

Thiol-linked alkylation of RNA to assess expression dynamics

Veronika A Herzog¹, Brian Reichholz¹, Tobias Neumann², Philipp Rescheneder³, Pooja Bhat¹, Thomas R Burkard¹, Wiebke Wlotzka¹, Arndt von Haeseler³, Johannes Zuber² & Stefan L Ameres¹ 

Gene expression profiling by high-throughput sequencing reveals qualitative and quantitative changes in RNA species at steady state but obscures the intracellular dynamics of RNA transcription, processing and decay. We developed thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM seq), an orthogonal-chemistry-based RNA sequencing technology that detects 4-thiouridine (s⁴U) incorporation in RNA species at single-nucleotide resolution. In combination with well-established metabolic RNA labeling protocols and coupled to standard, low-input, high-throughput RNA sequencing methods, SLAM seq enabled rapid access to RNA-polymerase-II-dependent gene expression dynamics in the context of total RNA. We validated the method in mouse embryonic stem cells by showing that the RNA-polymerase-II-dependent transcriptional output scaled with Oct4/Sox2/Nanog-defined enhancer activity, and we provide quantitative and mechanistic evidence for transcript-specific RNA turnover mediated by post-transcriptional gene regulatory pathways initiated by microRNAs and N⁶-methyladenosine. SLAM seq facilitates the dissection of fundamental mechanisms that control gene expression in an accessible, cost-effective and scalable manner.

The regulated expression of genetic information is required for cellular homeostasis and environmental adaptability and its transformation can cause human diseases¹. Underlying these fundamental biological processes are tightly regulated molecular events that control the relative kinetics of RNA transcription, processing and degradation. Understanding the molecular basis for gene regulation demands insight into the relative kinetics of RNA biogenesis and degradation in a transcript-specific and systematic manner².

Metabolic RNA labeling approaches that employ nucleotide-analogs enable tracking of RNA species over time without interfering with cellular integrity. Among these, 4-thiouridine (s⁴U) represents the most widely used nucleotide analog to study the dynamics of RNA expression because it is readily imported into metazoan cells by equilibrate nucleoside transporters³ and

provides unique physicochemical properties for thiol-specific reactivity and affinity, which enables biochemical separation by reversible biotinylation^{4–10}. Affinity-based RNA-purification following s⁴U labeling has been successfully applied to cultured cells of diverse biological and organismal origin, as well as *in vivo* in yeast and metazoan model organisms, including insects and mammals, using either s⁴U or 4-thiouracil following metabolic activation by uracil phosphoribosyltransferase (UPRT)^{4,5,9–11}. However, as with any biochemical separation method, the underlying protocols are laborious, require ample starting material and typically encounter the problem of low signal-to-noise performance, in part because of limited biotinylation efficiency⁷. Furthermore, analysis of labeled RNA species by sequencing requires extensive controls to provide integrative insights into gene expression dynamics and fails to report global effects unless spike-in strategies are applied^{8,12}. Alternative concepts for the direct identification of nucleotide analogs by sequencing have emerged from recent epitranscriptomics technologies, but current methods are incompatible with biologically inert nucleotide analogs (such as, s⁴U) and fail to report absolute stoichiometry^{13,14}.

Here, we report thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM seq), an orthogonal chemistry approach that uncovers s⁴U at single-nucleotide resolution by reverse-transcription-dependent thymine-to-cytosine (T > C) conversions in a high-throughput sequencing-compatible manner.

RESULTS

Detection of s⁴U by sequencing

In SLAM seq, we employed the primary thiol-reactive compound iodoacetamide (IAA), which covalently attaches a carboxyamidomethyl group to s⁴U by nucleophilic substitution (**Fig. 1a**). Quantitative s⁴U alkylation was confirmed by a shift in the characteristic absorbance spectrum of 4-thiouracil from ~335 to ~297 nm (**Fig. 1b**)¹⁵. Under optimal reaction conditions (**Supplementary Fig. 1**), absorbance at 335 nm decreased 50-fold compared with untreated 4-thiouracil, resulting in complete (≥98%) alkylation in 15 min (**Supplementary Fig. 1**). Mass spectrometry analysis

¹Institute of Molecular Biotechnology, Vienna Biocenter Campus, Vienna, Austria. ²Research Institute of Molecular Pathology, Vienna Biocenter Campus, Vienna, Austria. ³Center for Integrative Bioinformatics Vienna, Max F Perutz Laboratories, Medical University of Vienna, University of Vienna, Vienna Biocenter Campus, Vienna, Austria. Correspondence should be addressed to S.L.A. (stefan.ameres@imba.oeaw.ac.at).

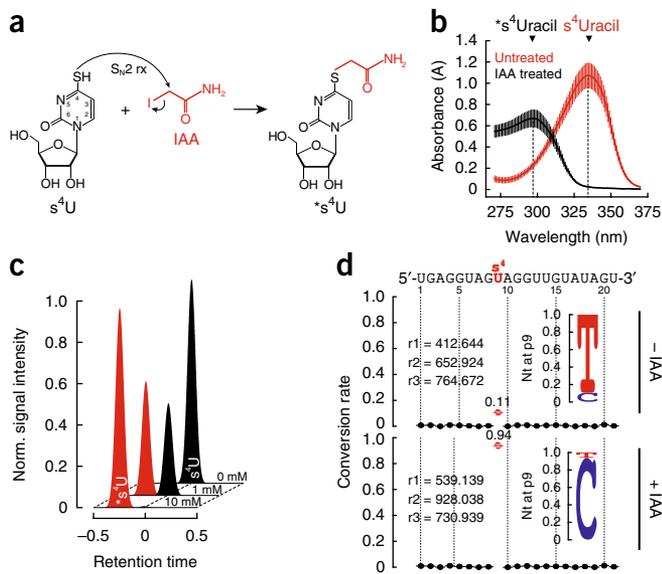


Figure 1 | Detection of s^4U by chemical derivatization and sequencing. (a) s^4U reacts with the thiol-reactive compound iodoacetamide (IAA), attaching a carboxyamidomethyl group to the thiol group in s^4U as a result of a nucleophilic substitution (S_N2) reaction. (b) Absorption spectra of 4-thiouracil ($s^4Uracil$) before and after treatment with IAA. Absorption maxima of educt ($s^4Uracil$; $\lambda_{max} \approx 335$ nm) and product (carboxyamido-methylated 4-thiouracil, $s^4Uracil$; $\lambda_{max} \approx 297$ nm) are indicated. Data represent mean (center line) \pm s.d. (whiskers) of independent experiments (untreated, $n = 13$; IAA treated, $n = 3$). (c) Normalized LC-MS extracted ion chromatograms of s^4U (black) and alkylated s^4U (red) at the indicated iodoacetamide concentrations. (d) Conversion rates for each position of a s^4U -containing RNA before or after IAA treatment. Average conversion rates (center line) \pm s.d. (whiskers) of three independent experiments (points) are shown. The number of sequenced reads in each replicate (r1–3) are indicated. Nucleotide identity at s^4U site (p9) is shown.

of thiol-specific alkylation in a ribose context confirmed these derivatization efficiencies (Fig. 1c and Supplementary Fig. 2). Because quantitative identification of s^4U by sequencing presumes that reverse transcriptase (RT) passes alkylated s^4U residues without drop-off, we determined the effect of s^4U alkylation on RT processivity in primer extension assays (Supplementary Fig. 3a). We did not observe a significant effect of s^4U alkylation on RT processivity when compared with a non- s^4U -containing oligo with identical sequence ($P > 0.05$; Supplementary Fig. 3b,c). To evaluate the effect of s^4U alkylation on RT-directed nucleotide incorporation, we isolated the full-length products of primer extension reactions, PCR amplified the cDNA and subjected the libraries to high-throughput sequencing (Fig. 1d and Supplementary Fig. 4). Although the presence of s^4U prompted a constant 10–11% T > C conversions already in the absence of alkylation (presumably as a result of base-pairing variations of s^4U tautomers), s^4U alkylation increased T > C conversions by 8.5-fold, resulting in a >0.94 conversion rate (Fig. 1d). Notably, iodoacetamide-treatment leaves conversion rates of any given non-thiol-containing nucleotide unaltered (Supplementary Fig. 4c).

SLAM-seq quantifies s^4U -labeled transcripts in mESCs

We subjected mouse embryonic stem cells (mESCs) to 100 μM s^4U labeling, a concentration far below the EC_{50} toxicity

value of s^4U in mESCs (Fig. 2a and Supplementary Fig. 5). After metabolic RNA labeling for 24 h, we prepared total RNA followed by thiol-alkylation and 3'-end mRNA sequencing (Quant-seq). Quant-seq provides rapid and quantitative access to mRNA expression profiles by generating Illumina-compatible libraries of the sequences close to the 3' end of polyadenylated RNA (Fig. 2b and Supplementary Fig. 6) (Moll, P., Ante, M., Seitz, A. & Reda, T., unpublished data). Hence, only one fragment per transcript is generated, which corresponds to polyadenylated mRNA 3'-end tags, rendering normalization of reads to gene length obsolete (Supplementary Fig. 6). Furthermore, 3'-end sequencing enables the cell-type-specific re-evaluation of untranslated region (UTR) annotations to conduct mRNA 3' isoform-specific expression analysis (Supplementary Fig. 7). Following generating SLAM-seq libraries through the Quant-seq protocol from mESCs 24 h after s^4U metabolic labeling, we observed a strong accumulation of T > C conversions when compared with unlabeled conditions (Fig. 2b). Transcriptome-wide analyses confirmed this observation (Fig. 2c). In the absence of s^4U metabolic labeling, we observed a median rate of $\leq 0.1\%$ for any given conversion, consistent with Illumina-reported sequencing error, whereas s^4U -labeling resulted in a statistically significant ($P < 10^{-4}$, Mann–Whitney test), >50-fold increase in T > C conversion rates (Fig. 2c), which distributed evenly across the covered genomic regions (Fig. 2d, and Supplementary Fig. 8). Notably, non-T > C conversions remained below the expected sequencing error rates (Fig. 2c), and treatment of total RNA with iodoacetamide in the absence of metabolic labeling did not affect quantitative gene expression analysis (Supplementary Fig. 8d). s^4U incorporation measured by mass spectrometry in polyadenylated-enriched RNA was comparable with SLAM-seq data (Supplementary Fig. 9).

Measuring the polyadenylated transcriptional output in mESCs

Next, we subjected mESCs to 45 min of s^4U pulse labeling (final concentration of 100 μM s^4U), followed by total RNA extraction, alkylation and mRNA 3'-end library preparation (Supplementary Fig. 10a). To identify newly made transcripts, we extracted background-error-subtracted T > C conversion-containing reads for individual transcripts (Supplementary Table 1). Indeed, initial inspection of selected transcripts with comparable steady-state abundance (~ 100 cpm) revealed transcript-specific differences in the number of recovered T > C reads (Fig. 3a). Although high levels of T > C reads were recovered for the ESC-specific transcription factor Sox2 and the inherently instable primary microRNA transcript from the miR-290-295 cluster, the house-keeping transcript Gapdh was associated with fewer T > C reads, presumably because its accumulation to high steady-state expression levels was achieved by high transcript stability (Fig. 3a).

Transcriptional output by RNA polymerase II is regulated by transcription factors that bind cis-acting regulatory elements known as enhancers¹⁶. In ESCs, the pluripotent state is largely governed by a small number of enhancer-associated master transcription factors, including Oct4, Sox2 and Nanog, which drive the expression of the target genes necessary to maintain the ESC state (Supplementary Fig. 10b)¹⁷. Transcriptional output measurements by SLAM-seq revealed that well-established Oct4, Sox2 and Nanog (OSN) target genes produced overall a larger number of T > C conversion-containing reads in

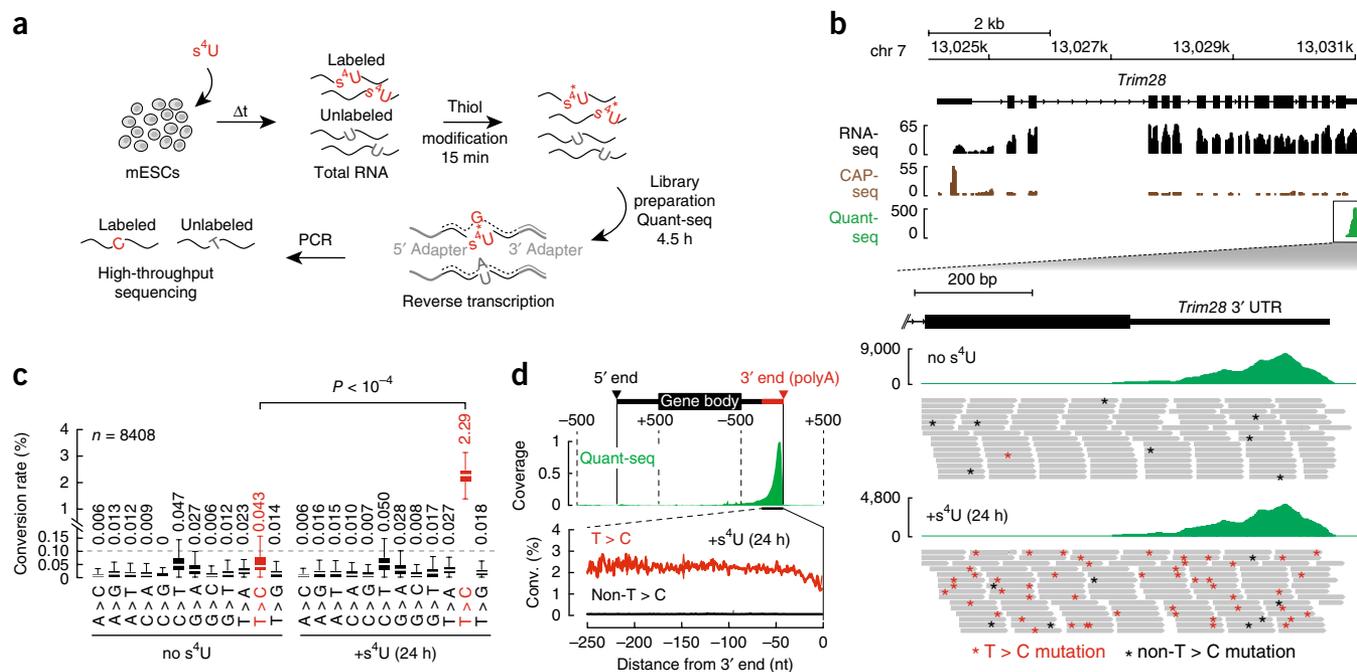


Figure 2 | Thiol-linked alkylation for the metabolic sequencing of RNA (SLAM seq). **(a)** Workflow of SLAM seq. The working times for alkylation and Quant-seq library preparation are indicated. **(b)** Representative genome browser screen shot for three independent mRNA libraries generated from total RNA of mESCs, prepared using standard mRNA sequencing (top), Cap-seq (middle) and mRNA 3' end sequencing (bottom; reads per million, RPM). A representative area in the mouse genome encoding the gene *Trim28* is shown. Bottom, magnification of the 3' UTR of *Trim28*. Unnormalized coverage plots of Quant-seq libraries prepared from untreated mESCs or mESCs subjected to s⁴U-metabolic labeling using 100 μM s⁴U for 24 h followed by SLAM seq. A random subset of individual reads underlying the coverage plots is depicted. Asterisks indicate T > C conversions (red) or any conversion other than T > C (black). **(c)** Conversion rates in defined counting window-mapping reads of Quant-seq libraries, prepared from mESCs before (no s⁴U) and after metabolic labeling for 24 h using 100 μM s⁴U (+s⁴U). Dashed line represents expected background sequencing error rate. Median conversion rate across the indicated number of transcripts (*n*) is shown above Tukey boxplots. Outliers are not shown. *P* value (Mann–Whitney test) is indicated. **(d)** Relative coverage across 8,408 transcripts in Quant-seq data sets. T > C conversion rate (Conv.) distributes evenly in Quant-seq-covered areas across 8,408 counting windows.

the s⁴U-pulse experiment (Fig. 3b, Supplementary Fig. 10b and Supplementary Table 1). When inspected globally, transcripts derived from the 2,029 expressed genes (>5 cpm steady state) with proximal OSN occupancy produced significantly more T > C reads when compared with 4,994 genes without proximal OSN enhancer (Mann–Whitney test, $P < 10^{-4}$; Fig. 3c)¹⁸. A subset of enhancers in mESCs were previously described to form arrays of regulatory elements (aka super or strong enhancer, SE) with unusually strong accumulations of transcriptional coactivators^{18,19}. In fact, the 156 genes next to strong enhancers exhibited the highest transcriptional output (Mann–Whitney test, $P < 10^{-4}$; Fig. 3c). In contrast, only genes proximal to strong enhancers were associated with above-average steady-state expression (Supplementary Fig. 10c). We concluded that SLAM seq provides a quantitative readout for enhancer activity in mESCs.

Given these findings, together with the fact that transcriptional output significantly correlated with data derived from global nuclear run-on experiments ($r = 0.41$, $P < 10^{-15}$; Supplementary Fig. 10d)²⁰, we conclude that SLAM seq uncouples transcriptional output from stability effects to globally measure RNA-polymerase-II-derived transcriptional activity.

Global and transcript-specific mRNA stability in mESCs

To directly measure mRNA transcript stabilities, we subjected mESCs to s⁴U metabolic RNA labeling (100 μM s⁴U) for 24 h,

followed by washout and chase using non-thiol-containing uridine, and prepared total RNA at various time points along the chase. Total RNA was then subjected to alkylation and Quant-seq (Supplementary Fig. 11a). Inspection of candidate genes revealed constant steady-state expression across the time course, whereas the number of T > C conversion-containing reads decreased over time in a transcript-specific manner (T > C reads; Supplementary Fig. 11b,c). After calculating the background-subtracted, U-content- and coverage-normalized T > C conversion rate for each transcript at every time point relative to 0 h chase, normalized T > C conversion rates fit well to single-exponential decay kinetics, enabling the determination of transcript half-life (Fig. 4a). As expected, RNA stabilities differed by more than one order of magnitude among individual transcripts (Fig. 4a). By fitting the data of 8,405 transcripts (steady-state expression >5 cpm) to single-exponential decay kinetics, we determined a median mRNA half-life of 3.9 h, corresponding to a cell-cycle-normalized half-life of 4.3 h (Fig. 4b and Supplementary Table 2). These measurements fall in the range of previously determined mRNA stabilities in mammalian cells²¹.

Previous studies proposed a close relationship between transcript-specific mRNA half-life and its physiological function^{1,10}. We therefore ranked the 6,665 transcripts for which half-life was determined at high accuracy ($r^2 > 0.6$) according to their relative stability and performed gene ontology enrichment analysis

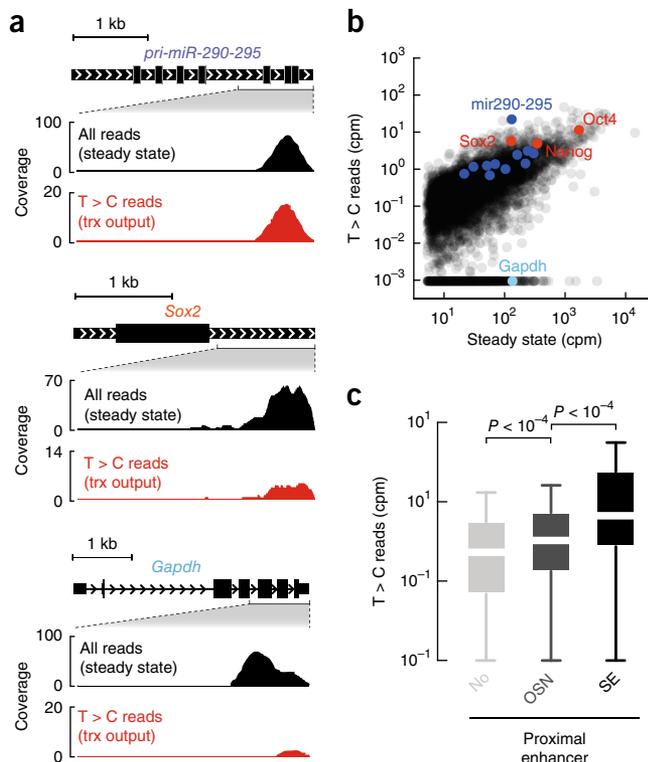


Figure 3 | Quantitative description of the polyadenylated transcriptional output in mESCs. (a) Genome browser plots of the indicated genes show SLAM seq data prepared from mESCs, subjected to s^4U metabolic RNA labeling. Black reads represent all mapped reads (steady state, in RPM); red reads represent T > C conversion-containing reads (*de novo* transcribed; trx output, in RPM). (b) Relative transcriptional output for 7,179 genes in mESCs. T > C reads represent abundance of *de novo* transcripts in counts per million (cpm); steady state represents sum of T > C and non-T > C containing reads. Core pluripotency transcription factors are highlighted in red, a subset of primary target genes for OSN in dark blue and a gene with housekeeping function in light blue. (c) Transcriptional output, as measured in number of T > C conversion containing reads, for expressed genes (steady state >5 cpm) without adjacent OSN enhancer (no, $n = 4,994$), proximal to canonical OSN enhancer ($n = 2,029$) or proximal to strong enhancers (SE, $n = 156$). Data are represented by Tukey boxplots without outliers. P values determined by Mann–Whitney test are indicated.

for the 666 most or least stable mRNAs (Fig. 4c). Transcripts with short half-life significantly enriched for regulators of RNA-polymerase-II-dependent transcription ($P < 10^{-3}$), whereas stable mRNAs associated with the gene ontology (GO) terms translation ($P < 10^{-14}$), respiratory electron transport ($P < 10^{-9}$) and oxidative phosphorylation ($P < 10^{-12}$). Together with gene set enrichment analyses (Supplementary Fig. 11d), SLAM seq measurements confirmed that transcripts encoding proteins with housekeeping functions tend to decay at low rates, perhaps reflecting evolutionary adaptation to energy constraints. In contrast, transcripts with a regulatory role tend to decay faster, most certainly because control over the persistence of genetic information facilitates adaptation to environmental changes¹.

We also examined global relationships between transcriptional output, mRNA stability and steady-state gene expression in mESCs as determined by SLAM seq pulse and pulse and chase experiments (Supplementary Fig. 12). Transcript biogenesis rates

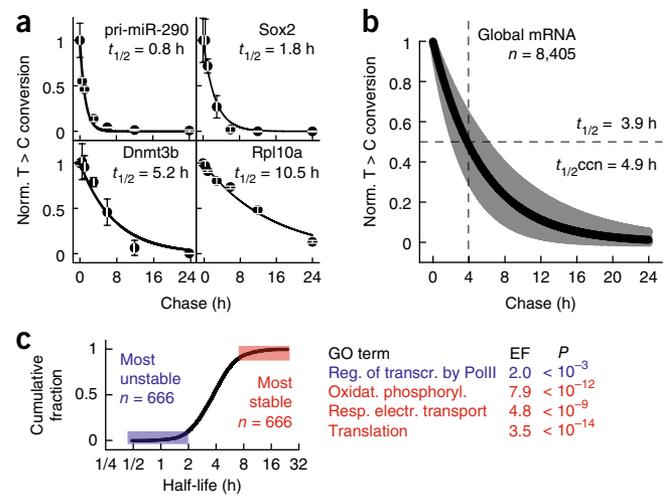


Figure 4 | Global and transcript-specific mRNA stability in mESCs. (a) Transcript stability for the indicated example genes as determined by SLAM seq. T > C conversion rates were determined for each time point of the s^4U pulse/chase and fit to a single-exponential decay model to derive half-life ($t_{1/2}$). Values are mean \pm s.d. of three independent cell cultures. (b) Global analysis of mRNA stability in mESCs. RNA half-life for 8,405 transcripts in mESCs determined as described in a. Single-exponential fit of median and interquartile range are shown. Median half-life before ($t_{1/2}$) or after ($t_{1/2}^{ccn}$) normalization to cell divisions. (c) Cumulative distribution of ranked high-confidence transcript stabilities for 6,665 transcripts. Enriched GO terms for the 666 most unstable (blue) and most stable (red) are indicated. Enrichment factor (EF) and P values are indicated.

and mRNA half-life both positively correlated with steady-state gene expression with correlation coefficients of 0.57 and 0.43, respectively. In contrast, the rates of mRNA biogenesis did not positively correlate with mRNA half-life ($r = -0.07$), but they showed high correlation with mRNA decay rates ($r = 0.66$). These results agree with a transcript-specific contribution of both mRNA synthesis and decay to the establishment of steady-state gene expression in mESCs. mRNA half-life measurements showed an overall good correlation with mRNA stabilities determined in mESCs after inhibition of transcription using actinomycin D ($r = 0.77$; Supplementary Fig. 11e).

SLAM-seq uncovers molecular determinants of mRNA stability

To further validate SLAM seq, we performed mechanistic studies on two specific post-transcriptional gene regulatory pathways with well-established biological functions in mESCs.

First, we focused on microRNAs (miRNAs), which act as key regulators of gene expression²². In mESCs, they contribute to cell-state maintenance and transitions by tuning the expression of ESC transcripts and promoting their clearance during differentiation¹⁷. At the molecular level, miRNAs act as guides for ribonucleoprotein complexes that target complementary sites, usually in the 3' UTR of mRNAs, as defined by miRNA seed sequence (nucleotides 2 to 7 or 8 of the miRNA)²². MicroRNAs elicit their function by repressing translation and/or promoting mRNA decay, although the relative contribution of repressive modes remains a matter of debate and may vary in different biological contexts²³. We determined the stability of miRNA targets in wild-type mESCs by inspecting the half-life of transcripts harboring in their 3' UTR target sites for the

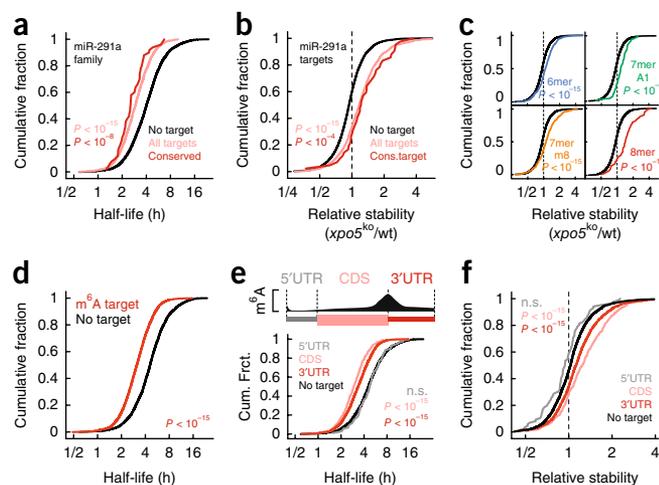


Figure 5 | Molecular determinants of mRNA stability in mESCs.

(a) Cumulative distribution of ranked mRNA stabilities. Plotted are distributions for transcripts that do (rose, $n = 1,450$) or do not (black, $n = 5,095$) contain at least one miR-291a family target site or contain at least one conserved miR-291a target site (red, $n = 50$). miR-291a family members were defined as described in **Supplementary Figure 13g**. P values were determined by KS test. (b) Cumulative distribution of mRNA stability changes in $xpo5^{ko}$ relative to wild-type mESCs. Plotted are distributions for transcripts that do (rose, $n = 1,288$) or do not (black, $n = 4,825$) contain at least one miR-291a target site or that contain at least one conserved miR-291a target site (red, $n = 42$). P values were determined by KS test. (c) Cumulative distribution of mRNA stability changes in $xpo5^{ko}$ relative to wild-type mESCs. Plotted are distributions for transcripts that contain exclusively one 6-mer (blue, $n = 493$), 7-mer-A1 (green, $n = 95$), 7-mer-m8 (yellow, $n = 325$) or 8-mer site (red, $n = 63$). Black shows transcripts without a miR-291a target site ($n = 4825$). P values were determined by KS test. (d) Cumulative distribution of ranked mRNA stabilities. Plotted are distributions for transcripts that do (red, $n = 3,492$) or do not (black, $n = 3,173$) contain the m⁶A mark, as previously mapped by m⁶A-RIP-seq²⁸. P values were determined by KS test. (e) Top, schematic distribution of m⁶A in mRNA (adapted from ref. 30). Bottom, cumulative distribution of ranked mRNA stabilities. Plotted are distributions for transcripts that do not (black, $n = 3,173$) or do contain m⁶A exclusively in the 5' UTR (gray, $n = 88$), the coding sequence (CDS, green, $n = 545$) or the 3' UTR (red, $n = 2,093$). P values were determined by KS test. (f) Cumulative distribution of mRNA stability changes in $mettl3^{ko}$ relative to wild-type mESCs. Plotted are distributions for transcripts that do not (black, $n = 3,118$) or do m⁶A exclusively in the 5' UTR (gray, $n = 86$), the coding sequence (CDS, green, $n = 518$) or the 3' UTR (red, $n = 2,017$). P values were determined by KS test.

miR-291-3p/294-3p/295-3p/302-3p and miR-292a-3p/467a-5p family (referred to as the miR-291a family), which share the same seed sequence and derive from the ESC-specific miR-290-295 cluster that gives rise to more than half of all small RNAs expressed in this cell type (**Supplementary Fig. 13b**). With a median half-life of 2.9 h ($n = 1,450$ transcripts), miR-291a family targets were significantly less stable than transcripts without sites ($t_{1/2} = 4.0$ h, $n = 5,095$ transcripts, Kolmogorov-Smirnov (KS) test, $P < 10^{-15}$; **Fig. 5a**). Transcripts with conserved sites exhibited an even shorter half-life ($t_{1/2} = 2.6$ h, $n = 50$ transcripts; **Fig. 5a**). To confirm the direct contribution of miRNAs to transcript destabilization, we determined changes in mRNA half-life by s⁴U pulse labeling followed by SLAM seq in mESCs depleted of the core miRNA biogenesis factor exportin-5 (Xpo5) by CRISPR-Cas9 (**Supplementary**

Fig. 13c,d). Depletion of Xpo5 reduced overall miRNA levels by more than 90%, and miR-291a-family members by more than 95%, as determined by northern hybridization (**Supplementary Fig. 13e**) and small RNA sequencing (Student's t test, $P < 10^{-4}$; **Supplementary Fig. 13f**). We observed a significant increase in relative mRNA stability for targets of the miR-291a family when compared with transcripts without target site (KS test, $P < 10^{-15}$ and $P < 10^{-4}$ for all or conserved sites, respectively; **Fig. 5b**). Notably, the degree of de-repression followed previously established rules for miRNA targeting (**Fig. 5c** and **Supplementary Fig. 13a**)²². Although each site type responded to Xpo5 depletion with a significant increase in mRNA stability (KS test, $P < 10^{-8}$), 6-mer target sites exhibited the weakest effects, followed the two 7-mer site types (**Fig. 5c**). 8-mer sites showed the strongest de-repression (**Fig. 5c**). Finally, by inspecting target mRNAs of less abundant miRNA families, we confirmed that miRNA function, as determined by target mRNA stability in wild-type mESCs and relief of repression following depletion of Xpo5, was directly dependent on small RNA abundance (**Supplementary Fig. 13g-i**), as previously described^{24,25}.

Second, we focused on N⁶-methyladenosine (m⁶A), the most abundant internal modification in mammalian mRNA, which has been implicated in the regulation of various physiological processes^{26,27}. In ESCs, m⁶A facilitates the resolution of naive pluripotency toward differentiation^{28,29}. At the mechanistic level, the m⁶A mark impinges on various aspects of mRNA processing, including mRNA stability³⁰ (**Supplementary Fig. 14a**). To estimate the effect of m⁶A on mRNA stability in mESCs, we first determined the general association of m⁶A targets, as mapped previously by m⁶A RNA immunoprecipitation and sequencing²⁸, with mRNA stability in wild-type cells. With a half-life of 3.1 h, m⁶A-containing transcripts ($n = 3,492$ transcripts) were significantly less stable than naive transcripts ($t_{1/2} = 4.6$ h, $n = 3,173$ transcripts, KS test, $P < 10^{-15}$; **Fig. 5d**). N⁶-methyladenosine marks do not distribute evenly in mRNAs, but are enriched in long exons, near stop codons and in 3' UTRs, although m⁶A also occurs in the coding region (CDS) and 5' UTR³⁰ (**Fig. 5e**). We investigated the relationship between the position of m⁶A in targeted mRNAs and its effect on RNA decay. We found that mRNAs containing m⁶A exclusively in the CDS ($n = 545$) or in the 3' UTR ($n = 2,093$ transcripts) were significantly less stable than naive transcripts (KS test, $P < 10^{-15}$; **Fig. 5e**). In contrast, mRNAs that contained m⁶A exclusively in the 5' UTR ($n = 88$) were not less stable than naive transcripts (KS test, $P > 0.05$; **Fig. 5e**). To confirm the causal contribution of m⁶A to transcript destabilization, we determined changes in mRNA half-life by s⁴U pulse labeling followed by SLAM seq in mESCs depleted of Mettl3, the catalytic subunit of the m⁶A RNA methylation complex, which resulted in the codepletion of its RNA-binding partner protein Mettl14 (**Supplementary Fig. 14b,c**)³¹. Consistent with a direct and position-dependent effect of m⁶A on mRNA decay, we observed a significant increase in relative mRNA stability for transcripts containing m⁶A in the CDS or 3' UTR (KS test, $P < 10^{-15}$), but not in the 5' UTR ($P > 0.05$; **Fig. 5f**). Similar results were obtained when re-investigating recently described m⁶A profiling data in mESCs (**Supplementary Fig. 14d-f**)³².

DISCUSSION

Recent efforts in decoding RNA modifications have led to the emergence of epitranscriptome sequencing technologies profiling

ribonucleotide modifications on a genomic scale^{13,14}. Here we present an orthogonal-chemistry-based sequencing strategy for the identification of s⁴U, which is widely used for *in vivo*, *ex vivo* and *in vitro* RNA labeling and represents a natural base modification in eubacterial and archaeal tRNA^{4–8,33}.

Combining SLAM seq with Quant-seq provides several advantages with important practical and conceptual implications. First, the specific sampling of poly-adenylated RNA species assigns kinetics to functional, fully processed RNA polymerase II transcripts (Fig. 2 and Supplementary Fig. 6). Second, it provides access to mRNA 3' isoform-specific expression dynamics (Supplementary Fig. 7). Third, by eliminating the requirement to normalize for transcript length, Quant-seq facilitates downstream data analysis (Supplementary Fig. 6). Fourth, Quant-seq produces highly reproducible results from as little as 100 pg of total RNA, without a requirement for rRNA depletion, thereby providing access to cellular systems for which starting material is limiting. Finally, high sequencing coverage across inherently U-rich 3' UTRs facilitates the robust quantification of T > C conversions. Note that Quant-seq restricts gene expression analysis to RNA polymerase II transcripts and fails to differentiate transcript variants such as splice isoforms. However, alternative sequencing methods may augment the applicability of SLAM seq, as s⁴U-identification by sequencing is in principle compatible with any RNA library preparation method that involves a reverse transcription step.

Studying intracellular RNA kinetics by s⁴U metabolic RNA labeling requires general and method-specific considerations to be taken into account; s⁴U-incorporation was previously linked to rRNA processing defects in human cancer cells³⁴. Because s⁴U uptake can vary between cell types, careful assessment of cell-type-specific toxicity is imperative to meet s⁴U-labeling conditions that do not affect gene expression or cell viability (Supplementary Fig. 5)^{5,35}. In mESCs, nontoxic concentrations of 100 μM s⁴U result in a median s⁴U incorporation of 2.29% across 8,408 transcripts following long-term metabolic labeling (that is, 24 h), corresponding to one s⁴U incorporation in every 43 uridines at steady-state labeling conditions (Fig. 2c and Supplementary Fig. 8c). Considering the U content of mRNA 3' UTRs (~31% in mESCs), SLAM seq recovers each s⁴U-labeled transcript at a probability of up to 35% or 70% in single-read 50 or 100 sequencing reactions, respectively, which enables labeled-transcript identification in even short s⁴U pulse-labeling conditions (Fig. 3). Note that the ability to determine *de novo* synthesized transcripts will depend on the cellular s⁴U uptake kinetics, the overall transcriptional activity of the cell type and the library sequencing depth. Thus, these parameters need to be taken into account when designing a SLAM seq experiment, particularly when employing short s⁴U pulse labeling, where sequencing depth demands adjustments to the given cellular parameters. In that respect, s⁴U-tagging approaches (such as, TT-seq) may provide some advantage over SLAM seq when analyzing transient RNA species that escape detection by standard RNA sequencing approaches⁸.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

We thank J. Jude (Research Institute of Molecular Pathology, IMP) for generously providing a modified version of pLenti-CRISPR-v2-GFP, G. Krssakova and K. Mechtler (IMP, Vienna) for high-performance liquid chromatography analysis, IMP/Institute of Molecular Biotechnology (IMBA) Biooptics facility for FACS support, and all laboratory members for support and discussions. Mass spectrometry was performed at the Vienna Biocenter Core Facilities (VBCF) Metabolomics unit (<http://www.vbcf.ac.at>), funded by the City of Vienna through the Vienna Business Agency. High-throughput sequencing was performed at the VBCF NGS Unit (<http://www.vbcf.ac.at>). This work was supported by grants from the European Research Council to S.L.A. (ERC-StG-338252) and J.Z. (ERC-StG-336860) and the Austrian Science Fund to S.L.A. (Y-733-B22 START, W-1207-B09 and SFB F43-22) and A.v.H. (W-1207-B09). The IMP is generously supported by Boehringer Ingelheim.

AUTHOR CONTRIBUTIONS

V.A.H. and S.L.A. conceived the approach and wrote the paper. V.A.H., B.R. and S.L.A. developed the methods, performed the experiments and analyzed the data. W.W. performed initial s⁴U-alkylation experiments. T.N., P.R., V.A.H., J.Z., A.v.H. and S.L.A. developed SLAM-DUNK. P.B., T.R.B., V.A.H. and S.L.A. performed mRNA 3'-end annotation.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
- Rabani, M. *et al.* High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014).
- Yao, S.Y.M., Ng, A.M.L., Cass, C.E., Baldwin, S.A. & Young, J.D. Nucleobase transport by human equilibrative nucleoside transporter 1 (hENT1). *J. Biol. Chem.* **286**, 32552–32562 (2011).
- Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R. & Boothroyd, J.C. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat. Biotechnol.* **23**, 232–237 (2005).
- Dölken, L. *et al.* High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959–1972 (2008).
- Miller, C. *et al.* Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* **7**, 458 (2011).
- Duffy, E.E. *et al.* Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Mol. Cell* **59**, 858–866 (2015).
- Schwab, B. *et al.* TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).
- Miller, M.R., Robinson, K.J., Cleary, M.D. & Doe, C.Q. TU-tagging: cell-type-specific RNA isolation from intact complex tissues. *Nat. Methods* **6**, 439–441 (2009).
- Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29**, 436–442 (2011).
- Gay, L. *et al.* Mouse TU tagging: a chemical/genetic intersectional method for purifying cell-type-specific nascent RNA. *Genes Dev.* **27**, 98–115 (2013).
- Sun, M. *et al.* Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res.* **22**, 1350–1359 (2012).
- Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **14**, 23–31 (2016).
- Heiss, M. & Kellner, S. Detection of nucleic acid modifications by chemical reagents. *RNA Biol.* **0**, 1–9 (2016).
- Shugart, L. Effect of selective chemical modification of 4-thiouridine of phenylalanine transfer ribonucleic acid on enzyme recognition. *Arch. Biochem. Biophys.* **148**, 488–495 (1972).
- Ong, C.-T. & Corces, V.G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–293 (2011).

17. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
18. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
19. Pott, S. & Lieb, J.D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).
20. Min, I.M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25**, 742–754 (2011).
21. Tani, H. & Akimitsu, N. Genome-wide technology for determining RNA stability in mammalian cells: historical perspective and recent advantages based on modified nucleotide labeling. *RNA Biol.* **9**, 1233–1238 (2012).
22. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
23. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12**, 99–110 (2011).
24. Ameres, S.L. & Zamore, P.D. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14**, 475–488 (2013).
25. Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods* **9**, 840–846 (2012).
26. Meyer, K.D. & Jaffrey, S.R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15**, 313–326 (2014).
27. Yue, Y., Liu, J. & He, C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev.* **29**, 1343–1355 (2015).
28. Batista, P.J. *et al.* m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).
29. Geula, S. *et al.* Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
30. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m⁶A RNA methylation. *Nat. Rev. Genet.* **15**, 293–306 (2014).
31. Wang, X. *et al.* Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. *Nature* **534**, 575–578 (2016).
32. Ke, S. *et al.* m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* **31**, 990–1006 (2017).
33. Limbach, P.A., Crain, P.F. & McCloskey, J.A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **22**, 2183–2196 (1994).
34. Burger, K. *et al.* 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol.* **10**, 1623–1630 (2013).
35. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

ONLINE METHODS

A step-by-step protocol is available as a **Supplementary Protocol** and as an open resource in *Protocol Exchange*³⁶.

Carboxyamidomethylation of s⁴U. If not indicated otherwise, carboxyamidomethylation was performed under standard conditions (50% DMSO, 10 mM iodoacetamide, 50 mM sodiumphosphate buffer pH8, for 15 min at 50 °C) using either 1 mM 4-thiouracil (Sigma), 800 μM 4-thiouridine (Sigma) or 5–50 μg total RNA prepared from s⁴U metabolic labeling experiments. The reaction was quenched by addition of excess DTT.

Adsorption measurements. 1mM 4-thiouracil was incubated under optimal reaction conditions (10mM iodoacetamide, 50% DMSO, 50 mM sodiumphosphate buffer pH 8, for 15 min at 50 °C) if not indicated otherwise. Reaction was quenched by the addition of 100 mM DTT and adsorption spectra were measured on a Nanodrop 2000 instrument (Thermo Fisher Scientific), followed by baseline subtraction of adsorption at 400 nm.

Mass spectrometry. 40 nmol 4-thiouridine were reacted in the absence or presence of 0.05, 0.25, 0.5 or 5 μmol iodoacetamide under standard reaction conditions (50 mM sodiumphosphate buffer, pH 8; 50% DMSO) at 50 °C for 15 min. The reaction was stopped with 1% acetic acid. Acidified samples were separated on a Ultimate U300 BioRSLC HPLC system (Dionex; Thermo Fisher Scientific), employing a Kinetex F5 Pentafluorophenyl column (150 mm × 2.1 mm; 2.6 μm, 100 Å; Phenomenex) with a flow rate of 100 μl/min. Nucleosides were on-line analyzed using a TSQ Quantiva mass spectrometer (Thermo Fisher Scientific) after electrospray ionization with the following SRMs: 4-thiouridine *m/z* 260 → 129, and alkylated 4-thiouridine *m/z* 318 → 186. Data were interpreted using the Trace Finder software suite (Thermo Fisher Scientific) and manually validated.

To determine s⁴U incorporation events in polyadenylated or total RNA by mass spectrometry, total RNA was either subjected to oligo(dT) enrichment using Dynabeads Oligo (dT)²⁵ (Ambion) following manufacturer's instructions to purify polyadenylated RNA or directly enzymatically degraded to monomeric ribonucleosides as described previously before Mass spectrometry analysis³⁷.

Primer extension assays. Primer extension assays were essentially performed as described previously³⁸. Briefly, template RNA oligonucleotides (5L-*let-7-3L* or 5L-*let-7-s⁴Up9-3L*; Dharmacon; see **Supplementary Table 3** for sequences) were deprotected according to the instructions of the manufacturer and purified by denaturing polyacrylamide gel-elution. 100 μM purified RNA oligonucleotides were treated with 10 mM iodoacetamide (+IAA) or EtOH (−IAA) in standard reaction conditions (50% DMSO, 50 mM sodiumphosphate buffer, pH8) for 15 min at 50 °C. The reaction was stopped by addition of 20 mM DTT, followed by ethanol precipitation. RT primer (see **Supplementary Table 3** for sequence) was 5' radiolabeled using γ-³²P-ATP (Perkin-Elmer) and T4-polynucleotide kinase (NEB), followed by denaturing polyacrylamide gel-purification. 640 nM γ-³²P-RT primer was annealed to 400 nM 5L-*let-7-3L* or 5L-*let-7-s⁴Up9-3L* in 2 × annealing buffer (500 mM KCl, 50 mM Tris pH 8.3) in a PCR machine (3 min 95 °C, 30 s 85 °C Ramp 0.5 °C/s, 5 min

25 °C Ramp 0.1 °C/s). Reverse transcription was performed using Superscript II (Invitrogen), Superscript III (Invitrogen), or Quant-seq RT (Lexogen) as recommended by the manufacturer. For dideoxynucleotide reactions, a final concentration of 500 μM ddNTP (as indicated) was added to RT reactions. Following completion, RT reactions were resuspended in formamide loading buffer (Gel loading buffer II, Thermo Fisher Scientific) and subjected to 12.5% denaturing PAGE. Gels were dried, exposed to storage phosphor screen (PerkinElmer), imaged on a Typhoon TRIO variable mode imager (Amersham Biosciences), and quantified using ImageQuant TL v7.0 (GE Healthcare). For analysis of RT drop-off, signal-intensities at p9 were normalized to preceding drop-off signal intensities (bg; **Supplementary Fig. 3b**) for individual reactions. Values reporting the change in drop off signal (+IAA/−IAA) for s⁴U-containing and non-containing RNA oligonucleotides were compared for the indicated reverse transcriptases.

HPLC analysis of s⁴U-labeled RNA. Analysis of s⁴U-incorporation into total RNA following metabolic labeling was performed as previously described³⁷.

Cell viability assay. 5,000 mESCs were seeded per 96 well the day before the experiment. After onset of the experiment, media containing the indicated concentration of s⁴U was replaced every three hours for a total of 12 h or 24 h. Cell viability was assessed by CellTiter-Glo Luminescent Cell Viability Assay (Promega) according to the instructions of the manufacturer. Luminescent signal was measured on Synergy (BioTek) using Gen5 Software (v2.09.1).

Cell culture. mESCs (clone AN3-12), derived from C57BL/6x129 F1 females, were obtained from IMBA Haplobank (Elling, U., Wimmer, R.A., Leibbrandt, A., Burkard, T., Michlits, G., Leopoldi, A., Micheler, T., Abdeen, D., Zhuk, S., Aspalter, I.M., Handt, C., Liebergesell, J., Hubmann, M., Husa, A.-M., Kinzer, M., Schuller, N., Wetzl, E., van der Loo, N., Arturo, J., Martinez, Z., Estoppey, D., Riedl, R., Yang, F., Fu, B., Dechat, T., Ivics, Z., Agu, C.A., Bell, O., Blaas, D., Gerhardt, H., Hoepfner, D., Stark, A. & Penninger, J.M., unpublished data) and cultured in 15% FBS (Gibco), 1 × penicillin-streptomycin solution (100 U/ml penicillin, 0.1 mg/ml streptomycin, Sigma), 2 mM L-glutamine (Sigma), 1 × MEM Non-essential amino acid solution (Sigma), 1 mM sodium pyruvate (Sigma), 50 μM 2-Mercaptoethanol (Gibco) and 20 ng/ml LIF (in-house produced). Cells were maintained at 37 °C with 5% CO₂ and passaged every second day. Cell doubling time of AN3-12 mES in presence of s⁴U cells as determined by cell counting was 14.7h. Prior to metabolic labeling experiments, mESCs were stained with Hoechst33342 and FACS-sorted to obtain a pure diploid population³⁹.

SLAM seq in mESCs. See *Protocol Exchange* for detailed information regarding SLAM seq³⁶. mESCs were seeded the day before the experiment at a density of 10⁵ cells/ml. s⁴U-metabolic labeling in mESCs was performed by incubating mESCs in standard medium but adding s⁴U (Sigma) to a final concentration of 100 μM and media exchange every 3 h for the duration of the pulse. For the uridine chase experiment, cells were washed twice with 1 × PBS and incubated with standard medium supplemented with 10 mM uridine (Sigma). At respective time points, cells were harvested followed by total RNA extraction using TRIzol (Ambion) following the manufacturer's instructions but including 0.1 mM DTT

(final concentration) during isopropanol precipitation. RNA was resuspended in 1 mM DTT. For a typical SLAM-seq experiment, 5 µg total RNA were treated with 10 mM iodoacetamide under optimal reaction conditions and subsequently ethanol precipitated and subjected to Quant-seq 3'-end mRNA library preparation.

RNA library preparation. Standard RNA-seq libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) following the instructions of the manufacturer. Cap-seq libraries were prepared as previously described⁴⁰. mRNA 3' end sequencing was performed using the Quant-seq mRNA 3' end library preparation kit (Lexogen) according to the instructions of the manufacturer. Small RNA libraries were generated as described before⁴¹, but adding total RNA from *Arabidopsis thaliana* unopened floral buds as spike-in before initial size-selection. Sequencing was performed on Illumina HiSeq 2500. Libraries were sequenced in SR50 mode except for transcriptional output measurements (Fig. 3), which were sequenced in SR100 mode.

Transcriptional inhibition by actinomycin D. 3×10^5 AN3-12 mESCs were seeded per 35-mm plate and grown overnight. To block transcription, actinomycin D (Sigma) was added to the medium at the concentration of 5 µg/ml. Cells were harvested at 0, 0.25 h, 0.5 h, 1 h, 3 h and 10 h after addition of actinomycin D by directly lysing them in TRIzol (Ambion). RNA was extracted following the manufacturer instructions and libraries were prepared using Quant-seq mRNA 3'-end library preparation kit (Lexogen) according to the instructions of the manufacturer.

CRISPR-Cas9 genome engineering. gRNAs were designed using WTSI Genome Editing⁴². gRNA oligonucleotides (see **Supplementary Table 3**) were cloned into pLenti-CRISPR-v2-GFP vector as described⁴³, but modified by replacing the puromycin resistance cassette with GFP. Prior to gRNA transfection targeting Xpo5 or Mett13, wildtype An3-12 mESCs were FACS sorted for haploid cells as described previously³⁹. 3×10^5 cells were seeded per six wells and transfected the next day with 3 µg pLenti-CRISPR-v2-GFP using Lipofectamine 2000 as recommended by the manufacturer. 48 h after transfection, GFP-positive cells were sorted by fluorescence-activated cell sorting (FACS) and 1,500 cells were subsequently seeded per 15-cm plate. Single colonies were picked after 10 d. DNA isolation, PCR amplification (for oligonucleotide sequences, see **Supplementary Table 3**) of the targeted locus and Sanger sequencing was performed to genotype the clonal cell lines. Protein depletion was confirmed by western blot analysis.

Western blotting. Protein lysates were separated on 10% SDS PAGE and transferred to PVDF membrane (BioRad). Antibodies were used at a dilution of 1:500 for anti-Exportin-5 (H-300, sc-66885, rabbit), 1:3,000 for anti-Mett13 (15073-1-AP, Proteintech, rabbit), 1:5,000 for anti-Mett14 (HPA038002, Sigma, rabbit) and 1:10,000 for anti-Actin (A2066, Sigma, rabbit) and detected by secondary HRP-antibody-conjugates G21040 (Invitrogen; dilution 1:10,000). Primary antibodies were incubated at room temperature for three hours and secondary antibodies were incubated at room temperature for two hours. Images were acquired on a ChemiDoc MP Imaging System (BioRad) using ImageLab v5.1.1 (Bio-Rad) or by Amersham Hyperfilm ECL (GE Healthcare).

Northern blotting. Northern hybridization experiments were performed as described previously⁴⁴. For northern probes, see **Supplementary Table 3**.

Bioinformatics and data analysis. Gel images were quantified using ImageQuant v7.0a (GE Healthcare). Curve fitting was performed according to the integrated rate law for a first-order reaction in Prism v7.0 (GraphPad) or R (v2.15.3) using the minpack.lm package.

For sequencing analysis of synthetic RNA samples (**Fig. 1d** and **Supplementary Fig. 4**) barcoded libraries were demultiplexed using Picard Tools BamIndexDecoder v1.13 allowing 0 mismatches in the barcode. Resulting files were converted to fastq using picard-tools SamToFastq v1.82. Cutadapt v1.7.1 was used to trim adapters (allowing for default 10% mismatch in adaptor sequence) and filter for sequences of 21-nt length. Resulting sequences were aligned to aligned to mature dme-let-7 sequence (TGAGGTAGTAGGTTGTATAGT) using bowtie v0.12.9 allowing for three mismatches and converted to bam using samtools v0.1.18. 'N'-containing sequences were filtered from alignment. Remaining alignments were converted to pileup format. Finally, fractions of each conversion per position were extracted from pileup. Output table was analyzed and plotted in Excel v15.22 (Microsoft) and Prism v7.0a (GraphPad).

For standard RNA sequencing data analysis, barcoded libraries were demultiplexed using Picard Tools BamIndexDecoder v1.13 allowing one mismatch in the barcode. Adapters were clipped using cutadapt v1.5 and reads were size-filter for ≥ 15 nucleotides. Reads were aligned to mouse genome mm10 using STAR aligner v2.5.2b⁴⁵. Alignments were filtered for alignment scores ≥ 0.3 and alignment identity ≥ 0.3 was normalized to read length. Only alignments with ≥ 30 matches were reported and chimeric alignments with an overlap ≥ 15 bp were allowed. Two-pass mapping was used. Introns < 200 kb were filtered and alignments containing non-canonical junctions were filtered. Alignment with a mismatch to mapped bases ratio ≥ 0.1 or with a max. number of 10 mismatches were excluded. The max number of gaps allowed for junctions by 1,2,3,N reads was set to 10, 20, 30 and 50 kb, respectively. The minimum overhang length for splice junctions on both sides for (1) non-canonical motifs, (2) GT/AG and CT/AC motif, (3) GC/AG and CT/GC motif, (4) AT/AC and GT/AT motif was set to 20, 12, 12, 12, respectively. 'Spurious' junction filtering was used and the maximum number of multiple alignments allowed for a read was set to 1. Exonic reads (Gencode) were quantified using FeatureCounts⁴⁶.

For Cap analysis gene expression (Cap-Seq), barcoded libraries were demultiplexed using Picard Tools BamIndexDecoder v1.13 allowing one mismatch in the barcode. The first 4nt of the reads were trimmed using seqtk. Reads were screened for ribosomal RNA by aligning with BWA (v0.6.1)⁴⁷ against known rRNA sequences (RefSeq). The rRNA subtracted reads were aligned with TopHat (v1.4.1)⁴⁸ against the *Mus musculus* genome (mm10). Maximum multihits was set to 1, segment-length to 18 and segment-mismatch to 1. Additionally, a gene model was provided as GTF (Gencode VM4).

For analysis of mRNA 3'-end sequencing (Quant-seq) data sets, reads were demultiplexed using Picard Tools BamIndexDecoder v1.13 allowing one mismatch in the barcode. Quant-seq data was processed using Digital Unmasking of Nucleotide conversion-containing k-mers (DUNK), SLAM-DUNK v0.2.4, a

T > C aware alignment software package based on NextGenMap⁴⁹ developed to recover T > C conversions from SLAM seq data sets (Neumann, T., Herzog, V.A., Muhar, M., von Haeseler, A., Zuber, J., Ameres, S.L. & Rescheneder, P., unpublished data). Briefly, adaptor-clipped reads were trimmed 12 bp from the 5' end (-5 12) and polyadenylated stretches (>4 subsequent As at the 3' end) were removed. Trimmed reads were aligned to the full reference genome (mm10) using local alignment scoring and up to 100 alignments were reported for multimapping reads (-n 100). In the filtering step, alignments with a minimum identity of 95% and a minimum of 50% of the read bases mapped were retained. Among multimappers, reads mapping to no or ambiguously to > 1 annotated UTR sequence (bed files provided in GEO data sets) were discarded (-fb). If a multimapping read mapped >1 time to the same annotated UTR sequence, one alignment was randomly picked. SNPs exceeding a coverage cutoff of 10× and a variant fraction cutoff of 0.8 were called using VarScan2.4.1 using default parameters⁵⁰. Non-SNP overlapping T > C-conversions with a base quality of Phred score >26 were identified. T > C containing reads and total reads aligning within the custom defined counting windows (bed files provided in GEO data sets) were reported. T > C conversion rate was determined for each position along the custom defined counting windows by normalizing to genomic T content and coverage of each position and averaged per UTR.

For extended mRNA 3'-end annotation, we assembled a pipeline to annotate 3' ends of mRNA transcripts using Quant-seq data sets (<https://github.com/AmeresLab/UTRannotation>). Quant-seq data was pre-processed as described above. To determine exact priming sites, reads with continuous 3' terminal A stretches (>4) and a length of at least 23 nt were retained. Polymeric A-stretches were trimmed from the 3' ends of reads and mapped to mm10 using SLAM-DUNK's map and filter module as described above but using global alignment scoring. Priming sites were identified based on mapping of > = 10 reads to genomic positions and consecutive positions were merged. Genomic A content of ≥ 0.36 and ≥ 0.24 was used to identify internal priming events (for polyA site-containing and no-polyA site-containing priming sites respectively, see **Supplementary Fig. 7** for PAS sequences). Priming sites overlapping with RefSeq and ENSEMBL 3' UTR annotations were considered for further analysis (UTRends). RNA-seq signal, mapped as described above, was used to identify intergenic ends. RNA-seq coverage was calculated using bedtools multicov in 200-nt bins separated by 20 nt starting from the last 200 nt of gene annotations. Bins were extended until RNA-seq coverage dropped below 10% compared to the first bin or until the bin overlapped another gene annotation. Priming sites overlapping identified counting bins were retained (intergenicEnds). For each gene, all identified 3' ends were ranked by underlying counts and ends that did not exceed 10% of the total signal were removed. RefSeq-annotated mRNA 3' ends were then included and 250-nt counting windows were created upstream of 3' ends. Overlapping counting windows were merged. Beyond protein coding mRNAs, counting windows were added for the following classes of non-coding RNAs: antisense, bidirectional_promoter_lincRNA, lincRNA, macro_lincRNA, processed_transcript, sense_intronic, sense_overlapping and primary miRNAs. To annotate 3' UTR start positions for *de novo* annotated 3' ends, each 3' end was assigned to the most proximal 3' UTR start annotation (RefSeq).

For comparison of Quant-seq and RNA-seq, we employed RefSeq transcripts of mm10 from UCSC's table browser (downloaded 2017-02-14) consisting of 35,805 transcripts which we mapped to 24,440 Entrez genes. All transcripts for a given gene were merged using bedtools⁵¹. Stranded coverage tracks for Quant-seq and RNA-seq samples were created using deeptools' *bamCoverage* command⁵², using a binSize of 1 and normalizing to RPKM. Next, the density matrix was calculated separately for + and - strand genes, with static windows 500 bp in both directions at TSS and TTS and dynamic binning for the remaining gene body. Stranded signal from the density matrix was plotted in composite plots.

For transcriptional output analysis, the number of normalized reads (in cpm; steady-state expression) and the number of normalized reads containing ≥ 1 T > C conversion (in cpm; transcriptional output) were obtained for every gene after aligning SLAM seq data with SLAM-DUNK to the mouse genome mm10. Background T > C reads (T > C reads observed without s⁴U labeling) were subtracted from the T > C reads in the 45-min time point and an expression threshold of >5 cpm for the mean of steady-state expression was set. Genes were classified as proximal to 'no', 'OSN' or 'strong/super' enhancer as described previously¹⁸.

GRO-seq data from mESCs was downloaded from GEO (GSE27037)²⁰. Reads were mapped to mm10 using bowtie allowing for uniquely mapping reads with at most two mismatches. Unmapped reads were reiteratively trimmed by one nucleotide and remapped until reaching a minimum length of 20 nucleotides. GRO-seq signal was assessed using featureCounts⁴⁶ for the full length gene omitting the first kilobase. Transcriptional output as determined by SLAM seq was then compared to GRO-seq for all genes that are expressed above 5cpm in Quant-seq data sets and detected in GRO-seq data sets.

To calculate RNA half-lives, T > C conversions were background-subtracted (no s⁴U treatment) and normalized to chase-onset. Curve fitting was performed according to the integrated rate law for a first-order reaction in R (v2.15.3) using the minpack.lm package. RNA half-lives > 24 h were set to 24 h. If not stated otherwise an R² cutoff of >0.6 was applied. To calculate RNA half-lives normalized to cell cycle length, T > C conversions were multiplied by $2^{(\text{time point}/14.7\text{h})}$.

To calculate RNA stabilities measured by polymerase II inhibition (ActD treatment), reads from the Actinomycin D-treated samples were aligned to mm10 using SLAM-DUNK. Transcripts were extracted that were expressed >5 cpm in the SLAM seq experiment. To correct for the relative increase in stable transcripts following global transcriptional inhibition, data was normalized to the 50 most stable transcripts. Half-lives were calculated by fitting data to a single-exponential decay model as described above.

GO terms enrichment analysis was performed using PANTHER database with a custom reference set consisting of genes expressed >5 cpm in mESCs⁵³. For gene-set enrichment analysis, gene-association with GO terms 'regulation of transcription' (GO:0006357), 'cell cycle' (GO:0007049), 'translation' (GO:0006412) and 'extracellular matrix' (GO:0031012) were derived from AmiGO⁵⁴. Transcripts were pre-ranked based on the difference half-life to the mean half-life after log₂-transformation. GSEAPreranked was performed using GSEA v.2.2.4 (refs. 55,56).

MicroRNA targets were predicted using Targetscan v7 (ref. 57). Briefly, we provided a 60-way multiple genome alignment against mm10 and our custom 3'-end annotation to create a tailored database of conserved miRNA targets. The output was then intersected with our data, filtered, and grouped according to site type. To determine site conservation, cutoffs for branch length score were set to ≥ 1.6 (7-mer-1a), ≥ 1.3 (7-mer-m8) and ≥ 0.8 (8-mer).

Relative RNA stabilities were determined by performing SLAM seq after 3 h and 12 h s^4U pulse labeling in wild-type or knockout cell lines. The background subtracted T > C conversion rates at 3 h were normalized to 12 h and relative stabilities for control (treated with non-targeting gRNA⁴⁴) and knockout cells were assessed from the following equation: $\ln(2) / \ln(1 - (T > C \text{ conversion [3h]} / T > C \text{ conversion [12h]})) / 3$.

N6-methyladenosine-targets were extracted from ref. 28 and batch coordinate conversion (liftOver) from mm9 to mm10 (UCSC) was performed, or from ref. 32. Tags in 3' UTRs were refined by overlapping the genomic coordinates with the custom mESC annotation.

Statistics. Statistical analyses (as indicated in text and figure legends) were performed in Prism v7.0a (GraphPad), Excel v15.22 (Microsoft) or R (v2.15.3 and v3.3).

Data availability statement. Sequencing data associated with this manuscript is available at GEO under the accession number [GSE99978](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99978). All main and supplementary figures have associated source data. A pipeline for extended mRNA 3' end annotation is available at github (<https://github.com/AmeresLab/UTRannotation>). The DUNK analysis pipeline for SLAM seq data analysis is available for download (<http://t-neumann.github.io/slamdunk/>).

A Life Sciences Reporting Summary is available.

36. Herzog, V.A. *et al.* Thiol-linked alkylation for the metabolic sequencing of RNA (SLAMseq). *Protoc. Exch.* <http://dx.doi.org/10.1038/protex.2017.105> (2017).
37. Spitzer, J. *et al.* PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation): a step-by-step protocol to the

- transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol.* **539**, 113–161 (2014).
38. Nilsen, T.W. RNA sequencing by primer extension. *Cold Spring Harb. Protoc.* **2013**, 1182–1185 (2013).
 39. Eling, U. *et al.* Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9**, 563–574 (2011).
 40. Mohn, F., Sienski, G., Handler, D. & Brennecke, J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* **157**, 1364–1379 (2014).
 41. Reimão-Pinto, M.M. *et al.* Uridylation of RNA hairpins by tailor confines the emergence of microRNAs in *Drosophila*. *Mol. Cell* **59**, 203–216 (2015).
 42. Hodgkins, A. *et al.* WGE: a CRISPR database for genome engineering. *Bioinformatics* **31**, 3078–3080 (2015).
 43. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
 44. Han, B.W., Hung, J.-H., Weng, Z., Zamore, P.D. & Ameres, S.L. The 3'-to-5' exonuclease Nibbler shapes the 3' ends of microRNAs bound to *Drosophila* Argonaute1. *Curr. Biol.* **21**, 1878–1887 (2011).
 45. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
 46. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
 47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
 48. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
 49. Sedlazeck, F.J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).
 50. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
 51. Quinlan, A.R. BEDTools: The swiss-army tool for genome feature analysis **27**, 11.12.1–11.12.34 (John Wiley & Sons, 2002).
 52. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W1, W160 (2016).
 53. Thomas, P.D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
 54. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).
 55. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
 56. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
 57. Friedman, R.C., Farh, K.K.-H., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

No statistical method to pre-determine sample size were used.

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization was used.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded to allocation during experiments and outcome assessment.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Please refer to the "Bioinformatics and Data analysis" (Online Methods) section for software and algorithms used in this study.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used for this study. Cell lines are available through IMBA Haplobank without any restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Please refer to "Western Blotting" in the Online Methods section for the information about antibodies used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Mouse embryonic stem cells (AN3-12 clone) was obtained from Haplobank (U. Elling; personal communication); please refer to "Cell culture" in the Online Methods section.

b. Describe the method of cell line authentication used.

Cnv, karyotyping, rna seq, morphology

c. Report whether the cell lines were tested for mycoplasma contamination.

The cell lines used in this study were tested on a regular basis (at least once a month) and were always tested negative for mycoplasma.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No cross-contaminations reported according to ICLAC for mouse embryonic stem cells (AN3-12).

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants.