



Figures and figure supplements

Transposase-assisted tagmentation of RNA/DNA hybrid duplexes

Bo Lu et al

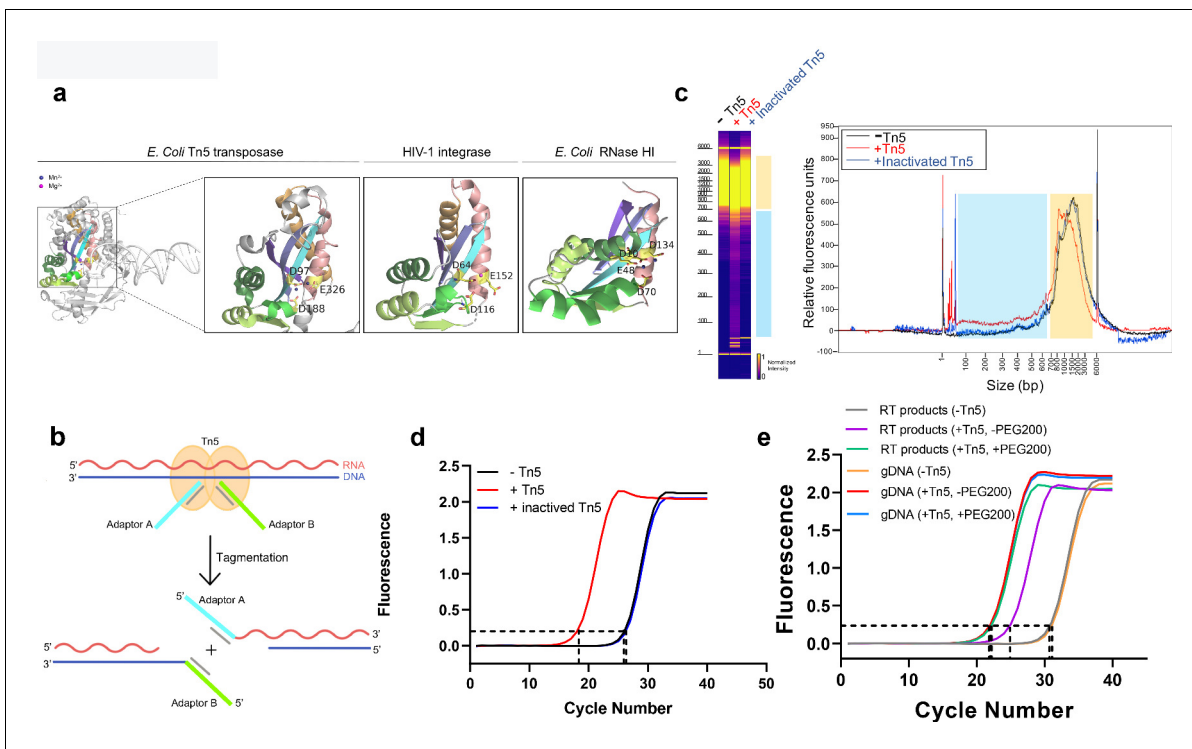


Figure 1. Tn5 transposome has direct tagmentation activity on RNA/DNA hybrid duplexes. (a) Crystal structure of a single subunit of *E. coli* Tn5 Transposase (PDB code 1MM8) complexed with ME DNA duplex, and zoom-in views of the conserved catalytic core of Tn5 transposase, HIV-1 integrase (PDB code 1BIU), and *E. coli* RNase HI (PDB code 1G15), all of which are from the retroviral integrase superfamily. Active-site residues are shown as sticks, and the Mn^{2+} and Mg^{2+} ions are shown as deep blue and magenta spheres. (b) Schematic of Tn5-assisted tagmentation of RNA/DNA hybrids. (c) Gel pictures (left) and peak pictures (right) represent size distributions of HEK293T mRNA-derived RNA/DNA hybrid fragments after incubation without Tn5 transposome, with Tn5 transposome, and with inactivated Tn5 transposome. The blue and orange patches denote small and large fragments, respectively. (d) qPCR amplification curve of tagmentation products of HEK293T mRNA-derived RT samples with Tn5 treatment, with inactivated Tn5 treatment, or without Tn5 treatment. Average Ct values of two technical replicates are 18.06, 26.25 and 26.41, respectively. (e) qPCR amplification curve of tagmentation products of HEK293T mRNA-derived RT products samples and gDNA samples under different conditions. (Average Ct values of three technical replicates: RT products sample without Tn5 treatment = 30.38; RT products sample with PEG200 = 21.94; RT products sample without PEG200 = 25.23; gDNA sample without Tn5 treatment = 30.71; gDNA sample with PEG200 = 21.15; gDNA sample without PEG200 = 21.19).

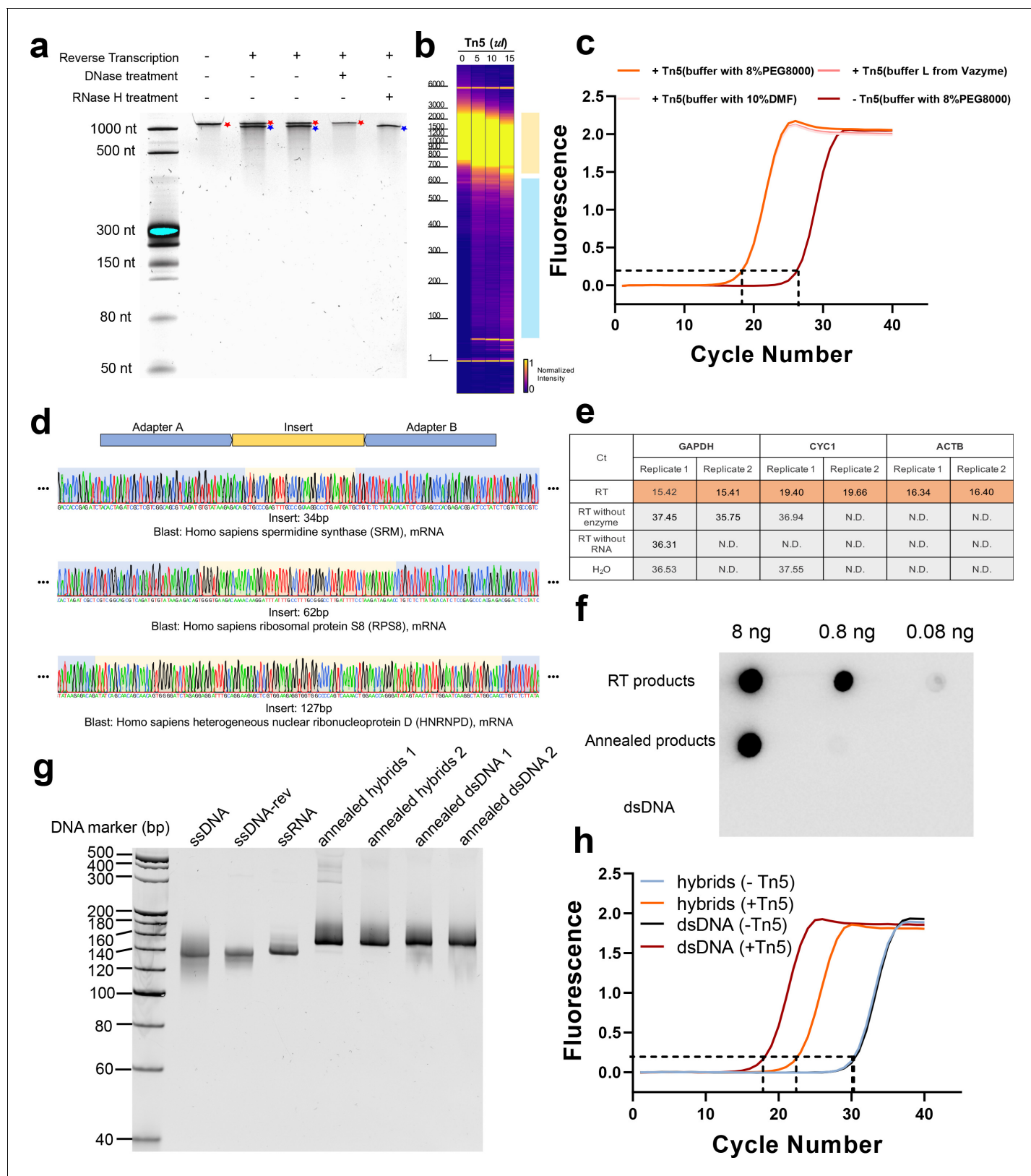


Figure 1—figure supplement 1. Tagmentation activity of Tn5 transposome on RNA/DNA hybrids. (a) Denaturing (8 M urea) polyacrylamide gel analysis of reverse transcription products of an in vitro transcribed mRNA (IRF9). Lane 1: ssRNA marker. Lane 2: in vitro transcribed mRNA (IRF9). Lane 3 and 4: reverse transcription products of an in vitro transcribed mRNA (IRF9). Lane 5: reverse transcription product treated with DNase I. Lane 6: reverse transcription product treated with RNase H. (b) Heatmap showing Tn5 tagmentation activity on RNA/DNA hybrids. (c) Real-time PCR amplification curves for RT products. (d) Sanger sequencing of Tn5 inserts. (e) Real-time PCR quantification of RT products. (f) Dot blot analysis of RT products. (g) Denaturing PAGE analysis of DNA markers and hybrids. (h) Real-time PCR amplification curves for hybrids and dsDNA.

Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

transcription product treated with RNase H. ssRNA and ssDNA is marked with a red asterisk and a blue pound sign, respectively. **(b)** Gel picture showing size distribution of RNA/DNA hybrids products of 50 μ l reaction systems without Tn5 transposome, and with 5 μ l, 10 μ l, and 15 μ l Tn5 transposome, respectively. The blue and orange patches denote small and large fragments, respectively. **(c)** qPCR amplification curve of tagmentation products without Tn5 treatment or with Tn5 treatment in three different buffers (see Methods). Average Ct values of two technical replicates are 26.41, 18.39, 18.33 and 18.34, respectively. **(d)** Sanger sequencing chromatograms of PCR products following RNA/DNA hybrid tagmentation and strand extension. Adaptor A and B sequences are highlighted with blue background color and insert sequences are highlighted with yellow background. **(e)** Assessment of gDNA contamination by qPCR of represented genes. **(f)** Dot blot analysis of a series of diluted samples using the anti-hybrid S9.6 antibody. S9.6 antibody showed no cross-reactivity with dsDNA and the successful hybrids productions were confirmed in CLuc annealed products and mRNA RT products. **(g)** Native PAGE analysis of 150 bp CLuc annealed products under different annealing conditions. (Annealed hybrids 1: RNA:DNA = 2:1; Annealed hybrids 2: RNA:DNA = 1.2:1; Annealed dsDNA 1: ssDNA:ssDNA-rev = 2:1; Annealed dsDNA 2: ssDNA:ssDNA-rev = 1.2:1; See Methods). **(h)** qPCR amplification curve of tagmentation products of CLuc annealed RNA/DNA hybrid and dsDNA products with Tn5 treatment or without Tn5 treatment. Average Ct values of three technical replicates of annealed hybrid products are 22.68 and 30.4, respectively. Average Ct values of three technical replicates of annealed dsDNA products are 18.08 and 30.83, respectively.

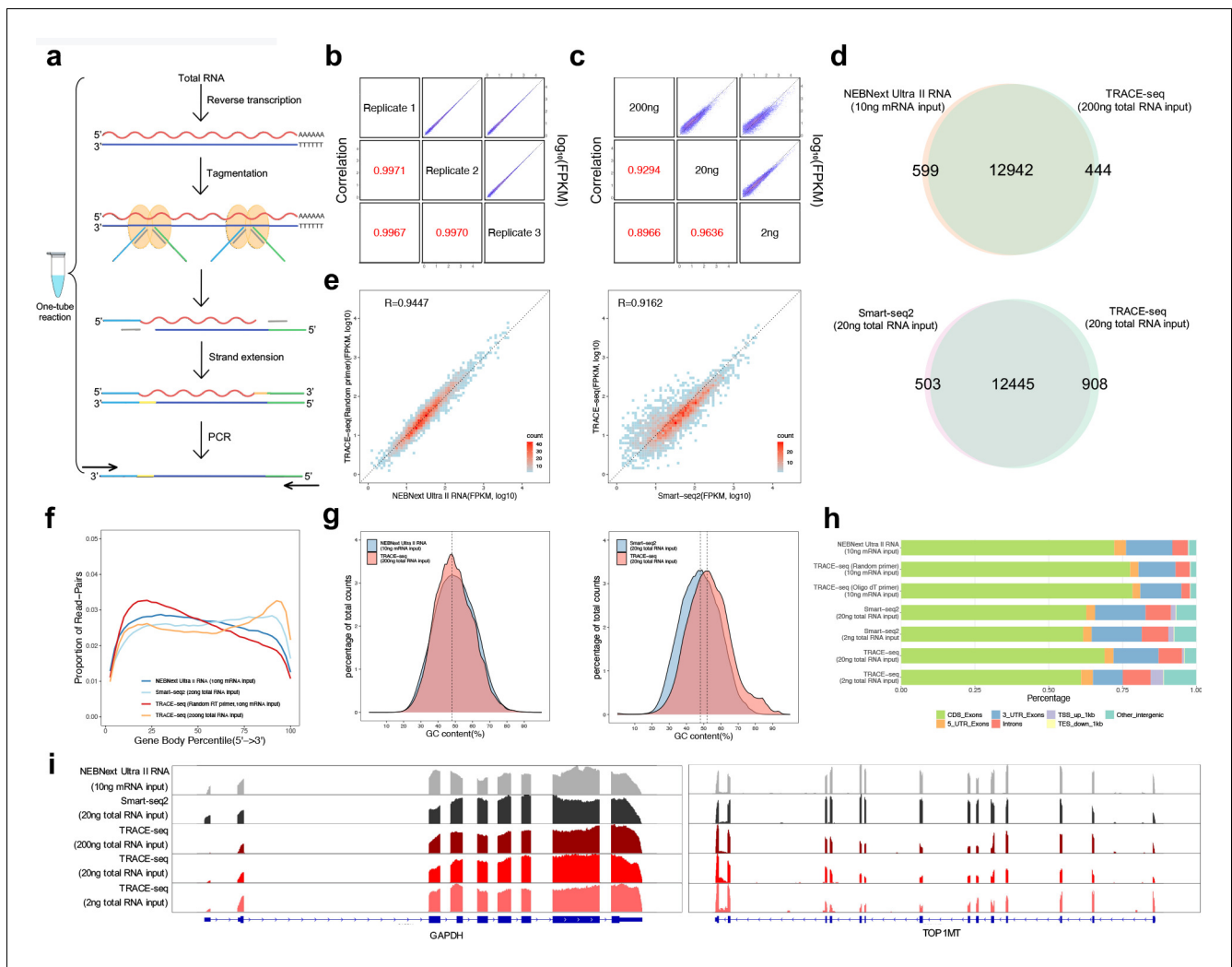


Figure 2. Workflow and evaluation of TRACE-seq. (a) Workflow of TRACE-seq. (b) Gene expression, measured by three technical replicates of TRACE-seq with 200 ng total RNA as input, are shown as scatter plots in the upper right half. Pearson's product-moment correlations are displayed in the lower left half. (c) Gene expression, measured by TRACE-seq using 200 ng, 20 ng and 2 ng total RNA as input, are shown as scatter plots in the upper right half. Pearson's product-moment correlations are displayed in the lower left half. (d) Venn diagrams of gene numbers detected by TRACE-seq with 200 ng total RNA as input and NEBNext Ultra II RNA kit with 20 ng mRNA as input (top) and by TRACE-seq with 20 ng total RNA as input and Smart-seq2 with 20 ng mRNA as input (below). (e) Scatterplots showing a set of housekeeping gene expression values for TRACE-seq and NEBNext Ultra II RNA kit with 10 ng mRNA as input (left), and for TRACE-seq with 10 ng mRNA as input and Smart-seq2 with 20 ng total RNA as input (right). Pearson's product-moment correlation is displayed in the upper left corner. (f) Comparison of read coverage over gene body for NEBNext Ultra II RNA kit, Smart-seq2 and TRACE-seq with different amount of RNA as input. The read coverage over gene body is displayed along with gene body percentile from 5' to 3' end. (g) Distribution of GC content of all mapped reads from TRACE-seq library with 200 ng total RNA as input and NEBNext Ultra II RNA library with 10 ng mRNA as input (left) or Smart-seq2 library with 20 ng total RNA as input (right). The vertical dashed lines indicate 48% (left) and 48% and 51% respectively (right). (h) Comparison of the distribution of reads across known genome features for NEBNext Ultra II RNA kit, Smart-seq2 and TRACE-seq with different amount of RNA as input. (i) IGV tracks showing the coverage of two representative transcripts (GAPDH and TOP1MT). The data come from NEBNext Ultra II RNA kit, Smart-seq2 and three sets of TRACE-seq with different amount of total RNA as input.

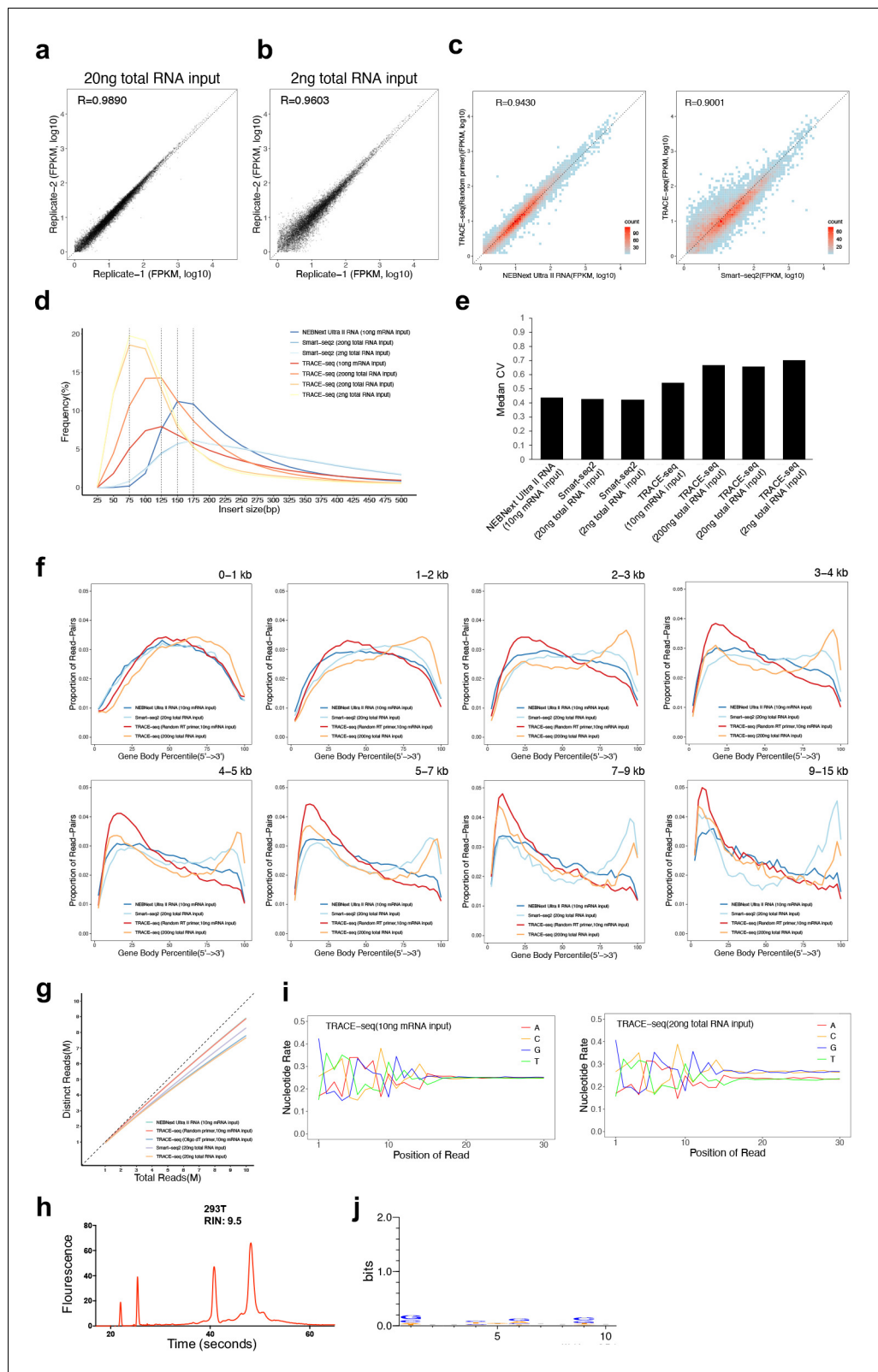


Figure 2—figure supplement 1. Quality assessment of TRACE-seq. (a) Gene expression measured by two technical replicates of TRACE-seq with 20 ng total RNA as input are shown as scatter plots. Pearson’s product-moment correlations are displayed in the upper left corner. (b) Gene expression measured by two technical replicates of TRACE-seq with 2 ng total RNA as input are shown as scatter plots. Pearson’s product-moment correlations are displayed in the upper left corner. (c) Gene expression measured by two technical replicates of TRACE-seq with 2 ng total RNA as input are shown as scatter plots. Pearson’s product-moment correlations are displayed in the upper left corner. (d) Frequency of insert sizes for various methods. (e) Median CV for various methods. (f) Proportion of read-pairs for various methods across different gene body regions. (g) Distinct reads vs total reads. (h) Fluorescence trace for 293T cells with RIN: 9.5. (i) Nucleotide rate vs position of read for TRACE-seq. (j) Bits for various methods. *Figure 2—figure supplement 1 continued on next page*

Figure 2—figure supplement 1 continued

correlations are displayed in the upper left corner. (c) Scatterplots showing gene expression values for TRACE-seq and NEBNext Ultra II RNA kit with 10 ng mRNA as input (left), and for TRACE-seq with 10 ng mRNA as input and Smart-seq2 with 20 ng total RNA as input (right). All expressed genes (FPKM >0.5) are included. Pearson's product-moment correlation is displayed in the upper left corner. (d) Distribution of the insert size in NEBNext Ultra II RNA library, Smart-seq2 libraries and TRACE-seq libraries with different amount of RNA as input, respectively. (e) Median coefficient of variation of gene coverage over the 1000 most highly expressed transcripts in NEBNext Ultra II RNA library, Smart-seq2 libraries and TRACE-seq libraries with different amount of RNA as input, respectively. (f) Comparison of read coverage over gene body for NEBNext Ultra II RNA kit, Smart-seq2 and TRACE-seq with different amount of RNA as input. Transcripts were grouped according to annotated lengths and analyzed separately, with the transcript length ranges indicated (top right). The read coverage over gene body is displayed along with gene body percentile from 5' to 3' end. (g) Library complexity for each library shown by plotting the number of uniquely occurring read-pairs with respect to total number of sampled read-pairs. (h) Assessment of RNA Integrity number (RIN). RNA of high RIN score (9.5) was used as input, which allows us to rule out the possibility that the 3' end bias of gene body coverage is due to RNA degradation. (i) Nucleotide versus cycle (NVC) plots showing percentage of observed bases at each position of the first 30 bases of each sequencing read from TRACE-seq library with 10 ng mRNA and 20 ng total RNA as input. (j) WebLogo plot of sequence conservations of the first 10 bases of all sequencing reads from TRACE-seq library with 10 ng mRNA as input. The overall height of the stack indicates the sequence conservation at that position (measured in bits), while the height of symbols within the stack indicates the relative frequency of each nucleic acid at that position.

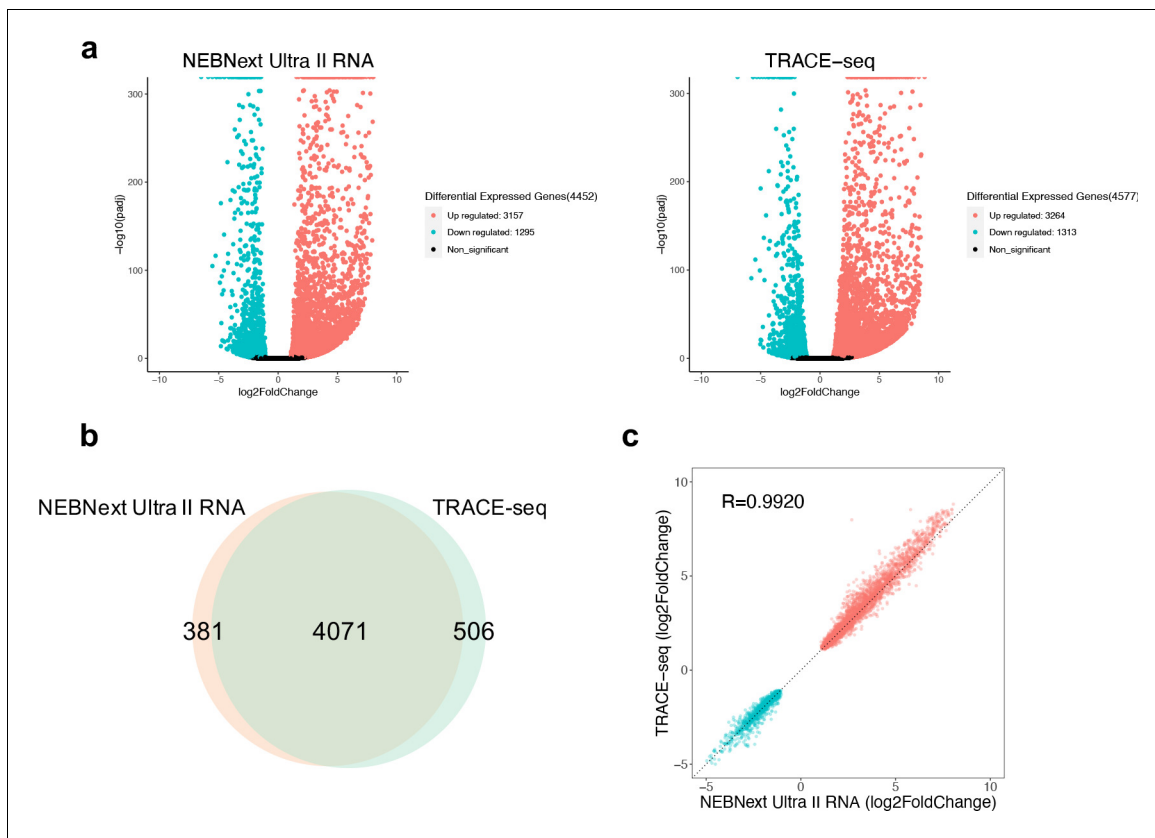


Figure 3. Performance of TRACE-seq in differential expression analysis. (a) Volcano plot showing differential expressed genes between undifferentiated and differentiated mESCs detected by NEBNext Ultra II RNA kit and TRACE-seq. Significantly up-regulated and down-regulated expressed genes ($p_{adj} < 0.05$, $|\log_2\text{FoldChange}| > 1$) are highlighted in red and blue, respectively. (b) Venn diagram of differentially expressed gene numbers detected by TRACE-seq and NEBNext Ultra II RNA kit. (c) Correlation between the fold change of the 4071 differentially expressed genes that overlap between NEBNext Ultra II RNA and TRACE-seq library.