

PALM-Seq: integrated sequencing of cell-free long RNA and small RNA

Xi Yang^{1,3,6}, Taifu Wang^{1,3,4,6}, Sujun Zhu^{2,6}, Juan Zeng², Yanru Xing^{1,3,4}, Qing Zhou^{1,3}, Zhongzhen Liu^{1,3}, Haixiao Chen^{1,3}, Jinghua Sun^{1,3,4}, Liqiang Li^{1,3}, Jinjin Xu^{1,3}, Chunyu Geng⁵, Xun Xu^{1,3}, Jian Wang^{1,3}, Huanming Yang^{1,3}, Shida Zhu^{1,3*}, Fang Chen^{1,3*}, Wen-Jing Wang^{1,3*}

1. BGI-Shenzhen, Shenzhen 518083, China
 2. Obstetrics Department, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, Guangdong Province, China
 3. China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China
 4. BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China
 5. MGI, BGI-Shenzhen, Shenzhen 518083, China.
 6. These authors contributed equally: Xi Yang, Taifu Wang, Sujun Zhu.
- *e-mail: zhushida@genomics.cn; fangchen@genomics.cn; wangwenjing@genomics.cn.

Abstract

Cell-free RNA, including both long RNA and small RNA, has been considered important for its biological functions and potential clinical usage, but the major challenge is to effectively sequence them at the same time. Here we present PolyAdenylation Ligation Mediated-Seq (PALM-Seq), an integrated sequencing method for cell-free long and small RNA. Through terminal modification and addition of 3' polyadenylation and 5' adaptor, we could get mRNA, long non-coding RNA, microRNA, tRNA, piRNA and other RNAs in a single library. With target RNA depletion, all these RNAs could be sequenced with relatively low depth. Using PALM-Seq, we identified pregnant-related mRNAs, long non-coding RNAs and microRNAs in female plasma. We also applied PALM-Seq to sequence RNA from amniotic fluids, leukocytes and placentas, and could find RNA signatures associated with specific sample type. PALM-Seq provides an integrated, cost-effective and simple method to characterize the landscape of cell-free RNA, and can stimulate further progress in cell-free RNA study and usage.

Introduction

After cell-free RNA (cfRNA) was discovered, its importance has been proven. CfRNA plays important roles in cell communications^{1,2}, and its biogenesis, uptake³ and distribution⁴ are correlated with physical and pathological situations⁵. Besides this, cfRNA carries information from human tissues^{6,7} or tumor^{8,9}. According to these reasons, cfRNA is considered as potential targets for disease interventions and useful biomarkers for disorder prediction.

CfRNA in plasma is usually made up of degradative small fragments with size smaller than 200nt, very low concentration (may lower than 10 ng/mL)¹⁰, and different terminal modification^{11,12}, and these properties make it difficult to research. Since 2014, next-generation sequencing was used for cfRNA studies^{13,14}, but some problems were not solved till now, including requirement of large volume for blood¹⁵, no uniform sequencing method for all cfRNA fractions¹², high cost for large scale cfRNA library preparation, and as well as low mapping rate.

Here, we presented a new method, PolyAdenylation Ligation Mediated-Seq (PALM-Seq), to get all sorts of RNA in one library with lower cfRNA input. Additionally, PALM-Seq provided a new possible way for entire transcriptome sequencing with low input of RNA samples. This approach could enable us to get more comprehensive results for large population studies and clinical usages of cfRNA.

Results

Principle of PALM-Seq.

For amount of cfRNA was low, in order to avoid RNA degradation or adsorption to the tube, 3' polyadenylation was added by E. coli Poly(A) Polymerase (PAP)¹⁶ in the first step. The terminal modification of cfRNA was different, however T4 Polynucleotide Kinase (T4 PNK) could make most part of them to 5' end phosphorylated and 3' end hydroxyl^{17,18} except those with 5' cap. Because T4 PNK and PAP could work in the same buffer with ATP, separate step of pretreatment for cfRNA with T4 PNK was unnecessary. Then, 5' adaptor was then ligated by T4 RNA Ligase 1¹⁹. In contrast to 3' DNA adaptor used in common small RNA seq, 3' polyadenylation could not be degraded by DNase I which was necessary for RNase H method²⁰. So, rRNA or other uninterested, abundant RNA could be easily removed by RNase H method. DNase I treatment could also prevent the possible contamination of cell-free DNA. Finally, RNA with 3' polyadenylation and 5' adaptor was reverse transcribed by oligo(dT) with 3' adaptor and amplified by PCR (Fig. 1a). Most part of the library was short fragments as expected. The processing pipeline of sequencing data was shown in Fig. 1b.

In order to test the effect of T4 PNK treatment (PNKT) and target depletion (TD) which was performed through RNase H method, and identify the best work condition for cfRNA, we used four samples, each of which was treated in four work conditions, with or without PNKT and with or without TD. We also wanted to know whether TD

could remove other abundant RNA, so Y RNA and Vault RNA which were rich in plasma cfRNA²¹ and not interested in this study, were also depleted for plasma samples.

PALM-Seq could capture plasma mRNA, lncRNA and small RNA with high complexity.

Complexity is especially important for low-quantity and low-quality RNA sequencing^{20,22}, and it is better to get more kinds of RNA in one library. To assess the performance of different treatments and methods, we analyzed the data to account for differences in number of different genes detected and contribution of different kinds of RNA. PALM-Seq with PNKT (TD or no TD) could detect larger number of mRNA (Fig. 2a) or lncRNA (Fig. 2b), and it could also detect miRNA (Fig. 2c), tRNA (Fig. 2d) and piRNA (Fig. 2e), while the fewer miRNAs were covered. PALM-Seq with no PNKT and no TD were more efficient to detect miRNA (Fig. 2c), tRNA (Fig. 2d) and piRNA (Fig. 2e), but it covered less mRNA (Fig. 2a) and lncRNA (Fig. 2b).

To further qualify our method, we referred to public data produced by SMARTer Seq (by using Clontech SMARTer Stranded Total RNA-Seq Kit - Pico)¹⁵, ScriptSeq (by using Illumina ScriptSeq v2 Kit)²³ and Small RNA Seq (by using NEBNext Small RNA Library Preparation Kit)²⁴ to quantify if PALM-seq was comparable or even better. While SMARTer Seq and ScriptSeq were designed for large fragment RNA, they were effective to detect mRNA or lncRNA (Fig. 2a-b), but they could not cover small RNA (Fig. 2c-e). In contrast, Small RNA Seq performed well in detecting

miRNA (Fig. 2c), but was not suitable for lncRNA or mRNA (Fig. 2a-b). RNA ratio also showed that mRNA and lncRNA were rich in PALM-Seq with PNKT (TD or no TD), SMARTer Seq or ScriptSeq, however miRNA was rich in PALM-Seq with no PNKT (TD or no TD) or Small RNA Seq (Fig. 2f). In addition, tRNA and piRNA ratio was higher in PALM-Seq (Fig. 2f).

Length contribution of different sort of RNA were also checked to further estimate the effect of PNKT. Large fragments of mRNA or lncRNA could be captured without PNKT, however, more small fragments of mRNA or lncRNA would be captured with PNKT, for small fragments might be produced by RNase (e.g. RNase A) with 3' phosphorylated and 5' hydroxyl. Length contribution of small RNA were less influenced by PNKT. The information provided by small fragments of mRNA or lncRNA would be important especially for small amount or long-stored samples without enough long fragments, so PNKT was acquired for mRNA or lncRNA analysis.

These results illustrated that PALM-Seq could be used for miRNA, tRNA and piRNA whether with or without PNKT, but PNKT was necessary for mRNA and lncRNA analysis.

TD could remove uninterested RNA efficiently and improve mapping rate.

Mapping rate is important for data analysis to detect more genes, splicing or other variations with lower sequencing depth²⁰. Depletion of uninformative abundant RNA (e.g. rRNA) can increase mapping rate and maximize the coverage of other transcripts

present in a sample. To test the efficiency and necessity of TD, we calculated rRNA ratio (Fig. 3a). We found that rRNA could be removed efficiently by TD, especially in PNKT condition, otherwise, the rRNA ratio would be extremely high. In no PNKT condition, TD was dispensable. In addition, Y RNA and Vault RNA could also be removed by TD. SMARTer Seq removed rRNA through CRISPR, so the rRNA ratio was low, but ScriptSeq or Small RNA Seq data contained high ratio of rRNA.

To further estimate informative data ratio, we calculated mRNA, lncRNA and miRNA mapping rate to total reads (Fig. 3b-d). We found that TD increased mRNA and lncRNA mapping ratio (Fig. 3b-c) and decreased the acquired sequencing depth²⁵ when PNKT was used (Fig. 3e-f). However, TD did not significantly influence miRNA mapping ratio.

Then we compared mapping rate among different methods. For mRNA and lncRNA mapping rate, PALM-Seq with PNKT and TD performed similarly with SMARTer Seq, and much better than Small RNA Seq. ScriptSeq performed best for mRNA but not well for lncRNA. For miRNA mapping rate, PALM-Seq with PNKT and TD performed similarly with Small RNA Seq, although lower than PALM-Seq with no PNKT and no TD.

These results suggested that TD was helpful, and it was necessary when PALM-Seq was used to detect both mRNA and miRNA, for rRNA ratio might be unacceptable if only PNKT was used.

We recommended PALM-Seq with PNKT and TD for the studies which needed both mRNA/lncRNA and miRNA, and PALM-seq with no PNKT and no TD for the

researches focused on miRNA.

Because both mRNA/lncRNA and miRNA were needed in our further work, PALM-Seq with PNKT and TD were used for other samples mentioned in this study.

PALM-Seq was a robust method for plasma RNA-Seq.

An ideal cfRNA sequencing method should not only display high yield, complexity and mapping rate, but also have high reproducibility²². We tested the reproducibility of PALM-Seq and found that PALM-Seq with PNKT and TD showed the similar consistency for mRNA as SMARTer Seq or ScriptSeq (Fig. 4a), but the consistency of lncRNA was higher for PALM-Seq (Fig. 4b). PALM-Seq with PNKT and TD or PALM-Seq with no PNKT and no TD gave more stable results compared with Small RNA Seq (Fig. 4c).

Coverage metrics might be critical for certain measurements. We calculated the variation in coverage of mRNA and lncRNA from 5' to 3' ²⁰, and PALM-Seq did not show significant bias from 5' to 3' by considering all detected genes. We then aligned the small RNAs to their respective precursors²⁶. As expected, the mature miRNAs were aligned to the two arms of the pre-miRNA hairpins (Fig. 4f). The tRNA-derived small RNAs (tsRNAs) were mainly derived from the two regions of the mature tRNA, but it showed slight difference with cell tsRNAs²⁶ (Fig. 4g).

To estimate the failure rate of this method, we totally built 120 libraries for plasma cfRNA, and 2 of them failed for low yield (<300ng).

For further usage of PALM-Seq, we assessed the influence of sequencing depth²⁵

and read length. First, we evaluated saturation plots for different transcript biotypes. It was shown that 20M clean reads (about 40M total reads) were enough for mRNA, lncRNA, miRNA and tRNA quantification. Due to the low piRNA ratio, more reads might be acquired for piRNA analysis. We then compared expression results with single read 35 cycles or 100 cycles, and there was no significant difference, because most of cfRNA fragments were short.

Placental-specific genes could be detected in plasma cfRNA from pregnant women.

One of the basic usages for RNA-Seq is differentially expressed gene (DEG) analysis²⁷, so PALM-Seq was tested for DEGs. During pregnancy, placenta releases nucleic acids into blood. To test whether PALM-Seq could capture placental specific RNA, we collected blood samples from 4 nonpregnant females and 4 pregnant females (three time points: 12, 24 and 36 gestational week).

Principal components analysis²⁸ (PCA) showed that there were no significant outliers. Then, we checked the expression of placenta-specific mRNA/lncRNA (Fig. 5a) or miRNA (Fig. 5b). As expected, these genes were only expressed in plasma from pregnant women. We verified 5 mRNA/lncRNA genes and 5 miRNAs. Protein coding gene such as *CGA*, *CSH1* and *CSH2*, and long non-coding gene such as *PLAC4*¹⁵, were very low or undetected in nonpregnant females, and increased during pregnancy. *S100A8* which are related to pregnant immunomodulation¹⁵, also increased (Fig. 5c). Placental-specific C17MC miRNA²⁹ such as mir-526b-5p,

mir-519d-3p, mir-515-5p and mir-512-3p, also existed in plasma of pregnant females; and we found mir-454-3p decreased (Fig. 5d). These results were validated by RT-qPCR (Fig. 5c-d). Together with the results above, it was suggested that PALM-Seq could provide reliable results for DEG analysis.

Discussion

Although the kits designed for total RNA such as Ribo-Zero Gold Kit NuGen's¹³ or RNA-Seq Ovation System Kit¹⁴, were used before, they are not proper for cell-free RNA. Now, Small RNA Seq, SMARTer Seq and ScriptSeq are considered as optimized choices^{15,22-24}. However, Small RNA Seq method requires 5' phosphorylated and 3' hydroxyl³⁶, and pretreatment for low input of samples with Tobacco Acid Pyrophosphatase (TAP) or RNA 5' Pyrophosphohydrolase (RppH) and T4 PNK may cause significant loss. SMARTer Seq and ScriptSeq could only capture large fragments, but small fragments are the main fraction of cfRNA, and they also may not be used for long-stored samples. There is no suitable method to capture small RNA and long RNA at the same time. Additionally, these kits are also expensive for large scale usage.

In contrast, PALM-Seq is a flexible, low cost method which could cover most of RNA in plasma except mRNA fragment with 5' Cap. Compared to commercial RNA-Seq kit, many details can be easily modified to capture certain fractions without additional purification. Different terminal modifications might be distinguished by the usage of T4 PNK, TAP or RppH; however, if TAP or RppH is used, buffer should be

optimized. Different target depletion could also be achieved just by change DNA oligo set. If certain length of fragments was required, size selection could be used after PCR amplification.

RNase H based target depletion²⁰ is recruited here because it does not need certain kit, and there is no sequence restriction as in CRISPR-based method (PAM sequence NGG is necessary for Cas9) used by SMARTer Seq. RNase H based target depletion method was not limited by fragment length as well (DSN-lite might have problems when the fragments were too short). Removing rRNA by magnetic bead capture was also not practical for cfRNA due to severely lost. CATS-Seq¹¹ (which shared the same principle of Clontech SMARTer smRNA-Seq Kit) which is similar to PALM-Seq, is not compatible for RNase H method, because the digestion production could not be removed and would be conjunct with 5' adaptor through template-switch.

PALM-Seq is also time-saving and easy to handle. One could finish library preparation for 24 samples in 10 hours or for 32 samples in 12 hours without multichannel pipettors or purification workstations, and it might also be used in automatic production line. In a word PALM-Seq is a powerful sequencing method suitable for large scale cfRNA study or clinical application.

Acknowledgements

We are grateful to volunteers who donated their samples to this study. We appreciate colleagues at BGI-Shenzhen for sequencing. This project is supported by the National

Natural Science Foundation of China (No.81300075), the Natural Science Foundation of Guangdong Province (No. 2014A030313795), the Shenzhen Municipal Government of China (No.JCYJ20170412152854656, JCYJ20180703093402288), the Shenzhen Peacock Plan (No. KQTD20150330171505310).

Author Contributions

W.J.W., F.C., and S.Z. conceived and designed the project. X.Y., S.Z., J.Z., Y.X., Z.L., H.C., J.S., L.L., C.G., collected samples and extracted plasma. X.Y. and Z.L. constructed PALM-Seq libraries and carried out other experiments. T.W., Y.X., Q.Z., and J.X. wrote the PALM-Seq analysis pipeline and analyzed PALM-Seq data. W.J.W., X.X., J.W., and H.Y. analyzed and interpreted all data. X.Y. and T.W. wrote the manuscript. W.J.W., F.C., and S.Z. revised the manuscript. All authors read and approved the manuscript for submission.

Competing interests

The authors declare no competing interests.

Ethical compliance

We complied with all relevant ethical regulations.

Code availability

The custom codes used to analyze PALM-Seq data are available from

<https://github.com/wonderful1/PALM-Seq-cfRNA>

Data availability

All the sequencing datasets generated in this study, including PALM-Seq test (with/without PNKT, with/ without TD), PALM-Seq for plasma from pregnant or nonpregnant women and PALM-Seq for amniotic fluid cfRNA or RNA from leukocytes and placentas have been deposited in CNSA (<https://db.cngb.org/cnsa/>) with accession CNP0000506.

References

- 1 Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* **9**, 654-659, doi:10.1038/ncb1596 (2007).
- 2 Turchinovich, A., Samatov, T. R., Tonevitsky, A. G. & Burwinkel, B. Circulating miRNAs: cell-cell communication function? *Front Genet* **4**, 119, doi:10.3389/fgene.2013.00119 (2013).
- 3 Mathieu, M., Martin-Jaular, L., Lavieu, G. & Thery, C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nat Cell Biol* **21**, 9-17, doi:10.1038/s41556-018-0250-9 (2019).
- 4 Ragusa, M. *et al.* Asymmetric RNA Distribution among Cells and Their Secreted Exosomes: Biomedical Meaning and Considerations on Diagnostic Applications. *Front Mol Biosci* **4**, 66, doi:10.3389/fmolb.2017.00066 (2017).
- 5 Yuan, T. *et al.* Plasma extracellular RNA profiles in healthy and cancer patients. *Sci Rep* **6**, 19413, doi:10.1038/srep19413 (2016).
- 6 Miyamoto, M. *et al.* Detection of cell-free, liver-specific mRNAs in peripheral blood from rats with hepatotoxicity: a potential toxicological biomarker for safety evaluation. *Toxicol Sci* **106**, 538-545, doi:10.1093/toxsci/kfn188 (2008).
- 7 Tsang, J. C. H. *et al.* Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc Natl Acad Sci U S A* **114**, E7786-E7795, doi:10.1073/pnas.1710470114 (2017).
- 8 Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* **11**, 426-437, doi:10.1038/nrc3066 (2011).
- 9 Larrea, E. *et al.* New Concepts in Cancer Biomarkers: Circulating miRNAs in Liquid Biopsies. *Int J Mol Sci* **17**, doi:10.3390/ijms17050627 (2016).
- 10 Spornraft, M. *et al.* Optimization of extraction of circulating RNAs from plasma--enabling small RNA sequencing. *PLoS One* **9**, e107259, doi:10.1371/journal.pone.0107259 (2014).
- 11 Turchinovich, A. *et al.* Capture and Amplification by Tailing and Switching (CATS). An ultrasensitive ligation-independent method for generation of DNA libraries for deep sequencing from picogram amounts of DNA and RNA. *RNA Biol* **11**, 817-828, doi:10.4161/ma.29304 (2014).
- 12 Wei, Z. *et al.* Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat Commun* **8**, 1145, doi:10.1038/s41467-017-01196-x (2017).
- 13 Tsui, N. B. *et al.* Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clin Chem* **60**, 954-962, doi:10.1373/clinchem.2014.221648 (2014).
- 14 Koh, W. *et al.* Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc Natl Acad Sci U S A* **111**, 7361-7366, doi:10.1073/pnas.1405528111 (2014).

- 15 Ngo, T. T. M. *et al.* Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science* **360**, 1133-1136, doi:10.1126/science.aar3819 (2018).
- 16 Yehudai-Resheff, S. & Schuster, G. Characterization of the E.coli poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Res* **28**, 1139-1144, doi:10.1093/nar/28.5.1139 (2000).
- 17 Das, U. & Shuman, S. Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res* **41**, 355-365, doi:10.1093/nar/gks977 (2013).
- 18 Wang, L. K., Lima, C. D. & Shuman, S. Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *EMBO J* **21**, 3873-3880, doi:10.1093/emboj/cdf397 (2002).
- 19 Middleton, T., Herlihy, W. C., Schimmel, P. R. & Munro, H. N. Synthesis and purification of oligoribonucleotides using T4 RNA ligase and reverse-phase chromatography. *Anal Biochem* **144**, 110-117 (1985).
- 20 Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* **10**, 623-629, doi:10.1038/nmeth.2483 (2013).
- 21 Yeri, A. *et al.* Total Extracellular Small RNA Profiles from Plasma, Saliva, and Urine of Healthy Subjects. *Sci Rep* **7**, 44061, doi:10.1038/srep44061 (2017).
- 22 Srinivasan, S. *et al.* Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. *Cell* **177**, 446-462 e416, doi:10.1016/j.cell.2019.03.024 (2019).
- 23 Pan, W. *et al.* Simultaneously Monitoring Immune Response and Microbial Infections during Pregnancy through Plasma cfRNA Sequencing. *Clin Chem* **63**, 1695-1704, doi:10.1373/clinchem.2017.273888 (2017).
- 24 Giraldez, M. D. *et al.* Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat Biotechnol* **36**, 746-757, doi:10.1038/nbt.4183 (2018).
- 25 Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: a matter of depth. *Genome Res* **21**, 2213-2223, doi:10.1101/gr.124321.111 (2011).
- 26 Faridani, O. R. *et al.* Single-cell sequencing of the small-RNA transcriptome. *Nat Biotechnol* **34**, 1264-1266, doi:10.1038/nbt.3701 (2016).
- 27 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
- 28 Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**, 896-902, doi:10.1038/nbt.2931 (2014).
- 29 Bentwich, I. *et al.* Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**, 766-770, doi:10.1038/ng1590 (2005).
- 30 Godoy, P. M. *et al.* Large Differences in Small RNA Composition Between

- Human Biofluids. *Cell Rep* **25**, 1346-1358, doi:10.1016/j.celrep.2018.10.014 (2018).
- 31 Murillo, O. D. *et al.* exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**, 463-477 e415, doi:10.1016/j.cell.2019.02.018 (2019).
- 32 Juzenas, S. *et al.* A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res* **45**, 9290-9301, doi:10.1093/nar/gkx706 (2017).
- 33 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
- 34 Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* **10**, R130, doi:10.1186/gb-2009-10-11-r130 (2009).
- 35 Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062-6067, doi:10.1073/pnas.0400782101 (2004).
- 36 Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858-862, doi:10.1126/science.1065062 (2001).

Figure 1 | Scheme of PALM-Seq. **a**, Principle of PALM-Seq. Cell-free RNAs (cfRNA) are terminal modified and added with 3' poly(rA) tail and 5' adaptor. Targeted RNAs are depleted. Finally, RNAs are converted to cDNA by reverse transcription and PCR amplification. **b**, Data Analysis. Low quality data are removed at first, and then adapter trimming is made. Reads shorter than 17bp, rRNA and/or other targeted RNA (e.g. Y RNA and Vault RNA in plasma) reads are filtered. The mapping is performed in the order of miRNA, tRNA/piRNA, mRNA/lncRNA and other RNA. The unmapped reads will be mapped to genome if necessary. Small RNAs are quantified by counts and Reads Per Million (RPM), and mRNA or lncRNA are quantified by counts and Transcripts Per Million (TPM).

Figure 2 | Complexity and RNA biotype contribution of different RNA-seq methods for plasma cfRNA. **a-e**, Number of mRNAs (**a**), lncRNA (**b**), miRNA (**c**), tRNA (**d**) and piRNA (**e**) expressed in each treatment or method. Four samples are used for each subgroup, the results are shown as mean $\pm$ S.E.M. (Standard Error of Mean). **f**, Average ratio (n=4) of mapped reads for each type of RNA in different cfRNA sequencing methods. PNKT: T4 polynucleotide kinase treatment. TD: Target depletion.

Figure 3 | Mapping rate of different cfRNA sequencing methods. **a-d**, Performance of each library with respect to rRNA ratio (**a**, normalized to read number after low-quality reads and <17bp reads have been filtered), mRNA mapping rate (**b**, normalized to total read number), lncRNA mapping rate (**c**, normalized to total read number) and miRNA mapping rate (**d**, normalized to total read number). Unpaired two-sided Student's t-test is used to assess significance. The original p values are shown here. **e-f**, Saturation plots for mRNA (**e**) and lncRNA (**f**) shows the influence of TD to the acquired sequencing depth when PNKT is used.

Figure 4 | Reproducibility and transcript coverage of PALM-Seq. **a-c**, Reproducibility is measured by coefficient of determination (R^2). PALM-Seq with PNKT and TD is compared with SMARTer-Seq or ScriptSeq for mRNA (**a**) and lncRNA (**b**). PALM-Seq with PNKT and TD or with no PNKT and no TD are compared with Small RNA Seq for miRNA (**c**). R^2 is calculated through simple linear regression of $\log_2(\text{TPM}+1)$ for mRNA and lncRNA or $\log_2(\text{RPM}+1)$ for miRNA. **d-e**, Normalized coverage by position for mRNA (**d**) and lncRNA (**e**). For each library, the average relative coverage is shown at each relative position along the transcripts' length. **f-g**, Coverage of miRNA (**f**) and tRNA-derived small RNAs (tsRNA, **g**) across the length of the precursor RNAs (divided into 40 bins). Different curves stand for different samples using PALM-Seq with PNKT and TD.

Figure 5 | PALM-Seq could detect pregnancy-related plasma cfRNA. **a**, The expression of placenta-specific mRNA or lncRNA in plasma during pregnancy. **b**, The expression of placenta-specific miRNA (C17MC cluster) in plasma during

pregnancy. **c**, TPM of PALM-Seq and Δ CT of RT-qPCR normalized to *B2M* of *CGA*, *CSH1*, *CSH2*, *PLAC4* and *S100A8*. **d**, RPM of PALM-Seq and Δ CT of RT-qPCR normalized to RN7SL of mir-526b-5p, mir-519d-3p, mir-515-5p, mir-512-3p and mir-454-3p. N: nonpregnant females; T1: females at 12 gestational week; T2: females at 24 gestational week; T3: females at 36 gestational week. N=4.

Figure 6 | Usage of PALM-Seq in amniotic fluid cfRNA and RNA from leukocyte or placenta. a, Average ratio (n=4) for each type of RNA in different samples. **b-d**, Principal components analysis for mRNA (**b**), lncRNA (**c**) and miRNA (**d**). **e**, 100 differently expressed miRNAs between leukocyte and placenta by PALM-Seq. **f**, 200 differently expressed mRNAs or lncRNAs between leukocyte and placenta by PALM-Seq.

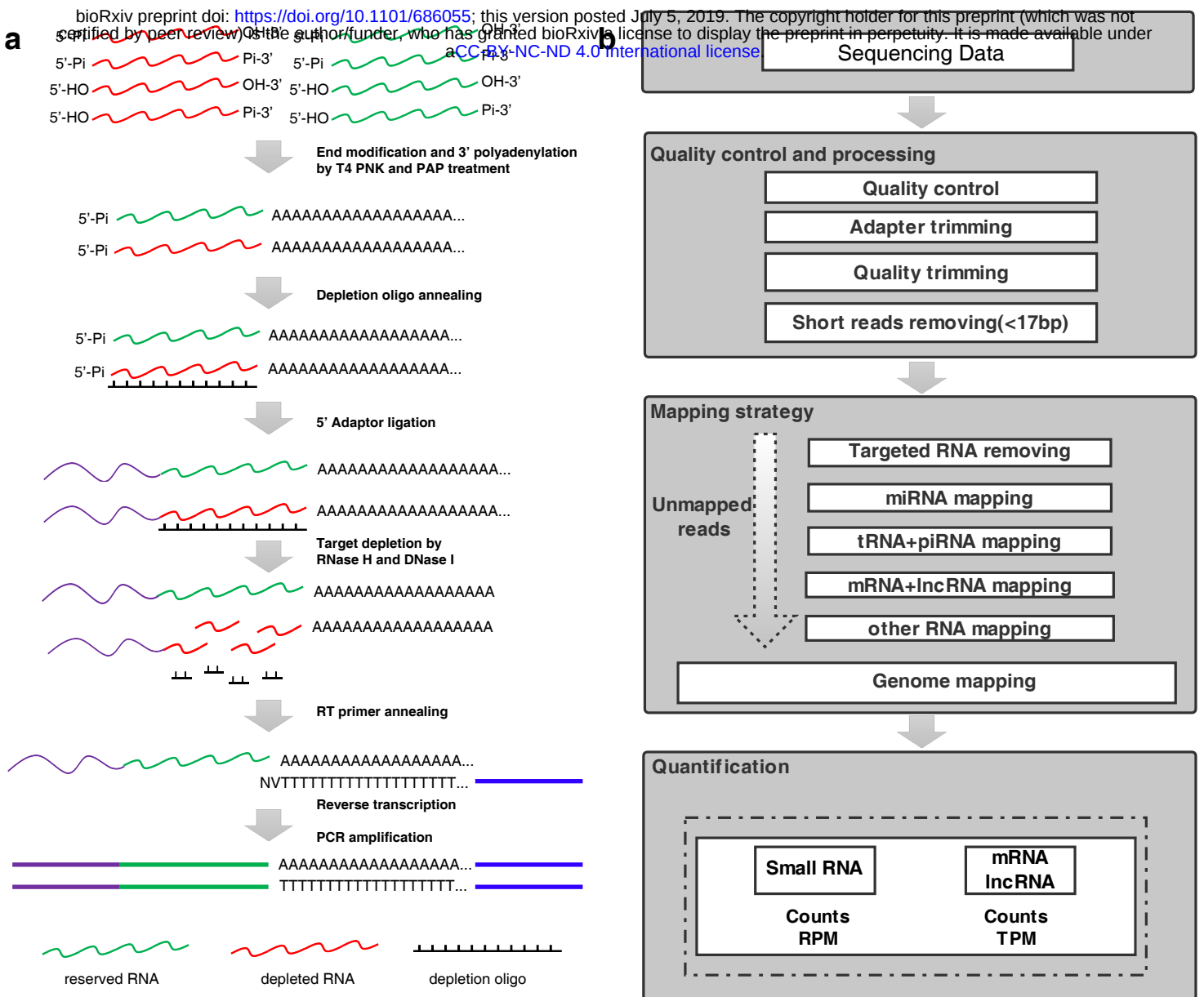


Figure 1 | Scheme of PALM-Seq.

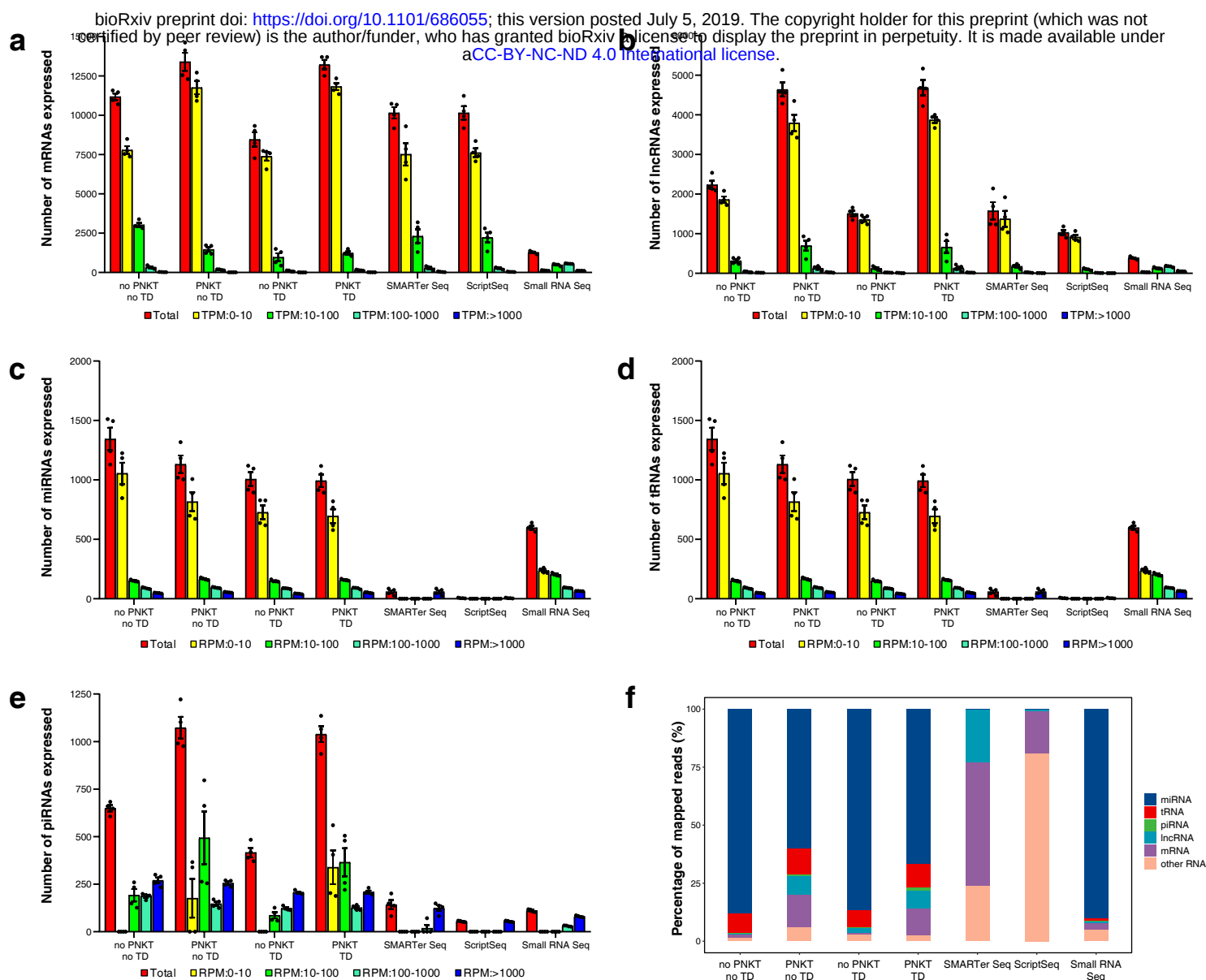


Figure 2 | Complexity and RNA biotype contribution of different RNA-seq methods for plasma cfRNA.

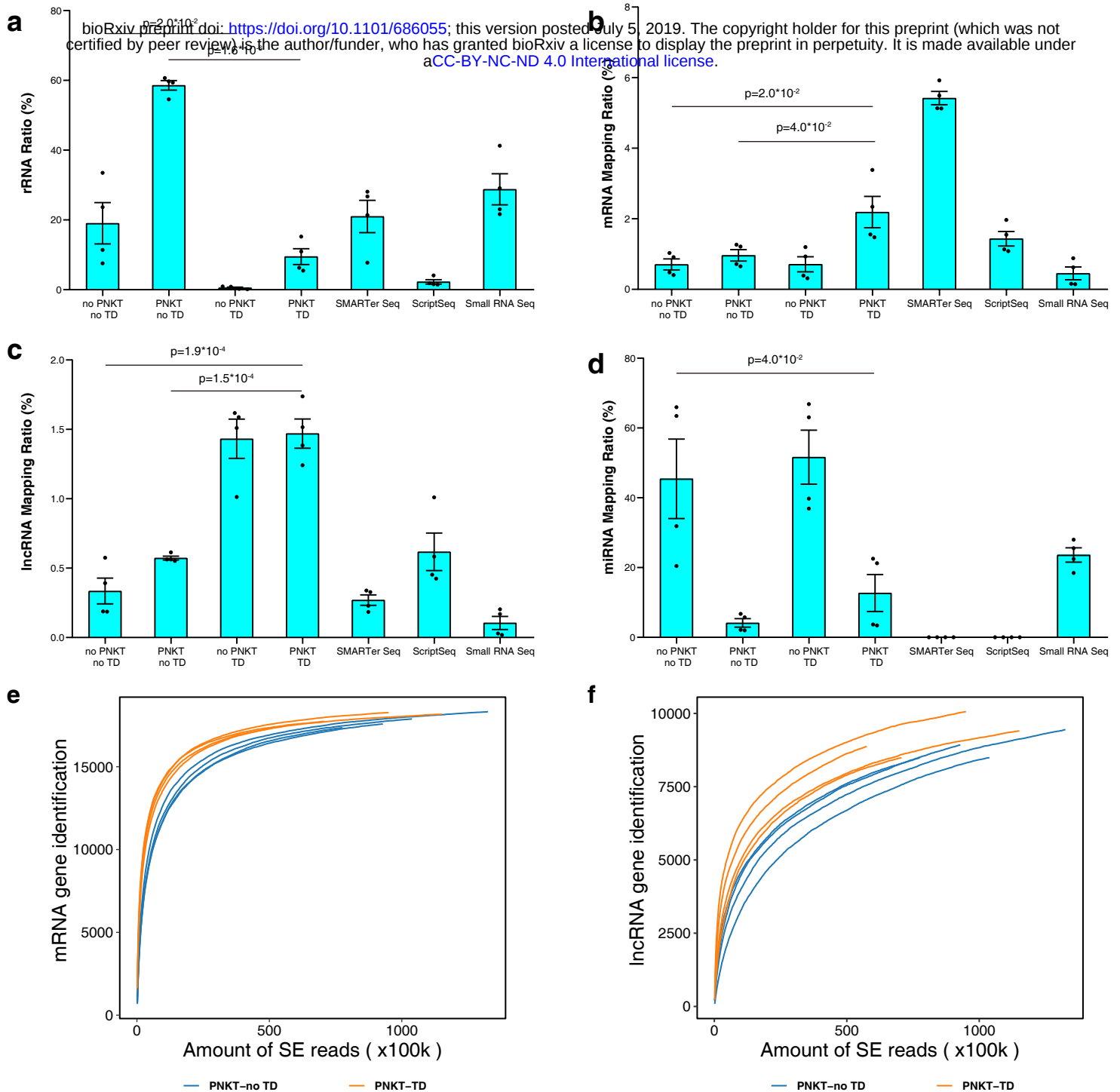


Figure 3 | Mapping rate of different cfRNA sequencing methods.

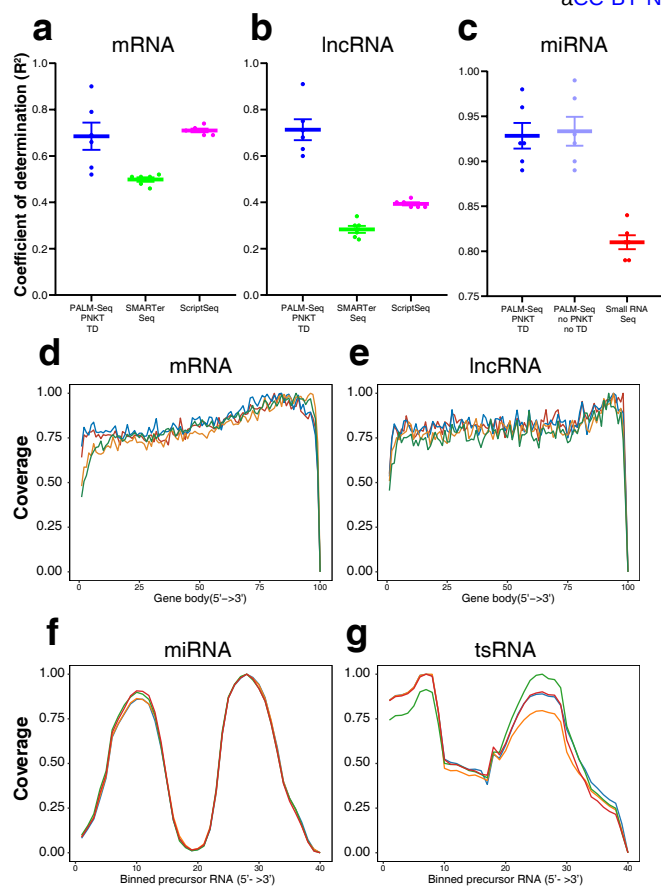


Figure 4 | Reproducibility and transcript coverage of PALM-Seq.

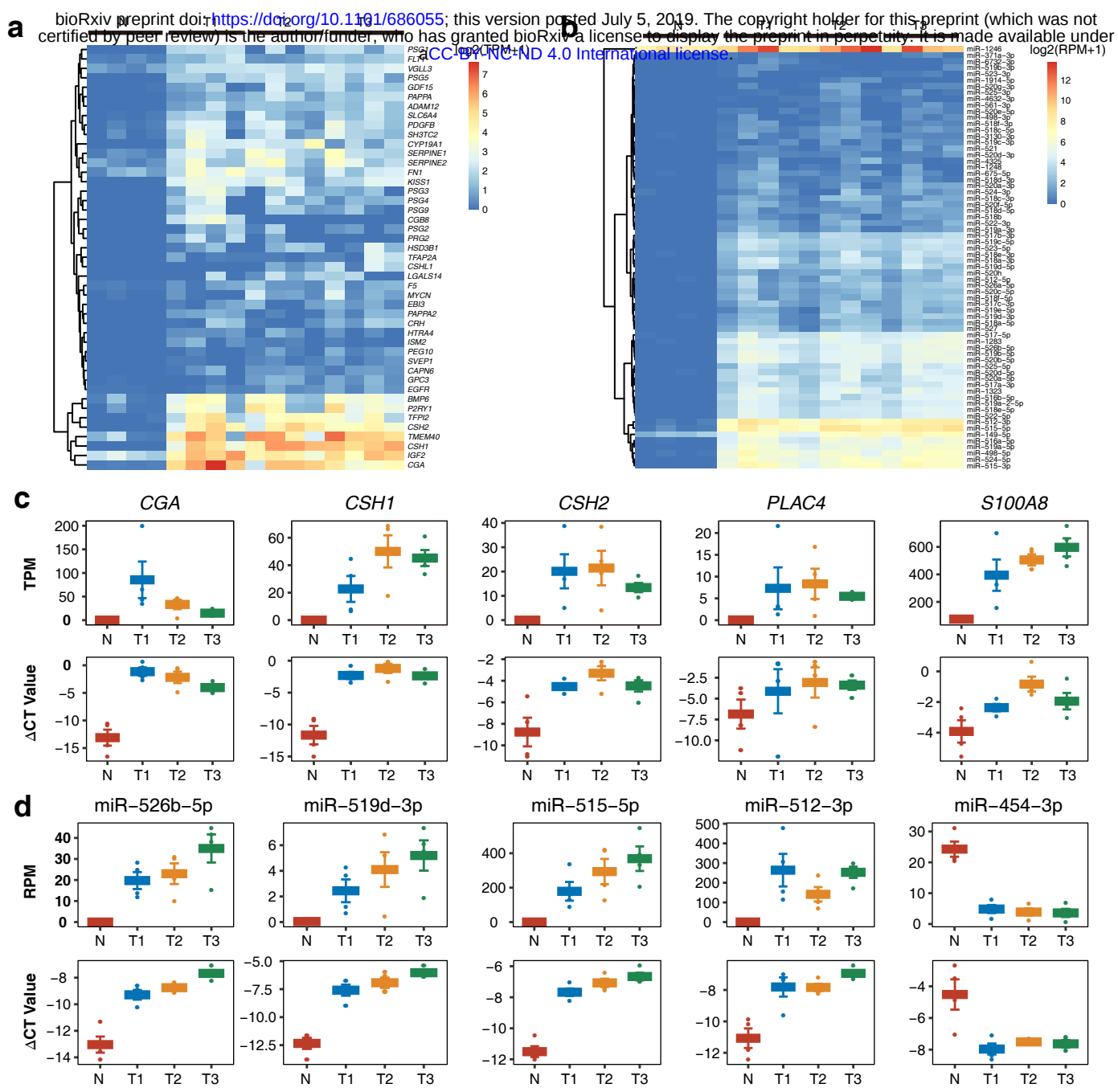


Figure 5 | PALM-Seq could detect pregnancy-related plasma cfRNA.