Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing

Saiful Islam[1], Una Kjällquist[1], Annalena Moliner[2], Pawel Zajac[1], Jian-Bing Fan[3], Peter Lönnerberg[1] & Sten Linnarsson[1]

[1]Laboratory for Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [2]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. [3]Illumina Inc., San Diego, California, USA. Correspondence should be addressed to S.L. (sten.linnarsson@ki.se).

Published online 5 April 2012; doi:10.1038/nprot.2012.022

Single-cell analysis of gene expression is increasingly important for the analysis of complex tissues, including cancer, developing organs and adult stem cell niches. Here we present a detailed protocol for quantitative gene expression analysis in single cells, by the sequencing of mRNA 5′ ends. In all, 96 cells are lysed, and their mRNA is converted to cDNA. By using a template-switching mechanism, a bar code and an upstream primer-binding sequence are introduced simultaneously with reverse transcription. All cDNA is pooled and then prepared for 5′ end sequencing, including fragmentation, adapter ligation and PCR amplification. The chief advantage of this approach is the great reduction in cost and time, afforded by the early bar-coding strategy. Compared with previous methods, it is more suitable for large-scale quantitative analysis, as well as for the characterization of transcription start sites, but it is unsuitable for the detection of alternatively spliced transcripts. Sample preparation takes 3 d, and two sets of 96 cells can be prepared in parallel. Finally, the sequencing and data analysis can take an additional 4 d altogether.

INTRODUCTION

Tissues are rarely homogeneous and gene expression is invariably heterogeneous even in similar cell types[1]. Any gene expression analysis on the tissue level therefore will blend the true expression profiles of constituent cells. In addition, cell-to-cell variation in gene expression due to stochastic effects in gene expression is now established[2]. Therefore, the best approach for analyzing complex tissues may be to perform analysis at the single-cell level.

Nearly two decades ago, the polymerase chain reaction was used to detect and amplify DNA from single cells[3]. There have been different approaches to detect gene expression at the single-cell level in the past decade, including microarray[4,5], quantitative PCR (qPCR)[6,7] and, recently, RNA-seq[8]. RNA-seq in particular opens the door for single-cell analysis more widely, and it therefore facilitates the understanding of the complexity of gene expression in single cells.

Outline of the protocol

Here we give a detailed protocol for single-cell tagged reverse transcription (STRT), a highly multiplexed method for single-cell RNA-seq on the Illumina platform[9]. See **Figure 1** for a schematic illustration of the protocol. We collect 96 single cells in a 96-well plate (each well contains a single cell) and incorporate a bar code using the template-switching mechanism during reverse transcription (RT). The resulting cDNA library is then processed for sequencing on the Illumina platform. In this method, the sequenced fragments correspond to a template-switching site located preferentially at the 5′ end of mRNA, which can be used to analyze promoter usage in single cells by CAGE (cap analysis of gene expression, e.g., NanoCAGE[10]). To confirm the 5′ location of reads, and to quantify the sensitivity and dynamic range of the method, we introduce eight synthetic polyadenylated RNAs into each well, with known concentrations and lengths ranging from 750 bp to 2,000 bp.

General principles of single-cell expression analysis

A typical single mammalian cell contains only picograms of total RNA, and only 1–5% of this is polyadenylated mRNA, corresponding to only about 10^5–10^6 mRNA molecules. Thus, the challenge is to efficiently isolate such a small number of molecules and then accurately quantify the number of copies present of each type of transcript. Furthermore, gene expression in single cells is intrinsically very noisy: genes are transcribed in infrequent, random bursts separated by periods of mRNA decay and dilution because of cell division[11,12]. As a consequence, even highly expressed genes will show considerable cell-to-cell variation. To attenuate this noise, it is necessary to analyze large numbers of single cells. Thus, a highly multiplexed protocol is required, which reduces cost and time of sample preparation.

In single-cell analysis, the capture efficiency—the fraction of mRNA molecules that are actually recovered and quantified—is as important as the quantification accuracy. This is in contrast to bulk RNA analysis, where losses can be tolerated as long as the remaining sample is still representative of the original. Thus, if 90% of a bulk RNA sample is lost during sample preparation, the remaining 10% could still yield an accurate gene expression profile. In contrast, if 90% of a single-cell RNA sample were lost, then all information on most genes expressed in the range of 1–10 copies per cell would be lost irretrievably.

The most straightforward way to estimate capture efficiency is to include one or more internal control mRNAs, which may be, for example, synthetic *in vitro*–transcribed RNAs such as those described in the present protocol. If these are added into the same tube as the single cell at a known concentration, they can be used to control for any losses or enzymatic inefficiencies throughout the protocol (but not, however, for the efficiency of cell lysis or any degradation happening inside a possibly dying cell prior to lysis). Although this has not been a common practice in the past, we suggest that internal controls be routinely used. The recently developed External RNA Controls Consortium (ERCC) RNA Spike-in Control Mixes[13] should facilitate wider adoption of this practice.

Once a single cell has been lysed and its mRNA efficiently converted to cDNA, the sample must be amplified. Although one could imagine directly analyzing unamplified cDNA (or even mRNA)

from single cells using single-molecule sequencing, this has not yet been demonstrated in practice. Amplification of single-cell cDNA has been commonly achieved by either PCR[14] or by several rounds of *in vitro* transcription (IVT)[15]. In both cases, an upstream sequence must be introduced at the 3′ end of the cDNA (5′ end of the mRNA) to serve as template for the amplification using a universal primer. The two common methods to introduce this sequence are by using a terminal transferase[4] (which introduces a homopolymer sequence) or by using the template-switching mechanism of Moloney murine leukemia virus (MMLV) reverse transcriptase[16]. The latter is based on the fact that MMLV reverse transcriptase introduces a few non-templated nucleotides (predominantly cytosines) at the 3′ end of the cDNA, which can anneal to a helper oligo having a few riboguanines at its 3′ end. Reverse transcriptase then switches template and continues to synthesize a copy of the helper oligo. Thus, template switching permits the introduction of an arbitrary sequence, directed by a synthetic oligonucleotide. Unfortunately, the absolute efficiencies (percentage of mRNA converted to amplifiable cDNA) of these methods have not been directly measured, or even compared side by side. The suitably modified cDNA is then amplified, either by PCR or by IVT. In both cases, amplification must be kept to a minimum in order to avoid introducing a quantitative bias.

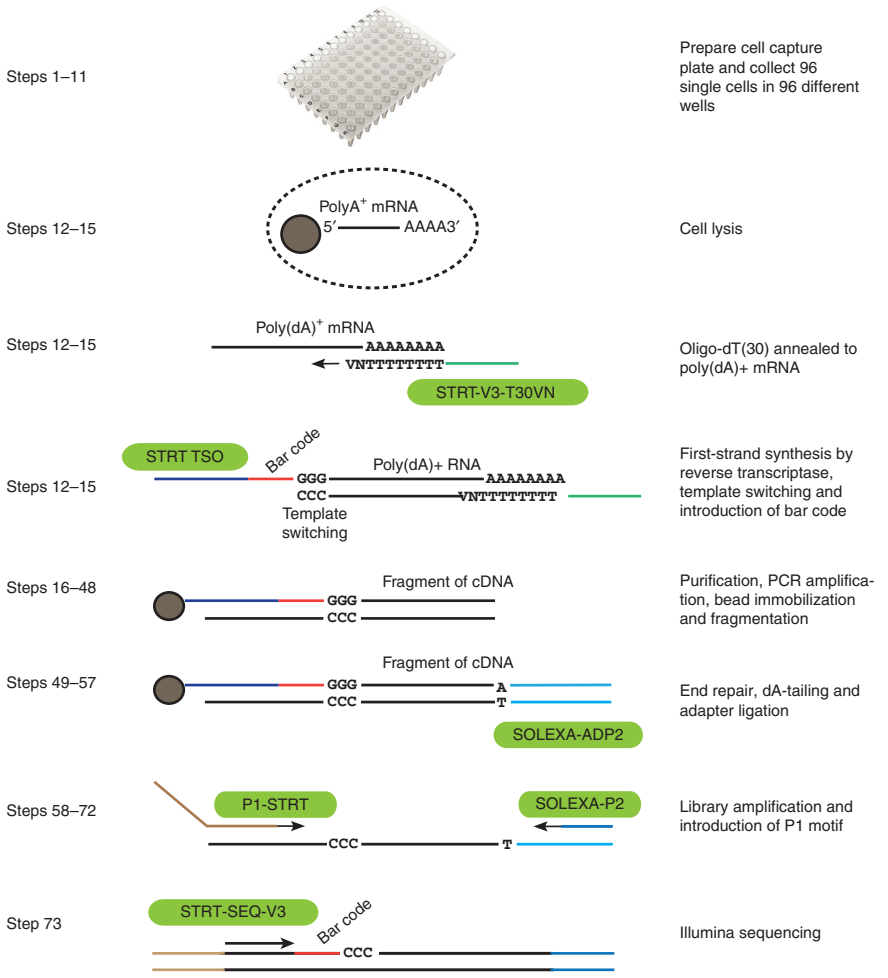

Figure 1 | Schematic overview of the method. Custom oligonucleotides are labeled when first introduced (named as in the REAGENT SETUP section, on green background) and are colored so that they can be followed through the protocol.

Finally, the amplified cDNA is analyzed, traditionally by microarray but more recently by high-throughput sequencing. In the latter case, two general approaches have been proposed. The first method, by the Surani group[8], essentially treats the cDNA as any double-stranded DNA (dsDNA) for sequencing, and subjects it to a standard sample preparation protocol including fragmentation, adapter ligation and library amplification. The entire procedure is repeated for each single cell to be analyzed. The resulting reads were distributed along the entire cDNA length (but strongly biased to the 3′ end), which permitted the analysis of alternative splice isoforms. However, it also complicates quantification, as the number of reads observed for a gene will be proportional to its effective length, which in turn is a function of the 3′ bias of RT. Strand information was lost, and thus it was not possible to determine the direction of transcription, and overlapping exons in opposite orientation could not be distinguished.

The second method, STRT[9], for which we here give a detailed and updated protocol, sequences only the 5′ end of each cDNA. This precludes analysis of splice isoforms, but gives several other advantages. First, it retains strand information and thus permits the distinction of overlapping genes transcribed from opposite strands[17]. Second, it identifies the actual transcription start site (TSS), which is often lost in methods that show 3′ bias. Knowledge of the TSS is crucial for locating promoter and enhancer elements that control transcription in single cells. Third, it facilitates quantification, as each mRNA molecule results in a single cDNA molecule, and thus the number of reads observed should be proportional to the number of mRNA molecules. Finally, it permits the introduction, already during RT, of a bar code for multiplexing. As a result, cDNAs from many cells can be pooled and amplified together, which greatly simplifies the preparation of large numbers of cells, and reduces cost. The protocol given here is for 96 single cells.

Alternative methods

As mentioned earlier, the Surani protocol[8] is more suitable for splice variant detection, especially in 3′ exons, whereas STRT is more suitable for detection of TSSs. By using the Surani protocol[8], one cell is prepared at a time, leading to a higher average cost of preparation; in fact, the cost of STRT is probably an order of magnitude lower, and it is thus more suitable for large-scale analyses. Surani libraries[8] are adapted for use on Applied Biosystems SOLiD, whereas STRT libraries are suitable for the more common Illumina instruments.

There are now also several kits commercially available for low input RNA amounts, generally down to 100 pg and corresponding to about 2–10 cells. The Ovation RNA-seq kit (NuGEN) requires 500 pg of total RNA, whereas SMARTer (Clontech) uses a

protocol very similar to STRT but requires 100 pg of purified total RNA. With appropriate modifications, either of these protocols may be suitable for the analysis of single cells, but this has not yet been demonstrated.

All of the alternative methods suffer from the fact that each cell must be prepared separately. Major concerns for large-scale studies are the cost of, and laboratory work required for, library preparation. STRT currently costs approximately $200 per 96 cells, not including sequencing, and it takes 2 d to prepare two such libraries in parallel. The commercial kits are more than $100 per single reaction, or almost $10,000 for 96 single cells. This shows clearly the cost advantage of the early bar-coding strategy, but there is also a corresponding reduction in the amount of laboratory work. The steps involved in one STRT library preparation are similar to those involved in a single SMARTer preparation, but as the latter can only be multiplexed at a late step (adapter ligation) each single cell is essentially one complete prep. Thus, where it takes 3 working days to prepare a total of 192 cells with STRT, the alternative protocols would take weeks or months to complete a project of similar size. Automation could mitigate this workload, but automation would also be applicable to STRT, and thus the advantage remains.

Future improvements

The method could be improved in several ways. Because we use many cycles of PCR (20 cycles to amplify single cell cDNA and 12 cycles for Illumina sample preparation), technical variation may be introduced as a result of PCR bias. However, we have recently shown that it is possible to reduce or even eliminate amplification bias by introducing unique molecular identifiers (UMIs), which are essentially bar codes of random sequence[18]. It will be interesting to see how the UMI approach could improve single-cell RNA sequencing.

Furthermore, the capture efficiency of conversion of mRNA into amplifiable cDNA could be improved, thereby resulting in greater sensitivity. Currently, we reliably detect only highly and medium-expressed genes, i.e., about ten copies per cell or more. For example, we detected 1,000–6,000 genes in embryonic stem cells and 2,000–8,000 genes in mouse embryonic stem cells. Consistent with the ten-molecule limit of detection, direct measures of the capture efficiency by qPCR revealed that 5–25% of all mRNA molecules were converted into amplifiable cDNA (data not shown). Increasing this number to 50% would lower the detection limit to about two molecules.

As a bar code is introduced at the 5′ end and all 96 samples are pooled before amplification, it is not possible to sequence the entire length of each transcript, as internal reads would not be bar coded. An alternative would be to amplify each sample individually and then prepare them separately for sequencing, similar to the Surani protocol[8]; however, this would greatly increase the reagent and labor cost of preparing 96 libraries.

Finally, we have noticed that the different bar codes do not work with the same efficiency[9]. We found no obvious sequence composition that could explain the differences, but a systematic search for more efficient bar code sequences might help increase the yield of cDNA in all wells.

Applications

The quantitative analysis of gene expression in single cells has been used to characterize the sources of transcriptional noise and bursting[19]; to characterize key events during the derivation of embryonic stem cells from blastocysts[20]; to study expression in early embryos, when the number of cells is naturally limited[21]; to elucidate the heterogeneity and somatic evolution of tumors[22]; and for many other applications. One attractive possibility is that the range of cell type diversity in complex tissues could be directly accessed using single-cell transcriptome analysis of thousands of cells[9]. This would permit *de novo* cell type discovery without the use of pre-existing markers.

Metaphorically, analyzing gene expression in a tissue sample is a lot like measuring the average personal income throughout Europe—many interesting and important phenomena are simply invisible at the aggregate level. Ultimately, we must develop the tools that allow us to count all the molecular components in thousands or even millions of single cells, which would give access to the full spectrum of biological phenomena. The method presented here is one small step on the road to that future.

MATERIALS

REAGENTS

General reagents

- Trizma hydrochloride solution (1.0 M, pH 8; Sigma, cat. no. T3038-1 l)
- PBS
- Tris-HCl (1.0 M, pH 7.60; Sigma, cat. no. T-2788)
- Trizma hydrochloride buffer solution (1.0 M, pH 7.5; Sigma, cat. no. T2319-100 ml)
- Potassium chloride solution (1.0 M; Sigma, cat. no. 60142)
- Magnesium chloride solution (1.0 M; Sigma, cat. no. M1028-100 ml)
- Tween-20 (MERCK Schuchardt, cat. no. S 35789) **! CAUTION** It is harmful; it is irritating to the skin, eyes and respiratory tract. Handle it using appropriate safety equipment.
- Manganese chloride solution (1.0 M; Sigma, cat. no. M-1787)
- Sodium chloride solution (5.0 M; Sigma, cat. no. S 5150-1L)
- Distilled water (GIBCO, cat. no. 10977)
- Dynabeads MyOne streptavidin C1 (Invitrogen, cat. no. 650.02)
- Agencourt Ampure XP beads (Beckman Coulter, cat. no. A 63881)
- EDTA disodium salt solution (0.5 M; Sigma, cat. no. E 7889) **! CAUTION** It is harmful; irritating to the eyes. Handle using appropriate safety equipment.
- Dithiothreitol (DTT; Invitrogen, cat. no. D-1532)
- Superscript II reverse trascriptase (Invitrogen, cat. no. 18064-014)
- USER enzyme (New England Biolabs, cat. no. M 5505L)
- Advantage 2 PCR buffer mix (10×; Clontech, cat. no. 639137)
- Advantage 2 polymerase mix (50×; Clontech, cat. no. 639201)
- Dynabeads MyOne carboxylic acid (Invitrogen, cat. no. 650.12)
- Polyethylene glycol 6000 (PEG-6000; BDH, cat. no. 442715K)
- E-gel (1.2%, wt/vol) with SYBR safe (Invitrogen, cat. no. G5218-01)
- E-gel (2%, wt/vol) with SYBR safe (Invitrogen, cat. no. G5218-02)
- dsDNA fragmentase (New England Biolabs, cat. no. M0348L)
- dsDNA fragmentase reaction buffer (New England Biolabs, cat. no. B0348 S)
- NEBNext end repair buffer (New England Biolabs, cat. no. E6052AA)
- NEBNext end repair enzyme mix (New England Biolabs, cat. no. E6051AA)
- *E. coli* DNA ligase for fragmentase (New England Biolabs, cat. no. M0349AA)
- NEBNext dA-tailing buffer (New England Biolabs, cat. no. E6055AA)
- Klenow fragment (3′-5′ exo-; New England Biolabs, cat. no. E6054AA)
- DNA ladder (100 bp; New England Biolabs, cat. no. N3231 L)
- NEBuffer 4 (New England Biolabs, cat. no. B7004S)
- T4 DNA ligase (Fermentas, cat. no. EL0011)

- SalI-HF (New England Biolabs, cat. no. R3138S)
- Phusion HF reaction buffer (Finnzymes, cat. no. F518)
- Phusion DNA polymerase (Finnzymes, cat. no. F530S)
- ArrayControl RNA spikes (Ambion, cat. no. AM1780)
- Firstchoice human brain reference total RNA (Ambion, cat. no. 6051)
- Buffer EB (Qiagen, cat. no. 19086)
- Deoxynucleotide triphosphate (New England Biolabs, cat. no. N0446)
- BSA (New England Biolabs, cat. no. B9001S)
- Adenosine triphosphate (New England Biolabs, cat. no. P0756L)
- Ethanol (99.5%, vol/vol; Kemetyl, cat. no. SN366915-06) **! CAUTION** It is flammable; handle it using appropriate safety equipment.
- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- QIAquick gel extraction kit (Qiagen, cat. no. 28706)
- RNase*Zap* (Ambion, cat. no. AM9780)
- RNase- and DNase-free water
- Spike-in control solution (**Box 1**)
- TSO stock plate (**Box 2**)

Custom oligonucleotides

- All oligos were ordered from commercial vendor Eurofins MWG operon, with HPLC purification (see REAGENT SETUP)

Adapter oligos

- Adapter oligos are annealed to form a double-stranded adapter (see REAGENT SETUP)

EQUIPMENT

- Thermo-Fast 96, semi-skirted PCR plate (Thermo Scientific, cat. no. AB-0900)
- 96-Well plate (an AbGene 0600 PCR plate; AbGene)
- Microcentrifuge tube, polyallomer (Beckman, cat. no. 357448)
- Falcon polystyrene conical tubes (15 ml; BD Biosciences, cat. no. 352095)
- Thermomixer comfort (Eppendorf, cat. no 5355 000.011)
- Thermal cycler (PTC-100; MJ Research)
- Vortex mixer (Vortex Genie 2; VMR)
- Thermo-strip with domed cap (0.2 ml; Thermo Scientific, cat. no. AB-0266)
- Magnatrix 1200 (Nordiag, cat. no. 8.21.01)
- Reservoir tubes for Magnatrix test rack (1,200/8,000; Nordiag, cat. no. 8.22.17)
- Tips, standard bulk (1,200/8,000; Nordiag, cat. no. 8.22.01)
- Biosphere filter tips (1,000; Sarstedt, cat. no. 70.762.211)
- Biosphere filter tips (200 nuetral; Sarstedt, cat. no. 70.760.211)
- Biosphere filter tips (10 neutral; Sarstedt, cat. no. 70.1115.210)
- Adhesive PCR foil seals (Thermo Scientific, cat. no. AB-0626)
- PS-Microplate, 96 wells, V-shape (Greiner Bio-One, cat. no. 651101)
- PP-Master block, 96 wells, V-shape (Greiner Bio-One, cat. no. 786201)
- E-gel opener (Invitrogen, cat. no. G5300-01)
- E-gel opener blades (Invitrogen, cat. no. 65300-01)
- Safe Imager viewing glass (Invitrogen, cat. no. S37103)
- Safe Imager blue-light transilluminator (Invitrogen, cat. no. G6600)
- Qubit assay tubes (Invitrogen, cat. no. Q32856)
- Qubit RNA assay kit (Invitrogen, cat. no. 32852)
- Qubit dsDNA HS assay kit (Invitrogen, cat. no. 32851)
- Qubit 1.0 fluorometer (Invitrogen, cat. no. Q32857)
- Disposable filter unit (0.45 µM; Schleicher & Schuell, cat. no. 10462100)
- Cell strainer with a 40-µm nylon mesh (BD, product number: 352340)
- CellEctor (Molecular Machines and Industries)
- A compatible Illumina DNA sequencing instrument (currently including the Genome Analyzer IIx, HiSeq 1000, HiSeq 2000, MiSeq and HiScanSQ).
- Data analysis pipeline software is freely available from the authors

REAGENT SETUP

STRT buffer, 5× Mix 100 mM Tris-HCl (pH 8), 375 mM KCl, 30 mM MgCl2 and 0.1% (vol/vol) Tween-20. This solution is prepared in RNase/DNase-free water. Store at 4 °C for 6 months.

STRT-RT buffer, 5× Mix 100 mM Tris-HCl, pH 8, 375 mM KCl and 0.1% (vol/vol) Tween-20. This solution is prepared in RNase/DNase-free water. Store at 4 °C for 6 months.

Binding and washing buffer, 2× (2× BWT) Mix 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl and 0.01% (vol/vol) Tween-20. Store at room temperature (23 °C) for 6 months.

Box 1 | Preparing spike-in solution as control ● TIMING 2 h

The eight ArrayControl RNA spikes are supplied as 100 ng µl^{-1} in a 10-µl volume. Ambion ArrayControl RNA includes eight *in vitro*–transcribed polyadenylated transcripts, ranging from 750 to 2,000 bp in length. We will make a tenfold dilution series containing pairs of controls at four different concentrations. For example, spike-1 and spike-2 will be 1 ng µl^{-1}, spike-3 and spike-4 will be 100 pg µl^{-1}, and so on.

Stock spike RNA solution:

First prepare fresh storage buffer (10 mM Tris, pH 7.6, 0.05% (vol/vol) Tween-20 in RNase/DNase-free water). Start to prepare dilution series by taking 1 µl each from spike-7 and spike-8 and add 8 µl of storage buffer to make dilution A. Now take 1 µl each from spike-5 and spike-6 and 1 µl of spike RNA from dilution A. Add 7 µl of storage buffer to these mixtures to make dilution B. Next, take 1 µl each from spike-3 and spike-4 and 1 µl of spike RNA from dilution B. Add 7 µl of storage buffer to make dilution C. In the next step, take 1 µl each from spike-1 and spike-2 and 1 µl of dilution C. Add 7 µl of storage buffer and mix to make dilution D. Finally, add 90 µl of storage buffer to dilution D. The final concentration of the dilution D is 1 ng µl^{-1} for spike-1 and spike-2; 100 pg µl^{-1} for spike-3 and spike-4; 10 pg µl^{-1} for spike-5 and spike-6; and 1 pg µl^{-1} for spike-7 and spike-8. Prepare a stock of control RNA by making aliquots of 1 µl of the final dilution in polyallomer tubes; store them at −80 °C for up to 6 months.

Working spike RNA solution:

Always freshly prepare working spike RNA solution from the stock control RNA solution. The final concentration of working spike solution will be 10 fg µl^{-1} (spike-1 and spike-2). This dilution is made from stock spike RNA solution where the concentration is 1 ng µl^{-1} (spike-1 and -2). Take one tube of 1-µl stock spike RNA solution from −80 °C and thaw on ice. Add 999 µl of freshly prepared storage buffer on it and mix extensively by vortexing. Then transfer 10 µl of this dilution to a new polyallomer tube and add 990 µl of freshly prepared storage buffer. Again, mix extensively by vortexing. The working spike solution is now ready for use.

▲ **CRITICAL** RNase is ubiquitous in nature and can easily degrade the RNA. Great care should be taken when preparing spike-in RNA, which is of very low concentration and is crucial for calibrating the exact number of molecules in each reaction. The water used for this purpose must be RNase-free water, and all other chemicals such as Tris-HCl and Tween-20 are also prepared in RNase-free water. In addition, the Tween-20 solution should be filtered with a 0.45-µm syringe filter. Wear disposable gloves throughout the RNA preparation and occasionally change the gloves as necessary to prevent any cross-contamination. Use filter tips in all pipetting steps throughout the experiment to prevent any cross-contamination. Use polyallomer tubes (Beckman) to prepare RNA (these tubes cause a minimum loss of RNA from binding in the surface of the polyallomer tubes) and perform the procedure on ice.

Note: The recently developed ERCC spike-in controls[13] may be a better choice for future experiments, as they contain a larger number of synthetic transcripts and were designed specifically for RNA-seq.

Box 2 | Preparing the 96-TSO stock plate ● TIMING 1 h

Prepare a stock plate of 96 TSOs (**Supplementary Table 1**) as follows. Take a V-shaped 96-well plate Greiner PP-Master block and add 190 µl of DNase/RNase-free water in all 96 wells. Add 10 µl of 100 µM TSO from each original oligo tube according to the layout in **Supplementary Table 2**. The final TSO concentration is 5 µM. Mix the reaction by pipetting up and down. The stock plate is enough to make many cell capture plates. It is kept at −20 °C for up to 6 months.
▲ CRITICAL Filter tips must be used to prevent accidental mixing of two different oligos. Mix the reaction very carefully and gently with a pipette. Be attentive to prevent any mixing between wells in any stage of the master plate preparation. Keep the plate on ice during the preparation.

Binding and washing buffer, 1× (1× BWT) Mix 5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 1 M NaCl and 0.02% (vol/vol) Tween-20. Store at room temperature for 6 months.
EBT buffer Combine 10 mM Tris-Cl (pH 8.5) and 0.02% (vol/vol) Tween-20. This solution is prepared in RNase/DNase-free water. Store it at room temperature for 6 months.
Tween-20 (10%, vol) stock Prepare 10% (vol/vol) Tween-20 from Tween-20 using water and then filter with a filter of 0.45-µm pore size. The solution is stored at room temperature for 6 months.
STRT template-switching oligo (TSO) These oligos (5′-AAGCAGTGGTATCAACGCAGAGTGCAG**U**GC**U**XXXXXXrGrGrG-3′) are used to introduce the bar code and upstream primer during RT. Each oligo carries a common primer sequence, a six-base bar code and three riboguanines to facilitate template switching. Furthermore, two uracil residues are included (marked in bold) in this oligo to facilitate the degradation of this oligo before PCR. This degradation prevents them from (cross-re)acting as primer during PCR. XXXXXX is the 6-bp error-correcting bar code. In all, 96 different bar codes are used for 96 different STRT TSOs (see **Supplementary Tables 1** and **2**). The bar codes are error correcting in the sense that no single-base substitution of a bar code can generate another valid bar code or a single-base substitution of another bar code. Thus, it takes two mutations to generate an invalid bar code, and a mutated bar code can be corrected by examining all possible single-base changes to find a valid bar code.
An example of a STRT TSO:
5′-AAGCAGTGGTATCAACGCAGAGTGCAG**U**GC**U**TTTAGGrGrGrG-3′
STRT-V3-T30VN oligo This oligo (5′-Bio-TTAAGCAGTGGTATCAACGCAGAGTCGACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3′) acts as a primer for capturing mRNA containing a polyA tail. ▲ CRITICAL This oligo must be 5′ biotinylated to prevent template switching at its 5′ end (which would generate useless background reads). The 3′ end of the DNA oligo contains VN where 'N' is a mix of all four bases and 'V' is a mix of 'A' or 'C' or 'G'. This VN sequence is important because it directs the primer to the beginning of the polyA tail. The underlined sequence in this oligo is the recognition site of restriction enzyme SalI.
STRT-PCR oligo This oligo (5′-Bio-AAGCAGTGGTATCAACGCAGAGT-3′) acts as a primer for amplification. ▲ CRITICAL This oligo must be 5′ biotinylated.
SOLEXA-P2 oligo This oligo (5′-CAAGCAGAAGACGGCATACGAG-3′) acts as a PCR primer in the library PCR stage.
P1-STRT oligo This oligo (5′-AATGATACGGCGACCACCGAGATCT **AAGCAGTGGTATCAACGCAGAGT**-3′) has two parts. The underlined part of the oligo contains the Solexa-P1 motif, required for sequencing, and the boldface part of the oligo is the STRT-PCR motif. It acts as a primer during the library PCR stage, and thus introduces the necessary P1 motif for sequencing.
STRT-SEQ-V3 oligo This (5′-AAGCAGTGGTATCAACGCAGAGTGCAGTG CT-3′) is the custom sequencing primer for Illumina sequencing.
▲ CRITICAL This is different from the conventional Illumina sequencer primer (which cannot be used).
SOLEXA-ADP2U adapter oligo The sequence of this adapter oligo is 5′-CAAGCAGAAGACGGCATACGAGCTCTTCCGATC**T**-3′.
▲ CRITICAL This oligo must have an overhanging 'T' at the 3′ end that will base-pair with the 'A' overhang of the cDNA.
SOLEXA-ADP2L adapter oligo The sequence of this adapter oligo is 3′-PHO-GTTCGTCTTCTGCCGTATGCTCGAGAAGGCTAG-PHO-5′.
▲ CRITICAL In this oligo, both ends are phosphorylated (PHO). 5′ end phosphorylation is necessary to ligate this end with 3′ end of the template. The 3′ end is modified to prevent the ligation from occurring in the wrong direction.
Adapter SOLEXA-ADP2, 10 µM Mix 10 µM SOLEXA-ADP2U and 10 µM SOLEXA-ADP2L in 10 mM Tris-HCl (pH 7.6) and 100 mM NaCl in RNase/DNase-free water. The solution is stored at −20 °C for 6 months.

PROCEDURE

Making STRT cell capture plates ● TIMING ~3 h

1| Clean the hood with RNAse*Zap* before starting to make cell capture plates.
▲ CRITICAL STEP The working area must be free from RNase to prevent degradation of RNA, and from DNA to prevent contamination by previously prepared samples. Experiments should be conducted in a dedicated single-cell laboratory under a UV-sterilized hood. Spray the pipette with RNAse*Zap* before starting the work.

2| Prepare a reaction mix as follows in a 15-ml Falcon tube—this is enough to prepare 11 cell capture plates:

Reagent	Volume	Final concentration
STRT-V3-T30 VN	27.5 µl	400 nM
STRT buffer (5×)	1,375 µl	1×
RNase/DNase-free water	4.9 ml	
Total	6.3 ml	

▲ CRITICAL STEP We use the Magnatrix 1200 liquid handling platform for pipetting in order to prepare the cell capture plates, but they can also be prepared manually.

3| Vortex the reaction mix vigorously.

4| Take a Greiner PP V-bottom 96-well plate and aliquot 57.5 µl of the reaction mix per well.

5| Now add 5 µl of 5 µM stock TSO (**Supplementary Table 1** and **Box 2**) to each well of the 96-well plate, for a final volume of 62.5 µl per well (for layout of template-switching oligos in the 96-well plate, see **Supplementary Table 2** and **Box 2**).
▲ **CRITICAL STEP** The stock TSO plate from **Box 2** must thaw completely before use. Change filter tips when adding TSO in each well.

6| Take a 96-well plate (an AbGene 0600 PCR plate) and aliquot 5 µl of reaction mix from Step 5 per well into the plate. This plate is called the 'cell capture plate'. Repeat the same procedure up to ten times to make 11 cell capture plates. Seal and store the plates at −20 °C for up to 6 months. Use adhesive PCR aluminum foil to seal the plates properly.
▲ **CRITICAL STEP** The plates must be properly sealed for long-term storage.

Collecting single cells on cell capture plate ● TIMING 1–3 h

7| Take out the cell capture plate from −20 °C and thaw the cell capturing solution.

8| Prepare a single-cell suspension for collection (different cells are prepared in different ways). After preparing the single-cell suspension, pass the cells through a cell strainer and resuspend them in 1× PBS. The cell strainer acts as a sieve with a 40-µm nylon mesh and is sterile. To obtain a more uniform single-cell suspension, pass the single-cell solution through the cell strainer using a 100-µl pipette. The cells collected after passing through this strainer are almost free from clumps and debris. Cells must be kept on ice until collected in the cell capture plate.
▲ **CRITICAL STEP** The time between preparing a single-cell suspension and cell collection should be as short as possible. It also varies between cell types. For example, cells such as midbrain dopaminergic neurons must be collected within 1 h. On the other hand, embryonic stem cells and fibroblasts stay in good condition for a longer time (2–3 h).

9| Remove the seal from the cell capture plate (from Step 7) just before the collection of single cells.
▲ **CRITICAL STEP** Adhesive PCR foil seals the plate very strongly, and thus it is very important to remove the seal carefully to avoid splashing. Before removing the seal, we recommend centrifuging the plate briefly to collect all the liquids at the bottom of the plate.

10| Collect single cells by fluorescence-activated cell sorting (FACS) or cell picker into the cell capturing plate from Step 9. By using FACS, unstained single cells can be identified and sorted according to their light-scattering properties. Alternatively, cells can be sorted on the basis of fluorescence after immunostaining or using an expressed fluorescent protein, which makes it possible to target specific subpopulations. Automated cell pickers based on micromanipulation under the microscope are also available (e.g., the CellEctor). Cells are then picked using a small capillary, and they can be selected on the basis of any feature that can be identified under a microscope, including size, morphology and fluorescence. We find that cell picking generally yields better results than FACS, but FACS is much quicker.

11| Seal the plate immediately after collecting the cells and freeze it immediately on dry ice.
▲ **CRITICAL STEP** The plate should be frozen even if it is to be used immediately, as this contributes to lysis. Seal the plate before placing it on dry ice, as the adhesive may fail at lower temperatures. Do not store the sealed plate for more than 2 weeks (the liquid may be lost to sublimation).

Lysis and RT ● TIMING ~3 h

12| Prepare the RT mix as follows:

Reagent	Volume (µl)	Final concentration
STRT-RT buffer (5×)	112	1×
DTT (20 mM)	112	2 mM
dNTP (10 mM)	112	1 mM
Mn^{2+} solution (1 M)	3.4	6 mM
Spike RNA (see **Box 1**)	5.0	
Superscript II (200 U μl^{-1})	11.0	2.5 U μl^{-1}
DNase/RNase-free H_2O	204.6	
Total	560	

▲ **CRITICAL STEP** Prepare everything in advance. Thaw all chemicals that are required for lysis and RT.
▲ **CRITICAL STEP** Including Mn^{2+} in the RT mix markedly improves the template-switching efficiency.

13| Let the cell capture plate (from Step 11) thaw at room temperature; wait until the last well has thawed, but do not wait longer. Vortex RT mix for 5 s and then spin down the liquid. Add 5 µl of RT mix to each well of the cell capture plate from Step 11. Mix each reaction mixture briefly by pipetting up and down once.
▲ **CRITICAL STEP** Change filter tips after every pipetting. It is important to mix the reaction mixture very slowly.

14| Immediately seal the cell capture plate and place the plate on ice. Before putting the cell capture plate on the MJ thermal cycler, spin down the plate (700*g* for 10 s at 4 °C) to collect all liquids at the bottom of the wells.
▲ **CRITICAL STEP** During pipetting, some wells might have bubbles and liquid dispersed inside the wall of the wells, which must be removed by centrifugation. However, the time between adding the RT mix to the plate and placing the plate on the thermal cycler should be as short as possible.

15| Transfer the plate to the thermal cycler and start the cycle for RT as follows. All the steps use the heated lid. The ramping speed of the PCR machine is 1.4 °C s^{-1}:

Cycle	Temperature (°C)	Time	Purpose
1	10	10 min	Annealing
2	42	45 min	Reverse transcription
3	4	Forever	Safe storage

Purification ● **TIMING ~3 h**
16| Take 100 µl of MyOne carboxylate beads in a 1.5-ml polyallomer tube. Capture the beads using a magnet and remove the supernatant carefully.

17| Resuspend the beads in 100 µl of elution buffer (EB), capture the beads with a magnet and remove and discard the supernatant.

18| Repeat Step 17 once more. Finally, resuspend the beads in 100 µl of EB and divide the beads into two 1.5-ml tubes, each containing 50 µl beads. Add 1,000 µl of 14% (wt/vol) PEG-6000 and 0.05% (vol/vol) Tween-20 in 0.9 M NaCl to each tube. This buffer is designed to capture fragments longer than about 100 bp (ref. 23). Resuspend the beads by pipetting up and down ten times.

19| After RT has ended, immediately pipette 20 µl of diluted beads from Step 18 to each well of the RT plate from Step 15 and mix the reaction mixture by pipetting.

20| Keep the plate at room temperature and incubate the reaction mixture for 20 min. Pool all the beads from the full plate into a new set of eight reservoir tubes using a multipipette. Thereafter, pool the beads from four reservoir tubes into a 1.5-ml polyallomer tube.
▲ **CRITICAL STEP** Collect every last drop of beads from the reservoir.

21| Capture the beads using a magnet held against the outside of the tube. Use a pipette to remove the supernatant from the tube without disturbing the beads.
▲ **CRITICAL STEP** The beads are visible as a brown precipitate. Wait for a long enough time to make sure that all the beads are attracted to the magnet. It takes ~10 min to collect all beads, but it depends on the quality of the magnet. We use strong neodymium magnets.

22| Repeat Step 21 with the other 1.5-ml polyallomer tube, collecting the beads into the same collection tube by using a pipette.

23| Add 1.5 ml of freshly prepared 70% (vol/vol) ethanol. Incubate the beads for 30 s in ethanol at room temperature. Remove the ethanol slowly using a 1-ml pipette.
▲ **CRITICAL STEP** Add ethanol by using a 1-ml pipette very slowly without disturbing the beads. Beads still must bind with the magnet. The best practice is to add 70% (vol/vol) ethanol by touching the pipette tip to the side of the tube and slowly releasing the ethanol.

24| Repeat Step 23 once more.

25| Let the beads dry in air with the lid open inside the hood, for around 10–20 min.
▲ **CRITICAL STEP** It is important to observe the pellet occasionally and not let the beads overdry. Cracks in the beads can be seen if they are overdried, and this may drastically reduce the solubility of the beads in EB.

26| Add 37 µl of EB and resuspend the beads. Incubate the beads for 5 min in EB. Finally, capture the beads and transfer the supernatant into a 1.5-ml microcentrifuge tube.
■ **PAUSE POINT** We do not recommend pausing the experiment at any other step before reaching this point. You can store the sample at −20 °C at this stage for up to 3 months.

Full-length cDNA amplification ● TIMING ~3 h
27| Prepare the reaction mix for full-length cDNA amplification. This reaction mix is prepared in a 0.2-ml PCR tube:

Reagent	Volume (µl)	Final concentration
Purified cDNA in EB (from Step 26)	37	
Advantage 2 PCR buffer (10×)	5	1×
dNTP (10 mM)	1	200 µM
STRT-PCR (10 µM)	1	200 nM
USER enzyme	5	
Total	49	

28| Mix the reaction mixture by pipetting up and down a few times without forming bubbles.

29| Incubate in a thermal cycler at 37 °C for 15 min using a heated lid, to allow USER enzyme to degrade the TSOs. Remove the PCR tube from the thermal cycler and immediately place it on ice. Wait until the temperature reaches 4 °C.

30| Now add 1 µl of Advantage 2 DNA polymerase mix (50×). Mix the enzyme properly using a pipette. Briefly spin down the PCR tube to collect all sample at the bottom of the PCR tube.

31| Perform the PCR as follows using a heated lid at 104 °C:

Cycle	Temperature (°C)	Time
1	95	1 min
Start 20 cycles		
2	95	15 s
3	65	30 s
4	68	4 min
End cycle		
5	4	Forever

32| Place the PCR tube on ice. Transfer 5 µl of PCR samples into new PCR tubes and keep the remaining PCR sample on ice.
▲ **CRITICAL STEP** This PCR sample can be stored at −20 °C for up to 3 months.

33| Add 45 µl of freshly prepared PCR mix to the 5-µl samples from Step 32. The PCR mix is as follows:

Reagent	Volume (µl)	Final concentration
DNase/RNase-free water	37.5	
Advantage 2 PCR buffer (10×)	5	1×
dNTP (10 mM)	1	200 µM
STRT-PCR (10 µM)	1	200 nM
Advantage 2 DNA polymerase mix (50×)	1	1×
Total	45	

34| Replace the PCR tube in the thermal cycler again for five more cycles. These extra five cycles of PCR enable the quality of the cDNA library to be checked. During optimization of this protocol, we found that after 20 cycles of PCR the amount of product is sufficient for downstream processing, but not enough for gel visualization. Therefore, to minimize amplification of the library, this aliquot is further amplified specifically for quality checking.

Cycle	Temperature (°C)	Time
1	95	5 min
Start 5 cycles		
2	95	15 s
3	65	30 s
4	68	4 min
End cycle		
5	4	Forever

Quality checking the cDNA library ● TIMING ~1 h

35| Purify the 25-cycle reaction from Step 34 using the Qiagen PCR purification kit according to the manufacturer's instructions, and elute it in 25 µl of EB.

36| Load the entire 25-µl sample on a 1.2% (wt/vol) agarose SYBR Safe e-gel and run it for 26 min according to the manufacturer's instructions. Check the gel in the Safe Imager blue-light transilluminator. If the library of cDNA looks good (**Fig. 2a**, lane 3), proceed to the next step.

? TROUBLESHOOTING

Binding amplified DNA with beads ● TIMING ~1 h

37| Take 20 µl of Dynabeads MyOne C1 streptavidin in a 1.5-ml polyallomer tube. Place the tube on the magnetic stand for at least 2 min or until the beads clearly separate from the liquid. Discard the supernatant and remove the tube from the

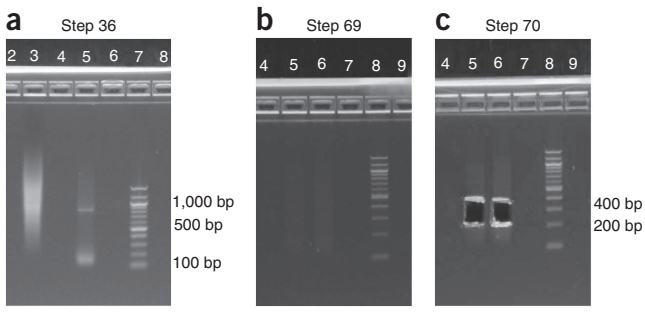

Figure 2 | Quality check of the cDNA library. (**a**) Gel picture from two libraries at Step 36, each from 96 cells and after 25 PCR cycles. In lane 3, the expected cDNA library covers the range between about 200 bp and >2,000 bp and the smear of cDNA is nearly uniform, indicating a good cDNA library. In lane 5, weak intensity or uneven smear in the gel indicates a poor cDNA library. In lane 7, the size marker is a 100-bp ladder with 500 bp and 1,000 bp emphasized (top two bands are 1,200 and 1,500 bp). (**b**) Gel picture after fragmentation, adapter ligation and 12 cycles of library PCR amplification, corresponding to Step 69. Lanes 5 and 6 contain the same PCR sample but split into two lanes. (**c**) The same gel picture (as in **b**) after excising, in Step 70, a gel slice between 200 and 400 bp. The same 100-bp ladder was used for all three gels run.

magnetic stand and add 50 µl of 2× BWT buffer. Resuspend the beads thoroughly by pipetting, and then place the tube again on a magnetic stand.

38| Repeat Step 37 twice and finally resuspend the beads in 25 µl of 2× BWT.

39| Add 25 µl of PCR-amplified DNA sample from Step 32 to 25 µl of beads (1:1 ratio) and mix them uniformly using a pipette. Incubate the reaction mix for 10 min at room temperature. This 10-min incubation is for binding the Dynabeads MyOne streptavidin with the specific biotinylated cDNA target. Keep the remaining 20-µl sample at −20 °C as a backup.

40| Place the tube on the magnetic stand to capture the beads on the magnet. Remove the supernatant very carefully. Remove the tube from the magnetic stand and add 50 µl of 1× BWT. Resuspend the beads thoroughly using a pipette and then place the tube again on the magnetic stand.

41| Repeat Step 40 two more times. Thereafter, repeat Step 40 two more times but use EBT buffer instead of 1× BWT buffer. Finally, resuspend the beads in 50 µl of EBT buffer.

Fragmenting the cDNA ● TIMING ~1.5 h

42| Set up the digestion reaction as follows and mix the reaction mixture carefully using a pipette:

Reagent	Volume (µl)	Final concentration
RNase/DNase-free water	39.5	
Fragmentase buffer (10×)	5	1×
BSA (100×)	0.5	1×
Total	45	

43| Place the tube from Step 41 onto the magnetic stand and remove the supernatant. Remove the tube from the magnet and add 45 µl of the above-mentioned reaction mix to the tube. Mix the reaction by pipetting without creating bubbles. Incubate the reaction mix on ice for 5 min.

44| Add 5 µl of dsDNA fragmentase to the reaction mix and mix it thoroughly using a pipette.
▲ **CRITICAL STEP** Before adding dsDNA fragmentase, mix the enzyme using a slow vortex and spin it down quickly.

45| Incubate the tube at 37 °C for 45 min.
▲ **CRITICAL STEP** To prevent the beads from settling, this reaction should be performed in a shaking incubator at 400 r.p.m. Alternatively, the tube can be manually mixed by pipetting at 15-min intervals to disperse the beads.

46| Place the tube on the magnetic stand and capture the beads on the magnet. Discard the supernatant. Resuspend the beads immediately with 50 µl of EBT buffer, and then place the tube again on the magnetic stand.

47| Repeat Step 46 three more times.
▲ **CRITICAL STEP** Ensure that all traces of dsDNA fragmentase are removed.

48| Finally, resuspend the beads in 25 µl of EBT buffer.
■ **PAUSE POINT** The protocol can be safely stopped in this stage. Store the sample at −20 °C for up to 3 months.

End repair ● TIMING ~1 h

49| Prepare the heating block at room temperature. Prepare the reaction mix for the end repair of DNA in a 1.5-ml polyallomer tube as follows:

Reagent	Volume (µl)	Final concentration
Beads containing DNA in EBT (from Step 48)	25	
NEBNext end repair buffer (10×)	5	1×
NEBNext end repair enzyme mix	2.5	
E. coli DNA ligase for fragmentase	0.5	
RNase/DNase-free water	17	
Total	50	

50| Mix the reaction well by pipetting slowly. Place the tube in a heat block and incubate the reaction for 30 min at room temperature.
▲ **CRITICAL STEP** The heat block should rotate at 400 r.p.m. to prevent the beads from sinking to the bottom of the tube. Alternatively, the tube can be manually mixed by pipetting at 15-min intervals to disperse the beads.

51| Place the tube on the magnetic stand to capture the beads and remove the solution. Add 50 µl of EBT buffer and resuspend the beads in EBT buffer by pipetting.

52| Repeat Step 51 once more to complete the washing. Finally, resuspend the beads in 21 µl of EBT.

dA tailing ● TIMING ~1 h
53| Prepare the heating block at 37 °C. Prepare the following reaction mix in a 1.5-ml reaction tube:

Reagent	Volume (µl)	Final concentration
Beads in EBT (from Step 52)	21	
NEBNext dA-tailing buffer (10×)	2.5	1×
Klenow exo$^-$ (5 U ml^{-1})	1.5	0.2 U µl^{-1}
Total	25	

54| Place the tube in the heat block and incubate the reaction mix for 30 min at 37 °C.
▲ **CRITICAL STEP** The heat block should rotate at 400 r.p.m. to prevent the beads from sinking to the bottom of the tube. Alternatively, the tube can be manually mixed by pipetting at 15-min intervals to disperse the beads.

55| Repeat Steps 51 and 52 and resuspend the beads in 25 µl of EBT.

Adapter ligation ● TIMING ~1 h
56| Prepare the heating block at 37 °C. Prepare the following reaction mix in a 1.5-ml reaction tube and mix the reaction properly by pipetting:

Reagent	Volume (µl)	Final concentration
Beads in EBT (from Step 55)	25	
NEBuffer 4 (10×)	5	1×
Adaptor SOLEXA-ADP2 (10 µM)	5	1 µM
T4 DNA ligase (5 U ml^{-1})	5	0.5 U µl^{-1}
SalI-HF (20 U ml^{-1})	2.5	2 U µl^{-1}
ATP (10 mM)	5	1 mM
RNase/DNase-free water	2.5	
Total	50	

▲ **CRITICAL STEP** Note that in this step only adapter 2 is ligated. In this way, the risk of generating adapter dimers is drastically reduced. Adapter 1 is introduced in the library PCR (below).

57| Repeat Steps 51 and 52, including an additional wash in EBT buffer to complete the washing, and finally resuspend the beads in 25 µl EBT.
▲ **CRITICAL STEP** It is very important to remove the liquid completely during washing. Completely removing adapter 2 helps prevent the formation of adapter dimers during PCR.

Library PCR amplification ● TIMING ~2 h
58| Prepare the PCR thermal cycler. Prepare the PCR mix in a 1.5-ml tube as follows:

Reagent	Volume (µl)	Final concentration
P1-STRT primer (10 µM)	10	1 µM
Solexa-P2 primer (10 µM)	10	1 µM
dNTP (10 mM)	2	200 µM
Phusion HF buffer (5×)	20	1×
RNase/DNase-free water	32	
Phusion polymerase (2 U ml^{-1})	1	0.02 U μl^{-1}
Total	75	

59| Place 25 µl of bead-immobilized DNA sample from Step 57 in a 200-µl PCR tube.

60| Add 75 µl of PCR mix into the PCR tube. Thereafter, distribute the PCR mix into two PCR tubes, each containing 50 µl. Immediately place the PCR tube in the thermal cycler.
▲ **CRITICAL STEP** Ensure that you start the PCR before the beads have a chance to settle at the bottom of the tube, which may inhibit amplification. After the first cycle of PCR, this is no longer crucial, as the amplified copy of DNA will be in solution.

61| Amplify by PCR using the following program with the heated lid:

Step	Temperature (°C)	Time
1	98	30 s
Start 12 cycles		
2	98	10 s
3	65	30 s
4	72	30 s
End cycle		
5	72	5 min
6	4	Forever

62| Transfer the PCR solutions into 1.5-ml microcentrifuge tubes. Place the tubes on a magnetic stand to capture the beads, and collect the solutions very carefully without disturbing the pellet.
▲ **CRITICAL STEP** All amplified DNA is now in the solutions. Accordingly, carefully collect all amplified DNA solution without disturbing the beads.

Purification ● TIMING ~30 min
63| Add 90 µl of Ampure beads to each 50 µl of PCR solution from Step 62 and mix by pipetting up and down ten times. Incubate the reactions for 5 min at room temperature.

64| Repeat Step 21.
▲ **CRITICAL STEP** Normally, Ampure beads take a longer time to bind to the magnet (~5 min).

65| Repeat Steps 23–25, using 200 µl of freshly prepared 70% (vol/vol) ethanol.
▲ **CRITICAL STEP** Drying times may vary substantially. Monitor the beads and do not overdry them, as this can drastically reduce their solubility in EB.

66| When the beads are dry, add 20 µl of EBT to each of the tubes. Pool the two reactions into a single tube and mix them together by pipetting. Incubate the reaction for 5 min at room temperature.

67| Place the tube on the magnetic stand to capture the beads. Collect the supernatant from the beads and transfer into a new 1.5-µl polyallomer tube.

68| Measure the concentration of DNA by Qubit according to the manufacturer's instructions.
▲ **CRITICAL STEP** The expected DNA concentration in this stage is 1–3 ng μl^{-1}.
■ **PAUSE POINT** You can keep the sample at −20 °C in this stage for up to 3 months.

Gel size selection ● TIMING ~1 h
69| Run the sample on a 2% (wt/vol) E-gel with SYBR Safe stain using a 100-bp DNA marker. Load 20 µl of sample from Step 67 into each of the two wells. The run time is 26 min. Fill the empty lane with water (according to the manufacturer's instructions). See **Figure 2b**.
▲ **CRITICAL STEP** To reduce the risk of cross-contamination, be very careful when loading, and leave one empty lane between each different library and next to the size marker.
? TROUBLESHOOTING

70| Break open the E-gel and place it on the Safe Imager transilluminator. Use a sterile scalpel to recover a gel piece between 200 and 400 bp and place it in a sterile 1.5-ml reaction tube. See **Figure 2c**.
▲ **CRITICAL STEP** A narrow size range may give better sequencing yield, but if the sample looks weak you can include up to 600 bp.

71| Purify on a single Qiaquick gel extraction column according to the manufacturer's instructions, but do not heat. Finally, recover in 30 µl of EB.
▲ **CRITICAL STEP** Do not heat to dissolve the gel in gel extraction medium, as you may lose AT-rich fragments. Instead, use a shaker at room temperature with vigorous shaking at 1,400 r.p.m. The gel will dissolve within 15 min of agitation.

72| Estimate the molar concentration. The molar concentration, needed to determine how much to load on the sequencer, can be estimated from the mass concentration (measured using Qubit in Step 68) and the fragment size distribution (from the gel image from Step 69). Alternatively, it can be directly measured using a Bioanalyzer or TapeStation (both from Agilent). The final expected concentration is 1–2 ng μl^{-1}, corresponding to about 1 nM.
▲ **CRITICAL STEP** The molar concentration determines the cluster density on the Illumina sequencing flowcell, and therefore the yield of sequencing. Every effort must be made to obtain as accurate an estimate as possible. If the estimate is too high, the yield will suffer proportionately. If it is too low, the sequencing flowcell may be overloaded and both the yield and sequence quality will drop markedly.

DNA sequencing ● TIMING ~3 d
73| Generate single-end sequencing reads from the final cDNA libraries by Illumina sequencing. Load the sample on multiple lanes to obtain a total of 200–500 million reads per 96-well library. We typically load 10–14 pM of each sample on our HiSeq 2000, and obtain 150–200 million reads per lane, but the optimal concentration will be different for individual instruments. Use only the custom sequencing primer STRT-SEQ-V3.
▲ **CRITICAL STEP** Do not use the standard Illumina sequencing primer; it will not work.

Data analysis ● TIMING ~1 d
74| Filter the reads output from the Illumina sequencer to remove low-quality reads according to the Illumina pass-filter score.

75| Remove reads that do not have a correct bar code, that lack the template switch GGG trinucleotide, that consist of less than six non-As followed by a polyA tail or that have an insert sequence shorter than 25 nt. The remaining reads are called 'Valid STRT' reads (**Fig. 3**).

76| Align the remaining sequences (which have the bar code and oligo-G) to the genome using the Bowtie short read aligner[24] while allowing for up to three mismatches. Bowtie is used because it is very fast.

77| Annotate the aligned reads to known transcript and repeat features with definitions obtained from the UCSC Genome Browser database[25]. Reads that align to exons and in the sense orientation are counted toward the expression value for that gene.

78| Finally, normalize the total hit count of every transcript as the number of reads per million mapped reads. This is different from the commonly used RPKM measure[26]: we do not divide by the gene length, because each mRNA gives rise to a single sequenceable fragment near its 5′ end. Please note that a more detailed description of the data analysis pipeline will be published elsewhere, and the software is freely available from the authors.

? TROUBLESHOOTING

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

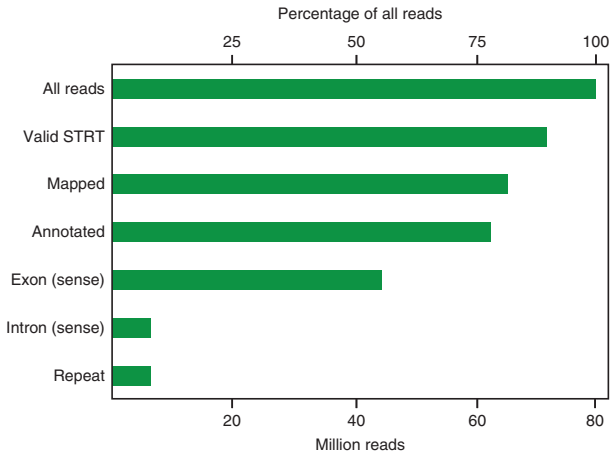

Figure 3 | Typical distribution of reads in a successful sample of 96 single cells. The bars show millions of reads (below the figure) and percentages of all reads (above the figure), assigned to different categories during data analysis. Percentages are given relative to all reads.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Possible solution
36	Smear of cDNA does not look uniform or cDNA library does not cover the range of expected size (200–2,000 bp)	Cells might be dead or dying before the reverse transcription reaction	Keep the time between making a single-cell suspension and collecting the last cells as short as possible (<60 min)
			Use collected cells as soon as possible. We do not recommend keeping them for more than 2 weeks at −80 °C
		Cells collected by FACS may not have hit the liquid in the wells (e.g., they may have ended up on the tube walls)	Make sure that the FACS sorter accurately hits the liquid of the wells, or use a manual or automated cell picker to make sure that single cells are properly collected
		The template-switching reaction might be inhibited by medium during cell collection	When making single-cell suspension, make sure to wash cells repeatedly with PBS to remove all the substances that might inhibit reverse transcription reaction
		If there is little mRNA in the sample because of fewer or smaller cells, the primer may become dominant in the reaction and give primer dimer after PCR	Collect cells again
69	No smear at all or smear looks very weak or smear is concentrated around 100 bp	Too little starting material (few or small cells)	Check the cDNA library after the first PCR, which indicates the starting material quantity
		cDNA is lost during washing of beads	Perform washing steps very carefully. Bind the beads in the magnet for a longer time
		The number of cycles (12) is insufficient to yield the least visible amplification in the gel	If needed, use 13 or 14 cycles of amplification to obtain enough material for sequencing
	The smear looks as long as it did after the Step 33	Fragmentation step is not working	Use fresh fragmentation enzyme and redo the experiment using the saved half of the sample from Step 36

● TIMING

Day 1

Steps 1–6, making cell capture plates: 3 h

Steps 7–11, collecting single cells: 1–3 h

Day 2

Steps 12–15, lysis and RT: 3 h

Steps 16–26, purification: 3 h

Steps 27–34, full-length cDNA amplification: 3 h

Steps 35 and 36, quality checking the cDNA library: 1 h

Day 3

Steps 37–41, binding amplified DNA with beads: 1 h

Steps 42–48, fragmentation by fragmentase: 1.5 h

Steps 49–52, end repair: 1 h

Steps 53–55, dA tailing: 1 h

Steps 56 and 57, adapter ligation: 1 h

Steps 58–62, library PCR amplification: 2 h

Steps 63–68, purification on Ampure beads: 30 min

Days 4–7

Steps 69–72, gel size selection: 1 h

Step 73, sequencing: 3 d

Steps 74–78, data analysis: 1 d

Box 1, preparing spike-in solution, 2 h

Box 2, preparing the TSO stock plate, 1 h

ANTICIPATED RESULTS

Step 36

This is one of the crucial steps of this experiment. The end result is dependent on the outcome of this step. **Figure 2a** shows two libraries, each from 96 cells and after 25 PCR cycles. A successful cDNA library (**Fig. 2a**, lane 3) covers the range between about 200 bp and >2,000 bp, and the smear of cDNA should be nearly uniform. Occasionally, a few bands will be observed in the smear, indicating the presence of some abundant housekeeping gene in the library. These bands should occur at different sizes in different cell types. There should be no strong smear in the lower part of the gel (below 200 bp), which would indicate the presence of primer or adapter dimers in the cDNA library. Weak intensity or uneven smearing in the gel indicates a poor cDNA library (**Fig.2a**, lane 5).

Step 69

In this step, we obtain the final cDNA library for sequencing after adapter ligation. We recommend using the least number of PCR cycles needed to obtain a visible smear in the gel. After optimization of Step 61, we found that most of the time 12 cycles yield just enough material for sequencing (**Fig. 2b**). The expected size of DNA range covers between 150 bp and 1,200 bp, concentrated around 200–400 bp. You should not see any bands in this smear. The gel is cut to recover between 200 and 400 bp; this is followed by gel extraction to obtain the sequenceable cDNA library (**Fig. 2c**). The length of the excised region should be used to estimate the library molar concentration for sequencing.

Step 78 (end result)

Figure 3 shows the typical distribution of reads in a successful sample of 96 single cells. The bars show millions of reads. A total of nearly 80 million reads were generated ('All reads'), of which 90% had a valid bar code and template-switching sequence ('Valid STRT'). Of the total, 82% were mapped to the reference genome, and 79% were mapped to some annotated genomic feature (an exon or a repeat); 56% of all reads were mapped to exons (or splice junctions) in the correct (sense) orientation, whereas 8% mapped to introns. The latter could be explained in some cases by the existence of unannotated alternative exons; 9% of all reads were mapped to repeats, many of which represented expressed repeat families such as *B2_Mm1*. Only reads assigned to exons (sense) contribute to the quantitative gene expression profile. A large drop in the percentage of 'Valid STRT' would indicate the presence of primer-dimer artifacts or other background amplification lacking proper bar codes. A drop for 'exon (sense)' would indicate the possible presence of genomic DNA contamination or poor-quality degraded mRNA. We typically accept libraries with 'Valid STRT' better than 70% and 'exon (sense)' better than 30%.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS We thank P. Sekyrova (Karolinska Institutet) for providing cultured cells. This work was supported by grants from the Swedish Foundation for Strategic Research (MDB09-0052) and from the European Research Council (261063).

AUTHOR CONTRIBUTIONS S.I. developed the method and performed optimization experiments, analyzed data and co-wrote the manuscript; U.K. performed initial experiments; A.M. performed cell culture and FACS sorting; P.Z. performed experiments to optimize the method; J.-B.F. performed sequencing experiments and advised on method development; P.L. developed the bioinformatics pipeline and analyzed data; S.L. conceived of the method, supervised the development, developed the bioinformatics pipeline and co-wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.natureprotocols.com/.

Published online at http://www.natureprotocols.com/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862 (2009).
2. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
3. Li, H.H. *et al.* Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335**, 414–417 (1988).
4. Kurimoto, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* **34**, e42 (2006).
5. Esumi, S. *et al.* Method for single-cell microarray analysis and application to gene-expression profiling of GABAergic neuron progenitors. *Neurosci. Res.* **60**, 439–451 (2008).
6. Taniguchi, K., Kajiyama, T. & Kambara, H. Quantitative analysis of gene expression in a single cell by qPCR. *Nat. Methods* **6**, 503–506 (2009).
7. Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
8. Tang, F. *et al.* mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
9. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
10. Salimullah, M., Sakai, M., Plessy, C. & Carninci, P. NanoCAGE: a high-resolution technique to discover and interrogate cell transcriptomes. *Cold Spring Harb. Protoc.* published online doi:10.1101/pdb.prot5559 (2011).
11. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
12. Chubb, J.R., Trcek, T., Shenoy, S.M. & Singer, R.H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006).
13. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).
14. Kurimoto, K., Yabuta, Y., Ohinata, Y. & Saitou, M. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat. Protoc.* **2**, 739–752 (2007).
15. Morris, J., Singh, J.M. & Eberwine, J.H. Transcriptome analysis of single cells. *J. Vis. Exp.* published online doi:10.3791/2634 (2011).
16. 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).
17. Levin, J.Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**, 709–715 (2010).
18. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).
19. Maheshri, N. & O'Shea, E.K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434 (2007).
20. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6**, 468–478 (2010).
21. Tang, F. *et al.* Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS ONE* **6**, e21208 (2011).
22. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* **29**, 1120–1127 (2011).
23. Lundin, S., Stranneheim, H., Pettersson, E., Klevebring, D. & Lundeberg, J. Increased throughput by parallelization of library preparation for massive sequencing. *PLoS ONE* **5**, e10029 (2010).
24. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
25. Fujita, P.A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011).
26. 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).