

RESEARCH PAPER



Wide-spectrum profiling of plasma cell-free RNA and the potential for health-monitoring

Xinxin Wang^{a,b,*}, Shaogang Li^{a,b,*}, Rijiang Ou^{b,*}, Wending Pang^{a,b}, Yingying Wang^b, Yifan Zhang^{b,c}, Yu Lin^b, Changlin Yang^{b,c}, Wei Chen^{b,d}, Changgui Lei^b, Guodan Zeng^{b,c}, Wenwen Zhou^b, Yeqin Wang^b, Jianhua Yin^b, Haiqiang Zhang^b, Xin Jin^{b,d}, and Yan Zhang^b

^aSchool of Biology and Biological Engineering, South China University of Technology, Guangzhou, China; ^bBGI Research, Shenzhen, China; ^cCollege of Life Sciences, University of Chinese Academy of Sciences, Beijing, China; ^dSchool of Medicine, South China University of Technology, Guangzhou, Guangdong, China

ABSTRACT

Circulating cell-free RNA (cfRNA) has emerged as a promising analyte for disease detection. However, the comprehensive profiling of diverse cfRNA types remains under-characterized. Here, we applied a new wide-spectrum cfRNA sequencing method and simultaneously captured rRNA, tRNA, mRNA, miRNA, lncRNA and all mitochondrial RNA. The cfRNA compositions, size distributions and highly abundant cfRNA genes were analysed for each type of cfRNA. We depicted the cfRNA cell types of origin profiles of 66 generally healthy individuals and found that BMI showed a significant impact on the kidney-derived cfRNA proportion. Three individuals with some liver problems were identified because of relatively high levels of hepatocyte-specific cfRNA. The abundance levels of different genes and RNA types, including mRNA, miRNA and lncRNA, were significantly correlated with the liver function test results. The genes of individual cfRNA variances were enriched in pathways associated with common diseases such as liver diseases, virus infections, cancers and metabolic diseases. This study provided a profiling of cfRNA and displayed the potential of cfRNA as a biomarker in health monitoring.

ARTICLE HISTORY

Revised 10 February 2025
Accepted 13 March 2025

KEYWORDS

Plasma RNA; liquid biopsy; healthy people; RNA types; tissue-of-origin

Introduction


Circulating free nucleic acids, including cell-free DNA (cfDNA) and cell-free RNA (cfRNA), are widely existed in body fluids, such as blood plasma and urine [1]. The high-throughput sequencing of cfDNA has been extensively explored in the context of early detection, diagnosis and monitoring of various diseases [2,3]. The cfDNA fragmentomic features, such as fragment size, end motif and coverage patterns, have been recognized as promising non-invasive biomarkers [4]. In comparison, research on cfRNA dropped behind, with much fewer samples being comprehensively studied. The understanding of cfRNA biology is limited. This is largely due to the very low level, the highly degraded fragments, the instability of cfRNA and the technical limitations [5]. Nonetheless, researchers have shown that cfRNA contains abundant physiological and pathological signatures of human bodies. For instance, significant cfRNA differences have been found in the plasma of patients with COVID-19, cancer, pregnancy-related diseases and brain disorders [6–9]. Besides, the tissue and cell-type specific RNA levels can be employed to estimate the cfRNA origins and to reflect the

status of multiple organs and tissues. For example, foetus/placenta-derived cfRNA in maternal plasma can be used to predict gestational age, estimate the risk of preterm birth and predict the onset of preeclampsia [8,10,11]. The brain-specific cfRNA levels in the plasma of Alzheimer's disease (AD) patients reflect the pathophysiological alterations [6], corresponding well with the dysregulated gene expression discovered in the post-mortem brain tissues of AD patients [6]. These research have demonstrated the great potential of cfRNA applications in non-invasive disease prediction, diagnosis and monitoring.

Many early studies focused primarily on miRNA because it is relatively stable in plasma [12–14]. In recent years, cell-free mRNA has received increasing research interest. These studies with important discoveries include those above about pregnancy-related conditions, AD and cancers [6,7,15,16]. However, our understanding of the cfRNA characteristics is still not comprehensive. For example, the relative abundance profile of different types of RNA and their size distributions were not fully and simultaneously investigated. Researchers have reported some cfRNA profiling results using different

CONTACT Yan Zhang  zhangyan15@genomics.cn; Haiqiang Zhang  zhanghaiqiang@genomics.cn  BGI Research, Building 11, Beishan Industrial Zone, No. 146 Beishan Road, Yantian District, Shenzhen 518083, China; Xin Jin  jinjin@genomics.cn  School of Medicine, South China University of Technology, Guangzhou 510006, China

*These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2025.2481736>

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

RNA-seq methods; however, each method has its preference and most researches focused only on limited cfRNA types [17–20]. The types of cfRNA in plasma are as diverse as those in cells, including ribosomal RNA (rRNA), messenger RNA (mRNA), transfer RNA (tRNA), microRNA (miRNA), long non-coding RNA (lncRNA) and others [21]. Conventional RNA sequencing methods target or enrich specific types of RNA. The experimental protocols for mRNA and miRNA are usually quite different [22–24]. There is a lack of simple, stable and standardized cfRNA analysing methods, and some basic cfRNA characteristics are still not well-documented. Several research groups have developed cfRNA sequencing methods that involve a step of adding poly(A) tails to RNA fragments, thus enabling the capture of various types of RNA in addition to mRNA [25,26]. However, these research mainly focused on the disease signatures [9,27]. There is still a lack of systematic cfRNA exploration in healthy individuals.

In this study, we recruited 66 generally healthy individuals and comprehensively analysed plasma cfRNA characteristics through a poly(A)-adding method. The cfRNA compositions, size profiles, high-abundance cfRNA and cell-type origins were analysed. In addition, we analysed the individual cfRNA differences, and we demonstrated the potential of cfRNA signature as a biomarker in liver health monitoring.

Results

Blood sampling and cfRNA sequencing for various RNA types

In this study, we recruited 66 generally healthy adults aged from 21 to 69, during their physical examination. We collected their plasma samples, their blood test results and their personal information through questionnaires, including sex, age, BMI and self-reported health status (Supplementary Table S1). The blood samples were collected with EDTA tubes, and the plasma was separated within 4 h after blood sampling. The plasma samples were stored at -80°C until cfRNA extraction. The cfRNA sequencing library was prepared with our poly(A)-adding method. Briefly, the residual DNA was removed by DNase I; the cfRNA samples were treated with T4 PNK for 3' hydroxyl modification [24]; and then a poly(A) tail was added to the 3' end of the RNA fragment; the modified RNA fragments were then reverse-transcribed into cDNA using a primer containing poly(T). After the digestion of poly(T), the cDNA product was directly ligated with 5' and 3' adapters using a single-stranded library preparation technique, which is also a strand-specific library preparation method [28,29], with the adaptors specifically tailored for the DNBSEQ platform. After the PCR amplification, the preparation of cfRNA sequencing libraries was finished. The libraries were sequenced on the DNBSEQ platform with paired-end 100 bp mode (Figure 1A). Notably, after the cfRNA extraction, the samples were rapidly processed, and the reverse transcription step was combined with the poly(A)-adding step in one-tube reaction.

The raw reads of sequencing were processed through several filtering steps to remove adapter sequences, low-quality

reads and short reads (<17 bp). The average proportion of clean reads was 84.40%. The clean reads were then aligned to reference human rRNA sequences. As expected, the rRNA reads accounted for the majority of the reads, with an average proportion of 84.51% of the mapping reads. In addition, mitochondrial rRNA (mt-rRNA) reads accounted for an average of 8.06% of the mapping reads (Figure 1B). This rRNA proportion is consistent with the results in previous studies, which reported a range of 66% to 93% [30,31]. Then, rRNA and mt-rRNA reads were further excluded for downstream analyses. After the removal of rRNA and mt-rRNA reads, we found that the remaining RNA reads included those from yRNA (29.08%), mRNA (28.3%), tRNA (13.57%), mitochondrial tRNA (mt-tRNA) (10.02%) and mitochondrial mRNA (mt-mRNA) (6.91%). In addition, we captured a small proportion of lncRNA (1.64%), miRNA (0.41%) and some other RNA types (Figure 1B).

We assessed the quality of data after rRNA removal by analysing sequencing coverage, mRNA gene coverage and strand specificity. The average exonic reads ratio in our samples is 49.86% ($\text{SD}=9.07\%$). Taking two mRNA genes of relatively high and low cfRNA abundances as examples (*HBB* and *NONO*), the coverage analysis (Figure 1C) showed that reads were concentrated in the exonic regions but not in the intronic regions, indicating that our method effectively captures mRNA in plasma while avoiding genomic DNA contamination. Next, we evaluated the gene body coverage of mRNA. As a result, the reads from 5' to 3' of gene bodies were in a generally uniform manner. There is a slightly increased coverage at the 3'-end region of the mRNA (Figure 1D), which may probably due to the natural presence of poly(A) sequences at the 3' ends of mRNA. The strand specificity index is calculated by Qualimap [32], and the average strand specificity of our cfRNA data reached 0.90 ($\text{SD}=0.05$), which is comparable to the results of other strand-specific library preparation methods [33]. Taken together, the basic quality control showed our sequencing data are of relatively good quality. We then calculated the cfRNA abundance at the gene level using the Counts Per Million (CPM) normalization. In total, we identified 10,062 mRNA genes, 1,215 lncRNA genes, 195 tRNA genes and 121 miRNA genes in our data. These results indicate that our poly(A)-adding method successfully captured a variety of RNA types, including mRNA, lncRNA, miRNA, tRNA, yRNA, mt-rRNA, mt-mRNA, mt-tRNA, scRNA, snRNA and snoRNA.

The capture of size information indicated that relatively long cfRNA fragments exist in plasma

Most previous studies did not systematically explore the size distribution of plasma cfRNA for different RNA types. The lack of suitable detection methods is the main reason. As our library preparation method is based on the direct ligation of adaptors to both the 5' and 3' ends of cDNA, the size information of cfRNA fragments could largely be retained. To assess the feasibility of the size evaluation, we tested two synthesized RNA fragments of 25 nt and 62 nt, respectively. We prepared the sequencing libraries and analysed the data.

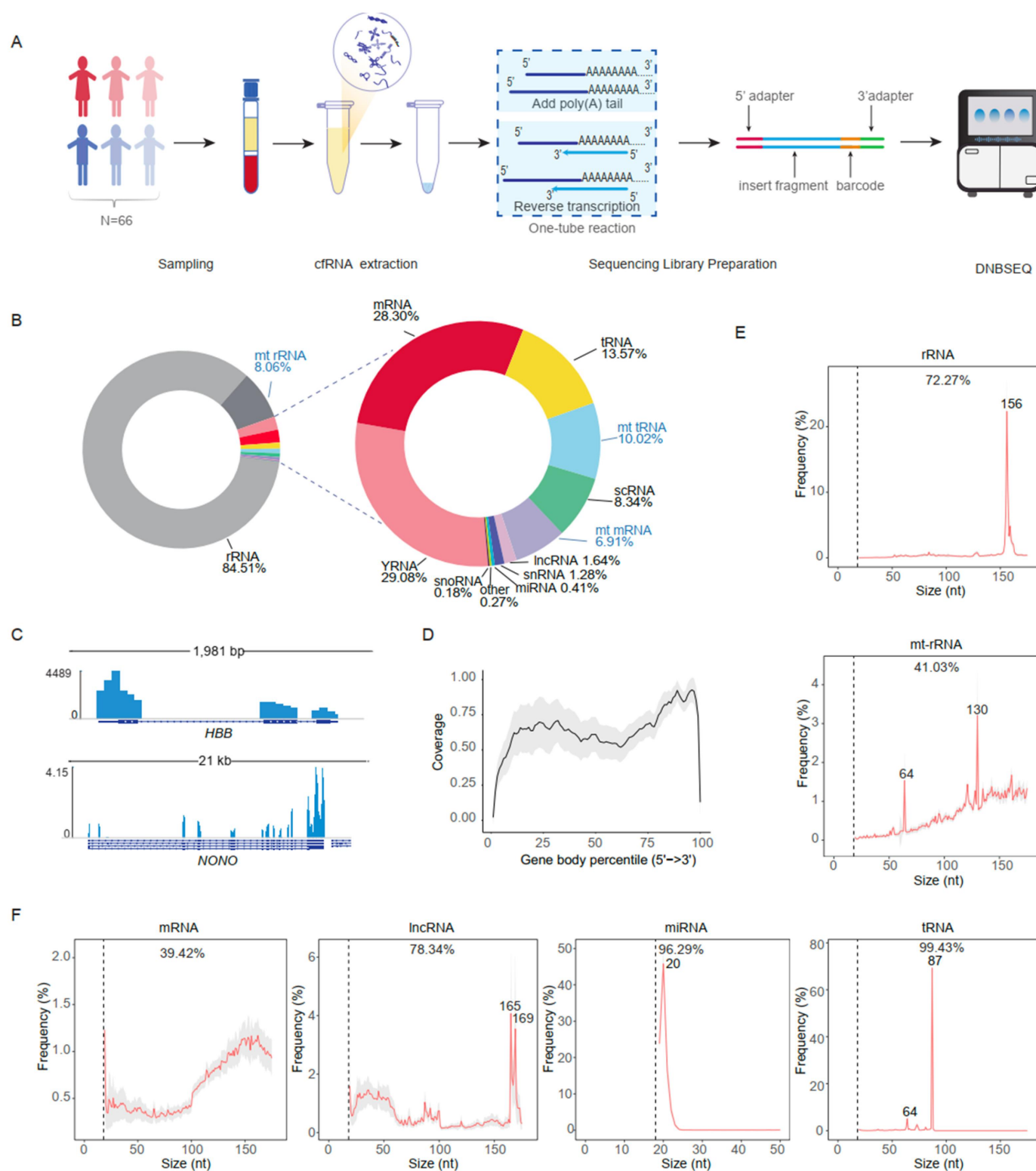


Figure 1. cRNA sequencing workflow and basic characterization. (A) blood sample collection and cRNA library preparation workflow. In this study, 66 generally healthy participants were recruited. The blood samples were collected by EDTA tubes and plasma was separated and stored at -80°C until cRNA extraction. The cRNAs were added with poly(a) tails by *E. coli* Poly(A) polymerase and then quickly reverse-transcribed into cDNA in the same reaction tube. After that, a single-stranded DNA library preparation technique based on splint ligation was utilized to add 5' and 3' adapters to the two ends of cDNA molecules. The ligation products were amplified by PCR to generate the sequencing libraries. The libraries were sequenced on the DNBSEQ platform with paired-end 100 bp reads. (B) the average read percentage of different types of RNA. (C) the sequencing coverage of two representative genes of high and low cRNA levels, *HBB* and *NONO*, were shown as examples, respectively. The blue bars represent the average sequencing coverage in samples, while the dark blue blocks and arrows at the bottom represent the exons and introns in the reference genome, respectively. (D) gene body coverage of mRNA, with the black line representing the average coverage and the grey shadow indicating the standard deviation, reflecting the uniformity of mRNA capturing during library preparation. The size distribution of cell-free (E) nuclear and mitochondrial rRNA; cell-free (F) mRNA, lncRNA, miRNA and tRNA, respectively. The red lines represent the average size distribution frequency of each RNA type, with the grey shadow indicating the standard deviation. The percentage numbers above the size profile indicate the proportion of cRNA fragments shorter than 175nt.

As a result, most of the reads contain the designed sequence and give the correct size information, with reads ratio of 80.37% for the 25 nt fragments and 88.87% for the 62 nt fragments, respectively. In the tests, the size of RNA fragments can be calculated by using either single-end reads or paired-end reads. These results demonstrated that our cfRNA sequencing method can be used to evaluate the size of short RNA fragments.

By merging PE 100 reads with the overlapping sequences between Read1 and Read2, we could accurately determine the fragment shorter than 175 nt. For longer fragments, the insert fragment can be estimated based on the read alignments on the reference genome (Supplementary Figure S1). The true fragment sizes may involve more complex factors such as RNA splicing, degradation, the quality of reverse transcription and adaptor ligation. In this study, we mainly focused on the more accurately estimated size range of 17–175 nt, and we also estimated the proportion of RNA fragments longer than 175 nt.

Previous studies generally considered cfRNA in plasma to be very short, with most cfRNA no longer than 50 nt [25,26]. However, a cfRNA study on breast and liver cancers reported that about 45% of cfRNA fragments exceeded 200 nt [31], and another study based on single-molecule nanopore sequencing also revealed the fraction of cfRNA longer than 200 nt [20]. In our cfRNA data, relatively long fragments were found for various RNA types. The proportion of RNA fragments longer than 175 nt were 27.73%, 58.97%, 60.58% and 21.66% for rRNA, mt-rRNA, mRNA and lncRNA, respectively (Figure 1E,F). The longest RNA fragments detected were more than 1000 nt for rRNA, 350 nt for mt-rRNA, 1000 nt for mRNA and 1000 nt for lncRNA (Supplementary Figure S1A, B, C). Within the range of 17 nt to 175 nt, the size distribution profiles were shown (Figure 1E,F). The size of mRNA fragments was more dispersedly distributed, and the majority of mRNA is longer than 100 nt, with a broad peak around 150 nt (Figure 1F). lncRNA has a proportion of relatively short fragments (<100 nt), and two narrow peaks at 165 nt and 169 nt, with the peaks contributed by the most highly abundant lncRNA (*LOC105370449*).

For miRNA and tRNA, relatively complete RNA fragments were captured. For example, miRNA fragments are typically in the range of 18–22 nt (Figure 1F), consistent with the length of mature miRNA [34]. In our result, tRNA fragments showed a dominant size peak at 87 nt, which is consistent with the size of tRNA (Figure 1F). The size distribution analysis indicates that miRNA and tRNA in plasma are protected by RNA binding protein complexes or other RNA carriers, such as extracellular vesicles (EVs) [35], and therefore largely remain full-length. It also demonstrated that our method is capable of effectively capturing complete RNA fragments. The rapid processing of plasma samples and the poly(A)-adding method provides valuable information about the size distributions of various RNA types.

Some of the top genes with relatively high cfRNA levels were less known before

The cfRNA levels of different genes vary widely. Here, the genes of mRNA, lncRNA and miRNA were analysed

separately. We ranked the genes by their average cfRNA abundance (CPM) in samples, and the genes were divided into four groups: top 10%, top 10–20%, top 20–50% and the remaining genes. The total cfRNA abundance (CPM) in each group was added up to calculate the ratio. The top 10% of genes for mRNA contribute to over 65% of the total mRNA abundance. For lncRNA and miRNA, this proportion is even higher, at 78.53% and 71.81%, respectively, indicating that a small subset of genes plays a predominant role in the overall abundance profile (Figure 2A).

In the mRNA abundance profile, the top five ranked genes are *HBB*, *HBA2*, *EEF1A1*, *WDR74* and *TMSB4X* (Figure 2B). The cfRNA abundances of the *HBB* and *HBA2* genes, which encode subunits of haemoglobin [36], are much higher than those of other genes. *EEF1A1*, *WDR74* and *TMSB4X* encode proteins crucial for translation elongation, ribosome assembly and actin polymerization, respectively [37–39]. The discovery of these mRNA top genes is not surprising, considering their important and basic roles in the blood and cells.

In the miRNA abundance profile (Figure 2C), the top 5 miRNA genes include *MIR15A*, *MIR16-1*, *MIR93*, *MIR16-2* and *MIR25*. These plasma cfRNA have been reported to be dysregulated in type 2 diabetes, hepatocellular carcinoma, lymphoma patients, postoperative colorectal cancer patients and oesophageal squamous cell carcinoma patients' plasma, respectively [40–44]. These miRNAs are all regarded as potential biomarkers due to their significant association with pathological conditions.

Regarding the highly abundant lncRNAs (Figure 2D), we have not yet found any plasma cfRNA reports about the top genes of *LOC105370449*, *LOC105379549*, *LOC107984865*, *HAR1A* and *LOC105375785*. Research on these genes in tissue and blood samples is also very limited. One study suggested that *LOC105375785* may be associated with the invasiveness of pituitary adenomas [45], and some other studies showed that *HAR1A* is dysregulated in multiple sclerosis, diffuse gliomas and non-small cell lung cancer [46–48].

Profiling of cell-free mitochondrial RNA

Current research on circulating free nucleic acids has only focused on mitochondrial DNA (mt-DNA) [49], while studies on mitochondrial RNA (mt-RNA) are relatively limited and usually focus on some specific mt-RNA [50–52]. In our study, by using the poly(A)-adding method, cfRNA from 37 genes in the mitochondria were all captured, including 2 rRNAs, 13 mRNAs and 22 tRNAs.

For all the mapping reads, mt-RNA reads accounted for 9.32%. Specifically, mt-rRNA reads represented 8.06% and the remaining mt-RNA reads accounted for 1.26%. After the removal of rRNA and mt-rRNA reads, mt-mRNAs and mt-tRNAs accounted for 6.91% and 10.02%, respectively (Figure 1B). The coverage profile of the mitochondrial genome is shown in Figure 3B, with all the gene regions covered by abundant sequencing reads, while the region without genes has no mapping reads. The mt-rRNA genes *RNR1* and *RNR2* exhibit the highest and most uniformly distributed sequencing

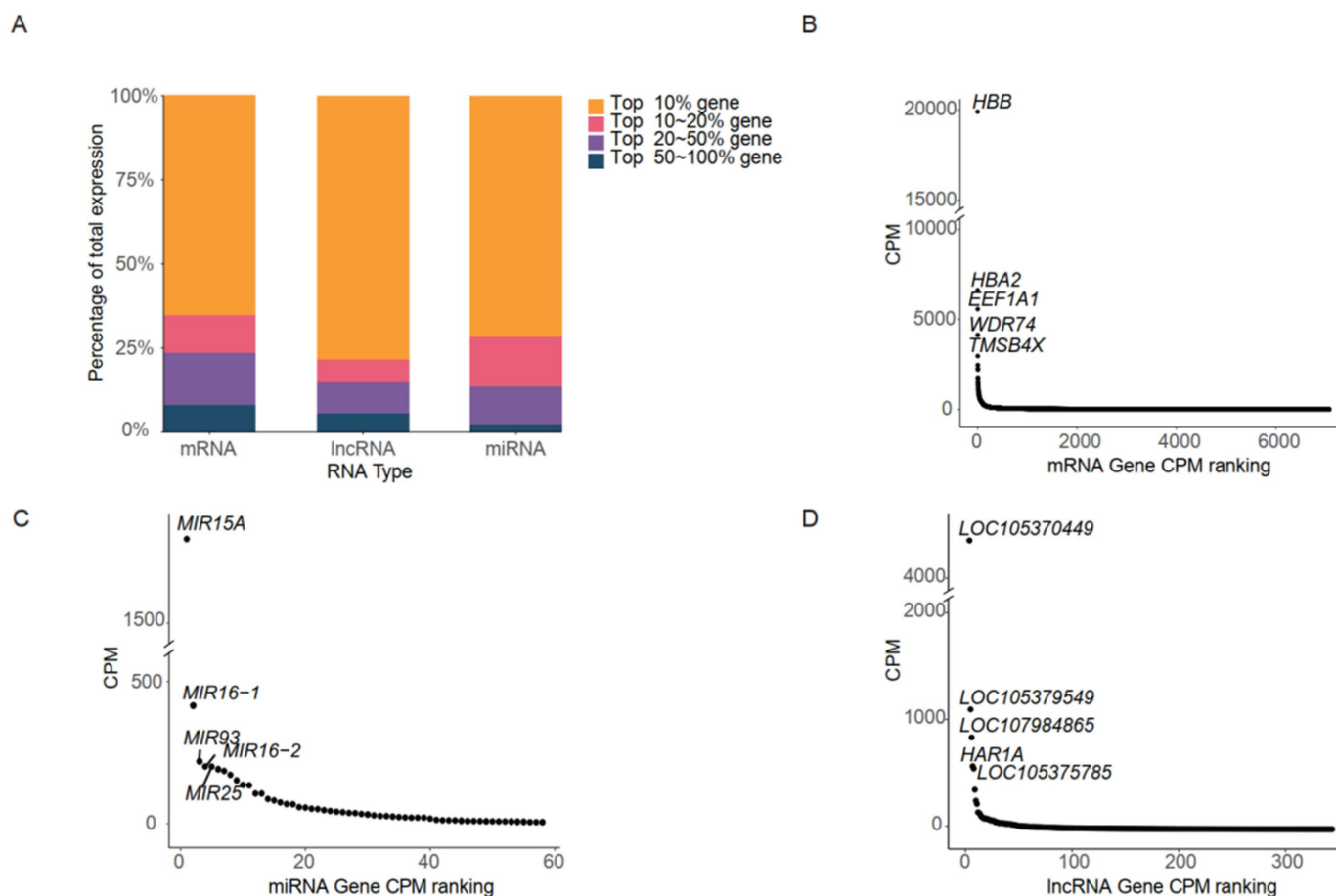


Figure 2. Relative abundance of cfRNA at the gene level. (A) the average cfRNA percentages of top-ranked (top 10%, 10–20%, 20–50%, 50–100%) genes in samples ($n = 66$). Only cfRNA genes presented in more than 50% of samples were selected for this analysis. The vast majority of the cfRNA was contributed by a few genes with high abundance levels. The abundance levels of mRNA, lncRNA and miRNA genes were ranked by the average abundance level of cfRNA read counts per million (CPM). Only genes with average cfRNA CPM larger than 5 were shown for (B) mRNA, (C) miRNA and (D) lncRNA. The names of the top five genes were labeled in the figures (B–D).

depth. In contrast, the sequencing depth of mt-mRNA and mt-tRNA varies among genes (Figure 3B). The gene body coverage curve of mt-mRNA was also relatively uniform (Figure 3C). In terms of CPM, the mt-mRNA and mt-tRNA abundance levels for all genes are shown in Figure 3D. As shown in Figure 3B,D some mt-tRNAs exhibited relatively low abundance levels and sequencing depth. This is due to their sequence similarity with some nuclear genomic regions, and these reads were filtered out during the data processing, to calculate the very specific mt-RNA fragments.

We then analysed the fragment size distribution of cell-free mt-RNA. In addition to the described mt-rRNA, we found that the size distribution of mt-mRNA is dispersed, with a higher proportion in the 130–160 nt range (Figure 3A). The size distribution profile of mt-mRNA is somewhat similar to that of cell-free mRNA (Figure 1F). The proportion of longer fragments decreases with increasing size. The proportion of mt-mRNA fragments exceeding 175 nt in size is 55.95%, with the longest fragments potentially exceeding 400 nt. For mt-tRNA, we were also able to capture relatively complete fragments, with mt-tRNA fragments concentrated in the size of 60–68 nt, which is close to the size of complete mt-tRNAs (58–74 nt) (Figure 3A).

Cell types of origins analysis revealed the impact of BMI and enabled the identification of abnormal individuals with some liver problems

To explore the cellular origins of cfRNA, we utilized a deconvolution method based on the Tabula Sapiens and cell-type-specific RNA [53]. As Figure 4A,B shows, most cfRNA is derived from Erythrocyte/erythroid progenitors, with an average proportion of 75.04%, followed by Platelet (4.72%), Salivary/bronchial secretory cell (3.05%), Monocyte (1.85%), B cell (1.14%) and Macrophage (1.13%). Immune cells such as Monocytes, B cells and Macrophages are important contributors to cfRNA, albeit at much lower proportions compared to Erythrocyte/erythroid progenitors and Platelets. This result of cfRNA origins is generally consistent with the findings in the previous study [53], but with a higher proportion of Erythrocyte/erythroid progenitors and a lower proportion of Platelets, which may be caused by preanalytical procedures, such as plasma preparation and sequencing library preparation bias.

Then, we explored the fragment size distribution of cfRNA with different tissue origins. Here, the cell-type-specific genes (1310 genes) are strictly defined as those exclusively expressed in specific cell types in the basis matrix based on the TSP

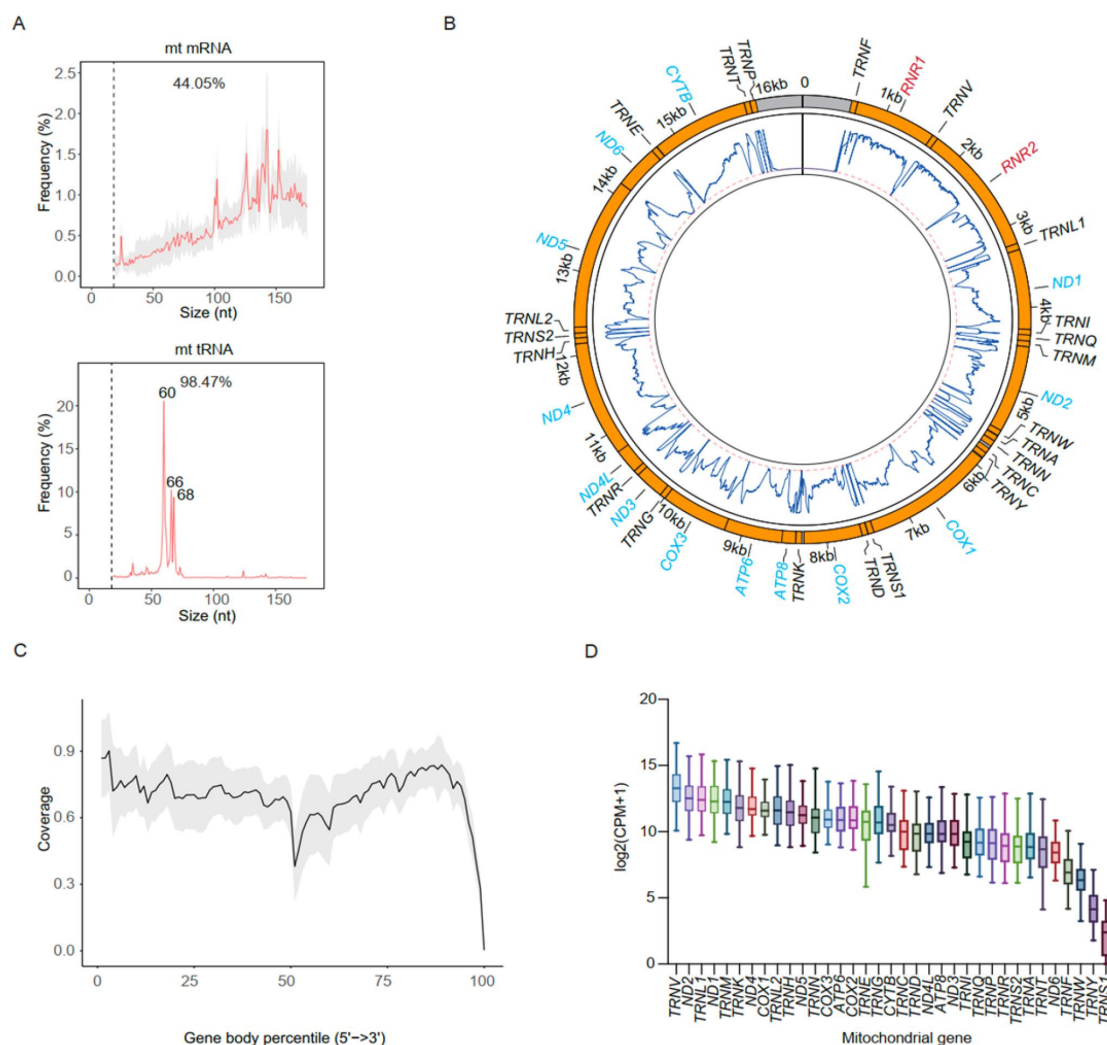


Figure 3. Mitochondrial RNA (mt-RNA) characteristics in plasma cfRNA. (A) size distribution of cell-free mitochondrial mRNA and tRNA. The red line represents the average size distribution frequency and the grey shadow indicates the standard deviation. The percentage numbers above the size profile indicate the proportion of cfRNA fragments shorter than 175nt. (B) CfRNA sequencing depth on the mitochondrial genome. The blue line indicates the average sequencing depth per base. Gene names were annotated with red (rRNA), blue (mRNA) and black (tRNA) colors. (C) gene body coverage of cell-free mitochondrial mRNA. The black line represents the average coverage, and the grey shadow indicates the standard deviation. (D) the abundance levels of all mitochondrial mRNA and tRNA genes.

database [53]. We selected cell-type-specific genes and grouped them into categories of blood, heart, kidney, liver and lung. We found that the detected blood-derived cfRNA was longer than organ-derived cfRNA (Figure 4C). For example, the liver-derived cfRNA fragments were significantly shorter and more narrowly distributed than blood-derived cfRNA. This result revealed that the cfRNA size was influenced by the tissue or cell-type origins. Such size influence of tissue origins was known in cfDNA [54].

By comparing the groups of people with different BMI, we found that the BMI ≤ 18.5 group ($n = 9$) had significantly higher proportions of Kidney epithelial cells than the other two BMI groups ($18.5 < \text{BMI} \leq 24$, BMI > 24) (Figure 4D). Notably, compared to other organ-specific cell types, the proportion of Kidney-epithelial-cell-derived cfRNA is the highest (Figure 4A). The kidney function biomarkers of serum creatinine and urea are significantly lower in the BMI ≤ 18.5 group than those with higher BMI (Supplementary Figure S2A, B). On one hand, the lower

levels of the two biomarkers may be because of diets low in protein and low muscle mass [55]. On the other hand, low creatinine levels may also indicate chronic kidney disease, kidney dysfunction or malnutrition [56,57]. Our results suggest that BMI has a significant impact on cfRNA origins and researchers would better consider this factor during cfRNA analysis.

Given that cfRNA is predominantly derived from blood cells, cfRNA derived from other organs and tissue cell types were very low in healthy people, and thus the background noise may interfere with the abundance analysis. For this reason, we focused on the more strictly defined cell-type-specific genes to investigate the organ-derived cfRNA, as mentioned above in the fragment size analysis. Specifically, 642 genes were detected in our plasma samples. These genes include those specific to Hepatocyte (62 genes), Ciliated cell/lung ciliated cell (53 genes), Cardiac muscle cell (37 genes), Kidney epithelial cell (31 genes), etc. The detected gene numbers of the top 15 cell types were shown (Figure 5A). These cell-type specific cfRNA

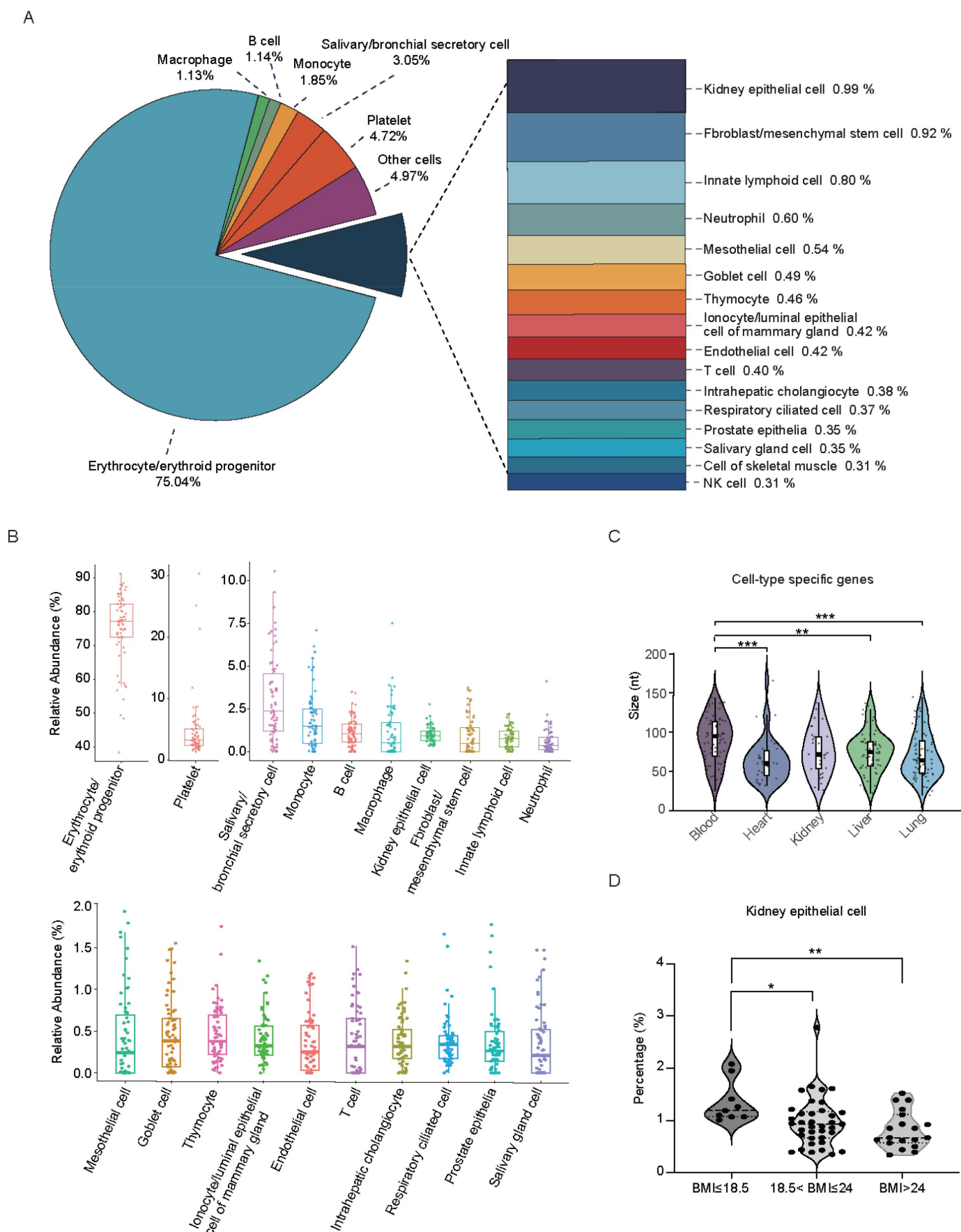


Figure 4. Cell-type decomposition of the plasma cfRNA samples from generally healthy people ($n = 66$). (A) the average fraction contributions of cell-type specific plasma cfRNA. (B) the boxplot of cfRNA fraction contributions in samples. (C) the fragment size distribution of cfRNA derived from different tissues. Only tissue/cell-type specific genes and cfRNA fragments shorter than 175 nt were analysed here. Each point in the violin plot signifies a specific gene. (D) the comparison of cfRNA derived from kidney epithelial cells between BMI groups.

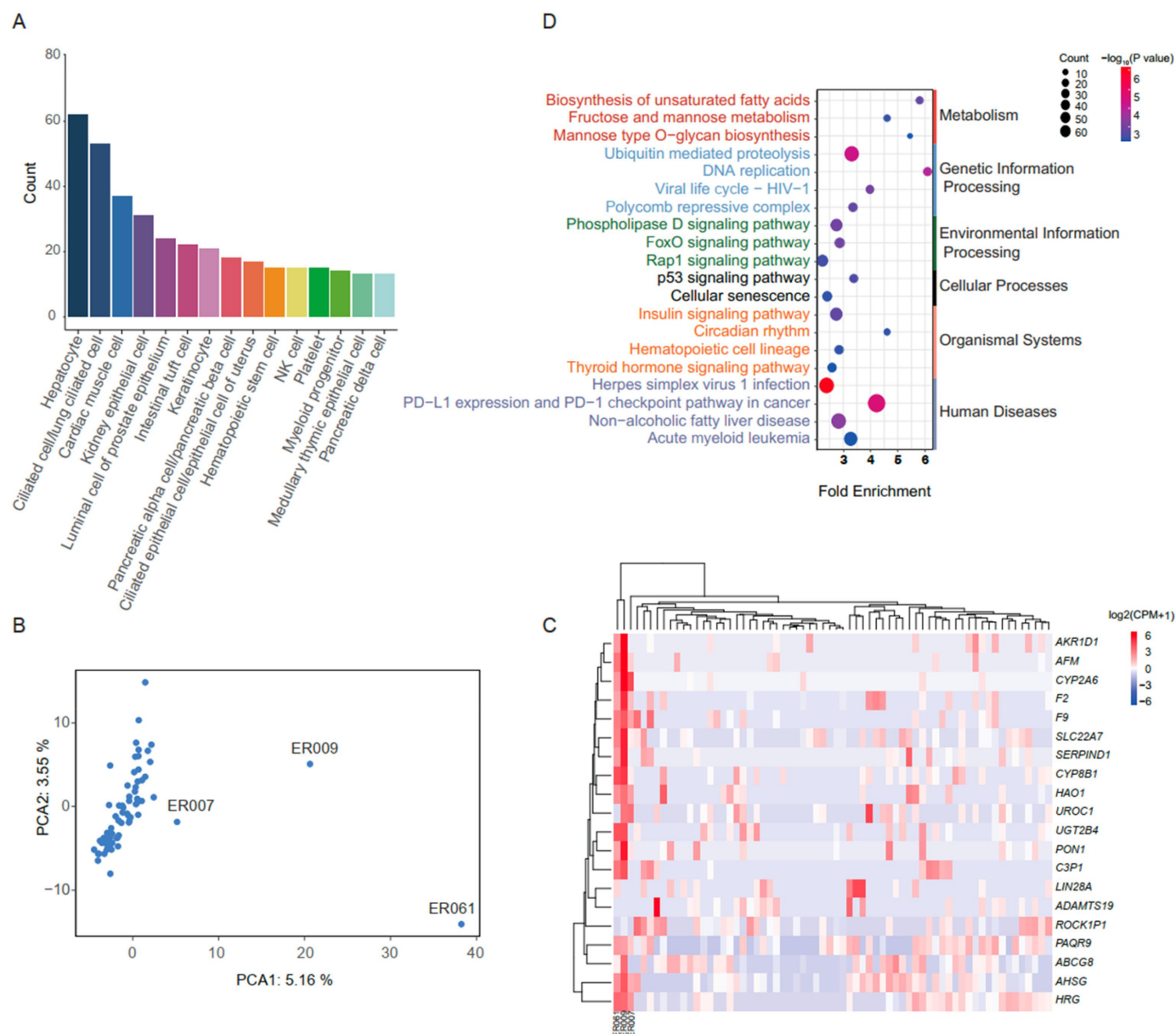


Figure 5. Cell-type specific and individual variance of plasma cfRNA. (A) the number of specific genes detected in the plasma cfRNA for different cell types. A total of 642 cell-type specific genes were detected in our data, and the top 15 cell types were shown. (B) a principal component analysis (PCA) of the cell-type-specific cfRNA of 66 samples. Three outlier samples were labeled. Detailed analysis revealed that their hepatocyte-specific gene abundance levels were abnormal compared to other samples. (C) a heatmap of the top 20 hepatocyte-specific cfRNA among samples. The three outliers had much higher levels of most of these genes compared to other samples. Validation through questionnaires and blood liver function results in physical examinations confirmed that these individuals exhibit abnormalities related to liver function. (D) The top 1000 cfRNA genes with the highest individual variations evaluated by coefficient of variation (CV) were used for the KEGG pathway enrichment analysis by Metascape. In Metascape, the overlapping genes of different pathways were subject to redundancy removal. For this reason, the Alcoholic liver disease pathway did not appear in the summarized figure. However, this pathway is significantly enriched (Padj = 0.0303) and was grouped into the pathway of PD-L1 expression and PD-1 checkpoint pathway in cancer (Supplementary Table S2).

may serve as non-invasive biomarkers and provide important clues for health assessments of organs.

To explore this idea, the cfRNA levels of the cell-type specific genes were utilized to perform principal component analysis (PCA). The PCA results showed three samples as outliers, denoted as ER007, ER009 and ER061 (Figure 5B). It was identified that these samples had much higher hepatocyte-derived cfRNA levels than those of other samples. The cfRNA levels of the top 20 hepatocyte-specific genes were shown in the heatmap (Figure 5C), and these hepatocyte-specific cfRNA levels clearly distinguished the three samples

from others. By checking the results of physical examination and questionnaires, we found that the individual of ER009 had blood liver function test results within the reference range, but this individual reported a chronic liver disease. Notably, this is the only participant in our study who had such a liver disease. Furthermore, the individual of ER061 was the only one with indirect bilirubin (IBIL) and total bilirubin (TBIL) levels exceeding the upper limit of the reference ranges, at 15.2 $\mu\text{mol/L}$ (reference range: 3.4–12.2 $\mu\text{mol/L}$) and 20.5 $\mu\text{mol/L}$ (reference range: 2–20.4 $\mu\text{mol/L}$), respectively. Additionally, among all the participants, only two had

gamma-glutamyl transferase (GGT) levels exceeding the upper limit of the reference range (reference range: 10–47 U/L). The individual of ER007 had the highest GGT level at 69.9 U/L. Meanwhile, his aspartate aminotransferase (AST) level was 60.4 U/L, exceeding the reference range of 0–40 U/L. The other participant with an abnormal GGT level had a value of 58.8 U/L, but the cfRNA expression did not show abnormality.

For liver function tests of the remaining participants, we found that 11 had alanine aminotransferase (ALT) and/or AST levels exceeding the reference ranges to some extent (Supplementary Table S3). For these individuals, their cfRNA profiles were not found to be different from other participants. Although ALT and AST are most abundant in the liver, they are also present in other organs, the leakage from other organs could also contribute to the higher levels in plasma [58]. To further explore the potential relationship, we used the total abundance of strictly defined hepatocyte-specific cfRNA to analyse the correlations with liver function indicators. However, still no significant correlation was found. Nevertheless, we discovered that the cfRNA level of *MIR483*, the only detected liver-specific miRNA reported by Guo et al. [59] in our data, was significantly higher in the group of people with biochemical indicators exceeding the normal ranges (Supplementary Figure S3B). Similarly, ER007, ER009 and ER061 also showed relatively high levels of this miRNA.

Then, we categorized the cfRNA levels of each gene into three groups (<60 nt, 60–120 nt, >120 nt) to analyse their correlations with blood biochemical parameters and evaluate fragment size effects. By filtering the fragment sizes of cfRNA, we identified that many cfRNA levels showed significant correlations with blood biochemical parameters (Supplementary Table S4). For instance, in the <60 nt cfRNA group, liver-specific mRNAs (genes: *ALB*, *APOC1*, *FGB* and *HP*), liver-specific miRNAs (genes: *MIR193B* and *MIR6777*) and liver-specific lncRNA (gene: *EGFR-AS1*) demonstrated significant positive correlations with biochemical parameters (Supplementary Figure S4). These miRNA and lncRNA were known to be related to liver diseases [59–62]. The Tissue Enrich analysis [60] showed that the identified genes in the cfRNA group (<60 nt) were significantly enriched in the liver ($P_{adj} < 0.001$), while the genes identified in the other two categories had no such enrichment ($P_{adj} > 0.05$). This result may be related to the discovery that liver-derived cfRNA was generally shorter than blood-derived cfRNA (Figure 4C). Our findings indicate that organ-derived cfRNA were more enriched in the short fraction, and the size-selection method might be helpful in organ health and disease studies. Many non-tissue-specific or under-characterized genes were also identified (Supplementary Table S4). These genes may be important to liver function as well and need to be studied in the future. In summary, cfRNA of multiple types reflects liver health and may be promising new liver biomarkers.

Individual cfRNA variances among generally healthy people are associated with common diseases

To understand the cfRNA variance among the participants, we calculated the coefficient of variation (CV) for the detected cfRNA levels and then selected the top 1,000 genes with the

highest CV for KEGG pathway enrichment analysis. As a result, the enriched pathways include those categories of metabolism, cellular processes, genetic and environmental information processing, organismal systems and human diseases. More specifically, some enriched pathways are associated with common diseases such as liver diseases, virus infections, cancers and metabolic diseases.

The significantly enriched pathways for all samples include liver-related pathways: Non-alcoholic fatty liver disease and Alcoholic liver disease. This enrichment result can be repeated after excluding the three abnormal samples (Figure 5D, Supplementary Table S2). There is a relatively high prevalence of liver diseases such as fatty liver and hepatitis B in China [61,62]. Our results displayed the common individual differences in liver status. Although only one participant had been diagnosed with liver disease in our study, individuals who have not met the diagnostic criteria may already have some gene expressions altered with an increased risk of liver diseases, which could be captured in our sequencing method.

The enriched pathways include Herpes simplex virus 1 (HSV-1) infection. HSV-1 infection is one of the most common viral infections [63], with the reported infection rate as high as 66.6% [64]. While most HSV-1 infections are very mild, sometimes they may lead to severe symptoms in the eyes and central nervous system, and even causing blindness, encephalitis and epilepsy [65–67]. Another virus-related pathway enriched in the analysis is Viral life cycle – HIV – 1. For cancer-related pathways, the PD-L1 expression and PD-1 checkpoint pathway in cancer play a key role in maintaining immune balance and in response to tumour cells [68]. Besides, the p53 signalling pathway is enriched, which is important in regulating the cell cycle, maintaining cellular homeostasis and suppressing tumour development. These enrichments may reflect the distinct baseline immune characteristics among people and the different susceptibility to viral infections and cancer development.

In the metabolic category, we identified enrichment in carbohydrate and lipid metabolic pathways including Fructose and Mannose metabolism, Mannose type O-glycan biosynthesis and Biosynthesis of unsaturated fatty acids. Additionally, the Insulin signalling pathway in another KEGG category is also enriched, which plays a vital role in maintaining blood sugar levels and energy balance. Our results demonstrate notable individual differences in metabolic pathways, which may be influenced by sex, age, genetics and lifestyles, and may also be related to the increasing metabolic disease burden in the general population. We showed that cfRNA might provide another perspective to assess the metabolism of individuals.

Discussion

In this study, we developed a new wide-spectrum cfRNA-sequencing method and comprehensively characterized plasma cfRNA compositions, size distributions, cell-type origins and individual differences in generally healthy individuals. The cell-free mRNA, miRNA, lncRNA and those cell-type specific cfRNA might provide useful information for health monitoring.

For future studies, further optimization of the experimental methods is possible. For instance, the use of cfRNA protection blood collection tubes or reagents may help to reduce the degradation of cfRNA. The plasma cfRNA concentrations were very low in healthy people, and it is difficult to quantify them accurately. There were notable differences among methods [69]. In our lab, the estimated plasma cfRNA concentrations in healthy people were mostly <150 ng/mL by using Qubit microRNA Assay Kit. We will further develop our sequencing-based method to accurately quantify cfRNA concentrations using synthetic RNA spike-in. To explore biomarkers, we suggest depleting the rRNA sequence before sequencing. Although the almost full spectrum of cfRNA is captured in our method, rRNA constitutes a large portion of the total reads and made the reads ratio of more informative cfRNA types relatively low. This will seriously affect the discovery of cfRNA biomarkers, especially for the cfRNA with very low abundances. Technically, it should be much better to deplete the rRNA sequence at the DNA level after the reverse transcription [70,71], for example, by target probe-based depletion or by CRISPR-Cas9 treatment [31,72]. Since the plasma cfRNA is highly fragmented and the level is very low, the commonly used rRNA removal methods for RNA samples from tissues and cells are not suitable for plasma cfRNA samples. Some other rRNA removal protocols, if involved in complex and time-consuming procedures, may easily cause RNA degradation and thus we do not recommend. Besides, the cost comparison of the extra experimental steps with the extra depth of sequencing is needed for researchers to choose their preferred methods [30].

Although many previous studies suggest that cfRNA fragments are typically less than 50 nt [25,26], our research supports the existence of relatively long cfRNA fragments in plasma [20,31]. These fragments are likely protected by RNA binding protein complexes and EVs [35]. Researchers have shown that the size distribution of cfDNA is an important fragmentomic feature to distinguish patients and healthy people [73]. Similarly, the size features and the compositions of various types of cfRNA may also represent valuable biological and pathological information, as the size effect of tissue origins demonstrated in our study. Any disturbance during cfRNA release and clearance would result in altered cfRNA profiles. Such a cfRNA sequencing method that captures multiple types of cfRNA and provides size information may be applied in future biomarker discovery studies. This RNA sequencing method might also be useful for some tissue or cellular RNA studies that are interested in the profiles of various RNA types (including mt-RNA and tRNA), the RNA fragment size and the end motif [26]. Moreover, this method is technically compatible with single-molecule long-read sequencing and could be modified and improved for a broader range of applications.

According to our results, the high-abundance mRNA in the plasma of healthy people is well-known and is closely related to the blood and basic cellular functions. In contrast, the high-abundance miRNA and lncRNA in plasma are not well studied. Some studies have demonstrated that these non-coding RNA are significantly changed in diseases, such as type 2 diabetes (*MIR15A*) [40], hepatocellular

carcinoma (*MIR16-1*) [41], lymphoma (*MIR93*) [42], colorectal cancer (*MIR16-2*) [43], oesophageal squamous cell carcinoma (*MIR25*) [44], pituitary adenomas (*LOC105375785*) [45], and small cell lung cancer, diffuse gliomas and multiple sclerosis (*HAR1A*) [46–48]. Their relatively high abundance in plasma will make them much easier to detect and the detection results more reliable compared to those with very low levels. Future research should pay more attention to these miRNA and lncRNA to explore their potential as biomarkers for health assessment and disease prediction, especially for generally healthy people.

There are various common diseases closely related to mitochondria dysfunction, including cardiovascular diseases, neurodegeneration, metabolic syndrome and cancer [74]. Some studies have also demonstrated that plasma cell-free mt-DNA can be used as an indicator and predictor of COVID-19 severity and mortality [49,75]. In cancers, the mutations of cell-free mt-DNA were found associated with cancer progression [76,77]. Dysregulated mt-RNA in tissues is closely related to cancer development and progression. One study discovered that cell-free mt-RNA could serve as a prognostic indicator of advanced prostate cancer [50]. Nevertheless, the lack of systematic analysis of mt-RNA has hampered the understanding of its role in diseases and hindered the discovery of related disease biomarkers of cell-free mt-RNA. In this study, we showed that cell-free RNA fragments of all mitochondrial genes can be effectively captured through high-throughput sequencing. This provides a new way to study mitochondrial-related diseases and to develop non-invasive biomarkers.

Moreover, various organ-derived cfRNA fragments were detected in the generally healthy participants, including those from hepatocytes, ciliated cells/lung ciliated cells, cardiomyocytes and kidney epithelial cells, although they are in relatively very low abundances. Interestingly, the cellular origin deconvolution results indicated that cfRNA derived from sex organs was very low, and no significant difference was found between males and females. However, by focusing on tissue-specific lncRNA of female reproductive organs, a significant difference was found (Supplementary Figure S5). Our findings suggest that cfRNA is a promising analyte for monitoring organ health status in the general population, as demonstrated in our study for indicating liver health. In addition, the previously discovered disease RNA biomarkers in tissues and cells might be rediscovered in plasma to show their potential as non-invasive cfRNA biomarkers. Tissue-specific mRNA, lncRNA and miRNA in plasma provide multiple perspectives to illustrate disease signals. Risk prediction models based on the combined cfRNA signatures may provide valuable information for health and disease management in the future.

Materials and methods

Blood sampling and cell-free RNA extraction

We recruited generally healthy participants during their physical examinations at a health examination centre in Shenzhen, China. A total of 34 males and 32 females were

enrolled. The exclusion criteria included pregnancy, lactation, recent surgery, fever symptoms, cancer and any other serious diseases. Additionally, a health questionnaire was utilized to collect basic information on gender, age, body mass index (BMI) and medical history. This study was approved by The Institutional Review Board on Bioethics and Biosafety of BGI (BGI-IRB 21,157-T2). Blood samples from the generally healthy participants were collected in EDTA tubes. The plasma samples were separated within 4 h after blood collection. The blood samples were centrifuged for 10 min at 1600 g at 4°C, and the supernatant was centrifuged again for another 10 min at 14 000 g at 4°C to collect the plasma samples. The plasma samples were stored at -80°C for less than 4 months before cfRNA extraction. The plasma samples were extracted using miRcute Serum/Plasma miRNA Isolation Kit (TIANGEN, DP503) according to the manufacturer's instructions.

Synthesis of designed RNA fragments

The RNA fragments were designed and synthesized by Sangon Biotech Co., Ltd. The 62 nt synthetic RNA sequence is 5'-NNNNNNUAAGAAGAGGAAUUGAACCCUGACUGAAAGCCUUAUUUCUCUUGUCCUNNNNNN-3', and the 25 nt synthetic RNA sequence is 5'-NNNNNNUAAGAA GAGGAAUNNNNNN-3'. In these sequences, 'N' represents any nucleotide (A, T, C or G), which allows for the splint ligation of adaptors with random nucleotides. The synthesized RNA samples were diluted to a concentration of 10 ng/μL. For each sample, 50 ng of the diluted synthetic RNA was used as input RNA for the test of the library preparation method.

Cell-free RNA library preparation and sequencing

- DNA digestion and T4 PNK treatment

For the removal of the remaining DNA and the end treatment by T4 PNK: 2 μL DNase I (New England Biolabs, M0303L), 2 μL 10× DNase I buffer, 0.5 μL T4 PNK (BGI, LS-EZ-E-000150) 1 μL ATP (New England Biolabs, P0756, 10 mm each), 0.5 μL RNase Inhibitor (THERMO, EO0382) and 14 μL cfRNA sample were mixed, and incubated at 37°C for 15 min and then at 95°C for 5 min. After the reaction, the tubes were immediately transferred to ice for 2 min.

- Poly(A) tailing, Reverse Transcription (RT) and Purification

The following reaction reagents were added to the samples for poly(A) tailing and reverse transcription: 1 μL BSA (MACGENE, PP001, 10 mg each), 10 μL 50% PEG8000 (Beyotime, R0056-2 ml), 2 μL dNTP Mix (Thermo Scientific, R0181, 10 mm each), 1 μL *E. coli* Poly(A) Polymerase (New England Biolabs, M0276L), 0.5 μL RNase Inhibitor (THERMO, EO0382), 1 μL HiScript III Reverse Transcriptase (VAZYME, R302-01), 6 μL 5 × HiScript III Buffer, 2 μL 10 μM reverse transcription primer (5'-TTTTTTTTTTTTTTTTTUVN-3') and 6.5 μL DEPC water (BBI, D1005-500). The reaction tubes were incubated at 37°C for 15 min, 50°C for 15 min and incubated at 80°C for 5 min. After the reaction, 100 μL DNA clean beads (VAZYME, N411-03) were added to purify cDNA.

- Deoxyuridylic acid Digestion

Deoxyuridylic acid digestion was used to remove the poly(T) sequence of cDNA. 5 μL 10× PNK reaction buffer (BGI, LS-EZ-E-000150), 2 μL Tris (PH 8.0) (SIGMA, T2694-100 ML, 440 mm each), 0.6 μL ET SSB (New England Biolabs, M2401S), 2 μL USER™ Enzyme (New England Biolabs, M5505L) and 20.4 μL DEPC water (BBI, D1005-500) were added to the reaction tubes. The reaction tubes were incubated at 37°C for 30 min and 95°C for 5 min, followed by immediate transfer to ice for 2 min.

- Adaptor ligation and purification

2 μL 5' adaptors (2 μM) and 2 μL 3' adaptors (2 μM) specifically tailored for the DNBSEQ platform were then added, mixed with 21 μL 2× adaptor Ligation Buffer, 5 μL T4 DNA ligase (Enzymatics, L6030-HC-L) and 2 μL Ligation enhancer Hexaamminecobalt(III) chloride (MERYER, M84231-25 G, 2.5 μM). DEPC water (BBI, D1005-500) was added to the volume of 82 μL. The reaction tubes were incubated at 23°C for 30 min. The ligation products were then purified with 160 μL of DNA Clean Beads (VAZYME, N411-03) according to the manufacturer's instructions.

- PCR amplification and purification

The purified ligation products were amplified by PCR. 2 μL PCR universal primers (40 μM) and 2 μL PCR barcode primers (40 μM) were added to the ligation product, along with 25 μL 2× Alpha High-Fidelity Ready Mix (BGI 1,000,004,283), in a total volume of 50 μL. PCR was performed with an initial denaturation at 98°C for 2 min, followed by 20 cycles of 98°C for 20 sec, 56°C for 1 min, and 72°C extension for 30 sec with a final extension of 72°C for 5 min. The PCR products were purified with 75 μL of DNA Clean Beads. The PCR products were circularized using the MGIEasy Rapid Circularization Module (MGI 1,000,005,282).

- Sequencing

Sequencing was performed using the DNBSEQ platform (MGI) to generate 100 bp pair-ended reads. The raw data generated for each sample were about 150 M paired-end reads.

Sequencing data processing

Data cleaning was performed using fastp [78] (v0.22.0) and seqtk [79] (1.3-r119-dirty) through the following steps (1): base correction, removal of sequencing adapters (fastp -c -adapter_sequence=AAGTCGGA -adapter_sequence_r2=AAGTCGGA) (2); trimming of the residual 5' poly(A) sequence of Read1 and the 3' poly(T) sequence of Read2 (3) discarding reads with more than 50% of the bases below a Phred quality score of 30 (fastp -qualified_quality_phred = 30 -unqualified_percent_limit = 50 -n_base_limit = 10) (4); discarding reads shorter than 17 bp

after adapter trimming and low-quality filtering (fastp – length_required = 17). Subsequently, clean reads were aligned to the rRNA reference sequence of the human reference genome (GRCh38.p14) in NCBI using HISAT2 [80] (v2.2.1) with parameters (-k 10 -rna-strandness RF -fr -dta -avoid-pseudogene -no-mixed -no-discordant). The remaining reads were aligned to the human reference genome (GRCh38.p14) in NCBI using the same parameters with HISAT2 (v2.2.1) for other RNA types. Paired reads uniquely aligned to positions were categorized and quantified using featureCounts [81] (v2.0.2) to generate a count table (genes × samples) with parameters (-t exon -g gene -B -C -p). Subsequently, all gene counts were normalized using Counts Per Million (CPM) to generate a cRNA abundance matrix. Quality control metrics such as exon ratios, strand specificity and other quality control metrics were analysed for uniquely aligned paired reads using the qualimap [32] (v2.2.2-dev) default parameter.

Size analysis of synthetic RNA fragments

The FASTQ files were aligned to the designed synthetic RNA sequences. The reads ratio was calculated by counting the number of reads containing the target sequences with one nucleotide mismatch allowed.

cRNA fragment characterization and gene body coverage

We utilized the fastp software (v0.22.0) to merge the paired-end reads, with the merging criteria being an overlap of at least 10 bases and no more than two mismatched bases within the overlapping region (parameters set to -overlap_len_require = 10 -overlap_diff_limit = 2). The length of the insert size was obtained based on the paired-end reads compared to the human reference genome. The distribution of merged reads and insert lengths was visualized using the ggplot2 package in R. GeneBody coverage profiles were produced using GeneBody_coverage.py [82] with default flag according to different RNA type regions.

Analysis of high cRNA level genes

We filtered the genes based on the following criteria: only genes detected in at least 50% of the samples and with an average CPM of more than 5 were considered. The mRNA, lncRNA and miRNA genes were ranked by their CPM for visualization. We used the ggplot2 package in R to visualize cRNA levels for mRNA, lncRNA and miRNA genes.

Cellular origin deconvolution and tissue/cell-type specific cRNA analysis

We used the support vector regression deconvolution method [83] proposed by Sevahn K. Vorperian et al. [53] to perform cell type-specific analysis on the cRNA samples. This method uses the Tabula Sapiens 1.0 (TSP, <https://tabula-sapiens-portal.ds.czbiohub.org>) database (a comprehensive cell atlas covering 24 different tissues and organs from multiple donors) and

defines a basis matrix that includes a minimum discriminatory gene set to distinguish different cell types [84]. Using the code and basis matrix provided by Sevahn et al. (https://github.com/sevahn/deconvolution/tree/master/deconvolve_crna_tutorial), we performed Support Vector Regression (SVR) deconvolution analysis on the CPM (Counts Per Million reads) matrix of our cRNA samples. Through this analysis, we successfully quantified the contributions of various cell types. From the basis matrix, 1310 genes that were exclusively expressed in specific cell types were selected (with basis matrix values represented as CPM + 1, and a value greater than 1 in only one cell type was considered as specifically expressed in that cell type) and defined here as cell-type-specific genes. In our samples, we detected a total of 642 genes that are defined as cell-type-specific genes.

In this study, we identified liver-specific miRNAs based on the report of Guo et al. [59]. We obtained tissue-specific lncRNAs from the miRNA Tissue Atlas 2025 database [85], using the following criteria: a Tissue Specificity Index (TSI) of 0.8 or higher.

In another way, TissueEnrich (<https://tissueenrich.gdcdb.ias.tate.edu/>) was used for the tissue-specific gene enrichment analysis based on the dataset of Human Protein Atlas [60].

Analysis of individual variances

We performed principal component analysis (PCA) using the FactoMineR package in R based on the cRNA abundance of 642 cell-type-specific genes in 66 samples. Additionally, we used the heatmap.2 function from the R gplots package to create heatmaps that display the cRNA abundance levels of the top 20 hepatocyte-specific genes in the 66 samples.

We calculated the Coefficient of Variation (CV) values for cRNA gene abundance levels (CPM) in the 66 samples. The cRNA genes in more than 50% of the samples and with an average CPM more than 5 were analysed here. The top 1000 genes with the highest CV values in abundance and conducted KEGG pathway analysis on these genes using the Metascape website (<http://metascape.org/>). Metascape addresses redundancies in descriptors and ontologies by removing overlapping terms. As a result, some pathways with overlapping genes may not be displayed in the summarized figure [86].

Statistical analysis

We used the Mann–Whitney U and Dunn's test to compare different groups. The levels of significance were determined as follows: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. In the analysis of liver function tests, the participants were divided into 'High' and 'Normal' groups. The 'High' group included participants with liver function test results exceeding the reference ranges and ER009 who had liver diseases. The remaining participants were in the 'Normal' group.

We correlated the size-related CPM values to the blood biochemical test results with Pearson correlation. The p values were adjusted by the Benjamini–Hochberg method. The cRNA levels of each gene (CPM) were divided into three values according to the fragment size categories (<60 nt, 60–120nt and >120nt).

Acknowledgment

We sincerely thank the support provided by China National GenBank (CNCB).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work has been supported by the National Natural Science Foundation of China [Nos. 32171441 and 32000398], National Key R&D Program of China [2022YFC2502402] and Natural Science Foundation of Guangdong Province, China [2017A030306026].

Author contributions

Yan Zhang: Conceptualization, Supervision, Methodology, Project administration, Writing-original draft, Writing-review & editing. Xin Jin: Conceptualization, Supervision, Resources, Writing-review & editing. Haiqiang Zhang: Supervision, Resources, Methodology, Writing-review & editing. Xinxin Wang: Investigation, Formal analysis, Visualization, Data curation, Project administration, Writing-original draft, Writing-review & editing. Shaogang Li: Formal analysis, Methodology, Visualization, Software, Data curation, Validation, Writing-original draft. Rijng Ou: Investigation, Methodology, Validation, Writing-original draft. Wending pang: Formal analysis, Visualization. Yingying Wang, Yifan Zhang, Yu Lin and Changgui Lei: Software. Changlin Yang, Wei Chen, Guodan Zeng, Wenwen Zhou and Yeqing Wang: Investigation. Jianhua Yin: Resources.

Data availability statement

The data supporting this study have been submitted to CNCB Nucleotide Sequence Archive (CNSA) of China National GeneBank DataBase (CNCBdb) with accession number CNP0005451 (<https://db.cncb.org/cnsa/>).

References

- [1] Pös Z, Pös O, Styk J, et al. Technical and methodological aspects of cell-free nucleic acids analyzes. *Int J Mol Sci.* 2020;21(22):8634. doi: [10.3390/ijms21228634](#)
- [2] Loy C, Ahmann L, De Vlaminc I, et al. Liquid biopsy based on cell-free DNA and RNA. *Annu Rev Biomed Eng.* 2024;26(1):169–195. doi: [10.1146/annurev-bioeng-110222-111259](#)
- [3] Jin X, Wang Y, Xu J, et al. Plasma cell-free DNA promise monitoring and tissue injury assessment of COVID-19. *Mol Genet Geno.* 2023;298(4):823–836. doi: [10.1007/s00438-023-02014-4](#)
- [4] Qi T, Pan M, Shi H, et al. Cell-free DNA fragmentomics: the novel promising biomarker. *Int J Mol Sci.* 2023;24(2):1503. doi: [10.3390/ijms24021503](#)
- [5] Zhong P, Bai L, Hong M, et al. A comprehensive review on circulating cfRNA in plasma: implications for disease diagnosis and beyond. *Diagnostics.* 2024;14(10):1045. doi: [10.3390/diagnostics14101045](#)
- [6] Toden S, Zhuang J, Acosta AD, et al. Noninvasive characterization of Alzheimer's disease by circulating, cell-free messenger RNA next-generation sequencing. *Sci Adv.* 2020;6(50):eabb1654. doi: [10.1126/sciadv.abb1654](#)
- [7] De Souza MF, Kuasne H, Barros-Filho MDC, et al. Circulating mRNA signature as a marker for high-risk prostate cancer. *Carcinogenesis.* 2020;41(2):139–145. doi: [10.1093/carcin/bgz129](#)
- [8] Rasmussen M, Reddy M, Nolan R, et al. RNA profiles reveal signatures of future health and disease in pregnancy. *Nature.* 2022;601(7893):422–427. doi: [10.1038/s41586-021-04249-w](#)
- [9] Wang Y, Li J, Zhang L, et al. Plasma cell-free RNA characteristics in COVID-19 patients. *Genome Res.* 2022;32(2):228–241. doi: [10.1101/gr.276175.121](#)
- [10] Moufarrej MN, Vorperian SK, Wong RJ, et al. Early prediction of preeclampsia in pregnancy with cell-free RNA. *Nature.* 2022;602(7898):689–694. doi: [10.1038/s41586-022-04410-z](#)
- [11] Ngo TT, Moufarrej MN, Rasmussen M-LH, et al. Noninvasive blood tests for fetal development predict gestational age and pre-term delivery. *Medicine (Baltimore).* 2018;360(6393):1133–1136. doi: [10.1126/science.aar3819](#)
- [12] Gahlawat AW, Witte T, Sinn P, et al. Circulating cf-miRNA as a more appropriate surrogate liquid biopsy marker than cfDNA for ovarian cancer. *Sci Rep.* 2023;13(1):5503. doi: [10.1038/s41598-023-32243-x](#)
- [13] Gahlawat AW, Fahed L, Witte T, et al. Total circulating microRNA level as an independent prognostic marker for risk stratification in breast cancer. *Br J Cancer.* 2022;127(1):156–162. doi: [10.1038/s41416-022-01756-z](#)
- [14] Schwarzenbach H, Nishida N, Calin GA, et al. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat Rev Clin Oncol.* 2014;11(3):145–156. doi: [10.1038/nrclinonc.2014.5](#)
- [15] Hannan NJ, Stock O, Spencer R, et al. Circulating mRNAs are differentially expressed in pregnancies with severe placental insufficiency and at high risk of stillbirth. *BMC Med.* 2020;18(1):1–16. doi: [10.1186/s12916-020-01605-x](#)
- [16] Munchel S, Rohrbach S, Randise-Hinchliff C, et al. Circulating transcripts in maternal blood reflect a molecular signature of early-onset preeclampsia. *Sci Transl Med.* 2020;12(550):eaaz0131. doi: [10.1126/scitranslmed.aaz0131](#)
- [17] Yao J, Wu DC, Nottingham RM, et al. Identification of protein-protected mRNA fragments and structured excised intron RNAs in human plasma by tigr-seq peak calling. *Elife.* 2020;9:e60743. doi: [10.7554/eLife.60743](#)
- [18] Chen S, Jin Y, Wang S, et al. Cancer type classification using plasma cell-free RNAs derived from human and microbes. *Elife.* 2022;11:e75181. doi: [10.7554/eLife.75181](#)
- [19] Ibarra A, Zhuang J, Zhao Y, et al. Non-invasive characterization of human bone marrow stimulation and reconstitution by cell-free messenger RNA sequencing. *Nat Commun.* 2020;11(1):400. doi: [10.1038/s41467-019-14253-4](#)
- [20] Reggiardo RE, Maroli SV, Peddu V, et al. Profiling of repetitive RNA sequences in the blood plasma of patients with cancer. *Nat Biomed Eng.* 2023;7(12):1627–1635. doi: [10.1038/s41551-023-01081-7](#)
- [21] Pös O, Biró O, Szemes T, et al. Circulating cell-free nucleic acids: characteristics and applications. *Eur J Hum Genet.* 2018;26(7):937–945. doi: [10.1038/s41431-018-0132-4](#)
- [22] Ma F, Fuqua BK, Hasin Y, et al. A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics.* 2019;20(1):1–12. doi: [10.1186/s12864-018-5393-3](#)
- [23] Heinicke F, Zhong X, Zucknick M, et al. Systematic assessment of commercially available low-input miRNA library preparation kits. *RNA Biol.* 2020;17(1):75–86. doi: [10.1080/15476286.2019.1667741](#)
- [24] Giraldez MD, Spengler RM, Etheridge A, et al. Phospho-RNA-seq: a modified small RNA-seq method that reveals circulating mRNA and lncRNA fragments as potential biomarkers in human plasma. *Embo J.* 2019;38(11):e101695. doi: [10.15252/embj.2019101695](#)
- [25] Wang J, Huang J, Hu Y, et al. Terminal modifications independent cell-free RNA sequencing enables sensitive early cancer detection and classification. *Nat Commun.* 2024;15(1):156. doi: [10.1038/s41467-023-44461-y](#)
- [26] Liu Z, Wang T, Yang X, et al. Polyadenylation ligation-mediated sequencing (palm-Seq) characterizes cell-free coding and non-coding RNAs in human biofluids. *Clin Transl Med.* 2022;12(7):e987. doi: [10.1002/ctm2.987](#)
- [27] Zhou S, Li J, Yang W, et al. Noninvasive preeclampsia prediction using plasma cell-free RNA signatures. *Am J Obstet Gynecol.* 2023;229(5):e553.1–e553.16. doi: [10.1016/j.ajog.2023.05.015](#)

- [28] Troll CJ, Kapp J, Rao V, et al. A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genomics*. 2019;20(1):1–14. doi: [10.1186/s12864-019-6355-0](#)
- [29] Raine A, Manlig E, Wahlberg P, et al. Splinted ligation adapter tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Res*. 2017;45(6):e36–e36. doi: [10.1093/nar/gkw1110](#)
- [30] Pan W, Ngo TT, Camunas-Soler J, et al. Simultaneously monitoring immune response and microbial infections during pregnancy through plasma cfRNA sequencing. *Clin Chem*. 2017;63(11):1695–1704. doi: [10.1373/clinchem.2017.273888](#)
- [31] Larson MH, Pan W, Kim HJ, et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat Commun*. 2021;12(1):2357. doi: [10.1038/s41467-021-22444-1](#)
- [32] Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32(2):292–294. doi: [10.1093/bioinformatics/btv566](#)
- [33] Naphade S, Bhatnagar R, Hanson-Smith V, et al. Systematic comparative analysis of strand-specific RNA-seq library preparation methods for low input samples. *Sci Rep*. 2022;12(1):1789. doi: [10.1038/s41598-021-04583-z](#)
- [34] Smith B, Treadwell J, Zhang D, et al. Large-scale expression analysis reveals distinct microRNA profiles at different stages of human neurodevelopment. *PLOS ONE*. 2010;5(6):e11109. doi: [10.1371/journal.pone.0011109](#)
- [35] Kim HJ, Rames MJ, Goncalves F, et al. Selective enrichment of plasma cell-free messenger RNA in cancer-associated extracellular vesicles. *Commun Biol*. 2023;6(1):885. doi: [10.1038/s42003-023-05232-z](#)
- [36] Hardison RC. Evolution of hemoglobin and its genes. *Cold Spring Harb Perspect Med*. 2012;2(12):a011627. doi: [10.1101/cshperspect.a011627](#)
- [37] Joung EK, Kim J, Yoon N, et al. Expression of *EEF1A1* is associated with prognosis of patients with colon adenocarcinoma. *J Clin Med*. 2019;8:1903. doi: [10.3390/jcm8111903](#)
- [38] Hiraishi N, Ishida Y-I, Sudo H, et al. WDR74 participates in an early cleavage of the pre-rRNA processing pathway in cooperation with the nucleolar AAA-ATPase NVL2. *Biochem Biophys Res Commun*. 2018;495(1):116–123. doi: [10.1016/j.bbrc.2017.10.148](#)
- [39] Yang Z, Luo J, Zhang M, et al. *TMSB4X*: A novel prognostic marker for non-small cell lung cancer. *Heliyon*. 2023;9(11):e21505. doi: [10.1016/j.heliyon.2023.e21505](#)
- [40] Sadeghzadeh S, Dehghani Ashkezari M, Seifati SM, et al. Circulating miR-15a and miR-222 as potential biomarkers of type 2 diabetes. *Diabetes, Metabolic Syndr Obes*. 2020;13:3461–3469. doi: [10.2147/DMSO.S263883](#)
- [41] Wei H, Luo S, Bi Y, et al. Plasma *microRNA-15a/16-1*-based machine learning for early detection of hepatitis B virus-related hepatocellular carcinoma. *Liver Res*. 2024;8(2):105–117. doi: [10.1016/j.livres.2024.05.003](#)
- [42] Khare D, Goldschmidt N, Bardugo A, et al. Plasma microRNA profiling: exploring better biomarkers for lymphoma surveillance. *PLOS ONE*. 2017;12(11):e0187722. doi: [10.1371/journal.pone.0187722](#)
- [43] Ristau J, Staffa J, Schrotz-King P, et al. Suitability of circulating miRNAs as potential prognostic markers in colorectal cancer. *Biomarker Prev*. 2014;23:2632–2637. doi: [10.1038/bjc.2014.451](#)
- [44] Komatsu S, Ichikawa D, Hirajima S, et al. Plasma microRNA profiles: identification of *miR-25* as a novel diagnostic and monitoring biomarker in oesophageal squamous cell carcinoma. *Br J Cancer*. 2014;111:1614–1624. doi: [10.1038/bjc.2014.451](#)
- [45] Peng C, Wang S, Yu J, et al. lncRNA-mRNA expression patterns in invasive pituitary adenomas: a microarray analysis. *Biomed Res Int*. 2022;2022(1):1380485. doi: [10.1155/2022/1380485](#)
- [46] Zou H, Wu L-X, Yang Y, et al. lncRNAs *PVT1* and *HARIA* are prognosis biomarkers and indicate therapy outcome for diffuse glioma patients. *Oncotarget*. 2017;8(45):78767. doi: [10.18632/oncotarget.20226](#)
- [47] Ma J, Cao K, Ling X, et al. lncRNA *HARIA* suppresses the development of non-small cell lung cancer by inactivating the STAT3 pathway. *Cancers (Basel)*. 2022;14(12):2845. doi: [10.3390/cancers14122845](#)
- [48] Akbarzadeh S, Tayefeh-Gholami S, Najari P, et al. The expression profile of *HARIA* and *HARIB* in the peripheral blood cells of multiple sclerosis patients. *Mol Biol Rep*. 2023;50(3):2391–2398. doi: [10.1007/s11033-022-08182-7](#)
- [49] Valdes-Aguayo JJ, Garza-Veloz I, Badillo-Almaraz JJ, et al. Mitochondria and mitochondrial DNA: key elements in the pathogenesis and exacerbation of the inflammatory state caused by COVID-19. *Cirurgia*. 2021;57(9):928. doi: [10.3390/medicina57090928](#)
- [50] Mehra N, Penning M, Maas J, et al. Circulating mitochondrial nucleic acids have prognostic value for survival in patients with advanced prostate cancer. *Clin Cancer Res*. 2007;13:421–426. doi: [10.1158/1078-0432.CCR-06-1087](#)
- [51] Reznik E, Wang Q, La K, et al. Mitochondrial respiratory gene expression is suppressed in many cancers. *Elife*. 2017;6:e21592. doi: [10.7554/eLife.21592](#)
- [52] Xu L, Ziegelbauer J, Wang R, et al. Distinct profiles for mitochondrial t-RNAs and small nucleolar RNAs in locally invasive and metastatic colorectal cancer. *Clin Cancer Res*. 2016;22(3):773–784. doi: [10.1158/1078-0432.CCR-15-0737](#)
- [53] Vorperian SK, Moufarrej MN, Quake SR. Cell types of origin of the cell-free transcriptome. *Nat Biotechnol*. 2022;40(6):855–861. doi: [10.1038/s41587-021-01188-9](#)
- [54] Chan KA, Zhang J, Hui AB, et al. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem*. 2004;50(1):88–92. doi: [10.1373/clinchem.2003.024893](#)
- [55] Salazar JH. Overview of urea and creatinine. *Lab Med*. 2014;45(1):e19–e20. doi: [10.1309/LM920SBNZPJRGUT](#)
- [56] Kamal A. Estimation of blood urea (BUN) and serum creatinine level in patients of renal disorder. *Indian J Fundamental Appl Life Sci*. 2014;4:199–202.
- [57] Stenvinkel P, Heimbürger O, Paultre F, et al. Strong association between malnutrition, inflammation, and atherosclerosis in chronic renal failure. *Kidney Int*. 1999;55(5):1899–1911. doi: [10.1046/j.1523-1755.1999.00422.x](#)
- [58] Kwo PY, Cohen SM, Lim JK. ACG clinical guideline: evaluation of abnormal liver chemistries. *Am J Gastroenterol*. 2017;112(1):18–35. doi: [10.1038/ajg.2016.517](#)
- [59] Guo Z, Maki M, Ding R, et al. Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Sci Rep*. 2014;4:5150. doi: [10.1038/srep05150](#)
- [60] Jain A, Tuteja G, Kelso J. TissueEnrich: tissue-specific gene enrichment analysis. *Bioinformatics*. 2019;35(11):1966–1967. doi: [10.1093/bioinformatics/bty890](#)
- [61] Zhou F, Zhou J, Wang W, et al. Unexpected rapid increase in the burden of NAFLD in China from 2008 to 2018: a systematic review and meta-analysis. *Hepatology*. 2019;70(4):1119–1133. doi: [10.1002/hep.30702](#)
- [62] Wang H, Men P, Xiao Y, et al. Hepatitis B infection in the general population of China: a systematic review and meta-analysis. *BMC Infect Dis*. 2019;19(1):1–10. doi: [10.1186/s12879-019-4428-y](#)
- [63] Li W, Wang X-H, Luo Z, et al. Traditional Chinese medicine as a potential source for HSV-1 therapy by acting on virus or the susceptibility of host. *Int J Mol Sci*. 2018;19(10):3266. doi: [10.3390/ijms19103266](#)
- [64] James C, Harfouche M, Welton NJ, et al. Herpes simplex virus: global infection prevalence and incidence estimates, 2016. *Bull World Health Organ*. 2020;98(5):315. doi: [10.2471/BLT.19.237149](#)
- [65] Bradshaw MJ, Venkatesan A. Herpes simplex virus-1 encephalitis in adults: pathophysiology, diagnosis, and management. *Neurotherapeutics*. 2016;13(3):493–508. doi: [10.1007/s13311-016-0433-7](#)

- [66] Wu D, Wang C, Pang P, et al. The association between herpes simplex virus type 1 infection and Alzheimer's disease. *J Clin Neurosci*. 2020;82:63–70. doi: [10.1016/j.jocn.2020.10.044](https://doi.org/10.1016/j.jocn.2020.10.044)
- [67] Liesegang TJ. Herpes simplex virus epidemiology and ocular importance. *Cornea*. 2001;20(1):1–13. doi: [10.1097/00003226-200101000-00001](https://doi.org/10.1097/00003226-200101000-00001)
- [68] Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res*. 2020;10:727. doi: [10.7150/ijbs.42935](https://doi.org/10.7150/ijbs.42935)
- [69] Garcia-Elias A, Alloza L, Puigdecane E, et al. Defining quantification methods and optimizing protocols for microarray hybridization of circulating microRNAs. *Sci Rep*. 2017;7(1):7725. doi: [10.1038/s41598-017-08134-3](https://doi.org/10.1038/s41598-017-08134-3)
- [70] Isakova A, Neff N, Quake SR. Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. *Proc Natl Acad Sci USA*. 2021;118(51):e2113568118. doi: [10.1073/pnas.2113568118](https://doi.org/10.1073/pnas.2113568118)
- [71] Gu W, Crawford ED, O'Donovan B, et al. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol*. 2016;17(1):1–13. doi: [10.1186/s13059-016-0904-5](https://doi.org/10.1186/s13059-016-0904-5)
- [72] Wang H, Zhan Q, Ning M, et al. Depletion-assisted multiplexed cell-free RNA sequencing reveals distinct human and microbial signatures in plasma versus extracellular vesicles. *Clin Transl Med*. 2024;14(7):e1760. doi: [10.1002/ctm2.1760](https://doi.org/10.1002/ctm2.1760)
- [73] Ding SC, Lo YD. Cell-free DNA fragmentomics in liquid biopsy. *Diagnostics*. 2022;12(4):978. doi: [10.3390/diagnostics12040978](https://doi.org/10.3390/diagnostics12040978)
- [74] Zong Y, Li H, Liao P, et al. Mitochondrial dysfunction: mechanisms and advances in therapy. *STTT*. 2024;9(1):124. doi: [10.1038/s41392-024-01839-8](https://doi.org/10.1038/s41392-024-01839-8)
- [75] Scozzi D, Cano M, Ma L, et al. Circulating mitochondrial DNA is an early indicator of severe illness and mortality from COVID-19. *JCI Insight*. 2021;6:e143299. doi: [10.1172/jci.insight.143299](https://doi.org/10.1172/jci.insight.143299)
- [76] Xia P, An H-X, Dang C-X, et al. Decreased mitochondrial DNA content in blood samples of patients with stage I breast cancer. *BMC Cancer*. 2009;9(1):1–7. doi: [10.1186/1471-2407-9-454](https://doi.org/10.1186/1471-2407-9-454)
- [77] Li LH, Kang T, Chen L, et al. Detection of mitochondrial DNA mutations by high-throughput sequencing in the blood of breast cancer patients. *Int J Mol Med*. 2014;33:77–82. doi: [10.3892/ijmm.2013.1559](https://doi.org/10.3892/ijmm.2013.1559)
- [78] Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–i890. doi: [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560)
- [79] Shen W, Le S, Li Y, et al. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS ONE*. 2016;11(10):e0163962. doi: [10.1371/journal.pone.0163962](https://doi.org/10.1371/journal.pone.0163962)
- [80] Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and hisat-genotype. *Nat Biotechnol*. 2019;37(8):907–915. doi: [10.1038/s41587-019-0201-4](https://doi.org/10.1038/s41587-019-0201-4)
- [81] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–930. doi: [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656)
- [82] Wang L, Nie J, Sicotte H, et al. Measure transcript integrity using RNA-seq data. *BMC Bioinformatics*. 2016;17(1):1–16. doi: [10.1186/s12859-016-0922-z](https://doi.org/10.1186/s12859-016-0922-z)
- [83] Awad M, Khanna R. Efficient learning machines: theories, concepts, and applications for engineers and system designers. London: Springer Nature. 2015:67–80.
- [84] Consortium*, T.T.S.Jones RC, Karkanas J, et al. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Medicine (Baltimore)*. 2022;376:eabl4896. doi: [10.1126/science.abl4896](https://doi.org/10.1126/science.abl4896)
- [85] Rishik S, Hirsch P, Grandke F, et al. miRnatisueatlas 2025: an update to the uniformly processed and annotated human and mouse non-coding RNA tissue atlas. *Nucleic Acids Res*. 2025;53(D1):D129–D137. doi: [10.1093/nar/gkae1036](https://doi.org/10.1093/nar/gkae1036)
- [86] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. doi: [10.1038/s41467-019-09234-6](https://doi.org/10.1038/s41467-019-09234-6)