

1 BIOLOGICAL SCIENCES – Applied Biological Sciences

2

3 **RNA Sequencing by Direct Tagmentation of RNA/DNA Hybrids**

4

5 Lin Di,^{a,#} Yusi Fu,^{a,1,#} Yue Sun,^a Jie Li,^b Lu Liu,^b Jiacheng Yao,^b Guanbo
6 Wang,^{c,d} Yalei Wu,^e Kaiqin. Lao,^e Raymond W. Lee,^e Genhua Zheng,^e
7 Jun Xu,^f Juntaek Oh,^f Dong Wang,^f X. Sunney Xie,^{a,d,*} Yanyi Huang,^{a,d,g,*}
8 and Jianbin Wang^{b,*}

9

10 ^a Beijing Advanced Innovation Center for Genomics (ICG), Biomedical
11 Pioneering Innovation Center (BIOPIC), School of Life Sciences, and Peking-
12 Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.

13 ^b School of Life Sciences, and Tsinghua-Peking Center for Life Sciences,
14 Tsinghua University, Beijing 100084, China.

15 ^c School of Chemistry and Materials Science, Nanjing Normal University,
16 Nanjing, Jiangsu Province, China

17 ^d Institute for Cell Analysis, Shenzhen Bay Laboratory, Shenzhen, Guangdong
18 Province, China

19 ^e XGen US Co, South San Francisco, CA

20 ^f Department of Cellular and Molecular Medicine, Skaggs School of Pharmacy
21 and Pharmaceutical Sciences, University of California, San Diego, La Jolla,
22 CA 92093

23 ^g College of Engineering, Peking University, Beijing 100871, China

24

25 [#] These authors contributed equally to this work.

26 ¹ Present address: Department of Molecular and Human Genetics, Baylor
27 College of Medicine, Houston, TX 77030, USA

28 * Corresponding authors: jianbinwang@tsinghua.edu.cn (J.W.),
29 yanyi@pku.edu.cn (Y.H.) and sunneyxie@pku.edu.cn (X.S.X.)

30

31 Keywords: single cell | RNA-seq | Tn5 transposase

32 **Abstract**

33 Transcriptome profiling by RNA sequencing (RNA-seq) has been widely used to
34 characterize cellular status but it relies on second strand cDNA synthesis to generate
35 initial material for library preparation. Here we use bacterial transposase Tn5, which has
36 been increasingly used in various high-throughput DNA analyses, to construct RNA-seq
37 libraries without second strand synthesis. We show that Tn5 transposome can randomly
38 bind RNA/DNA heteroduplexes and add sequencing adapters onto RNA directly after
39 reverse transcription. This method, Sequencing HEteRo RNA-DNA-hYbrid (SHERRY), is
40 versatile and scalable. SHERRY accepts a wide range of starting materials, from bulk
41 RNA to single cells. SHERRY offers a greatly simplified protocol, and produces results
42 with higher reproducibility and GC uniformity compared with prevailing RNA-seq methods.

43

44

45

46

47 **Significance Statement**

48 RNA sequencing is widely used to measure gene expression in biomedical research;
49 therefore, improvements in the simplicity and accuracy of the technology are desirable.
50 All existing RNA sequencing methods rely on the conversion of RNA into double-stranded
51 DNA through reverse transcription followed by second strand synthesis. The latter step
52 requires additional enzymes and purification, and introduces sequence-dependent bias.
53 Here, we show that Tn5 transposase, which randomly binds and cuts double-stranded
54 DNA, can directly fragment and prime the RNA/DNA heteroduplexes generated by
55 reverse transcription. The primed fragments are then subject to PCR amplification. This
56 provides a new approach for simple and accurate RNA characterization and quantification.

57 **Introduction**

58 Transcriptome profiling through RNA sequencing (RNA-seq) has become routine in
59 biomedical research since the popularization of next-generation sequencers and the
60 dramatic fall in the cost of sequencing. RNA-seq has been widely used in addressing
61 various biological questions, from exploring the pathogenesis of disease (1, 2) to
62 constructing transcriptome maps for various species (3, 4). RNA-seq provides informative
63 assessments of samples, especially when heterogeneity in a complex biological system
64 (5, 6) or time-dependent dynamic processes are being investigated (7–9). A typical RNA-
65 seq experiment requires a DNA library generated from mRNA transcripts. The commonly
66 used protocols contain a few key steps, including RNA extraction, poly-A selection or
67 ribosomal RNA depletion, reverse transcription, second strand cDNA synthesis, adapter
68 ligation, and PCR amplification (10–12).

69

70 Although many experimental protocols, combining novel chemistry and processes, have
71 recently been invented, RNA-seq is still a challenging technology to apply. On one hand,
72 most of these protocols are designed for conventional bulk samples (11, 13–15), which
73 typically contain millions of cells or more. However, many cutting-edge studies require
74 transcriptome analyses of very small amounts of input RNA, for which most large-input
75 protocols do not work. The main reason of this incompatibility is because the purification
76 operations needed between the main experimental steps cause inevitable loss of the
77 nucleic acid molecules.

78 On the other hand, many single-cell RNA-seq protocols have been invented in the past
79 decade (16–19). However, for most of these protocols it is difficult to achieve both high
80 throughput and high detectability. One type of single-cell RNA-seq approach, such as
81 Smart-seq2 (17), is to introduce pre-amplification to address the low-input problem but
82 such an approach is likely to introduce bias and to impair quantification accuracy. Another
83 type of approach is to barcode each cell's transcripts and hence bioinformatically assign
84 identity to the sequencing data that is linked to each cell and each molecule (20–23).
85 However, the detectability and reproducibility of such approaches are still not ideal (24).
86 An easy and versatile RNA-seq method is needed that works with input from single cells
87 to bulk RNA.

88

89 Bacterial transposase Tn5 (25) has been employed in next generation sequencing, taking
90 advantage of the unique ‘tagmentation’ function of dimeric Tn5, which can cut double-
91 stranded DNA (dsDNA) and ligate the resulting DNA ends to specific adaptors.
92 Genetically engineered Tn5 is now widely used in sequencing library preparation for its
93 rapid processivity and low sample input requirement (26, 27). For general library
94 preparation, Tn5 directly reacts with naked dsDNA. This is followed by PCR amplification
95 with sequencing adaptors. Such a simple one-step tagmentation reaction has greatly
96 simplified the experimental process, shortening the workflow time and lowering costs.
97 Tn5-tagmentation has also been used for the detection of chromatin accessibility, high-
98 accuracy single-cell whole-genome sequencing, and chromatin interaction studies (28–
99 32). For RNA-seq, the RNA transcripts have to undergo reverse transcription and second
100 strand synthesis, before the Tn5 tagmentation of the resulting dsDNA (33, 34).

101

102 In this paper we present a novel RNA-seq method using Tn5 transposase to directly
103 tagment RNA/DNA hybrids to form an amplifiable library. We experimentally show that,
104 as an RNase H superfamily member (35), Tn5 binds to RNA/DNA heteroduplex similarly
105 as to dsDNA and effectively fragments and then ligates the specific amplification and
106 sequencing adaptor onto the hybrid. This method, named Sequencing HEteRo RNA-
107 DNA-hYbrid (SHERRY), greatly improves the robustness of low-input RNA-seq with a
108 simplified experimental procedure. We also show that SHERRY works with various
109 amounts of input sample, from single cells to bulk RNA, with a dynamic range spanning
110 six orders of magnitude. SHERRY shows superior cross-sample robustness and
111 comparable detectability for both bulk RNA and single cells compared with other
112 commonly used methods and provides a unique solution for small bulk samples that
113 existing approaches struggle to handle. Furthermore, this easy-to-operate protocol is
114 scalable and cost effective, holding promise for use in high-quality and high-throughput
115 RNA-seq applications.

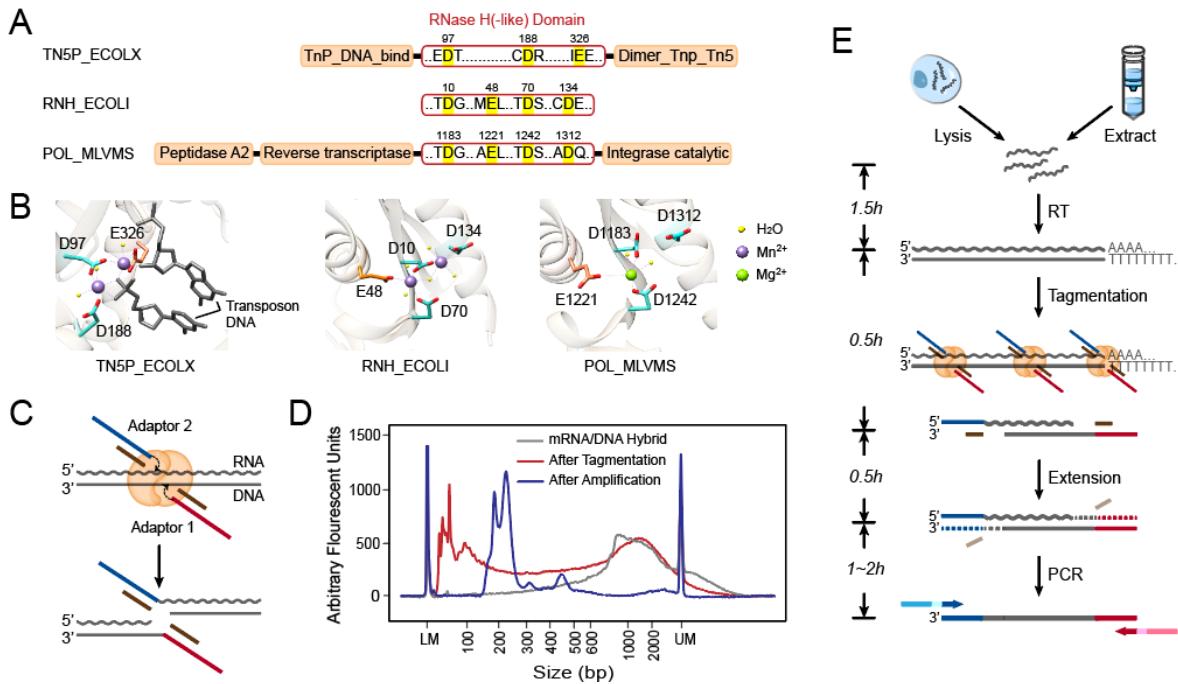
116

117

118 **Results**

119 **New RNA-seq strategy using RNA/DNA hybrid tagmentation.** Because of its
120 nucleotidyl transfer activity, transposase Tn5 has been widely used in recently developed
121 DNA sequencing technologies. Previous studies (36, 37) have identified a catalytic site
122 within its DDE domain (**Fig. S1A**). Indeed, when we mutated one of the key residues
123 (D188E) (38) in pTXB1 Tn5, its fragmentation activity on dsDNA was notably impaired
124 (**Fig. S1A, B**). Increased amounts of the mutated enzyme showed no improvement in
125 tagmentation, verifying the important role of the DDE domain in Tn5 tagmentation. Tn5 is
126 a member of the ribonuclease H-like (RNHL) superfamily, together with RNase H and
127 MMLV reverse transcriptase (35, 39, 40); therefore, we hypothesized that Tn5 is capable
128 of processing not only dsDNA but also RNA/DNA heteroduplex. Sequence alignment
129 between these three proteins revealed a conserved domain (two Asps and one Glu) within
130 their active sites, termed the RNase H-like domain (**Fig. 1A**). The two Asp residues (D97
131 and D188) in the Tn5 catalytic core were structurally similar to those of the other two
132 enzymes (**Fig. 1B**). Moreover, divalent ions, which are important for stabilizing substrate
133 and catalyzing reactions, also occupy similar positions in all three proteins (**Fig. 1B**) (39).
134 We determined the nucleic acid substrate binding pocket of Tn5 according to charge
135 distribution. We then docked double-stranded DNA and RNA/DNA heteroduplex into this
136 predicted pocket and showed that the binding site had enough space for an RNA/DNA
137 duplex (**Fig. S1C**). These structural similarities among Tn5, RNase H and MMLV reverse
138 transcriptase and the docking results further supported the possibility that Tn5 can
139 catalyze the strand transfer reaction on RNA/DNA heteroduplex (**Fig. 1C**).
140

141 To validate our hypothesis, we purified Tn5 using the pTXB1 plasmid and corresponding
142 protocol. We prepared RNA/DNA hybrids using mRNA extracted from HEK293T cells.
143 Using a typical dsDNA tagmentation protocol, we treated 15 ng of RNA/DNA hybrids with
144 0.6 μ l pTXB1 Tn5 transposome. Fragment analysis of the tagmented RNA/DNA hybrids
145 showed an obvious decrease (~1000bp) in fragment size compared with that of untreated
146 control, validating the capability of Tn5 to fragment the hybrid (**Fig. 1D**).
147
148



149

150 **Fig. 1** Tn5 tagmentation activity on double-stranded hybrids and the experimental process of SHERRY.
151 (A) RNase H-like (RNHL) domain alignment of Tn5 (TN5P_ECOLX), RNase H (RNH_ECOLI) and M-
152 MLV reverse transcriptase (POL_MLVMS). Active residues in the RNHL domains are labeled in bright
153 yellow. Orange boxes represent other domains. (B) Superposition of the RNHL active sites in these
154 three enzymes. PDB IDs are 1G15 (RNase H), 2HB5 (M-MLV) and 1MUS (Tn5). (C) Putative
155 mechanism of Tn5 tagmentation of an RNA/DNA hybrid. Crooked arrows represent nucleophilic
156 attacks. (D) Size distribution of mRNA/DNA hybrids with and without Tn5 tagmentation and after
157 amplification with index primers. (E) Workflow of sequencing library preparation by SHERRY. The input
158 can be a lysed single cell or extracted bulk RNA. After reverse transcription with oligo-dT primer, the
159 hybrid is directly tagmented by Tn5, followed by gap-repair and enrichment PCR. Wavy and straight
160 gray lines represent RNA and DNA, respectively. Dotted lines represent the track of extension step.

161

162 Based on the ability of the Tn5 transposome to fragment RNA/DNA heteroduplexes, we
163 propose SHERRY (Sequencing HEtRo RNA-DNA-hYbrid), a rapid RNA-seq library
164 construction method (Fig. 1E). SHERRY consists of three components: RNA reverse
165 transcription, RNA/cDNA hybrid tagmentation, and PCR amplification. The resulting
166 product is an indexed library that is ready for sequencing. Specifically, mRNA is reverse
167 transcribed into RNA/cDNA hybrids using d(T)₃₀VN primer. The hybrid is then tagmented
168 by the pTXB1 Tn5 transposome, the adding of partial sequencing adaptors to fragment
169 ends. DNA polymerase then amplifies the cDNA into a sequencing library after initial end
170 extension. The whole workflow only takes approximately four hours with hand-on time
171 less than 30 minutes.

172

173 To test SHERRY feasibility, we gap-repaired the RNA/DNA tagmentation products
174 illustrated in **Fig. 1D** (red line) and then amplified these fragments with library construction
175 primers. Amplified molecules (**Fig. 1D**, blue line) were approximately 100~150 bp longer
176 than the tagmentation products, which matched the extra length of adaptors added by
177 gap-repair and index primer amplification. (**Fig. S2**). Thus, direct Tn5 tagmentation of
178 RNA/DNA hybrids offers a new strategy for RNA-seq library preparation.

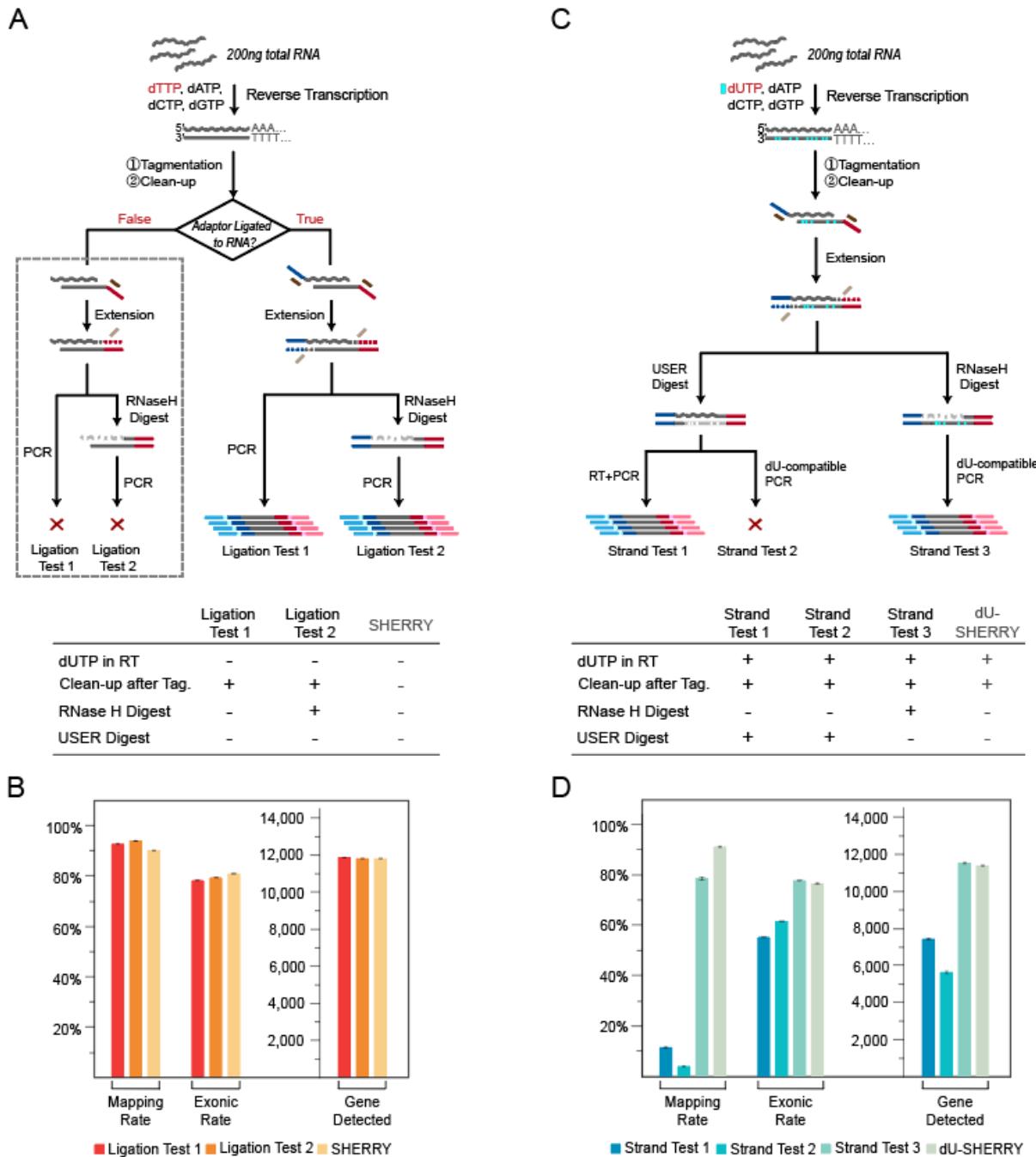
179

180 **Tn5 has ligation activity on tagmented RNA/DNA hybrids.** To further investigate the
181 detailed molecular events of RNA/DNA hybrid tagmentation, we designed a series of
182 verification experiments. First, we wanted to verify that the transposon adaptor can be
183 ligated to the end of fragmented RNA (**Fig. 2A**). In brief, we prepared RNA/DNA hybrids
184 from HEK293T RNA by reverse transcription. After tagmentation with the Tn5
185 transposome, we purified the products to remove Tn5 proteins and free adaptors. We
186 assumed that Tn5 ligated the adaptor to the fragmented DNA. At the same time, if Tn5
187 ligated the adaptor (**Fig. 2A**, dark blue) to the RNA strand, the adaptor could serve as a
188 template in the subsequent extension step. After extension the DNA strand should have
189 a primer binding site on both 5' and 3' ends for PCR amplification. RNase H treatment
190 should not affect production of the sequencing library. If Tn5 failed to ligate the adaptor to
191 the RNA strand, neither strand of the heteroduplex would be converted into a sequencing
192 library.

193

194 After PCR amplification, we obtained a high quantity product regardless of RNase H
195 digestion, indicating successful ligation of the adaptor to the fragmented DNA.
196 Sequencing results from both reaction test conditions as well as from SHERRY
197 showed >90% mapping rate to the human genome with ~80% exon rate and nearly
198 12,000 genes detected, validating the transcriptome origin of the library (**Fig. 2B**, **Fig.**
199 **S3A**). The additional purification step after reverse transcription and/or RNase H digestion
200 before PCR amplification did not affect the results, probably because of the large amount
201 of starting RNA. We examined the sequencing reads with an insert size shorter than 100
202 bp (shorter than the sequence read length), and 99.7% of them contained adaptor
203 sequence. Such read-through reads directly proved ligation of the adaptor to the

204 fragmented RNA (**Fig. S3B and C**). In summary, we confirmed that Tn5 transposome can
 205 fragment both DNA and RNA strands of RNA/DNA heteroduplexes.



206
 207 **Fig.2** Verification of Tn5 fragmentation of RNA/DNA heteroduplexes. **(A)** Procedures of two Ligation
 208 Tests. Gray dotted box indicates negative results. The table below lists key experimental parameters
 209 that are different from standard SHERRY. **(B)** Comparison of two Ligation Tests and standard SHERRY
 210 with respect to mapping rate, exon rate and number of genes detected. Each test consisted of two
 211 replicates of 200 ng HEK293T total RNA. **(C)** Strand Test procedures. **(D)** Comparison among three
 212 Strand Tests and dU-SHERRY with respect to mapping rate, exon rate and number of genes detected.

213 Each test consisted of two replicates of 200 ng HEK293T total RNA.

214

215 **Tagmented cDNA is the preferred amplification template.** Next, we investigated
216 whether RNA and DNA strands could be amplified to form the sequencing library (**Fig.**
217 **2C**). We replaced dTTP with dUTP during the reverse transcription and then purified the
218 tagmented products to remove free dUTP and Tn5 proteins. Bst 2.0 Warmstart DNA
219 polymerase was used for extension because it is able to use RNA as a primer and to
220 process the dU bases in the template. The product fragments were then treated with
221 either USER enzyme or RNase H to digest cDNA and RNA, respectively. We performed
222 RT-PCR with the USER-digested product, to test the efficiency of converting tagmented
223 RNA for library construction (Strand Test 1). To exclude interference from undigested DNA,
224 we performed PCR amplification with the USER-digested fragments using dU-compatible
225 polymerase (Strand Test 2). We also used dU-compatible PCR to test the efficiency of
226 converting tagmented cDNA for library construction (Strand Test 3). For comparison, we
227 included a control experiment with the same workflow as Strand Test 3 except that the
228 RNase H digestion step was omitted (dU-SHERRY) to ensure that Tn5 can recognize
229 substrates with dUTP.

230

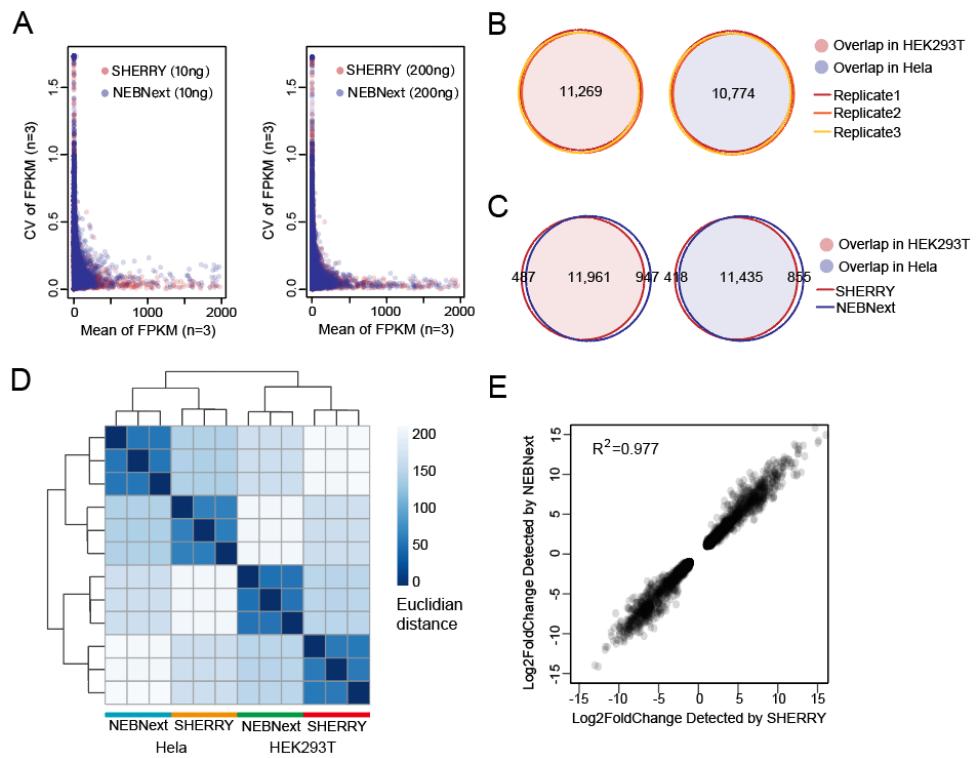
231 Sequencing results of Strand Test 1 showed a low mapping rate and gene detection count
232 that were only slightly higher than those of Strand Test 2. In contrast, Strand Test 3
233 demonstrated a similar exon rate and gene count to dU-SHERRY and SHERRY (**Fig. 2D,**
234 **Fig. S3A**). Based on these results, we conclude that the tagmented cDNA contributes to
235 the majority of the final sequencing library, likely because of inevitable RNA degradation
236 during the series of reactions.

237

238 **SHERRY for rapid one-step RNA-seq library preparation.** We tested different reaction
239 conditions to optimize SHERRY with 10 ng total RNA as input (**Fig. S4A**). We evaluated
240 the impact of different crowding agents, different ribonucleotide modifications on
241 transposon adaptors, and different enzymes for gap filling. We also included purification
242 after certain steps to remove primer dimers and carry-over contaminations. Sequencing
243 results showed little change in performance from most of these modifications, indicating
244 that SHERRY is robust under various conditions.

245

246 We then compared the optimized SHERRY with NEBNext® Ultra™ II, a commercially
247 available kit, for bulk RNA library preparation. This NEBNext kit is one of the most
248 commonly used kits used for RNA-seq experiments, with 10 ng total RNA being its
249 minimum input limit. We therefore tested the RNA-seq performance with 10 ng and 200
250 ng total RNA inputs, each condition having three replicates. SHERRY demonstrated
251 comparable performances with NEBNext for both input levels (**Fig. S4B**). For the 10 ng
252 input tests, SHERRY produced more precise gene expression measurements across
253 replicates (**Fig. 3A**), probably because of the simpler SHERRY workflow.



254

255 **Fig.3** Performance of SHERRY with large RNA input. **(A)** Coefficient of variation (CV) across three
256 replicates was plotted against the mean value of each gene's FPKM (Fragments Per Kilobase of
257 transcript per Million mapped reads). All experiments used HEK293T total RNA as input. **(B)** Genes
258 detected by SHERRY in three replicates of 200 ng HEK293T or HeLa total RNA are plotted in Venn
259 Diagrams. Numbers of common genes are indicated. **(C)** Common genes detected by SHERRY and
260 NEBNext in the three replicates of 200 ng HEK293T or HeLa total RNA. **(D)** Distance heatmap of
261 samples prepared by SHERRY or NEBNext for three replicates using 200 ng HEK293T or HeLa total
262 RNA. The color bar indicates the Euclidian distance. **(E)** Correlation of gene expression fold-change
263 identified by SHERRY and NEBNext. Involved genes are differentially expressed genes between
264 HEK293T and HeLa detected by both methods.

265

266 Next, we compared the ability to detect differentially expressed genes between HEK293T

267 and HeLa cells using SHERRY and NEBNext. In all three replicates, SHERRY detected
268 11,269 genes in HEK293T cells and 10,774 genes in HeLa cells, with high precision
269 (correlation coefficient $R^2=0.999$) (**Fig. 3B, Fig. S5A**). The numbers of detected genes
270 and their read counts identified by SHERRY and NEBNext were highly concordant (**Fig.**
271 **3C, Fig. S5B**). This excellent reproducibility of SHERRY ensured the reliability of
272 subsequent analyses. Then we plotted a heatmap of the distance matrix (**Fig. 3D**)
273 between different cell types and library preparation methods. Libraries from the same cell
274 type were clustered together as expected. Libraries from the same method also tended
275 to cluster together, indicating internal bias in both methods.

276

277 We then used DESeq2 to detect differentially expressed genes (P -value $<5\times10^{-6}$,
278 $|\log_2 \text{fold change}| >1$). In general, the thousands of differentially expressed genes
279 detected by both methods were highly similar (**Fig. S5C**) and their expression fold-change
280 was highly correlated (correlation coefficient $R^2=0.977$) between SHERRY and NEBNext
281 (**Fig. 3E**). Examination of the genes that showed differential expression in only one
282 method, revealed the same trend of expression change in the data from the other method
283 (**Fig. S5D and E**). We conclude that SHERRY provides equally reliable differential gene
284 expression information as NEBNext, but with a much faster and less labor-intensive
285 process, specifically saving around two hours hand-on time (**Fig. S7A**).

286

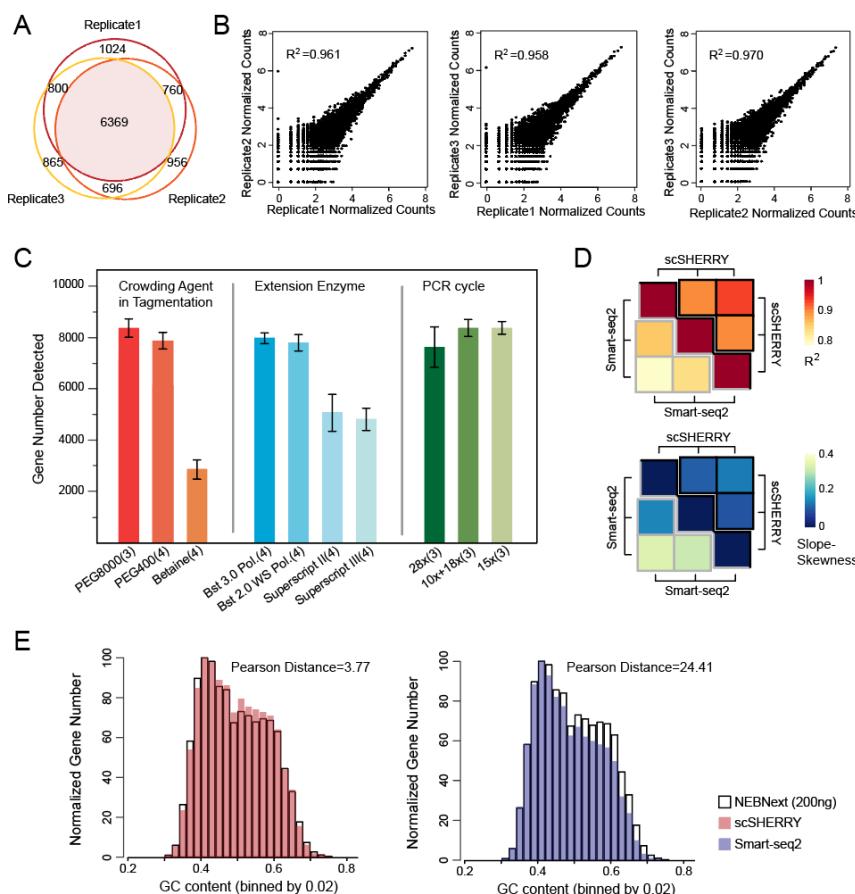
287 **SHERRY using trace amounts of RNA or single cells.** We next investigated whether
288 SHERRY could construct RNA-seq libraries from single cells. First, we reduced the input
289 to 100 pg total RNA, which is equivalent to RNA from about 10 cells. SHERRY results
290 were high quality, with high mapping and exon rates and nearly 9,000 genes detected
291 (**Fig. S6A**). 72% of these genes were detected in all three replicates, demonstrating good
292 reproducibility (**Fig. 4A**). The expression of these genes showed excellent precision with
293 R^2 ranging from 0.958 to 0.970. (**Fig. 4B**).

294

295 To further push the detection limit, we carried out single-cell SHERRY experiments
296 (scSHERRY) using HEK293T cell line. In contrast to the experiments with purified RNA,
297 scSHERRY required several optimizations to the standard protocol (**Fig. 4C, Fig. S6B**).

298 Although we found no positive effect by replacing betaine in the standard protocol with
299 other crowding agents during optimization, we found that addition of a crowding reagent
300 with a higher molecular weight improved the library quality from single cells. Therefore,
301 we used PEG8000 for the following scSHERRY experiments. For the extension step, the
302 use of Bst 3.0 or Bst 2.0 WarmStart DNA polymerases detected more genes than the use
303 of Superscript II or Superscript III reverse transcriptases. This is probably because of the
304 stronger processivity and strand displacement activity of Bst polymerases, and better
305 compatibility with higher reaction temperatures to open the secondary structure of RNA
306 templates. We also tried to optimize the PCR strategy because extensive amplification
307 can lead to strong bias. Compared to the continuous 28-cycle PCR, the incorporation of
308 a purification step after 10 cycles, or simply reducing the total cycle number to 15
309 increased the mapping rate and the number of genes detected. Therefore, we performed
310 15-cycle PCR without extra purification to better accommodate high-throughput
311 experiments.

312



313

314 **Fig.4** Performance of SHERRY with micro-input samples. **(A)** Genes detected by SHERRY in three
315 replicates with 100 pg HEK293T total RNA. **(B)** Correlations of normalized gene read counts between
316 replicates with 100 pg HEK293T total RNA. **(C)** Gene number detected by scSHERRY under various
317 experimental conditions in single HEK293T cells. Each condition involved 3–4 replicates. **(D)** The
318 heatmap of R^2 calculated from scSHERRY and Smart-seq2 replicates, and slope deviation in a linear
319 fitting equation for the two methods. **(E)** Normalized gene numbers with different GC content.
320

321 The optimized scSHERRY was capable of detecting 8,338 genes with a 50.17% mapping
322 rate (**Fig. S6D**), and the gene read counts correlated well (correlation coefficient $R^2=0.600$)
323 with Smart-seq2, the most prevalent protocol in the single-cell RNA amplification field.
324 Besides, scSHERRY showed better reproducibility than Smart-seq2 (**Fig.4D**). Compared
325 with Smart-seq2, the gene number and coverage uniformity of the scSHERRY-generated
326 library was slightly inferior (**Fig. S6D, E**) because Smart-seq2 enriches for full-length
327 transcripts via a preamplification step. However, this enrichment step of Smart-seq2 also
328 introduced bias (**Fig. 4E**). We used 200 ng of HEK293T total RNA to construct a
329 sequencing library using the NEBNext kit, expecting to capture as many genes as
330 possible. Besides, the NEBNext protocol used short RNA fragments for reverse
331 transcription and fewer cycles of PCR amplification, which should introduce less GC bias
332 than the other protocols. We then compared the GC distribution of genes detected by
333 scSHERRY or Smart-seq2 with NEBNext results. scSHERRY, which is free from second-
334 strand synthesis and preamplification, produced a distribution similar to the standard.
335 However, the library from Smart-seq2-amplified scRNA showed clear enrichment for
336 genes with lower GC content. Genes with high GC content were less likely to be captured
337 by Smart-seq2, which may cause biased quantitation results. Overall, compared with
338 Smart-seq2, scSHERRY produced libraries of comparable quality and lower GC-bias.
339 Moreover, the scSHERRY workflow spares pre-amplification and QC steps before
340 tagmentation, saving around four hours (**Fig. S7A**), and the one tube strategy is promising
341 for high throughput application.
342

343 **Discussion**

344 We found that the Tn5 transposome has the capability to directly fragment and tag
345 RNA/DNA heteroduplexes and, therefore, we have developed a quick RNA amplification
346 and library preparation method called SHERRY. The input for SHERRY could be RNA
347 from single cell lysate or total RNA extracted from a large number of cells. Comparison of

348 SHERRY with the commonly used Smart-seq2 protocol for single-cell input or the
349 NEBNext kit for bulk total RNA input, showed comparable performance for input amount
350 spanning more than five orders of magnitude. Furthermore, the whole SHERRY workflow
351 from RNA to sequencing library consists of only five steps in one tube and takes about
352 four hours, with hands-on time of less than 30 min. Smart-seq2, requires twice this
353 amount of time and an additional library preparation step is necessary. The ten-step
354 NEBNext protocol is much more labor-intensive and time-consuming (**Fig. S7A**).
355 Moreover, the SHERRY reagent cost is five-fold less compared with that of the other two
356 methods (**Fig. S7B**). For single cells, the lower mapping rate of SHERRY compared to
357 Smart-seq2 could increase sequencing cost. However, both methods reached a plateau
358 of saturation curve with 2 million total reads (**Fig. S8**), which costs less than \$5. Therefore,
359 SHERRY has strong competitive advantages over conventional RNA library preparation
360 methods and scRNA amplification methods.

361

362 In our previous experiments, we assembled a Tn5 transposome using home-purified
363 pTXB1 Tn5 and synthesized sequencer-adapted oligos. To generalize the SHERRY
364 method, and to confirm Tn5 tagmentation of RNA/DNA heteroduplexes, we tested two
365 commercially available Tn5 transposomes, Amplicon Tagment Mix (abbreviated as ATM)
366 from the Nextera XT kit (Illumina, USA) and TTE Mix V50 (abbreviated as V50) from the
367 TruPrep kit (Vazyme, China). We normalized the different Tn5 transposome sources
368 according to the enzyme processing 5 ng of genomic DNA to the same size under the
369 same reaction conditions (**Fig. S9A**). The tagmentation activity of our in-house pTXB1
370 Tn5 was 10-fold higher than V50 and 500-fold higher than ATM when using transposome
371 volume as the metric. The same units of enzyme were then used to process RNA/DNA
372 heteroduplexes prepared from 5 ng mRNA to confirm similar performance on such hybrids
373 (**Fig. S9B**). The RNA-seq libraries from all three enzymes showed consistent results,
374 demonstrating the robustness of SHERRY (**Fig. S9C**).

375

376 During DNA and RNA/DNA heteroduplex tagmentation, the Tn5 transposome reacted
377 with these two substrates in different patterns. We tagmented 5 ng DNA or mRNA/DNA
378 hybrids with 0.02 μ l, 0.05 μ l or 0.2 μ l pTXB1 Tn5 transposome (**Fig. S10**). As the amount

379 of Tn5 increased, dsDNA was cut into overall shorter fragments. While for the hybrid, Tn5
380 cut the template ‘one by one’, because only hybrids above a certain size became shorter
381 and most were too short to be cut. We supposed that such phenomenon might attribute
382 to the different conformation of dsDNA and RNA/DNA hybrid, since diameter of the latter
383 is larger. Thus, the binding pocket or catalytic site of Tn5 would be tuned to accommodate
384 the hybrid strands and cause different fragmentation pattern.

385

386 Despite its ease-of-use and commercial promise, the library quality produced by SHERRY
387 may be limited by unevenness of transcript coverage. Unlike the NEBNext kit, which
388 fragments RNA before reverse transcription, or Smart-seq2, which performs pre-
389 amplification to enrich full-length cDNAs, SHERRY simply reverse transcribes full-length
390 RNA. Reverse transcriptase is well known for its low efficiency and when using polyT as
391 the primer for extension, it is difficult for the transcriptase to reach the 5' end of the RNA
392 template. This can cause coverage imbalance across transcripts, making the RNA-seq
393 signal biased toward the 3' end of genes (**Fig. S11A**). In an attempt to solve this problem,
394 we added TSO primer, the sequence and concentration of which was the same as Smart-
395 seq2 protocol (17), to the reverse transcription buffer to mimic the Smart-seq2 reverse
396 transcription conditions. The resulting hybrid was then fragmented and amplified following
397 standard SHERRY workflow. This produced much improved evenness across transcripts
398 (**Fig. S11A and B**), although some of the sequencing parameters dropped accordingly
399 (**Fig. S11C**). We believe that with continued optimization SHERRY will improve RNA-seq
400 performance.

401

402 **Author contributions**

403 Y.F., K.L., X.S.X., Y.H. and J.W. conceived the project; L.D., J.X., J.O. and D.W. performed
404 structural analysis; Y.S., J.L. and G.W. performed Tn5 purification and characterization;
405 L.D., Y.F. and L.L. conducted research; L.D., Y.F., L.L., J.Y., Y.W., R.L., G.Z., Y.H. and J.W.
406 analyzed the data; L.D., X.S.X., Y.H. and J.W. wrote the manuscript with the help from all
407 other authors.

408

409 **Conflict of interest statement**

410 XGen US Co has applied for a patent related to this work. X. Sunney Xie, Kaijin. Lao and
411 Yalei Wu are shareholders of XGen US Co. Kai Q. Lao, Yalei Wu, Raymond W. Lee and
412 Genhua Zheng are employees of XGen US Co.

413

414 **Acknowledgement**

415 We thank Dr. Yun Zhang and BIOPIC sequencing platform at Peking University for the
416 assistance of high-throughput sequencing experiments. This work was supported by
417 National Natural Science Foundation of China (21675098, 21525521), Ministry of Science
418 and Technology of China (2018YFA0800200, 2018YFA0108100, 2018YFC1002300),
419 2018 Beijing Brain Initiation (Z181100001518004), Beijing Advanced Innovation Center
420 for Structural Biology, and Beijing Advanced Innovation Center for Genomics.

421

422 **References**

- 423 1. Y. Liu, Y. L. Lightfoot, N. Seto, C. Carmona-Rivera, E. Moore, R. Goel, L. O'Neil, P.
424 Mistry, V. Hoffmann, S. Mondal, P. N. Premnath, K. Gribbons, S. Dell'Orso, K. Jiang, P. R.
425 Thompson, H. W. Sun, S. A. Coonrod, M. J. Kaplan, Peptidylarginine deiminases 2 and
426 4 modulate innate and adaptive immune responses in TLR-7-dependent lupus. *JCI Insight* **3**, e124729 (2018)
- 427 2. E. Schmidt, I. Dhaouadi, I. Gaziano, M. Oliverio, P. Klemm, M. Awazawa, G. Mitterer,
428 E. Fernandez-Rebollo, M. Pradas-Juni, W. Wagner, P. Hammerschmidt, R. Loureiro, C.
429 Kiefer, N. R. Hansmeier, S. Khani, M. Bergami, M. Heine, E. Ntini, P. Frommolt, P. Zentis,
430 U. A. Ørom, J. Heeren, M. Blüher, M. Bilban, and J. W. Kornfeld, LincRNA H19 protects
431 from dietary obesity by constraining expression of monoallelic genes in brown fat. *Nat. Commun.* **9**, 3622 (2018).
- 432 3. O. Wurtzel, R. Sapra, F. Chen, Y. Zhu, B. A. Simmons, R. Sorek, A single-base
433 resolution map of an archaeal transcriptome. *Genome Res.* **20**, 133-141 (2010).
- 434 4. A. A. Penin, A. V. Klepikova, A. S. Kasianov, E. S. Gerasimov, M. D. Logacheva,
435 Comparative Analysis of Developmental Transcriptome Maps of *Arabidopsis thaliana* and
436 *Solanum lycopersicum*. *Genes* **10**, 50 (2019).
- 437 5. P. Civita, S. Franceschi, P. Aretini, V. Ortenzi, M. Menicagli, F. Lessi, F. Pasqualetti, A.
438 G. Naccarato, C. M. Mazzanti, Laser Capture Microdissection and RNA-Seq Analysis:
- 439
- 440

- 441 High Sensitivity Approaches to Explain Histopathological Heterogeneity in Human
442 Glioblastoma FFPE Archived Tissues. *Front Oncol.* **9**, 482 (2019).
- 443 6. D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner,
444 N. Cohen, S. Jung, A. Tanay, I. Amit, Massively Parallel Single-Cell RNA-Seq for Marker-
445 Free Decomposition of Tissues into Cell Types. *Science* **343**, 776-779 (2014)
- 446 7. Y. Chen, Y. Zheng, Y. Gao, Z. Lin, S. Yang, T. Wang, Q. Wang, N. Xie, R. Hua, M. Liu,
447 J. Sha, M. D. Griswold, J. Li, F. Tang, M. H. Tong, Single-cell RNA-seq uncovers dynamic
448 processes and critical regulators in mouse spermatogenesis. *Cell Res.* **28**, 879-896
449 (2018).
- 450 8. M. Wang, X. Liu, G. Chang, Y. Chen, G. An, L. Yan, S. Gao, Y. Xu, Y. Cui, J. Dong, Y.
451 Chen, X. Fan, Y. Hu, K. Song, X. Zhu, Y. Gao, Z. Yao, S. Bian, Y. Hou, J. Lu, R. Wang, Y.
452 Fan, Y. Lian, W. Tang, Y. Wang, J. Liu, L. Zhao, L. Wang, Z. Liu, R. Yuan, Y. Shi, B. Hu,
453 X. Ren, F. Tang, X. Y. Zhao, J. Qiao, Single-Cell RNA Sequencing Analysis Reveals
454 Sequential Cell Fate Transition during Human Spermatogenesis. *Cell Stem Cell* **23**, 599-
455 614 (2018).
- 456 9. H. Hochgerner, A. Zeisel, P. Lönnerberg, S. Linnarsson, Conserved properties of
457 dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA
458 sequencing. *Nat. Neurosci.* **21**, 290-299 (2018).
- 459 10. U. Gubler, B. J. Hoffman, A simple and very efficient method for generating cDNA
460 libraries. *Gene* **25**, 263-269 (1983).
- 461 11. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and
462 quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621-628 (2008).
- 463 12. P. Cui, Q. Lin, F. Ding, C. Xin, W. Gong, L. Zhang, J. Geng, B. Zhang, X. Yu, J. Yang,
464 S. Hu, J. Yu, A comparison between ribo-minus RNA-sequencing and polyA-selected
465 RNA-sequencing. *Genomics* **96**, 259-265 (2010).
- 466 13. N. Cloonan, A. R. Forrest, G. Kolle, B. B. Gardiner, G. J. Faulkner, M. K. Brown, D. F.
467 Taylor, A. L. Steptoe, S. Wani, G. Bethel, A. J. Robertson, A. C. Perkins, S. J. Bruce, C.
468 C. Lee, S. S. Ranade, H. E. Peckham, J. M. Manning, K. J. McKernan, S. M. Grimmond,
469 Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**,
470 613-619 (2008).
- 471 14. C. D. Armour, J. C. Castle, R. Chen, T. Babak, P. Loerch, S. Jackson, J. K. Shah, J.

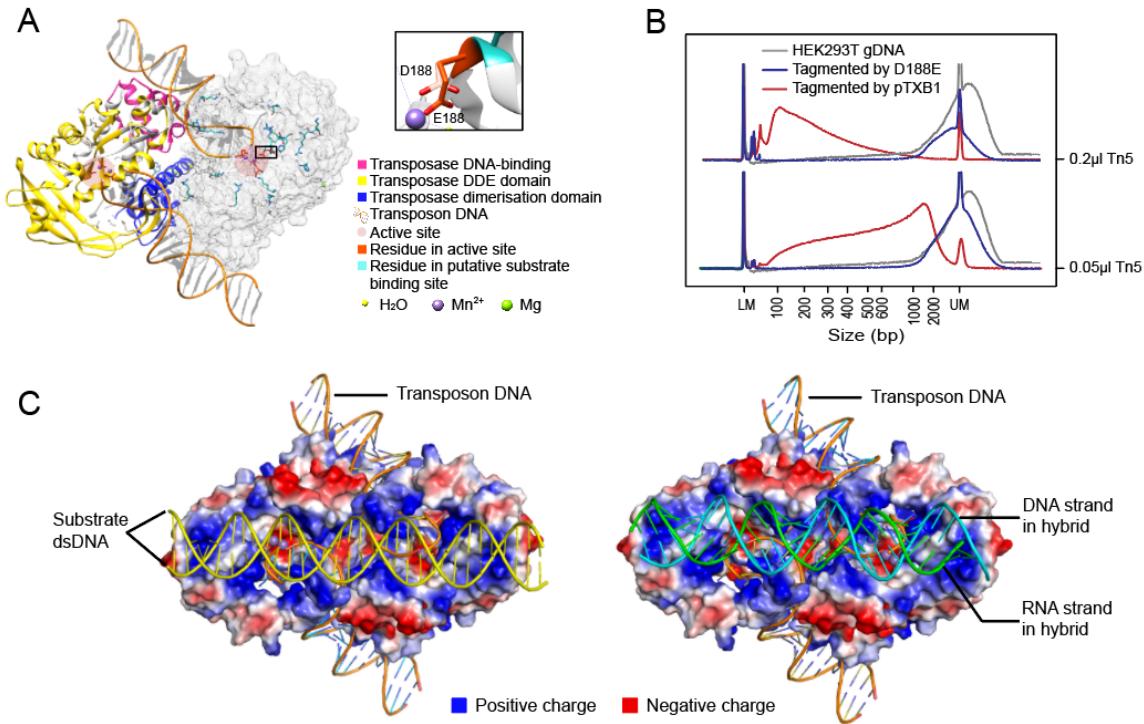
- 472 Dey, C. A. Rohl, J. M. Johnson, C. K. Raymond, Digital transcriptome profiling using
473 selective hexamer priming for cDNA synthesis. *Nat. Methods* **6**, 647-649 (2009).
- 474 15. D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H.
475 Lehrach, A. Soldatov, Transcriptome analysis by strand-specific sequencing of
476 complementary DNA. *Nucleic Acids Res.* **37**, e123 (2009).
- 477 16. F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B.
478 B. Tuch, A. Siddiqui, K. Lao, M. A. Surani, mRNA-Seq whole-transcriptome analysis of a
479 single cell. *Nat. Methods* **6**, 377-382 (2009).
- 480 17. S. Picelli, O. R. Faridani, A. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-
481 length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171-181 (2014).
- 482 18. T. Hashimshony, N. Senderovich, G. Avital, A. Klochendler, Y. de Leeuw, L. Anavy, D.
483 Gennert, S. Li, K. J. Livak, O. Rozenblatt-Rosen, Y. Dor, A. Regev, I. Yanai, CEL-Seq2:
484 sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
- 485 19. Y. Fu, H. Chen, L. Liu, Y. Huang, Single Cell Total RNA Sequencing through Isothermal
486 Amplification in Picoliter-Droplet Emulsion. *Anal. Chem.* **88**, 10795-10799 (2016).
- 487 20. G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo,
488 T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G.
489 Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M.
490 Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K.
491 R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J.
492 Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat.*
493 *Commun.* **8**, 14049 (2017).
- 494 21. T. M. Gierahn, M. H. Wadsworth, T. K. Hughes, B. D. Bryson, A. Butler, R. Satija, S.
495 Fortune, J. C. Love, A. K. Shalek, Seq-Well: portable, low-cost RNA sequencing of single
496 cells at high throughput. *Nat. Methods* **14**, 395-398 (2017).
- 497 22. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee,
498 S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure,
499 Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*
500 **357**, 661-667 (2017).
- 501 23. A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T.
502 Graybuck, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G.

- 503 Seelig, Single-cell profiling of the developing mouse brain and spinal cord with split-pool
504 barcoding. *Science* **360**, 176-182 (2018).
- 505 24. P. See, J. Lum, J. Chen, F. Ginhoux, A Single-Cell Sequencing Guide for
506 Immunologists. *Front Immunol.* **9**, 2425 (2018).
- 507 25. I. Y. Goryshin, W. S. Reznikoff, Tn5 in vitro transposition. *J. Biol. Chem.* **273**, 7367–
508 7374 (1998).
- 509 26. A. Adey, H. G. Morrison, Asan, X. Xun, J. O. Kitzman, E. H. Turner, B. Stackhouse, A.
510 P. MacKenzie, N. C. Caruccio, X. Zhang, J. Shendure, Rapid, low-input, low-bias
511 construction of shotgun fragment libraries by high-density in vitro transposition. *Genome*
512 *Biol.* **11**, R119 (2010).
- 513 27. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5
514 transposase and tagmentation procedures for massively scaled sequencing projects.
515 *Genome Res.* **24**, 2033-2040 (2014).
- 516 28. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition
517 of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
518 binding proteins and nucleosome position. *Nat. Methods* **10**, 1213-1218 (2013).
- 519 29. C. Chen, D. Xing, L. Tan, H. Li, G. Zhou, L. Huang, X. S. Xie, Single-cell whole-
520 genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**,
521 189-194 (2017).
- 522 30. S. Rohrback, C. April, F. Kaper, R. R. Rivera, C. S. Liu, B. Siddoway, J. Chun,
523 Submegabase copy number variations arise during cerebral cortical neurogenesis as
524 revealed by single-cell whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **115**,
525 10804-10809 (2018).
- 526 31. L. Tan, D. Xing, C. H. Chang, H. Li, X. S. Xie, Three-dimensional genome structures
527 of single diploid human cells. *Science* **361**, 924-928 (2018).
- 528 32. B. Lai, Q. Tang, W. Jin, G. Hu, D. Wangsa, K. Cui, B. Z. Stanton, G. Ren, Y. Ding, M.
529 Zhao, S. Liu, J. Song, T. Ried, K. Zhao, Trac-looping measures genome structure and
530 chromatin accessibility. *Nat. Methods* **15**, 741-747 (2018).
- 531 33. J. Gertz, K. E. Varley, N. S. Davis, B. J. Baas, I. Y. Goryshin, R. Vaidyanathan, S.
532 Kuersten, R. M. Myers, Transposase mediated construction of RNA-seq libraries.
533 *Genome Res.* **22**, 134-141 (2011).

- 534 34. S. Brouilette, S. Kuersten, C. Mein, M. Bozek, A. Terry, K. R. Dias, L. Bhaw-Rosun, Y.
535 Shintani, S. Coppen, C. Ikebe, V. Sawhney, N. Campbell, M. Kaneko, N. Tano, H. Ishida,
536 K. Suzuki, K. Yashiro, A simple and novel method for RNA-seq library preparation of single
537 cell cDNA analysis by hyperactive Tn5 transposase. *Dev. Dyn.* **241**, 1584-1590 (2012).
- 538 35. K. A. Majorek, S. Dunin-Horkawicz, K. Steczkiewicz, A. Muszewska, M. Nowotny, K.
539 Ginalska, J. M. Bujnicki, The RNase H-like superfamily: new members, comparative
540 structural analysis and evolutionary classification. *Nucleic Acids Res.* **42**, 4160-4179
541 (2014).
- 542 36. D. R. Davies, I. Y. Goryshin, W. S. Reznikoff, I. Rayment, Three-dimensional structure
543 of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77-85 (2000).
- 544 37. W. S. Reznikoff, Transposon Tn5. *Annu. Rev. Genet.* **42**, 269-286 (2008).
- 545 38. G. Peterson, W. Reznikoff, Tn5 transposase active site mutations suggest position of
546 donor backbone DNA in synaptic complex. *J. Biol. Chem.* **278**, 1904-1909 (2003).
- 547 39. M. Nowotny, S. A. Gaidamakov, R. J. Crouch, W. Yang, Crystal structures of RNase
548 H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. *Cell*
549 **121**, 1005-1016 (2005)
- 550 40. D. Lim, G. G. Gregorio, C. Bingman, E. Martinez-Hackert, W. A. Hendrickson, S. P.
551 Goff, Crystal structure of the Moloney murine leukemia virus RNase H domain. *J. Virol.*
552 **80**, 8379-8389 (2006).
- 553
- 554

555 **Supplementary Figures**

556

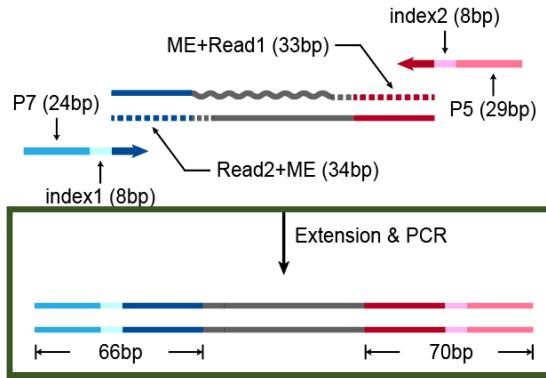


557

558

559 **Fig.S1** Structure of Tn5 and D188E mutation in Tn5. **(A)** Structure of pTXB1 Tn5 (PDB ID: 1MUS). Left monomer marked domains in different colors. Right monomer marked residues of catalytical core and putative substrate binding site in atom form. [Referred to D. R. Davies, I. Y. Goryshin, W. S. Reznikoff, I. Rayment, Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289, 77-85 (2000).] Black box showed D188E mutation. **(B)** Size distribution of genomic DNA with no treatment or tagmented by D188E mutant Tn5 or tagmented by pTXB1 Tn5. **(C)** Model of docking double-stranded DNA (Left) or RNA/DNA heteroduplex (Right) in predicted substrate binding sites of Tn5.

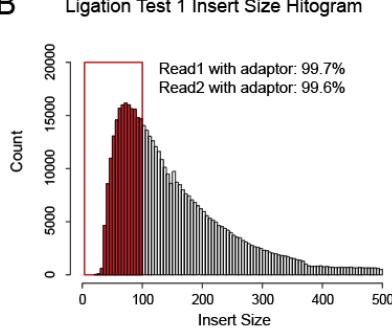
567



A

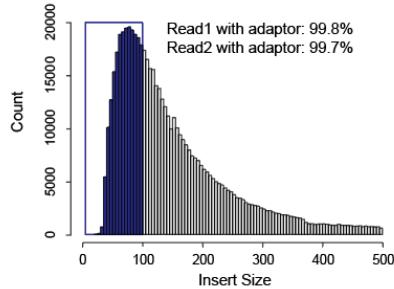
	Mapping Rate	Exonic Rate	Gene Detected
Ligation Test 1	92.20±0.20%	77.76±0.18%	11,825±7
Ligation Test 2	93.30±0.00%	78.98±0.10%	11,784±1
SHERRY	89.77±0.05%	80.32±0.04%	11,784±8
Strand Test 1	11.25±0.05%	54.92±0.06%	7,384±13
Strand Test 2	4.05±0.15%	61.21±0.32%	5,566±81
Strand Test 3	78.20±0.50%	77.38±0.10%	11,470±14
du-SHERRY	90.90±0.10%	76.31±0.00%	11,335±40

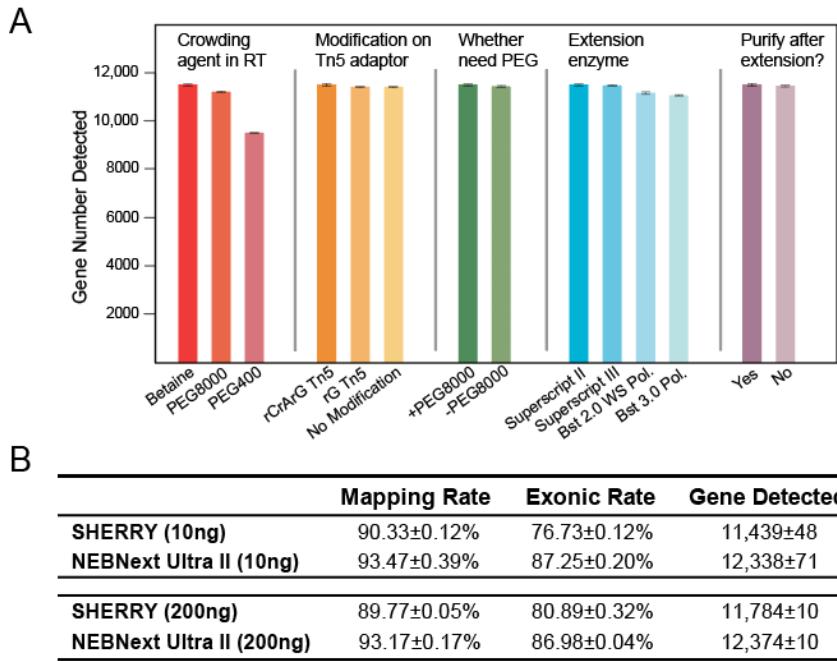
B



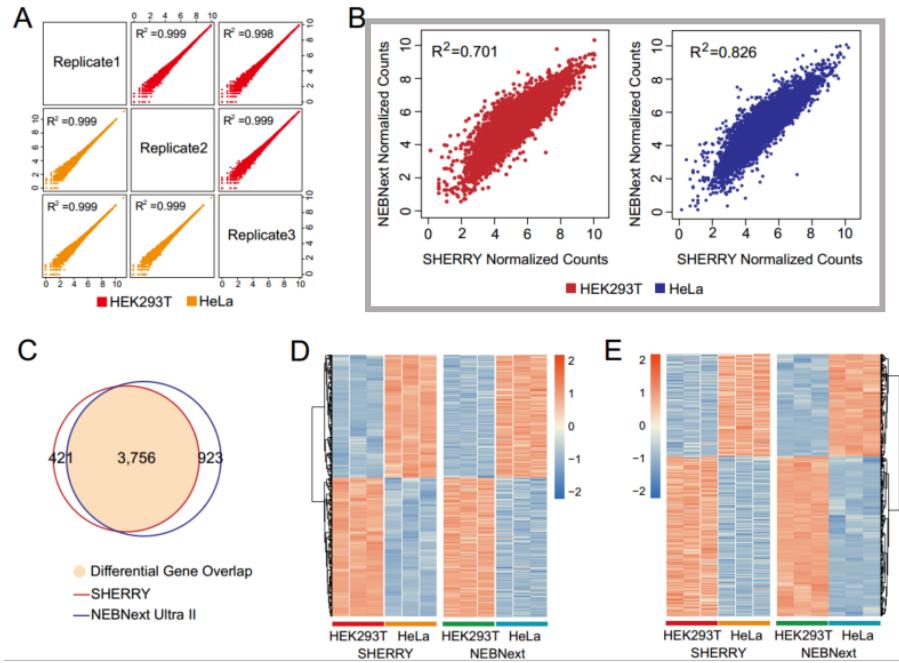
C

Ligation Test 2 Insert Size Histogram





580
581
582 **Fig.S4** Optimization of SHERRY and comparison with NEBNext. **(A)** Gene number detected by
583 SHERRY under various experimental conditions. Each condition consisted of three replicates of 10 ng
584 HEK293T total RNA. **(B)** Comparison of sequencing indicators between SHERRY and NEBNext with
585 10 ng and 200 ng HEK293T total RNA input. Each condition consisted of three replicates and down-
586 sampled to 2 million total reads.
587



588

589

590 **Fig.S5** Functional comparison between SHERRY and NEBNext. **(A)** Correlation of normalized gene
591 counts among duplicates of SHERRY, which start from 200 ng HEK293T total RNA input. **(B)**
592 Correlation of normalized genes counts (average of three replicates) between SHERRY and NEBNext
593 within the two cell types. The input was 200 ng total RNA. **(C)** Differentially expressed genes of HeLa
594 and HEK293T detected by SHERRY and NEBNext kit (200 ng input) are plotted into Venn Diagram.
595 Colored area represents genes identified by both methods. Gene numbers are listed on corresponding
596 part. **(D)** Heatmap of differentially expressed genes detected by SHERRY while missed by NEBNext
597 kit. The Color bar indicates Z-score. **(E)** Heatmap of differentially expressed genes detected by
598 NEBNext kit while missed by SHERRY. The Color bar indicates Z-score.

599

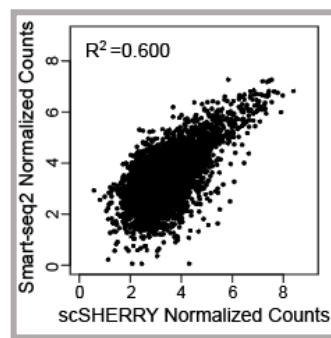
A

	Mapping Rate	Exonic Rate	Gene Detected
Replicate1	64.60%	62.60%	8,955
Replicate2	61.80%	59.90%	8,783
Replicate3	64.80%	63.00%	8,733

B

		Mapping Rate	Exonic Rate	Gene Detected
Crowding	8% PEG8000 (m/v)	50.17 \pm 2.72%	44.28 \pm 6.98%	8,338 \pm 345
Agent in	20% PEG400 (v/v)	29.98 \pm 6.92%	56.53 \pm 2.13%	7,835 \pm 326
Fragmentation	20% Betaine (v/v)	67.63 \pm 2.51%	43.47 \pm 5.38%	2,846 \pm 367
Extension	Bst 3.0	34.53 \pm 5.49%	60.57 \pm 3.30%	7,966 \pm 209
Enzyme	Bst 2.0 Warmstart	34.45 \pm 3.62%	58.08 \pm 7.79%	7,790 \pm 321
	SuperScript™ II	33.25 \pm 6.67%	55.11 \pm 4.74%	5,065 \pm 307
PCR Cycle	SuperScript™ III	33.20 \pm 13.68%	48.33 \pm 4.34%	4,803 \pm 550
	28 cycle	24.43 \pm 7.65%	52.46 \pm 7.11%	7,590 \pm 778
	10 cycle + 18 cycle	50.17 \pm 2.72%	44.28 \pm 6.98%	8,338 \pm 345
	15 cycle	31.97 \pm 2.85%	53.94 \pm 9.83%	8,352 \pm 255

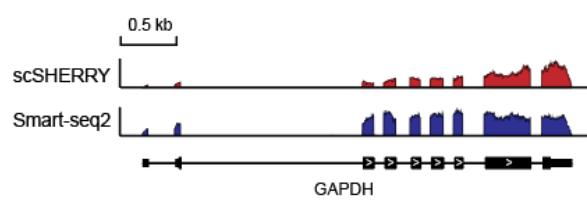
C



D

	Mapping Rate	Exonic Rate	Gene Detected
scSHERRY	50.17 \pm 2.72%	44.28 \pm 6.98%	8,338 \pm 345
Smart-seq2	91.23 \pm 0.46%	72.78 \pm 2.26%	9,761 \pm 577

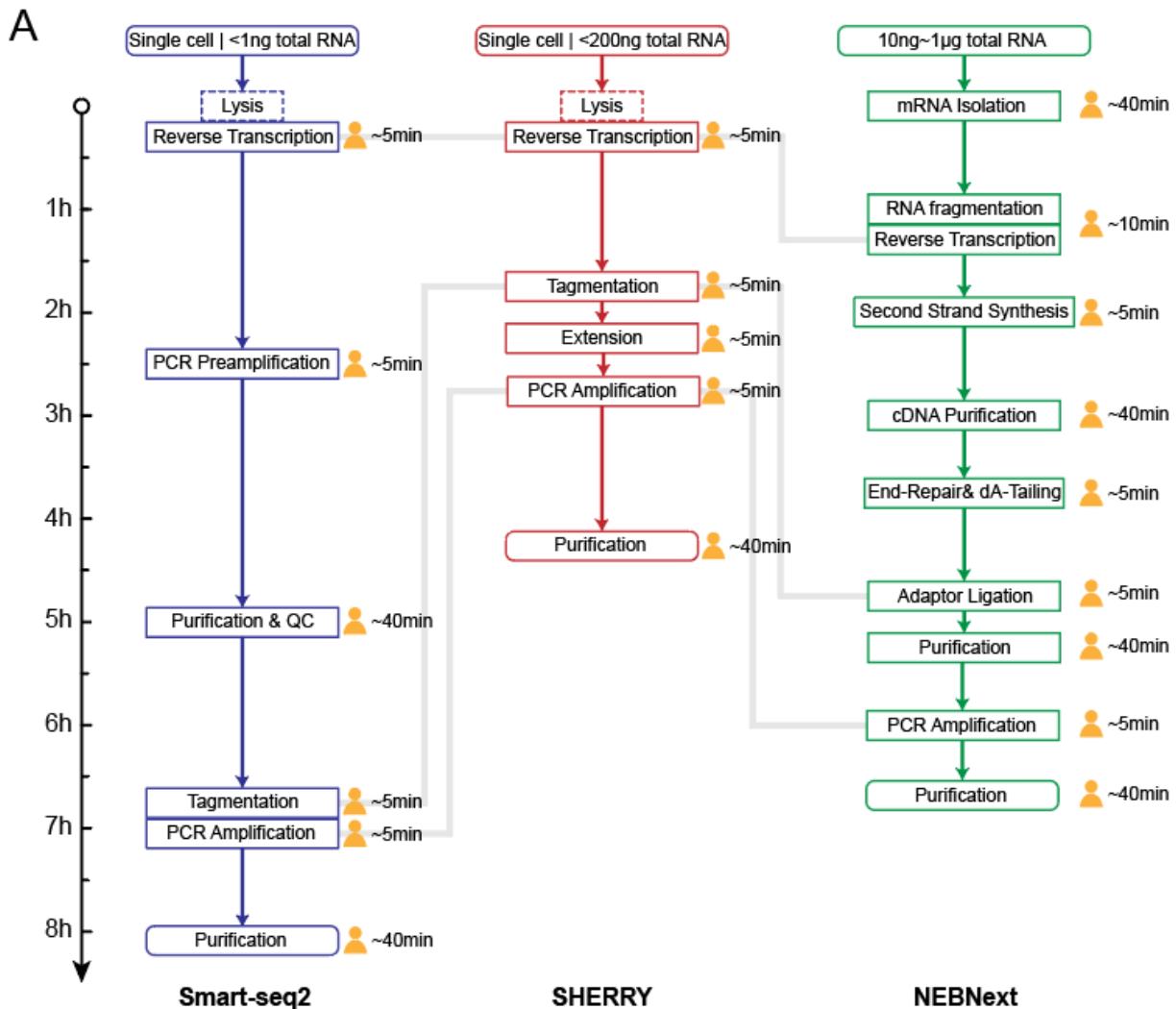
E



600

601

602 **Fig.S6** Optimization of micro-input SHERRY and comparison with Smart-seq2. (A) Sequencing
603 indicators of SHERRY (n=3) starting from 100 pg HEK293T total RNA input. (B) Comparison of
604 scSHERRY library quality under various experiment conditions. Each condition used single HEK293T
605 cell as input and consisted of 3-4 replicates. (C) Correlation of normalized gene counts (average of
606 three replicates) between scSHERRY and Smart-seq2. (D) Comparison of sequencing indicators
607 between scSHERRY (n=3) and Smart-seq2 (n=4). Both of them used single HEK293T cells as input.
608 (E) The coverage of GAPDH transcript calculated from scSHERRY and Smart-seq2.



B

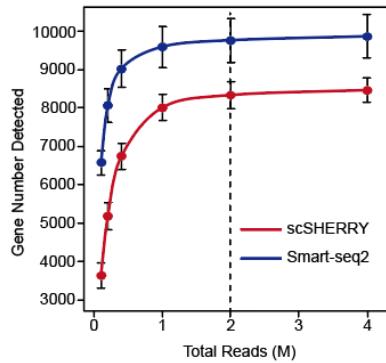
	Smart-seq2	SHERRY	NEBNext
Lysis+RT	~\$4.04	~\$3.94	mRNA Isolation ~\$2.79
Pre-amplification	~\$2.09	-	-
Library Prep	~\$42.10	~\$6.46	Library Prep ~\$45.42
Total	~\$48.23	~\$10.40	Total ~\$48.21

Fig.S7 Comparison of workflow and cost among Smart-seq2, SHERRY and NEBNext kit. **(A)** Workflow of Smart-seq2, SHERRY and NEBNext kit. Length of arrow indicates time consumed for each step. The human-shaped icon indicated hands-on time. Dotted box means this step is alternative. Gray line connects corresponding key step in each method. **(B)** Cost list of Smart-seq2, SHERRY and NEBNext kit.

609
610

611 **Fig.S7** Comparison of workflow and cost among Smart-seq2, SHERRY and NEBNext kit. **(A)** Workflow
612 of Smart-seq2, SHERRY and NEBNext kit. Length of arrow indicates time consumed for each step.
613 The human-shaped icon indicated hands-on time. Dotted box means this step is alternative. Gray line
614 connects corresponding key step in each method. **(B)** Cost list of Smart-seq2, SHERRY and NEBNext
615 kit.

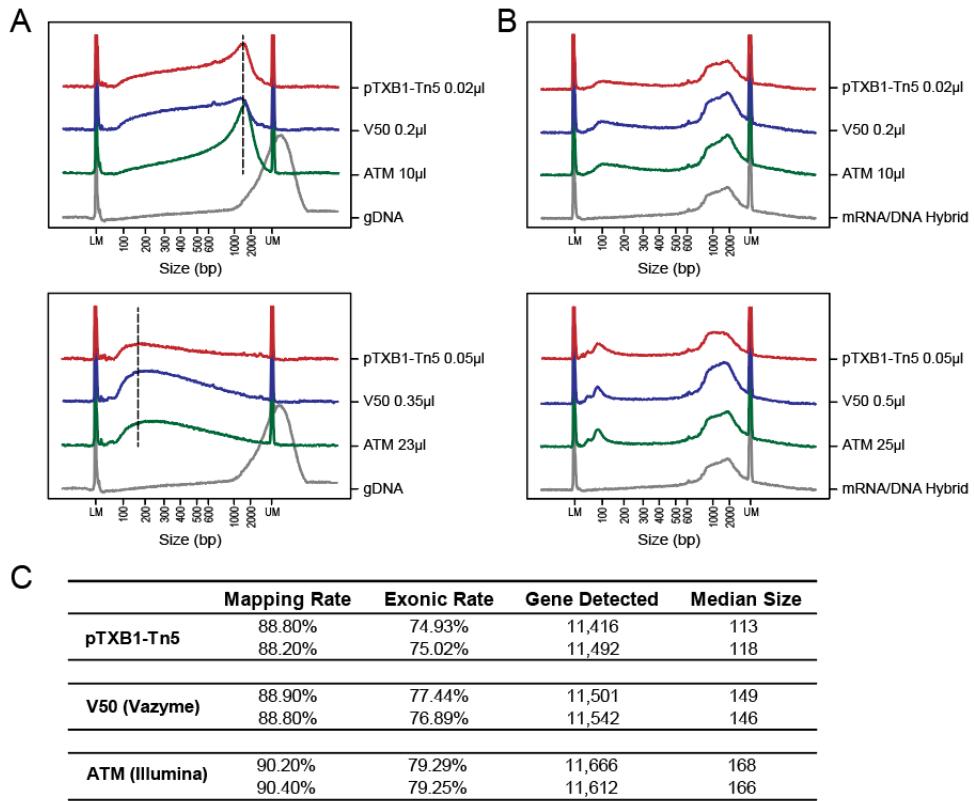
616
617



618
619

620 **Fig.S8** Saturation curve of scSHERRY (n=3) and Smart-seq2 (n=4). Both methods used single
621 HEK293T cell as input. Dotted line indicated 2 million total reads.

622
623
624

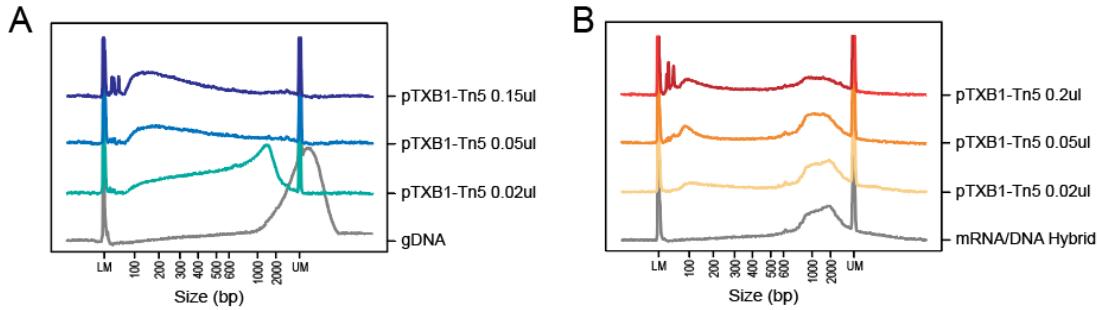


625

626

627 **Fig.S9** Hybrid fragmentation activity of commercial Tn5. (A) Size distribution of genomic DNA with no
628 treatment or fragmented by different volumes of pTXB1 Tn5, V50 or ATM. The dotted black line
629 indicates peak of fragment size. (B) Size distribution of mRNA/DNA hybrid with no treatment or
630 fragmented by different volumes of pTXB1 Tn5, V50 or ATM. (C) Sequencing indicators of SHERRY
631 library constructed by three Tn5.

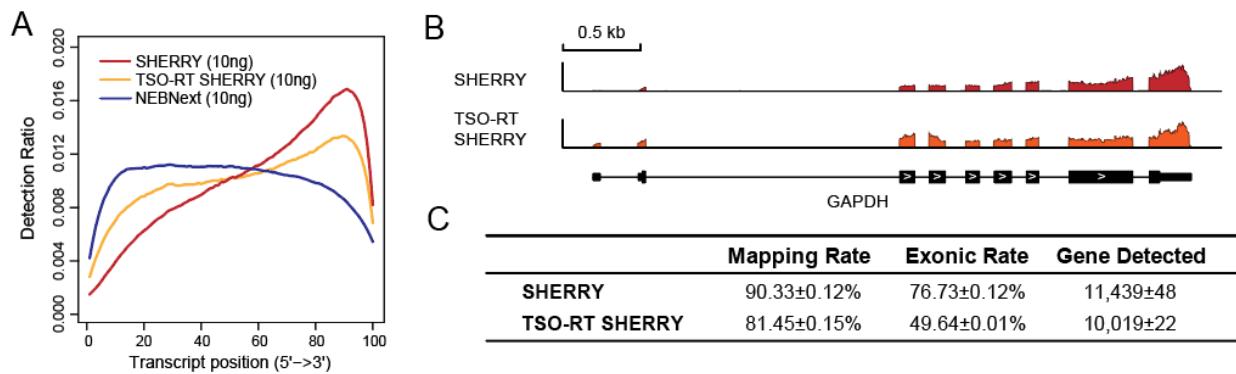
632



633
634

635 **Fig.S10** Titration of Tn5 transposome for fragmentation. **(A-B)** Size distribution of 5ng genomic DNA
636 or mRNA/DNA hybrid with no treatment or fragmented by different gradients of pTXB1 Tn5.

637
638
639
640
641
642
643



644
645

646 **Fig.S11** Coverage evenness optimization of SHERRY (10 ng HEK293T total RNA input). **(A)**
647 Normalized transcript coverage of standard SHERRY, SHERRY using TSO-RT method and NEBNext
648 kit. **(B)** The coverage of GAPDH transcript calculated from SHERRY and TSO-RT SHERRY. **(C)**
649 Comparison of sequencing indicators between SHERRY (n=3) and TSO-RT SHERRY (n=2).

650
651
652
653
654
655
656

657 **Materials and Methods**

658

659 **Purification of pTXB1 Tn5 and D188E mutation**

660 The pTXB1 cloning vector, which introduced hyperactive E54K and L372P mutation into
661 wildtype Tn5, was acquired from Addgene. The pTXB1 Tn5 and its mutant were
662 expressed and purified mainly according to the protocol published by Picelli S et al. [1] To
663 construct D188E mutation into Tn5, pTXB1 vector was firstly amplified into two parts by
664 two sets of primers. Mutagenesis primers used for the first part (3771-7979) which
665 contained site 188 were 5'-GGCAGCATGATGAGCAACGTGATTGCGGTGTGCGAACG
666 TGAAGCGGATATTCATGC-3' and 5'-TATCAGCTCACTCAAAGG-3'. Amplified primers
667 for the remaining part were 5'-GTATTACCGCCTTGAGT-3' and 5'-CAATCACGTTGCTC
668 ATCA-3'. The purified PCR products were then assembled into intact plasmid using
669 Gibson Assembly Master Mix (NEB, Cat.No. E2611). The newly assembled plasmid was
670 transformed into E.coli Trans5 α chemically competent cells (Transgene, CD201-01). After
671 growing overnight on LB medium plate, single colony was picked and shaken in SOC
672 liquid medium for at least 9 hours. Plasmid was extracted by PurePlasmid Mini Kit
673 (CW BIO, Cat.No.CW0500S) and confirmed carrying D188E mutation by Sanger
674 sequencing. Then the plasmid was transformed into E.coli Transetta (DE3) chemically
675 competent cells (Transgene, CD801-01) for further protein expression and purification.

676

677 **Cell culture**

678 HEK293T and HeLa cell lines are acquired from ATCC. Both of them were cultured in
679 Dulbecco's Modified Eagle Medium (Gibco, Cat.No.11965092), supplemented with 10%
680 fetal bovine serum (Gibco, Cat.No.1600044) and 1% penicillin-streptomycin (Gibco,
681 Cat.No.15140122). The cell incubator (Thermo Scientific) was set at temperature of 37°C
682 with 5% CO₂ injected.

683 Adherent cells were washed twice by DPBS (Gibco, Cat.No.14190136) and detached by
684 0.05% Trypsin-EDTA (Gibco, Cat.No.25300062) at 37°C for 4min. Then double volume of
685 culture media was added to terminate trypnization. Cells were collected by centrifugation
686 at 200g for 5min and resuspended for downstream experiment or passage cultivation.

687

688 **Nucleic acids extraction and messenger RNA isolation**

689 Genomic DNA was extracted using PureLink Genomic DNA Mini Kit (Invitrogen,
690 Cat.No.K182002), and total RNA was extracted using RNeasy Mini Kit (Qiagen,
691 Cat.No.74104). The resulting total RNA was then reacted with 10ul DNase I (NEB,
692 Cat.No.M0303) to remove remaining DNA thoroughly, and concentrated by RNA Clean &
693 Concentrator-5 kit (Zymo Research, Cat.No.R1015). The quality of extracted DNA and
694 RNA was assessed by the Fragment Analyzer Automated CE System (AATI) and
695 quantification was done by Qubit 2.0 (Invitrogen, Cat.No.Q33230/ Q32852).
696 We followed standard protocol of NEBNext Poly(A) mRNA Magnetic Isolation Module
697 (NEB, Cat.No.E7490) to isolate messenger RNA from the purified total RNA and stored
698 them at -80°C.

699

700 **Single cell preparation**

701 We used pipette tips to form some drops made up of PBS (containing 1% BSA) (Thermo
702 Scientific, Cat.No.37525) on a clean petri dish. The cell resuspension was pipetted up
703 and down gently to disperse into single cells and we took ~5 μ l of them diluted in one of
704 the drops. The mouth pipette with 50 μ m inside diameter was then used to pick one cell in
705 the drop and release it in another clean drop. The picked cell was passed by at least three
706 clean drops in order to wash away any debris and confirm that only one cell was in the
707 last drop. We then aspirated the cell with as little buffer as possible and blew it into 4 μ l
708 lysis buffer [4 units of Recombinant RNase Inhibitor (Takara, Cat.No.2313), 2.5 μ M
709 poly(T)₃₀VN primer (Sangon), 2.5mM dNTP (NEB, Cat.No.N0447) and 0.48% Triton X-
710 100 (Sigma, Cat.No.T9284)]. Successful transfer was confirmed by blowing mouth pipette
711 again in a clean drop and no cell was to be seen in visual field. Reaction was carried out
712 at 72°C for 3min after violent vortex.

713

714 **mRNA/DNA hybrid formation**

715 Total RNA was reverse transcribed into mRNA/DNA hybrid mainly referring to Smart-seq2
716 protocol [2], but with several modifications: 1) The ISPCR part in Oligo-dT primer was
717 removed; 2) TSO was omitted, but in TSO-RT SHERRY, it should be kept; 3) The reaction
718 was performed at 42°C for 1.5h without cycling. If input was purified RNA, Triton X-100
719 was omitted. When inputting more than 10ng total RNA, we would slightly upregulate

720 amount of dNTP, poly(T)₃₀VN primer and Superscript II (Invitrogen, Cat.No.18064014).

721

722 **Tn5 transposome in vitro assembly and tagmentation**

723 Functional mosaic-end (ME) oligonucleotides (5'-CTGTCTCTTATACACATCT-3', 100μM)
724 was separately annealed with equal amounts of Adaptor A (5'-TCGTCGGCAGCGTCAG
725 ATGTGTATAAGAGACAG-3', 100μM) and Adaptor B (5'-GTCTCGTGGGCTCGGAGATG
726 TGTATAAGAGACAG-3', 100μM). The concentration of purified Tn5 was quantified by
727 Qubit Protein Assay Kit (Invitrogen, Cat.No.Q33212) and took around 100μg for
728 transposome assembly. We mixed the Tn5 transposase with annealed ME-Adaptor A or
729 B (20μM) in 45% glycerol (Sigma, Cat.No.G5516) thoroughly and incubated the mixture
730 at 30°C for one hour. These two resulting transposomes (assembled with ME-Adaptor
731 A/B) were then mixed together, ready for tagmentation or stored at -20°C. Specifically, to
732 assemble rCrArG Tn5 in **Fig.S4**, ribonucleotide modifications were made on the three
733 terminal bases at 3'-end of Adaptor A/B. And for rG Tn5, the last base at 3'-end of adaptors
734 was modified.

735 The dsDNA tagmentation was performed in 1xTD buffer [10mM Tris-Cl (pH 7.6,
736 ROCKLAND, Cat.No.MB-003), 5mM MgCl₂ (Invitrogen, Cat.No.AM9530G), 10% N,N-
737 Dimethylformamide (Sigma, Cat.No.D4551)]. The reaction was incubated at 55°C for
738 30min.

739 As for RNA/DNA hybrid, tagmentation was performed in buffer containing 10mM Tris-Cl
740 (pH 7.6), 5mM MgCl₂, 10% N,N-Dimethylformamide, 9% PEG8000 (VWR Life Science,
741 Cat.No.97061), 0.85mM ATP (NEB, Cat.No.P0756). In SHERRY library preparation, we
742 used 0.05 μl, 0.006 μl and 0.003 μl Tn5 transposome for input of 200 ng, 10 ng and 100
743 pg total RNA, respectively. scSHERRY also used 0.003 μl Tn5 transposome.

744 The transposome could be diluted in 1xTn5 dialysis buffer [50mM Hepes (pH 7.2,
745 Leagene, Cat.No.CC064), 0.1M NaCl (Invitrogen, Cat.No. AM9759), 0.1 mM EDTA
746 (Invitrogen, Cat.No.AM9260G), 1 mM DTT, 0.1% Triton X-100, 10% glycerol]. The reaction
747 was incubated at 55°C for 30min.

748 Commercial Tn5 transposomes were available in Nextera XT DNA Library Prep Kit
749 (Illumina, Cat.No.FC-131-1024) and TruePrep DNA Library Prep Kit V2 for Illumina
750 (Vazyme, Cat.No.TD501).

751

752 **SHERRY library preparation and sequencing**

753 To construct scSHERRY library, the single cell tagmentation product was mixed well with
754 4 units of Bst 3.0 DNA Polymerase (NEB, Cat.No.M0374) and indexed common primers
755 (Vazyme, Cat.No.TD202) in 1 x Q5 High-Fidelity Master Mix (NEB, Cat.No.M0492). Then
756 index PCR was performed as follow: 72°C 15min, 98°C 30s, 10 cycles of [98°C 20s, 60°C
757 20s, 72°C 2min], 72°C 5min. The PCR product was purified with 0.85:1 ratio by VAHTS
758 DNA Clean Beads (Vazyme, Cat.No.N411) and eluted in 30μl nuclease-free water
759 (Invitrogen, Cat.No.AM9937) for another 18 cycles of PCR. When performing high-
760 throughput experiment, each sample could be amplified by 15 cycles, then merged for
761 beads purification and library quality check.

762 As for purified 10ng or 200ng total RNA input, the tagmentation product was firstly gap-
763 filled with 100 units of Superscript II and 1 x Q5 High-Fidelity Master Mix at 42°C for 15min,
764 then Superscript II was inactivated at 70°C for 15min. When inputting 100pg total RNA,
765 the extension enzyme was replaced with 4 units of Bst 2.0 Warmstart DNA Polymerase
766 (NEB, Cat.No.M0538). Correspondingly, the reaction temperature was upregulated to
767 72°C and inactivation was performed at 80°C for 20min. After that, indexed common
768 primers were added to perform PCR. We performed 12, 15 and 25 cycles of PCR for input
769 of 200 ng, 10 ng and 100 pg total RNA, respectively.

770 The resulting library was purified with 1:1 ratio by VAHTS DNA Clean Beads.
771 Quantification was done by Qubit 2.0 and quality check was done by Fragment Analyzer
772 Automated CE System. The sequencing platform we used was Illumina NextSeq 500 or
773 HiSeq 4000.

774

775 **NEBNext and SmartSeq2 library preparation**

776 NEBNext RNA-Seq library preparation starting from 10ng and 200ng total RNA was
777 performed using NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB, Cat.No.E7770).
778 Single cell SmartSeq2 library was constructed as previously reported [2].

779

780 **Ligation Test and Strand Test**

781 In Ligation Tests, tagmentation products from 200ng HEK293T total RNA were purified by

782 DNA Clean & Concentrator-5 (Zymo Research, Cat.No.D4013) and eluted with 20 μ l
783 nuclease-free water. After gap-filling with Superscript II, Ligation Test 1 was processed
784 directly to index PCR while product in Ligation Test 2 was digested by 12.5 units of RNase
785 H (NEB, Cat.No.M0297) at 37°C for 20min before index PCR.
786 In Strand Tests, we used dUTP (Thermo Scientific, Cat.No.R0133), dATP, dCTP, dGTP
787 (NEB, Cat.No.N0446) mix, each of them at equal concentration, to incorporate in cDNA
788 during reverse transcription step. 0.15 μ l Tn5 transposome was used to tagment the
789 resulting hybrid. Fragments were then column-purified and gap-filled by Bst 2.0 Warmstart
790 DNA Polymerase, and column purification was again applied. For Strand Test 1, 3 units
791 of USER enzyme (NEB, Cat.No.M5505) and 40units of recombinant rnase inhibitor was
792 added into elution products and incubated at 37°C for 20min for DNA strand digestion.
793 Indexed common primers added with digestion product in 1 x Q5 High-Fidelity Master Mix
794 were then reacted at 85°C for 30s, followed by 60°C for 2min and temperature went down
795 to 4°C slowly. After that, reverse transcription was performed with 200 units of Superscript
796 II added at 42°C for 30min, then transferred to index PCR program. For Strand Test 2,
797 the USER digestion product was directly performed index PCR with 1 x KAPA HiFi
798 HotStart Uracil+ ReadyMix (Kapa Biosystems, Cat.No.KK2801). Protocol of Strand Test
799 3 was almost same as Strand Test 2, except replacing USER enzyme with RNase H. For
800 dU-SHERRY, the USER enzyme digestion was omitted compared with Strand Test 2
801 workflow.

802

803 **Docking model**

804 To generate the substrate-transposon DNA-Tn5 structure model, a 32bp dsDNA or
805 DNA/RNA hybrid were generated by 3D-NuS (3-Dimensional Nucleic Acid Structures)
806 web server. Then the substrate dsDNA or DNA/RNA was manually docked to the
807 transposon DNA-Tn5 structure, PDB ID: 1MUS, based on charge and shape
808 complimentary.

809

810 **Data analysis**

811 Sequencing adaptors or poly(T/A) positioned at end of paired reads were recognized and
812 removed by Cutadapt v1.15 [3]. The trimmed reads which length was shorter than 20bp

813 were filtered. Remaining reads were down sampled to 2 million (except that library with
814 200ng total RNA input used for differential gene expression analysis was 10 million) total
815 reads, and aligned with index built from human(hg38) genome and known transcript
816 annotations by Tophat2 v2.1.1 [4]. The mapped reads were then used to calculate FPKM
817 value for each known gene (annotation acquired from UCSC) by Cufflinks v2.2.1 with
818 multi-mapped reads correction. Gene with FPKM more than 1 was considered to be
819 detected. The exonic rate, duplicate rate and insert size of library were all calculated by
820 Picard Tools v2.17.6.

821 General coverage across known transcripts was plotted by RSeQC v.2.6.4 [5]. For
822 specific transcript, depth of mapped reads overlapped with transcript position was
823 calculated by Samtools v1.3.1.

824 We used DESeq2 v1.22.2 [6] to perform differential gene expression analysis with raw
825 count-matrix acquired by HTSeq v.0.11.0 [7]. Differentially expressed genes should meet
826 following criteria: 1) FPKM value >1; 2) significant p-value <5x10⁻⁶; 3) absolute value of
827 log₂(Fold Change) >1. Counts in correlation plot were normalized mainly according to
828 DESeq2 normalization method [6], which considered library size and library
829 compensation. Correlation efficient R² and slope of linear fitting equations were calculated
830 by least square method. The slope-skewness between two replicates was defined as |k₁-
831 1|+|k₂-1|, k₁ or k₂ was slope when one of the replicates was conducted as X or Y axis.
832 GC content distribution of detected genes was plotted by custom Perl script. GC content
833 was binned by 2% and gene number at each bin was normalized by dividing maximum
834 gene number of one bin. And for each gene, only the longest transcript isoform was
835 calculated.

836

837 **Data deposition**

838 The sequence reported in this paper has been deposited in the Genome Sequence
839 Archive (accession no. CRA002081).

840

841 **References**

842 [1] S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5
843 transposase and tagmentation procedures for massively scaled sequencing projects.

- 844 Genome Res. 24, 2033-2040 (2014).
- 845 [2] S. Picelli, O. R. Faridani, A. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-
- 846 length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171-181 (2014).
- 847 [3] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing
- 848 reads. Embnet Journal 17, 200 (2011).
- 849 [4] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2:
- 850 accurate alignment of transcriptomes in the presence of insertions, deletions and gene
- 851 fusions. Genome Biol. 14, R36 (2013).
- 852 [5] L. Wang, S. Wang, W. Li, RSeQC: quality control of RNA-seq experiments.
- 853 Bioinformatics 28, 2184-2185 (2012).
- 854 [6] M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion
- 855 for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
- 856 [7] S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-
- 857 throughput sequencing data. Bioinformatics 31, 166-169 (2015).
- 858
- 859