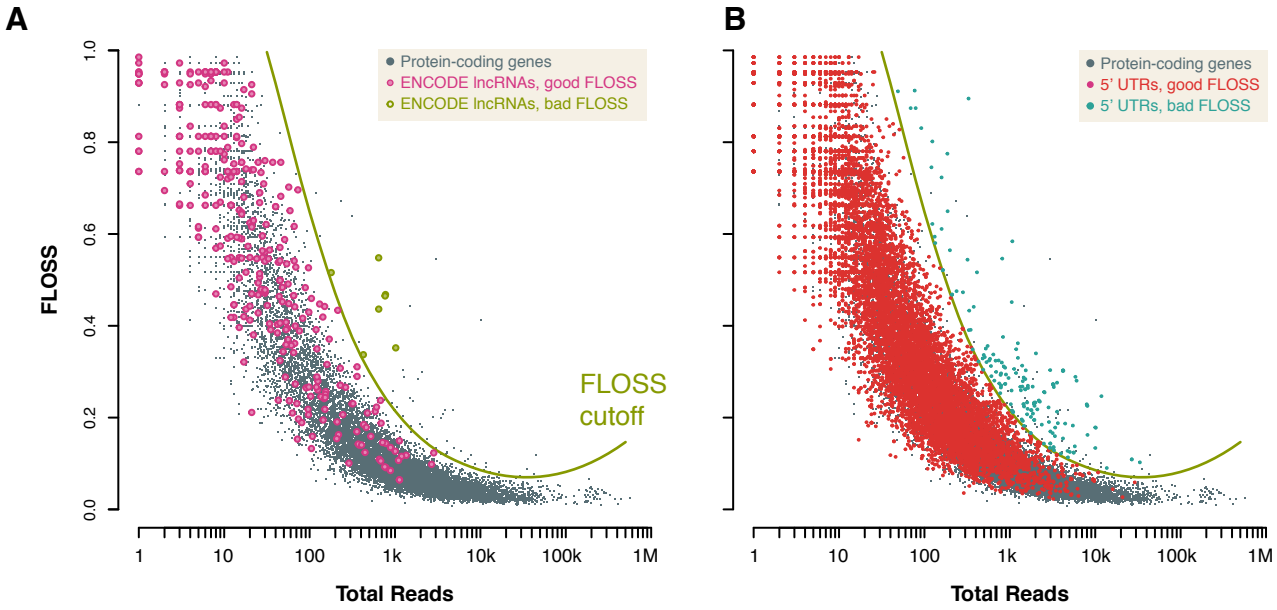

Figure S1. Classification of individual lncRNAs and 5' UTRs using the FLOSS. Related to Figure 1. (A, B) Fragment length analysis of mESC nuclear coding sequences showing a score cutoff based on the extreme outlier threshold of 3 x interquartile range over the 3rd quartile along a rolling 200 gene window. Individual (A) lncRNAs and (B) 5' UTRs are classified according to the FLOSS cutoff.

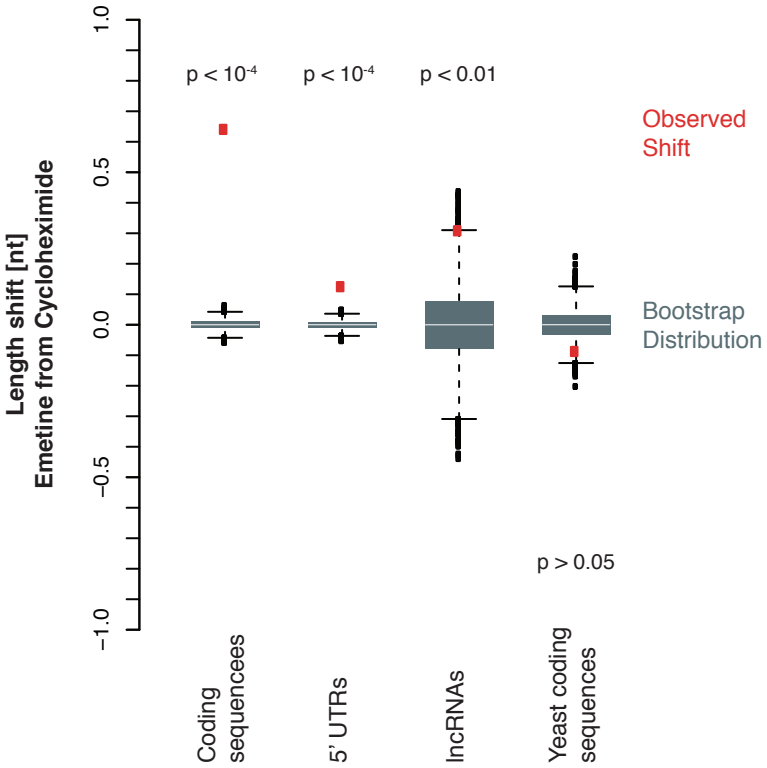

Figure S2. Bootstrap analysis of drug-dependent fragment length shifts. Related to Figure 2. The actual shift in cumulative length distribution between cycloheximide and emetine is shown, along with the distribution of shifts in bootstrap trials (N = 10,000). Bootstrapping was performed by assembling two sets of transcripts through selecting (with replacement) individual transcripts from a combined set of cycloheximide- and emetine-treated transcript data without regard for their drug treatment condition. Length distribution was computed for each of these bootstrapped transcript sets, and the length shift between the two randomly selected transcript sets was used as the result for that individual bootstrap trial.

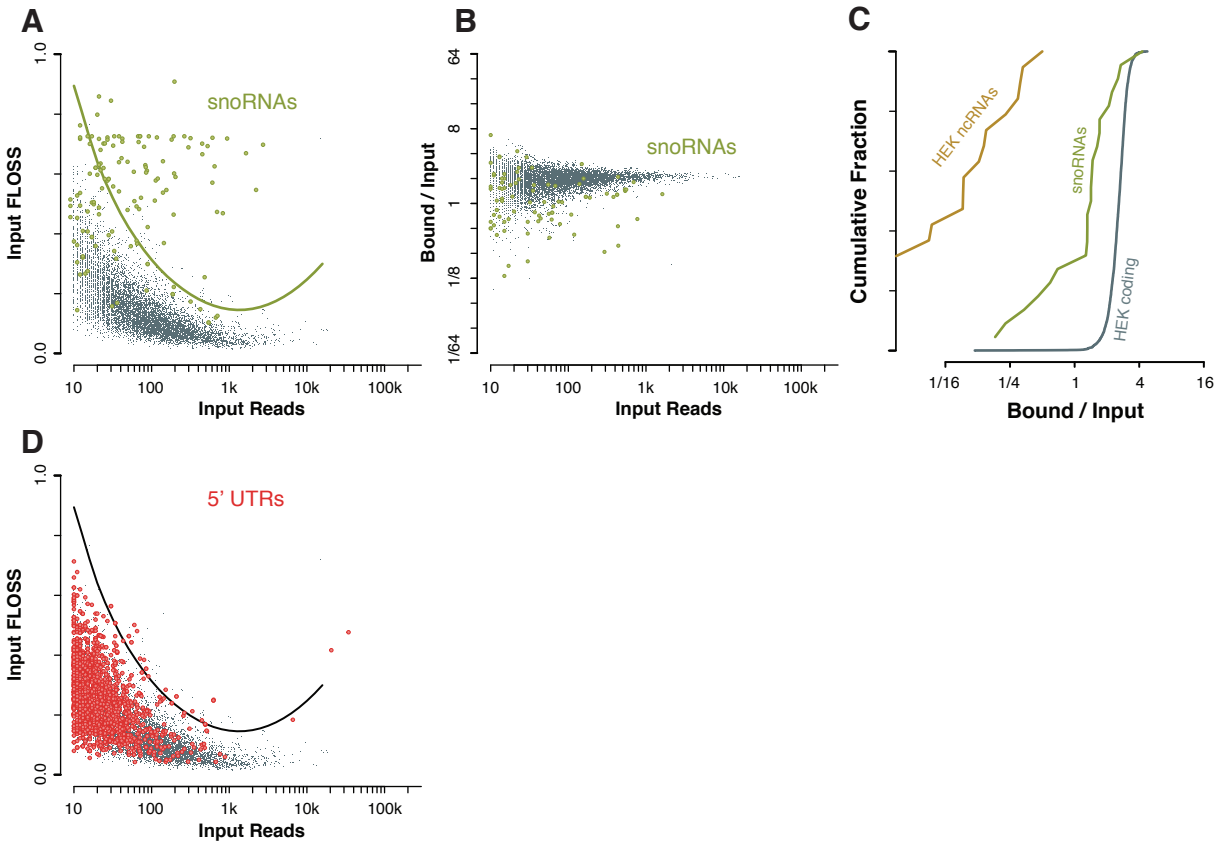

Figure S3. FLOSS analysis detects snoRNA-derived background that co-purifies with the ribosome. Related to Figure 3. (A) FLOSS analysis distinguishes snoRNA-derived background from true ribosome footprints. (B, C) SnoRNAs are substantially retained during ribosome affinity purification. (D) FLOSS analysis confirms that nearly all 5' UTRs resemble coding sequences in total HEK cell ribosome profiling.

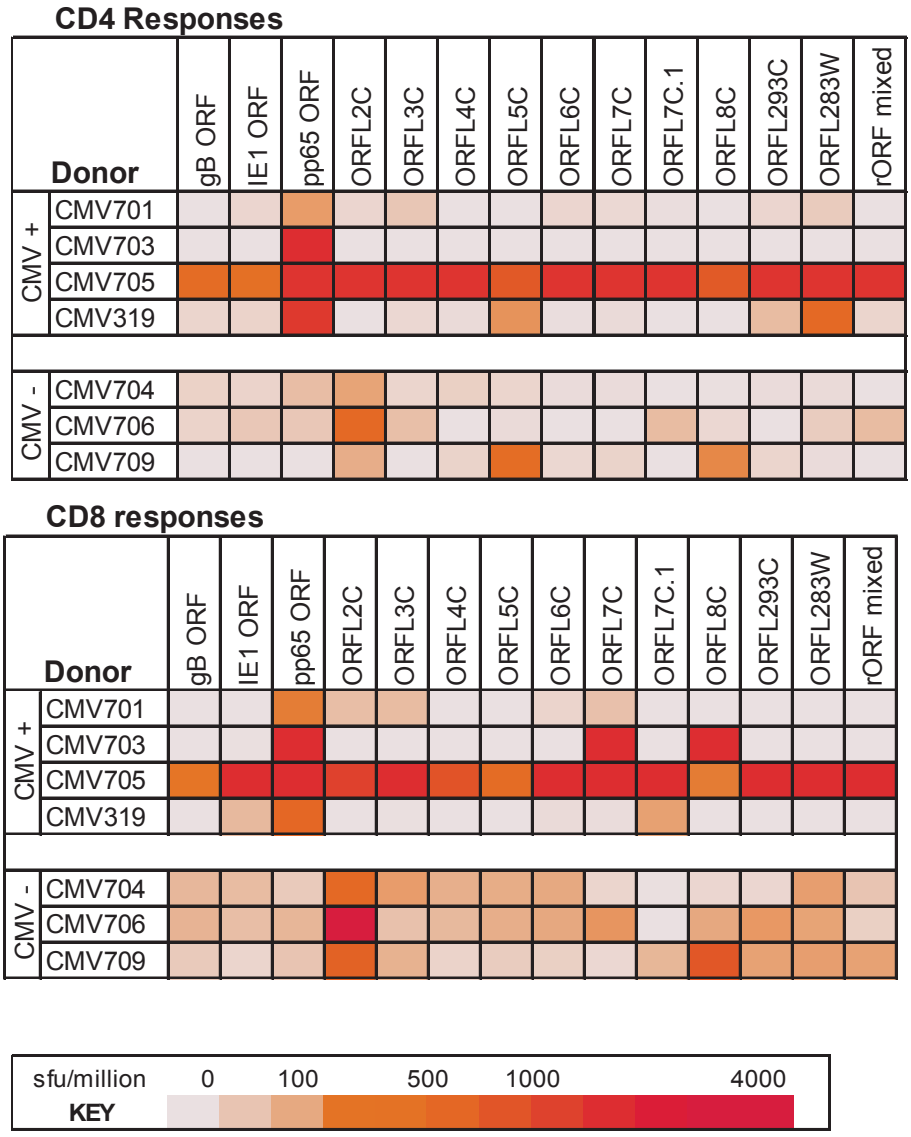

Figure S4. Antigen responses to novel CMV ORFs in CMV-positive but not CMV-negative donors. Related to Figure 6. Quilt plot summary of ELISPOT assays of four CMV positive donors and three CMV negative control.

Table S1. Fragment length analysis on mouse lncRNAs. Related to Figure 1.

Fragment length analysis data (read counts, FLOSS values, and classification relative to typical protein-coding genes) are shown for each mouse lncRNA transcript, along with genome annotation data.

Table S2. Affinity purification and fragment length analysis on human lncRNAs. Related to Figure 3. Affinity purification data (total input read counts and affinity-purified bound read counts) and fragment length analysis data (read counts and FLOSS values in total input profiling data, and classification relative to typical protein-coding genes) are shown for human lncRNAs with aligned footprint reads, along with genome annotation data for those transcripts. Fragment length analysis read counts include all aligned reads regardless of their length, whereas read counts for expression comparison include only reads in the 26 nt – 31 nt size range corresponding to ribosome footprints in this sample.

Table S3. Initiation site detection and reading frame bias on human lncRNAs. Related to Figure 4. Fragment length analysis data from each mouse lncRNA transcript are provided, as in Table S1, accompanied by data regarding the detection of harringtonine-marked AUG initiation sites and the bias of ribosome footprints for the downstream reading frames.

EXTENDED EXPERIMENTAL PROCEDURES

Mouse ES Cell Footprinting

Cell Growth and Lysis
E14 mouse ES cells were seeded at 4 x 10^4 cells per cm^2 in gelatin-coated tissue culture dishes in ESGRO Complete medium (Millipore), which was changed after 24 hours. Cells were grown at 37º C in 5% CO_2. At 48 hours, cells were dissociated with ESGRO Accutase (Millipore), counted, and re-plated at 4 x 10^4 cells per cm^2 in gelatin-coated tissue culture dishes in ESGRO Complete medium. Each profiling sample was prepared from one 10 cm dish of ES cells plated 48 hours prior to harvesting. Immediately prior to harvesting, cycloheximide (100 µg / ml final from a 100 mg / ml stock in DMSO) or emetine (50 µg / ml final from a 50 mg / ml stock) was added by withdrawing ~1 ml of media, mixing it with the drug stock, and returning this to the dish. Drug treatment was conducted for 1 minute in an incubator. Dishes were removed from the incubator, media was aspirated, and the dish was placed on ice while adherent cells were washed once in 10 ml ice-cold 1x PBS containing drug. The wash was aspirated thoroughly and 400 µl lysis buffer (ribosome buffer (see below) + 0.5% Triton X-100, 25 U / ml Turbo DNase (Life Technologies)) was dripped onto the dish. The dish was tilted slightly and cells in lysis buffer were scraped and then pipetted off, yielding ~550 µl lysate per dish. Lysate was transferred to a clean non-stick RNase-free tube and pipetted extensively, then incubated 10 min. on ice. Lysate was then trituated several times through a 26-gauge needle and debris was pelleted by centrifugation for 10 minutes at 20,000 x g, 4º C. The supernatant was recovered. Ribosome buffer is 20 mM Tris pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, and 100 µg / ml cycloheximide.

Ribosome Footprinting
RNA concentrations in mouse ES cell lysates were estimated using the Quant-iT RNA Assay kit (Molecular Probes) with fluorescence measurements on a SpectraMax M5 (Molecular Devices). The RNA in 2.0 µl of a 5-fold dilution was measured at 60.6 ng (Chx) or 56.8 ng (Emet), indicating an undiluted RNA concentration of ~0.15 µg / µl in the lysate. A lysate volume containing 56.8 µg RNA was diluted to 400 µl with lysis buffer and 12.0 µl of a diluted yeast lysate, prepared as described (Ingolia, 2010), was added. Ribosome footprinting was performed by adding 2.0 µl E. coli RNase I (10 U /µl; Epicentre) and incubating 45 min. on a room temperature nutator. Digestion was stopped by the addition of 10.0 µl SuperAseIn (20 U / µl; Life Techologies) and samples were placed on ice.

Total Ribosome Recovery
Mouse ES cell footprinting reactions were loaded into 13x51 mm thick-walled ultracentrifuge tubes (Beckman Coulter) and an 0.90 ml sucrose cushion (1 M sucrose in ribosome buffer) was underlaid. Ribosomes were precipitated by ultracentrifugation for 4 hours at 70,000 rpm, 4º C in a TLA100.4 rotor. The supernatant was discarded and a glassy ribosome pellet was recovered. The ribosome pellet was dissolved in 700 µl QIAzol (Qiagen) and total RNA including small (< 200 nt) RNAs was purified from the dissolved pellet using the miRNEasy Micro Kit (Qiagen) according to the manufacturer's

instructions. RNA was eluted in 100 µl total volume of nuclease-free water, yielding ~600 ng / µl RNA.

Footprint Library Generation

This RNA was precipitated by adding 1.0 µl GlycoBlue (Life Technologies), 12 µl 3M sodium acetate, and 150 µl isopropanol, followed by overnight precipitation on dry ice. Precipitated RNA was recovered by centrifugation for 30 minutes at 20,000 x g, 4º C, the supernatant was removed, and the RNA pellet was air dried for ~15 minutes. The RNA was then resuspended in 5 µl 10 mM Tris pH 8 and 5 µl denaturing loading dye (98% formamide, 10 mM EDTA, 300 µg / ml brompehnol blue) and separated by electrophoresis on a 15% polyacrylamide denaturing TBE-Urea gel (Life Technologies) at 200 V for 65 minutes. The gel was stained for 3 minutes in 1x Sybr Gold (Life Technologies) and visualized on a blue light transilluminator (Clare Chemical). A gel region between 26 nt and 33 nt was excised based on the migration of marker RNA oligos. The gel slice was transferred to a non-stick nuclease-free microfuge tube and 400 µl of RNA gel extraction buffer (300 mM sodium acetate pH 5.5, 1 mM EDTA, 0.25% SDS) was added. The gel slice was frozen on dry ice for 30 minutes and then transferred to a room temperature rocker overnight to extract RNA. Gel extraction buffer was recovered from the elution, 1.5 µl GlycoBlue was added, followed by 500 µl isopropanol, and RNA was precipitated and recovered as described above.

Size selected RNA was resuspended in 10 µl 10 mM Tris pH 8 and transferred to a new non-stick tube with 33 µl nuclease-free water. The same was denatured 90 s at 80º C and then equilibrated at 37º C. Dephosphorylation was carried out by adding 5.0 µl 10x T4 polynucleotide kinase buffer (New England Biolabs), 1.0 µl SuperAseIn, and 1.0 µl T4 polynucleotide kinase (New England Biolabs) and incubating 1 hour at 37º C. The enzyme was inactivated by heating for 10 min. at 70º C, and then RNA was precipitated by adding 39 µl water, 1.0 µl GlycoBlue, and 10 µl 3M sodium acetate pH 5.5 to each reaction, followed by 150 µl isopropanol. Precipitation and RNA recovery was carried out as described above.

Dephosphorylated RNA was resuspended in 8.5 µl 10 mM Tris pH 8 and 1.5 µl preadenylylated linker at 0.5 µg / µl (AIR adenylated linker A, Bioo Scientific) was added. This sample was denatured 90 s at 80º C and returned to room temperature. 2.0 µl 10x T4 RNA ligase buffer (New England Biolabs), 6.0 µl 50% polyethylene glycol, 1.0 µl SuperAseIn, and 1.0 µl T4 Rnl2(tr) (New England Biolabs) was added and mixed well. Ligation was carried out for 3 hours at room temperature. RNA was precipitated by the addition of 156 µl water, 20 µl 3M sodium acetate pH 5.5, 2.0 µl GlycoBlue, followed by 300 µl isopropanol, followed by chilling and centrifugation as described above. Ligation products were resuspended in 5.0 µl 10 mM Tris pH 8 and 5.0 µl 2x denaturing loading dye was added. Samples were separated by electrophoresis on a 15% polyacrylamide denaturing TBE-Urea gel and ligation products were excised. RNA was recovered from excised gel slices as described above.

Gel-purified ligation products were resuspended in 10.0 µl 10 mM Tris pH 8 and 2.0 µl reverse transcription primer at 1.25 µM was added. Nucleic acid samples were denatured 90 s at 80º C and then placed on ice. Reverse transcription reactions were set up with 4.0

µl 5x first-strand buffer, 1.0 µl dNTPs 10 mM each, 1.0 µl 0.1 M DTT, and 1.0 µl SuperAseIn, followed by 1.0 µl M-MuLV reverse transcriptase (New England Biolabs), and mixed well. Reverse transcription was carried out for 30 min. at 48º C and then RNA was hydrolyzed by adding 2.2 µl 1N NaOH and heating 20 min. at 98º C. Reverse transcription products were purified using an RNA Clean and Concentrator-5 column (Zymo Research) with elution into 7.0 µl 10 mM Tris pH 8. The first-strand cDNA was mixed with 7.0 µl 2x denaturing loading dye and separated by electrophoresis on a 15% polyacrylamide denaturing TBE-Urea gel. The reverse transcription product band was excised and DNA was extracted as described above, except using a DNA extraction buffer (300 mM NaCl, 10 mM Tris pH 8, 1 mM EDTA) in place of RNA extraction buffer.

First-strand cDNA was resuspended in 5.0 µl 10 mM Tris pH 8 and transferred to a new tube alogn with 1.0 µl 1 mM ATP, 1.0 µl 50 mM $MnCl_2$, 2.0 µl 10x CircLigase buffer, 10.0 µl water, and 1.0 µl CircLigase 2. Circularization was carried out for 60 min. at 60º C and then the enzyme was inactivated 10 min. at 80º C. Circles containing rRNA fragments were depleted by mixing 5.0 µl circularization reaction with 1.0 µl depletion oligo pool (10 µM each), 1.0 µl 20x SSC, and 3.0 µl water. Samples were heated to 100º C for 90 s and then annealed at 0.1ºC / s to 37º C, followed by a 15 minute incubation at 37º C. MyOne C1 DynaBeads (Life Technologies), 25 µl per sample, were prepared by washing three times in 1x bind/wash buffer (1M NaCl, 0.5 mM EDTA, 2.5 mM Tris pH 7.5, 0.1 % Triton X-100) and resuspended in 10 µl 2x bind/wash buffer per sample. Annealing reactions were combined with 10 µl washed beads and incubated 15 min. at 37º C with agitation. Beads were collected by placing the tubes in a magnetic rack and ~17.5 µl supernatant was recovered. DNA was recovered from the supernatant by adding 1 µl GlycoBlue, 6 µl 5M NaCl, 75 µl water, and 150 µl isopropanol, followed by precipitation as described above.

Depleted circles were resuspended in 10.0 µl 10 mM Tris pH 8 and used to prepare PCR amplification reactions with 12.0 µl 5x HF buffer, 1.2 µl dNTPs, 3.0 µl forward primer at 10 µM, 3.0 µl reverse indexed primer at 10 µM, 30 µl water, and 0.6 µl Phusion polymerase (New England Biolabs). Aliquots of 16.7 µl were amplified by thermal cycling (denaturation for 30s at 98º C followed by cycles of 10 s at 98º C, 10 s at 65º C, and 5 s at 72º C). Samples were amplified for 10 cycles of PCR and then separated by electrophoresis on an 8% non-denaturing polyacrylamide gel. The sequencing library band (~175 bp) was excised and the dsDNA library was extracted as described above. The library was validated and quantified using the High Sensitivity DNA kit on the BioAnalyzer 2000 (Agilent).

Library Generation Primers
Reverse transcription primer (NI-NI-9):
[Phos]AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCG
C[Sp-C18]CACTCA[Sp-
C18]TTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCTACAG
Forward primer (NI-NI-2): AATGATACGGCGACCACCGAGATCTACAC

Reverse index primer for Chx sample (NI-NI-6):
CAAGCAGAAGACGGCATACGAGATATGCTGGTGACTGGAGTTCAGACGTGTG
CTCTTCCG
Reverse index primer for Emet sample (NI-NI-7):
CAAGCAGAAGACGGCATACGAGATACGTCGGTGACTGGAGTTCAGACGTGTG
CTCTTCCG

Subtraction oligos:

NI-NI-21	[BioTEG]tcccggggctacgcctgtctgag
NI-NI-23	[BioTEG]gggcccaagtccttctgatcg
NI-NI-24	[BioTEG]gcctctccagtccgccgagg
NI-NI-63	[BioTEG]ACTCGCCGAATCCCGGGGCCGA
NI-NI-64	[BioTEG]GCGACCGGCTCCGGGACGGCT
NI-NI-65	[BioTEG]TTCACTGACCCGGTGAGGCGG
NI-NI-66	[BioTEG]CCTGGATACCGCAGCTAGGAATAA
NI-NI-356	[BioTEG]CTCGGTTGGCCYCGGATAGCCGG
NI-NI-357	[BioTEG]CTCGCTTCTGGCGCCAAGCGCCCG
NI-MH-1	[BioTEG]GAAGCCGAGCGCACGGGGTCGG
NI-MH-2	[BioTEG]GTCGGGGTTTCGTACGTAGCAGAGC
NI-MH-3	[BioTEG]CGATCTATTGAAAGTCAGCCCTCG
NI-MH-4	[BioTEG]GACTCTAGATAACCTCGGGCCGATC

Here [Phos] indicates 5’ phosphorylation, [Sp-C18] indicates an 18-atom hexa-ethylene glycol spacer, and [BioTEG] indicates a 5’ biotin attached by a tetra-ethylene glycol spacer.

Ribosome Affinity Purification

Ribosome Affinity Tagging

The ribosome affinity tag construct comprised human ribosomal protein L1 fused to the biotin acceptor peptide (Beckett et al., 1999; de Boer et al., 2003) with a 27 amino acid flexible linker, along with a human codon optimized biotin ligase co-expressed and separated by a 2A ribosome skipping peptide (de Felipe et al., 2006). This *RPL1-GS-TEVsite-Avi-T2A-hBirA* fusion was transgenically expressed from the pNTI194 expression vector. This plasmid is derived from the pcDNA5/FRT expression vector (Invitrogen) with CMV promoter and BGH poly-(A) site. The human *RPL1* sequence was amplified from human cell cDNA and fused translationally to a linker and affinity tag sequence encoding the polypeptide GGSSGSGSSGGSGSSGSSGGSENLYFQ**GLNDIFEAQ*K*IEWHE**, where the underlined amino acids comprise a TEV recognition site not used in this study, the bold peptide sequence comprises a biotinylation motif (Beckett et al., 1999; de Boer et al., 2003) for the *E. coli birA* biotin ligase, and the bold italic lysine is the biotinylated residue. The biotin acceptor peptide is followed by the T2A translational skipping peptide (de Felipe et al., 2006; Szymczak-Workman et al., 2012) and then by *hBirA*, a human codon optimized version of the *E. coli birA331,825* biotin ligase with Ala34Glu and

Arg33Leu mutations that abolish DNA binding without affecting enzymatic activity (Buoncristiani et al., 1986).

Stable pNTI194 expression was established by co-transfection into Flp-In HEK cells (Invitrogen) with pOG44, which expresses the Flp recombinase and selection for hygromycin resistance (150 µg / ml) according to the manufacturer's instructions. HEK cells expressing the affinity-tagged ribosome were propagated in the presence of hygromycin (150 µg / ml; InvivoGen) according to standard procedures. For ribosome profiling experiments, cells were plated in 10 cm dishes in the absence of hygromycin and expanded to ~50% confluency. Growth media was supplemented with 1 mM biotin (Sigma) 16 hours prior to cell harvesting. Cell harvesting and lysate preparation were performed as described for mouse ES cells. Affinity-tagged HEK lysates were mixed with yeast lysate containing an equivalent amount of total RNA and this mixture was subjected to nuclease footprinting as described above. Following nuclease digestion, lysates were loaded onto a Sepharcryl S-400 gel filtration spin column (Boca Scientific) and the flow-through was collected. One aliquot of flow-through was bound to streptavidin-coated magnetic beads (Invitrogen) and RNA was recovered by Trizol extraction directly from beads after washing. Another aliquot was used directly. Extracted RNA from affinity purification (bound) and ultracentrifugation (input) were subsequently converted into deep sequencing libraries according to the same procedures used for total ribosome footprint profiling.

Sequencing Read Counts

Cells	Treatment	Mapped Reads
mES Cells	Cycloheximide	254,516,745
	Emetine	170,514,739
HEK + Yeast	Input	26,071,571
	Bound	14,723,220
mES Cells (Ingolia et al. 2011)	Harringtonine 120s	33,355,973
	Harringtonine 150s	47,320,855

Fragment Length Organization Similarity Score

The fragment length organization similarity score (FLOSS) was computed by calculating a histogram of read lengths for all footprints that aligned to a specific transcript or reading frame, collapsing those below 26 nt or above 34 nt into the 26nt – 34nt range used in physical fragment size selection. A reference histogram was produced by summing individual raw counts (without normalization) for each annotated nuclear protein-coding transcript, excluding those whose gene overlapped a gene annotated as non-coding. The FLOSS was defined as

$$0.5 \times \sum_{l=26}^{34} \left\| f(l) - f_{\mathrm{ref}}(l) \right\|$$

where *f(l)* is the fraction of reads at length *l* in the transcript histogram and $f_{ref}(l)$ is the corresponding fraction in the reference histogram. The FLOSS cutoff, calculated as a function of the total number of reads in the transcript histogram, was established by considering a rolling window of individual annotated genes and the computing the upper extreme outlier cutoff for each window using Tukey's method (Q3 + 3*IQR, where Q3 is the 3rd quartile and IQR is the interquartile range).

T Cell Response Assays

Peptides constructed as sequential 15 amino acid peptides with 10 amino acid overlap, spanning ORFL2C, ORFL3C, ORFL4C, ORFL5C, ORFL7C, ORFL8C, ORFL293C and ORFL283W were obtained from JPT Peptide Technologies. Peptides were reconstituted to give a storage concentration of 40 mg/ml. Individual peptides were further diluted in RPMI 1640 to create a stock of 1 mg/ml. Peptide pools were made and were constructed to give 2 µg/ml of each individual peptide.

Blood was collected in heparin sodium (100 IU/ml), diluted 1:2 with RPMI-1640 containing no serum supplemented with 100,000 IU/ml penicillin, 100 mg/ml streptomycin, and 2 mmol/ml L-glutamine (RPMI-wash). Peripheral blood mononuclear cells (PBMC) were isolated by Lymphoprep (Axis-Shield, Norway) centrifuged at 800 g for 15 minutes.

ELISPOT plates were prepared, coated and blocked according to manufacturer's instruction (EBioscience). PBMC directly ex vivo, previously frozen, or depleted of either CD4+ or CD8+ T cells by magnetic activated cell sorting (MACS), were plated 3.0×105 cells in 100 µl RPMI-10 per well (of a 96 well Multiscreen IP sterile plate (Millipore, UK)). Plates were incubated for 48 hours at 37°C 5% CO2, and developed according to manufacturer's instruction.

PBMC were depleted of either CD4+ or CD8+ T cells by MACS using either anti-CD4+ or anti-CD8+ direct beads (Miltenyi, U.K.), according to manufacturer's instructions and separated on LS columns (Miltenyi, U.K.). Efficiency of depletion was determined by staining cells with either anti-CD4 or anti-CD8 antibodies and analysed by flow cytometry. Depletions performed in this manner resulted in 0.1–0.8% CD4+ cells and 0.3–0.8% CD8+ cells, respectively.

Supplemental References

Beckett, D., Kovaleva, E., and Schatz, P.J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci *8*, 921-929.
Buoncristiani, M.R., Howard, P.K., and Otsuka, A.J. (1986). DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (BirA) of Escherichia coli. Gene *44*, 255-261.
de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A *100*, 7480-7485.
de Felipe, P., Luke, G.A., Hughes, L.E., Gani, D., Halpin, C., and Ryan, M.D. (2006). E unum pluribus: multiple proteins from a self-processing polyprotein. Trends Biotechnol *24*, 68-75.
Ingolia, N.T. (2010). Genome-wide translational profiling by ribosome footprinting. Methods in enzymology *470*, 119-142.
Szymczak-Workman, A.L., Vignali, K.M., and Vignali, D.A. (2012). Design and construction of 2A peptide-linked multicistronic vectors. Cold Spring Harbor protocols *2012*, 199-204.