Enhancement of synthetic mRNA translation efficiency through engineered poly(A) tails

Yusheng Liu^{1,4} ✉, Hu Nie^{1,2,4}, Rongrong Sun^{1,2}, Jiaqiang Wang³ ✉, and Falong Lu^{1,2} ✉

¹State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China.

²University of Chinese Academy of Sciences, Beijing 100049, China.

³College of Life Science, Northeast Agricultural University, Harbin 150030, China.

⁴These authors contributed equally: Yusheng Liu, Hu Nie.

✉e-mail: liuys1126@foxmail.com; wangjiaqiang@neau.edu.cn; flflu@genetics.ac.cn

SUMMARY

In vitro transcribed (IVT) mRNA represents a new class of drug in both therapeutics and vaccines. Improving the translation efficiency of IVT mRNA remains a core challenge for mRNA-based applications. Here, using IVT mRNAs with poly(A) tails containing non-A residues which were recently revealed to be widespread in RNA poly(A) tails^{1,2}, we unexpectedly find that non-A residues can effectively promote the mRNA translation. To further support our finding, we provide evidence that non-A residues associated with enhanced mRNA translation efficiency transcriptome-wide in mouse and human cells. Together, our study provides a novel approach to enhance mRNA translation efficiency by inclusion of non-A residues in the mRNA poly(A) tails, holding great potential to promote mRNA-based therapeutics and vaccines.

INTRODUCTION

In recent year, synthetic mRNA has become a promising approach for therapeutic gene delivery and vaccination^{3,4}. Tremendous success has been achieved in mRNA vaccine against the ongoing global Covid-19 pandemic^{5,6}. Great efforts have been invested to modify the mRNA to improve its stability and translation efficiency, including the 5' cap, 5'UTR, 3'UTR, codon optimization, modified nucleosides and poly(A) tail³. The presence of a poly(A) tail is essential for mRNA stability, nuclear export and translation⁷⁻¹⁶. However, previous effort mainly focused on the length of the poly(A) tail, as previous knowledge of poly(A) tail mainly considers it as a structure component there and certain length ensures its translation and stability³.

Poly(A) tail information is lost in most transcriptome analysis because of the intrinsic difficulty of sequencing long homopolymeric stretches using second-generation sequencing platforms. Therefore, the roles of RNA poly(A) tails in regulating the fate and functions of RNAs are highly understudied. Researchers have used the TAIL-seq and PAL-seq (poly(A)-tail length profiling by sequencing) methods to estimate the lengths of poly(A)

tails on the Illumina platform^{17,18}. Studies of RNA poly(A) tails in oocytes of multiple species, including *Drosophila*, *Xenopus* and mice, revealed a clear link between the length of the poly(A) tail and the efficiency of mRNA translation¹⁸⁻²². Surprisingly, the length of a poly(A) tail, as long as it is longer than 20 nt, minimally affects mRNA translation in somatic cells²³, leading to the question of whether poly(A) tails can regulate translation other than being a structural component of mRNA in somatic cells.

TAIL-seq revealed that the 3' ends of mRNA poly(A) tails can be modified by end extension with other bases that can regulate the RNA stability, constituting another layer of mRNA stability regulation^{17,24-28}. Short deadenylated poly(A) tails can be terminally U modified by TUT4/7, which can mark the target RNA for degradation²⁴. In contrast, G modification catalyzed by TENT4A/B can protect the poly(A) tail from active deadenylation by the CCR4-NOT complex to stabilize the modified mRNA²⁵.

The base composition within the body of poly(A) tails remained largely unknown until the recent development of the PAIso-seq and FLAM-seq methods on the PacBio platform, which allow long homopolymers to be sequenced accurately^{1,2,29}. These methods revealed a novel group of RNA modifications, non-A residues in the RNA poly(A) tails (referred as non-A residues here after), that are widespread in the RNA poly(A) tails in mouse oocytes and human cell lines^{1,2}, suggesting the existence a novel layer of RNA post-transcriptional regulation through non-A residues. Here, we reveal that non-A residues can enhance translation efficiency of IVT mRNA. We further provide evidence for this translation-enhancing function of non-A residues using unbiased genome-wide translation analysis. Together, we reveal a novel approach to enhance the translation of IVT mRNA using non-A residues, which will pave a new way of great potential to promote mRNA-based therapeutics and vaccines.

RESULTS

Non-A residues promote translation efficiency of IVT mRNA

Immediately after we and the Rajewsky lab identified non-A residues in mouse and human cells^{1,2}, we asked what are the roles of these newly identified RNA residues in mRNA function. We synthesized mRNAs encoding fluorescent reporters with poly(A) tails of the same length but with different numbers of non-A residues using IVT (Fig. 1a). Then we transfected these IVT mRNAs into HeLa cells and monitored the encoded reporter protein and mRNA levels (Fig. 1b). RNA poly(A) tail is traditionally known as pure A sequence that can support translation. Therefore, our initial hypothesis was that mixed incorporation of nucleotides identified in the poly(A) tails might impair the role of poly(A) tail in supporting translation, leading to decreased or aborted translation. However, we unexpectedly observed obvious higher levels of EGFP fluorescence in cells transfected with EGFP mRNAs containing non-A residues compared with those transfected with EGFP mRNAs with a pure A tail (Fig. 1c), which was opposite to our initial hypothesis. Then, we used fluorescent-activated cell sorting (FACS) to quantify the level of fluorescent reporter translated from mRNA with non-A98-2 poly(A) tail and pure A tail. We found higher level of EGFP fluorescence for mRNA with the non-A98-2 tail; the same observation was seen when the mCherry reporter was used instead of EGFP (Fig. 1d). These results demonstrate that the non-A residues increase the protein production of multiple IVT mRNAs which is not limited to one special mRNA coding sequence.

To investigate whether the increase of protein production by non-A residues is through increased translation efficiency or RNA stability, we then quantified the protein and mRNA levels within these cells by western blot and quantitative PCR (qPCR), respectively. Consistent with the observations under the microscope and FACS, the cells transfected with mRNAs containing non-A residues showed an increased amount of mCherry and EGFP protein production (Fig. 1e). qPCR analysis did not show higher mRNA levels for samples containing non-A residues, indicating that increased protein production was caused by increased translation efficiency rather than increased mRNA stability (Fig. 1e). To further confirm this effect of non-A residues in promoting translation, we also tested

different time points after mRNA transfection and found consistent stronger fluorescence from 6 to 120 hours (Extended Data Fig. 1). Taken together, these data clearly demonstrate that non-A residues can greatly enhance the translation of synthetic mRNA.

non-A residues are associated with high translation efficiency *in vivo*

The above findings showed enhanced translation efficiency by non-A residues in IVT mRNAs. In order to support the general function of non-A residues in enhancing translation that we unexpectedly discovered, we asked whether non-A residues presented in endogenous mRNAs can enhance the translation efficiency of these mRNAs. Therefore, we investigated the relationship between translation and non-A residues using the available ribosome engagement data and the non-A residues data for mouse GV oocytes and HeLa cells^{1,2,19,23,30}.

Consistent with previous observations^{18,19,23,31}, poly(A) tail length was positively correlated with translation efficiency as measured by ribosome engagement in mouse GV oocytes (Fig. 2a), whereas no such correlation was observed in HeLa cells (Fig. 2b). This global positive association between poly(A) tail length and translation efficiency is seen only in reproductive cells but not in somatic cells²³. In addition, global translation efficiency varies a lot during different stages of mitotic cell cycle, while the poly(A) tail length changed only moderately for a small proportion of genes²³. However, we noticed extensive changes in non-A residues during somatic cell cycle³². Since poly(A) tail length does not contribute to translational regulation in somatic cells as seen in reproductive cells, we asked whether the non-A residues, the base composition of poly(A) tails, affect the translation efficiency.

Interestingly, genes with G and C residues showed significantly higher translation efficiencies in both germ cells and somatic cells (Fig. 2a, b). However, U residues were not always associated with increased translation efficiency, especially in mouse GV oocytes (Fig. 2a, b). This may be related to the much higher abundance of U residues in GV oocytes,

especially consecutive oligo-U residue (Extended Data Fig. 2), which might result in uridylation mark RNA for degradation^{24,27}. Given that mRNA translation ultimately results in the accumulation of proteins, we investigated the relationship between non-A residues and protein abundance in mouse GV oocytes and HeLa cells, for which the mass-spectrometry based proteomic data were available^{33,34}. The results showed that proteins encoded by genes with non-A residues showed significantly increased abundance in both GV oocytes and HeLa cells, which is consistent with the ribosome engagement data (Fig. 2c, d). These results reveal that the non-A residues are indeed prevalent regulators of translation of endogenous mRNAs in both germline and somatic cells, demonstrating that poly(A) tails can be critical regulators of mRNA translation rather than being simple structural components. Therefore, these unbiased genome-wide analysis of endogenous mRNA non-A residues and translation well support our finding that non-A residues promote translation of IVT mRNA.

DISCUSSION

In recent years, mRNA-based therapeutics have become a new focus in the pharmaceutical industry^{3,4}. Compared with protein-based therapy, mRNA-based therapy is much more cost-effective because mRNAs are easy to design and produce. mRNA vaccines allow the fastest pre-clinical development, and have proven great potential in responding to emerging outbreaks of infectious diseases^{5,6,35,36}. For example, mRNA vaccines designed and synthesized within one month from Moderna and BioNTech were among the first vaccines progressing into phase 1/2 clinical trials during the current ongoing global Covid-19 pandemic³⁷⁻⁴⁰. Moreover, these mRNA vaccines are the first approved vaccines and proven of exceptional efficacy in preventing Covid-19^{5,6}. Improving translation efficiency has been one of the main technical challenges in mRNA-based therapeutics and vaccines^{3,41}. In this study, we demonstrate that non-A residues are conserved and abundant novel RNA modifications in diverse eukaryotes that can enhance mRNA translation (Fig. 2e). The

discovery that non-A residues enhance translation of IVT mRNA provides a completely new way to promote the development of mRNA-based therapeutics and vaccines.

In order to further confirm our finding that non-A residues enhance translation of IVT mRNAs, we provide evidence that non-A residues associate with enhanced mRNA translation in human and mouse cells, revealing the non-A residues as endogenous mRNA translation regulating factors. Our recent study reveal the conservation and prevalence of non-A residues in mRNA poly(A) tails across eukaryotes, suggesting universal functions of them in regulating diverse biological processes²⁹. As the above data reveal that non-A residues enhance mRNA translation in human and mouse cells, the prevalence of non-A residues suggest that non-A residues may share similar function in enhancing mRNA translation across eukaryotic species. Therefore, the including of non-A residues in the synthetic mRNAs will likely be a novel approach to enhance the protein production in mRNA-based applications across the eukaryotic species. What enzymes are responsible for the incorporation of these non-A residues is an interesting question to answer for mechanistic study *in vivo*. The mechanisms by which non-A residues promote mRNA translation and whether G, C and U residues are using different mechanisms are the next important questions awaiting further investigations.

Previous methods to improve IVT mRNA stability, translational properties and immunogenicity largely relies on chemically modified nucleosides in mRNA, such as pseudouridine, 1-methyl-pseudouridine 2-thiouridine, 5-methyluridine, 5-methylcytidine or N6-methyladenosine^{3,4}. However, chemically modified nucleosides such as pseudouridine and N6-methyladenosine severely decrease synthesis efficiency and increase error rate associated during IVT mRNA synthesis with the use of chemically modified nucleosides⁴². Here, we discover a completely new way to improve mRNA translation by mixed incorporation of non-A residues within the mRNA poly(A) tails. These non-A residues are regular nucleotides, which can be easily designed into the mRNA synthesis template to ensure 100% reproducible incorporation into the final full length

mRNA products. Moreover, as this new approach is different from all the current available approaches in improving the pharmacokinetics of mRNA drugs, it can be readily incorporated into every existing mRNA developing pipelines together with other methods, such as combined with the use of modified nucleosides that reduce immunogenicity and increase stability. In summary, the new approach described in the current study using non-A residues to enhance IVT mRNA translation will have considerable translational potential in pushing forward the mRNA therapeutics and vaccines.

METHODS

mRNA preparation, cell culture and transfection

The template used for *in vitro* transcription (IVT) were amplified by PCR from the pcDNA3.1-EGFP/mCherry-3XFlag plasmids using T7-F forward primer and different reverse primers (Supplementary Table 1). The capped mRNAs were *in vitro* synthesized with HiScribe T7 ARCA mRNA Kit (NEB) and purified with an RNA Clean & Concentrator-25 kit (Zymo Research) according to the manufacturers' instructions. The purified mRNA was dissolved in RNase-free water and stored at -80 °C until use. HeLa cells were grown in DMEM (Invitrogen) containing 10% FBS (Gibco) at 37 °C in a humidified 5% CO₂ incubator. For mRNA transfection, HeLa cells were transfected with Lipofectamine MessengerMAX reagent (Invitrogen). The transfected cells were used for imaging and FACS analysis or harvested for other downstream experiments at the indicated time points.

Western blot analysis

HeLa cells were lysed with SDS loading buffer containing protease inhibitor cocktail (Roche), boiled and separated by SDS-PAGE. The samples were then transferred onto PVDF membrane. The membrane was incubated with primary antibody (mouse anti-FLAG antibody (Sigma, F1804; 1:1000); mouse anti-Tubulin (Sigma, T9026; 1:1000) overnight at 4 °C. The next day, after washing 3 times with TBST, the membrane was incubated with fluorescent dye-labeled secondary antibody (LI-COR, 925-32210, 1:1000) for 1 hour at RT. The membrane was then scanned using an Odyssey infrared scanner (LI-COR).

Quantitative real-time RT-PCR

The total RNA was isolated from HeLa cells with Direct-zol RNA MicroPrep (Zymo Research) according to the manufacturer's instructions. The SuperScript II Reverse Transcriptase (Invitrogen) was used to synthesize cDNA from RNA. Quantitative PCR was performed using TB Green Premix Ex Taq (Takara) with specific primers (Supplementary Table 1). Measurements were performed with three independent biological replicates.

Detection of non-A residues in poly(A) tails

To minimize errors introduced by sequencer, we used clean CCS reads with at least ten passes to call non-A residues in poly(A) tails. G, C, U (presented as T in CCS reads) were counted in poly(A) tail of each CCS reads. The percentage of transcripts with non-A residues of a gene was the number of CCS reads with at least one G, C or U residue divided by the total number of CCS reads with poly(A) tails of length at least 1 nt for the given gene.

TE and the protein abundance analysis

Translation efficiency (TE) was calculated based on ribosome engagement data and related RNA-seq data (GSE135525 for mouse GV oocyte RiboTag/RNA-seq¹⁹, GSE79664 for HeLa Ribo-seq²³). Genes with at least ten reads measured in both ribosome engagement and RNA-seq data were included in the analysis. CPM (counts per million) were calculated to represent the relative gene expression levels. Translation efficiency was calculated by dividing the CPM of ribosome engagement fragments to that of the total mRNA. Protein copy-number estimates assigned to each gene were used to represent the protein abundance in HeLa cells³⁴. The number of observed spectra assigned to each gene were used to represent the protein abundance in mouse GV oocytes³³. Other than the PAIso-seq data obtained in this study, HeLa cell poly(A) tail information from the previous poly(A) tail analysis were used for the translation efficiency and protein abundance analysis¹.

Data availability

The ccs data in bam format from PAIso-seq1 experiments (GV oocytes) will be available at Genome Sequence Archive hosted by National Genomic Data Center. This study also includes analysis of the following published data: Legnini et al. (Gene Expression Omnibus database (GEO) accession no. GSE126465), Liu et al. (Sequence Read Archive database (SRA) accession no. PRJNA529588), Luong et al. (GEO accession no. GSE135525), Park et al. (GEO accession no. GSE79664), and Bekker-Jensen et al. (ProteomeXchange Consortium no. PXD004452). Custom scripts used for data analysis will be available upon

request.

Acknowledgements

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24020203), the National Key Research and Development Program of China (2018YFA0107001), National Natural Science Foundation of China (31970588), Natural Science Foundation of Heilongjiang province (YQ2020C003), the China Postdoctoral Science Foundation (2020M670516, 2020T130687), and the State Key Laboratory of Molecular Developmental Biology.

Author Contributions

Yusheng Liu, Jiaqiang Wang and Falong Lu conceived the project and designed the study. Yusheng Liu and Rongrong Sun performed the synthetic mRNA study in culture cells. Yusheng Liu, Hu Nie, Jiaqiang Wang and Falong Lu analyzed the sequencing data. Yusheng Liu and Jiaqiang Wang organized all figures. Yusheng Liu, Jiaqiang Wang and Falong Lu supervised the project. Yusheng Liu, Jiaqiang Wang and Falong Lu wrote the manuscript with the input from the other authors.

Competing interests: Y.L. and F.L. are named inventors on a patent application filed by Institute of Genetics and Developmental Biology covering the discovery in this manuscript.

REFERENCES

- 1 Legnini, I., Alles, J., Karaikos, N., Ayoub, S. & Rajewsky, N. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat Methods* **16**, 879-886, doi:10.1038/s41592-019-0503-y (2019).
- 2 Liu, Y., Nie, H., Liu, H. & Lu, F. Poly(A) inclusive RNA isoform sequencing (PALso-seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. *Nat Commun* **10**, 5292, doi:10.1038/s41467-019-13228-9 (2019).
- 3 Sahin, U., Kariko, K. & Tureci, O. mRNA-based therapeutics--developing a new class of drugs. *Nat Rev Drug Discov* **13**, 759-780, doi:10.1038/nrd4278 (2014).
- 4 Kariko, K. In vitro-Transcribed mRNA Therapeutics: Out of the Shadows and Into the Spotlight. *Mol Ther* **27**, 691-692, doi:10.1016/j.ymthe.2019.03.009 (2019).
- 5 Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **384**, 403-416, doi:10.1056/NEJMoa2035389 (2021).
- 6 Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **383**, 2603-2615, doi:10.1056/NEJMoa2034577 (2020).
- 7 Barabino, S. M. L. & Keller, W. Last but not least: Regulated poly(A) tail formation. *Cell* **99**, 9-11, doi:10.1016/S0092-8674(00)80057-4 (1999).
- 8 Sachs, A. The role of poly(A) in the translation and stability of mRNA. *Curr Opin Cell Biol* **2**, 1092-1098, doi:10.1016/0955-0674(90)90161-7 (1990).
- 9 Eckmann, C. R., Rammelt, C. & Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* **2**, 348-361, doi:10.1002/wrna.56 (2011).
- 10 Groisman, I., Huang, Y. S., Mendez, R., Cao, Q. & Richter, J. D. Translational control of embryonic cell division by CPEB and maskin. *Cold Spring Harb Symp Quant Biol* **66**, 345-351, doi:10.1101/sqb.2001.66.345 (2001).
- 11 Groisman, I., Jung, M. Y., Sarkissian, M., Cao, Q. & Richter, J. D. Translational control of the embryonic cell cycle. *Cell* **109**, 473-483, doi:10.1016/s0092-8674(02)00733-x (2002).
- 12 Nicholson, A. L. & Pasquinelli, A. E. Tales of Detailed Poly(A) Tails. *Trends Cell Biol* **29**, 191-200, doi:10.1016/j.tcb.2018.11.002 (2019).
- 13 Richter, J. D. Cytoplasmic polyadenylation in development and beyond. *Microbiol Mol Biol Rev* **63**, 446-456 (1999).
- 14 Richter, J. D. CPEB: a life in translation. *Trends Biochem Sci* **32**, 279-285, doi:10.1016/j.tibs.2007.04.004 (2007).
- 15 Villalba, A., Coll, O. & Gebauer, F. Cytoplasmic polyadenylation and translational control. *Curr Opin Genet Dev* **21**, 452-457, doi:10.1016/j.gde.2011.04.006 (2011).
- 16 Weill, L., Belloc, E., Bava, F. A. & Mendez, R. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol* **19**, 577-585, doi:10.1038/nsmb.2311 (2012).
- 17 Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* **53**, 1044-1052, doi:10.1016/j.molcel.2014.02.007 (2014).
- 18 Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals

an embryonic switch in translational control. *Nature* **508**, 66-71, doi:10.1038/nature13007 (2014).

- 19 Luong, X. G., Daldello, E. M., Rajkovic, G., Yang, C. R. & Conti, M. Genome-wide analysis reveals a switch in the translational program upon oocyte meiotic resumption. *Nucleic Acids Res* **48**, 3257-3276, doi:10.1093/nar/gkaa010 (2020).
- 20 Lim, J., Lee, M., Son, A., Chang, H. & Kim, V. N. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev* **30**, 1671-1682, doi:10.1101/gad.284802.116 (2016).
- 21 Eichhorn, S. W. *et al.* mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. *eLife* **5**, doi:10.7554/eLife.16955 (2016).
- 22 Yang, F., Wang, W., Cetinbas, M., Sadreyev, R. I. & Blower, M. D. Genome-wide analysis identifies cis-acting elements regulating mRNA polyadenylation and translation during vertebrate oocyte maturation. *RNA* **26**, 324-344, doi:10.1261/rna.073247.119 (2020).
- 23 Park, J. E., Yi, H., Kim, Y., Chang, H. & Kim, V. N. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol Cell* **62**, 462-471, doi:10.1016/j.molcel.2016.04.007 (2016).
- 24 Lim, J. *et al.* Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* **159**, 1365-1376, doi:10.1016/j.cell.2014.10.055 (2014).
- 25 Lim, J. *et al.* Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science* **361**, 701-704, doi:10.1126/science.aam5794 (2018).
- 26 Morgan, M. *et al.* A programmed wave of uridylation-primed mRNA degradation is essential for meiotic progression and mammalian spermatogenesis. *Cell Res* **29**, 221-232, doi:10.1038/s41422-018-0128-1 (2019).
- 27 Morgan, M. *et al.* mRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* **548**, 347-351, doi:10.1038/nature23318 (2017).
- 28 Chang, H. *et al.* Terminal Uridyltransferases Execute Programmed Clearance of Maternal Transcriptome in Vertebrate Embryos. *Mol Cell* **70**, 72-82 e77, doi:10.1016/j.molcel.2018.03.004 (2018).
- 29 Liu, Y., Nie, H., Zhang, Y., Lu, F. & Wang, J. Comprehensive analysis of mRNA poly(A) tail reveals complex and conserved regulation. *bioRxiv Processing* (2021).
- 30 Liu, Y. *et al.* Abundant non-A residues in the poly(A) tail orchestrate the mouse oocyte-to-embryo transition. *bioRxiv Processing* (2021).
- 31 Liu, Y. *et al.* Poly(A) tail length is a major regulator of maternal gene expression during the mammalian oocyte-to-embryo transition. *bioRxiv Processing* (2021).
- 32 Liu, Y., Nie, H. & Lu, F. Dynamic RNA 3' Uridylation and Guanylation during Mitosis. *iScience* **23**, 101402, doi:10.1016/j.isci.2020.101402 (2020).
- 33 Wang, S. *et al.* Proteome of mouse oocytes at different developmental stages. *Proc Natl Acad Sci U S A* **107**, 17639-17644, doi:10.1073/pnas.1013185107 (2010).
- 34 Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* **4**, 587-599 e584, doi:10.1016/j.cels.2017.05.009 (2017).

- 35 Deweerdt, S. Rna Therapies Explained. *Nature* **574**, S2-S3, doi:DOI 10.1038/d41586-019-03068-4 (2019).
- 36 Zhong, Z. F. *et al.* mRNA therapeutics deliver a hopeful message. *Nano Today* **23**, 16-39, doi:10.1016/j.nantod.2018.10.005 (2018).
- 37 Thanh Le, T. *et al.* The COVID-19 vaccine development landscape. *Nat Rev Drug Discov*, doi:10.1038/d41573-020-00073-5 (2020).
- 38 Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines - a new era in vaccinology. *Nature Reviews Drug Discovery* **17**, 261-279, doi:10.1038/nrd.2017.243 (2018).
- 39 Wang, F., Zuroske, T. & Watts, J. K. RNA therapeutics on the rise. *Nat Rev Drug Discov*, doi:10.1038/d41573-020-00078-0 (2020).
- 40 Mulligan, M. J. *et al.* Phase 1/2 Study to Describe the Safety and Immunogenicity of a COVID-19 RNA Vaccine Candidate (BNT162b1) in Adults 18 to 55 Years of Age: Interim Report. *medRxiv*, 2020.2006.2030.20142570, doi:10.1101/2020.06.30.20142570 (2020).
- 41 Zhang, C., Maruggi, G., Shan, H. & Li, J. Advances in mRNA Vaccines for Infectious Diseases. *Front Immunol* **10**, 594, doi:10.3389/fimmu.2019.00594 (2019).
- 42 Potapov, V. *et al.* Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Research* **46**, 5753-5763, doi:10.1093/nar/gky341 (2018).

Fig. 1: Non-A residues promote translation of IVT mRNA

FIGURE LEGENDS

Figure 1. non-A residues promote translation of IVT mRNA.

- a**, Schematic of reporter mRNAs with different non-A residues in poly(A) tails of the same length.
- b**, Schematic of the translation efficiency reporter system based on transcripts with different types of poly(A) tails.
- c**, Images of HeLa cells 18 hours post transfection with reporter mRNAs with different poly(A) tails. Scale bar, 100 μ m.
- d**, FACS quantification of reporter mRNA expression in HeLa cells 42 hours post transfection.
- e**, Quantification of mRNA level by qRT-PCR (left) and protein level by western blot (right, anti-FLAG antibody was used to detect the FLAG tag for all reporters) 42 hours post transfection of HeLa cells with the reporter mRNA.

Fig. 2: Endogenous non-A residues are associated with enhanced translation efficiency

Figure 2. Endogenous non-A residues are associated with enhanced translation efficiency.

a, b, Genome-wide relationship between poly(A) tail length^{1,2} and ribosome engagement (TE) (left) and cumulative distribution function (CDF) plots of mRNA-normalized TE for genes with or without non-A residues (right) in mouse GV oocytes¹⁹ (gene number: with G, n = 4,104; without G, n = 1,363; with C, n = 4,109; without C, n = 1,358; with U, n = 4,116; without U, n = 1,351) (**a**), and HeLa cells²³ (gene number: with G, n = 2,075; without G, n = 810; with C, n = 2,323; without C, n = 562; with U, n = 1,884; without U, n = 1,001) (**b**). For GV oocyte sample, genes with 25% or more poly(A)+ mRNA containing G, C, or U residues were considered as genes with G, C, or U residues; while those containing less than 25% were considered as genes without G, C, or U residues. For HeLa sample, the overall level of non-A residues was much lower than that in GV oocytes. Therefore, genes with or without poly(A)+ mRNA containing G, C, or U residues were considered as genes with or without G, C, or U residues. Genes with at least 10 ccs reads were included in the analysis. In the violin plots, median values are represented by red lines and the first and third quartiles are represented by black, dashed lines. *p* value was calculated by two-tailed Student's *t* test for the relationship between length and TE and by Kolmogorov-Smirnov (KS) test for the relationship between non-A residues and TE.

c, d, CDF plots of protein abundance quantified by mass spectrometry for genes with or without non-A residues in GV oocytes³³ (gene number: with G, n = 2,307; without G, n = 768; with C, n = 2,308; without C, n = 767; with U, n = 2,306; without U, n = 769) (**c**), and HeLa cells³⁴ (gene number: with G, n = 2,398; without G, n = 1,029; with C, n = 2,694; without C, n = 733; with U, n = 2,166; without U, n = 1,261) (**d**). The *p* value was calculated by KS test.

e, A model for the full view of a complete single mRNA molecule. The magnified part highlights the non-A residues in RNA poly(A) tails, which are conserved in eukaryotes. Other known RNA modifications such as m6A, m5C, ac4C and terminal non-A

Extended Data Fig. 1. Quantification of fluorescent reporter translated from IVT mRNA at different time points.

Extended Data Fig. 1. Quantification of fluorescent reporter translated from IVT mRNA at different time points.

b

modifications are also labeled. The non-A residues can enhance the translation of the mRNA.

Extended Data Fig. 1|. Quantification of fluorescent reporter translated from IVT mRNA at different time points.

a, Quantification of reporter EGFP mRNA expression in HeLa cells by FACS at different time points post transfection. Control: cells not transfected with reporter mRNA. PURE-A98 and Non-A98-2: cells transfected with reporter mRNAs with the PURE-A98 and Non-A98-2 poly(A) tails, respectively, as shown in Fig. 1a.

b, Relative fluorescence intensity quantification of reporter mRNA expression in HeLa cells at different time points by FACS shown in panel a.

a

b

Extended Data Fig. 2|. Abundance of non-A residues.

Proportion of the transcripts containing non-A residues **(a)**, and mono- and oligo-Us **(b)** in mouse GV oocytes² and HeLa cells¹.