

1 **Rapid and sensitive single cell RNA sequencing with SHERRY2**

2 Lin Di,^{1,2,3,4,#} Bo Liu,^{5,6,#} Yuzhu Lyu,¹ Shihui Zhao,^{2,7} Yuhong Pang,² Chen Zhang,¹ Jianbin
3 Wang,^{8,*} Hai Qi,^{5,6,*} Jie Shen,^{1,*} Yanyi Huang^{2,4,9,10,*}

4 ¹ School of Basic Medical Sciences, Beijing Key Laboratory of Neural Regeneration and Repair,
5 Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing
6 100069, China

7 ² Biomedical Pioneering Innovation Center, Beijing Advanced Innovation Center for Genomics,
8 Peking University, Beijing 100871, China

9 ³ School of Life Sciences, Peking University, Beijing 100871, China

10 ⁴ Institute for Cell Analysis, Shenzhen Bay Laboratory, Guangdong 528107, China

11 ⁵ Laboratory of Dynamic Immunobiology, Institute for Immunology, Tsinghua University, Beijing
12 100871, China

13 ⁶ Department of Basical Medical Sciences, School of Medicine, Tsinghua University, Beijing
14 100871, China

15 ⁷ Peking University–Tsinghua University–National Institute of Biological Sciences Joint
16 Graduate Program (PTN), Peking University, Beijing 100871, China

17 ⁸ School of Life Sciences, Beijing Advanced Innovation Center for Structural Biology, Tsinghua
18 University, Beijing 100084, China

19 ⁹ College of Chemistry and Molecular Engineering, Beijing National Laboratory for Molecular
20 Sciences, Peking University, Beijing 100871, China

21 ¹⁰ Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
22 Paste the full affiliation list here.

23 [#] These authors contributed equally.

24 * Corresponding authors: yanyi@pku.edu.cn (Y.H.), shenjie@ccmu.edu.cn (J.S.),
25 qihai@tsinghua.edu.cn (H.Q.) and jianbinwang@tsinghua.edu.cn (J.W.)

26

27 **Abstract**

28 Prevalent single cell transcriptomic profiling (scRNA-seq) methods are mainly based on
29 synthesis and enrichment of full-length double-stranded complementary DNA. These
30 approaches are challenging to generate accurate quantification of transcripts when their
31 abundance is low or their full-length amplifications are difficult. Based on our previous finding
32 that Tn5 transposase can directly cut-and-tag DNA/RNA hetero-duplexes, we present
33 SHERRY2, a specifically optimized protocol for scRNA-seq without second strand cDNA
34 synthesis. SHERRY2 is free of pre-amplification and eliminates the sequence-dependent
35 bias. In comparison with other widely-used scRNA-seq methods, SHERRY2 exhibits
36 significantly higher sensitivity and accuracy even for single nuclei. Besides, SHERRY2 is
37 simple and robust, and can be easily scaled up to high-throughput experiments. When testing
38 single lymphocytes and neuron nuclei, SHERRY2 not only obtained accurate countings of
39 transcription factors and long non-coding RNAs, but also provided bias-free results that
40 enriched genes in specific cellular components or functions, which outperformed other
41 protocols. With a few thousand cells sequenced by SHERRY2, we confirmed expression and

42 dynamics of *Myc* in different cell types of germinal centers, which were previously only
43 revealed by gene-specific amplification methods. SHERRY2 is able to provide high sensitivity,
44 high accuracy, and high throughput for those applications that require high number of genes
45 identified in each cell. It can reveal the subtle transcriptomic difference between cells and
46 facilitate important biological discoveries.

47
48 **Keywords:** single cell | RNA-seq | Tn5 transposase

49 **Background**

50 Many experimental methods for transcriptome profiling by next generation sequencing (RNA-
51 seq) have been developed to cover various scales of input samples, ranging from bulk samples
52 [1, 2] to single cells [3-5] or even subcellular components [6, 7]. High quality single-cell RNA-
53 seq (scRNA-seq) data can be used to reveal the kinetic details of gene expression and
54 transitions between cell states or types [8-10]. Prevalent scRNA-seq methods mainly rely on
55 template switching and pre-amplification of complementary DNA (cDNA). However, large-scale
56 scRNA-seq techniques, commonly operated in micro-droplets or wells, have relatively low
57 sensitivity [11]. Single-tube based scRNA-seq approaches can typically produce higher
58 coverage for low-abundance genes, but they still suffer from quantification bias due to
59 insufficient reverse transcription and GC imbalance during amplification. Besides, their complex
60 experimental methods are generally unsuitable for large-scale studies.

61 We have reported a highly reproducible and rapid library preparation method for RNA-seq,
62 SHERRY, which can be applied to minute amount of RNA samples [12]. The development of
63 SHERRY was based on the recent discovery that Tn5 transposase can bind and cut RNA/DNA
64 hetero-duplexes directly. With slight modifications, SHERRY could also be applied to various
65 clinical metatranscriptome applications, such as identification of SARS-CoV-2 and other
66 pathogens [13].

67 Although SHERRY was applied to process single cells and achieved less biased
68 quantification of gene expression in comparison with other scRNA-seq methods, the results still
69 exhibited clear coverage bias toward the 3' -ends of transcripts, relatively low sensitivity, and
70 low tolerance to endogenous DNA. In this work, we present an optimized method, SHERRY2,
71 which addresses the limitations of SHERRY and is fully compatible with single cells and single
72 nuclei with low RNA content. In comparison with prevalent RNA-seq methods, SHERRY2
73 showed higher sensitivity, better concordance with reference data, greater reproducibility
74 between replicates, and superior scalability, allowing the method to be used to process a few
75 thousand single cells per batch and thus reducing the time required to conduct experiments.

76 **Results**

77 **SHERRY2 provides high sensitivity and even coverage across gene bodies for scRNA-
78 seq.**

79 For scRNA-seq, RNA degradation and incompleteness of reverse transcription (RT) are two
80 major factors that reduce gene detection sensitivity and coverage evenness. Although adding
81 random RT primers facilitates the coverage of long transcripts, it requires removal of ribosomal

82 RNA, which is incompatible with scRNA-seq [13]. Spiking template-switching oligonucleotides
83 also provides more uniform coverage, but this strategy has limited detection sensitivity and
84 specificity [12].

85 We altered various experimental parameters of the original SHERRY protocol for both bulk
86 (**Additional file 1: Fig. S1-2, Additional file 2-3**) and single cell inputs (**Additional file 1: Fig.**
87 **S3A**). To protect RNA from degradation, we lowered the concentration of free Mg²⁺, either by
88 reducing the amount of total Mg²⁺ or adding more dNTP to chelate Mg²⁺ ions [14], and observed
89 significant improvement of the coverage evenness of RNA-seq. To facilitate cDNA synthesis,
90 we screened different reverse transcriptases and found that SuperScript IV (SSIV), working at
91 a relatively high temperature with a low Mg²⁺ concentration, could better overcome the
92 secondary structure of RNA and hence simultaneously enhanced the sensitivity and uniformity
93 of RNA-seq.

94 When RNA-seq was conducted using pictogram-level RNA inputs, sufficient amount of
95 Tn5 transposome was important for high sensitivity, and Bst 3.0 DNA polymerase filled the gap
96 left by Tn5 fragmentation more effectively than other enzymes. The protocol was insensitive to
97 many experimental conditions, including the usage of single strand DNA binding proteins [15],
98 Tn5 inactivation, the concentration of extension polymerase, and the usage of hot-start
99 polymerase.

100 We named this optimized protocol SHERRY2. Using RNA extracted from HEK293T cells
101 as input, we compared the performance of SHERRY2 and the original SHERRY protocol. At
102 the 10-*ng* level, both protocols identified more than 11,000 genes at saturation. At the 100-*pg*
103 level, SHERRY2 performed better than SHERRY and detected 5.0% more genes at 0.6-million
104 reads (**Additional file 1: Fig. S2A**). In addition, SHERRY2 greatly diminished 3'- end coverage
105 bias (**Additional file 1: Fig. S2B**) and increased the unique mapping rate for 10-*ng* and 100-*pg*
106 inputs (**Additional file 1: Fig. S2C**). We also constructed a bias-free RNA-seq library using
107 200-*ng* total RNA input via the conventional fragmentation-and-ligation method with the
108 NEBNext E7770 kit (NEBNext). For 100-*pg* input, the gene overlap between NEBNext and
109 SHERRY2 was greater than that between NEBNext and SHERRY (81.7% vs 78.4%)
110 (**Additional file 1: Fig. S2D**), and the gene expression results of NEBNext and SHERRY2 were
111 also more closely correlated ($R=0.70$ vs $R=0.65$) (**Additional file 1: Fig. S2E**).

112 The SHERRY2 protocol for scRNA-seq contains only four steps: reverse transcription, Tn5
113 fragmentation, gap-filling through extension, and PCR amplification. The entire SHERRY2
114 protocol can be completed within 3 hours, one hour less than the original SHERRY protocol,
115 and still held its competence in costs (**Additional file 1: Fig. S3B**). Other high-sensitivity
116 scRNA-seq methods such as SmartSeq2 may require much more time and more steps to be
117 completed [3] (**Fig. 1A**). The one-tube workflow of SHERRY2 is readily scalable to high-
118 throughput applications. SHERRY2 was able to detect 10,024 genes (FPKM >1) on average
119 within a single HEK293T cell at 1-million reads. When subsampling to 0.2-million reads,
120 SHERRY2 still detected 8,504 genes on average, which was 1,622 (23.6%) more than
121 SHERRY and 886 (11.6%) more than SmartSeq2 (**Fig. 1B**). In addition, the reproducibility of
122 SHERRY2 was significantly higher than that of SHERRY or SmartSeq2 (**Fig. 1C**) due to its
123 simplified workflow and stable performance. Moreover, the evenness of gene body coverage
124 for SHERRY2 was much higher than that of the original SHERRY protocol (0.84 vs 0.72) and
125 was comparable to that of SmartSeq2 (0.84) (**Fig. 1D**). The exonic rate of SHERRY2 was also

126 improved in comparison with that of SHERRY, likely due to the higher RT efficiency of the newly
127 developed method (**Fig. 1E**).

128 Last but not least, scRNA-seq with SHERRY2 exhibited superior accuracy, as
129 demonstrated by the significantly higher correlation between the SHERRY2 gene expression
130 results and NEBNext libraries in comparison with that of SmartSeq2 ($R=0.71$ vs $R=0.67$) (**Fig.**
131 **1F**), since NEBNext fragmented mRNA before cDNA synthesis and amplified cDNA with very
132 limited cycles which theoretically resulted in negligible bias at transcriptome level. Especially,
133 SHERRY2 showed high tolerance to GC content and was insensitive to the length of transcripts
134 (**Additional file 1: Fig. S4**). Unlike SmartSeq2, for which the gene overlap and expression
135 correlation with bulk RNA-seq showed clear declines when GC-content was greater than 40%,
136 SHERRY2 maintained these parameters at high and constant levels (82.6% overlap and
137 $R=0.76$) until the GC content reached 60%. Transcript length did not influence the accuracy of
138 SHERRY2 or SmartSeq2, although SmartSeq2 exhibited a small degree of intolerance for
139 transcripts longer than 800 bases.

140

141 **scRNA-seq for low RNA-content cells.**

142 For low RNA-content cells, such as immune cells [16], we found that removal of intergenic DNA
143 contaminations by DNase treatment was especially crucial for SHERRY2 scRNA-seq. In such
144 cells, the open DNA regions of disassembled chromatin might be favored over RNA/DNA
145 hybrids during Tn5 fragmentation. When DNase was omitted from the SHERRY2 protocol, more
146 than 50% of reads sequenced from single mouse lymphocytes (**Additional file 1: Fig. S5A**)
147 were mapped to intergenic regions, and only around 10% of reads were exonic reads (**Fig. 2A**).

148 Different DNases performed differently in SHERRY2 scRNA-seq. We tested five DNases
149 (**Additional file 1: Fig. S6A**) and found three (NEB, Ambion, and TURBO DNase I) that worked
150 and inactivated at higher temperatures increased the intergenic rate unexpectedly, and this
151 effect was probably due to RNA degradation at high temperatures with excess Mg^{2+} in the
152 reaction buffer. In contrast, AG DNase I and gDW mix, which worked at room temperature,
153 yielded ideal results.

154 We confirmed that all the five DNases could digest more than 99.5% of DNA (30- μ g) by
155 simply utilizing divalent ions of their respective storage buffer (**Additional file 1: Fig. S6B**,
156 **Additional file 4**). Without adding extra divalent ions, the intergenic rates of single germinal
157 center (GC) B cells for all DNases were less than 20% (**Fig. 2C**). Among the DNases, AG
158 DNase I retained high sensitivity for gene detection, and more than 60% of reads were mapped
159 to exon regions (**Fig. 2D**), while the evenness of coverage was not affected (**Fig. 2E**).

160 Next, dU-containing carrier DNA, which could not be amplified by dUTP-intolerant
161 polymerase, was added to further improve the efficiency of fragmentation of RNA/DNA hybrids.
162 With carrier DNA, SHERRY2 detected 3,200 genes at saturation (0.6-million reads) for single
163 GC B cells (**Fig. 2F**), and the number of detectable genes increased from 2,301 to 2,393 on
164 average for single lymphocytes, with an exonic ratio comparable to that of SmartSeq2 (**Fig. 2A**,
165 **2B**). Moreover, we examined the genes that were only detected by one method for single GC
166 B cells, and found that SmartSeq2 was preferential to capture genes participated in
167 mitochondrial function (**Fig. 2G**). Based on these results, chromatin digestion and the addition
168 of carrier DNA were included in the standard SHERRY2 protocol and the step of chromatin
169 digestion would consume another 20 minutes.

170

171 **Selection dynamics in germinal centers profiled by SHERRY2.**

172 SHERRY2 can be easily scaled to thousands of single cells per batch, owing to its simplified
173 procedure. The GC is a transient structure that supports antibody affinity maturation in response
174 to T cell-dependent antigens, and it contains diverse cell types with complex dynamics.
175 Histologically, the GC can be separated into two micro-compartments, the dark zone and the
176 light zone [17, 18]. By surface phenotyping, cells in the two compartments can be distinguished
177 through CXCR4, CD83 and CD86 markers [19-21], with light-zone cells being
178 CXCR4⁺CD83⁺CD86⁺ while dark-zone cells CXCR4⁺CD83⁺CD86⁺. GC cells cycle between the
179 dark zone and light zone states. Dark zone cells are highly proliferative and undergo somatic
180 hypermutation, which generates a range of affinities against antigens. In the light zone, these
181 B cells compete with each other for survival factors and help signals, which are mainly derived
182 from follicular helper T cells. Those B cells that have acquired higher-affinity B cell receptors
183 are selected to differentiate into plasma cells (PC) or memory B cells (MBC) or cycle back to
184 the dark zone [18, 22-24]. Recently, gray zone, consisting of CXCR4⁺CD83⁺ cells with distinct
185 gene expression patterns, was discovered and found to be involved in GC recycling [25]. The
186 complex spatiotemporal dynamics of the GC and their underlying mechanisms are incompletely
187 understood. To this end, sensitive scRNA-seq methods that can be used to detect gene
188 expression with less bias are highly desirable.

189 We profiled 1,248 sorted CXCR4⁺CD86⁺ GC light zone cells with SHERRY2, and 1,231
190 (98.6%) high-quality cells were retained for downstream analysis (**Additional file 1: Fig. S5B**).
191 The gene expression levels of *Cd19*, *Ccnd3*, *Fas*, *Cd86* and *Cxcr4* were consistent with flow
192 cytometry gating (**Additional file 1: Fig. S7A**), and no batch effect was observed (**Additional**
193 **file 1: Fig. S7B**).

194 Unsupervised clustering identified seven clusters (**Fig. 3A**), two of which belonged to the
195 gray zone, which was defined by co-expression of *Cxcr4* and *Cd83*, as well as the on-going cell
196 division (enriched *Ccnb1*) [25] (**Fig. 3B**). We observed the expected down-regulation of *Bcl6*
197 and *S1pr2*, the signature genes of GC B cells [26, 27], in memory B cell precursors (MPs) and
198 plasma cell precursors (PPs). Specifically, *Ccr6* was exclusively enriched in MPs [28], while *Irf4*
199 was up-regulated in PPs, which was known to be mediated by NF- κ B pathway downstream of
200 *Cd40* stimulation [24]. It's worth noting that our results exhibited such *Cd40* signaling effects as
201 well (**Additional file 1: Fig. S7C**). Besides, *Icam1* and *Slam1* which were reported to be
202 activated by *Cd40* [29] were also observed (**Additional file 1: Fig. S7D**, **Additional file 5**). The
203 relatively low expression levels of *Prdm1* (not shown) and *Gpr183* in PPs were consistent with
204 the early stage of plasma cell development. In total, 1,071 genes significantly up- or down-
205 regulated in specific clusters were identified.

206 The high sensitivity of SHERRY2 enabled detection of *Myc* in 588 (47.8%) single GC light
207 zone B cells. Using fluorescent protein reporting, *Myc* was found to mark light-zone cells
208 destined for dark zone re-entry [30], although *Myc* expression *per se* had been difficult to
209 identify in specific cell types by low-sensitivity scRNA-seq approaches [31]. Consistent with
210 previous findings [25, 29], *Myc* expression was significantly higher in PPs (**Fig. 3B**, **Additional**
211 **file 1: Fig. S7E**) and active in the gray zone cells (**Fig. 3C**). Light Zone-1 also had a relatively
212 higher portion of *Myc*⁺ cells, which are probably those destined for cyclic re-entry to the dark
213 zone [30]. MPs also contained some cells that expressed *Myc*.

214 RNA velocity analysis (**Fig. 3D**) suggested that Light Zone-1 contained cells selected for
215 dark zone re-entry, which were migrating to the gray zone and had *Myc* expression
216 characterized by burst kinetics (**Additional file 1: Fig. S7F**). In addition, cells that appeared to
217 have just entered the light zone were also identified. A few velocity vectors that moved to MPs
218 were mixed in PPs, and these vectors were in the same direction with the down-regulation of
219 *Myc*. According to the velocity analysis, the aforementioned *Myc*-expressing MPs seemed to
220 have a tendency to cycle back to the GC, suggesting that some MPs with *Myc* up-regulation
221 have the potential to re-participate in GC reactions.

222 We then assembled the BCR sequence for each cell to screen the usage of IgH variable
223 sequences, which were assigned in 1,101 (89.4%) cells. As expected [32], IGHV1-72
224 dominated the NP-reactive GC response, and the coupled light chain was mainly IgL rather
225 than IgK (**Additional file 1: Fig. S8A, S8B**). In addition, we identified CDR1 and CDR2 regions
226 in 269 (24.4%) and 493 (44.8%) cells in which IgH variable sequences were assigned,
227 respectively (**Additional file 1: Fig. S8C**).

228 SHERRY2 revealed differences in the usage frequencies of exons across cell types. The
229 usage of a particular exon (chr11: 51,601,750-51,601,890) within the *Hnrnpab* transcript (**Fig.**
230 **3E**), which is widely expressed and encodes a protein that mainly functions in processing pre-
231 mRNAs, was significantly biased among GC clusters. As shown in **Fig. 3F**, Light Zone-1 cells
232 favored inclusion of this exon.

233

234 **Superior performance of SHERRY2 applied in snRNA-seq.**

235 Single nucleus RNA-seq (snRNA-seq) has gained popularity since fresh and intact single cells
236 are challenging to obtain in many applications. Hence, we tested the performance of SHERRY2
237 on snRNA-seq using single nuclei isolated from HEK293T cells. SHERRY2 detected 10,137
238 genes (RPM>1) on average at 1-million reads, which was 4,330 (74.6%) more than SmartSeq2,
239 demonstrating that SHERRY2 had superior sensitivity for single nuclei (**Fig. 4A**). SHERRY2
240 still exhibited superior accuracy as it was significantly more correlated with NEBNext
241 quantification results in comparison with SmartSeq2 ($R=0.41$ vs $R=0.39$) (**Fig. 4B**).

242 The high accuracy and sensitivity of SHERRY2 allowed better distinction between
243 HEK293T cells and their nuclei, which had minimal differences. We performed principal
244 component analysis (PCA) using RNA-seq data from NEBNext, SHERRY2 and SmartSeq2
245 (**Fig. 4C**). Single cells and nuclei prepared by SHERRY2 were much closer in distance to the
246 bulk RNA results in comparison with those prepared with SmartSeq2. In addition, the
247 expression pattern of the differential genes identified by SHERRY2 was more similar to that of
248 NEBNext in comparison with SmartSeq2 (**Additional file 1: Fig. S9**).

249 Furthermore, we compared the performance of these two methods with hippocampal
250 neurons since snRNA-seq is a popular method for studies of brain tissue due to the technical
251 challenge of isolating intact single neurons. We constructed snRNA-seq libraries of frozen and
252 freshly prepared hippocampus with SHERRY2 and SmartSeq2. For both samples, SHERRY2
253 detected significantly more genes than SmartSeq2 (6,600 vs 5,331 at 1-million reads for frozen
254 samples, 6,686 vs 5,769 at 1-million reads for fresh samples) (**Fig. 4D**). And still, Smart-seq2
255 tended to detect genes functionized in mitochondrion (**Additional file 1: Fig. S10A**). Next, we
256 sequenced a small number of fresh hippocampal neurons (176 nuclei) (**Additional file 1: Fig.**
257 **S10B**) with SHERRY2 and classified their cell types correctly. The nuclei were non-

258 supervisedly clustered into 4 distinct groups (**Fig. 4E**), after which they were re-clustered using
259 marker genes identified by sNuc-Seq [33] (**Additional file 1: Fig. S10C**). The two clustering
260 results were highly consistent. Neurons within dentate gyrus (DG) and CA1, which occupy large
261 areas of hippocampus, could be assigned to Cluster 0 and Cluster 1 respectively, according to
262 the high expression of *Dock10*, *Slc4a4* and high expression of *Pex5l* and *Hs6st3* (**Fig. 4F**).
263 However, CA3 pyramidal cells were not shown in our results, probably due to the small number
264 of samples. Cluster 3 that were featured with enriched *Arx* and *Lhx6* could be annotated as
265 GABAergic cells, which migrated from medial ganglionic eminence (MGE). Except for the
266 aforementioned markers, the expression patterns of these three clusters acquired from sNuc-
267 seq and SHERRY2 were very similar (**Additional file 1: Fig. S10D**). Cluster 2 was found to
268 consist of cells with relatively high expression of *Dpp10* and *Tshz2*, inferring that it might be
269 contamination of cortex neurons. Moreover, our results revealed a long non-coding RNA
270 (lncRNA) cluster [34] containing *Meg3*, *Rian* (*Meg8*) and *Mirg* (*Meg9*), which showed higher
271 density in CA1 pyramidal cells and GABAergic cells while was relatively sparse in DG granule
272 cells (**Additional file 1: Fig. S10E**).

273 Discussion

274 SHERRY2 is a major improvement of our previously developed SHERRY [12], a Tn5
275 transposase-based RNA-seq method that eliminates the second-strand complementary DNA
276 synthesis. Although the original SHERRY protocol has shown satisfactory simplicity to
277 construct RNA-seq libraries using low amount of starting material, the coverage bias at 3'-ends
278 of transcripts and fragmentation-prone DNA contaminant make it challenging to work with single
279 cells. In MINERVA [13], a derivative of SHERRY that specifically designed to work for
280 metatranscriptome of COVID-19 clinical samples, we have explored the various conditions to
281 reduce the DNA coverage. In SHERRY2, we further optimized the DNA reduction process and
282 lead to a new protocol that can work for single cells and single nuclei, providing uniform
283 coverage of whole transcripts and resists to DNA contents.

284 There are three major advantages that SHERRY2 holds. First, SHERRY2 exhibits superior
285 sensitivity and accuracy compared with SmartSeq2, a prevalent scRNA-seq method. What's
286 more, from sequencing data of single GC B cells and single neuron nuclei, we found that
287 SmartSeq2 biasedly detected genes involved in mitochondrial components. Though more genes
288 were obtained by SHERRY2, there was no specific functional enrichment of these genes (**Fig.**
289 **2G, Additional file 1: Fig. S10A**). Thus, SHERRY2 would have more chance to facilitate
290 biological discoveries that relied on subtle changes. Recently SmartSeq3 [35], the upgraded
291 protocol of SmartSeq2, has been reported to increase the gene detection sensitivity. We have
292 also compared the scRNA-seq data of HEK293T cells produced by SHERRY2 and SmartSeq3.
293 SHERRY2 is able to detect over 10,000 genes at around 1 million reads, while SmartSeq3
294 cannot acquire same number of genes even at 3-fold of sequencing depth (**Additional file 1:**
295 **Fig. S11A**). Second, SHERRY2 retains great simplicity and expeditiousness, with the entire
296 workflow taking around 3 hours and with all reactions performed in one tube. The swift
297 experimental pipeline ensures less RNA degradation, eliminates the operational errors, and
298 saves costs of supplies and labor. Third, SHERRY2 is highly robust and scalable. Procedural
299 simplification not only reduces error cascade through step-wise operations, but also increases
300 the tolerance of pipetting by offering easily-handled volumes, leading to a significantly higher

301 repeatability when in comparison with SmartSeq3 (**Additional file 1: Fig. S11B**). Besides,
302 SHERRY2 contains richer information about exon junctions and coding regions across full
303 length transcripts, probably because SmartSeq3 is specifically optimized to quantify 5'-end of
304 transcripts (**Additional file 1: Fig. S11C, S11D**).

305 SHERRY2 can be further developed to uncover more information from single cells. The
306 simplicity and tolerance of protocol make it an ideal component to be incorporated into multi-
307 omics studies. Moreover, since SHERRY2 actually contains the strand-specific information of
308 the transcript since it builds libraries from RNA/DNA duplex directly. Therefore, SHERRY2 can
309 be potentially modified to differentiate the transcriptional strand of DNA. In addition, barcoded
310 Tn5 tagmentation [36, 37] may also be applied to SHERRY2 to realize assembling full-length
311 RNA molecules. Interestingly, when examining reads generated by SHERRY2 and SmartSeq2,
312 we find that the cleavage sites of Tn5 tend to exhibit different sequence bias on substrate DNA
313 and RNA/DNA duplex, which might give hints to understand Tn5 mechanism (**Additional file**
314 **1: Fig. S12**).

315 There are a few remaining hitches of current SHERRY2 protocol that need to be fixed in
316 the future. The slightly unsatisfactory mapping rate may be compensated by slightly more
317 sequencing reads. Without cDNA enrichment, the exogenous DNA from environment or
318 reagents still can be introduced after lysis step and easily tagged by Tn5, thus impairing the
319 performance of RNA-seq of low RNA content single cells or nuclei. Besides, it's still challenging
320 to capture the complete 5'-end regions of transcripts for the limited processivity of reverse
321 transcriptase. For example, CDR1 and CDR2 sequence in Ig variable regions cannot be
322 acquired for all GC cells (**Additional file 1: Fig. S8C**).

323 **Conclusions**

324 We present SHERRY2, an RNA-seq method designed for single cells and single nuclei.
325 SHERRY2 is based on the direct tagmentation function of Tn5 transposase for RNA/DNA
326 hetero-duplexes, and overthrows prevalent single cell RNA-seq chemistries which typically
327 require pre-amplification of full-length transcripts, thus greatly improving the sensitivity of gene
328 detection and eliminating the sequence-dependent bias. As a result, SHERRY2 can reveal
329 expression dynamics of transcription factors and lncRNAs, both of which typically harbor
330 essential biological functions while at low abundance. Meanwhile, SHERRY2 maintains the
331 simplicity of operation, with whole process completed in one pot within 3 hours, and hence
332 elevates the throughput to a few thousand single cells/nuclei per experimental batch. As the
333 simplest protocol of large-depth scRNA-seq, SHERRY2 has been validated in various
334 challenging samples, and can be seamlessly integrated into wide range of applications.

335 **Methods**

336 **Cell culture**

337 HEK293T cell line was purchased from ATCC and incubated at 37°C with 5% CO₂ in Dulbecco's
338 Modified Eagle Medium (DMEM) (Gibco, 11965092), which was supplemented with 10% fetal
339 bovine serum (FBS) (Gibco, 1600044) and 1% penicillin-streptomycin (Gibco, 15140122). Cells
340 were dissociated by 0.05% Trypsin-EDTA (Gibco, 25300062) at 37°C for 4min and washed by
341 DPBS (Gibco, 14190136).

342 For DNA or RNA extractions, we took ~10⁶ suspended cells, and followed the guideline of
343 PureLink Genomic DNA Mini Kit (Invitrogen, K182002) or RNeasy Mini Kit (Qiagen, 74104).
344 The extracted RNA was further dealt with 20U DNase I (NEB, M0303) for removal of DNA and
345 re-purified by RNA Clean & Concentrator-5 kit (Zymo Research, R1015).

346 For single nuclei preparation, we followed the guideline of Nuclei EZ Prep kit (Sigma, NUC-
347 101) and resuspended the nuclei into DPBS. Both single cells and single nuclei were sorted by
348 FACS Aria SORP flow cytometer (BD Biosciences).

349

350 **Mice**

351 For samples of germinal center B cells, C57BL/6 mice were originally from the Jackson
352 Laboratory. 6-12 week-old, age- and sex-matched mice were used for the experiments.

353 For samples of hippocampus nuclei and lymphocytes, aged (2-months old) male C57BL/6
354 mice were used and obtained from Charles River Laboratories.

355 All mice were maintained under specific pathogen-free conditions and used in accordance
356 of governmental, Tsinghua University and Capital Medical University guidelines for animal
357 welfare.

358

359 **GC light zone B cells preparation and sorting**

360 To generate T-cell dependent GC responses in B6 mice, 100µg NP-KLH (Biosearch
361 Technologies, N-5060-5) plus 1µg LPS (Sigma, L6143) emulsified in 100µl 50% alum (Thermo,
362 77161) were utilized for intraperitoneal immunization.

363 Spleens isolated from 4 mice of 13-days post immunization were placed on a 70µm cell
364 strainer (Falcon, 08-771-2), which was previously wetted with MACS buffer (1% FBS and 5mM
365 EDTA in PBS), and minced by flat end of the plunger of 2ml syringes (Becton Dickinson,
366 301940). The splenocytes passed through the strainer with MACS buffer into a 50ml-tube. The
367 mixed red blood cells were then lysed by ACK lysis buffer (Thermo, A1049201). The cell
368 suspension was further incubated with biotinylated 4-Hydroxy-3-iodo-5-nitrophenylacetyl
369 (NIP)15-BSA (Biosearch Technologies, N-1027-5) for 1.5h, and enriched by Anti-biotin cell
370 isolation kit (Miltenyi Biotec, 130-090-485) to get NP-reactive cells.

371 The enriched cells were blocked with 20µg/ml 2.4G2 antibody (BioXCell, BE0307) and
372 subsequently stained with APC-Cy7 (anti-B220, BD Biosciences, 552094), PE-Cy7 (anti-CD95,
373 BD Biosciences, 557653), eF450 (anti-GL7, eBioscience, 48-5902-82), APC (anti-CD86,
374 eBioscience, 17-0862-82) and PE (anti-CXCR4, BioLegend, 146505). Also, 7-AAD (Biotium,
375 40037) was stained to exclude dead cells. All staining reactions were incubated in MACS
376 staining buffer (1% FBS and 5mM EDTA in PBS) for 30min on ice, followed by 3 times of
377 washings. As gated in **Additional file 1: Fig. S5B**, single GC Light Zone B cells (B220⁺ GL7⁺
378 Fas⁺ CD86⁺ CXCR4⁻) were sorted into lysis buffer using Aria III flow cytometer (BD
379 Biosciences).

380

381 **Lymphocyte cells preparation and sorting**

382 The retro-orbital blood was taken from the eyeball of ether-anesthetized mice and dipped into
383 K2EDTA tube (BD Vacutainer, 367525). PBS was added to dilute blood at ~50%. 1ml diluted
384 blood was transferred into a clean 15ml-tube and incubated with 9ml 1x red blood cells lysing
385 solution (BD Pharm Lyse, 555899) at room temperature for 15min avoiding light. The resulted

386 cell suspension was washed twice by PBS containing 1% BSA at 200g for 5min, followed by
387 staining with SYTOX green (Thermo, S7020) to identify intact cells. Single lymphocytes were
388 sorted with FACSaria SORP flow cytometer according to the gates shown in **Additional file 1:**
389 **Fig. S5A.**

390

391 **Hippocampal nuclei preparation and sorting**

392 The isolated hippocampus tissue was transferred into a Dounce homogenizer (Sigma, D8938)
393 containing 2ml of EZ Lysis Buffer (Sigma, NUC-101). The tissue was carefully dounced for 22
394 times with pestle A followed by 22 times with pestle B, then transferred to a 15ml-tube. Next,
395 1ml of EZ lysis buffer was added into the Dounce homogenizer to resuspend residual nuclei,
396 then transferred to the same 15ml tube. The samples were centrifuged at 300g for 5 min.
397 Supernatant was removed and the pellet was resuspended in 100 μ l of ice-cold PBS (Gibco,
398 10010023) with 1% BSA (NEB, B9000S) and 20U RRI (Takara, 2313). 40 μ m FlowMi cell
399 strainers were firstly wetted with PBS and filtered the resuspended nuclei into 1.5 ml Eppendorf
400 tubes. The nuclei were further washed by PBS (1% BSA).

401 To enrich neuron nuclei, 1,000-fold diluted mouse Anti-NeuN antibody (Millipore, MAB377)
402 was added into 0.5ml nuclei suspension and incubated with the nuclei at 4°C for 30min. The
403 nuclei were then stained with 1000-fold diluted Goat anti-Mouse IgG (H&L) antibody (Abcam,
404 ab150113) and washed with PBS (1% BSA). The whole process was on ice. As gated in **Fig.**
405 **S10B**, single neuron nuclei were sorted with FACSaria SORP flow cytometer.

406 For frozen samples, hippocampus tissues were previously flash frozen in liquid nitrogen,
407 and stored in -80°C. Before single nuclei preparation, they were thawed on ice totally.

408

409 **DNA carrier preparation**

410 100-ng pTXB1 plasmids were firstly linearized by 10U XbaI (NEB, R0145S) at 37°C for 1h and
411 purified by Zymo columns. Then we took 0.5-ng linearized plasmids for multiple displacement
412 amplification (MDA), with all dTTPs replaced by dUTPs. Specifically, the 1 μ l DNA was mixed
413 with 22 μ l reaction buffer containing 1x phi29 reaction buffer (NEB, M0269S), 20 μ M random
414 primers (Thermo, SO181) and 1mM dNTP (NEB, N0446S and N0459S), then they were
415 incubated at 98°C for 3min and immediately cooled down at 4°C for 20min. 2 μ l phi29 DNA
416 polymerase was added and the amplification was carried out at 30°C for 5h, terminated at 65°C
417 for 10min. The products were purified by Zymo columns.

418

419 **Generation of RNA-seq library**

420 We constructed NEBNext libraries with 200- and 10-ng RNA by following the guideline of
421 NEBNext Ultra II RNA Library Prep Kit for Illumina kit (NEB, E7770). SmartSeq2 libraries with
422 single cells were prepared following the protocol that was reported by Picelli, S. *et al* [3]. 10X
423 libraries of 10,000 single hippocampal nuclei were constructed by Chromium Single Cell 3'
424 Reagent Kits (v3.1).

425 For scRNA-seq library of SHERRY2, single cells were sorted into 96-well plates containing
426 2 μ l lysis buffer which consisted of 0.5% Triton X-100 (Sigma, T9284), 2U SUPERaseIn RNase
427 Inhibitor (Thermo, AM2694), 0.2U AG DNase I (Thermo, 18068015). The plates were
428 immediately spun down and incubated at 20°C for 10min for DNA digestion. The plates could
429 be stored at -80°C or proceeded with next step. 2 μ l inactivation buffer containing 5 μ M OligoTs

430 (T₃₀VN, Sangon), 5mM dNTPs and 1mM EDTA (Thermo, AM9260G) was then added and the
431 reaction was incubated at 65°C for 10min and 72°C for 3min to facilitate RNA denaturation at
432 the same time. Next, RT was performed by adding 6μl RT mix (70U SuperScript IV (Thermo,
433 18090050), 1.7x SSIV buffer, 8.3mM DTT, 10U RRI, 1.7M Betaine (Sigma, B0300)) and
434 incubated at 50°C for 50min, then inactivated the reverse transcriptase at 80°C for 10min. The
435 resulted RNA/DNA hybrids mixed with 10-pg DNA carriers were tagmented by 0.05μl TTE Mix
436 V50 (Vazyme, TD501) at 55°C for 30 min, through adding 10μl reaction mix containing 2x TD
437 buffer (20mM Tris-HCl (ROCKLAND, MB-003), 10mM MgCl₂ (Thermo, AM9530G), 20% N,N-
438 Dimethylformamide (Sigma, D4551)), 16% PEG8000 (VWR Life Science, 97061), 0.5mM ATP
439 (NEB, P0756), 8U RRI. 6U Bst 3.0 DNA polymerase (NEB, M0374M) within 1 x Q5 high-fidelity
440 master mix were utilized to repair the gap left by V50 at 72°C for 15min, followed by 80°C for
441 5min to terminate the reaction. Finally, 3μl indexed primers mix (Vazyme, TD203) and 3μl Q5
442 mix were added to perform PCR amplification. PCR cycled as following: 98°C 30s for initial
443 denaturation, 21 cycles of 20s at 98°C, 20s at 60°C and 2min at 72°C, 72°C 5min for final
444 extension. The indexed products were merged and purified at 0.75x with VAHTS DNA Clean
445 Beads (Vazyme, N411).

446 Libraries were quantified with Qubit 2.0 and their fragment length distributions were
447 checked by Fragment Analyzer Automated CE System. Libraries were sequenced by Illumina
448 NextSeq 500 or NovaSeq S4.

449

450 **RNA-seq data analysis**

451 **Data quality.** Adaptors, poly(A/T) sequences were trimmed, bases with quality less than 20
452 and reads shorter than 20 bases were removed from the raw sequencing data by Cutadapt
453 (v1.15) [38]. Clean reads were mapped to indexed genome (human: Gencode.v31, mouse:
454 Gencode.vM23) by STAR (2.7.1a) [39]. Only unique alignment was utilized for downstream
455 analysis. The mitochondrial and ribosomal ratios were counted with samtools (v1.10) [40]. The
456 ratios of coding region, UTR, intron and intergenic region were counted with Picard tools
457 (v2.17.6). Exonic rate was defined as sum of coding region and UTR ratios. For cells, Cufflinks
458 (v2.2.1) [41] with exon annotations of protein coding genes were used to count gene number
459 (FPKM>1). For nuclei, genes (RPM>1) were counted by featureCounts (v1.5.1) [42] with
460 transcript annotations. Coverage across gene body was calculated by RSeQC (v.2.6.4) [43].
461 The coverage uniformity was integral area between coverage curve and x-axis normalized by
462 100.

463 **Gene ontology analysis.** We used genes that were detected in cell A while missed by cell B
464 as “study”, and combined the “study” genes with genes detected by cell B as “background”. The
465 gene ontology analysis was performed by GOATOOLS (v1.2.3) [44] and repeated between
466 every two cells from different methods. GO terms (excluding electronic annotations) with
467 adjusted p-value less than 0.01 were counted. All cells were firstly downsampled to 500K or
468 1M total reads.

469 **Clustering and marker genes.** For scRNA-seq and snRNA-seq, clustering followed the basic
470 tutorials of Scanpy (v1.8.1) [45]. The cell type annotations were through manually checking
471 expression of well-known marker genes. Marker genes that identified by SHERRY2 should
472 satisfy following conditions: 1) adjusted p-values calculated by Mann-Whitney-U test were less
473 than 1e-3; 2) foldchanges were greater than 1.5 or less than 0.67; 3) The average normalized

474 counts of up-regulated gene in the cell type, or down-regulated gene in the rest of cell types
475 was greater than 0.3. For NEBNext, DESeq2 (v1.22.2) [46] was utilized to identify the
476 differentially expressed genes (adjusted p-value < 1e-4, foldchange > 2).

477 **RNA Velocity.** Splicing and unslicing mRNA were quantified by Velocyto (v0.17.17) [10] with
478 unique alignment. The generated loom file was utilized by scVelo (v0.2.4) [47] to profile velocity
479 dynamics based on clustering results of Scanpy.

480 **BCR assembly.** BCR sequences of each cell was assembled by MIXCR (v3.0.13) [48] with
481 clean reads. The assembled BCR were realigned by IgBlast (v1.17.1) [49] to determine clone
482 types.

483 **Exon usage.** The frequency of exon usage in each cell was calculated by BRIE (v2.0.5) [50].
484 For each exon, cells satisfying following conditions were retained: 1) counts of gene which
485 included the exon were greater than 10; 2) exon regions sided by the specific exon should be
486 covered by greater than 50% with uniquely aligned reads; 3) at least one read should detect
487 junctions involved in this exon splicing events. Pairwise comparison of exon usage frequency
488 was made between cell types which contained greater than 10 cells using Mann-Whitney-U
489 test. The exons with p-value less than 0.05 was further checked in IGV viewer to check whether
490 transcript coverage was consistent with usage frequency. The passed ones were considered
491 as significantly biased among cell types.

492 **SmartSeq3 data reanalysis.** SmartSeq3 [35] sequencing data of 117 single HEK293T cells
493 was downloaded from ArrayExpress. The UMI and tag sequences at 5'-end were firstly
494 removed. Merged 5'-end reads and internal reads were then analyzed using pipeline described
495 in Data quality.

496 **Declarations**

497 **Ethics approval and consent to participate**

498 All mice were maintained under specific pathogen-free conditions and used in accordance of
499 governmental, Tsinghua University and Capital Medical University guidelines for animal
500 welfare.

501

502 **Consent for publication**

503 Not applicable

504

505 **Availability of data and materials**

506 All data generated or analysed during this study are included in this published article, its
507 supplementary information files and publicly available repositories. The sequence data reported
508 in this study have been deposited in the NCBI Sequence Read Archive (assessment no.
509 PRJNA879104).

510

511 **Competing interests**

512 The authors declare that they have no competing interests.

513

514 **Funding**

515 This work was supported by the National Key Research and Development Program of China
516 (2018YFA0108100 to Y.H.); the National Natural Science Foundation of China (22050002,

517 21927802 to Y.H.); the Beijing Municipal Science and Technology Commission
518 (Z201100005320016 to Y.H.); and the Shenzhen Bay Laboratory. Funding for open access
519 charge: Beijing Municipal Science and Technology Commission.

520

521 **Authors' contributions**

522 Y.H. and J.W. conceived the study; L.D., B.L., Y.L., S.Z., Y.P., and J.S. performed experiments;
523 L.D. performed data analyses; C.Z. and J.S. provided samples; L.D., B.L., J.W., H.Q., J.S., and
524 Y.H. wrote manuscript with input from all authors; Y.H., J.W., and H.Q. supervised all aspects
525 of this study.

526

527 **Acknowledgements**

528 We thank Chenyang Geng and Yan Chen from the Peking University High-throughput
529 Sequencing Center and Biomedical Pioneering Innovation Center for experimental assistance.
530 This work was supported by the National Key Research and Development Program of China
531 (2018YFA0108100 to Y.H.), National Natural Science Foundation of China (22050002,
532 21927802 to Y.H.), Beijing Municipal Science and Technology Commission
533 (Z201100005320016 to Y.H.), Beijing Advanced Innovation Center for Genomics, and
534 Shenzhen Bay Laboratory.

535 **References**

- 536 1. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying
537 mammalian transcriptomes by RNA-Seq. *Nature methods*. 2008;5(7):621-8.
- 538 2. Cui P, Lin Q, Ding F, Xin C, Gong W, Zhang L, et al. A comparison between ribo-minus
539 RNA-sequencing and polyA-selected RNA-sequencing. *Genomics*. 2010;96(5):259-65.
- 540 3. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2
541 for sensitive full-length transcriptome profiling in single cells. *Nature methods*.
542 2013;10(11):1096-8.
- 543 4. Bagnoli JW, Ziegenhain C, Janjic A, Wange LE, Vieth B, Parekh S, et al. Sensitive and
544 powerful single-cell RNA sequencing using mcSCRB-seq. *Nature communications*.
545 2018;9(1):1-8.
- 546 5. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel
547 digital transcriptional profiling of single cells. *Nature communications*. 2017;8(1):1-12.
- 548 6. Perez JD, tom Dieck S, Alvarez-Castelao B, Tushev G, Chan IC, Schuman EM. Subcellular
549 sequencing of single neurons reveals the dendritic transcriptome of GABAergic
550 interneurons. *Elife*. 2021;10:e63092.
- 551 7. Oguchi Y, Ozaki Y, Abdelmoez MN, Shintaku H. NanoSINC-seq dissects the isoform
552 diversity in subcellular compartments of single cells. *Science Advances*.
553 2021;7(15):eabe0317.
- 554 8. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IW, Ng LG, et al. Dimensionality
555 reduction for visualizing single-cell data using UMAP. *Nature biotechnology*.
556 2019;37(1):38-44.
- 557 9. Trapnell C, Cacchiarelli D. Monocle: Differential expression and time-series analysis for
558 single-cell RNA-Seq and qPCR experiments. *Bioconductor* Fmrp Usp Br. 2014;2.
- 559 10. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity

- 560 of single cells. *Nature*. 2018;560(7719):494-8.
- 561 11. Wang X, He Y, Zhang Q, Ren X, Zhang Z. Direct comparative analyses of 10X Genomics
562 Chromium and smart-seq2. *Genomics, Proteomics & Bioinformatics*. 2021.
- 563 12. Di L, Fu Y, Sun Y, Li J, Liu L, Yao J, et al. RNA sequencing by direct fragmentation of
564 RNA/DNA hybrids. *Proceedings of the National Academy of Sciences*. 2020;117(6):2886-
565 93.
- 566 13. Chen C, Li J, Di L, Jing Q, Du P, Song C, et al. MINERVA: a facile strategy for SARS-CoV-
567 2 whole-genome deep sequencing of clinical samples. *Molecular cell*. 2020;80(6):1123-34.
568 e4.
- 569 14. Gerard GF, Collins S, Smith MD. Excess dNTPs minimize RNA hydrolysis during reverse
570 transcription. *Biotechniques*. 2002;33(5):984-90.
- 571 15. Chandler DP, Wagnon CA, Bolton Jr H. Reverse transcriptase (RT) inhibition of PCR at
572 low concentrations of template and its implications for quantitative RT-PCR. *Applied and
573 environmental microbiology*. 1998;64(2):669-77.
- 574 16. Yamawaki TM, Lu DR, Ellwanger DC, Bhatt D, Manzanillo P, Arias V, et al. Systematic
575 comparison of high-throughput single-cell RNA-seq methods for immune cell profiling.
576 *BMC genomics*. 2021;22(1):1-18.
- 577 17. MacLennan IC. Germinal centers. *Annual review of immunology*. 1994;12(1):117-39.
- 578 18. Victora GD, Nussenzweig MC. Germinal centers. *Annual review of immunology*.
579 2012;30:429-57.
- 580 19. Allen CD, Ansel KM, Low C, Lesley R, Tamamura H, Fujii N, et al. Germinal center dark
581 and light zone organization is mediated by CXCR4 and CXCR5. *Nature immunology*.
582 2004;5(9):943-52.
- 583 20. Caron G, Le Gallou S, Lamy T, Tarte K, Fest T. CXCR4 expression functionally
584 discriminates centroblasts versus centrocytes within human germinal center B cells. *The
585 Journal of Immunology*. 2009;182(12):7595-602.
- 586 21. Victora GD, Schwickert TA, Fooksman DR, Kamphorst AO, Meyer-Hermann M, Dustin ML,
587 et al. Germinal center dynamics revealed by multiphoton microscopy with a
588 photoactivatable fluorescent reporter. *Cell*. 2010;143(4):592-605.
- 589 22. Victora GD. SnapShot: the germinal center reaction. *Cell*. 2014;159(3):700-. e1.
- 590 23. Mesin L, Ersching J, Victora GD. Germinal center B cell dynamics. *Immunity*.
591 2016;45(3):471-82.
- 592 24. De Silva NS, Klein U. Dynamics of B cells in germinal centres. *Nature reviews immunology*.
593 2015;15(3):137-48.
- 594 25. Kennedy DE, Okoreeh MK, Maienschein-Cline M, Ai J, Veselits M, McLean KC, et al. Novel
595 specialized cell state and spatial compartments within the germinal center. *Nature
596 immunology*. 2020;21(6):660-70.
- 597 26. Green JA, Suzuki K, Cho B, Willison LD, Palmer D, Allen CD, et al. The sphingosine 1-
598 phosphate receptor S1P 2 maintains the homeostasis of germinal center B cells and
599 promotes niche confinement. *Nature immunology*. 2011;12(7):672-80.
- 600 27. Huang C, Melnick A. Mechanisms of action of BCL6 during germinal center B cell
601 development. *Science China Life sciences*. 2015;58(12):1226-32.
- 602 28. Suan D, Kräutler NJ, Maag JL, Butt D, Bourne K, Hermes JR, et al. CCR6 defines memory
603 B cell precursors in mouse and human germinal centers, revealing light-zone location and

- 604 predominant low antigen affinity. *Immunity*. 2017;47(6):1142-53. e4.
- 605 29. Ise W, Fujii K, Shiroguchi K, Ito A, Kometani K, Takeda K, et al. T follicular helper cell-
606 germinal center B cell interaction strength regulates entry into plasma cell or recycling
607 germinal center cell fate. *Immunity*. 2018;48(4):702-15. e4.
- 608 30. Dominguez-Sola D, Victora GD, Ying CY, Phan RT, Saito M, Nussenzweig MC, et al. c-
609 MYC is required for germinal center selection and cyclic re-entry. *Nature immunology*.
610 2012;13(11):1083.
- 611 31. Laidlaw BJ, Duan L, Xu Y, Vazquez SE, Cyster JG. The transcription factor Hhex
612 cooperates with the corepressor Tle3 to promote memory B cell development. *Nature
613 immunology*. 2020;21(9):1082-93.
- 614 32. Jacob J, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-
615 nitrophenyl) acetyl. II. A common clonal origin for periarteriolar lymphoid sheath-associated
616 foci and germinal centers. *The Journal of experimental medicine*. 1992;176(3):679-87.
- 617 33. Habib N, Li Y, Heidenreich M, Swiech L, Avraham-David I, Trombetta JJ, et al. Div-Seq:
618 Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*.
619 2016;353(6302):925-8.
- 620 34. Tan MC, Widagdo J, Chau YQ, Zhu T, Wong JJ-L, Cheung A, et al. The activity-induced
621 long non-coding RNA Meg3 modulates AMPA receptor surface expression in primary
622 cortical neurons. *Frontiers in cellular neuroscience*. 2017;11:124.
- 623 35. Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks G-J, Larsson AJ, et
624 al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature
625 Biotechnology*. 2020;38(6):708-14.
- 626 36. Chen H, Yao J, Fu Y, Pang Y, Wang J, Huang Y. Tagmentation on microbeads: restore
627 long-range DNA sequence information using Next Generation Sequencing with library
628 prepared by surface-immobilized transposomes. *ACS applied materials & interfaces*.
629 2018;10(14):11539-45.
- 630 37. Zhang F, Christiansen L, Thomas J, Pokholok D, Jackson R, Morrell N, et al. Haplotype
631 phasing of whole human genomes using bead-based barcode partitioning in a single tube.
632 *Nature biotechnology*. 2017;35(9):852-7.
- 633 38. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
634 *EMBnet journal*. 2011;17(1):10-2.
- 635 39. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
636 universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
- 637 40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
638 alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
- 639 41. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and
640 transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature
641 protocols*. 2012;7(3):562-78.
- 642 42. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
643 assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-30.
- 644 43. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*.
645 2012;28(16):2184-5.
- 646 44. Klopfenstein D, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, et al.
647 GOATOOLS: A Python library for Gene Ontology analyses. *Scientific reports*. 2018;8(1):1-

- 648 17.
- 649 45. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data
650 analysis. *Genome biology*. 2018;19(1):1-5.
- 651 46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
652 seq data with DESeq2. *Genome biology*. 2014;15(12):1-21.
- 653 47. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell
654 states through dynamical modeling. *Nature biotechnology*. 2020;38(12):1408-14.
- 655 48. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al.
656 MiXCR: software for comprehensive adaptive immunity profiling. *Nature methods*.
657 2015;12(5):380-1.
- 658 49. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain
659 sequence analysis tool. *Nucleic acids research*. 2013;41(W1):W34-W40.
- 660 50. Huang Y, Sanguinetti G. BRIE: transcriptome-wide splicing quantification in single cells.
661 *Genome biology*. 2017;18(1):1-11.
- 662

663 **Figures**

664 **Figure 1.The workflow and general performance of SHERRY2 on single cell RNA-seq.** (A)
665 The workflow of SHERRY2 for scRNA-seq. Poly(A) tailed RNA is firstly released from single
666 cells and reverse transcribed. The resulting RNA/cDNA hetero-duplex is then fragmented by
667 Tn5 transposome, followed by gap-repair and indexed PCR. Optionally, chromatin can be
668 digested during lysis. The entire protocol is performed in one tube and takes 3 hours. (B) Gene
669 number (FPKM>1) with SmartSeq2, SHERRY2 and SHERRY when subsampling reads to 0.1,
670 0.2, 0.4, 0.6, 0.8 and 1 million reads. (C) Pairwise correlation of gene expression within
671 replicates for the three scRNA-seq protocols. The correlation R-value was calculated by a linear
672 fitting model with normalized counts of overlapped genes. (D) Gene body coverage detected by
673 the three scRNA-seq protocols. The gray region represents the standard deviation of the
674 normalized depth among replicates. (E) Components of reads that were mapped to different
675 regions of the genome using the three scRNA-seq protocols. The error bars show the standard
676 deviation. (F) Gene expression correlation between single HEK293T cells and 200-ng RNA
677 extracted from HEK293T cells. Single-cell data were acquired by the three scRNA-seq
678 protocols. Bulk RNA results were acquired by the standard NEBNext protocol. The correlation
679 R-value was calculated by a linear fitting model with normalized gene counts. The samples in
680 (B-F) are single HEK293T cells. The p-values in (B, C, F) were calculated by the Mann-Whitney-
681 U test.

682

683 **Figure 2.scRNA-seq of low RNA-content samples with SHERRY2.** (A) Proportions of
684 genome regions covered by reads from SHERRY2 without DNase treatment, SHERRY2 with
685 AG DNase I addition, SHERRY2 with AG DNase I and DNA carrier addition, and SmartSeq2.
686 (B) Gene number (FPKM>1) detected by SHERRY2 with AG DNase I addition, SHERRY2 with
687 AG DNase I and DNA carrier addition, and SmartSeq2 when subsampling to 20, 50, 100, 200,
688 400, and 600 thousand reads. Only samples with intergenic rate lower than 25% were counted.
689 Samples in (A, B) were single lymphocyte cells from murine eyeball blood. (C) Library quality
690 of SHERRY2 tested with different DNases, including gene number (FPKM>1) at 0.25-million
691 reads, coverage uniformity across gene body and percentage of reads that were mapped to
692 intergenic regions. The labels below the figure indicate the amounts and names of the DNases,
693 as well as the EDTA concentration that was added during DNase inactivation. SmartSeq2 was
694 also performed as a reference. (D) Components of reads covering different genome regions
695 detected by SHERRY2 without DNase treatment, SHERRY2 with optimized AG DNase I, and
696 SmartSeq2. (E) Gene body coverage detected by SHERRY2 (with AG DNase I) and
697 SmartSeq2. The gray region shows the standard deviation of the normalized depth among
698 replicates. (F) Gene number (FPKM>1) detected by SHERRY2 (with AG DNase I and DNA
699 carrier) and SmartSeq2 when subsampling to 20, 50, 100, 200, 400, and 600 thousand reads.
700 (G) Gene ontology analysis of genes that only detected by SHERRY2 (left) or SmartSeq2
701 (right). The top 20 most commonly occurred GO terms were shown. Samples in (C-G) were
702 single B cells isolated from murine GC light zones. The p-values in (B, F) were calculated by
703 the Mann-Whitney-U test. The error bars in (A, D) show the standard deviation.

704

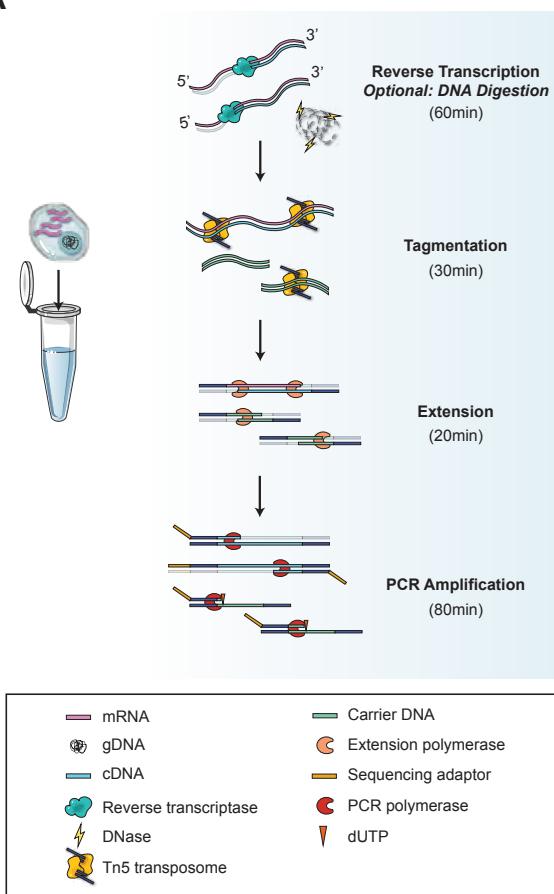
705 **Figure 3.Mouse germinal center profiled by scRNA-seq through SHERRY2.** (A) Clustering
706 of single B cells from murine GC light zones visualized by UMAP plot. The library was prepared
707 by SHERRY2 (with AG DNase I and DNA carrier). Different colors indicate distinct cell types.
708 (B) Cell cycle and marker gene expression of different cell types marked on a UMAP plot. The
709 gradient colors correspond to the normalized counts of a specific gene ranging from 0 (white)
710 to 1 (blue). (C) Distribution of *Myc* gene expression in different cell types. Different colors
711 indicate different intervals of normalized *Myc* counts. The percentages of cells within the
712 clusters falling into corresponding intervals were counted. (D) Dynamic process of the GC light
713 zone indicated by vector fields of RNA velocity on a UMAP plot. The expanded region shows
714 the velocity vector of each cell. The colors correspond to the same cell types as annotated in
715 (A). (E) Isoforms of the *Hnrnpab* gene. The top two lines show isoforms from two example cells

716 that rarely and preferentially used the highlighted exon in *Hnrnpab* transcripts. The bottom two
717 lines show the isoform structures of *Hnrnpab* transcripts that include or exclude the exon. (F)
718 Inclusion ratio distribution of the highlighted exon in (E) in different cell types. Only cell types
719 represented by more than 10 cells after filtering are shown.

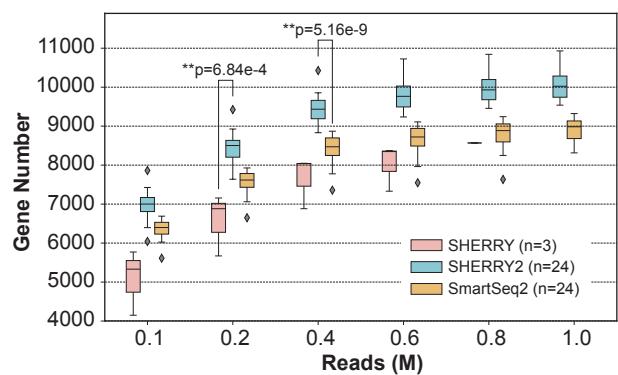
720

721 **Figure 4. Sensitivity and accuracy of SHERRY2.** (A) Gene number (RPM>1) of single
722 HEK293T nuclei detected by SHERRY2 and SmartSeq2 when subsampling reads to 0.1, 0.2,
723 0.4, 0.6, 0.8 and 1 million reads. (B) Gene expression correlation between single HEK293T
724 nuclei and 200-ng RNA extracted from HEK293T nuclei. Single-nucleus data were acquired by
725 SHERRY2 and SmartSeq2. Bulk RNA results were acquired by the standard NEBNext protocol.
726 The correlation R-value was calculated by a linear fitting model with normalized gene counts.
727 (C) Clustering of HEK293T cellular and nuclear RNA-seq data from SHERRY2, SmartSeq2 and
728 NEBNext using principal component analysis. The analysis utilized differentially expressed
729 genes (adjusted p-value < 1e-4 and fold change > 2) between cells and nuclei detected by
730 NEBNext. (D) Gene number (RPM>1) of single neuron nuclei detected by SHERRY2 and
731 SmartSeq2 when subsampling reads to 0.1, 0.2, 0.4, 0.6, 0.8 and 1 million reads. The nuclei
732 were isolated from mouse hippocampi that were freshly prepared or previously frozen at -80°C.
733 (E) Clustering of single hippocampal neuron nuclei visualized by UMAP plot. The snRNA-seq
734 library was prepared by SHERRY2. The analysis utilized genes expressed (counts > 0) in more
735 than 4 nuclei. (F) Marker gene expression of different cell types on UMAP plot from (E). The
736 gradient colors correspond to the normalized counts of a specific gene ranging from 0 to 1. The
737 p-values in (A, B, D) were calculated by the Mann-Whitney-U test.

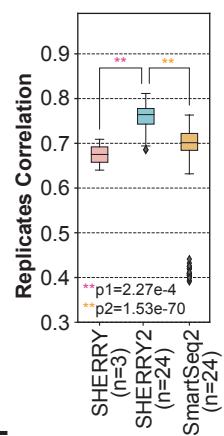
A



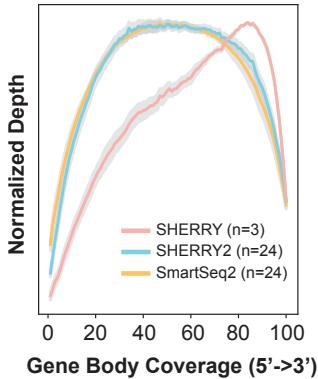
B



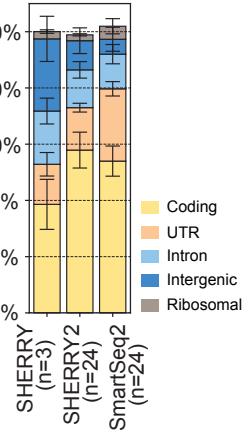
C



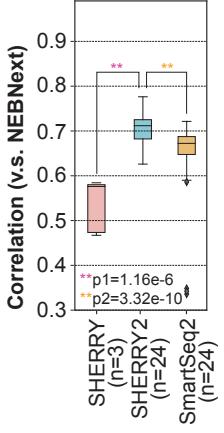
D



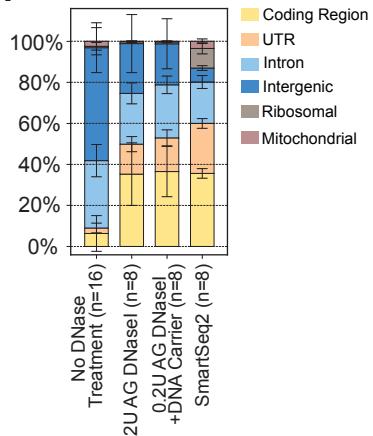
E



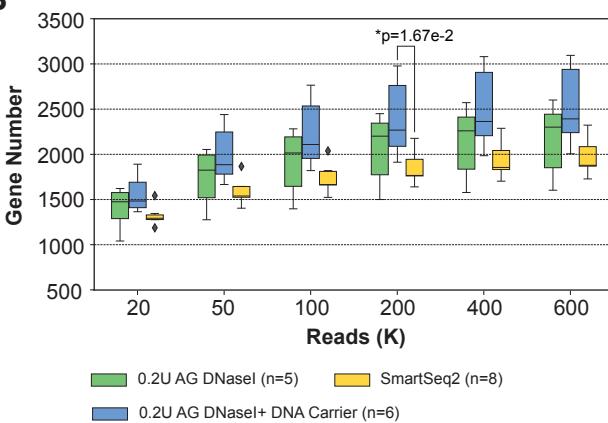
F



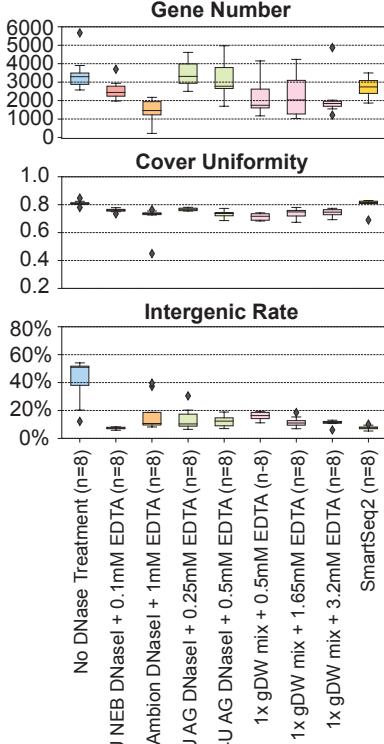
A



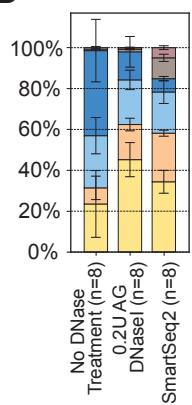
B



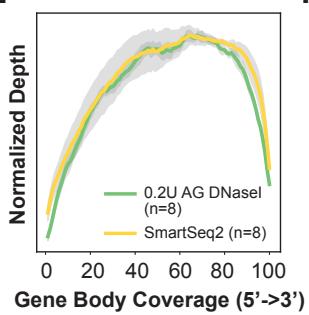
C



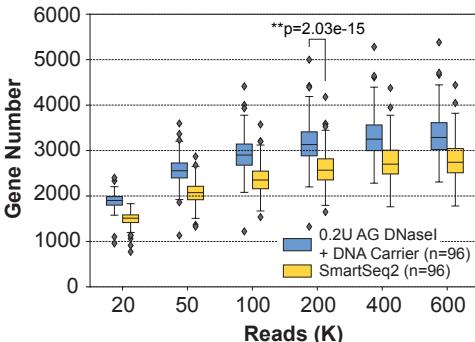
D



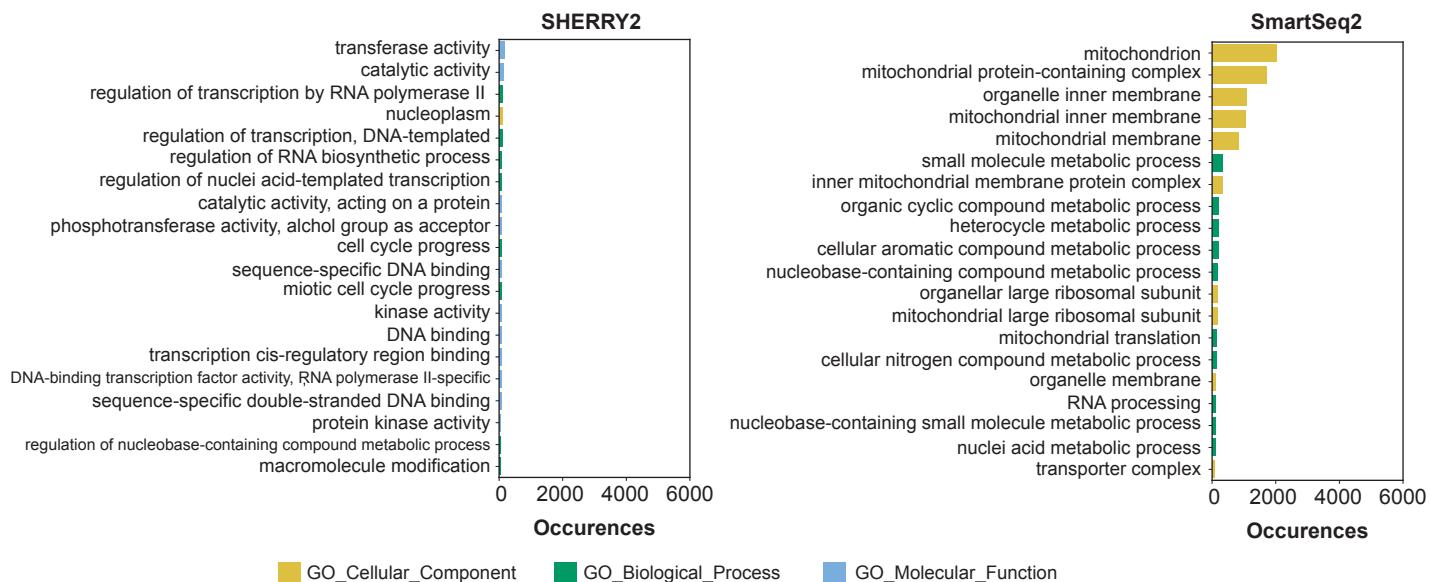
E



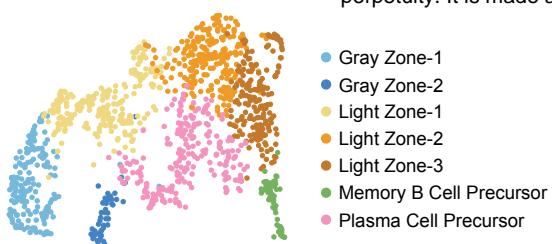
F



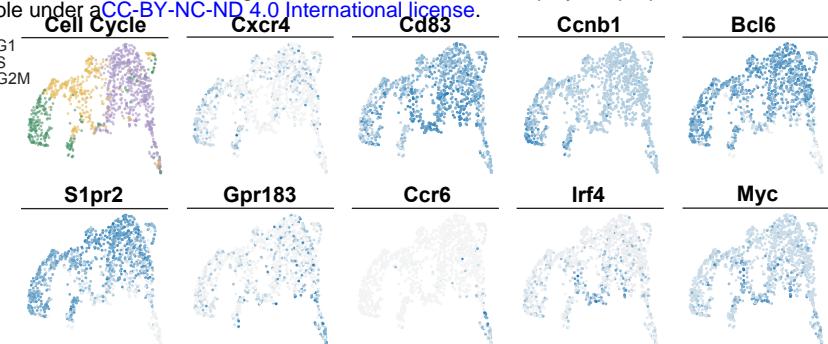
G



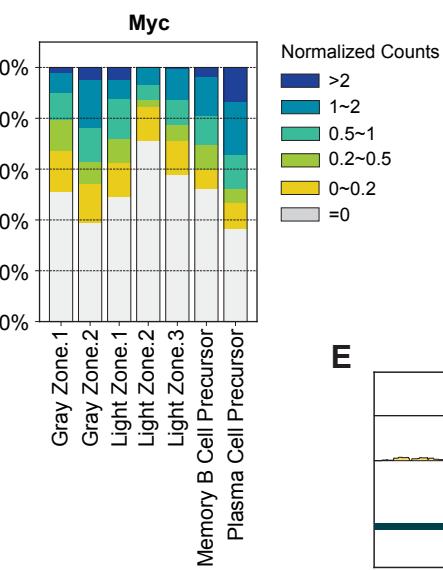
A



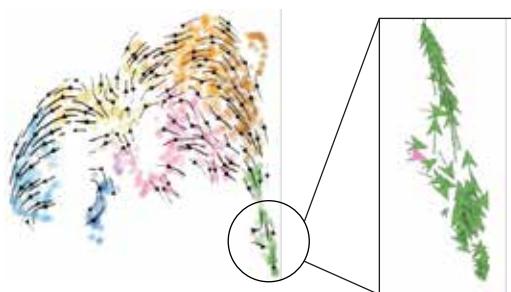
B



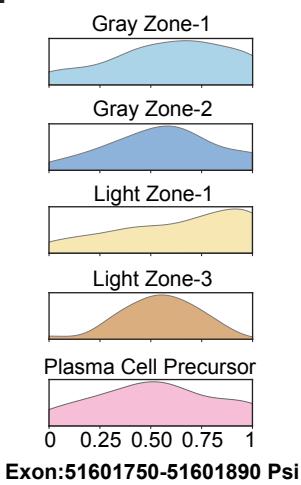
C



D



F



E

