Published in final edited form as:

Methods Mol Biol. 2011 ; 733: 51–61. doi:10.1007/978-1-61779-089-8_4.

Transcriptome Profiling Using Single-Molecule Direct RNA Sequencing

Fatih Ozsolak and Patrice M. Milos

Abstract

Methods for in-depth characterization of transcriptomes and quantification of transcript levels have emerged as valuable tools for understanding cellular physiology and human disease biology, and have begun to be utilized in various clinical diagnostic applications. Today, current methods utilized by the scientific community typically require RNA to be converted to cDNA prior to comprehensive measurements. However, this cDNA conversion process has been shown to introduce many biases and artifacts that interfere with the proper characterization and quantitation of transcripts. We have developed a direct RNA sequencing (DRS) approach, in which, unlike other technologies, RNA is sequenced directly without prior conversion to cDNA. The benefits of DRS include the ability to use minute quantities (e.g. on the order of several femtomoles) of RNA with minimal sample preparation, the ability to analyze short RNAs which pose unique challenges for analysis using cDNA-based approaches, and the ability to perform these analyses in a low-cost and high-throughput manner. Here, we describe the strategies and procedures we employ to prepare various RNA species for analysis with DRS.

Keywords

RNA sequencing; Single-molecule sequencing; Transcriptome profiling; Polyadenylation site mapping

1. Introduction

The emergence of microarray (1-4) and high-throughput DNA/ cDNA sequencing technologies (5-10) and their application to understanding biological processes and human disease initially provided a relatively simplistic view of transcriptomes which has since been replaced with a larger, more complicated view of genome-wide transcription. We now have a much more comprehensive view of the genome in which a large fraction of transcripts emanate from unannotated parts of the genome (reviewed in (11)), and has highlighted our limited, yet rapidly emerging, knowledge of the transcriptome and the intimate role RNA plays in health and disease (12-17). New technologies and methods, which offer unique approaches to transcriptome characterization and quantitation, with particular emphasis on minimizing the inherent biases seen with existing methods and the ability to work with minute quantities of cellular RNA are critical to fully explore transcriptome biology.

DNA/cDNA sequencing has eliminated some of the technical challenges posed by earlier hybridization-based microarray strategies, including limited dynamic range of detection and

© Springer Science+Business Media, LLC 2011

⁸For small RNA profiling applications, enrichment for small RNA species can be performed using various methods. We have successfully used the miRNeasy (Qiagen, 217004) and the mirVana (Ambion, AM1560) kits following the manufacturer's guidelines to obtain and profile <200 nts small RNA species. Other gel-based RNA size selection methods can also be used to enrich for RNA species with a more limited size distribution, but has not been tested for DRS.

the relatively high background due to cross-hybridization, but several fundamental shortcomings still remain which prevent us from understanding the “true” nature of transcriptomes. One limitation of cDNA-based approaches is the tendency of various reverse transcriptases (RT) to generate spurious second-strand cDNA due to their DNA-dependent DNA polymerase activities (18-20). This is thought to occur through either a hairpin loop at the 3' end of the first-strand cDNA or by specific or nonspecific re-priming, involving either RNA fragments or primers used for the first-strand synthesis. This effect confounds analyses aimed at identifying the strand of the genomic DNA which gives rise to the RNA (e.g. detecting sense vs. antisense transcripts) (21). Another cDNA limitation, known as template switching (22-25), occurs during the process of reverse transcription in which the nascent cDNA being synthesized can sometimes dissociate from the template RNA and re-anneal to a different stretch of RNA with a sequence similar to the initial template. This event creates an arti-factual cDNA that comprises the 5' region of the initial template attached to the 3' region of the second template. In addition to causing difficulties in RNA quantification, template switching causes problems in the identification of exon-intron boundaries and true chimeric transcripts. RTs are also known to synthesize cDNAs in a primer independent manner, thought to be caused by self-priming due to RNA secondary structure, resulting in the generation of random cDNAs (26, 27). While this self-priming may only occur at a frequency between 2 and 10% of cDNA molecules, these self-primed products are thought to be a major source of error in achieving accurate detection and quantification of RNA expression. Furthermore, RTs are error-prone due to their lack of proofreading mechanisms (28, 29) and yield low quantities of cDNA, necessitating the use of large quantities of input RNA and relatively high levels of amplification. Finally, most commercial technologies for gene expression/whole transcriptome analyses add further artifacts to their measurements by requiring double-stranded cDNA for subsequent amplification by PCR, which can eliminate information regarding which DNA strand is encoding the RNA. While strand-specific libraries can be prepared, they are laborious with many steps (30) or involve RNA-RNA ligation, which is highly inefficient (31).

Since almost all RNA analysis technologies in use today suffer from the limitations briefly summarized above, there is a great need for a technology that can eliminate the difficulties associated with reverse transcription, amplification, ligation, and other cDNA synthesis-based artifacts. To address these difficulties, we have developed the first direct RNA sequencing (DRS) technology (32) using single-molecule sequencing enabled by the Helicos® Genetic Analysis System. A short read technology, DRS currently produces alignable reads up to 55 nts in length with a mean read length of 33-34 nts. Each DRS run contains two flow cells with 50 independent channels, and produce between 800,000 and 8,000,000 aligned reads (≥ 25 nts) per channel on an average, depending on the requested run time and fields of view (FOV) (e.g. imaging quantity) per channel. The DRS sample preparation step involves only polyadenylation of 3' ends of RNA molecules without the need for complicated and potentially biased steps such as ligation or PCR amplification of cDNAs (Fig. 1). For several applications, such as gene expression profiling of polyA+ RNA species or polyadenylation site mapping, the natural poly-A tails that are already present on the RNAs provide the hybridization template sufficient for sequencing. This ensures that any biases that may be introduced by sample preparation steps are reduced or eliminated. The simplicity of DRS sample preparation, requiring femtomole-level RNA quantities, combined with its strand-specific and quantitative nature free from reverse transcription-associated artifacts will make DRS the method of choice for transcriptome analyses. As such, we here provide the readers with a detailed description of the methods utilized for our early studies of DRS.

2. Materials

2.1. Reagents

1. Poly(A) Tailing Kit (Ambion). This kit contains *Escherichia coli* PolyA polymerase (2 U/μl), 5× *E. coli* PolyA polymerase buffer, 10 mM ATP solution, and 25 mM MnCl₂.
2. 10 mM 3'dATP (Axxora/Jena Biosciences).
3. Nuclease-free water (Ambion).
4. 5 mg/ml Linear Acrylamide (Ambion).
5. 5 M Ammonium acetate (Ambion).
6. Phenol/Chloroform/Isoamyl alcohol.
7. 100% Ethanol.
8. 70% Ethanol.
9. Helicos RNA sequencing kit (Helicos BioSciences Corporation).

2.2. Equipment

1. Thermal cycler (of choice).
2. Refrigerated microcentrifuge.
3. Aluminum blocks (VWR).
4. HeliScope™ Single Molecule Sequencer (Helicos BioSciences Corporation).

3. Methods

We here outline the protocols routinely employed to prepare RNA samples for DRS using the Helicos® Genetic Analysis System. In the first section, we describe the methodology utilized for gene expression profiling and polyadenylation site mapping applications. For these studies, we benefit from the naturally occurring polyadenylation of RNA and require only 3' blocking followed directly by flow cell hybridization and subsequent sequencing by synthesis ((32); Subheading 3.1). In many cases, a broader view, or qualitatively different view, of the transcriptome requires the RNA molecules to be synthetically polyadenylated prior to hybridization and sequencing. These applications are exemplified by studies involving small RNA sequencing or whole transcriptome analyses, where in each case an additional 3' polyadenylation step is performed (Subheading 3.2).

3.1. Gene Expression Analyses and Polyadenylation Site Mapping with DRS

Proceed with this method if the RNA of interest is polyadenylated or part of the total RNA in which the RNA species of interest are polyadenylated.

1. Prepare RNA to be blocked in nuclease-free water in a 22 μl volume (see Notes 1 and 2).

¹ RNAs isolated with various techniques and commercial kits can be used. We have successfully used total RNA isolated with Trizol® (Invitrogen) and RNeasy Plus Mini Kit (Qiagen). The removal of any potential contaminating genomic DNA from the RNA samples is strongly recommended. RNAs enriched for the polyA+ species can also be used instead of total RNA. RNA quantification can be performed using UV absorbance. For the quantification of low-quantity RNA samples, in our experience, the Quant-iT™ RiboGreen® Reagent (Invitrogen, R11490) provides the most accurate results.

2. Heat the RNA at 85°C for 1 min in a thermocycler, followed by rapid cooling in a prechilled aluminum block kept in an ice and water slurry (~0°C, see Note 3).
3. Add the following reagents in the indicated order while keeping the denatured sample in the cooled aluminum block: 8 µl of 5× *E. coli* PolyA polymerase buffer; 4 µl of 25 mM MnCl₂; 2 µl of 10 mM 3'dATP, and 4 µl of 2 U/µl *E. coli* PolyA polymerase. Mix well by pipetting gently up and down at least five times without vortexing. Incubate for 1 h at 37°C (see Notes 4 and 5).
4. Transfer the samples to a 1.5-ml tube (see Note 6). Add 260 µl nuclease-free water and 300 µl phenol/chloroform/ isoamyl alcohol. Vortex vigorously for 30 s.
5. Centrifuge at room temperature for 5 min at maximum speed (~16,000 × g). Transfer ~280 µl from the upper aqueous phase to a fresh 1.5-ml tube.
6. Add 30 µl 5 M ammonium acetate, 4 µl 5 mg/ml linear acryl-amide, and 900 µl 100% ethanol. Incubate for a minimum of 30 min at -80°C.
7. Centrifuge down at 4°C for 30 min at top speed (~16,000 × g). Remove the supernatant with a 1,000-µl pipette tip. Avoid touching the pellet with the pipette tip.
8. Wash the pellet once by adding 500 µl of 70% ethanol. Centrifuge the tube at 4°C for 5 min at top speed (~16,000 × g) followed by removal of the supernatant. Keep the pellet at room temperature for 10 min to allow the evaporation of remaining ethanol and water.
9. Resuspend the pellet in 20 µl water.

3.2. Small RNA and Transcriptome Profiling with DRS

The following method is utilized when the RNA of interest is not polyadenylated and thus requires the addition of a poly-A tail for subsequent hybridization and sequencing.

1. Prepare 5 pmol of RNA to be tailed and blocked in nuclease-free water in a 24 µl volume (see Notes 7-9).
2. Heat the RNA at 85°C for 1 min in a thermocycler, followed by rapid cooling in a prechilled aluminum block kept in an ice and water slurry (~0°C, see Note 3).
3. Add the following reagents in the indicated order while keeping the sample on the cooled aluminum block: 8 µl of 5× *E. coli* PolyA polymerase buffer; 4 µl of 25 mM

²The HeliScope™ Single Molecule Sequencer generally requires 2-20 ng total RNA to load each of its 50 channels at optimal levels for DRS. The required total RNA quantity is determined by the number of polyA+ RNA molecules in each sample, which varies depending on the cell type and organism being studied. The OptiHyb assay system from Helicos BioSciences allows determination of polyA+ RNA molarity in a sample. We recommend HeliScope Sequencer users to employ this assay to achieve the optimal results.

³It is essential to chill the block to 0°C in an ice and water slurry, and to cool the sample as quickly as possible to 0°C to minimize re-annealing of the denatured RNA strands.

⁴The blocking reaction can be used for RNA sample quantities as high as 10 pmol.

⁵A cleanup step is not required if the sample will be used for sequencing within several days of tailing/blocking. After the completion of the tailing/blocking step, the sample should be supplemented with 5 µl 100 mM EDTA prior to storage at -80°C until HeliScope DRS flow cell preparation. For long-term storage of the blocked sample, we recommend that the sample be cleaned prior to storage.

⁶The cleanup step can be performed using various approaches. We have successfully used the Qiagen RNeasy MinElute Cleanup Kit (Qiagen, 74204) following manufacturer's instructions and the standard phenol/chloroform method described here.

⁷The RNA molarity should be estimated using the concentration information and RNA size distribution. RNA quantities less than the recommended 5 pmol level can be used (down to 0.05 pmol) with this protocol, but ATP and 3' dATP levels should be adjusted accordingly to maintain the 200:1 ATP:template RNA and 4,000:1 3'dATP:template RNA ratios.

⁹For whole transcriptome applications, RNA species generally need to be fragmented prior to polyadenylation and sequencing. Fragmentation can be done in multiple ways, such as divalent cation-catalyzed hydrolysis (14). Depending on the fragmentation method used, phosphatase treatment of the RNA prior to 3' polyadenylation and blocking may be necessary to make 3' RNA ends available for tailing.

MnCl₂; 1 µl of 1 mM ATP, and 3 µl of 2 U/µl *E. coli* PolyA polymerase. Mix well by gently pipetting up and down at least five times without vortexing. Incubate for 10 min at 37°C.

4. While keeping the sample on the 37°C thermocycler block, add 2 µl of 10 mM 3'-dATP to the sample (total volume is now 42 µl) and mix thoroughly. Incubate the sample for an additional 50 min at 37°C.
5. Sample cleanup can be performed as described in steps 4-9 of Subheading 3.1.

3.3. Flow Cell Hybridization and Single-Molecule Sequencing

3.3.1. Flow Cell Hybridization—Following the methods described in Subheading 3.1 or 3.2 above, the RNA molecules are ready for flow cell hybridization and subsequent sequencing. Hybridization of samples to Helicos flow cell channels is performed in 15-75 µl volume and chosen by the user. The RNA samples are mixed 50:50 with 2× hybridization buffer provided in the Helicos RNA Sequencing Kit (Helicos BioSciences Corporation). The volume of nuclease-free water to be used to resuspend the RNA sample should be determined considering the input RNA quantity and the volume of hybridization cocktail preferred to be used. In general, 0.5-2 fmol of RNA material is required to optimally load each sequencing channel. Following hybridization, the RNA molecules are “filled and locked” (Fig. 1) and are ready for sequencing. The detailed protocols and reagents for flow cell preparation and sequencing are provided in the Helicos RNA Sequencing Kit (Helicos BioSciences Corporation).

3.3.2. Single-Molecule Direct RNA Sequencing—Following hybridization, the Helicos flow cells are moved to the HeliScope™ Single Molecule Sequencer and sequencing by synthesis initiated using prespecified scripts which offer the user the opportunity to run one or two flow cells and define the numbers of FOV from 110 FOV to the standard condition of 1,100 FOV for maximal sequencing yield.

3.4. Data Analysis

The various Unix-based programs and pipelines for the filtering, alignment and downstream analyses of the HeliScope DNA sequencing and DRS data can be downloaded freely at <http://open.helicosbio.com/mwiki/index.php/Releases>. The detailed descriptions of the programs and functions available are described at http://open.helicosbio.com/helisphere_user_guide/2009-R1/index.html.

Briefly, an initial filtering step is performed on the raw DRS reads before initiating their alignment to reference sequences. This filtering step involves the following read selection steps:

1. DRS generates reads between 6 and 60 nts in length. Depending on the experimental goals, the reference sequence complexity, and size, a user-defined minimum read length cutoff is employed to remove short reads that cannot be aligned reliably. While we routinely use reads ≥25 nts for alignment to human and mouse genomes, for smaller genomes such as *Saccharomyces cerevisiae* and *E. coli*, DRS reads as short as 18 nts can be used.
2. Any 5' polyT stretches in DRS reads are trimmed. The likely cause of such homopolymeric stretches is the infrequent incomplete “fill” which may occur during dTTP addition step (Fig. 1). Therefore, polyT trimming is preferred to minimize potential misalignment events.

3. Because of flow cell surface imperfections and/or imaging errors, artifactual reads that have a repetition of the base-addition order sequence (CTAG) may appear. Such reads are eliminated during the filtering step.

Total raw base error rate for DRS is currently in the range of 4-5%, dominated by missing base errors typical of single molecule sequencing (2-3%), while the insertion rate is 1-2% and the substitution error rate is 0.1-0.3%. Given that the majority of sequencing errors are due to indels, an aligner that is tolerant to these types of errors should be employed. We highly recommend the use of the indexDPgenomic aligner freely available from the Helicos BioSciences HeliSphere (<http://open.helicosbio.com/mwiki/index.php/Releases>). While multiple aligners are available and can successfully align Helicos sequence reads, including BWA (33) and SHRiMP (34), use of these aligners will result in a significant reduction in actual aligned reads due to their reduced ability to deal effectively with indels.

Acknowledgments

We thank our colleagues at the Helicos BioSciences Corporation for technical assistance and discussions.

References

1. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science*. 1991; 251:767–773. [PubMed: 1990438]
2. Lennon GG, Lehrach H. Hybridization analyses of arrayed cDNA libraries. *Trends Genet*. 1991; 7:314–317. [PubMed: 1781028]
3. Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*. 1996; 6:639–645. [PubMed: 8796352]
4. Southern EM, Maskos U, Elder JK. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligo-nucleotides: evaluation using experimental models. *Genomics*. 1992; 13:1008–1017. [PubMed: 1380482]
5. Bennett ST, Barnes C, Cox A, Davies L, Brown C. Toward the 1,000 dollars human genome. *Pharmacogenomics*. 2005; 6:373–382. [PubMed: 16004555]
6. Deamer DW, Branton D. Characterization of nucleic acids by nanopore analysis. *Acc Chem Res*. 2002; 35:817–825. [PubMed: 12379134]
7. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*. 2003; 299:682–686. [PubMed: 12560545]
8. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirake KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437:376–380. [PubMed: 16056220]
9. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. 2005; 309:1728–1732. [PubMed: 16081699]
10. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res*. 2008; 18:1051–1063. [PubMed: 18477713]
11. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet*. 2007; 8:413–423. [PubMed: 17486121]

12. Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F. Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* 2008; 9:R175. [PubMed: 19087247]
13. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 2008; 18:1509–1517. [PubMed: 18550803]
14. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628. [PubMed: 18516045]
15. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science.* 2008; 320:1344–1349. [PubMed: 18451266]
16. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science.* 2008; 321:956–960. [PubMed: 18599741]
17. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature.* 2008; 453:1239–1243. [PubMed: 18488015]
18. Gubler U. Second-strand cDNA synthesis: classical method. *Methods Enzymol.* 1987; 152:325–329. [PubMed: 3657578]
19. Gubler U. Second-strand cDNA synthesis: mRNA fragments as primers. *Methods Enzymol.* 1987; 152:330–335. [PubMed: 3309563]
20. Spiegelman S, Burny A, Das MR, Keydar J, Schlom J, Travnicek M, Watson K. DNA-directed DNA polymerase activity in oncogenic RNA viruses. *Nature.* 1970; 227:1029–1031. [PubMed: 4317810]
21. Wu JQ, Du J, Rozowsky J, Zhang Z, Urban AE, Euskirchen G, Weissman S, Gerstein M, Snyder M. Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. *Genome Biol.* 2008; 9:R3. [PubMed: 18173853]
22. Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. *Genomics.* 2006; 88:127–131. [PubMed: 16457984]
23. Mader RM, Schmidt WM, Sedivy R, Rizovski B, Braun J, Kalipciyan M, Exner M, Steger GG, Mueller MW. Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA. *J Lab Clin Med.* 2001; 137:422–428. [PubMed: 11385363]
24. Roy SW, Irimia M. When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. *Bioessays.* 2008; 30:601–605. [PubMed: 18478540]
25. Roy SW, Irimia M. Intron mis-splicing: no alternative? *Genome Biol.* 2008; 9:208. [PubMed: 18304372]
26. Haddad F, Qin AX, Bodell PW, Zhang LY, Guo H, Giger JM, Baldwin KM. Regulation of antisense RNA expression during cardiac MHC gene switching in response to pressure overload. *Am J Physiol Heart Circ Physiol.* 2006; 290:H2351–2361. [PubMed: 16415074]
27. Haddad F, Qin AX, Giger JM, Guo H, Baldwin KM. Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. *BMC Biotechnol.* 2007; 7:21. [PubMed: 17480233]
28. Roberts JD, Preston BD, Johnston LA, Soni A, Loeb LA, Kunkel TA. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. *Mol Cell Biol.* 1989; 9:469–476. [PubMed: 2469002]
29. Varadaraj K, Skinner DM. Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. *Gene.* 1994; 140:1–5. [PubMed: 8125324]
30. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods.* 2008; 5:613–619. [PubMed: 18516046]

31. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*. 2008; 133:523–536. [PubMed: 18423832]
32. Ozsolak F, Platt AR, Jones DR, Reifenger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. *Nature*. 2009; 461:814–818. [PubMed: 19776739]
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
34. Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M. SHRiMP: accurate mapping of short color-space reads. *PLoS Comput Biol*. 2009; 5:e1000386. [PubMed: 19461883]

Fig. 1.

DRS sample preparation. RNA species which contain a 3' poly-A tail require 3' end blocking as described. Other RNA species are enzymatically 3' polyadenylated and 3' blocked. The blocking step is performed to prevent “downward” nucleotide additions to the 3' end of the template during the sequencing process (details of the sequencing strategy and chemistry have been described previously (32)). Polyadenylated RNA is captured on the sequencing flow cell surfaces coated with poly(dT) oligonucleotides through hybridization. A “fill” step is performed with dTTP and polymerase, and then the templates are “locked” in position with fluorescently labeled proprietary Virtual Terminator™ (VT)-A, -C and -G sequencing nucleotide analogs. VT analogs are nucleotides used for sequencing, containing a fluo-

cent dye and chemically cleavable groups that prevent the addition of another nucleotide. These “fill and lock” steps correct for any misalignments that may be present in poly-A/T duplexes, and ensure that the sequencing starts in the template rather than the poly-A tail.