Published in final edited form as:
Nat Methods. 2017 December ; 14(12): 1198–1204. doi:10.1038/nmeth.4435.

Thiol-linked alkylation of RNA to assess expression dynamics

Veronika A. Herzog[1], **Brian Reichholf**[1], **Tobias Neumann**[2], **Philipp Rescheneder**[3], **Pooja Bhat**[1], **Thomas R. Burkard**[1], **Wiebke Wlotzka**[1], **Arndt von Haeseler**[3], **Johannes Zuber**[2], and **Stefan L. Ameres**[1,*]

[1]Institute of Molecular Biotechnology (IMBA), Vienna Biocenter Campus (VBC), 1030 Vienna, Austria

[2]Research Institute of Molecular Pathology (IMP), Vienna Biocenter Campus (VBC), 1030 Vienna, Austria

[3]Center for Integrative Bioinformatics Vienna, Max F Perutz Laboratories, Medical University of Vienna, University of Vienna, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria

Abstract

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 (s^4U)-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 enables 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 scales 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 N^6-methyladenosine. SLAM-seq facilitates the dissection of fundamental mechanisms that control gene expression in an accessible, cost-effective, and scalable manner.

Introduction

The regulated expression of genetic information imperatively stipulates cellular homeostasis and environmental adaptability and its transformation can cause human diseases1.

*Correspondence to: stefan.ameres@imba.oeaw.ac.at.

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 s^4U-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 interest:
VAH, BR, and SLA declare competing financial interest. A patent application related to this work has been filed.

Europe PMC Funders Author Manuscripts

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 insights into the relative kinetics of RNAs 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 (s^4U) 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 the biochemical separation by reversible biotinylation4–10. Affinity-based RNA-purification upon s^4U-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 4-thiouridine or 4-thiouracil upon metabolic activation by uracil phosphoribosyltransferase (UPRT)4,5,9–11. However, like 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 in order to provide integrative insights into gene expression dynamics and fails to report global effects unless spike-in strategies are applied8,12. Alternative concepts for the direct identification of nucleotide-analogs by sequencing emerge from recent epitranscriptomics-technologies, but current methods are incompatible with biologically inert nucleotide-analogs (i.e. s^4U) and fail to report absolute stoichiometry13,14.

Here, we report thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq), an orthogonal chemistry approach that uncovers s^4U at single-nucleotide-resolution by reverse-transcription-dependent thymine-to-cytosine-conversions in a high-throughput sequencing-compatible manner.

Results

Detection of 4-thiouridine by sequencing

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

s^{4}U-containing oligo with identical sequence (Supplementary Fig.3b, c). To evaluate the effect of s^{4}U-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). While the presence of s^{4}U prompted a constant ten to eleven percent T>C-conversions already in the absence of alkylation (presumably due to base-pairing variations of s^{4}U-tautomeres), s^{4}U-alkylation increased T>C-conversions by 8.5-fold, resulting in a >0.94 conversion rate (Fig.1d). Importantly, iodoacetamide-treatment leaves conversion rates of any given non-thiol-containing nucleotide unaltered (Supplementary Fig.4c).

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

We subjected mouse embryonic stem cells (mESCs) to 100 μM s^{4}U-labeling, a concentration far below the EC_{50} toxicity value of s^{4}U in mESCs (Supplementary Fig.5). After metabolic RNA labeling for 24h, 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, Supplementary Fig.6)16. 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)16. Furthermore, 3′ end sequencing enables the cell-type-specific re-evaluation of UTR-annotations to conduct mRNA 3′ isoform-specific expression analysis (Supplementary Fig.7). Upon generating SLAM-seq libraries through the Quant-seq protocol from mESCs 24h after s^{4}U metabolic labeling, we observed a strong accumulation of T>C-conversions when compared to unlabeled conditions (Fig.2b). Transcriptome-wide analyses confirmed this observation (Fig.2c): In the absence of s^{4}U metabolic labeling, we observed a median rate of ≤0.1% for any given conversion, consistent with Illumina-reported sequencing error, whereas s^{4}U-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). Importantly, 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^{4}U incorporation measured by mass spectrometry in poly(A)-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 s^{4}U-pulse labeling (final conc.:100 μM s^{4}U) 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 Table1). 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): While high levels of T>C reads were recovered for the ES cell-specific transcription factor Sox2, and the inherently instable primary microRNA transcript from the miR-290-295

cluster, the house-keeping transcript Gapdh associated with fewer T>C reads, presumably because its accumulation to high steady-state expression levels is achieved by high transcript stability (Fig.3a).

Transcriptional output by Pol II is regulated by transcription factors that bind cis-acting regulatory elements known as enhancers17. 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 target genes necessary to maintain the ESC state (Supplementary Fig.10b)18. Transcriptional output measurements by SLAM-seq revealed that well-established Oct4/Sox2/Nanog-target genes produced overall a larger number of T>C-conversion-containing reads in the s^4U-pulse experiment (Fig.3b, Supplementary Fig.10b, Supplementary Table1). When inspected globally, transcripts derived from the 2029 expressed genes (>5cpm steady-state) with proximal OSN-occupancy produced significantly more T>C reads when compared to 4994 genes without proximal OSN enhancer (Mann-Whitney test, $p<10^{-4}$, Fig.3c)19. A subset of enhancers in mESCs were previously described to form arrays of regulatory elements (aka “super” or strong enhancer, SE) with unusually strong accumulation of transcriptional coactivators19,20. In fact, the 156 genes next to strong enhancers exhibited highest transcriptional output (Mann-Whitney test, $p<10^{-4}$, Fig.3c). In contrast, only genes proximal to strong enhancers associated with above average steady-state expression (Supplementary Fig.10c). We concluded that SLAM-seq provides a quantitative readout for enhancer activity in mESCs.

Together with the fact that transcriptional output significantly correlated with data derived from global nuclear run-on experiments (Supplementary Fig.10d)21, we concluded that SLAM-seq uncouples transcriptional output from stability effects to globally measure Pol II-derived transcriptional activity.

Global and transcript-specific mRNA stability in mESCs

To directly measure mRNA transcript stabilities, we subjected mESCs to s^4U metabolic RNA labeling (100 μM s^4U) for 24h, 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, while T>C-conversion-containing reads decreased over time in a transcript-specific manner (Supplementary Fig.11b,c, T>C reads). After calculating the background-subtracted, U-content- and coverage-normalized T>C-conversion rate for each transcript at every time point relative to oh 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 8405 transcripts (steady-state expression >5cpm) to single-exponential decay kinetics, we determined a median mRNA half-life of 3.9h, corresponding to a cell-cycle normalized half-life of 4.3h (Fig.4b, Supplementary Table2). These measurements fall within the range of previously determined mRNA stabilities in mammalian cells22.

Previous studies proposed a close relationship between transcript-specific mRNA half-life and its physiological function1,10. We therefore ranked the 6665 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 for the 666 most or least stable mRNAs (Fig. 4c). Transcripts with short half-life significantly enriched for regulators of Pol II-dependent transcription ($p<10^{-3}$), while stable mRNAs associated with the 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 house-keeping function tend to decay at low rates, perhaps reflecting the 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/chase experiments (Supplementary Fig.12): Transcript biogenesis rates 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 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 transcription-inhibition 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 expression23. In mESCs, they contribute to cell state maintenance and transitions by tuning the expression of ES cell transcripts and promoting their clearance during differentiation18. At the molecular level, miRNAs act as guides for ribonucleoprotein complexes that target complementary sites, usually within the 3′ UTR of mRNAs, as defined by miRNA seed sequence (nucleotides two to seven or eight of the miRNA)23. 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 contexts24. We determined the stability of miRNA targets in wild-type mESCs by inspecting half-life of transcripts harboring in their 3′ UTR target sites for the miR-291-3p/294-3p/295-3p/302-3p and miR-292a-3p/467a-5p-family (referred to as 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 2.9h (n=1450), miR-291a-family targets were significantly less stable compared to transcripts without sites ($t_{½}$=4.0h; n=5095;

KS-test, $p<10^{-15}$; Fig.5a). Transcripts with conserved sites exhibited even shorter half-life ($t_{½}$=2.6h; n=50; Fig.5a). To confirm the direct contribution of miRNAs to transcript destabilization, we determined changes in mRNA half-life by s^4U-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 to 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)23: While each site-type responded to Xpo5-depletion with a significant increase in mRNA stability (KS-test, $p<10^{-8}$), 6mer target sites exhibited the weakest effects, followed the two 7mer site types (Fig.5c). 8mer sites showed 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 upon depletion of Xpo5, is directly dependent on small RNA abundance (Supplementary Fig.13g-i), as described previously25,26.

Second, we focused on N^6-methyladenosine (m^6A), the most abundant internal modification in mammalian mRNA, implicated in the regulation of various physiological processes27,28. In ESCs, m^6A facilitates the resolution of naïve pluripotency towards differentiation29,30. At the mechanistic level, the m^6A mark impinges on various aspects of mRNA processing, including mRNA stability31 (Supplementary Fig.14a). To estimate the effect of m^6A on mRNA stability in mESCs, we first determined the general association of m^6A targets, as mapped previously by m^6A-RNA-immunoprecipitation and sequencing29, with mRNA stability in wild-type cells. With a half-life of 3.1h, m^6A-containing transcripts (n=3492) were significantly less stable compared to naïve transcripts ($t_{½}$=4.6h, n=3173, KS-test, $p<10^{-15}$, Fig.5d). N^6-methyladenosine marks do not distribute evenly within mRNAs but are enriched in long exons, near stop codons, and in 3′ untranslated regions (UTRs), although m^6A also occurs in the coding region (CDS) and 5′ UTR31 (Fig.5e top). We investigated the relationship between the position of m^6A within targeted mRNAs and its impact on RNA decay. We found that mRNAs containing m^6A exclusively in the CDS (n=545) or in the 3′ UTR (n=2093) were significantly less stable compared to naïve transcripts (KS-test, $p<10^{-15}$, Fig.5e bottom). In contrast, mRNAs that contained m^6A exclusively in the 5′ UTR (n=88) were not less stable compared to naïve transcripts (KS-test, $p>0.05$, Fig.5e). To confirm the causal contribution of m^6A to transcript destabilization we determined changes in mRNA half-life by s^4U-pulse labeling followed by SLAM-seq in mESCs depleted of Mettl3, the catalytic subunit of the m^6A RNA methylation complex, which resulted in the co-depletion of its RNA-binding partner protein Mettl14 (Supplementary Fig.14b,c)32. Consistent with a direct and position dependent impact of m^6A on mRNA decay, we observed a significant increase in relative mRNA stability for transcripts containing m^6A 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 m^6A profiling data in mESCs (Supplementary Fig.14d-e)33.

Discussion

Recent efforts in decoding RNA modifications led to the emergence of epitranscriptome sequencing technologies profiling ribonucleotide modifications on a genomic scale13,14. Here, we present an orthogonal chemistry-based sequencing strategy for the identification of 4-thiouridine (s^4U), which is widely used for in vivo, ex vivo and in vitro RNA labeling and represents a natural base modification in eubacterial and archaeal tRNA4–8,34.

Combining SLAM-seq with Quant-seq provides several advantages with important practical and conceptual implications: (1) The specific sampling of poly-adenylated RNA species assigns kinetics to functional, fully processed RNA polymerase II transcripts (Fig.2 and Supplementary Fig.6). (2) It provides access to mRNA 3′ isoform-specific expression-dynamics (Supplementary Fig.7). (3) By eliminating the requirement to normalize for transcript length, Quant-seq facilitates downstream data analysis (Supplementary Fig.6). (4) Quant-seq produces highly reproducible results from as little as 100 pg total RNA without requirement for rRNA depletion, hence provides access to cellular systems for which starting material is limiting16. (5) 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. But alternative sequencing methods may augment the applicability of SLAM-seq, because s^4U-identification by sequencing is in principle compatible with any RNA library preparation method that involves a reverse transcription step.

Studying intracellular RNA kinetics by s^4U-metabolic RNA labeling requires general and method-specific considerations to be taken into account: s^4U-incorporation was previously linked to rRNA processing defects in human cancer cells35. Because s^4U-uptake can vary between cell types, careful assessment of cell-type-specific toxicity is imperative to meet s^4U-labeling conditions that do not affect gene expression or cell viability (Supplementary Fig.5)5,36. In mESCs non-toxic concentrations of 100 μM s^4U result in a median s^4U-incorporation of 2.29% across 8408 transcripts upon long-term metabolic labeling (i.e. 24h), corresponding to one s^4U 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 s^4U-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 even in short s^4U pulse labeling conditions (Fig.3). Note, that the ability to determine de novo synthesized transcripts will depend on (1) the cellular s^4U uptake kinetics, (2) the overall transcriptional activity of the cell type and (3) the library sequencing depth. Hence, these parameters need to be taken into account when designing a SLAM-seq experiment, particularly when employing short s^4U pulse labeling, where sequencing depth demands adjustments to the given cellular parameters. In that respect, s^4U-tagging approaches (i.e. TT-seq) may provide some advantage over SLAM-seq when analyzing transient RNA species that escape detection by standard RNA sequencing approaches8.

Online Methods

A step-by-step protocol is available as a Supplementary Protocol an open resource in *Protocol Exchange*37.

Carboxyamidomethylation of s^4U

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 s^4U 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 pH8, 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 minutes. 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 x 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 s^4U 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 prior to Mass Spectrometry analysis38.

Primer extension assays

Primer extension assays were essentially performed as described previously39. Briefly, template RNA oligonucleotides (5L-*let-7*-3L or 5L-*let-7*-s^4Up9-3L; 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-^{32}P-ATP (Perkin-Elmer) and T4-polynucleotide kinase (NEB), followed by denaturing polyacrylamide gel-purification. 640 nM γ-^{32}P-RT primer was annealed to 400 nM 5L-*let-7*-3L or 5L-*let-7*-s^{4}Up9-3L in 2 x annealing buffer (500 mM KCl, 50 mM Tris pH 8.3) in a PCR machine (3 min 95°C, 30 sec 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. Upon completion, RT reactions were resuspended in formamide loading buffer (Gel loading buffer II, Thermo Fisher Scientific) and subjected to 12.5% denaturing polyacrylamide gel electrophoresis. 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 s^{4}U-containing and non-containing RNA oligonucleotides were compared for the indicated reverse transcriptases.

HPLC analysis of s^{4}U-labeled RNA

Analysis of s^{4}U-incorporation into total RNA following metabolic labeling was performed as previously described38.

Cell viability assay

5000 mESCs were seeded per 96 well the day before the experiment. After onset of the experiment, media containing the indicated concentration of s^{4}U 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

Mouse embryonic stem (mES) cells (clone AN3-12), derived from C57BL/6x129 F1 females, were obtained from IMBA Haplobank (U. Elling et al., accepted for publication in Nature) and cultured in 15 % FBS (Gibco), 1x Penicillin-Streptomycin solution (100 U/ml Penicillin, 0.1 mg/ml Streptomycin, Sigma), 2 mM L-Glutamine (Sigma), 1x 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% CO_2 and passaged every second day. Cell doubling time of AN3-12 mES in presence of s^{4}U 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 population40.

SLAM-seq in mESCs

See *Protocol Exchange* for detailed information regarding SLAM-seq37. mESCs were seeded the day before the experiment at a density of 10^5 cells/ml. s^{4}U-metabolic labeling in

mESCs was performed by incubating mESCs in standard medium but adding s^4U (Sigma) to a final concentration of 100 μM and media exchange every 3 hours for the duration of the pulse. For the uridine chase experiment, cells were washed twice with 1x 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 0.1mM DTT (final conc.) 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 described41. 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 before42, 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 x 10^5 AN3-12 mESCs were seeded per 35 mm plate and grown over night. To block transcription, actinomycin D (Sigma) was added to the medium at the concentration of 5 μg/ml. Cells were harvested at 0, 0.25h, 0.5h, 1h, 3h 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 Editing43. gRNA oligonucleotides (see Supplementary Table 3) were cloned into pLenti-CRISPR-v2-GFP vector as described44, but modified by replacing the puromycin resistance cassette with GFP. Prior to gRNA transfection targeting Xpo5 or Mettl3, wildtype An3-12 mESCs were FACS sorted for haploid cells as described previously40. 3 x 10^5 cells were seeded per 6 well and transfected the next day with 3 μg pLenti-CRISPR-v2-GFP using Lipofectamine 2000 as recommended by the manufacturer. 48h after transfection, GFP positive cells were sorted by fluorescence-activated cell sorting (FACS) and 1500 cells were subsequently seeded per 15 cm plate. Single colonies were picked after 10 days. 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-Mettl3 (15073-1-AP, Proteintech, rabbit), 1:5,000 for anti-Mettl14 (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 (BioRad) or by Amersham Hyperfilm ECL (GE Healthcare).

Northern Blotting

Northern hybridization experiments were performed as described previously45. 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.1i 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 adapter sequence) and filter for sequences of 21nt length. Resulting sequences were aligned to aligned to mature dme-let-7 sequence (TGAGGTAGTAGGTTGTATAGT) using bowtie v0.12.9 allowing for 3 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, fraction 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 1 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 v2.5.2b46. 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. 2-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 kb, 20kb, 30kb 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 FeatureCounts47.

For Cap analysis gene expression (Cap-Seq), barcoded libraries were demultiplexed using Picard Tools BamIndexDecoder v1.13 allowing 1 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)48 against known rRNA sequences (RefSeq). The rRNA subtracted reads were aligned with TopHat (v1.4.1)49 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) datasets, reads were demultiplexed using Picard Tools BamIndexDecoder v1.13 allowing 1 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 NextGenMap50 developed to recover T>C-conversions from SLAM-seq data sets (Neumann T., *et al.*, in preparation). Briefly, adapter-clipped reads were trimmed 12 bp from the 5′ end (-5 12) and poly(A) 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 datasets) 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 10x and a variant fraction cutoff of 0.8 were called using VarScan2.4.1 using default parameters51. 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 datasets) 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 datasets (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 nts long 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 200nt bins separated by 20nts starting from the last 200nts 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 250nt 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_lncRNA, lincRNA, macro_lncRNA, 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 bedtools52. Stranded coverage tracks for Quant-seq and RNA-seq samples were created using deeptools' *bamCoverage* command53, 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 s^4U labeling) were subtracted from the T>C reads in the 45min 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 according to Whyte et al.19.

GRO-seq data from mESCs was downloaded from GEO (GSE27037)21. Reads were mapped to mm10 using bowtie allowing for uniquely mapping reads with at most 2 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 featureCounts47 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 datasets and detected in GRO-seq datasets.

To calculate RNA half-lives, T>C-conversions were background-subtracted (no s^4U 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 > 24h were set to 24h. If not stated otherwise an R^2 cutoff of > 0.6 was applied. To calculate RNA half-lives normalized to cell cycle length, T>C-conversions were multiplied by $2^{(\text{timepoint}/14.7\text{h})}$.

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 > 5cpm 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 > 5cpm in mESCs (n=8533)54. 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 AmiGO55.Transcripts were pre-ranked based on the difference half-life to the mean half-life after $\log_2$-transformation. GSEAPreranked was performed using GSEA v.2.2.456,57.

MicroRNA targets were predicted using Targetscan v758. 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 by site type. To determine site conservation, cutoffs for branch length score were set to ≥ 1.6 (“7mer-1a”), ≥ 1.3 (“7mer-m8”) and ≥ 0.8 (“8mer”).

Relative RNA stabilities were determined by performing SLAM-seq after 3h and 12h s^4U pulse labelling in wildtype or knock-out cell lines. The background subtracted T>C conversion rates at 3h were normalized to 12h 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 conversion [3h] / T>C conversion [12h]))/3.

N6-methyladenosine-targets were extracted from Batista et al., 201429 and batch coordinate conversion (liftOver) from mm9 to mm10 (UCSC) was performed, or from Ke et al., 201733. Tags in 3′ UTRs were refined by overlapping the genomic coordinates with the custom mES cell 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

Sequencing data associated with this manuscript is available at GEO under the accession number GSE99978. All main and supplementary figures have associated with 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/).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Julian Jude for generously providing a modified version of pLenti-CRISPR-v2-GFP, Gabriela Krssakova for HPLC analysis, the IMP/IMBA Biooptics facility for FACS support, and all laboratory members for support and discussions. Mass spectrometry was performed at the VBCF Metabolomics unit (www.vbcf.ac.at), funded by the City of Vienna through the Vienna Business Agency. HTP sequencing was performed at the VBCF NGS Unit (www.vbcf.ac.at). This work was supported by 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.

References

1. Schwanhäusser B, et al. Global quantification of mammalian gene expression control. Nature. 2011; 473:337–342. [PubMed: 21593866]
2. Rabani M, et al. High-Resolution Sequencing and Modeling Identifies Distinct Dynamic RNA Regulatory Strategies. Cell. 2014; 159:1698–1710. [PubMed: 25497548]
3. Yao SYM, Ng AML, Cass CE, Baldwin SA, Young JD. Nucleobase transport by human equilibrative nucleoside transporter 1 (hENT1). J Biol Chem. 2011; 286:32552–32562. [PubMed: 21795683]
4. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat Biotechnol. 2005; 23:232–237. [PubMed: 15685165]
5. Dölken L, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA. 2008; 14:1959–1972. [PubMed: 18658122]
6. Miller C, et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol. 2011; 7:458–458. [PubMed: 21206491]
7. Duffy EE, et al. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Mol Cell. 2015; 59:858–866. [PubMed: 26340425]
8. Schwalb B, et al. TT-seq maps the human transient transcriptome. Science. 2016; 352:1225–1228. [PubMed: 27257258]
9. Miller MR, Robinson KJ, Cleary MD, Doe CQ. TU-tagging: cell type–specific RNA isolation from intact complex tissues. Nat Methods. 2009; 6:439–441. [PubMed: 19430475]
10. Rabani M, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat Biotechnol. 2011; 29:436–442. [PubMed: 21516085]
11. Gay L, et al. Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes Dev. 2013; 27:98–115. [PubMed: 23307870]
12. Sun M, et al. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. 2012; 22:1350–1359. [PubMed: 22466169]
13. Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods. 2016; 14:23–31. [PubMed: 28032622]
14. Heiss M, Kellner S. Detection of nucleic acid modifications by chemical reagents. RNA Biol. 2016; 0:1–9.
15. Shugart L. Effect of selective chemical modification of 4-thiouridine of phenylalanine transfer ribonucleic acid on enzyme recognition. Arch Biochem Biophys. 1972; 148:488–495. [PubMed: 4553420]
16. Moll P, Ante M, Seitz A, Reda T. QuantSeq 3′ mRNA sequencing for RNA quantification. Nat Methods. 2014; 11
17. Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 2011; 12:283–293. [PubMed: 21358745]
18. Young RA. Control of the Embryonic Stem Cell State. Cell. 2011; 144:940–954. [PubMed: 21414485]

19. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]
20. Pott S, Lieb JD. What are super-enhancers? Nature Genetics. 2015; 47:8–12. [PubMed: 25547603]
21. Min IM, et al. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 2011; 25:742–754. [PubMed: 21460038]
22. 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. 2012; 9:1233–1238. [PubMed: 23034600]
23. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136:215–233. [PubMed: 19167326]
24. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet. 2011; 12:99–110. [PubMed: 21245828]
25. Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol. 2013; 14:475–488. [PubMed: 23800994]
26. Mullokandov G, et al. High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. Nat Methods. 2012; 9:840–846. [PubMed: 22751203]
27. Meyer KD, Jaffrey SR. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol. 2014; 15:313–326. [PubMed: 24713629]
28. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev. 2015; 29:1343–1355. [PubMed: 26159994]
29. Batista PJ, et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell. 2014; 15:707–719. [PubMed: 25456834]
30. Geula S, et al. Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science. 2015; 347:1002–1006. [PubMed: 25569111]
31. Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m^6A RNA methylation. Nat Rev Genet. 2014; 15:293–306. [PubMed: 24662220]
32. Wang X, et al. Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. Nature. 2016; 534:575–578. [PubMed: 27281194]
33. Ke S, et al. m 6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. 2017; 31:990–1006. [PubMed: 28637692]
34. Limbach PA, Crain PF, McCloskey JA. Summary: the modified nucleosides of RNA. Nucleic Acids Research. 1994; 22:2183–2196. [PubMed: 7518580]
35. Burger K, et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. 2013; 10:1623–1630. [PubMed: 24025460]
36. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010; 141:129–141. [PubMed: 20371350]
37. Herzog VA, et al. Thiol-linked alkylation for the metabolic sequencing of RNA (SLAMseq). Protoc Exch. 2017; doi: 10.1038/protex.2017.105
38. 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. Meth Enzymol. 2014; 539:113–161. [PubMed: 24581442]
39. Nilsen TW. RNA sequencing by primer extension. Cold Spring Harb Protoc. 2013; 2013:1182–1185. [PubMed: 24298029]
40. Elling U, et al. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. Cell Stem Cell. 2011; 9:563–574. [PubMed: 22136931]
41. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell. 2014; 157:1364–1379. [PubMed: 24906153]
42. Reimão-Pinto MM, et al. Uridylation of RNA Hairpins by Tailor Confines the Emergence of MicroRNAs in Drosophila. Mol Cell. 2015; 59:203–216. [PubMed: 26145176]
43. Hodgkins A, et al. WGE: a CRISPR database for genome engineering. Bioinformatics. 2015; 31:3078–3080. [PubMed: 25979474]

44. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014; 11:783–784. [PubMed: 25075903]
45. Han BW, Hung J-H, Weng Z, Zamore PD, Ameres SL. The 3'-to-5' Exoribonuclease Nibbler Shapes the 3' Ends of MicroRNAs Bound to Drosophila Argonaute1. Curr Biol. 2011; 21:1878–1887. [PubMed: 22055293]
46. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. [PubMed: 23104886]
47. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–930. [PubMed: 24227677]
48. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
49. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111. [PubMed: 19289445]
50. Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics. 2013; 29:2790–2791. [PubMed: 23975764]
51. Koboldt DC, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568–576. [PubMed: 22300766]
52. Quinlan, AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Vol. 27. John Wiley & Sons, Inc; 2002. p. 11.12.1-11.12.34.
53. Ramírez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research. 2016; 44:W160–5. [PubMed: 27079975]
54. Thomas PD, et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 2003; 13:2129–2141. [PubMed: 12952881]
55. Carbon S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009; 25:288–289. [PubMed: 19033274]
56. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–15550. [PubMed: 16199517]
57. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics. 2003; 34:267–273. [PubMed: 12808457]
58. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009; 19:92–105. [PubMed: 18955434]

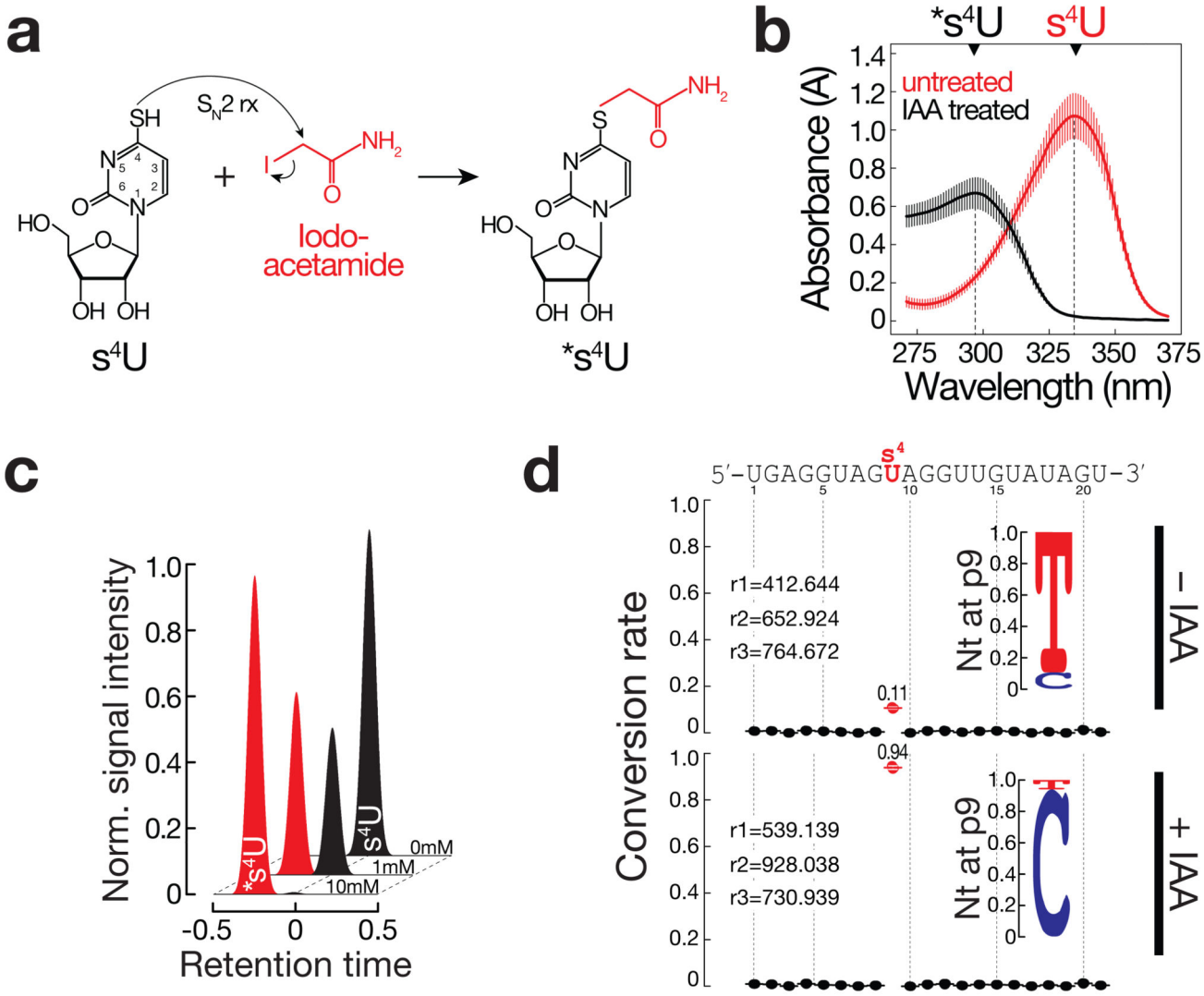

Figure 1. Detection of 4-thiouridine (s^4U) by chemical derivatization and sequencing.

(a) 4-thiouridine (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^4U) before and after treatment with iodoacetamide (IAA). Absorption maxima of educt (4-thiouracil; s^4U; $\lambda_{max} \approx 335$ nm) and product (carboxyamido-methylated 4-thiouracil; $*s^4U$; $\lambda_{max} \approx 297$ nm) are indicated. Data represents mean (center line) ± SD (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 iodoacetamide (IAA) treatment. Average conversion rates (center line) ± SD (whiskers) of three independent experiments (points) are shown. Number of sequenced reads in each replicate (r1-r3) are indicated. Nucleotide identity at s^4U site (p9) is shown.

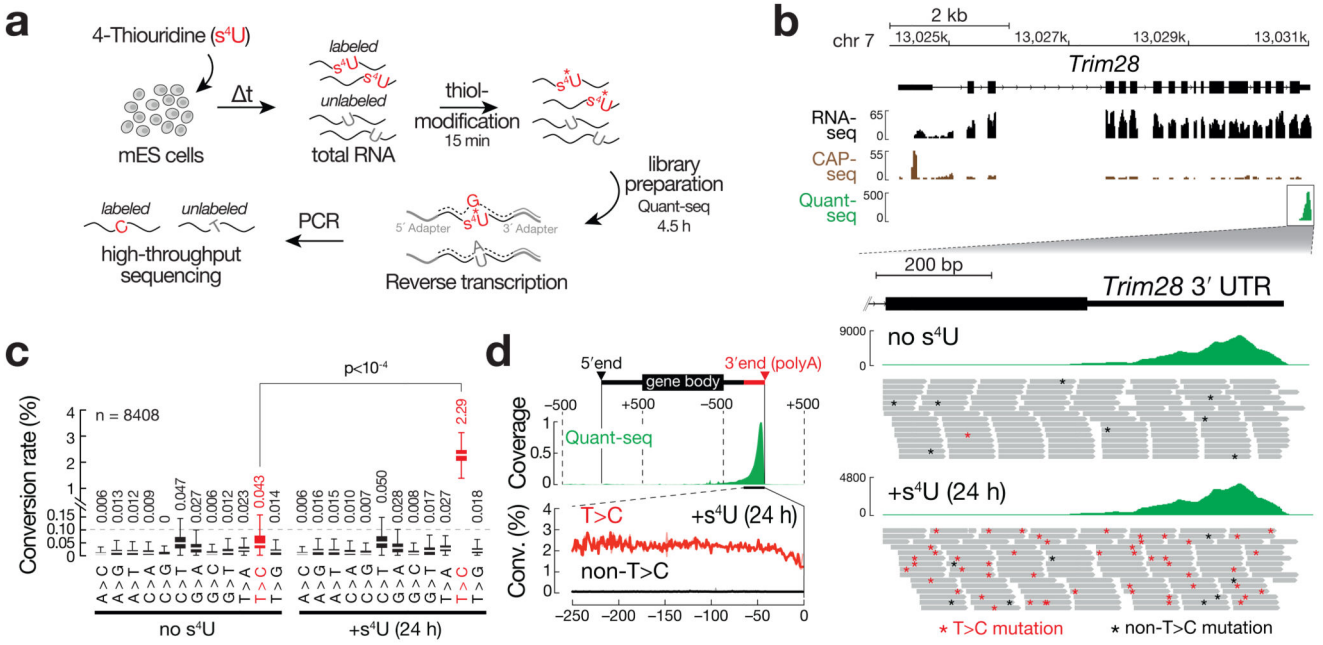

Figure 2. Thiol-linked alkylation for the metabolic sequencing of RNA (SLAM-seq).

(a) Workflow of SLAM-seq. Working time 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 panel), Cap-seq (middle panel) and mRNA 3′ end sequencing (bottom panel; RPM, reads per million). A representative area in the mouse genome encoding the gene Trim 28 is shown. Bottom shows zoom into 3′ UTR of Trim28. Unnormalized coverage plots of Quant-seq libraries prepared from untreated mESCs or mESCs subjected to s^4U-metabolic labeling using 100 μM s^4U for 24 h followed by SLAM-seq. A random subset of individual reads underlying the coverage plots are 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 s^4U) and after metabolic labeling for 24 h using 100 μM s^4U (+s^4U). 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 8408 transcripts in Quant-seq datasets. T>C conversion rate (Conv.) distributes evenly within Quant-seq-covered areas across 8408 counting windows.

Europe PMC Funders Author Manuscripts

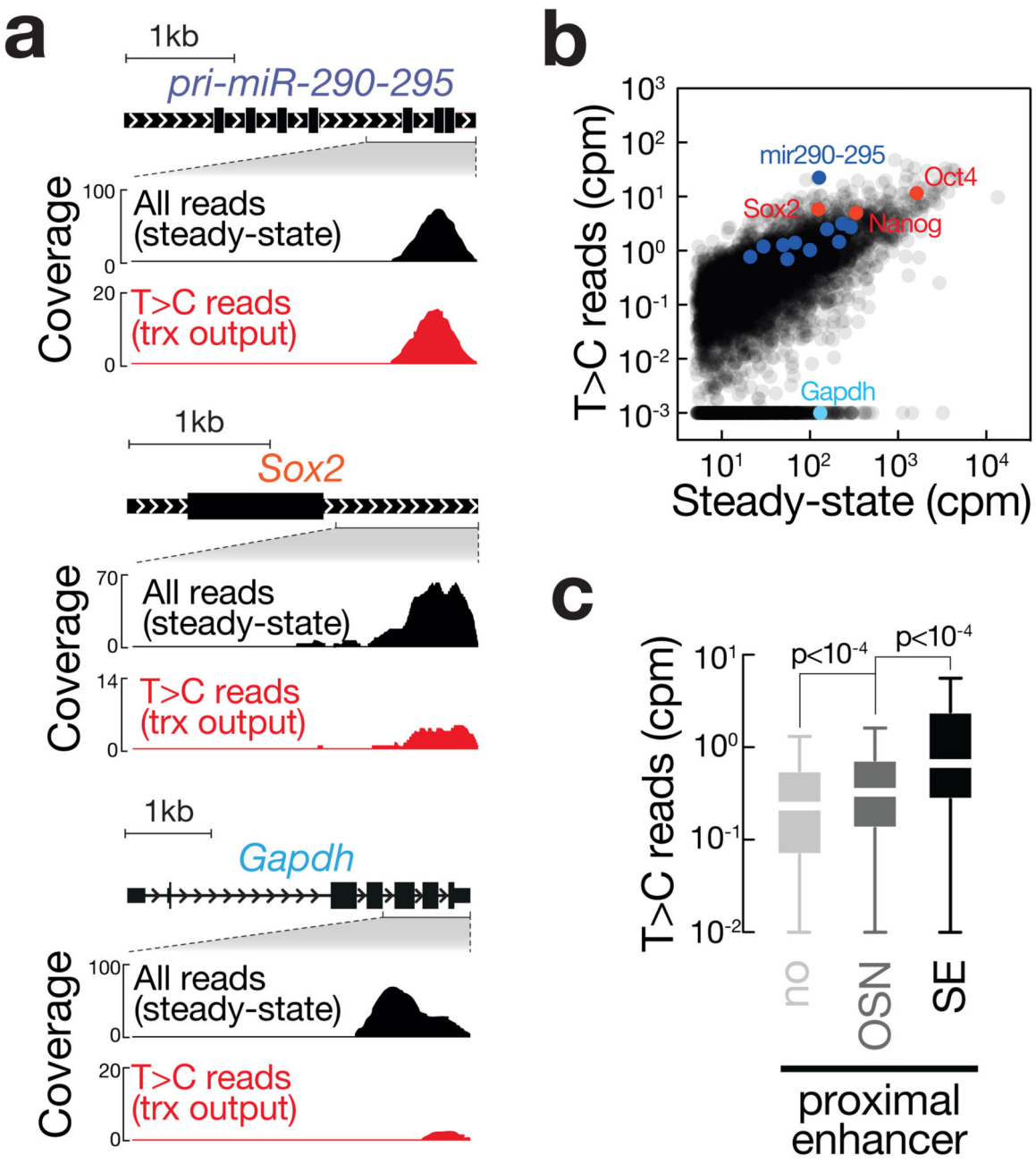

Figure 3. Quantitative description of the polyadenylated transcriptional output in mESCs.

(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 7179 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 Oct4/Sox2/

Nanog (OSN) in dark blue and a gene with house-keeping function in light-blue. **(c)** Transcriptional output, as measured in number of T>C conversion containing reads, for expressed genes (steady-state >5cpm) without adjacent OSN enhancer (no, n=4994), proximal to canonical Oct4/Sox2/Nanog enhancer (OSN, n=2029) or proximal to strong enhancers (SE, n=156). Data is represented by Tukey-boxplots without outliers. P-values determined by Mann-Whitney test are indicated.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

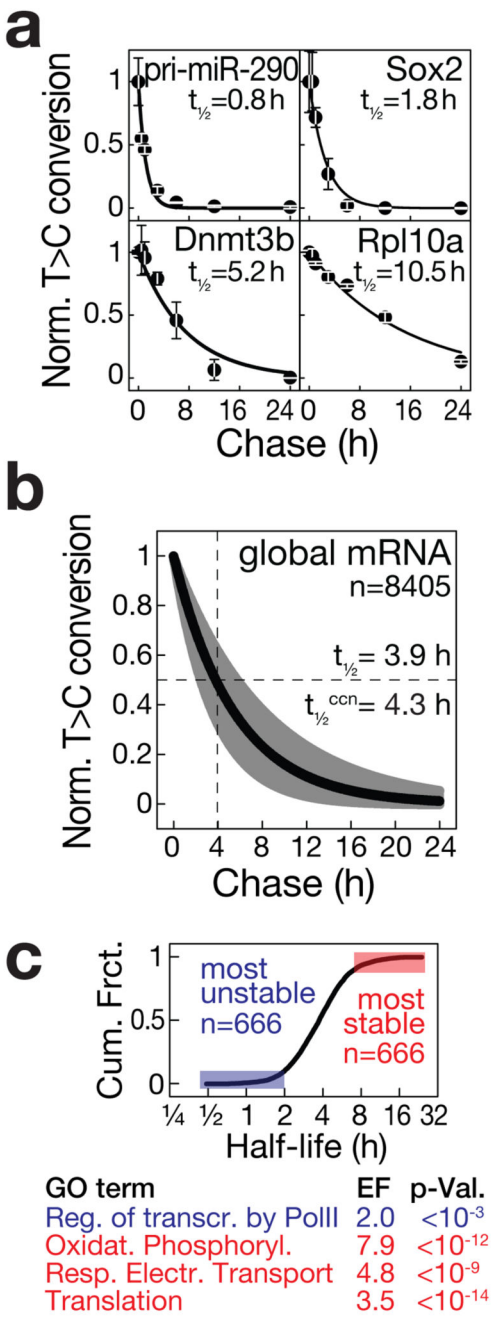

Figure 4. Global and transcript-specific mRNA stability in mESCs.

(a) Transcript stability for the indicated example genes as determined by SLAM-seq. T>C-conversion rates were determined for each timepoint of the s^4U-pulse/chase and fit to a single-exponential decay model to derive half-life ($t_{½}$). Values are mean ± SD of three independent cell cultures. **(b)** Global analysis of mRNA stability in mESCs. RNA half-life for 8405 transcripts in mESCs determined as described in (a). Single-exponential fit of median and interquartile range are shown. Median half-life before ($t_{½}$) or after ($t_{½}^{ccn}$) normalization to cell divisions. **(c)** Cumulative distribution of ranked high-confidence

transcript stabilities for 6665 transcripts. Enriched gene ontology (GO) terms for the 666 most unstable (blue) or most stable (red) are indicated. Enrichment factor (EF) and p-values (p-Val.) are indicated.

Europe PMC Funders Author Manuscripts

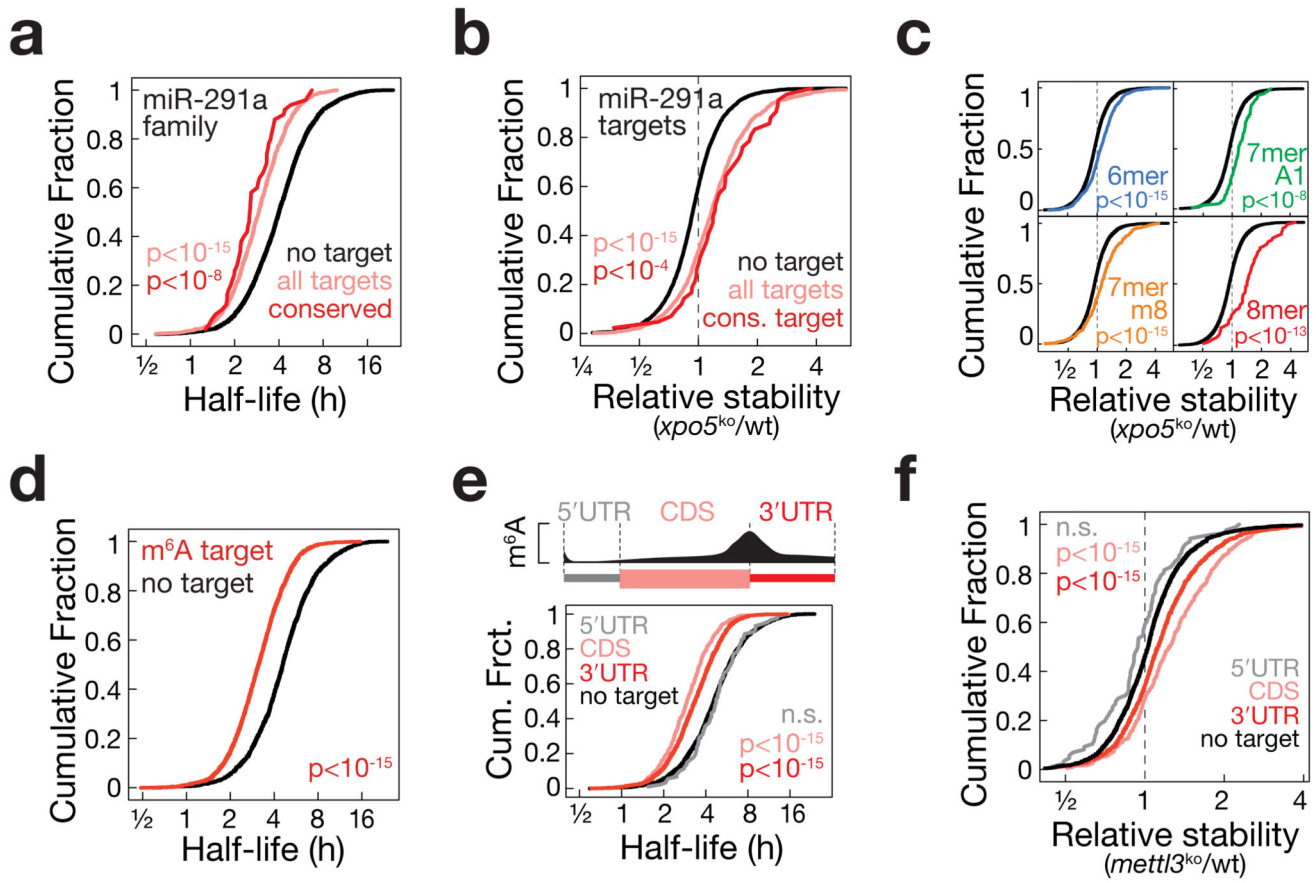

Figure 5. Molecular determinants of mRNA stability in mESCs.

(a) Cumulative distribution of ranked mRNA stabilities. Plotted are distributions for transcripts that do (rose, n=1450) or do not (black, n=5095) 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 as defined in Supplementary Fig.13g. P-value was determined by KS-test. **(b)** Cumulative distribution of mRNA stability changes in *xpo5*^{ko} relative to wt mESCs. Plotted are distributions for transcripts that do (rose, n=1288) or do not (black, n=4825) contain at least one miR-291a target site or that contain at least one conserved miR-291a target site (red, n=42). P-value was determined by KS-test. **(c)** Cumulative distribution of mRNA stability changes in *xpo5*^{ko} relative to wt mESCs. Plotted are distributions for transcripts that contain exclusively one 6mer (blue, n=493), 7mer-A1 (green, n=95), 7mer-m8 (yellow, n=325), or 8mer site (red, n=63). Black shows transcripts without any miR-291a target site (n=4825). P-value was determined by KS-test. **(d)** Cumulative distribution of ranked mRNA stabilities. Plotted are distributions for transcripts that do (red, n=3492) or do not (black, n=3173) contain the m^{6}A mark, as previously mapped by m^{6}A-RIP-seq29. P-value was determined by KS-test. **(e)** Top: Schematic distribution of m^{6}A within mRNA (adapted from31). Bottom: Cumulative distribution of ranked mRNA stabilities. Plotted are distributions for transcripts that do not (black, n=3173) or do contain m^{6}A exclusively in the 5′ UTR (grey, n=88), the coding sequence (CDS, green, n= 545) or the 3′ UTR (red, n=2093). P-value was determined by KS-test. **(f)**

Europe PMC Funders Author Manuscripts

Cumulative distribution of mRNA stability changes in *mettl3*^{ko} relative to wt mESCs. Plotted are distributions for transcripts that do not (black, n=3118) or do m^{6}A exclusively in the in the 5′ UTR (grey, n=86), the coding sequence (CDS, green, n= 518) or the 3′ UTR (red, n=2017). P-value was determined by KS-test.

Europe PMC Funders Author Manuscripts