# The m(1)A landscape on cytosolic and mitochondrial mRNA at single-base resolution

Safra, Modi; Sas-Chen, Aldema; Nir, Roni



Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmanith, W., Stern-Ginossar, N., & Schwartz, S. (2017). The m(1)A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature, 551(7679), 251–255. https://doi.org/10.1038/nature24456
Document Version: Accepted

Published Version: https://doi.org/10.1038/nature24456

# The m(1)A landscape on cytosolic and mitochondrial mRNA at single-base resolution

# The m$^1$A landscape on cytosolic and mitochondrial mRNA at single base resolution

Modi Safra[1], Aldema Sas-Chen[1], Ronit Nir[1], Roni Winkler[1], Aharon Nachshon[1], Dan Bar-Yaacov[1], Matthias Erlacher[2], Walter Rossmanith[3], Noam Stern-Ginossar[1], Schraga Schwartz[1,#]

[1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

[2]Division of Genomics and RNomics, Biocenter Innsbruck, Medical University of Innsbruck, Innrain 80/82, Innsbruck, 6020, Austria.

[3] Center for Anatomy & Cell Biology, Medical University of Vienna, Vienna 1090, Austria

#Corresponding author. Email: schwartz@weizmann.ac.il

Modifications on mRNA offer the potential of regulating mRNA fate post-transcriptionally. Recent studies have suggested the widespread presence of N1-methyladenosine ($m^1A$), a modification disrupting Watson-Crick basepairing, at internal sites of mRNAs[1,2]. These studies lacked the resolution of identifying individual modified bases, and did not identify specific sequence motifs undergoing the modification or an enzymatic machinery catalyzing them, rendering it challenging to validate and functionally characterize putative sites. Here we develop an approach allowing the semi-quantitative mapping of $m^1A$ at single nucleotide resolution and in a transcriptome-wide manner. Within the cytosol, $m^1A$ is present in a low number of mRNAs, typically at low stoichiometries, and almost invariably in tRNA T-loop-like structures, where it is introduced by the TRMT6/TRMT61A complex. In contrast, we find a single $m^1A$ site in the mitochondrial ND5 mRNA, catalyzed by TRMT10C, with methylation levels that are highly tissue specific and tightly developmentally controlled. We show that $M^1A$ leads to translational repression, likely via a mechanism involving ribosomal scanning or translation. Our findings suggest that $m^1A$ on mRNA, likely due to its disruptive impact on base pairing, leads to translational repression, and is generally avoided by cells, while revealing one case in mitochondria where tight spatiotemporal control over $m^1A$ levels was adopted as a potential means of post-transcriptional regulation.

We developed a protocol for mapping $m^1A$ at single nucleotide resolution ($m^1A$-seq), relying on the property of $m^1A$ to lead to typical misincorporation and truncation profiles upon reverse transcription[3]. Key steps in the protocol include: (1) antibody-based enrichment of $m^1A$-comprising mRNA fragments[1,2], (2, optional step) Dimroth rearrangement, by which $m^1A$ sites are converted to N6-methyladenosine (m6A) [2,4], which eliminates the $m^1A$ misincorporation and truncation patterns, (3) Reverse transcription using either a highly processive reverse transcriptase (TGIRT) with a greater tendency to lead to misincorporations at $m^1A$ sites [5,6], or a less processive one (SuperScript III (SS)) tending to lead to premature truncations (**Fig. 1a, Note S1)**.

We applied $m^1A$-seq-SS and $m^1A$-seq-TGIRT to RNA derived from HEK293T cells. Each dataset included samples sequenced either directly ('Input'), or following $m^1A$-IP ('IP'), or following both $m^1A$-IP and Dimroth rearrangement ('IP+Dimroth'). We enhanced these two datasets with 8 input and IP sample pairs from Ref. [2], and 14 sample pairs from Ref. [1]. These experimental datasets collectively comprised an unprecedented depth of >2 billion reads **(Table S1).** We developed a single, common analytical pipeline to identify $m^1A$-specific misincorporation profiles, relying on statistical tests assessing whether misincorporation rates are significantly higher in (1) IP samples, compared to input, or (2) IP compared to IP+Dimroth . In addition, we required putatively modified sites to have at least two distinct

types of misincorporation (e.g. A->T and A->G), and employed the RT truncation rate as an optional filter **(Fig. S1a, Methods)**. Collectively, we identified 205 high-confidence putative $m^1A$ sites, and 72 additional lower-ranking ones **(Fig. 1b-e, Table S2)**. Quantifications of RT-truncation and misincorporation levels were highly reproducible among replicates **(Fig. S1b-d).** Each site was detected, on average, in ~11 independent *samples* (interquartile range: 8-16), and 115 sites were independently identified in ≥2 *datasets* **(Fig 1b-c)**. As expected, use of TGIRT-based $m^1A$-seq led to higher misincorporation and lower truncation rates at the detected sites compared to the SuperScript based libraries **(Fig. S1e-f)**; IPs highly enriched for misincorporation rates, and Dimroth rearrangement reduced them **(Fig. S1g-h)**. We established our approach as highly sensitive and specific, as it allowed de-novo discovery of almost all known classes of $m^1A$ sites in cytosolic tRNAs and rRNA, with very few false positives (**Note S2a**). We establish misincorporation rates as a quantitative, *relative* estimate of $m^1A$ levels, providing a lower boundary on $m^1A$ stoichiometry **(Fig. S1i, Note S2b)**.

Surprisingly, our dataset comprised predominantly well-established sites in rRNA and tRNA **(Fig. 1c)**, and only 15 sites in mRNA and lncRNAs, 10 of which in cytosolic transcripts and 5 in mitochondria **(Fig. 1d)**. Most sites had very low misincorporation rates (<2.5%) in the absence of antibody mediated enrichment **(Fig. 1d)**, suggesting that they are modified at low stoichiometries. Two notable exceptions, both previously shown to harbor RNA:DNA sequencing differences of unknown nature, were in **(1)** tRNA like mascRNA (MALAT1-associated small cytoplasmic RNA)[7], and **(2)** mitochondrially encoded ND5 transcript [8,9] **(Note S3)**. Misincorporation levels at the detected mRNA sites in the cytosol across thousands of samples from >50 tissues in the GTEx collection[10] similarly revealed very low misincorporation levels, but high levels in rRNAs and in mascRNA serving as controls (**Fig. S2, Note S4**), suggesting that the low $m^1A$ levels are not specific only to tumorous cell lines. We further exclude that the low number of sites in mRNAs is due to decreased detection power in more lowly expressed mRNAs **(Note S5, Fig. S3a-b)**. In addition, we observe the previously described 5' biased distribution of $m^1A$-seq peaks[1,2] **(Fig. S4a-d, Table S3)**, but find no $m^1A$-specific misincorporation profiles within these regions **(Note S5)**. It remains to be established whether these enrichments originate from complex modification patterns at the first transcribed nucleotide, or are experimental artefacts. Thus, with few exceptions, $m^1A$ is rarely observed at internal sites on mRNA, and typically at low stoichiometry.

All sites in cytosolic mRNAs and in Malat1 mascRNA comprised a single motif, or slight deviations thereof, consisting of the sequence GUUCN**A**NNC (underlined A = $m^1A$) within a strong hairpin structure typically comprising a 5 bp stem and a 7 bp loop **(Fig. 2a; see also Note S6, Fig. S5)** . This sequence and structural motif is identical to the T loop of tRNAs, where $m^1A$ at position 58 is catalyzed

by the TRMT6/TRMT61A complex at precisely the same relative position [11,12]. Consistently, TRMT6/TRMT61A knockdown resulted in elimination of $m^1A$ from T-loop harboring mRNAs **(Fig. 2b)**. Conversely, overexpression of TRMT6/TRMT61A followed by $m^1A$-seq-TGIRT dramatically increased misincorporation rates at the detected sites **(Fig. 2c)**, and led to accumulation of $m^1A$ at 384 cytosolic mRNA and lncRNA sites **(Fig. 2d, Table S4)**, a massive enrichment with respect to the 10 originally identified sites**.** The GUUCNANNC motif was highly enriched at these sites (present in 193/384 sites) **(Fig. 2e)**, which were further enriched for a stable T-loop like structure **(Fig. 2f)**, typically consisting of a 7 bp loop **(Fig. 2g)** and a 6-7 bp stem **(Fig. 2h)**. A subset of the peaks could be further validated by seeking $m^1A$ sites that had converted to m6A following Dimroth treatment of mRNA extracted from TRMT6/TRMT61A overexpressing cells **(Fig. S6)**. To directly explore the determinants of specificity of TRMT6/TRMT61A, we utilized a plasmid library comprising thousands of T-loop sequences and systematically mutated counterparts[13], all cloned as 3' UTR elements downstream of a reporter. Co-transfection of this plasmid pool with TRMT6/TRM61A into HEK293T cells allowed reconstituting the $m^1A$ misincorporation signal precisely at the predicted position **(Fig. 2i).** Systematic point-mutation of each position in the GUUCNANNC motif allowed functionally reconstructing the consensus required for modification via the TRMT6/TRMT61A **(Fig. 2j),** highlighting requirements for G-C base-pairing at positions -5 and +3 , and the requirements for a pyrimidine, a cytosine, and a purine at positions -3, -2 and -1, respectively. Systematic structural mutants and compensatory mutations demonstrated direct dependency of misincorporation on stem stability **(Fig. 2k)**. These analyses demonstrate that within the cytosol, $m^1A$ is catalyzed at T-loop like elements via TRMT6/TRMT61A.

We next focused on the site in the mitochondrially encoded ND5 gene, harboring the highest levels of modification in our dataset, with misincorporation rates of ~25% in poly(A) mRNA from HEK293T cells **(Note S8)**. The absence of misincorporations in reads from the transcribed antisense strand **(Fig. S7a)** and in DNA **(Fig. S7b)** ruled out DNA heteroplasmy as their source. Misincorporation levels in ND5 RNA across thousands of GTEx samples were highly tissue specific: essentially absent in muscle and in heart but ~30% (median) in ovary and pituitary gland **(Fig. 3a).** Targeted sequencing of the ND5 locus in human muscle and ovary samples confirmed these findings **(Fig. 3b)**. The relatively high $m^1A$ levels observed in the human ovary samples prompted an exploration of ND5 methylation in development. Strikingly, single-cell RNA-seq (scRNA-seq) from 1,529 individual cells[14] revealed misincorporation levels >75% at the 8 cell stage **(Fig. 3c)**, roughly equivalent to those observed in 16S rRNA, which was previously shown to be methylated at nearly 100% stoichiometry[15]. Levels of misincorporation decreased with developmental progression, and by day 7 reached ~12.5% **(Fig. 3c)**. scRNA-seq analysis of 124 single cells, spanning a wider developmental range [16] extended these findings, and revealed that from the

metaphase II oocyte to the 4-cell embryo, methylation levels at ND5 are likely close to 100% **(Fig. 3d, Note S9)** followed by a precipitous decrease **(Fig. 3d)**. As zygotic mitochondrial transcription (ZMT) begins around the 8-cell developmental stage [17], we speculated that $m^1A$ might mark particularly stable maternal transcripts that persist up to ZMT. Indeed, transcriptional arrest in HEK293T cells using actinomycin D or ethidium bromide led to a ~4-fold increase in ND5 misincorporation levels (**Fig. 3e and S7c**, respectively), confirming the association between ND5 stability and methylation. Thus, $m^1A$ in ND5 is highly tissue and developmentally specific, and may serve as a mark of a stable subset of ND5 transcripts, that are maternally inherited and dominate until ZMT at roughly the 8 cell stage.

We hypothesized that $m^1A$ in ND5 is catalyzed by TRMT10C, which catalyzes methylation at position 9 of mitochondrial tRNAs [18]. Indeed, TRMT10C knockdown led to almost complete abolishment of methylation at ND5:1374; TRMT10C overexpression resulted in a 50% increase in methylation levels **(Fig. 3f)**. Analysis of the GTEx data revealed that ND5 methylation levels are under genetic control, as they are strongly correlated across different tissues from the same individual **(Fig. S7d-e)**. Detailed analysis revealed a relatively common SNP (G13708A) two bases upstream of the ND5 site (at 13710), severely reducing the ability of ND5 to undergo methylation **(Fig. 3g)**. Targeted sequencing of the ND5 locus in lymphoblastoid cell lines from individuals harboring this SNP compared to controls confirmed these results **(Fig. S7f)**. G13708A is among the defining SNPs of the Eurasian J haplogroup, and is thought to affect the clinical expression of Leber's hereditary optic neuropathy (LHON) in western Eurasians [19–22]. Our results indicate that this haplotype has lost the ability of undergoing efficient ND5 methylation.

We speculated that the Watson-Crick disruptive nature of $m^1A$ would prevent effective translation of modified codons. Polysome fractionation experiments revealed a substantial and highly significant reduction in misincorporation levels in ND5 in the heavier fractions relative to lighter ones **(Fig. 4a, S8a)**, suggesting repressed translation of $m^1A$ harboring transcripts. Consistently, upon overexpression of TRMT6/TRMT61A, we found a striking and highly significant depletion of $m^1A$-modified mRNA in the ribosome-heavy fractions relative to the ribosome-poor fraction **(Fig. 4b)**. This was observed for cytosolic sites that were present either in the 5' UTR or within the CDS **,** but not for a site present in the 3' UTR **(Fig. 4b)**. We next cloned a 60-bp region harboring the $m^1A$ sites in the PRUNE gene in frame and upstream of a firefly luciferase coding region **(Fig. 4c)**. Co-transfection of this construct with TRMT6/TRM61A led to high $m^1A$ levels **(Fig. 4d)**, which were eliminated upon disrupting the sequence or structure, but restored via a compensatory structural mutation **(Fig. 4c-d).** The point-mutation of the $m^1A$ site in PRUNE or structural disruption led to ~2-fold increased luciferase levels in comparison to the

sequences harboring intact consensus sequences and T-loop structures **(Fig. 4e)**. Conversely, no decrease - and even a slight increase (P=0.03) - was observed when introducing the WT element into the 3' UTR upon overexpression of TRMT6/TRMT61A in comparison to controls **(Fig. S8b-d)**. The consistent translational repression associated with $m^1A$ sites within the 5' UTR or CDS, but not in the 3' UTR, suggest that it may be dependent on ribosomal scanning or translation **(Note S10, Fig. S9)**.

Collectively, the ability to map and quantify $m^1A$ at single nucleotide resolution allowed redefining its genome-wide distribution, diverging substantially from previous reports [1,2], and to address its biogenesis, functions and potential mechanisms of action. Similarly to pseudouridine [13,23–25] and 5-methylcytosine [26], $m^1A$ is also catalyzed via co-opting of the tRNA/rRNA modifying machineries. The repressive impact of $m^1A$ on translation likely underlies its scarcity in cytosolic mRNAs. The dramatic reduction in ND5 methylation in mitochondria following the the 8-cell stage coincides with activation of zygotic transcription at this stage, an increase in oxygen consumption [27–29] and changes in mitochondrial morphology[30], collectively suggestive of a developmental, regulatory role for $m^1A$ methylation at this position.

**Figure legends**

**Figure 1.** Establishment of M$^1$A-seq and characterization of 205 putative m$^1$A-containing sites. **(a)** Scheme depicting the m$^1$A-seq pipeline. Following poly(A) selection, mRNA is fragmented into ~100 nt-long fragments. M$^1$A-containing fragments undergo enrichment using an anti-m$^1$A antibody, following which they are subjected to reverse transcription using either TGIRT or SuperScript enzymes, leading predominantly to misincorporation and premature truncation, respectively. As controls, the immunoprecipitation is either omitted (Input) or followed by a Dimroth rearrangement which converts m$^1$A to m6A (IP+Dimroth). **(b)** Venn-diagram depicting the overlap between sites detected across the four datasets analyzed here. The number of putative m$^1$A sites within all classes of RNA are indicated in black; sites within tRNA molecules are indicated in parenthesis in red. **(c)** Pie-chart summarizing the classes of RNAs in which the 205 putative m$^1$A containing sites were observed. Within tRNAs, the sites are further classified based on the position within tRNA harboring the sites. **(d)** All detected putative m$^1$A sites excluding ones in tRNA molecules. For each site, we present the percentage of misincorporation in Input and IP experiments, which were calculated as the mean respective values across the datasets in which these sites were detected, the % truncation as estimated from m$^1$A-seq-SS upon m$^1$A-IP, the number of samples and experiments in which the site was independently detected, and the sequence surrounding the putative m$^1$A site. We highlight two sequence motifs that are reproducibly found within them in red and green. The putative m$^1$A site is highlighted in yellow. **(e)** Misincorporation, truncation and coverage plots for putative m$^1$A sites identified in this study. The graphical representation, inspired by [3], depicts the truncation rate (black line, left y axis), misincorporation rate (stacked barplot, left y axis), and the overall coverage (grey shade, right y axis) in a sequence window surrounding the putative m$^1$A site.

**Figure 2:** Cytosolic m$^1$A sites share a sequence and structural motif and are strongly induced upon overexpression of TRMT6/TRMT61A: **(a)** Predicted secondary structures of sequence environment surrounding putative m$^1$A-containing sites in cytoplasmic mRNA and lncRNAs; note the almost invariable 7-bp loop stabilized by a relatively strong stem with terminal G-C base pair and UUCNANY loop. **(b)** Misincorporation levels at the indicated sites, measured via targeted sequencing, in cells depleted of TRMT6/TRMT61A via siRNAs, compared to mock treated controls. Error bars represent binomial confidence intervals. **(c)** Frequencies of misincorporation rates obtained in input versus IP RNA fractions, in WT versus upon overexpression of TRMT6/TRMT61A. Values are shown for four putative m$^1$A sites. Error bars reflect binomial confidence intervals. **(d)** Pie chart as in **Fig. 1c**, depicting the distribution into RNA classes of ~500 putative m$^1$A sites, of which 384 are within mRNA and lncRNA

molecules, detected upon overexpression of TRMT6/TRMT61A. **(e)** Sequence motif obtained via unbiased sequence analysis of the sequences surrounding the putative sites in mRNA upon overexpression. **(f-h)** Distributions of predicted free energies **(f)**, loop length **(g)** and stem length **(h)** among the 384 putative $m^1A$ sites in mRNA + lncRNAs, in comparison to randomly shuffled controls. These values are derived based on parsing of the predicted secondary structures obtained using RNAfold. Note that lower free energies are indicative or more thermodynamically stable structures. **(i)** Misincorporation percentages across three regions that were monitored in the massively parallel reporter assay. Positions are numbered with respect to the predicted $m^1A$ site, and the '0' position with predicted $m^1A$ is highlighted in yellow. **(j)** Functionally reconstructed sequence motif, based on measurement of misincorporation rates at each of the nucleotides in each of the displayed positions into every other nucleotides across each of the 74 T-loops (Methods). **(k)** Misincorporation rates following abolishment and gradual restoration of the stem. The number of consecutive complementary bases in the stem region (beginning with the T-loop proximal bases) are indicated in the X axis.

**Figure 3:** Tissue and development specific methylation in ND5 catalyzed via TRMT10C. **(a)** Highly tissue-specific distribution of misincorporation rates at ND5:1374 across 29 tissues based on >9000 RNA-seq datasets obtained from the GTEx collection, with particularly high levels in the pituitary gland and ovary. In comparison, misincorporation levels, pooled across all tissues, are presented for chrM:2617, a position in the mitochondrial 16S rRNA; This position was previously demonstrated to be methylated close to 100% [15], and thus allows a rough calibration of the readout. **(b)** Misincorporation levels at ND5:1374 in human muscle and ovary samples, based on targeted, strand-specific sequencing of the ND5 locus in poly(A) RNA. Error bars represent the binomial confidence interval. **(c)** Distributions of misincorporation levels across 1,529 individual cells from 88 human preimplantation embryos ranging from developmental day 3 (corresponding to the 8 stage phase) to day 7 [14]; In this study the authors used SuperScript II for RT. 16S rRNA methylation levels are shown in comparison as in (a). **(d)** Distributions of misincorporation levels across 124 cells spanning a developmental range from metaphase II oocytes to late blastocyst [16]; In this study the authors used SuperScript III for RT. **(e)** Misincorporation levels at ND5:1374 measured at the indicated time-points following actinomycin D-mediated transcriptional arrest. **(f)** Misincorporation levels at ND5:1374 in HEK293T cells siRNA mediated knockdown or overexpression of TRMT10C, or in cells treated with a control siRNA. Error bars represent the binomial confidence interval. **(g)** Misincorporation levels at ND5:1374 across 102 samples from human ovaries, color-coded based on the presence of a SNP at position 13708 (red - WT; blue - SNP).

**Figure 4:** $M^1A$-containing mRNAs are inefficiently translated. **(a)** Misincorporation rates at the ND5 locus across the polysomal fractions, measured using strand specific targeted sequencing (n=3). Dots represent the measurements, and the red bar represents the mean. **(b)** Misincorporation rates at four selected loci following overexpression of TRMT6/TRMT61A, measured across four polysomal fractions (n=3), displayed as in (a). **(c)** Depiction of the four designed variants, perturbing either the modified site or the secondary structure, on the basis of the methylated site in the PRUNE gene. The methylated position is plotted in red, the perturbed position in magenta. **(d)** Misincorporation percentages at the designed site, quantified across the four constructs using targeted sequencing (n=3). Dots represent measurements, the red bar represents the mean. Note that for the 'site mutation' variant the misincorporation rate reflects the fraction of reads not harboring a 'T, in contrast to all remaining variants in which it reflects the fraction not harboring an 'A'. **(e)** Renilla-normalized firefly luciferase levels in TRMT6/TRMT61A overexpressing cells divided by the corresponding ratio in non-overexpressing control for the four tested constructs. Dots represent the measurements, red bars the mean. T-test based P values are presented.

## Methods

**Cell culture for knockdown and overexpression experiments:** Human HEK293T cells ( ATCC; passage number 5-15; no further verification of cell line identity was performed; screened negatively for mycoplasma) were plated in 6-well plates at 20% confluency. siRNAs targeting TRMT10C (Thermo Fisher: s29784), TRMT6 (s28400) and TRMT61A (s41859) were transfected using Lipofectamine RNAiMAX (Life Technologies, catalog no. AB4427037) following the manufacturer's protocols, with two siRNA boosts with a 48 hours interval between them; siRNA targeting TRMT6/TRMT61A were co-transfected using half of the recommended amount of siRNA per each. As negative controls, we used Ambion® In Vivo Negative Control #2 siRNA (catalog number: 4390846). Cells were harvested at 96 hours. For overexpression, plasmids encoding full length of TRMT6/TRMT61A under CMV promoter were obtained from ORIGENE. The plasmids were transfected into HEK293T cell using PolyJet (SignaGene) with one boost of the plasmid at 24 hours. Cells were harvested 48 hours following transfection.

**Blocking transcription or translation:** Transcription was blocked using using Actinomycin D (Sigma) at a concentration of 10µg/mL or Ethidium Bromide (Amresco) at a concentration of 0.4µg/mL. Translation was blocked using cycloheximide (Sigma) at concentration of 100µg/mL.

**Human RNA:** Total RNA extracted from ovary and muscles from a human donor were obtained from Takara. Lymphoblastoid cell lines were obtained from Coriell.

**RNA preparation for $m^1A$-seq:** RNA was extracted from cells using NucleoZOL (MACHEREY-NAGEL). Enrichment of polyadenylated RNA (polyA+ RNA) from total RNA was performed using Oligo(dT) dynabeads (Invitrogen) according to the manufacturer's protocol. The mRNA was chemically fragmented into ~100-nt-long fragments using RNA fragmentation reagent (Ambion). The sample was cleaned using Dynabeads (Life Technologies), and resuspended in 20 µl of IPP buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.5).

**$m^1A$-seq and $m^6A$-seq:** Transcriptome-wide mapping of $m^1A$: The protocol we developed for mapping $m^1A$ is based on our previously published protocol for mapping N6-methyladenosine (m6A) [31]. Briefly, 40 µl of protein-G magnetic beads were washed and resuspended in 200 µl of IPP buffer, and tumbled with 5 µl of affinity purified anti-$m^1A$ polyclonal antibody (Synaptic Systems) at room temperature for 30 minutes. RNA was added to the antibody-bead mixture, and incubated for 2 h at 4°C. The RNA was then

washed twice in 200 µl of IPP buffer, twice in low-salt IPP buffer (50 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.5), and twice in high-salt IPP buffer (500 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.5), and eluted in 30 µl RLT (Qiagen). To purify the RNA, 20 µl MyOne Silane Dynabeads (Life Technologies) were washed in 100 µl RLT, resuspended in 30 µl RLT, and added to the eluted RNA. 60 µl 100% ethanol was added to the mixture, the mixture attached to the magnet and the supernatant discarded. Following two washes in 100 µl of 70% ethanol, the RNA was eluted from the beads in 10 µl $H_20$. Dimroth rearrangements were performed as described in [2]. For mapping $m^6A$ following Dimroth conversion of $m^1A$, RNA extracted from TRMT6/TRMT61A overexpressing cells was first Dimroth rearranged as described in [2], and subjected to **$m^6A$**-seq as previously published [31].

**Library preparation:** Strand-specific $m^1A$ RNA-seq libraries were generated on the basis of the protocol described in [32,33]. Briefly, RNA was first subjected to FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), followed by a 3' ligation of an RNA adapter using T4 ligase (New England Biolabs). Ligated RNA was reverse transcribed either using SuperScript-III (Invitrogen) or using TGIRT-III (InGex), and the cDNA was subjected to a 3' ligation with a second adapter using T4 ligase. The single-stranded cDNA product was then amplified for 9-12 cycles in a PCR reaction. Libraries were sequenced on Illumina Nextseq platforms generating short paired end reads, ranging from 25-55 bp from each end.

**Identification of putative $m^1A$ sites:** A human reference genome was generated on the basis of the hg19 assembly of the human genome, supplemented with tRNA, rRNA and snRNAs, obtained from the modomics database [34]. Non-enriched (Input), $m^1A$-enriched (IP) and Dimroth treated $m^1A$-enriched (IP+Dimroth) samples were aligned to the genome, using STAR aligner [35] with an increased stringency allowing only up to 3 mismatches per each read pair ('--outFilterMismatchNmax 3'). All duplicates were marked using picard tools MarkDuplicates.jar', and non-primary alignments were removed. The identity of each nucleotide at each genomic position was extracted using 'samtools mpileup', using max per-file depth settings of '-d 100000'. A custom script was employed to parse the pileup format into a tabular format summarizing the abundance of each nucleotide at each position. All positions harboring an 'A' in the annotated sense strand, with at least 2 mismatches, occurring in at least 10% of the reads overlapping it were recorded.

All sites, recorded across any replicates across any of the samples were pooled into a single dataset. Misincorporation rates at each of these pooled sites were subsequently re-extracted for each of the experiments (allowing analysis of each site in each experiment, even if did not pass the initial thresholds in that particular experiment). A bona-fide $m^1A$ site is expected to have higher mismatch rates in IP

compared to Input, and lower levels in IP+Dimroth compared to IP. We hence used a chi-squared test on the basis of the aggregated misincorporation counts (across replicates), to test the hypotheses that (1) the number of mismatches in the IP sample is higher than Input, and (2) the number of mismatches in IP is higher than in IP+Dimroth. The following criteria were then used for identification of putative $m^1A$ sites: (1) At least one of the two calculated P values was significant ($P<0.05$), (2) The product of the P values was < 0.01, (3) The difference in misincorporation rate between the sample with the lowest levels of misincorporation and the highest level ≥0.2, (4) The sites had to be covered by ≥10 reads in at least two samples, (5) ≥1% of all reads mapping to the site (across all replicates and samples) had to be 'A', ≥1% of all reads mapping to the site (across all replicates and samples) had to be 'T', and ≥1% of all reads mapping to the site (across all replicates and samples) had to be 'C' or 'G'. The latter criteria were set to help discriminating SNPs (where only one alternative to an 'A' is expected) and an $m^1A$ site (where typically more than one type of misincorporation is observed). Sites harboring identical sequences in a 24 bp window (12 bp upstream + 11 bp downstream) surrounding the putative site were filtered, to retain only a single one. Nonetheless, due to merging of sites from different datasets and the multiple loci from which identical or nearly identical tRNAs are transcribed, a subset of duplication was retained - and are flagged as such - within tRNA entries in **Table S2**.

This pipeline was applied to four batches of samples: (1) Input, IP and IP+Dimroth (two replicates each) to which we applied $m^1A$-seq-SS, (2) Input, IP and IP+Dimroth (two replicates each) to which we applied $m^1A$-seq-TGIRT, (3) 16 samples generated by Dominissini and coworkers [2] with RNA-seq readouts from HepG2 cells in m6A-IP or Input, with and without Dimroth rearrangements downloaded from GEO (accession: GSE70485), (4) 28 samples generated by Li and coworkers [1], with $m^1A$ mapped under different genetic perturbations and upon different stimuli downloaded from GEO (GSE73941). The 'high confidence' dataset of 205 sites comprises all sites for which at least two significant P values were obtained across any of the comparisons performed across any of the datasets; The 'low confidence' sites comprised all sites associated with a single significant P value. Of note, to accommodate the distinct experimental design in the datasets obtained from Li et al and Dominissini et al, we adapted the precise sets of comparisons that were performed by the analytical pipeline. Specifically, in addition to assessing whether IP differed from Input, and from IP+Dimroth, in the Dominissini et. al dataset we further derived chi-squared based P values to assess whether the Dimroth rearrangement in the Input samples led to reduced mismatch levels than in its absence. In the dataset produced by Li et al [1], the authors did not use a Dimroth rearrangement, but instead relied on treating the RNA with AlkB, an E. coli derived demethylase that eliminates $m^1A$. Li et al further compared measurements upon knockout of AlkBH3, which they find to demethylate $m^1A$. Accordingly, we performed four statistical tests, examining differences in mutation

rates between (1) IP versus Input in WT samples, (2) IP versus Input across stress conditions (H2O2, starvation) (2) IP versus IP+AlkB, (3) IP in WT cells versus IP in ALKBH3 knockout cells. Of note, given that any site passing any of the statistical tests was considered a putative site, effectively our criteria for identifying putative $m^1A$ sites in the Dominissini and Li datasets are more *lenient* compared to the criteria we applied for the two datasets we generated. Finally, in the dataset generated upon overexpression of TRMT6/TRMT61A we did not perform Dimroth rearrangements, and instead performed the following two tests, examining differences in misincorporation rates between: (1) Input samples and IP samples, (2) Input samples following overexpression of TRMT6/TRMT61A compared to Controls. All analyses were performed utilizing the identical computational pipeline, into which we fed, as parameters, the precise comparisons to be made.

Although our $m^1A$-seq approach provides strand specific data, in the initial analyses in figures 1 and 2 strand specificity was inferred from the genomic annotation rather than the read. This allowed applying an identical pipeline to the data generated in this study and the two previously published datasets. For analyzing $m^1A$ at the ND5 locus, we subsequently called mutations in a strand specific manner (separately inferring mutations on the + and - strands), to prevent dilution of misincorporation signal from the antisense strand.

**Quantification of RT truncations:** For each of the putatively identified $m^1A$ sites, we calculated the rate of transcription termination at position +1 with respect to the site. This was performed by first artificially merging all read pairs into a single, artificial read extending from the beginning of one read to the end of its mate, and then using bedtools to count the number of reads *beginning* and *overlapping* each position. The ratio between the two was defined as the stop rate at that position, as performed in [23].

**Identification of $m^1A$ peaks:** Peak detection was performed based on our previously published approach for detecting peaks in m6A-seq data [31,36]. Specifically, an in-house script was first used to project all reads aligning to the genome upon the human transcriptome. Only reads fully matching a transcript structure, as defined by the 'UCSC Known Genes' transcriptome annotation, were retained. Such reads were computationally extended in transcriptome space from the beginning of the first read to the end of its mate, and coverage in transcriptome-space was calculated for each nucleotide across all transcripts.

Putative $m^1A$ sites were identified using a 3 step-approach. **(1) Peak detection within genes.** To search for enriched peaks in the m6A IP samples, we scanned each gene using sliding windows of 100 nucleotides with 50 nucleotides overlap. Each window was assigned a Peak Over Median (POM) score,

defined as mean coverage in the window / median coverage across the gene. Windows with POM scores greater than 4 (*i.e.*, greater than 4-fold enrichment) and with a mean coverage >10 reads were retained. Overlapping windows were merged together, and for each disjoint set of windows in transcriptome space we recorded its start, end, and peak position, corresponding to the position with the maximal coverage across the window. **(2) Ensuring that peaks were absent in input.** We repeated the procedure in step (1) for the input sample. We eliminated from all subsequent analysis all windows that were detected in both step (1) and (2). **(3) Comparison of multiple samples**. To search for consistently occurring peaks across different samples, we first merged the coordinates of all windows from all samples passing step (1) and (2), to define a set of disjoint windows passing these filters in at least one of the samples. For each such window, we recalculated the peak start, end, peak position, and POM score (as defined above) across each of the samples using the approach in step (1). In addition, for each window we calculated a Peak Over Input (POI) score, corresponding to the fold-change of coverage across the window in the IP sample over the coverage in the input sample. To account for differences in sample depth, we estimated the mean difference between IP and input samples across the 500 most highly expressed genes, which we used as an estimate for background. We subtracted this background from the POI score.

Based on careful examination of the peaks at the beginning of transcripts which revealed that in many cases the peaks originated from the first transcribed nucleotide, we utilized the approach we described in [36]. Briefly, this approach relies on calculating the fold-change upon $m^1A$-IP, compared to input, in reads *beginning* at each of the first 50 annotated positions in each transcript. For the analysis displayed in **Fig. S4D**, we first calculated these ratios across the set of SuperScript IP and Input samples. We then integrated the quantifications of fold-changes by extracting the maximum fold-change per each position per each transcript. We then quantified the proportion of pileups harboring an 'A' as a function of this fold-change, revealing that pileups beginning with As were more frequent at the higher-confidence sites, harboring stronger fold-changes.

**mRNA expression analysis:** To estimate expression levels, reads were aligned against the human genome using RSEM (version 1.2.31) with default parameters [37]. For robust comparison between different samples, we used TMM normalization of the RSEM read counts[38] as implemented by the edgeR package[39] in R.

**Prediction of RNA secondary structure:** For predicting secondary structure in the region surrounding putative $m^1A$ sites, we extracted a sequence window of 24 bp, including 12 bp upstream of the modified site and 11 bp downstream. Free energy calculations and predicted secondary structures were calculated

using RNAfold version 2.1.5. The secondary structures were subsequently parsed, using an in-house script, to quantify the stem and loop lengths **(Fig. 2)**.

**Massively parallel reporter assay:** The design and cloning of the MPRA library into a plasmid were described in [13]. A 10cm plate of HEK-293T cells was transiently transfected with equal amounts of TRMT6, TRMT61A and the library plasmid using PolyJet (SignaGene). RNA was purified using Nucleozol reagent (Macherey Nagel). Sequence specific $m^1A$-seq-TGIRT was performed on total RNA essentially as described in [13], except reverse transcription from the constant sequence stemming from the library plasmid (AGCATTAACCCTCACTAAAGGGAAAGG) was carried out using TGIRT-III (InGex). Adapter ligation and PCR enrichment with an inner plasmid specific primer (GGTCCGATATCGAATGGCGC), were carried out as described [13].

Alignment of the MPRA data was performed as described in [13]. For quantifying misincorporations we used 'samtools mpileup', as described above. For the sequence logo depicted in **Fig. 2j**, we first extracted the 75th percentile of misincorporation rates following point-mutation of each of the indicated sites across each of the assayed 74 T-loops into each of the four nucleotides. For each position, this value was then divided by the sum of this value across all four nucleotides, to yield 'relative misincorporation rate' (summing up to 1, at each position). The height of each nucleotide at each position was then plotted in direct proportion to its relative misincorporation rate.

**Annotation of mitochondrial and tRNA sites:** All reads were aligned to the chrM assembly forming part of the human hg19 assembly, and supplementary tables provide positions with respect to it. For consistency with the mitochondrial community, within the manuscript we refer to positions with respect to the slightly more updated chrM_rCRS assembly. For tRNAs, we refer to all positions in the figures based on the standard tRNA nomenclature (so that the anticodon nucleotides are always numbered 34–36, and the T loop between positions 54 and 60).

**Quantification of misincorporation in GTeX and single cell RNA-seq data**: Raw fastq files were obtained for each of the files in these datasets, and aligned using STAR (as above). Mpileup was applied to quantify misincorporation levels across the positions detected in this study. For Fig. 4C-D, we filtered out all single cells in which a SNP was observed at position 13708, as this SNP severely reduces methylation. A SNP was called in this position, based on the RNA-seq data, if >80% of the reads corresponded to the SNP nucleotide.

**Polysome fractionation:** Polysome fractionation was done as specified in [40], with one exception: we used 10%-50% sucrose gradient instead of 5%-50% .

**Targeted sequencing of m$^1$A amplicons:** For targeted measurement of m$^1$A levels at specific loci, reverse transcription was done on 1μg of RNA using random hexamers and TGIRT-III (InGex) reverse transcriptase. Amplicons were PCR amplified using a nested PCR approach, involving a first amplification step with gene-specific primers and a partial Illumina adapter tail, and a second amplification leading to the incorporation of the full-length illumina adapters. For ND5 amplicons, RT was done using a strand specific primer instead of random hexamers, to avoid contamination by the ND5 antisense transcript. The resultant amplicon was amplified with primers including the full-length Illumina adapters in a single step. All primers can be found in **Table S5**.

**Luciferase assay:** For the luciferase experiments two plasmids were used: (1) pGL4.73 - for expression of Renilla luciferase under an SV40 promoter, and (2) a plasmid encoding an ATG start codon followed by 60-bp surrounding the PRUNE m$^1$A site ( GCGGAGGCCGATTCGCCGTGTGGCGGGTTCGAGTCCCGCCTCCTGACTCTGGCCTCTAGTC) followed by firefly luciferase, all driven by a CMV promoter. We constructed three derivatives based on this plasmid, point mutating the sequence and structure, as described in the text. These plasmids were pooled together and transfected into cells with either control DNA or with TRMT6/61A overexpression plasmids. The luciferase assay was done with promega kit according to the manufacturer's instructions.

**Data availability:** All datasets generated in this manuscript were uploaded to Gene Expression Omnibus (GEO, Accession GSE97419).

## Supplementary Figure legends

**Figure S1**. **(a)** Overview of the analytical pipeline, which was applied to each of the four datasets. The pipeline utilizes two key statistical tests, to identify 'A' containing sites that either harbor misincorporations at higher levels following immunoprecipitation with an anti-m$^1$A antibody compared to Input samples or compared to Dimroth treated samples. In addition the pipeline requires that minimum levels of at least two distinct types of misincorporations be observed for a given site. **(b-d)** Truncation and misincorporation rates are reproducible between replicates, both with SuperScript and with TGIRT enzymes. Shown are scatter-plots, on the basis of 205 putative m$^1$A sites, indicating the correlation between replicates of truncation rates using SuperScript **(b)**, misincorporation rates using SuperScript **(c)** and misincorporation rates using TGIRT **(d)**. **(e)** Comparison of truncation rates using TGIRT (y axis) and SuperScript (x axis), on the basis of the 205 detected sites. Sites are color-coded based on the indicated classes of RNA. **(f)** Comparison of misincorporation rates using TGIRT and SuperScript, plotted as in (e). **(g)** Comparison of misincorporation rates in IP samples compared to IP+Dimroth, in

both cases using TGIRT. **(h)** Comparison of misincorporation rates in IP samples compared to Input, using TGIRT. (**i**) Serial dilutions of RNA extracted from cells overexpressing TRMT6/TRMT61A within RNA extracted from cells depleted of TRMT6/TRMT61A. The X axis denotes the percentage of the RNA originating from cells overexpressing TRMT6/TRMT61A, and the Y axis captures the misincorporation rate, as measured via targeted sequencing of the three presented sites.

**Figure S2.** Misincorporation levels across 8 lncRNA + mRNA and 2 rRNA sites on the basis of a random sampling of 4114 GTEx RNA-seq samples.

**Figure S3.** $M^1A$ detection as a function of gene expression, on the basis of analysis of sites identified following TRMT6/TRMT61A overexpression. **(a)** All genes were divided into five bins, based on the indicated levels of expression. The number of $m^1A$ harboring sites in each bin is plotted. **(b)** As in **(a)**, but normalized by number of genes within each bin. Error bars indicate the binomial confidence intervals.

**Figure S4:** Peak detection following $m^1A$-seq reveals a strong enrichment towards 5' terminus. **(a)** Representative meta-gene profiles of read coverage in $m^1A$-IP (top) or Input (bottom). Shown are the first 500 nucleotides immediately following the transcription start site (left) and immediately preceding the annotated 3' terminus of the gene (right). Genes are divided into five bins, based on expression levels. **(b)** Distribution of expression levels (based on Input samples) of all genes harboring high-confidence peaks. On this distribution are overlaid, in red, the genes in which putative $m^1A$ sites were detected. **(c)** Consistently identified peaks are highly enriched towards the 5' terminus of the gene. Each peak was classified into one of five segments, as in [41], in the following order: Transcription start site (TSS) if the peak is present in the first 200 nt of the gene, 5' UTR if in the 5' UTR region but outside of the TSS region, stop codon region, comprising 200 nt on both sides of the stop window, CDS region for peaks within the CDS, and 3' UTR for remaining 3' UTR peaks. Each peak is scored based on the number of experiments in which it was detected, whereby more robustly identified ones should be considered as peaks of higher confidence. The stacked bar plots summarize the relative proportion of peaks in each segment, and the right-most bar plots the relative amount of space taken up by each of these segments. **(d)** Analysis of transcription start site peaks. The number of reads *beginning* (rather than overlapping) at each of the first 50 annotated transcribed bases was calculated across IP and Input samples (Methods), and the log fold-change between the two was derived. All fold-changes were binned into six bins (as plotted) and the fraction of positions harboring an 'A' are plotted as a function of this binned fold-change, revealing that positions at the transcription start site that are enriched in IP samples over Input samples are biased towards beginning with an 'A'.

**Figure S5.** Common secondary structure surrounding mitochondrial sites; Note the 'UAAA' motif in the loop, stabilized by a stem.

**Figure S6.** Validation of putative $m^1A$ sites, via m6A-seq of Dimroth converted RNA extracted from TRMT6/TRMT61A mRNA. **(a)** Coverage plot along the cytosolic 28S rRNA across the four indicated samples. The known $m^1A$ and m6A sites are indicated. **(b)** Coverage plots as in (a) for the mitochondrial 16S rRNA. The normalized coverage levels are indicated to the left of the track (note the orders of magnitude higher coverage at the $m^1A$ sites upon m6A-IP in Dimroth treated samples, compared to the controls). **(c)** Quantification of percentage of coverage in a 100-nt centered region around putative $m^1A$ sites out of overall coverage of the gene in two biological replicates of mRNA extracted from cells overexpressing TRMT6/TRMT61A; This mRNA was subjected to Dimroth treatment followed by m6A-seq. Quantifications were obtained for 40 sites exhibiting misincorporation levels >10% upon TRMT6/TRMT61A overexpression in addition to the two sites on the 16S and 28S rRNA molecules. **(d)** Comparison of '% Coverage in peak' (as in **C**) between RNA subjected to Dimroth treatment and m6A-IP compared to RNA only subjected to m6A-IP. **(e)** Misincorporation levels across the indicated conditions across 8 sites (of the 42 tested ones) in which a significant P value ($P<0.05$) was obtained when comparing Dimroth + m6a-IP samples to their corresponding input (significance indicated via '*'), and/or when comparing Dimroth + m6A-IP to no-Dimroth + m6A-IP (significance indicated via '#').

**Figure S7.** ND5 methylation levels are genetically determined, are in part controlled by a SNP two bases upstream, and are increased in stable transcripts. **(a)** Snapshot of randomly sampled reads aligned to ND5 locus. Reads originating from the heavy strand are depicted in red, reads from the light strand in purple. Top panel reflects an IP sample, lower panel reflects an Input sample. Misincorporations are apparent only in reads originating from the heavy strand. **(b)** Misincorporation levels at ND5 locus in DNA and RNA samples of five individuals. For RNA, the distribution of misincorporation reads are shown across all tissues; For DNA, the measurement consists of a single measurement available in GTEx. **(c)** Misincorporation levels at ND5:1374 measured at the indicated time-points following ethidium bromide mediated transcriptional arrest. **(d)** Correlations between misincorporation levels at ND5:1374 in skin versus brain samples, from the same individuals. **(e)** A histogram of all pairwise correlation coefficients between tissues (but from the same individuals); note that values are centered around 0.5, rather than around 0 if they were independent of each other. **(f)** Misincorporation rates at ND5:1374 obtained via strand-specific targeted sequencing of the ND5 locus across 6 lymphoblastoid cell lines, two harboring a G13708A SNP and 4 WT samples.

**Figure S8.** $M^1A$ represses translation. **(a)** Representative sucrose gradient, indicating the division into fractions based on the number of polysomes associated with them. **(b)** Scheme of experimental design. The WT $m^1A$ containing stretch from the PRUNE gene was cloned either in-frame and upstream of firefly luciferase (CDS construct) or as a 3' UTR element (3' UTR construct). Control or TRMT6/TRMT61A overexpressing cells were co-transfected with each of these plasmids along with a plasmid expressing renilla. **(c-d)** Renilla-normalized firefly luciferase levels in TRMT6/TRMT61A overexpressing cells, standardized by this value in the non-overexpressing (control) cells. Error bars represent standard deviation of the mean (n=3). Note that for the CDS construct the presented data is identical to the one in **Figure 4**, and is re-plotted for convenience.

**Figure S9. (a)** Comparison of expression levels of all genes acquiring robust levels of $m^1A$ (defined as misincorporation levels >10%) upon overexpression of TRMT6/TRMT61A, in comparison to WT (non-overexpressing) counterparts. **(b)** Misincorporation levels in the four indicated genes, in cells overexpressing TRMT6/TRMT61A, in a 6 hour timecourse following cycloheximide treatment.

**Legends of Supplementary Tables**

**Table S1.** Alignment statistics for all four analyzed samples. Presented are total number of reads and the percentage of aligned reads.

**Table S2.** Dataset of 277 putative $m^1A$ sites, obtained after intersection of the four analyzed datasets. 205 of the peaks scored as significant ($P<0.01$) based on two or more comparisons and form the basis for the analyses in the paper; These are marked as 'higher confidence'. The 72 remaining sites are considered 'lower confidence', as they were only associated with a single significant P value. For each site, we list its genomic coordinates, its gene annotation, the genomic sequence surrounding it, the predicted secondary structure and corresponding free energy associated with the site. We further indicate the ratio of WT nucleotides observed at the sites across each of the experiments (the misincorporation rate is thus 1 minus this value), the P values achieved for each of the comparisons across each of the datasets.

**Table S3.** Dataset of 690 consistently identified peaks, used in the context of analyses presented in Figure S3 revealing strong bias towards 5' termini of genes. The dataset is in bed format.

**Table S4.** Dataset of 495 putative $m^1A$ sites, obtained following overexpression of TRMT6/TRMT61A.

**Table S5.** Primer sequences used in this study.

**Statements**

Reprints and permissions information is available at www.nature.com/reprints".

**Competing financial interests**: None.

Correspondence and requests for materials should be addressed to schwartz@weizmann.ac.il.

**Contributions**: M.S., A.S, R.N, D.B, M.E, W,R, N.SG, and S.S designed the experiments. M.S and A.S performed the experiments. R.N performed the oligoarray experiment. R.W. assisted in the polysomal fractionation. A.N. assisted in the GTeX analysis. S.S. wrote the manuscript with input from all authors.

**Funding**: This project was supported by the Israel Science Foundation (543165), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 714023), by the Abisch-Frenkel-Stiftung, by research grants from The Abramson Family Center for Young Scientists, the David and Fela Shapell Family Foundation INCPM Fund for Preclinical Studies, the Estate of David Turner, and the Berlin Family Foundation New Scientist Fund.

**References**


1. Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat. Chem. Biol.* (2016). doi:10.1038/nchembio.2040
2. Dominissini, D. *et al.* The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* (2016). doi:10.1038/nature16998
3. Hauenschild, R. *et al.* The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res.* **43,** 9950–9964 (2015).
4. Jones, J. W. & Robins, R. K. Purine nucleosides. III. Methylation studies of certain naturally occurring purine nucleosides. *J. Am. Chem. Soc.* **85,** 193–201 (1963).
5. Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19,** 958–970 (2013).
6. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* (2016). doi:10.1038/nmeth.4057
7. Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135,** 919–932 (2008).
8. Hodgkinson, A. *et al.* High-resolution genomic analysis of human mitochondrial RNA sequence variation. *Science* **344,** 413–415 (2014).
9. Bar-Yaacov, D. *et al.* RNA-DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. *Genome Res.* **23,** 1789–1796 (2013).
10. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).
11. Ozanick, S., Krecic, A., Andersland, J. & Anderson, J. T. The bipartite structure of the tRNA $m^1A58$ methyltransferase from S. cerevisiae is conserved in humans. *RNA* **11,** 1281–1290 (2005).
12. Jühling, F. *et al.* tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **37,** D159–62 (2009).

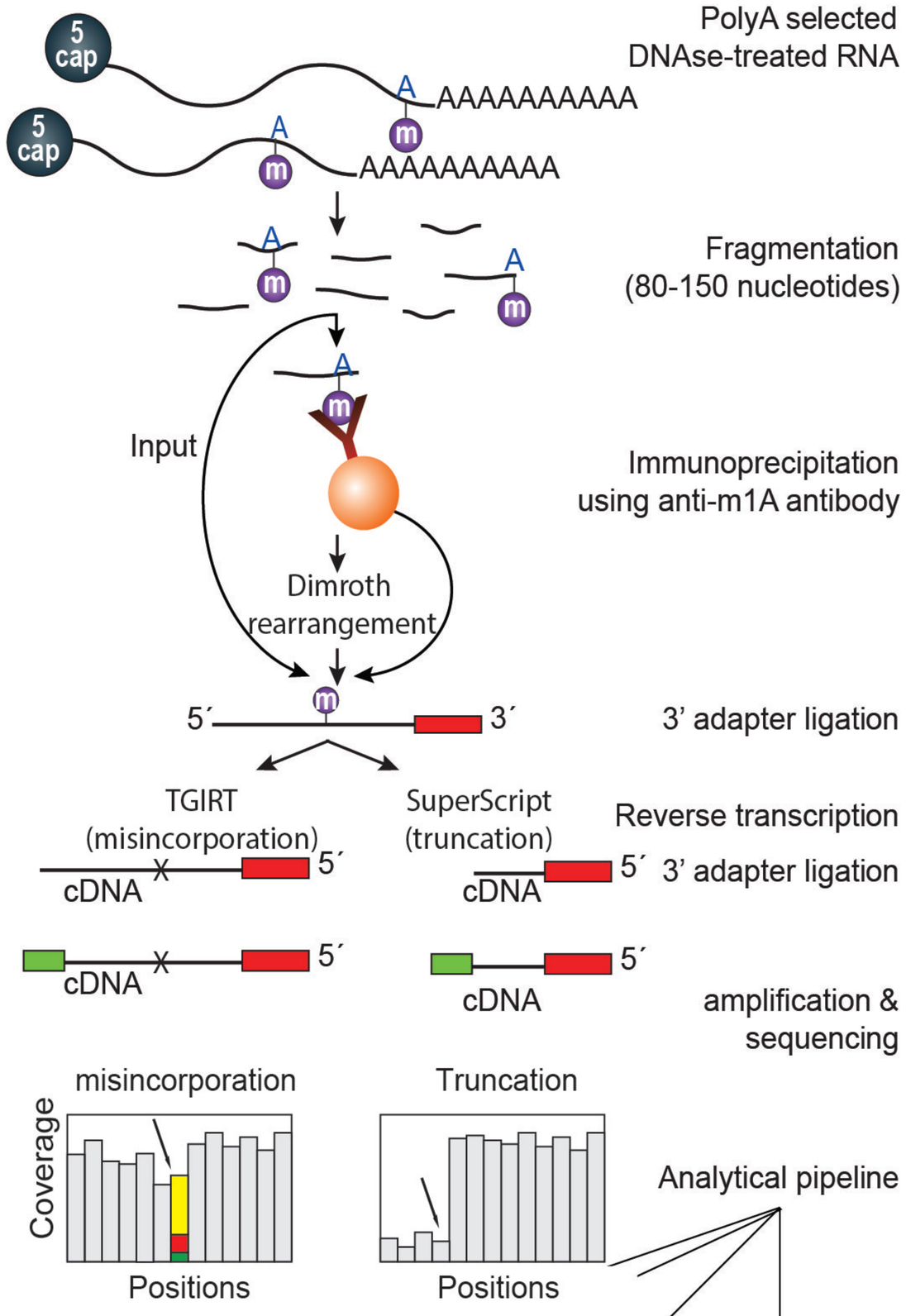
13. Safra, M., Nir, R., Farouq, D. & Schwartz, S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res.* (2017). doi:10.1101/gr.207613.116
14. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165,** 1012–1026 (2016).
15. Bar-Yaacov, D. *et al.* Mitochondrial 16S rRNA Is Methylated by tRNA Methyltransferase TRMT61B in All Vertebrates. *PLoS Biol.* **14,** e1002557 (2016).
16. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20,** 1131–1139 (2013).
17. Pikó, L. & Taylor, K. D. Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. *Dev. Biol.* **123,** 364–374 (1987).
18. Vilardo, E. *et al.* A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase--extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res.* **40,** 11583–11593 (2012).
19. Hudson, G. *et al.* Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *Am. J. Hum. Genet.* **81,** 228–233 (2007).
20. Torroni, A. *et al.* Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am. J. Hum. Genet.* **60,** 1107–1121 (1997).
21. Carelli, V. *et al.* Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. *Am. J. Hum. Genet.* **78,** 564–574 (2006).
22. Brown, M. D., Sun, F. & Wallace, D. C. Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage. *Am. J. Hum. Genet.* **60,** 381–387 (1997).
23. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159,** 148–162 (2014).

24. Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515,** 143–146 (2014).
25. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One* **9,** e110799 (2014).
26. Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-Wide Mapping of 5-methylcytidine RNA Modifications in Bacteria, Archaea, and Yeast Reveals m 5 C within Archaeal mRNAs. *PLoS Genet.* **9,** e1003602 (2013).
27. Mills, R. M. & Brinster, R. L. Oxygen consumption of preimplantation mouse embryos. *Exp. Cell Res.* **47,** 337–344 (1967).
28. Ginsberg, L. & Hillman, N. ATP metabolism in cleavage-staged mouse embryos. *J. Embryol. Exp. Morphol.* **30,** 267–282 (1973).
29. Biggers, J. D. & Borland, R. M. Physiological aspects of growth and development of the preimplantation mammalian embryo. *Annu. Rev. Physiol.* **38,** 95–119 (1976).
30. Wassarman, P. M. & Josefowicz, W. J. Oocyte development in the mouse: an ultrastructural comparison of oocytes isolated at various stages of growth and meiotic competence. *J. Morphol.* **156,** 209–235 (1978).
31. Schwartz, S. *et al.* High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155,** 1409–1421 (2013).
32. Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* **12,** 323–325 (2015).
33. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341,** 1237973 (2013).
34. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41,** D262–7 (2013).
35. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
36. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at

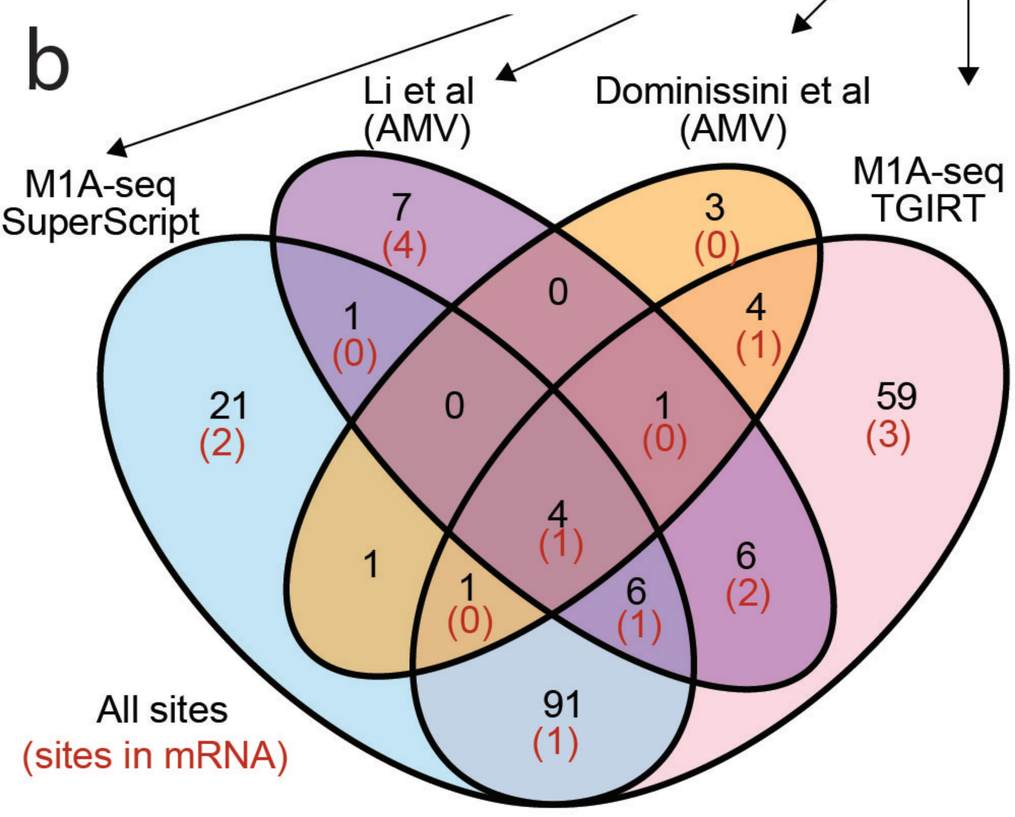
internal and 5’ sites. *Cell Rep.* **8,** 284–296 (2014).

37. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
38. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).
39. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
40. Gandin, V. *et al.* Polysome fractionation and analysis of mammalian translatomes on a genome-wide scale. *J. Vis. Exp.* (2014). doi:10.3791/51455
41. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).

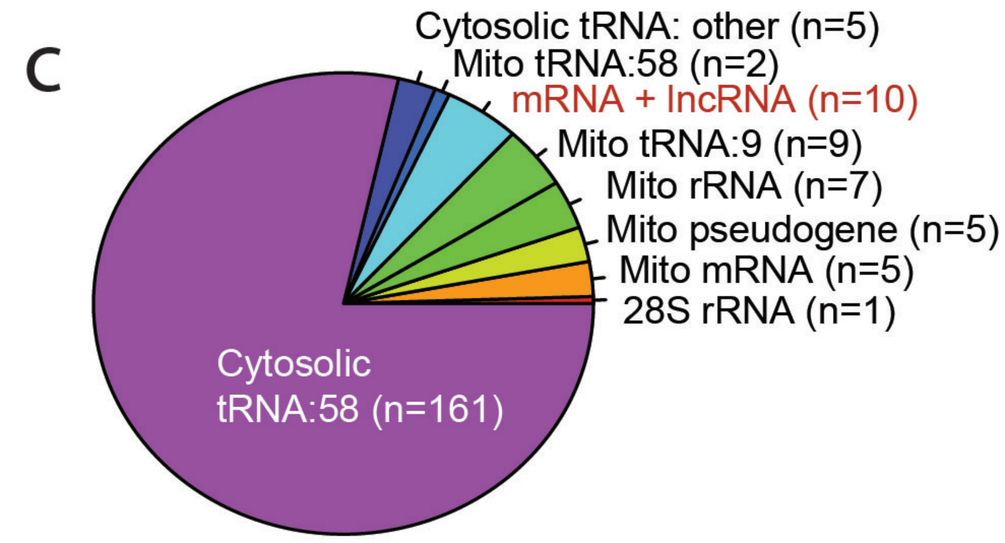
a




b




C




d

| | coords | class | % Mis (Input) | % Mis (IP) | % Trunc | # Samp | # Exp. | sequence |
|---|---|---|---|---|---|---|---|---|
| cytosolic RNA | 28S:1302 | rRNA | 70.7% | 76.8% | 87.2% | 55 | 4 | CCCGTCTTGAAACACGG |
| | MALAT1:8398 | lncRNA | 81.3% | 76.2% | 75.1% | 8 | 2 | CGGGGTTCAAATCCCTG |
| | C9orf100:4328 | mRNA | 25.0% | 84.2% | 41.8% | 4 | 1 | CTGGGTTCAAATCCCAG |
| | PRUNE:10 | mRNA | 2.3% | 86.5% | 17.0% | 13 | 4 | GCGGGTTCGAGTCCGGC |
| | ZNF664:1530 | mRNA | 0.0% | 44.3% | 73.8% | 4 | 1 | CAGAGTTCGAGCCTCTG |
| | GTF3C2:2873 | mRNA | 0.0% | 50.0% | 37.9% | 3 | 1 | GCTGGTTCGAATCCATT |
| | BRD2:119 | mRNA | 0.0% | 62.2% | 13.4% | 4 | 1 | GAGGATTCGAAAACCCT |
| | ATAD3B:1300 | mRNA | 0.0% | 80.0% | 3.2% | 2 | 1 | ACTGGGCCAATACCAGC |
| | TP53I13:891 | mRNA | 0.0% | 27.3% | 0.0% | 2 | 1 | GTCTGTTCAAGTCAGGC |
| | SRSF1:518 | mRNA | 17.1% | 52.6% | 0.0% | 7 | 1 | CGGGGGTGGAGGTGGCG |
| | UBC:2288 | mRNA | 15.1% | 44.4% | 0.0% | 6 | 1 | TGGGATGCAAATCTTCG |
| mitochondrial RNA | 16S:947 | rRNA | 83.4% | 90.3% | 41.2% | 52 | 4 | GTTCCTTAAATAGGGAC |
| | 12S:490 | rRNA | 1.3% | 42.6% | 29.5% | 18 | 3 | CTGCTCGCCAGAACACT |
| | 12S:282 | rRNA | 0.6% | 30.6% | 8.1% | 18 | 3 | GTCAATAGAAGCCGGCG |
| | 16S:363 | rRNA | 0.6% | 29.5% | 20.5% | 13 | 2 | CCAAGATAGAATCTTAG |
| | 16S:1360 | rRNA | 0.5% | 16.4% | 14.9% | 12 | 2 | CGCTATTAAAGGTTCGT |
| | 12S:433 | rRNA | 0.4% | 22.0% | 9.1% | 9 | 1 | TGGGATTAGATACCCCA |
| | 16S:1162 | rRNA | 0.1% | 3.8% | 19.7% | 6 | 1 | TCGGAGCAGAACCCAAC |
| | ND5:1374 | mRNA | 12.5% | 82.4% | 26.1% | 23 | 3 | ACGCCTGGCAGCCGGAA |
| | COX2:297 | mRNA | 0.7% | 28.7% | 25.2% | 14 | 2 | CATCAAATCAATTGGCC |
| | COX1:1472 | mRNA | 0.6% | 32.2% | 3.0% | 15 | 2 | CCTCCATAAACCTGGAG |
| | ND6:449 | mRNA | 0.5% | 22.1% | 16.9% | 5 | 1 | CCCCCATAAATAGGAGA |
| | COX3:707 | mRNA | 0.5% | 30.1% | 5.6% | 8 | 2 | TTGGCTTCGAAGCCGCC |

m¹A

e

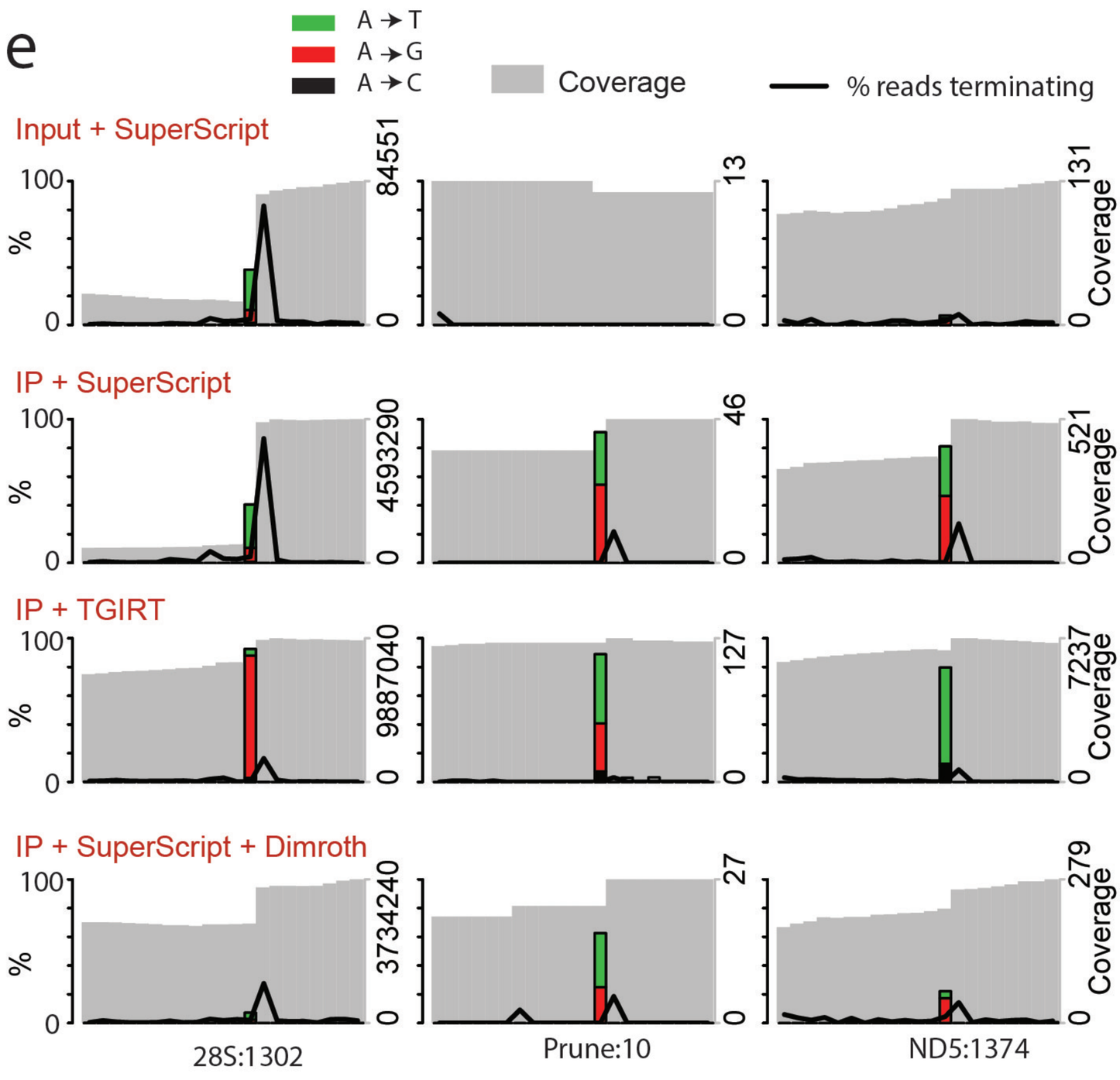



Figure 1

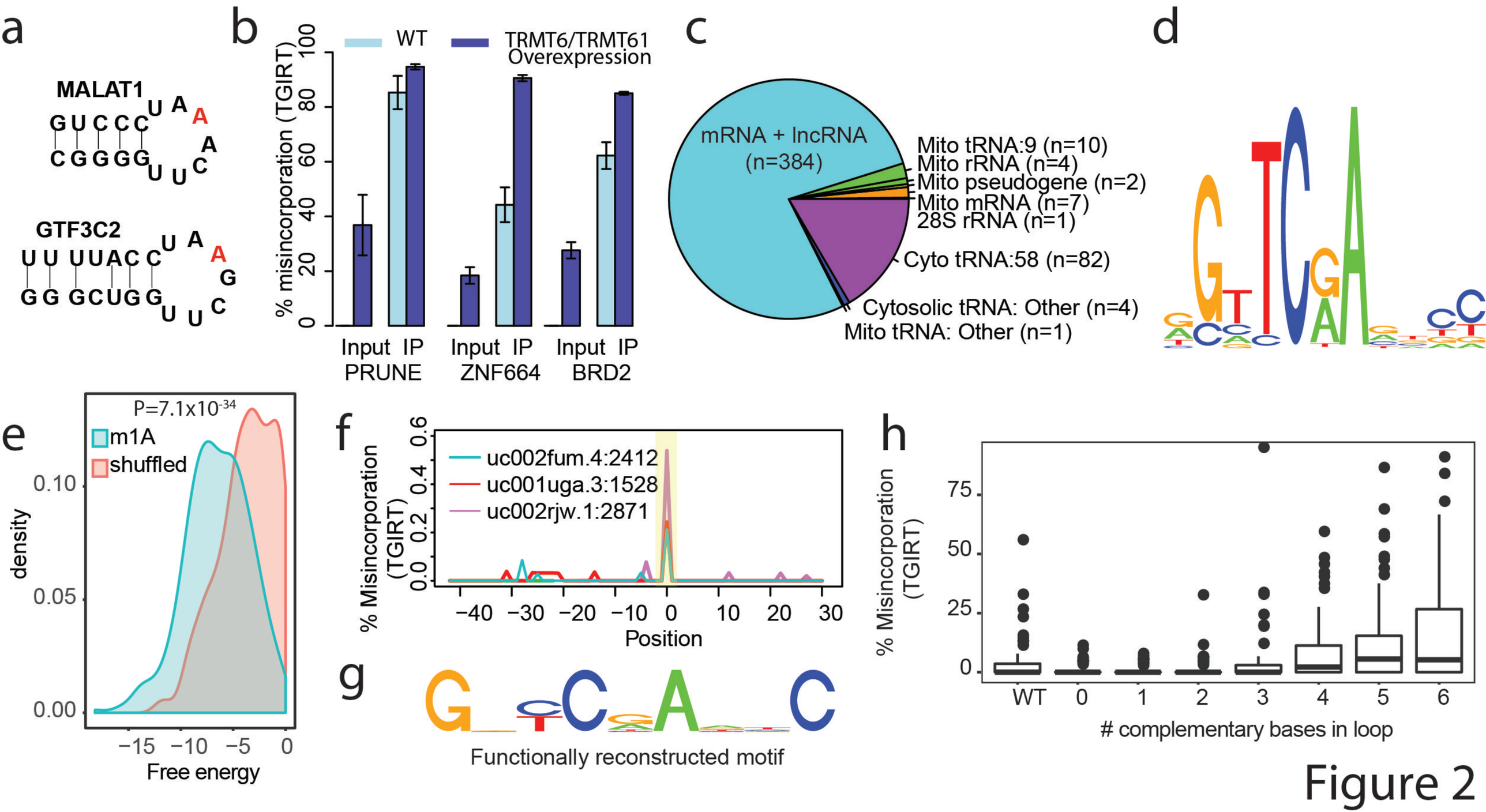

a
MALAT1
GUCCCUAA
CGGGGUUC
GTF3C2
UUUUACCUA
GGGCUGGUUC
b
WT
TRMT6/TRMT61 Overexpression
% misincorporation (TGIRT)
0
20
40
60
80
100
Input IP
PRUNE
Input IP
ZNF664
Input IP
BRD2
c
mRNA + lncRNA (n=384)
Mito tRNA:9 (n=10)
Mito rRNA (n=4)
Mito pseudogene (n=2)
Mito mRNA (n=7)
28S rRNA (n=1)
Cyto tRNA:58 (n=82)
Cytosolic tRNA: Other (n=4)
Mito tRNA: Other (n=1)
d
e
P=7.1x10-34
m1A
shuffled
density
0.10
0.05
0.00
−15 −10 −5 0
Free energy
f
% Misincorporation (TGIRT)
0.0 0.2 0.4 0.6
uc002fum.4:2412
uc001uga.3:1528
uc002rjw.1:2871
−40 −30 −20 −10 0 10 20 30
Position
g
Functionally reconstructed motif
h
% Misincorporation (TGIRT)
0
25
50
75
WT 0 1 2 3 4 5 6
# complementary bases in loop

Figure 2

a

GTeX Project

ND5 (chrM: 13710)

16S rRNA (chrM:2617)

% Misincorporation (SS-II)

Muscle, Bone Marrow, Heart, Adipose Tissue, Pancreas, Testis, Skin, Adrenal Gland, Breast, Liver, Nerve, Kidney, Colon, Lung, Small Intestine, Blood, Salivary Gland, Stomach, Spleen, Thyroid, Blood Vessel, Esophagus, Prostate, Uterus, Brain, Vagina, Bladder, Ovary, Pituitary, Pool

b

% misincorporation (TGIRT)

Muscle, Ovary

c

Petropoulos et al (SS-II)

% Misincorporation

Embryonic day (8 cell)

d

Yan et al (SS-III)

ND5 (chrM: 13710)

16S rRNA (chrM:2617)

Oocyte, Zygote, 2 cell, 4 cell, 8 cell, Morulae, Late blastocyst, hESC (p. 0), hESC (p. 10), Pool

e

% Misincorporation (TGIRT)

hours following actinomycin

f

% misincorporation (TGIRT)

Control, siTRMT10C, OE TRMT10C

g

GTeX, Ovary

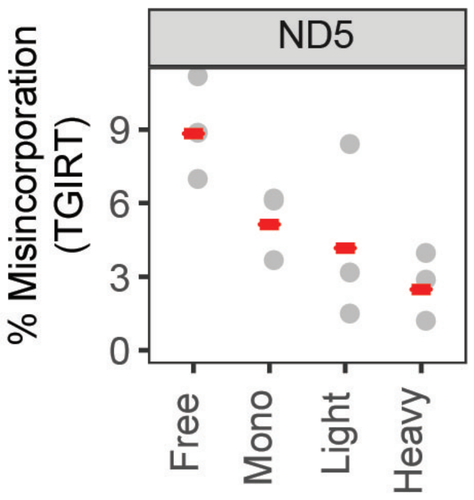
% Misincorporation (SS-II)

G (WT) A(SNP)

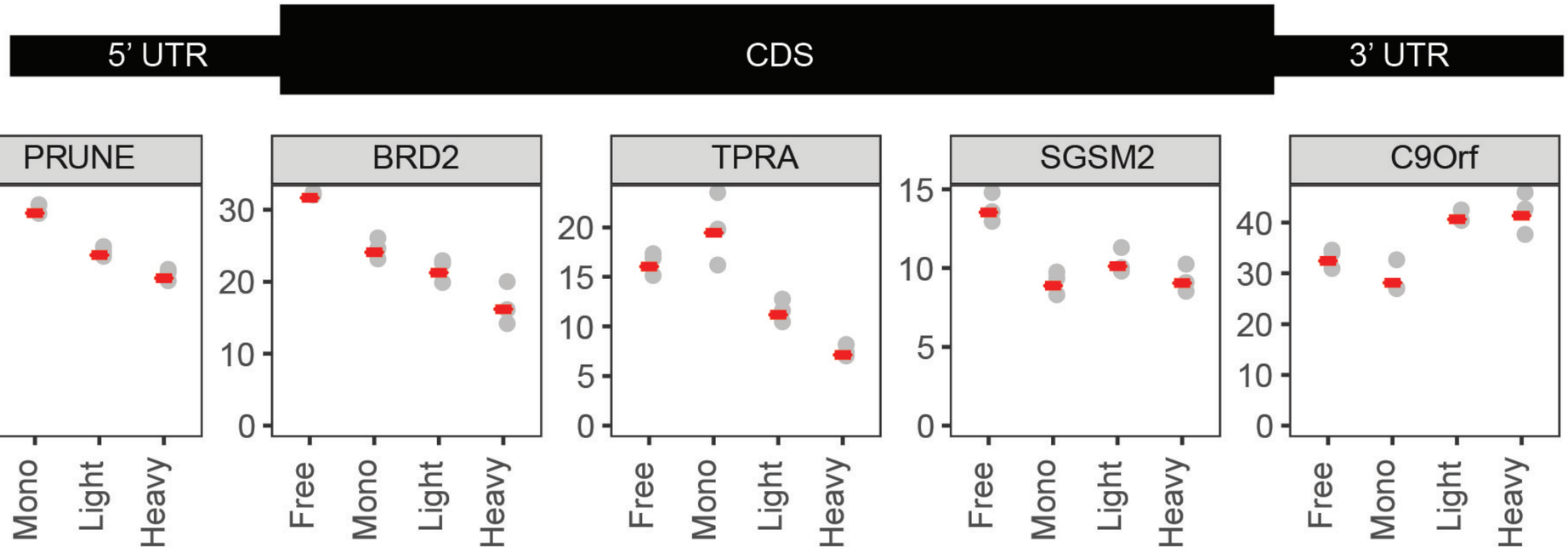
chrM:13708

Figure 3

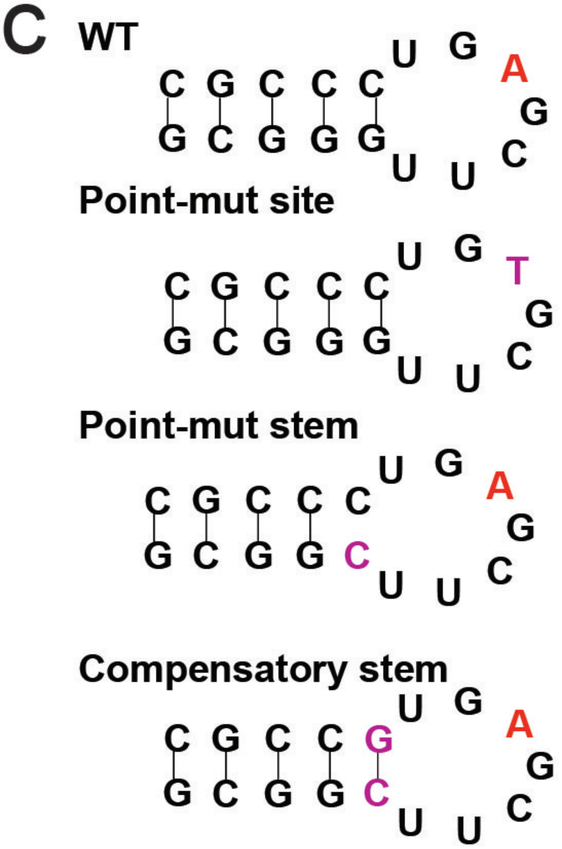
a



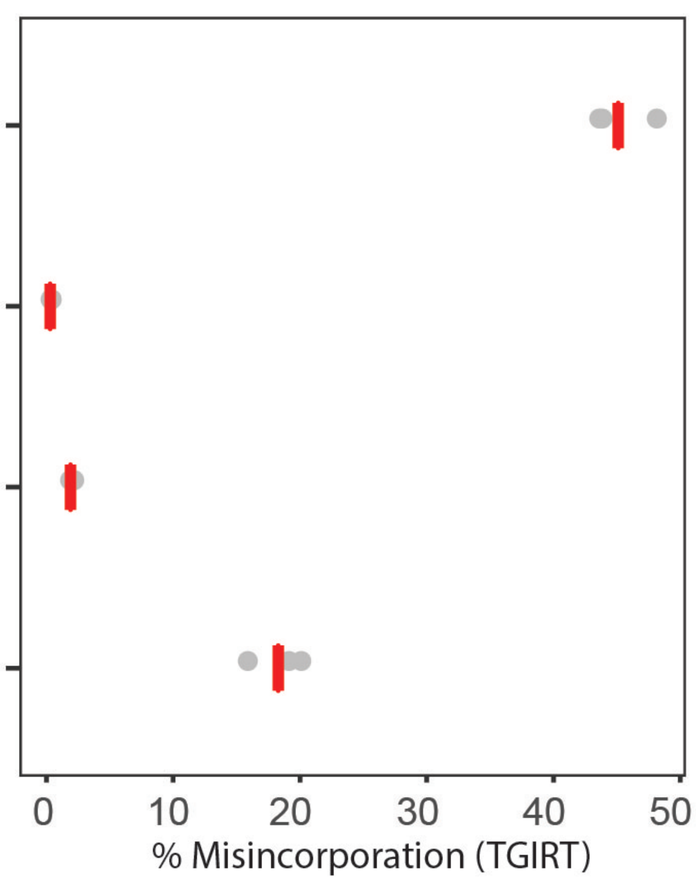

b




c




d




e

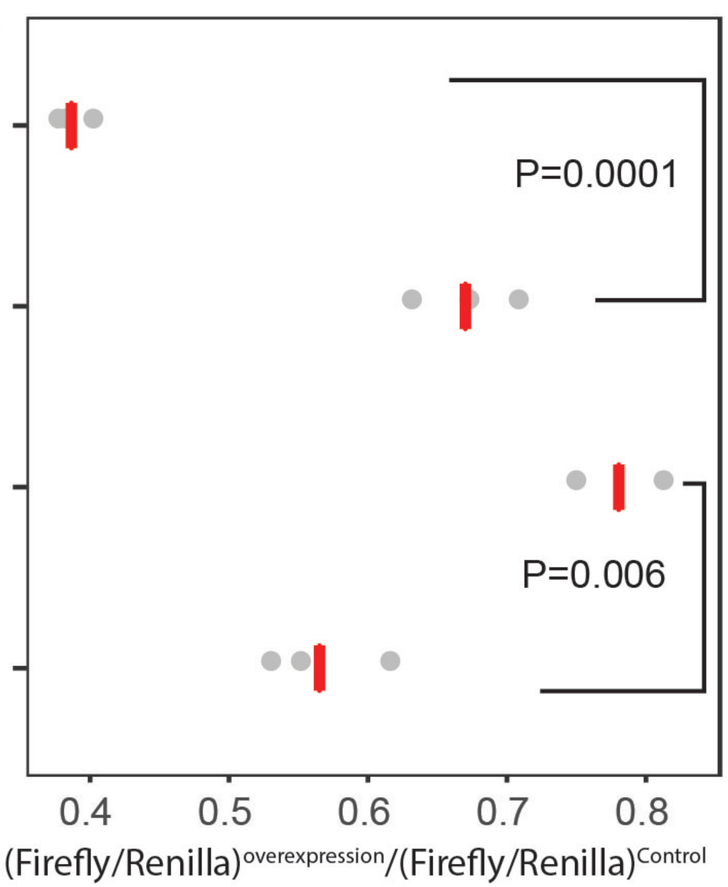



Figure 4