Methods 63 (2013) 126–134

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

A new method for stranded whole transcriptome RNA-seq

David F.B. Miller [a], Pearlly S. Yan [b], Aaron Buechlein [c], Benjamin A. Rodriguez [b], Ayse S. Yilmaz [b], Shokhi Goel [a], Hai Lin [d], Bridgette Collins-Burow [e], Lyndsay V. Rhodes [e], Chris Braun [f], Sunila Pradeep [g], Rajesha Rupaimoole [g], Mehmet Dalkilic [f], Anil K. Sood [g], Matthew E. Burow [e], Haixu Tang [h], Tim H. Huang [i], Yunlong Liu [d,*], Douglas B. Rusch [c,*], Kenneth P. Nephew [a,*]

[a] Medical Sciences, Indiana University School of Medicine, 1001 East 3rd St., Bloomington, IN 47405, United States
[b] Department of Internal Medicine, OSUCCC-Illumina Core, 460 West 12th St., Columbus, OH 43210, United States
[c] Indiana University Center for Genomics and Bioinformatics, Bloomington, IN 47405, United States
[d] Indiana University School of Medicine, Center for Computation Biology and Bioinformatics, Indianapolis, IN 46202, United States
[e] Hematology and Medical Oncology, Tulane University, New Orleans, LA 70112, United States
[f] Indiana University School of Informatics, 910 East 10th St., Bloomington, IN 47408, United States
[g] University of Texas MD Anderson Cancer Center, Department of Gynecologic Oncology, Houston, TX 77030, United States
[h] Indiana University School of Informatics and Computing, 150 S Woodlawn Ave., Bloomington, IN 47405, United States
[i] University of Texan Health Science Center, Department of Molecular Medicine, 7703 Floyd Curl Dr., San Antonio, TX 78229, United States

ARTICLE INFO

Article history:
Available online 1 April 2013

Keywords:
RNA-seq
Transcriptome
Duplex-specific nuclease
Gene expression

ABSTRACT

This report describes an improved protocol to generate stranded, barcoded RNA-seq libraries to capture the whole transcriptome. By optimizing the use of duplex specific nuclease (DSN) to remove ribosomal RNA reads from stranded barcoded libraries, we demonstrate improved efficiency of multiplexed next generation sequencing (NGS). This approach detects expression profiles of all RNA types, including miRNA (microRNA), piRNA (Piwi-interacting RNA), snoRNA (small nucleolar RNA), lincRNA (long non-coding RNA), mtRNA (mitochondrial RNA) and mRNA (messenger RNA) without the use of gel electrophoresis. The improved protocol generates high quality data that can be used to identify differential expression in known and novel coding and non-coding transcripts, splice variants, mitochondrial genes and SNPs (single nucleotide polymorphisms).

1. Introduction

Next-generation sequencing (NGS) technologies represent the most advanced molecular tools currently available to interrogate the complexities of the human transcriptome in its entirety [1–3]. RNA-seq NGS has the capability and capacity to detect all types of gene expression, *i.e.*, the whole transcriptome, including small noncoding RNAs, mRNAs, lincRNAs, novel transcripts, and repetitive elements in a manner capable of identifying splice variants and SNPs [4–6]. A whole transcriptome analysis requires the reduction of rRNA products for efficient NGS while incorporating barcoded adapters for multiplexing. While current RNA-seq library generation protocols utilize poly(A+) selection, RNase H, Ribozero and random hexamer priming for rRNA reduction [3,7–10], here we describe an improved protocol for capturing the strand-specific whole transcriptome of gene expression. This new method utilizes re-association kinetics and duplex specific nuclease (DSN) to preferentially remove highly abundant rRNA species. The use of specially designed barcoded TrueSeq adapters facilitates multiplexing and sequencing of samples on the Illumina HiSeq platform [11,12].

1.1. Work flow for RNA-seq

The key steps used to generate barcoded, stranded, whole transcriptome RNA-seq libraries are illustrated in Fig. 1. Beginning with total RNA, samples are fractionated by size to separate RNA species

Abbreviations: smRNA, small total RNA <200 nt; LgRNA, large total RNA >200 nt; FLgRNA, fragmented LgRNA; DSN, duplex-specific nuclease.

* Corresponding authors. Fax: +1 812 855 4436 (K.P. Nephew), fax: +1 317 278 9217 (Y. Lui), fax: +1 812 855 8043 (D. Rusch).

E-mail addresses: millerdf@indiana.edu (D.F.B. Miller), Pearlly.Yan@osumc.edu (P.S. Yan), abuechle@cgb.indiana.edu (A. Buechlein), Benjamin.Rodriguez@bcm.edu (B.A. Rodriguez), AyseSelen.Yilmaz@osumc.edu (A.S. Yilmaz), shgoel@indiana.edu (S. Goel), linhai@umail.iu.edu (H. Lin), bcollin1@tulane.edu (B. Collins-Burow), lvanhoy@tulane.edu (L.V. Rhodes), cbbraun@imail.iu.edu (C. Braun), SPradeep@mdanderson.org (S. Pradeep), RRupaimoole@mdanderson.org (R. Rupaimoole), dalkilic@indiana.edu (M. Dalkilic), asood@mdanderson.org (A.K. Sood), mburow@tulane.edu (M.E. Burow), hatang@indiana.edu (H. Tang), huangt3@uthscsa.edu (T.H. Huang), yunliu@iupui.edu (Y. Liu), drusch@cgb.indiana.edu (D.B. Rusch), knephew@indiana.edu (K.P. Nephew).

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2013.03.023

<200 nt (smRNA) from large RNA >200 nt (LgRNA). The LgRNAs are chemically fragmented to an average of 200 nt in length [13]. The ends are prepared and the adapters are sequentially added to the ends to retain strandedness. Following PCR amplification, the libraries are normalized using hybridization kinetics, duplex specific nuclease (DSN), and PCR amplification with the addition of barcoded index sequences for multiplexing samples in a single flow cell lane [11,12]. This approach allows the capture of all RNA species independent of size or poly(A) status.

The improved method described in this report is practical for multiplexed RNA-seq on the Illumina HiSeq platform. For Illumina multiplex sequencing, it is critical to plan the index sequence combinations in advance due to limitations with this platform. Not all index sequences are compatible for de-multiplexing and should be carefully reviewed prior to the addition of these sequences [11]. Our approach utilizes DSN normalization yet allows flexibility for addition of the barcodes later, with the option of generating multiple versions of the same library with different barcoded adapter sequences.

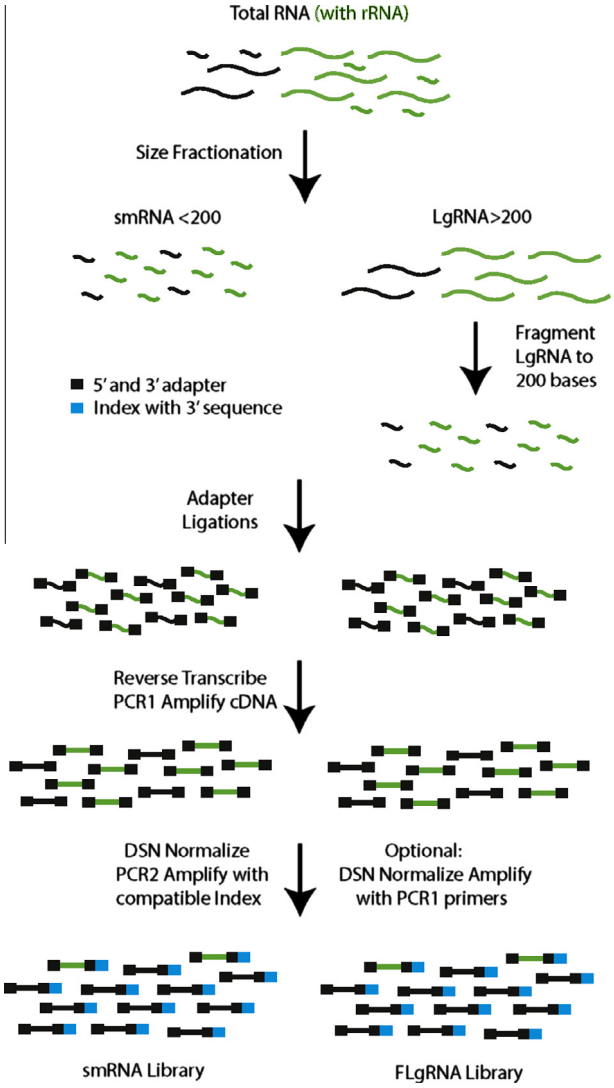

Fig. 1. Library generation workflow. Starting with total RNA, size fractionation into smRNA (<200 nt) and LgRNA (>200 nt) generates whole transcriptome coverage in two fractions. Following fragmentation of LgRNA (FLgRNA), protocols to generate stranded RNA-seq libraries are essentially the same for both fractions. Duplex specific nuclease (DSN) is utilized to remove rRNA sequences while retaining all other transcript library elements. Index sequences are added after DSN treatment, during PCR2, to generate libraries that can be used for multiplexed next generation sequencing with the Illumina platform.

2. Materials and methods

2.1. Cell culture

For this study, we used a pair of isogenic epithelial ovarian cancer cell lines (A2780 and A2780-C1R5) [14]. The culture conditions for parental A2780 line (sensitive to the chemotherapy cisplatin) A2780-C1R5 (derived from A2780 after several rounds of cisplatin selection and thus represents acquired drug resistance) have been described previously [14]. Briefly, cells were grown in RPMI 1640 media (ThermoFisher Scientific, Wilmington, MA 01887, cat. No. MT-15-040-CV) supplemented with 10% fetal bovine serum (Atlanta Biological, Norcross, GA 30093, cat. No. S11050) and 2 mM glutamine, 100 units U/mL penicillin and 100 μg/mL Streptomycin (ThermoFisher Scientific, Wilmington, MA 01887, cat. No. MT30-009-CI) at 37 °C in 5% CO_2 in a Steri-Cycle CO_2 incubator (ThermoFisher Scientific, Wilmington, MA 01887, model 370 Series).

2.2. RNA isolation

RNA was isolated from trypsinized cells utilizing a modified version of the Qiagen RNeasy AllPrep method (Qiagen Sciences, Valencia, CA 91355, cat. No. 80004). Pellets were harvested at 4 °C, 1000 rpm (129 g) for 5 min in an Eppendorf fixed angle rotor (Hauppauge, NY 11788, cat. No. F34-6-38) using a Falcon 15 mL conical (ThermoFisher Scientific, Wilmington, MA 01887, cat. No. 14-959-70C) in an Eppendorf 5810R centrifuge at 4 °C. Cells were lysed in 600 μL RNeasy RLT buffer (Qiagen Sciences, Valencia, CA 91355, cat. No. 79216) with 1% BME (vol/vol) by vortexing or raking the tube across a micro-centrifuge tube rack about 20 times. Cell lysis was completed by centrifugation through the Qiashredder spin column (Qiagen Sciences, Valencia, CA 91355, cat. No. 79654) for 2 min at 13,300 rpm (17,000 g) in an AccuSpin Micro17 centrifuge (ThermoFisher Scientific, Wilmington, MA 01887, cat. No. 13-100-675). The flow-thru was collected and passed over an AP-DNA column (10,200 rpm, 10,000 g) for 20 s to remove DNA. Ethanol (900 μL) was added to a 60% final concentration (vol/vol) to the RNA/protein fraction before RNA capture using an RNeasy spin column (centrifugation at 10,200 rpm for 20 s). The protein flow-thru fraction was stored at −80 °C for later use. The RNeasy column was washed with Qiagen RWT buffer (Qiagen Sciences, Valencia, CA 91355, cat. No. 1067933) containing ethanol (as recommended by the manufacturer) for 20 s at 10,200 rpm. The RNeasy column was then washed with 500 uL RPE buffer by centrifugation for 20 s at 10,200 rpm. The RNeasy column was also washed with 500 μL of 80% (vol/vol) ethanol by centrifugation for 5 min at 13,300 rpm. The column was transferred to a clean collection tube and centrifuged for an additional 5 min at 13,300 rpm to dry the column. Elution of the total RNA fraction into a clean collection tube was accomplished by soaking the column matrix with 50 uL DIW (supplied in the kit) for one minute followed by centrifugation for 1 min at 13,300 rpm. Total RNA samples were stored at −80 °C until further use. RNA integrity was monitored using a Bioanalyzer and the Agilent RNA 6000 Pico assay (Agilent Technologies, Santa Clara, CA, 95051, cat. No. 5067-1513). An RNA integrity number (RIN) greater than eight is recommended for all samples.

2.3. RNA size fractionation

Size fractionation of total RNA samples around 200 nt was accomplished utilizing a supplemental Qiagen protocol [13]. For