Network and Systems Medicine 2025 | Vol. 1 | Iss. 1 | Pages: 4–19
DOI: 10.14293/NSM.25.1.0001

RESEARCH ARTICLE

Multi-omic Signatures Relate to the Severity of Pulmonary Outcome in Neonates Traced into Adult Disease

Juan Henao[1], Alida Kindt[2], Tanja Seegmüller[3], Kai Foerster[3,4,5], Andreas W. Flemmer[4], Juergen Behr[6], Nikolaus Kneidinger[6,7], Marion Frankenberger[8], Fabian Theis[1,9], Benjamin Schubert[1,a], Markus List[10,11,a]*, Anne Hilgendorff[3,5,8]*

[1]Institute for Computational Biology, Helmholtz Zentrum München, Member of the German Lung Research Center (DZL), Munich, Germany
[2]Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, 2333 CC Leiden, The Netherlands
[3]Center for Comprehensive Developmental Care at the iSPZ Hauner, Dr. von Haunersche Children's Hospital, Hospital of the Ludwig Maximilian University, Member of the German Lung Research Center (DZL), Munich, Germany
[4]Division of Neonatology, Dr. von Hauner Children's Hospital, LMU University Hospital, Member of the German Lung Research Center (DZL), LMU Munich, Germany
[5]Comprehensive Pneumology Center (CPC-M), Member of the German Center for Lung Research (DZL), Munich, Germany
[6]Department of Medicine V, University Hospital, Comprehensive Pneumology Center, LMU Munich, Member of the German Lung Research Center (DZL), Munich, Germany
[7]Division of Pulmonology, Department of Internal Medicine, Lung Research Cluster, Medical University of Graz, Graz, Austria
[8]Institute of Lung Health and Immunity and Comprehensive Pneumology Center with the CPC-M bioArchive, Helmholtz Zentrum München, Member of the German Lung Research Center (DZL), Munich, Germany
[9]Department of Mathematics, Technical University of Munich, 85748 Garching bei München, Germany
[10]Data Science in Systems Biology, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany
[11]Munich Data Science Institute (MDSI), Technical University of Munich, 85748 Garching, Germany

[a]Equal contributions.
***Correspondence to:** Markus List, Data Science in Systems Biology, TUM School of Life Sciences, Technical University of Munich, Maximus-von-Imhof Forum 3, 85354 Freising. E-mail: markus.list@tum.de
Anne Hilgendorff, Dr. v. Hauner University Children's Hospital, LMU Hospital, Member of the German Lung Research Center (DZL), Munich, Germany; Comprehensive Pneumology Center, Helmholtz Zentrum München, Member of the German Lung Research Center (DZL), Munich, Germany; School of Medicine and Health Sciences, Department of Paediatrics, Section of Neonatology, Pediatric Intensive Care, Pediatric Cardiology, Pediatric Pneumology and Allergology and Research Centre Neurosensory Science, Carl von Ossietzky University Oldenburg, Germany. E-mail: anne.hilgendorff@med.uni-muenchen.de

Received March 26, 2024; accepted November 22, 2024; published online May 23, 2025.

ABSTRACT

Chronic lung disease (CLD) i.e., bronchopulmonary dysplasia (BPD) is the most common long-term complication after preterm birth. This clinically heterogeneous disease is characterized by impaired development of the gas exchange area and the bronchial tree. The identification of disease endotypes or indicators of disease onset early after birth would allow for individualized monitoring and treatment. In a cohort of 55 preterm infants phenotypically described by detailed clinical data

on pregnancy, birth, and neonatal intensive care unit care until discharge, and a complete assessment of pulmonary and extrapulmonary morbidities, we analyzed 1120 proteins and 213 metabolites in samples obtained in the first weeks of life to characterize biological signatures of BPD. Latent factor analysis highlighted seven factors, three of which linked proteomic and metabolomic data, highlighting a common inflammatory/immune signature but no independent endotypes. We next used abundance patterns of differentially abundant proteins and metabolites and successfully identified biomarker candidates associated with disease severity including PC(O-36:5), CCL22, KIR3DL2, SCGF-alpha, and SCGF-beta. Confirmation of the discriminatory power of these biomarkers in adult CLD patients (n=44) using matched proteomic profiling suggests CCL22, KIR3DL2, and SCGF-beta as shared biomarker candidates of BPD and adult CLD.

KEYWORDS

bronchopulmonary dysplasia (BPD); preterm neonatal lung; lung development; disease endotypes; disease severity; biomarkers; chronic obstructive pulmonary disease (COPD); random forest; differential abundance analysis; multi-omics latent factor analysis.

INTRODUCTION

Bronchopulmonary dysplasia (BPD) is the most prevalent long-term complication of prematurity,[1,2] with a significant impact on long-term morbidity and mortality.[3] The disease is characterized by the impaired development of both the gas exchange area and the bronchial tree, caused by the impact of a combination of antenatal and postnatal risk factors, including infection as well as the exposure to life-saving treatments such as oxygen supplementation and mechanical ventilation.[1,4] Clinically, the most urgent need is to identify disease at the earliest stage possible and to better stratify and monitor patients according to their therapeutic needs.

The significant heterogeneity of the clinical picture together with the broad variety of the pathophysiologic mechanisms discovered suggest potential underlying disease endotypes.[2] Disease endotypes are defined as different pathophysiological mechanisms that converge into the same cluster of symptoms and manifest as a single (clinical) phenotype. The characterization of these endotypes likely results in improved patient stratification to enable personalized risk monitoring and treatment targeting distinct mechanisms in patient subpopulations.[5,6]

Different attempts were made to stratify such BPD endotypes clinically using structural and functional information generated by computer tomography, lung function as well as techniques for the assessment of vascular complications such as echocardiography and right heart catheter.[7–9] Although varying degrees of lung scaffold remodeling, vascular impairment, and airway disease characterized individuals across different BPD severity grades,[10–12] meaningful clinical endotypes were not deducted thus far and likely need to engage additional biomarkers to reach clinical significance. As a result, the NICHD/NHLBI/ORD (National Institute of Child Health and Human Development, National Heart, Lung and Blood Institute, Office of Rare Diseases, respectively) consensus definition,[4,13] frequently used in clinical and scientific settings, as well as other definitions used[14,15] are based on only few clinical parameters, resulting in limitations for outcome prediction and therapeutic strategies.

Pathophysiologically, an array of pathways was demonstrated to play a role in BPD pathogenesis, reflecting the relevance of developmental processes, oxidative stress, and inflammatory processes resulting in an imbalance of the immune system.[16–20] All pathways were found to form a tightly knit network and the ongoing efforts to understand and delineate their function did not reveal distinct endotypes thus far. Nonetheless, indicators of these processes have significant potential to inform clinical decision-making when they allow for early detection of risk for disease development while providing insight into relevant disease processes at the same time. The integration of molecular signatures into data from deep clinical phenotyping might identify underlying mechanisms and allow for risk stratification or patient subgroup identification.[21,22]

Diversifying omic-layers[23–26] has been shown to increase the predictive power of multi-omic analysis when used for endotype identification and can provide important pathophysiologic insights.[27–29]

We, therefore, took advantage of early postnatal multi-omic data comprising 1120 proteins and 213 metabolites available in a cohort of n=55 preterm infants born before 32 weeks postmenstrual age (PMA), phenotyped by detailed clinical data on pregnancy, birth, and neonatal intensive care unit until discharge and a complete assessment of pulmonary and extrapulmonary morbidities. We met the clinical need to early identify disease through risk stratification by addressing the following aims: First, we pursued the identification of clinically relevant disease endotypes by one or more omics layers and their integrative analysis. Specifically, we employed latent factor analysis via multi-omics latent factor analysis (MOFA)[30,31] to explore whether such subtypes or endotypes could emerge and, if so, which molecular features might characterize them. Second, we focused on differential abundance analysis to highlight molecular features that may differentiate disease states early after birth and thus serve as potential biomarkers for clinical use. Supervised analysis with random forest corroborated these findings as it revealed the same biomarkers, showing that these are robust predictive features. This analysis enabled us to contextualize the predictive potential in comparison to the sole use of clinical variables, and to investigate whether omics data could explain additional variance in a combined model. Finally, we traced BPD biomarker candidates into adult lung diseases that share important features with the diseased preterm lung, i.e., chronic obstructive pulmonary disease (COPD)[32–34] and idiopathic pulmonary fibrosis (IPF).[35,36]

Table 1. Clinical Characteristics of Patients Included for Molecular Analysis.

N (patients)	55
GA (weeks)	26.31±1.8
Birth weight (g)	804.73±235.1
Gender (F/M)	28/27
ANCS (yes/no)	46/7
Early-onset infection (yes/no)	43/12
RDS ≥grade 3 (yes/no)	46/8
MV (days)	53.52±25
O2 (days)	58.85±44.9
Postnatal steroid (days)	2.35±4.2
ROP ≥grade 3	3.67±1.3
IVH ≥grade 3	4±1.4
BPD	
None	14 (25%)
Mild	21 (38%)
Moderate	6 (11%)
Severe	14 (25%)
PDA (yes/no)	27/28

Description of clinical variables of the patients (n=55) considered for proteomic and metabolomic analysis.
Variables presented as mean and standard deviation or percent of total, respectively.
ANCS, antenatal corticosteroids; BPD, bronchopulmonary dysplasia; GA, gestational age; IVH, intraventricular hemorrhage; MV, mechanical ventilation; O2, oxygen supplementation; PDA, patent ductus arteriosus; RDS, respiratory distress syndrome diagnosed according to Couchard et al.;[37] ROP, retinopathy of prematurity.

METHODS

Data Collection

Preterm Cohort

In the first weeks of life, 55 plasma samples were obtained from n=175 preterm infants born at <32 weeks gestational age (GA) at the Perinatal Center in Munich (attention to infants at respiratory risk (AIRR)) after written informed parental consent (**Table 1**). The study received approval from the ethics committee of the Medical Faculty at Ludwig Maximilian University in Munich (Ethical vote #195-07) and is registered at the German Registry for Clinical Trials (No. 00004600; https://www.drks.de).

The BPD diagnosis was based on the NICHD/NHLBI/ORD consensus definition.[13] The time point of assessment was at 36 weeks PMA, and the diagnosis was stated if the neonate (<32 weeks GA) has been treated with oxygen (fraction of inspired oxygen (FiO_2) >0.21) for at least 28 days. The severity of the disease was defined as *mild* if supplemental oxygen (FiO_2 >0.21) was required for at least 28 days, but no need of oxygen supplementation persisted at 36 weeks PMA; *moderate* if oxygen supplementation (FiO_2 <0.30) was needed at 36 weeks PMA; and *severe* if oxygen supplementation (FiO_2 >0.30) and/or positive pressure ventilation/continuous positive pressure was received at 36 weeks PMA.[13] Each treatment refers to continuous application and oxygen supplementation for >12 hours equaling 1 day of treatment.[13] Cranial ultrasound was performed in all children (n=175) and brain MRI was performed in a subgroup (n=111). The findings were adjusted for structural abnormalities (e.g., mild asymmetry of the lateral ventricles), age-inappropriate maturation markers (e.g., delayed gyration), bleeding events (e.g., intraventricular hemorrhage (IVH) or periventricular leukomalacia (PVL)), and/or congenital brain malformations (e.g., septum pellucidum agenesis). In the event of one (or more) out of the aforementioned brain pathologies, the findings were classified as suspicious.

Neonatal Clinical Descriptors

To acknowledge the complex pathophysiological and clinical nature of the disease, we used different levels of clinical phenotyping. Disease grouping was performed according to the clinical BPD diagnosis based on the NICHD/NHLBI/ORD consensus definition.[13] As especially mild BPD comprises a broad variety of patients with regard to severity of disease, finer granulated stratification was achieved using the duration of oxygen supplementation or mechanical ventilation, here referred to as BPD descriptors. Next, main risk variables for BPD development were included in the analysis, i.e., GA and birth weight, which were correlated with disease descriptors (Pearson coefficient (n=175): mechanical ventilation r=−0.85 and −0.77, oxygen supplementation r=−0.70 and −0.65, respectively), and deep clinical phenotyping including comorbidity profiles as well as information about antenatal and postnatal treatment regimen (**Figure 1**; **Table 2**).

Multi-Omics Data Acquisition

We worked on two molecular levels obtained from blood plasma samples of patients with (n=41) and without (n=14) BPD.

Metabolomic data were obtained from ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) (Biomedical Metabolomics Facility Leiden, The Netherlands) run in four assays. The choice of platforms for metabolomic measurements considered sample volume limitations and reflected on the goal to highlight (i) potentially related pathways of cardiomyopathy, rhabdomyolysis, and metabolic acidosis which could impact BPD development (acylcarnitine assay); (ii) amines as building blocks for proteins, important signaling molecules, and immune functions; and (iii) the energy household, membrane integrity, and lung fluid alignment (lipid panels). Assays quantifying k=51 amines (no BPD=8, BPD=25) and k=26 acylcarnitines (no BPD=7, BPD=28) employed 5.0 μL and 10.0 μL of each sample to be mixed with an internal standard solution. The metabolites were precipitated by the addition of MeOH. After the reaction, the samples were transferred to vials placed in an autosampler tray and cooled to 4°C and 10°C, respectively. 1.0 μL of the reaction mixture was injected into an Agilent 1290 Infinity II LC System (San Jose, CA, USA) on an AccQ-Tag Ultra column with a flow of 0.7 mL/min over an 11-minute gradient. The UPLC was coupled to electrospray ionization on a triple quadrupole mass spectrometer (AB SCIEX QTRAP 6500, San Jose, CA, USA). The analytes were detected in the positive ion mode and monitored in multiple reaction monitoring using nominal mass resolution. Assays quantifying k=43 positive lipids (triglycerides (no BPD=8, BPD=26)) and k=93 non-triglycerides (no BPD=7, BPD=22) were

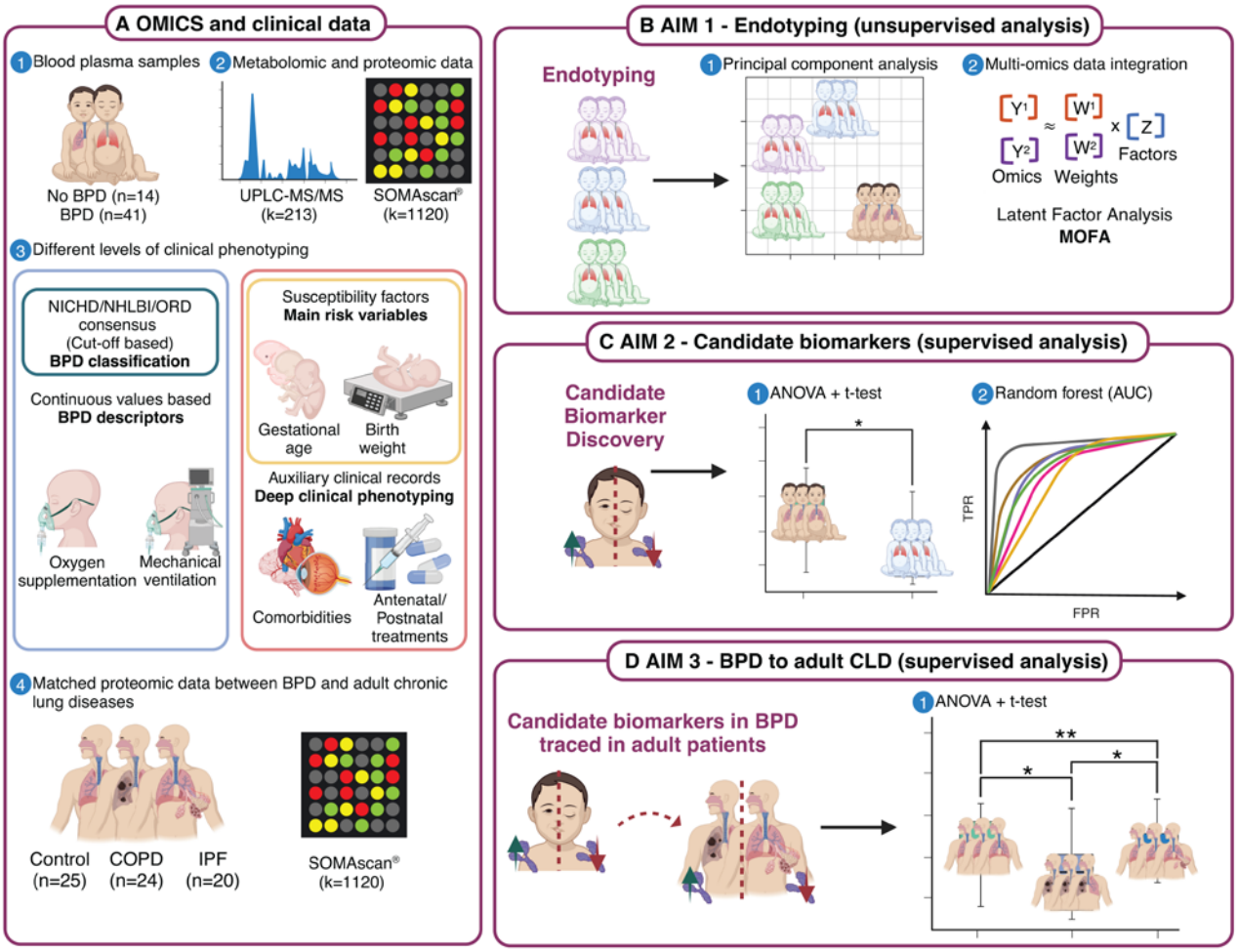

Figure 1. Project overview. (A) We used proteomic (SOMAscan®; n=35; k=1120) and metabolomic (UPLC-MS/MS; amines n=33, acylcarnitines n=35, positive lipids triglycerides n=34 and non-triglycerides n=29; k=213) data from blood plasma (1) for a cohort of n=55 neonates with (n=41) and without BPD (n=14) (2) with different levels of clinical data stratification including BPD predictors, main risk factors, and deep clinical phenotyping (3). We included a cohort of n=69 samples obtained from patients with adult CLD or respective controls (control n=25, COPD=24, and IPF=20) with proteomics data (SOMAscan®; k=1120) matched to the BPD proteomics profiles (4). (B) The discovery of potential BPD endotypes relied on unsupervised analyses such as principal component analysis by evaluating normalized abundances (PCA, 1) and multi-omics latent factor analysis (MOFA, 2) for data integration. (C) Candidate biomarker discovery used supervised analysis such as ANOVA to detect significant differences between the three severities (no, mild, moderate/severe BPD; adjusted p-value <0.1) and t-test for significant differences between no BPD and moderate/severe BPD (mild BPD reclassified in one of both severities, see the Methods section for details) (adjusted p-value <0.1) (1). Furthermore, we used the random forest for multi-class classification (three classes: no, mild, and moderate/severe BPD) and binomial (no BPD and moderate/severe BPD with mild cases reclassified) (2). (D) The evaluation of detected biomarkers in BPD as significant features in adult CLD included ANOVA with t-test as post-hoc analysis (adjusted p-value <0.05) (1). ANOVA, analysis of variance; AUC, area under the receiving operating characteristic; BPD, bronchopulmonary dysplasia; CLD, chronic lung disease; COPD, chronic obstructive pulmonary disease; IPF, idiopathic pulmonary fibrosis; NICHD, National Institute of Child Health and Human Development; NHLBI, National Heart, Lung and Blood Institute; ORD, Office of Rare Diseases; PCA, principal component analysis; UPLC-MS/MS, ultra-performance liquid chromatography tandem mass spectrometry. Asterisks denote the significance level (*P<0.05; **P<0.01).

performed using 10 µL of pre-processed sample mixed with 1000 µL IPA containing standards at C4 level and transferred to vials for LC-MS analysis. The chromatographic separation was achieved on an ACQUITY UPLC™ (Waters, Ettenleur, The Netherlands) with an HSS T3 column (1.8 µm, 2.1*100 mm) with a 0.4 mL/min flow over a 16-minute gradient. The analysis was performed using reference mass correction using a UPLC-ESI-Triple-TOF (Sciex 6600+) high-resolution mass spectrometer. Positive lipids were detected in full scan in the positive ion mode. All plasma samples were taken at the same time interval (median=4, range: 0–60 days) for all metabolomic assays.

Proteomic abundances were obtained from the SOMAscan® assay (SomaLogic®, Boulder, CO, USA) (no BPD=11, BPD=24; sampling time (median=4, range: 0–30 days)) for k=1124 individual high-affinity proteins (SOMAmer®) quantified on a custom Agilent hybridization array.[38,39] The abundances were analyzed in two batches (first batch n=9 (no BPD=2, BPD=7) and second batch n=26 (no BPD=9, BPD=17)). Proteomic targets included all measureable

Table 2. Levels of Clinical Phenotyping.

Levels	Variable	No BPD	Mild BPD	Mod./Sev. BPD
BPD descriptors (n=175)	O2 (days)	7.65±11	44.09±19.4	90.31±34
	MV (days)	19.78±14.6	50.98±16.6	71.49±22.1
Main risk variables (n=175)	GA (weeks)	29.03±1.7	26.65±1.8	25.71±1.9
	Birth weight (g)	1190.08±305.3	861.75±227.3	702.73±206.6
Deep clinical phenotyping(n=175)	Gender (F/M)	36/27	20/37	29/26
	Multiple births (yes/no)	29/34	14/43	16/39
	Preeclampsia (yes/no)	8/55	7/50	9/46
	Premature rupture of membranes (yes/no)	17/46	23/34	17/38
	AIS (yes/no)	29/34	36/21	31/24
	Tocolysis (yes/no)	34/29	31/26	24/31
	ANCS (yes/no)	58/5	48/9	46/9
	Apgar 5 minutes	8.05±0.9	7.96±1.1	7.65±1.4
	Surfactant administration (yes/no)	37/26	53/4	50/5
	RDS ≥grade 3 (yes/no)	34/29	45/12	50/5
	Postnatal steroid (days)	0.38±1.1	1.19±2.5	4.44±7
	Early-onset infection (yes/no)	35/28	42/15	43/12
	PDA (yes/no)	13/50	23/34	28/27
	ROP (yes/no)	5/58	14/43	27/28
	ROP ≥grade 3 (yes/no)	3/60	11/46	21/34
	IVH (yes/no)	6/57	10/47	11/44
	IVH ≥grade 3 (yes/no)	4/59	5/52	5/50
	PVL (yes/no)	2/61	4/53	1/54
	Brain pathology[a] (yes/no)	18/45	24/33	20/35
	Hospitalization (days)	53.95±19.6	77.79±17.9	99.4±23.5
	NICU (days)	48.49±18.7	71.09±22.4	93.8±25.1

[a]Brain pathology summarizes the presence of any pathology in cranial ultrasound and/or brain MRI.
Clinical variables are summarized by BPD severity showing mean±SD or number of cases, respectively.
We grouped clinical variables according to different levels of phenotyping.
AIS, amniotic infection syndrome; ANCS, antenatal corticosteroids; BPD, bronchopulmonary dysplasia; GA, gestational age; IVH, intraventricular hemorrhage; Mod./Sev., moderate/severe; MV, mechanical ventilation; NICU, neonatal intensive care unit; O2, oxygen supplementation; PDA, patent ductus arteriosus; PVL, periventricular leukomalacia; RDS, respiratory distress syndrome diagnosed according to Couchard et al.;[37] ROP, retinopathy of prematurity; SD, standard deviation.

proteins on the Somalogic® platform at that time given the limitations in sample volume.

Adult CLD Cohort

Proteomics data (SOMAscan®) in a cohort of adult chronic lung disease (CLD) patients (n=44) were obtained after informed consent (CPC-M bioArchive, Munich, Ethics Committee of the Medical Faculty of Ludwig Maximilian University in Munich (Ethical vote #19-629)). This included patients with (i) COPD (number of patients: n=24, median age 50 years (range: 14–74), 58.3% males), (ii) IPF (number of patients n=20, median age 56 years (range: 30–73), 70% males), (iii) control samples from the KORA (Cooperative Health Research in the Region Augsburg) cohort (number of patients: n=25, median age 60 years (range: 53–67), 52% males).[40] KORA is a regional research platform for population-based surveys and subsequent follow-up studies focusing on diabetes, cardiovascular diseases, and lung diseases, including the impact of environmental factors.

Data Analysis

Pre-processing

We removed clinical variables (k=73) with 20% missing values, keeping 25 variables for further analysis. The remaining missing values were imputed using the missRanger V2.1.3 R package, a random forest approach using 100 trees and default parameters[41,42] (**Supplementary Tables 1 and 2**).

We pre-processed data from no BPD and BPD cases together. Data from the four metabolomic assays were pre-processed separately. We applied variance stabilization normalization (VSN) using the vsn V3.64.0 R package.[43] Missing values were imputed

using the missRanger V2.1.3 R package with the same parameters used in clinical data imputation. In addition, we used hierarchical clustering and principal component analysis (PCA) to detect batch effects. We used the first two principal components as coordinates to find and remove extreme outliers by visual inspection (amines=5/38, acylcarnitines=4/39, positive lipids triglycerides=1/35, non-triglycerides=6/35; four samples shared in three metabolomic assays, one sample shared in two metabolomic assays, and two unique samples) (**Supplementary Figure 1A–D**) resulting in a total of 40 samples with metabolic information for further analyses. We used the proBatch V1.11.0 R package for both analyses.[44]

We pre-processed proteomic abundances following the same protocol applied to metabolomic data, resulting in four samples being removed (no overlap with metabolomic assay outliers) and a total of 35 samples with proteomic information for further analyses. We resolved technical replicates by matching proteomic and metabolomic samples according to sampling time. For the remaining replicates, we selected the sample with the lowest coefficient of variation. From the initial 1124 proteins, we selected proteins annotated as *human* in the organism category from the SOMAscan® report, excluding four proteins related to the human-virus infection process. We detected a batch effect regarding the two separated experimental runs and corrected it using the limma V3.52.2 R package, applying a linear model-based approach (removeBatchEffect function)[45] (**Supplementary Figure 1E**).

Unsupervised Analysis

To identify BPD endotypes and thereby enable improved patient stratification, we inferred latent factors using MOFA from the MOFA2 V1.6.0 R package. This method uses a spike-and-slab prior distribution to provide sparsity in the factor scores and weights, iteratively approaching the original normalized abundances via variational inference.[30,31]

We used samples with data representation in at least three out of five molecular assays (four UPLC-MS/MS assays for metabolomics and one SOMAscan® assay for proteomics) resulting in 33 samples in total (no BPD=8, BPD=25) with 13 samples sharing all, 16 sharing at least two metabolomics and the proteomics layer, and 17 sharing at least three metabolomics but no proteomics layer (**Supplementary Figure 2A**).

We trained the model using a Gaussian prior distribution, including normalized abundance matrices scaling, 2% of variance explained per factor as the threshold, and nine factors as the initial number of factors. We repeated model training 10 times using different random seeds to provide robustness analysis, choosing the model with the highest evidence lower bound value for further analyses. Each omics layer contributed to explaining the variance of latent factors. We only considered omics layers that accounted for at least 5% of the variance to ensure they accurately reflect the biological signature of each latent factor. For each factor and omics layer, we selected features explaining 90% of the variance as the most explanatory and used those to further characterize the latent factors. For latent factors with proteomic loadings, we conducted enrichment analysis on the top explanatory proteins using the gprofiler2 V0.2.1 R package (database from 07-05-2021). We searched for REACTOME pathways with an adjusted p-value (gSCS method) of 0.05.[46,47]

Significance Analysis

To identify potential biomarkers for early diagnosis and risk assessment, we performed differential abundance analysis between BPD severities via analysis of variance (ANOVA) with a paired t-test as the post-hoc method. ANOVA was adjusted for gender, sampling time, and the BPD main risk factors: GA and birth weight.

In both cases, ANOVA and t-test, p-values were adjusted using the Benjamini and Hochberg correction. We considered features with adjusted p-values below 0.1 for further analysis. Significance analysis, normalization, and unsupervised analysis were performed in R V4.2.3.

Reclassification of Mild BPD Cases

The consensus definition published by Jobe and Bancalari[13] classifies moderate and severe BPD together with a heterogeneous group of mild BPD cases (see the Methods section for disease definition). Mild BPD comprises patients with a history of rather short-term to prolonged durations of hyperoxia or mechanical ventilation exposure while not correcting for the degree of immaturity. To re-stratify this heterogeneous patient group, we used their history of oxygen supplementation and need for mechanical ventilation (duration in days) as more adequate indicators of the underlying degree of lung disease to reassign them into either no or moderate/severe BPD. The supervised approach used a random forest model with nested five-fold cross-validation (no BPD and moderate/severe BPD cases) to reclassify cases diagnosed with mild BPD (n=57) into the no BPD (n=63) or moderate/severe BPD (n=55) category. We evaluated the number of estimators (50, 100, and 150), the maximum depth of the tree (maximum growth, 10, and 20), the minimum number of samples to split a node (2, 5, and 10), and the minimum number of leaves per tree (1, 2, and 4) and selected the hyperparameter configuration with the highest inner-loop area under the receiving operating characteristic curve (AUC). The remaining random forest parameters were kept at their default. For this classification task, we used the imbalance-learn V0.11.0 Python library.[48]

Confirmatory BPD Severity Classification

To corroborate the identified biomarkers and associate the latent factors with disease severity, we performed multi-class random forest classification (three classes: no, mild, and moderate/severe BPD) and feature importance analysis on the full set of metabolites and proteins, respectively, or latent factors. Next, to assess the biomarkers' ability to improve patient classification using a random forest model trained in a three-fold cross-validation, we combined them with our sets of clinical data and compared their performance against models that were solely based on clinical parameters. The

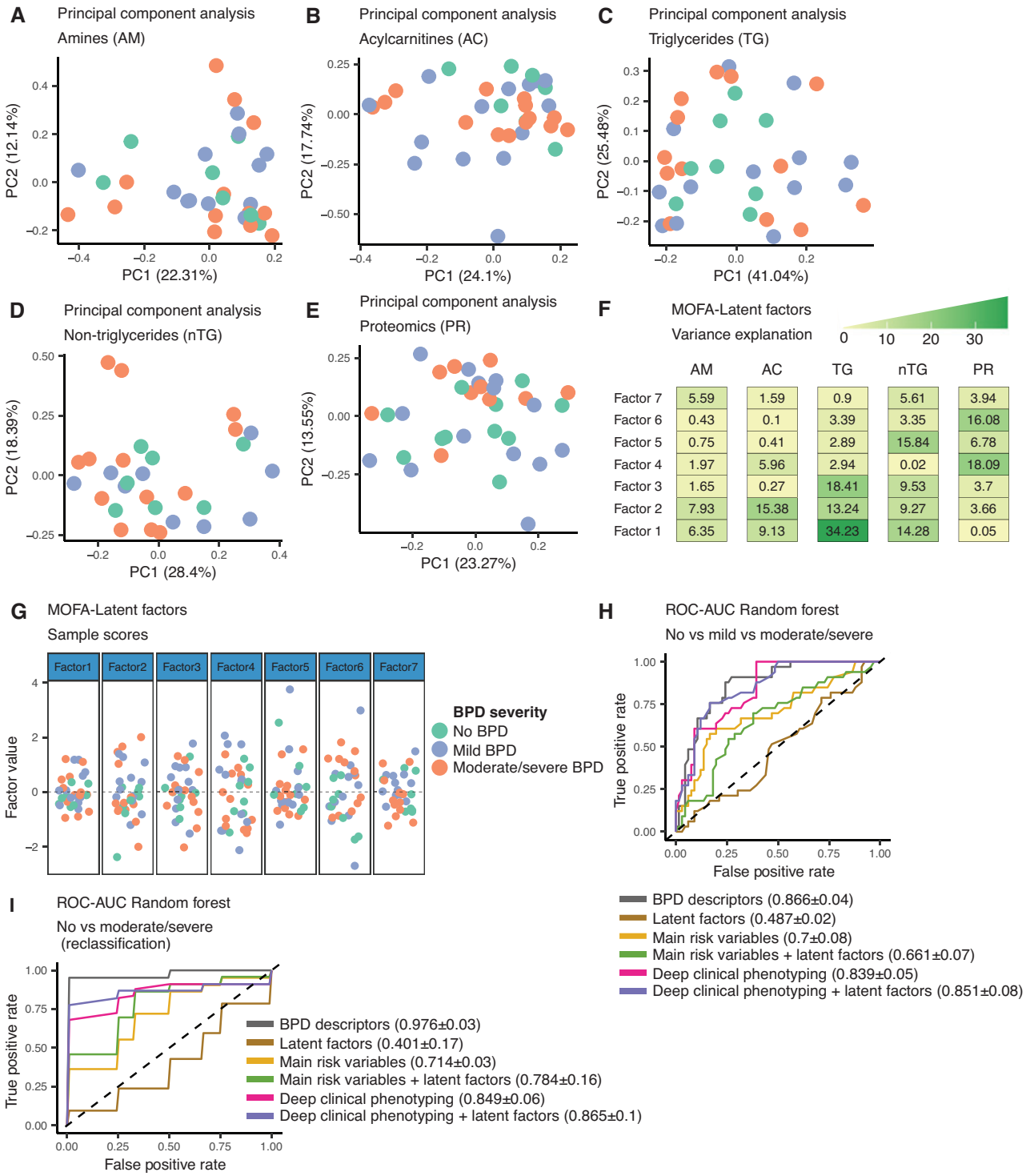

Figure 2. Metabolomic and proteomic data integration. (A) Principal component analysis (PCA) for normalized amine abundances. (B) PCA for normalized acylcarnitine abundances. (C) PCA for normalized positive lipids triglyceride abundances. (D) PCA for normalized positive lipids non-triglyceride abundances. (E) PCA for normalized proteomics abundances. (F) Heatmap showing the variance explanation for the seven latent factors detected by MOFA (5% minimal explanatory threshold). (G) Dot-plot displaying the distribution of patients per latent factor colored by BPD severities according to factor scores (H) ROC-AUC curve for six multi-class random forest models (three classes: no, mild, and moderate/severe BPD), using oxygen supplementation and mechanical ventilation (BPD descriptors) as the reference model; the legend indicates micro AUC averages and their standard deviation. (I) ROC-AUC curve for six binomial random forest models (no, moderate/severe BPD and reclassified mild BPD cases, see the Methods section for details), using oxygen supplementation and mechanical ventilation (BPD descriptors) as the reference model; the legend indicates AUC mean±standard deviation. (A-E) The dots represent patients colored by BPD severities. AUC, area under the receiving operating characteristic; BPD, bronchopulmonary dysplasia; MOFA, multi-omics latent factor analysis; ROC, receiving operating characteristic.

approach enabled us to assess whether the identified biomarkers explain additional variance of the data the clinical variables could not capture alone. The performances of individual models were measured using micro and macro mean±standard deviation of receiving operating characteristic (ROC)-AUC scores. We used the mean accumulation of impurity decrease per tree as a criterion for feature importance calculation. All supervised analyses were performed using sckit-learn V1.3.0[49] in Python V3.11.0.

Adult CLD Analysis

For proteomic analysis performed in adult CLD cases, we followed the pre-processing steps described above including technical replicates selection, VSN, batch effect detection, and outlier removal (3/72 outliers removed; n=69; k=1120) (**Supplementary Figure 3**). We pre-processed the three conditions (COPD (n=24), IPF (n=20), and controls (n=25)) together. ANOVA with paired t-test as post-hoc was applied afterwards to detect significant abundant proteins between the three conditions, using gender and smoke status as confounders. P-values for ANOVA and t-tests were adjusted using the Benjamini–Hochberg correction.

Availability and Reproducibility

All the scripts used to perform the different analyses, including the generation of figures, tables, and supplementary material, as well as the intermediate data produced during this study, are available at Zenodo: https://zenodo.org/records/15283709.

A detailed report about methods reproducibility including pre-processing steps and machine learning training process is available at AIMe: https://aime.report/LSZJul.

RESULTS

Our study aimed to identify (i) early postnatal multi-omic signatures that yield endotypes in neonatal CLD, i.e., BPD, (ii) candidate biomarkers that explain clinical severity, and (iii) to link adult with neonatal CLDs through the identified biomarkers (**Figure 1**). To this end, we analyzed metabolomic and proteomic samples obtained in the first weeks of life (mean 7.64±9.8 days) and sets of clinical variable with increasing detail in a cohort of infants with BPD (n=41) and age-matched preterm infants without BPD (n=14) though unsupervised and supervised analyses.

Metabolomic and Proteomic Profiles Identified Shared Immune Phenomena as Disease Indicators While Distinct Disease Endotypes Could Not Be Revealed

For the identification of disease endotypes to enhance individualized monitoring and treatment (aim 1), we performed unsupervised data analysis with a focus on integrative latent factor analysis based on MOFA. PCA did not reveal disease endotypes in any early postnatal molecular layers (**Figure 2A–E**). MOFA revealed seven latent factors that collectively explained the variance associated with BPD (**Figure 2F**). The model was able to explain 23.76% of the variance for amines (n=33), 35.62% for acylcarnitines (n=31), 74.76% for triglycerides (n=31), 53.44% for non-triglycerides (n=29), and 51.54% for proteomics (n=16) (**Supplementary Figure 2B and C**). However, none of the seven latent factors effectively separated samples into endotypes (**Figure 2G**). Likewise, K-means analysis did not cluster individual molecular layers or latent factors (**Supplementary Figures 4 and 5**).

Three latent factors (factor 4, 5, and 6, **Figure 2F**) shared variance between early postnatal proteomics and metabolomics abundance levels. Furthermore, Pearson correlation analysis suggested a shared biological signature between factors 4 and 6 (r=0.88) but an independent signal for factor 5 (**Supplementary Figure 2D**). Enrichment analysis on the set of proteins that loaded most strongly on those three latent factors to provide precise biological descriptions (pathways) highlighted "cytokine signaling in immune system," "signaling by interleukins," and "interleukin-4 and interleukin-13 signaling" as shared relevant pathways representing a common inflammatory disease component (**Supplementary Table 3**).

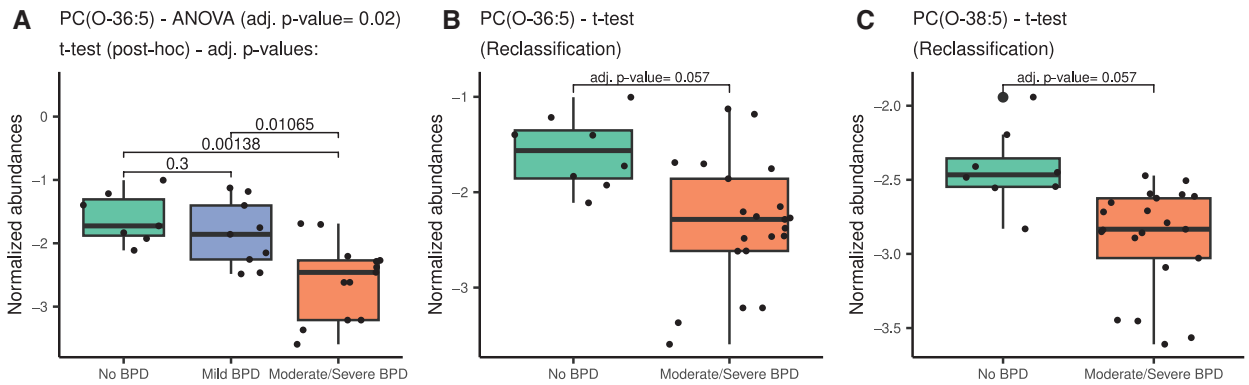

Figure 3. BPD metabolomic analysis. (A) Boxplot demonstrating significant differences (ANOVA adjusted p-value <0.1) between PC(O-36:5) abundances obtained for no, mild, and moderate/severe BPD. (B) Boxplot showing significant differences (t-test adjusted p-value <0.1) between abundances for PC(O-36:5) in no and moderate/severe BPD cases. (C) Boxplot showing significant differences (t-test adjusted p-value <0.1) between abundances for PC(O-38:5) in no and moderate/severe BPD. (B and C) Mild BPD cases were reclassified into no or moderate/severe BPD groups by random forest trained with the BPD predictors of the entire AIRR cohort dataset (n=175; no BPD=63, mild BPD=57, moderate/severe BPD=55). ANOVA, analysis of variance; AIRR, attention to infants at respiratory risk; BPD, bronchopulmonary dysplasia.

Acylcarnitines and proteomics explained latent factor 4 (5.96% and 18.09%, respectively). Enrichment analysis based on the most explanatory proteins (k=497) allowed the association of this latent factor with fibroblast growth factor receptor (FGFR) signaling processes such as "signaling by activated point mutants of FGFR1" (adj. p-value=6.146e-07) and "FGFR4 ligand binding and activation" (adj. p-value=1.348e-06) (**Supplementary Table 4**). Among the most explanatory acylcarnitines in factor 4, we found isobutyrylcarnitine (weight=0.37), followed by acetylcarnitine and hexanoylcarnitine (weights: −0.29 and −0.29, respectively) (**Supplementary Figure 6A**). In proteomics, the most explanatory protein for factor 4 was SSRP1, a subunit involved in the formation of chromatin elongation factor (FACT) (weight=−1.03), followed by CRISP3 (weight=−1.0), a protein related to the innate immune system, and RPS6KA5 (weight=−0.93), a protein with serine/threonine kinase activity (**Supplementary Figure 6B**).

Factor 5 was explained by positive lipids non-triglycerides (lipids ionized to get a positive charge during the MS process including ceramides, diglycerides, phosphatidylcholines, lysophosphatidylcholines, phosphatidylethanolamines, lysophosphatidylethanolamines, and sphingomyelins) (15.84%), and proteomics (6.78%). Enrichment analysis of the most explanatory proteins (k=103) indicated "immunoregulatory interactions between a lymphoid and a non-lymphoid cell" (adj. p-value=0.001) as the biological signature of this latent factor (**Supplementary Table 5**). Among the most explanatory positive lipids non-triglycerides for factor 5, we found a long chain ceramide with the highest weight (−1.07), followed by a lysophosphatidylcholine (LPC(20:4); weight=0.95) and phosphatidylethanolamine PE(O-38:5) (weight=−0.73) (**Supplementary Figure 6C**). Furthermore, among the most explanatory proteins for this latent factor, we found the enzymes CA2 and DDX19B (weights: 2.91 and 2.54, respectively) and the RNA-binding protein HNRNPK (weight=2.38) (**Supplementary Figure 6D**).

In contrast, latent factor 6 was exclusively explained by proteomics (16.08%). Its biological signal was more diverse, with enrichment analysis of the 158 explanatory proteins revealing

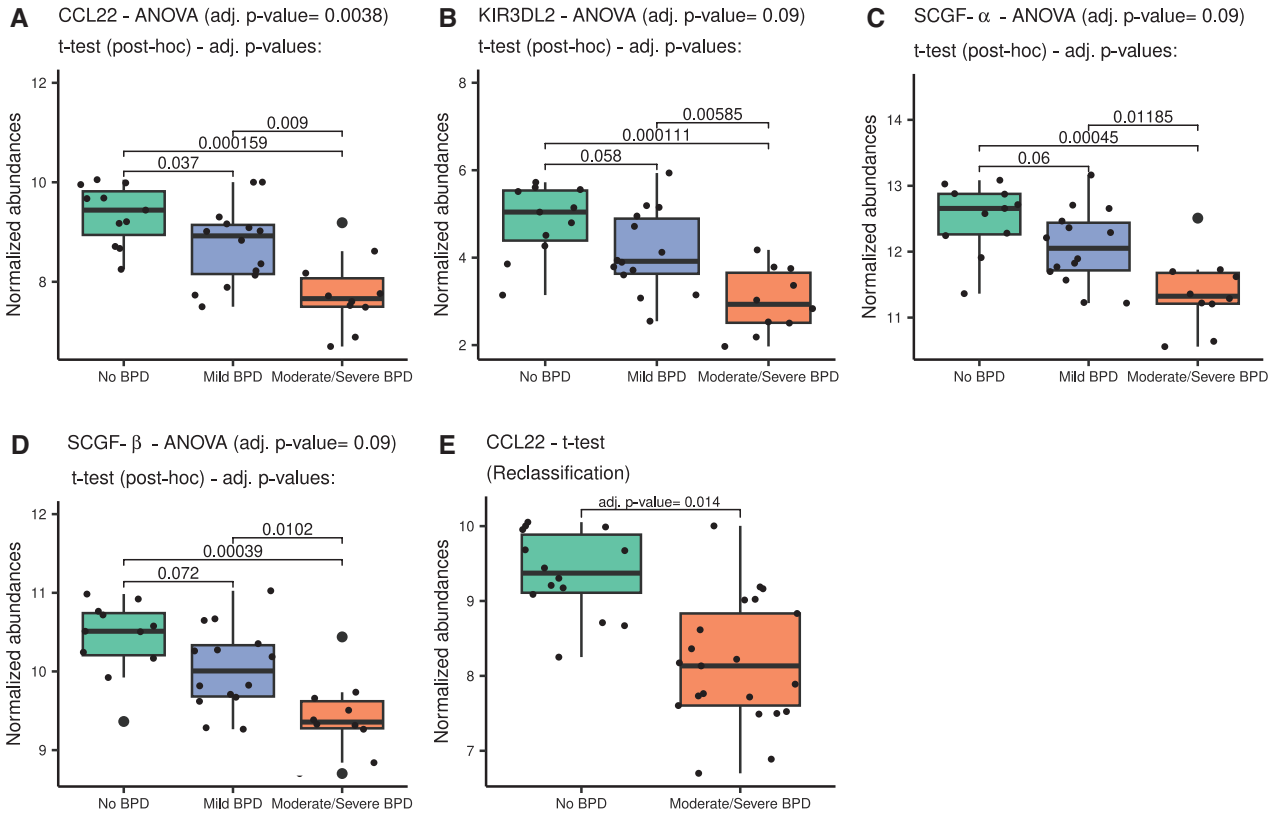

Figure 4. BPD proteomic analysis. (A) Boxplot displaying significant differences between no, mild, and moderate/severe BPD according to ANOVA analysis (adjusted p-value <0.1) for CCL22. (B) Boxplot displaying significant differences between no, mild, and moderate/severe BPD according to ANOVA analysis (adjusted p-value <0.1) for KIR3DL2. (C) Boxplot displaying significant differences between no, mild, and moderate/severe BPD according to ANOVA analysis (adjusted p-value <0.1) for SCGF-alpha. (D) Boxplot displaying significant differences between no, mild, and moderate/severe BPD according to ANOVA analysis (adjusted p-value <0.1) for SCGF-beta. (E) Boxplot displaying significant differences between no and moderate/severe BPD severities according to t-test analysis (adjusted p-value <0.1) for CCL22[a].

[a]Mild BPD cases were reclassified into no or moderate/severe BPD groups by random forest trained with the BPD predictors in the entire AIRR cohort dataset (n=175; no BPD=63, mild BPD=57, moderate/severe BPD=55). ANOVA, analysis of variance; AIRR, attention to infants at respiratory risk; BPD, bronchopulmonary dysplasia.

pathways such as "activation of NMDA receptors and postsynaptic events" (adj. p-value=2.898e-09), "homeostasis" (adj. p-value=1.861e-07), and "intrinsic pathway for apoptosis" (adj. p-value=4.147e-06) to be significantly enriched (**Supplementary Table 6**). Among the most explanatory proteins for factor 6 were SRC (weight=2.55), a tyrosine-protein kinase activity protein, followed by the translation initiation mediator EIF4G2 (weight=2.11), and PDPK1 (weight=2.03), a protein with phospholipase activity (**Supplementary Figure 6E**).

To associate the latent factors to disease severity, we performed a multi-class random forest (three classes: no, mild, moderate/severe BPD; training n=22, test n=11) using the latent factor scores as predictors. The general performance (macro AUC average=0.462±0.05) did not support using an embedded representation of proteomics and metabolomics for BPD severity classification (**Supplementary Figure 7A**). However, combining latent factor scores and deep clinical phenotyping variables improved prediction performance (micro AUC averages=0.851±0.08 and 0.839±0.05, respectively), indicating that the latent factors orthogonally explain variability to the clinical variables (**Figure 2H**; **Supplementary Figure 7B**). We also conducted a binomial random forest analysis (no vs moderate/severe BPD; training n=22, test n=11) after reclassifying mild BPD cases. Similar to the multi-class model, latent factors alone performed poorly (mean AUC=0.401±0.17) but improved when combined with deep clinical phenotyping (mean AUC=0.856±0.1), outperforming all other model combinations (**Figure 2I**; **Supplementary Figure 7C**). Pearson correlation analysis showed no strong associations between latent factors and key clinical BPD parameters, with the highest correlation (r=−0.45) observed between factor 4 and multiple births (**Supplementary Figure 2E**).

Supervised Analysis Highlights Phosphatidylcholine PC(O-36:5) and CCL22 as Biomarker Candidates for BPD Severity

For the discovery of clinically relevant biomarkers that allow early postnatal risk stratification (aim 2), we conducted supervised analyses including differential abundance analysis and random forest severity prediction.

We identified significantly different abundant proteins and metabolites between BPD severities by ANOVA analysis. The phosphatidylcholine PC(O-36:5) (adjusted p-value 0.021) was the only significant metabolite detected (**Figure 3A**). Post-hoc test (t-test) showed significant differences between no and moderate/severe BPD (adjusted p-value=0.001), as well as between mild and moderate/severe BPD (adjusted p-value=0.011) (**Figure 3A**; **Supplementary Tables 7–10**).

After reclassification of cases diagnosed with mild BPD, we performed a t-test comparing abundance levels between no BPD and moderate/severe BPD groups. This analysis identified only two metabolites with significantly different abundances: PC(O-36:5) and PC(O-38:5) (adjusted p-values 0.057 each; **Figure 3B and C**; **Supplementary Tables 11–14**).

When focusing the analysis on the first 4 weeks of life (samples ≤28 days after birth; amines n=30, acylcarnitines n=31, positive lipids: triglycerides n=31, and non-triglycerides n=26), we confirmed PC(O-36:5) (adjusted p-value=0.012) as differentially abundant metabolites between the three BPD severities, followed by PC(O-38:4), PC(40:4), PC(36:6), and PC(O-36:4) (adjusted p-values=0.022, 0.031, 0.041, and 0.041, respectively) (**Supplementary Figure 8A–E**; **Supplementary Tables 15–18**). After mild BPD reclassification, t-test analysis for no BPD and moderate/severe BPD in this subset of samples confirmed PC(O-36:5) and PC(O-38:5) as differentially abundant metabolites (adjusted p-values=0.074 and 0.082, respectively). Furthermore, we identified the acylcarnitines hexadecenoylcarnitine, palmitoylcarnitine, and isovalerylcarnitine as additional differentially abundant metabolites (adjusted p-values=0.057, 0.064, and 0.096, respectively) (**Supplementary Figure 8F–J**; **Supplementary Tables 19–22**).

The correlation between the metabolites of interest, PC(O-36:5) and PC(O-38:5), and key clinical variables in BPD (including descriptors and main risk factors) revealed that the most significant metabolite, PC(O-36:5), showed weak correlations with days of oxygen supplementation (r=−0.62) and birth weight (r=0.67). In contrast, PC(O-38:5) was not correlated with any of the primary clinical variables associated with BPD (**Supplementary Figure 9A**).

ANOVA-based differential abundance analysis of proteomics data across BPD severities (no, mild, and moderate/severe) identified four differentially expressed proteins: CCL22 (adjusted p-value 0.004, post-hoc t-tests significant for all comparison levels), KIR3DL2, SCGF-alpha, and SCGF-beta (adjusted p-values of 0.09) (**Figure 4A–D**; **Supplementary Table 23**). After reclassifying mild BPD cases, CCL22 remained as the unique, significantly differentially abundant protein differentiating no BPD from moderate/severe BPD cases (adjusted p-value=0.014, **Figure 4E**; **Supplementary Table 24**).

Analysis of samples from the first 4 weeks of life (samples ≤28 days after birth, n=34) confirmed CCL22's differential abundance across BPD severities via ANOVA (adjusted p-value 0.042, **Supplementary Figure 10A**; **Supplementary Table 25**), and between no and moderate/severe BPD (after mild BPD reclassification) by t-test (adjusted p-value 0.051, **Supplementary Figure 10B**; **Supplementary Table 26**).

Pearson correlation analysis between significant proteins and key BPD clinical variables revealed moderate correlations for CCL22 with days of oxygen supplementation (r=−0.63) and birth weight (r=0.60), and stronger correlations with days of mechanical ventilation (r=−0.77) and GA (r=0.73). In contrast, SCGF-alpha, SCGF-beta, and KIR3DL2 showed weak correlations with these clinical variables (**Supplementary Figure 9B**).

To corroborate our findings, we performed random forest-based BPD severity prediction and feature importance analysis based on the entire measured metabolome and proteome, respectively, thereby scrutinizing the identified biomarkers in the context of all measured metabolites and proteins. We expected the identified biomarkers to rank in the top percentile of importance features.

Multi-class models (no, mild, moderate/severe BPD; training n=20, test n=9) based on positive lipids non-triglycerides demonstrated comparable performance to the model trained with BPD

risk variables (micro AUC average=0.703±0.07 and 0.701±0.18; **Figure 5A**). The performance decreased when metabolomic data were combined with clinical parameters. Feature importance revealed PC(O-36:5) among the most important features (importance=2.46%), followed by PC(40:4), PC(O-38:5), and PC(O-44:5) (importance=5.36%, 4.92%, and 4.09%, respectively) (**Figure 5B**).

Binomial random forest models trained on proteomics data (no vs moderate/severe BPD, after mild BPD reclassification; training n=24, test n=11; k=1120) performed well (mean AUC=0.783±0.08). Again, the performance was reduced when proteomic data were combined with clinical parameters (**Figure 5C**). Feature importance displayed CCL22 as the most important feature (importance=2.5%), followed by ERBB2 and CST6 (importance=1.87% and 1.58%, respectively) (**Figure 5D**). Detailed experimental results can be found in **Supplementary Figures 11 and 12**.

To confirm the biomarkers' added value for clinical decision-making over the sole use of clinical descriptors, we trained random forest classifiers (three classes: no, mild, moderate/severe BPD) combining clinical variables with the identified biomarkers. Inclusion of the identified phosphatidylcholines into the different sets of clinical variables demonstrated an increase in performance ranging between 0.02 and 0.11 ROC-AUC (**Figure 6A**), indicating that the differentially abundant metabolites explained variance that the clinical variables could not explain alone. Feature importance analysis on the top performing model also highlighted PC(0-38:5) as the top most important feature (**Figure 6B**). Similar behavior was demonstrated for the proteomic biomarkers with an improved ROC-AUC of 0.05–0.06 AUC (**Figure 6C**). Feature importance of the top performing model ranked CCL22, SCGF-beta, and KIR3DL2 within the top six important features (**Figure 6D**).

Differential Abundance Analysis Confirmed CCL22 and KIR3DL2 as Biomarker Candidates in COPD Patients

Lastly, we reevaluated the BPD biomarker candidates (CCL22, SCGF-alpha, SCGF-beta, KIR3DL2) in adult patients with CLD, i.e., COPD (n=24) and IPF (n=20) to link neonatal pathophysiology to adult CLDs that share histopathological pattern (aim 3). The diseases mirror structural characteristics of BPD, namely emphysema

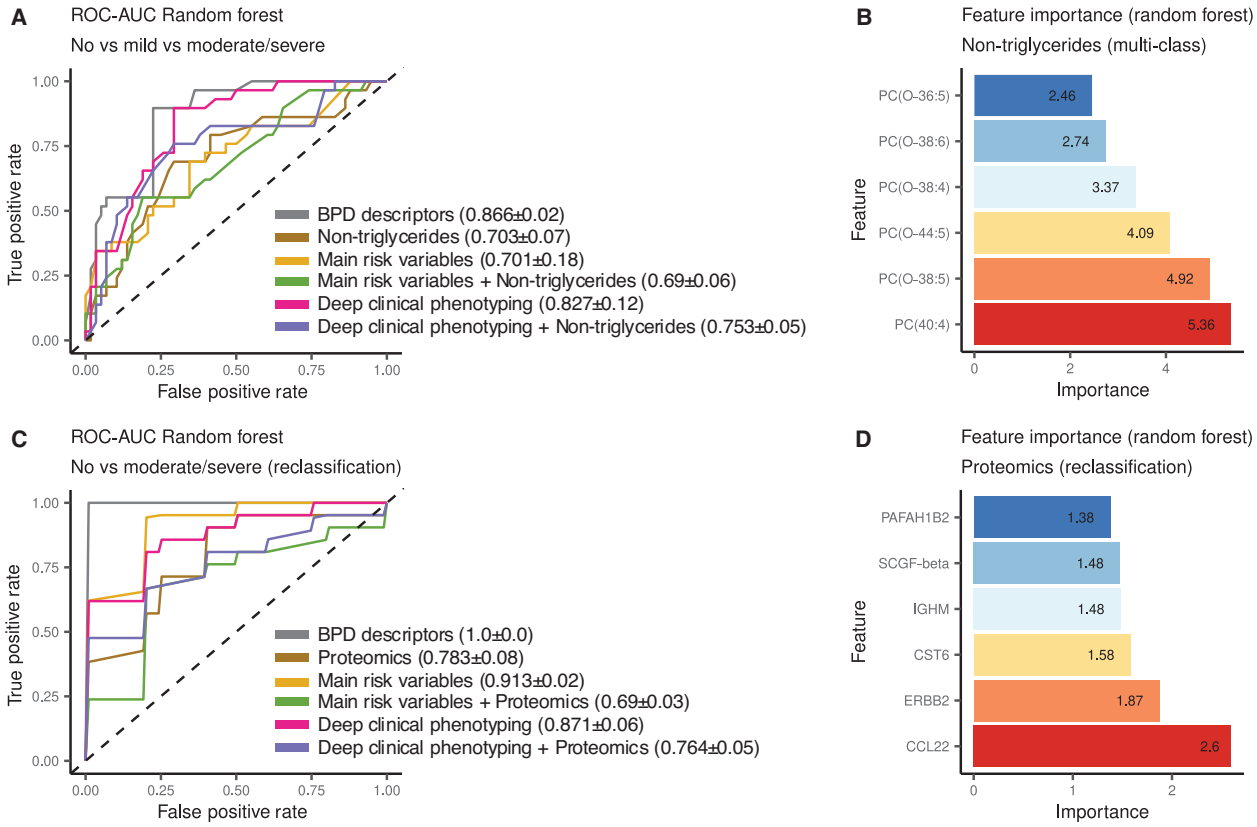

Figure 5. Corroborating BPD biomarker candidates through random forest. (A) ROC-AUC curve for six multi-class random forest models (three classes: no, mild, moderate/severe BPD) trained using positive non-triglyceride lipids and their combination with different sets of clinical parameters, using oxygen supplementation and mechanical ventilation (BPD descriptors) as the reference model. The legend indicates micro AUC averages and their standard deviation. (B) Top six feature importance for the model trained using positive non-triglyceride lipids. (C) ROC-AUC curves for six binomial random forest models (no and moderate/severe BPD, after reclassification of mild BPD cases, see the Methods section for details) trained using proteomics and their combination with different sets of clinical parameters, using oxygen supplementation and mechanical ventilation (BPD descriptors) as the reference model; the legend shows the mean±standard deviation of AUC. (D) Top six feature importance for the model trained using proteomics. AUC, area under the receiving operating characteristic; BPD, bronchopulmonary dysplasia; ROC, receiving operating characteristic.

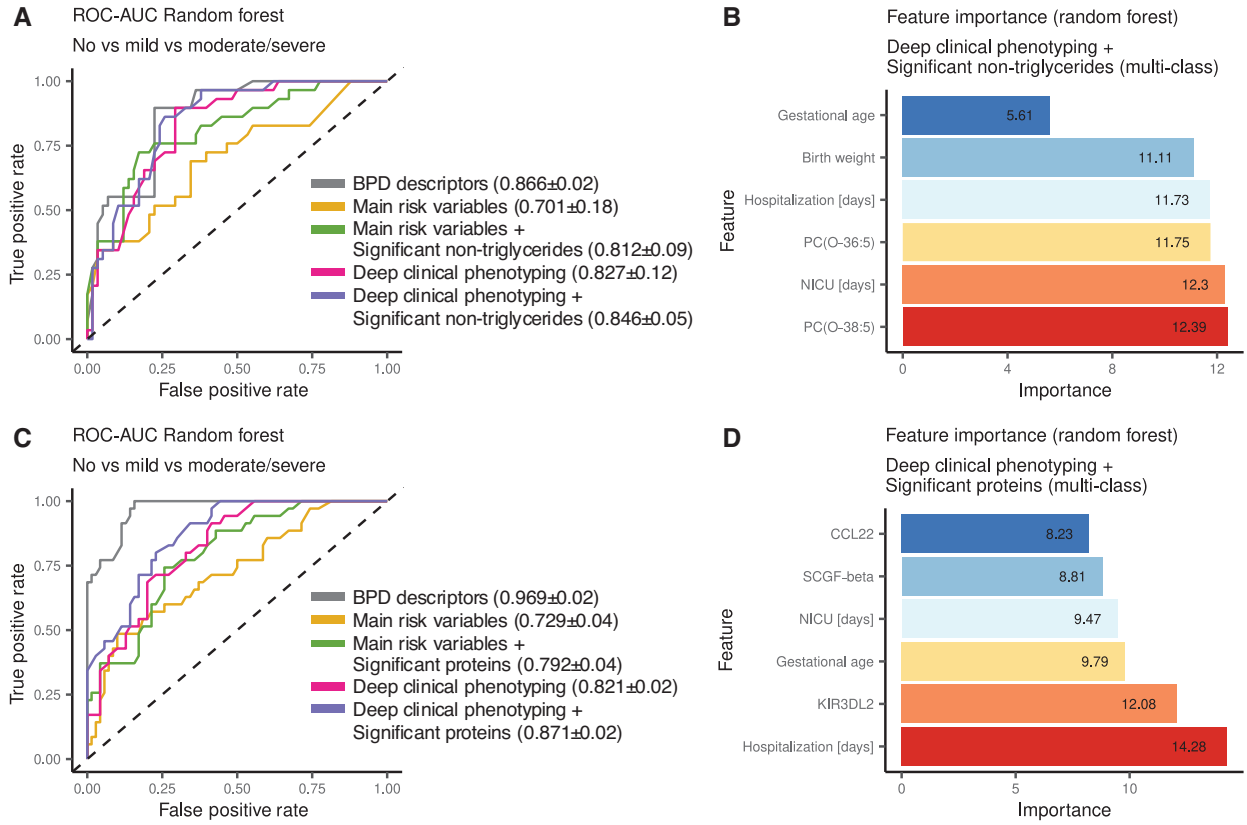

Figure 6. Classification power of significant metabolites and proteins in BPD. (A) ROC-AUC curve for five multi-class random forest models (three classes: no, mild, moderate/severe BPD) including combinations of different sets of clinical parameters and significant metabolites in BPD, using oxygen supplementation and mechanical ventilation (BPD descriptors) as the reference model; the legend indicates micro AUC averages and their standard deviation. (B) Top six features based on random forest inferred importance for the model trained combining deep clinical phenotyping and significant positive lipid non-triglycerides. (C) ROC-AUC curve for five multi-class random forest models (three classes: no, mild, moderate/severe BPD) including combinations of different sets of clinical parameters and significant proteins in BPD, using oxygen supplementation and mechanical ventilation (BPD descriptors) as the reference model; the legend indicates micro AUC averages and their standard deviation. (D) Top six feature importance for the model trained combining deep clinical phenotyping and significant proteins. AUC, area under the receiving operating characteristic; BPD, bronchopulmonary dysplasia; ROC, receiving operating characteristic.

and pulmonary fibrosis. ANOVA revealed significant differential abundance for CCL22, KIR3DL2, and SCGF-beta (adjusted p-values= 9e-05, 4e-03, and 5e-04, respectively). Post-hoc t-tests highlighted CCL22 and KIR3DL2 as significantly different in COPD patients (adjusted p-values=0.003 and 2.4e-04 vs controls, respectively; **Figures 7A and B**). SCGF-beta showed significant differences in both COPD and IPF patients when compared to individuals without CLD (adjusted p-values=0.007 and 9.6e-06, respectively; **Figure 7C**). In both neonatal and adult CLD, the proteins were downregulated; however, this was less pronounced in adult CLD (**Supplementary Table 27**).

DISCUSSION

BPD, one of the most prevalent morbidities in preterm infants, impacts on pulmonary and extrapulmonary health in the long term.[3] Impairing predictive power and the development of therapeutic strategies, the diagnosis solely relies on selected clinical parameters[4,13] while lacking comprehensive insight into underlying pathophysiological mechanisms. The need for novel approaches to improve disease stratification was reflected by different modifications to disease definitions as well as the emerging potential of omic data integration to improve the diagnostic process.[2,50–52]

We met the clinical need to improve risk stratification by outlining two project aims. First, we applied integrated latent factor analysis to test the idea that the combination of molecular characterization and deep clinical phenotyping can identify endotypes that pay tribute to disease heterogeneity while overcoming non-granularized grouping to thereby allow for individualized monitoring and treatment.[21,22] Second, we aimed at early disease identification by the use of biomarkers that—at the same time—provide insight into relevant pathophysiological processes. Tracing these biomarkers into adult CLDs that share striking histopathological features with BPD underscored the biomarkers' potential for risk stratification.

We employed multi-omic analysis in early postnatal samples in an age-matched, deeply phenotyped cohort of very immature preterm infants with and without BPD. By global unsupervised multi-factor analysis, we detected a shared immune response and inflammation signature across factors and identified candidate

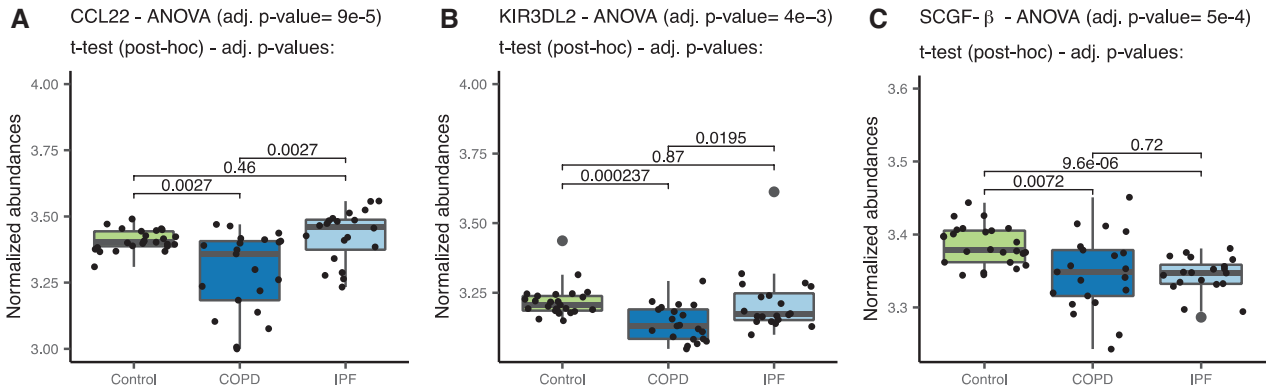

Figure 7. Tracing the BPD protein signature into COPD and IPF. Boxplot displaying significant differences between protein abundance levels obtained in control (n=25), COPD (n=24), and IPF (n=20) samples (adult patients) according to ANOVA (adjusted p-value <0.1). (A) CCL22; (B) KIR3DL2; (C) SCGF-beta. ANOVA, analysis of variance; BPD, bronchopulmonary dysplasia; COPD, chronic obstructive pulmonary disease; IPF, idiopathic pulmonary fibrosis.

biomarkers for BPD. The signature detected in the first weeks after birth aligns with the characteristic pro-inflammatory BPD signature identified by different studies[53–55] and furthermore highlights the early, potentially prenatal, development of disease and determination of its course.

The differential abundance of PC(O-36:5) and CCL22 as well as SCGF-alpha, SCGF-beta, and KIR3DL2 among disease grades, and their validation through unfiltered random forest feature importance analysis highlights critical pathophysiological changes, making them plausible biomarker candidates. Confirmation of the biomarker candidates identified when restricting the analysis to measurements obtained in the first 4 weeks of life and the improved predictive performance when combined with clinical variables in a random forest-based model, underscored the potential for early disease detection. The added value of the biomarker for the diagnostic process, however, is not only explained by parameters of sensitivity and specificity, but as well based on the objectivity and standardizability of these measurements as opposed to often more subjective clinical measures that vary between centers.

The decrease in circulating CCL22 levels in BPD infants highlights a shift in the immune response of BPD infants as the cytokine's role in orchestrating the interaction with regulatory T cells suggests a subsequent impairment in adaptive immunity and a predisposition to inflammatory disease development.[56] Upregulation of this protein in lung tissue samples of BPD mouse models likely mirrors the early injury phases studied, where the recruitment of monocytic cells to the lung plays a pivotal role.[57,58] Alongside CCL22, the downregulation of KIR3DL2 in BPD reflects on changes in adaptive immunity[59] and separates BPD cases from controls when combined with deep clinical phenotyping. Negative regulation of SCGF-alpha and -beta as hematopoietic stem/progenitor cell growth factors[60,61] highlights a possible relation to impaired lung development and may serve as a target of current therapeutic efforts.[62,63]

Successful identification of the BPD biomarker candidates, including their regulatory pattern in adult lung disease, underscores the notion that BPD represents an early form of adult CLD. Interestingly, main features were shared between BPD and COPD, i.e., CCL22, KIR3DL2, and SCGF-beta, supporting discussion of shared drivers of disease.[64] In IPF only SCGF-beta demonstrated a shared pattern of differential abundance.

Identifying BPD endotypes by integrating multi-omic layers and deep phenotyping parameters through global unsupervised multi-factor analysis remains challenging as proven by our and other studies. A variety of factors are likely caused by limited sample size, the time frame of sample acquisition, and the limitations of current clinical phenotyping. The exclusion of distinct structural and functional changes results in limited strategies to determine disease severity in a diverse patient population and heterogeneous disease presentation. To enable the formation of robust clusters and the identification of multi-omics factors for a discernible separation of the disease, future studies require larger, extensively phenotyped cohorts including age-matched controls, meeting or exceeding the efforts of the present study, that undergo unbiased high-throughput multi-omic screening at different time points including early postnatal as well as possible prenatal sampling.[7,9,65] Early disease processes might be "silenced" after birth when metabolomic and proteomic signatures are significantly driven by therapeutic interventions or uniformly converge into an inflammatory signature overshadowing other signals.

The strong impact of clinical confounders such as GA and birth weight[25,66] might require further stratification of cohorts for the degree of immaturity while considering the presence of growth retardation. On the other hand, although known as an important risk factor, immaturity alone does not fully explain morbidity development in the preterm infant and the varying impact of interdependent risk factors such as infections, genetic background, or therapy (and subsequent complications) on outcome is challenging to account for in multi-omic profiles. The challenges are equally presented by previous studies, where BPD endotype identification through hierarchical clustering of transcriptomic data based on a fixed number of expected clusters[67] was mainly attributed to differences in GA and birth weight, suggesting the identified markers as surrogates for the clinical variables rather than indicators of robust endotypes.

Studies might need to include alternative approaches as exemplified by genetic studies where holistic analyses were limited in

success,[68] whereas "reverse" approaches confirmed risk factors identified in other diseases.[64] Likewise, the focus on disease subgroups such as BPD-associated pulmonary hypertension might help to obtain first results.[69]

Our choice of a less stringent significance threshold to select metabolites and proteins for further analysis (<0.1) paid tribute to the limitations in sample number and noise.[70–72] However, the candidate biomarkers were identified through multiple subset analyses as well as different approaches such as random forest-based importance analysis. Despite this indication of robustness toward analysis techniques together with the biologically meaningful processes indicated by the biomarkers profile, the predictive power of the biomarker candidates needs to be re-assessed in an independent cohort while using different platforms for abundance detection, thereby overcoming bias through the pre-selection of metabolites and proteins at the same time.

While not allowing for endotype identification, latent factors with proteomic explanation demonstrated BPD features with pathophysiological relevance, underscoring the biological insight provided. Biomarkers identified by well-trained analysis strategies beyond the background of disease complexity can likely serve as robust markers for patient stratification in future studies next to clinical criteria.

CONFLICTS OF INTEREST

Fabian Theis consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, Cellarity, and has ownership interest in Dermagnostix GmbH and Cellarity. All other authors declare no conflicts of interest.

FUNDING INFORMATION

This work was funded by the Young Investigator Grant NWG VH-NG-829 by the Helmholtz Foundation and the Helmholtz Zentrum Munich, Germany, the German Center for Lung Research (DZL) (Federal Ministry of Education and Research in Germany (BMBF)) as well as the Research Training Group "Targets in Toxicology" (GRK2338) of the German Science and Research Organization (DFG). Additional financial support was provided by the Stiftung AtemWeg (LSS AIRR) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), CRC/TRR 359 (Project number 491676693) "Perinatal Development of Immune Cell Topology (PILOT)."

AUTHOR CONTRIBUTIONS

Juan Henao: Formal analysis, data curation, visualization, writing—original draft. Alida Kindt: Resources, data analysis. Tanja Seegmüller: Resources, writing—review & editing. Kai Foerster: Resources. Andreas W. Flemmer: Resources. Juergen Behr: Resources. Nikolaus Kneidinger: Resources. Marion Frankenberger: Resources. Fabian Theis: Supervision. Benjamin Schubert: Supervision, conceptualization. Markus List: Supervision, writing—review & editing. Anne Hilgendorff: Conceptualization, supervision, writing—review & editing, funding acquisition.

REFERENCES

1. Jobe AH. The new bronchopulmonary dysplasia. Curr Opin Pediatr. 2011;23: 167–172. doi: 10.1097/MOP.0b013e3283423e6b.
2. Wang S-H, Tsao P-N. Phenotypes of bronchopulmonary dysplasia. Int J Mol Sci. 2020;21: 6112. doi: 10.3390/ijms21176112.
3. McGeachie MJ, Yates KP, Zhou X, et al. Patterns of growth and decline in lung function in persistent childhood asthma. N Engl J Med. 2016;374: 1842–1852. doi: 10.1056/NEJMoa1513737.
4. Davidson LM, Berkelhamer SK. Bronchopulmonary dysplasia: Chronic lung disease of infancy and long-term pulmonary outcomes. J Clin Med Res. 2017;6: 4. doi: 10.3390/jcm6010004.
5. Battaglia M, Ahmed S, Anderson MS, et al. Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. Diabetes Care. 2020;43: 5–12. doi: 10.2337/dc19-0880.
6. Agache I, Akdis CA. Precision medicine and phenotypes, endotypes, genotypes, regiotypes, and theratypes of allergic diseases. J Clin Invest. 2019;129: 1493–1503. doi: 10.1172/JCI124611.
7. Carlton EF, Sontag MK, Younoszai A, et al. Reliability of echocardiographic indicators of pulmonary vascular disease in preterm infants at risk for bronchopulmonary dysplasia. J Pediatr. 2017;186: 29–33. doi: 10.1016/j.jpeds.2017.03.027.
8. Apitz C, Hansmann G, Schranz D. Hemodynamic assessment and acute pulmonary vasoreactivity testing in the evaluation of children with pulmonary vascular disease. Expert consensus statement on the diagnosis and treatment of paediatric pulmonary hypertension. The European Paediatric Pulmonary Vascular Disease Network, endorsed by ISHLT and DGPK. Heart. 2016;102(Suppl 2): ii23–ii29. doi: 10.1136/heartjnl-2014-307340.
9. Häfner F, Kindt A, Strobl K, et al. MRI pulmonary artery flow detects lung vascular pathology in preterms with lung disease. Eur Respir J. 2023;62: 2202445. doi: 10.1183/13993003.02445-2022.
10. Pierro M, Van Mechelen K, van Westering-Kroon E, Villamor-Martínez E, Villamor E. Endotypes of prematurity and phenotypes of bronchopulmonary dysplasia: Toward personalized neonatology. J Pers Med. 2022;12: 687. doi: 10.3390/jpm12050687.
11. Pierro M, Villamor-Martinez E, van Westering-Kroon E, Alvarez-Fuente M, Abman SH, Villamor E. Association of the dysfunctional placentation endotype of prematurity with bronchopulmonary dysplasia: A systematic review, meta-analysis and meta-regression. Thorax. 2022;77: 268–275. doi: 10.1136/thoraxjnl-2020-216485.
12. Wu KY, Jensen EA, White AM, et al. Characterization of disease phenotype in very preterm infants with severe bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2020;201: 1398–1406. doi: 10.1164/rccm.201907-1342OC.
13. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2001;163: 1723–1729. doi: 10.1164/ajrccm.163.7.2011060.
14. Jeon GW, Oh M, Lee J, Jun YH, Chang YS. Comparison of definitions of bronchopulmonary dysplasia to reflect the long-term outcomes of extremely preterm infants. Sci Rep. 2022;12: 18095. doi: 10.1038/s41598-022-22920-8.
15. Vyas-Read S, Logan JW, Cuna AC, et al. A comparison of newer classifications of bronchopulmonary dysplasia: Findings from the Children's Hospitals Neonatal Consortium Severe BPD Group. J Perinatol. 2021;42: 58–64. doi: 10.1038/s41372-021-01178-4.
16. Jing X, Jia S, Teng M, et al. Cellular senescence contributes to the progression of hyperoxic bronchopulmonary dysplasia. Am J Respir Cell Mol Biol. 2024;70: 94–109. doi: 10.1165/rcmb.2023-0038OC.
17. Parikh P, Britt RD Jr, Manlove LJ, et al. Hyperoxia-induced cellular senescence in fetal airway smooth muscle cells. Am J Respir Cell Mol Biol. 2019;61: 51–60. doi: 10.1165/rcmb.2018-0176OC.

18. Asikainen TM, White CW. Pulmonary antioxidant defenses in the preterm newborn with respiratory distress and bronchopulmonary dysplasia in evolution: Implications for antioxidant therapy. Antioxid Redox Signal. 2004;6: 155–167. doi: 10.1089/152308604771978462.
19. Cyr-Depauw C, Hurskainen M, Vadivel A, Mižíková I, Lesage F, Thébaud B. Characterization of the innate immune response in a novel murine model mimicking bronchopulmonary dysplasia. Pediatr Res. 2021;89: 803–813. doi: 10.1038/s41390-020-0967-6.
20. Windhorst AC, Heydarian M, Schwarz M, et al. Monocyte signature as a predictor of chronic lung disease in the preterm infant. Front Immunol. 2023;14: 1112608. doi: 10.3389/fimmu.2023.1112608.
21. Ofman G, Caballero MT, Alvarez Paggi D, et al. The discovery BPD (D-BPD) program: Study protocol of a prospective translational multicenter collaborative study to investigate determinants of chronic lung disease in very low birth weight infants. BMC Pediatr. 2019;19: 227. doi: 10.1186/s12887-019-1610-8.
22. Shukla VV, Ambalavanan N. Recent advances in bronchopulmonary dysplasia. Indian J Pediatr. 2021;88: 690–695. doi: 10.1007/s12098-021-03766-w.
23. Bhandari A, Bhandari V. Biomarkers in bronchopulmonary dysplasia. Paediatr Respir Rev. 2013;14: 173–179. doi: 10.1016/j.prrv.2013.02.008.
24. Piersigilli F, Lam TT, Vernocchi P, et al. Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. Metabolomics. 2019;15: 20. doi: 10.1007/s11306-019-1482-9.
25. Kindt ASD, Förster KM, Cochius-den Otter SCM, et al. Validation of disease-specific biomarkers for the early detection of bronchopulmonary dysplasia. Pediatr Res. 2023;93: 625–632. doi: 10.1038/s41390-022-02093-w.
26. Ahmed S, Odumade OA, van Zalm P, et al. Urine proteomics for noninvasive monitoring of biomarkers in bronchopulmonary dysplasia. Neonatology. 2022;119: 193–203. doi: 10.1159/000520680.
27. Zielinski JM, Luke JJ, Guglietta S, Krieg C. High throughput multi-omics approaches for clinical trial evaluation and drug discovery. Front Immunol. 2021;12: 590742. doi: 10.3389/fimmu.2021.590742.
28. Dar MA, Arafah A, Bhat KA, et al. Multiomics technologies: Role in disease biomarker discoveries and therapeutics. Brief Funct Genomics. 2023;22: 76–96. doi: 10.1093/bfgp/elac017.
29. Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The need for multi-omics biomarker signatures in precision medicine. Int J Mol Sci. 2019;20: 4781. doi: 10.3390/ijms20194781.
30. Argelaguet R, Velten B, Arnol D, et al. Multi-omics factor analysis – A framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14: e8124. doi: 10.15252/msb.20178124.
31. Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21: 111. doi: 10.1186/s13059-020-02015-1.
32. Um-Bergström P, Hallberg J, Pourbazargan M, et al. Pulmonary outcomes in adults with a history of bronchopulmonary dysplasia differ from patients with asthma. Respir Res. 2019;20: 102. doi: 10.1186/s12931-019-1075-1.
33. McGrath-Morrow SA, Collaco JM. Bronchopulmonary dysplasia: What are its links to COPD? Ther Adv Respir Dis. 2019;13: 1753466619892492. doi: 10.1177/1753466619892492.
34. Roos AB, Berg T, Nord M. A relationship between epithelial maturation, bronchopulmonary dysplasia, and chronic obstructive pulmonary disease. Pulm Med. 2012;2012: 196194. doi: 10.1155/2012/196194.
35. Ota C, Baarsma HA, Wagner DE, Hilgendorff A, Königshoff M. Linking bronchopulmonary dysplasia to adult chronic lung diseases: Role of WNT signaling. Mol Cell Pediatr. 2016;3: 34. doi: 10.1186/s40348-016-0062-6.
36. Sucre JMS, Deutsch GH, Jetter CS, et al. A shared pattern of β-catenin activation in bronchopulmonary dysplasia and idiopathic pulmonary fibrosis. Am J Pathol. 2018;188: 853–862. doi: 10.1016/j.ajpath.2017.12.004.
37. Couchard M, Polge J, Bomsel F. [Hyaline membrane disease: Diagnosis, radiologic surveillance, treatment and complications]. Ann Radiol. 1974;17: 669–683.
38. Rohloff JC, Gelinas AD, Jarvis TC, et al. Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. Mol Ther Nucleic Acids. 2014;3: e201. doi: 10.1038/mtna.2014.49.
39. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One. 2010;5: e15004. doi: 10.1371/journal.pone.0015004.
40. Holle R, Happich M, Löwel H, Wichmann HE; MONICA/KORA Study Group. KORA – A research platform for population based health research. Gesundheitswesen. 2005;67(Suppl 1): S19–S25. doi: 10.1055/s-2005-858235.
41. Stekhoven DJ, Bühlmann P. MissForest – Non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28: 112–118. doi: 10.1093/bioinformatics/btr597.
42. Wright MN, Ziegler A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw. 2017;77: 1–17. doi: 10.18637/jss.v077.i01.
43. Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 2002;18(Suppl 1): S96–S104. doi: 10.1093/bioinformatics/18.suppl_1.s96.
44. Čuklina J, Lee CH, Williams EG, et al. Diagnostics and correction of batch effects in large-scale proteomic studies: A tutorial. Mol Syst Biol. 2021;17: e10240. doi: 10.15252/msb.202110240.
45. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47. doi: 10.1093/nar/gkv007.
46. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 – An R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. 2020;9: ELIXIR-709. doi: 10.12688/f1000research.24956.2.
47. Gillespie M, Jassal B, Stephan R, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50: D687–D692. doi: 10.1093/nar/gkab1028.
48. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res. 2016;18: 1–5.
49. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12: 2825–2830. doi: 10.5555/1953048.2078195.
50. Higgins RD, Jobe AH, Koso-Thomas M, et al. Bronchopulmonary dysplasia: Executive summary of a workshop. J Pediatr. 2018;197: 300–308. doi: 10.1016/j.jpeds.2018.01.043.
51. Jensen EA, Dysart K, Gantz MG, et al. The diagnosis of bronchopulmonary dysplasia in very preterm infants. An evidence-based approach. Am J Respir Crit Care Med. 2019;200: 751–759. doi: 10.1164/rccm.201812-2348OC.
52. Isayama T, Lee SK, Yang J, et al. Revisiting the definition of bronchopulmonary dysplasia: Effect of changing panoply of respiratory support for preterm neonates. JAMA Pediatr. 2017;171: 271–279. doi: 10.1001/jamapediatrics.2016.4141.
53. Heydarian M, Schulz C, Stoeger T, Hilgendorff A. Association of immune cell recruitment and BPD development. Mol Cell Pediatr. 2022;9: 16. doi: 10.1186/s40348-022-00148-w.
54. Ambalavanan N, Carlo WA, D'Angio CT, et al. Cytokines associated with bronchopulmonary dysplasia or death in extremely low birth weight infants. Pediatrics. 2009;123: 1132–1141. doi: 10.1542/peds.2008-0526.
55. Eldredge LC, Creasy RS, Presnell S, 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–L56. doi: 10.1152/ajplung.00060.2019.

56. Rapp M, Wintergerst MWM, Kunz WG, et al. CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes. J Exp Med. 2019;216: 1170–1181. doi: 10.1084/jem.20170277.
57. Kandasamy J, Roane C, Szalai A, Ambalavanan N. Serum eotaxin-1 is increased in extremely-low-birth-weight infants with bronchopulmonary dysplasia or death. Pediatr Res. 2015;78: 498–504. doi: 10.1038/pr.2015.152.
58. Shrestha D, Ye GX, Stabley D, et al. Pulmonary immune cell transcriptome changes in double-hit model of BPD induced by chorioamnionitis and postnatal hyperoxia. Pediatr Res. 2021;90: 565–575. doi: 10.1038/s41390-020-01319-z.
59. Euchner J, Sprissler J, Cathomen T, et al. Natural killer cells generated from human induced pluripotent stem cells mature to $CD56^{bright}CD16^{+}NKp80^{+/-}$ in-vitro and express KIR2DL2/DL3 and KIR3DL1. Front Immunol. 2021;12: 640672. doi: 10.3389/fimmu.2021.640672.
60. Alachkar N, Ugarte R, Huang E, et al. Stem cell factor, interleukin-16, and interleukin-2 receptor alpha are predictive biomarkers for delayed and slow graft function. Transplant Proc. 2010;42: 3399–3405. doi: 10.1016/j.transproceed.2010.06.013.
61. Wang Y, Khan A, Heringer-Walther S, Schultheiss H-P, Moreira M da CV, Walther T. Prognostic value of circulating levels of stem cell growth factor beta (SCGF beta) in patients with Chagas' disease and idiopathic dilated cardiomyopathy. Cytokine. 2013;61: 728–731. doi: 10.1016/j.cyto.2012.12.018.
62. Bouzina H, Rådegran G. Low plasma stem cell factor combined with high transforming growth factor-α identifies high-risk patients in pulmonary arterial hypertension. ERJ Open Res. 2018;4: 00035-2018. doi: 10.1183/23120541.00035-2018.
63. Thébaud B, Lalu M, Renesme L, et al. Benefits and obstacles to cell therapy in neonates: The INCuBAToR (Innovative Neonatal Cellular Therapy for Bronchopulmonary Dysplasia: Accelerating Translation of Research). Stem Cells Transl Med. 2021;10: 968–975. doi: 10.1002/sctm.20-0508.
64. Nissen G, Hinsenbrock S, Rausch TK, et al. Lung function of preterm children parsed by a polygenic risk score for adult COPD. NEJM Evid. 2023;2: EVIDoa2200279. doi: 10.1056/EVIDoa2200279.
65. Goss KN, Beshish AG, Barton GP, et al. Early pulmonary vascular disease in young adults born preterm. Am J Respir Crit Care Med. 2018;198: 1549–1558. doi: 10.1164/rccm.201710-2016OC.
66. Twisselmann N, Pagel J, Künstner A, et al. Hyperoxia/hypoxia exposure primes a sustained pro-inflammatory profile of preterm infant macrophages upon LPS stimulation. Front Immunol. 2021;12: 762789. doi: 10.3389/fimmu.2021.762789.
67. Moreira AG, Arora T, Arya S, Winter C, Valadie CT, Kwinta P. Leveraging transcriptomics to develop bronchopulmonary dysplasia endotypes: A concept paper. Respir Res. 2023;24: 284. doi: 10.1186/s12931-023-02596-y.
68. Shaw GM, O'Brodovich HM. Progress in understanding the genetics of bronchopulmonary dysplasia. Semin Perinatol. 2013;37: 85–93. doi: 10.1053/j.semperi.2013.01.004.
69. Siddaiah R, Oji-Mmuo C, Aluquin V, et al. Multi-omics endotype of preterm infants with bronchopulmonary dysplasia and pulmonary hypertension. medRxiv. 2022. doi: 10.1101/2022.11.03.22281890.
70. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. Am Stat. 2019;73: 235–245. doi: 10.1080/00031305.2018.1527253.
71. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p < 0.05". Am Stat. 2019;73: 1–19. doi: 10.1080/00031305.2019.1583913.
72. Amrhein V, Trafimow D, Greenland S. Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. Am Stat. 2019;73: 262–270. doi: 10.1080/00031305.2018.1543137.