How strandedness is maintained during library preparation

Summary

- The Invitrogen™ Collibri™ Stranded RNA Library Prep Kit for Illumina produces libraries that are >98% stranded
- The Collibri Stranded RNA Library Prep Kit increases confidence in genetic allele assignment due to the high fidelity of strand-specific results

Introduction

RNA sequencing (RNA-Seq) data are growing in importance for the understanding of disease onset and progression. Over the past several years, RNA-Seq has become the preferred method for better understanding the dynamic nature of gene expression and gene profiling. The abilities to detect which strand of the genome is encoding the gene transcript and to differentiate between two distinct alleles encoded on opposite strands of the genome are highly valuable. Stranded RNA-Seq has been shown to more accurately quantify the expression levels of thousands of genes than other metods [1]; very few methods can match the ability of RNA-Seq to provide a detailed and quantitative view of strand specificity, alternative gene splicing, and gene fusions.

Even though RNA-Seq has become ubiquitous in the study of the molecular basis for disease, the process remains challenging. The general workflow for preparing an RNA sample for sequencing depends on the goal of the experiment. Two primary methods, based on (1) whole-transcriptome RNA and (2) messenger RNA (mRNA) present in the sample at the time of collection, have emerged. RNA is harvested from samples such as tissue, cell culture, or blood under the appropriate experimental conditions and depleted of abundant sequences such as ribosomal RNA (rRNA) if whole-transcriptome RNA-Seq is required. If selection of mRNA is desired, then poly(A) mRNA is selected using oligo(dT) enrichment. Following depletion or selection, the resulting RNA is converted into a sequencing library (Figure 1).

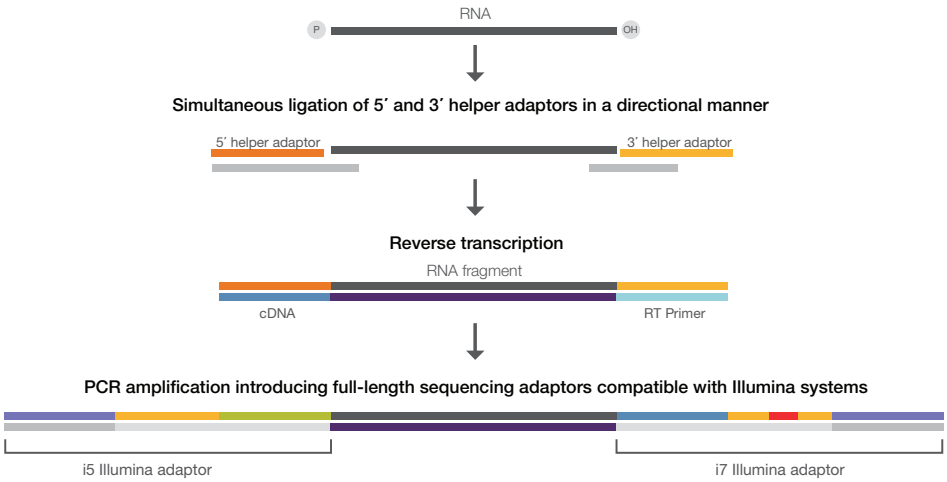

Figure 1. The Collibri protocol generates libraries directly from RNA, rather than from cDNA, for a more accurate representation of sample complexity.

ThermoFisher SCIENTIFIC

After the sample is harvested and any further procedures are completed, the Collibri Stranded RNA Library Prep Kit for Illumina begins with enzymatic fragmentation of the RNA to create fragments suitable for library preparation. Following fragmentation, the RNA sample is hybridized with a helper adaptor mix containing a set of RNA/DNA oligonucleotides with a single-stranded degenerate sequence at one end and a defined sequence at the other end. Ligation enzyme mix is then added to the mixture to ligate the hybridized adaptors. Next, the RNA population with ligated adaptors is reverse-transcribed in a directional manner to generate cDNA. The cDNA is then PCR-amplified, employing helper adaptor sequences as priming sites. As a result, full-length adaptors compatible with Illumina™ systems are generated in single-indexed, ready-to-sequence libraries compatible with single-read or paired-end sequencing. Optimized cleanup steps efficiently remove residual primers and adaptor dimers while preserving high library yields.

Methods

RNA-Seq libraries were prepared from Agilent™ Universal Human Reference RNA (UHRR) and formalin-fixed, paraffin-embedded (FFPE) tissue from liver and lymph node samples. The FFPE material was secured from a European registry with appropriate consents. The FFPE material was highly degraded, resulting in an RNA integrity number (RIN) of 3.1 and 2.4 for liver and lymphoid tissue, respectively (Figure 2). These samples were prepared using the Collibri Stranded RNA Library Prep Kit for Illumina (Figure 1), the KAPA™ RNA HyperPrep Kit with RiboErase (HMR), and the Illumina™ TruSeq™ Stranded Total RNA Library Prep Gold Kit according to manufacturer protocols. All libraries were sequenced on an Illumina™ HiSeq™ 4000 system.

The resulting libraries were demultiplexed using Illumina™ bcl2fastq2 Conversion Software v2.19 and were trimmed using the bbduk v.37.90 trimming and filtering tool. Down-sampling to 47 million reads and 62 million reads was performed for UHRR and FFPE samples, respectively, using the seqtk v1.0-r31 sequence processing toolkit. Alignment of the data was completed using STAR v2.5.3a software with chimeric read detection “on”; otherwise, all default options were used. The alignment was completed using the hg19 reference genome supplemented with the sequences of the Invitrogen™ External RNA Controls

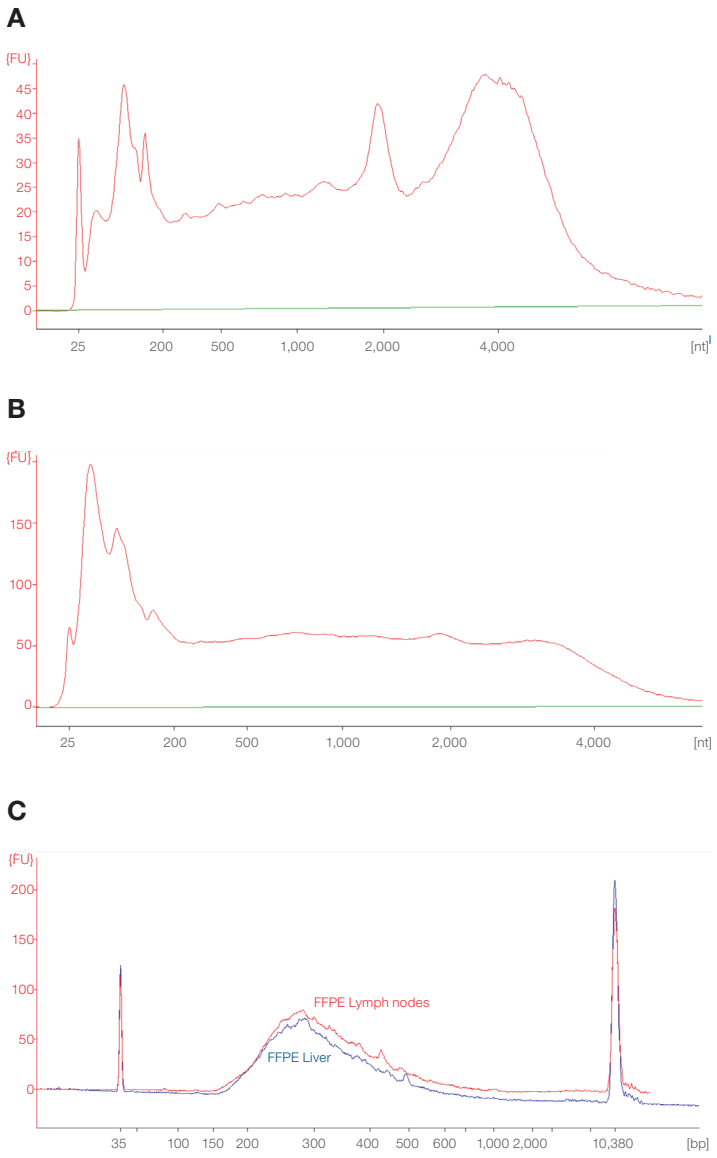

Figure 2. Agilent™ Bioanalyzer™ system traces of highly degraded FFPE RNA and resultant libraries prepared using the Collibri Stranded RNA Library Prep Kit for Illumina. The traces for **(A)** FFPE liver RNA (RIN of 3.1), **(B)** FFPE lymph tissue RNA (RIN of 2.4), and **(C)** libraries prepared using the Collibri Stranded RNA Library Prep Kit for Illumina are shown. Libraries were diluted 3-fold prior to loading onto the Bioanalyzer system.

Consortium (ERCC) spike-in control, and annotations were taken from the ENCODE project. Quality measurements were completed and read counts per gene were determined using the QoRTs v1.1.8 program. Read distributions across introns, exons, and ribosomal fractions were characterized with RSeQC package v2.6.4. Finally, the differential gene expression was measured using DESeq2 package v1.16.

Results

Total RNA-Seq libraries were prepared from UHRR or FFPE samples and the percentage of stranded reads were determined bioinformatically (Figure 3, panels A and B). Greater than 98% of reads aligned to the proper RNA strand from the Collibri Stranded RNA Library Prep Kit even from highly degraded FFPE samples.

Helper adaptors in the Collibri Stranded RNA Library Prep Kit for Illumina anneal directly to RNA and so read one (R1) reports the original RNA template strand and read two (R2) reports the complementary strand. The parameter to detect stranded reads in common analysis tools include:

- TopHat, Cufflinks, or Cuffdiff: library-type fr-secondstrand
- HISAT2: --rna-strandness FR (F in case of single reads)
- HTSeq: --stranded yes
- QoRTS: --stranded_fr_secondstrand

Conclusions

The Collibri Stranded RNA Library Prep Kit provides a high degree of confidence in genetic allele assignment. Through the generation of RNA libraries with highly specific stranded information, greater than 98% accurately stranded data can be achieved from a range of sample types including highly degraded FFPE materials.

Reference

1. Zhao S, Zhang Y, Gordon W et al. (2015) Comparison of stranded and non-stranded RNA-Seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* 16:675.

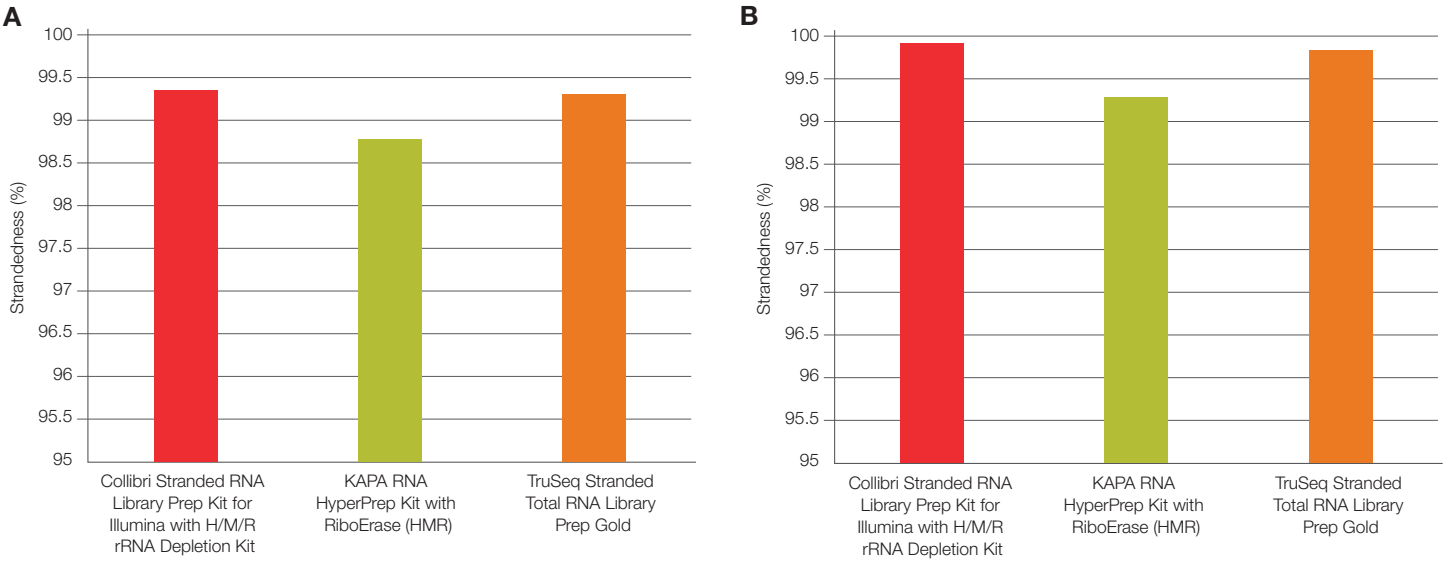

Figure 3. Strandedness results obtained using the Collibri Stranded RNA Library Prep Kit for Illumina with H/M/R rRNA Depletion Kit compared with the KAPA RNA HyperPrep Kit with RiboErase (HMR) and the TruSeq Stranded Total RNA Library Prep Gold. The percent strandedness is shown for libraries prepared from **(A)** Universal Human Reference RNA and **(B)** highly degraded FFPE RNA. All samples were sequenced on an Illumina HiSeq 4000 system and down-sampled to 47 million reads and 62 million reads per sample for Universal Human Reference RNA and FFPE RNA, respectively.

invitrogen

Ordering information

Product	Quantity	Cat. No.
RNA-Seq kits		
Collibri Stranded RNA Library Prep Kit for Illumina*	24 preps	A38994024
	96 preps	A38994096
Collibri Stranded RNA Library Prep Kit for Illumina with H/M/R rRNA Depletion Kit*	24 preps	A39003024
	96 preps	A39003096
ERCC RNA Spike-In Mix	1 kit	4456740
ERCC ExFold RNA Spike-In Mixes	1 kit	4456739
Library quantification		
Collibri Library Quantification Kit*	100 rxns	A38524100
	500 rxns	A38524500
Qubit 4 Fluorometer	1 instrument	Q33238
Qubit 1X dsDNA HS Assay Kit	100 assays	Q33230
Qubit 4 NGS Starter Kit	1 kit	Q33240
Library amplification		
Platinum SuperFi Library Amplification Master Mix	50 rxns	A38539050
	250 rxns	A38539250
Platinum SuperFi Library Amplification Master Mix with Primer Mix*	50 rxns	A38540050
	250 rxns	A38540250
Purification		
PureLink RNA Mini Kit	25 preps	12183020
	250 preps	12183025
MagMAX *mir*Vana Total RNA Isolation Kit	96 preps	A27828
Thermo Scientific accessories		
KingFisher Flex Purification System with 96 Deep-Well Head	1 system	5400630
Applied Biosystems accessories		
Veriti 96-Well Thermal Cycler	1 instrument	4375786
ProFlex 96-Well PCR System	1 instrument	4484075
MicroAmp EnduraPlate Optical 96-Well Clear Reaction Plates with Barcode	20 plates	4483354
MicroAmp Optical 96-Well Reaction Plate	10 plates	N8010560
MicroAmp Clear Adhesive Film	100 films	4306311
MicroAmp 8-Tube Strip with Attached Domed Caps, 0.2 mL	125 strips	A30589

H = human, M = mouse, R = rat.
* Not all kits are available in all countries.

Find out more at **thermofisher.com/collibri**

ThermoFisher SCIENTIFIC