This is the pre-peer reviewed version of the following article: Chen, W., Gardeux, V., Meireles-Filho, A. & Deplancke, B. Profiling of Single-Cell Transcriptomes. *Current Protocols in Mouse Biology* 7, 145–175 (2017), which has been published in final form at <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpmo.30>

Profiling of Single cell transcriptome

Wanze Chen¹, Vincent Gardeux¹, Antonio Meireles-Filho¹, Bart Deplancke^{1,2}

¹Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland

²For correspondence: bart.deplancke@epfl.ch

ABSTRACT

Complex biological systems are composed of multiple cell types whose transcriptional activity can vary due to differences in cell state, environmental stimulation or intrinsic programs. Using conventional bulk analysis capture the average transcriptional programs of the cell population, thus missing the unique cellular signature of each single cell. During the last seven years, the development of single-cell RNA-sequencing (scRNA-seq) technologies provided a powerful approach to dissect the cellular heterogeneity of complex biological systems. However, such approaches required specialized equipment or were considerably costly. In this protocol, we describe a Smart-seq2 based method to profile the transcriptome of hundreds of single cells simultaneously, without utilizing commercial kits or requiring any specialized single-cell capture/library preparation tools. Moreover, we introduce a data analysis pipeline, named Automated Single-cell Analysis Pipeline (ASAP), which allows researchers without strong computational expertise to explore scRNA-seq data using a wide range of commonly used algorithms and sophisticated visualization tools.

Keywords: single-cell, transcriptomics, scRNA-seq, Smart-seq2, gene expression, Tn5, tagmentation, ASAP, web-tool, analysis pipeline

INTRODUCTION

Conventional bulk RNA sequencing (RNA-seq) profiles the transcriptome of thousands to millions of cells. However, it only provides a population gene expression profile, neglecting cell to cell differences in the population. Recently, due to unprecedented advancements in library preparation methods, next-generation sequencing technologies and bioinformatics, it is now possible to explore the transcriptome landscape of every single cell in an organism, allowing the identification of new cell types and states, including those that are rare.

Here, Basic protocol 1 describes an improved version of the Smart-seq2 protocol, which profiles single-cell transcriptomes without using either whole transcriptome amplification. In addition, this protocol uses an in-house prepared Tn5 transposase (see below) that evict the use of expensive library preparation kits, which are especially costly when processing hundreds or thousands of cells.

Support protocol 1 introduces the generation of in-house Tn5 transposase, which is necessary for Basic protocol 1 and orders of magnitude cheaper than the commercial alternative (Nextera XT DNA library prep kit).

Alternate protocol 1 is a replacement of Basic protocol 1, for users who prefer using the commercial transposase Nextera XT DNA library prep kit.

Basic protocol 2 provides a straightforward bioinformatics pipeline to analyze the resulting single-cell RNA-seq data, specifically designed for users without a strong computational data analysis background.

Of note, the protocol is scaled for a 96-well PCR plate of cells. It can be easily downscaled to several single cells or upscaled to hundreds and thousands of cells. Although not required, automatic liquid handling platforms can be incorporated to increase the scale and throughput.

BASIC PROTOCOL 1: MODIFIED SMART-SEQ2 SINGLE-CELL RNA-seq

The SMART template switching technology utilizes the template-switching activity of the Moloney murine leukemia virus (MMLV) reverse transcriptase to convert mRNA to full-length cDNA (Figure 1) (Zhu et al., 2001). Briefly, an oligo-dT with an anchor sequence in the 5'-end is used to initiate the reverse transcription when paired with polyA sequences from mRNA. MMLV RTase is able to add additional nucleotides (mostly -CCC) to the 3'-end of the nascent cDNA upon reaching the 5'-end of the RNA template. When an oligo with -rGrGrG ribonucleotides on its 3'-end is provided, this oligo (termed as template-switching oligo, or TSO) pairs with the 3'-end nucleotides (-CCC) of the cDNA. Subsequently, the reverse transcriptase switches to this TSO and continues the reverse transcription. In this way, the 5' anchor sequence in oligo-dT and the TSO sequence are incorporated into the 5'- and 3'-end of the first strand cDNA, respectively, which is subsequently amplified by PCR. The amplified DNA is subjected to Tn5 transposase mediated tagmentation, a process where the DNA is simultaneously fragmented and tagged with PCR adaptors (Adey et al., 2010).

By combining the SMART technology, Tn5 tagmentation and next generation sequencing, Smart-seq was developed and applied to single cell mRNA sequencing (Ramskold et al., 2012). The updated version of the method, Smart-seq2, further refined reverse transcription (RT), template switching and preamplification conditions to improve the sensitivity and reduce variability (Picelli et al., 2013).

The protocol described here is a further improvement of Smart-seq2 with several modifications. The main ones are:

- 1) Blocking of the 5'-end of the TSO and oligo-dT with a biotin modification. The reason of this choice is that the template-switching can occur multiple times, which leads to TSO concatemers containing multiple TSO sequence in the cDNA, such as TSO-TSO_(n). The 5'-biotinylated TSO can almost completely prevent the concatemers. The oligo-dT can function as a TSO, although at very low efficiency.
- 2) Replacement of the Nextera XT DNA library prep kit with an in-house prepared Tn5 transposase, to drastically reduce the library preparation costs.
- 3) Optimization of the reverse transcriptase and its concentration in the reaction. The MMLV reverse transcriptase from different brands are differentially engineered to improve the reverse transcription. However, among the six reverse transcriptases tested in our hands, only two of them successfully generated high quality single-cell cDNA, potentially due to differential template-switching activities. While too little enzyme reduces the reverse transcription efficiency, too much enzyme leads to high rRNA sequence reads in the sequenced library.
- 4) Application of stricter purification conditions to reduce the primer dimer carryover.
- 5) Optimization of library normalization. We sub-pooled several libraries before the final library purification and normalization, which simplifies the protocol and reduces reagent consumption.

Figure 1. Flowchart of the optimized Smart-seq2 procedure.

Materials

Oligos (from IDT): Oligos are dissolved in nuclease free water and stored in 100 μ M aliquots at -80 $^{\circ}$ C. Avoid repeated freeze-thaw cycles.

HPLC purified oligos:

biotin-oligo-dT30VN:

/5Biosg/AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNN

biotin-ISPCR oligo: /5Biosg/AAGCAGTGGTATCAACGCAGAGT

biotin-TSO: /5Biosg/AAGCAGTGGTATCAACGCAGAGTACATrGrGrG

biotin-TSO-LNA: /5Biosg/AAGCAGTGGTATCAACGCAGAGTACATrGrG+G

biotin-TSO: /5Biosg/AAGCAGTGGTATCAACGCAGAGTACATrGrGrG

Desalted oligos:

Ad1.1: AATGATACGGCGACCACCGAGATCTACAC**TAGATCGC**TCGTCGGCAGCGTCAGATGTG
Ad1.2: AATGATACGGCGACCACCGAGATCTACAC**CTCTCTAT**TCGTCGGCAGCGTCAGATGTG
Ad1.3: AATGATACGGCGACCACCGAGATCTACAC**TATCCTCT**TCGTCGGCAGCGTCAGATGTG
Ad1.4: AATGATACGGCGACCACCGAGATCTACAC**AGAGTAGAT**TCGTCGGCAGCGTCAGATGTG
Ad1.5: AATGATACGGCGACCACCGAGATCTACAC**GTAAGGAG**TCGTCGGCAGCGTCAGATGTG
Ad1.6: AATGATACGGCGACCACCGAGATCTACAC**ACTGCATAT**TCGTCGGCAGCGTCAGATGTG
Ad1.7: AATGATACGGCGACCACCGAGATCTACAC**AAGGAGTAT**TCGTCGGCAGCGTCAGATGTG
Ad1.8: AATGATACGGCGACCACCGAGATCTACAC**CTAAGCCT**TCGTCGGCAGCGTCAGATGTG
Ad1.9: AATGATACGGCGACCACCGAGATCTACAC**CGTCTAAT**TCGTCGGCAGCGTCAGATGTG
Ad1.10: AATGATACGGCGACCACCGAGATCTACAC**TCTCTCCG**TCGTCGGCAGCGTCAGATGTG
Ad1.11: AATGATACGGCGACCACCGAGATCTACAC**TCGACTAG**TCGTCGGCAGCGTCAGATGTG
Ad1.12: AATGATACGGCGACCACCGAGATCTACAC**TTCTAGCT**TCGTCGGCAGCGTCAGATGTG
Ad1.13: AATGATACGGCGACCACCGAGATCTACAC**CCTAGAGT**TCGTCGGCAGCGTCAGATGTG
Ad1.14: AATGATACGGCGACCACCGAGATCTACAC**GCGTAAGAT**TCGTCGGCAGCGTCAGATGTG
Ad1.15: AATGATACGGCGACCACCGAGATCTACAC**CTATTAAG**TCGTCGGCAGCGTCAGATGTG
Ad1.16: AATGATACGGCGACCACCGAGATCTACAC**AAGGCTAT**TCGTCGGCAGCGTCAGATGTG
Ad1.17: AATGATACGGCGACCACCGAGATCTACAC**GAGCCTTAT**TCGTCGGCAGCGTCAGATGTG
Ad1.18: AATGATACGGCGACCACCGAGATCTACAC**TTATGCGA**TCGTCGGCAGCGTCAGATGTG

Ad2.1: CAAGCAGAAGACGGCATAACGAGAT**TCGCCTTAG**TCTCGTGGGCTCGGAGATGT
Ad2.2: CAAGCAGAAGACGGCATAACGAGAT**CTAGTACG**TCTCGTGGGCTCGGAGATGT
Ad2.3: CAAGCAGAAGACGGCATAACGAGAT**TTCTGCCT**GTCTCGTGGGCTCGGAGATGT
Ad2.4: CAAGCAGAAGACGGCATAACGAGAT**GCTCAGGAG**TCTCGTGGGCTCGGAGATGT
Ad2.5: CAAGCAGAAGACGGCATAACGAGAT**AGGAGTCC**GTCTCGTGGGCTCGGAGATGT
Ad2.6: CAAGCAGAAGACGGCATAACGAGAT**CATGCCTAG**TCTCGTGGGCTCGGAGATGT
Ad2.7: CAAGCAGAAGACGGCATAACGAGAT**GTAGAGAG**TCTCGTGGGCTCGGAGATGT
Ad2.8: CAAGCAGAAGACGGCATAACGAGAT**CCTCTCTG**TCTCGTGGGCTCGGAGATGT
Ad2.9: CAAGCAGAAGACGGCATAACGAGAT**AGCGTAGC**GTCTCGTGGGCTCGGAGATGT
Ad2.10: CAAGCAGAAGACGGCATAACGAGAT**CAGCCTCG**TCTCGTGGGCTCGGAGATGT
Ad2.11: CAAGCAGAAGACGGCATAACGAGAT**TGCCTCTT**GTCTCGTGGGCTCGGAGATGT
Ad2.12: CAAGCAGAAGACGGCATAACGAGAT**TCCTCTAC**GTCTCGTGGGCTCGGAGATGT
Ad2.13: CAAGCAGAAGACGGCATAACGAGAT**ATCACGAC**GTCTCGTGGGCTCGGAGATGT
Ad2.14: CAAGCAGAAGACGGCATAACGAGAT**ACAGTGGT**GTCTCGTGGGCTCGGAGATGT
Ad2.15: CAAGCAGAAGACGGCATAACGAGAT**CAGATCCA**GTCTCGTGGGCTCGGAGATGT
Ad2.16: CAAGCAGAAGACGGCATAACGAGAT**ACAAACGG**TCTCGTGGGCTCGGAGATGT
Ad2.17: CAAGCAGAAGACGGCATAACGAGAT**ACCCAGCA**GTCTCGTGGGCTCGGAGATGT
Ad2.18: CAAGCAGAAGACGGCATAACGAGAT**AACCCCTC**GTCTCGTGGGCTCGGAGATGT
Ad2.19: CAAGCAGAAGACGGCATAACGAGAT**CCCAACCT**GTCTCGTGGGCTCGGAGATGT
Ad2.20: CAAGCAGAAGACGGCATAACGAGAT**CACCACAC**GTCTCGTGGGCTCGGAGATGT
Ad2.21: CAAGCAGAAGACGGCATAACGAGAT**GAAACCCA**GTCTCGTGGGCTCGGAGATGT

Ad2.22 : CAAGCAGAAGACGGCATAACGAGAT**TGTGACCA**GTCTCGTGGGCTCGGAGATGT

Ad2.23 : CAAGCAGAAGACGGCATAACGAGAT**AGGGTCAA**GTCTCGTGGGCTCGGAGATGT

Ad2.24 : CAAGCAGAAGACGGCATAACGAGAT**AGGAGTGG**GTCTCGTGGGCTCGGAGATGT

The Ad1.X and Ad2.X oligos are adapted from ATAC-seq libraries(Buenrostro et al., 2013). The bases in bold in Ad1.X oligos are the i5 indexes, and those in Ad2.X are the i7 indexes. The combination of the 18×i5 indexes with 24×i7 indexes allows to multiplex up to 432 cells in one sequencing run.

RNaseZap (Ambion, cat. no. AM9780)

Nuclease-free water (Invitrogen, cat. no. 10977)

Triton X-100 (Sigma-Aldrich, cat. no. T9284)

Control RNA (either human or mouse purified RNA, with RNA Integrity Number > 9, diluted to 10 pg/μl before use)

ERCC RNA Spike-In Mix (ThermoFisher, cat. no. 4456740), aliquot and store in -80 °C, avoid repeated freeze-thaw cycles.

dNTP mix (10 mM each; Invitrogen, cat. no. 18427013)

5× RT buffer (ThermoFisher, cat. no. EP0751. Formula: 250 mM Tris-HCl (pH 8.3 at 25°C), 375 mM KCl, 15 mM MgCl₂, 50 mM DTT)

Maxima H Minus Reverse Transcriptase (200 U/μl) (ThermoFisher, cat. no. EP0751)

Recombinant RNase inhibitor (40 U/μl) (Clontech, cat. no. 2313A)

Betaine (BioUltra ≥99.0%; Sigma-Aldrich, cat. no. 61962)

Magnesium chloride (Sigma-Aldrich, cat. no. M8266)

Ethanol (analytical reagent grade)

KAPA HiFi HotStart ReadyMix (2×; KAPA Biosystems, cat. no. KK2601)

KAPA HiFi PCR Kits (cat. no. KK2102, note that this kit does not contain hotstart activity)

5 × TAPS-DMF buffer: see “REAGENTS AND SOLUTIONS”

Agencourt Ampure XP beads (Beckman Coulter, cat. no. A 63881)

Eppendorf™ LoBind PCR Plates (96-well; Eppendorf, cat. no. 0030129504)

Microseal 'B' PCR Plate Sealing Film (Bio-Rad, cat. no. MSB1001)

Microseal 'F' PCR Plate Seal (Bio-Rad, cat. no. MSF1001), or other foil seal

Filtered pipette tips

40 μM cell strainer (Corning, cat. no. 352340)

DynaMag™-96 Side Magnet stand (ThermoFisher, cat. no. 12331D)

Fragment analyzer (Advanced Analytical Technologies) or 2100 Bioanalyzer (Agilent Technologies) or similar instrument

Qubit Fluorometer or similar instrument

Laminar flow hood

BRAND 8-channel manifold (Sigma-Aldrich, cat. no. BR704526-1EA)

Lysis buffer preparation

1. Clean the pipettes and hood with RNaseZap. Sterilize the hood with UV-light.

*Due to the low amount of RNA present in a single cell (the order of magnitude is picograms/cell), the protocol is prone to contamination and degradation. All the procedures before **cDNA preamplification** shall be performed in a laminar flow hood. Avoid bringing any post-PCR product inside the hood. Use filter tips and change powder-free gloves regularly.*

2. Dilute the biotin-oligo-dT30VN stock (100 μ M) into 10 μ M with nuclease-free water.

3. Dilute ERCC RNA Spike-In Mix 1:1000000 with nuclease-free water.

4. Add the following lysis buffer mastermix into a 96 well PCR plate (volumes per sample):

dNTP (10 mM each dNTP)	1 μ l
biotin-oligo-dT30VN (10 μ M)	1 μ l
0.2% Triton-X100	2 μ l
Rnase Inhibitor (40 U/ μ l)	0.1 μ l
ERCC RNA Spike-In Mix (diluted)	0.1 μ l
In total	4.2 μ l

5. Add 0.5 μ l Control RNA (10 pg/ μ l) in two wells, which serve as positive controls.

6. Seal the 96 well PCR plate with a foil seal. Centrifuge 1 min at 1000 $\times$ g. Place on ice for single-cell isolation.

Single-cell isolation

Single cells from tissues or culture dishes can be dispensed into the 96 well plate by different methods according to the user's application, e.g. FACS, manual microcapillary pipetting, laser capture microdissection. Keep the volume as low as possible (preferably ≤ 0.5 μ l), as some buffers and cell culture mediums can strongly inhibit the reverse transcription. Here we use FACS as an example, accessible in most universities and research institutes.

7. Dissociate the tissue or culture cells into single-cell suspension. In order to prevent changes in transcriptional state after dissociation, subsequent steps have to be performed on ice.

8. Pellet the cell by centrifuging 3 min at 300 $\times$ g. Resuspend the cells in pre-chilled DMEM medium (or the other complete mediums).

9. Filter the cell with 40 μ m cell strainer.

10. Stain with cell viability indicator (e.g. Calcein AM) or cell death indicator (e.g. Propidium Iodide).

Optionally, staining for cell surface marker can be applied here.

11. Subject the cell suspension to FACS sorting, disposing one single-cell into each 92 wells. Don't add cells to the two positive control wells and two additional wells which will serve as negative controls.

Depending on the FACS machines and application, the settings are different. In principle, forward scatter height versus area criteria were used to discriminate doublets and gate only singlets. Positive of cell viability indicator and/or negative of cell death indicator shall be applied to remove dead cell.

The FACS needs to be carefully calibrated to make sure the sorting droplet reaches the bottom of each well on the PCR plate, where the lysis buffer is located. As the volume of the sorting droplet is in the nanoliter range it dries quickly if misplacement, subsequently resulting in RNA degradation.

12. Seal the 96 well PCR plate with a Microseal 'B' film. Centrifuge 1 min at $1000 \times g$, 4 °C. Place the PCR plate on ice and proceed to reverse transcription immediately.

If the plate will be processed at a later time point, seal the plate with Microseal 'F' film and store at -80 °C. The samples are stable for at least three months.

Reverse transcription

13. Place the PCR plate on a thermal cycler for 3 min, at 72 °C. Transfer it back on ice immediately.

If the PCR plate is transferred from -80 °C, remove the Microseal 'F' film and re-seal with the Microseal 'B' film before being placed on thermal cycler.

14. Remove the sealing film and add the following RT master mix to each well (volumes per sample):

Nuclease-free water	1.31 µl
MgCl ₂ (1 M)	0.06 µl
Betaine (5 M)	2 µl
biotin-TSO-LNA (100 µM)	0.1 µl
5* Maxima H-buffer	2 µl
Rnase Inhibitor (40U/ µl)	0.25 µl
Maxima H- RTase (200U/ µl)	0.08 µl
In total	5.8 µl

The locked nucleic acid (LNA) in the 3'-end of the biotin-TSO-LNA oligo increases the template-switching efficiency (Picelli et al., 2013), probably due to higher binding affinity between LNA and DNA. However, we found that it also increases the proportion of cDNA which are initiated by TSO rather than oligo-dT. In some rare cases, non-polyA RNA sequences (mostly rRNA) can be predominant in the sequencing reads. If the ratio of the sequencing reads mapped to rRNA is high, Biotin-TSO can be used instead of biotin-TSO-LNA. SuperScript IV Reverse Transcriptase is an alternative enzyme suitable to be used here, among the 6 reverse transcriptases we tested.

15. Seal the 96 well plate with a Microseal 'B' film, then vortex for 10 second and centrifuge for 1 min at $1000 \times g$, 4 °C.

16. Place the 96 well PCR plate on thermal cycler with a heated lid using following program:

- step 1 42 °C 90 min
- step 2 50 °C 2 min
- step 3 42 °C 2 min back to step 2 for 4 cycles
- step 4 85 °C 5 min
- step 5 4 °C hold

cDNA preamplification

17. Remove the sealing film and add the following PCR master mix to each well (volumes per sample):

2 × KAPA HiFi HotStart ReadyMix	12.5 µl
biotin-ISPCR (10 µM)	0.25 µl
Nuclease-free water	2.25 µl
Total	15 µl

18. Seal the 96 well plate with a Microseal 'B' film, then vortex for 10 second and centrifuge for 1 min at $1000 \times g$, 4 °C.

19. Place the 96 well PCR plate on thermal cycler with a heated lid using following program:

step 1	98 °C	3 min
step 2	98 °C	20 sec
step 3	67 °C	15 sec
step 4	72 °C	6 min back to step 2 for 18 cycles
step 5	72 °C	5 min
step 5	4 °C	hold

18 cycles are optimal for most cells. For cells with low RNA content (e.g. T cells), increase the PCR cycles up to 22 cycles. The samples can be safely stored at -20 °C for several months.

cDNA Purification

Starting at this point of the protocol, all steps shall be performed out of the hood used previously.

To minimize the carry-over of the primer dimers and oligo concatemers, the cDNA is purified twice as following:

20. Equilibrate AMPure XP beads and cDNA samples for 15 minutes at room temperature on the bench. Resuspend the AMPure XP beads with several seconds of vortexing.

21. Add 15 µl of resuspended AMPure XP beads (beads:sample (V/V) = 0.6:1) to each well and mix by pipetting up and down. Avoid bubble formation. Incubate for 8 min at room temperature.

22. Place the PCR plate on the 96-well Magnetic stand for 5 min.

23. Remove the liquid carefully using multichannel pipette without disturbing the beads.

Pipette from the opposite site of the bead pellet. If the tips carry the beads, simply dispose the liquid back to the wells, wait for 2 min and remove the liquid again.

24. Add 200 µl freshly prepared 80% ethanol solution to each well.

The DynaMag™-96 Side Magnet has 13 columns instead of 12. So it is possible to wash the beads by shifting the 96-well plate back and forth several times from the right-most position to the left-most position.

25. Remove the ethanol solution using multichannel pipette without disturbing the beads.

Alternately, use an 8-channel manifold connected to a vacuum pump for liquid aspiration, which can largely increase the efficiency in the washing step here and after. Make sure before to adjust the pressure so no beads are aspirated.

26. Wash the beads again by repeating step 24 and 25. Then remove the ethanol completely and air-dry for 4 min.

27. Move the PCR plate off the magnet stand. Add 25 µl nuclease-free water on the top of the bead pellet each well and resuspend by pipetting up and down. Incubate for 2 min at room temperature to elute the cDNA.

28. Purify the eluted cDNA again by repeating step 21-26. Note that the beads in the eluted cDNA are not necessary to be removed.

29. Move the PCR plate off the magnet stand. Add 17.5 µl nuclease-free water on the top of the beads each well and resuspend by pipetting up and down. Incubate for 2 min at room temperature to elute the cDNA.

30. Place the PCR plate on the DynaMag™-96 Side Magnet stand for 2 min.

31. Transfer 15 µl of the supernatant (purified cDNA) to a new 96 well PCR plate carefully, avoid carryover of beads.

If the tips carry the beads, simply dispose the liquid back to the well, wait for 2 min and transfer the supernatant again. The cDNA can be stored in -20 °C for at least 6 months.

cDNA Quality control

32. Measure the cDNA concentration using Qubit fluorometer, according to the manufacturer's instructions. Exclude the wells with concentration comparable the negative control for further processing.

The yield varies on different cells. For the positive control (5 pg purified RNA as input), the yield is among 5-15 ng when amplified for 18 cycles.

33. Measure the size distribution of the cDNA on Fragment analyzer or Bioanalyzer. A good cDNA profile is supposed to have significant signal across 500 bp to ~5000 bp, with a peak around 1.8 kb. An example is shown in **Figure 10**.

It is practical to just choose several samples to measure the size distribution, instead of measuring all of them. The user might choose 7 single-cell samples, 2 positive and 2 negative controls, which sum up 11 that fits into a single Fragment analyzer lane or one Bioanalyzer chip.

cDNA tagmentation by Tn5

34. Dilute the single cell cDNA in the 96-well PCR plate to 0.3 ng/ul with nuclease-free water.

35. Transfer 3.5 µl of the cDNA to a new 96-well PCR plate and place it on ice for subsequent tagmentation.

36. Add the following tagmentation master mix to each well (volumes per sample):

5 × TAPS-DMF buffer	1 µl
Tn5 transposase (12.5 µM)	0.5 µl

Tn5 is active at room temperature. To avoid over-tagmentation, the reactions shall be kept on ice until moved to thermal cycler at 55°C.

37. Seal the 96 well plate with a Microseal 'B' film, then vortex for 10 second and centrifuge for 1 min at $1000 \times g$, 4°C .

38. Place the 96 well PCR plate on thermal cycler with a heated lid using following program:

step 1 55°C 7 min

step 2 4°C hold

39. Move the plate on ice. Stop the tagmentation reaction by adding $1.5\ \mu\text{l}$ 0.2% SDS, vortex and incubate at room temperature for 5 min.

The volume of SDS added needs to be precise, as too much SDS will inactivate the DNA polymerase used for subsequent library amplification and too little SDS will fail to inactivate the Tn5 transposase.

Library amplification

40. Add the following library amplification PCR mastermix to the tagmented DNA (volumes per sample):

dNTP (10 mM)	1.5 μl
5 $\times$ KAPA HiFi buffer	10 μl
KAPA HiFi polymerase (1U / μl)	1 μl
Nuclease-free water	21 μl
In total	33.5 μl

As the 3'-end on both strands of the tagmented DNA needs to be repaired by DNA polymerase before PCR amplification, don't use a hotstart polymerase here which has only low polymerase activity before the first 98°C denature step.

41. Add 5 μl Ad1.x primer (3 μM) and 5 μl Ad2.x primer (3 μM) to each well as shown in **Figure 2**:

Figure 2. Adding of the index containing primers for library amplification.

42. Seal the 96 well plate with a Microseal 'B' film, then vortex for 10 second and centrifuge for 1 min at $1000 \times g$, at room temperature.

43. Place the 96 well PCR plate on thermal cycler with a heated lid using following program:

- step 1 72 °C 3 min (3'-end repair)
- step 2 98 °C 30 sec
- step 3 98 °C 10 sec
- step 4 63 °C 30 sec
- step 5 72 °C 1 min back to step 3 for 11 cycles
- step 6 72 °C 5 min
- step 7 12 °C hold

Library purification

44. Pool the libraries from the same column of the 96 well plate (e.g. A1, B1, C1...) in one 1.5 microcentrifuge tube.

While pooling libraries is less laborious and save reagents, it can lead to uneven representation of each library in case some of them have higher amounts of DNA than others. Alternatively, individual library purification as the “cDNA Purification” process, step 20 - 31, can be performed. Yet this significantly increases the workload and cost of this procedure. Comparing both sides, pooling 8 libraries from each column seems a good compromise.

45. Purify the sub-pooled libraries as described in step 20 - 31. Make sure to use the following bead:sample (v/v) ratio = 0.6:1.

Library quality control

46. Measure the library concentration using Qubit fluorometer, according to the manufacturer's instructions.

The yield should be around 80 ng/sub-pooled libraries.

47. Measure the size distribution of the sub-pooled libraries on the fragment analyzer or bioanalyzer. A good library shall contain no signal below 200 bp. The average size of the library ranges around 300-800 bp. An example is shown in **Figure 11**.

Library pooling and sequencing

48. Calculate the molarity of the sub-pooled libraries, based on the concentration (ng/μl) obtained from the Qubit and the average size obtained from the Fragment analyser, using the following formula.

$$\text{Molarity (nM)} = \text{Concentration (ng/}\mu\text{l)} \times 10^6 / ((\text{average size (bp)} \times 607.4) + 157.9)$$

49. Adjust the concentration of each sub-pooled library to 4 nM (or 2 nM or 1 nM, or 0.5 nM).

50. Pool an equal amount of each sub-pooled library together.

51. Re-check the concentration of the final pooled libraries using Qubit, then adjust to 4 nM (or 2 nM or 1 nM, or 0.5 nM) if necessary. This final library is ready for sequencing on Illumina sequencer.

52. Denature, dilute and sequence the final pooled library on Illumina sequencer according to the manufacturer's instruction.

Both the i5 and i7 indices need to be sequenced to assign the reads to each single cell. 1 million sequencing reads per cell is sufficient to apply differential gene expression analysis. For example 400

single cell libraries can be pooled in one Illumina Nextseq-500, lane. Make sure each of the individual single-cell library has a unique Ad1.X and Ad2.X combination.

SUPPORT PROTOCOL 1: IN-HOUSE TN5 PREPARATION

This protocol describes the purification of high-quality, tagmentation-ready Tn5 transposase that has comparable activity to the commercial Nextera XT DNA Library Prep Kit (Illumina), which is established by Picelli et al (Picelli et al., 2014). Briefly, following cell lysis, the Tn5-intein tag fusion protein is captured by the chitin resin, and non-specific proteins are washed away. Oligonucleotides containing the hyperactive Mosaic End (ME) sequences are annealed on-column to Tn5, and the excess of un-annealed oligonucleotides is washed. Tn5 is then released from the chitin column by dithiothreitol (DTT)-induced cleavage of the intein tag.

Materials

Oligonucleotides (desalted): dissolve to 200 μ M in 5 mM Tris, pH 8; 5 mM NaCl; 0.25 mM EDTA.

→ Tn5MErev: 5'-[phos]CTGTCTCTTATACACATCT-3'

→ FC-121-1030: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'

→ FC-121-1031: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'

Competent *E. coli* strain T7 Express *lysY/T^r* (NEB, cat. no. C3013)

pTXB1-Tn5 construct (order from Addgene: <https://www.addgene.org/60240/>)

Luria-Bertani (LB) broth medium

Ampicillin (amp; 100 mg/ml)

LB agar + amp (100 μ g/ml) plates

Isopropyl β -D-1-thiogalactopyranoside (IPTG; Sigma, cat. no. I6758)

Poly-ethyleneimine solution (PEI 50%; Sigma, cat. no. P3143)

Econo-Pac® Chromatography Columns (Biorad, cat. no. 7321010)

Chitin resin (NEB, cat. no. S6651)

5 l Erlenmeyer flasks

Shaker incubator

500-ml centrifuge tubes

Centrifuge with swinging bucket

HEGX buffer (see “**REAGENTS AND SOLUTIONS**”)

Tn5 dialysis Buffer (see “**REAGENTS AND SOLUTIONS**”)

Tn5 storage Buffer (see “**REAGENTS AND SOLUTIONS**”)

Oligonucleotides (see “**REAGENTS AND SOLUTIONS**”)

50-ml centrifuge conical tubes

15-ml centrifuge conical tubes

1.5-ml microcentrifuge tubes

Probe sonicator with a 10-mm tip

Spectra/Por 4 Membrane Dialysis, 12-14 kD (Spectrum Labs, cat. no. 132706)

Additional reagents and equipment for SDS-PAGE, Coomassie blue staining of protein gels, and Bradford protein assay.

Transformation

Freezing transformed E. coli lysY/T^q clones for later use is not recommended since the cells are RecA positive and the plasmid may not be stable in these cells.

1. Thaw a tube of T7 Express *lysY/T^q* competent *E. coli* cells on ice until the last ice crystals disappear. Cut a 200 μ l tip with sterile scissors, mix the cells gently (do not pipet) and carefully transfer 50 μ l into a transformation tube on ice.
2. Add 1 μ l pTXB1-Tn5 plasmid (1-1000 pg/ μ l) to the cells. Mix carefully the DNA and cells; do not vortex or pipet.
3. Place the mixture on ice for 30 minutes.
4. Heat shock at 42°C for 10 seconds.
5. Place on ice for 5 minutes.
6. Add 950 μ l of room temperature LB into the mixture.
7. Incubate the cells at 37°C/250 rpm in a horizontal shaker for 60 minutes. In the meanwhile, pre-warm LB ampicillin plates to 37°C.
8. Spread 50-100 μ l onto an ampicillin plate and incubate overnight at 37°C.

Growth and induction

9. Resuspend a single colony in 10 ml LB with 100 μ g/mL ampicillin and incubate at 37°C overnight.
10. Add the 10 ml pre-culture into 1 l LB with ampicillin and incubate at 37°C until OD₆₀₀ reaches ~ 0.75.
11. Chill the culture on ice to about 10°C.
12. Induce Tn5 expression by adding 250 μ l of a 1M stock of IPTG (final concentration 0.25 mM) and shake at 23°C/250 rpm until A₆₀₀ = 2.1 (takes ~ 4h).
13. Centrifuge for 10 min at 4,000 $\times$ g, 4°C; discard the medium, weigh the pellet and freeze it at -80°C.

1L of induced E. coli lysY/T^q cells yield between 2.5 to 4 g cells

Sonication

14. Thaw the frozen pellet on ice and resuspend in 60 ml ice-cold 1 $\times$ HEGX buffer.
15. Sonicate with a tip sonicator in ice. We use 8-9 times of 40 sec at 80% duty cycle, output 7 in a Branson Sonifier. Temperature should not exceed 10°C. Cool in salt-ice between each sonication cycle to -1°C. For fast cooling, use stainless steel beakers.

As several sonicators exists and even the same brand can give different results in different labs, the best sonication conditions should be determined empirically. To check lysis, save 200 μ l of unsonicated and sonicated cell suspension after 2, 4, 6 and 8 sonication cycles in 1,5 mL tubes and pellet at 15K $\times$ g, 2 min in a microcentrifuge. Mix 100 μ l supernatant of each fraction with 50 μ l 3X SDS Sample Buffer. After boiling for 5 minutes, spin it down and use 10 - 20 μ l in a SDS-PAGE run followed by Coomassie staining. The optimal condition shows the strongest Tn5 band with relative low sonication cycles.

16. Immediately centrifuge for 30 min at 23,000 $\times$ g, 4°C; collect the cleared 60 ml lysate and discard the pellet, which may contain about 20% of the protein in inclusion bodies.

PEI mediated DNA precipitation

PEI DNA precipitation reduces E. coli DNA reads in SMART-seq2 libraries.

17. Transfer the cleared 60 ml lysate into a beaker on a magnetic stirrer.
18. Dilute the 50% PEI to 10% with H₂O and add dropwise 1.6 ml into the lysate while stirring.
19. Centrifuge for 10 min at 15,000 × g, 4°C; discard the pellet and transfer the cleared cell lysate into a new flask.

Bind Tn5 to chitin

The binding capacity of chitin resin is about 2 mg/ml of fusion protein, and 3 g of cell pellet yields about 10 - 15 mg Tn5-itein, so 5 ml of chitin is used to have the Tn5 as concentrated as possible.

20. Homogenize the chitin resin by inverting the bottle several times. Then transfer 5 ml of chitin resin slurry to a 50 ml capped conical tube.
21. Allow the resin to settle and discard the supernatant.
22. Wash the resin by adding 50 ml ice-cold 1× HEGX, invert the tube several times. Next, allow the resin to settle and discard the supernatant.
23. Repeat the wash step (step 22) once more for a total of two washes.
24. Add cell lysate to the washed chitin resin and rotate overnight at 4°C.
25. Load the lysate/chitin mixture on the chromatography column.
26. Wash with 200 ml ice-cold 1× HEGX buffer. Stop the flow when the buffer surface reaches the top of the resin bead (the resin should never dry off).

On column assembly of Tn5 with oligos

This protocol requires big amounts of oligos (around 100 nmoles). It is advisable to ensure that a sufficient amount of oligo is available prior to start the subsequent steps of this protocol.

27. Prepare oligonucleotides solution as follows:

- Mix 750 µl ME-Rev (200 µM) + 750 µl FC-121-1030 (200 µM) = 1.5 ml
- Mix 750 µl ME-Rev (200 µM) + 750 µl FC-121-1031 (200 µM) = 1.5 ml
- Mix both solutions in a 15 ml capped conical tube

This is about 2x molar excess of oligos over the protein binding capacity of 5 ml chitin.

28. Transfer the tube into a beaker with 500 ml 95°C water and leave on benchtop to cool to around 50°C (monitor the water temperature with a thermometer).

This allows the ME-Rev oligo to anneal to each of the longer FC-121-1030/FC-121-1031 oligos, which is needed for assembling into Tn5.

29. After the water temperature reaches around 50°C, add 3 ml of 2× HEGX buffer and cool on ice. This solution is now ready to apply to the Tn5-chitin column.
30. Add the 6 ml of oligo solution to the top of the closed column.
31. Open the valve and let the column buffer drip off.
32. Stop the flow when the oligo solution drains down to the top of the resin beds (the resin should never dry off). There should just be a small layer of buffer still on top and around 5.5-6 ml of buffer in the collection tube.
33. Incubate the column at 4°C overnight.

34. Open the valve, let the column buffer drip off and add 100 ml ice-cold 1× HEGX to remove free oligonucleotides.

Cleave Tn5 from the Intein tag with DTT

35. Close the valve and add 14 ml ice-cold 1× HEGX with 100 mM DTT to the column.

36. Open the column valve and run out 7 ml of the HEGX buffer (to fill the column with the HEGX-DTT buffer).

37. Stop the flow by closing the valve and leave column at 4°C overnight.

Tn5 elution and molarity adjustment

38. Open the valve to elute the Tn5 protein solution. Collect the first five fractions with five drops each, so that the start of elution peak can be determined, and the subsequent fractions can be about collected in 40 drops (1 ml).

39. Check the presence of Tn5 in each fraction by mixing 10 µl of each fraction with 5 µl 3 × SDS Sample Buffer. After boiling for 5 minutes, spin it down and use 15 µl in a SDS-PAGE run followed by Coomassie staining. An example is shown in **Figure 3**.

Figure 3. SDS-PAGE gel stained with Coomassie. The 8 collected fractions are numbered; the arrowhead shows the band containing the eluted Tn5 in fractions 5, 6, 7 and 8.

40. Pool the fractions with highest concentration (about 5 ml).

In the example above, pool fractions 5, 6, 7 and 8.

41. Dialysis: transfer around 5 mL of the eluted Tn5 into a Spectra/por dialysis tubing (12-14 kD cut-off) and dialyze against 1 l 1 × Tn5 dialysis buffer at 4°C overnight.

42. Recover the dialysed Tn5 from the Spectra/por dialysis tubing.

43. Determine Tn5 concentration using the Bradford protein assay:

- Generate a standard curve with six dilutions of bovine serum albumin (BSA) ranging from 2 to 0.25 mg/ml. Add 5 µl of each sample into 250 µL of Bio-Rad protein assay solution (diluted 1:5 from stock). For the blank use 5 µl of 1 x HEGX.

44. Calculate Tn5 molarity (molecular weight of 53300 Da); it should be between 25-35 µM. For final use, adjust Tn5 concentration to 12.5 µM in Tn5 storage buffer.

Concentration methods such as Centricon should be avoided as Tn5 might precipitate during the process.

45. Aliquot and flash freeze in liquid nitrogen and store at -80°C.

Activity test

Test Tn5's ability to fragment DNA (DNA ladder tagmentation) and to generate amplifiable fragments (Tn5 tagmented library amplification), as shown below.

Tn5 DNA ladder tagmentation

46. Dilute the 1kb + DNA ladder (Fermentas) on PCR tube to 20 ng/ul with nuclease-free water.

47. Add the following tagmentation master mix to each tube (volumes per sample):

5X TAPS-DMF buffer	4.0 µl _[SEP]
1kb + DNA ladder	2.5 µl _[SEP]
Tn5 transposase	x µl (final concentration 1nM to 625nM)
Nuclease-free water	up to 20 µl _[SEP]
Total	20 µl _[SEP]

Tn5 have some activity at room temperature. To avoid of over tagmentation, the reactions shall be kept on ice until moved to thermal cycler at 55°C.

48. Mix well and incubate at 55°C for 7 min in a thermal cycler with a heated lid.

49. Move the tubes on ice and immediately purify DNA by phenol-chloroform extraction.

50. Subject to electrophoresis on a 1 % agarose gel. An example is shown in **Figure 4**.

Figure 4. 2% agarose gel stained with SYBR green. (A) Five different Tn5 concentrations were tested to tagment 50 ng of the 1Kb + DNA ladder (fermentas). (B) Same as (A), but with a naked Tn5 (not loaded with oligonucleotides). Note that differently from the loaded Tn5, even at high concentrations such as 125 and 625 nM the naked Tn5 is inert to DNA.

Tn5 tagmented library amplification

51. Take 5 μ l of the reactions with the most concentrated Tn5 (625 nM and 125 nM) and amplify using the library amplification PCR mastermix (volumes per sample):

NEBNext High-Fidelity 2x PCR Master Mix (NEB Cat #M0541)	12.5 μ l _[SEP]
transposed DNA	5.0 μ l _[SEP]
Nextera PCR Primer 1	1.25 μ l _[SEP]
Nextera PCR Primer 2	1.25 μ l _[SEP]
100x SYBR Green I (Invitrogen Cat #S-7563)	0.15 μ l
Nuclease-free water	4.85 μ l
Total volume	25 μ l _[SEP]

52. Run a qPCR using the following program (an example is shown in **Figure 5**):

- step 1 72 °C 5 min
- step 2 98 °C 30 sec
- step 3 98 °C 10 sec
- step 4 63 °C 30 sec
- step 5 72 °C 30 sec back to step 3 for 39 cycles

step 6 12 °C hold

Figure 5. PCR amplification plot. Samples 1 and 2 are libraries generated with 625nM and 125nM Tn5, respectively, while samples 3 and 4 are the control samples generated with naked Tn5 at 625nM and 125nM, respectively.

ALTERNATE PROTOCOL 1: LIBRARY PREPARATION USING NEXTERA XT KIT

Here is an alternate for the basic protocol 1, if the user prefers using the Nextera XT DNA library preparation kit instead of the in-house Tn5 preparation.

Additional Materials (also see Basic protocol 1)

Nextera XT DNA sample preparation kit, 96 samples (Illumina, cat. no. FC-131-1096)

Single-cell cDNA preparation

1. Same as step 1-33 in the basic protocol 1.

cDNA tagmentation by Nextera XT DNA library preparation kit

2. Dilute the single cell cDNA on the 96-well PCR plate to 0.3 ng/ul with nuclease-free water.
3. Transfer 3.5 µl of the cDNA (1 ng in total) to a new 96-well PCR plate and place it on ice.
4. Add the following tagmentation mastermix to each well (volumes per sample):

Nuclease-free water	1.5 µl
Tagment DNA Buffer (2 ×)	10 µl
Amplicon Tagment Mix	5 µl
In total	16.5 µl

Tn5 have some activity at room temperature. To avoid over-tagmentation, keep reactions on ice until moving to a thermal cycler at 55°C.

5. Seal the 96 well plate with a Microseal 'B' film, then vortex for 10 seconds and centrifuge for 1 min at 1000 × g, 4 °C.
6. Place the 96 well PCR plate on a thermal cycler with a heated lid using following program:
  - step 1 55°C 5 min
  - step 2 4°C hold
7. Move the plate on ice. Stop the tagmentation reaction by adding 5 µl Neutralize Tagment Buffer. Pipette to mix and incubate at room temperature for 5 min.

Library amplification

8. Add each tagmented DNA with 15 µl Nextera PCR master mix on RT.
9. Add 5 µl Index 1 primers and 5 µl Index 2 primers to each well in the same way as **Figure 2**.
10. Seal the 96 well plate with a Microseal 'B' film, then vortex for 10 second and centrifuge for 1 min at 1000 × g, at room temperature.
11. Place the 96 well PCR plate on thermal cycler with a heated lid using following program:
  - step 1 72 °C 3 min (3'-end repairing)
  - step 2 95 °C 30 sec
  - step 3 95 °C 10 sec
  - step 4 55 °C 30 sec
  - step 5 72 °C 30 sec back to step 3 for 10 cycles
  - step 6 72 °C 5 min
  - step 7 12 °C hold

Library purification

12. The same as “Basic protocol 1, step 44- 45”.

Library quality control

13. The same as “Basic protocol 1, step 46- 47”.

Library pooling and sequencing

14. The same as “Basic protocol 1, step 48- 52”.

For the library pooling steps, an alternative is to use the normalization beads according to the manual of Nextera XT DNA Library Prep kit.

BASIC PROTOCOL 2: DATA ANALYSIS

In this section we provide the bioinformatics pipeline required to pre-process the output data from the Illumina sequencer and the basic QC and analyses that need to be done when analyzing single-cell RNA-seq data. All code is provided as Unix shell instructions to be run on Unix/Linux distributions. Path to files/folders are shortcutted as `$path` for better readability, but should correspond to real path in the final script. Required software and data are listed in the Materials section.

The first pre-processing section can be performed more automatically using online web portals such as Illumina BaseSpace (basespace.illumina.com), Galaxy, or others. However, we will detail in the current protocol the different steps required to be run as well as providing existing tools that can be used to perform the pre-processing if such kind of websites were not available to the user. The pre-processed data can then be analyzed in an Automated Single-Cell Analysis Pipeline (ASAP), which is web-based and user friendly.

Materials

bcl2fastq: Software for trimming/demultiplexing *.bcl* files (raw sequencing files from the sequencing machine) and creating *.fastq* files. It can be freely downloaded from [Illumina website](https://www.illumina.com/sequencing/analysis/illumina-software/bcl2fastq.html). For demultiplexing, bcl2fastq requires a [SampleSheet.csv](#) file in the root folder containing the mapping between the barcodes and the sample names. The user can download an example of a [SampleSheet.csv](#) file at <https://github.com/DeplanckeLab/SmartSeq2>

fastQC: Software for QC of *.fastq* or *.bam* files. This software assess the quality of the reads in terms of sequencing quality, duplicate number, adapter contamination, repetitive sequence contamination, and GC content. It can be freely downloaded from its [main page](https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The website contains informative examples of [good](https://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc.html) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc.html) and [bad](https://www.bioinformatics.babraham.ac.uk/projects/fastqc/bad_sequence_fastqc.html) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/bad_sequence_fastqc.html) quality data.

STAR: software for read alignments on reference genome (Dobin et al., 2013). Can be downloaded from [Github](https://github.com/alexdobin/STAR) (<https://github.com/alexdobin/STAR>).

Samtools: software suite for SAM/BAM files (here used for indexing). Can be downloaded from [here](http://samtools.sourceforge.net/) (<http://samtools.sourceforge.net/>).

IGV (Integrative Genomics Viewer): a portable Genome Browser-like software (Robinson et al., 2011). Can be downloaded from [here](http://software.broadinstitute.org/software/igv/) (<http://software.broadinstitute.org/software/igv/>).

ASAP (Automatic Single-cell Analysis Pipeline): a web tool to analyze single-cell RNA-seq data (Gardeux et al., 2016). It does not require installation and is readily accessible at asap.epfl.ch

Demultiplexing

(If the sequencer provides already demultiplexed *.fastq* files, then the user can skip the Demultiplexing section and directly jump to the next section)

The sequencer provides raw sequencing files in the form of .bcl files that requires to be demultiplexed according to the Illumina barcodes the user chose to use (refer to Basic protocol 1 - Materials) for identifying its libraries.

1. Locate the folder where the sequencer files are generated/stored. We will call this folder **\$bclfolder**.

This folder should be automatically named following this pattern: “YYMMDD_machinename_XXXX_FC” and should contain a subfolder Data/Intensities which further contains all the .bcl files.

2. At the root of **\$bclfolder** you should create a **SampleSheet.csv** file with the association between your samples/barcodes (see **Materials** for an example of the file)

3. Output files from the demultiplexing are .fastq files that require to be created and stored in a novel location on the hard drive. We will call this folder **\$fastqfolder**.

4. .bcl files can be demultiplexed using the bcl2fastq2 software (see Materials). .fastq files are text files (can be gzipped / compressed) that contains information about all sequencing reads. **\$read1Length** and **\$read2Length** are the number of base pairs of the reads. The following command will generate as many .fastq files as there are samples to demultiplex.

```
bcl2fastq --ignore-missing-bcls --ignore-missing-filter --ignore-missing-positions --
ignore-missing-controls --no-lane-splitting --barcode-mismatches 1 --use-bases-mask
y$read1Length,i8,i8,y$read2Length -R $bclfolder -o $fastqfolder
```

For example, if you performed a paired-end sequencing with reads of length 75 (both), then the --use-bases-mask option should be set to y75,i8,i8,y75

In case you decided to perform single-end sequencing, then change the option --use-bases-mask y\$read1Length,i8,i8

Depending if the study is single-end or paired-end, the user will end with files containing only the ‘R1’ tag (single-end), or also the ‘R2’ tag (paired-end). Therefore, in case of paired-end sequencing the command will provide twice the number of .fastq files.

5. The **\$fastqfolder** will also contain a **Report** folder with a **index.html** web page. Check this report for the quantity of reads that were demultiplexed.

Sequencing quality check

6. Run fastQC (see **Materials**) on every of the .fastq files. The output folder is called **\$QCdir**. It will contain an html file and a folder containing the report from fastQC.

```
fastqc --outdir=$QCdir/ $fastqfolder/MyProject/MySample_R1.fastq.gz
fastqc --outdir=$QCdir/ $fastqfolder/MyProject/MySample_R2.fastq.gz
```

7. Check fastQC reports to assess the quality of the samples (see **Materials**).

“per base sequence content” and “kmer content” are rarely green. For single-cell data, often the “sequence duplication levels” are high (can go up to around 70% of duplicates). In this protocol, RNA-seq duplicates are not removed (it could be the case if the protocol contains UMI). If there is some “adapter contamination” / “overrepresented sequence” detected in the data, it may not be an issue (if the effect is limited to <10~20%). These are lost reads but most of them will be filtered out during the next step.

Alignment

The .fastq files then require to be aligned on a reference genome. For doing that, STAR (Dobin et al., 2013) is currently one of most efficient software. It contains a “soft-clipping” tool that automatically cut

the beginning or the end of the reads to improve the mapping efficiency, thus allowing to skip the step of trimming the reads for adapter contamination.

(Following steps 8-11 are required to be done only once)

8. Download the proper genome assembly (.fasta file, here called `$fastafile`) and gene annotation (.gtf file, here called `$gtffile`) from Ensembl or UCSC. For example, for a human assembly, we can get the following two files:

```
wget ftp://ftp.ensembl.org/pub/release-87/fasta/homo_sapiens/dna/$fastafile
gzip -d $fastafile # Because it is gzipped

wget ftp://ftp.ensembl.org/pub/release-87/gtf/homo_sapiens/$gtffile
gzip -d $gtffile # Because it is gzipped
```

Try to download the primary_assembly fasta file if available (without the 'sm' or 'rm' tags). If not available, download the top_level assembly. For the gtf, download the one that do not have the 'chr' or 'abinitio' tags.

9. Download the ERCC spike ins .fasta and .gtf files with all ERCC sequences and annotations (ERCC92.fa and ERCC92.gtf thereafter called `$erccfastafile` and `$erccgtffile`, respectively) from <https://github.com/DeplanckeLab/SmartSeq2>

10. Concatenate the downloaded .fasta file from Ensembl with the ERCC .fasta file into a final merged .fasta file (called `$mergedfastafile`) that will contain both the genome assembly and the ERCC spike ins.

```
cat $fastafile $erccfastafile >$mergedfastafile
```

11. Concatenate the downloaded .gtf file from Ensembl with the ERCC .gtf file into a final merged .gtf file (called `$mergedgtffile`) that will contain both the genes annotations from the genome of reference and the 92 entries of the ERCC spike ins.

```
cat $gtffile $erccgtffile >$mergedgtffile
```

12. Create an index for the genome assembly (the output folder containing the index files is called `$genomedir`).

```
STAR --runMode genomeGenerate --genomeDir $genomedir --genomeFastaFiles $mergedfastafile -
-sjdbGTFfile $mergedgtffile --runThreadN 8
```

The --runThreadN parameter can be modified depending on the number of cores available on your machine. The larger is this number, the more parallelized/fast will be the indexing.

(In case of single-end analysis) the following steps 12.a-15.a needs to be repeated for each .fastq file)

13.a. Align the reads to the indexed reference genome. This step will output .bam files in the folder called `$bamdir`.

```
STAR --runMode alignReads --runThreadN 8 --genomeDir $genomedir --outFilterMultimapNmax 1
--readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --outFileNamePrefix $bamdir/ -
-readFilesIn $fastqfolder/MyProject/MySample_R1.fastq.gz
```

14.a. Check the quality of the alignment in the “`$bamdir/Log.final.out`” file.

The most important statistic is the proportion of “Uniquely mapped reads” which should be greater than 50-60%.

15.a. [Optional] Indexing the BAM file can be useful for the visualization of the reads/coverage using IGV or the genome browser.

```
samtools index $bamdir/Aligned.sortedByCoord.out.bam
```

16.a. Create the final count matrix (called `counts.txt`) by running `htseq-count` (see **Materials**) for counting the number of reads that fall into each of the annotated genes (the annotation is in the GTF file).

```
htseq-count -s no -m union -f bam $bamdir/Aligned.sortedByCoord.out.bam $gtffile > counts.txt
```

(In case of paired-end analysis the following steps 12.b-15.b needs to be repeated for each pair of R1/R2 fastq files)

13.b. Align the reads on the indexed reference genome. This step will output `.bam` files (one for each pair of reads) in the folder called `$bamdir`.

```
STAR --runMode alignReads --runThreadN 8 --genomeDir $genomedir --outFilterMultimapNmax 1 --readFilesCommand zcat --outSAMtype BAM Unsorted SortedByCoordinate --outFileNamePrefix $bamdir/ --readFilesIn $fastqfolder/MyProject/MySample_R1.fastq.gz $fastqfolder/MyProject/MySample_R2.fastq.gz
```

14.b. Check the quality of the alignment in the “`$bamdir/Log.final.out`” file.

The most important statistic is the proportion of “Uniquely mapped reads” which should be greater than 50-60%.

15.b. [Optional] Indexing the BAM file can be useful for the visualization of the reads/coverage using IGV or the genome browser.

```
samtools index $bamdir/Aligned.sortedByCoord.out.bam
```

16.b. Creating the final count matrix (called `counts.txt`) by counting the number of reads that fall into each of the annotated genes (the annotation is in the GTF file).

```
htseq-count -s no -m union -f bam $bamdir/Aligned.out.bam $gtffile > counts.txt
```

(The following is the final stage of the pre-processing step; it **should be run both for single- and paired-end sequencing**)

16. The different `counts.txt` files needs to be merged into a unique file, thus generating the raw count matrix that will be used further in the analysis stage. This can be simply done within an Excel sheet or with custom scripts. Usually genes are in rows and samples are in columns. The created matrix will be called `matrix.raw.counts.txt`

Single-cell RNA-seq analysis

The analysis of the raw count matrix can be complicated to perform if the user is not familiar with R or a programming language in general. Therefore, we recommend using a tool that we developed: ASAP (Automated Single-Cell Analysis Pipeline) (Gardeux et al., 2016). This is a web-tool that do not necessitate any software to be installed and can be run directly from the raw count matrix that we just generated.

The next steps will be performed assuming that the user will use ASAP for the analysis of his/her data.

18. Open a web browser and go to <https://asap.epfl.ch>

19. The user can register and log in if he/she wishes to conserve his/her analysis. Or just create a sandbox project if he/she wants to analyze his/her data and do not store it on the server.

20. Click on “Create new project” and upload the count matrix [matrix.raw.counts.txt](#)

21. On the “Parsing results” page, check the percentage of reads that map to mitochondrial genes and ribosomal genes. The more %protein coding, the better quality is the data. Usually this number is greater than 80%. There can be some minor mitochondrial / ribosomal contamination.

22. Click on “Filtering” and filter out genes that are lowly expressed across the samples. The user can use the “Pagoda” (Kharchenko et al., 2014) tool with default parameters or just keep the most variable genes (Brennecke et al., 2013) using the “Most variable genes” tool.

23. Check the output of the filtering step and the plots. The “number of expressed genes per cell” plot (**Figure 6**) gives information about which cells have a sufficient number of genes detected as expressed. If some cells have a too low number (e.g. <1000), they should be excluded from the [matrix.raw.counts.txt](#)

Figure 6. Barplot showing the number of expressed genes per cell.

24. Click on “Normalization” to normalize data across samples. The scLVM (Buettner et al., 2015) tool specifically uses ERCCs for data normalization.

25. Click on “Visualization” to have a first glimpse of the data and clustering behavior (**Figure 7**). This should show some correlation with the biological signal that is expected by the user according to his/her hypothesis. If this is not the case then try to change the filtering (e.g. using the “variance 5%” tool) or the normalization and check again these plots. If the samples cluster according to technical batches (for e.g. different days the samples were extracted/sequenced/...) then try inputting a batch file and correct the data in the normalization step.

Figure 7. PCA plots are generated automatically after normalization. Both 2D and 3D plots are available and are fully interactive (zooming, rotating, coloring, selection of cells). The PC on the axis can also be changed. Coloring can be done according to clustering, selection, or gene expression.

26. Click on “Correlation” to perform correlations between technical replicates or samples from similar conditions. Usually the correlation of single-cell RNA-seq data shows high variation and contains drop-out events, visualized as shown in **Figure 8**.

Figure 8. Correlation plots help identify outlier samples and also helps to assess the global quality of the replicates. Here it shows drop-out events and high dispersion / variance that is typical in single-cell analyses.

27. Click on “Clustering” to group the cells/samples according to the similarity of their transcriptome. SC3 (Kiselev et al., 2017) is a very robust tool that can be run without prior knowledge of the number of clusters to expect (Auto mode). Then, the clusters can be added to selections in order to perform DE analysis (find Differentially Expressed genes).

Of note, the user can also select groups of cells directly on the plot by using the selection or the lasso tools. These selections can be renamed and edited as needed.

28. Click on the “Differential Expression” tool to perform a DE analysis using limma(Ritchie et al., 2015). The user can perform whether a search for marker genes (i.e. genes that are marker of a group of samples/cells), or between two selected groups of cells. The result table lists the genes differentially expressed according to FDR (recommended at least FDR 5%) and Fold-change (recommended at least 1.5) cutoffs.

29. The user can look at the expression of the DE genes directly on the plots in the “Visualization” tab. Or the “Enrichment” tool can provide a biological explanation of the genes that are found differentially expressed. As such, the user can select a gene set database (such as KEGG (Kanehisa and Goto, 2000), GO (Ashburner et al., 2000) or Gene Atlas) to enrich the DE genes into.

30. Finally, a heatmap can be used to plot the results of the DE step. DE genes expression can thus be plotted to highlight the differences between the groups of samples (**Figure 9**).

Figure 9. Heatmaps show gene expression differences between samples/cells. They help highlighting the up- and down-regulation and cluster the genes/samples according to their similarity.

REAGENTS AND SOLUTIONS

Use Milli-Q water in all recipes and protocol steps.

0.2% Triton X-100 solution (v/v)

dilute Triton X-100 to 0.2% with nuclease-free water.

5 x TAPS-DMF buffer:

50 mM TAPS-NaOH, pH 8.3 (RT),

25 mM MgCl₂,

50% DMF

2 x HEGX

40mM Hepes-KOH (pH 7.2)

1.6M NaCl

2 mM EDTA

20% Glycerol

0.4% TX100

Complete protease inhibitors (Roche)

2xTn5 dialysis Buffer

100 mM Hepes, pH 7.2

200 mM NaCl

0.2 mM EDTA

2 mM DTT

0.2% Triton X-100

20% Glycerol

Tn5 storage Buffer

50 mM Hepes, pH 7.2

100 mM NaCl

0.1 mM EDTA

1 mM DTT

0.1% Triton X-100

50% Glycerol

COMMENTARY

Background Information

Due to the heterogeneity of biological system, studies based on bulk cell populations are prone to dilute or miss cell types or states among the whole population. To detect the property of each individual cells, several traditional techniques had been implemented to profile the single cell characters, such as FACS, single cell qPCR, RNA FISH, etc. However, these techniques suffer from important limitations, either in term of the number of targets that can be detected, or the number of cells that can be processed. For instance, FACS can only detect several channels/genes simultaneously, and FISH is limited by the

time-consuming probe optimization and its low-throughput (it is needed to mention that the very recently reported multiplex FISH methods have greatly improved the throughput (Moffitt et al., 2016).)

Recently, several single cell RNA-seq methods have been developed, which allow detecting the whole transcriptome of hundreds to thousands of cells (or even more, e.g. 10⁴ Genomics Chromium platform). Many cells types have been successfully subjected to single cell RNA-seq, from culture cells to primary cells, from bacteria to mammalian cells (Wu et al., 2017).

As the low amount of RNA in single cells, the cDNA needs to be amplified to meet the need of the sequencer. The first single cell RNA-seq study use polyA tailing to add a stretch of A to prime the second strand synthesis, and then the cDNA is amplified by PCR using the priming size introduced before (Tang et al., 2009). Later, multiple methods have been developed mainly based on the template switching mechanism (e.g. STRT-seq (Islam et al., 2011), Smart-seq, SCRB-seq (Soumillon et al., 2014), Drop-seq (Macosko et al., 2015) , etc.) or in vitro transcription (IVT) (Cel-seq (Hashimshony et al., 2012), MARS-seq (Jaitin et al., 2014), etc.). The template switching based methods utilize the template-switching activity of the MMLV reverse transcriptase to add a priming sequence in the 3' of the cDNA, while another priming sequence is incorporated in the 5' in the oligo-dT (Figure 1). Then the cDNA are amplified using the corresponding primers. In order to reduce the bias introduced by PCR, the IVT based methods linearly amplify the cDNA by IVT and then reverse transcribed to cDNA again later. These methods offer different data characteristics, as such Smart-seq2 covering the full length of the transcripts, the other methods enrich either the 5' or the 3' end of the RNA sequence; some methods (SCRB-seq, Drop-seq, Cel-seq2, etc) multiplex the cells at the beginning, while Smart-seq2 only multiplex after the tagmentation steps; some use the unique molecular identifiers "UMI" (Islam et al., 2014) to eliminate the amplification bias while some not. Comprehensive comparison of the methods have been done (Svensson et al., 2017; Wu et al., 2014; Ziegenhain et al., 2017). As there is no superior method so far, the user can choose according to the needs.

New library preparation methods based on the Tn5 transposase are highly efficient in time, throughput and initial input requirements as they simultaneously fragment the DNA and incorporate adaptors for amplification in a several minutes reaction (Adey et al., 2010). This has been particularly relevant for single-cell based protocols, where diminutive amounts of RNA are an important limiting factor. As in Smart-seq2 each cell has to be handled individually, each cell is prepared individually, increasing substantially the library preparation costs. Therefore, the production of home-made Tn5 not only diminishes drastically the costs associated with library preparation, but importantly enables protocol optimization and innovation.

The analysis of single-cell RNA-seq data is very similar to the analysis of bulk RNA-seq data. There are specific methods that exist specifically for scRNA-seq data that we suggested in the protocol (such as SC3 or scLVM), but the overall analysis is similar. We also saw that drop-out events and high variability was expected with single-cell data, so it should be kept in mind when analyzing the data. ASAP does not cover all possible aspects of the analysis of single-cell data so it is possible that extra analysis would be needed. ASAP allows the user to download the normalized data so that the user can load the data in R if he/she needs additional analysis.

Critical Parameters

Clean work environment and consumable

Single cell RNA-seq is very sensitive and thus little contamination of RNA, DNA or Rnase can lead to bias or even completely failure of the experiment. The hood and bench shall be clean and wiped with RNaseZAP. All the reagents and consumables used before the cDNA amplification shall be free of RNA, DNA, Rnase and Dnase.

Cell sample preparation

The transcriptome of the cells can change during the sample preparation process. Make sure to minimize the sample preparation time and site the cells on ice whenever possible. Use mild condition to dissociate the cells from tissues to minimize the cell death during this process. Too much cell death will lead to the cell suspension to be very viscous and hard to isolation.

FACS

Improper FACS calibration, especially the positioning of the 96 well (or 386 well) plate can cause the cDNA amplification to fail completely.

Oligos quality

High quality oligos with HPLC (or PAGE) purification are required for the reverse transcription and template-switching.

Reagent quality

High quality reagents are required. Reagents shall be aliquoted to avoid freeze and thaw cycling.

Tn5 tagmentation

The library size after the Tn5 tagmentation strongly depends on the amount of input cDNA. It is important to have the same amount of single cell cDNA to make sure the same size distribution of the libraries, which is crucial to reduce the potential bias.

Tn5 production

Tn5 production has been reported to fail in multiple laboratories, so we suggest the user to carefully follow the instruction given here. Crucially, plasmid transformation should be done prior to every purification. Tn5 concentration during the process should be kept high, so the user is advised to follow the volumes suggested here as diluted Tn5 might have lower activity. Additionally, concentration steps with centricon or amicon should be avoided as the Tn5 frequently precipitates in these circumstances.

Troubleshooting

Problems		Potential reasons	Solutions
No yield/low yield of single cell cDNA	The positive controls work	The cells are not sorted directed to the lysis buffer	Calibrate the FACS
	The positive controls fail	Failure of RT or the subsequent PCR	Change reagents and oligos

	Low yield of cDNA	Too little input RNA (e.g. small cells)	Increase the PCR cycles
The negative controls have cDNA amplified		contamination	Change reagents, prepare the RT and PCR reactions in a clean laminar hood, improve personal practices
The size of sequencing library too large/small		less/over tagmentation	Increase/reduce the tagmentation time or Tn5 dose
Too many rRNA reads detected		Use biotin-TSO instead of biotin-TSO-LNA	Reduce the reverse transcriptase dose
Tn5 production unsuccessfully	No Tn5 detected after induction with IPTG	Culture started from glycerol stocks of transformed <i>E. coli</i> <i>lysY/T^q</i> clones	Transform fresh cells for every purification
Inefficient library tagmentation	Tn5 not active	Tn5 was too diluted during the purification process	Adjust pellet resuspension volume according to its weight
		Tn5 precipitated during concentration steps after dialysis	Only pool concentrated fractions after column elution and avoid posterior concentration

Anticipated Results

Figure 10: cDNA profile from single brown pre-adipocyte. Left is an example of high quality cDNA and the right is a typical TSO concatemers dominant cDNA profile.

Figure 11. The library profile after Tn5 tagmentation and purification.

50% ~ 70% of the reads mapping to the genome is expected. Depending on the cell type, around 5,000 to 8000 genes are normally identified. The Pearson correlation between two positive controls is usually around 0.7.

Time Considerations

For 96 samples, the library preparation takes 3 days, sequencing takes 1 day (on Illumina platform), data pre-processing takes between 1 and 2 days. Tn5 production would take 6 days. Data analysis times are variable and depend on the hypothesis required to be validated, which can require extended periods of time. However, with the ASAP tool, the basic data analysis should be achievable within hours.

Acknowledgements

We would like to thank Johannes Bues and Riccardo Dainese for critical reading of the manuscript. Wanze Chen has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under grant agreement 665667 (call 2015). Antonio Meireles-Filho was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme. This work has been supported by institutional support from the EPFL.

Literature Cited

- Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., *et al.* (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 11, R119.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25-29.
- Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., *et al.* (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* 10, 1093-1095.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (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.
- Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* 33, 155-160.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.

Gardeux, V., David, F., Shajkofci, A., Schwalie, P.C., and Deplancke, B. (2016). ASAP: a Web-based platform for the analysis and inter-active visualization of single-cell RNA-seq data. *bioRxiv*.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2, 666-673.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 21, 1160-1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 11, 163-166.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., *et al.* (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776-779.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27-30.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat Methods* 11, 740-742.

Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R., *et al.* (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods*.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., *et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202-1214.

Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P., and Zhuang, X. (2016). High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* 113, 11046-11051.

Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10, 1096-1098.

Picelli, S., Bjorklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24, 2033-2040.

Ramskold, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., *et al.* (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 30, 777-782.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat Biotechnol* 29, 24-26.

Soumillon, M., Cacchiarelli, D., Semrau, S., van Oudenaarden, A., and Mikkelsen, T.S. (2014). Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv*.

Svensson, V., Natarajan, K.N., Ly, L.H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A., and Teichmann, S.A. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat Methods* 14, 381-387.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., *et al.* (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6, 377-382.

Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F., *et al.* (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods* 11, 41-46.

Wu, A.R., Wang, J., Streets, A.M., and Huang, Y. (2017). Single-Cell Transcriptional Analysis. *Annual review of analytical chemistry*.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* 30, 892-897.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* 65, 631-643 e634.

Internet Resources

ASAP website: <https://asap.epfl.ch>

Github repository for required files: <https://github.com/DeplanckeLab/SmartSeq2/>