**Title:** A sensitive and integrated approach to profile messenger RNA from samples with low cell numbers

**Authors:** Sandy Lisette Rosales1,4, Shu Liang1,4, Isaac Engel1, Benjamin Joachim Schmiedel1, Mitchell Kronenberg1,2, Pandurangan Vijayanand1,3, Grégory Seumois1

**Affiliations:**

1 La Jolla Institute for Allergy and Immunology, La Jolla, U.S.A.

2 Division of Biologic Sciences, University of California San Diego, La Jolla, U.S.A.

3 Clinical and Experimental Sciences, National Institute for Health Research, Southampton Respiratory Biomedical Research Unit, University of Southampton, Faculty of Medicine, Southampton, U.K.

4 These authors contributed equally to this work

Corresponding author: G. Seumois, gregory@lji.org

**Running head:** RNA-Seq assay for small cell number samples or single cells.

**Correspondence:** 9420 Athena Circle, La Jolla, CA 92037, U.S.A.

**A sensitive and integrated approach to profile messenger RNA from samples with low cell numbers**

Sandy Lisette Rosales, Shu Liang, Isaac Engel, Benjamin Joachim Schmiedel, Mitchell Kronenberg, Pandurangan Vijayanand, and Grégory Seumois

**Abstract**

Transcriptomic profiling by RNA sequencing (RNA-Seq) represents the preferred approach to measure genome-wide gene expression for understanding cellular function, tissue development, disease pathogenesis, as well as to identify potential biomarkers and therapeutic targets. For samples with small cell numbers, multiple methods have been described to increase the efficiency of library preparation and to reduce hands-on time and costs. This chapter reviews our approach, which combines flow-cytometry and the most recent high-resolution techniques to perform RNA-Seq for samples with low cell numbers as well as for single-cell samples. Our approach reduces technical variability while increasing sensitivity and efficiency. Thus, it is well-suited for large-scale gene expression profiling studies with limited samples for basic and clinical studies.

**Keywords:** RNA-Seq, Single cell, low-input, transcriptomics, Smart-Seq 2

**1 Introduction**

Large-scale transcriptomic profiling studies constitute an important first step to understanding diverse biological phenomena such as tissue development, cellular differentiation and functionality, and disease pathogenesis [1-3]. They also represent a promising tool for development of diagnostics and therapeutics [4]. However, such studies can be limited by sample availability and heterogeneity, especially in immunology [5]. The heterogeneous and complex cellular composition of tissue samples can confound analyses by increasing the biological noise and masking subtle but biologically relevant changes in mRNA expression. Diverse methods have been used to address these issues, for example, using more accessible sources of cells, such as blood, or using *ex-vivo* culture to expand rare cell populations derived from tissue samples [6,7]. However, multiple reports have shown that the data derived by these alternative methods do not always reflect the exact processes that occur *in vivo*. Recently developed protocols in RNA sequencing for samples with low cell number or single-cell RNA-Seq hold great potential for exploring biological systems with unprecedented resolution [8]. Among these protocols, Smart-Seq2 has shown significant advantages, such as increased efficiency of reverse-transcription and library preparation, as well as reduced hands-on time and costs [9]. Smart-Seq2 has been recently appreciated as the most reproducible and sensitive method for low-input RNA-Seq and single-cell RNA-Seq [10]. The protocol consists of an oligo-based Poly-A tailed mRNA capture followed by a high-fidelity reverse transcription using locked nucleic acids (LNA)-oligos to capture full-length transcripts. mRNA complementary (c)DNA strands are then amplified using a PCR-based process and sequencing adaptors integrated into the cDNA fragments in a single step using Tn5 transposase technology [9].

Our work explores the merits of both bulk RNA-Seq and single-cell RNA-Seq in revealing disease-associated patterns or changing paradigms in development of rare immune cell subsets [2,3]. To overcome problems arising from tissue paucity and heterogeneity, we designed a flow cytometry-based method for isolating pure populations from dispersed tissue samples, coupled with the highly sensitive, medium-throughput RNA-Sequencing Smart-Seq2 protocol. The procedure allows for the performance of both bulk RNA-Sequencing and single-cell RNA-Seq assays from the same cell sorting experiment. We have successfully applied our protocol to multiple research projects. In the study by *Engel et al.,* we performed RNA-Seq analysis at the bulk and single-cell levels in three thymic *i*NKT cell subsets [3]. Analysis showed an extensive and unexpected diversity of global gene expression levels revealing unique cell type-specific functional molecular patterns. We have also applied our protocol to investigate qualitative differences in Th2 cells from subjects with allergic asthma and allergic rhinitis. Although both diseases share clinical and pathological features characterized by an exaggerated Th2 type inflammation, allergic rhinitis patients do not develop asthma, suggesting a divergence in disease mechanisms. Analysis of circulating Th2 cells isolated from both disease groups and from healthy subjects revealed differentially expressed genes involved in cell survival, metabolic pathways and activation persistence [2]. In this chapter, we provide a detailed step-by-step description of the entire procedure from cell sorting to final library quantification for either bulk or single-cell RNA-Seq. It is an integrated, multi-sample and highly-sensitive approach. Fig.1 displays anoverview of critical steps, quality control steps, and hands-on time estimations for the planning of experiments.

**2 Materials**

**2.1 Long term common stock solutions** (*see* **Note 1**)

1. 1 M Tris pH 8.0\*
2. 0.5 M EDTA pH 8.0\*
3. TE buffer 1\*: 10 mM Tris pH 8.0; 1 mM EDTA pH 8.0, store at 4 ˚C.
4. TE buffer 2\*: 1 mM Tris pH 8.0, 100 µM EDTA pH 8.0, store at 4 ˚C.
5. TE buffer 3\*: 10 mM Tris pH 8.0, 100 µM EDTA pH 8.0, store at 4 ˚C.

**2.2 Common reagents and specific laboratory equipment**

1. Molecular biology grade water (DNase, RNase, Protease, endotoxin – free; referred as ultrapure water).
2. RNase away solution (see **Note 2**).
3. 70 % and 80 % ethanol solutions (prepare fresh).
4. Recombinant RNase inhibitor, store at -20 ˚C.
5. Ethanol 200 proof, anhydrous 99.5 %
6. 10 mM dNTP, store at -20 ˚C.
7. Collecting tube: 1.5 mL Axygen maxymum recovery PCR tubes (Fisher Scientific)
8. Plate 1: 96-well semi-skirted PCR plates (BioRad) (*see* **Note 3**)
9. Plate 2: Hard-shell thin-wall 96-well skirted PCR plates
10. 0.2 mL RNase-free PCR 8-tube strips
11. Plate 3: MicroAmp Fast 96-Well Reaction Plate (Life Technologies, *see* **Note 4**)
12. qPCR MicroAmp optical adhesive film (Life Technologies)
13. MicroAmp Clear Adhesive Film (Life Technologies)
14. 96-well magnetic device (Axygen)
15. Repeater micropipette
16. Thermomixer
17. Real-time quantitative PCR system
18. Capillary DNA-RNA electrophoresis equipment (Advance Analytical Fragment analyzer or Agilent Bioanalyzer).

**2.3 Specific reagents and buffers**

**2.3.1 Sample Collection from size-limited samples**

 For samples with more than 5,000 cells:

1. TRIzol LS, store at room temperature (*see* **Note 5**).
2. 1.5 mL sterile tubes: RNase-free, individually wrapped.

For single-cell or low cell number collection:

1. 2x low input cell lysis buffer (LI-LB)\*: 0.2 % Triton X-100 (vol/vol) (*see* **Note 6**)

**2.3.2 RNA extraction**

1. miRNeasy micro kit (Qiagen) - contains the following: miRNeasy MinElute spin columns, RWT and RPE buffers. The RPE and RWT buffers have to be reconstituted with 100 % ethanol.
2. RNase-free DNase I kit (Qiagen) – contains the following: RNase-free DNase I, RNase-free Buffer RDD and Rnase-free water.
3. DNAse mix: 1:8 vol:vol ratio of reconstituted RNase-free DNase I and Rnase-free Buffer RDD.
4. Chloroform
5. 0.5 mL Nunc cryobank vials (ThermoFisher)

**2.3.3 RNA quality and quantity measurement**

1. Superscript III reverse transcriptase kit (Invitrogen), store at -20 ˚C.
2. Shaved ice from -80 °C freezer.
3. Annealing mix: 2.5 µM OligodT(20) primers, 5 ng/µL of random hexamers, and 1.2 U/µL of RNase-OUT for a total volume of 3.5 μL per sample (*see* **Note 7**).
4. Superscript III reverse transcription mix: 1x strand buffer, 5 mM MgCl2, 10 mM DTT, 0.5 mM dNTP, 1.2 U/μL of RNase-OUT, and 10 U/μL SuperScript III enzyme for a total volume of 5.6 µL per sample.
5. Human *β2m* housekeeping gene primers: 25 nM FWD: 5’- CTG CCG TGT GAA CCA TGT GAC TTT-3’; 25 nM REV: 5’-TGC GGC ATC TTC AAA CCT CCA TGA-3’ (*see* **Notes 8** and **9**).
6. SybrGreen FAST universal 2x qPCR master mix (Roche)
7. Housekeeping gene qPCR mix: 1x SybrGreen Master Mix, 0.31 µM for each primers, complete to 7.5 µL with ultrapure water.
8. Applied Biosystems QuantStudio 6 Flex Real-Time PCR System

**2.3.4 cDNA synthesis by high fidelity reverse transcription**

1. 5 M Betaine (*see* **Notes 1** and **10**).
2. Superscript II reverse transcriptase (RT) kit: includes RT enzyme and 25 mM MgCl2 solution (Life Technologies), store at -20 ˚C.
3. Template Switching Oligo (TSO): 5’-AAG CAG TGG TAT CAA CGC AGA GTA CAT rGrG+G-3’, HPLC purified (Exiqon); dilute with TE buffer 1 to 100 µM (*see* **Notes 8** and **9**).
4. oligo-dT30VN: 5’-ACA AGC AGT GGT ATC AAC GCA GAG TAC T(30)VN-3’, HPLC purified (Integrated DNA Technologies); dilute with TE buffer 1 to 100 µM (*see* **Notes 8** and **9**).
5. Reverse transcription mix (RT-mix): 100 U/μL Superscript II RT, 10 U/μL RNase Inhibitor, 1x SuperScript II first strand buffer, 5 mM DTT, 1 M betaine, 0.6 mM MgCl2, 1 μM template switching oligo (TSO).

**2.3.5 cDNA amplification**

1. SybrGreen dye, store at -20 ˚C.
2. 50x SybrGreen solution = 1:1000 dilution of stock.
3. 50x Rox reference dye (Invitrogen), store at -20 ˚C.
4. KAPA HiFi HotStart ready mix kit, store at -20 ˚C.
5. Oligo ISPCR: 5’-AAG CAG TGG TAT CAA CGC AGA-3’, HPLC purified (Integrated DNA Technologies); dilute with TE buffer 1 to 100 µM (*see* **Notes 8** and **9**).

**2.3.6 cDNA size selection**

1. Agencourt Ampure XP beads, store at 4 ˚C (Beckman Coulter).

**2.3.7 cDNA Quantification**

1. Bacteriophage lambda DNA (Life Technologies). Sonicate and dilute up to 40 ng / µL with TE buffer 1 (*see* **Note 11**).
2. Quant-iT PicoGreen dsDNA reagent (Life Technologies), store at -20 ˚C.
3. Prepare PicoGreen solution by diluting Quant-iT PicoGreen dsDNA reagent 500-fold with TE buffer 3 (*see* **Note 12**).
4. Plate 4: 96-well plate, flat bottom, black for fluorescence measurements.
5. Top-read 96-well plate reader as for example SpectraMax M2 Multi-Mode Microplate Reader

**2.3.8 cDNA library preparation**

1. Nextera XT DNA library preparation kit: 96 reactions, store at -20 ˚C.
2. Nextera XT index kit: 96 indexes, store at -20 ˚C.
3. Tagmentation mix: Illumina Nextera XT reagents amplicon tagmentation buffer and tagmentation DNA mix in a 1:2 ratio, respectively***.***
4. Agencourt Ampure XP beads (Beckman Coulter), store at 4 ˚C.

**3 Methods**

Before starting any bench work, U.V. sterilize instruments (pipettes, racks, etc.) for at least 15 minutes, clean bench and equipment with 70 % ethanol solution and RNase away solution (*see* **Note 2**).

**3.1 Sample Collection from size-limited samples**

Depending on the cell number and the scientific objectives, we suggest three distinct approaches to sort cells (*see* Fig. 1). For sample sizes ranging between 5,000 to 200,000 cells, start from **3.1.1**. For single cells or cell numbers less than 5,000 cells, proceed directly to **3.1.2** (*see* **Note 13**).

**3.1.1 Sample collection for samples with more than 5,000 cells.**

If the size of the cell population of interest ranges between 5,000 to 200,000 cells, we suggest performing the sorting directly into TRIzol LS lysis buffer. Proceed with the following steps:

1. Directly sort cells into 750 μL of TRIzol LS in RNase-free individually wrapped 1.5 mL tubes.
2. During the sort, either after every 5 minutes of sorting, or after every 10,000 cells sorted, pause sorting, close tubes, and vortex for about 15 seconds. Place tubes back on the sorting device and resume sorting (*see* **Note 14**).
3. After sorting, vigorously pipette each sample to lyse cells properly.
4. Fill up the volume to 1 mL with ultrapure water, vortex for 30 seconds, incubate at room temperature for 1 minute, then briefly spin and store samples at -80 ˚C (*see* **Note 15**). Once all samples are collected, proceed with the RNA extraction (**3.2**).

**3.1.2 Single-cell or low cell number collection**

1. For less than 5,000 cells, we recommend sorting a fixed cell number for all cell populations of interest into a fixed volume of complete LI-LB (*see* **Note 16**). Sort cells directly into 0.2 mL RNase-free tubes. For single cells, we recommend sorting cells directly in 96- plates with 4 µL of complete LI-LB (Plate 1).
2. Dispense the appropriate volumes (see Table 1, **Note 17**) of complete LI-LB into 0.2 ml PCR tubes or 4 μL into each plate well for single cell sorts.

[Table 1]

1. After sort collection, promptly close tubes or seal plates firmly with adhesive seal. Vortex with care for 15 seconds and spin at 3,000 g-force for 2 minutes.
2. Samples are ready for cDNA synthesis or can be stored at -80 °C until needed.

* 1. **RNA extraction**

RNA extraction is performed with Qiagen miRNeasy micro kits (*see* **Note 18**). We have slightly modified the manufacturer’s protocol to increase extraction yield.

1. Add 100 µL (1:5 vol:vol) of chloroform to 500 µL of cells lysed in TRIzol LS and close tube securely. Vortex the tube for 30 seconds.
2. Incubate at room temperature for 2 to 3 minutes. Centrifuge for 15 minutes at 12,000 g-forceat 4 °C.
3. Transfer the upper aqueous phase to a fresh 1.5 mL collecting tube (*see* **Note 19**).
4. Measure volume and add 1.5x volumes of 100% ethanol to the tube. Mix thoroughly by pipetting (*see* **Note 20**).
5. Place up to 700 μL of sample, including any precipitate, into a miRNeasy MinElute spin column.
6. Place the loaded miRNeasy MinElute spin column in a 2 mL tube (from Qiagen kit). Close the lid and centrifuge at 10,000 g-forcefor 30 seconds at room temperature.
7. Re-load the flow-through into the miRNeasy MinElute column, close the lid and centrifuge at 10,000 g-forcefor 30 seconds at room temperature.
8. Discard the flow-through by pipetting. Repeat **steps 5**-**8** until all of the sample (and flow-through) has been put through the column.
9. Add 350 μL of Qiagen RWT buffer into the column. Spin at 10,000 g-forcefor 30 seconds at room temperature. Discard the flow-through.
10. Prepare DNase mix (*see* **2.3.2**). Dispense 80 μL of DNase mix on every column. Incubate for 15 minutes at room temperature.
11. Add 500 μL of RWT buffer to the RNeasy MinElute spin column, close lid and spin the column at 10,000 g-forcefor 30 seconds at room temperature.
12. Pipette the flow-through and re-load it into the RNeasy MinElute column for a second time. Close lid and spin the column at 10,000 g-forcefor 30 seconds at room temperature. Discard the flow-through by pipetting.
13. Pipette 500 μL of Qiagen RPE buffer onto the RNeasy MinElute spin column, close lid, and centrifuge for 30 seconds at 10,000 g-force at room temperature*.* Discard the flow-through by pipetting.
14. Add 500 μL of freshly prepared 80 % ethanol to the RNeasy MinElute spin column, close the lid, and centrifuge for 2 minutes at 10,000 g-force. Discard the flow-through and the collection tube.
15. Place the RNeasy MinElute spin column in a new 2 mL collection tube (provided with the kit).
16. Open the lid of the spin column and centrifuge at full speed (≈17,000 g-force) for 5 minutes at room temperature to dry the column matrices. Discard flow-through and 2 mL collection tubes.
17. Prepare TE buffer 2 and pre-warm to 60 ˚C.
18. Place the RNeasy MinElute spin column in a fresh 1.5 mL collecting tube and add 16 μL of warmed TE buffer 2 directly to the center of column matrix. Close the lid gently, and centrifuge for 1 minute at full speed to elute the RNA.
19. Pipette the RNA flow-through back into the column a second time, centrifuge at full speed for another minute.
20. Collect RNA in Nunc CryoBank cryogenic vials and store at -80 °C.
	1. **RNA quality and quantity measurement**

All the following steps describe how to determine the quantity and quality of total RNA extracted from samples sorted into TRIzol LS (*see* **Notes 21** and **22**).

1. Thaw Superscript III reagents on ice. Keep all reagents on ice.
2. Transfer 1.5 μL (10 %) of total RNA from each sample to a Plate 1.
3. Prepare Annealing mix and Superscript III reverse transcription mix for all samples and for the standard RNA sample (*see* **2.3.3**, Table 2 and 3, **Note 23**).

[Table 2]

[Table 3]

1. Dispense 3.3 µL of Annealing mix in sample wells containing the 1.5 µL of total RNA.
2. Place the plate for 5 minutes at 65 °C on a thermoblock to allow the unfolding of RNA secondary structures.
3. Immediately cool down samples using shaved ice from a -80 °C freezer for at least 1 minute before proceeding to reverse transcription. These step allows OligodT(20) primers and random hexamers to bind efficiently to unfolded RNAs.
4. Dispense 5.3 μL of Superscript III mix to both sample well and in the standard RNA well. Vortex and pulse-spin the plate.
5. Run the following program on a thermoblock to generate cDNA: 25 ˚C for 10 minutes, 50 ˚C for 50 minutes, 85 ˚C for 5 minutes, and 4 °C on hold.
6. Take samples out of the thermoblock and add 15 μL of ultrapure water to every 10 µL cDNA sample. Use 2.5 μL (10 %) of diluted cDNA for the *β2m* qPCR quantification experiment (*see* **Notes 23** and **24**).
7. Regarding the standard samples with known concentration, dilute cDNA to reach 6.4 ng/μL and prepare a 7-point 1:4 standard serial dilution. Use 2.5 μL of each dilution for a qPCR reaction in a Plate 3. (*see* Note 23).
8. Prepare Housekeeping gene qPCR mix (*see* **Notes 25** and Table 4).

[Table 4]

1. Dispense 7.5 μL of housekeeping gene qPCR mix to standards and sample wells in a Plate 3. Vortex and pulse-spin plate.
2. Perform qPCR program: 95 °C for 10 minutes; 40 cycles: 95 °C for 15 seconds, and 65 °C for 1 minute.
3. Based on Ct values, determine the RNA sample quantity and concentration and run approximately 1 ng (but no more than 10 %) of the sample on a capillary DNA-RNA electrophoresis equipment following manufacturer’s recommendations. Determine the RNA integrity number (RIN) number as well as the ratios between the 28S- and 18S-rRNA peaks (*see* **Note 26**; Fig. 2a, b).

**3.4 cDNA synthesis by high fidelity reverse transcription**

All the following steps are common for all samples independently of the collection method. All reactions are performed at room temperature unless otherwise noted. The following steps are adapted from the Smart-Seq2 method described by *Picelli et al.* [9]. For convenience, as minor changes are necessary in the procedure depending on the method chosen (bulk RNA-Seq from TRIzol LS, bulk-low input in Li-LB or single-cell RNA-Seq), we have clearly annotated these changes to the appropriate method throughout the steps (*see* **Note 27**).

1. Prepare RNA:
	1. *For samples extracted from TRIzol LS*: Into a fresh Plate 1, transfer from the RNA plate the appropriate amount of RNA to reach 1.25 ng/μL in 4 μL. Dilute with ultrapure water. If the RNA concentration is lower than the recommended 1.25 ng/uL, uniformly dilute RNA to the closest working concentration acceptable for a majority of samples (use Plate 1).
	2. *For low-input samples (all with the same number of cells):* Transfer 4 µL volumes into Plate 1.
	3. *For single-cells*: Take the plate out of -80 ˚C and centrifuge at 2,000 g-force for 5 minutes at 4 ˚C to thaw plates.
2. Prepare a fresh solution of 10 μM oligodT30-VN from a 10 µL aliquot of a 100 μM stock with ultrapure water.
3. Prepare the following Poly-T oligo mix (see Table 5):
	1. *For samples extracted from TRIzol LS and low-input samples:* Dispense 5.2 μL of oligo mix to each sample containing 4 µL of diluted RNA.
	2. *For single-cells:* add 1 µL of 10 μM oligodT30-VN directly into every well containing 4 µL of LI-LB

[Table 5]

1. Seal the plate firmly, mix well by vortexing and pulse-spin the plate to collect liquid to the bottom of wells.
2. Place the plate at 72 °C on thermomixer to allow denaturation of RNA for 3 minutes and then cool the plate down to room temperature by centrifuging the plate at 2,000 g-force for 4 minutes (*see* **Note 28**).
3. Prepare reverse transcription mix (RT-mix) (*see* **2.3.4**, Table 6).

*6.1 For samples extracted from TRIzol LS and low-input samples:* Dispense 11.2 μL of RT-mix to 9.2 μL of oligo-dT hybridized samples.

*6.2 For single-cells:* Dispense 5.6 μL of RT-mix to each sample.

[Table 6]

1. Place plate on a thermocycler and perform the following program for full-length cDNA synthesis: 42 °C for 90 minutes; 10 cycles of: 50 °C for 2 minutes and 42 °C for 2 minutes; 72 °C for 15 minutes; and 4 °C on hold.

**3.5 cDNA amplification**

1. In a Plate 3, take 3 μL (15 %) of cDNA template from each low-input RNA sample or 3 single-cell samples from each population sorted for single-cell RNA sequencing (*see* **Note 29**).
2. Prepare 10 μM oligo ISPCR primers from a 100 μM stock aliquot with ultrapure water.
3. Prepare the following ISPCR mix (*see* **Note 17***,* Table 7).

[Table 7]

1. Dispense 4.5 μL of the ISPCR mix into a 96-well plate (Plate 3) containing 3 μL of cDNA to make the final PCR volume to 7.5 μL. Seal the plates, vortex briefly and spin for 1 minutes up to 1,000 g-force at 4°C.
2. Perform qPCR cycle following this program: 98 °C for 3 minutes; 26 cycles of: 98 °C for 20 seconds, 67 °C for 15 seconds, 72 °C for 6 minutes (*see* **Note 30**).
3. Determine the number of cycles required for optimal cDNA amplification (*see* Fig. 3a): For samples extracted from TRIzol LS and low-input samples, first determine the Ct value for each sample corresponding to half of the maximum fluorescence intensity. The appropriate cycle number to be performed during amplification should be within the range of (Ct-3, Ct+1) for most samples. For single cell RNA-Seq, the appropriate number of cycles correspond when the amplification curve peaks or start to reach the plateau phase (*see* **Note 31**).
4. Prepare the Amplification mix (*see* **Note 17**,Table 8).

[Table 8]

1. Based on the qPCR analysis, perform cDNA amplification as follows:

*8.1 For samples extracted from TRIzol LS and low-input samples*: dispense 24 μL of amplification mix to the remaining 16 μL of RT product.

*8.2 For single-cells:* Dispense 15 μL of amplification mix to 10 μL RT product.

1. Seal the plates, vortex briefly and spin for 1 minute up to 1,000 g-force.
2. Perform cDNA amplification program using a thermocycler for the number of cycles determined previously by the CtD qPCR as follows: 98 °C for 4 minutes; CtD number of cycles: 98 °C for 20 seconds, 67 °C for 15 seconds, 72 °C for 6 minutes; 72 °C for 15 minutes, and 4 °C on hold (*see* **Note 32**).

**3.6 cDNA size selection**

The following steps describe cDNA fragment purification protocol using the AmpureXP beads approach.

1. Allow the Agencourt Ampure XP beads solution (stored at 4 ˚C) to equilibrate to room temperature, and vortex thoroughly to homogenize well.
2. Equilibrate the cDNA samples to room temperature before proceeding with the Ampure XP clean-up steps.
3. Prepare the 80 % ethanol washing solution (*see* **Note 33**).
4. Dispense 0.9:1 (vol:vol) of AmpureXP beads to each well with a cDNA sample using a repeat pipetter.
5. Mix carefully with a pipette, seal the plate and let it sit at room temperature for 2 minutes.
6. Place the plate on the Axygen 96-well magnet. Capture beads for 5 minutes and then discard supernatant with a 200ul pipetter.
7. Maintaining the plate on the magnet, add 200 µL of 80 % ethanol washing solution to each well. Discard supernatant without disturbing the pellet by pipetting.
8. Repeat washing step twice.
9. Remove all 80 % ethanol washing solution on the last wash (*see* **Note 34**).
10. Leave the beads to dry for about 10 minutes until cracks appear on the pellets (*see* **Note 35**).
11. Pre-warm TE buffer 3 at 60 ˚C.
12. With plate still on the magnet, elute DNA by dispensing 24 µL (16 µL for single cells) of pre-warmed TE buffer 2 to each well.
13. Seal plate and take out the plate from the magnet.
14. Vortex plate thoroughly (avoid spill over on the seal) and briefly pulse-spin plate to 1,000 g-force.
15. Place plate on the magnet for bead capture and transfer the supernatant to a new full skirted 96 well-PCR plate.
16. Repeat the cDNA size selection protocol a second time for single-cell samples (*see* **Note 36**).
	1. **cDNA Quantification**

Quantification of cDNA is performed using the PicoGreen assay. The fluorometric DNA quantification assay PicoGreen dsDNA enables highly sensitive measurement of amplified cDNA and final DNA libraries (*see* **Note 37**).

1. Take an approximate volume of TE buffer 3 corresponding to 0.2 mL x number of samples. For example, 48 samples will require preparation of roughly 10 mL of TE buffer 3.
2. To prepare the 1:4 serial dilution of standard DNA, calculate the amount of sonicated bacteriophage lambda DNA (40 ng/µL): 4.4 μl (160 ng) of bacteriophage lambda DNA for each plate.
3. Dilute the bacteriophage lambda DNA to 0.4 ng/μL with 39.6 µL of TE buffer 3 for the top standard, and prepare a serial 4-fold dilution (11 in 33 µL) for 6 consecutive standard points (for practical purposes, use an 8-tube strip).
4. Make a 10-fold sample dilution by adding 3 μL of purified product to 27 μL of TE buffer 3.
5. Use TE buffer 3 neat as the blank sample.
6. Prepare PicoGreen solution (*see* **2.3.7**).
7. Fill up a Plate 4 with 190 μL PicoGreen solution (*see* Fig. 3b).
8. In each plate, load 10 μL of 1:4 standard dilution series and blank in triplicate.
9. Load 10 μL of diluted samples in triplicate (*see***step 4**).
10. Read and record data using a top-read 96-well plate reader.
11. Determine sample concentrations and total quantities using the standard curve constructed with fluorescence values measured for the bacteriophage lambda DNA dilutions. Subtract blank values (*see* **Note 38**)
12. Based on quantity values, run approximately 1 to 5 ng of samples (no more than 10 % of total) on a capillary DNA-RNA electrophoresis equipment following manufacturer recommendations and determine the median size of cDNA fragments as well as the fraction of fragments with a length shorter than 200 bp (*see* **Note 39** andFig. 3c).

**3.8 cDNA library preparation**

Integration of sequencing adaptors is performed by “tagmentation” using the tn5 DNA insertion method commercialized by Illumina (Nextera XT reagents and barcoded adaptors kits). The following sequence of steps describes the method for library preparation using the pre-cited kits (*see* **Note 40**).

1. Prepare tagmentation mix (*see* **Note 17** and Table 9).

[Table 9]

1. Prepare the ***dilution plate*** using a Plate 2 on ice, diluting samples with ultrapure water to reach:

*3.1 For samples extracted from TRIzol LS and low-input samples:* at least 5 µL of cDNA at 0.2 ng/μL.

*3.2 For single-cell samples:* at least 5 µL of cDNA at 0.5 ng/μL.

1. Prepare a separate ***tagmentation plate*** using Plate 1 on ice, dispensing:

*4.1 For sample extracted from TRIzol LS and low-input samples:* 15 µL of tagmentation mix.

*4.2 For single-cell samples*: 2.4 µL of tagmentation mix*.*

1. On ice, transfer the diluted cDNA samples from the ***dilution plate*** to the ***tagmentation plate*** containing the tagmentation mix as follows:

 5.1 *For sample extracted from TRIzol LS and low-input samples:* 5 µL of samples diluted at 0.2 ng/µL (final quantity is 1 ng of cDNA sample).

*5.2 For single-cell samples*: 0.8 µL of sample diluted at 0.5 ng/µL (final quantity is 0.4 ng of cDNA sample).

1. Seal plates, and promptly incubate samples at 55 °C for exactly 10 minutes.
2. To stop the tagmentation reaction, quickly proceed as follows:

*7.1 For sample extracted from TRIzol LS and low-input samples*: Add 5 μL of the NT buffer to strip off the Tn5 enzyme. Vortex and quickly centrifuge to 1,000 g-force at room temperature.

*7.2 For single-cell samples*: add 0.8 μL of NT buffer. Vortex and quickly centrifuge plate to 1,000 g-force at room temperature.

1. Add the Nextera XT PCR mix as follows:

*8.1 For samples extracted from TRIzol LS and low-input samples*: Add to the tagmented samples, sequentially, 15 µL of Nextera XT PCR master mix and 5 µL each of Illumina N7xx and S5xx primers (*see* **Note 41**).

*8.2 For single-cell samples:* Combine in a separate Plate 1: 2.4 µL Nextera XT PCR master mix with index combination containing 0.8 µL each of Illumina N7xx and S5xx primers. Transfer 4 µL of mix to tagmented samples (*see* **Note 42**).

1. Seal plates and vortex with care. Spin briefly up to 1,000 g-force at 4°C.
2. Using a thermocycler, perform the following program to generate Illumina libraries for sequencing: 72 °C for 3 minutes, 95 °C for 30 seconds; number of cycles: 95 °C for 10 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds; 72 °C for 5 minutes; and 4°C on hold (*see* **Note 43)**.
3. For single-cell samples only: add 16 μL of water per sample after library preparation.
4. From this point, perform Ampure XP bead clean-up (*see* **3.6**), using a 1:1 volume ratio of Ampure XP beads to sample.
5. Quantify 1:10 dilutions of the cDNA using PicoGreen dye (*see* **3.7**, **Note 44**).
6. Run approximately 1 to 5 ng of samples (no more than 10 % of total) on a capillary DNA-RNA electrophoresis equipment following manufacturer recommendations and determine either the median size of cDNA fragments and the fraction of fragments with a length shorter than 200 bp or longer than 1,000 bp (*see* **Note 45**).
7. For each library determine the molar concentration using this formula:

*[Library concentration (ng/µl) \* 106] / [660 \* Median size (bp)] = Library Molarity (nM)*

1. Libraries are now ready to be pooled in equimolar amount and sequenced following standard procedures on Illumina sequencing platforms (*see* Fig.5 and**Note 46**).

**4 Notes**

1. All stock solutions (\*) are prepared with ultrapure water and autoclaved prior to use in experimental buffers. For practical purposes, all constituents of buffers in mixes are given as final concentration.
2. RNase away solution may cause irritation and redness if it contacts the skin. Avoid working with this solution without proper personal protection equipment.
3. These PCR plates have wells with higher volume capacity, up to 0.35 mL.
4. We used those plates to perform qPCR experiments with the Applied Biosystems™ QuantStudio™ 6 Flex Real-Time PCR System. The qPCR MicroAmp optical adhesive films are only used for qPCR experiment, otherwise we recommend to use MicroAmp Clear Adhesive Films.
5. Take caution when working with TRIzol LS as it contains phenol, guanidine isothiocyanate, and other health hazardous chemicals that are corrosive/toxic. Always use TRIzol LS reagent in a chemical fume hood to limit exposure.
6. Autoclave the lysis buffer (LI-IB) and keep sterile at 4 ˚C for up to 12 months.
7. Components are part of the Superscript III kit.
8. Prepare aliquots of 5 µL, store at -80 ˚C and avoid freeze/thaw cycles.
9. Primers are usually delivered in lyophilized form. Upon reception, dissolve and make a 100 µM stock solution with TE Buffer 1. Then prepare a working stock solution at 5 µM with TE buffer 2. Keep at -20 ˚C.
10. Make 1M betaine aliquots and store at -20 ˚C.
11. Sonicate Lambda DNA in order to obtain a majority of fragments within length ranging between 300 bp and 800 bp. Adjust concentration with TE buffer 1, aliquot and store at -20°C.
12. Prepare solution fresh at room temperature and avoid exposure to light.
13. Regarding sample collection, it is critical to sort cell(s) directly into the appropriate lysis buffer (TRIzol LS or complete LI-LB) as promptly as possible at 4 ˚C. Make sure the sort stream is well calibrated such that cells are hitting the liquid interface of the lysis solution and not the side of the tube. It is also important to ensure that cells and nuclei are thoroughly lysed before storing samples at -80 ̊C. Due to the complexity of the cell sorting, the number and relative abundance of cells, sample collection time can be quite long. We recommend running a "dry" experiment to finalize all details of sample collection, including harvesting of tissue, preparation of cell suspensions, staining, and sorting to avoid any delays. It is crucial to prevent RNA degradation by optimizing the sorting time to its maximum efficiency.
14. We detect significant RNA degradation if cells are not properly lysed and if they lay on the top of the TRIzol buffer for too long. It is crucial to calibrate the cell sorter in such a way that cells will be sorted directly into lysis buffer.
15. For all samples collected in TRIzol LS, we suggest splitting the volume into two RNase-free individually wrapped 1.5 mL tubes containing 500 μL aliquots. One aliquot will serve as safety backup aliquot in case of failed amplification. Avoid multiple freeze-thaw cycles for all samples.
16. Standard sorting strategies to recapitulate cell heterogeneity in low input samples will require cell to volume ratios that are limited to 200 to 400 cells in 8 μL of lysis buffer or, if possible, by sorting 1,000 cells in 20 μL of lysis buffer directly into 0.2 mL PCR tubes. We recommend sorting at a minimum, 200 cells in triplicates or 1,000 cells in duplicates. It is crucial to calibrate the cell sorter in such a way that cells will be sorted directly into lysis buffer.
17. When preparing mixes, we increase reagent volume by 5 % to account for dispensing volumes error.
18. We use the Qiagen miRNAeasy Micro kits for consistent extraction of pure, concentrated RNA for downstream RNA-SmartSeq input requirements. In our hands, the MinElute columns have enabled purification of RNA ranging from 200,000 to as low as 2,500 cells. For these results, we perform a double elution of RNA eluted into 16 μL. We estimate the procedure to take 3 hours for up to 16 samples.
19. When transferring the upper aqueous avoid transferring any interphase as this could lead to downstream RNA degradation.
20. Do not spin the samples at this step, as the RNA will precipitate and will be difficult to load on to columns.
21. When preparing samples for low input RNA-Seq or single-cell RNA-Seq, we bypass the RIN determination because estimated quantities are outside the sensitivity of our equipment. The fact that we sort an equal number of cells allows us to also bypass the quantification steps.
22. RNA has to be accurately quantified to successfully proceed to the downstream steps. It is important to start the reverse transcription and cDNA amplification reactions with a similar amount of RNA. We have designed a method of quantification by qPCR [11] using a standard with known amount of RNA(*see* **Note 24**). Currently, we use Superscript III kit reagents and the manufacturer’s protocol with slight modifications. For ease of reference we made a table listing the reagent volumes needed for 48 RNA samples (*see* Table 2 and 3).
23. Consider using a non-essential RNA sample extracted from a cell population very similar to that of the cell population of interest as standard, which has been accurately quantified by standard spectrometry methods. For instance, for our T cell work, we prepared a good quality CD3+ T cell RNA aliquot for the standard dilution with an initial concentration greater than 25 ng/μL. Use 3 μL of RNA for annealing mix. After reverse transcription, we dilute RNA in TE buffer 2 based on the initial quantity to get 6.4 ng/μL in 20 µL as top standard. We estimate RNA quantification qPCR and analysis will take approximately 3 hours.
24. The remaining Superscript III synthesized cDNA could be used to detect expression of other genes of interest.
25. β2-microglobulin (B2m) serves as a housekeeping gene that could be used for quantification of T cells.
26. Calculate total RNA quantities based on qPCR data, and determine RIN for every RNA sample. Regarding quantity, one primary T cell typically has between 0.5 to 1 pg of total RNA, while more metabolically active cells or cell lines with of larger size will have up to 10 pg of total RNA/cell. If RNA quantity is less than 20% or more than 200% of expected quantities, try to determine the underlying cause of the noted difference. Consider eliminating samples if the RIN is less than 7.5 and if the 28S/18S ratio is less than 1 (Fig 2a, b).
27. We estimate that cDNA synthesis by high fidelity reverse transcription for up to 48 or 96 RNA samples will take approximately 3.5 hours.
28. Denaturation of RNA at 72 °C will allow the binding of Poly-T oligos to bind to Poly-A mRNA tail templates.
29. Once the reverse transcription is completed, given the extremely low amount of cDNA present, we recommend directly proceeding with the cDNA amplification to preserve the integrity of samples. However, it is important to amplify cDNA to the correct amount, as too little or too much amplification will compromise the quality of the downstream steps and could lead to saturation of reagent or loss of diversity (jackpotting). We recommend initially performing a real-time quantitative PCR to optimally determine the number of cycles of amplification to decrease risk of PCR biases and omit steps afterward when satisfied with amplification results. This experiment is called CtD qPCR (Ct threshold determination for cDNA amplification). Typically for bulk RNA-seq the CtD for 5ng is 15 cycles and for single cell the CtD ranges from 19-22. The CtD qPCR program requires about 4 hours.
30. Including a melt-curve in the qPCR program for essential detection of unspecific amplification.
31. The number of cycles chosen should correspond to an intensity present in the linear range of the PCR amplification (avoid going beyond the peak of intensity and plateau phase).
32. After cDNA amplification program samples can be stored at -20 ˚C for up to a week.
33. Prepare the 80 % washing solution fresh before starting the cDNA Ampure cleanup and Nextera XT library prep cleanup. For the bulk cDNA sample Ampure clean up prepare 600 μL for each sample to perform three washes; for single cell samples prepare twice as much for second round of washes. For Nextera XT steps prepare 400 μL for a total of two washes.
34. Use 20 µL tips to remove small drops of leftover washing buffer.
35. Do not over-dry beads, otherwise pellets will be difficult to resuspend.
36. For single cells, the required number of cDNA amplification cycles generates significant amounts of small fragment side products that may interfere with the tagmentation efficiency if not removed with a second purification step.
37. The assay takes 1 hour to prepare and quantitate a 96-well plate.
38. The expected cDNA yield is approximately 50 ng when starting with 5 ng of RNA from T lymphocytes in resting conditions. For low input RNA samples (200 cells) and single-cell samples expect approximately 10 ng.
39. Samples with more than 10 % of low molecular weight cDNA (<500 bp) should be eliminated as it is indicative of RNA degradation. Samples can be stored at -20 ˚C for a week or more, however we recommend to continue the downstream procedure within 72 hours.
40. We estimate it will take 4 hours to perform all steps manually. It is important to keep all tagmentation reagents at 4 °C. It is recommended to keep the NT buffer readily available before starting the tagmentation incubation, as these reactions are time sensitive.
41. Final PCR volume is 25 μL.
42. Final PCR volume is 8 μL.
43. We use 9 cycles for sample extracted from TRIzol LS and Low-Input and 12 cycles for single-cell samples.
44. The expected total yield is between 50 ng to 150 ng (see Fig. 4).
45. Based on this information, samples with less than 85 % of fragments size ranging between 200 to 1,000 bp will have to go through another round of clean-up or will be eliminated (see Fig. 4). The presence of low size fragments or free adapters will increase risks of index misalignment and reduce the quality of sequencing [12,13].
46. We usually perform 50 bp single-end read sequencing and aim to generate 15 million mapping reads per sample. This protocol achieves greater than 80 % mapping to reference genome and uniquely mapped reads. As an example, Fig. 5a shows the fraction of reads that map with coding, intronic, and intergenic regions. Figs. 5b,c show example of consistency of RNA-Seq results achieved with this procedure. It illustrates read profiles distribution along different gene loci for two independent sets of bulk RNA-Seq libraries for CD4 T cells (b) and 20 CD8 T single-cell RNA-Seq (at higher resolution). UCSC genome browser tracks; Rpkm, reads per kilobase per million mapped.

REFERENCES

1. Costa V, Aprile M, Esposito R, Ciccodicola A (2013) RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet 21 (2):134-142. doi:10.1038/ejhg.2012.129

2. Seumois G, Zapardiel-Gonzalo J, White B, Singh D, Schulten V, Dillon M, Hinz D, Broide DH, Sette A, Peters B, Vijayanand P (2016) Transcriptional Profiling of Th2 Cells Identifies Pathogenic Features Associated with Asthma. J Immunol 197 (2):655-664. doi:10.4049/jimmunol.1600397

3. Engel I, Seumois G, Chavez L, Samaniego-Castruita D, White B, Chawla A, Mock D, Vijayanand P, Kronenberg M (2016) Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. Nat Immunol 17 (6):728-739. doi:10.1038/ni.3437

4. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW (2016) Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet 17 (5):257-271. doi:10.1038/nrg.2016.10

5. Zhao Y, Simon R (2010) Gene expression deconvolution in clinical samples. Genome Med 2 (12):93. doi:10.1186/gm214

6. Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, van den Berg LH, Ophoff RA, Horvath S (2010) Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11:589. doi:10.1186/1471-2164-11-589

7. Mohamad-Fauzi N, Ross PJ, Maga EA, Murray JD (2015) Impact of source tissue and ex vivo expansion on the characterization of goat mesenchymal stem cells. J Anim Sci Biotechnol 6 (1):1. doi:10.1186/2049-1891-6-1

8. Proserpio V, Lonnberg T (2016) Cutting-edge single-cell genomics and modelling in immunology. Immunol Cell Biol 94 (3):224. doi:10.1038/icb.2015.117

9. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9 (1):171-181. doi:10.1038/nprot.2014.006

10. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W (2017) Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell 65 (4):631-643 e634. doi:10.1016/j.molcel.2017.01.023

11. Seumois G, Vijayanand P, Eisley CJ, Omran N, Kalinke L, North M, Ganesan AP, Simpson LJ, Hunkapiller N, Moltzahn F, Woodruff PG, Fahy JV, Erle DJ, Djukanovic R, Blelloch R, Ansel KM (2012) An integrated nano-scale approach to profile miRNAs in limited clinical samples. Am J Clin Exp Immunol 1 (2):70-89

12. Wright ES, Vetsigian KH (2016) Quality filtering of Illumina index reads mitigates sample cross-talk. BMC Genomics 17 (1):876. doi:10.1186/s12864-016-3217-x

13. Illumina (2017) Effects of Index Misassignment on Multiplexing and Downstream Analysis. Illumina Technical Note

**Fig. 1. Overview of the method.** The different sorting methods depending on the sample availability for bulk or single-cell RNA-Seq are indicated on the top of the schematic. All the major steps of the procedure including quality control steps and timing are shown on the left for easy planning of the experiment. In parallel with the procedure, we display a schematic of the molecular process.

**Fig. 2. RNA extraction quality and quantity controls.** (a) Bioanalyzer traces of TotRNA sample profiles for high-quality RNA with a clear presence of rRNAs peaks (top) and a degraded RNA sample with absence of the 28S rRNA peak (bottom). RIN and 28S/18S ratios for both samples and the suggested threshold values for the elimination of samples are shown. (b) Correlation plot between the number of cells sorted into Trizol LS and the TotRNA quantity (ng) estimated by qPCR for B2m house-keeping gene. The dotted line indicates the 5,000 cells threshold value.

**Fig. 3. cDNA amplification.** (a) Interpretation of the qPCR amplification curves to determine the amplification number of cycles for each sample after reverse transcription. For bulk-RNA-Seq samples (Trizol LS or LI-LB), the number of cycles is equal to the Ct value corresponding to mid-fluorescence intensity minus 2 or 3 depending on the other samples ran concomitantly. For the single cell RNA-Seq, the number of cycles corresponds to the Ct value for the peak of the amplification curve. (b) Picogreen assay plate layout. (c) cDNA profile of size distribution after amplification. A standard well-amplified sample and a poorly amplified sample characterized by low molecular weight peaks are shown. Every sample with more than 5 % of fragments with less than 200 bp is eliminated.

**Fig. 4. Final library quality and quantity controls.** (a) Fragment size distribution for Bulk RNA-Seq and single-cell RNA-Seq libraries. The top graph shows good fragment distributions with a majority (> 80 %)of fragments ranging in size between 200 to 1000 bp. The bottom graphs show unfit library size distribution; those samples would be eliminated. Numbers indicate our threshold values for elimination. (b) Yields quantity of libraries for Bulk RNA-Seq libraries (left) and single-cell RNA-Seq (right). Values show a great consistency of yield between samples.

**Fig. 5. Quality control post-sequencing.** (a) The mappability of around 10 to 15 million single-ended 50 bp sequencing reads to the reference genome (hg19) for two independent sets of bulk RNA-Seq libraries (left) and single-cell RNA-Seq (right). Percentage rates shown are for a total number of reads (total), for reads mapping only to one and unique genomic location, reads mapping to coding regions (mRNA), reads mapping to intronic regions, reads mapping to intergenic regions compared to the total number of reads passing sequencing filters. The red dashed lines indicate the threshold values for elimination. (b & c) UCSC genome browser tracks of examples of the sequencing read profiles along different gene loci for two independent sets Bulk RNA-Seq libraries for CD4 T cells (b) and 20 CD8 T cells single-cell RNA-Seq (at higher resolution)(c). Rpkm, reads per kilobase per million mapped.