

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

**Loss of the multifunctional RNA-binding protein RBM47 as a  
source of selectable metastatic traits in breast cancer**

Sakari Vanharanta<sup>1, 2, 5</sup>, Christina B. Marney<sup>3, 5</sup>, Weiping Shu<sup>1</sup>, Manuel Valiente<sup>1</sup>,  
Yilong Zou<sup>1</sup>, Aldo Mele<sup>3</sup>, Robert B. Darnell<sup>3,4,\*</sup> and Joan Massagué<sup>1,4,\*</sup>

<sup>1)</sup> Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer  
Center, New York, New York 10065, USA

<sup>2)</sup> MRC Cancer Unit, MRC/Hutchison Research Centre, University of Cambridge,  
Cambridge, CB2 0XY, United Kingdom

<sup>3)</sup> Laboratory of Molecular Neuro-Oncology, The Rockefeller University, 1230  
York Avenue, New York, NY 10021, USA

<sup>4)</sup> Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA

<sup>5)</sup> These authors contributed equally to this work

\* For correspondence: massaguj@mskcc.org (J.M.) or  
darnelr@mail.rockefeller.edu (R.B.D)

## SUMMARY

The mechanisms through which cancer cells lock in altered transcriptional programs in support of metastasis remain largely unknown. Through integrative analysis of clinical breast cancer gene expression datasets, cell line models of metastatic breast cancer, and mutation data from cancer genome resequencing studies, we have identified *RNA binding motif protein 47 (RBM47)* as a suppressor of breast cancer progression and metastasis. RBM47 inhibited breast cancer re-initiation and growth in experimental models. Transcriptome-wide HITS-CLIP analysis revealed widespread RBM47 binding to mRNAs, most prominently in introns and 3'UTRs. RBM47 altered splicing and abundance of a subset of its target mRNAs. Some of the mRNAs stabilized by RBM47, as exemplified by *dickkopf WNT signaling pathway inhibitor 1 (DKK1)*, inhibit tumor progression downstream of RBM47. Our work identifies RBM47 as a multifunctional RNA-binding protein that can suppress breast cancer progression. More generally, these findings demonstrate how the inactivation of a broadly targeted RNA chaperone can establish a platform for the selection of a pro-metastatic transcriptomic state.

## INTRODUCTION

Cancers arise through an evolutionary process that feeds from stochastic genetic alterations and selection [1]. The identities of the alterations that get selected for are rapidly coming to light through large-scale resequencing efforts. For example, several independent studies have characterized the mutational complement of breast cancer, one of the most common human malignancies [2-8]. Besides confirming previously known cancer genes, such as *TP53* and *PIK3CA*, most studies also report a long tail of rarely mutated genes. While many of these mutations are likely to be passenger events, some of them are potential mediators of tumor phenotypes. How to identify such low-frequency driver mutations remains a challenge.

In addition to mutations that directly promote tumorigenesis through specific alterations in cell signaling and repair pathways, many aberrations found in cancers do not affect cell signaling pathways directly, but rather, they support the stabilization of altered transcriptomic profiles that facilitate the emergence of pro-tumorigenic and metastatic traits. Mutations in epigenetic regulators fall into this category [9]. The phenotypic output of such alterations would depend on the activity of already existing signaling processes. Indeed, examples of epigenetic alterations that result in a phenotypic trait in the presence of a specific transcriptional program have been described [10]. Analogously, aberrations in the multi-step mRNA processing and turnover cascade [11] could also lock in aberrant transcriptomic states.

Precise regulation of RNA metabolism is fundamental in the generation of biological complexity in both normal and disease states [12, 13]. The concerted action of multiple RNA binding proteins (RBPs) regulate the spatial, temporal and functional dynamics of the transcriptome via alternative splicing, alternative polyadenylation and transcript stability [11]. While malignancy-associated dysregulation of RNA metabolism via aberrant microRNA expression is relatively well established [14, 15], a growing body of evidence indicates a prominent role also for RBPs in both the development and progression of cancer. For example,

76 upregulation of the splicing regulator SRSF1 is associated with multiple tumor  
77 types [16], and is necessary for the oncogenic activity of MYC in lung cancer [17].  
78 Conversely, RBM5 has been shown to be tumor suppressive in several cancer  
79 models [18-20].

80 We hypothesized that combining cancer genome resequencing data with gene  
81 expression information from both clinical data sets and experimental model  
82 systems of metastasis would allow the identification of rarely mutated cancer  
83 genes with potential functional significance. This approach identified *RNA binding*  
84 *motif protein 47 (RBM47)* as a suppressor of breast cancer progression. By  
85 analyzing the transcriptome-wide RBM47 binding patterns we demonstrate that  
86 RBM47, a previously uncharacterized RNA-binding protein, modulates mRNA  
87 splicing and stability. Loss of RBM47 function thus provides a specific example of  
88 the power of global RNA modulatory events in the selection of pro-metastatic  
89 phenotypic traits.

90  
91

## RESULTS

### RBM47 inactivation associated with breast cancer progression

We combined gene expression data from triple negative metastatic breast cancer models [21, 22] and a cohort of 368 untreated clinical breast cancer cases [22, 23] with mutational data from a brain metastasis that originated from a basal breast cancer [2] (Figure 1A). Specifically, we looked for genes that had reduced mRNA expression in functionally metastatic cancer cells, evidence for low mRNA expression associated with poor patient outcome in clinical samples, and an enriched mutation in the brain metastasis sequenced by Ding et al. [2]. *RBM47*, a gene encoding a previously uncharacterized putative RNA-binding protein was the only one that fulfilled all these criteria [24-26]. We confirmed the lower expression of *RBM47* mRNA in the highly metastatic cells (Figure 1B). This translated into a comparable difference at the protein level (Figure 1C). In the clinical data sets, low *RBM47* mRNA expression was significantly associated with relapse to brain and lung (Figure 1D-E) but not to bone (Figure 1F). In multivariate analysis combining *RBM47* expression with estrogen, progesterone and HER2 receptor status (ER, PR and HER2), the association with brain metastasis remained statistically significant (Figure 1 – figure supplement 1A). We further characterized the expression patterns of *RBM47* in the TCGA cohort of 748 breast cancer samples studied by RNA-seq [3, 27]. We found that low *RBM47* expression was significantly associated with claudin-low and basal breast cancers (Figure 1G), two subtypes of poor prognosis [28, 29].

The *RBM47*<sup>J281fs</sup> mutation first reported in a brain metastasis truncates the protein from the third RNA recognition motif (RRM) onwards (Figure 1H). As this mutation was already present in a minority subpopulation of the corresponding primary tumor [2], we looked for additional evidence of genetic *RBM47* aberrations in primary breast cancer cohorts. The catalogue of somatic mutations in cancer (COSMIC) database [30] reported 9 non-synonymous mutations in *RBM47*, three of which were frameshift mutations truncating one or more of the RRM domains (Figure 1 – figure supplement 1B). Furthermore, analysis of the

data from the TCGA cohort revealed that in basal breast cancer, *RBM47* was targeted by a mutation or homozygous deletion in ~8% of the cases (Figure 1 – figure supplement 1C). Moreover, heterozygous loss of the *RBM47* locus was present in 30% of the TCGA cohort [27]. These correlative analyses of multiple different breast cancer data sets, both experimental systems and large clinical patient cohorts, suggested that reduced expression or function of *RBM47* is associated with breast cancer progression already within primary tumors, and that clones with reduced *RBM47* function may display enhanced lung and brain metastatic fitness.

### ***RBM47* suppresses breast cancer progression**

In order to test the role of *RBM47* as a suppressor of breast cancer progression, we stably introduced both wild type *RBM47* and *RBM47*<sup>l281fs</sup> in the lung metastatic (LM2) and brain metastatic (BrM2) derivatives of the MDA-MB-231 triple negative breast cancer cells (MDA231 for short), respectively (Figure 2 – figure supplement 1A) [21, 22]. Of note, despite robust mRNA expression, the mutant *RBM47*<sup>l281fs</sup> protein levels were low, indicating that the mutant protein was unstable and therefore inactive (Figure 2 – figure supplement 1B). As determined by *in vivo* bioluminescence in experimental metastasis assays, wild type *RBM47* inhibited lung colonization by the MDA231-LM2 cells, when compared to *RBM47*<sup>l281fs</sup> (Figure 2A-B). Similarly, we observed extended brain metastasis free survival in mice inoculated with the MDA231-BrM2 cells expressing *RBM47* when compared to those expressing *RBM47*<sup>l281fs</sup> (Figure 2C). This difference translated into reduced brain metastatic burden as determined by ex vivo imaging (Figure 2D-E and Figure 2 – figure supplement 1C). We then tested the effects of inhibiting endogenous *RBM47* in the metastatic colonization of the parental MDA231 cells. With comparable effects of *RBM47* re-expression seen in both the lung and brain metastasis models (Figure 2A and 2E), we chose the lung colonization assay for loss-of-function studies as this allowed the simultaneous analysis of a greater number of tumor re-initiation events. Two *RBM47*-targeting shRNA constructs significantly shortened the lung metastasis free survival in

mice when compared to controls (Figure 2F and Figure 2 – figure supplement 1D), as determined by bioluminescence imaging (Figure 2G-H). This result was confirmed with a second more indolent cancer cell clone (Figure 2 – figure supplement 1E-G).

### Clonal heterogeneity in RBM47 sensitivity

The initial functional experiments suggested that the tumor suppressive effect of RBM47 on the overall population of metastatic cancer cells was modest. This could reflect either weak tumor suppressive function of RBM47 in general, heterogeneity in the sensitivity to RBM47 among different cancer cell subpopulations, or in the case of RBM47 reintroduction, loss of transgene expression. In order to distinguish between these possibilities, we used immunohistochemistry to assess RBM47 expression in the lung nodules formed either by wild type *RBM47* or *RBM47*<sup>l281fs</sup> expressing cells. This revealed that some cancer clones were able to form robust lung metastasis even in the presence of RBM47 (Figure 3A and Figure 3 – figure supplement 1A), but that many of the metastases formed after the inoculation of wild type RBM47-expressing cells had avoided or suppressed the expression of RBM47 (Figure 3B and Figure 3A – figure supplement 1A). As expected, the *RBM47*<sup>l281fs</sup>-expressing tumors contained only weakly staining cancer cells intermingled with small cells with strong RBM47 expression (Figure 3C), similar to those seen in normal lung parenchyma (Figure 3D). The rate of proliferation as determined by Ki67 immunohistochemistry did not correlate with the level of RBM47 expression (Figure 3 – figure supplement 1A). This was in line with the idea that some clones were more sensitive to RBM47 than others, but also that a strong selective pressure led metastatic cells to lose RBM47, a finding consistent with the initial observation of RBM47 loss in metastatic cell populations (Figure 1C).

In order to allow better experimental control over RBM47 expression we utilized a conditional expression system. By focusing on the two brain metastatic cell lines that expressed the lowest levels of endogenous RBM47 (Figure 1C) we generated single cell-derived clones with doxycycline-inducible expression of

either wild type RBM47 (henceforth WT10 and WT6, respectively) or the patient-derived mutant RBM47<sup>I281fs</sup> (henceforth MUT3). All clones exhibited dose-dependent RBM47 mRNA upregulation upon doxycycline induction (Figure 3 – figure supplement 2A-C) that translated into increased RBM47 protein expression in the wild type-expressing clones (Figure 3 – figure supplement 2D-E). Cancer cells also tolerated RBM47 in vitro (Figure 3 – figure supplement 2F). After confirming the feasibility of doxycycline-mediated conditional gene activation in metastatic brain lesions (Figure 3 – figure supplement 2G), we inoculated WT6, WT10 and MUT3 cells into immunocompromized mice and assessed their brain-metastatic phenotype. All clones formed robust brain metastases under doxycycline-free conditions (Figure 3E-J). The induction of *RBM47* expression with doxycycline in the diet inhibited robustly brain colonization of both metastatic cell clones, WT6 and WT10 (Figure 3E-G). However, the patient-derived mutant *RBM47*<sup>I281fs</sup> did not show any tumor suppressive effects (Figure 3H-J). Collectively, these results demonstrated that RBM47 was able to strongly inhibit metastatic functions of some cancer clones, whereas other clones were able to form metastasis despite the presence of RBM47.

### Transcriptome-wide identification of RBM47 binding sites

To determine whether RBM47 can directly bind RNA *in vivo*, we made use of the ability of UV-irradiation at 254nm to induce chemical crosslinks between RNA and proteins that are in direct contact [31].  $\gamma$ -<sup>32</sup>P-labeled RNA was detected by autoradiogram at ~76kD, the predicted size of Flag-RBM47, in Flag-immunoprecipitates from UV-irradiated, doxycycline-treated MDA231-BrM2 WT Flag-RBM47 cells, but not doxycycline-treated empty vector control or non-irradiated WT Flag-RBM47 expressing cells (Figure 4A). To identify directly bound RBM47 targets, a modified version of the high throughput sequencing and cross-linking immunoprecipitation (HITS-CLIP) protocol [32, 33] was carried out in duplicate on Flag-RBM47 expressing MDA231-BrM2 cells treated with doxycycline. (Figure 4B, see Materials and Methods). RBM47-bound HITS-CLIP reads were mapped to the human genome, yielding ~7.7x10<sup>6</sup> and ~2.0x10<sup>6</sup>



unique reads (tags) per replicate. Seventy-five percent of the tags mapped to regions corresponding to UCSC/Refseq genes, with a high degree of reproducibility of binding observed between replicates (Spearman correlation coefficient  $R^2=0.998$ , total tags per gene, **Figure 4C**).

To identify robust and reproducible RBM47-binding sites, tags were clustered to return regions with evidence of binding in both replicates (biological complexity, BC=2) with increasingly stringent filters (**Figure 4D**): tags per cluster (tags $\geq$ 2, 617,026 clusters; tags $\geq$ 10, 94,966 clusters), a previously described significant peak threshold (significant peak height  $\geq$ 10, 29,562 clusters [34]), and a ranked reproducibility chi-squared score ( $p\leq 0.01$ , 19,433 clusters [35]). As has been previously described for the n-ELAV proteins [36], identification of the most robust binding through increased stringency of cluster definition led to an increase in proportional binding in 3'UTR regions. For RBM47 this occurred due to loss of the majority of reproducible, yet relatively small intronic binding sites (**Figure 4E**), which may reflect the relative abundance of pre- and mature RNA message. Motif analysis (MEME, [37]) failed to identify an enriched RBM47-binding sequence in  $\pm$ 10nt footprints centered on the top 3,000 significant peaks (data not shown), but revealed a polyU sequence enriched in RBM47-binding sites containing cross-link induced mutations (deletion CIMS, [38], **Figure 4F**). The apparent lack of an enriched RBM47-binding motif within robust CLIP-derived binding sites may reflect the broad binding patterns observed, as exemplified by the predominantly 3'UTR binding in *DKK1* and *IL8* (**Figure 4G**).

#### **RBM47 regulates alternative splicing**

Reproducible binding of RBPs in intronic regions has proven to be predictive of a role in pre-mRNA processing for multiple proteins. To explore the relationship between RBM47 intronic binding and alternative splicing, RNA-seq was carried out in triplicate to compare MDA231-BrM2 cells and RBM47-expressing WT10 cells. Reads were mapped to cassette (CA) exon junctions as described previously for Mbnl2 [39], and an average inclusion rate (IR) calculated for each cell type to allow for the identification of reciprocal splicing changes while

normalizing for changes in RNA stability [40] (Figure 5A). The average change in inclusion rate ( $\Delta I$ ) was then calculated such that positive  $\Delta I$  indicates RBM47-dependent cassette exon inclusion. This analysis revealed 121 and 140 CA exons with significant RBM47-dependent inclusion and exclusion, respectively. To assess whether RBM47-binding occurred in the vicinity of these splice sites, HITS-CLIP tags in BC2 tags $\geq 5$  clusters (to account for lower levels of intronic binding seen in Figure 4E) in the region of the alternative splice were calculated. Forty-eight RBM47-bound included and 49 RBM47-bound excluded CA exons were identified in a total of 84 genes (Figure 5B and Supplementary file 1). RBM47-dependent splicing changes were confirmed via RT-PCR as shown for *SLK*, *MDM4*, *LIMCH1*, *MBNL1* and *SEC31A* (Figure 5C and Figure 5 – figure supplement 1).

By mapping normalized RBM47 CLIP tags associated with RBM47-dependent splicing changes on a composite transcript [32] we generated an RNA binding map of consensus binding sites within 1kb of exon-intron boundaries, with respect to exon inclusion or exclusion (Figure 5D). The resulting normalized complexity map reveals enriched RBM47 binding in the vicinity of splice acceptor sites of the CA exon and 3' CE for included alternative isoforms, while relative enrichment of RBM47 binding was seen at the 5' CE splice donor and 3'CE splice acceptor of excluded isoforms.

## **RBM47 affects mRNA steady state levels**

To further study the functional consequences of RBM47-mRNA binding events, we determined global steady state mRNA levels in the RBM47 expressing brain metastatic cells by RNA-seq. We took advantage of the clonal doxycycline-inducible cell line systems, which facilitated the analysis of RBM47 dose-dependence. First, genome-wide analyses showed that the mRNA level of several more genes correlated significantly with increasing concentrations of doxycycline in both cell lines expressing the wild type RBM47 (WT6 and WT10), when compared to cells expressing the mutant (Figure 6A). This result indicated that RBM47 elicits dose-dependent changes in mRNA levels. The correlation

coefficients followed a pattern that suggested the existence of mRNA species that correlated both positively and negatively with wild type, but not mutant, RBM47 expression, i.e. few genes in MUT3 cells had correlation coefficients close to 1 or -1, whereas in WT6 and WT10 cells such genes were abundant (Figure 6B). Encouraged by these observations, we looked for mRNAs that fulfilled the following criteria: (i) p-value of correlation less than 0.01 in both WT6 and WT10 cells, (ii) mRNA expression change detectable already at the lowest levels of RBM47 expression in both cell clones and (iii) no significant correlation with RBM47<sup>l281fs</sup> expression. This revealed a set of 102 mRNAs that were upregulated and 92 that were downregulated, respectively, in cells expressing the wild type RBM47 (Figure 6C and Figure 6 – figure supplement 1A-B and Supplementary file 2). Importantly, these changes were observed already with the lowest expression level of RBM47 that was comparable or lower than those detected in endogenously RBM47 expressing cells (Figure 3 – figure supplement 2D-E).

To determine whether the 194 RBM47-responsive genes displayed clinically meaningful expression patterns, we conducted an unsupervised hierarchical clustering analysis of the TCGA cohort of breast cancer specimens. Two main clusters emerged, one of which harbored characteristics of the RBM47-low phenotype ('Cluster 2', Figure 6 – figure supplement 1C). First, this subgroup, Cluster 2, had significantly lower RBM47 expression levels when compared to Cluster 1 (Figure 6D). Second, the genes that were induced upon RBM47 reintroduction tended to have lower expression in Cluster 2, whereas genes with reduced expression level in the RBM47-expressing cells tended to show higher expression in Cluster 2 (Figure 6E). These observations validated in clinical tumor samples the RBM47-dependent gene expression correlations identified *in vitro*.

To identify directly regulated targets of RBM47, we combined the mRNA expression data with the HITS-CLIP-derived RBM47 transcriptome-wide binding data. Of the 2498 strongest binding partners with >100 tags per transcript (total

tags in BC2 tags $\geq$ 10 clusters; [Supplementary file 3](#)), 25 were among the 102 RBM47-upregulated genes and 17 among the 92 RBM47-downregulated genes ([Figure 6F](#)), indicating no significant binding preference for either groups. This was reflected in the similar overall RBM47 binding profiles in both the up- and downregulated genes ([Figure 6G](#)). Two of the top-scoring RBM47 binding mRNAs, *IL8* and *DKK1* (17,079 and 16,208 tags in clusters per gene, respectively, [Figure 4G](#)), were among the upregulated genes. This increase in mRNA levels was associated with increased protein secretion as determined by ELISA, whereas VEGFA, the mRNA of which was bound but not upregulated by RBM47, showed no change in protein secretion ([Figure 6H](#)).

### **RBM47 modulates *DKK1* mRNA stability**

Nuclear RNA binding proteins typically function in large multiprotein complexes that regulate mRNA biogenesis [41]. Our data from both genome-wide HITS-CLIP and RNA-seq analysis was compatible with RBM47 being a member of these RNA chaperone units. This suggested that RBM47 may not necessarily have a direct tumor suppressive signaling function. Rather, it raised the possibility that loss of RBM47, leading to subtle changes in multiple mRNAs, either through stabilization, destabilization or alternative splicing, could be selected for if the net effect of both growth-promoting and growth-inhibiting changes would be beneficial for cancer cells under the stress of dissemination to and colonizing distant organs. In line with this, both the up and downregulated target genes of RBM47, as well as the genes that were targets of RBM47-mediated alternative splicing, contain genes that have previously been shown to either promote or inhibit tumor phenotypes ([Supplementary file 1 and Supplementary file 3](#)). For example, *DKK1* [42-44], *HTATIP2* [45], *HBP1* [46, 47], *MXI1* [48] and *CASP7* [49], all bound by RBM47 and upregulated upon RBM47 reintroduction, have known tumor suppressive functions. Similarly, RBM47-induced splicing changes were seen in genes such as *SLK* [50], *MDM4* [51] and *TNC* [52], all of which are genes with known functions in cancer.

Focusing on *DKK1*, one of the most robustly bound RBM47 target transcripts identified by HITS-CLIP, we investigated the possible role of RBM47 as a modulator of mRNA abundance. As predicted by our results in WT6 and WT10 cells, knockdown of RBM47 in two additional breast cancer cell lines expressing high levels of endogenous RBM47, SKBR3 and ZR-75-30, reduced *DKK1* mRNA levels (Figure 7A and Figure 7 – figure supplement 1A). This validated RBM47 as a modulator of *DKK1* mRNA level in breast cancer cells.

The fact that RBM47 bound to *DKK1* mRNA 3'UTR and increased *DKK1* mRNA levels suggested the possibility that RBM47 had the capability of stabilizing *DKK1* mRNA. We tested this by treating cancer cells with actinomycin D, a general inhibitor of transcription, and measuring *DKK1* mRNA levels in the following hours. This demonstrated that wild type *RBM47*, but not the *RBM47*<sup>Δ281fs</sup> mutant, was able to increase the half-life of *DKK1* mRNA by up to four-fold (Figure 7B).

#### **RBM47 protects *DKK1* 3'UTR from destabilizing factors**

As RBM47 binding to *DKK1* was concentrated on the 3'UTR, a known region of regulatory activity [53], we considered the possibility that RBM47 could compete with microRNAs or other mRNA destabilizing factors that target the 3'UTR [54, 55]. To test this experimentally, we generated miR-30-based shRNA-miR constructs that targeted different regions of the *DKK1* transcript, two at the 5' end with minimal RBM47 binding and three at the 3' end with abundant RBM47 signal (Figure 7C) [56]. All constructs knocked the *DKK1* transcript level down by 65-80% in the WT6 cells when no doxycycline was present (Figure 7D). The induction of RBM47 expression did not have a significant effect on the efficiency on the 5'-targeting shRNA-miRs (Figure 7E). In contrast, RBM47 inhibited the capability of the 3'-targeting shRNA-miRs to keep the *DKK1* mRNA level down (Figure 7E). Similar observations were made with the WT10 cells (Figure 7F-G). Argonaute HITS-CLIP in the MDA231-BrM2 cells (C.B.M. et al., unpublished data) indicates robust binding to the *DKK1* 3'UTR, supporting the regulatory role of this region. We conclude that the effects of RBM47 on *DKK1* mRNA levels

may be due to the ability of RBM47 to protect this mRNA from destabilizing factors, possibly through direct interaction with the *DKK1* mRNA.

### **RBM47 inhibits tumor progression by suppressing Wnt activity**

DKK1 is an inhibitor of Wnt signaling, a pathway with a well-established role in regulating stem cell characteristics in both normal and malignant cells [57]. Indeed, DKK1 as well as other Wnt antagonists have been shown to inhibit breast cancer progression [42-44, 58]. The fact that RBM47 was able to increase DKK1 secretion therefore suggested that RBM47 may also inhibit Wnt signaling and consequently reduce the tumorigenic fitness of metastatic breast cancer cells. We first tested the effects of RBM47 on cancer cell Wnt responsiveness by treating WT6 cells with recombinant Wnt3A and subsequently measuring the expression of *AXIN2*, a common TCF/ $\beta$ -catenin target gene and a general marker of Wnt activity [57, 59]. Doxycycline-induced expression of RBM47 in WT6 cells led to a dampened Wnt3A-dependent *AXIN2* induction (Figure 8A). This inhibition of *AXIN2* expression was at least partially dependent on DKK1 (Figure 8B). In agreement with these results, human breast cancers with low RBM47 expression had in general higher levels of Wnt transcriptomic activity when compared to tumors with high RBM47 expression in the TCGA cohort (Figure 8C).

We then tested the possibility that RBM47 would suppress tumorigenesis by implanting low numbers of WT6 cells at the orthotopic site. The induction of RBM47 by doxycycline was able to significantly delay the emergence of mammary tumors (Figure 8D), and the tumors that formed were smaller in size (Figure 8E). This observation was confirmed in WT10 cells (Figure 8 – figure supplement 1A). In line with this, inhibition of DKK1 expression in the MDA231-BrM2 cells promoted tumor formation in the orthotopic site (Figure 8F and S8A). Finally, we used the two exon 1 DKK1 shRNAmiR constructs (Figure 7C) to reduce DKK1 levels in the WT6 cells (Figure 8 – figure supplement 1B) and tested how this affected the brain metastatic ability of these cells in the presence of RBM47. Only 1/9 mice (16%) injected with control cells developed brain



metastasis, whereas 8/17 mice (47%) inoculated with DKK1 RNAi construct transduced cells developed metastasis (Figure 8G-H and Figure 8 – figure supplement 1C). Taken together, these observations suggested that RBM47-dependent suppression of tumor progression was partially mediated by its ability to increase the production of the Wnt antagonist DKK1, a secreted protein that can inhibit tumor phenotypes in metastatic cancer cells.

## DISCUSSION

Cancer genomes contain numerous genes with low-frequency mutations of unknown functional significance. We have studied one such gene, the previously uncharacterized *RBM47*, and demonstrate that it has tumor suppressive functions in breast cancer. RBM47 acts as a multifunctional RBP modulating alternative splicing and the abundance of several mRNAs, which can lead to inhibition of cancer progression (Figure 8I-J). These results highlight the significance of infrequent mutations in cancer, the importance of integrated experimental approaches to identify such functionally relevant mutations, and the role of broadly targeted mRNA chaperones as determinants of cancer progression.

RBM47 contains three classical RNA recognition motifs (RRM domains). The closest homologs of RBM47 are Apobec1 complementation factor (A1CF) and hnRNP-Q, which regulate RNA editing [60] and splicing and transcript stability, respectively [61, 62]. In general, RBPs bind to and influence the function and fate of both pre-mRNAs and mRNAs [41]. They can operate in large multiprotein complexes that dynamically regulate all the steps of mRNA biogenesis, nuclear export, stability and translation. Individual subunits of these complexes can therefore have diverse phenotypic roles depending on the exact protein complex they are in [63]. Our observations on RBM47 are in line with these known general principles of RBP function.

Our data demonstrate widespread and reproducible RBM47 binding to target mRNAs predominantly in introns and 3'UTRs, with the most robust binding occurring in 3'UTRs. Recent RNA-compete studies have proposed a binding motif for RBM47 *in vitro* [26]. Our *in vivo* HITS-CLIP data does not suggest a clear nucleotide binding specificity for exogenous Flag-tagged RBM47, although some preference was observed for polyU stretches around CIMS sites. It has been shown that the presence of a canonical motif is neither necessary nor sufficient to predict HITS-CLIP binding sites of FUS in both mouse and human brain [64], while specific sites suggested by *in vitro* RNA selection experiments are not enriched in HITS-CLIP derived FMRP binding sites *in vivo* [35]. This would suggest that other factors such as RNA accessibility, secondary structure or protein-protein interactions may modulate RBM47 target choice [65]. Further work is therefore needed for a comprehensive understanding of the determinants of RBM47-mRNA interactions.

We find that RBM47 binds robustly to ~2500 gene transcripts in human breast cancer cells, with only a subset showing steady state level change or alternative splicing upon RBM47 reintroduction. Given the stringent criteria used to define RBM47-bound and regulated targets and the generally low level of intronic RNA in a cell, it is likely that this subset is an underestimation of the number of RBM47-regulated transcripts, and does not take into account the potential for regulatory events, such as re-localization, that may not alter steady state transcript levels. The complex interplay of RBPs in agonistic and antagonistic modulation of mRNA is becoming increasingly apparent. For example, the RRM-domain containing protein HuR modulates the destabilizing effects of miRNAs [54, 55, 66] and AUF1 [67] on common target transcripts. The data presented here suggest that target-specific RBM47 regulation may arise through modulation of accessibility of other factors to a common mRNA transcript.

RBM47 binds and regulates transcripts that encode for proteins of several different biological functions. The effects of reduced RBM47 activity on cancer cell fitness, determined by the sum phenotypic output of all regulated target



transcripts, may therefore vary depending on the context. Such pleiotropic effects could target multiple steps of cancer progression. Indeed, even though RBM47 loss was associated with metastatic cancer clones in our model systems, evidence for selection against RBM47 was detected already in primary breast cancers. One of the most highly bound RBM47 mRNA targets, the secreted Wnt inhibitor *DKK1*, is stabilized by RBM47 and partially mediates RBM47 tumor suppressive function. Interestingly, *rbm47* knockdown in zebra fish embryos leads to a headless phenotype mediated via up-regulation of the *wnt8a* pathway [68]. In addition to *DKK1*, our analysis identified a number of potential mediators of RBM47 effects for future studies.

From a general perspective, the present findings illuminate two concepts. First, we show that low-frequency cancer mutations can give rise to tumorigenic phenotypes. Our work highlights the power of orthogonal approaches for the analysis of cancer genome resequencing data. Second, we show that loss of a broadly targeted and multifunctional RBP can increase the fitness of certain cancer cell clones in support of metastasis. This complements previous findings of RNA-binding proteins as mediators of oncogenic phenotypes [16, 17, 69-71]. Deregulation of RNA-binding proteins is thus emerging as a prominent source of complex transcriptomic diversity that can serve as a platform for the selection of metastatic traits during tumor progression.

469

## 470 **EXPERIMENTAL PROCEDURES**

471 **Cell lines.** The metastatic breast cancer cell lines have been previously  
472 described [21, 22, 72]. SKBR3, ZR-75-30 and HCC1954 cells were obtained from  
473 ATCC. For retrovirus and lentivirus production, GPG29 and 293T cells,  
474 respectively, were utilized. All cell lines were maintained under standard tissue  
475 culture conditions. Single cell-derived clones were isolated utilizing fluorescence-  
476 activated cell sorting from genetically engineered 231-BrM2 (WT10, MUT3) and  
477 CN34-BrM2 (WT6) cells.

478 **cDNA expression and RNAi.** For RBM47 restoration, *RBM47* was cloned into  
479 the pBABE-puro retroviral expression vector. The *RBM47*<sup>J281fs</sup> was generated by  
480 site-directed mutagenesis. Virus was generated in the GPG29 packaging cells.  
481 For the generation of doxycycline-inducible expression constructs, both wild type  
482 and mutant FLAG-RBM47 were cloned into the pRetroX-Tight-Pur expression  
483 system (Clontech).

484 For RNAi-mediated gene silencing, RBM47 pGIPZ shRNA constructs were  
485 obtained from Open Biosystems. The DKK1 shRNAmiR constructs were  
486 designed based on the rules described by [56] and cloned into the pGIPZ vector  
487 as described [73] using the following oligonucleotide templates:

488 DKK1miRm:

489 TGCTGTTGACAGTGAGCGACGGGTCTTTGTCGCGATGGTATAGTGAAGCCA  
490 CAGATGTATACCATCGCGACAAAGACCCGGTGCCTACTGCCTCGGA

491 DKK1miRn:

492 TGCTGTTGACAGTGAGCGACACCTTGAACCTCGGTTCTCAATAGTGAAGCCA  
493 CAGATGTATTGAGAACCGAGTTCAAGGTGGTGCCTACTGCCTCGGA

494 DKK1miRo:

495 TGCTGTTGACAGTGAGCGCCAACTCAATCCTAAGGATATATAGTGAAGCCAC  
496 AGATGTATATATCCTTAGGATTGAGTTGATGCCTACTGCCTCGGA

497 DKK1miRp:  
498 TGCTGTTGACAGTGAGCGACAGTAAATTACTGTATTGTAATAGTGAAGCCAC  
499 AGATGTATTACAATACAGTAATTTACTGCTGCCTACTGCCTCGGA

500 DKK1miRq:  
501 TGCTGTTGACAGTGAGCGAAACGGAAGTGTGATATGTTTATAGTGAAGCCA  
502 CAGATGTATAAACATATCACACTTCCGTTCTGCCTACTGCCTCGGA

503 The pGIPZ empty vector was used as a control.

504 **Animal studies.** All animal experiments were performed in accordance with a  
505 protocol approved by MSKCC Institutional Animal Care and Use Committee.  
506 Lung metastasis assays were conducted in 5-7 week old female NOD/SCID  
507 mice. Brain metastasis and mammary tumor assays were carried out using 5-7  
508 week old female athymic nude mice. *In vivo* bioluminescence imaging was  
509 performed using the IVIS Spectrum Xenogen machine (Caliper Life Sciences).  
510 For orthotopic mammary tumor assays, cells were mixed with Matrigel. Injections  
511 were confirmed and tumor growth was followed by bioluminescent imaging.  
512 Statistical significance of tumor and metastasis free survival was assessed by the  
513 log-rank test. Differences in raw and normalized bioluminescence signal was  
514 assessed by the Student's t-test and Wilcoxon rank-sum test, respectively. Brain  
515 images were acquired with a Leica SP5 up-right confocal microscope and Zeiss  
516 AxioVert 200M using 20X and 5X objectives. Image analysis was performed with  
517 Metamorph and ImageJ softwares.

518 **mRNA and protein detection.** Total RNA was extracted using PrepEase RNA  
519 spin kit (USB). We used Transcriptor First Strand cDNA Synthesis Kit (Roche) for  
520 cDNA synthesis. Quantitative PCR was performed using predesigned Taqman  
521 gene expression assays (Life Technologies) and the 7900HT or ViiA 7 real-time  
522 PCR systems (Applied Biosystems/Life Technologies). *TBP* was used as a  
523 housekeeping control gene. mRNA stability was assessed by performing  
524 quantitative PCR after actinomycin D treatment. For immunoblotting, antibodies  
525 recognizing RBM47 (HPA006347, Sigma),  $\alpha$ -tubulin (11H10, Cell Signaling) and

526 ACTB (Sigma) were utilized. Secondary antibodies were HRP (Pierce) or  
527 fluorescence (LiCor) conjugated. Immunostaining for RBM47 (HPA006347,  
528 Sigma) was performed according to standard protocols in the MSKCC Molecular  
529 Cytology Core Facility on paraffin embedded tissue blocks. Secreted protein was  
530 detected by ELISA (R&D).

531 **HITS-CLIP.** 231BrM2 tet-on FLAG-RBM47 cells were treated with 1ng/ml  
532 doxycycline for 3 days before 254nm UV crosslinking at 400mJ/cm<sup>2</sup> on a bed of  
533 ice (Stratalinker2400, Stratagene). Samples were processed for HITS-CLIP as  
534 previously described [32] using an anti-Flag antibody (Sigma F3165), and  
535 omitting 3' linker ligation in favor of direct labeling of protein-bound RNA with <sup>32</sup>P-  
536 γ-ATP. Non-crosslinked cells were used as a negative control, with IPed Flag-  
537 RBM47 protein detected using a second anti-Flag antibody (Sigma F7425).  
538 Purified RBM47-bound RNA fragments were polyA tailed (E-PAP, NEB), and  
539 reverse transcribed (Superscript III, Invitrogen) in the presence of Br-dUTP  
540 (Sigma). Unique polydT-NV RT-primers were used per replicate, containing  
541 Solexa sequences separated by an abasic furan (that serves as an Apel cut site),  
542 a 6nt degenerate region and a 6nt index sequence to allow for multiplexing  
543 during sequencing.

544 RT Primer 1

545 pGCACTGTTN<sub>6</sub>GATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGGC  
546 ATACGAT<sub>20</sub>VN

547 RT Primer 2

548 pGCGAAACTN<sub>6</sub>GATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGGC  
549 ATACGAT<sub>20</sub>VN

550 BrdU-cDNA was stringently purified by IP (Santa Cruz, sc-32323) using ProteinG  
551 dynabeads (Invitrogen), eluted from the beads via BrdU competitive elution  
552 (Sigma), and re-immunoprecipitated. cDNA was circularized on bead (CircLigase  
553 ssDNA Ligase II, Epicentre), washed and digested with Apel (NEB) to relinearize.  
554 cDNAs were eluted from the beads by heating to 98C in Phusion HF Buffer

(NEB), then PCR amplified using Phusion DNA polymerase (NEB) and SYBR Green I (Invitrogen) in an iQ5 real-time PCR machine in order to monitor amplification, with the samples being removed when the RFU signal reached ~1000.

P5 – aatgatacggcgaccaccgacaggttcagagttctacagtccgacg

P3 - caagcagaagacggcata

PCR products were purified using MinElute columns (Qiagen) as per manufacturers instructions and quantified using Quant-It (Invitrogen). cDNA was multiplexed and sequenced using Illumina Hi-Seq (small RNA sequencing primer – cgacaggttcagagttctacagtccgacgacg). All data analysis was done using the Galaxy platform [74], as previously described [32, 34, 35, 38].

#### **RT-PCR validation**

RT-PCR validation was carried out using total RNA from MDA231-BrM2 (Control) and WT10 cells (iScript, Bio-Rad, Accuprime Pfx Supermix 1, Life Technologies) as previously described [75]. In all cases lanes 1-3 contain an equal mixture of control and WT10 cell cDNA amplified at  $n-1$ ,  $n$  and  $n+1$  cycles, lane 4 contains an absence of reverse transcriptase control (-RT), with all other lanes corresponding to replicate samples of the indicated cell type amplified to  $n$  cycles. IR and  $\Delta I$  calculated using ImageJ [76].

**Bioinformatic analysis.** All analyses were conducted using R. Microarray data from human untreated tumor data sets (GSE2603 [22] and GSE2034 [23]) were preprocessed as described [77]. For the RNA-seq TCGA data set, normalized mRNA z-scores were downloaded from the TCGA cBio portal [27]. The microarray data from metastatic cell lines [21, 22] were processed with GCRMA together with updated probe set definitions using R packages *affy*, *gcrma* and *hs133ahsentrezgcdf* (version 10). Unsupervised hierarchical clustering was performed using the function *heatmap.2* with Pearson's correlation coefficient as the similarity metric. For survival analysis, a Cox proportional hazards model was

utilized as implemented in the *coxph* function in the R-package *survival*. For RNA-seq analysis of the cell lines, raw paired-end sequencing data were mapped to human genome (hg19 build) with STAR2.3.0 [78] using standard options. Reads mapped to each transcript were counted by HTSeq v0.5.4 [79] with default settings. The read count table was normalized to library size by DESeq [79]. Correlation with RBM47 expression was assessed by Pearson's correlation coefficient utilizing the R functions *cor* and *cor.test*. Wnt pathway activity in clinical tumors was assessed utilizing a curated list of Wnt target genes in breast cancer [58] and calculating sums of z-scores for each tumor.

**ACCESSION NUMBERS**

RNA-seq data has been deposited to the Gene Expression Omnibus under the accession number GSE53779.

**ACKNOWLEDGEMENTS**

We thank members of the Massagué lab for discussion. This work was supported by a grant from the Starr Cancer Consortium to R.B.D. and J.M. J.M. and R.B.D are investigators of the Howard Hughes Medical Institute.

**LIST OF SUPPLEMENTARY MATERIALS**

- Figure 1 – figure supplement 1
- Figure 2 – figure supplement 1
- Figure 3 – figure supplement 1
- Figure 3 – figure supplement 2
- Figure 5 – figure supplement 1
- Figure 6 – figure supplement 1
- Figure 7 – figure supplement 1
- Figure 8 – figure supplement 1
- Supplementary file 1
- Supplementary file 2
- Supplementary file 3

## FIGURE LEGENDS

### Figure 1. *RBM47* expression associated with breast cancer progression.

(A) A schematic of the analytical approach. Genes identified as mutated in a breast cancer brain metastasis by Ding et al. (2010) were compared to metastasis-associated gene expression traits in both clinical data sets and experimental model systems. This identified *RBM47* as a putative breast cancer suppressor gene.

(B) *RBM47* mRNA expression measured by quantitative real-time RT-PCR in two cell line systems of breast cancer metastasis. In both panels, the data are normalized to the parental cell line (Par). Error bars represent 95% confidence intervals obtained from multiple PCR reactions. LM, lung metastatic derivative; BoM, bone metastatic derivative; BrM, brain metastatic derivative.

(C) *RBM47* protein expression measured by Western blotting. Samples as in (B). Tubulin used as a loading control.

(D) Brain metastasis free survival in a cohort of 368 untreated breast cancer patients. Cases classified based on *RBM47* mRNA expression, bottom 1/3 in blue, top 2/3 in red. P-value derived from a Cox proportional hazards model using *RBM47* expression as a continuous variable.

(E) Lung metastasis free survival in a cohort of 368 untreated breast cancer patients. Cases classified based on *RBM47* mRNA expression, bottom 1/3 in blue, top 2/3 in red. P-value derived from a Cox proportional hazards model using *RBM47* expression as a continuous variable.

(F) Bone metastasis free survival in a cohort of 368 untreated breast cancer patients. Cases classified based on *RBM47* mRNA expression, bottom 1/3 in blue, top 2/3 in red. P-value derived from a Cox proportional hazards model using *RBM47* expression as a continuous variable.



**(G)** RBM47 expression as measured by RNA-seq in the TCGA data set of 748 patients. Samples grouped based on breast cancer molecular subtype: luminal A, luminal B, Her2 positive, claudin low and basal. RBM47 expression is lower in claudin low and basal subtypes, both of which are associated with poor patient outcome.

**(H)** A schematic showing the predicted protein structure of RBM47, its closest homologue A1CF, a known RNA-binding protein, and the predicted structure of RBM47<sup>I281fs</sup> mutant. Blue diamonds represent RRM motifs, pink rectangle represents a truncated RRM motif.

**Figure 1 – figure supplement 1.**

**(A)** Multivariate Cox proportional hazards models for both brain and lung metastasis free survival in a cohort of 368 untreated breast cancer patients. The model incorporates *RBM47* expression as a continuous variable, estrogen receptor (ER) expression, progesterone receptor (PR) expression and Her2 expression.

**(B)** A list of protein-altering *RBM47* breast cancer mutations in the Catalogue of Somatic Mutations in Cancer database (COSMIC, <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>) as of June 2013.

**(C)** Percentage of genetic alterations, non-synonymous mutations or homozygous deletions, in *RBM47* in the TCGA cohort of 748 breast cancers classified based on molecular subtypes: luminal A, luminal B, Her2 positive, claudin low and basal. A significant enrichment is observed in the basal subtype ( $P = 0.00015$ , Fisher's exact test).

**Figure 2. RBM47 suppresses metastatic breast cancer progression.**

670 **(A)** Normalized lung photon flux in mice after tail vein inoculation of MDA231-  
671 LM2 cells expressing either wild type *RBM47* or *RBM47*<sup>Δ281fs</sup>. P-value calculated  
672 utilizing repeated measures two-way ANOVA. N=6 for *RBM47*<sup>Δ281fs</sup>, N=8 for  
673 *RBM47*.

674 **(B)** Representative bioluminescence images from the experiment shown in  
675 (B). The color scale shows bioluminescence (photons/second).

676 **(C)** Brain metastasis free survival as determined by *in vivo* bioluminescence  
677 imaging in mice after intracardiac inoculation of MDA231-BrM2 cells expressing  
678 either wild type *RBM47* or *RBM47*<sup>Δ281fs</sup>. P-value calculated using the log-rank test.  
679 N=9 for *RBM47*<sup>Δ281fs</sup>, N=7 for *RBM47*.

680 **(D)** Representative bioluminescence images from the experiment shown in  
681 (C). The color scale shows bioluminescence (photons/second).

682 **(E)** Ex vivo quantification of bioluminescence from brain metastasis on day 42  
683 of the experiment shown in panel C. P-value calculated by two-tailed Student's t-  
684 test.

685 **(F)** Lung metastasis free survival as determined by *in vivo* bioluminescence  
686 imaging in mice after tail vein inoculation of parental MDA-MB-231 cells  
687 expressing either control vector (pGIPZ) or hairpins against *RBM47* (shRNA1  
688 and shRNA2). P-values calculated using the log-rank test. N=8 for Ctrl group,  
689 N=6 for shRNA1, N=5 for shRNA2.

690 **(G)** *In vivo* bioluminescence imaging on day 36 of the experiment shown in (E)  
691 demonstrates earlier emergence of detectable lung metastasis for the *RBM47*  
692 knockdown groups when compared to the control animals. The color scale shows  
693 bioluminescence (photons/second).

694 **(H)** Quantification of bioluminescence for the time point shown in (F). Data  
695 normalized to signal on day 0 for each animal. P-value calculated using the  
696 Wilcoxon rank-sum test.

**Figure 2 – figure supplement 1.**

**(A)** *RBM47* mRNA expression measured by quantitative real-time RT-PCR in 231-BrM2 and 231-LM2 cells transduced with either wild type *RBM47* or *RBM47*<sup>l280fs</sup>-expressing cDNA constructs, respectively. The data are normalized to the untransduced control cell lines. Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(B)** Flag-RBM47 protein expression determined by Western blotting in non-clonal 231BrM2 tet-on Flag-RBM47 cells. No RBM47 or Flag protein is detected in the cells expressing Flag-RBM47<sup>l281fs</sup> (predicted size 34kDa) treated with the same doxycycline doses.

**(C)** Quantification of bioluminescence from the head of mice on day 0 of the experiment shown in Figure 2A. P-value calculated by two-tailed Student's t-test.

**(D)** *RBM47* mRNA and protein expression measured by quantitative real-time RT-PCR and Western blotting, respectively, in parental MDA-231 cells expressing either control vector (pGIPZ) or hairpins against *RBM47* (shRNA1 and shRNA2). Error bars represent 95% confidence intervals obtained from multiple PCR reactions. Tubulin used as a protein loading control.

**(E)** *RBM47* mRNA and protein expression measured by quantitative real-time RT-PCR and Western blotting, respectively, in parental CN34 cells expressing either control vector (pGIPZ) or hairpins against *RBM47* (shRNA1 and shRNA2). Error bars represent 95% confidence intervals obtained from multiple PCR reactions. Tubulin used as a protein loading control.

**(F)** Quantification of lung cancer cell burden on day 0 by *in vivo* bioluminescence imaging in mice after tail vein inoculation of parental CN34 cells expressing either control vector (pGIPZ) or hairpins against *RBM47* (shRNA1 and shRNA2). P-values calculated using two-sided Student's t test. N=5 for all groups.

**(G)** Quantification of lung cancer cell burden on day 141 by *in vivo* bioluminescence imaging in mice after tail vein inoculation of parental CN34 cells expressing either control vector (pGIPZ) or hairpins against RBM47 (shRNA1 and shRNA2). P-values calculated using two-sided Student's t test. N=4 for Ctrl group, N=5 for shRNA1, N=4 for shRNA2.

**Figure 3. Clonal heterogeneity in RBM47 sensitivity.**

**(A)** A lung metastatic nodule with strong RBM47 expression in a mouse inoculated with RBM47-transduced MDA231-LM2 cells. RBM47 protein expression detected by immunohistochemistry using an antibody against RBM47.

**(B)** A lung metastatic nodule with weak RBM47 expression in a mouse inoculated with RBM47-transduced MDA231-LM2 cells. RBM47 protein expression detected by immunohistochemistry. Note the clearly reduced staining when compared to panel (A).

**(C)** A lung metastatic nodule with weak RBM47 expression in a mouse inoculated with *RBM47*<sup>Δ280fs</sup>-transduced MDA231-LM2 cells. RBM47 protein expression detected by immunohistochemistry. Staining intensity similar to that seen in panel (B).

**(D)** RBM47 expression in normal mouse lung. RBM47 protein expression detected by immunohistochemistry. Note small cells with strong RBM47 expression, similar to those seen in panels (B) and (C).

**(E)** Brain metastasis free survival as determined by *in vivo* bioluminescence imaging in mice after intracardiac inoculation of WT10 cells. The RBM47 group received doxycycline in diet. P-value calculated using the log-rank test. N=9 for both groups.

**(F)** Representative ex vivo bioluminescence images from brain metastasis of the experiment shown in (E). The color scale shows bioluminescence (photons/second).

**(G)** Ex vivo quantification of bioluminescence from brain metastasis with and without RBM47 reintroduction. WT10 data from the experiment shown in (E). WT6 data from a similar experimental setup. P-value calculated by two-tailed Student's t-test. N=9 for Ctrl group, N=10 for RBM47 group.

**(H)** Brain metastasis free survival as determined by *in vivo* bioluminescence imaging in mice after intracardiac inoculation of MUT3 cells. The l281fs group received doxycycline in diet. P-value calculated using the log-rank test. N=9 for Ctrl group, N=7 for RBM47<sup>l281fs</sup> group.

**(I)** Representative ex vivo bioluminescence images from brain metastasis of the experiment shown in (H). The color scale shows bioluminescence (photons/second).

**(J)** Ex vivo quantification of bioluminescence from brain metastasis of the experiment shown in (H). P-value calculated by two-tailed Student's t-test.

#### **Figure 3 – figure supplement 1.**

**(A) Examples of** lung metastatic nodules in a mouse inoculated with RBM47-transduced 231-LM2 cells. Some metastases have strong expression of RBM47 (left), whereas others show only weak RBM47 expression (right). Staining against human vimentin and Ki67 is shown for the same lesions.

#### **Figure 3 – figure supplement 2.**

773 **(A)** *RBM47* mRNA expression measured by quantitative real-time RT-PCR in  
774 WT6 cells treated with increasing concentrations of doxycycline. Error bars  
775 represent 95% confidence intervals obtained from multiple PCR reactions.

776 **(B)** *RBM47* mRNA expression measured by quantitative real-time RT-PCR in  
777 WT10 cells treated with increasing concentrations of doxycycline. Error bars  
778 represent 95% confidence intervals obtained from multiple PCR reactions.

779 **(C)** *RBM47* mRNA expression measured by quantitative real-time RT-PCR in  
780 MUT3 cells treated with increasing concentrations of doxycycline. Error bars  
781 represent 95% confidence intervals obtained from multiple PCR reactions.

782 **(D)** RBM47 protein expression detected by Western blotting in WT6 cells  
783 treated with increasing concentrations of doxycycline. ACTB used as a loading  
784 control. Quantification of signal shown in the lower panel, normalized to both  
785 ACTB loading control and the level of endogenous RBM47 detected in HCC1954  
786 cells.

787 **(E)** RBM47 protein expression detected by Western blotting in WT10 cells  
788 treated with increasing concentrations of doxycycline. ACTB used as a loading  
789 control. Quantification of signal shown in the lower panel, normalized to both  
790 ACTB loading control and the level of endogenous RBM47 detected in HCC1954  
791 cells.

792 **(F)** Proliferation of WT6 cells assessed under standard tissue culture  
793 conditions with and without doxycycline (2.5ng/ml).

794 **(F)** Demonstration of doxycycline-inducible gene induction in brain metastatic  
795 lesions formed by 231-BrM2 cells transduced with an inducible RFP construct.  
796 Cancer cells express constitutively GFP. Doxycycline treatment started on day 14  
797 after intracardiac cancer cell inoculation. Doxycycline-inducible RFP expression  
798 can be seen in cancer cells.

799

**Figure 4. HITS-CLIP identifies genome-wide RBM47 binding maps.**

**(A)** Radiolabelled RNA is detectable in RBM47-expressing 231-BrM2 metastatic cells that have been UV-irradiated indicating *in vivo* RNA binding ability. No RNA is detected in non-crosslinked cells despite the presence of ample immunoprecipitated Flag-RBM47 protein. No RNA or protein is detected in control 231-BrM2 transduced with empty vector. Samples run in duplicate.

**(B)** Schematic of the modified HITS-CLIP protocol showing autoradiogram of duplicate Flag-RBM47 samples used. Purified RBM47-bound RNA fragments (green) were polyA tailed and reverse transcribed in the presence of Brd-U using a polydT-NV primer encoding the full sense sequence of the Illumina reverse sequencing primer (blue), an abasic furan that serves as an Apel cut site (⊙), a partial reverse complement to the Illumina forward sequencing primer (orange), and two hexamer sequences (purple): a known-sequence index for multiplexing and a degenerate barcode used to distinguish unique cDNA clones from PCR duplicates. cDNA were stringently purified, circularized and linearized using Apel to bring Illumina sequenced into the correct orientation with respect to the cloned fragment, and the samples PCR amplified and deep sequenced.

**(C)** RBM47 HITS-CLIP is highly reproducible between replicate experiments at the level of unique CLIP tags per transcript.

**(D)** Increasing the stringency of biologically reproducible RBM47 binding site definition reveals predominant binding in 3'UTRs and intronic regions of target transcripts, with the most robust binding (tags per binding site) evident in 3'UTRs.

**(E)** Distribution of tags number per biologically reproducible cluster in coding and non-coding regions of RBM47-targeted transcripts reveals a bimodal binding pattern between 3'UTRs and introns, with the latter having large numbers of reproducible yet less robust binding.

**(F)** MEME analysis reveals an enrichment for polyU sequences (50 sites,  $P = 2.4e^{-16}$ ) in the +/-10nt foot print region surrounding reproducible RBM47 deletions CIMS (357 sites with  $\geq 5$  mutations,  $FDR \leq 0.01$ , see Ref. [38]).

**(G)** Widespread RBM47 binding is seen in target transcripts, as exemplified by binding patterns seen in the 3'UTRs of DKK1 and IL8.

**Figure 5. RBM47 regulates alternative splicing.**

**(A)** Schematic showing the method used to calculate alternative exon inclusion rates from paired-end RNA-seq data. 5'CE - 5' flanking constitutive exon, 3'CE - 3' flanking constitutive exon, 5'FI - 5' Flanking intron, 3'FI - 3' Flanking intron.

**(B)** Scatter plot of all expressed alternatively spliced CA exons showing RBM47-dependent change in inclusion (black,  $\geq 10$  RNA-seq reads spanning exon-exon junctions,  $\Delta I \geq |0.2|$ ,  $p \leq 0.05$ ) with orange points indicating RBM47-bound and included CA exons, and blue points indicating bound and excluded exons, respectively. CA exons were considered bound given a total of tags  $\geq 10$  in BC2 tags  $\geq 5$  clusters mapping to the region spanning the start of the 5'CE to the end of the 3'CE. P-values calculated by Fisher's exact test using total isoform 1 and total isoform 2 RNA-seq reads in each condition.

**(C)** Left panel shows a section of the SLK transcript (blue) that includes a CA exon (grey box). The top two panels show RNA-seq data from WT10 (green) and control cells (red), with RBM47 HITS-CLIP tags mapping to the region shown beneath in black. Increased RNA-seq signal corresponding to the CA exon is seen in the presence of RBM47 expression, while robust binding is evident in the 5'FI. Independent RT-PCR validation of this splice is shown in the right panel, with IR calculated using ImageJ analysis of autoradiograms.

**(D)** Normalized complexity map of RBM47-dependent alternative splicing of CA exons. Orange and blue peaks represent binding associated with RBM47-dependent exon inclusion and exclusion, respectively.



854

855 **Figure 5 – figure supplement 1.**

856 **(A)** RBM47-dependent exclusion of exon 6 of MDM4, as in Figure 5C.

857 **(B)** RBM47-dependent exclusion of a CA in LIMCH1, as in Figure 5C.

858 **(C)** RBM47-dependent inclusion of exon 5 in MBNL1, as in Figure 5C.

859 **(D)** RBM47-dependent inclusion of two CA exons in SEC31A, as in Figure 5C

860

861 **Figure 6. RBM47-induced changes in mRNA levels.**

862 **(A)** Distribution of P-values from correlation analysis of doxycycline  
863 concentration and gene expression for all genes in WT6, WT10 and MUT3 cell  
864 lines, respectively. Global gene expression determined by RNA-seq.

865 **(B)** Distribution of correlation coefficients between doxycycline concentration  
866 and gene expression in WT6, WT10 and MUT3 cell lines, respectively.

867 **(C)** Heat maps showing the top 102 positively (UP) correlated and 92  
868 negatively (DOWN) correlated genes with RBM47 expression in WT6 and WT10  
869 cells. The expression of these genes does not correlate with *RBM47*<sup>Δ280fs</sup>  
870 expression in the MUT3 cells.

871 **(D)** *RBM47* mRNA expression in the TCGA cohort of breast cancer samples  
872 classified by the clusters shown in Figure 6 – figure supplement 1C. P-value  
873 determined by two-tailed Student's t-test.

874 **(E)** Fold change between Cluster 1 and Cluster 2 shown for the 102 positively  
875 and 92 negatively correlated RBM47 target genes, respectively. Positive fold  
876 change shows higher expression in Cluster 2, which has lower expression of  
877 RBM47. The genes that are induced upon RBM47 reintroduction tend to have  
878 lower expression in Cluster 2, and the genes that show lower expression upon

RBM47 reintroduction tend to have higher expression in Cluster 2. This is in line with the experimental results shown in panel C. P-value determined by two-tailed Student's t-test.

**(F)** Pie charts showing the fraction of target genes with more than 100 RBM47 tags for both the 102 positively and 92 negatively correlated RBM47 target genes.

**(G)** Tags per transcript plotted for both the positively and negatively correlated RBM47 target genes that showed more than 1 tag. The only two binding partners with  $>10^4$  tags represent *DKK1* and *IL8*, respectively.

**(H)** Secreted DKK1 and IL8 protein levels determined by ELISA in WT6 cells treated with the indicated doxycycline concentrations. VEGFA used as a control.

#### **Figure 6 – figure supplement 1.**

**(A)** Normalized average RNA-seq reads per transcript for the 102 positively correlated RBM47 target genes shown for WT6, WT10 and MUT3 cells treated with the indicated doxycycline concentrations.

**(B)** Normalized average counts for the 92 negatively correlated RBM47 target genes shown for WT6, WT10 and MUT3 cells treated with the indicated doxycycline concentrations.

**(C)** A heatmap showing unsupervised hierarchical clustering in the TCGA cohort of 748 samples using the 194 RBM47 target genes identified by RNA-seq. Two main clusters are detected.

#### **Figure 7. RBM47 modulates *DKK1* mRNA stability.**

**(A)** *RBM47* and *DKK1* mRNA expression measured by quantitative real-time RT-PCR in SKBR3 cells expressing either control vector (pGIPZ) or hairpins

against *RBM47* (shRNA1 and shRNA2). Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(B)** *DKK1* mRNA stability determined by measuring mRNA levels by quantitative real-time RT-PCR in WT6, WT10 and MUT3 cells, treated with or without doxycycline, after inhibition of transcription with actinomycin D. Data normalized to time point 0. Error bars represent 95% confidence intervals obtained from multiple PCR reactions. WT6:  $T_{1/2}$  Ctrl = 2.3h,  $T_{1/2}$  Dox = 9.8h; WT10:  $T_{1/2}$  Ctrl = 2.4h,  $T_{1/2}$  Dox = 5.4h; MUT3:  $T_{1/2}$  Ctrl = 2.1h,  $T_{1/2}$  Dox = 2.2h.

**(C)** Schematic showing the locations of different *DKK1* miR-constructs in relation to *DKK1* genomic locus and *RBM47* binding patterns. miRm and miRn target exon 1 that is not bound by *RBM47*. miRo, miRp and miRq target *DKK1* 3'UTR that is strongly bound by *RBM47*.

**(D)** *DKK1* mRNA expression measured by quantitative real-time RT-PCR in WT6 cells expressing either control vector (pGIPZ) or the five *DKK1*-targeting miR constructs shown in panel (E). Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(E)** *DKK1* mRNA expression measured by quantitative real-time RT-PCR in WT6 cells expressing either control vector (pGIPZ) or the five *DKK1*-targeting miR constructs, with or without doxycycline treatment. Data normalized to the non-treated control for each cell line separately. Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(F)** *DKK1* mRNA expression measured by quantitative real-time RT-PCR in WT10 cells expressing either control vector (pGIPZ) or the five *DKK1*-targeting miR constructs shown in panel (E). Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(G)** *DKK1* mRNA expression measured by quantitative real-time RT-PCR in WT10 cells expressing either control vector (pGIPZ) or the five *DKK1*-targeting miR constructs, with or without doxycycline treatment. Data normalized to the

non-treated control for each cell line separately. Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**Figure 7 – figure supplement 1.**

**(A)** *RBM47* and *DKK1* mRNA expression measured by quantitative real-time RT-PCR in ZR-75-30 cells expressing either control vector (pGIPZ) or a hairpin against *RBM47* (shRNA2). Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**Figure 8. RBM47 suppresses tumor progression via Wnt inhibition.**

**(A)** *AXIN2* mRNA levels determined by quantitative real-time RT-PCR in WT6 cells treated with recombinant WNT3A in the presence of increasing concentrations of doxycycline. Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(B)** Normalized level of *AXIN2* mRNA inhibition as determined by quantitative real-time RT-PCR in WT6 cells transduced with control vector or shRNAmiR constructs targeting the first exon of *DKK1*. The cells were treated with recombinant WNT3A in the presence of increasing concentrations of doxycycline. Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(C)** Wnt pathway activity assessed in the TCGA cohort of primary breast tumors, grouped by *RBM47* expression tertiles (L, low; M, medium; H, high). Wnt signature value calculated as sum of z-scores for a curated set of 16 Wnt target genes in breast cancer. P-value determined by linear regression analysis.

**(D)** Mammary tumor re-initiation assay. Five thousand WT6 cells implanted orthotopically in mice. *RBM47* induced by doxycycline feed. Tumor growth

detected by bioluminescence. P-value determined by the log-rank test. N=20 tumors for each group.

**(E)** Quantification of mammary tumor burden by *in vivo* bioluminescence imaging on day 33 of the experiment shown in (D). Data normalized to day 0 for each tumor. P-value calculated by the Wilcoxon rank-sum test.

**(F)** Quantification of mammary tumor burden by *in vivo* bioluminescence imaging in mice inoculated with 231-Brm2 cells transduced with either control (pGIPZ) or DKK1-targeting shRNAmiR constructs. Data normalized to day 0 for each tumor. P-value calculated by one-tailed Wilcoxon rank-sum test.

**(G)** Quantification of *ex vivo* brain bioluminescence shown for mice inoculated intracardiacly with WT6 cells transduced with either control (pGIPZ) or DKK1-targeting shRNAmiR constructs in the presence of RBM47, i.e. doxycycline in diet. One out of 9 (11%) control mice developed robust brain metastasis whereas 8/17 (47%) mice in the DKK1 knockdown groups showed metastasis. P-value calculated by one-tailed Student's t-test.

**(H)** Representative images of coronal brain sections analyzed for GFP immunofluorescence from the experiment shown in panel (G). Lesion contours are marked in white. Arrowheads indicate the lesions shown in higher magnification on the right; a similar brain area is shown for the control group. Scale bar 500µm.

**(I)** At the global level, RBM47 binds to ~2500 target mRNAs. However, the abundance or alternative splicing of only a fraction of these change depending on RBM47 status. The target genes represent molecules from various signaling pathways. The net effect of growth promoting and inhibiting alterations determine whether RBM47 loss is beneficial for a particular cancer clone.

**(J)** At the target mRNA level, the effects of RBM47 are dependent on the presence of other factors that modulate mRNA processing. Hence, the phenotype of RBM47 loss depends on the intracellular molecular milieu on a per

transcript basis. This is exemplified by the interaction of RBM47 with *DKK1* mRNA.

**Figure 8 – figure supplement 1.**

**(A)** Quantification of mammary tumor burden by *in vivo* bioluminescence imaging on day 36 after five thousand WT10 cells were implanted orthotopically in mice. RBM47 induced by doxycycline feed. Data normalized to day 0 for each tumor. P-value calculated by the Wilcoxon rank-sum test.

**(B)** *DKK1* mRNA measured by quantitative real-time RT-PCR in 231-BrM2 cells transduced with either control (pGIPZ) or DKK1-targeting shRNAmiR constructs. Error bars represent 95% confidence intervals obtained from multiple PCR reactions.

**(C)** Secreted DKK1 protein levels determined by ELISA in WT6 cells transduced with either control (pGIPZ) or DKK1-targeting shRNAmiR constructs, with or without doxycycline treatment.

**SUPPLEMENTARY FILES**

**Supplementary file 1** – RBM47-bound alternatively spliced exons showing RBM47-dependent  $\Delta I \geq |0.2|$ .

**Supplementary file 2** – mRNAs changed upon RBM47 reintroduction.

**Supplementary file 3** – RBM47 bound transcripts.

**REFERENCES**

1. Vogelstein, B., et al., *Cancer genome landscapes*. Science, 2013. **339**(6127): p. 1546-58.

- 1014 2. Ding, L., et al., *Genome remodelling in a basal-like breast cancer metastasis and*  
1015 *xenograft*. Nature, 2010. **464**(7291): p. 999-1005.
- 1016 3. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast*  
1017 *tumours*. Nature, 2012. **490**(7418): p. 61-70.
- 1018 4. Banerji, S., et al., *Sequence analysis of mutations and translocations across*  
1019 *breast cancer subtypes*. Nature, 2012. **486**(7403): p. 405-9.
- 1020 5. Stephens, P.J., et al., *The landscape of cancer genes and mutational processes in*  
1021 *breast cancer*. Nature, 2012. **486**(7403): p. 400-4.
- 1022 6. Shah, S.P., et al., *The clonal and mutational evolution spectrum of primary*  
1023 *triple-negative breast cancers*. Nature, 2012. **486**(7403): p. 395-9.
- 1024 7. Stephens, P.J., et al., *Complex landscapes of somatic rearrangement in human*  
1025 *breast cancer genomes*. Nature, 2009. **462**(7276): p. 1005-10.
- 1026 8. Shah, S.P., et al., *Mutational evolution in a lobular breast tumour profiled at*  
1027 *single nucleotide resolution*. Nature, 2009. **461**(7265): p. 809-13.
- 1028 9. Shen, H. and P.W. Laird, *Interplay between the cancer genome and epigenome*.  
1029 Cell, 2013. **153**(1): p. 38-55.
- 1030 10. Vanharanta, S., et al., *Epigenetic expansion of VHL-HIF signal output drives*  
1031 *multiorgan metastasis in renal cancer*. Nat Med, 2013. **19**(1): p. 50-6.
- 1032 11. Moore, M.J. and N.J. Proudfoot, *Pre-mRNA processing reaches back to*  
1033 *transcription and ahead to translation*. Cell, 2009. **136**(4): p. 688-700.
- 1034 12. Sharp, P.A., *The centrality of RNA*. Cell, 2009. **136**(4): p. 577-80.
- 1035 13. Licatalosi, D.D. and R.B. Darnell, *RNA processing and its regulation: global*  
1036 *insights into biological networks*. Nat Rev Genet, 2010. **11**(1): p. 75-87.
- 1037 14. Pencheva, N. and S.F. Tavazoie, *Control of metastatic progression by microRNA*  
1038 *regulatory networks*. Nat Cell Biol, 2013. **15**(6): p. 546-54.
- 1039 15. Di Leva, G., M. Garofalo, and C.M. Croce, *MicroRNAs in Cancer*. Annu Rev  
1040 Pathol, 2013.
- 1041 16. Karni, R., et al., *The gene encoding the splicing factor SF2/ASF is a proto-*  
1042 *oncogene*. Nat Struct Mol Biol, 2007. **14**(3): p. 185-93.
- 1043 17. Das, S., et al., *Oncogenic splicing factor SRSF1 is a critical transcriptional target*  
1044 *of MYC*. Cell Rep, 2012. **1**(2): p. 110-7.
- 1045 18. Mourtada-Maarabouni, M., et al., *Candidate tumor suppressor LUCA-*  
1046 *15/RBM5/H37 modulates expression of apoptosis and cell cycle genes*. Exp Cell  
1047 Res, 2006. **312**(10): p. 1745-52.
- 1048 19. Oh, J.J., et al., *3p21.3 tumor suppressor gene H37/Luca15/RBM5 inhibits*  
1049 *growth of human lung cancer cells through cell cycle arrest and apoptosis*.  
1050 Cancer Res, 2006. **66**(7): p. 3419-27.
- 1051 20. Oh, J.J., et al., *A candidate tumor suppressor gene, H37, from the human lung*  
1052 *cancer tumor suppressor locus 3p21.3*. Cancer Res, 2002. **62**(11): p. 3207-13.
- 1053 21. Bos, P.D., et al., *Genes that mediate breast cancer metastasis to the brain*.  
1054 Nature, 2009. **459**(7249): p. 1005-9.
- 1055 22. Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung*. Nature,  
1056 2005. **436**(7050): p. 518-24.
- 1057 23. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-*  
1058 *node-negative primary breast cancer*. Lancet, 2005. **365**(9460): p. 671-9.

- 1059 24. Baltz, A.G., et al., *The mRNA-bound proteome and its global occupancy profile*  
1060 *on protein-coding transcripts*. Mol Cell, 2012. **46**(5): p. 674-90.
- 1061 25. Castello, A., et al., *Insights into RNA biology from an atlas of mammalian*  
1062 *mRNA-binding proteins*. Cell, 2012. **149**(6): p. 1393-406.
- 1063 26. Ray, D., et al., *A compendium of RNA-binding motifs for decoding gene*  
1064 *regulation*. Nature, 2013. **499**(7457): p. 172-7.
- 1065 27. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for*  
1066 *exploring multidimensional cancer genomics data*. Cancer Discov, 2012. **2**(5):  
1067 p. 401-4.
- 1068 28. Smid, M., et al., *Subtypes of breast cancer show preferential site of relapse*.  
1069 Cancer Res, 2008. **68**(9): p. 3108-14.
- 1070 29. Lu, S., et al., *Claudin expression in high-grade invasive ductal carcinoma of the*  
1071 *breast: correlation with the molecular subtype*. Mod Pathol, 2013. **26**(4): p.  
1072 485-95.
- 1073 30. Forbes, S.A., et al., *COSMIC (the Catalogue of Somatic Mutations in Cancer): a*  
1074 *resource to investigate acquired mutations in human cancer*. Nucleic Acids  
1075 Res, 2010. **38**(Database issue): p. D652-7.
- 1076 31. Darnell, R.B., *HITS-CLIP: panoramic views of protein-RNA regulation in living*  
1077 *cells*. Wiley Interdiscip Rev RNA, 2010. **1**(2): p. 266-86.
- 1078 32. Licatalosi, D.D., et al., *HITS-CLIP yields genome-wide insights into brain*  
1079 *alternative RNA processing*. Nature, 2008. **456**(7221): p. 464-9.
- 1080 33. Weyn-Vanhentenryck, S.M., et al., *HITS-CLIP and Integrative Modeling Define*  
1081 *the Rbfox Splicing-Regulatory Network Linked to Brain Development and*  
1082 *Autism*. Cell Rep, 2014. **6**(6): p. 1139-52.
- 1083 34. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction*  
1084 *maps*. Nature, 2009. **460**(7254): p. 479-86.
- 1085 35. Darnell, J.C., et al., *FMRP stalls ribosomal translocation on mRNAs linked to*  
1086 *synaptic function and autism*. Cell, 2011. **146**(2): p. 247-61.
- 1087 36. Ince-Dunn, G., et al., *Neuronal Elav-like (Hu) proteins regulate RNA splicing*  
1088 *and abundance to control glutamate levels and neuronal excitability*. Neuron,  
1089 2012. **75**(6): p. 1067-80.
- 1090 37. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization*  
1091 *to discover motifs in biopolymers*. Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p.  
1092 28-36.
- 1093 38. Zhang, C. and R.B. Darnell, *Mapping in vivo protein-RNA interactions at single-*  
1094 *nucleotide resolution from HITS-CLIP data*. Nat Biotechnol, 2011. **29**(7): p.  
1095 607-14.
- 1096 39. Charizanis, K., et al., *Muscleblind-like 2-mediated alternative splicing in the*  
1097 *developing brain and dysregulation in myotonic dystrophy*. Neuron, 2012.  
1098 **75**(3): p. 437-50.
- 1099 40. Ule, J., et al., *Nova regulates brain-specific splicing to shape the synapse*. Nat  
1100 Genet, 2005. **37**(8): p. 844-52.
- 1101 41. Dreyfuss, G., V.N. Kim, and N. Kataoka, *Messenger-RNA-binding proteins and*  
1102 *the messages they carry*. Nat Rev Mol Cell Biol, 2002. **3**(3): p. 195-205.
- 1103 42. Bafico, A., et al., *An autocrine mechanism for constitutive Wnt pathway*  
1104 *activation in human cancer cells*. Cancer Cell, 2004. **6**(5): p. 497-506.



- 1105 43. Cowling, V.H., et al., *c-Myc transforms human mammary epithelial cells*  
1106 *through repression of the Wnt inhibitors DKK1 and SFRP1*. Mol Cell Biol, 2007.  
1107 **27**(14): p. 5135-46.
- 1108 44. Mikheev, A.M., et al., *Dickkopf-1 mediated tumor suppression in human breast*  
1109 *carcinoma cells*. Breast Cancer Res Treat, 2008. **112**(2): p. 263-73.
- 1110 45. Zhao, J., et al., *TIP30/CC3 expression in breast carcinoma: relation to*  
1111 *metastasis, clinicopathologic parameters, and P53 expression*. Hum Pathol,  
1112 2007. **38**(2): p. 293-8.
- 1113 46. Li, H., et al., *miR-17-5p promotes human breast cancer cell migration and*  
1114 *invasion through suppression of HBP1*. Breast Cancer Res Treat, 2011. **126**(3):  
1115 p. 565-75.
- 1116 47. Paulson, K.E., et al., *Alterations of the HBP1 transcriptional repressor are*  
1117 *associated with invasive breast cancer*. Cancer Res, 2007. **67**(13): p. 6136-45.
- 1118 48. Lahoz, E.G., et al., *Suppression of Myc, but not E1a, transformation activity by*  
1119 *Max-associated proteins, Mad and Mxi1*. Proc Natl Acad Sci U S A, 1994.  
1120 **91**(12): p. 5503-7.
- 1121 49. Hudson, R.S., et al., *MicroRNA-106b-25 cluster expression is associated with*  
1122 *early disease recurrence and targets caspase-7 and focal adhesion in human*  
1123 *prostate cancer*. Oncogene, 2013. **32**(35): p. 4139-47.
- 1124 50. Roovers, K., et al., *The Ste20-like kinase SLK is required for ErbB2-driven*  
1125 *breast cancer cell motility*. Oncogene, 2009. **28**(31): p. 2839-48.
- 1126 51. Wade, M., Y.C. Li, and G.M. Wahl, *MDM2, MDMX and p53 in oncogenesis and*  
1127 *cancer therapy*. Nat Rev Cancer, 2013. **13**(2): p. 83-96.
- 1128 52. Oskarsson, T., et al., *Breast cancer cells produce tenascin C as a metastatic*  
1129 *niche component to colonize the lungs*. Nat Med, 2011. **17**(7): p. 867-74.
- 1130 53. Zhou, A.D., et al., *beta-Catenin/LEF1 transactivates the microRNA-371-373*  
1131 *cluster that modulates the Wnt/beta-catenin-signaling pathway*. Oncogene,  
1132 2012. **31**(24): p. 2968-78.
- 1133 54. Bhattacharyya, S.N., et al., *Relief of microRNA-mediated translational*  
1134 *repression in human cells subjected to stress*. Cell, 2006. **125**(6): p. 1111-24.
- 1135 55. Young, L.E., et al., *The mRNA stability factor HuR inhibits microRNA-16*  
1136 *targeting of COX-2*. Mol Cancer Res, 2012. **10**(1): p. 167-80.
- 1137 56. Fellmann, C., et al., *Functional identification of optimized RNAi triggers using a*  
1138 *massively parallel sensor assay*. Mol Cell, 2011. **41**(6): p. 733-46.
- 1139 57. Clevers, H. and R. Nusse, *Wnt/beta-catenin signaling and disease*. Cell, 2012.  
1140 **149**(6): p. 1192-205.
- 1141 58. Matsuda, Y., et al., *WNT signaling enhances breast cancer cell motility and*  
1142 *blockade of the WNT pathway by sFRP1 suppresses MDA-MB-231 xenograft*  
1143 *growth*. Breast Cancer Res, 2009. **11**(3): p. R32.
- 1144 59. Lustig, B., et al., *Negative feedback loop of Wnt signaling through upregulation*  
1145 *of conductin/axin2 in colorectal and liver tumors*. Mol Cell Biol, 2002. **22**(4): p.  
1146 1184-93.
- 1147 60. Mehta, A., et al., *Molecular cloning of apobec-1 complementation factor, a novel*  
1148 *RNA-binding protein involved in the editing of apolipoprotein B mRNA*. Mol Cell  
1149 Biol, 2000. **20**(5): p. 1846-54.

- 1150 61. Chen, H.H., et al., *The RNA binding protein hnRNP Q modulates the utilization*  
1151 *of exon 7 in the survival motor neuron 2 (SMN2) gene*. Mol Cell Biol, 2008.  
1152 **28**(22): p. 6929-38.
- 1153 62. Weidensdorfer, D., et al., *Control of c-myc mRNA stability by IGF2BP1-*  
1154 *associated cytoplasmic RNPs*. RNA, 2009. **15**(1): p. 104-15.
- 1155 63. Chaudhury, A., P. Chander, and P.H. Howe, *Heterogeneous nuclear*  
1156 *ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's*  
1157 *multifunctional regulatory roles*. RNA, 2010. **16**(8): p. 1449-62.
- 1158 64. Lagier-Tourenne, C., et al., *Divergent roles of ALS-linked proteins FUS/TLS and*  
1159 *TDP-43 intersect in processing long pre-mRNAs*. Nat Neurosci, 2012. **15**(11): p.  
1160 1488-97.
- 1161 65. Li, X., et al., *Finding the target sites of RNA-binding proteins*. Wiley Interdiscip  
1162 Rev RNA, 2014. **5**(1): p. 111-30.
- 1163 66. Kim, H.H., et al., *HuR recruits let-7/RISC to repress c-Myc expression*. Genes  
1164 Dev, 2009. **23**(15): p. 1743-8.
- 1165 67. Zou, T., et al., *Polyamines regulate the stability of JunD mRNA by modulating*  
1166 *the competitive binding of its 3' untranslated region to HuR and AUF1*. Mol Cell  
1167 Biol, 2010. **30**(21): p. 5021-32.
- 1168 68. Guan, R., et al., *rbm47, a novel RNA binding protein, regulates zebrafish head*  
1169 *development*. Dev Dyn, 2013.
- 1170 69. Richard, S., et al., *Sam68 haploinsufficiency delays onset of mammary*  
1171 *tumorigenesis and metastasis*. Oncogene, 2008. **27**(4): p. 548-56.
- 1172 70. Sommer, G., et al., *The RNA-binding protein La contributes to cell proliferation*  
1173 *and CCND1 expression*. Oncogene, 2011. **30**(4): p. 434-44.
- 1174 71. Wang, J., et al., *Predictive and prognostic significance of cytoplasmic expression*  
1175 *of ELAV-like protein HuR in invasive breast cancer treated with neoadjuvant*  
1176 *chemotherapy*. Breast Cancer Res Treat, 2013. **141**(2): p. 213-24.
- 1177 72. Kang, Y., et al., *A multigenic program mediating breast cancer metastasis to*  
1178 *bone*. Cancer Cell, 2003. **3**(6): p. 537-49.
- 1179 73. Dow, L.E., et al., *A pipeline for the generation of shRNA transgenic mice*. Nat  
1180 Protoc, 2012. **7**(2): p. 374-93.
- 1181 74. Hillman-Jackson, J., et al., *Using Galaxy to perform large-scale interactive data*  
1182 *analyses*. Curr Protoc Bioinformatics, 2012. **Chapter 10**: p. Unit10 5.
- 1183 75. Licatalosi, D.D., et al., *Pthp2 represses adult-specific splicing to regulate the*  
1184 *generation of neuronal precursors in the embryonic brain*. Genes Dev, 2012.  
1185 **26**(14): p. 1626-42.
- 1186 76. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25*  
1187 *years of image analysis*. Nat Methods, 2012. **9**(7): p. 671-5.
- 1188 77. Zhang, X.H., et al., *Latent bone metastasis in breast cancer tied to Src-*  
1189 *dependent survival signals*. Cancer Cell, 2009. **16**(1): p. 67-78.
- 1190 78. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner*. Bioinformatics,  
1191 2013. **29**(1): p. 15-21.
- 1192 79. Anders, S. and W. Huber, *Differential expression analysis for sequence count*  
1193 *data*. Genome Biol, 2010. **11**(10): p. R106.
- 1194
- 1195

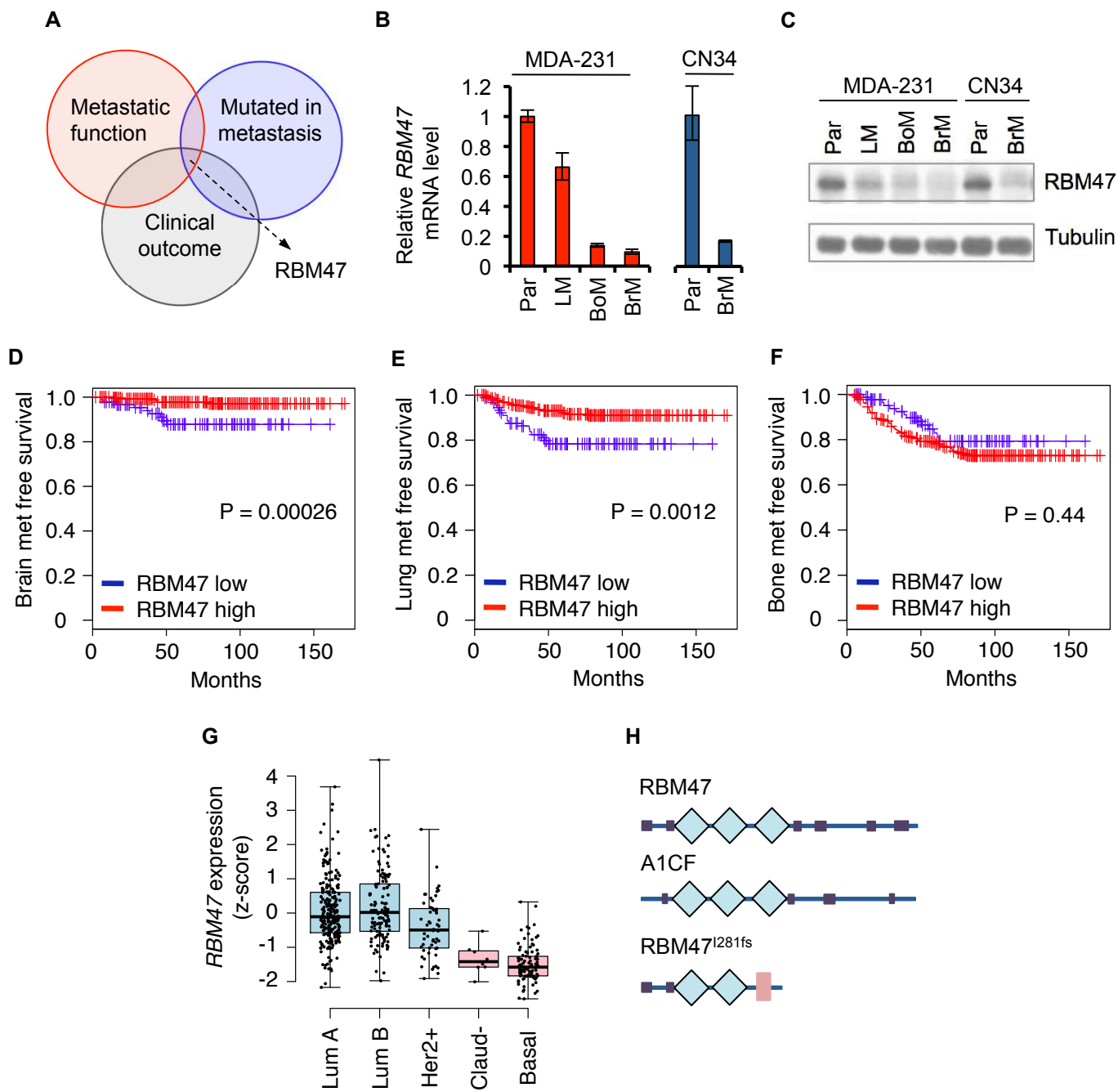
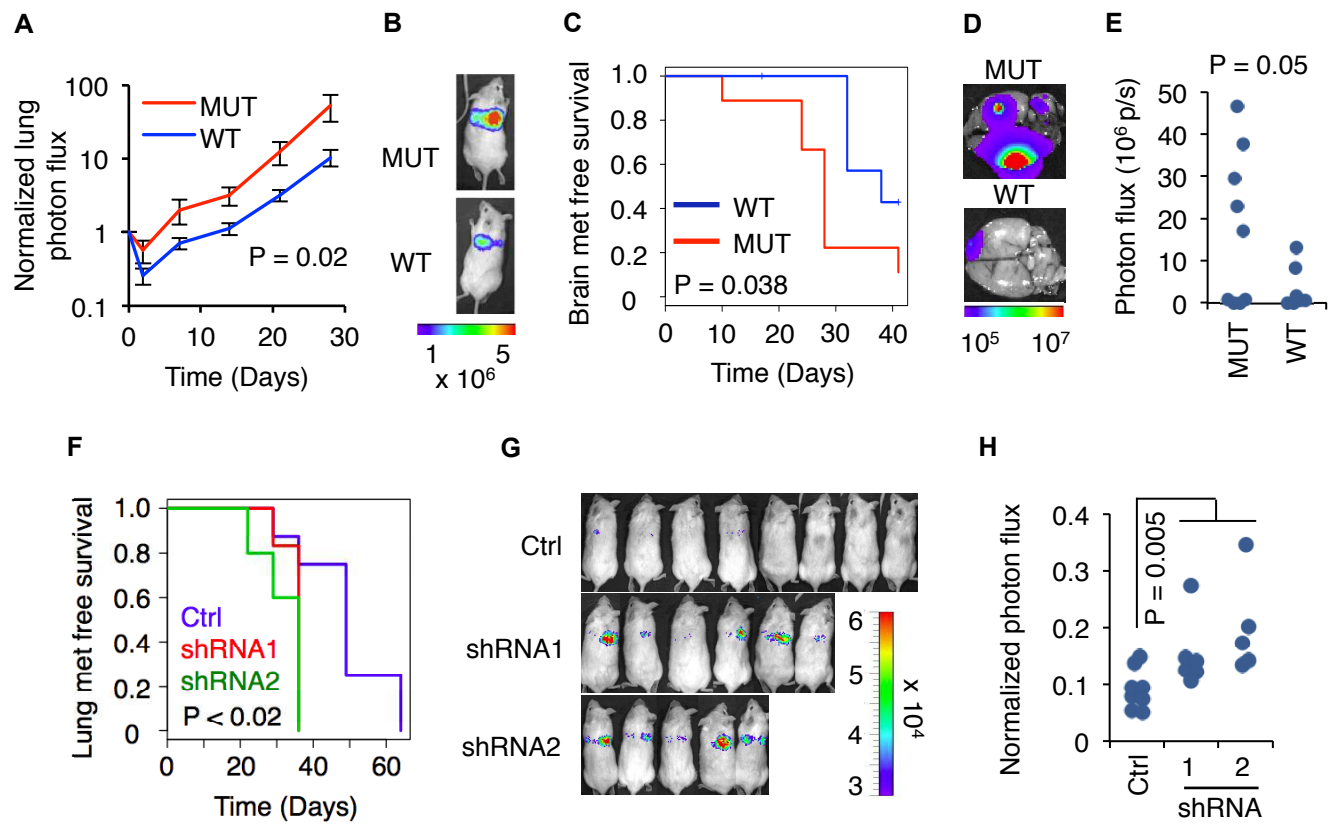


Figure 1. RBM47 expression is associated with breast cancer progression.



**Figure 2. RBM47 suppresses metastatic breast cancer progression.**

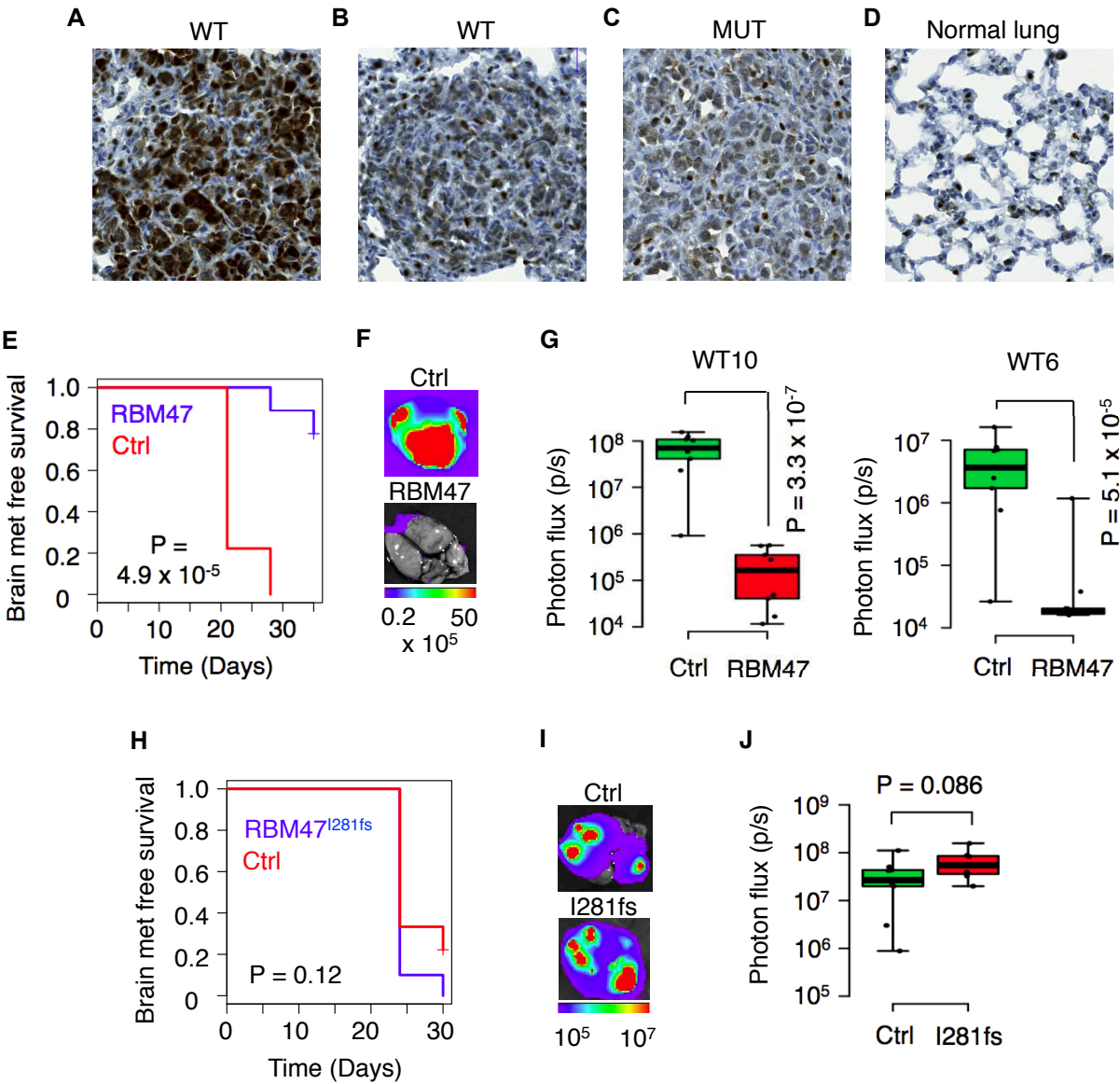


Figure 3. Clonal heterogeneity in RBM47 sensitivity.

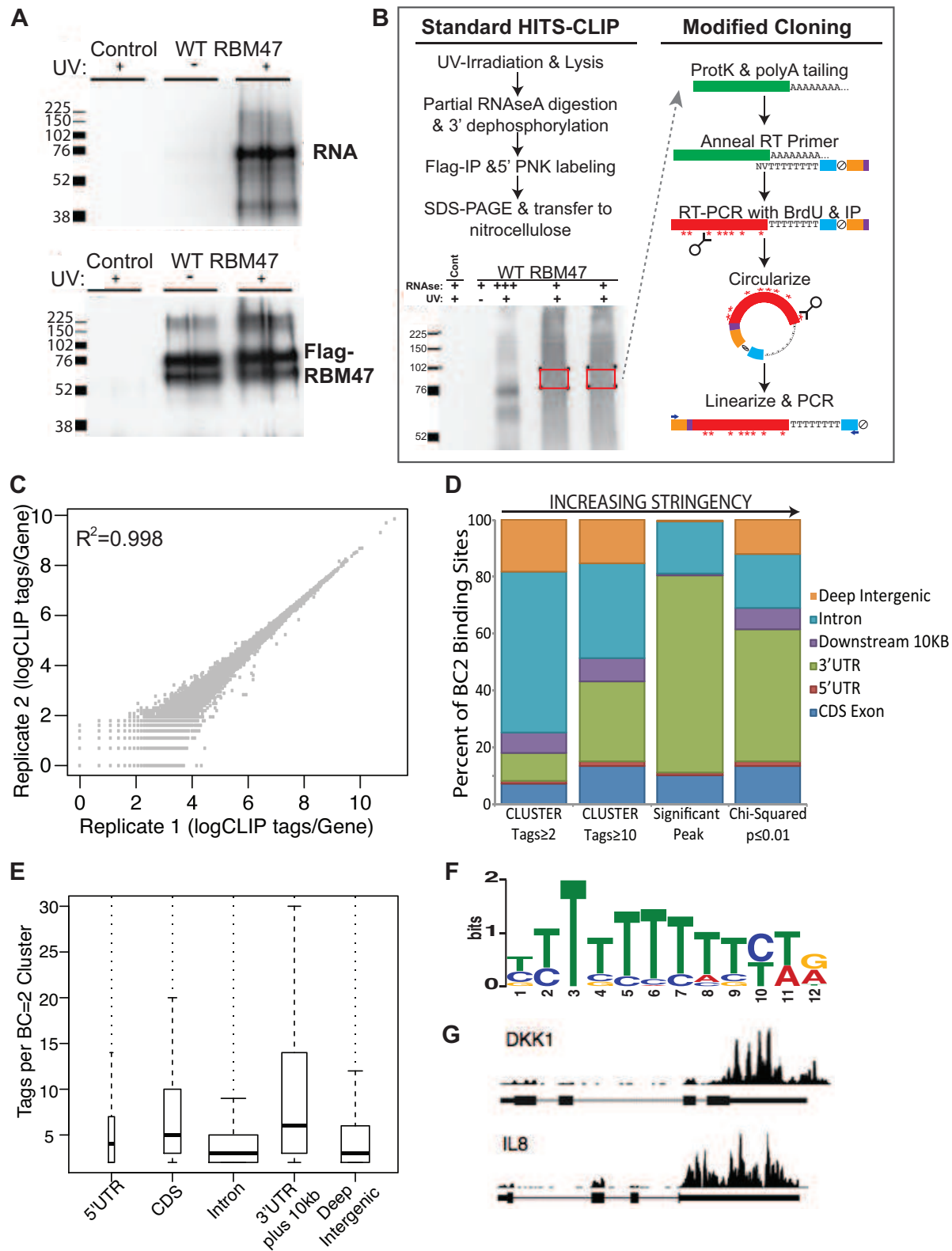


Figure 4. HITS-CLIP identifies genome-wide RBM47 binding maps.

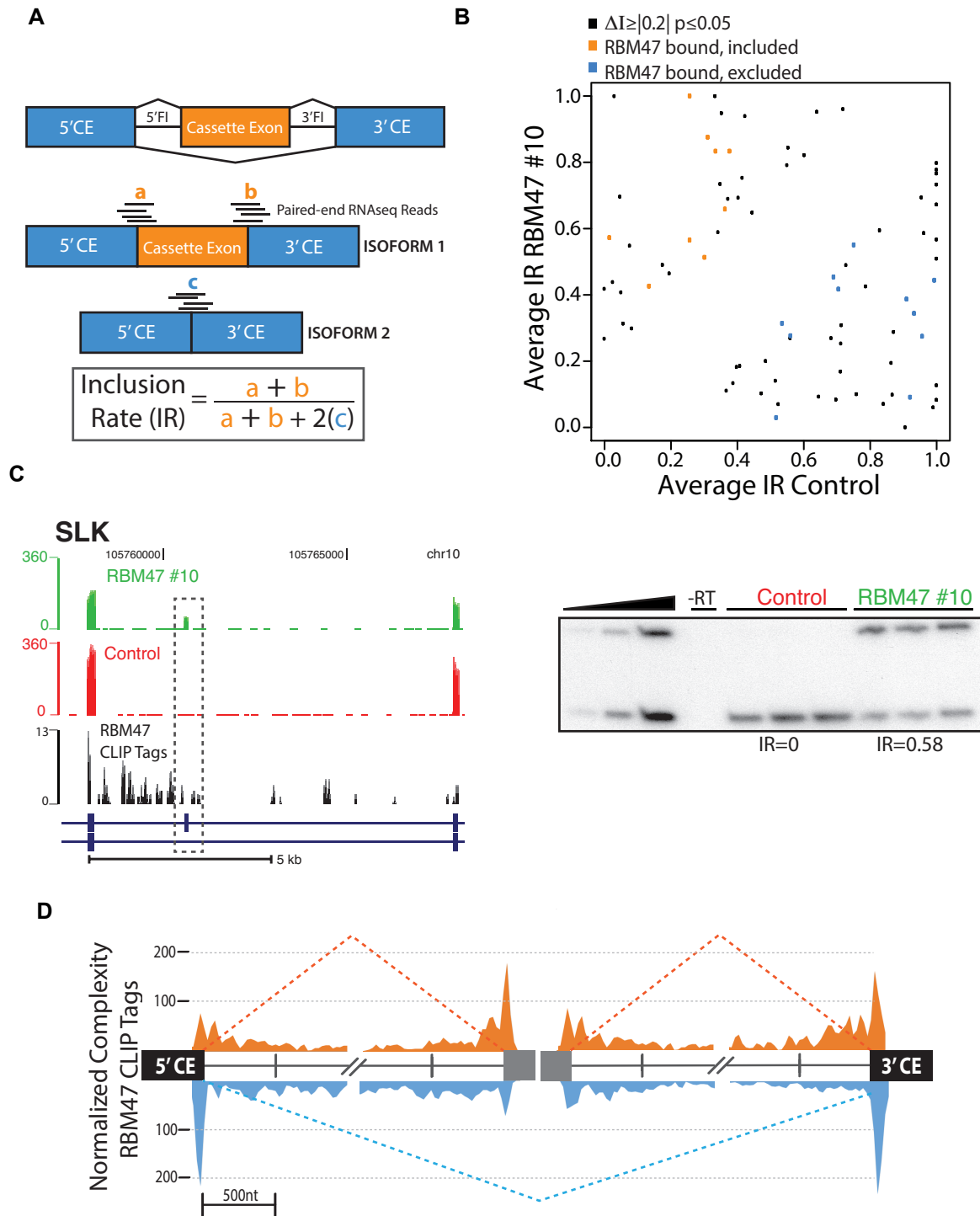


Figure 5. RBM47 regulates alternative splicing.

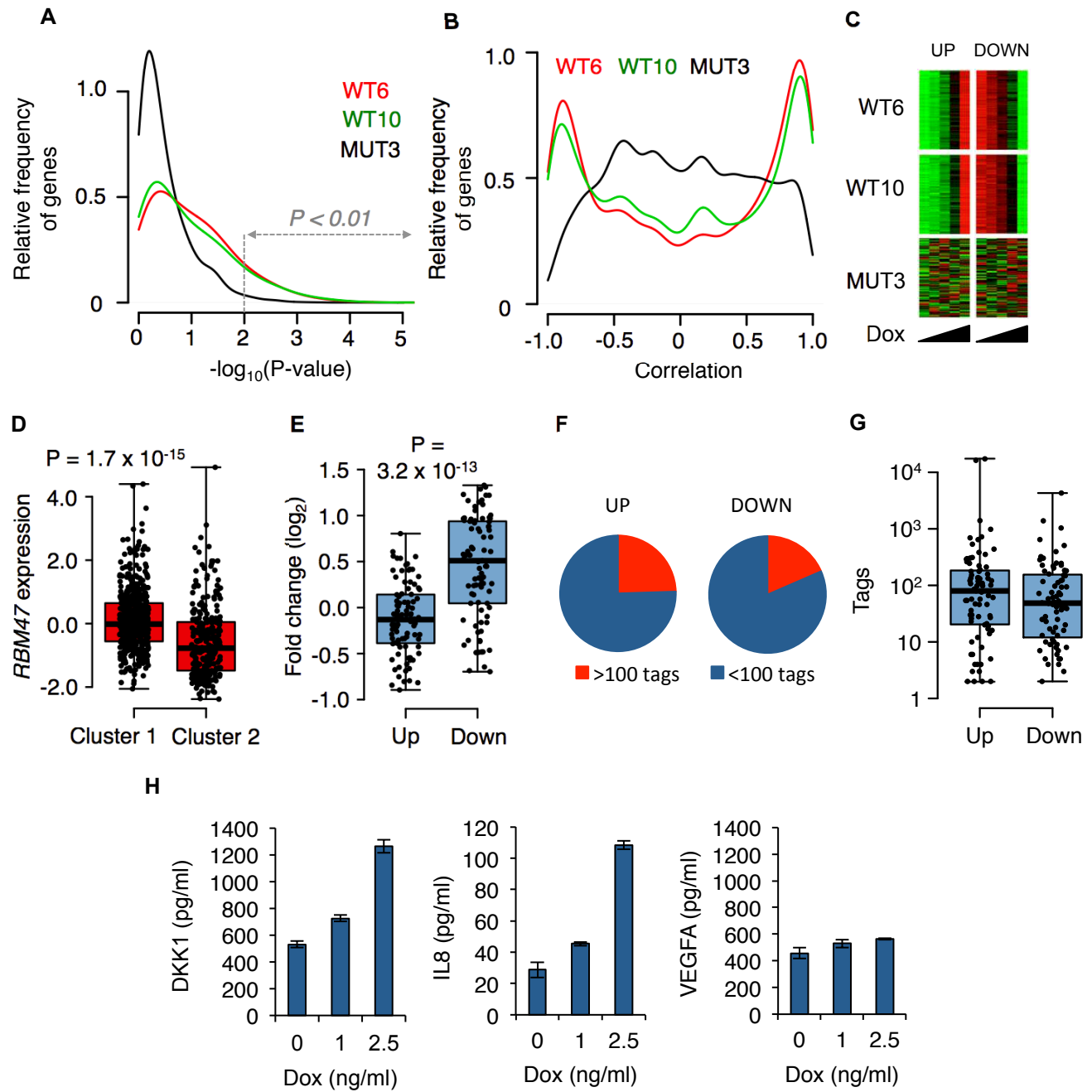


Figure 6. RBM47-induced changes in mRNA levels.



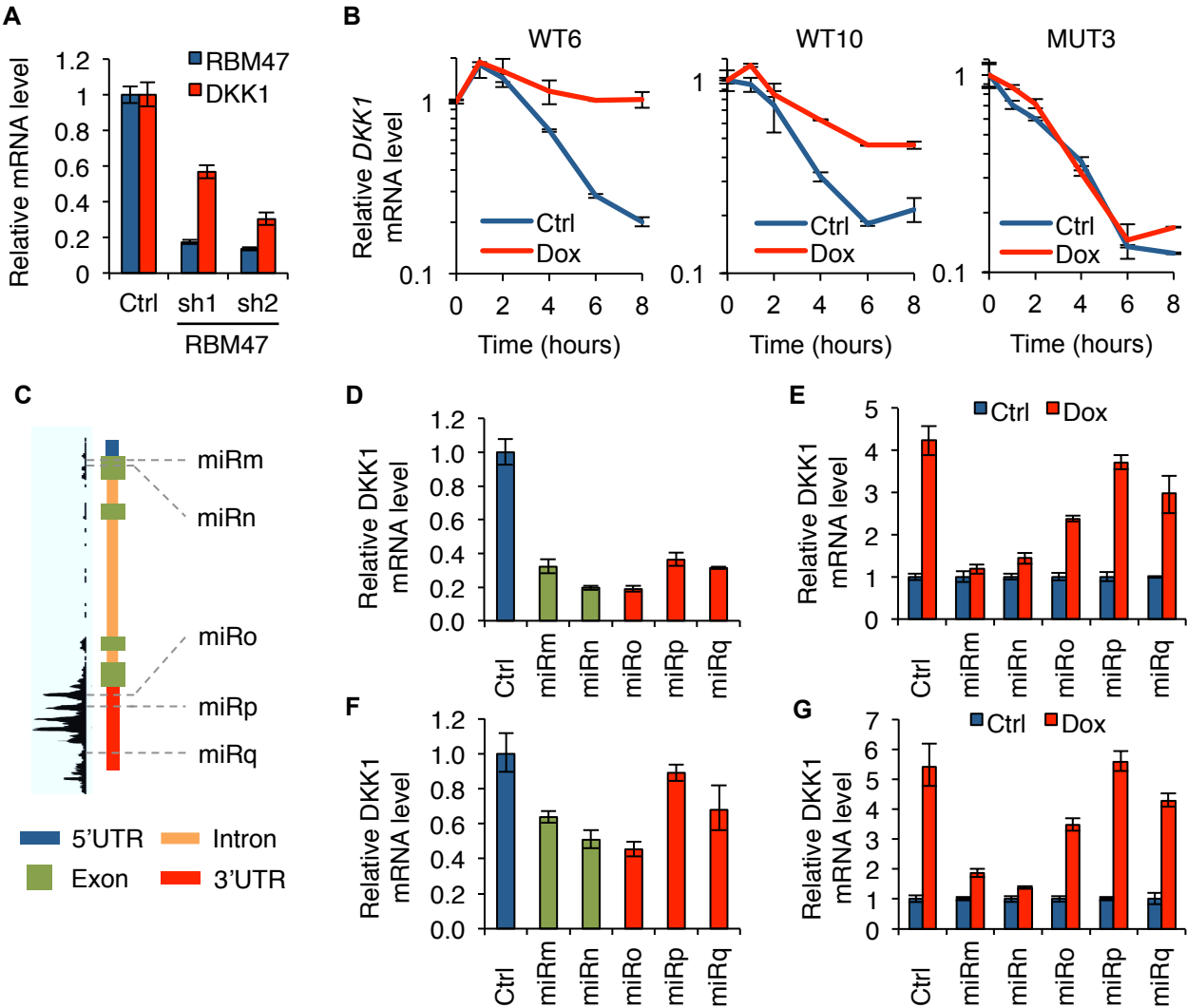
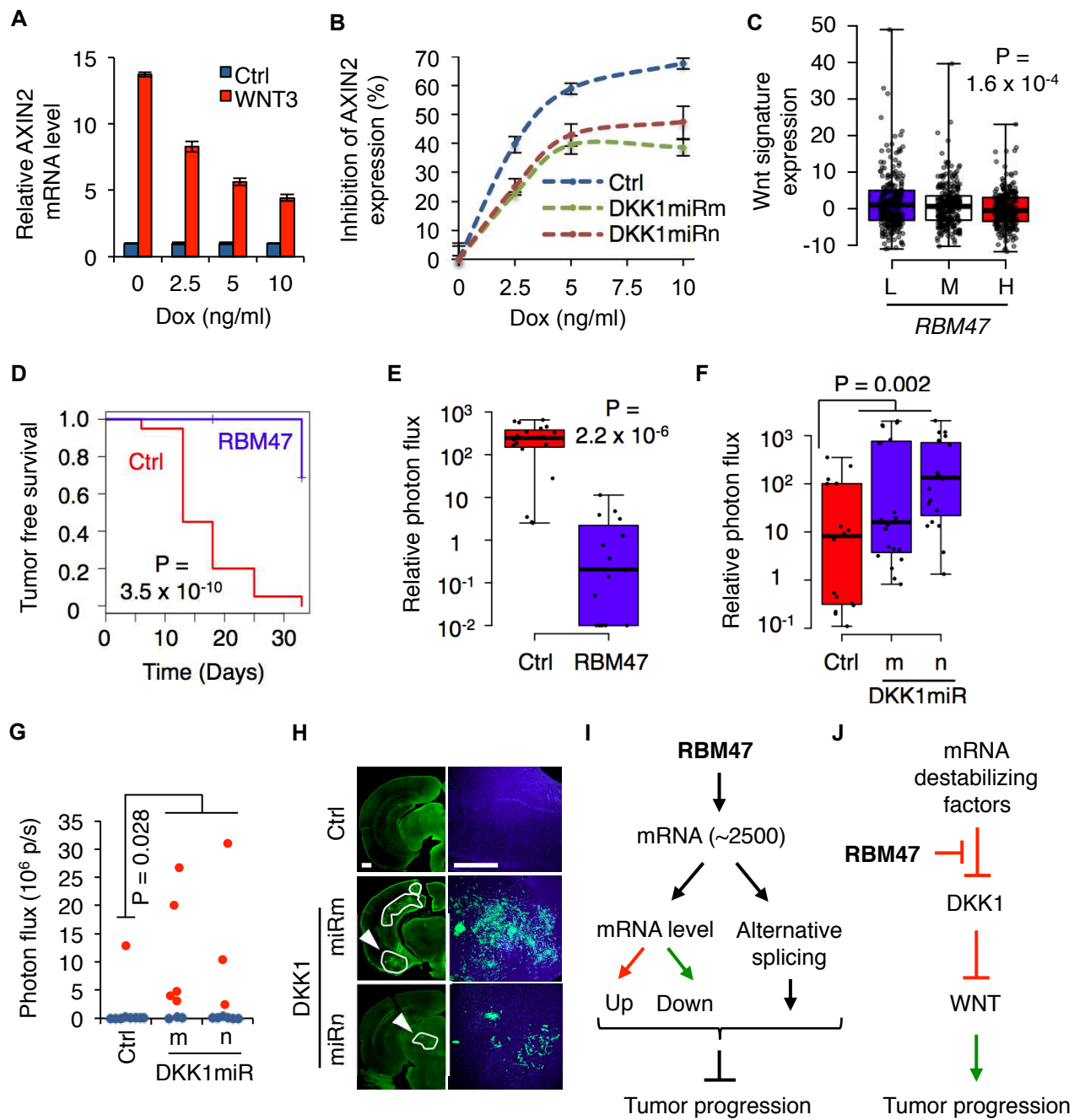


Figure 7. RBM47 modulates *DKK1* mRNA stability.



**Figure 8. RBM47 suppresses tumor progression via Wnt inhibition.**