Identification and exploration of 2'-O-methylation sites in rRNA and mRNA with a novel RNase based platform

Yao Tang[1], Yifan Wu[1], Ruilin Xu[1], Xiangwen Gu[1], Yan Wu[1], Jian-Qun Chen[1*], Qiang Wang[1*], Qihan Chen[1, 2*]

1 The State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, Jiangsu 210023, China.

2 Nanjing University Medical School, Nanjing, Jiangsu 210093, China.

*Corresponding author

To whom correspondence should be addressed. E-mail: chenqihan@nju.edu.cn, wangqiang1997@gmail.com, chenjq@nju.edu.cn.

Abstract

Epigenetic modifications of RNA play a critical role in RNA structure and gene regulation. Ribose 2'-O-methylation (Nm) is one of the most common modifications among them, featured with one methyl group added to the 2' hydroxyl of the ribose moiety of RNA nucleosides. Although 2'-O-methylation was widely present in different type active RNAs, such as mRNAs, tRNAs, rRNAs, and even miRNAs, its biological functions and significance are still largely unknown due to the lack of accuracy and efficient identification screening tools. In this study, we first time established an enzyme-based high throughput Nm sites detection method with base resolution, which was named NJU-seq (Nm site Judge Universally Sequencing). We effectively identified Nm sites in human cultured cell lines, plants and animals with this new sequencing screening technique. Through this approach, we observed high abundance and complexity of Nm sties in cellular RNAs cross different species samples. Our study suggested that the function and regulation of Nm sites are more complicated than previous studies revealed and worth further mechanistic investigation.

Introduction

More than 100 types of RNA modification have been identified, which play important roles in regulating diverse biological processes [1]. Among these RNA modifications, 2'-O-methylation (also known as Nm), where a methyl group is added to the 2' hydroxyl of the ribose moiety of a nucleoside, could decorate RNA without limitation of the base.

Indeed, Nm is one of the most common and widespread types of RNA modification commonly found in ribosomal RNA (rRNA), transfer RNA (tRNA), small nucleolar RNA (snRNA) and, as reported recently, in messenger RNA (mRNA) [1-5]. The recent study from Dr. He Chuan has indicated that Nm is abundant among mRNAs of HeLa and HEK293t cell lines [6, 7]. Besides, several sporadic published studies demonstrated the importance of Nm in multiple biological processes and function as an important signaling regulator, involved in RNA structure stability [8, 9], immune recognition [5], mRNA alternative splicing [3, 10, 11] and protein synthesis [6]. However, due to the lack of screening tools to efficiently detect Nm sites on a large scale, the exploration of biological functions stagnates and is still largely unclear.

In the past few years, three high throughput Nm detection methods had been developed, which were 2'OMe-seq [12, 13], RiboMethSeq [14-16] and Nm-seq [7]. The principle of 2'OMe-seq was based on the block of RNA reverse transcription at a low concentration of dNTPs. Nevertheless, this method showed quite high false-positive outcomes sites because of recognizing RNA modifications with low specificity [13]. RiboMethSeq showed some advantage since the Nm site is specific resistant to RNA alkaline lysis. However, this process was inferred by the 3'ends of RNA fragments. The matrix effect of this approach is significantly influenced by the sample size and data coverage. Indeed, this was a barrier to the effective application of RiboMethSeq in the study of mRNA or other complicated modification circumstances [13, 15-17]. Nm-seq was designed as a multiple cyclic operation of Sodium Metaperiodate oxidation and sample purification. This method made a big step forward that showed a high sensitivity in mRNA Nm site detection. However, the large sample size demand, the complex operation steps and the low consistency for investigation of mRNA modifications diminished the possibility to establish a high throughput approach based on Nm-seq process [7]. Therefore, the development of a new effective technique for comprehensive and accurate profiling Nm sites is the emerging need for further research of the Nm mechanism.

Here, we established a breakthrough approach NJU-seq as a high throughput tool to efficiently identify Nm sites on rRNAs as well as mRNAs. The method relies on an RNase from *Mycoplasma genitalium* (MgR), which has both helicase and RNA hydrolase activities, involved in digestion processes of ssRNA from 3 'to 5'direction [18]. Since MgR is highly sensitive to Nm modification, the digested products are produced at every Nm+1. Sequencing these products by NGS, we efficiently demonstrated the Nm sites on rRNAs

in human cultured HeLa cells, which showed promising consistent results with the previous findings[19]. Also, we mapped the Nm sites on mRNAs in multiple human cultured cell lines, plant samples and animal tissues (most data was not shown in this article). Furthermore, in this study, we profiled mRNA Nm sites in stem tissues of *Arabidopsis*, and the first time showed NJU-seq as an efficient screening technique to characterize Nm features on RNAs.

Result

1. Principal and Application of NJU-seq

MgR, a member of the RNase R family from *Mycoplasma genitalium*, is highly sensitive to Nm sites, exhibiting both helicase and RNA hydrolase activities. Taking advantage of this unique feature, we combined MgR Nm site-specific hydrolysis activity with NGS to develop the new high throughput Nm screening method.

We first expressed and purified MgR protein in vitro and tested the enzyme activity as described in Methods (Fig. 1a). We furtherly tested whether such Nm sensitivity applies to all four kinds of ribonucleotides from the different sequence contexts by performing reaction of MgR with the synthesized 22nt ssRNA substrate carrying two mixed 2'-O-methylation modification sites NNNmNN (N=A/G/C/U) (Fig. 1b). From the PAGE gel, the 18nt and 14nt ssRNA products were accumulated with the increase of MgR concentration, which confirmed the versatility of the enzyme activity (Fig. 1b). In addition, we also performed a reaction of MgR with the synthesized 40nt ssRNA substrate specifically carrying only Am site at position 28 to demonstrate the exact breakpoint of MgR activity. The reaction products were TA cloned and sequenced. Importantly, we discovered most of the ssRNA products ended at the Am+1 position (19/20) (Fig. 1c), suggesting conjecture of Nm sites by accumulated 3 ’end of MgR reaction products. Moreover, a range of MgR usage, pH, condition, and ion concentration was tested to optimize the appropriate reaction activity (Supplementary Fig. 1).

We established this novel approach as NJU-seq with the following main steps: 1. RNA fragmentation. 2. An adaptor was ligated to 3 ’end of RNA fragments. 3. MgR digestion. 4. 5’end phosphorylation. 5. NGS library construction. 6. Nm site profiling and identification (Fig.1d). A sample preparing with the same process but no MgR treatment was highly recommended as a NC (negative group) to control the false-positive sites caused by strong RNA secondary structure. We performed a mapping analysis of 3’ ends of each fragment from NGS to reference sequence (Fig. 2a, 2b). Here the situations of 18s rRNA from position 157 to 169 were shown as an example, we found the fragments ended

at position 160 and 167 significantly accumulated in cultured HeLa cell samples after MgR digestion comparing with NC results, indicating the two possible Am sites at position 159 and 166 (Fig. 2c, 2d). To further optimize our data analysis approach to control possible interference between Nm modifications, we developed two algorithms for scoring Nm modification sites in differential RNA samples to preserve the high accuracy and efficiency of NJU-seq assay. To analyze the complex circumstance with high Nm density and strong RNA secondary structure (such as rRNA), we investigated both 5 ’and 3’ends information to score Nm sites. To test the separated Nm sites with no interference (such as mRNA), we only collected 3 ’ends information for scoring. The details of calculation and quantification steps were described in Methods.

2. Nm sites on rRNA

2.1. Detection of rRNA Nm sites in HeLa cells by NJU-seq.

We first screened the Nm sites on the rRNA of HeLa cell line with three independent sample repeats. Human ribosomal DNA complete repeating unit (Genbank accession: U13369) was used as a reference for data integration. From the score of points, the results were consistent among the three sample repeats. Since the Nm sites in 18s and 28s rRNA of HeLa cell line were well identified in several previous studies, we compared our results with reported Nm sites to estimate the coverage and accuracy of the results generated by NJU-seq platform. Here we listed sites with top 40 scores in 18s rRNA and top 50 scores in 28s rRNA (Fig. 2e). In 18s rRNA samples, 20/30/35 sites were reported to be Nm sited in top score 20/30/40 sites as we identified. In 28s rRNA samples, 20/30/39/43 sites were reported to be Nm sites in top score 20/30/40/50 sites as we demonstrated. To verify the specificity of MgR for Nm modification, we performed statistics on the other eight kinds of common epigenetic modifications reported in the previous studies on rRNA[19], and Nm showed an obvious reads 3’ end accumulation at Nm+1 site (Supplementary Fig. 2). These results, with such high consistency, strongly indicated the NJU-seq is a promising new alternative tool for Nm detection of rRNAs.

2.2. Detection of rRNA Nm sites in the stem tissue of *Arabidopsis* by NJU-seq.

Based on the results of HeLa cell line, we expanded NJU-seq to other cell lines, different tissues from animals and plants that commonly used in the lab (data not shown). Here, we presented the analysis of the stem tissue of *Arabidopsis* as the proof-of-concept case study example. *Arabidopsis* ribosomal DNA complete repeating unit (Genbank accession: X16077 and X52320) was used as a reference for data integration. We applied a

similar data processing approach as we describe in the HeLa cell line sample study. In 18s rRNA samples, 16/22/24 sites were reported or predicted to be Nm sited in top score 20/30/40 sites as we found. In 25S rRNA samples, 12/22/29/35 sites were reported or predicted to be Nm sites in top score 20/30/40/50 sites as we confirmed (Fig. 2f). Because less previous study reported Nm site profiles in *Arabidopsis*, validation of our discovery in stem tissue of *Arabidopsis* would request more results from future *Arabidopsis* study.

2.3. Investigation of cross-species conserved rRNA Nm sites

Since rRNAs are evolutionary conserved RNAs, we hypothesized that the existence of cross-species conserved 2'-O-methylation sites on rRNAs. Demonstration of cross-species conserved Nm sites would not only further validate the accuracy of NJU-seq platform but also bring novel insights into mechanistic RNA 2'-O-methylation study. Indeed, we further performed rRNA Nm sites comparison analysis between HeLa cell line and the stem tissue of *Arabidopsis*. We first analyzed and labeled the conserved regions of 18s rRNA and 25s/28s rRNA, and calculated the distribution of all 180 Nm sites we identified. Surprisingly, most of Nm sites were actually located in the conserved regions (Fig. 2g, 2h, 2i). There were 14 and 16 pairs of Nm sites highly conserved and both reported in the previous studies in 18s rRNA and 25s/28s rRNA, which were linked by green lines. Intriguingly, we also found 6 pairs of Nm sites highly conserved but only reported in human samples previously, which were linked by yellow lines (Fig. 2g, 2h). Additionally, a certain number of unpaired Nm sites were also discovered, which might suggest the sample-specific rRNA 2'-O-methylation features.

3. Nm sites on mRNA

3.1. Detection of mRNA Nm sites in stem tissue of *Arabidopsis* by NJU-seq

Since the above promising results from rRNA analysis indicated that NJU-seq was an efficient tool for profiling Nm sites, our study took one step further extending to the Nm site profile in mRNAs. As mRNAs have more diverse cellular functions, the 2'-O-methylation might be more excessively regulated. Firstly, we confirmed mRNAs from cell lines, different tissues from animals and plants that commonly used in the lab applied to NJU-seq platform (data not shown). Here, we still illustrated the data of stem from *Arabidopsis* as the proof-of-concept case study example. To simply evaluate the quality of Nm sites data we extracted, we analyzed the consistency of top score 50/500/1000 sites from three independent samples, which turned out to be 86% to 63% (Fig. 3a, 3b, 3c). The overlap ratio tended to be higher among the sites with a higher score, which resembled

rRNA. With this Nm sites map, we tried to uncover the unique feature of Nm sites on mRNAs.

By analyzing the distribution and pattern of approximate top 1000 sites shared in all three samples (1010 sites due to the same scores), Am/Gm/Cm/Um occupied 122/248/316/324 sites (Fig. 3d, Supplementary Fig. 3). Comparing with the total A/G/C/U ratio besides modification in total mRNA, we addressed that this distribution was not random (Supplementary Fig. 4). When we calculated Nm-1 and Nm+1 ribonucleotides with Nm together as a Tri-ribonucleotide unit, the result showed an obvious uneven distribution (Fig. 3d). Notably, GCmU, GUmG, CUmG revealed remarkable higher ratios than other combinations. This trend was quite different from the natural distribution of *Arabidopsis* mRNA Tri-ribonucleotide unit, and was certainly not random (Supplementary Fig. 4). To further review the Nm distribution of functional locations on mRNA, we investigated 401 Nm sites on mRNA from *Arabidopsis* stem with full reference mRNA information (5' UTR/CDS/3'UTR) calculated by relative position. Nm sites density in CDS was higher than UTR regions, which may suggest the gene expression regulation function of 2'-O-methylation in mRNA. In addition, while Nm sites in 5' UTR or CDS had no obvious trend, the sites in 3' UTR showed to be accumulated in the 3' end, which implied its potential role in mRNA stability and maintenance (Fig. 3e).

To further address the correlation between Nm sites and gene functions, we listed the information of the top 300 Nm sites detected by NJU-Seq, including the Nm position and location, gene name and function annotation (Supplementary Table 1). Interestingly, we found that Nm seemed involved in metabolism widespread. Using Metascape to classify the selected genes by biological functions, we further identified most of these genes were gathered in four branches: DNA repair, water stress, signal transduction and defense response (Fig. 3f). This data strongly indicated the importance of Nm sites in the regulation of mRNA responding to environmental stress. On the other hand, the enrichment by molecular function demonstrated a significant accumulation in the domain of ATP binding and metal ion binding (Fig. 3f). This data suggested Nm sites may regulate protein confirmation by promoting the molecular-molecular interactions indirectly.

3.2. Nm sites and alternative splicing.

Recently, several studies showed evidence of the 2'-O-methylation on Adenine, the branch point in alternative splicing exon, which played an important role in regulating the alternative splicing process of pre-mRNA [11]. To further unveil the specific character of this genetic feature, we classified the 1010 Nm sites from the stem of Arabidopsis to two

groups by the location: CSE (Constitute Splicing Exon) and ASE (Alternative Splicing Exon). To our surprise, the Nm sites in CSE were much more than those in ASE. Furthermore, Am occupied only approximately 11% in all Nm sites in ASE and CSE (Fig. 4a). However, the enrichment of Nm sites in two regions exhibited a similar trend as well (Fig. 4b). This finding suggested that the branch point regulator might be just partially involved in 2'-O-methylation modification on mRNA regulation. Despite the ratio of four Nm ribonucleotides, the distribution of Tri-ribonucleotides unit also showed different patterns in CSE and ASE. Especially in Cm and Tm, CSE and ASE had almost opposite bias in Tri-ribonucleotides unit, which implied quite different modification mechanisms restricted in the differential mRNA function regions (Fig. 4c, 4d).

3.3. Nm sites and snoRNA.

Finally, because of previous studies demonstrating most rRNA Nm sites methylated by snoRNA and methyltransferases complex [20], we tested the possible influence of snoRNA on the Nm site profile. We screened the top 1010 Nm sites we identified from *Arabidopsis* stem with snoRNA database and investigated the potential impact. However, there were only 24 of them found to be potentially modified by known snoRNA (table 1). This result highly suggested the majority of Nm sites were mediated by some other machinery other than snoRNA. Also, our primary Nm site profiling results could serve to pursue further mechanistic research objectives.

Discussion

In this study, we established NJU-seq approach as a base-resolution Nm site detection method with the advantages of high throughput, easy to use, and accurate, which made it possible to efficiently profile Nm sites and lead to uncovering the role and mechanism of 2'-O-methylation. This tool was proved to screen most Nm sites not only in rRNA but also in mRNA and other RNAs. Comparing with the previous methods, NJU-seq had no extra sample waste caused by chemical reactions or extra RNA purifications, which made the method superior in coverage, consistency, and cost. Data quality of NJU-seq was effected by sample quality, NGS bias, library length distribution, and NGS coverage, which were critical to the integration of results. Importantly, the algorithms were needed to be trained and optimized for different analytic purposes based on objectives.

The entire function of 2'-O-methylation remained unclear, which was limited to a research tool. The previous study suggested that Nm seemed to involved in immune response via mRNA 5' cap Nm sites, mRNA alternative splicing mediated by intron branch point, maintaining the secondary structure of rRNA, and regulating protein translation by interfering tRNA-mRNA or mRNA-rRNA interactions [5, 6, 11, 14, 21]. Based on our preliminary data shown in this article, the roles of Nm sites in cells were much more prevalent than we expected. It seems 2'-O-methylation commonly involved in most biological processes in the cell. Especially, we first time discovered Nm seemed to play an important role in stress response.

The Nm sites may be methylated in two ways, one is a group of methyltransferases recognizing targeted sites with the specific feature of secondary structures such as 5' cap or the 3' end of small RNAs, the other is a group of methyltransferases recognizing targeted sites by the guidance of an RNA partially complementary to the target sequence around such as snoRNA and other unknown RNAs [5, 17, 22, 23]. Because we found most of Nm sites were not correlated with snoRNA regulation, our study strongly implies the involvement of other major Nm sites methylation unknown mechanisms. With more Nm sites maps generated by NJU-seq, studies focused on 2'-O-methylation would be accelerated to uncover the mysteries of "writer", "reader", and "eraser".

Additionally, we first time compared RNAs crossing the kingdoms from the aspect of 2'-O-methylation. While highly conserved rRNAs from Homo sapiens and plants shared most ribonucleotides in conserved regions, the top 50 Nm sites we detected were all located in those conserved regions. However, only about 40% of Nm sites were one-to-one correspondence, which implies the difference larger than the sequence. More data about other RNAs, or from other species were important and helpful to further investigate the evolution aspect. With the limited time and stipend, we only drew maps of Nm sites at a certain time point of each sample of different tissues. A study sample from multiple time points during a development range would be interesting to observe the dynamic change of Nm sites. Despite the biological function, Nm sites also had great potential to be makers or footprints for clinical diagnosis or other biomedical applications.

Materials and Methods

1. Cloning, expression, and purification

RNase R from *Mycoplasma genitalium* (accession number: WP_009885662.1) was synthesized by GenScript Biotech Co. and inserted into pET28a vector (+) with digestion and ligation at NdeI and BamHI sites, followed by confirmation of sanger sequencing. The

plasmid was transferred into the *E.Coli* BL21 strain (DE3). Cells were grown in 1L LB media under 37°C. 0.3mM IPTG was used to every 1L cells to induce the expression of protein at 16°C for 18h in the shaker incubator. Cell pellets were harvested and lysed by French press with lysis buffer (50 mM Tris-HCl at pH 8.0, 250 mM NaCl, 0.5 mM TCEP), followed by centrifuging at 10000 rpm for 1h. MgR protein was purified by passing through a column Nickle resin (GenScript Biotech Co.), and washed by using 3 column volumes of wash buffer (50 mM Tris-HCl at pH 8.0, 250 mM NaCl, 20mM imidazole, 0.5 mM TCEP), then eluted with elution buffer (50 mM Tris-HCl at pH 8.0, 250 mM NaCl, 100mM imidazole, 0.5 mM TCEP). The elution was concentrated and aliquoted into multiple small tubes, and stored at -80°C for further experiments. Protein was confirmed by SDS-PAGE gel, and its concentration was determined by standard BSA protein on the same gel.

2. In vitro assay of RNA hydrolysis

Nm modified ssRNA substrate was synthesized by GenScript Biotech Co. (5′-Fam-CAAAACAAAACAAAmNCAAAACAAAA-3′). Briefly, various amounts of purified proteins were reacted with 200nM ssRNA substrate in reaction buffer (20 mM Tris-HCl pH 8.0, 100mM KCl, 0.01 mM $ZnCl_2$) at 37°C for 30mins. The reaction was terminated by heating to 85°C for 10mins. The products were separated on 20% urea-PAGE, visualized with Tanon 3500, and quantified with Image-J software. To explore the best reaction condition of MgR, series reactions were designed to test the range of pH, reaction duration, and ion based on the purposes.

Another Nm modified ssRNA substrate was synthesized by GenScript Biotech Co. (AGCCGCCUGGAUACCGCAGCUAGAAUGAmGCUGAGAUGAAA) to test the consistency of hydrolysis. After reacted with MgR, reaction products were purified by RNA Clean & Concentrator-5 kit (Zymo Research), followed by 5'phospheration, re-purification, both end adaptors ligation. The final products were TA cloned and analyzed by sanger sequencing.

3. Cell culture, plants, and RNA extraction

HeLa cells (ATCC) were grown in DMEMs media with 10% FBS (Genial) and 5% antibiotic at 37°C in an incubator containing 5% CO_2. The Columbia eco-type *Arabidopsis thaliana* seeds were planted in a mixture of vermiculite: nutrient soil of 1: 3 sterilized for 30 minutes at high temperature, covered with a thin film to maintain high humidity for about 1 week. The culture was performed in a light incubator at 22°C day and 20°C night with a photoperiod of 14h light/ 10h dark and a parameter of 150 μmol/(m2·s) until *Arabidopsis thaliana* blooms. Total RNA from cells and tissues were extracted by the Qiazol

RNA extraction kit (QIAGEN). Extracted total RNA was confirmed of quantity by Nanodrop (ThermoFisher) and quality by 1% agarose gel.

4. NJU-seq

5ug total RNA was fragmented by RNA fragmentation reagents kit (ThermoFisher) at 95°C for 10 mins and then purified by 10ul 3M NaOAc (pH 5.2), 24ul 5ug/ul glycogen and 372ul 96% cold ethanol. After resoluble in water, every 1ug RNA solution was reacted with 30ug MgR in 50ul final volume with 20 mM Tris-Cl (pH 8.5), 100 mM KCl, 0.01 mM $ZnCl_2$ at 37°C for 30mins. The products were purified by RNA Clean & Concentration-5 kit (Zymo Research), followed by 3'end 5'phosphoration (39ul RNA, 5ul (10x T4 PNK Reaction Buffer), 5ul ATP (10 mM), 1ul T4 PNK (3' phosphatase minus). The NGS library was constructed following the NEBNext Small RNA Library Prep Set for Illumina (New England BioLabs) protocol with the products from the last step. Fragments smaller than 50nt were extracted from 6% PAGE gel and sent for NGS using Illumina X-Ten. The sample of NC (negative control group) was prepared by the same process without MgR protein.

5. NGS data analysis and statistics

Reference

1. Motorin, Y. and M. Helm, *RNA nucleotide methylation.* Wiley Interdiscip Rev RNA, 2011. **2**(5): p. 611-31.
2. Jockel, S., et al., *The 2'-O-methylation status of a single guanosine controls transfer RNA-mediated Toll-like receptor 7 activation or inhibition.* J Exp Med, 2012. **209**(2): p. 235-41.
3. Deryusheva, S., et al., *Post-transcriptional modification of spliceosomal RNAs is normal in SMN-deficient cells.* RNA, 2012. **18**(1): p. 31-6.
4. Tycowski, K.T., et al., *A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14480-5.
5. Zust, R., et al., *Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5.* Nat Immunol, 2011. **12**(2): p. 137-43.
6. Elliott, B.A., et al., *Modification of messenger RNA by 2'-O-methylation regulates gene expression in vivo.* Nat Commun, 2019. **10**(1): p. 3401.
7. *Corrigendum: Nm-seq maps 2'-O-methylation sites in human mRNA with base precision.* Nat Methods, 2018. **15**(3): p. 226-227.
8. Kumar, S., K. Mapa, and S. Maiti, *Understanding the effect of locked nucleic acid and 2'-O-methyl modification on the hybridization thermodynamics of a miRNA-mRNA pair in the presence and absence of AfPiwi protein.* Biochemistry, 2014. **53**(10): p. 1607-15.

9. Yildirim, I., et al., *Interplay of LNA and 2'-O-methyl RNA in the structure and thermodynamics of RNA hybrid systems: a molecular dynamics study using the revised AMBER force field and comparison with experimental results.* J Phys Chem B, 2014. **118**(49): p. 14177-87.
10. Zhao, X. and Y.T. Yu, *Targeted pre-mRNA modification for gene silencing and regulation.* Nat Methods, 2008. **5**(1): p. 95-100.
11. Ge, J., H. Liu, and Y.T. Yu, *Regulation of pre-mRNA splicing in Xenopus oocytes by targeted 2'-O-methylation.* RNA, 2010. **16**(5): p. 1078-85.
12. Maden, B.E., et al., *Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA.* Biochimie, 1995. **77**(1-2): p. 22-9.
13. Huang, C., J. Karijolich, and Y.T. Yu, *Detection and quantification of RNA 2'-O-methylation and pseudouridylation.* Methods, 2016. **103**: p. 68-76.
14. 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, 2017. **114**(49): p. 12934-12939.
15. Birkedal, U., et al., *Profiling of ribose methylations in RNA by high-throughput sequencing.* Angew Chem Int Ed Engl, 2015. **54**(2): p. 451-5.
16. Krogh, N., et al., *Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity.* Nucleic Acids Res, 2016. **44**(16): p. 7884-95.
17. Gumienny, R., et al., *High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq.* Nucleic Acids Res, 2017. **45**(5): p. 2341-2353.
18. Lalonde, M.S., et al., *Exoribonuclease R in Mycoplasma genitalium can carry out both RNA processing and degradative functions and is sensitive to RNA ribose methylation.* RNA, 2007. **13**(11): p. 1957-68.
19. Taoka, M., et al., *Landscape of the complete RNA chemical modifications in the human 80S ribosome.* Nucleic Acids Res, 2018. **46**(18): p. 9289-9298.
20. Shubina, M.Y., Y.R. Musinova, and E.V. Sheval, *Nucleolar Methyltransferase Fibrillarin: Evolution of Structure and Functions.* Biochemistry (Mosc), 2016. **81**(9): p. 941-50.
21. Choi, J., et al., *2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation.* Nat Struct Mol Biol, 2018. **25**(3): p. 208-216.
22. Ringeard, M., et al., *FTSJ3 is an RNA 2'-O-methyltransferase recruited by HIV to avoid innate immune sensing.* Nature, 2019. **565**(7740): p. 500-504.
23. Yu, B., et al., *Methylation as a crucial step in plant microRNA biogenesis.* Science, 2005. **307**(5711): p. 932-5.

Figure legends

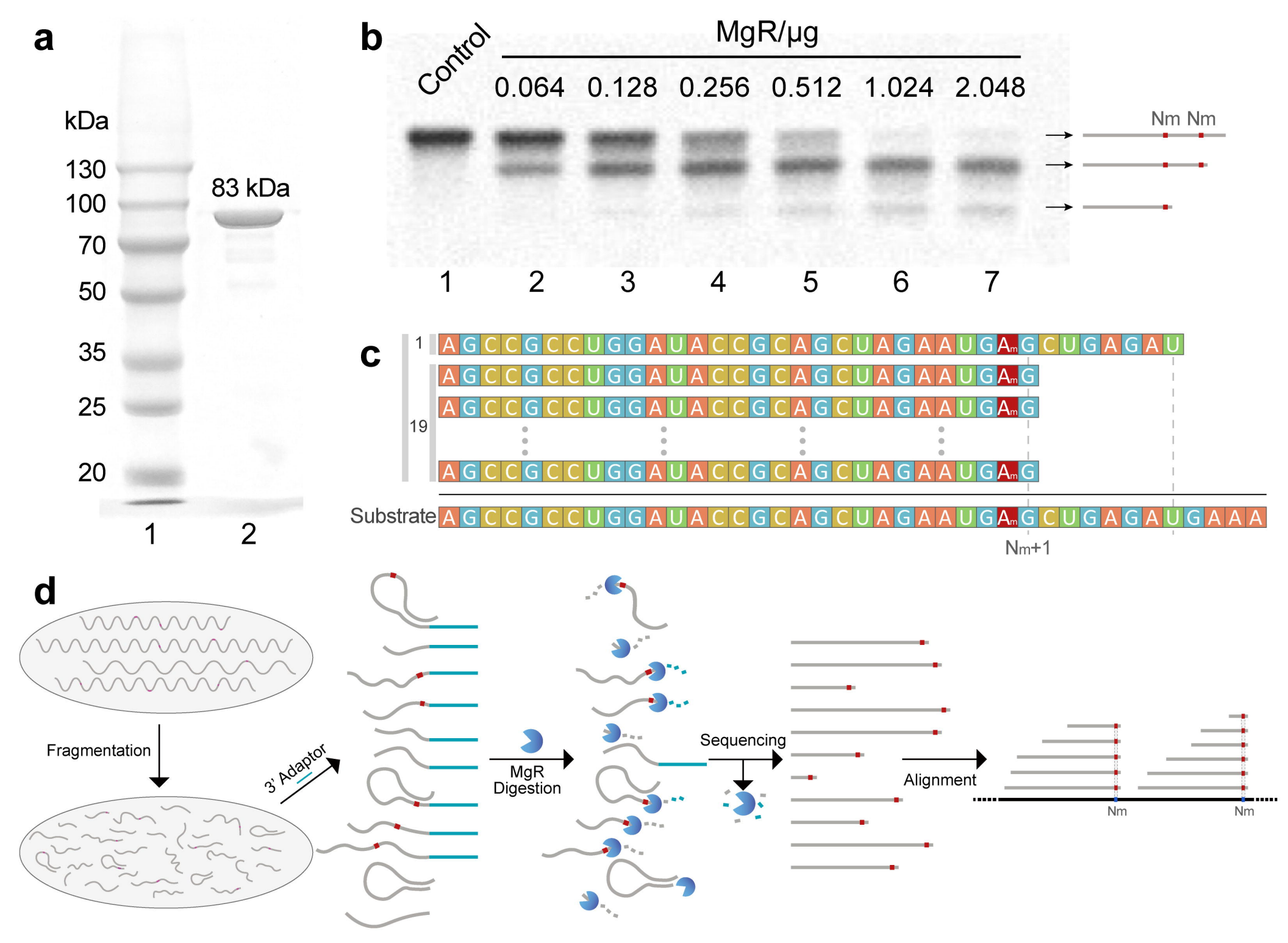
Fig. 1 | NJU-seq, an enzyme-based high throughput Nm sites detection method with base resolution. a, Purified MgR protein. The MgR protein was expressed in *E. Coli* and purified by Ni-NTA resin. The purified protein was separated by 12% SDS-PAGE gel and stained by coomassie blue. (lane1) Ladder; (lane2) MgR protein (83kDa). **b**, In vitro ssRNA hydrolysis assay. The ssRNA substrate UAACCUAUGAAGNmNNUNmNNCUC (N represents the mix of A\G\C\U) was reacted with different amount of MgR proteins as shown. Lane1-7:0, 0.064ug, 0.128ug, 0.256ug, 0.512ug, 1.024ug, 2.048ug MgR protein. **c**, TA clone MgR hydrolyzed ssRNA. ssRNA substrate (AGCCGCCUGGAUACCGCAGCUAGAAUGAmGCUGAGAUGAAA) was reacted with MgR protein, followed by TA cloned into *E. Coli*. The results demonstrated 20 sequences inserted from independent clones by sanger sequencing. 19 of 20 ssRNA products were ended at 1 ribonucleotide downstream of the Am site. **d**, NJU-seq process. 1) total RNA fragmentation. 2) 3' tail ligation. 3) MgR reaction. 4)NGS library construction. 5) Data analysis.

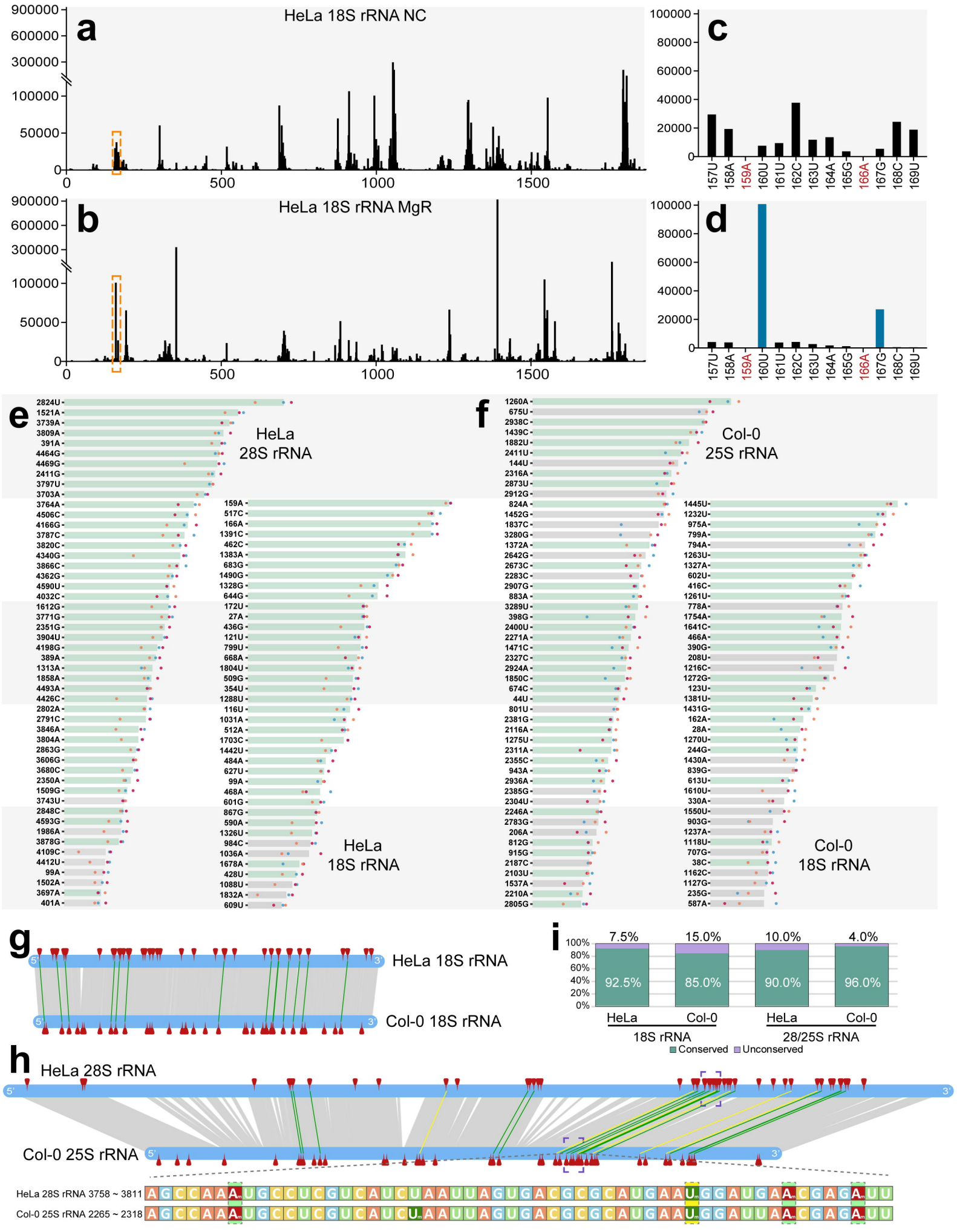
Fig. 2 | Nm sites in rRNA. a-d, Distribution of 3' ends of NGS fragments. a, **b**, The pictures illustrated 3' ends of NGS fragments mapped to 18S rRNA from NC and MgR treated group. **c**, **d**, The regions from position 157 to 169 were zoomed in. **e**, **f**, Top Nm sites detected of rRNA. The score of each site in rRNA was calculated by our algorism and listed from the highest. Each bar was calculated by three independent sample repeats showing with three dots. The green bar means the site was reported in the previous study, while the grey bar means the opposite. **e**, Top Nm sites detected of HeLa cell line rRNA. Top 50 of 28S rRNA and top 40 of 18S rRNA were presented above. **f**, Top Nm sites detected of *Arabidopsis* stem rRNA. Top 50 of 25S rRNA and top 40 of 18S rRNA were presented above. **g**, **h**, The comparison of Nm sites on rRNA from *Homo Sapiens* and *Arabidopsis*. Gray lines represented conserved regions between two species. Red triangles represented the location of Nm sites listed above. Each green line linked two Nm sites that were paired and both reported in the previous study. Each yellow line linked two Nm sites that were paired but only the Nm site from humans was reported. **g**, The comparison of Nm sites on 18s rRNA. **h**, The comparison of Nm sites on HeLa 28s rRNA and *Arabidopsis* 25s rRNA. A region of 25S/28S rRNA were zoomed in and shown. Four paired Nm sites and one unpaired Um site were included. **i**, Distribution of Nm sites in rRNA. Whether the top 50 in 25/28S rRNA and top 40 in 18S rRNA Nm sites were located in a conserved region or not was shown above. The conserved regions were determined by sequences alignment.

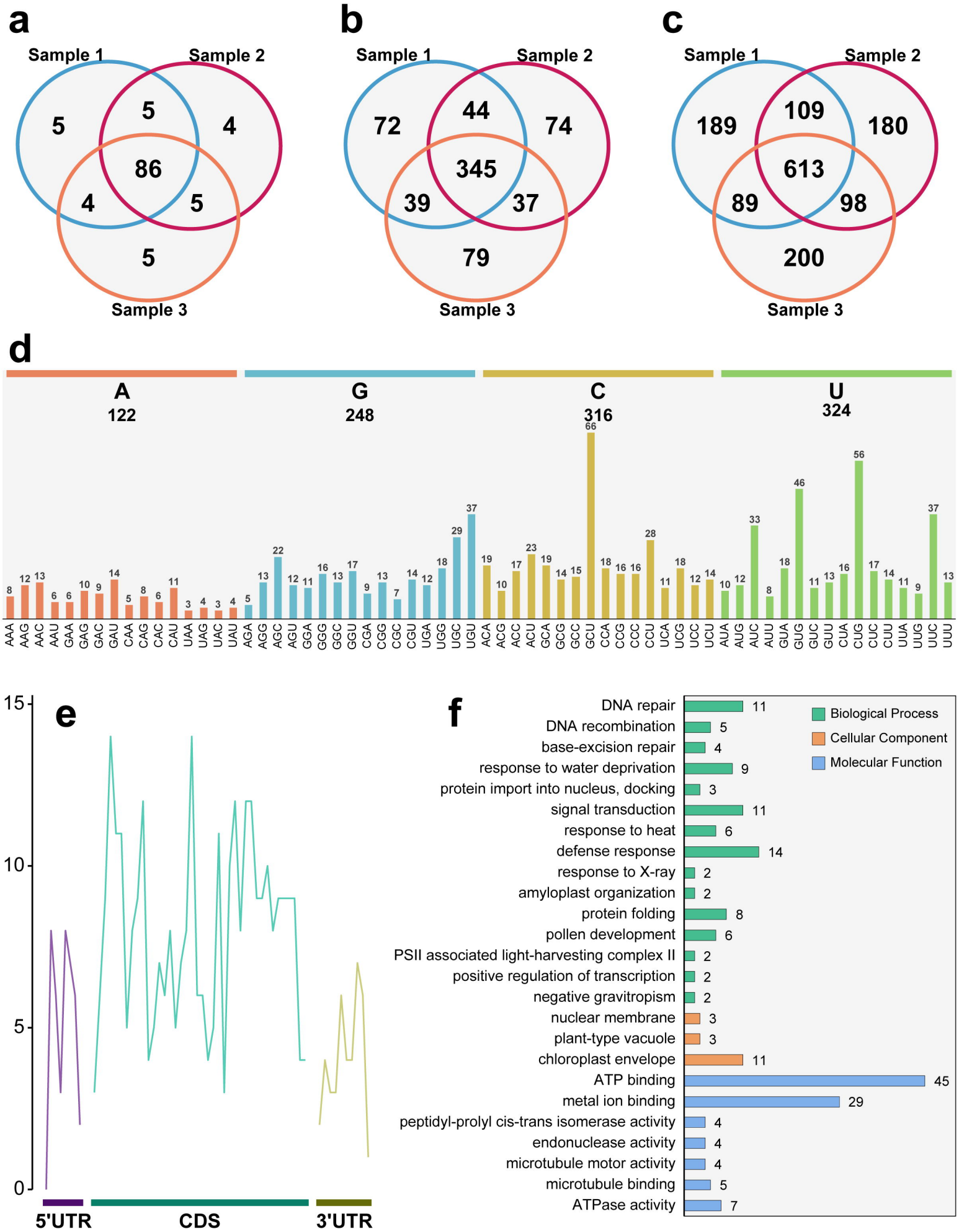
Fig. 3 | Nm modifications presented in mRNA. a-c, The consistency analysis of mRNA Nm sites from *Arabidopsis* stem. The top 100, 500 and 1000 Nm sites on mRNA were extracted from three

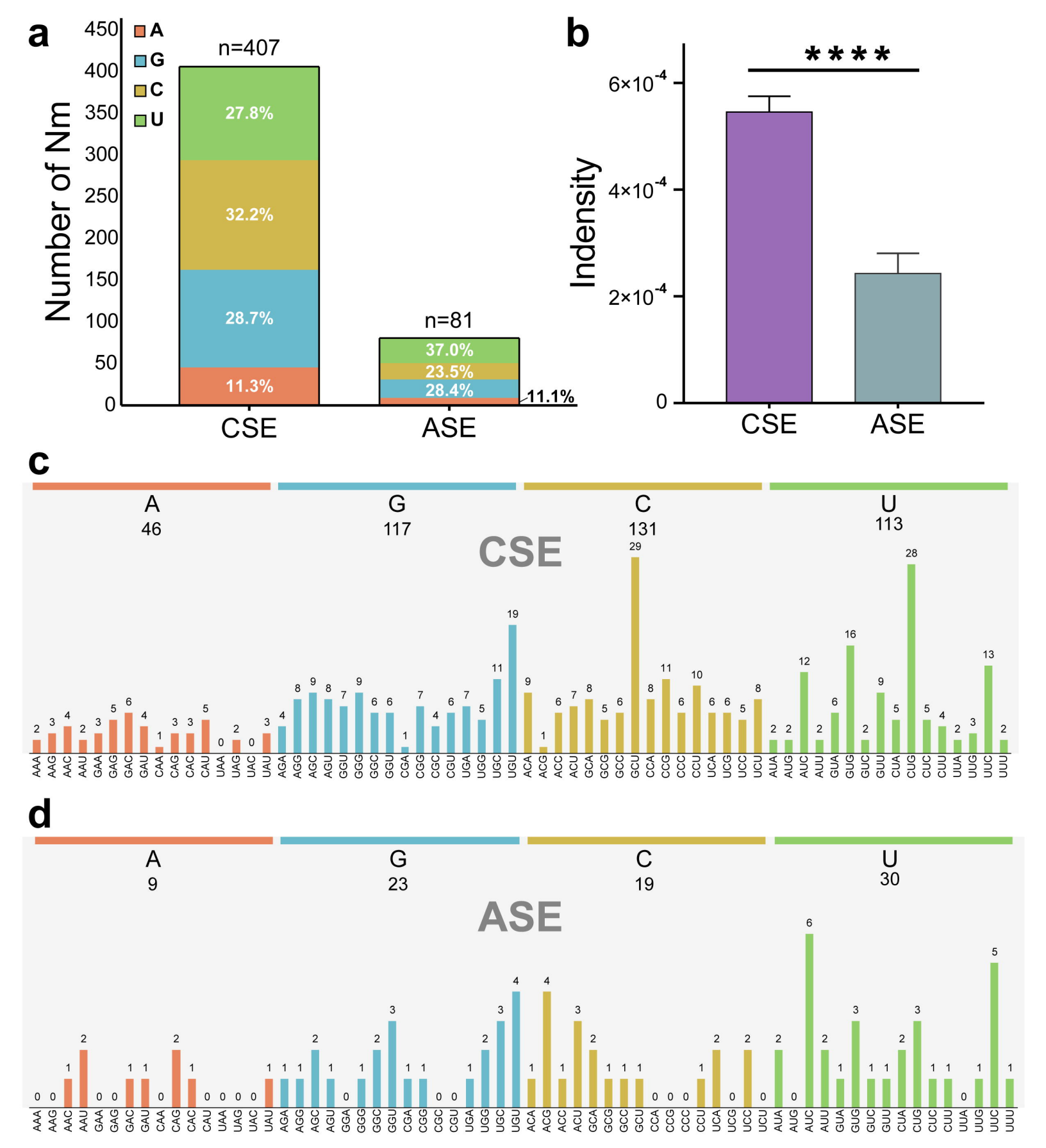
independent results of *Arabidopsis* stem. The common Nm sites shared with all three samples were 86 (86.0%), 345 (70.8%) and 613 (63.4%) as shown above. **d**, Nm sites Tri-ribonucleotides distribution in mRNA. Top 1010 Nm sites on mRNA shared in three samples of *Arabidopsis* stem were calculated by Nm, Nm-1 and Nm+1 sites as a Tri-ribonucleotides unit. **e**, Nm sites density on mRNA in different regions. 401 Nm sites shared with three samples on mRNA from *Arabidopsis* stem with full reference mRNA information (5' UTR/CDS/3'UTR) were calculated by relative position. **f**, Enrichment analysis of top 300 Nm sites located mRNAs. The functions of the top 300 Nm sites located mRNAs were enriched (listed in Supplementary Table 1). Green, orange and blue bars presented enrichment results by biological process, Cellular component and Molecular Function.

Fig. 4 | Nm participates in alternative splicing. a, Nm sites distribution in CSE and ASE regions. 488 Nm sites located in genes with multiple transcripts were calculated by the locations of CSE (Constitute Splicing Exon) or ASE (Alternative Splicing Exon). **b**, Nm sites density in CSE and ASE regions. The density of 488 Nm sites located in genes with multiple transcripts was calculated by the locations and total lengths of CSE or ASE. Difference analysis was done by paired t-test, $p<0.0001$. **c**, **d**, Ribonucleotides, and Tri-ribonucleotides units were also further calculated.


a
kDa
130
100
70
50
35
25
20
83 kDa
1
2
b
Control
MgR/μg
0.064
0.128
0.256
0.512
1.024
2.048
Nm Nm
1
2
3
4
5
6
7
c
1
19
Substrate
Nm+1
d
Fragmentation
3' Adaptor
MgR
Digestion
Sequencing
Alignment
Nm
Nm


a
HeLa 18S rRNA NC
b
HeLa 18S rRNA MgR
c
d
157U 158A 159A 160U 161U 162C 163U 164A 165G 166A 167G 168C 169U
e
HeLa
28S rRNA
HeLa
18S rRNA
f
Col-0
25S rRNA
Col-0
18S rRNA
g
HeLa 18S rRNA
Col-0 18S rRNA
i
7.5% 15.0% 10.0% 4.0%
92.5% 85.0% 90.0% 96.0%
HeLa Col-0 HeLa Col-0
18S rRNA 28/25S rRNA
Conserved Unconserved
h
HeLa 28S rRNA
Col-0 25S rRNA
HeLa 28S rRNA 3758 ~ 3811
Col-0 25S rRNA 2265 ~ 2318


a
Sample 1
Sample 2
Sample 3
5
5
4
86
4
5
5
b
Sample 1
Sample 2
Sample 3
72
44
74
345
39
37
79
c
Sample 1
Sample 2
Sample 3
189
109
180
613
89
98
200
d
A
122
G
248
C
316
U
324
AAA 8
AAG 12
AAC 13
AAU 6
GAA 6
GAG 10
GAC 9
GAU 14
CAA 5
CAG 8
CAC 6
CAU 11
UAA 3
UAG 4
UAC 3
UAU 4
AGA 5
AGG 13
AGC 22
AGU 12
GGA 11
GGG 16
GGC 13
GGU 17
CGA 9
CGG 13
CGC 7
CGU 14
UGA 12
UGG 18
UGC 29
UGU 37
ACA 19
ACG 10
ACC 17
ACU 23
GCA 19
GCG 14
GCC 15
GCU 66
CCA 18
CCG 16
CCC 16
CCU 28
UCA 11
UCG 18
UCC 12
UCU 14
AUA 10
AUG 12
AUC 33
AUU 8
GUA 18
GUG 46
GUC 11
GUU 13
CUA 16
CUG 56
CUC 17
CUU 14
UUA 11
UUG 9
UUC 37
UUU 13
e
15
10
5
0
5'UTR
CDS
3'UTR
f
Biological Process
Cellular Component
Molecular Function
DNA repair 11
DNA recombination 5
base-excision repair 4
response to water deprivation 9
protein import into nucleus, docking 3
signal transduction 11
response to heat 6
defense response 14
response to X-ray 2
amyloplast organization 2
protein folding 8
pollen development 6
PSII associated light-harvesting complex II 2
positive regulation of transcription 2
negative gravitropism 2
nuclear membrane 3
plant-type vacuole 3
chloroplast envelope 11
ATP binding 45
metal ion binding 29
peptidyl-prolyl cis-trans isomerase activity 4
endonuclease activity 4
microtubule motor activity 4
microtubule binding 5
ATPase activity 7


a
Number of Nm
A
G
C
U
n=407
27.8%
32.2%
28.7%
11.3%
n=81
37.0%
23.5%
28.4%
11.1%
CSE
ASE
b
Indensity
****
CSE
ASE
c
A
46
G
117
C
131
U
113
CSE
d
A
9
G
23
C
19
U
30
ASE