. For the molecular subtypes either a Mann-Whitney (MW)-test (TGFBa-like signature) or Kruskal Wallis (KW)-test (CMS classifier) was used. Spearman's rank-order correlation (SpCorr) served to measure the correlation between the S% in PT and the S% in LM. Δ S% was defined as the difference in S% between matched tissue pairs and calculated as Δ S% = S% of LM – S% of PT. An independent *t*-test was used to investigate the relationship between Δ S% and a Boolean variable indicating a switch or not in molecular subtype of the TGFBa-like signature or the CMS classifier.

Survival analyses were performed using R (version 3.6.1) and the packages survival (version 2.44) and survminer (version

FIGURE 1 Workflow design of the study. A total of 55 fresh samples and 110 FFPE samples were collected, ending in 82 matched pairs. For 11 matched pairs, both fresh and FFPE tissue were collected. Upon RNA quality control and tumor content evaluation, samples that fulfill the criteria were processed on the array. Only successfully profiled PT with their corresponding matched LM were further analyzed for the different signatures. Further, clustering of these 11 pairs based on the genes belonging to the MSI-like and TGFBa-like signature as well as the CMS classification was performed. Because differences observed among those 11 pairs were mainly due to intratumor heterogeneity, the 11 FFPE matched pairs were retained for further analyses and combined with 13 fresh and 27 FFPE pairs. Therefore 51 matched pairs were finally molecularly classified based on the MSI-like signature, the BRAF-like signature, the ABC/TGFBa-like signature and the CMS classification

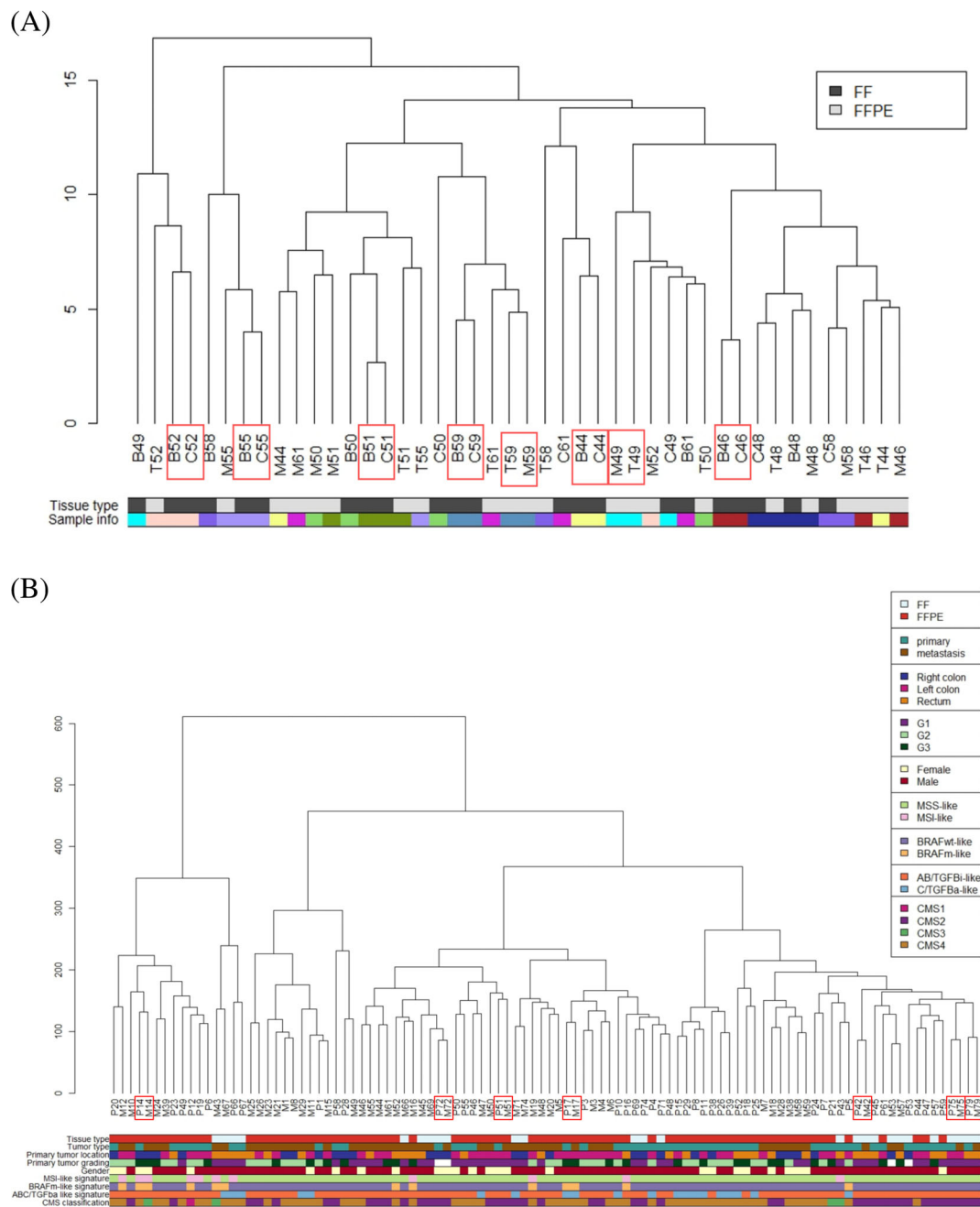


FIGURE 2 Clustering based on genes belonging to the MSI-like signature and transcriptome-wide gene expression. A, Clustering of the 11 matched pairs for which we received both fresh and FFPE tissue based on the genes belonging to the MSI-like signature (number genes = 64). Red rectangles highlight matched pairs that cluster together. T: primary tumor, FFPE tissue; M: matched liver metastasis, FFPE tissue; B: primary tumor, fresh tissue; C: matched liver metastasis, fresh tissue; Dendrogram bars: Tissue type: fresh tissue (dark gray), FFPE tissue (light gray); sample info: each color indicates samples belonging to the same patient. B, Unsupervised clustering based on genome wide transcriptomic profile of 51 matched pairs (13 fresh pairs and 38 FFPE pairs). Red rectangles highlight matched pairs that cluster together P: primary tumor; M: matched liver metastasis. Dendrogram bars: tissue type (fresh, FFPE); Tumor type (primary, metastasis); Primary tumor location (right colon, left colon, rectum); Primary tumor grading (G1, G2, G3); Gender (male, female); MSI-like signature (MSS-like, MSI-like); BRAFm-like signature (BRAFwt-like, BRAFm-like); ABC/TGFBa-like signature (AB/TGFBi-like, C/TGFBa-like); CMS classification (CMS1, CMS2, CMS3, CMS4)

0.4). The Cox proportional hazards model was used to analyze the association of molecular subtypes with overall survival (OS), which was defined as the time from surgery until

death from any cause. Kaplan-Meier curves were used to compare the survival distributions of the molecular subtypes with OS.

3 | RESULTS

3.1 | Study population

To gain insights into the concordance of the transcriptomic profiles of PT and their matched LM, we collected 82 matched mCRC samples. As summarized in Figure 1, 48 (= 24 pairs) fresh tissue samples were processed and all passed quality control (QC). The FFPE tissue cohort contained 76 samples (= 38 pairs) that were available for molecular subtyping. When we compared the success rate of sample processing on the gene expression array, we did not observe statistically significant differences between the fresh (94.4%) and the FFPE (87.8%) cohorts ($P = .259$).

We next sought to investigate if tissue preservation could have an influence on the gene expression read-out. To this end, we looked at the expression of genes belonging to the MSI-signature, the TGF β -like signature and the CMS classification in the 11 patients' pairs for which we received both fresh and FFPE tissues. Unsupervised clustering of these pairs showed that samples derived from the same patients were clustering together irrespective of tissue type (Figures 2A and S1A,B). This effect was most apparent when considering the MSI-like and TGF β -like signatures (Figures 2A and S1A). Therefore, we concluded that gene expression differences between samples from the same patient were mainly due to intratumor heterogeneity rather than tissue preservation method, as previously reported for other solid malignancies.²⁵

3.2 | Primary CRC and matched liver metastasis differ at gene expression level

We next aimed to investigate if transcriptomic profiles of the PT differed from those of their matched LM. Considering that the tissue preservation method did not influence the transcriptomic profiles, we combined the fresh and FFPE pairs. As reported in Figure 1, the final cohort of 51 successfully profiled matched pairs derived from 13 fresh pairs and 38 FFPE pairs, combined together. For patient characteristics, see Table 1. Overall, the distribution of the major clinical-pathological characteristics was similar between the three centers, except for tumor grading ($P < .001$), with samples from ICO being mainly characterized by well-differentiated tumors.

Furthermore, unsupervised clustering of the transcriptome profiles of these 51 matched pairs showed two major clusters without any obvious correlation with molecular subtyping calls or categorical clinical-pathological variables. As reported in Figure 2B, we neither observed a clear separation of PT from LM, since each cluster was characterized by both tissue types, nor a homogeneous clustering of the PT and their matching LM. Only 13% of the matched pairs (7 out of 51) clustered together indicating differences in the overall gene expression profiles. This exploratory analysis suggested that primary mCRC differ from their matched LM at the transcriptome-wide level.

3.3 | The mesenchymal profile of primary tumors is not always retained in their matched liver metastasis

Next, we aimed at comparing four established molecular gene signatures with potential clinical utility in PT and their matched LM. In particular, both PT and their matched LM were classified as MSI-like or MSS-like¹³ and as BRAF m-like or BRAF wt-like.¹⁰⁻¹² Tumors were also classified as being TGF β -like or TGF β I-like. For this purpose, we used the ABC classification⁹ for the fresh tissue samples and the TGF β -like signature for the FFPE tissue samples. It is important to note that the genes belonging to the C-group of the ABC classification and the TGF β -like group are highly overlapping and they both identify tumors showing an epithelial-mesenchymal transition (EMT) phenotype. Therefore, we classified tumors as AB/TGF β I-like or C/TGF β -like to give a uniform nomenclature for fresh and FFPE samples. Finally, the CMS classification¹⁵ was applied both to PT and their matched LM. A schematic overview of the concordance and changes of the different molecular subtypes between PT and their matched LM is reported in Figure 3 as well as in Table S1.

Overall, we observed high concordance for the BRAF-like and the MSI-like signatures between PT and LM, while lower concordance was observed for the TGF β -like signature and the CMS classification. Four PT were classified as BRAF wt-like while their matched LM were classified as BRAF m-like. One PT was classified as BRAF m-like while its matched LM was classified as BRAF wt-like (Figure 3A). The overall concordance in terms of BRAF-like signature between PT and LM was 90.2%; the number of switches was not statistically significant ($P = .177$) (Table S1). Only one matched pair was not concordant in terms of MSI signature, with the PT classified as MSI-like and its matched LM as MSS-like (Figure 3A). The overall concordance of MSI-like signature between PT and LM was 98%; the number of switches was again not statistically significant ($P = .313$; Table S1).

Two pairs switched from AB/TGF β I-like in the PT to C/TGF β -like in the LM (Figure 3A). More importantly, 10 out of 14 pairs (71%), whose PT were classified as C/TGF β -like, were classified as AB/TGF β I-like in their matched LM showing an overall concordance of 76.5% (Table S1). This significant switch ($P = .020$) was also observed for the CMS4 classification (Figure 3B). Thirteen out of 32 pairs (40.6%), whose PT were classified as CMS4, were classified as CMS2 in their corresponding LM. One pair, whose PT was classified as CMS4, was classified as CMS3 in its matched LM. Furthermore, one pair switched from CMS1 in the PT to CMS4 in the LM and one pair switched from CMS3 in the PT to CMS2 in the LM. Within 16 PT that were classified as CMS2, four pairs switched to CMS4. These results indicated a 60.8% overall concordance for the CMS classification between PT and matched LM, with major significant switches ($P = .050$) regarding the mesenchymal subtype. Overall, the switches observed, both for the C/TGF β -like signature and the CMS classification, indicate that the mesenchymal profile of 41% or 71%, depending on the classification used, is not retained in their matched LM.

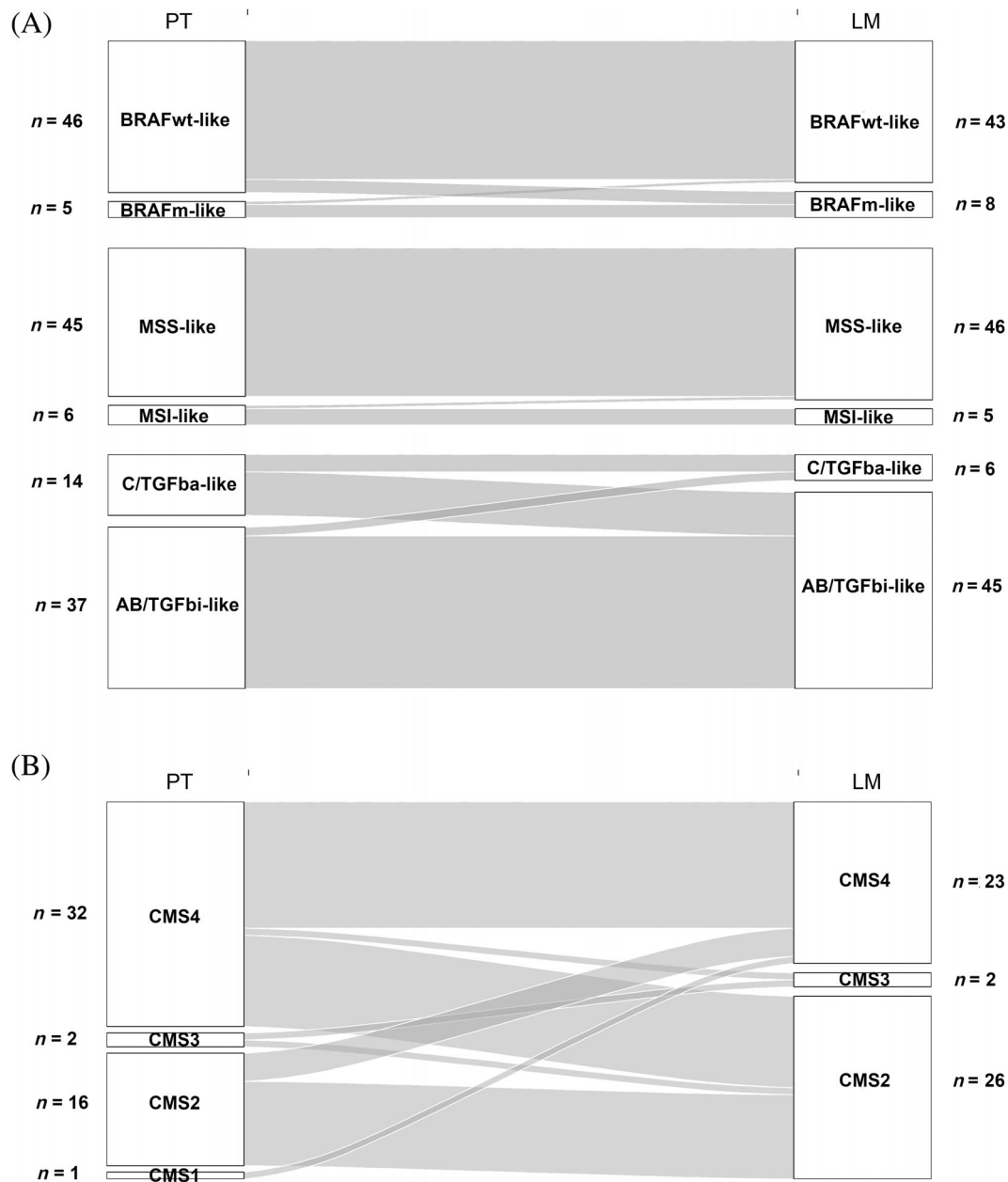


FIGURE 3 Molecular subtypes switch between PT and matched LM. A, Sankey plots showing the switch between PT and LM, in terms of MSI-like signature, BRAF-like signature, ABC/TGFba-like signature. The molecular classification of the PT, with the corresponding number of pairs classified as such, is reported on the left of the Sankey plot. The molecular classification of the matched LM, with the corresponding number of pairs classified as such, is reported on the right of the Sankey plot. B, Sankey plots showing the switch between PT and LM in terms of CMS classification. The molecular classification of the PT, with the corresponding number of pairs classified as such, is reported on the left of the Sankey plot. The molecular classification of the matched LM, with the corresponding number of pairs classified as such, is reported on the right of the Sankey plot

3.4 | The loss of mesenchymal profile between primary tumors and their matched liver metastases is independent of the tumor microenvironment

Tumor microenvironment and in particular tumor stroma might play a role in determining the mesenchymal transcriptional profile of CRC.²⁰ We therefore investigated the composition of the microenvironment

in order to better understand the switches of the mesenchymal profile observed between PT and their matched LM. To this end, we first quantified the stromal percentage in our samples. Because of a lack of further available tissue from the fresh pairs, we only considered the FFPE matched pairs for this analysis.

Overall, we observed similar population means of stroma percentages in PT and LM ($P = .097$). Also, no correlation was observed

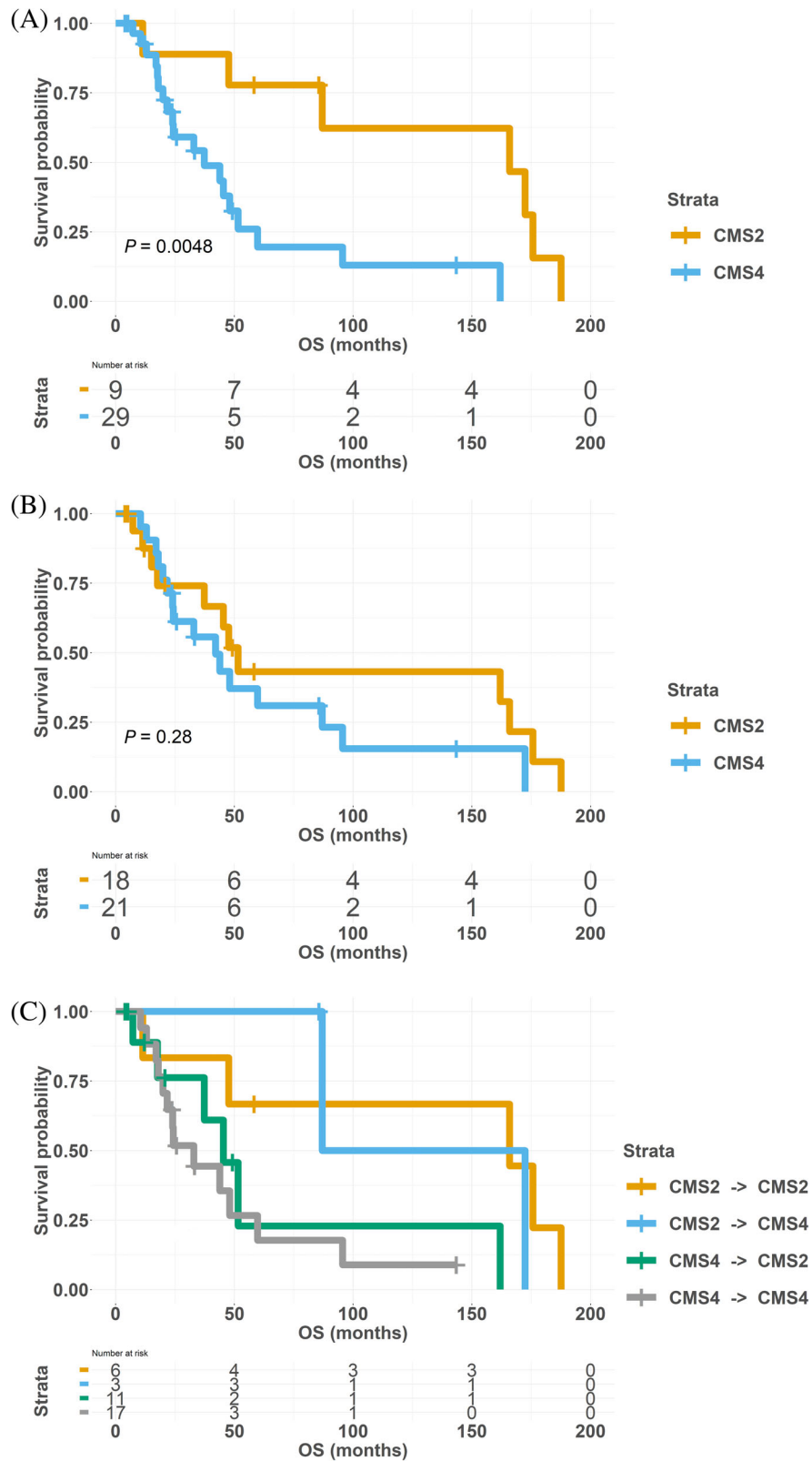


FIGURE 4 Estimate survival curves for the CMS classification. Kaplan-Meier plot of overall survival in months. A, Subjects were divided based on their primary tumor sample's CMS classification. CMS1 and CMS3 were excluded due to the small number of samples classified as such. The table below the survival plot contains the numbers of samples remaining in each group (strata) at each time point. B, Subjects were divided based on their metastatic sample's CMS classification; C, subjects were classified based on the CMS classification change from primary tumor samples to the matched metastatic sample. A two-sided P -value was not applied to C due to the small sample size of the groups

between stroma percentages in the matched pairs ($\rho = -.196$, $P = .252$). We did not observe an association between stroma percentage (S%) and CMS classification ($P = .127$). However, we observed a significantly higher S% with the TGFBA-like subtype ($P = .04$). To understand if a difference in S% could be associated with a switch from PT to LM in CMS or TGFBA classification, we analyzed if the difference in stroma percentage between PT and LM calculated as the $S\%_{LM} - S\%_{PT}$ ($\Delta S\%$) was different between switchers and non-switchers. We did not observe a significant difference in $\Delta S\%$ between nonswitchers and switchers neither for the TGFBA-like signature ($P = .607$) nor for the CMS classification ($P = .076$), indicating that the difference in S% is not associated with a switch in molecular subtype classification. To confirm this observation, we further used MCP-counter in order to robustly quantify the absolute abundance of both immune and stromal cells using the transcriptomic data of our 51 matched samples. As reported in Figure S2, we did not observe a systematic difference in microenvironment composition between PT and LM. More importantly, no pattern was observed among the samples based on the CMS classification.

In addition, because a dedicated translation of the CMS classification to metastatic organs of CRC remains pending and by considering that gene expression signals might be strongly influenced by the organ of origin, we used the CMScaller classification to compare tumor classification between PT and LM. Overall, as reported in Figure S3, we observed a better distribution of the different subtypes both in the PT and LM, with more tumors classified as CMS3 and CMS1 as compared to the CMS classification. Nevertheless, the subtype assignments varied significantly (Fisher test, $P = .0033$) between primary and metastasis.

In summary, these results indicate that the switch observed between PT and LM was not influenced by the stromal component, both evaluated as stromal percentage and by transcriptomic prediction, and that the classification of PT and matched LM are significantly different both by using the CMS classification and a classification that is based on cancer cell-intrinsic gene markers as the CMScaller.

3.5 | The molecular profile of the primary tumor determines the outcome of mCRC patients independent from the molecular profile of their matched liver metastases

Next, we investigated if differences observed in CMS classification and TGFBA-like signature between PT and LM could affect patient overall survival (OS). Median OS (mOS) for PT was 165.9 months vs 37.3 months in CMS2 and CMS4, respectively (HR = 5.2, 95% CI = 1.5-18.5, $P = .0048$; Figure 4A) and 51.6 vs 24.0 months for TGFBI-like vs TGFBA-like, respectively (HR = 2.5, 95% CI = 1.1-5.6, $P = .028$; Figure S4A). These results confirmed that tumors classified as positive for a mesenchymal phenotype have a worse prognosis when compared to tumors classified as nonmesenchymal.^{14,15} In contrast, no mOS differences were observed among LM classified as CMS2 vs CMS4 (51.6 vs 42.1 months, respectively, HR = 1.5, 95%

CI = 0.7-3.5, $P = .28$; Figure 4B) and TGFBA-like vs TGFBI-like (59.7 vs 45.4 months, respectively; Figure S4B). Finally, when we compared matched pairs that switched phenotype with the ones that did not switch phenotype, we did not observe major differences. Even if exploratory, these analyses confirmed previous observations,^{14,15} that also report mesenchymal-like tumors to have a worse outcome compared to nonmesenchymal tumors. Interestingly, this effect was independent of the transcriptomic profile of their matched LM (Figures 4C and S4C).

4 | DISCUSSION

Currently, the treatment of mCRC is based on the molecular profile of the archived primary tissue and this is sufficient in most cases to identify mutations in genes that are predictive of response to conventional biological agents.^{3,26} Nevertheless, it has been shown that primary colon tumors and their matched metastases might differ in terms of copy number alterations.^{7,8} This raises the possibility that LM could have different actionable targets as compared to their matched PT. In addition, this could imply that the transcriptomic profile of PT and their matched LM might also differ. Different molecular classifications for CRC are currently under investigation for their predictive role in response to specific treatment strategies. It is therefore important to understand if the transcriptomic profile of archived primary tumor is sufficiently informative to predict the efficacy of certain treatment or the gene expression profile of their matched LM is required.

In this retrospective study, we analyzed the concordance of gene expression signatures with potential treatment implications in 51 matched samples of primary colon tumors and their matched synchronous LM. We observed that PT did not cluster together with their matched LM on transcriptome-wide gene expression level, indicating that the biology of PT might differ from the biology of their matched LM. When we looked at the concordance of different molecular subtypes, we found that both the BRAF-like and the MSI-like signatures were highly concordant between PT and matched LM. In contrast, major discordances were observed for the CMS classification and the TGFBA-like signature. Indeed, 41% of PT that were classified as CMS4 and 70% of PT that were classified as TGFBA-like lost their mesenchymal profile in the matched LM. These differences were statistically significant. Nevertheless, possibly due to the limited sample size, no major differences were observed for the other CMS subgroups. Both the TGFBA-like signature and CMS4 are characterized by high mesenchymal gene expression, which could be attributed to stromal cells as well as to cancer cells.^{20,27-29} When we looked at the differences in stroma percentage in our FFPE cohort, we did not observe a statistically significant difference in terms of stroma percentage in LM compared to their matched PT. We found that stroma percentage was statistically significantly associated with TGFBA-like signature, but not with the CMS classification. Nevertheless, this was not associated with a change in the mesenchymal expression phenotype meaning that the switch observed between PT and LM was not influenced by the tumor stromal component. In addition, as also reported by

Sandberg et al³⁰ there was no linear association between stroma percentage and CMS classification. Because a quantification of the stromal content represents a limited description of the tumor microenvironment, we additionally applied transcriptomic signatures to quantify the stromal contribution. The MCP-counter results showed no systematic differences in microenvironment composition between PT and LM, thus validating our findings of the visual stromal quantification of the FFPE samples. Importantly, CMScaller, which was designed to focus on expression of tumor cell-specific genes, also indicated that subtype assignments of many matched PT and LM were different as we have seen using the CMS classification. We are aware that a dedicated translation of CMS classifiers to colorectal tumors from different metastatic organs remains pending and that the CMScaller, as highlighted by Eide et al,²³ in its implementation is not recommended for use with samples with a different human stromal component than primary, like biopsies and metastatic tissue. Nevertheless, based on these results we can conclude that independent of the classification used, most of the PT classified as mesenchymal by gene expression lose this phenotype in their matched LM and this is independent of the tissue in which the tumor arises and its intrinsic microenvironment.

By looking at OS differences among molecular subgroups, we could confirm that PT classified as CMS2 and TGFBI-like have significantly longer mOS as compared to CMS4 and TGFBI-like PT tumors, respectively. Surprisingly, this effect was lost when the analyses were performed using LM as the basis for subgroup classification. Finally, no substantial differences were observed in terms of mOS between PT that switched their transcriptomic profile in the matched LM from epithelial to mesenchymal and from mesenchymal to epithelial, compared to tumors that did not change their expression profile. We are aware that the survival analyses need to be considered with caution because of the small sample size. No conclusions could be derived for other molecular subgroups due to low numbers of tumors classified as MSI-like and/or CMS1 and BRAF m-like and/or CMS3. In addition, survival estimates were not adjusted for relevant clinical variables, such as kind of treatment, radical resection of liver metastasis and the presence of other metastatic lesions. Due to our inclusion criteria, patient selection did not follow predefined criteria with respect to the treatment received. Moreover, 90% of patients received liver resection while 10% of patients received liver biopsies, thus implying a potential selection bias. Finally, with respect to the molecular classification, grouping our patients by considering other clinical variables would have led to even smaller subgroups and to inconclusive results. Despite these limitations, our cohort represents a unique series of synchronous mCRC where only LM were analyzed. By keeping in mind the limitations above reported, our data suggest that the transcriptomic profile of the PT is the driver of patient outcome rather than the profile of their matched LM. This may indicate that the PT has intrinsic properties that are constant despite changes induced by a different microenvironment. Our data argue in favor of using the PT rather than the distant metastases, for molecular analyses of mCRC.

ACKNOWLEDGEMENTS

We want to particularly acknowledge the patients and the Biobank HUB-ICO-IDIBELL (PT17/0015/0024) integrated in the Spanish National Biobanks Network for their collaboration and the patients of INT and IOV. This work was supported by the INTRACOLOR grant funded under the programme ERA-NET on Translational Cancer Research (TRANSCAN 2). Dr L. V. is a participant in the BIH-Charité Clinician Scientist Program funded by the Charité-Universitätsmedizin Berlin and the Berlin Institute of Health.

CONFLICT OF INTEREST

I. J. B., M. H. J. S., L. M., A. E., R. B. and A. M. G are employed by Agendia NV. S. T., A. M. G. and R. B. are one of the named inventors on patents for the gene signatures used in our study. R. B. is also shareholder in Agendia. C. D. is consultant by Agendia NV. A. E., L. M. and A. M. G reports grants from the European Union during the conduct of the study. A. S. is employee and shareholder of Bayer AG. L. V. is a member of the GI connect group, spouse is employee and shareholder of Bayer AG. U. K. reports grants and personal fees from Astra Zeneca, Bayer, BMS, GlycoTope Merck Serono, MSD, Novartis, Pfizer, outside the submitted work. M. F. reports grants from Astellas and QED. M. F. is advisory board of Astellas and Tesaro. B. D., R. S., F. L., F. P., C. S.-V., M. M. V.-C., A. V., X. S., A. M., G. F., M. S. and S. L. declare no conflict of interest. Authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

DATA ACCESSIBILITY

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.

ORCID

Andreas Schlicker  <https://orcid.org/0000-0002-6588-7883>

Fotios Loupakis  <https://orcid.org/0000-0001-9651-0395>

Filippo Pietrantonio  <https://orcid.org/0000-0002-8530-8420>

Giovanni Fucà  <https://orcid.org/0000-0002-1560-2253>

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2014;61:69-90.
2. Cidón E. The challenge of metastatic colorectal cancer. *Clin Med Insights Oncol*. 2010;4:55-60.
3. Vakiani E, Janakiraman M, Shen R, et al. Comparative genomic analysis of primary versus metastatic colorectal carcinomas. *J Clin Oncol*. 2012;30:2956-2962.
4. Udali S, De Santis D, Ruzzenente A, et al. DNA methylation and hydroxymethylation in primary colon cancer and synchronous hepatic metastasis. *Front Genet*. 2017;8:229.
5. Konishi K, Watanabe Y, Shen L, et al. DNA methylation profiles of primary colorectal carcinoma and matched liver metastasis. *PLoS One*. 2011;6:e27889.

6. Bullman S, Pedamallu CS, Sicinska E, et al. Analysis of fusobacterium persistence and antibiotic response in colorectal cancer. *Science*. 2017;358:1443-1448.
7. Mamlouk S, Childs LH, Aust D, et al. DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer. *Nat Commun*. 2017;8:14093.
8. Kawamata F, Patch A-M, Nones K, et al. Copy number profiles of paired primary and metastatic colorectal cancers. *Oncotarget*. 2018;9:3394-3405.
9. Roepman P, Schlicker A, Tabernero J, et al. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int J Cancer*. 2013;134:552-562.
10. Tian S, Simon I, Moreno V, et al. A combined oncogenic pathway signature of BRAF, KRAS and PI3KCA mutation improves colorectal cancer classification and cetuximab treatment prediction. *Gut*. 2013;62:540-549.
11. Popovici V, Budinska E, Tejpar S, et al. Identification of a poor-prognosis BRAF-mutant - like population of patients with colon cancer. *J Clin Oncol*. 2012;30:1288-1295.
12. In't Veld S, Duong K, Snel M, et al. A computational workflow translates a 58-gene signature to a formalin-fixed, paraffin-embedded sample-based companion diagnostic for personalized treatment of the BRAF-mutation-like subtype of colorectal cancers. *High-Throughput*. 2017;6:16.
13. Tian S, Roepman P, Popovici V, et al. A robust genomic signature for the detection of colorectal cancer patients with microsatellite instability phenotype and high mutation frequency. *J Pathol*. 2012;228:586-595.
14. Huang S, Hölzel M, Knijnenburg T, et al. Europe PMC funders group MED12 controls the response to multiple cancer drugs through regulation of TGF- β receptor signaling. *Cell*. 2013;151:937-950.
15. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015;21:1350-1356.
16. Available from: <https://cordis.europa.eu/project/id/635342>.
17. Pantaleo MA, Astolfi A, Nannini M, et al. Gene expression profiling of liver metastases from colorectal cancer as potential basis for treatment choice. *Br J Cancer*. 2008;99:1729-1734.
18. Vignot S, Lefebvre C, Frampton GM, et al. Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: evaluation of concordance between genomic and transcriptional profiles. *Eur J Cancer*. 2015;51:791-799.
19. Trumpi K, Ubink I, Trinh A, et al. Neoadjuvant chemotherapy affects molecular classification of colorectal tumors. *Oncogenesis*. 2017;6:e357.
20. Isella C, Terrasi A, Bellomo SE, et al. Stromal contribution to the colorectal cancer transcriptome. *Nat Genet*. 2015;47:312-319.
21. Dutch Federation of Medical Scientific Societies. *Code for Proper Secondary Use of Human Tissues*. Rotterdam: Dutch Federation of Medical Scientific Societies; 2001 Available from: https://www.federa.org/sites/default/files/images/print_version_code_of_conduct_english.pdf.
22. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2017. Available from: <https://www.r-project.org/>
23. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep*. 2017;7:16618.
24. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17:218.
25. Beumer I, Witteveen A, Delahaye L, et al. Equivalence of MammaPrint array types in clinical trials and diagnostics. *Breast Cancer Res Treat*. 2016;156:279-287.
26. Van Cutsem E, Cervantes A, Adam R, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol*. 2016;27:1386-1422.
27. Vellinga TT, den Uil S, Rinkes IHB, et al. Collagen-rich stroma in aggressive colon tumors induces mesenchymal gene expression and tumor cell invasion. *Oncogene*. 2016;35:5263-5271.
28. Calon A, Lonardo E, Berenguer-Illergo A, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet*. 2015;47:320-329.
29. Becht E, de Reyniès A, Giraldo NA, et al. Immune and stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clin Cancer Res*. 2016;22:4057-4066.
30. Sandberg TP, Oosting J, van Pelt GW, Mesker WE, Tollenaar RAEM, Morreau H. Molecular profiling of colorectal tumors stratified by the histological tumor-stroma ratio - increased expression of galectin-1 in tumors with high stromal content. *Oncotarget*. 2018;9:31502-31515.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Schlicker A, Ellappalayam A, Beumer IJ, 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. <https://doi.org/10.1002/ijc.33003>