CHAPTER FIVE

A Strand-Specific Library Preparation Protocol for RNA Sequencing

Tatiana Borodina, James Adjaye, *and* Marc Sultan

Contents

Abstract

The analysis of transcriptome, which was over the past decade based mostly on microarray technologies, is now being superseded by so-called next generation sequencing (NGS) systems that changed the way to explore entire transcriptome. RNA sequencing (RNA-Seq), one application of NGS, is a powerful tool, providing information not only about the expression level of genes but also further about the structure of transcripts as it enables to unequivocally identify splicing events, RNA editing products, and mutations in expressed coding sequences within a single experiment. Herein, we describe step by step the deoxy-UTP (dUTP) strand-marking protocol [Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37**(18), e123], which has been recently reviewed as the leading protocol for strand-specific RNA-Seq library preparation [Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A., Regev, A. (2009). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**(9), 709–715]. The procedure starts with the isolation of the polyA fraction (mRNA) within a pool of total RNA, followed by its fragmentation. Then double-stranded (ds) cDNA synthesis is performed with

Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Ihnestrasse 63-73, Berlin, Germany

Methods in Enzymology, Volume 500
ISSN 0076-6879, DOI: 10.1016/B978-0-12-385118-5.00005-0

the incorporation of dUTP in the second strand. The ds cDNA fragments are further processed following a standard sequencing library preparation scheme tailored for the Illumina sequencing platform: end polishing, A-tailing, adapter ligation, and size selection. Prior to final amplification, the dUTP-marked strand is selectively degraded by Uracil-DNA-Glycosylase (UDG). The remaining strand is amplified to generate a cDNA library suitable for sequencing.

1. Introduction

1.1. NGS

The next generation sequencing (NGS) has revolutionized molecular biology and is currently one of the most promising and fastest developing technologies. Since the first instrument from Roche appeared on the market in 2004, enabling the sequencing of thousands of DNA fragments of 100 base pair (bp) in parallel, the throughput and quality of the generated sequences has continuously increased. Today's Genome Sequencer (FLX Titanium) from Roche is able to generate about 1 million of 400 bp sequences per run. This system belongs to the so-called long-read technologies. Though it is "next generation" in terms of throughput (0.4–0.6 Gb per run), it is only ~20 times cheaper than the conventional Sanger sequencing approach. The major breakthrough came from the so-called short-read sequencing technologies, which widely broadened sequencing applications and significantly reduced the sequencing costs. In 2006, Solexa (acquired by Illumina in 2007) was the first company to introduce a short-read sequencing system to the market, based on sequencing by synthesis approach. A year later, the ligation-based SOLiD system from ABI appeared on the market. Both systems have considerably evolved over the passed years, increasing the throughput, optimizing the machine operation, and constantly decreasing the reagents costs (up to twofold per year so far). Currently, both the Illumina and SOLiD systems are comparable in terms of sequencing costs and throughput: ~25 Gb of mappable reads per day at ~70€ per Gb for the Illumina HiSeq2000 and ~6 Gb of mappable reads per day at ~45€ per Gb for SOLiD4 systems[1] (180 and 280 times cheaper than Sanger sequencing, respectively). Both systems have similar applications spectrum and are now widely used for genomic *de novo* and resequencing, transcriptome analysis, epigenetics studies, genotyping, etc. The two platforms have made NGS an affordable tool for routine laboratory research and rendered challenging large-scale projects like the 1000 genomes project, possible.

[1] The throughput and costs information are derived from the official announcements of the respective companies.

Recently, authors began to distinguish second generation sequencing (Roche, Illumina and ABI platforms) and third generation single molecule sequencing technologies. The single molecule sequencer Heliscope from Helicos Bioscience is already on the market. Several platforms are being developed and should be available soon on the market (e.g., sequencers from Pacific Biosciences and Oxford Nanopore).

1.2. RNA sequencing

Among all NGS applications, high-throughput complementary DNA (cDNA) sequencing (RNA-Seq) is probably one of most challenging. RNA-Seq is a powerful tool for whole-transcriptome analysis, providing information not only on the expression level of genes but also on the structure of transcripts as it enables the identification of splicing events and RNA editing products, mutations in expressed coding sequences, and allele-specific transcription. The RNA-Seq procedure itself is straightforward and has a large dynamic range as well as high sensitivity. RNA-Seq has clear advantages over previous high-throughput approaches, such as microarray hybridization, gene-specific and tiling arrays, or SAGE analyses (Stoughton, 2005; Velculescu *et al.*, 1995). In contrast to SAGE, RNA-Seq does not depend on the presence of specific restriction sites within the cDNA. Unlike hybridization-based approaches, RNA-Seq does not require prior information on transcript sequences, allowing the detection of novel transcripts. It has a very low background noise and provides a dynamic range ($\sim 10^5$) typically an order of magnitude higher than one can achieve with hybridization-based arrays. The digital character of RNA-Seq data allows easily comparing and integrating results obtained from different laboratories. The applications and advantages of RNA-Seq have already been shown in several studies (e.g., Denoeud *et al.*, 2008; Marioni *et al.*, 2008; Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008; Sultan *et al.*, 2008; Wilhelm *et al.*, 2008). An exhaustive review about RNA sequencing was recently published by Costa *et al.* (2010).

The RNA-Seq procedure can be divided into four main steps: (i) preparation of the sequencing library, (ii) clonal amplification of library molecules on the sequencing support (glass surface in the case of the Illumina platform; magnetic beads in Roche and ABI schemes), (iii) the sequencing itself, and (iv) data analysis. The second and the third steps are sequencer-dependent and are therefore highly standardized. Library preparation and data analysis allow more freedom in adjusting the protocol for a particular purpose leading to a great variety of methods, where standards and guidelines are still evolving. In this chapter, we discuss the first step of the RNA-Seq procedure, which is the preparation of the cDNA library.

Until recently, the common strategy of an RNA-Seq library preparation was to convert single-stranded RNA molecules into double-stranded (ds)

DNA fragments of a certain size flanked by platform-specific adapter sequences. Originally, RNA-Seq was not strand specific, that is, reads corresponding to original RNA molecule or complementary to it were indistinguishable, which created difficulties for the data analysis. Transcription can occur in both directions, genes being located on either DNA strand, sometimes overlapping. In eukaryotes, a complementary RNA molecule to a given mRNA can also be transcribed: this has been described as antisense transcription and these molecules are involved in regulatory mechanism (He *et al.*, 2008; Kapranov *et al.*, 2007). Knowing from which DNA strand the RNA molecule originates from is an important piece of information, which helps resolving annotation ambiguities for known and novel genes, provides hints to the function of the studied RNA, and helps to correctly predict the expression levels of a given transcript.

Several strand-specific RNA-Seq (ssRNA-Seq) protocols based either on the ligation of adapters to the RNA molecules or on modifications of the ds cDNA synthesis procedure were suggested recently. Schemes that involve adapter ligation directly to single-stranded RNA molecules (Lister *et al.*, 2008) have no restrictions in the RNA length and are the only choice for the analysis of short RNA molecules like micro-RNA. This principle is used in many commercial preparation kits,[2] but these methods are laborious and are sensitive to ribosomal RNA (rRNA) contamination, so the RNA fraction of interest (e.g., mRNA, micro-RNA, or short transcripts) must be preselected.

In alternative directional transcriptome profiling protocols, the ds cDNA synthesis procedure is modified. Cloonan and colleagues used a tagged random hexamer primer in the synthesis of the first cDNA strand and DNA–RNA template-switching primers during the synthesis of the second cDNA strand (Cloonan *et al.*, 2008). Another approach is based on substituting all cytidine residues in RNA molecules to uridines by bisulfite treatment prior to cDNA synthesis (He *et al.*, 2008). Thus, the first strand contains no guanines, while the second strand has no cytidines. This method requires a tailored data analysis scheme and leads to the loss of $\sim$30% of sequences that can be mapped uniquely to the genome, as the transformation of a four-bases code to a three-bases code partly results in skewing of the complexity of the genome under investigation. The directional RNA-Seq procedure described herein is based on the incorporation of deoxy-UTP (dUTP) during either the first- or second-strand cDNA synthesis and the subsequent selective destruction of this strand (Parkhomchuk *et al.*, 2009).

Recently, Levin and coworkers compared seven ssRNA-Seq protocols in terms of strand specificity, library complexity, evenness and continuity of coverage, agreement with known annotations, and accuracy for expression

[2] For example, DGE small RNA Sample Preparation Kit (Illumina); directional mRNA-Seq library preparation kit (Illumina); SOLiD small RNA expression kit (Applied Biosystems); total RNA-Seq kit (Applied Biosystems).

level analysis (Levin *et al.*, 2009). Based on their conclusion, the dUTP approach (Parkhomchuk *et al.*, 2009) provided the most compelling overall balance across criteria, followed closely by the Illumina RNA ligation protocol (Directional mRNA-Seq Library Preparation Kit, Illumina). Despite a wide range of strand-specific protocols, many researchers still often rely on nonstrand-specific RNA-Seq library preparation, where RNA is first converted into ds cDNA, and then processed into a sequencing library, mostly because it is simple, straightforward, and stable (Wilhelm *et al.*, 2010). In this respect, the dUTP strand-marking method is a good alternative, as it adds only minor modifications to the standard cDNA synthesis workflow, and is compatible with commercially available kits.

2. ssRNA-Seq Protocol

In this chapter, we present the detailed protocol for the preparation of an ssRNA-Seq library from mRNA suitable for Illumina-based sequencing. The protocol relies on the incorporation of dUTP during the second-strand synthesis, allowing subsequent selective destruction of this strand by Uracil-DNA-Glycosylase (UDG; Fig. 5.1).

The initial step of this part is to generate a cDNA molecule to the initial mRNA, in a reaction known as first-strand cDNA synthesis. The results of this reaction are DNA–RNA hybrids, which will serve as templates for the generation of the second cDNA strand. During the latter reaction, the RNA is degraded and the second cDNA strand is synthesized using a nucleotide mixture in which dTTP has been replaced by dUTP. The next steps enable the ligation of adaptor sequences on both ends of the ds cDNAs. The adaptor sequences have a Y-shaped structure, which preserves the directionality of the molecules. Finally, the second strand is being degraded using an enzyme called UDG that selectively removes the incorporated uracil from the DNA. The remaining intact first-strand cDNAs complemented with Y-shaped adapter sequences enable the directional sequencing of all molecules while preserving the strand information of the RNA.

The protocol proceeds over 3.5 days and includes several steps, which are detailed below and in Fig. 5.2.

Day 1:. mRNA purification from total RNA, DNase treatment, RNA fragmentation.

Day 2:. Synthesis of ds cDNA with incorporation of dUTP within the second strand.

Day 3:. Preparation of the Illumina Paired End sequencing library (end polishing, A-Tailing, adapter ligation, size selection), degradation of the dUTP-marked strand.

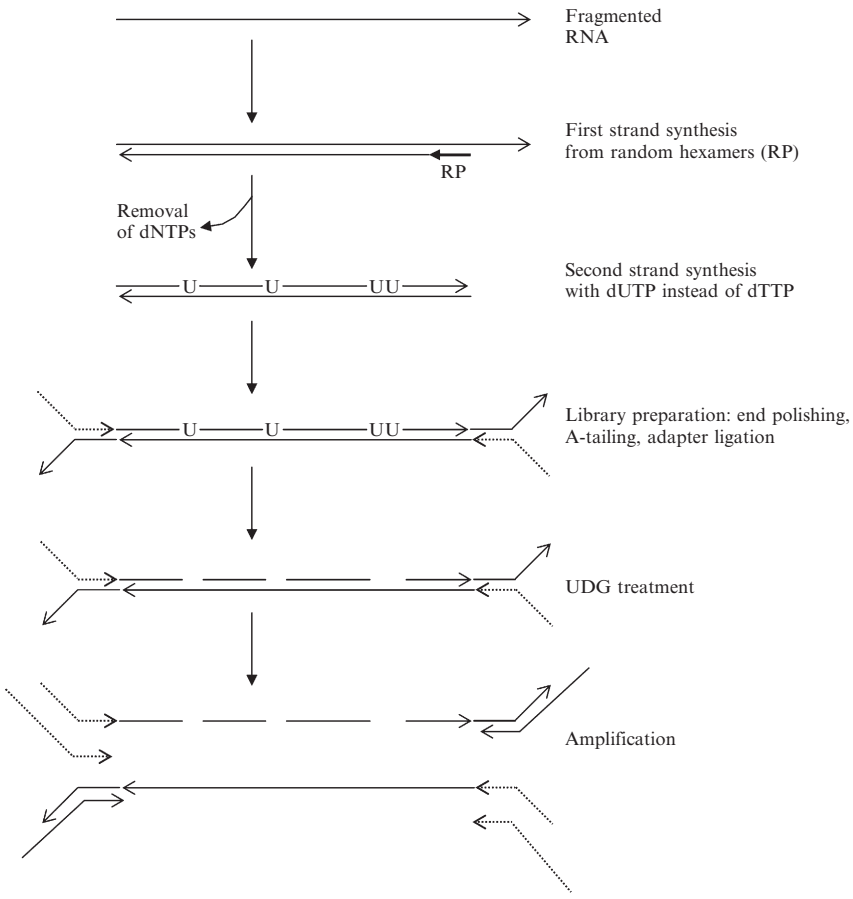

Figure 5.1 **Principle of the dUTP strand-marking ssRNA-Seq method.**

Day 4:. Amplification of the cDNA library.

All protocols presented below with a possibility of automatic calculation for different volumes of reagents are available online (Zbio-wiki, 2010).

2.1. General issues

2.1.1. Selection of fragmentation approach

The presented protocol exploits chemical RNA fragmentation before cDNA synthesis. Alternatively, the ds cDNA can be first synthesized, and then sheared (preferably with ultrasound). Both procedures work well with the dUTP strand-marking method.

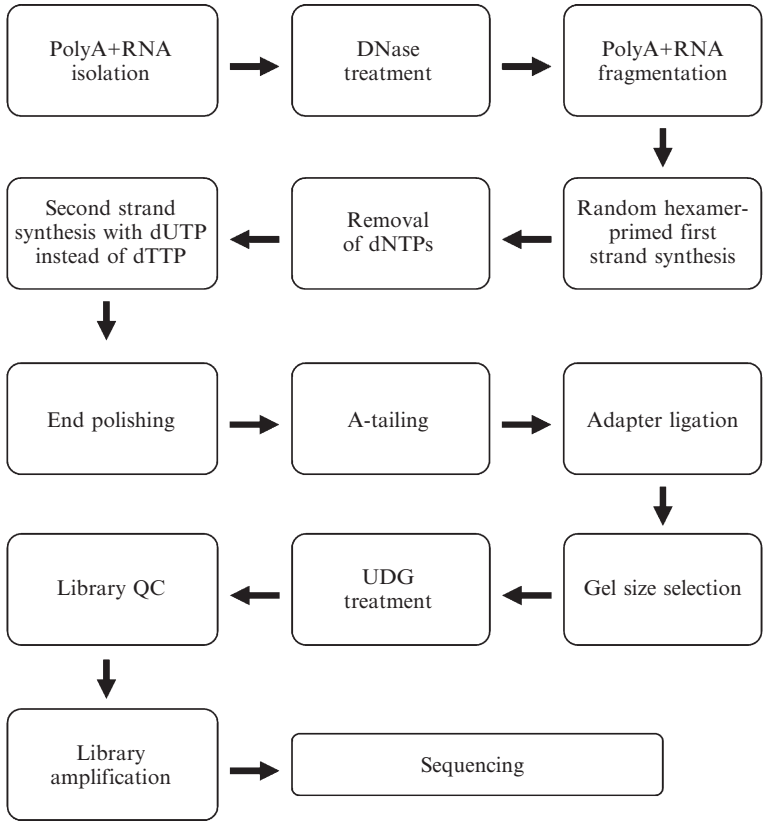

Figure 5.2 **Workflow of the Illumina PE sequencing library preparation integrating the dUTP strand-marking method.**

The advantage of the RNA fragmentation approach is a significantly more even reads distribution along the transcript. In the cDNA shearing scheme, RNA secondary structures distort locally the cDNA synthesis, which leads to nonuniform reads distribution along the RNA and, if oligo (dT) primer was used for the first-strand synthesis together with random primers, to a 3′-bias. Also, the RNA fragmentation is much quicker.

The cDNA shearing scheme is advantageous for samples where rRNA is highly represented (e.g., preselection of polyA$^+$RNA fraction is not possible like in prokaryotic transcriptomes). The rRNA has strong secondary structures which suppress cDNA synthesis. For example, if cDNA is synthesized from total RNA, where the rRNA fraction constitutes 85–95% of the material, ribosomal reads would be ~50% of the sequencing output. After one round of polyA$^+$RNA isolation on Oligo d(T)$_{25}$ Dynabeads, about 50% of the RNA washed off from the Dynabeads is rRNA. Using the

cDNA shearing approach would result in just a small percent of rRNA reads in the sequencing data. For RNA fragmentation, it is obligatory to perform an additional round of polyA$^+$RNA isolation to get rid of rRNA.

If compared to RNA chemical fragmentation, ultrasonic shearing of cDNA is a safer procedure in terms of over digestion. Ultrasound gives 150–250 bp fragments size distribution. This range would not change if the shearing time is increased. Chemical over digestion, or fragmentation of partly degraded samples, would lead to formation of smaller than necessary fragments (below 100 bp instead of ~170 bp expected).

When performing the protocol, it is necessary to consider that shorter RNA fragments require higher concentration of random primers for an effective primer extension: 2.4 per 0.5 μg of RNA for the RNA fragmentation approach and 40 ng per 0.5 μg for the cDNA shearing scheme.

2.1.2. Required materials

Here, the items are listed which are necessary throughout the whole protocol.

- Water (Ambion, e.g., # AM9930)
- DNA LoBind 1,5-ml tubes (Eppendorf, #0030.108.051)
- DNA LoBind 0.5-ml tubes (Eppendorf, #0030.108.035)
- 0.2 ml thin-wall tubes (e.g., Applied Biosystems, #N8010540)
- Filtered tips (e.g., Axygen, Biozym)

Note 1: We recommend to avoid using DEPC-treated reagents for the protocol: DEPC may slightly inhibit the cDNA synthesis.

Note 2: Powder-free gloves should be worn when performing the protocol.

2.2. Purification of polyA$^+$RNA

The first step of an RNA-Seq experiment is to extract good quality intact RNA from the tissue or cells of interest. For this purpose, one of many standard protocols for total RNA extraction (Trizol, Qiagen Kits, etc.) can be employed. Depending on the task, the sequencing library can be prepared either from total RNA or from a distinct RNA fraction (e.g., mRNA, noncoding RNA, small RNA, etc.). Here, we focus on the selective extraction of the polyadenylated RNA fraction (polyA$^+$RNA or mRNA) as it is the most commonly used RNA fraction for gene expression studies. This protocol was also successfully applied for the sequencing ribo minus[3] RNA samples.

[3] Ribo minus RNA is total RNA treated with, for example, RiboMinus™ Eukaryote Kit for RNA-Seq (Invitrogen, SKU# A10837-08).

2.2.1. Required materials

Equipment

- Water bath: 65, 70, 37 °C
- Magnetic stand for 1.5-ml Eppendorf tubes (e.g., Applied Biosystems, #AM10055)
- Rotator for 1.5-ml Eppendorf tubes (room temperature, ~10 rpm required)
- PCR thermal cycler for 0.2-ml tubes
- Centrifuge for 1.5–2 ml tubes (4 °C, 14,000 rpm required)

Reagents/enzymes

- Dynabeads® mRNA Purification Kit for mRNA purification from total RNA preps (Invitrogen SKU# 610–06)
- TURBO™ DNase 2 u/μL (Ambion, #AM2238)
- 10× Fragmentation Reagent and Stop Solution (Ambion, #AM8740) or, alternatively, 5× Fragmentation Buffer and Fragmentation Stop Solution (included in the Illumina mRNA-Seq 8-Sample Prep Kit, #RS-100-0801). Stop Solution is 0.2 M EDTA, pH 8.0
- Glycogen, 10 μg/μl
- 3 M Sodium acetate, pH 5.2
- Ethanol, 100%
- 10 mM Tris–Cl, pH 8.5
- RNeasy MinElute Cleanup Kit (Qiagen, #74204)

Additional material

- Intact total RNA sample of high quality (RIN $\geq$8 measured on an Agilent Bioanalyzer). We recommend starting with 30–50 μg of total RNA. It is possible to go down to 10 μg.

2.2.2. Isolation of polyA$^+$RNA

Preparation of Oligo d(T)$_{25}$ Dynabeads:

1. Resuspend Dynabeads by vortexing. Transfer 0.5 μl of suspension per 1 μg of total RNA (minimum 50 μl of Oligo d(T)$_{25}$ Dynabeads) to a 1.5-ml LoBind tube.
2. Vortex the tube for 30 s and then keep the tube on the magnetic stand for 1 min. Remove the supernatant.
3. Resuspend the beads in Binding Buffer—take the same volume as the volume of Dynabeads suspension.
4. Repeat steps 2 and 3.

Preparation of the total RNA sample and binding polyA$^+$RNA to Dynabeads:

1. Transfer your RNA to the LoBind tube. Place the sample at 65 °C for 2 min to denature the secondary structures, then place the sample immediately on ice.
2. If necessary, add Tris–HCl, 10 mM pH7.5 or Binding Buffer.

   Note: Optimal hybridization conditions are obtained when $V_{\mathrm{RNA}} = V_{\mathrm{Dynabeads}}$, where V is the volume. If the RNA sample volume is smaller than the volume of the Dynabeads suspension, fill it up with Tris–HCl, 10 mM pH7.5. If the RNA sample volume is larger, then add Binding Buffer.
3. Combine the Dynabeads and the total RNA sample, mix carefully by inverting the tube several times.
4. Rotate the tube for 30–45 min at room temperature at ~10 rpm.
5. Put the tube on the magnetic stand for 1 min. Transfer the supernatant to a new LoBind tube.

   Note: We advise the retention of the non-polyA$^+$RNA fraction at least until it is clear that the polyA$^+$RNA fraction has been successfully isolated.
6. Wash the Dynabeads twice with Washing Buffer (volume = $V_{\mathrm{RNA}} + V_{\mathrm{Dynabeads}}$).
7. Resuspend beads in 50 µl of Tris–HCl, 10 mM pH 7.5. Place the tube in a 70 °C water bath for 2 min (to wash off RNA from beads), then put immediately on ice. Add 50 µl of Binding Buffer and mix carefully by inverting the tube. Repeat steps 4–6.

   Note: Second round of mRNA purification on Dynabeads is obligatory for the RNA fragmentation approach and optional for cDNA shearing approach.

polyA$^+$RNA elution from Dynabeads:

1. Resuspend the Dynabeads in 10 µl of 10 mM Tris–HCl, pH7.5.
2. Place the tube in the 70 °C water bath for 2 min.
3. Heat the magnetic stand in the 70 °C water bath for ~5 min, take it out shortly before taking out the tube with Dynabeads.

   Note: Be careful not to heat the magnetic stand to 80 °C or above, it would loose its magnetic properties.
4. Place the tube with Dynabeads on the preheated magnetic stand for ~15 s, collect the eluate into a new 0.5-ml LoBind tube.
5. Repeat steps 1–4.
6. Combine the eluates from step 3. Measure the concentration of the polyA$^+$RNA.

   Note: Typically polyA$^+$RNA constitutes about 1–5% of total RNA.

7. For safe storage, snap freeze the tube with polyA$^+$RNA in liquid nitrogen, place at −20 °C for short-term storage, otherwise store the tube at −70 °C.

2.2.3. DNase treatment

DNase treatment should be performed systematically for all RNA-Seq samples to minimize nonspecific sequencing reads, which might also interfere with strand specificity and increase background noise. We recommend you to perform DNase treatment after mRNA purification because the RNA might be partly fragmented during the DNase treatment, which would lead to an underrepresentation of 5′ ends of mRNA. We further recommend the use of TURBO^{TM} DNase (Ambion), as it is much more efficient for low concentration of DNA in solution, than DNase from other vendors.

1. Prepare the reaction mix: RNA (up to ~0.5 μg/μl), 1× TURBO^{TM} DNase Buffer, TURBO^{TM} DNase (take 0.2 U of TURBO^{TM} DNase per 1 μg of RNA, minimum—0.2 U per reaction).
2. Incubate the reaction at 37 °C (water bath) for 30 min.
3. Proceed immediately to the RNA cleanup. We recommend the RNeasy MinElute Cleanup Kit.
4. Elute the RNA in RNase-free water. Check the sample concentration (optional).
5. For safe storage, snap freeze the tube with polyA$^+$RNA in liquid nitrogen, place at −20 °C for short-term storage, otherwise store the tube at −70 °C.

2.2.4. RNA fragmentation

Chemical RNA fragmentation yields ~170 nt RNA fragments

1. Preparation of the reaction mix: RNA (up to ~2.5 μg/μl) in 1× Fragmentation Buffer.
2. Incubate the reaction at 70 °C (in a preheated PCR thermal cycler) for 5 min. Then place the tube on ice and add 2 μl of Stop Solution.

   Note: If the Illumina Fragmentation Buffer is used, incubate at 94 °C for exactly 5 min.
3. Transfer the reaction into a LoBind tube.
4. Perform RNA precipitation with salt/2.5 V ethanol. We use 0.3 M NaOAc, pH 5.2 as a salt component. If the amount of RNA is less than 1 μg, add 20 μg of glycogen as a coprecipitant. Incubate ~30 min at −20 °C before centrifugation.

5. Elute the RNA in RNase-free water. Check the sample concentration (optional).

2.3. cDNA synthesis

2.3.1. Required materials

Equipment

- PCR machine for 0.2-ml tubes
- Water bath: 16 °C

Reagents/enzymes

- SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, #18080-051)
- Actinomycin D (Sigma, #A1410), 120 ng/μl
- 1 mM Tris–HCl, pH7.0
- dNTP Set, 100 mM each A,C,G,T (GE Healthcare, #28-4065-51)
- dUTP, 100 mM (GE Healthcare, #28-4065-41)
- Second-Strand Buffer (Invitrogen, #10812-014)
- *Escherichia coli* DNA ligase (New England BioLabs, # M0205)—Optional
- DNA Polymerase I (*E. coli*) (New England BioLabs, #M0209)
- RNase H (Invitrogen Cat #18021–071)
- Illustra ProbeQuant G-50 Micro Columns (GE Healthcare, #28-9034-08)
- QIAquick PCR Purification Kit (Qiagen, #28106)

2.3.2. First-strand synthesis

First-strand synthesis is performed in 1× reverse transcription buffer, 0.5 mM dNTPs, 5 mM $MgCl_2$ and 10 mM DTT. $(dN)_6$ primers (for fragmented polyA$^+$RNA: 40 ng per 0.5 μg of mRNA, min. 40 ng; for full-length polyA$^+$RNA: 2.4 μg per 0.5 μg of mRNA, min. 2.4 μg), RNase OUT (20 U per 1 μg of RNA, min. 20 u) and SuperScript III polymerase (100 U per 1 μg of RNA, min. 100 U) are added to the reaction. In case cDNA is synthesized from nonfragmented mRNA, oligo(dT) are added to the mix (25 pmol per 0.5 μg of mRNA, min. 25 pmol). Usually we start with 0.2–0.5 μg of mRNA and perform the first-strand synthesis in 20 μl.

1. Combine in a 0.2 ml thin-wall tube, all reaction components except enzymes.
2. Add mRNA (up to 50 ng/μl).
3. Place the reaction in PCR machine and run the following program:
   - 70 °C for 5 min
   - cooling down to 15 °C, at 0.1 °C/s rate

– 15 °C—pause

Note 1: In case cDNA is synthesized from nonfragmented mRNA, we recommend to hold mRNA at 98 °C for 1 min to melt RNA secondary structures, and then go down to 70 °C.

Note 2: Slow cooling allows annealing of primers being as reproducible as possible.

4. Add Actinomycin D at a concentration of 6 ng/μl while the tube is still within the PCR machine (optional step: if missed, proceed further to step 5).

   Note: The addition of Actinomycin D specifically inhibits DNA-dependent, but not RNA-dependent, DNA synthesis (Ruprecht *et al.*, 1973). This prevents the synthesis of artifact second-strand cDNA, which is an important source of false antisense transcription (Perocchi *et al.*, 2007).

5. Add RNase OUT and SuperScript III polymerase to the reaction while the tube is still within the PCR machine.

   Note: It is convenient to have a premix of these enzymes, stored at −20 °C.

6. Run the following program:
   – Heating from 15 to 25 °C, 0.1 °C/s
   – 25 °C for 10 min
   – Heating from 25 to 42 °C, 0.1 °C/s
   – 42 °C for 45 min
   – Heating from 42 to 50 °C, 0.1 °C/s
   – 50 °C for 25 min
   – 75 °C for 15 min to inactivate the SuperScript III polymerase

   Note: The temperature of the reverse transcription reaction is increased gradually as a compromise between the stability of primer annealing (for 6-mers — ∼25 °C), biological activity of the enzyme (active up to 60 °C, optimal temperature ∼42 °C), and denaturation of RNA secondary structures.

2.3.3. Removal of dNTPs

The labeling of the second cDNA strand with uridines requires the removal of dTTP from the reaction. To remove dNTPs, we recommend the use of Illustra ProbeQuant G-50 Micro Columns. It is also possible to clean up the first-strand synthesis reaction by applying a PCIA extraction followed by ethanol precipitation (Levin *et al.*, 2009). We recommend gel filtration because it is faster.

1. Change buffer in the gel filtration column: pass 1.5 ml of 1 mM Tris–HCl, pH 7.0 through the column under gravity.

   Note: This step is required to change the buffer in the column. Illustra ProbeQuant G-50 Micro Columns are originally filled with the STE Buffer (100 mM NaCl, 10 mM Tris–HCl, 1 mM EDTA). For

the second-strand synthesis, the cDNA has to be eluted in 1 mM Tris–HCl, pH 7.0.
2. Add 10 mM Tris–HCl, pH 8.5, to the first-strand synthesis reaction up to 50 μl. Load the reaction on the column and spin down at 2000 rpm for 1 min.
3. Transfer the eluate to a new 1.5-ml LoBind tube, put the tube on ice.

2.3.4. Second-strand synthesis

Second-strand synthesis reaction is performed in 0.13× reverse transcription buffer, 0.26 mM dNTPs with dTTP replaced by dUTP, 0.67 mM $MgCl_2$ and 1.33 mM DTT and 1× second-Strand Synthesis Buffer.[4] *E. coli* DNA polymerase I (0.27 U/μl) and RNase H (0.013 U/μl) are added to the reaction. If ds cDNA synthesis is performed on nonfragmented polyA$^+$RNA, *E. coli* ligase (0.07 U/μl) is added to the reaction.

Note: The substitution of dTTP by dUTP during second-strand synthesis does not change DNA synthesis rate.

1. Combine in a LoBind tube all reaction components. Usually, we prepare 25 μl of the reaction mixture: 15 μl of 5× second-Strand Synthesis Buffer, 1 μl of 10× reverse transcription buffer, 2 μl of 10 mM dNTPs with dTTP replaced by dUTP, 0.5 μl of 100 mM $MgCl_2$, 1 μl of 100 mM DTT, 2 μl of *E. coli* DNA polymerase I (10 U/μl), and 0.5 μl of RNase H (2 U/μl).
2. Add the reaction mixture to the cDNA (50 μl).
3. Incubate the reaction at 16 °C for 2 h.
4. Perform the reaction cleanup with Qiagen PCR Purification Kit.
5. Check the sample concentration (optional).

Note: typical outcome of ds cDNA synthesis is 70–90% of the starting mRNA amount.

2.4. ssRNA-Seq library preparation

The procedure of sequencing library preparation for Illumina sequencing platform that we describe below is based on the original Illumina protocol of library construction from genomic DNA. But additionally, uridine digestion by UDG is performed before the final library amplification. Thus, the uridine-containing strand is destroyed and is not amplified. We, further, routinely perform a check of the library concentration by real-time PCR to accurately determine the number of cycles for the final amplification.

[4] We use the Invitrogen system for cDNA synthesis, in which the first-strand synthesis reaction is diluted in 7.5 times in the second-strand synthesis reaction.

2.4.1. Required materials

Equipment

- PCR machine for 0.2-ml tubes
- Real-time PCR machine
- Water bath: 20, 37 °C
- Power supply
- Agarose electrophoresis chamber

Reagents/enzymes

- 10× T4 DNA Ligase Reaction Buffer (New England BioLabs, # B0202)
- T4 DNA Polymerase (New England BioLabs, # M0203)
- DNA Polymerase I, Klenow large fragment (New England BioLabs, # M02103)
- T4 Polynucleotide Kinase (New England BioLabs, # M0201)
- Klenow fragment (3′→5′ exo-) (New England BioLabs, # M0212)
- NEBuffer 2 (New England BioLabs, #B7002S)
- 10 mM dATP
- Quick Ligation™ Kit (New England BioLabs, #M2200)
- Low range ultra agarose 125 g (Bio-Rad, #161-3107)
- Phusion™ High-Fidelity PCR Master Mix with HF Buffer (New England BioLabs, #F-531S)
- PE genomic adapters, 12 μM (sequences are available)
- PE PCR primers (sequences are available)

  Note: all enzymes and oligos mentioned above are included in the Illumina Genomic DNA Sample Prep Kit (e.g., #PE-102-1001).

- SYBR® Green PCR Core Reagents (Applied Biosystems, #4304886)
- UDG (New England BioLabs, #M0280)
- QIAquick PCR Purification Kit (Qiagen, #28106)
- MinElute Reaction Cleanup Kit (Qiagen, #28206)
- QIAquick Gel Extraction Kit (Qiagen, #28706)
- Oligos for real-time PCR:
  - # PE adapter primer 2:
    5′-CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3′
  - # adapter primer 1:
    5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
- 50 bp DNA Ladder (New England BioLabs, #N3236)

Disposables

- Scalpels
- Optical plastic (tubes, strips or plates) for the real-time PCR machine

2.4.2. End-repair

1. Prepare the reaction mix in a 1.5-ml LoBind tube: 30 μl of ds cDNA, 1× T4 DNA Ligase Buffer with 1 mM ATP, 0.25 mM dNTPs, T4 DNA Polymerase (0.1 U/μl), DNA Polymerase I, Klenow large fragment (0.03 U/μl), T4 Polynucleotide kinase (0.3 U/μl).

   Note: We perform the end-repair reaction in 50 μl. We recommend a final concentration of ds cDNA up to 20 ng/μl. The volume of the enzymes added should not exceed 10% of the volume of the reaction.
2. Incubate the reaction at 20 °C for 30 min.

   Note: It is important not to increase the temperature as this would otherwise increase the exonuclease activity of T4 polymerase.
3. Perform the reaction cleanup with Qiagen PCR Purification Kit, elute in 30 μl EB.

   Note: The presence of artificial hybrid molecules in the library poses a problem when searching for rare structural rearrangements. Such molecules may appear during adapter ligation when some fragments may ligate to each other. To avoid *dimers*, a gel size selection should be performed after end-repair. It is necessary to select the cut-out range of this first size selection (before adapter ligation) in such a way that, during the second size selection (after adapter ligation), the doubled smallest fragments reliably differ from the largest fragments: $2 \times L_{\text{small}} \gg L_{\text{large}}$.

2.4.3. A-Tailing

1. Prepare the reaction mix in a 1.5-ml LoBind tube: 30 μl of ds cDNA, 1× NEBuffer 2, 1 mM dATP, Klenow fragment; $3' \rightarrow 5'$ exo- (1 U per ~160 ng DNA, min. 2.5 I per reaction).

   Note: The A-tailing reaction is performed in a volume of 50 μl. The volume of the enzymes added should not exceed 10% of the volume of the reaction.
2. Incubate the reaction at 37 °C for 30 min.
3. Perform the reaction cleanup using the MinElute Reaction Cleanup Kit and following the manufacturer instructions. Elute in 10 μl EB.

2.4.4. Adapter ligation

1. Prepare the reaction mix in a 1.5-ml LoBind tube: 10 μl of ds cDNA, PE genomic adapter (10× molar excess over the amount of ds cDNA, min. 0.5 μl), 1× Quick Ligase Buffer, Quick Ligase (10% of the volume of the reaction).

   Note: To minimize nonspecific ligation, we first mix the cDNA with the adapters, then mix thoroughly with the buffer, and then add ligase.

2. Incubate the reaction at 20 °C for 30 min.
3. Perform the reaction cleanup with Qiagen PCR Purification Kit, elute in 30 μl EB.

2.4.5. Size selection

1. Load the sample on a 2% Low Range Ultra Agarose 1× TAE gel along with an appropriate marker (e.g., 50 bp marker, NEB). Run the gel until the marker bands in the range 100–400 bp are well separated (the 100 bp band should be about 5 cm from the well).

   Note: There should be at least one empty well between sample and marker and at least two empty wells between two samples. To minimize cross contaminations, we recommend avoiding loading samples differing more than 10-fold in concentration. For precious samples, it is worth running one gel per sample.
2. Cut out the gel band corresponding to a 180–220 nt fragment.

   Note 1: The selected size range corresponds to a 110–150 bp insertion (PE adapters add ∼70 bp). We use this size range for the transcriptome analysis of human samples. For the analysis of transcript structure, it is desirable that PE reads belong to different exons. We have chosen the range comparable with the average size of an exon (∼170 bp).

   Note 2: If the library is prepared exclusively for expression profiling, insert size variation is not important. If structural analysis would be performed, the size range should be as small as possible.

   Note 3: We recommend cutting out neighboring band (220–250 bp) and to freeze it in liquid nitrogen and store at −20 °C. Such a reserve might be helpful especially for precious samples.
3. Perform the reaction cleanup with Qiagen Gel Extraction Kit and following the manufacturer instructions. Elute the library in 30 μl EB.

2.4.6. UDG treatment

1. Add 3.4 μl of 10× UDG Buffer to the library. Then add UDG (1 U per 1 μg of DNA, min. 0.5 U per reaction).
2. Incubate the reaction tube at 37 °C for 15 min. No cleanup is required.

2.4.7. Real-time quality check

1. Prepare the 20 μl real-time PCR reaction. Combine in an appropriate tube 12.7 μl of water, 2 μl of 10× SYBRGreen Buffer, 2 μl of 12.5 nM dNTPs with dUTP, 2 μl of 25 mM $MgCl_2$, 0.4 μl of each of 20 μM

oligos #PE adapter primer 2 and #adapter primer 1, and 0.1 μl of AmpliTaq Gold (5 U/μl). Add 0.5 μl of library per reaction.

2. Place the reaction in real-time PCR machine and run the following program:
   - 95 °C for 10 min
   - 40 cycles: 95 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s
3. Determine the number of cycles A, which corresponds to the middle of the logarithmic phase.

   Note: Over-amplification leads to the distortion of the original proportion of different DNA fragments.
4. Calculate the number of cycles for the final library amplification N, taking into account the volume of final amplification V_{amp} and volume of library taken for amplification

$$V_{\text{library}} : N = A - \log_2\left\{\left(V_{\text{library}}[\mu\text{l}]/V_{\text{amp}}[\mu\text{l}]\right)/(0.5\,\mu\text{l}/20\,\mu\text{l})\right\}$$

2.4.8. Amplification of the library

1. Prepare the amplification reaction mix in a 0.2 ml thin-wall tube: 1× Phusion HF Master Mix, 0.4 μM oligos #PE PCR primer 1.1 and #PE PCR primer 2.1. Take 10–30 μl of the library for the reaction.

   Note:We recommend the use of 1/3 to 1/2 of the material for the final amplification. Taking less would result in an unreasonable reduction of the complexity of the library. If all the material is used, there would be no reserve to repeat the amplification in case of an accident (like unsuccessful amplification, broken well in the gel, etc.).

   Note: The Phusion DNA polymerase used for the library amplification is strongly suppressed by uridines in the template (Hogrefe *et al.*, 2002). So, even in case that UDG occasionally would not remove the uridine base efficiently, the molecule will still not be amplified.
2. Place the reaction in a PCR machine and run the following program:
   - 95 °C for 30 s
   - Selected (by real-time library check) number of cycles: 95 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s.
3. Load the entire PCR reaction on an agarose gel and excise cDNAs ranging in size between ~230 and 270 bp.

   Note: this additional gel selection step results in a higher quality library giving a better fragment size distribution and eliminating PCR by-products like primer dimers.
4. Determine the library concentration.

Proceed to the flowcell preparation, cluster generation, and sequencing reaction on the Illumina Cluster Station and Genome Analyzer according to the manufacturer's instructions

REFERENCES

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5,** 613–619.

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *J. Biomed. Biotechnol.* 2010: 853916.

Denoeud, F., Aury, J. M., Da Silva, C., Noel, B., Rogier, O., Delledonne, M., Morgante, M., Valle, G., Wincker, P., Scarpelli, C., Jaillon, O., and Artiguenave, F. (2008). Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* **9**(12), R175.

He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., and Kinzler, K. W. (2008). The antisense transcriptomes of human cells. *Science* **322,** 1855–1857.

Hogrefe, H. H., Hansen, C. J., Scott, B. R., and Nielson, K. B. (2002). Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. *Proc. Natl. Acad. Sci. USA* **99,** 596–601.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermuller, J., Hofacker, I. L., *et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316,** 1484–1488.

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A., and Regev, A. (2009). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**(9), 709–715.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133,** 523–536.

Marioni, J., Mason, C., Mane, S., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18,** 1509–1517.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320,** 1344–1349.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37**(18), e123.

Perocchi, F., Xu, Z., Clauder-Münster, S., and Steinmetz, L. M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.* **35,** e128.

Ruprecht, R. M., Goodman, N. C., and Spiegelman, S. (1973). Conditions for the selective synthesis of DNA complementary to template RNA. *Biochim. Biophys. Acta* **294,** 192–203.

Stoughton, R. B. (2005). Applications of DNA microarrays in biology. *Annu. Rev. Biochem.* **74,** 53–82.

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., *et al.* (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321,** 956–960.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* **270,** 484–487.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., and Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453,** 1239–1243.

Wilhelm, B. T., Marguerat, S., Goodhead, U., and Bähler, J. (2010). Defining transcribed regions using RNA-Seq. *Nat. Protoc.* **5,** 255–266.

Zbio-wiki (2010). Next-generation_sequencing. Available at:http://www.zbio.net/wiki/Next-generation_sequencing(accessed 11 November 2010).