Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing

John J. Trombetta,1 David Gennert,1 Diana Lu,1 Rahul Satija,1 Alex K. Shalek,2 and Aviv Regev1,3

1Broad Institute of MIT and Harvard, Cambridge, Massachusetts
2Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, Cambridge, Massachusetts
3Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

For the past several decades, due to technical limitations, the field of transcriptomics has focused on population-level measurements that can mask significant differences between individual cells. With the advent of single-cell RNA-Seq, it is now possible to profile the responses of individual cells at unprecedented depth and thereby uncover, transcriptome-wide, the heterogeneity that exists within these populations. This unit describes a method that merges several important technologies to produce, in high-throughput, single-cell RNA-Seq libraries. Complementary DNA (cDNA) is made from full-length mRNA transcripts using a reverse transcriptase that has terminal transferase activity. This, when combined with a second “template-switch” primer, allows for cDNAs to be constructed that have two universal priming sequences. Following preamplification from these common sequences, Nextera XT is used to prepare a pool of 96 uniquely indexed samples ready for Illumina sequencing. Curr. Protoc. Mol. Biol. 107:4.22.1-4.22.17. © 2014 by John Wiley & Sons, Inc.

Keywords: SMART-Seq • single-cell RNA-Seq • transcriptomics • cell type • single cell • low-input RNA-Seq • SMART-Seq2 • Next-generation sequencing • template-switching

INTRODUCTION

The development of single-cell RNA-Seq provides new opportunities to study complex cellular systems at unprecedented resolution. While previous RNA-Seq methods could only capture the average behavior of a population, single-cell methods now enable studies of cellular heterogeneity, as well as the discovery of novel cell types and states. These methods employ a variety of techniques to address the technical challenges posed by single cells—most notably, the ultra-low input quantity of RNA.

SMART-Seq SINGLE-CELL RNA-Seq (METHOD)

Here, the terminal transferase activity of the SMARTScribe reverse transcriptase is exploited in conjunction with a “template-switch” primer to make cDNAs, that have PCR priming sites on both ends, directly from full-length mRNA (Fig. 4.22.1). First, single cells are lysed in a guanidine thiocyanate solution. Second, a bead cleanup is used to isolate RNA. Third, cDNA is synthesized from the nearly full-length mRNA transcripts. Fourth, these cDNA products are amplified by PCR and normalized. Fifth, the amplified cDNA is incubated with a Tn5 transposase to fragment nearly full-length double-stranded cDNAs and append adapters on each molecule. Sixth, each single-cell library is individually barcoded by PCR with index primers, and the sample set is pooled. Finally, the resulting pool is purified by a bead cleanup and is then sent for Illumina sequencing.
Figure 4.22.1 Single-cell RNA-Seq library construction. Illustrated are the steps required to convert mRNA to cDNA with sequencing adapters.

In this protocol, Microseal F is used when sealing for long-term storage and Microseal B when thermal cycling.

**Materials**

- 2-Mercaptoethanol, ≥99.0%
- TCL buffer (Qiagen, cat. no. 1031576)
- Cell suspension in complete medium
- Dry ice
- RNeasy Plus Micro Kit (Qiagen)
- RNA-SPRI beads (Agencourt RNAClean XP RNA-SPRI beads, Beckman Coulter, cat. no. A63987)
- RNAse decontamination solution (RNaseZap, Life Technologies, cat. no. AM9780)
- RNase-free water
- 80% ethanol
- SMARTer Ultra-Low Input RNA Kit for Illumina Sequencing-HV, 96 reactions (Clontech, cat. no. 634828) containing:
  - Dilution buffer
  - 3' SMART CDS Primer II A (12 µM): 5'
    - AAGCAGTGGTATCAACGCAGAGTACT(30)VN
  - RNase inhibitor (40 U/µl)
  - 5× first-strand buffer (RNase-free)
  - Dithiothreitol (DTT; 100 mM)
  - SMARTer dNTP Mix (dATP, dCTP, dGTP, and dTTP, each at 20 mM)
  - SMARTer II A oligonucleotide (12 µM): 5'
    - AAGCAGTGGTATCAACGCAGAGTACATrGrGrG
  - SMARTScribe reverse transcriptase (100 U/µl)
Nuclease-free water
IS PCR primer (12 μM): 5’ AAGCAGTGGTATCAACGCAGAGT
Purification buffer (10 mM Tris-Cl, pH 8.5)
Control total RNA (1 μg/μl)
Advantage 2 PCR Kit, 100 reactions (Clontech, cat. no. 639206) containing:
10× Advantage 2 PCR buffer
50× dNTP Mix (dATP, dCTP, dGTP, and dTTP, each at 10 mM)
50× Advantage 2 polymerase mix
PCR-grade water
10× Advantage 2 SA PCR buffer
Control DNA template (100 ng/μl)
Control primer mix, 5’ primer and 3’ primer (10 μM each)
Agencourt AMPure XP SPRI beads (Beckman Coulter, cat. no. A63881)
1 × TE buffer (see recipe)
Agilent High Sensitivity DNA Kit (Agilent Technologies, cat. no. 5067-4626)
Qubit dsDNA HS assay kit, 100 reactions (Life Technologies, cat. no. Q32851)
Taqman Fast Advanced Master Mix, 100 reactions (Life Technologies, cat. no. 4444557)
Taqman Probe set (choose housekeeping genes that are appropriate for specific system)
Nextera XT DNA Sample Preparation Kit, 96 samples (Illumina, cat. no. FC-131-1096) containing:
Amplicon Tagment Mix (ATM)
Tagment DNA Buffer (TD)
Nextera PCR Master Mix (NPM)
Neutralize Tagment Buffer (NT)
Resuspension Buffer
Library Normalization Additives 1
Library Normalization Wash 1
Hybridization Buffer
Library Normalization Beads 1
Library Normalization Storage Buffer 1
Nextera XT Index Kit, 96 indices, 384 samples (Illumina, cat. no. FC-131-1002) containing:
Index primers, S501 to S508 (8 tubes)
Index primers, N701 to N712 (12 tubes)
96-well PCR plates, skirted (Eppendorf, cat. no. 951020401)
1.5-ml RNAse- and DNAse-free microcentrifuge tubes
Microseal F foil (Bio-Rad, cat. no. MSF-1001)
FACS machine
Plate centrifuge
Vortex
DynaMag-96 side-skirted magnet (Life Technologies, cat. no. 12027)
8-channel pipettor
Thermal cycler
Microseal B adhesive seals (Bio-Rad, cat. no. MSB-1001)
2100 Electrophoresis Bioanalyzer Instrument (Agilent Technologies, cat. no. G2939AA)
Qubit assay tubes (Life Technologies, cat. no. Q32856)
Qubit 2.0 fluorometer (Life Technologies, cat. no. Q32866)
TruSeq Index Plate Fixture Kit (Illumina, cat. no. FC-130-1005)
DynaMag-2 magnet (Life Technologies, cat. no. 12321D)
Illumina sequencer

Preparation and Analysis of RNA
4.22.3
**Prepare single-cell lysates**

1. Prepare a 1% solution of 2-mercaptoethanol (by volume) in TCL buffer and distribute 10 µl of this solution into each well of a skirted-side 96-well PCR plate, and 350 µl into a 1.5-ml RNase-free microcentrifuge tube. Cover the plate with Microseal F and keep at room temperature until ready for single-cell isolation.

2. Prepare a cell suspension in complete medium and use a FACS machine to sort a single cell into each well of the abovementioned 96-well plate containing TCL buffer, and an additional sample of ≥10,000 cells into the 1.5-ml microcentrifuge tube to use as a population control. Once sorting is complete, seal plate with Microseal F foil and centrifuge 1 min at 800 × g, room temperature. Immediately freeze plate and population control on dry ice and keep up to 1 year at −80°C until ready for lysate cleanup.

   To improve yield, sorting single cells on both the presence of a positive viability indicator (e.g., Calcein AM, Life Technologies) and the absence of a cell death marker (e.g., the membrane-impermeant DNA stain EthD-1, Life Technologies) is recommended.

**Perform lysate cleanup and reverse transcription of mRNA species**

The following steps can be performed inside a biosafety cabinet or a RNA workstation (if available); otherwise, they can be carefully performed on a standard benchtop.

3. Purify RNA from the population control sample using the RNeasy Plus Micro Kit according to the manufacturer’s recommendations. Keep purified RNA at 4°C until step 18 (−80°C for up to 1 year).

4. Bring RNA-SPRI beads to room temperature (allow 30 min) and use RNaseZap to clean workbench and all equipment used to process RNA.

5. Thaw lysate plate on ice, and then centrifuge 1 min at 800 × g, room temperature.

6. Thoroughly vortex RNA-SPRI beads to ensure a uniform suspension. Add 2.2 vol RNA-SPRI beads to each well of lysate (2.2 × 10 µl cell lysate = 22 µl of beads are added, and 2.2 × 15 µl = 33 µl) and mix well by pipetting up and down ten times. If each solution is not in the bottom of its respective well, centrifuge plate 1 min at 300 × g, room temperature.

   The following steps are done at room temperature.

   Do not centrifuge plates containing SPRI beads at >300 × g. Beads will adhere to the well wall and impair RNA recovery.

7. Incubate lysate and bead suspension for 10 min on bench.

   If this procedure is performed on a standard benchtop (i.e., not in an RNA fume hood), keep the plate covered (e.g., using the lid from a fresh box of pipet tips, placed slightly ajar) to prevent dust and debris from falling into samples during this and subsequent incubation steps.

8. Place plate on 96-well plate magnet (e.g., DynaMag-96 side-skirted magnet) and incubate for 5 min. Remove supernatant from each column of the plate, being careful not to aspirate the beads collected on the side of each well.

   The magnet described draws the beads to alternating sides of each column of the plate. It is therefore recommended to use an eight-channel pipettor to remove the supernatant.

9. Wash beads by adding 75 µl of 80% ethanol (prepared same day with nuclease-free water) to each well. Wait ~30 sec. Aspirate ethanol and repeat wash two additional times.
10. Aspirate final ethanol wash. Leave the plate on the magnet and allow beads to dry for ~10 min at room temperature. Keep plate loosely covered with the lid from a fresh box of pipet tips, placed slightly ajar.

Allowing beads to dry for too long will impair recovery of RNA. Immediately proceed with protocol when visible cracks appear in bead pellet.

11. Once bead pellet has dried, remove plate from magnet and elute RNA from beads by resuspending dried beads in 4.5 µl of the following mix, using reagents from the SMARTer Ultra-Low Input RNA Kit (volumes per sample):
   - 2.375 µl dilution buffer
   - 1 µl RNase-free H₂O
   - 1 µl 24 µM 3’ SMART CDS Primer II A
   - 0.125 µl RNase inhibitor.

   Do not remove eluent from original plate.

   For population control, substitute water with 1 µl purified RNA.

12. Seal and incubate 1 min at room temperature.

13. Centrifuge 1 min at 800 × g, room temperature.

14. Incubate 3 min at 72°C to anneal 3’ SMART CDS Primer II A.

15. Place plate on ice immediately following incubation.

16. Remove seal and add the following as a mastermix (volumes per sample):
   - 2 µl first strand buffer
   - 0.25 µl 100 mM DTT
   - 1 µl 10 mM dNTP mix
   - 1 µl 12 µM SMARTer II A oligonucleotide
   - 0.25 µl RNase inhibitor
   - 1 µl SMARTScribe reverse transcriptase.

   Mix well by pipetting up and down, reseal with a new Microseal B adhesive seal, and centrifuge 1 min at 800 × g, room temperature.

17. Carry out the following reverse transcription in a thermal cycler using the following conditions:

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Time</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90 min</td>
<td>42°C (initial step)</td>
</tr>
<tr>
<td></td>
<td>10 min</td>
<td>72°C (inactivate)</td>
</tr>
<tr>
<td></td>
<td>Hold</td>
<td>4°C (cool)</td>
</tr>
</tbody>
</table>

Following this step, the product can be stored overnight at 4°C.

Residual SPRI beads remain in each well with the libraries until step 28.

The protocol provided with the SMARTer kit describes an SPRI cleanup at this stage of the protocol, which is omitted here. It is important to avoid losing material before amplification, and it has been found that the cleanup is not essential to generate a high-quality sequencing library (Picelli et al., 2013; Shalek et al., 2013).

Perform whole transcriptome amplification and post-PCR cleanup

18. Centrifuge plate 1 min at 800 × g, room temperature and add the following from the Advantage 2 PCR kit as a mastermix (volumes per sample):
   - 5 µl 10× Advantage 2 PCR buffer
   - 2 µl 10 mM dNTP mix
   - 2 µl 12 µM IS PCR primer (from Clontech Ultra-Low Input RNA Kit)
   - 2 µl 50× Advantage 2 polymerase mix
   - 29 µl H₂O.

Preparation and Analysis of RNA

4.22.5
Mix well by pipetting up and down, seal with Microseal B adhesive seal, and centrifuge 1 min at 800 × g, room temperature.

19. Carry out the whole transcriptome amplification in a thermal cycler using the following conditions:

<table>
<thead>
<tr>
<th>Cycle Type</th>
<th>Time</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 cycle</td>
<td>1 min</td>
<td>95°C (initial denaturation)</td>
</tr>
<tr>
<td>5 cycles</td>
<td>20 sec</td>
<td>95°C (denaturation)</td>
</tr>
<tr>
<td></td>
<td>4 min</td>
<td>58°C (annealing)</td>
</tr>
<tr>
<td></td>
<td>6 min</td>
<td>68°C (extension)</td>
</tr>
<tr>
<td>9 cycles</td>
<td>20 sec</td>
<td>95°C (denaturation)</td>
</tr>
<tr>
<td></td>
<td>30 sec</td>
<td>64°C (annealing)</td>
</tr>
<tr>
<td></td>
<td>6 min</td>
<td>68°C (extension)</td>
</tr>
<tr>
<td>7 cycles</td>
<td>20 sec</td>
<td>95°C (denaturation)</td>
</tr>
<tr>
<td></td>
<td>30 sec</td>
<td>64°C (annealing)</td>
</tr>
<tr>
<td></td>
<td>7 min</td>
<td>68°C (extension)</td>
</tr>
<tr>
<td>1 cycle</td>
<td>10 min</td>
<td>72°C (final extension)</td>
</tr>
<tr>
<td></td>
<td>Hold</td>
<td>4°C (cool)</td>
</tr>
</tbody>
</table>

Following this step, the product can be stored overnight at 4°C.

For this amplification, a few modifications to the thermal cycling steps described in the original SMARTer protocol have been made. These modifications improve annealing, denaturation, and extension for the diverse transcript sets with which the authors work.

20. Bring DNA SPRI beads (Agencourt AMPure XP) to room temperature.

21. Centrifuge whole transcriptome amplification plate 1 min at 800 × g, room temperature.

22. Unseal and add 0.8 vol of DNA SPRI beads to each well and mix well by pipetting up and down.

23. Cover with a clean lid from a pipet tip box and incubate bead suspension for 5 min on bench.

24. Place plate on 96-well plate magnet and incubate for an additional 5 min. Carefully remove supernatant.

25. Wash beads by adding 100 µl of 80% ethanol to each well and shift the plate left and right on the magnet to move the beads from side to side in each well. Continue shifting plate on magnet for ~30 sec. Aspirate ethanol and repeat wash two additional times.

26. Aspirate final ethanol wash. Leave the plate on the magnet and allow beads to dry ~10 min at room temperature. Keep plate covered with the lid of a pipet tip box (left slightly ajar) to prevent dust and debris from falling into samples.

27. Once bead pellet has dried, remove plate from magnet and elute DNA from beads by resuspending dried beads with 20 µl of 1× TE buffer. Transfer eluent to a new 96-well plate.

Following this step, the new product plate can be sealed and stored overnight at 4°C or for months at −20°C.

28. Use 1 µl of the purified PCR product to measure the fragment size distribution using the Agilent HS DNA BioAnalyzer or similar instrument according to the manufacturer’s instructions and 1 µl to estimate the library concentration using the Qubit dsDNA HS Assay kit with Qubit assay tubes and the Qubit fluorometer or similar instrument according to the manufacturer’s instructions.
29. Use 1 µl of the purified PCR product to measure single-cell viability with the Taqman qPCR assay.

In most single-cell sample sets, a small subset of libraries will have low complexity and will lack expression of highly expressed housekeeping genes. Since there are, in most systems, a set of housekeeping genes that should be expressed in every cell (Shalek et al., 2013), use qPCR of two housekeeping genes (usually ACTB and B2M) to identify viable samples. Samples with low or no expression of these genes can be removed from Nextera library construction to reduce cost.

30. For each sample, dilute the purified PCR product to a concentration between 0.1 and 0.2 ng/µl with TE buffer in preparation for Nextera XT library construction.

**Construct Nextera XT sequencing library**

31. Before beginning library construction, thaw index primers and mix by brief vortexing followed by centrifugation. Arrange Nextera index tubes in the TruSeq Index Plate Fixture such that each slot in the fixture holds one index tube. The 12 index 1 (i7) primers, which have orange caps, should be arranged horizontally in order, such that each tube corresponds to a column (Fig. 4.22.2). The column 1 slot should hold N701 and the column 12 slot should hold N712. Similarly, the index 2 (i5) primers should be arranged vertically in order, such that S501 is in the slot for row A and S508 is in the slot for row H. Using a multichannel pipettor, distribute 10 µl of each index into the corresponding row/column (i7 primers will be distributed down rows, while i5 primers will be distributed across columns). Seal, vortex mildly to mix, and then centrifuge 1 min at 800 × g, room temperature.

These mixed index primers can be used immediately, or stored long-term at −20°C.

There is enough combined index primer solution in each well to generate 8 × 96 single-cell libraries.

32. Using a new 96-well plate, add to each well:
   - 2.5 µl TD Buffer
   - 1.25 µl ATM
   - 1.25 µl diluted PCR product from SMART RT-PCR.

---

**Figure 4.22.2** Layout to prepare the combined dual-index primer plate that is used in Nextera XT library construction.
Mix well by pipetting up and down, seal, and centrifuge 1 min at 800 × g, room temperature.

33. Carry out tagmentation using the following conditions:

<table>
<thead>
<tr>
<th>Initial step:</th>
<th>10 min</th>
<th>55°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cool:</td>
<td>hold</td>
<td>10°C</td>
</tr>
</tbody>
</table>

34. Unseal and immediately add 1.25 µl NT Buffer to each well and mix well by pipetting up and down. Wait 5 min at room temperature.

35. Add 3.75 µl NPM and 2.5 µl combined Index primer solution to each well. Mix well by pipetting up and down, seal, and centrifuge 1 min at 800 × g, room temperature.

36. Carry out the amplification in a thermal cycler using the following conditions:

1 cycle: 3 min 72°C (initial annealing)
1 cycle: 30 sec 95°C (denaturation)
12 cycles: 10 sec 95°C
30 sec 50°C
60 sec 72°C
1 cycle: 5 min 72°C
hold 4°C.

Following this step, the product can be stored overnight at 4°C.

Pool DNA and perform DNA SPRI bead cleanup

37. Bring DNA SPRI beads (Agencourt AMPure XP) to room temperature.

38. Centrifuge plate 1 min at 800 × g, room temperature, unseal, and pool 2.5 µl from each well into a single 1.5-ml microcentrifuge tube. Measure total volume of resulting pool, add 0.9 volumes SPRI beads to tube, and mix well by pipetting up and down.

39. Incubate bead suspension for 5 min on bench.

40. Place tube on DynaMag-2 magnet, and incubate for 5 min. Carefully remove supernatant.

41. Wash beads by adding 500 µl of 80% ethanol to the tube and rotate it on the magnet to move the beads from front to back across the tube. Aspirate ethanol and repeat wash one additional time.

42. Aspirate final ethanol wash. Leave the tube on the magnet and allow beads to dry ~10 min at room temperature. Keep the tube covered to prevent dust and debris from falling into the sample.

43. Once bead pellet has dried, remove the tube from the magnet and elute DNA from beads using 30 µl TE buffer (consider smaller volumes if working with <96 samples). Transfer eluent to a new 1.5-ml tube and repeat 0.9 × SPRI bead cleanup one additional time.

44. Use 1 µl of purified PCR product to measure the fragment size distribution using the Agilent HS DNA BioAnalyzer or similar instrument according to the manufacturer’s instructions and 1 µl to estimate the library concentration using the Qubit dsDNA HS Assay kit with Qubit assay tubes and Qubit fluorometer or similar instrument according to the manufacturer’s recommendations.

Perform sequencing

45. Sequence libraries on an Illumina sequencer using paired-end reads.
The authors typically run two times 25 bp for expression quantification. Expression estimates will saturate \(~1 \times 10^6\) aligned reads, so one lane of HiSeq should be sufficient for 96 single cells.

**SMART-Seq2 SINGLE-CELL RNA-Seq MODIFIED METHOD**

In single-cell RNA-Seq, small amounts of sample loss during a number of steps can lead to significant decreases in transcript detection sensitivity. A decrease in assay sensitivity results in data that is only accurate and reproducible for highly expressed genes, limiting the scope and confidence of gene expression analyses. Further complications in assay sensitivity arise from an uneven distribution of sequencing reads along a transcript; usually, in SMARTer, there is a bias towards more reads at the 3′ end of the transcript. Even coverage along a transcript improves the accuracy of analytical tools used to quantify gene expression and transcript isoform abundance. SMART-Seq2 published by Picelli et al. (2013) modified the traditional SMARTer protocol to address this by improving transcript detection, coverage, accuracy, yield, and cost. Following the same strategy as the SMARTer library construction, SMART-Seq2 uses several alternative reagents to generate whole-transcriptome, full-length cDNA libraries.

Avoiding small-volume, bead-based SPRI cleanup of each sample is an effective way of reducing loss and increasing assay sensitivity. Lysing single cells in a guanidine thiocyanate buffer necessitates SPRI cleanup due to the protein denaturing effects of the compound, which will affect downstream reactions, like reverse transcription. Multiple alternative lysis buffers exist that address this. The Ambion single cell lysis buffer (Life Technologies, cat. no. 4458235), often used for single-cell RT-PCR, only requires the addition of a stop solution to inactivate its lytic activity before subsequent reactions. A hypotonic lysis buffer with small amounts of RNase-inhibitor and surfactant, as described in SMART-Seq2, is the preferred buffer due to the lack of a need for a post-lysis cleanup or the addition of a stop solution prior to reverse transcription. However, the optimal lysis strategy will depend on the experimental system being analyzed.

SMART-Seq2 takes additional steps to minimize sample loss during library construction. The reverse transcription is improved by the addition of betaine and additional magnesium chloride to the reaction mix and by the use of a template-switch oligonucleotide with one locked nucleic acid (LNA) riboguanosine base. These improvements facilitate the hybridization between the template-switch oligonucleotide and the cDNA product, thereby increasing the probability of successfully introducing a second PCR adapter onto the cDNA product (see Fig. 4.22.1). A second key improvement was made in the preamplification PCR step, which can be heavily biased against either long transcripts or those containing regions with high G/C content. Picelli et al. (2013) found that the preamplification PCR is improved by using the KAPA HiFi HotStart ReadyMix, which dramatically improved coverage and sensitivity, particularly for GC-rich transcripts.

**Additional Materials (also see Basic Protocol)**

- Triton X-100 (Sigma-Aldrich, cat. no. T8787)
- SUPERase-In RNase inhibitor, 40 U/µl (Life Technologies, cat. no. AM2694)
- 10 µM reverse transcription DNA oligonucleotide primer (custom synthesized by Integrated DNA Technologies):
  - AAGCAGTGGTATCAACGCAGAGTACT(30)VN
- dNTP mix (dATP, dCTP, dGTP, and dTTP, each at 10 mM) (Thermo-Fisher, cat. no. R0192)
- SuperScript II Reverse Transcriptase Kit (Life Technologies, cat. no. 18064-014) containing:
5 × first-strand buffer
100 mM DTT
SuperScript II reverse transcriptase
5 M betaine
100 mM MgCl₂
10 µM SMARTer template switching oligonucleotide (TSO, custom synthesized by Exiqon): 5′ AAGCAGTGGTATCAACGCAGAGTACATGrGrG + G
RNase-free water
KAPA HiFi HotStart PCR ReadyMix (KAPA Biosystems, cat. no. KK2601)
10 µM PCR oligonucleotide primer (custom synthesized by Integrated DNA Technologies): AAGCAGTGGTATCAACGCAGAGT

Prepare single-cell lysates
1. Prepare a mild hypotonic lysis buffer of 0.2% Triton X-100 and 2 U/µl RNase inhibitor and distribute 4 µl of this solution to each well of a 96-well plate.
2. Sort single cells with a FACS machine as described in the Basic Protocol.

Reverse transcribe mRNA species
3. Prepare work space and thaw plate as described in the Basic Protocol. Add 1 µl of 10 µM anchored oligo-dT RT primer and 1 µl of 10 mM dNTP mix to each lysate well, seal plate with Microseal B, and centrifuge 1 min at 800 × g, room temperature.
4. Incubate 3 min at 72°C and then immediately place on ice.
5. Centrifuge plate 1 min at 800 × g, room temperature, unseal, and add the following as a mastermix:
   - 2 µl 5 × SuperScript II first-strand buffer
   - 0.25 µl 100 mM DTT
   - 2 µl 5 M betaine
   - 0.9 µl 100 mM MgCl₂
   - 1 µl 10 µM template switching oligonucleotide
   - 1 µl RNase-free H₂O
   - 0.25 µl 40 U/µl RNase-inhibitor
   - 0.5 µl 200 U/µl SuperScript II reverse transcriptase.
   Mix well by pipetting up and down, seal plate, and centrifuge 1 min at 800 × g, room temperature.
6. Carry out the following reverse transcription in a thermal cycler using the following conditions:
   - Initial step: 90 min 42°C
   - 10 cycles: 2 min 50°C
   - 2 min 42°C
   - 1 cycle: 15 min 70°C (inactivation)
   - Hold 4°C (cool).

Perform WTA and post-PCR cleanup
7. Add the following as a mastermix:
   - 13 µl KAPA HiFi HotStart ReadyMix
   - 1 µl 10 µM IS PCR primer.
8. Carry out the amplification in a thermal cycler using the following conditions:
Initial step: 3 min 98°C
20 cycles: 15 sec 98°C
20 sec 67°C
6 min 72°C
1 cycle: 5 min 72°C (extension)
hold 4°C (cool).

Following this step, the product can be stored overnight at 4°C.

9. Carry out DNA SPRI bead cleanup of preamplification product as described in Basic Protocol steps, 16 through 24.

Construct Nextera XT sequencing library
10. Carry out Nextera XT library construction as described in Basic Protocol, steps 25 through 30.

Pool products and perform SPRI bead cleanup
11. Carry out pooling and SPRI bead cleanup as described in Basic Protocol, steps 31 through 38.

SINGLE-CELL RNA-Seq EXPRESSION ANALYSIS
Following sequencing of the cDNA libraries on an Illumina sequencer, data is generated as a series of files in the FASTQ format. For each unique sample specified in the sequencing sample sheet, four files are generated: one containing the “left-hand” read data (one end of the paired-end reads), one containing the “right-hand” read data (the other end of the pair), one containing the “left-hand” Nextera indexing read data, and one containing the “right-hand” Nextera indexing read data. RNA-Seq analysis uses computational tools to match each read pair, align the read pair to the genome sequence, and quantify the number of reads that align within each annotated gene.

The GenomeSpace Web portal was developed to assist researchers with minimal computational analysis experience. Using its drag-and-drop interface, data sets and modules of pre-built analytic tools can be organized into customizable pipelines for numerous applications. Despite its ease of use, GenomeSpace uses cloud storage and computing power, making it less efficient for a large number of sequencing analyses or if a researcher has access to higher computing power at their own institution; as an alternative strategy for higher throughput, a Unix-based workflow is also provided.

Using GenomeSpace for expression analysis
2a. Upload each of the raw FASTQ files from the sequencing run into the home directory of the GenomeSpace interface via drag-and-drop onto the GenomeSpace interface.
3a. Under the Recipes drop-down menu on the GenomeSpace interface, select Analyzing Data with GenomeSpace Tools.
4a. Select the appropriate application for which the data will be analyzed, and follow the instructions to generate an analysis pipeline using the tools available through GenomeSpace.

Using Unix command line for expression analysis
1b. Ensure that the following programs are installed and ready to use on the computer or server that will run the analysis:
2b. Run the program TopHat to match each of the paired-end reads with its mate and align the reads to the desired reference genome (Trapnell et al., 2009). The following files are required:

(i) Reference genome index: download at ftp://ftp.ccb.jhu.edu/pub/data/bowtie_indexes/
(ii) Transcriptome reference annotation file (.GTF): see http://tophat.cbcb.umd.edu/igenomes.shtml for up-to-date reference annotations
(iii) Left reads FASTQ file (either compressed or uncompressed)
(iv) Right reads FASTQ file (either compressed or uncompressed)

% tophat [options] [genome index base] [~/Left-Reads.fastq.gz] [~/RightReads.fastq.gz]
For full list of [options], see http://tophat.cbcb.umd.edu/manual.shtml
(v) TopHat will create several files, including ~/accepted_hits.bam. The data in this file contains the alignment and pairing information for all reads that successfully paired and aligned in the BAM file format

3b. Sort the ~/accepted_hits.bam file using the Picard tool SortSam.jar, which organizes the BAM file data based on the aligned reads’ locations in the reference genome.

% java --jar ~/picard/SortSam.jar
   I=~/accepted_hits.bam O=~/accepted_hits.sort.bam
   SO=coordinate

4b. The Picard suite of command-line tools contains programs that provide important metrics regarding the sequencing data. Run various tools on the ~/accepted_hits.bam file to analyze the sequencing quality. Helpful tools include the following:

CollectAlignmentSummaryMetrics.jar—view statistics on the number of reads that correctly align to the reference genome
CollectInsertSizeMetrics.jar—view statistics on the lengths of the sequenced cDNA fragments

% java -jar ~/picard/CollectAlignmentSummaryMetrics.jar
   [options] I=~/accepted_hits.bam O=~/AlignmentMetrics.txt
% java -jar ~/picard/CollectInsertSizeMetrics.jar [options]
   H=~/InsertSizeHistogramChart.txt I=~/accepted_hits.bam
For full list of tools and [options], see http://picard.sourceforge.net/command-line-overview.shtml.

5b. Index the sorted BAM file using the Samtools suite of command-line tools (Li et al., 2009).

% samtools index ~/accepted_hits.bam
6b. For graphical visualization of the data, load the aligned sequencing reads into the program Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2012). Open the IGV program, select the appropriate reference genome, and load the ~/accepted_hits.sort.bam file. This will display a graph for each file loaded with the location within the genome on the horizontal axis and the number of reads on the vertical axis.

7b. For a quantitative measure of gene or transcript abundance, run the program Cufflinks on the ~/accepted_hits.sort.bam file. The two primary output files generated by Cufflinks represent quantified expression level estimates at the gene and transcript isoform level. Each of these presents the expression level estimates as a table relating each gene or transcript to its relative abundance in fragments per kilobase of exon per million mapped fragments (FPKM).

Files required:
BAM data file
Transcriptome reference annotation file (.GTF)
(see http://cufflinks.cbcb.umd.edu/igenomes.html for up-to-date reference annotations)

% cufflinks [options] ~/accepted_hits.sort.bam

For full list of [options], see http://cufflinks.cbcb.umd.edu/manual.html (Trapnell et al., 2012).

REAGENTS AND SOLUTIONS

Use Milli-Q purified water or equivalent in all recipes and protocol steps. For common stock solutions, see APPENDIX 2; for suppliers, see APPENDIX 4.

**TE buffer, 1×**

10 mM Tris-Cl, pH 8.0 (APPENDIX 2)
1 mM EDTA
Store indefinitely at room temperature

COMMENTARY

**Background Information**

It has long been known that population-level observations of cellular transcription do not always accurately represent transcriptional patterns across the single cells that comprise them (Eberwine et al., 1992; Shapiro et al., 2013; Sandberg, 2014). Earlier methods developed to explore this transcriptional heterogeneity in cellular populations first interrogated individual genes of interest in single cells (Eberwine et al., 1992), and later allowed for highly multiplexed qRT-PCR of up to 96 genes simultaneously across 96 single cells (see, for example, Sanchez-Freire et al., 2012).

Recent technical advances that make high-throughput construction of next-generation sequencing libraries possible is the introduction of uniform adapter sequences to the transcripts of interest, allowing for the amplification and sequencing of many mRNA transcripts. A number of strategies have been developed to efficiently add these adapters. Pioneering work on mRNA-Seq utilized poly(A) tailing of reverse transcribed products followed by poly(T)-primed PCR amplification (Tang et al., 2009). CEL-Seq, a more recently published method, relies on in vitro transcription (IVT) to linearly amplify reverse transcribed products, followed by ligation of adapter sequences to the 3’ end of amplified RNA (Hashimshony et al., 2012). Presented here, the SMARTer protocol leverages the terminal transferase activity of an
M-MLV-derived reverse transcriptase to reverse transcribe mRNA and then, with a template-switch primer, add an adapter sequence in a single reaction (Zhu et al., 2001). Each method has its own distinct advantages, disadvantages, and biases specific to the biochemical reactions underlying each protocol. For example, CEL-Seq avoids biases introduced by PCR amplification of reverse transcription products by linearly amplifying its reverse transcription products with IVT; this, however, necessitates a cleanup of both reverse transcription products and IVT amplification products prior to subsequent reactions (Hashimshony et al., 2012). With varying levels and tolerances for biases arising in specific experimental contexts, selecting the appropriate method for a given study depends on the system being assayed.

As the throughput and power of single-cell transcriptome profiling increases, these methodologies can be applied to increasingly complex biological systems. Whole-transcriptome sequencing across cell types can reveal transcriptional profiles specific to each type, as well as the heterogeneity within them. Identification of cell types from a population of cells of unknown composition based solely on transcriptional profiles is now possible, and is increasingly easy as more single-cell RNA-Seq profiling data is added to the knowledge base (Shapiro et al., 2013). Meanwhile, primary cells, or rare subtypes of cells, desirable for study, may not appear in extracted populations in sufficient frequency to conduct RNA-Seq at a population level, necessitating single-cell methods. The transcriptional profiles of a patient’s tumor cells may elucidate the origin, development, and function of various cell lineages within a tumor. Other questions about cellular response to stimuli (Shalek et al., 2013), transcriptional fluctuation across time (Shapiro et al., 2013), and differential gene expression in disease or pathological tissue (Sandberg, 2014) are now being explored; it is envisioned that many more topics will be fruitfully investigated with single-cell RNA-Seq as it becomes a more widely used tool.

**Critical Parameters and Troubleshooting**

Proper sample preparation is key when preparing successful single-cell RNA-Seq libraries, as degradation and contamination of input material are the most common sources of error in downstream steps. Prior to sorting, sample integrity can be improved by minimizing the time spent preparing cells for sorting, and then freezing lysate plates no more than 10 min after sorting has concluded. When preparing for cDNA synthesis, it is vitally important to ensure that all workspaces, equipment, and reagents are RNase free by using products such as RNase-ZAP and RNase inhibitors. Cross contamination is also of great concern throughout this protocol, and care should be taken to avoid sharing tips between samples. Additionally, samples should be covered whenever possible to prevent airborne contamination.

Because sample integrity is a common source of error, and because small quantities of starting material are frequently used in this protocol, several parameters must be assayed following cDNA synthesis, amplification, and cleanup to ensure library quality. Average fragment size, cDNA concentration, and the presence of transcripts for housekeeping genes are effective measures of library quality. Measuring, via qPCR, at least two well-established housekeeping genes is particularly important at this stage. In virtually any set of single-cell libraries, there will be samples of low complexity that lack expression of key housekeeping genes, and the results of qPCR measurements can be used to eliminate these samples.

The Tn5 transposase used in Nextera XT library construction fragments dsDNA at a rate proportional to [dsDNA] in the specific library. Thus, it is crucial to normalize the concentration of your SMARTer single-cell amplified cDNAs prior to Nextera XT library construction. This ensures that dsDNA from each single-cell sample will yield tagmentation products with similar distributions of fragment length, and further that after pooling, cleanup, and sequencing the distribution of reads across each sample will be unbiased.

Table 4.22.1 describes some problems commonly encountered with these protocols along with descriptions of possible causes and suggestions for overcoming or avoiding these problems.

**Anticipated Results**

Following RT, PCR, and SPRI cleanup, successful single-cell libraries can have concentrations ranging from 0.1 to 50 ng/µl. At this stage in the protocol, each library should have an average fragment size of ~2000 bp, with few or no species <400 bp as shown in Figure 4.22.3. When running qPCR, genes in viable libraries will amplify in fewer than ~26 cycles. Contaminated samples may yield libraries that align poorly to a corresponding reference genome, which can often
<table>
<thead>
<tr>
<th>Problem</th>
<th>Possible cause</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little or no WTA yield</td>
<td>RNA degradation</td>
<td>Start with high-quality RNA; use RNA with RNA integrity number &gt; 8; begin with a single RNA sample and work quickly; work in an RNase-free environment</td>
</tr>
<tr>
<td></td>
<td>RNA-SPRI failure</td>
<td>Make sure beads have been fully resuspended prior to use; make sure beads are never exposed to RNases; use freshly prepared ethanol; do not overdry beads</td>
</tr>
<tr>
<td></td>
<td>Reverse transcription failure</td>
<td>Make sure all primers are added to the correct reaction</td>
</tr>
<tr>
<td></td>
<td>Too few PCR cycles</td>
<td>Increase number of PCR cycles (using final cycling profile, i.e., 7→9)</td>
</tr>
<tr>
<td></td>
<td>PCR failure</td>
<td>Poor library generation, possibly due to inactive enzymes/reagents, replace reagents</td>
</tr>
<tr>
<td></td>
<td>PCR-SPRI failure</td>
<td>Make sure beads have been fully resuspended prior to use</td>
</tr>
<tr>
<td></td>
<td>Cell sorting failure</td>
<td>Make sure cell sorter dispenses cells accurately to center of each well</td>
</tr>
<tr>
<td>Presence of WTA products that are shorter than expected (&lt;500 bp)</td>
<td>RNA degradation</td>
<td>Start with high-quality RNA; use RNA with RNA integrity number &gt; 8; begin with a single RNA sample and work quickly; work in an RNase-free environment</td>
</tr>
<tr>
<td></td>
<td>PCR-SPRI failure or ineffective cleanup</td>
<td>Make sure beads have been fully resuspended prior to use; repeat 0.8× SPRI cleanup</td>
</tr>
<tr>
<td>Regularly-spaced spiky peaks on WTA Bioanalyzer trace</td>
<td>Contamination</td>
<td>Make sure to prepare samples and reactions in a clean and sterile environment</td>
</tr>
<tr>
<td>Small fragment size after Nextera</td>
<td>Tagmentation reaction ran too long</td>
<td>Reduce reaction time of Nextera tagmentation</td>
</tr>
<tr>
<td></td>
<td>Low tagmentation input cDNA concentration</td>
<td>Make sure tagmentation input cDNA is at the correct concentration; make sure the cDNA Bioanalyzer traces show little primer contamination (&lt;500 bp peaks) after Nextera</td>
</tr>
<tr>
<td>Low alignment rate of reads to genome</td>
<td>Contamination</td>
<td>Make sure to prepare samples and reactions in a clean and sterile environment</td>
</tr>
<tr>
<td>Low library complexity/number of genes detected</td>
<td>Too many PCR cycles</td>
<td>Reduce number of PCR cycles</td>
</tr>
<tr>
<td></td>
<td>Loss of material prior to PCR</td>
<td>Make sure loss of material is limited during the initial RNA-SPRI cleanup</td>
</tr>
</tbody>
</table>
Figure 4.22.3 Typical best-case results following RT-PCR and cleanup. The ideal average WTA product is ~2000 bp.

Figure 4.22.4 Regularly-spaced spiked peaks indicate sample contamination.

be identified by regularly spaced spiky peaks between 100 and 600 bp on a Bioanalyzer trace, as shown in Figure 4.22.4.

After a Nextera library has been pooled and cleaned up, its concentration can range from 2 to 100 ng/µl, depending largely on the volume used for elution after the final SPRI cleanup. These libraries will have average fragment sizes of 350 to 550 bp with a somewhat flat peak as shown in Figure 4.22.5.

Time Considerations
The SMARTer library construction and normalization take 2 to 3 days, and Nextera XT library construction and pooling requires 1 day.
Figure 4.22.5  Typical best-case results following Nextera library construction, pooling, and normalization. The ideal average fragment size for a Nextera library is 350 to 500 bp.

Acknowledgements

Work was supported by an NIH Postdoctoral Fellowship (1F32HD075541-01, RS), an NIH CEGS Award (1P50HG006193-01, AR), NIH Pioneer Awards (DP1OD003958-01 to AR), the Broad Institute (AR), HHMI (AR), and the Klarman Cell Observatory at the Broad Institute (AR).

Literature Cited


Preparation and Analysis of RNA

4.22.17