

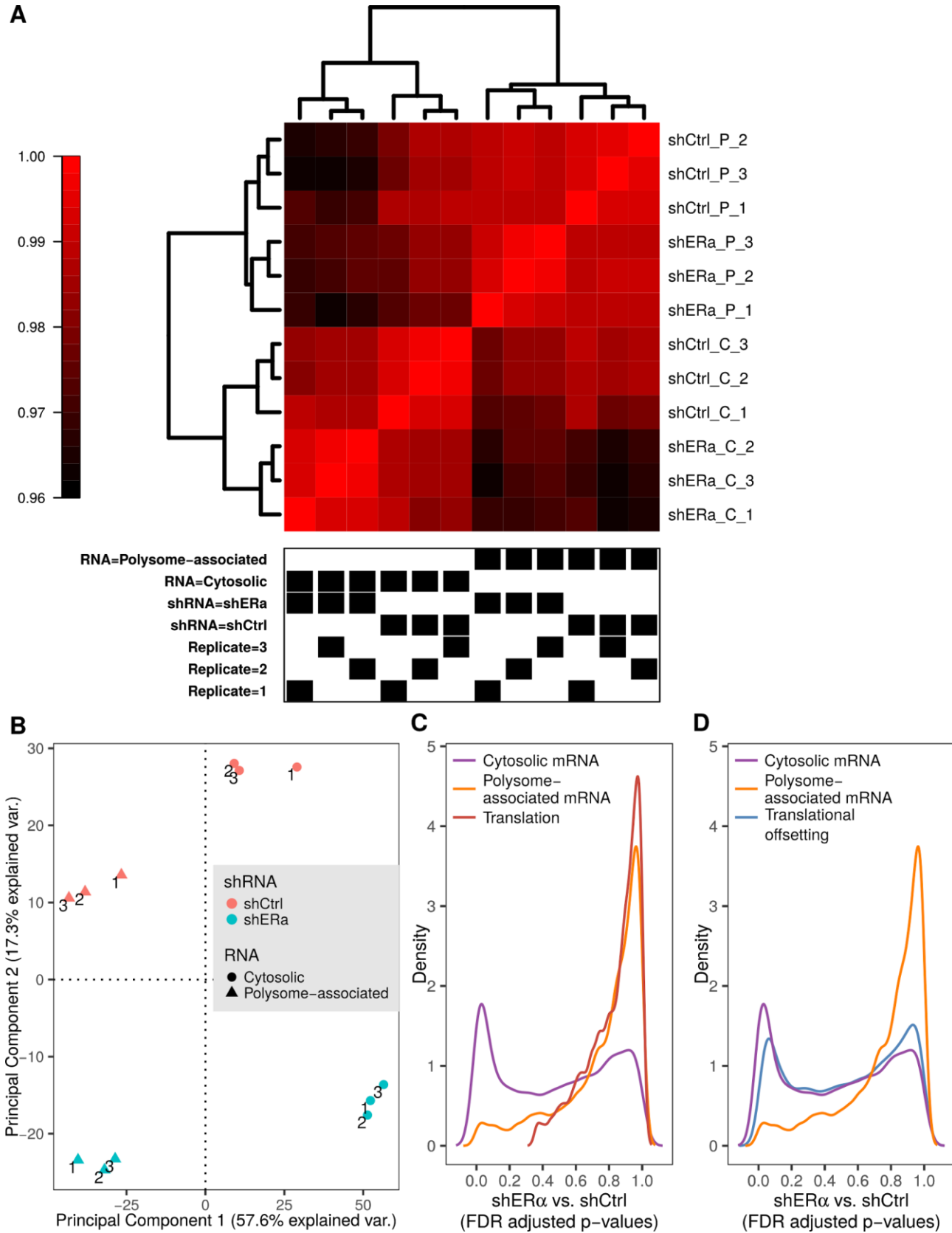
Translational offsetting as a mode of estrogen receptor α -dependent regulation of gene expression

APPENDIX

APPENDIX FIGURES	2
Appendix Figure S1	2
Appendix Figure S2	3
Appendix Figure S3	4
Appendix Figure S4	5
Appendix Figure S5	6
Appendix Figure S6	6
Appendix Figure S7	8
APPENDIX TABLES	9
APPENDIX SUPPLEMENTARY METHODS	11
Analysis of polysome-profiling data	11
Cycloheximide Chase Assay	11
Nanostring gene expression quantification and analysis	11
Preprocessing and analysis of RNAseq data	12
GO enrichment analysis	12
nanoCAGE library preparation and sequencing	12
nanoCAGE data processing and analysis	13
Preprocessing and analysis of RNAseq of small RNAs data	14
Analysis of codon usage	15
Quantification and analysis of tRNA levels	16
Analysis of public dataset for E2 dependent expression of ELP3, ALKBH8 and CTU2	16
Western blot	16
Analysis of codon usage in a public dataset of Tamoxifen-sensitive and resistant cell lines	17
REFERENCES	18

APPENDIX FIGURES

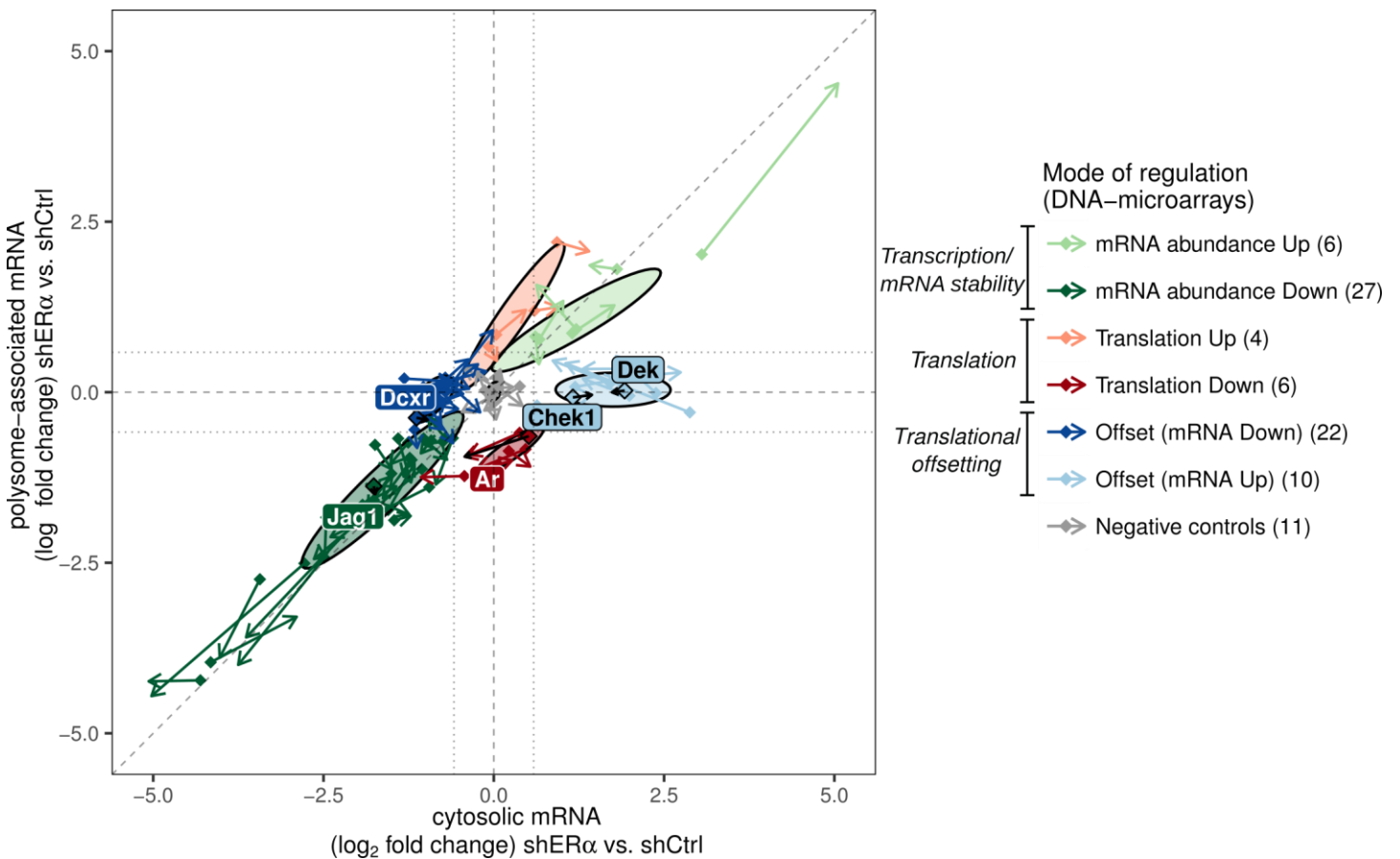
Appendix Figure S1



(Appendix Figure S1. Legend on next page)

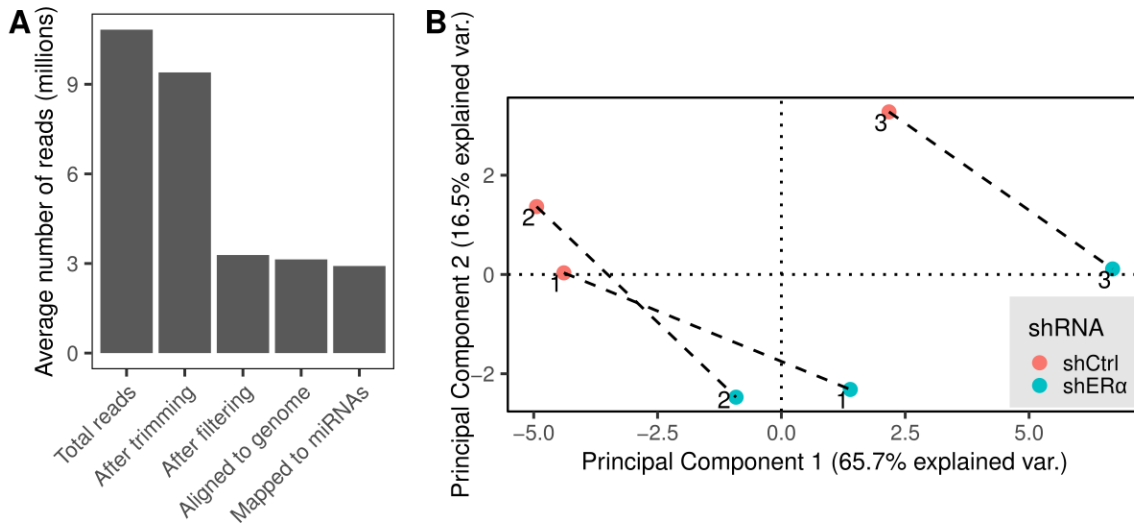
Appendix Figure S1. Quality control for DNA-microarrays data set assessing transcriptome-wide effects on gene expression programs using polysome-profiling (n = 3). (A) Heatmap of a correlation matrix between all samples from the transcriptome wide study quantified using DNA-microarrays (lower correlations in black and higher correlations in red). Samples were reordered according to the hierarchical clustering. (B) Samples from the transcriptome-wide polysome-profiling of ER α depletion (normalized data) were projected on the 2 first components of a centered principal component analysis. (C) Distributions (quantified by kernel density estimation) of FDR-adjusted p-values for differential expression between shER α and shCtrl BM67 cells using data from polysome-associated mRNA (orange) or cytosolic mRNA (purple) and for analysis of differences in translational efficiencies leading to altered protein levels (red). (D) Same plot as in (C) but including the distribution of FDR-adjusted p-values from analysis of translational offsetting (blue).

Appendix Figure S2



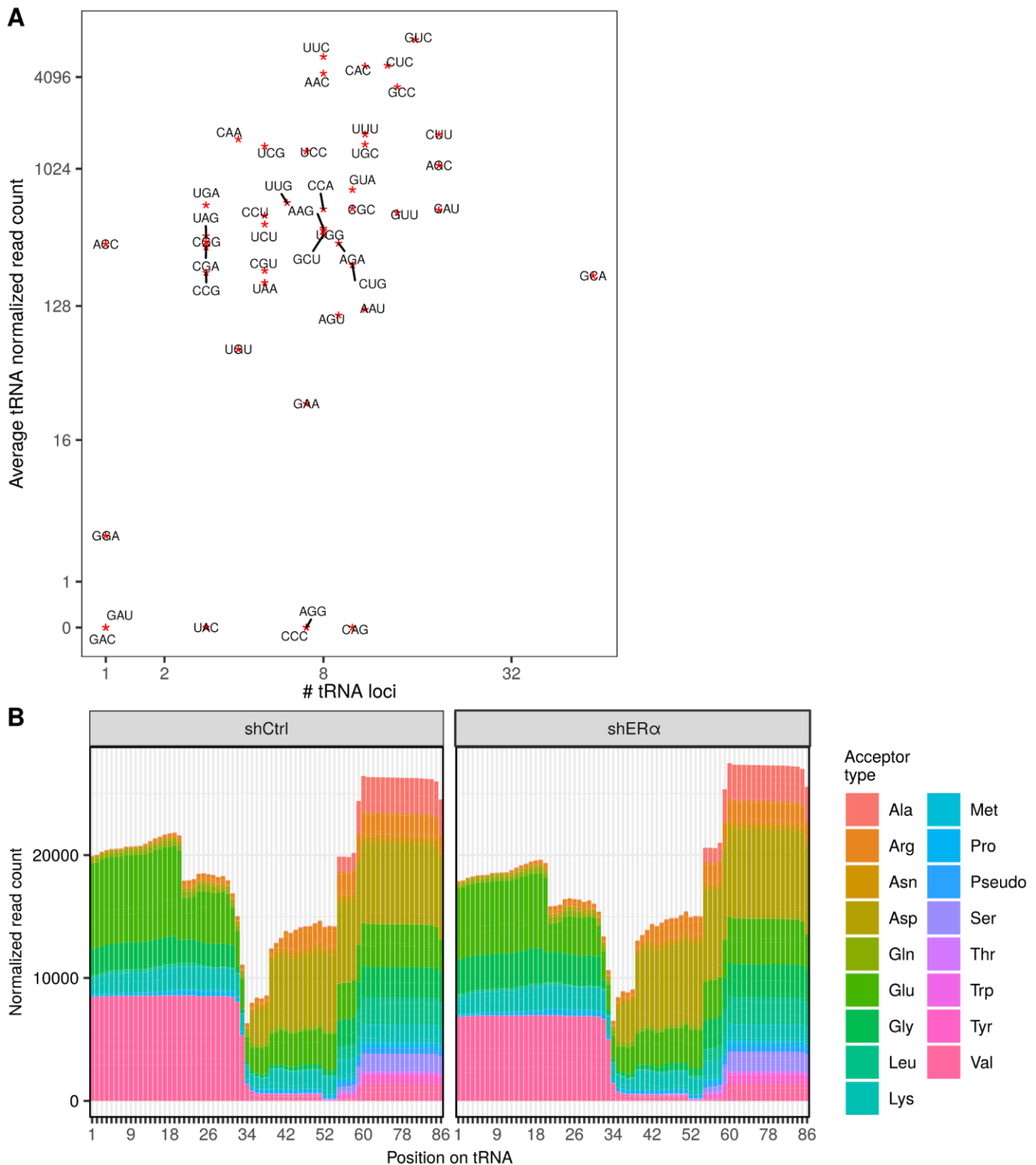
Appendix Figure S2. Validation of translational offsetting by Nanostring (n = 3). (A) Scatter plot of polysome-associated mRNA vs. cytosolic mRNA log₂ fold-change (shER α vs. shCtrl). Confidence ellipses are overlaid (based on fold-changes from Fig 1E). For each gene assessed using Nanostring, an arrow is plotted where the start of the arrow indicates fold-changes for the transcriptome wide study (i.e. Fig 1E) and the tip of the arrow by fold-changes as measured by Nanostring (i.e. Fig 2A). Genes and arrows are colored according to their mode of regulation.

Appendix Figure S3



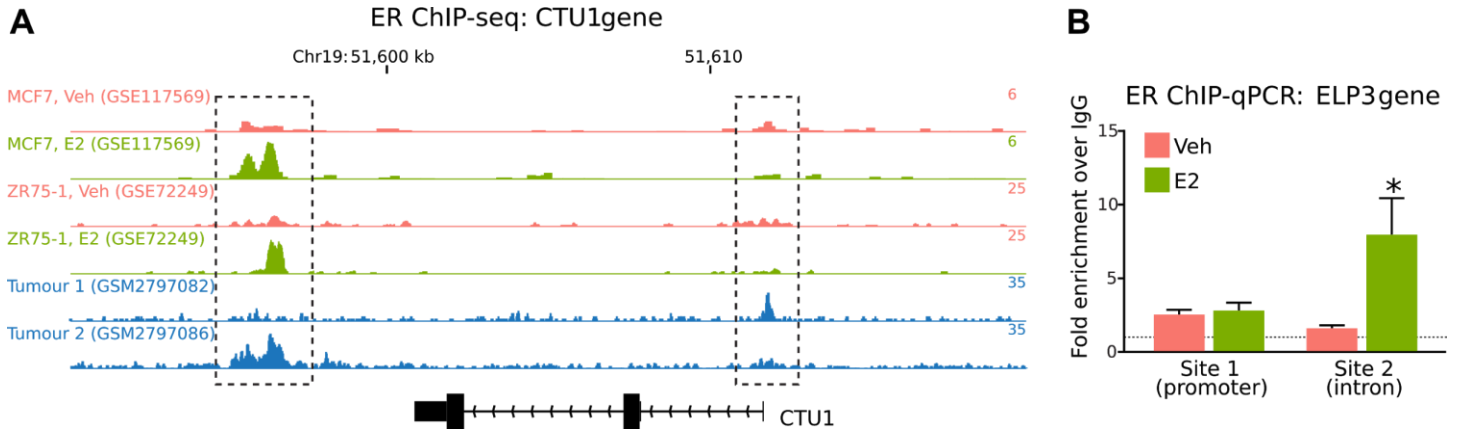
Appendix Figure S3. Quality control for small RNA-seq data on BM67 shER α and control cells (n = 3). (A) Average (over the 6 samples) number of reads after each preprocessing step. (B) Triplicate samples (1-3) from the small RNA-seq data were projected on the 2 first components of a centered principal component analysis. miRNAs with less than 30 counts on average have been excluded and the data were normalized using DESeq2 size factors.

Appendix Figure S4



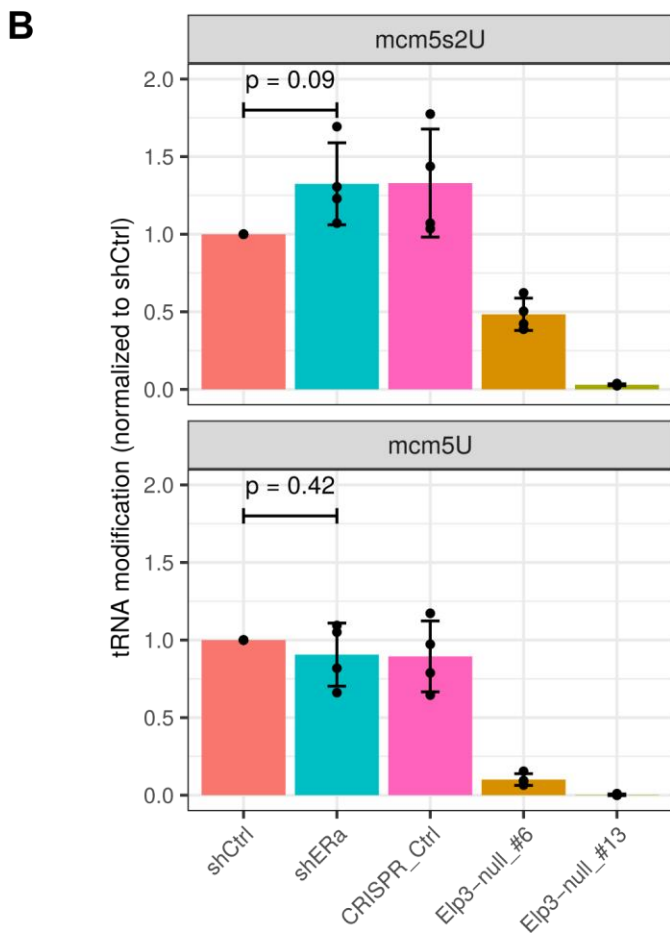
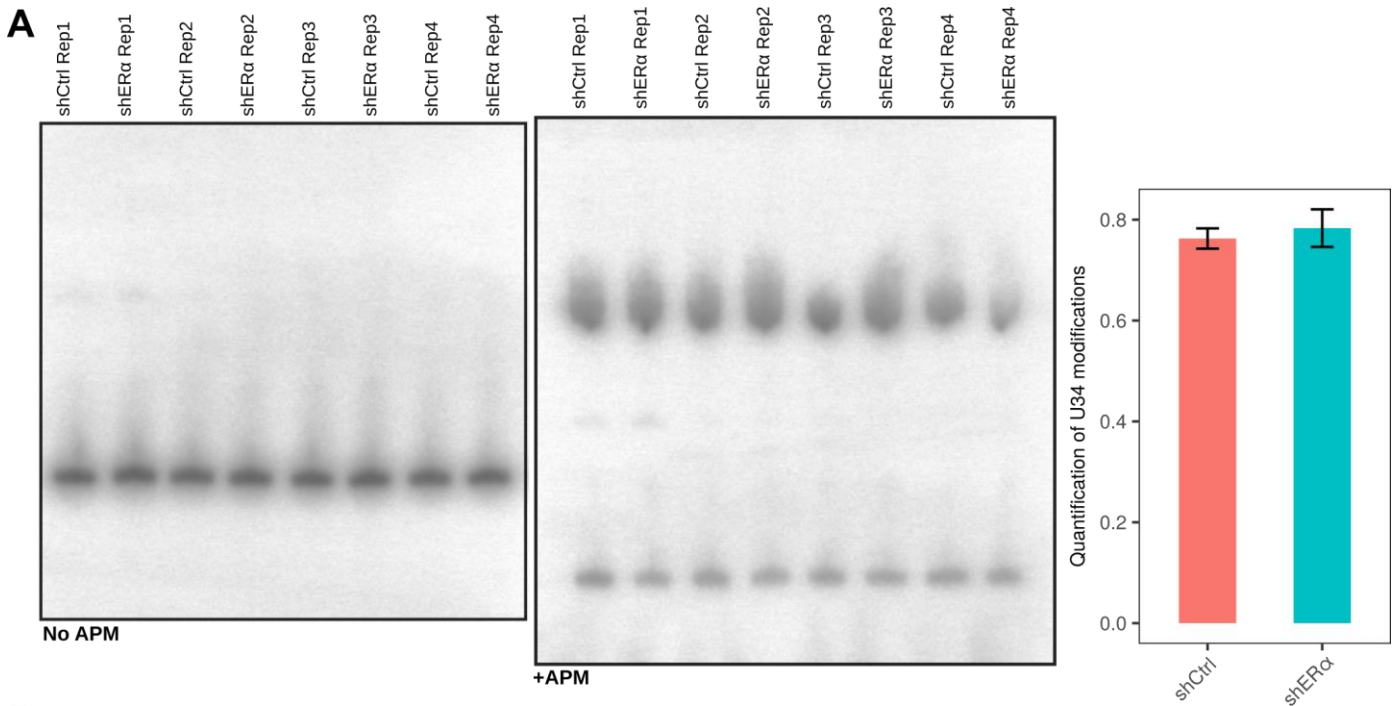
Appendix Figure S4. Quality control for quantification of tRNA levels from small RNA sequencing. (A) DESeq2 normalized count of all mature tRNAs were summed per anticodon and the average across the 6 samples is plotted against number of tRNA loci (retrieved from UCSC table browser). Quantitative data could be retrieved for most tRNAs (counts > 100), (B) tRNA coverage as measured by the ARM-Seq pipeline. Expected hard-stop modifications can be observed at positions 34 and 58.

Appendix Figure S5



Appendix Figure S5. (A) ER binding sites proximal to the CTU1 gene. Each track represents a distinct ChIP-seq dataset. **(B)** Validation of putative ER binding sites (1 and 2, shown below the ChIP-seq tracks in Fig 7C) by ChIP-qPCR in ZR-75-1 cells treated for 4 hours with Vehicle (Veh) or Estradiol (E2). Error bars = \pm SEM. *, $p < 0.05$ (t test, E2 versus Veh).

Appendix Figure S6



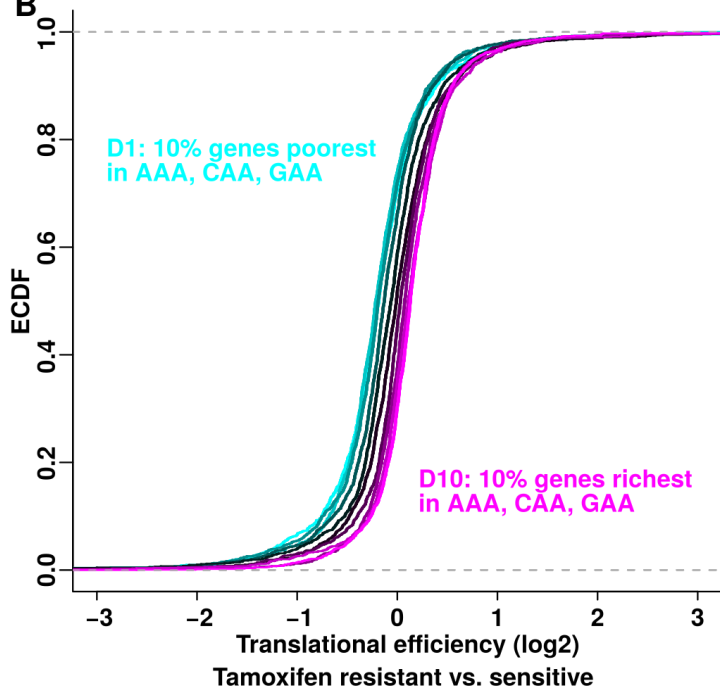
Appendix Figure S6. Quantification of tRNA modifications. (A) APM-PAGE and subsequent northern blot analysis of tE(UUC) isolated from 4 replicates of BM67 shERα and control cells. Right panel shows quantification of the northern blot (mean ± sd of relative amount of modified tRNA as compared to total tRNA; Student t-test p = 0.37). **(B)** Quantification of mcm⁵s²-U and mcm⁵-U levels in ERα knockdown and Elp3-null BM67 cells by liquid chromatography-mass spectrometry (LC-MS) analysis. Graphs represent mean ± SD of n=4. Data were analysed by Student t test.

Appendix Figure S7

A

	log2 fold change	p-value	FDR-adjusted p-value
ELP3	-0.06	0.74	0.87
ALKBH8	0.58	0.02	0.16
CTU1	0.08	0.69	0.84
CTU2	0.42	0.09	0.30

B



Appendix Figure S7. Translational regulation via the U34-modification pathway may promote resistance to tamoxifen in breast cancer cells. (A) Analysis of ELP3, ALKBH8, CTU1/2 expression between Tamoxifen resistant and Tamoxifen sensitive breast cancer cells (data from (Geter *et al*, 2017; Data ref: Geter *et al*, 2017), see Supplementary Methods). The analysis is performed on polysome-associated mRNA. **(B)** Human genes were categorized into deciles of AAA, CAA, GAA codon frequencies in their coding sequences. Cumulative distributions for changes in translational efficiency are plotted for each decile (Tamoxifen resistant vs. tamoxifen sensitive). Statistics are given for the comparison of each decile to the first decile (D1).

signature	Wilcox_pval	q25	q50	q75
D2 vs. D1	3.3e-01	0.016	-0.027	-0.037
D3 vs. D1	6.8e-01	0.034	-0.0042	-0.013
D4 vs. D1	4.6e-05	0.11	0.057	0.031
D5 vs. D1	1.8e-16	0.16	0.12	0.095
D6 vs. D1	4.9e-33	0.21	0.18	0.17
D7 vs. D1	3.7e-52	0.29	0.22	0.2
D8 vs. D1	2.9e-72	0.32	0.26	0.24
D9 vs. D1	2.3e-87	0.35	0.3	0.28
D10 vs. D1	5.5e-98	0.39	0.32	0.26

APPENDIX TABLES

<i>Overrepresented mode of regulation</i>	<i>residuals</i>	<i>Overrepresented codon</i>	<i>anticodon</i>	<i>U34 modification</i>
Offset (mRNA Up)	73.02678253	AAA	UUU	mcm5-s2-U
Offset (mRNA Up)	63.41464656	GAA	UUC	mcm5-s2-U
Offset (mRNA Up)	36.61552086	CAA	UUG	mcm5-s2-U
Offset (mRNA Up)	42.15970308	AGA	UCU	mcm5-U
Offset (mRNA Up)	29.64877454	UCA	UGA	ncm5-U
Offset (mRNA Up)	44.92368964	UUA	UAA	ncm5-Um
Offset (mRNA Up)	31.25744675	AUA	UAU	pseudouridine
Offset (mRNA Up)	39.04013545	AAU	AUU	
Offset (mRNA Up)	33.14535926	CAU	AUG	
Offset (mRNA Down)	37.80295594	CUG	CAG	
Offset (mRNA Down)	29.85645598	GGC	GCC	
Offset (mRNA Down)	29.6877003	GCC	GGC	

Appendix Table S1. Codons identified as enriched in translationally offset mRNAs. The residuals of a test of independence between codon count and mode for regulation of gene expression are given together with the most overrepresented mode of regulation for the first quintile positive residuals. Most of these codons are read by U34 modified tRNAs. The corresponding U34 modification is provided in this table.

Gene Name	Species	Sequence Fwd (5' --> 3')	Sequence Reverse (5' --> 3')
DCXR	mouse	ACTGTGCTGGCGTTGAAGG	CGGGTCCCACATTGCTTAGG
DEK	mouse	GGGCACAGTGTCTCGTTG	CGCCTGACCTCTCTAAATCAAG
	human	AACTGCTTTACAACAGGCCAG	ATGGTTTGCCAGAAGGCTTTG
CHK1	mouse	GTTAAGCCACGAGAATGTAGTGA	GATACTGGATATGGCCTTCCCT
	human	CCAGATGCTCAGAGATTCTTCCA	TGTTCAACAAACGCTCACGATTA
Gene Name	Species	Sequence Fwd (5' --> 3')	Sequence Reverse (5' --> 3')
controls			
beta actin	mouse	GGCTGTATTCCCCTCCATCG	CCAGTTGGTAACAATGCCATGT
	human	gATgAgATTggCATggCTTT	CACCTTCACCgTTCCAgTTT
GAPDH	mouse	AGGTCGGTGTGAACGGATTTG	TGTAGACCATGTAGTTGAGGTC A
NONO	mouse	GCCAGAATGAAGGCTTGA	TATCAGGGGGAAGATTGCCCA

Appendix Table S2. Sequences of qPCR primers used in this study

APPENDIX SUPPLEMENTARY METHODS

Analysis of polysome-profiling data

To verify whether the reproducibility of data from polysome-associated mRNA and cytosolic mRNA samples was comparable, we visualized 2 by 2 Pearson correlation coefficients in a clustered heatmap. The clustering was done using default parameters of the `annHeatmap2` function (Heatplus 2.26.0 R/Bioconductor package (Ploner, 2015)). We also performed a centered Principal Component Analysis after variance filtering (exclusion of the first quartile).

Anota2seq (Oertlin *et al*, 2019) classifies transcripts into three modes for regulation of gene expression: changes in mRNA abundance (i.e. similar changes in cytosolic and polysome-associated mRNA), changes in translational efficiency leading to altered protein levels, and translational buffering (i.e. offsetting) of changes in cytosolic mRNA levels. The following significance thresholds were used in the `anota2seqSelSigGenes` function: $\text{minEff} = \log_2(1.5)$; $\text{maxPAdj} = 0.1$ (i.e. $\text{FDR} < 0.1$ [for analysis of changes in translational efficiency leading to altered protein levels, no genes passed this threshold. To allow all modes of regulation to be evaluated in downstream analyses, this threshold was relaxed and $\text{maxP} = 0.05$ i.e. $\text{p-value} < 0.05$ was used instead]), $\text{selDeltaPT} = \text{selDeltaTP} = \text{selDeltaP} = \text{selDeltaT} = \log_2(1.5)$; $\text{maxSlopeTranslation} = 1.5$; $\text{minSlopeTranslation} = -1$; $\text{maxSlopeBuffering} = -1.5$; $\text{minSlopeBuffering} = 1$. For genes belonging to more than one regulatory mode, a priority order was established such that mRNAs identified as changing their translational efficiency leading to altered protein levels will belong to the translation group and no other group; mRNAs that change their mRNA abundance but are not in the translation group will be in the mRNA abundance group; and buffered (offset) mRNAs are not in the former two groups (implemented in `anota2seqRegModes`). Genes that were not allocated to any regulatory pattern were called “non regulated”. The different regulatory modes are visualized in scatter plots of polysome-associated mRNA \log_2 Fold Change (shER α vs. shCtrl) vs. cytosolic mRNA \log_2 Fold Change (shER α vs. shCtrl) using confidence ellipses as implemented in the `ggplot2` R package (`stat_ellipse` function) (Wickham, 2009). The smooth scatterplot in Fig 1C was performed using the `stat_density2d` function in `ggplot2`. The plotting area was divided in 16384 squares (128^2) of equal size and the density was computed using a two-dimensional kernel density estimation (`kde2d`; (Venables & Ripley, 2002)). The higher the number of genes there are in each square, the darker the color is on a scale from white to dark blue. Because the estimation is based on smooth distributions, the color of a square also depends on the number of genes in its neighboring squares.

Cycloheximide Chase Assay

BM67 shCtrl or shER α cells (600 k) were seeded in a 10 cm plate (RPMI + 10% FBS, P/S/G) 24 hours prior to treatment with cycloheximide (100 $\mu\text{g/ml}$). Cells were placed on ice for 10 minutes prior to lysis at the times indicated. Lysates (20 μg) were run and blotted for the indicated antibodies.

Nanostring gene expression quantification and analysis

Nanostring was performed as described previously (William *et al*, 2018) on 3 replicates of cytosolic and polysome-associated mRNA (isolated as described above) from both shER α and control BM67 cells using a custom panel of 145 genes. Position of samples on the lanes of 2 different cartridges were randomized.

The resulting Nanostring counts were \log_2 transformed and normalized using the geometric mean of reference genes as normalization factors. *Actb*, *Emc10* and *Mcm5p* were used as reference genes. Genes for which less than 3 samples were quantified above a background level set at 6 (\log_2 scale) were excluded. This resulted in a total of 75 mRNAs belonging to

one of the modes of regulation (from Fig 1E) and 11 pre-selected negative controls. Anot2seq analysis was performed on the normalized Nanostring data with the same parameters and thresholds as for the DNA-microarray analysis with one modification: the cartridge identifier was added as a batch effect in the models.

Preprocessing and analysis of RNAseq data

RNAseq reads were preprocessed using the nf-core/rnaseq v1.3 pipeline available at <https://github.com/nf-core/rnaseq> (preprint: Ewels *et al*, 2019). HISAT2 was selected as alignment tool and all other parameters were kept as default.

Genes with multiple loci were excluded from the analysis. Counts from protein-coding genes which had at least one read in each sample were scaled using TMM normalized library sizes (Robinson & Oshlack, 2010). Log2 counts per million were then computed using the voom function of the limma R package (Ritchie *et al*, 2015) as implemented in anota2seq. Similarly as for DNA-microarrays and Nanostring data, anota2seq analysis was performed on the normalized counts. The same significance thresholds were considered except for the FDR threshold for analysis of changes in translational efficiency leading to altered protein levels which was not relaxed but kept at FDR < 0.01, as for the other two modes for regulation of gene expression.

GO enrichment analysis

Gene to GO term associations (retrieved from org.Mm.eg.db_version 3.6.0 (Carlson, 2018)) based only on inference from electronic annotation (IEA) were excluded. GO terms having 5 to 1000 genes were included in the analysis. Genes were ranked based on their signed (by direction of regulation [i.e. up/down]) \log_{10} FDR-adjusted p-values. We used Wilcoxon-Mann-Whitney tests as enrichment tests in both directions separately (i.e. up or down regulation) within the Generally Applicable Gene-set Enrichment Analysis (GAGE) R/Bioconductor package (Luo *et al*, 2009). GO term enrichment results of offset (mRNA down) and cytosolic Down were visualized using the Enrichment Map plug-in in Cytoscape version 3.6.0 (Shannon *et al*, 2003; Merico *et al*, 2010) using the generic analysis and an FDR threshold of 0.15. All other parameters were kept as default. In this network visualization, each node is a GO term connected to other GO terms whose genes overlap. The size of each node reflects the number of genes associated to its corresponding term and the width of the edge illustrates the size of the overlap between two connected nodes. Groups of GO terms were clustered (black circles on Fig EV2) using the AutoAnnotate plug-in used with default settings (Kucera *et al*, 2016).

nanoCAGE library preparation and sequencing

shER α BM67 cells (3 replicates) were grown until 80% confluency before RNA was extracted with TRI-reagent (Sigma-Aldrich, Stockholm, Sweden) and further purified using the RNeasy MinElute Cleanup kit (Qiagen, Sollentuna, Sweden). RNA quantity and quality were determined by Qubit (Life Technologies, Stockholm, Sweden) and Agilent Bioanalyzer (Agilent Technologies, Kista, Sweden), respectively. The integrity number (RIN) of all samples was greater than 8. Reverse transcription was carried out by mixing 1 μ l of cytosolic mRNA (75ng) with 1 μ l 0.66 M D-Trehalose, 3.3 M of D-sorbitol, 10 μ M MS-RanN6 (Sigma-Aldrich; oligo sequences listed below) and an equimolar mixture of two Template Switching Oligonucleotides (TSO; Integrated DNA Technologies, Leuven, Belgium) with different barcodes (ACAGAT and GTATGA). As described previously (Poulain *et al*, 2017), unique molecular identifiers (UMIs) consisting of 8 random nucleotides were included in the sequence of the TSO in order to distinguish multiples copies of a transcript from PCR duplicates. After heat-denaturation for 5 min at 65 $^{\circ}$ C, following components were added to each sample until a total volume of 10 μ l: 1x first-strand buffer (Life Technologies, Stockholm, Sweden), 10 mM DTT (Life Technologies, Stockholm, Sweden), 1 M betaine (Sigma-Aldrich, Stockholm, Sweden), 1 mM dNTP mix (TaKaRa, Malmo, Sweden) and 100 units of SuperScript III (Life

Technologies, Stockholm, Sweden). Samples were incubated at 25° C for 5 min, 42° C for 60 min, 72° C for 15 min and chilled on ice. Samples were cleaned of primers and enzymes by using AMPure XP magnetic beads (Beckman Coulter, Sweden) according to the manufacturer's instructions and cDNA was eluted in 30 µl nuclease free water.

A diagnostic qPCR to determine the adequate number of temperature cycles was performed prior to the semi-suppressive PCR (ssPCR) needed for cDNA amplification. For this diagnostic qPCR, 8.5 µl of Takara SYBR Premix Ex Taq (1x, TaKaRa, Malmo, Sweden) containing 100 nM of MS-dir 1F and MS-dir 1R each was mixed with 1.5 µl cDNA (in triplicate) and amplified by qPCR (95° C for 5 min, [65° C for 15 s and 68° C for 2 min] x 40 cycles; CFX96 Touch, Bio-Rad, Sweden). For the ssPCR, 20 µl of the purified first-strand cDNA was mixed with 0.5 µl of MS-dir 1F and MS-dir 1R each (100 nM), 4 µl of water and 25 µl of Kapa HiFi HotStart Ready Mix (2x, Kapa Biosystems). [N] cDNA PCR cycles were performed for each sample, where N was based on the Cycle threshold (Ct) value obtained from the diagnostic qPCR. After ssPCR, the samples were cleaned by adding 90 µl AMPure XP beads and washed three times according to the manufacturer's protocol. cDNA PCR products were eluted in 25 µl of water. The size distribution of the cDNA PCR products was checked with the Agilent BioAnalyzer and diluted to 70 pg/µl. For the subsequent tagmentation (slightly modified from (Poulain *et al*, 2017)), which simultaneously fragments and tags input DNA with adapter sequences, 1 µl of sample was mixed with 5 µl Tagment DNA Buffer (Illumina, US), 1 µl PEG, 2.5 µl Amplicon Tagment mix (Illumina, US) and 0.5 µl of PCR grade water. Samples were incubated at 55° C for 10 min and chilled to 10° C. After reaching 10° C, 2.5 µl of NT buffer was added to strip the enzyme of the cDNA and the sample was incubated at room temperature for 5 min. Next, the fragments were amplified using a limited-cycle PCR. Additionally, this PCR also added index sequences to the 3' end of the DNA fragments, thereby enabling the sequencing of pooled libraries on an Illumina Sequencing System. For this, 1 µl nanoCAGE custom S-series primer (10 µM) and 1 µl of a Nextera XT N-Series Index primer (N7xx, 10 µM), 7.5 µl Nextera PCR Mastermix and 3 µl water were added to each sample and amplified by PCR (72° C for 3 min, 95° C for 3 min, [95° C for 10 s, 55° C for 30 s, and 72° C for 1 min] x 15 cycles, 72° C for 5 min; CFX96 Touch, Bio-Rad, Sweden). The amplified libraries were cleaned with 22.5 µl AMPure XP magnetic beads (Beckman Coulter, Sweden) according to the manufacturer's instructions and cDNA was eluted in 20 µl nuclease free water. cDNA quantity was determined by Qubit and final library size distribution was checked with the High-sensitivity DNA kit on an Agilent Bioanalyzer. Prior to sequencing, library concentrations were adjusted to 10 nM and pooled together. Sequencing was done on HiSeq2500 (Illumina, US) in Rapid Run mode, producing 100bp single end reads.

MS-RanN6: 5'-TAGTCGAACTGAAGGTCTCCGAACCGCTCTTCCGATCTNNNNNN-3'

Template Switching Oligo: 5'-TAGTCGAACTGAAGGTCTCCAGCA[barcode][UMI]TATA(rG)(rG)(rG)-3'

MS-dir1F: 5'-TAGTCGAACTGAAGGTCTCCAGC-3'

MS-dir1R: 5'-TGACGTCGTCTAGTCGAACTGAAGGTCTCCGAACC-3'

nanoCAGE Custom S-series: 5'-AATGATACGGCGACCACCGAGATCTACACTAGTCGAACTGAAGG-3'

Nextera XT N-Series: 5'-CAAGCAGAAGACGGCATACGAGAT[index]GTCTCGTGGGCTCGG-3'

nanoCAGE data processing and analysis

nanoCAGE reads were extracted using TagDust (Lassmann, 2015) (version 2.31) and those mapping to ribosomal RNA were removed. Reads were trimmed to remove Nextera XT N-Series 3' adaptors using cutadapt version 1.9.1 (Martin, 2012) with the maximum read-adaptor mapping error rate set to 0.15, the minimum overlap length set to 1 and trimming of up to 4 adaptors was allowed. Remaining reads shorter than 25 bases were excluded. Duplicated reads with identical UMI and

sequence (in the first 25 bp) were assumed to be PCR duplicates and only one was kept for further analysis. A first alignment to the mouse genome mm10 was performed using Bowtie (version 0.12.7) (Langmead *et al*, 2009) allowing no more than 2 mismatches in the first 28 bp on the 5' end of the read. Reads which aligned uniquely (selected with parameters -m 1 – best –strata) and reads which did not align underwent a second alignment to a custom index of 5' UTR sequences. For this index, refSeq 5'UTR sequences (O'Leary *et al*, 2016) were extended with 78 bases upstream genomic and downstream mRNA sequences (78 is the maximum length of nanoCAGE reads after trimming of all adaptors). All best strata alignments with a maximum of 3 mismatches in the first 25 bases of the read were reported. One or two bases were trimmed from the 5' end of each read when alignment mismatches were present at the first or second position, respectively. Strand invasion artefacts were removed using the perl script provided in Tang *et al*. (Tang *et al*, 2013) with minimum edit distance set to 2.

Transcription start site peaks (defined as 5' end mapping positions of reads) observed in less than 4 out of 6 samples (3 replicates and 2 barcodes per replicate) were excluded. Reads from all replicates were then merged. RefSeqs with less than 10 reads were excluded. The complexity was verified by sampling increasing numbers of reads (delta of 150 000) and counting the number of peaks (>2 reads) and number of RefSeqs (>20 reads) detected. An average out of 3 iterations was used. For each gene entrez identifier, the RefSeq(s) with most reads was (were) selected (for multiple RefSeqs with the same maximum counts, an average was considered for each 5' UTR characteristics). The 5' UTR length was computed as a mean length along the RefSeq weighted by the read count of each observed peak. We then extracted the 5'UTR sequence corresponding to the most prevalent TSS position and computed GC content, fold energy and identified potential uORF with strong Kozak context (a start codon and in-frame stop codon within the 5' UTR; the start codon being in a [A,G]..ATGG context). Fold energies were obtained using mfold version 3.6 with default parameters (Zuker, 2003). We considered in the analysis herein the minimum folding energy after re-evaluation by the efn algorithm (see mfold manual). Twenty-four UTRs failed to be in silico folded by mfold (these sequences were amongst the shortest ones) and positive fold energy results (n = 24) were excluded.

The weighted length, GC content and fold energy distribution of 5'UTR were compared between mRNAs regulated via different modes and visualized as boxplots. Wilcoxon-Mann-Whitney tests were used to assess differences between translationally offset and non-offset (mRNA abundance) mRNAs. Presence of strong uORFs is shown as a bar chart and differences tested using Fisher's exact tests. A global Bonferroni adjustment was applied on p-values.

Preprocessing and analysis of RNAseq of small RNAs data

The 3' adapter TGG AATTCTCGGGTGCCAAGG was trimmed when present using fastx_clipper (part of FASTX Toolkit version 0.0.14). Reads which were less than 17nt-long were excluded. Reads longer than 23 nt were also filtered out. Remaining reads were aligned to the mm10 mouse genome using Bowtie version 1.0.1 (Langmead *et al*, 2009) allowing one mismatch anywhere in the sequence. All alignments falling in the best alignment stratum were reported (parameters -v 1 –best –strata -a). miRNA levels were quantified using the countOverlaps function of the GenomicAlignments R/Bioconductor package (Lawrence *et al*, 2013) and the miRBase definition of miRNA coordinates Release 22 (Kozomara & Griffiths-Jones, 2014). For miRNAs mapping to multiple genomic locations, the quantification with minimum total count was considered. Only mature miRNA counts were kept in the analysis.

miRNAs with average count below 30 were excluded. Differential expression analysis upon ER α depletion was performed using DESeq2 (Love *et al*, 2014) on raw miRNA counts without independent filtering. The replicate identifier was included as a covariate in the models. miRNAs were considered significantly differentially expressed for FDR<0.2 and abs(log₂ Fold Change) > log₂(1.25).

Datasets of targets of conserved mouse miRNAs were downloaded from TargetScanMouse release 7.1, January 2016 (Lewis *et al*, 2005) from tables “miR Family” and “Predicted Conserved Targets”. Only highly conserved (code 2) and conserved (code 1) miRNA families were selected. Each mouse miRNA was then mapped to mRNAs having conserved target site(s) for this miRNA. Next, we tested whether downregulated miRNAs would have target sites overrepresentation among upregulated but translationally offset vs. upregulated non-offset mRNAs using Fisher exact tests. A Bonferroni correction was applied on Fisher exact p-values. To visualize this overrepresentation, each upregulated mRNA upon ERα depletion (mRNA abundance Up or Offset (mRNA Up)) was assigned an angle given by arc-tangent (polysome-associated mRNA log2 Fold Change / cytosolic mRNA log2 Fold Change). This corresponds to the angle between the horizontal reference line and the position of this mRNA in Fig 1E (see Fig 4E, blue arrow). For each miRNA, the distribution of its targets’ angles is plotted together with the background (mRNAs non targets of any downregulated miRNAs). We proceeded with the same approach for upregulated miRNAs and downregulated mRNAs. We also compared the distributions of angles of mRNA targets of any upregulated miRNAs (10 miRNAs were upregulated) with mRNA targets of groups of 10 randomly selected miRNAs (we generated 30 such groups of 10).

Analysis of codon usage

For all mRNAs categorized under the different modes for regulation of gene expression, the longest coding sequence was retrieved from the consensus coding sequence project database release 21 for NCBI mouse annotation release 106 and Ensembl release 86 (2016-12-08, (Pruitt *et al*, 2009)). Once the average number of each codon across mRNAs under each mode for regulation of gene expression was computed; methionine, tryptophan and stop codons were removed and a Chi-Square test of independence was performed. The standardized residuals from this test were visualized using a heatmap. Rows and columns of the heatmap were ordered according to unsupervised hierarchical clustering using the 1 - Pearson correlation distance and average linkage. All p-values were adjusted for multiple testing using the Bonferroni correction and the significance threshold for being colored in the heatmap was 0.01. The top 12 codons with highest positive residuals (i.e. most over-represented codons in a mode of regulation) were selected for further analysis. This resulted in 9 codons over-represented among Offset (mRNA Up) and 3 codons over-represented among offset (mRNA Down). Five measures of codon bias were performed at the mRNA level using the codonW software version 1.4.4: the Codon Adaptation Index (Sharp & Li, 1987), the Frequency of Optimal codons (Ikemura, 1981), the Codon Bias Index (Bennetzen & Hall, 1982), the effective number of codons (Wright, 1990) and the G+C content at the 3rd position of synonymous codons. The tAI was calculated using the tAI R package (dos Reis *et al*, 2003, 2004). The tAI measures the adaptation of the codon composition of the tRNA pool. Relative tRNA levels are estimated from copy numbers of each tRNA in the genome. tRNA genes from GRCm38/mm10 (Dec. 2011) were retrieved using the UCSC table browser and used as input (Karolchik *et al*, 2004). The tAI was also computed on coding sequences divided into sextiles. The distribution of these codon measures was compared between the different modes of regulation.

Codon composition of GO terms (Biological processes only) was plotted in the two first dimensions of a correspondence analysis using the FactoMineR R package (Lê *et al*, 2008). GO terms with less than 40 genes or more than 1000 genes were excluded and the R/Bioconductor package org.Mm.eg.db version 3.6.0 (Carlson, 2018) was queried to get go_id to entrez id associations (genes with IEA as only association were excluded as described above). Codon composition of each mode for regulation of gene expression was projected on this correspondence analysis. The FactoMineR package was also used to perform a gene-level correspondence analysis of codon composition of all regulated genes (from Fig 1E).

When analyzing codon usage normalized by amino acid (each codon count was divided by its corresponding amino acid count), the data was scaled to account for the fact that amino acids are encoded by different numbers of synonymous codons.

In order to analyze which codons showed general co-usage in the same mRNAs, we used the Spearman rank correlation of each pair of codons within genes in all modes of regulation.

Quantification and analysis of tRNA levels

tRNA quantification was performed using the RNAseq of small RNAs using the ARM-seq bioinformatics pipeline from Cozen *et al.* (Cozen *et al.*, 2015) with few modifications. Briefly, the tRNAscan-SE algorithm was used to predict the presence of tRNA loci in the mm10 genome. These tRNA loci were “in-silico processed” (introns are removed, CCA is added, 5-prime G base is added to Histidine tRNAs, N-bases are added to the ends of each mature sequence) using the maketrnadb script. All tRNA loci and mature tRNAs were added to the mm10 genome and a Bowtie2 index was created for alignment of the RNAseq reads of small RNAs after adaptor trimming (same trimming method as for the miRNA pre-processing). We used the default Bowtie2 alignment strategy (i.e. search for multiple alignments, report the best one, with upper limits for parameters $-D$ and $-R$ set at 100 (Langmead & Salzberg, 2012)). The quantification was performed using the ARM-seq countreads.py script. All reads mapping tRNAs were retained, regardless of where in the tRNA they mapped. DEseq2 was used for read count normalization (quantification of non coding RNA features retrieved from RNAcentral Release 10 (The RNAcentral Consortium, 2017) was provided to estimate size factors) and analysis of differential expression (Love *et al.*, 2014). Mature tRNA raw counts were summed for each anticodon and the differential expression analysis was performed on all non coding RNA with a minimum average count of 30. tRNA coverage was computed using the ARM-seq pipeline script.

Analysis of public dataset for E2 dependent expression of ELP3, ALKBH8 and CTU2

CEL files were downloaded from the GSE35428 GEO repository (Wardell *et al.*, 2012; Data Ref: McDonnell *et al.*, 2012). Gene expression was normalized using Robust Multiarray Averaging and annotated with the custom probeset definition HGU133Plus2_Hs_ENTREZG (<http://mbni.org/customcdf/21.0.0/entrezg.download/>). Gene expression of *ELP3*, *ALKBH8* and *CTU2* was presented as boxplots and compared between conditions with Ethanol, E2, ICI-182,780 and E2+ICI-182,780 (10 replicates per condition). Wilcoxon-Mann-Whitney tests were performed for all 3 genes and a global Bonferroni adjustment was applied on p-values.

Western blot

Samples were collected from cells grown to 70% confluence and lysed in RIPA buffer (Sigma) containing 5 mM NaF, 1 mM Na₃VO₄ and 1X complete protease inhibitor cocktail (Roche). Cell lysates were quantitated using a detergent compatible protein assay kit (BioRad). Cell extracts were denatured and separated by SDS-PAGE. Blots were incubated with the following antibodies overnight at 4°C, on a platform shaker; Estrogen Receptor alpha (ER α , 1:2000, [6F11] Leica), DEK (1:1000 [5E3A6], Protein Tech), DEK (1:1000, [E1L3V], Cell signaling), ELP3 (1:1000, [D5H12], Cell signaling), MYC (1:1000, [Y69], abcam), β -Actin (1:100,000, [AC15], Abcam), JAG1 (1:2000, D4Y1R, Cell signaling), CHK1 (1:1000, #2360, Cell signaling), DCXR (1:1000, [15188-1-AP], Protein Tech), Androgen Receptor (AR, 1:10,000, [D6F11], Cell Signaling), GAPDH (1:100,000, [6C5], abcam), α -Tubulin (1:200,000, [DM1A], ThermoFisher), MYC-Tag (1:2000, [9B11], Cell signaling), HA-Tag (1:10,000, [12CA5], Sigma). Blots were then incubated for 1 hour with 5% skim milk and anti-rabbit or anti-mouse, horse radish peroxidase (HRP) conjugated secondary antibody HRP (Dako) amplification. Blots were incubated

with Immobilon Western Chemiluminescent HRP Substrate (Merck) and Chemiluminescence was detected with amersham hyperfilm ECL (GE Healthcare) film. Antibodies have been validated by knockdown and/or overexpression of target proteins (except for commercially well validated antibodies e.g. β -Actin, GAPDH).

Analysis of codon usage in a public dataset of Tamoxifen-sensitive and resistant cell lines

Pre-processed RNAseq reads were downloaded from the GEO repository (Geter *et al*, 2017; Data ref: Geter *et al*, 2017). We only considered data from heavy polysome mRNA (mRNA associated with > 4 ribosomes in this study) and total mRNA. Filtering of genes with zero counts, normalization (default settings) and analysis of changes in translational efficiency leading to altered protein levels was performed with anota2seq (Oertlin *et al*, 2019). The models were fitted on data from all samples (2 replicates of 3 conditions: tamoxifen sensitive, tamoxifen resistant, tamoxifen resistant with eIF4E silencing) and only the “Tamoxifen-resistant vs sensitive” contrast was considered for codon usage analysis. The frequency of AAA, CAA and GAA codons were calculated for the coding sequences of all human genes. These genes were categorized into deciles of such frequencies. Empirical Cumulative Distribution Function of the changes in translational efficiency between Tamoxifen resistant and sensitive cells was plotted for each decile. Wilcoxon-Mann-Whitney tests were performed to compare each AAA, CAA, GAA codon frequency decile to the first decile.

REFERENCES

- Bennetzen JL & Hall BD (1982) Codon selection in yeast. *J. Biol. Chem.* **257**: 3026–3031
- Carlson M (2018) Org.Mm.eg.db: Genome Wide Annotation for Mouse. R package version 3.6.0. Available at: <https://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html> [Accessed June 15, 2018]
- Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM & Lowe TM (2015) ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **12**: 879–884
- Ewels PA, Peltzer A, Fillinger S, Alneberg J, Patel H, Wilm A, Garcia MU, Tommaso P Di & Nahnsen S (2019) nf-core: Community curated bioinformatics pipelines. bioRxiv: 610741
- Geter PA, Ernlund AW, Bakogianni S, Alard A, Arju R, Giashuddin S, Gadi A, Bromberg J & Schneider RJ (2017) Hyperactive mTOR and MNK1 phosphorylation of eIF4E confer tamoxifen resistance and estrogen independence through selective mRNA translation reprogramming. *Genes Dev.* **31**: 2235–2249
- Geter P, Schneider RJ, Ernlund A (2017) Gene Expression Omnibus GSE107590 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107590>) [DATASET]
- Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol.* **151**: 389–409
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D & Kent WJ (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**: D493–496
- Kozomara A & Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**: D68–73
- Kucera M, Isserlin R, Arkhangorodsky A & Bader GD (2016) AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research* **5**: 1717
- Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359
- Langmead B, Trapnell C, Pop M & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25
- Lassmann T (2015) TagDust2: a generic method to extract reads from sequencing data. *BMC Bioinformatics* **16**: 24
- Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT & Carey VJ (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**: e1003118
- Lê S, Josse J & Husson F (2008) FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **25**: 1–18
- Lewis BP, Burge CB & Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20
- Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**: 550

- Luo W, Friedman MS, Shedden K, Hankenson KD & Woolf PJ (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**: 161
- McDonnell DP, Wardell SE, Kazmin DA (2012) Gene Expression Omnibus GSE35428 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35428> [DATASET]
- Martin M (2012) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Bioinforma. Action* **17**: 10–12
- Merico D, Isserlin R, Stueker O, Emili A & Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**: e13984
- O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**: D733--745
- Oertlin C, Lorent J, Murie C, Furic L, Topisirovic I & Larsson O (2019) Generally applicable transcriptome-wide analysis of translation using anota2seq. *Nucleic Acids Res.* **47**: e70–e70
- Ploner A (2015) Heatplus: Heatmaps with row and/or column covariates and colored clusters. <https://github.com/alexploner/Heatplus>.
- Poulain S, Kato S, Arnaud O, Morlighem J-É, Suzuki M, Plessy C & Harbers M (2017) NanoCAGE: A Method for the Analysis of Coding and Noncoding 5'-Capped Transcriptomes. In pp 57–109. Humana Press, New York, NY
- Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ, Hart E, Suner M-M, Landrum MJ, Aken B, Ayling S, Baertsch R, Fernandez-Banet J, Cherry JL, Curwen V, Dicuccio M, et al (2009) The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* **19**: 1316–1323
- dos Reis M, Savva R & Wernisch L (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**: 5036–5044
- dos Reis M, Wernisch L & Savva R (2003) Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* **31**: 6976–6985
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W & Smyth GK (2015) limma powers differential expression analyses for {RNA}-sequencing and microarray studies. *Nucleic Acids Res.* **43**: e47
- Robinson MD & Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**: R25
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504
- Sharp PM & Li WH (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**: 1281–1295

- Tang DTP, Plessy C, Salimullah M, Suzuki AM, Calligaris R, Gustincich S & Carninci P (2013) Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Res.* **41**: e44
- The RNAcentral Consortium (2017) RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.* **45**: D128–D134
- Venables WN & Ripley BD (2002) *Modern Applied Statistics with S* New York, NY: Springer New York
- Wardell SE, Kazmin D & McDonnell DP (2012) Research resource: Transcriptional profiling in a cellular model of breast cancer reveals functional and mechanistic differences between clinically relevant SERM and between SERM/estrogen complexes. *Mol. Endocrinol.* **26**: 1235–1248
- Wickham H (2009) *Ggplot2 : elegant graphics for data analysis* Springer
- William M, Leroux L-P, Chaparro V, Lorent J, Graber TE, M'Boutchou M-N, Charpentier T, Fabié A, Dozois CM, Stäger S, van Kempen LC, Alain T, Larsson O & Jaramillo M (2018) eIF4E-Binding Proteins 1 and 2 Limit Macrophage Anti-Inflammatory Responses through Translational Repression of IL-10 and Cyclooxygenase-2. *J. Immunol.* **200**: 4102–4116
- Wright F (1990) The 'effective number of codons' used in a gene. *Gene* **87**: 23–29
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406–3415