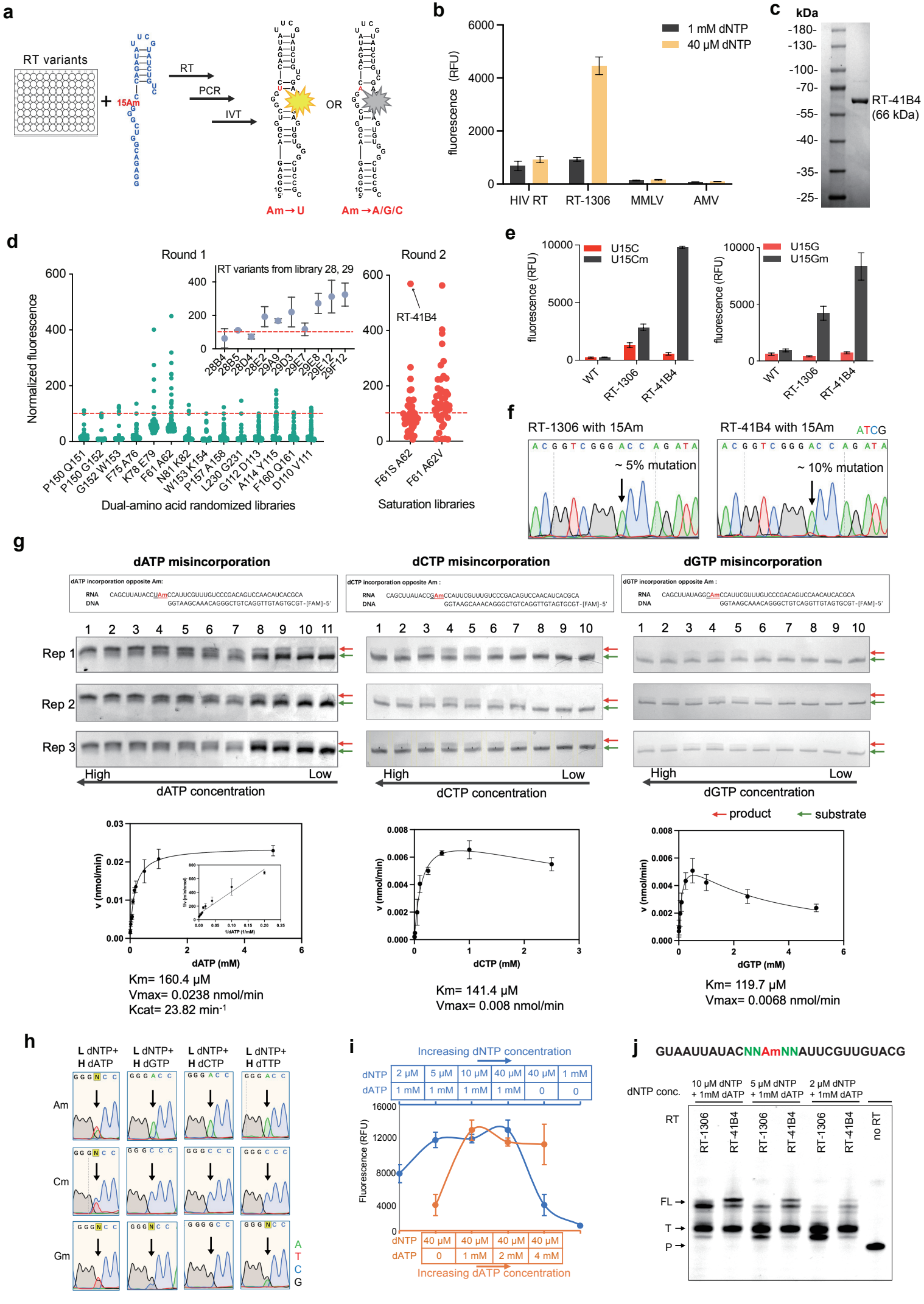
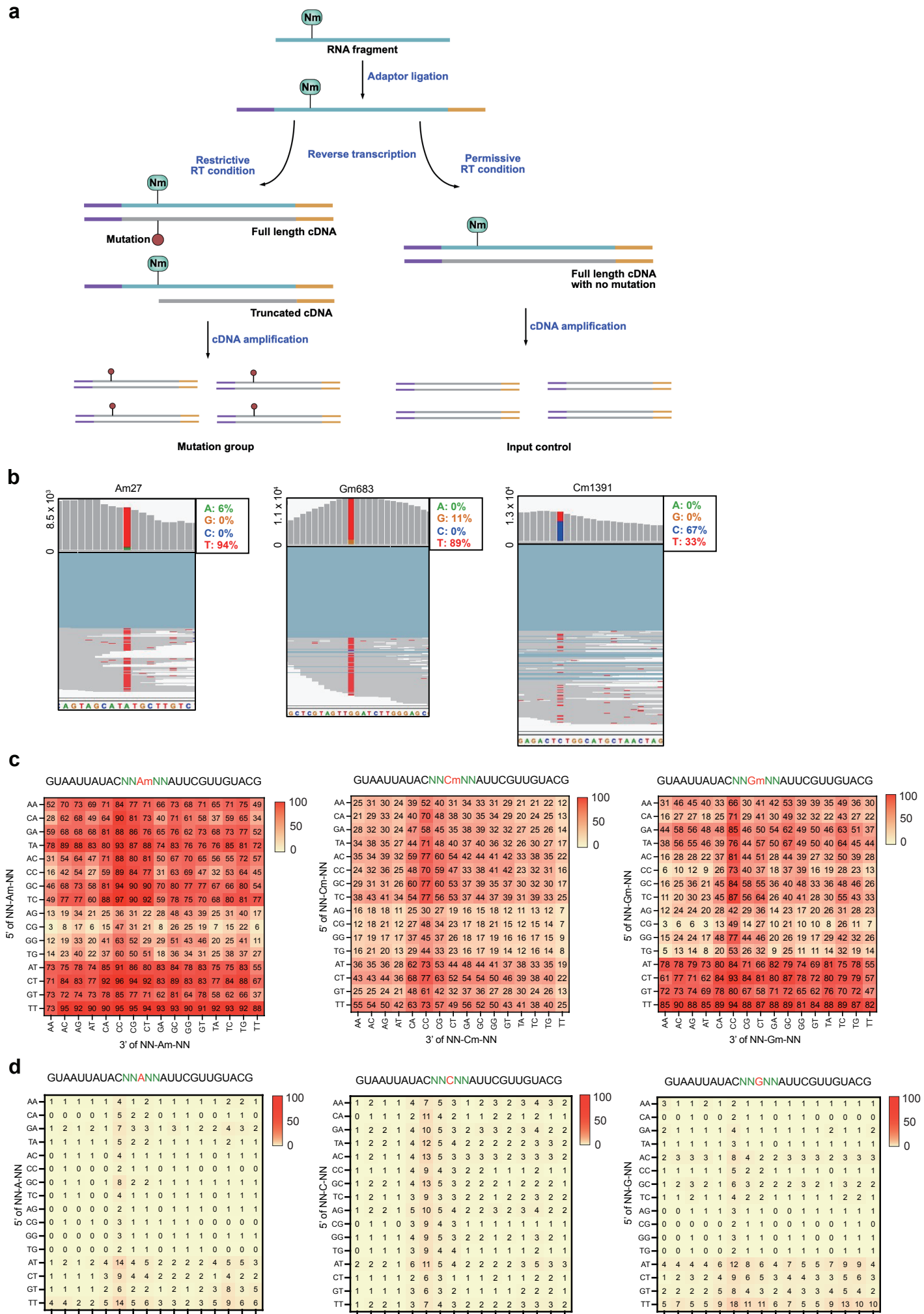
Supplementary information, Fig. S1

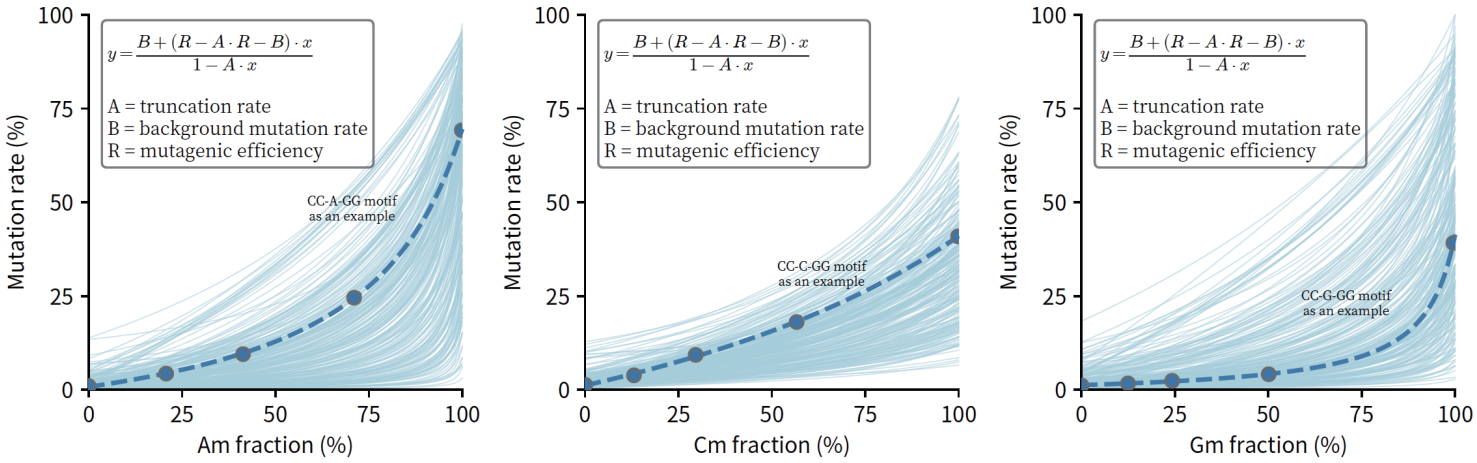


Supplementary information, Fig. S2

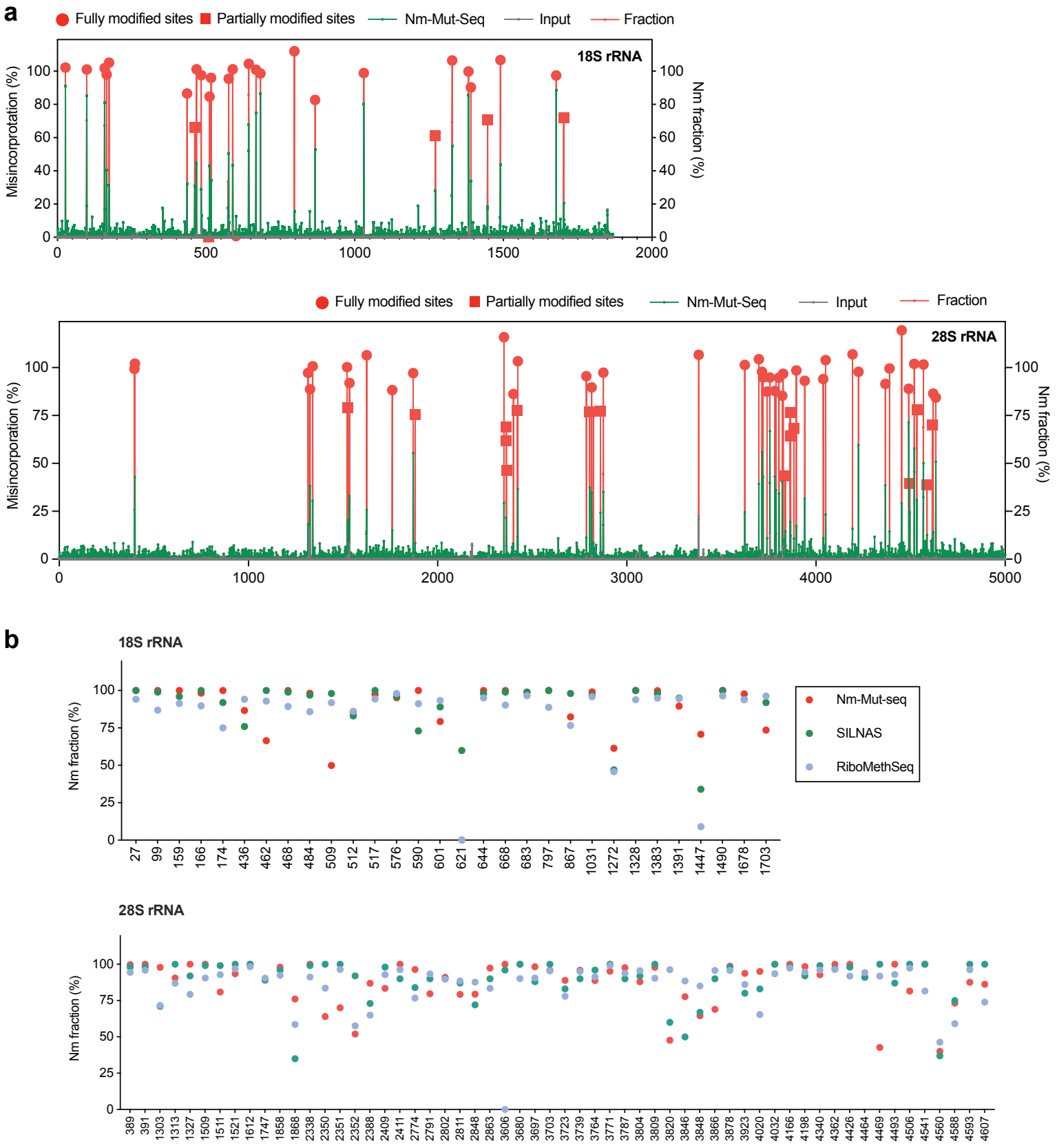


Supplementary information, Fig. S2-continued

e

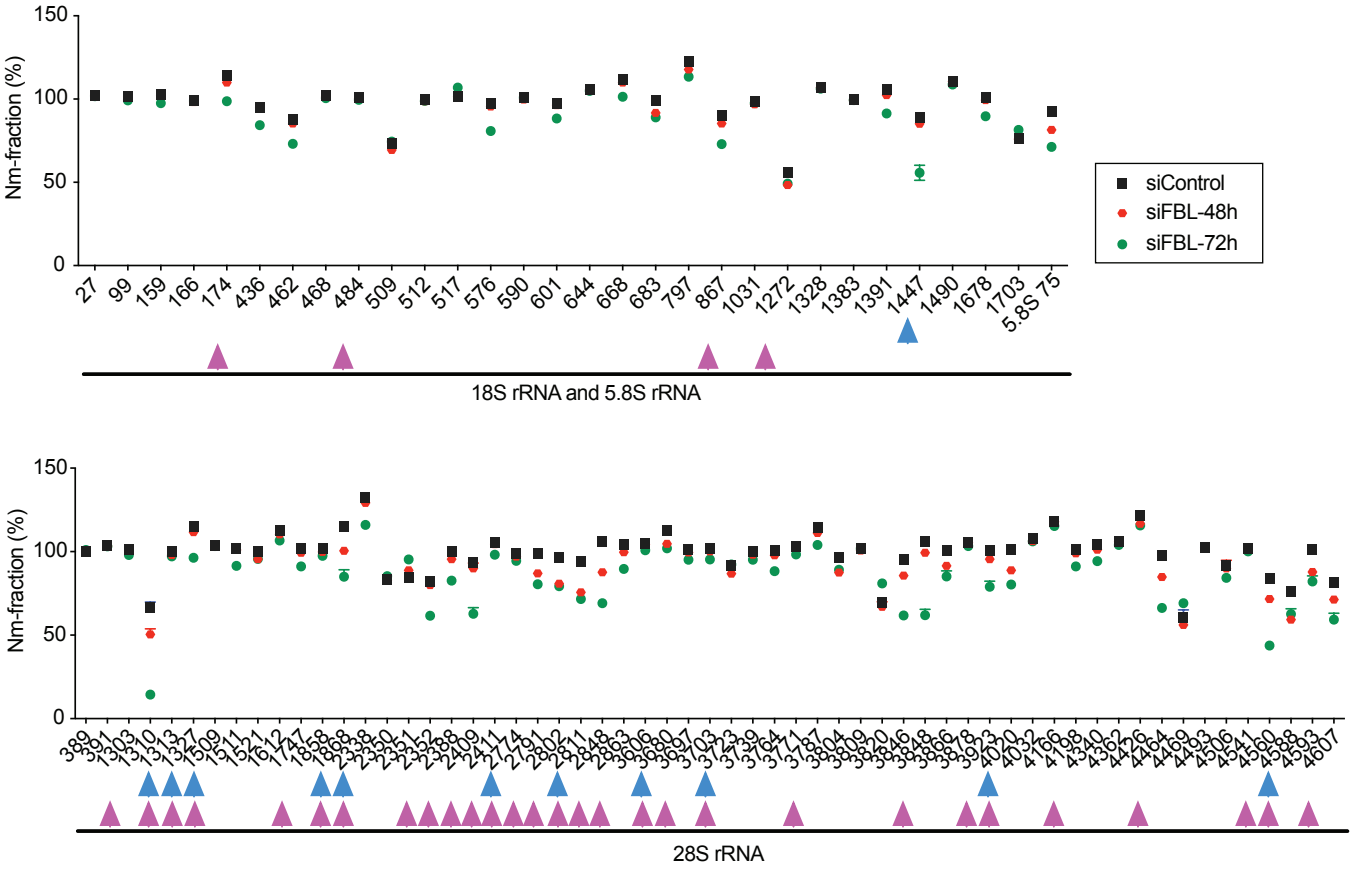


Supplementary information, Fig. S3

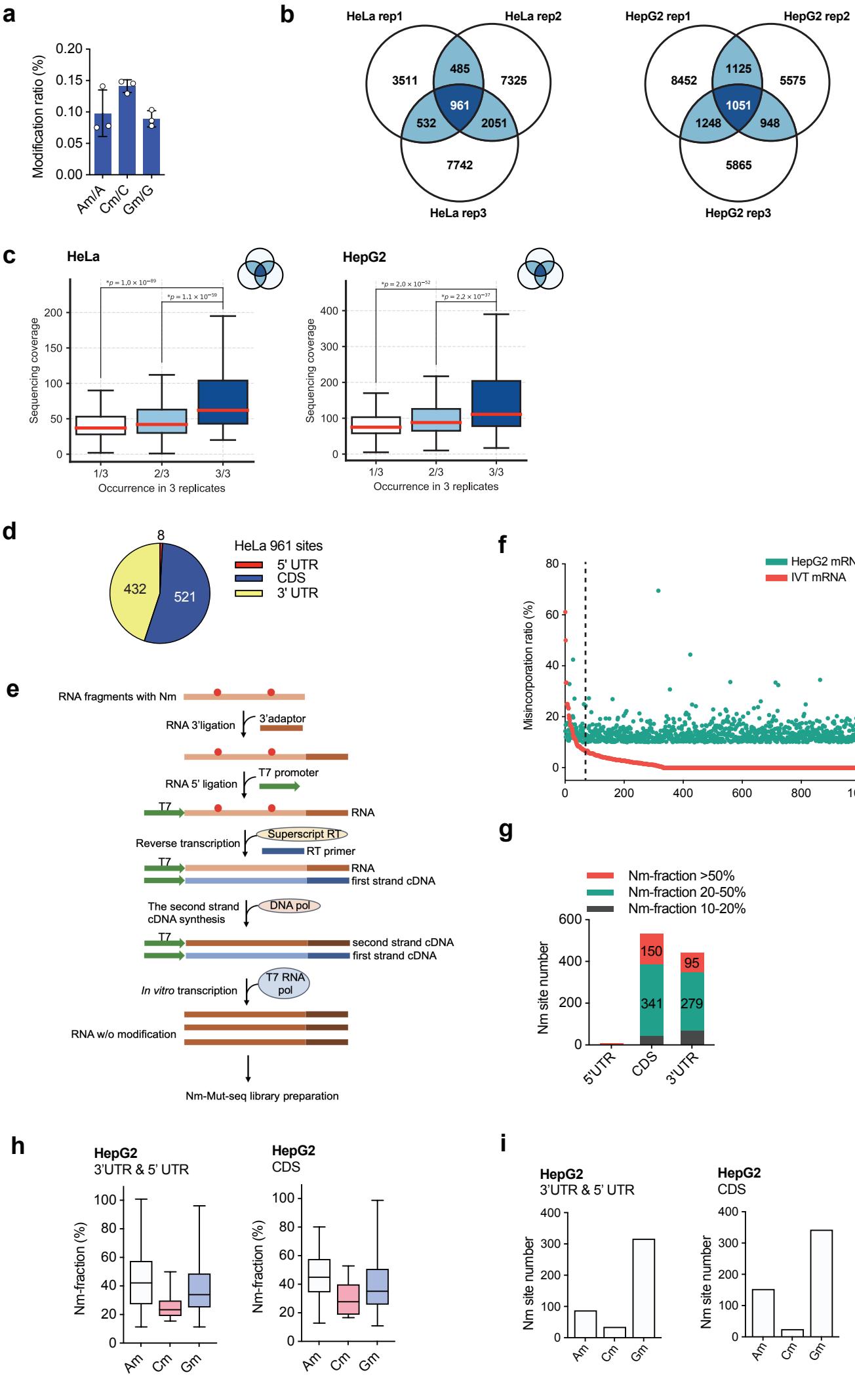


Supplementary information, Fig. S3-continued

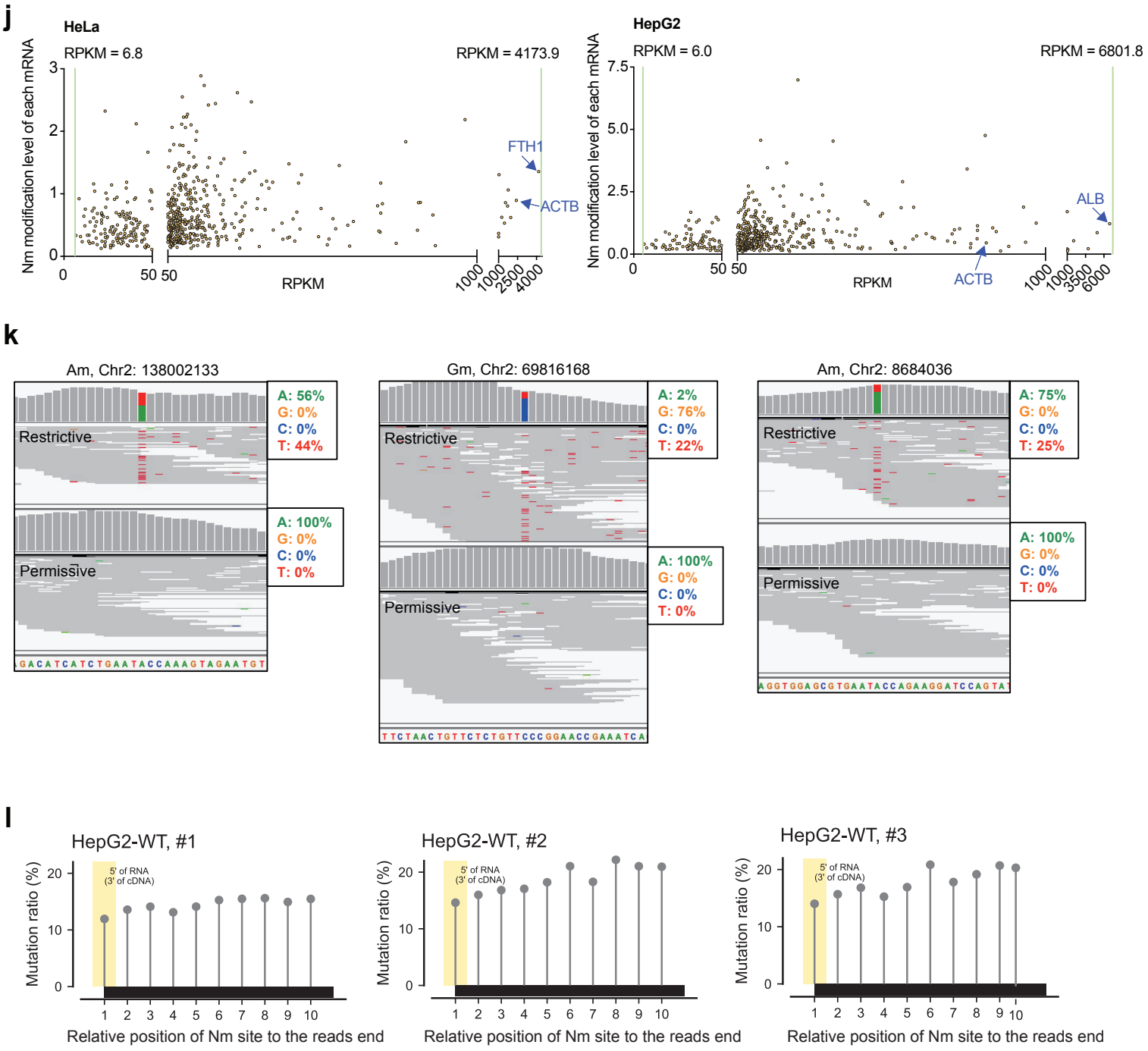
d



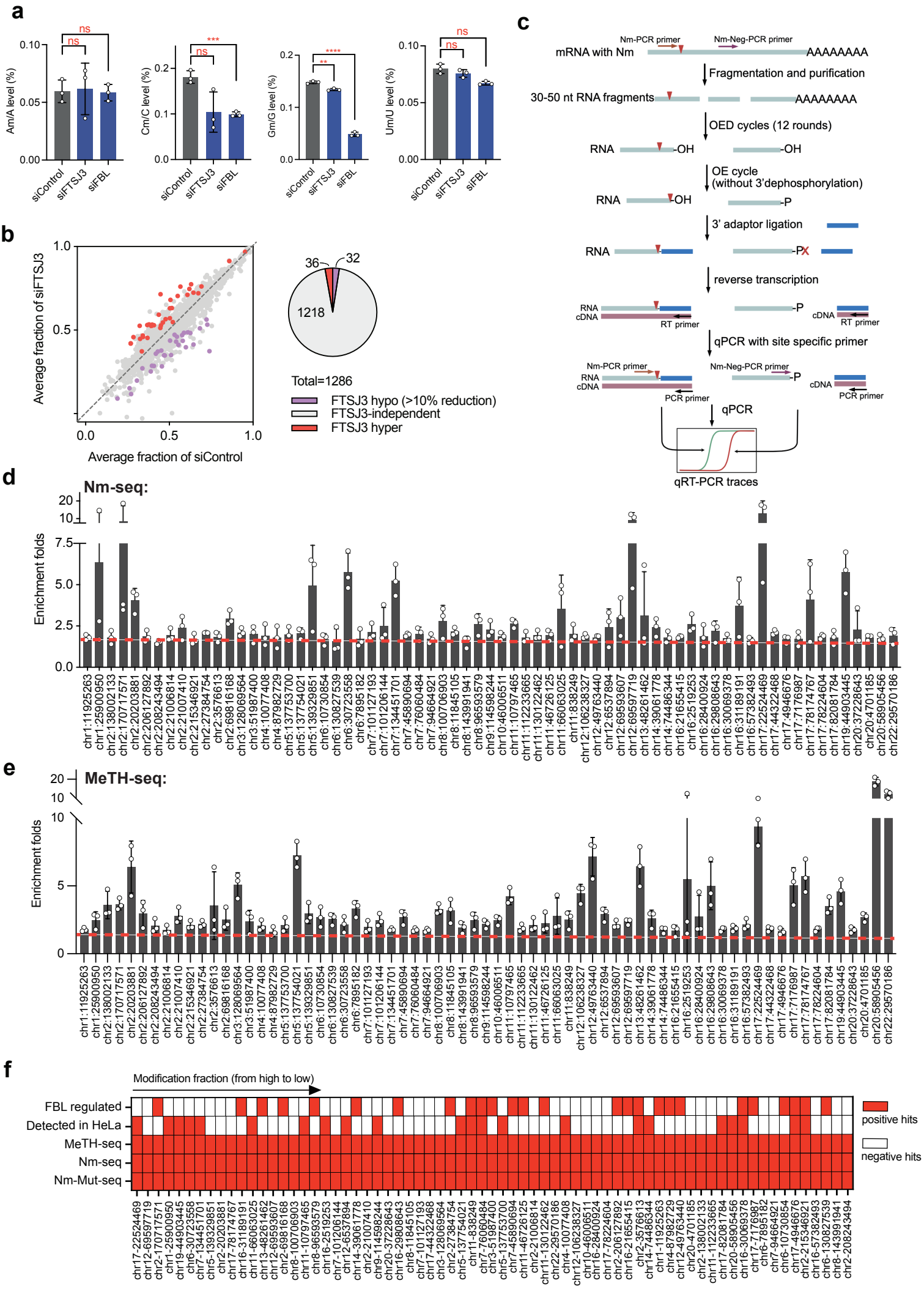
Supplementary information, Fig. S4



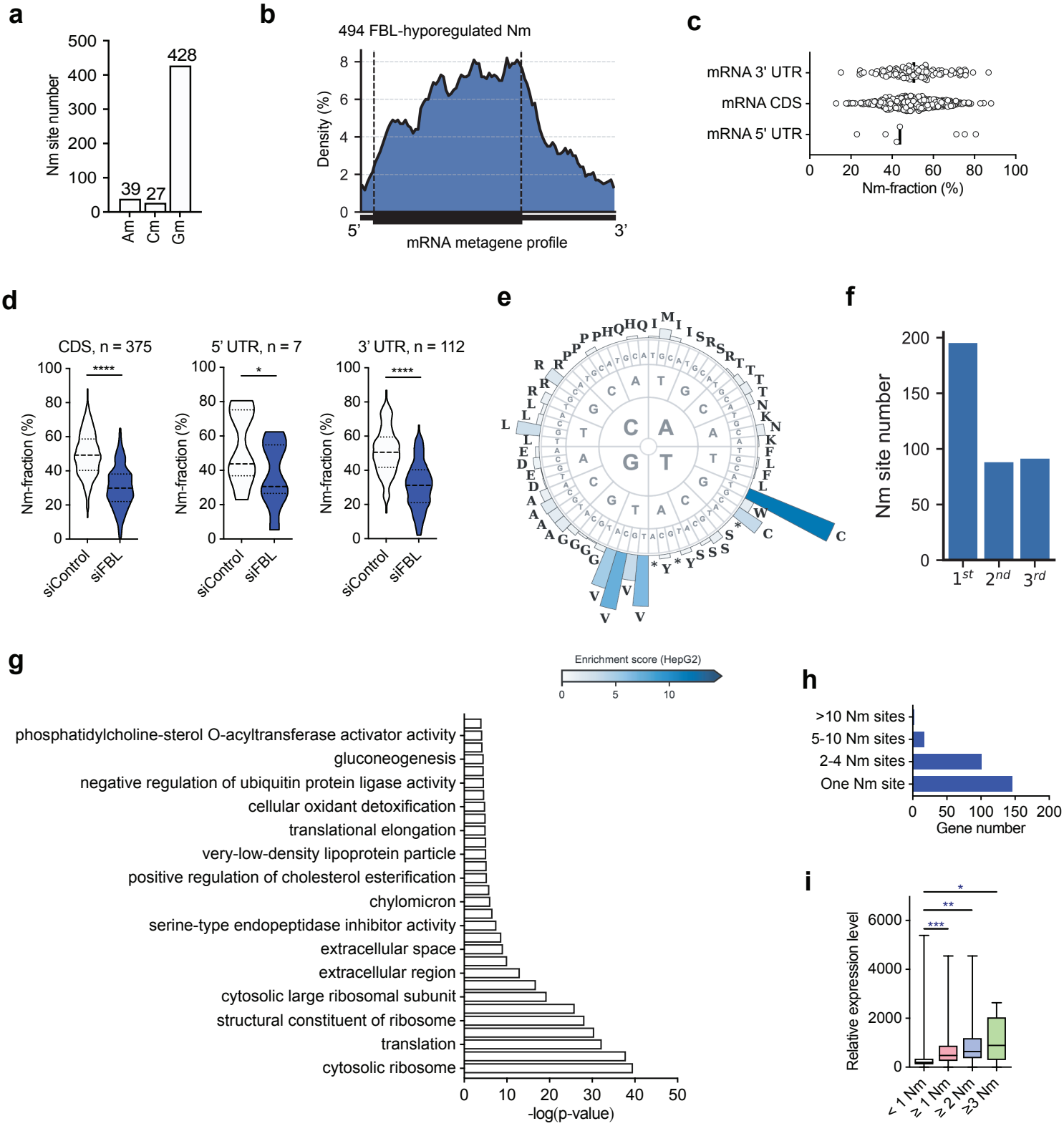
Supplementary information, Fig. S4-continued



Supplementary information, Fig. S5

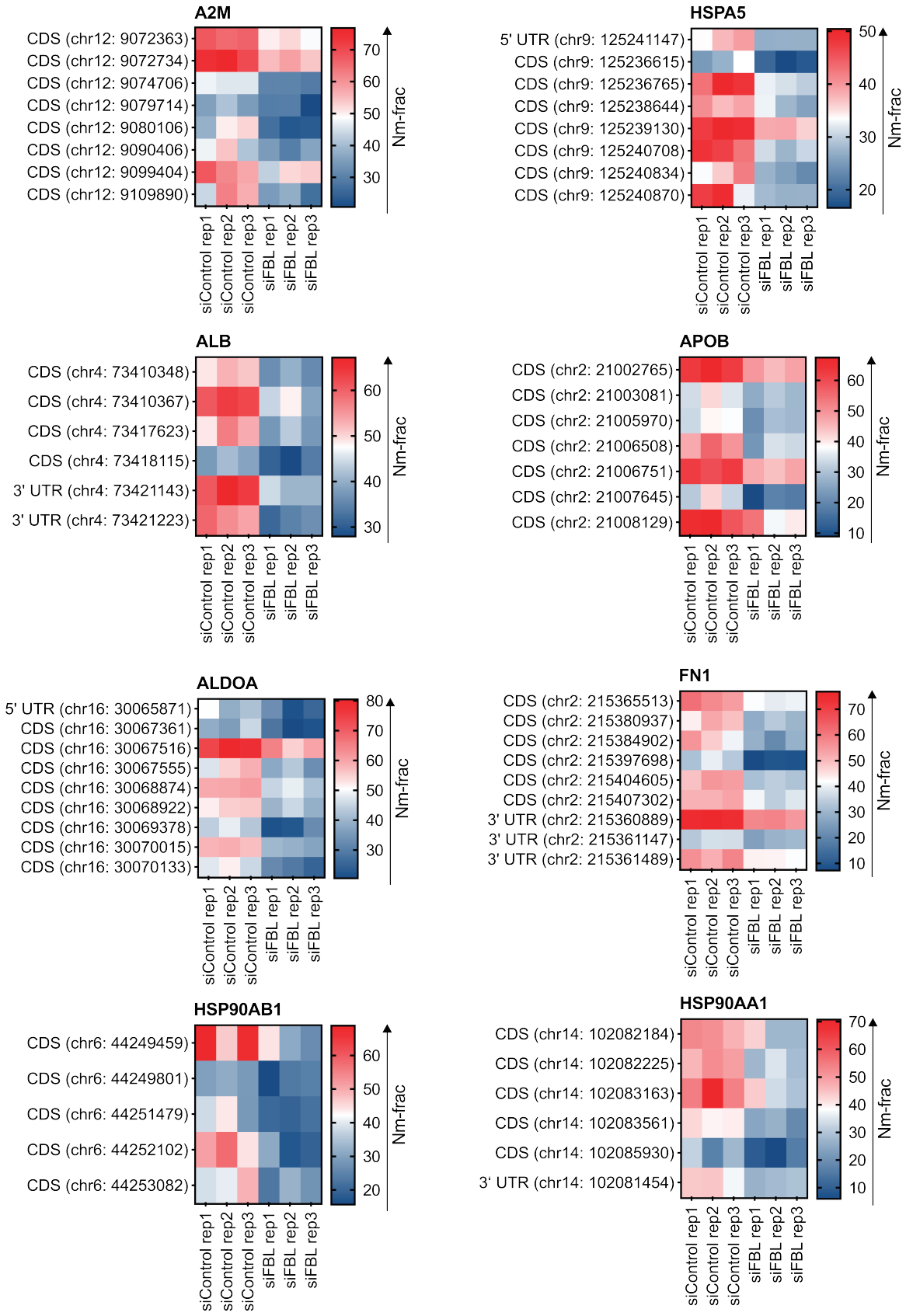


Supplementary information, Fig. S6

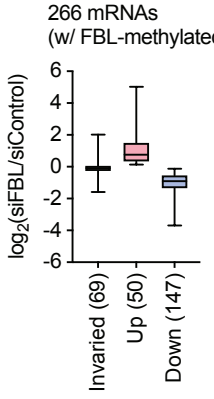


Supplementary information, Fig. S7

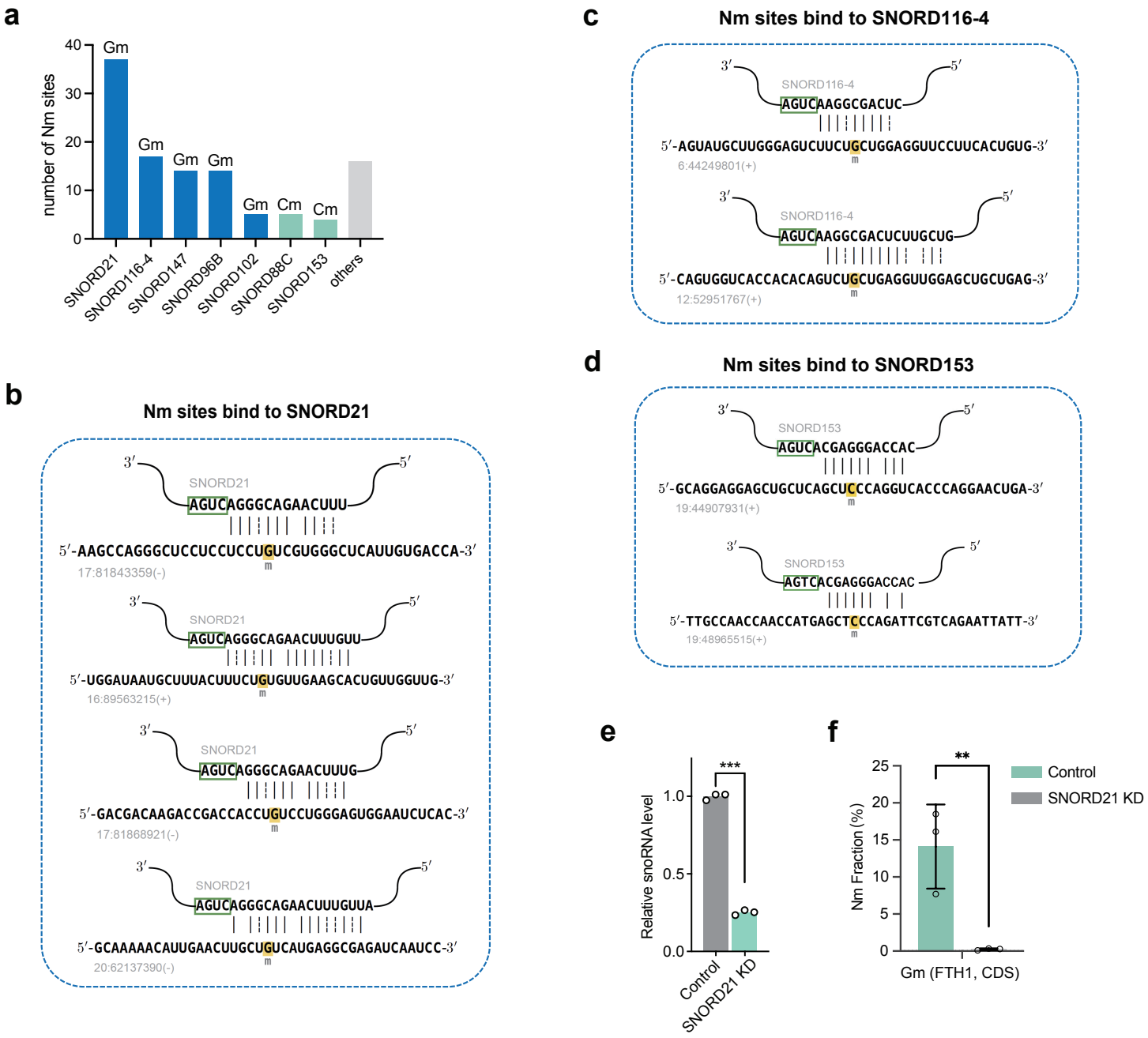
a



b



Supplementary information, Fig. S8



Supplementary information, Fig. S1 Screening of evolved RTs to map internal 2'-*O*-methylation modifications as misincorporation signatures. a Schematic outline of RT-PCR-IVT assay. The Am modification at position 15 of the Broccoli RNA probe is used to screen RT enzymes that are sensitive to Am modification. **b** Fluorescence intensity generated by four different RT enzymes using Am15 RNA substrate under 1 mM dNTP (black) or 40 μM dNTP (grey) concentration. **c** SDS-PAGE analysis of the purified, evolved RT-41B4. **d** Normalized fluorescence intensity measured during IVT of all variants from 16 RT mutation libraries (14 randomized libraries in the first round and 2 saturation libraries in the second round). The red dotted lines indicate the fluorescence response of the starting enzyme RT-1306. **e** Comparison of fluorescence intensity generated by HIV RT WT, RT-1306 and RT-41B4 pure protein with 33mer RNA probe. Left: with probes containing U15C (in red) or U15Cm (in grey) as substrate. Right: with probes containing U15G (in red) or U15Gm (in grey) as substrate. **f** Sanger sequencing of cDNA products generated by RT-1306 and RT-41B4 with 15Am RNA probe under 40 μM dNTP concentration. The 15Am position is highlighted. **g** Kinetic analysis of nucleotide misincorporation by the evolved RT-41B4. The steady-state kinetics of misincorporated elongation at Am site of the 5'-6-FAM-labeled Primer/template (shown on the top) were measured by examining dATP (Left) /dCTP (Middle) /dGTP (Right) concentration dependence on the steady-state rate. The elongation products were separated by denaturing Urea-PAGE gel and quantified across three replicates. The steady-state parameter K_m of different nucleotides (dATP /dCTP /dGTP) misincorporation at Am sites was similar; however, the V_{max} of dATP incorporation is ~3-4 fold higher than that of dCTP and dGTP, suggesting that dATP is the optimal substrate for RT-41B4 to induce misincorporations. **h** Sanger sequencing results of cDNA products generated by RT-41B4 with 15Am, 15Cm, and 15Gm RNA probes. Different nucleotide (dATP/dCTP/dGTP/dTTP) were examined for screening high misincorporation efficiency at Nm site in the RT step ('H' = 1 mM, 'L' = 40 μM). The modification position is indicated by an arrow. Results demonstrated that low dNTP with high dATP is the optimal condition for Am, Cm, and Gm detection. **i** Fluorescence intensity generated by RT-41B4 with different dNTP or dATP concentrations. **j** Results of primer

extension assay with the RNA probe containing randomized bases up and downstream of the Am site (-NNAmNN-). P, T, and FL are the labels for RT primers, truncated cDNA products, and full-length cDNA products, respectively.

Supplementary information, Fig. S2 Mapping 2'-*O*-methylation in human rRNA and the misincorporation patterns at Nm sites within different sequence contexts. a Detailed schematic illustration of Nm-Mut-seq library construction. **b** IGV coverage traces and alignment tracks of three annotated Nm sites in rRNA, exhibiting Nm signals captured as internal mutation signatures. **c** Heatmap of sequence-context-dependent mutation rates at Am, Cm, or Gm under "restrictive" RT condition in the –NN(Nm)NN- NGS libraries. **d** Heatmap of sequence context-dependent mutation rates at regular A, C, or G under "restrictive" RT condition in the –NN(N)NN- NGS libraries. **e** Sequence context-dependent calibration curves.

Supplementary information, Fig. S3 FBL knockdown affects the methylation level of known Nm sites in human rRNA, characterized by the stoichiometry change after sequence-context dependent calibration. a Misincorporation and Nm fraction detected by Nm-Mut-seq in human 18S rRNA and 28S rRNA. The mutation rates under "restrictive" RT conditions are shown in green, sites which passed the Nm detection criteria were further marked with Nm fraction (red). The annotated Nm sites are highlighted with red dots (round dots and square dots illustrate fully modified and partially modified Nm sites, respectively). **b** Comparison of the methylation levels at annotated Nm sites in rRNA detected by Nm-Mut-seq and previous methods (Red dots: Nm-Mut-seq; Green dots: SILNAS; Light blue dots: RiboMethSeq). **c** Left: western blot of siFBL 48-hour knockdown efficiency in HepG2 cells (normalized to Calnexin). Right: Western blot of siFTSJ3 48-hour knockdown efficiency in HepG2 cells (normalized to GAPDH). **d** Nm fractions of the full set of annotated Nm sites in rRNA before and after FBL knockdown. The figure shows Nm level for each Nm site from FBL-depleted HepG2 cells (red, 48 h; dark green, 72 h) *vs*. siControl (black). Values are presented as mean values ± SD (n = 3, biologically independent samples). The blue and pink

arrows indicate the Nm sites vulnerable to FBL depletion reported by Sharma *et al*.[1] and Erales *et al*.[2]

Supplementary information, Fig. S4 Nm-Mut-seq reveals Nm sites in human mRNA. a LC-MS/MS quantification of Am/A, Cm/C, and Gm/G ratios in wild-type HepG2 polyA$^+$ RNA. **b** Venn diagram showing the overlapped Nm sites among three biologically independent replicates, as detected by Nm-Mut-seq, in HeLa and HepG2 cells, respectively. **c** The sequencing coverage of Nm sites uniquely detected in one replicate (1/3), Nm sites detected in dual replicates (2/3), and overlapped Nm sites (3/3) detected by Nm-Mut-seq in polyA$^+$ RNA from HeLa (Left panel) and HepG2 (Right panel) cells, respectively. The overall sequencing depth of non-reproduced Nm hits is statistically lower than the reproduced Nm sites, indicating the replicate-specific hits may be attributed to variations in sequencing depth among samples. **d** Distribution of 961 HeLa Nm sites in three mRNA segments with the Nm fraction ≥ 10%. **e** Schematic outline of generating a modification-free RNA using *in vitro* transcription (IVT). **f** Mutation rates of the 1,051 HepG2 mRNA Nm site candidates detected by Nm-Mut-seq, in both wild-type HepG2 mRNA (shown in green) and *in vitro* transcribed (IVT) mRNA (shown in red). Of those, 75 sites showing a high mutation rate in the IVT library were determined to not be 2'-*O*-methylated (indicated as the sites on the left of the dash line). **g** Distribution of Nm site count and the Nm fraction in three segments of mRNA from HeLa cells. **h** Distribution of Nm stoichiometry in three types of Nm modifications (Am, Cm, and Gm) in UTR and CDS of HepG2 mRNA. **i** Distribution of modification site number in three types of Nm modifications (Am, Cm, and Gm) in UTR and CDS of HepG2 mRNA. **j** The correlation of Nm methylation level of each mRNA vs. RPKM value of the corresponding mRNA, in HeLa and HepG2 cells, respectively. The most abundant Nm-modified transcripts in each cell line, like FTH1 and ALB, are indicated with arrows respectively. ACTB, an endogenous housekeeping gene, was also indicated with arrows as an expression level reference gene. This result indicated that Nm-Mut-seq can detect Nm methylations on low-abundance mRNA that are ~400-fold less abundant than abundant transcripts like ACTB. **k** IGV coverage traces and alignment tracks of three putative Nm sites in mRNA, exhibiting Nm signals captured as internal mutation signatures. **l**

The mutation ratio distribution within the 10-nt region at the end of the mapped reads. The reads containing detected Nm sites were analyzed and the average mutation rate at each position was calculated among three biological replicates. The 5'-ends of the reads are highlighted in yellow color.

Supplementary information, Fig. S5 Validation of human mRNA Nm sites. a LC-MS/MS quantification of Am/A, Cm/C, Gm/G and Um/U ratios in HepG2 polyA$^+$ RNA, with the depletion of FBL or FTSJ3 *vs*. the control. P values were determined by paired two-tailed t-test; *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; data are presented as mean values ± SD. **b** Left: Scatter plot of Nm-Mut-seq data shows the reduced Nm fraction at Nm sites in FTSJ3 knockdown cells. The FTSJ3-dependent Nm sites must exhibit a notable >10% reduction in methylation after FTSJ3 depletion, with a P-value <0.05 for three biologically independent samples (3 siFTSJ3 *vs*. 3 siControl). Right: Pie chart of FTSJ3 hypo-regulated, hyper-regulated, and independent Nm sites. **c** Schematic outline of RT-qPCR validation assay based on Nm-seq. For each Nm candidate site validation, two site-specific RT-qPCR primers were designed (Nm-PCR-primer and Nm-Neg-PCR primer). Fragmented RNA is subjected to iterative oxidation–elimination–dephosphorylation (OED) cycles that remove 2′-hydroxylated nucleotides in the 3′-to-5′ direction to expose Nm sites at 3′ ends of fragments. Fragments ending with 2′-hydroxyl were blocked by an incomplete cycle (OE) and could not be ligated to the adaptor in the following adaptor ligation step. Conversely, RNA fragments ending with Nm could be ligated to the adaptor. During the subsequent RT-qPCR step, the ligated RNA produced a unique cDNA that could be amplified by qPCR with site-specific primer pairs; however, the non-Nm fragment could not be amplified because of the failure of adaptor ligation. The Nm sites were indicated in red color. **d** Validation results of the 69 Nm sites by Nm-seq-based RT-qPCR. Fold change values are presented as mean values ± SD; n = 3, biologically independent replicates. The red dotted line indicats the cutoff value. **e** Validation results of the 69 Nm sites by MeTH-seq-based RT-qPCR. Fold change values are presented as mean values ± SD; n = 3, biologically independent replicates. The red dotted line indicats the cutoff value. **f** 69 highly confident Nm

sites detected in HepG2mRNA with different verification approaches. Positive hits (red) and negative hits (white) are indicated in the figure.

Supplementary information, Fig. S6 FBL regulates Nm modifications on hundreds of mRNAs in HepG2 cells. a The Am, Cm, and Gm composition of 494 FBL hypo-regulated Nm sites (above 20% reduction in Nm fraction). **b** The metagene profiles of 494 FBL hypo-regulated Nm sites in HepG2 mRNA. **c** Nm fraction distribution of 494 FBL hypo-regulated Nm sites (above 20% reduction in Nm fraction). **d** 494 FBL hypo-regulated Nm sites distributed in CDS, 5' UTR and 3' UTR of HepG2 mRNA. P values were determined by paired two-tailed t-test; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; data are presented as mean values ± SD. **e** Codon enrichment of 494 FBL hypo-regulated Nm sites in HepG2 mRNA. The enrichment was calculated as the ratio between the number of individual Nm codon occurrences and the total number of each codon in mRNA. Optimal codons, which are those that are more frequently modified by 2'-*O*-methylations, are denoted by dark blue bars. **f** The distribution of 494 FBL hypo-regulated Nm sites within three positions of one specific codon, in HepG2 mRNA. **g** Enriched GO clusters of 266 adequately expressed HepG2 genes carrying 494 FBL hypo-regulated Nm sites. The clusters are ranked by p-value and shown as top 20. **h** The transcripts with one or more Nm sites. Nearly 50% of the transcripts contain more than one Nm site per mRNA. **i** The relative expression level of the transcripts grouped by different Nm modification strengths. A higher expression level was observed when the mRNA displayed a higher Nm modification strength. The x-axis shows the Nm modification sites number of each mRNA.

Supplementary information, Fig. S7 A group of human mRNAs contains multiple Nm methylations per mRNA. a Heatmaps showing the changes in Nm fraction of representative abundant human mRNA genes, before and after FBL knockdown. **b** 266 adequately expressed mRNA genes containing FBL-regulated Nm sites were used for studying the expression level change upon FBL depletion. 147 out of these 266 genes showed decreased expression levels

after FBL depletion. P values were determined by paired two-tailed t-test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; data are presented as mean values ± SD.

Supplementary information, Fig. S8 snoRNA-guided Nm methylation in human mRNA. **a** Top seven snoRNAs predicted to bind a portion of FBL-regulated mRNA Nm sites. **b-d** predicted base pairing between snoRNAs and putative Nm sites in mRNA, showing SNORD21, SNORD116-4, and SNORD153, respectively. **e** Normalized SNORD21 expression level before and after SNORD21 knockdown. **f** Nm fraction changes of one Nm site (Gm-Chr11: 61967318) in FTH1 mRNA before and after SNORD21 knockdown. **g** Predicted base pairing between SNORD21 and putative Nm sites in mRNA.

SUPPLEMENTARY INFORMATION

MATERIALS AND METHODS

Enzyme preparation:

RT enzymes AMV and MMLV were purchased from New England Biolabs, Inc., and Promega Corporation, respectively. The RT-1306 and RT-41B4 enzymes were overexpressed as pET30a N-terminal His_6-tagged fusion proteins in *E.coli* BL21(DE3). The detailed protocol for enzyme overexpression and purification is described in the Extended protocol.

The T7 RNA polymerase was prepared using our protocol reported previously[3].

DNA and RNA oligonucleotides preparation:

DNA primers and RNA oligonucleotides for *in vitro* assays, such as 33 mer U15Am RNA, U15A RNA, U15Cm RNA, U15C RNA, U15Gm RNA and U15G RNA were purchased from Integrated DNA Technologies, and purified by RNase free HPLC. The 35mer -NNNmNN- RNA oligonucleotides used in Nm-Mut-seq library preparation were also ordered from IDT, with RNase free HPLC. The siRNAs (siFTSJ3, siFBL and siControl) for methyltransferase gene knockdown were purchased from Qiagen (Supplementary information, Table S11).

Preparation and Screening of RT random mutation library:

RT mutation library and cell lysate preparation. Mutant RT polymerase libraries were prepared as described by Zhou *et al* [3]. Briefly, the RT-1306 protein overexpression vector containing NNK to replace the codon of targeted amino acids was constructed by Gibson assembly. Single colonies from transformation were picked and grown in 96-deepwell plates to induce and overexpress RT variants. The cells were then collected and lysed, and crude cell lysate was used for RT screening.

RT variant screening assays. RT screening assays were performed as described previously[3]. Briefly, the reverse transcription reaction was carried out with 0.5 μM RT primer and 15Am RNA probes, 40 μM each dNTP and 3 μL cell lysate in 1X RT reaction buffer. The reaction mixture was incubated at 37°C for 1 h followed by 80°C for 10 min. 1 μL of RT product was then subjected to 10 μL PCR reaction for 15 cycles amplification. Finally, 7 μL of the PCR product was added to the *in vitro* transcription reaction, containing 2mM rNTP, 25 mM Mg^{2+}, 10mM DTT, 50 μM DFHBI-1T and 0.5 μL T7 RNA polymerase in 1X T7 reaction buffer.

Steady-state kinetics of nucleotide misincorporation opposite Am sites in the presence of RT-41B4:

RT reaction mixture (10 μL volume) containing 500 nM RT-41B4 enzyme, 5 μM 5'-6-FAM-labeled Primer/template, 1X RT Reaction Buffer and different concentrations of dATP/dCTP/dGTP. The reaction was performed in triplicates and initiated by adding RT enzyme at 37°C for 10 min, followed by heating at 80°C for 5 min to inactivate the enzyme. The primer extension products were separated by denaturing polyacrylamide gel electrophoresis (15% acrylamide/bisacrylamide (19:1), 8.0 M urea). The amount of primer extension product was quantified by ImageJ software. The K_m and V_{max} values were estimated by nonlinear regression using Prism software (version 9.1.1).

Mutation rate-Nm methylation fraction calibration curves:

Equimolar amounts of three 35-mer RNA probes with -NNAmNN-, -NNCmNN- and -NNGmNN- were mixed together for the "100% Nm" standard. RNA oligos containing -NNANN-, -NNCNN- and -NNGNN- were pooled for the "0% Nm". The "100% Nm" and "0%

Nm" standards were combined to generate five oligo mixtures at different methylation levels (100% Nm, 75% Nm, 50% Nm, 25% Nm, 0% Nm). All oligo mixtures were subjected to Nm-Mut-seq and the mutation rate pattern of each sequence context was analyzed and a fitting curve was plotted based on the relationship of observed mutation rate and methylation fraction. The observed mutation rate *y* and Nm fraction *x* can be expressed in the following equation:

$$y = \frac{B + (R - A \cdot R - B) \cdot x}{1 - A \cdot x}$$

Sequence context dependent mutagenic efficiency (R) is a parameter of the model (formula), which is directly related to the actual mutation rate detected from synthetic oligo (-NNNmNN-) libraries. The background rate (B) is directly related to the actual mutation rate detected from "0% Nm" standard and the truncation rate (A) represents the pause frequency of RT enzyme in each sequence context containing Nm modification. The most opitmal combination of these parameters was determined by Least Squares Fitting, for each motif**.** The numerator of the formula represents the number of sequences that exhibit mutations, while the denominator subtracts the product of the truncation rate (A) and the modification ratio (x), indicating how many sequences were lost during the RT. The parameters of A, B and R for each sequence context are provided in Supplementary information, Table S3.

Quantification of Nm by LC-MS/MS:

Roughly 150-200 ng polyA+ enriched mRNA was digested with nuclease P1 (Sigma, 1U) in 35 μL 1× reaction buffer containing 10 mM NH4Ac (PH=5.3), 2.5 mM ZnCl2 at 37 °C for 2 h. Then rSAP treatment was carried out by adding 1 μL of thermosensitive alkaline phosphatase (NEB) and 4 μL 10 × FastAP buffer (Thermo Fisher) and incubating at 37 °C for another 2 h. After incubation, the sample was diluted to 60 μL and filtered by a 0.22 μm filter (4mm diameter, Millipore). The chromatographic instrument was a 1290 series LC coupled with an Agilent 6410 QQQ triple-quadrupole LC mass spectrometer which equipped with an electrospray ionization interface. The filtered sample was injected into a C18 reverse phase column (Agilent, ZORBAX Eclipse XDB-C18, Rapid Resolution Ht, 2.1× 50 mm, 1.8 μm) and then analyzed by QQQ in positive electrospray ionization mode. For the quantification, a mock

control sample containing only digestion buffers and enzymes was used for baseline signal deduction. The 2'-O-methyl containing nucleosides were quantified with retention time and the nucleoside-to-base ion mass transition of 268-136 (A), 282-136 (Am), 244-112 (C), 258-112 (Cm), 284-152 (G), 298-152 (Gm). Nucleoside quantification was performed in comparison with the standard curve, obtained from pure nucleoside standards running on the same batch of samples.

Cell culture and antibodies:

Cell culture. HeLa and HepG2 were used for Nm analysis in this study. HeLa cells were maintained in DMEM medium (4.5 g/L glucose) with 10% FBS and 1% penicillin/streptomycin. HepG2 was maintained in low glucose DMEM medium (1 g/L glucose) with 10% FBS and 1% penicillin/streptomycin.

Antibodies for Western blot: mouse monoclonal anti-FBL antibody (Proteintech, 66985-1-lg, WB 1:2000), rabbit polyclonal anti-FTSJ3 antibody (Bethyl A304-199A, WB 1:2000), Calnexin-HRP antibody (Santa Cruz, sc-46669 HRP, WB, 1:1000), GAPDH-HRP antibody (Proteintech HRP-60004, WB 1:5000).

NGS Sequencing library preparation:

Synthetic oligonucleotide library: The Illumina sequencing libraries were prepared using a NEBNext Small RNA Library Prep Set. 1 μg of synthetic RNA oligo containing different methylation levels were end repaired/5'-phosphorylation by PNK treatment (T4 polynucleotide kinase, Thermo Fisher): 40 ul RNA was mixed with 5μL 10× T4 PNK buffer and 0.5 μl PNK enzyme, incubated at 37°C for 30 min, then add 5 μL 10 mM ATP into the mixture and incubated another 30 min. The end-repaired RNA fragments were cleaned up with an OCC kit (Zymo Research). Starting with the 3' adaptor ligation, the RNA fragments were incubated with 0.5 μL 12 μM customized adaptors containing 5-mer, randomized nucleotide sequence (NNNNN), 3' ligation buffer and enzyme at 25°C for 2 h. Then the SR RT primer was added to the ligation mixture to hybridize the 3' adaptor. The 5' adaptor ligation was then carried out by adding 5' adaptor, 5' ligation buffer and enzyme to the previous reaction mixture and

incubating at 25°C for 2 h followed by 16°C for 4 h. The resultant RNA was cleaned up by an OCC kit, eluted with 20 μL RNase-free water and subjected to reverse transcription. Two independent RT reactions each containing 10 μl of RNA from the last step were conducted using evolved RT-41B4 and optimized dNTP concentration. One RT reaction ("input") was conducted using the condition of 1mM dNTP which does not generate mutations on Nm sites while the other reaction (as "Nm-Mut-seq") with 2 μM dNTP + 1 mM dATP installs mutations on Am, Cm and Gm sites. The RT reactions were performed at 37°C for 1 h, followed by 10min at 80°C, to inactivate the RT enzyme. In the last step, PCR was used to amplify the cDNA products which were further purified, and size selected (160-200bp) by gel recovery.

Biological RNA library: The step-by-step protocols for Nm-Mut-seq are described in the Extended protocols.

rRNA library preparation: isolated RNA was fragmented into ~50 nt and ligated with sequencing adaptors at both ends prior to reverse transcription. The RNA was then split in half and subjected to reverse transcription by RT-41B4 under either Nm-optimized condition (restrictive condition), to generate misincorporation of dATP opposite of Nm sites, or normal RT condition, which do not generate mutations (permissive condition). Importantly, only the full-length cDNA from the RT step is amplified by PCR, which eliminates signal generated by truncation.

PolyA$^+$ RNA library preparation: For mRNA library preparation, ~50 μg of total RNA extracted from cell lines or tissue was used as the starting material. We performed the pre-treatment steps of RNA following the protocol described previously[3]. Briefly, we first carry out DNaseI treatment on total RNA, then purified and removed small RNA species using the MEGAclear Transcription Cleanup Kit (Invitrogen). In the next step, we performed two-rounds of polyA+ RNA enrichment with 100 μg of the resulting RNA from the last step using Dynabeads^{TM} mRNA purification Kits (Thermo Fisher) followed by two-rounds of rRNA depletion using RiboMinus^{TM} Eukaryote Kits (Thermo Fisher). The yield of mRNA in this step is roughly 0.5 μg. Next, the mRNA was fragmented to 50-80nt by heating at 95°C for 8 min in 1M NaCO3/NaHCO3 buffer (PH 9.2), then the mRNA fragments were cleaned up by Oligo Clean

& Concentrator (Zymo Research) and end-repaired using PNK treatment. In the 3' and 5' adaptor ligation step, 50-100ng of RNA was used to ligate the customized adaptors which contains 5-mer randomized nucleotides (NNNNN). After ligation, the RNA was cleaned up again by Oligo Clean & Concentrator (Zymo Research) and eluted with 40µl of RNase-free water. We then performed two independent RT reactions each with 10 µl of RNA from the last step using evolved RT-41B4 and optimized dNTP concentrations. One of the RT reactions was conducted using the condition of 1 mM dNTP which does not generate mutations on Nm sites while the other reaction with 2 µM dNTP + 1 mM dATP would install mutations on Am, Cm and Gm sites. The RT reactions were performed at 37°C for 1 h, followed by 10 min at 80°C, to inactivate the RT enzyme. In the last step, PCR was used to amplify the cDNA products which were further purified, and size selected (160-200bp) by gel recovery. All libraries were sequenced on Illumina NextSeq 550 with single-end 75 bp read length.

Methyltransferase siRNA knockdown:

HepG2 cells were transfected using siRNAs purchased from Qiagen targeting FBL (Qiagen SI04164951) or FTSJ3 (Qiagen SI03197509) or non-targeting sequences (Qiagen, AllStars Neg. Control, SI03650318). To prepare the siRNA/RNAiMAX solution for a 15 cm plate, 120 pmol of siRNAs were diluted in 1.75 mL of OPTI-MEM and 50 µL of RNAiMAX (Thermo Fisher) was diluted in 1.75 mL of the same media in a separate tube. The siRNA and RNAiMAX were mixed together and incubated at room temperature for 15 min. The resulting 3.5 mL transfection solution was added into the HepG2 cell culture which had been cultured for 12 hours. RNA was extracted for further analysis 48 h or 72 h after the transfection.

Targeting snoRNA target prediction:

C/D box snoRNA of the human were subset from snoDB[4], and the locations of both C-box and D-box were annotated by the snoReport tool[5]. For each Nm site, the upstream and downstream sequences within ±20 nt window were extracted to build Nm motifs. For each Nm motif, the binding affinity to each annotated snoRNA was calculated by the RNAplex tool[6]. Minimum

free energy (MFE) threshold for binding was set to -7.7, and the modified nucleotide on the target RNA was restricted to the 5th position upstream of the D-box.

snoRNA knockdown and targeted libraries preparation

The antisense oligonucleotide (ASO)-mediated snoRNA depletion in HepG2 cells was performed as described by Liang *et al* [7]. HepG2 cells were transfected for 48 hours using self-designed ASOs synthesized from IDT targeting SNORD21 (mU*mC*mC*mC*mG*T*C*T*T*G*A*A*A*C*A*mA*mU*mU*mA*mU) or non-targeting sequence (Qiagen, AllStars Neg. Control, SI03650318). RNA was purified as for Nm-Mut-seq and the fold change in snoRNA expression level in siControl and ASO treated samples was quantified by RT-qPCR. The polyA$^+$ RNA samples from three biologically independent replicates were used to prepare Nm-Mut-seq libraries.

Nm sites validation by site-specific RT-qPCR based on Nm-seq

Total RNA of wild-type HepG2 cells was fragmented using RNA Fragmentation Reagent (Thermo Fischer Scientific) at 95 °C for 5 min and cleaned up by RNA Clean & Concentrator column (Zymo Research). The RNA fragments were further purified by denaturing polyacrylamide gel electrophoresis (12% acrylamide/bisacrylamide, 8.0 M urea). Gel-purified RNA fragments (~50 nt, 10 μg) were dephosphorylated with CIP (NEB). Twelve cycles of standard OED were performed as described in the Nm-seq method. Then a final round of OE cycle was performed without dephosphorylation. Since 2'-*O*-methylation prevents periodate cleavage, the RNA fragments on the 5'-side of the Nm sites could be enriched after multiple OED cycles. The harvested RNA fragments were purified by Oligo Clean & Concentrator (Zymo Research) and ligated to a 3' adaptor. The reverse transcription was then performed using AMV RT (Promega) with a primer hybridized to the 3'-adaptor sequence. The cDNA generated by RT reaction was further quantified by RT-qPCR with site-specific PCR primers. The resulting Ct values were used to calculate fold changes. We first tested this method by validating 19 known Nm sites in 18S rRNA. All the sites displayed the fold change at >1.5, and

16 out of 19 sites are >2, indicating that this method could be used to validate Nm sites in mRNA. Therefore, candidates were classified as Nm sites if the fold change was above 1.5.

Nm site validation using MeTH-seq

115 putative Nm sites were subjected to the MeTH-seq assay, which was performed as described by Bartoli *et al*[8]. The polyA$^+$ RNA from wild-type HepG2 cells was subjected to RT reaction with locus-specific PCR primers. The reverse transcription was carried out using AMV RT (Promega) with 6 mM $MgCl_2$ or 20 mM $MgCl_2$ at 42°C for 1 hour. Then the cDNAs generated from the two RT reactions were quantified by RT-qPCR with site-specific PCR primer pairs. The resulting Ct values were used to calculate fold changes. Sites were defined as Nm sites if the fold change was above 1.5.

Modification-free mRNA fragments by *in vitro* transcription (IVT)

The polyA$^+$ RNA was fragmented using RNA Fragmentation Reagents (Thermo Fischer Scientific) at 95 °C for 3.5 min and cleaned up by RNA Clean & Concentrator column (Zymo Research). Then RNA fragments were end-repaired using PNK and then ligated with a 3'-adaptor. The 5'-adaptor containing T7 promoter sequence was then ligated to the 5'-end of RNA fragments. Next, the first and the second-strand cDNA was synthesized using Superscript IV and Q5 DNA polymerase, respectively. Double-stranded cDNA was purified using the DNA Clean & Concentrator (Zymo Research). The T7 transcription step was carried out using T7 RNA polymerase (NEB) at 37°C for 1 hour. The dsDNA in the reaction mixture was then cleaned up using TURBO DNase (Thermo Fischer Scientific) and the resulting IVT RNA was purified by RNA Clean & Concentrator (Zymo Research). The IVT RNA was then used for Nm-Mut-seq library preparation.

NGS sequencing data analysis

Sequencing was carried out on Illumina NextSeq550 according to the manufacturer's instructions, with a single-end 75 bp read length. The sequencing data was all trimmed with the cutadapt tool to remove adapters and low-quality reads. PCR duplicates were removed with BBMap tool (v.38.73), 5-mer random barcodes at reads ends were trimmed, and low-quality or

short reads (less than 20 nt) were removed using cutadapt tool (v.1.15). Remaining reads were aligned to hg38 genome using Tophat2 (v.2.1.1) and bowtie2 (v.2.4.0) allowing at most three mismatches. The generated .bam files were split into positive and negative strands and sorted using Samtools (v.1.9). Sequence variants were identified by measuring the base composition at each position using bam-readcount software (v.0.8.0). The generated bam-readcount results were parsed and analyzed by in-house scripts. Internal A → T or C → T or G → T mutation ratio at each Nm candidate site was calculated by data output from bam-readcount pipeline and confirmed by direct IGV visualization (v.2.8.0).

To identify high-confidence Nm mutation signatures, we set the Nm detection criteria as follows: (1) At one Nm candidate site, the A/C/G-to-T mutation rate is above 10% (with mutation count above five) in Nm-Mut-seq libraries; (2) At one Nm candidate site, the mutation rate in Nm-Mut-seq libraries is above 3-fold over that in 'Input' libraries; (3) At one Nm candidate site, the total reads coverage depth is above 20 in both Nm-Mut-seq and 'Input' libraries; (4) At one Nm candidate site, the mutation rate in Nm-Mut-seq libraries is above 1.5-fold over the background in any given sequence motif (defined as the mutation rates detected from RNA probes containing 0% Nm, as in Supplementary Fig. 2d); (5) We excluded mutation sites at neighboring nucleotide 3′ or 5′ to known Nm sites.

For mRNA, a consistent cutoff with mutation rate > 10% in "Nm-Mut-seq" libraries was set to call transcriptome-wide Nm candidates. Then the mutation ratios of all Nm candidate sites were converted into the methylation fraction based on the Nm calibration curves. The confident sites with stoichiometry above 10% in wild-type cells were used for downstream analysis.

To confidently classify Nm-modified sites dependent on a specific methyltransferase, the A/C/G-to-T mutation ratios in siControl were converted into Nm stoichiometry (based on the calibration curves). The candidate sites with >10% stoichiometry in siControl were used for downstream analysis. For FBL and FTSJ3, we defined the hyporegulated Nm sites that displayed a notable ≥ 10% reduction in methylation fraction after FBL- or FTSJ3-depletion ($P < 0.05$, for three biologically independent samples, siControl versus siFBL or siFTSJ3).

Particularly, for FBL, we focused on the subset that exhibited a notable ≥ 20% reduction in methylation fraction after FBL depletion, for downstream analysis.

The input of Nm-Mut-seq is equivalent to regular RNA-seq, therefore we quantified the gene-level read counts of input samples that aligned to hg38 for differential gene expression analysis. Cufflinks tool (v2.2.1) was used to make inferential tests where differentially expressed genes were identified at FDR < 0.05 cut-off. Gene ontology analysis was performed using the online analysis software DAVID 2021.

Statistics and reproducibility

For the Nm-Mut-seq libraries, three biologically independent replicates were used in each experiment with cultured cells. Immunoblots are the representative images from at least three rounds of independent experiments. Data are presented as the mean ± standard deviation (s.d.), with two-tailed Student's *t*-tests on the statistical significance of differences between groups. All statistical analysis and data graphing was done in Prism software (version 9.1.1).

No statistical methods were applied to pre-evaluate the sample size. No data were excluded from analysis. Samples in this study were not randomized. Blinding was not used for this study because for each experiment cell culture, sample preparation, reagents, and experimental settings were kept consistent.

SUPPLEMENTARY DATA

Data S1. Fluorescence-based RT selection platform

This system utilizes a synthetic RNA oligo containing a component of the fluorogenic Broccoli aptamer [9], with a modified RNA base of interest at a defined site (Supplementary information, Fig. S1a). An RT variant was then used for reverse transcription (RT) of RNA into cDNA, followed by PCR amplification and the installment of a T7 promoter and the rest of the Broccoli sequence. Finally, *in vitro* transcription by T7 RNA polymerase was conducted to produce RNA in the presence of the fluorophore that fluoresces upon binding to the aptamer. Critically,

the product is fluorescent only when a mutation could be installed at the specific site of the RNA modification during the initial RT reaction.

Data S2. The selection of starting point for RT evolution

To select an evolved RT sensitive to Nm detection, we synthesized the 33-mer Broccoli RNA fragment with an internal Am15 modification. We then tested whether a subset of commonly used RTs which may already induce mutations at Am site using the "RT-PCR-IVT" assay (Supplementary information, Fig. S1a). At two different dNTP concentrations, we tested wild-type HIV RT (the catalytic p66 subunit), M-MLV RT, and AMV RT, as well as our previously evolved RT-1306[3] which was evolved from HIV RT p66 subunit in m^1A-quant-seq. M-MLV RT and AMV RT have been used in Nm detection at low dNTP concentration (e.g., RTL-P method)[10]. Intriguingly, we found that RT-1306 provided the highest mutation signal in this assay, particularly under lower dNTP concentrations in the RT step. These results indicated that the enzyme evolution in endowing RT-1306 with m^1A sensitivity could also enable it to detect Am to some degree (Supplementary information, Fig. S1b). Therefore, we decided to select RT-1306, an evolved variant of the p66 subunit of HIV RT with six mutations (D76A, R78K, W229Y, M230L, V75F and F77A), as the starting point for further optimization and evolution of an Am-sensitive RT.

Data S3. RT mutation library selection

To search for putative amino acid residues that affect the mutagenic activity of the evolved RT on Nm detection, we designed 14 libraries of RT-1306, screening for mutation amino acids that are located near the dNTP-binding regions of the template protein. Each library contained random mutations of two selected amino acid positions, thereby yielding a theoretical library size of 400 variants per library, screened in crude cell lysate. In the first round of selection, we assayed 90 variants from each library for a total of 1,260 variants screened (Supplementary information, Fig. S1d). The libraries based on mutations at K78/E79 and F61/A62 both yielded many hits with a higher fluorescence response to the Am selection compared to RT-1306. Secondary validation of hits revealed that the variants from the F61/A62 library gave the most

robust and reproducible results. Therefore, in the second round of selection, we saturated the two amino acids from the F61/A62 library to find the optimal combination of mutations at these sites. This second round of selection yielded our optimal RT, variant RT-41B4, which contains F61S and A62V substitutions, in addition to the mutagenic background of template RT-1306 (Fig. 1c).

Data S4. The possibility for m^6Am mapping

m^6Am modifications are known to exist in the 5'-cap regions of mammalian cytoplasmic mRNAs. Our RT-41B4 can theoretically recognize m^6Am and read it out as a mutation signature, thus potentially expanding the application of Nm-Mut-seq. However, since our current Nm-Mut-seq protocol does not include a decapping step for 5'-cap removal before RNA 5'-adaptor ligation, m^6Am modifications adjacent to mRNA 5'-cap cannot be detected by Nm-Mut-seq. In the future, researchers can add one decapping step before mRNA end repair, thereby allowing the m^6Am sites to be uncovered through Nm-Mut-seq method.

To determine if the internal Am sites identified through Nm-Mut-seq are in fact m^6Am, we compared our HeLa mRNA data with m6A sites revealed by two m6A sequencing methods of single-base resolution, m^6A-SAC-seq[11] and eTAM-seq[12]. From m^6A-SAC-seq, which specifically maps m^6A sites as mutations, we were unable to find an overlap with our data (Am sites, N = 241). eTAM-seq, which mutates A sites but leaves m^6A unchanged, provides the dataset of m^6A sites with FTO-treated RNA and IVT RNA as background controls. By comparing Am sites with m^6A sites in eTAM-seq, there is also no overlap. Overall, these data suggest that internal Am sites on HeLa mRNA are unlikely to be m^6Am.

Data S5. Limitations of the study

Although Nm-Mut-seq has promoted the quality of Nm mapping, the RT and the method protocol can still be further improved. First, for some sequence contexts, such as -NG(Am)NN-, the RT enzyme is less robust at generating misincorporations at 2'-O-methylated sites, therefore the actual Nm modification within these sequence contexts may be underestimated, especially

when the sequencing depth is moderate. As indicated from our sequencing of synthetic oligo probes containing –NNNmNN-, our method generates background mutations on **N**CC motif, for which we set up cutoff-based analysis pipeline to remove these false positives. Our method also probably underestimates Nm modification in the GNm sequence context. For example, Elliott *et al* [13] reported a Nm site (Am3150) in *Pxdn* mRNA in HeLa and HEK 293T cells. However, we detected an A-to-T mutation rate of only 6% at Am3150 in one of our HeLa replicates, possibly due to the low mutagenic efficiency of RT-41B4 at the -GGAmAG- motif around this site. Second, the average Nm modification fraction detection limit (across all sequence contexts) is around 35%, indicating that some low fraction Nm sites might be missed, resulting in underestimation of the total number of detected Nm sites. Third, this method doesn't distinguish m^6Am from Am, but comparing Nm-Mut-seq data with m^6A mapping data could help extract putative m^6Am sites. Fourth, the stoichiometry estimation based on the 5-bp window centered by the modified sites may not fully reflect the actual modification level of Nm sites. Although we demonstrated that our sequence context-based calibration curves are overall accurate for most rRNA sites, we do not exclude the possibility that mutation rates may not be accurate for a small portion of sequence contexts, especially when the mutation-to-stoichiometry curve tends to be more horizontal at some motif contexts. Fifth, we observed that the accuracy of Nm-Mut-seq could be affected by local RNA structures or neighboring RNA modifications around the Nm-modified sites. Using the known Nm methylated sites on rRNA as an example, when comparing Nm-Mut-seq with other methods like SILNAS and RiboMethSeq, we have observed a disagreement on Nm stoichiometry at 6 Nm sites, including 18S Cm462/Gm509/Gm1447 and 28S Gm1303/Gm2351/Gm4469 (Supplementary information, Fig. S3b). For 18S Gm1447, 18S Cm462, 18S Gm509, and 28S Gm1303, these 4 rRNA Nm sites are located within strong local RNA structures. For 28S Gm2351 and 28S Gm4469, these 2 rRNA Nm sites are installed between other RNA modifications: Am2350 and Cm2352 are located at the neighboring positions of Gm2351; Um4468 and Ψ4470 are located at the neighboring positions of Gm4469. Finally, as previously stated, the current RT-41B4 cannot be used to map Um sites. Future development of a new Um-sensitive RT enzyme is

therefore reasonably justified, and may in turn uncover FTSJ3-mediated Um modifications in human mRNA. Our RT selection and optimization strategy, validated by m^1A-quant-seq and the success in this study, stands to provide a robust method for creating customized RTs for mapping technologies of diverse RNA modifications.

SUPPLEMENTARY REFERENCES

1. Sharma, S., Marchand, V., Motorin, Y. & Lafontaine, D.L.J. Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.* **7**, 11490 (2017).
2. Erales, J. *et al.* Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. U S A* **114**, 12934-12939 (2017).
3. Zhou, H. *et al.* Evolution of a reverse transcriptase to map N(1)-methyladenosine in human messenger RNA. *Nat. Methods* **16**, 1281-1288 (2019).
4. Bouchard-Bourelle, P. *et al.* snoDB: an interactive database of human snoRNA sequences, abundance and interactions. *Nucleic Acids Res.* **48**, D220-D225 (2020).
5. Hertel, J., Hofacker, I.L. & Stadler, P.F. SnoReport: computational identification of snoRNAs with unknown targets. *Bioinformatics* **24**, 158-164 (2008).
6. Tafer, H. & Hofacker, I.L. RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* **24**, 2657-2663 (2008).
7. Liang, X.H., Vickers, T.A., Guo, S. & Crooke, S.T. Efficient and specific knockdown of small non-coding RNAs in mammalian cells and in mice. *Nucleic Acids Res.* **39**, e13 (2011).
8. Bartoli, K.M., Schaening, C., Carlile, T.M. & Gilbert, W.V. Conserved methyltransferase Spb1 targets mRNAs for regulated modification with 2′ -O-methyl ribose. *BioRxiv.* (2018).
9. Filonov, G.S., Moon, J.D., Svensen, N. & Jaffrey, S.R. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136**, 16299-16308 (2014).
10. Dong, Z.W. *et al.* RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res.* **40**, e157 (2012).
11. Hu, L. *et al.* m(6)A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat. Biotechnol* **40**, 1210-1219 (2022).
12. Xiao, Y.L. *et al.* Transcriptome-wide profiling and quantification of N(6)-methyladenosine by enzyme-assisted adenosine deamination. *Nat. Biotechnol* (2023).
13. Elliott, B.A. *et al.* Modification of messenger RNA by 2'-O-methylation regulates gene expression in vivo. *Nat. Commun.* **10**, 3401 (2019).

Protocol for RT-41B4 enzyme overexpression and purification.

RT-41B4 Overexpression in *E.coli BL21*:

1. Expression vector transformation: transform vector (pET30a/RT-41B4) into *E. coli* BL21 competent cells. Plate cell culture onto LB agar plate with Kanamycin (50 μg/ml).
2. Starter culture: on the following day, pick a single colony of the expression strain from the transformation plate into 50 mL LB media with Kanamycin (50 μg/ml), shake at 37 °C overnight.
3. Expansion of start culture: expand the overnight culture by transfer 5mL to 500mL LB with Kanamycin. The cell culture was then grown at 37 °C for ~3 h until the optical density at 600 nm (OD600) reached 0.6– 0.8.
4. Induction: Cool down the cell culture to room temperature by placing the flask in iced water bath after it has reached OD600 0.6- 0.8. Induce expression by adding IPTG to a final concentration of 0.5 mM. Induce the protein expression at 16 °C for 12-16 h.
5. Cell collection: The bacteria were then harvested by centrifugation (4000 rpm) for 20 min. Cell pellets were transferred to a 50 mL centrifuge tube and weighed. In general, 1L cell culture produces 5-10 g cell. The cell can be used to purify protein immediately or be stored at -80°C for 1 month.

RT-41B4 Purification:

6. Resuspend the cell: Obtain the cell pellet from -80°C and thaw at room temp until half melted. Prepare cell **Lysis Buffer** containing 100mM Tris-HCl, 150 mM NaCl, pH 8.0, and dissolve half-tablet of the proteinase inhibitor cocktail (Pierce). Next, add 0.5 mg/mL (final conc.) lysozyme freshly and mix until it is fully dissolved. All the buffer used are making by RNase/DNase free water.
7. Lysis the cell: The cell pellet was resuspended in lysis buffer (5-10 mL buffer for 1g *E.coli* cells). The mixture was stirred slowly at room temp for 10 min to lyse the cell and then on ice for 20 min. The cells were then lysed by sonication and centrifuged at 12,000 r.p.m. for 40 min at 4 °C.
8. Nucleic acid removal: Pour the supernatant from the previous step into a chilled glass beaker on ice. To eliminate the DNA and RNA in the supernatant, add 1% streptomycin sulfate (w/v): for 50 mL supernatant, add 0.5 g streptomycin sulfate. Stir the mixture slowly on ice for several minutes until it becomes cloudy, and then centrifuged at 12,000 r.p.m. for 15 min at 4 °C.
9. Prepare the Ni-NTA resin: Fill the column with Ni-NTA resin to prepare a bed volume of 2 mL. Equilibrate the column with 20-30 mL of equilibration buffer.
10. Apply the sample to the column by gravity flow. Immediately following the sample having passed through the resin, wash the column with 30 mL **Washing Buffer** (100 mM Tris-HCl, 150 mM NaCl, pH 8.0)
11. To wash out nonspecifically bound proteins: wash the column using 15 mL washing buffer containing 50 mM, 75 mM imidazole one after another.

12. To elute RT-41B4, wash the column with 15 mL washing buffer containing 100 mM, 250 mM imidazole one after another, collecting the elution fractions in 1.5 mL eppendorf tubes.
13. Run SDS-PAGE gel to check the purity of protein in each collection tube and combine the elution with pure protein.
14. To get rid of imidazole, the eluted protein fractions were concentrated and run through a desalting column (PD-10, GE Healthcare).
15. (Optional) The protein could be further purified by Q ion-exchange chromatography: The fractions were subjected to Mono Q ion-exchange chromatography, where the protein was injected onto the column flushing with 97.5% Buffer A and 2.5% Buffer B (50 mM Bis-tris pH 7.0, 1 M NaCl) and the protein was recovered in the flow-through portion. The ion-exchange purification was found to be essential for effectively removing nuclease contamination. All purification steps were carried out at 4 °C or on ice. Fractions containing the expressed protein were combined and concentrated to 2.5 mL with a 30-kDa cut-off centrifugal filter (Millipore), run through the desalting column again, and eluted with the storage buffer (50 mM Tris-HCl, 25 mM NaCl, 1 mM EDTA, 10% glycerol, pH 7.0).
16. The protein concentration was determined by both BCA protein assay (Thermo Fisher) and Bradford reagent (Sigma).
17. Purified proteins were concentrated to 100 μM using a 30-kDa cut-off centrifugal filter (Millipore), aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C. To make a ready-to-use protein solution (at 10 μM), the concentrated protein was mixed with a buffer containing 50 mM Tris-HCl, 25 mM NaCl, 1 mM EDTA, 60% glycerol, pH 7.0. The protein solution can be kept at −20 °C for six months.

Quality control of purified RT enzyme:

18. The RT read-through assays were performed to test the quality of RT enzyme.

The reactions were performed in 10-μl reaction volumes containing:

1× RT buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl and 2 mM MgCl2),
4 pmol RNA substrates (example:
rGrGrArGrArCrGrGrUrCrGrGrGmArCrCrArGrArUrArUrUrCrGrUrArUrCrUrGrUrC)
2 pmol DNA primer with 5′-FAM label
(example:/56FAM/CGGAGCCCACACTCTACTCGACAGATACGAATAT),
1 mM of each dNTP
2 μM purified RT.

The RNA substrate and DNA primer were added first and heated to 70 °C for 2 min and then annealed on ice for another 5 min. The RT reactions were then carried out at 37 °C for 1 h, followed by heating at 80 °C for 10 min to denature the RT. 4 μl of the reaction was mixed with RNA loading dye and

heated to 95 °C for 5 min. 7 µl of that mixture was then immediately loaded onto 10 % denaturing PAGE gel and run at 200 V for 50 min. The gel was then imaged on a Bio-Rad imager. The full-length cDNA product can be clearly observed on the gel without any degradation, indicating that the RT enzyme is active and there is no DNase or RNase contamination.

Detailed step-by-step Nm-Mut-seq protocol:

Culture and harvest HeLa and HepG2 cells:
(Growth media: HeLa: DMEM + 10% FBS + Pen/step antibiotics; HepG2: DMEM (low glucose) + 10% FBS + Pen/step antibiotics.)

1) Seeding: Begin with a 10 cm plate containing ~90% confluent cells. Treat with 1 mL trypsin for 5 min and transfer ~30 % cells to a new 10-cm plate with 9 mL media. Let cells grow for 2 days.
2) Harvesting: When the cells have reached ~90% confluency, remove the culture media and wash the cells with 5 mL DPBS. Treat with 1 mL trypsin for 5 min and collect cells in a 1.5 mL Eppendorf tube. Extract total RNA with TRIzol reagent or RNeasy Mini Kit (Qiagen).

Day 1:

1. Total RNA DNase I treatment

80-100 µg Total RNA extracted from the cells are treated by DNase I (Thermo Fisher) to remove DNA:

85 µL RNA in RNase-free water
10 µL DNaseI buffer with $MgCl_2$
5 µL DNase I

Mix well and incubate at 37°C for 45 min. Add 10 µL 50mM EDTA and heat at 65°C for 2 min to inactivate the enzyme. Immediately follow with MEGAclear™ kit (Thermo Fisher) to clean up.

2. Small RNA removal

Use 1 column per sample and follow the manufacturer's instructions to remove small RNA. Adjust the total elution volume to 100 µL with RNase-free water.

3. Poly A^+ RNA enrichment using DynaBeads™ mRNA DIRECT™ purification Kit (ThermoFisher)

Process 50 µg of the RNA sample (in a volume of 100 µL) from the previous step in 2 rounds of mRNA purification.

1) Prepare beads:
   a) Use 130 µL × 2 = 260 µL beads per sample.
   b) Wash the beads with binding buffer, and place on magnet until the solution is clear.

c) Remove the supernatat and add binding buffer to adjust the total volume to 200 μL (for each sample, concentrate the original 130 μL beads to 100 μL).

2) Add 100 μL beads to 100 μL RNA sample, rotate on a roller for 20 min at room temperature.
3) Place the tubes on magnet for ~ 2min, remove the supernatant.
4) Wash the beads twice with washing buffer B.
5) After removing the washing buffer, add 100 μL 10 mM Tris-HCl pH 7.5 and heat at 75°C for 2 min.
6) Place the tube immediately on magnet for 2 min and transfer the supernatant into a new tube.
7) Perform a second round of mRNA enrichment, using new beads and repeat steps 2-6.
8) For the final elution, use 50 μL Tris buffer (pH = 7.5) to elute the RNA.
9) Concentrate polyA$^+$ RNA by RNA Clean & Concentrator Kits (Zymo)

The yield of polyA$^+$ RNA should be ~1-2% of the total RNA.

Day 2:

1. RNA fragmentation

1) Preheat the thermo block at 95°C.
2) Dilute 1 μg polyA$^+$ RNA sample from the last step into a 45 μL final volume with RNA water and mix with 5 μL fragmentation buffer (1M Na_2CO_3-$NaHCO_3$, PH = 9.2).
3) Heat the sample at 95°C for 8 min then put on ice immediately for 1 min.
4) Perform Oligo Clean & Concentrator (Zymo) purification.
5) Elute RNA with 40 μL RNA water.

2. PNK treatment

1) Prepare the PNK reaction mixture:
   40 μL RNA sample
   5 μL PNK buffer
   1 μL enzyme
   Mix together, incubate at 37°C for 30 min.
2) Add 5 μL ATP (10 mM stock) and 1 μL PNK enzyme, 37°C incubate for another 30 min.
3) Heat the sample at 65°C for 10 min to inactivate the enzyme.
4) Perform Oligo Clean & Concentrator (Zymo) purification.
5) Elute RNA with a suitable volume of RNase-free water (the final concentration of RNA should be over 100 ng / 6 μL).

3. Adaptor ligation

Use NEBNext Multiplex Small RNA library prep set (for illumine) to ligate the sequencing adaptors.

3'-NNadaptor: 5'rApp-NN NNN <u>TTA GGC</u> AG ATC GGA AGA GCA CAC GTC T-3'biotin (IDT, RNase-free HPLC purification, 12 μM)

5'-NNadaptor: 5'-GU UCA GAG UUC UAC AGU CCG ACG AUC NNN NN-3' (IDT, RNase-free HPLC purification, 12 μM)

1) 3' ligation
   a) Add 3 μL RNA and 0.5 μL 3' adaptor in a PCR tube
   b) Incubate at 70°C for 2 min and put on ice immediately.
   c) Add 5 μL 3' ligation buffer and 1.5 μL ligation enzyme, mix well.
   d) Incubate the ligation mixture at 25°C for 2 hr.
2) Hybridize the reverse primer.
   a) Dilute 0.5 μL SR RT primer to 2.75 μL, then add to the ligation mixture.
   b) Incubate the sample at 75°C for 5 min, then 37°C for 15 min, followed by 25°C for 15 min.
3) 5' ligation
   a) Aliquot the 5' adaptor into a separate PCR tube and preheat at 70°C for 2 min and put on ice immediately.
   b) Add 1 μL adaptor, 0.5 μL 5' ligation buffer and 1.25 μL ligation enzyme to the ligation mixture from last step and incubate at 25°C for 2 h or overnight.
   c) Perform Oligo Clean & Concentrator (Zymo) purification.
   d) Elute RNA with 21 μL of RNA water.

Day 3:

1. Reverse transcription

Split the resulting RNA into equal halves, preheat to 70°C for 2 min and put on ice immediately.

RT mixture:

RT reaction mixture	volume
RNA sample	10 ul
5× RT buffer	4 ul
dNTP	4 ul
RNase inhibitor	0.5 ul
Evolved RT enzyme (10 uM stock)	2 ul

Restrictive and Permissive conditions differ in dNTP concentration:

Permissive condition: 2 μL 10 mM dNTP + 2 μL water (no mutation at Nm sites);

Restrictive condition: 2 μL 20 μM dNTP + 2 μL 10 mM dATP (mutation signature at Nm site).

Incubate at 37°C for 1 h, and then inactivate the enzyme at 80°C for 10 min.

Note: For the 20 μM dNTP stock preparation, the 100 mM dNTP set solution was purchased from Thermo Scientific (R0181). 10 μL of each dNTP was taken from each solution tube, then combined

and 60 µL of RNase free water was added to create a 10 mM dNTP mixture solution. Subsequently, the 10mM dNTP mixture solution was diluted with RNase free water to a final conc of 20 µM.

2. qPCR inspection on library quality.

1 µL cDNA from last step.

10 µL 2× FastStart SYBR Master Mix (Roche)

1 µL SR primer

1 µL indexed primer

7 µL water.

Run the following protocol:

Pre-incubation	95°C; 600 s
~30 cycles	95°C; 20 s
	60°C; 20 s
	72°C; 20 s
Melting	95°C; 10 s
	65°C; 60 s
	97°C; 1 s

3. PCR amplification.

Use 10 µL cDNA to perform PCR with 0.625 µL indexed primer. Add 14.38 µL of PCR mix to each sample and run following protocol (15 cycles for “Permissive condition” and 21 cycles for “Restrictive condition”, based on the qPCR results).

Pre-heating	94°C; 30 s
12~15 cycles	94°C; 15 s
	62°C; 30 s
	70°C; 15 s
Extension	70°C; 5 min
Incubation	4°C

PCR mix:

	Volume (uL)

2X LongAmp Taq	12.5 uL
SR primer	0.625 uL
Water	1.25 uL
Total	14.38 uL

4. Gel Electrophoresis.

1) Prepare a ~2.5% low melting agarose gel.
2) Add 5 µL 6× loading buffer (NEB, #B7024) to the samples.
3) Load the samples together with the ladder (Quick Load pBR322 DNA-MspI Digest NEB #N3032) and run the gel at 90V for 90 min.
4) Cut bands between 160 and 200 bp and perform DNA extraction using the MinElute Gel Extraction Kit (QIAGEN, #28606).
5) Elute with 13 µL water.
6) Use 1 µL to measure the concentration by Qubit.

   The library concentration should between 1 – 5 ng/µL.