Journal of Perinatology (2021) 41:551–561
https://doi.org/10.1038/s41372-020-00868-9

ARTICLE

# Tracheal aspirate transcriptomic and miRNA signatures of extreme premature birth with bronchopulmonary dysplasia

Christiana N. Oji-Mmuo[1] · Roopa Siddaiah[1] · Deborah T. Montes[2] · Melody A. Pham[1] · Debra Spear[1] · Ann Donnelly[1] · Nathalie Fuentes[1] · Yuka Imamura-Kawasawa[3] · Judie A. Howrylak[4] · Neal J. Thomas[1] · Patricia Silveyra[1,2]

Received: 28 February 2020 / Revised: 6 October 2020 / Accepted: 28 October 2020 / Published online: 11 November 2020


## Abstract

**Objective** Extreme preterm infants are a growing population in neonatal intensive care units who carry a high mortality and morbidity. Multiple factors play a role in preterm birth, resulting in major impact on organogenesis leading to complications including bronchopulmonary dysplasia (BPD). The goal of this study was to identify biomarker signatures associated with prematurity and BPD.

**Study design** We analyzed miRNA and mRNA profiles in tracheal aspirates (TAs) from 55 infants receiving invasive mechanical ventilation. Twenty-eight infants were extremely preterm and diagnosed with BPD, and 27 were term babies receiving invasive mechanical ventilation for elective procedures.

**Result** We found 22 miRNAs and 33 genes differentially expressed (FDR < 0.05) in TAs of extreme preterm infants with BPD vs. term babies without BPD. Pathway analysis showed associations with inflammatory response, cellular growth/proliferation, and tissue development.

**Conclusions** Specific mRNA-miRNA signatures in TAs may serve as biomarkers for BPD pathogenesis, a consequence of extreme prematurity.

## Introduction

The routine use of antenatal steroids has resulted in the survival of extremely preterm infants. Because prematurity disrupts the development of all the organ systems in the body, the mortality and morbidity of prematurely born infants continue to be high. Bronchopulmonary dysplasia (BPD) is a form of chronic lung disease that develops in neonates as a consequence of preterm birth, and is characterized by altered growth factor expression, immature immune system, and arrest of pulmonary alveologenesis and vasculogenesis [1–3]. BPD is the second most common complication of extreme preterm birth (39–42%), only to follow retinopathy of prematurity (63%) [4]. Preclinical studies suggest that alteration in the expression of certain growth factors plays a critical role in the arrest of lung development, now referred to as the "New BPD" [1, 2, 5].

First described in 1967 by Northway and colleagues, BPD is characterized by clinical, pathological and radiographic features [6]. Approximately 10,000–15,000 neonates are diagnosed with BPD per year in the United States, with annual healthcare costs of up to $1.5 billion [7]. While BPD

These authors contributed equally: Christiana N. Oji-Mmuo, Roopa Siddaiah

**Supplementary information** The online version of this article (https://doi.org/10.1038/s41372-020-00868-9) contains supplementary material, which is available to authorized users.

✉ Patricia Silveyra
patry@email.unc.edu

[1] Department of Pediatrics, Penn State College of Medicine, Hershey, PA, USA

[2] Biobehavioral Laboratory, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3] Institute for Personalized Medicine, Departments of Biochemistry and Molecular Biology and Pharmacology, Penn State College of Medicine, Hershey, PA, USA

[4] Division of Pulmonary and Critical Care Medicine, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA

patients are at heightened risks for adverse short-and long-term outcomes, infants with severe BPD (as defined by the National Heart, Lung, and Blood Institute, NHLBI) [8] are associated with high mortality and morbidity in the first 2 years of life, including longer length of stay in the neonatal intensive care unit (NICU), increased number of days on ventilatory support, and high hospital readmissions once discharged home [9]. The etiology of BPD is multifactorial, and it has become evident that early preventive strategies have the greatest potential to modify the disease progression [10]. Among factors that have been implicated in BPD are antenatal/postnatal infection, placental insufficiency, postnatally prolonged mechanical ventilation (volutrauma), oxidative stress, poor nutrition, and racial and genomic influences [11]. More recently, genomics, microbiomics, proteomics, and metabolomics studies have begun to provide newer insights into BPD pathogenesis, with great potential to change clinical management via biomarker discovery and development of therapeutic targets [12].

MicroRNAs (miRNAs) are small non-coding RNAs that control gene expression when interacting with specific sequences in the untranslated regions (UTRs) of mRNAs. In general, miRNAs bind to 3'UTRs to negatively affect protein levels by inhibiting translation and/or targeting mRNAs for degradation. There are about 2300 human miRNAs identified to date which are annotated and cataloged in searchable web-based databases [13]. Due to the nature of their interaction, each miRNA is predicted to have hundreds of mRNA targets; thus, it is likely that ~30% of human genes are regulated by miRNAs [14]. MiRNAs have emerged as regulatory elements involved in the control of immune cellular development and homeostasis, as well as in responses to oxidative stress caused by hyperoxia in BPD [15, 16]. In addition, studies have reported unique miRNA expression profiles in inflammatory lung diseases like cystic fibrosis, chronic obstructive pulmonary disease, asthma, idiopathic pulmonary fibrosis, and lung cancer, and recent studies have reported links between altered miRNA levels and BPD [17–19]. The relatively stable nature of these extracellular miRNAs in several biological fluids, including plasma, urine, serum, saliva, and tracheal aspirates (TAs) in disease states characterizes them as potential candidates for diagnostic biomarkers [20]. Moreover, increasing evidence has implicated miRNAs in the regulation of key genes in a variety of human pulmonary diseases, driving current investigations into their potential as therapeutic targets [21].

Transcriptomic analyses of samples obtained from patients with BPD are also beginning to uncover targeted genes that are likely disease specific [22]. For example, autopsy examination followed by genome wide transcriptional analysis of lung tissue obtained from BPD and control patients identified 159 differentially expressed genes, including genes related to connective tissue and mast cells function [23]. However, very few studies have evaluated transcriptomic profiles in newborn's TAs who are born prematurely [24]. The goals of our current study were to identify and characterize differences in the transcriptomic and miRNA profiles in TAs obtained from extremely preterm infants with BPD, using term infants as controls, as well as to identify regulatory gene–miRNA networks associated with BPD pathogenesis and prematurity.

## Materials and methods

### Patient population

The study was approved by the Penn State College of Medicine Institutional Review Board. Following written informed consent from the parents, we prospectively enrolled extreme preterm infants receiving invasive mechanical ventilation at 7 days of life or older and at risk for developing BPD ($n = 28$) at the Penn State Health Children's Hospital NICU. BPD was defined based on the NHLBI consensus conference classification [8], and the severity of BPD was determined based on the degree of respiratory support and or oxygen requirement at 28 days of life and 36 weeks postmenstrual age (PMA) [25]. As control subjects (C), we also enrolled term infants ($n = 27$) who were receiving invasive mechanical ventilation for other disease conditions, including neuromuscular or congenital cardiac defects correction. These infants showed no evidence of chronic lung disease. There was no randomization and no intervention. We excluded neonates with major congenital malformations, chromosomal anomalies, and complex congenital heart defects from the BPD group. We also excluded patients with PMA higher than 60 weeks at the time of sample collection.

### Specimen collection and processing

Tracheal aspirates were collected from all subjects following routine suctioning. Pertinent clinical information was obtained from the electronic medical record, including gestational age (GA) at birth, birth weight, PMA and fraction of inspired oxygen ($FiO_2$) at the time of sample collection, type of respiratory support, and information relevant to the patient's lung health. Samples were stored at −80 °C until analysis.

### MiRNA purification and quality control

MiRNAs were purified from 500 µl of TAs using the Norgen miRNA Purification Kit, after addition of a spike-in control (cel-mir-39, QIAGEN). This kit is optimized for

purification of small RNAs from a variety of samples that are low-yield and contain high concentrations of inhibitors. Purified miRNAs were quantified by Nanodrop, and miRNA quality was confirmed by bioanalyzer analysis at the Penn State College of Medicine Genome Sciences Core Facility.

## MicroRNA profiling

Small RNAs were retrotranscribed with the miScriptII RT kit (QIAGEN) from 250 ng of RNA. The expression of 1066 human miRNAs was quantified using miScript miRNA PCR custom arrays (QIAGEN, MIHS-3216Z). This kit enables sensitive high-throughput expression profiling from minimal amounts of starting material. PCR arrays were run on a QuantStudio 12K Flex system (Life Technologies, Carlsbad, CA) following the manufacturer's protocol. For each miRNA, cycle threshold (Ct) values were normalized to the average Ct of five endogenous controls, and fold changes were calculated using the $2^{-\Delta\Delta CT}$ method [26]. Datasets were uploaded to the Gene Expression Omnibus under project GSE156055 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE156055.

## MiRNA expression validation assays

The top differentially expressed miRNAs were selected, and their expression was validated in a subset of samples using the All-in-One™ miRNA qRT-PCR Reagent Kits and validated primers (GeneCopoeia Inc.), from 6 ng of mRNA, following the manufacturer's protocol. The following miRNAs were selected for validation: hsa-miR-627-5p, hsa-miR-3132, hsa-miR-3131, hsa-miR-183-3p, and hsa-miR-16-1-3p, which was used as normalization control [27]. Differential expression was determined using the $2^{-\Delta\Delta CT}$ method [26]. Significant differences were determined by *t*-test using the GraphPad Prism software.

## Library preparation and sequencing for mRNA

Total RNA was extracted from 500 µl of TA using the Plasma/Serum RNA Purification kit (Norgen Biotek) as per the manufacturer's instructions. Optical density values of extracted RNA were measured using NanoDrop (Thermo Scientific) to confirm an A260:A280 ratio above 1.9. RNA integration number was measured using BioAnalyzer (Agilent Technologies) RNA 6000 Nano Kit. Library preparation and sequencing were performed at the Genome Sciences and Bioinformatics Core of the Institute of Personalized Medicine at the Penn State College of Medicine. The cDNA libraries were prepared using the NuGEN's Ovation® SoLo RNA-Seq System (Tecan Genomics, Inc.). This system offers a built-in Unique Molecular Identifier (UMI) application as well as a depletion of rRNA and other high-abundant transcripts using the proprietary InDA-C technology. UMI was used to eliminate possible PCR duplicates in sequencing datasets and therefore facilitate unbiased gene expression profiling. The basic principle behind the UMI deduplication step is to collapse reads with identical mapping coordinates and UMI sequences. This step helps increase the accuracy of sequencing read counts for downstream analysis of gene expression levels. The processed libraries were assessed for its size distribution and concentration using BioAnalyzer High Sensitivity DNA Kit (Agilent Technologies). The libraries were pooled and diluted to 3 nM using 10 mM Tris-HCl, pH 8.5 and then denatured using the Illumina protocol. The denatured libraries were loaded onto an S1 flow cell on an Illumina NovaSeq 6000 (Illumina) and run for $2 \times 50$ cycles according to the manufacturer's instructions.

## Gene expression validation studies

The expression of selected differentially expressed genes was measured in a subset of samples, from 25 ng of cDNA by Real Time PCR with TaqMan™ assays (Life Technologies). The following probes were used: TFF3 (assay Hs00902278), HLA-DQB1 (assay Hs03054971), SCGB3A1 (assay Hs00369360), CD177 (assay Hs00360669), FFAR2 (assay Hs00271142), and NFKBIA (assay Hs00355671). A housekeeping gene, 18S (assay Hs03003631) was measured as a normalization control, from 2.5 ng of cDNA. The reactions were conducted in duplicate using the TaqMan™ Fast Advanced Master Mix in 10 µl of final volume, following the manufacturer's protocol. Expression results (Ct values) were monitored and extracted using the QuantStudio 12K Flex Software, and data were analyzed using the relative quantification method [26].

## Data analysis

For miRNA arrays, statistical analyses were performed with the R software using the Bioconductor *limma* package, and differential expression was defined as a Benjamini-Hochberg False Discovery Rate (FDR) < 0.05. For RNA-seq, de-multiplexed and adapter-trimmed sequencing reads were generated using Illumina bcl2fastq (released version 2.18.0.12) allowing no mismatches in the index read. BBDuk was used to trim/filter low quality sequences using "qtrim=lr trimq=10 maq=10" option. Alignment of the filtered reads to the human reference genome (GRCh38) was done using HISAT2 (version 2.1.0) [28], applying no-mixed and no-discordant options. UMI specific workflow that was developed and distributed by NuGEN was used to extract reads that were free from PCR artifacts (i.e., deduplication). The resulting deduplicated reads were

summarized to each gene using HTSeq [29] by supplementing Ensembl gene annotation (GRCh38.78). Differential gene expression analysis was conducted with the edgeR package on R [30]. Data were normalized with the TMM method [31]. Differential expression was defined as a Benjamini-Hochberg FDR < 0.05. Heatmaps were generated with the nonnegative matrix factorization package on R [32]. We uploaded datasets to the Gene Expression Omnibus under project GSE156028 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156028.

### Ingenuity pathway analysis (IPA)

The IPA software (IPA, Qiagen Redwood City, www.qiagen.com/ingenuity) was used to assess mRNA and miRNA-regulated target genes, gene interaction networks, top diseases, and molecular functions based on prediction scores [33]. IPA general settings for core analyses were: Ingenuity Knowledge base genes only (as reference set), and experimentally observed and/or highly predicted (as confidence filter). Correlation of expression patterns of miRNAs and differentially expressed transcripts were performed with logarithmic fold changes and adjusted *P* values.

## Results

### Patient demographics

We obtained TAs from 28 preterm infants at increased risk for developing BPD who needed invasive ventilatory support even after 28 days of life. The TA samples from controls ($n = 26$) were collected at any age below 60 weeks PMA while the infants were receiving invasive mechanical ventilation due to other conditions. The maternal and infant characteristics of the subjects are shown in Table 1, and individual data are visualized in Supplementary fig. S1. As expected, the BPD group had younger GA at birth (mean GA 25.7 weeks ± 1.6, $P < 0.0001$) and very low birth weight (mean BW 744 g ± 212, $P < 0.0001$) in comparison to the C group. The PMA at the time of sample collection was also significantly lower in the BPD vs. C group. The $FiO_2$ at the time of sample collection was similar for both groups. There was a slight male preponderance in the full sample (53.6% and 74.1% for the BPD and C groups, respectively, $P = 0.1625$) (Table 1).

**Table 1** Patient characteristics at study enrollment.

| Characteristic | Control ($n = 27$) | Severe BPD ($n = 28$) | *p* value |
|---|---|---|---|
| Gestational Age (GA) at Birth, weeks (mean ± SD) | 38.2 ± 1.7 | 25.7 ± 1.6 | **<0.0001** |
| Postmenstrual age at sample collection time, weeks (mean ± SD) | 41.2 ± 5.0 | 34.5 ± 6.6 | **<0.0001** |
| Birth Weight, g (mean ± SD) | 3156 ± 583 | 744 ± 212 | **<0.0001** |
| Male Sex (%) | 74.1 | 53.6 | 0.1625 |
| Racial/Ethnic Group (%) | | | |
| Non-Hispanic White | 85.2 | 42.9 | **0.0018** |
| Non-Hispanic Black | 3.7 | 3.6 | >0.9999 |
| Hispanic | 7.4 | 32.1 | **0.0403** |
| Other | 3.7 | 21.4 | 0.1012 |
| Antenatal Steroids Exposure (%) | n/a | 75 | – |
| Delivered via C-section (%) | 56 | 82 | **0.0437** |
| FiO2 at sample collection time | 0.34 ± 0.18 | 0.36 ± 0.13 | 0.7125 |

### MiRNA expression

Analysis of PCR arrays detected expression of 928 miRNAs in TAs from 51 samples (25 BPD, 26C). A heatmap of miRNA expression values is shown in Fig. 1A. Differential expression analysis revealed 36 miRNAs with at least twofold change (FC) between BPD vs. C and FDR < 0.05 (Table 2). Of these, 31 miRNAs were upregulated (log FC > 0.3, FDR < 0.05) in BPD vs. C samples, and 5 were downregulated (log FC < −0.3, FDR < 0.05) (Table 2). Log FC and FDR data for all miRNAs are also shown in Supplementary Table S1, and a volcano plot is shown in Fig. 1B. Validation experiments by Real-Time PCR confirmed upregulation of the top differentially expressed miRNAs hsa-miR-627-5p, hsa-miR-3132, and downregulation of hsa-miR-3131 (Fig. 1C). The expression of hsa-miR-323b-3p and hsa-miR-183* was not significantly different in the subset of samples used for validation, although it followed the expected trend (Fig. 1C).

### Transcriptomics analysis

RNAseq analysis detected expression of 64,253 transcripts in TAs from 38 samples (21 BPD, 17C). A summary of RNAseq data statistics is shown in Supplementary Table S2. A heatmap showing log2 of read counts per sample for the 33,310 transcripts with >20 read counts is shown in Fig. 2A. Differential expression analysis revealed 33 transcripts (11 downregulated, 22 upregulated) with FDR < 0.05 (Table 3). A volcano plot of differential transcript expression between BPD and C samples is shown in Fig. 2B. The majority of the differentially expressed transcripts were long intergenic non-coding RNAs and pseudogenes, and few were protein-coding genes, which likely represent the cellular fraction of the TA samples [34, 35]. For the protein-coding transcripts found, we validated the expression of two upregulated (TFF3, SCGB3A1) and 3 downregulated genes (FFAR2, CD177, NKBIA) in a subset

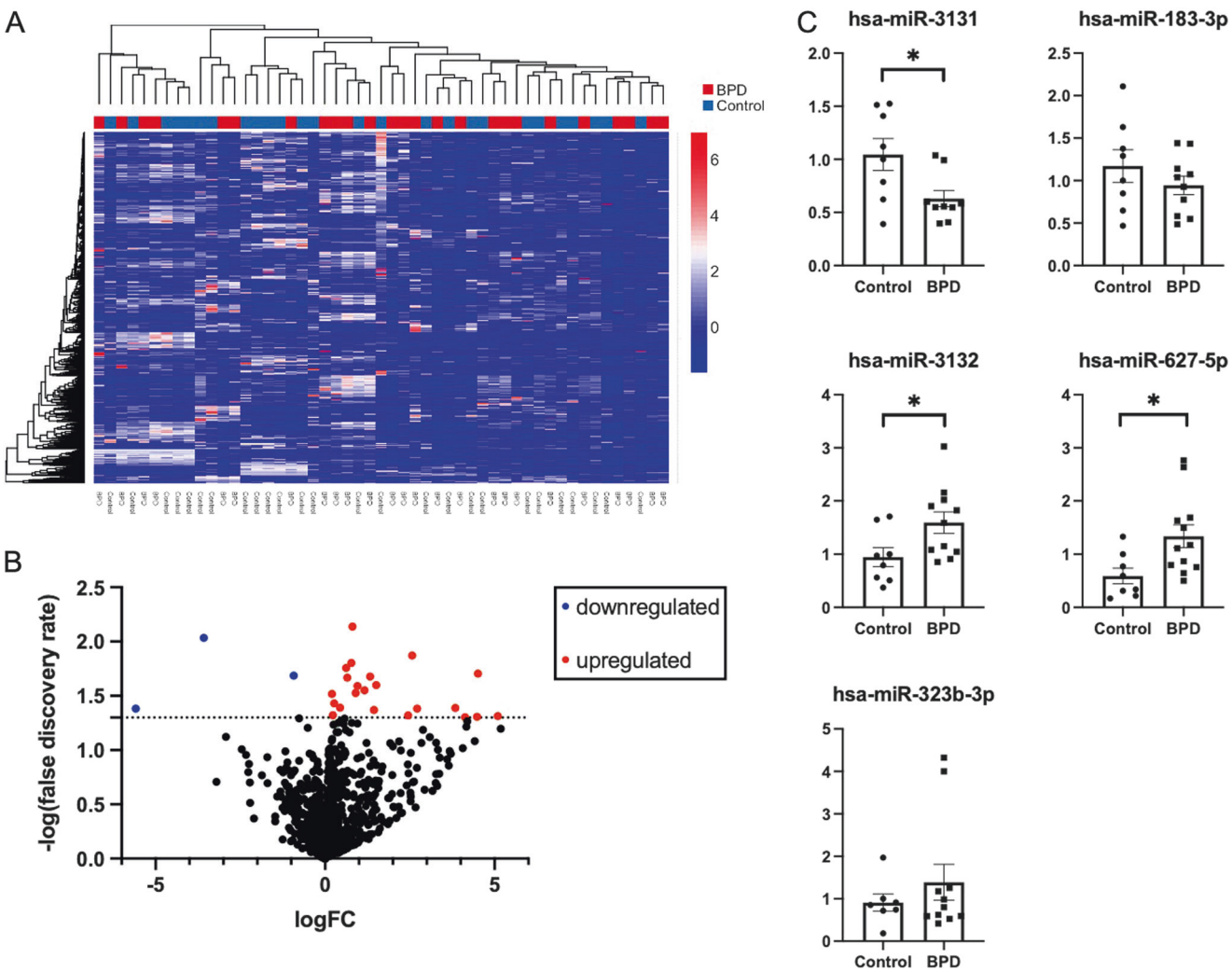


**Fig. 1 MiRNA expression in newborn tracheal aspirates. A** Heatmap of expression of 928 miRNAs (normalized by global mean) in TA samples from BPD ($n = 25$) and C ($n = 26$) subjects. **B** Volcano plot (red dots = downregulated miRNAs, blue dots = upregulated miRNAs). The dotted line indicates FDR = 0.05. **C** MiRNA expression validated by Real Time PCR in BPD ($n = 11$) and C ($n = 8$) samples.

of 26 samples (11 C, 16 BPD) (Fig. 2C). The expression of HLA-Q1 could not be validated due to low amplification.

## Pathway analysis

IPA analysis of differentially expressed miRNAs in BPD vs. control revealed significant associations with cell cycle, growth, morphology, and development functions (Table 4). The top physiological pathways associated with differentially expressed miRNAs are organismal development and survival, embryonic development, tissue morphology, and organismal function (Table 4). The top network including most differentially expressed miRNAs in BPD vs. C (six upregulated and one downregulated) is associated with inflammatory disease and inflammatory response (Table 4, Fig. 3A). Similarly, IPA analysis of differentially expressed genes identified by RNAseq also showed molecular and cellular functions associated with cell cycle, growth and proliferation, gene expression, and cell signaling (Table 5). The top physiological functions included tissue morphology and development, as well as organismal functions. The top associated networks were cell-to-cell signaling and interaction and immune cell trafficking, encompassing eight downregulated and six upregulated genes in BPD vs. C (Fig. 3B), as well as cellular assembly and organization and embryonic development, encompassing three upregulated genes (Table 5, Supplementary Fig. S2).

MiRNA-mRNA networks: When combining miRNA and mRNA pathway analyses, we found that the top network associated with the combined dataset was also Inflammatory disease and inflammatory response, with a score of 18. In addition, IPA identified the OX40 Signaling Pathway, B cell development, and iCOS-iCOSL Signaling in T helper cells, among the top canonical pathways

**Table 2** Differentially expressed miRNAs in BPD vs. C.

| Upregulated miRNAs in BPD vs. C ($n = 19$) | | | |
|---|---|---|---|
| miRNA ID | Log FC | Avg Exp | FDR |
| hsa-miR-323b-3p | 5.09 | 7.9 | 0.049 |
| hsa-miR-627 | 4.51 | 5.04 | 0.02 |
| hsa-miR-3132 | 4.48 | 8.23 | 0.049 |
| hsa-miR-26b* | 3.85 | 5.79 | 0.041 |
| hsa-miR-890 | 2.72 | 3.13 | 0.042 |
| hsa-miR-505* | 2.57 | 2.81 | 0.013 |
| hsa-miR-32* | 2.45 | 3.02 | 0.048 |
| hsa-miR-1911 | 1.51 | 2.49 | 0.025 |
| hsa-miR-3682-3p | 1.44 | 2.53 | 0.043 |
| hsa-miR-3921 | 1.33 | 1.2 | 0.021 |
| hsa-miR-548k | 1.16 | 1.42 | 0.028 |
| hsa-miR-320a | 0.96 | 1.24 | 0.026 |
| hsa-miR-1293 | 0.9 | 0.77 | 0.03 |
| hsa-miR-145* | 0.81 | 0.83 | 0.007 |
| hsa-miR-514b-5p | 0.77 | 0.91 | 0.016 |
| hsa-miR-25* | 0.66 | 0.56 | 0.022 |
| hsa-miR-100 | 0.62 | 0.75 | 0.017 |
| hsa-miR-20a* | 0.45 | 0.6 | 0.041 |
| hsa-miR-539 | 0.27 | 0.34 | 0.037 |
| Downregulated miRNAs in BPD vs. C ($n = 3$) | | | |
| miRNA ID | Log FC | Avg Exp | FDR |
| hsa-miR-223 | −5.59 | 10.95 | 0.041 |
| hsa-miR-3131 | −3.58 | 2.87 | 0.009 |
| hsa-miR-183* | −0.92 | 0.93 | 0.021 |

associated with these differentially expressed genes and miRNAs. Combined, these differentially expressed molecules displayed associations with several miRNAs via gene regulatory networks. An analysis of miRNAs associated with respiratory disease are shown in Supplementary Table S3.

## Discussion

Although BPD results from arrested lung growth in preterm birth, it is a multifactorial disease resulting from complex interplay of antenatal, perinatal and postnatal insults [36]. It is also becoming clearer that genetic, epigenetic, and environmental factors are all likely contributory factors to the disease severity [37]. The underlying mechanisms in the development of BPD are yet to be fully elucidated. In this pilot exploratory study, we evaluated miRNA expression and transcriptomics in TAs and uncovered differential regulatory pathways in BPD vs. C groups. We found differential expression of miRNAs in TAs of preterm infants with BPD versus term controls. The pathways uncovered in our analyses show molecules that are involved in cellular development, growth, and proliferation, embryonic development and tissue morphology, as well as gene expression regulation. We also found associated network functions including immunological and inflammatory disease, inflammatory response, cellular development, cellular growth and proliferation, and embryonic development. Together, these findings reveal that TAs express markers of pathways known to be associated with BPD and extreme prematurity and the complex interplay of inflammation and lung growth and development arrest as hallmarks of the disease.

Our study revealed 22 differentially expressed miRNAs in TAs from BPD vs. C neonates. These included several miRNAs that were previously reported in BPD clinical studies and animal models (miR-548, miR-3131, miR-152, mir-3129-5p, and miR-20) [17, 18, 38–41]. We also found a few novel miRNAs, never reported in BPD (miR-183, miR-32, and miR-890). Some of the canonical pathways associated with these miRNAs were related to cellular morphology, growth, proliferation, and organ development. In a recent systematic review and meta-analysis, miRNA expression from lung tissue versus control animal models of BPD was analyzed in three independent studies [42]. The authors reported that 89 miRNAs were differentially expressed between the groups (77 were up-regulated and 12 down-regulated), all playing a role in lung development and homeostasis. On the other hand, only a few published studies have reported miRNA expression obtained from biological samples in extremely premature human neonates at risk for developing BPD, with very few using TAs [18, 40].

Cell division is very critical in the pathogenesis of BPD, due to its involvement in lung development in the developing fetus. The severe BPD phenotype noted in our cohort of extremely born preterm infants (23-28 weeks GA) likely results from the consequences of disruption of lung development in the canalicular and saccular stages. Our data revealed 3 miRNAs (miR-3682-3p, miR-505-3p, and miR-323b-3p) associated with cell proliferation pathways in the BPD vs. C TAs, indicating that these miRNAs could be involved in lung development disruption resulting from prematurity and/or BPD. While we know through observation studies that BPD is more common in those infants with intrauterine growth restriction [43, 44], we have also learned that infants with BPD continue to have poorer growth until post discharge from the neonatal ICU [45]. Our transcriptomic data also revealed cell cycle, cell development, cell-to-cell signaling/interaction, and embryonic and organ development as top gene networks associated with the 33 differentially expressed genes in BPD vs. C. The majority of these genes were involved in inflammatory responses and organismal injury and abnormalities, and we

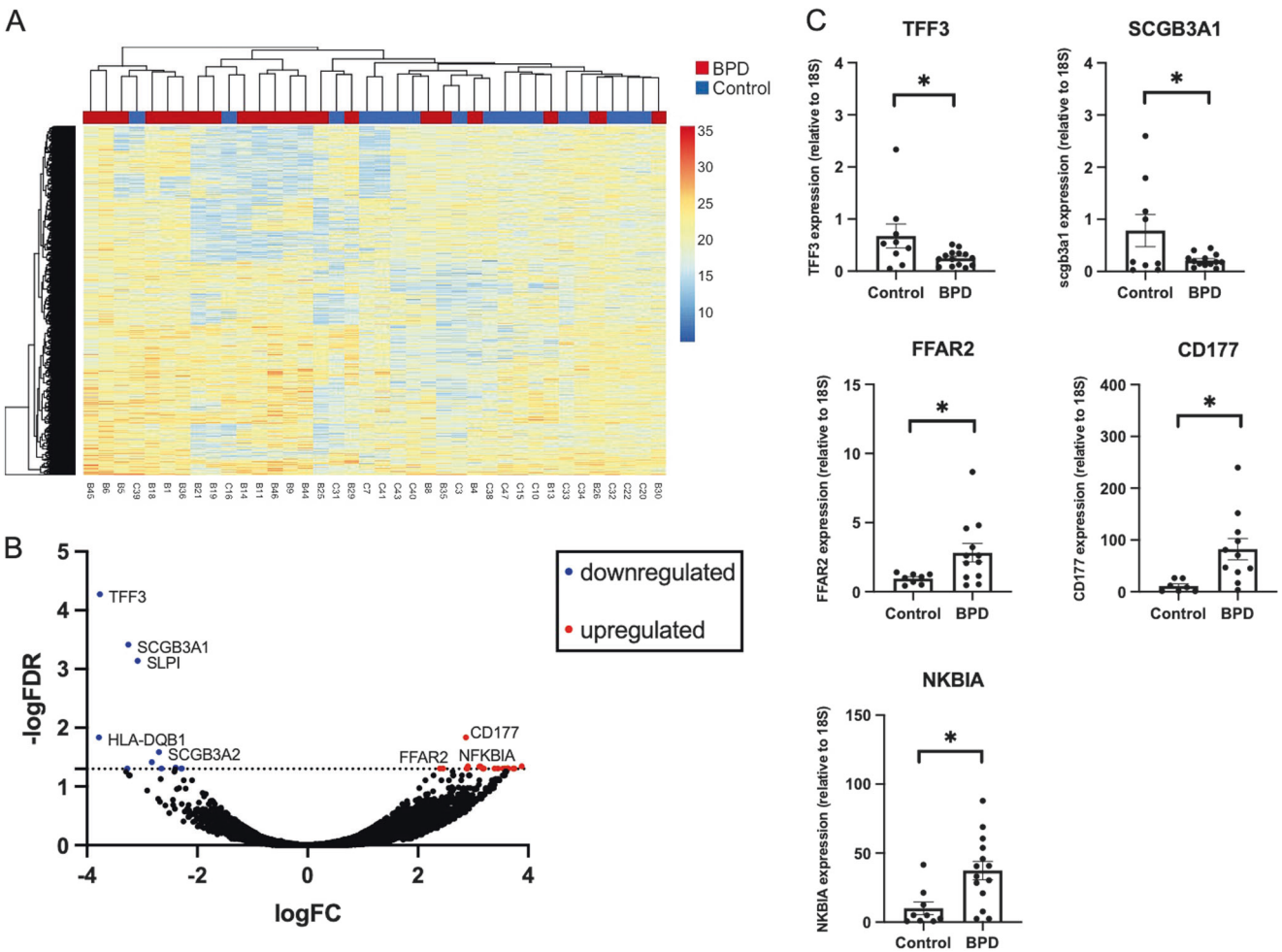


**Fig. 2 Gene expression in newborn tracheal aspirates. A** Heatmap of log2(read counts >10) for 33,310 transcripts in TAs samples from BPD ($n = 21$) and C ($n = 17$) subjects. **B** Volcano plot (red dots = downregulated miRNAs, blue dots = upregulated genes). The dotted line indicates FDR = 0.05. **C** mRNA expression validated by Real Time PCR in BPD ($n = 14$) and C ($n = 10$) samples.

also found involvement of the OX40 canonical pathway, better known as the tumor necrosis factor receptor superfamily, member 4 (TNFRSF4 or CD134) pathway. Because delay in alveolar development is a major characteristic of BPD, the OX40 signaling pathway may play an important role in the disease development and progression [46]. OX40 also plays a key role in the survival and homeostasis of effector and memory T cells in autoimmunity [47]. It has also been reported that during the alveolarization phase of lung development, neonatal dendritic cells (DCs) display increased levels of OX40L and this is related to their preferential ability to promote $T_H2$ responses [47, 48]. While we lack information on the cell populations present in the TAs, it is likely that the differentially expressed transcripts (and miRNAs) found are representative of specific cell populations [35].

To address relevant miRNA–mRNA interactions in BPD pathogenesis, we conducted pathway analysis of differentially expressed miRNAs and mRNAs. We found multiple associations with lung injury and inflammatory response. Inflammatory markers have been studied extensively in animal models of BPD [49], pre and postnatal-infection, volutrauma, atelectotrauma, and oxygen toxicity. For example, the pro-inflammatory cytokines MCP-1 and MIP-1α can negatively affect the formation of alveolar capillaries [50]. More recently, miR199a-5p/miR-3129-5p expression in hyperoxia-exposed mice lungs, cells, and in TA samples from preterm infants with evolving BPD showed a reduction in the expression of its target, caveolin-1. It was also noted that treatment with miR199a-5p-mimic worsened lung injury secondary to hypoxemia with increased levels of inflammatory markers [38]. Furthermore, there is evidence of airway microbiome altering the metagenomic and metabolomic signatures in BPD [51]. We believe the miRNA mRNA data from our study represent the thumb print of complex interplay of inflammation, organogenesis, evolving airway microbiome and ongoing repair processes associated with severe BPD and extreme prematurity.

**Table 3** Differentially expressed genes in BPD vs. C.

| gene ID | log FC | FDR |
|---|---|---|
| Downregulated transcripts in BPD vs. C ($n = 11$) | | |
| HLA-DQB1 | −3.783 | 1.46E-02 |
| TFF3 | −3.765 | 5.31E-05 |
| Y_RNA | −3.268 | 4.95E-02 |
| SCGB3A1 | −3.252 | 3.85E-04 |
| SLPI | −3.084 | 7.24E-04 |
| AC114498.1 | −2.825 | 3.83E-02 |
| SCGB3A2 | −2.697 | 2.59E-02 |
| RPS23P8 | −2.652 | 4.95E-02 |
| MT-TN | −2.395 | 4.77E-02 |
| CRNN | −2.291 | 4.95E-02 |
| MT-TA | −2.281 | 4.95E-02 |
| Upregulated transcripts in BPD vs. C ($n = 22$) | | |
| RP11-547C13.1 | 3.881 | 4.54E-02 |
| RP1-38C16.2 | 3.748 | 4.95E-02 |
| SIGLEC31P | 3.727 | 4.95E-02 |
| RP11-78O7.1 | 3.639 | 4.95E-02 |
| RP11-32A1.2 | 3.627 | 4.95E-02 |
| FTH1P5 | 3.625 | 4.95E-02 |
| RP4-594L9.2 | 3.623 | 4.95E-02 |
| ACTBP6 | 3.614 | 4.95E-02 |
| CTD-2026G22.1 | 3.574 | 4.95E-02 |
| SNORA11C | 3.542 | 4.92E-02 |
| TCTEX1D4 | 3.452 | 4.95E-02 |
| C2CD4B | 3.39 | 4.95E-02 |
| ISPD-AS1 | 3.194 | 4.95E-02 |
| WASH2P | 3.175 | 4.95E-02 |
| RP11-358H9.1 | 3.136 | 4.58E-02 |
| RP11-131M11.2 | 3.112 | 4.58E-02 |
| CKAP4 | 2.903 | 4.54E-02 |
| CLEC2B | 2.891 | 4.95E-02 |
| TRIM63 | 2.885 | 4.95E-02 |
| CD177 | 2.868 | 1.46E-02 |
| FFAR2 | 2.454 | 4.95E-02 |
| NFKBIA | 2.396 | 4.95E-02 |

**Table 4** IPA pathways associated with differentially expressed miRNAs in BPD vs. C.

| Top molecular and cellular functions | *P* value |
|---|---|
| Cell cycle | 2.49E-05–2.49E-05 |
| Cell development | 1.65E-02–8.26E-04 |
| Cellular growth and proliferation | 8.26E-04–8.26E-04 |
| Cell morphology | 4.47E-03–4.47E-03 |
| Gene expression | 1.27E-02–1.27E-02 |
| **Top Physiological system development and function** | ***P* value** |
| Organismal development | 2.56E-03–2.56E-03 |
| Organismal survival | 1.46E-02–1.46E-02 |
| Embryonic development | 1.65E02–1.65E-02 |
| Organismal functions | 2.66E-02–2.53E-02 |
| Tissue morphology | 3.34E-02–3.34E-02 |
| **Top associated networks** | **Score** |
| Glomerular injury, inflammatory disease, inflammatory response | 18 |

The current study has several strengths. We prospectively enrolled infants born extremely preterm and at highest risk of developing BPD, who needed invasive ventilation for at least 28 days after birth. Furthermore, we utilized TAs, a readily available medical waste collected routinely from intubated infants receiving invasive mechanical ventilation, as the biological fluid of choice. One advantage of TAs vs. other biofluids such as serum samples is that they may represent a source of biomarkers local to the pathogenesis of the respiratory system, rather than systemic markers of other comorbidities associated with prematurity. Because the control group consisted of term infants without chronic lung disease, these markers may represent specific mechanisms of lung disease of prematurity.

Our study also has several limitations. This was a single-center study that was limited by a small sample size affecting the generalizability of the study. There was also a male preponderance in our sample, which could be due to sampling bias or the fact that male neonates tend to have higher morbidity compared to their female counterparts. Furthermore, our study population included the sickest cohort since they were dependent on invasive mechanical ventilation in order to obtain TA samples. We were unable to collect TAs from preterm infants without BPD or with milder forms of the disease that only received non-invasive mechanical ventilation throughout their hospitalization. Thus, our results may be confounded by the significant difference in GA and PMA at sample collection for both groups. We also lacked data on postnatal infections, postnatal growth failure, maternal chorioamnionitis, and the receipt of glucocorticoids after enrollment and before 36 weeks PMA. Additional limitations to the generalizability of this study include the fact that a large proportion of the enrolled BPD subjects were of Hispanic ethnicity, whereas the C subjects were primarily Caucasians of European descent. Although unlikely, it is possible that the miRNA and gene expression shown in the study could have been affected by race. Lastly, we have not determined whether the transcripts and miRNAs detected are originated from cells present in the TAs vs. extracellular vesicles, or are a result of expression from the oral microbiota, as suggested by the

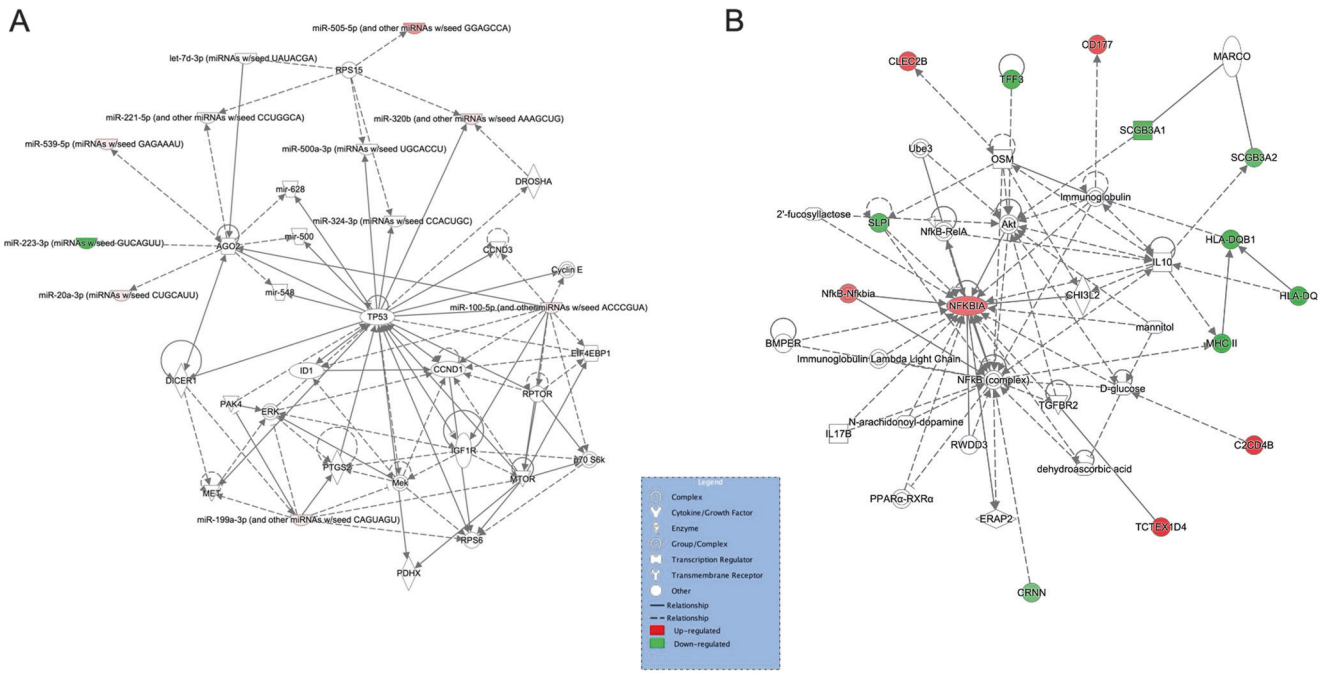


**Fig. 3 Network analysis of differentiallly expressed genes and miRNAs.** Ingenuity Pathway Analysis top associated networks for differentially expressed miRNAs (**A**) and transcripts (**B**) between BPD vs. C samples. **A** Glomerular injury, inflammatory disease, inflammatory response pathway. **B** Cell-to-cell signaling and interaction, immune cell trafficking pathway.

**Table 5** IPA pathways associated with differentially expressed mRNA in BPD vs. C.

| Top molecular and cellular functions | *P* value |
|---|---|
| Cell cycle | 4.10E-02–3.94E-05 |
| Gene expression | 4.10E-02–3.94E-04 |
| Cellular movement | 4.73E-02–7.88E-04 |
| Cell-to-cell signaling and interaction | 4.47E-02–1.18E-03 |
| Cell death and survival | 2.25E-02–2.52E-03 |
| **Top Physiological system development and function** | ***P* value** |
| Organismal functions | 1.57E-03–4.46E-04 |
| Tissue morphology | 2.49E-02–4.46E-04 |
| Skeletal and muscular system development | 1.41E02–1.57E-03 |
| Tissue development | 3.72E-02–2.05E-03 |
| **Top associated networks** | **Score** |
| Cell-to-cell signaling and interaction, immune cell trafficking | 30 |
| Cellular assembly and organization, embryonic development | 6 |

high alignment rate to bacteria in our RNAseq data [52]. Future experiments using different RNA extraction, sample processing, and sample storage methods that enrich for specific cell types will determine whether the results presented here are the result of sample quality issues and/or technical matters. Our RNAseq experiments also showed very low alignment rates to mRNA and coding regions, and a large proportion of alignment to intergenic regions and introns. Thus, our transcriptomics results have a large proportion of non-coding RNAs and pseudogenes. However, the main findings from our analysis were validated using specific human Taqman probes and miRNA assays.

In summary, we showed differential miRNA and mRNA expression in TA samples obtained from extreme preterm infants with BPD and their associations with pathways involved in their complex phenotype. We found unique candidate genes and miRNAs expressed specifically in TAs of extreme preterm infants with BPD vs. term controls, and the predicted pathways they regulate. Future studies using similar samples are needed to evaluate the significance of posttranscriptional regulation in BPD, as well as the clinical significance, contributions to disease progression, and potential use as therapeutic targets of the mRNA and miRNAs identified here.

## Code availability

The code used for data analysis in the current study can be found in the Silveyra lab repository, available at http://psilveyra.github.io/silveyralab/.

**Acknowledgements** The authors thank Susan DiAngelo for assistance with sample and data collection. The authors thank the Pennsylvania State University College of Medicine Genome Sciences Core Facility for RNAseq analysis and Real Time PCR equipment.

**Funding** This work was supported by grants from Children's Miracle Network (PS, CNO), Center for Research for Women and Newborn Health (PS, NF, RS), and Penn State College of Medicine faculty endowment funds (CNO, PS, RS).

**Author contributions** CNO participated in design of work, acquired samples, analyzed and interpreted data, drafted the work, and revised the work; RS participated in design of work, acquired samples, analyzed data, interpreted work, drafted the work, and revised the work; DTM purified samples, conducted validation studies, analyzed data, interpreted work, drafted the work, prepared figures, and revised the work; MAP purified and annotated samples, obtained array data, analyzed the work, and revised the work; DS acquired samples, obtained data, and revised the work; AD acquired samples, obtained data, and revised the work; NF conducted pathway analysis, created tables, and revised the work; YIK conducted RNA sequencing experiments and revised the work, JAH participated in design of work, and revised the work, NJT participated in design of work, and revised the work; PS led all stages of work from initial conceptualization, study design, data generation, transcriptomics, bioinformatics analysis and interpretation, and manuscript generation and revision.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

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

## References

1. Been JV, Debeer A, van Iwaarden JF, Kloosterboer N, Passos VL, Naulaers G, et al. Early alterations of growth factor patterns in bronchoalveolar lavage fluid from preterm infants developing bronchopulmonary dysplasia. Pediatr Res. 2010;67:83–9.
2. Jobe AJ. The new BPD: an arrest of lung development. Pediatr Res. 1999;46:641–3.
3. Stenmark KR, Abman SH. Lung vascular development: implications for the pathogenesis of bronchopulmonary dysplasia. Annu Rev Physiol. 2005;67:623–61.
4. Glass HC, Costarino AT, Stayer SA, Brett CM, Cladis F, Davis PJ. Outcomes for extremely premature infants. Anesthesia analgesia. 2015;120:1337–51.
5. Jankov RP, Keith Tanswell A. Growth factors, postnatal lung growth and bronchopulmonary dysplasia. Paediatr Respir Rev. 2004;5:S265–75.
6. Northway WH, Rosan RC, Porter DY. Pulmonary disease following respirator therapy of hyaline-membrane disease. Bronchopulmonary dysplasia. N Engl J Med. 1967;276:357–68.
7. Jensen EA, Schmidt B. Epidemiology of bronchopulmonary dysplasia. Birth Defects Res A Clin Mol Teratol. 2014;100:145–57.
8. McEvoy CT, Jain L, Schmidt B, Abman S, Bancalari E, Aschner JL. Bronchopulmonary dysplasia: NHLBI workshop on the primary prevention of chronic lung diseases. Ann Am Thorac Soc. 2014;11:S146–53.
9. Akangire G, Manimtim W, Nyp MF, Noel-MacDonnell J, Kays AN, Truog WE, et al. Clinical outcomes among diagnostic subgroups of infants with severe bronchopulmonary dysplasia through 2 years of age. Am J Perinatol. 2018;35:1376–87.
10. Bancalari E, Jain D. Bronchopulmonary dysplasia: 50 years after the original description. Neonatology. 2019;115:384–91.
11. Ryan RM, Feng R, Bazacliu C, Ferkol TW, Ren CL, Mariani TJ, et al. Black race is associated with a lower risk of bronchopulmonary dysplasia. J Pediatr. 2019;207:130–5.e132.
12. Rivera L, Siddaiah R, Oji-Mmuo C, Silveyra GR, Silveyra P. Biomarkers for bronchopulmonary dysplasia in the preterm infant. Front Pediatr. 2016;4:33.
13. Alles J, Fehlmann T, Fischer U, Backes C, Galata V, Minet M, et al. An estimate of the total number of true human miRNAs. Nucleic Acids Res. 2019;47:3353–64.
14. Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, et al. Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol. 2009;7:e1000238.
15. Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ. Biological functions of microRNAs: a review. J Physiol Biochem. 2011;67:129–39.
16. Zhang X, Chu X, Gong X, Zhou H, Cai C. The expression of miR-125b in Nrf2-silenced A549 cells exposed to hyperoxia and its relationship with apoptosis. J Cell Mol Med. 2020;24:965–72.
17. Wang J, Yin J, Wang X, Liu H, Hu Y, Yan X, et al. Changing expression profiles of mRNA, lncRNA, circRNA, and miRNA in lung tissue reveal the pathophysiological of bronchopulmonary dysplasia (BPD) in mouse model. J Cell Biochem. 2019;120:9369–80.
18. Lal CV, Olave N, Travers C, Rezonzew G, Dolma K, Simpson A, et al. Exosomal microRNA predicts and protects against severe bronchopulmonary dysplasia in extremely premature infants. JCI Insight. 2018;3:e93994.
19. Sayed D, Abdellatif M. MicroRNAs in development and disease. Physiol Rev. 2011;91:827–87.
20. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. Clin Chem. 2010;56:1733–41.
21. Tomankova T, Petrek M, Kriegova E. Involvement of microRNAs in physiological and pathological processes in the lung. Respir Res. 2010;11:159.
22. Bhattacharya S, Go D, Krenitsky DL, Huyck HL, Solleti SK, Lunger VA, et al. Genome-wide transcriptional profiling reveals connective tissue mast cell accumulation in bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2012;186:349–58.
23. Pietrzyk JJ, Kwinta P, Wollen EJ, Bik-Multanowski M, Madetko-Talowska A, Günther CC, et al. Gene expression profiling in preterm infants: new aspects of bronchopulmonary dysplasia development. PLoS ONE. 2013;8:e78585.
24. Fulton CT, Cui TX, Goldsmith AM, Bermick J, Popova AP. Gene expression signatures point to a male sex-specific lung mesenchymal cell PDGF receptor signaling defect in infants developing bronchopulmonary dysplasia. Sci Rep. 2018;8:17070.
25. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2001;163:1723–9.
26. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25:402–8.
27. Schwarzenbach H, da Silva AM, Calin G, Pantel K. Data normalization strategies for microRNA quantification. Clin Chem. 2015;61:1333–42.
28. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
29. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

31. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.
32. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinform. 2010;11:1–9.
33. Fuentes N, Roy A, Mishra V, Cabello N, Silveyra P. Sex-specific microRNA expression networks in an acute mouse model of ozone-induced lung inflammation. Biol Sex Differ. 2018;9:18.
34. Nathe KE, Mancuso CJ, Parad R, Van Marter LJ, Martin CR, Stoler-Barak L, et al. Innate immune activation in neonatal tracheal aspirates suggests endotoxin-driven inflammation. Pediatr Res. 2012;72:203–11.
35. Eldredge LC, Creasy RS, Presnell S, Debley JS, Juul SE, Mayock DE, et al. Infants with evolving bronchopulmonary dysplasia demonstrate monocyte-specific expression of IL-1 in tracheal aspirates. Am J Physiol Lung Cell Mol Physiol. 2019;317:L49–56.
36. Merritt TA, Deming DD, Boynton BR. The 'new' bronchopulmonary dysplasia: challenges and commentary. Semin Fetal Neonatal Med. 2009;14:345–57.
37. Ambalavanan N, Cotten CM, Page GP, Carlo WA, Murray JC, Bhattacharya S, et al. Integrated genomic analyses in bronchopulmonary dysplasia. J Pediatr. 2015;166:531–7.e513.
38. Alam MA, Betal SGN, Aghai ZH, Bhandari V. Hyperoxia causes miR199a-5p-mediated injury in the developing lung. Pediatr Res. 2019;86:579–88.
39. Zhang X, Peng W, Zhang S, Wang C, He X, Zhang Z, et al. MicroRNA expression profile in hyperoxia-exposed newborn mice during the development of bronchopulmonary dysplasia. Respir Care. 2011;56:1009–15.
40. Wu YT, Chen WJ, Hsieh WS, Tsao PN, Yu SL, Lai CY, et al. MicroRNA expression aberration associated with bronchopulmonary dysplasia in preterm infants: a preliminary study. Respir Care. 2013; 58:1527–35.
41. Dravet-Gounot P, Morin C, Jacques S, Dumont F, Ely-Marius F, Vaiman D, et al. Lung microRNA deregulation associated with impaired alveolarization in rats after intrauterine growth restriction. PLoS ONE. 2017;12:e0190445.
42. Yang Y, Qiu J, Kan Q, Zhou XG, Zhou XY. MicroRNA expression profiling studies on bronchopulmonary dysplasia: a systematic review and meta-analysis. Genet Mol Res. 2013;12:5195–206.
43. Lio A, Rosati P, Pastorino R, Cota F, Tana M, Tirone C, et al. Fetal Doppler velocimetry and bronchopulmonary dysplasia risk among growth-restricted preterm infants: an observational study. BMJ Open. 2017;7:e015232.
44. Jung YH, Park Y, Kim BI, Choi CW. Length at birth z-score is inversely associated with an increased risk of bronchopulmonary dysplasia or death in preterm infants born before 32 gestational weeks: a nationwide cohort study. PLoS ONE. 2019;14: e0217739.
45. DeMauro SB, Jensen EA, Bann CM, Bell EF, Hibbs AM, Hintz SR, et al. Home oxygen and 2-year outcomes of preterm infants with bronchopulmonary dysplasia. Pediatrics. 2019;143: e20182956.
46. Collins JJP, Tibboel D, de Kleer IM, Reiss IKM, Rottier RJ. The future of bronchopulmonary dysplasia: emerging pathophysiological concepts and potential new avenues of treatment. Front Med. 2017;4:61.
47. Croft M, So T, Duan W, Soroosh P. The significance of OX40 and OX40L to T-cell biology and immune disease. Immunol Rev. 2009;229:173–91.
48. Jember AG, Zuberi R, Liu FT, Croft M. Development of allergic inflammation in a murine model of asthma is dependent on the costimulatory receptor OX40. J Exp Med. 2001;193:387–92.
49. Dong J, Carey WA, Abel S, Collura C, Jiang G, Tomaszek S, et al. MicroRNA-mRNA interactions in a murine model of hyperoxia-induced bronchopulmonary dysplasia. BMC Genom. 2012;13:204.
50. Compernolle V, Brusselmans K, Acker T, Hoet P, Tjwa M, Beck H, et al. Loss of HIF-2alpha and inhibition of VEGF impair fetal lung maturation, whereas treatment with VEGF prevents fatal respiratory distress in premature mice. Nat Med. 2002;8:702–10.
51. Lal CV, Kandasamy J, Dolma K, Ramani M, Kumar R, Wilson L, et al. Early airway microbial metagenomic and metabolomic signatures are associated with development of severe bronchopulmonary dysplasia. Am J Physiol Lung Cell Mol Physiol. 2018;315:L810–5.
52. Pan Y, Du L, Ai Q, Song S, Tang X, Zhu D, et al. Microbial investigations in throat swab and tracheal aspirate specimens are beneficial to predict the corresponding endotracheal tube biofilm flora among intubated neonates with ventilator-associated pneumonia. Exp Ther Med. 2017;14:1450–8.