TECHNICAL REPORT

Detection of Genomic DNA in Human RNA Samples for RNA-Seq

Introduction

Most methods of RNA preparation produce RNA samples containing unknown and variable levels of contaminating genomic DNA (gDNA). This gDNA may contribute to erroneous reads in RNA-Seq experiments conducted using NuGEN's Ovation® Human Blood RNA-Seq Library Systems, Ovation Human FFPE RNA-Seq Library Systems and Encore® Complete Library Systems. In order to address this problem, we used quantitative real-time PCR (qPCR) to measure the level of gDNA contamination in several whole blood total RNA samples prepared using the Ovation Human Blood RNA-Seq Multiplex System 1–8. We found a correlation between gDNA contamination and the percentage of intergenic reads in RNA-Seq libraries.

Materials and Methods

Real-time PCR

The two primers described below were ordered from Integrated DNA Technologies (IDT) at 100 nmol scale with HPLC purification.

- 5′-GAGGCTGAGGCAGGAGAATCG
- 5′-GTCGCCCAGGCTGGAGTG

Real-time qPCR reactions were 12 µL total volume, consisting of 6 µL 2X QuantiTect® Master Mix (QIAGEN Cat. #204141), 1.8 µL of 2 µM each primer (final concentration = 300 nM) and 4 µL of either genomic DNA standard or RNA sample.

A standard curve was generated using human female genomic DNA (Promega Cat. #G1521) serially diluted 4-fold from 200 pg/µL to 0.06 pg/µL. For each dilution, 4 µL was added to the 12-µL PCR reaction. RNA samples were diluted to 2 ng/µL. For each of these dilutions, 4 µL was also used per 12 µL qPCR reaction. No-template controls (NTC) using 4 µL of water were also included. A 384-well plate was used with an Applied Biosystems 7900HT Fast Real-time PCR System programmed for 95°C – 15 minutes, followed by 40 cycles of 95°C – 15 seconds, 60°C – 30 seconds.

DNA Contamination

Percent DNA contamination was calculated as nanograms of genomic DNA (as measured by qPCR) per nanogram of RNA sample (as measured using a Nanodrop spectrophotometer, assuming 100% RNA) x 100.

RNA-Seq

The Ovation Human Blood RNA-Seq Multiplex System 1–8 (NuGEN® Part No. 0337-32) was used to make libraries starting with 100 ng of total RNA from human whole blood. Libraries were sequenced using an Illumina Genome Analyzer IIx sequencer, producing 40-nt single-end reads. These reads were aligned to the hg18 human genome reference sequence using the Bowtie short read aligner.[1] Using RefSeq, all reads mapping to 40 or fewer locations in the genome, (M40 reads) were classified as exon or intron. The remaining M40 reads were classified as intergenic. The percent of intergenic reads was calculated by dividing the number of intergenic reads by the total number of M40 reads.

Results and Discussion

In order to increase the sensitivity of qPCR and improve the uniformity of detection across the genome, we chose primers that amplify the *Alu* repeat (~1 million copies per genome) rather than a single copy target. While this *Alu* repeat is expressed, *Alu* RNA is not amplified during PCR.[2] Therefore, any amplification detected is due to gDNA contamination.

Real-time qPCR produced a standard curve of log(pg/µL input) vs Ct with a slope of –3.37 and R^2 = 0.998, with the lowest input (0.06 pg/µL) well above the NTC background. This indicates a PCR efficiency of 98%, with excellent sensitivity as low as 1/100th of a cell equivalent of gDNA.

TABLE 1. **Percentage of gDNA contamination and intergenic reads in different samples**

RNA sample	gDNA contamination	Intergenic Reads
1	0.15%	21%
2	0.01%	15%
3	0.44%	37%

NuGEN
imagine more from less®

Three total RNA samples isolated from human whole blood were analyzed by qPCR for gDNA contamination and prepared for RNA-Seq using the Ovation Human Blood RNA-Seq Multiplex System 1–8. **Table 1** presents gDNA contamination and intergenic reads found in each sample.

Figure 1 shows the correlation between gDNA contamination and intergenic reads. Reads from random fragments of genomic DNA are expected to map randomly across the genome, including intergenic locations.

Conclusion

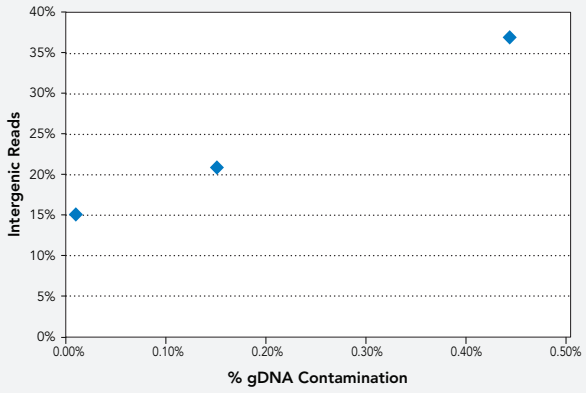
The fraction of RNA-Seq reads assigned as intergenic is strongly dependent on the amount of contaminating genomic DNA present in the RNA sample. The qPCR assay described above can easily measure gDNA contamination in all human RNA samples. Samples containing more than 0.01% genomic DNA by weight can be treated with DNase I (RNase-free) and re-purified to minimize this contamination.

References

1 http://bowtie-bio.sourceforge.net/manual.shtml

2 Marullo et al. *Genome Biology* 2010, 11:R9

FIGURE 1. Increase in intergenic reads with increased gDNA contamination



NuGEN Technologies, Inc.

Headquarters USA

201 Industrial Road, Suite 310
San Carlos, CA 94070 USA
Toll Free Tel: 888.654.6544
Toll Free Fax: 888.296.6544
custserv@nugeninc.com
techserv@nugeninc.com

Europe

P.O. Box 109
9350 AC Leek
The Netherlands
Tel: +31-13-5780215
Fax: +31-13-5780216
europe@nugeninc.com

For our international distributors contact information, visit our website

www.nugeninc.com

M01355 v1