Proteomic characterization of transcription and splicing factors associated with a metastatic phenotype in colorectal cancer

Sofía Torres¹⁺, Irene García-Palmero¹⁺, Consuelo Marín-Vicente^{1,2}, Rubén A. Bartolomé¹, Eva Calviño¹, María Jesús Fernández-Aceñero³ and J. Ignacio Casal^{1*}

+. Equal authorship

1. Functional proteomics. Centro de Investigaciones Biológicas (CIB-CSIC). Ramiro de Maeztu 9. Madrid. Spain.
2. Proteomic facilities. CIB-CSIC. Madrid. Spain
3. Department of Pathology. Hospital Clínico. Madrid. Spain

Running title: Transcription factors in metastatic colorectal cancer

Keywords: SRSF3, transcription factors, splicing factors, metastasis, colorectal cancer

*. Corresponding author

J. Ignacio Casal

Department of Cellular and Molecular Medicine

Centro de Investigaciones Biológicas (CIB-CSIC)

Ramiro de Maeztu, 9

28040 Madrid, Spain

Phone: +34 918373112

Fax: +34 91 5360432

Email: icasal@cib.csic.es

ABSTRACT

We investigated new transcription and splicing factors associated with the metastatic phenotype in colorectal cancer. A concatenated tandem array of consensus transcription factors (TFs)-response elements was used to pull down nuclear extracts in two different pairs of colorectal cancer cells, KM12SM/KM12C and SW620/480, genetically-related but differing in metastatic ability. Proteins were analyzed by label-free LC-MS and quantified with MaxLFQ. We found 240 proteins showing a significant dysregulation in highly-metastatic KM12SM cells relative to non-metastatic KM12C cells and 257 proteins in metastatic SW620 versus SW480. In both cell lines there were similar alterations in genuine TFs and components of the splicing machinery like UPF1, TCF7L2/TCF-4, YBX1 or SRSF3. However, a significant number of alterations were cell-line specific. Functional silencing of MAFG, TFE3, TCF7L2/TCF-4 and SRSF3 in KM12 cells caused alterations in adhesion, survival, proliferation, migration and liver homing, supporting their role in metastasis. Finally, we investigated the prognostic value of the altered TFs and splicing factors in cancer patients. SRSF3 and SFPQ showed significant prognostic value. We observed that SRSF3 displayed a gradual loss of expression associated to cancer progression. Loss of SRSF3 expression was significantly associated to poor survival and shorter disease-free, particularly at early stages, in colorectal cancer.

INTRODUCTION

The transcription factors (TFs) proteome and the spliceosome are attractive targets to understand human disease given its role in the control of almost all biological processes. They represent 3–6% of the human genome ¹. To regulate transcription, TFs bind to specific DNA sequences, typically 6–12 base pairs, called the response element (RE)². However, TFs are resistant to proteomic characterization because they are frequently masked by more abundant proteins³. Less than 5% of the transcription factors (TFs) have ever been purified and characterized ⁴. The very low abundance of TFs complicates their purification, but new techniques are becoming available to make purification and analysis much less challenging and more rational ⁵. The high affinity of TFs for REs on gene promoters enabled the enrichment for low abundance TFs prior to proteomic analysis ⁵.

TFs play an essential role in the cancer progression and metastasis by regulating, among other processes, the epithelial-mesenchymal (EMT) transition ⁶⁻⁷ and the chromatin remodeling that regulates gene expression in the different metastatic steps ⁸. We aim to investigate the role of TFs in colorectal cancer, particularly in late stages. The process of liver colonization in colorectal cancer remains poorly characterized. In particular, the role of TFs dysregulation in metastasis is not well understood. Colorectal cancer is a heterogeneous disease consisting of, at least, five different consensus molecular subtypes ⁹. However, two major subgroups MSI and MSS have been defined according to microsatellite stability ¹⁰. To cover these subgroups, we have used two genetically-related metastatic cellular models, KM12C/SM and SW480/620 for the characterization of the colorectal cancer TF metastatic proteome. They represent MSI (microsatellite instable) and

MSS (microsatellite stable) types of colorectal cancer, respectively. MSI status is characterized by a deficiency in mismatch repair (MMR) ability, being MSI⁺ cells MMR deficient. MSI cancers are associated to better prognosis compared with MSS tumors ¹¹. MSI cancers tend to be more proximal, poorly differentiated, mucinous and with marked lymphocyte infiltration. In contrast SW480/620 cells are MSS cells, containing extensive deletions in APC, aneuploidy, and mutations in p53 and KRAS. On the other hand, while KM12SM cells develop strong liver metastasis ¹², SW620 cells were isolated from the lymphatic nodes of a cancer patient and are not particularly efficient for liver metastasis ¹³.

In this study, we investigated alterations in transcription and splicing regulators that play a role in colorectal cancer metastasis and might be used as prognostic factors. To that end, we used DNA sequences containing a concatenated tandem array of consensus transcription factor response elements (catTFRE) for TFs enrichment ⁵. Nuclear extracts of both cell pairs were catTFRE-pulled down for label-free quantitative proteomic analysis. A robust statistical analysis enabled the detection of minor abundance changes. A significant enrichment in transcription factors, spliceosome components and other DNA/RNA binding proteins was obtained. We identified different TFs playing a role in the metastatic progression. Finally, the splicing factor SRSF3 seems to play a significant role in the progression of colorectal cancer metastasis and as prognostic biomarker.

MATERIALS AND METHODS

Clinical samples for biomarker validation

Clinical and demographic characteristics of human samples used in the validation are indicated in Supporting Information Table S1. All the tissue samples were obtained from the Surgical Pathology Department (Fundación Jiménez Díaz, Madrid) and reviewed by two pathologists specialized in Gastrointestinal Pathology to confirm pathological stage and grade of the tumors. Colorectal carcinoma cases were randomly selected from those diagnosed in the Hospital between January 2003 and December 2009. Demographic features have been retrieved from the electronic files of the patients. Patients with TNM stages I and II did not receive further therapy after surgery, while stages III and IV received adjuvant chemotherapy, based in oxaliplatin and 5-fluorouracil. Informed written consent was obtained from all participants, as required and approved by the Research Ethics Committee of the Hospital. To investigate TFs as prognostic biomarkers, first we used *in silico* analysis with two GEO datasets (see below). For SRSF3 validation, we used 144 RNA paired samples for quantitative PCR and 187 samples for immunohistochemistry in a tissue microarray (TMA) with two 1-mm cores from each patient to get a wide representation of the whole tumor.

Cell cultures and siRNA transfections

KM12C and KM12SM human colon cancer cells were obtained from Dr Fidler's laboratory (MD Anderson Cancer Center, Houston, TX). SW480 and SW620 were purchased from the ATCC. Cells were grown in DMEM supplemented with 10% fetal calf serum (FCS) (Invitrogen), 2 mM L-glutamine and 40 µg/ml gentamicin at 37 °C in a humidified

atmosphere containing 5% CO₂. For transient silencing, KM12C cells were transfected with siRNAs targeting specifically SRSF3 (#1: SASI_Hs01_00018992; #2: 00018995) (Sigma-Aldrich), or control siRNAs using JetPrime (Polyplus). KM12SM cells were transfected with siRNAs targeting specifically MAFG (#1: SASI_Hs01_00015605; #2: 00015606), TFE3 (#1: SASI_Hs02_00341761; #2: SASI_Hs01_00111844), TCF7L2/TCF-4 (#1: SASI_Hs01_00197691; #2: 00197690), or control siRNAs. Cellular assays were performed at 48 h after transfection, as previously described ¹⁴.

Preparation of nuclear extracts

Nuclear extracts were obtained as previously described ⁷. Briefly, cells were collected in 1 ml of PBS and centrifuged at 1,500 rpm for 10 min at 4 °C. The resulting pellets containing intact cells were homogenized with ice-cold Buffer A (10 mM HEPES pH 7.5; 10 mM KCl; 0.1 mM EDTA; 0.1 mM EGTA; 1 mM DTT, 1 mM PMSF and proteases and phosphatases inhibitors (100 mM NaF, 1 mM Na₃VO₄)) by gently pipetting. After 15 min incubation on ice, NP-40 was added to a final concentration of 0.6 % (v/v) and lysates were centrifuged at 15,000 rpm for 10 min at 4 °C. The resulting supernatant containing the non-nuclear fraction (NNF) was transferred to another tube and stored at -80 °C. Nuclear pellets were washed twice with ice-cold Buffer A and resuspended in ice-cold Buffer B (20 mM Hepes pH 7.5; 0.4 M NaCl; 1 mM EDTA; 1 mM EGTA; 1 mM DTT, 1 mM PMSF, phosphatase and protease inhibitor cocktail tablets (Roche) and incubated for 30 min at 4 °C in a shaking platform to extract nuclear proteins. After centrifugation at 15,000 rpm for 10 min at 4 °C, the supernatant of the nuclear fraction (NF) was transferred to another tube and stored on ice. Nuclear pellet was re-extracted with ice-cold Buffer B repeating the same

process. Both supernatants were combined and stored at -80 °C. Protein concentration of nuclear extracts was determined by Bradford protein assay (Bio-Rad, Hercules, CA, USA).

Preparation of catTFRE Pull-Down

The catTFRE plasmid was a kind gift of Dr Qin (BPRC, China)⁵. The plasmid included a synthetic DNA containing a concatenated tandem array of 100 consensus TFREs (catTFRE) covering the majority of TF families. The catTFRE fragment was amplified by PCR with biotinylated primers as described⁵. The resulting biotinylated DNA (15 pM) was preimmobilized on streptavidin Dynabeads (M-280) and mixed with 400 µg of nuclear protein extracts from KM12C or KM12SM cells and incubated for 2 h at 4 °C in rotation.

Sample preparation for LC-MS/MS analysis

The resulting protein-catTFRE complexes were reduced, alkylated and digested first with 4 µg of Lys-C (Wako) for 2 h at 37 °C and then with 5 µg of trypsin overnight at 37 °C. Resulting peptides were desalted using C18 OMIX tips (Agilent Technologies) and kept at -80°C until LC/MS/MS analysis. Vacuum-dried peptides were reconstituted in 0.1 % formic acid. Then, three technical replicates of each sample were injected onto a precolumn C18-A1 EASY-Column (2 cm, 100 µm, 5µm) (Thermo Scientific) and run with a linear gradient of 2–35 % ACN in 0.1% aqueous solution of formic acid. The gradient was performed over 100 min using an Easy-nLC (Proxeon) at a flow-rate of 200 nl/min onto a Biosphere C18 column (15 cm, 75 µm, 3 µm) (NanoSeparations). Then, peptides were scanned and fragmented with a linear ion trap-Orbitrap Velos (Thermo Scientific). The Orbitrap Velos was operated in data-dependent mode to automatically switch between MS and MS/MS. Survey full-scan MS spectra were acquired from m/z 400 to 2000 after

accumulation to a target value of 10^6 in the Orbitrap, at a resolution of 60,000 at m/z 400. For internal mass calibration, we used the 445.0 ion for lock mass. Charge state screening was enabled and precursors with charge state unknown or 1 were excluded. After the survey scan, the 10 most intense precursor ions were selected for CID MS/MS fragmentation. For CID fragmentation, the target value was set to 10,000 and normalized collision energy to 35 %. Dynamic exclusion was applied during 30 s.

Mass spectrometry data analysis

MS data were searched against the human SwissProt database (version October 2013 with 20187 protein entries) using the Andromeda search engine integrated into the MaxQuant environment software, version 1.4.0.5¹⁵. Enzyme was set to trypsin with two missed cleavages. Carbamidomethylation of cysteines was set as fixed modification and methionine oxidation and N-terminal protein acetylation as variable modifications. For peptide identification, precursor and fragment mass tolerance were set to 20 ppm and 0.5 Da, respectively. Peptide FDR was specified as 0.01 and protein identifications were filtered for common contaminants and reverse-identified proteins. Protein quantification was calculated on the basis of the normalized spectral protein intensity using the MaxLFQ algorithm implemented in the MaxQuant environment¹⁶, setting the minimal ratio count at two and the option “match between runs”. Minimum 1 razor + unique peptides were used for quantification¹⁶. Lineal regression in scatter plot graphs and histograms were performed with Perseus (v.1.5.1.6)¹⁷ for checking reproducibility and protein abundance distribution in each replicate.

Abundance information rendered by MaxQuant software was compiled. Sample and control abundances were compared in pairwise comparisons using Student T-test for statistical significance. P-value threshold was set to 0.05. Quantitative proteomics p-value calculator¹⁸ was used to match the statistically significant proteins, based on p-value permutations, with proteins with different fold-changes.

Bioinformatic Analysis

Enrichment analysis on biological processes and subcellular location were calculated based on gene ontology terms by using DAVID (Database for Annotation, Visualization and Integrated Discovery, v6.7)¹⁹. Interaction analyses were performed using STRING (Search Tool for the Retrieval of INteracting Genes/Proteins v10)²⁰. All identified proteins were searched against the TFcheckpoint database to select transcription factors and co-factors²¹. Significantly altered transcription factors were uploaded into Genomatix pathway system (GePS) to analyze the disease association of deregulated transcription factors in KM12SM vs KM12C using “Diseases MeSH annotation”. “Characterization of gene set” option was performed. Enriched categories of neoplasm related diseases from Disease MeSH annotation were selected. Threshold p-value was set to 0.05, adjusted p-value option was enabled and also set to 0.05. *In silico* gene regulation network analyses were performed with MOTIFMAP²²⁻²³. Each database was used for a specific purpose, as stated above. Specific p-values were obtained for each of the individual tests according to the utilities recommendations.

Quantitative PCR

A set of 144 histological samples obtained from patients followed for more than five years at the Hospital Fundación Jiménez Díaz were used to validate our findings in the public databases. RNA from formalin-fixed paraffin-embedded (FFPE) tissue was isolated from 3x10µm-sections using the NucleoSpin totalRNA FFPE XS kit (Macherey-Nagel). In addition, total RNA was isolated from cellular cultures using the mirVana isolation kit (Ambion). cDNA was obtained by using SuperScript II First Strand Synthesis System with random hexamers (Invitrogen). Real-time quantitative PCR (qPCR) was performed using the FastStart Master Mix (Roche) with probes from the Universal Probe Library Set (Roche). Amplifications were run in a LightCycler® 480 (Roche). Each value was adjusted using 18S RNA levels as reference. Primer sequences and probes are shown in *Supporting Information Table S2*. Expression of SRSF3 was considered positive when the tumor/normal ratio showed a fold-change ≤ 0.5 .

Prognostic analyses using public datasets

For the public database analysis, we used two independent external cohorts in the prognostic study as previously described^{5, 24}. For biomarker training, we used the GSE17538 serie, which contains a cohort of 232 patients of colon cancer (Moffitt and VMC samples). As a training dataset, we used the GSE39582 series containing a cohort of 585 patients of colon cancer. Expression levels for all probes within each sample were transformed to Z-score and hierarchical clustering was performed using the average method combined with Pearson correlation with R software version 3.3.2. Kaplan-Meier survival curves were plotted with IBM SPSS Statistics 20.0.

Cellular assays

Cell adhesion, proliferation and invasion assays were carried out as previously described²⁵⁻²⁶. For survival assays to oxidative stress, cells were incubated with 0.5 or 1 mM H₂O₂ for 16 h in DMEM with 10% FCS. Then, cells were incubated for 1 h with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (0.6 mg/mL, Sigma-Aldrich). Cell number was determined by absorbance at 560 nm and compared with non-treated control cells.

For anoikis determination²⁷, 1x10⁵ KM12C cells were detached with trypsin, resuspended in 1 mL of DMEM with 10 % FCS and placed on a rotatory wheel at 37° C for 5 or 20 h. Then, cells were allowed to attach to a culture plate and incubated for 1 h with MTT. Cell number was determined as above and compared with control cells collected *ad initium*.

In vivo liver homing experiments

The Ethical Committee of the Consejo Superior de Investigaciones Científicas (Madrid, Spain) approved the protocols used for experimental work with mice. Swiss nude mice were tested for liver homing of metastatic cells as previously described²⁶. Spleens were surgically removed 24 h after inoculation, to prevent the local formation of tumors in the spleen.

Immunohistochemistry

TMA samples were deparaffined for antigen retrieval using citrate sodium buffer at pH 6.0 for 20 min and subsequent incubation with the primary antibody against SRSF3 (1:250; ab198291; Abcam). Then, we used EnVision Dual Link system-HRP (Dako), a horseradish peroxidase labelled polymer conjugated with secondary antibodies. The reaction was

revealed using diaminobenzidine (DAB) as chromogen and hematoxylin for counterstaining. Sections from normal colonic mucosa distant from the tumor site were used as negative controls. In order to analyze possible associations between the SRSF3 expression with different clinical characteristics, and taking into account that normal colon showed intense staining, tumoral epithelial expression was categorized as high (intense staining) and low expression (moderate and negative staining).

Statistical analysis

All the variables of the study have been described using mean (standard deviation) or percentages. Associations between variables were analyzed with the Chi-squared or Student's t test for percentage or mean comparison, respectively. The univariate survival analysis was investigated using the Kaplan-Meier plots and compared by the log-rank test. For the aim of the present study we first performed a univariate survival model to test the effect on disease-free survival of SRSF3 expression. We compared Kaplan-Meier survival curves with the log rank test to measure prognostic influence of SRSF3 immunohistochemical expression. Then, we segmented the data base and repeated this analysis according to stage, to test whether stage influences the prognostic value of SRSF3 expression. The significance was established at $p < 0.05$ for all tests. All the analyses were performed with SPSS 20.0 (IBM).

RESULTS

Proteomic characterization of transcription and splicing factors associated with a metastatic phenotype in colorectal cancer

To investigate alterations in transcription and splicing factors in colorectal cancer metastatic cells, we separated the nuclear fractions (NFs) from the non-nuclear fractions (NNFs) in KM12 and SW cell lines (Fig. 1A). The quality of the subcellular separation was confirmed by western blot using MTA2 and lamin B antibodies as markers for the nuclear fractions and RhoGDI for the non-nuclear fractions. Then, KM12 and SW nuclear extracts were pulled down with biotinylated catTFRE DNA retrieved after PCR amplification, digested with Lys-C and trypsin, and analyzed by LC-MS. MS data were searched using Andromeda and proteins were quantified with the MaxLFQ algorithm implemented in the MaxQuant environment (Fig. 1B).

The quality of the mass spectrometry data and the reliability of the identifications and quantifications of the protein groups in the catTFRE-pulled down fractions of KM12 and SW cells were examined by comparing three replicates for each analysis. The distribution of intensities among the different replicates was similar for both cell types (*Supporting Information Fig. S1A, C*). The correlation of the three replicates showed a linear trend. The R^2 values were around 0.992 for KM12C, 0.993 for KM12SM, 0.993 for SW480 and 0.992 for SW620, indicating a high correlation between the replicates (*Supporting Information Fig. S1B, D*). The comparison of linear regressions for the metastatic cell lines versus the non-metastatic KM12C and SW480 produced R^2 values around 0.950 for KM12SM/KM12C and 0.934 for SW620/SW480 cells, indicating a higher variability between metastatic and non-metastatic cells.

A total of 972 proteins were identified in the catTFRE-pulled down fractions of KM12 (931 proteins) and SW (943 proteins) cells (Supporting Information Table S3). From the total, 201 proteins were identified with 1 razor + unique peptide in at least one replicate (Supporting Information Table S4). About 690 and 755 proteins were quantified in KM12 and SW, respectively (Supporting Information Table S5) (Fig. 2A). We selected 240 and 257 proteins with a significant alteration in metastatic relative to non-metastatic cells ($p < 0.05$) (Fig. 2A) (Supporting Information Table S6). The highest overlap between statistically significant proteins and proteins with a specific fold change was obtained for a p-value of 0.05 and a fold-change of 1.4 for the upregulated proteins and 0.7 for the downregulated proteins. Those proteins significantly deregulated ($p \leq 0.05$) were represented using a volcano plot display (Fig. 2B). Although there was a significant overlap in the quantified proteins ($n=650$) for KM12 and SW cells, the similarity between the deregulated proteins was much lower ($<50\%$) (Fig. 2C), probably due to significant differences in the genetic background of the cells.

We performed GO analysis of the molecular functions associated to the altered proteins using DAVID. For KM12 cells, the top deregulated functions according to statistical significance were “*transcription, chromosome organization, response to stress, DNA damage and cell cycle*” (Supporting Information Fig. S2A). For SW cells, we found “*regulation of transcription, RNA processing, translation, intracellular transport and response to stress*” (Supporting Information Fig. S2A). Transcription and regulation of transcription were the most deregulated functions in both cell types. Other functions associated to nucleic acid-binding proteins as, for instance, components of the splicing machinery, DNA damage and repair were also highly significant.

In silico analysis of deregulated TFs after catTFRE fractionation

To look for TFs and co-factors, we interrogated the catTFRE deregulated proteins against the TFcheckpoint database ²¹, which contains TFs that form or regulate the pre-initiation transcription complex. For KM12SM/KM12C and SW620/SW480, we found 85 and 87 deregulated transcription and co-transcription factors, respectively (Supporting Information Table S7). Besides canonical TFs, the final list also included co-activators, co-repressors, histone modifiers and chromatin remodeling proteins. We analyzed the deregulated proteins in KM12SM cells with the Genomatix Pathway System (GePS) for disease association. There was a significant enrichment in TFs associated to neoplasias and tumoral development. Digestive system neoplasms were among the top related diseases (Supporting Information Fig. S2B).

STRING analysis reported a complex network of protein-protein interactions among the deregulated TFs together with the molecular functions significantly enriched ($p < 0.05$) after these proteins in both cell lines (Fig. 3). Within the KM12 up-regulated TFs, the top molecular functions of the protein networks were basic-leucine zipper domain, regulation of transcription, gene expression, high mobility group box domain, transcriptional misregulation in cancer and regulation of cell proliferation. KM12 down-regulated proteins were related with RNA metabolic process, RNA binding and processing, including a significant number of proteins of the spliceosome, and in general RNA recognition motifs. Regarding up-regulated proteins in SW620, the main molecular functions were “gene expression” and “RNA processing”. For down-regulated proteins in SW620, proteins related with “RNA splicing”, but different to KM12, and “transcriptional misregulation in cancer” (another set of proteins). In summary, there were multiple similarities in the molecular functions between KM12 and SW cells but based on different proteins.

Validation of dysregulated TFs and splicing factors

A heat-map based on similarities in the expression levels of TFs between KM12 and SW cells revealed some similarities between KM12SM and SW620, i.e. TCF7L2/TCF-4, UPF1, ETV6, SRSF3, YBX1/2, THRAP3, XPC, XRCC5/6 and MAX (Fig. 4A). In contrast, a significant number of TFs showed opposed expression ratios between KM12SM/C and SW620/480, like TFE3, MAFG, FOSL2 or CEPB. To note that SW620 cells were isolated from lymph nodes, whereas KM12SM cells were recovered from metastatic liver. Therefore, these differences might be related to organ tropism and colonization.

We verified these results by qPCR and western blot. The qPCR confirmed the up-regulation of TCF7L2/TCF4 and down-regulation of SFPQ, SRSF3 and YBX1 in both cell lines (Fig. 4B). However, TFE3, FOSL2, MAFG, and CEPBPB mRNAs were upregulated in KM12SM, but downregulated in SW620 cells. By western blot, the basal expression of MAFG and TFE3 in SW480 cells was significantly higher than in KM12C cells. In contrast, TCF7L2/TCF-4, SRSF3 and p53 expression was less variable among both cell lines (Fig. 5C). Several protein bands were observed for TCF7L2/TCF-4 and TFE3, which might correspond to the phosphorylated and non-phosphorylated forms. For p53, we also noticed the presence of two proteoforms. The low molecular weight band might correspond to the p53 β isoform. In the case of SRSF3, we observed some discrepancies between qPCR, western blot and MS values. Recent CPTAC studies in multiple cancer types showed that proteins involved in RNA splicing tend to have low mRNA-protein correlation²⁸, which seems to be the case for SRSF3, probably due to its own autoregulation²⁹.

Next, we carried out an “in silico” analysis of the regulatory networks among the altered TFs using MotifMap and Cytoscape for visualization. Each binding interaction predicted by MotifMap corresponds to a regulatory hypothesis. In the regulatory networks, we observed the presence of MAFG/MAFB, TCF-4, TFE3 or LEF-1 as major key nodes interconnected among them (Fig. 5). They regulate the expression of multiple TFs and co-factors as well as several other genes relevant in colorectal carcinogenesis. Therefore, these TFs were selected for further characterization as they seem to play a relevant role in colorectal cancer progression.

Role of altered TFs in metastasis

Next, we investigated whether knocking down TFs expression affects the pro-metastatic properties of the KM12 colorectal cancer cells. TFE3, MAFG, TCF-4 or SRSF3 were related with the critical Wnt/ β catenin pathway in colorectal cancer. Transient silencing of TFE3, TCF-4 and MAFG TFs in metastatic KM12SM and SRSF3 in KM12C cells using two different siRNAs for each TF was specific and did not affect each other expression (Fig. 6A, Supporting Information Fig. S3A). Knock-down of TFs expression after the RNAi experiments was between 70 and 80% of the original expression in KM12 cells (Fig. 6A). KM12SM metastatic cells knocked down for TFE3, TCF-4 and MAFG showed a clear reduction in cell proliferation, adhesion and, particularly, invasion (Fig. 6B, Supporting Information Fig. S3B). In contrast, SRSF3 silencing in KM12C cells caused a slight but significant and reproducible increase in adhesion (Fig. 6B, Supporting Information Fig. S3B). To further investigate the role of SRSF3, we carried out apoptosis and adhesion-related assays. These two cellular functionalities have been previously associated to SRSF3³⁰⁻³¹. SRSF3 knocked down KM2C cells exhibited a significant increase in

resistance to apoptosis due to oxidative stress and to anchorless anoikis (Fig. 6C). Finally, to confirm the pro-metastatic effects of these factors we carried out “in vivo” liver homing experiments in mouse (Fig. 6D, Supporting Information Fig. S3C). TFs-silenced cells were inoculated in the spleen of Swiss nude mice. After RT-PCR amplification of liver RNA, human GAPDH, a surrogate marker, exhibited an increase in the liver of mice inoculated with SRSF3 knocked-down KM12C cells. In contrast, KM12SM cells silenced for MAFG, TCF-4 or TFE3 lost homing capacity, as shown by a decrease in the amount of GAPDH. Collectively, these results indicate that these TFs promote pro-metastatic functionalities and liver colonization in colorectal cancer.

Correlation of TFs expression with disease-free progression in colorectal cancer

Next, we investigated whether the deregulated transcription and splicing factors could be used to predict the outcome of colorectal patients, including the relapse and the survival of the patients. Using GSE17538, a GEO dataset with information from 232 patients at stages I-IV, we carried out an *in silico* analysis of the survival prediction capacity for 85 TFs altered in KM12 cells. Eight factors showed significant association ($p \leq 0.05$) to disease-free survival: JUP, SFPQ, SRSF3, H1FX, JUND, RING1, HP1BP3 and SUB1 (Supporting Information Fig. S4A). The prognostic results were in agreement with the observed dysregulation values. Poor survival was associated to low expression for down-regulated and to high expression for up-regulated TFs. Then, we used the GSE39582 dataset containing 585 patients as a training panel. Only SRSF3 (Serine/Arginine Rich Splicing Factor 3) and SFPQ maintained a statistically significant prognostic capacity ($p=0.003$ and $p<0.0001$, respectively) to disease-free survival, associated to low expression (Supporting Information Fig. S4B).

SRSF3 expression is decreased in colorectal cancer metastasis and associates to poor prognosis

To confirm the relevance of SRSF3 as a biomarker in human patients, we assessed SRSF3 expression in tissue microarrays containing 187 patients of colorectal cancer at different stages using immunohistochemistry. There was a high expression of SRSF3 in normal colon and different levels of expression in the cancer tissues, from negative to strong expression (Fig. 7A). SRSF3 expression was inversely proportional to cancer progression. High expression was more abundant in stages I and II than in stage III or IV, in a statistically significant way. At metastatic stage IV, 80% of the samples were negative or weakly positive for SRSF3 expression ($p = 0.0001$) (Fig. 7B). As a further validation, we tested by qPCR a collection of 144 samples collected at two hospitals of Madrid (Supporting Information Table S1). We observed a significant association of low SRSF3 expression to more recurrence ($p \leq 0.010$) (Fig. 7C). In conclusion, patients with low expression of SRSF3 presented a worse outcome than those with high expression. Next, we built Kaplan-Meier curves to estimate the disease-free survival using the immunohistochemistry data. Similarly, a significant association between low expression of SRSF3 and shorter disease free survival was observed for the whole series (Fig. 7D). Then, we carried out a univariate survival analysis according to the progression stage (Fig. 7E). Stratification analysis showed an excellent prognostic association for early stages (I and II). Finally, a multivariable analysis using Cox regression was performed to study whether the correlation was stage-independent. Results showed a clear trend to significance ($p \leq 0.110$), which would be only significant in early stages of disease (TNM stages I and II) and not in advanced stages (III and IV) (Fig. 7F). Collectively, these data indicate an association

between low expression of SRSF3 and poor survival, particularly at early stages where relevant clinical decisions have to be taken.

DISCUSSION

TFs are essential for the regulation of gene expression in human health and disease. We used a novel proteomic approach, named catTFRE, to recover a highly enriched fraction of transcription, splicing factors and DNA/RNA-binding proteins in two pairs of genetically-related colorectal cancer cell lines, KM12SM/C and SW620/480 cells, differing in metastatic ability. Although the catTFRE approach was initially proposed⁵ for enrichment of TFs due to the presence of TF concatemer binding regions, in our hands many classes of RNA/DNA binding, chromatin-binding and DNA damage proteins were also enriched in the pull down procedure. Therefore, the strategy was not particularly selective for TFs. In any case, we identified a significant number of altered TFs in both cell lines. There were relatively few similarities between them, which might reflect the different metastatic ability and genetic background. This relatively low similarity among deregulated proteins agrees well with previous transcriptomic results for these cell lines³². In addition, this heterogeneity remarks the convenience of using different cell lines for proteomic analysis followed by orthogonal validations. Functional silencing of MAFG, TFE3, TCF7L2/TCF-4 and SRSF3 confirmed their pro-metastatic role in proliferation, adhesion, migration, survival and liver homing.

We observed a gradual loss of SRSF3 expression with cancer progression, which was significantly associated to poor survival and shorter disease-free. The prognostic value of SRSF3 was excellent for early stages of the disease. This is the group of patients in whom more prognosticators are needed, as they are not usually candidates to chemotherapy, although it is well known that about 15-20% of them will recur and eventually die due to disease progression. The identification of prognostic factors that can

help select patients amenable to adjuvant therapy remains a hot topic of research and we feel our results identify a new prognostic factor for this group of early stage tumors. However, it would be necessary to perform prospective clinical trials with and without chemotherapy to know whether this factor can gain widespread clinical applicability in this context.

SRSF3, also known as SRp20, is a multifunctional protein involved in alternative splicing and multiple cellular functions, such as DNA repair, RNA export, alternative RNA polyadenylation and protein translation ³³. Aberrant alternative splicing is being increasingly associated to oncogenesis and cancer progression ³⁴. Even mild changes in the expression of splicing factors might perturb alternative splicing ³¹. SRSF3 has been described as a proto-oncogene in several types of cancer ³⁵⁻³⁷. However, recent reports showed that loss of SRSF3 predisposes to hepatocellular carcinoma ³⁸ and acute myeloid leukemia ³⁹. In hepatocellular carcinoma, SRSF3 deletion caused an increase in proliferation through alterations in the insulin receptor splicing and the activation of ERK1/2 ³⁸. Our results are not conflicting with the consideration of SRSF3 as oncogene. Indeed, we observed mostly high SRSF3 expression in early stages of cancer, in contrast to the loss of SRSF3 expression at advanced and metastatic stages. The loss of SRSF3 was necessary for metastatic cells to colonize the liver microenvironment in mice. Here, we did not observe an effect of SRSF3 on cell proliferation, but in cell adhesion, resistance to oxidative stress and anoikis. SRSF3 promotes alternative splicing of some adhesion-related molecules such as CD44 ⁴⁰ and fibronectin ⁴¹ together with apoptosis-related proteins like insulin receptor⁴² and caspase-2³⁰, which may explain these changes in cell adhesion and resistance to oxidative stress and anoikis.

SRSF3 KO mice showed an increase of peroxisome proliferator activated receptor gamma (*Pparg*) expression ³⁸. Interestingly, we noticed an increase in the abundance of PPAR γ and CEP/ β in metastatic KM12SM cells. Therefore, the loss of SRSF3 may facilitate the upregulation of PPAR γ and CEP/ β , which are targets of the β -catenin pathway ⁴³. We found TCF7L2/TCF-4, a protein involved in the Wnt/ β -catenin signaling, to be upregulated in both cell lines. The β -catenin/TCF-4 pathway can modulate SRSF3 expression to regulate alternative splicing in colorectal cancer cell lines, confirming the coordinated action between transcription and alternative splicing events ⁴⁰. In summary, both mechanisms work together as two different layers for the control of gene expression ⁴⁴.

UPF1 (up-frameshift protein 1) is one of the common up-regulated proteins in KM12 and SW metastatic cells. UPF1 is the key effector of the nonsense-mediated mRNA decay (NMD) process. NMD is an important mRNA quality control mechanism, it degrades aberrant mRNAs containing premature termination codons ⁴⁵. UPF1 might constitute another layer of regulation for gene expression. A major class of splicing regulators targeted by NMD is serine–arginine (SR) proteins, including SRSF3, to regulate their own expression ⁴⁶. Therefore, high expression of UPF1 might regulate the gradual loss of SRSF3 and, probably, to downregulate gene expression as a consequence of the substantial increase of aberrant splicing in metastatic cells. In summary, TCF-4, SRSF3 and UPF1 might be interconnected for a fine regulation of transcription, alternative splicing and mRNA control.

Another mechanism for cancer progression and metastasis is through alternative splicing of gene suppressors, such as p53 or retinoblastoma ³⁴. SRSF3 participates in the

alternative splicing of p53⁴⁷. Notably, p53 isoforms are expressed differently from tumor to tumor⁴⁸. Downregulation of SRSF3 induces p53 β , a p53 isoform that promotes cellular senescence⁴⁹. Interestingly, we observed an increase of a p53 isoform, compatible with p53 β by mass spectrometry, in the metastatic colorectal cancer cells. Moreover, an increase in senescence has been associated to the dormancy state of metastatic cells to persist for months or years before final colonization⁵⁰, which might contribute to explain the association of low SRSF3 expression to disease-free prediction and survival.

Among the TFs found only in metastatic KM12SM cells, TFE3 is a basic helix-loop-helix/leucine zipper that works through dimerization with itself or with other MiT/TFE family members, such as TFEB or MITF. TFE3 promotes the expression of genes downstream of TGF β signaling and is involved in osteoclast and macrophage differentiation⁵¹. TFE3 has been involved in multiple translocations. Many of these translocations in renal cell carcinoma and PEComas involve SFPQ⁵², which was down-regulated in KM12SM cells. According to MotifMap, TFE3 regulates the expression of MAFG. In BRAF-positive human colorectal cancer cell lines and tumors, MAFG binds the promoters of MLH1 and other CpG island methylator phenotype (CIMP) genes resulting in hypermethylation and transcriptional silencing⁵³. Genes that regulate chromatin are mutated in the CIMP colorectal cancer; the highest rates of mutation were observed in CHD7 and CHD8, which encode members of the chromodomain helicase/ATP-dependent chromatin remodeling family⁵⁴. In support of this, KM12SM cells have been related to abnormal methylation and the CIMP phenotype⁵⁵, which might explain the specific alterations in TFE3 and MAFG.

CONCLUSIONS

The decline of SRSF3 expression in advanced colorectal cancer stages together with the upregulation of different TFs (mainly UPF1, TCF-4, TFE3 and MAFG) drives diverse alterations in the gene expression and the proteome of the metastatic colorectal cancer cells. These changes might affect a number of molecular mechanisms involved in adhesion, survival, invasion and metastasis. Moreover, the loss of SRSF3 expression has shown a clear and significant association to poor survival in early cancer stages and could be used as prognostic biomarker in colorectal cancer for patient risk-stratification.

Associated content

The following files are available free of charge at ACS website <http://pubs.acs.org>:

Table S1. Clinical and pathological parameters of patients with colorectal carcinoma used for qPCR and IHC validation. Table S2. Forward and reverse primers used for qPCR of selected targets. Table S3. Identified proteins in catTFRE-pulled down nuclear extracts of KM12SM/C and SW620/480. Table S4: Proteins identified with 1 razor + unique peptide in at least one replicate of catTFRE-pulled down nuclear extracts of KM12SM/C and SW620/480 Table S5. Proteins quantified by MaXLFQ in KM12SM/C and SW620/480 after catTFRE pull down. Table S6. Deregulated proteins in KM12SM vs KM12C and SW620 vs SW480. Table S7. Deregulated transcription and co-transcription factors in KM12SM vs KM12C and SW620 vs SW480 (XLSX)

Figure S1. Quality control analysis of replicates for label free quantification. Figure S2. *In silico* analysis for molecular functions and top diseases associated to catTFRE deregulated proteins. Figure S3. Functional assays to determine the effect of dysregulated TFs in metastasis. Figure S4. Association of dysregulated transcription factors expression with recurrence in colorectal cancer patients. (PDF)

ACKNOWLEDGEMENTS

We thank the Proteomics facility at the CIB, a member of the Proteored platform, for excellent technical assistance. This research was supported by grants BIO2015-66489-R from the MINECO, Foundation Ramón Areces and PRB2 (IPT13/0001-ISCIII-SGEFI/FEDER).

CONFLICT OF INTEREST

The authors have no conflict of interest to declare

DATA AVAILABILITY

Raw data files have been deposited to the ProteomeXchange Consortium (<http://www.proteomexchange.org/>) via the PRIDE partner repository. Data are available with identifier PXD006662.

REFERENCES

- (1). Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.;

- Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. The sequence of the human genome. *Science* **2001**, *291*, 1304-51.
- (2). Pan, Y.; Tsai, C. J.; Ma, B.; Nussinov, R. How do transcription factors select specific binding sites in the genome? *Nat Struct Mol Biol* **2009**, *16*, 1118-20.
- (3). Simicevic, J.; Schmid, A. W.; Gilardoni, P. A.; Zoller, B.; Raghav, S. K.; Krier, I.; Gubelmann, C.; Lisacek, F.; Naef, F.; Moniatte, M.; Deplancke, B. Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat Methods* **2013**, *10*, 570-6.
- (4). Nagore, L. I.; Nadeau, R. J.; Guo, Q.; Jadhav, Y. L.; Jarrett, H. W.; Haskins, W. E. Purification and characterization of transcription factors. *Mass Spectrom Rev* **2013**, *32*, 386-98.
- (5). Ding, C.; Chan, D. W.; Liu, W.; Liu, M.; Li, D.; Song, L.; Li, C.; Jin, J.; Malovannaya, A.; Jung, S. Y.; Zhen, B.; Wang, Y.; Qin, J. Proteome-wide profiling of activated transcription factors with a concatenated tandem array of transcription factor response elements. *Proc Natl Acad Sci U S A* **2013**, *110*, 6771-6.
- (6). Barrallo-Gimeno, A.; Nieto, M. A. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* **2005**, *132*, 3151-61.
- (7). Garcia-Palmero, I.; Torres, S.; Bartolome, R. A.; Pelaez-Garcia, A.; Larriba, M. J.; Lopez-Lucendo, M.; Pena, C.; Escudero-Paniagua, B.; Munoz, A.; Casal, J. I. Twist1-induced activation of human fibroblasts promotes matrix stiffness by upregulating palladin and collagen alpha1(VI). *Oncogene* **2016**, *35*, 5224-5236.
- (8). Gregory, R. I.; Shiekhhattar, R. Chromatin modifiers and carcinogenesis. *Trends Cell Biol* **2004**, *14*, 695-702.

- (9). Guinney, J.; Dienstmann, R.; Wang, X.; de Reynies, A.; Schlicker, A.; Soneson, C.; Marisa, L.; Roepman, P.; Nyamundanda, G.; Angelino, P.; Bot, B. M.; Morris, J. S.; Simon, I. M.; Gerster, S.; Fessler, E.; De Sousa, E. M. F.; Missiaglia, E.; Ramay, H.; Barras, D.; Homicsko, K.; Maru, D.; Manyam, G. C.; Broom, B.; Boige, V.; Perez-Villamil, B.; Laderas, T.; Salazar, R.; Gray, J. W.; Hanahan, D.; Tabernero, J.; Bernardis, R.; Friend, S. H.; Laurent-Puig, P.; Medema, J. P.; Sadanandam, A.; Wessels, L.; Delorenzi, M.; Kopetz, S.; Vermeulen, L.; Tejpar, S. The consensus molecular subtypes of colorectal cancer. *Nat Med* **2015**, *21*, 1350-6.
- (10). Thibodeau, S. N.; French, A. J.; Cunningham, J. M.; Tester, D.; Burgart, L. J.; Roche, P. C.; McDonnell, S. K.; Schaid, D. J.; Vockley, C. W.; Michels, V. V.; Farr, G. H., Jr.; O'Connell, M. J. Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. *Cancer Res* **1998**, *58*, 1713-8.
- (11). Popat, S.; Hubner, R.; Houlston, R. S. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* **2005**, *23*, 609-18.
- (12). Morikawa, K.; Walker, S. M.; Jessup, J. M.; Fidler, I. J. In vivo selection of highly metastatic cells from surgical specimens of different primary human colon carcinomas implanted into nude mice. *Cancer Res* **1988**, *48*, 1943-8.
- (13). Ding, Q.; Chang, C. J.; Xie, X.; Xia, W.; Yang, J. Y.; Wang, S. C.; Wang, Y.; Xia, J.; Chen, L.; Cai, C.; Li, H.; Yen, C. J.; Kuo, H. P.; Lee, D. F.; Lang, J.; Huo, L.; Cheng, X.; Chen, Y. J.; Li, C. W.; Jeng, L. B.; Hsu, J. L.; Li, L. Y.; Tan, A.; Curley, S. A.; Ellis, L. M.; Dubois, R. N.; Hung, M. C. APOBEC3G promotes liver metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic metastasis. *J Clin Invest* **2011**, *121*, 4526-36.
- (14). Barderas, R.; Mendes, M.; Torres, S.; Bartolome, R. A.; Lopez-Lucendo, M.; Villar-Vazquez, R.; Pelaez-Garcia, A.; Fuente, E.; Bonilla, F.; Casal, J. I. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. *Mol Cell Proteomics* **2013**, *12*, 1602-20.
- (15). Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26*, 1367-72.

- (16). Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **2014**, *13*, 2513-26.
- (17). Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **2016**, *13*, 731-40.
- (18). Chen, D.; Shah, A.; Nguyen, H.; Loo, D.; Inder, K. L.; Hill, M. M. Online quantitative proteomics p-value calculator for permutation-based statistical testing of peptide ratios. *J Proteome Res* **2014**, *13*, 4184-91.
- (19). Huang da, W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **2009**, *4*, 44-57.
- (20). Snel, B.; Lehmann, G.; Bork, P.; Huynen, M. A. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **2000**, *28*, 3442-4.
- (21). Chawla, K.; Tripathi, S.; Thommesen, L.; Laegreid, A.; Kuiper, M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics* **2013**, *29*, 2519-20.
- (22). Daily, K.; Patel, V. R.; Rigor, P.; Xie, X.; Baldi, P. MotifMap: integrative genome-wide maps of regulatory motif sites for model species. *BMC Bioinformatics* **2011**, *12*, 495.
- (23). Xie, X.; Rigor, P.; Baldi, P. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics* **2009**, *25*, 167-74.
- (24). Torres, S.; Garcia-Palmero, I.; Herrera, M.; Bartolome, R. A.; Pena, C.; Fernandez-Acenero, M. J.; Padilla, G.; Pelaez-Garcia, A.; Lopez-Lucendo, M.; Rodriguez-Merlo, R.; de Herreros, A. G.; Bonilla, F.; Casal, J. I. LOXL2 Is Highly Expressed in Cancer-Associated Fibroblasts and Associates to Poor Colon Cancer Survival. *Clin Cancer Res* **2015**, *21*, 4892-902.
- (25). Bartolome, R. A.; Pelaez-Garcia, A.; Gomez, I.; Torres, S.; Fernandez-Acenero, M. J.; Escudero-Paniagua, B.; Imbaud, J. I.; Casal, J. I. An RGD motif present in cadherin 17 induces integrin activation and tumor growth. *J Biol Chem* **2014**, *289*, 34801-14.
- (26). Bartolome, R. A.; Garcia-Palmero, I.; Torres, S.; Lopez-Lucendo, M.; Balyasnikova, I. V.; Casal, J. I. IL13 Receptor alpha2 Signaling Requires a Scaffold

Protein, FAM120A, to Activate the FAK and PI3K Pathways in Colon Cancer Metastasis. *Cancer Res* **2015**, 75, 2434-44.

(27). Swan, E. A.; Jasser, S. A.; Holsinger, F. C.; Doan, D.; Bucana, C.; Myers, J. N. Acquisition of anoikis resistance is a critical step in the progression of oral tongue cancer. *Oral Oncol* **2003**, 39, 648-55.

(28). Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M. C.; Zimmerman, L. J.; Shaddox, K. F.; Kim, S.; Davies, S. R.; Wang, S.; Wang, P.; Kinsinger, C. R.; Rivers, R. C.; Rodriguez, H.; Townsend, R. R.; Ellis, M. J.; Carr, S. A.; Tabb, D. L.; Coffey, R. J.; Slebos, R. J.; Liebler, D. C.; Nci, C. Proteogenomic characterization of human colon and rectal cancer. *Nature* **2014**, 513, 382-7.

(29). Saeki, K.; Yasugi, E.; Okuma, E.; Breit, S. N.; Nakamura, M.; Toda, T.; Kaburagi, Y.; Yuo, A. Proteomic analysis on insulin signaling in human hematopoietic cells: identification of CLIC1 and SRp20 as novel downstream effectors of insulin. *Am J Physiol Endocrinol Metab* **2005**, 289, E419-28.

(30). Jang, H. N.; Lee, M.; Loh, T. J.; Choi, S. W.; Oh, H. K.; Moon, H.; Cho, S.; Hong, S. E.; Kim, D. H.; Sheng, Z.; Green, M. R.; Park, D.; Zheng, X.; Shen, H. Exon 9 skipping of apoptotic caspase-2 pre-mRNA is promoted by SRSF3 through interaction with exon 8. *Biochim Biophys Acta* **2014**, 1839, 25-32.

(31). Corbo, C.; Orru, S.; Salvatore, F. SRp20: an overview of its role in human diseases. *Biochem Biophys Res Commun* **2013**, 436, 1-5.

(32). Hegde, P.; Qi, R.; Gaspard, R.; Abernathy, K.; Dharap, S.; Earle-Hughes, J.; Gay, C.; Nwokekeh, N. U.; Chen, T.; Saeed, A. I.; Sharov, V.; Lee, N. H.; Yeatman, T. J.; Quackenbush, J. Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res* **2001**, 61, 7792-7.

(33). Zhou, Z.; Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **2013**, 122, 191-207.

(34). Oltean, S.; Bates, D. O. Hallmarks of alternative splicing in cancer. *Oncogene* **2014**, 33, 5311-8.

(35). Jia, R.; Li, C.; McCoy, J. P.; Deng, C. X.; Zheng, Z. M. SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance. *Int J Biol Sci* **2010**, 6, 806-26.

- (36). He, X.; Arslan, A. D.; Pool, M. D.; Ho, T. T.; Darcy, K. M.; Coon, J. S.; Beck, W. T. Knockdown of splicing factor SRp20 causes apoptosis in ovarian cancer cells and its expression is associated with malignancy of epithelial ovarian cancer. *Oncogene* **2011**, *30*, 356-65.
- (37). Peiqi, L.; Zhaozhong, G.; Yaotian, Y.; Jun, J.; Jihua, G.; Rong, J. Expression of SRSF3 is Correlated with Carcinogenesis and Progression of Oral Squamous Cell Carcinoma. *Int J Med Sci* **2016**, *13*, 533-9.
- (38). Sen, S.; Langiewicz, M.; Jumaa, H.; Webster, N. J. Deletion of serine/arginine-rich splicing factor 3 in hepatocytes predisposes to hepatocellular carcinoma in mice. *Hepatology* **2015**, *61*, 171-83.
- (39). Liu, J.; Huang, B.; Xiao, Y.; Xiong, H. M.; Li, J.; Feng, D. Q.; Chen, X. M.; Zhang, H. B.; Wang, X. Z. Aberrant expression of splicing factors in newly diagnosed acute myeloid leukemia. *Onkologie* **2012**, *35*, 335-40.
- (40). Goncalves, V.; Matos, P.; Jordan, P. The beta-catenin/TCF4 pathway modifies alternative splicing through modulation of SRp20 expression. *RNA* **2008**, *14*, 2538-49.
- (41). de la Mata, M.; Kornblihtt, A. R. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol* **2006**, *13*, 973-80.
- (42). Sen, S.; Talukdar, I.; Webster, N. J. SRp20 and CUG-BP1 modulate insulin receptor exon 11 alternative splicing. *Mol Cell Biol* **2009**, *29*, 871-80.
- (43). Jansson, E. A.; Are, A.; Greicius, G.; Kuo, I. C.; Kelly, D.; Arulampalam, V.; Pettersson, S. The Wnt/beta-catenin signaling pathway targets PPARgamma activity in colon cancer cells. *Proc Natl Acad Sci U S A* **2005**, *102*, 1460-5.
- (44). Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **2006**, *126*, 37-47.
- (45). Karam, R.; Wengrod, J.; Gardner, L. B.; Wilkinson, M. F. Regulation of nonsense-mediated mRNA decay: implications for physiology and disease. *Biochim Biophys Acta* **2013**, *1829*, 624-33.
- (46). Lareau, L. F.; Inada, M.; Green, R. E.; Wengrod, J. C.; Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **2007**, *446*, 926-9.

- (47). Okumura, N.; Yoshida, H.; Kitagishi, Y.; Nishimura, Y.; Matsuda, S. Alternative splicings on p53, BRCA1 and PTEN genes involved in breast cancer. *Biochem Biophys Res Commun* **2011**, *413*, 395-9.
- (48). Bourdon, J. C.; Fernandes, K.; Murray-Zmijewski, F.; Liu, G.; Diot, A.; Xirodimas, D. P.; Saville, M. K.; Lane, D. P. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev* **2005**, *19*, 2122-37.
- (49). Tang, Y.; Horikawa, I.; Ajiro, M.; Robles, A. I.; Fujita, K.; Mondal, A. M.; Stauffer, J. K.; Zheng, Z. M.; Harris, C. C. Downregulation of splicing factor SRSF3 induces p53beta, an alternatively spliced isoform of p53 that promotes cellular senescence. *Oncogene* **2013**, *32*, 2792-8.
- (50). Aguirre-Ghiso, J. A. Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer* **2007**, *7*, 834-46.
- (51). Zanicco-Marani, T.; Vignudelli, T.; Parenti, S.; Gemelli, C.; Condorelli, F.; Martello, A.; Selmi, T.; Grande, A.; Ferrari, S. TFE3 transcription factor regulates the expression of MAFB during macrophage differentiation. *Exp Cell Res* **2009**, *315*, 1798-808.
- (52). Kauffman, E. C.; Ricketts, C. J.; Rais-Bahrami, S.; Yang, Y.; Merino, M. J.; Bottaro, D. P.; Srinivasan, R.; Linehan, W. M. Molecular genetics and cellular features of TFE3 and TFEB fusion kidney cancers. *Nat Rev Urol* **2014**, *11*, 465-75.
- (53). Fang, M.; Ou, J.; Hutchinson, L.; Green, M. R. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Mol Cell* **2014**, *55*, 904-15.
- (54). Tahara, T.; Yamamoto, E.; Madireddi, P.; Suzuki, H.; Maruyama, R.; Chung, W.; Garriga, J.; Jelinek, J.; Yamano, H. O.; Sugai, T.; Kondo, Y.; Toyota, M.; Issa, J. P.; Estecio, M. R. Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology* **2014**, *146*, 530-38 e5.
- (55). Kasuya, K.; Nagakawa, Y.; Hosokawa, Y.; Sahara, Y.; Takishita, C.; Nakajima, T.; Hijikata, Y.; Soya, R.; Katsumata, K.; Tsuchida, A. RhoA activity increases due to hypermethylation of ARHGAP28 in a highly liver-metastatic colon cancer cell line. *Biomed Rep* **2016**, *4*, 335-339.

LEGEND TO FIGURES

Figure 1. Cellular fractionation and summary of experimental workflow. (A) Nuclear fractions from non-metastatic (KM12C and SW480) and metastatic cells (KM12SM and SW620) were isolated. Concentrated samples were separated on SDS-PAGE gels, transferred to nitrocellulose membranes, and probed with the indicated antibodies. MTA2 and lamin B were used as specific markers of the nuclear fractions and RhoGDI as specific cytoplasmic marker. (B) Nuclear fractions from poorly metastatic and highly metastatic cells were isolated and incubated with a biotinylated catTFRE DNA fragment in order to bind the transcription factors differentially expressed in these cell lines. TFs were pulled down, digested with Lys-C and trypsin and analyzed by LC-MS in a LTQ-orbitrap Velos. Label-free MS analysis was performed using MaxQuant.

Figure 2. Transcription factors and co-factors differentially-expressed in the nuclear fractions of colorectal cancer cell lines. (A) Venn's diagram representing the number of proteins identified, quantified and deregulated in the nuclear fraction of KM12SM vs KM12C, or SW620 vs SW480. (B) Volcano plot showing the distribution of $-\log_{10}$ p-value versus $\log_2$ LFQ ratios of KM12SM vs KM12C and SW620 vs SW480 quantified proteins. Statistically significant ($p < 0.05$) up and down-regulated proteins are indicated in red and green, respectively, while unaltered proteins are represented in gray. (C) Venn's diagrams representing the number of proteins quantified or dysregulated in the nuclear fractions of KM12SM/KM12C and SW620/SW480. Overlapping proteins are indicated.

Figure 3. STRING analysis of dysregulated transcription factors in metastatic cells. Networks of predicted protein-protein interactions among the up- and down-regulated

transcription factors and co-factors in both cell lines were analyzed using STRING. Proteins classified in the most relevant functional categories are represented as colored nodes. Connection lines represent known and predicted protein-protein interactions according to STRING conventions. Areas within broken dashes contain those proteins functionally enriched in the indicated functional categories.

Figure 4. Validation of identified and quantified proteins. (A) Differentially-expressed proteins in KM12SM/KM12C and SW620/SW480 were represented by heat-maps of protein-expression levels according to label-free quantification. Color scale represents the value of the fold-change variation. Up-regulated proteins (red), down-regulated proteins (green), unaltered proteins (gray) and non-quantified proteins (black). (B) cDNA from non-metastatic (KM12C and SW480) and highly metastatic cells (KM12SM and SW620) was used for qPCR analysis using specific primers for selected genes and 18S rRNA for normalization. Data represent the mean $\pm$ SD of three experiments (**, $p < 0.01$; ***, $p < 0.001$). (C) Nuclear extracts of KM12C and KM12SM were analyzed by Western blot to confirm proteomic expression changes in KM12C versus KM12SM cells using indicated antibodies. Lamin B was used as loading control.

Figure 5. *In silico* prediction of regulatory networks affecting altered TFs. Predicted gene targets of the altered transcription factors and regulators of these genes were analyzed using the MotifMap system. Arrows are directed from a transcription factor to a gene. Described interactors of the dysregulated factors were obtained using GeneCards. Default parameters were applied according to the system. TFE3, MAFG, TCF4 and LEF1 compose 4 major key nodes interconnected among them, represented in rectangles. TFs and co-factors identified in our proteomics analyses are represented in dark gray, whereas not

identified TFs are represented in light gray. Cytoscape v3.5.1 was used for representing the data.

Figure 6. In vitro functional studies to determine the effect of dysregulated TFs in metastasis. (A) siRNA silencing was tested by western blot in nuclear fractions of transiently transfected cells with #1 siRNA. Lamin B was used as loading control. (B) After 24h of transfection, KM12 cells silenced for the indicated proteins, or with control siRNA, were incubated with or without serum for 24 h and subjected to MTT assays. Bar plot represents the mean $\pm$ standard deviation (***, $p < 0.001$). Adhesion assays were performed after 48 h of transfection (**, $p < 0.01$; ***, $p < 0.001$). For invasion assays, at 24h of transfection, silenced cells were seeded on top of Matrigel-coated transwell plates. Invasive cells were counted with crystal violet after 48h (***, $p < 0.001$). (C) SRSF3 knocked down KM12C cells were treated with the indicated concentrations of H₂O₂ (above) or kept in suspension for the indicated times (below) for determination of survival to oxidative stress and anoikis, respectively. Silencing of SRSF3 reduces significantly cell apoptosis (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$) compared to control cells. (D) Swiss nude mice were inoculated intrasplenically with KM12C or KM12SM cells transfected with the indicated siRNAs and sacrificed 48 h after injection. RNA was isolated from the liver and subjected to RT-PCR to amplify human GAPDH. A representative experiment (out of three) is shown. Murine β -actin was amplified as loading control.

Figure 7. Low levels of SRSF3 staining are associated to cancer progression and poor survival of patients with colorectal cancer. Immunohistochemical analysis of SRSF3 expression was performed in TMAs containing 187 colorectal cancer patients. (A) Representative images showing negative, moderate or intense staining of different

colorectal carcinomas and intense normal paired mucosa. Counterstaining was made with hematoxylin. Photographs were taken at 200X magnification. (B) Percentage of patients with negative, moderate or intense staining according to cancer stage. (C) Kaplan–Meier analyses of disease-free survival of patients according to SRSF3 expression quantified by immunohistochemistry. Differences between curves were performed using the log-rank test. (D) Further validation was carried out by qPCR of SRSF3 in a different cohort of 148 colorectal cancer patients. Kaplan Meier survival curve was built for disease-free survival. Differences between curves were performed using the log-rank test. (E) Univariate survival analysis. Kaplan-Meier survival curves were compared with the long rank test to measure prognostic influence of SRSF3 immunohistochemical expression. (F) Cox regression analysis to test whether stage influences the prognostic value of SRSF3 expression. B, regression coefficient. SE, standard error. Wald, Wald statistics $(B/SE)^2$. P, probability. CI, confidence interval.

A **B** **Figure 1**

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7