Tn5 tagments and transports oligos to single-strand DNA for strand-specific RNA sequencing

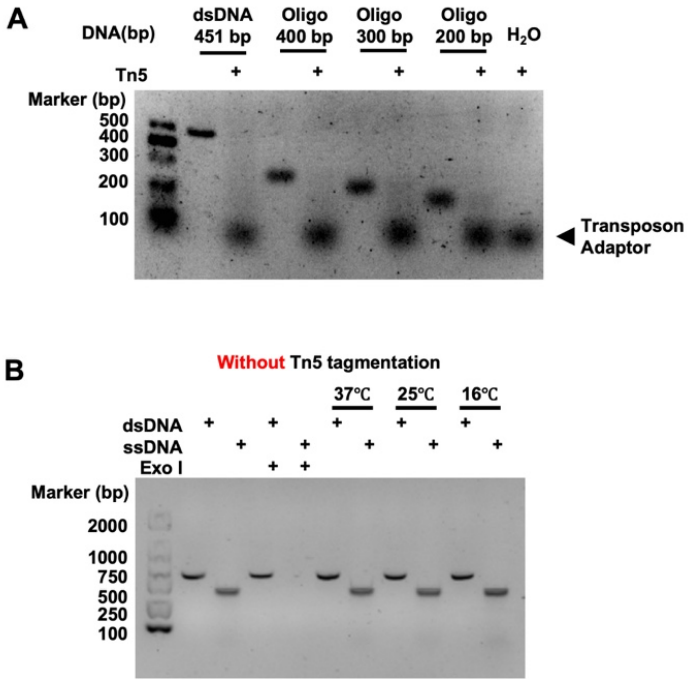
Yanjun Zhang[1, #], Yin Tang[1, #], Zhongxing Sun[1], Junqi Jia[1], Yuan Fang[1], Xinyi Wan[1], and Dong Fang[1, 2, *]

[1] Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China.

[2] Department of Medical Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.

[#] These authors contribute equally

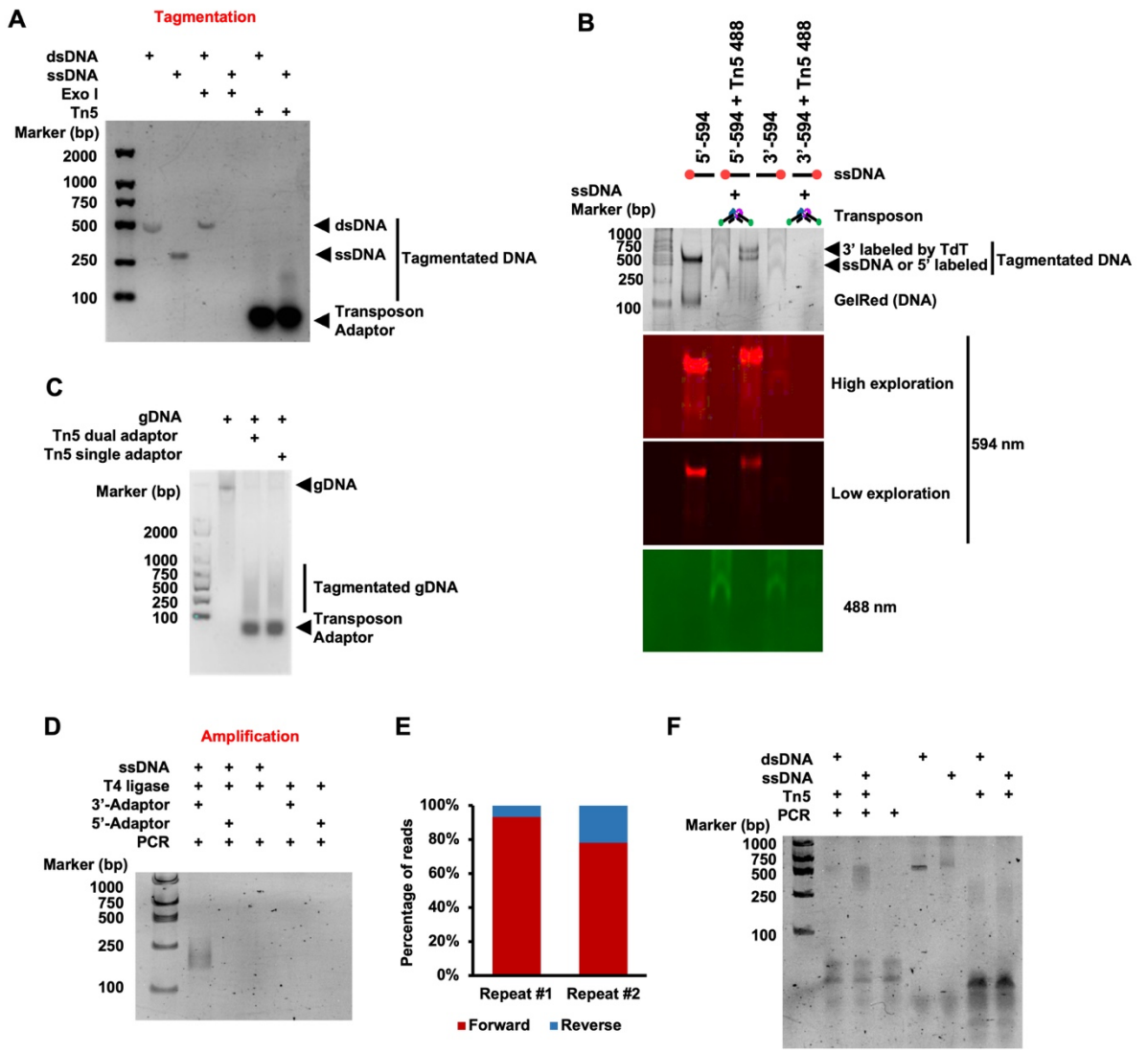
[*] Correspondence: dfang@zju.edu.cn



Supplemental Figure S1. Tn5 transposon tagments ssDNA.

(**A**) Tn5 transposon tagments synthesized oligos. Different lengths of oligos were used for Tn5 tagmentation. dsDNA was used as a positive control for the tagmentation.

(**B**) Purified dsDNA and ssDNA were incubated under different temperatures for 5 min.



Supplemental Figure S2. Tn5 transposon is used to generate the sequencing library from ssDNA.

(**A**) Gel result presenting the tagmentation of dsDNA and ssDNA. Exo I was used to indicate the ssDNA.

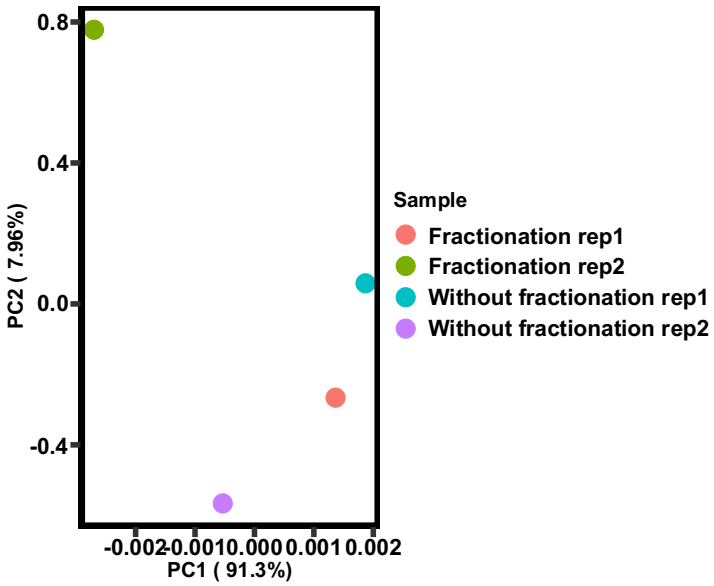
(**B**) The Tn5 transposon can add oligos to the 5' end of ssDNA. ssDNA with Alexa Fluor 594 labeled 5' or 3' end was tagmented by the Tn5 transposon with Alexa Fluor 488 labeled adaptors.

(**C**) Gel result showing the transposition effect of Tn5 with dual and single adaptors. gDNA, genomic DNA from 293T cells.

(**D**) Gel result presenting TABLE-seq method could amplify the tagmented ssDNA by using adaptors ligated to the 3' end but not adaptors ligated to the 5' end.

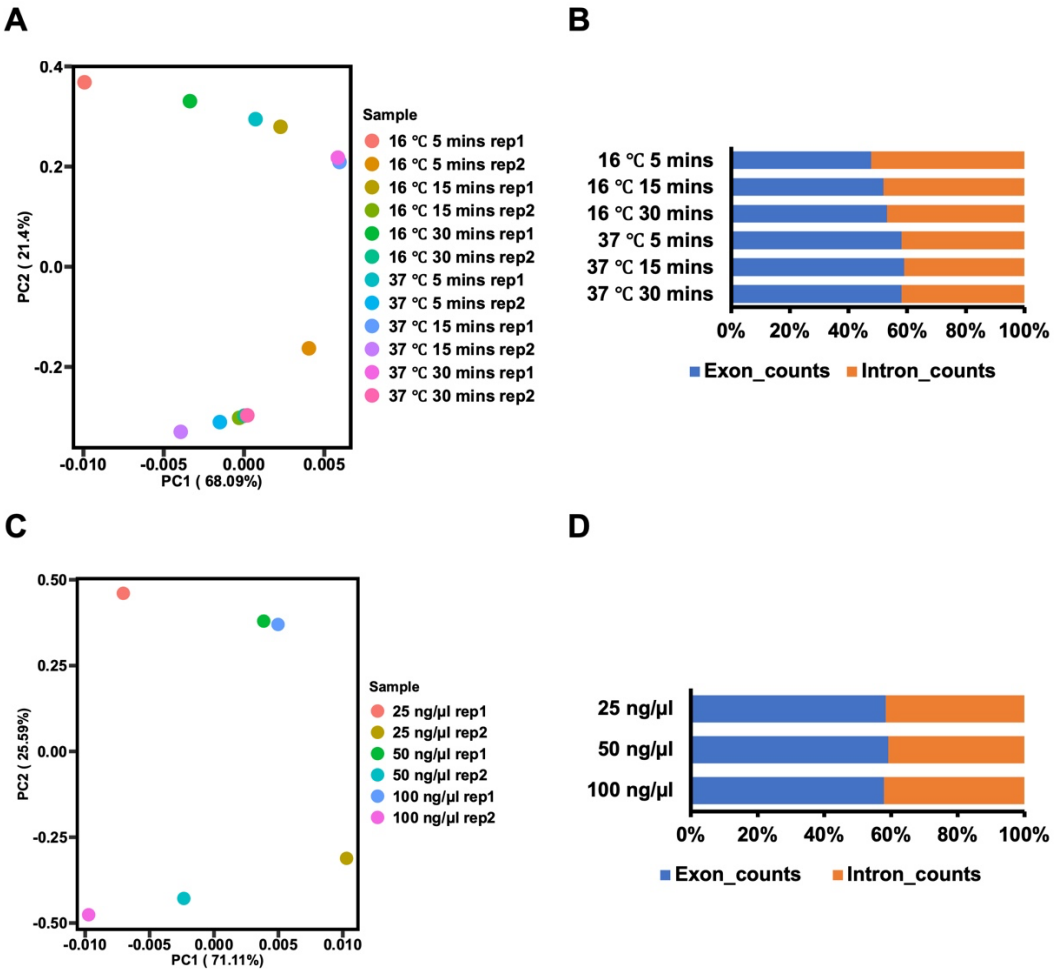
(**E**) Sequencing reads from TABLE-seq were mapped to the forward direction of ssDNA. The percentage of reads mapped to the forward and reverse directions of ssDNA was shown.

(**F**) Gel result presenting TABLE-seq method could amplify the tagmented ssDNA.



Supplemental Figure S3. Fragmented and un-fragmented RNA are both suitable for TABLE-seq based RNA sequencing.

Gene expression PCA plot showing the principal component between different sequencing results.



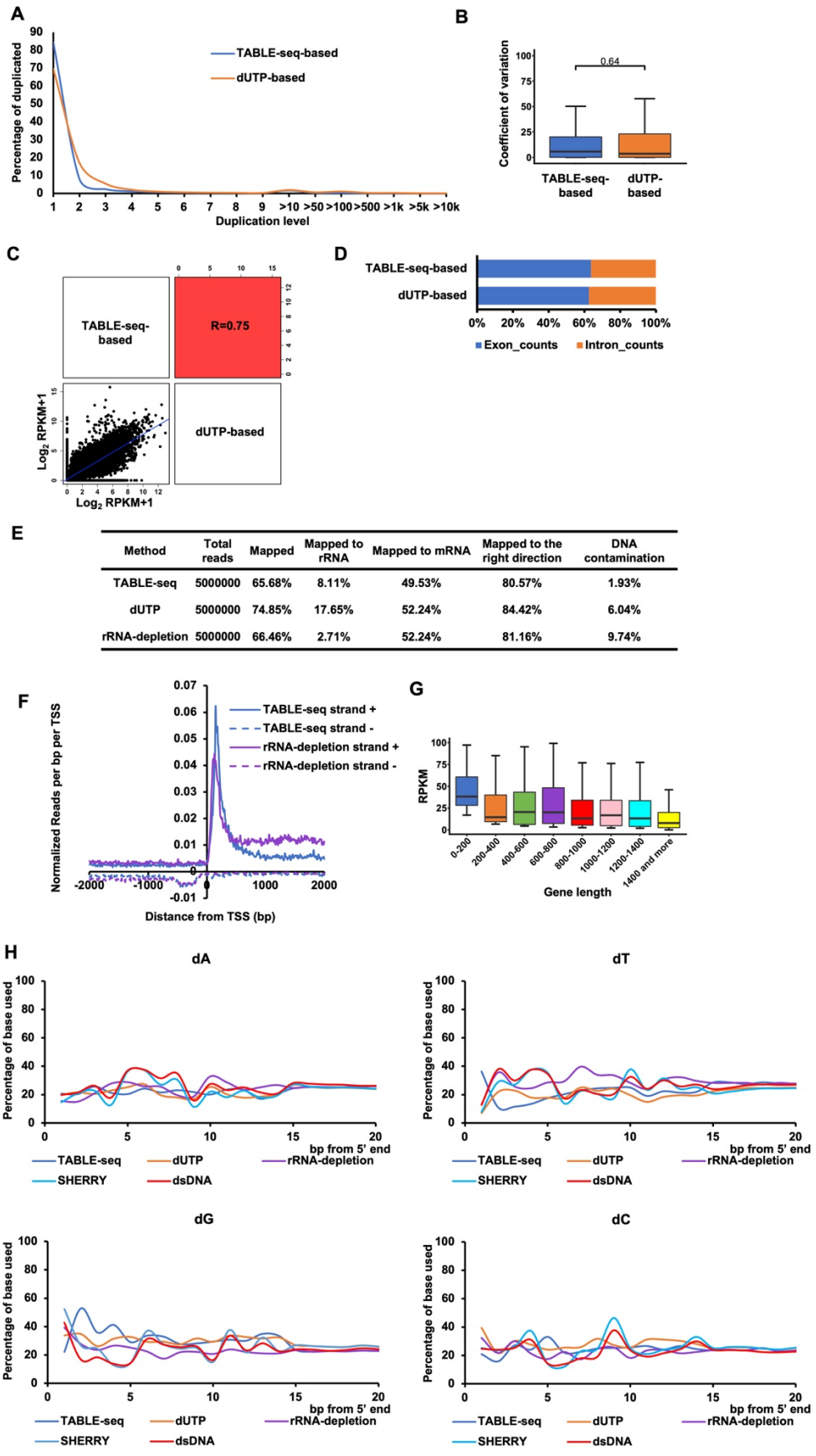
Supplemental Figure S4. Sequencing libraries generated by different reaction conditions show similar gene expression profiles.

(**A**) Gene expression PCA plot showing the principal component between different sequencing results with the indicated reaction time and temperature.

(**B**) Bar graph showing ratios of detected reads at exons and introns. The number of reads mapped to exons or introns was calculated by Homer. Libraries with the indicated reaction time and temperature were sequenced.

(**C**) Gene expression PCA plot showing the principal component between different sequencing results with the indicated amounts of Tn5.

(**D**) Bar graph showing ratios of detected reads at exons and introns among the libraries generated by the indicated amounts of Tn5.



Method	Total reads	Mapped	Mapped to rRNA	Mapped to mRNA	Mapped to the right direction	DNA contamination
TABLE-seq	5000000	65.68%	8.11%	49.53%	80.57%	1.93%
dUTP	5000000	74.85%	17.65%	52.24%	84.42%	6.04%
rRNA-depletion	5000000	66.46%	2.71%	52.24%	81.16%	9.74%

Supplemental Figure S5. TABLE-seq and dUTP based sequencing libraries exhibit similar gene expression profiles.

(**A**) The level of reads duplicates in TABLE-seq- and dUTP-based methods.

(**B**) Coefficient of variation of TABLE-seq- and dUTP-based methods. The variation between each gene in two replicates was calculated.

(**C**) The Pearson's correlations among libraries generated by the indicated library construction methods. The scatter plots showed all detected genes. RPKM, reads per kilobase of exon per million reads mapped.

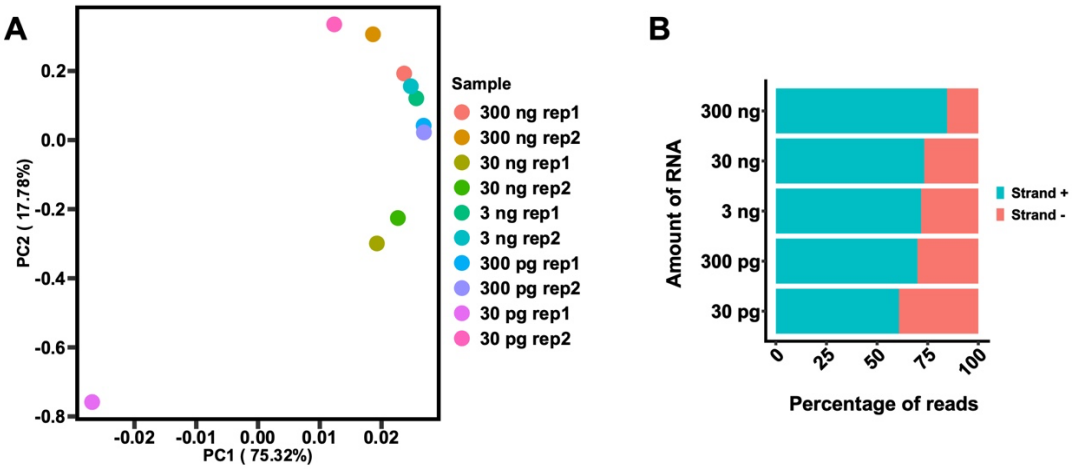
(**D**) Bar graph showing ratios of detected reads at exons and introns. The number of reads mapped to exons or introns was calculated by Homer. Libraries constructed with the indicated methods were sequenced.

(**E**) Library quality of TABLE-seq-, dUTP-, and rRNA depletion-based methods.

(**F**) Comparison of sequencing signals between libraries generated by the indicated methods. Sense (strand +) and antisense (strand -) transcripts associated with transcription start site (TSS) were shown.

(**G**) The expression levels of genes with different lengths in mESC. The sequencing data used were from TABLE-seq. RPKM, reads per kilobase of exon per million reads mapped.

(**H**) The base preferences at first 20 base pairs among different sequencing methods.



Supplemental Figure S6. TABLE-seq can be applied to strand-specific RNA sequencing.

(**A**) Gene expression PCA plot showing the principal component among different sequencing results.

(**B**) Bar graph showing the ratio of reads mapped to sense (+) and antisense (-) strands among different sequencing results.

Supplemental Table S1. Oligonucleotides used in the study.

Name	5'- 3'	Application
Tn5ME-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	Transposons Assembly
Tn5ME-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	Transposons Assembly
Tn5ME-A-5'AF488	[AF488]TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	Transposons Assembly
Tn5MErev	[phos]CTGTCTCTTATACACATCT	Transposons Assembly
5ph_Tn5a	[phos]CTGTCTCTTATACACATCTGACGCTGCCGACGA	Adaptor anneal
Tn5a_N6_invert_dT	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNN[invert dT]	Adaptor anneal
Tn5a	CTGTCTCTTATACACATCTGACGCTGCCGACGA	Adaptor anneal
5'_N6_Tn5a_invert_dT_3'	NNNNNNTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG[invert dT]	Adaptor anneal
H3.1F-5'AF594	[AF594]ATATAGGATCCATGGCTCGTACGAAGCAAAC	Asymmetric PCR
H3.1R	TATAACTCGAGCTAGGCGTAGTCGGGCACGTC	Asymmetric PCR
ss400F	GCGTGAAGAAACCTCATCGC	Asymmetric PCR
ss300F	CCAGCGCCTGGTGCGAGAAATC	Asymmetric PCR
ss200F	GTGGGACTCTTCGAAGACAC	Asymmetric PCR
ss160F	CTAAACGCGTCACCATCATG	Asymmetric PCR
ss130F	TCCAGCTGGCACGTCGCATC	Asymmetric PCR
ss400bpoligo	ACACTTTTGTGTGTGCTCTCATTGCAAATGGCTCGTACGAAGCAAACAGCTCGCAAGTCTACCGGCGGCAAAGCTCCGCGCAAGCAGCTTGCTACTAAAGCAGCCCGTAAGAGCGCTCCGGCCACCGGTGGCGTGAAGAAACCTCATCGCTACCGCCCGGGCACCGTGGCCTTGCGCGAAATCCGTCGCTACCAGAAGTCCACCGAGCTGCTGATCCGGAAGCTGCCGTTCCAGCGCCTGGTGCGAGAAATCGCCCAGGACTTCAAAACCGACCTGCGTTTCCAGAGCTCTGCGGTGATGGCGCTGCAGGAGGCTTGTGAGGCCTACCTGGTGGGACTCTTCGAAGACACCAATCTGTGCGCTATTCACGCTAAACGCGTCACCATCATGCCCAAAGATA	Tn5 tagmentation

ss300oligo	ACACTTTTGTGTGTGCTCTCATTGCAAATGG CTCGTACGAAGCAAACAGCTCGCAAGTCTA CCGGCGGCAAAGCTCCGCGCAAGCAGCTTG CTACTAAAGCAGCCCGTAAGAGCGCTCCGG CCACCGGTGGCGTGAAGAAACCTCATCGCT ACCGCCCGGGCACCGTGGCCTTGCGCGAAA TCCGTCGCTACCAGAAGTCCACCGAGCTGC TGATCCGGAAGCTGCCGTTCCAGCGCCTGG TGCGAGAAATCGCCCAGGACTTCAAAACCG ACCTGCGTTTCCAGAGCTCTGCGGTGATG	Tn5 tagmentation
ss200oligo	ACACTTTTGTGTGTGCTCTCATTGCAAATGG CTCGTACGAAGCAAACAGCTCGCAAGTCTA CCGGCGGCAAAGCTCCGCGCAAGCAGCTTG CTACTAAAGCAGCCCGTAAGAGCGCTCCGG CCACCGGTGGCGTGAAGAAACCTCATCGCT ACCGCCCGGGCACCGTGGCCTTGCGCGAAA TCCGTCGCTACCAGAAGTC	Tn5 tagmentation
ss110oligo	GGTATCAACGCAGAGTACATGTCATTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTCTTG TACAGCTCGTCC	Tn5 tagmentation
Index-F1	AATGATACGGCGACCACCGAGATCTACACTA GATCGCTCGTCGGCAGCGTCAGATGTGTAT	Forward primer for library
Index-F2	AATGATACGGCGACCACCGAGATCTACACCT CTCTATTCGTCGGCAGCGTCAGATGTGTAT	Forward primer for library
Index-F3	AATGATACGGCGACCACCGAGATCTACACTA TCCTCTTCGTCGGCAGCGTCAGATGTGTAT	Forward primer for library
Index-F4	AATGATACGGCGACCACCGAGATCTACACA GAGTAGATCGTCGGCAGCGTCAGATGTGTAT	Forward primer for library
Index-F5	AATGATACGGCGACCACCGAGATCTACACGT AAGGAGTCGTCGGCAGCGTCAGATGTGTAT	Forward primer for library
Index-F6	AATGATACGGCGACCACCGAGATCTACACA CTGCATATCGTCGGCAGCGTCAGATGTGTAT	Forward primer for library
Index-R1	CAAGCAGAAGACGGCATACGAGATTCGCCT TAGTCTCGTGGGCTCGGAGATGTG	Reverse primer for library
Index-R2	CAAGCAGAAGACGGCATACGAGATCTAGTA CGGTCTCGTGGGCTCGGAGATGTG	Reverse primer for library
Index-R3	CAAGCAGAAGACGGCATACGAGATAGGAGT CCGTCTCGTGGGCTCGGAGATGTG	Reverse primer for library

Supplemental Table S2. Correlations between two replicates. A 10 bp sliding window across the ssDNA region was used to calculate the Pearson's product moment correlation for ssDNA sequencing. The correlation of gene expression repeats was calculated by the coefficient of determination (R^2) of their shared genes' RPKM values.

Samples	Correlation
ssDNA	0.99
dUTP_based	0.99
TABLE-seq_frag	0.98
TABLE-seq_unfrag	0.99
TABLE-seq_RNA 300ng	0.86
TABLE-seq_RNA 30ng	0.85
TABLE-seq_RNA 3ng	0.82
TABLE-seq_RNA 300pg	0.60
TABLE-seq_RNA 30pg	0.45
TABLE-seq_16°C 5min	0.84
TABLE-seq_16°C 15min	0.82
TABLE-seq_16°C 30min	0.78
TABLE-seq_37°C 5min	0.80
TABLE-seq_37°C 15min	0.83
TABLE-seq_37°C 30min	0.85
TABLE-seq_TN5_25ng/μl	0.79
TABLE-seq_TN5_50ng/μl	0.78
TABLE-seq_TN5_100ng/μl	0.75