



Improved RNA-seq of blood-derived RNA increases gene discovery and coverage

RNA-seq of blood-derived RNA can aid discovery of the cause of disease, as well as preclinical research. However, the high content of globin mRNA and ribosomal RNA (rRNA) in blood samples limits the detection of rare transcripts and splice variants. The ScriptSeq™ Complete Kit (Blood) is a new method to prepare directional RNA-seq libraries that are virtually free of globin mRNA and rRNA contamination and exhibit high detection of reference and novel genes.

Introduction

Transcriptome analysis by RNA-seq has rapidly become the method of choice for gene-expression profiling as it is not limited to detecting transcripts that correspond to existing genomic sequences. Blood is a particularly attractive sample for disease studies because of its relative ease of collection, unlike tissues, and its high yield of RNA per sample. Epicentre has developed a new method called ScriptSeq™ Complete Gold Kit (Blood) that produces RNA-seq libraries that are virtually free of globin mRNA and rRNA from fragmented or intact RNA derived from whole blood or erythroid cells. The entire ScriptSeq™ Complete method can be completed in ~7 h and produces directional RNA-seq libraries with high detection of reference and novel genes.

The ability to identify novel genes, splice variants and isoforms is particularly important to discovery, preclinical and clinical research, for which the ultimate goal is to develop a prevention or treatment strategy for over 6,000 rare diseases and 12,000 disease categories. Blood as a sample type, however, presents unique challenges because 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. Removing globin mRNA and rRNA from a blood RNA sample enables deeper sequencing for discovery of rare transcripts and splice variants and reduces the number of expensive sequencing reads that are wasted because they do not lead to prevention or treatment of disease.

Jim Pease & Cris Kinross

Epicentre (an Illumina company), Madison, Wisconsin, USA. Correspondence should be addressed to C.K. (cristine.kinross@epibio.com).

Methods overview

The ScriptSeq™ Complete Gold Kit (Blood) is composed of two modules: the Globin-Zero™ module, which removes both rRNA and globin mRNA from the sample, and the ScriptSeq™ v2 library preparation module, which 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 RNA. A low-input version of the kit has also been developed for use with precious samples of limited amounts of RNA (100 ng to 1 µg RNA). Both kits prepare directional cDNA libraries in less than 1 d from fragmented or intact blood RNA (RNA integrity number = 3–10) for single-read, paired-end and multiplex sequencing on Illumina sequencers.

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™ kit or a competitive kit, here called Method L. Depletion was assessed by quantitative PCR for each of the three globin subunits and six rRNA subunits. Results showed >99% depletion of all globin mRNA and rRNA subunits by Globin-Zero™ in both fragmented and intact samples (**Fig. 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™

One of RNA-seq's greatest strengths is the ability to detect reference and novel genes for the discovery of underlying causes of disease. Choices made during sample preparation for RNA-seq impact the number of genes that can be detected by this new technology. One microgram of RNA derived from human blood was treated with either Globin-Zero™ or an older method available from a different company ("Method L"). The resulting RNA was converted into a sequencing library with ScriptSeq™ v2 and sequenced on a paired-end, 2× 101 bp run on an Illumina sequencer. Resulting reads were

APPLICATION NOTE

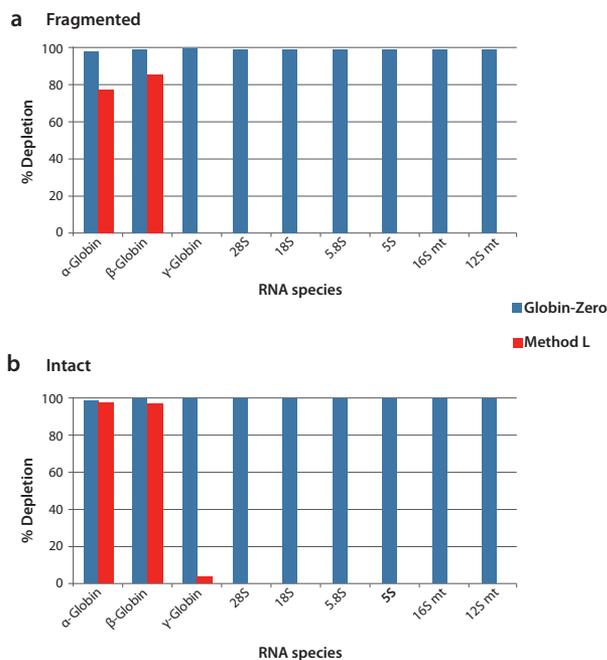


Figure 1 | Globin-Zero™ depletes >99% of rRNA and globin mRNA from all samples. **(a,b)** Five micrograms of **(a)** fragmented or **(b)** intact RNA purified from human erythroid cells (BioChain) was treated with either Globin-Zero™ (blue) or Method L (red), an older depletion method.

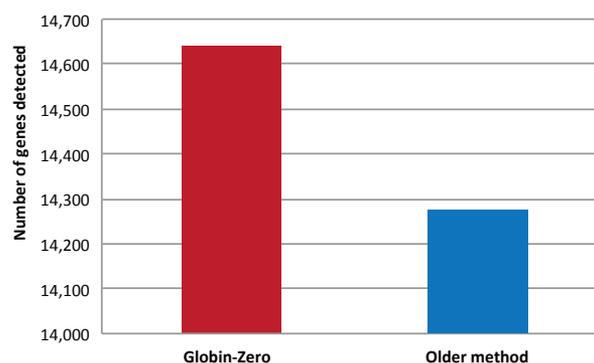


Figure 2 | Detect more genes with ScriptSeq™ Complete Gold Kit (Blood) compared to an older alternative, Method L. One microgram of total RNA derived from blood was treated either with ScriptSeq™ Complete Gold Kit (Blood) to deplete rRNA and globin mRNA or with Method L, which depletes only some globin mRNA. Resulting libraries were sequenced on an Illumina sequencer, and genes were detected by Cufflinks. ScriptSeq™ Complete Gold Kit (Blood) libraries contained more annotated and novel genes than libraries created with Method L.

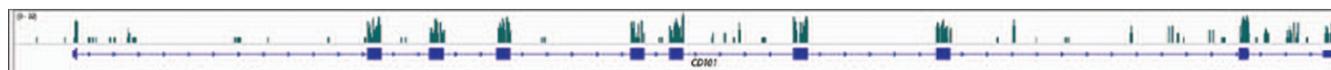


Figure 3 | Enhanced transcript coverage increases identification of reference genes and discovery of new genes. Using ScriptSeq™ Complete Gold Kit (Blood)–Low Input, 100 ng of blood RNA was prepared for Illumina sequencing. The sample was sequenced on an Illumina sequencer, and genes were identified by Cufflinks. Coverage for the *CD101* gene is shown as an example of gene coverage obtained with ScriptSeq™ Complete Gold Kit (Blood)–Low Input.

aligned to the UCSC hg19 annotation using Tophat, and reference and novel genes were detected by Cufflinks. Reference and novel genes are shown combined for a total gene count for each method (**Fig. 2**). Globin-Zero™ detects more reference genes and more novel genes than the older Method L for a combined greater detection of over 360 genes.

Directional RNA-seq with enhanced transcript coverage

The ScriptSeq™ Complete library preparation procedure utilizes a unique terminal-tagging process (Pease, J. & Sooknanan, R. A rapid, directional RNA-seq library preparation workflow for Illumina® sequencing. *Nat. Methods* Application Note, 9, i–ii, 2012) that preserves the transcript orientation and generates directional RNA-seq libraries. Directional RNA-seq libraries enable 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.

Figure 3 shows transcript coverage for the *CD101* gene from 100 ng of blood RNA prepared by ScriptSeq™ Complete Gold Kit (Blood)–Low Input. Enhanced transcript coverage enables identification of reference genes and discovery of new genes and splice variants.

Conclusion

ScriptSeq™ Complete Gold Kit (Blood) is a new method to maximize biologically informative RNA-seq information from blood samples. Comprehensive transcript coverage facilitates 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 by ScriptSeq™ Complete Gold Kit (Blood) yields enhanced transcript coverage for identification of critical genes missed by the older Method L.

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.