Title

RBMS1 Suppresses Colon Cancer Metastasis through Targeted Stabilization of Its mRNA Regulon

Journal

Cancer Discovery, 10(9)

ISSN

2159-8274

Authors

Yu, Johnny
Navickas, Albertas
Asgharian, Hosseinali
et al.

Publication Date

2020-09-01

DOI

10.1158/2159-8290.cd-19-1375

Peer reviewed

HHS Public Access

Author manuscript

Cancer Discov. Author manuscript; available in PMC 2021 March 01.

Published in final edited form as:

Cancer Discov. 2020 September ; 10(9): 1410–1423. doi:10.1158/2159-8290.CD-19-1375.

RBMS1 suppresses colon cancer metastasis through targeted stabilization of its mRNA regulon

Johnny Yu[1,2,3,†], **Albertas Navickas**[1,2,3,†], **Hosseinali Asgharian**[1,2,3,†], **Bruce Culbertson**[1,2,3], **Lisa Fish**[1,2,3], **Kristle Garcia**[1,2,3], **John Paolo Olegario**[1,2,3], **Maria Dermit**[7], **Martin Dodel**[7], **Benjamin Hänisch**[1,2,3], **Yikai Luo**[1,2,3], **Ethan M Weinberg**[4], **Rodrigo Dienstmann**[5], **Robert S Warren**[3,6], **Faraz K. Mardakheh**[7], **Hani Goodarzi**[1,2,3,*]

[1]Department of Biochemistry & Biophysics, University of California, San Francisco, San Francisco, California, USA.

[2]Department of Urology, University of California, San Francisco, San Francisco, California, USA.

[3]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA.

[4]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA.

[5]Medical Oncology Department, Vall d'Hebron University Hospital, Barcelona, Spain.

[6]Department of Surgery, University of California, San Francisco, San Francisco, California, USA.

[7]Centre for Cancer Cell & Molecular Biology, Barts Cancer Institute, Queen Mary University of London, Charterhouse square, London EC1M 6BQ, The United Kingdom.

Abstract

Identifying master regulators that drive pathological gene expression is a key challenge in precision oncology. Here, we have developed an analytical framework, named PRADA, that identifies oncogenic RNA-binding proteins through the systematic detection of coordinated changes in their target regulons. Application of this approach to data collected from clinical samples, patient-derived xenografts, and cell line models of colon cancer metastasis revealed the RNA-binding protein RBMS1 as a suppressor of colon cancer progression. We observed that

[*]**Corresponding author**: Hani Goodarzi, PhD, University of California, San Francisco, Genentech Hall, S312D, 600 16th Street, San Francisco, CA 94158, hani.goodarzi@ucsf.edu, 415-230-5189.

Contributions

H.G. conceived and designed the study. RNA sequencing and analysis was performed by J.Y., J.P.O., and H.G. Cell lines and mouse experiments were carried out by B.C., J.Y., K.G., E.M.W. and H.G. B.C performed immunofluorescent staining. Reporter experiments were conducted by A.N. and B.H. L.F. performed RBMS1 CLIP-seq and western blotting. A.N. performed RBMS1 Co-IP and western blotting. M.D. and M.D. performed MS analyses. F.K.M. designed and supervised MS analyses. E.M.W. generated PDX data. PRADA was developed by H.G. and H.A. R.S.W. provided patient tumor samples. Clinical analyses were performed by Y.L. and R.D. and H.G. Manuscript was written by J.Y., A.N., H.A., L.F. and H.G. H.G. supervised the project.

[†] These authors contributed equally.

Disclosure of Potential Conflicts of Interest:

No potential conflicts of interest are declared by the authors.

Data availability

The sequencing data generated as part of this study are accessible through Gene Expression Omnibus (GSE147749). All mass spectrometry raw files and their associated MaxQuant output data were deposited on ProteomeXchange Consortium via the PRIDE partner repository (PXD019478 and PXD019479).

silencing *RBMS1* results in increased metastatic capacity in xenograft mouse models, and that restoring its expression blunts metastatic liver colonization. We have found that RBMS1 functions as a post-transcriptional regulator of RNA stability by directly binding its target mRNAs. Together, our findings establish a role for RBMS1 as a previously unknown regulator of RNA stability and as a suppressor of colon cancer metastasis with clinical utility for risk stratification of patients.

Keywords

RBMS1; metastasis; colon cancer; RNA binding protein; RNA stabilization

Introduction

Metastatic progression in colorectal cancer (CRC) is accompanied by widespread gene expression reprogramming. Cancer cells often co-opt post-transcriptional regulatory mechanisms to achieve pathological expression of gene networks that drive metastasis (1–3). Colorectal cancer is the third most commonly diagnosed cancer (4), therefore understanding the underlying regulatory programs that drive metastatic progression in this disease is a crucial step towards improving patient outcomes. Notably, there are not many predictive computational methods aimed at the discovery of unknown regulatory networks. By relying on annotated post-transcriptional regulatory pathways, e.g. those mediated by microRNAs, existing methods fail to capture previously unknown regulatory interactions. To tackle this problem, we have developed a computational approach called PRADA that identifies post-transcriptional master regulators responsible for aberrant mRNA stability and gene expression in cancer cells. By applying this tool to a large compendium of gene expression profiling data from patient samples, patient-derived xenograft models, and established colon cancer cell lines, we identified a novel regulatory program involved in CRC metastasis. We find that this previously unknown regulatory pathway, which controls mRNA stability and is mediated by the RNA-binding protein RBMS1, is often silenced in highly metastatic CRC tumors. We demonstrate that RBMS1 stabilizes transcripts by binding to the last exons of target mRNAs in concert with the RNA-binding protein ELAVL1, and that in highly metastatic CRC cells and patient tumors, RBMS1 downregulation is associated with poor clinical outcome. We identify mRNA targets of RBMS1 that are functionally relevant to metastasis and reveal that *RBMS1* silencing can be accomplished through epigenetic dysregulation. This study not only describes a disease-relevant post-transcriptional regulatory mechanism that governs the stability of a sizeable regulon, but also demonstrates the value of bottom-up discovery strategies like PRADA that do not solely rely on prior knowledge of annotated regulatory programs.

Results

PRADA identifies RBMS1 as a novel regulator of CRC progression:

Metastasis is a complex multistep process and requires modulation of many cellular pathways and functions. As such, increased metastatic capacity often involves broad reprogramming of the gene expression landscape in cancer cells. Thus, a mechanistic dissection of cancer progression relies on the systematic identification of the underlying

regulatory programs that drive pathologic cellular states. To accomplish this, we developed PRADA (Prioritization of Regulatory Pathways based on Analysis of RNA Dynamic Alterations) that uses regulatory network predictions to identify the key RNA-binding proteins (RBPs) that may act as master regulators of mRNA stability and gene expression. Conceptually, PRADA solves a regression problem to predict changes in gene expression as a function of the expression of RBPs and their known or predicted targets across the transcriptome. In this customized regression analysis, the coefficient assigned to each RBP reflects its strength as a regulator of gene expression and the direction of its effect (i.e. activator or repressor). Given our limited knowledge of post-transcriptional regulatory pathways and their role in human disease, PRADA provides an opportunity for the discovery of previously unknown disease pathways. To assess the contribution of post-transcriptional regulatory programs to colon cancer metastasis, we took advantage of a publicly available set of gene expression profiles from established colon cancer cell lines (GSE59857). We first categorized these cell lines as poorly metastatic or highly metastatic based on their liver colonization capacity in xenograft mouse models (Supplementary Fig. 1A; (5)). We then performed differential gene expression analysis to compare gene expression changes across the transcriptomes of the two groups. Finally, we applied PRADA to this dataset to find RNA-binding proteins whose differential activity is most informative of metastasis-associated gene expression changes. PRADA identified the protein RBMS1 as the RNA-binding protein with the strongest regulatory potential in this dataset (Fig. 1A). The size and the direction of the regulatory coefficient assigned to RBMS1 implies that *RBMS1* levels are strongly informative of changes in the expression of transcripts with sequence matches to RBMS1 binding sites. To confirm this, we performed motif enrichment analysis using FIRE, which uses mutual information to capture the association between the presence and absence of a given binding site and genome-wide transcriptomic measurements (6). As shown in Fig. 1B, the RBMS1 consensus binding site is strongly informative of gene expression differences between poorly and highly metastatic cells, showing a highly significant enrichment among the transcripts that have decreased expression in highly metastatic cells. To ensure that this result is not sensitive to the choice of a specific RBMS1 binding motif, we tested two additional sequence models: (i) an independently derived representation of RBMS1 binding preferences (7), and (ii) predictions based on DeepBind models (8) (Supplementary Table 1). Each of these three predicted RBMS1 binding motifs gave largely identical results, and we observed a similar decrease in the expression of the putative RBMS1 regulon derived from each motif (Supplementary Fig. 1B). For consistency, we have used the DeepBind-derived RBMS1 regulon (referred to as the RBMS1 regulon hereafter) in our subsequent analyses as DeepBind is the state-of-the-art approach for predicting protein-nucleic acid interactions. Importantly, concordant with the results reported by PRADA and the lower expression of its regulon, the expression of *RBMS1* was also significantly lower in highly metastatic cells (Fig. 1C). Moreover, consistent with RBMS1 acting as a post-transcriptional regulator of these genes, we observed a significant correlation between the expression of *RBMS1* and the average expression of its regulon in multiple independent datasets (Fig. 1D, Supplementary Fig. 1C–D).

In order to further assess the association between RBMS1 and CRC metastasis in more directly disease-relevant models, we took advantage of patient-derived xenograft (PDX)

models of colon cancer metastasis to liver. We used RNA-seq data from three parental CRC PDX models (CLR4, CLR27, and CLR32) and their highly liver metastatic derivatives (CLR-LvMs). The highly metastatic CLR-LvM models were derived from the parental PDXs by repeated splenic delivery and growth of cells in the liver of immunocompromised mice (NOD scid gamma), recapitulating the major site of CRC metastasis in humans (9). We observed that the increase in the metastatic capacity of the CLR-LvM PDX models compared to their parental PDXs was accompanied by a significant reduction in the expression of the RBMS1 regulon in the highly metastatic CLR-LvM models (Fig. 1E). More importantly, all three models also showed a significant and concomitant decrease in the expression of *RBMS1* (Fig. 1F).

In addition to these cell line and PDX models of CRC liver metastasis, we also performed RNA-seq on matched primary tumors and liver metastases biopsied and frozen from two patients. As shown in Fig. 1G, in both cases, the RBMS1 regulon was enriched among transcripts that were downregulated in the metastases relative to their primary tumors. Interestingly, while the gene expression changes between the primary tumors and metastases were not generally correlated ($R = 0.02$, $P = 0.1$), the RBMS1 regulon was independently downregulated in both metastatic tumors (Supplementary Fig. 1E). Consistently, *RBMS1* was also strongly silenced in both metastases (Fig. 1H). Collectively, these findings in cell lines, PDXs, and clinical samples implicate RBMS1 silencing in both the downregulation of its putative regulon and in CRC metastatic progression.

RBMS1 acts as a post-transcriptional regulator of RNA stability:

The correlated expression of *RBMS1* and its putative regulon defined from its binding site preferences implies that RBMS1 modulates gene expression. To further assess the potential role of RBMS1 as a post-transcriptional regulator, we performed RNA sequencing of *RBMS1* knockdown and control SW480 colon cancer cells. We chose the SW480 cell line for this experiment because: (i) *RBMS1* expression in SW480 is among the highest in colon cancer lines tested by us and others, and (ii) it is an established xenograft model of liver metastasis and is considered to be poorly metastatic (relative to other CRC lines listed in Supplementary Fig. 1A). We used two independent shRNAs to silence *RBMS1* in SW480 cells and confirmed its knockdown with both qPCR and Western blotting (Supplementary Fig. 2A). A 2.5-fold reduction in RBMS1 protein expression induced major changes in gene expression (Supplementary Fig. 2B) and resulted in a significant decrease in the expression of RBMS1 regulon (Fig. 2A). Since we are focused on the role of RBMS1 as an RNA-binding protein, we reasoned that its effect on gene expression is likely through regulation of the stability of its target RNAs. To test this hypothesis, we took advantage of REMBRANDTS, a computational framework we have developed to estimate RNA stability from RNA-seq data (10). We have previously established that REMBRANDTS accurately recapitulates experimental RNA stability measurements (10). Application of this method to RNA-seq data from *RBMS1* knockdown and control cells found a significant enrichment of RBMS1 targets among transcripts destabilized upon *RBMS1* silencing (Fig. 2B). REMBRANDTS relies on the comparison of exonic and intronic reads to measure changes in RNA stability. As shown in Supplementary Fig. 2C, *RBMS1* silencing results in lower expression of its putative regulon as measured by exonic reads; however, intronic reads,

which reflect changes in pre-mRNA levels (and transcription rates) do not significantly change in response to RBMS1 depletion. To further strengthen our claim that RBMS1 functions as a post-transcriptional regulator of RNA stability rather than as a transcriptional activator, we also performed whole-genome RNA stability measurements in control and *RBMS1* knockdown cells. For this, we used α-amanitin-mediated inhibition of RNA-polymerase II followed by RNA sequencing at two time points (0-hr and 9-hr). Consistent with our analyses with REMBRANDTS, we observed a similar reduction in the stability of the putative RBMS1 regulon upon *RBMS1* silencing (Fig. 2C). We used RNA-seq data from the three matched PDX models (CLR panel) to compare RNA stability in highly liver metastatic models to that of their parental PDXs (using REMBRANDTS). As shown in Fig. 2D, lower *RBMS1* expression in the LvM models (Fig. 1F) accompanies a reduction in the stability of its target regulon in all three independently derived models. Together, these findings establish RBMS1 as a post-transcriptional regulator of RNA stability with broad functional consequences for the transcriptome and clear implications for CRC progression.

RBMS1 CLIP-seq reveals 3' UTR binding of target mRNAs:

To this point, our analyses were performed using *in silico* predicted binding targets of RBMS1. In order to create a transcriptome-wide snapshot of RBMS1 binding sites in colon cancer cells at nucleotide resolution, we performed UV crosslinking immunoprecipitation followed by sequencing (CLIP-seq). We carried out irCLIP (11) for endogenous RBMS1 in SW480 cells, and identified hundreds of high-confidence RBMS1 binding sites across the transcriptome (Fig. 3A). Importantly, we observed that 90% of the CLIP-identified RBMS1 targets overlapped with our computationally-derived putative RBMS1 regulon (Supplementary Table 1). Consistently, these RBMS1-bound transcripts in our irCLIP data (RBMS1 targets hereafter) follow the same expression patterns as the computationally-derived RBMS1 regulon described above, exhibiting both decreased stability and expression upon *RBMS1* silencing (Fig. 3B; Supplementary Fig. 3A).

In our analysis of RBMS1 binding sites, we noted a strong enrichment of RBMS1 binding to the last exon (on or close to 3' UTRs; Fig. 3A) of mRNAs. This is consistent with RBMS1 acting as a post-transcriptional regulator of RNA stability, and suggested that RBMS1 3' UTR binding could result in increased RNA stability. To assess this possibility, we built a bi-directional CMV promoter that drives the expression of both *GFP* and *mCherry*. We then cloned nine CLIP-seq-derived RBMS1 binding sites, plus ~150 nucleotides flanking the binding sites, downstream of *GFP*, and asked whether there was a reduction in *GFP* mRNA relative to *mCherry* mRNA upon RBMS1 depletion. As shown in Fig. 3C, there was a decrease in *GFP* reporter transcript levels in almost all instances. To further ensure that this effect was RBMS1-dependent, we also tested the two reporters with the strongest reduction in transcript levels (*DAG1* and *CD9*) in LS174T cells, where *RBMS1* is endogenously silenced. In this instance, as expected, there was no response in reporter mRNA levels upon transfection of RBMS1-targeting siRNAs (Supplementary Fig. 3B).

In order to gain insight into how RBMS1 regulates the stability of its targets, we carried out an unbiased search for its interacting protein partners. We performed immunoprecipitation of RBMS1, along with an IgG control, from SW480 cells and analyzed the resulting protein

complexes by mass spectrometry (Fig. 3D). From this data, we found that the Gene Ontology terms RNA binding, RNA processing and negative regulation of mRNA metabolism were overrepresented among the RBMS1-interacting proteins (Supplementary Table 2). Of these proteins, we have confirmed an RNA-dependent interaction between RBMS1 and two well-characterized RBPs that predominantly bind 3' UTRs, PABPC1 and ELAVL1 (Fig. 3E). These findings suggest that RBMS1 stabilizes its targets at least partially in concert with ELAVL1 which preferentially binds A/U-rich elements (AREs) (12). In line with this hypothesis, we found a significant enrichment of poly(U) motifs in the 3' UTRs of RBMS1 targets (Fig. 3F). As shown in Fig. 3G, 72% of RBMS1-bound 3' UTRs are also bound by ELAVL1 *in vivo* (as determined by PAR-CLIP (13)). Furthermore, knockdown of *ELAVL1* in SW620 colon cancer cells (of the same genetic background as SW480 cells) resulted in a significant downregulation of RBMS1 targets, as determined by RNA-seq (14) (Fig. 3H). Together, these results indicate that direct binding of RBMS1 to mRNA 3' UTRs, together with other stabilizing factors such as ELAVL1, results in increased mRNA stability.

RBMS1 is a suppressor of EMT and metastatic liver colonization in colon cancer cells:

In order to carry out a functional study of RBMS1 and its downstream regulon in CRC metastasis, we picked an additional cell line model to complement the SW480 line. We chose the LS174T line because *RBMS1* is almost completely silenced in this line as measured by qPCR and Western blotting (Supplementary Fig. 4A–B). LS174T is also an established xenograft model and is considered a highly liver metastatic cell line, ~100x more metastatic than the SW480 line as measured by liver colonization assays (15). We first performed RNA sequencing of both SW480 and LS174T lines and consistent with *RBMS1* silencing in the LS174T cells, we observed a significant reduction in the expression of the RBMS1 regulon in this line (Supplementary Fig. 4C).

To establish a causal link between *RBMS1* silencing and higher metastatic capacity, we performed liver colonization assays by splenically injecting *RBMS1* knockdown and control SW480 cells and measuring metastatic burden in the livers of mice over time. While *RBMS1* silencing did not have a strong effect on *in vitro* cell proliferation (Supplementary Fig. 4D), we observed a significant increase in liver colonization upon silencing *RBMS1*, based on *in vivo* bioluminescence measurements and gross liver mass at the experimental endpoint (Fig. 4A; Supplementary Fig. 4E). To control for possible off-target effects of the shRNAs, we repeated the experiment with an additional *RBMS1*-targeting hairpin and observed a similarly significant increase in metastatic liver colonization (Supplementary Fig. 4F). We also performed a gain-of-function experiment by expressing *RBMS1* in the LS174T line, where *RBMS1* is endogenously silenced. Consistent with our earlier findings, we observed that expressing *RBMS1* in LS174T cells results in a marked reduction in their liver colonization capacity (Fig. 4B; Supplementary Fig. 4G). It should be noted that as expected, expressing shRNAs targeting *RBMS1* in highly metastatic cells in which *RBMS1* is already silenced, namely LS174T, HCT116, and Colo320, did not have an impact on their liver colonization capacity in xenograft mouse assays, further supporting the on-target effect of the shRNAs used in these experiments (Supplementary Fig. 4H). In contrast, knockdown of *RBMS1* in WiDr cells, in which *RBMS1* is endogenously expressed, resulted in a significant

increase in liver colonization in xenograft mouse models without an overt effect on *in vitro* cell proliferation (Supplementary Fig. 4I–J).

The metastatic cascade is a complex multistep process, involving cellular processes that not only impact cancer cell growth and survival, but also cell migration and invasion. As mentioned earlier, we did not observe a significant change in *in vitro* cell proliferation rates upon RBMS1 knockdown. Additionally, trans-well invasion assays of *RBMS1* knockdown and control cells did not detect a significant role for RBMS1 in cancer cell invasion (Supplementary Fig. 4K). Therefore, to gain a better understanding of the key step(s) in the metastatic cascade that are regulated by RBMS1, we performed a systematic analysis of known and predicted gene-sets to identify those that may be modulated upon *RBMS1* silencing. We performed this analysis using iPAGE, an information-theoretic framework we have developed for this type of analysis (16) (Supplementary Fig. 4L). Downregulation of genes associated with negative regulation of epithelial-mesenchymal transition (EMT(−) signature gene set defined by various studies (17–19)) was the most significant pathway identified in *RBMS1* knockdown cells (Fig. 4C; Supplementary Fig. 4L–M). For example, the canonical EMT marker E-Cadherin (*CDH1*) is significantly downregulated by ~2-fold in *RBMS1* knockdown cells based on RNA-seq data (Supplementary Fig. 2B). We confirmed this by performing immunofluorescence staining for E-Cadherin in RBMS1 knockdown and control cells. We observed both a reduction in E-cadherin expression in *RBMS1* knockdown cells, and also noted the appearance of spindle-like cell morphology that is associated with EMT (Fig. 4D). Moreover, the EMT(−) signature genes were also expressed at relatively lower levels in the LS174T cell line, where *RBMS1* is endogenously silenced (Supplementary Fig. 4M). Finally, analysis of the TCGA-COAD dataset revealed significantly lower E-cadherin expression in tumor samples with low *RBMS1* expression (~20% reduction, P=0.006) and a significant general correlation between *RBMS1* and E-Cadherin expression (Rho=0.1, P=0.03), which is indicative of the clinical relevance of this regulatory axis in CRC.

AKAP12 and SDCBP are functional downstream targets of RBMS1:

In order to identify genes that are regulated by RBMS1 and act as suppressors of metastasis in CRC, we used an integrated analytical approach, which relies on mining relevant datasets to prioritize target genes based on their direct interaction with RBMS1 (3’ UTR binding), their RBMS1-dependent magnitude of change in expression, the robustness of their response to *RBMS1* silencing, as well as their lower expression in metastatic clinical samples. Using this approach, we prioritized two target genes, *AKAP12* and *SDCBP*, that satisfy these criteria: (i) destabilized and downregulated in highly metastatic CRC cells and PDXs, (ii) destabilized and downregulated in *RBMS1* knockdown cell lines, (iii) directly bound by RBMS1 based on our irCLIP data (Fig. 3A), and (iv) downregulated in liver metastases relative to primary colon cancers in a publicly available dataset (20). Consistently, silencing *RBMS1* in SW480 cells resulted in the downregulation and destabilization of *AKAP12* and *SDCBP* mRNAs (Fig. 5A–B). Conversely, overexpression of *RBMS1* in LS174T cells resulted in upregulation and stabilization of these targets (Fig. 5A–B). Additionally, in line with the concerted action of RBMS1 and ELAVL1, both *AKAP12* and *SDCBP* 3’ UTRs are

bound by ELAVL1 *in vivo* (13). These observations establish *AKAP12* and *SDCBP* as direct targets of RBMS1 that are post-transcriptionally regulated by this RBP.

To test the possible role of *AKAP12* and *SDCBP* in driving CRC metastatic progression, we silenced them individually in SW480 cells using CRISPR interference (CRISPRi) (21) and performed liver colonization assays. As shown in Supplementary Fig. 5A, silencing these genes did not have a significant impact on *in vitro* proliferation of SW480 cells. However, reduced expression of these genes resulted in increased metastatic liver colonization in mice (Fig. 5C). To confirm that *AKAP12*, which, of the two targets tested, elicited a stronger phenotype when silenced, indeed functions downstream of RBMS1, we performed an *in vivo* epistasis experiment by generating cells with knockdown of both *AKAP12* and *RBMS1* in SW480 cells. As shown in Supplementary Fig. 5B, unlike in Fig. 5C, further silencing of *AKAP12* failed to increase metastatic liver colonization in *RBMS1* knockdown cells, in which *AKAP12* is already downregulated. To evaluate the contribution of AKAP12 to the phenotype of *RBMS1* silencing, we compared *AKAP12* knockdown and the double knockdown cells. As shown in Supplementary Fig. 5C, we observed a slight additional increase (albeit not statistically significant) in metastatic liver colonization upon *RBMS1* silencing in *AKAP12* knockdown cells. This suggests that AKAP12 is a major effector downstream of RBMS1, and that other downstream targets, such as SDCBP, may also play minor roles in this process. Importantly, and consistent with AKAP12 acting downstream of RBMS1, gene expression profiling of *AKAP12* knockdown and control cells revealed a significant induction of an EMT signature (Fig. 5D). Notably, this included the upregulation of canonical EMT transcriptional repressors, *SNAI1* and *SNAI2* (Supplementary Fig. 5D).

As an orthogonal approach, to assess if the effects of *RBMS1* silencing on the transcriptome are reflected in the proteome, we used TMT labelling and tandem mass spectrometry to compare the global proteomes of SW480 cells with shRNA-mediated RBMS1 knockdown and control cells (Supplementary Fig. 5E, Supplementary Table 2). As shown in Supplementary Fig. 5F, we observed a positive and significant correlation between changes in mRNA abundance (as determined by RNA-seq) and changes in protein abundance. We found that RBMS1 targets, as determined by irCLIP, were enriched among the downregulated proteins in the RBMS1 knockdown cells (Supplementary Fig. 5G). Most notably, AKAP12, which we have identified as the downstream effector of RBMS1 in EMT regulation, was significantly downregulated at the protein level as well, mirroring our transcriptomic results (Supplementary Fig. 5H). Taken together, our findings demonstrate that the RBMS1-AKAP12 regulatory axis acts as a suppressor of EMT and liver metastasis in models of CRC progression.

RBMS1 silencing and the downregulation of its targets is associated with CRC progression:

To further assess the clinical relevance of this previously unknown regulatory pathway, we performed a variety of measurements in clinical samples as well as analysis of clinical data to evaluate RBMS1 activity in CRC metastasis. First, we performed qRT-PCR for *RBMS1* in two independent clinical cohorts, one cohort stratifying patients based on their tumor stage (n=96), and another comparing samples from normal mucosa, primary CRC tumors, and

liver metastases (n=91). As shown in Fig. 6A and 6B, we observed a significant reduction in *RBMS1* expression as the disease progresses, with stage IV clinical samples showing the lowest *RBMS1* expression levels. We also carried out survival analysis in a large clinical cohort with publicly available expression data and clinical outcomes (22). We observed a significant association between *RBMS1* silencing and reduced relapse-free survival as well as overall survival in colon cancer patients (Fig. 6C). A multivariate Cox proportional-hazards model revealed that this association with *RBMS1* silencing remains statistically significant even when controlled for other known clinical factors that may contribute to patient survival (Supplementary Fig. 6A). This observation emphasizes the relevance of *RBMS1* silencing as an effective read-out for risk stratification of patients based on samples collected from their primary tumors.

Our results indicate that AKAP12 acts as a suppressor of metastasis downstream of RBMS1, and therefore is expected to show a similar association with metastatic disease. To test this hypothesis, we performed qRT-PCR measurements in the clinical cohorts mentioned above, and we observed a significant reduction in *AKAP12* expression as a function of disease progression (Fig. 6D–E). Consistent with RBMS1 acting as a regulator of *AKAP12*, we observed a highly positive and significant correlation between the expression of these two genes (Supplementary Fig. 6B). Interestingly, our analyses indicate that the identified RBMS1 targets provide a robust gene signature that, similar to RBMS1, is highly informative of clinical outcomes. For this analysis, we defined an RBMS1 target score as an aggregate measure of expression for the RBMS1 80-gene signature set (average normalized expression of target mRNAs that (i) interact *in vivo* with RBMS1 on or within 200 nucleotides of their 3' UTRs, and (ii) are downregulated upon *RBMS1* knockdown; this is referred to as the RBMS1 80-gene signature hereafter (Supplementary Table 1)). We then stratified patients based on this score and we observed that lower expression of the RBMS1 80-gene signature is associated with lower relapse-free and overall survival in colon cancer patients (Fig. 6F). Consistently, the RBMS1 signature score remained a significant covariate in a multivariate Cox proportional-hazards model of relapse-free survival (Supplementary Fig. 6C). Moreover, we have not observed significant association between the RBMS1 signature score and validated clinico-pathological features or other molecular markers (such as microsatellite instability (MSI) status, left- or right-sidedness, or CRC consensus molecular subtypes (CMS)). Importantly, similar to our initial observation with the RBMS1 regulon (Fig. 1D), the RBMS1 signature score was significantly correlated with *RBMS1* expression in a large TCGA pan-cancer dataset (Fig. 6G). We also observed a high correlation between the RBMS1 signature score and *ELAVL1* expression (Fig. 6H), consistent with RBMS1 and ELAVL1 acting in concert to promote mRNA stabilization. The association between the RBMS1 regulon and CRC metastasis is not limited to a single dataset. Consistent with our initial observations in matched samples (Fig. 1G), comparison of gene expression changes in liver metastases relative to CRC primary tumors in a publicly available dataset (23), revealed a significant reduction in the expression of the RBMS1 regulon (Supplementary Fig. 6D–E). Furthermore, analysis of publicly available RNA-seq data using REMBRANDTS to infer changes in mRNA stability from a set of matched primary and metastatic samples (24), also confirmed (i) *RBMS1* silencing in multiple liver metastases and (ii) a concordant reduction in the stability of the RBMS1 regulon

(Supplementary Fig. 6F). Together, these results establish the clinical relevance of our findings and the importance of *RBMS1* silencing both as a prognostic factor and as a suppressor of liver metastasis in patients.

HDAC1-mediated promoter deacetylation leads to RBMS1 silencing in LS174T cells:

Our findings described above establish a previously unknown regulatory pathway driven by the RNA-binding protein RBMS1 that plays a functional role in CRC metastasis to liver. Regulatory pathways can be expanded by uncovering the upstream pathways that influence their activity. We have developed computational and data analytical tools designed to integrate publicly available datasets from a variety of biological sources to identify such upstream regulatory mechanisms (16). Using these tools, we sought to extend the RBMS1 regulatory pathway by exploring the mechanisms for *RBMS1* silencing observed in the highly metastatic LS174T colon cancer cells. First, we found *RBMS1* expression to be strongly associated with promoter acetylation, but did not identify any specific transcription factors associated with *RBMS1* expression (Fig. 7A). Consistent with this observation, analysis of the connectivity map dataset (25), which reports the impact of hundreds of small molecule treatments on gene expression, identified the histone deacetylase inhibitor Trichostatin A (TSA) as an activator of *RBMS1* expression and of the RBMS1 regulon (Supplementary Fig. 7A–B). As *RBMS1* is endogenously silenced in LS174T cells and expressed in the SW480 line, we tested the relative acetylation levels of the *RBMS1* promoter in these two lines by performing H3K27Ac ChIP-qPCR. This data showed relatively low levels of H3K27 acetylation at the *RBMS1* promoter in the LS174T line compared to the SW480 line, consistent with the differences in the level of *RBMS1* in the two lines (Fig. 7B). TSA simultaneously inhibits multiple HDACs, however, we noted that among this group, *HDAC1* is the most significantly upregulated in highly metastatic cells generally, and in LS174T cells compared to SW480 cells in particular (Fig. 7C–D; Supplementary Fig. 7C). Moreover, we observed that *RBMS1* and *HDAC1* expression are negatively correlated in colon cancer clinical samples (Supplementary Fig. 7D). Furthermore, we observed that silencing *HDAC1* increases *RBMS1* expression (Fig. 7E; Supplementary Fig. 7E–F). Increased expression of *HDAC1*, and other class I HDACs, has been reported as a strong predictor of survival in colon cancer patients (26) and modulations in HDAC activity result in widespread gene expression reprogramming, including repression of tumor suppressor genes (27). Our observations here indicate that HDAC1-mediated transcriptional repression may be one possible mechanism for *RBMS1* silencing in colon cancer cells.

Discussion

Complex human diseases, including cancer, often accompany broad reprogramming of gene expression. In these cases, a comprehensive understanding of the disease state requires not only the identification of the differentially expressed genes, but also understanding the underlying regulatory pathways that explain the dysregulated expression patterns. A number of approaches have been developed by us (16,28) and others (6,29) to tackle this problem. These methods formalize the association between expression and/or activity of master regulators with those of their regulons. Since direct and indirect associations are challenging

to disentangle, knowledge of binding preference (*in vitro* or *in vivo*) is also used to better define putative RNA-RBP interactions. Here, we have introduced PRADA to facilitate and formalize the discovery of post-transcriptional regulators that are involved in normal cell physiology and disease. RNA-binding proteins fall into families with highly similar binding preferences (7) and therefore have similar putative regulons. As a result, direct modeling of RBP-target interactions results in unstable predictions (i.e. the model fails to converge on similar outcomes for repeated runs) where the link between a given RBP and its putative regulon is not clear. PRADA solves this problem by integrating the knowledge of changes in the expression of RBPs (used as a proxy for their activities) directly into the analysis. This approach provides a one-step and often stable solution that effectively reveals the RNA-binding proteins whose differential expression is highly informative of changes in gene expression. As RBPs continue to emerge as key regulators of RNA dynamics with crucial roles in human disease, systematic methods like PRADA provide a suitable approach for studying this class of regulators. In this study, we used this approach to discover a previously unknown regulatory network that counteracts colon cancer progression through RBMS1-mediated RNA stabilization.

We demonstrate that *RBMS1* silencing accompanies CRC progression in cell line and PDX models as well as clinical samples, resulting in destabilization of RBMS1-bound transcripts (Fig. 7F). In line with this, we found several well-known RNA-stabilizing factors interacting with RBMS1. Importantly, we found that RBMS1-bound transcripts contain ELAVL1 motifs and binding sites in their 3’ UTRs, and that *ELAVL1* downregulation resulted in the decrease of RBMS1 target abundance, consistent with a model where RBMS1 and ELAVL1 have an RNA-dependent functional interaction that promotes mRNA stability. Furthermore, we observed that *ELAVL1* expression is highly corelated with the expression of the RBMS1 80-gene signature set in the TCGA pan-cancer dataset. ELAVL1 has a well-documented role in stabilizing mRNA by a mechanism that is incompletely characterized. It is suggested that ELAVL1 competes for 3’ UTR ARE binding with RNA destabilizing factors, such as TTP or BFR1. It is also proposed that ELAVL1 acts in counteracting microRNA-mediated decay and translational shutdown, thereby indirectly stabilizing mRNAs (12). *ELAVL1* upregulation has been broadly associated with oncogenesis and cancer progression (30), suggesting a functional interplay between RBMS1 and ELAVL1.

Metastasis often hijacks developmental pathways to reprogram gene expression. EMT is such a pathway with a central role in cancer progression, including CRC (31). Our results show that *RBMS1* downregulation results in the suppression of negative EMT regulators, thereby promoting the mesenchymal features of the tumor cells. One of the direct targets of RBMS1, AKAP12, participates in the protein kinase A and C signaling cascades. The association between *AKAP12* silencing and colon cancer recurrence and prognosis has been previously described (32). Interestingly, AKAP12 has been described as a negative regulator of SNAI1 (33), a transcription factor that represses epithelial markers during EMT. In line with this, we found that *AKAP12* knockdown caused a global shift towards an EMT transcriptional program, including the upregulation of *SNAI1/2*. One of the main targets of SNAI1 is the epithelial marker E-cadherin, which is in agreement with our observation that E-cadherin is downregulated in RBMS1-deficient cells and clinical samples. This suggests

that the RBMS1-AKAP12-SNAI1/2-E-cadherin axis is a potential route to EMT onset during CRC metastasis.

The Wnt/β-catenin signaling pathway has also been implicated in both EMT and CRC metastasis. β-catenin, a central factor in the canonical Wnt signaling cascade, has a dynamic role: it is known to interact with E-cadherin at the cell membrane, with TCF/LEF transcription factors in the nucleus to activate the downstream transcriptional program, and with APC, a member of the destruction complex, that targets β-catenin for proteasomal degradation in the absence of pathway activation (34). *APC* is a known mutational hotspot in CRC, and the SW480 cells used in this study harbor an inactivating *APC* mutation, indicating that the Wnt pathway is misregulated in this cell line (35). We have shown that *RBMS1* silencing results in the downregulation of E-cadherin and have detected upregulated β-catenin levels in *RBMS1* knockdown cells (Supplementary Fig. 7G). Interestingly, one of the direct RBMS1 targets validated in this study is syndecan binding protein (*SDCBP*), which interacts both with Frizzled family proteins (canonical Wnt receptors) and syndecans (Wnt Co-receptors) (36). As our results suggest multiple crossover points, future studies will be required to elucidate the role of RBMS1 in Wnt signaling cascades and their promotion of EMT.

Mirroring our observations from the CRC TCGA dataset, we found the expression of *RBMS1* and its signature gene set were highly corelated in the TCGA pan-cancer dataset. However, when examined individually, *RBMS1* and its signature gene set expression was informative for specific tumor types. *RBMS1* silencing was associated with poor outcome in the CRC cohort, as discussed above (Fig. 6F); in contrast, different outcomes were observed in other cancer types, ex. in the thyroid cancer cohort, patients with tumors with relatively low expression of *RBMS1* had better outcomes (Supplementary Fig. 7H–I). We observed a similar trend in other pan-cancer cohorts, notably in a dataset of metastatic tumors (MET500) (37). While *RBMS1* and its signature expression remained highly corelated, we found reduced *RBMS1* levels specifically in colorectal, prostate and gall bladder cancer metastases. Interestingly, when stratified by the metastasis site, *RBMS1* expression was significantly downregulated in liver and bone marrow (independently of the primary tumor type), reflecting the principal organotropisms of metastatic CRC (Supplementary Fig. 7J–K). It is thus highly plausible that RBMS1 represents a regulatory node particularly suited for disease progression in the liver, potentially via activation of downstream signaling pathways as discussed above. This finding has implications for other cancer types that can metastasize to the liver.

Our work in the LS174T cell line model indicated that increased activity of histone deacetylase(s), such as HDAC1, could be responsible for *RBMS1* silencing. HDAC inhibitors have been used in the clinic to treat metastatic CRC. A number of studies have previously noted increased expression of *HDAC1/2* in colon cancer (27) and in cases where *RBMS1* silencing also occurs, HDAC inhibitors could be useful as they may also restore *RBMS1* expression. While HDAC inhibitors are not specific in their effects, and their effects are not solely mediated through RBMS1, we speculate that in some cases their impact on RBMS1 may lead to better clinical outcomes. Identifying such cases, however, will require a

deeper understanding of the mechanisms through which *RBMS1* is silenced and the degree to which its expression can be effectively controlled.

Given our limited knowledge of the pathways and processes that regulate the RNA life-cycle in the cell, analytical tools that mine quantitative measurements of mRNA dynamics to identify key regulatory interactions can provide an effective avenue for identifying previously unknown molecular mechanisms with critical functions in health and disease. However, these computational strategies must be paired with rigorous experimentation to functionally validate and characterize the putative regulons and their regulators. Using one such framework, we have established a novel functional association between RBMS1 silencing and increased liver metastatic capacity in colon cancer.

Methods

Prioritization of Regulatory Pathways based on Analysis of RNA Dynamics Alterations (PRADA).

PRADA is a customized variation on lasso regression (least absolute shrinkage and selection operator). Therefore, PRADA can simultaneously perform feature selection and apply a custom penalty function as part of its regularization step. The goal of PRADA is to identify RNA-binding proteins whose differential expression explains changes in the expression of their targets that is observed in the data. We first generated an RBP-RNA interaction matrix based on binding preferences of RNA-binding proteins, as previously reported (7,38). For this, we scanned mRNA sequences for matches to RBP-derived regular expression (as previously described (6)). Given this interaction matrix, our goal is to identify RBPs whose change in expression is predictive of global changes in gene expression of their targets. In other words, we aim to solve the following: $\Delta Exp(g) = \sum_i \alpha_i \cdot t_{i,g} \cdot \Delta Exp(RBP_i) + c_g$, where $\Delta Exp(g)$ stands for changes in the expression of gene g based on dataset of interest, $t_{i,g}$ is a binary variable, 1 if RBP_i is predicted to bind g and 0 otherwise, $\Delta Exp(RBP_i)$ marks changes in the expression of RBPs themselves as a proxy for their differential activity. The resulting α_i represents the strength of regulatory interactions. To ensure that the most informative RBPs are selected, we introduced a customized penalty term that extends the lasso regression framework: $\min \frac{1}{2n} \left[\| \alpha X - Exp \| + \lambda \sum_i \frac{|\alpha_i|}{|\Delta Exp(RBP_i)|} \right]$. Here, the coefficients are penalized by $|\Delta Exp(RBP_i)|^{-1}$, which is $1/\left|\hat{\beta}\right|$ estimate of the linear model used to compare gene expression between the two groups. Other variations of this penalty can also be used, for example standard effect size can account for confidence in $\hat{\beta}$ as well $\left(SE/\left|\hat{\beta}\right|\right)$. This custom penalty term ensures that RBPs whose activity does not change are not selected by the model. This also stabilizes the resulting model, which would otherwise be a major issue as RBPs that belong to the same family often have very similar binding preferences resulting in correlated features in the interaction matrix. After running this regression analysis, the RBPs with the largest assigned coefficients are prioritized for further study. In this study, we used PRADA to compare poorly and highly metastatic colon cancer lines, which revealed RBMS1 as the strongest candidate. The computational and experimental data flow used in this study is summarized in Supplementary Material 1. The PRADA benchmarking

summary is available in Supplementary Material 2. All code and notebooks are available on Github at www.github.com/goodarzilab/PRADA.

Tissue Culture

SW480, LS174T, WiDr, HCT116, and COLO320 cell lines were obtained from ATCC. 293LTV cells were from Cell Biolabs. All cell lines were routinely screened for Mycoplasma infection by PCR (once a month in general and before every animal experiment in particular). SW480 and HCT116 cells were cultured in McCoy's 5A supplemented with 10% FBS, sodium pyruvate, and L-glutamine. LS174T and 293LTV cells were cultured in DMEM supplemented with 10% FBS, COLO320 cells were cultured in RPMI-1640 supplemented with 10% FBS, and WiDr cells were grown in EMEM supplemented with 10% FBS. All commonly used cell lines were routinely authenticated using STR profiling at UC Berkeley Sequencing Facility.

Animal studies

All animal studies were performed according to a protocol approved by UCSF Institutional Animal Care and Use Committee (AN179718). NOD/SCID gamma male mice (The Jackson Laboratory), aged 8 to 10 weeks, were used in all experiments. For splenic (portal circulation) injections, cells were injected directly into the spleen followed by splenectomy (250k for SW480 and LS174T lines and 100k for WiDr cells). *In vivo* bioluminescence was measured by retro-orbital injection of luciferin (Perkin Elmer) followed by imaging with an IVIS instrument (Perkin Elmer). For *ex vivo* liver imaging, mice were injected with luciferin prior to liver extraction, and the liver was then imaged and weighed after rinsing with PBS.

RBMS1 irCLIP

RBMS1 irCLIP was performed as described (11).

RNA-seq library preparation

Unless otherwise specified below, RNA sequencing libraries were prepared using RNA that had been rRNA depleted using Ribo-Zero Gold (Illumina) followed by ScriptSeq-v2 (Illumina), and sequenced on an Illumina HiSeq4000 at UCSF Center for Advance Technologies. Matched primary and liver metastases were sequenced using SENSE RNA-seq library preparation kit (Lexogen). AKAP12 knockdown cells were profiled using QuantSeq 3' mRNA-Seq library prep kit fwd (Lexogen).

To measure RNA stability, SW480 *RBMS1* knockdown and control cells were treated with 10 mg/mL α-amanitin (final concentration in the medium). After 9 hours, total RNA was harvested from the cells using the Norgen Cytoplasmic and Nuclear RNA Purification Kit per the manufacturer's protocol. RNA-seq libraries were then prepared and log-fold changes in RNA stability were measured by comparing log(t9/t0) in control and *RBMS1* knockdown cells.

Measurements in clinical samples.

For clinical samples, 96 samples across all stages of the disease were acquired from Origene (HCRT104 and HCRT105), 14 normal, 8 stage I, 25 stage II, 32 stage III, and 17 stage IV.

HPRT was used as endogenous control, and relative *RBMS1* and *AKAP12* expression levels were respectively measured across the samples (using the primers listed above). We also extracted RNA from another 100 normal mucosa, primary tumors, and liver metastases. Roughly 90 of these samples yielded sufficient RNA for qPCR, which was carried out as previously described.

Clinical association studies.

Patients profiled and documented in GSE39582 were first stratified into two groups based on their *RBMS1* expression: silenced (bottom 5%) and expressed (otherwise). 5% was selected as the cut-off because it is close to two standard deviations away from the average *RBMS1* expression across all samples. A multivariate Cox Proportional Hazard model (R package survival) was then used to evaluate 5-year disease-free survival. Univariate survival analyses, both disease-free and overall, were also carried out with this stratification. For the target regulon, the RBMS1 signature score was calculated as the median expression of target genes in each sample (normalized gene expression data). Disease-free survival was defined as the time from diagnosis to relapse or death due to any cause. We performed Cox Proportional Hazards multivariable modeling using the RBMS1 score as a continuous variable. Variables added to the model included stage, microsatellite status, *KRAS* and *BRAF* mutation status, and Consensus Molecular Subtypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the UCSF Center for Advanced Technology (CAT) for high throughput sequencing and other genomic analyses. We thank Byron Hann and the Preclinical Therapeutics core as well as the Laboratory Animal Resource Center (LARC) at UCSF. We are also grateful for the genomic data contributed by the TCGA Research Network, including donors and researchers. We acknowledge support from our colleagues at the Helen Diller Family Comprehensive Cancer Center and Rockefeller University. We specifically acknowledge Dr. Sohail Tavazoie for providing access to data from PDX models and for reading an earlier version of this manuscript. We also acknowledge Jonathan Weissman and Luke Gilbert for CRISPRi constructs. We acknowledge Dr. David Erle for providing the backbone construct used for the reporter assay. This work was supported by grants to H.G. (NIH: R00CA194077 and R01CA24098; 2017 AACR-Takeda Oncology NextGen Grant for Transformative Cancer Research: 17-20-38-GOOD). J.Y. was supported by an NSF Graduate Research Fellowship, A.N. was supported by a DoD PRCRP Horizon Award (DoD: W81XWH-19-1-0594), H.A. and L.F. were supported by NIH training grants, F32GM133118 and T32CA108462-15 respectively. M.D., M.D. and F.K.M. were supported by MRC Career Development Award (MR/P009417/1).

Financial support: This work was supported by grants to H.G. (NIH: R00CA194077 and R01CA24098; 2017 AACR-Takeda Oncology NextGen Grant for Transformative Cancer Research: 17-20-38-GOOD). J.Y. was supported by an NSF Graduate Research Fellowship, A.N. was supported by a DoD PRCRP Horizon Award (DoD: W81XWH-19-1-0594), H.A. and L.F. were supported by NIH training grants, F32GM133118 and T32CA108462-15 respectively. M.D., M.D. and F.K.M. were supported by MRC Career Development Award (MR/P009417/1).

References

1. Goodarzi H, Zhang S, Buss CG, Fish L, Tavazoie S, Tavazoie SF. Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. Nature. 2014;513:256–60. [PubMed: 25043050]

2. Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell. 2015;161:790–802. [PubMed: 25957686]

3. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell. 2016;165:1416–27. [PubMed: 27259150]

4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7–30. [PubMed: 29313949]

5. Hamada K, Monnai M, Kawai K, Nishime C, Kito C, Miyazaki N, et al. Liver metastasis models of colon cancer for evaluation of drug efficacy using NOD/Shi-scid IL2Rgammanull (NOG) mice. Int J Oncol. 2008;32:153–9. [PubMed: 18097554]

6. Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all Genomes and data types. Mol Cell. 2007;28:337–50. [PubMed: 17964271]

7. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013;499:172–7. [PubMed: 23846655]

8. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol. 2015;33:831–8. [PubMed: 26213851]

9. Yamaguchi N, Weinberg EM, Nguyen A, Liberti MV, Goodarzi H, Janjigian YY, et al. PCK1 and DHODH drive colorectal cancer liver metastatic colonization and hypoxic growth by promoting nucleotide synthesis. eLife. 2019;8.

10. Alkallas R, Fish L, Goodarzi H, Najafabadi HS. Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. Nat Commun. 2017;8:909. [PubMed: 29030541]

11. Zarnegar BJ, Flynn RA, Shen Y, Do BT, Chang HY, Khavari PA. irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods. 2016;13:489–92. [PubMed: 27111506]

12. Hinman MN, Lou H. Diverse molecular functions of Hu proteins. Cell Mol Life Sci. 2008;65:3168–81. [PubMed: 18581050]

13. Lebedeva S, Jens M, Theil K, Schwanhäusser B, Selbach M, Landthaler M, et al. Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. Mol Cell. 2011;43:340–52. [PubMed: 21723171]

14. Fischl H, Neve J, Wang Z, Patel R, Louey A, Tian B, et al. hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. Nucleic Acids Res. 2019;47:7580–91. [PubMed: 31147722]

15. Loo JM, Scherl A, Nguyen A, Man FY, Weinberg E, Zeng Z, et al. Extracellular metabolic energetics can promote cancer progression. Cell. 2015;160:393–406. [PubMed: 25601461]

16. Goodarzi H, Elemento O, Tavazoie S. Revealing global regulatory perturbations across human cancers. Mol Cell. 2009;36:900–11. [PubMed: 20005852]

17. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci. 2010;107:15449–54. [PubMed: 20713713]

18. Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA. Loss of E-Cadherin Promotes Metastasis via Multiple Downstream Transcriptional Pathways. Cancer Res. 2008;68:3645–54. [PubMed: 18483246]

19. Jechlinger M, Grunert S, Tamir IH, Janda E, Lüdemann S, Waerner T, et al. Expression profiling of epithelial plasticity in tumor progression. Oncogene. 2003;22:7155–69. [PubMed: 14562044]

20. Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A. 2009;106:7131–6. [PubMed: 19359472]

21. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. eLife. 2016;5.

22. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10:e1001453. [PubMed: 23700391]
23. Del Rio M, Molina F, Bascoul-Mollevi C, Copois V, Bibeau F, Chalbos P, et al. Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. J Clin Oncol Off J Am Soc Clin Oncol. 2007;25:773–80.
24. Kim S-K, Kim S-Y, Kim J-H, Roh SA, Cho D-H, Kim YS, et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. Mol Oncol. 2014;8:1653–66. [PubMed: 25049118]
25. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. Science. 2006;313:1929–35. [PubMed: 17008526]
26. Weichert W, Röske A, Niesporek S, Noske A, Buckendahl A-C, Dietel M, et al. Class I histone deacetylase expression has independent prognostic impact in human colorectal cancer: specific role of class I histone deacetylases in vitro and in vivo. Clin Cancer Res Off J Am Assoc Cancer Res. 2008;14:1669–77.
27. Li Y, Seto E. HDACs and HDAC Inhibitors in Cancer Development and Therapy. Cold Spring Harb Perspect Med. 2016;6:a026831. [PubMed: 27599530]
28. Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature. 2012;485:264–8. [PubMed: 22495308]
29. Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst Biol. 2010;6:377. [PubMed: 20531406]
30. Schultz CW, Preet R, Dhir T, Dixon DA, Brody JR. Understanding and targeting the disease-related RNA binding protein human antigen R (HuR). WIREs RNA. 2020;11.
31. Vu T, Datta P. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. Cancers. 2017;9:171.
32. Xu G, Zhang M, Zhu H, Xu J. A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. Gene. 2017;604:33–40. [PubMed: 27998790]
33. Cha J-H, Wee H-J, Seo JH, Ahn BJ, Park J-H, Yang J-M, et al. Prompt meningeal reconstruction mediated by oxygen-sensitive AKAP12 scaffolding protein after central nervous system injury. Nat Commun. 2014;5:4952. [PubMed: 25229625]
34. Schatoff EM, Leach BI, Dow LE. WNT Signaling and Colorectal Cancer. Curr Colorectal Cancer Rep. 2017;13:101–10. [PubMed: 28413363]
35. El-Bahrawy M, Poulsom SR, Rowan AJ, Tomlinson IT, Alison MR. Characterization of the E-cadherin/catenin complex in colorectal carcinoma cell lines: E-cadherin/catenin complex and colorectal carcinoma. Int J Exp Pathol. 2004;85:65–74. [PubMed: 15154912]
36. Wawrzak D, Luyten A, Lambaerts K, Zimmermann P. Frizzled–PDZ scaffold interactions in the control of Wnt signaling. Adv Enzyme Regul. 2009;49:98–106. [PubMed: 19534027]
37. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548:297–303. [PubMed: 28783718]
38. Oikonomou P, Goodarzi H, Tavazoie S. Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts. Cell Rep. 2014;7:281–92. [PubMed: 24656821]
39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. [PubMed: 25516281]

Significance

By applying a new analytical approach to transcriptomic data from clinical samples and models of colon cancer progression, we have identified RBMS1 as a suppressor of metastasis and as a post-transcriptional regulator of RNA stability. Notably, *RBMS1* silencing and downregulation of its targets are negatively associated with patient survival.

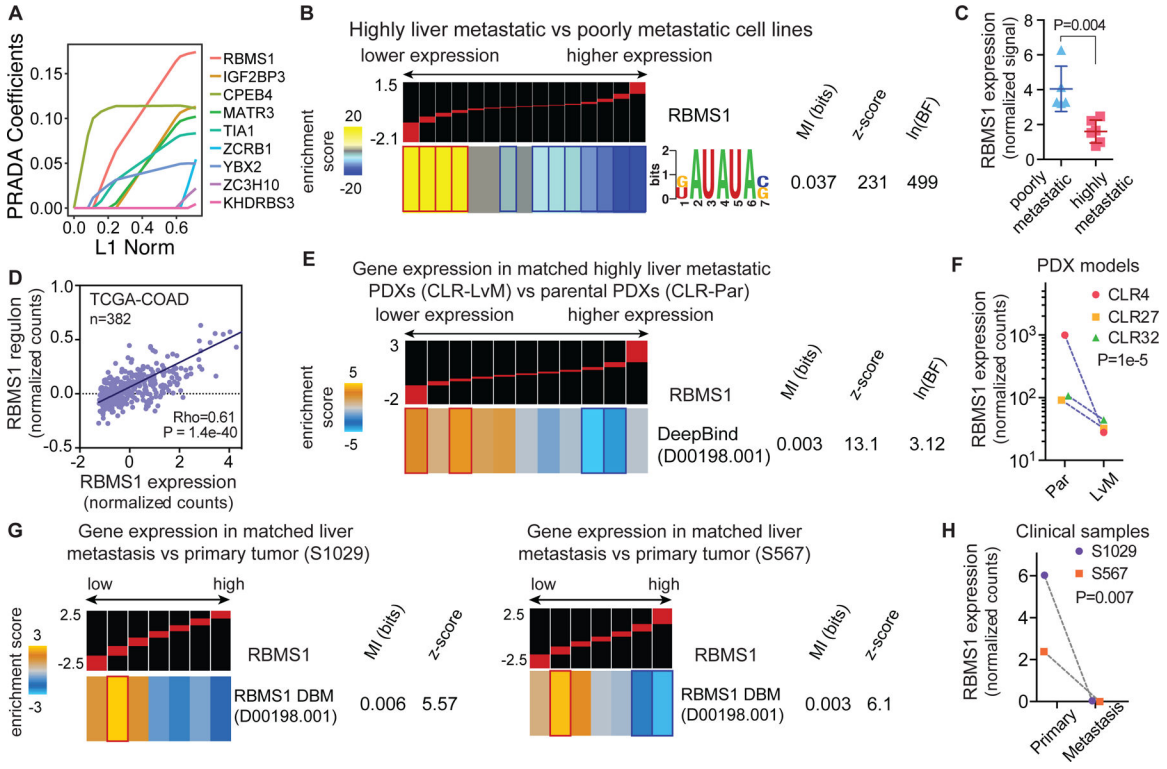

Figure 1. *RBMS1* silencing in metastatic cells is associated with lower expression of RBMS1 targets.

(A) Regression coefficients set by PRADA as a function of the *l1*-norm of the coefficient vector. Each line is associated with an RNA-binding protein and the magnitude of its coefficient is a measure of its strength as a putative regulator of gene expression. Here, the first ten non-zero coefficients are shown as a function of *l1*-norm (i.e. sum of the magnitude of all coefficients). **(B)** Analysis of RBMS1 recognition sites across gene expression changes between poorly and highly metastatic colon cancer lines. In this analysis, transcripts are first ordered based on their log-fold changes from left (lower expression in metastatic cells) to right (higher expression) and then partitioned into equally populated bins (~1000 transcripts per expression bin). The red bars on the black background show the range of values in each bin (with the minimum and maximum values, i.e. −1.5 and 2, presented on the left). As shown here, genes that are expressed at a lower level in highly metastatic cells were significantly enriched for the RBMS1 binding motif (KAUAUAS) (38). In this heatmap, gold represents overrepresentation of putative RBMS1 targets while blue indicates underrepresentation. Enrichment and depletions that are statistically significant (based on hypergeometric distribution) are marked with red and dark blue borders, respectively. Also included are the logo representation of the RBMS1 binding motif, its mutual information (MI) value, the associated z-score and the Bayes factor (BF) (for details of this analysis see (28)). **(C)** *RBMS1* expression in colon cancer lines grouped based on their metastatic capacity (Supplementary Fig. 1A). *P*-value calculated using a two-tailed Mann-Whitney *U*-test. **(D)** Linear regression analysis of *RBMS1* expression versus average normalized expression of its putative regulon in TCGA-COAD dataset (cbioportal; N = 382). Shown are the Spearman correlation coefficient and the associated *p*-value. **(E)** Enrichment and

depletion patterns of the RBMS1 regulon in PDX models of CRC liver metastasis. For this analysis, log-fold changes between parental (CLR-Par) and liver metastatic (CLR-LvM) were averaged across three independent PDX models, CLR4, CLR27, and CLR32. D00198.001 is the unique identifier containing the binding site information of human RBMS1 obtained from DeepBind (8) analysis. The distribution of RBMS1 targets was then assessed using mutual information, its associated *z*-score and Bayes factor. The enrichment and depletion patterns were visualized as described in (B). **(F)** Relative RBMS1 levels in matched poorly metastatic (Par) and highly liver metastatic (LvM) derivatives for three PDX models (CLR4, CLR32, and CLR27). *P*-value was calculated using DESeq2 (39). **(G-H)** Expression of *RBMS1* and RBMS1 targets in matched primary tumor and liver metastases from two patients (S1029 and S567). The results are presented as described in (B). The *P*-value for *RBMS1* silencing was calculated using DESeq2 (controlled for genetic background).

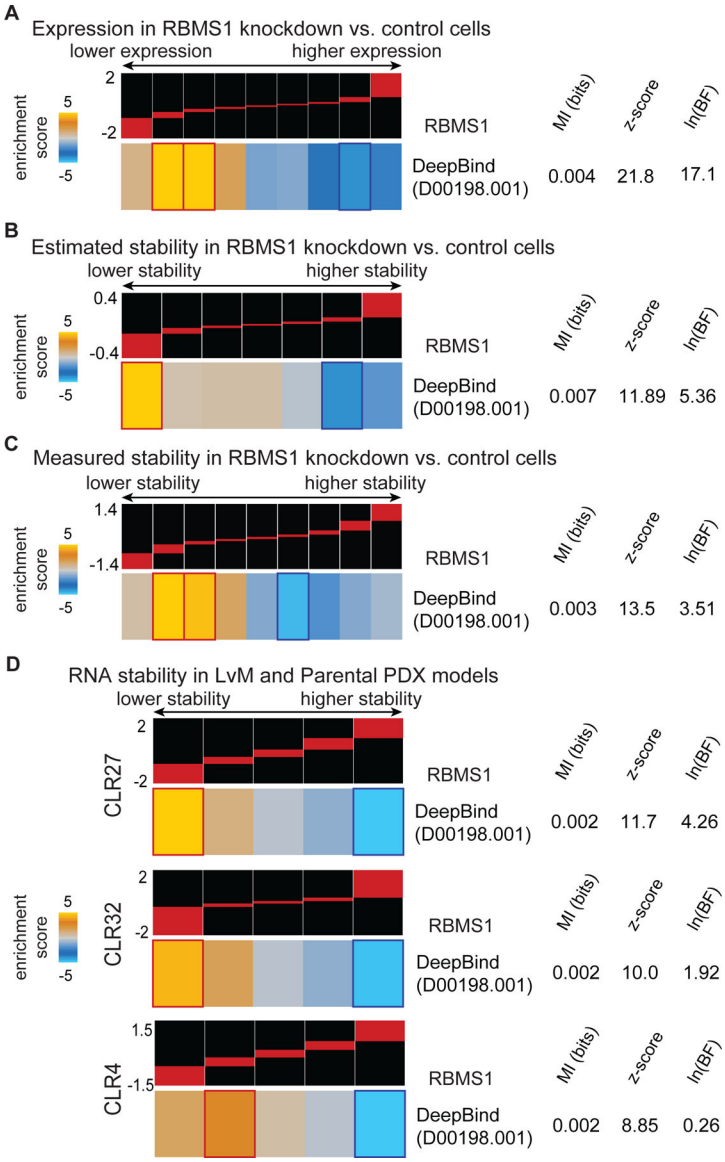

Figure 2. RBMS1 post-transcriptionally regulates the stability and expression of its targets.

(A) Enrichment and depletion patterns of the RBMS1 regulon in *RBMS1* knockdown cells relative to control (~2.5-fold knockdown). **(B)** We used our computational tool, called REMBRANDTS, to estimate changes in RNA stability upon *RBMS1* silencing. These differential stability estimates were then used to assess the enrichment patterns of the RBMS1 targets across the changes in RNA decay. **(C)** Experimental RNA stability changes were measured using α-amanitin treatment as previously described (1). The enrichment and depletion patterns of the RBMS1 regulon was then assessed among the transcripts destabilized or stabilized upon *RBMS1* knockdown. **(D)** We used REMBRANDTS to measure changes in RNA stability between poorly and highly metastatic PDX models from three independent PDX models (CLR27, CLR32, and CLR4). As shown here, consistent with the silencing of *RBMS1* in LvM PDX models (Fig. 1F) and the down-regulation of its

regulon (Fig. 1E), the RBMS1 regulon is destabilized in these three independent models of CRC metastasis.

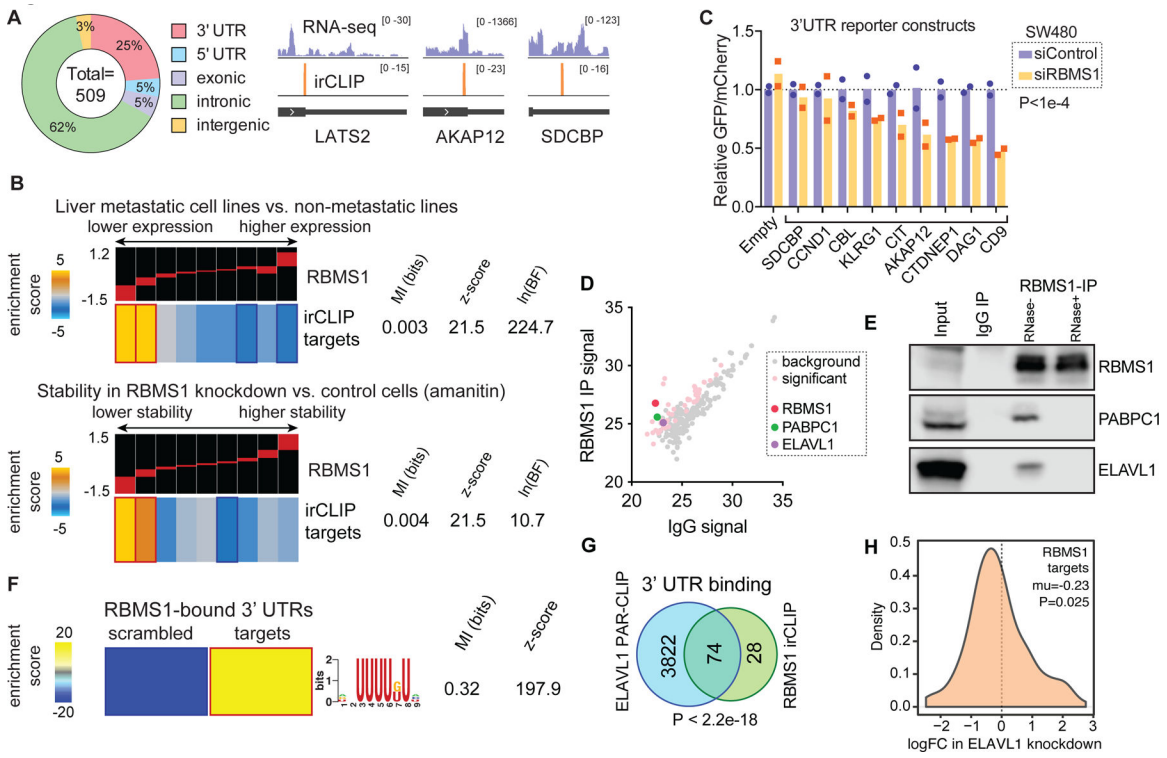

Figure 3. RBMS1 irCLIP identifies direct RBMS1 targets in colon cancer cells.

(A) 509 RBMS1 binding sites were found using irCLIP, with a significant enrichment of binding to the last exon/3' UTR (relative to the total length of genomic features). Last exons from *LATS2*, *AKAP12*, and *SDCBP* are shown as examples of RBMS1 binding patterns. **(B)** Enrichment of the RBMS1-bound mRNAs among those that are downregulated in highly metastatic cells (top) and those destabilized upon *RBMS1* silencing (bottom). **(C)** qRT-PCR was used to measure changes in *GFP* mRNA levels upon cloning RBMS1 binding sites of the listed genes downstream of the *GFP* ORF. *mCherry* was expressed from the same bidirectional promoter as *GFP*, and *mCherry* levels were used to normalize *GFP* measurements. A one-sample Wilcoxon signed rank test was used to assess whether the ratios in siRBMS1 samples were significantly below 1.0. **(D)** Scatter plot of mass spectrometry data showing proteins that co-immunoprecipitate with RBMS1 versus control IgG in SW480 cells. Shown are the average of three replicates across all detected proteins. Proteins enriched in the RBMS1 co-IP samples are shown in pink. RBMS1, PABPC1 and ELAVL1 are highlighted in red, green and violet, respectively. **(E)** RBMS1, PABPC1 and ELAVL1 were detected by western blot in input and eluate samples from RBMS1 and IgG immunoprecipitations (SW480 cell lysates). For RBMS1 immunoprecipitation, the lysates were additionally treated by RNaseA. **(F)** Heatmap showing the enrichment of poly(U) sites in the 3' UTRs of RBMS1-bound mRNAs. **(G)** Venn diagram showing the overlap between RBMS1- and ELAVL1-bound mRNA 3' UTRs. ELAVL1 targets were determined by PAR-CLIP (13). *P*-value calculated using hypergeometric test. **(H)** Density plot showing the log fold-change in RBMS1 target expression (determined by RNA-seq (14)) upon *ELAVL1* knockdown in SW620 cells. Median value (*mu*) is indicated. *P*-value calculated using one-tailed Wilcoxon signed rank test.

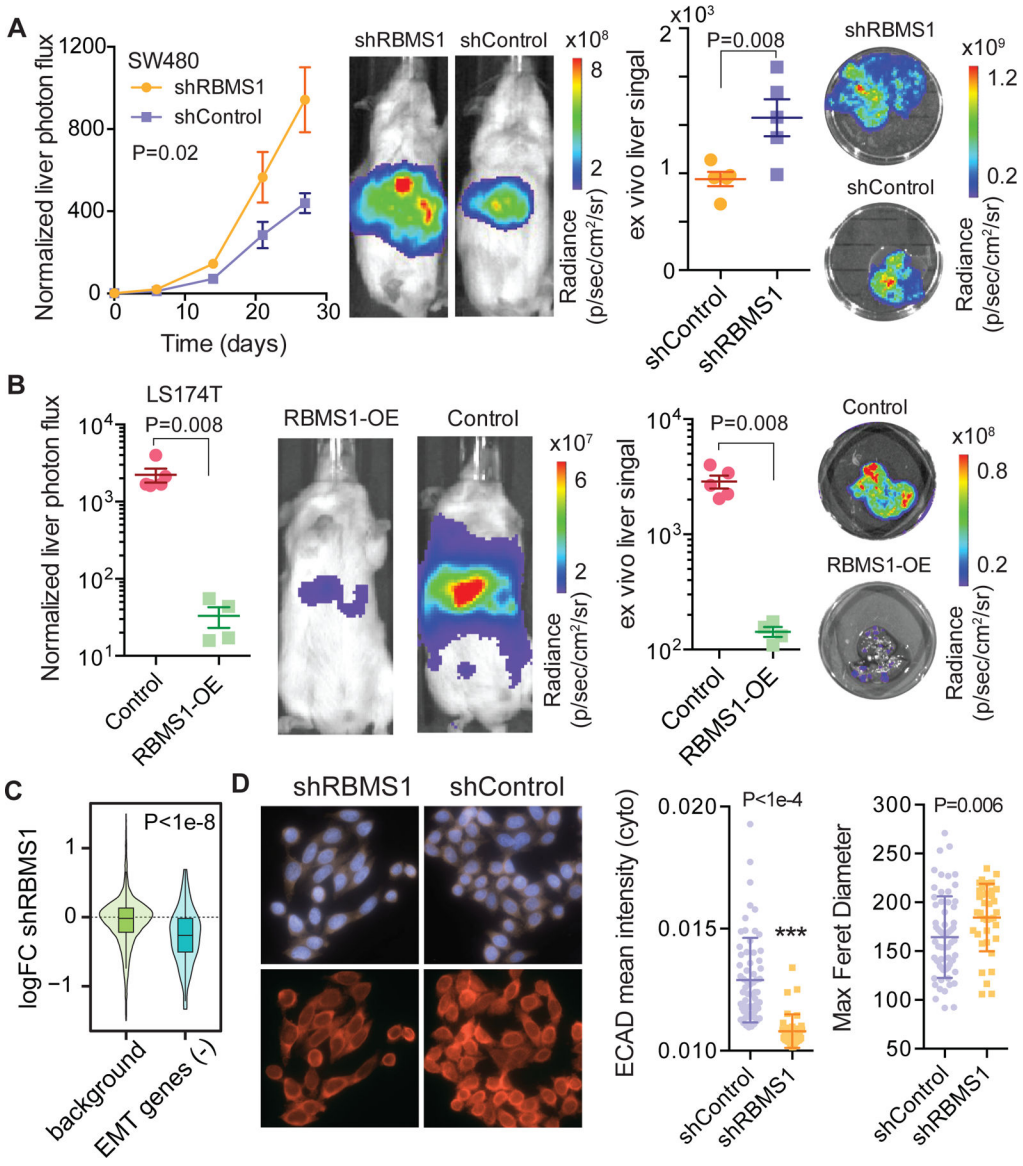

Figure 4. RBMS1 is a suppressor of epithelial-mesenchymal transition (EMT) and metastatic liver colonization.

(A) Bioluminescence imaging plot of liver colonization by *RBMS1* knockdown or control cells; N = 5 in each cohort. Two-way ANOVA was used for statistical testing. *Ex vivo* liver signal was also measured and compared using a one-tailed Mann-Whitney *U*-test. Also shown are representative (median signal) mice and livers. **(B)** Splenic injection of LS174T highly metastatic colon cancer cells overexpressing *RBMS1* and those expressing *mCherry* as a control. Day 21 signal (normalized to day 0) was plotted and compared for both *in vivo* and *ex vivo* signal (N = 4–5; Mann-Whitney *U*-test). **(C)** Downregulation of EMT(−) signature genes upon *RBMS1* knockdown in SW480 cells. 160-gene EMT(−) signature set (17) was compared to the rest of the transcriptome using a Mann-Whitney *U*-test. **(D)** Immunofluorescence staining for E-Cadherin (red) in *RBMS1* knockdown and control cells (SW480 background). Note the lower expression of E-Cadherin and the more spindle-like cellular morphology in *RBMS1* knockdown (top panels show DAPI signal). ECAD intensity

and maximum Feret Diameter (a measure of length of the cell) for cells in control and knockdown samples (N = 64 and 37, respectively). Two-tailed Mann-Whitney *U*-test was used to compare measurements.

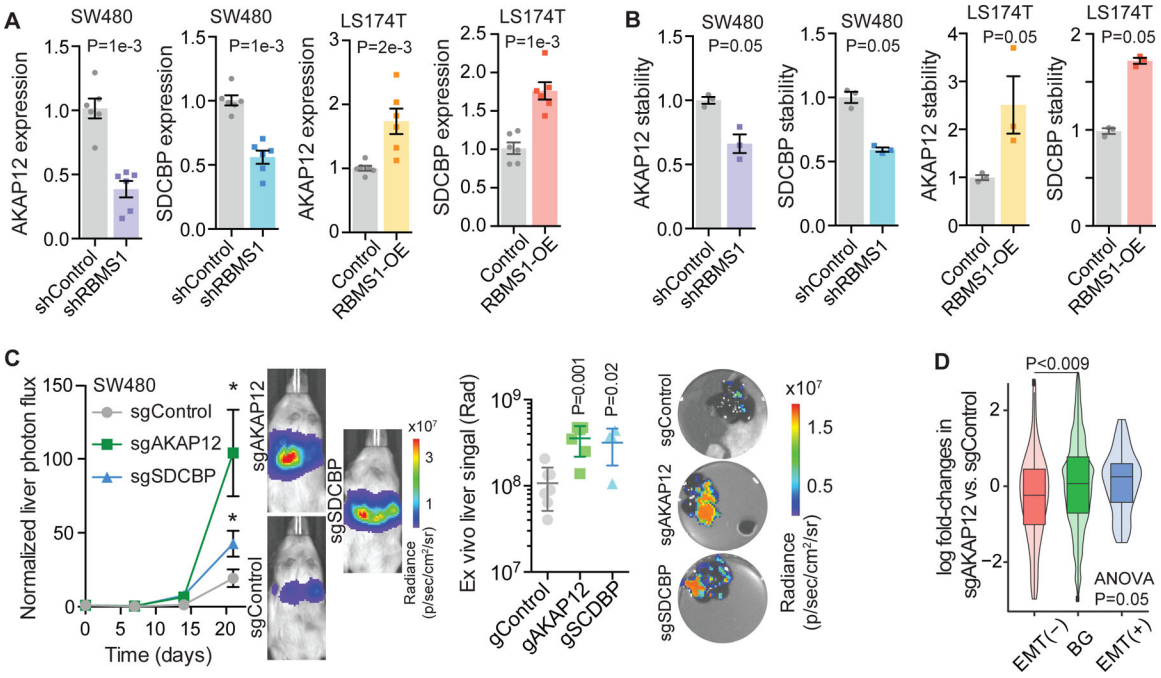

Figure 5. AKAP12 and SDCBP act downstream of RBMS1 to suppress CRC metastasis.

(A) Changes in the expression of *AKAP12* and *SDCBP* mRNAs upon *RBMS1* knockdown in SW480 cells and *RBMS1* over-expression in LS174T cells. The expression of target genes was determined by qRT-PCR, normalized to *HPRT* internal control and shown as relative fold change over shControl or OE-Control. *P*-value was calculated using one-tailed Mann-Whitney *U*-test. **(B)** Changes in the stability of *AKAP12* and *SDCBP* mRNAs upon *RBMS1* knockdown in SW480 cells and *RBMS1* over-expression in LS174T cells. RNA stability was calculated by comparing mRNA levels with and without treatment with α-amanitin, and shown as relative fold change over shControl or OE-Control. The relative abundance of target genes was determined by qRT-PCR, normalized to 18S RNA (RNA Pol-I transcript insensitive to α-amanitin). *P*-value was calculated using one-tailed Mann-Whitney *U*-test. **(C)** *In vivo* liver colonization assays were used to measure the impact of CRISPRi-mediated silencing of the RBMS1 targets *AKAP12* and *SDCBP* on liver metastasis (N = 6). Also shown are representative mice from each cohort. Two-way ANOVA was used to compare cohorts to control (P = 0.01 and 0.02 for sgAKAP12 and sgSDCBP respectively). Livers were also extracted and their tumor burden was measured *ex vivo.* Mann-Whitney *U*-test was used to compare measurements. **(D)** The expression of EMT(−) and EMT(+) signature genes relative to background (BG) in *AKAP12* knockdown (CRISPRi) and control cells (measurements using 3'-end RNA-seq). Shown are the ANOVA *p*-value and a Mann-Whitney comparison between EMT(−) and background genes.

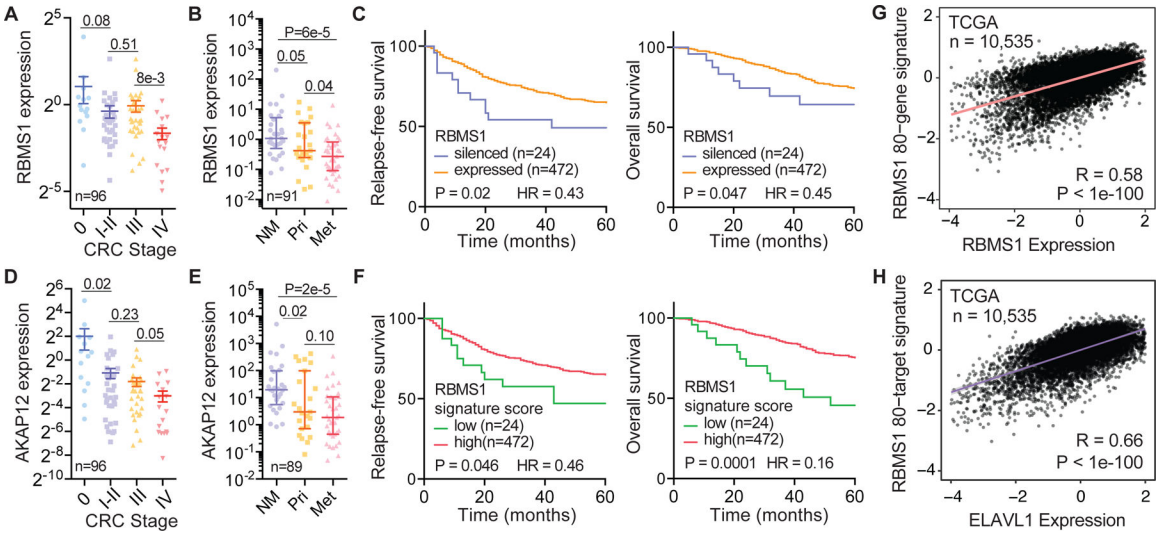

Figure 6. RBMS1 and its target gene signature score are associated with colon cancer metastasis and reduced survival in CRC patients.

(A) *RBMS1* qPCR (relative to *HPRT*) in 96 colon cancer samples stratified based on tumor stage. **(B)** *RBMS1* qPCR (relative to *HPRT*) in 29 normal mucosa, 25 primary colon cancer, and 37 liver metastases. **(C)** *RBMS1* silencing was observed in ~5% of primary CRC tumors (see methods), and these patients showed substantially lower relapse-free and overall survival. Reported are Mantel-Haenszel hazard ratios (HR) and *p*-values from Gehan-Breslow-Wilcoxon tests. **(D-E)** *AKAP12* expression in clinical samples (similar to (A) and (B)). **(F)** Patient primary tumors were scored based on aggregate expression of RBMS1 signature genes and the resulting values were used to perform survival analyses similar to those in (C). As shown here, lower RBMS1 signature score was significantly associated with lower relapse-free and overall survival. Also reported are Mantel-Haenszel hazard ratios (HR) and *p*-values from Gehan-Breslow-Wilcoxon tests. **(G-H)** Regression analysis comparing the expression of *RBMS1* (G) or *ELAVL1* (H) and RBMS1 80-gene signature set in TCGA pan-cancer dataset. Shown are the Spearman correlation coefficient and the associated *p*-value.

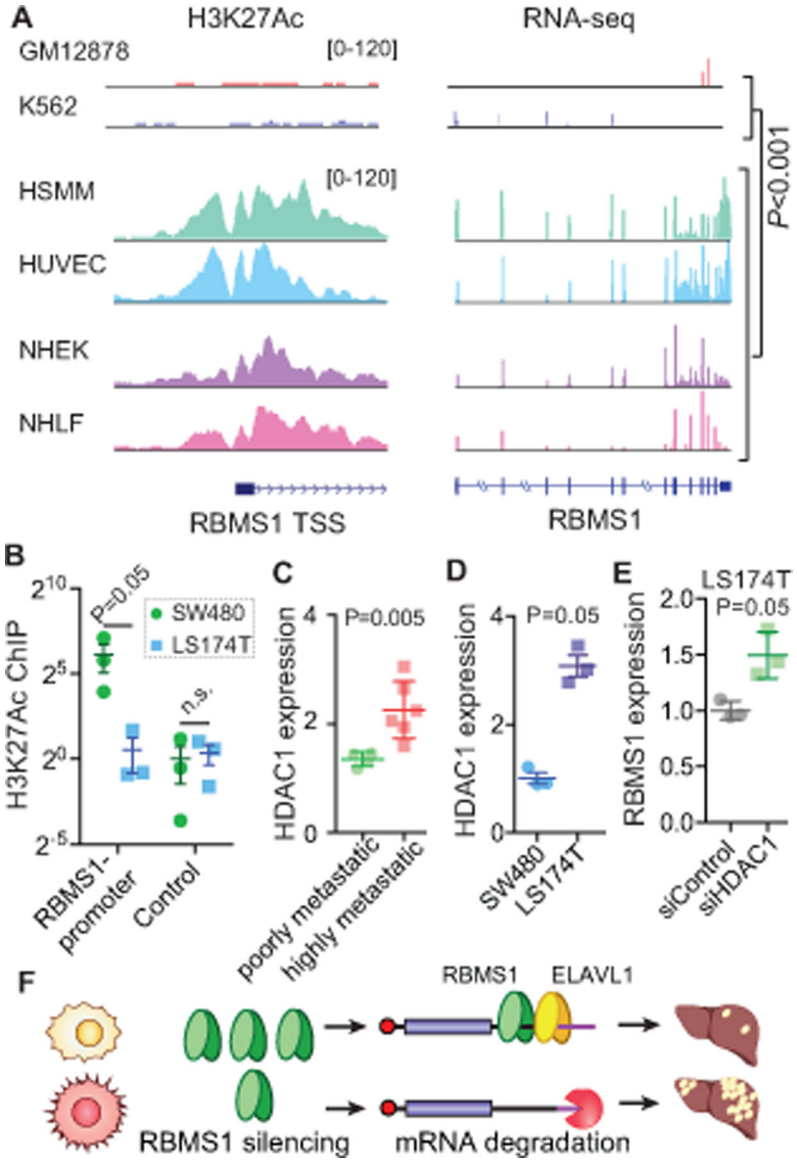

Figure 7. HDAC-mediated promoter deacetylation results in *RBMS1* silencing.

(A) *RBMS1* shows dynamic expression and acetylation changes across different cell types. Also shown here is the association between *RBMS1* expression and its promoter acetylation (source from ENCODE). **(B)** *RBMS1* promoter acetylation levels were measured in SW480 and LS174T cells using H3K27Ac ChIP-qPCR. An unacetylated region ~40kb away from the promoter was used as control (N = 3). **(C)** *HDAC1* expression in colon cancer lines stratified based on their metastatic capacity. One-tailed *U*-test was used to compare the two groups. Cell line names are listed in Supplementary Fig. 1A. **(D)** Quantitative PCR to compare *HDAC1* levels in the highly metastatic LS174T cells (with silenced *RBMS1*) relative to poorly metastatic SW480 cells. **(E)** RT-qPCR was used to measure *RBMS1* mRNA levels in LA174T cells with RNAi-mediated *HDAC1* silencing. **(F)** A schematic

model of *RBMS1* silencing and its role in suppressing colon cancer metastasis. One-tailed Mann-Whitney *U* tests were used to assess statistical significance for all panels.