LEXOGEN

Enabling complete transcriptome sequencing

Globin Block Modules for QuantSeq: Globin depletion during library Prep for low-cost, low-input gene expression profiling of blood

The Globin Block Modules for QuantSeq provide an ideal low-cost, low-input approach to library preparation for gene expression profiling of blood. Globin Block enables the generation of globin-depleted, ready-to-sequence 3' mRNA-Seq libraries from as little as 50 ng of total RNA from whole blood. Globin depletion is integrated into the QuantSeq protocol, with no additional steps required, resulting in a significant reduction of globin-mapping reads and enhanced gene detection sensitivity.

Introduction

Blood is a highly accessible and informative tissue that carries great potential for the discovery and monitoring of biomarkers for disease states and physiological functions. RNA Sequencing enables transcriptome-wide gene expression profiling of blood RNA samples, however, the presence of highly-abundant globin messenger RNA (mRNA) in whole blood presents additional challenges for RNA-Seq library preparation.

The most abundant globin mRNAs present in whole blood are transcribed from the hemoglobin alpha and beta globin chain genes (*HBA1*, *HBA2*, and *HBB*). These globin mRNAs account for 50 – 80 % of total RNA in whole blood and therefore, sequester the majority of sequencing reads, severely limiting gene detection and quantification sensitivity [1]. Existing methods deplete globin mRNAs from total RNA prior to RNA-Seq library preparation. However, these require high amounts of total RNA input (typically >1 µg), add additional hands-on time for pre-processing of RNA prior to library preparation, and incur additional cost [1,2].

In contrast, Lexogen's Globin Block Modules for QuantSeq enable globin depletion during the library prep itself, through a simple solution exchange. Lower input amounts starting from 50 ng of total RNA from blood can be used and no additional pre-processing or protocol steps are required.

Workflow

The Globin Block Modules are specifically designed for use with the QuantSeq 3' mRNA-Seq Library Prep Kits for Illumina and work with both the FWD and REV Kit versions (Cat. No's. 015, 016). Globin Blockers are introduced during the QuantSeq protocol in a modified RNA removal solution (RS-Globin Block), which replaces the standard RNA Removal solution (RS) from the QuantSeq Kits (Fig. 1). The Globin Blockers bind to globin first strand cDNA, and prevent the generation of double-stranded cDNA from globin mRNAs during second strand synthesis (Fig. 1).

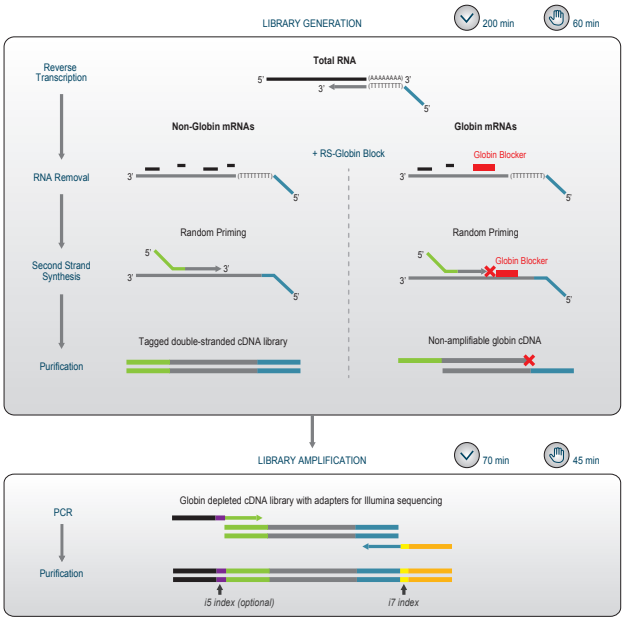

Figure 1 | Workflow for QuantSeq 3' mRNA-Seq Library Prep using Globin Block Modules. Ready-to-sequence libraries are generated in approximately 5 hours. Second strand cDNA synthesis is blocked for globin first strand cDNA. Therefore, globin cDNA fragments lack complete sequencing adapters and cannot be amplified during PCR.

Significant Reduction of Globin Mapping Reads in Human and Pig Whole Blood

Globin Block Modules for QuantSeq are available for human (RS-Globin Block, *Homo Sapiens*, Cat. No. 070.96) and pig (RS-Globin Block, *Sus scrofa*, Cat. No. 071.96). Globin reduction was evaluated for both species by comparing RNA-Seq mapping statistics and gene detection rates (Tables 1 and 2) between libraries prepared with the standard QuantSeq protocol (Standard), and globin-depleted libraries prepared with QuantSeq and Globin Block Modules (+Globin Block).

Library Preparation and Sequencing

Replicate libraries (n=2 or 4) were prepared using the QuantSeq 3′ mRNA-Seq Library Prep Kit for Illumina (FWD), with and without Globin Block. Human whole blood total RNA (RIN 6.9, 50 ng and 250 ng), human leukocyte-enriched blood total RNA (RIN 8.5, 50 ng), and pig whole blood total RNA (RIN 6.5, 100 ng) were used as input for library preparation (see figure legends for RNA extraction details). The standard RNA Removal Solution (RS) was used for the Standard libraries, and the RS-Globin Block, *Homo sapiens* (RS-GBHs) and RS-Globin Block, *Sus scrofa* (RS-GBSs) solutions were used for the human and pig +Globin Block libraries, respectively. As globin mRNAs constitute a majority of mRNA present in whole blood samples, depletion causes a significant reduction in the library yield. To compensate for this and equalize yields, one additional PCR cycle was used for the amplification of +Globin Block libraries compared to Standard libraries (Table 1). Human and pig libraries were sequenced in separate pools on a NextSeq 500 with SR 75 (High Output 75 cycle cartridge: 2.2 pM loaded). Cluster densities were 205-212 K/mm^2 and 256-260 K/mm^2, respectively. For both runs pass filter rates were >88.3 % and %≥Q30 scores were >87.9%.

Kit Performance

Quality control of Standard and +Globin Block libraries showed a reduction in the presence of major globin peaks at 197 bp, 212 bp, 222 bp, 235 bp, and 312 bp for human blood (Fig. 2A), and reduced peaks between 150 – 260 bp for pig blood (Fig. 2B).

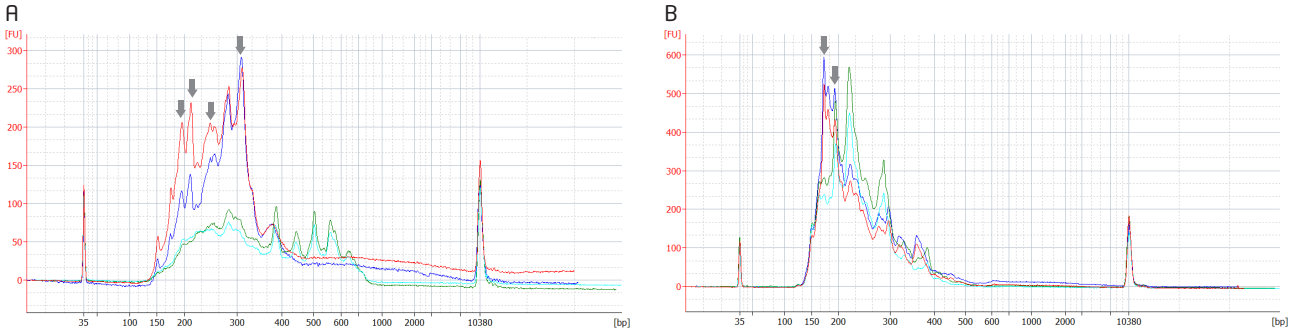

Figure 2 | Bioanalyzer traces for QuantSeq FWD libraries prepared with (+Globin Block) and without (Standard) the RS-Globin Block solution. (A) Human whole blood RNA libraries prepared in duplicate from 250 ng of total RNA, extracted using SPLIT RNA Extraction Kit without red blood cell lysis (Lexogen). **(B)** Libraries prepared from 100 ng of total RNA from pig whole blood, extracted using the Preserved Blood RNA Purification Kit I (Norgen Biotek). Grey arrows indicate distinct peaks in **Standard** libraries (RS Standard, dark blue and red traces) that are reduced in **+Globin Block** libraries (+RS-GBHs and +RS-GBSs, green and light blue traces).

Table 1 | Summary of average mapping statistics and percentages of reads uniquely mapping to globin genes for human and pig blood RNA libraries prepared with Standard QuantSeq or QuantSeq +Globin Block.

Species	Sample Type	Total RNA Input	Standard					+Globin Block					
			PCR cycles	Library Yield (ng/ul)	No. Uniquely Mapped Reads (M)	% Uniquely Mapped Reads	% Total Globin Mapped Reads	PCR cycles	Library Yield (ng/ul)	No. Uniquely Mapped Reads (M)	% Uniquely Mapped Reads	% Total Globin Mapped Reads	% Globin Reduction
Human	Whole Blood	50 ng[1]	15	2.63	7.12	73.2 %	80.7 %	16	2.67	4.68	60.8 %	11.5 %	85.8 %
		250 ng[1]	13	3.27	5.51	81.9 %	75.5 %	14	1.62	3.72	73.6 %	9.7 %	87.2 %
	Leukocyte-enriched Blood	50 ng[2]	15	4.22	3.36	68.5 %	50.1 %	16	3.36	3.07	57.6 %	4.2 %	91.6 %
		50 ng[3]	16	3.99	4.53	71.9 %	12.8 %	17	3.86	4.86	68.7 %	0.7 %	94.3 %
Pig	Whole Blood	100 ng[4]	16	7.19	15.5	92.3 %	61.4 %	17	7.22	13.7	86.5 %	14.4 %	76.5 %

RNA extraction methods: [1] SPLIT without red blood cell lysis. [2] PAXgene® Blood RNA Kit (Qiagen, includes 24 hour partial red blood cell lysis). [3] SPLIT with red blood cell lysis. [4] Preserved Blood RNA Purification Kit I + Dnase I kit (Norgen Biotek).

Globin Block Reduces Globin Mapped Reads down to 0.7 %

Libraries prepared with Globin Block Modules resulted in a significant reduction of total globin mapped read percentages, compared to libraries prepared with Standard QuantSeq (Fig. 3, Table 1). The Globin Block modules specifically target the hemoglobin alpha and beta chain mRNAs (*HBA1*, *HBA2*, and *HBB*). As expected, *HBB*, the most abundant globin mRNA in both human and pig Standard QuantSeq libraries (from whole blood), accounting for >40 % of all mapped reads, was reduced to <6.3 % in +Globin Block libraries. Reads mapping to *HBA1/2* were also reduced to <4.3 % for human and 11.5 % for pig (Table 1).

Total globin mapped read percentages dropped to as low as 0.7 % for leukocyte-enriched blood, and 9.7 % for whole blood in +Globin Block libraries (Table 1, Fig. 3). The percent-reduction of globin mapping reads was slightly better for 250 ng than 50 ng input amounts for human whole blood, and highest for leukocyte-enriched blood extracted using the SPLIT RNA Extraction kit with red blood cell lysis. For pig whole blood, total globin mapped reads were reduced from 61.4 % in Standard QuantSeq libraries, down to 14.4 % in +Globin Block libraries.

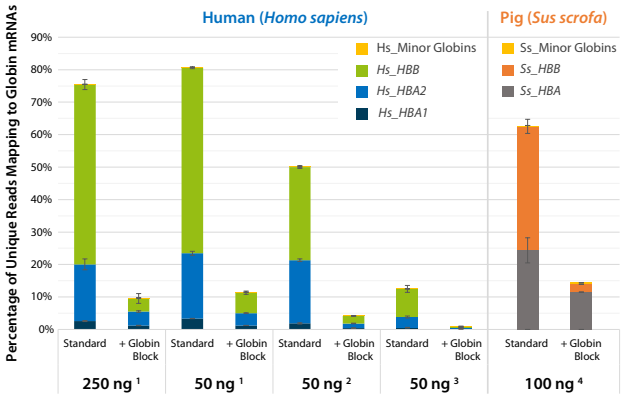

Figure 3 | Percentage of reads uniquely mapping to human and pig globin mRNAs. Libraries were prepared from whole blood RNA with the **Standard** QuantSeq FWD protocol, versus QuantSeq **+Globin Block**. [1] SPLIT RNA Extraction Kit without red blood cell lysis (Lexogen), [2] PAXgene® Blood RNA Kit (Qiagen, includes red blood cell lysis). [3] SPLIT RNA Extraction Kit with red blood cell lysis, [4] Preserved Blood RNA Purification Kit I (Norgen Biotek, without red blood cell lysis).

Enhanced Gene Detection Rates in Globin Block Libraries

As QuantSeq libraries generate one tag per transcript, genes with ≥1 uniquely mapping read can be detected and included for quantification (Table 2). However, to account for variable library size, gene detection rates were calculated after normalization of read counts to counts per million (CPM).

Libraries prepared with 50 ng input from whole blood (without red blood cell lysis), showed higher numbers of additional genes, while 50 ng input leukocyte-enriched blood libraries (SPLIT + red blood cell lysis) showed the smallest increase of only 106 genes. However, the overall number of detected genes was higher for leukocyte-enriched blood libraries compared to those from whole blood (Table 2). In general, up to 2,120 additional genes (raw reads ≥1, or 3,929 at CPM >0.5), were detected in +Globin Block libraries for human, and up to 384 additional genes (raw reads ≥1, or 1,929 at CPM >0.5) were detected in pig +Globin Block libraries (Table 2).

Figure 4 summarizes the unique and overlapping genes detected in Standard and +Globin Block libraries. At CPM threshold 0.5, the majority of genes were similarly detected in Standard and +Globin Block libraries, with up to 3,690 uniquely detected in +Globin Block libraries only. In contrast, only 14, 17, and 136 transcripts were uniquely detected in Standard libraries.

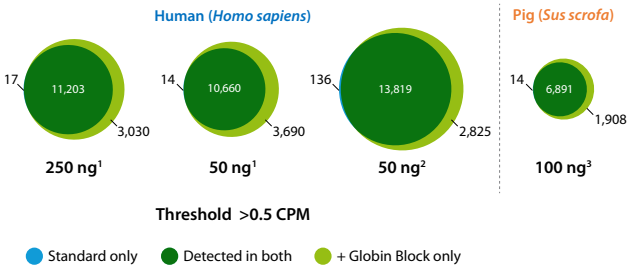

Figure 4 | Increased gene detection in human and pig blood QuantSeq libraries using Globin Block. Libraries were prepared from whole blood RNA with the **Standard** QuantSeq FWD protocol or QuantSeq +Globin Block. Number of detected genes was calculated from CPM-normalized read counts (threshold value >0.5). Gene lists were compared to determine the overlap (dark green), versus genes uniquely detected in Standard (blue) or **+Globin Block** (light green) libraries. Raw numbers of detected genes before normalization are given in Table 2. [1] SPLIT RNA Extraction Kit without red blood cell lysis (Lexogen), [2] PAXgene® Blood RNA Kit (Qiagen, includes red blood cell lysis). [3] Preserved Blood RNA Purification Kit I (Norgen Biotek, without red blood cell lysis).

Table 2 | Summary of average gene detection rates for pig blood RNA libraries prepared with and without Globin Block Modules for QuantSeq. Numbers are given for all genes with ≥1 uniquely-mapped read as well as for CPM-normalized read counts, with different thresholds.

Species	Sample Type	Total RNA Input	Numbers of Detected Genes						Increase in Detected Genes with Globin Block		
			Standard			+Globin Block					
			Raw Reads ≥1	CPM >0.5	CPM >1	Raw Reads ≥1	CPM >0.5	CPM >1	Raw Reads ≥1	CPM >0.5	CPM >1
Human	Whole Blood	50 ng[1]	14,923	10,487	8,667	17,043	14,416	12,721	2,120	3,929	4,054
		250 ng[1]	14,959	11,218	9,570	16,707	14,362	12,668	1,748	3,144	3,098
	Leukocyte-enriched Blood	50 ng[2]	15,912	14,722	12,814	16,955	17,174	15,087	1,043	2,452	2,273
		50 ng[3]	18,138	15,301	13,932	18,244	16,079	14,368	106	778	435
Pig	Whole Blood	100 ng[4]	9,613	6,882	5,741	9,997	8,811	7,933	384	1,929	2,192

[1] SPLIT without red blood cell lysis. [2] PAXgene® Blood RNA Kit (Qiagen, includes 24 hour partial red blood cell lysis). [3] SPLIT with red blood cell lysis. [4] Preserved Blood RNA Purification Kit I + Dnase I kit (Norgen Biotek).

Evaluation of Reproducibility and Detection Limits Using Spike-in Controls

The SIRV-Set 3 Spike-in RNA Variant Controls (Lexogen, Cat. No. 051), containing 69 SIRV isoforms and 92 ERCC transcripts were spiked into each of the human blood RNA libraries (Standard and +Globin Block) for an expected read share of 1 %. Of the 92 ERCCs, 52-58 on average were detected in Standard libraries, while 55-70 were detected in +Globin Block libraries (Table 3). Percentages increased in +Globin Block libraries, reflecting an enrichment of spike-in control mapped reads. This is expected given the gross depletion of globin library fragments (Fig. 2). ERCC input-output correlations were also analyzed for +Globin Block versus Standard libraries (Table 3, Fig. 5). Although, low abundance ERCCs were not consistently detected due to the lower read depth (Table 1), the overall input-output correlations were high and increased in whole blood +Globin Block versus Standard libraries (Table 3).

Table 3 | Summary of ERCC and SIRV detection for QuantSeq Standard and +Globin Block libraries for human blood RNA. Average person correlation coefficients were calculated from ERCC input-output correlations.

Species	Sample Type	Total RNA Input	Standard				+Globin Block			
			No. ERCCs	Pearson Cor.	ERCC %	SIRV %	No. ERCCs	Pearson Cor.	ERCC %	SIRV %
Human	Whole Blood	50 ng[1]	58	0.933	0.47 %	0.49 %	70	0.963	2.83 %	2.79 %
		250 ng[1]	55	0.956	0.62 %	0.63 %	66	0.969	3.24 %	3.05 %
	Leukocyte-enriched Blood	50 ng[2]	52	0.963	0.81 %	0.88 %	55	0.956	2.92 %	3.04 %
		50 ng[3]	56	0.949	1.04 %	1.14 %	58	0.947	1.63 %	1.61 %

[1] SPLIT without red blood cell lysis (Lexogen). [2] PAXgene® Blood RNA Kit (Qiagen, includes 24 hour partial red blood cell lysis). [3] SPLIT with red blood cell lysis.

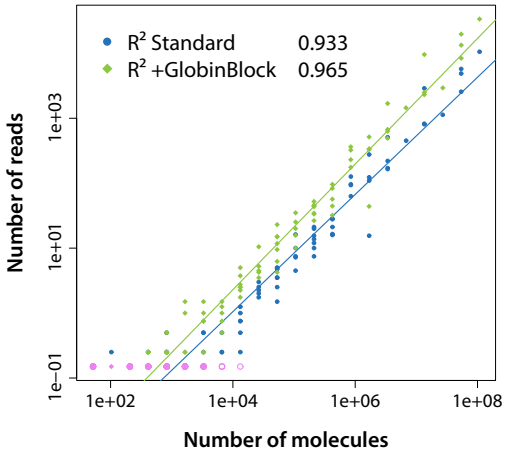

Figure 5 | Input-output correlation for QuantSeq libraries from human whole blood (250 ng). Correlations compared observed ERCC read counts to the expected number of molecules of each ERCC transcript (92). The plots show the correlation for **Standard** (blue) and **+Globin Block** (green) libraries. Levels are plotted for individual replicates (n=4) for each condition, lines represent the linear correlation (R^2). Slope values are 0.890 and 0.930 for Standard and +Globin Block libraries, respectively. Pink circles (bottom left) indicate ERCCs that were not detected in sequencing data.

Globin Depletion for Low Quality RNA

Globin Block Modules for QuantSeq can also be used for globin depletion with low quality RNA samples. Additional tests performed with RNA extracted from frozen blood (RIN 3.2, 50 ng) including red blood cell lysis prior to RNA extraction, resulted in significant globin reduction from 18.1 % down to 0.77 % for +Globin Block libraries (Average unique mapping rates were 63. 2 % for Standard, and 61.6 % for +Globin Block libraries, n=4).

Automated Library Preparation

QuantSeq FWD and REV library prep protocols are also fully automation friendly, with scripts available for a range of liquid handling platforms (from www.lexogen.com). The simple solution exchange to incorporate Globin Block into the QuantSeq workflow requires no alteration to autoQuantSeq protocols, enabling automated, high throughput preparation of globin-depleted libraries from blood RNA.

Advantages

The integrated use of Globin Block Modules within the QuantSeq Library Prep protocol simplifies the preparation of libraries from blood RNA for RNA-Seq-based gene expression profiling. Key advantages as demonstrated by the current results include:

- Significant reduction of globin-mapping reads in QuantSeq libraries prepared from pig whole blood, and human whole and leukocyte-enriched blood.
- Increased detection of up to 1,929 and 3,929 additional genes for pig and human whole blood libraries, respectively.
- The inclusion of red blood cell lysis prior to RNA extraction enhances globin reduction rates, and enables the detection of more genes in total from QuantSeq +Globin Block libraries.
- The enhanced gene detection in human +Globin Block libraries compares well to existing studies that identified 2,112 [1], and 4,671 [3] additional genes in human blood RNA libraries prepared after depletion of globin mRNA from total RNA.
- Where globin depletion prior to library preparation requires input amounts of 1.5 µg and 7.5 – 10 µg of total RNA [1,2,3], similar improvements in gene quantification sensitivity can be achieved using Globin Block with as little as 50 - 250 ng of total RNA input.

References

1. Mastrokolias A., *et al.*, Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics*. 13:28 (2012).
2. Pease J. and Kinross C., Improved RNA-seq of blood-derived RNA increases gene discovery and coverage, Application Note, *Nature Methods*, July. (2013).
3. Shin *et al.*, Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy Individuals with and without globin depletion. *PLOS ONE* 9(3): e91041 (2014).

Acknowledgements

Lexogen wish to thank Claire Rogel-Gaillard and Fany Blanc (INRA, France) for providing the pig blood RNA, and Jack Dekkers (Iowa State University, USA) for initiating the analysis of pig globin depletion.

Ordering Information

Catalog Numbers:
070 (RS-Globin Block, *Homo sapiens*, 96 rxn)
071 (RS-Globin Block, *Sus scrofa*, 96 rxn)
015 (QuantSeq 3′ mRNA-Seq Library Prep Kit for Illumina (FWD))
016 (QuantSeq 3′ mRNA-Seq Library Prep Kit for Illumina (REV) with Custom Sequencing Primer)
020 (PCR Add-on Kit for Illumina)
022 (Purification Module with Magnetic Beads)
047 (i5 Dual Indexing Add-on Kit for QuantSeq/SENSE for Illumina)
051 (SIRV-Set 3 (Iso Mix E0/ERCC))

Find more about Globin Block Modules for Quantseq at www.lexogen.com
Contact us at info@lexogen.com or +43 1 345 1212-41

QUANT SEQ™
Sequencing that counts