Title
Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins

Journal
Nature, 513(7517)

ISSN
0028-0836

Authors
Goodarzi, Hani
Zhang, Steven
Buss, Colin G
et al.

Publication Date
2014-09-11

DOI
10.1038/nature13466

Peer reviewed

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2015 May 21.

Published in final edited form as:

Nature. 2014 September 11; 513(7517): 256–260. doi:10.1038/nature13466.

Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins

Hani Goodarzi[1], **Steven Zhang**[1], **Colin G. Buss**[1], **Lisa Fish**[1], **Saeed Tavazoie**[2,*], and **Sohail F. Tavazoie**[1,*]

[1]Laboratory of Systems Cancer Biology, Rockefeller University, 1230 York Avenue, New York, New York 10065, USA

[2]Department of Biochemistry and Molecular Biophysics, and Department of Systems Biology, Columbia University, New York, New York 10032, USA

Abstract

Aberrant regulation of RNA stability plays an important role in many disease states[1,2]. Deregulated post-transcriptional modulation, such as that governed by microRNAs targeting linear sequence elements in mRNAs, has been implicated in the progression of many cancer types[3-7]. A defining feature of RNA is its ability to fold into structures. However, the roles of structural mRNA elements in cancer progression remain unexplored. We performed an unbiased search for post-transcriptional modulators of mRNA stability in breast cancer by conducting whole-genome transcript stability measurements in poorly and highly metastatic isogenic breast cancer lines. Using a computational framework that searches RNA sequence and structure space[8], we discovered a family of GC-rich structural *cis*-regulatory RNA elements, termed sRSE for structural RNA stability element, that is significantly over-represented in transcripts displaying reduced stability in highly metastatic cells. By integrating computational and biochemical approaches, we identified TARBP2, a double-stranded RNA binding protein implicated in micro-RNA processing as the *trans* factor that binds the sRSE family and similar structural elements—collectively termed TARBP2-binding structural elements (TBSE)—in transcripts. TARBP2 is overexpressed in metastatic cells and metastatic human breast tumours and destabilizes transcripts containing TBSE instances. Endogenous TARBP2 promotes metastatic cell invasion and colonization by destabilizing amyloid precursor protein (*APP*) and *ZNF395* transcripts, two genes previously associated with Alzheimer's and Huntington's disease, respectively. We reveal these

[*]Corresponding authors: Sohail F. Tavazoie, Leon Hess Assistant Professor, Head, Laboratory of Systems Cancer Biology, Rockefeller University, Box 16, 1230 York Avenue, New York, NY 10065 USA, Phone: 212-327-7208 Fax: 212-327-7209, stavazoie@mail.rockefeller.edu, Saeed Tavazoie, Professor, Department of Biochemistry and Molecular Biophysics and Department of Systems Biology, Columbia University, New York, NY 10032., st2744@columbia.edu.

Author Contributions

H.G., S.T., and S.F.T. conceived the project and designed the experiments. S.F.T. and S.T. supervised all research. H.G., S.Z., S.T., and S.F.T. wrote the manuscript. H.G., S.Z., C.G.B., and L.F. performed the experiments. S.Z. performed cancer cell invasion and reporter assays. L.F. and H.G. performed the HITS-CLIP experiment. C.G.B. and H.G. performed lung colonization assays. H.G. performed the analyses.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

The authors declare no competing financial interests.

RNA-seq and microarray data have been deposited in the Gene Expression Omnibus database under accession number GSE49649.

genes to be novel metastasis suppressor genes in breast cancer. The cleavage product of APP, extracellular α-amyloid peptide, directly suppresses invasion while ZNF395 transcriptionally represses a pro-metastatic gene expression program. The expression levels of *TARBP2*, *APP*, and *ZNF395* in human breast carcinomas support their experimentally uncovered roles in metastasis. Our findings establish a non-canonical and direct role for TARBP2 in mammalian gene expression regulation and reveal that regulated RNA destabilization through protein-mediated binding of mRNA structural elements can govern cancer progression.

Gene expression studies, in principle, measure steady-state transcript levels. However, such measurements obscure the role of dynamic post-transcriptional programs, from splicing to nuclear export to transcript stability[9]. In order to study the dynamics of the RNA life-cycle in cancer, we isolated transcript stability from other aspects of RNA regulation. We used a non-invasive method—based on 4-thiouridine labeling and capture[8,10] followed by high-throughput sequencing—to determine the decay rates for roughly 13,000 transcripts expressed by the parental MDA-MB-231 (MDA) breast cancer cell line and its *in vivo*-selected highly metastatic MDA-LM2 sub-line (biological quadruplicates spanning 4 time-points composed of 32 total samples). A *t*-test derived statistic (*t*-score), ranging from -1 (more stable in parental MDA) to 1 (more stable in highly metastatic MDA-LM2), was used as a measure for differential decay rates between the two cell lines. We then employed a mutual-information (MI) based computational approach[8] to identify the *cis*-regulatory elements that may mediate the differences in transcript stability observed between the metastatic MDA-LM2 cell line and its parental MDA line. We discovered a family of structural RNA stability elements, termed sRSE, embedded in transcripts that displayed reduced stability in highly metastatic cells (Figure 1a and Extended Data Fig. 1a). As we show below, this broad family of GC-rich stem-loops, with a median stem length of 9bp and median loop length of 7nt, is bound by a common *trans* factor. Consistent with their higher decay rates in metastatic cells, sRSE-carrying transcripts displayed significantly reduced steady-state expression in metastatic MDA-LM2 cells relative to less metastatic MDA parental cells (Figure 1b and Extended Data Fig. 1b). Moreover, the significantly correlated expression of these transcripts in three independent human gene-expression datasets raised the possibility of their co-regulation through a common post-transcriptional pathway mediated by this structural element (Extended Data Fig. 2a-c).

To directly assess the transcriptome-wide functionality of the identified sRSE instances, we performed an in-culture cellular titration experiment in which synthetic RNA oligonucleotides harboring tandem instances of sRSE1, the most informative representative member of the sRSE family (Extended Data Fig. 1a), were used as intracellular decoys that would bind the putative *trans* factor, preventing it from targeting endogenous transcripts[11] (Extended Data Fig. 2d). Consistent with our hypothesis, the expression levels of endogenous sRSE-carrying transcripts were significantly up-regulated in cells transfected with synthetic decoys relative to their levels in cells transfected with scrambled controls (Figure 1c and Extended Data Fig. 2e). These findings suggest that the sRSE-binding *trans* factor, when competitively titrated by the decoy, promotes transcript destabilization.

We then chose an sRSE instance, matching the motif definition of sRSE1 on a differentially destabilized transcript, for further analysis. The secondary structure of this element, determined *in silico* (M-fold[12]) and *in vitro* through differential S1/V1 nuclease digestion sequence analysis[13], matches that of the sRSE motif (Extended Data Fig. 3a). Additionally, mCherry reporter constructs carrying this element and its modified versions in their 3' untranslated regions (UTR) were used to test its functionality and establish the necessity of its underlying stem-loop structure (Extended Data Fig. 3b-c). We compared mCherry-encoding transcripts (using GFP as internal control) carrying four different forms of the structural element in their 3' UTR: an sRSE1 versus scrambled pair to reveal whether the element has a functional role in transcript stability and expression, and a structured versus unstructured mimetic pair to establish if its secondary structure is essential for its functionality (Figure 1d). This analysis revealed that this sRSE instance is sufficient for suppression of expression and that its structure—not simply its sequence—is the key regulatory determinant.

We next sought to identify the sRSE-binding *trans* factor by computationally searching for candidate RNA-binding proteins (RBPs) whose expression levels correlated with sRSE-carrying transcripts across breast cancer gene expression profiles[14]. Using this approach, we identified three candidate RBPs, namely TARBP2, HEXIM1, and PPRC1, as potential post-transcriptional regulators of this regulon based on their correlated expression (Extended Data Fig. 4a). RNAi-mediated knock-down followed by transcriptomic profiling revealed that silencing one of these RNA binding proteins, TARBP2, yielded a significant up-regulation of sRSE-carrying transcripts (Figure 2a and Extended Data Fig. 4b). A similar up-regulation was observed upon *TARBP2* knockdown in CN-LM1a cells, an *in vivo* selected highly metastatic breast cancer line derived from another patient's breast tumor (CN34)[6], and 293T kidney epithelial cells, suggesting a more general and physiological regulatory link between TARBP2 and sRSE-carrying transcripts (Extended Data Fig. 4c-d). Importantly, the enhanced expression of this regulon was concomitant with enhanced transcript stability upon α-amanitin treatment (Figure 2b and Extended Data Fig. 4e). Conversely, over-expression of TARBP2 in MDA-231 parental cells resulted in a significant down-regulation of sRSE-carrying transcripts (Figure 2c and Extended Data Fig. 4f). Moreover, the gene expression profile resulting from *TARBP2* knock-down in metastatic cells was significantly correlated to that obtained from in-culture cellular titration, consistent with the titration of TARBP2 upon sRSE-decoy RNA transfection (Extended Data Fig. 4g). TARBP2, first identified based on its ability to bind the double-stranded stem portions of HIV's TAR RNA[15], was subsequently found (through gel mobility-shift assays of synthetic RNA variants) to prefer double-stranded RNAs (*e.g.* stem loops) with high GC content[16]—closely resembling sRSE. TARBP2 was later found to play a physiological role in the cell by binding microRNA hairpin precursors on their path to maturation[17]. To biochemically assess direct binding of the sRSE by TARBP2, we performed UV cross-linking and coimmunoprecipitation (HITS-CLIP[18]) of tagged TARBP2 and found a significant enrichment of sRSE among the TARBP2 co-immunoprecipitated mRNAs (Figure 2d and Extended Data Fig. 5a-e). More importantly, the TARBP2 binding sites determined using HITS-CLIP enabled us to go beyond sRSE and provide a direct experimental description of the broader underlying structural elements, which we collectively termed TBSE (TARBP2-

binding structural elements). TARBP2 binds both exonic and intronic TBSE elements with a preference towards intronic sites. TARBP2 binding elements showed a high GC-content along with a higher propensity to form secondary structures than expected by chance (Extended Data Fig. 5f-g). Moreover, transcript measurements from mCherry-fused sRSE/scrambled reporters showed that the sRSE-dependent transcript down-regulation was abrogated in the setting of TARBP2 knock-down (Figure 2e and Extended Data Fig. 5h). More importantly, while we initially used sRSE as a proxy to study the behavior of TARBP2 regulon in metastatic cells, TARBP2 HITS-CLIP provided us with the set of transcripts that are bound by TARBP2 *in vivo* that can be studied directly. These TARBP2-bound transcripts showed significant enrichment amongst the differentially destabilized transcripts in MDA-LM2 cells; they were also significantly up-regulated and stabilized in the context of TARBP2 knock-down and were down-regulated in cells over-expressing TARBP2 (Extended Data Fig. 5i-j). *In vitro* binding assays using purified tagged-TARBP2 and short oligo-ribonucleotides also supported our *in vivo* observations: (i) TARBP2 directly interacts with sRSE1-element or its structured mimetic but minimally interacts with scrambled and unstructured variants (Extended Data Fig. 6a-c); (ii) TARBP2-mediated co-precipitation of a large randomized RNA population followed by high-throughput sequencing resulted in an enrichment for GC-rich structured RNA variants closely resembling sRSE elements (Extended Data Fig. 6d-h). Taken together, our findings reveal that TARBP2 binds TBSE—a family of GC-rich apical and internal stem-loop elements within endogenous transcripts—and negatively regulates their stability and expression *in vivo* (Extended Data Fig. 6i-k).

Consistent with the observed down-regulation of TARBP2-bound transcripts in metastatic MDA-LM2 cells, TARBP2 transcript and protein levels were found to be expressed at higher levels in MDA-LM2 and CN-LM1a cells relative to their parental lines (Figure 2f and Extended Data Fig. 7a). The increased expression of TARBP2 in multiple patients' metastatic derivative lines suggested that it may play a role in metastatic progression. We also observed that primary human breast tumours that metastasized (stage IV) displayed significantly higher expression of TARBP2 relative to early-stage (stages I and II) tumours, which display low metastasis rates (Extended Data Fig. 7b).

Consistent with these clinical associations, TARBP2 knock-down significantly reduced lung metastatic colonization in CN-LM1a (~7-fold) and MDA-LM2 (~2.5-fold) cell lines as assessed by tail-vein lung colonization assays (Figure 3a and Extended Data Fig. 7c). We also observed a significant reduction in the number of metastatic nodules in lungs of mice injected with TARBP2 knock-down cells (Figure 3b and Extended Data Fig. 7d). These effects on metastasis were not due to enhanced growth as the *in vitro* proliferation rate of cells was not significantly reduced upon TARBP2 knock-down (Extended Data Fig. 7e). TARBP2 knock-down also reduced invasiveness (~2-fold)—a phenotype required for efficient metastasis (Figure 3c and Extended Data Fig. 7f). TARBP2 silencing did not reduce primary tumour growth in cells growing in the mammary fat pads of mice (Figure 3d). TARBP2 depletion did, however, significantly reduce orthotopic metastasis to the lungs (Figure 3e and Extended Data Fig. 7g). Our findings, taken together, reveal a role for

endogenous TARBP2 as a promoter of metastatic cell invasion and colonization in breast cancer.

To identify the metastasis suppressor genes that are post-transcriptionally repressed by TARBP2, we performed an unbiased analysis of the gene-expression and stability profiles of TARBP2 knock-down cells to find transcripts that are directly bound by TARBP2 and are sensitive to modulations in TARBP2 levels. We identified four transcripts that were directly bound by TARBP2 *in vivo* (see Extended Data Fig. 8a) as well as substantially destabilized and down-regulated in highly metastatic MDA-LM2 and CN-LM1a cells (validated through qRT-PCR), which express higher levels of TARBP2 than their isogenic parental lines (Extended Data Fig. 8b-d). These genes consisted of amyloid precursor protein (*APP*), zinc finger protein 395 (*ZNF395*), the adaptor-related protein complex 2 (*AP2A2*), and laminin beta 1 (*LAMB1*). Importantly, these transcripts displayed both higher stability and abundance in the context of TARBP2 knock-down (Figure 4a and Extended Data Fig. 8e-f). High-grade tumours, which display higher relapse rates, from three independent gene-expression compendia (N=821)[19] exhibited a significantly lower aggregate expression of these four genes relative to low-grades tumours (Figure 4b). Additionally, patients whose tumours displayed a reduced aggregate expression for these genes experienced significantly lower overall metastasis-free survival compared to those whose tumours expressed higher aggregate expression for these genes (Figure 4c). These clinical associations and their inverse relationship to *TARBP2* transcript expression further highlight these genes' potential roles as suppressors of metastatic progression in human breast cancer and provide additional support for TARBP2 as their upstream regulator.

In order to identify the primary modulators of metastasis suppression among these four targets, we performed functional epistasis experiments, in which each of these genes was stably silenced in the background of *TARBP2* knock-down. In both breast cancer lines, silencing *APP* or *ZNF395* significantly enhanced metastatic colonization (Figure 4d and Extended Data Fig. 9a-c) as well as cellular invasion by cells depleted of TARBP2 (Figure 4e and Extended Data Fig. 9d-e). APP—previously implicated in Alzheimer's disease—is a membrane protein that is proteolytically cleaved to yield soluble products (*i.e*. soluble-APPα)[20], while ZNF395 is a poorly characterized transcription factor involved in the transcriptional activation of Huntington's disease gene Huntingtin[21]. Consistent with these functional findings, patients whose primary breast tumours displayed reduced expression levels of *APP* and *ZNF395* experienced significantly lower survival rates than those whose tumours displayed higher levels of these genes in several independent clinical cohorts (p<1e-6, Extended Data Fig. 9f). These findings reveal *ZNF395* and *APP* to be downstream functional targets of TARBP2 and to associate with survival of breast cancer patients.

To determine if amyloid peptide products could directly mediate the suppressive effects of APP mRNA on invasion, we added soluble APPα and β to MDA-LM2 cells and quantified their invasive capacity. Interestingly, only APPα treatment reduced the invasiveness of cancer cells (Figure 4f and Extended Data Fig. 10a). Consistent with our findings, APP has also been previously implicated in ovarian cancer cell invasion[22]. Our similar observations in breast cancer metastasis suggest a more general role for *APP* as a suppressor of cancer

invasion as well as a suppressor of metastatic progression by a metastatic cancer type—breast cancer.

On the other hand, transcriptomic profiling of ZNF395 knock-down cells revealed up-regulation of a number of established promoters of cancer progression (Extended Data Fig. 10b). Among the most highly up-regulated genes, we identified and validated by qRT-PCR several known promoters of metastatic colonization and cell invasion, including *IL8*[23] (~13-fold), *IL1B*[24] (~40-fold) and *COX2*[25] (~20-fold; Extended Data Fig. 10c). Given their well-established roles in cancer progression and metastasis, we hypothesize that *ZNF395* silencing enhances metastasis in part through de-repression of these genes.

TARBP2, a component of RISC-loading complex (RLC)[17,26], has been previously implicated as a modest suppressor of tumor growth in Ewing's sarcoma through its role in miRNA maturation[27]. In contrast, our findings reveal TARBP2 to be a robust promoter of metastasis in breast cancer. Furthermore, the TARBP2-mediated post-transcriptional regulation of *APP* and *ZNF395*, and TARBP2-bound transcripts in general, proved to be independent of microRNA-mediated regulation as *DICER* knock-down did not increase transcript levels of these genes (Extended Data Fig. 10d-f). More importantly, neither *DICER* silencing nor its genetic inactivation in *DICER*$^{-/-}$ murine cells precluded TARBP2-mediated suppression of TARBP2-bound transcripts and *APP* and *ZNF395* in particular, consistent with this network constituting a TARBP2-dependent and DICER-independent pathway (Extended Data Fig. 10g-i). In summary, our findings establish a novel post-transcriptional regulatory network whereby TARBP2 binding of structural hairpins contained in specific metastasis suppressor transcripts leads to their destabilization (Figure 4g). We reveal that structural information contained in transcripts can govern cancer progression by acting as binding sites for a destabilizing RNA-binding *trans* factor. We also implicate the *TARBP2/APP/ZNF395* pathway in metastatic progression through loss-of-function, epistasis, and clinical pathologic correlation analyses. Additionally, our findings reveal a miRNA-independent role for TARBP2 in gene expression regulation through mRNA destabilization. While we have focused on transcript stability, several other aspects of the RNA life-cycle, such as alternative splicing and RNA localization, are impacted by structural RNA elements and may be systematically studied in the context of cancer progression using a similar approach. We speculate that small molecule or oligonucleotide-based interference with interactions between such *trans* factors and structural elements in transcripts may constitute novel routes towards therapeutic modulation of disease states.

Methods

Cell culture

CN34 and MDA-MB-231 cells and their respective sub-lines, CN-LM1a and MDA-LM2, were propagated as previously described[6]. HEK293T cells were cultured in DMEM-based media supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin and fungizone. RNAi and DNA transfections were performed using Lipofectamine 2000 (Invitrogen) and TransIT-293 (Mirus), respectively. Cells were regularly tested for mycoplasma contamination.

RNA decay rate measurements

4-Thio-uridine labeling and RNA preparation were performed as described elsewhere[8] for 0, 2, 4, and 7hr time-points. The collection was performed in biological quadruplicate for each cell line (32 samples total). Samples were barcoded (4-fold multiplexed) for sequencing using ScriptSeq RNA-seq Library Preparation Kit (Epicentre) and sequencing-by-synthesis was performed using the HiSeq 2000 sequencer (Illumina). Read alignment to human transcriptome (build hg19) and quantification were performed using TopHat[28] and Cufflinks[29], respectively. For each transcript, the reported RPKM values were first corrected based on the amount of labeled RNA recovered at each time-point and were then used to estimate decay rates ($r = -\ln(S_t/S_0)/t$)[8], which resulted in four decay rate measurements (biological quadruplicates) in MDA parental cells and four in MDA-LM2 cells. These values were then used to calculate the probability of a difference between the two cell lines (Student's *t*-test). The calculated probabilities were subsequently transformed into a directional statistic for all transcripts (i.e. *t*-score = $s(1-p)$, where s is the direction of the change), ranging from -1 (significantly stable in MDA-231) and 1 (significantly stable in MDA-LM2). For quantized analyses, a *t*-score<-0.8 threshold (~10% of transcripts) was used to mark the set of transcripts destabilized in MDA-LM2 cells.

Discovering the structural *cis*-regulatory elements

Post-transcriptional *cis*-regulatory elements rely on contributions from the local secondary structure as well as their underlying sequence to interact with their binding partners[8,30-34]. We recently introduced a computational framework called TEISER which captures both structural and sequence information contained in RNA sequences[8]. While regulatory elements are enriched in the un-translated regions (UTRs), many examples of RNA elements fall within the boundaries of the coding sequence. To employ an unbiased search of the transcript sequence space, we employed TEISER to discover structural regulatory elements across mature transcripts (along the coding sequences as well as the 5' and 3' UTRs) that were informative of differential transcript stability measurements.

We identified a large number of GC-rich structural elements that were enriched in the transcripts destabilized in MDA-LM2 cells. Upon their identification by TEISER, informative structural elements were grouped based on their similarity (defined based on conditional information ratio[35]) and were found to be highly similar with conditional information ratio of less than two relative to the element with the highest mutual information value. Accordingly, TEISER classified these elements into a single family of GC-rich structural elements, named sRSE for structural RNA stability element. The element with the highest MI value (sRSE1, Extended Data Fig. 1) was subsequently used as a representative of the entire family for all transcriptome-wide analyses (in Extended Data Fig. 1, we have included sRSE1 as well as a few members of this family sorted based on their similarity to sRSE1). In order to represent the entire family, which contains 318 distinct elements, we extracted all instances of all elements across the transcriptome. We then used RNAfold[36] to determine the secondary structure of the extracted sequences along with 20bp flanking sequences. Sequences whose *in silico* secondary structures were capped with their corresponding structural elements with a low folding free-energy (one standard deviation below average folding energy of all extracted sequences) were labeled as sRSE instances

and the host transcripts (carrying at least one such instance) were identified as the sRSE-carrying transcripts or the sRSE regulon. The extracted sRSE instances were also used to calculate the GC-content at select positions across the stem and the loop (4bp of stem and 4nt of loop) from which a graphical representation of the entire family across the transcriptome was created (*e.g*. Figure 1a). TEISER in non-discovery mode was used to show enrichment/depletion patterns of the sRSE-regulon in various types of data (*e.g*. gene expression profiles, in-culture cellular titration experiments, knock-down profiles, and stability measurements)[8].

Hypergeometric distribution-based enrichment scores were defined as described previously[8]. To ensure that the behavior of the entire family matches that of representative members, we have also included results in which sRSE1 was used as the most informative member in our analyses.

In-culture sRSE-decoy cellular titration experiment

Decoy and scrambled sequences (Extended Data Fig. 3c) were synthesized *in vitro* (IDT) and PCR amplified using forward-T7 linked primers. Capped and polyadenylated RNAs were generated using *in vitro* transcription (mMESSAGE mMACHINE, Ambion; poly-A polymerse, NEB). MDA-LM2 cells were then transfected with decoy and scrambled RNA molecules, in biological replicates, as detailed previously[8]. Forty-eight hours post-transfection, total RNA was collected from each sample (Norgen Biotek), labeled (TargetAmp-Nano Labeling Kit, Epicentre), and hybridized to human microarray slides (HumanHT-12, Illumina). The intensities were measured using GenomeStudio 1.0.6 (Illumina), and the Lumi package in R was used to transform and normalize signal intensities. The logFC between average decoy and scrambled signal was used as a measure of differential expression for each transcript. For quantized analyses, a logFC>0.2 threshold was used to mark the set of up-regulated transcripts.

In vitro probing of RNA secondary structure

We used a medium-throughput approach, based on PARS[37], to probe the secondary structure of the given sRSE decoy sequence. Upon digestion and precipitation[37] of 1μg of sRSE RNA, we used poly-A polymerase (NEB) for 15 min at 37 °C to add a poly-A tail to RNA fragments. We then used an oligo-dT primer linked to a KpnI site to make the first-strand cDNA (Superscript III, Invitrogen). A KpnI-linked 5’ specific primer along with a reverse primer matching the reverse transcriptase primer was used to PCR amplify the cDNA fragments with intact 5’ ends. The PCR product was then digested with KpnI and ligated (for concatenation) before cloning into KpnI digested pUC19 vector. Colonies were then randomly picked for each nuclease, and the 3’ digestion sites were mapped through Sanger sequencing.

mCherry/GFP reporter system

A Gateway (Invitrogen) destination vector was designed by cloning an R1-R2 gateway cloning site in the 3’ UTR of mCherry in the pcDNA5/FRT backbone[38] (Invitrogen). A bi-directional promoter (pBI-CMV, Clontech) drives both mCherry and GFP (internal control) transcription (Extended Data Fig. 4b). Oligonucleotides carrying different variants were

PCR-amplified using primers linked to B1 and B2 sequences. BP Clonase enzyme mix (Invitrogen) was used to clone PCR products into pDONR-221 vector. LR Clonase (Invitrogen) was subsequently used to subclone the variants into the destination reporter construct. The reporters were then transiently transfected into MDA-LM2 cells and subjected to RNA extraction (Norgen Biotek RNA extraction kit), cDNA synthesis and qRT-PCR for both mCherry and GFP samples 48hr post-transfection as detailed previously[8]. Cells were seeded and transfected in biological replicates, and qPCR was performed in technical replicates. Technical replicates were averaged prior to statistical comparisons (performed using a one-tailed Mann Whitney test).

Identifying potential sRSE-binding proteins

We used a compendium of breast cancer gene expression profiles[14] to probe correlations between sRSE-carrying transcripts and putative RNA binding proteins. First, we focused on sRSE-carrying transcripts that were also destabilized in MDA-LM2 cells (differential decay rate score of less than -0.8) to enrich for functional sRSE instances. Using a “centroid linkage” k-means method, we then clustered the matrix of expression values across these transcripts to identify clusters with low intra-cluster and high inter-cluster distance as a core module. We then correlated the expression levels of putative/validated double-stranded RNA binding protein to those of the identified cluster both individually as also as an aggregate (average of variance normalized values). The list of RNA-binding proteins with a significantly non-zero distribution of correlations ($p < 0.01$) and significant correlations with the aggregate profile was further narrowed down by selecting the candidates that displayed expression changes in the metastatic lines relative to their parental cells. Using these criteria, we identified TARBP2, HEXIM1, and PPRC1 as potential candidates that bind and regulate sRSE-carrying transcripts (Extended Data Fig. 4a). We should highlight that this short list of candidate RBPs was not sensitive to the choice of regulon, method or dataset. For example, a similar analysis performed for sRSE1-carrying transcripts, using the same criteria as above, resulted in the identification of TARBP2, HEXIM1, and PPRC1 as potential candidates that bind and regulate sRSE1-carrying transcripts.

Gene expression and stability measurements for RNAi-mediated knock-down and TARBP2 over-expression

For TARBP2, the following siRNAs were used (anti-sense strand):

si1. UACUGAAUUUCGUAGAGAAUCCCAGGU

si2. GCAGCAUAUUUAUUCCAGGCUCUAGUA

si3. GCUGAGGCUCUGUGGUCCUUGGCUCCU

Sequences 2 and 3 were used for MDA-LM2 and 1 and 2 for CN-LM1a. For HEXIM1 we used: UGAUUUAUUGAUUUGUGUAUGCUUCUU and GUCCUCUUCAGUUUGUAAUUCCUGCAG. For PPRC1, we used: UUCCACACAGGCUGCUGGCUGCUCCUU and AGAACUAGUCUCUAGCGAGUCAUCCAG. The siRNAs were transfected as described previously[8]. Total RNA was extracted (Norgen Biotek RNA extraction kit) 48hr post-transfection, and sufficient knock-down (>5-fold) was tested using qRT-PCR[8]. The samples

were then labeled and hybridized as described above. The Lumi package in R was used to transform and normalize signal intensities. Each siRNA was counted as a biological replicate (two biological replicates per target gene) and BLOCK-iT Fluorescent Oligo (IDT), used as the control (siControl), was transfected in biological replicate and prepared in parallel. The logFC between average siRNA and siControl signal was used as a measure of differential expression for each transcript. For quantized analyses, a logFC>0.2 threshold was used to mark the set of up-regulated transcripts.

Transcript stability was measured for siTARBP2 and siControl samples using α-amanitin (Sigma) as an inhibitor of RNA polymerase II[8]. Briefly, RNA from α-amanitin-treated cells was collected at 0 and 18hr and hybridized (as described above). The normalized transcript levels were then corrected based on the amount of labeled RNA achieved per unit of input total RNA. The ratio between average abundance (averaged between biological replicates) at 18 and 0 hour samples was used as a measure of stability for each sample (all transfections were performed in biological replicates). The log-ratio between siTARBP2 and siControl samples was used to identify the transcripts that are stabilized in TARBP2 depleted cells. A cut-off of 0.32, which corresponds to ~10% of all transcripts, was used to mark the stabilized transcripts (Figure 2b). This step also ensures that the identified transcripts were not by-products of the 4sU methodology and can be recapitulated using RNA polymerase inhibitors.

For transient *TARBP2* over-expression, we used a pCMV6-TARBP2 vector transfected into MDA-LM2 cells along with a pCMV6-GFP vector as control. Total RNA was extracted 72hr post-transfection, labeled, and hybridized as described above. The logFC between average pCMV6-TARBP2 and pCMV6-GFP signal was used as a measure of differential expression for each transcript. A logFC<-0.2 threshold was used to mark the set of down-regulated transcripts.

TARBP2 HITS-CLIP

HITS-CLIP was performed as previously described[39] with the following modifications. MDA-LM2 cells were transiently transfected with a vector expressing myc-tagged TARBP2 (Origene). Forty-eight hours post-transfection, cells were crosslinked with 400mJ UV. An anti-myc antibody (Millipore 4A6 #05-724) was conjugated to protein G dynabeads (Life Technologies) and used to immunoprecipitate the RNase A treated protein-RNA complexes. After CIP treatment, the RNAs bound to the immunoprecipitated proteins were labeled with ^{32}P by PNK. After elution from the beads, the labeled protein-RNA complexes were resolved by SDS-PAGE and visualized by blotting onto a nitrocellulose membrane and exposing the membrane to film. After isolation from the nitrocellulose membrane, the RNA was prepared into a four-fold multiplexed Illumina compatible sequencing library using the Nextflex small RNA kit (Bioo Scientific). We sequenced low (1:3,000 dilution) and medium (1:1,500 dilution) RNase A-treated samples and used the medium treatment samples for the results presented here. We used strand-specific ChIPseeqer[8,40] to identify TARBP2 binding sites common between the two biological replicates. To correct for the peak positions obtained from single-end reads, we appended an additional 50bp downstream of the identified peaks. We used a 1st order Markov model (provided as part of ChIPseeqer[40]) to

generate shuffled sequences from the identified binding sites. These randomized sequences were used as background for measuring the enrichment of sRSE instances among the identified sites. Similar to the case for the entire transcriptome, to find sRSE matches, we used TEISER in non-discovery mode to locate all potential instances of the entire family. *In silico* folding was then used to identify the sequence matches that also strictly followed the structural attributes. The set of these matches were defined as sRSE instances, which were then used to measure enrichment among the binding sites. In parallel, we also used crosslinking induced mutation sites (CIMS[41]) to ensure that our findings were not sensitive to the choice of analytical pipeline. Similarly, we also used sRSE1 as the representative member of the family to run TEISER in non-discovery mode on TARBP2-bound sequences.

Identifying the TARBP2 bound transcripts

TARBP2 binding sites identified using strand-specific ChIPseeqer[8,40] were used to obtain a comprehensive list of TARBP2 bound transcripts (~5000 transcripts). We subsequently calculated the enrichment/depletion of this regulon in our whole-genome measurement datasets, namely: (i) differential stability measurements in MDA-LM2 vs. MDA-231 cells, (ii) expression and stability of transcripts in RNAi-mediated TARBP2 knock-down, and (iii) expression in TARBP2 over-expression in MDA-LM2 cells.

Western blotting

Cellular lysates were prepared by lysing cells (5-10 million) in ice-cold RIPA buffer containing protease and phosphatase inhibitors (Roche). Cellular debris was removed by centrifugation (12,000 rpm) for 20 min at 4 °C. Samples were denatured in loading buffer, separated using SDS–PAGE, transferred to a PVDF membrane (Pierce), blocked, and probed using an anti-TARBP2 antibody (1:1,000; H00006895-PW1, Abnova). Bound antibodies were chemiluminescently detected using horseradish peroxidase–conjugated secondary antibodies (1:10,000), ECL Western Blotting Substrate (Pierce) and the SRX-101A (Konica Minolta) developer, according to the manufacturer's instructions. The membrane was then stripped (Restore Western Blot Stripping Buffer, Pierce), re-probed (at RT for 1 hr) and re-developed (similar to the previous step) using an anti-tubulin antibody (1: 1,000; Sigma) as an internal control.

Analysis of TARBP2 transcript expression in cDNA arrays

Breast cancer TissueScan cDNA plates I and IV (Origene) were re-suspended in 44μL of water. 10 μL of dissolved cDNA was quantified (in technical quadruplicate) for *ACTB* expression per manufacturer's instructions. Another 10μL was used for TARBP2 quantification as described before[6]. 5'-CAGGAGTATGGGACCAGAATAGG and ACCCGGAAGGTGAAATTAGGC were used as primers.

Metastatic colonization assays

Stable *TARBP2* knock-down cells were generated by stable transfection and expression of shRNAs in MDA-LM2 and CN-LM1a lines. TARBP2-sh1 and -sh2 in MDA-LM2 corresponds to TRCN0000019339 and TRCN0000330578 (Sigma) and in CN-LM1a to TRCN0000330642 and TRCN0000019339. In all cases, we achieved ~8-fold reduction in

TARBP2 transcript levels measured using qRT-PCR. Seven- to eight-week-old age-matched female NOD/SCID gamma mice (Jackson Labs) were used for lung metastasis assays[6]. In all cases, 2×10^4 cells in 100 μL PBS were injected via tail-vein along with cells expressing a neutral hairpin as control. Metastatic growth was tested using both a two-way ANOVA as a function of time and sub-line identity and also a Mann-Whitney test mediated comparison of area under the curve for each mouse.

Histology

For gross macroscopic metastatic nodule visualization, three lungs from representative mice (from each cohort) were extracted ~30 days post-injection and 5 μm thick lung tissue sections were hematoxylin and eosin (H&E) stained. The number of macroscopic nodules was then recorded for each section. An unpaired *t*-test and a Mann-Whitney test were used to test for significant variations.

Cancer cell proliferation

Roughly 1×10^4 cells were seeded into three 6-well plates. They were subsequently trypsinized and viable cells were counted using a hemocytometer at days 1, 3 and 5. An exponential model was then used to fit a growth rate for each sample ($\ln(N_{t-1}/N_1)= rt$ where t is measured in days). The experiment was performed in biological quadruplicates and a one-tailed Mann Whitney test was used to test for significant variations.

Invasion assays

Invasion assays were performed as previously described[42]. For APP peptide pre-treatment, 100 ng of APPα, APPα* (304-612), or APPβ (Sigma) were added to trans-well invasion chambers (BD Biosciences) along with BSA (Sigma) as control.

Orthotopic metastasis

For orthotopic metastasis assay, we prepared 5×10^5 cells in a 100 μL of 1:1 mixture of PBS with Matrigel and injected them into the mammary fat pad of 7-8 week old NOD-SCID gamma female mice (the two bottom mammary fat pads in each mouse were used). After one week, wound clips were removed and tumour size was measured every three days beginning from day 20 post-injection to the day 32. At day 32, tumours were extracted and one week later, the metastasis to the lungs were measured by bioluminescence imaging. The lungs were then extracted for *ex vivo* imaging. Tumour growth rates (day^{-1}) were estimated based on an exponential model starting at day 20.

Generating double-knock down cell lines

The four TARBP2 targets (i.e. *APP*, *ZNF395*, *AP2A2*, and *LAMB1*) were chosen based on the following criteria: higher expression and stability in MDA-LM2 *TARBP2* knock-down (siRNA-mediated), down-regulation in MDA-LM2 cells relative to the parental MDA line, and carrying significant tag clusters in TARBP2 HITS-CLIP. For each of the four targets, independent shRNAs expressed in pLKO.1-Blast vector were stably transfected into TARBP-sh1 in MDA-LM2 and TARBP2-sh1 in CN-LM1a (which correspond to different hairpins as mentioned above). For the secondary hairpin, we used TRCN0000381611,

TRCN0000183104, TRCN0000006705, and TRCN0000083428 in the MDA-LM2-shTARBP2 background and used TRCN0000065123, TRCN0000184700, TRCN0000006706, and TRCN0000083429 in CN-LM1a-shTARBP2 subline. Double knock-downs were subsequently validated using qRT-PCR. A control hairpin in pLKO.1-Blast was used as the reference. These sub-lines were subsequently used for *in vivo* colonization assays, histology and *in vitro* invasion assays upon validating sufficient (>3-fold) knock-down in the targeted transcripts.

Clinical associations

For the aggregate target gene-signature, *i.e. APP, ZNF395, AP2A2*, and *LAMB1*, we used a compendium of breast cancer gene expression datasets (BR comp.)[19] to generate Kaplan-Meier curves. For the *APP-ZNF395* combined signature, GSE25066[43] was also used, which records survival after taxane-anthracycline chemotherapy to highlight that the association of these genes were not remedied by chemotherapy. The *p*-values were calculated using a one-tailed log-rank test. For association between expression and tumour grade, we used three independent datasets—GSE5460[44], GSE2109, and the breast cancer compendium above—to stratify tumours based on BR-grades (high versus low) and employed a one-tailed Mann-Whitney test to calculate the associated *p*-values.

Gene expression profiling of ZNF395 knock-down in TARBP2-depleted background

Total RNA samples from shZNF395 and shControl in MDA-LM2-shTARBP2 background were prepared for hybridization as detailed before (in quadruplicates).

Statistics and animal studies

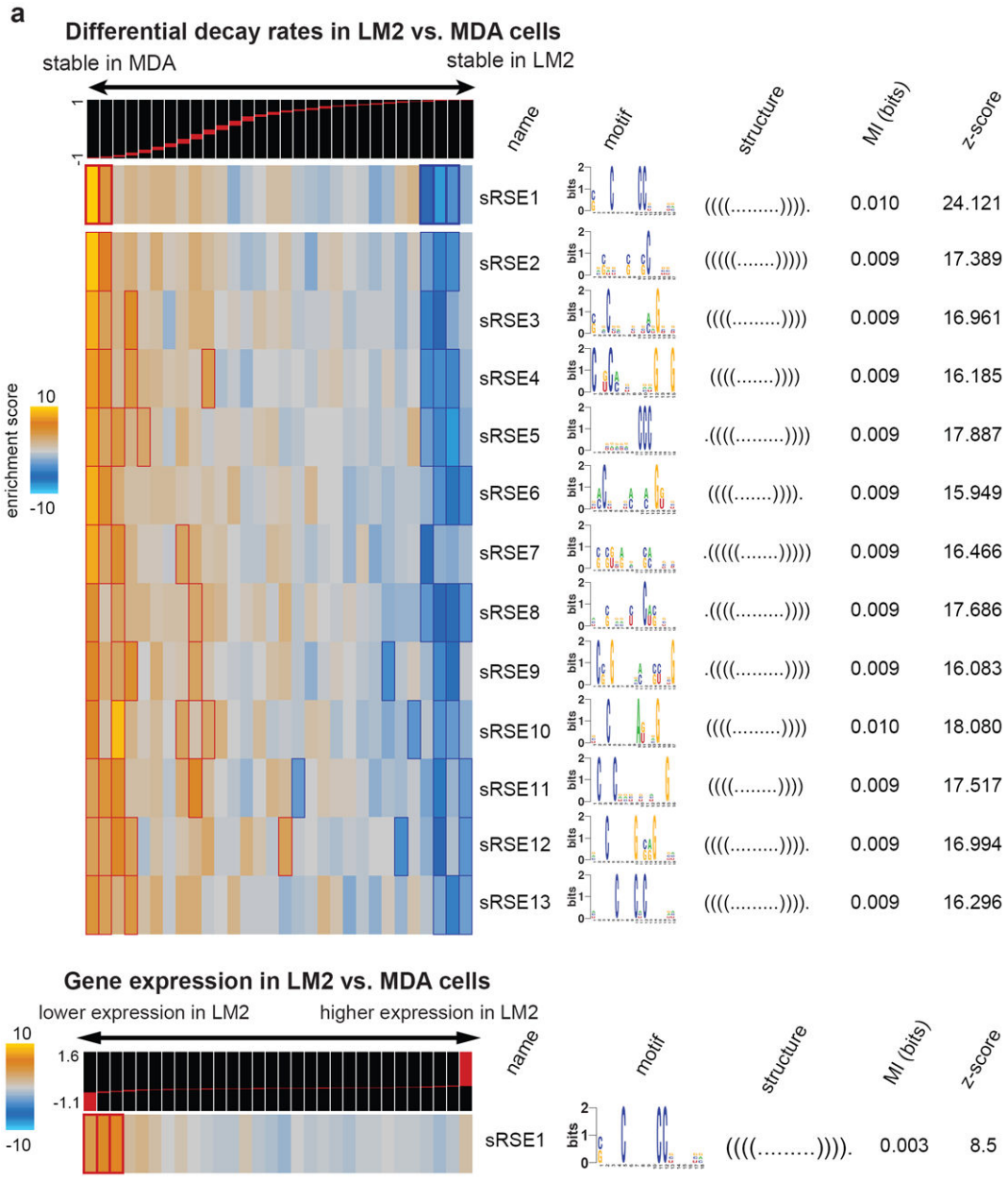
All animal work was conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee at The Rockefeller University. As mentioned above, animal experiments were performed in cohorts of approximately five mice each. This number has been previously shown to provide sufficient power to detect larger than 2-fold changes in metastatic burden in the lungs. Independent cell-lines and hairpins were used to ensure that the observed effects were neither cell-line specific nor due to off-target effects. Criteria for exclusion included accidental death prior to the completion of the experiment for causes unrelated to the experiment (*e.g.* death due to injection). A two-way ANOVA and a one-tailed Mann-Whitney were used to compare cohorts (as described above). No blinding was used for the animal studies. Mice were randomly assigned to the sample and control cages prior to injections.

The statistical tests used to compare samples and cohorts in this study are previously described and well-established. In cases in which parametric *t*-test was used, we first checked normality using a Shapiro-Wilk normality test or a KS normality test (depending on the number of measurements available). We were cautious in making as few assumptions as possible in our data analysis, often utilizing non-parametric and/or multiple statistical tests for each experiment/analysis.

Data availability

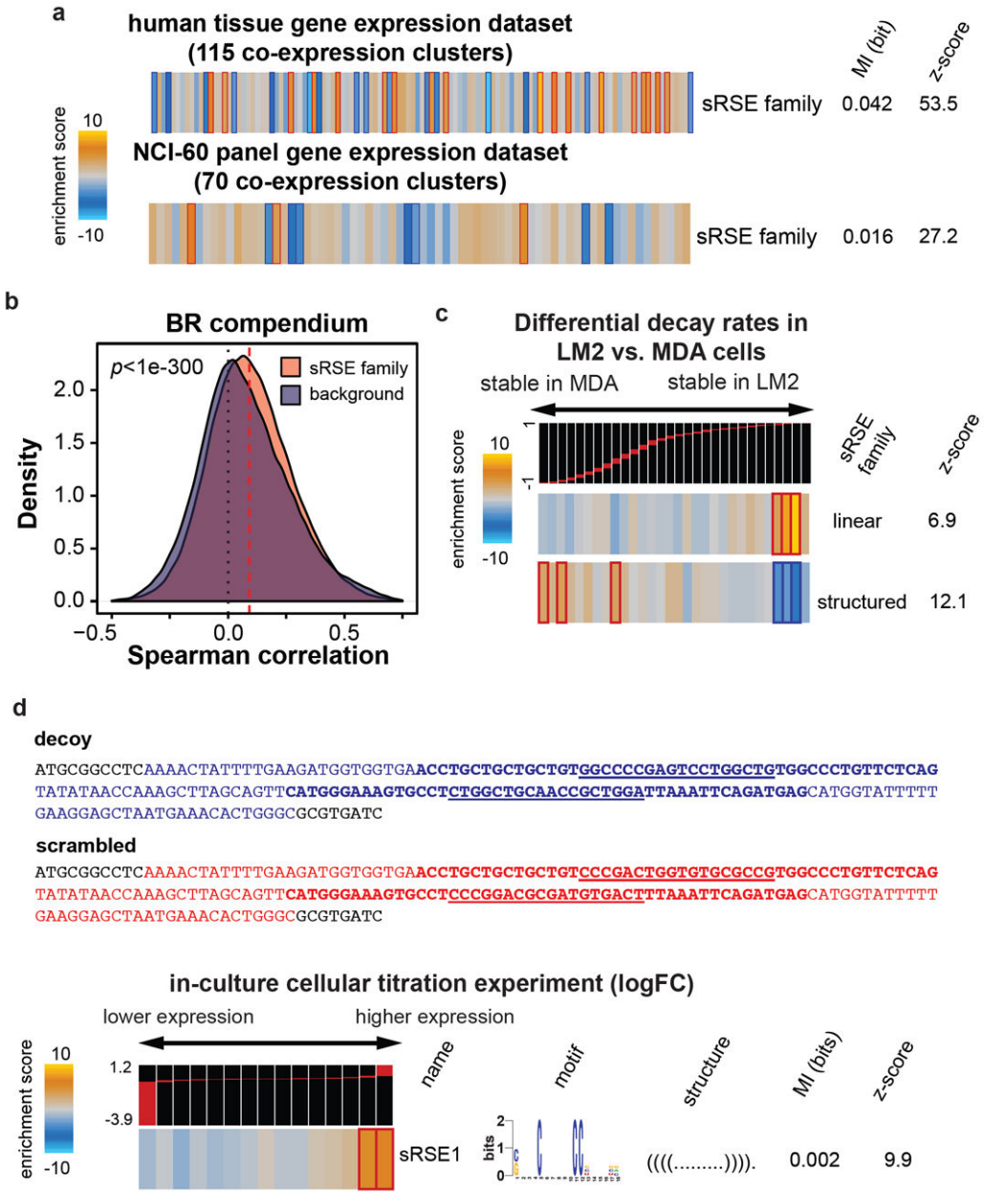
Raw datasets are deposited in the Gene Expression Omnibus database under accession number GSE49649. The processed files and associated results are also available at https://tavazoielab.c2b2.columbia.edu/TBSE; TEISER is available at https://tavazoielab.c2b2.columbia.edu/TEISER.

Extended Data



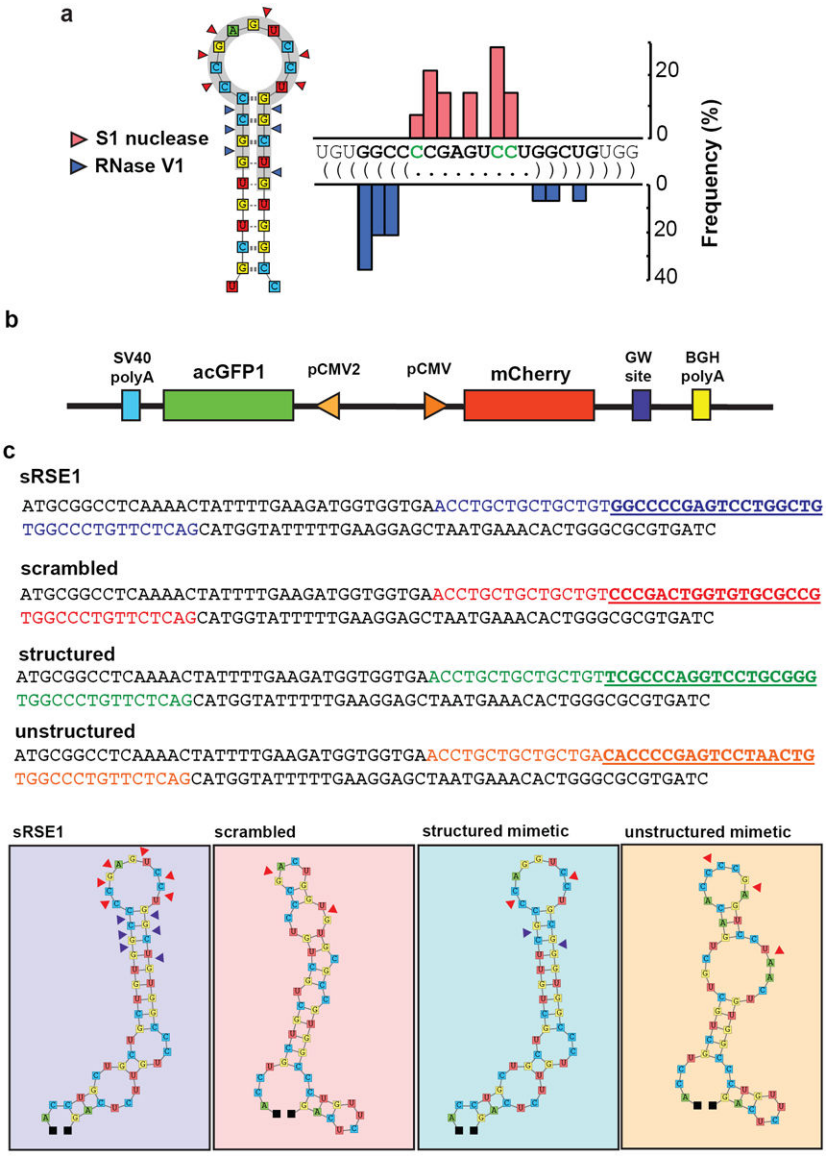
Extended Data Figure 1. The sRSE family of structural *cis*-regulatory RNA elements is enriched in transcripts exhibiting reduced stability in metastatic MDA-LM2 cells

a, The sRSE family is composed of 318 GC-rich structural elements with sRSE1 serving as the most informative representative. These elements were enriched in the transcripts destabilized in MDA-LM2 cells. Here, we have included a subset of structural elements that reside in this family. These motifs display a lower statistical significance relative to sRSE1 and have a conditional information ratio of less than 2.0 with respect to sRSE1; in other words, these motifs form an overlapping family of structural elements for which sRSE1 is the best representative. For the analyses presented in this study, we have provided enrichment/depletion patterns for the entire sRSE family as well as sRSE1 as the representative. **b,** sRSE1 was significantly enriched in transcripts down-regulated in MDA-LM2 cells relative to the parental population.



Extended Data Figure 2. The sRSE-carrying transcripts show correlated expression in human gene expression datasets and are modulated in in-culture cellular titration experiments

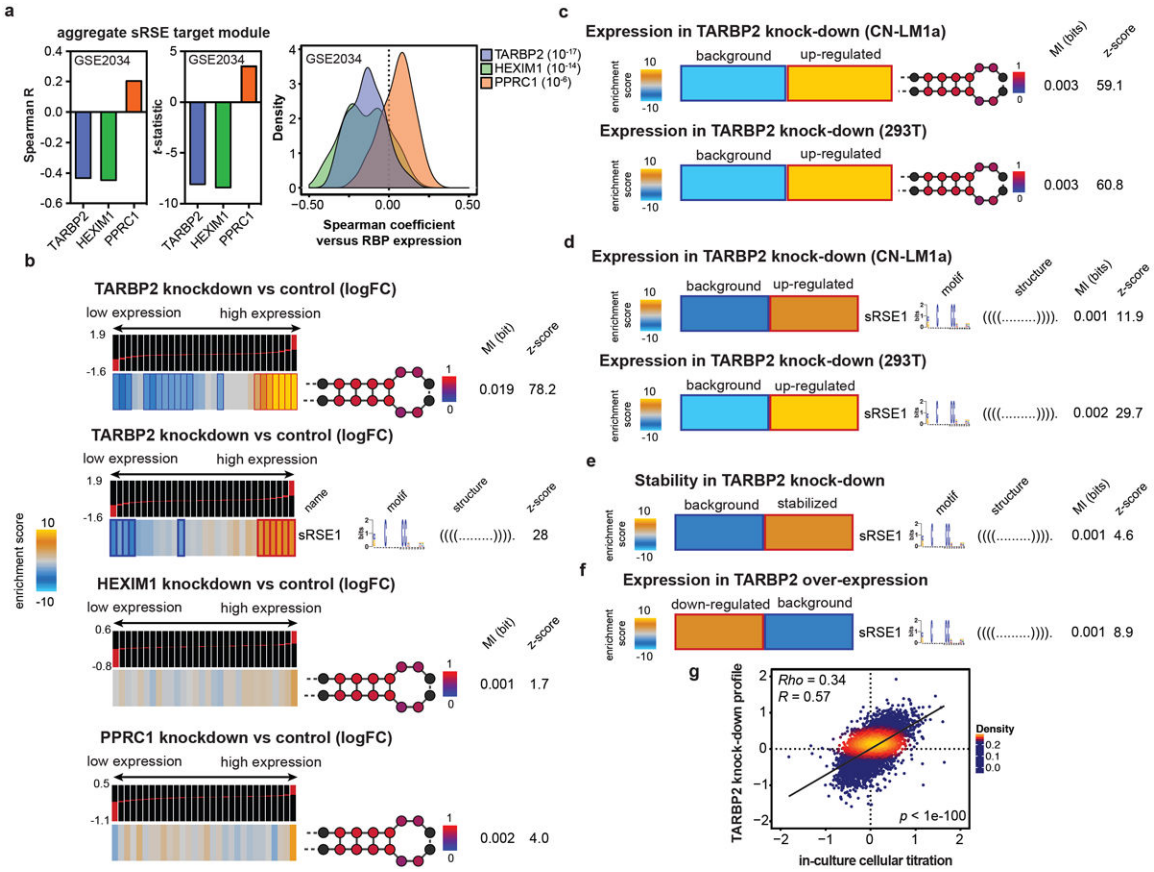
a, The enrichment pattern of the sRSE regulon in two human whole-genome co-expression datasets. The significant enrichment/depletion pattern in different co-expression clusters further supports the role of sRSEs as potentially functional regulatory elements. **b,** The distribution of Spearman correlation coefficients between the expression of active sRSE-carrying transcripts across a compendium of breast cancer profiles[19]. Active targets were defined as transcripts that carry the sRSE instances and showed a differential decay rate score of less than - 0.8. For reference, we also generated a background distribution by selecting a random set of transcripts (of equal size) and calculated pair-wise correlations. The sRSE-carrying transcripts showed a significantly higher correlated expression than expected by chance (the calculated *p*-value is based on a one-tailed Mann-Whitney test). **c,** The transcripts with sequence but no structural matches to sRSEs (the sequence component of sRSEs) were stratified based on whether or not the sequence match also satisfies the structural restrictions (labeled as structured or linear). As shown here, the sequence restrictions alone showed no enrichment among the destabilized transcripts. **d,** Decoy (blue) and scrambled (red) sequences were chosen based on an identical randomly generated backbone sequence with the bold sequences matching the chosen sRSE1 sites (along with ~30nt flanking sequences). Only the underlined regions, which correspond to exact matches to the sRSE1 motif description, were shuffled to generate the scrambled sequences. **e,** sRSE1 was significantly enriched among transcripts up-regulated in MDA-LM2 cells upon the transfection of decoy RNA in comparison to scrambled oligonucleotides.



Extended Data Figure 3. ***In vitro*** **secondary structure probing of the sRSE1 decoy and measuring the functionality of exemplary sRSE variants in reporter systems**

a, We used a medium-throughput approach to probe the secondary structure of a given sRSE1 decoy sequence (see Methods). Shown here are *in silico* and *in vitro* secondary structure prediction for a functional sRSE instance. The histogram for each nuclease digestion shows the percentage of clones (from a total of ~15 tested clones) with truncation at each specific site. The *in silico* folding was constrained based on the digestion sites. **b,** The reporter construct used for testing the functionality of the engineered elements. A Gateway site in the 3' UTR of mCherry was used to insert elements downstream of the mCherry coding sequence; acGFP served as the internal control. qRT-PCR of mCherry and GFP transcripts was then used to assay the effects of the cloned sRSE variants. **c,** Decoy, scrambled, structured and unstructured sequences (along with their predicted secondary structure) that were used in the mCherry/GFP reporter system. In the scrambled sequence, the nucleotides were shuffled in order to compromise both the sequence and the structure of

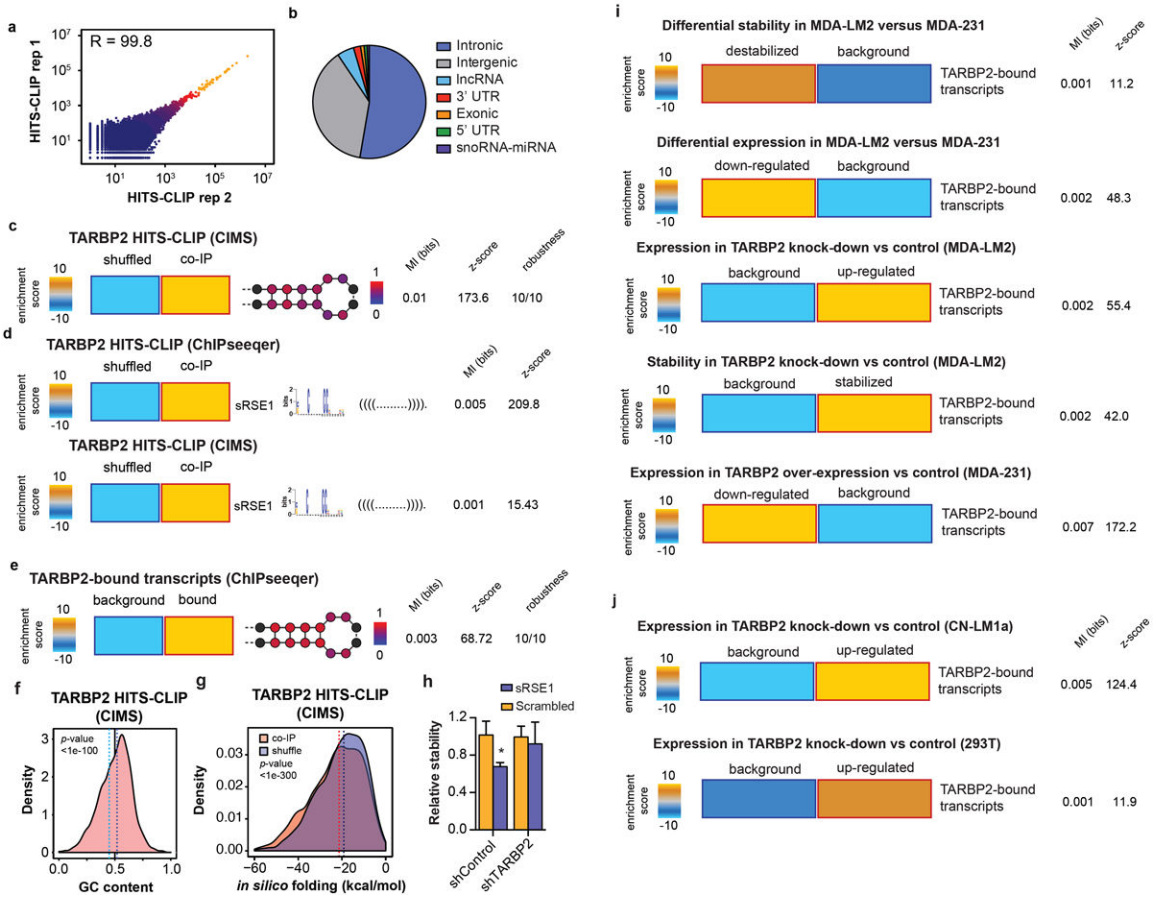
the element. In the structured mimetic sequence, the nucleotides in the stem are swapped while maintaining the sequence of the loop, ensuring that the structure of the stem is maintained while its sequence is changed. In the unstructured mimetic sequence, only the nucleotides in the stem are mutated to disrupt the structure while maintaining the sequence identity of the loop. The *in silico* folding was guided based on the information provided by batch V1/S1 digestion of RNA variants followed by cloning and sequencing.



Extended Data Figure 4. The identification of TARBP2 as the potential modulator of the sRSE regulon and the regulatory consequences of TARBP2 modulations

a, Three RNA binding proteins (RBPs), TARBP2, HEXIM1 and PPRC1, were selected as potential regulators of the sRSE regulon based on their positive/negative correlations with this regulon at the transcript level (Methods). Here, the Spearman correlations (and their associated *t*-statistics) between the sRSE target signature and each of the candidate RBPs have been shown. We have also included the distribution of Spearman correlations for each candidate RBP along with the associated *p*-values (see Methods). **b,** Unlike *TARBP2*, RNAi-mediated knock-down of *PPRC1* and *HEXIM1* did not result in a significant deregulation of the sRSE-carrying transcripts. **c,** In addition to MDA-LM2 cells, a significant increase in the expression sRSE-carrying transcripts was also observed in RNAi-mediated *TARBP2* knock-down cells in CN-LM1a and 293T backgrounds. **d,** The sRSE1 element showed a similar enrichment pattern as that of the entire sRSE family with significant over-representations among the transcripts up-regulated upon TARBP2 silencing in both CN-LM1a and 293T

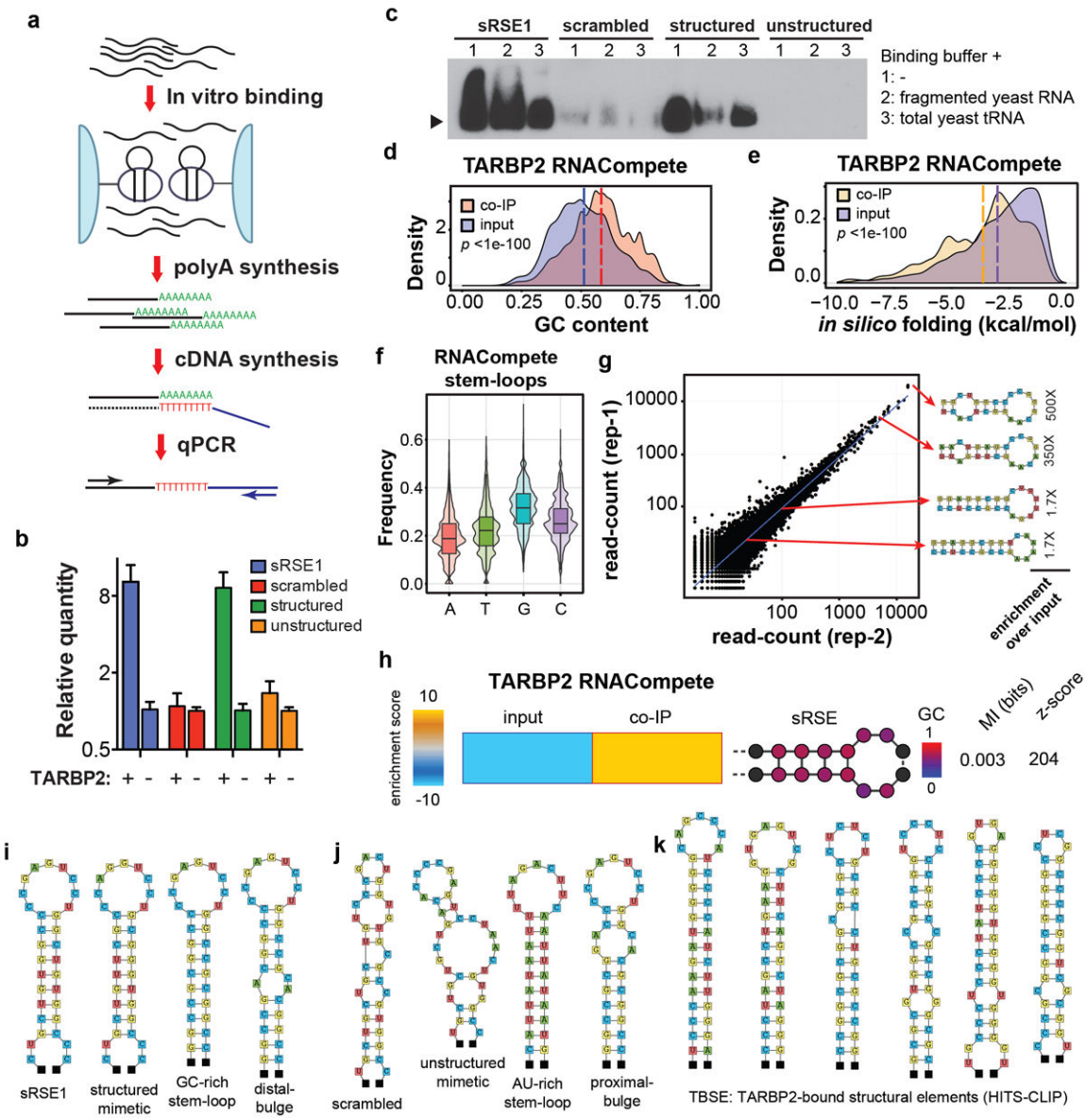
backgrounds. **e-f,** The enrichment/depletion patterns of sRSE1-carrying transcripts, as the representative of the sRSE family, in transcript stability and gene expression profiles of cells in which *TARBP2* levels were modulated. **g,** Linear regression analysis of gene-expression profiles from *TARBP2* siRNA knock-down and sRSE1 in-culture cellular titration experiments. The smoothed density at each point in the scatterplot has been shown as a colormap. The *p*-value represents the significance of the reported Spearman correlation (Rho).



Extended Data Figure 5. TARBP2 directly interacts with and modulates the sRSE regulon *in vivo*

a, TARBP2 HITS-CLIP results were significantly correlated between the two replicates. **b,** TARBP2-bound sequences (determined by HITS-CLIP) were annotated and presented as a pie-chart. The majority of interactions were intronic; however, direct interactions with exons, sno/miRNAs and lnc-RNAs were also observed. **c,** CIMS analysis of TARBP2 HITS-CLIP data (Methods), similar to the results obtained from ChIPseeqer identified binding sites, also showed an enrichment of the sRSE family in the TARBP2 binding site relative to their shuffled counterparts (1st-order Markov model; see Methods). **d,** Similar to the case for the entire sRSE family, sRSE1 was also significantly enriched among the TARBP2 binding sites identified by ChIPseeqer and CIMS respectively. **e,** We observed a highly significant overlap between transcripts that we identified as direct interacting partners of TARBP2 (using HITS-CLIP; see methods) and the sRSE-carrying transcripts (53% overlap,

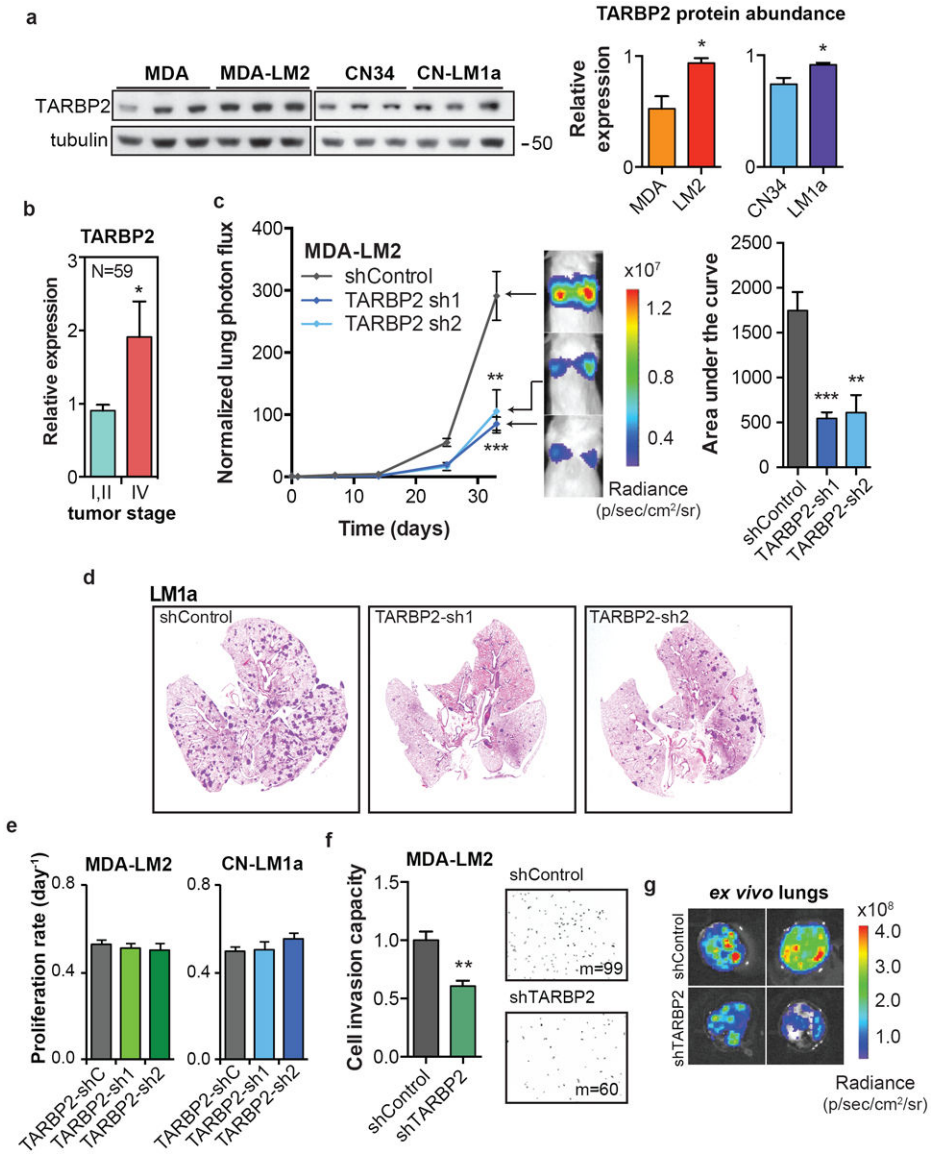
hypergeometric *p*-value < 10^{-22}). **f,** The TARBP2 HITS-CLIP clusters, identified as the cross-linking induced mutations sites (CIMS) common between the two biological replicates, were folded *in silico* and the resulting terminal stem-loops were extracted. Consistent with the GC-rich nature of sRSE elements, these stem-loops displayed a significant shift towards higher GC-contents. The associated *p*-value was calculated using a Wilcoxon rank-sum test with μ=0.45 (GC content of annotated genes in human genome marked by the light blue line) as the null hypothesis. **g,** *In silico* folding of the identified TARBP2 CIMS binding sites (Vienna RNA Package[45]) displayed more stable secondary structures formed by real sequences relative to their shuffled counterparts (1^{st}-order Markov model). The associated *p*-value was calculated using a paired Mann-Whitney test between free-energy estimates for real and shuffled sequences. **h,** The relative stability of sRSE1/scrambled-fused mCherry reporters assayed in control and TARBP2 knock-down cells. Samples in quadruplicate were treated with α-amanitin and total RNA was extracted at 0- and 8hr time-points. Following DNase digestion and cDNA synthesis, transcript levels were quantified using qPCR in order to measure relative stability for mCherry (with GFP as the endogenous control). Error bars indicate s.e.m. *, $P<0.05$ using a Mann-Whitney test. **i-j,** In this study, we initially used the sRSE-family as a proxy to track the behavior of a broader regulon. Upon identification of TARBP2 as the *trans* factor, however, we can directly assess the targets that are bound by TARBP2 *in vivo*. As shown here, consistent with our previous observations, these transcripts showed an enrichment pattern identical to that of sRSE-carrying transcripts: they are enriched among the transcripts that are down-regulated/destabilized in MDA-LM2 cells, up-regulated and stabilized in *TARBP2*-knockdown cells, and down-regulated when TARBP2 is over-expressed. We observed similar up-regulations upon TARBP2 knock-down in CN-LM1a and 293T cells. These observations further support the coherence of our findings and the use of sRSE as a surrogate for probing this regulon.



Extended Data Figure 6. TARBP2 interacts with GC-rich stem-loops *in vitro*

a-b, A qRT-PCR based assay for quantification of RNA variants immunoprecipitated in the presence and absence of tagged TARBP2. This assay, which has been widely adopted for miRNA qRT-PCR, relies on poly-A tailing and oligo-dT annealing for cDNA synthesis followed by a universal reverse primer and variant specific forward primers for quantitative PCR. We combined our four stem-loop variants presented in Extended Data Fig. 5c (i.e. sRSE1, scrambled, structured and unstructured mimetics) in equal amounts and subjected them to co-purification and precipitation in the presence and absence of His-TARBP2 (20nmol of RNA plus 20μg of TARBP2 per sample in triplicate) as described before[46]. Relative quantity of each RNA variant was then measured using the assay described above. For the sRSE1 and structured variants, we observed an ~8-fold enrichment after TARBP2 co-purification. **c,** RNA electrophoretic mobility shift assays for the four variants also support the binding of the sRSE1 and structured variants to TARBP2 *in vitro*. 200 picomoles

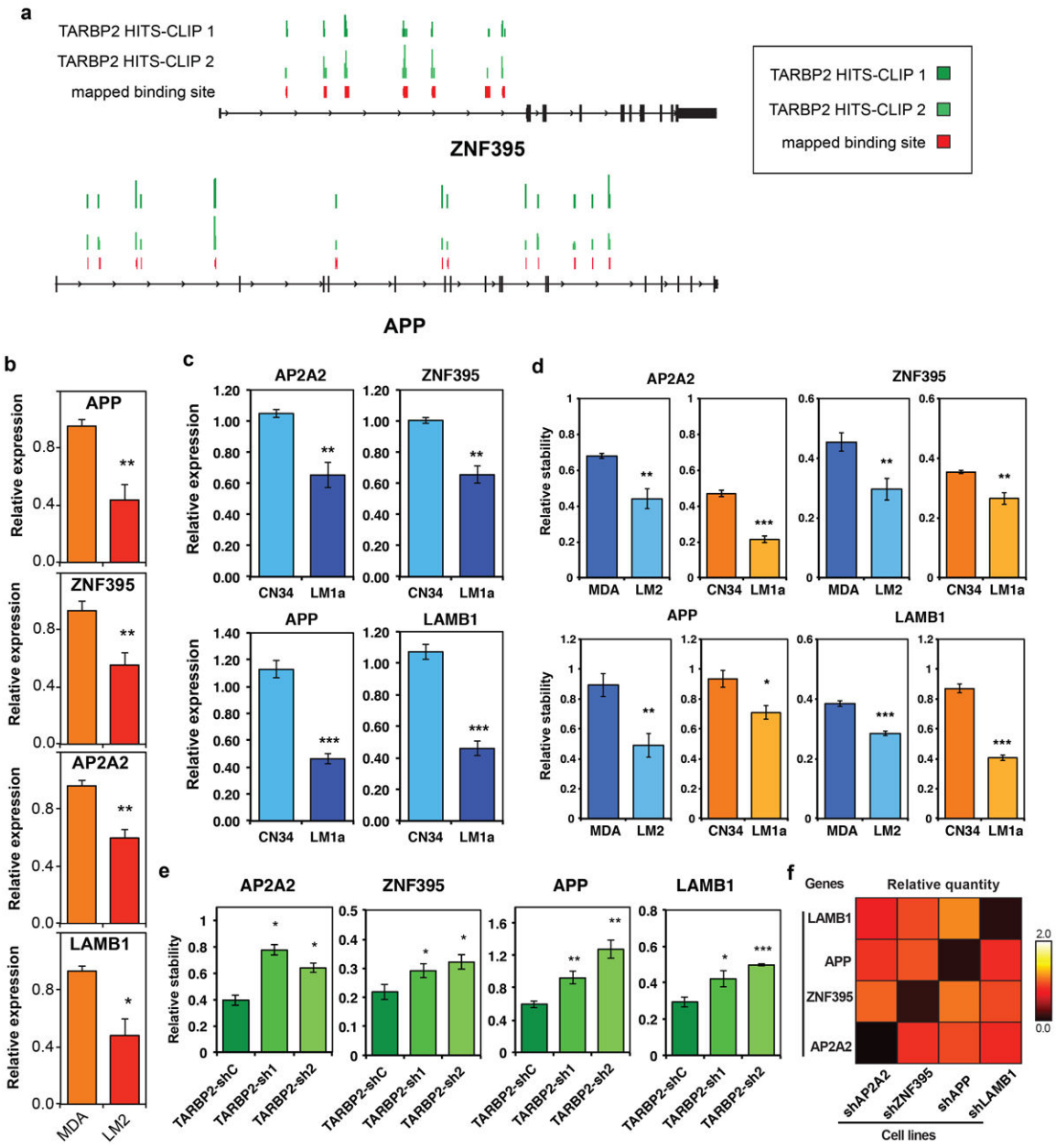
of 3' biotinylated RNAs were incubated in the presence of 200ng of TARBP2 and three different non-specific competitors (i.e. 10ng/μL DNA, 10ng/μL DNA+1μg/μL fragmented yeast RNA, and 10ng/μL DNA+0.1μg/μL yeast tRNA) as described before[16]. **d,** We also used a degenerate library of 5' phosphorylated 24-nt long RNA molecules to select for sequences that preferentially bind TARBP2 *in vitro* in one round of selection (a strategy termed RNAcompete[34]). Upon co-precipitating 20 nmoles of randomized RNA with 20μg of TARBP2 (in duplicate), we prepared the input as well as the selected libraries for small-RNA high-throughput sequencing. Upon linker removal, the 24-nt reads from the two replicates were combined and collapsed into unique sequences (~35 million unique reads in the input). Calculating the GC-content of the selected versus input library showed a significant enrichment of sequences with higher GC-content. **e,** *In silico* folding (using RNAfold) of the input and selected sequences also showed a significant shift towards sequences with higher propensity to form secondary structures. **f,** The break-down of nucleotide frequencies in the selected sequences showing a significantly higher preference of TARBP2 for GC-rich sequences. **g,** The number of reads for each sequence was generally consistent between the two populations. *In silico* folding of sequences with high read-counts (>80% sub-optimality) form GC-rich stem-loops similar to those described for the HITS-CLIP tags. For example, we have included a number of these stem-loops along with their fold enrichment over the input library. **h,** sRSEs were highly enriched among the *in vitro* selected RNA population relative to the input. Together, these *in vitro* binding experiments show that TARBP2 binds GC-rich stem-loops. **i,** GC-rich structural elements validated in *in vitro* TARBP2 binding experiments. The first two stem-loops (sRSE1 and structured mimetic) were also tested using reporter assays. **j,** Scrambled and unstructured mimetic showed minimal or no binding to TARBP2 *in vitro* and were not functional in reporter assays. An AT-rich stem-loop and a GC-rich stem-loop with a proximal bulge showed significantly diminished binding to TARBP2 in our *in vitro* binding assays (17% and 12% higher than scrambled control relative to >2-fold for functional stem-loops, respectively). **k,** A set of exemplary sRSE instances likely bound by TARBP2 discovered among the TARBP2-bound sequences (HITS-CLIP).



Extended Data Figure 7. TARBP2 silencing suppresses metastatic colonization and results in fewer metastatic nodules formed in the lungs of xenografted mice

a, Western blots were performed for TARBP2 in MDA, LM2, CN34 and LM1a cells using Tubulin as an internal control. TARBP2 relative protein abundance was then quantified with each sample measured in triplicate. Quantification was performed using ImageJ. Statistical significance was assessed using a Mann-Whitney test (*, P=0.05). **b,** Relative expression levels of *TARBP2* mRNA in metastatic (stage IV) tumours compared to early-stage (stages I and II) tumours (n=59; see Methods). Error bars indicate s.e.m. *, P<0.05 by a one-tailed Student's t-test. **c,** Bioluminescence imaging plot of lung metastasis by MDA-LM2 cells expressing short hairpins targeting *TARBP2* (shTARBP2) or a control hairpin (shControl); n=5 and 6 respectively. Area-under-the-curve was also calculated for each mouse. **d,** Haematoxylin-and-eosin-stained images of representative lung sections from a tail-vein injected mouse cohort extracted at day 33 post-injection (Figure 3b). **e,** *In vitro* proliferation rates (day^{-1}) for cells expressing shControl or *TARBP2* shRNAs. **f,** Cell invasion capacity of

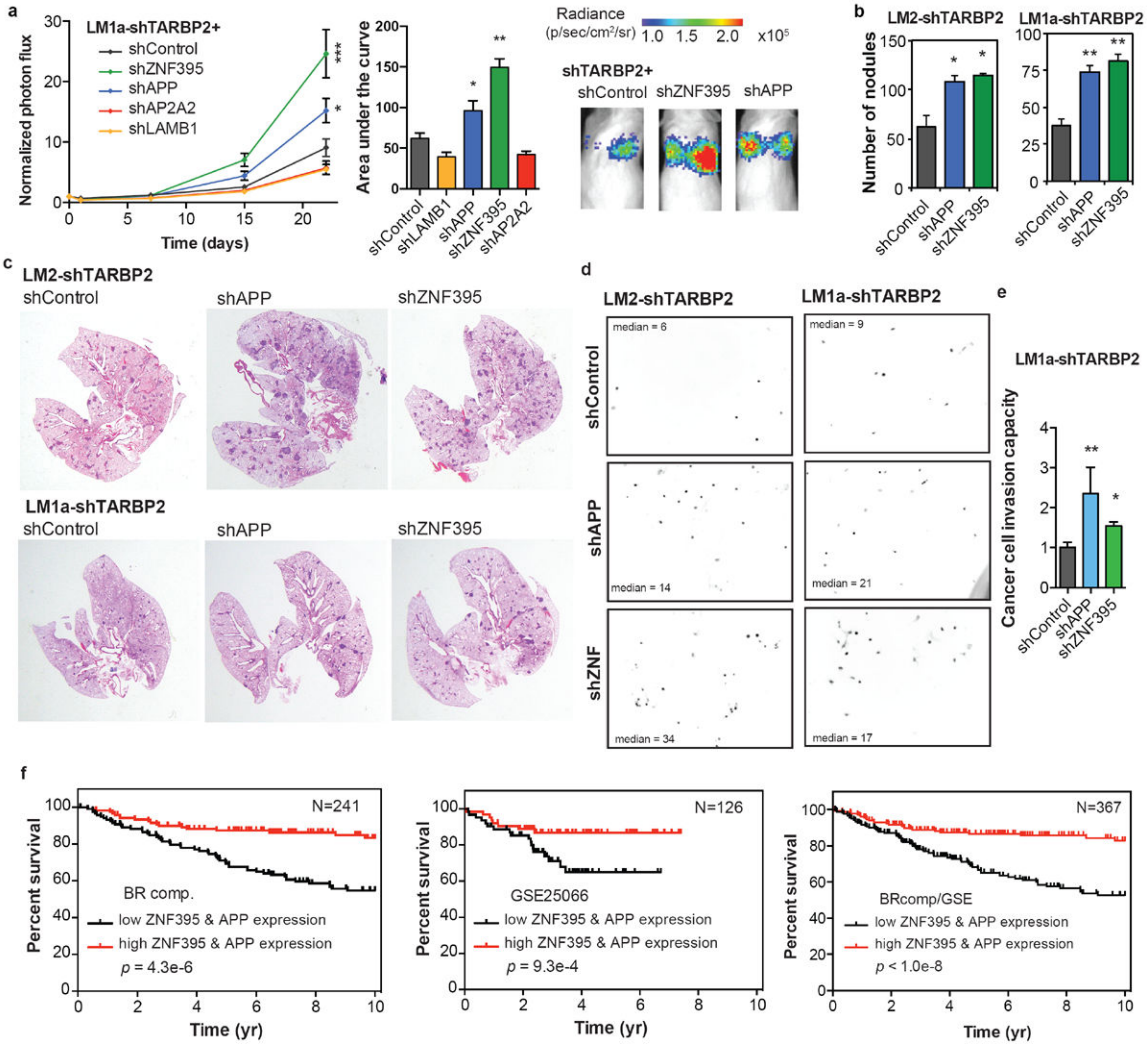
MDA-LM2 cells quantified upon trans-well matrigel invasion assays; n=8 for each sample comparing shTARBP2 to shControl cells (two independent sets of biological quadruplicates). Also shown are representative images of trans-well inserts along with the median count (m) for each experiment. **, $P<0.01$ by a one-sample Mann-Whitney test. **g,** Lung bioluminescence (seven days post tumour extraction) measured upon extraction *ex vivo* (n=3 for shControl and 4 for shTARBP2 cohorts). Error bars indicate s.e.m.



Extended Data Figure 8. AP2A2, ZNF395, APP and LAMB1 are post-transcriptionally regulated by TARBP2

a, Shown are the HITS-CLIP cluster-tags common in both replicates for the identified TARBP2 binding site. These sites mark the interaction sites between TARBP2 and its

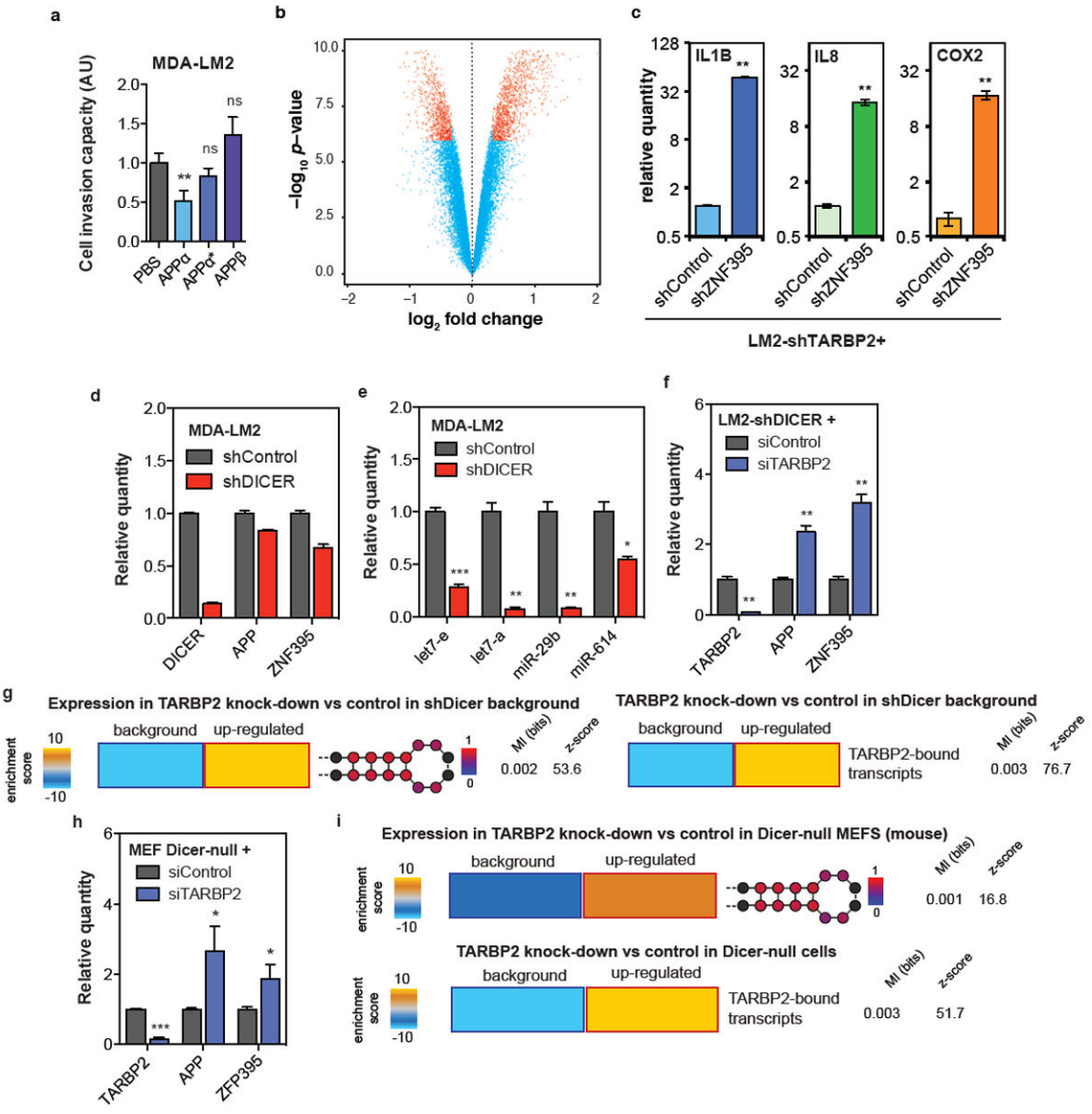
targets, *APP* and *ZNF395*. **b-c,** Relative expression levels of direct TARBP2 targets *APP*, *ZNF395*, *AP2A2*, and *LAMB1* in MDA, MDA-LM2, CN34, and CN-LM1a cells; n=4 in each sample. **d,** The observed down-regulation of these targets resulted from reduced stability as measured by α-amanitin-mediated inhibition of transcription. For this, relative abundance of each transcript was measured at 0- and 18-hr time-points (using 18S rRNA as endogenous control). **e,** These targets displayed significantly higher stability in the context of *TARBP2* knock-down in LM2 cells (shC is shControl; sh1 and sh2 are two independent hairpins). **f,** Epistatic interactions were not observed between the TARBP2-regulated genes as the knock-down of each gene did not significantly modulate the levels of the other transcripts. Error bars indicate s.e.m. *, $P<0.05$, **, $P<0.01$, and ***, $P<0.001$ by a one-tailed Student's *t*-test.



Extended Data Figure 9. APP and ZNF395 suppress metastatic cell invasion and colonization

a, Bioluminescence imaging plot of lung metastasis by cells expressing short hairpins targeting one of four *TARBP2* targets or a control hairpin (shControl) in a *TARBP2* knock-down background in CN-LM1a lines; n=5 in each cohort except for shZNF395 where n=4.

Area-under-the-curve summaries are also shown for each cohort along with bioluminescence images from representative mice from significant cohorts (*APP* and *ZNF395*). **b,** The number of nodules quantified per lung section for each cohort in both MDA-LM2 and CN-LM1a backgrounds. *, $P<0.05$, **, $P<0.01$, by a one-tailed Student's *t*-test. **c,** Haematoxylin-and-eosinstained images of representative lung sections show an increase in the number of metastatic nodules in shAPP and shZNF395 cells. **d,** *APP* and *ZNF395* were individually knocked-down in LM2 and LM1a cells expressing shRNAs targeting *TARBP2*. The resulting populations were subjected to trans-well invasion assays using a control hairpin in parallel (n=8 for each sample). Representative (median) fields are shown here. *, $P<0.05$, **, $P<0.01$, by one-tailed Mann-Whitney test. **e,** Cancer cell invasion capacity relative to shControl cells measured for *APP* and *ZNF395* knockdown cells in the *TARBP2* knock-down background; n=8 for each sample. **f,** Tumours with lower than median expression of both *APP* and *ZNF395* displayed a significantly lower survival rate in various cohorts of patients[14,19,47]. Error bars indicate s.e.m.



Extended Data Figure 10. TARBP2-mediated down-regulation of APP and ZNF395 is DICER-independent

a, Cancer cell invasion capacity measured for MDA-LM2 cells with exogenously added APPα, APPα* (*i.e.* α-APP$^{304-612}$), APPβ, and BSA as control; n=6 or 8 for each sample. **b,** A volcano plot, based on the microarray analysis of *ZNF395* knock-down cells versus control samples, depicting a large number of probes that were significantly deregulated in *ZNF395* knock-down cells. **c,** qRT-PCR validation for three pro-metastatic/cell invasion genes up-regulated in *ZNF395* knock-down samples. Error bars indicate s.e.m. **, $P<0.01$ by a one-tailed Student's *t*-test. **d,** We achieved ~8-fold *DICER* knock-down in the MDA-LM2 cells expressing a short-hairpin targeting *DICER* (TRCN0000290486, Sigma). This level of *DICER* knock-down, unlike *TARBP2* knock-down, did not result in an up-regulation of *APP* and *ZNF395*. **e,** To ensure functional knock-down of Dicer in these cells, we performed qRT-PCR for four microRNAs (TaqMan MicroRNA Assays per manufacturers

instructions in triplicate) ranging from low (miR-614) to medium (miR-29b) to high expression (let7). In all cases, we observed a significant and substantial reduction in mature miRNA levels in this *DICER* knock-down background. A one-tailed Student's *t*-test was used to calculate the associated *p*-values. **f,** We achieved ~10-fold *TARBP2* transcript knock-down using siRNA transfections in the *DICER* knock-down background (determined using qPCR in triplicate). We also performed qRT-PCR for *ZNF395* and *APP*, which were significantly up-regulated (expression of APP and ZNF395 was not up-regulated upon *DICER* silencing). **g,** The observed up-regulation of TARBP2 targets is not limited to *ZNF395* and *APP*. As shown here, the sRSE-regulon and TARBP2-bound transcripts showed a similar up-regulation pattern as the *TARBP2* knockdown in LM2 cells (up-regulated transcripts in shDICER background were defined as logFC>0.23). **h,** We achieved ~7-fold TARBP2 knock-down in Dicer-null mouse embryonic fibroblasts[48], which in turn resulted in 2.5-fold and 2-fold up-regulation in *APP* and *ZFP395* (an ortholog of *ZNF395*), respectively. **i,** Gene expression profiling of *TARBP2* knock-down cells showed an enrichment of sRSE-carrying transcripts and the transcripts whose human homologs were bound by *TARBP2* among the up-regulated genes. We used two independent siRNAs (biological replicates) and three siControl-transfected samples as control. Up-regulated transcripts were defined as the set with logFC>0.1 (Illumina MouseRef-8 arrays were used for profiling, and the Lumi package in R was used for normalization). The sRSE-regulon was re-generated for the mouse transcriptome and was found to be significantly enriched among the up-regulated transcripts. The homologous TARBP2-bound transcripts in mice were used to ensure their enrichment among the up-regulated set as well. Error bars indicate s.e.m. *, $P<0.05$, **, $P<0.01$, and ***, $P<0.001$ by a one-tailed Student's *t*-test.

Acknowledgments

We thank C. Alarcon, N. Halberg, N. Pencheva, P. Furlow, and A. Nguyen for technical assistance and comments on previous versions of this manuscript. We thank H. Lee for her help with orthotopic injections. We thank P. Oikonomou for kindly providing the reporter constructs. We thank the Darnell lab for their input on the HITS-CLIP methodology. We are also grateful to M. McManus and Z. Mourelatos for providing Dicer-null cells. We thank C. Zhao, W. Zhang and S. Dewell of the Rockefeller Genomics Resource Center for assistance with next-generation RNA-sequencing and microarray profiling. H.G. is an Anderson Cancer Center Fellow at Rockefeller University. S.F.T. is a Department of Defense Era of Hope Scholar and a Department of Defense innovator award recipient. S.T. was supported by NHGRI (R01 HG003219). We also acknowledge the efforts of ExpO and IGC in providing public access to breast cancer gene expression profiles and annotations.

References

1. Hollams EM, Giles KM, Thomson AM, Leedman PJ. MRNA stability and the control of gene expression: implications for human disease. Neurochemical research. 2002; 27:957–980. [PubMed: 12462398]
2. Wilusz CJ, Wormington M, Peltz SW. The cap-to-tail guide to mRNA turnover. Nature reviews Molecular cell biology. 2001; 2:237–246.10.1038/35067025
3. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nature genetics. 2007; 39:673–677.10.1038/Ng2003 [PubMed: 17401365]
4. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA 10b in breast cancer. Nature. 2007; 449:682–U682.10.1038/Nature06174 [PubMed: 17898713]
5. Song SJ, et al. MicroRNA-Antagonism Regulates Breast Cancer Stemness and Metastasis via TET-Family-Dependent Chromatin Remodeling. Cell. 2013; 154:311–324.10.1016/j.cell.2013.06.026 [PubMed: 23830207]

6. Tavazoie SF, et al. Endogenous human microRNAs that suppress breast cancer metastasis. Nature. 2008; 451:147–152.10.1038/nature06487 [PubMed: 18185580]
7. Xiao CC, Rajewsky K. MicroRNA Control in the Immune System: Basic Principles. Cell. 2009; 136:26–36.10.1016/J.Cell.2008.12.027 [PubMed: 19135886]
8. Goodarzi H, et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature. 2012; 485:264–268.10.1038/nature11013 [PubMed: 22495308]
9. Keene JD. RNA regulons: coordination of post-transcriptional events. Nature reviews Genetics. 2007; 8:533–543.10.1038/nrg2111
10. Dolken L, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. Rna. 2008; 14:1959–1972.10.1261/Rna.1136108 [PubMed: 18658122]
11. Cutroneo KR, Ehrlich H. Silencing or knocking out eukaryotic gene expression by oligodeoxynucleotide decoys. Critical reviews in eukaryotic gene expression. 2006; 16:23–30. [PubMed: 16584380]
12. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research. 2003; 31:3406–3415. [PubMed: 12824337]
13. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome through RNA structure. Nature reviews Genetics. 2011; 12:641–655.10.1038/nrg3049
14. Wang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 2005; 365:671–679.10.1016/S0140-6736(05)17947-1 [PubMed: 15721472]
15. Gatignol A, Buckler-White A, Berkhout B, Jeang KT. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. Science. 1991; 251:1597–1600. [PubMed: 2011739]
16. Gatignol A, Buckler C, Jeang KT. Relatedness of an RNA-binding motif in human immunodeficiency virus type 1 TAR RNA-binding protein TRBP to human P1/dsI kinase and Drosophila staufen. Molecular and cellular biology. 1993; 13:2193–2202. [PubMed: 8455607]
17. Chendrimada TP, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature. 2005; 436:740–744.10.1038/nature03868 [PubMed: 15973356]
18. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008; 456:464–U422.10.1038/Nature07488 [PubMed: 18978773]
19. Korpal M, et al. Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. Nature medicine. 2011; 17:1101–1108.10.1038/nm.2401
20. Kuhn PH, et al. ADAM10 is the physiologically relevant, constitutive alpha-secretase of the amyloid precursor protein in primary neurons. The EMBO journal. 2010; 29:3020–3032.10.1038/emboj.2010.167 [PubMed: 20676056]
21. Tanaka K, Shouguchi-Miyata J, Miyamoto N, Ikeda JE. Novel nuclear shuttle proteins, HDBP1 and HDBP2, bind to neuronal cell-specific cis-regulatory element in the promoter for the human Huntington's disease gene. The Journal of biological chemistry. 2004; 279:7275–7286.10.1074/jbc.M310726200 [PubMed: 14625278]
22. Fan X, et al. miR-20a promotes proliferation and invasion by targeting APP in human ovarian cancer cells. Acta biochimica et biophysica Sinica. 2010; 42:318–324. [PubMed: 20458444]
23. Strieter RM, et al. Interleukin-8. A corneal factor that induces neovascularization. The American journal of pathology. 1992; 141:1279–1284. [PubMed: 1281615]
24. Soria G, et al. Inflammatory mediators in breast cancer: coordinated expression of TNFalpha & IL-1beta with CCL2 & CCL5 and effects on epithelial-to-mesenchymal transition. BMC cancer. 2011; 11:130.10.1186/1471-2407-11-130 [PubMed: 21486440]
25. Singh B, et al. COX-2 involvement in breast cancer metastasis to bone. Oncogene. 2007; 26:3789–3796.10.1038/sj.onc.1210154 [PubMed: 17213821]
26. MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. In vitro reconstitution of the human RISC-loading complex. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:512–517.10.1073/pnas.0710869105 [PubMed: 18178619]
27. De Vito C, et al. A TARBP2-dependent miRNA expression profile underlies cancer stem cell properties and provides candidate therapeutic reagents in Ewing sarcoma. Cancer cell. 2012; 21:807–821.10.1016/j.ccr.2012.04.023 [PubMed: 22698405]

28. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111.10.1093/bioinformatics/btp120 [PubMed: 19289445]
29. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 2010; 28:511–515.10.1038/nbt.1621
30. Eddy SR. Computational Genomics of noncoding RNA genes. Cell. 2002; 109:137–140.10.1016/S0092-8674(02)00727-4 [PubMed: 12007398]
31. Foat BC, Stormo GD. Discovering structural cis-regulatory elements by modeling the behaviors of mRNAs. Mol Syst Biol. 2009; 510.1038/Msb.2009.24
32. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Bio. 2007; 8:479–490.10.1038/Nrm2178 [PubMed: 17473849]
33. Ray D, et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nature biotechnology. 2009; 27:667–670.10.1038/nbt.1550
34. Ray D, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013; 499:172–177.10.1038/nature12311 [PubMed: 23846655]
35. Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all genomes and data types. Molecular cell. 2007; 28:337–350.10.1016/j.molcel.2007.09.027 [PubMed: 17964271]
36. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic acids research. 1981; 9:133–148. [PubMed: 6163133]
37. Kertesz M, et al. Genome-wide measurement of RNA secondary structure in yeast. Nature. 2010; 467:103–107.10.1038/nature09322 [PubMed: 20811459]
38. Oikonomou P, Goodarzi H, Tavazoie S. Systematic Identification of Regulatory Elements in Conserved 3' UTRs of Human Transcripts. Cell Rep. 2014; 7:281–292.10.1016/j.celrep.2014.03.001 [PubMed: 24656821]
39. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 2009; 460:479–486.10.1038/nature08170 [PubMed: 19536157]
40. Giannopoulou EG, Elemento O. An integrated ChIP-seq analysis platform with customizable workflows. BMC bioinformatics. 2011; 12:277.10.1186/1471-2105-12-277 [PubMed: 21736739]
41. Zhang CL, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nature biotechnology. 2011; 29:607–U686.10.1038/Nbt.1873
42. Pencheva N, et al. Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. Cell. 2012; 151:1068–1082.10.1016/j.cell.2012.10.028 [PubMed: 23142051]
43. Hatzis C, et al. A Genomic Predictor of Response and Survival Following Taxane-Anthracycline Chemotherapy for Invasive Breast Cancer. Jama-J Am Med Assoc. 2011; 305:1873–1881.
44. Lu XS, et al. Predicting features of breast cancer with gene expression patterns. Breast cancer research and treatment. 2008; 108:191–201.10.1007/S10549-007-9596-6 [PubMed: 18297396]
45. Hofacker IL, et al. Fast Folding and Comparison of Rna Secondary Structures. Monatsh Chem. 1994; 125:167–188.10.1007/Bf00818163
46. Manley JL. SELEX to identify protein-binding sites on RNA. Cold Spring Harbor protocols. 2013; 2013:156–163.10.1101/pdb.prot072934 [PubMed: 23378656]
47. Hatzis C, et al. A genomic predictor of response and survival following taxaneanthracycline chemotherapy for invasive breast cancer. JAMA : the journal of the American Medical Association. 2011; 305:1873–1881.10.1001/jama.2011.593
48. Harfe BD, McManus MT, Mansfield JH, Hornstein E, Tabin CJ. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:10898–10903.10.1073/pnas.0504834102 [PubMed: 16040801]

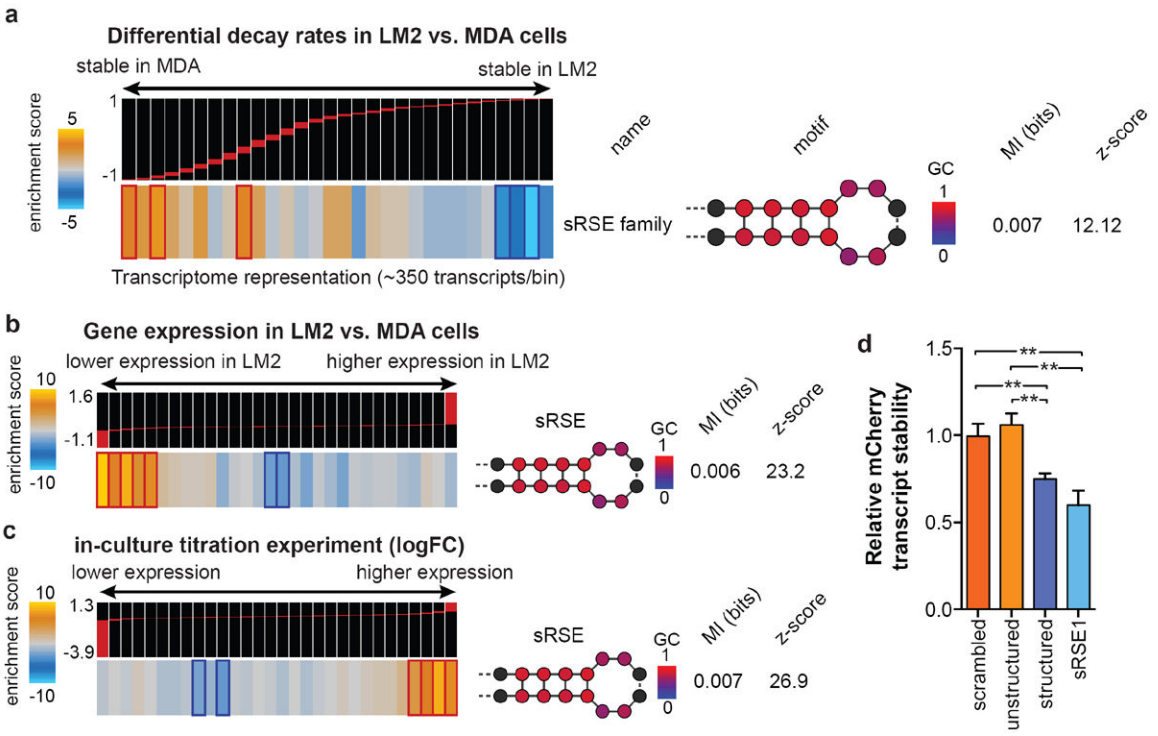


Figure 1. A family of GC-rich structural *cis*-regulatory RNA elements are enriched in transcripts destabilized in metastatic breast cancer cells

a, The sRSE family of mRNA elements is significantly informative of differential transcript stability measurements between the metastatic MDA-LM2 and its parental MDA-231 breast cancer line. Each bin contains differential decay rate measurements for roughly 350 transcripts. From left (more stable in MDA) to right (more stable in MDA-LM2), sRSE-carrying transcripts were enriched among those destabilized in MDA-LM2 cells. In the heatmap representation of enrichment scores[8], gold entries correspond to bins with over-representation of sRSE-carrying transcripts, while blue bins mark under-representation. Statistically significant enrichments and depletions are marked with red and dark-blue borders, respectively. The sRSE elements are collectively depicted as a generic stem-loop with blue and red circles marking nucleotides with low and high GC-content, respectively (black positions are unconstrained regarding the number and identity of nucleotides from these positions onward; also see Methods). Also included are mutual information (MI) values and their associated *z*-scores. **b,** The significant enrichment of sRSE-carrying transcripts among the genes with lower expression in metastatic MDA-LM2 cells relative to the parental MDA. Transcripts were sorted according to the logFC of their expression levels in MDA-LM2 versus MDA cells and divided into equally populated bins from lower expression in MDA-LM2 cells (left) to higher expression (right). **c,** The enrichment of sRSE-carrying transcripts among the genes up-regulated upon transfection of decoy RNA molecules harboring engineered sRSE1 instances compared to scrambled controls. **d,** Transcript stability quantification for mCherry (normalized to GFP as internal control) carrying four different forms of the sRSE, namely, sRSE1, structured mimetic, unstructured mimetic, and scrambled control. α-amanitin treatments at 0- and 8-hr time-points followed by total RNA extraction and cDNA synthesis were used to estimate relative stability

between variants (n=6 per sample per time-point; two independent sets of biological triplicate). Error bars indicate s.e.m. **, $P<0.01$ by a one-tailed Mann Whitney test.

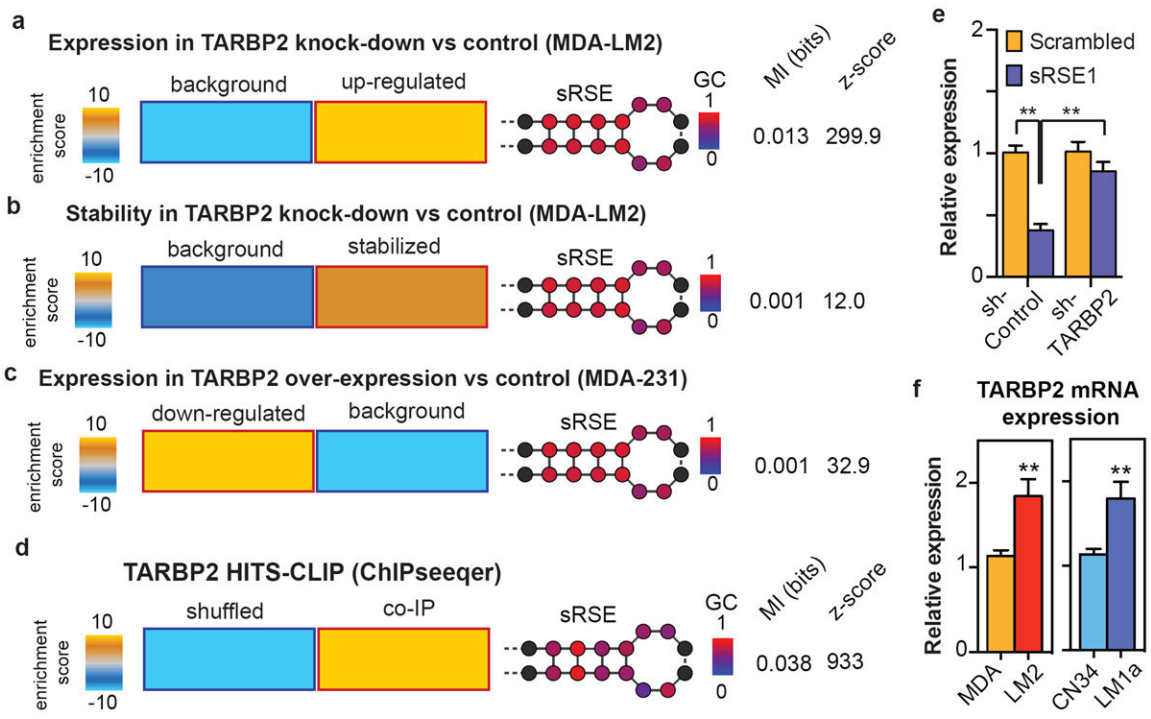


Figure 2. TARBP2 binds and post-transcriptionally destabilizes sRSE-carrying transcripts

a, sRSE-carrying transcripts were enriched among those up-regulated in the RNAi-mediated *TARBP2* knock-down in MDA-LM2 cells (compared to siControl-transfected cells). Transcripts were divided into up-regulated and background groups based on their expression levels in *TARBP2* knock-down cells relative to control (see Methods). **b,** The enrichment of sRSEcarrying transcripts among those whose stabilities were enhanced in *TARBP2* knock-down cells. Samples taken at 0- and 18-hr time-points post α-amanitin treatment were used to estimate relative stability (see Methods). **c,** The sRSE-regulon was enriched among transcripts down-regulated in MDA-LM2 cells over-expressing *TARBP2* (relative to GFP-transfected cells). **d,** Significant enrichment of sRSE elements among the TARBP2 binding sites (determined using HITS-CLIP). **e,** The expression levels of sRSE/scrambled-fused mCherry reporters assayed in control and *TARBP2* knock-down cells. (n=6 per sample; two independent sets of biological triplicates). **f,** Relative *TARBP2* mRNA expression in MDA/MDA-LM2 and CN34/CN-LM1a cells determined using qRT-PCR (n=7 per sample; three independent sets of two biological replicates and a triplicate). In this figure, error bars indicate s.e.m. **, $P<0.01$ by a one-tailed Mann-Whitney test unless otherwise specified.

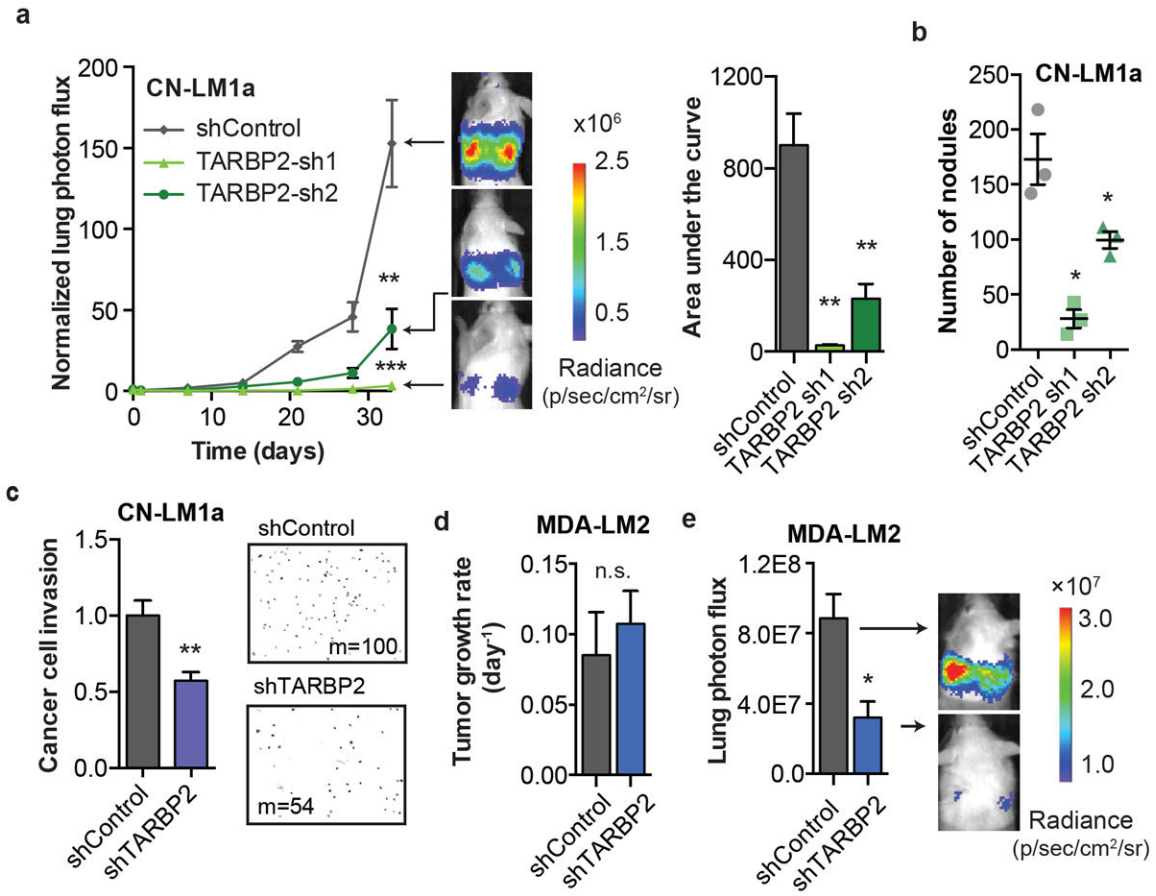


Figure 3. Endogenous TARBP2 promotes metastatic colonization

a, Bioluminescence imaging plot of lung metastasis by CN-LM1a cells expressing short hairpins targeting *TARBP2* (shTARBP2) or a control hairpin (shControl); n=6 in each cohort. Area-under-the-curve was also calculated for each mouse. **b,** The number of nodules recorded per lung section for three representative mice in each cohort (n=3 per sample, extracted at day 33). *P*=0.05 by a one-tailed Mann-Whitney test. **c,** Cell invasion capacity of shTARBP2 and shControl cells in the CN-LM1a background quantified using trans-well matrigel invasion assays (normalized to shControl); n=8 for each sample comparing shTARBP2 to shControl cells (two independent sets of biological quadruplicates). Also shown are representative images of trans-well inserts along with the median count (m) for each experiment. **d,** Tumour growth rate for LM2 cells expressing shTARBP2 or shControl injected into the mammary fat pads of mice (day^{-1}; n=8 and 6 respectively). **e,** Lung bioluminescence (seven days post tumour extraction) measured *in vivo* (n=3 and 4 for shControl and shTARBP2 respectively in MDA-LM2 background). Error bars indicate s.e.m. *, *P*<0.05 and **, *P*<0.01 by a one-tailed Mann-Whitney test unless specified otherwise.

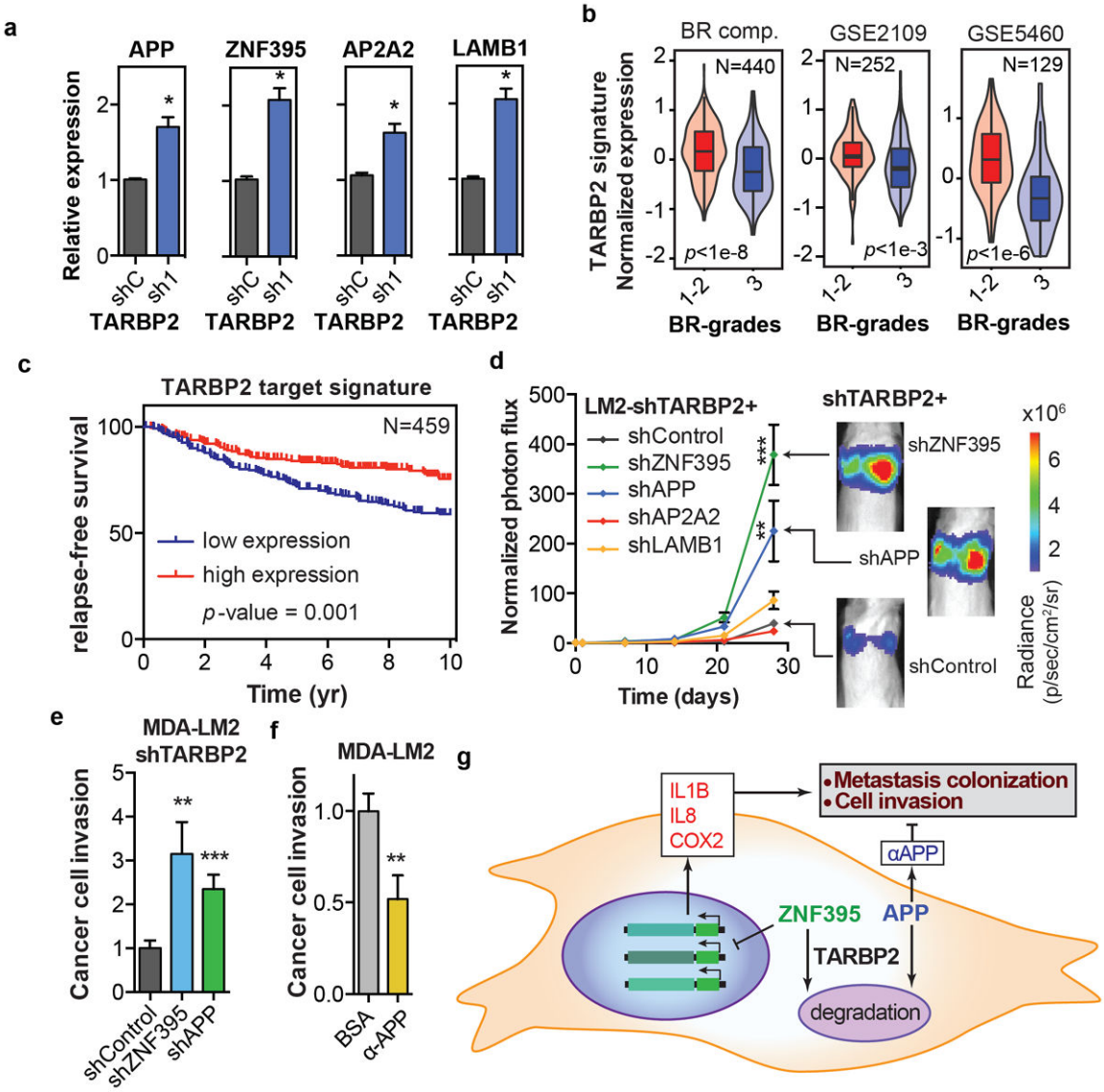


Figure 4. TARBP2 promotes metastatic cell invasion and colonization by destabilizing *APP* and *ZNF395* transcripts

a, Relative expression of TARBP2 target transcripts *APP*, *ZNF395*, *AP2A2*, and *LAMB1* in *TARBP2* knock-down cells versus control (MDA-LM2 background); n=4 per sample (two independent sets of biological replicates). **b,** Distribution of expression levels for the TARBP2 target signature (defined as the aggregate expression of the four targets) in tumours with BR grades 1 and 2 compared to 3 in three separate datasets (BR compendium[19], expO compendium/GSE2109, and GSE5460). Box plots: the bottom and the top of the box are the first and third quartile, respectively, and the central line is the median; probability density plots are superimposed. **c,** Kaplan–Meier curves for a compendium of breast cancer patients[19] (n=459) showing metastasis-free-survival as a function of TARBP2 target aggregate expression. *P*-value is based on a Mantel-Cox log-rank test. **d,** Bioluminescence imaging plot of lung metastasis by cells expressing short hairpins targeting one of four TARBP2 targets or a control hairpin (shControl) in MDA-LM2 shTARBP2

background; n=5 in each cohort except for shZNF395, in which n=4. Also shown are bioluminescence images from representative mice. **e,** Cell invasion capacity measured for *APP* and *ZNF395* knockdown cells (relative to shControl) in MDA-LM2 shTARBP2 background; n=8 for each sample (two independent sets of quadruplicates). **f,** Cell invasion capacity measured for MDA-LM2 cells with exogenously added APPα relative to BSA as control; n=8 for each sample (two independent sets of quadruplicates). **g,** Schematic of TARBP2-mediated enhancement of invasion and metastatic colonization through *APP* and *ZNF395* transcript destabilization. Error bars indicate s.e.m. *, $P<0.05$, **, $P<0.01$, and ***, $P<0.001$ by a one-tailed Mann-Whitney test.