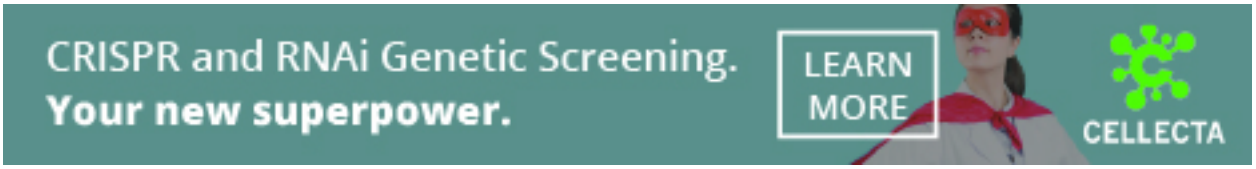
Tn5 tagments and transposes oligos to single-stranded DNA for strand-specific RNA sequencing

Yanjun Zhang, Yin Tang, Zhongxing Sun, et al.

Genome Res. 2023 33: 412-426 originally published online March 23, 2023
Access the most recent version at doi:10.1101/gr.277213.122



Method

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

Yanjun Zhang,[1,3] Yin Tang,[1,3] 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 310009, China*

Tn5 transposon tagments double-stranded DNA and RNA/DNA hybrids to generate nucleic acids that are ready to be amplified for high-throughput sequencing. The nucleic acid substrates for the Tn5 transposon must be explored to increase the applications of Tn5. Here, we found that the Tn5 transposon can transpose oligos into the 5′ end of single-stranded DNA longer than 140 nucleotides. Based on this property of Tn5, we developed a tagmentation-based and ligation-enabled single-stranded DNA sequencing method called TABLE-seq. Through a series of reaction temperature, time, and enzyme concentration tests, we applied TABLE-seq to strand-specific RNA sequencing, starting with as little as 30 pg of total RNA. Moreover, compared with traditional dUTP-based strand-specific RNA sequencing, this method detects more genes, has a higher strand specificity, and shows more evenly distributed reads across genes. Together, our results provide insights into the properties of Tn5 transposons and expand the applications of Tn5 in cutting-edge sequencing techniques.

[Supplemental material is available for this article.]

Because of advances in high-throughput sequencing and its low cost, the method is extensively available in many aspects of life science research, especially for analyzing genome-wide events (Metzker 2010). New high-throughput sequencing–based methodologies and approaches are being developed for gene expression analysis, epigenetic information profiling, chromatin structure analysis, noncoding RNA evaluation, high-throughput screening, mutation detection, and so on (Kolmar et al. 2022; Severins et al. 2022).

To use the sequencing power of high-throughput sequencing, the DNA of interest needs to be ligated with specific adaptors on both the 5′ and 3′ ends, and if the amount of DNA is low, the DNA must be amplified by PCR (Quail et al. 2008; Kozarewa et al. 2009). This procedure is usually called library preparation, which generally requires the following steps: DNA fragmentation and the addition of adaptors to DNA. Two major strategies are widely used in high-throughput library preparation. One is the DNA ligation–based technique, which fragments target DNA by sonication or enzyme digestion, adds an adenine to DNA ends, and then ligates specific primers to the target DNA by T-A ligation (Oyola et al. 2012). The other method uses transposons to fragment and insert specific primers into the enriched DNA of interest (Adey et al. 2010; Hennig et al. 2018).

Transposons, which were discovered in maize by Barbara McClintock, can "jump" from one location to another within a genome (McClintock 1950). The Tn5 transposons, which are found in *Escherichia coli*, consist of two inverted IS50 sequences. These sequences contain two pairs of 19-bp outside ends (OEs) and inside ends (IEs) (Berg et al. 1975; York and Reznikoff 1996). The OEs are inverted to be the target sites of Tn5 transposase (Johnson and Reznikoff 1983). Tn5 transposases bind to OEs to form a synaptic transposon complex by dimerizing two Tn5 transposases through the C-terminal domain. Then, Tn5 transposases use water molecules to catalyze the hydrolysis of the phosphodiester bond of DNA, forming a nucleophilic hydroxyl group at the 5′ ends of DNA. The hydroxyl group attacks the complementary DNA strand to form a hairpin structure, which is further activated by another water molecule to create a blunt end (Davies et al. 1999, 2000; Steiniger-White et al. 2004). Through this process, Tn5 transposons bind to the target DNA sequence and transfer the transposon DNA strands into both strands, leaving a 9-bp gap at the 3′ end of targeting sites. Through this "cut and paste" process, Tn5 transposons "jump" from one location to other locations (Zhou and Reznikoff 1997; Reznikoff 2003).

Tn5 transposases that are assembled with OEs can tagment and transpose OEs into target DNA in vitro; as a result, owing to the tagmentation of Tn5 with specifically designed sequences in OEs, DNA is ready for PCR amplification (Goryshin and Reznikoff 1998; Naumann and Reznikoff 2002). In addition, the tagmented DNA is fragmented by Tn5 transposition, and the step of DNA fragmentation in high-throughput sequencing is not needed. Using these fragmentation and adaptor ligation properties, Tn5 is widely used in genomic research (Cain et al. 2020; Adey 2021). Tn5 transposons are assembled by mixing Tn5 transposases and designed adaptors with OE sequences. When Tn5 transposons recognize the target DNA, the transposons fragment the target DNA and insert the designed adaptors into each end of the fragmented DNA (Bourque et al. 2018; Adey 2021). The

[3]**These authors contributed equally to this work.**
Corresponding author: dfang@zju.edu.cn
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.277213.122.

33:412–426 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/23; www.genome.org

resulting DNA can be amplified by adding barcodes and sequencing adaptors for high-throughput sequencing (Kia et al. 2017).

We and others have found that Tn5 can tagment cDNA/RNA hybrids to generate PCR-ready products to quickly construct RNA sequencing libraries (Di et al. 2020; Lu et al. 2020; Sun et al. 2021). However, little is known about how Tn5 transposon tagments single-stranded DNA (ssDNA). Here, we applied Tn5 to tagment ssDNA to uncover the potential applications of this widely used enzyme in omics studies.

Results

Tn5 tagments ssDNA

We and others have shown that Tn5 tagments and transposes oligos to the DNA/RNA hybrids (Di et al. 2020; Lu et al. 2020; Sun et al. 2021). To explore how Tn5 tagments different nucleic acid substrates, we treated double-stranded DNA (dsDNA), total RNA, and ssDNA with the Tn5 transposon.

As previously reported (Naumann and Reznikoff 2002), Tn5 transposon tagmented dsDNA, which was purified as genomic DNA from HEK293T cells (Fig. 1A). The Tn5 transposon showed little ability to tagment total RNA that was purified from HEK293T cells (Fig. 1B). To synthesize ssDNA, we used asymmetric PCR, which used an unbalanced concentration of forward and reverse primers (Heiat et al. 2017), and amplified an H3.1 DNA sequence from a plasmid as the template. After PCR, we detected an upper and a lower DNA band (Fig. 1C). Moreover, when subjected to ExoI treatment, which digests only ssDNA but not dsDNA, the lower DNA band was abolished, indicating that the lower band represented the synthesized ssDNA. As expected, Tn5 tagmented the upper band that represented the dsDNA band. To our surprise, the Tn5 treatment also tagmented the lower band that represented the ssDNA band. To further validate the identities of synthesized DNA, we purified the upper and lower bands of DNA and treated this DNA with ExoI and DNase I, respectively (Fig. 1D). Unlike ExoI, which only digests ssDNA, DNase I catalyzes the removal of nucleotides from both ssDNA and dsDNA (Korada et al. 2013). The upper band of DNA was abolished only by DNase I but not by ExoI, and the lower band of DNA was eliminated by both DNase I and ExoI, further showing the observation that the upper band of DNA was dsDNA and the lower band of DNA was ssDNA.

We speculate that Tn5 requires a certain length of ssDNA to be bound by both monomers (Davies et al. 2000; Steiniger-White et al. 2004). To test this hypothesis, we synthesized ssDNA with different lengths by asymmetric PCR and then subjected this ssDNA to ExoI and Tn5 treatment (Fig. 1E). The synthesized ssDNA was fully digested by ExoI, confirming that the DNA was single-stranded. In addition, ≥160-nt ssDNA was tagmented by Tn5, whereas 130-nt and 110-nt ssDNA were not tagmented, indicating that a length >130 nt was necessary for the tagmentation of Tn5 against ssDNA. In addition to the ssDNA generated by asymmetric PCR, we further tested whether synthesized oligos could be tagmented by Tn5 transposon. dsDNA was used as a positive control (Supplemental Fig. S1A). Oligos with a length >200 nt were tagmented after being incubated with Tn5 transposon. Together, these data suggest that Tn5 tagments ssDNA. We further analyzed the tagmentation efficiency of Tn5 transposons against similar lengths of dsDNA and ssDNA. When subjected to 5 min at 37°C, 25°C, and 16°C, respectively, the dsDNA and ssDNA were tagmented to similar degrees (Fig. 1F). In addition, the tagmented ssDNA at 16°C showed a higher band, which was not generated by the process of incubation (Supplemental Fig. S1B), suggesting the ssDNA was ligated with transposon adaptors.

Tn5 transposes oligos into ssDNA

DNA fragmentation by Tn5 transposon relied on nucleophilic attacks by a 3′ hydroxyl group of Tn5 adaptors (Davies et al. 1999, 2000; Steiniger-White et al. 2004). In addition, Tn5 catalyzes the translocation of transposable elements through a "cut and paste" mechanism, leading to a 9-bp gap between the nontransferred strand and the target DNA (Picelli et al. 2014). This attacking direction and 9-bp gap could inhibit the transposition of oligos at the 3′ end of ssDNA (Fig. 2A). To further examine whether Tn5 could "paste" oligos to both ends or a single end of ssDNA, we designed to tagment ssDNA by Tn5 transposons with dual adaptors and then used Tn5 transposon-specific primers to perform PCR amplification on the resulting DNA. The tagmentation product was amplified by PCR when Tn5 transposed oligos to both ends of ssDNA and was not amplified when only one end of ssDNA was transposed with oligos (Fig. 2B). As previously observed, Tn5 tagmented both dsDNA and ssDNA (Supplemental Fig. S2A). However, only the dsDNA tagmentation product, but not the ssDNA product, could be amplified by PCR (Fig. 2C). To further confirm the ligation of Tn5 adaptors to the ssDNA, we designed a tagmentation experiment using 5′ or 3′ end immunofluorescence-labeled ssDNA as the substrate. The 5′ end–labeled ssDNA was generated by asymmetric PCR with 5′ end Alexa Fluor 594–labeled forward primer. The 3′ end immunofluorescence–labeled ssDNA was synthesized by terminal deoxynucleotidyl transferase, which added the Alexa Fluor 594–labeled dUTP to the 3′ end of ssDNA. In addition, Alexa Fluor 488–labeled Tn5 adaptors were used to assemble the Tn5 transposons. This red Alexa Fluor 594 5′ end–labeled or 3′ end–labeled ssDNA was then tagmented by the Tn5 transposon with green Alexa Fluor 488–labeled adaptors. If the ligation of Tn5 adaptors happened at the 5′ end or the 3′ end of ssDNA, the red immunofluorescence signal at the 5′ end or 3′ end of ssDNA would be removed, and green immunofluorescence–labeled Tn5 adaptors were added to the tagmentation products. If the ligation happened to both ends, both of the red signals at the 5′ and 3′ ends of ssDNA would be removed after tagmentation and only green signals would be detected. We successfully labeled the ssDNA with red Alexa Fluor 594 immunofluorescence (Supplemental Fig. S2B). The mobility shift in the 3′ end–labeled ssDNA was because the addition of dUTP increased the length of ssDNA. The green immunofluorescence signals corresponding to the tagmented ssDNA indicated the ligation of Tn5 adaptors to both 3′ and 5′ Alexa Fluor 594–labeled ssDNA. The Alexa Fluor 594 signals at the 5′ end, but not the ones at the 3′ end, were abolished after tagmentation, indicating that the Tn5 transposon adaptors were mainly ligated to the 5′ end of ssDNA.

Because the transposed Tn5 oligo is likely at the 5′ end of ssDNA, we designed a dsDNA adaptor with a 3′ protruding end for ligation at the 3′ end of ssDNA (Fig. 2D). The random hexamer at the 3′ protruding end was designed to be annealed to the ssDNA to facilitate the ligation of the designed adaptor and tagmented ssDNA. The phosphorylation of the 5′ end at the adaptor was the necessary terminal modification for the ligation of the DNA strand. In addition, we added a reversed dT to the 3′ end of the adaptor to prevent misligation and repress the extension of random hexamer-containing stands, which would increase the fidelity and strand specificity.

Tn5 transposons with two different adaptors were used to tagment dsDNA and insert two different adaptors for the PCR

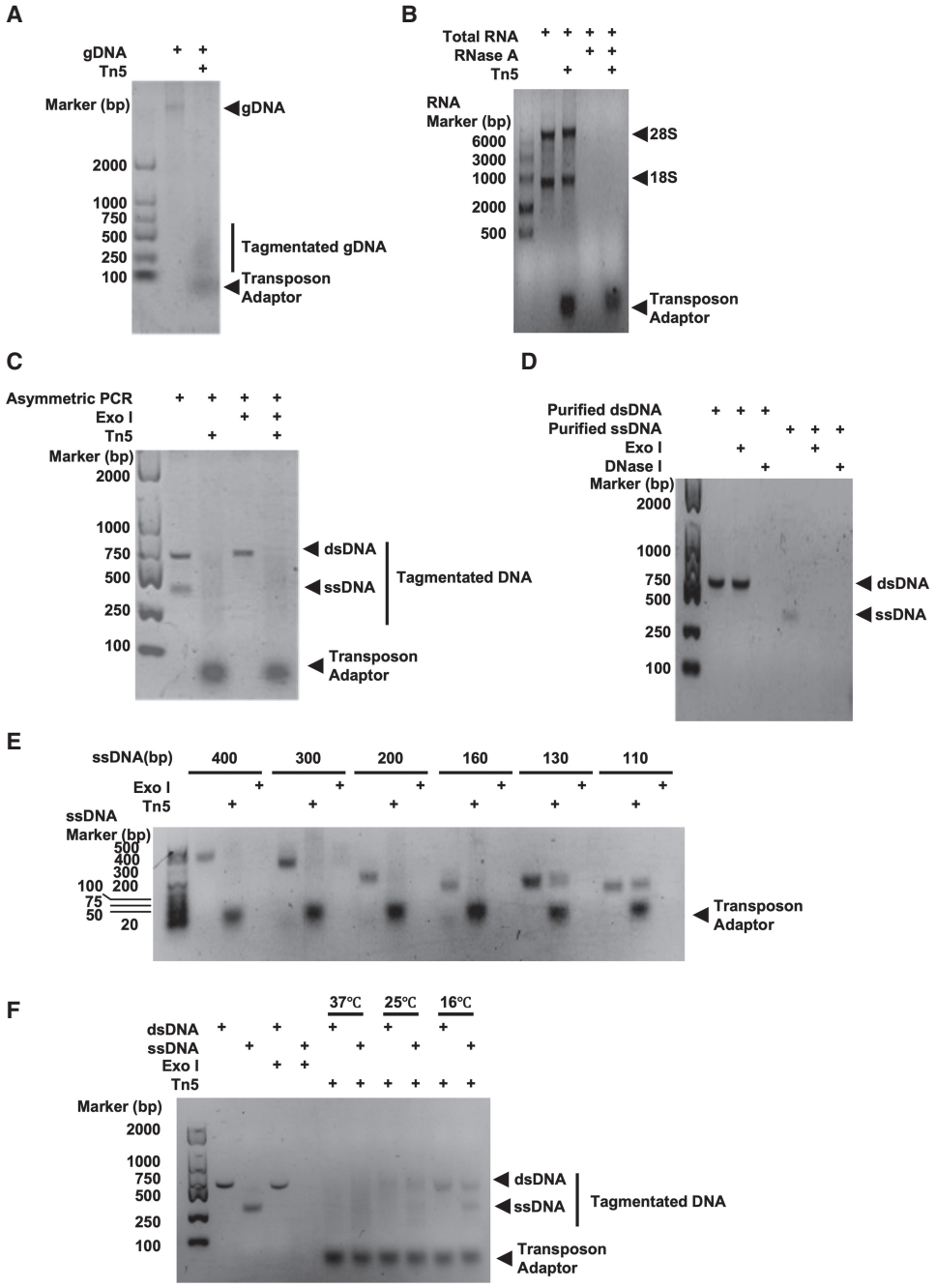


Figure 1. Tn5 transposon tagments ssDNA. (*A*) Tn5 transposon tagmented genomic DNA (gDNA). gDNA from HEK293T cells was subjected to Tn5 transposon tagmentation. (*B*) Tn5 transposon had little activity to tagment total RNA. Total RNA purified from HEK293T cells was subjected to Tn5 tagmentation reaction. RNase A, which digested RNA, was used as the control for the existence of RNA. (*C*) Tn5 transposon tagmented both dsDNA and ssDNA. dsDNA and ssDNA, which were generated from asymmetric PCR, were used for Tn5 tagmentation. ExoI, which only digested ssDNA, was used to indicate the ssDNA. (*D*) dsDNA and ssDNA, which were generated from asymmetric PCR, were purified and treated with different nucleases. ExoI, which only digested ssDNA, was used to indicate the ssDNA. DNase I, which digested both ssDNA and dsDNA, was used to indicate the DNA. (*E*) Tn5 transposon could not tagment ssDNA <130 nt. Different lengths of ssDNA, which were generated from asymmetric PCR, were used for Tn5 tagmentation. (*F*) Tn5 transposon tagmented dsDNA and ssDNA with similar efficiencies. Purified dsDNA and ssDNA were subjected to Tn5 transposon tagmentation under different temperature for 5 min.

amplification (Metzker 2010). However, these two different adaptors lead to a loss of strand specificity in ssDNA amplification, so this system is not suitable for ssDNA amplification with strand specificity. We then synthesized the Tn5 transposon with a single kind of adaptor that could tagment dsDNA as the Tn5 transposon with two different adaptors (Supplemental Fig. S2C). We then used the Tn5 transposon with a single kind of adaptor to tagment ssDNA, which was synthesized from asymmetric PCR. The tagmentation product was subjected to adaptor ligation and PCR amplification. We called this approach tagmentation-based and ligation-enabled

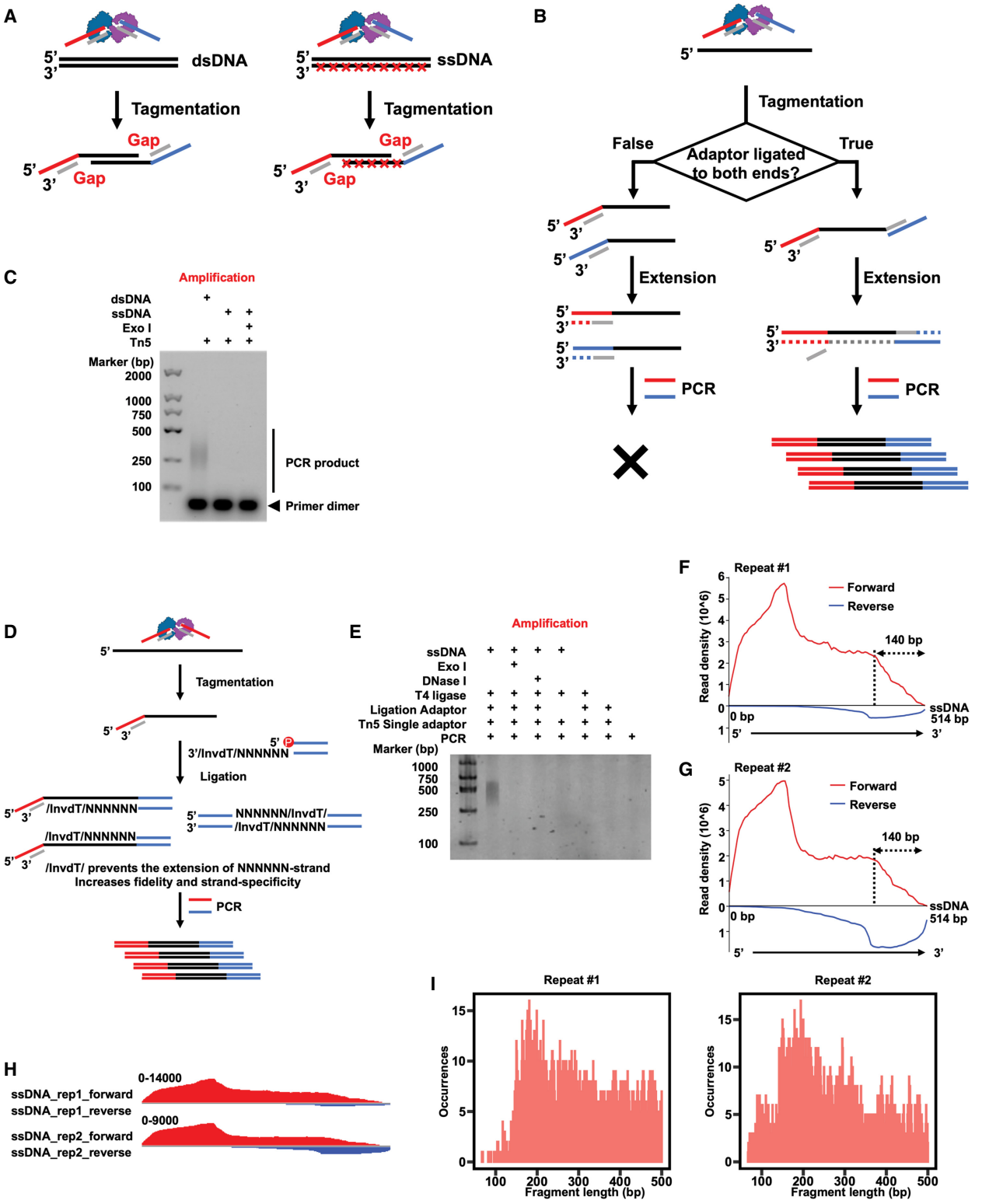


Figure 2. TABLE-seq-generated strand-specific sequencing library from ssDNA. (*A*) Illustration of how Tn5 transposon tagmented dsDNA and ssDNA. (*B*) Illustration showing the design for PCR analysis of oligo transposition in ssDNA. (*C*) PCR results presenting only tagmented dsDNA can be amplified. (*D*) Illustration showing the design for TABLE-seq. (*E*) Gel result presenting TABLE-seq method could amplify the tagmented ssDNA. (*F*,*G*) Profiles showing the distributions of reconstructed paired-end sequencing fragments forwardly and reversely mapped to the ssDNA. Two independent replicates from repeat #1 (*F*) and repeat #2 (*G*) are shown. The *x*-axis indicates the 5′-to-3′ direction of the template ssDNA. Forward, reconstructed paired-end sequencing reads mapped to the forward direction of ssDNA. Reverse, reconstructed paired-end sequencing reads mapped to the reverse direction of ssDNA. (*H*) IGV viewing presenting the forward and reverse reads mapped across the ssDNA. (*I*) The distribution of fragment length in ssDNA TABLE-seq. Two independent replicates from repeat #1 and repeat #2 are shown.

ssDNA sequencing (TABLE-seq). The PCR results showed that the ligation product could be amplified (Fig. 2E). Moreover, when we digested ssDNA with ExoI before tagmentation, no amplified products could be detected, further confirming the specific amplification of tagmented ssDNA but not the dsDNA contamination. When using samples treated with DNase I or without ssDNA as the starting material, the PCR results showed no detectable signal. To further confirm the ligation of transposon oligos to the 5′ end, we designed oligos that could be ligated to the 5′ end and performed the same PCR analysis (Supplemental Table S1). The results showed that only the oligos ligated to the 3′ end could be amplified (Supplemental Fig. S2D), further supporting the idea that the Tn5 transposons ligated oligos to the 5′ end of ssDNA.

To further verify the transposition of Tn5 oligos to ssDNA, we sent the amplified ssDNA from two biological repeats for high-throughput sequencing. The sequencing reads were mainly mapped to the forward direction of ssDNA (Supplemental Fig. S2E). In addition, we profiled the distribution of reconstructed paired-end sequencing fragment reads mapped throughout the ssDNA. The mapped reads were highly enriched in the forward direction, further indicating the strand specificity of the constructed libraries (Fig. 2F,G). More importantly, the read densities were higher at the 5′ end than at the 3′ end of the ssDNA. The strong total 5′ bias was owing to the decrease of tagmentation effect when the ssDNA was fragmented to a shorter length. It is possible that when ssDNA was tagmented to a shorter length, Tn5 could not transpose adaptors to the ssDNA so the signal increased from the 5′ end to the 3′ end and then dropped at the very end of the 3′ end. We further analyzed the point that led to a sharp drop in signals in the forward direction. The results showed that the read densities decreased largely at ~140 nt from the 3′ end of ssDNA, suggesting that the minimum length necessary for Tn5 tagmentation is 140 nt. The enrichment of reads was further reviewed by Integrative Genomics Viewer (IGV) views (Fig. 2H). In line with this observation, the occurrence of sequencing reads with a fragment length of <140 nt was much lower than that of others (Fig. 2I). Moreover, the transposition of ssDNA may be the consequence of transient dsDNA secondary structures. Our TABLE-seq method used a single-kind adapter-coupled Tn5, which transposed the same sequence of DNA to both ends of the dsDNA so the dsDNA, if tagmented by a single-kind adapter-coupled Tn5, could not be amplified by PCR. We then analyzed the amplification results of tagmented dsDNA and ssDNA by the TABLE-seq method. The mobility of ssDNA on PAGE gel is not the same as that of dsDNA, so the ssDNA ran slower than the dsDNA, showing a high molecular weight. The samples we used for the PCR were from the tagmented results, which were shown in lane 6. The dsDNA was fully tagmented, and the PCR products were likely from a nonspecific PCR result. The tagmented ssDNA but not dsDNA could be amplified, indicating the tagmentation was directly on ssDNA (Supplemental Fig. S2F). Altogether, these results suggest that Tn5 transposes oligos to the 5′ end of ssDNA and that the transposed ssDNA can be ligated with specific adaptors at the 3′ end for TABLE-seq-based high-throughput sequencing.

Tn5 can be used for strand-specific RNA sequencing

Because Tn5 tagmented and transposed oligos to ssDNA, we proposed that Tn5 was capable of constructing the strand-specific RNA sequencing library from single-strand cDNA, which omitted the second-strand synthesis.

Many different high-throughput RNA sequencing technologies have been developed to profile enriched RNA (Stark et al. 2019). Fragmented RNA, which was enriched for RNA modifications, protein-binding RNA, focused sets of RNA, etc., and full-length RNA, which was used to profile the whole transcriptome, were usually used as starting materials for RNA sequencing. We designed fragmented and unfragmented RNA to mimic the starting materials while constructing the RNA sequencing library (Fig. 3A). In addition, we used RNase A and RNase H to digest RNA and denatured the reverse-transcribed cDNA by heat to prevent the potential contamination of the RNA/DNA hybrid. The resulting single-strand cDNA was subjected to TABLE-seq, respectively. Because the tagmentation of Tn5 requires a minimum nucleic acid length, the ligation would increase the length of the ssDNA, which would potentially increase the efficiency of tagmentation. We ligated the ssDNA with adaptors and then tagmented the DNA to achieve the maximal recovery of reads. Both fragmented and unfragmented RNA were used to generate sequencing libraries (Fig. 3B). When the samples were treated with ExoI after RNase digestion and denaturation, the sequencing library could not be amplified, further showing that the amplified DNA was from single-strand cDNA.

To further analyze the quality of the constructed sequencing library, we sent the libraries for high-throughput sequencing. Two independent replicates from fragmented and unfragmented RNA were constructed and sequenced to increase the sequencing strength. These two replicates showed good correlations and were merged for further analyses (Supplemental Table S2). Sequencing results from fragmented RNA and unfragmented RNA showed high correlations (Fig. 3C). A similar number of genes (Fig. 3D) and similar gene expression profiles were detected (Supplemental Fig. S3). Additionally, the ratios of identified reads at exons and introns were similar between these two kinds of libraries (Fig. 3E).

To analyze the strand specificity of sequencing reads, we profiled sequencing reads at the sense and antisense strands spanning 2 kb of the transcription start site (TSS) of all genes (Fig. 3F). Reads mapped to the sense strand were highly enriched on the gene transcription side, and reads mapped to the antisense strand were enriched on the opposite side, indicating that the sequencing results were highly strand specific. The read profile of unfragmented RNA showed a preference for distal sites from TSS compared with that of fragmented RNA, which was consistent with the method that uses oligo(dT) to reverse-transcribe RNA from the distal sites from TSS. Together, these data suggest that TABLE-seq can use both fragmented and unfragmented RNA for strand-specific RNA sequencing.

The strand-specific RNA sequencing library can be constructed in a reproducible and time-saving manner

Subsequently, we tested different reaction conditions to improve the reproducibility and stability of strand-specific RNA-seq libraries.

First, we examined the effects of tagmentation temperature and time on the sequencing results. We synthesized single-strand cDNAs by performing reverse transcription with the total RNA extracted from mouse embryonic stem cells (mESCs). We previously found that when the DNA/RNA hybrid was treated at a high temperature and a prolonged reaction time, the mapped reads would be enriched in the introns (Sun et al. 2021). So, we used mild digestion conditions for the TABLE-seq. The RNase-treated and heat-denatured single-strand cDNA was then subjected to TABLE-seq, in which Tn5 tagmentation was conducted for 5 to 30 min at 16°C

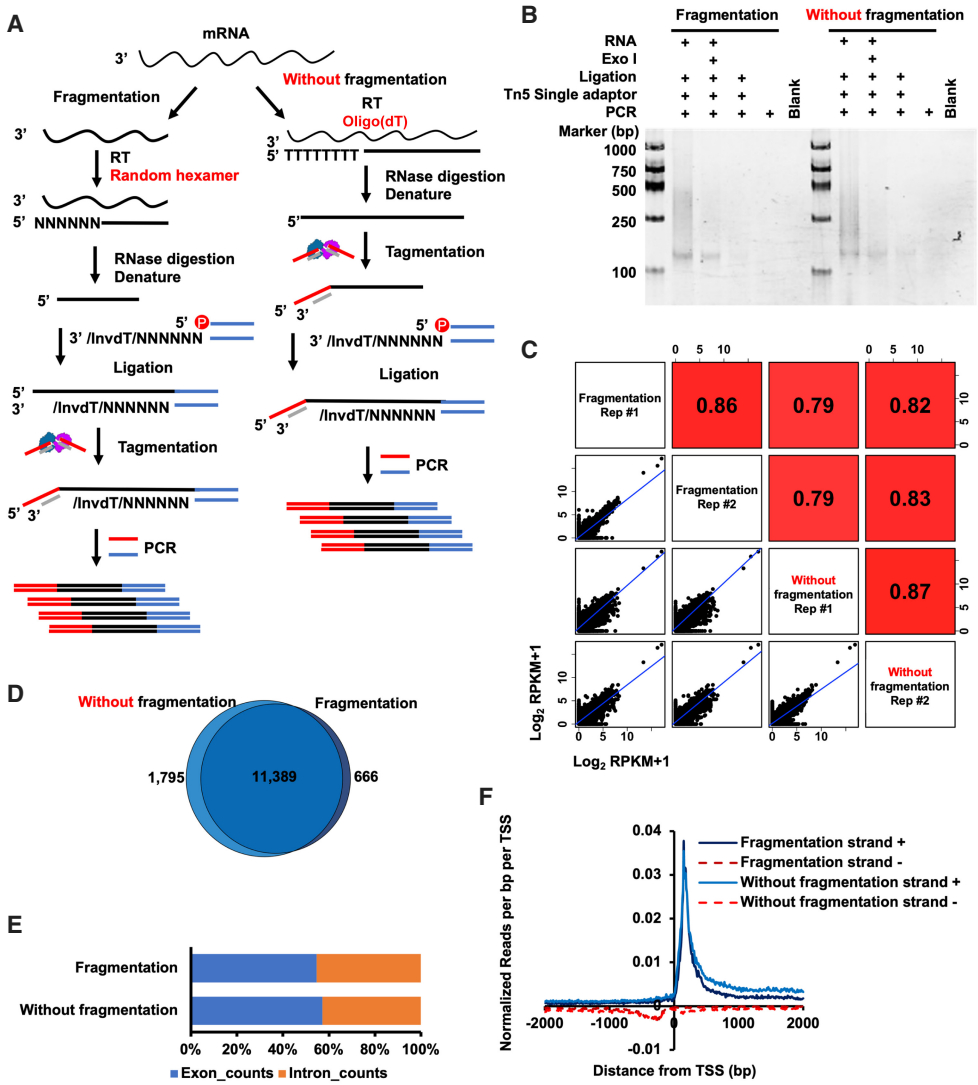


Figure 3. Tn5 is used for strand-specific RNA sequencing. (*A*) Illustration showing the strategy for strand-specific RNA sequencing. (*B*) Fragmented and unfragmented RNA were used to generate strand-specific RNA sequencing libraries. (*C*) The Pearson's correlations between the libraries generated from fragmented and unfragmented RNA. The scatter plots show all detected genes. (RPKM) Reads per kilobase of exon per million reads mapped. (*D*) Venn diagram showing the overall detected genes between libraries generated from fragmented and unfragmented RNA. (*E*) Bar graph showing ratios of detected reads at exons and introns. The number of reads mapped to exons or introns was calculated by HOMER. (*F*) Comparison of sequencing signals between libraries generated from fragmented and unfragmented RNA. Sense (strand +) and antisense (strand −) transcripts associated with transcription start site (TSS) are shown.

and 37°C. With increased temperature and prolonged incubation time, the fragments in the libraries became shorter (Fig. 4A). Two independent replicates were constructed and sequenced to increase the sequencing strength. These two replicates showed good correlations and were merged for further analyses (Supplemental Table S2). Strong correlations were observed among each library, whereas tagmentation for 5 min at 16°C showed the lowest correlation with other tagmentation conditions (Fig. 4B). A similar number of genes and similar gene expression profiles were detected among these libraries (Fig. 4C,D; Supplemental Fig. S4A). The ratios of detected reads at exons compared with those at introns increased from 5 min to 30 min when tagmented at 16°C and remained similar when tagmented at 37°C (Supplemental Fig. S4B). In addition, we detected strong strand specificities in all the libraries. A lower temperature and shorter time generated higher signals close to TSS (Fig. 4E). It is possible that the lower temperature and shorter time decreased the enzymatic activities of Tn5, leading to a preference for tagmenting shorter cDNAs. To save experimental time, prevent a loss of reads in exons, and recover distal cDNA from TSS, we chose to perform tagmentation for 5 min at 37°C to conduct the subsequent experiments.

Next, we assessed how the amount of Tn5 affected the sequencing results of strand-specific RNA-seq. Single-strand cDNA was synthesized from 500 ng total RNA and then subjected to TABLE-seq with 25, 50, and 100 ng/µL Tn5 treatment for 5 min at 37°C. We found that higher amounts of Tn5 slightly reduced the length of the fragmented cDNA (Fig. 4F). Two independent replicate libraries were then constructed and sequenced to increase the sequencing strength. These two replicates showed good correlations and were merged for further analyses (Supplemental Table

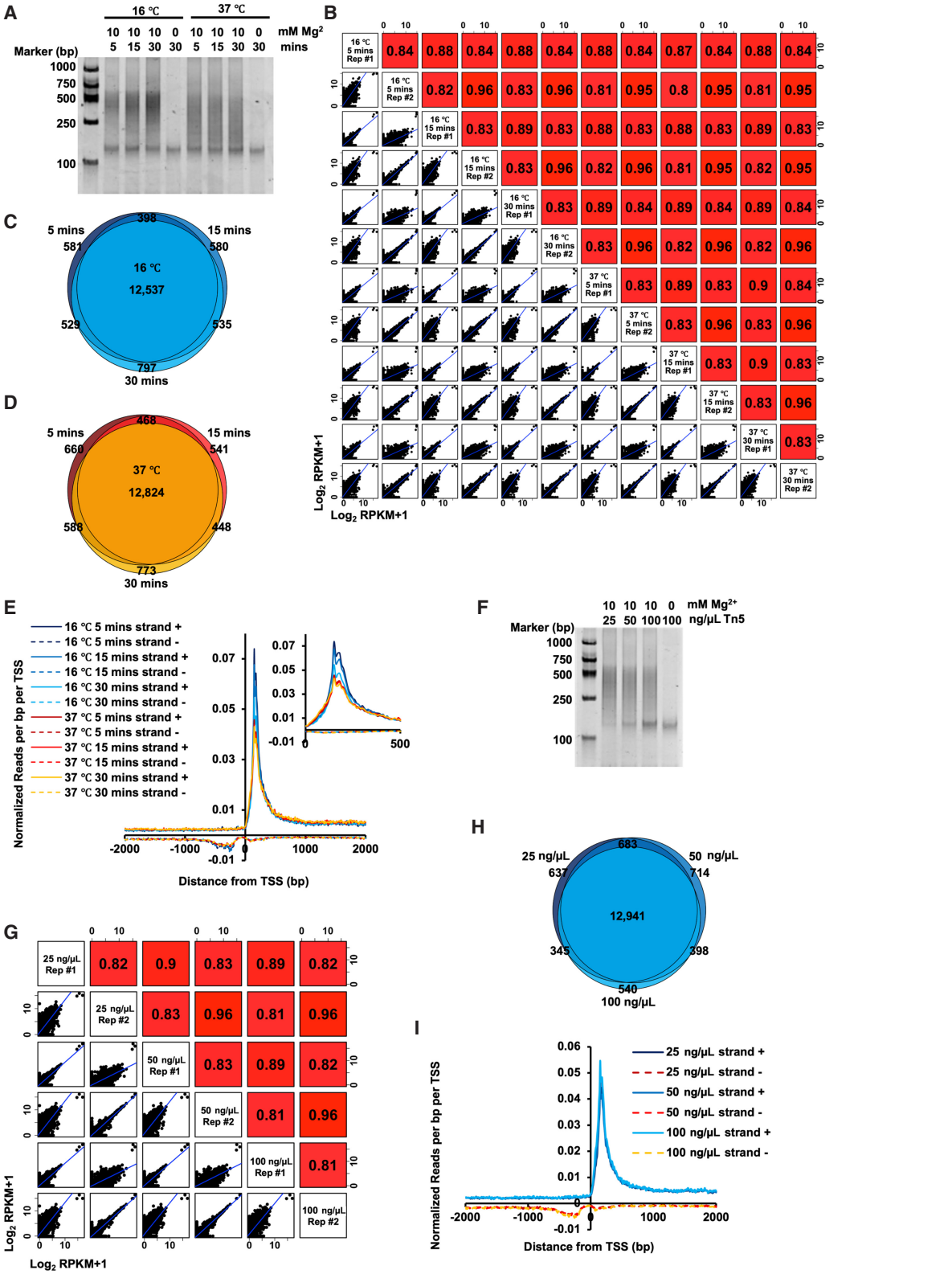


Figure 4. Analyses of tagmentation conditions in strand-specific RNA sequencing. (*A*) PCR results of TABLE-seq-constructed libraries at the indicated time and temperature are shown. cDNAs were tagmented at the indicated time and temperature, PCR amplified, and analyzed by TBE gel. Mg^{2+} was required for the enzymatic activity of Tn5 transposon. Samples treated without Mg^{2+}, which inhibited the Tn5 transposon activity, were used as negative controls. (*B*) The Pearson's correlations among libraries generated by the indicated reaction time and temperature. The scatter plots show all detected genes. (RPKM) Reads per kilobase of exon per million reads mapped. (*C*) Venn diagram showing the overall detected genes among libraries generated by reactions at 16°C with the indicated reaction time. (*D*) Same as *C*, except libraries generated by reactions at 37°C are shown. (*E*) Comparison of sequencing signals among libraries generated by the indicated time and temperature. Sense (strand +) and antisense (strand −) transcripts associated with TSS are shown. The distributions at the first 500 bp are enlarged. (*F*) PCR results of TABLE-seq-constructed libraries by different amounts of Tn5 are shown. cDNAs were tagmented with the indicated amounts of Tn5, PCR amplified, and analyzed by TBE gel. Mg^{2+} was required for the enzymatic activity of Tn5 transposon. Samples treated without Mg^{2+}, which inhibited the Tn5 transposon activity, were used as negative controls. (*G*) The Pearson's correlations among libraries generated by the indicated amounts of Tn5. The scatter plots show all detected genes. (*H*) Venn diagram showing the overall detected genes among libraries generated by the indicated amounts of Tn5. (*I*) Comparison of sequencing signals between libraries generated by the indicated amounts of Tn5. Sense (strand +) and antisense (strand −) transcripts associated with TSS are shown.

S2). The correlations between the sequencing results from different concentrations of Tn5 were high (Fig. 4G). The numbers of detected genes, gene expression profiles, and the ratios of detected reads at exons compared with those at introns were similar among the tested libraries (Fig. 4H; Supplemental Fig. S4C,D). We also noticed a slight batch effect when different amounts of Tn5 were used (Supplemental Fig. S4C). More importantly, we detected similar strand-specific gene expression in all the sequenced libraries (Fig. 4I). These data suggest that during the tagmentation of single-strand cDNA, the concentration of Tn5, ranging from 25 to 100 ng/μL, does not affect the sequencing results. Together, we used Tn5 to construct the strand-specific RNA sequencing library through TABLE-seq.

TABLE-seq shows higher sequencing power than that of canonical dUTP-based strand-specific RNA sequencing

To evaluate how the strand-specific RNA sequencing library constructed by TABLE-seq is comparable with that constructed by a widely used method, we benchmarked the TABLE-seq-based and canonical dUTP–based methods (Stark et al. 2019). cDNA was reverse-transcribed from 500 ng of total RNA and split for TABLE-seq-based and dUTP-based strand-specific RNA sequencing. Two independent replicates with high correlations were performed and merged for further analyses (Supplemental Table S2). The dUTP-based method showed a higher ratio of reads mapped to rRNA (~18%) than the TABLE-seq-based method (~8%). It is possible that the dUTP-based method required a second-strand synthesis step, which increased the enrichment of abundant rRNA. To reduce the impact of rRNA read contamination and avoid the potential impact of sequencing depth on the benchmark analysis, we first excluded the rRNA reads and then randomized 5 million sequencing reads from each sequencing result to perform the comparison.

The duplication rate of reads was similar between the TABLE-seq-based and dUTP-based sequencing results (Supplemental Fig. S5A). To further compare the similarity between sequencing repeats, we analyzed the coefficient of variation between each gene in two replicates. The TABLE-seq-based method showed a similar coefficient to that in the dUTP-based method (Supplemental Fig. S5B). Gene expression profiles and ratios of detected reads at exons compared with those at introns were similar between the TABLE-seq-based and canonical dUTP-based methods (Supplemental Fig. S5C,D). There were slightly more genes detected by the TABLE-seq-based method than by the dUTP-based method (Fig. 5A). The GC contents between these two methods were similar (Fig. 5B). To further analyze the detected gene number by sequencing reads, we calculated the saturation curves using randomly gradient sampled reads. At about 1 million total reads, the two analyzed methods reached a plateau of saturation, and the TABLE-seq-based method resulted in a higher detected gene number when the same amount of sequencing reads was used (Fig. 5C). In addition, the complexity of the TABLE-seq-based method was higher than that of the dUTP-based method (Fig. 5D). Moreover, we calculated the RPKM per gene. The median PRKM per gene in the TABLE-seq-based method was similar (Fig. 5E).

We then analyzed the strand specificities of these two tested methods. Strand specificities of both methods were detected in read profiles and IGV views (Fig. 5F,G). Most sequencing reads were mapped to the sense strand of cDNA (Fig. 5G). To further analyze the coverage of sequencing reads spanning the genes, we calculated the read densities in detected genes. Read distributions spanning the gene region were evenly distributed in the TABLE-seq-based method, indicating that this method recovered genes without 5′ or 3′ end preference (Fig. 5H). Consistent with previous observations (Agarwal et al. 2015), the sequencing reads were biasedly enriched at the 5′ end in the dUTP-based method. Moreover, to benchmark our sequence results with published strand-specific RNA sequencing, we compared the sequencing quality with the data set (obtained from the NCBI Gene Expression Omnibus [GEO; https://www.ncbi.nlm.nih.gov/geo/] under accession number GSE29278) that used rRNA depletion before constructing the strand-specific RNA sequencing library in mESCs (Shen et al. 2012). The mapping ratio, reads mapped to mRNA, and reads mapped to the right direction were similar among TABLE-seq-based, dUTP-based, and rRNA depletion–based methods (Supplemental Fig. S5E). As expected, the reads mapped to rRNA were lower with rRNA depletion. The reads distributions around TSS regions were similar between TABLE-seq-based and rRNA depletion–based methods (Supplemental Fig. S5F). The transcription reads were also close to the TSS sites. We speculated that the enrichments of reads at TSS were caused by the higher expression of short genes in mESCs. To test this idea, we calculated the expression levels of genes with different lengths. Indeed, the results showed that the expression levels of shorter genes were higher than those of longer genes (Supplemental Fig. S5G). Previous results have shown that the tagmentation of dsDNA has a specific bias in the preferential bases (Gertz et al. 2012). We also analyzed bases usage in the first 20 bases of dsDNA tagmentation (Picelli et al. 2014), TABLE-seq, cDNA/RNA hybrid tagmentation (Di et al. 2020), dUTP-based sequencing, and rRNA depletion–based sequencing (Supplemental Fig. S5H; Shen et al. 2012). The Tn5-based double-strand nucleic acid tagmentation showed a preference for dG at the first, 11th, and 13th bases; dA/T preference at the second to eighth bases; dC preference at the ninth base; and dT preference at the 10th base. The TABLE-seq showed a dG preference at the first 5 bp, which were different from the double-strand nucleic acid tagmentation. Together, these data suggest that the TABLE-seq-based strand-specific RNA sequencing method has a comparable, and even higher, sequencing power than that of the dUTP-based method.

30 pg of RNA is sufficient to construct a strand-specific RNA sequencing library by TABLE-seq

Because we constructed strand-specific RNA sequencing libraries through TABLE-seq, we further evaluated the minimum amount of total RNA necessary to construct the sequencing libraries. To construct the libraries, we used a series of 10-fold dilutions starting from 300 ng of total RNA as the initiating materials. In addition, we did not scale the reagents during library preparation to avoid the influence of other potential factors. Total RNA as low as 30 pg, which was equal to the total RNA from a single mESC (van den Hurk et al. 2018), was sufficient to be amplified by PCR after TABLE-seq library construction (Fig. 6A).

To analyze how the amount of RNA affected the sequencing results, we sent all the samples for sequencing. Two independent replicates were conducted and showed high correlations (Supplemental Table S2). We then merged these replicates for further analysis. The library constructed from 30 pg of RNA showed the lowest correlation with other libraries (Fig. 6B,C; Supplemental Fig. S6A). The correlations among other sequencing results were high, and the ratios of reads mapped to exons and introns were similar. A total of 12,879, 10,739, 9387, 2773, and 2020 genes were

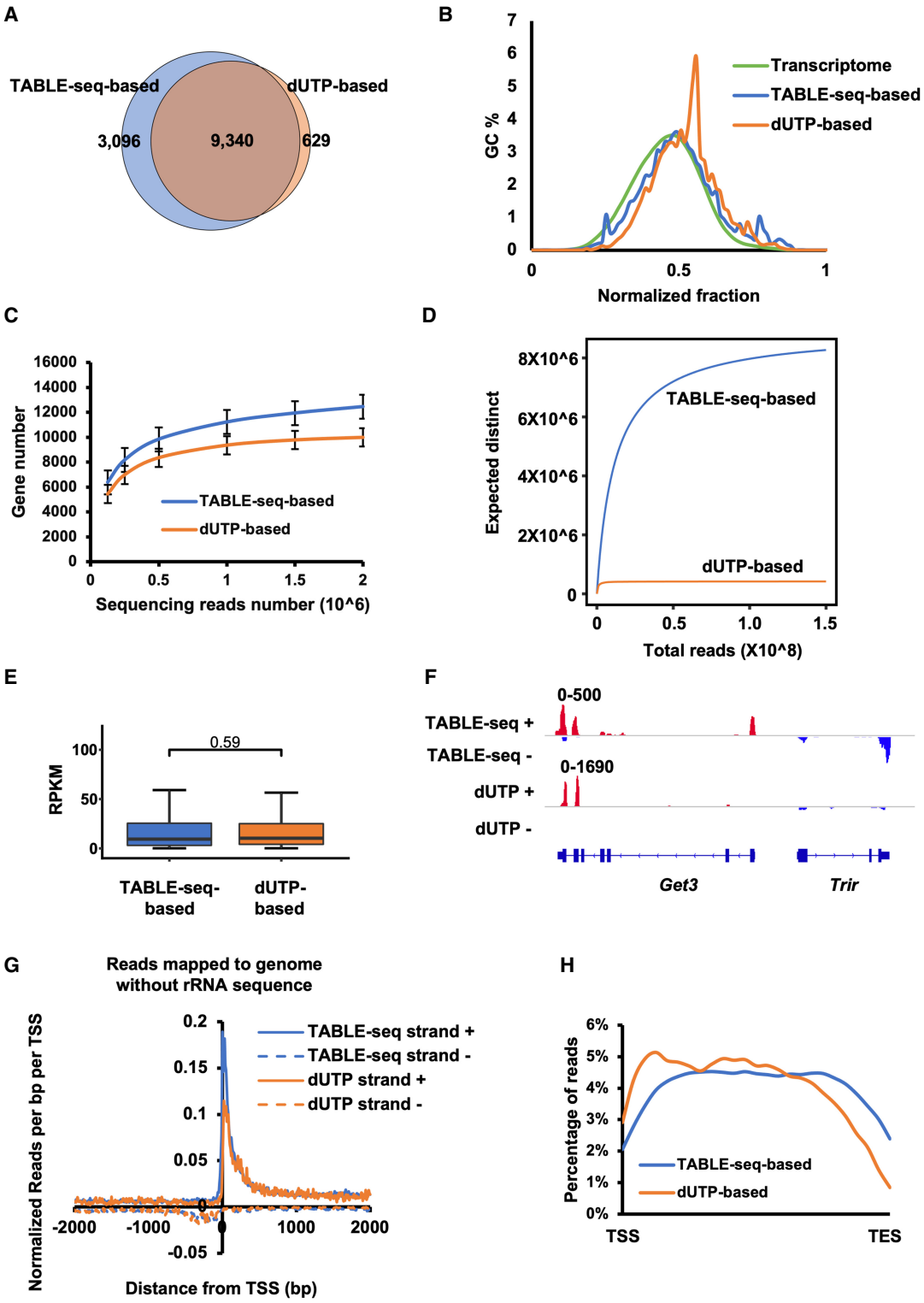


Figure 5. Benchmark between TABLE-seq-based and dUTP-based strand-specific RNA sequencing. (*A*) Venn diagram showing the overall detected genes between libraries generated by TABLE-seq-based and dUTP-based methods. (*B*) Percentage of GC content in TABLE-seq-based and dUTP-based methods. (*C*) TABLE-seq-based and dUTP-based methods shared a plateau of saturation at around 1 million total reads. Saturation curves were constructed with random gradients. N = 2 M, 1.5 M, 1 M, 0.5 M, and 0.25 M reads from total reads. Data were mean ± SD. (*D*) The library complexity of TABLE-seq-based and dUTP-based methods. (*E*) Boxplot showing the read counts per gene in libraries generated by the indicated methods. (*F*) IGV views showing the enrichments of mapped reads at the sense (strand +) and antisense (strand −) transcripts. (*G*) Comparison of sequencing signals between libraries generated by the indicated methods. Reads were mapped to the genome without ribosomal RNA sequences. Sense (strand +) and antisense (strand −) transcripts associated with TSS are shown. (*H*) The reads distributions spanning the gene regions. (TES) Transcription end site.

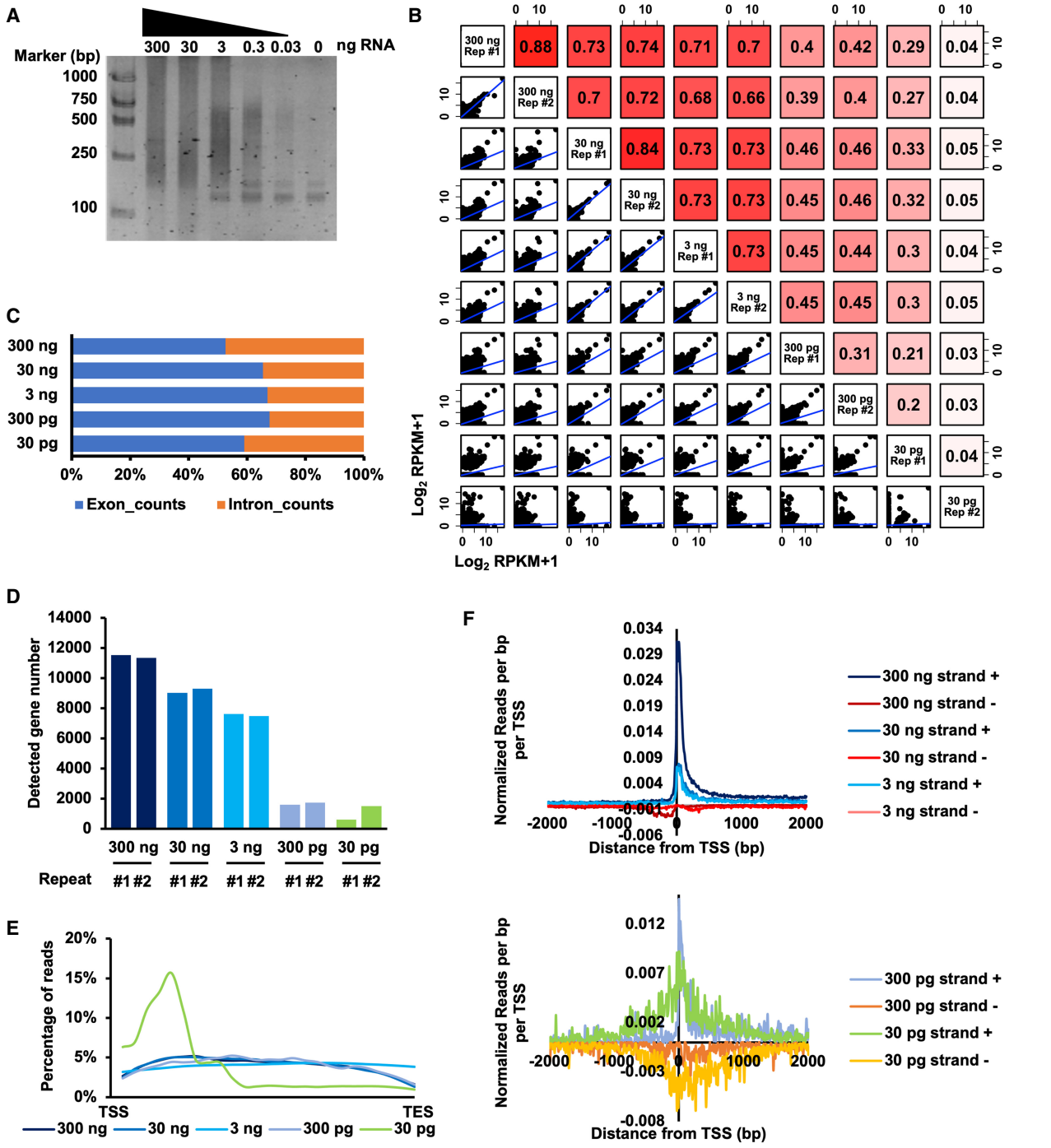


Figure 6. The 30 pg of RNA is sufficient to construct a strand-specific RNA sequencing library. (*A*) PCR results of libraries constructed from different amounts of RNA are shown. (*B*) The Pearson's correlations among libraries generated by the indicated amounts of RNA. The scatter plots show all detected genes. (RPKM) Reads per kilobase of exon per million reads mapped. (*C*) Bar graph showing ratios of detected reads at exons and introns among the libraries generated by the indicated amounts of RNA. (*D*) Bar graph showing the overall detected genes among libraries generated by the indicated amounts of RNA. (*E*) The reads distributions spanning the gene regions among libraries generated by the indicated amounts of RNA. (TES) Transcription end site. (*F*) Comparison of sequencing signals between libraries generated by the indicated amounts of RNA. Sense (strand +) and antisense (strand –) transcripts associated with TSS are shown.

detected from 300 ng, 30 ng, 3 ng, 300 pg, and 30 pg of total RNA, respectively (Fig. 6D). Fewer genes were identified when less RNA was used. The TABLE-seq-based method detected approximately 2000 genes from the total RNA similar to one single mESC (van den Hurk et al. 2018). In addition, the reads enrichments were distributed evenly across genes except for the 30 pg, which showed a higher enrichment at the 5′ end (Fig. 6E). The strand specificity gradually decreased when less RNA was used for sequencing.

When the amount of total RNA decreased to 30 pg, we could still detect a high enrichment (although the amount was less than those identified from a large amount of RNA) of reads mapped to the sense strand compared with reads mapped to the antisense strand (Fig. 6F; Supplemental Fig. S6B). These data suggest that as little as 30 pg of RNA, which is similar to the amount of RNA from a single mESC (van den Hurk et al. 2018), is sufficient to construct strand-specific RNA sequencing libraries by TABLE-seq.

Discussion

Acting as natural DNA transfer vehicles, transposons are widely used as genetic tools in many aspects of life science research, including transgenesis of genetic elements into cells, transgenic genetic screening, and mouse generation (Bourque et al. 2018). The most widely used transposon systems are *Sleeping Beauty*, *piggyBac*, and *Tol2*. Because transposons tagment and transpose designed adaptors into target nucleic acids, resulting in PCR-ready DNA fragments for high-throughput sequencing, transposon-dependent library construction strategies draw increasing attention to the increased need for omics data (Hennig et al. 2018). Here, we explored the new molecular properties of Tn5 transposons in substrate recognition. Tn5 can tagment and transpose the designed adaptor into the 5′ end of ssDNA. Using this feature, we designed a high-throughput sequencing library construction strategy, TABLE-seq, to apply Tn5 as a new methodological tool. We applied TABLE-seq to strand-specific RNA sequencing, which shows similar, or even higher, sequencing power compared with that of the canonical dUTP-based method. This TABLE-seq strategy is suitable for any applications using ssDNA and RNA as the starting materials, including eSPAN (Li et al. 2020), ReIN-Map (Xu et al. 2021), DNA bisulfite sequencing (Gu et al. 2011), Okazaki fragment mapping (Kit Leng Lui et al. 2021), and GRO-seq (Lopes et al. 2017).

Because the relocation of transposon elements can cause mutations in the genome, which subsequently leads to genome instability, the transposition activity of Tn5 transposons is very weak in the cells (Cordaux and Batzer 2009). In vitro, wild-type Tn5 has no transposition effect, likely owing to the inhibitory effect of the N-terminal domain. Luckily, a hyperactive mutation of Tn5 is generated and shows strong tagmentation activity in vitro (Goryshin and Reznikoff 1998). Tn5 transposons are assembled with Tn5 transposase and OE sequences to facilitate a direct tagmentation of target nucleic acids. Unlike the in vivo system in which Tn5 needs to bind with OE sequences to "cut" the transposon sequence out of the genome, Tn5 is preloaded with OE sequences to attack the target nucleic acids in vitro. This preloading of OE sequences may increase the diversity of nucleic acid recognition that is tagmented.

Tn5 transposases bind to OEs to form a synaptic transposon complex by dimerizing two Tn5 transposases through the C-terminal domain. Then, Tn5 transposases use water molecules to catalyze the hydrolysis of the phosphodiester bond of DNA, forming a nucleophilic hydroxyl group at the 5′ ends of DNA. The hydroxyl group attacks the complementary DNA strand to form a hairpin structure, which is further activated by another water molecule to create a blunt end (Davies et al. 1999, 2000; Steiniger-White et al. 2004). Through this process, Tn5 transposons bind to the target DNA sequence and transfer the transposon DNA strands into both strands, leaving a 9-bp gap at the 3′ end of targeting sites. Through this "cut-and-paste" process, Tn5 transposons "jump" from one location to other locations (Zhou and Reznikoff 1997; Reznikoff 2003).

Tn5 is able to tagment a RNA/cDNA hybrid, leading to a transposition of oligos to both 5′ and 3′ ends to generate a PCR-ready sequence (Di et al. 2020). Using this property, SHERRY, which reduces preparation steps and costs during library preparation, is conducted for RNA sequencing. The SHERRY is based on direct tagmentation of RNA/cDNA hybrid; the information on strand specificity would be lost by the transposition of transposon adaptors at both the 5′ and 3′ ends. With some modifications, such as denaturing the tagmented RNA/cDNA hybrid, digesting the RNA strand, or adding a ligation step, SHERRY can be used to generate strand-specific RNA sequencing libraries. Compared with SHERRY, the TABLE-seq is designed to generate a strand-specific RNA sequencing library by tagmentation on the cDNA. Moreover, TABLE-seq can be used for ssDNA sequencing in vitro and probably in vivo by coupling Tn5 with protein A, which is similar to CUT&Tag (Carter et al. 2019).

We do not perform rRNA depletion or poly(A) enrichment before strand-specific RNA sequencing because we intend to perform TABLE-seq on single cells. The depletion of rRNA by oligo hybridization and the enrichment of poly(A) by oligo(dT) binding will result in RNA loss before reverse transcription, especially when the input amount of RNA is low. Oligo(dT) was used in reverse transcription in SMART-Seq3 to enrich poly(A)-containing RNA (Hagemann-Jensen et al. 2020). To reduce rRNA contamination, we also use oligo(dT) in reverse transcription to mimic the SMART-Seq3 method. When performing bulk strand-specific RNA-seq, rRNA depletion or poly(A) enrichment would increase the detected mRNA/rRNA ratio and the number of genes.

When a very low amount of RNA is used to construct a strand-specific RNA sequencing library, the strand specificity decreases a lot (Fig. 6E; Supplemental Fig. S6B). One possibility for this decrease in strand specificity is library preparation. The commercially used library preparation primers contain a short DNA sequence of OEs that is essential for the tagmentation of Tn5 (Supplemental Table S1). These sequences in the forward and reverse PCR primers are reverse complementary to each other. When a low amount of template DNA is used for the PCR amplification, the specificity of PCR primers at either end may decrease, leading to the complementary use of PCR primers at the opposite end of the template DNA. The other possibility for this decrease in strand specificity is reverse transcription. The reverse transcriptase uses RNA as the temple to synthesize first-strand cDNA for strand-specific sequencing. However, the reverse transcriptase has the minus-strand transfer property, which will use synthesized first-strand cDNA as the template to synthesize the second-strand cDNA (Svarovskaia et al. 2000). So, actinomycin D is used to inhibit the minus-strand transfer step in reverse transcription (Perocchi et al. 2007). When a low amount of input RNA is used, the relative ratio of reverse transcriptase to RNA has to be high enough to catalyze a successful reverse transcription. It is possible that the minus-strand transfer effect of reverse transcriptase is not fully inhibited, leading to the generation of second-strand cDNA and subsequent loss of strand specificity.

To keep the single variable, we do not change the reagents we use, such as Tn5, when we compare TABLE-seq performance on different amounts of input RNA. However, if a low amount of input RNA is used, the amount of Tn5 could be systematically adjusted and controlled to get a maximal recovery of cDNAs. In the future, a series of condition tests can be conducted to increase the sequencing power of TABLE-seq during low-input RNA experiments. In addition, the enzymatic activity should be tightly controlled during the preparation and storage of Tn5 to get stable

results from batch-to-batch experiments (Wu et al. 2021; Wei et al. 2022; Zhang et al. 2022).

In conventional strand-specific RNA sequencing experiments, dUTP is usually incorporated during the second-strand cDNA synthesis (Zhang et al. 2012). After the ligation of PCR adaptor, this dUTP-containing second-strand cDNA is digested by UDGase treatment, leaving an ssDNA that is amplified by strand direction information. In the TABLE-seq-based RNA-seq method, the fragmentation and second-strand cDNA synthesis steps are omitted to reduce the time of library preparation and prevent the biased enrichment of cDNA. Furthermore, total RNA is reverse-transcribed by oligo(dT), which reduces the amount of rRNA synthesized. In TABLE-seq, adaptors at the 3′ end of ssDNA are ligated by T4 DNA ligase and a splint DNA oligo. Other methods can be further tested to increase the ligation efficiency, for example, using other ligation enzymes, using a DNA/RNA hybrid with a 3′ protruding end that can be ligated with the Splint R enzyme, using ssDNA as adaptors to be directly ligated to tagmented fragments, and adding polyethylene glycol (PEG) into the ligation reaction. We speculate that with the improvement in ligation efficiency, Tn5 could be used to recover more genes from a low amount of total RNA. Because Tn5 transposon tagments dsDNA, ssDNA, and DNA/RNA hybrids, it would also be interesting to evaluate the substrate preference of Tn5.

In cells, ssDNA is frequently generated to serve as a template for various cellular processes (Ashton et al. 2013). Naked ssDNA, if exposed without the protection of proteins, is harmful to cells. Compared with dsDNA, ssDNA is more sensitive to nucleolytic and chemical radicals, which are common reactors causing mutations in the genome (Maréchal and Zou 2015). In addition, the thermodynamic stability of DNA decreases with unwinding, leading to the spontaneous development of duplex secondary structures that inhibit DNA replication. Cells bypass the potential damage from ssDNA by expressing ssDNA-binding proteins that stabilize ssDNA and recruit other partner proteins for DNA repair, recombination, and replication. During DNA replication, the lagging strand is synthesized discontinuously and is composed of multiple fragments, the Okazaki fragments, which generate ssDNA at the replication bubble (Kit Leng Lui et al. 2021). It will be interesting to apply TABLE-seq to profile the enrichment in ssDNA and ssDNA-binding proteins.

Methods

Cell cultures

E14 mESCs were cultured on 0.1% gelatin-coated dishes in DMEM supplemented with 15% fetal bovine serum (FBS), 1% antibiotic solution (penicillin/streptomycin), 1% GlutaMAX, 1% MEM nonessential amino acids, 1% sodium pyruvate, 0.1 mM beta-mercaptoethanol, and 1000 U/mL recombinant leukemia inhibitory factor (LIF). HEK293T cells were cultured in DMEM supplemented with 10% FBS and 1% antibiotic solution (penicillin/streptomycin). All cells were grown at 37°C and 5% CO_2.

Oligos

Oligos were listed in Supplemental Table S1.

Asymmetric PCR

Asymmetric PCR was used to generate ssDNA (Heiat et al. 2017). The Tsin-H3.1-flag plasmid was used as the template. The forward primer and reverse primer amounts were used at a 1:10 ratio. One microliter of 10 ng/μL plasmid, 1 μL 1 μM forward primer, 1 μL 10 μM reverse primer, 12.5 μL 2× HiFi PCR master mix (NEB M0541L), and 10.5 μL H_2O were mixed and incubated for 5 min at 95°C; 34 cycles for 30 sec at 95°C, for 30 sec at 55°C, and for 30 sec at 72°C; then 10 min at 72°C; and 16°C to hold. To separate dsDNA and ssDNA from amplification products, PCR samples were subjected to agarose gel electrophoresis. The bands corresponding to dsDNA and ssDNA were cut from the gel and purified by the DNA gel purification kit (Tiangen DP209-03), respectively. To synthesize different lengths of ssDNA, forward primers' annealing positions were designed accordingly, whereas the same reverse primer was used.

RNA extraction

RNA extraction was performed by RNeasy plus mini kit (Qiagen 74134). In brief, 10^6 mESCs were harvested and washed with PBS once. Cells were lysed with 350 μL buffer RLT and vortexed for 30 sec. Then the lysates were added to a gDNA eliminator spin column and centrifuged at 10,000 rpm for 30 sec. The flow-through was mixed with 350 μL of 70% ethanol before being transferred to an RNeasy spin column. Samples were bound to the column by centrifuging at 10,000 rpm for 15 sec. The RNA-bound column was washed with 700 μL buffer RW1 once and 500 μL buffer RPE twice. RNA was eluted with 30 μL of RNase-free H_2O. Extracted RNA was treated with 1 μL of DNase I (NEB M0303S) for 1 h at 37°C to further eliminate genomic DNA. Treated RNA was precipitated by 75% ethanol and dissolved in RNase-free H_2O. The purified RNA was stored at −80°C until usage.

Reverse transcription

Reverse transcription was performed with SuperScript IV kit (Thermo Fisher Scientific 18091050). Different amounts of RNA were annealed with 1 μL 50 μM oligo$(dT)_{20}$ and 1 μL 10 mM dNTP for 5 min at 65°C and on ice immediately for 2 min. Then each sample was mixed with 4 μL 5× SSIV buffer, 1 μL 100 mM DTT, 1 μL ribonuclease inhibitor, and 1 μL SuperScript reverse transcriptase. One microliter of 100 μg/mL actinomycin D was also added to inhibit the minus-strand transfer step in the reverse transcription. H_2O was added up to a total of 20 μL volume. Samples were incubated for 90 min at 42°C; 10 cycles of 2 min at 50°C and 2 min at 42°C; then 5 min at 85°C, and 16°C to hold. After reverse transcription, 1 μL ExoI (NEB M0293S), 3 μL 10× reaction buffer, and 6 μL H_2O were added to digest oligo$(dT)_{20}$ primers for 30 min at 37°C. After inactivation of ExoI for 15 min at 80°C, 1 μL RNase A (Takara 2158) and 1 μL RNase H (Thermo Fisher Scientific 18091050) were added to digest RNA for 30 min at 37°C.

Tn5 transposon assembly and treatment

Tn5 transposase was purified as described previously (Picelli et al. 2014; Hennig et al. 2018). One microliter of 100 μM Tn5ME-A or Tn5ME-B and 1 μL 100 μM Tn5MErev were added to 8 μL H_2O and annealed as Tn5-A or Tn5-B, respectively. Ten microliters of Tn5-A and 10 μL Tn5-B were incubated with 70 μL of 200 μg/mL Tn5 transposase for 1 h at 25°C to assemble the dual-adaptor Tn5 transposon, whereas 20 μL Tn5-B was incubated with 70 μL of 200 μg/mL Tn5 transposase to assemble the single-adaptor Tn5 transposon. The assembled transposons were stored at −20°C until usage. Nucleic acids were treated as indicated in the paper, and then the treated samples were run by 6× gel loading dye (NEB B7024S), which contained 0.48% SDS, to dissociate Tn5 with DNA.

TABLE-seq

A step-by-step protocol was deposited on protocols.io (https://www.protocols.io/view/table-seq-for-strand-specific-rna-sequencing-8epv5jrw6l1b/v1) and as Supplemental Protocol.

Fluorescence labeling of ssDNA and transposon adaptors

5′ Alexa Fluor 594–labeled ssDNA was generated by asymmetric PCR with Alexa Fluor 594–labeled forward primer (Supplemental Table S1). 3′ Alexa Fluor 594–labeled ssDNA was synthesized by terminal deoxynucleotidyl transferase. In brief, 1 µM ssDNA was labeled with a TUNEL assay kit (fluorescence, 594 nm) in a 50-µL reaction for 30 min at 37°C. To extract DNA, 52.5 µL chloroform/isoamyl alcohol (24:1) was added to the reaction, vortexed, and centrifuged at 12,000 rpm for 2 min. Samples were then purified by a MinElute PCR purification kit (Qiagen 28004). Transposons with Alexa Fluor 488–labeled adaptors were assembled as follows: 1 µL 100 µM Tn5ME-A-5′AF488 and 1 µL 100 µM Tn5MErev were added to 8 µL H_2O and annealed as Tn5-A-488. Twenty microliters of Tn5-A-488 was incubated with 70 µL of 200 µg/mL Tn5 transposase for 1 h at 25°C.

Tagmentation of cDNA

The reverse-transcribed cDNA, which was treated by RNase A and RNase H, was denatured for 10 min at 98°C and on ice immediately for 5 min. Fourteen microliters of cDNA was mixed with 4 µL 5× DMF buffer (50% DMF, 50 mM Tris-HCl at pH 7.5, 10 mM $MgCl_2$) and 2 µL transposon Tn5 A + B or Tn5 B + B, respectively. Tagmentation reactions were incubated for 5 min at 37°C and stopped by adding stop buffer (1.5 µL 0.5 M EDTA, 1.8 µL 10% SDS, and 0.5 µL 20 mg/mL Proteinase K) and incubating for 30 min at 55°C and then for 20 min at 70°C. The tagmented DNA was purified by 0.9× of VAHTS DNA clean beads (VAHTS N411-03) and eluted into 10 µL H_2O.

TABLE-seq library construction

Two microliters of 100 µM 5ph_Tn5a and 2 µL 100 µM Tn5a_N6_invert_dT were added to 16 µL H_2O. Samples were annealed for 5 min at 95°C and ramped to 25°C at 5°C/min. One microliter of ExoI, 3 µL 10× ExoI reaction buffer, and 6 µL H_2O were added and incubated for 30 min at 37°C followed by 15 min at 80°C. The annealed adaptors were purified by 2× of VAHTS DNA clean beads, eluted into 20 µL H_2O, and stored at −20°C until usage.

Ten microliters of tagmented ssDNA or cDNA was mixed with 2 µL annealed adaptors, 0.5 µL T4 DNA ligase (Takara 2011A), 2 µL 10× ligation buffer, 2.5 µL 40% PEG6000, and 3 µL H_2O and incubated for 1 h at 20°C. The ligated DNA was purified by 0.8× VAHTS DNA clean beads and eluted into 10 µL H_2O. The purified DNA was amplified by primers with sequencing indexes. One microliter of 10 µM forward primer, 1 µL 10 µM reverse primer, 8 µL DNA, and 10 µL 2× HiFi PCR master mix (NEB M0541) were mixed and amplified for 5 min at 72°C, for 30 sec at 98°C, for 20 cycles of 10 sec at 98°C and 10 sec at 63°C, then 1 min at 72°C, and 16°C to hold. PCR products were purified by 0.9× VAHTS DNA clean beads and eluted into 20 µL H_2O. The concentrations of DNA were detected by Equalbit 1× dsDNA HS working solution (Vazyme 121-01-AA). The libraries were sequenced on an Illumina NovaSeq platform with pair-end reads of 150 bp.

dUTP-based library construction

First-strand cDNA was synthesized as described previously in reverse transcription. Starting from second-strand cDNA synthesized, the RNA-seq library prep kit (VAHTS NR604) was used under the manufacturer's instructions. Twenty-five microliters of first-strand cDNA was mixed with 25 µL second-strand synthesis buffer with dUTP and 15 µL second-strand synthesis enzyme super mix. Samples were incubated for 30 min at 16°C followed by 15 min at 65°C. Adaptors were then ligated to dsDNA by using 25 µL rapid ligation buffer, 5 µL rapid DNA ligase 2, and 5 µL 10 mM adaptors for 15 min at 20°C. Adaptors were then removed by a sequential 0.45× and 1× VAHTS DNA clean beads purification. Samples were dissolved in 19 µL H_2O and mixed with 5 µL PCR primer mix, 25 µL NEBNext high-fidelity 2× PCR master mix (NEB M0541S), 1 µL heat-labile UDG. Samples were amplified for 10 min at 37°C; for 30 sec at 98°C; 10 cycles of 10 sec at 98°C, 30 sec at 60°C, and 30 sec at 72°C; followed by 5 min at 72°C. PCR product was purified by 0.9× VAHTS DNA clean beads before sending to sequencing.

Data sets

Tn5-based genomic DNA sequencing data were downloaded from the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) database under accession number GSE58652. rRNA-depleted strand-specific RNA sequencing data were downloaded from the GEO database under accession number GSE29278. TnRNA-seq data were downloaded from the GEO database under accession number GSE32307. SHERRY sequencing data were downloaded from the Genome Sequence Archive database (GSA; https://ngdc.cncb.ac.cn/gsa/) under accession number CRA002081.

Strand-specific data processing and analysis

For removing adaptors and low-quality reads, Trim Galore! (version 0.6.4) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used with the parameter "--paired." Bowtie 2 (version 2.3.5.1) (Langdon 2015) was used to map trimmed reads to the mouse reference genome mm9 and Tsin-H3.1-flag plasmid construct reference. The reference genome used in this study is available in the UCSC Genome Browser under mouse reference genome mm9. Realigning the reads to mm10 would not significantly affect the conclusions because the coordinates of most genomic regions could be uniquely converted between these two genome assemblies, and our study compares samples across different treatments and does not perform absolute analyses. PCR duplicates were removed by GATK4 (version 4.1.4.0) (Poplin et al. 2018) with the parameter "--REMOVE_DUPLICATES = true" in cDNA data.

For visualizing the read densities, normalizing the mapped reads, and calculating the coverage of features across the genome, BEDTools (version 2.92.2) (Quinlan 2014) and bedGraphToBigWig (version 4) (Kent et al. 2010) were used with the following parameters: "genomecov --scaleFactor 10,000,00/(the number of reads mapped to mm9 or Tsin-H3.1-flag plasmid construct reference)." Normalized signals were visualized in IGV (version 2.6.3) (Thorvaldsdottir et al. 2013). The correlations between two repeats were summarized in Supplemental Table S2. The correlations were calculated by the Pearson's product moment correlation.

Forward reads and reverse reads were filtered by SAMtools (version 1.11) (Sun et al. 2021) with the parameters "samtools view -F 0 × 10" and "samtools view -f 0 × 10." deepTools (version 3.5.1) (Sun et al. 2021) functions "computeMatrix" and "plotProfile" were used to draw the profile in ssDNA data. HOMER (version 4.11) (Heinz et al. 2010) was used to draw forward and reverse read densities profiles in cDNA and benchmark data.

Gene expression results were counted using featureCounts (version 2.0.0) (Liao et al. 2014) and normalized as reads per kilobase per million mapped reads (RPKM) by edgeR (version 3.28.1) (Robinson et al. 2010). Heatmaps were drawn by the R package pheatmap (version 1.0.12) (https://github.com/raivokolde/pheatmap). R (version 4.0) (R Core Team 2021) was used for data analysis.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE208316. The code used in this study is available at GitHub (https://github.com/Fanglab-zju/TABLE-seq) and as Supplemental Code. The bigWig files for the sequencing were uploaded as Supplemental Results.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Hangjun Wu in the Center of Cryo-Electron Microscopy (CCEM), Zhejiang University, for his technical assistance on the computing node setup. We thank our colleagues at the core facility of the Life Sciences Institute for their assistance with the computing node setup. We thank Dr. Jun Huang and Dr. Bin Zhao at the mentor program of Life Sciences Institute for the discussion in preparing the manuscript. This research was partly supported by grants from the National Key Research and Development Program of China (grant no. 2022YFA1300034), the National Natural Science Foundation of China (grants no. 32222017 and no. 81874153), the Fundamental Research Funds for the Central Universities (grant no. 2019QN81005), and the Opening Research Fund from Shanghai Key Laboratory of Stomatology, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine (grant no. 2022SKLS-KFKT002).

Author contributions: Y.Z., Y.T., and D.F. conceived the project. Y.Z., Z.S., J.J., Y.F., and X.W. performed the experiments. Y.T. performed the bioinformatics analysis. Y.Z., Z.S., J.J., Y.F., X.W., and D.F. analyzed the data. Y.Z., Y.T., and D.F. prepared the first draft.

References

Adey AC. 2021. Tagmentation-based single-cell genomics. *Genome Res* **31:** 1693–1705. doi:10.1101/gr.275223.121

Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, et al. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol* **11:** R119. doi:10.1186/gb-2010-11-12-r119

Agarwal S, Macfarlan TS, Sartor MA, Iwase S. 2015. Sequencing of first-strand cDNA library reveals full-length transcriptomes. *Nat Commun* **6:** 6002. doi:10.1038/ncomms7002

Ashton NW, Bolderson E, Cubeddu L, O'Byrne KJ, Richard DJ. 2013. Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Mol Biol* **14:** 9. doi:10.1186/1471-2199-14-9

Berg DE, Davies J, Allet B, Rochaix JD. 1975. Transposition of R factor genes to bacteriophage lambda. *Proc Natl Acad Sci* **72:** 3628–3632. doi:10.1073/pnas.72.9.3628

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. *Genome Biol* **19:** 199. doi:10.1186/s13059-018-1577-z

Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. 2020. A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* **21:** 526–540. doi:10.1038/s41576-020-0244-x

Carter B, Ku WL, Kang JY, Hu G, Perrie J, Tang Q, Zhao K. 2019. Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat Commun* **10:** 3747. doi:10.1038/s41467-019-11559-1

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10:** 691–703. doi:10.1038/nrg2640

Davies DR, Mahnke Braam L, Reznikoff WS, Rayment I. 1999. The three-dimensional structure of a Tn*5* transposase-related protein determined to 2.9-Å resolution. *J Biol Chem* **274:** 11904–11913. doi:10.1074/jbc.274.17.11904

Davies DR, Goryshin IY, Reznikoff WS, Rayment I. 2000. Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. *Science* **289:** 77–85. doi:10.1126/science.289.5476.77

Di L, Fu Y, Sun Y, Li J, Liu L, Yao J, Wang G, Wu Y, Lao K, Lee RW, et al. 2020. RNA sequencing by direct tagmentation of RNA/DNA hybrids. *Proc Natl Acad Sci* **117:** 2886–2893. doi:10.1073/pnas.1919800117

Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, Kuersten S, Myers RM. 2012. Transposase mediated construction of RNA-seq libraries. *Genome Res* **22:** 134–141. doi:10.1101/gr.127373.111

Goryshin IY, Reznikoff WS. 1998. Tn*5 in vitro* transposition. *J Biol Chem* **273:** 7367–7374. doi:10.1074/jbc.273.13.7367

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* **6:** 468–481. doi:10.1038/nprot.2010.190

Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks G-J, Larsson AJM, Faridani OR, Sandberg R. 2020. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol* **38:** 708–714. doi:10.1038/s41587-020-0497-0

Heiat M, Ranjbar R, Latifi AM, Rasaee MJ, Farnoosh G. 2017. Essential strategies to optimize asymmetric PCR conditions as a reliable method to generate large amount of ssDNA aptamers. *Biotechnol Appl Biochem* **64:** 541–548. doi:10.1002/bab.1507

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, Besir H, Remans K, Steinmetz LM. 2018. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3 (Bethesda)* **8:** 79–89. doi:10.1534/g3.117.300257

Johnson RC, Reznikoff WS. 1983. DNA sequences at the ends of transposon Tn*5* required for transposition. *Nature* **304:** 280–282. doi:10.1038/304280a0

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. Bigwig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26:** 2204–2207. doi:10.1093/bioinformatics/btq351

Kia A, Gloeckner C, Osothprarop T, Gormley N, Bomati E, Stephenson M, Goryshin I, He MM. 2017. Improved genome sequencing using an engineered transposase. *BMC Biotechnol* **17:** 6. doi:10.1186/s12896-016-0326-1

Kit Leng Lui S, Keegan S, Tonzi P, Kahli M, Chen YH, Chalhoub N, Coleman KE, Fenyo D, Smith DJ, Huang TT. 2021. Monitoring genome-wide replication fork directionality by Okazaki fragment sequencing in mammalian cells. *Nat Protoc* **16:** 1193–1218. doi:10.1038/s41596-020-00454-5

Kolmar L, Autour A, Ma X, Vergier B, Eduati F, Merten CA. 2022. Technological and computational advances driving high-throughput oncology. *Trends Cell Biol* **32:** 947–961. doi:10.1016/j.tcb.2022.04.008

Korada SK, Johns TD, Smith CE, Jones ND, McCabe KA, Bell CE. 2013. Crystal structures of *Escherichia coli* exonuclease I in complex with single-stranded DNA provide insights into the mechanism of processive digestion. *Nucleic Acids Res* **41:** 5887–5897. doi:10.1093/nar/gkt278

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* **6:** 291–295. doi:10.1038/nmeth.1311

Langdon WB. 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min* **8:** 1. doi:10.1186/s13040-014-0034-0

Li Z, Hua X, Serra-Cardona A, Xu X, Gan S, Zhou H, Yang WS, Chen CL, Xu RM, Zhang Z. 2020. DNA polymerase α interacts with H3-H4 and facilitates the transfer of parental histones to lagging strands. *Sci Adv* **6:** eabb5820. doi:10.1126/sciadv.abb5820

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Lopes R, Agami R, Korkmaz G. 2017. GRO-seq, a tool for identification of transcripts regulating gene expression. *Methods Mol Biol* **1543:** 45–55. doi:10.1007/978-1-4939-6716-2_3

Lu B, Dong L, Yi D, Zhang M, Zhu C, Li X, Yi C. 2020. Transposase-assisted tagmentation of RNA/DNA hybrid duplexes. *eLife* **9:** e54919. doi:10.7554/eLife.54919

Maréchal A, Zou L. 2015. RPA-coated single-stranded DNA as a platform for post-translational modifications in the DNA damage response. *Cell Res* **25:** 9–23. doi:10.1038/cr.2014.147

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci* **36:** 344–355. doi:10.1073/pnas.36.6.344

Metzker ML. 2010. Sequencing technologies: the next generation. *Nat Rev Genet* **11:** 31–46. doi:10.1038/nrg2626

Naumann TA, Reznikoff WS. 2002. Tn*5* transposase active site mutants. *J Biol Chem* **277:** 17623–17629. doi:10.1074/jbc.M200742200

Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, Macinnis B, Kwiatkowski DP, Swerdlow HP, et al. 2012. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics* **13:** 1. doi:10.1186/1471-2164-13-1

Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* **35:** e128. doi:10.1093/nar/gkm683

Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24:** 2033–2040. doi:10.1101/gr.177881.114

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv doi:10.1101/201178

Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5:** 1005–1010. doi:10.1038/nmeth.1270

Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics* **47:** 11.12.1–11.12.34. doi:10.1002/0471250953.bi1112s47

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Reznikoff WS. 2003. Tn*5* as a model for understanding DNA transposition. *Mol Microbiol* **47:** 1199–1206. doi:10.1046/j.1365-2958.2003.03382.x

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140. doi:10.1093/bioinformatics/btp616

Severins I, Joo C, van Noort J. 2022. Exploring molecular biology in sequence space: the road to next-generation single-molecule biophysics. *Mol Cell* **82:** 1788–1805. doi:10.1016/j.molcel.2022.04.024

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. 2012. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488:** 116–120. doi:10.1038/nature11243

Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. *Nat Rev Genet* **20:** 631–656. doi:10.1038/s41576-019-0150-2

Steiniger-White M, Rayment I, Reznikoff WS. 2004. Structure/function insights into Tn*5* transposition. *Curr Opin Struct Biol* **14:** 50–57. doi:10.1016/j.sbi.2004.01.008

Sun Z, Tang Y, Zhang Y, Fang Y, Jia J, Zeng W, Fang D. 2021. Joint single-cell multiomic analysis in Wnt3a induced asymmetric stem cell division. *Nat Commun* **12:** 5941. doi:10.1038/s41467-021-26203-0

Svarovskaia ES, Delviks KA, Hwang CK, Pathak VK. 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. *J Virol* **74:** 7171–7178. doi:10.1128/JVI.74.15.7171-7178.2000

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192. doi:10.1093/bib/bbs017

van den Hurk M, Erwin JA, Yeo GW, Gage FH, Bardy C. 2018. Patch-seq protocol to analyze the electrophysiology, morphology and transcriptome of whole single neurons derived from human pluripotent stem cells. *Front Mol Neurosci* **11:** 261. doi:10.3389/fnmol.2018.00261

Wei X, Xiang Y, Peters DT, Marius C, Sun T, Shan R, Ou J, Lin X, Yue F, Li W, et al. 2022. HiCAR is a robust and sensitive method to analyze open-chromatin-associated genome organization. *Mol Cell* **82:** 1225–1238.e6. doi:10.1016/j.molcel.2022.01.023

Wu SJ, Furlan SN, Mihalas AB, Kaya-Okur HS, Feroze AH, Emerson SN, Zheng Y, Carson K, Cimino PJ, Keene CD, et al. 2021. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* **39:** 819–824. doi:10.1038/s41587-021-00865-z

Xu Z, Feng J, Li Q. 2021. Measuring genome-wide nascent nucleosome assembly using ReIN-Map. *Methods Mol Biol* **2196:** 117–141. doi:10.1007/978-1-0716-0868-5_10

York D, Reznikoff WS. 1996. Purification and biochemical analyses of a monomeric form of Tn*5* transposase. *Nucleic Acids Res* **24:** 3790–3796. doi:10.1093/nar/24.19.3790

Zhang Z, Theurkauf WE, Weng Z, Zamore PD. 2012. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly (A) selection. *Silence* **3:** 9. doi:10.1186/1758-907X-3-9

Zhang B, Srivastava A, Mimitou E, Stuart T, Raimondi I, Hao Y, Smibert P, Satija R. 2022. Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nat Biotechnol* **40:** 1220–1230. doi:10.1038/s41587-022-01250-0

Zhou M, Reznikoff WS. 1997. Tn*5* transposase mutants that alter DNA binding specificity. *J Mol Biol* **271:** 362–373. doi:10.1006/jmbi.1997.1188

Received August 18, 2022; accepted in revised form February 1, 2023.