



OPEN

Delayed processing of blood samples impairs the accuracy of mRNA-based biomarkers

Chace Wilson, Nicholas W. Dias, Stefania Pancini, Vitor Mercadante & Fernando H. Biase

The transcriptome of peripheral white blood cells (PWBCs) are indicators of an organism's physiological state, thus making them a prime biological sample for mRNA-based biomarker discovery. Here, we designed an experiment to evaluate the impact of delayed processing of whole blood samples on gene transcript abundance in PWBCs. We hypothesized that storing blood samples for 24 h at 4 °C would cause RNA degradation resulting in altered transcriptome profiles. There were no statistical differences in RNA quality parameters among samples processed after one, three, six, or eight hours post collection. Additionally, no significant differences were noted in RNA quality parameters or gene transcript abundance between samples collected from the jugular and coccygeal veins. However, samples processed after 24 h of storage had a lower RNA integrity number value ($P = 0.03$) in comparison to those processed after one hour of storage. Using RNA-sequencing, we identified four and 515 genes with differential transcript abundance in samples processed after storage for eight and 24 h, respectively, relative to samples processed after one hour. Sequencing coverage of transcripts was similar between samples from the 24-h and one-hour groups, thus showing no indication of RNA degradation. This alteration in transcriptome profiles can impair the accuracy of mRNA-based biomarkers, therefore, blood samples collected for mRNA-based biomarker discovery should be refrigerated immediately and processed within six hours post-sampling.

Blood is a fluid connective tissue that links the entire biological system of an individual, and is composed of plasma and red and white blood cells¹. Liew and colleagues¹ coined the idea of the “sentinel principle”, whereby blood can harbor molecular indicators of physiological changes in organs, tissues, and cells. Gene transcripts in peripheral white blood cells (PWBCs) are among these molecular indicators. The transcriptome profile of PWBCs is distinct from one individual to another^{2,3}, and the profile is as dynamic as the physiological changes that an individual experiences^{4–6}. Most importantly, changes in gene expression are detected in the blood relative to several environmental and pathological factors (reviewed in^{1,7}).

Liquid biopsy, including from blood, has emerged as a powerful source of biological material for studying messenger RNA (mRNA) based biomarkers⁸. For instance, mRNAs have been associated with chemo-sensitivity in advanced gastric cancer patients⁹, non-small cell lung cancer¹⁰, acute ischemic stroke¹¹, neuroendocrine tumor¹², prostate cancer¹³, hepatocellular carcinoma^{14,15}, and Huntington's disease¹⁶. In reproductive health, several studies have focused on changes in genes expressed in PWBCs. A cohort of women who were enrolled in the PREVIENI project¹⁷, and identified as infertile, presented altered levels of diverse genes expressed in the PWBCs relative to fertile women^{18,19}. Recently, we have identified several genes that are differentially expressed when contrasting heifers of different pregnancy outcomes (pregnant by AI, pregnant by natural breeding, or not pregnant)^{20,21}. Therefore, the analysis of mRNAs is one possible avenue for the determination of bloodborne molecules that serve as biomarkers of health.

The processing of blood samples for the separation of the buffy coat followed by resuspension in TRIzol Reagent and immediate cryopreservation at -80 °C is very efficacious for the extraction of RNAs with high quality and purity^{20–23} suitable for producing data by RNA-sequencing^{20,21,24,25}. However, when the site of collection cannot be used to process the blood samples, there is a window of time between sampling and collection. Malentacchi et al. detected the alteration of transcript abundance of one gene, out of seven tested by polymerase chain reaction, when samples were stored for 24 h at 4 degrees Celsius (°C)²⁶. To date, no study has carried out a systematic interrogation of the transcriptome of PWBCs to understand the consequences of storing blood on the alteration of transcript abundance.

School of Animal Sciences, Virginia Polytechnic Institute and State University, 175 W Campus Dr., Blacksburg, VA 24061, USA. email: fbiase@vt.edu

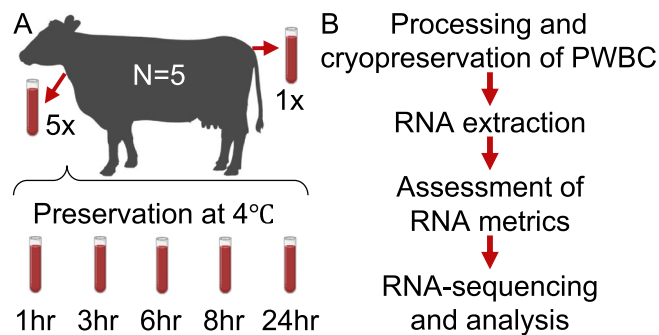


Figure 1. Experimental design and workflow. (A) Overview of the experimental design with the number of subjects and tubes processed at different times after storage at 4 °C. (B) Workflow of the assays from sample processing to RNA-sequencing. Elements of this figure were created with BioRender.com.

Processing Time	RIN		A_{260}		Ratio(A_{260}/A_{280})	
	\bar{x}	$\hat{\sigma}$	\bar{x}	$\hat{\sigma}$	\bar{x}	$\hat{\sigma}$
Coccygeal						
1 h	8.48	0.16	13.27	4.02	1.93	0.02
Jugular						
1 h	8.52	0.37	15.99	2.75	1.93	0.02
3 h	8.60	0.23	17.74	6.34	1.93	0.01
6 h	8.66	0.23	13.76	4.84	1.93	0.02
8 h	8.34	0.33	12.40	6.85	1.91	0.06
24 h	8.00	0.37	14.48	2.52	1.92	0.03

Table 1. RNA parameters obtained from peripheral white blood cells. \bar{x} : average; $\hat{\sigma}$: standard deviation.

Here, we designed an experiment to systematically interrogate the consequences of storing blood samples at different periods up to 24 h at 4 °C on RNA degradation and the transcriptome profile in PWBCs. We hypothesized that the preservation of blood samples for 24 h at 4 °C would lead to RNA degradation, which would result in an alteration in the transcriptome profile of PWBCs.

Results

Overview of the experimental design. We collected blood samples from five estrus-synchronized heifers. Ten mL of blood were drawn from the coccygeal vein and five samples of 10 mL were drawn from the jugular vein within seconds among all samplings within each animal. All samplings were performed within 45 min. All tubes were preserved on ice and the samples from the jugular vein were randomly assigned to different groups for delayed processing (one, three, six, eight or 24 hours (h), Fig. 1A). At the assigned time, PWBCs were isolated, pelleted and resuspended in TRIzol™ Reagent for cryopreservation at –80 °C. We extracted total RNA from all samples in one batch, assessed quantity and quality and submitted all samples for library preparation prior to freezing (Fig. 1B).

Parameters of total RNA based on sampling location and processing delay. We extracted total RNA from 30 samples in one batch, with an average yield of $11.8 \mu\text{g} \pm 4.5$. There was no difference ($P > 0.05$) of the parameters from the samples obtained from the coccygeal versus jugular vein (Table 1, Supplementary Fig. S1). We then compared the effect of delayed processing on the parameters from samples obtained from the jugular vein. There was no difference ($P > 0.05$) for values of absorbance (A_{260} and A_{280}) and the ratio (A_{260}/A_{280}). However, there was an effect ($P = 0.03$) of the time for delayed processing on the RNA integrity number (RIN). The samples processed 24 h post-collection presented lower RIN relative to the samples processed 1 h post-collection ($\bar{x}_{RIN,1hr} = 8.52 \pm 0.37$, $\bar{x}_{RIN,24hr} = 8 \pm 0.37$, $P = 0.03$, Z-test, Table 1, Supplementary Fig. S2).

Quality of libraries produced based on sampling location and processing delay. Because the lowest value for RIN was 7.4, which is suitable for transcriptome analysis²⁷, we proceeded with RNA-sequencing and produced genome-wide transcriptome data for all 30 samples. On average, we produced $29,871,716 \pm 3,365,045$ pairs of reads per sample (ranging from 21,139,000 to 34,856,707, median 29,941,861, Table 2).

There was no difference ($P > 0.05$) on library parameters of 3'/5' bias, efficiency of reads assigned to the annotation, number of genes in relationship to the location of blood sampling, nor the delayed processing of the samples (Table 2, Supplementary Fig. S3 and S4). We noted, that among the samples with delayed processing, the library with the lowest efficiency of reads assigned to the annotation (40%) did not originate from the sample with

Processing Time	Reads Produced		Genes detected before filtering		Genes detected after filtering		Proportion of reads matching annotation		3'/5' Bias	
	\bar{x}	$\hat{\sigma}$	\bar{x}	$\hat{\sigma}$	\bar{x}	$\hat{\sigma}$	\bar{x}	$\hat{\sigma}$	\bar{x}	$\hat{\sigma}$
Coccygeal										
1 h	30,214,519	4,270,887	13,135	167	12,490	51	0.56	0.04	0.51	0.02
Jugular										
1 h	28,452,995	5,098,116	13,055	165	12,462	74	0.58	0.03	0.51	0.03
3 h	32,132,121	1,968,340	13,118	128	12,499	49	0.56	0.03	0.50	0.01
6 h	28,263,832	2,338,260	13,130	139	12,497	57	0.59	0.02	0.51	0.03
8 h	30,483,625	3,040,358	13,102	132	12,478	54	0.57	0.04	0.52	0.02
24 h	29,683,203	2,546,563	13,112	105	12,502	40	0.53	0.08	0.52	0.01

Table 2. Metrics for RNA-sequencing data produced from peripheral white blood cells.

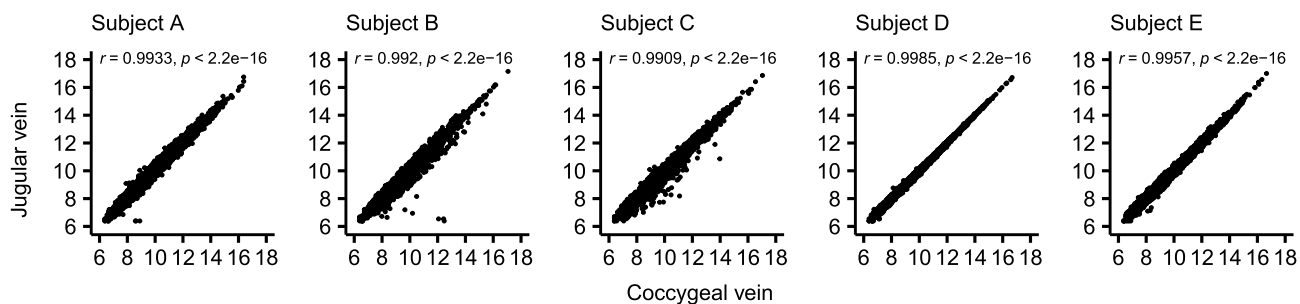


Figure 2. Transcript abundances obtained from the jugular and coccygeal veins from different subjects. The values presented are variance stabilizing read counts obtained from the “DESeq2” package⁵⁵.

lowest RIN (7.4), but both samples were processed after 24 h of preservation at 4 °C. Further interrogation of the relationship between RIN and library percentage of reads assigned to the annotation showed only a moderate correlation between these two metrics (Pearson’s $r = 0.3799$, $P = 0.0610$).

There was no difference ($P > 0.05$) for number of genes detected pre- or post-filtering in relationship to sampling location or delayed processing (Table 2, Supplementary Fig. S3 and S4). After filtering for lowly expressed genes (FPKM > 1 and CPM > 1 in five or more samples), we quantified transcript abundance for 12,414 protein-coding genes, followed by 287 long non-coding RNAs and 109 pseudogenes.

Differential transcript abundance based on sampling location and processing delay. First, we tested whether transcript abundance would be distinguishable based on the location of sampling. The results show no difference (FDR > 0.05) between transcripts from samples obtained from coccygeal or jugular veins. Next, we assessed the consistency of transcript abundance within each animal by calculating the correlation of the transcript abundance between the two sources of sampling. The Pearson’s correlation coefficients were greater than 0.99 for all subjects. Both results convergently show no variation in transcript abundance within subject based on sampling source (Fig. 2). Thus, mRNA quantitation data collected from liquid biopsies are consistent regardless of which vein is used for sampling.

Second, we asked if the transcript abundance in PWBCs would change if blood samples remained stored at 4 °C for different periods of time, relative to the processing of the blood samples within one hour of collection. There was no differential transcript abundance between the samples stored for three or six hours at 4 °C relative to the samples processed within one hour of collection (FDR > 0.05 , Fig. 3A).

By comparison, we identified four and 515 genes with differential transcript abundance between samples stored for eight and 24 h, respectively, at 4 °C relative to the samples processed within one hour of collection (Fig. 3A). Notably, the four genes detected in the ‘8 h vs 1 h’ contrast were also detected in the ‘24 h vs 1 h’ contrast with higher transcript abundance in the preserved samples relative to those processed within one hour of collection (Fig. 3B). Furthermore, 291 and 224 genes presented greater and lower abundance, respectively, for the ‘24 h vs 1 h’ contrast (please see Supplementary Table S1 for the lists of genes, and Supplementary Fig. S5 for the individual graphs of transcript abundance for all genes for the contrast ‘24 h vs 1 h’). These results show that storage of blood samples for ≥ 8 h prior to cryopreservation of PWBCs causes significant changes in the transcriptome profile.

Analysis of the relationship between the decline in transcript abundance and mRNA coverage. Considering the results of differential transcript abundance, we asked if lower values of FPKM for transcripts in the samples processed after 24 h of storage at 4 °C were caused by reduced transcript coverage, which

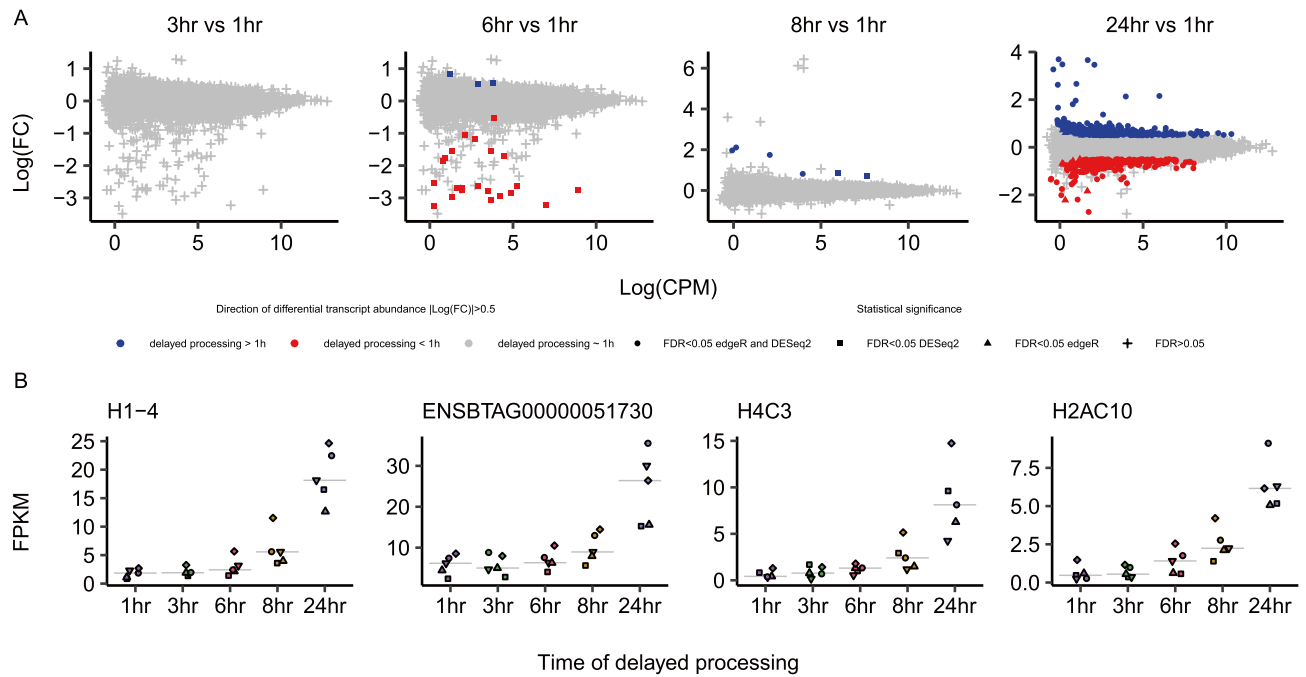


Figure 3. Comparison of transcript abundance in PWBCs from blood samples processed after prolonged storage at 4 °C. **(A)** M-A plots of the contrasts between each of the prolonged storage times versus samples processed within one hour. Gray shapes indicate genes whose transcript abundance are not significantly altered following the preservation of blood samples (FDR > 0.05). Squares indicate genes whose transcript abundance were significantly altered (FDR < 0.05) based on the DESeq2 algorithm. Triangles indicate genes whose transcript abundance were significantly altered (FDR < 0.05) based on the edgeR algorithm. Circles indicate genes whose transcript abundance were significantly altered (FDR < 0.05) based on both DESeq2 and edgeR algorithms. **(B)** Transcript abundance for genes with differential abundance in both the ‘8 h vs 1 h’ and ‘24 h vs 1 h’ contrasts.

can be indicative of RNA degradation²⁸. First, we inspected whether there was a global trend of transcripts to have prominent reduced coverage in one of the extremities (3’ or 5’). Coverage plots for the 224 genes with lower transcript abundance at 24 h of delayed processing showed a relative nucleotide sequencing depth similar to the 12,264 genes that were not differentially abundant (Fig. 4A, B).

We further tested whether the nucleotide coverage of the 224 genes with lower transcript abundance at 24 h of delayed processing was statistically different from distribution of the same genes observed at 1 h of processing (Fig. 4B). First, we calculated the Kolmogorov–Smirnov D-statistic²⁹ for the relative nucleotide coverage of genes from samples obtained from the jugular and coccygeal veins (both processed within 1 h of blood collection), which we referred to as ($D_{j,c,1hr}$). We also calculated the D-statistic for the relative nucleotide coverage of genes from samples obtained from the jugular vein processed at 24 h and 1 h post-sampling, which we referred to as ($D_{j,24h,1hr}$). Next, we calculated the difference between the two statistics ($\Delta D = (D_{j,24h,1hr}) - (D_{j,c,1hr})$). We reasoned that, for a given gene, ΔD would approximate to zero if the variation in the sequencing coverage was similar between the samples processed at different times (24 h vs 1 h) and the samples processed at the same time (1 h). Indeed, only seven out of the 224 genes (24 h < 1 h) had ΔD within the range of -0.25 and 0.25 (Fig. 4C, left plot). Furthermore, the range of ΔD calculated for the genes with lower transcript abundance at 24 h of storage (24 h < 1 h) was within the range of ΔD calculated for the genes with no transcript variation with the passing of 24 h post-collection (Fig. 4C, center plots). Altogether, these results provided strong evidence that the overall sequencing coverage of transcripts was similar between samples processed after 24 h storage at 4 °C and within one hour of sampling.

Gene ontology enrichment analysis of differentially expressed genes. Because we did not observe a systematic reduction in transcript coverage, we reasoned that the differential transcript abundance was a cellular regulatory response to the preservation of blood samples ex vivo. It was noteworthy that three out of four genes with greater transcript abundance at ‘8 h vs 1 h’ (*H1-4*, *H2AC10* and *H4C3*) were involved in chromatin configuration, specifically annotated with the gene ontology terms ‘nucleosome assembly’ (*H1-4* and *H4C3*), and ‘chromatin silencing’ (*H2AC10*).

Further interrogation of the 291 genes that had greater abundance at ‘24 h vs 1 h’ also revealed an enrichment of the category ‘nucleosome assembly’ with a series of histone related genes (*H1-4*, *H2BC12*, *H2BC13*, *H2BC14*, *H2BC18*, *H2BC4*, *H2BU1*, *H4C14*, *H4C3*, *H4C4* and *H4C8*, fold enrichment = 9.78, Fig. 5A, please see Supplementary Table S2 for a complete list of categories and gene annotation). All these genes overlapped with the molecular function ‘DNA binding’ (*BSX*, *DNMT3B*, *H1-4*, *H2AC21*, *H2AC6*, *H2AW*, *H2BC12*, *H2BC13*,

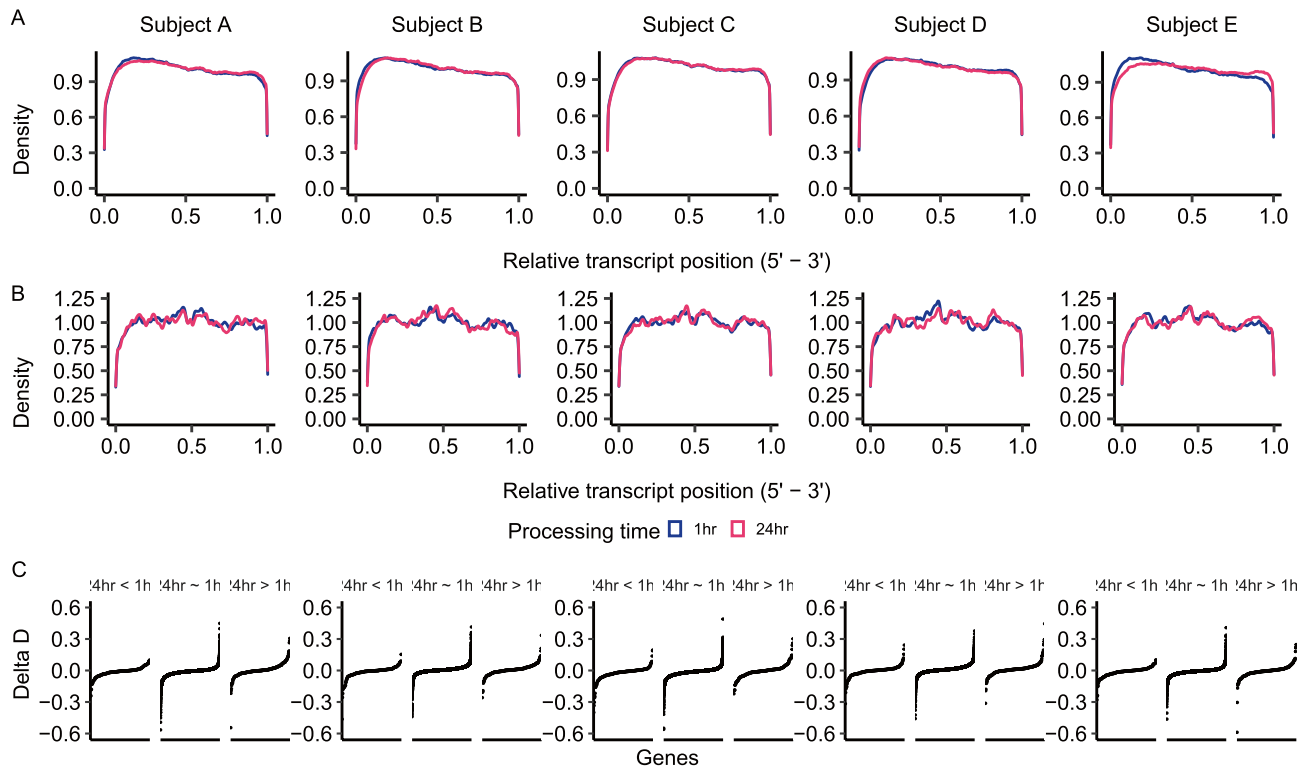


Figure 4. Comparison of overall transcript sequencing coverage from samples that were processed within 1 h and 24 h after blood collection. **(A)** Genes with no change in transcript abundance 24 h of storage at 4 °C. **(B)** Genes with lower transcript abundance after 24 h of storage at 4 °C. **(C)** Plots of the Delta D-statistic for all genes separated by their results of the contrast '24 h vs 1 h'.

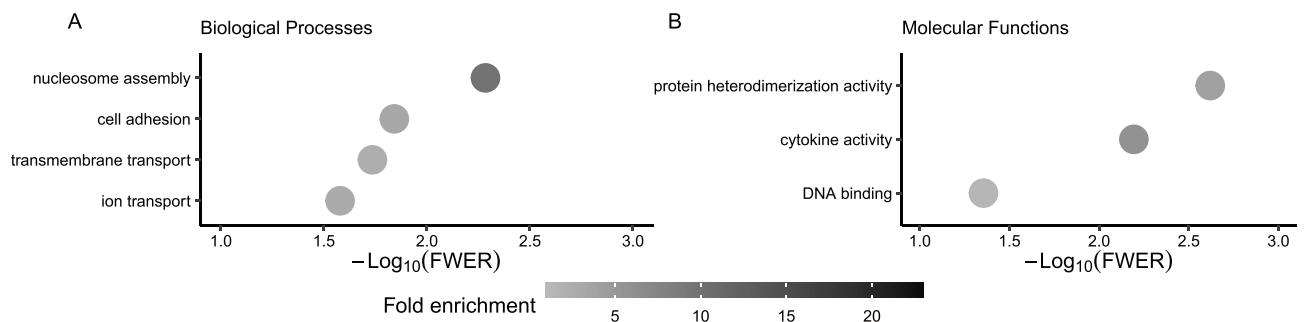


Figure 5. Gene ontology categories enriched in the genes with greater transcript abundance at 24 h versus one h. **(A)** Biological processes and **(B)** Molecular functions. To improve readability, only categories with more than five genes are displayed on the graphs, please, see Supplementary Table S2 and S3 for a full list of categories and the associated genes with their annotation.

H2BC14, H2BC18, H2BC4, H2BU1, H3C6, H4C14, H4C3, H4C4, H4C8, MSH5, PROX2, SNAPC4, TEAD3 and *TERT*, fold enrichment = 1, Fig. 5B, Supplementary Table S3). Other biological processes significantly enriched were 'neutrophil chemotaxis' (fold enrichment = 7.34), 'cell adhesion' (fold enrichment = 3.19), and 'transport membrane' (fold enrichment = 2.4) (Fig. 5A, Supplementary Table S2).

We also asked if there was enrichment of gene ontology categories within the 224 genes that had less transcript abundance after 24 h of storage at 4 °C, and there were several categories significantly enriched (FWER < 0.05, Fig. 6A, please see Supplementary Table S4 for a complete list of categories and gene annotation). Notably, there were a series of signaling related categories such as 'positive regulation of interferon-gamma production', 'positive regulation of interleukin-8 production', 'negative regulation of interferon-gamma production', 'positive regulation of ERK1 and ERK2 cascade', 'positive regulation of MAPK cascade', 'positive regulation of interleukin-1 beta production', 'positive regulation of interleukin-12 production', 'positive regulation of interleukin-6 production', 'positive regulation of NF-kappaB transcription factor activity', 'positive regulation of NIK/NF-kappaB signaling', 'positive regulation of peptidyl-tyrosine phosphorylation', and 'positive regulation of phosphatidylinositol 3-kinase signaling'.

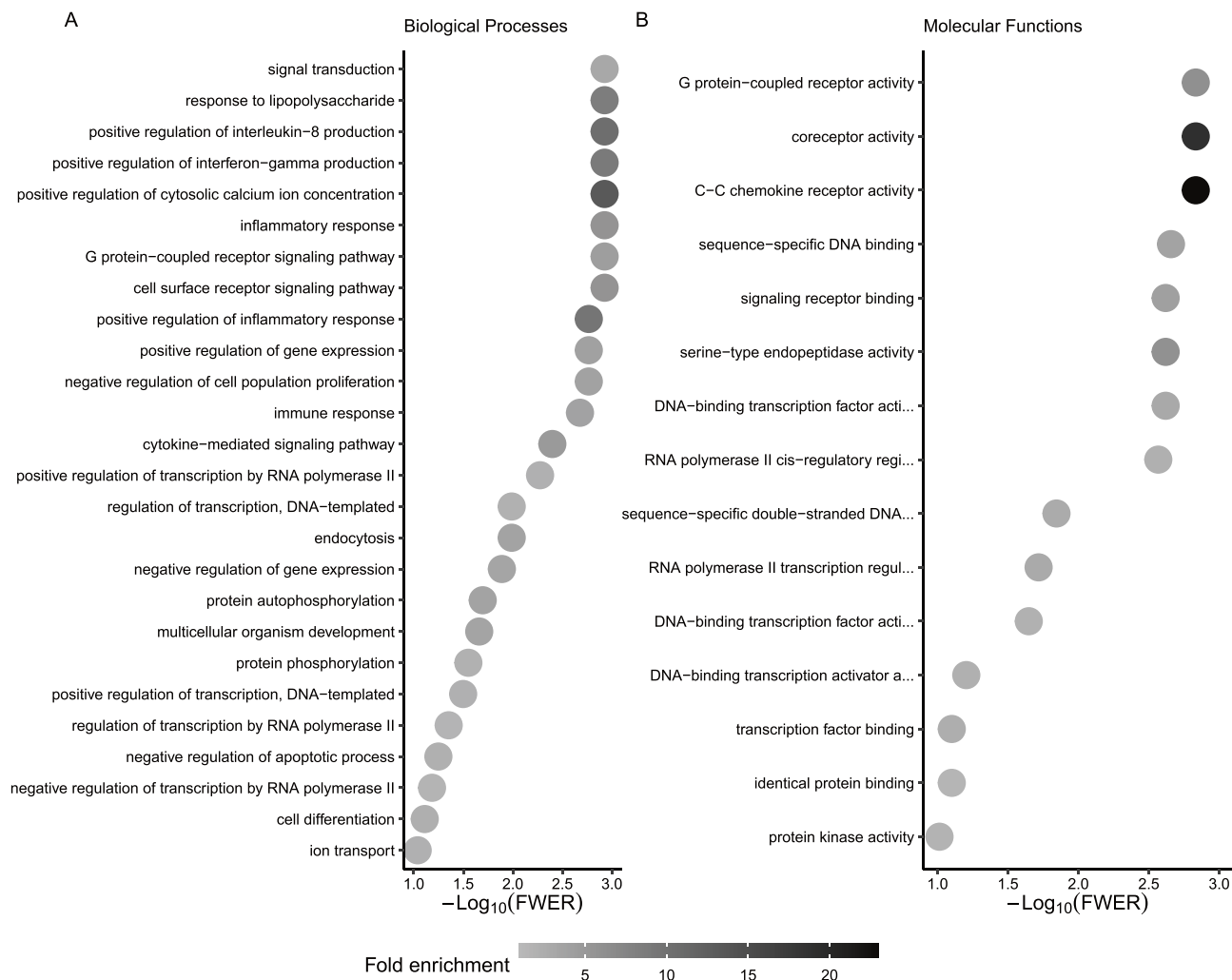


Figure 6. Gene ontology categories enriched in the genes with lower transcript abundance at 24 h versus one h. **(A)** Biological processes and **(B)** Molecular functions. To improve readability, only categories with more than five genes are displayed on the graphs, please see Supplementary Table S4 and S5 for a full list of categories and the associated genes with their annotation.

Also within the 224 genes that had less transcript abundance after 24 h of storage at 4 °C, there was significant enrichment of a series of categories involved in regulation of transcription and gene expression ($\text{FWER} < 0.05$, Fig. 6A, Supplementary Table S4), such as ‘positive regulation of transcription by RNA polymerase II’, ‘regulation of transcription, DNA-templated’, ‘regulation of transcription by RNA polymerase II’, ‘positive regulation of gene expression’, and ‘positive regulation of transcription, DNA-templated’.

In parallel with the identification of the significant enrichment of the categories involved in regulation of signaling and gene expression, the test for enrichment of molecular functions identified that many of those 224 genes were associated with functions that involve interaction with DNA to regulate gene expression, such as ‘RNA polymerase II cis-regulatory region sequence-specific DNA binding’, ‘DNA-binding transcription factor activity, RNA polymerase II-specific’, and ‘DNA-binding transcription factor activity’ (Fig. 6B, Supplementary Table S5).

Discussion

The main purpose of our study was to understand the dynamics of RNA degradation and the consequences of this RNA degradation on the quantification of transcript abundance in PWBCs from samples stored in the fridge (4 °C). We collected multiple samples from the same subject and proceeded with a strategic delay in the processing of samples, followed by immediate cryopreservation of PWBCs. Our methodical interrogation of the RNA quality and systematic analysis of transcriptome data lead us to identify critical factors related to the short-term preservation of blood samples for RNA analysis: (i) the vein used for sampling blood is not a source of significant and systematic changes in the transcriptome profiling of PWBCs; (ii) storing blood samples under refrigeration for 24 h does reduce their RIN values by approximately one unit, however the drop in RIN values does not interfere with the quantification of transcripts from protein-coding genes or long non-coding RNAs produced in PWBCs; (iii) even if blood samples are refrigerated, the abundance of gene transcripts produced in PWBCs starts to drop irregularly as early as three hours past blood sampling, but changes are consistent across

samples after eight hours of refrigeration; and (iv) the transcriptome of PWBCs is severely altered after blood samples are refrigerated for 24 h post-collection.

According to our hypothesis, we expected that storage of blood tubes at 4 °C for a long period of time would reduce the RNA quality through degradation. Indeed, there was a reduction in RIN values from RNA obtained from PWBCs after blood samples were preserved at 4 °C for 24 h (from 8.52 ± 0.37 at 1 h to 8 ± 0.37 at 24 h). The relatively high values of RIN after the preservation of blood samples at 4 °C for 24 h are similar to RIN values reported elsewhere²³. However, the results observed from the RNA-sequencing did not show indication of reduced RNA quality. The values of the 3'/5' bias for all libraries ranged from 0.47 to 0.56, with no effect of the processing time on the averages. In samples with degraded RNA, there is a bias towards transcript coverage on the 3' end, whereas samples with 3'/5' bias values close to 0.5 have balanced coverage of RNA extremities and are only observed in samples with high RNA quality³⁰. Thus, there was no systematic coverage bias towards the 3' end of polyadenylated transcripts in our samples.

Also based on our hypothesis, we anticipated that transcripts with significantly lower quantification would be a consequence of RNA degradation following a period of storage of blood samples at 4 °C. Here we reasoned that coverage plots for the 224 genes with lower abundance at 24 h in the contrast '24 h vs 1 h' would be distinct between the libraries produced from the 24 h group versus the 1 h group. Contrary to our expectation, the coverage charts (Fig. 4A) showed a virtually identical coverage of transcripts with significantly lower quantification whether on samples processed within one hour or 24 h of collection. Furthermore, a comparison of the distributions using the Kolmogorov–Smirnov test confirmed no significant changes in transcript coverage based on the amount of time that samples were preserved.

A possible explanation for the discrepancy between the significantly lower values of RIN for samples preserved for 24 h and the consistent sequencing coverage across transcripts is on the source of data. The RIN values are computed based on data collected from a series of features of an electropherogram, most of which involve information from ribosomal RNAs (5S, 5.8S, 18S and 28S)³¹. On the other hand, RNA-sequencing libraries were produced with enrichment of polyadenylated transcripts, and thus, the results from 3'/5' bias nor coverage plots do not account for ribosomal RNA. Our results indicate that, although a correlation between transcript coverage and RIN values have been identified^{28,31}, this relationship may be prominent in samples with RIN values less than 8.

Our results show that there is a prominent systematic alteration of transcript abundance in PWBCs when blood samples are preserved for 24 h at 4 °C, which is aligned with previous reports^{23,26}. Interestingly, we determined that this alteration of transcript abundance across individuals starts as early as eight hours post-collection.

Because we could not find indication that RNA degradation was a cause of these alterations, we reasoned that the alteration in transcript abundance was a consequence of the PWBCs responding to the cold temperature (4 °C) and lack of oxygen. The consequences of long-term exposure of mammalian cells at 4 °C have not been well studied, but Al-Fageeh and Smales³² proposed that the active transcription of a selected group of genes would cause a wide-spread reduction in transcription activity. Well-aligned with this possible mechanism, three out of four genes up regulated in PWBCs after the storage of blood samples for eight hours at 4 °C have a role in chromatin organization, including nucleosome assembly, which can be related to a compaction of the chromatin and reduction in transcriptional activity.

The alteration of transcript abundance in PWBCs after storing blood samples at 4 °C for 24 h has been observed before²⁶. However, our genome-wide transcriptome analysis shows that the changes are more prominent after 24 h of storage of blood samples at 4 °C. The genes with greater transcript abundance after 24 h of storage of blood samples at 4 °C relative to those processed within one hour of collection seem to be enriched for few biological processes and again with a high enrichment for genes involved in nucleosome assembly. It is possible that the cells increase the transcription of histone related genes to increase the genome-wide compaction of chromatin. The greater number of biological processes enriched for genes with lower transcript abundance after 24 h of storage of blood samples at 4 °C relative to those processed within one hour of collection corroborate the notion of a global silencing in transcription.

Considering the results of significant differential transcript abundance observed in the present study, we reasoned that the prolonged storage of blood samples at 4 °C would be relevant for investigations searching for mRNA markers in PWBCs. The overlap of our results with transcript abundance of genes also expressed in PWBCs and previously associated with fertility in heifers^{20,21} identified two genes (*NKG2A*, *PPP1R3B*, Supplementary Fig. S6) whose analysis of differential transcript abundance would have been compromised by storage of blood samples at 4 °C for eight hours or longer. These results strongly indicate that blood samples collected for studies of mRNA biomarkers should: (i) be preserved on ice as soon as they are collected and processed as early as possible, preferably within six hours of collection, for the proper cryopreservation of PWBCs, or (ii) if possible, collected in tubes that allow for the immediate preservation of RNA transcript abundance in the whole blood. However, we note that the chemical or cryopreservation of whole blood for RNA extraction requires further depletion of hemoglobin transcripts if the samples will be used for RNA-sequencing^{33,34}.

The transcriptome of PWBCs changes after blood sampling, even if the samples are refrigerated. A systemic alteration is detected at eight hours post blood collection and follows a pattern where PWBCs increase the transcription of genes related to chromatin compaction. This compaction is likely to reduce the transcription of several genes that function across multiple cellular processes in PWBCs. It is evident that this alteration in transcriptome profiles after prolonged storage can mask the transcriptome signature of a specific physiological phenomenon.

Our findings can be used as a guide for the establishment of protocols for blood processing when samples are supposed to be used for genome-wide quantification of transcripts in PWBCs. Blood samples collected for mRNA-based biomarker discovery should be refrigerated immediately and processed within six hours post-sampling. This recommendation can be considered by investigators working in diverse several areas of life sciences.

Methods

The reporting in this study follows the recommendations in the ARRIVE guidelines³⁵. Please see Supplementary table S6 for catalog number of kits and reagents used in this work.

Animal handling and sample collection. All animal handling and use was approved by the Institutional Animal Care and Use Committee (IACUC) at Virginia Tech. All procedures involving animal handling were performed in accordance with IACUC guidelines and regulations.

Eleven crossbred beef heifers (Angus x Simmental cross), averaging 14 months of age, located at Kentland Farm (Virginia Tech, Blacksburg, VA) were subjected to estrus synchronization. On day zero we administered an intramuscular injection of gonadotrophin-releasing hormone (GnRH, 100 µg; Factrel[®]; Zoetis Incorporated, Parsippany, NJ) and inserted a controlled internal drug release (CIDR, 1.38 g Progesterone; Eazi-Breed[™] CIDR[®]; Zoetis Inc.) device in each heifer. On day seven we removed the CIDR insert and administered an intramuscular injection of prostaglandin F2alpha (PGF2α, 25 µg; Lutalyse[®]; Zoetis Inc.), which was followed by a second injection of GnRH on day ten of the protocol. We used estrus synchronization to mitigate possible effects that the stages of the estrus cycle may have on gene expression³⁶.

We collected blood samples from heifers that expressed estrus ($n = 5$) at the time artificial insemination would normally be performed. Fifty mL of blood were sampled from the jugular vein and 10 mL from the coccygeal vein of each heifer using vacutainers containing 18 mg K2 EDTA (Becton, Dickinson, and Company, Franklin Lakes, NJ). Each tube was inverted several times to prevent blood coagulation and placed on ice immediately until processing.

Experimental design and blood processing. Blood tubes were sprayed thoroughly with a disinfectant (Lysol[®]) prior to storage. While on ice, tubes containing samples from the jugular vein were randomly assigned into five groups: 1 h, 3 h, 6 h, 8 h, and 24 h, which correspond to the time the samples remained at 4 °C prior to processing. We processed blood samples from the coccygeal vein in group 1 h for comparison of gene expression with samples from the jugular vein.

The buffy coat was separated from whole blood by centrifugation for 20 min at 2000xg at 4 °C. The buffy coat of each sample was aspirated and dispensed into 14 mL of a red blood cell lysis buffer solution (1.55 M ammonium chloride, 0.12 M sodium bicarbonate, 1 mM EDTA, Cold Spring Harbor Protocols). The mixture was gently mixed on a rocker for 10 min at room temperature, and then centrifuged for 10 min at 800xg at 4 °C. The supernatant was removed, and each sample was mixed with 200 µL of TRIzol[™] Reagent (Invitrogen[™], Thermo Fisher Scientific, Waltham, MA). The mixture of TRIzol[™] and PWBCs was transferred into cryotubes (Corning Incorporated, Corning, New York) and then snap frozen in liquid nitrogen prior to storage at -80 °C^{20,21}.

Total RNA extraction. Total RNA was extracted from the PWBCs using the acid guanidinium thiocyanate-phenol-chloroform procedure^{37,38}, with the aid of Phasemaker[™] tubes (Invitrogen[™], Thermo Fisher Scientific, Waltham, MA), following the manufacturer's instructions. Briefly, the samples were thawed on ice and 800 µL of TRIzol[™] was added to each. Once homogenized, the mixture was transferred into Phasemaker[™] tubes, where it was mixed with 200 µL of chloroform and centrifuged for 5 min at 12,000xg at 4 °C to complete phase separation. Next, the aqueous phase was collected into 1.7 mL microtubes and mixed with 0.5 µL of glycoblue. Then, 500 µL of 100% isopropanol was added to each tube and they were centrifuged for 10 min at 12,000xg at 4 °C to precipitate the RNA. The RNA pellet was collected and washed twice with 1 mL of 75% ethanol and centrifuged for 2 min at 7,500xg at 4 °C. Then, the RNA pellet was air-dried briefly and eluted in nuclease free water and maintained on ice for quantification and assessment of quality.

We quantified the total RNA concentration (A_{260}) and purity (A_{260}/A_{280} ratios) using a NanoDrop[™] 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). We also quantified the RNA using a Qubit RNA High Sensitivity Assay Kit (Invitrogen[™], Thermo Fisher Scientific, Waltham, MA) assayed on a Qubit 4 Fluorometer (Invitrogen[™], Thermo Fisher Scientific, Waltham, MA). Next, we evaluated the RNA integrity by assaying a sample on an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) using the Agilent RNA 6000 Pico Kit (Agilent, Santa Clara, CA).

Library preparation and high-throughput sequencing. We diluted the RNA samples to 1 ng/mL for library preparation and confirmed the concentration using a Qubit RNA High Sensitivity Assay Kit (Invitrogen[™], Thermo Fisher Scientific, Waltham, MA) and Qubit 4 Fluorometer (Invitrogen[™], Thermo Fisher Scientific, Waltham, MA). Five hundred ng were used as starting material for library preparation using the TruSeq[®] Stranded mRNA Library Prep (Illumina, Inc, San Diego, CA) and the IDT-ILMN TruSeq UD indexes. Sequencing was assayed in a NovaSeq 6000 sequencing platform (Illumina, Inc, San Diego, CA) using a NovaSeq 6000 SP Reagent Kit v1.5, to produce paired-end reads 150 nucleotides long. Preparation of libraries and sequencing assays was performed by staff at the Virginia Tech Genomics Sequencing Center.

Alignment of sequences and filtering. We removed the sequencing adapters using cutadapt (v. 2.8) and the sequences indicated by the manufacturer (Illumina, Inc, San Diego, CA). Next, we aligned the sequences to the cattle genome^{39,40} (Bos_taurus.ARS-UCD1.2.99) obtained from the Ensembl database⁴¹ using hisat2 (v. 2.2.0⁴²) with the -very-sensitive parameter. Using samtools (v. 1.10⁴³), we filtered the alignment to remove unmapped reads, secondary alignments, alignments whose reads failed quality control, and duplicates. We then utilized biobambam2 (v. 2.0.95⁴⁴) to mark and eliminate duplicates.

For the estimation of transcript coverage, we aligned the sequences trimmed from adapters to transcript sequences obtained from the Ensembl database⁴¹ with bowtie2 (v.2.4.2⁴⁵) using the `-very-sensitive-local` parameter.

Quantification of transcript abundance and gene filtering. We used featureCounts (subread v. 2.0.1⁴⁶) to count the fragments that matched to the Ensembl cattle annotation gene (Bos_taurus.ARS-UCD1.2.103). Genes annotated as protein coding, long non-coding RNA and pseudogene were retained. Following the calculation of counts per million (CPM) and reads per million per kilobase (FPKM) we retained genes that presented FPKM and CPM greater than one in five or more samples.

Quantification of library properties. We calculated the 3'/5' bias in our libraries using RNA-SeQC (v. 2.4.2³⁰), and the proportion of reads assigned to annotation by dividing the number of reads mapped to the Ensembl annotation divided by the total number of reads sequenced.

Statistical analyses. *RNA metrics (RIN, A_{280} , A_{260} and A_{280}/A_{260}) and number of genes detected per library.* We used paired Student's t ^{47,48} and Wilcoxon⁴⁹ tests to access the null hypothesis of no difference between two sampling locations ($H_0: \mu_{\text{jugular}} = \mu_{\text{coccygeal}}$). Within the samples obtained from the jugular vein, we used a generalized linear mixed model to access the null hypothesis of no difference between groups of delayed processing ($H_0: \mu_{(T_{1h})} = \mu_{(T_{2h})} = \dots = \mu_{(T_{24h})}$). The model included time of processing ($T_{(1h, 3h, 6h, 8h \text{ or } 24h)}$) as fixed effect and animal as random variable ($A_{(1,2,3,4 \text{ or } 5)}$). When the model indicated significance of the fixed effect ($P < 0.05$), we used the Z-test⁵⁰ and the Dunnett's approach⁵¹ for simultaneous tests for general linear hypothesis^{52,53} to compare the average of the groups $T_{(3h, 6h, 8h \text{ or } 24h)}$ with the baseline $T_{(1h)}$. Averages were inferred as statistically different when Bonferroni adjusted $P < 0.05$.

Library 3'/5' bias, proportion of reads assigned to annotation and genes detected. We used a generalized linear mixed model, with a binomial family and a logistic regression function to access the null hypothesis of no difference between groups of delayed processing ($H_0: \mu_{(T_{1h})} = \mu_{(T_{2h})} = \dots = \mu_{(T_{24h})}$). The time of processing was included in the model as fixed effect and animal was set as random effect. Averages were inferred as statistically different when $P < 0.05$.

Differential transcript abundance. We compared the transcript abundance from samples obtained from the jugular and coccygeal veins by using a paired-sample structure ($H_0: \mu_{\text{jugular}} = \mu_{\text{coccygeal}}$). Next, we compared the transcript abundance from samples obtained from the jugular vein that were processed at different times. The analyses were performed with the R packages 'edgeR'⁵⁴ using the quasi-likelihood F-test and 'DESeq2'⁵⁵ using the Wald's test. In the case of the delayed processing, we set up contrasts to compare the different processing times versus T_{1h} ($H_0: \mu_{(T_{1h})} = \mu_{(T_{2h})}; \dots; H_0: \mu_{(T_{1h})} = \mu_{(T_{24h})}$). We adjusted nominal P values for multiple hypothesis testing using the Benjamini–Hochberg false discovery rate⁵⁶. We assumed a difference in transcript abundance to be significant when $FDR < 0.05$ in the results obtained by both 'edgeR' and 'DESeq2' packages and absolute $\text{Log}_{(\text{fold-change})} > 0.5$. We utilized this approach to report robust results of differential transcript abundance independent of algorithm biases or limitations^{20,21,57,58}.

Gene ontology enrichment analysis. We tested lists of genes for enrichment of gene ontology terms using the R package 'Goseq'⁵⁹ and the genes retained after filtering as a background list^{60,61}. Nominal P values were adjusted for multiple hypothesis testing by family wise error rate^{62,63}.

Contrasts of transcript coverage. We quantified the relative position of each nucleotide in relation to the total number of nucleotides in the transcript, given in percentage. In addition, we calculated the relative proportion of occurrence of each nucleotide in relation to the total coverage of the gene. Then, for each gene in different groups, in a pair-wise manner, we compared the relative position of each nucleotide weighed by the relative coverage using the weighted Kolmogorov–Smirnov test, as described elsewhere²⁹.

Data availability

The raw data generated and analyzed during the current study are available in the GEO NCBI repository, under accession GSE192530 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192530>). To make our work fully reproducible the code utilized for the bioinformatics pipeline and analytical procedures is deposited as Supplementary Methods S1, in the figshare repository (<https://doi.org/10.6084/m9.figshare.17886068>)⁶⁴ and can also be accessible at https://biase-lab.github.io/rna_temporal_expression_PWBC/index.html⁶⁵.

Received: 28 January 2022; Accepted: 3 May 2022

Published online: 17 May 2022

References

- Liew, C. C., Ma, J., Tang, H. C., Zheng, R. & Dempsey, A. A. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J. Lab. Clin. Med.* **147**, 126–132. <https://doi.org/10.1016/j.lab.2005.10.005> (2006).
- Radich, J. P. *et al.* Individual-specific variation of gene expression in peripheral blood leukocytes. *Genomics* **83**, 980–988. <https://doi.org/10.1016/j.ygeno.2003.12.013> (2004).
- Whitney, A. R. *et al.* Individuality and variation in gene expression patterns in human blood. *P Natl. Acad. Sci. USA* **100**, 1896–1901. <https://doi.org/10.1073/pnas.252784499> (2003).

4. Garrett-Bakelman, F. E. *et al.* The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science* <https://doi.org/10.1126/science.aau8650> (2019).
5. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307. <https://doi.org/10.1016/j.cell.2012.02.009> (2012).
6. Piening, B. D. *et al.* Integrative personal omics profiles during periods of weight gain and loss. *Cell Syst.* **6**, 157–170. <https://doi.org/10.1016/j.cels.2017.12.013> (2018).
7. Mohr, S. & Liew, C. C. The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol. Med.* **13**, 422–432. <https://doi.org/10.1016/j.molmed.2007.08.003> (2007).
8. Sole, C., Arnaiz, E., Manterola, L., Otaegui, D. & Lawrie, C. H. The circulating transcriptome as a source of cancer liquid biopsy biomarkers. *Semin. Cancer Biol.* **58**, 100–108. <https://doi.org/10.1016/j.semcancer.2019.01.003> (2019).
9. Shen, J. *et al.* Plasma mRNA as liquid biopsy predicts chemo-sensitivity in advanced gastric cancer patients. *J. Cancer* **8**, 434–442. <https://doi.org/10.7150/jca.17369> (2017).
10. Chang, J. W. *et al.* Transcriptomic analysis in liquid biopsy Identifies circulating PCTAIRE-1 mRNA as a biomarker in NSCLC. *Cancer Genomics Proteomics* **17**, 91–100. <https://doi.org/10.21873/cgp.20170> (2020).
11. Wijerathne, H., Witek, M. A., Baird, A. E. & Soper, S. A. Liquid biopsy markers for stroke diagnosis. *Expert Rev. Mol. Diagn.* **20**, 771–788. <https://doi.org/10.1080/14737159.2020.1777859> (2020).
12. Modlin, I. M. *et al.* Early identification of residual disease after neuroendocrine tumor resection using a liquid biopsy multigenomic mRNA signature (NETest). *Ann. Surg. Oncol.* **28**, 7506–7517. <https://doi.org/10.1245/s10434-021-10021-1> (2021).
13. Boerrigter, E. *et al.* Liquid biopsy reveals KLK3 mRNA as a prognostic marker for progression free survival in patients with metastatic castration-resistant prostate cancer undergoing first-line abiraterone acetate and prednisone treatment. *Mol. Oncol.* **15**, 2453–2465. <https://doi.org/10.1002/1878-0261.12933> (2021).
14. Morimoto, O. *et al.* Association between recurrence of hepatocellular carcinoma and alpha-fetoprotein messenger RNA levels in peripheral blood. *Surg. Today* **35**, 1033–1041. <https://doi.org/10.1007/s00595-005-3077-5> (2005).
15. Jeng, K. S. *et al.* Prognostic significance of preoperative circulating vascular endothelial growth factor messenger RNA expression in resectable hepatocellular carcinoma: a prospective study. *World J. Gastroenterol.* **10**, 643–648. <https://doi.org/10.3748/wjg.v10.i5.643> (2004).
16. Andrade-Navarro, M. A. *et al.* RNA sequencing of human peripheral blood cells indicates upregulation of immune-related genes in huntington's disease. *Front. Neurol.* **11**, 573560. <https://doi.org/10.3389/fneur.2020.573560> (2020).
17. Guerranti, C. *et al.* Biomonitoring of chemicals in biota of two wetland protected areas exposed to different levels of environmental impact: results of the “PREVIENI” project. *Environ. Monit. Assess.* **189**, 456. <https://doi.org/10.1007/s10661-017-6165-2> (2017).
18. Caserta, D. *et al.* Correlation of endocrine disrupting chemicals serum levels and white blood cells gene expression of nuclear receptors in a population of infertile women. *Int. J. Endocrinol.* **2013**, 510703. <https://doi.org/10.1155/2013/510703> (2013).
19. La Rocca, C. *et al.* Exposure to endocrine disrupters and nuclear receptor gene expression in infertile and fertile women from different Italian areas. *Int. J. Environ. Res. Public Health* **11**, 10146–10164. <https://doi.org/10.3390/ijerph111010146> (2014).
20. Moore, S. E. *et al.* Rewiring of gene expression in circulating white blood cells is associated with pregnancy outcome in heifers (Bos taurus). *Sci. Rep.* **10**, 16786. <https://doi.org/10.1038/s41598-020-73694-w> (2020).
21. Dickinson, S. E. *et al.* Transcriptome profiles in peripheral white blood cells at the time of artificial insemination discriminate beef heifers with different fertility potential. *BMC Genomics* <https://doi.org/10.1186/s12864-018-4505-4> (2018).
22. Liu, X. *et al.* Comparison of six different pretreatment methods for blood RNA extraction. *Biopreserv. Biobank* **13**, 56–60. <https://doi.org/10.1089/bio.2014.0090> (2015).
23. Gautam, A. *et al.* Investigating gene expression profiles of whole blood and peripheral blood mononuclear cells using multiple collection and processing methods. *PLoS ONE* **14**, e0225137. <https://doi.org/10.1371/journal.pone.0225137> (2019).
24. Dickinson, S. E. *et al.* Evaluation of age, weaning weight, body condition score, and reproductive tract score in pre-selected beef heifers relative to reproductive potential. *J. Anim. Sci. Biotechnol.* **10**, 18. <https://doi.org/10.1186/s40104-019-0329-6> (2019).
25. Dickinson, S. E. & Biase, F. H. Transcriptome data of peripheral white blood cells from beef heifers collected at the time of artificial insemination. *Data Brief* **18**, 706–709. <https://doi.org/10.1016/j.dib.2018.03.062> (2018).
26. Malentacchi, F. *et al.* Effects of transport and storage conditions on gene expression in blood samples. *Biopreserv. Biobank* **14**, 122–128. <https://doi.org/10.1089/bio.2015.0037> (2016).
27. Gallego Romero, I., Pai, A. A., Tung, J. & Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* **12**, 42. <https://doi.org/10.1186/1741-7007-12-42> (2014).
28. Feng, H., Zhang, X. & Zhang, C. mRIN for direct assessment of genome-wide and gene-specific mRNA integrity from large-scale RNA-sequencing data. *Nat. Commun.* **6**, 7816. <https://doi.org/10.1038/ncomms8816> (2015).
29. Monahan, J. F. *Numerical Methods of Statistics*. (Cambridge University Press, 2011).
30. Graubert, A., Aguet, F., Ravi, A., Ardlie, K. G. & Getz, G. RNA-SeQC 2: Efficient RNA-seq quality control and quantification for large cohorts. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btab135> (2021).
31. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3. <https://doi.org/10.1186/1471-2199-7-3> (2006).
32. Al-Fageeh, M. B. & Smales, C. M. Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochem. J.* **397**, 247–259. <https://doi.org/10.1042/BJ20060166> (2006).
33. Jang, J. S. *et al.* Comparative evaluation for the globin gene depletion methods for mRNA sequencing using the whole blood-derived total RNAs. *BMC Genomics* **21**, 890. <https://doi.org/10.1186/s12864-020-07304-4> (2020).
34. Harrington, C. A. *et al.* RNA-Seq of human whole blood: evaluation of globin RNA depletion on Ribo-Zero library method. *Sci. Rep.* **10**, 6271. <https://doi.org/10.1038/s41598-020-62801-6> (2020).
35. du Sert, N. P. *et al.* Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. *PLoS Biol* **18**, e3000411. <https://doi.org/10.1371/journal.pbio.3000411> (2020).
36. Ioannidis, J. & Donadeu, F. X. Circulating microRNA Profiles during the Bovine Oestrous Cycle. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0158160> (2016).
37. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat. Protoc.* **1**, 581–585. <https://doi.org/10.1038/nprot.2006.83> (2006).
38. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159. <https://doi.org/10.1006/abio.1987.9999> (1987).
39. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* <https://doi.org/10.1093/gigascience/giaa021> (2020).
40. Elsik, C. G. *et al.* The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528. <https://doi.org/10.1126/science.1169588> (2009).
41. Flicek, P. *et al.* Ensembl 2014. *Nucl. Acids Res.* **42**, D749–755. <https://doi.org/10.1093/nar/gkt1196> (2014).
42. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915. <https://doi.org/10.1038/s41587-019-0201-4> (2019).
43. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).

44. Tischler, G. & Leonard, S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code Biol. Med.* <https://doi.org/10.1186/1751-0473-9-13> (2014).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
46. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2014).
47. Student. The probable error of a mean. *Biometrika* **6**, 1–25 (1908).
48. Kalpić, D., Hlupić, N. & Lovrić, M. in *International Encyclopedia of Statistical Science* (ed Miodrag Lovric) 1559–1563 (Springer Berlin Heidelberg, 2011).
49. Neuhäuser, M. in *International Encyclopedia of Statistical Science* (ed Miodrag Lovric) 1656–1658 (Springer Berlin Heidelberg, 2011).
50. Luke, S. G. Evaluating significance in linear mixed-effects models in R. *Behav Res Methods* **49**, 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y> (2017).
51. Tallarida, R. J. & Murray, R. B. in *Manual of Pharmacologic Calculations: With Computer Programs* 145–148 (Springer New York, 1987).
52. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom J* **50**, 346–363. <https://doi.org/10.1002/bimj.200810425> (2008).
53. Bretz, F., Hothorn, T. & Westfall, P. *Multiple Comparisons using R*. (Chapman and Hall/CRC, 2016).
54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. <https://doi.org/10.1093/bioinformatics/btp616> (2010).
55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/PREACCEPT-8897612761307401> (2014).
56. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Met.* **57**, 289–300 (1995).
57. Seyednasrollah, F., Laiho, A. & Elo, L. L. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform.* **16**, 59–70. <https://doi.org/10.1093/bib/bbt086> (2015).
58. Costa-Silva, J., Domingues, D. & Lopes, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **12**, e0190152. <https://doi.org/10.1371/journal.pone.0190152> (2017).
59. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14. <https://doi.org/10.1186/gb-2010-11-2-r14> (2010).
60. Timmons, J. A., Szkop, K. J. & Gallagher, I. J. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol.* **16**, 186. <https://doi.org/10.1186/s13059-015-0761-7> (2015).
61. Rhee, S. Y., Wood, V., Dolinski, K. & Draghici, S. Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* **9**, 509–515. <https://doi.org/10.1038/nrg2363> (2008).
62. Meijer, R. J. & Goeman, J. J. Multiple testing of gene sets from gene ontology: possibilities and pitfalls. *Brief Bioinform.* **17**, 808–818. <https://doi.org/10.1093/bib/bbv091> (2016).
63. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. Stat. Theory Appl.* **6**, 65–70 (1979).
64. Biase, F. H. & Wilson, C. Supplementary code and files to Delayed processing of blood samples impairs the accuracy of mRNA-based biomarkers. figshare. 01/06/2022.
65. Biase, F. H. & Wilson, C. *Supplementary Material S1 - code*, <https://biase-lab.github.io/rna_temporal_expression_PWBC/index.html> (2022).

Acknowledgements

We thank Chad Joines (Director of Beef Operations at Virginia Tech) and the staff from the Beef Cattle Center for the support with animal handling.

Author contributions

F.B. conceived, supervised, and obtained funding for the study. C.W. and F.B. processed the samples, analyzed the data, and wrote the paper. V.M. supervised the reproductive management and estrus synchronization of the heifers and sample collection. N.D. and S.P. contributed to the management and estrus synchronization of the heifers and sample collection. All authors read and approved the final manuscript.

Funding

This project was partially supported by Agriculture and Food Research Initiative Competitive Grant No. 2020-67015-31616 from the USDA National Institute of Food and Agriculture. The funding agency had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12178-5>.

Correspondence and requests for materials should be addressed to F.H.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022