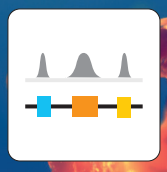
AnyDeplete®– a customizable depletion solution for RNA-Seq workflows.

Application Note



IMPROVING EFFICIENCY OF WHOLE BLOOD MRNA-SEQ USING ANYDEPLETE

INTRODUCTION

Analysis of whole blood can provide important insights into the real-time physiological and pathological status of an individual, and is therefore an ideal, easily accessible sample type for clinical surveillance studies (1,2). Transcriptome analysis from whole blood using next generation sequencing (NGS) is an unbiased method which can help identify biomarkers to determine disease presence or progression. However, one of the major challenges when performing whole blood mRNA-Seq is the presence of significant amounts of globin mRNA (up to 80 percent) that usurps sequencing space without providing useful biological information. Consequently, it becomes challenging to uncover biologically relevant transcripts, particularly if they are expressed at low levels (3,4). Deeper sequencing can potentially address these challenges, but this comes with a substantial increase in overall sequencing costs per sample.

One approach to improve the overall efficiency of RNA-Seq has been through removal of unwanted transcripts. Methods such as GLOBINclear™ (ThermoFisher) and Globin-Zero® (Illumina) can specifically remove globin mRNA from a population of total RNA (5,6). These methods involve manipulating the input RNA to remove globin transcripts using a hybrid capture method. A second selection, using oligo(dT) beads to isolate polyadenylated transcripts, can then be used to study the remaining mRNA transcripts. This tedious and time-consuming workflow poses multiple challenges, including the requirement for a large amount of starting material, increased library preparation costs and multiple manipulations of the input RNA, which can result in the introduction of bias or loss of input material due to RNA instability or RNase contamination.

This application note describes a streamlined workflow for mRNA-Seq from whole blood using the Universal Plus mRNA-Seq library prep kit with Human Globin AnyDeplete (Figure 1). AnyDeplete is a targeted probe-based approach for removal of unwanted transcripts such as globin. Unlike the previously described methods, AnyDeplete-mediated transcript depletion occurs after adaptor ligation, targeting the more stable cDNA, and therefore does not carry the same risks associated with hybrid capture methods. Furthermore, AnyDeplete is customizable to any transcript, and a combination of probes can be created to deplete all unwanted transcripts at one time. This application note details the utility of the Universal Plus mRNA-Seq workflow for whole blood mRNA-Seq studies.

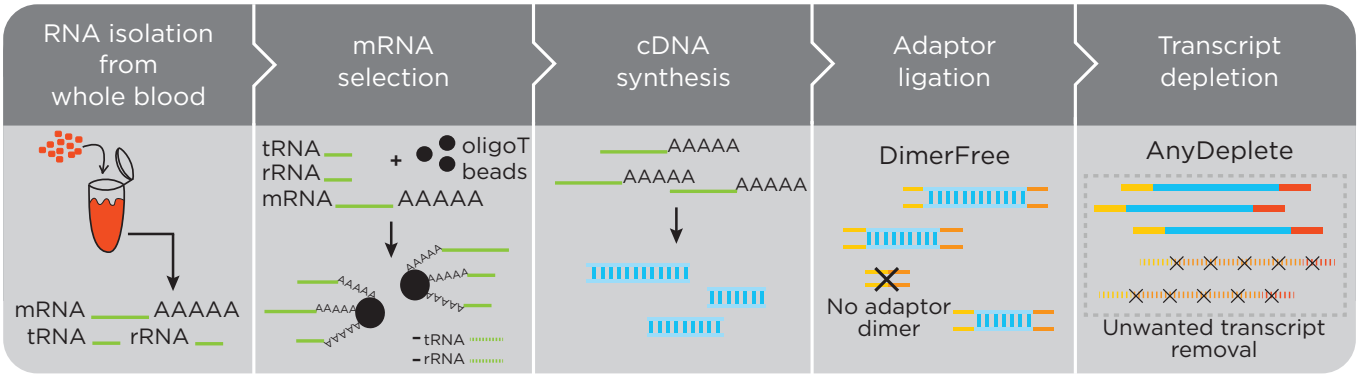
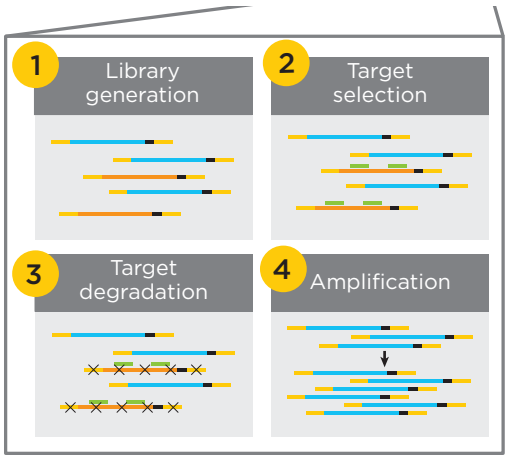


Figure 1: Universal Plus mRNA-Seq is a complete mRNA-Seq workflow, from mRNA selection to library construction with optional targeted depletion using AnyDeplete.



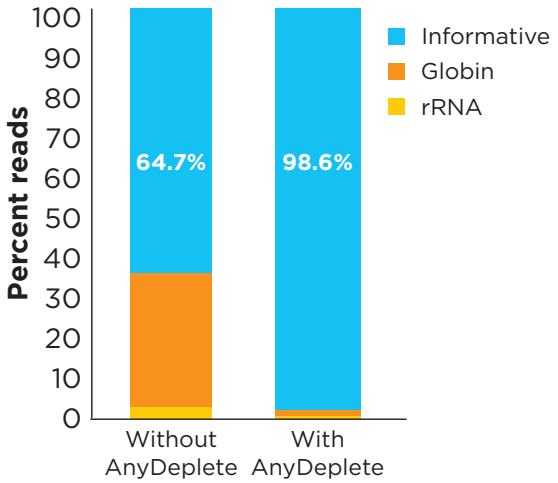


Figure 2: AnyDeplete efficiently depletes globin and increases informative reads. Read distribution metrics from 100 ng whole blood. Informative reads include reads aligning to exons, introns and intergenic regions.

METHOD OVERVIEW

Total RNA was isolated from whole cord blood collected from three full-term infants, as well as peripheral blood from one adult, using standard RNA isolation methods. The Universal Plus mRNA-Seq kit was used to generate mRNA-Seq libraries from 100 ng of adult blood total RNA and 500 ng of fetal blood total RNA. The fetal blood total RNA input was increased to compensate for the higher abundance of globin transcripts, and to allow comparison with existing workflows. Total RNA was added to oligo(dT) beads to select for polyadenlyated transcripts. The selected transcripts were then fragmented and converted to cDNA using random and oligo(dT) priming. After second strand synthesis, the ends of the cDNA were repaired and the adaptors were ligated. Unwanted globin transcripts were depleted using an AnyDeplete probe mix targeting adult globin transcripts (*HBA1, HBA2, HBB,* and *HBD*) or a custom AnyDeplete probe mix targeting both adult and fetal globin (*HBG1*) transcripts. The resulting libraries were PCR amplified, quantified and sequenced on a HiSeq® 4000 (Illumina).

Sequencing data was analyzed using standard methods. Briefly, after sample demultiplexing, reads were trimmed to remove adaptors and low quality sequences. The trimmed reads were aligned to rRNA and globin sequences using Bowtie and removed. The remaining reads were then aligned to the human genome (hg19) using the STAR aligner. After alignment, FPKM values were calculated with Cufflinks for all annotated RefSeq genes (7). Genes with an FPKM ≥1 are reported.

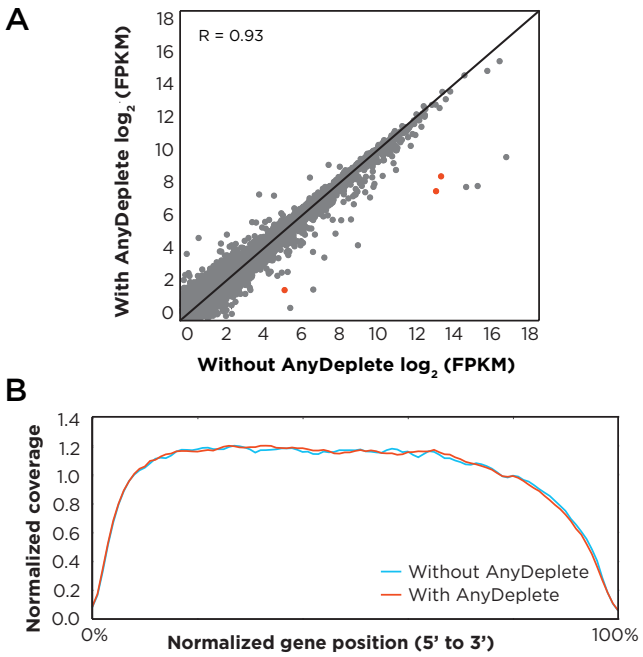


Figure 3: **(A)** FPKM correlation plot comparing samples with and without AnyDeplete. Significantly reduced globin genes shown in red. **(B)** Gene coverage plot comparing samples with and without AnyDeplete.

RESULTS AND DATA ANALYSIS

The adult peripheral blood sample served as a control. Starting with 100 ng total RNA, the standard libraries showed the presence of 35 percent globin reads making only ~65 percent of the total reads available for downstream analysis (Figure 2). A parallel library that included targeted depletion of globin with AnyDeplete showed a reduction of globin to fewer than 2 percent of reads, increasing the number of biologically informative reads to ~99 percent – an increase of over 30 percent when compared to the library without globin depletion (Figure 2).

To ensure that the globin depletion did not introduce bias, the FPKM correlation of RefSeq genes between the libraries with and without AnyDeplete were compared. This showed high correlation between the two libraries (Figure 3A). Globin genes showed a statistically significant difference in expression between the two conditions (red dots, Figure 3A). Gene coverage between libraries with and without AnyDeplete were also compared, and demonstrated even 5' to 3' coverage regardless of depletion (Figure 3B). Taken together, this data demonstrates the reliable and unbiased performance of AnyDeplete.

The fetal cord blood mRNA-Seq libraries showed significantly more globin reads than the adult peripheral blood control. All three fetal blood samples resulted in

similar library metrics. Data from Sample 1 is represented in Figures 4 and 5. In the absence of AnyDeplete, globin and rRNA represented 79 percent of the total reads with only 21 percent of the reads providing useful information (Figure 4). Depletion of adult globin genes with AnyDeplete reduced the proportion of unwanted globin reads by ~33 percent (and rRNA by 2.6 percent), increasing the number of informative reads to ~56 percent (Figure 4). The AnyDeplete probe set was modified to also target fetal globin, allowing further reduction of unwanted reads. Custom depletion of adult and fetal globin sequences reduced the number of these unwanted reads to ~11 percent of the total, and increased the number of informative reads to ~89 percent (Figure 4). Moreover, libraries prepared with the custom fetal and adult AnyDeplete probe set resulted in the detection of over 1,500 additional genes in this sample (Figure 5). This provided more data for differential expression, and more useful biological data for downstream studies.

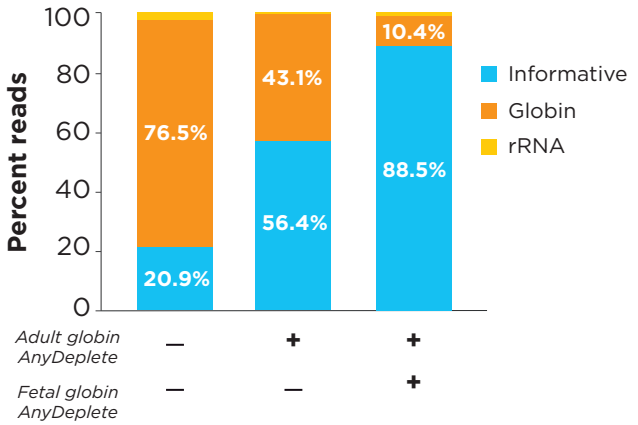


Figure 4: Sequencing read metrics from 500 ng whole fetal cord blood without depletion, with adult globin depletion and with fetal and adult globin depletion for Sample 1. Informative reads include reads mapping to both exons and introns.

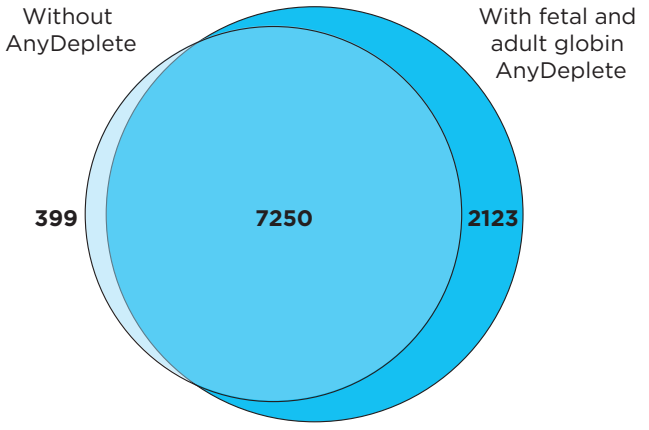


Figure 5: RefSeq genes with FPKM ≥1 detected with and without fetal and adult AnyDeplete probes.

CONCLUSIONS

Whole blood provides an easily accessible, non-invasive sample type that can offer valuable insights into patient health. While targeted methods, such as amplicon sequencing or target enrichment, can provide relevant data of known markers, NGS analysis of blood samples is an unbiased approach for biomarker discovery and detection. This application note demonstrates the use of the Universal Plus mRNA-Seq kit as a straightforward, inexpensive method for mRNA-Seq studies from whole blood total RNA. This kit offers a simple, efficient library construction method for mRNA-Seq studies, with the ability to integrate AnyDeplete-mediated transcript depletion. The AnyDeplete technology provides several key advantages over other depletion methods. Because transcript depletion occurs after adaptor ligation, this eliminates the need for multiple manipulations of the input RNA and reduces the total amount of input RNA required for library construction. A second benefit of AnyDeplete is that it can be customized to any transcript, allowing depletion of fetal globin in this instance.

While the overall cost of sequencing is steadily decreasing, it still represents a significant part of the NGS workflow. mRNA-Seq studies provide high quality gene expression data with reduced sequencing, data storage and data analysis costs compared to whole transcriptome analysis. However, current methods for whole blood mRNA-Seq studies still need significant sequencing depth to overcome the presence of globin reads, resulting in higher costs. Eliminating unwanted globin reads with AnyDeplete yielded more informative data with the same sequencing depth, providing more biologically relevant data or reducing sequencing costs.

REFERENCES

1) Liew CC, et al. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med*, 2006, **147**, 126-132.

2) Mohr S, and Liew CC. The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends in Mol Med*, 2007, **13**, 422-432.

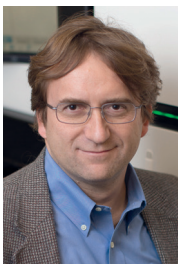
3) Mastrokolias A, et al. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics*, 2012, **13**, 28.

4) Shin H, et al. Variation in RNA-Seq Transcriptome Profiles of Peripheral Whole Blood from Healthy Individuals with and without Globin Depletion. *PLoS ONE*, 2014, **9**, e91041.

5) Liu J, et al. Effects of Globin mRNA Reduction Methods on Gene Expression Profiles from Whole Blood. *J Mol Diag*, 2006, **8**, 551-558.

6) Krjutškov K, et al. Globin mRNA reduction for whole-blood transcriptome sequencing. *Scientific Reports*, 2016, **6**, 31584.

7) Trapnell C, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotech*, 2010, **28**, 511-515.



About the authors

The study described here was done in collaboration with Dr. Piotr Mieczkowski and Dr. Neeta Vora. Dr. Miechzkowski is the Director of the High Throughput Sequencing Facility at the University of North Carolina, Chapel Hill.

For Research Use Only. Not for use in diagnostic procedures.

Australia +61 3 9647 4100 **Austria** +43 62 46 89 330 **Belgium** +32 15 42 13 19 **China** +86 21 220 63 206 **France** +33 4 72 76 04 80 **Germany** +49 79 51 94 170 **Italy** +39 02 92 44 790 **Japan** +81 44 556 73 11 **Netherlands** +31 18 34 48 17 4 **Nordic** +46 8 750 39 40 **Singapore** +65 644 41 886 **Spain** +34 93 595 25 31 **Switzerland** +41 44 922 89 22 **UK** +44 118 9300 300 **USA** +1 919 361 5200 **Other countries** +41 44 922 81 11

401103 V1.0, 2019-07

www.tecan.com

TECAN.