Supplemental Methods

RNA-sequencing cDNA library preparation

RNA samples depleted from ribosomal RNAs were fragmented using RNA fragmentation reagent (Ambion, Cat: AM8740) for 3 minutes and 30 seconds at 70°C followed by inactivation with the provided “Stop” buffer. Fragmented RNAs were then dephosphorylated at their 3’ end using PNK (New England Biolabs, Cat: M0201) in MES buffer (100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 300 mM NaCl) at 37 °C for 3 h. RNA fragments with a 3′-OH were ligated to a preadenylated DNA adaptor. Following this, ligated RNAs were reverse transcribed with Superscript III (Invitrogen) with a barcoded reverse-transcription primer that anneals to the preadenylated adaptor. After reverse transcription, cDNAs were resolved in a denaturing gel (10% acrylamide and 8 M urea) for 1 h and 45 min at 35 W. Gel-purified cDNAs were then circularized with CircLigase I (Lucigen, Cat: CL4111K) and PCR-amplified with Illumina's paired-end primers 1.0 and 2.0. PCR amplicons (12-14 cycles for RNA-seq and 4-6 cycles for ribosome profiling) were gel-purified and submitted for sequencing on the Illumina HiSeq 2500 platform.

Ribosome profiling

Cells (15 million for each biological replicate) were incubated with cycloheximide (100 µg/mL final) for 5 min at 37°C. Cells were then washed two times in ice-cold PBS + cycloheximide (100 µg/mL) and scraped in 1 ml of PBS + cycloheximide (100 µg/mL). Cells were pelleted at 500g for 5 min at 4 °C and lysed in 1 ml of lysis buffer (10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 µg/mL cycloheximide and 1× Protease-Inhibitor Cocktail EDTA-free (Roche)). Lysate was homogenized with a P1000 pipette by gentle pipetting up and down for a total of eight strokes and incubated at 4 °C for 10 min. The lysate was centrifuged at 1,300g for 10 min

at 4 °C, the supernatant recovered and the absorbance at 260 nm measured. For the footprinting, 5 A260 units of the cleared cell lysates were incubated with 300 units of RNase T1 (Fermentas) and 500 ng of RNase A (Ambion) for 30 min at RT. After this, samples were loaded on top of a 10–50% (w/v) linear sucrose gradient (20 mM HEPES-KOH, pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 2 mM DTT and 100 µg/mL of cycloheximide) and centrifuged in a SW-40ti rotor at 35,000 r.p.m. for 2 h 40 min at 4 °C. The collected 80S fraction was complemented with SDS to 1% final and Proteinase K (200 µg/mL) and then incubated at 42 °C for 45 min. After Proteinase K treatment, RNA was extracted with one volume of phenol (pH 4.5)/chloroform/isoamyl alcohol (25:24:1). The recovered aqueous phase was supplemented with 20 µg of glycogen, 300 mM sodium acetate, pH 5.2, and 10 mM $MgCl_2$. RNA was precipitated with three volumes of 100% ethanol at −20 °C overnight. After a wash with 70% ethanol, RNA was resuspended in 5 µl of water and the 3′ ends dephosphorylated with PNK (New England BioLabs) in MES buffer (100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 300 mM NaCl) at 37 °C for 3 h. Dephosphorylated RNA footprints were then resolved on a 15% acrylamide (19:1), 8 M urea denaturing gel for 1 h 30 min at 35 W and fragments ranging from 26 nt to 32 nt size-selected from the gel. Size-selected RNAs were extracted from the gel slice by overnight nutation at RT in RNA elution buffer (300 mM NaCl, and 10 mM EDTA). The recovered aqueous phase was supplemented with 20 µg of glycogen, 300 mM sodium acetate, pH 5.2, and 10 mM $MgCl_2$. RNA was precipitated with three volumes of 100% ethanol at −20 °C overnight. After a wash with 70% ethanol, RNA was resuspended in 5 µl of water and subjected to cDNA library construction as described above.

Poly(A)-site sequencing

Poly(A) + RNA enriched by oligoT hybridization. Poly(A) + RNA samples were then fragmented to 60–80 nt via chemical hydrolysis and reverse transcribed with anchored

oligoT oligonucleotides containing forward and reverse Illumina sequencing primer sites separated by a hexa-ethyleneglycol spacer (Sp18) linker. At the 5′ end, each oligonucleotide began with 5′p-GG to promote ligation (Heyer et al. 2015), followed by 5 random nucleotides (unique molecular index, UMI) to enable PCR duplicate removal. Each primer also harbored a unique 5 nt Hamming barcode (BC), allowing for sample multiplexing. Following cDNA circularization with CircLigase I, libraries were PCR amplified (12–14 cycles) and subjected to single end 100 nt sequencing on Illumina's HiSeq platform.

Transcript database creation

The rationale to build the database was to eventually select a single transcript isoform to which quantify expression and ribosome density for further analyses and obtain transcript information such as the length of UTRs, coding-sequence, codon usage and other features used to build the random forest model. For this we decided to rely on the APPRIS annotation in order to select the principal transcript isoform for each gene, further guided by our PAS-seq data in case multiple principal transcript isoforms were described. For this, the mouse transcript database was generated using the genome assembly and sequences files gencode.vM23.annotation.gff3 and GRCm38.p6.genome.fa obtained from https://www.gencodegenes.org/ (downloaded on 06/05/2019). A SQLite database was first generated using the Python package GffUtils (https://github.com/daler/gffutils). Then a csv file was generated for each transcript with an associated coding sequence in order to collect all available components or attributes (UTRs, CDS, introns, exons, associated gene information) and their basic properties (genomic start/end points, sequence, length, GC percentage). When available, the APPRIS level of each transcript was also recovered.

All available experimental data (ribo-density, m6A-seq, PAS-seq) or bio-informatics analysis (TDD indexes, G-quadruplexes, specific codon stretches), were then assigned to

the corresponding transcripts files. When data was only associated to the gene, all corresponding transcripts files were updated. The PASeq peaks genomic positions and counts were identified in triplicates for both cell status: Resting and Activated and already associated to a specific gene. Thus the peaks and their relative counts in each experimental condition were associated to the most suitable transcripts of the corresponding gene (for example: in an exon *vs* an intron, or to the closest 3'UTR end when downstream). Depending on the number of PASeq peaks attributed to a transcript and their positions and counts, a weighted experimental 3'UTR length (corresponding to the mean 3'UTR length of the different 3'UTR isoforms taking into consideration their relative expression level) was then possible to calculate for each experimental condition.

As the RNA-seq approach attributes reads to genes, the best transcript for each experimental condition was determined. In this purpose all the transcripts of a single gene were ranked on 4 criteria: pass the initial quality control, APPRIS level, number of PASeq peaks (*i.e.* present in the experimental condition), and length.