



Plasma-Free Blood as a Potential Alternative to Whole Blood for Transcriptomic Analysis

Qingwang Chen¹ · Xiaorou Guo¹ · Haiyan Wang¹ · Shanyue Sun² · He Jiang¹ · Peipei Zhang¹ · Erfei Shang¹ · Ruolan Zhang¹ · Zehui Cao¹ · Quanne Niu¹ · Chao Zhang¹ · Yaqing Liu¹ · Leming Shi^{1,3} · Ying Yu¹ · Wanwan Hou¹ · Yuanting Zheng¹

Received: 27 April 2023 / Revised: 29 June 2023 / Accepted: 13 July 2023 / Published online: 13 September 2023
© The Author(s) 2023

Abstract

RNA sequencing (RNAseq) technology has become increasingly important in precision medicine and clinical diagnostics, and emerged as a powerful tool for identifying protein-coding genes, performing differential gene analysis, and inferring immune cell composition. Human peripheral blood samples are widely used for RNAseq, providing valuable insights into individual biomolecular information. Blood samples can be classified as whole blood (WB), plasma, serum, and remaining sediment samples, including plasma-free blood (PFB) and serum-free blood (SFB) samples that are generally considered less useful byproducts during the processes of plasma and serum separation, respectively. However, the feasibility of using PFB and SFB samples for transcriptome analysis remains unclear. In this study, we aimed to assess the suitability of employing PFB or SFB samples as an alternative RNA source in transcriptomic analysis. We performed a comparative analysis of WB, PFB, and SFB samples for different applications. Our results revealed that PFB samples exhibit greater similarity to WB samples than SFB samples in terms of protein-coding gene expression patterns, detection of differentially expressed genes, and immunological characterizations, suggesting that PFB can serve as a viable alternative to WB for transcriptomic analysis. Our study contributes to the optimization of blood sample utilization and the advancement of precision medicine research.

Keywords RNA sequencing · Whole blood · Plasma-free blood · Serum-free blood · Transcriptomic analysis

Abbreviations

RNAseq	RNA sequencing
WB	Whole blood
PFB	Plasma-free blood

Qingwang Chen, Xiaorou Guo, and Haiyan Wang have contributed equally to this work.

- ✉ Ying Yu
ying_yu@fudan.edu.cn
- ✉ Wanwan Hou
houwanwan2008@163.com
- ✉ Yuanting Zheng
zhengyuanting@fudan.edu.cn

¹ State Key Laboratory of Genetic Engineering, School of Life Sciences and Human Phenome Institute, Shanghai Cancer Center, Fudan University, Shanghai 200438, China

² Shandong Provincial Hospital, Shandong First Medical University, Jinan 250021, China

³ The International Human Phenome Institutes, Shanghai 200438, China

SFB	Serum-free blood
PBMC	Peripheral blood mononuclear cell
DEG	Differential expressed gene
EDTA	Ethylenediaminetetraacetic acid
H0	Holding 0.5 h
H6	Holding 6 h
RIN	RNA integrity number
FPKM	Fragments per kilobase of exon model per million mapped fragments
TPM	Transcripts per kilobase of exon model per million mapped reads
PVCA	Principal variance component analysis
CV	Coefficient of variation
PCA	Principal component analysis
HCA	Hierarchical clustering analysis
RPS4Y1	Ribosomal protein S4 Y-linked 1
DDX3Y	DEAD-box helicase 3 Y-linked
EIF1AY	Eukaryotic translation initiation factor 1A Y-linked
KDM5D	Lysine demethylase 5D
TXLNGY	Taxilin gamma pseudogene, Y-linked

XIST	X inactive specific transcript
TSIX	TSIX transcript, XIST antisense RNA
SD	Standard deviation
PCC	Pearson correlation coefficient
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
CIBERSORT	Cell-type identification by estimating relative subsets of RNA transcripts
PC	Principal component
NK	Nature killer
HPA	Human protein atlas

Introduction

RNA sequencing (RNAseq) technology (Wang et al. 2009; Stark et al. 2019) for transcriptomic analysis has become increasingly important in the fields of precision medicine (De Ruysscher et al. 2017) and clinical diagnostics (Ruan et al. 2021). RNAseq has also emerged as a powerful tool for identifying protein-coding genes (Wang et al. 2009; Hong et al. 2019), performing differential expression gene (DEG) analysis for signature discovery (Soneson et al. 2015), and inferring immune cell composition (Newman et al. 2015). Human peripheral blood samples are widely used for RNAseq due to their availability and convenience of collection (Mohr and Liew 2007; Qi et al. 2023), offering valuable insights into the biomolecular information of the individual. As the transcriptome of blood samples has been intensively studied (Hong et al. 2019), it has been shown that it can help improve the diagnosis of patients with rare diseases and identify new disease-related genes (Fresard et al. 2019). The analysis of blood transcriptome can provide prospective scientific guidance for disease prevention, diagnosis (de Almeida Chuffa et al. 2022), treatment (He et al. 2019b), and prognosis (Krishnan and Thomas 2022).

Blood samples can be classified into whole blood (WB), plasma, serum, blood cells, and other remaining sediment samples according to post-collection processing (Bayot and Tadi 2022). Plasma samples are obtained by centrifugation of WB samples with anticoagulants, while serum samples are obtained by centrifugation of WB samples without anticoagulants or with procoagulants (Sotelo-Orozco et al. 2021). WB, plasma, and serum samples are commonly used for transcriptomic studies due to their sample availability and convenience of collection under ideal conditions (Qin et al. 2016; Mjelle et al. 2019; Husseini et al. 2022). However, the remaining sediment samples after plasma or serum separation are often discarded or stored for biological resource preservation in real-world scenarios. These include plasma-free blood (PFB) and serum-free blood (SFB) samples, which also contain abundant RNAs that reflect individual information. WB

samples contain various immune cell types, such as peripheral blood mononuclear cells (PBMCs), polymorphonuclear cells (granulocytes), and other anucleate cells (e.g., erythrocytes and platelets) according to the number of cell nuclei (Uhlen et al. 2019). PFB samples are like WB samples but without plasma. Moreover, the cells in SFB samples are all enclosed in a blood clot, which may lead to differences in their expression when compared to WB and PFB samples.

For retrospective cohort studies that rely on biobank WB samples from authoritative institutions and hospitals for transcriptomic research, PFB and SFB samples may serve as alternative sources of RNA for sequencing when WB samples are insufficient. Nevertheless, the feasibility of using PFB and SFB samples for transcriptome analysis remains unclear. Moreover, as multi-omics studies such as genomics, transcriptomics, proteomics, metabolomics, and microbiomics require adequate blood samples (Wu et al. 2021a), it is essential to maximize the utilization of blood samples including PFB and SFB for identifying biomarkers in the blood for achieving personalized treatment (De Ruysscher et al. 2017; Kamali et al. 2022).

Given these concerns, our study aimed to assess the suitability of employing either PFB or SFB samples as alternative RNA sources in transcriptomic analysis. Previous researches have demonstrated the influence of pre-analytical factors on RNA quality and gene expression (Debey-Pascher et al. 2011; Mastrokolas et al. 2012; Dvinge et al. 2014; Shin et al. 2014; Zhao et al. 2014; Huang et al. 2017; Reust et al. 2018; Shen et al. 2018; Donohue et al. 2019; Gautam et al. 2019; He et al. 2019a; Harrington et al. 2020; Xing et al. 2021; Chebbo et al. 2022; Husseini et al. 2022). However, a comprehensive understanding of the similarities and differences among WB, PFB, and SFB samples in transcriptomic profiles remains elusive, making it crucial to evaluate their suitability for various applications using reliable performance metrics.

In the comparative study, we analyzed the transcriptomic profiles of WB, PFB, and SFB samples collected from two healthy donors at two different holding times. We assessed expression patterns, differentially expressed genes (DEGs), and immunological characterizations for each blood type. Our results indicate that PFB samples may be used as a potential alternative to WB samples for transcriptomic analysis, because they exhibit expression patterns that closely resemble to those of WB samples based on various performance metrics. This would provide a valuable resource for maximizing blood sample utilization in research and clinical applications. Further research is needed to confirm the suitability of PFB and SFB samples for various specific applications.

Methods

Blood Sample Collection and Preparation

Informed consents were obtained from two healthy donors (P10: male; P11: female) by following the study protocol approved by the Ethics Committee of Fudan University. We collected 5 mL of blood from each donor using anti-coagulation tubes with K2-Ethylenediaminetetraacetic acid (EDTA) (Shanghai Aoxiang Medical Technology Co., Ltd.) and procoagulation tubes (Jiangsu Yuli Medical Instruments Co., Ltd.). One group of blood samples was processed after holding 0.5 h (H0) and another group was processed after holding 6 h (H6) at room temperature to simulate the real turnaround times. PFB samples were obtained by centrifuging blood samples collected in anticoagulation tubes and separating the plasma supernatants, while SFB samples were obtained by centrifuging blood samples collected in pro-coagulation tubes and separating the serum supernatants. For each donor, three tubes with 150 μ L WB samples, three tubes with 100 μ L PFB samples, and one tube with 1.20 mL SFB samples were used for RNA extraction, and the remaining samples were snap frozen in liquid nitrogen and stored at -80°C in the refrigerator. The separated plasma and serum samples in this study were used in another parallel study for miRNA analysis using small RNA sequencing.

RNA Extraction and cDNA Library Construction

RNA samples were extracted by QIAzol lysis reagents (QIAGEN) with QIAcube Connect (QIAGEN) according to the manufacturer's manual. RNA concentration and integrity were measured by Qubit 3.0 Fluorometer (Thermo Fisher Scientific) and Agilent 4200 TapeStation (Agilent Technologies (China) Co., Ltd.), respectively. We normalized RNA concentrations according to the volume ratio of PFB vs WB samples and SFB vs WB samples to make them comparable among different types of blood samples. RNA samples with a total RNA amount ≥ 100 ng and RNA integrity number (RIN) score ≥ 3 were qualified to construct Ribozero RNAseq libraries for sequencing (Li et al. 2014).

Globin and ribosomal RNAs were depleted using Ribo-off Globin & rRNA Depletion Kit (H/M/R) (Vazyme #N408), and cDNA libraries were constructed using VAHTS® Universal V8 RNAseq Library Prep Kit for Illumina (Vazyme #NR605). The concentrations of cDNA libraries were measured using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific). The distribution of cDNA fragment size was measured by Qsep 100 Advance (BiOptic).

RNA Sequencing and Data Preprocessing

cDNA libraries were sequenced on the NovaSeq 6000 sequencing platform (Illumina) to generate paired-end reads (150 bp). Then we followed a previously published protocol (Chen et al. 2022) for processing and quality control of raw FASTQ reads. Briefly, adapter sequences were trimmed using fastp v0.19.6 (Chen et al. 2018) and read quality was assessed, while after trimming using FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Potential contamination from other species or junction primers was detected by extracting the first 10,000 reads from the clean FASTQ files using FastQ Screen v0.12.0 (Wingett and Andrews 2018). The mapping quality of 10% bam file per sample was calculated using Qualimap v2.0.0 (Okonechnikov et al. 2016) for efficiency and cost-effectiveness purposes. Reads were aligned to GRCh38_snp_tran genome reference and quantified using HISAT v2.1 (Kim et al. 2019), SAMtools v1.3.1 (Li et al. 2009), and StringTie v1.3.4 (Pertea et al. 2015) pipeline (Pertea et al. 2016) with Ensembl gene models (version: Homo_sapiens.GRCh38.93.gtf).

Gene expression was normalized by fragments per kilobase of exon model per million mapped fragments (FPKM) and transcripts per kilobase of exon model per million mapped reads (TPM) to remove the effect of gene length and library size (Zhao et al. 2021). The counts were used for correlation analysis and to define detected genes for the Jaccard index (Mukherjee et al. 2021) comparison, principal variance component analysis (PVCA), as well as DEG analysis. FPKM was used for coefficient of variation (CV) calculation, sex check, principal component analysis (PCA), and hierarchical clustering analysis (HCA). TPM values were deconvoluted to infer the proportion of immune cell expression. Only protein-coding genes were used for CV calculation, PCA, and correlation analysis. To avoid infinite values, a minimum value of 0.01 was added to the FPKM value of each gene before the log₂ transformation.

Sex-Specific Gene Expression and Sex Check

Sex check was performed using sex-specific genes identified from previous studies (Chen et al. 2022), including five male-specific genes (*RPS4Y1: Ribosomal Protein S4 Y-Linked 1*, *DDX3Y: DEAD-Box Helicase 3 Y-Linked*, *EIF1AY: Eukaryotic Translation Initiation Factor 1A Y-Linked*, *KDM5D: Lysine Demethylase 5D*, *TXLNGY: Taxilin Gamma Pseudogene, Y-Linked*) and two female-specific genes (*XIST: X Inactive Specific Transcript*, *TSIX: TSIX Transcript*, *XIST Antisense RNA*). A distance matrix using the Euclidean method was calculated to measure the distance of the samples and genes, whereas Ward linkage was used for HCA using the R package pheatmap v1.0.12.

Human Protein-Coding Gene Detection in Different Blood Types

Similarities in protein-coding gene expression of all samples were compared using PCA. PCA was conducted with the univariance scaling, using the `prcomp` function of R package `stats` v4.2.1. Moreover, the clustering of protein-coding genes in the top 1000 standard deviation (SD) ranking across all samples was compared using HCA. HCA was performed using the R package `pheatmap` v1.0.12 with clustering based on a distance matrix calculated using the correlation distance metric. PVCA was performed using the R packages `pvca` v1.36.0 and `Biobase` v2.56.0 to assess the impact of different factors on the SD TOP 1000 protein-coding gene expression data. Pearson correlation coefficient (PCC) was calculated for the intra-group and inter-group to show correlations between technical replicates vs correlations between different blood types. The detected genes of the individual sample were defined as the genes with \geq three counts in \geq two technical replicates of three. The similarity of detected protein-coding gene sets between different groups was evaluated using the Jaccard index (Mukherjee et al. 2021). Finally, the detected protein-coding genes were filtered and compared between blood types with the same donor and the same holding time using R package `eulerr` v6.1.1 to produce Venn diagrams.

Differential Expression Analysis and Enrichment Analysis

The R package `limma` v3.52.4 (Ritchie et al. 2015) was used to identify DEGs with a $p < 0.05$ and $| \log_{2}FC | \geq 1$. Venn diagrams created by the R package `eulerr` v6.1.1 were used to compare and visualize the results. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed using the R package `ClusterProfiler` 4.0 v4.4.4 (Roncaglia et al. 2017; Wu et al. 2021b) and `KOBAS-I` (<http://kobas.cbi.pku.edu.cn/genelist/>) (Bu et al. 2021) to assess the enrichment of DEGs in each group based on blood types or donors as variables.

Immune Cell Cluster Characterization and Cell Abundance

The cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) algorithm (Newman et al. 2015) was used to estimate the cell-type composition of each sample using the R package `IOBR` v0.99.9 (Zeng et al. 2021), which translated the TPM-normalized gene expression matrix of different blood types into the relative proportion of immune cells. Euclidean distance was performed to assess the correlations between immune cell subsets using R package `pheatmap` v1.0.12.

Immune Cell-Specific Gene Expression

The top 60 genes specifically expressed in six major cell types of human blood (granulocytes, monocytes, dendritic cells, nature killer (NK) cells, B cells, and T cells) were collected from the "IMMUNE CELL" section of the website for Human Protein Atlas (HPA) (Uhlen et al. 2019), which can be accessed at <https://www.proteinatlas.org/humanprotein/immune+cell>. TPM values of these genes were used for cluster and comparative analysis in different blood types.

Statistical Analysis

All experiments were performed with three replicates. Statistical analysis and graphical work were performed using R v4.2.1 (<https://cran.r-project.org/>, R development core team) and a suite of R packages. Student's *t* test and Wilcoxon test were used to compare RNAseq quality control metrics and gene expression profiles between groups with different blood holding times, donors, and blood types. All statistical tests were two tailed, and a *p* value less than 0.05 was considered statistically significant.

Results

Study Overview

In this study, we performed a comparative transcriptomic analysis of three blood types, including WB, PFB, and SFB, collected from two healthy donors at two different holding times (Fig. 1a, top). WB samples consist of diverse immune cell types, including PBMCs, polymorphonuclear cells (granulocytes), as well as anucleate cells (e.g., erythrocytes and platelets) based on their number of cell nuclei. PFB samples are similar to WB samples, while the cells in SFB samples are all enclosed in a blood clot (Fig. 1a, bottom). The RNA samples were subjected to library preparation and sequencing (Fig. 1b). The sequencing data were aligned to the hg38 reference genome, and transcript abundance was estimated for each sample. The quality check was performed using various metrics, including mapping ratio, genome region, CV, and sex check. To compare expression profiles across blood types, we employed various analytical methods such as PCA, HCA, PVCA, PCC, and the Jaccard index. Additionally, we identified DEGs and compared immunological characterizations across the three blood types using different groups (Fig. 1c).

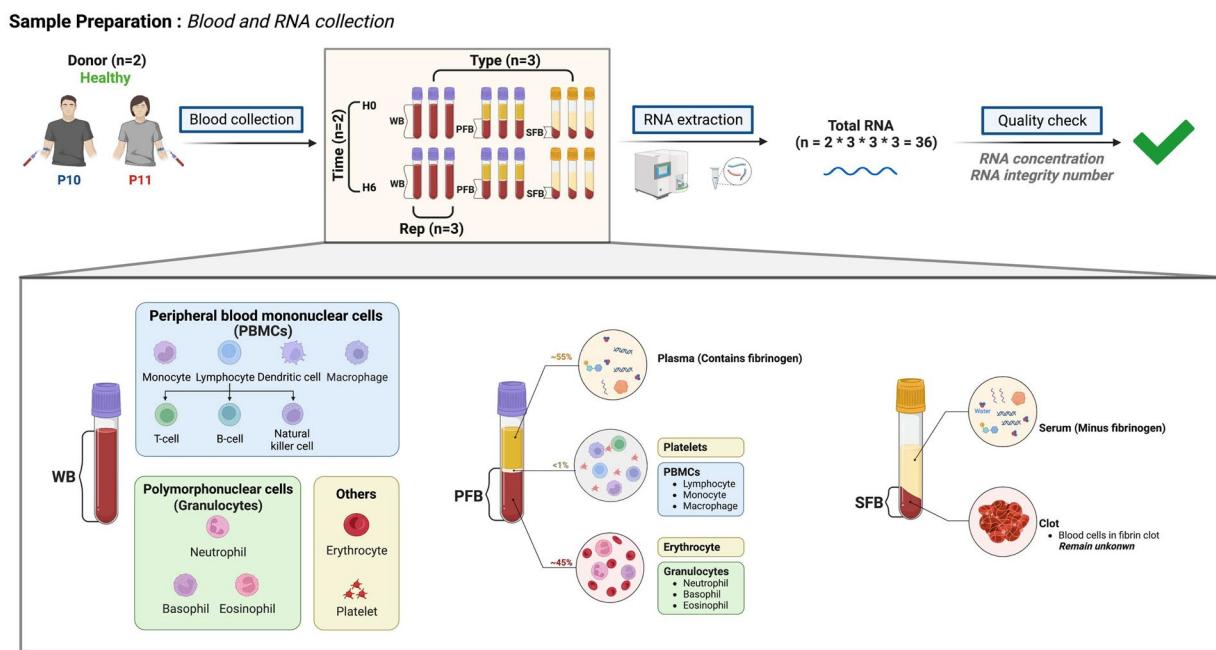
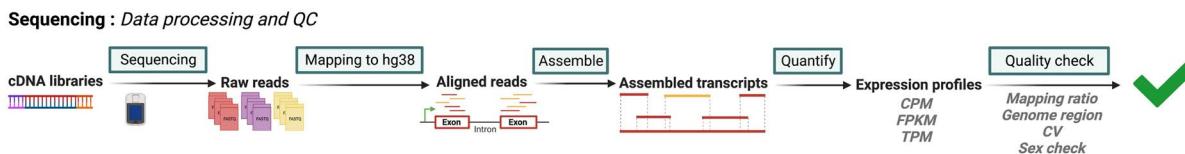
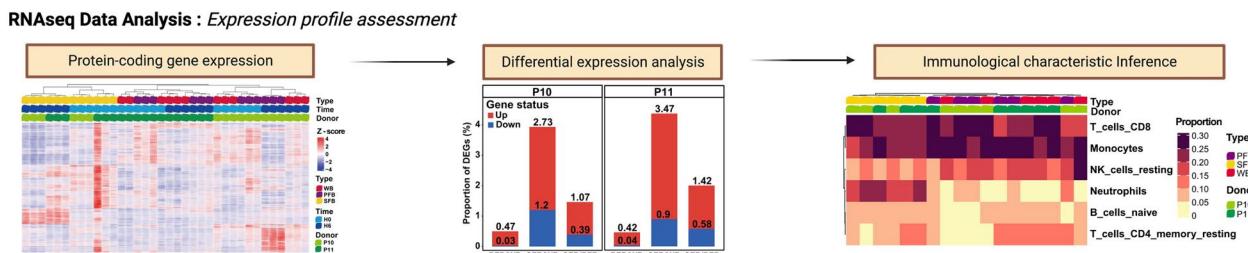
a**b****c**

Fig. 1 Overview of study design and analysis. **a** Top: Sample collection and processing. Whole blood (WB), plasma-free blood (PFB), and serum-free blood (SFB) samples were collected from two healthy donors (P10 and P11) through two holding times (H0 and H6). (P10: male donor, P11: female donor. H0: processed after holding 0.5 h, H6: processed after holding 6 h.) Bottom: Diagram of estimated immune cell composition in three blood sample types. Total RNA was extracted from each blood sample and assessed for quality, including RNA concentration and RNA integrity number (RIN). **b** RNA sequencing and quality control. The RNA samples were sub-

jected to library preparation and sequencing. The sequencing data were aligned to the hg38 reference genome. Transcript abundance was estimated for each sample, and the quality check was performed using various metrics, including mapping ratio, genome region, coefficient of variation, and sex check. **c** Bioinformatic analysis of RNAseq data. The expression of protein-coding genes across all samples was shown by HCA. Differential expression gene analysis was conducted between different blood types and donors. Immune cell composition was inferred for each blood type using gene expression data

High-Quality RNA and Sequencing Data Obtained from Each Blood Sample

We assessed RNA quality by measuring RNA concentration and RIN score for each sample. The minimum RNA

concentration required for cDNA library construction was 1 ng/ μ L (Li et al. 2014). All samples had RNA concentrations above this threshold with a median value of 34.50 ng/ μ L (Fig. 2a, left; Table 1). Furthermore, there was no significant difference in normalized RNA concentrations

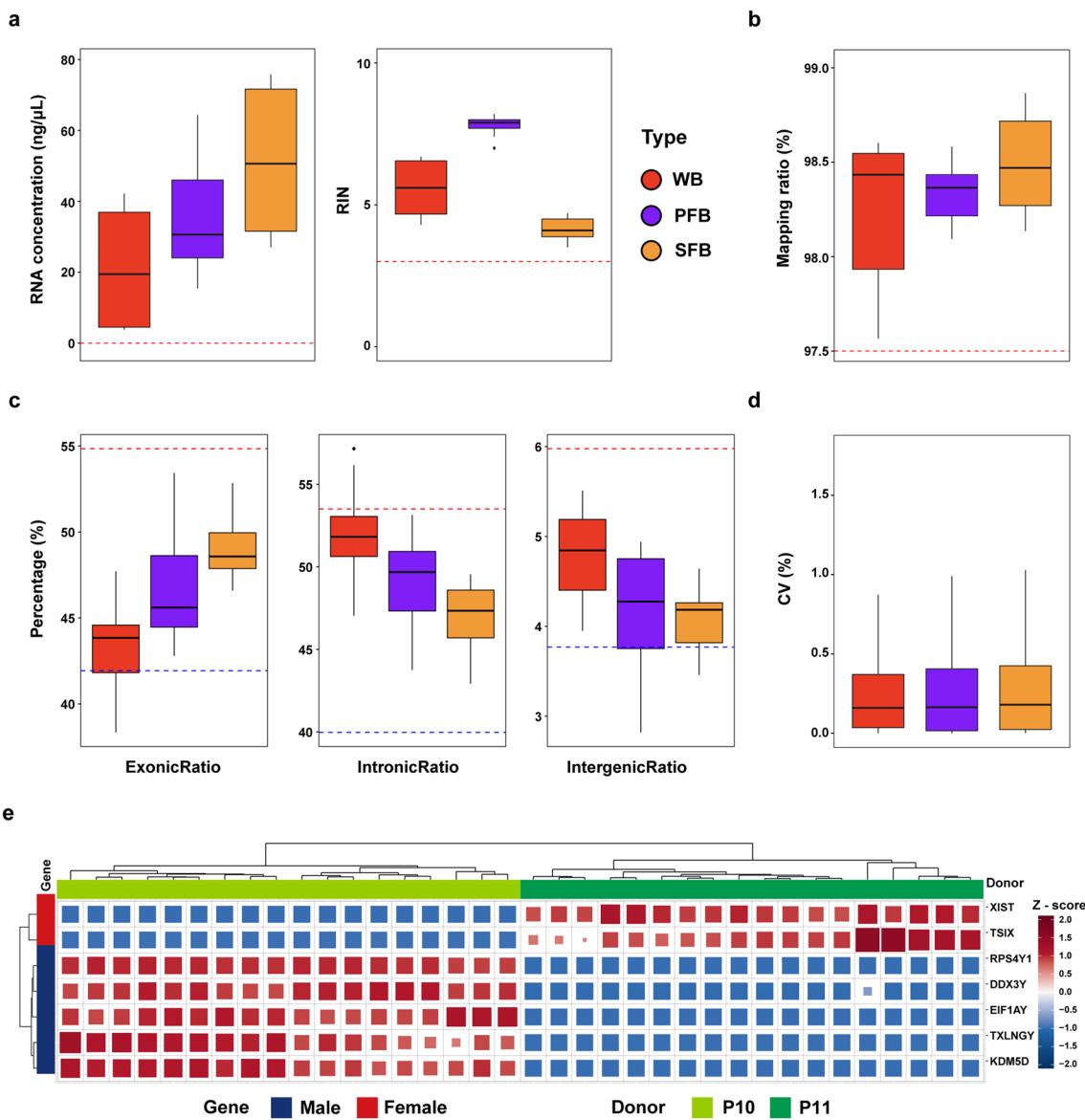


Fig. 2 Quality assessment of RNAseq data. **a** RNA concentrations (ng/μL) and RIN values in three types of blood samples (left red line value: 1; right red line value: 3). **b** The mapping ratio (%) in three types of blood samples (The red line represents the median value (97.51%) corresponding to the RiboZero library mapping ratio in the Quartet dataset). **c** The distribution of reads across the human genome. The dotted line represents the mean \pm SD values of the corresponding gene regions of the RiboZero library in the Quartet RNA reference dataset. Red indicates mean \pm SD value (ExonicRatio: 54.85, IntronicRatio: 53.51, and IntergenicRatio: 5.98), and blue indicates mean $-$ SD value (ExonicRatio: 41.93, IntronicRatio: 39.97, and IntergenicRatio: 3.77). **d** The coefficient of variation (CV) values of protein-coding gene expression in three blood types **e** Heatmap and hierarchical clustering of the expression levels of seven sex-specific genes in all samples, including five male-specific genes (*RPS4Y1*, *DDX3Y*, *EIF1AY*, *KDM5D*, and *TXLNGY*) and two female-specific genes (*XIST* and *TSIX*). Blue indicates male-specific genes, and red indicates female-specific genes. Lime green indicates P10, and dark green indicates P11

cates mean $-$ SD value (ExonicRatio: 41.93, IntronicRatio: 39.97, and IntergenicRatio: 3.77). **d** The coefficient of variation (CV) values of protein-coding gene expression in three blood types **e** Heatmap and hierarchical clustering of the expression levels of seven sex-specific genes in all samples, including five male-specific genes (*RPS4Y1*, *DDX3Y*, *EIF1AY*, *KDM5D*, and *TXLNGY*) and two female-specific genes (*XIST* and *TSIX*). Blue indicates male-specific genes, and red indicates female-specific genes. Lime green indicates P10, and dark green indicates P11

between the WB and SFB samples (Fig. S1). Similarly, the minimum RIN score required for sequencing was three (Li et al. 2014). All samples had RIN scores above this threshold, with a median value of 5.6 (Fig. 2a, right; Table 1). These results indicated that all samples were suitable for RiboZero library construction and sequencing.

The average mapping ratio of all samples was 98.36% (Fig. 2b), all above the median mapping ratio of 97.51% for the RiboZero libraries in the Quartet RNA reference datasets (Yu et al. 2023), which suggested the successful alignment of the reads to the reference genome. The gene region distribution was consistent with the expected characteristics of RiboZero libraries in the Quartet RNA

Table 1 RNA quantity and quality

Donor	Time	Type	Conc	RIN
P10	H0	WB	35.30±5.13	4.50±0.17
P10	H0	PFB	60.27±6.47	7.27±0.23
P10	H0	SFB	74.87±2.84	5.82±1.09
P10	H6	WB	40.47±3.04	5.86±0.47
P10	H6	PFB	39.07±5.14	7.97±0.12
P10	H6	SFB	69.13±6.60	5.89±1.12
P11	H0	WB	6.00±1.92	5.57±1.03
P11	H0	PFB	21.93±5.40	7.87±0.06
P11	H0	SFB	31.33±2.89	5.85±1.02
P11	H6	WB	7.19±3.42	6.60±0.10
P11	H6	PFB	23.00±5.55	8.07±0.12
P11	H6	SFB	34.07±5.75	4.66±1.76

Conc. RNA concentration (ng/µL), RIN RNA integrity number

reference datasets (Yu et al. 2023) across the three blood types (Fig. 2c). The median CV values of the sample replicates for the three blood types were all less than 0.17%, indicating good reproducibility of expression data (Fig. 2d). Additionally, a sex check was performed to validate the gender of the sample donors. Specifically, seven sex-specific genes were analyzed, including five male-specific genes and two female-specific genes. The results of the sex check were found to be consistent with the expected gender of each donor (Fig. 2e), thus affirming the quality of our expression data and providing confidence in the subsequent analysis.

Moreover, we performed additional quality control analysis on various metrics of the raw and mapped data (Fig. S2). Our findings revealed that: (1) high Phred quality scores were observed across all bases and positions in the FASTQ files (Fig. S2a); (2) no contamination from extraneous sources was detected in any of the samples (Fig. S2b); (3) the GC content exhibited a normal distribution in all samples (Fig. S2c). Details of the quality control metrics are provided in Table S1. Overall, high-quality RNA and sequencing data could be obtained from each blood type, by employing rigorous quality control measures. These results established a strong foundation for further investigation into comparative gene expression profiling.

Consistent Expression of Human Protein-Coding Genes in PFB and WB Samples

Due to the limited sequencing depth of our data, we primarily focused on the examination of highly expressed protein-coding genes as compared to non-coding genes in our expression analysis to ensure the reliability of our findings. We performed PCA of the protein-coding gene expression profiles to evaluate the effects of the donor,

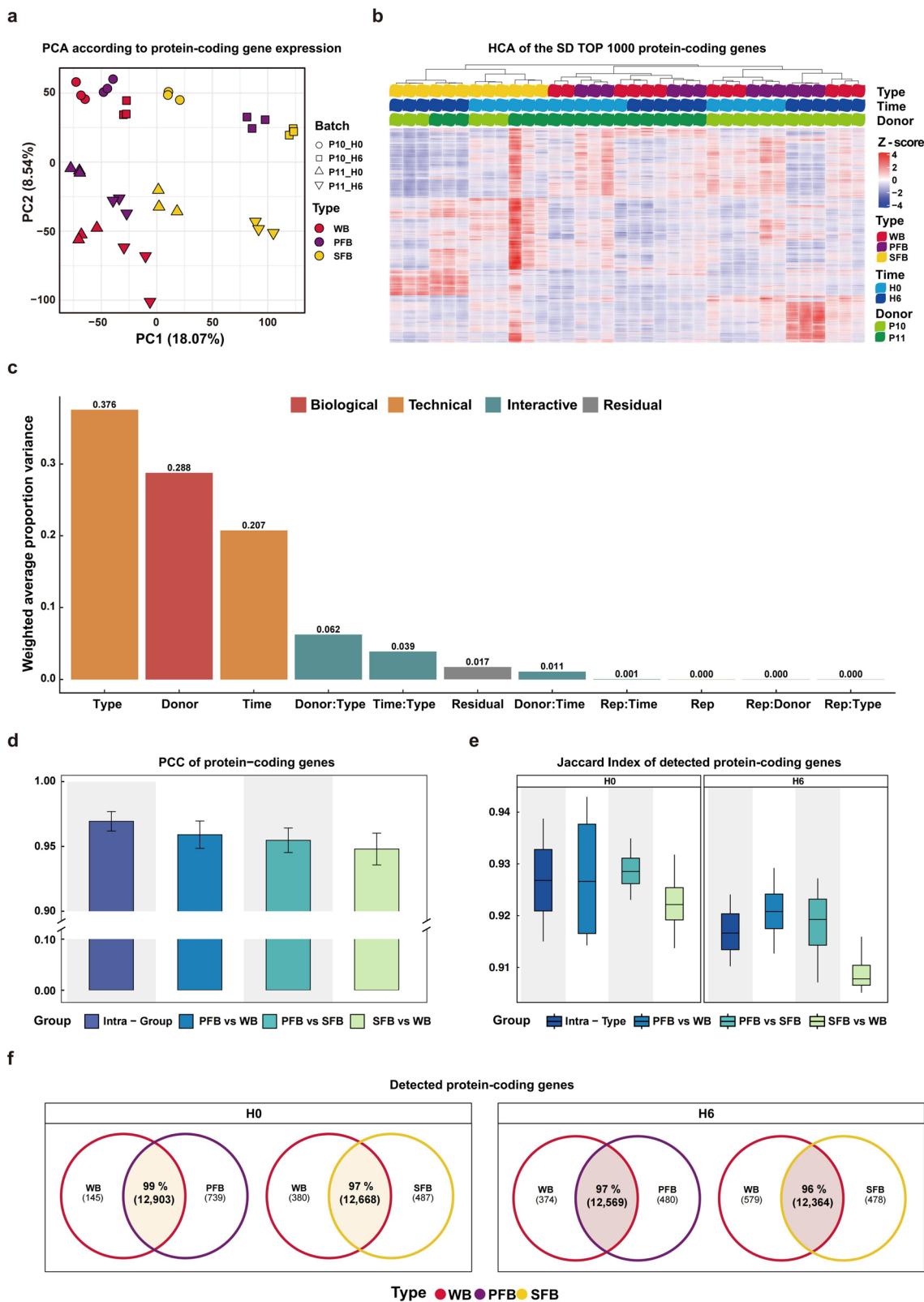
holding time, and blood type. The result showed that data were primarily separated by blood type on the first principal component (PC1) and by the donor on the second principal component (PC2), while samples from the same donor and holding time were clustered together by blood type (Fig. 3a). Notably, the PFB and WB samples of donor P10 from H0 were more closely related to each other than to other samples, and the same was true for the PFB and WB samples of donor P11, indicating a higher similarity between PFB and WB samples. To further investigate this similarity, we performed HCA on the SD TOP 1000 protein-coding genes in each sample, revealing that PFB and WB samples clustered together, whereas SFB samples clustered separately (Fig. 3b). Additionally, we employed PVCA to evaluate the contribution of different factors to transcriptional profile heterogeneity. The main factor was blood type (technical factor), followed by the donor (biological factor) and holding time (technical factor) (Fig. 3c).

Then PCC was calculated for the protein-coding gene expression profiles (counts) in different sample groups. The expression levels of protein-coding genes were highly correlated across blood types, where the correlations between PFB and WB samples were closest to those among technical replicates (positive control: Intra-Group) (Fig. 3d). Detailed correlation analysis for all samples is presented in Fig. S2. Additionally, the Jaccard index was used to measure the consistency of gene detection among groups. Our analysis revealed that all groups had Jaccard index values above 0.91 (Fig. 3e) and that the Jaccard index values between PFB and WB samples were closest to those between technical replicates, indicating the similarity in protein-coding gene detection between PFB and WB samples. Last, we employed Venn diagrams to compare the number and proportion of detected protein-coding genes among WB, PFB, and SFB samples at H0 and H6. The results showed that over 96% of the protein-coding genes were detectable in PFB vs WB samples and SFB vs WB samples (Fig. 3f) at H0 and H6, demonstrating that each blood type was amenable to human protein-coding gene detection.

Above all, we demonstrated the consistent expression of human protein-coding genes in PFB and WB samples, highlighting their high similarity. However, to exclude the effect of holding time on our main findings (Fig. 3c), we only selected samples from H0 in subsequent analysis.

Less Variation Between PFB and WB Samples in Differential Expression Analysis

To investigate the utility of the three blood types, we conducted a DEG analysis on WB, PFB, and SFB samples derived from two donors (P10 and P11) across three blood type comparisons (PFB vs WB, SFB vs WB, SFB vs PFB). As expected, we found that the DEGs proportion



between the three types of blood samples was low, ranging from a minimum of 0.46% (PFB vs WB in P11) to a maximum of 4.37% (SFB vs WB in P11) of all detected

genes. Specifically, we observed that the smallest proportion of DEGs was observed between PFB and WB samples, accounting for approximately 0.5% (Fig. 4a), which further

Fig. 3 Human protein-coding gene expression analysis. **a** Principal component analysis (PCA) based on FPKM values of protein-coding genes in each sample (red represents WB, purple represents PFB, and yellow represents SFB. The shape indicates the donor and the holding time of each sample.). **b** Heatmap of the top 1000 most variable (SD top 1000) protein-coding genes across all samples, selected and normalized by genes. “Type” represents the blood type, “Donor” represents the donor, and “Time” represents the holding time. **c** Principal variance component analysis (PVCA) based on log₂-transformed CPM values (+0.5) of the top 1000 SD protein-coding genes across all samples. **d** Barplot of the Pearson correlation coefficients for the protein-coding gene expression profiles (counts) in different sample groups, with the intra-group indicating the same blood type at different holding times. **e** Boxplot of Jaccard index of protein-coding genes expression levels (counts) in different groups at H0 and H6, with intra-type indicating samples from the same blood type in the same holding time. **f** Venn diagram of the number of detected expression genes in different blood types at H0 and H6

indicated that PFB and WB samples may have similar gene expression profiles.

Moreover, we identified the DEGs between the two donors across three blood types (P10_WB vs P11_WB, P10_PFB vs P11_PFB, and P10_SFB vs P11_SFB) to assess the ability to detect biological differences. To determine the extent of overlap between DEGs, we compared PFB and SFB with WB samples. PFB samples covered 61% (88 in PFB / 144 in WB) DEGs of protein-coding genes and 73% (27 in PFB / 37 in WB) DEGs of non-coding genes in WB samples, while SFB samples can cover 53% (76 in SFB / 144 in WB) DEGs of protein-coding genes and 65% (24 in SFB / 37 in WB) DEGs of non-coding genes in WB samples (Fig. 4b). The results showed that PFB samples shared more DEGs with WB than SFB did, indicating that PFB samples were closer to WB samples in terms of differential gene expression analysis. Generally, a limited number of DEGs are observed between healthy individuals, and these are typically not utilized in pathway enrichment. However, the enrichment analysis of DEGs in WB samples revealed that only one pathway (external side of plasma membrane) was significantly enriched between donors, suggesting that inter-individual differences in gene expression may not have significant biological implications. These findings suggested that PFB samples exhibited a greater similarity to WB samples in terms of differential gene expression analysis compared to SFB samples.

Similar Properties of Immune Cell Expression in PFB and WB Samples

To compare the three blood types more comprehensively, we analyzed the composition of immune cells and the expression of related genes in WB, PFB, and SFB samples. The abundance of 22 infiltrating immune cells in each sample was estimated by CIBERSORT (Newman et al. 2015).

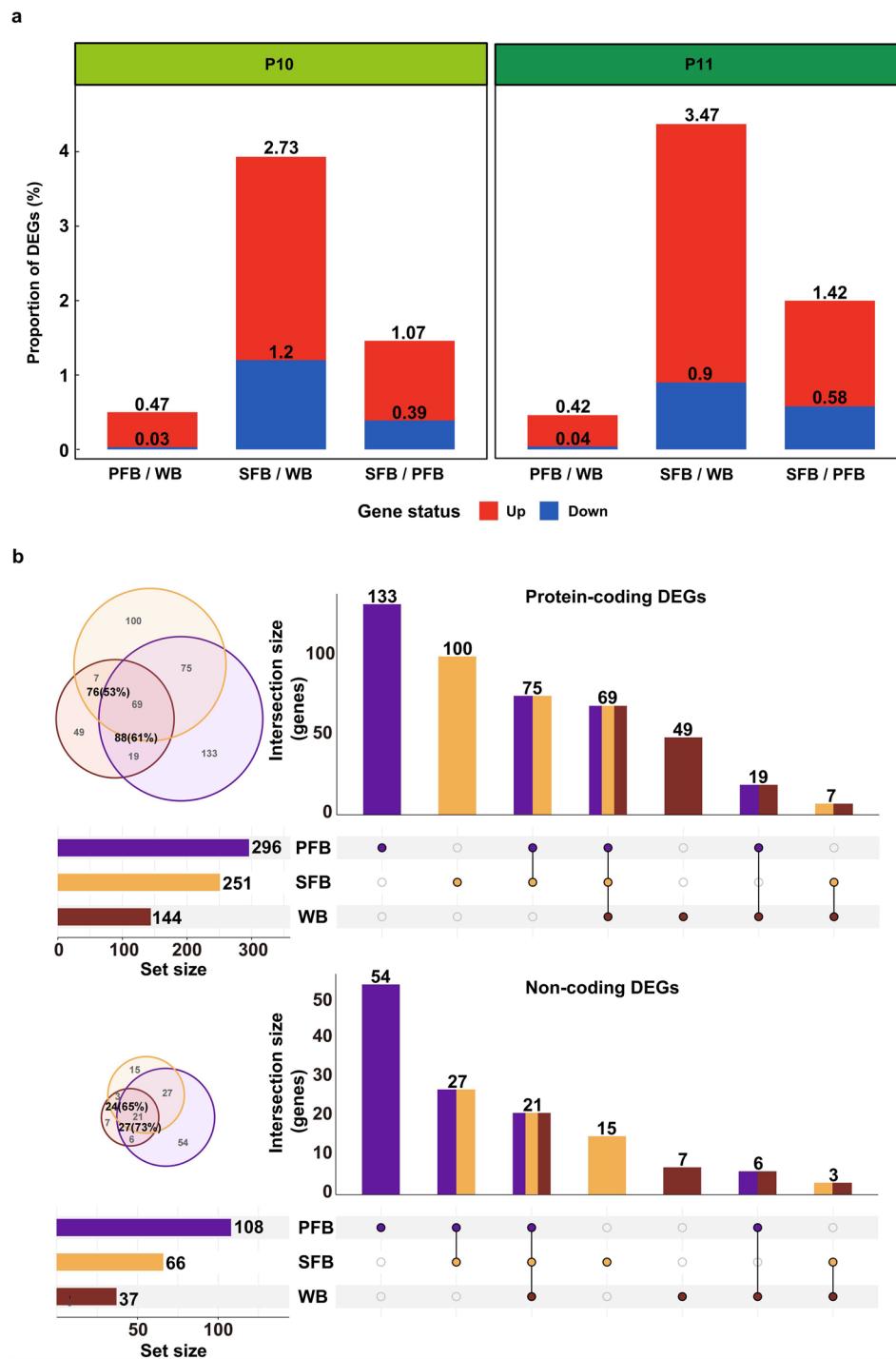
We selected six cell types that had an average abundance of more than 1/22 across all samples for further analysis. HCA showed that WB and PFB samples clustered together while SFB samples formed a separate cluster (Fig. 5a). In addition, the inferred composition of immune cells in PFB samples showed greater similarity to that in WB samples, as compared to SFB samples, across the six immune cell types analyzed. The comparison between groups revealed that WB and PFB samples had no significant difference in immune cell composition except for neutrophils. SFB samples had higher proportions of neutrophils and lower proportions of resting NK cells than other samples (Fig. 5b). This may be due to the influence of platelets and other components in WB, PFB, and SFB samples that affected neutrophils content and expression (Ruf and Ruggeri 2010).

Furthermore, we collected 60 immune cell-related genes specific to major blood cell types from the HPA database. As different methods for classifying immune cells may vary, these six cell types of HPA may not entirely match the cell types from CIBERSORT results. After filtering out genes with unusually high or low expression levels across all samples, a subset of retained 21 genes was selected for subsequent clustering and comparative analysis. The clustering analysis produced consistent findings with the prior observations, signifying that WB and PFB samples also clustered together while SFB samples constituted a distinct cluster (Fig. 5c). Moreover, the comparative analysis between groups demonstrated that most immune cell-specific gene expressions were analogous between PFB and WB samples, whereas SFB samples exhibited certain discrepancies (Fig. 5d, Fig. S2). These results suggested that PFB and WB samples shared similar immunological characterizations.

Discussion

Enhancing the utilization of blood samples can promote the advancement of precision medicine. In this study, we comprehensively compared transcriptomic profiles across three types of blood samples obtained from two healthy donors collected at two different holding times. The results implied that PFB samples may serve as a potential alternative to WB samples for specific applications where WB samples are impractical or unfeasible in retrospective and multi-omics studies. We confined the study to two donors to maintain the reproducibility of inter-individual biological differences. Our findings exhibited a high correlation between PFB and WB samples at the expression profile level, with fewer DEGs between PFB and WB samples compared to SFB and WB samples. Notably, the individual difference (donor) profiles also demonstrated greater similarity between PFB and WB samples than between SFB and WB samples. Besides, we showed that PFB and WB samples had analogous gene

Fig. 4 Differential expression genes analysis in different groups. **a** The percentage of differentially expressed genes (DEGs) between different blood samples from two donors. The comparisons are PFB vs WB, SFB vs WB, and SFB vs PFB. Red indicates significantly upregulated genes, and blue indicates significantly down-regulated genes. **b** The number of DEGs between two different donors for the three blood types. The comparisons are P10_WB vs P11_WB, P10_PFB vs P11_PFB, and P10_SFB vs P11_SFB. The Venn diagram and upset plot provide an overview of the number of DEGs from the three blood types, including both protein-coding (up) and non-coding DEGs (down)



expression traits of immune cell proportion and immune cell-specific gene sets. These observations bore substantial biological relevance, underscoring the feasibility of employing PFB samples as a practical substitute for WB samples

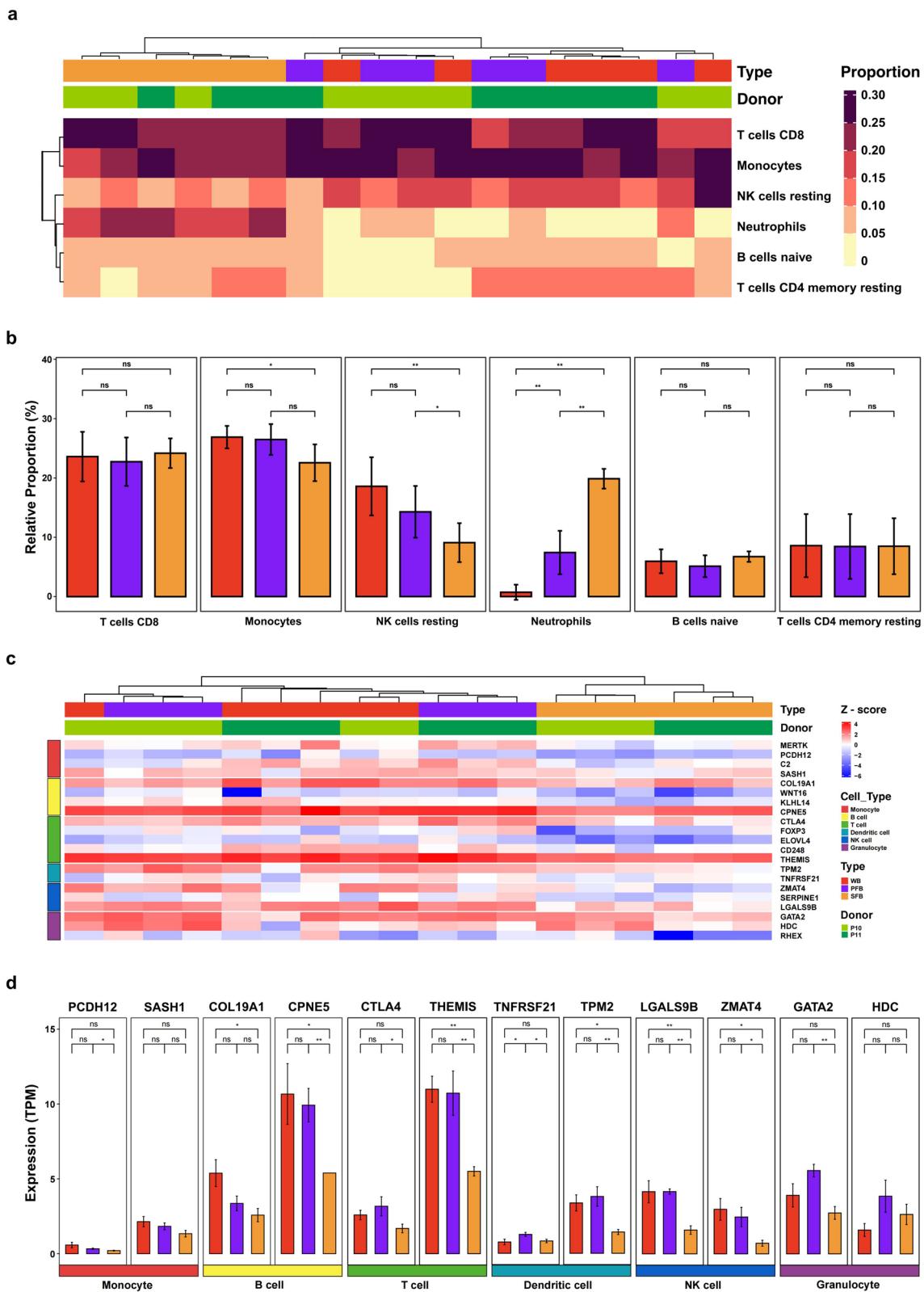
in investigations about immune functionality and disease research (Feng et al. 2022). In summary, PFB samples could serve as an alternative to WB samples for RNAseq, thereby augmenting the significance of PFB samples in biobanks.

The reason for more overlapping DEGs between two donors for PFB and WB samples (Fig. 4b) could be the higher number of DEGs in PFB samples compared to SFB samples. This discrepancy in DEG numbers stems from the fact that PFB samples have the highest number of detected genes (Fig. 3f). The increased number of detected genes in PFB samples can be attributed to the higher RIN score (Fig. 2a, right), indicating better RNA quality. The improved RNA quality in PFB samples enhances the sensitivity of gene detection, enabling the identification of more DEGs (Li et al. 2014). Additionally, only one pathway (external side of the plasma membrane) was significantly enriched between donors in WB samples, indicating that the inter-individual differences in gene expression between two healthy individuals might not carry substantial biological implications. Moreover, we identified the DEGs (Table S3) between WB and PFB samples for two donors and performed GO and KEGG pathway enrichment analysis. These common pathways are implicated in important biological processes, including immune regulation, cellular signaling, metabolic regulation, hormone synthesis, and action (Tables S4–S5). Further exploration of the regulatory mechanisms underlying these DEGs within specific pathways is crucial to enhance our understanding of the biological distinctions between PFB and WB samples. In summary, these differences may be due to the lack of plasma-specific components in PFB samples, such as liver-derived proteins or cytokines, which can contribute to the plasma-specific gene expression changes in WB samples (Benjamin and McLaughlin 2012; Rosenberg-Hasson et al. 2014; Liu et al. 2021). These results indicate the potential limitations of utilizing PFB samples as substitutes for WB samples in transcriptome analysis, particularly when investigating plasma-specific gene expression changes associated with altered immune responses or inflammatory conditions. Besides, discrepancies in neutrophil expression were observed between PFB and WB samples, suggesting that PFB samples might have certain limitations when employed as a surrogate for WB samples in the context of specific neutrophil-associated diseases (Wang et al. 2018; George et al. 2022; Gungabeesoon et al. 2023). Further investigation is required to clarify these differences and evaluate their potential consequences on the study of diseases about neutrophil function (Liew and Kubes 2019).

SFB samples are not recommended to be used as a substitute for WB samples in transcriptomic research unless they are the only available blood type. SFB samples may have differences in composition and cellular components compared to WB samples, which can affect the reliability and generalizability of the research findings. Additionally, in specific scenarios where PFB samples are considered as an alternative to WB samples, caution should be exercised

when the research primarily focuses on cell-free molecules. PFB samples lack the plasma component, which may impact the analysis and interpretation of results related to cell-free molecules. Researchers should be mindful of these considerations to ensure the appropriate selection and utilization of blood samples in transcriptomic studies. Apart from that, the availability of a broader range of blood samples for transcriptomics, including PFB samples, may expand the opportunities for investigating gene expression in diverse contexts. Several alternative blood types, such as resected clots from large vessel acute ischemic stroke (Tutino et al. 2021), dried blood spots (Reust et al. 2018), placenta and umbilical cord blood from pregnant women and newborns (Lu et al. 2022), have demonstrated promise in transcriptome analysis, yielding valuable insights into disease mechanisms and pathways. Consequently, the continued investigation of these alternative blood types is of paramount importance.

It is important to acknowledge the limitations of this study, which are as follows. First, the study utilized a relatively modest sample size, comprising blood samples from merely two healthy donors, each with three technical replicates. To bolster the validity of our findings, we advocated for the enlargement of the participant pool in subsequent research endeavors, incorporating multiple paired-donor comparative analysis. Second, our analysis primarily focused on comparing transcriptomic results of PFB and WB samples from healthy donors, and we did not conduct a systematic analysis of the transcriptome specifically in the context of disease or stimulus for the blood samples. However, it is important to note that the extent to which various factors influence the similarity of transcriptomic analysis results between WB and PFB samples may vary depending on the specific disease, stimulus, and biological context. Further research is needed to comprehensively evaluate the utility and applicability of PFB samples in disease-related transcriptomic studies. Third, despite prior studies indicating that blood samples held for 6 h did not significantly impact RNA concentration (Kim et al. 2007) and integrity (Huang et al. 2017), our RNA expression profiles showed a discernible degree of differences between inter-type samples and intra-type samples held for 0.5 h and 6 h. The causal mechanism of this observation remains unclear and needs further investigation. Last, as transcriptome analysis of blood samples predominantly involved protein-coding genes (Wang et al. 2020), our study mainly focused on protein-coding genes, and the expression patterns of non-coding genes in different blood types remain to be studied.



◀Fig. 5 Inferred immunological characterizations. **a** Heatmap and hierarchical clustering of inferred relative abundance of immune cells in three blood types at H0 by CIBERSORT. Lime green indicates P10, and dark green indicates P11. **b** The proportion of six major immune cell types in each blood type at H0. **c** Heatmap and hierarchical clustering representing the expression levels of 21 immune cell-specific genes selected from the Human Protein Atlas (HPA) across six major immune cell types at H0 (pale peach: monocyte, deep pink: B_cell, magenta: T_cell, violet: dendritic_cell, dark purple: NK cell, black: granulocyte). **d** Comparison of expression levels of immune cell-specific genes between three blood types in six immune cell types

Conclusion

This study confers a substantial understanding pertaining to the feasibility of deploying PFB samples as a viable substitute for WB samples in RNAseq applications. The findings present crucial perspectives and potential approaches for increasing the use of infrequently collected blood samples in biospecimen repositories for retrospective studies and maximizing the utilization of blood samples in multi-omics cohort studies. Ultimately, the insights gleaned from this research contribute to the refinement of blood sample utilization in transcriptomic studies and the progression of precision medicine research.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43657-023-00121-1>.

Acknowledgements This study was supported in part by the National Natural Science Foundation of China (31720103909 and 32170657), the National Key R&D Project of China (2018YFE0201603, 2018YFE0201600, and 2021YFF1201305), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), State Key Laboratory of Genetic Engineering (SKLGE-2117), and the 111 Project (B13016). Fig. 1 has been created with <https://www.Biorender.com>.

Authors' Contributions YZ, LS, and WH conceived the research and constructed the experimental design. YZ, LS, YY, and WH managed the project. QC, XG, ES, and CZ analyzed the data and participated in the verification and interpretation of data. WH, HW, SS, PZ, HJ, and QN conducted blood sample collection, RNA purification, cDNA library construction, and quality control experiments. YL and YY provided some critical ideas for data analysis and manuscript preparation. XG and QC drafted the initial version of the manuscript. ZC and RZ contributed to the final revision of the paper. All authors have thoroughly reviewed and approved the final manuscript.

Data Availability The blood samples, RNA materials, and datasets generated during the current study are available from the corresponding authors upon reasonable request.

Code Availability Our pipeline applications include all the Workflow Description Language (WDL, <https://github.com/openwdl/wdl>) processes required for RNAseq upstream analysis before the R analyses. Please note that the parameters in the pipeline applications are fixed, and the sample files can be directly processed. We have made the upstream analysis of RNAseq available on our GitHub repository (<https://github.com/fudan-tnbc>), and the dockers used in the upstream analysis can be obtained from Docker Hub (<https://hub.docker.com/u/chengqingwang>). The code used for pre-processing the upstream result

data and drawing the figures can be found on our GitHub repository (<https://github.com/xiaorouguo/c3bta>). Quality control metrics for all data were collected in the metadata tables and visualized using R v4.2.1 (<https://cran.r-project.org/>, R development core team).

Declarations

Conflict of Interest The authors declare no competing financial interests. ZC and RZ are the editorial operation team members of *Phenomics*, and they were not involved in reviewing this paper.

Ethical Approval This study was approved by the ethics committee of the School of Life Sciences of Fudan University (Shanghai, China) (ID: BE2050).

Consent to Participate Written informed consent was obtained from the participants.

Consent for Publication All the participants approved to publish.

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/>.

References

- Bayot ML, Tadi P (2022) Laboratory tube collection. StatPearls, Treasure Island
- Benjamin RJ, McLaughlin LS (2012) Plasma components: properties, differences, and uses. *Transfusion* 52(8):19S-19S. <https://doi.org/10.1111/j.1537-2995.2012.03622.x>
- Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, Fang S, Cao W, Yi L, Zhao Y, Kong L (2021) KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res* 49(W1):W317–W325. <https://doi.org/10.1093/nar/gkab447>
- Chebbo M, Assou S, Pantesco V, Duez C, Alessi MC, Chanze P, Gras D (2022) Platelets purification is a crucial step for transcriptomic analysis. *Int J Mol Sci.* <https://doi.org/10.3390/ijms23063100>
- Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen Q, Liu Y, Gao Y, Zhang R, Hou W, Cao Z, Jiang YZ, Zheng Y, Shi L, Ma D, Yang J, Shao ZM, Yu Y (2022) A comprehensive genomic and transcriptomic dataset of triple-negative breast cancers. *Sci Data* 9(1):587. <https://doi.org/10.1038/s41597-022-01681-z>
- de Almeida Chuffa LG, Freire PP, Dos Santos SJ, de Mello MC, de Oliveira NM, Carvalho RF (2022) Aging whole blood transcriptome reveals candidate genes for SARS-CoV-2-related vascular and immune alterations. *J Mol Med (berl)* 100(2):285–301. <https://doi.org/10.1007/s00109-021-02161-4>

- De Ruysscher D, Jin J, Lautenschlaeger T, She JX, Liao Z, Kong FS (2017) Blood-based biomarkers for precision medicine in lung cancer: precision radiation therapy. *Transl Lung Cancer Res* 6(6):661–669. <https://doi.org/10.21037/tlcr.2017.09.12>
- Debey-Pascher S, Hofmann A, Kreusch F, Schuler G, Schuler-Thurner B, Schultze JL, Staratschek-Jox A (2011) RNA-stabilized whole blood samples but not peripheral blood mononuclear cells can be stored for prolonged time periods prior to transcriptome analysis. *J Mol Diagn* 13(4):452–460. <https://doi.org/10.1016/j.jmoldx.2011.03.006>
- Donohue DE, Gautam A, Miller SA, Srinivasan S, Abu-Amara D, Campbell R, Marmar CR, Hammamieh R, Jett M (2019) Gene expression profiling of whole blood: a comparative assessment of RNA-stabilizing collection methods. *PLoS ONE* 14(10):e0223065. <https://doi.org/10.1371/journal.pone.0223065>
- Dvigne H, Ries RE, Ilagan JO, Stirewalt DL, Meshinchi S, Bradley RK (2014) Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc Natl Acad Sci U S A* 111(47):16802–16807. <https://doi.org/10.1073/pnas.1413374111>
- Feng Q, Lin S, Liu H, Yang B, Han L, Han X, Xu L, Xie Z (2022) Meta-analysis of whole blood transcriptome datasets characterizes the immune response of respiratory syncytial virus infection in children. *Front Cell Infect Microbiol* 12:878430. <https://doi.org/10.3389/fcimb.2022.878430>
- Fresard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, Bonner D, Kernohan KD, Marwaha S, Zappala Z, Balliu B, Davis JR, Liu B, Prybol CJ, Kohler JN, Zastrow DB, Reuter CM, Fisk DG, Grove ME, Davidson JM, Hartley T, Joshi R, Strober BJ, Utiramerur S, Lind L, Ingelsson E, Battle A, Bejerano G, Bernstein JA, Ashley EA, Boycott KM, Merker JD, Wheeler MT, Montgomery SB (2019) Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med* 25(6):911–919. <https://doi.org/10.1038/s41591-019-0457-8>
- Gautam A, Donohue D, Hoke A, Miller SA, Srinivasan S, Sowe B, Detwiler L, Lynch J, Levangie M, Hammamieh R, Jett M (2019) Investigating gene expression profiles of whole blood and peripheral blood mononuclear cells using multiple collection and processing methods. *PLoS ONE* 14(12):e0225137. <https://doi.org/10.1371/journal.pone.0225137>
- George PM, Reed A, Desai SR, Devaraj A, Faiez TS, Laverty S, Kanwal A, Esneau C, Liu MKC, Kamal F, Man WD-C, Kaul S, Singh S, Lamb G, Faizi FK, Schuliga M, Read J, Burgoine T, Pinto AL, Micallef J, Bauwens E, Candiracci J, Bougoussa M, Herzog M, Raman L, Ahmetaj-Shala B, Turville S, Aggarwal A, Farne HA, Dalla Pria A, Aswani AD, Patella F, Borek WE, Mitchell JA, Bartlett NW, Dokal A, Xu X-N, Kelleher P, Shah A, Singanayagam A (2022) A persistent neutrophil-associated immune signature characterizes post-COVID-19 pulmonary sequelae. *Sci Transl Med* 14(671):eab05795. <https://doi.org/10.1126/scitranslmed.abo5795>
- Gungabeesoon J, Gort-Freitas NA, Kiss M, Bolli E, Messemaker M, Siwicki M, Hicham M, Bill R, Koch P, Cianciaruso C, Duval F, Pfirschke C, Mazzola M, Peters S, Homicsko K, Garris C, Weissleder R, Klein AM, Pittet MJ (2023) A neutrophil response linked to tumor control in immunotherapy. *Cell* 186(7):1448–1464.e1420. <https://doi.org/10.1016/j.cell.2023.02.032>
- Harrington CA, Fei SS, Minnier J, Carbone L, Searles R, Davis BA, Ogle K, Planck SR, Rosenbaum JT, Choi D (2020) RNA-Seq of human whole blood: evaluation of globin RNA depletion on Ribo-Zero library method. *Sci Rep* 10(1):6271. <https://doi.org/10.1038/s41598-020-62801-6>
- He D, Yang CX, Sahin B, Singh A, Shannon CP, Oliveria J-P, Gauvreau GM, Tebbutt SJ (2019a) Whole blood vs PBMC: compartmental differences in gene expression profiling exemplified in asthma. *Allergy Asthma Clin Immunol* 15(1):67. <https://doi.org/10.1186/s13223-019-0382-x>
- He Y, Zhou Y, Ma W, Wang J (2019b) An integrated transcriptomic analysis of autism spectrum disorder. *Sci Rep* 9(1):11818. <https://doi.org/10.1038/s41598-019-48160-x>
- Hong S, Banchereau R, Maslow BL, Guerra MM, Cardenas J, Baisch J, Branch DW, Porter TF, Sawitzke A, Laskin CA, Buyon JP, Merrill J, Sammaritano LR, Petri M, Gatewood E, Cepika AM, Ohouo M, Obermoser G, Anguiano E, Kim TW, Nulsen J, Nehar-Belaid D, Blankenship D, Turner J, Banchereau J, Salmon JE, Pascual V (2019) Longitudinal profiling of human blood transcriptome in healthy and lupus pregnancy. *J Exp Med* 216(5):1154–1169. <https://doi.org/10.1084/jem.20190185>
- Huang LH, Lin PH, Tsai KW, Wang LJ, Huang YH, Kuo HC, Li SC (2017) The effects of storage temperature and duration of blood samples on DNA and RNA qualities. *PLoS ONE* 12(9):e0184692. <https://doi.org/10.1371/journal.pone.0184692>
- Husseini AA, Derakhshandeh M, Tatlisu NB (2022) Comprehensive review of transcriptomics (RNAs) workflows from blood specimens. *Sep Purif Rev* 51(1):57–77. <https://doi.org/10.1080/1542219.2020.1831537>
- Kamali Z, Keaton JM, Haghjooy Javanmard S, International Consortium Of Blood P, Million Veteran P, e QC, Bios C, Edwards TL, Snieder H, Vaez A (2022) Large-scale multi-omics studies provide new insights into blood pressure regulation. *Int J Mol Sci*. <https://doi.org/10.3390/ijms23147557>
- Kim SJ, Dix DJ, Thompson KE, Murrell RN, Schmid JE, Gallagher JE, Rockett JC (2007) Effects of storage, RNA extraction, genechip type, and donor sex on gene expression profiling of human whole blood. *Clin Chem* 53(6):1038–1045. <https://doi.org/10.1373/clinc hem.2006.078436>
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37(8):907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Krishnan A, Thomas S (2022) Toward platelet transcriptomics in cancer diagnosis, prognosis and therapy. *Br J Cancer* 126(3):316–322. <https://doi.org/10.1038/s41416-021-01627-z>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, Viale A, Wright C, Schweitzer PA, Gao Y, Kim D, Boland J, Hicks B, Kim R, Chhangawala S, Jafari N, Raghavachari N, Gandara J, Garcia-Reyero N, Hendrickson C, Roberson D, Rosenfeld J, Smith T, Underwood JG, Wang M, Zumbo P, Baldwin DA, Grills GS, Mason CE (2014) Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol* 32(9):915–925. <https://doi.org/10.1038/nbt.2972>
- Liew PX, Kubis P (2019) The neutrophil's role during health and disease. *Physiol Rev* 99(2):1223–1248. <https://doi.org/10.1152/physrev.00012.2018>
- Liu C, Chu D, Kalantar-Zadeh K, George J, Young HA, Liu G (2021) Cytokines: from clinical significance to quantification. *Adv Sci (weinh)* 8(15):e2004433. <https://doi.org/10.1002/advs.202004433>
- Lu S, Wang J, Kakongoma N, Hua W, Xu J, Wang Y, He S, Gu H, Shi J, Hu W (2022) DNA methylation and expression profiles of placenta and umbilical cord blood reveal the characteristics of gestational diabetes mellitus patients and offspring. *Clin Epigenet* 14(1):69. <https://doi.org/10.1186/s13148-022-01289-5>
- Mastrokolias A, den Dunnen JT, van Ommen GB, t Hoen PAC, van Roon-Mom WMC (2012) Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in

- human blood RNA. *BMC Genomics* 13(1):28. <https://doi.org/10.1186/1471-2164-13-28>
- Mjelle R, Dima SO, Bacalbasa N, Chawla K, Sorop A, Cucu D, Herlea V, Saetrom P, Popescu I (2019) Comprehensive transcriptomic analyses of tissue, serum, and serum exosomes from hepatocellular carcinoma patients. *BMC Cancer* 19(1):1007. <https://doi.org/10.1186/s12885-019-6249-1>
- Mohr S, Liew CC (2007) The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol Med* 13(10):422–432. <https://doi.org/10.1016/j.molmed.2007.08.003>
- Mukherjee P, Burgio G, Heitlinger E (2021) Dual RNA sequencing meta-analysis in plasmodium infection identifies host-parasite interactions. *mSystems*. <https://doi.org/10.1128/mSystems.00182-21>
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12(5):453–457. <https://doi.org/10.1038/nmeth.3337>
- Okonechnikov K, Conesa A, Garcia-Alcalde F (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32(2):292–294. <https://doi.org/10.1093/bioinformatics/btv566>
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33(3):290–295. <https://doi.org/10.1038/nbt.3122>
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11(9):1650–1667. <https://doi.org/10.1038/nprot.2016.095>
- Qi D, Geng Y, Cardenas J, Gu J, Yi SS, Huang JH, Fonkem E, Wu E (2023) Transcriptomic analyses of patient peripheral blood with hemoglobin depletion reveal glioblastoma biomarkers. *NPJ Genom Med* 8(1):2. <https://doi.org/10.1038/s41525-022-00348-3>
- Qin Y, Yao J, Wu DC, Nottingham RM, Mohr S, Hunicke-Smith S, Lambowitz AM (2016) High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA* 22(1):111–128. <https://doi.org/10.1261/rna.054809.115>
- Reust MJ, Lee MH, Xiang J, Zhang W, Xu D, Batson T, Zhang T, Downs JA, Dupnik KM (2018) Dried blood spot RNA transcriptomes correlate with transcriptomes derived from whole blood RNA. *Am J Trop Med Hyg* 98(5):1541–1546. <https://doi.org/10.4269/ajtmh.17-0653>
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* 43(7):e47. <https://doi.org/10.1093/nar/gkv007>
- Roncaglia P, van Dam TJP, Christie KR, Nacheva L, Toedt G, Huynen MA, Huntley RP, Gibson TJ, Lomax J (2017) The gene ontology of eukaryotic cilia and flagella. *Cilia* 6:10. <https://doi.org/10.1186/s13630-017-0054-8>
- Rosenberg-Hasson Y, Hansmann L, Liedtke M, Herschmann I, Maecker HT (2014) Effects of serum and plasma matrices on multiplex immunoassays. *Immunol Res* 58(2–3):224–233. <https://doi.org/10.1007/s12026-014-8491-6>
- Ruan QL, Yang QL, Gao YX, Wu J, Lin SR, Zhou JY, Shao LY, Wang S, Liu QQ, Gao Y, Jiang N, Zhang WH (2021) Transcriptional signatures of human peripheral blood mononuclear cells can identify the risk of tuberculosis progression from latent infection among individuals with silicosis. *Emerg Microbes Infect* 10(1):1536–1544. <https://doi.org/10.1080/2221751.2021.1915184>
- Ruf W, Ruggeri ZM (2010) Neutrophils release brakes of coagulation. *Nat Med* 16(8):851–852. <https://doi.org/10.1038/nm0810-851>
- Shen Y, Li R, Tian F, Chen Z, Lu N, Bai Y, Ge Q, Lu Z (2018) Impact of RNA integrity and blood sample storage conditions on the gene expression analysis. *Onco Targets Ther* 11:3573–3581. <https://doi.org/10.2147/ott.S158868>
- Shin H, Shannon CP, Fishbane N, Ruan J, Zhou M, Balshaw R, Wilson-McManus JE, Ng RT, McManus BM, Tebbutt SJ (2014) Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS ONE* 9(3):e91041. <https://doi.org/10.1371/journal.pone.0091041>
- Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4:1521. <https://doi.org/10.12688/f1000research.7563.2>
- Sotelo-Orozco J, Chen SY, Hertz-Pannier I, Slupsky CM (2021) A comparison of serum and plasma blood collection tubes for the integration of epidemiological and metabolomics data. *Front Mol Biosci* 8:682134. <https://doi.org/10.3389/fmolb.2021.682134>
- Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: the teenage years. *Nat Rev Genet* 20(11):631–656. <https://doi.org/10.1038/s41576-019-0150-2>
- Tutino VM, Fricano S, Frauens K, Patel TR, Monteiro A, Rai HH, Waqas M, Chaves L, Poppenberg KE, Siddiqui AH (2021) Isolation of RNA from acute ischemic stroke clots retrieved by mechanical thrombectomy. *Genes (basel)*. <https://doi.org/10.3390/genes12101617>
- Uhlen M, Karlsson MJ, Zhong W, Tebani A, Pou C, Mikes J, Lakshminikanth T, Forsstrom B, Edfors F, Odeberg J, Mardinoglu A, Zhang C, von Feilitzen K, Mulder J, Sjostedt E, Hober A, Oksvold P, Zwahlen M, Ponten F, Lindskog C, Sivertsson A, Fagerberg L, Brodin P (2019) A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science*. <https://doi.org/10.1126/science.aax9198>
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63. <https://doi.org/10.1038/nrg2484>
- Wang X, Yousefi S, Simon HU (2018) Necroptosis and neutrophil-associated disorders. *Cell Death Dis* 9(2):111. <https://doi.org/10.1038/s41419-017-0058-8>
- Wang D, Chen Z, Zhuang X, Luo J, Chen T, Xi Q, Zhang Y, Sun J (2020) Identification of circRNA-associated-ceRNA networks involved in milk fat metabolism under heat stress. *Int J Mol Sci.* <https://doi.org/10.3390/ijms21114162>
- Wingett SW, Andrews S (2018) FastQ screen: a tool for multi-genome mapping and quality control. *F1000Res* 7:1338. <https://doi.org/10.12688/f1000research.15931.2>
- Wu P, Chen D, Ding W, Wu P, Hou H, Bai Y, Zhou Y, Li K, Xiang S, Liu P, Ju J, Guo E, Liu J, Yang B, Fan J, He L, Sun Z, Feng L, Wang J, Wu T, Wang H, Cheng J, Xing H, Meng Y, Li Y, Zhang Y, Luo H, Xie G, Lan X, Tao Y, Li J, Yuan H, Huang K, Sun W, Qian X, Li Z, Huang M, Ding P, Wang H, Qiu J, Wang F, Wang S, Zhu J, Ding X, Chai C, Liang L, Wang X, Luo L, Sun Y, Yang Y, Zhuang Z, Li T, Tian L, Zhang S, Zhu L, Chang A, Chen L, Wu Y, Ma X, Chen F, Ren Y, Xu X, Liu S, Wang J, Yang H, Wang L, Sun C, Ma D, Jin X, Chen G (2021a) The trans-omics landscape of COVID-19. *Nat Commun* 12(1):4543. <https://doi.org/10.1038/s41467-021-24482-1>
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G (2021b) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (cambridge)* 2(3):100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Xing Y, Yang X, Chen H, Zhu S, Xu J, Chen Y, Zeng J, Chen F, Johnson MR, Jiang H, Wang WJ (2021) The effect of cell isolation methods on the human transcriptome profiling and microbial transcripts of peripheral blood. *Mol Biol Rep* 48(4):3059–3068. <https://doi.org/10.1007/s11033-021-06382-1>
- Yu Y, Hou W, Liu Y, Wang H, Dong L, Mai Y, Chen Q, Li Z, Sun S, Yang J, Cao Z, Zhang P, Zi Y, Liu R, Gao J, Zhang N, Li J, Ren L, Jiang H, Shang J, Zhu S, Wang X, Qing T, Bao D, Li B, Li B,

- Suo C, Pi Y, Wang X, Dai F, Scherer A, Mattila P, Han J, Zhang L, Jiang H, Thierry-Mieg D, Thierry-Mieg J, Xiao W, Hong H, Tong W, Wang J, Li J, Fang X, Jin L, Xu J, Qian F, Zhang R, Shi L, Zheng Y (2023) Quartet RNA reference materials improve the quality of transcriptomic data through ratio-based profiling. *Nat Biotechnol*. <https://doi.org/10.1038/s41587-023-01867-9>
- Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, Zhou R, Qiu W, Huang N, Sun L, Li X, Bin J, Liao Y, Shi M, Liao W (2021) IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front Immunol* 12:687975. <https://doi.org/10.3389/fimmu.2021.687975>
- Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM (2014) Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* 15(1):419. <https://doi.org/10.1186/1471-2164-15-419>
- Zhao Y, Li MC, Konate MM, Chen L, Das B, Karlovich C, Williams PM, Evrard YA, Doroshow JH, McShane LM (2021) TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *J Transl Med* 19(1):269. <https://doi.org/10.1186/s12967-021-02936-w>