Full-length RNA-seq from single cells using Smart-seq2

Simone Picelli[1], Omid R Faridani[1], Åsa K Björklund[1,2], Gösta Winberg[1,2], Sven Sagasser[1,2] & Rickard Sandberg[1,2]

[1]Ludwig Institute for Cancer Research, Stockholm, Sweden. [2]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden. Correspondence should be addressed to R.S. (rickard.sandberg@ki.se).

Published online 2 January 2014; doi:10.1038/nprot.2014.006

Emerging methods for the accurate quantification of gene expression in individual cells hold promise for revealing the extent, function and origins of cell-to-cell variability. Different high-throughput methods for single-cell RNA-seq have been introduced that vary in coverage, sensitivity and multiplexing ability. We recently introduced Smart-seq for transcriptome analysis from single cells, and we subsequently optimized the method for improved sensitivity, accuracy and full-length coverage across transcripts. Here we present a detailed protocol for Smart-seq2 that allows the generation of full-length cDNA and sequencing libraries by using standard reagents. The entire protocol takes ~2 d from cell picking to having a final library ready for sequencing; sequencing will require an additional 1–3 d depending on the strategy and sequencer. The current limitations are the lack of strand specificity and the inability to detect nonpolyadenylated (polyA$^-$) RNA.

INTRODUCTION

It is becoming increasingly apparent that seemingly homogeneous cell populations *in vivo* or cell cultures *in vitro* can show considerable heterogeneity in expression patterns[1], owing to both intrinsic stochastic processes and extrinsic factors, such as the surrounding microenvironment. Although high-throughput technologies, such as microarray and, more recently, RNA-seq, have given us a better understanding of complex eukaryotic transcriptomes (e.g., see refs. 2–4), the RNA input requirements have limited analyses to complex mixtures of often tens to hundreds of thousands of cells. As the results from expression analyses over tissues or large cell populations reflect the cell population average, it does not inform on cellular variability nor does it allow for the discovery of cellular subtypes. Instead, a single-cell resolution is needed to increase our understanding of cell-to-cell variability.

During the past few years, several groups have developed new sequencing-based methods for single-cell transcriptome analysis[5–13]. Our group has recently demonstrated that Smart-seq is able to generate quantitative and reproducible data from both single cells and small amounts of purified RNA[8]. With Smart-seq2, we further refined reverse transcription (RT), template switching and preamplification to obtain an increased cDNA yield from single cells, as well as higher sensitivity, fewer technical biases and less variability[12]. Here we present a detailed protocol for Smart-seq2 (ref. 12) that entirely relies on off-the-shelf reagents.

Overview of the protocol

This protocol was developed for single cells but works equally well on purified total RNA of a few picograms or more. A flowchart of the protocol is provided in **Figure 1**. Cells should first be disaggregated into a single-cell suspension, and you should aim for rapid processing at near-physiological conditions. A lengthy procedure can lead to unwanted alterations in gene expression or, worse, to the death of a large fraction of the cell population. Therefore, cells should be picked promptly (in <30 min if possible) and immediately placed in lysis buffer containing a ribonuclease inhibitor that blocks RNA degradation and stabilizes the RNA. When working with precious samples, manual cell picking with micro capillary pipettes is probably the best available option, although automated or semiautomated commercial solutions are emerging. Tubes or plates containing single cells should be kept on ice or on an IsoFreeze PCR rack during collection and stored at −80 °C if not processed immediately.

In this protocol we use a relatively mild (hypotonic) lysis buffer[8] capable of lysing cells without interfering with or inhibiting the RT reaction. The lysis solution also contains free dNTPs and tailed oligo-dT oligonucleotides (30-nt poly-dT stretch and a 25-nt universal 5′ anchor sequence) that are responsible for the priming of the RT reaction on polyadenylated RNA sequences. The free dNTPs are added in the initial step to improve the yield of RT-PCRs[12,14], probably through mechanisms that stabilize RNA-primer hybridizations.

The RT reaction is normally performed at 42 °C for 90 min when using commercial Smart-seq. However, some RNAs form secondary structures (such as hairpins or loops) that might cause the enzyme to terminate chain elongation owing to steric hindrance. This undesired effect[15–17] can, to some extent, be overcome by the addition of trehalose and betaine (N,N,N-trimethylglycine)[18,19]. Although some studies have shown that trehalose without other additives, or in combination with betaine, is capable of increasing cDNA yield[18,20], we obtained better results with betaine alone[12]. Therefore, our protocol uses betaine, a methyl group donor with two important effects: it increases the thermal stability of proteins and reduces or even eliminates the base pair composition dependence of DNA thermal melting transitions by destabilizing the DNA helix[21–23]. It was also shown that betaine is more effective in destabilizing G-C pairs than A-T pairs[23]. Although its beneficial effect on RT was established several years ago, it was not until recently that it was shown to improve the yield of full-length cDNA when using template-switching oligos (TSOs)[20].

We also found that the increase in cDNA yield after adding betaine (1 M) required an increased concentration of magnesium chloride[12]. Mg^{2+} forms ion pairs with the carboxylate anion of the betaine molecule, which under physiological conditions is actually a stabilizing osmolyte that helps the cell respond to

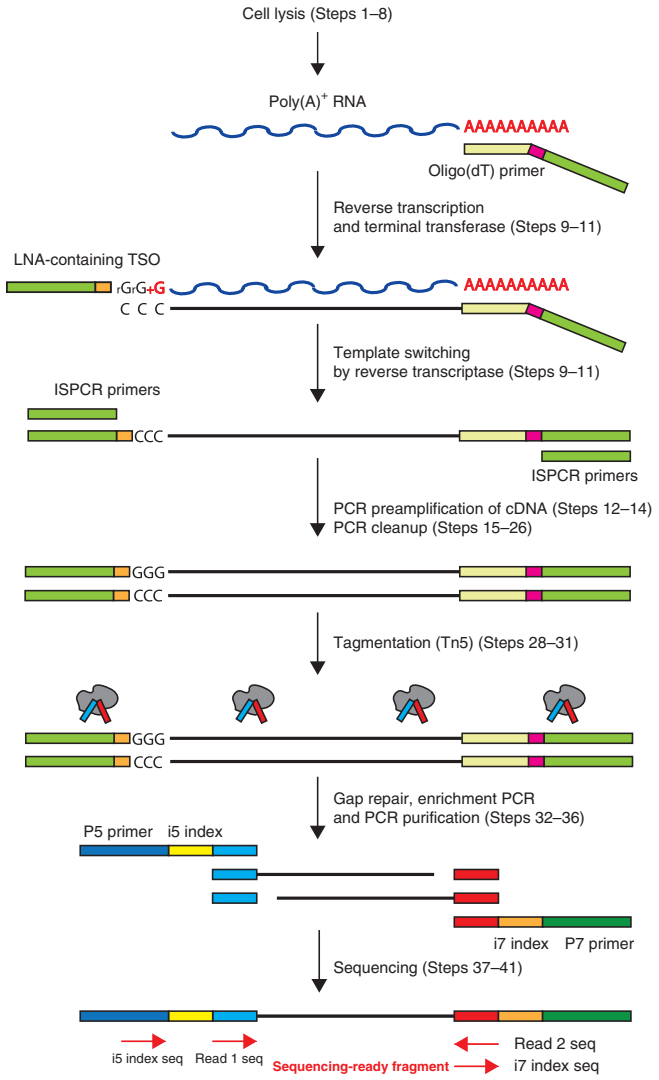

Figure 1 | Flowchart for Smart-seq2 library preparation. Outline of the protocol and the corresponding procedure steps. The oligo-dT primer, TSO and ISPCR primer are described in the main text, whereas tagmentation uses primers that are included in the Nextera XT sample preparation and index kits.

changes in osmotic pressure. A recent study showed that betaine becomes a DNA-destabilizing agent above 1 mM $MgCl_2$ (ref. 24). In agreement with that study, we found that a final concentration of 9–12 mM $MgCl_2$ was necessary for maximizing cDNA yield[12]. Although it has been reported that too much $MgCl_2$ has a negative effect on the fidelity of the PCR[25], we did not observe any negative influence either for the Superscript II (Invitrogen) used in the RT or the KAPA HiFi DNA polymerase used in the PCR.

Through the addition of betaine, we could also exploit its protein thermal-stabilization properties to modify the temperature at which the RT is performed. After 90 min at 42 °C (the optimal conditions for Superscript II, according to the manufacturer), we raised the temperature to 50 °C (the upper limit before inactivation) for 2 min to promote the unfolding of any RNA secondary structures. We then lowered it to 42 °C again for 2 min to allow the completion of the RT and repeated this cycle a total of ten times to maximize cDNA yield. We chose ten cycles as suggested by Carninci *et al.*[18], even though that experiment used no TSOs. Incidentally, between 10 and 15 cycles appeared to maximize the yield with our setup as well (data not shown).

The template-switching reaction relies on 2–5 untemplated nucleotides that are added to the cDNA 3′ end when the RT reaches the 5′ end of the RNA. In its original version, the reaction contained a TSO carrying three riboguanosines at its 3′ end[15,26]. The Moloney murine leukemia (M-MLV) reverse transcriptase is thus able to switch template and synthesize a complementary sequence to the TSO. Every full-length cDNA carries the entire 5′ end of the transcript and an additional artificial sequence, which in this case is the same as the one located at the 5′ end of the oligo-dT primer (i.e., the 3′-end of the mRNA), making possible the subsequent PCR using a single primer.

Smart-seq2 uses a TSO carrying two riboguanosines in the third- and second-last positions and a modified guanosine to produce a locked nucleic acid (LNA) as the last base at the 3′ end. These locked nucleotides are characterized by an internal bond between the O2′ and the C4′ of the furanose ring, linked by a methylene group. The modification introduces a conformational lock in the molecule that, however, still retains the physical properties of the native nucleic acids. Two interesting properties of LNAs are likely to be advantageous for our protocol: the enhanced thermal stability of the LNA monomers; and their ability to anneal strongly to the untemplated 3′ extension of the cDNA[27].

After the first-strand reaction, the cDNA is amplified using a limited number of cycles, usually 18 when working with single cells or just as many as needed to get enough material for the following steps (one or few nanograms of cDNA are sufficient). To simplify the protocol, to improve sensitivity and to adapt for liquid handling automation, we eliminated the bead (AMPure XP)-mediated purification of first-strand cDNA before the PCR. This was made possible by the use of the KAPA HiFi HotStart ReadyMix, which replaced the Advantage 2 polymerase mix used in the SMARTer protocol. Incidentally, KAPA HiFi was recently tested in the sequencing of bacterial genomes and found to introduce the least bias when compared with unamplified samples[28].

We used tagmentation to quickly and efficiently construct sequencing libraries from the amplified cDNA. The tagmentation reaction takes advantage of a hyperactive derivative of the Tn5 transposase that catalyzes *in vitro* integration of predetermined oligonucleotides into target DNA[29]. The main advantage of this approach is that DNA fragmentation and adapter ligation occur in a single step, and size selection is not necessary. A comparison of sequence composition surrounding the fragmentation sites identified a bias signature that was stronger than that associated with mechanical shearing methods[29]. Nonetheless, this was found to have little effect on the coverage for the human genome. Thus, we could reasonably assume that the fragments generated after digestion of full-length cDNA evenly cover the entire length of the transcript[5].

The fragments in the tagmented DNA library have an average size that usually ranges from 200 to 600 bp and are ready for enrichment PCR. We followed the protocol developed by Illumina that allows the pooling of up to 96 samples through the use of a dual-index strategy, referred to as index 1 (i7) and index 2 (i5). Samples were pooled after enrichment PCR and sequenced together on the same lane of the Illumina sequencers.

Strengths and limitations of the protocol

This method has several advantages, briefly summarized below:

- It gives the possibility of analyzing hundreds of cells for a fraction of the cost of currently available commercial kits. Compared with the widely used SMARTer kit, our protocol can generate improved-quality libraries for ~12% of the cost.
- It allows the recovery of full-length cDNAs, owing to the preference of M-MLV reverse transcriptase for full-length over truncated cDNAs as a substrate for its terminal transferase activity. It is therefore possible to analyze all the exons of each transcript and to detect the different splice variants, a big advantage over previous methods[5–11,13]. It also enables comprehensive SNP and mutation analysis, widening its field of application.
- It allows a high degree of multiplexing; up to 96 samples can be pooled and sequenced on a single lane of an Illumina sequencer.

However, some limitations remain:

- This method is selective for polyadenylated RNA, precluding analyses of polyA$^-$ RNA.
- The reads do not reflect the strand-specific nature of mRNAs.
- Despite high multiplexing at sequencing, the pooling of samples is performed after the adapter-ligated enrichment PCR step. The relatively high manual labor compared with early multiplexing methods (e.g., CEL-seq[9] or STRT[7,10]) can be decreased by the adaptation of the protocol for automated liquid handling platforms (such as the Agilent Bravo) or microfluidics solutions (such as the Fluidigm C1).

Future applications

Numerous applications are made feasible because of the reduction in cost per sample and the increase in throughput, including (but not limited to) the following:

- Isolation of single cells from different regions of a primary tumor to gain information on cellular heterogeneity. It has been shown that solid tumors are often composed of multiple clonal subpopulations[30–32], which confounds clinical analysis of bulk RNA or DNA samples from hundreds of thousands of cells. The gene expression profiles from such mixed populations of tumor cells may be strongly biased by the contributions from different cell populations within the tumor, making it impossible to discern the characteristics that determine the malignant phenotype[33]. Moreover, preliminary exome-sequencing studies at the single-cell level have revealed that solid tumors and hematopoietic tumors can show a very different clonal origin[34,35]. It is reasonable to assume that larger numbers of single cells from different tumor regions need to be analyzed to better understand tumor evolution and progression.
- The study of the different cell types that constitute a developing embryo, a tissue or a whole animal, with the goal of establishing the developmental hierarchy of stem cells, progenitors and differentiated cells.

MATERIALS

REAGENTS

- Dulbecco's phosphate-buffered saline (DPBS), no calcium, no magnesium (Gibco, cat. no. 14190-144)
- 1× TrypLE Express enzyme, no phenol red (Gibco, cat. no. 12604-021)
- RNaseZap (Ambion, cat. no. AM9780)
- DNA-OFF (Takara Bio, cat. no. 9036)
- Triton X-100 (Sigma-Aldrich, cat. no. T9284) **! CAUTION** Triton X-100 is harmful if swallowed. It causes serious eye damage. Handle it using appropriate safety equipment.
- dNTP mix (10 mM each; Fermentas, cat. no. R0192)
- First-strand buffer (5×; 250 mM Tris-HCl, pH 8.3, at room temperature (25 °C); 375 mM KCl; 15 mM $MgCl_2$; Invitrogen, cat. no. 18064-014)
- DTT (Invitrogen, cat. no. 18064-014) **! CAUTION** DTT is toxic when ingested. Avoid inhaling fumes or contact with the skin. Handle it using appropriate safety equipment.
- Superscript II reverse transcriptase (Invitrogen, cat. no. 18064-014)
- Recombinant RNase inhibitor (Clontech, cat. no. 2313A)
- Betaine (BioUltra ≥99.0%; Sigma-Aldrich, cat. no. 61962)
- Magnesium chloride ($MgCl_2$; anhydrous; Sigma-Aldrich, cat. no. M8266)
- Distilled water (Gibco, cat. no. 10977)
- KAPA HiFi HotStart ReadyMix (2×; KAPA Biosystems, cat. no. KK2601) **▲ CRITICAL** A HotStart DNA polymerase is necessary to minimize the background amplification when working with single cells and is more practical when working with automated liquid-handling platforms.
- Agencourt Ampure XP beads (Beckman Coulter, cat. no. A 63881)
- Ethanol 99.5% (vol/vol); Kemethyl, cat. no. SN366915-06) **! CAUTION** It is flammable; handle it using appropriate safety equipment.
- EB solution (10 mM Tris-Cl, pH 8.5; Qiagen, cat. no. 19086)
- TruSeq dual-index sequencing primer kit for single-read runs (Illumina, cat. no. FC-121-1003) or paired-end runs (Illumina, cat. no. PE-121-1003)
- Nextera XT DNA sample preparation kit, 96 samples (Illumina, cat. no. FC-131-1096)
- Nextera XT 24-index kit, 96 samples (Illumina, cat. no. FC-131-1001)
- Adapter oligos (See Reagent Setup). All oligos except the LNA-modified TSO were ordered from commercial vendor Biomers.net (http://www.biomers.net/) and were subjected to HPLC purification. LNA-modified TSO was ordered from Exiqon (http://www.exiqon.com/).
- CASAVA 1.8.2 (Illumina) software, freely available for the Linux environment

EQUIPMENT

- Center-well organ culture dish, 60 × 15 mm polystyrene dish (Corning, cat. no. 353037)
- Microcentrifuge tubes, polyallomer (Beckman Coulter, cat. no. 357448)
- Microcentrifuge Safe-Lock tubes, polypropylene (Eppendorf, cat. no. 0030 120.086)
- Falcon polystyrene conical tube (50 ml, BD Biosciences, cat. no. 352095)
- Mini vortexer (VWR, cat. no. 82019-170)
- Thermal cycler (S1000, Bio-Rad)
- Magnetic stand 96 (Ambion, cat. no. AM10027)
- 8-strip, nuclease-free, 0.2-ml, thin-walled PCR tubes with caps (Sarstedt, cat. nos. 72.985.002 and 65.989.002)
- V-bottom plates (96 well; VWR, cat. no. 47743-996)
- Filter tips: 10, 20, 100 and 200 µl (Gilson, cat. nos. F171203, F171303, F171403 and F171503)
- Qubit assay tubes (Invitrogen, cat. no. Q32856)
- Qubit dsDNA high-sensitivity (HS) kit (Invitrogen, cat. no. Q32851)
- Qubit 2.0 fluorometer (Invitrogen, cat. no. Q32866)
- Agilent 2100 Bioanalyzer (Agilent Technologies, cat. no. G2938C)
- Agilent high-sensitivity DNA kit (Agilent Technologies, cat. no. 5067-4626)
- A compatible Illumina DNA sequencing instrument (instruments currently on the market are the MiSeq, HiSeq 2000, HiSeq 2500)

REAGENT SETUP

Mammalian cell isolation Single cells are obtained by trypsin treatment of adherent cell cultures. To facilitate manual picking with a micro capillary pipette, a small volume of the cell suspension is added to a culture dish containing DPBS and TrypLE Express enzyme in a ratio of 1:1.

Cell lysis buffer Combine 0.2% (vol/vol) Triton X-100 and 2 U μl^{-1} RNase inhibitor. This buffer can be stored at 4 °C for 6 months.

TSO (5′-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3′) At the 5′ end, this TSO carries a common primer sequence, whereas, at the 3′ end, there are two riboguanosines (rG) and one LNA-modified guanosine (+G) to facilitate template switching. Dissolve TSO in TE buffer. TSO can be stored in 100 µM aliquots at −80 °C for 6 months. Avoid repeated freeze-thaw cycles.

Oligo-dT_{30}VN (5′–AAGCAGTGGTATCAACGCAGAGTACT_{30}VN-3′) This anneals to all the RNAs containing a poly(A) tail. The 3′ end of this oligonucleotide contains 'VN', where 'N' is any base and 'V' is either A, C or G. The two terminal nucleotides are necessary for anchoring the oligonucleotide to the beginning of the polyA tail and to avoid unnecessary amplification of a long stretch of adenosines. Dissolve the oligonucleotide in TE buffer, to a final concentration of 100 µM. Store this oligo at −20 °C for 6 months.

ISPCR oligo (5′-AAGCAGTGGTATCAACGCAGAGT-3′) This oligo acts as PCR primer in the amplification step after RT. Dissolve the oligonucleotide in TE buffer to a final concentration of 100 µM. This oligo can be stored at −20 °C for 6 months.

PROCEDURE

Preparing single-cell samples ● TIMING ~15 min

▲ **CRITICAL** Use a thermal cycler with a heated lid set to 105 °C for all incubations throughout this protocol.

1| Clean the hood with RNaseZap and DNA-OFF solutions before setting up the working plates. Spray pipettes with RNaseZap.

▲ **CRITICAL STEP** All the experiments must be performed under a UV-sterilized hood with laminar flow, and all the surfaces must be free from RNase to prevent degradation of RNA and from DNA to prevent cross-contamination from previous samples. The hood must be used only for single-cell experiments up to (but excluding) the cDNA amplification step (Step 12). An ideal scenario would be to place the hood in a separate room with a positive air pressure to prevent any contaminants from being carried inside, where they might affect the experiments. The room should be equipped with a garmenting area in which the user changes into a fresh disposable lab coat, hair net, dust mask, shoe covers and vinyl gloves (powder-free).

Single-cell lysis ● TIMING ~15 min (for eight-strip tubes)

2| Dilute the oligo-dT_{30}VN primer to 10 µM by adding 10 µl of 100 µM oligo-dT primers and 90 µl of nuclease-free water to a tube and mix well.

3| Prepare cell lysis buffer by adding 1 µl of RNase inhibitor to 19 µl of a 0.2% (vol/vol) Triton X-100 solution. If you are working with purified RNA, this step can be omitted and a corresponding volume of water can be used instead.

4| Isolate single cells in the lowest possible volume (preferably ≤0.5 µl, possibly 0.3 µl) or pipet the appropriate amount of RNA into a 0.2-ml thin-walled PCR tube. Single cells can be obtained either by using a micro capillary pipette or via FACS.

5| Place each single cell into a 0.2-ml thin-walled PCR tube containing 2 µl of cell lysis buffer, 1 µl of oligo-dT primer and 1 µl of dNTP mix.

6| Quickly vortex the tube to mix, and then spin down the solution (700*g* for 10 s at room temperature) and immediately place it on ice.

7| Incubate the samples at 72 °C for 3 min and immediately put the tube back on ice.

8| Spin down the samples (700*g* for 10 s at room temperature) to collect the liquid at the bottom of the tubes, and then put them immediately back on ice. The oligo-dT primer is now hybridized to the poly(A) tail of all the mRNA molecules.

Reverse transcription ● TIMING 3 h

9| Prepare the RT mix for all reactions plus one additional reaction by combining and mixing the reagents listed in the table below.

▲ **CRITICAL STEP** Thaw all the reagents in advance and assemble the RT mix while performing denaturation (Step 7) to minimize bias.

▲ **CRITICAL STEP** The addition of $MnCl_2$ to improve the template-switching efficiency is not necessary[26]. It actually decreases the final yield of cDNA and causes misincorporation during RT[12].

Component	Volume (µl)	Final concentration
SuperScript II reverse transcriptase (200 U µl^{-1})	0.50	100 U
RNAse inhibitor (40 U µl^{-1})	0.25	10 U
Superscript II first-strand buffer (5×)	2.00	1×
DTT (100 mM)	0.50	5 mM
Betaine (5 M)	2.00	1 M
$MgCl_2$ (1 M)	0.06	6 mM
TSO (100 µM)	0.10	1 µM
Nuclease-free water	0.29	–
Total volume	5.70	–

10| Add 5.7 µl of the RT mix to the samples from Step 9 to obtain a final reaction volume of 10 µl. Mix the reaction by gently pipetting up and down a few times without forming bubbles.

11| Spin down the samples (700*g* for 10 s at room temperature) to collect the liquid at the bottom of the tubes, and incubate the reaction in a thermal cycler with a heated lid, as detailed below. Note that the ten cycles after the initial incubation at 42 °C for 90 min are not crucial for a successful reaction, but they give a marginal increment in yield.

Cycle	Temperature (°C)	Time	Purpose
1	42	90 min	RT and template-switching
2–11	50	2 min	Unfolding of RNA secondary structures
	42	2 min	Completion/continuation of RT and template-switching
12	70	15 min	Enzyme inactivation
13	4	Hold	Safe storage

PCR preamplification ● TIMING 3 h

12| Prepare the PCR mix for all reactions plus one additional reaction by combining and mixing the following components:

Component	Volume (µl)	Final concentration
First-strand reaction	10	–
KAPA HiFi HotStart ReadyMix (2×)	12.50	1×
IS PCR primers (10 µM)	0.25	0.1 µM
Nuclease-free water	2.25	–
Total volume	25	–

13| Add 15 µl of PCR mix to each tube from Step 12, which contains the first-strand reaction. Vortex the tubes to mix, and then spin them down (700*g* for 10 s at room temperature) to collect the liquid at the bottom of the tubes.

14| Perform the PCR in a thermal cycler by using the following program:

Cycle	Denature	Anneal	Extend	Hold
1	98 °C, 3 min	–	–	–
2–19	98 °C, 20 s	67 °C, 15 s	72 °C, 6 min	–
20	–	–	72 °C, 5 min	–
21	–	–	–	4 °C

▲ **CRITICAL STEP** The number of PCR cycles depends on the input amount of RNA. We typically use 18 cycles for single eukaryotic cells to obtain ~1–30 ng of amplified cDNA. The number of cycles can be increased for smaller cells (with less RNA content) or lowered for large cells (with more RNA).
▲ **CRITICAL STEP** The KAPA HiFi HotStart ReadyMix buffer has a higher salt concentration compared with common PCR buffers. The buffer composition affects DNA melting, and it is therefore important to perform the denaturation Step at 98 °C and not at 95 °C. Similarly, the optimal annealing temperature for the ISPCR primers is different compared with other DNA polymerases. In our hands, the KAPA HiFi has proven to be a very robust enzyme, giving satisfactory results in the range 64–68 °C for the annealing step, with no relevant differences in cDNA yield or length.
■ **PAUSE POINT** PCR product can be stored at –20 or –80 °C for 6 months or longer.

PCR purification ● **TIMING ~45 min**

15| Before commencing the purification steps, equilibrate Ampure XP beads at room temperature for 15 min, and then vortex well for several seconds.

16| Add 25 µl of Ampure XP beads (1:1 ratio) to each sample from Step 14 and mix by pipetting up and down ten times or until the solution appears homogeneous. Transfer solutions to a 96-well plate with compatible magnet stand.
▲ **CRITICAL STEP** Do not increase the volume of beads in the purification step above the 1:1 ratio. A less-than-standard amount of beads ensures that primer dimer carryover is kept to a minimum.

17| Incubate the mixture for 8 min at room temperature to let the DNA bind to the beads.

18| Place the 96-well plate on the magnetic stand for 5 min or until the solution is clear and the beads have been collected at one corner of the well.

19| While samples are on the magnet, carefully remove the liquid without disturbing the beads.

20| Wash the beads with 200 µl of 80% (vol/vol) ethanol solution. Incubate the samples for 30 s and then remove the ethanol.
▲ **CRITICAL STEP** It is important that the ethanol solution is freshly prepared every time, as ethanol absorbs moisture from the environment, thus changing the final concentration.

21| Repeat Step 20 once more.

22| Remove any trace of ethanol and let the beads dry completely, leaving the plate at room temperature for 5 min or until a small crack appears on the surface of the beads.
▲ **CRITICAL STEP** Avoid overdrying the beads because this will make their resuspension in the designated buffer more difficult. As a precaution, cover the plate during this step or protect it from any possible source of contamination or air flows that might disperse the beads around the well, thus leading to cross-contamination between adjacent wells (especially, when the beads are overdried).

23| Add 17.5 µl of EB solution (or nuclease-free water). Mix ten times to resuspend the beads.

24| Incubate the plate off the magnet for 2 min.

25| Place the plate on the magnetic stand and leave it for 2 min or until the solution appears clear and beads have accumulated in a corner of the well.

26| Set the volume of the pipette to 15 µl, collect the supernatant without disturbing the beads and transfer it to a fresh 0.2-ml thin-walled PCR tube.
▲ **CRITICAL STEP** Avoid aspirating the whole volume of EB solution or nuclease-free water. Leaving 2.5 µl in the well ensures that bead carryover is kept to a minimum.

Quality check of the cDNA library ● TIMING 1 h
27| Check the size distribution on an Agilent high-sensitivity DNA chip. A good library should be free of short (<500 bp) fragments and should show a peak at ~1.5–2 kb.
? TROUBLESHOOTING

Tagmentation reaction ● TIMING ~10 min
28| Tagmentation is carried out by using the Illumina Nextera XT DNA sample preparation kit. Set up the tagmentation reaction on ice as follows and mix the solution carefully with a pipette:

Component	Volume (µl)	Final concentration
Tagment DNA buffer (TD, 2×)	10	1×
Amplicon tagment mix	5	–
DNA from PCR	Variable	–
Nuclease-free water	Variable	–
Total volume	20	–

▲ **CRITICAL STEP** When you are using the Nextera XT DNA sample preparation kit, the input should never exceed 1 ng. An excess of DNA leads to incomplete cutting and libraries with longer average size.
▲ **CRITICAL STEP** We recommend assembling the reaction on ice. The Tn5 enzyme has optimal activity at 55 °C, but some activity also occurs at room temperature.

29| Perform the tagmentation reaction in a thermal cycler by using the following program:

Cycle	Temperature (°C)	Time
1	55	5 min
2	4	Hold

Stripping Tn5 transposase off the tagmented DNA ● TIMING 5 min
30| To strip off the Tn5, add 5 µl of NT buffer (Nextera XT DNA sample preparation kit) to the 20-µl solution containing the tagmented DNA (from Step 29). Pipet the solution up and down a few times to mix.

31| Incubate the mixture for 5 min at RT. DNA is now ready for the final enrichment PCR.

Amplification of adapter-ligated fragments ● TIMING 1 h
32| Prepare the enrichment PCR in a 0.2-ml tube using the reagents from the Nextera XT kit.

Component	Volume (µl)	Final concentration
DNA	25	–
Nextera PCR master mix	15	–
Index 1 primers (N7xx)	5	–
Index 2 primers (N5xx)	5	–
Total volume	50	–

33| Next, perform the PCR by using the following program:

Cycle	Denature	Anneal	Extend	Hold
1	–	–	72 °C, 3 min	–
2	95 °C, 30 s	–	–	–
3–14	95 °C, 10 s	55 °C, 30 s	72 °C, 30 s	–
15	–	–	72 °C, 5 min	–
16	–	–	–	4 °C

Amount (pg)	No. of cycles
500–1,000	6–8
100–500	10–12
50–100	12–14
10–50	14–16

▲ **CRITICAL STEP** The number of cycles depends on the amount of DNA used for tagmentation. If we are starting from 100 pg of amplified cDNA, we usually perform 12 PCR cycles. The optimal number of cycles depends on the sample and the experiment. It may be helpful to run a range of cycles to determine the best conditions. Above are cycling guidelines on the basis of the input DNA used for tagmentation.

PCR purification ● TIMING ~45 min

34| Repeat Steps 15–24, but use only 30 µl of AMPure XP beads in a ratio of 0.6:1. This will minimize the carryover of primer dimers.

35| Place the tube on the magnetic stand to capture the beads for 2 min or until the solution is clear. Set the volume of the pipette to 15 µl, collect the supernatant and transfer it into a new 1.5-ml polyallomer tube to minimize DNA absorption to the tube during long-term storage.

Quality check of the final cDNA library ● TIMING 1 h

36| Measure the concentration of each library by Qubit according to the manufacturer's instructions. Check the size distribution on an Agilent high-sensitivity chip. Size distribution varies according to the amount of DNA used in the tagmentation. As a rule of thumb, a broad peak with an average size of 300–800 bp will be observed.

? TROUBLESHOOTING

■ **PAUSE POINT** The sample can be stored after this step at –20 °C for 3 months.

Library pooling ● TIMING ~30 min

37| Perform a dilution of the library, aiming to get a 2 nM solution. Use the concentration (ng μl^{-1}) obtained with the Qubit and the average size obtained on the Bioanalyzer to calculate the molarity of the final library.

38| Measure the diluted library by Qubit and adjust the concentration to get 2 nM, if necessary.

39| Pool equal nanomoles of each sample, ensuring that none of them has the same combination of N5xx and N7xx adapters. If you are pooling just a few samples (especially fewer than six), make sure that you maintain the color balance for each base of the index read, or else the index read sequencing might fail because of registration failure. Check the Nextera DNA sample preparation guide (available at http://www.illumina.com) to determine the compatibility of each index.

40| Measure once again the concentration of the pooled libraries by Qubit, and then adjust the concentration to 2 nM, if necessary.

DNA sequencing ● TIMING ~1–2 d

41| Perform single-end or paired-end sequencing of the libraries according to the manufacturer's protocol and by using the TruSeq dual-index sequencing primers on either a HiSeq 2000, 2500 or MiSeq instrument; a 50-bp single-end sequencing on a HiSeq 2000 takes ~2 d, whereas that on a MiSeq takes less than 1 d. The degree of multiplexing per lane depends on the read depth obtained on each instrument. Multiplex sequencing of 50 libraries on a single HiSeq 2000, from which 200 million reads are regularly obtained, would give an average sequence depth of 4 million reads per cell. If the experiment is designed to identify SNPs, mutations or RNA splice variants, one might require a higher sequence depth per cell by pooling fewer libraries per lane.

Data analysis ● **TIMING ~1 d**

42| Export raw sequences and quality scores in FastQ format for all clusters that passed filtering (i.e., pass filter clusters) using CASAVA software.

43| Use the index reads to demultiplex the sequenced reads into correct sample origin (i.e., cells).

44| Align the reads to the genome and transcriptome by using an aligner developed for RNA-seq data.

45| Estimate gene expression levels for each gene or transcripts by following dedicated protocols, e.g., ref. 36.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
27	No product after cDNA preamplification	Cell was dead or damaged at the moment of picking	Keep the cells in conditions that are as close as possible to those present in the tissue/culture of origin. Check the quality of cells (i.e., by staining for dead cells) and avoid leaving them in DPBS or cell medium for too long before picking
	Noisy amplification with a tail of short fragments (see **Fig. 2c**, 'hedgehog pattern')	Large excess of TSO in relation to the mRNA is present in the cell (appears more frequently with cells that have a low transcriptional activity)	Decrease the amount of TSO used (a titration may be necessary) or use a 5′-blocked TSO that does not form concatamers
	Overall yield is low or lower than expected	Sample quality was poor	Keep cells in conditions that are as physiological as possible before and during the picking. Work fast, immediately put the cells in lysis buffer containing RNase inhibitor. Keep the cells that have already been picked on ice until you start the RT
	Amount of cDNA is much larger than usual	Contamination occurred during the first steps	Perform all the experiments in a dedicated area that has been thoroughly cleaned and decontaminated from RNases and DNases Use a dust mask, gloves, hair net, dedicated lab coat and change gloves often
		Overamplification of cDNA after RT	The number of PCR cycles can vary from sample to sample. Start with 18 cycles and adjust it later, if necessary
36	Tagmented DNA shows a wide distribution with a tail of longer fragments	Too much DNA was used in the tagmentation reaction	Use less DNA and never exceed 1 ng for optimal results
		Partial inhibition of Tn5 transposase	Elution from AMPure XP beads in the previous step was done using a solution containing a detergent or a buffer with a too high salinity. Even a small amount of detergent completely inhibits the enzyme
	Enough DNA is used as input for tagmentation but no or very low peaks after enrichment PCR is observed	Too few cycles of PCR	Increase the number of PCR cycles
	Primer dimers are present on the Bioanalyzer trace	Overamplification or an excessive amount of adaptors relative to DNA is used for tagmentation	Reduce either the number of PCR cycles or the amount of adaptors. Use a lower ratio of AMPure beads:DNA (never exceed a 1:1 ratio, preferably use a 0.6:1 ratio)

● TIMING

Day 1

Steps 1–8, preparing single-cell samples, picking and lysis: ~30 min (for eight-strip tubes)
Steps 9–11, reverse transcription: 3 h
Steps 12–14, PCR preamplification: 3 h

Day 2

Steps 15–26, PCR purification: ~45 min
Step 27, quality check of the cDNA library: 1 h
Steps 28 and 29, tagmentation reaction: ~10 min
Steps 30 and 31, purification of tagmented DNA: 5 min
Steps 32 and 33, amplification of adapter-ligated fragments: 1 h
Steps 34 and 35, PCR purification: ~45 min
Step 36, quality check of the final cDNA library: 1 h
Steps 37–40, library pooling: ~30 min

Days 3–5

Step 41, DNA sequencing: 1–2 d
Steps 42–45, data analysis: ~1 d

ANTICIPATED RESULTS

Step 27

After cDNA pre-amplification and purification, one has the first opportunity to determine the overall quality of the initial cell. This is normally performed with an Agilent Bioanalyzer. A good library should have an average size of 1.5–2 kb and a small amount of short fragments (**Fig. 2a**). In the case of suboptimal preamplification, the peak of primer dimers will be higher than the cDNA library (**Fig. 2b**) or, in extreme cases, mostly amplification of primer concatamers will be observed (**Fig. 2c**). A library prepared with degraded RNA is characterized by a shift toward short fragments, even though some amplification of longer fragments could take place (**Fig. 2d**).

Step 36

After this step, the final cDNA library is ready for sequencing. The cDNA has been tagmented and amplified with index primers that specifically enrich the final library for fragments carrying adapters on both ends. The expected size of the library is ~300–800 bp, and the amount of adapter dimers is low (**Fig. 3**).

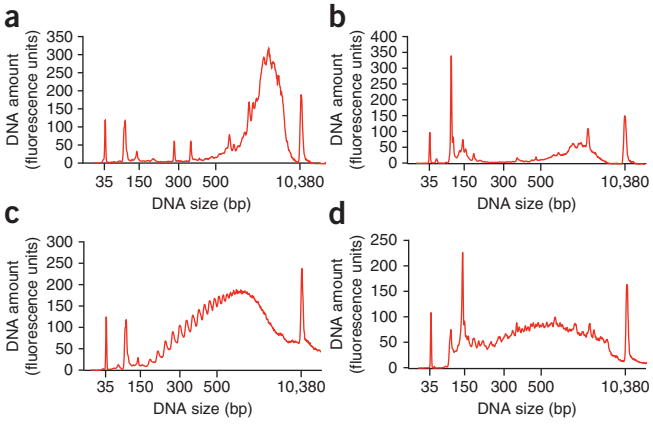

Figure 2 | Bioanalyzer electropherograms of pre-amplified cDNA libraries. (**a**) Representative example of the cDNA size distribution obtained from the successful cDNA preamplification from a single MEF cell (mouse). The profile has a peak ~1.5–2 kb with a small number of fragments below 500 bp and a small amount of primer dimers. (**b**) Representative example of the primer dimer–dominated cDNA profile from a single HEK293T cell (human). Amplification of full-length cDNA is visible but a large peak of primer dimers that were not removed by bead purification dominates the profile. (**c**) Data from a single T cell (human) with a 'hedgehog' pattern, most probably due to the formation of TSO concatamers. (**d**) Data from a single interneuron (mouse), with a wide distribution of fragments indicative of RNA degradation before library preparation.

Step 46 (end result)

Sequence reads from each individual cell are normally in the range of 1–20 million, depending on the level of multiplexing in the sequencing. When sequencing 50-bp single-end reads, we find that normally 60% of reads map uniquely to the genome (20% multimapping and 20% with no match); of the uniquely mapping reads, >60% of the reads map to annotated RefSeq exons, 20% intronic and 20% intergenic, but these values depend on the completeness of the gene annotations. The read coverage across transcripts should be even. Representative values for single-cell libraries prepared from HEK293T and DG-75 cells are listed in **Supplementary Table 1**.

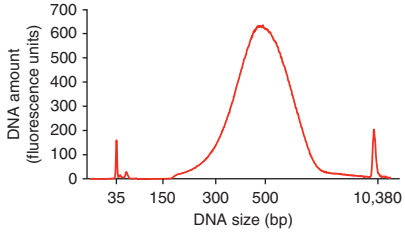

Figure 3 | Bioanalyzer electropherogram of a sequencing library after tagmentation. Representative profile of a library derived from a single HEK293T cell (human) prepared according to the Nextera XT protocol. Amplification (12 cycles) was carried out according to manufacturer's instructions, as described in Step 33.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS This work was supported by the European Research Council Starting grant no. 243066 (R.S.), the Swedish Foundation for Strategic Research FFL4 (R.S.), the Åke Wiberg Foundation (R.S.) and the Swedish Research Council grant nos. 2008-4562 (R.S.) and 2010-6844 (Å.K.B.).

AUTHOR CONTRIBUTIONS S.P. developed and wrote the protocol; O.R.F. conceived and designed LNA-based oligos; Å.K.B. performed computational analyses; S.S. picked cells; G.W. contributed to protocol development; R.S. designed the study and wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

1. Wilkinson, D.J. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.* **10**, 122–133 (2009).
2. Wilhelm, B.T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243 (2008).
3. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
4. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).
5. Tang, F. *et al.* mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
6. Tang, F. *et al.* RNA-seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516–535 (2010).
7. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
8. Ramskold, D. *et al.* Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
9. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
10. Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nat. Protoc.* **7**, 813–828 (2012).
11. Sasagawa, Y. *et al.* Quartz-seq: a highly reproducible and sensitive single-cell RNA-seq reveals non-genetic gene expression heterogeneity. *Genome Biol.* **14**, R31 (2013).
12. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
13. Yan, L. *et al.* Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
14. Huang, L., Lee, J., Sitaraman, K. & Gallego, A. A new highly sensitive two-step RT-PCR system. *Focus* **22**, 6–7 (2000).
15. Abbotts, J., Bebenek, K., Kunkel, T.A. & Wilson, S.H. Mechanism of HIV-1 reverse transcriptase. Termination of processive synthesis on a natural DNA template is influenced by the sequence of the template-primer stem. *J. Biol. Chem.* **268**, 10312–10323 (1993).
16. Brooks, E.M., Sheflin, L.G. & Spaulding, S.W. Secondary structure in the 3′ UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR. *Biotechniques* **19**, 806–812, 814–805 (1995).
17. Kuo, K.W., Leung, M.F. & Leung, W.C. Intrinsic secondary structure of human TNFR-I mRNA influences the determination of gene expression by RT-PCR. *Mol. Cell Biochem.* **177**, 1–6 (1997).
18. Carninci, P. *et al.* Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. USA* **95**, 520–524 (1998).
19. Spiess, A.N. & Ivell, R. A highly efficient method for long-chain cDNA synthesis using trehalose and betaine. *Anal. Biochem.* **301**, 168–174 (2002).
20. Pinto, F.L. & Lindblad, P. A guide for in-house design of template-switch-based 5′ rapid amplification of cDNA ends systems. *Anal. Biochem.* **397**, 227–232 (2010).
21. Santoro, M.M., Liu, Y., Khan, S.M., Hou, L.X. & Bolen, D.W. Increased thermal stability of proteins in the presence of naturally occurring osmolytes. *Biochemistry* **31**, 5278–5283 (1992).
22. Rajendrakumar, C.S., Suryanarayana, T. & Reddy, A.R. DNA helix destabilization by proline and betaine: possible role in the salinity tolerance process. *FEBS Lett.* **410**, 201–205 (1997).
23. Rees, W.A., Yager, T.D., Korte, J. & von Hippel, P.H. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**, 137–144 (1993).
24. Lambert, D. & Draper, D.E. Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA-Mg^{2+} interactions. *J. Mol. Biol.* **370**, 993–1005 (2007).
25. Eckert, K.A. & Kunkel, T.A. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18**, 3739–3744 (1990).
26. Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. & Siebert, P.D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897 (2001).
27. Petersen, M. & Wengel, J. LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol.* **21**, 74–81 (2003).
28. Quail, M.A. *et al.* Optimal enzymes for amplifying sequencing libraries. *Nat. Methods* **9**, 10–11 (2012).
29. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119 (2010).
30. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
31. Torres, L. *et al.* Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. *Breast Cancer Res. Treat* **102**, 143–155 (2007).
32. Shipitsin, M. *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11**, 259–273 (2007).
33. Navin, N. & Hicks, J. Future medical applications of single-cell sequencing in cancer. *Genome Med.* **3**, 31 (2011).
34. Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148**, 886–895 (2012).
35. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873–885 (2012).
36. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).