Efficient and quantitative high-throughput tRNA sequencing

Guanqun Zheng1,7, Yidan Qin2,7, Wesley C Clark1, Qing Dai3, Chengqi Yi1,8, Chuan He1,3–5, Alan M Lambowitz2,8 & Tao Pan1,4,8

Despite its biological importance, tRNA has not been adequately sequenced by standard methods because of its abundant post-transcriptional modifications and stable structure, which interfere with cDNA synthesis. We achieved efficient and quantitative tRNA sequencing in HEK293T cells by using engineered demethylases to remove base methylations and a highly processive thermostable group II intron reverse transcriptase to overcome these obstacles. Our method, DM-tRNA-seq, should be applicable to investigations of tRNA in all organisms.

High-throughput RNA sequencing (RNA-seq) has revolutionized our understanding of gene expression. Widely used RNA-seq methods start with adaptor ligation and cDNA synthesis of biological RNA samples followed by PCR amplification to generate sequencing libraries1. These standard methods work well for most cellular RNAs such as mRNA, long noncoding RNA, microRNA, or fragments derived from tRNA, small nuclear RNA and small nucleolar RNA. tRNA is the only class of small cellular RNA for which the standard sequencing methods cannot yet be applied efficiently and quantitatively, although attempts have been made (for example, ref. 2). Significant obstacles for the sequencing of tRNA include the presence of numerous post-transcriptional modifications and its stable and extensive secondary structure, which interfere with cDNA synthesis and adaptor ligation. tRNAs are essential for cells, and their synthesis is under stringent cellular control. Recent evidence shows that tRNA expression and mutations are associated with various diseases such as neurological pathologies and cancer development3,4. The lack of efficient and quantitative tRNA sequencing methods has hindered biological studies of tRNA.

We applied two strategies to eliminate or substantially reduce the obstacles of tRNA modification and structure for efficient and quantitative tRNA sequencing (Fig. 1a). First, an enzyme mix -

---

1Department of Biochemistry & Molecular Biology, University of Chicago, Chicago, Illinois, USA. 2Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA. 3Department of Chemistry, University of Chicago, Chicago, Illinois, USA. 4Institute of Biophysical Dynamics, University of Chicago, Chicago, Illinois, USA. 5Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois, USA. 6Present address: School of Life Sciences, Peking University, Beijing, China. *These authors contributed equally to this work. *These authors jointly supervised this work. Correspondence should be addressed to T.P. (taopan@uchicago.edu) or A.M.L. (lambowitz@austin.utexas.edu).

RECEIVED 10 MARCH; ACCEPTED 1 JUNE; PUBLISHED ONLINE 27 JULY 2015; DOI:10.1038/NMETH.3478
tRNA database when using gel-purified tRNA or total RNA as template, respectively (Supplementary Fig. 2a). These variations in read numbers were derived from sample handling, as similar proportions of reads were obtained for the added internal tRNA standards (Supplementary Fig. 2b). For the biological replicate from HEK293T cells, both untreated and treated samples showed similarly high reproducibility with \( r^2 \) values of 0.985–0.991 (Supplementary Fig. 3). As already indicated by the cDNA bands (Fig. 1c), the read length and in particular the proportion of reads corresponding to full-length tRNA increased substantially after demethylase treatment, as shown by the position plots along individual tRNA genes (Fig. 2a,b). The highest fraction of reads in a LeuAAG tRNA in the untreated samples were RT stops due to the \( m^1G37 \) modification; after demethylase treatment, all stops at this position were removed and there was a corresponding increase of full-length tRNA reads (Fig. 2a).

**Figure 1** | Demethylase-thermostable group II intron RT tRNA sequencing (DM-tRNA-seq). (a) Schematic representation. (b) Demethylation efficiency for total tRNA as measured by triple quadrupole liquid chromatography–mass spectrometry. (c) RT reaction for both purified tRNA and total RNA as template with (+) or without (−) demethylase treatment. The blue line shows the gel region excised for library construction.

**Figure 2** | Sequencing results. (a) 5′-position sequencing plot of a LeuAAG tRNA containing an \( m^1G37 \) modification. (b) 5′-position sequencing plot of a GlnCTG tRNA containing an \( m^3G9 \) modification. For a and b, \( m^3A58 \) stops correspond to very short DNA fragments that are outside the range of our sequenced cDNA and hence cannot be visualized here. (c) Correlation of tRNA isoacceptor expression and gene copy number in HEK293T cells. (d) Comparison of array fluorescence signals and sequencing reads for the Arg-tRNA family. Error bars, s.d.; \( n = 4 \). (e) Mismatches and stops with and without demethylase treatment. Shown are modification positions of \( m^3A58 \) in ValCAC, \( m^3C32 \) in ThrAGT, \( m^1G37 \) in ProTGG and \( m^1G9 \) in GlnCTG. The asterisk indicates stops for ValCAC corresponding to very short reads that were not determined in this experiment. Error bars, s.d.; \( n = 4 \). (f) Sequence logo at the same modified positions as in e, centered at the modified residue. Untreated samples (−) are at the top; demethylase-treated samples (+) are at the bottom.
Similarly, after demethylase treatment, the substantial RT stop at m^3G9 of a GlnCTG tRNA was markedly reduced, which occurred together with a substantially increased generation of full-length reads (Fig. 2b). We still detected a strong stop at the m^2G26 (N^2,N^2-dimethylguanosine) residue in LeuAAG tRNA that remained unchanged upon demethylase treatment (Fig. 2a). This result indicates that our demethylase mixture was not effective in removing m^2G modifications, which are present in ~20% of tRNAs. Nevertheless, our results show that the demethylase treatment was very effective in producing longer reads; this property is crucial for the ability to adequately map the mammalian tRNAome at single-base resolution.

We performed additional analysis to further demonstrate the usefulness of our sequencing method. Plotting each tRNA isoacceptor against its gene copy number showed a poor correlation (Fig. 2c), which is consistent with the known tissue-specific tRNA expression in humans. We compared the read fraction of the tRNA isoacceptors to the fluorescence hybridization signals of the Arg-tRNA probes from tRNA microarrays that were obtained through hybridization without the need for cDNA synthesis. The sequencing and array results showed the same trend of isoacceptor abundance, thus validating the quantitative nature of tRNA abundance obtained independently through sequencing- and hybridization-based approaches (Fig. 2d). We also compared RT stops and misincorporations at known modification positions with and without demethylase treatment. In the case of m^1A58 in ValCAC and m^3C32 in ThrAGT, the demethylases completely removed these modifications as demonstrated by the reversion to cognate sequence at these positions (Fig. 2e,f). In the case of m^3G37 in ProTGG and m^3G9 in GlnCTG, the demethylase treatment removed a large majority of the modification, so the mismatch and stops were substantially reduced (Fig. 2e,f). Therefore, our DM-tRNA-seq (demethylase-thermostable group II intron RT tRNA sequencing) method can determine differences in the modification dynamics of m^1A, m^3G and m^3C at single-base resolution, as well as potentially infer positions of non-demethylated modifications.

We also examined the expression of unique tRNA genes from chromosome 6. Human chromosome 6 contains approximately one-third of all tRNA genes, and over 150 genes are clustered within a 2.7-Mbp region near the class I major histocompatibility genes. We found that the tRNA expression levels within the tRNA gene cluster were higher compared to those of tRNA genes outside of the cluster (Supplementary Fig. 4). The expression level of tRNA genes in the cluster was uneven, suggesting that the expression of tRNA genes was not coordinated throughout the entire cluster in HEK293T cells.

Our approach described here makes efficient and quantitative tRNA sequencing feasible and has an additional advantage of being able to study modifications in a high-throughput manner.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** NCBI Gene Expression Omnibus: sequencing data have been deposited under accession GSE66550.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

This work was supported by US National Institutes of Health (NIH) grant DP1GM105386 to T.P., NIH grants R01GM07949 and GM37951 to A.M.L., K01HG006699 to G.Z., NIH MCB Training Grant (T32GM007185) to W.C.C. and a Chicago Biomedical Consortium Postdoctoral Research Grant Award to G.Z.

We thank L. Zhang, W.J. Chen and I.A. Gagnon for technical assistance. C.H. is supported by Howard Hughes Medical Institute as an investigator.

**AUTHOR CONTRIBUTIONS**

G.Z., Y.O., W.C.C., Q.D. and T.P. designed and performed experiments and analyzed data. G.Z. and T.P. conceived the project. G.Z., C.Y. and C.H. designed the demethylase constructs. G.Z., Y.O., W.C.C., A.M.L. and T.P. wrote the paper.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online methods.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wang, Z., Gerstein, M. & Snyder, M. Nat. Rev. Genet. 10, 57–63 (2009).
2. Pang, Y.L., Abo, R., Levine, S.S. & Dedon, P.C. Proc. Natl. Acad. Sci. USA 111, 12025–12030 (2014).
3. Kirchner, S. & Ignatova, Z. Nat. Rev. Genet. 16, 98–112 (2015).
4. Abbott, J.A., Francklyn, C.S. & Robey-Bond, S.M. Front. Genet. 5, 158 (2014).
5. Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. & Sedgwick, B. Nature 419, 174–178 (2002).
6. Falnes, P.B., Johansen, R.F. & Seeberg, E. Nature 419, 178–182 (2002).
7. Zheng, G., Fu, Y. & He, C. Chem. Rev. 114, 4602–4620 (2014).
8. Katibah, G.E. et al. Proc. Natl. Acad. Sci. USA 111, 12025–12030 (2014).
9. Shen, P.S. et al. Science 347, 75–78 (2015).
10. Mohr, S. et al. RNA 19, 958–970 (2013).
11. Goedenbour, J.M. & Pan, T. Nucleic Acids Res. 34, 6137–6146 (2006).
12. Chan, P.P. & Lowe, T.M. Nucleic Acids Res. 37, D93–D97 (2009).
13. Dittmar, K.A., Goedenbour, J.M. & Pan, T. PloS Genet. 2, e221 (2006).
14. Gingold, H. et al. Cell 158, 1281–1292 (2014).
15. Pavon-Eternod, M. et al. Nucleic Acids Res. 37, 7268–7280 (2009).
16. Horton, R. et al. Nat. Rev. Genet. 5, 889–899 (2004).
Design of the D135S AlkB mutant. The substrates for the wild-type AlkB (wtAlkB), m1A and m3C are positively charged. A close examination of the structures of substrate-bound AlkB revealed that these positively charged substrates are favorably positioned in the active site by interacting with the negatively charged carboxylic group of Asp135 (D135) (Supplementary Fig. 1a). We reasoned that single mutants at the D135 position might allow better accommodation of m1G in the active site (Supplementary Fig. 1a), thereby enhancing the demethylation efficiency of m1G. Truncated but catalytically active AlkB-ΔN11 was used, and mutants with single-amino-acid replacement at D135 were screened. Of several mutations tested, AlkB D135S gave the best demethylation yields toward m1G (Supplementary Fig. 1b). In the AlkB-m1G structure model constructed by computationally mutating the AlkB structure (Supplementary Fig. 1a), the shorter side chain of S135 seems to allow more room to accommodate m1G while sustaining the crucial hydrogen bond with m1G, which may explain the improvement in activity.

We also performed pH-activity profiling for the demethylation of m1A and m1G in tRNA. Decreasing demethylation activity was observed with increasing pH for both m1A and m1G substrates (Supplementary Fig. 1c). Therefore, we chose pH 5.0 for detailed kinetic analysis and subsequent experiments.

Cloning, expression and purification of wild-type and mutant AlkB. A truncated AlkB with deletion of the amino (N)-terminal 11 amino acids was cloned into a pET30a vector (Novagen) and overexpressed in E. coli BL21(DE3). The proteins were purified following procedures published previously. Briefly, cells were grown at 37 °C in the presence of 50 µM kanamycin until the OD600 reached 0.6–0.8. After the addition of IPTG (1 mM) and FeSO4 (5 µM), the cells were incubated for an additional 4 hours at 30 °C. The cells were collected, pelleted and then resuspended in lysis buffer (10 mM Tris, pH 7.4, 300 mM NaCl, 5% glycerol, 2 mM CaCl2, 10 mM MgCl2, 10 mM 2-mercaptoethanol). The cells were lysed by sonication and then centrifuged at 17,418g for 20 min. The soluble proteins were first purified using a Ni-NTA superflow cartridge (Qiagen) and then further purified by ion-exchange (Mono S GL, GE Healthcare) and gel-filtration (Superdex-200, Pharmacia) chromatography. All protein purification steps were performed at 4 °C. The Asp-to-Ser mutation was introduced using the QuikChange Site-Directed Mutagenesis Kit (Agilent). The mutant protein was expressed and purified following the same procedure as that for the wild-type protein.

Mammalian cell culture and RNA preparations. Human embryonic kidney HEK293T (CRL-11268) cells were obtained from the American Type Culture Collection (ATCC) and used without further validation. Cells were cultured in DMEM (Thermo) medium supplemented with 10% FBS and 1% 100× penicillin-streptomycin (Gibco). Cells were routinely checked for mycoplasma contamination every 3–6 months using the Universal Mycoplasma Detection Kit (ATCC).

Total RNA was isolated by using a mirVana miRNA Isolation Kit (Life Technologies). Purified total RNA was premixed with the T7 RNA polymerase transcripts of three tRNA standards (0.01 pmol each standard per µg of total RNA) and deacylated by incubating in 0.1 M Tris–HCl, pH 9, at 37 °C for 45 min. Although not necessary for studies of mature tRNAs, which all end with 3′ CCA, deacylated RNAs with or without demethylation treatment could be treated with T4 polynucleotide kinase (Epicentre) at 37 °C for 30 min to further warrant a free 3′ hydroxyl group for template-switching.

When necessary, total tRNA was subsequently isolated using a denaturing 10% polyacrylamide gel followed by passive gel elution and ethanol precipitation.

Thermostable group II intron RT template-switching. Template-switching reactions were performed as described. Briefly, we used an initial template-primer substrate consisting of a 41-nt RNA oligonucleotide (5′-AGU UCG GAA GAG CAC ACG UCU AGU UCAG CGA GCA CGA UC/3SpC3/-3′) that contains Illumina Read1 and Read2 primer-binding sites and a 3′ blocking group (three-carbon spacer; Integrated DNA Technologies, Inc.) annealed to a complementary 32P-labeled DNA primer with a single-nucleotide 3′ overhang, T, which facilitates the template switch to full-length tRNAs that mostly contain a 3′ CCA end. For TGIRT template-switching reactions, typically 100 ng of demethylated tRNAs or 1 µg of demethylated total RNA were mixed with the initial template-primer substrate (100 nM) and 500 nM TGIRT (Gst-IIC MalE rigid fusion RT) in reaction medium containing 450 mM NaCl, 5 mM MgCl2, 20 mM Tris–HCl, pH 7.5, and 5 mM DTT. The reactions were preincubated at room temperature for 30 min, initiated by adding 25 mM dNTPs (an equimolar mix of 25 mM dATP, dCTP, dGTP and dTTP) to a final concentration of 1 mM and incubating at 60 °C for 30 min. The reactions were terminated by adding 5 M NaOH to a final concentration of 0.25 M, incubating at 95 °C for 3 min and neutralizing with 5 M HCl. The cDNAs resulting from template-switching were analyzed by denaturing 6% polyacrylamide gel, electroeluted using a D-tube Dialyzer Maxi with MWCO of 6–8 kDa (EMD Millipore) and ethanol precipitated with 0.3 M sodium acetate, pH 5.2, in the presence of 25 µg of linear acrylamide (Life Technologies) carrier. The purified cDNAs were then circularized with CircLigase II (Epicentre) using the manufacturer’s protocol with an extended incubation time of 5 h at 60 °C, extracted with phenol–chloroform–isoamyl alcohol (25:24:1), ethanol precipitated and amplified with Phusion–HF (Thermo Scientific) using Illumina multiplex (5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC)
GAC GAT C-3′) and barcode (5′-CAA GCA GAA GAC GGC ATA CGA GAT BARCODE GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′) primers for 12 cycles of 98 °C for 5 s, 60 °C for 10 s and 72 °C for 10 s. The PCR products were sequenced on an Illumina HiSeq system.

Quantitative analysis of modification levels using LC-MS/MS. Quantitative analysis of modified nucleotides was done as previously described20. Briefly, 100 ng of tRNA were digested by nuclease P1 (2 U) in 30 µl of buffer containing 25 mM NaCl and 2.5 mM ZnCl2 at 37 °C for 1 h, followed by the addition of NH4HCO3 (100 mM) and alkaline phosphatase (0.5 U). After an additional incubation at 37 °C for 1 h, the solution was diluted to 60 µl, and 10 µl of the solution were injected for LC-MS/MS. Nucleosides were separated by reversed-phase ultraperformance liquid chromatography on a C18 column with on-line mass spectrometry detection using an Agilent 6410 QQQ triple quadrupole LC mass spectrometer in positive electrospray ionization mode. The nucleosides were quantified using the nucleoside-to-base ion mass transitions of 282 to 150 (m1A), 268 to 166 (A), 298 to (m1G), 284 to 152 (G), 258 to 126 (m3C) and 244 to 112 (C). Quantification was performed by comparison with the standard curve obtained from pure nucleoside standards running in the same batch of samples. Modification levels were compared by the ratios of methylated base (m1A, m1G, m3C) over regular base (A, G, C).

tRNA microarrays. The tRNA microarray assay consists of four steps starting from purified tRNA or total RNA without the need of cDNA synthesis: (i) deacylation to remove all 3′-attached amino acids, (ii) selective fluorophore labeling of tRNA using oligonucleotide ligation with T4 DNA ligase to the 3′ CCA of all tRNAs, (iii) hybridization and (iv) data analysis. The reproducibility of the tRNA microarray method and validation of the results have been extensively described previously13,15.

Sequencing read mapping. Sequencing reads were aligned using Bowtie to a modified hg19 genomic tRNA database12. A single mismatch was allowed in order to identify potential modification misincorporations at a modification site. Briefly, a tRNA library was adapted from the tRNAscan-SE library by appending CCA to tRNAs from the genomic tRNA database (http://gtrnadb.ucsc.edu/Hsapi19/). IsoDecoders with identical scores were consolidated for ease of identity assignment, decreasing the number of reference genes and pseudogenes from 625 to 462. Prior to mapping, reads were processed using Trimmmomatic v0.32. Sequences greater than 15 bp were then aligned to the aforementioned culled tRNA library using Bowtie2 with sensitive options. Reads mapping to multiple isodecoders owing to length of fragment were discarded. Modification fractions were determined by analyzing at a putative position n the number of correct reads, the number of misincorporations, as well as the number of reads stopped at the n + 1 position.

17. Holland, P.J. & Hollis, T. PloS ONE 5, e8680 (2010).
18. Yu, B. et al. Nature 439, 879–884 (2006).
19. Mishina, Y., Chen, L.X. & He, C. J. Am. Chem. Soc. 126, 16930–16936 (2004).
20. Zheng, G. et al. Mol. Cell 49, 18–29 (2013).