Improved RNA-Seq of Blood-Derived RNA

The Globin-Zero™ Gold Kit increases gene discovery and coverage.

Introduction

Transcriptome analysis by RNA-Seq has rapidly become the method of choice for gene expression profiling, because it is not limited to detecting transcripts that correspond to existing genomic sequences. Blood is a particularly attractive sample for disease studies due to its relative ease of collection, unlike tissues, and its high yield of RNA per sample.

However, blood samples present unique challenges. The collection methods cause fragmentation of the RNA, leading to complications during sample preparation for next-generation sequencing. In addition, as much as 70% of the mRNA in a blood total RNA sample can be globin mRNA, with the remaining total RNA composed of greater than 90% rRNA. Neither globin mRNA nor rRNA contribute high-value RNA-Seq information.

Illumina has developed the Globin-Zero Gold Kit to remove both rRNA and globin mRNA from blood-derived samples. The kit enables deeper sequencing for discovery of rare transcripts and splice variants and reduces the number of wasted sequencing reads.

Methods Overview

In this application, the Globin-Zero Gold Kit was used as part of the ScriptSeq™ Complete Gold Kit (Blood), which is composed of two modules:

1. The Globin-Zero Gold module removes both rRNA and globin mRNA from the sample.
2. The ScriptSeq v2 library preparation module produces directional RNA-Seq libraries from the Globin-Zero–treated sample.

The standard ScriptSeq Complete Gold Kit (Blood) utilizes 1–5 µg of DNA-free blood total RNA. A low-input version of the kit has also been developed for use with precious samples (100 ng to 1 µg of total RNA). Both kits prepare directional cDNA libraries in less than 1 day from fragmented or intact blood total RNA (RNA integrity number [RIN] = 3–10) for single-read, paired-end, or multiplex sequencing on Illumina sequencing systems.

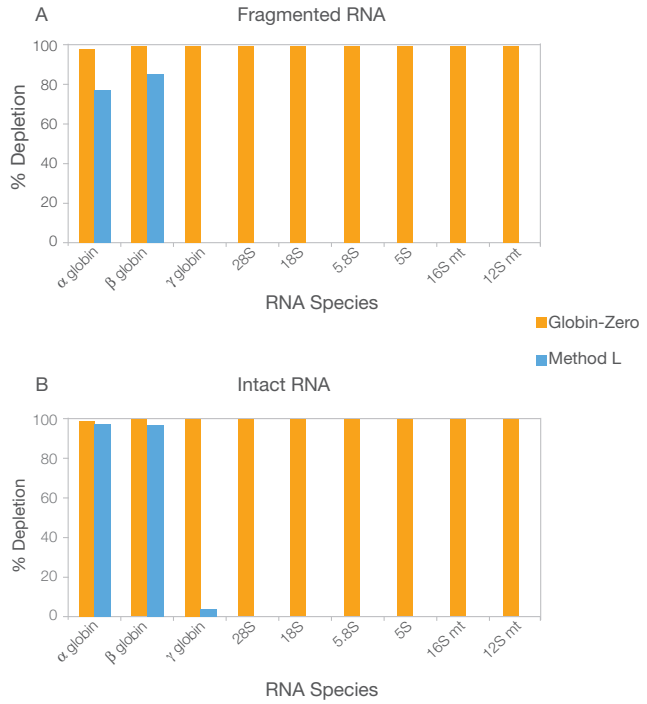

Figure 1: Better Depletion With the Globin-Zero Gold Kit. Five micrograms of (A) fragmented or (B) intact total RNA purified from human erythroid cells (BioChain) was treated using either the Globin-Zero Gold Kit or Method L, an older depletion method. Depletion was assessed by qPCR. The Globin-Zero Gold Kit depleted >99% of rRNA and globin mRNA from all samples.

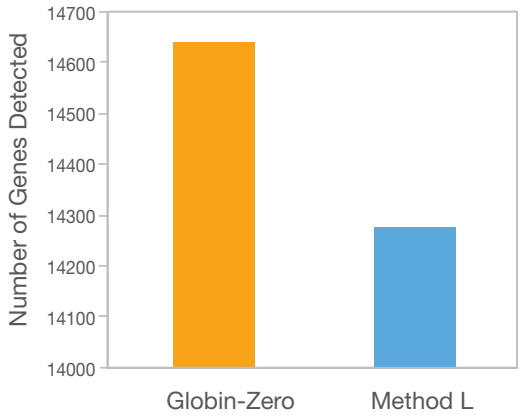

Figure 2: More Genes Detected After Globin-Zero Gold Treatment. One microgram of total RNA derived from blood was treated either with the Globin-Zero Gold Kit to deplete rRNA and globin mRNA, or with Method L, which depletes only some globin mRNA. RNA-Seq libraries were prepared with the ScriptSeq v2 Kit, sequenced on an Illumina Genome Analyzer$_{II}$, and genes were detected by Cufflinks. The RNA treated with Globin-Zero produced libraries with more annotated and novel genes than libraries created after depletion with Method L.

Results and Data Analysis

Significant Reduction in Both Globin mRNA and rRNA

Fragmented or intact RNA samples, derived from human erythroid cells, were treated with either the Globin-Zero Gold Kit or a competitive kit (Method L). Depletion was assessed by quantitative PCR for each of the 3 globin subunits and 6 rRNA subunits. Results showed >99% depletion of all globin mRNA and rRNA subunits by the Globin-Zero procedure in both fragmented and intact samples (Figure 1). For Method L, depletion of globin mRNA from fragmented samples was insufficient, and rRNA was not depleted at all.

More Genes Detected With Globin-Zero Treatment

One microgram of total RNA derived from human blood was treated with either the Globin-Zero Gold Kit or Method L. The resulting RNA was converted into a library with the ScriptSeq v2 Kit and sequenced using a paired-end, 2 × 101 bp run on an Illumina Genome Analyzer$_{II}$. Reads were aligned to the UCSC hg19 annotation using TopHat,[1] and reference and novel genes were detected by Cufflinks.[2] Reference and novel genes are shown combined for a total gene count for each method (Figure 2). Samples treated with the Globin-Zero Gold Kit showed more reference genes and more novel genes than those with Method L, for a combined greater detection of over 360 genes.

Directional RNA-Seq With Enhanced Transcript Coverage

The ScriptSeq v2 library preparation procedure utilizes a unique terminal-tagging process that preserves the transcript orientation and generates directional RNA-Seq libraries.[3] Directionality enables identification of the DNA strand from which the RNA was transcribed—information that is important for discovery and annotation of new transcripts, *de novo* transcriptome assembly, accurate gene expression analysis, and identification of antisense RNAs that are often involved in gene regulation. Enhanced transcript coverage enables identification and discovery of rare transcripts and of splice variants. An example of coverage for the *CD101* gene is shown in Figure 3.

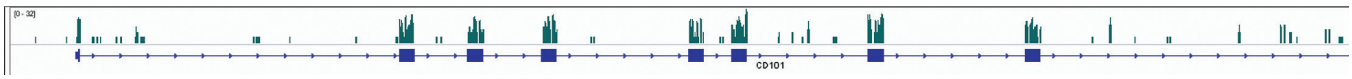

Figure 3: Enhanced Transcript Coverage. Using the ScriptSeq Complete Gold Kit (Blood)–Low Input, 100 ng of blood RNA was prepared for Illumina sequencing. Genes were identified by Cufflinks. The kit increases identification of reference genes and discovery of new genes. Coverage for the *CD101* gene is shown as an example.

AATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTA
AACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAC
ACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGT
CGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACA
AATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTA
TTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACGAA
TATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTAAC

Conclusions

The Globin-Zero Gold Kit is a new method to maximize biologically relevant RNA-Seq information from blood samples. Comprehensive transcript coverage facilitates the discovery of reference and new genes for all phases of clinical research. Elimination of globin mRNA and rRNA from fragmented and intact blood RNA samples yields enhanced transcript coverage for identification of critical genes missed by competitive methods.

References

1. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25(9):1105-1111.
2. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28(5):511-515.
3. Pease J, Sooknanan R. A rapid, directional RNA-Seq library preparation workflow for Illumina sequencing. *Nat Methods* 2012;9:i-ii.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2014 Illumina, Inc. All rights reserved.
Illumina, Globin-Zero, ScriptSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 770-2014-049 Current as of 26 November 2014

illumina®