Identification of Globin Switching-Related Genes and RNAs via Transcriptomic Profiling of Nucleated Red Blood Cells Isolated from the Cord Blood of Preterm and Full-Term Newborns

Yuanyuan Han
Guizhou University https://orcid.org/0000-0003-1507-0506
Ling Huang
Guizhou Provincial People's Hospital
Man Zhou
Guizhou Provincial People's Hospital
Shangjin Gong
Beijing Institute of Genomics Chinese Academy of Sciences
Zhaojun Zhang
Beijing Institute of Genomics Chinese Academy of Sciences
Tingting Jin
Guizhou University
Juan Liu
Guizhou University
Li Hu
Guizhou University
Wenqiu Zhang
Guizhou University
Xiangdong Fang
Beijing Institute of Genomics Chinese Academy of Sciences
Yankai Jia
GENEWIZ suzhou
Shengwen Huang (hsw713@sina.com)
Guizhou University, Guizhou Provincial People's Hospital https://orcid.org/0000-0003-0394-240X

Research article

Keywords: RNA sequencing, transcriptome, nucleated red blood cell, umbilical cord blood

Posted Date: September 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-45443/v1

Abstract

Background: β-thalassemia and sickle cell disease (SCD) are chronic hemolytic anemias caused by mutation or deletion of the β-globin gene. Reactivation of fetal γ-globin gene (*HBG*) expression can ameliorate the clinical course of β-hemoglobinopathies. Our current understanding of the globin switching mechanism is not comprehensive, and research into the roles of long non-coding RNAs (lncRNAs) in this process remains limited.

Results: We collected umbilical cord blood samples from preterm and full-term infants, from which we enriched and isolated nucleated red blood cells (NRBCs) for transcriptome sequencing. We identified mRNAs, lncRNAs, and microRNAs that were differentially expressed in the preterm and full-term samples. Gene Ontology analysis and Kyoto Encyclopedia of Genes and Genomes analysis revealed the biological significance of differentially expressed mRNAs and the targets of differentially expressed lncRNAs and microRNAs. Our analyses revealed that the Fanconi anemia pathway may be involved in the regulation of globin switching at birth. Using quantitative polymerase chain reaction, we confirm that *BCL11A* is more highly expressed in full-term samples than in preterm samples and is statistically differentially expressed based on biological sex. Results of RNA sequencing and quantitative polymerase chain reaction demonstrated that the lncRNA *H19* was expressed at a statistically lower level in full-term samples than in preterm samples.

Conclusions: Our results suggest that the Fanconi anemia pathway and the *H19*/*let-7*/*LIN28B* axis may be involved in globin switching in umbilical cord blood. The differentially expressed transcripts and their related pathways may constitute novel therapeutic targets for β-thalassemia.

Background

Sickle cell disease (SCD) and β-thalassemia are chronic hemolytic anemias caused by point mutations or large deletions in the β-globin gene. Patients with β-thalassemia major rely on a life-sustaining combination of blood transfusions and iron chelation therapy. At present, hematopoietic stem cell (HSC) transplantation is the only curative treatment for these diseases. However, finding a sufficiently matched donor is difficult, and transplantation is extremely complex and expensive, which limits its wide application. Gene therapy, on the other hand, may be a promising new way to treat β-thalassemia and SCD. Based on the evidence that patients with hereditary persistence of fetal hemoglobin (HPFH) are clinically asymptomatic due to persistently high levels of fetal hemoglobin (HbF), increasing HbF levels is one of the gene therapy strategies for the treatment of severe β-thalassemia [1].

Five functional genes are arranged in the order 5'-ε-^{G}γ-^{A}γ-δ-β-3' in the β-globin cluster on chromosome 11. The temporal expression of these genes during human development follows the order in which they are arranged from 5' to 3'. Immediately before and after birth, the expression of γ-globin genes gradually decreases and expression of β-globin gradually increases. The adult hemoglobin (HbA, $\alpha_2\beta_2$) gradually

replaces HbF ($\alpha_2\gamma_2$) as the major hemoglobin in the blood. Some studies have indicated that transcription factors, such as GATA-1, KLF1, BCL11A, and ZBTB7A [2–3], play key regulatory roles in globin switching.

Up to 76% of the human genome is transcribed, but the proportion of protein-encoding mRNAs among the transcriptional products is less than 2%; the remaining transcripts are non-coding RNAs [4]. In recent years, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), circular RNAs, and other non-coding RNAs have been found to have important gene transcriptional regulatory functions. In the last few years, the roles of non-coding RNAs, such as *HMI-LNCRNA*, *BGLT3*, and *HBBP1*, in globin gene switching have gained considerable attention [5–7]. However, the current understanding of globin switching is far from comprehensive, particularly with the regard to the roles of lncRNAs. Further studies on the roles of lncRNAs in γ- to β-globin gene switching may reveal unknown mechanisms and provide targets for activating γ-globin gene expression, which could be exploited for the treatment of β-thalassemia and SCD.

Comparing RNA expression differences between conditions or samples is an effective method to reveal mechanistic differences between them. Some studies have reported transcriptome changes associated with γ- to β-globin switching using *in vitro* samples. To date, few studies have reported the transcriptome profiles related to globin switching based on fresh ex vivo samples. Li et al. [8] analyzed transcriptome profiles of erythroid progenitors at different stages during in vitro erythroid maturation and identified groups of genes with distinct expression profiles, which had roles in metabolic pathways related to cell survival, hematopoiesis, blood cell activation, and inflammatory responses. Ma et al. [9] compared the lncRNA expression profiles of peripheral blood samples from patients with thalassemia minor and major via microarray. They constructed a coding-noncoding co-expression network and performed Gene Ontology (GO) biological process analysis, thereby identifying several signaling pathways related to the β-thalassemia phenotype that are involved in the hematologic, skeletal, and hepatic systems. Most studies of the erythropoietic developmental transcriptome have been based on *in vitro*-cultured CD34 + cells. Since CD34 + cells cultured *in vitro* express higher baseline levels of HbF compared with erythroid cells *in vivo*, this system cannot completely reflect erythroid proliferation and differentiation *in vivo* [10]. The differentially expressed genes associated with the globin switching process can be better reflected by using nucleated red blood cells (NRBCs) from the umbilical cord blood of neonates at different gestational stages.

In this study, we collected umbilical cord blood samples from births that occurred at 30–35 weeks or ≥ 40 weeks of gestation to identify differentially expressed genes that may be involved in globin switching at birth.

Results

Hemoglobin electrophoresis and thalassemia gene detection

The hemoglobin electrophoretic patterns of the six cord blood samples used for RNA sequencing (RNA-Seq) analysis are shown in Supplementary Figure S1-2. The average HbF levels of the full-term female (FF) group and premature female (PF) group statistically differed (84.5% ± 8.1% and 57.8% ± 8.1%, respectively; $P = 0.016$). The gene detection results showed that none of the six newborns carried any of the three α-globin gene deletion mutations (–SEA, -α 3.7, or -α 4.2), three non-deletion mutations (Hb CS, Hb QS, or Hb WS), or 17 β-globin gene mutations most commonly seen in the Chinese population (Supplementary Figure S3).

Isolation Of Nrbcs

NRBCs were successfully isolated from each cord blood sample using magnetic bead technology based on the CD71 and CD45 cell surface markers. All isolated cells were analyzed by flow cytometry to determine purity. The average purity of all samples was > 92% (Fig. 1a, b).

The isolated cells were stained with Wright's stain and observed under a microscope. Most of the cells in the field of vision were NRBCs; the NRBCs were scattered, and mostly comprised orthochromatic normoblasts with rare polychromatic erythroblasts, basophilic erythroblasts, and proerythroblasts (Fig. 1c). The enucleated red blood cells were further identified as reticulocytes by Brilliant cresyl blue staining (Fig. 1d). NRBCs and reticulocytes accounted for > 99% of the sorted cells.

Differentially Expressed Small Rnas

More than 23 million clean 50-bp single-end reads were generated for each sample. A Poisson distribution method was used to screen the differentially expressed small RNAs between samples, and NOIseq software was used to screen the differentially expressed small RNAs between the two groups of newborns. We detected 1,767 differentially expressed miRNAs between the FF and the PF group, including 1,033 miRNAs that were upregulated and 734 miRNAs that were downregulated in the FF group compared with the PF group (Fig. 2a).

With a fold change threshold of ≥ 2 and a deviation probability value of ≥ 0.6 as our screening criteria, we identified 79 differentially expressed miRNAs in the FF group compared with the PF group, including 53 that were upregulated and 26 that were downregulated. The volcano map (Fig. 2b) and hierarchical clustering heat map (Fig. 3) were constructed to illustrate the distribution of the differentially expressed miRNAs; some of the significantly differentially expressed miRNAs are marked on the volcano map. Several members of the *let-7* miRNA family— *let-7a-3p*, *let-7b-3p*, *let-7b-5p*, *let-7f-5p*, and *let-7 g-3p*—were more highly expressed in the FF group, whereas *miR-675-3p* and *miR-675-5p* were expressed at significantly lower levels (Table 1).

Table 1
Differential expression of the *let-7* family and *miR-675*

GENE ID	Expression in FF	Fold change	Probability
hsa-let-7a-3p	Upregulated	7.58	0.85
hsa-let-7b-3p	Upregulated	22.41	0.95
hsa-let-7b-5p	Upregulated	10.49	0.88
hsa-let-7d-3p	Upregulated	2.51	0.56
hsa-let-7f-5p	Upregulated	6.58	0.83
hsa-let-7f-1-3p	Upregulated	1.24	0.27
hsa-let-7 g-3p	Upregulated	14.30	0.94
hsa-let-7i-3p	Upregulated	2.63	0.57
hsa-miR-675-3p	Downregulated	38.12	0.88
hsa-miR-675-5p	Downregulated	26.03	0.82

GO enrichment analysis was performed for the target genes of the miRNAs that were differentially expressed between the two groups. We identified 49 significant enrichment terms at the biological process level, including histone modification, covalent chromatin modification, negative regulation of protein phosphorylation, positive regulation of neuron projection development, cell growth, transforming growth factor beta receptor signaling pathway, protein methylation, stem cell differentiation, and Wnt signaling pathway. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was also performed for the target genes of the differentially expressed miRNAs; it revealed enrichment for involvement in two pathways ($Q \leq 0.05$): morphine addiction and nicotine addiction (Supplementary Figure S5).

Differentially Expressed Mrna And Lncrna Analysis

A total of 114–164 million clean 150-bp paired-end reads were generated per sample. Between the FF and PF groups, there were 12,793 differentially expressed mRNAs (11,464 known mRNAs and 1,329 novel mRNAs), and 9,121 differentially expressed lncRNAs (8,417 known lncRNAs and 704 novel lncRNAs). The analysis parameters to determine differential expression with the DEGseq software were set to a fold change of ≥ 2 and a *P*-value of ≤ 0.001. We found 7,166 differentially expressed mRNAs (6,549 known mRNAs and 617 novel mRNAs) and 3,243 differentially expressed lncRNAs (3,009 known lncRNAs and 234 novel lncRNAs) that met these criteria (Supplementary Figure S6). Volcano maps illustrating the distribution of the differentially expressed mRNAs and lncRNAs are shown in Fig. 4.

Table 2
Differential expression of globin genes and related genes

GENE ID	Expression in FF	Fold change (FF/PF)	*P*-value	*Q*-value
HBB	Upregulated	2.78	0	0
HBD	Upregulated	21.53	0	0
HBE1	Downregulated	26.45	0	0
HBG2	Downregulated	1.56	0	0
BCL11A	Upregulated	11.26	7.53E-29	1.19E-28
ZBTB7A	Upregulated	2.75	2.18E-117	1.21E-116
KLF1	Upregulated	1.43	9.32E-273	1.28E-271
SOX6	Upregulated	1.19	9.83E-16	1.03E-15
GATA1	Upregulated	1.36	1.70E-33	3.01E-33
HBS1L	Upregulated	1.41	9.57E-81	3.76E-80
MYB	Upregulated	4.81	2.01E-35	3.74E-35
NR2C2	Upregulated	2.09	1.21E-31	2.05E-31
LIN28B	Downregulated	5.83	1.35E-02	5.73E-03
HMGA2	Upregulated	4.42	2.35E-04	1.27E-04
HMGB2	Upregulated	11.79	0	0
SIRT6	Upregulated	3.50	2.76E-16	2.97E-16
RUNX1	Upregulated	2.25	1.01E-03	5.09E-04
RUNX1-T1	Upregulated	4.28	8.39E-07	5.62E-07
JAZF1	Upregulated	3.07	3.86E-252	4.80E-251
CA1	Upregulated	9.78	0	0

Among the differentially expressed transcripts, *HBB* (encoding β-globin) expression was 2.78-fold higher, *HBG2* (encoding ^{G}γ- globin) expression was 1.56-fold lower, *HBG1* (encoding ^{A}γ- globin) expression was 1.47-fold lower, and *HBD* (encoding δ-globin) expression was 21.52-fold higher in the FF samples than in the PF samples (Table 2). The genes for the transcription factors *BCL11A*, *ZBTB7A*, *KLF1*, and *SOX6*, which are known to be involved in γ-globin repression, were 11.26-, 2.75-, 1.43-, and 1.19-fold higher, respectively, in the FF compared with the PF samples. Expression of *LIN28B*, which is associated with high HbF levels [11], was 5.83-fold lower in the FF samples. In addition, we found that *H19* was expressed at much lower levels (85.28-fold) in the FF samples (Table 3).

Table 3
Differential expression of lncRNAs

GENE ID	Expression in FF	Fold change (FF/PF)	*P*-value	*Q*-value
DANCR	Upregulated	3.87	3.33E-24	4.64E-24
H19	Downregulated	85.28	0	0
RN7SK	Upregulated	3.69	0	0
SBF2-AS1	Downregulated	11.87	8.66E-159	6.54E-158

GO enrichment analysis of the 6,549 known differentially expressed mRNAs, with a corrected *P*-value of ≤ 0.05 as the threshold, yielded 967 terms for enriched biological processes (Fig. 5a), including DNA replication, DNA conformation changes, and RNA splicing. KEGG pathway enrichment analysis revealed that the differentially expressed mRNAs were enriched in 57 pathways (Q ≤ 0.05) related to the spliceosome, RNA transport, cell cycle, systemic lupus erythematosus (SLE), the Fanconi anemia pathway, DNA replication, and other critical regulatory processes (Fig. 5b).

We also performed GO enrichment analysis of the target genes of the differentially expressed lncRNAs. For biological processes, 457 GO terms were enriched, such as DNA conformation changes, RNA splicing, DNA replication, and RNA transport (Fig. 6a). KEGG pathway enrichment analysis of the differentially expressed lncRNA target genes revealed enrichment of 11 pathways (Q ≤ 0.05), including the spliceosome, ubiquitin mediated proteolysis, SLE, the mRNA surveillance pathway, and chronic myeloid leukemia (Fig. 6b).

Quantitative Polymerase Chain Reaction Analysis

We collected 75 umbilical cord blood samples from which we isolated NRBCs to verify the differential expression of seven mRNAs and lncRNAs by quantitative polymerase chain reaction (qPCR). Compared with the preterm group, the full-term group had higher mRNA expression levels of *BCL11A*, *HBB*, and *HMGB2* (3.24-, 3.52-, and 1.59-fold, respectively). The expression of *H19* and *LIN28B* was 69.98- and 2.38-fold lower in the full-term group. The results were consistent with those of the RNA-SEq. We observed no significant differences in the expression of *DANCR* and *SIRT6*. In addition, we found that *BCL11A* expression significantly differed based on biological sex. Cord blood from preterm male neonates had 4.25-fold higher expression than cord blood from preterm female neonates, and cord blood from full-term male neonates had 2.64-fold higher expression than cord blood from full-term female neonates (Fig. 7).

Discussion

There are few studies on transcriptome changes during the important switch from γ- to β-globin gene expression immediately before and after birth, especially using fresh *ex vivo* samples. In this study, we

applied RNA-Seq technology to NRBCs collected from the umbilical cord blood of preterm and full-term newborns to comprehensively analyze their mRNA, lncRNA, and small RNA profiles, with the aim of guiding the study of fetal-to-adult hemoglobin switching.

Our results demonstrated that the expression levels of globin-encoding genes significantly differed in preterm and full-term cord blood. Compared with their expression in the preterm group, the mRNA expression levels of the β-globin-encoding *HBB* and δ-globin-encoding *HBD* genes were significantly higher in the full-term group, whereas the expression of the γ-globin-encoding *HBG1* and *HBG2* genes was lower. Our results, which are consistent with the gradual replacement of γ-globin chains by adult β- and δ-globin chains during the switch from fetal to adult hemoglobin, confirm the selection of appropriate samples for study. In addition, the expression of transcription factors genes, such as *BCL11A* and *ZBTB7A*, was also higher in the full-term samples. BCL11A and ZBTB7A are known to be important γ-globin gene repressors [3]. Furthermore, among the seven genes selected for qPCR analysis to verify their expression levels, five showed results consistent with the transcriptome sequencing results. Overall, these findings confirm the reliability of our transcriptome sequencing results.

Interestingly, significant sex-based differences were observed by qPCR analysis in the mRNA expression of *BCL11A* in both premature and full-term cord blood samples. The expression of *BCL11A* mRNA was significantly higher in male neonates than in female neonates. To our knowledge, this is the first report of sex-based differences in the expression of *BCL11A*. However, our previous study indicated that in the peripheral blood of adults, the reference range of HbF in normal male adults (< 0.5%) was significantly lower than that in female adults (< 1.1%) [12]. Whether the difference between the HbF levels in male and female adults is related to the sex-based differential expression of *BCL11A* requires further exploration.

GO and KEGG pathway enrichment analyses revealed that the differentially expressed RNAs were linked to biological processes such as DNA replication, DNA conformation changes, RNA splicing, RNA transport, cell cycle, SLE, and the Fanconi anemia pathway. The genes related to these GO terms and pathways may be involved in the regulation of globin switching. Among them, the Fanconi anemia pathway is of particular interest. Fanconi anemia is a rare autosomal recessive genetic disease characterized by congenital aplastic anemia, accompanied by multiple congenital malformations. The loss of Fanconi anemia pathway gene function causes Fanconi anemia; at least 22 pathogenic genes involved in Fanconi anemia have been reported [13]. Due to their genetic defects, the cells from patients with Fanconi anemia are unable to repair damaged DNA, putting the patients at high risk of developing a variety of cancers. We found that the Fanconi anemia genes, including *FANCA*, *FANCB*, and *FANCC*, were significantly more highly expressed in the FF group, in which the level of HbF is lower, than in the PF group. Meanwhile, high levels of HbF are often observed in FA patients, with an average increase of 11.3% over the levels in healthy individuals [14], suggesting a direct or indirect relationship between the upregulation of Fanconi anemia genes and the downregulation of the γ-globin gene.

In recent years, miRNAs have been found to regulate gene expression by silencing or degrading their target mRNA molecules, thereby playing an important role in development and disease. A relationship

between miRNAs and the fetal-to-adult hemoglobin switch has also been discovered. Sankaran et al. [15] demonstrated that increased expression of *miR-15a* and *mir-16-1* resulted in elevated HbF gene expression by downregulating the expression of *MYB*. Our results indicated that *miR-16-1-3p* was downregulated, which could lead to the upregulation of *MYB* gene expression, thereby inhibiting the expression of the gene for γ-globin. Previous studies have demonstrated that elevated levels of *miR-210* correlate with high HbF levels in patients with β-thalassemia, and induction of the expression of *miR-210* can promote the synthesis of γ-globin [16]. On the other hand, some miRNAs inhibit expression of the gene for γ-globin. *miR-96* is more highly expressed in reticulocytes from adult peripheral blood than in those from umbilical cord blood, and genetic knockdown of endogenous *miR-96* increased γ-globin expression by 20% [17]. However, we found no significant differences in the expression of *miR-210*, *miR-96-3p*, or *miR-96-5p* in the FF and PF groups. This discrepancy may be due to the different research subjects evaluated in this study and others. The effects of these miRNAs on the expression of the gene for γ-globin remain to be confirmed in future studies.

We also observed that several members of the *let-7* miRNA family were more highly expressed in the full-term group than in the preterm group, and their expression inversely correlated with γ-globin expression. Recent studies have found that silencing the miRNA *let-7a* in adult CD34 + cells increased HbF levels by up to 38%, and diminished the expression of *BCL11A*. On the contrary, no significant change was observed in *let-7* miRNA expression after *BCL11A* was knocked out, indicating that *let-7* is an upstream regulator of *BCL11A* expression [18].

lncRNAs also have important functions in gene expression regulation, and they closely interact with miRNAs. At present, only a few lncRNAs have been reported to be involved in the regulation of globin switching. Morrison et al. [5] found that downregulation of *HMI-LNCRNA* in the immortalized human erythroid progenitor HUDEP-2 cell line significantly upregulated *HBG* expression both at the mRNA and protein levels, although no changes were found in the expression of *BCL11A* and other key transcription factors known to modulate *HBG* expression. The lncRNA *BGLT3* locus located in the γ–δ intergenic region can promote the production of γ-globin in various ways; the *BGLT3* transcript alone is insufficient to rescue γ-globin transcription and either the locus or its transcription is also necessary [6]. In addition, lncRNAs are also involved in the regulation of erythroid development [4].

Our results demonstrated that the expression of the lncRNA *H19* was significantly lower in the FF group compared with the PF group. *H19* has no reported relationship to hemoglobin switching, but it interacts with γ-globin regulator *LIN28B* [19]; thus, the role of this lncRNA in γ-globin expression requires further investigation. Paternally imprinted and maternally expressed *H19*, one of the earliest reported lncRNAs, has an important role in embryonic development and growth control [20]. *H19* expressed in long-term HSCs regulates embryonic HSC production and adult HSC quiescence [21, 22]. It is highly expressed in developing embryos, mainly in the mesoderm and endoderm. After birth, the expression of *H19* in most tissues is significantly downregulated [19], with obvious spatiotemporal expression characteristics. Interestingly, this expression pattern is comparable to that of the β-globin gene cluster. The γ-globin genes

(*HBG1* and *HBG2*) are expressed throughout most of fetal development, and their expression is gradually replaced by β-globin gene (*HBB*) expression after birth. Thus, the spatiotemporal expression of *H19* may be related to the changes in β-globin cluster expression.

A recent study analyzed the transcriptomic differences between fetal liver- and adult bone marrow-derived HSCs during erythroid differentiation using RNA-SEq. The findings indicated that *H19* expression was lower, *let-7* miRNA family member expression was higher, and *LIN28B* expression was lower in adult compared with fetal erythroblasts [23], which is consistent with our transcriptome sequencing results. Our qPCR results also verified that the expression patterns of *LIN28B* and *BCL11A* were consistent with those reported in previous studies. Therefore, we speculate that *H19* regulates hemoglobin switching in various ways. Firstly, *H19* can act as a competitive endogenous RNA sponge to adsorb *let-7* miRNAs, freeing *let-7* to participate in the negative regulation of γ-globin via several mechanisms. During the differentiation of CD34 + cells, *let-7* and *LIN28B* affect the expression of the gene for γ-globin by upregulating and downregulating *BCL11A*, respectively [11]. *HMGA2* is a downstream target of the *let-7* family miRNAs, and over-expression of *HMGA2* in CD34 + cells resulted in increased γ-globin expression at the gene and protein level [24].

H19 transcripts also serve as precursors for *miR-675*, an important miRNA involved in gene expression regulation. In our sequencing results, *H19*, *miR-675-3p*, and *miR-675-3p* expression were lower in the FF group than in the PF group. The main downstream regulatory genes of *miR-675* are *RUNX1* and *RUNX1-T1*, which both have elevated expression in the FF group. *RUNX1* is an essential transcription factor that regulates the expression of several important target genes, and plays critical roles in the process of hematopoiesis. The expression of RUNX1a in embryoid bodies promotes definitive hematopoiesis that generates β-globin–producing erythrocytes [25]. Therefore, *H19* may also have an important influence on γ-globin expression through regulation of the *miR-675*/*RUNX1* axis. Taken together, these findings suggest that *H19* may be an important regulator of γ- to β-globin gene switching, but the specific regulatory mechanism is yet to be determined.

Conclusions

Our study identified differentially expressed transcripts in NRBCs freshly isolated from the cord blood of preterm and full-term newborns using transcriptome sequencing. These differentially regulated transcripts, as well as their related genes and pathways, may participate in regulating the switch from γ- to β-globin gene expression. As such, they may provide important insight into globin switching mechanisms, thereby guiding subsequent studies and the development of therapies for hemoglobinopathies.

Methods

Subjects

For the purpose of RNA-Seq analysis, cord blood was collected from three randomly selected healthy female preterm newborns (gestational age: 30–35 weeks) and three randomly selected healthy female full-term newborns (gestational age: ≥40 weeks) who were delivered in the Obstetrical Department of Guizhou Provincial People's Hospital. In addition, a total of 75 samples from healthy newborns (32 preterm newborns [15 female and 17 male] and 43 full-term newborns [21 female and 22 male]) were collected to verify the differential expression of transcripts by qPCR.

Separation Of Nrbcs From Umbilical Cord Blood

We collected 12-ml umbilical cord blood samples from each newborn, which were used for isolation of NRBCs (10 mL), and analysis of hemoglobin by capillary electrophoresis with a CAPILLARYS 2 instrument (Sebia, Lisses, France) and detection of the thalassemia gene (2 mL) with a kit provided by Yishengtang Biological Enterprise (Shenzhen, China). Samples from newborns carrying thalassemia mutations were excluded.

NRBCs were separated with a two-step method. First, mononuclear cells (MNCs) were enriched by density-gradient centrifugation, then NRBCs (CD71 + CD45−) were isolated from MNCs using the midiMACS magnetic separation system (Miltenyi Biotech, Bergisch Gladbach, Germany). After incubation with anti-CD71–FITC and anti-CD45–PE antibodies, the NRBCs were assessed for purity by flow cytometry (BD FACSAria™ II, Becton, Dickinson and Company, Franklin Lakes, NJ, USA). The flow cytometry data were analyzed with FlowJo software (Treestar, Ashland, Oregon, USA). The isolated NRBCs were stained with Wright's stain and Brilliant cresyl blue to facilitate the observation of cell morphology.

Extraction of total RNA and construction of the sequencing library

Total RNA was extracted from NRBCs using TRIzol® Reagent (Invitrogen, Carlsbad, CA, USA), per the manufacturer's instructions. The RNA integrity was measured with an Agilent 2100 Bioanalyzer (Santa Clara, CA, USA). Sequencing was performed if the samples had RNA integrity numbers > 8 and the ratio of 28S/18S was > 1.5. Then, a small RNA library (BGI RNA-Seq kit, BGI, Shenzhen, China) and a ribosomal RNA depletion library (Ribo-Zero™ rRNA Removal Kit and TruSeq® Stranded Total RNA Sample Preparation Kit, Illumina, San Diego, CA, USA) were constructed for RNA-Seq.

Rna-seq And Data Analysis

A BGIseq-500 sequencer and Illumina HiSeq 4000 platform were used for small RNA sequencing and lncRNA sequencing, respectively. Single-end 50-bp reads and 2 × 150-bp paired-end reads were generated. The raw data were filtered to get clean data, on which the subsequent analysis is based. The multiBamSummary program of the deepTools suite was used to compute the read coverage of the

genomic regions for the different samples. Then, we performed principal component analysis with plotPCA of the deepTools suite.

For small RNAs, clean reads were mapped to a reference genome and other small RNA databases using Bowtie2, then classified and annotated. MiRDeep2 was used to predict unknown miRNAs. The expression level of each small RNA was calculated and standardized based on the transcripts per kilobase million value. We used miRanda and TargetScan together to predict miRNA targets. A Poisson distribution method was used to screen differentially expressed small RNAs among the samples, according to the following default criteria: fold change ≥ 2 and false discovery rate ≤ 0.001. NOISeq was used to screen differentially expressed small RNAs between the groups, according to the following default criteria: fold change ≥ 2 and diverge probability ≥ 0.6. We performed hierarchical clustering analysis using the R package pheatmap, and visualized the data using Treeview. We used ClusterProfiler to perform GO and KEGG pathway enrichment analysis of the differentially expressed miRNA targets.

For mRNAs and lncRNAs, clean reads were mapped to the GRCh37/hg19 reference genome using HISAT. StringTie was used to assemble and compare the transcripts with known mRNAs or lncRNAs. Cuffmerge was used to merge the transcripts, and the merged transcripts were used for further analysis. Three pieces of prediction software (CPC, txCdsPredict, and CNCI) and the database pfam were used to predict protein-encoding ability and distinguish mRNAs from lncRNAs. The reads were aligned to GRCh37/hg19 using Bowtie2, then RSEM was used to calculate the fragments per kilobase of transcript per million mapped reads of each transcript, which represents the expression level of each gene. We performed differential expression analysis between samples and groups, respectively, using the Poisson distribution and DEGSeq methods. We predicted the target genes of lncRNAs and miRNA precursors, respectively, and performed GO and KEGG enrichment analysis of the differentially expressed mRNAs and target genes of the differentially expressed lncRNAs.

Qpcr Analysis

We isolated NRBCs from the 75 umbilical cord blood samples described above. Total RNA was extracted from the NRBCs, then cDNA was synthesized using a reverse transcription system (TAKARA, Shiga, Japan). qPCR assays were performed using a SYBR® Green qPCR kit (KAPA Biosystems, Wilmington, MA, USA). *GAPDH* was used as an endogenous control gene. The primer sequences are shown in Supplementary Table S1. The qPCR conditions were: pre-denaturation at 95 °C for 6 min, and 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and 65 °C for 5 s, increasing by 0.5 °C/5 s up to 95 °C. We derived expression levels using the $2^{-\Delta\Delta Ct}$ method. Student's *t*-test (two-tailed) was used to analyze the data. Statistical significance was set at a threshold of $P < 0.05$.

Abbreviations

GO: Gene Ontology; HbA: adult hemoglobin; HbF: hemoglobin; HPFH: hereditary persistence of fetal hemoglobin; HSCs: Hematopoietic stem cells; KEGG: Kyoto Encyclopedia of Genes and Genomes; lncRNA: long non-coding RNA; miRNA: microRNA; MNCs: mononuclear cell; NRBCs: nucleated red blood cells; qPCR: quantitative polymerase chain reaction; RNA-Seq: RNA sequencing; SCD: sickle cell disease; SLE: systemic lupus erythematosus.

Declarations

Ethics approval and consent to participate

This project was approved by the ethics committee of Guizhou Provincial People's Hospital. All parents of the donor infants gave written informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during the current study are available in the Genome Sequence Archive (GSA) for human, with Project ID: HRA000232 (https://bigd.big.ac.cn/gsa-human/s/83dXf6X3).

Competing interests

The authors declare that they have no competing interest.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81660023, 81960040), Guizhou Province Scientific and Technological Innovation Project (Grant No. 20165670 20205011). The funders played no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Authors' contributions

SWH and YKJ conceived and designed the study; LH, MZ and TTJ prepared experimental samples; YYH, SJG and LH contributed to bioinformatic analysis and data analysis; JL and WQZ performed the qPCR validation experiments; ZJZ and XDF conceived and supervised the study; YYH drafted the manuscript; and SWH, YKJ, and XDF finalized the manuscript. All authors discussed and interpreted results, and all authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to all the donors for their interest and cooperation.

References

1. Wu Y, Zeng J, Roscoe BP, et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nat Med. 2019;25(5):776–83.
2. Liu N, Hargreaves VV, Zhu Q, et al. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell. 2018;173(2):430–42.e17.
3. Masuda T, Wang X, Maeda M, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science. 2016;351(6270):285–9.
4. Gallagher PG. Long noncoding RNAs in erythropoiesis. Blood. 2014;123(4):465–6.
5. Morrison TA, Wilcox I, Luo HY, et al. A long noncoding RNA from the HBS1L-MYB intergenic region on chr6q23 regulates human fetal hemoglobin expression. Blood Cells Mol Dis. 2018;69:1–9.
6. Ivaldi MS, Diaz LF, Chakalova L, Lee J, Krivega I, Dean A. Fetal γ-globin genes are regulated by the BGLT3 long noncoding RNA locus. Blood. 2018;132(18):1963–73.
7. Huang P, Keller CA, Giardine B, et al. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. Genes Dev. 2017;31(16):1704–13.
8. Li B, Ding L, Li W, Story MD, Pace BS. Characterization of the transcriptome profiles related to globin gene switching during in vitro erythroid maturation. BMC Genom. 2012;13:153.
9. Ma J, Liu F, Du X, Ma D, Xiong L. Changes in lncRNAs and related genes in β-thalassemia minor and β-thalassemia major. Front Med. 2017;11(1):74–86.
10. Vinjamur DS, Bauer DE, Orkin SH. Recent progress in understanding and manipulating haemoglobin switching for the haemoglobinopathies. Br J Haematol. 2018;180(5):630–43.
11. Lee YT, de Vasconcellos JF, Yuan J, et al. LIN28B-mediated expression of fetal hemoglobin and production of fetal-like erythrocytes from adult human erythroblasts ex vivo. Blood. 2013;122(6):1034–41.
12. Han WP, Huang L, Li YY, et al. Reference intervals for HbA2 and HbF and cut-off value of HbA2 for β-thalassemia carrier screening in a Guizhou population of reproductive age. Clin Biochem. 2019;65:24–8.
13. Jung M, Ramanagoudr-Bhojappa R, van Twest S, et al. Association of clinical severity with FANCB variant type in Fanconi anemia. Blood. 2020;135(18):1588–602.
14. Sousa R, Gonçalves C, Guerra IC, et al. Increased red cell distribution width in Fanconi anemia: a novel marker of stress erythropoiesis. Orphanet J Rare Dis. 2016;11(1):102.
15. Sankaran VG, Menne TF, Šćepanović D, et al. MicroRNA-15a and – 16 – 1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc Natl Acad Sci U S A. 2011;108(4):1519–24.
16. Bianchi N, Finotti A, Ferracin M, et al. Increase of microRNA-210, decrease of raptor gene expression and alteration of mammalian target of rapamycin regulated proteins following mithramycin treatment of human erythroid cells. PLoS One. 2015;10(4):e0121567.

17. Azzouzi I, Moest H, Winkler J, et al. MicroRNA-96 directly inhibits γ-globin expression in human erythropoiesis. PLoS One. 2011;6(7):e22838.
18. De Vasconcellos JF, Byrnes C, Lee YT, et al. Tough decoy targeting of predominant let-7 miRNA species in adult human hematopoietic cells. J Transl Med. 2017;15(1):169.
19. Peng F, Li TT, Wang KL, et al. H19/let-7/LIN28 reciprocal negative regulatory circuit promotes breast cancer stem cell maintenance. Cell Death Dis. 2017;8(1):e2569.
20. Gabory A, Jammes H, Dandolo L. The H19 locus: role of an imprinted non-coding RNA in growth and development. Bioessays. 2010;32(6):473–80.
21. Zhou J, Xu J, Zhang L, et al. Combined Single-Cell Profiling of lncRNAs and Functional Screening Reveals that H19 Is Pivotal for Embryonic Hematopoietic Stem Cell Development. Cell Stem Cell. 2019;24(2):285–98.e5.
22. Venkatraman A, He XC, Thorvaldsen JL, et al. Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. Nature. 2013;500(7462):345–9.
23. Lessard S, Beaudoin M, Orkin SH, Bauer DE, Lettre G. 14q32 and let-7 microRNAs regulate transcriptional networks in fetal and adult human erythroblasts. Hum Mol Genet. 2018;27(8):1411–20.
24. De Vasconcellos JF, Lee YT, Byrnes C, Tumburu L, Rabel A, Miller JL. HMGA2 Moderately Increases Fetal Hemoglobin Expression in Human Adult Erythroblasts. PLoS One. 2016;11(11):e0166928.
25. Ran D, Shia WJ, Lo MC, et al. RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells. Blood. 2013;121(15):2882–90.

Figures

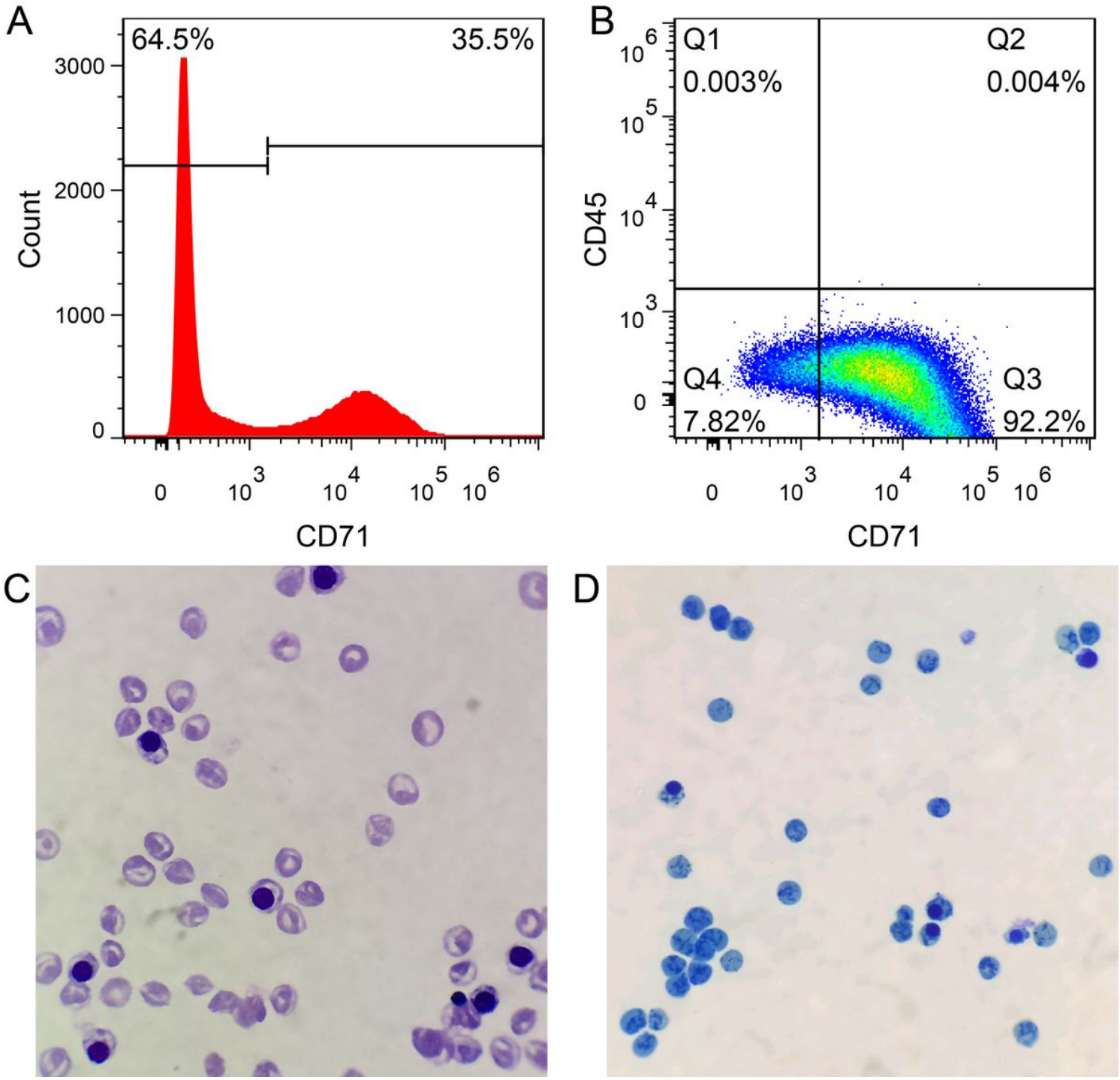

Figure 1

Flow cytometry analysis and cellular morphology. (a) CD71–FITC-stained MNCs were divided into two groups: CD71 positive and negative. (b) Flow cytometry analysis results of isolated NRBC, in which Q3 corresponds to CD71+ CD45- cells. (c) Wright's staining (100×). (d) Brilliant cresyl blue staining (100×).

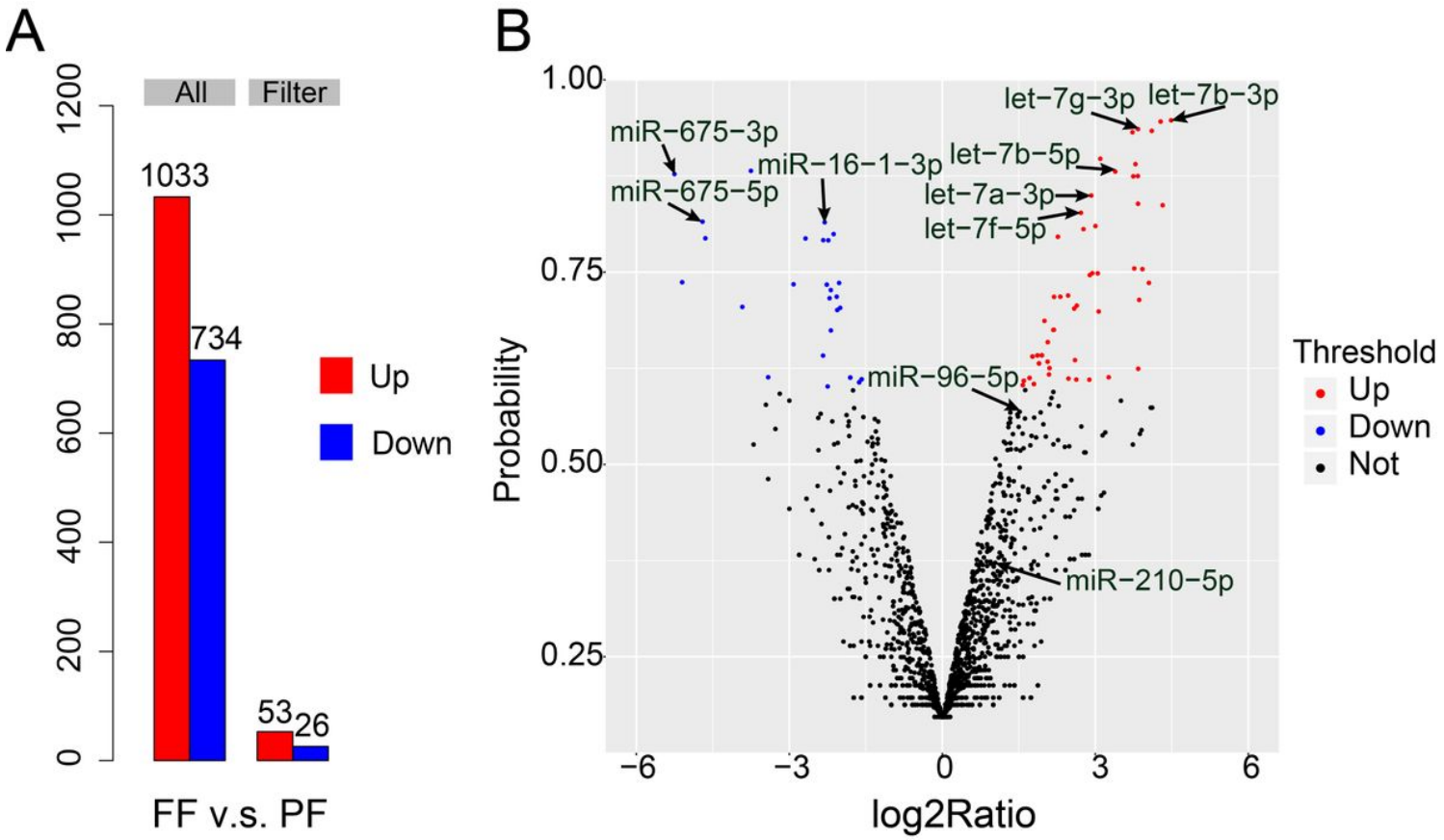

Figure 2

Differentially expressed miRNA. Blue color indicates downregulation, and red color indicates upregulation. (a) The X-axis represents the comparison groups, and the Y-axis represents the corresponding number of small RNAs. (b) Volcano map of all miRNAs between the preterm and full-term groups.

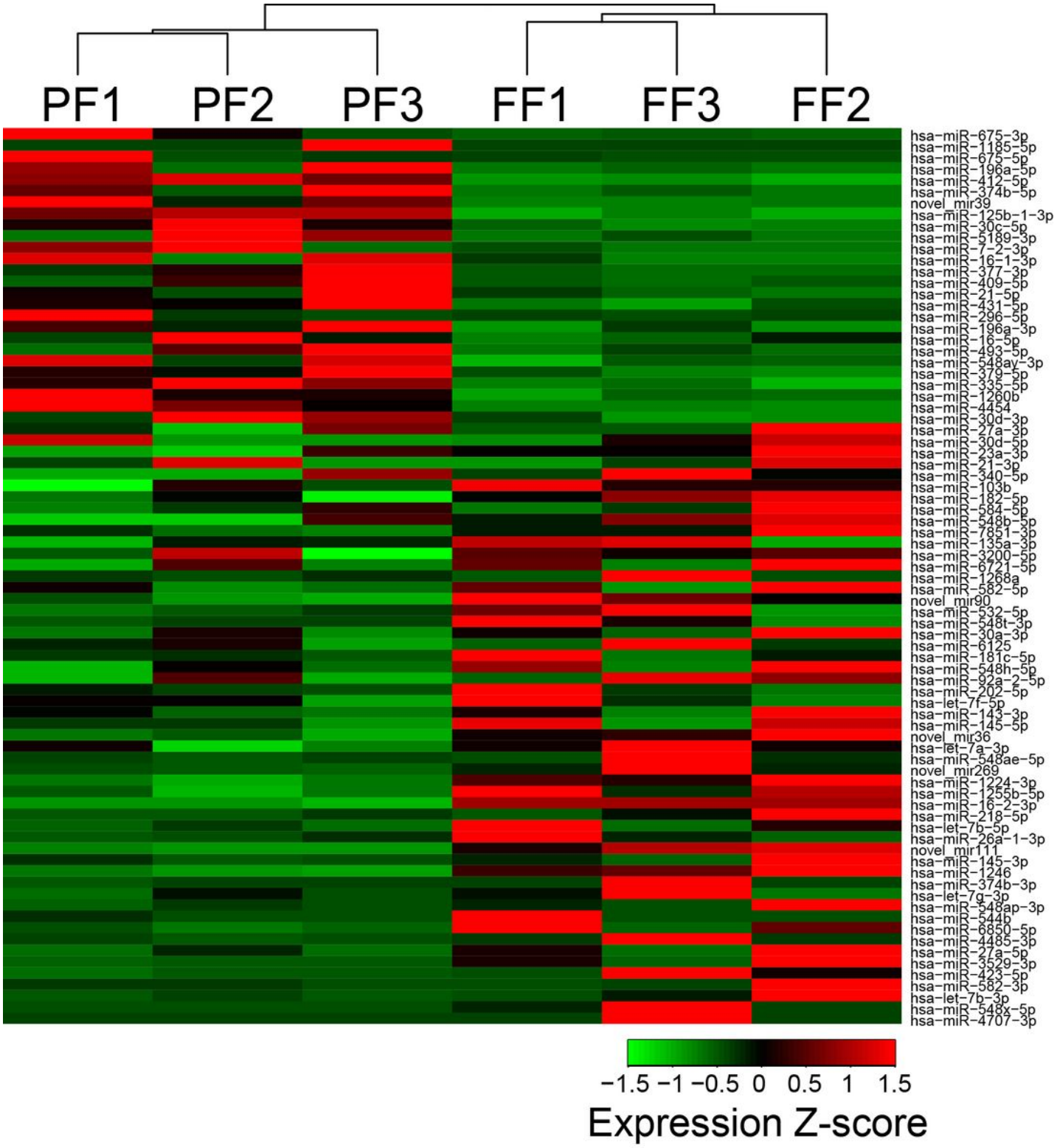

Figure 3

Hierarchical clustering heat map of 79 differentially expressed miRNAs. The color scale represents z-scores.

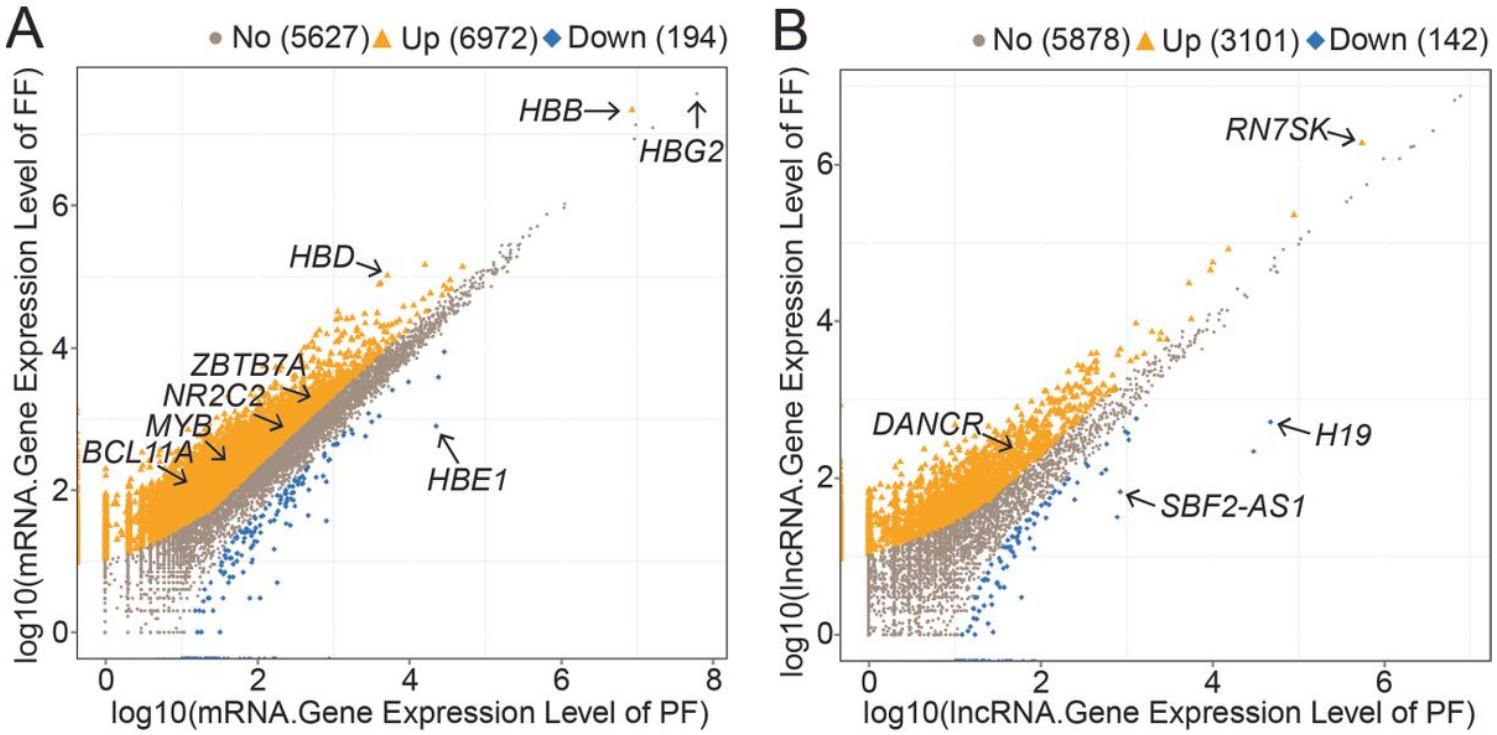

Figure 4

Differential expression profiles of known mRNA and lncRNA in FF-vs-PF. The yellow triangles represent differentially upregulated genes, blue diamonds represent differentially downregulated genes, and gray dots represent genes that are not different. (a) mRNA expression profile. (b) lncRNA expression profile.

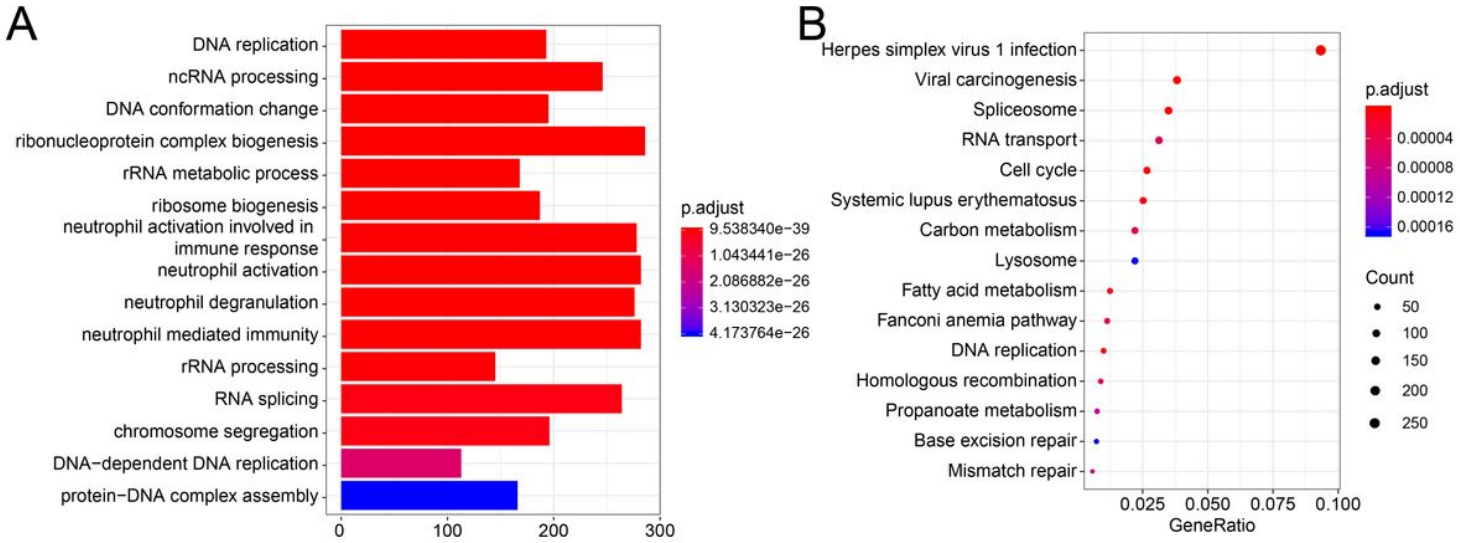

Figure 5

GO and KEGG enrichment analysis of differentially expressed mRNA. (a) GO enrichment analysis. The X-axis represents the number of genes, and the Y-axis represents GO term. (b) KEGG pathway enrichment analysis. The X-axis represents the Gene Ratio, and the Y-axis represents the path name. The color depth represents the degree of enrichment.

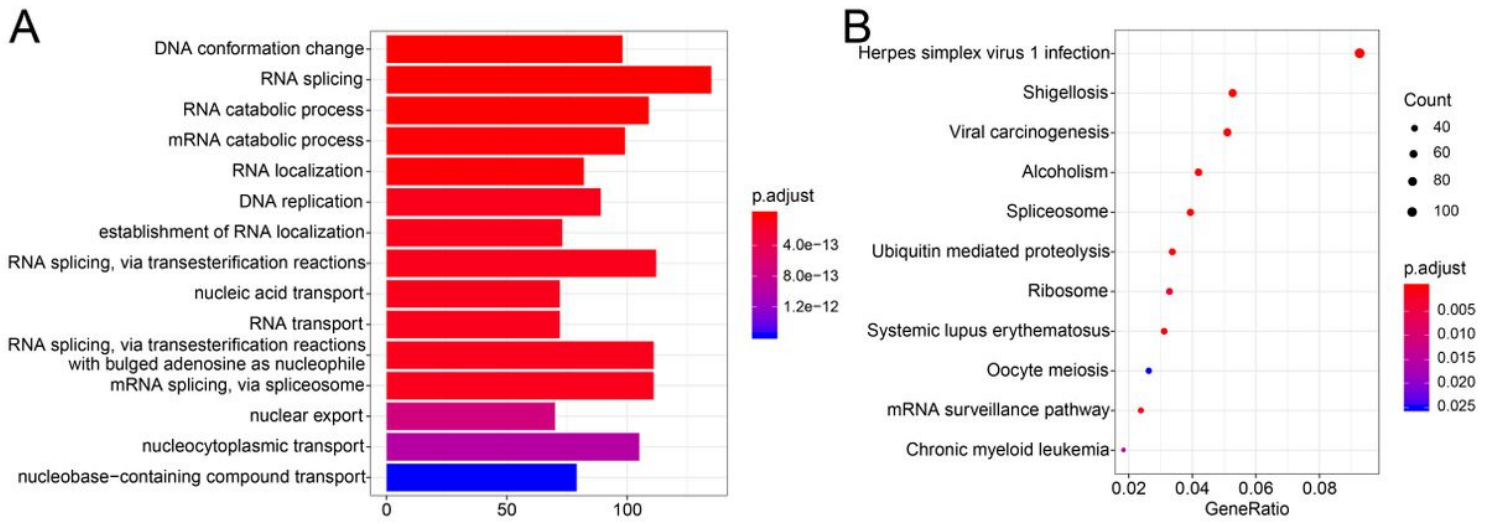

Figure 6

GO and KEGG enrichment analysis of differentially expressed lncRNA target genes. (a) GO enrichment analysis. The X-axis represents the number of genes, and the Y-axis represents GO term. (b) KEGG pathway enrichment analysis. The X-axis represents the Gene Ratio, and the Y-axis represents the path name. The color depth represents the degree of enrichment.

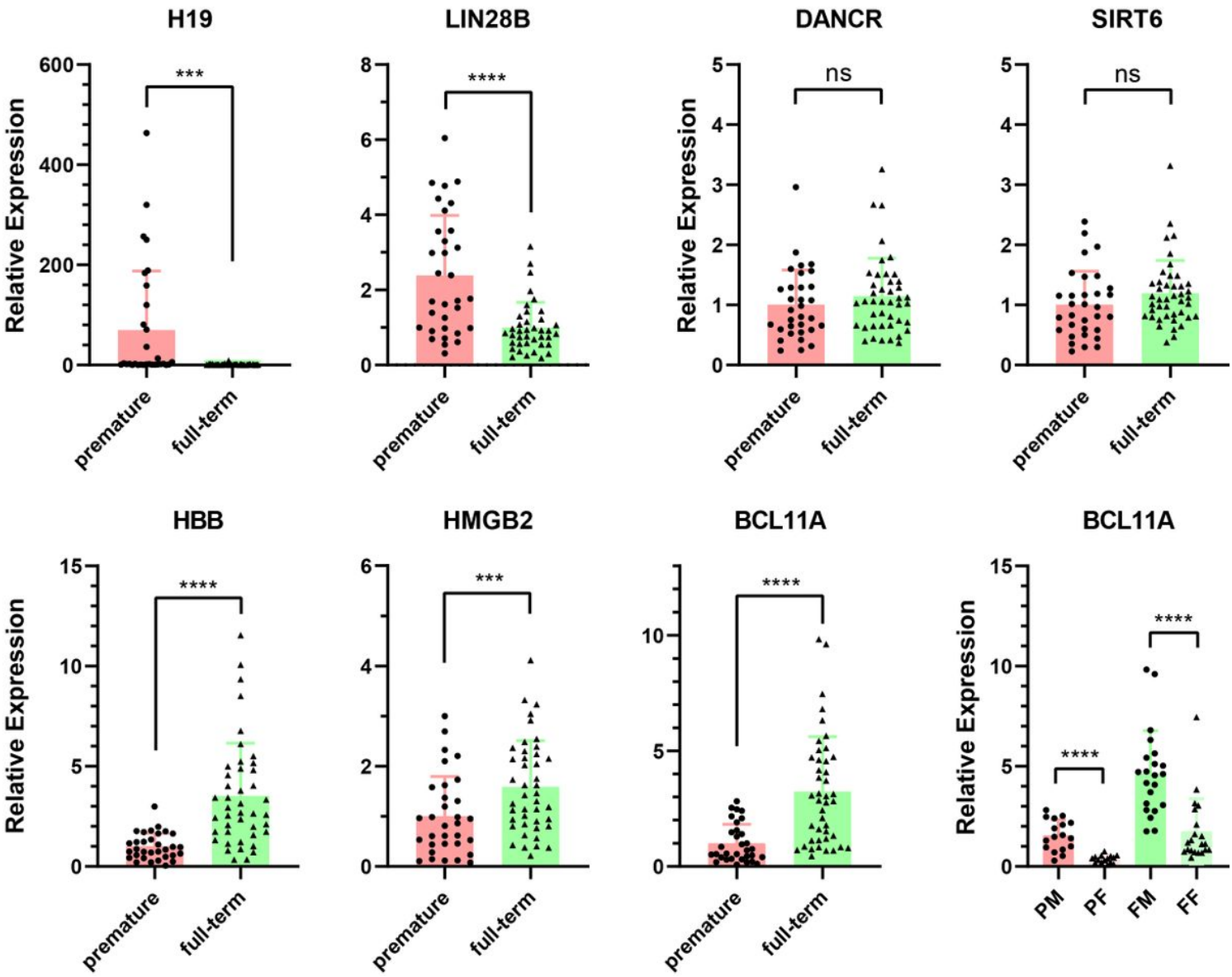

Figure 7

qPCR-based detection of mRNA and lncRNA relative expression changes. PM: preterm male; PF: preterm female; FM: full-term male; FF: full-term female. The abscissa represents different groups, and the ordinate represents the fold change of relative quantity. Shown is the mean ± SEM (ns: $P \geq 0.05$; ***$P < 0.001$; ****$P < 0.0001$).

Supplementary Files

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

- SupplementaryFigure7.docx
- SupplementaryTable.docx
- SupplementaryFigure16.docx