Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation of human-specific RNA enrichment kits for transcriptome sequencing in a non-model organism: the African savanna elephant (*Loxodonta africana*)

Stacey Caitlin Engel
engelsc@sun.ac.za

Stellenbosch University
Tracey Jooste
Stellenbosch University
Tanya Jane Kerr
Stellenbosch University
Gian D van der Spuy
Stellenbosch University
Peter Erik Buss
South African National Parks
Léanie Kleynhans
The University of Queensland
Michele Ann Miller
Stellenbosch University

Research Article

Keywords: African savanna elephants, globin mRNA, next generation sequencing, pre-processing, ribosomal RNA, RNA enrichment, RNA sequencing, transcriptomics

Posted Date: August 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10341759/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The use of whole blood to study immune responses in non-model organisms, like African savanna elephants (*Loxondota africana*), has been limited by the lack of species-specific resources. In whole blood, overrepresented globin mRNA and ribosomal RNA can dominate transcriptomic analyses, masking important biologically informative transcripts and therefore need to be removed for downstream applications such as next generation sequencing. Although kits to deplete globin mRNA and rRNA are commercially available for humans, rats and mice, similar kits are not available for non-model species. Therefore, this study aimed to evaluate the efficiency of cross-species globin mRNA (GLOBINclear™-Human Kit) and ribosomal RNA (NEBNext® rRNA Depletion Kit v2) depletion strategies for African savanna elephant whole blood-derived RNA, with the goal of establishing an optimized preprocessing workflow for downstream RNA sequencing.

Results

Following bioinformatic processing, an average of 127.9–278 million paired-end reads per sample were retained with Phred quality scores ranging from Q34 to Q38 across all enriched and unprocessed total samples. Depletion reduced the proportion of rRNA reads from 57.5–73.1% in unprocessed samples to < 3.5% in the enriched samples. However, no significant reduction was observed in the proportion of globin mRNA reads following enrichment. Despite this, genome alignment metrics improved substantially following RNA enrichment, with alignment efficiency increasing from 25.3–40.5% in the unprocessed samples, to 86.2–90.5% in enriched samples. BLAST results revealed moderate sequence similarity between human and African savanna elephant haemoglobin genes (identity: 73.8–79.6%) and high sequence similarity for the rRNA gene targets (identity: 76 to 100% %).

Conclusions

Human-specific RNA enrichment kits effectively depleted African savanna elephant ribosomal RNA prior to sequencing, resulting in a substantial increase in the proportion of uniquely mapped reads to the elephant reference genome. This pre-processing approach enhances the quality of RNA-seq data and supports downstream transcriptomic analyses in African savanna elephant, which will enable investigation of immune-related pathways in this endangered species. The application of this optimized workflow has potential to improve genomic reference resources and enhance the utility of transcriptomics for advancing the understanding of immune responses in African savanna elephants.

Background

Peripheral whole blood serves as a primary medium for transport and communication within the host immune system, comprising a diverse array of circulating immune cells. As such, it has been widely used to study immune responses across a range of diseases and is often preferred over peripheral blood mononuclear cells (PBMCs) because it more closely reflects the physiological environment [1, 2]. However, whole blood contains other components which may interfere with certain downstream applications.

A major challenge associated with whole blood transcriptomics is the high abundance of erythrocytes, which contain large quantities of haemoglobin transcripts. Globin messenger RNA (mRNA) is highly expressed and stable, comprising a substantial proportion of total RNA in blood samples and reducing the sensitivity of detecting less abundant transcripts [3–5]. Consequently, globin depletion strategies have been developed for human samples to improve transcriptomic profiling [6], although their application in non-model organisms remains poorly characterized.

Similarly, ribosomal RNA (rRNA), which constitutes approximately 80–90% of total cellular RNA [7–9], can further limit the detection of informative mRNA species. Methods to remove rRNA, including probe-based hybridisation and commercial depletion kits, are routinely applied in human and model organism RNA sequencing workflows to enrich for coding transcripts [10–12]. However, the effectiveness of these approaches in non-model species is not well established.

In African savanna elephants (*Loxodonta africana*), knowledge of immune responses to diseases with a significant impact on population health, like tuberculosis (TB) or Elephant Endotheliotropic Herpesvirus (EEHV) haemorrhagic disease (EEHV-HD), remains limited [13–15]. Most gene expression studies in this species have relied on targeted PCR-based approaches [16–18], which restrict analyses to predefined genes and may overlook novel or unanticipated targets. Conversely, whole blood transcriptomics offers a powerful, unbiased alternative for characterising immune responses at a systems level in organisms like African savanna elephants.

Despite its potential, whole blood transcriptomics are underutilized in non-model organisms due to technical challenges and limited availability of species-specific reagents. The proportions of globin mRNA and rRNA in African savanna elephant blood are currently unknown but expected to be substantial and kits for such model organisms are not readily available. Given the conservation of many genes between humans and African savanna elephants [19, 20], commercially available human, mouse, and rat depletion kits represent a practical starting point. Therefore, this study aimed to evaluate the efficiency of cross-species globin mRNA and rRNA depletion strategies in African savanna elephant whole blood RNA, with the goal of establishing an optimized preprocessing workflow for downstream transcriptome analyses.

Methods

Elephant whole blood sample collection

Whole blood samples included in this study were obtained from four free-ranging prime adult (> 25 years) bull African savanna elephants in the Kruger National Park (KNP), South Africa, as part of a registered disease surveillance project which includes formal approval and informed consent. All animal procedures, including chemical immobilisation and blood collection, were performed by South African Veterinary Council (SAVC)-registered wildlife veterinarians, as previously described [21, 22].

Whole blood (~ 2.5 mL) was collected from the auricular vein into PAXgene® Blood RNA Tubes (PreAnalytiX® GmbH, Hombrechtikon, 8634, Switzerland), according to the manufacturer's recommendations. After collection, the tubes were inverted several times and maintained upright at room temperature (20–24°C) for 2 hours, followed by storage at -20°C until processing.

RNA extraction and quality control

Blood samples were thawed at 4°C overnight, followed by 2 hours at room temperature prior to RNA extraction. Total RNA was extracted from the PAXgene® tubes using the complementary PAXgene® Blood RNA Kit (IVD) (PreAnalytiX® GmbH). Extractions were performed according to manufacturer's instructions, with modifications to the centrifugation speeds. Initial centrifugation steps to pellet and lyse the cells were conducted at 4 000 x g, while RNA washing and elution steps were performed at 16 000 x g. The extracted RNA was eluted in 75 µL of the elution buffer (BR5) provided in the kit.

Extracted total RNA purity and quantity was evaluated using the NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts 02451, USA) and the RNA High Sensitivity Qubit™ assay (Invitrogen™, Thermo Fisher Scientific) on a Qubit™ 4 Fluorometer (Thermo Fisher Scientific), respectively. In addition, total RNA integrity (quality) was assessed using a TapeStation 4150 (Agilent, Santa Clara, California 95050, USA) with the RNA ScreenTape and associated reagents (Agilent). Extracted total RNA with a yield > 2000 ng and a RNA Integrity Number equivalent (RINe) ≥ 7.0 was used for downstream library preparation and subsequent sequencing.

RNA enrichment

Prior to library preparation, the extracted RNA samples (n = 4) were enriched, as summarized in Fig. 1. First, 1000 ng of total RNA was used as input for the GLOBINclear™-Human Globin mRNA Removal Kit (Invitrogen™), following the manufacturer's instructions to remove globin transcripts. The resulting "globin-depleted" RNA was quantified using the Qubit™ 4 Fluorometer, and its integrity was assessed on the TapeStation 4150, as described above.

Subsequently, approximately 270 ng of "globin-depleted" RNA was used as input for the NEBNext® rRNA Depletion Kit v2 (Human/Mouse/Rat) (New England BioLabs (NEB), Ipswich, Massachusetts 01938, USA), following the manufacturer's instructions for rRNA removal. For the final elution, the rRNA depleted RNA was eluted in 10 µL nuclease-free water and used directly as input for next-generation sequencing (NGS) library preparation.

NGS library preparation and sequencing

Aliquots of eluted enriched RNA (n = 4) and total RNA (n = 2) underwent library preparation using the MGIEasy RNA Library Prep Set (MGI Tech, Shenzhen, 518083, China), following the manufacturer's guidelines. Total RNA aliquots from two of the four elephants was not subjected to depletion and were included in the library preparation and subsequent sequencing as controls to verify the efficacy of the depletion processes (Fig. 1). Briefly, RNA samples were enzymatically fragmented at 87°C for 6 minutes, converted to cDNA, and subjected to a series of enzymatic reactions, including end-repair and A-tailing, adapter ligation, and PCR amplification. Following size distribution assessment and quantification, PCR products from the four enriched RNA samples and two total RNA samples were pooled in an equimolar fashion and converted into a single-stranded, circularized DNA library. The resulting libraries were sequenced on the DNBSEQ-G400RS (MGI Tech) platform at the SAMRC Genomics Platform (Cape Town, South Africa) to generate approximately 127.9 to 278 million paired-end (PE) reads per library with a read length of 100 bp (PE100).

Bioinformatic analysis

All sequence data processing and quality control were performed using the computational infrastructure at the Centre for High Performance Computing (CHPC), South Africa. No custom code was used and all software and parameters employed are listed in Additional file 1 (Table S1).

The quality of the raw sequencing data was assessed using *FastQC* (v 0.12.1) [23], and results were compiled using *MulitQC* (v 1.23) [24]. Subsequently, reads were trimmed to remove adapter sequences (first 10 bp), bases with a Phred quality score below Q20, and reads shorter than 20 bp using *Trimmomatic* (v 0.36) [25]. The quality of the resulting reads was then re-evaluated to ensure that trimming was successful.

The African savanna elephant globin mRNA reference dataset was produced by searching "hemo" in the gene list and included all 16 hemoglobin-related genes that were known or predicted at the search time point [26] available on NCBI (National Centre for Biotechnology Information) [27]. The rRNA dataset comprised 576 ribosomal RNA sequences either annotated or predicted in the species as of April 2024 [28]. This reference file was obtained from NCBI searching for rRNA biomolecules in the African savanna elephant species.

The proportion of globin mRNA and rRNA sequences within the trimmed reads was determined using HTStream (v 1.3.1) [29], using the reference rRNA and globin mRNA datasets described above. In addition, reads were aligned to the African savanna elephant reference genome (mLoxAfr1.hap2; GCF_030014295.1) using *STAR* (v 2.7.7a) [30]. All data visualisation were performed using ggplot2 package (v 4.0.1) [31] in R (v 4.5.1) [32] within the RStudio (v 2025.5.1.513) [33] environment.

To assess the similarity between target sequences of the depletion kits and African savanna elephant genes, BLASTn [34] searches were performed on NCBI using the default parameters. Human target genes

were identified and queried against the RefSeq database, limited to African savanna elephant as the target organism to determine the potential specificity and effectiveness of the kits in this species. For the globin mRNA depletion kit, target genes included human *haemoglobin subunit alpha 1* (NC_000016.10, NC_060940.1), *haemoglobin subunit alpha 2* (NC_000016.10, NC_060940.1) and *haemoglobin subunit beta* (NC_000011.10, NC_060935.1). Similarly, the targets for the rRNA depletion kit, including cytoplasmic *5S* (NC_060925.1), *5.8S* (NC_060937.1), *18S* (NC_060937.1), *28S* (NC_060937.1), *human Internal Transcribed Spacer (ITS)* (NC_000021.9), *E26 transformation-specific (ETS)* (NC_000011.10) and mitochondrial *12S/MT-RNR1* (NC_012920.1) and *16S/MT-RNR2* (NC_012920.1) rRNA genes were screened against African savanna elephant sequences. The BLAST results were ranked based on query coverage and percentage sequence identity.

Results

Quality of RNA

African savanna elephant whole blood RNA was successfully isolated from samples collected in PAXgene® Blood RNA Tubes. The purity, quantity, and integrity of the extracted RNA is summarized in Additional File 2: Table S2. The A260/280 ratios ranged from 2.09–2.28 absorbance units (AU), indicating high RNA purity across all samples. Although the A260/230 ratios (1.13–6.49 AU) in some cases, fell outside the optimal range (2.0–2.20 AU), all samples were retained for downstream analyses (Additional File 2: Table S2). Total RNA yields ranged from 2730 ng to 7065 ng, and RINe values ≥ 7.0 were observed across all samples (Additional File 2: Table S2). The results demonstrate that sufficient quantity of high-quality and intact RNA were obtained for library preparation and subsequent sequencing.

The quality of enriched RNA was assessed post-globin depletion to ensure that high RNA integrity was maintained prior to rRNA depletion and library preparation. Following globin depletion, all samples remained of high quality with RINe values ≥ 7.3 and yields ≥ 384 ng. As expected, a reduction in RNA yield was observed across globin depleted samples, noted by a decrease in the area under the electropherogram peaks (Additional file 3:Figure S1).

Although RNA quality was not assessed after rRNA depletion, the quality of the final sequencing libraries was evaluated prior to sequencing. All libraries generated from both enriched and non-depleted samples exhibited a final yield > 200 ng and a peak fragment size ranging between 261 and 291 bp. These libraries met the sequencing quality and quantity requirements for the PE100 sequencing protocol and were sequenced as an equimolar pool.

Sequence data quality

The raw sequence data comprised a total of 656.2 million paired-end reads (100 bp in length) for the four RNA enriched elephant samples with Phred quality scores ranging from Q34 to Q38 (Additional file 4: Figure S2A and S2B). A total of 433.2 million paired-end reads (100 bp in length) was generated from the

unprocessed total RNA of two elephants. While total RNA samples had similar Phred quality scores (Additional file 4: Figure S2A and S2B), they displayed a markedly higher percentage of duplication (65.2–77.5%) than enriched RNA samples (40.6–51.6).

Reads were further trimmed to remove low-quality bases (below Q20) and reads shorter than 20 bp. After trimming, a total of 641.25 million enriched RNA and 424.51 million total RNA paired-end reads remained, with lengths ranging from 20 to 90 bp (mean length: 89.4 bp), Phred scores of Q34 to Q38, and a GC content of 49–54% (Table 1, Additional File 4: Figure S2C and S2D). The total RNA reads also showed a higher proportion of duplicated sequences (67.5–76.4%) compared to the enriched samples (45.5–49.5%, Table 1).

Table 1
Summary of sequencing statistics for trimmed RNA-seq reads from African savanna elephant whole blood samples.

RNA Enrichment Strategy	Elephant	Total number of reads	Sequence duplication (%)	Uniquely mapped reads (%)	Un-mapped reads (%)	Multi-mapped reads (%)
Enriched RNA	1	130 199 835	43.1	86.7	8.5	4.9
	2	127 896 858	49.5	90.4	6.4	3.3
	3	188 381 104	48.7	89.3	7.2	3.5
	4	194 768 298	42.3	89.0	7.3	3.8
Total RNA	3	146 513 402	67.5	40.4	57.5	2
	4	277 994 979	76.4	25.2	72.4	2.4

Raw reads were generated using paired-end sequencing (100bp) on the DNBSEQ-G400RS platform. Enriched RNA underwent globin mRNA and ribosomal RNA (rRNA) depletion, while Total RNA did not undergo any depletion. After trimming, basic quality metrics were assessed with *FastQC* (v 0.12.1), and mapping efficiencies to the African savanna elephant reference genome (mLoxAfr1.hap2; GCF_030014295.1) were calculated using *STAR* (v 2.7.7a). bp - base pair.

The trimmed reads were mapped to the African savanna elephant reference genome using *STAR* to confirm that majority of the reads were of elephant origin and that minimal contamination was present. All enriched and total unprocessed samples had a mismatch error rate < 0.3% and similar proportion of multi-mapped reads (2–4.9%, Table 2). Mapping efficiencies for the enriched dataset ranged from 86.7 to 90.4% uniquely mapped reads across all four elephants, while the total RNA samples only had 25.3 to

40.3% uniquely mapped reads (Table 1). Additionally, the total RNA samples had a higher proportion of unmapped reads (57.5–72.4%) compared to the enriched samples (6.4–8.5%, Table 1).

Depletion of rRNA and globin mRNA

To evaluate the efficiency of the depletion kits, RNA samples were subjected to globin mRNA and rRNA depletion strategies to remove these abundant transcripts and ensure that less frequently expressed transcripts could be sequenced. The efficiency of these enrichment strategies was assessed post-sequencing using *HTStream* by quantifying the number of reads aligning to globin mRNA and rRNA sequences in the African savanna elephant reference genome.

Following globin mRNA depletion, the proportion of globin transcripts in the sequencing data ranged from 5.5% to 6.2%, compared to 0 to 3.7% in the unprocessed total RNA sample (Fig. 2). Although this difference was not significant, globin levels remained relatively low. Nevertheless, all samples retained sufficient RNA to proceed with subsequent rRNA depletion.

Following rRNA depletion, the proportion of rRNA was markedly reduced to < 3.6%, compared to 57.5–73.1% in the unprocessed total RNA samples (Fig. 2). This reduction resulted in a greater proportion of reads of biological interest with minimal rRNA and globin contamination in the enriched samples relative to the unprocessed total RNA samples. Together, these results indicated the successful depletion of highly abundant globin mRNA and rRNA transcripts in elephant whole blood samples.

To evaluate the specificity of the depletion kits for African savanna elephants, the target sequences were compared to corresponding genes from elephants to assess sequence similarity. The BLAST results for each depletion target were ranked according to their query coverage followed by percentage sequence identity, and the highest ranked African savanna elephant gene was recorded (Table 2). African savanna elephant and human haemoglobin alpha and beta subunits showed moderate to high levels of similarity (Table 2). Interestingly both human *haemoglobin subunit alpha 1* and *2* showed the highest similarity to the African elephant *haemoglobin subunit alpha 2*. In contrast, the rRNA targets had comparatively higher similarity relative to African elephant genes with a query coverage ranging from 94 to 100% and sequence identity ranging from 76 to 100% for cytoplasmic and mitochondrial rRNA species (Table 2). These findings highlight the conserved nature of rRNA genes across species.

Table 2
Sequence similarity between human globin mRNA and rRNA target sequences and the African savanna elephant genome based on BLAST alignment metrics.

Human gene	Accession numbers	Query coverage (%)	Percent Identity (%)	African savanna elephant gene accession
haemoglobin subunit alpha 1	NC_000016.10, NC_060940.1	93	73.8	NC_087353.1
haemoglobin subunit alpha 2	NC_000016.10, NC_060940.1	93	74.1	NC_087353.1
haemoglobin subunit beta	NC_000011.10, NC_060935.1	82	79.6	NC_087348.1
5S ribosomal RNA	NC_060925.1	100	100	NC_087362.1
5.8S ribosomal RNA	NC_060937.1	100	100	NC_087352.1
18S ribosomal RNA	NC_060937.1	100	99.7	NC_087352.1
28S ribosomal RNA	NC_060937.1	94	92.1	NW_026975449.1
ITS	NC_000021.9	54	78.1	NC_087361.1
ETS	NC_000011.10	79	77.7	NC_087356.1
12S (MT-RNR1)	NC_012920.1	99	76.0	NC_087369.1
16S (MT-RNR2)	NC_012920.1	100	76.8	NC_087354.1

Query coverage (%) and percent identity (%) were determined using BLASTn via the NCBI web interface using default parameters. Human gene accession numbers corresponding to each target gene were used as query sequences and searched against the RefSeq database, with the search limited to the African savanna elephant as the target organism. *ETS - E26 transformation-specific, ITS - internal transcribed spacer*

Discussion

To successfully utilize transcriptomics in non-model organisms, such as African savanna elephant, overrepresented transcripts, such as globin mRNA and rRNA, must be removed. This study evaluated the efficiency of a human globin mRNA removal kit and a human/mouse/rat rRNA depletion kit applied to African savanna elephant whole blood RNA prior to NGS. The depleted RNA samples produced high-quality paired-end reads, with a high proportion of uniquely mapped reads to the African savanna elephant genome. In contrast, the unprocessed total RNA samples exhibited a much lower proportion of uniquely mapped reads, despite displaying overall high quality reads (Phred score). While globin mRNA depletion did not result in a significant difference in the read coverage of globin transcripts between enriched and

total RNA samples, the rRNA depletion kit markedly reduced the proportion of rRNA in all enriched samples. Overall, this study demonstrates the successful use of a human/mouse/rat rRNA depletion kit to effectively remove rRNA transcripts from African savanna elephant whole blood total RNA prior to sequencing.

High-quality RNA was successfully extracted from African savanna elephant whole blood stored in PAXgene Blood RNA tubes. The PAXgene system has been widely used in various wild animal studies and field settings and is considered one of the top-performing methods for preserving gene expression and enabling downstream NGS analyses [5, 35–37]. Although PAXgene tubes have previously been used in Asian elephants (*Elephas maximus*), RNA quality metrics were not reported [16]. In the present study, the A260/230 ratios fell outside the expected range, suggesting possible residual salt contamination from the extraction reagents used. However, all samples were retained as the RNA integrity remained high and multiple downstream processing steps were expected to mitigate these effects. The integrity of the RNA remained high following globin depletion, further supporting its use in this workflow and subsequent use for the rRNA depletion assay and NGS.

The overrepresented globin mRNA transcripts were removed using a human-specific depletion kit prior to library preparation. These kits have been successfully applied in human studies to reduce globin transcript abundance [38–40]. High levels of globin mRNA can mask biologically relevant transcripts and pathways [4, 41–43], highlighting the importance of their removal prior to downstream analyses such as differential gene expression or transcriptome assembly. As an alternative to removal, masking of globin mRNA using species-specific oligonucleotides has been successful in humans and holds potential in other species to achieve increased transcript detection [43].

Use of a human-specific globin removal kit did not result in a major reduction in African savanna elephant globin mRNA following treatment. However, a reduction in RNA quantity was observed after globin depletion, indicating that some RNA species were removed during the process. Since the human haemoglobin kit targets showed a relatively moderate similarity to the African savannah haemoglobin genes, these results suggest that some or partial globin mRNA sequences may have been removed without impacting other mRNA of interest. However, this effect may have been obscured by the subsequent rRNA depletion step. Future studies should evaluate globin mRNA depletion independently of rRNA depletion to accurately determine the efficiency of this specific kit to remove globin transcripts in this species.

In contrast, the human/rat/mouse rRNA depletion kit was highly effective in reducing the rRNA transcripts in African savanna elephant whole blood samples. Ribosomal sequences are highly conserved across species [44], and therefore efficient cross-species depletion was expected. This was also supported by the high similarity between the human and African savanna elephant rRNA genes assessed. However, as this kit employs an RNAse H-based depletion strategy, certain transcripts may be under-represented following treatment [40]. The impact of rRNA depletion on the broader transcriptome profile in African savanna elephants therefore warrants further investigation.

In non-model organisms, rRNA removal is often achieved using species-specific probes, which can be costly and time-consuming to develop and optimize [11]. At the time of this study design, the African savanna elephant genome remained incompletely annotated, with many rRNA sequences only predicted and the genome assembly largely at the scaffold level. These limitations supported the use of commercially available human rRNA depletion kits. An alternative approach for RNA enrichment involves the poly(A)-tail selection of mRNA. While some studies have reported lower residual rRNA levels in rRNA-depleted libraries compared to poly(A)-enriched libraries [11], others have shown comparable performance between the two methods [10].

As part of the pre-processing optimisation, a mRNA poly(A) enrichment approach was also attempted using the Dynabeads® mRNA Purification kit (Invitrogen™, Thermo Fisher Scientific), but this proved unsuccessful (unpublished data). This suggests that the enrichment kit may be specific for only human mRNA transcripts, excluding a significant proportion of elephant mRNA, and that the use of a rRNA depletion strategy may be more preferential. Consequently, the depletion strategies evaluated in this study may represent a more robust and suitable approach for preprocessing African savanna elephant RNA prior to transcriptomic analysis.

The number of uniquely mapped reads to the African savanna elephant reference genome (mLoxAfr1.hap2; GCF_030014295.1) not only significantly improved after RNA enrichment but the quality of the reads were also maintained. Without preprocessing, the proportion of unmapped reads was increased due to multiple reads mapping to the repetitive rRNA and globin portions of the genome. With evidence of high sequence duplication in these samples, the contaminating globin mRNA and rRNA present may be responsible for the significant decrease in the proportion of uniquely mapped reads in these unprocessed samples, rather than the reads being of non-elephant origin.

This whole blood RNA preprocessing approach provides a practical alternative to the use of PBMCs in African savanna elephants for transcriptomic analyses, especially when specialised laboratory infrastructure is not available in field-based studies. In human studies, globin mRNA reduction has also been favoured over PMBCs for certain disease contexts like sickle cell disease [39]. Although some transcripts are inevitably lost during globin depletion, the removal of highly abundant transcripts can unmask a greater number of biologically informative genes, thereby enhancing downstream analyses [4]. The broader effects of these depletion strategies on the gene expression profiles in African savanna elephants remains to be explored.

Limitations of this study included the relatively small number of technical replicates per animal and limited number of unprocessed control samples. Although four animals and two sequencing repeats produced consistent results, further validation across multiple samples from the same individuals would strengthen the robustness of these findings. Future work should focus on increasing the sample size, performing these depletion strategies independently and designing species-specific probes for globin mRNA and rRNA removal in African savanna elephant whole blood samples. The application of these optimized workflows to improve genomic reference resources will enhance the utility of transcriptomics for understanding immune responses in African savanna elephants.

Conclusion

The results presented demonstrate the successful application of human globin mRNA and rRNA depletion kits to RNA from a non-model organism, specifically African savanna elephants, to remove overrepresented transcripts. The proportion of rRNA was greatly reduced following the use of the human/mouse/rat depletion kit. Removal of these abundant transcripts increased the potential to detect low levels of expressed genes and supports the use of these kits prior to RNA sequencing for gene expression analyses in non-model organisms. This study also provides a useful framework for the future processing of whole blood RNA in African savanna elephants.

Abbreviations

AU
Absorbance Unit
BLAST
Basic Local Alignment Search Tool
bp
base pair
DNA
deoxyribonucleic acid
EEHV
Elephant Endotheliotropic Herpesvirus
EEHVHD
Elephant Endotheliotropic Herpesvirus Haemorrhagic Disease
GC
guanine (G) and cytosine (C)
KNP
Kruger National Park
mRNA
messenger RNA
NCBI
National Centre for Biotechnology Information
ng
nanogram
NGS
next generation sequencing
PBMC
peripheral blood mononuclear cells
RINe
RNA integrity number
RNA

ribonucleic acid
rRNA
ribosomal RNA
TB
tuberculosis

Declarations

Ethics approval and consent to participate

Ethical approval for this study was granted by the Stellenbosch University Animal Care and Use Research Ethics Committee (SU-ACU-2021-21572; SU-ACU-2024-30181) and the Stellenbosch University Biological and Environmental Safety Research Ethics Committee (SU-BEE-2021-22561). Section 20 approval was granted by the South African Department of Agriculture (DA 12/11/1/7/6; 12/11/1/7/6E-(JD); 12/11/1/7/6E (6192SdlR)). A Threatened or Protected Species (TOPS) permit was obtained through the South African Department of Environmental Affairs (DEA Standing Permit S02556; S65805; S67119 and DEA Registration Certificate 29416; 02256). An approved Research Agreement (SANParks Research Agreement-BUSP1511) was obtained from South African National Parks (SANParks), which included approval by SANParks Animal Care and Use Committee. SANParks is the owner and legal custodian of the elephants used in this study and provided formal approval for their inclusion in the research. All study activities were conducted under SANParks authorization and in compliance with relevant permits and ethical requirements

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the South African Medical Research Council (SAMRC) Centre for Tuberculosis Research, National Research Foundation (NRF) South African Research Chair Initiative (SARChI Grant 86949), NRF Competitive Program for Rated Researcher Grant (CPRR Grant 129268), NRF Black Academics Advancement Programme (BAAP Grant NFSG23040589700), The Crick Africa Network: African Career Acceleration Award (CAN ACA Grant), International Elephant Foundation (IEF), American Association of Zoo Veterinarians (AAZV) Wild Animal Health Fund (WAHF) (AAZV WAHF Grant 2024 #7) and The Tennessee Elephant Sanctuary. Funding to Stacey Engel was provided through a Department of Science and Innovation (DSI) – National Research Foundation (NRF) Doctoral Scholarship.

Author Contribution

SCE performed laboratory procedures, analysed data and drafted the manuscript. TJ assisted in training SCE with experimental procedures and data analysis, provided supervision, guided study design, funding acquisition and edited the manuscript. TJK assisted with sample collection and processing, provided supervision, guided study design and edited the manuscript. GDS provided supervision and edited the manuscript. PEB assisted with sample collection requiring veterinarian skills and manuscript review. LK provided supervision and edited the manuscript. MAM assisted with sample collection from elephants, provided supervision, guided study design, funding acquisition and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors wish to acknowledge Simomotfo Intimate Shongwe, Guy Hausler, Tebogo Manamela, Leana Freese, and Dr. Lufuno Netshitavhadulu from the South African National Parks (SANParks) Veterinary Wildlife Services, Kruger National Park (KNP), Skukuza, South Africa. The authors would also like to thank the South African Medical Research Council (SAMRC) Genomics Platform for providing the sequencing resources and infrastructure. The authors further acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project.

Data Availability

The datasets generated and analyzed during the current study are publicly available on the NCBI Sequencing Read Archive (SRA) under the BioProject accession number PRJNA1499697 and SRR accession numbers SRR39780548 - SRR39780553. The data used to quantify the number of reads aligning to globin mRNA [26] and rRNA [28] sequences are publicly available on NCBI using the link provided in the references. No custom code or unreported software were used in this study. All software and parameters used to analyze the data are stated in the methods section of this manuscript and further summarized in Additional File 1 (Tabe S1).

References

1. Silva D, Ponte CGG, Hacker MA, Antas PRZ. A whole blood assay as a simple, broad assessment of cytokines and chemokines to evaluate human immune responses to *Mycobacterium tuberculosis* antigens. Acta Trop. 2013. doi.org/10.1016/j.actatropica.2013.04.002.
2. Müller S, Kröger C, Schultze JL, Aschenbrenner AC. Whole blood stimulation as a tool for studying the human immune system. Eur J Immunol. 2024. doi.org/10.1002/eji.202350519.
3. Russell JE, Morales J, Liebhaber SA. The Role of mRNA Stability in the Control of Globin Gene Expression. In: Moldave K, editor. Progress in Nucleic Acid Research and Molecular Biology. Academic; 1997. pp. 249–87.

4. Field LA, Jordan RM, Hadix JA, Dunn MA, Shriver CD, Ellsworth RE, et al. Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. Clin Biochem. 2007. doi.org/10.1016/j.clinbiochem.2007.01.004.
5. Mastrokolias A, den Dunnen JT, van Ommen GB, 't Hoen PA, van Roon-Mom WM. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. BMC Genomics. 2012. doi.org/10.1186/1471-2164-13-28.
6. Liu J, Walter E, Stenger D, Thach D. Effects of globin mRNA reduction methods on gene expression profiles from whole blood. J Mol Diagn. 2006. doi.org/10.2353/jmoldx.2006.060021.
7. Blobel G, Potter VR. Studies on free and membrane-bound ribosomes in rat liver: I. Distribution as related to total cellular RNA. J Mol Biol. 1967. doi.org/10.1016/0022-2836(67)90297-5.
8. Wilhelm BT, Landry J-R. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. Methods. 2009. doi.org/10.1016/j.ymeth.2009.03.016.
9. Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? Front Genet. 2015;6:2.
10. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics. 2014. doi.org/10.1186/1471-2164-15-419.
11. Kim IV, Ross EJ, Dietrich S, Döring K, Sánchez Alvarado A, Kuhn C-D. Efficient depletion of ribosomal RNA for RNA sequencing in planarians. BMC Genomics. 2019. doi.org/10.1186/s12864-019-6292-y.
12. Gu H, Sun YH, Li XZ. Novel rRNA-depletion methods for total RNA sequencing and ribosome profiling developed for avian species. Poult Sci. 2021. doi.org/10.1016/j.psj.2021.101321.
13. Long SY, Latimer EM, Hayward GS. Review of elephant endotheliotropic herpesviruses and acute hemorrhagic disease. ILAR J. 2016. doi.org/10.1093/ilar/ilv041.
14. Paudel S, Sreevatsan S. Tuberculosis in elephants: Origins and evidence of interspecies transmission. Tuberculosis. 2020. doi.org/10.1016/j.tube.2020.101962.
15. Rajbhandari RM, De La Fuente J, Karmacharya D, Mathema S, Maharjan B, Dixit SM, et al. Understanding *Mycobacterium tuberculosis* complex in elephants through a one health approach: a systematic review. BMC Vet Res. 2022. doi.org/10.1186/s12917-022-03356-8.
16. Landolfi JA, Schultz SA, Mikota SK, Terio KA. Development and validation of cytokine quantitative, real time RT-PCR assays for characterization of Asian elephant immune responses. Vet Immunol Immunopathol. 2009. doi.org/10.1016/j.vetimm.2009.03.012.
17. Landolfi JA, Miller M, Maddox C, Zuckermann F, Langan JN, Terio KA. Differences in immune cell function between tuberculosis positive and negative Asian elephants. Tuberculosis. 2014. doi.org/10.1016/j.tube.2014.03.001.
18. de Waal CR, Kleynhans L, Parsons SDC, Goosen WJ, Hausler G, Buss PE, et al. Development of a cytokine gene expression assay for the relative quantification of the African elephant (*Loxodonta africana*) cell-mediated immune responses. Cytokine. 2021. doi.org/10.1016/j.cyto.2021.155453.
19. Frönicke L, Wienberg J, Stone G, Adams L, Stanyon R. Towards the delineation of the ancestral eutherian genome organization: comparative genome maps of human and the African elephant

(*Loxodonta africana*) generated by chromosome painting. Proc Biol Sci. 2003. doi.org/10.1098/rspb.2003.2383.

20. Goodman M, Sterner KN, Islam M, Uddin M, Sherwood CC, Hof PR, et al. Phylogenomic analyses reveal convergent patterns of adaptive evolution in elephant and human ancestries. Proc Natl Acad Sci U S A. 2009. doi.org/10.1073/pnas.0911239106.
21. Steyrer C, Miller M, Hewlett J, Buss P, Hooijberg EH. Markers of inflammation in free-living African elephants (*Loxodonta africana*): Reference intervals and diagnostic performance of acute phase reactants. Vet Clin Pathol. 2023. doi.org/10.1111/vcp.13197.
22. Engel SC, Kerr TJ, van der Spuy GD, Jooste T, Buss PE, Johns JL, et al. Optimisation of bronchoalveolar lavage fluid preparation for mononuclear cell isolation and cytologic evaluation in free-ranging African elephants (*Loxodonta africana*). Vet Immunol Immunopathol. 2025. doi.org/10.1016/j.vetimm.2025.110974.
23. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
24. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016. doi.org/10.1093/bioinformatics/btw354.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014. doi.org/10.1093/bioinformatics/btu170.
26. National Center for Biotechnology Information [Internet]. Bethesda (MD). National Library of Medicine (US). Haemoglobin mRNA; Accessed 22 April 2024. Available from: https://www.ncbi.nlm.nih.gov/datasets/gene/taxon/9785/?search=hemo
27. Sayers EW, Beck J, Bolton EE, Brister JR, Chan J, Connor R, et al. Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Res. 2025. doi.org/10.1093/nar/gkae979.
28. National Center for Biotechnology Information [Internet]. Bethesda (MD): National Library of Medicine (US). Ribosomal RNA; Accessed 22 April 2024. Available from: https://www.ncbi.nlm.nih.gov/nuccore?term=txid9785%5BOrganism%3Aexp%5D%20AND%20biomol_rrna%5BPROP%5D&cmd=DetailsSearch
29. UC Davis Bioinformatics Core. HTStream: A high throughput sequence read toolset using a streaming approach. GitHub. 2025. Available from: HTStream. https://github.com/s4hts/HTStream
30. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013. doi.org/10.1093/bioinformatics/bts635.
31. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016.
32. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
33. Posit team. Studio: Integrated Development Environment for R. 2026.
34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997.

doi.org/10.1093/nar/25.17.3389.

35. Schwochow D, Serieys LE, Wayne RK, Thalmann O. Efficient recovery of whole blood RNA - a comparison of commercial RNA extraction protocols for high-throughput applications in wildlife species. BMC Biotechnol. 2012. doi.org/10.1186/1472-6750-12-33.
36. Bowen L, Miles AK, Waters S, Meyerson R, Rode K, Atwood T. Gene transcription in polar bears (*Ursus maritimus*) from disparate populations. Polar Biol. 2015. doi.org/10.1007/s00300-015-1705-0.
37. Bowen L, Manlove K, Roug A, Waters S, LaHue N, Wolff P. Using transcriptomics to predict and visualize disease status in bighorn sheep (*Ovis canadensis*). Conserv Physiol. 2022. doi.org/10.1093/conphys/coac046.
38. Wright C, Bergstrom D, Dai H, Marton M, Morris M, Tokiwa G, et al. Characterization of globin RNA interference in gene expression profiling of whole-blood samples. Clin Chem. 2008. doi.org/10.1373/clinchem.2007.093419.
39. Raghavachari N, Xu X, Munson PJ, Gladwin MT. Characterization of whole blood gene expression profiles as a sequel to globin mRNA reduction in patients with sickle cell disease. PLoS ONE. 2009. doi.org/10.1371/journal.pone.0006484.
40. Jang JS, Berg B, Holicky E, Eckloff B, Mutawe M, Carrasquillo MM, et al. Comparative evaluation for the globin gene depletion methods for mRNA sequencing using the whole blood-derived total RNAs. BMC Genomics. 2020. doi.org/10.1186/s12864-020-07304-4.
41. Shin H, Shannon CP, Fishbane N, Ruan J, Zhou M, Balshaw R, et al. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. PLoS ONE. 2014. doi.org/10.1371/journal.pone.0091041.
42. Das R, Anu A, Jamwal M, Sharma P, Sachdeva MUS, Bansal D, et al. Gene expression profiling of reticulocytes in patients with hereditary spherocytosis after depleting globin gene transcripts. Blood. 2016. doi.org/10.1182/blood.V128.22.1255.1255.
43. Krjutškov K, Koel M, Roost AM, Katayama S, Einarsdottir E, Jouhilahti E-M, et al. Globin mRNA reduction for whole-blood transcriptome sequencing. Sci Rep. 2016. doi.org/10.1038/srep31584.
44. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics. 2007. doi.org/10.1534/genetics.107.071399.

Figures

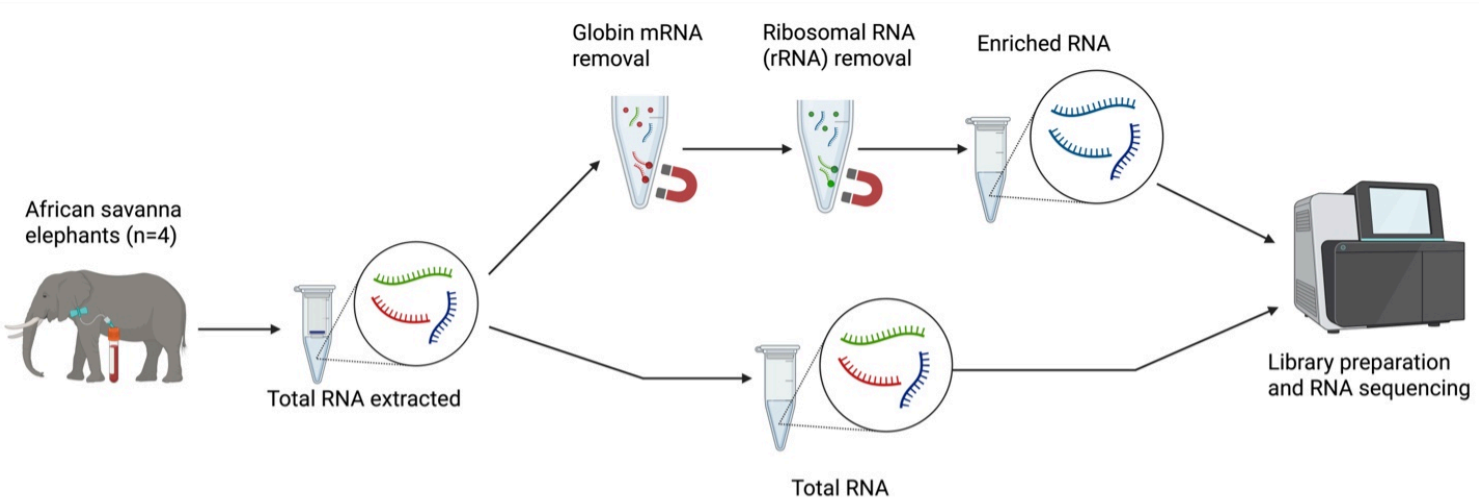


Figure 1

Experimental processing workflow of African savanna elephant whole blood samples. Total RNA was extracted from whole blood collected in PAXgene blood RNA tubes. The RNA enrichment strategy included globin messenger RNA (mRNA) and ribosomal RNA (rRNA) removal using the GLOBINclear™-Human Kit and NEBNext® rRNA Depletion Kit v2 (human/mouse/rat), respectively. Aliquots of eluted enriched RNA (n = 4) and total RNA (n = 2) underwent library preparation and next generation sequencing. Created in https://BioRender.com.

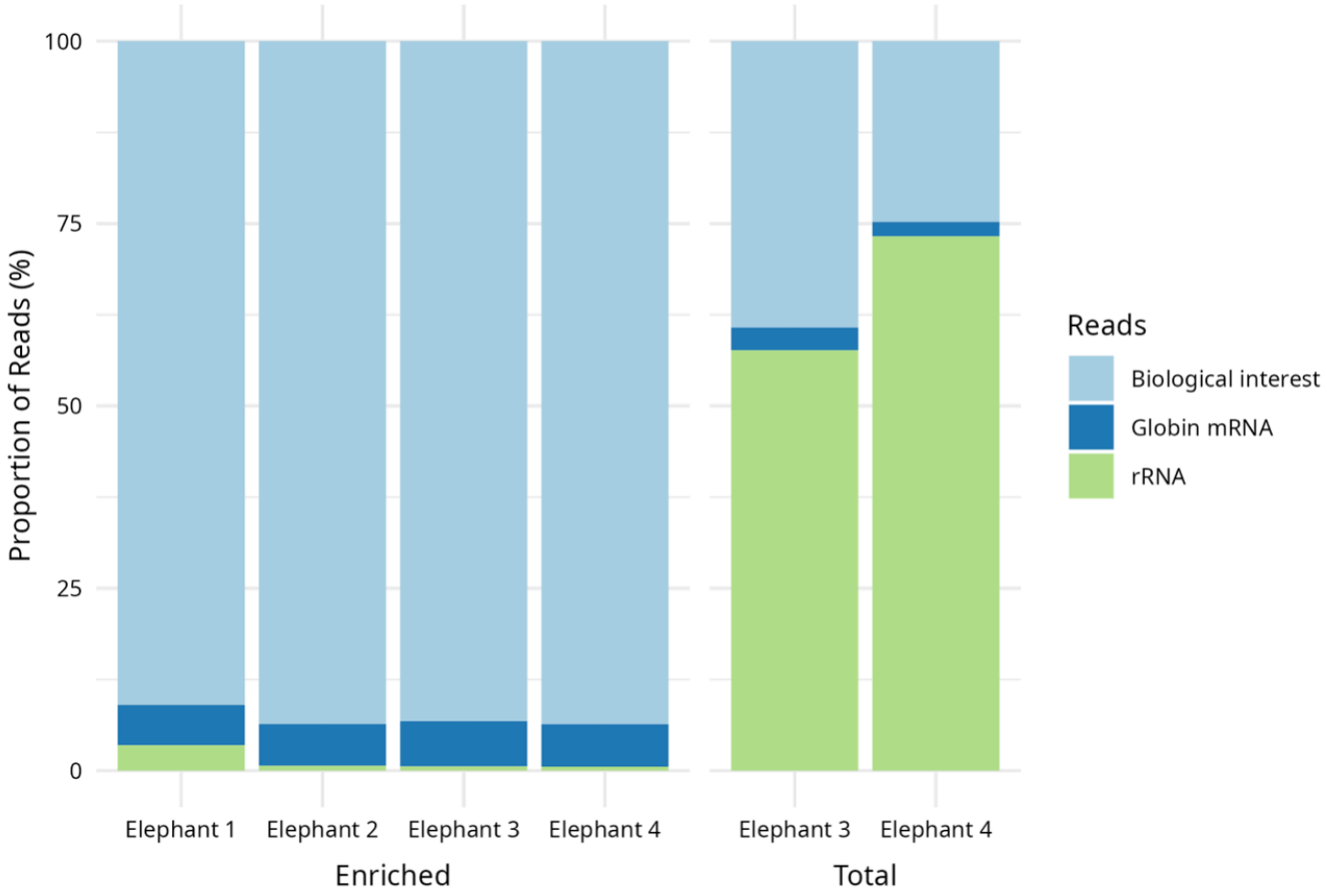


Figure 2

Proportion (%) of reads before (total) and after RNA enrichment (enriched) for RNA derived from African savanna elephant whole blood. Stacked bar plots show the percentage of globin messenger (mRNA) reads, ribosomal RNA (rRNA) reads, and reads of biological interest for each sample.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- UnLinkedSubmittedBMCmanuscriptSupplementary130726.docx