Supplementary Materials

Materials and methods

Keratinocyte Cell Culture, Viral Infection, and Growth assays

Keratinocytes were isolated from neonatal epidermis as described (Riemondy *et al*, 2015). Keratinocytes were initially plated on a NIH-3T3 subclone J2 feeder layer for three passages, after which continuously proliferating keratinocytes were propagated in the absence of feeder layer support. Keratinocytes were grown in the presence of E-Low Media with 0.05 mM Ca^{++} at 37°C in 5% CO_2. Lentiviral shRNA knockdowns were performed using PLKO shRNA constructs (See Table S2 for shRNA information). Lentivirus was generated by transiently transfecting HEK-293FT cells with pVSVG, psPAX.2, and relevant PLKO constructs at a 1 : 2.5 : 4 mass ratio using MiRus Bio LT1 (Mirus Bio LLC). Retroviral Msi2 overexpression was performed using MIGR-Vector and MIGR-Msi2 constructs (Jackson *et al*, 2013). Retrovirus was generated by transiently transfecting HEK-293FT cells with pCL-Eco, pAdvantage, and relevant MIGR constructs at a 1 : 1.7 : 8.7 mass ratio using MiRus Bio LT1. Viral supernatant was harvested 24-72 hours post transfection and filtered with a 0.45 μM filter. Puromycin was added (2 μg/ml) to select for shRNA producing keratinocytes. Cellular Growth Curves were calculated by counting the number of cells post-plating on a hemocytometer. Colony formation assays were performed by plating 1,000 cells per well of a 6 well plate in triplicate and culturing for 5-10 days. Colony density was calculated using Image J (images set to 8-bit to convert to grayscale, threshold set and applied across all colony images to convert to black (colony) and white (background), wells were selected with circle tool and particles

were analyzed to generate area covered by colonies. Colony area was normalized to PLKO-shScr/MIGR-Vec control).

Flow Cytometry

Click-IT EdU cell proliferation assays were performed following the manufacturer's instructions (Thermo Fisher). Keratinocytes were pulsed with 10 μM EdU for 1 hour prior to harvesting for EdU detection. DNA content was detected by Hoechst Dye staining. For detecting apoptotic cells, keratinocytes were harvested via trypsinization, then washed in complete E-Low media, followed by a wash in ice cold PBS, and 1x Annexin-V-binding buffer (10 mM HEPES, 140 mM NaCl, 2.5 mM CaCl2, pH 7.4). Cells were then re-suspended in 100 μl of 1x Annexin-V-binding buffer. 5 μl of Alexa-Fluor 488 Annexin-V and 1 μl of 100 μg/ml Propidium Iodide solution was next added and the cells were incubated for 15 minutes at room temperature. 400 μl of 1x Annexin-V-binding buffer was added and the cells were then analyzed on a BD Cyan flow cytometer. Flow cytometry profiles were analyzed with FlowJo software (www.flowjo.com).

Western Blotting and qPCR

Western blotting was performed with 20-30 μg of protein lysate run on a 10% SDS-PAGE gel. Proteins were transferred to 0.2 μm PVDF and blotted for *Musashi 2* and *β-tubulin* (See Table S2 for antibody information). Blots were incubated overnight with primary antibodies and detected using anti-rabbit HRP-conjugated secondary antibodies and Amersham ECL-Plus reagents (GE Healthcare Life

Sciences). X-ray films were used to detect the signal, scanned and processed with Image J software. qPCR was performed on Trizol extracted RNA using SuperScript III RT kit (Life Technologies) and a BioRad CFX-384 machine. Relative expression were computed using ΔΔCq method normalized to *Hprt* and *Gapdh* values with error bars denoting standard error of the mean. See Table S2 for qPCR primer information.

RIP-qPCR

RNA Immunoprecipitation followed by qPCR was performed as follows. One 15cm plate of cultured keratinocytes was harvested via scraping and stored at -80˚C until lysis with non-denaturing lysis buffer (2mM EDTA, 20mM Tris-HCl, 137mM NaCl, 1% NP-40, 1x Pierce Protease Inhibitor Cocktail, 40U/μL RNase OUT). The lysate was split and added to 25μL Protein G Dynabeads (Thermo Fisher) pre-bound with 5μg Anti-Msi2 antibody or Rabbit IgG (Table S2 for antibody information). Antibody-bead, lysate mix was allowed to incubate for 2 hours at 4˚C. The supernatant was removed and the beads were washed three times with NT2 buffer (50mM Tris-HCl, 150mM NaCl, 1mM $MgCl_2$, 0.05% NP-40) before the RNA was extracted using Trizol (Thermo Fisher). An aliquot of Total lysate and supernatant was taken for RNA extraction via Trizol and western blotting for determining the efficiency of the IP. RNA samples were subsequently used for qPCR (Table S2 for qPCR primer sequences). Relative enrichment was calculated as enrichment over IgG control using *Gapdh* and *Hprt* as non-target controls.

Quantifying Cellular Motility

Keratinocytes were transiently transfected with the pREX-H2B-mCherry plasmids and plated at a density of 400 cells/mm^2 on fibronectin coated glass bottom 96-well plates (Matrical Bioscience, MGB096-1-2-LG-L). After 24 hours, time lapse imaging was performed using an ImageXpress MicroXL (Molecular Devices), where an image of the mCherry channel (Excitation=562/40 nm, Emission= 641/75 nm) was acquired every 15 minutes for 4 hours for each condition in parallel. Time-lapse videos were analyzed as previously described using the Pathfinder program (Chapnick *et al*, 2013). Displaying of tracks was achieved using Pathfinder positional outputs for each cell, where each track for each cell was centered at coordinate (0, 0) using MATLAB in order to display normalized tracks. Data displayed represents one of two independent experiments done in triplicate, where each trial measures at least 200 cells. Error displayed depicts the standard error of the mean for three trials. The *p* value was calculated using t-test assuming equal variances in Excel.

Immunofluorescence and Immunocytochemistry

OCT skin tissue sections (10 μM) were fixed for 10 minutes with 4% Paraformaldehyde at room temperature, washed three times for 5 minutes with PBS, and blocked for 10 minutes using Gelatin Block (0.1% Triton X-100, 2% gelatin, 2.5% normal goat serum, 2.5% normal donkey serum, and 1% BSA in PBS). Sections were incubated with the following primary antibodies in blocking buffer overnight at 4°C: *Keratin 5* (*Krt5*), *E-cadherin* (*Cdh1*), and *Musashi-2* (*Msi2*) (See Table S2 for antibody information). Following three 10 minute washes in PBS, sections were incubated with appropriate Alexa-Fluor secondary antibodies (1:2000)

for 1 hour and washed two times for 5 minutes with PBS. Sections were incubated with Hoechst Dye for 10 minutes and mounted in Vectashield Anti-fade solution.

For focal adhesion staining, keratinocytes were sparsely plated on 1 μg/ml Fibronectin coated #1.5 thickness coverslips. Cells were fixed in 4% Paraformaldehyde with 0.5% Triton-X-100 in PBS at 37˚C for 10 minutes, followed by three PBS washes and storage at 4˚C for at least overnight. Cells were next blocked for 10 minutes in the Gelatin Blocking Buffer at room temperature, then incubated with mouse anti-Vinculin antibody in Gelatin Blocking buffer for 2 hours at room temperature (See Table S2 for antibody information). Coverslips were next washed 3 times with PBS for 10 minutes, followed by co-incubation with Alexaflour-488 anti-mouse IgG1a secondary antibody at 1:2000 and Phalloidin-Alexafluor-647 at 1:50 in Gelatin Blocking Buffer for 1 hour. Coverslips were washed twice with PBS for 10 minutes followed by incubation with Hoechst Dye for 10 minutes and mounted with VectaShield anti-fade.

Microscopy

Microscopy images for skin sections were taken using a Leica DM5500B microscope with a Hamamatsu C10600-10B camera and processed with the Leica image analysis suite, MetaMorph (MDS Analytical Technologies). The ImageJ/FIJI software package was used to stich together > 40 images using the Pairwise Stitching function to produce the skin wound images presented.

Microscopy for detecting focal adhesions was performed on a Nikon A1 laser scanning confocal using a 100x/1.49 NA objective lens at the CU Boulder Light

Microscopy Core Facility. For each image acquired 9 optical slices of 0.125 microns were taken through the focal plane containing focal adhesions. Each Z-stack was converted into a 2D representation by taking the maximum intensity at each pixel. Presented micrographs were further processed to adjust brightness and contrast values. Identical microscope settings and image analysis parameters were used for all images presented. For quantifying focal adhesions, an ImageJ/FIJI macro was constructed that counted focal adhesions in the following manner. To exclude partially imaged cells on the image boundary the following steps were performed. Maximum intensity projections of Phallodin staining were auto contrasted ("Enhance Contrast", "saturated=0.35"), then thresholded using the mean method (setAutoThreshold("Mean dark")). Thresholded images were converted to masks, and the binary operations Dilate, Close-, and Fill Holes were used to identify cellular regions. The Analyze Particles function was then applied to exclude regions that fell below a size threshold (size=1200-Infinity circularity=0.00-0.99). These masked regions were next overlaid on the Hoechst and Vinculin images to exclude non-cellular regions (setPasteMode("AND")). Next, the number of cells within each image was calculated as follows. The DAPI maximum intensity projection was processed with CLAHE to enhance local contrast ("Enhance Local Contrast (CLAHE)", "blocksize=75 histogram=256 maximum=6 mask=*None* fast_less(accurate)") then auto contrasted ("Enhance Contrast", "saturated=0.35"), and thresholded using the default method (setAutoThreshold("Default dark")). Nuclei counts were obtained by performing the Analyze Particles function ("size=25-Infinity circularity=0.00-1.00 show=Masks display summarize"). The nuclei counts were

manually inspected after analysis and corrected when errors occurred. Lastly, the number of focal adhesions was calculated using methods from on a previously published protocol for counting focal adhesions (Horzum *et al*, 2014). Briefly, the maximum intensity projections of Vinculin staining underwent the following processing steps:

1) Background subtraction ("Subtract Background...", "rolling=50 sliding")
2) CLAHE ("Enhance Local Contrast (CLAHE)", "blocksize=19 histogram=256 maximum=6 mask=*None* fast_less(accurate)")
3) Exponentiation ("Exp")
4) Autocontrasting ("Enhance Contrast", "saturated=0.35")
5) Log3D filtering ("LoG 3D", "sigmax=5 sigmay=5")
6) Thresholding (setAutoThreshold("Default dark")
7) Focal Adhesion Counting ("Analyze Particles...", "size=2-2000 circularity=0.00-0.99 exclude")

The number of focal adhesions per image was then divided by the number of nuclei to obtain a focal adhesion count per cell per image count.

Ribosome Profiling and RNA-seq Analysis

Ribosome profiling raw reads were first trimmed to remove 3' adaptor sequence (CutAdapt, default settings) and filtered to only retain sequences at least 20 nucleotides in length. Filtered reads were then aligned to remove contaminating ncRNAs by sequential Bowtie alignment (default settings) to an rRNA database (Illumina iGenomes, mm10), tRNA database (UCSC table browser, mm10), and

ncRNA database (Ensemble). Unaligned reads were then aligned to the mouse genome with Tophat with a supplied .gtf transcript annotation file (Illumina iGenomes, mm10) (settings = --bowtie1). Alignments uniquely overlapping coding sequences were counted with HTSeq Count (settings = -s yes –t CDS). Differential expression was next calculated with DESeq using default settings. Alignment statistics are shown in Figure S10.

RNA-seq raw reads were first trimmed to remove 3' adaptor sequence (GATCGGAAGAGCACACGTCTGAACTCCAGTC, CutAdapt, default settings). Reads were then aligned to the mouse genome with Tophat with a supplied .gtf transcript annotation file (Illumina iGenomes, mm10) (settings = --bowtie1 --library-type fr-firststrand). Alignments uniquely overlapping coding sequences were counted with HTSeq Count (settings = -s reverse –t CDS). Differential expression was calculated with DESeq using default settings. Alignment statistics are shown in Figure S10. Transcripts with at least a BaseMean of 10 in the shScr libraries in both Ribosome profiling and RNA-seq were used for downstream analysis. Translation efficiency was calculated as the ratio of the average reads per million for Ribosome profiling and RNA-seq data. Cumulative fractions were computed using the ecdf() function in R. *Msi2* targets were defined as the subset of genes that were upregulated in both the RNA-seq and Ribosome Profiling datasets (FDR <0.05) and contain a high-confidence 3'UTR *Msi2* HITS-CLIP peak. Target genes were ranked by ranking genes based up degree of upregulation in the Ribosome Profiling data and the RNA-seq. The ranks for each dataset were summed, and targets genes were then ranked by

the gene with the high summed rank with ties being decided by the gene with the largest number of total 3'UTR HITS-CLIP Reads.

Msi2-HITS-CLIP peak identification

Raw reads were first filtered to keep only high quality sequences (CIMS package, fastq_filter.pl, settings = -f mean:0-24:20), then trimmed to remove 5' and 3' adapter sequences (Cutadapt, default settings) (Table S2 for oligo sequences). Sequence duplicates were next discarded to remove reads arising from PCR duplicates, followed removal of the 2 NN's introduced by the adaptors from the 5' and 3'. Reads were next aligned to the mouse genome (mm10) using Novoalign (autothreshold alignment with -s 1 setting), requiring a minimum 20 nucleotide alignment length. Uniquely mapped alignments were next filtered to remove duplicate alignments with identical start and stop coordinates which likely represented PCR duplicates with sequencing errors or PCR introduced mutations. Continuous alignments were next clustered to identify Msi2 binding sites (Bedtools merge, setting = -s). Binding sites were then split to resolve multiple summits within long peaks using PeakSplitter (-v 0.6). Peaks were then filtered to keep only peaks with at least 10 total reads, with reads present in at least 4 of 5 libraries (Biological Complexity >=4). Alignment statistics for each library are shown in (Figure S10). Peaks were annotated to the genome using combined Gencode (VM4) and UCSC (mm10) annotations downloaded from UCSC table browser with regions selected based on at least a single nucleotide overlap with first 3'UTRs, followed by CDS, 5'UTRs, ncRNAs, introns from protein coding genes, and lastly intergenic (bedtools

intersect, setting = –s –u). ncRNA regions were defined by selecting Gencode transcripts with a non-coding transcript class ID, and by selecting UCSC knownGene transcripts with identical cdsStart and cdsEnd coordinates. 3'UTR regions were extended 5 Kbp or up to the next downstream genomic feature if a feature occurred closer than 5 Kbp downstream. Peak summits were identified by calculating reads coverage across the peak, then identifying the region with the maximum reads coverage. If the read coverage was equivalent across a region larger than a single nucleotide, then the middle of the region was selected as the summit (Bedtools coverage, setting = -s –d). For all analysis steps that do not reference a particular software package, in-house R, Python, or Bash scripts were used to perform the analysis.

Metagene Profiles

Msi2 binding profiles along a scaled average mRNA were computed using a Python script. Briefly, 5'UTR, CDS, and 3'UTR regions were extracted from protein-coding Gencode transcripts that contained 5'UTR, CDS, and 3'UTR attributes and were expressed in keratinocytes (BaseMean >= 10 in shScr libraries). The longest transcript for each region was taken as representative for each gene. Reads coverage was then calculated across each exonic region excluding intronic sequences (pyBedtools bed12tobed6 and coverage, setting = -s –d). Reads coverage vectors for each region (5'UTR, CDS, and 3'UTR) per transcript were then constructed and normalized by library size to reads per million mapped, averaged across 100 bins (Scipy binned_statistic) and averaged over all transcripts. Peak coverage vectors

were computed analogously and averaged over all transcripts without normalization for library size.

Motif Searching

Msi2 HITS-CLIP 3'UTR binding sites were used to identify *de novo* motifs. A region extending 20 nucleotides 5' and 3' of the peak summit was selected for motif searching. *De novo* motif searching for 3-9mers was performed using HOMER with sequence auto-normalization methods disabled with a supplied background 3'UTR database (findMotifsGenome.pl–size given –norevopp –rna –noweight –nlen 0 –bg –chopify). *De novo* motif searching within 5'UTR and CDS regions was performed in the same fashion. For intronic regions the background database was subset to included only 20,000 randomly selected regions to reduce computational burden for the motif searching. High resolution motif logos were generated by converting the HOMER motif output with Weblogo. The top motif generated for each length (3-9mer) were defined as Msi2 binding motifs (UAG, UAGU, UAGUA, UUAGUA, UAGUAGU, UAGUAGUA, GUAGUAGUA). The consensus sequence for each motif was then used to plot the positional distribution of motifs surrounding each peak summit, and to classify binding sites as motif containing or not motif containing. Positional motif densities were computed by enumerating consensus motif match start positions across an interval 110 nucleotides 5' and 3' of the peak summit and plotted with a bin size of 5 nucleotides. Randomized 3'UTR backgrounds were constructed by shuffling the genomic intervals throughout 3'UTR coordinates (bedtools shuffle, run individually for each strand). For assessing the

multiplicity of UAG motifs within *Msi2* binding sites, the number of UAGs was enumerated in 50 nucleotide windows +/- 225 nucleotide from the the peak summit position. 225 nucleotides of sequence 5' and 3' of the 450 nucleotide window were analyzed as flanking regions as a background control in addition to randomly selected 3'UTR regions which were defined as described above.

Analysis of publically available datasets:

Msi2-HITS-CLIP analyzed datasets were downloaded from GSE64388 and GSE62115 GEO datasets. For GSE64388 datasets, all 3'UTR HITS-CLIP peaks identified in all wild-type and Msi2 transgenic HITS-CLIP datasets were used for the analysis. For GSE62115 datasets, human gene symbols were converted to mouse homologs using a database downloaded from Ensemble BioMart and only genes with 3'UTR localized Msi2 binding sites were used for the analysis. GO analysis was performed using DAVID with Gene symbols as input.

IPA analysis

Ingenuity Pathway Analysis was performed on datasets filtered for upregulated genes in RNA-seq, Ribo-seq, HITS-CLIP, and combined datasets with RNA/Ribo-seq and RNA/Ribo-seq HITS-CLIP targets as illustrated in Figure 4. Only experimentally observed relationships were considered. IPA build version 346717M was used for the analysis. Analysis references were set to Ingenuity Knowledge Base (Genes Only) with all cell types and mammals selected. Direct and

indirect relationships were included. Data shown was taken directly from Summary of Analysis output in IPA.

Animals

Mice were bred and housed according to the guidelines of IACUC at a pathogen-free facility at the University of Colorado (Boulder, CO, USA). Tissues were harvested for immunofluorescence as previously described (Yi *et al*, 2006). All procedures involving mice were approved under IACUC protocol 1408.01.

Statistical Analyses

All statistics were computed using either R or Excel. Statistical tests performed for each experiment is listed in the figure legends. For two sample, student t-tests, normality was assumed and two-way equal variance tests were performed. For qPCR experiments, the mean of all technical replicates was computed and used to represent one independent value for statistical testing. For analyses of empirical cumulative distributions, two-way Kolmogorov-Smirnov tests were used unless indicated in the figure legend. Sample sizes were chosen based prior knowledge of variability in each assay. The investigators were not blinded to sample identity during sample collection, processing, and analysis.

Supplementary Figure legends

Figure S1. *Msi2* but not *Msi1* is enriched in the stem and progenitor cells of mouse skin. (A-E) *Keratin 5* and *Msi2* immunostaining of postnatal day 6 (P6) skin

in the bulge (B) bulge stem cell compartment (C) outer root sheath progenitors (D) interfollicular epidermis (E) transit-amplifying cells in the Matrix. (F) RNA-seq, Ribo-seq and Mass Spectrometry measurements of *Msi1* and *Msi2* in keratinocytes and total epidermal samples. For the microscopy panels, the asterisk denotes the auto-fluorescent hair shaft and stratum corneum. Epi = epidermis, Der = dermis, HF = hair follicle, ORS = outer root sheath, Mx = matrix. Scale bars represent 50 μM for Panels A-E.

Figure S2. *Msi2*-HITS-CLIP library preparation. (A) Schematic of the HITS-CLIP protocol used in this study. (B) Autoradiogram of the ^{32}P labeled *Msi2*-RNA complexes isolated and used for library preparation. The red boxes indicate regions of the nitrocellulose membrane excised for library preparation. The white numbers indicate the replicate number of the library. Note that this is the full autoradiogram for the experiment shown in Figure 1A. (C) Reads distribution for each replicate or all the libraries in aggregate. The reads displayed were first processed prior to plotting to remove low quality sequence, trimmed to remove 5' and 3' adaptors, collapsed to remove PCR duplicates, and further trimmed to remove 5' and 3' NN randomized dinucleotides introduced by the adaptors. Red bars indicate median read lengths. Blue bars indicate mean read lengths.

Figure S3. Motifs identified in *Msi2* HITS-CLIP peaks in non-3'UTR regions (A) *De novo* motif searching was performed for 3-9mer motifs on a genomic region +/-20 nucleotides surrounding the peak summit of those peaks localized to 5'UTR

regions. The top motif identified for each motif length is displayed (B) Same as Panel A except searching CDS regions. (C) Same as Panel C except searching intronic regions.

Figure S4. Positional distribution of *Msi2* motifs identified by HITS-CLIP and positional distribution of *Msi2* motifs identified by Msi1 SELEX. (A) Positional enrichment for the top 3-9mer motifs identified by HITS-CLIP was calculated in a window surrounding the HITS-CLIP peaks summit. A background distribution was also calculated by shuffling the position of the searched window randomly within 3'UTR sequences. (B) Positional enrichment for the Msi1 motif predicted by SELEXC, G/AUn_{1-3}AG. Otherwise as described in A.

Figure S5. Gene tracks of select Msi2 HITS-CLIP targets. Gene tracts of a set of Msi2 targets from HITS-CLIP data. Targets were selected as a representative set of genes and includes genes validated in this study as *Msi2* targets. Peak summit inset displayed below tracts. UAG core motifs and the reverse complement motif CUA are marked in red across the 3'UTR of targets and in the inset of the peak summits. Blue marks positive strand reads. Green marks negative strand reads. Scale bar = 500 nucleotides. Y-axis is peak height

Figure S6. Ribosome profiling read phasing. Metagene of Ribo-seq reads density across the start codon or termination codon for coding regions. Reads overlap was only assessed for the 5' nucleotide of the read. 29 nucleotide long reads from all

Ribo-seq libraries were used to assess phasing. Bar chart displays fraction of reads with 5' nucleotide aligning to frame 1, frame 2, or frame 3 of the gene coding sequence displayed.

Figure S7. *Msi2* targets are regulated by changes in mRNA abundance (A) Cumulative distributions of Log_2 Fold Changes in Translation Efficiency for genes that contain *Msi2* 3'UTR HITS-CLIP peaks, with 1-2 UAG motifs (Red), 3-4 UAG motifs (Orange), 5 or more UAG motifs (Black), no UAG motif (Green), or do not contain *Msi2* 3'UTR HITS-CLIP peaks. P-values were calculated by one-way KS tests assessing if translational efficiency is increased by the presence of UAG motifs. If a gene contained multiple peaks with differing numbers of UAGs, the gene was assigned to the highest UAG category that is found in the peaks. (B) *Msi2* HITS-CLIP data for an established *Msi2* target, *Jag1*. (C-D) Log_2 Ribo and RNA-seq Fold Changes for *Jag1* (C) and novel *Msi2* targets detected in this study and a negative control gene *Src* (D). Asterisk indicates FDR <0.05.

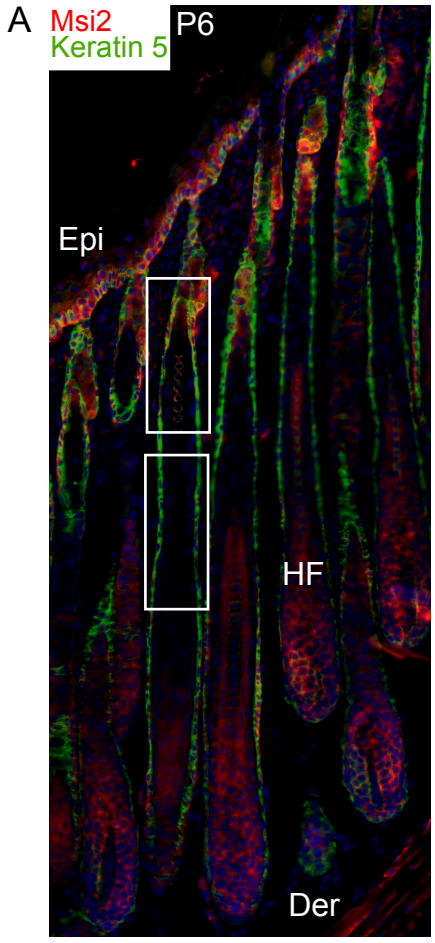
Figure S8. Regulatory impact of *Msi2* binding to non-3'UTR regions. (A) Cumulative distributions of RNA-seq and Ribo-seq Log_2 Fold changes for those genes bound by *Msi2* in the 5'UTR (shown in red) or genes not bound by *Msi2* in the 5'UTR (blue) (B) Same as Panel A, except genes with *Msi2* binding in the CDS were plotted. (C) Same as Panel A, except genes with *Msi2* binding in intronic regions were plotted. For all panels, if the gene was bound by *Msi2* in multiple regions

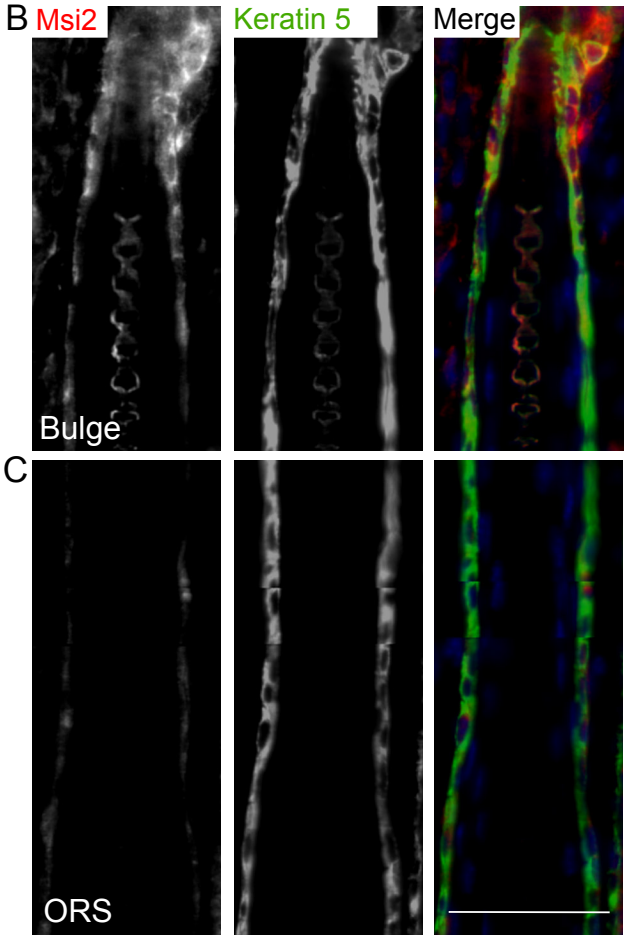
(5'UTR, CDS, 3'UTR, or intronic) these genes were classified as not bound to avoid conflating *Msi2* regulatory effects due to binding outside of the regions examined.

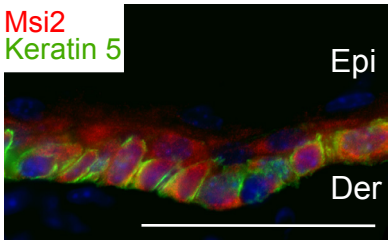
Figure S9. Identification of cell-type specific *Msi2* targets. (A) Venn-diagrams comparing *Msi2* targets identified in *Msi2*-HITS CLIP from keratinocytes (this study), intestinal stem cells (GSE64388), and a leukemia cell line (GSE62115). P-values were derived from the hypergeometric test (B) GO analysis of genes only found in keratinocytes. (C) GO analysis of genes shared between the intestine and the leukemia studies. (D) GO analysis of the common genes found in all three studies. The top ten enriched biological process pathways from each comparison are displayed with the annotated –log10 p-value.

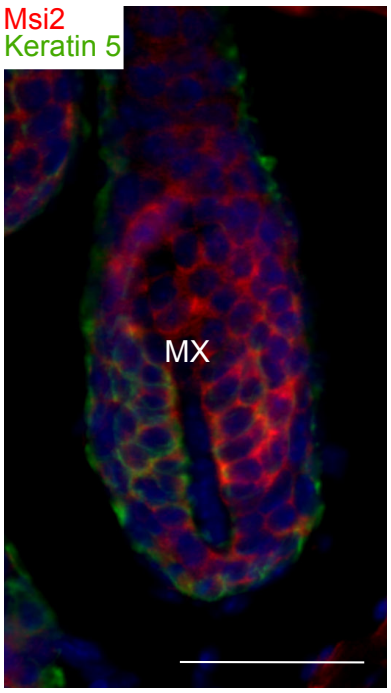
Figure S10. Mapping statistics for all sequencing libraries.

S1





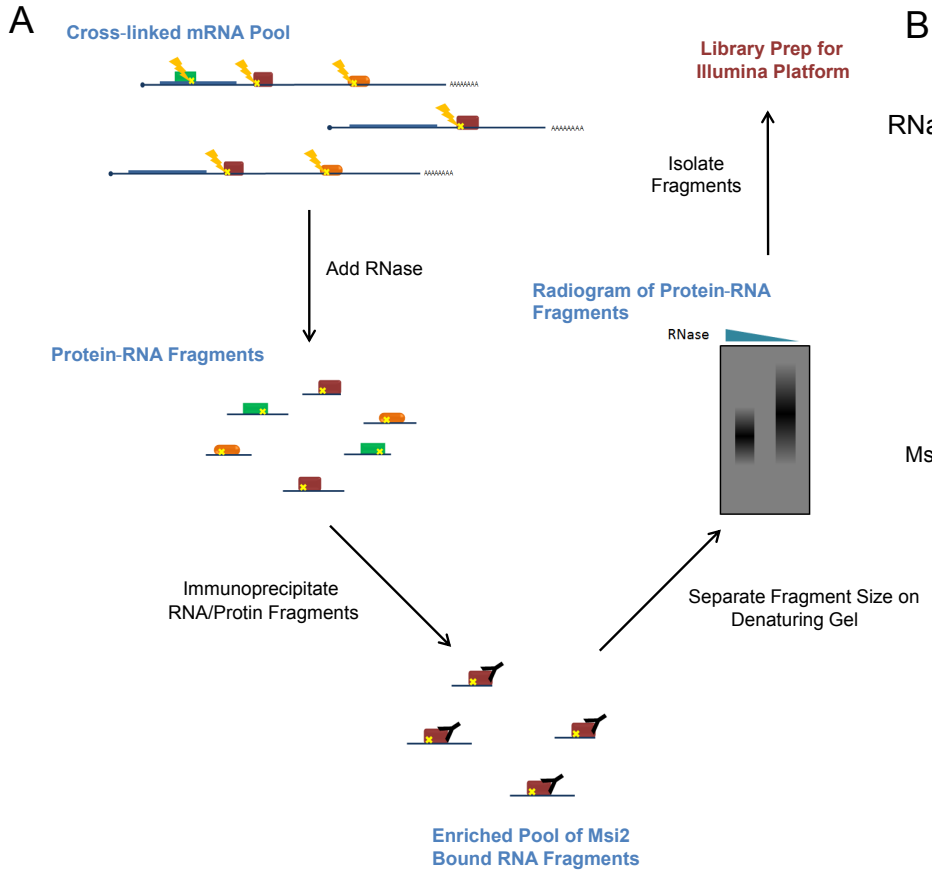


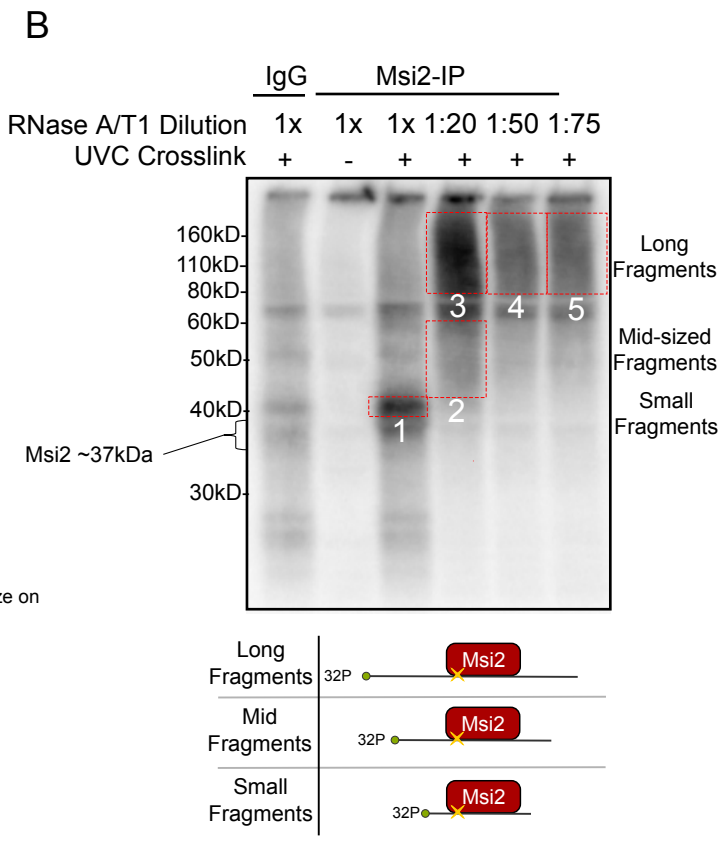


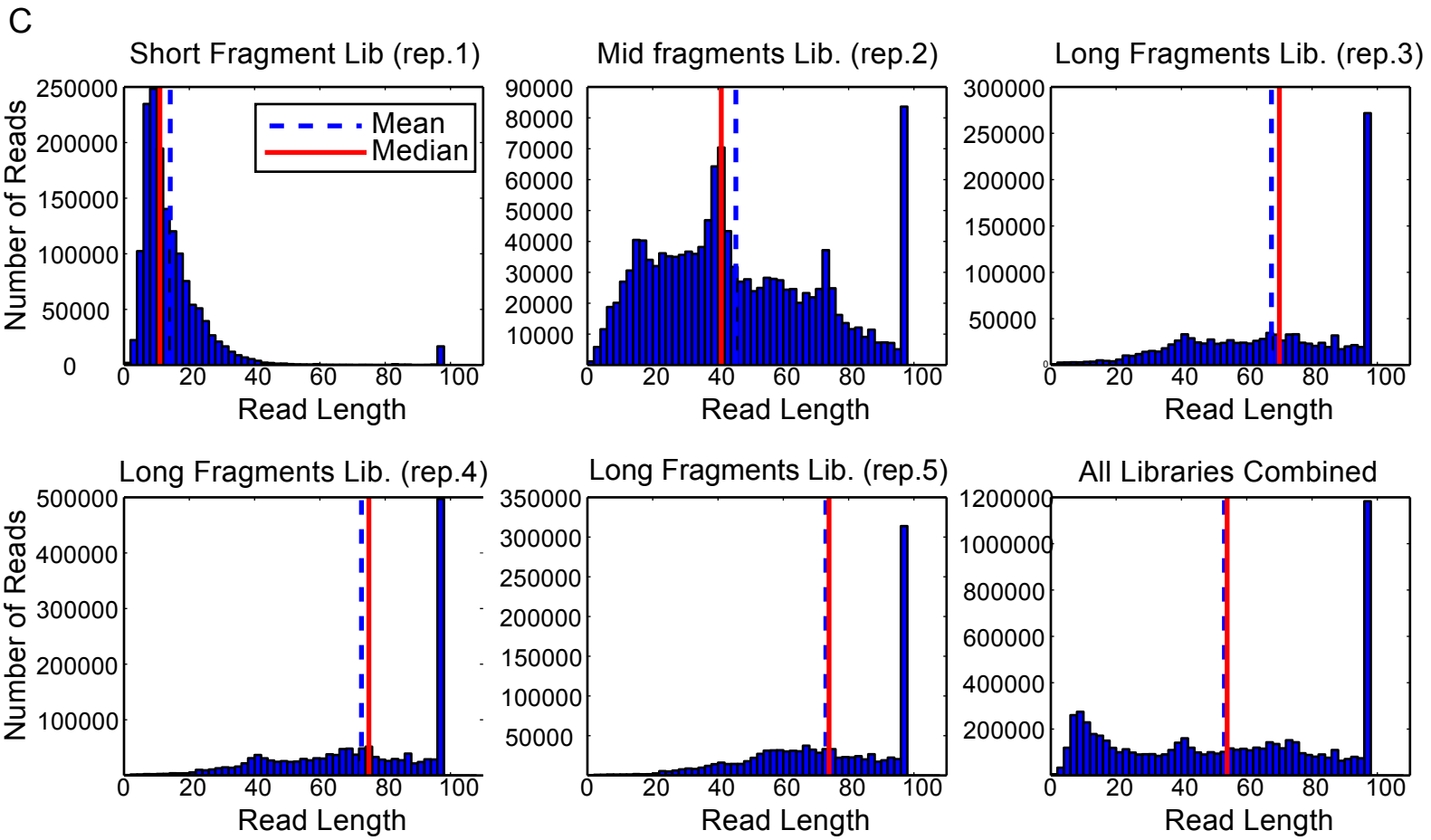
F

	RNA-Seq Total Epidermis FPKM	RNA-Seq Keratinocytes FPKM	Ribo-Seq Keratinocytes RPKM	Mass Spec Total Epidermis Spectral Counts
Msi1	1.65	0.35	0.96	N/D
Msi2	15.08	4.53	21.38	4
$\frac{\text{Msi2}}{\text{Msi1}}$ ratio	9.1	12.92	22.2	N/A

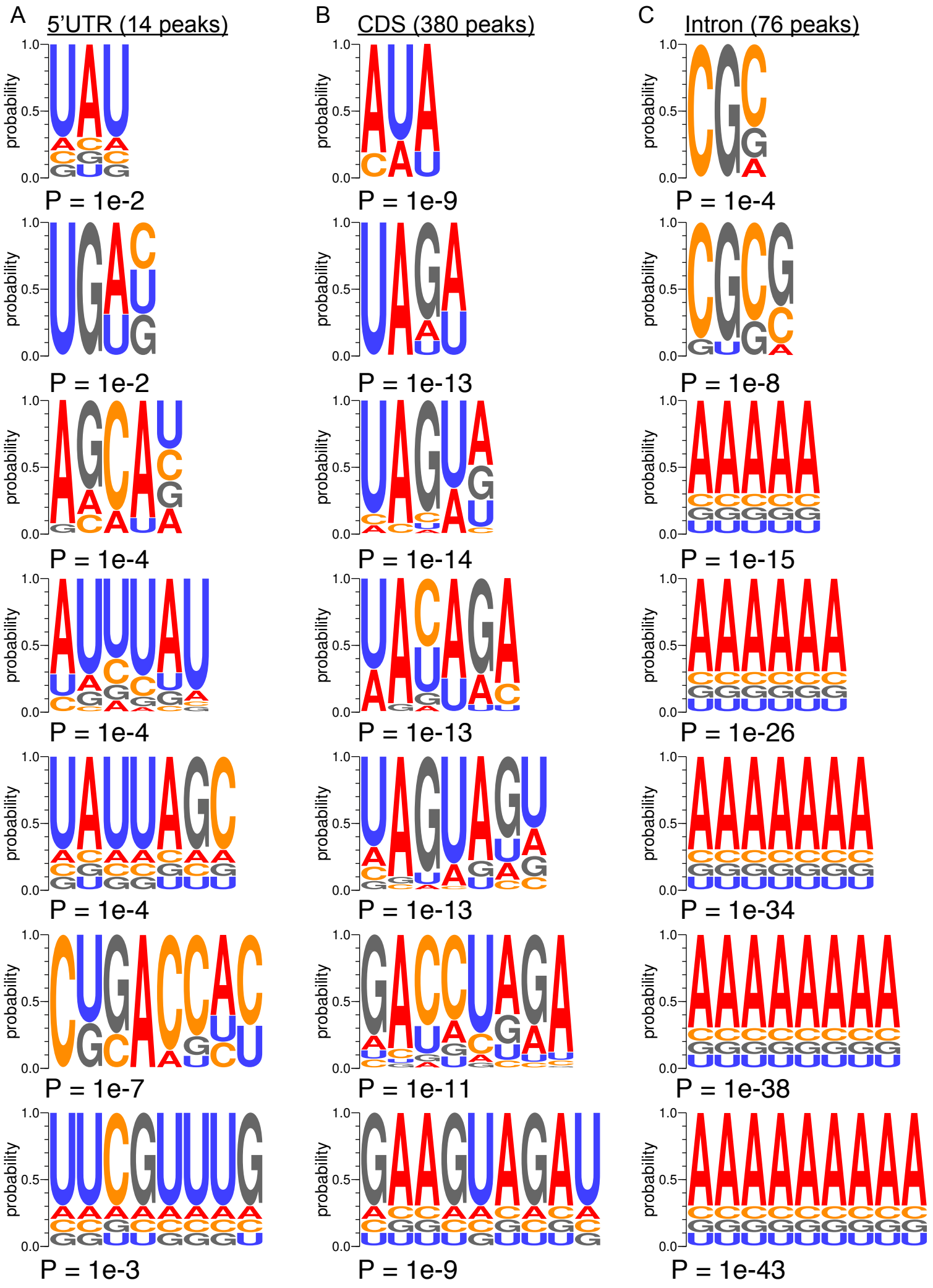
S2


A
Cross-linked mRNA Pool
Library Prep for
Illumina Platform
Isolate
Fragments
Add RNase
Radiogram of Protein-RNA
Fragments
RNase
Protein-RNA Fragments
Immunoprecipitate
RNA/Protin Fragments
Separate Fragment Size on
Denaturing Gel
Enriched Pool of Msi2
Bound RNA Fragments

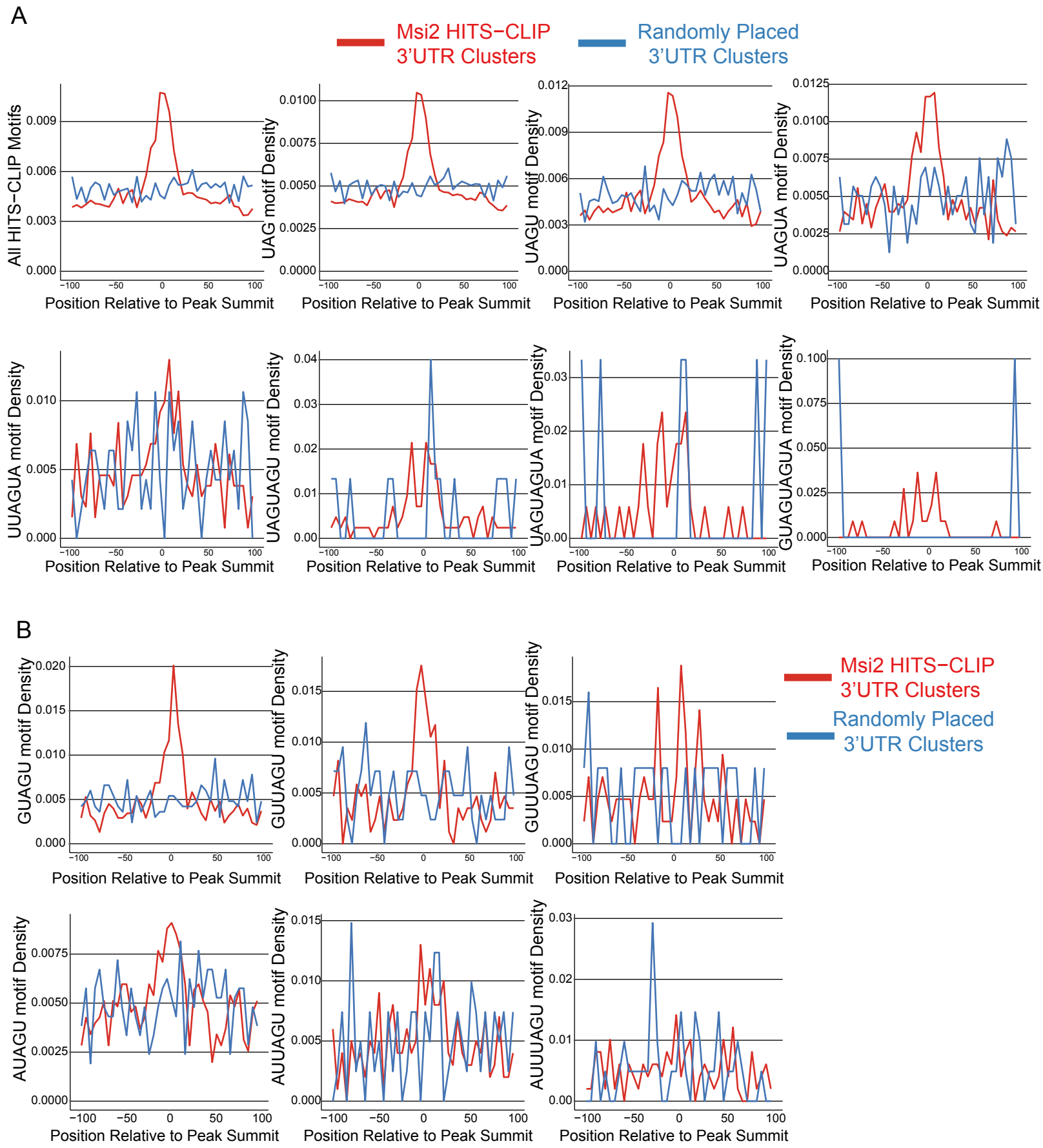

B
IgG
Msi2-IP
RNase A/T1 Dilution 1x 1x 1x 1:20 1:50 1:75
UVC Crosslink + - + + + +
160kD
110kD
80kD
60kD
50kD
40kD
30kD
Msi2 ~37kDa
1
2
3
4
5
Long
Fragments
Mid-sized
Fragments
Small
Fragments
Long
Fragments
Mid
Fragments
Small
Fragments
32P
Msi2


C
Short Fragment Lib (rep.1)
Mid fragments Lib. (rep.2)
Long Fragments Lib. (rep.3)
Long Fragments Lib. (rep.4)
Long Fragments Lib. (rep.5)
All Libraries Combined
Mean
Median
Number of Reads
Read Length

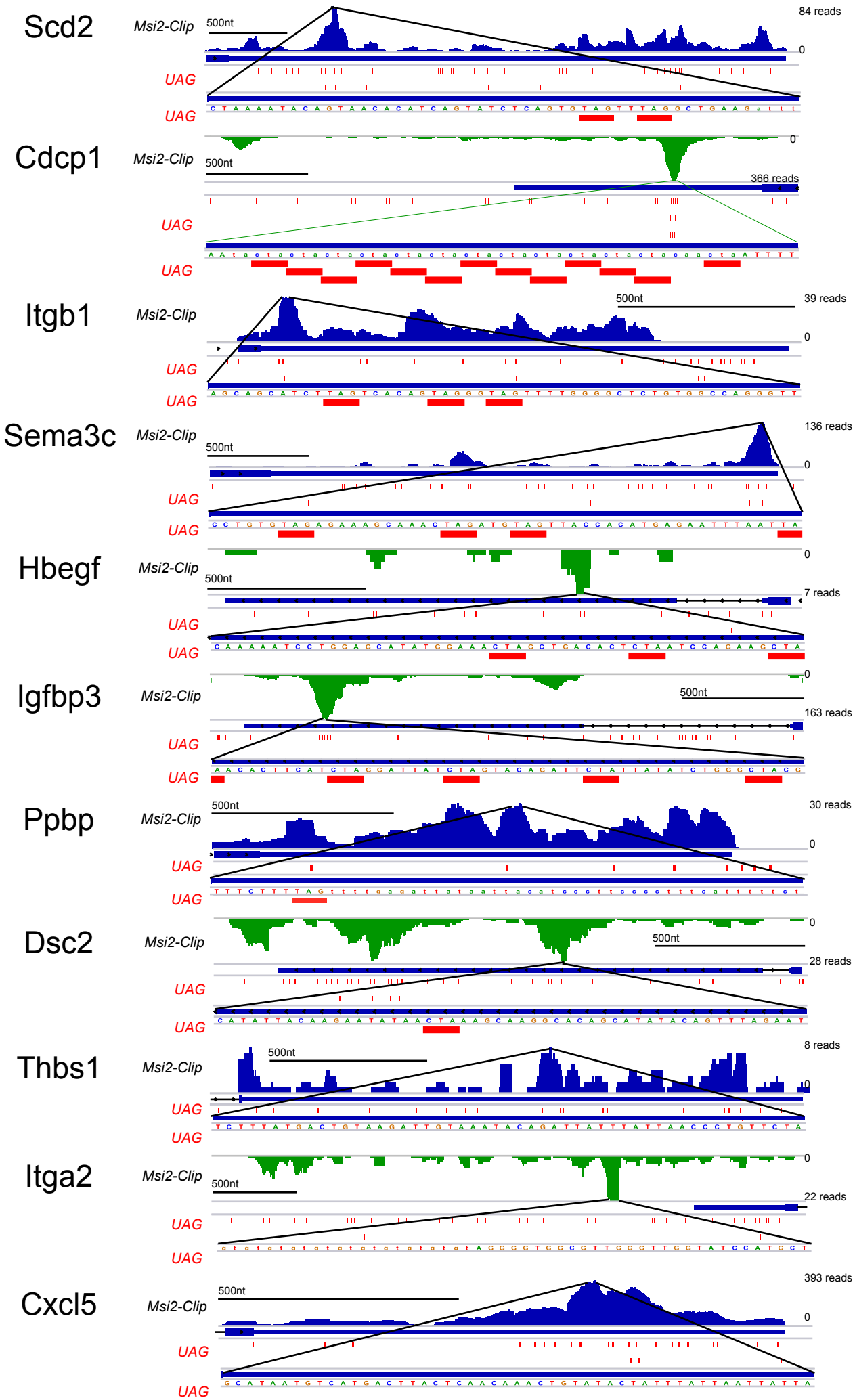
S3

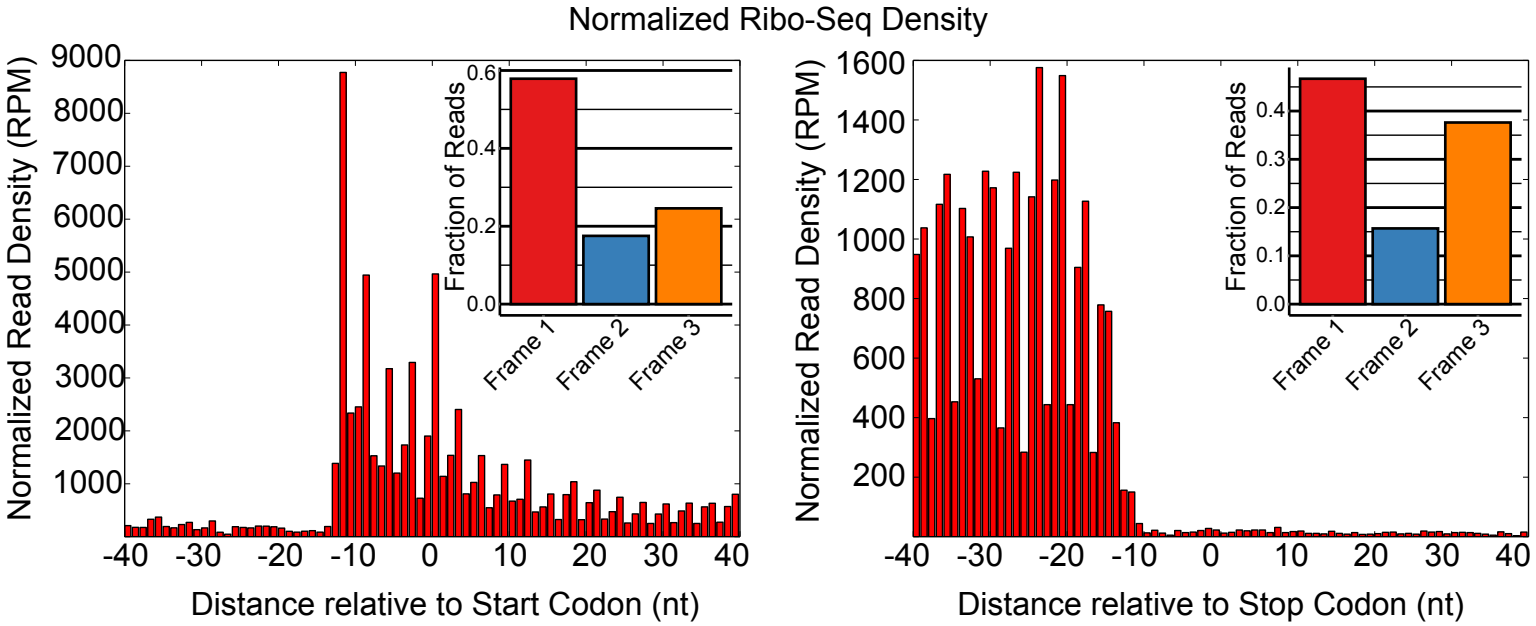

A
5'UTR (14 peaks)
B
CDS (380 peaks)
C
Intron (76 peaks)
probability
1.0
0.5
0.0
P = 1e-2
P = 1e-9
P = 1e-4
P = 1e-2
P = 1e-13
P = 1e-8
P = 1e-4
P = 1e-14
P = 1e-15
P = 1e-4
P = 1e-13
P = 1e-26
P = 1e-4
P = 1e-13
P = 1e-34
P = 1e-7
P = 1e-11
P = 1e-38
P = 1e-3
P = 1e-9
P = 1e-43

S4

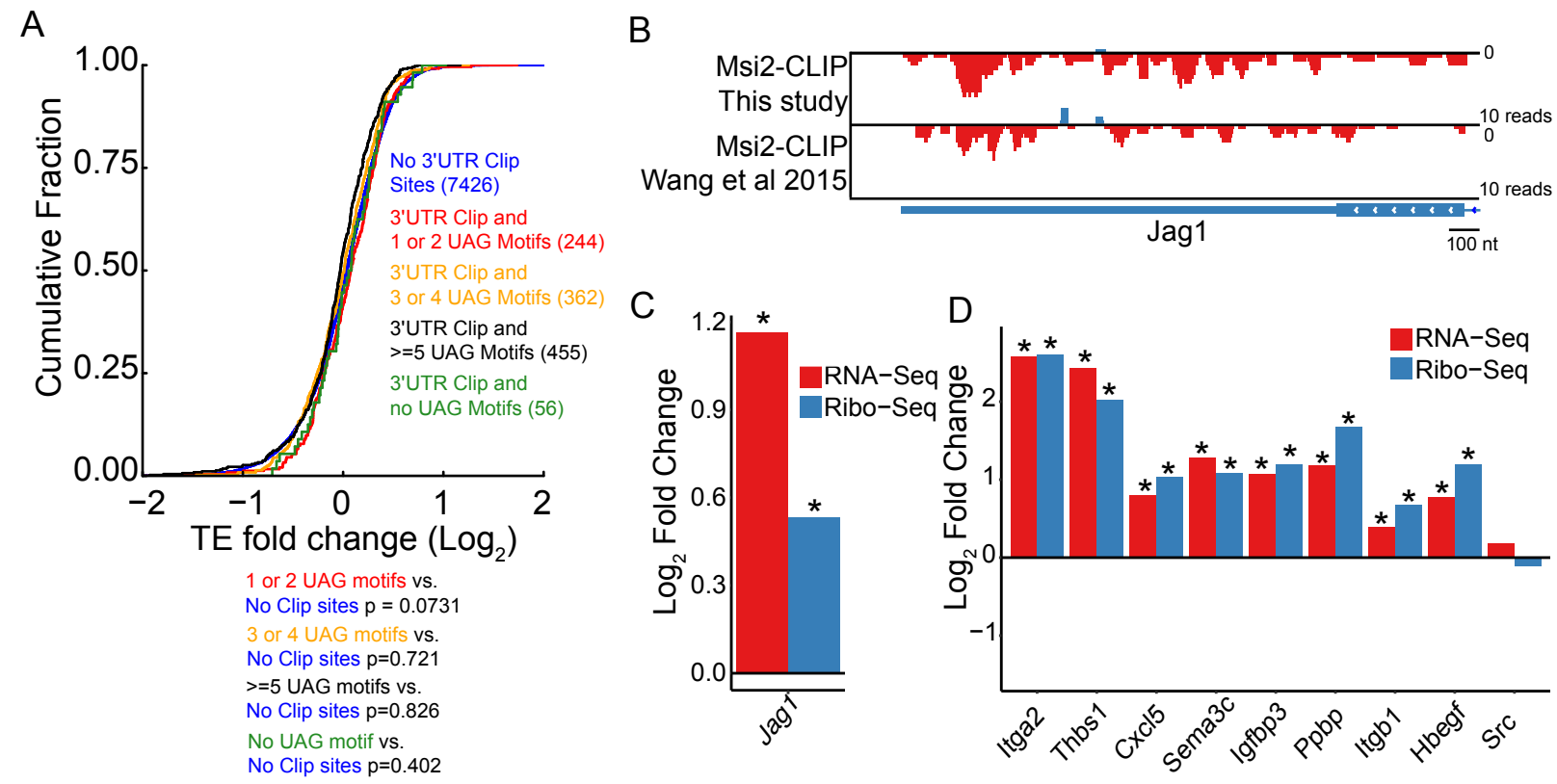

A
Msi2 HITS−CLIP 3'UTR Clusters
Randomly Placed 3'UTR Clusters
All HITS−CLIP Motifs
UAG motif Density
UAGU motif Density
UAGUA motif Density
UUAGUA motif Density
UAGUAGU motif Density
UAGUAGUA motif Density
GUAGUAGUA motif Density
Position Relative to Peak Summit
B
GUAGU motif Density
GUUAGU motif Density
GUUUAGU motif Density
AUAGU motif Density
AUUAGU motif Density
AUUUAGU motif Density
Msi2 HITS−CLIP 3'UTR Clusters
Randomly Placed 3'UTR Clusters
Position Relative to Peak Summit

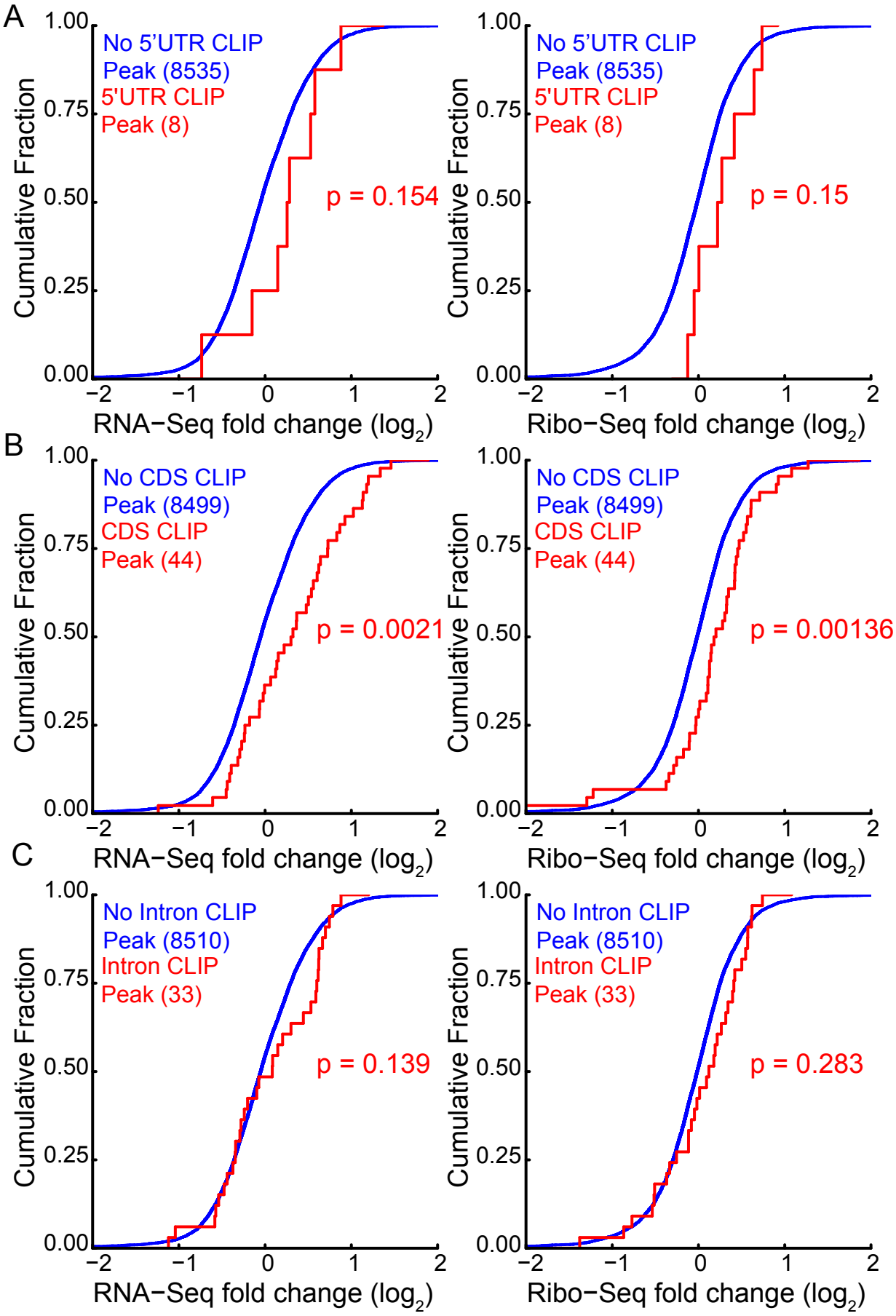
S5


Scd2
Msi2-Clip
500nt
84 reads
0
UAG
UAG
Cdcp1
Msi2-Clip
500nt
366 reads
0
UAG
UAG
Itgb1
Msi2-Clip
500nt
39 reads
0
UAG
UAG
Sema3c
Msi2-Clip
500nt
136 reads
0
UAG
UAG
Hbegf
Msi2-Clip
500nt
7 reads
0
UAG
UAG
Igfbp3
Msi2-Clip
500nt
163 reads
0
UAG
UAG
Ppbp
Msi2-Clip
500nt
30 reads
0
UAG
UAG
Dsc2
Msi2-Clip
500nt
28 reads
0
UAG
UAG
Thbs1
Msi2-Clip
500nt
8 reads
0
UAG
UAG
Itga2
Msi2-Clip
500nt
22 reads
0
UAG
UAG
Cxcl5
Msi2-Clip
500nt
393 reads
0
UAG
UAG

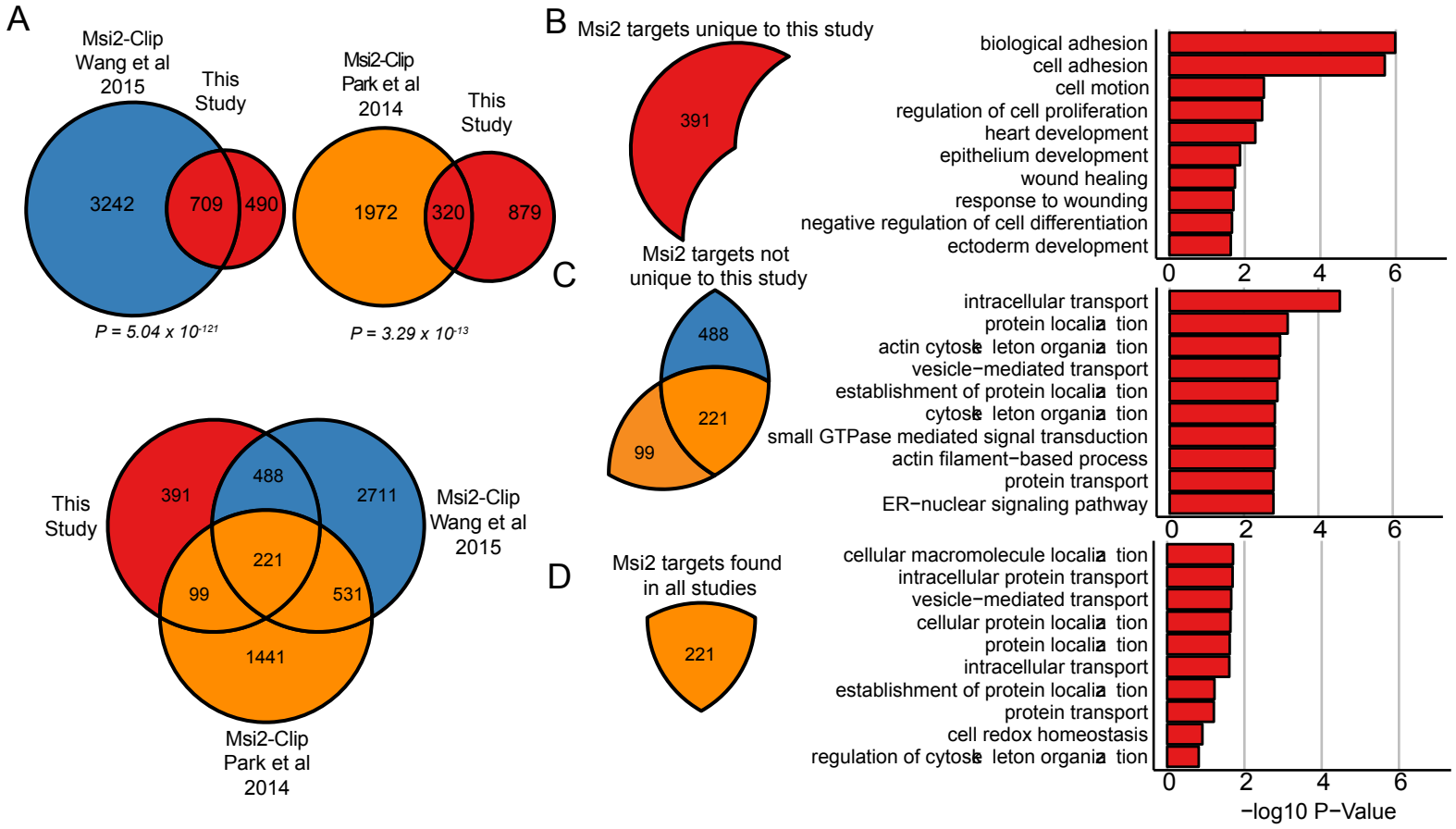

Normalized Ribo-Seq Density
Normalized Read Density (RPM)
9000
8000
7000
6000
5000
4000
3000
2000
1000
-40 -30 -20 -10 0 10 20 30 40
Distance relative to Start Codon (nt)
Fraction of Reads
0.6
0.4
0.2
0.0
Frame 1
Frame 2
Frame 3
Normalized Read Density (RPM)
1600
1400
1200
1000
800
600
400
200
-40 -30 -20 -10 0 10 20 30 40
Distance relative to Stop Codon (nt)
Fraction of Reads
0.4
0.3
0.2
0.1
0.0
Frame 1
Frame 2
Frame 3

S7


A
Cumulative Fraction
1.00
0.75
0.50
0.25
0.00
−2
−1
0
1
2
TE fold change (Log2)
No 3'UTR Clip Sites (7426)
3'UTR Clip and 1 or 2 UAG Motifs (244)
3'UTR Clip and 3 or 4 UAG Motifs (362)
3'UTR Clip and >=5 UAG Motifs (455)
3'UTR Clip and no UAG Motifs (56)
1 or 2 UAG motifs vs. No Clip sites p = 0.0731
3 or 4 UAG motifs vs. No Clip sites p=0.721
>=5 UAG motifs vs. No Clip sites p=0.826
No UAG motif vs. No Clip sites p=0.402
B
Msi2-CLIP This study
Msi2-CLIP Wang et al 2015
0
10 reads
0
10 reads
Jag1
100 nt
C
Log2 Fold Change
1.2
0.9
0.6
0.3
0.0
RNA−Seq
Ribo−Seq
Jag1
D
Log2 Fold Change
2
1
0
−1
RNA−Seq
Ribo−Seq
Itga2
Thbs1
Cxcl5
Sema3c
Igfbp3
Ppbp
Itgb1
Hbegf
Src


A
No 5'UTR CLIP
Peak (8535)
5'UTR CLIP
Peak (8)
p = 0.154
p = 0.15
B
No CDS CLIP
Peak (8499)
CDS CLIP
Peak (44)
p = 0.0021
p = 0.00136
C
No Intron CLIP
Peak (8510)
Intron CLIP
Peak (33)
p = 0.139
p = 0.283
Cumulative Fraction
RNA−Seq fold change (log2)
Ribo−Seq fold change (log2)

S9


A
Msi2-Clip Wang et al 2015
This Study
3242
709
490
P = 5.04 x 10-121
Msi2-Clip Park et al 2014
This Study
1972
320
879
P = 3.29 x 10-13
This Study
391
488
2711
Msi2-Clip Wang et al 2015
221
99
531
1441
Msi2-Clip Park et al 2014
B
Msi2 targets unique to this study
391
C
Msi2 targets not unique to this study
488
221
99
D
Msi2 targets found in all studies
221
biological adhesion
cell adhesion
cell motion
regulation of cell proliferation
heart development
epithelium development
wound healing
response to wounding
negative regulation of cell differentiation
ectoderm development
0
2
4
6
intracellular transport
protein localization
actin cytoskeleton organization
vesicle−mediated transport
establishment of protein localization
cytoskeleton organization
small GTPase mediated signal transduction
actin filament−based process
protein transport
ER−nuclear signaling pathway
0
2
4
6
cellular macromolecule localization
intracellular protein transport
vesicle−mediated transport
cellular protein localization
protein localization
intracellular transport
establishment of protein localization
protein transport
cell redox homeostasis
regulation of cytoskeleton organization
0
2
4
6
−log10 P−Value

RNA-Seq Mapping Statistics

	Scrambled shRNA			Msi2 shRNA		
	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3
Raw Reads	34,915,136	46,588,226	21,659,727	18,678,045	25,786,006	56,068,489
Uniquely Mapped	31,838,134	42,016,475	19,181,729	16,317,461	21,791,747	48,595,619
Unique Alignment (%)	91.19	90.19	88.56	87.36	84.51	86.67
Unique Alignment to exons (%)	84.67%	83.11%	79.97%	76.77%	71.76%	75.37%
Unique Alignment to CDS (%)	56.93%	53.52%	52.13%	49.37%	46.95%	47.77%

Ribo-Seq Mapping Statistics

	Scrambled shRNA			Msi2 shRNA		
	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3
Raw reads	21,330,185	30,971,905	14,418,386	38,448,389	21,734,079	25,236,826
Reads remaining after trimming	20,782,378	29,549,952	14,306,681	38,320,475	21,692,332	25,150,111
Reads remaining after rRNA mapping	19,010,247	27,374,135	13,943,666	37,382,044	21,237,040	24,360,146
Reads remaining after tRNA mapping	12,178,359	20,926,648	10,688,405	28,808,082	17,371,276	19,633,843
Reads remaining after ncRNA mapping	7,436,835	14,298,292	4,431,590	12,183,993	7,644,116	7,199,619
Reads Uniquely Mapped to CDS	2,196,405	2,959,625	702,602	2,069,440	512,256	998,028
Unique CDS Alignment (%)	10.30%	9.56%	4.87%	5.38%	2.36%	3.95%

Msi2 HITS-CLIP Mapping Statistics

	Keratinocytes					
	rep. 1	rep. 2	rep. 3	rep. 4	rep.5	Total
Raw Reads	7,694,905	8,130,804	4,646,423	7,886,026	5,696,326	34,054,484
Trimmed and Not sequence duplicate reads	451,785	1,188,587	1,091,490	1,556,367	1,145,519	5,433,748
Reads uniquely aligned ≥ 20 nt	58,738	738,408	857,567	1,202,388	851,987	3,709,088
Reads with unique alignment positions						358,074
Unfiltered Peaks						126,250
Filtered Peaks						4,051
5'UTR Peaks						14
CDS Peaks						380
3'UTR Peaks						2,671
Intronic Peaks						76
ncRNA Peaks						548
Intergenic Peaks						359