Chapter 11

Simultaneous Single-Cell Profiling of the Transcriptome and Accessible Chromatin Using SHARE-seq

Samuel H. Kim, Georgi K. Marinov, S. Tansu Bagdatli, Soon Il Higashino, Zohar Shipony, Anshul Kundaje, and William J. Greenleaf

Abstract

The ability to analyze the transcriptomic and epigenomic states of individual single cells has in recent years transformed our ability to measure and understand biological processes. Recent advancements have focused on increasing sensitivity and throughput to provide richer and deeper biological insights at the cellular level. The next frontier is the development of multiomic methods capable of analyzing multiple features from the same cell, such as the simultaneous measurement of the transcriptome and the chromatin accessibility of candidate regulatory elements. In this chapter, we discuss and describe SHARE-seq (**S**imultaneous **h**igh-throughput **A**TAC, and **R**NA **e**xpression with **seq**uencing) for carrying out simultaneous chromatin accessibility and transcriptome measurements in single cells, together with the experimental and analytical considerations for achieving optimal results.

Key words scRNA-seq, scATAC-seq, Multiomics, Chromatin accessibility, Transcriptomics, Split–pool

1 Introduction

The basic unit of biological organization is the individual cell. In combination with the surrounding cellular microenvironments within the context of a multicellular organism, each cell integrates across internal and external stimuli to maintain or alter its state for biological function. Understanding the cellular state at the single-cell resolution, therefore, is critical to defining the regulatory processes driving health and disease. A key advancement toward understanding cellular states has been in the development of transcriptomic methods. With the advent of high-throughput sequencing methods in the late 2000s, RNA-seq was developed to profile transcriptomes at base-pair resolutions [1–4]. Subsequently, the molecular biology approaches that enabled ever improved RNA-seq sensitivity have led to the development of

Georgi K. Marinov and William J. Greenleaf (eds.), *Chromatin Accessibility: Methods and Protocols*, Methods in Molecular Biology, vol. 2611, https://doi.org/10.1007/978-1-0716-2899-7_11,

single-cell RNA-seq (scRNA-seq) to measure transcriptomes at the single-cell level. The first scRNA-seq methods [5–8] were very low throughput, only able to measure a few cells at a time. Further technical advancements utilized microfluidics- and plate-based approaches to increase throughput to the 10^2–10^3 range [9, 10], while droplet- and bead-based methods later boosted it to the 10^4–10^5 range [11–14]. However, the approach that holds the most promise for ultra-high-throughput single-cell measurements is combinatorial indexing. The core concept of these approaches is to dynamically assign barcodes through multiple rounds of splitting and pooling cells to create a combinatorial set of barcodes that can be used to uniquely identify each cell. Specifically, a set of cells can be split into a 96- or 384-well plates, each well given a specific barcode, and then pooled back together to be randomly split into another set of plates. Iteratively performing these split–pool rounds with an optimal number of input cells, barcodes, and the number of rounds of barcoding, one can create a sufficient diversity of barcodes to uniquely assign each cell to a combination of barcodes. In comparison to physical isolation of each cell in a droplet or a well, combinatorial indexing provides a scalable platform for single-cell measurements. This is the basis of all "sci" (single-cell combinatorial indexing) methods, such as sci-RNA-seq [15] and SPLiT-seq [16].

While scRNA-seq measures the current amount of transcripts in a given cell, it does not provide insight into how that transcriptional state is achieved and maintained through regulation. Mapping active *cis*-regulatory elements (cREs) provides key insight to address this need. A common property of active cREs, originally recognized more than four decades ago [17–19], is that they are depleted of nucleosomes and exhibit an open, "accessible" conformation. This property has been the basis for numerous methods that have been developed over the years to profile these elements [20], which rely on the preferential enzymatic cleavage or labeling of open chromatin regions. ATAC-seq [21, 22] (**A**ssay for **T**ransposase-**A**ccessible **C**hromatin using **seq**uencing) has emerged as the most versatile instance of such assays. ATAC-seq takes advantage of the preferential insertion of a hyperactive Tn5 [23] transposase, preloaded with sequencing adapters, into open chromatin. Tn5 had been previously adapted and successfully used for the generation of high-throughput sequencing libraries from low-input DNA samples [24]. The realization that it can also be used to tag open chromatin regions with ready-for-amplification sequencing adapters in a single reaction allowed for chromatin accessibility profiling to be carried out in bulk on very low-input samples (typically 50,000 cells, but also down to just a few thousand [21]), and eventually in single cells, in the form of scATAC-

seq, in the mid-2010s[25]. As with scRNA-seq, the throughput of scATAC-seq has also been dramatically increased over the years, using combinatorial indexing (sciATAC-seq [26–28]), microwell plates (μATAC-seq [29]), droplet-based methods [30], and combinations of combinatorial indexing and droplets (dsciATAC-seq [31]).

Techniques such as scRNA-seq and scATAC-seq have provided unprecedented insights into the diversity of cell types, their developmental dynamics, and cellular responses to external stimuli in a wide variety of context. However, the ideal measurements would provide information about all relevant aspects of the state from the same cell. To this end, a variety of single-cell multiomic methods, measuring multiple such modalities in the same individual cells, have been under active development in recent years. These include methods for sequencing the genomes and transcriptomes of single cells (G & T-seq [32], PRDD-seq [33], DNTR-seq [34], sci-L3-RNA/DNA [35], TARGET-seq [36], and others), for sequencing methylomes and transcriptomes (scTrio-seq [37], scMT-seq [38], and scM & T-seq [39]), for mapping accessible chromatin and methylomes (e.g., scNOMe-seq [40]), for measuring proteins and transcripts (REAP-seq [41], CITE-seq [42], QBC [43], inCITE-seq [44], iNS-seq [45], using methylation-based labeling of open chromatin to map accessible DNA and transcripts (COOL-seq [46], scNMT-seq [47], scNOMeRe-seq [48], snmC2T-seq [49]), mapping protein occupancy and transcriptomes (CoTECH [50], Paired-Tag [51], scDam & T-seq [52]), for quantifying proteins levels and mapping open chromatin (PHAGE-ATAC [53], ASAP-seq [54]), for quantifying proteins and transcriptome levels and mapping open chromatin (DOGMA-seq [54], TEA-seq [55]), and others [56].

As regulatory elements and RNA levels are the two perhaps most informative modalities, joint scATAC-seq + scRNA-seq methods are the most sought after multiomic assays. A number of these have been developed in recent years—sci-CAR-seq [57], Paired-seq [58], ASTAR-seq [59], SNARE-seq [60], SHARE-seq [61], and others. The ideal such assay should capture as many of the transcripts present in each cell as possible and also as many of the open chromatin regions in the nucleus, with high specificity and little noise. The SHARE-seq assay, which is based on the combinatorial indexing described above, provides high-quality and high-throughput transcriptome and accessible chromatin measurements in the same single cells.

In this chapter, we describe in detail the SHARE-seq procedure and discuss the key optimization points and considerations for the generation of high-quality scATAC+scRNA-seq datasets.

2 Materials

2.1 DNA Oligos and Primers

All oligonucleotides can be obtained through IDT. The exact scale and purification methods are listed below:

1. Round 1 linker (1 μmol scale, standard desalting):
   CCGAGCCCACGAGACTCGGACGATCATGGG
2. Round 2 linker (1 μmol scale, standard desalting):
   CAAGTATGCAGCGCGCTCAAGCACGTGGAT
3. Round 3 linker (1 μmol scale, standard desalting):
   AGTCGTACGCCGATGCGAAACATCGGCCAC
4. Round 1 blocking (1 μmol scale, standard desalting):
   CCCATGATCGTCCGAGTCTCGTGGGCTCGG
5. Round 2 blocking (1 μmol scale, standard desalting):
   ATCCACGTGCTTGAGCGCGCTGCATACTTG
6. Round 3 blocking (1 μmol scale, standard desalting):
   GTGGCCGATGTTTCGCATCGGCGTACGACT
7. Read 1 (100 nmol scale, HPLC purified):
   TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
8. Template Switching Oligo (TSO) (100nmol scale, HPLC purified):
   AAGCAGTGGTATCAACGCAGAGTGAATrGrG+G
9. RNA PCR primer (100 nmol scale, standard desalting):
   AAGCAGTGGTATCAACGCAGAGT
10. P7 primer (100 nmol scale, standard desalting):
    CAAGCAGAAGACGGCATACGAGAT
11. Phosphorylated Read2 (100 nmol scale, HPLC purified):
    /5Phos/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG—
12. Reverse transcription primer (RT primer) (100 nmol scale, HPLC purified)
    /5Phos/GTCTCGTGGGCTCGGAGATGTGTATAAGAGA-
    CAGNNNNNNNNN/iBiodT/TTTTTTTTTTTTTTVN
13. Blocked_ME_Comp (100 nmol scale, HPLC purified):
    /5Phos/C*T*G* T*C*T* C*T*T* A*T*A* C*A*/3ddC/
14. Pool–split ligation Plate R1 (*see* **Note 4**:
    /5Phos/CGCGCTGCATACTTG[8-bp-barcode]
    CCCATGATCGTCCGA
15. Pool–split ligation Plate R2 (*see* **Note 4**:
    /5Phos/CATCGGCGTACGACT[8-bp-barcode]
    ATCCACGTGCTTGAG
16. Pool–split ligation Plate R3 (*see* **Note 4**:
    CAAGCAGAAGACGGCATACGAGAT[8-bp-barcode]
    GTGGCCGATGTTTCG

17. PCR Library indexing primers plate:
    AATGATACGGCGACCACCGAGATCTACAC[8bp-index]
    TCGTCGGCAGCGTCAGATGTGTAT

    An example set of 96 barcodes are listed below:

    AACGTGAT AAGGTACA CACTTCGA GATAGACA TGGAACAA ATCATTCC
    AAACATCG ACACAGAA CAGCGTTA GCCACATA TGGCTTCA ATTGGCTC
    ATGCCTAA ACAGCAGA CATACCAA GCGAGTAA TGGTGGTA CAAGGAGC
    AGTGGTCA ACCTCCAA CCAGTTCA GCTAACGA TTCACGCA CACCTTAC
    ACCACTGT ACGCTCGA CCGAAGTA GCTCGGTA AACTCACC CCATCCTC
    ACATTGGC ACGTATCA CCGTGAGA GGAGAACA AAGAGATC CCGACAAC
    CAGATCTG ACTATGCA CCTCCTGA GGTGCGAA AAGGACAC CCTAATCC
    CATCAAGT AGAGTCAA CGAACTTA GTACGCAA AATCCGTC CCTCTATC
    CGCTGATC AGATCGCA CGACTGGA GTCGTAGA AATGTTGC CGACACAC
    ACAAGCTA AGCAGGAA CGCATACA GTCTGTCA ACACGACC CGGATTGC
    CTGTAGCC AGTCACTA CTCAATGA GTGTTCTA ACAGATTC CTAAGGTC
    AGTACAAG ATCCTGTA CTGAGCCA TAGGATGA AGATGTAC GAACAGGC
    AACAACCA ATTGAGGA CTGGCATA TATCAGCA AGCACCTC GACAGTGC
    AACCGAGA CAACCACA GAATCTGA TCCGTCTA AGCCATGC GAGTTAGC
    AACGCTTA GACTAGTA CAAGACTA TCTTCACA AGGCTAAC GATGAATC
    AAGACGGA CAATGGAA GAGCTGAA TGAAGAGA ATAGCGAC GCCAAGAC

2.2 General Reagents

1. Eppendorf ThermoMixer C (96-well plate adapter)
2. Tabletop centrifuge
3. Swing bucket centrifuge with temperature control
4. Thermal cycler
5. Cold room
6. qPCR machine (QuantStudio 3)
7. Qubit fluorometer or equivalent
8. E-gel electrophoresis system (Thermo Fisher Scientific)
9. TapeStation (Agilent) or equivalent, e.g., BioAnalyzer (Agilent).
10. Multichannel pipettes or liquid handling instruments
11. gentleMACS Dissociator (Miltenyi Biotec)
12. Automated cell counter, e.g., Countess 3 (Thermo Fisher Scientific) or equivalent.

2.3 General Equipment

1. 1× PBS buffer solution (Thermo Fisher Scientific, Cat #10010049)
2. Bovine Albumin Fraction V (7.5% solution) (Thermo Fisher Scientific, Cat #15260037)
3. Trypan Blue Stain (0.4%) (Thermo Fisher Scientific, Cat #T10282)

4. Enzymatic RI (Qiagen, Cat #Y9240L)
5. SUPERase RI (Thermo Fisher Scientific, Cat #AM2696)
6. Lucigen RI (Lucigen Cat # 30281-2)
7. Protector RI (Sigma Aldrich Cat # 3335399001)
8. 16% FA (Thermo Fisher Scientific, Cat # 28906)
9. Glycine (Sigma Aldrich, Cat #50049)
10. 1 M Tris HCl pH 7.5 (Thermo Fisher Scientific, Cat #15567027)
11. 1 M Tris HCl pH 8.0 (Thermo Fisher Scientific, Cat #15568025)
12. 5 M NaCl (Thermo Fisher Scientific, Cat #AM9760G)
13. 1 M $MgCl_2$ (Sigma Aldrich, Cat #63069)
14. 1 M $CaCl_2$ (Sigma Aldrich, Cat #21115-100ML)
15. DMF (Dimethyl Formamide) (Sigma, Cat #227056)
16. 0.2 M Tris-acetate pH 7.8 (Bioworld, Cat #40120265-2)
17. 5 M Potassium acetate (Sigma Aldrich, Cat #95843-100ML-F)
18. 1 M Magnesium acetate (Sigma Aldrich, Cat #63052-100ML)
19. 10% NP-40 (Thermo Fisher Scientific, Cat #28324)
20. Buffer EB (Qiagen, Cat #19086)
21. PEG 6000 (Sigma Aldrich, Cat #528877)
22. Maxima H Minus Reverse Transcriptase with buffer (Thermo Fisher Scientific, Cat #EP075)
23. 10 mM dNTPs (NEB, Cat #N0447L)
24. T4 DNA Ligase (NEB, Cat #M0202L)
25. Additional 10× T4 Ligase buffer (NEB, Cat #B0202S)
26. Proteinase K (20 mg/mL) (NEB, Cat #P8107S)
27. 20% SDS (VWR, Cat #97062+440)
28. 100 mM PMSF/IPA (Sigma Aldrich, Cat # P7626)
29. cOmplete Protease Inhibitor Cocktail (Sigma Aldrich, Cat # 11697498001)
30. 0.5 M EDTA (Sigma Aldrich, Cat #AM9260G)
31. Tween-20 (Sigma Aldrich, Cat #P9416-100ML)
32. Digitonin (Promega, Cat #G9441)
33. MyOne C1 Dynabeads (Thermo Fisher Scientific, Cat #65001)
34. Ficoll PM-400 (20%) (Sigma Aldrich, Cat #F5415-25ML)
35. KAPA HiFi 2× mix (Fisher Scientific, Cat #NC0295239)
36. SPRIselect beads (Beckman Coulter, Cat #B23318)

37. 100% EtOH
38. 100 mM DTT (Thermo Fisher Scientific, Cat #707265ML)
39. NEBnext 2× Mix (NEB, Cat #M0541L)
40. Glycerol (Thermo Fisher Scientific, Cat #15514011)
41. TD buffer from Nextera kit
42. SYBR Green I Nucleic Acid Gel Stain (Thermo Fisher Scientific, Cat #S7563)
43. EVAGreen Dye, 20x in water (Biotium, Cat #31000)
44. Nuclease-free H_2O
45. 96-well plates (Eppendorf, Cat #0030129300) (preferably low protein and DNA binding; *see* **Note 5**)
46. 1.5-mL microcentrifuge tubes, preferably low protein and DNA binding (*see* **Note 5**)
47. 2-mL, 15-mL, and 50-mL tubes
48. gentleMACS M-Tubes (Miltenyi Biotec, Cat #130-093-236)
49. 30 μm Sterile single-pack CellTrics filters (Sysmex, Cat #04-004-2326)
50. 200-μL PCR tubes
51. Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Cat # Q32851)
52. TapeStation D1000 and D5000 tape and reagents (Agilent)
53. Tn5 transposase (*see* **Note 1**)
54. MinElute PCR Purification Kit (Qiagen Cat# 28004/28006), Zymo DNA Clean and Concentrator Kit (Zymo Cat# D4013/D4014), or equivalent

2.4 Buffers and Reagents

Make all buffers using ultrapure molecular biology-grade ddH_2O:

1. 2.5M Glycine (50 mL)

   9.375 g Glycine (powder)

   1× PBS up to 50 mL

   Filter through a 0.22 μM filter. Store at room temperature.

2. Tissue Dissociation (MACS) buffer

   10 mM Tris-HCl pH 8.0

   5 mM $CaCl_2$

   5 mM EDTA

   3 mM MgAc

   0.6 mM DTT

   cOmplete Protease Inhibitor

   Make fresh every time.

3. Nuclei Isolation Buffer (NIB)
   10 mM Tris-HCl pH 7.4
   10 mM NaCl
   3 mM $MgCl_2$
   0.1% IGEPAL CA-630
   Store at 4 °C.
4. 2× TD buffer
   20 mM Tris-HCl pH 7.6
   10 mM $MgCl_2$
   20% Dimethyl Formamide
   Store at − 20°C.
5. PEG 6000 50%
   Mix equal mass of PEG6000 and H_2O, heat to 65 °C) for 4 min, and then cool down to room temperature.
6. 2× RCB buffer
   100 mM Tris pH 8.0
   100 mM NaCl
   0.40% SDS
   Store at room temperature.
7. 2× BW buffer
   10 mM Tris pH 8.0
   2 M NaCl
   1 mM EDTA
   Store at 4 °C.
8. 1× B & W-T Buffer
   5 mM Tris pH 8.0
   1 M NaCl
   0.5 mM EDTA
   0.05% Tween-20
   Store at 4 °C.
9. Oligo resuspension buffer (IDTE)
   10 mM Tris pH 8.0
   0.1 mM EDTA
   Store at room temperature.
10. Oligo annealing buffer (STE)
    10 mM Tris pH 8.0
    50 mM NaCl
    1 mM EDTA
    Store at room temperature.

11. Dilution buffer
    50% glycerol
    50 mM Tris pH 7.5
    100 mM NaCl
    0.1 mM EDTA
    0.1% NP-40
    Store at −20°C.

2.5 Software Packages

1. `Bowtie` [62] (http://bowtie-bio.sourceforge.net/index.shtml).
2. `SAMtools` [63]: http://www.htslib.org/
3. `PicardTools` https://broadinstitute.github.io/picard/
4. UCSC Genome Browser [64, 65] utilities: http://hgdownload.cse.ucsc.edu/admin/exe/
5. `STAR` [66] https://github.com/alexdobin/STAR
6. `R`: https://www.r-project.org/
7. `Python` (version 2.7 or higher) https://www.python.org/
8. `ArchR` [67]: https://www.archrproject.com/
9. `Seurat` [68]: https://satijalab.org/seurat/
10. Additional scripts: https://github.com/georgimarinov/GeorgiScripts. Contains python scripts used in the examples shown below; some of the scripts depend on having `pysam` (https://pysam.readthedocs.io/en/latest/index.html) and `pyBigWig` (https://github.com/deeptools/pyBigWig) installed.

3 Methods

The general outline of the SHARE-seq assay is shown in Fig. 1. The first of the two basic ideas behind SHARE-seq and other pool–split-based assays is to label molecules originating from each cell with a unique combination of barcodes that are added serially and randomly by pooling cells and then randomly redistributing them across subsequent sets of barcodes, thus ensuring that statistically each cell can be identified through a unique combination of barcodes. The second is the separation of chromatin and transcriptome molecules through the use of a biotinylated reverse transcription (RT) primer, which can then be used for a streptavidin pulldown of the transcriptome.

In brief, before the beginning of a SHARE-seq experiment, the needed barcode plates and transposases are prepared and stored. The experiment itself begins with the isolation of nuclei from cells in culture or from tissues (*see* **Note 2**. Nuclei are then crosslinked,

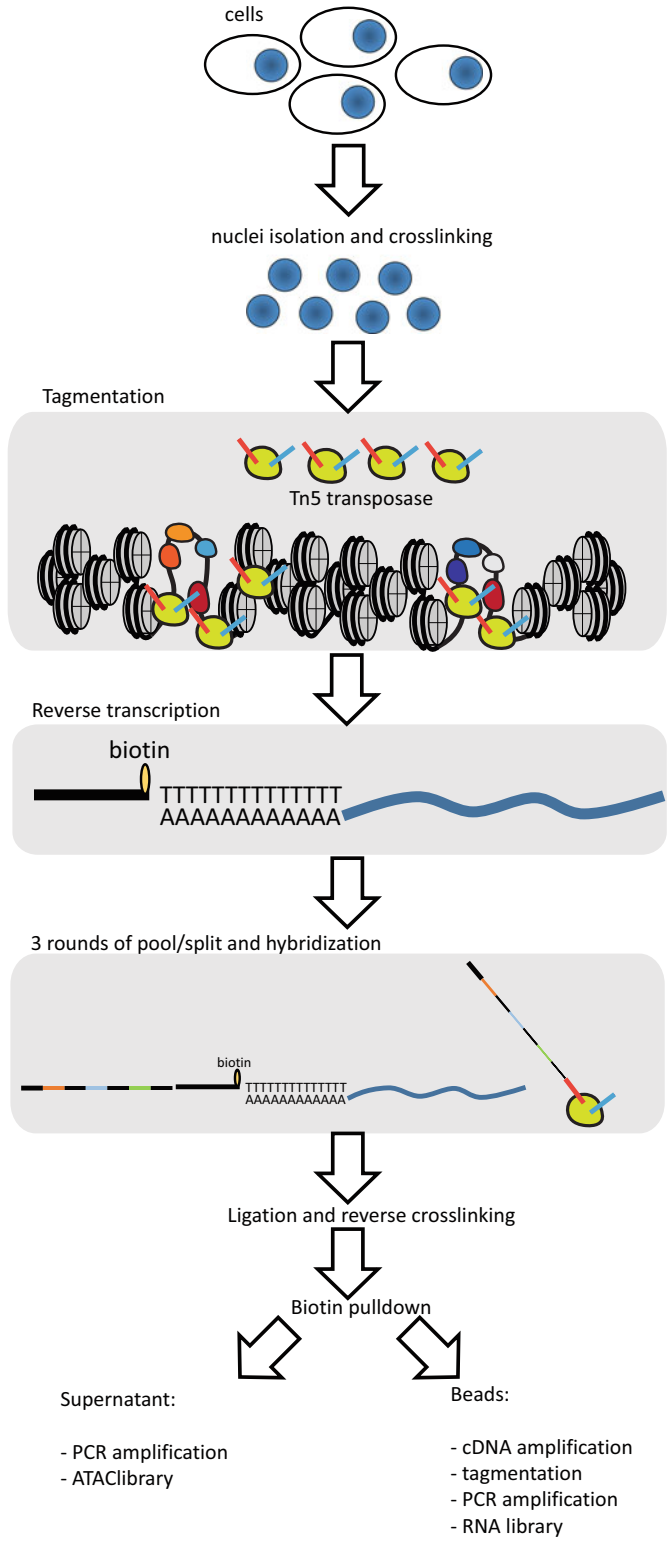

Fig. 1 Outline of the SHARE-seq assay. Nuclei are isolated from cells or tissues and crosslinked. Transposition is then carried out on chromatin, followed by reverse transcription with a biotinylated RT primer. Three pool–split rounds of hybridization of barcode oligos are then performed. Hybridized barcodes are then ligated, and crosslinks are reversed. The ATAC and RNA portions are separated by streptavidin pulldown. The ATAC is directly amplified, and the RNA is subjected to cDNA amplification, tagmentation, and final library amplification

usually lightly (*see* **Note 3**). Transposition is then carried out, followed by reverse transcription using a biotinylated RT primer containing a random unique molecular identifier (UMI). Three rounds of pool–split hybridization and blocking are then carried out, after which the hybridized oligos are ligated into single molecules to each other and to the transposed chromatin fragments and reverse transcribed mRNA. Crosslinks are then reversed, and streptavidin pulldown is used to separate the chromatin from the transcriptome. ATAC libraries are directly amplified from the supernatant. The transcriptome is first amplified on-beads into cDNAs, which are then tagmented into sequenceable fragments and PCR-amplified into final libraries.

The resulting library structures for ATAC and RNA are shown in Fig. 2. ATAC libraries contain three barcodes, while RNA libraries also include the UMI. Note that with many Illumina-based sequencing readouts, the first barcode to be read is actually the third one added during the pool–split procedure.

3.1 Determining the Optimal Cell Number

It is important to carefully track the number of cells going into the SHARE-seq assays and being retained at each key step of the procedure. Pool–split assays rely on the statistical uniqueness of barcode combinations through which cells pass, which in turn means that having too many cells entering the pool–split procedure will lead to an unacceptably high rate of doublets (two or more cells with the same barcode). In the same time, some of the reactions have an efficiency-imposed limit on the number of cells that can enter them and need to be distributed into parallel reactions for optimal results. This applies to the initial transposition and reverse transcription reactions, as well as to the final amplification, where the existing protocol is optimized for libraries of size ~20,000 cells, which means that after the final pooling cells are split into separate subpools of that size and processed into individual sublibraries.

Figure 3 shows the theoretical number of detected cells and doublet rate for different pool–split setups with three rounds, accounting for a certain level of cell loss during repeated handling. Based on these calculations and empirical experience, we usually start the pool–split rounds with $\sim 5 \times 10^5$ cells for a 96 × 96 × 96 pool–split experiment.

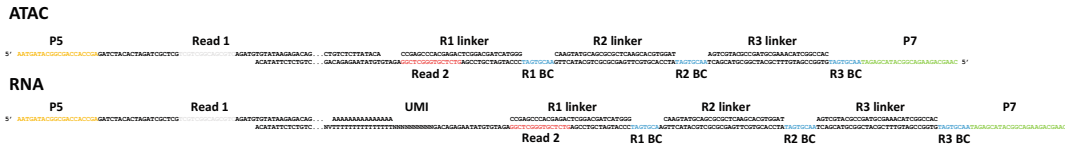

Fig. 2 Structure of final SHARE-seq libraries. ATAC (top) and RNA (bottom). Dots represent the actual library insert

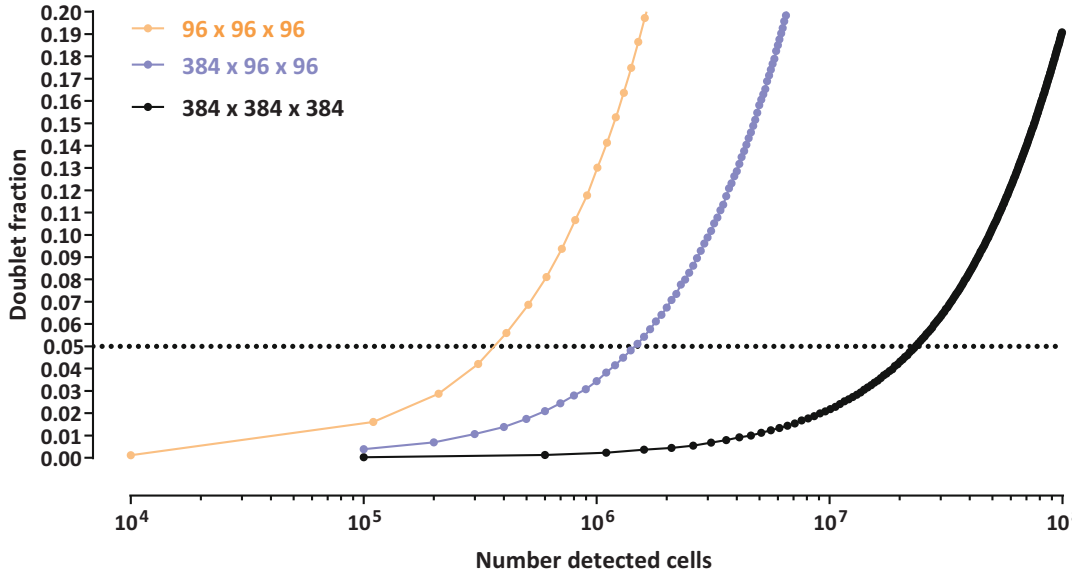

Fig. 3 Combinatorial indexing and SHARE-seq's throughput. Shown is the number of cells that can be detected at a given doublet rate; the pool–split process was simulated as a random Poisson loading at a 50% loss of cells during each pool–split round

3.2 Annealing of Oligo Plates

In this step, barcode containing oligonucleotides for each round of split–pool is annealed and distributed into 96-well plates prior to the actual assay. These plates can be stored at − 20°C indefinitely. It is advisable for the purposes of time saving to prepare sufficiently many such plates in advance to support multiple experiments. It is critical to thaw these plates to room temperature prior to use.

See **Note 4**.

1. Dilute Round 1 linker oligos (120 μL at 1 mM concentration) with 11,880 μL STE buffer.
2. Mix 90 μL diluted Round 1 linker oligo with 10 μL Round 1 oligo (at 100 μM) in the wells of a multiwell plate.
3. Dilute Round 2 linker oligos (120 μL at 1 mM concentration) with 9480 μL STE buffer.
4. Mix 88 μL diluted Round 2 linker oligo with 12 μL Round 2 oligo (at 100 μM) in the wells of a multiwell plate.
5. Dilute Round 3 linker oligos (144 μL at 1 mM concentration) with 9360 μL STE buffer.
6. Mix 86 μL diluted Round 3 linker oligo with 14 μL Round 3 oligo (at 100 μM) in the wells of a multiwell plate.
7. Anneal the Round 1, Round 2, and Round 3 plates as follows in a thermocycler:

2 min at 95 °C

Slow ramp at − 1°C per minute to 20 °C

2 min at 20 °C

Indefinitely at 4 °C

8. Check if there has been significant water evaporation for wells situated at the corners. If yes, add water to equalize volumes.
9. Aliquot 10 μL of the annealed oligos to new plates. This should be enough for 9 experiments. Store these plates at − 20°C.

3.3 Anneal Adapter Oligos

In this step, Tn5 adapters are prepared for both transposition of chromatin and tagmentation during cDNA library preparation:

1. Dilute the Phosphorylated Read2, Read1, and Blocked ME Comp oligos to a 100 μM concentration with the IDTE buffer.
2. Prepare the transposition adapter mix in a PCR tube as follows:

   6.5 μL 100 μM Phosphorylated Read2 oligo

   6.5 μL 100 μM Read1 oligo

   13 μL 100 μM Blocked ME Comp oligo

   0.26 μL 1 M Tris pH 8.0

   0.26 μL 5 M NaCl
3. Prepare the tagmentation adapter mix in a PCR tube as follows:

   13 μL 100 μM Read1 oligo

   13 μL 100 μM Blocked ME Comp oligo

   0.26 μL 1 M Tris pH 8.0

   0.26 μL 5 M NaCl
4. Anneal oligos as follows in a thermocycler:

   2 min at 85 °C

   Slow ramp at − 1°C per minute to 20 °C

   2 min at 20 °C

   Indefinitely at 4 °C
5. Heat glycerol to 65 °C, and then equilibrate to room temperature.
6. Mix 25 μL glycerol with 25 μL of annealed oligo.

The annealed adapters can be immediately used or stored at − 20°C.

3.4 Transposome Assembly

In this step, Tn5 transposomes are assembled together with the annealed adapter oligos:

1. Assemble Tn5 transposomes by mixing the following components:

0.625× *N* 1× home-made Tn5

0.625× *N* dilution buffer

1.25× *N* annealed transposition adapter with glycerol

Total volume: 2.5× *N*

2. Incubate at room temperature for 30 min.

   The assembled transposome can be stored at − 20°C for up to 2 weeks.

3.5 Tissue Dissociation

Here, we describe an example tissue dissociation protocol that has worked successfully in our hands for several human embryonic tissues. However, users should be aware that generally each tissue requires separate optimization of dissociation conditions, and it is likely that a different protocol will have to be adapted in most situations.

1. Set swing bucket centrifuge to 4 °C Fast Temp and thaw 1M DTT.
2. Transfer tissue samples onto dry ice.
3. Prepare MACS buffer (2 mL for each sample) as described above. Make sure the buffer is cold on ice.
4. Add 10 μL Protector RNase Inhibitor for each 1 mL in GentleMACS M-tubes. Add 1 mL of MACS buffer to each GentleMACS M-tube and chill on ice.
5. Transfer 30–50 mg of tissue into each GentleMACS M-tube containing 1 mL MACS buffer.
6. Allow the tissue to thaw in buffer. Transition to a cold room.
7. Homogenize using a `Protein_01_01` dissociation protocol on a GentleMACS Tissue Dissociator instrument.
8. Filter the homogenate through 30 μm CellTrics filter into a 2mL DNA LoBind tube by pipetting directly onto the top of the filter and gently tapping to allow flow.
9. Wash the GentleMACS M-tube with 1 mL MACS buffer and filter the wash again through the 30 μm CellTrics filter.
10. Spin down the homogenate in a swing bucket centrifuge at 500 *g* for 5 min at 4 °C (ramp up and down both at 3/9).
11. Remove and discard supernatant.
12. Resuspend in 1mL PBS-2RI.
13. Count cells/nuclei and proceed with a desired number of cells/nuclei.

3.6 Fixation of Cells in Culture and of Dissociated Nuclei from Tissue

The next step, if starting with a dissociated tissue, is to fix the nuclei. This is also the first step if starting with cells in culture. The procedure used is generally the same, with the difference that with nuclei the first step is directly the fixation:

1. Prepare PBS-2RI Buffer (~4 mL) by mixing the following:

   4 mL 1× PBS

   21.4 μL 7.5% BSA

   10 μL Enzymatic RI

   5 μL SUPERase RI

   Keep on ice.

2. Prepare NIB-RI Buffer (~8 mL) by mixing the following:

   8 mL NIB

   20 μL Enzymatic RI

   20 μL SUPERase RI

   Keep on ice.

3. Spin down cells at 500 *g*.
4. Wash cells with 0.5 ml PBS-2RI.
5. Count cells with Trypan blue.
6. Resuspend cells with cold PBS-2RI at a concentration of 1×10^6 cells/mL.
7. For each 1 mL of cells in PBS-2RI, add 66.7 μL of 1.6% FA (final concentration 0.1% FA) for cells or 66.7 μL of 3.2% FA for tissues. Mix and incubate at room temperature for 5 min.
8. Quench the reaction by adding to each 1 mL of cells in PBS-2RI the following:

   56.1 μL 2.5 M Glycine

   50 μL 1M Tris pH 8.0

   13.3 μL of 7.5% BSA

   Mix well and incubate on ice for 5 min.

9. Spin down at 500 *g*. Remove supernatant, and add 0.5 mL PBS-2RI without disturbing the cell pellet.
10. Prepare RSB-RI by mixing the following:

    2.5 μL 1 M Tris-HCl pH 7.5

    0.5 μL 5 M NaCl

    0.75 μL 1 M $MgCl_2$

    2.5 μL 10% Tween-20

    2.5 μL 10% NP-40

    2.5 μL 1% Digitonin

    33.3 μL 7.5% BSA

    0.25 μL 1 M DTT

    204 μL Ultrapure water

    1.25 μL Enzymatic RI

11. Spin down again at 500 *g*. Remove supernatant, and resuspend cells in 100 μL RSB-RI and incubate on ice for 3 min.
12. Prepare RSB-T by mixing the following:

    25 μL 1 M Tris-HCl pH 7.5

    5 μL 5 M NaCl

    7.5 μL 1 M $MgCl_2$

    25 μL 10% Tween-20

    333.3 μL 7.5% BSA

    2.5 μL 1 M DTT

    2089.5 μL Ultrapure water

    12.5 μL Enzymatic RI
13. Pipette 1 mL of RSB-T to cells and mix. Spin down at 500 *g* for 5 min.

3.7 ATAC Reaction

In this step, transposition of the entire sample is performed by splitting it into 10,000–20,000 cells in 50-μL reactions each in a 96-well plate. The smaller volume and the number of cells per reaction improve the quality of transposition.

The cell lysis conditions described here are adapted from the omniATAC bulk ATAC protocol [22] (*see* **Note 7**):

1. Prepare PBS-RI by mixing the following:

    800 μL PBS

    2 μL Enzymatic RI
2. After the last centrifugation, remove supernatant and resuspend the cells with PBS-RI to 2×10^6 cells/mL.
3. Prepare 2× TB buffer (sufficient for 96 reactions) by mixing the following:

    874.5 μL 0.2 M Tris-acetate

    70 μL 5 M Potassium acetate

    53 μL 1 M Magnesium acetate

    53 μL 10% Tween-20

    53 μL 1% Digitonin

    848 μL 100% DMF

    698.5 μL H_2O
4. Prepare 1× TB buffer according to the number of reactions N to be carried out. $N = 1$ corresponds to 10^4 input cells.

    25× N 2× TB

    16.45× N H_2O

0.2× *N* PIC

0.85× *N* Enzymatic RI

Total volume: 42.5× *N*.

5. Aliquot 5× *N*μL of the diluted cells to a new tube, e.g., for 10×10^5 cells, $N=10$, so aliquot 50 μL cells to a new tube.
6. Add 42.5× *N*μL 1× TB to sample.
7. Add 2.5× *N*μL of assembled Tn5 to sample. Mix well.
8. Aliquot 50 μL of sample in the wells of a 96- or 384-well plate.
9. Seal the plate and incubate with shaking at 500 rpm for 30 min at 37 °C.
10. Pool the reactions and spin down at 500 *g*.
11. Add 0.5 mL NIB-RI without disturbing the pellet and spin down again at 500 *g*.
12. Resuspend the cells in 60 μL EB.

3.8 Reverse Transcription

In this step, reverse transcription is performed in situ. The conditions are optimized for 1×10^5 cells entering each 50-μL reaction:

1. Prepare the reverse transcription (RT) mix (sufficient for 6 reactions) as follows:

   70 μL 5× RT buffer

   2.19 μL Enzymatics RNase Inhibitor

   4.38 μL SUPERase RI

   17.5 μL dNTPs

   35 μL RT Primer

   10.94 μL H2O

   105 μL 50% PEG

   35 μL Maxima H Minus Reverse Transcriptase (add right before RT reaction)

   Total volume: 280 μL.
2. Add 240 μL RT mix to 60 μL cells in EB.
3. Aliquot 50 μL to 6 PCR wells.
4. Start thawing the oligo plates, while the RT is ongoing.
5. Run the reverse transcription reaction in a thermocycler as follows:

   50 °C for 10 min

   3 cycles of:

   8 °C for 12 s

   15 °C for 45 s

   20 °C for 45 s

30 °C for 30 s
42 °C for 2 min
50 °C for 3 min
50 °C for 5 min

6. Pool samples and mix with 500 μL NIB-RI.
7. Spin down at 500 *g*.
8. Wash with 1000 μL NIB.
9. Spin down at 500 *g*.
10. Resuspend with 1152 μL NIB-RI.

3.9 Hybridization–Ligation and Pool–Split

In this step, cells/nuclei are iteratively split into individual wells to dynamically create a combinatorial index statistically unique to each cell. All handling is performed at room temperature so make absolutely sure that oligo plates have been fully thawed before proceeding.

If different samples are multiplexed in a single run, they can be individually identified based on the first-round barcodes. If such a strategy is deployed, each sample needs to be processed through transposition and reverse transcription separately and then loaded into specified positions in the first-round plate(s).

1. Prepare 3456 μL hybridization buffer as follows:
   2761.9 μL H_2O
   576 μL 10× T4 ligase buffer
   14.4 μL SUPERase RI 20 U/μL
   46.08 μL Enzymatics RI 40 U/μL
   57.60 μL 10% NP40
2. Mix 1152 μL of sample with 3456 μL hybridization buffer. Keep the sample at RT.
3. Aliquot 40 μL of mixture to a Round 1 plate.
4. Mix and shake at 300 rpm for 30 min at RT.
5. Prepare 1152 μL Blocking Oligo 1 mix as follows:
   253.4 μL 100 μM Round 1 blocking oligo
   211.2 μL 10× T4 DNA Ligase buffer
   687.4 μL H_2O
6. Add 10 μL Blocking Oligo 1 mix to each well.
7. Mix and shake at 300 rpm for 30 min at RT.
8. Pool samples from all wells.
9. Aliquot 50 μL of mixture to a Round 2 plate.
10. Mix and shake at 300 rpm for 30 min at RT.

11. Prepare 1152 μL Blocking Oligo 2 mix as follows:
    304.1 μL 100 μM Round 2 blocking oligo
    211.2 μL 10× T4 DNA Ligase buffer
    636.7 μL H_2O
12. Add 10 μL Blocking Oligo 2 mix to each well.
13. Mix and shake at 300 rpm for 30 min at RT.
14. Pool samples from all wells.
15. Aliquot 60 μL of mixture to a Round 2 plate.
16. Mix and shake at 300 rpm for 30 min at RT.
17. Prepare 1152 μL Blocking Oligo 3 mix as follows:
    265.0 μL 100 μM Round 3 blocking oligo
    11.5 μL 10% NP-40
    875.5 μL H_2O
18. Add 10 μL Blocking Oligo 1 mix to each well.
19. Mix and shake at 300 rpm for 30 min at RT.
20. Pool samples from all wells.
21. Spin down at 500 *g* 5 min.
22. Wash with 1 mL NIB-RI.
23. Spin down at 500 *g* 5 min.
24. Resuspend in 80 μL NIB-RI.
25. Prepare 320 μL Ligation mix as follows:
    3.2 μL Enzymatics RI
    1.00 μL SUPERase RI
    40 μL 10× T4 DNA Ligase Ligation buffer
    20 μL T4 DNA Ligase 400 U/μL
    251.8 μL H2O
    4 μL 10% NP40
26. Mix sample with the 320 μL Ligation mix.
27. Aliquot 8× 50 μL in PCR tubes.
28. Shake at 300 rpm for 30 min at RT.
29. Pool samples from all tubes.
30. Spin down at 500 *g* 5 min.
31. Wash with 1 mL NIB-RI.
32. Spin down at 500 *g* 5 min.
33. Resuspend in 400 μL NIB-RI.
34. Count the number of nuclei.

Note: If fewer cells are preferred per sub-library, count cells to desired concentration and add more NIB to make the volume up to 50 μL per sub-library.

3.10 Reverse Crosslinking

In this step, cells are reverse crosslinked to release DNA from the bound proteins so that the ATAC libraries can be amplified. As the crosslinking is relatively gentle (at 0.1 or 0.2%), a milder reverse crosslinking condition of 1 h incubation at 55 °C is generally sufficient.

Further reverse crosslinking optimization might be needed if the crosslinking protocol has been modified:

1. For each N of 50-μL sub-library, add the following:
   50 μL 2× RCB
   2 μL Proteinase K
   1 μL SUPERase RI
2. Incubate at 55 °C for 1 h.
3. Add 5 μL 100 mM PMSF/IPA.
4. Incubate at room temperature for 10 min.

Note: this is an optional stopping point. The reverse crosslinked product can be stored at − 80°C for a few days.

3.11 Pulldown

In this step, the cDNA is separated from the transposition products by pulling down on the biotin that is part of the reverse transcription primer. The supernatant constitutes the transposition products and is processed separately from the cDNA:

1. Prepare 1× B & W-T/RI buffer by mixing the following:
   400× (N + 1) μL 1× B & W-T buffer
   4× (N + 1) μL SUPERase RI
2. Prepare 1× B & W/RI buffer by mixing the following:
   100× (N + 1) μL 1× BW buffer
   2× (N + 1) μL SUPERase RI
3. Prepare 1× STE/RI buffer by mixing the following:
   200× (N + 1) μL 1× STE buffer
   N + 1 μL SUPERase RI
4. In a fresh tube, mix 10× NμL MyOne C1 Dynabeads with 100× NμL 1× B & W-T.
5. Separate on a magnetic rack and remove supernatant.
6. Wash twice with 100× NμL B & W-T without RI.
7. Wash once with 100× NμL B & W-T/RI.
8. Resuspend beads in 100× NμL 2× B & W/RI.

9. Add 100 μL beads to each sample.
10. Incubate at room temperature on a rotator for 60 min.
11. Place the tube on a magnetic rack.
12. Transfer the supernatant (which contains chromatin fragments) to a new tube for ATAC library preparation. The ATAC fragments are stable for a few hours at room temperature and can be processed concurrently or after cDNA library construction is complete.
13. Wash cDNA/RNA-bound beads three times with 100 μL 1× B & W-T/RI.
14. Wash with 100 μL 1× STE/RI without resuspending beads.

3.12 ATAC Library Preparation

In this step, ATAC fragments are purified and amplified into a final library ready for sequencing:

1. Clean up the ATAC part of the sample using Zymo DNA Clean and Concentrate. Elute in 11 μL EB buffer, and then elute again with additional 11 μL EB buffer (a total of 22 μL EB buffer).
2. Prepare ATAC PCR Master Mix by mixing the following:

   225 μL 2× NEBnext Master Mix

   9 μL P7 primer 25 μM

   27 μL H_2O
3. Mix the following:

   ~20 μL sample

   29 μL ATAC PCR Master Mix

   1 μL of 25 μM Adapter 1 Primer (from the PCR Library indexing primers plate)
4. Run PCR for 5 cycles as follows:

   72 °C for 5 min

   98 °C for 30 s

   5 cycles of:

   98 °C for 10 s

   65 °C for 30 s

   72 °C for 30 s
5. Determine additional cycles using qPCR. Add 5 μL of the pre-amplified reaction to 10 μL qPCR Master Mix for a total qPCR reaction of 15 μL as follows:

   5 μL NEBnext Master Mix

   0.2 μL 25 μM Adapter 1.1

   0.2 μL 25 μM P7

0.9 μL 10x SYBR Green

3.7 μL H_2O

6. Assess the amplification profiles and determine the required number of additional cycles to amplify. Please refer to Figure 2 in Buenrostro et al. [25].
7. Carry out final amplification by placing the remaining 45 μL in a thermocycler and running the following program:

   N_{add} cycles of:

   98 °C for 10 s

   65 °C for 30 s

   72 °C for 30 s

   where N_{add} is the number of additional cycles.
8. Clean up the final library using Zymo DNA Clean & Concentrate, eluting in 15 μL.

3.13 RNA Library Preparation Step 1. Template Switching

In this step, RNA library generation is initiated by carrying out template switching on the pulled down cDNA:

1. Prepare the Template switch mix by mixing the following:

   11.25 μL H_2O

   125 μL 50% PEG 6000

   90 μL 5× Maxima RT buffer

   90 μL Ficoll PM-400 (20%)

   45 μL 10 mM dNTPs

   45 μL RNase inhibitor (Lucigen)

   11.25 μL 100 μM TSO oligo

   22.5 μL Maxima RT Rnase H Minus (add last right before reaction)
2. Remove all supernatant. Be careful to avoid drying the beads.
3. Resuspend beads in 50 μL Template switch mix.
4. Incubate samples for 30 min at room temperature with rotation.
5. Incubate samples for 90 min at 42 °C at 300 rpm. Resuspend every 30 min by pipetting up and down.

3.14 RNA Library Preparation Step 2. Amplification of cDNA

The next step is to amplify the individual cDNA molecules.

1. Prepare cDNA PCR Mix by mixing the following:

   247.5 μL KAPA HiFi 2× mix

   7.92 μL 25 μM RNA PCR primer

   7.92 μL 25 μM P7 primer

   231.7 μL H_2O

2. Mix samples with 100 μL H_2O.
3. Separate beads on magnet. Wash with 200 μL STE without resuspending the beads.
4. Mix beads with 55 μL cDNA PCR Mix and transfer to PCR tubes/plates.
5. Run PCR as follows:

   95 °C for 3 min

   5 cycles of:

   98 °C for 20 s

   65 °C for 45 s

   72 °C for 3 min
6. Determine additional cycles using qPCR. Add 2.5 μL of the pre-amplified reaction to 7.5 μL qPCR Master Mix in a total qPCR reaction of 10 μL as follows:

   3.75 μL KAPA HiFi 2× mix

   0.12 μL 25 μM RNA PCR primer

   0.12 μL 25 μM P7 primer

   0.5 μL 20x EVAgreen

   3.01 μL H_2O
7. Determine additional cycles as described above for ATAC libraries.

   5 cycles of:

   98 °C for 20 s

   65 °C for 45 s

   72 °C for 3 min
8. Purify using SPRI beads. Mix the reaction with 0.8× volume of SPRI beads and incubate at room temperature for 10 min. Separate the beads on magnet and wash twice with 200 μL freshly prepared 70% EtOH. Make sure to remove all liquid, and elute in 20 μL.
9. Optional: check size of the cDNA using the D5000 TapeStation.

3.15 RNA Library Preparation Step 3. Tagmentation

The next step is to tagment the amplified cDNA, which will prepare it for the final library amplification step:

1. Quantify cDNA concentration using Qubit.
2. Dilute cDNA to a concentration of 5 ng/μL for tagmentation.

   Note: Expect more than 50 ng cDNA. If cDNA amount is low, it can get away with tagmenting 20 ng cDNA; in this case, adjust the volume of H2O and cDNA accordingly.

3. Prepare tagmentation transposome by mixing the following:
   11.25 μL 1× Tn5
   11.25 μL Dilution Buffer
   22.5 μL annealed tagmentation adapter with glycerol
4. Mix the following:
   10 μL 5 ng/μL cDNA
   10 μL H_2O
   25 μL 2× TD buffer
   5 μL assembled Tn5
5. Incubate for 5 min at 55 °C.
6. Purify tagmented library using the Zymo kit (use 250 μL binding buffer). Elute twice with 11 μL EB (a total of 22 μL).

3.16 RNA Library Preparation Step 4. Final Amplification

Final libraries are generated by PCR.

1. Prepare post-tagmentation PCR mix by mixing the following:
   20 μL sample
   25 μL 2× NEB Next Master Mix
   1 μL 25 μM P7 primer
   1 μL 25 μM Adapter 1 Primer (from the PCR Library indexing primers plate)
   3 μL H_2O
2. Run PCR as follows:
   72 °C for 5 min
   9 cycles of:
   98 °C for 10 s
   65 °C for 30 s
   72 °C for 60 s

3.17 Library Quantification and Evaluation of Library Quality

Before libraries can be sequenced, they need to be properly quantified and be subjected to quality evaluation. This is done by first, evaluation of the insert distribution, and second, quantification:

1. Examination of library size distribution. This step can be carried out using several different instruments, such as a TapeStation or a BioAnalyzer. We prefer to use a TapeStation (with the D1000 or HS D1000 kits) due to flexibility, ease of use, and rapid turnaround time.
2. Quantification of library concentration. For most high-throughput sequencing applications, this step is standardly carried out using a Qubit fluorometer. While this works well for libraries with a unimodal fragment-length distribution, ATAC libraries typically exhibit a multimodal fragment

distribution and also often contain fragments of length higher than what can be sequenced on standard Illumina instruments. As a result, effective library concentrations often differ from apparent library concentrations measured using Qubit, and the optimal way for estimating effective library concentration is qPCR.

3. Estimation of effective library concentration using qPCR. Standard Illumina library quantification kits can be used to quantify the concentration of the library that will be able to be sequenced. Products from NEB or KAPA are appropriate for this use.

3.18 Sequencing

The protocol described here generates libraries designed to be sequenced on Illumina sequencers, the most widely available of which is the NextSeq. On a NextSeq, SHARE-seq libraries are sequenced as follows using a 150-cycle kit:

For the RNA libraries, use a 50 bp × 10 bp × 99 bp × 8 bp configuration (Read 1 × Read 2 × Index1 × Index2, respectively).

For the ATAC libraries, use a 30 bp × 30 bp × 99 bp × 8 bp configuration (Read 1 × Read 2 × Index1 × Index2, respectively).

For RNA, the 10bp of Read 2 captures the UMI, and the 50 bp captures the actual RNA sequence.

For ATAC, fragments are sequenced in a 2 × 30 bp format.

The 8 bp of Index 2 captures the library barcode (if more than one library is sequenced in a single run). The 99 bp of Index 1 captures the pool–split barcodes.

For other Illumina instruments, different configurations can be used. For example, using a 200-cycle kit on NovaSeq, run ATAC libraries in 55 bp × 55 bp × 99 bp × 8 bp configuration and RNA libraries in a 100 bp × 10 bp × 99 bp × 8 bp configuration.

An important consideration to take into account before sequencing is that the standard Illumina run recipes do not allow for the 99-bp index read configuration that is necessary for SHARE-seq libraries. This necessitates the creation of custom recipes in which the limits on the length of the index reads are increased accordingly. However, different methods for creating these custom recipes are necessary depending on the Illumina instrument used and the versions of the control software that the machine is equipped with; resolving this issues may on occasions require seeking help from Illumina's customer support service.

4 Computational Processing

At present there is no standard tool for analyzing pool–split-based multiomics datasets. The pipeline presented here is the one we have been using in our practice. Its objective is to take the raw SHARE-

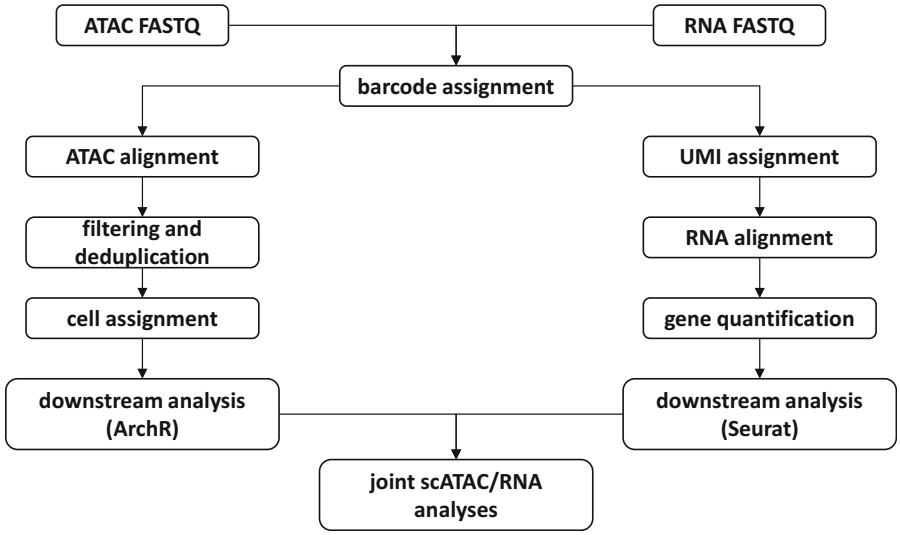

Fig. 4 Outline of the SHARE-seq computational processing procedures. As a first step, cell barcodes are annotated for all reads in both ATAC and RNA FASTQ files. Subsequently, UMIs are consolidated and assigned to reads in the RNA set. RNA reads are then aligned against the genome, and gene expression is quantified in single cells, resulting in a final data matrix that can be analyzed in Seurat (or other scRNA-seq) tools. ATAC reads are aligned against the genome, filtered (removing mitochondria-mapping reads), and deduplicated within each barcode. Alignments are then annotated with their cell barcodes and can be used as input for further analysis in ArchR. Further joint analysis of the ATAC and RNA can be carried out downstream

seq reads and to produce object that can be used for further analysis with established tools for scRNA-seq/scATAC-seq processing such as `Seurat` and `ArchR` (e.g., sparse matrices and BAM files). The outline of the processing is shown in Fig. 4. For both ATAC and RNA, reads are first assigned their cellular barcodes. RNA reads are additionally annotated with the sequenced UMIs. RNA reads are aligned against the genome, a quantification is carried out for each gene in each cell, and a final sparse matrix is created. For ATAC, reads are mapped against the genome, then filtered, and deduplicated within each cell, and a final BAM file with cellular barcodes appended to each alignment is created.

4.1 RNA

1. As a first step in the RNA processing, annotated barcodes for each read pair, using the `SHARE-seq-barcode-annotate.py` script.

```
python SHARE-seq-barcode-annotate.py
            BC1file fieldID pos1 lenBC1 BC2file
            fieldID2 pos2 lenBC2 BC3file fieldID3
            pos3 lenBC3 [-BCedit N] [-revcompBC]
```

The script is flexible and can be used to assign barcodes to almost any kind of pool–split experiment in which the indexes are in Index Read 1. It takes as input files containing the barcodes for each round of pool–split and the column positions of the barcode sequences in each file (0-based), their position in Index Read 1 (0-based), their length, and their orientation (use the `[-revcompBC]` option if the sequences are reverse complement, depending on the exact format of the sequencing). Use the `[-BCedit]` option to increase/decrease the stringency of matching barcode sequences to the master list (the default value is 1). In this case, the barcode files are in the following format:

```
#WellPosition   Name    Sequence
A1      Round1_01       AACGTGAT
B1      Round1_02       AAACATCG
C1      Round1_03       ATGCCTAA
[...]
```

And barcodes are assigned in a single step as follows:

```
python PEFastqToTabDelimited.py RNA.end1.fastq.gz
    RNA.end2.fastq.gz | python SHARE-seq-barcode-annotate.py
    Plate_R1.tsv 2 15 8 Plate_R2.tsv 2 53 8 Plate_R3.tsv
    2 91 8 -revcompBC
    | PEFastqToTabDelimited-reverse.py -
    RNA.barcodes_annotated
```

This will produce FASTQ files with headers looking as follows:

```
@[readID]:::[GTTAGCCT+TAGTCTTG+TACCGAGC] 1:N:0:
TGGGGNCACAGAGCCAAACCATATCAGCTG
+
AAAAA#EEEEEAEEEEEEEEEEEEEEEE
```

In which barcode combinations have been appended to the read headers, with `nan` if no matching barcode was found due to sequencing errors or other issues, e.g.:

```
@[readID]:::[GACGGATT+GATAGAGG+nan] 1:N:0:
ACCAANCTGTGCACAAGCGTGAATCAACCT
+
6AAAA#E/EEEEEEEEEAEEEEEEEEEEE
```

Note that it is considerably faster to split the FASTQ files into smaller pieces and process them in parallel.

2. Compress the output files:

```
gzip RNA.barcodes_annotated.barcodes_annotated.end1.fastq
gzip RNA.barcodes_annotated.barcodes_annotated.end1.fastq
```

3. Annotated UMIs using the `SHARE-seq-RNA-UMI-Add.py` script, which is also flexible and can read UMIs of different lengths in each read in the pair:

```
python SHARE-seq-RNA-UMI-Add.py UMIlen read1|read2
```

As follows:

```
python PEFastqToTabDelimited.py
                    RNA.barcodes_annotated.end1.fastq.gz
                   RNA.barcodes_annotated.end2.fastq.gz |
                 python SHARE-seq-RNA-UMI-Add.py 10 read2 |
                  python PEFastqToTabDelimited-reverse.py -
                  RNA.barcodes_annotated.RNA_UMI
```

This step will append the UMI sequence to the cell barcodes in the read ID:

```
@[readID]:::[TGACCACT+GGTCGTGT+TGCTGATA+TTTATGATAG]
CCTCTNGCTCAGCCTATATACCGCCATCTTCAGCAAACCCTGATGAAGGC
+
AAAAA#EEEEEEEEEEEEEEEEEEEEEEEEEEEE/EEEEEEEEEEEE
```

4. Compress the output files:

```
gzip RNA.barcodes_annotated.RNA_UMI.end1.fastq
gzip RNA.barcodes_annotated.RNA_UMI.end2.fastq
```

5. Merge the individual files:

```
cat RNA_*.barcodes_annotated.RNA_UMI.end1.fastq >
     RNA.barcodes_annotated.RNA_UMI.end1.fastq.gz
cat RNA_*.barcodes_annotated.RNA_UMI.end2.fastq >
     RNA.barcodes_annotated.RNA_UMI.end2.fastq.gz
```

6. Align the Read 1 FASTQ file against the genome using STAR as follows (the commands given here use the standard ENCODE Project Consortium[69] STAR settings):

```
STAR --limitSjdbInsertNsj 10000000 --genomeDir genome/STAR
--outFileNamePrefix RNA.end1.STAR/
--readFilesIn RNA.barcodes_annotated.RNA_UMI.end1.fastq.gz
```

```
--runThreadN 20 --outSAMunmapped Within --outFilterType
BySJout --outSAMattributes NH HI AS NM MD
--outFilterMultimapNmax 50 --outSAMstrandField intronMotif
--outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax
0.04 --alignIntronMin 10 --alignIntronMax 1000000
--alignMatesGapMax 1000000 --alignSJoverhangMin 8
--alignSJDBoverhangMin 1 --sjdbScore 1 --readFilesCommand
zcat --outSAMtype BAM SortedByCoordinate --outWigStrand
Stranded --twopassMode Basic --twopass1readsN -1
--limitBAMsortRAM 500000000000
```

7. Index the output BAM file:

```
samtools index
 RNA.end1.STAR/Aligned.sortedByCoord.out.bam
```

8. Calculate global mapping statistics:

```
python SAMstats.py
    RNA.end1.STAR/Aligned.sortedByCoord.out.bam
    SAMstats-RNA.end1.STAR.hg38
    -bam genome.chrom.sizes samtools
```

   This script will output the number of mapped reads in various categories (uniquely mapping, spliced, etc.) as well as the molecular complexity of the alignment.

9. Calculate read distribution relative to the genome annotation:

```
python SAM_reads_in_genes3_BAM.py annotation.gtf
     RNA.end1.STAR/Aligned.sortedByCoord.out.bam
     genome.chrom.sizes
     sam_reads_genes-RNA.end1.STAR -nomulti
```

   This script will output the fraction of exonic, intronic, and intergenic reads. This is important information for single-cell assays for evaluating to what extent the cytoplasm (which is enriched for exonic reads relative to the nucleus) is captured in the final libraries.

10. Make a RPM-normalized (Reads Per Million mapped reads) global coverage track:

```
python makewigglefromBAM-NH.py title
    RNA.end1.STAR/Aligned.sortedByCoord.out.bam
    genome.chrom.sizes
    RNA.end1.STAR/Aligned.sortedByCoord.out.wig -RPM
```

11. Evaluate read coverage along transcripts:

```
python gene_coverage_wig_gtf.py annotation.gtf
   RNA.end1.STAR/Aligned.sortedByCoord.out.wig
   1000 coverage-RNA -normalize -singlemodelgenes
```

This script run with these settings will output the average read profile over all genes with only a single transcript annotated (in order to avoid confounding by the presence of multiple isoforms) and ≥1000 bp in length. Use a simple annotation with few isoforms, such as refSeq to get as many genes meeting these requirements as possible.

12. Calculate UMI counts per gene and per cell barcode combination using the `SHARE-seq_RNA_counts.py`. For faster processing, run this on each chromosome in parallel, as follows (shown is chr1):

```
python SHARE-seq_RNA_counts.py
   RNA.end1.STAR/Aligned.sortedByCoord.out.bam
   annotation.gtf.chr1 genome.chrom.sizes
   RNA.SHARE-seq_RNA_counts.chr1 -UMIedit 1
```

The `[-UMIedit]` option can be used to tweak the level of UMI collapsing (in this case UMIs within an edit distance of 1 from each other will be collapsed into a single UMI).

13. Calculate per-cell statistics by merging the individual outputs using the `SHARE-seq-RNA-BC-sum-across-files.py` script as follows:

```
python SHARE-seq-RNA-BC-sum-across-files.py
      list_of_per_chromosome_outputs
      RNA.SHARE-seq_RNA_counts.UMIs_per_cell
```

This will output a file in the following format:

```
#BC1+BC2+BC3            rank3 UMIs3 Aligned Positions genes
GCCAATGT+CAGATCTG+TAACGCTG 1  64660 171969          8369
GTTGTCGG+TAAGCGTT+GATCAGCG 2  47079 123008          7864
TGACCACT+GGTCGTGT+TGCTGATA 3  45034 109960          7652
```

which shows the number of UMIs and the number of detected genes for each cell barcode combination.

14. Extract cell barcode combinations above a desired threshold, e.g., ≥500 UMIs into a separate file.

15. Create final sparse matrix format files that can be used as input to Seurat for further analysis with the `SHARE-seq-RNA-UMIs-sum-across-files.py` script:

```
python SHARE-seq-RNA-UMIs-sum-across-files.py
   list_of_per_chromosome_outputs
   RNA.SHARE-seq_RNA_counts.UMIs_per_cell.min500 0
   RNA.SHARE-seq_RNA_counts.UMIs_per_cell.min500.sparse
  -sparse
```

4.2 ATAC

The first steps of the ATAC processing are analogous to those of the RNA pipeline:

1. First, annotate cellular barcodes:

```
python PEFastqToTabDelimited.py
  ATAC.end1.fastq.gz ATAC.end2.fastq.gz |
  python SHARE-seq-barcode-annotate.py
  Plate_R1.tsv 2 15 8 Plate_R2.tsv 2 53 8 Plate_R3.tsv 2
  91 8 -revcompBC |
  PEFastqToTabDelimited-reverse.py -
  ATAC.barcodes_annotated
```

   Note as before that it is considerably faster to split the FASTQ files into smaller pieces and process them in parallel.

2. Compress the output files:

```
gzip ATAC.barcodes_annotated.end1.fastq
gzip ATAC.barcodes_annotated.end1.fastq
```

3. Merge the individual files:

```
cat ATAC_*.barcodes_annotated.end1.fastq >
     ATAC.barcodes_annotated.end1.fastq.gz
cat ATAC_*.barcodes_annotated.end2.fastq >
     ATAC.barcodes_annotated.end2.fastq.gz
```

4. Align reads against the mitochondrial genome with `Bowtie` as follows:

```
python PEFastqToTabDelimited.py
  ATAC.barcodes_annotated.end1.fastq.gz
  ATAC.barcodes_annotated.end2.fastq.gz -trim 30 30 |
  bowtie bowtie-indexes/chrM -p 20 -v 2 -a -t --best
  --strata -q -X 1000 --sam --12 - |
  samtools view -F4 -bT genome.fa - |
  samtools sort - ATAC.2x30mers.chrM
```

This step is for the purpose of evaluating the extent of mitochondrial contamination in the overall library.

5. Align reads against the full genome with `Bowtie` and filter out mitochondrial reads as follows:

```
python PEFastqToTabDelimited.py
 ATAC.barcodes_annotated.end1.fastq.gz
 ATAC.barcodes_annotated.end2.fastq.gz
 -trim 30 30 | bowtie bowtie-indexes/genome
 -p 20 -v 2 -k 2 -m 1 -t --best --strata -q
 -X 1000 --sam --12 - | egrep -v chrM |
 samtools view -F4 -bT genome.fa - | samtools sort -
 ATAC.2x30mers.unique.nochrM
```

Adjust accordingly if working a genome in which the mitochondrial chromosome/contigs are named differently or there are multiple contigs to be filtered out (e.g., in plants where there is also a plastid in addition to the mitochondrion).

6. Index the resulting BAM files.

```
samtools index ATAC.2x30mers.unique.nochrM.bam
samtools index ATAC.2x30mers.chrM.bam
```

7. Calculate mapping statistics for the two sets of alignments.

```
python SAMstats.py ATAC.2x30mers.chrM.bam
    SAMstats-ATAC.2x30mers.chrM
   -bam genome.chrom.sizes samtools
   -paired -noNHinfo
python SAMstats.py ATAC.2x30mers.unique.nochrM.bam
    SAMstats-ATAC.2x30mers.unique.nochrM
    -bam genome.chrom.sizes samtools
    -paired -uniqueBAM
```

8. Calculate the mitochondrial reads fraction *MRF* as follows:

$$MRF = \frac{|R_M|}{|R_M| + |R_N|} \tag{1}$$

where R_M is the total number of reads that map to the mitochondrial genome and R_N is the number of reads that map to the nuclear genome after filtering out mito-mapping reads.

9. Evaluate the fragment size distribution over the nuclear genome:

```
python PEInsertDistFromBAM.py
    ATAC.2x30mers.unique.nochrM.bam
    genome.chrom.sizes
    ATAC.2x30mers.unique.nochrM.InsLen
    -uniqueBAM -normalize
```

10. Create a normalized genome coverage track:

```
python makewigglefromBAM-NH.py title
   ATAC.2x30mers.unique.nochrM.bam
   genome.chrom.sizes ATAC.2x30mers.unique.nochrM.wig
   -notitle -RPM -uniqueBAM
```

11. Create a BigWig file using the `wigToBigWig` program from the UCSC Genome Browser utilities suite.

```
wigToBigWig ATAC.2x30mers.unique.nochrM.wig
    genome.chrom.sizes
    ATAC.2x30mers.unique.nochrM.bigWig
```

12. Calculate the global TSS enrichment. The TSS enrichment TSS_E is the most informative ATAC-seq and is based on generating an average read distribution profile around annotated transcription start sites for protein coding genes and then calculating the ratio between the number of reads in the immediate neighborhood of the TSS and the number of reads falling in the regions on the flanks of the TSS peak. The advantage of the TSS_E metric is that it is an internal to the dataset measure independent of peak calling. We use a TSS window of ±100 bp and a TSS flank distance of 2000 bp, i.e., TSS_E is calculated as follows:

$$TSS_E = \frac{|R \in [TSS \pm 100]|}{|R \in [TSS - 2050, TSS - 1950]| + |R \in [TSS + 1950, TSS + 2050]|} \quad (2)$$

First, generate the TSS metaprofile:

```
python signalAroundCoordinate-BW.py
   annotation.TSS-0bp.bed 0 1 3 4000
   ATAC.2x30mers.unique.nochrM.bigWig
   ATAC.2x30mers.unique.nochrM.TSS_profile -normalize
```

Note that you need a `BED` file containing the start positions and the strands of annotated TSSs in the genome, e.g.,

```
#chr TSS  TSS strand geneName
chr1 1000  1000  +       GENE1
```

Second, calculate the TSS score:

```
python ATACTSSscore.py
   ATAC.2x30mers.unique.nochrM.TSS_profile
   100 2000 >> ATACTSSscore.txt
```

13. Deduplicate the BAM file. Note that this step is different from the typical deduplication carried out in most high-throughput sequencing pipelines, based on tools such as `MarkDups` in `picard`. Here, we perform deduplication of fragments only within the same cell barcode, i.e., for two fragments to be collapsed, they need to have the same coordinates, orientation, and cell barcode.

```
python SHARE-seq_ATAC_dedup.py
   ATAC.2x30mers.unique.nochrM.bam
   genome.chrom.sizes
   ATAC.2x30mers.unique.nochrM.BC_dedup.bam
   -addBC
```

Use the `[-addBC]` to append the cell barcodes to each alignment as a `BC` tag, making these final files ready to use with `ArchR`.

14. Index the deduplicated BAM file:

```
samtools index ATAC.2x30mers.unique.nochrM.BC_dedup.bam
```

15. Calculate alignment stats for the deduplicated BAM file:

```
python SAMstats.py ATAC.2x30mers.unique.nochrM.BC_dedup.bam
       SAMstats-ATAC.2x30mers.unique.nochrM.BC_dedup
       -bam genome.chrom.sizes samtools -paired -uniqueBAM
```

16. Calculate fragment count and TSS enrichment statistics for each cell barcode.

```
python SHARE-seq_ATAC_stats_per_cell.py
  ATAC.2x30mers.unique.nochrM.BC_dedup.bam
  genome.chrom.sizes annotation.TSS-0bp.bed 0 1 2000 200
  ATAC.2x30mers.unique.nochrM.BC_dedup.per_cell_stats
```

This script will output a file containing information about the number of fragments and TSS enrichment for each barcode that can be used to filter barcodes for downstream analysis.

More sophisticated filtering, in addition to these simple metrics, i.e., of doublet cells, can be performed in `ArchR` [67].

5 Expected Results

5.1 Sequencing Libraries

Figure 5 shows the typical fragment profiles for ATAC and RNA SHARE-seq libraries. ATAC libraries are expected to show a nucleosomal signature, with a prominent subnucleosomal, mononucleosomal, and perhaps dinucleosomal peaks, shifted to the right by the length of the adapters and barcodes added to the original fragments. In contrast, RNA libraries are primarily unimodal in length.

5.2 Species Mixing Experiments

A customary experiment to be carried out when testing, adopting, or developing any new single-cell protocol is the species mixing experiment, in which cells from two different species, usually mouse and human, are mixed together, and the extent of crosstalk/contamination of individual barcodes or of doublet formation (in which two cells are processed together with the same barcode)

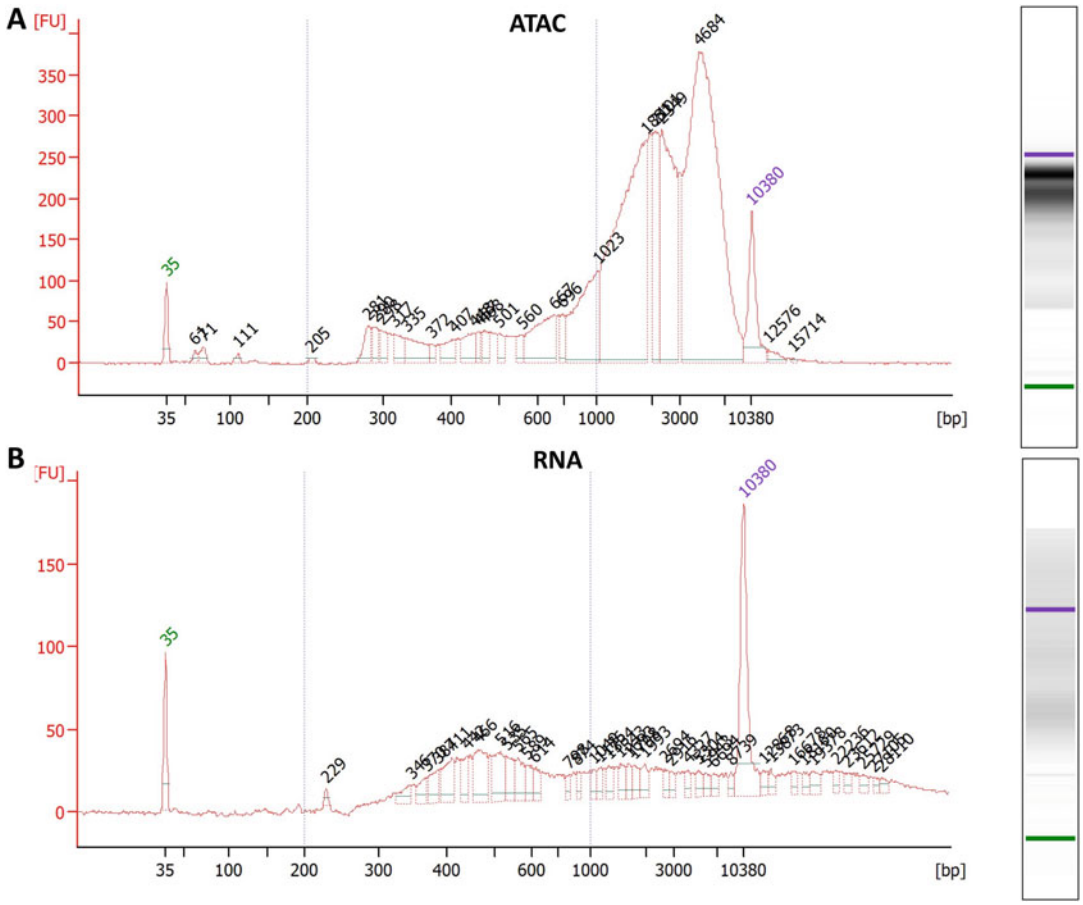

Fig. 5 Typical fragment-length profiles of SHARE-seq libraries. (**a**) BioAnalyzer profile of a SHARE-seq ATAC library. (**b**) BioAnalyzer profile of a SHARE-seq RNA library

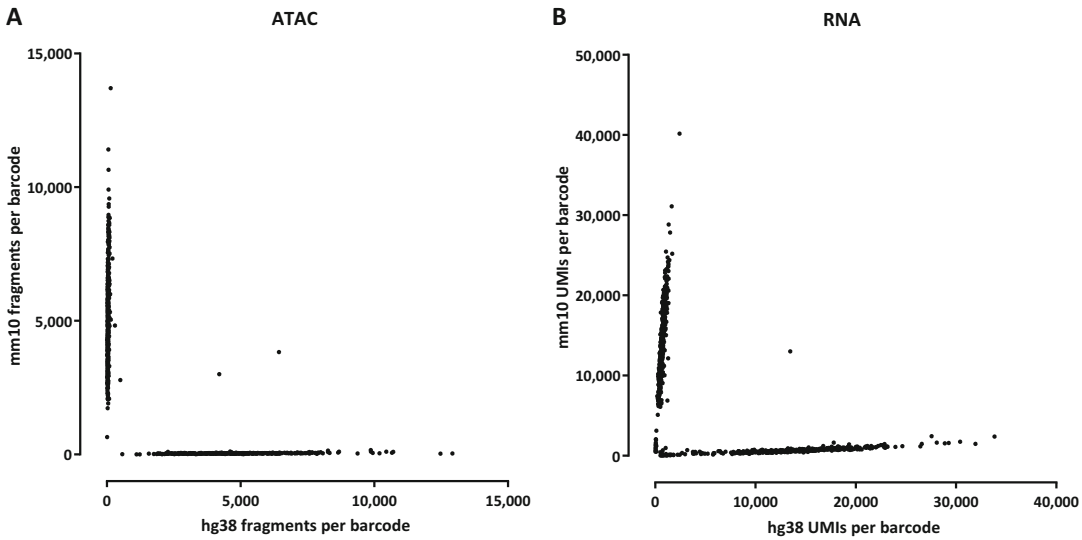

Fig. 6 Typical results from a species mixing SHARE-seq experiment. Human HEK293 and mouse embryonic fibroblast (MEF) cells were mixed in equal proportions and carried through the SHARE-seq workflow. (**a**) ATAC fragments per cell. (**b**) RNA UMIs per cell

is assessed based on how many reads in each barcode map to each species. Ideally, all barcodes should feature reads coming from only one of the two species. Doublet arise from loading of multiple cells in the same droplets/wells (depending on the method used) or from physical clumping of cells early in the protocol that then are processed together throughout the rest of the procedure.

Figure 6 shows typical species mixing results for a SHARE-seq experiment. We note that in our hands ATAC experiments usually show virtually no crosstalk between barcodes and very few doublets. On the other hand, pool–split RNA experiments in general often exhibit a small fraction of reads resulting from "leakage," likely because of some cells opening up during cell handling and releasing their content into the general reaction pool. This issue does not significantly affect most analyses, but it should be kept in mind in the cases in which it could be a confounding factor.

5.3 ATAC Post-sequencing Quality Evaluation

Figure 7 shows the key ATAC-seq bulk-level metrics. The fragment-length distribution (Fig. 7a) usually shows strong subnucleosomal and nucleosomal peaks as well as a weaker dinucleosomal one. High TSS enrichment is desirable; in this case (Fig. 7b), it is very high ($TSS_E \geq 25$). *See* **Note 8** for more details. Figure 7c shows the fraction of mitochondrial reads in the human and mouse cells in the species mixing experiment. Note that the fraction can vary greatly depending on the properties of the cell type (cancer cell lines and highly metabolically active cells tend to have more mitochondria [70]) and not just on the experimental variation (which in this case is completely minimized as the cells were processed together).

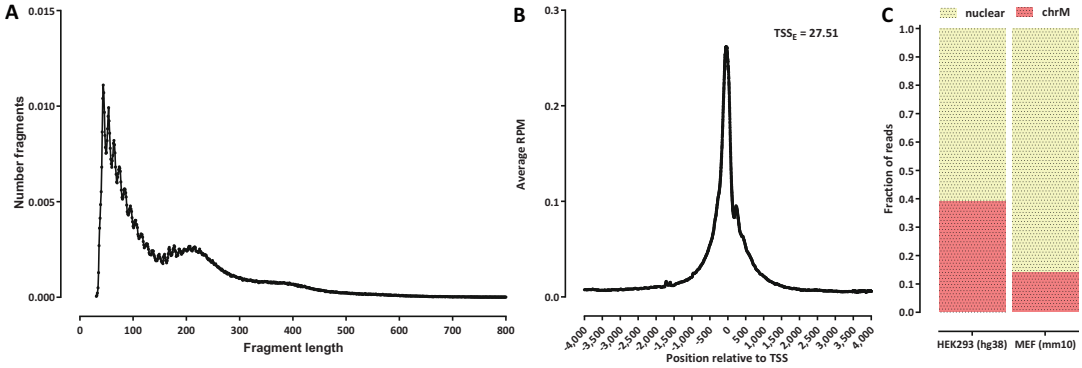

Fig. 7 Basic evaluation of bulk-level ATAC quality and enrichment. (**a**) Fragment-length distribution. (**b**) TSS enrichment. Shown are the same experiments as those featured in Fig. 6. (**c**) Mitochondrial read fraction for each species in this experiment

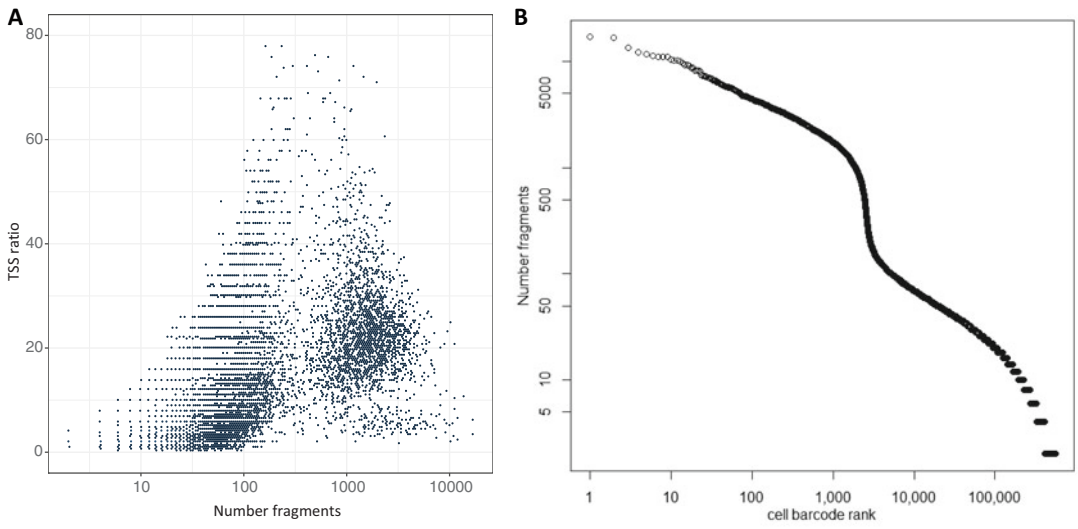

Fig. 8 Basic evaluation of scATAC-seq-level quality and enrichment. (**a**) Number fragments per cell barcode vs. TSS enrichment. (**b**) Cell barcode rank (by fragment counts) vs. fragment counts per cell barcode

Figure 8 shows the key scATAC metrics. One such metric is the relationship between the number of fragments per cell barcode and the TSS enrichment within each cell barcode (Fig. 8a). Another is the curve of the number of fragments per cell barcode plotted against the rank (by the number of fragments per cell barcode) of the cell barcodes (Fig. 8b). Ideally, there should be a clear inflection point between the cell barcodes with high fragment counts and the cell barcodes with low fragment counts, indicating that a set of high-quality cells have been captured and preserved intact through the full pool–split procedure. A flatter, diagonal-like shape of that curve can be indicative of loss of cell integrity during handling and is potentially concerning regarding the biological interpretability of the experiment if the lack of inflection is too extreme.

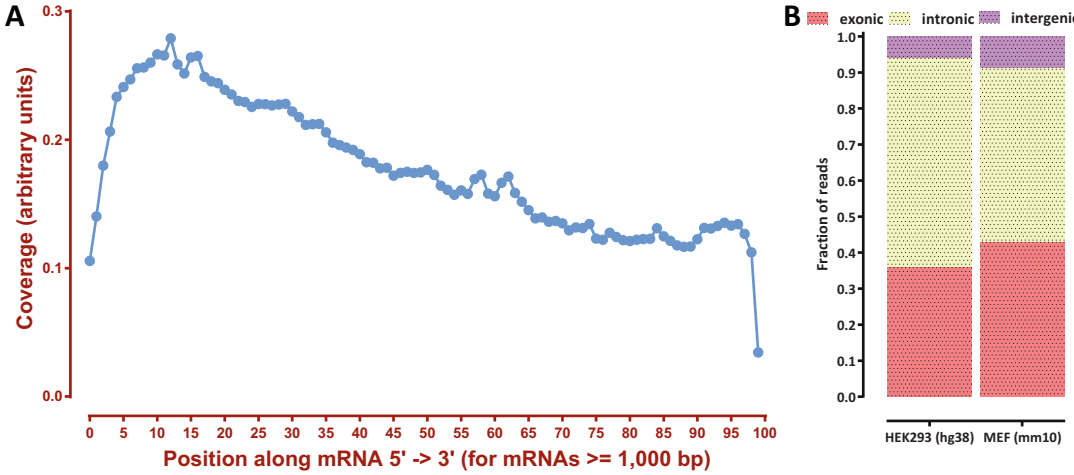

Fig. 9 Basic evaluation of the bulk-level RNA-seq properties. (**a**) Read distribution along transcript lengths. (**b**) Read distribution relative to the exonic, intronic, and intergenic genomic spaces

5.4 RNA Post-sequencing Quality Evaluation

Figure 9 shows the typical parameters to be evaluated for a bulk-level RNA-seq dataset. One is the distribution of reads along transcripts (Fig. 9a). SHARE-seq is not a 3'-tagging experiment the way some scRNA-seq approaches are as it attaches UMIs to the 3' end of transcripts, but cDNAs are tagmented at random after cDNA amplification; thus the first reads of the RNA part of a SHARE-seq dataset can be some distance away from the 3' end.

Another is the distribution of reads relative to the annotation (Fig. 9b). As is often observed in scRNA-seq datasets, SHARE-seq RNA libraries contain a significant portion of reads originating from introns, presumably from unspliced transcripts present in the nucleus. This is likely due to the fact that the ATAC reaction has to happen first in the workflow, and thus a substantial portion of the cytoplasm is lost and the final libraries are enriched for nuclear material.

Figure 10 shows the key metric for evaluating the success of the RNA portion of a SHARE-seq experiment. As with ATAC above, the curve of the number of UMIs per cell barcode plotted against the rank (by the number of UMIs per cell barcode) of the cell barcodes should ideally feature a clear inflection point between the cell barcodes with high UMI counts and the cell barcodes with low UMI counts (Fig. 10a). There should also be a concordance between the cell barcodes with high ATAC fragment counts and those with high UMI counts, i.e., the same cells are of high quality in both modalities, and are thus usable for joint analysis (Fig. 10b).

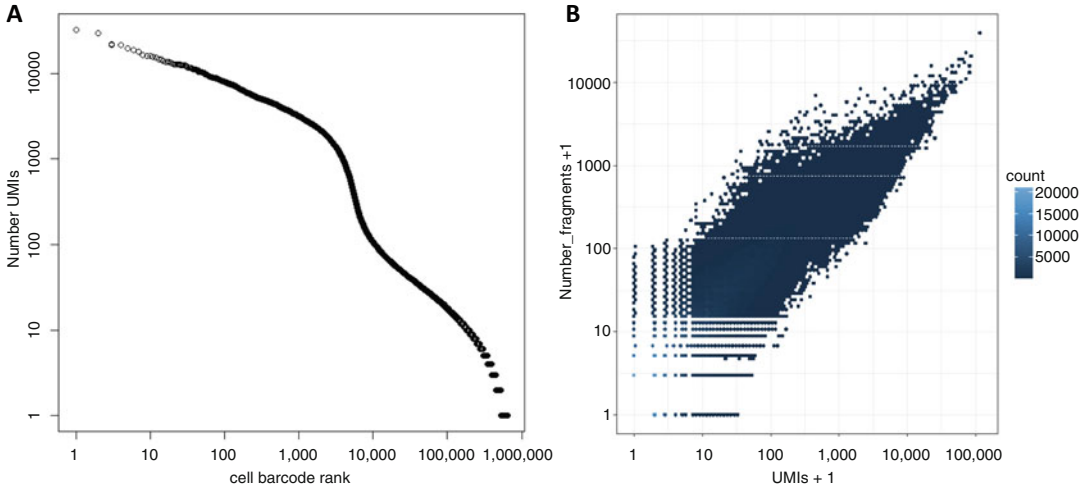

Fig. 10 Basic evaluation of SHARE-seq RNA single-cell-level quality and enrichment. (**a**) Cell barcode rank (by UMI counts) vs. UMI counts per cell barcode. (**b**) UMI counts per barcode vs. ATAC fragment counts per barcode.

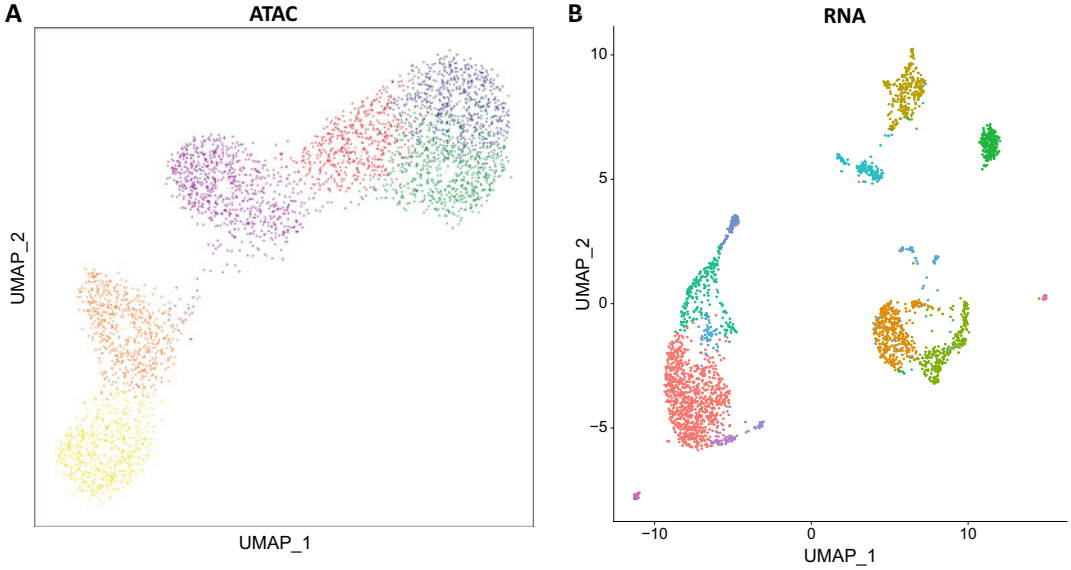

Fig. 11 Example SHARE-seq output on human embryonic lung samples. (**a**) ArchR iterative LSI UMAP on the ATAC-seq dataset. (**b**) Seurat UMAP on the RNA dataset. Individual ArchR- and Seurat-defined clusters are colored separately

5.5 Dimensionality Reduction and Cell Type/Cluster Identification

Following initial data processing, clusters and cell types can be identified using standard tools for that purpose such as `Seurat` [68] and/or `ArchR` [67]. Figure 11 shows typical such output in UMAP space for both the ATAC and RNA sides of a SHARE-seq experiment from a human embryonic lung tissue sample.

6 Notes

1. The details of the production of hyperactive transposition are beyond the scope of this chapter. However, detailed instructions for how to carry it out can be found in Picelli et al. 2014 [71].
2. In this chapter, we presented one of many available protocols for tissue dissociation and nuclei isolation that has worked in our hands in some contexts. However, the variety of tissues and their properties that can be encountered in different organisms is vast, making it practically impossible to have one common such protocol for all situations. Thus novel optimal procedures for tissue dissociation often have to be empirically devised or adapted.
3. The protocol we described here used light 0.1% FA crosslinking. This does not mean that optimal results will be obtained in all contexts with the same conditions, and crosslinking may have to be optimized depending on the specifics of the experimental system being studied.
4. The protocol described here is for a 96 × 96 × 96 indexing. However, it can be expanded to more cycles and/or more barcodes, e.g., to a 3-round 384 × 384 × 384 indexing, or 4-round or 5-round 96/384 × 96/384 × 96/384. Pick the optimal design based on the availability of robotic liquids handlers (it is generally not practical to carry out pipetting of 384-well plates by hand), the desired throughput, and other considerations. Note that additional barcodes and linker would have to be designed so that they are compatible with each other and with further rounds of barcoding. Aim for as much distance in sequence space between the 8-bp barcodes (or increase their length, if the sequencing format allows for it). The set of 8-bp barcodes can be identical throughout all rounds of indexing.
5. Low-binding tubes are preferable for all reactions in order to ensure maximum yields.
6. It is optimal in terms of effort to anneal a sufficient amount of oligos for multiple experiments on many separate plates. These can then be used immediately when cells/tissues become available, saving a considerable amount of experimental time.
7. The TB buffer described here is modified from the original omniATAC protocol with the addition of acetate. In our experience, this provides superior results compared to the traditional buffer formulation.

8. In our (and not only ours) experience, experiments in cell lines always produce much higher quality ATAC datasets than those obtained from tissues, especially frozen tissues. This is not limited to SHARE-seq but is what has been observed by numerous previous studies mapping chromatin accessibility in tissue samples in contexts such as cancer, development, and adult tissues [27, 28, 72, 73]. This is likely due to the extensive handling and freezing and thawing of tissues leading to the breaking up of nuclei and the release of unprotected free DNA that is tagmented by Tn5, increasing the background fragments and decreasing the signal to noise. Whether future protocol optimizations can resolve these issues or they are fundamentally insurmountable is not known at present.

Acknowledgements

The authors thank Sai Ma and Jason Buenrostro for helpful discussion regarding the SHARE-seq protocol. This work was supported by NIH grants (P50HG007735, RO1 HG008140, U19AI057266 and UM1HG009442 to W.J.G., 1UM1HG009436 to W.J.G. and A.K., 1DP2OD022870-01 and 1U01HG009431 to A.K., and HG006827 to C.H.), the Rita Allen Foundation (to W.J.G.), the Baxter Foundation Faculty Scholar Grant, and the Human Frontiers Science Program grant RGY006S (to W.J.G). W.J.G is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative. S.K. is supported by MSTP training grant T32GM007365 and the Paul and Daisy Soros Fellowship. Fellowship support also provided by the Stanford School of Medicine Dean's Fellowship (G.K.M.), by the EMBO Long-Term Fellowship EMBO ALTF 1119-2016, and by the Human Frontier Science Program Long-Term Fellowship HFSP LT 000835/2017-L (Z.S.).

References

1. Mortazavi A, Williams BA, McCue K et al. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628
2. Nagalakshmi U, Wang Z, Waern K et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320(5881):1344–1349
3. Sultan M, Schulz MH, Richard H et al. (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321(5891):956–960.
4. Wilhelm BT, Marguerat S, Watt S et al. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453(7199):1239–1243.
5. Tang F, Barbacioru C, Wang Y et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6(5):377–382.
6. Islam S, Kjällquist U, Moliner A et al. (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res 21(7):1160–1167.
7. Ramsköld D, Luo S, Wang YC et al. (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30(8):777–782

8. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2(3): 666–673.
9. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H, Regev A (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498(7453):236–240.
10. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343(6172):776–779
11. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161(5): 1187–1201
12. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161(5): 1202–1214
13. Zheng GX, Terry JM, Belgrader P et al. (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049
14. Han X, Wang R, Zhou Y et al. (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 172(5):1091–1107.e17
15. Cao J, Packer JS, Ramani V et al. (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357:661–667
16. Rosenberg AB, Roco CM, Muscat RA et al. (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360:176–182
17. McGhee JD, Wood WI, Dolan M et al. (1981) A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. Cell 27:45–55
18. Keene MA, Corces V, Lowenhaupt K et al. (1981) DNase I hypersensitive sites in Drosophila chromatin occur at the 5′ ends of regions of transcription. Proc Natl Acad Sci U S A 78:143–146
19. Wu C (1980) The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. Nature 286(5776):854–860
20. Minnoye L, Marinov GK, Krausgruber T et al. (2021) Chromatin accessibility profiling methods. Nat Rev Meth Primers 1:10.
21. Buenrostro JD, Giresi PG, Zaba LC et al. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218
22. Corces MR, Trevino AE, Hamilton EG et al. (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959–962
23. Reznikoff WS (2008) Transposon Tn5. Annu Rev Genet 42:269–286
24. Adey A, Morrison HG, Asan et al. (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11(12): R119
25. Buenrostro JD, Wu B, Litzenburger UM et al. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523:486–490
26. Cusanovich DA, Daza R, Adey A et al. (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348:910–914
27. Cusanovich DA, Reddington JP, Garfield DA et al. (2018) The *cis*-regulatory dynamics of embryonic development at single-cell resolution. Nature 555:538–542
28. Preissl S, Fang R, Huang H et al. (2018) Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci 21(3):432–439
29. Mezger A, Klemm S, Mann I et al. (2018) High-throughput chromatin accessibility profiling at single-cell resolution. Nat Commun 9(1):3647
30. Satpathy AT, Granja JM, Yost KE et al. (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 37:925–936
31. Lareau CA, Duarte FM, Chew JG et al. (2019) Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol 37:916–924
32. Macaulay IC, Haerty W, Kumar P, et al. 2015. G & T-seq: parallel sequencing of single-cell

genomes and transcriptomes. Nat Methods 12(6):519–522

33. Huang AY, Li P, Rodin RE et al. (2020) Parallel RNA and DNA analysis after deep sequencing (PRDD-seq) reveals cell type-specific lineage patterns in human brain. Proc Natl Acad Sci U S A 117(25):13886–13895
34. Zachariadis V, Cheng H, Andrews N, Enge M (2020) A highly scalable method for joint whole-genome sequencing and gene-expression profiling of single cells. Mol Cell 80(3):541–553.e5
35. Yin Y, Jiang Y, Lam KG et al. (2019) High-throughput single-cell sequencing with linear amplification. Mol Cell 76(4):676–690.e10
36. Rodriguez-Meira A, Buck G, Clark SA et al. (2019) Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel rna sequencing. Mol Cell 73(6):1292–1305.e8
37. Hou Y, Guo H, Cao C et al. (2016) Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res 26(3): 304–319
38. Hu Y, Huang K, An Q, Du G, Hu G, Xue J, Zhu X, Wang CY, Xue Z, Fan G (2016) Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol 17:88
39. Angermueller C, Clark SJ, Lee HJ et al. (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 13(3):229–232
40. Pott S (2017) Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. Elife 6: e23203
41. Peterson VM, Zhang KX, Kumar N et al. (2017) Multiplexed quantification of proteins and transcripts in single cells. Nat Biotechnol 35(10):936–939
42. Stoeckius M, Hafemeister C, Stephenson W et al. (2017) Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 14(9):865–868
43. O'Huallachain M, Bava FA et al. (2020) Ultra-high throughput single-cell analysis of proteins and RNAs by split-pool synthesis. Commun Biol 3(1):213
44. Chung H, Parkhurst CN, Magee EM et al. (2021) Simultaneous single cell measurements of intranuclear proteins and gene expression. https://doi.org/10.1101/2021.01.18.427139
45. Katzenelenbogen Y, Sheban F, Yalin A et al. (2020) Coupled scRNA-seq and intracellular protein activity reveal an immunosuppressive role of TREM2 in cancer. Cell 182(4): 872–885.e19
46. Guo F, Li L, Li J et al. (2017) Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell Res 27(8):967–988
47. Clark SJ, Argelaguet R, Kapourani CA et al. (2018) scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun 9(1): 781
48. Wang Y, Yuan P, Yan Z et al. (2021) Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. Nat Commun 12(1):1247
49. Luo C, Liu H, Xie F et al. (2019) Single nucleus multi-omics links human cortical cell regulatory genome diversity to disease risk variants. bioRxiv 2019.12.11.873398
50. Xiong H, Luo Y, Wang Q et al. (2021) Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. Nat Methods 18(6):652–660
51. Zhu C, Zhang Y, Li YE et al. (2021) Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods 18(3):283–292
52. Markodimitraki CM, Rang FJ, Rooijers K et al. (2020) Simultaneous quantification of protein-DNA interactions and transcriptomes in single cells with scDam & T-seq. Nat Protoc 15(6): 1922–1953
53. Fiskin E, Lareau CA, Eraslan G et al. (2020) Single-cell multimodal profiling of proteins and chromatin accessibility using PHAGE-ATAC. bioRxiv 2020.10.01.322420
54. Mimitou EP, Lareau CA, Chen KY et al. (2021) Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat Biotechnol. https://doi.org/10.1038/s41587-021-00927-2
55. Swanson E, Lord C, Reading J et al. (2021) Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. eLife 10:e63632
56. Kearney CJ, Vervoort SJ, Ramsbottom KM et al. (2021) SUGAR-seq enables simultaneous detection of glycans, epitopes, and the transcriptome in single cells. Sci Adv 7(8): eabe3610
57. Cao J, Cusanovich DA, Ramani V et al. (2018) Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 361:1380–1385
58. Zhu C, Yu M, Huang H et al. (2019) An ultra high-throughput method for single-cell joint

analysis of open chromatin and transcriptome. Nat Struct Mol Biol 26:1063–1070

59. Xing QR, Farran CAE, Zeng YY et al. (2020) Parallel bimodal single-cell sequencing of transcriptome and chromatin accessibility. Genome Res 30(7):1027–1039
60. Chen S, Lake BB, Zhang K (2019) High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol 37(12):1452–1457
61. Ma S, Zhang B, LaFave LM et al. (2020) Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell 183: 1103–1116.e20
62. Langmead B, Trapnell C, Pop M et al. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
63. Li H, Handsaker B, Wysoker A et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
64. Kuhn RM, Haussler D, Kent WJ (2013) The UCSC Genome Browser and associated tools. Brief Bioinform 14:144–161
65. Kent WJ, Zweig AS, Barber G et al. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26: 2204–2207
66. Dobin A, Davis CA, Schlesinger F et al. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21.
67. Granja JM, Corces MR, Pierce SE et al. (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 53(3):403–411
68. Hao Y, Hao S, Andersen-Nissen E et al. (2021) Integrated analysis of multimodal single-cell data. Cell 184(13):3573–3587.e29
69. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
70. Marinov GK, Wang YE, Chan DC, Wold BJ (2014) Evidence for site-specific occupancy of the mitochondrial genome by nuclear transcription factors. PLoS ONE 9(1):e84713. link
71. Picelli S, Björklund AK, Reinius B et al. (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24:2033–2040
72. Domcke S, Hill AJ, Daza RM et al. (2020) A human cell atlas of fetal chromatin accessibility. Science 370(6518):eaba7612
73. Corces MR, Granja JM, Shams S et al. (2018) The chromatin accessibility landscape of primary human cancers. Science 362(6413): eaav1898