Supplementary Materials

Systematic Discovery of Structural Elements Governing Mammalian mRNA Stability

Hani Goodarzi[1,2,7], Hamed S. Najafabadi[3,4,8], Panos Oikonomou[1,2,7], Todd M. Greco[2], Lisa Fish[6], Reza Salavati[3,4,5], Ileana M. Cristea[2], Saeed Tavazoie[1,2,7,*]

[1]Lewis-Sigler Institute for Integrative Genomics, [2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08540, USA

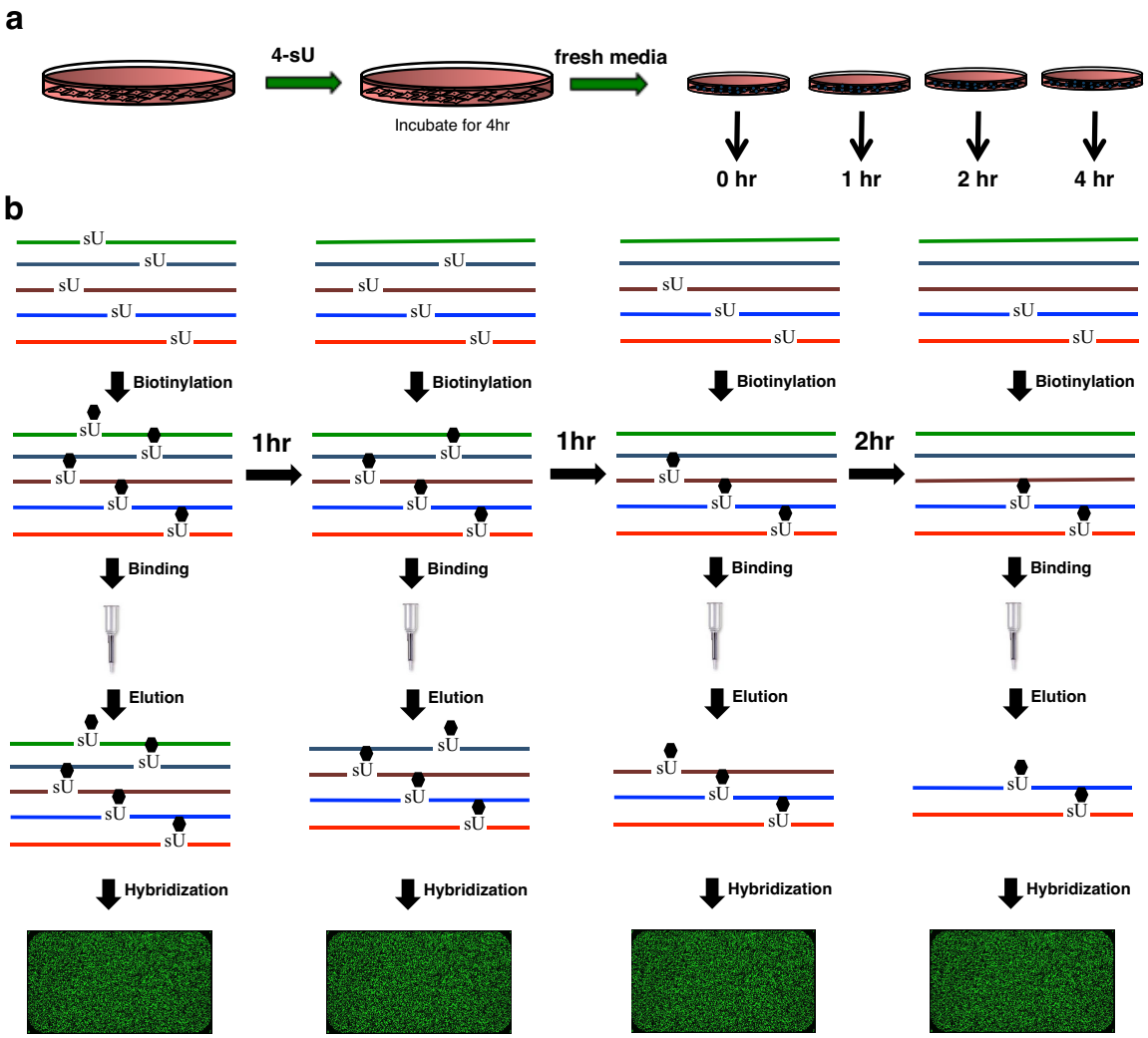
[3]Institute of Parasitology, [4]McGill Centre for Bioinformatics, [5]Department of Biochemistry, McGill University, Montreal, Quebec H3G1Y6, Canada

[6]Laboratory of Systems Cancer Biology, Rockefeller University, New York, NY 10065, USA.

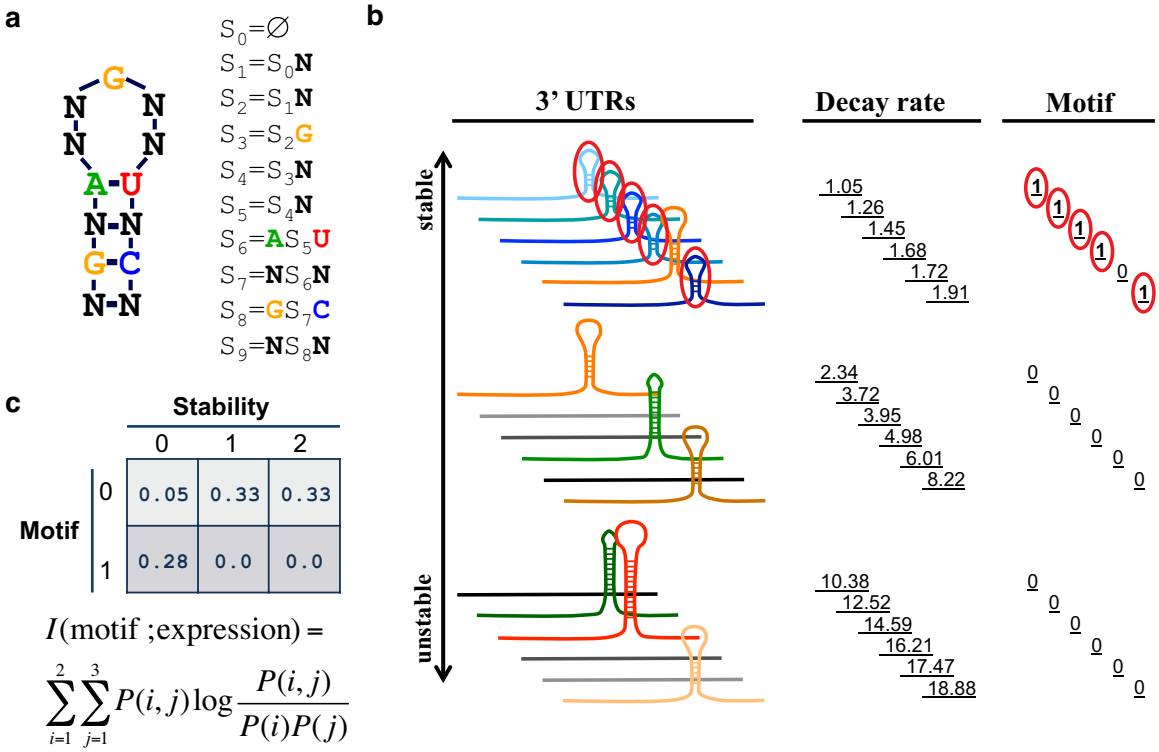
[7]Present address: Department of Biochemistry and Molecular Biophysics & the Initiative in Systems Biology, Columbia University, New York, NY 10032, USA

[8]Present address: The Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S3E1, Canada

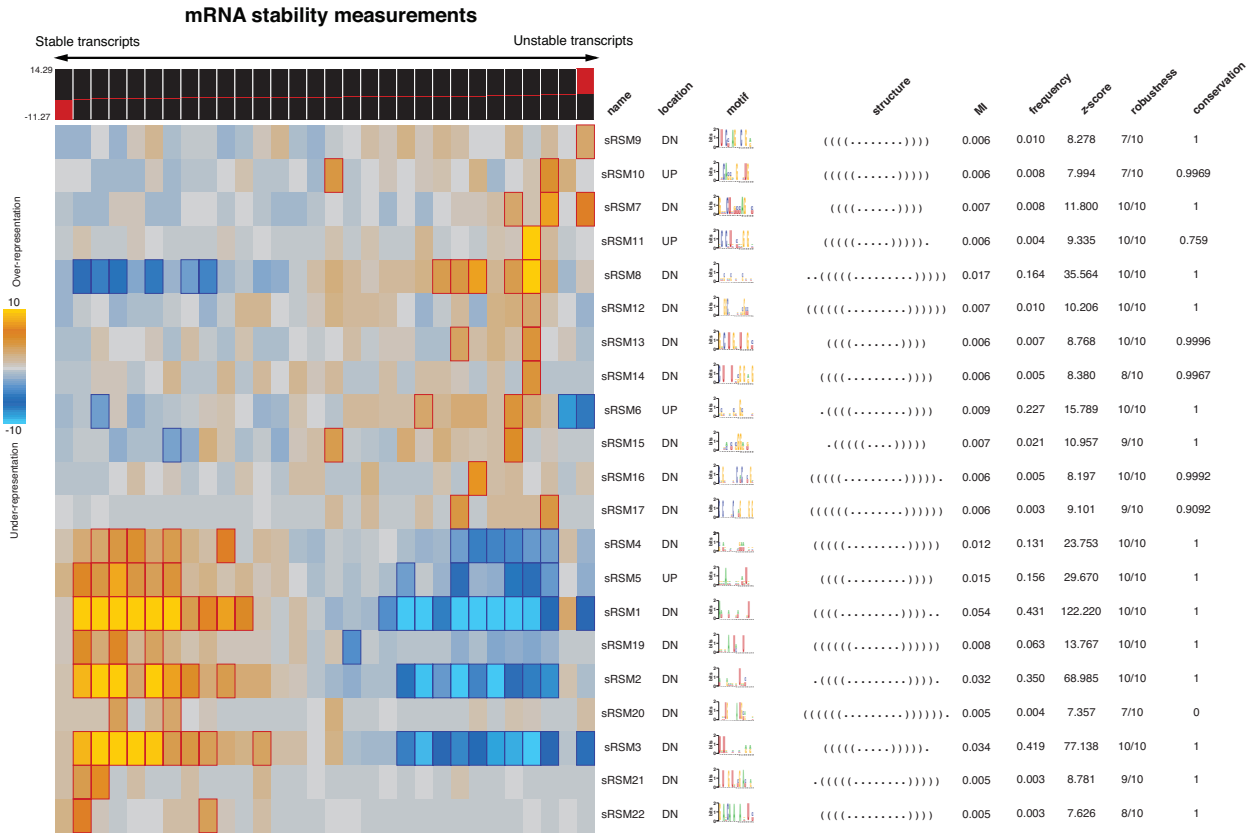
[*]To whom correspondence should be addressed: st2744@columbia.edu



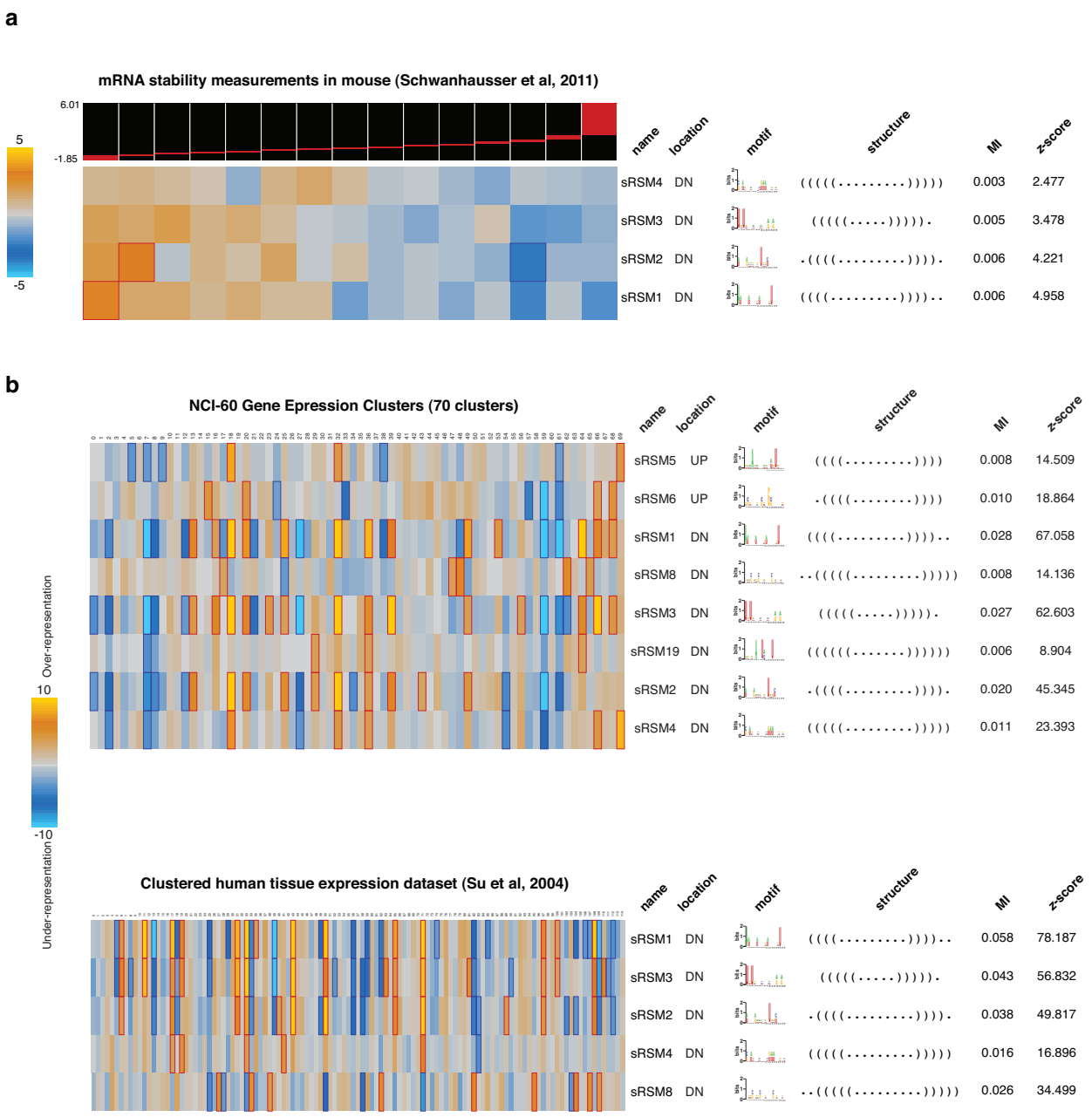
Supplementary Figure 1. mRNA stability measurements. a, Genome-wide transcript stability measurements are acquired through pulse-chase experiments using 4-thiouridine incorporation into cellular RNA[1]. The biotinylation of the 4sU labeled transcripts enables us to distinguish between the newly synthesized mRNAs and the ones transcribed during the pulse period in the presence of 4sU in the media. For this, cells were grown in the presence of 25μM 4sU for 4 hours after which they were incubated in fresh media with excess Uridine (without 4sU). Samples were taken at 0, 1, 2 and 4 hours after 4sU removal, during which 4sU-containing transcripts are degraded and replaced by newly synthesized (unlabeled) mRNAs. **b,** Upon RNA extraction from each sample, the thiol groups on 4sU were biotinylated (EZ-Link Biotin-HPDP, Pierce) and captured using streptavidin micro-beads (Miltenyi Biotec)[1]. The resulting RNA samples were labeled (single-color using low-input quick-amp labeling kit, Agilent) and hybridized to whole-genome human microarrays (Agilent's 4×44 human gene expression arrays). The rate at which the signal drops for each transcript was used as a measure of its decay rate (r): $r = -\ln\left(\frac{S_t}{S_0}\right)\Big/t$.



Supplementary Figure 2. Structural motif discovery schematic. a, Structural motifs are produced based on a context-free grammar (CFG) that represents short stem-loop structures along with a known primary sequence. CFGs provide an efficient representation of RNA secondary structures (see Methods for details). **b,** TEISER initially relies on a large number of such structural seeds (see Methods). The RNA sequences, both 5' and 3' UTRs, are then scanned for the occurrence of these motifs. The result is a binary vector denoting the presence and absence of a given motif. These vectors are then tested to assess whether they are informative of the input whole-genome measurements. **c,** This dependency is captured through the notion of Mutual Information (MI). Mutual information is a non-parametric measure, which can be defined for both continuous and discrete variables, hence providing a framework that is generally suitable for all types of whole-genome measurements, *e.g.* single microarray measurements (continuous) or co-expression clusters (discrete). Upon finding the structural motifs that are informative of the input data, the motifs are further refined and elongated based on a greedy algorithm, which requires the MI values to increase throughout the optimization process. In the end, the robustness of the identified motifs is also evaluated through a three-fold jackknifing test where only two-thirds of the data is used to determine the significance of the given motifs. The jackknifing is repeated 10 times and robustness is reported as x/10 where x is the number of trials in which the motif maintained its significance in the partial dataset. In order to provide a graphical representation of the results, hypergeometric distribution is utilized to calculate the *p*-value associated with the enrichment or depletion of a predicted element in a given bin or cluster[2]. In this study, we have also required a significant Spearman correlation ($p<0.001$) between the stability bins and the enrichment/depletion scores (calculated based on log-transformed hypergeometric *p*-values). This step ensured a very low false-discovery rate for the discovered structural elements. The following single letter nucleotide code was also used to illustrate structural motifs: Y=[UC], R=[AG], K=[UG], M=[AC], S=[GC], W=[AU], B=[GUC], D=[GAU], H=[ACU], V=[GCA] and N=any nucleotide. TEISER is available online at https://tavazoielab.c2b2.columbia.edu/TEISER and through our iGET portal at https://iget.c2b2.columbia.edu/.



Supplementary Figure 3. TEISER graphical output. Shown is the list of structural elements that are informative of RNA stability in MDA-MB-231 cells under static conditions. This list is generated at FDR of 0.1 compared to the list in Fig. 1 (main text) with associated FDR of <0.01. In this figure, the robustness threshold is set to 6/10 and the Spearman correlation *p*-value to 0.01 (for Fig. 1 the Spearman *p*-value cut-off is 0.001). The output here contains an id, the location of the motif (UP for 5'UTR and DN for 3'UTR), the primary sequence of the motif in a weblogo[3] format and its secondary structure. The calculated mutual information values (MI, see Suppl. Fig. S2c) are also shown. The frequency column shows the fraction of the transcripts that carry an instance of the given motif. We also provide *z*-scores associated with a randomization-based statistical test and a robustness score from a three-fold jackknifing test[2]. Network-level conservation[4] between human and mouse genomes was used to provide conservation scores ranging from 0 to 1 (1 being highly conserved). For this, orthologous pairs of transcripts (from mouse and human genomes) were scanned for the presence of each motif. The mutual information between the resulting binary vectors was used as a measure of conservation. For each motif, the associated *p*-value was calculated using a randomization based statistical test and the corresponding conservation score was reported as 1-*p*.



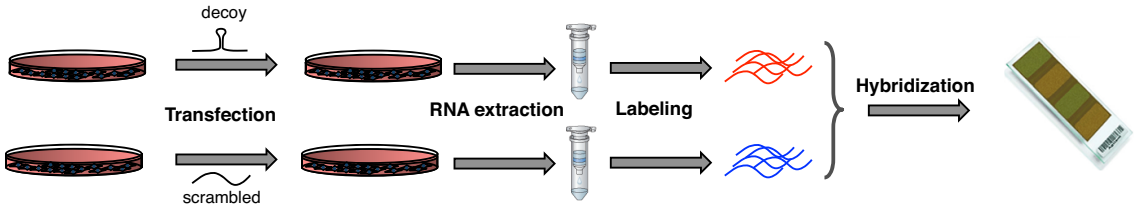
Supplementary Figure 4. Predicted RNA structural stability motifs convey information across diverse datasets. a, Structural motifs identified in the RNA stability dataset were tested on a small dataset of mRNA stability measurements in mouse[5] by applying TEISER in non-discovery mode. As shown here, four of the eight elements are significantly informative of mRNA stability highlighting the conservation of these elements and their functionality across distant mammals. **b,** We also tested the identified motifs in other human whole-genome datasets[6,7]. For each motif, we looked at the expression of its target genes across different co-expression clusters. Listed are the structural motifs that showed significant mutual information with these datasets. Robust enrichment and depletion patterns for many of these motifs suggest a strong underlying regulatory program for their target genes and provides additional support for their functionality. It should be noted that random motif representations, obtained through random shuffling of the CFG statements from the original motifs, are not significantly informative of any of these datasets.

Set 1

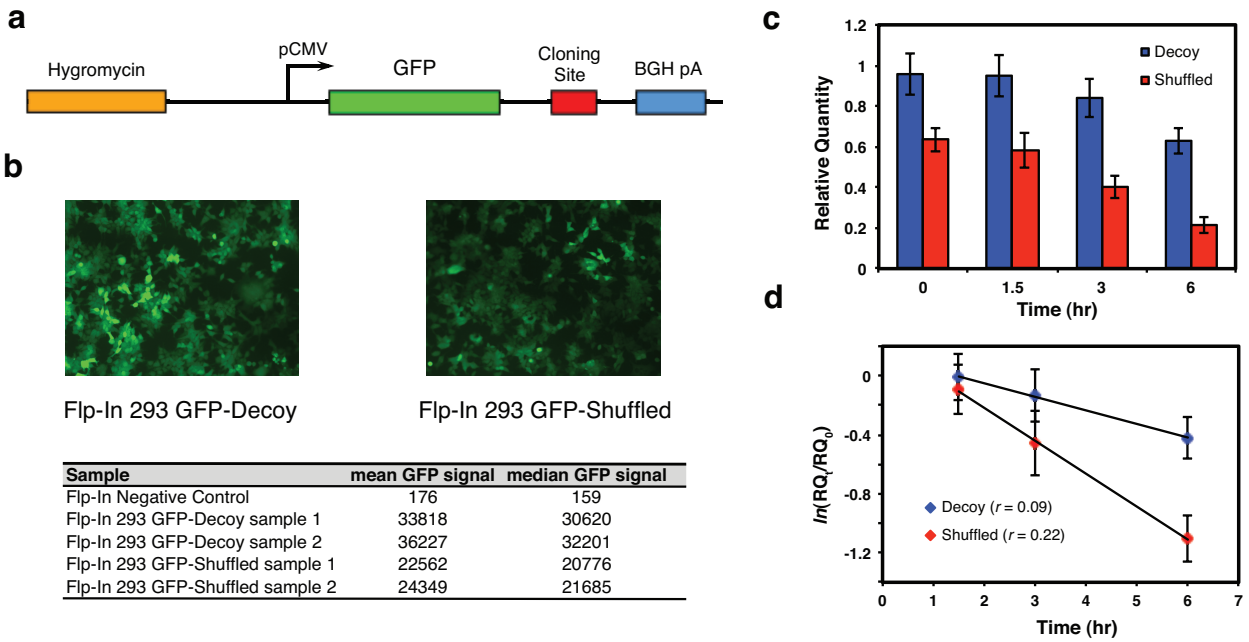
AAAACTATTTTGAAGATGGTGGTGA**GCTGCAAAATAGCTGGATGGATTTGAATGATTGGGATGATACATCATTGAACTGCA**
CTTTATATAACCAAAGCTTAGCAGTT**TGTTAGATAAGAGTCTATGTATGTCTCTGGTTAGGATGAAGTTAATTTTATGTTT**
TTAACATGGTATTTTTGAAGGAGCTAATGAAACACTGG
AAAACTATTTTGAAGATGGTGGTGA**TAGGGACTACGTCATTAGCTAAAGCCATGGGCATTTGAGAGATGGAACATTATTAT**
TGTTATATAACCAAAGCTTAGCAGTT**GAAGTGTTGTATTTTATATAGTGATGCATGTATCATTTATGTATGAGAGATTTAT**
TCGTCATGGTATTTTTGAAGGAGCTAATGAAACACTGG

Set 2

ATTGTTTC**TGGAAACTGCTTGCCAAGACAACATTTATTAACTGTTAGAACACTTGCTTTATGTTTGT**GTGTACATATTTTC
CACAAATGTTATAATTTATATAGTGTGGTTGAACAGGATGCAATCTTTTGTTG**TCTAAAGGTGCTGCAGTTAAAAAAAAAA**
CAACCTTTTCTTTCAATATGGCATGTAGTGGAGTTTTT
ATTGTTTC**TGCTGAAAGTTTCACCTGTGATTGAATAAACAGCTTTTCTTTTCAGAGTCATGTACATA**GTGTACATATTTTC
CACAAATGTTATAATTTATATAGTGTGGTTGAACAGGATGCAATCTTTTGTTG**CTTGCGCAACTCGTTACATATTGCTGCA**
AATTATAAATTTGGAAATTAAATGACGGAGAAGTTTTT



Supplementary Figure 5. Synthetic-decoy titration experiments. Sequences with real (black) and shuffled (red) instances of sRSM1 structural motif were synthesized. The experiment was designed with two different sets of sequences each carrying two instances of sRSM1 in their native context with ~25 nucleotides on either side (bold sequences). The scrambled sequences were designed so that the backbone sequences remained identical and only the motif instances (bold sequences) were shuffled. Upon placing these constructs downstream of a T7 promoter, we synthesized decoy and scrambled RNAs (capped and polyadenylated) that were subsequently transfected into MDA-MB-231 cells. After 48 hours, we performed RNA extraction, labeling, and two-color hybridization in duplicates.



Sample	mean GFP signal	median GFP signal
Flp-In Negative Control	176	159
Flp-In 293 GFP-Decoy sample 1	33818	30620
Flp-In 293 GFP-Decoy sample 2	36227	32201
Flp-In 293 GFP-Shuffled sample 1	22562	20776
Flp-In 293 GFP-Shuffled sample 2	24349	21685

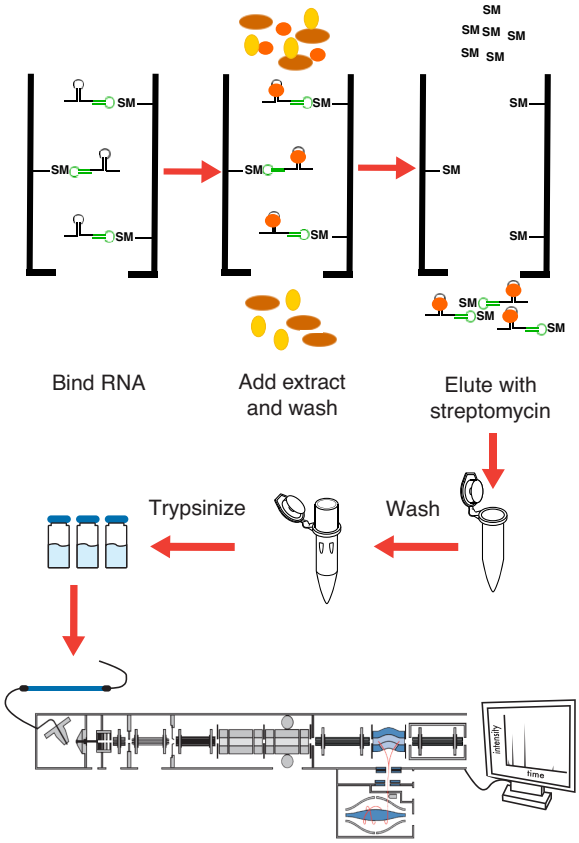
Supplementary Figure 6. The effect of sRSM1 on stability and expression of a reporter gene. a, We constructed a GFP reporter system in the pcDNA5/FRT/TOPO backbone (Invitrogen) containing a gateway cloning site in its 3' UTR. We cloned real instances of sRSM1 (decoy) downstream of GFP as well as shuffled sequences (scrambled) as controls (see Supplementary Figure 5). We then used the Flp-In methodology to stably clone these reporters into Flp-In 293 cell line (Invitrogen). **b,** GFP expression is visibly higher in the cell line that contains sRSM1 cloned downstream of GFP (Flp-In 293 GFP-Decoy) compared to the shuffled control (Flp-In 293 GFP-Shuffled). Accordingly, flow cytometric measurements also showed a 50% increase in the GFP signal in the decoy cell line (Table). **c,** To measure the stability of GFP in the presence and absence of sRSM1, we used α-Amanitine mediated inhibition of RNA polymerase II. Upon incubation in the presence of 5μg/mL Amanitine, we extracted RNA at 0, 1.5, 3 and 6 hour time-points in duplicates (biological replicates). We then performed qPCR reactions (Fast SYBR Green Master Mix, Applied Biosystems) with primers against GFP and 18S rRNA as endogenous controls (in quadruplicates). Relative quantity of GFP transcripts was then reported based on ΔΔCt values (compared to 18SrRNA normalized at time 0 for Flp-In 293 GFP-Decoy). For each decoy and shuffle set, 8 relative quantities were reported at each time point (2 biological replicates × 4 technical replicates). In accordance with FACS measurements, at time 0 hr, the abundance of GFP is ~50% higher in the presence of sRSM1 (p-value < 1e-5) highlighting the regulatory effect of this structural motif. **d,** For each run, $\ln(RQ_t/RQ_0)$ was plotted as a function of time (where RQ_t is Relative Quantity at time t) and the -1×slope of the line was reported as the decay rate (r). The decoy runs showed, on average, more than 2 fold increase in their stability (p-value < 1e-5). These observations support the model that the upregulation of the GFP transcript in the decoy construct is due to the presence of sRSM1 and its contribution to increased stability.

a

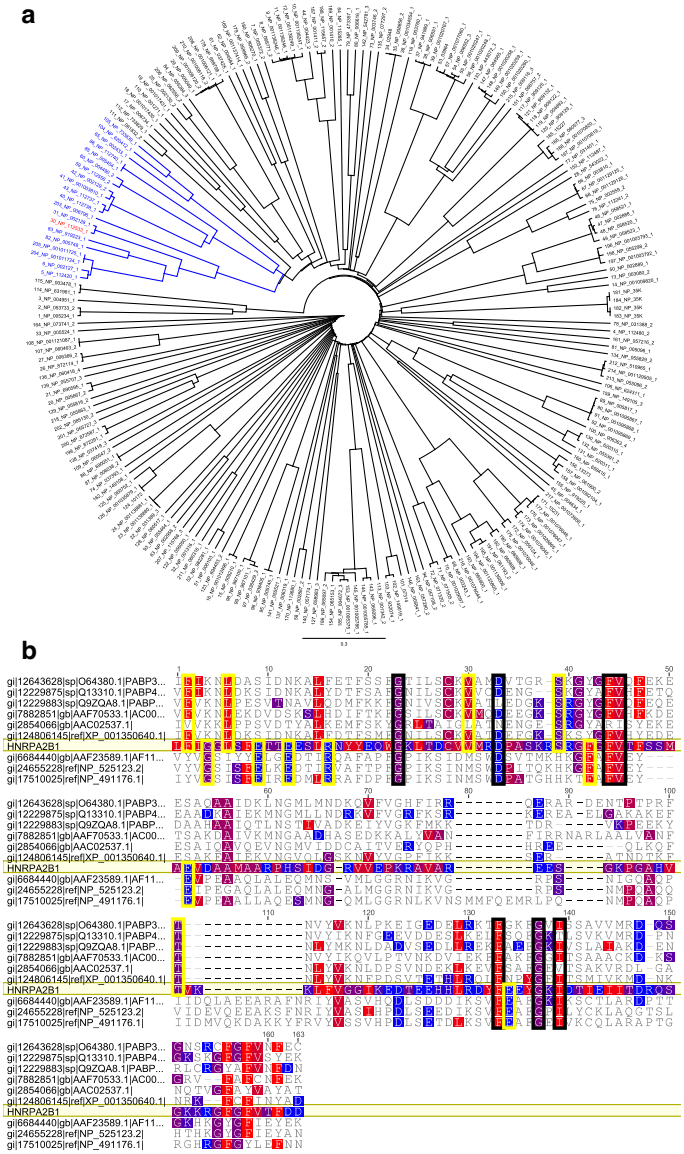
AAAACTATTTTGAAGATGGTGGT
GA**GCTGCAAAATAGCTGGATGGA**
TTTGAATGATTGGGATGATACAT
CATTGAACTGCACTTTATATAAC
CAAAGCTTAGCAGTT**TGTTAGAT**
AAGAGTCTATGTATGTCTCTGGT
TAGGATGAAGTTAATTTTATGTT
TTTAACATGGTATTTTTGAAGGA
GCTAATGAAACACTGG**GGATCGC**
ATTTGGACTTCTGCCCGCAAGGG
CACCACGGTCGGATCC

TCCACGAGGGCAGGGCCTCT**CGG**
TTCGTGTGGCGCGCGAAAGCCGA
GTTGGGCCAGACGCCCTTCTGCC
TGTGCACTTTCCCTGTGGGGTCT
AGTCAAGACAGTCAGGACCTGAC
CCGTCTC**TTCGTATCGCCTGCAC**
CCAGCGTTGACGTGTGGTGAACC
GCGCGCCAGGTCGCTAGGGCAGA
AACCAGCAACTCCCTC**GGATCGC**
ATTTGGACTTCTGCCCGCAAGGG
CACCACGGTCGGATCC

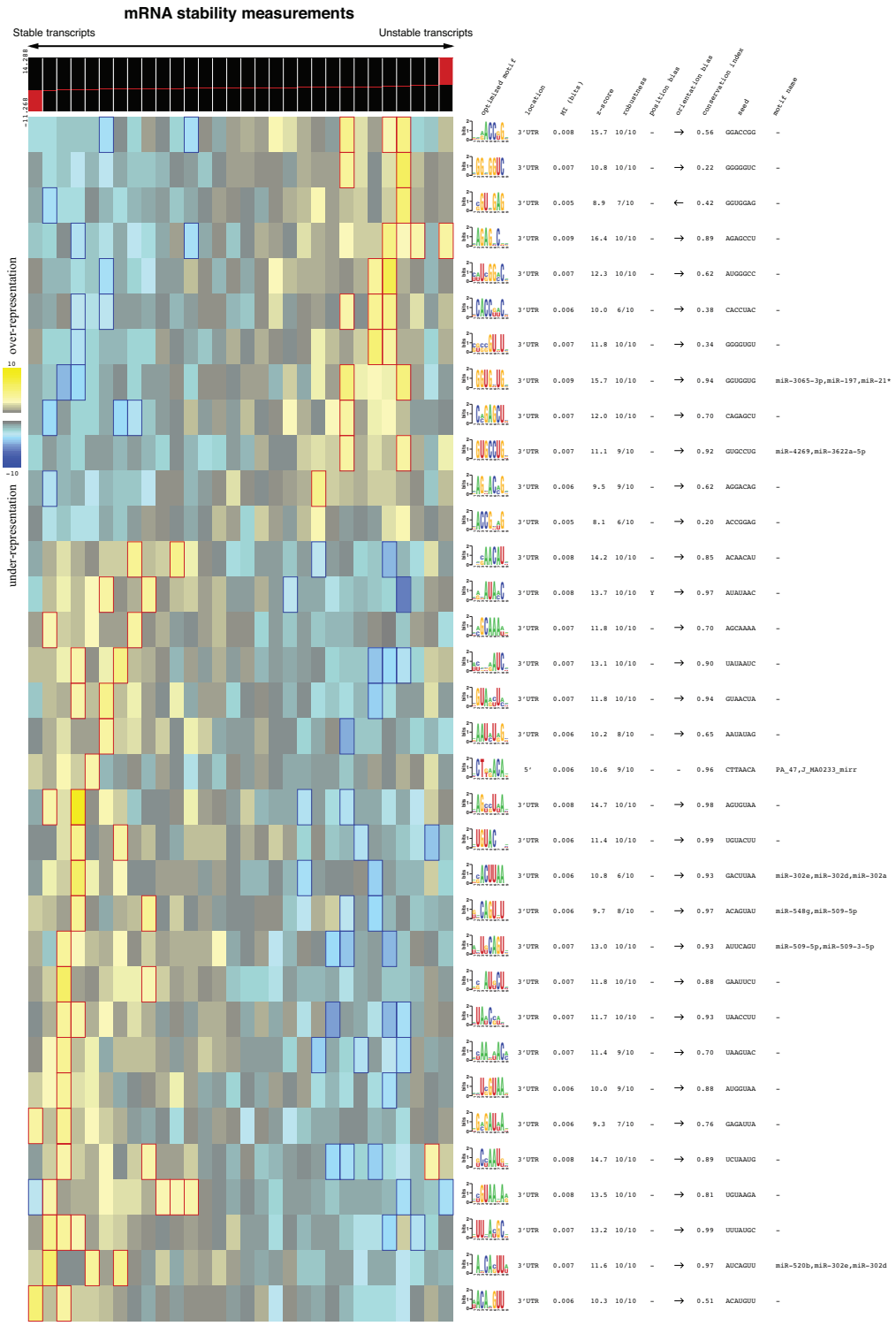
b



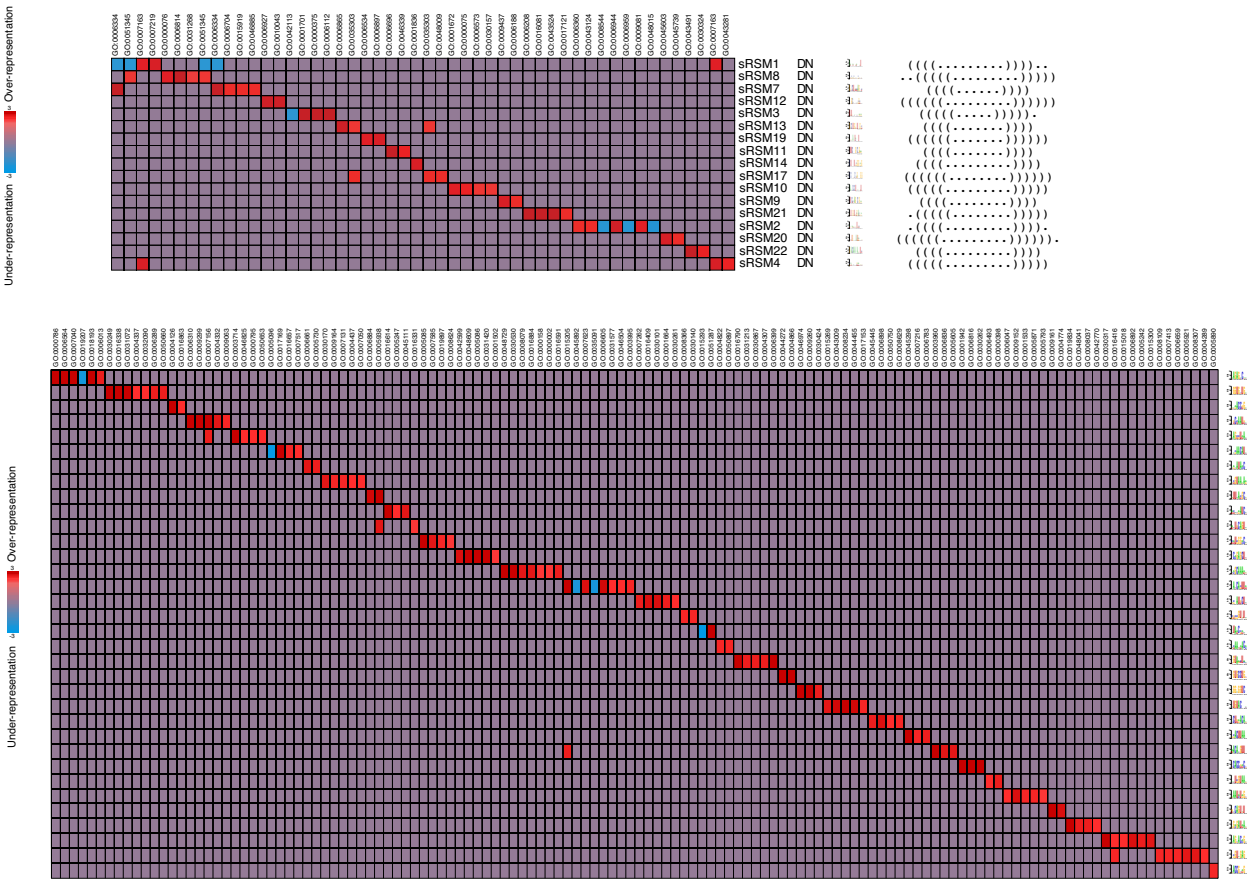
Supplementary Figure 7. Streptomycin-binding RNA aptamer immobilization coupled with mass spectrometry. a, The decoy and scrambled sequences from Set 1 (Supplementary Fig. 5) were used for affinity purification of potential binding candidates. For this, StreptoTag, a streptomycin-binding sequence, was added to the 3' end of the sequences (the green sequence). **b,** A previously published method[8] was employed to identify the candidates that bind the sRSM1 structural motif. The RNA sequences carrying StreptoTag were immobilized on a streptomycin-coated column. Then, protein extract was added, washed and the binding proteins were eluted using streptomycin. The resulting protein sample was precipitated, trypsinized and analyzed by mass-spectrometry (see online Methods).



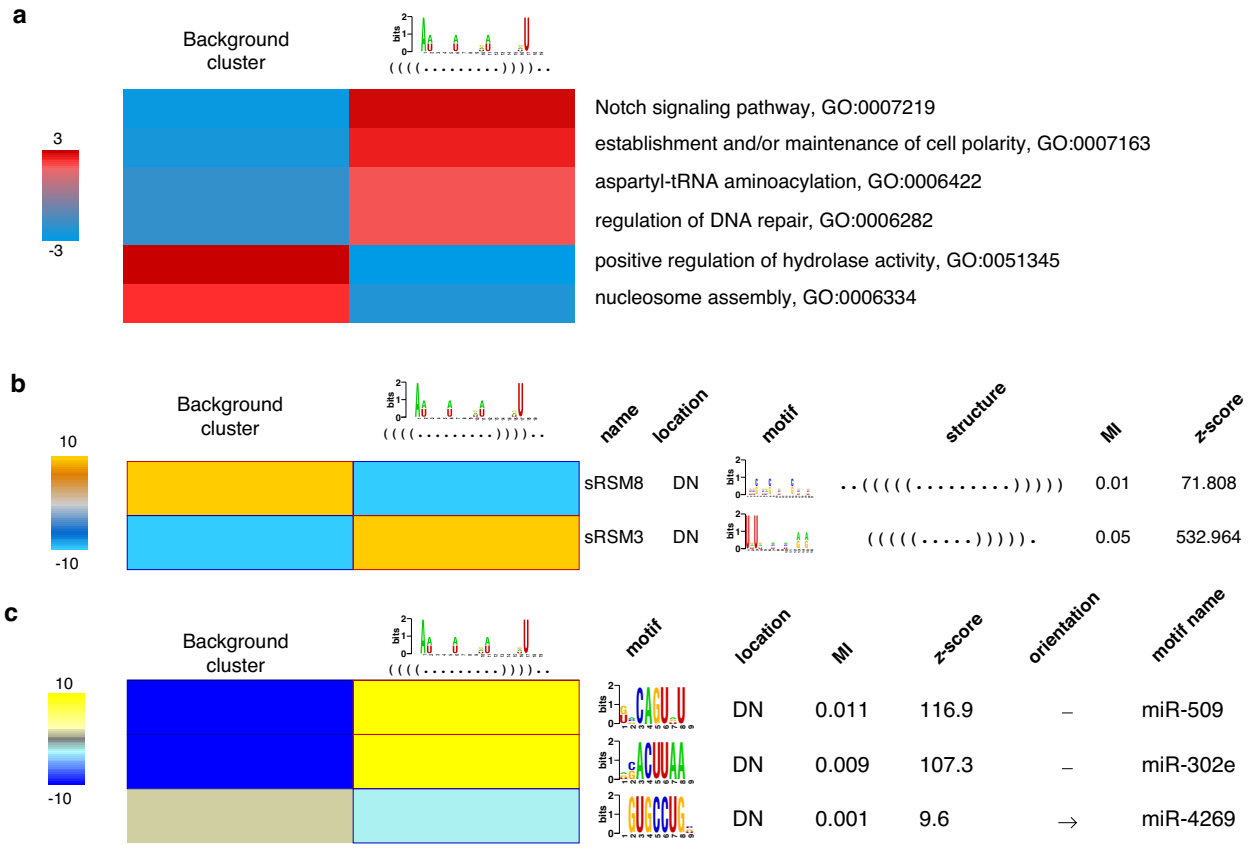
Supplementary Figure 8. Phylogenetic analysis of HNRPA2B1. a, Phylogenetic analysis groups HNRPA2B1 together with heterogeneous nuclear ribonucleoproteins. Blue branches represent a distinct cluster of proteins that almost entirely consists of heterogeneous nuclear ribonucleoproteins, which are all RNA-binding proteins, except for NP_733839, NP_620412 and NP_002433. Only human proteins with similarity to HNRPA2B1 (BLASTP E-value<1) are included in this tree. HNRPA2B1 is highlighted in red. **b,** The sequence of HNRPA2B1 matches the signatures of two RNA-binding motifs: poly-A binding protein types 1, 2, 3, 4 family (TIGR01628) and poly-U binding splicing factor half-pint family (TIGR01645). Sequence comparison of these two families suggests potential HNRPA2B1 residues that are involved in specific binding to A or U as well as residues that potentially contribute to RNA-binding capability of HNRPA2B1 regardless of the RNA sequence composition. In this alignment, proteins that are above HNRPA2B1 belong to polyadenylate binding protein types 1, 2, 3, 4 family (TIGR01628), and proteins below HNRPA2B1 belong to poly-U binding splicing factor half-pint family (TIGR01645). Yellow boxes represent residues that are highly conserved in at least one of the two families, are similar to HNRPA2B1, and are different between the two families. Thus, yellow boxes above HNRPA2B1 highlight residues that are potentially involved in binding to A nucleotides, and yellow boxes below HNRPA2B1 represent potential U-binding residues. Black boxes are used to mark residues that are conserved and similar to HNRPA2B1 in both families, likely contributing to RNA-binding capability of HNRPA2B1 in general.



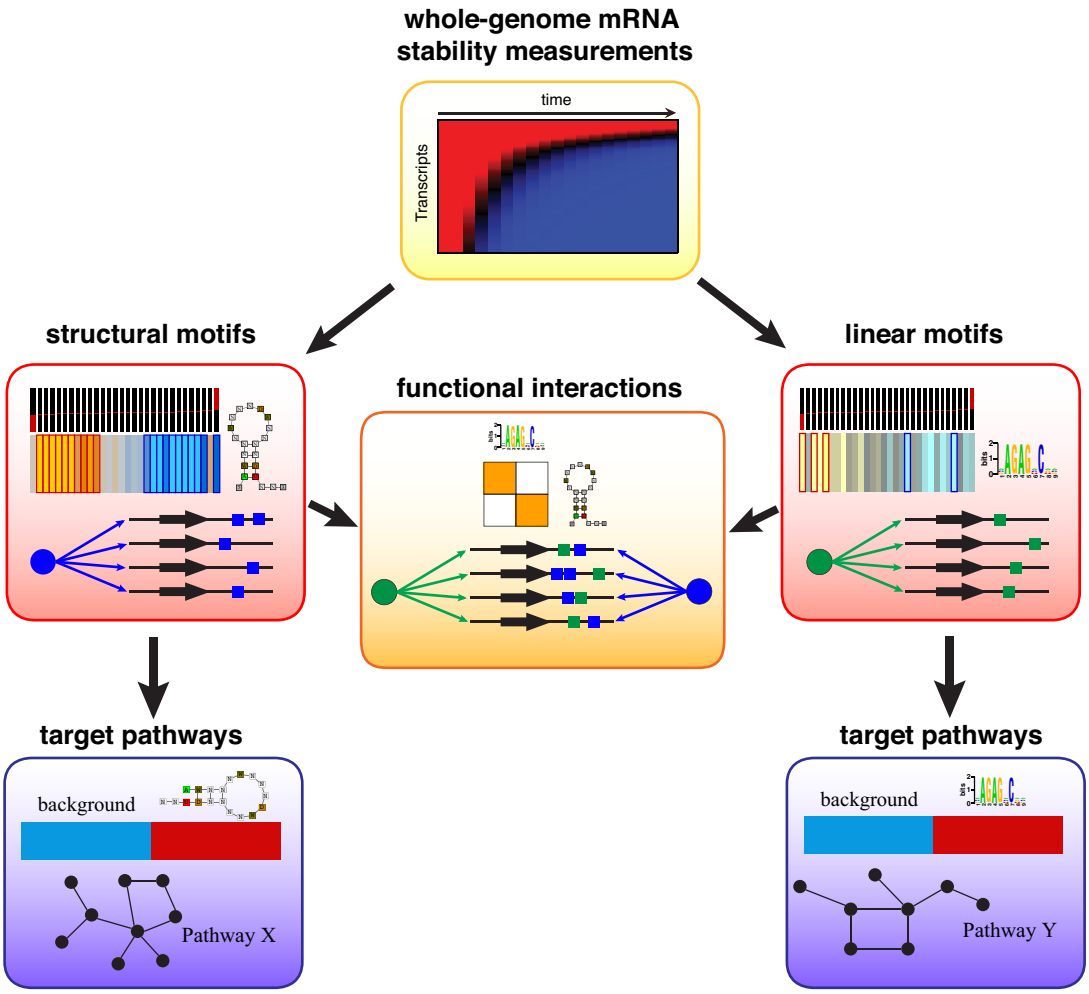
Supplementary Figure 9. Discovering linear RNA structural motifs (lRSMs). FIRE[2] discovered a large set of informative linear motifs (lRSM for linear RNA Stability Motif) that potentially contribute to mRNA stability. The majority of these motifs are in the 3' UTR sequences, reaffirming the post-transcriptional nature of these elements. It should be noted that FIRE's false-discovery rate is very low and these elements are highly significant. In addition to MI, *z*-score and robustness of the motifs, FIRE also tests for orientation and/or position biases across the target transcripts.



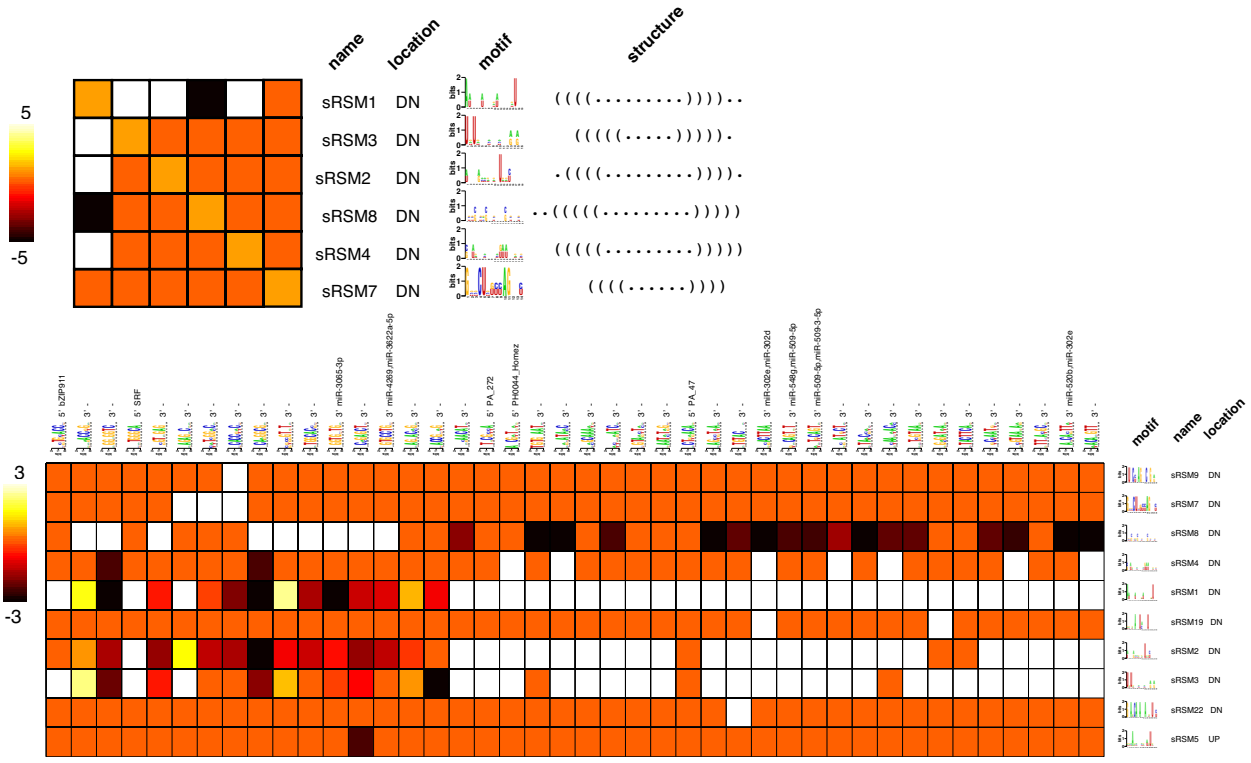
Supplementary Figure 10. Identifying the pathways likely targeted by the discovered regulatory elements. iPAGE[9] was used to identify the pathways that are over-represented among the putative targets of each motif. The top heatmap shows these targets for the structural motifs identified by TEISER (Supplementary Figure 3), whereas the bottom one shows the target pathways for linear motifs identified by FIRE (Supplementary Figure 9).



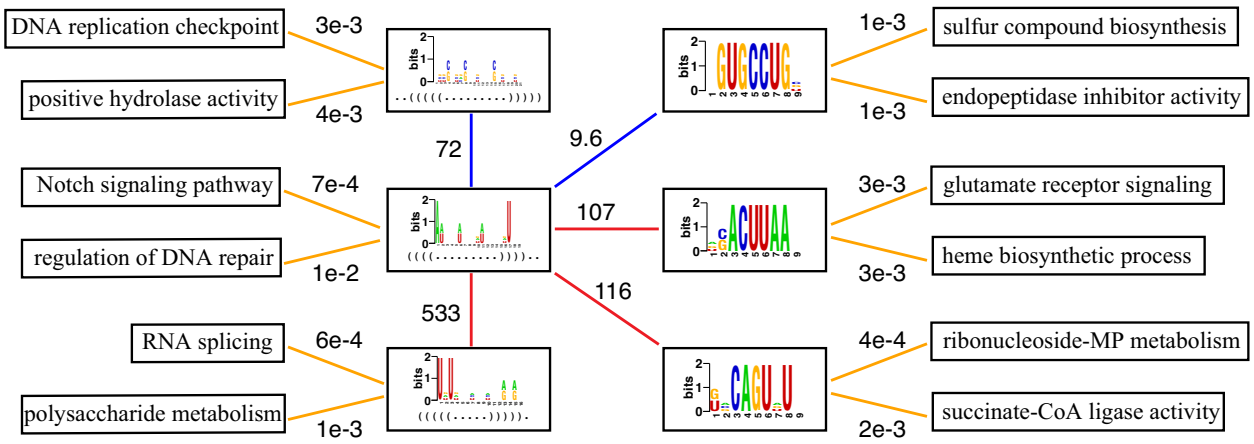
Supplementary Figure 11. Discovering motif-motif and motif-pathway interactions. a, We have grouped the transcripts that carry instances of sRSM1 motif in their 3' UTR into one cluster, while placing all the rest in a background cluster. iPAGE[9] was employed to discover the pathways that are likely targeted by the sRSM1 regulatory elements. The red/blue entries show the enrichment/depletion of the pathways in the corresponding clusters. **b,** TEISER was then used to assess interactions between sRSM1 and other structural motifs. This figure shows two such interactions between sRSM1 and sRSM8 and also sRSM1 and sRSM3. **c,** Similarly, FIRE[2] was used to discover significant interactions between structural motifs and the linear motifs discovered by FIRE. This figure shows three such motifs, interacting positively or negatively with sRSM1. Yellow entries denote the enrichment of the associated motif in the corresponding cluster; whereas blue entries denote significant depletion.



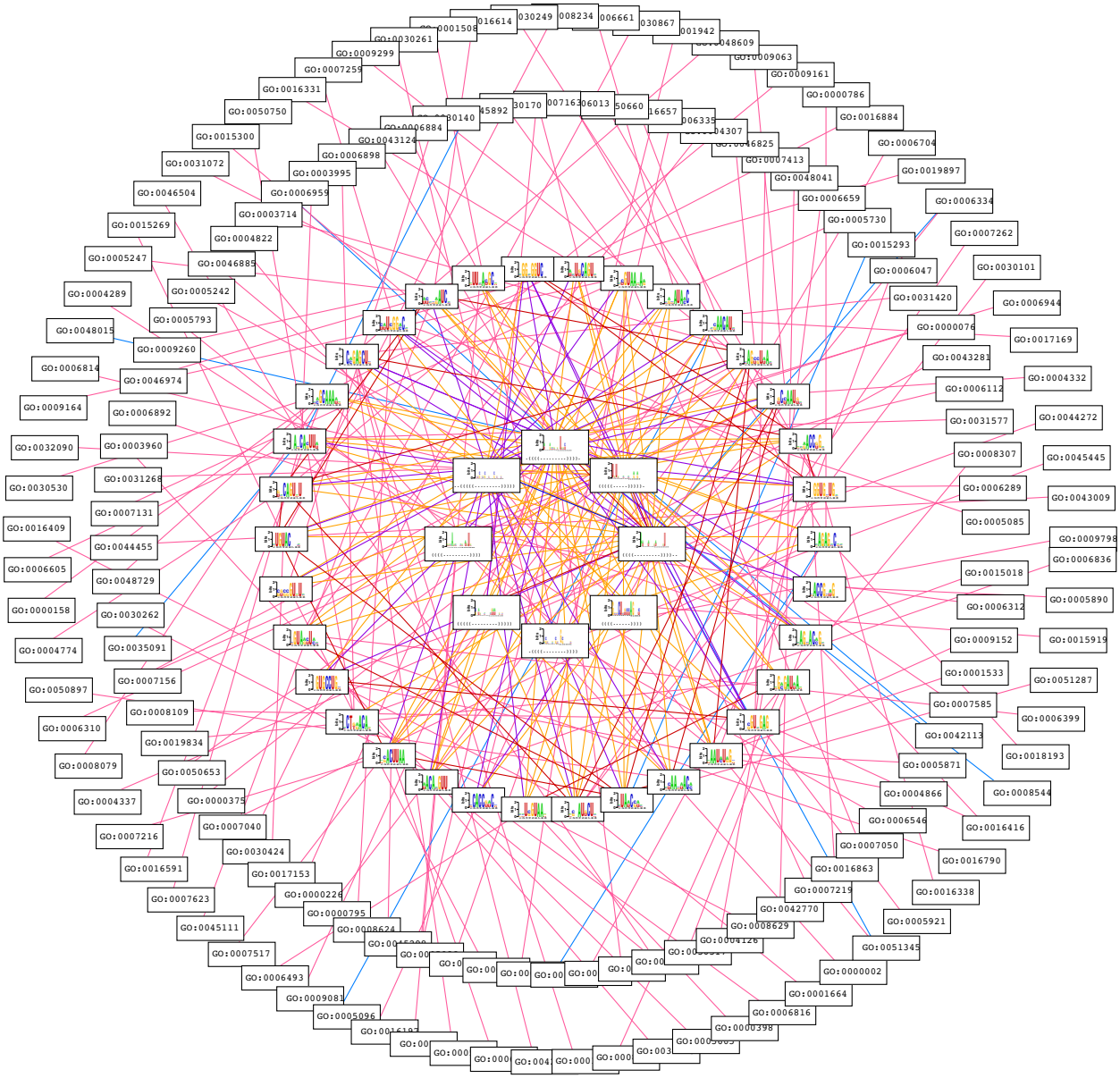
Supplementary Figure 12. Schematic of our analysis pipeline. In the first step, we identify the putative structural and linear motifs that are informative of a given whole-genome dataset. The resulting putative motifs then serve as building blocks for the regulatory modules formed based on their potential interactions and the pathways they most likely target.



Supplementary Figure 13. Motif-motif interaction maps. Mutual information enables the discovery of a variety of dependencies, including regulatory element interactions in the sense that the presence and absence of a given motif would be informative of the presence and absence of another[2]. The top figure shows such a map for the structural motifs where sRSM1 tends to show significant positive and negative interactions with other motifs. For example, sRSM3 is highly enriched in sRSM1-carrying transcripts, whereas sRSM8 significantly avoids these mRNAs. The bottom figure shows a similar map for structural motifs versus the linear motifs discovered by FIRE. In addition, we observed a complex map of pairwise interactions between these elements.



Supplementary Figure 14. An sRSM1-centered regulatory module. This figure schematically presents an exemplary regulatory module discovered through our framework. In the center, we show sRSM1 with a highly positive interaction with sRSM3 (bottom left, *z*-score=533) and a negative interaction with sRSM8 (top left, *z*-score=72). Similarly, two positive interactions (*z*-score 107 and 116) and one negative interaction (*z*-score=9.6) between sRSM1 and linear elements are shown. For each element in this regulatory module, we have also provided the potential target pathways along with their associated *p*-values.

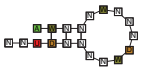


Supplementary Figure 15. Post-transcriptional regulatory network underlying transcript stability. Combining results from TEISER, FIRE and the regulatory interaction maps, we created a network of dependencies that comprises an inferred post-transcriptional regulatory network based on our observations. We also used iPAGE[9] to identify the pathways that are likely targeted by the identified elements. In this figure, the red-blue edges show positive and negative interactions between structural and linear motifs, respectively. The orange-purple edges show the interactions between structural motifs and linear ones, respectively. The pink-light blue edges connect each element with its target pathways. A pink edge means that the genes in a given pathway carry the specified motif more than estimated by chance, whereas, light blue edges show that the motif is under-represented in the genes of the corresponding pathway.

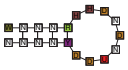
Supplementary Table 1. Mass-spectrometry of sRSM1 RNA binding samples. Proteins were identified from co-isolations with decoy or scrambled (Scr; to control for non-specific associations) and their respective number of unique peptides. Among these proteins, only HNRPA2B1 (highlighted) contains RNA binding domains and is specific to the decoy sample. Hence, we chose this protein as the likely candidate that directly interacts with sRSM1 structural motif.

Category	Protein Description	Accession #	MW, kDa	Decoy	Scr
RNA-binding protein	Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	37	2	0
Other	Uncharacterized protein C3orf63 (RAP140)	Q9UK61	189	2	0
	Cytoplasmic dynein 2 heavy chain 1	Q8NCM8	493	3	0
Commonly Observed Contaminants	Desmoplakin	P15924	332	6	0
	Junction plakoglobin	P14923	82	4	0
	Desmoglein-1	Q02413	114	3	0
	Plakophilin-1	Q13835	83	3	0
	Protein S100-A8	P05109	11	3	0
	Human serum albumin	P02768	69	2	0
	Protein-glutamine gamma-glutamyltransferase E	Q08188	77	2	0
	Keratinocyte proline-rich protein	Q5T749	64	10	1
	Putative annexin A2-like protein	A6NMY6	39	2	1
Enriched in Control (Scr)	Vimentin	P08670	54	4	8
	Actin, cytoplasmic 1	P60709	42	3	6
	Alpha-enolase	P06733	47	0	3
	Polyubiquitin-B	P0CG47	26	1	2
	60S acidic ribosomal protein P2	P05387	12	1	2
	10 kDa heat shock protein, mitochondrial	P61604	11	1	2
	Dynein heavy chain 17, axonemal	Q9UFH2	512	1	2
	Hemicentin-1	Q96RW7	613	0	2
	Antigen KI-67	P46013	359	0	2

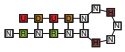
Supplementary Table 2. Position weight matrices associated with each of the sRSMs. The structural logos of the sRSMs are derived from regular expressions. In regular expressions, at each position, the possibilities are treated equally (*i.e.,* each nucleotide has a probability of 0, 0.25, 0.33, 0.5 or 1 at every position). However, upon identifying the likely targets for each structural motif (see Methods), a weight-matrix can be generated from these potentially functional instances as a post-processing step.



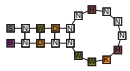
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	1.000	0.513	0.328	0.295	0.334	0.483	0.324	0.323	0.335	0.384	0.452	0.310	0.315	0.280	0.232	0.310	0.000	0.266	0.285
G	0.000	0.000	0.267	0.235	0.169	0.000	0.205	0.162	0.164	0.208	0.000	0.189	0.161	0.190	0.172	0.177	0.000	0.194	0.150
C	0.000	0.000	0.052	0.056	0.177	0.000	0.141	0.149	0.136	0.000	0.000	0.141	0.147	0.072	0.077	0.000	0.000	0.137	0.160
U	0.000	0.487	0.353	0.413	0.320	0.516	0.331	0.366	0.365	0.408	0.547	0.360	0.378	0.459	0.519	0.513	1.000	0.403	0.405



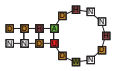
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	0.450	0.270	0.310	0.319	0.520	0.383	0.416	0.335	0.266	0.318	0.247	0.000	0.290	0.273	0.000	0.253	0.202	0.282	0.28
G	0.000	0.266	0.275	0.256	0.479	0.000	0.000	0.206	0.183	0.191	0.184	0.000	0.220	0.236	0.000	0.171	0.213	0.182	0.22
C	0.000	0.053	0.061	0.045	0.000	0.231	0.171	0.000	0.149	0.000	0.156	0.000	0.000	0.000	0.150	0.080	0.081	0.088	0.15
U	0.550	0.410	0.353	0.379	0.000	0.386	0.413	0.459	0.402	0.491	0.414	0.999	0.489	0.490	0.850	0.496	0.505	0.448	0.36



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	0.000	0.153	0.000	0.198	0.199	0.300	0.369	0.322	0.317	0.391	0.363	0.375	0.632	0.416	0.653	0.352
G	0.000	0.224	0.000	0.227	0.189	0.158	0.000	0.157	0.166	0.000	0.249	0.200	0.368	0.208	0.347	0.193
C	0.000	0.000	0.000	0.000	0.073	0.181	0.176	0.142	0.170	0.163	0.065	0.069	0.000	0.075	0.000	0.154
U	1.000	0.623	1.000	0.574	0.539	0.362	0.455	0.379	0.346	0.446	0.323	0.357	0.000	0.301	0.000	0.301



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	0.001	0.274	0.453	0.253	0.244	0.263	0.384	0.302	0.267	0.285	0.360	0.001	0.391	0.464	0.255	0.299	0.345	0.235	0.001
G	0.737	0.205	0.001	0.283	0.226	0.190	0.001	0.171	0.186	0.141	0.001	0.325	0.001	0.001	0.275	0.165	0.201	0.286	0.261
C	0.261	0.134	0.001	0.001	0.088	0.194	0.210	0.149	0.171	0.201	0.135	0.001	0.001	0.001	0.059	0.112	0.001	0.065	0.237
U	0.001	0.388	0.545	0.463	0.441	0.353	0.405	0.378	0.376	0.373	0.504	0.673	0.607	0.533	0.411	0.424	0.453	0.413	0.501



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	0.279	0.315	0.445	0.998	0.397	0.429	0.316	0.304	0.361	0.366	0.305	0.450	0.335	0.001	0.260	0.250	0.336
G	0.224	0.290	0.001	0.001	0.219	0.001	0.161	0.164	0.001	0.185	0.182	0.001	0.223	0.001	0.295	0.145	0.160
C	0.001	0.001	0.130	0.001	0.001	0.167	0.136	0.170	0.193	0.001	0.160	0.001	0.001	0.001	0.001	0.108	0.072
U	0.496	0.394	0.425	0.001	0.383	0.403	0.387	0.362	0.445	0.448	0.353	0.548	0.441	0.998	0.445	0.497	0.432

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	0.190	0.001	0.001	0.087	0.076	0.151	0.155	0.001	0.001	0.167	0.001	0.001	0.171	0.048	0.072	0.080	0.001
G	0.461	0.559	0.456	0.425	0.456	0.476	0.358	0.548	0.422	0.354	0.998	0.547	0.321	0.420	0.416	0.464	0.440
C	0.348	0.440	0.308	0.286	0.285	0.371	0.306	0.450	0.402	0.279	0.001	0.452	0.349	0.305	0.277	0.312	0.373
U	0.001	0.001	0.235	0.202	0.182	0.001	0.181	0.001	0.176	0.200	0.001	0.001	0.159	0.227	0.235	0.145	0.186

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.933	0.022	0.489	0.133	0.022
G	0.933	0.156	0.200	0.022	0.022	0.244	0.156	0.022	0.022	0.022	0.933	0.289	0.689	0.022
C	0.022	0.622	0.289	0.933	0.022	0.111	0.022	0.156	0.844	0.022	0.022	0.111	0.089	0.756
U	0.022	0.200	0.489	0.022	0.933	0.622	0.800	0.800	0.111	0.022	0.022	0.111	0.089	0.200

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.165	0.001	0.001	0.001	0.148	0.001	0.203	0.001	0.207	0.183	0.001	0.180	0.164	0.215	0.001	0.206	0.001	0.142	0.133	0.001	0.170
G	0.310	0.344	0.435	0.599	0.339	0.511	0.556	0.478	0.232	0.333	0.348	0.273	0.345	0.297	0.504	0.204	0.241	0.346	0.380	0.400	0.395
C	0.278	0.332	0.176	0.400	0.218	0.187	0.241	0.521	0.324	0.266	0.383	0.308	0.293	0.238	0.494	0.338	0.315	0.288	0.200	0.312	0.246
U	0.247	0.323	0.388	0.001	0.295	0.301	0.001	0.001	0.237	0.219	0.268	0.239	0.198	0.250	0.001	0.252	0.444	0.224	0.286	0.288	0.189

References

1 Dolken, L. *et al.* High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *Rna* **14**, 1959-1972, doi:Doi 10.1261/Rna.1136108 (2008).

2 Elemento, O., Slonim, N. & Tavazoie, S. A universal framework for regulatory element discovery across all Genomes and data types. *Mol Cell* **28**, 337-350, doi:Doi 10.1016/J.Molcel.2007.09.027 (2007).

3 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).

4 Chan, C. S., Elemento, O. & Tavazoie, S. Revealing posttranscriptional regulatory elements through network-level conservation. *PLoS Comput Biol* **1**, e69, doi:10.1371/journal.pcbi.0010069 (2005).

5 Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

6 Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* **24**, 227-235, doi:10.1038/73432 (2000).

7 Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *P Natl Acad Sci USA* **101**, 6062-6067, doi:10.1073/pnas.0400782101 (2004).

8 Windbichler, N. & Schroeder, R. Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat Protoc* **1**, 638-U634, doi:Doi 10.1038/Nprot.2006.95 (2006).

9 Goodarzi, H., Elemento, O. & Tavazoie, S. Revealing global regulatory perturbations across human cancers. *Mol Cell* **36**, 900-911, doi:10.1016/j.molcel.2009.11.016 (2009).