SUPPLEMENTAL FIGURE LEGENDS

Figure S1, Related to Figure 1. Genome-wide protein synthesis in mouse ES cells
(A) Metagene analysis of footprints at the start codon used to calibrate the A site position. Ribosome footprints prepared with no drug treatment were stratified by length, and total footprint count in annotated mouse genes, aligned at their beginning (0 is the first nucleotide of the start codon), shown for the predominant length categories. The A site codon begins at nucleotide +3.
(B) Metagene analysis of footprints at the stop codon. As (A), for genes aligned at the end (0 is the last nucleotide of the stop codon). The A site codon begins at nucleotide -2.
(C) Genome-wide protein synthesis. Total translation for each mouse gene was computed from the ribosome footprint density in the coding sequence of the canonical isoform (Table S1A). A histogram of log-scaled translation levels is shown.
(D) Genome-wide translational efficiency. Translational efficiency for each mouse gene with detectable mRNA abundance was computed from the ratio of ribosome footprint density to mRNA abundance from mRNA-seq data (Table S1C). A histogram of log-scaled translational efficiencies is shown.

Figure S2, Related to Figure 4. Identification of translation initiation sites and characterization of protein products
(A) Position-specific accuracy of initiation site prediction. As in Figure 4C, separating the non-start-codon test set into specific classes based on their codon position relative to the start codon. Many bars represent data from two positions with similar behavior, grouped together.
(B) Reading frame products by start codon. As in Figure 4F, for initiation sites separated into AUG (top) and near-cognate (bottom).
(C) Patterns of initiation and translation on two transcripts of the *Igf2* gene. The exon structure is shown with thin gray rectangles for the 5' UTR and thick gray rectangles for the annotated coding sequence. An mRNA-seq read profile is shown as well on an inverted y axis. Sequencing data for isoform-specific transcript positions are shown in dark colors and data for non-isoform-specific positions are shown with faint colors.
(D) Pattern of initiation and translation on the *Etv5* transcript. As in Figure 4H. Three alternate AUG reading frames are highlighted, including one that encodes a truncated protein, as well as a prominent short CUG reading frame overlapping the internal AUG initiation site. Two weak non-AUG reading frames are shown in blue.

Figure S3, Related to Figure 6. Translation on Alternate *Pih1d1* Transcripts
Patterns of initiation and translation on two *Pih1d1* transcripts. As Figure 6E, for the 5' end of the *Pih1d1* transcripts. Ribosome footprints from the annotated start codon overlap the alternative 5' splice junctions, so it is possible to determine which 5'UTR is associated with a footprint from an initiating ribosome.

Figure S4, Related to Figure 7. Translation During ES Cell Differentiation
(A) Schematic of samples during the differentiation timecourse.
(B) Changes in marker gene expression during differentiation. The total translation (sample-normalized ribosome footprint density) for several marker genes was scaled to

the maximum translation seen in any sample and the scaled translation levels are plotted. Expected expression periods are shown above (Leahy et al., 1999; Niwa et al., 2000; Niwa et al., 2009).
(C) Changes in translation during differentiation. The distribution of $\log_2$ fold-changes of translation (sample-normalized ribosome footprint density) is shown for all genes, 36 hours and 8 days after LIF withdrawal (see Tables S5A and S5D).
(D) Induction of developmental genes in embryoid bodies. As (C), showing the distribution for all genes as well as for genes with the "multicellular organismal development" GO annotation.
(E) Translational regulation of localized proteins in embryoid bodies. The cumulative distribution of log2 fold-changes in translational efficiency is shown for four broad GO localization categories.

SUPPLEMENTAL TABLE LEGENDS

Table S1, Related to Figure 1. Translation in ES cells
(A) Translation levels of genes in mouse ES cells. The representative transcript of each mouse gene is listed, with its UCSC identifier, ribosome footprint density (reads per base from cycloheximide-treated mouse ES cells), gene name, and gene description.
(B) Gene ontology analysis of translation. GO categories were tested for significant differences in translation by a Mann-Whitney test (corrected $p < 0.01$). All significant categories are shown, with the median $\log_2$ footprint density difference for genes in the category and genes not in the category used as a measure of the magnitude of the effect.
(C) Translation efficiency levels of expressed mRNAs in mouse ES cells. Each mouse gene with at least 50 mRNA-seq reads is listed, with its UCSC identifier, CDS ribosome footprint count, CDS mRNA-seq read count, translation efficiency ratio (ribosome footprint read count / mRNA-seq read count), $\log_2$ translation efficiency ratio, gene name, and gene description. The high translational efficiency of histone genes is an artifact, as histone transcripts are not polyadenylated and are thus under-represented in poly-(A) purified mRNA samples.
(D) Gene ontology analysis of translational efficiency. As (B), using $\log_2$ translational efficiency.

Table S2, Related to Figure 2. Ribosome pause sites
(A) Well-translated genes. Genes selected for pausing analysis and metagene profiles are listed, with the UCSC identifier of the representative transcript, the ribosome footprint density (reads per codon in the sample with no drug pre-treatment), and the gene name.
(B) Internal pause sites. CDS coordinates of 1543 ribosome pause sites, along with median CDS ribosome footprint count per codon, pause codon ribosome footprint counts, peptide sequences and nucleotide sequences surrounding these sites.
(C) Ribosome density at stop codons. Genes selected for pausing analysis with annotated 3'UTRs longer than 20 nucleotides are listed, along with the median CDS ribosome footprint count per codon, the ratio of stop codon ribosome footprint counts to the CDS median, the length in nucleotides of the CDS and the annotated 3'UTR, the peptide sequence at the end of the gene, the nucleotide sequence flanking the stop codon, and the gene name.

Table S3, Related to Figure 4. Sites of Translation initiation
Transcript coordinates of 13454 initiation sites, along with candidate initiation codons with flanking sequence context, initiation site predictor data, and reading frame classifications.

Table S4, Related to Figure 5. Translation of lincRNAs
For each detected lincRNA, the chromatin-defined lincRNA locus (Guttman et al., 2009) and the reconstructed transcript (Guttman et al., 2010) is listed along with the transcript coordinates of the window with maximum ribosome footprint coverage, per-nucleotide ribosome footprint read statistics (mean, median, and fraction non-zero in the 90 nt window), and per-nucleotide mRNA-seq read statistics (total count and average reads per nucleotide).

Table S5, Related to Figure 7. Changes in translation during differentiation
(A) Translation differences after 36 hours of LIF withdrawal. The representative transcript of each mouse gene is listed, with its UCSC identifier, $\log_2$ translation change (normalized footprint count in 36 hours –LIF divided by normalized footprint count in ES cells), $\log_2$ translation efficiency change (translational efficiency in 36 hours –LIF divided by translational efficiency in ES cells), gene name, and description. Normalized footprint count is the ribosome footprint count in the CDS divided by the total number of ribosome footprint reads that align to any region of any annotated transcript. Translational efficiency is the normalized footprint count divided by the normalized mRNA-seq read count, calculated in the same way. All ratios are comparisons between the same CDS in different conditions and so it is not necessary to normalize by the length of the CDS. In order to avoid statistically unreliable ratios, all translation comparison measurements require at least 100 reads total between the two conditions compared, and all translation efficiency measurements additionally require at least 50 mRNA reads in each sample.
(B) Gene ontology analysis of changes in translation. As Table S1B, using $\log_2$ translation change, 36 hours –LIF versus ES cells.
(C) Gene ontology analysis of changes in translational efficiency. As (B), using $\log_2$ translational efficiency change, -LIF versus ES cells.
(D) Translational differences in embryoid bodies. As (A), using embryoid bodies versus ES cells.
(E) Gene ontology analysis of changes in translation. As (B), using embryoid bodies versus ES cells.
(F) Gene ontology analysis of changes in translational efficiency. As (C), using embryoid bodies versus ES cells.

EXTENDED EXPERIMENTAL PROCEDURES

Cell Culture and Drug Treatment

E14 mESCs were plated on 15 cm dishes coated with gelatin (0.1% in Dulbecco's phosphate-buffered saline without calcium or magnesium salts) and grown in GMEM supplemented with 10% FBS, non-essential amino acids, glutamine, pyruvate, β-mercaptoethanol, and leukemia inhibitory factor (Tremml et al., 2008). Harringtonine (LKT Laboratories) treatment was performed by adding the drug to a final concentration of 2 μg / ml from a 2 mg /ml stock in DMSO. Cells were returned to 37°C following drug addition. Cycloheximide (Sigma) treatment was performed by adding the drug to a final concentration of 100 μg / ml from a 50 mg /ml stock in 100 percent EtOH. Cells were returned to 37° C for 1 minute following drug addition. Emetine (Sigma) treatment was performed in a similar manner, adding the drug to a final concentration of 20 μg / ml from a 100 mg/ml stock in DMSO.

Lysis

Medium was aspirated from dishes, which were immediately placed on ice and rinsed with 10 ml ice cold PBS supplemented with drugs used in pre-treatment of the cells. PBS was aspirated and 800 μl ice-cold lysis buffer consisting of polysome buffer (20 mM Tris pH 7.4, 250 mM NaCl, 15 mM $MgCl_2$, 1 mM dithiothreitol) supplemented with 0.5% Triton X-100 and 24 U / ml Turbo DNase (Ambion, AM2239), along with any drugs used for sample treatment) was dripped onto dishes. Cells were scraped extensively and triturated through a pipette. The lysate was removed and incubated 10 min on ice with periodic agitation. The lysate was then clarified by centrifugation for 10 min at 20,000x g, 4° C and ~1.1 ml supernatant was recovered.

Ribosome Footprinting

A 600 μl aliquot of lysate was treated with 15 μl RNase I 100 U / μl (Ambion, AM2295) for 45 minute at room temperature with gentle agitation. The digestion was stopped by the addition of 30 μl SuperaseIn 20 U / μl (Ambion, AM2696). Digestions were then immediately loaded onto a 1 M sucrose cushion, prepared in polysome buffer containing 0.1 U / μl SuperaseIn. Ribosomes were pelleted by centrifugation for 4 hr at 70,000 rpm, 4°C in a TLA-110 rotor. The liquid was removed and the pellet was resuspended in 570 μl 10 mM Tris pH 7, followed by the immediate addition of 30 μl 20% SDS. The sample was heated to 65°C and RNA was extracted using two rounds of acid phenol / chloroform followed by chloroform alone. RNA was precipitated from the aqueous phase by adding sodium acetate to a final concentration of 300 mM followed by at least one volume of isopropanol. Precipitation was carried out at -30°C for 30 min and RNA was then pelleted by centrifugation for 30 min at 20,000x g, 4° C. The supernatant was discarded, the pellet was air-dried, and the RNA was resuspended in 150 μl Tris pH 7. The typical RNA yield was 100 to 200 μg.

Size Selection

RNA was first filtered through a Microcon YM-100 (Millipore) to remove of RNAs greater than 100 nt. Subsequent experiments have shown that this filtration is unnecessary and that samples can be prepared by performing electrophoretic separation

directly on resuspended footprinting pellets (our unpublished observations). To perform filtration, a 50 μg RNA aliquot was diluted to a total volume of 500 μl and SuperaseIn was added to a final concentration of 0.1 U / μl. The sample was loaded onto a Microcon YM-100 (Millipore) and 50 U of SuperaseIn was placed in the collection tube. The sample was centrifuged for roughly 30 min at 510x g, to recover 425 μl filtrate containing only small RNAs. RNA was then precipitated from the filtrate as described above, except that 30 μg GlycoBlue (Ambion AM9515) was added as a coprecipitant. Filtered RNA was resuspended in 10 μl 10 mM Tris pH 7. Samples were separated by denaturing polyacrylamide gel electrophoresis (PAGE) in a 15% polyacrylamide gel with urea. Marker oligos oNTI199 and oNTI265 were used to demarcate the 28– 34nt region, inclusive, that was excised. The gel slices were physically disrupted by centrifugation through a needle hole from an inner 0.5ml microfuge tube nested in an outer 1.5ml collection microfuge tube. The gel was extracted in 200 μl RNase-free water for 10 min. at 70° C. The eluate was recovered by loading the gel slurry onto a Spin-X column (Corning 8160) and centrifuging to recover the eluate in the collection tube. RNA was then precipitated from the filtered eluate as described above, including the use of a co-precipitant.

mRNA-Seq Fragment Preparation

A 300 μl aliquot of lysate was diluted with 300 μl 10 mM Tris pH 7 and RNA was extracted as described above. Poly-(A)+ mRNA was purified from the total RNA sample using the Oligotex mRNA Mini kit (Qiagen) according to the manufacturer's instructions. The resulting mRNA was fragmented by partial hydrolysis in a bicarbonate buffer as previously described (Ingolia et al., 2009). The fragmented mRNA was separated by denaturing PAGE as described above and fragments of 50 – 80 nt length were selected. The fragmented mRNA was recovered from the gel as described above.

Library Generation

RNA samples were dephosphorylated using T4 polynucleotide kinase, which also possesses a 3'-phosphatase activity that is capable of removing 2',3'-cyclic phosphodiesters. Ribosome footprints and mRNA fragments were resuspended in 25 μl 10 mM Tris pH 8 and denatured for 2 min at 75° C. Samples were then equilibrated to 37° C and brought to a volume of 50 μl in 1x T4 polynucleotide kinase reaction buffer with 25 U T4 polynucleotide kinase (NEB M0201S) and 12.5 U SuperaseIn. This dephosphorylation reaction was incubated 1 hr at 37° C and heat-inactivated 10 min at 70° C. Dephosphorylated RNA was then purified by precipitation as described above. Linker attachment was then performed either by linker ligation or by enzymatic polyadenylation.

Linker ligation was carried out in a 20 μl reaction consisting of dephosphorylated RNA, 12.5% w/v PEG 8000 (prepared from a 50% w/v stock no more than 1 month old), 10% DMSO, 1x T4 Rnl2(tr) reaction buffer, 20 U SuperaseIn, 500 ng preadenylylated miRNA cloning linker (IDT, Linker #1), and 200 U T4 Rnl2(tr) (NEB, M0242L). The ligation was incubated for 2.5 hr at 37° C. Ligation products were separated by denaturing PAGE as described above, the product bands were excised, and RNA was extracted and precipitated.

Ribosome footprint samples were then treated by subtractive hybridization using biotinylated oligos that were reverse complements of abundant rRNA contaminants observed in preliminary sequencing experiments. RNA was suspended in 30 µl 2x SSC and 250 pmol total biotinylated subtraction oligos were added. The sample was denatured 2 min. at 70° C, transferred to 37° C, and 20 U SuperaseIn was added. Hybridization was performed for 30 min. at 37° C. Biotinylated oligos were then recovered using MyOne streptavidin C1 DynaBeads (Invitrogen) according to the manufacturer's instructions, using 1 mg magnetic beads per sample. RNA was then precipitated as described above.
Reverse transcription was then carried out on all samples. Reactions were prepared in a 18.0 µl volume using SuperScript III (Invitrogen) according to the manufacturer's instructions, using 50 pmol oNTI225-Link1 primer (for linker ligation samples) or oNTI225 (for polyadenylated samples). Reactions were denatured 5 min at 65°C and then equilibrated at 48° C, at which point 10 U SuperaseIn was added along with DTT and SuperScript III enzyme. Reverse transcription was carried out for 30 min at 48° C. RNA template was destroyed by adding 2.0 µl 1N NaOH and incubating 20 min. at 98° C.
Reverse transcription products were then purified by denaturing PAGE, incorporating an unextended primer control to facilitate the identification of extended products. These extended products were excised, extracted from the gel, and precipitated.
Reverse transcription products were circularized by carrying out a 20 µl CircLigase (Epicentre) reaction according to the manufacturer's instructions, using the entire extracted sample as a substrate.
Some circularized first-strand cDNA libraries were subject to a second round of subtractive hybridization as described above, except that 60 pmol total biotinylated subtraction oligos were used and no SuperaseIn was added. The subtractive oligos used with cDNA libraries were forward strand sequences derived from abundant rRNA contaminants.
One quarter of the circularized cDNA was used as a template for PCR amplification using Phusion (NEB). Reactions were prepared according to the manufacturer's instructions in a 100 µl volume using oligos oNTI230 and oNTI231. Reactions were split into five aliquots of 16.7 µl and amplified with a 30 s denaturation at 98° C followed by cycles of 10 s denaturation at 98° C, 10 s annealing at 65°C, and 5 s extension at 72°C. Reactions were carried out for 6, 8, 10, 12, and 14 cycles. Reactions were then separated by non-denaturing PAGE on an 8% polyacrylamide gel in 1x TBE. Product bands were excised from reactions selected to achieve reasonable yield without saturation, which manifests as reannealed library fragments that migrate slowly due to their imperfect complementarity. DNA was extracted as described above, except elution was carried out by soaking overnight in STE at 4°C. Extracted DNA was resuspended in 10 µl 10 mM Tris pH 8 and verified using the High Sensitivity DNA assay on the Agilent Bioanalyzer.

Oligonucleotide Sequences

oNTI199 (RNA): 5'-AUGUACACGGAGUCGACCCGCAACGCGA
oNTI225-Link1 (DNA): 5'-(P)
GATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT(Sp18)
CACTCA(Sp18)CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
oNTI230 (DNA): 5'-AATGATACGGCGACCACCGA

oNTI231 (DNA): 5’-CAAGCAGAAGACGGCATACGA
oNTI265 (RNA): 5’-AUGUACACGGAGUCGAGCUCAACCCGCAACGCGA
oNTI269 (DNA): (biotin)- TGGCGCCAGAAGCGAGAGCC
oNTI270 (DNA): (biotin)-AGACAGGCGTAGCCCCGGGA
oNTI298 (DNA): (biotin)-GGGGGGATGCGTGCATTTATCAGATCA
oNTI299 (DNA): (biotin)-TTGGTGACTCTAGATAACCTCGGGCCGATCGCACG
oNTI300 (DNA): (biotin)-GAGCCGCCTGGATACCGCAGCTAGGAATAATGGAAT
oNTI301 (DNA): (biotin)-TCGTGGGGGGCCCAAGTCCTTCTGATCGAGGCCC
oNTI303m (DNA): (biotin)-GGGGCCGGGCCGCCCCTCCCACGGCGCG
oNTI304m (DNA): (biotin)-
CCCAGTGCGCCCCGGGCGTCGTCGCGCCGTCGGGTCCCGGG
oNTI305 (DNA): (biotin)-TCCGCCGAGGGCGCACCACCGGCCCGTCTCGCC
oNTI306 (DNA): (biotin)-AGGGGCTCTCGCTTCTGGCGCCAAGCGT
oNTI307m (DNA): (biotin)-GAGCCTCGGTTGGCCCCGGATAGCCGGGTCCCCGT
oNTI308 (DNA): (biotin)-TCGCTGCGATCTATTGAAAGTCAGCCCTCGACACA
oNTI309 (DNA): (biotin)-TCCTCCCGGGGCTACGCCTGTCTGAGCGTCGCT

(P) designates 5’ phosphorylation, (Sp18) designates a hexa-ethyleneglycol spacer, and (biotin) designates biotin attached to the 5’ terminus by a C6 spacer.

Footprint Sequence Alignment

Ribosome footprints were aligned as described in (Ingolia, 2010). The first 26 nucleotides of each read were aligned using Bowtie (Langmead et al., 2009) and this alignment was then extended with reference library sequence followed by library generation linker sequence CTGTAGGCACCATCAATTCGTATGCCGTCTTCTGCTTGAA. The length of the reference sequence extension was chosen to minimize the number of mismatches between the read query sequence and the constructed reference-linker target sequence. The use of a 3' linker, rather than polyadenylation as described previously, avoided most ambiguity in the length of the reference sequence alignment. Alignments with up to two mismatches between the query and target sequence were accepted, and a small fraction of query sequences with over 255 alignments were suppressed.

Alignments were carried out first against a library of transcripts consisting of the mm9 UCSC Known Genes transcript sequences, downloaded on 2009 Dec 9, combined with sequences derived from the genome using the coordinates of the reconstructed transcripts in (Guttman et al., 2010). Reads with no acceptable alignment to these transcript sequences were then aligned against the mm9 genomic sequence.

Footprint alignments were assigned to a specific A site nucleotide based on the length of the fragment. The offset from the 5' end of the alignment was: 29-30 nt long, +15; 31-33 nt long, +16; 34-35 nt long, +17 (Figures S1A and S1B). Transcript density profiles were constructed by determining the number of sequencing reads whose A site was assigned to each nucleotide position.

Metagene Profiles

Footprint profiles within coding sequences were produced by assigning reads to a codon when they mapped to nucleotides at positions -1, 0, and +1 relative to the first nucleotide of the codon. Mean and median footprint levels were computed across coding sequences,

excluding the first fifteen and the last five codons. Non-overlapping five-codon windows were tiled across the body of the gene and well-translated genes were selected based on a median value of at least 2 reads per window in a cycloheximide-treated ES cell sample (Table S2A). Only the representative splice isoform of each gene was considered. Metagene analyses were performed by normalizing the footprint profile of each well-translated gene by the average footprint density across the body of the gene, again excluding the first fifteen and the last five codons. The scaled profiles of all well-translated genes were then averaged at each position, excluding genes from the average at positions that lay outside the bounds of their transcripts.

Translation and Translational Efficiency Calculations

The translation level of a gene was computed as the number of ribosome footprint reads mapping to the gene's coding sequence, divided by the length of the coding sequence in nucleotides, which we call the "ribosome footprint density" in the gene. When comparing translation measurements between different samples (see Figure 7A, Figures S4B, and S4B and Tables S5A and S5D), these measurements were normalized by dividing by the total number of ribosome footprint reads that align to any region of any annotated mouse transcript, which we call the "normalized ribosome footprint density." This normalization accounts for differences in the total number of sequencing reads and the extent of rRNA contamination in different samples.

The translational efficiency of a gene was computed as the ratio of the normalized ribosome footprint density to the normalized mRNA-seq read density. The latter is computed from mRNA-seq data just as normalized ribosome footprint density is computed and should reflect an estimate of mRNA abundance.

Quantitation of translation was highly reproducible in our experiments—the comparison presented in Figure 1A represents a biological replication, convolved with any cycloheximide-induced differences in translation. When the total number of footprints involved in a comparison is small, statistical fluctuations will dominate the ratio. When there are sufficient footprints for reliable comparison—in the case of Figure 1A, we require 200 in total—then the ratio reflects underlying biological and experimental variability. The typical inter-replicate difference of 15% indicates that measurements of footprint density are highly reproducible.

Gene Ontology Analysis

The Mann-Whitney U test was used to test for significant differences in the translation, translational efficiency, or change in translation or translational efficiency of genes within a certain GO category relative to the full list of genes analyzed. GO categories with at least 16 genes analyzed were tested and and the threshold for significance was an uncorrected $p < 1.7e\text{-}5$, which corresponds to a Bonferroni-corrected $p < 0.01$. Note that the Mann-Whitney is a non-parametric, rank-based test, so the monotonic transformation between raw values and log2 values does not affect the result. The median value for genes in the GO category and for genes not in the GO category was also computed and the difference was used as a measurement of the magnitude; median values are not directly used in the Mann-Whitney test.

Ribosome Pausing

The restricted set of genes used for metagene analysis was selected for pausing analysis as well (Table S2A) and each footprint profile was then scaled by the median footprint count on the transcript. Pauses were identified as codons whose ribosome footprint count was at least 25-fold the gene median. In order to avoid artifactually high read density caused by reads mapping to degenerate positions in the genome, pause sites were excluded when they aligned with a 28mer hypothetical footprint, beginning 15 nt upstream of the pause site, that had a perfect match in a distinct transcript. A meta-gene analysis was performed on these pauses as described above. The nucleotide and peptide sequence flanking these non-degenerate pause sites was used to generate a peptide sequence motif logo.

Ribosome Depletion Analysis

Metagene analysis was performed on well-translated genes following harringtonine treatment using a 5-codon window average as described above. Profiles were then scaled by the average along the metagene profile between codons 800 and 1000 inclusive. Harringtonine depletion did not extend beyond 750 codons at any timepoint used in this experiment, so the average read density beyond 800 codons serves as an appropriate normalization for overall sequencing coverage in a sample. The half-recovery point for each scaled profile was then defined as the position, beyond codon 40, where the profile first exceeded 0.5.

Low and high expression genes were defined as the lowest and highest quintile of well-translated genes sorted by mean ribosomes per codon in the coding sequence. Low and high TAI genes were defined as the lowest and highest quartile of well-translated genes sorted by TAI. Only genes of at least 750 codons were used for stratifying genes by length; the shorter subset of genes were less than 1000 codons and the longer subset was more than 1000 codons. Secreted proteins were defined based on SignalP predictions in the Ensembl database.

Initiation Site Prediction

Initiation site score vectors were constructed from footprint profiles in four harringtonine-treated samples: the 90 s, 120 s, and 150 s profiles described above, plus a fourth profile from a sample treated with 0.5 μg / ml harringtonine for 180 s in a preliminary experiment. An initiation site scoring vector was constructed for a nucleotide position by first collecting a per-codon footprint profile from the per-nucleotide footprint profile by summing reads at nucleotide positions at -1, +0, and +1 relative to the first nucleotide of a codon. Codons at the following positions, relative to the candidate initiation codon at position 0, were then summed as follows: [[-2, -1], [0], [1], [2], [3, 4], [5, 6, 7], [8, 9, 10], [11, 12, 13]]. This provided eight read counts per sample, which were concatenated to produce a 32-element initiation site profile vector. With the initiation codon defined as position 0, the A site position showing the strongest footprint accumulation was position +1. The training set was constructed using the first and second of every three genes in the set of well-translated genes (Table S1). The nucleotide position of the annotated start codon was used as a positive example and the following nucleotide positions, relative to the start codon, were used as negative examples: [-6, -3, 3, 9, 18, 30, 60, 90, 120, 150]. Positions without at least 18 nucleotides on each side of the initiation site scoring window, which covers nucleotides -7 through +40 relative to

the candidate initiation site, were excluded. The vectors were used to train the "svm_learn" program within "svm_light" using a radial basis kernel, $\gamma = 2.4$, error/margin trade-off $C = 2.0$, relative positive example weighting of 4.0 (though there were 10-fold more negative than positive examples in the training corpus), using iterative removal and retraining. The model was then tested on the third of every three genes in the set of well-translated genes (Table S2A), using the same positive and negative nucleotide positions. Initiation sites were predicted based on a minimum score of 0.75 and at least 50 harringtonine footprint reads overall in the scoring vector (covering forty-eight nucleotides per profile across four profiles). Contiguous blocks of initiation site nucleotide positions were collected into a single initiation site, whose width is given in the "Harr Peak Width" column of Table S3. The site was then assigned to an initiation codon whose first nucleotide was located within the site. If an AUG codon was present, then the initiation site was assigned to it. If no AUG was present but a near-cognate codon was present, then the site was assigned to it. Sites with no candidate initiation codon were excluded from further analysis because it was not typically possible to predict the reading frame being decoded.

Alternative isoform analysis

In order to identify initiation sites affected by alternative splicing, UCSC genomic coordinates and cluster IDs were used to determine if each initiation site was present in all splice isoforms of a gene, or only found in a subset of isoforms. We identified 1827 initiation sites that were absent in some splice isoforms and either annotated as a uORF in all isoforms containing the site, or as an overlapping uORF in all isoforms containing the site.

To identify differences in 5' UTR translation of distinct isoforms, ribosome footprint and mRNA-seq reads were mapped to the transcripts and classified as alternative or constitutive by the following method. We first created an index of isoform-specific sequence positions in the transcriptome by mapping each 28-nt window of each UCSC transcript with Bowtie against all other transcripts; matches to other transcripts of the same gene indicated that a window was shared by multiple isoforms of the gene. 28-nt windows that matched multiple genomic positions, *e.g.*, repetitive sequences, were discarded. The result was an annotation of each transcript position as isoform-specific or shared by multiple isoforms of the same gene. We also classified each position as 5' UTR, CDS, or 3' UTR in each isoform. Most positions were shared between multiple isoforms, leaving only a fraction of positions that could distinguish between isoforms. (The motivation of this indexing method was to count the total number of isoform-specific positions, not just those observed in sparse ribosome footprint data, to allow calculations of footprint density, *i.e.*, footprints per base.)

We then took the Bowtie alignments of all ribosome footprint and mRNA-seq reads to the transcriptome and tallied the number of reads falling in each category: isoform A 5' UTR, isoform A CDS + isoform B CDS, etc.

The 906 genes with alternative uORF initiation sites and sufficient isoform-specific sequence (at least 21 isoform-specific 5' UTR windows for at least two isoforms of the gene) were analyzed for differences in translation. A χ^2 test was used to test each gene for significant differences between ribosome and mRNA-seq read counts across each isoform-specific 5'UTR region, *e.g.*,

	footprint reads	mRNA-seq reads

5' UTR isoform A		
5' UTR isoform B		
5' UTR isoform C		

263 genes had a significant difference in ribosome footprint to mRNA ratios between isoforms, with a false discovery rate of 1%. (χ^2 results for genes with expected counts of 5 or less in any cell of the contingency table were discarded as not having enough data for a reliable comparison. This undercounts the biologically interesting examples by discarding genes with two highly expressed, differentially translated 5' UTR isoforms and a third isoform that is never observed in this cell type.)

LincRNA Analysis

Profiles of ribosome footprints on lincRNAs were computed using only alignments with a unique genomic origin in order to exclude the mis-assignment of reads that actually derived from a known protein-coding transcript. The genomic origin was defined by determining the genomic coordinate corresponding to the position of each transcript alignment. The 90 nucleotide window with the most nucleotide positions occupied by ribosome footprints was identified for each lincRNA profile, and the ribosome footprint and mRNA-Seq read density was determined for this window. The translational efficiency of the window was defined as the ratio of ribosome footprint read density to mRNA-Seq read density. As a comparison, the same analysis was performed for windows contained entirely within the codon sequence of the set of well-translated protein-coding genes, and similarly for the 3' UTRs of these genes.

SUPPLEMENTAL REFERENCES

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Leahy, A., Xiong, J.W., Kuhnert, F., and Stuhlmann, H. (1999). Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. J Exp Zool *284*, 67-81.

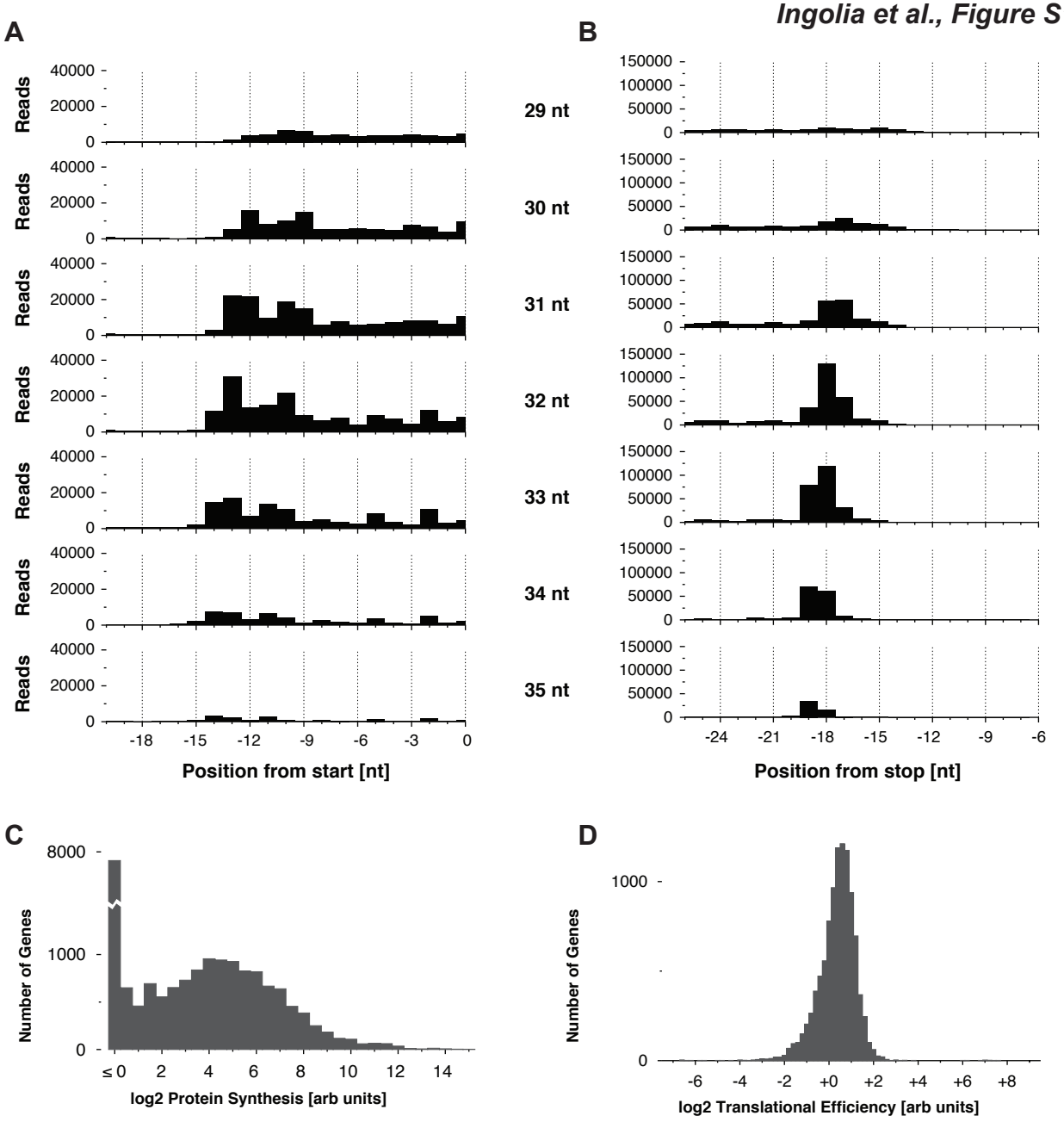
Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118-122.

Tremml, G., Singer, M., and Malavarca, R. (2008). Culture of mouse embryonic stem cells. Curr Protoc Stem Cell Biol *Chapter 1*, Unit 1C 4.

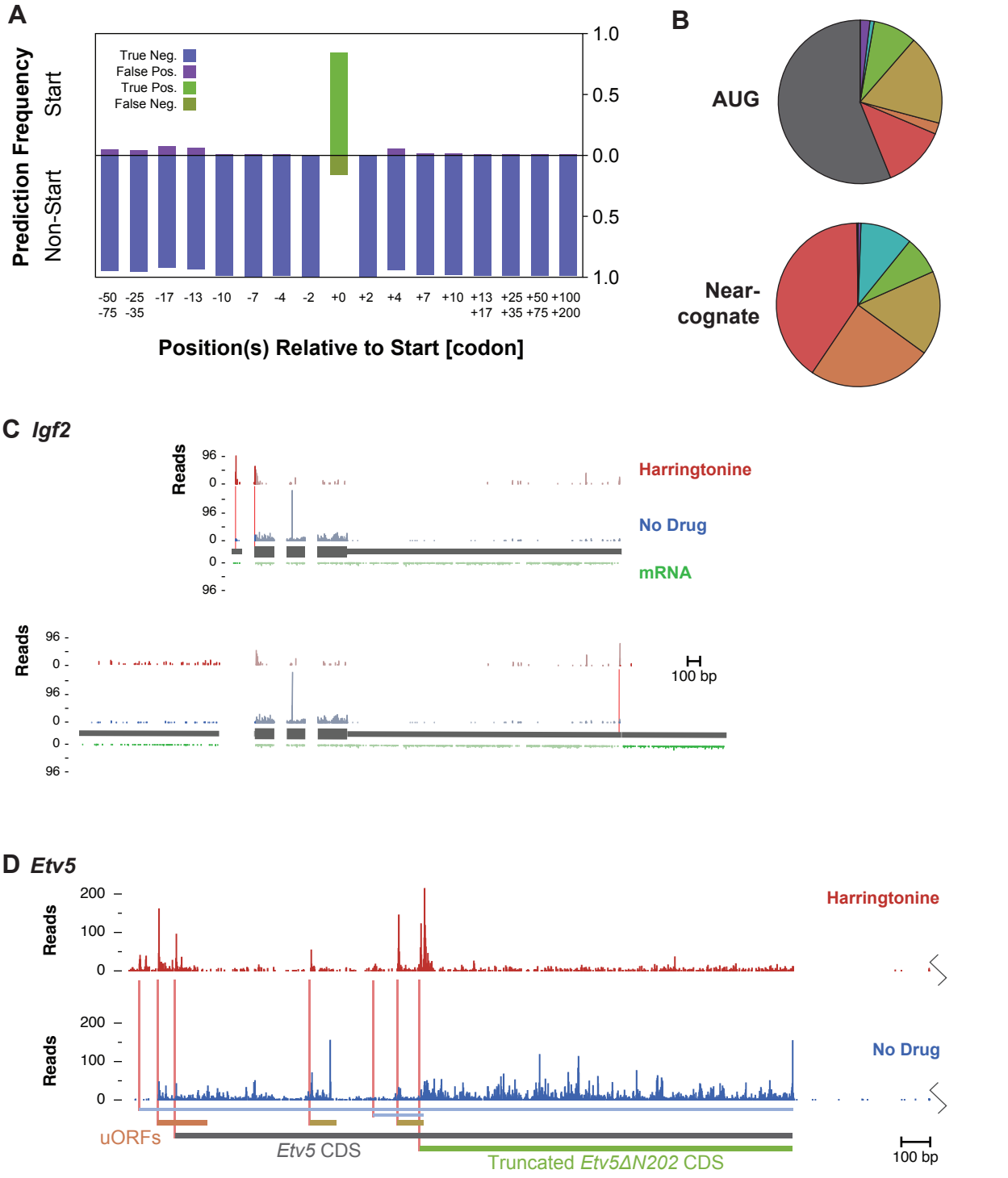
Supplemental Figure

Ingolia et al., Figure S1



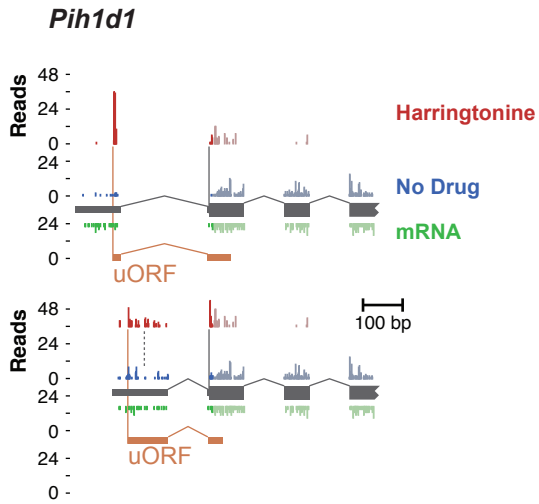
Supplemental Figure

Ingolia et al., Figure S2



Supplemental Figure

Ingolia et al., Figure S3



Supplemental Figure

Ingolia et al., Figure S4