

1 **CASB: A concanavalin A-based sample barcoding strategy for**  
2 **single-cell sequencing**

3

4 Liang Fang<sup>1,2#</sup>, Guipeng Li<sup>1,2#</sup>, Qionghua Zhu<sup>1</sup>, Huanhuan Cui<sup>1,2</sup>, Yunfei Li<sup>1</sup>, Zhiyuan Sun<sup>1</sup>, Weizheng  
5 Liang<sup>1</sup>, Wencheng Wei<sup>1</sup>, Yuhui Hu<sup>1,2</sup>, Wei Chen<sup>1,2\*</sup>

6

7 1. Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong  
8 518005, China

9 2. Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology,  
10 Shenzhen, Guangdong 518005, China

11

12 #These authors contributed equally: Liang Fang, Guipeng Li

13 \*Corresponding author. Tel: +86 755 88018449; E-mail: chenw@sustech.edu.cn

14 **Abstract**

15 Sample multiplexing facilitates single cell sequencing by reducing costs, revealing subtle difference  
16 between similar samples, and identifying artifacts such as cell doublets. However, universal and cost-  
17 effective strategies are rather limited. Here, we reported a Concanavalin A-based Sample Barcoding  
18 strategy (CASB), which could be followed by both single-cell mRNA and ATAC (assay for transposase  
19 accessible chromatin) sequencing techniques. The method involves minimal sample processing,  
20 thereby preserving intact transcriptomic or epigenomic patterns. We demonstrated its high labeling  
21 efficiency, high accuracy in assigning cells/nuclei to samples regardless of cell type and genetic  
22 background, as well as high sensitivity in detecting doublets by two applications: 1) CASB followed by  
23 scRNA-seq to track the transcriptomic dynamics of a cancer cell line perturbed by multiple drugs, which  
24 revealed compound-specific heterogeneous response; 2) CASB together with both snATAC-seq and  
25 scRNA-seq to illustrate the IFN- $\gamma$ -mediated dynamic changes on epigenome and transcriptome profile,  
26 which identified the transcription factor underlying heterogeneous IFN- $\gamma$  response.

## 27 **Introduction**

28 Single-cell mRNA sequencing (scRNA-seq) and single-nucleus assay for transposase-accessible  
29 chromatin using sequencing (snATAC-seq) have emerged as powerful technologies for interrogating  
30 the heterogeneous transcriptional profiles and chromatin landscapes of multicellular subjects  
31 (Buenrostro, Wu et al., 2015, Cusanovich, Daza et al., 2015, Hashimshony, Wagner et al., 2012,  
32 Ramskold, Luo et al., 2012). Early scRNA/snATAC-seq workflows were limited to analyzing tens to  
33 hundreds of individual cells at a time. With the latest development of single-cell sequencing technologies  
34 based on microwells (Gierahn, Wadsworth et al., 2017), combinatorial indexing (Rosenberg, Roco et al.,  
35 2018) and droplet-microfluidics (Klein, Mazutis et al., 2015, Macosko, Basu et al., 2015), the parallel  
36 analysis of thousands of single cells or nuclei has become routine. The increase in throughput does not  
37 only lower the reagent costs per cell, but also enable the analysis of whole organs or entire organisms  
38 in one experimental run.

39

40 Recently, with the ever-increasing throughput, these technologies have also been used to reveal the  
41 temporal response of heterogeneous cell population under diverse perturbations, which require tens of  
42 samples to be processed in parallel (Hurley, Ding et al., 2020, Weinreb, Rodriguez-Fraticelli et al., 2020).  
43 Based on existing methods, sample-specific barcodes (for example, Illumina library indices) are often  
44 incorporated at the very end of standard library preparation workflow. Such workflow requires parallel  
45 processing of multiple individual samples until the final step, therefore not only is labor-intensive and  
46 limits the number of samples, but also increase the reagent costs if a small number of cells would be  
47 sufficient to characterize the heterogeneity of each individual sample. To overcome this, alternative  
48 multiplexing approaches should label cells from each sample with distinct barcodes before pooling for  
49 single-cell sequencing experiment. The sample-specific barcodes could then be linked to cell barcodes  
50 during single-cell sequencing library preparation. Several methods have been developed in this  
51 endeavor, which introduce sample barcodes using either genetic or non-genetic mechanisms.  
52 Genetically, researchers have used various strategies to express an exogenous gene with sample-  
53 specific barcodes at its 3' UTR, which can be captured similarly as endogenous genes (Hurley et al.,  
54 2020, Weinreb et al., 2020); non-genetically, people have used oligonucleotide containing a sample  
55 barcode followed by a poly-A sequences, which can be immobilized on the cell or nuclear membrane  
56 through anchoring molecules (e.g. antibody and lipid) (McGinnis, Patterson et al., 2019, Stoeckius,  
57 Hafemeister et al., 2017) or chemical cross-linking reaction (Gehring, Hwee Park et al., 2020), or

58 defused into permeabilized nuclei (Srivatsan, McFaline-Figueroa et al., 2020), and then captured during  
59 reverse transcription. Although being already quite powerful, each of these methods has still its own  
60 liabilities, including issues with scalability, universality or the potential to introduce artefactual  
61 perturbations. Moreover, all of these methods have only been combined with scRNA/snRNA-seq, and  
62 have not yet been applied and are likely incompatible with snATAC-seq.

63 Here, we developed a Concanavalin A-based Sample Barcoding strategy (CASB) that overcomes many  
64 of these limitations. Taking advantage of the glycoprotein-binding ability of concanavalin A (ConA),  
65 CASB was used to label cell or nucleus with biotinylated single-strand DNA (ssDNA) through a  
66 streptavidin bridge. CASB could be easily adapted into scRNA/snATAC-seq workflows, and showed  
67 high accuracy in assigning cells or nuclei regardless of genetic background as well as in resolving cell  
68 doublets. The application of CASB in samples with time-series experiments, followed by scRNA- and/or  
69 snATAC-seq, allows revealing diverse transcriptome/epigenome dynamics.



## 70 **Results**

### 71 **CASB enables cell and nucleus labeling with ssDNA**

72 The CASB consists of three components: biotinylated ConA, streptavidin and biotinylated ssDNA as  
73 barcoding molecules. Both ConA and streptavidin form homo-tetramer autonomously, allowing the  
74 assembly of ConA-streptavidin-ssDNA complex (Fig 1A). Relying on the glycoprotein-binding ability of  
75 ConA, such assembled complex can be immobilized on the cell or nuclear membrane (Fig 1A). To  
76 measure how many ssDNA molecules can be immobilized on the cell membrane, a biotinylated ssDNA  
77 with 5' and 3' PCR handles flanking an eight-nucleotide (N8) random sequence was used to label the  
78 cells (Fig 1A). After incubation with different quantity of preassembled ConA-streptavidin-ssDNA  
79 complex in PBS at 4 °C (Methods), the number of ssDNA molecules immobilized on mouse embryonic  
80 stem cells (mESC) was quantified using qPCR. As shown in Figure 1B, the amount of ssDNA  
81 immobilized on cells increased with the increased usage of ConA-streptavidin-ssDNA complex, and  
82 could reach as many as 50,000 molecules per cell. To test whether ssDNA may fall off from labeled  
83 cells and cause cross-contamination during sample pooling, a mouse embryonic fibroblast (MEF) cell  
84 population expressing mCherry fluorescent proteins was labeled with the ssDNA and then mixed with  
85 another MEF cell population expressing GFP fluorescent proteins, which was only coated with “empty”  
86 ConA (Methods). After 30 min incubation in PBS at 4 °C, mCherry and GFP positive cells were  
87 separated using FACS and subjected to qPCR measurement. As shown in Figure 1C, the ssDNA  
88 immobilized on mCherry<sup>+</sup> cells was not detectable from GFP<sup>+</sup> cells, demonstrating the stability of CASB  
89 labeling. In addition to labeling the whole cell, we also measured the labeling efficiency of CASB for cell  
90 nucleus, in which nuclei were labeled with preassembled ConA-streptavidin-ssDNA complex in nuclear  
91 extraction buffer at 4 °C (Methods). As shown in Figure EV1A, the amount of ssDNA immobilized on  
92 nuclei increased with the increased usage of ConA-streptavidin-ssDNA complex, and reached at least  
93 120,000 molecules per nucleus. Taken together, these results demonstrated that CASB is able to stably  
94 label both cell and nucleus with ssDNA.

95

### 96 **CASB enables scRNA-seq sample multiplexing**

97 In scRNA-seq, cell specific barcodes are attached to the cDNA during reverse transcription (RT) by  
98 using primers consisting of a cell barcode sequence, a unique molecular identifier (UMI) sequence and  
99 a poly-T sequence that anchors to the poly-A tail of mRNA molecule. To make our CASB compatible

100 with the standard scRNA-seq workflow, we designed a biotinylated barcoding ssDNA with a 5' PCR  
101 handle followed by a N8 barcode and a 30 nt poly-A tail, which can be captured by a RT primer consisting  
102 of a PCR handle followed by a 30 nt poly-T tail (Fig 2A). After CASB labeling, MEF cells were directly  
103 lysed and subjected to RT reaction (Methods). The barcoding ssDNA immobilized on cell membrane  
104 was quantified together with the endogenous housekeeping gene ActB using qPCR. As shown in Figure  
105 EV1B, both barcoding ssDNA and ActB gene can be efficiently capture by RT primer. Therefore, CASB  
106 could be easily adapted into scRNA-seq workflow with high efficiency.

107

108 To demonstrate the strength of CASB in scRNA-seq, a breast cancer cell line MDA-MB-231 was  
109 perturbed with 5 different compounds, collected at 3 different time points after treatment and pooled with  
110 3 other breast cancer cell lines as well as MEF cells after separate sample labelling using CASB (Fig  
111 2B). Unlabeled MDA-MB-231 cells were also added into the sample pool to measure the potential  
112 influence of CASB on transcriptome profile. Sample pool was then subjected to scRNA-seq using the  
113 10x Genomics Chromium system with minor modifications: 1) in order to examine the efficiency of CASB  
114 to detect doublets, we intentionally overloaded the system (~20,000 instead of ~10,000 cells  
115 recommended by the manufacturer) to create more cell doublets; 2) CASB barcode and transcriptome  
116 library were separated by size selection before next-generation sequencing library construction,  
117 enabling pooled sequencing at user-defined proportions (Methods).

118

119 As a result, a total of 12068 cells were captured with sufficient reads for transcriptome analysis. For  
120 each cell, the reads derived from each of the 20 different sample barcodes were counted and used to  
121 demultiplex the samples using HTODemux method (Stoeckius, Zheng et al., 2018) (Methods). A total  
122 of 483 cells were assigned as 'Unlabeled', as expected due to the inclusion of unlabeled MDA-MB-231  
123 cells (Fig EV1C). Among the remaining ones, 3962 cells were assigned as cell doublets encapsulated  
124 in the same droplet, as they contained two or three major barcodes (Fig EV1C). Indeed, the doublets  
125 consisting of both mouse and human cells, which could be unambiguously detected based on their  
126 mapping results, could also be efficiently identified based on the mixture of CASB barcodes. As shown  
127 in Figure 2C, out of 110 mouse-human doublets, 107 (97.3%) were defined as doublets based on our  
128 CASB data. When compared with singlets, more UMI derived from both CASB barcode and mRNA  
129 transcripts were detected in doublets (Fig EV1D), further validating the correct assignment of cell  
130 doublets. Within 7623 singlets, the number of detected UMI from CASB per cell ranged from 245 to

131 2134 (5-95 percentile), and significantly correlated with UMI detected for endogenous transcripts in the  
132 same cell (Fig EV1E), suggesting a similar cell-specific capture efficiency between CASB barcode and  
133 endogenous transcripts, and that CASB did not impair mRNA capture. Based on the qPCR quantification,  
134 the same amount of CASB mixture could label cells with ~20,000 ssDNA (Fig 1B), mean UMI (1051)  
135 detected in the scRNA-seq indicates a ~5% capture efficiency at current sequencing depth (25 million  
136 total sequencing reads). To determine the variation of labeling efficiency among different cells, given  
137 the cell-specific capture efficiency, we first normalized the number of UMI numbers from CASB by that  
138 of UMI derived from the endogenous transcripts in the same cell. As shown in Figure EV1F, the CASB  
139 barcoding manifested a good uniformity of labeling efficiency among all singlets and within individual  
140 cell samples. Taken together, our CASB strategy could achieve high sensitivity in detecting cell identity  
141 and doublets in scRNA-seq experiments.

142

143 For the 7623 cells with unambiguously assigned sample origin, we then clustered them based on their  
144 scRNA-seq profiles. As shown in Figure 2D, different human and mouse cells formed 5 distinct cell  
145 clusters, respectively. Each cluster was composed of cells from individual cell line labeled with distinct  
146 CASB indices (Fig 2D). We also compared the untreated MDA-MB-231 cells with to those without CASB  
147 labelling. As shown in Figure EV1G and H, single cell profiles were intermingled together and their  
148 cumulative transcriptome were highly correlated, demonstrating a negligible influence of CASB labeling  
149 on transcriptome profile. Within the MDA-MB-231 cell population, all 16 sample barcodes can be  
150 detected (Fig EV1I). Cells associated with 24 h-treatment of Niraparib, Rucaparib and OSI-027 could  
151 be well distinguished from untreated cells, whereas those with LCL161 and Fludarabine could not (Fig  
152 EV1J). As expected, Niraparib- and Rucaparib-treated cells were intermingled due to their common  
153 molecular target PARP.

154

155 MDA-MB-231 is of triple-negative breast cancer origin, which lacks efficient targeted therapy. As  
156 intratumoral heterogeneity has been associated with therapy resistance, we investigated whether drug  
157 treatments could lead to heterogeneous response in the MDA-MB-231 cells. We focused on compound  
158 OSI-027, as it induced the largest transcriptomic changes (Fig EV1J). As shown in Figure 2E and EV2A,  
159 in which cells treated with OSI-027 were plotted with untreated cells, there indeed existed three cell  
160 populations with distinct transcriptomic responses. Whereas one showed clearly time-dependent  
161 transcriptomic changes (Fig 2E, cluster 0 circled in pink), the other two had limited alteration in gene

162 expression even after 24 h (Fig 2E, cluster 1&2 circled in green and blue, respectively), suggesting that  
163 the latter were less sensitive to the OSI-027. Neighbor proportion analysis also confirmed that, untreated  
164 cells were well separated from treated cells in cluster 0, while it is not the case for cluster 1 and 2 (Fig  
165 EV2B). As cellular heterogeneity existed already in the untreated MDA-MB-231 cells (Fig EV2C), we  
166 sought to further check whether the insensitive cell populations were also resistant to other effective  
167 compounds. Indeed, as shown in Figure 2F and EV2D, the same two cell clusters also appeared less  
168 sensitive to Niraparib and Rucaparib, suggesting the intrinsic multidrug insensitivity. To explore the  
169 underlying factors, we perform function enrichment analysis on genes that were commonly up- or  
170 downregulated in cluster 1 and 2 compared to cluster 0 using IPA software (Methods). Interestingly,  
171 these genes were highly enriched in the cellular compromise and movement pathways (Table EV1 &  
172 Fig EV2E). Importantly, many of them (e.g. VIM) were upregulated in cluster 1 and 2, and known to  
173 promote cellular movement (Fig EV2F). In tumor cells, increased cell motility mediated by epithelial-  
174 mesenchymal transition (EMT) is also highly associated with drug resistance (Singh & Settleman, 2010,  
175 Zhang & Weinberg, 2018). Our results suggested that the intrinsic multidrug insensitivity of MDA-MB-  
176 231 cells may result from the activated EMT. More interestingly, when overlap the potential insensitivity-  
177 causing genes in cluster 1 and 2 with OSI-27-regulated genes, we observe that many genes, including  
178 VIM, SQSTM1, NPM1 and RACK1, were also upregulated by OSI-027 in sensitive cells (Fig 2G). Given  
179 these genes are involved in promoting EMT and potentially also drug resistance (Fig EV2F), this  
180 observation indicated the potential of OSI-027 treatment in inducing acquired therapy resistance.

181

## 182 **CASB enables snATAC-seq sample multiplexing**

183 So far, no sample multiplexing method has been developed for droplet-based snATAC-seq. In droplet-  
184 based snATAC-seq, cell nuclei are firstly incubated with transposase in bulk, where genomic DNA is  
185 fragmented and tagged with adapter sequences. Afterwards, single cells are encapsulated, and cell  
186 barcodes are added to DNA fragments during PCR in individual droplets using primers targeting the  
187 adapter sequences. To adapt CASB into snATAC-seq workflow, we designed a 222 nt barcoding ssDNA  
188 with S5-ME and S7-ME adapter sequences flanking a sequence containing sample barcodes (Fig 3A).  
189 S5-ME and S7-ME adapter sequences were used as primer anchoring sites during snATAC-seq library  
190 amplification (Methods). The labeling efficiency using such ssDNA was measured similarly as before, in  
191 which nuclei were directly labeled with preassembled ConA-streptavidin-ssDNA complex in nuclear  
192 extraction buffer at 4 °C (Methods). As shown in Figure EV3A, the amount of ssDNA immobilized on

193 nuclei increased with the increased usage of ConA-streptavidin-ssDNA complex, and could reached at  
194 least 80,000 molecules per nucleus.

195

196 To demonstrate the application of CASB in snATAC-seq, we sought to monitor the temporal chromatin  
197 changes induced by interferon-gamma (IFN- $\gamma$ ) in HAP1 cells. IFN- $\gamma$  is an important cytokine in the host  
198 defense against infection by viral and microbial pathogens (Shtrichman & Samuel, 2001). It mediates  
199 innate immunity through regulating effector gene expression, which is accompanied by substantial  
200 changes at epigenetic level (Ivashkiv, 2018). However, how heterogeneously and dynamically cells  
201 respond to IFN- $\gamma$  stimulation at the chromatin level has remained elusive. Taking advantage of CASB,  
202 we analyzed the changes in chromatin accessibility of HAP1 cells at 7 different time points after IFN- $\gamma$   
203 stimulation using snATAC-seq. MDA-MB-231 cells were added into the pool here as an outlier control.  
204 After sequencing, a total of 2890 cells were obtained with sufficient reads, 305 of which were identified  
205 as cell doublets that have at least two major CASB barcodes (Fig 3B, marked in black), and 23 cells  
206 were unlabeled (Fig EV3B). MDA-MB-231 cells with its specific CASB barcode presented as an isolated  
207 cluster (Fig EV3C). As revealed by UMAP projection, HAP1 cells showed a continuous shift in chromatin  
208 profile from 0 to 12 h (Fig 3C). To uncover the key transcription factors (TFs) that mediate IFN- $\gamma$ -induced  
209 chromatin remodeling, we analyzed their binding motifs on the ATAC-peaks across all cells and  
210 observed that peaks containing motifs of IRF, STAT and NF- $\kappa$ B TF showed large variation in their  
211 intensity, indicating their functions in modulating IFN- $\gamma$  response (Fig 3D) (Methods). Indeed, IRF and  
212 STAT peaks showed continuous activation (Fig 3E&F). This is expected as IFN- $\gamma$  is able to activate  
213 JAK/STAT signaling through binding to its receptor, which in term activates the expression of IFN- $\gamma$ -  
214 responsive genes, including transcription factors IRFs (Leonard & O'Shea, 1998). Interestingly, the large  
215 variation of NF- $\kappa$ B peak intensity did not result from IFN- $\gamma$  treatment, but was instead largely due to the  
216 heterogeneity of HAP1 cells (Fig 3F). It is known that NF- $\kappa$ B can be activated by IFN- $\gamma$  and able to  
217 facilitate the transcription activation of IFN- $\gamma$  targets, including CXCLs (Pfeffer, 2011, Qin, Roberts et al.,  
218 2007). This result suggested that heterogeneous NF- $\kappa$ B activity may give rise to heterogeneous IFN- $\gamma$   
219 response.

220

221 To evaluate whether heterogeneous NF- $\kappa$ B activity causes heterogeneous IFN- $\gamma$  response, IFN- $\gamma$   
222 treated samples from the same time points were also analyzed using CASB followed by scRNA-seq. A  
223 total of 3407 cells were captured, 294 of which were identified as cell doublets and 9 cells were

224 unlabeled (Fig EV4A&B). As shown in Figure 4A, HAP1 cells showed a continuous shift in the  
225 transcriptome profile from 0 to 12 h on UMAP projection. To globally evaluate the correlation between  
226 chromatin accessibility and gene expression, we analyzed the dynamic expression patterns of predicted  
227 IRF and STAT target genes. In consistent with the activity of IRF and STAT observed in snATAC-seq  
228 (Fig 3C), the expression of their target genes also exhibited continuous upregulation (Fig 4B).

229

230 As revealed by unsupervised clustering with Louvain method, cells at later time points (4, 6, 8 and 12 h)  
231 were clustered into two populations, one of which exhibited more divergent transcriptome profile from  
232 earlier time points (Fig 4C, cluster 2, circled in red). To see whether this is associated with  
233 heterogeneous NF- $\kappa$ B activity identified in snATAC-seq, the expression of predicted NF- $\kappa$ B target genes  
234 was compared between the two cell populations. Consistent with the heterogeneous NF- $\kappa$ B activity, its  
235 target genes also exhibit heterogeneous expression, and were expressed at a higher level in cluster 2  
236 at later time points (Fig 4D). The high induction of CXCL10 and 11, well-known targets of IFN- $\gamma$ , only in  
237 cluster 2 cells further corroborate that the heterogeneous NF- $\kappa$ B activity indeed results in differential  
238 responses to IFN- $\gamma$  in HAP1 cells (Fig 4E&S4C).

## 239 **Discussion**

240 CASB is a flexible sample barcoding approach, ready to be prepared in an average molecular biology  
241 laboratory. CASB could be used to label cells and nuclei of different cell types and from different species.  
242 Moreover, the binding of CASB molecules to the subject is fast and stable, and takes place even at low  
243 temperature, which is critical to preserve sample integrity. Importantly, the design of CASB barcoding  
244 ssDNA is extremely flexible, which can be easily adapted to different single cell sequencing workflows.

245  
246 CASB allows scalable sample multiplexing by solely increasing the variety of barcoding ssDNAs. In this  
247 study, we tested CASB's scalability by performing a 20-plex perturbation assay followed by scRNA-seq,  
248 which revealed new information about drug response of triple-negative breast cancer cells. Specifically,  
249 it demonstrated the different response dynamics of different compounds, and different response of  
250 different cell subpopulations to the same drug. The scalability of CASB could be potentially enhanced  
251 by combinatorial indexing. When integrated with automated cell handling system, CASB following by  
252 scRNA-seq could serve as a powerful platform for single-cell sequencing-based drug screens.

253  
254 Cell doublets, i.e. two or more cells encapsulated in a same droplet, posed a challenge for single cell  
255 sequencing data analysis. Without sample barcoding, cell doublets from the same species could only  
256 be estimated mathematically with certain ambiguity (DePasquale, Schnell et al., 2019). To reduce the  
257 doublet rate, an often-sought strategy is to limit the number of cells loaded in the microwell- or droplet-  
258 based systems. As demonstrated in this study, CASB could reveal cell doublets in high accuracy and  
259 as such its application would allow to increase the throughput of single-cell sequencing systems by  
260 loading more cells. Indeed, similar strategy has been proposed by using antibody-based barcoding  
261 approach (Stoeckius et al., 2018). However, the efficiency of doublet identification correlated to the  
262 diversity of sample barcodes. By increasing the sample barcodes to hundreds or even thousands, which  
263 could be easily achieved using CASB, we would envisage a much higher doublet detection efficiency,  
264 which allows the further optimization of cell loading rate.

265  
266 Due to the universal presence of glycoprotein on plasma membrane, CASB is applicable to any sample  
267 with an accessible plasma membrane. Worth noting, after 1 h transposition reaction at 37 °C, the CASB  
268 barcodes remained abundant and showed minimal cross contamination. We believe that CASB can  
269 become compatible with many other single-cell sequencing technologies such as CITE-seq (Stoeckius

270 et al., 2017) and SNARE-seq (Chen, Lake et al., 2019), and for samples preserved in different ways  
271 such as flash-frozen and formalin-fixed ones. Moreover, by using biotin and fluorophore dual-labeled  
272 barcoding ssDNA, one could further enrich cells that are successfully barcoded.

273

274 In summary, CASB allows to incorporate additional layers of information into single-cell sequencing  
275 experiments. With the ever-increasing throughput of single-cell sequencing technologies, CASB does  
276 not only reduce reagent costs, improve data analysis, minimize the batch effect, but also can become a  
277 versatile tool in this field by incorporating more diverse types of information, including time-points,  
278 treatment conditions and potentially also spatial coordinates. With further improvement, such as using  
279 ConA-Streptavidin fusion protein or fluorophore-labeled ssDNA, it will facilitate more novel applications  
280 of single cell sequencing technology.



## 281 **Materials and Methods**

### 282 **Experimental materials and methods**

#### 283 **Cell culture and pre-processing**

284 The MDA-MB-231, MDA-MB-453, T-47D and MCF7 cells were obtained from the ATCC, while HAP1  
285 from Horizon discovery. The MEF cells and mESC were kindly gifted by the Qi Zhou's and Wei Li's Lab  
286 at the Institute of Zoology, Chinese Academy of Sciences. MDA-MB-231, MDA-MB-453, T-47D, MCF7,  
287 HAP1 and MEF cells were cultured in DMEM (11995040, Gibco) with 10% FBS (10270106, Gibco) and  
288 1% P/S (15070063, Gibco) with 5% CO<sub>2</sub> at 37°C, while the mESC were cultured in Neuralbasal (21103-  
289 049, Gibco)-DMEM/F12 (11330-032, Gibco) based medium with N2 (17502048, Gibco), B27(17504-  
290 044, Gibco), PD0325901 (s1036, Selleck), Chir99021 (s1263, Selleck) and mLIF (ESG1107, Millipore)  
291 with 5% CO<sub>2</sub> at 37°C. For stimulation with IFN- $\gamma$ , HAP1 cells were treated with 100 ng/mL IFN- $\gamma$  (#300-  
292 02, Peprotech) for 2, 4, 6, 8 and 12 hours. For scRNA-seq related experiments, cells were trypsinized  
293 and washed once with PBS, while for snATAC-seq related experiments, after washing with PBS, cells  
294 were cryopreserved in 200  $\mu$ L cryo-medium (10% DMSO, 40% FBS, 50% culture medium) and kept in  
295 -80 °C.

#### 296 **Compounds and treatment**

297 The compounds used in this study include LCL161 targeting XIAP, Fludarabine inhibiting DNA synthesis,  
298 OSI-027 blocking mTOR, Rucaparib and Niraparib targeting PARP1, which were chosen based on their  
299 selective inhibitory effect on MDA-MB-231 cells (Garnett, Edelman et al., 2012, Iorio, Knijnenburg et al.,  
300 2016). Compounds LCL161, Fludarabine, OSI-027, Niraparib and Rucaparib (HY-15518, HY-B0069,  
301 HY-10423, HY-10619 and HY-10617) were obtained from MedChemExpress and dissolved in DMSO.  
302 For scRNA-seq experiment, 0.1  $\mu$ M of LCL161, 0.15  $\mu$ M of Fludarabine, 2.5  $\mu$ M of OSI-027, 15  $\mu$ M of  
303 Rucaparib and 12.5  $\mu$ M of Niraparib were used to treat the cells for 4, 8 or 24 h.

#### 304 **Design and synthesis of CASB barcoding ssDNA**

305 For measuring the number of ssDNA molecules immobilized on cell or nuclear membrane, a 5'-  
306 biotinylated ssDNA with 5' and 3' PCR handles flanking a N8 random sequence was designed: 5'-  
307 TCGTCGGCAGCGTCAGATGTGTATA-NNNNNNNN-TATACACATCTCCGAGCCCACGAGAC-3'. For  
308 scRNA-seq related experiments, 5'-biotinylated ssDNAs with a 5' PCR handle followed by a N8 barcode  
309 and a 30 nt poly-A tail were designed: 5'-GCTGCGCTCGATGCAAATA-NNNNNNNN-  
310 BAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-3'. For snATAC-seq related experiments, a 5'-

311 biotinylated forward primer (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-  
312 CTTGTGGAAAGGACGAAACACCG-3') and a reverse primer (5'-  
313 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-GTGTCTCAAGATCTAGTTACGCCAAGC-3')  
314 were used to amplify 222 bp fragments from CROPseq-Guide-Puro plasmids (#86708, Addgene) which  
315 have been inserted with different gRNA sequences. To generate ssDNAs, purified PCR products were  
316 denatured at 95 °C for 2 mins and immediately put on ice. Information of CASB barcode sequences and  
317 their corresponding samples can be found in Table EV2.

### 318 **Assembly of CASB barcode**

319 Biotinylated ConA (C2272, Sigma-Aldrich) and streptavidin (CS10471, Coolaber) were dissolved in 50%  
320 glycerol at concentration of 1.6  $\mu$ M and store in -20 °C, while different biotinylated ssDNAs were diluted  
321 at concentration of 100 nM in nuclease-free water and stored in -20 °C. To assemble CASB barcode,  
322 streptavidin was firstly mixed with biotinylated ssDNA at molar ratio of 4:1 and incubated for 10 min at  
323 room temperature. Afterwards, biotinylated ConA was added to the streptavidin-ssDNA mix at molar  
324 ratio of 1:1 and incubated for 10 min at room temperature.

### 325 **Cell labeling with CASB**

326 Half a million cells were collected and suspended in 0.5 mL PBS. Indicated amount of CASB barcode  
327 was added to the cells and incubated for 10 mins on ice after thorough mixing. For labeling mCherry+  
328 MEF cells, 2.5  $\mu$ L assembled CASB was used. For RT-qPCR and scRNA-seq, 5  $\mu$ L assembled CASB  
329 barcode was used. Afterwards, cells were washed once with PBS and then subjected to direct qPCR,  
330 direct RT reaction or mixed for scRNA-seq. For qPCR quantification, three independent biological  
331 replications were performed for each experiment.

### 332 **Nuclei labeling with CASB**

333 Cells were thawed by adding 800  $\mu$ L warm culture medium and collected by centrifugation. Afterwards,  
334 cells were resuspended in 0.5 mL nuclei extraction buffer (NUC101-1KT, Sigma-Aldrich), incubated for  
335 5 min on ice, and collected by centrifugation (500 g). Extracted nuclei were incubated again with 0.5 mL  
336 nuclei extraction buffer, in which indicated amount of CASB barcode was added, for another 5 min on  
337 ice. For snATAC-seq, 2.5  $\mu$ L assembled CASB barcode was used. Nuclei were then washed once with  
338 nuclei wash buffer (10 mM Tris 7.4, 10 mM NaCl, 3mM MgCl<sub>2</sub>, 1% BSA, 0.1% Tween-20) and then  
339 subjected to direct qPCR or mixed for snATAC-seq. For qPCR quantification, three independent  
340 biological replications were performed for each experiment.

### 341 **Quantification of ssDNA immobilized on cell or nuclear membrane**

342 For all quantification experiments using qPCR, standard curves were always first drawn using serially  
343 diluted pure ssDNA for calculating the precise number of ssDNA in each reaction. For each reaction of  
344 qPCR, 2000 labeled cells in 5  $\mu$ L PBS were directly mix with 5  $\mu$ L of primer mix (1  $\mu$ M) and 10  $\mu$ L of  
345 qPCR master mix (11201ES03, Yeasen). Primers used for quantifying barcoding ssDNA are 5'-  
346 TCGTCGGCAGCGTCAG-3' and 5'-GTCTCGTGGGCTCGGAG-3'.

347 For measuring the ssDNA with the polyA tail, 20000 cells were directly lysed in 6  $\mu$ L of 0.17% TrionX-  
348 100 (T8787, Sigma-Aldrich) for 3 min at 72  $^{\circ}$ C, and then revers transcribed with 1<sup>st</sup> strand cDNA  
349 synthesis kit (11119ES60, Yeasen) using RT primer 5'-CACGCACTGACTGACAGAC-  
350 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTV-3' (final concentration 2.5  $\mu$ M). The qPCR was performed  
351 with ssDNA-specific forward primer 5'-GCTGCGCTCGATGCAAATA-3' and Actb-specific forward  
352 primer 5'-GTGACAGCATTGCTTCTGTGTAAAT-3' combining with common reverse primer 5'-  
353 CACGCACTGACTGACAGACT-3'.

#### 354 **scRNA-seq and snATAC-seq**

355 The scRNA-seq experiments were performed according to the standard protocol of single cell 3' reagent  
356 kits v2 (PN-120237, 10X Genomics) with following modifications. During cDNA amplification, additional  
357 primer (5'-GCTGCGCTCGATGCAAATA-3', 0.1  $\mu$ M) was added to amplify CASB barcode. To capture  
358 amplified CASB barcode, during "post cDNA amplification reaction cleanup", the amplified full-length  
359 cDNA library was purified with 2X SPRIselect Reagent (B23318, Beckman Coulter) and eluted in 40  $\mu$ L  
360 of nuclease-free water, 10  $\mu$ L of which was subject to PCR with primer pair 5'-  
361 AATGATACGGCGACCACCGAGATCT-ACACTCTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-  
362 CAAGCAGAAGACGGCATAACGAGAT-CTGATC-TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-  
363 GCTGCGCTCGATGCAAATA-3' using PrimeSTAR Max PCR master mix (R045A, Takara) to add  
364 sequencing adapter sequences to CASB barcode.

365 The snATAC-seq experiments were performed according to the standard protocol of single cell ATAC  
366 reagent kits (PN-1000111, 10X Genomics) with no modification.

#### 367 **Next generation sequencing**

368 All sequencing experiments were performed with Illumina NovaSeq 6000 System. The service for  
369 scRNA-seq was provided by Haplox genomics center, while for snATAC-seq by Genergy Bio. For  
370 scRNA-seq, paired-end 150 bp with i7 8 bp sequencing strategy was used, while for snATAC-seq,  
371 paired-end 150 bp with i7 8 bp and i5 16 bp sequencing strategy was applied. All sequencing data were  
372 submitted to GEO under the accession number GSE153116.

373

## 374 **Computational methods**

### 375 **CASB barcode analysis**

376 For scRNA-seq, raw barcode library FASTQ files were converted to barcode UMI count matrix using  
377 custom script leveraging the pysam (Li, Handsaker et al., 2009) package ([https://github.com/pysam-](https://github.com/pysam-developers/pysam)  
378 [developers/pysam](https://github.com/pysam-developers/pysam)). This procedure was similar with a previous method (McGinnis et al., 2019). Briefly,  
379 raw FASTQ files were first parsed to use only the reads where the first 16 bases of R1 perfectly match  
380 any of the cell barcodes predefined by Cell Ranger. Then, reads where the 20-28 bases of R2 align with  
381 at most 1 mismatch to any predefined sample barcodes were used. Reads were grouped by cell  
382 barcodes and duplicated UMIs were identified as reads where 17-26 bases of R1 exactly matched.

383 In snATAC-seq, sample barcodes were in R2 reads and cell barcodes were in R1 reads. Reads with  
384 sample and cell barcodes were first extracted from raw FASTQ files of snATAC library using custom  
385 script to get the cell-by-sample count matrix.

386 HTODemux method (Stoeckius et al., 2018) implemented in Seurat package was used to define  
387 'doublets', 'singlets' and 'negatives'.

### 388 **scRNA-seq gene expression analysis**

389 FASTQ files were processed using Cell Ranger (10X Genomics, v3.1.0). The reads were aligned to the  
390 concatenated hg38-mm10 or hg38 reference using STAR (Dobin, Davis et al., 2013). Cell-associated  
391 barcodes were defined by Cell Ranger. Gene expression UMI count matrix (h5 file) was obtained using  
392 Cell Ranger with default parameters.

393 After the pre-processing, RNA UMI count matrices were prepared for scRNA-seq analysis using the  
394 'Seurat' R package (Butler, Hoffman et al., 2018). Cells with no more than 4000 reads or 2000 expressed  
395 genes were removed. Outlier cells with elevated mitochondrial gene expression were visually defined  
396 and discarded. Ribosomal genes and mitochondrial genes were then filtered out.

397 'sctransform' R package (Hafemeister & Satija, 2019) was used to normalize the RNA UMI count data  
398 and find highly variable genes. These variable genes were then used during principal component  
399 analysis (PCA). Elbow plot was used to select the top principal components. Then these principal  
400 components were used for dimensionality reduction with UMAP and unsupervised clustering with  
401 Louvain method. Differential gene expression analysis was performed using the 'FindMarkers' function  
402 in Seurat with 'MAST' method (Finak, McDavid et al., 2015). To quantify the magnitude of perturbation  
403 induced by drug on gene expression, we compared the proportion of each cell's  $k$  ( $k=9$ ) nearest

404 neighbors in principal component space with the 'knn.covertree' R package. The proportion was  
405 normalized by the cell numbers of different groups.

#### 406 **snATAC-seq data analysis**

407 After filtering out the reads with sample barcodes, Cellranger-atac (version 1.2.0) was used to process  
408 the raw FASTQ files. The reads were aligned to hg38 reference using BWA-MEM (Li & Durbin, 2009).  
409 The filtered peak-by-cell matrix (h5 file) obtained after cellranger-atac processing was used in the  
410 subsequent analysis. The matrix was first binarized. Cells of low quality (no more than 2000 peaks or  
411 more than 500000 peaks, percent of reads in peaks  $\leq$  30%, percent of peaks in ENCODE black list  $>$ 5%)  
412 were filtered out. Only cells defined by HTODemux as 'singlet' were used for subsequent analysis. ATAC  
413 peaks with low coverage (less than 50 cells) or ultra-high coverage (more than 2000 cells) were also  
414 removed. The binarized count matrix was normalized by term frequency inverse document frequency  
415 (TF-IDF). Latent semantic indexing analysis was performed as applying singular value decomposition  
416 (SVD) on the normalized count matrix. Only the 2nd-50th dimensions after the SVD were passed to  
417 UMAP for 2D visualization. Motif analysis was performed using chromVAR (Schep, Wu et al., 2017).  
418 The predicted target genes of TF were defined by the nearest genes within 100 kb of the activated ATAC  
419 peaks with the TF motif at 12h. The activated ATAC peaks were called by using FindMarkers function  
420 with parameter test.use='LR'. The average expression levels of these predicted target genes were  
421 calculated using 'AddModuleScore' function in Seurat package.

#### 422 **Gene function enrichment and network analysis using IPA software**

423 The commonly up- or downregulated genes in cluster 1 and 2 compared to cluster 0 of untreated MDA-  
424 MB-231 cells were subjected to the Ingenuity Pathway Analysis (IPA) (QIAGEN) (Kramer, Green et al.,  
425 2014) to gain insights into the gene functions. The "Diseases & Functions" module under the "Expression  
426 Analysis" were used for this purpose. In "Diseases & Functions" module, the analysis was restricted to  
427 "Molecular and Cellular Functions".

428 **Reference**

- 429 Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ  
430 (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523: 486-90  
431 Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data  
432 across different conditions, technologies, and species. *Nature Biotechnology* 36: 411-420  
433 Chen S, Lake BB, Zhang K (2019) High-throughput sequencing of the transcriptome and chromatin  
434 accessibility in the same cell. *Nat Biotechnol* 37: 1452-1457  
435 Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C,  
436 Shendure J (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular  
437 indexing. *Science* 348: 910-4  
438 DePasquale EAK, Schnell DJ, Van Camp PJ, Valiente-Alandi I, Blaxall BC, Grimes HL, Singh H,  
439 Salomonis N (2019) DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data.  
440 *Cell Rep* 29: 1718-1727 e8  
441 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR  
442 (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21  
443 Finak G, McDavid A, Yajima M, Deng JY, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ,  
444 Prlic M, Linsley PS, Gottardo R (2015) MAST: a flexible statistical framework for assessing  
445 transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome*  
446 *Biol* 16  
447 Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR,  
448 Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson  
449 JA, Barthorpe S et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer  
450 cells. *Nature* 483: 570-5  
451 Gehring J, Hwee Park J, Chen S, Thomson M, Pachter L (2020) Highly multiplexed single-cell RNA-seq  
452 by DNA oligonucleotide tagging of cellular proteins. *Nat Biotechnol* 38: 35-38  
453 Gierahn TM, Wadsworth MH, 2nd, Hughes TK, Bryson BD, Butler A, Satija R, Fortune S, Love JC,  
454 Shalek AK (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat*  
455 *Methods* 14: 395-398  
456 Hafemeister C, Satija R (2019) Normalization and variance stabilization of single-cell RNA-seq data  
457 using regularized negative binomial regression. *Genome Biol* 20  
458 Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear  
459 amplification. *Cell Rep* 2: 666-73  
460 Hurley K, Ding J, Villacorta-Martin C, Herriges MJ, Jacob A, Vedaie M, Alysandratos KD, Sun YL, Lin  
461 C, Werder RB, Huang J, Wilson AA, Mithal A, Mostoslavsky G, Oglesby I, Caballero IS, Guttentag SH,  
462 Ahangari F, Kaminski N, Rodriguez-Fraticelli A et al. (2020) Reconstructed Single-Cell Fate Trajectories  
463 Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung  
464 Progenitors. *Cell Stem Cell* 26: 593-608 e8  
465 Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Goncalves E, Barthorpe  
466 S, Lightfoot H, Cokelaer T, Greninger P, van Dyk E, Chang H, de Silva H, Heyn H, Deng X, Egan RK,  
467 Liu Q, Mironenko T et al. (2016) A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* 166:  
468 740-754  
469 Ivashkiv LB (2018) IFN $\gamma$ : signalling, epigenetics and roles in immunity, metabolism, disease and  
470 cancer immunotherapy. *Nat Rev Immunol* 18: 545-558  
471 Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW  
472 (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161: 1187-  
473 1201  
474 Kramer A, Green J, Pollard J, Jr., Tugendreich S (2014) Causal analysis approaches in Ingenuity  
475 Pathway Analysis. *Bioinformatics* 30: 523-30  
476 Leonard WJ, O'Shea JJ (1998) Jaks and STATs: biological implications. *Annu Rev Immunol* 16: 293-  
477 322  
478 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.  
479 *Bioinformatics* 25: 1754-1760  
480 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD  
481 (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079  
482 Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N,  
483 Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly  
484 Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161: 1202-  
485 1214



486 McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, Hu JL, Murrow LM,  
487 Weissman JS, Werb Z, Chow ED, Gartner ZJ (2019) MULTI-seq: sample multiplexing for single-cell  
488 RNA sequencing using lipid-tagged indices. *Nat Methods* 16: 619-626  
489 Pfeffer LM (2011) The role of nuclear factor kappaB in the interferon response. *J Interferon Cytokine*  
490 *Res* 31: 553-9  
491 Qin H, Roberts KL, Niyongere SA, Cong Y, Elson CO, Benveniste EN (2007) Molecular mechanism of  
492 lipopolysaccharide-induced SOCS-3 gene expression in macrophages and microglia. *J Immunol* 179:  
493 5966-76  
494 Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF,  
495 Laurent LC, Schroth GP, Sandberg R (2012) Full-length mRNA-Seq from single-cell levels of RNA and  
496 individual circulating tumor cells. *Nat Biotechnol* 30: 777-82  
497 Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, Graybuck LT, Peeler DJ, Mukherjee  
498 S, Chen W, Pun SH, Sellers DL, Tasic B, Seelig G (2018) Single-cell profiling of the developing mouse  
499 brain and spinal cord with split-pool barcoding. *Science* 360: 176-182  
500 Schep AN, Wu BJ, Buenrostro JD, Greenleaf WJ (2017) chromVAR : inferring transcription-factor-  
501 associated accessibility from single-cell epigenomic data. *Nature Methods* 14: 975-+  
502 Shtrichman R, Samuel CE (2001) The role of gamma interferon in antimicrobial immunity. *Curr Opin*  
503 *Microbiol* 4: 251-9  
504 Singh A, Settleman J (2010) EMT, cancer stem cells and drug resistance: an emerging axis of evil in  
505 the war on cancer. *Oncogene* 29: 4741-51  
506 Srivatsan SR, McFaline-Figueroa JL, Ramani V, Saunders L, Cao J, Packer J, Pliner HA, Jackson DL,  
507 Daza RM, Christiansen L, Zhang F, Steemers F, Shendure J, Trapnell C (2020) Massively multiplex  
508 chemical transcriptomics at single-cell resolution. *Science* 367: 45-51  
509 Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija  
510 R, Smibert P (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*  
511 14: 865-868  
512 Stoeckius M, Zheng SW, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, Smibert P, Satija R (2018)  
513 Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell  
514 genomics. *Genome Biol* 19  
515 Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM (2020) Lineage tracing on transcriptional  
516 landscapes links state to fate during differentiation. *Science* 367  
517 Zhang Y, Weinberg RA (2018) Epithelial-to-mesenchymal transition in cancer: complexity and  
518 opportunities. *Front Med* 12: 361-373  
519

520 **Acknowledgments**

521 **Funding**

522 This work was supported by the Shenzhen Science and Technology Program (Grant No.  
523 KQTD20180411143432337, JCYJ20190809154407564 and JCYJ20180504165804015) and the  
524 National Natural Science Foundation of China (Grant No. 31970601, 31701237 and 31900431).  
525 Computational resource was supported by the Center for Computational Science and Engineering of  
526 Southern University of Science and Technology.

527 **Author contributions**

528 W.C. and L.F. developed the concept of the project. L.F., Q.Z., Y.L. and H.C. designed and performed  
529 experiments. G.L. performed bioinformatic analysis. Z.S., W.L. and W.W. assisted in performing  
530 experiments. W.C., L.F., G.L. and Y.H. reviewed and discussed results. W.C., L.F. and G.L. wrote the  
531 manuscript.

532 **Competing interests**

533 The authors declare no conflict of interest.

534 **Data and materials availability**

535 All next-generation sequencing data were submitted to GEO under the accession number GSE153116.



536 **Figures**

537 **Figure 1. Cell labeling with CASB. A**, An illustration of CASB. Biotinylated ssDNA was immobilized on  
538 glycoprotein on cell/nuclear membrane through streptavidin and biotinylated ConA. The ssDNA contains  
539 5' and 3' PCR handles that flank an 8 nt random sequence. **B**, mESC were labeled with different quantity  
540 of CASB, and the number of ssDNA molecules immobilized on mESC was quantified used qPCR. The  
541 amount of ssDNA immobilized on cells increased with the increased usage of ConA-streptavidin-ssDNA  
542 complex, and reach as many as 50,000 molecules per cell. **C**, CASB-labeled mCherry+ MEF cells were  
543 incubated with unlabeled GFP+ MEF cells. The number of ssDNA molecules immobilized on mCherry+  
544 and GFP+ cells was quantified used qPCR after FACS separation. The ssDNA immobilized on  
545 mCherry+ cells was not detectable from GFP+ cells. "n" means number of qPCR reactions. Error bars  
546 represent SD.

547  
548 **Figure 2. CASB enables scRNA-seq sample multiplexing. A**, An illustration of CASB used in scRNA-  
549 seq. A biotinylated barcoding ssDNA with a 5' PCR handle followed by an 8 nt barcode and a 30 nt poly-  
550 A tail was used to mimic the endogenous transcripts. **B**, The design of the experiment. MDA-MB-231  
551 cells were perturbed with 5 different compounds, collected at 3 different time points, CASB labeled, and  
552 then pooled with 3 other breast cancer cell lines and MEF cells. **C**, Scatter plot depicting the number of  
553 UMIs associated with transcripts from human or mouse genome. Cell doublets revealed by CASB were  
554 marked in black. Out of 110 mouse-human doublets, 107 were detected as doublets by CASB barcodes.  
555 Three interspecies cell doublets that were not detected by CASB were circled in red. Beside interspecies  
556 cell doublets, cell doublets from one species were also detected by CASB. **D**, Transcriptome-based  
557 UMAP of cells captured in scRNA-seq. Cells were colored according to the CASB barcodes, and  
558 doublets were excluded. Different human and mouse cells formed 5 distinct cell clusters, respectively.  
559 **E&F**, Transcriptome-based UMAP of untreated and **(E)** OSI-027-, **(F)** Niraparib- and Rucaparib-treated  
560 MDA-MB-231 cells. Three cell populations with distinct transcriptomic responses were observed in each  
561 UMAP: sensitive cell subpopulation was circled in red, while insensitive ones in green and blue,  
562 respectively. **G**, Transcriptome-based UMAP of untreated and OSI-027-treated MDA-MB-231 cells.  
563 Sensitive cell subpopulation was circled in red, while insensitive ones in green and blue, respectively.  
564 Expression level of VIM, SQSTM1, NPM1 and RACK1 are indicated by color code, which were  
565 expressed in untreated insensitive cell populations and induced by OSI-027 in sensitive cells.

566

567 **Figure 3. CASB enables snATAC-seq sample multiplexing.** **A**, An illustration of CASB used in  
568 snATAC-seq. A biotinylated barcoding ssDNA with S5-ME and S7-ME adapter sequences flanking a  
569 sequence containing sample barcodes was used to mimic the transposed genomic DNA. **B**, t-SNE  
570 projection based on the CASB barcode reads captured in snATAC-seq. Cells were colored according  
571 to the CASB barcodes, and doublets were marked in black. **C**, ATAC-based UMAP of all HAP1 cells  
572 captured in snATAC-seq. Cells were colored according to the CASB barcodes, and doublets were  
573 excluded. HAP1 cells showed a continuous shift in chromatin profile from 0 to 12 h. **D**, Dot plot revealing  
574 the TFs with the most variable activity across all cells including IRF, STAT and NF- $\kappa$ B. **E**, ATAC-based  
575 UMAP of all HAP1 cells, in which the TF activity was presented by bias-corrected deviation z-score  
576 across all cells in color code. **F**, Violin plots demonstrating the deviation z-score of different TFs across  
577 different cells at different time points. IRF and STAT associated peaks showed continuous activation  
578 upon IFN- $\gamma$  stimulation, while the variation of NF- $\kappa$ B peak intensity was largely due to the heterogeneity  
579 within HAP1 cells.

580

581 **Figure 4. Transcriptomic heterogeneity within HAP1 cells.** **A**, Transcriptome-based UMAP of HAP1  
582 cells captured in scRNA-seq, in which cells were colored according to the CASB barcodes. HAP1 cells  
583 showed globally a continuous shift from 0 to 12 h. **B**, Violin plots demonstrating the continuous  
584 transcriptional activation of predicted IRF and STAT target genes across different cells. **C**,  
585 Transcriptome-based UMAP of HAP1, in which cells were unsupervised clustered and colored  
586 according to the transcriptomic feature revealed by Louvain algorithm. Cells were clustered into two  
587 populations at 4-12 h, one of which exhibited more divergent transcriptome profile from earlier time  
588 points. **D**, Violin plots comparing the expression of predicted NF- $\kappa$ B target genes between cluster 0 and  
589 2 at 4-12 h. Predicted target genes were heterogeneously expressed and more actively induced in  
590 cluster 2. **E**, Transcriptome-based UMAP of HAP1 cells, in which the expression of CXCL10 were  
591 presented with color code and showed activation only in cluster 2 (circled in red) at later time points.

592 **Supplementary Materials**

593 **Supplementary Figures**

594 **Figure EV1. CASB facilitates scRNA-seq sample multiplexing. A**, The number of ssDNA molecules  
595 immobilized on mESC nuclei was quantified using qPCR. The amount of ssDNA immobilized on nuclei  
596 increased with the increased usage of ConA-streptavidin-ssDNA complex, and reached at least 120,000  
597 molecules per nucleus. **B**, The poly-A ssDNA molecules immobilized on MEF were detected using RT-  
598 qPCR. Both poly-A ssDNA and ActB transcripts can be efficiently captured by RT primer. As expected,  
599 barcoding ssDNA can be detected by qPCR even without RT reaction. Error bars represent SD. **C**,  
600 Heatmap showing the detected levels of each CASB barcode in individual cells in scRNA-seq. A total  
601 of 12068 cells with sufficient reads were captured; 3962 cells that contained at least two major barcodes  
602 were assigned as cell doublets; 483 cells were assigned as 'unlabeled', as expected due to the inclusion  
603 of unlabeled MDA-MB-231 cells. **D**, Boxplot demonstrating the number of UMI derived from both CASB  
604 barcode and mRNA transcripts in cell doublets and singlets. Comparing with singlets, more UMI derived  
605 from both CASB barcode and mRNA transcripts were detected in doublets. **E**, Scatterplot illustrating the  
606 significant positive correlation between the number of detected UMI from CASB and endogenous  
607 transcripts among individual cells. "R" means Pearson's correlation coefficient. **F**, Distribution of  
608 normalized CASB UMI counts of singlets and individual cell samples. The CASB barcoding manifested  
609 a good uniformity of labeling efficiency (5-95 percentile: 2.1% - 21.8%) (upper panel); comparing with  
610 human cell samples, MCF7 cells had slightly lower labeling efficiency (lower panel). **G**, Transcriptome-  
611 based UMAP comparing labeled and unlabeled untreated MDA-MB-231 cells, in which two cell  
612 populations were intermingled. **H**, Scatterplot demonstrating the well correlated gene expression profiles  
613 between labeled and unlabeled untreated MDA-MB-231 cells. "R" means Pearson's correlation  
614 coefficient. **I**, t-SNE projection based on the CASB barcode reads captured in scRNA-seq. Cells were  
615 colored according to the CASB barcodes, and doublets were marked in black. All 20 sample barcodes  
616 can be detected. **J**, Transcriptome-based UMAP of all MDA-MB-231 cells captured in scRNA-seq.  
617 Untreated and 24-h treated cells were highlighted. Cells associated with 24 h-treatment of Niraparib,  
618 Rucaparib and OSI-027 could be well distinguished from untreated cells, whereas those with LCL161  
619 and Fludarabine could not.

620

621 **Figure EV2. Intrinsic heterogeneity of MDA-MB-231 cells. A,** Transcriptome-based UMAP of  
622 untreated and OSI-027-treated MDA-MB-231 cells. Cells were unsupervised clustered into three distinct  
623 groups with Louvain method. **B,** Neighbor proportion analysis of untreated and OSI-027-treated MDA-  
624 MB-231 cells. In cluster 0, untreated cells were distant from treated cells, while, in cluster 1 and 2,  
625 untreated cells were 50% neighbored with treated cells. **C,** Transcriptome-based UMAP of untreated  
626 MDA-MB-231 cells. Cells were unsupervised clustered into three distinct groups with Louvain method.  
627 **D,** Transcriptome-based UMAP of untreated and Niraparib- and Rucaparib-treated MDA-MB-231 cells.  
628 Cells were unsupervised clustered into three distinct groups with Louvain method. **E&F,** Function  
629 analysis of genes that were commonly up- or downregulated in insensitive cell populations. **E,** Gene  
630 function enrichment revealed that these genes were highly enriched in the cellular compromise and  
631 movement pathways. **F,** Genes that were upregulated in insensitive cell populations and predicted to  
632 promote cell movement, including VIM, SQSTM1, NPM1 and RACK1.

633

634 **Figure EV3. CASB facilitates snATAC-seq sample multiplexing. A,** The number of ATAC-barcode  
635 molecules immobilized on mESC nuclei was quantified used qPCR. The amount of ssDNA immobilized  
636 on nuclei increased with the increased usage of ConA-streptavidin-ssDNA complex, and could reached  
637 at least 80,000 molecules per nucleus. Error bars represent SD. **B,** Heatmap showing the detected  
638 levels of each CASB ATAC-barcode in individual cells in snATAC-seq. A total of 2890 cells were  
639 obtained with sufficient reads, 305 of which were identified as cell doublets and 23 cells were unlabeled.  
640 **C,** ATAC-based UMAP of MDA-MB-231 and HAP1 cells. Cells were colored according to the cell line  
641 specific barcodes. MDA-MB-231 cells with its specific CASB barcode presented as an isolated cluster.

642

643 **Figure EV4. CASB helps to reveal the dynamic transcriptome change in HAP1 cells. A,** Heatmap  
644 showing the detected levels of each CASB barcode in individual cells in scRNA-seq. A total of 3407  
645 cells were captured, 294 of which were identified as cell doublets and 9 cells were unlabeled. **B,** t-SNE  
646 projection based on the CASB barcode reads captured in scRNA-seq. Cells were colored according to  
647 the CASB barcodes, and doublets were marked in black. **C,** Transcriptome-based UMAP of HAP1 cells,  
648 in which the expression of CXCL11 were presented with color code. At later time points, CXCL11 were  
649 only actively induced in cluster 2 (circled in red).

650

651 **Supplementary Tables**

652 **Table EV1. Genes commonly up- or down-regulated in cluster 1 and 2 of untreated MDA-MB-231**  
653 **cells.**

654

655 **Table EV2. CASB barcode sequences for samples in scRNA/snATAC-seq.**

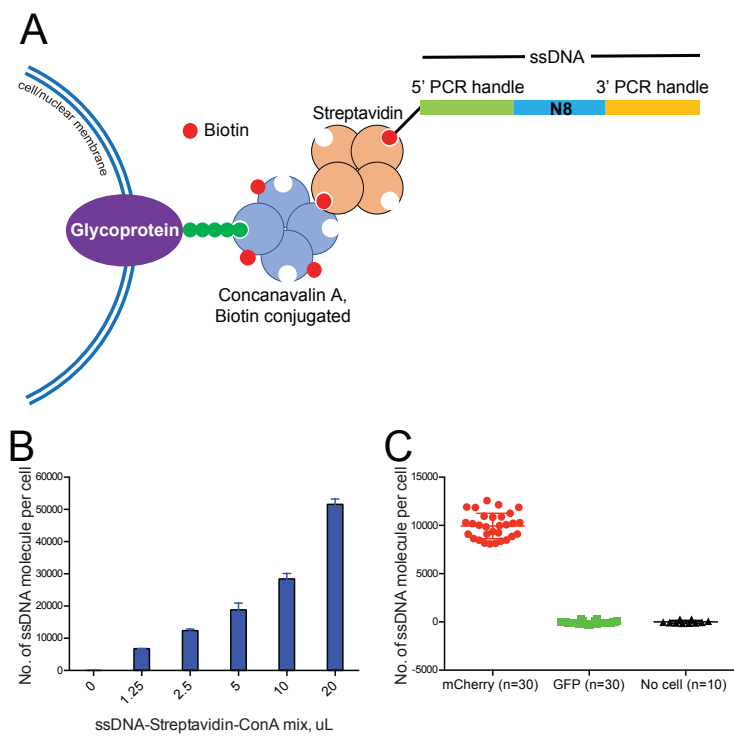


Figure 1

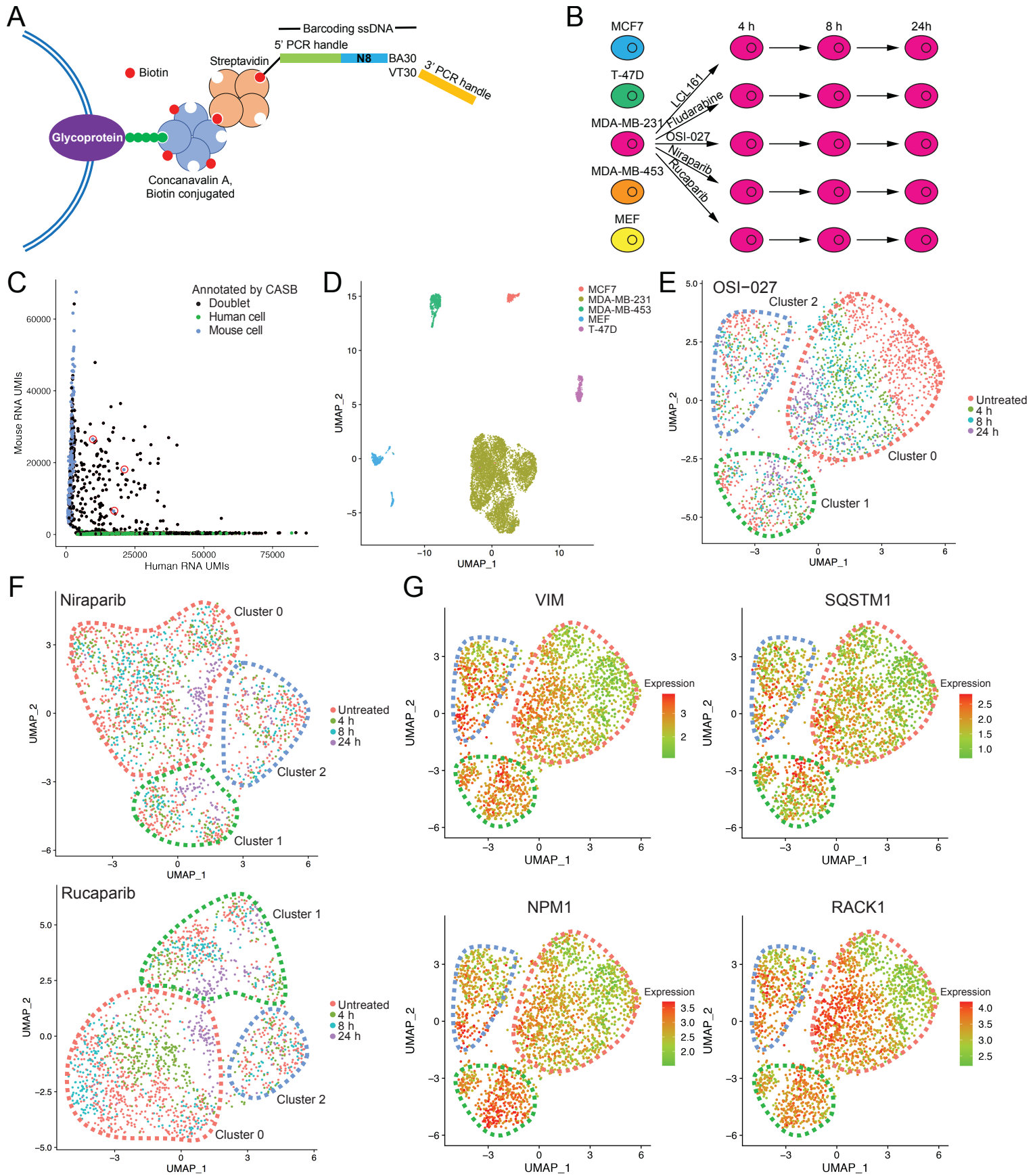
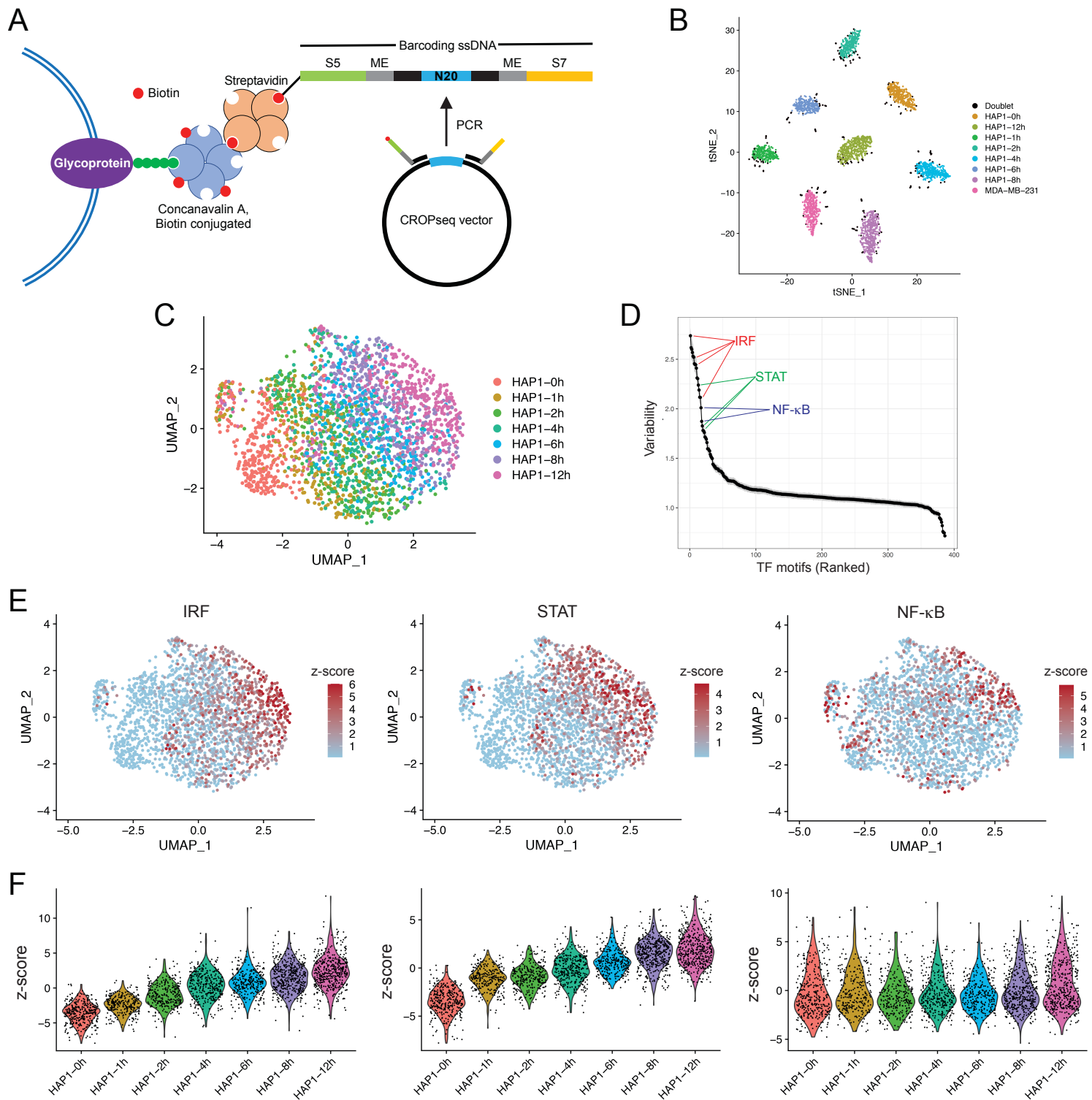


Figure 2







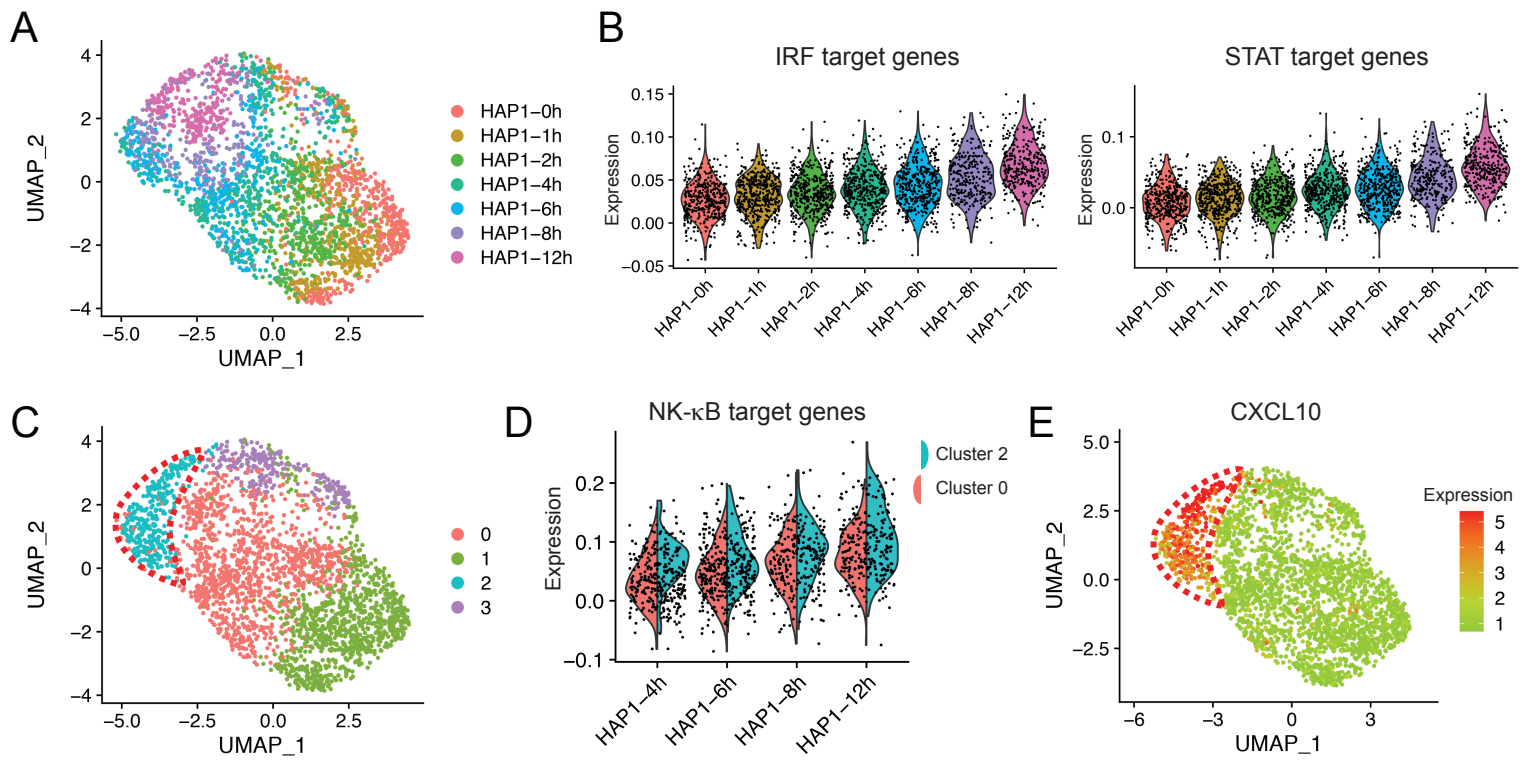


Figure 4