High-throughput Minaturized RNA-Seq Library Preparation

Samuel Mildrum,1 Austin Hendricks,1 Alexei Stortchevoi, Noelani Kamelamela, Vincent L. Butty, and Stuart S. Levine*

Department of Biology, MIT BioMicro Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Advances in next-generation sequencing technologies have allowed RNA sequencing to become an increasingly time efficient, cost-effective, and accessible tool for genomic research. We present here an automated and miniaturized workflow for RNA library preparation that minimizes reagent usage and processing time required per sample to generate Illumina compatible libraries for sequencing. The reduced-volume libraries show similar behavior to full-scale libraries with comparable numbers of genes detected and reproducible clustering of samples.

KEY WORD: Sequencing, automation, Illumina

INTRODUCTION

The continued development of next-generation sequencing technologies has made RNA sequencing (RNA-Seq) an increasingly powerful tool and readily available technique for transcriptome analysis.1, 2 Despite the continuous decrease in sequencing costs over the years, Illumina library preparation using commercially available kits has remained expensive and is a significant proportion of experimental costs.3, 4 Because of this, many next-generation sequencing experiments sacrifice replicates, which consequently reduces the statistical power of the data.5 A solution to offset the high costs would be to miniaturize reagent volumes and reduce hands-on labor through the utilization of liquid handlers in Illumina library preparation; however, standard liquid handlers often create a bottleneck in the miniaturization process due to the inability to accurately dispense volumes under 4 μl.6 Several groups have demonstrated miniaturization for DNA,7, 8 RNA,9 and single-cell RNA10 library preparation below the standard volume limit using nanoliter liquid handlers. Yet, a potential drawback of this is that library preparation at reduced volumes is still prone to introducing sampling error as a consequence of the decreased input mass proportional to the scale of miniaturization necessary to maintain stoichiometric ratios of the chemistries.

In this work, we combined the low-volume SPT Labtech Mosquito HV and the Tecan Freedom EVO 150 liquid handlers to establish a cost-effective workflow at reduced volumes for both a bead-based poly(T) mRNA enrichment and an enzymatic rRNA depletion for RNA-Seq. These methods allowed for greatly reduced reagent usage and increased sample throughput as compared to manual preparations, resulting in overall library preparation savings on a per sample basis while preserving library quality and maintaining an appreciable dynamic range of input. We validated this method using Universal Human Reference (UHR) total RNA as well as experimental samples. Analysis of RNA-Seq yielded data of similar quality and reproducibility to that of manually prepared full-reagent volume libraries by sequence content, rRNA contamination, directionality, and gene detection.

MATERIALS AND METHODS

Input material

UHR RNA (740000; Stratagene, San Diego, CA, USA) was used as standard input. Total RNA of experimental samples was derived from primary human colonic epithelia and human T cells (Trapecar et al., unpublished results). RNA quality was assessed on the 5300 Fragment Analyzer System (Agilent Technologies, Santa Clara, CA, USA) and concentration was determined by the ND-1000 Spectrophotometric Nano Drop (Thermo Fisher Scientific, Waltham, MA, USA).

RNA enrichment

Poly(A) mRNA isolation

mRNA was captured from 100 ng of UHR total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490; New England BioLabs, Ipswich, MA, USA) at a 1/20th reaction volume (Fig. 1). 2.5 μl of UHR total RNA (40 ng/μl) in a 384-well PCR plate (HSP3801; Bio-Rad, Hercules, CA, USA) was used for input. Oligo d(T)25 beads

1These authors contributed equally to this work.
*ADDRESS CORRESPONDENCE TO: Stuart Levine, Massachusetts Institute of Technology, 77 Massachusetts Ave., 68-304D, Cambridge, MA 02139, USA (Tel: 617-452-2949; E-mail: slevine@mit.edu).

This article includes supplemental data. Please visit http://www.fasebj.org to obtain this information.
https://doi.org/10.7171/jbt.20-3104-004
were conditioned in a single tube then aliquoted into a PCR microplate (PCR-96-FS-C; Axygen, Corning, NY, USA). The Freedom Evo 150 (Tecan, Männedorf, Switzerland) was used to transfer bead suspension to the UHR total RNA. An extra round of elution and re-binding of mRNA was included with increased incubation time (Table 1). The remainder of the purification process was carried out as per protocol at a 1/20th miniaturized scale up until mRNA fragmentation.

**rRNA depletion**

rRNA was depleted from 100 ng UHR total RNA using the NEBNext rRNA Depletion Kit (Human/Mouse/Rat) (E6350; New England BioLabs) at a 1/6th reaction volume (Fig. 1). 2 μl of UHR total RNA (50 ng/μl) in a 384-well PCR plate (Bio-Rad) was used for input. Master mixes were prepared and aliquoted before each reaction into a column of a 384-well LVSD plate (reagent plate) (SPT Labtech). Reagents were transferred to the samples using the Mosquito HV (SPT Labtech). Ribosomal depleted RNA was cleaned using the provided RNA sample purification beads on a Freedom Evo 150 (Tecan) and eluted in 1.5 μl of nuclease-free water on the Mosquito HV (Table 1).

**Library preparation and sequencing**

**cDNA synthesis**

Purified mRNA was processed using the NEBNext Ultra II Directional RNA Library Prep Kit (E7760; New England BioLabs) at a 1/10th reaction volume (Fig. 1). Poly(A) isolated mRNA was resuspended in 1.5 μl fragmentation mix while 750 nl of fragmentation mix was added to 750 nl of the ribosomal depleted RNA. RNA was fragmented for 10 min at 94°C then placed on ice for 2 min. First-strand cDNA synthesis was performed using 1 μl of the fragmented RNA with remaining material lost as dead volume. Thermocycling was extended during the incubation at 42°C from 15 to 50 min (Table 1). Second-strand synthesis was performed using the NEBNext Ultra II Linear Amplification Kit (E7330; New England BioLabs) at a 1/10th reaction volume.
synthesis and cDNA purification were carried out per protocol at the miniaturization scale.

**Illumina library generation**

Libraries were generated using the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs) at a 1/10th reaction volume starting with 5 μl of cDNA (Fig. 1). Adaptors were ligated by adding 500 nl of NEBNext Adaptor diluted to 0.75 μM in adaptor dilution buffer (Table 1). The adaptor-ligated material was PCR amplified (14 cycles). The indexed libraries were purified by a two-part clean, first with 1.2× PCRClean DX SPRI beads (C-1003-250; ALine Biosciences, Woburn, MA, USA) followed by 0.9× SPRI buffer [20% PEG-8000 (w/v)] (NC0107553; Thermo Fisher Scientific), 2.5 M NaCl (AM9759; Thermo Fisher Scientific), 10 mM Tris-HCl (pH 8.0) (J22638AE; Thermo Fisher Scientific), 5 mM EDTA (J15694AE; Thermo Fisher Scientific)] (Table 1). Concentrations of the libraries were determined fluorometrically using SYBR Green I nucleic acid gel stain (S7585; Thermo Fisher Scientific) and spot checked on Fragment Analyzer (Agilent Technologies) to assess size distribution. Libraries were pooled equimolar for sequencing and concentration was determined by quantitative PCR using the Kapa SYBR Fast qPCR kit (Kapa Biosystems, KK4601, Wilmington, MA, USA) and internal standards. The pooled libraries were sequenced by single-end 50 base-pair run on the Illumina HiSeq2000 (Illumina, San Diego, CA, USA).

**RNA-Seq data analysis**

**Quality control**

Reads were aligned against the hg19 (Feb 2009) human genome assembly using bwa mem v. 0.7.12-r1039 [http://bio-bwa.sourceforge.net/] with flags –t 16 –f and mapping
rates, fraction of multiply-mapping reads, number of unique 20-mers at the 5’ end of the reads, insert size distributions, and fraction of rRNAs were calculated using bedtools v. 2.25.0. 11 In addition, read density across genomic features was estimated for RNA-Seq specific quality-control metrics by randomly down-sampling each resulting bam file to a million reads and then aligned against hg19.

RNA-Seq mapping and quantitation

Reads were aligned against GRCh38/ENSEMBL 89 annotation using STAR v. 2.5.3a 12 with the following flags -runThreadN 8 –runMode alignReads –outFilterType BySJout –outFilterMultiMapNmax 20 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –outFilterMismatchNmax 999 –alignIntronMin 10 –alignIntronMax 1000000 –alignMatesGapMax 1000000 –alignIntronMax 1000000 –outSAMtype BAM SortedByCoordinate –quantMode TranscriptomeSAM with –genomeDir pointing to a 75 nt-junction GRCh38 STAR suffix array. Gene expression was quantitated using RSEM v. 1.3.0 13 with the following flags for all libraries: rsem-calculate-expression –calc-pme –alignments –p 8 –forward-prob 0 against an annotation matching the STAR SA reference. Posterior mean estimates of counts and estimated Reads Per Kilobase of transcript, per Million (RPKM) were retrieved.

RESULTS AND DISCUSSION

Illumina RNA-Seq library preparation can be subdivided into 4 stages: isolation of desired RNA species from total RNA, fragmentation of the isolated RNA, conversion of this RNA to cDNA, and library construction by adapter ligation and PCR amplification. Standard RNA-Seq experiments typically begin with the removal of abundant rRNAs through either selection of poly(A) transcripts by hybridization or targeted depletion of rRNA using oligo probes. Separation of mRNA by poly(T) beads was performed at a 20-fold reduction in volume, reducing the total volume from 100 to 5 μl. Miniaturization of a RNaseH-based depletion of rRNA using single-stranded DNA probes was possible as well and could be done using a 6-fold reduction in volume. The subsequent steps were all similarly miniaturized by 10-fold with the final PCR reaction volume reduced from 50 to 5 μl (Fig. 1). Miniaturization scale for each stage was based upon the lowest reproducible transfer volume of the liquid handlers used in this work that provided successful reactions based on Fragment Analyzer.

![FIGURE 2](image_url)

Performance of kit is sustained following the miniaturized protocol. Comparison of quality-control metrics between manual (Gray) and automated (Blue) libraries. A) Example Fragment Analyzer traces of libraries following poly(A) isolation from 100 ng UHR RNA. B) Example Fragment Analyzer traces of libraries following rRNA depletion from 100 ng UHR RNA. C) Percent average rRNA contamination. D) Directionality of libraries determined by average sense/antisense ratio. E) Average number of genes detected. F) Transcriptomic mapping efficiency demonstrated by the average relative abundance of exonic to intergenic reads.
traces. Deviations of the protocol from simple proportional decrease in reagent volumes are listed in Table 1.

In order to assess the performance of this workflow, we compared the behavior of the miniaturized reaction to full-volume reaction prepared by hand. UHR total RNA (Stratagene) was used as a control RNA. For both the poly(A)- and the RNaseH-based kit, traces of libraries prepared following both the automated and full-volume protocols showed consistent size distributions with minimal primer- and adaptor-dimer contamination (Fig. 2A, B).

Sequencing of miniaturized libraries shows a high degree of similarity to full-volume manual preparations. Several RNA-Seq quality-control metrics were evaluated of libraries produced from control UHR total RNA following both the manual and automated protocols for each depletion method (Fig. 2). Both the manual and automated protocols demonstrated similar depletion efficiency of rRNA between each isolation method (Fig. 2C). All libraries showed a pronounced strand bias with minimal deviation in directionality between manual and automated preparation methods (Fig. 2D). Furthermore, total number of genes detected was similar across each library type (Fig. 2E). Within both isolation methods, manual and automated libraries showed comparable exon/intergenic ratios with slightly higher average exonic ratio observed in the automated poly(A) libraries (Fig. 2F).

The miniaturized library preparation workflow was then applied to biological RNA as a pilot experiment as a means to demonstrate its functional capacity with more practical sample conditions encountered in the experimental setting. In order to properly assess this, we prepared parallel libraries with both manual and automated rRNA depletion protocols and analyzed the ability of the miniaturized libraries to recapitulate RNA-Seq data produced by full-volume libraries of real experimental samples used in the study by Trapecar and colleagues (Trapecar et al., unpublished results). To analyze the consistency of our method, we conducted hierarchical clustering of RNA-Seq

**FIGURE 3**

Integrity of sequencing data is maintained following miniaturized prep of real RNA samples. Heat map and hierarchical clustering analysis of protein-coding gene expression from biological samples (n = 10) following manual and automated library preparation (depicted as log2-transformed reads per kilobase of modeled exon per million mapped reads, RPKM). Expression levels are represented as percentile of RPKM values across the whole dataset (corresponding raw RPKM values are indicated under the color scale: blue, low expression; yellow, high expression). Sample labels (cultured cells, orange and purple; T cells, pink and green) and library preparation methods are indicated at the top (manual, gray; automated, blue). Dendrograms represent hierarchical clustering of samples and genes. Solid white line indicates clustering according to cell type and dotted white line indicates separate cluster due to sample dropout.
data between both the manual and automated method, focusing on the expression profiles of each biological sample (Fig. 3). Visualization of gene expression levels shows slightly increased expression of certain gene clusters with the automated libraries, but overall, there is highly consistent color distribution between preparation methods of samples within each cell type. The hierarchical clustering analysis showed that the samples do first split according to cell type, which then mainly clustered according to library preparation technique with one sample that clustered independently and was determined to be a sample dropout.

In summary, we have demonstrated an automated RNA library preparation method that produces libraries for Illumina sequencing at reduced reaction volumes while not sacrificing library performance. With the high-throughput capacity and low-volume capability of the Mosquito HV liquid handling system, we are able to efficiently produce libraries at reduced volumes while simultaneously increasing the number of samples processed at once (Supplemental Table S1), retaining library quality and the integrity of the sequencing data.

ACKNOWLEDGMENTS

The authors are thankful for members of the BioMicro Center at MIT for their constructive comments and discussions on this manuscript. We are grateful for Martin Trapecar and the Griffith lab for contributing to the validation of the work performed here. We thank SPT Labtech for providing the Mosquito HV used in this work. This work was funded by the National Cancer Institute of the U.S. National Institutes of Health (NIH) under Grant P30-CA14051 and by the National Institute of Environmental Health Sciences of the U.S. NIH under Grant P30-ES002109. Samples derived from human cells were used in this study. The authors declare no conflicts of interest.

REFERENCES

1. Kukurba KR, Montgomery SB. RNA sequencing and analysis. Cold Spring Harb Protoc. 2015;2015:951–969.
2. Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.
3. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31–46.
4. Shoner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! Genome Biol. 2011;12:125.
5. Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics. 2014;30:301–304.
6. Andersen D, Rasmussen B, Linnet K. Validation of a fully automated robotic setup for preparation of whole blood samples for LC-MS toxicology analysis. J Anal Toxicol. 2012;36:280–287.
7. Suckling L, McFarlane C, Sawyer C, et al. Miniaturisation of high-throughput plasmid DNA library preparation for next-generation sequencing using multifactorial optimisation. Synth Syst Biotechnol. 2019;4:57–66.
8. Minich JJ, Humphrey G, Benitez RAS, et al. High-throughput miniaturized 16S rRNA amplicon library preparation reduces costs while preserving microbiome integrity. mSystems. 2018;3:e00166-18.
9. Mayday MY, Khan LM, Chow ED, Zinter MS, DeRisi JL. Miniaturization and optimization of 384-well compatible RNA sequencing library preparation. PLoS One. 2019;14:e0206194.
10. Mora-Castilla S, To C, Vaezeslami S, et al. Miniaturization technologies for efficient single-cell library preparation for next-generation sequencing. J Lab Autom. 2016;21:557–567.
11. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–842.
12. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
13. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.