Biological Mechanisms Underlying Prematurity-Associated Lung Disease in School-Aged Children

A thesis submitted in consideration for the degree of:

Doctor of Philosophy (PhD)

By

Dr Christopher William Course (BSc, MBBCh, MRCPCH)

Department of Child Health

School of Medicine

Cardiff University

Cardiff

CF14 4XN

2024

Summary

This thesis uses data collected from the Respiratory Outcomes in Neonates (RHiNO) Study and aimed to identify the underlying mechanisms of prematurity-associated lung disease (PLD) and the effect of inhaler treatments on any implicated biological processes. Exhaled breath condensate (EBC) and urine samples were collected from preterm- and term-born school-aged children both at baseline and after twelve weeks of inhaled therapies. Untargeted proteomic and metabolomic analyses have been used to investigate the mechanisms underlying a range of phenotypes of PLD, based upon neonatal history of bronchopulmonary dysplasia (BPD) and current spirometry patterns, as well as post-treatment effects on any alterations identified.

The results demonstrated that preterm-born children with a history of BPD had detectable EBC proteome changes indicative of pulmonary structural alterations, including a reduced abundance of desmosome-constituent proteins. These changes were reversed with combined inhaler therapy (corticosteroid and long-acting β_2 agonists), increasing these protein's abundances to levels comparable to term-born subjects. The EBC metabolome suggested changes in pulmonary antioxidant mechanisms in those with a history of BPD, including significant reductions in metabolites associated with glutathione metabolism, although not revealing any associations with current lung function.

Urinary analyses demonstrated proteomic and metabolomic alterations associated with current spirometry patterns. Prematurity-associated obstructive lung disease (POLD) was associated with proteomic changes linked with increases in neutrophil activity and tissue-remodelling proteases, with metabolomic changes also suggestive of impairments in glutathione metabolism. The urinary proteome of those with prematurity-associated preserved ratio impaired spirometry (pPRISm) showed changes associated with inflammatory

processes and altered T-lymphocyte biology, however minimal metabolomic changes were identified for this group.

Overall, these results demonstrated differing biological mechanisms underlying different phenotypes of PLD, including both for those with a history of BPD and those with current impaired lung function results, with evidence that some of these mechanisms are potentially modifiable with inhaled therapies.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I have many people to acknowledge for their help and support in undertaking this research project and thesis. Firstly, I would like to offer my sincere thanks to Professor Sailesh Kotecha for the opportunities to be involved in both the RHiNO and AZTEC studies, and to undertake this PhD project, and for several years of supporting my research ambitions to reach this point. I have greatly valued his advice, help, guidance and mentorship throughout this journey.

I would also like to thank my co-supervisors Dr Sarah Kotecha and Dr John Watkins. Sarah has been a valuable source of day-to-day support during my time at the Department of Child Health, and John has always shown generosity in giving his time to help with my statistical questions. I would also like to acknowledge my other colleagues from the department, Dr John Lowe and Dr Ali Aboklaish for their support and advice.

I must acknowledge Dr Michael Cousins and Dr Kylie Hart, who recruited the children to the RHiNO study, undertook their clinical assessments and spirometry measurements, and collected exhaled breath condensate and urine samples from the participants. I used these samples to arrange the '-omics' analyses, with my thanks to the teams at University of Bristol Proteomics Facility and West Coast Metabolomics Centre, University of California Davis for supplying the mass spectrometry data I used for the data analyses contained in this thesis.

Next, I must specifically thank Dr Phil Lewis and Dr Kate Heesom from the Proteomics Facility at the University of Bristol. Phil has been a guiding light through much of this project, and his patience and encouragement in helping me to learn R coding cannot be understated. Both Phil and Kate have shown great generosity of spirit and time in answering questions, giving their advice, and sharing their experience through some of the more challenging aspects of this project.

Dedication

To Kate, for her unwavering love and support,

and for my daughters, Naomi and Jasmine.

List of Publications

Peer-reviewed Publications

Course, C. W. Lewis P., Kotecha, S.J., et al. 2024. Similarities of metabolomic disturbances in prematurity-associated obstructive lung disease to chronic obstructive pulmonary disease. Scientific Reports (14), article number: 23294 (10.1038/s41598-024-73704-1).

Course C.W., Kotecha E.A., Course K, Kotecha S. 2024. The respiratory consequences of preterm birth: from infancy to adulthood. British Journal of Hospital Medicine. 85(8) https://doi.org/10.12968/hmed.2024.0141

Watkins W.J., **Course C.W.**, Cousins M., et al. 2024. Impact of ambient air pollution on lung function in preterm-born school-aged children. Thorax 79:553-563. (*Joint first author)*

Course, C. W., Kotecha, S. J. and Kotecha, S. 2024. Evolving treatment for prematurity-associated lung disease. Translational Pediatrics 13(1), pp. 1-5. (10.21037/tp-23-505)

Course, C. W. Lewis P., Kotecha, S.J., et al. 2023. Evidence of abnormality in glutathione metabolism in the airways of preterm born children with a history of bronchopulmonary dysplasia. Scientific Reports 13(1), article number: 19465. (10.1038/s41598-023-46499-w)

Course, C. W. Lewis P., Kotecha, S.J., et al. 2023. Characterizing the urinary proteome of prematurity-associated lung disease in school-aged children. Respiratory Research 24, article number: 191. (10.1186/s12931-023-02494-3)

Gibbons, J. T., **Course, C.W.**, Evans, E. E., et al. 2023. Increasing airway obstruction through life following bronchopulmonary dysplasia: a meta-analysis. ERJ Open Research 9(2), article number: 00046-2023. (10.1183/23120541.00046-2023)

Course, C. W. Lewis P., Kotecha, S.J., et al. 2023. Modulation of pulmonary desmosomes by inhaler therapy in preterm-born children with bronchopulmonary dysplasia. Scientific Reports 13, article number: 7330. (10.1038/s41598-023-34233-5)

Course, C. W., Kotecha, S. J., Cousins, M., et al. 2023. Association of gestation and fetal growth restriction on cardiovascular health in preterm-born children. Journal of Pediatrics 255, pp. 42-49. (10.1016/j.jpeds.2022.09.057)

Kotecha, S. J., Gibbons, J. T. D., **Course, C. W.**, et al. 2022. Geographical differences and temporal improvements in forced expiratory volume in 1 second of preterm-born children. JAMA Pediatrics 176(9), pp. 867-877. (10.1001/jamapediatrics.2022.1990)

Kotecha, S. J., **Course, C. W.**, Jones, K. E., et al. 2022. Follow-up study of infants recruited to the randomised, placebo-controlled trial of azithromycin for the prevention of chronic lung disease of prematurity in preterm infants: study protocol for the AZTEC-FU study. Trials 23, article number: 796. (10.1186/s13063-022-06730-x)

Oral Presentations

Course C. W. 'Neonatal Research as a Trainee in Wales', Pop Watkins Guest Lecture, Welsh Paediatric Society 50th Anniversary Conference, Lake Vyrnwy, Powys, Wales 15th-16th June 2023

Poster Presentations

Course, C.W., Lewis P., Kotecha, S.J., Cousins, M., Hart, K., Watkins, W.J., Heesom, K.J., Kotecha, S. Pulmonary desmosome protein expression increases following inhaler therapy in preterm-born children with bronchopulmonary dysplasia. Poster presentation at the European Respiratory Society Congress 2024, Vienna, Austria, 7th-11th September 2024.

Course, C.W., Lewis P., Kotecha, S.J., Cousins, M., Hart, K., Heesom, K.J., Watkins, W.J., Kotecha, S. Impaired responses to oxidative stress in the airways of preterm-born school-aged children with a history of bronchopulmonary dysplasia. Poster presentation at the European Respiratory Society Congress 2024, Vienna, Austria, 7th-11th September 2024.

Course, C.W., Watkins, W.J., Kotecha, S.J., Kotecha, S. Association of air pollution exposure at birth and in childhood with lung function in preterm-born school-aged children. Poster presentation at the European Respiratory Society Congress 2022, Barcelona, Spain, 4th-6th September 2022

Course, C.W., Gallacher, D., Aboklaish, A., Zhang, L., Marchesi, J., Kotecha, S. Azithromycin resistance in the stool of preterm infants at risk of chronic lung disease of prematurity. Poster presentation at the European Respiratory Society Congress 2022, Barcelona, Spain, 4th-6th September 2022

Course, C.W., Kotecha, S.J., Cousins, M., Hart, K., Watkins, W.J., Kotecha, S. Cardiovascular response to exercise in preterm-born children with and without low lung function. Poster presentation at the European Respiratory Society Congress 2022, Barcelona, Spain, 4th-6th September 2022

Abbreviations

ANOVA	Analysis of Variance
ATS	American Thoracic Society
BPD	Bronchopulmonary Dysplasia
COPD	Chronic Obstructive Pulmonary Disease
EBC	Exhaled Breath Condensate
ELF	Epithelial Lining Fluid
ERS	European Respiratory Society
FEV_1	Percent Predicted Forced Expiratory Volume in One Second
FEV_1/ FVC	Ratio between forced expiratory volume in the first 1 second of a forced expiratory manoeuvre and forced vital capacity
FVC	Percent Predicted Forced Vital Capacity
GC	Gas Chromatography
GLI	Global Lung Initiative
HPLC	High Performance Liquid Chromatography
ICS	Inhaled Corticosteroids
ICS/LABA	Inhaled Corticosteroids in Combination with Long Acting β_2 Agonist
IPA	Ingenuity Pathways Analysis®
IUGR	Intrauterine Growth Restriction
LC	Liquid Chromatography
Log_2FC	Logarithm to Base 2-Fold Change
m/z	Mass to Charge Ratio
MS	Mass Spectrometry
MSEA	Metabolite Set Enrichment Analysis
NHS	National Health Service

NICHD	National Institute for Health and Human Development
PDA	Patent Ductus Arteriosus
PLD	Prematurity-associated Lung Disease
PMA	Post Menstrual Age
POLD	Prematurity-associated Obstructive Lung Disease
pPRISm	Prematurity-associated Preserved Ratio with Impaired Spirometry
PT_c	Preterm-born Control
PT_{low}	Preterm-born with Low Lung Function ($FEV_1 \leq 85\%$ predicted)
RANOPS	Respiratory and Neurological Outcomes in Children Born Preterm Study
RCT	Randomised Controlled Trial
RDS	Respiratory Distress Syndrome
RHiNO	Respiratory Health Outcomes in Neonates Study
SEM	Standard Error of the Mean
TMT	Tandem Mass Tag
TOF-MS	Time-of-flight Mass Spectrometry
WIMD	Welsh Index of Multiple Deprivation

1 Introduction

The survival of infants born prematurely (<37 week's gestation) has improved significantly over the last thirty years in line with advances in modern antenatal care and neonatal intensive care. Despite continuing improvements in mortality, survivors still experience a considerable burden of long-term morbidity, with the respiratory system being one of the most affected, along with neurodevelopmental outcomes. These co-morbidities extend beyond infancy and will impact a proportion of survivors into childhood, adolescence, and later adult life. There is a growing need to understand the biological mechanisms underlying this pulmonary dysfunction to identify individuals at risk and to develop potentially targeted therapeutic agents.

In this chapter, I will provide an overview of preterm birth, and the short- and long-term respiratory consequences associated with preterm birth. In addition, I will discuss two "-omics" methods for analysing biological samples, namely proteomics and metabolomics. Both are gaining increasing interest for understanding mechanisms of respiratory diseases, identifying biomarkers for clinical phenotypes, and assessing response to treatments. Finally, I shall detail my hypotheses, and the specific research aims this thesis will address.

1.1 Preterm birth

1.1.1 Epidemiology of Preterm Birth

Preterm birth is defined by the World Health Organisation as birth before 37 weeks of completed gestation, or less than 259 days since the first day of the mother's last menstrual period (Howson et al., 2012). The degree of preterm birth is further subdivided into three

categories dependent on the gestational age at birth (World Health Organisation, 2019), as shown in Table 1-1.

Gestation	Definition
<28 weeks	Extremely preterm
28 to <32 weeks	Very preterm
32 to <37 weeks	Moderate to late preterm

Table 1-1: WHO Classifications of preterm birth (Howson et al., 2012)

Current estimates suggest that the worldwide average rate of preterm birth is 10%, with marked geographical variation. Of those infants born preterm, the majority (approximately 85%) are in the moderate or late preterm group, with a minority (approximately 5%) in the extremely preterm group. The rate of preterm birth is increasing globally, especially in industrialised nations, where delivery at a preterm gestation may be recommended for maternal or fetal reasons, and where there are more healthcare resources to care for these vulnerable infants. In developed countries rates of preterm birth range from 5-8% in European nations to 12% in the USA (Blencowe et al., 2012, Ohuma et al., 2023).

1.1.2 Causes of Preterm Birth

The reasons for preterm birth are multifactorial in origin. Preterm birth can be broadly categorised into either spontaneous or following medical intervention, such as induction of labour or delivery by Caesarean section. Spontaneous preterm births following premature rupture of membranes (PROM) and onset of labour before 37 week's completed gestation occurs in approximately 3% of all pregnancies (Medina and Hill, 2006). Spontaneous preterm labour accounts for approximately 70% of all preterm births, with the remaining 30% being medically indicated deliveries, for either maternal or fetal health reasons (Goldenberg et al., 2008). The increase in preterm births in singleton pregnancies is largely driven by medical

intervention, whereas assisted reproductive technologies account for the increase in preterm births in multiple pregnancies (Goldenberg et al., 2008).

Of those spontaneous preterm births, approximately 40% follow a prolonged PROM. Prelabour PROM is defined as a spontaneous rupture of the membranes at <37 weeks gestation and at least one hour prior to the onset of contractions (Goldenberg et al., 2008).

The reasons for PROM in most cases are unknown, but asymptomatic intrauterine infection is thought to be a common cause. The majority of women with prelabour ROM begin labour within a few days, but some mother's will not begin contractions for weeks or months following premature ROM, increasing the risk of ascending infection to the fetus, as the protective barrier for the fetus is no longer present (Romero et al., 1988).
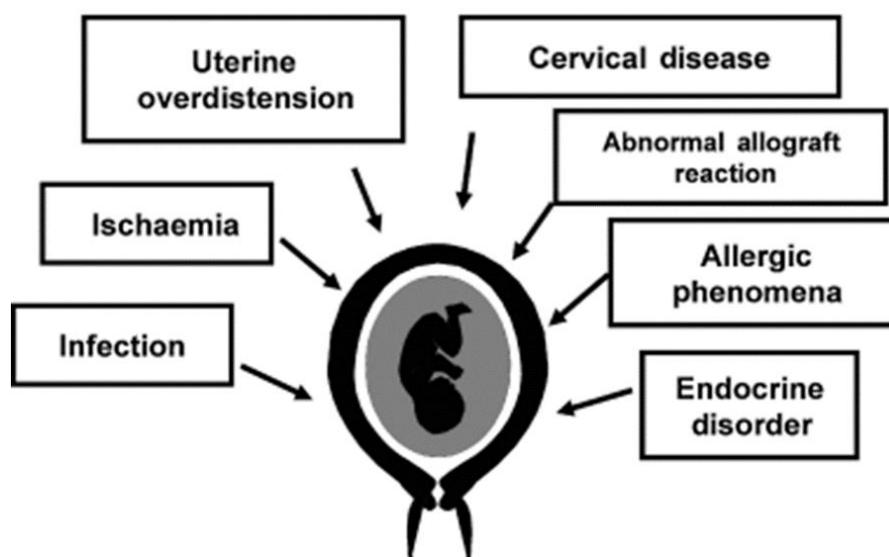
Figure 1-1: Pathological processes implicated in the onset of preterm labour (Romero et al., 2006). Image reproduced with permission of the rights holder.

The mechanisms underlying the onset of preterm labour remain incompletely understood. Whilst the process of labour is the same for term and preterm deliveries, namely increased

uterine contractility, cervical 'ripening' (dilatation and effacement), and decidua/membrane activation, it is thought that whilst term labour results from physiological activation, preterm labour is a pathological process that results from activation of one or more parts of the labour process (Romero et al., 2006). Pathophysiological causes for the initiation of preterm labour are detailed in Figure 1-1.

There are also several maternal risk factors which are associated with an increased risk of preterm birth. These include black ethnicity, low socioeconomic status, short time interval between pregnancies, previous preterm birth, multiple pregnancy, low body mass index (BMI), maternal nutritional deficiencies (vitamins and minerals such as zinc, folate and iron), and tobacco smoking (Goldenberg et al., 2008). Intrauterine infection is a frequent and important cause of preterm labour and birth, with an established causal link and clear, defined molecular mechanism. Infections can either be maternal (extrauterine) or intrauterine, which may be a subclinical infection or one that produces symptoms in the mother and fetus (chorioamnionitis), especially in the context of prolonged prelabour ROM (Romero et al., 2006, Romero et al., 1988). Potential routes for intrauterine infection are given in Figure 1-2.

Attempts have been made to reduce the modifiable risk factors associated with preterm birth or prevent or delay the underlying mechanisms responsible for preterm labour. A review of evidence for interventions targeted towards reducing preterm labour found that smoking cessation (relative risk 0.84; 95% confidence interval 0.72 - 0.98) and progesterone in high-risk women (0.65; 0.54 - 0.79) were the only effective interventions for reducing the risk of preterm birth (Barros et al., 2010).
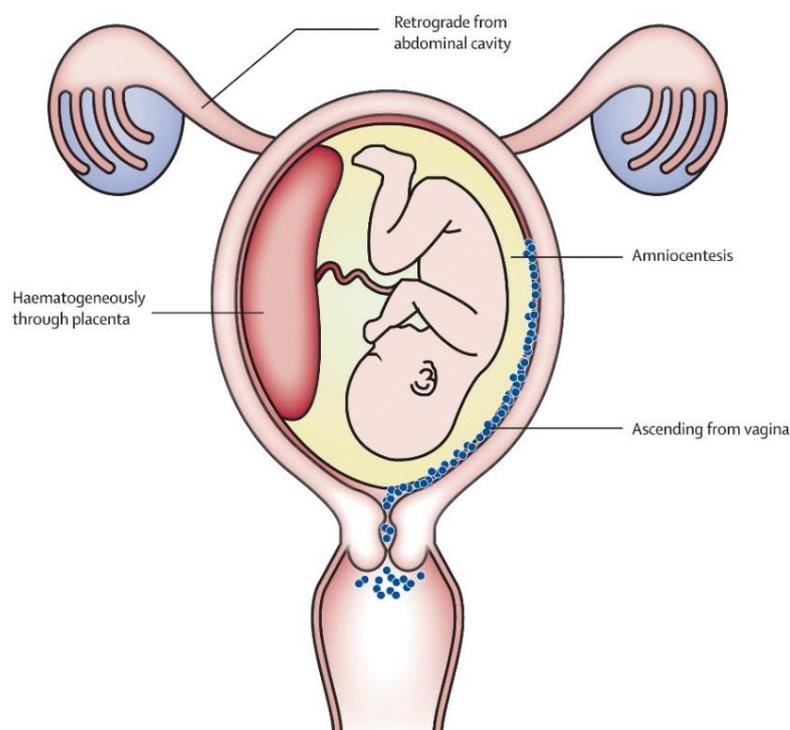
Figure 1-2: Potential routes for introduction of intrauterine infection

(Goldenberg et al., 2008). Image reproduced with permission of the rights holder.

Tocolysis (administering medications to slow or prevent contractions in labour) is effective only in the context of delaying birth to allow successful administration of antenatal corticosteroids, which is a strongly evidenced intervention for improving survival (0.78; 0.70 - 0.87) and immediate postpartum respiratory health (0.71; 0.65 - 0.78) of preterm infants (McGoldrick et al., 2020).

1.2 Morbidity and Mortality associated with Preterm Birth

1.2.1.1 *Mortality associated with preterm birth.*

Preterm birth is the leading cause of death in the neonatal population (defined as the first 28 days of life). In 2015, there were 2.7 million deaths globally in the neonatal period, with preterm birth accounting for approximately 943,000 (35%). Preterm birth remains the leading cause of childhood mortality up to the age of five years, with a cause specific mortality of 7.6

per 1000 live births (Liu et al., 2016). The limit of viability, the lowest gestational age at which it is possible for a fetus to born at and survive, has continued to reduce, especially over the last two decades. The British Association of Perinatal Medicine (BAPM) supports healthcare professionals in offering resuscitation to infants born at and above 22 weeks' gestation in the United Kingdom (BAPM, 2019). The EPICure studies from 1995 and from 2006 examined the outcomes for babies born extremely preterm between 22- and 26-weeks' gestation in the United Kingdom. EPICure found that even though there remained a high mortality for live-born infants (47%), there was a significant improvement in survival of 13% for these infants with advances in neonatal intensive care, from 40% in 1995 to 53% in 2006. Those infants born at the most preterm gestations (22 weeks) had the lowest survival (2% of live births) rising rapidly with each week of gestation thereafter (77% survival of live births at 26 weeks gestation) (Costeloe et al., 2012). Santhakumaran et al. have published more recent data covering survival of infants born <32 weeks' gestation who were admitted to English neonatal units between 2008 and 2014 (Figure 1-3). The authors found that overall survival to discharge continued to increase over this period from 88.0% to 91.3%. Overall, survival to discharge of 22-week gestation infants was 17.9% increasing up to 98.1% for 31-week gestation infants (Santhakumaran et al., 2018). Although these figures seem markedly higher than the EPICure data, it is worth noting that the data used by Santhakumaran et al. comes from the National Neonatal Research Database (NNRD) which collects data from all admissions to neonatal units in the UK but does not include stillbirths or delivery room deaths.

6
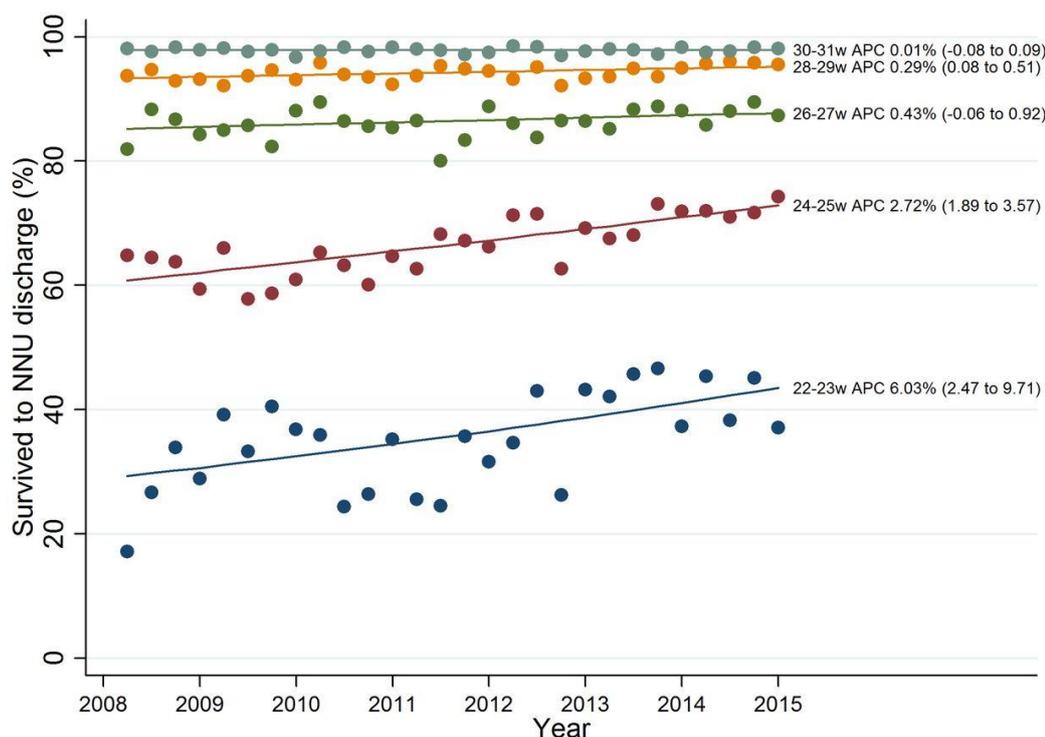
Figure 1-3: Survival rates to discharge in preterm born infants admitted to English neonatal units at different gestational ages between 2008 and 2015

(Santhakumaran et al., 2018). Figure available under Creative Commons CC BY 4.0 license.

1.2.1.2   Morbidity associated with preterm birth.

Preterm infants are vulnerable to numerous pathologies affecting the respiratory (neonatal respiratory distress syndrome), cardiovascular (patent ductus arteriosus, haemodynamic instability), neurological (intraventricular haemorrhage, periventricular leukomalacia) and gastrointestinal (necrotizing enterocolitis) systems immediately following their birth and for the subsequent weeks until reaching term-corrected age. In addition, owing to an immature immune system, they are vulnerable to systemic infection from bacterial, viral and fungal pathogens. Even with advances in medical management over the last twenty years, there remains a significant disease burden in this population (Berrington et al., 2012, Edwards et al., 2024). As a preterm infant is born whilst their major body systems are at an immature stage, they are vulnerable to pathological insult from ex utero organ growth and

development, which can be compounded by iatrogenic injury resulting from medical interventions required to support their survival, such as mechanical ventilation and supplemental oxygen.

In addition, for those preterm infants who survive to discharge home, preterm birth represents a significant cause of morbidity in later childhood. Preterm born children are at particular risk from longer term respiratory and neurological/neurodevelopmental pathologies, as well as issues with their vision, hearing, and psychosocial skills (Raju et al., 2017).

1.3 Lung Development

Lung development begins early in embryonic life and spans a series of tightly regulated developmental stages as shown in Figure 1-4. As early as three weeks post-conception, the respiratory tract begins forming as an out-growth of the ventral wall of the primitive foregut endoderm. By around seven weeks of embryonic life, the epithelial cells of the foregut have invaded the mesenchyme of the forming trachea, and the first phases of airway branching have occurred, forming the left and right main bronchi, and the lobar and segmental bronchi.

The pseudoglandular phase lasts until 17 weeks' gestation, during which there is further branching of the respiratory tree and evolving pulmonary vasculature, with 70% of the total airway generated by 14 weeks' gestation. There is further differentiation of cell types to begin forming the structures of the adult airway, such as cartilage, submucosal glands, bronchial smooth muscle, and epithelial cells. By 17 weeks, all the conducting airways and terminal bronchioles have formed (Joshi and Kotecha, 2007).

Foetal stage

Embryonic
0-7 weeks

Pseudoglandular
7-17 weeks

Canalicular
17-27 weeks

Saccular/Alveolar
28 wks - term

pleura

Lung structure

Trachea	24 days
Extrapulmonary Main Bronchus	28 days
Bronchi 8 - 13 generations	4 -12 weeks
Bronchioli 3 - 10 generations	12 - 16 weeks
Terminal bronchiolus 1 generation	16 - 17 weeks
Respiratory bronchioli 3 - 5 generations	18 - 25 weeks
Alveolar ducts 2 - 3 generations	25 wk - 1yr
Alveoli 300 - 600 million ie 1000/acinus	30 wk - 2-3yr

Acinus

Figure 1-4: Diagram illustrating the stages of normal human lung development.

(Chakraborty et al., 2010). Image reproduced with permission of the rights holder.

The canalicular phase spans 17 weeks to 27 weeks' gestation and is characterised by the formation of the acinar structures of the airways, namely the respiratory bronchioles, alveolar ducts, and the development of primitive alveoli. The acini form the functional gas exchange unit of the lung. The canalicular phase is marked by two crucial steps in the development of the lung.

Firstly, type I and II pneumocytes begin to develop in the primitive alveolar structures (Kotecha, 2000). Type I pneumocytes form most of the flat epithelial lining of the alveolus

and provide a thin barrier between the air space and underlying capillary network. Type II pneumocytes are cubic structures that generally lie isolated amongst the type I pneumocytes. These cells are filled with lamellar bodies that release pulmonary surfactant into the alveolus (Konrad Morgenroth, 2008). Pulmonary surfactant is a complex mix of phospholipids, neutral lipids and apoproteins. Phosphatidylcholine is the major phospholipid component of mammalian surfactant and is the primary constituent responsible for lowering the surface tension within the alveolus at the air-liquid interface. This facilitates efficient respiratory mechanics and gas exchange (Chakraborty and Kotecha, 2013).

Secondly, the alveolar capillary barrier forms as pulmonary capillary walls thin sufficiently to allow gas exchange between the blood and acinar air spaces. Once these two processes have occurred, the blood-gas barrier is sufficiently effective to maintain life in an extremely preterm infant (Hislop, 2002). The saccular stage follows and lasts between 28 – 36 weeks' gestation. By this time, the branching of the airways is nearly complete, and lung growth is comprised of enlarging peripheral airways, with the airway walls continuing to thin, and further development of the acinar units, with dilatation of the acinar tubules to form saccules. This further increases the surface area for gas exchange. The alveolar lining continues to develop with type II pneumocytes increasing their lamellar body production.

The alveolar stage of lung development is the last to occur in utero and lasts up to at least two years of postnatal life. This stage is defined by further septations of the terminal airways and the alveoli begin to form their classic cup-shaped appearance. There is further postnatal lung growth, with the number of alveoli increasing at least until the age of 2-3 years, and size and surface area of acinar units increasing until after adolescence. The number of alveoli at

birth ranges from 20 to 50 million, with this number increasing to 300 to 800 million in adults (Joshi and Kotecha, 2007).

1.4 Respiratory Consequences of Preterm Birth

Respiratory insufficiency is the principal pathology affecting the preterm infant immediately after birth. This manifests early as neonatal respiratory distress syndrome (RDS), which can either resolve or progress to bronchopulmonary dysplasia (BPD), also known as chronic lung disease of prematurity (CLD).

1.4.1 Respiratory Distress Syndrome

RDS is the commonest respiratory morbidity affecting infants born prematurely, with increasing frequency and severity being associated with lower gestational age of birth and lower birthweight (Wen et al., 2019). Additional risk factors for development of RDS include male sex, white ethnicity, maternal diabetes mellitus, perinatal hypoxia and ischaemia and delivery in the absence of active labour (Li et al., 2019, Reuter et al., 2014). The primary underlying pathophysiology of RDS is a deficiency of pulmonary surfactant, resulting in microatelectasis and low lung volumes. RDS usually presents within the first few hours of birth and improves over the first three to four days of life, following diuresis and reduction of pulmonary oedema, and as type II pneumocytes begin endogenous surfactant production. Some infants with milder disease can be managed with non-invasive respiratory support, such as Continuous Positive Airway Pressure (CPAP) or High Flow Nasal Cannulae (HFNC), whereas others will require exogenous pulmonary surfactant replacement, either by minimally invasive techniques or via an endotracheal tube. If respiratory failure is severe, then infants often require mechanical ventilation (Reuter et al., 2014, Course and Chakraborty, 2020). The introduction of routine administration of antenatal corticosteroids to mothers with

threatened preterm labour has markedly improved the severity of RDS in preterm born infants (McGoldrick et al., 2020). For some preterm-born infants with RDS who require a prolonged period of ventilatory support, there is a risk of developing chronic respiratory pathology, termed BPD (or CLD). However, not all infants who develop BPD have a history of RDS.

1.4.2 Bronchopulmonary Dysplasia

BPD was first described by Northway et al. in 1967 (Northway et al., 1967). Their paper described injury to the airways and lung parenchyma secondary to positive pressure ventilation and high fractions of inspired oxygen over the first few weeks of life. This cohort of infants had a high rate of mortality and an average gestational age of 34 weeks. The lung injury was heterogeneous and characterized by marked pulmonary fibrosis and epithelial damage, with areas of hyperinflation and atelectasis. In addition, there was hyperplasia of the airway smooth muscle and hypertensive changes to the pulmonary vasculature. Over the subsequent years, the management of preterm infants has evolved, with the introduction of neonatal ventilators, use of maternal antenatal corticosteroid administration and administration of exogenous pulmonary surfactant. The result of these interventions, amongst others, means that infants born at increasingly immature gestations, and therefore increasingly immature stages of lung development, are now surviving (Edwards et al., 2024). Consequently, the clinical and pathological presentation of BPD has changed. Rather than the injury to relatively well-developed lungs described by Northway et al. (Northway et al., 1967), contemporary BPD is a disease caused by alteration of normal lung development (Thebaud et al., 2019).

As described above, lung development in utero progresses through four distinct phases, embryonic (0-7 week's gestation), pseudoglandular (7-17 week's gestation), canalicular (17-27 week's gestation) and saccular/alveolar (28 week's gestation to term) (Chakraborty et al., 2010). It is during the canalicular phase that gas exchange within the lung becomes possible, and surfactant producing type II pneumocytes start to appear (Kotecha, 2000, Hislop, 2002). Preterm birth interrupts this normal lung development, and histological specimens from animal models and infants who have died from contemporary BPD demonstrate a more homogenous lung disease of alveolar simplification, whereby the alveoli appear larger, but fewer in number, with decreased septation, interstitial fibrosis and impaired growth of the pulmonary microvasculature (Jobe and Bancalari, 2001). Despite ongoing advances in neonatal medicine, the incidence of BPD is continuing to rise, unlike many other neonatal outcomes (Stoll et al., 2015, Bonadies et al., 2023), likely related to the improving survival of infants born at increasingly immature gestational ages (Edwards et al., 2024).

1.4.2.1   *Pathogenesis and risk factors for bronchopulmonary dysplasia*

Prematurity and low birth weight are the most important risk factors for the incidence of BPD, with the incidence being inversely proportional to both gestational age and weight at birth (Stoll et al., 2010). Whether there is a genetic predisposition to BPD is unclear. Although twin studies have demonstrated that genetic factors play an important role in the susceptibility to BPD, with an estimated heritability between 50-80% for moderate to severe BPD, study sample sizes have overall been small from single populations, and identifying candidate genes has been so far unsuccessful (Shaw and O'Brodovich, 2013).

The pathogenesis of BPD is multifactorial, with pulmonary inflammation being a common mechanism by which the lung is injured, and the normal lung development trajectory altered.

Chorioamnionitis, inflammation of the chorion and amnion within the uterus during pregnancy, usually secondary to infection, is a common precursor to preterm birth. In clinical chorioamnionitis the mother becomes symptomatic with symptoms of fever, a tender uterus, and purulent amniotic fluid. Histological chorioamnionitis can be asymptomatic, more common and is likely under-reported as it requires routine examination of the placenta post-partum (Papagianis et al., 2019). This intra-uterine infection and prenatal inflammation in the fetus can have both beneficial and harmful effects on the fetuses' respiratory health. Chorioamnionitis is associated with a reduced rate of RDS in infants born at <30 week's gestation (Lahra et al., 2009), but may be associated with a longer duration of mechanical ventilation and pulmonary hypertension (Yum et al., 2018). Whilst histological chorioamnionitis appears to increase the risk of developing BPD (Jain et al., 2024), this relationship is complicated by birthweight and postnatal events, such as prolonged mechanical ventilation and episodes of sepsis (Papagianis et al., 2019).

Mechanical ventilation has been shown to induce lung inflammation in both animal and clinical studies. In a preterm lamb model, pulmonary inflammation is evident after two hours of mechanical ventilation with upregulation of Interleukin (IL)-1β, IL-6 and IL-8, all pro-inflammatory cytokines (Hillman et al., 2010, Brew et al., 2011). Neutrophil and macrophage infiltration of the lung parenchyma is evident after three-to-four weeks of mechanical ventilation with abnormal vascular development and heterogenous lung inflation (Albertine et al., 1999). Mechanical trauma to the lung from artificial ventilation has also been demonstrated in animal models, with evidence from a preterm lamb model of lung injury from artificial lung inflation with physiological inspiratory volumes in surfactant deficient lungs (Bjorklund et al., 1997). Evidence from a murine model demonstrated that prolonged mechanical ventilation (for 24 hours) reduced alveolar number by 50%, reduced alveolar

septation and resulted in a five-fold increase in cellular apoptosis within the lung (Mokres et al., 2010). In human studies, preterm infants who develop BPD have also been shown to have persistently elevated proinflammatory cytokines in bronchioalveolar lavage fluid (Chakraborty et al., 2013) and blood (Paananen et al., 2009). The role of oxygen toxicity to the preterm lung has been long established since Northway's original description of BPD (Northway et al., 1967) and has since been shown in both experimental models and clinical trials. As the preterm lung tissue is not meant to be exposed to ambient room air concentrations of oxygen, or therapeutically elevated levels of inspired oxygen, it is thought that oxygen mediated injury is a combination of the production of reactive oxygen species, coupled with immature intracellular antioxidant mechanisms, leading to DNA damage, lipid peroxidation and protein oxidation. High inspired oxygen fractions in a preterm baboon model have shown the development of fewer and larger alveoli, whilst high inspired oxygen fractions in a murine model have shown increased neutrophil infiltration and prostaglandin synthesis in lung tissue after seven days (Buczynski et al., 2013). However, in clinical practice how much oxygen to administer to prevent tissue hypoxia whilst also preventing toxicity is less clear. A large meta-analysis demonstrated that whilst targeting a lower range of oxygen saturations (as defined by pulse oximetry measurement) reduced the risk of BPD diagnosis, it also increased the risk of death at 18-24 months of life, and increased the risk of other major neonatal morbidities such as necrotizing enterocolitis (Askie et al., 2018).

Whilst postnatal systemic and pulmonary infections increase the risk for BPD (Beeton et al., 2011), there is increasing evidence that pulmonary colonisation with *Ureaplasma urealiticum* is a significant risk factor for the development of BPD, with estimates from metanalysis showing an increased odds ratio of 2.2 (95% confidence interval 1.42-3.47) for those infants positive for *Ureaplasma* after birth, an effect which does not appear to be related to

gestational age (Lowe et al., 2014). Macrolide antibiotics appear to be effective at eliminating *Ureaplasma* species, and decolonizing the infant lung (Viscardi et al., 2020); however, whether this translates to reducing the risk of BPD is under investigation (Lowe et al., 2020) and the effect on longer-term respiratory and neurological outcomes is unclear (Viscardi et al., 2021, Kotecha et al., 2022a). The role of the pulmonary microbiome in the development of BPD is also of current interest. It appears that a dysbiotic environment, a microbiome with reduced bacterial diversity and individual species predominance, is associated with an increase in IL-6 and IL-8 inflammatory markers in respiratory tract samples, and an increased risk for the development of BPD (Gallacher et al., 2020).

A patent ductus arteriosus (PDA) is a common finding in infants born prematurely, and following lung inflation and falling pulmonary arterial pressures, if the PDA remains large a significant amount of blood flow can shunt from the systemic to the pulmonary circulation. Increased pulmonary blood flow can compromise lung compliance and gas exchange and precipitate the development of pulmonary hypertension and right ventricular dysfunction (El-Khuffash et al., 2023). Whilst a PDA has been inconsistently associated with an increased risk of BPD across multiple studies (Hamrick et al., 2020), trials examining both the effect of medical prophylactic closure (Mitra et al., 2022), early, late (Clyman et al., 2019) and targeted treatment of haemodynamically significant PDAs (Hundscheid et al., 2023, Mitra et al., 2020, Gupta et al., 2024) have all shown no effect on reducing the risk of developing BPD, with some suggestion of an increased risk of BPD in those actively treated.

The pathogenesis of BPD is clearly multifactorial, and the underlying mechanisms remain incompletely understood. Although this makes preventing and treating such a condition clinically challenging, it also presents the opportunity for research to identify underlying

common pathways and mechanisms of injury, and the potential for multiple therapeutic targets.

1.4.2.2 Diagnosis and assessing severity of bronchopulmonary dysplasia

BPD is an unusual disease in that it is defined by its treatment, as opposed to a diagnostic test result or histopathological appearance. The diagnostic criteria for BPD have evolved as the disease process has changed with advances in neonatal intensive care (Ryan, 2006). BPD was first described by Northway et al. in 1967 as a form of prolonged neonatal respiratory distress syndrome in preterm infants (mean gestational age ~32 weeks, mean birthweight of ~1900g) with granularity and opacity on chest radiographs and development of hyaline membranes, alveolar atelectasis and emphysematous changes on histology (Northway et al., 1967). Bancalari et al. defined BPD clinically in 1979 focusing on the need for positive pressure ventilation for at least the first week of life and a persistent supplemental oxygen requirement at 28 days of age (Bancalari et al., 1979). In 1988, Shennan et al. showed that assessment of supplemental oxygen requirement at 36 week's post menstrual age (PMA) was a better predictor of respiratory pathology at two year's corrected age (Shennan et al., 1988).

With the introduction of the routine use of maternal antenatal corticosteroid administration in cases of anticipated preterm labour or delivery, exogenous pulmonary surfactant replacement, titrated supplementary oxygen administration and modern neonatal ventilators, the pathophysiological features underlying BPD have evolved over the last thirty years, transitioning from a disease of more mature infants to largely those born at <30 weeks' gestation and with a birth weight <1000g, markedly more immature infants from those reported by Northway et al. Thus, it became clear that improved diagnostic criteria were required.

Currently, the most widely accepted definition of BPD comes from a National Institute of Child Health and Human Development (NICHD) workshop, published in 2001, which defined BPD as requirement for supplemental oxygen for at least the first 28 days of life. This definition also focuses on defining severity when the preterm-born infant born at less than 32 weeks' gestation reaches 36 weeks' PMA. Those who were supplemental oxygen dependent at 28 days of age but who no longer require respiratory support or supplementary oxygen are classed as having "mild BPD"; those who require <30% oxygen have "moderate BPD"; and infants who require ≥30% supplemental oxygen or require any form of positive pressure respiratory support (either invasive or non-invasive) are classed as having "severe BPD". For infants born at or over 32 weeks' gestation, this assessment is made at 56 days of life (Table 1-2). This classification system acknowledges that some infants on supplemental oxygen therapy at 28 days of age but not at 36 weeks' PMA may also have underlying residual lung pathology (Jobe and Bancalari, 2001). A validation study by Ehrenkranz et al. in 2005 (Ehrenkranz et al., 2005) supported this classification system but recommended dividing the 'severe BPD' category into those infants who continue to need either invasive or non-invasive respiratory support and those who do not.

There have been further attempts to improve upon the 2001 NICHD criteria acknowledging the introduction of newer modalities of respiratory support, such as heated and humidified high-flow nasal cannula therapy (Higgins et al., 2018). This definition acknowledged both the degree of positive pressure support required and quantity of supplement oxygen received to grade moderate and severe BPD more precisely (Table 1-2).

As the diagnosis of BPD is based upon supplemental oxygen requirement, and methods of assessing supplemental oxygen requirements are not standardised between neonatal units

and physicians (Ellsbury et al., 2002), the concept of the supplemental oxygen reduction test has been introduced in many units, and especially in clinical trials, to diagnose BPD more accurately. This a physiological test performed at 36 weeks' PMA, in infants who require supplemental oxygen of <30% and are not receiving positive pressure support. A trial without supplemental oxygen is attempted to firmly confirm an absolute requirement for supplemental oxygen dependency with subsequent classification into mild or moderate BPD (Walsh et al., 2004). This standardisation approach is increasingly used to permit epidemiological comparison between centres and in randomised controlled trials (RCT) to objectively report diagnosis and severity of BPD. The optimal time to assess oxygen requirement to diagnose BPD has also been questioned. Data from the Canadian Neonatal Network suggest that assessing BPD severity at 40 week's PMA, using both oxygen requirement and degree of positive pressure respiratory support required, possibly provides better predictive ability than at 36 week's PMA, for future respiratory morbidity (including home oxygen and recurrent respiratory-related hospital admissions in the first 21 months of life, or having a tracheostomy) as well as later neurosensory impairments (Isayama et al., 2017).

Other proposed diagnostic criteria to diagnose BPD focus on the underlying pathophysiological phenotype. In 2020, Wu et al. proposed a set of subgroup criteria for infants with severe BPD, after extensive phenotyping including bronchoscopy/tracheoscopy, chest computed tomography with angiography (CTA) and echocardiography between 40- and 50-week's PMA. The infants were subsequently categorised according to the presence or absence of large airway disease (tracheomalacia/bronchomalacia), moderate-severe parenchymal lung disease (including hyper-expansion, air cysts, bullae and consolidation amongst others), and pulmonary arterial hypertension (Wu et al., 2020a). Although many phenotypes overlapped, the presence of pulmonary arterial hypertension and/or large airway disease was strongly associated with a need for tracheostomy or use of pulmonary

vasodilators and with death, but such associations were not noted with parenchymal lung disease. The presence of pulmonary arterial hypertension in severe BPD was significantly associated with an increased risk of pre-discharge mortality. These in-depth phenotyping classifications highlight that BPD is not a disease limited to the lung parenchyma and may have benefit in prognosticating for those infants most severely affected.

In summary, the definition of BPD has evolved with the NICHD definition still most commonly used. However, it is increasingly recognised that better definitions are required, especially as the current definitions are poor predictors of future lung disease, including lung function deficits in childhood and beyond.

NICHD 2001 Workshop Definition		
Gestational Age	**<32 weeks**	**≥32 weeks**
Assessment timepoint	36 weeks PMA or discharge home	>28 days but <56 days postnatal age/at discharge home
Treatment with supplementary oxygen for 28 days **plus**		
Mild BPD	Breathing room air	
Moderate BPD	Need for <30% supplementary oxygen	
Severe BPD	Need for ≥ 30% oxygen and/or positive pressure respiratory support	
NICHD 2016 Workshop Revised Definition for Moderate/Severe BPD		
Treatment with supplementary oxygen for 28 days plus requiring respiratory support at 36 weeks PMA		
Grade I	- nCPAP, NIPPV or HFNC ≥3 litres/min with FiO_2 0.21 - NC 1 - <3 litres/min, hood O_2 with FiO_2 0.22-0.29 - NC <1 litres/min with FiO_2 0.22-0.70	
Grade II	- IPPV with FiO_2 0.21 - nCPAP, NIPPV or HFNC ≥3 litres/min with FiO_2 0.22-0.29 - NC 1 - <3 litres/min, hood O_2 with FiO_2 ≥0.30 - NC <1 litres/min with FiO_2 0.70	
Grade III	- IPPV with FiO_2 >0.21 - nCPAP, NIPPV or HFNC ≥3 litres/min with FiO_2 ≥0.30	
Grade IIIa	Death between 14 days of postnatal age and 36 weeks owing to persistent parenchymal lung disease and respiratory failure that cannot be attributable to other neonatal morbidities	

Table 1-2: NICHD 2001 BPD Classification modified from (Jobe and Bancalari, 2001) NICHD 2016 BPD Classification modified from (Higgins et al., 2018).

IPPV: Invasive positive pressure ventilation, nCPAP: nasal continuous positive airway pressure, NIPPV: non-invasive intermittent positive pressure ventilation, HFNC: High-flow nasal canulae, NC: Nasal cannula, FiO_2: fraction of inspired oxygen

1.4.3   Respiratory consequences of preterm birth in infancy and childhood

Historically, most research on the longer-term respiratory outcomes for preterm-born individuals has focused on those subjects with a neonatal diagnosis of BPD. As discussed above, the assessment of BPD severity at 36 week's corrected gestational age was chosen as it provided the highest positive predictive value for respiratory morbidity over the first two years of life (Shennan et al., 1988). Despite the advances in perinatal care and improvement in survival of infants born extremely preterm (especially those born at <28 weeks' gestation), BPD rates remain largely unchanged or may even be increasing (Stoll et al., 2015). Although lung function tests in infants are challenging to perform, evidence suggests that over the first 12 months of life infants with severe BPD can develop obstructive, restrictive or mixed lung function deficits (Shepherd et al., 2018), which is likely related to changes related to impaired alveolar growth, small airways disease and gas trapping (Broughton et al., 2007, Iles and Edmunds, 1997, Shepherd et al., 2018). In addition, infants with BPD are at heightened risk of sleep-disordered breathing and sleep hypoxaemia, which may be clinically silent (Moyer-Mileur et al., 1996). Overnight polysomnography is useful for detecting these issues and titrating supplementary oxygen accordingly. Preventing hypoxic episodes improves airway resistance (Tay-Uyboco et al., 1989) and pulmonary arterial resistance (Balfour-Lynn et al., 2009), reversing some of the components of pulmonary arterial hypertension. A small proportion of infants with severe BPD will require long-term positive pressure ventilation, either due to significant parenchymal disease and/or larger airway disease, including tracheobrochomalacia (Hysinger et al., 2017) or subglottic stenosis secondary to prolonged periods of endotracheal intubation and invasive mechanical ventilation. In these cases, insertion of a tracheostomy can provide safe management of the airway and optimise ventilation (Wasserzug and DeRowe, 2016).

Infants with BPD are at higher risk of severe respiratory infections, especially over the first

two years of life. Commonly, these infections are viral in nature, with respiratory syncytial virus (RSV) and rhinovirus being common causative agents. Up to 50% of infants with BPD will require hospitalisation in the first two years of life for respiratory illness (Bhandari and Panitch, 2006), with BPD severity and duration of oxygen-dependency being associated with length of hospital stay. RSV infection during the first two years of life can be particularly severe in those who had BPD in the neonatal period, with increased need for intensive care unit admission and increased healthcare costs (Deshpande and Northern, 2003). Many centres routinely use the anti-RSV monoclonal antibody palivizumab especially targeting high-risk preterm infants, including those with BPD (Quinn et al., 2021). Given its half-life, palivizumab is administered monthly over the winter season. A newer monoclonal antibody, nirsevimab, has extended half-life, only requiring a single administration. It has been shown to be effective in reducing hospital admissions for RSV infection in both term and near-term infants (Hammitt et al., 2022) and in preterm-born infants (Domachowske et al., 2022).

BPD is associated with the development of pulmonary vascular disease and secondary pulmonary arterial hypertension, caused by pulmonary vascular remodelling. This leads to increased pulmonary vascular resistance which can precipitate right sided heart failure. Pulmonary arterial hypertension in infants with BPD is associated with high mortality in the first two years of life (Farrow and Steinhorn, 2012). Maintaining oxygen saturations >95% for infants with pulmonary arterial hypertension can potentially prevent progression of the disease. Sildenafil (a phosphodiesterase-5 inhibitor, which induces vasodilation by acting on the nitric oxide pathway) is commonly used in the management of pulmonary arterial hypertension to reduce pulmonary blood pressure, however there is a lack of robust data on its long-term use in pulmonary arterial hypertension with BPD (Hansmann et al., 2021).

Beyond infancy, preterm-born individuals have heightened risk of poor respiratory outcomes

in childhood and adolescence. Again, most published work to date focuses on individuals with a neonatal history of BPD. A recent large systematic review and meta-analysis of 86 studies demonstrated that across a wide range of ages (3 to 52 years old) preterm birth was associated with a 9.2% (95% confidence interval 8.0, 10.4%) reduction in forced expiratory volume in 1 second (FEV_1), but this increased to a 15.2% (14.2, 17.6%) reduction for those with a history of mild or moderate/severe BPD (Kotecha et al., 2022b). The studies included in this analysis spanned a broad range of years, with subjects born between 1961 and 2017, with most subjects studied being in childhood or adolescence. A meta-regression analysis demonstrated that these spirometry deficits appeared to be improving with advances in neonatal care over the years, but only for those preterm-born individuals with a history of BPD. There also appeared to be geographical variation for these outcomes around the world, with Scandinavian countries having better outcomes than Western European countries and the United States of America. Whether these differences are due to genetic, environmental, or wider socio-economic factors requires further investigation.

There is also evidence of increased risk of adverse respiratory outcomes for those born prematurely without a BPD diagnosis in childhood. Data from the UK-based Respiratory Health Outcomes in Neonates (RHiNO) study, which will be discussed in greater depth later in this chapter and throughout this thesis, examined lung function in over 500 preterm-born (<34 weeks' gestation) school-aged children who had experienced a contemporary standard of neonatal care. Within this cohort increasing gestational immaturity at birth and a history of intrauterine growth restriction (IUGR) were better predictors of childhood lung function than a history of BPD (Hart et al., 2022). Data from the large Avon Longitudinal Study of Parents and Children (ALSPAC) cohort demonstrated spirometry impairments were present at 8 – 9 years of age not only in those born extremely prematurely regardless of BPD status,

but also in those born in the moderate preterm category (33 – 34 weeks' gestation), whereas those born late preterm (35 – 36 weeks' gestation) had similar lung function values to those born at term. Those moderately preterm born individuals who had required mechanical ventilation in the neonatal period appeared to have greater spirometry impairments than those who did not. By 14 – 17 years of age, these spirometry deficits appeared to have generally improved, however the moderate preterm group still had significantly reduced values for mid-expiratory airflow measurements (FEF_{25-75}) and the ratio between FEV_1/Forced vital capacity (FVC) when compared to those born at term (Kotecha et al., 2012). Given that with modern neonatal care BPD is a disease predominantly seen in those born <30 weeks' gestation, this study suggests that later preterm birth and delivery at a saccular stage of lung development still impacts later lung function in childhood regardless of early neonatal outcomes. A large systematic review and metanalysis collating data on over 1,600 moderate-late preterm born individuals, who were predominantly school-aged children and adolescents, showed modest, but persistent deficits for FEV_1, FVC, FEV_1/FVC and FEF_{25-75} when compared to both term-born controls and internationally standardized reference values (Global Lung Initiative [GLI] references (Quanjer et al., 2012)), highlighting that individuals born moderate-late preterm are not catching up to population physiological levels and may fail to achieve optimal peak lung function in late adolescence/early adulthood (Du Berry et al., 2022).

1.4.4 Respiratory consequences of preterm birth in adulthood

With advances in neonatal care and improving survival over the last thirty years, more extremely preterm-born individuals are now reaching adulthood, and the respiratory consequences of prematurity in later life are becoming apparent. An individual participant data meta-analysis of spirometry data from eleven cohort studies, including over 900

preterm-born individuals (mean gestational age of 28 weeks, and 1054g birthweight) with a mean age of 21 years, showed significant impairments to expiratory flows (FEV$_1$ z-score -0.78 [95% confidence interval -0.96 to -0.61, p=<0.001] and FEF$_{25-75}$ z-score -0.88 [-1.12 to -0.65, p <0.001]) and an increased likelihood of an obstructive spirometry pattern when compared to term-born controls (Doyle et al., 2019a). Most of these individuals were born prior to the routine use of antenatal corticosteroids and exogenous surfactant replacement therapy. Longitudinal studies of preterm-born individuals from childhood to adulthood (aged 25 years) born at <28 weeks in the post-surfactant era have demonstrated a larger and persistent deficit in FEV$_1$ (mean z-score -0.97 [-1.23 to -0.71] p <0.001) and airway obstruction, which appear more pronounced in those with a neonatal history of BPD, when compared to term-born individuals (Doyle et al., 2019b). A large meta-analysis has also demonstrated that preterm-born individuals with a history of BPD have an increasingly obstructive spirometry phenotype with age (Gibbons et al., 2023). These finding are supported by another recent meta-analysis of data from sixteen studies of lung function in adults born at <28 week's gestation, finding a mean FEV$_1$ deficit of 14% (or one z-score), and a mean FEV$_1$/FVC near the lower limit of normal. These findings were more pronounced in those with neonatal history of BPD. Interestingly the authors also found no difference between those born before and after the introduction of pulmonary surfactant replacement (Lahn-Johannessen Lillebøe et al., 2024).

However, these findings are not exclusive to extremely preterm born individuals, with a longitudinal study of moderate-to-late (32 to <37 week's gestation) preterm-born young adults (aged 16 and 24 years at each respective assessment) demonstrating a smaller, but persistent deficit in FEV$_1$ (mean z-score -0.28 [-0.56 to -0.01], p=0.05) in males and an obstructive spirometry pattern in both sexes when compared to term-born controls, regardless of smoking status (Lundberg et al., 2024).

It is increasingly recognised that using BPD, a diagnosis that describes a troublesome clinical course in preterm infants, to label respiratory pathology many decades later is problematic and ignores many other risk factors (for example degree of prematurity, male sex, intrauterine growth restriction amongst others) and exposures through the rest of life (including lower respiratory tract infections, tobacco smoke exposure, air pollution exposure and socio-economic deprivation) that may help to stratify the risk and degree of prematurity-associated lung disease (Simpson et al., 2023, Watkins et al., 2024). The concept of the 'expiratory airflow trajectome' can be used to understand the impact of preterm birth on later lung function (

Figure 1-5). FEV_1 increases through childhood and adolescence, to reach a peak in the early-to-mid-twenties, followed by a gradual physiological decline with increasing age (Jakeways et al., 2003, Agusti and Faner, 2019). Those with a history of preterm birth may have altered postnatal lung development and risk not achieving the same peak of lung function, with either an earlier or accelerated decline in lung function as they age, and are likely at heightened risk of early development of chronic obstructive pulmonary disease (COPD) as they enter adult life (Simpson et al., 2023).
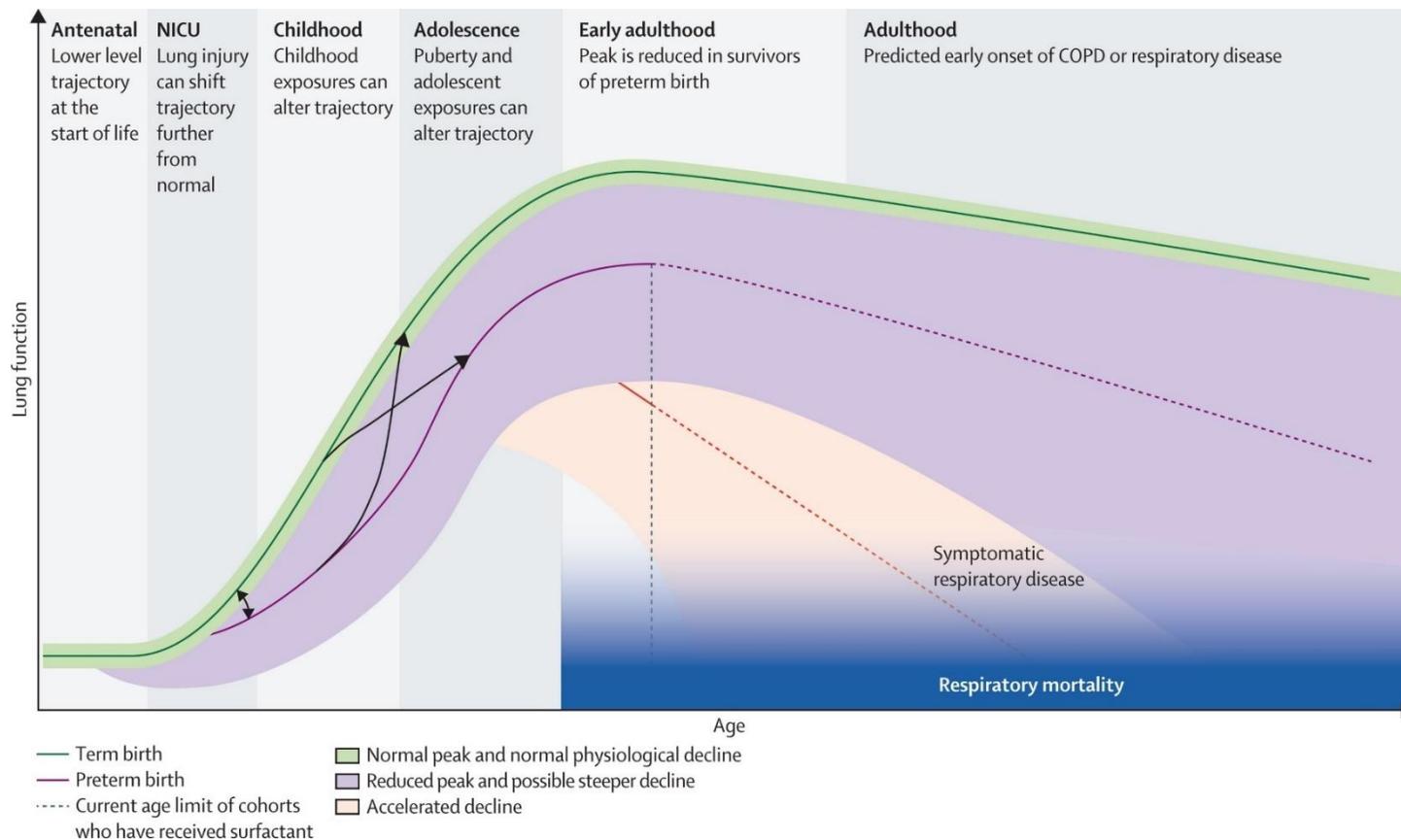
Figure 1-5: A series of potential expiratory airflow trajectories for individuals born prematurely compared with those born at term over the lifespan.

Many survivors of preterm birth will have abnormal lung development, reduced peak lung function, and potentially an increased rate of lung function decline, each of which places them at increased risk of chronic respiratory disease. Trajectories can be crossed through the life course (arrows). COPD: chronic obstructive pulmonary disease. NICU: neonatal intensive care unit (Simpson et al., 2023). Image reproduced with permission of the rights holder.

1.5 Proteomics and Metabolomics

The development of human disease involves a highly dynamic and interactive system of biological processes, involving innate biological features (such as genetics, transcription products such as proteins and peptides, and metabolites) and influences by environmental factors. Advances in high-throughput analysis techniques, based primarily on mass spectrometry and nuclear resonance spectroscopy, have advanced rapidly over the last thirty years allowing analysis of the genome, transcriptome, proteome and metabolome on large cohorts of human subjects (Sun and Hu, 2016). Genomics is a well-established methodology for examining the complete genetic complement of a tissue or biological samples and has proved useful in identifying the genetic loci associated with the development of human disease (Lonsdale et al., 2013). Following the advances in genome analysis, proteomics and metabolomics are two methods which have been developed to perform widespread analysis of biological processes occurring within a tissue or sample.

The proteome includes the entire set of proteins expressed by the genome (Wilkins et al., 1996), and most of the functional information of genes is characterised by the proteome (Aslam et al., 2017). Proteomic analysis can detect amino acid mutations, peptide isoforms and post-translational modifications to proteins that may all have an influence over protein function and cellular physiology (Nilsson et al., 2010). The metabolome represents the total low molecular weight (<2000 Da) metabolite content that is produced by a cell during metabolism, including amino acids, sugars, fatty acids, lipids and steroids, providing a direct representation of cellular biochemistry and physiological status (Chen et al., 2022). The metabolome reflects the combined internal effects of the genome, transcriptome and proteome (Moitra et al., 2023), including pathological factors, as well as the external effects of lifestyle and environmental factors on cellular function (Sun and Hu, 2016).

Therefore, some of the advantages of proteomic and metabolomic techniques over genomics relate to a greater understanding of which genetic products are being used by a cell or tissue (Figure 1-6), which biological processes are being upregulated or supressed and how the environment that organism is experiencing is affecting the biological processes of the cell or tissue. This is especially true of proteomics, as many biological processes are transmitted through proteins and alterations of the proteome can give detailed insights into disease states (Sun and Hu, 2016).
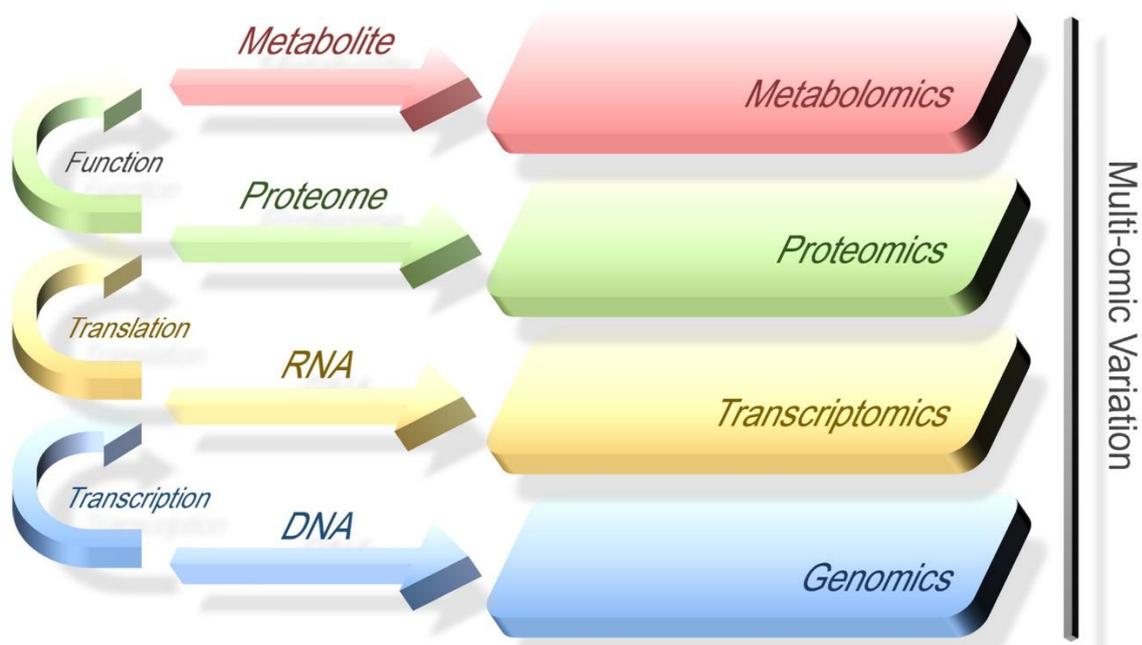
Figure 1-6: A graphic representation of the relationships between the "-omics" fields (Cong and Endo, 2022). Figure available under Creative Commons CC BY 4.0 license.

Whilst the whole human genome has been mapped, and analysis of the whole genome has become accessible as part of clinical medicine with standardisation of analytical techniques and reducing costs, the whole human 'proteome' is as yet unknown, and there is as yet no single proteomics analysis technique which provides sufficient sensitivity and resolution to completely map the entire proteome of a particular sample type (Dupree et al., 2020). Similarly, the human 'metabolome' is a highly dynamic system which is in a state of near-

constant change depending upon innate and environmental factors, in addition to the effect of disease processes if present.

Despite these challenges proteomics and metabolomics both represent an attractive opportunity to identify biomarkers of disease, to aid in rapid and early diagnostics, for identifying mechanisms of disease and targeting therapeutic interventions. In addition, a 'multi-omic' approach allows for a more comprehensive understanding of the biological processes occurring in a cell or tissue, especially in disease states, and how different '-omic layers' interact with one another (Sun and Hu, 2016).

1.5.1 Analytical methods in proteomics and metabolomics

There are many similarities between proteomic and metabolomic analysis workflows, as described in Figure 1-7. Both techniques can be applied to a wide variety of sample types, including those obtained invasively such as tissue, whole cell, blood/serum, and broncheoalveolar lavage fluid, as well as those obtained non-invasively such as urine and respiratory samples such as exhaled breath. Mass spectrometry (MS)-based analysis platforms are used in both proteomic and metabolomic high-throughput workflows (Nilsson et al., 2010, Aslam et al., 2017), with nuclear magnetic resonance analysis platforms also being employed in some metabolomic workflows (Patti et al., 2012, Chen et al., 2022). The following description of proteomic and metabolomic workflows will focus on MS-based platforms as this is most relevant to my research topic and thesis.

1.5.1.1 *Proteomic Analysis Workflows*

Most proteomics research involves an untargeted, or "bottom-up" approach. This means that rather than specific, known proteins being "sought out" in a sample, the entire protein

content of a sample is analysed using a high-throughput technique, such as mass spectrometry, and then individual peptides and proteins identified.

There are several different approaches taken towards proteomics analysis depending upon the research question being posed and the sample type for analysis, however a typical proteomics workflow consists of several major steps. Firstly, the protein content of the studied biological material needs to be isolated, and the total protein content quantified. Different sample types will require different degrees of preparation to be suitable for analysis, but the aim is to ensure all the available proteins are suspended in a stable medium ready for analysis (Norman et al., 2018). The protein content is fractionated, typically using liquid chromatography (LC). After fractionation, the proteins are proteolytically cleaved by enzymes into peptides. Trypsin is the most used proteolytic enzyme for digestion as it cleaves peptides with high specificity at the C terminal end of arginine and lysine residues generating predictable peptide fragments. Peptide fragments can then be further fractionated based on mass, charge, polarity or hydrophobicity (Pappireddi et al., 2019). Peptides are ionized and sent through the mass spectrometer. Ionized peptide signal intensities, in the form of mass-to-charge ratio, are then used to search databases to identify proteins from predicted *in silico* peptide sequences (Dupree et al., 2020).

While signal intensities from MS analysis can be used to identify peptides and thereby protein content of a sample, quantification of peptide content within samples and, more importantly between samples, is unreliable in simple MS. This is due to variable ionisation yields of different peptides, variability in proteolytic digestion, and variable MS detector response (Nilsson et al., 2010). Therefore, the number of ions present in the mass spectrometer is not a direct reflection of the protein content of the original sample (Pappireddi et al., 2019).

Additional techniques have developed to reliably quantify protein content to compare protein abundance between samples. This is particularly useful when comparing samples from different clinical groups or where an intervention has been made as part of a study and there is a need to determine whether there is a treatment effect detectable. There are several approaches to quantitative proteomics, and they can be divided into label-free and labelled methods, and into whether absolute (determining the absolute protein concentration in a sample) or relative (determining the relative ratio of protein concentration between two or more samples) protein quantification is being determined. Absolute protein quantification methods are usually an extension of relative protein abundance and utilise a known standard – a substance which is added to the MS run with a known absolute concentration with which peptide signal intensities can be compared. These standards tend to be costly, therefore limiting their application to smaller sample sets (Pappireddi et al., 2019). Label-free methods use the relative area-under-the-curve of signal intensities of detected peptides to compare protein abundances between MS runs. Whilst relatively cheaper, it is less accurate and reproducible than labelled methods and can have issues in quantifying relative protein abundances when peptides are not detected in every sample or every MS run. Therefore, labelled methods are generally preferred.

Multiplexed proteomics with isobaric tag labelling aims to produce a high throughput, accurate and reproducible relative protein abundances for the full spectra and range of signal intensities of peptides detected in samples. After proteolytic digestion, isobaric tags are added to the sample which covalently bind to peptide fragments. These tags have a known and constant mass and attach to predictable peptide sequences. After the first MS run where peptide signal intensities are noted, the tags are counted after a second MS run and these tag counts are used to determine the relative abundance of proteins between samples and

MS runs (Pappireddi et al., 2019). This isobaric tag-labelling methodology allows highly accurate quantitation of individual protein abundances and allows robust comparison of protein abundances between samples in the same experiment, and potentially allows comparison of protein abundances between studies (Zecha et al., 2019).

Following sample injection into the mass analyser, software is used to clean and normalize the raw data, using quality control (QC) samples from each MS run. Software is also used to identify peptides from the detected mass/charge (m/z) ratios and retention times from a reference library, and then proteins are identified *in silico* from the peptide sequences identified in the mass spectrum (Aslam et al., 2017, Orsburn, 2021). Given the large number of detected proteins typically identified in an untargeted proteomic workflow, there are several statistical techniques that can be employed in identifying significantly altered proteins between groups. These include traditional univariable and multivariable approaches that use mathematical approaches to identify proteins that have significantly different abundances between groups, such as t-tests, ANOVA and regression modelling (Lualdi and Fasano, 2019). These traditional statistical methods do not take into account biological relationships between different detected proteins (Reshetova et al., 2014), and setting appropriate levels for statistical significance are important when performing multiple comparisons, aiming to reduce the risk of false positives balanced against the risk of removing important, biologically relevant discoveries (Lualdi and Fasano, 2019). Unsupervised statistical methods, such as principle component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) aim to overcome some of the issues of more traditional statistical techniques by reducing the dimensionality of the dataset, looking for significant variations in the data and clustering of data points (Lualdi and Fasano, 2019).

Once the significantly altered proteins have been identified, the next step of data analysis aims to identify the functional relationships of these proteins, and the biological processes involved. The most commonly performed method in proteomics is over-representation analysis, the premise of which if that relevant pathways can be detected if the proportion of differentially expressed proteins belonging to a given pathways exceeds the proportion of proteins that would be expected to be seen by chance alone (Garcia-Campos et al., 2015, Norman et al., 2018). This approach can help to overcome the issue of false positive results from the initial statistical analysis, as if a proportion of proteins are biologically related and mapped to the same significantly altered biological pathway, those proteins that are not mapped to a pathway are more likely to be false positives (Lualdi and Fasano, 2019).

1.5.1.2 Metabolomic Analysis Workflows

As discussed above, both MS and nuclear magnetic resonance (NMR) platforms are utilised in metabolomic analysis workflows. The main advantages of MS-based workflows include high resolution analysis with reliable metabolite identification and the ability to perform selective qualitative and quantitative analysis. MS analysis is also relatively quick, with analysis time ranging from 5 to 140 minutes. Despite it being a more costly analysis platform, with expensive instruments and more extensive sample preparation prior to analysis, it remains the most commonly used workflow for metabolome analysis (Chen et al., 2022). Despite there being some advantages to NMR-based workflows, including sample preservation and reproducibility, overall, it has lower sensitivity than mass spectrometry meaning lower concentrations of potentially important compounds can be masked by higher concentration, and therefore larger, spectral peaks.
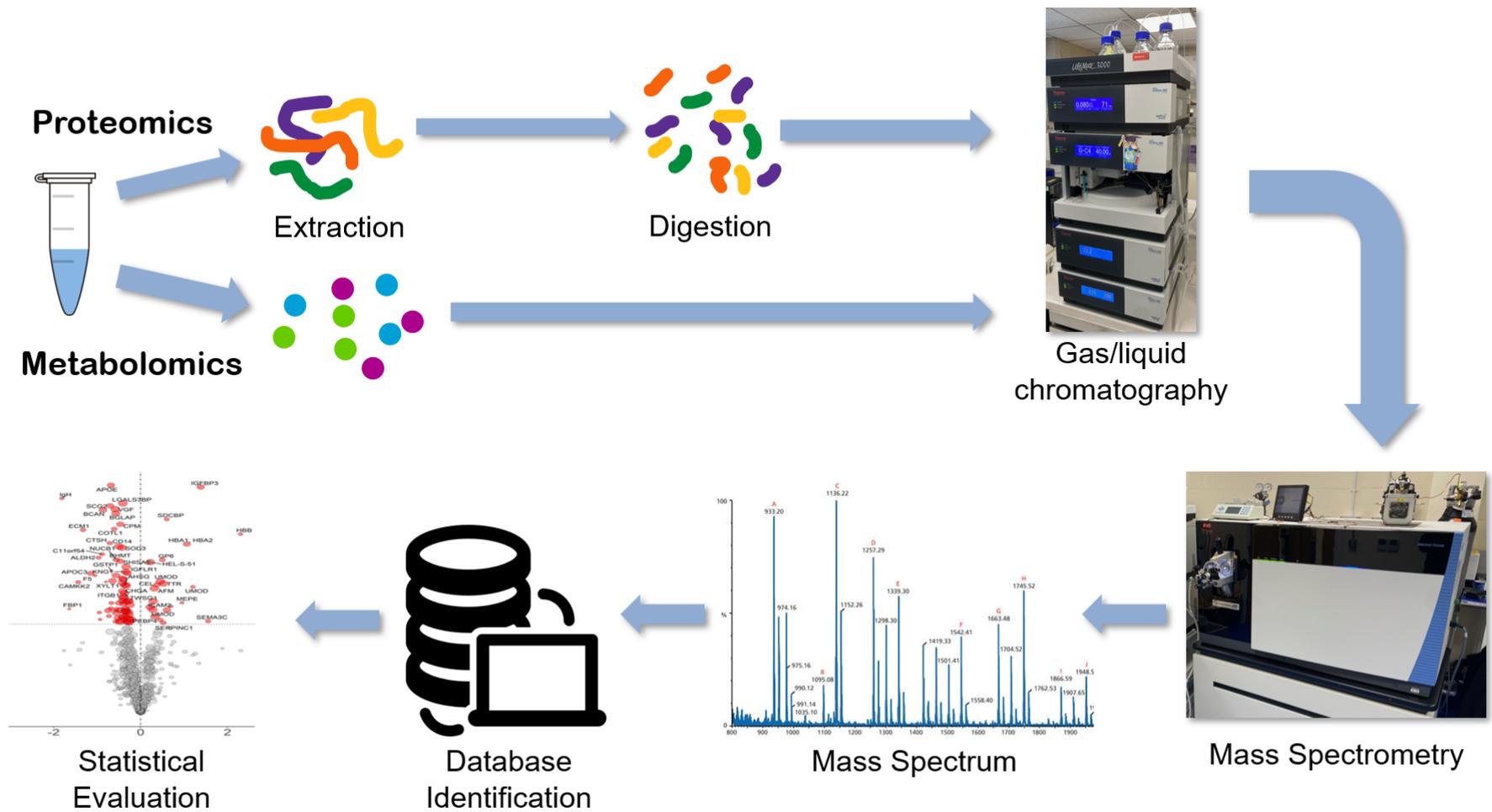
Figure 1-7 Graphical representation of typical mass spectrometry-based proteomic and metabolomic analysis workflows.

The stages to MS-based metabolomic workflows are similar to proteomics workflows in many respects. The first step is to isolate and separate the individual metabolites present within the sample, reducing sample complexity and allowing identification of different metabolites at the same time. Metabolomic workflows use either liquid or gas chromatography (GC) to achieve this, with each separation method having advantages over the other depending on sample type and biological question being investigated (Chen et al., 2022). LC is more suited for detection of moderately and high polarity molecules such as fatty acids, alcohols, phenols, vitamins, organic acids, polyamines, nucleotides, terpenes, flavonoids, and lipids (Theodoridis et al., 2011). Conversely, GC is suited to analysis of volatile metabolites, or compounds that can be derivatized into volatiles, including amino acids, organic acids, fatty acids, sugars, polyols, and amines, often termed volatile organic compounds (Lai and Fiehn, 2018, Theodoridis et al., 2011, Course et al., 2021). LC and GC methods have different resolution and sensitivity for detecting metabolites, and their selection is based upon the physical and chemical qualities of not only the sample type being analysed, but also the physical and chemical properties of the hypothesized target metabolites, as well as whether targeted or untargeted metabolomics analysis is required (Chen et al., 2022).

During the next steps in the analysis workflow, the raw signal output from the MS is processed by specific software to allow quantitative analysis of compounds. Broadly, this step includes noise reduction, retention time correction, peak detection and integration and chromatographic alignment (Smith et al., 2006). Quality control samples are used to correct for variations between plate runs on the mass spectrometer, further correct for spectral noise and determine the variance of metabolite features (Mattoli et al., 2023) and the data is normalised to reduce systematic biases or technical variations and ensure accurate metabolite identification and quantification (Chen et al., 2022). Following these data cleaning

and normalisation steps, metabolite identification is performed, by comparing spectral peak data to authenticated reference data, typically using an in-house library. In the absence of an in-house library, international databases, such as the Human Metabolome Database (HMDB) (Wishart et al., 2018) can be used to identify metabolites from spectral data.

A variety of statistical approaches are employed in the analysis of metabolomics and are similar to the approaches employed in proteomics. Univariable and multivariable comparisons between samples enable identification of metabolites that undergo abnormal changes (Sugimoto et al., 2012), however traditional statistical approaches based solely upon mathematical criteria do not consider biological connections between metabolites and is recognized as a limitation of these approaches (Reshetova et al., 2014). This can be overcome using multiple alternative statistical techniques. Considering this, determining an appropriate p-value threshold for determining significance is important as it can affect the ultimate biological interpretation of the analysis. Once the significantly altered metabolites have been identified, enrichment and pathways analysis software tools are used to identify impacted biological processes (Chen et al., 2022).

1.5.2 Proteomics in respiratory disease

Proteomics has been applied to the study of numerous respiratory diseases, predominantly lung cancer, but also including asthma, COPD and cystic fibrosis (CF), mainly in adult populations. Some studies have focused on paediatric and neonatal lung diseases, including the pathogenesis of BPD. A wide range of sample types have been employed that either assess the lung directly or assess systemic changes. These include invasively obtained samples such as whole tissue, blood/serum and broncheoalveolar lavage, and non-invasively obtained samples such as saliva, breath and urine (Norman et al., 2018).

Proteomic tissue and serum studies of adults with lung cancer have, for example, identified potential biomarkers for predicting premalignant stages (Rice et al., 2015) and tumors at high risk of relapse (Pernemalm et al., 2013). Urine proteomic studies have also been able to discriminate between a wide range of respiratory diseases, including acute viral infection, asthma and bronchiectasis (Martelo-Vidal et al., 2022), highlighting the different biological pathways involved with differing pathogenesis of disease. Whilst applied extensively to adult respiratory disease, proteomic technologies have been applied much less frequently to the study of paediatric respiratory disease, where most studies focus on the study of respiratory tract infection, cystic fibrosis and asthma, mainly focused on biomarker discovery and unravelling disease pathogenesis (Pereira-Fantini and Tingay, 2016).

Proteomic technologies have recently now started to be applied to the neonatal population, and a proteomic analysis of urine from infants born at <28 weeks' gestation, collected in the first three days of life, identified a distinct pattern of protein changes that predicted later BPD (Ahmed et al., 2022). Some of the proteins alterations observed, such as an increase in matrix metalloprotease 9 (MMP9) have been observed in invasively collected airway samples from other studies using more traditional analytical methods (Davies et al., 2010). A small proteomic study of bronchoalveolar lavage fluid from twelve preterm infants found alterations in calcium-signaling proteins in those who later develop BPD (Magagnotti et al., 2013). Urine proteomics in neonates has also shown the ability to discriminate between infectious and non-infectious respiratory pathologies (Starodubtseva et al., 2016).

As the technology for proteome analysis has developed rapidly over the past decade, it has increasingly been applied to a wide range of respiratory diseases and is delivering greater

information on respiratory disease pathogenesis, biomarker identification and risk stratification, although many of these findings come from studies including a small number of participants, use differing proteome analysis protocols, and the majority of findings are yet to come into routine clinical use (Norman et al., 2018).

1.5.3 Metabolomics in respiratory disease

Similarly to proteomics, metabolomics has also been extensively applied to the study of respiratory diseases, again focusing mainly on the adult population (Moitra et al., 2023), where asthma, COPD, lower respiratory tract infection and acute lung injury have been studied most extensively. These studies have demonstrated an ability to discriminate between asthma phenotypes, and evidence of medication altering the EBC metabolome in CF.

A urine metabolomic study of paediatric asthma identified specific pathways (tyrosine and glutathione metabolism) associated with corticosteroid resistance (Park et al., 2017), with a paired urine and serum study finding multiple discriminating metabolic processes (including amino acid metabolism, citrate cycle and pyruvate metabolism) for identifying treatment effect from inhaled therapies during an acute asthma exacerbation (Quan-Jun et al., 2017).

Metabolomic techniques have been more frequently applied to the neonatal population than proteomic studies. In a predominantly term-born cohort, changes to the infant metabolome (analysed from umbilical cord blood) were associated with gestational age, delivery mode, infant sex, and maternal pregnancy complications such as gestational diabetes (Mansell et al., 2022). Within preterm neonates, a small metabolomic study of infants with RDS identified

25 metabolites which appeared over-represented following surfactant administration and mechanical ventilation (Fabiano et al., 2011). A study of amniotic fluid from 32 mothers taken between 21- and 28-week's gestation found a specific metabolomic profile which could predict later preterm birth and later BPD with reasonable accuracy (Baraldi et al., 2016). Overall, urine metabolomic studies have implicated changes in oxidative stress, and cord blood studies have implicated alterations to lipid metabolism in the pathogenesis of BPD (Piersigilli et al., 2019). One MS-based metabolomic study of adolescent survivors of preterm birth with BPD also suggested alterations in lipids related to surfactant production, although could not show a link to current lung function (Carraro et al., 2015).

As with proteomics, metabolomics offers the opportunity to study the pathogenesis of respiratory disease both from invasively and non-invasively obtained samples, assess treatment response, and predict risk of later disease development. Whilst there is an extensive proteomics and metabolomic literature in adult respiratory disease, there has been less focus on the paediatric population, especially those who are survivors of preterm birth.

1.6 Respiratory Health Outcomes in Neonates (RHiNO) study

The Respiratory Health Outcomes in Neonates (RHiNO) Study was funded by the Medical Research Council (MRC) to study a group of preterm-born (≤34 weeks' gestation) school-aged children (aged 7-12 years) with the overall aim to establish the phenotypes and to study underlying mechanisms of prematurity-associated lung disease (PLD), and to establish whether these phenotypes do or do not respond to inhaled therapies. The RHiNO study utilised a cohort developed from a previous cross-sectional population-level questionnaire study called RANOPS (Respiratory and Neurological Outcomes of Children born Preterm study) (Edwards et al., 2015), which had responses from families for approximately 7000

(4200 preterm-born and 2800 term-born) children in Wales born between 2003 and 2011. To supplement this cohort, NHS Wales records for infants born ≤34 weeks' gestation who had been cared for at one of four neonatal units in South Wales between 2004 and 2012 were accessed and families invited to participate.

The inclusion criteria for participation in the RHiNO Study were:

- Child aged 7-12 years old at the time of screening.

- For preterm-born subjects, a gestational of ≤34 weeks' post menstrual age (PMA) at birth; for term-born controls, delivery at a gestational age of ≥37 weeks' PMA.

- Resident within the south Wales area, and suitable for follow-up.

- Fully informed proxy consent from parents/guardians and assent from child where possible.

Exclusion criteria for participation included:

- Respiratory tract infection within the last three weeks (but family/child would be asked to consider participation at a later date).

- Congenital abnormalities.

- Severe cardiopulmonary defects, neuromuscular disease or severe neurodevelopmental impairment which would preclude compliance with testing protocol.

The primary objective of the RHiNO study was to assess whether 12-weeks of treatment with either a long-acting beta agonist bronchodilator (salmeterol) and inhaled corticosteroid (fluticasone), or inhaled corticosteroid alone modifies the underlying mechanisms and/or improves lung function when compared to placebo. Secondary objectives included in-depth characterisation of prematurity-associated lung disease phenotypes assessing the role of atopy, airway inflammation, bronchodilator responses, structural abnormalities, and

cardiorespiratory exercise responses. Samples and data used in my thesis were obtained during the course of this study.

1.6.1 Structure of the RHiNO Study

The RHiNO study was conducted in three parts, as described in Figure 1-8, and an overview is provided below. Initially, those families who returned their questionnaire were invited to participate in a home- or clinic-based screening visit conducted by two trained research nurses, termed part one of the study. During this visit, the child's medical history was obtained, questionnaire responses confirmed, and a short physical examination as performed. Anthropometric measures, including weight measurement on bio-impedance floor scales and height measurement on stadiometer, were obtained. A cardiovascular assessment was performed, and urine and saliva samples obtained. Lung function was measured with a calibrated portable spirometer (Microloop, CareFusion®, Wokingham, UK) and raw values were corrected for age, sex and height as per Global Lung Initiative (GLI) reference ranges (Quanjer et al., 2012).

All preterm-born children with a percentage predicted forced expiratory volume in 1 second (%FEV_1) ≤85% identified from part one of the study were offered the opportunity to participate in part two of the study, along with a select number of preterm-born children with a %FEV_1 >85% and term-born children with a %FEV_1 >90% who served as control subjects. Part two of the study comprised of two visits to the Children's Research Facility at the Children's Hospital for Wales, Cardiff, conducted by a trained paediatrician and research nurse.
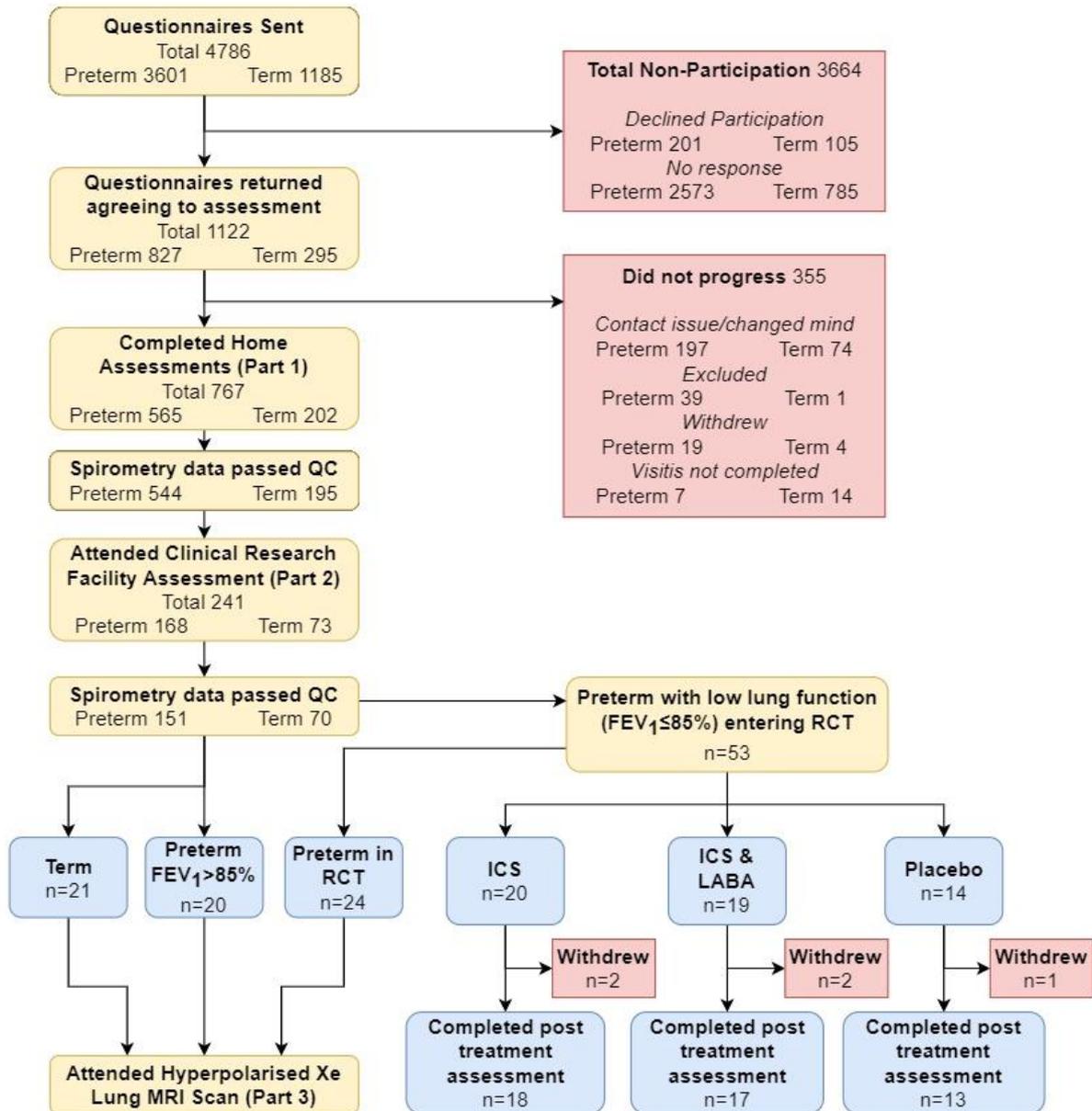
Figure 1-8: Flow diagram showing recruitment to Respiratory Health Outcomes in Neonates (RHiNO) Study.

QC: Quality control. FEV_1 percent predicted forced expiratory volume in 1 second. RCT: Randomised controlled trial. ICS: Inhaled corticosteroid. LABA: Long-acting beta agonist. MRI: Magnetic resonance imaging

During the first visit, consent and assent was reconfirmed, as well as reconfirming questionnaire responses. A physical examination and updated anthropometry measures were collected using the same techniques as the part one visit. Saliva and urine samples were also repeated, as well as a repeat cardiovascular examination and FE_{NO} measurement. Lung function was assessed with laboratory-based spirometers, including whole body plethysmography and inert gas washout techniques to give measures of lung volumes, all performed to ATS/ERS standards and corrected using GLI reference equations as before. Participants underwent skin-prick testing (to assess for atopy) and were subjected to an exercise test using a cycle ergometer to assess maximal exercise capacity. Additional biological samples were obtained including exhaled breath condensate (EBC) and induced sputum collection.

Those preterm-born infants who were confirmed to have $\%FEV_1 \leq 85\%$ after lab-based spirometry were offered the opportunity to enrol into a twelve-week randomised controlled trial (RCT) of inhaled corticosteroids (ICS) (fluticasone), ICS and long-acting β_2 agonists (LABA) in combination(fluticasone/salmeterol) or placebo. Children were monitored during the RCT with daily peak expiratory flow measurements and telephone contact to monitor adherence and any adverse events. Following the twelve-week treatment protocol, children returned for their second visit to the children's research facility, where all the visit one evaluations were repeated, except for skin-prick testing. To avoid ethical conflict, children receiving ICS prior to starting the trial were not randomised to the placebo arm.

In part three of the study, a subset of the children who participated in the RCT, as well as preterm-born and term-born control children, were invited to attend the University of Sheffield to undergo a hyperpolarised xenon magnetic resonance imaging (MRI) scan of their

chests to assess lung ventilation patterns. Findings from the RHiNO study have been extensively published, are refenced throughout this thesis where appropriate and give further methodological details.

1.6.2 Lung function phenotypes from the RHiNO cohort

As I have described above, the children who participated in the RHiNO study underwent extensive lung function testing, and this has allowed the identification of distinct respiratory phenotypes during childhood in the preterm-born population. I will reference these phenotypes throughout this thesis during the following results and discussion chapters, and I give an overview of how they have been defined below.

1.6.2.1 Bronchopulmonary Dysplasia

This classification distinguishes those preterm-born children who received a diagnosis in the neonatal period of BPD from those who did not, and represents one of the most commonly studied phenotypes in the long-term follow-up of lung function in preterm-born individuals (Simpson et al., 2018, Galderisi et al., 2019, Kotecha et al., 2022b, Um-Bergström et al., 2022). In the RHiNO study, BPD was defined according to NICHD 2001 criteria (Ehrenkranz et al., 2005) of oxygen requirement at 28 days of age, with an assessment of severity at 36 weeks of corrected gestation for those born at <32 weeks of gestation, and at 28 and 56 days of age for those born at ≥32 weeks gestation, as described above. Parent-reported history of BPD was supplemented and corroborated with medical notes.

1.6.2.2   Low lung function: Percent predicted FEV₁ ≤85%

Those preterm-born children entering the RCT phase of the RHiNO trial were selected based on a %FEV$_1$ ≤85% and are termed the PT$_{low}$ group. Using a cut off of 85% effectively discriminated between those preterm-born children with and without lung function impairment, as those preterm-born children with %FEV$_1$ >85% (PT$_c$) had comparable spirometry, bronchodilator response and FE$_{NO}$ levels to term-born children (Hart et al., 2022). Therefore, this cut-off proved appropriate in identifying which children may benefit from treatment and therefore should enter the RCT stage of the RHiNO study.

1.6.2.3   Global Lung Initiative: Lung Function below Lower Limit of Normal (LLN)

The Global Lung Initiative (GLI) reference equations are used to correct raw spirometry values for age, sex, ethnicity, and height (Quanjer et al., 2012). The GLI references classify abnormal spirometry as those which fall below the lower limit of normal (LLN), defined as the fifth percentile. Therefore, another classification of low lung function in the preterm-born children in RHiNO is an %FEV$_1$<LLN. This group has been further divided by using the FEV$_1$/forced vital capacity (FVC) ratio. Those preterm-born children with an FEV$_1$<LLN and an FEV$_1$/FVC<LLN have an obstructive spirometry pattern and are termed prematurity-associated obstructive lung disease (POLD). Those with an FEV$_1$<LLN and an FEV$_1$/FVC≥LLN are termed prematurity-associated preserved ratio impaired spirometry (pPRISm). Whilst obstructive lung function patterns have previously been described in preterm-born children (Doyle et al., 2019b, Gibbons et al., 2023), PRISm is a relatively new concept in adult pulmonology (Marott et al., 2021, Wan et al., 2021), and the RHiNO study was the first to describe this pattern in preterm-born children (Cousins et al., 2023).

1.6.3 Findings from the RHiNO Study

The RHiNO study enrolled one of the largest cohorts of preterm-born children who have experienced modern standards of neonatal care, with high maternal antenatal corticosteroid exposure and routine use of exogenous pulmonary surfactant administration. Owing to the large amount of data generated during the course of the study, there have thus far been multiple publications regarding the clinical characteristics and lung function phenotypes. These will be referenced throughout this thesis in context with my results, but I have also briefly summarised them below to provide some background to my research questions and aims.

1.6.3.1 The role of early life factors in predicting later lung function in the RHiNO cohort

As I have mentioned previously in this chapter (sections 1.4.3 and 1.4.4), much of the existing literature regarding later respiratory outcomes in preterm-born individuals focuses on those with a neonatal diagnosis of BPD. Analysis of the portable spirometry data from part one of the RHiNO study, using a mediation model, revealed that whilst gestational age may significantly predict the development of BPD, low lung function (PT_{low}) in later childhood was predicted by increasing gestational immaturity at birth and intra-uterine growth restriction (IUGR) but not BPD (Figure 1-9) (Hart et al., 2022).

1.6.3.2 Prematurity-associated lung dysfunction phenotypes in childhood

Further analysis of the spirometry data available from part one of the RHiNO study allowed the identification of specific lung function patterns and phenotypes using the GLI LLN cut-offs (Figure 1-10), as I have described above. This identified that 22.6% of preterm-born children had some form of lung dysfunction on spirometry, including both fixed- and reversible-POLD,
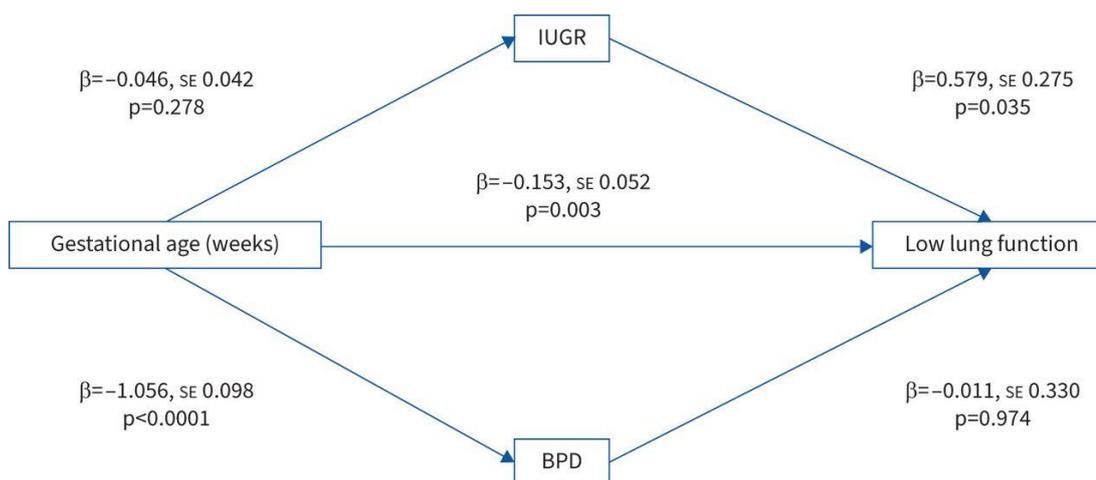
Figure 1-9: Mediation model demonstrating associations between early life factors and low lung function in childhood.

(Hart et al., 2022). Figure available under Creative Commons BY 4.0 license

pPRISm and prematurity-associated dysanapsis (pDysanapsis, $FEV_1 \geq$ LLN, $FEV_1/FVC <$ LLN). These spirometry phenotypes had differential associations between early and current life factors.

POLD, subdivided into reversible and fixed by the spirometry response to inhaled bronchodilators, was associated with a neonatal history of BPD, with the POLD-reversible group also have a significant association with raised FE_{NO} and IUGR. A minority of the pDysanapsis group also showed a bronchodilator response and increased FE_{NO}, and pDysanapsis was associated with postnatal weight gain. pPRISm showed no association with FE_{NO}, despite a subgroup of 13% responding to bronchodilators, and a near significant negative association with body mass index ($\beta=-0.23$, p=0.064) (Cousins et al., 2023). Laboratory-based spirometry performed in part two of the RHiNO study demonstrated that those with a pPRISm phenotype had increased functional residual capacity (FRC) and residual volume (RV) (Cousins et al., 2022).

1.6.3.3   Hyperpolarised ^{129}Xenon Lung MRI

Hyperpolarised ^{129}Xe lung MRI revealed significant ventilation defects and ventilation heterogeneity in approximately half of those preterm-born children with a POLD phenotype, as described above. Whether this represents further phenotypic differences within the POLD group, where for some children there is a persistent ongoing disease process, such as inflammation, occurring requires further investigation. In addition, in those children with a history of BPD, alveolar size appeared increased. Importantly, those preterm-born children with a preserved FEV_1 showed comparable MRI imaging to Term-born controls. Children with a pPRISm phenotype also showed comparable ventilation patterns to preterm- and term-born controls, however only four children underwent MRI with this phenotype, and therefore this result should be interpreted with caution (Chan et al., 2023).

1.6.3.4   RCT of Inhaler Therapies for Prematurity-associated lung disease

Of the 53 preterm-born children with an $FEV_1 \leq 85\%$ enrolled to the RCT, 48 completed the 12-week trial of inhaler therapies. Compared to placebo treatment, those children treated with combination inhaler therapy of ICS and LABA showed an increase in FEV_1 of 14.1% (95% CI 7.3 to 21.0, p=0.002), whilst those treated with ICS alone showed a smaller increase of 7.7% (-0.3 to 15.7, p=0.16) which did not reach statistical significance. Treatment with ICS, either as monotherapy or in combination with LABA, also significantly decreased FE_{NO} and improved postexercise bronchodilator response.

1.7   Conclusions

In this chapter, I have detailed the significant impact that preterm birth has on later health outcomes, particularly related to respiratory health and lung function. Whilst much of the existing literature has focussed on the longer-term outcomes of those with a neonatal

diagnosis of BPD, more recent work has highlighted that a significant proportion of preterm-born subject without BPD are also at risk of lung function deficits. In addition, there are multiple phenotypes of prematurity-associated lung disease with differential associations with early- and current-life factors and clinical parameters, highlighting the complexity of this spectrum of pathologies, increasingly termed prematurity-associated lung disease (PLD) (Simpson et al., 2023). There is a subset of these children who may potentially benefit from inhaled therapies. Childhood represents an opportunity to intervene for preterm-born children with reduced lung function, aiming to maximise the lung function 'peak' in early adulthood, and reduce the risk of COPD.

Advances in mass spectrometry technologies and the '-omics' sciences, namely proteomics and metabolomics, has allowed a greater depth of understanding of the pathogenesis of a range of respiratory diseases in both adults and children. Whilst these techniques have been used to study into respiratory pathology in the neonatal period, there is a paucity of studies applying these methods to preterm-born subjects in later life. What is currently lacking from the literature is an understanding of the biological mechanisms and pathways that underlie these lung function phenotypes. Not only would this help to inform the pathogenesis of these phenotypes but may help to identify those that may respond to existing available treatments, or aid in the development of new treatments.
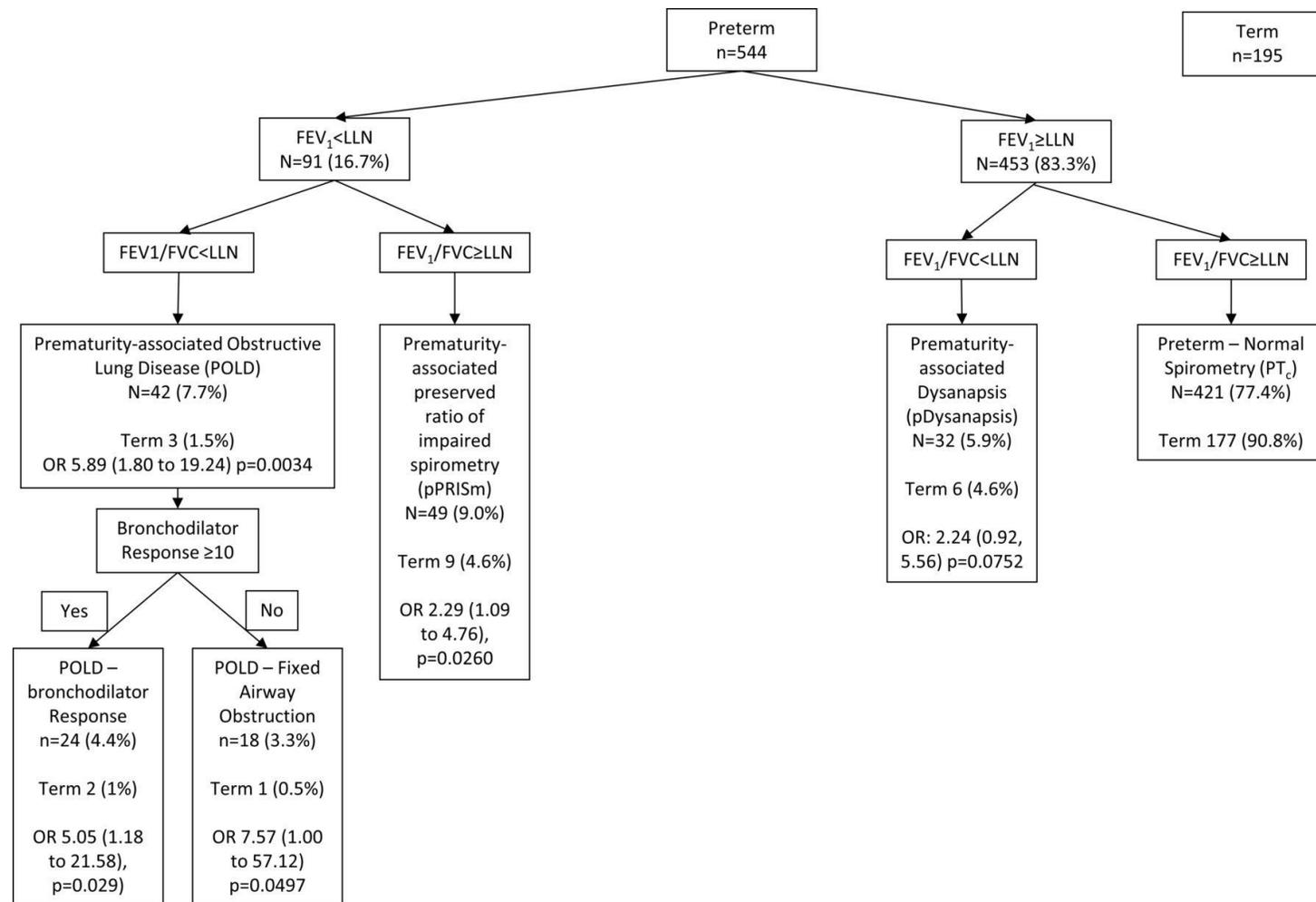
Figure 1-10: Classification of preterm and term groups based on their spirometry measures and bronchodilators responses.

FEV_1, forced expiratory volume in 1s; FVC, forced vital capacity; LLN, lower limit of normal.(Cousins et al., 2023). Figure available under Creative Commons BY 4.0 license

1.8 Research Questions and Aims

As I have highlighted above, there is a need to understand the biological mechanisms underlying the prematurity-associated lung disease phenotypes that have recently been identified through the RHiNO study. Therefore, using the biological samples collected from preterm- and term-born subjects during parts one and two of the RHiNO study, namely the urine and exhaled breath condensate, which were collected on a large majority of children who attended the respective parts, I aim to address the following hypotheses and aims:

1.8.1 Hypotheses:

1. Preterm-born children with a neonatal history of BPD will have a specific pattern of changes in the EBC and urine proteome and metabolome compared to preterm-born and term-born controls.

2. Preterm-born children with PLD will have a specific pattern of changes in the EBC and urine proteome and metabolome compared to preterm-born and term-born controls.

3. For those children with PLD treated with inhaled therapies, there will be corresponding changes in the EBC and urine proteome and metabolome.

1.8.2 Specific Aims:

EBC and urine proteomics:

- Is there evidence of an altered proteome in preterm-born children who develop the neonatal lung disease bronchopulmonary dysplasia (BPD) and/or have evidence of prematurity-associated lung disease and those who do not, and when compared to term-born children.

- Can any profile of proteomic differences detected at baseline be modified by treatment with inhaled corticosteroids and/or a combination of inhaled corticosteroids and long-acting beta-agonist?

EBC and urine metabolomics:

- Is there evidence of an altered metabolome in preterm-born children who develop the neonatal lung disease bronchopulmonary dysplasia (BPD) and/or have evidence of prematurity-associated lung disease and those who do not, and when compared to term-born children.

- Can any profile of metabolomic differences detected at baseline be modified by treatment with inhaled corticosteroids and/or a combination of inhaled corticosteroids and long-acting beta-agonist?

2 Proteomic Analysis of Exhaled Breath Condensate

2.1 Introduction

As discussed in my Introduction (section 1.4), preterm-born children are at greater risk of respiratory morbidity (Saigal and Doyle, 2008, Kotecha et al., 2022b, Simpson et al., 2023), however the mechanisms underlying this longer-term respiratory pathology in those with BPD and/or PLD remain incompletely understood. Whilst many of these children are diagnosed as having asthma, it is increasingly becoming apparent that there are more complex respiratory phenotypes resulting after preterm birth (Hart et al., 2022, Goulden et al., 2021, Cousins et al., 2023); however, the underlying pathophysiology is poorly characterised (Course et al., 2019). It is important to understand the underlying mechanisms of respiratory pathology for these children to identify the underlying endotypes so appropriate therapeutic interventions can be developed. In this chapter, I began my exploration of the underlying mechanisms of PLD by examining the proteome of exhaled breath condensate (EBC) from participants in the RHiNO study.

2.1.1 Exhaled Breath Condensate

EBC provides a useful sample to study in children due to its ease in collection. EBC is composed of droplets of the epithelial lining fluid (ELF), evolved from the airway during turbulent airflow from all compartments of the lung as part of normal tidal breathing, and held in a matrix of condensed moisture from the breath. It is a complex mixture of DNA, RNA, proteins, metabolites and volatile organic compounds which reflect the biological processes occurring in the lung tissue (Davis et al., 2012). EBC is of interest for identifying and understanding respiratory pathology due to its simple, non-invasive and easily-repeatable method of collection (Horvath et al., 2005), and has been used to study mechanisms in lung

cancer (Campanella et al., 2019), COPD (Borrill et al., 2008), asthma (Cavaleiro Rufo et al., 2019) and bronchiolitis (Demirkan et al., 2018).

Identifying the proteome in EBC has been challenging, but recent developments aid identification and accurate quantification of a large array of proteins. Proteomics refers to the study of the entire protein complement of a biological sample, which represents a down-stream product of the genotype and reflects an organism's phenotype. Proteomic methods simultaneously analyse the entire protein content of a sample and have gained interest clinically as a potential tool for unravelling pathogenesis of various diseases and potentially to identify specific biomarker (Monti et al., 2019, Bloemen et al., 2011, Lopez-Sanchez et al., 2017).

2.1.2   Aims

In this analysis, I aimed to identify differences in the EBC proteome of preterm-born school-aged children, with and without BPD diagnosed in infancy, when compared with term-born controls. In addition, I aimed to identify whether there was modulation of the EBC proteome with inhaled therapies from preterm-born children with low lung function, who participated in the randomised control trial (RCT) stage of the RHiNO study (Goulden et al., 2021).

2.2 Methods

2.2.1   Study Participants

This analysis was conducted on a cohort of children recruited to the Respiratory Health Outcomes in Neonates study (RHiNO, EudraCT: 2015-003712-20) which has been described previously in this thesis (section 1.6) and publications (Hart et al., 2022, Goulden et al., 2021). Briefly, children from the previous Respiratory and Neurological Outcomes in children born

Preterm (RANOPs) questionnaire study (Edwards et al., 2015, Edwards et al., 2016) were supplemented with additional preterm-born children, sourced via the NHS Wales Informatics Service (NWIS) and sent a respiratory and neurology questionnaire if they were born ≤34 or ≥37 week's gestation and were aged 7-12 years. BPD was defined as oxygen-dependency of 28 days or greater for those born <32 weeks' gestation and at 56 days of age for those born ≥32 weeks' gestation (Ehrenkranz et al., 2005). Neonatal history was corroborated with medical records. Intrauterine growth restriction (IUGR) was defined as birthweight <10th percentile adjusted for sex and gestation (LMS Growth version 2.77, Medical Research Council, UK). Children with congenital malformations, significant cardiopulmonary disorder or neuromuscular disease were excluded. Ethical approval was obtained from the South-West Bristol Research Ethics Committee (15/SW/0289). Parents gave informed written consent and children provided assent.

2.2.2   Lung function assessment

Responders underwent a home or hospital assessment by two research nurses. A subset attended the Children's Hospital for Wales, Cardiff, UK for comprehensive respiratory testing conducted by a trained research nurse and paediatrician. Spirometry (MasterScreen Body and PFT systems, Vyaire Medical, Germany) was performed as per European Respiratory Society/American Thoracic Society guidelines (Miller et al., 2005) and normalised using Global Lung Initiative (GLI) reference equations. Those preterm-born children with low lung function (PT_{low}) defined as percent predicted forced expiratory volume in 1 second (%FEV_1) of ≤85% were enrolled into the RCT (Goulden et al., 2021). Term-born children who had %FEV_1 >90% were included as term controls.

PT_{low} participants were offered the opportunity to participate in a twelve-week randomised controlled trial (RCT), which has been described in detail previously in this thesis (section

1.6.3.4), and the results of which have been published (Goulden et al., 2021). In brief, children received either monotherapy with inhaled corticosteroids (ICS) (50μg fluticasone propionate, two actuations taken twice a day), combination therapy of ICS and long-acting beta-agonists (LABA) (50μg fluticasone propionate and 25μg salmeterol xinafoate, two actuations taken twice a day) or placebo for twelve weeks. All trial medications were administered via a spacer device (Volumatic®, Allen & Hanburys, UK). Active trial medications were produced by St Mary's Pharmaceutical Unit, Cardiff, UK. Placebo inhalers were sourced from GlaxoSmithKline (GSK). All inhaler canisters were placed in identical plain-coloured actuators sourced from GSK to ensure adequate blinding to trial group. Participants who had previously been receiving inhaled steroid treatment prior to enrolment were weaned off steroid inhalers over a four-week washout period prior to starting the trial. All trial participants received training on good inhaler technique by the research team. To avoid ethical conflict, children receiving ICS prior to starting the trial were not randomised to the placebo arm, as described in section 1.6.1. Following treatment, RCT participants underwent repeat EBC sampling.

2.2.3 EBC sampling

EBC was collected using a cooling tube (RTube®, Respiratory Research Inc. Texas, USA) over a period of 10 minutes of passive tidal breathing whilst the participant wore a nose clip, stopping briefly to swallow saliva if needed. The RTube® is a single-patient, single-use design, preventing cross contamination, and features a large 'Tee' section to separate saliva from exhaled breath, thereby ensuring collection of airway lining fluid and not secretions from the oropharynx. Environmental temperature and humidity remained stable during sampling. Once collected, samples were immediately separated into aliquots and stored at -80°C pending analysis. Samples were taken during the baseline assessment, and for those participating in the RCT, EBC sampling was repeated 12-weeks later after completing the

treatment protocol. EBC samples were collected by a trained paediatrician (Dr Michael Cousins) and trained research nurse (Dr Kylie Hart) during the RHiNO trial.

2.2.4 EBC Analysis:

2.2.4.1 TMT Labelling

EBC samples were analysed at the University of Bristol Proteomics Facility. An equal volume of each sample (ensuring that no sample contained more than 50µg of protein) was digested with trypsin (1.25µg trypsin; 37°C, overnight), labelled with Tandem Mass Tag (TMT) eleven plex reagents according to the manufacturer's protocol (Thermo Fisher Scientific, Loughborough, UK) and the labelled samples pooled. The pooled sample was desalted using a SepPak cartridge according to the manufacturer's instructions (Waters, Milford, Massachusetts, USA). Eluate from the SepPak cartridge was evaporated to dryness and resuspended in 1% formic acid prior to analysis by nano-LC MSMS using an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific).

2.2.4.2 Nano-LC Mass Spectrometry

The TMT-labelled pool was fractionated using an Ultimate 3000 nano-LC system in line with an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) (Figure 2-1). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid peptides were resolved on a 250 mm × 75 µm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1min., 6-15% B over 58min., 15-32% B over 58min., 32-40% B over 5min., 40-90% B over 1min., held at 90% B for 6min and then reduced to 1% B over 1min.) with a flow rate of 300 nl min−1. Solvent A was 0.1% formic acid, and Solvent B was

aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.0kV using a stainless-steel emitter with an internal diameter of 30 μm (Thermo Scientific) and a capillary temperature of 300°C.

All spectra were acquired using an Orbitrap Fusion Lumos mass spectrometer controlled by Xcalibur 3.0 software (Thermo Scientific) and operated in data-dependent acquisition mode using an SPS-MS3 workflow (Figure 2-2). FTMS1 spectra were collected at a resolution of 120,000 with an automatic gain control (AGC) target of 200,000 and a max injection time of 50ms. Precursors were filtered with an intensity threshold of 5,000, according to charge state (to include charge states 2-7) and with monoisotopic peak determination set to Peptide. Previously interrogated precursors were excluded using a dynamic window (60s +/-10ppm). The MS2 precursors were isolated with a quadrupole isolation window of 0.7m/z. ITMS2 spectra were collected with an AGC target of 10,000, max injection time of 70ms and CID collision energy of 35%.

For FTMS3 analysis, the Orbitrap was operated at 50,000 resolution with an AGC target of 50,000 and a max injection time of 105ms. Precursors were fragmented by high energy collision dissociation (HCD) at a normalised collision energy of 60% to ensure maximal TMT reporter ion yield. Synchronous Precursor Selection (SPS) was enabled to include up to 5 MS2 fragment ions in the FTMS3 scan.

Figure 2-1: Photograph of High-Performance Liquid Chromatography (HPLC) Fractionation Column (Ultimate 3000 nano-LC system) used for the proteomic analyses in this thesis.
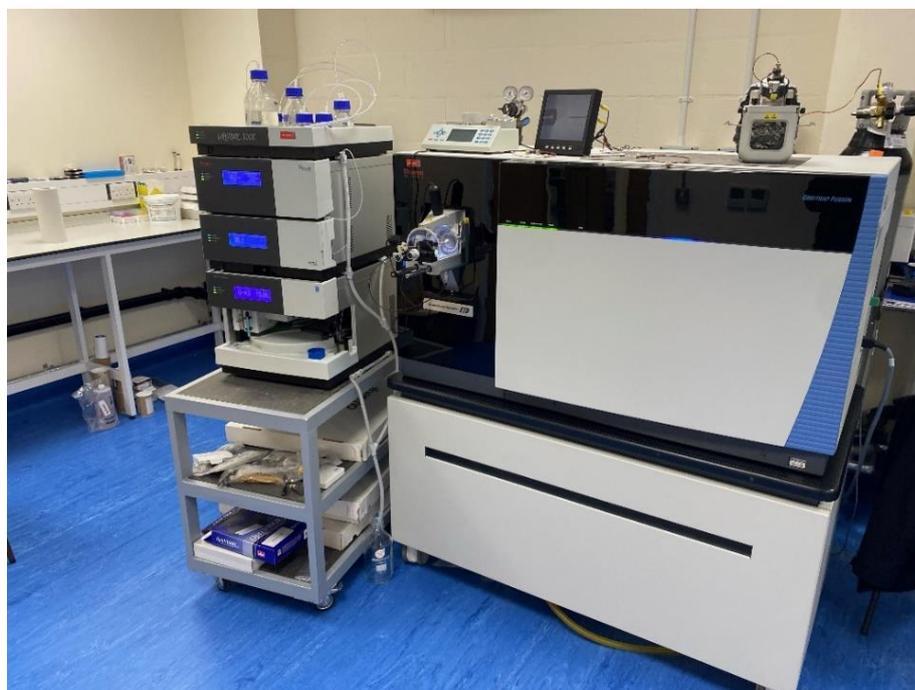
Figure 2-2 Photograph of Nano-LC system (left) and Orbitrap Fusion Mass Spectrometer (right) used for the proteomic analyses in this thesis.

2.2.4.3 Raw Proteomic Data Processing

The raw data files were processed and quantified using Proteome Discoverer software v2.1 (Thermo Scientific) and searched against the UniProt Human database (downloaded October 2019: 150786 entries) using the SEQUEST HT algorithm by the University of Bristol Proteomics Facility. Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.6Da. Search criteria included oxidation of methionine (+15.995Da), acetylation of the protein N-terminus (+42.011Da) and Methionine loss plus acetylation of the protein N-terminus (-89.03Da) as variable modifications and carbamidomethylation of cysteine (+57.021Da) and the addition of the TMT mass tag (+229.163Da) to peptide N-termini and lysine as fixed modifications. Searches were performed with full tryptic digestion and a maximum of 2 missed cleavages were allowed. The reverse database search option was enabled, and all data was filtered to satisfy false discovery rate (FDR) of 5%.

2.2.5 Statistical Analysis

Baseline population and RCT group characteristics were compared using Chi-squared, t-test or one-way ANOVA with Bonferroni correction as appropriate. Replicate numbers (number of samples in which a particular protein was detected) were calculated. Relative protein abundances, determined from the quantity of TMT-tag counts at each detected peptides spectral peak, were $\log_2$-transformed and fold changes ($\log_2$FC) between groups were compared, and the data inspected for normality. Welch's t-test/ANOVA with post-hoc correction was used for baseline samples as appropriate, and paired samples t-test for pre/post-RCT samples. $p < 0.05$ was considered statistically significant. I performed all statistical analyses using R v4.0.4 (R Core Team, 2021) utilising the R packages *"stringr", "dplyr", "ggplot2", "ggpubr"* and *"ellipsis"*. Dr P Lewis (Bioinformatician at University of Bristol Proteomics Facility) supported my learning of the R coding language and of commonly used statistical approaches to analysing TMT-based proteomics datasets. Gene name is used

synonymously with protein name. Gene names were unavailable for four proteins. I used WebGestalt to perform functional enrichment analysis (Liao et al., 2019), which maps functional and biological processes to over-represented proteins/genes. Ingenuity Pathways Analysis (IPA, Qiagen®, Germany) identified relationships between significantly different proteins using network maps, which were reproduced in Cytoscape v3.9 (Shannon et al., 2003) for this thesis. Linear models were created to identify relationships between proteins of interest and participant characteristics.

2.3 Results

2.3.1 Participants

From a total of 1,426 returned questionnaires, 768 children participated in the home screening visit. 241 attended for baseline assessment, and 53 children entered the RCT. EBC was successfully collected and analysed from 218 (91%) children at baseline. 48 of the 53 RCT participants completed treatment and 46/48 (96%) post-treatment EBC samples were successfully collected and analysed. Participant demographics for children at baseline and for those participating in the RCT who provided EBC are given in Table 2-1. As expected, the preterm-born group were delivered at a significantly more immature gestational age (mean 30.9±2.8 weeks vs 40.2±1.1 weeks, p=<0.001) and with a significantly lower birthweight (mean 1,613±587g vs 3,521±518g, p<0.001). At baseline, no significant differences were noted in current characteristics between the preterm-born and term-born children apart from age at testing (mean 11.01±1.24 years vs 10.43±1.09, p=0.001) and asthma diagnosis (34 (23%) vs 5 (7%), p=0.007). Thirty-seven (25%) of the preterm-born children had a neonatal diagnosis of BPD and 53 (36%) were classed PT_{low}, all of whom joined the RCT. There were no differences for asthma diagnosis (10 (27%) vs 24 (21%); p=0.67) or IUGR (8 (22%) vs 19 (17%); p=0.70) between the preterm BPD and No BPD groups, nor for between the PT_{low} and PT_c

groups (asthma: 16 (30%) vs 18 (19%); p=0.19; IUGR: 12 (23%) vs 15 (16%); p=0.35 respectively). Marginally more EBC was collected from term-born children compared to preterm-born (1.13ml vs 1.28ml, p=0.001), but no significant differences were noted between the preterm groups (BPD or PT$_{low}$ vs preterm controls, p=1.0]) or between the three RCT groups. Demographics were similar for the three RCT groups. However, the placebo group produced more EBC after treatment (p=0.02), but not for ICS or ICS/LABA (p>0.1).

2.3.2 Proteins Detected in EBC

A total of 210 different proteins were identified with details given in the Appendix in Table 8-1 together with the number of samples in which the proteins were detected. The distribution of detected proteins across all samples is given in a heatmap in Figure 2-3. Figure 2-4 gives results of functional enrichment analysis, which was possible for 192 of the detected proteins. 28 proteins were identified with a significant difference between one or more of the group comparisons and functional enrichment analysis was possible for 27 of these. Most proteins with significantly differing abundances were functionally related to protein/ion binding and cell structure.

2.3.3 Comparison of EBC proteome in baseline samples

Nineteen proteins were detected in all 218 baseline EBC samples (Table 2-2). Cytokeratins were the most detected protein class. Only increased abundance of two keratins, type II cytoskeletal 5 (KRT5) (0.12, p=0.03) and 6A (KRT6A) (0.14, p=0.02) was observed when the all preterm-born and term-born groups were compared.

Exploratory analyses of proteins not detected in every sample but with a significant abundance difference between the groups are shown in Table 2-3, ordered by decreasing replicate number. Eleven proteins were detected with significant differences between

preterm- and term-born children, nine between BPD and No BPD, and seven between PT_{low}

and PT_c groups. Figure 2-5 shows all significantly different protein abundances between the

BPD and No BPD groups; and PT_{low} and PT_c groups.

Baseline		
Variable	Preterm born ($\leq$34/40) n = 149	Term born ($\geq$37/40) n= 69
Sex (male), n(%)	71 (48)	36 (52)
Ethnicity (white), n(%)	140 (94)	68 (99)
Gestational age (weeks), mean (SD)	30.9 (2.8)	40.2 (1.1)***
Birthweight (g), mean (SD)	1613 (587)	3521 (518)***
Bronchopulmonary dysplasia, n(%)	37 (25)	0 (0)***
Age at testing (years), mean (SD)	11.0 (1.2)	10.4 (1.1)**
Weight (kg), mean (SD)	39.1 (10.7)	37.9 (10.5)
Body Mass Index (kg/m^2), mean (SD)	18.2 (3.5)	18.0 (3.2)
Asthma diagnosis, n(%)	34 (23)	5 (7)**
Low lung function (FEV$_1$$\leq$85%pred), n(%)	53 (36)	0 (0)
Post RCT Samples		
Variable	Preterm born ($\leq$34/40) with low lung function (FEV$_1$ $\leq$85%) n = 46		
	Placebo n = 12	ICS n = 17	ICS/LABA n = 17
Sex (male), n(%)	5 (39)	6 (35)	8 (47)
Ethnicity (white), n(%)	13 (100)	14 (82)	17 (100)
Gestational age (weeks), mean (SD)	29.6 (3.3)	29.4 (3.0)	30.9 (2.8)
Birthweight (g), mean (SD)	1394 (612)	1282 (545)	1470 (570)
Bronchopulmonary dysplasia, n(%)	7 (54)	7 (41)	6 (35)
Age at testing (years), mean (SD)	11.0 (1.2)	10.7 (1.4)	10.7 (1.2)
Weight (kg), mean (SD)	38.9 (11.3)	36.7 (12.4)	36.5 (9.7)
Body Mass Index (kg/m^2), mean (SD)	18.0 (3.6)	17.9 (3.9)	17.4 (2.3)
Asthma diagnosis, n(%)	2 (17)	8 (47)	4 (24)

Table 2-1 EBC Proteomics participant demographics.

Preterm born vs Term born: *p<0.05, **p<0.01 ***p<0.001. Treatment groups: *p<0.05, **p<0.01 ***p<0.001
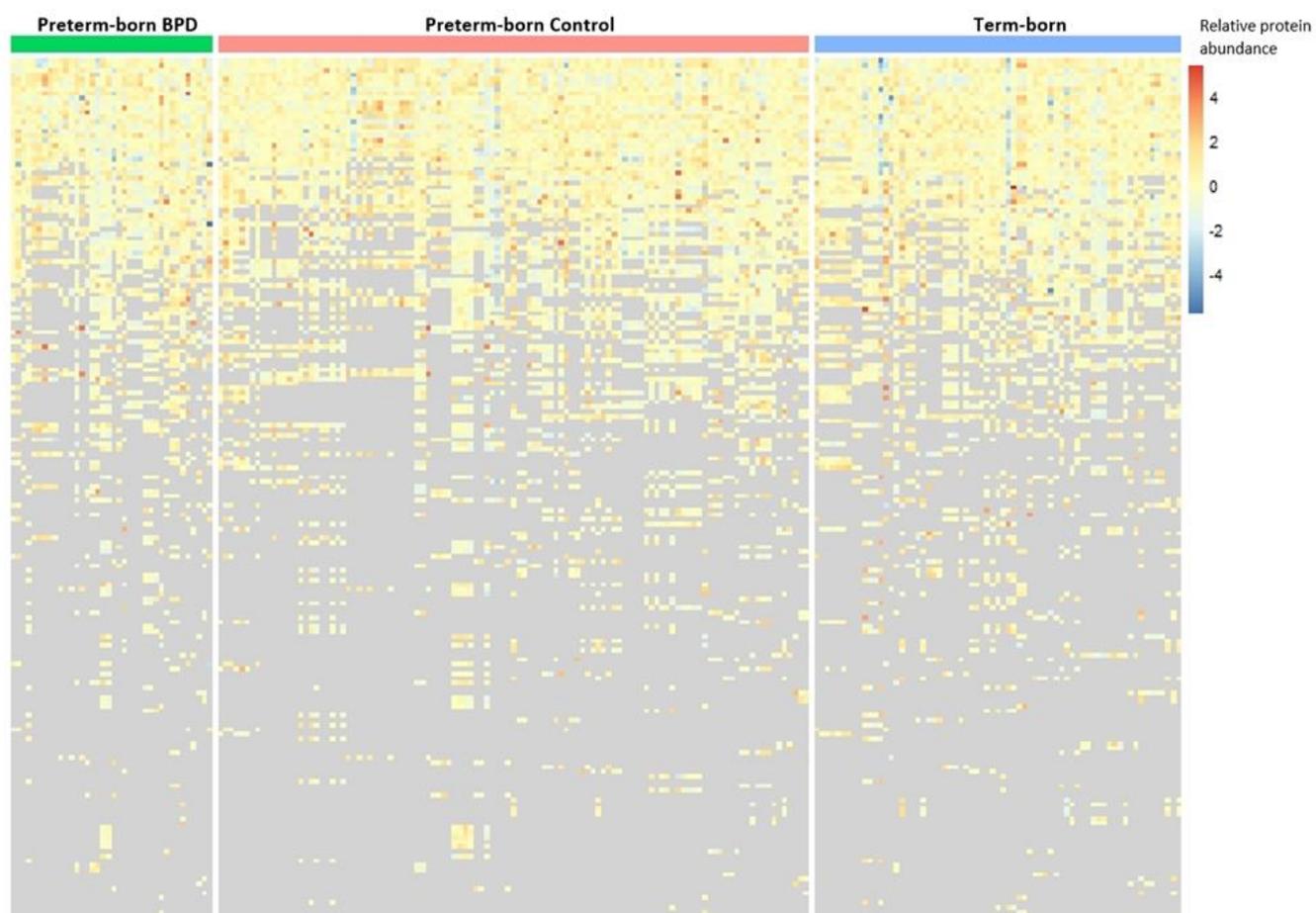
Figure 2-3: Heatmap displaying Protein Content of Baseline EBC Samples.

Individual samples represented in each column. Coloured areas represent relative protein abundance. Grey areas represent proteins not detected within that sample
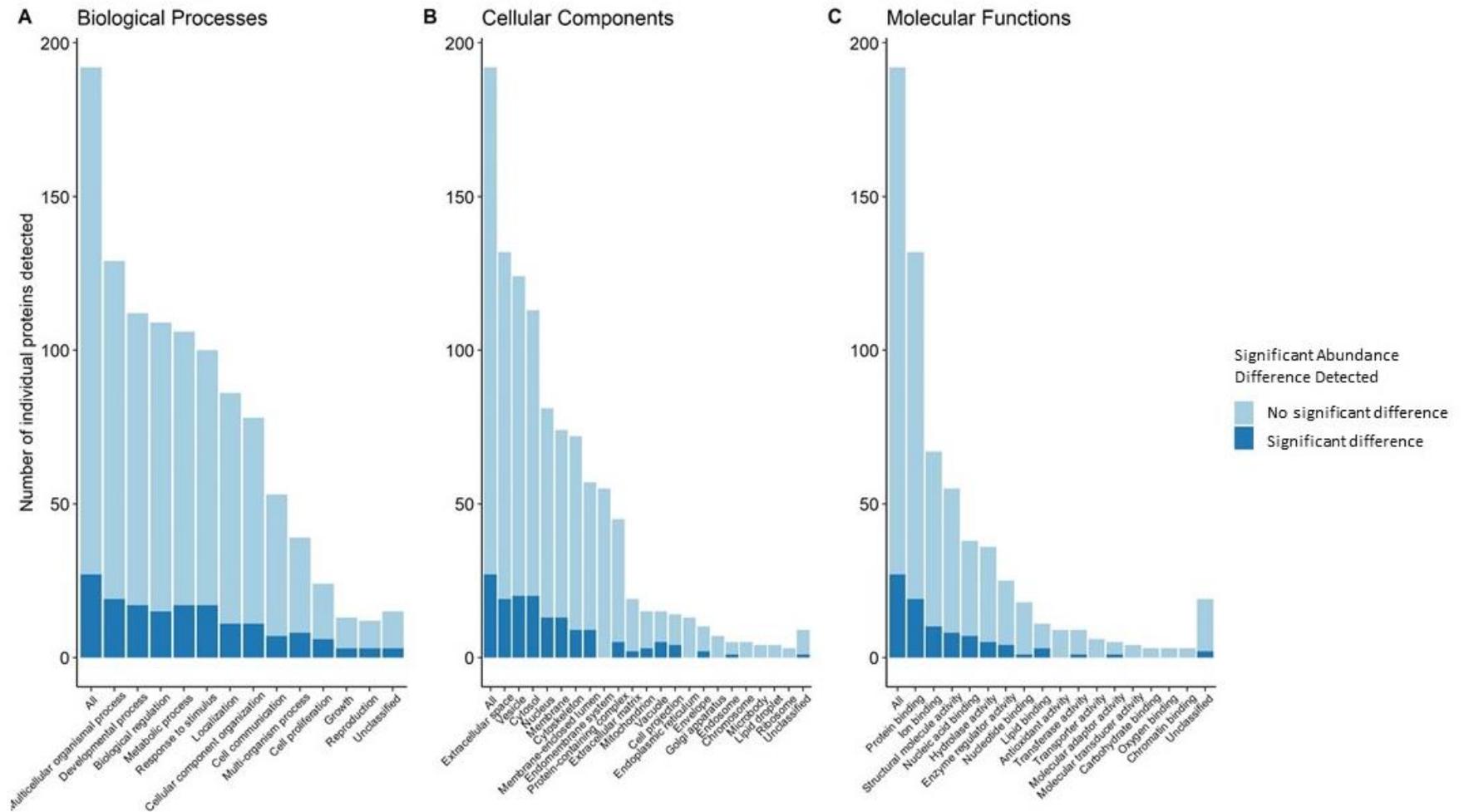
Figure 2-4: Functional enrichment analysis of proteins with and without significant abundance difference between any clinical groups in EBC samples.

2.3.3.1 BPD vs No BPD:

For proteins detected in every sample, significantly decreased abundance of the desmosome proteins desmoglein-1 (DSG1) (Log_2FC -0.26, p=0.02), desmocollin-1 (DSC1) (-0.27, p=0.02) and junctional plakoglobin (JUP) (-0.23, p=0.04), and increased abundance of KRT6A (0.68, p=0.01) was observed when the BPD and No BPD groups were compared (Figure 2-5). No significant differences were noted for DSG1, DSC1 and JUP between the No BPD and Term groups (Figure 2-6).

Protein network maps highlighting significant protein pathways (including proteins detected in all or only some samples) comparing BPD and No BPD groups are shown in Figure 2-7. For proteins detected in a proportion of samples, dermcidin (DCD) was detected in n=146 (98%) samples and was less abundant in the BPD group (Log_2FC -0.43, p=0.03), and was related to DSG1 and DSC1 in the network map. As with KRT6A, KRT6B was detected in 133 (89%) samples being more abundant in the BPD group (0.76, p=0.03) when compared to the No BPD group. Small proline-rich protein 2E (SPRR2E), secretory leukocyte peptidase inhibitor (SLPI) and gamma-glutamyl hydrolase (GGH) were all significantly less abundant in the BPD group (-0.92, p=0.04; -2.08, p=0.04; -0.79, p=0.03 respectively); however, these were detected in <25% of the samples.

Univariable linear regression models using BPD as a binary outcome variable for DSG1, DSC1 and JUP identified that a history of BPD had a significant association with each of these three proteins (β-0.23, p=0.014; β-0.27, p=0.019; β-0.23, p=0.008 respectively; Table 2-4). Using BPD history and current lung function status as interaction terms in the model identified that the reduced abundance of DSG1, DSC1 and JUP was statistically significant/near significant in those with both BPD and PT_{low} (β-0.35, p=0.012; β-0.30, p=0.06, β-0.30, p=0.01 respectively) but not for those with BPD and current normal lung function (Table 2-4).

Accession Number	Gene Name	Protein Name	Protein Function	Preterm vs Term n= 149 vs 69			BPD vs No BPD n = 37 vs 112			PT$_{low}$ vs PT$_c$ n = 53 vs 97		
				Log$_2$FC	Ratio	p	Log$_2$FC	Ratio	p	Log$_2$FC	Ratio	p
Q08554	DSC1	Desmocollin-1	Cell-cell junction	0.02	1.01	0.84	**-0.27**	**0.83**	**0.02***	-0.08	0.95	0.43
Q02413	DSG1	Desmoglein-1	Cell-cell junction	**-0.17**	**0.89**	**0.03***	**-0.26**	**0.84**	**0.02***	-0.16	0.90	0.10
P15924	DSP	Desmoplakin	Cell-cell junction, Cytoskeleton	-0.005	0.997	0.94	-0.08	0.95	0.35	0.11	1.08	0.13
P14923	JUP	Junction plakoglobin	Plasma membrane protein complex	-0.08	0.95	0.30	**-0.23**	**0.85**	**0.04***	-0.10	0.93	0.27
H6VRG2	KRT1	Cytokeratin-1	Cytoskeleton	0.06	1.04	0.45	-0.15	0.90	0.13	-0.04	0.97	0.62
H6VRG3	KRT1	Cytokeratin-1	Cytoskeleton	0.13	1.09	0.24	-0.06	0.96	0.68	-0.01	0.99	0.94
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal	Cytoskeleton	-0.18	0.88	0.06	0.11	1.08	0.45	0.06	1.04	0.64
P13647	KRT5	Keratin, type II cytoskeletal 5	Cytoskeleton	**0.12**	**1.09**	**0.03***	0.13	1.09	0.14	0.05	1.04	0.52
P02538	KRT6A	Keratin, type II cytoskeletal 6A	Cytoskeleton	**0.41**	**1.33**	**0.02***	**0.68**	**1.60**	**0.01***	0.14	1.10	0.55
P35527	KRT9	Keratin, type I cytoskeletal 9	Cytoskeleton	0.19	1.14	0.06	-0.20	0.87	0.08	-0.09	0.94	0.38
P13645	KRT10	Keratin, type I cytoskeletal 10	Structural protein extracellular space	-0.08	0.95	0.39	0.07	1.05	0.49	0.04	1.03	0.64
P02533	KRT14	Keratin, type I cytoskeletal 14	Cytoskeleton	-0.006	0.996	0.91	0.14	1.10	0.14	0.06	1.04	0.45
P08779	KRT16	Keratin, type I cytoskeletal 16	Cytoskeleton	0.31	1.24	0.10	0.40	1.32	0.17	0.28	1.21	0.25
Q04695	KRT17	Keratin, type I cytoskeletal 17	Intermediate filament cytoskeleton	-0.04	0.97	0.79	0.36	1.28	0.09	0.10	1.07	0.58
Q8N1N4	KRT78	Keratin, type II cytoskeletal 78	Cytoskeleton	0.03	1.02	0.73	0.07	1.05	0.44	-0.02	0.99	0.86
P31944	CASP14	Caspase-14	Protease	0.13	1.09	0.4	0.01	1.01	0.95	0.11	1.08	0.62
P01040	CSTA	Cystatin-A	Protease inhibitor	0.13	1.09	0.27	-0.19	0.88	0.22	0.05	1.04	0.72
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a	Structural component of ribosome	0.09	1.06	0.41	-0.31	0.81	0.06	0.30	1.23	0.05
P25311	AZGP1	Zinc-alpha-2-glycoprotein	Major histocompatibility complex protein	-0.02	0.99	0.86	-0.27	0.83	0.11	0.18	1.13	0.27

Table 2-2: Proteins detected in every EBC sample.

BPD: Preterm-born with history of bronchopulmonary dysplasia; PT$_c$: Preterm-born control; PT$_{low}$: Preterm born with low lung function; Log$_2$FC: Log2 fold-change between groups. *Denotes p-vale <0.05

Accession Number	Gene Name	Protein Name	Protein Function	Preterm vs Term n = 149 v 69				BPD vs No BPD n = 37 v 112				PT_low vs PT_c n = 53 v 97			
				n	Log₂FC	Ratio	p	n	Log₂FC	Ratio	p	n	Log₂FC	Ratio	p
P81605	DCD	Dermcidin	Peptidase Antimicrobial activity	146 v 64	-0.03	0.98	0.83	**37 v 109**	**-0.42**	**0.75**	**0.03***	52 v 94	0.23	1.17	0.16
P04259	KRT6B	Keratin, type II cytoskeletal 6B	Cytoskeleton	**133 v 60**	**0.54**	**1.45**	**0.03***	**33 v 100**	**0.76**	**1.69**	**0.03***	46 v 87	0.34	1.27	0.25
Q01469	FABP5	Fatty acid-binding protein 5	Lipid transporter	122 v 68	-0.19	0.88	0.27	27 v 95	0.06	1.04	0.77	**38 v 84**	**-0.39**	**0.76**	**0.04***
P62736	ACTA2	Actin, aortic smooth muscle	Muscle protein	**115 v 50**	**0.54**	**1.45**	**0.01***	31 v 84	-0.43	0.74	0.10	46 v 69	-0.07	0.95	0.77
P05089	ARG1	Arginase-1	Hydrolase Antimicrobial activity	104 v 48	-0.03	0.98	0.86	29 v 75	-0.25	0.84	0.17	**37 v 67**	**-0.39**	**0.76**	**0.03***
Q6UWP8	SBSN	Suprabasin	Unknown	**99 v 54**	**-0.33**	**0.80**	**0.02***	**25 v 74**	**-0.34**	**0.79**	**0.049***	36 v 63	-0.12	0.92	0.43
Q6KB66	KRT80	Keratin, type II cytoskeletal 80	Cytoskeleton	**72 v 50**	**-0.27**	**0.83**	**0.01***	18 v 54	0.09	1.06	0.52	24 v 48	0.02	1.01	0.89
P04083	ANXA1	Annexin A1	Protease inhibitor Anti-inflammatory activity	50 v 31	-0.09	0.94	0.75	14 v 36	0.19	1.14	0.57	**15 v 35**	**-0.77**	**0.59**	**0.02***
P23490	LORICRIN	Loricrin	Cytoskeleton	**42 v 19**	**-0.42**	**0.75**	**0.03***	10 v 32	-0.08	0.95	0.67	14 v 28	-0.23	0.85	0.20
P29508	SERPINB3	Serpin B3	Protease inhibitor	37 v 21	0.59	1.51	0.19	12 v 25	0.23	1.17	0.62	**12 v 25**	**-0.86**	**0.55**	**0.01***
P22531	SPRR2E	Small proline-rich protein 2E	Peptide cross-linking	36 v 7	0.07	1.05	0.88	**12 v 24**	**-0.92**	**0.53**	**0.04***	16 v 20	-0.40	0.76	0.33
P06733	ENO1	Alpha-enolase	Transcription regulation	**29 v 10**	**-0.64**	**0.64**	**0.01***	6 v 23	0.52	1.43	0.19	7 v 22	0.36	1.28	0.36
Q9HCM4	EPB41L5	Band 4.1-like protein 5	Cytoskeleton	22 v 11	-0.38	0.77	0.35	**4 v 18**	**0.68**	**1.60**	**0.04***	7 v 15	0.61	1.53	0.21
Q9C075	KRT23	Keratin, type I cytoskeletal 23	Cytoskeleton	**22 v 8**	**-1.19**	**0.44**	**0.01***	6 v 16	-0.05	0.97	0.91	7 v 15	0.51	1.42	0.35
P22735	TGM1	Protein-glutamine gamma-glutamyltransferase K	Acyltransferase Peptide cross-linking	**15 v 9**	**-0.62**	**0.65**	**0.047***	4 v 11	-0.56	0.68	0.36	5 v 10	0.32	1.25	0.55

P63104	YWHAZ	14-3-3 protein zeta/delta	Signalling regulator	**11 v 9**	**-0.70**	**0.62**	**0.04***	3 v 8	0.05	1.04	0.88	2 v 9	-0.87	0.55	0.47
P01036	CST4	Cystatin-S	Protease inhibitor	13 v 5	-0.23	0.85	0.79	**5 v 8**	**-1.23**	**0.43**	**0.004***	6 v 7	-0.29	0.82	0.58
P03973	SLPI	Secretory leukocyte antipeptidase inhibitor	Protease inhibitor Antimicrobial activity	11 v 7	0.99	1.99	0.18	**2 v 9**	**-2.08**	**0.24**	**0.04***	**2 v 9**	**-2.32**	**0.20**	**0.03***
Q96QA5	GSDMA	Gasdermin-A	Pore-forming protein	**11 v 7**	**1.03**	**2.04**	**0.02***	4 v 7	-0.06	0.96	0.93	6 v 5	-0.92	0.53	0.13
Q6ZUA9	MROH5	Maestro heat-like repeat family member 5	Unknown	6 v 11	0.01	1.01	0.99	2 v 4	0.18	1.13	0.82	**3 v 3**	**-1.48**	**0.36**	**0.01***
Q92820	GGH	Gamma-glutamyl hydrolase	Peptidase	12 v 3	-0.66	0.63	0.30	**2 v 10**	**-0.79**	**0.58**	**0.03***	2 v 10	0.08	1.06	0.81
Q9BZE2	PUS3	tRNA pseudouridine(38/39) synthase	Isomerase	**7 v 2**	**-1.66**	**0.32**	**0.004***	**2 v 5**	**1.66**	**3.16**	**0.004***	2 v 5	0.03	1.02	0.98
P60900	PSMA6	Proteasome subunit alpha type-6	Peptidase	7 v 0	NA	NA	NA	2 v 5	-0.35	0.78	0.37	**2 v 5**	**1.05**	**2.07**	**0.02***

Table 2-3: Detected proteins not present in all EBC samples but with a significant abundance difference between clinical groups.

*BPD: Preterm-born with history of bronchopulmonary dysplasia; PT_c: Preterm-born control; PT_{low}: Preterm born with low lung function; Log_2FC: Log2 fold-change between groupings. *Denotes p-vale <0.05*
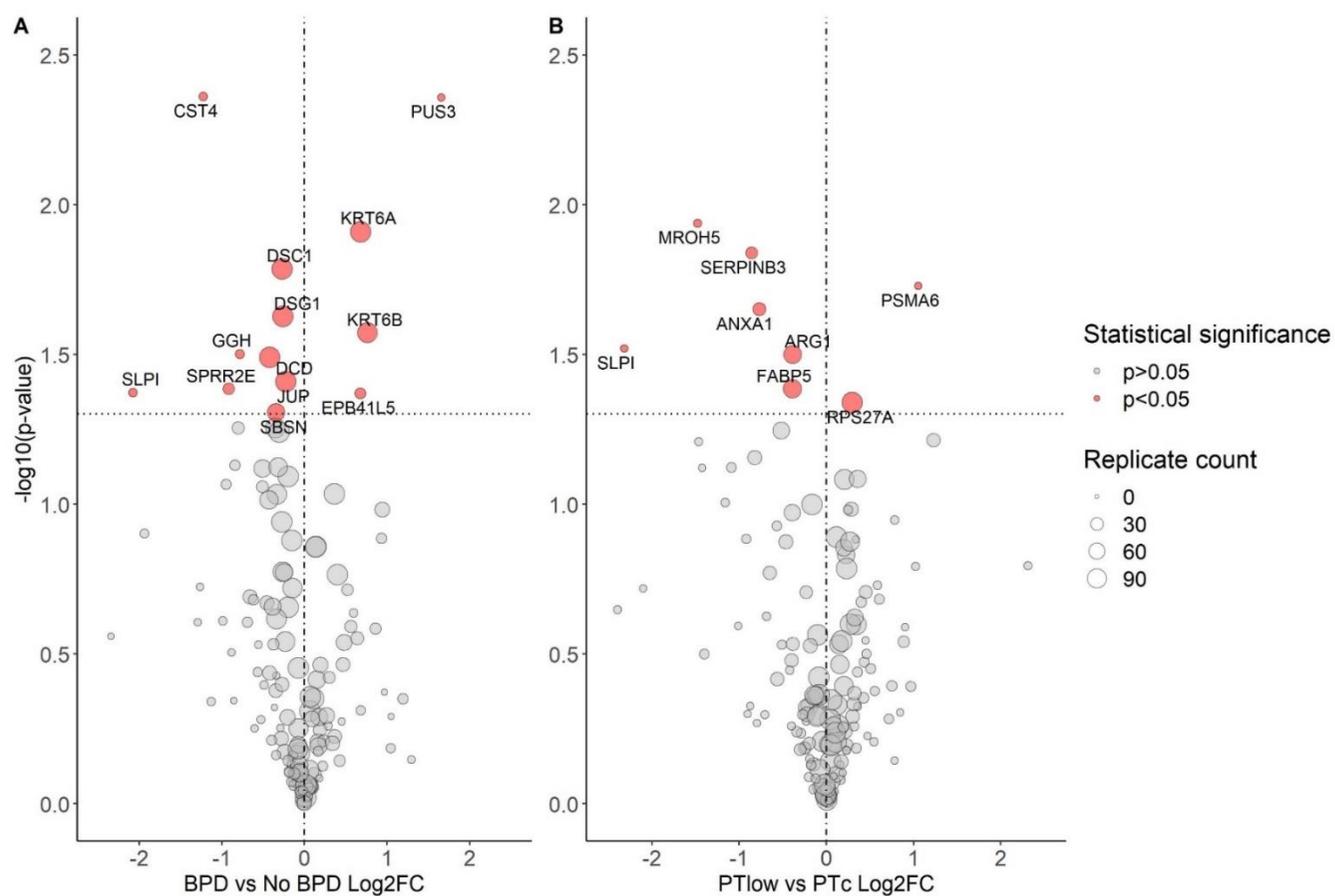
Figure 2-5: Volcano Plots demonstrating baseline protein abundance in EBC by BPD and Lung Function Status for Preterm-born children.

BPD: Bronchopulmonary dysplasia; PT_{low}: Preterm-born with low lung function; PT_c: Preterm-born control; Log2FC: Log_2 fold-change between groups Vertical line represents a Log_2FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to replicate number. Gene name associated with protein given if p<0.05
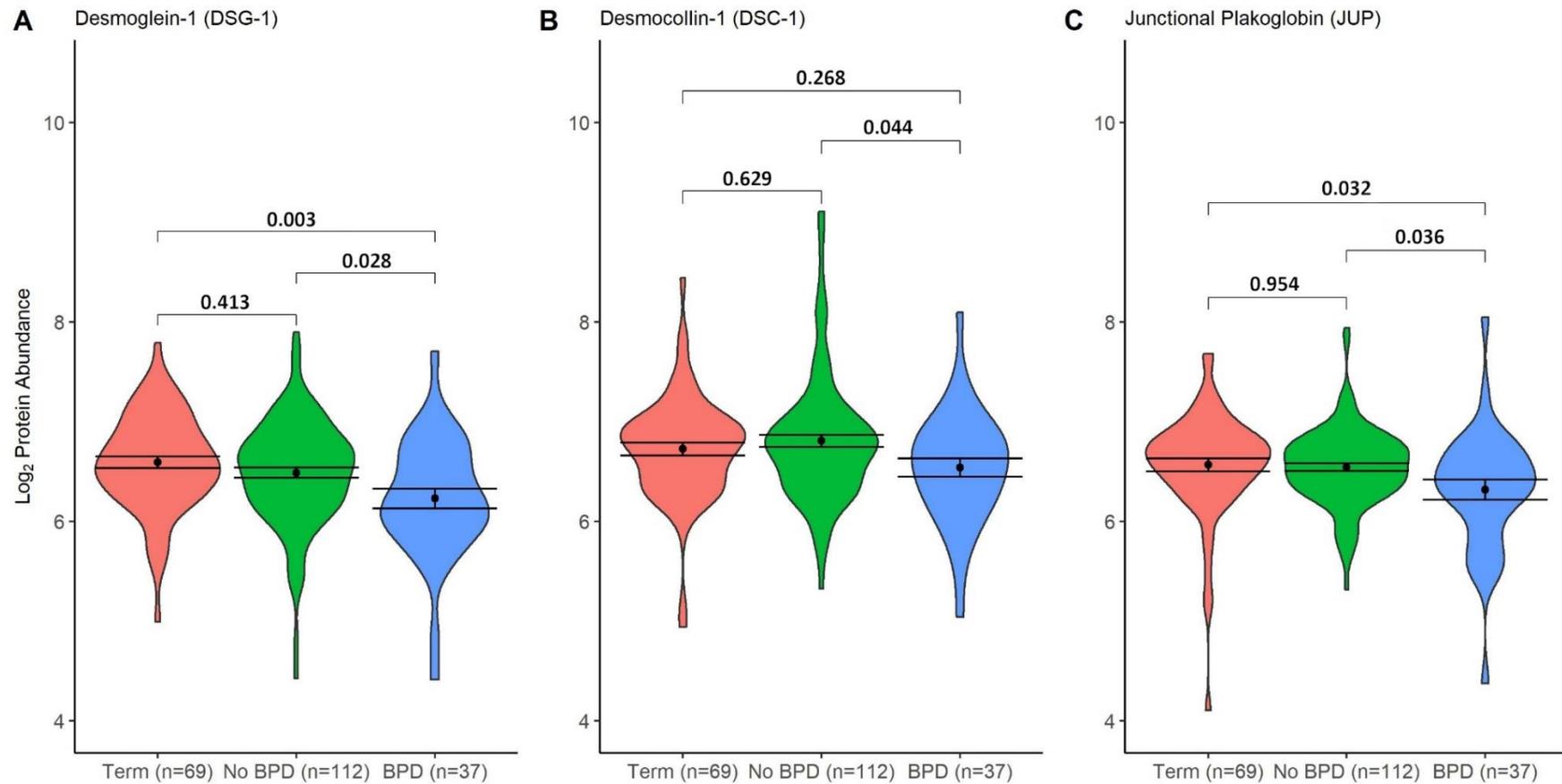
Figure 2-6: Violin plots: Desmosome/Cell Adhesion Baseline Protein Abundances in Children with history of BPD.

Term: Term-born control; No BPD: Preterm-born without BPD; BPD: Bronchopulmonary dysplasia; Dot and bars represent mean and standard error (SEM); Comparison bars between violin plots give p-values by ANOVA with post hoc Tukey's correction for multiple comparisons. Coloured areas represent distribution of sample values.
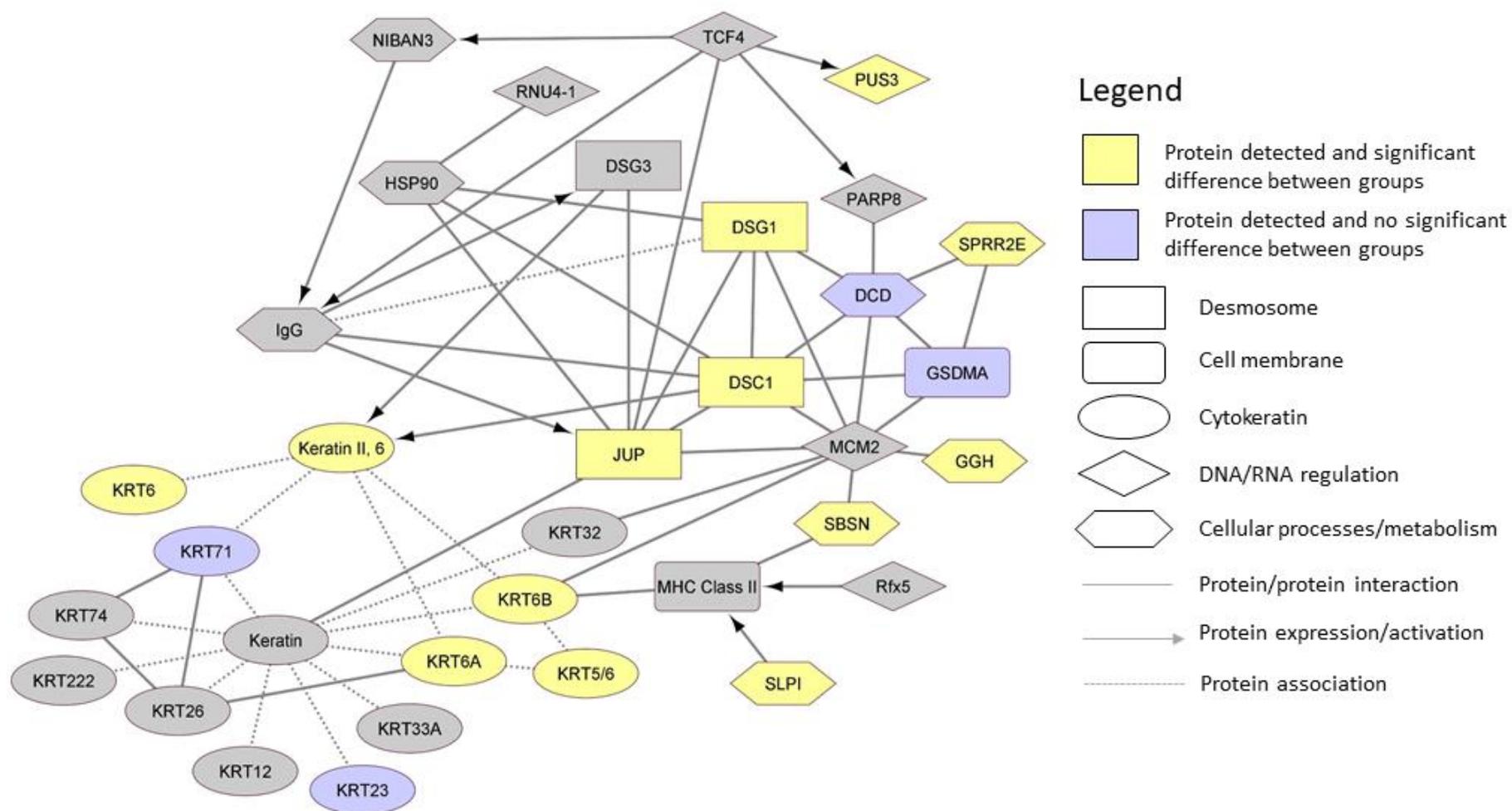
Figure 2-7: Protein network map of significant protein differences in EBC between BPD and No BPD preterm-born children.
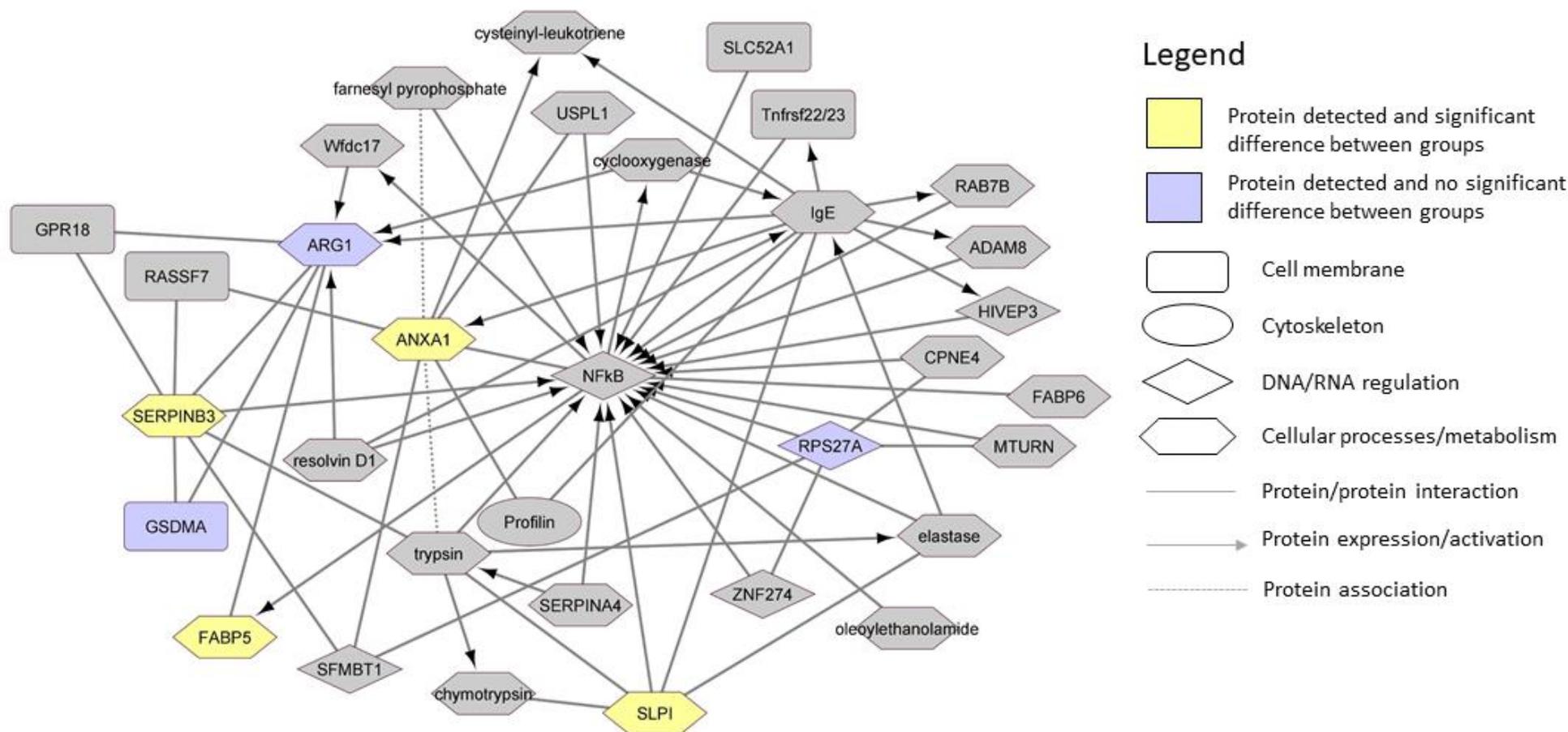
Figure 2-8: Protein network map of significant protein differences in EBC between PT_{low} and PT_c groups.

Univariable Analysis		
Variable	**Beta**	**p-value**
Desmoglein-1 (DSG1)		
Sex (Ref = Male)	-0.10	0.26
Age	-0.04	0.34
BPD (Ref = No BPD)	-0.23	**0.014***
Low Lung Function (Ref = PT$_c$)	-0.16	0.10
Asthma (Ref = No)	-0.05	0.66
Desmocollin-1 (DSC1)		
Sex (Ref = Male)	0.07	0.52
Age	-0.03	0.45
BPD (Ref = No BPD)	-0.27	**0.019***
Low Lung Function (Ref = PT$_c$)	-0.09	0.43
Asthma (Ref = No)	-0.09	0.49
Junctional Plakoglobin (JUP)		
Sex (Ref = Male)	-0.02	0.77
Age	-0.06	*0.06*
BPD (Ref = No BPD)	-0.23	**0.008***
Low Lung Function (Ref = PT$_c$)	-0.10	0.21
Asthma (Ref = No)	-0.15	*0.09*

Interaction Modelling						
	DSG1		**DSC1**		**JUP**	
	Beta	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**
No BPD * PTc	Ref	Ref	Ref	Ref	Ref	Ref
No BPD * PT$_{low}$	-0.12	0.30	-0.04	0.73	-0.04	0.66
BPD * PT$_c$	-0.23	0.11	-0.27	*0.09*	-0.18	0.12
BPD * PT$_{low}$	-0.35	**0.012***	-0.30	*0.06*	-0.30	**0.01***

Table 2-4: Linear regression analyses for relationships between desmosome proteins detected in EBC and early- and current-life factors in preterm-born children.

*p<0.05, italic = p<0.1. BPD: Bronchopulmonary dysplasia, PT$_{low}$: Preterm-born low lung function, PT$_c$: Preterm-born controls.

2.3.3.2 *Preterm-born children with low lung function (FEV$_1$≤85%) vs Preterm-born controls:*

For proteins detected in all samples, no significant differences were noted between the PT$_{low}$ group and PT$_c$ groups at baseline. A protein network map including all detected proteins with significant differences between PT$_{low}$ and PT$_c$ groups is given in Figure 2-8. Three antiproteases (Annexin A1 [ANXA1], Serpin B3 [SERPINB3], SLPI) were less abundant in the PT$_{low}$ group (-0.77, p=0.02; -0.86, p=0.01; -2.32, p=0.03 respectively), with reduced abundance of fatty acid-binding protein 5 (FABP5) (-0.39, p=0.04) when compared to the PT$_c$ group. The network map (Figure 8) did not demonstrate any direct links between these proteins.

2.3.4 Comparison of EBC proteome before and after inhaled therapies:

Figure 2-9, which includes proteins detected in all or some samples, shows significant differences before and after the three inhaler treatments. Table 2-5 shows the changes observed in the RCT treatment groups for proteins detected in all samples. Significant increases in abundance of DSG1 (0.58, p=0.003), DSC1 (0.47, p=0.048), JUP (0.52, p=0.002), KRT2 (0.32, p=0.047) and KRT10 (0.27, p=0.04) occurred after ICS/LABA treatment. For proteins not detected in every sample, increases in Protein-glutamine gamma-glutamyltransferase-E (TGM3) (log2 fold change 1.82, p=0.005), Filaggrin-2 (FLG2) (0.76, p=0.007) and Rab5 GDP/GTP exchange factor (RABGEF1) (0.76, p=0.02), and a decrease in Heat shock protein beta-1 (HSPB1) (-3.09, p=0.04) abundances were noted after ICS/LABA treatment. A protein network map demonstrating the relationships between these altered proteins for ICS/LABA treatment is shown in Figure 2-10.

Following ICS treatment, significant increase in abundance of cytokeratin-1 (KRT1) (0.34, p=0.03) and decreased abundances of cystatin-A (CSTA) (-0.66, p=0.01) and Zinc-alpha-2-

glycoprotein (AZGP1) (-0.70, p=0.03) was seen. Protein network map is shown in Figure 2-11. No differences were observed for proteins detected in every sample after placebo treatment, but immunoglobulin kappa constant (IGKC) (-2.02, p=0.04), Lipocalin-1 (LCN1) (-1.35, p=0.03), Plakophilin-1 (PKP1) (-0.80, p=0.03) and Catalase (CAT) (-0.33, p=0.04) decreased but were only noted in some samples.

Figure 2-12 shows significant increases in DSG1, DSC1 and JUP after ICS/LABA treatment which were not noted after ICS intervention. The PT_{low} group who had BPD in infancy had significant increases in abundance of all three proteins after ICS/LABA treatment, whereas PT_{low} without BPD only had significantly increased JUP abundance. Following ICS/LABA treatment in the PT_{low} with BPD group, levels of DSG1, DSC1 and JUP were comparable to the term control group at baseline (p = 0.56, 0.12, 0.06 respectively). Figure 2-10 demonstrates the biological links between these proteins and the changes observed for TGM3, FLG2, HSPB1, KRT2 and KRT10 as described above.
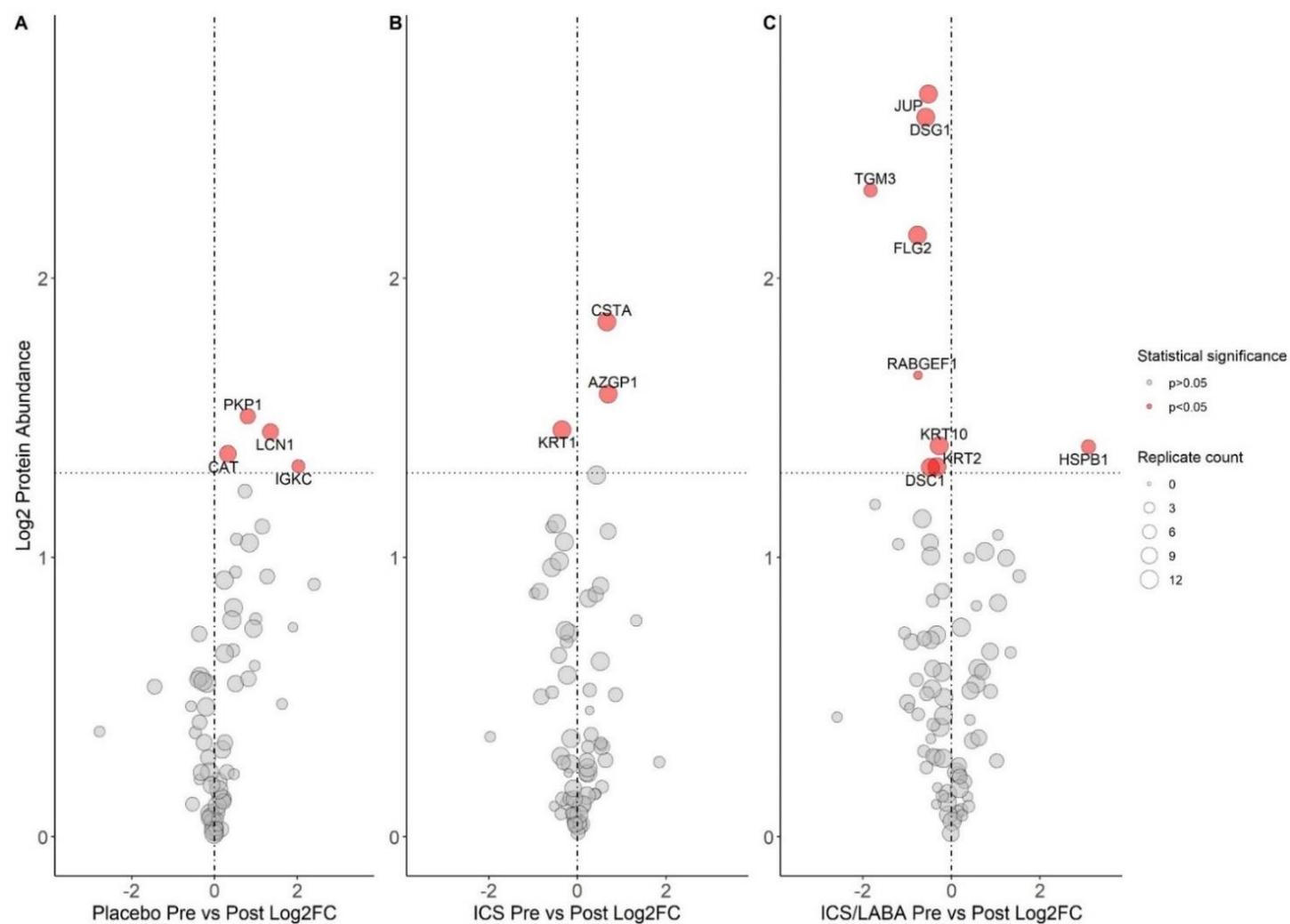
Figure 2-9: Volcano Plots demonstrating protein abundances in EBC pre- and post-RCT treatment.

Log_2FC: Log_2 fold-change between groups. Vertical line represents a Log_2FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to number of samples in which protein was detected. Gene name associated with protein given if $p<0.05$

Accession Number	Gene Name	Protein Name	Protein Function	Placebo n=12 Pre vs Post			ICS n=17 Pre vs Post			ICS/LABA n=17 Pre vs Post		
				Log$_2$FC	Ratio	p	Log$_2$FC	Ratio	p	Log$_2$FC	Ratio	p
Q08554	DSC1	Desmocollin-1	Cell-cell junction	0.24	1.18	0.12	0.04	1.03	0.83	**-0.47**	**0.72**	**0.048***
Q02413	DSG1	Desmoglein-1	Cell-cell junction	-0.20	0.87	0.34	-0.47	0.72	0.08	**-0.58**	**0.67**	**0.003***
P15924	DSP	Desmoplakin	Cell-cell junction, Cytoskeleton	0.07	1.05	0.77	-0.18	0.88	0.19	-0.21	0.86	0.26
P14923	JUP	Junction plakoglobin	Plasma membrane protein complex	-0.12	0.92	0.59	-0.41	0.75	0.10	**-0.52**	**0.70**	**0.002***
H6VRG2	KRT1	Cytokeratin-1	Cytoskeleton	-0.18	0.88	0.28	**-0.34**	**0.79**	**0.03***	-0.17	0.89	0.32
H6VRG3	KRT1	Cytokeratin-1	Cytoskeleton	-0.34	0.79	0.27	-0.58	1.49	0.11	-0.33	0.80	0.19
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal	Cytoskeleton	0.10	1.07	0.68	-0.16	0.90	0.55	**-0.32**	**0.80**	**0.047***
P13647	KRT5	Keratin, type II cytoskeletal 5	Cytoskeleton	-0.03	0.98	0.89	0.02	1.01	0.92	0.02	1.01	0.88
P02538	KRT6A	Keratin, type II cytoskeletal 6A	Cytoskeleton	0.19	1.14	0.73	0.24	1.18	0.59	0.60	1.52	0.25
P35527	KRT9	Keratin, type I cytoskeletal 9	Cytoskeleton	-0.28	0.82	0.28	-0.30	0.81	0.09	0.11	1.08	0.59
P13645	KRT10	Keratin, type I cytoskeletal 10	Structural protein extracellular space	0.09	1.06	0.64	-0.14	0.91	0.45	**-0.27**	**0.83**	**0.04***
P02533	KRT14	Keratin, type I cytoskeletal 14	Cytoskeleton	0.03	1.02	0.82	-0.22	1.16	0.26	0.22	1.16	0.18
P08779	KRT16	Keratin, type I cytoskeletal 16	Cytoskeleton	-0.11	0.93	0.83	0.07	1.05	0.90	0.55	1.46	0.28
Q04695	KRT17	Keratin, type I cytoskeletal 17	Intermediate filament cytoskeleton	-0.01	0.99	0.98	0.24	1.18	0.57	0.76	1.69	0.10
Q8N1N4	KRT78	Keratin, type II cytoskeletal 78	Cytoskeleton	0.24	1.18	0.22	-0.27	0.83	0.18	-0.17	0.89	0.37
P31944	CASP14	Caspase-14	Protease	0.85	1.80	0.09	0.52	1.43	0.24	-0.44	0.74	0.30
P01040	CSTA	Cystatin-A	Protease inhibitor	-0.07	0.95	0.85	**0.66**	**1.58**	**0.01***	-0.25	0.84	0.41
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a	Structural component of ribosome	-0.02	0.99	0.95	0.44	1.36	0.05	-0.18	0.88	0.52
P25311	AZGP1	Zinc-alpha-2-glycoprotein	Major histocompatibility complex protein	0.46	1.38	0.15	**0.70**	**1.62**	**0.03***	-0.10	0.93	0.75

Table 2-5: Treatment effect on proteins detected in every EBC sample.

ICS: Inhaled corticosteroids; ICS/LABA: Inhaled corticosteroids/long-acting beta agonists; Log$_2$FC: Log2 fold-change between groups. *Denotes p-vale <0.05
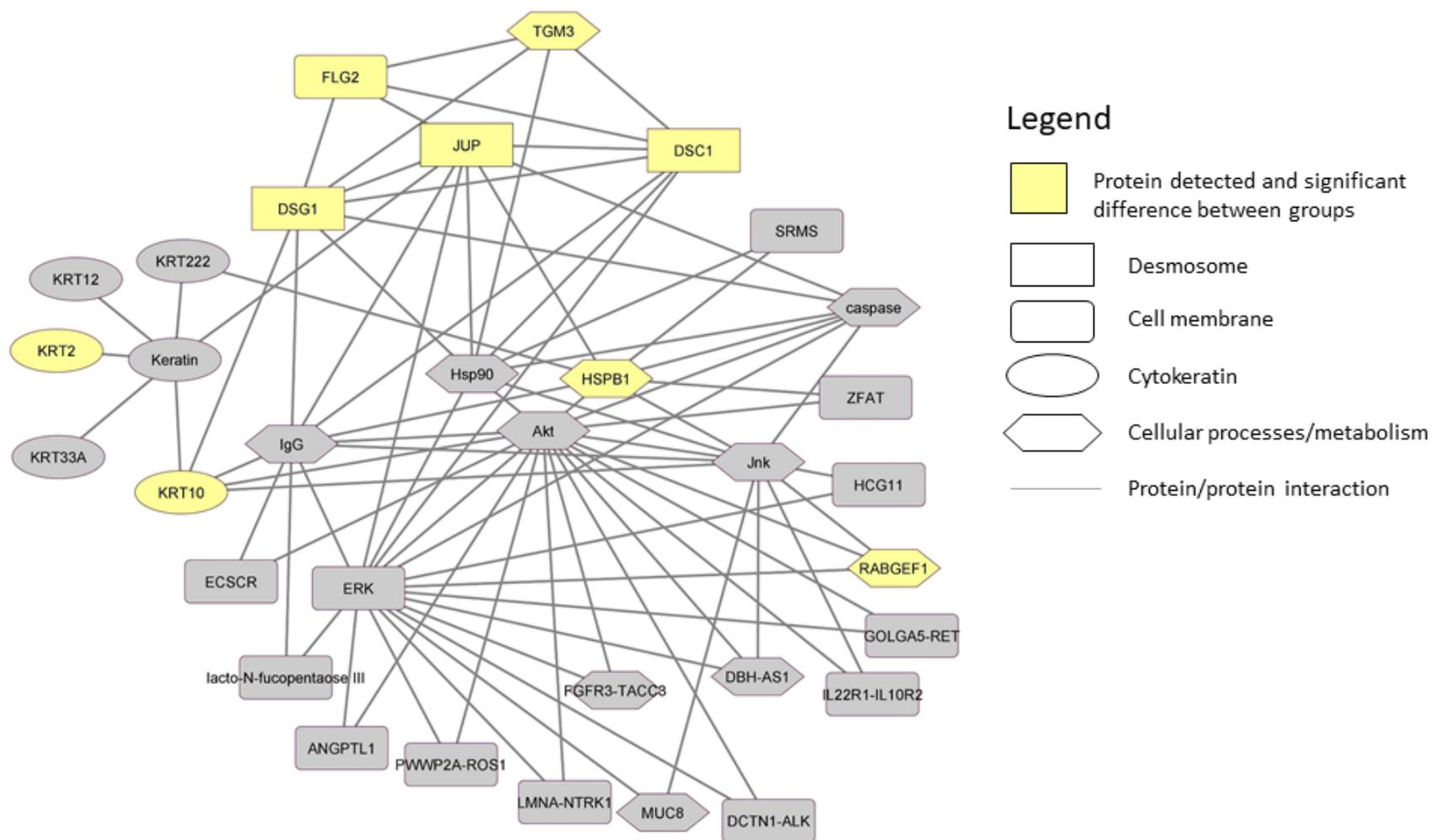
Figure 2-10: Protein network map of significant protein differences in EBC before and after ICS/LABA treatment.
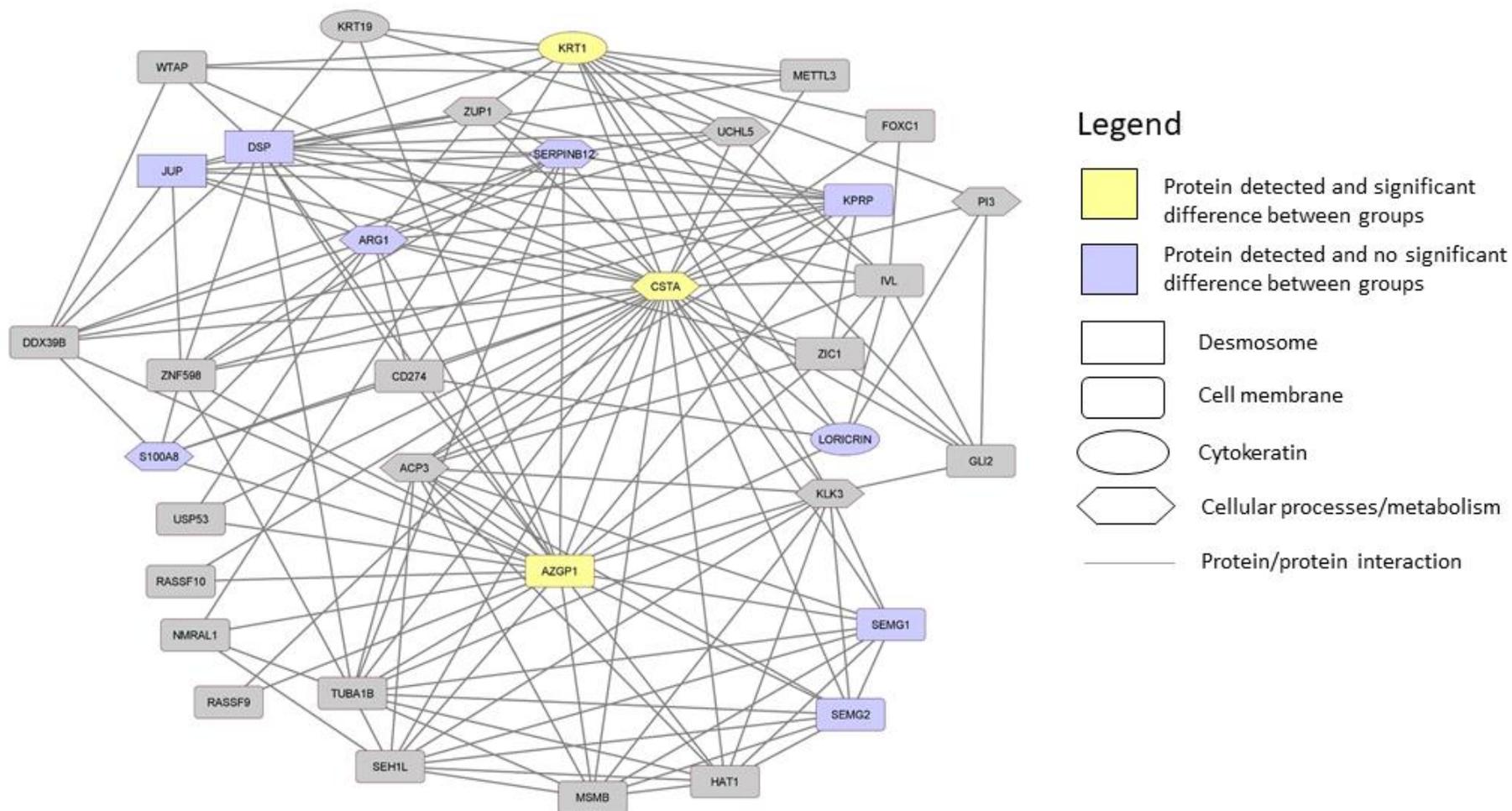
Figure 2-11 Protein network map of significant protein differences in EBC before and after ICS treatment.
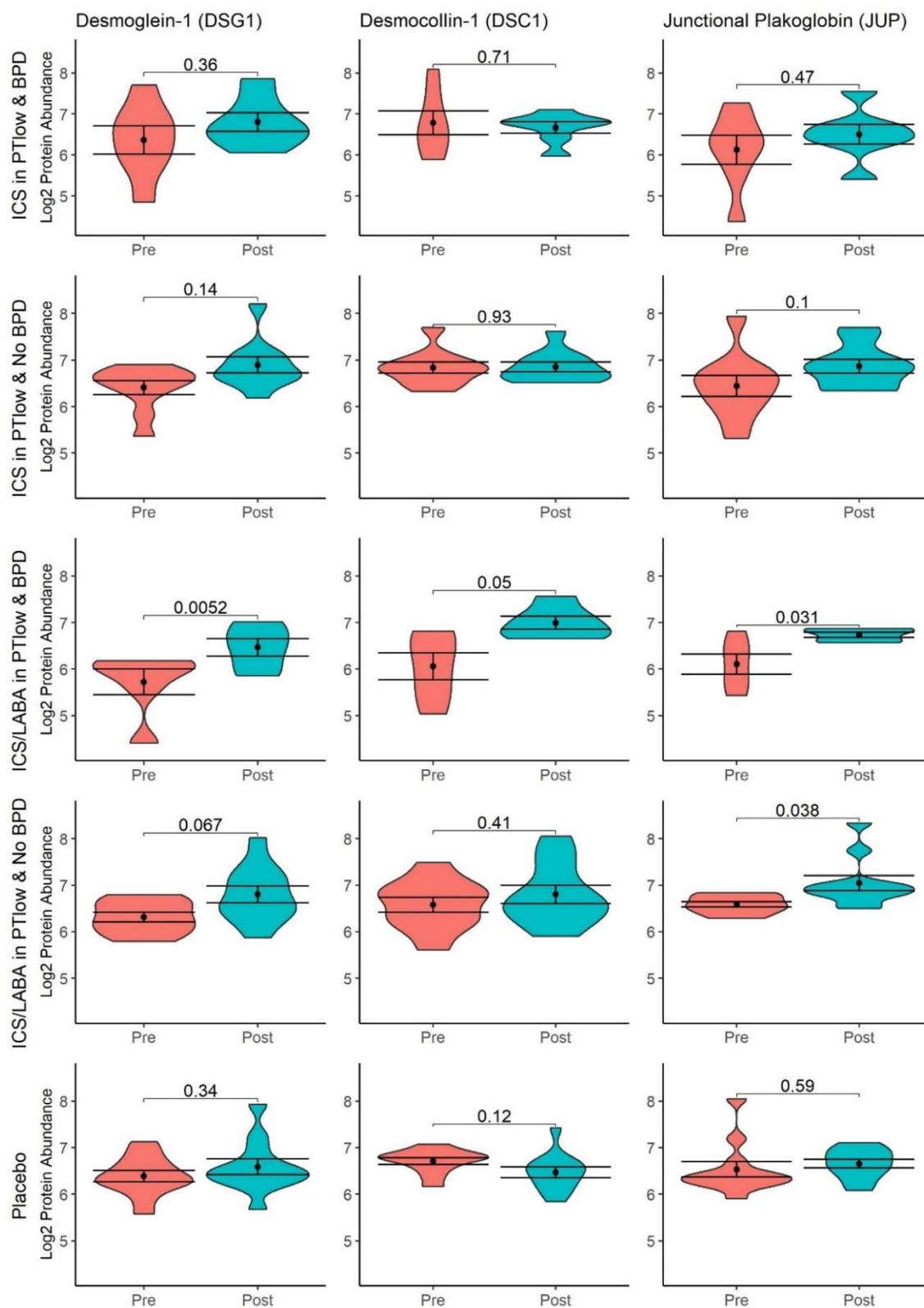
Figure 2-12: Violin plots of Desmosome Proteins before and after treatment with Placebo, ICS or ICS/LABA by BPD status.

BPD: Bronchopulmonary dysplasia; Coloured areas represent distribution of sample values. Dot and bars represent mean and standard error (SEM); Comparison bars between violin plots give p-values by paired samples t-test.

2.4 Discussion

This chapter has examined the changes in the detectable proteome in exhaled breath condensate using LC/MS with TMT in preterm-born children both before and after inhaled therapies. I have described the effect of preterm-birth and its major respiratory complication, BPD, on the EBC proteome of school-aged children. I have shown that those individuals born preterm with a history of BPD continue to have detectable significant differences in protein abundances for key structural proteins involved in desmosome and cytoskeleton formation, several years after the initial pulmonary insult. I have also demonstrated that the reduced abundance of desmosome proteins, namely DSG1, DSC1 and JUP, seen in preterm-born school-aged children with low lung function (FEV$_1$ ≤85%) can be reversed with 12-weeks of combination ICS/LABA inhaler therapy. This effect was predominantly noted in those children in the PT$_{low}$ group with a history of BPD, with linear modelling confirming that those in both BPD and PT$_{low}$ groups have a significantly reduced abundance of these three proteins at baseline. Previously published data from the RHiNO trial has demonstrated that combination inhaler therapy is also the most effective treatment for preterm-born children with low lung function, increasing FEV$_1$ by over 14% on average (Goulden et al., 2021).

The mechanism by which some preterm-born children continue to experience lung function deficits remains incompletely understood, and there is evidence of developmental structural changes to the lung parenchyma and chronic active inflammatory processes (Thebaud et al., 2019). Desmosomes have historically been thought to provide inert structural support to tissues through their strong cell-to-cell adhesion properties, however more recent evidence has shown that they have an active role in cell signalling, proliferation, migration, and apoptosis (Holthofer et al., 2007, Green et al., 2010). Despite minimal published evidence taking a proteomics approach, a reduction in desmosome proteins has been implicated in respiratory pathologies. In a murine asthma model, analysis of bronchial wall tissue found

reduced expression of DSG1 following asthma exacerbation and reduced epithelial barrier integrity, potentially predisposing to further exacerbation (Bao et al., 2019). Desmosome size and number have also been seen to be reduced in bronchial biopsies taken from asthmatic adults (Shahana et al., 2005), and two studies examining cultured human bronchial epithelium reported that pro-inflammatory cytokines (TNF-α and INF-γ) reduced expression of desmosomes and JUP, an effect which could be reversed by the administration of corticosteroids (Andersson et al., 2010, Carayol et al., 2002). The reasons why I observed changes in DSG1, DSC1 and JUP in the ICS/LABA therapy group but not in the ICS group is unclear, however the underlying pathology of BPD is different to that of asthma, with infant post-mortem evidence of smooth muscle extension into the peripheral airways (Bush et al., 1990), and airway biopsies demonstrating peri-bronchial fibrosis, neoangiogenesis, and a predominantly CD8+ T lymphocyte epithelial infiltrate in adolescent survivors of BPD (Galderisi et al., 2019).

I also noted an increase in abundance in the cytokeratins KRT6A and KRT6B in those children with a history of BPD. Cytokeratins comprise the intracytoplasmic cytoskeleton of epithelial tissues and form an important component of intermediate filaments, which connect to desmosomes, helping cells and tissues resist mechanical stress (Herrmann et al., 2007). Whilst detection of these proteins can represent epidermis sample contamination, cytokeratins have previously been shown as the most abundant protein type in EBC (Kurova et al., 2009), and both KRT6A and KRT6B have been identified in EBC proteomic and gene expression analyses as potential biomarkers for lung carcinomas (Lopez-Sanchez et al., 2017, Xiao et al., 2017). In addition, studies examining EBC in mechanically ventilated adults and smokers have identified an increased total cytokeratin abundance as a marker of airway stress and damage to the lung parenchyma (Gessner et al., 2008, Gianazza et al., 2004). Additionally, neonatal studies have demonstrated an increased abundance of cytokeratin-19 fragments in

the serum of preterm infants who required mechanical ventilation and in those who later developed BPD or died (Panahabadi et al., 2021). In conjunction with the changes in DSG1, DSC1 and JUP, the increased abundance of KRT6A and KRT6B in the BPD group suggest that these children have persistent parenchymal structural changes secondary to airway inflammation and tissue stress.

I have focused the majority my analysis on proteins detected in every sample to capitalise on our large sample size and ensure robust findings. Overall, the protein content of EBC was low, as previously reported (Bloemen et al., 2011, Kurova et al., 2009, Gianazza et al., 2004), and close to the limits of detection. As a result, I performed exploratory analyses of proteins detected only in a proportion of samples, as the TMT methodology allowed robust quantification of these proteins in multiple replicates, most of which exceed sample sizes of many other published proteomic studies. DCD was detected in a very high proportion of our samples (98%) noting significantly decreased abundance in the BPD group. DCD, a peptidase with antimicrobial activity, has been described in EBC samples previously (Kurova et al., 2009), and increased detection was weakly associated with asthma in a small paediatric proteomic study (Bloemen et al., 2011). In addition, I observed reduced abundance of several protease inhibitors in the BPD or PT$_{low}$ groups, including ANXA1, SERPINB3, CSTA and SLPI, with reduced abundance of SLPI being noted in both the BPD and PT$_{low}$ groups compared to controls. Our group have previously demonstrated an imbalance between protease and antiprotease activity, and subsequent tissue remodelling, may been implicated early in the pathogenesis of BPD (Davies et al., 2010) but this has not been reported in later life. Animal models of BPD pathogenesis/oxidative stress have shown significant increase in SLPI over the first 10 days of life (Wagenaar et al., 2004); however, tracheobronchial aspirates from ventilated preterm neonates showed a relative deficit of SLPI with elevated protease activity observed over the first few weeks of life (Watterberg et al., 1994, Sveger et al., 2002). This result should be

interpreted with caution as SLPI was detected in a minority of samples. ANXA1, a protease inhibitor also known to have innate immune properties, which was decreased in the PT_{low} group but not in the BPD group, has been implicated in early lung injury in neonatal mouse models (Raffay et al., 2013). This decrease in anti-proteases suggest that there may be an imbalance in protease/antiprotease activity, but additional work will be required to estimate both proteases and anti-proteases in more appropriate samples (e.g. bronchoalveolar lavage or induced sputum) which are more invasive and ethically more challenging to obtain in this cohort.

It is well established that survivors of preterm birth, both with and without BPD, are at risk of lung function deficits in later life (Kotecha et al., 2022b), and there is increasing evidence that BPD is a poor predictor of lung function in later life (Hart et al., 2022, Corwin et al., 2018). In this cohort, I saw fewer differences in biologically related proteins at baseline when comparing PT_{low} and PT_c groups in comparison to those with and without BPD, and less than half of the children in the RCT had BPD. It is most likely that the decrease observed in DSG1, DSC1 and JUP seen in the BPD group, which is reversed by combination inhaler therapy, is due to cellular injury secondary to continuing airway inflammation (Teig et al., 2012, Filippone et al., 2012), although further work is needed to clarify this relationship. Overall, those children treated with ICS/LABA had significantly decreased HSPB1 following treatment. HSPB1 is a member of the small heat-shock protein family, which are molecular chaperones controlling protein folding and preventing aggregation. HSPB1 has been shown to have an important role in the cellular response to oxidative stress, preventing apoptosis and regulating inflammation (Acunzo et al., 2012), adding further to the suggestion of chronic airway inflammation contributing to low lung function. Previous studies have also demonstrated evidence of persistent airway inflammation in children several years after very preterm birth (<32 weeks' gestation), with raised neutrophil and IL-8 values in induced

sputum (Teig et al., 2012); however, the link between this chronic inflammation and lung function parameters was not clear. This EBC proteomic data was unable to show as clear a pattern for differences in proteins for preterm-born children who had low lung function but did not have BPD in infancy. In addition, I did also observe a change in protein abundances after placebo treatment which is challenging to fully explain. This is the first time to my knowledge that proteomics has been performed on EBC before and after a trial containing a placebo arm.

2.4.1   Strengths and Limitations

This study represents one of the largest proteomic analyses of EBC, and the first time, to my knowledge, that preterm-born children have been studied. By using EBC, there has been direct sampling of ELF, representative of the biochemistry of the airways, in a simple, well-tolerated and non-invasive manner. I and the team at University of Bristol Proteomics Facility have demonstrated that not only is it technically possible to perform a quantitative proteomic analysis of EBC using Tandem Mass Tagging on a large sample size, but also to identify meaningful changes within our clinical groups. By restricting the primary analysis to proteins detected in every sample, I have achieved robust findings, strengthened further by the modulation of these proteins of interest after inhaler treatment in the RCT. By using an untargeted approach and performing exploratory analysis of less frequently detected proteins, I have also implicated potentially important protein relationships, including protease/antiprotease dynamics, that future work should explore. Limitations of this analysis include the overall low protein content of EBC, as discussed above, and the relatively low number of proteins detected in every sample, which limited the statistical analysis approaches I could undertake such as Principal Component Analysis. There may have been very low levels of some proteins in the samples which did not reach the limit of detection for the TMT methodology utilised. In addition, it is not possible to normalise the EBC samples as

there is currently no universally agreed internal or external control value to adjust for differing EBC volumes between subjects (Horvath et al., 2017), however this is unlikely to have affected the results, as EBC volumes did not appear to demonstrate a significant association with my findings in linear regression modelling.

2.5 Conclusions

In conclusion, the proteomic analysis of EBC from preterm-born school-aged children has revealed a significantly reduced abundance of three key desmosome proteins, DSG1, DSC1 and JUP, in those with a history of BPD, in addition to an increase in cytokeratins. In linear modelling, using BPD and PTlow as interaction terms, DSG1 and JUP were significantly reduced in those children in the BPD group with current low lung functions. These reduced protein abundances were reversed for those children with reduced lung function who entered a 12-week clinical trial and received ICS/LABA inhaler therapy. I can hypothesise that the changes seen in these desmosome constituents, as well as the increased abundance of cytokeratins, are related to an ongoing inflammatory process within the airways of those individuals with BPD altering parenchymal structure, and potentially reducing lung function, especially as there is also a suggestion that there is a protease/anti-protease imbalance.

3 Metabolomic Analysis of Exhaled Breath Condensate

3.1 Introduction:

As discussed in my Introduction, BPD (also known as CLD) is the one of the commonest respiratory consequences of preterm birth and, despite advances in neonatal care over the last twenty years, rates of BPD are continuing to rise (Stoll et al., 2015, Jensen et al., 2021), likely related to improving survival from increasingly immature gestational ages at birth (Edwards et al., 2024). The pathogenesis of BPD is multifactorial, but pulmonary injury secondary to oxygen free radical production and inflammation form an important common pathway leading to altered lung development (Buczynski et al., 2013, Chakraborty et al., 2010). As I have discussed previously (sections 1.4.3 and 1.4.4) history of preterm birth, both with and without a history of BPD, has been consistently associated with poorer lung function in later life (Kotecha et al., 2022b), and there is growing evidence that those with a history of BPD risk the development of chronic obstructive pulmonary disease (COPD) in early adulthood (Gibbons et al., 2023, Doyle et al., 2019b), as well as being diagnosed with asthma (Pulakka et al., 2023), although there is increasing recognition that prematurity-associated lung disease (PLD) has different underlying mechanism to asthma (Cousins et al., 2023).

Lung function continues to develop throughout childhood and adolescence, with increasing number of alveoli, airway size and lung volume (Stocks et al., 2013), before declining after early adulthood (Agusti and Faner, 2019). Preterm-born individuals have been shown to have significantly lower forced expiratory volume in 1 second (FEV_1) than those born at term (Bardsen et al., 2022), therefore providing a potential therapeutic window of opportunity for optimizing peak lung function, and highlights the importance of understanding the underlying mechanisms of PLD during childhood.

As I discussed in Chapter 2, Exhaled breath condensate (EBC) is a useful sample type to study in children as it is easily and non-invasively collected (Section 2.1.1). EBC is composed of droplets of epithelial lining fluid (ELF), evolved from all compartments of the lung during tidal breathing. ELF is a complex matrix of compounds, which includes metabolites, and reflects lung tissue biology (Davis et al., 2012). As discussed in section 1.5, metabolomic methods simultaneously analyse the entire low-molecular weight (<2000Da) metabolite content of biological samples and have been applied extensively to EBC in the study of both adult and paediatric respiratory diseases such as COPD, asthma and CF (Maniscalco et al., 2019), showing the ability to discriminate between asthma phenotypes, and evidence of medication altering the EBC metabolome in CF. It is clear metabolomics offers a tool to unravel mechanisms of disease pathogenesis and progression and identify potential groups of biomarkers in respiratory pathologies. Metabolomic techniques have previously been used on tracheal aspirates obtained during the neonatal period to study the pathogenesis of BPD (Piersigilli et al., 2019), finding an increase in metabolites related to hypoxic stress and nitric oxide synthesis.

3.1.1 Aims

After finding significant associations between the EBC proteome and BPD and PLD (Chapter 2), I hypothesized that the EBC of preterm-born school-aged children who had a history of BPD in infancy would show altered metabolite profiles when compared to preterm-born and term-born control children. Therefore, in this exploratory analysis, I aimed to characterise the metabolome of preterm-born school-aged children with a history of BPD compared to preterm-born children without a history of BPD and term-born controls.

3.2 Methods:

3.2.1 Participants:

This analysis was conducted on a cohort of children recruited to the Respiratory Health Outcomes in Neonates study (RHiNO, EudraCT: 2015-003712-20) which has been described extensively previously in the literature (Goulden et al., 2021, Hart et al., 2022, Course et al., 2023a) and in this thesis (section 1.6 and 2.2.1). In brief, children from a previous study (Edwards et al., 2015) were supplemented with additional preterm-born children identified by the NHS Wales Informatics Service and sent a respiratory and neurodevelopmental questionnaire if they were born ≤34 or ≥37 weeks' gestation and were aged 7-12 years. Children with significant congenital malformations, cardiopulmonary or neuromuscular disease were excluded. Ethical approval was obtained from the South-West Bristol Research Ethics Committee (15/SW/0289). Parents gave informed written consent and children provided assent.

As I described in section 2.2.2, following a home assessment, a subset of responders attended the hospital-based children's research facility for comprehensive clinical examination and respiratory testing including collection of EBC, conducted by a trained nurse and paediatrician between January 2017 and November 2019. Spirometry (MasterScreen Body and PFT systems, Vyaire Medical, Germany) was performed to ATS/ERS guidelines (Miller et al., 2005) and normalised using GLI references (Quanjer et al., 2012). Any respiratory medications were withheld prior to their assessment (short- and long-acting β2-agonists for 8- and 48-hours respectively; inhaled corticosteroids for 24 hours; and leukotriene receptor antagonists for 48 hours) and children were free of respiratory infections for at least three weeks prior to testing. Term-born children who had %FEV_1 >90% were included as term controls. BPD was defined as oxygen-dependency of 28 days or greater for those born <32 weeks' gestation and at 56 days of age for those born ≥32 weeks' gestation (Ehrenkranz et al., 2005). Intrauterine

growth restriction (IUGR) was defined as birthweight <10th percentile adjusted for sex and gestation (LMS Growth version 2.77, Medical Research Council, UK). Doctor-diagnosed asthma was self-reported by parents. Neonatal history was corroborated with medical records. Socioeconomic status was assessed using the Welsh Index of Multiple Deprivation (StatsWales, 2019) scores from 2019, the most contemporaneous available for this cohort.

3.2.2 EBC Sampling:

As I described in section 2.2.3, EBC was collected during the research unit visit in the RHiNO trial, using a cooling tube (RTube®, Respiratory Research Inc., Texas, USA), that was pre-cooled to −20 °C for at least two hours prior to use, during 10 minutes of passive tidal breathing, with the participant wearing a nose clip, stopping briefly to swallow saliva if needed. EBC was collected immediately prior to spirometry and once collected, samples were immediately separated into aliquots and stored at -80°C pending analysis.

3.2.3 Metabolome analysis:

EBC samples were analysed using Gas Chromatography Time-of-Flight Mass Spectrometry (GCTOF-MS) by the West Coast Metabolomics Centre (University of California, Davis), who have previously published their analytical method (Fiehn et al., 2008). 50μL of each sample was fractionated using an Agilent 6890 gas chromatograph (Agilent, Santa Clara, CA, USA), controlled using Leco ChromaTOF software v2.32 (LECO, St. Joseph, MI, USA), in a Rtx-5Sil MS (Restek, Bellafonte, PA, USA) column (30m length x 0.25mm internal diameter with 0.25μm film made of 95% dimethyl/5%diphenylpolysiloxane). Column temperature was maintained between 50-330°C, with a helium mobile phase. Injection volumes of 0.5μL were used, with injection temperatures starting at 50°C, ramped up to a maximum temperature of 250°C by 12°Cs^{-1}. Oven temperature program was set to 50°C for 1 min, then ramped at 20°C min^{-1} to

330°C, and held constant for 5 min. The analytical GC column was protected by a 10m long empty guard column which was cut by 20cm intervals whenever the reference mixture QC samples indicated problems caused by column contamination. This sequence of column cuts has previously been validated, with no detrimental effects being detected with respect to peak shapes, absolute or relative metabolite retention times or reproducibility of quantifications. This chromatography method yields excellent retention and separation of primary metabolite classes (amino acids, hydroxyl acids, carbohydrates, sugar acids, sterols, aromatics, nucleosides, amines and miscellaneous compounds) with narrow peak widths of 2–3s and very good within-series retention time reproducibility of better than 0.2s absolute deviation of retention times. Automatic liner exchanges after each set of 10 injections were used, which reduces sample carryover for highly lipophilic compounds.

All spectra were acquired using a Leco Pegasus IV (LECO, St. Joseph, MI, USA) time of flight mass spectrometer, with unit mass resolution at 17 spectra s^{-1} from 80-500Da at -70eV ionization energy and 1800V detector voltage with a 230°C transfer line and a 250°C ion source. Raw data files were normalised to QC/pool samples using the systematic error removal by random forest (SERRF) method(Fan et al., 2019). Raw data files were processed and metabolites identified with the BinBase metabolomics database (Lai et al., 2018) by West Coast Metabolomics Centre, using an in-house algorithm based on the following: validity of chromatogram (<10 peaks with intensity>10^7 counts s^{-1}), unbiased retention index marker detection (MS similarity >800, validity of intensity range for high m/z marker ions), retention index calculation by 5th order polynomial regression. Spectra were cut to 5% base peak abundance and matched to database entries from most to least abundant spectra using the following matching filters: retention index window ±2,000 units (equivalent to about ±2s retention time), validation of unique ions and apex masses (unique ion must be included in

apexing masses and present at >3% of base peak abundance), mass spectrum similarity fitted criteria dependent on peak purity and signal/noise ratios and a final isomer filter. Quantification of metabolites were reported as spectral peak height of the unique ion detected (m/z value) at the specific retention index. Peak heights are more precise for low abundant metabolites than peak areas, due to the larger influence of baseline determinations on areas compared to peak heights.

3.2.4 Statistical analysis:

Sample demographics were compared using chi-squared or one-way ANOVA with Bonferroni correction tests as appropriate. Metabolite quantities were log_{10} transformed and visually inspected for normality. Metabolites with mean and median peak intensities below the mass spectrometer's limit of detection were removed from further analysis to ensure robust statistical comparisons between clinical groups. Fold changes between groups were calculated and log_2 transformed (log_2FC) for visualization. Independent t-test/ANOVA with post-hoc Bonferroni correction was used to compare metabolite quantities between groups. Metabolite Set Enrichment Analysis (MSEA; identifying biological processes linked to over-represented metabolites) was performed on all metabolites identified with a significantly different quantity between groups using the Small Molecule Pathways Database (SMPDB) (Jewison et al., 2014), which is based on the Human Metabolome Database (HMDB). Univariable and multivariable linear regression models were used to identify associations between participant characteristics and metabolites of interest identified by MSEA. $p < 0.05$ was considered statistically significant. I performed all statistical analyses using R v4.0.4 (R Foundation for Statistical Computing, Austria) and R packages *"dplyr", "ggplot2", "ggpubr"*, and MetaboAnalyst v5.0 (www.metaboanalyst.ca) (Pang et al., 2022).

3.3 Results:

As described in section 1.6 total of 241 children underwent detailed assessment at the research facility, performed by Dr Michael Cousins and Dr Kylie Hart. EBC was successfully collected and analysed from 214 (89%) children with adequate spirometry. Sample demographics are shown in Table 3-1. 34 preterm-born children had a diagnosis of BPD (13 mild BPD, 21 moderate/severe BPD) (Ehrenkranz et al., 2005). Preterm-born children with a history of BPD (BPD) were born at a significantly lower gestational age compared to those without BPD (No BPD) (mean±SD 27.1±2.1 weeks vs 31.8±1.9, p=<0.001), with a significantly lower birthweight (1029±415g vs 1817±493, p=<0.001). No BPD group was significantly older than the Term group at assessment (10.3±1.1 years vs 9.7±1.1, p=0.002) but there was no significant age difference between the BPD and the No BPD and Term groups. There were no differences in socioeconomic deprivation score between the three groups. Percent predicted forced expiratory volume in 1sec (%FEV_1) was significantly lower in the BPD group compared to both the No BPD (86.9±15.9 vs 93.2±14.1, p=0.036) and Term groups (86.9±15.9 vs 104.3±7.1, p=<0.001). FEV_1/FVC ratio was also significantly lower in the BPD group compared to the No BPD (0.77±0.10 vs 0.82±0.09, p=0.002) and Term groups (0.77±0.10 vs 0.85±0.06, p=<0.001). Percent predicted mid-expiratory flows (%FEF_{25-75}) were also significantly lower in the BPD group compared to No BPD (64.0±25.8 vs 79.3±24.6, p=0.003) and Term groups (64.0±25.8 vs 94.7±19.1, p=<0.001).

Variable	Preterm BPD n = 34	Preterm No BPD n = 110	Term n= 70
Sex (male), n(%)	15 (44.1)	54 (49.1)	37 (52.9)
Ethnicity (white), n(%)	32 (94.1)	103 (93.6)	69 (98.6)
Gestational age (weeks), mean (SD)	27.1 (2.1)***†††	31.8 (1.9)†††	40.0 (1.1)
Birthweight (g), mean (SD)	1029 (415)***†††	1817 (493)†††	3528 (518)
Birthweight (z-score), mean (SD)	-0.06 (1.29)	0.17 (1.38)	0.08 (0.97)
Antenatal Steroids, n(%)	28 (84.8)†††§	94 (88.7)†††§	0 (0)
Intrauterine growth restriction, n(%)	8 (23.5)††	18 (16.4)†	4 (5.7)
Age at testing (years), mean (SD)	9.9 (1.4)	10.3 (1.1)††	9.7 (1.1)
Weight (kg), mean (SD)	36.3 (13.1)	37.7 (9.0)	36.6 (10.5)
Weight (z-score), mean (SD)	0.08 (1.54)	0.31 (1.02)	0.46 (1.02)
Body Mass Index (kg/m^2), mean (SD)	18.1 (4.1)	18.0 (3.1)	17.9 (3.2)
Body Mass Index (z-score), mean (SD)	0.14 (1.52)	0.14 (1.24)	0.30 (1.08)
Asthma diagnosis, n(%)	9 (36.0)	22 (20.0)	5 (7.1)**
WIMD 2019 Rank, mean (SD)	1019 (507)	1052 (545)	1178 (520)
FEV_1 (%predicted), mean (SD)	86.9 (15.9)*†††	93.2 (14.1)†††	104.3 (7.1)
FVC (%predicted), mean (SD)	99.2 (10.7)††	99.1 (11.6)†††	107.6 (8.8)
FEV_1/FVC, mean (SD)	0.77 (0.10)**†††	0.82 (0.09)	0.85 (0.06)
$FEF_{25-75\%}$ (%predicted), mean (SD)	64.0 (25.8)**†††	79.3 (24.6)†††	94.7 (19.1)
Total volume of EBC collected (ml), mean (SD)	1.1 (0.3)†	1.1 (0.4)	1.3 (0.3)

Table 3-1: Participant demographics for samples used in EBC metabolomics analysis

Comparisons by ANOVA with Bonferroni correction/Chi-squared test as appropriate at baseline.
*$p<0.05$, **$p<0.01$, ***$p<0.001$ compared to Preterm No BPD
†$p<0.05$, ††$p<0.01$, †††$p<0.001$ compared to Term-born
§AN Steroids data missing for 5 cases (4 Preterm-born No BPD, 1 Preterm-born BPD)
WIMD: Welsh Index of Multiple Deprivation 2019 Rank Scores
BPD: Bronchopulmonary Dysplasia; FEV_1: Forced Expiratory Volume in 1 second; FVC: Forced Vital Capacity

3.3.1 Detected metabolites:

For the analysis I used a total of 235 fully annotated metabolites related to primary metabolism (carbohydrates and sugar phosphates, amino acids, hydroxyl acids, free fatty acids, purines, pyrimidines, aromatics, and exposome-derived chemicals) that had been successfully detected and identified from the BinBase database by West Coast Metabolomics Centre. 128 (54.5%) of these metabolites had a mean and median peak intensity greater than the limit of detection, and 38 (16.2%) metabolites were detected above the limit of detection in every sample analysed. Details of all detected metabolites and the number of samples in which they were present are given in the Appendix in Table 8-2. Overall, the metabolite content of EBC was relatively low, with several metabolites close to the limit of detection in multiple samples.

3.3.2 Metabolomic differences between BPD and No BPD groups:

Significant log2FC differences were noted between BPD and No BPD groups for ten metabolites (Figure 3-1,

Table 3-2). Alanine was reduced in the BPD group (log$_2$FC -1.71, p=0.025) and octadecanol increased (0.17, 0.026), with both metabolites detected in every sample. Urea (-2.52, 0.012), pyroglutamic acid (-1.78, 0.012), valine (-1.98, 0.014), ornithine (-2.69, 0.033) and serine (-2.62, 0.035) were all detected in >98% of samples, all with a significantly lower quantity in the BPD group. MSEA (Table 3-3) linked alanine, ornithine and urea with a significant alteration of urea cycle metabolism (p=<0.001) and alanine and pyroglutamic acid with an alteration of glutathione metabolism (p=0.008) (Figure 3-2). Ornithine and urea were also significantly linked with an alteration of arginine and proline metabolism (p=0.047).

Univariable, unadjusted, linear regression models of the preterm-born cohort studying demographic and lung function characteristics identified that alanine (beta -0.18, p=0.025) and urea (-0.29, 0.013) were only significantly associated with a history of BPD (Table 3-4). In univariable linear regression models for pyroglutamic acid and ornithine, female sex and history of BPD were both significantly related to a reduced quantity of these metabolites. When combined into a multivariable linear regression model, BPD remained significantly associated with a reduced quantity of pyroglutamic acid (beta -0.24, p=0.016) and ornithine (-0.24, 0.039) (Table 3-4). No significant associations were noted between these metabolites and current lung function in univariable linear regression models with these metabolites of interest. There were minimal significantly altered metabolites when comparing the mild and moderate/severe BPD groups (Figure 3-3).

3.3.3   Metabolomic differences between BPD and Term groups:

Significant log$_2$FC were observed between BPD and Term groups for 14 metabolites (Figure 3-1,

Table 3-2). As in the preterm-born cohort, significantly reduced quantities of valine (log$_2$FC -1.33, p=0.006), alanine (-0.94, 0.017), serine (-1.2, 0.039), pyroglutamic acid (-0.93, 0.039) and urea (-4.43, 0.043) were seen in the BPD group when compared to term-born children. Glycine was detected in every sample, again with a significantly decreased quantity in the BPD group (-0.45, 0.031). Oleamide was detected in >90% of samples, with a significantly reduced quantity in the BPD group (-0.2, 0.034). MSEA (Table 3-3) linked alanine, glycine and pyroglutamic acid with a significant alteration of glutathione metabolism (p=<0.001), and alanine, glutamic acid and urea with a significant alteration of urea cycle metabolism (p=0.002) (Figure 3-2). There was also a significant alteration of Glucose-Alanine cycle (p=<0.001) and Alanine metabolism (p=<0.001); however, glutamic acid was implicated in both processes and was detected in <20% of samples.
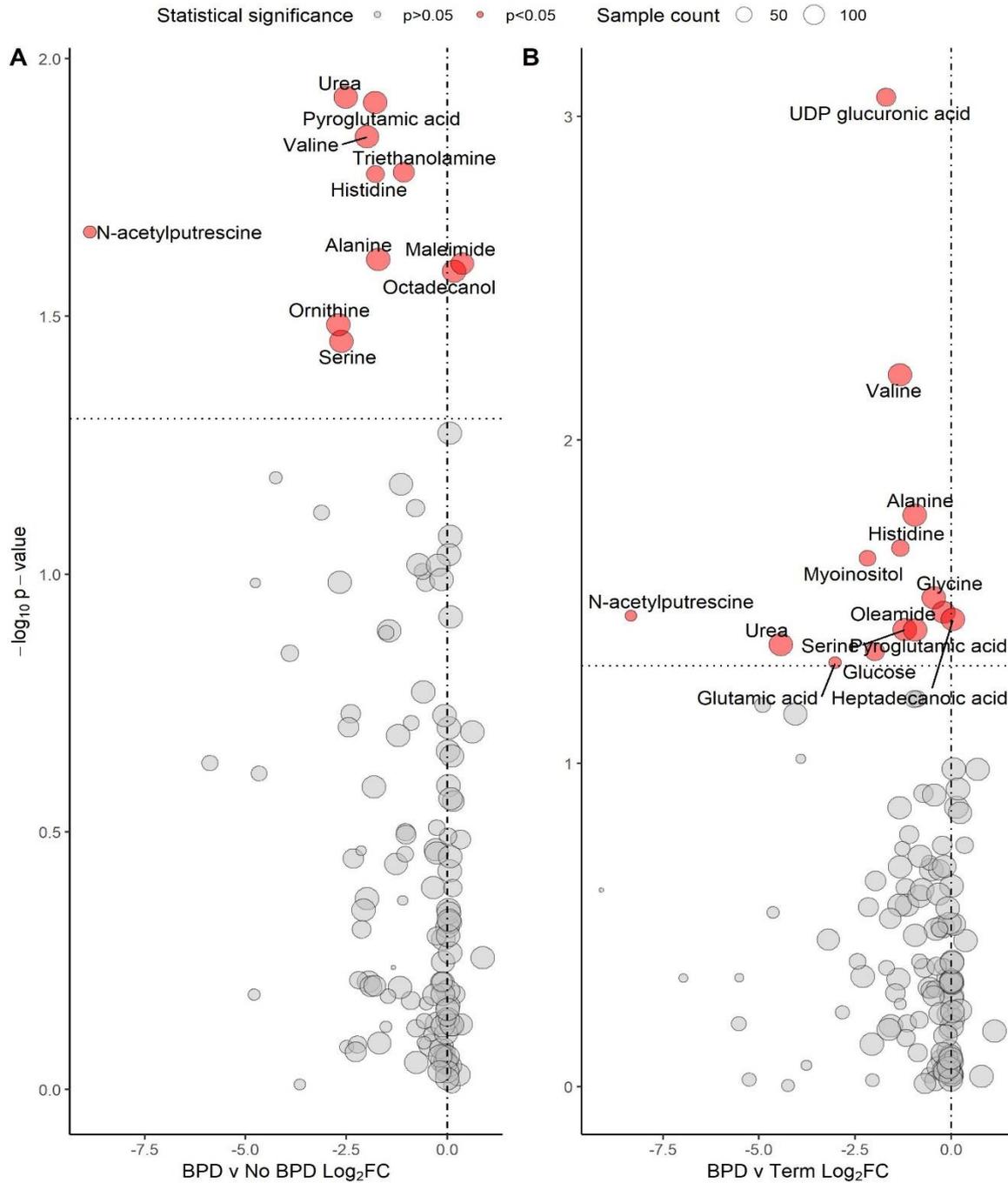
Figure 3-1 Volcano plots demonstrating significantly altered EBC metabolites between preterm-born infants with BPD and control groups

(A) BPD vs No BPD (B) BPD vs Term**.** Vertical line represents a Log$_2$FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to number of samples in which metabolite was detected. Metabolite name given if p<0.05. BPD: Bronchopulmonary dysplasia; Log$_2$FC: Log$_2$ fold-change between groups.

Metabolite	Retention Index	m/z	PubChem ID	% of samples	Fold Change	log$_2$FC	p value
BPD vs No BPD							
Urea	323728	189	1176	99.3	0.17	-2.52	0.012
Pyroglutamic acid	485935	156	7405	99.3	0.29	-1.78	0.012
Valine	313502	144	6287	99.3	0.25	-1.98	0.014
Triethanolamine	531892	262	7618	73.6	0.47	-1.08	0.017
Histidine	663790	154	6274	49.3	0.29	-1.78	0.017
Alanine	244189	116	5950	100	0.31	-1.71	0.025
Maleimide	245118	154	10935	89.9	1.30	0.37	0.025
Octadecanol	755409	327	8221	100	1.13	0.17	0.026
Ornithine	619196	142	88747248	99.3	0.15	-2.69	0.033
Serine	395020	204	5951	98.6	0.16	-2.62	0.035
BPD vs Term							
UDP-glucuronic acid	585473	217	17473	60.6	0.31	-1.68	0.0009
Valine	313502	144	6287	99.0	0.40	-1.33	0.006
Alanine	244189	116	5950	100	0.52	-0.94	0.017
Histidine	663790	154	6274	48.1	0.40	-1.31	0.022
Myoinositol	730022	305	892	42.3	0.22	-2.17	0.023
Glycine	368707	248	750	100	0.73	-0.45	0.031
Oleamide	849710	144	5283387	92.3	0.87	-0.20	0.034
N-acetylputrescine	595523	174	122356	16.3	0.003	-8.33	0.035
Heptadecanoic acid	751309	117	10465	100	1.04	0.06	0.036
Serine	395020	204	5951	99.0	0.43	-1.20	0.039
Pyroglutamic acid	485935	156	7405	99.0	0.52	-0.93	0.039
Urea	323728	189	1176	99.0	0.05	-4.43	0.043
Glucose	659798	319	64689	59.6	0.25	-1.98	0.045
Glutamic acid	529100	246	33032	19.2	0.12	-3.01	0.049

Table 3-2: Significantly different metabolites detected in EBC between infants with BPD and control groups

BPD: Bronchopulmonary dysplasia. m/z: mass-to-charge ratio. Log$_2$FC: Log$_2$ fold change. p values represent between group comparisons using t-test.

Process	Enriched Metabolites	Enrichment Ratio	p value	FDR
BPD vs No BPD				
Urea Cycle	Alanine, Ornithine, Urea	15.2	0.0006	0.065
Glutathione Metabolism	Alanine, Pyroglutamic acid	13.9	0.008	0.39
Methylhistidine Metabolism	Histidine	36.6	0.027	0.89
Arginine and Proline Metabolism	Ornithine, Urea	5.5	0.047	1.0
BPD vs Term				
Glutathione Metabolism	Alanine, Glutamic Acid, Glycine, Pyroglutamic acid	19.5	0.00003	0.003
Glucose-Alanine Cycle	Alanine, Glucose, Glutamic acid	23.6	0.0002	0.009
Alanine Metabolism	Alanine, Glycine, Glutamic acid	18.1	0.0004	0.014
Urea Cycle	Alanine, Glutamic acid, Urea	10.6	0.002	0.05
Ammonia Recycling	Glycine, Glutamic acid, Histidine	9.6	0.003	0.06
Glutamate Metabolism	Alanine, Glycine, Glutamic acid	6.3	0.010	0.16
Arginine and Proline Metabolism	Glycine, Urea	5.8	0.012	0.17
Glycine and Serine Metabolism	Glycine, Glutamic acid, Urea	5.2	0.016	0.20
Methylhistidine Metabolism	Histidine	25.6	0.038	0.40
Beta-alanine Metabolism	Glutamic acid, Histidine	6.0	0.041	0.40

Table 3-3 Metabolite Set Enrichment Analysis demonstrating altered biological processes implicated by significantly altered EBC metabolite quantities in the BPD group.

Univariable Models											
Variable	**Pyroglutamic acid**			**Ornithine**			**Alanine**			**Urea**		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**
Sex (ref=Male)	-0.20	0.09	**0.020***	-0.22	0.10	**0.031***	-0.11	0.07	0.11	-0.14	0.10	0.16
Birthweight (z-score)	0.001	0.03	0.96	-0.0001	0.04	0.99	-0.03	0.03	0.26	-0.01	0.04	0.76
IUGR (Ref = No IUGR)	-0.09	0.11	0.45	-0.05	0.13	0.68	0.001	0.09	0.99	-0.03	0.13	0.81
BPD (ref=No BPD)	-0.25	0.10	**0.013***	-0.25	0.12	**0.033***	-0.18	0.08	**0.025***	-0.29	0.11	**0.013***
Age (years)	-0.03	0.04	0.48	-0.04	0.04	0.40	-0.03	0.03	0.29	-0.02	0.04	0.71
Weight (z-score)	0.05	0.04	0.17	0.04	0.04	0.40	0.02	0.03	0.53	0.03	0.04	0.42
BMI (z-score)	0.02	0.03	0.64	0.004	0.04	0.92	-0.001	0.03	0.96	0.01	0.04	0.75
Asthma (Ref = No Asthma)	-0.10	0.11	0.34	-0.12	0.12	0.34	0.03	0.08	0.70	0.01	0.12	0.93
FEV_1 (% predicted)	<0.001	0.003	0.89	<0.001	0.003	0.99	-0.001	0.002	0.85	0.001	0.003	0.81
FVC (% predicted)	<0.001	0.004	0.92	-0.001	0.004	0.73	-0.001	0.003	0.84	-0.002	0.004	0.53
FEV_1/FVC	-0.13	0.47	0.78	-0.08	0.55	0.89	-0.16	0.37	0.66	0.35	0.54	0.52
$FEF_{25-75\%}$ (%predicted)	0.001	0.002	0.42	0.001	0.002	0.75	<0.001	0.001	0.92	0.002	0.002	0.34
EBC volume collected (ml)	0.02	0.11	0.88	0.04	0.14	0.77	-0.03	0.10	0.80	0.12	0.13	0.36
Multivariable Models											
Sex (ref=Male)	-0.20	0.08	**0.021***	-0.21	0.10	**0.038***						
BPD (ref=No BPD)	-0.24	0.10	**0.016***	-0.24	0.12	**0.039***						

Table 3-4: Linear regression models demonstrating associations between EBC metabolites of interest and early- and current-life factors

Univariable regression models where p<0.1 included in multivariable model. Multivariable models not created where only one variable has a p<0.1 in univariable models.

IUGR: Intrauterine growth restriction, BPD: Bronchopulmonary Dysplasia, BMI: Body Mass Index, FEV_1: Forced Expiratory Volume in 1 second, FVC: Forced Vital Capacity, EBC: Exhaled Breath Condensate

**denotes p<0.05
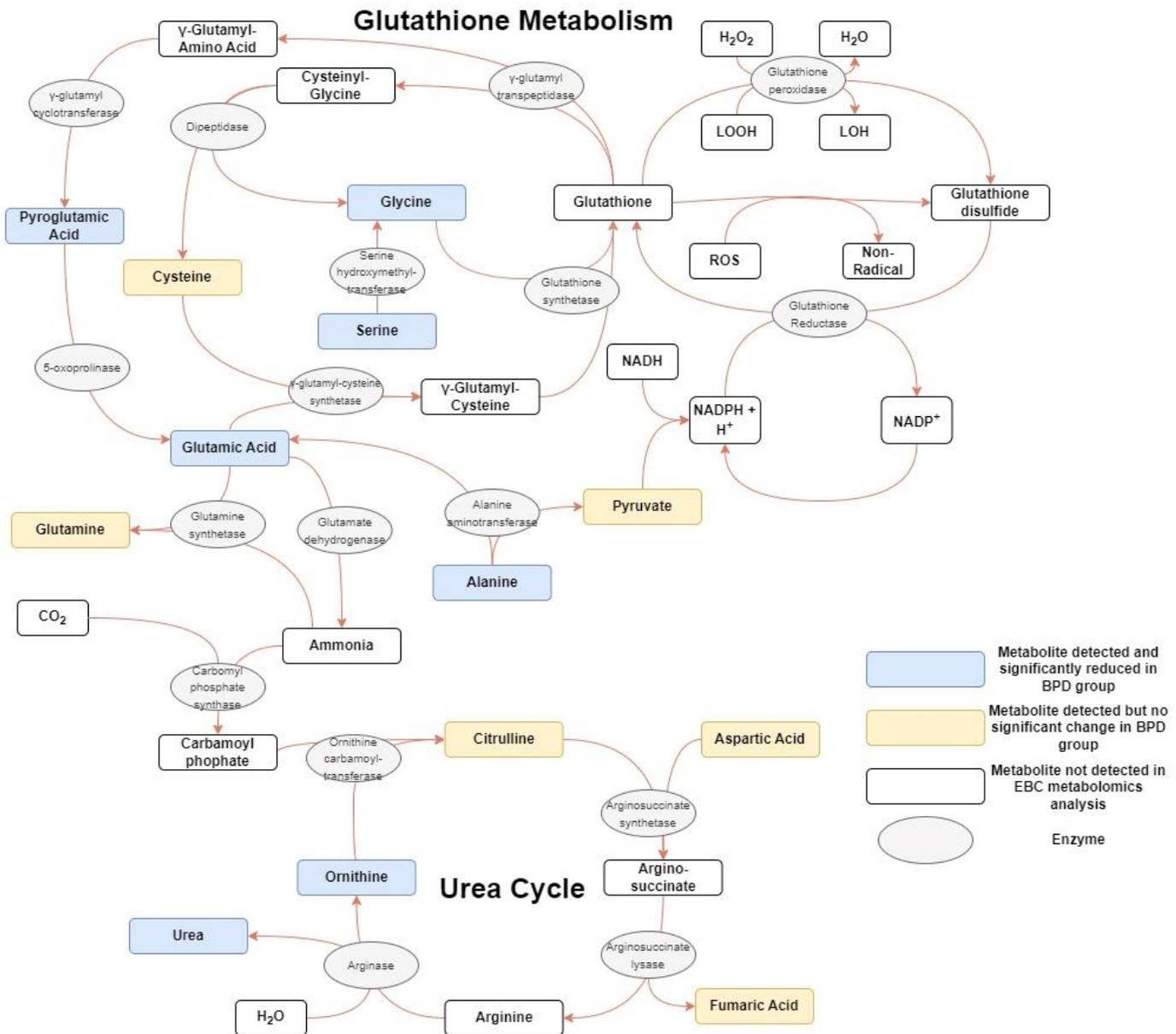
Figure 3-2: Graphic representation of glutathione metabolism and the urea cycle, highlighting metabolites detected in this analysis and those with a significantly reduced concentration in the BPD group.

LOOH: Lipid hydroperoxide; LOH: Lipid hydroxide; ROS: Reactive oxygen species; NADP: Nicotinamide adenine dinucleotide phosphate; NADPH: nicotinamide adenine dinucleotide phosphate hydrogen; NADH: nicotinamide adenine dinucleotide + hydrogen.
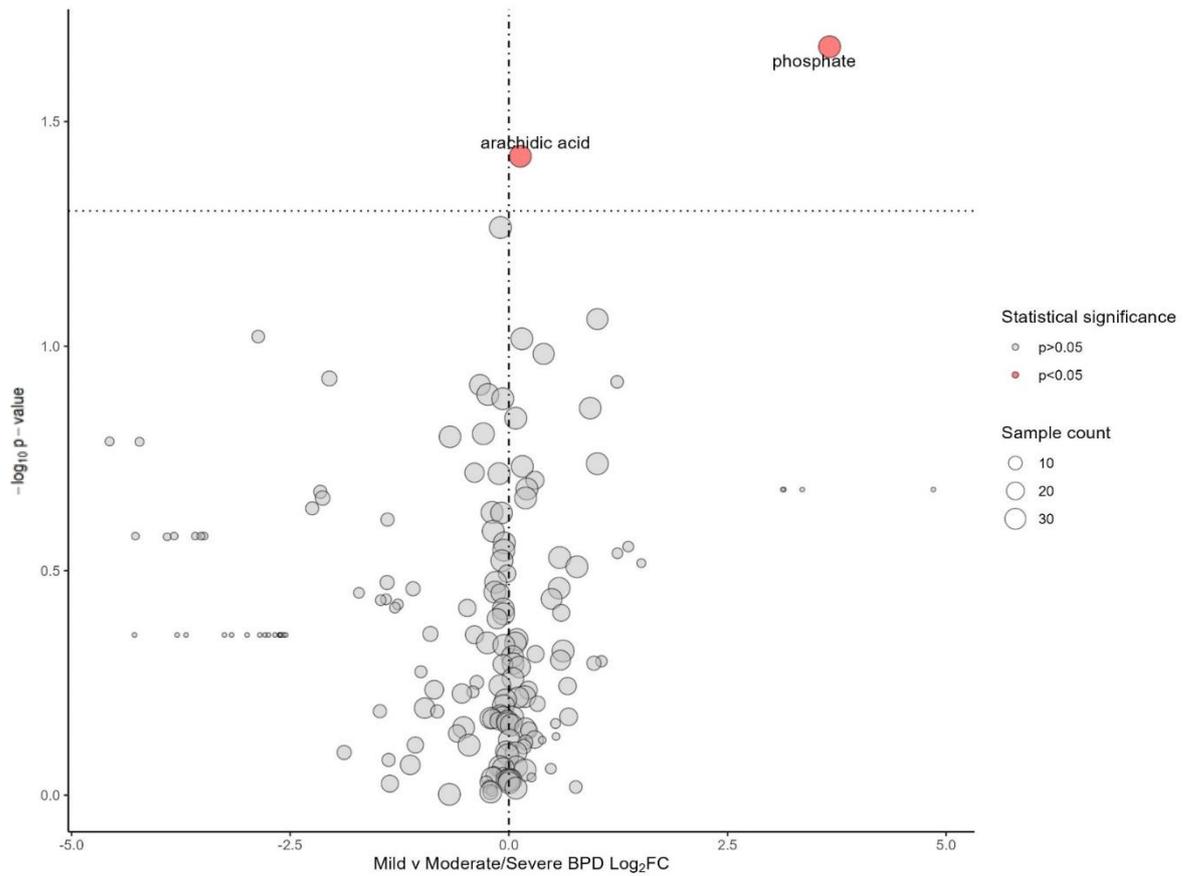
Figure 3-3: Volcano plot showing significantly altered metabolites between the mild and moderate/severe BPD groups.

Vertical line represents a Log_2FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to number of samples in which metabolite was detected. Metabolite name given if $p<0.05$. BPD: Bronchopulmonary dysplasia; Log_2FC: Log_2 fold-change between groups.
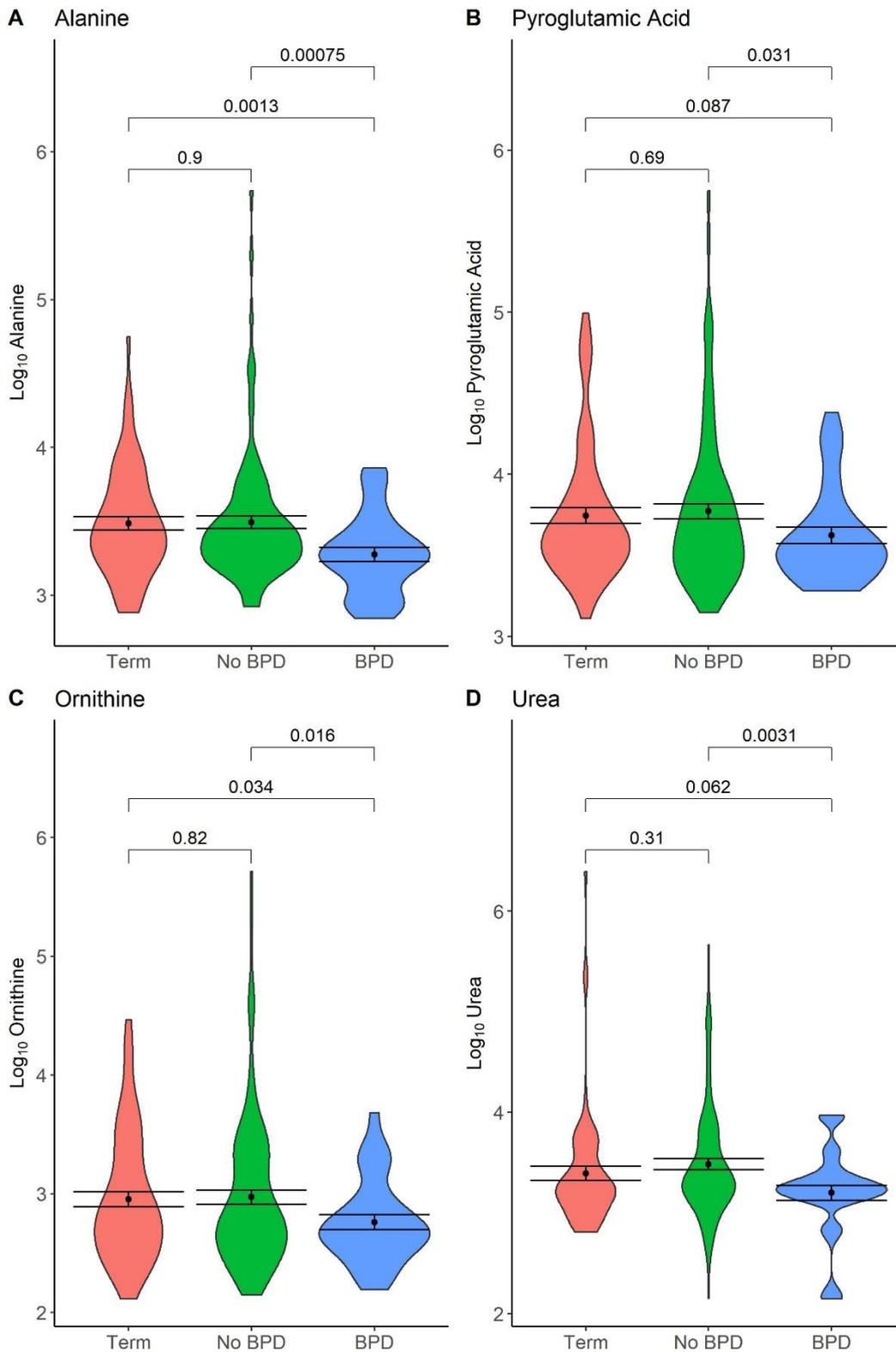
Figure 3-4: Violin Plots of Significantly Altered Metabolites of interest detected in the EBC.

Black dot and bars give mean and standard error of the mean (SEM). Bars give p-values from

ANOVA with post-hoc Bonferroni correction for between group comparisons

Figure 3-4 shows results for ANVOA with post-hoc Bonferroni comparisons between BPD, No BPD and Term groups for alanine, pyroglutamic acid, ornithine and urea. All four metabolites showed a consistent trend of the lowest quantities being present in the BPD group, with both alanine and ornithine having a significantly lower quantity in the BPD group when compared to both the No BPD ($p=<0.001$, $p=0.016$ respectively) and Term (0.0013, 0.034 respectively) groups. Pyroglutamic acid and urea had a significantly lower quantity in the BPD group compared to No BPD (0.031, 0.0031 respectively), with a near-significant difference when compared to the Term group (0.087, 0.062 respectively).

3.3.4 Comparison of EBC metabolome before and after inhaled therapies

Paired pre- and post-treatment EBC samples were available for 40 RCT participants, 14 in both the ICS and ICS/LABA groups and 12 in the placebo group. Significantly altered metabolites within each treatment group are detailed in Figure 3-5 and Table 3-5. Four metabolites were significantly altered by ICS treatment, three of which (3-hydroxypropionic acid, caprylic acid and caprylic acid) were detected in every sample analysed, and all three were significantly higher quantity in the pre-treatment samples.

Six metabolites were significantly altered by ICS/LABA treatment, five of which (butane-2,3-diol, 1-monostearin, tyrosine, hydroxycarbamate and 2-hydroxypyrazinyl-2-propenoic acidethylester) were detected in every sample analysed. Butane-2,3-diol, 1-monostearin and tyrosine were all significantly higher quantity in the pre-treatment samples, with hydroxycarbamate and 2-hydroxypyrazinyl-2-propenoic acidethylester having a significantly lower quantity pre-treatment. Four metabolites were significantly altered in the placebo group, with only one metabolite, 2-ketoisocaproic acid, being detected in every sample analysed and having a significantly lower quantity pre-treatment. MSEA did not reveal any significantly enriched metabolic processes in any of the three treatment groups.
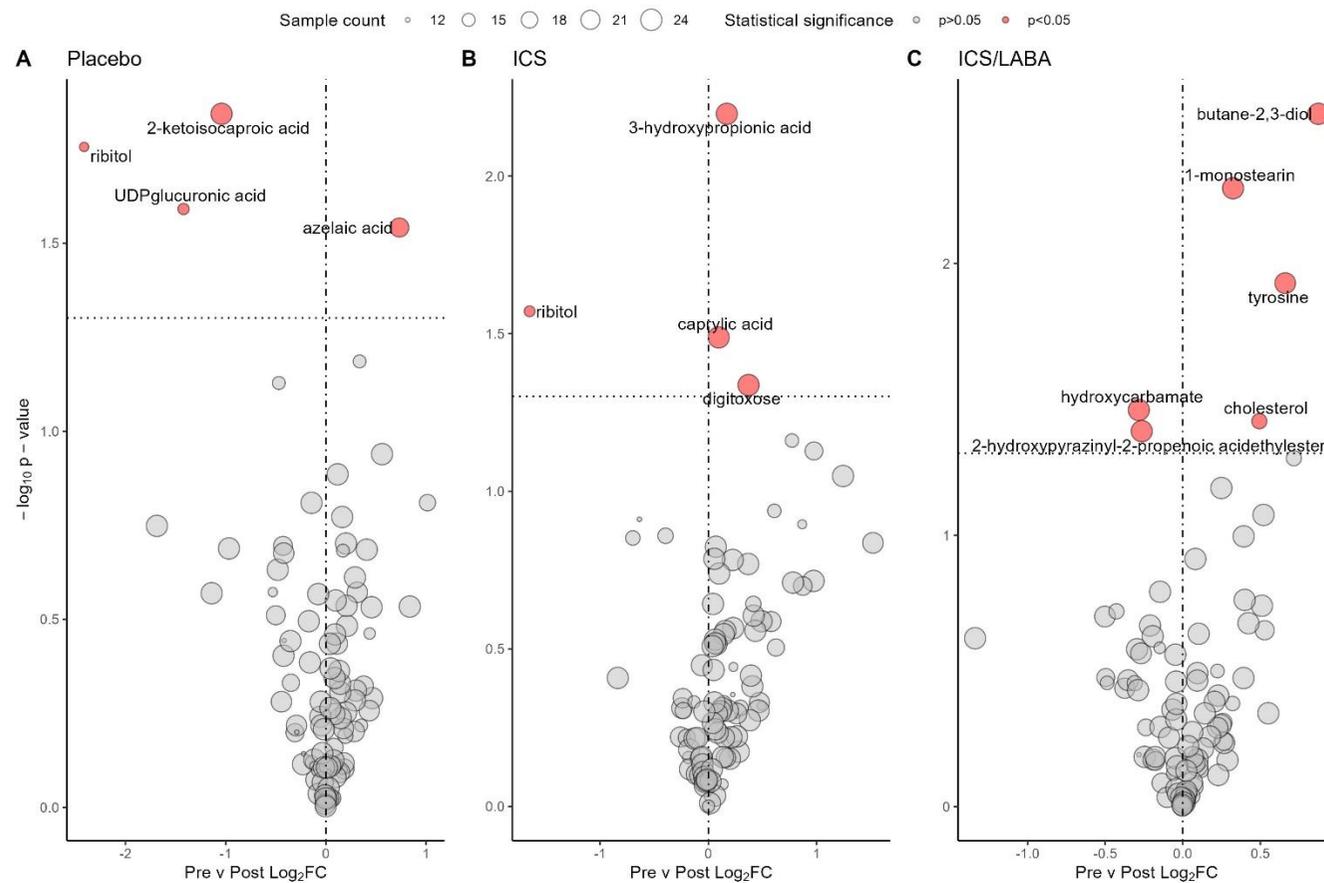
Figure 3-5: Volcano plots comparing EBC metabolome before to after treatment in the three RCT inhaler groups.

Vertical line represents a Log_2FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to number of samples in which metabolite was detected. Metabolite name given if p<0.05. ICS: Inhaled corticosteroid. LABA: Long-acting β2 agonist; Log_2FC: Log_2 fold-change between groups

Metabolite	Retention Index	m/z	PubChem ID	% of participants	Fold Change	log$_2$FC	p value
Inhaled Corticosteroid (n=14)							
3-hydroxypropionic acid	269265	177	68152	100	1.13	0.17	0.006
ribitol	575497	217	827	57.1	0.32	-1.65	0.027
caprylic acid	343457	201	379	100	1.07	0.10	0.032
digitoxose	521798	117	94168	100	1.29	0.37	0.046
Inhaled Corticosteroid/Long-Acting ß$_2$-agonist (n=14)							
butane-2,3-diol	205778	117	262	100	1.83	0.87	0.003
1-monostearin	959214	203	24699	100	1.25	0.32	0.005
tyrosine	671252	218	6057	100	1.58	0.66	0.012
hydroxycarbamate	325318	278	16639161	100	0.82	-0.28	0.035
cholesterol	1078536	129	5997	64.3	1.41	0.49	0.038
2-hydroxypyrazinyl-2-propenoic acidethylester	493127	121	5371086	100	0.83	-0.26	0.041
Placebo (n=12)							
2-ketoisocaproic acid	290473	89	70	100	0.49	-1.04	0.014
ribitol	575497	217	827	58.3	0.19	-2.41	0.018
UDP-glucuronic acid	585473	217	17473	58.3	0.37	-1.42	0.026
azelaic acid	610551	317	19347555	91.7	1.66	0.73	0.029

Table 3-5: Significantly altered EBC metabolites in the three RCT treatment groups.

m/z: mass-to-charge ratio. Log$_2$FC: Log$_2$ fold change. p values represent between group comparisons using a paired samples t-test.

3.4 Discussion:

In this exploratory analysis of the EBC metabolome of preterm-born school-aged children, I have demonstrated significant differences in several metabolites from those with a history of BPD. On comparison to both preterm-born and term-born controls, levels of alanine, pyroglutamic acid, serine, urea, and valine were all significantly lower in the BPD group. These five metabolites were detected in >98% of samples. Alanine and pyroglutamic acid were significantly associated with an alteration of glutathione metabolism. Alanine and urea were significantly associated with an alteration in the urea cycle (with ornithine also being associated when compared to preterm-born controls). Linear regression analyses demonstrated that alanine, pyroglutamic acid, ornithine and urea remained significantly associated with BPD in the preterm-born group when other participant characteristics were also considered. However, linear regression analyses did not show a significant association between any of these metabolites and current lung function parameters. Whilst there were changes in metabolite levels in the three RCT treatment groups, with the ICS/LABA group showing the largest number of altered metabolites, these did not map to any significantly altered metabolic processes.

Pyroglutamic acid, also known as 5-oxoproline, is an intermediary in glutathione synthesis and recycling. Glutathione is a potent antioxidant, and under conditions of oxidative stress, where glutathione is consumed, pyroglutamic acid levels also become low (Kumar and Bachhawat, 2012). Alanine concentration also appears to be reduced in metabolomic studies of murine models of pulmonary inflammation, along with pyroglutamic acid (Ambruso et al., 2021). Alanine is a non-essential amino acid that is a constituent of nearly all proteins. Whilst it is not a direct precursor to glutathione, alanine can be converted to pyruvate, a key intermediate of glucose metabolism (Brosnan, 2003).

Glucose metabolism is an important source of reducing substances, such as NADPH, which are essential in glutathione synthesis (Wu et al., 2004, Meister and Anderson, 1983). Alanine can also be converted to other amino acids, such as serine, which is a precursor to glycine (Meister and Anderson, 1983). Glycine is a key, rate-limiting amino acid for glutathione synthesis. Glutathione consumption increases metabolism of glycine, as well as inflammatory conditions reducing glycine availability (Wu et al., 2004, Meister and Anderson, 1983). I noted a significant decrease in serine in our BPD group when compared with the No BPD group, as well as a significant reduction in glycine when compared to the Term group. Taken together, the metabolomic differences observed in the BPD group suggest decreased glutathione levels, and thereby suggesting persistent oxidative stress, in the airways of preterm-born children with a history of BPD.

Glutathione has previously been shown to provide first-line defense against pulmonary oxidative injury. Adult studies have shown that glutathione concentrations in the airway's ELF are many times greater than that seen in plasma (Cantin et al., 1987), and animal models have shown that pulmonary glutathione depletion enhances oxidant toxicity (Deneke et al., 1985). In the paediatric population, alterations of glutathione metabolism have been linked with respiratory pathology. A study of ELF in children with severe asthma reported significantly decreased concentration of glutathione, with evidence of glutathione consumption by oxidative stress, further supported by increased levels of hydrogen peroxide (H_2O_2), a potent oxidant. However, there was no significant association between markers of impaired glutathione metabolism and FEV_1 (Fitzpatrick et al., 2009). Impaired glutathione metabolism has also been associated with impaired macrophage function in the airways of children with severe asthma, with glutathione supplementation restoring macrophage

111

function (Fitzpatrick et al., 2011). Reduced concentration of glutathione secondary to oxidative stress has also been described in bronchoalveolar lavage fluid (BAL) obtained from children with CF, with those experiencing an infective exacerbation having a further decreased concentration (Kettle et al., 2014).

Previous mechanistic studies of BPD in later childhood and adulthood have implicated cytotoxic CD8+ T-lymphocytes (Um-Bergström et al., 2022), elevated neutrophils and pro-inflammatory cytokines (Teig et al., 2012), and thickened basement membranes and airway remodeling (Galderisi et al., 2019). However, to my knowledge, this is the first time the EBC metabolome of preterm-born children with a history of BPD has been studied. Previous studies in preterm infants in the neonatal period have linked increases in pulmonary oxidative stress and glutathione metabolism to the development of BPD. In a small study of BAL from preterm infants born at <34 week's gestation, lower glutathione levels on the first day of life were associated with increased development of BPD at 36 weeks' gestation (Grigg et al., 1993). This study was performed before the routine use of surfactant replacement therapy. A further study in the post-surfactant era of infants born at <32 weeks' gestation reported lower BAL glutathione levels in the first 24 hours of life in those who later developed BPD, lower glutathione levels in those who had delayed surfactant administration, and higher concentrations of malondialdehyde, suggestive of oxidative damage, in the BPD group (Collard et al., 2004). Animal models have further supported the role of glutathione in lung injury, with glutathione deficient mice having impaired tolerance of oxidative stress, and abnormal early lung development (Robbins et al., 2021). Whilst glutathione metabolism has been implicated in our study by MSEA, we did not detect glutathione in either its reduced or oxidized form. However, glutathione has a short half-life of approximately ten minutes, which can make it's detection challenging (Hong et al., 2005). One previous study has examined

oxidative stress in the airways of preterm-born adolescents, measuring 8-isoprostane (a product of lipid peroxidation in the presence of oxygen free radicals) in EBC, both with and without BPD. These individuals, born during the peri-surfactant era, demonstrated evidence of persistent airway oxidative stress, with increased 8-isoprostane when compared to term-born controls (Filippone et al., 2012). However, in this study no significant difference was observed for 8-isoprostane between BPD and No BPD groups, and no correlation was seen between 8-isoprostane levels and spirometry values.

Previous metabolomic studies of BPD, with similar analytical techniques to our study, have focused on preterm-born infants in the neonatal period. A study of tracheal aspirates taken from infants born <30 weeks' gestation using metabolomic techniques reported that 19 metabolites discriminated infants who subsequently did or did not develop BPD (Piersigilli et al., 2019), including alterations in amino acids (citrulline and symmetric dimehtlyarginine) involved in nitric oxide metabolism, as well as an increase in serine. The authors also observed increased acylcarnitines which are released after β-oxidation of fatty acids during oxidative stress. In contrast, another metabolomic study of tracheal aspirates from infants ≤28 weeks' gestation noted that early decreases in fatty-acid metabolism, particularly the fatty acid β-oxidation pathway, may predispose infants to developing BPD (Lal et al., 2018). A nuclear magnetic resonance metabolomics study of urine from 18 infants born <28 weeks' gestation showed decreased glycine levels in those who subsequently developed BPD, similarly to this analysis, also suggesting impaired glutathione metabolism. The authors also found increased alanine in the BPD group, which they attributed to increased cellular metabolism demands, due to an inflammatory process (Pintus et al., 2018). These authors studied infants in the first week of life, where the pathophysiology is respiratory distress syndrome and pulmonary

surfactant deficiency, as opposed to the chronic inflammation seen in BPD, potentially explaining the different alanine levels I observed in this analysis.

I observed significant decreases in alanine, ornithine, and urea in the BPD group, with MSEA linking these changes to a significant decrease in urea cycle metabolism. The urea cycle removes ammonia, produced during protein catabolism, preventing cellular toxicity. Animal models have demonstrated increased ammonia levels lead to intracellular production of reactive oxygen species, induction of cellular apoptosis in bovine epithelial cells, increased inflammatory cytokines and repression of DNA repair-related genes in porcine Type II alveolar epithelial cells (Li et al., 2023, Wang et al., 2018). I also observed a significant reduction in histidine in our BPD group when compared to both the No BPD and Term groups. Histidine, an essential amino acid, is metabolised to histamine and can affect the contractility of bronchial smooth muscle and cause airway oedema (Yamauchi and Ogasawara, 2019). Histidine itself has also been reported to have antioxidant properties, being a scavenger of free hydroxyl and singlet oxygen radicals and inhibiting fatty-acid oxidation during in vitro studies (Wade and Tucker, 1998). Lower quantities of urea and ornithine in the BPD group were also significantly linked to arginine and proline metabolism. Arginine is metabolised into either nitric oxide by nitric oxide synthetase (NOS) or into urea and ornithine by arginase. Increased arginase activity is thought to play a role in childhood asthma pathogenesis, with a consequent increase in proline production leading to collagen deposition(Salam et al., 2009). Whilst elevated FE_{NO} is used clinically as a biomarker of asthma (Menzies-Gow et al., 2020), methylated products of arginine can inhibit NOS activity and contribute to airway oxidative stress in specific phenotypes of asthma associated with obesity (Holguin, 2013). The reduced quantities of metabolites linked to protection from reactive oxidant species in those with BPD implied by my results suggest a persistent deficit since the neonatal period, especially for

glutathione, however, this is speculative. I did not observe any associations between metabolic processes involved with oxidative stress and current lung function. This may be due to other processes such as functional and structural abnormalities also significantly contributing to the development of PLD as I have published previously and discussed elsewhere in this thesis (Course et al., 2023b, Course et al., 2023c) (Chapters 2 and 4). I also did not observe a consistent pattern of metabolite differences in the three RCT treatment groups of preterm-born children with low lung function. Despite the ICS/LABA group showing the largest number of altered metabolites, as well as this group having a significant improvement in lung function following treatment (as described in section 1.6.3.4), MSEA was unable to implicate a biological process. Butane-2,3-diol levels were elevated pre-treatment in the ICS/LABA group, and this metabolite has been associated with accentuated pulmonary inflammation in a small study of a rodent model of CF with pseudomonas colonization (Nguyen et al., 2016). How this metabolite relates to PLD pathogenesis and treatment requires further evaluation. One previous study using superoxide dismutase did not decrease rates of BPD in the neonatal period but was associated with decreased respiratory symptoms at one year of age (Suresh et al., 2001). Whether such treatments are beneficial for PLD will require further study. In addition, longitudinal metabolome analysis beginning in the neonatal period or infancy may reveal emerging mechanisms related to lung function decline.

3.4.1 Strengths and Limitations

This analysis represents one of the largest metabolomic studies of a clinical cohort. The significantly altered metabolites of interest in the BPD group were detected in all or nearly all samples analyzed, ensuring robust results. Using an untargeted metabolomics approach, I have been able to identify patterns of changes in multiple metabolites which I have been able

to link with biological processes. I have studied a cohort of children who had experienced contemporary standards of neonatal care, and by using EBC, I have been able to directly sample ELF, which is representative of airway biochemistry in a simple and well-tolerated manner. Although EBC volumes collected varied between the study groups, linear regression analyses did not reveal a significant association between metabolite quantities and EBC volume. As the overall metabolite content of EBC was low, as has been previously reported (Aksenov et al., 2017), there may have been metabolites present that were below the limit of detection for the mass spectrometry method used. I combined the mild and moderate/severe BPD groups as few differences were noted for metabolites between these two groups thus it was reasonable to combine these two groups to improve statistical power to detect biologically relevant differences. Dietary intake has recently been shown to modulate the breath metabolome (Neyrinck et al., 2022), however, I had insufficient nutritional data to adjust for this potential confounder. Similarly, whilst data was initially collected on antenatal and household smoking, overall rates in the preterm-born cohort were low suggesting high recall bias, and therefore were not included in my analyses. Whilst a reference metabolite would be useful to normalize metabolite concentrations, as in urine metabolomics (Li et al., 2022), this is currently not possible with EBC samples (Horvath et al., 2017). Ideally, these findings should be replicated using a validation cohort, however this is limited by the number of large contemporaneous cohorts available to study.

3.5 Conclusions

In conclusion, in this chapter, I have presented results from the exploratory mass spectrometry-based analysis of the EBC metabolome, revealing significant reductions in metabolites related to antioxidant pathways in the airways of preterm-born school-aged children with a history of BPD, many years after the initial pulmonary insult. However, these

changes did not appear to be associated with current lung function, nor show any alteration

following inhaled therapies.

4 Proteomic Analysis of Urine

4.1 Introduction:

Historically, efforts to understand the mechanisms underlying PLD focussed on those with BPD, with evidence of smooth muscle extension into the distal airways in post-mortem samples from infants (Bush et al., 1990), and peri-bronchial fibrosis and CD8+ T-lymphocyte epithelial infiltrate in adolescent (Galderisi et al., 2019) and adult (Um-Bergström et al., 2022) survivors of BPD. However, as previously discussed in sections 1.4.3 and 1.4.4, it is increasingly recognised that many preterm-born survivors with BPD do not develop later lung disease, and infants born at later gestations (33-34 weeks), who are at low-risk for BPD, still develop PLD (Hart et al., 2022). Work by other authors on the RHiNO cohort (as described in section 1.6.3) has demonstrated a range of PLD phenotypes, with obstructive (both fixed and reversible), preserved ratio-impaired spirometry (PRISm) and dysanaptic low lung function patterns (Cousins et al., 2023). There is evidence that those individuals with BPD risk developing an obstructive PLD phenotype in childhood (Cousins et al., 2023), and there is also increasing evidence that PLD is likely to progress to an earlier onset of chronic obstructive pulmonary disease (COPD) in adulthood (Bolton et al., 2015, Gibbons et al., 2023). PRISm phenotypes have been associated with an increased risk of developing COPD, cardiovascular disease, and all-cause mortality in middle- and old-aged adult populations (Wan et al., 2021, Higbee et al., 2022), but little is known about this phenotype in paediatric populations.

As discussed in section 1.6.3.4, a proportion of those with PLD will respond to inhaled therapies (Goulden et al., 2021), however, a clearer understanding of the biological pathways underlying these PLD-associated phenotypes will aid their identification and the ability to target therapeutic interventions.

4.1.1 Urine Proteomics in Respiratory Disease

Urine proteomics is emerging as a tool for understanding respiratory diseases and identifying biomarkers (Martelo-Vidal et al., 2022) with the advantage that it can be sampled easily and non-invasively. A study of the urinary proteome in 42 neonates born extremely preterm (<29 week's gestation) identified 94 proteins significantly altered in those who developed BPD, with biological process mapping implicating a reduction in myeloid leukocyte activation and neutrophil degranulation (Ahmed et al., 2022). A label-free proteomic study found distinct urinary proteome compositions for preterm infants with infectious (congenital pneumonia) and non-infectious (respiratory distress syndrome, transient tachypnoea of the newborn) respiratory pathologies (Starodubtseva et al., 2016), with a majority of these proteins being linked to inflammatory processes.

Urine lacks the same homeostatic mechanisms as blood, meaning that systemic changes can accumulate and the urinary proteome may show alterations prior to clinical manifestations, serum proteome changes or histopathological changes to the lung tissue, reflecting an earlier stage of disease (Wu and Gao, 2015). This makes urine analysis particularly attractive when studying the RHiNO cohort, as these children are likely at a pre-symptomatic stage of impaired lung function.

4.1.2 Aims

I, therefore, performed an exploratory analysis of the urinary proteome from the RHiNO cohort of preterm-born, school-aged children, with term-born matched controls, to elucidate the biological mechanisms underlying PLD phenotypes.

4.2 Methods:

4.2.1 Participants

This study was conducted on a cohort of children recruited to the Respiratory Health Outcomes in Neonates study (RHiNO, EudraCT: 2015-003712-20) which has been described previously in this thesis (section 1.6) and in the published literature (Hart et al., 2022, Goulden et al., 2021, Cousins et al., 2022, Cousins et al., 2023). Briefly, children from a previous study (Edwards et al., 2016) were supplemented with additional preterm-born children sourced from NHS Wales healthcare records and sent a respiratory and neurodevelopmental questionnaire if they were born ≤34 or ≥37 weeks' gestation and were aged 7-12 years. Children with significant congenital malformations, cardiopulmonary or neuromuscular disease were excluded. Ethical approval was obtained from the South-West Bristol Research Ethics Committee (15/SW/0289). Parents gave informed written consent and children provided assent. Recruitment took place prospectively between November 2016 and September 2019.

4.2.2 Lung function assessment

Responders underwent spirometry (Microloop, Care Fusion, UK), performed to ATS/ERS guidelines (Miller et al., 2005) and normalised using GLI references (Quanjer et al., 2012) by trained research nurses. Any respiratory medications were withheld prior to their assessment (short- and long-acting β2-agonists for 8- and 48-hours respectively; inhaled corticosteroids for 24 hours; and leukotriene receptor antagonists for 48 hours) and children were free of respiratory infections for at least three weeks prior to testing. Low lung function in preterm-born children (PLD) was defined as FEV_1 less than the lower limit of normal (LLN) as per GLI references (Quanjer et al., 2012). Those with PLD were further categorised according to the pattern of their respiratory impairment (as previously described in the RHiNO cohort in section 1.6.2.3 (Cousins et al., 2023)); Prematurity-associated Preserved Ratio of Impaired

Spirometry (pPRISm) defined as FEV_1<LLN and FEV_1/FVC≥LLN, and Prematurity-associated Obstructive Lung Disease (POLD) defined as an FEV_1<LLN with an FEV_1/FVC<LLN. Preterm-born control (PT_c) and term-born children had FEV_1≥LLN.

BPD was defined as oxygen-dependency of 28-days or greater for those born <32 weeks' gestation and at 56 days of age for those born ≥32 weeks' gestation (Ehrenkranz et al., 2005). Intrauterine growth restriction (IUGR) defined as birthweight <10th percentile adjusted for sex and gestation (LMSgrowth v2.77, Medical Research Council, UK). Neonatal history was corroborated with medical records.

4.2.3 Urine sample collection and analysis

Urine samples were obtained at the time of spirometry by trained research nurses (Gill Willets and Louise Yendell, members of the RHiNO study team) during the home visit, and by Dr Michael Cousins and Dr Kylie Hart during the clinical research facility visit. Samples were immediately placed on ice, and aliquoted and stored at -80°C as soon as possible on the day of collection until analysis.

4.2.3.1 TMT Labelling

I arranged for analysis of the urine samples at the University of Bristol Proteomics Facility. 190µl of urine was digested with trypsin (1.25µg trypsin; 37°C, overnight), labelled with Tandem Mass Tag (TMT) eleven plex reagents according to the manufacturer's protocol (Thermo Fisher Scientific, Loughborough, UK) and the labelled samples pooled. The pooled sample was desalted using a SepPak cartridge according to the manufacturer's instructions (Waters, Milford, Massachusetts, USA). Eluate from the SepPak cartridge was evaporated to dryness and resuspended in 1% formic acid prior to analysis by nano-LC MSMS using an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific).

4.2.3.2 Nano-LC Mass Spectrometry

The TMT-labelled pool was fractionated using an Ultimate 3000 nano-LC system in line with an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid peptides were resolved on a 250 mm × 75 μm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1min, 6-15% B over 58min, 15-32% B over 58min, 32-40% B over 5min, 40-90% B over 1min, held at 90% B for 6min and then reduced to 1% B over 1min) with a flow rate of 300 nl min^{-1}. The TMT-labelled pool underwent a further fractionation to try and maximise peptide yield. The second fractionation used the above methodology again with a different gradient protocol: 6 gradient segments (1-6% solvent B over 1min, 6-25% B over 118min, 25-40%B over 3min, 40-90%B over 1min, held at 90%B for 6min and then reduced to 1%B over 1min.) again with a flow rate of 300 nl min^{-1}. Solvent A was 0.1% formic acid, and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid for both fractionation processes. Peptides were ionized by nano-electrospray ionization at 2.0kV using a stainless-steel emitter with an internal diameter of 30μm (Thermo Scientific) and a capillary temperature of 300°C.

All spectra were acquired using an Orbitrap Fusion Lumos mass spectrometer controlled by Xcalibur 3.0 software (Thermo Scientific) and operated in data-dependent acquisition mode using an SPS-MS3 workflow. FTMS1 spectra were collected at a resolution of 120,000, with an automatic gain control (AGC) target of 200,000 and a maximum injection time of 50ms. Precursors were filtered with an intensity threshold of 5,000, according to charge state (to include charge states 2-7) and with monoisotopic peak determination set to Peptide.

Previously interrogated precursors were excluded using a dynamic window (60s±10ppm). The MS2 precursors were isolated with a quadrupole isolation window of 0.7m/z. ITMS2 spectra were collected with an AGC target of 10,000, maximum injection time of 70ms and CID collision energy of 35%.

For FTMS3 analysis, the Orbitrap was operated at 50,000 resolution with an AGC target of 50,000 and a maximum injection time of 105ms. Precursors were fragmented by high energy collision dissociation (HCD) at a normalised collision energy of 60% to ensure maximal TMT reporter ion yield. Synchronous Precursor Selection (SPS) was enabled to include up to 5 MS2 fragment ions in the FTMS3 scan. All mass spectrometry runs were performed consecutively on the mass spectrometer with blank runs in between to prevent carry over from one experiment to the next.

4.2.3.3 Data Analysis

The raw data files were processed and proteins quantified using Proteome Discoverer software v2.1 (Thermo Scientific) and searched against the UniProt Human database (downloaded October 2019: 150786 entries) using the SEQUEST HT algorithm by the University of Bristol Proteomics Facility. Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.6Da. Search criteria included oxidation of methionine (+15.995Da), acetylation of the protein N-terminus (+42.011Da) and Methionine loss plus acetylation of the protein N-terminus (-89.03Da) as variable modifications and carbamidomethylation of cysteine (+57.021Da) and the addition of the TMT mass tag (+229.163Da) to peptide N-termini and lysine as fixed modifications. Searches were performed with full tryptic digestion and a maximum of 2 missed cleavages were allowed.

The reverse database search option was enabled, and all data was filtered to satisfy false discovery rate (FDR) of 5%.

4.2.4   Data Normalisation:

An equal volume of urine (190µl) was analysed from each subject. I investigated whether there was variation in detected total protein abundance by the mass spectrometer within the samples. I used the summed total protein abundance detected for each sample, after scaling to pool samples from each mass spectrometer run, firstly by examining the range of total abundances detected in every sample (Figure 4-1) and then as to whether there were differences in total protein abundances detected based on any of the key clinical groups being studied (Figure 4-2). As can be seen Figure 4-1 panel A, there was a wide variation in total median protein abundance between samples with an 8.36 $\log_2$ fold change between the sample with lowest compared to the sample with the highest detected median protein abundance (range -0.15, 8.21). There was a statistically significant lower abundance of proteins as detected by the mass spectrometer in the preterm-born group compared to the term-born group, but not within the major preterm phenotypes (Figure 4-2). Additionally, a test analysis of the initial proteomic dataset also showed a strong skewing of $\log_2$-fold changes between relative individual protein abundances towards the term-born group (Figure 4-3 panel A). This may be potentially due to an intrinsic difference in kidney function between preterm- and term-born children, however, previous studies of preterm-born school-aged children and adults have not found any consistent significant difference in urine protein excretion when compared to term-born individuals (Rakow et al., 2008, Kistner et al., 2000).

As this finding is also plausibly related to a dilutional effect on the urine from differing hydration status of each participant at the time of sampling, I investigated potential methods

to normalise the urine sample data to correct for any dilutional effect and differing protein load analysed by the mass spectrometer. This is a key consideration, as the TMT-labelled proteomic analysis performed is based upon relative abundance of individual proteins detected between samples, therefore, any dilutional effect resulting in differing total sample protein quantities being loaded into the mass spectrometer between samples would potentially lead to false positive/negative results.

4.2.4.1 Extrinsic Factor Normalisation Methods:

Urine samples from RHiNO trial participants had also undergone biochemistry analysis at the University Hospital of Wales (UHW) clinical biochemistry laboratories. I investigated whether total urine protein (g/L) or urine creatinine (mmol/L) could be used as a normalisation factor for the urine proteomics dataset. The method used to determine total protein content (turbidimetric method using benzethonium chloride (Iwata and Nishikaze, 1979)) returned results which were too imprecise to use as a normalisation factor. As overall protein quantities were relatively low, and no subjects had gross proteinuria, the total urine protein results returned in g/L became almost a categorical variable (Figure 4-4 panel A) which would be unsuitable for use as a normalisation factor in the urine proteomics dataset, which had been analysed using mass spectrometry, a far more sensitive method. Next, I investigated whether urinary creatinine could be used as a normalization factor.

Urinary creatinine is commonly used clinically in renal pathology to quantify severity of disease and create corrected values for total protein and electrolyte excretion. Therefore, urinary creatinine appeared the most sensible correction factor to normalise the dataset. However, there was a wide range of urine creatinine results (0.44 - 24.0 mmol/L), and in addition there were significant (p<0.05) or near-significant (p<0.1) differences between the

major clinical groups (Figure 4-5). When applied as a normalization factor, it increased the range of median log_2 abundances observed between samples (Log_2FC 11.78, range -2.91, 8.87) (Figure 4-1 panel B), markedly reduced the number of significant protein differences between groups (Figure 4-3 panel B), and, therefore, was an unsuitable extrinsic normalisation factor.

After discussions with the proteomics facility at University of Bristol, I was informed that they use a Bradford protein assay (Bradford, 1976) to quantify the total protein content of their samples. The Bradford assay is more sensitive than the turbidimetric method used by the hospital clinical biochemistry laboratory, reporting results for protein content in µg/µL. I therefore arranged for all 304 urine samples to have their total protein content analysed by Bradford assay at the proteomics facility at University of Bristol. The Bradford assay results gave a greater degree of discrimination between samples (Figure 4-4 panel B), which demonstrated neither a strong nor significant correlation with the clinical biochemistry lab total protein measurements (Figure 4-4 panel C). The Bradford assay total urine protein did not appear to be significantly different between the major clinical groupings (Figure 4-6). However, when the Bradford total protein assay results were used as a normalisation factor, there remained a large variation in observed protein abundances between the samples (Figure 4-1 panel C), with all relative abundances markedly elevated with a greater range (Log_2FC 10.12, range 4.17, 14.29), and a persistent skewing of relative individual protein abundances (Figure 4-3 panel C).
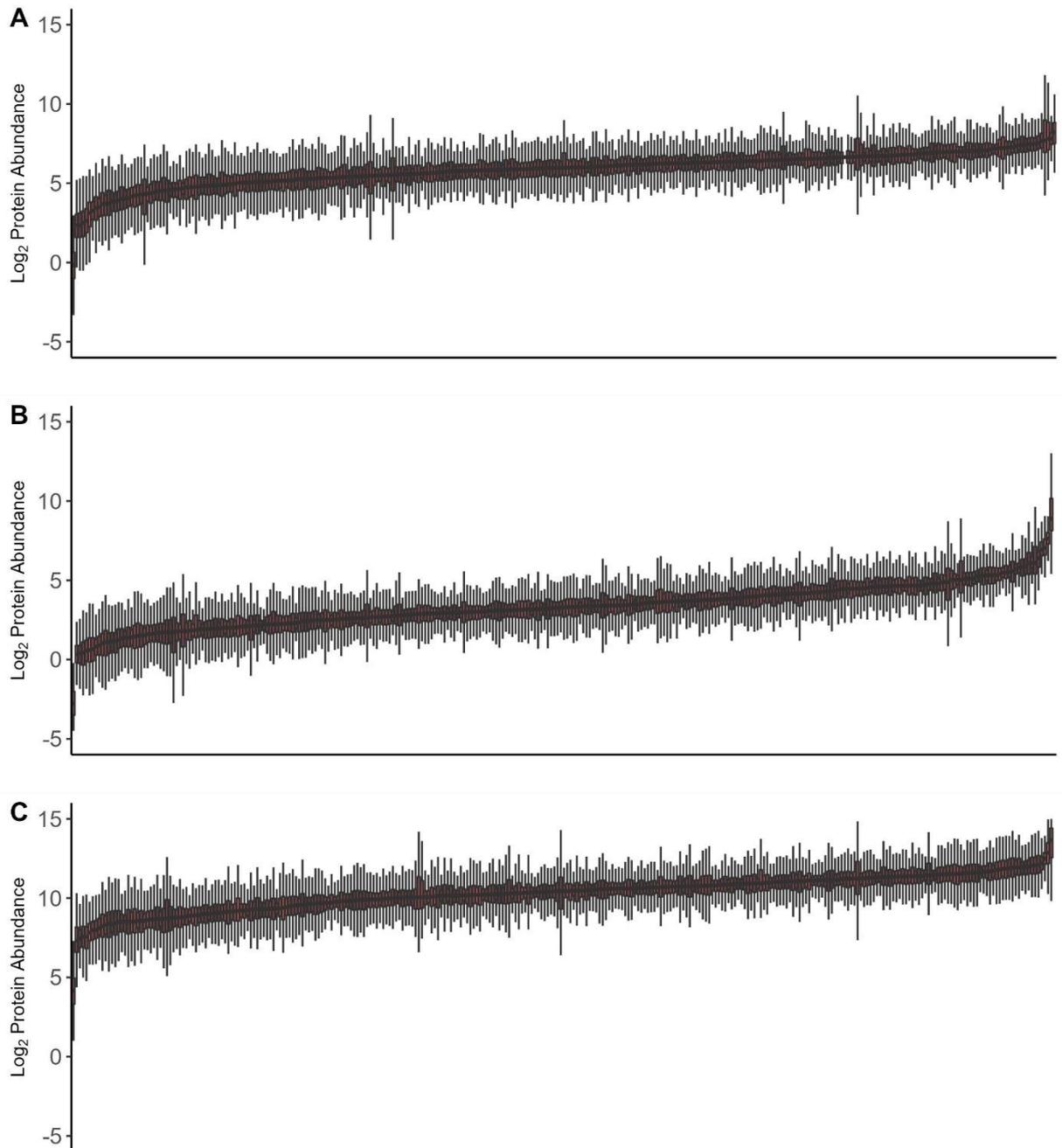
Figure 4-1: Boxplot of all Log$_2$ protein abundances for each urine sample analysed with various normalisation methods applied.

Samples ordered by median protein abundance. (A) No normalisation. (B) Normalised to urine creatinine. (C) Normalised to total protein (Bradford assay).
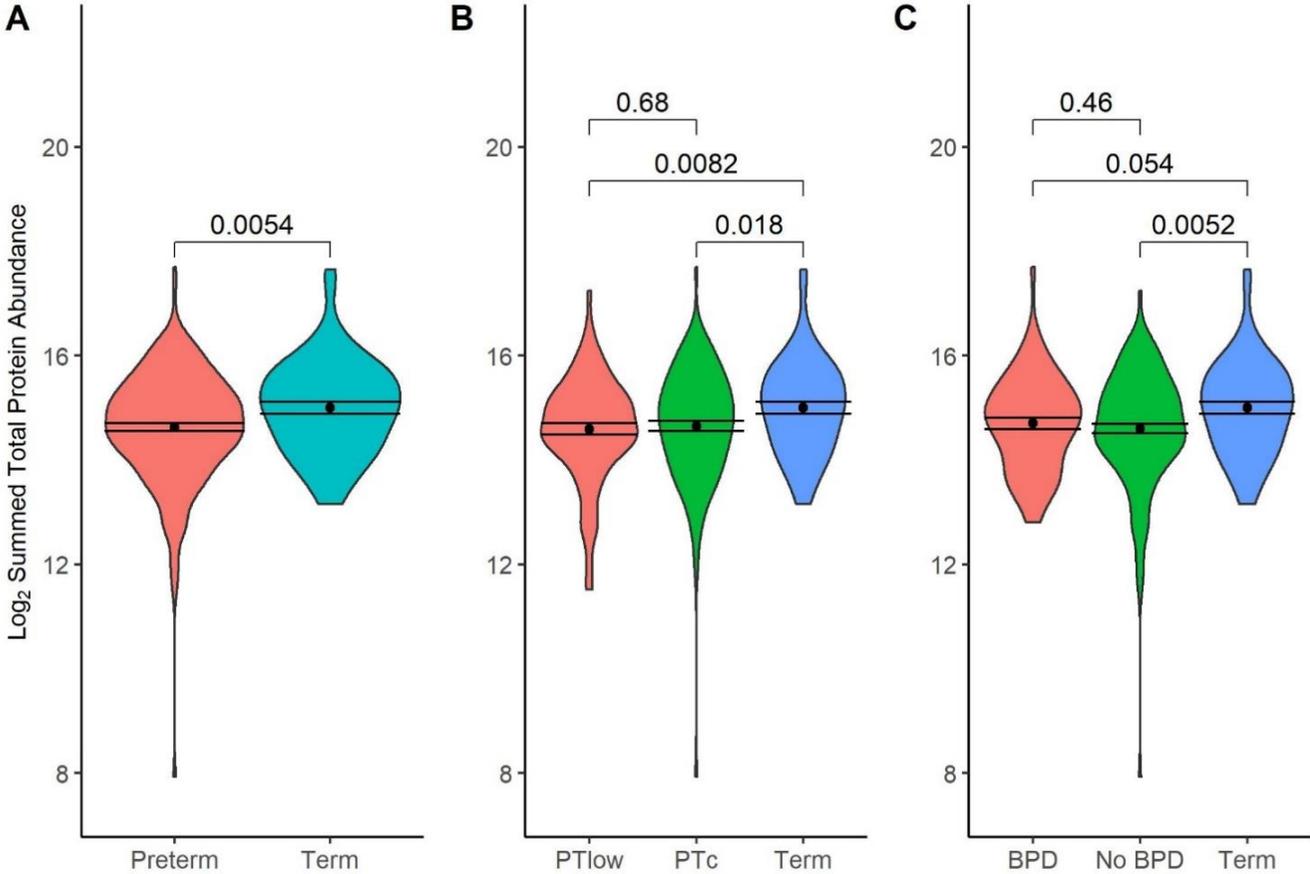
Figure 4-2: Violin plots demonstrating log$_2$ total protein abundance as detected by mass spectrometer between major clinical groupings.

Bar's give p-values by Welch's t-test between group comparisons.
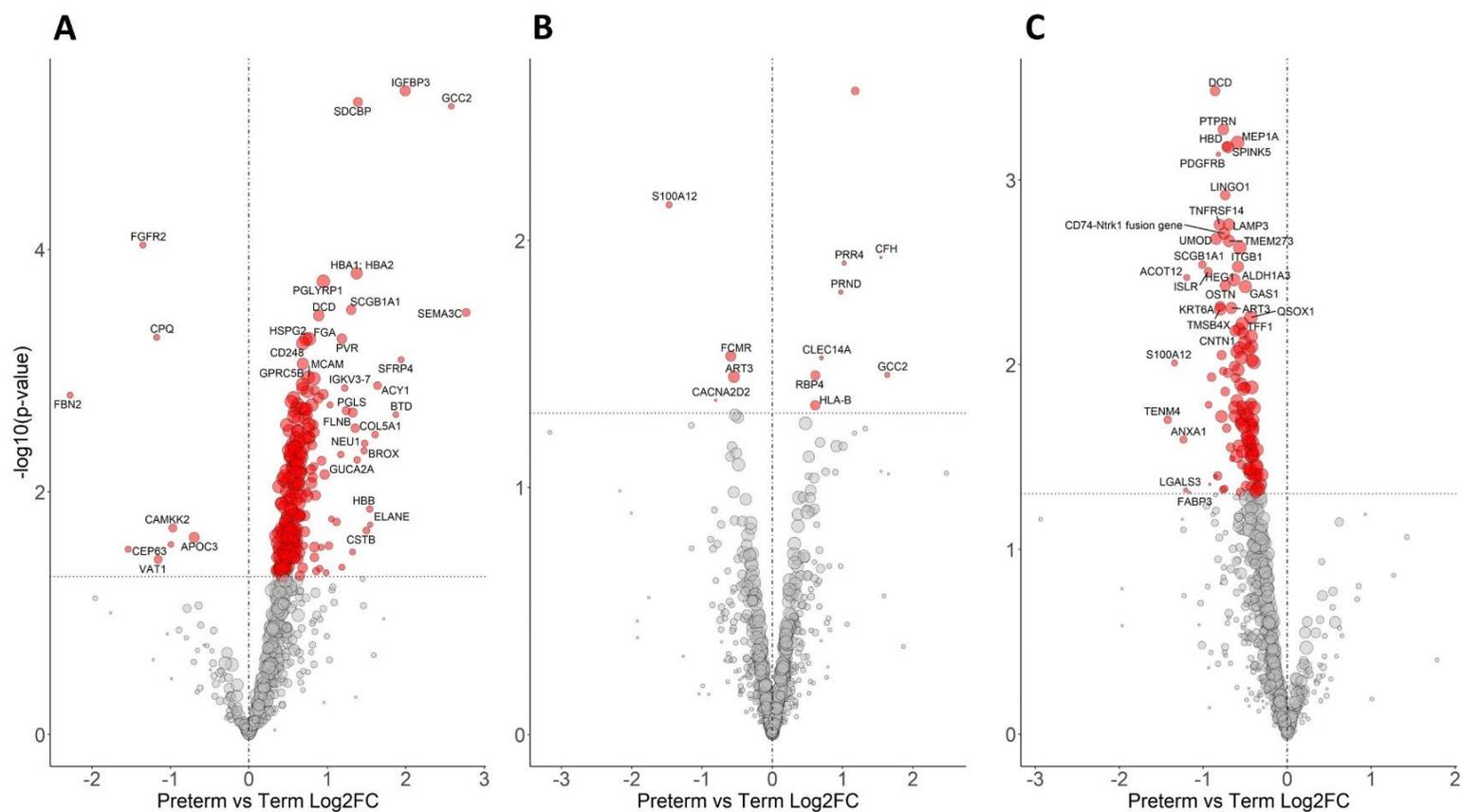
Figure 4-3: Test Volcano Plots comparing Preterm and Term groups by different normalisation methods.

(A) Log$_2$; (B) Normalised to urine creatinine; (C) Normalised to total protein by Bradford assay;. Protiens with a significant (p<0.05) log$_2$ fold change (Log2FC) between groups highlighted in red. Size of point represents replicate number (number of samples in which the protein was detected). Vertical line represents a Log2FC of 0. Horizontal line represents p<0.05
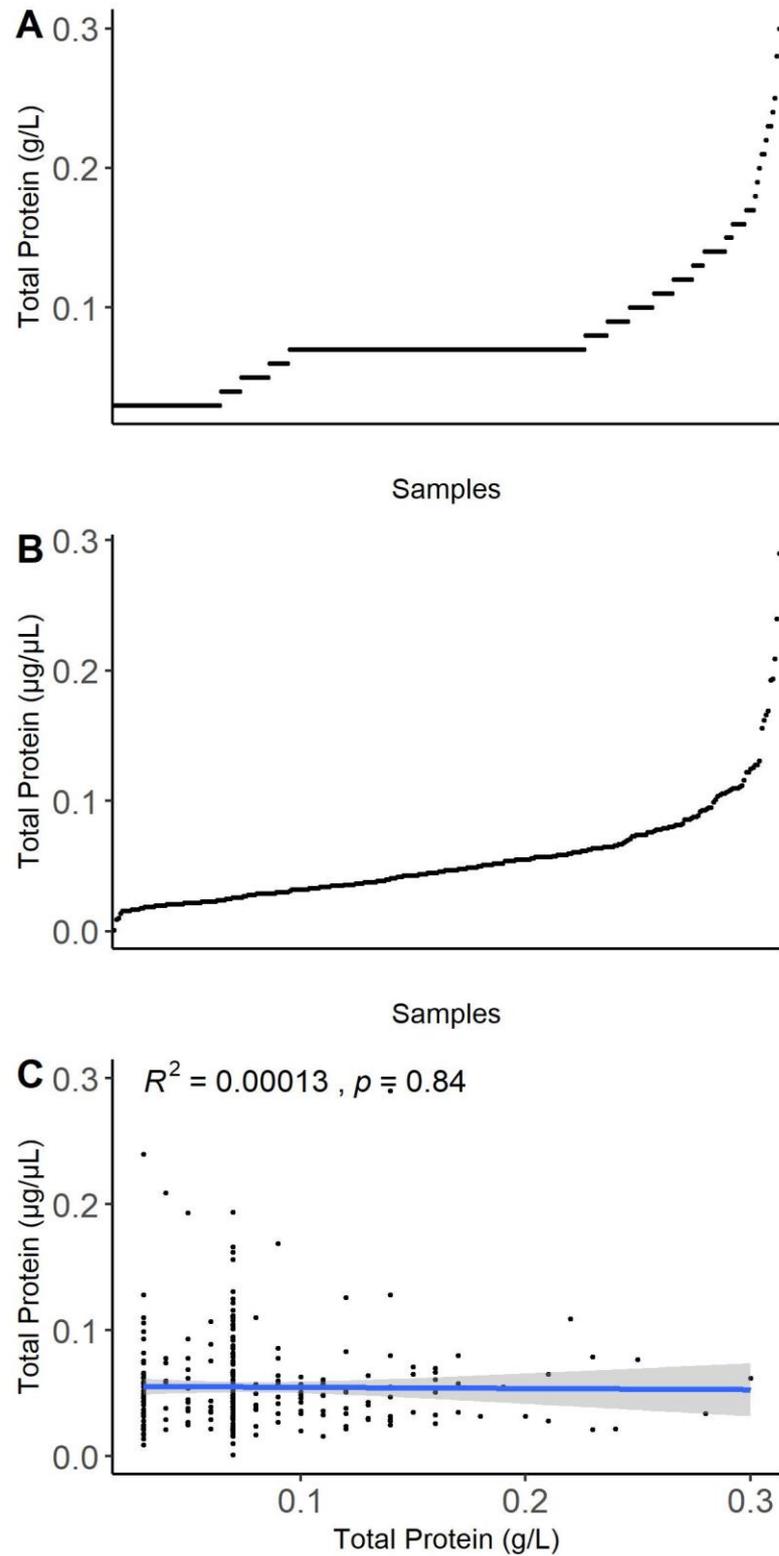
Figure 4-4: Scatterplots showing (A) Total urine protein content by clinical biochemistry assay, (B) Total urine protein content by Bradford Assay and (C) Relationship between two measurement methods.

Blue line represents a linear model regression line, with grey shaded area representing 95% confidence interval. R2 and p-value for regression line given.
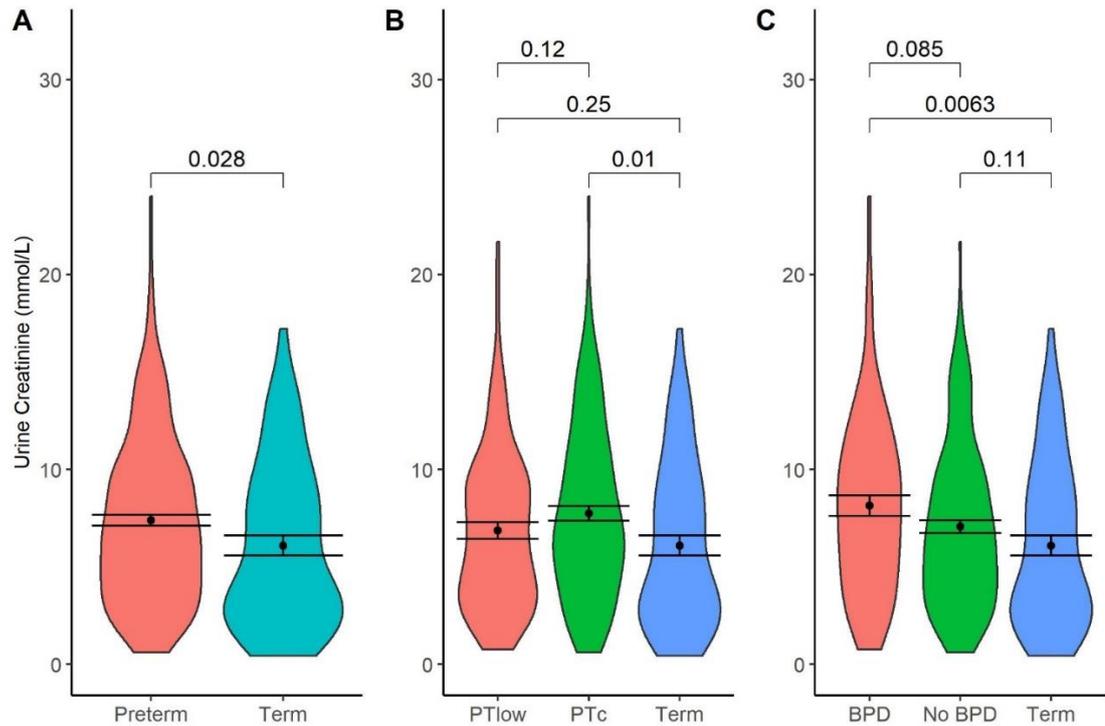
Figure 4-5: Violin plots demonstrating differences in urine creatinine between major clinical groupings.

Bar's give p-values by Welch's t-test between group comparisons.
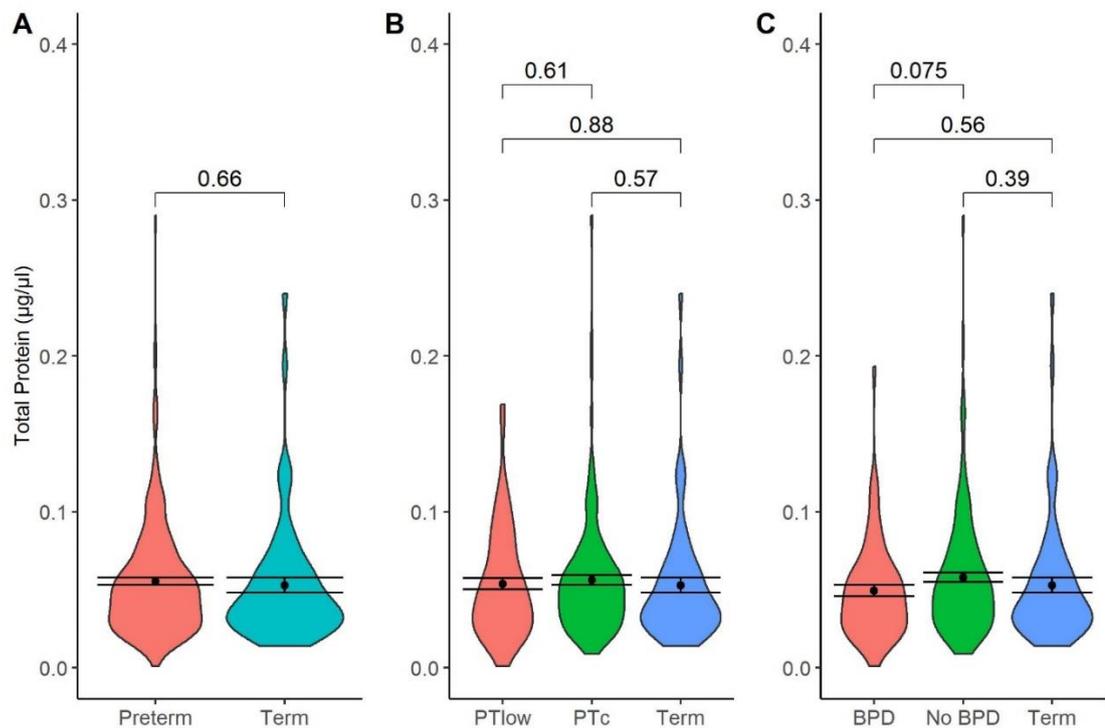
Figure 4-6: Violin plots demonstrating total urine protein content as analysed by Bradford assay between major clinical groupings.

Bar's give p-values by Welch's t-test between group comparisons.

4.2.4.2 Intrinsic Factor Normalization Methods

Owing to the persistent difficulties in using an extrinsic factor to normalise the dataset, I searched the published literature for methods utilised in other experiments for overcoming dilution/protein load differences in mass spectrometry-based proteomic studies. An untargeted, unlabeled proteomic experiment examining the urine of preterm-born infants at risk of BPD had utilised a similar methodology to my experiment, whereby an equal volume of urine, with an unknown protein content was loaded into the mass spectrometer (Ahmed et al., 2022). The authors of this study had performed a normalisation process of their data to account for differing protein loads using a central tendency method, normalising to the median of the total protein abundance of all of their samples, as described by Callister et al. (Callister et al., 2006).

This technique appeared promising and after further discussions with Dr Phil Lewis, Bioinformatician at the Proteomics Facility at University of Bristol, I spoke with Dr Michaela Scigelova, who is an LC/MS application specialist and works with the Proteome Discoverer software at Thermo Fischer Scientific, seeking her advice regarding normalisation of a labelled proteomics dataset. She agreed that a median scale normalisation was an appropriate approach to use, however, rather than using the median of the whole dataset, to use the median of each sample's respective mass spectrometer (MS) run, as this would better account for inter-run technical variations. Samples could then be scaled to each run's pooled sample to allow comparison of samples between mass spectrometry runs. This is the same method that the internal normalisation method in Proteome Discoverer software uses for smaller differences in protein load between samples/runs. Dr Scieglova advised that accurate protein quantitation with Proteome Discoverer v2.1 software became less reliable if there were large differences in total protein abundance between samples, and within her team at

Thermo Fischer Scientific, they would have reservations interpreting results where there was a three-to-four-fold difference in abundance between them. I, therefore, calculated the median protein abundances for each MS run, and then normalised the detected protein abundance in each sample to the median protein abundance of the sample's respective MS run. I then excluded samples where the total protein abundance was either two-fold higher or lower than the median MS run protein abundance. Of the 304 samples analysed, 75 were excluded as outliers using this process (Figure 4-7). Following this, all samples were scaled by their respective MS runs pool samples. The results of this normalisation process are shown in Figure 4-8 and Figure 4-9. This process markedly improved the range of median protein abundances between samples (Log$_2$FC 6.67, Range -1.14, 13.6). A test analysis comparing preterm and term groups demonstrated a more even distribution of significant protein differences (Figure 4-10).

4.2.5   Statistical analysis

Baseline population characteristics were compared using Chi-squared or t-test or as appropriate. Replicate numbers (number of samples in which a particular protein was detected) were calculated. Scaled protein abundances were log$_2$-transformed and fold changes (log$_2$FC) between groups were compared using Welch's t-test. I performed all analyses were performed using R v4.0.4(R Core Team, 2021), using R packages *"stringr"*, *"dplyr"*, *"ggplot2"*, *"ggpubr"*, *"pROC"* and *"reportROC"*. Gene name is used synonymously with protein name throughout. Functional enrichment analysis (identifying changes in classes of proteins present) was performed with Webgestalt (Liao et al., 2019). Ingenuity Pathways Analysis (IPA, Qiagen®, Germany) software was used to identify functional relationships between significantly different protein abundances between groups, highlighting altered biological processes. I generated Receiver Operator Characteristic (ROC) curves for

biologically related proteins, with high replicate numbers, between study groups (identified by IPA) in R using the *"pROC"* and *"reportROC"* packages. By doing this I aimed to assess potential biomarker performance (assessed by area under the curve (AUC)), using two linear models, one based on the whole cohort and one using a leave-one-out cross validation (LOOCV) method. I also analysed these proteins of interest using univariable linear regression models to ascertain associations between these proteins and other early and current life factors. Associations with a p-value < 0.1 were combined into a multivariable model to examine the overall combined influence of each association. As in Chapters 2 and 3, owing to the exploratory nature of the analysis, a p-value of <0.05 was considered significant.

4.3 Results:

4.3.1 Participants

Urine samples were analysed from 271 RHiNO study participants with valid spirometry. 64 (23.6%) samples were excluded as outliers, as described above in section 4.2.4.2. Demographic details for those participants included in the analysis (n=207) are given in Table 4-1. Preterm-born children were significantly older than the term-born children (mean 10.4±1.4 years vs 9.9±1.1, p=0.02) and had higher rates of asthma (39 (24.2%) vs 2 (4.3%), p=<0.001). 47 (29.2%) of the preterm-born subjects had received a neonatal diagnosis of mild/moderate/severe BPD, and 50 (31.1%) had an FEV_1<LLN. Of those, 27 (54%) were classified as pPRISm and 23 (46%) as POLD. Between included and excluded subjects, POLD had significantly higher rates of asthma and lower rates of BPD in those included in the analysis compared to those excluded (Table 4-2), otherwise there were no significant differences between those included and excluded from the analysis.
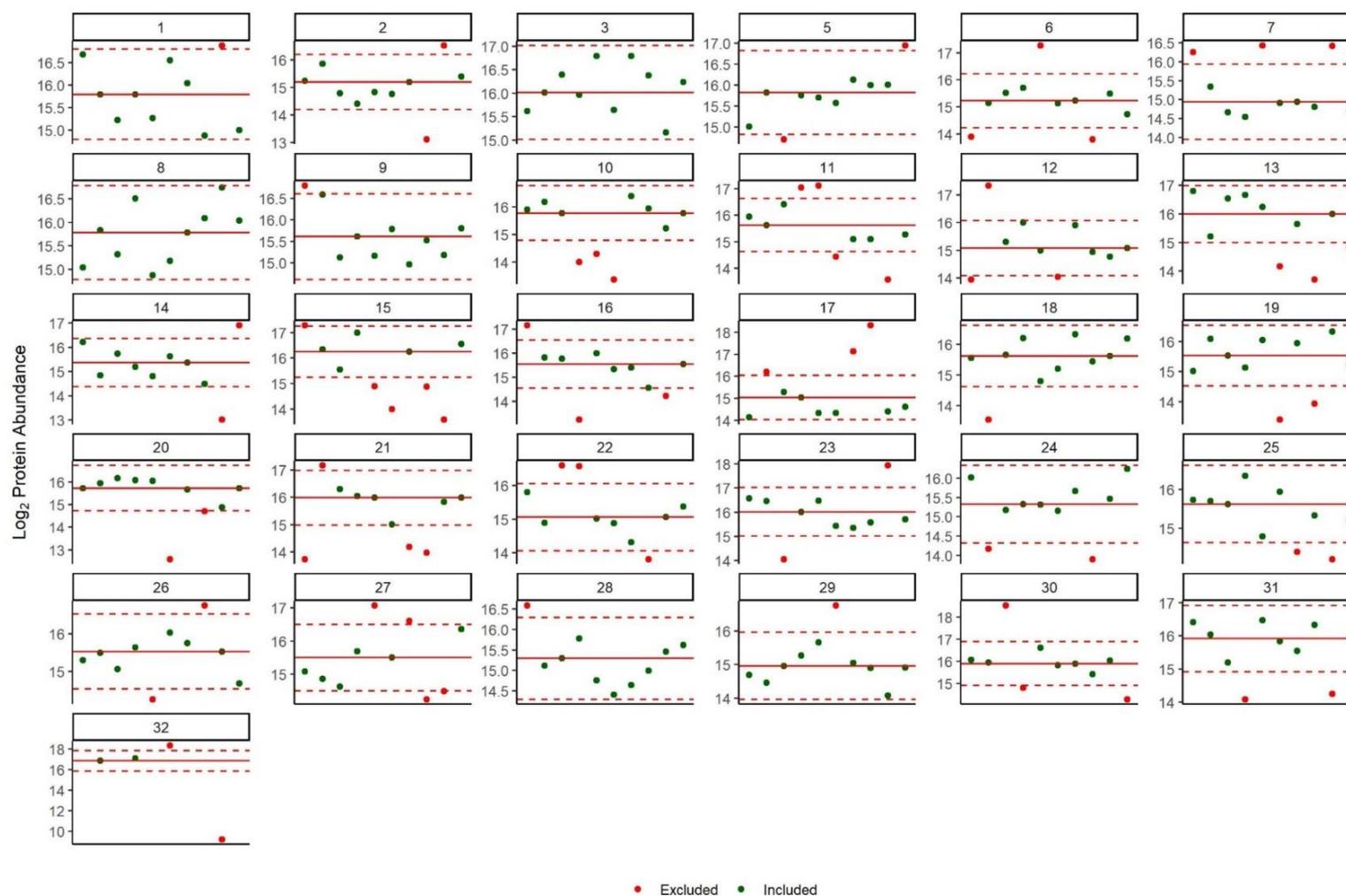
Figure 4-7: Log₂ total protein abundance of each sample separated by mass spectrometry (MS) run (given in number above each graph).

Solid red line represents median of total protein abundance of each MS run, with dashed lines representing ±2-fold difference of MS run median total protein abundance. Samples with a total protein abundance >±2-fold difference from median of total protein abundance of the MS run highlighted in red as outliers and excluded from further analysis.
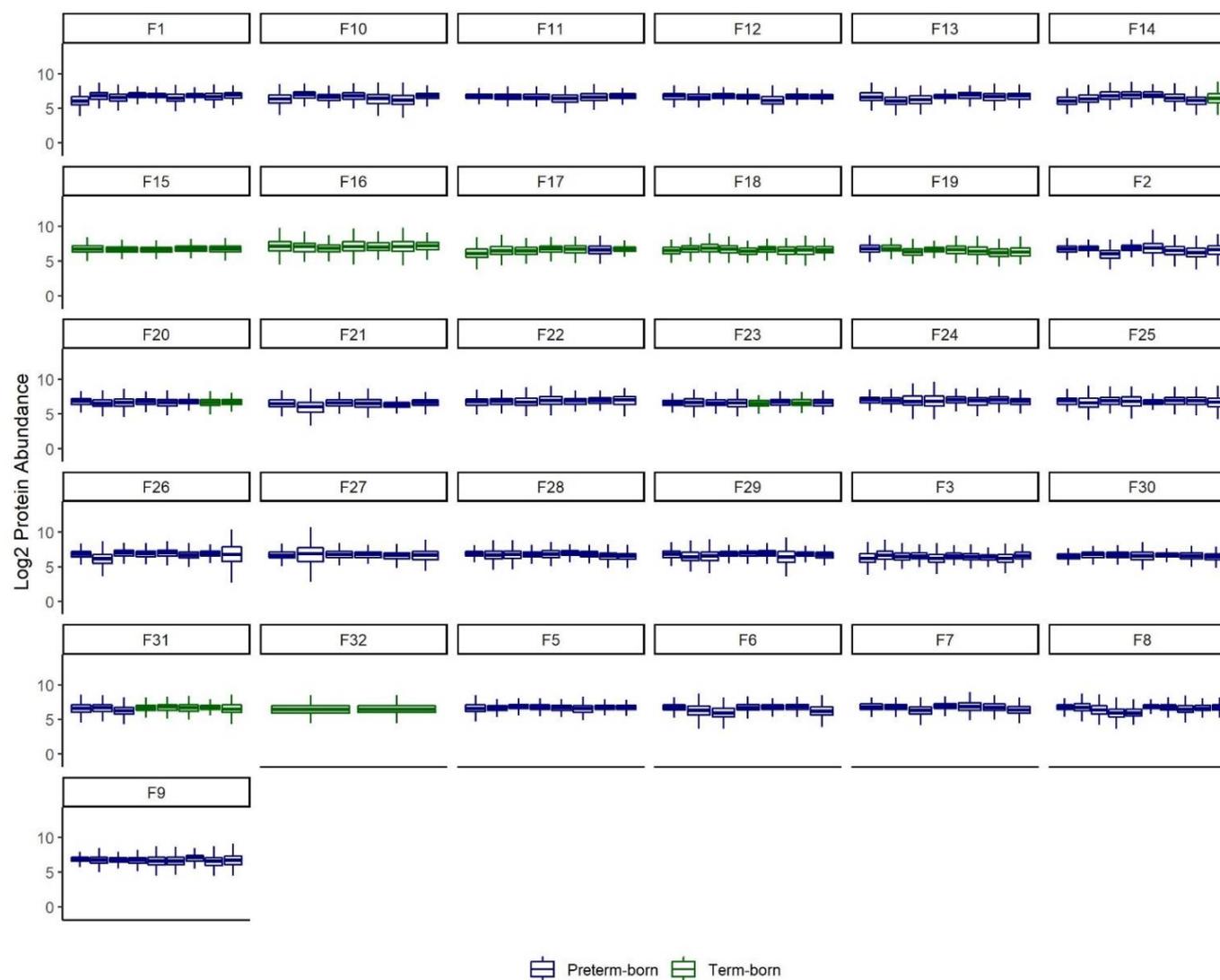
Figure 4-8: Log₂ normalized and scaled protein abundances for each sample given by boxplot by MS run. Preterm-born subjects in navy blue, term-born subjects in green.
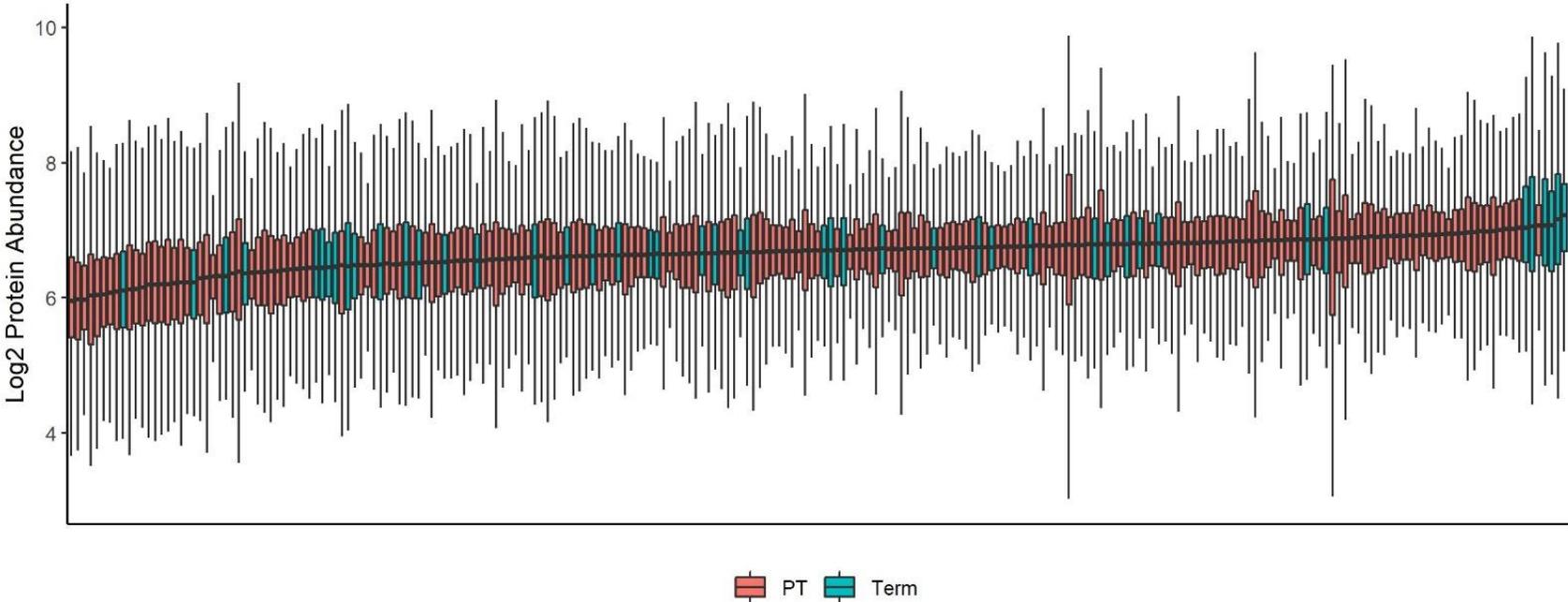
Figure 4-9: Log$_2$ normalised and scaled protein abundances

PT: Preterm-born participants. Term: Term-born participants
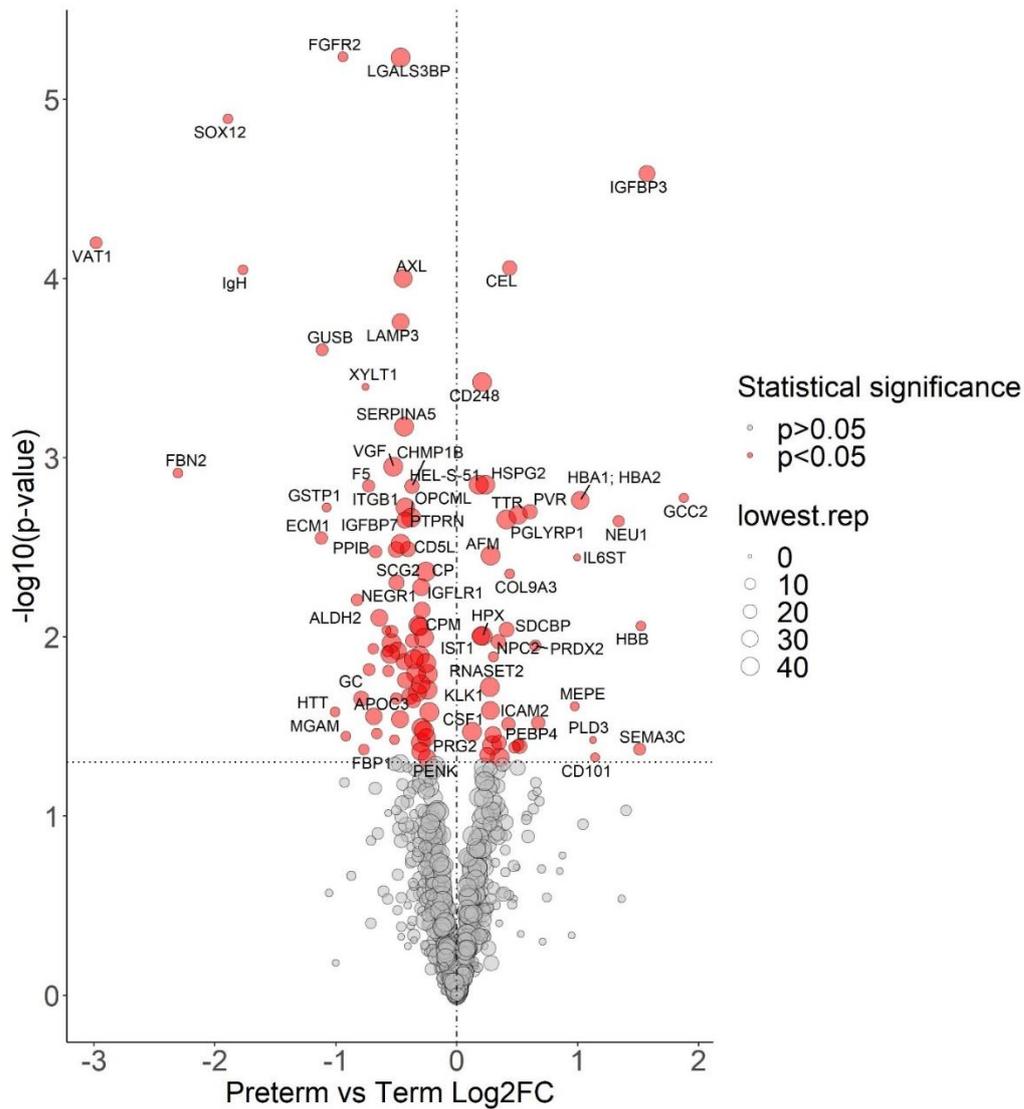
Figure 4-10: Test volcano plot showing protein differences between preterm- and term-born children following median scale normalisation and scaling of dataset.

Protiens with a significant (p<0.05) $\log_2$ fold change (Log2FC) between groups highlighted in red. Size of point represents replicate number (number of samples in which the protein was detected). Vertical line represents a Log2FC of 0. Horizontal line represents p<0.05

4.3.2 Detected proteins

A total of 785 proteins were detected, 735 (93.6%) of which were mapped to official gene symbols (gene names). The remaining 50 proteins could not be mapped to known genes. 129 proteins were present in all samples analysed. Functional enrichment analysis (Liao et al., 2019) was possible for 681 (86.8%) of the detected proteins (Figure 4-11). 288 proteins were significantly different between any of the phenotype comparisons, and functional enrichment analysis was possible for 255 (88.5%). Overall, an enrichment of proteins related to metabolic processes, hydrolase activity and extracellular space/cell membrane activities was observed in the preterm-born groups.

4.3.3 Comparison between the pPRISm group with preterm- and term-born control groups

There were no significant differences in demographics between the pPRISm and PT_c groups (Table 4-1). 37 (5.3%) proteins had significantly different abundance when compared to PT_c (Figure 4-12; Table 4-3), and 62 (8.9%) when compared to the term-born group (Figure 4-12;Table 4-4). 14 proteins were common between the two comparisons. IPA linked 16 significantly altered proteins in pPRISm compared to PT_c to six biological processes (Table 4-5; Figure 4-13):

- *Inflammation of body cavity* (PGLYRP1, DNASE1, MYH9, SERPINA3, CTSV, AGT, ANXA1, CLEC4G, SCGB1A1, B2M, CD14) (p=0.042).

- *Apoptosis of myeloid cells* (SERPINA3, ANXA1, ANPEP, CD14) (p=0.038).

- *Quantity of leucocytes* (GLA, CLEC11A, PGLYRP1, DNASE1, CTSV, AGT, ANXA1, CLEC4G, SCGB1A1, B2M, ANPEP) (p=0.038).

- *Quantity of T-lymphocytes* (PGLYRP1, DNASE1, CTSV, AGT, ANXA1, CLEC4G, B2M, ANPEP) (p=0.015).

IPA-calculated activation z-scores suggested upregulation of these processes (Figure 4-13). There was also a significant link between these proteins and the quantities of CD4+ (p=0.008) and CD8+ (p=0.005) T-lymphocytes, with a suggestion of a downregulation of CD4+ T-lymphocytes (activation z-score -0.73). IPA analysis of significantly different protein abundances in pPRISm group compared to the Term-born group linked six proteins (AGT, CD14, CSF1, FABP5, HBB, ANXA1) with *Synthesis of prostaglandin* (p=0.038, activation z-score 1.23)). Five proteins (PRG2, MGAM, CD14, LGALS3BP, ANXA1) were significantly linked with neutrophil activation (p=0.038, z-score -0.64).

ROC analysis (Table 4-6; Figure 4-14) demonstrated that DNASE1, PGLYRP1, B2M and SERPINA3 in combination had the highest predictive ability for identifying pPRISm from within the preterm group (AUC: 0.73 (95% confidence interval 0.61, 0.84), sensitivity 0.80 (0.64, 0.96), specificity 0.73 (0.64, 0.82), p=<0.001). Using the LOOCV model, the predictive ability of this protein panel was AUC 0.65 (0.52, 0.78), p=0.01 (Table 4-7; Figure 4-15). Results from univariable and multivariable linear regression modelling for these proteins are shown in Table 4-8. DNASE1, PGLYRP1, B2M remained significantly associated with pPRISm in multivariable modelling (p-values 0.008, 0.011, 0.018 respectively) with B2M also being significantly associated with a history of BPD in the multivariable model (p=0.003). No other life factors were significantly associated with SERPINA3 on univariable models, with pPRISm being highly significant (p=0.005).

Variable	Term born (≥37/40) n = 46	Preterm born (≤34/40) n = 160	Preterm born Controls n = 112	pPRISm n = 27	POLD n = 23
Sex (male), n(%)	23 (50.0)	76 (47.5)	52 (46.4)	17 (63)	7 (33.3)
Ethnicity (white), n(%)	45 (97.8)	152 (95)	105 (93.8)	27 (100)	20 (95.2)
Age at testing (years), mean (SD)	9.9 (1.1)	10.4 (1.4)*	10.4 (1.3)	10.6 (1.6)	10.1 (1.7)
Weight (kg), mean (SD)	36.3 (10.3)	35.9 (9.7)	36.4 (9.5)	35.5 (9.1)	33.9 (11.8)
Body Mass Index (kg/m^2), mean (SD)	18.1 (3.2)	17.5 (3.1)	17.7 (3.0)	17.0 (2.9)	17.0 (3.9)
Wheeze-ever, n(%)	12 (26.1)	97 (56.9)***	64 (57.1)	15 (55.6)	18 (85.7)†ⱸ
Doctor-diagnosed asthma, n(%)	2 (4.3)	39 (24.4)**	21 (18.8)	7 (25.9)	11 (52.4)††
Neonatal Characteristics					
Gestational age (weeks), mean (SD)	40.1 (1.2)	30.5 (2.8)***	30.4 (2.9)	30.9 (2.9)	30.2 (2.5)
Birthweight (g), mean (SD)	3499 (576)	1549 (594)***	1577 (607)	1587 (543)	1352 (572)
Birthweight (z-score), mean (SD)	0.1 (0.99)	0.11 (1.37)	0.3 (1.42)	-0.12 (0.9)	-0.45 (1.46)
Intrauterine growth restriction, n(%)	3 (6.5)	25 (15.6)	15 (13.4)	2 (7.4)	8 (38.1)††ⱸⱸ
Antenatal Steroids, n(%)	0 (0)	137 (85.6)***	99 (88.4)	22 (81.5)	16 (76.2)
Mechanical ventilation, n(%)	0 (0)	70 (43.8)***	53 (47.3)	7 (25.9)	10 (47.6)
Bronchopulmonary Dysplasia (BPD), n(%)	0 (0)	47 (29.2)***	34 (30.4)	7 (25.9)	6 (28.6)
Antenatal smoking, n(%)	3 (6.5)	22 (13.8)	16 (14.3)	4 (14.8)	2 (9.5)

Table 4-1: Urine Proteomics Sample Demographics.

Preterm vs Term: *p<0.05, **p<0.01, ***p<0.001. pPRISm/POLD vs Preterm born control: †p<0.05, ††p<0.01, †††p<0.001. pPRISm vs POLD:
ⱸp<0.05, ⱸⱸp<0.01, ⱸⱸⱸp<0.001

pPRISm: Prematurity-related preserved ratio with impaired spirometry. POLD: Prematurity-related obstructive lung disease. BPD:
Bronchopulmonary dysplasia.

Variable	Term born (≥37/40)		Preterm born (≤34/40)		Preterm born Controls		POLD		pPRISm	
	Included	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included	Excluded
	n = 46	n = 20	n = 160	n = 44	n = 112	n = 28	n = 23	n = 6	n = 27	n = 10
Sex (male), n(%)	23 (50.0)	11 (55)	76 (47.5)	22 (50)	52 (46.4)	14 (50)	7 (33.3)	3 (50)	17 (63)	5 (50)
Ethnicity (white), n(%)	45 (97.8)	20 (100)	152 (95)	42 (95.5)	105 (93.8)	26 (92.9)	20 (95.2)	6 (100)	27 (100)	10 (100)
Age at testing (years), mean (SD)	9.9 (1.1)	10.3 (1.0)	10.4 (1.4)	10.4 (1.3)	10.4 (1.3)	10.0 (1.3)	10.1 (1.7)	10.9 (1.2)	10.6 (1.6)	11.3 (1.0)
Weight (kg), mean (SD)	36.3 (10.3)	36.1 (10.4)	35.9 (9.7)	37.2 (10.9)	36.4 (9.5)	36.7 (11.7)	33.9 (11.8)	36.0 (9.0)	35.5 (9.1)	39.2 (10.6)
Body Mass Index (kg/m^2), mean (SD)	18.1 (3.2)	17.3 (3.0)	17.5 (3.1)	18.4 (3.6)	17.7 (3.0)	18.6 (3.8)	17.0 (3.9)	17.7 (2.8)	17.0 (2.9)	18.2 (3.7)
Wheeze-ever, n(%)	12 (26.1)	5 (25)	97 (56.9)	22 (50)	64 (57.1)	16 (57.1)	18 (85.7)	3 (50)	15 (55.6)	3 (30)
Doctor-diagnosed asthma, n(%)	2 (4.3)	1 (5)	39 (24.4)	5 (11.4)	21 (18.8)	3 (10.7)	**11 (52.4)**	**0 (0)***	7 (25.9)	2 (20)
Neonatal Characteristics										
Gestational age (weeks), mean (SD)	40.1 (1.2)	40.4 (1.1)	30.5 (2.8)	30.5 (3.0)	30.4 (2.9)	30.8 (3.1)	30.2 (2.5)	28.6 (2.7)	30.9 (2.9)	31.0 (2.5)
Birthweight (g), mean (SD)	3499 (576)	3650 (627)	1549 (594)	1575 (650)	1577 (607)	1642 (672)	1352 (572)	1050 (428)	1587 (543)	1702 (590)
Birthweight (z-score), mean (SD)	0.1 (0.99)	0.2 (1.2)	0.11 (1.37)	0.15 (1.52)	0.3 (1.42)	0.3 (1.6)	-0.45 (1.46)	-0.82 (1.20)	-0.12 (0.9)	0.3 (1.4)
Intrauterine growth restriction, n(%)	3 (6.5)	2 (10)	25 (15.6)	7 (15.9)	15 (13.4)	3 (10.7)	8 (38.1)	3 (50)	2 (7.4)	1 (10)
Antenatal Steroids, n(%)	0 (0)	0 (0)	137 (85.6)	39 (88.6)	99 (88.4)	24 (85.7)	16 (76.2)	6 (100)	22 (81.5)	9 (90)
Mechanical ventilation, n(%)	0 (0)	0 (0)	70 (43.8)	18 (40.9)	53 (47.3)	11 (39.3)	10 (47.6)	5 (83.3)	7 (25.9)	2 (20)
BPD, n(%)	0 (0)	0 (0)	47 (29.2)	14 (31.8)	34 (30.4)	7 (25)	**6 (28.6)**	**5 (83.3)***	7 (25.9)	2 (20)
Antenatal smoking, n(%)	3 (6.5)	0 (0)	22 (13.8)	5 (1.4)	16 (14.3)	4 (14.3)	2 (9.5)	0 (0)	4 (14.8)	1 (10)

Table 4-2: Comparing included and excluded samples from urine proteomics analysis.

Independent samples T-test/Chi-squared test; Included vs Excluded: *p<0.05, **p<0.01, ***p<0.001

POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. BPD: Bronchopulmonary dysplasia.
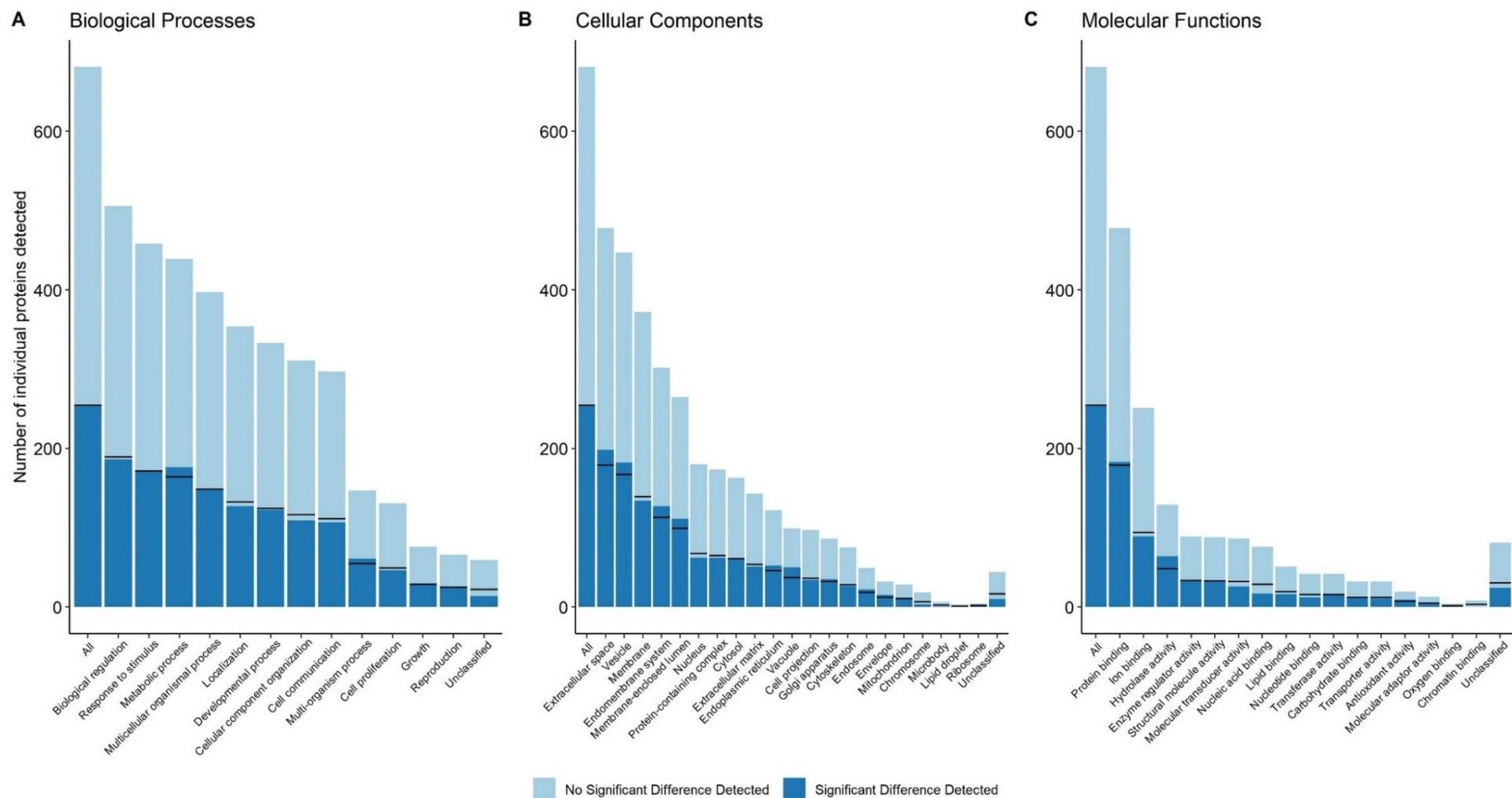
Figure 4-11: Functional enrichment analysis of detected urine proteome.

Black bar on each column represents the expected background enrichment level.
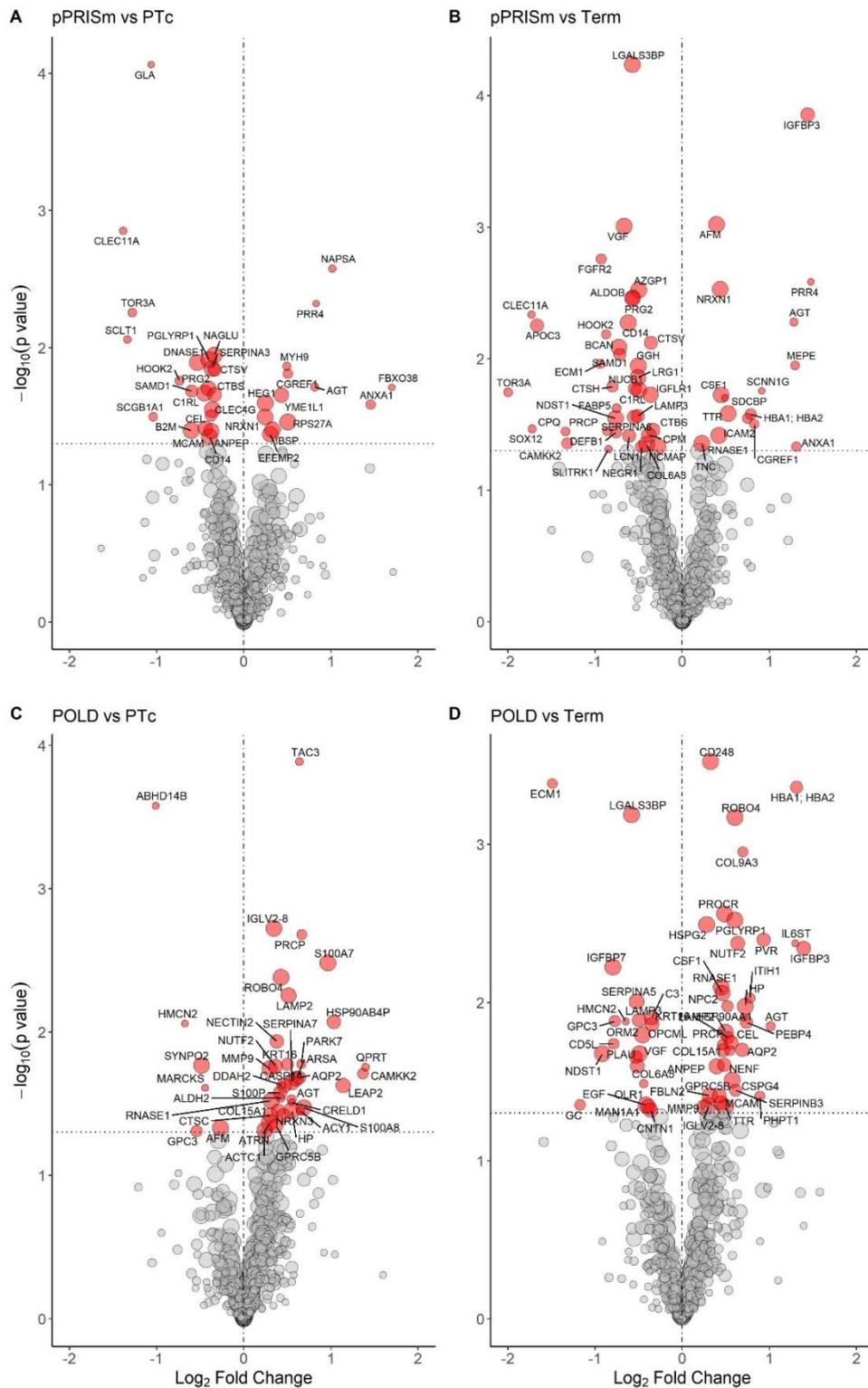
Figure 4-12: Volcano Plots showing significant urine proteome differences for pPRISm and POLD phenotypes compared to PT$_c$ and Term groups.

Vertical line represents log$_2$ fold change of 0. Horizontal line equivalent to a p-value of 0.05. Proteins with a significant difference between groups highlighted and labelled with respective gene name. Size of circle relative to replicate number. POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PTc: Preterm-born controls

UniProt Accession	Gene Name	Protein Name	Protein Function	pPRISm vs PTc n = 27 v 111		
				Replicates	Log$_2$FC	p
P06280	GLA	Alpha-galactosidase A	Lipid metabolism	2 v 6	-1.06	0.0001
Q9Y240	CLEC11A	C-type lectin domain family 11 member A	Osteogenesis	3 v 15	-1.39	0.001
A0A2U8J8Y8	IgH	Ig heavy chain variable region	Immune response	2 v 6	-1.86	0.002
O96009	NAPSA	Napsin-A	May be involved in processing of pneumocyte surfactant precursors.	3 v 17	1.01	0.003
Q16378	PRR4	Proline-rich protein 4	Extracellular space/Visual perception	2 v 15	0.83	0.005
Q9H497	TOR3A	Torsin-3A	Nucleotide binding	5 v 17	-1.28	0.006
Q96NL6	SCLT1	Sodium channel and clathrin linker 1	Anchors basal body to plasma membrane	3 v 11	-1.34	0.009
P54802	NAGLU	Alpha-N-acetylglucosaminidase	Glycosidase/hydrolase	25 v 107	-0.34	0.011
O75594	PGLYRP1	Peptidoglycan recognition protein 1	Innate immunity	27 v 111	-0.40	0.012
P24855	DNASE1	Deoxyribonuclease-1	Serum endonuclease	27 v 111	-0.53	0.013
P35579	MYH9	Myosin-9	Cell adhesion/cell shape	4 v 8	0.49	0.014
P01011	SERPINA3	Alpha-1-antichymotrypsin	Serine protease inhibitor	25 v 100	-0.36	0.014
O60911	CTSV	Cathepsin L2	Thiol protease	16 v 73	-0.33	0.015
Q99674	CGREF1	Cell growth regulator with EF hand domain protein 1	Cell adhesion, cell cycle	7 v 22	0.51	0.016
Q96ED9	HOOK2	Protein Hook homolog 2	Protein transport	6 v 24	-0.74	0.017
P01019	AGT	Angiotensinogen (Serpin A8)	Regulation of blood pressure	4 v 28	0.81	0.019
Q6PIJ6	FBXO38	F-box only protein 38	Adaptive immune response	2 v 11	1.70	0.020
P13727	PRG2	Bone marrow proteoglycan	Immune response, antimicrobial	22 v 91	-0.41	0.020
Q6SPF0	SAMD1	Sterile alpha motif domain-containing protein 1	Chromatin regulator	13 v 63	-0.60	0.021
Q9NZP8	C1RL	Complement C1r subcomponent-like protein	Serine protease, complement pathway	25 v 105	-0.45	0.021
Q01459	CTBS	Di-N-acetylchitobiase	Glycosidase	19 v 87	-0.34	0.022
Q96TA2	YME1L1	ATP-dependent zinc metalloprotease YME1L1	Metalloprotease	21 v 80	0.43	0.022
Q9ULI3	HEG1	Protein HEG homolog 1	Developmental protein, heart and vessel formation	27 v 111	0.25	0.025
P04083	ANXA1	Annexin A1	Inflammatory/immune response	7 v 29	1.46	0.026
Q6UXB4	CLEC4G	C-type lectin domain family 4 member G	Substrate binder, cell receptor for virus entry	27 v 106	-0.36	0.028
P19835	CEL	Bile salt-activated lipase	Serine esterase, lipid degradation	12 v 64	-0.38	0.031
B1AHL2	FBLN1	Fibulin-1	Cell adhesion/migration	4 v 24	0.49	0.031

P11684	SCGB1A1	Uteroglobin	Phospholipase A2 inhibitor	5 v 33	-1.04	0.032
P58400	NRXN1	Neurexin-1-beta	Cell adhesion	26 v 107	0.24	0.032
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a	Ribonucleoprotein	27 v 111	0.51	0.035
A0A0J9YVZ3	MGAM	Maltase-glucoamylase, intestinal	Carbohydrate metabolism	27 v 111	-0.35	0.038
P43121	MCAM	Cell surface glycoprotein MUC18	Cell adhesion	21 v 87	-0.45	0.039
P21815	IBSP	Bone sialoprotein 2	Biomineralization, cell adhesion	26 v 101	0.33	0.040
P61769	B2M	Beta-2-microglobulin	Class I major histocompatibility complex	27 v 111	-0.60	0.040
P15144	ANPEP	Aminopeptidase N	Aminopeptidase	23 v 103	-0.37	0.040
O95967	EFEMP2	EGF-containing fibulin-like extracellular matrix protein 2	Elastic fibre formation	25 v 101	0.30	0.043
P08571	CD14	Monocyte differentiation antigen CD14	Immune and inflammatory response	27 v 111	-0.39	0.045

Table 4-3: Significantly altered urine protein abundances between pPRISm vs PTc.

UniProt Accession	Gene Name	Protein Name	Protein Function	pPRISm vs Term n = 27 v 46		
				Replicates	Log$_2$FC	p
Q08380	LGALS3BP	Galectin-3-binding protein	Cell adhesion	27 v 46	-0.57	0.0001
P17936	IGFBP3	Insulin-like growth factor-binding protein 3	Growth regulation	18 v 31	1.44	0.0001
P43652	AFM	Afamin	Protein transport	27 v 46	0.39	0.001
A0A2U8J953	IgH	Ig heavy chain variable region (Fragment)	Immune response	9 v 7	-2.45	0.001
O15240	VGF	Neurosecretory protein VGF	Neurogenesis/neuroplasticity	27 v 46	-0.67	0.001
P21802	FGFR2	Fibroblast growth factor receptor 2	Cell proliferation regulator	8 v 8	-0.93	0.002
Q16378	PRR4	Proline-rich protein 4	Extracellular space/Visual perception	2 v 7	1.48	0.003
P58400	NRXN1	Neurexin-1-beta	Cell adhesion	26 v 46	0.44	0.003
Q9UL78	Unknown	Myosin-reactive immunoglobulin light chain variable region (Fragment)	Unknown	22 v 46	-0.44	0.003
P25311	AZGP1	Zinc-alpha-2-glycoprotein	Lipid degradation	27 v 46	-0.50	0.003
P05062	ALDOB	Fructose-bisphosphate aldolase B	Glycolysis	23 v 46	-0.57	0.003
P13727	PRG2	Proteoglycan 2	Immune response, antimicrobial	22 v 38	-0.57	0.004
Q9Y240	CLEC11A	C-type lectin domain family 11 member	Osteogenesis	3 v 9	-1.73	0.005
P01019	AGT	Angiotensinogen (Serpin A8)	Regulation of blood pressure	4 v 8	1.28	0.005
P08571	CD14	Monocyte differentiation antigen CD14	Immune and inflammatory response	27 v 46	-0.62	0.005
P02656	APOC3	Apolipoprotein C	Lipid metabolism	16 v 26	-1.67	0.006
Q96ED9	HOOK2	Protein Hook homolog 2	Protein transport	6 v 16	-0.87	0.007
O60911	CTSV	Cathepsin L2	Thiol protease	16 v 39	-0.36	0.008
Q96GW7	BCAN	Brevican core protein	Nervous system development	26 v 34	-0.73	0.008
Q8TE24	MGAM	Maltase-glucoamylase	Carbohydrate metabolism	5 v 7	-1.31	0.008
Q6SPF0	SAMD1	Sterile alpha motif domain-containing protein 1	Chromatin regulator	13 v 28	-0.72	0.009
Q99985	SEMA3C	Semaphorin-3C	Developmental protein	6 v 14	2.72	0.010
Q16610	ECM1	Extracellular matrix protein 1	Multifunctional, protease binding	6 v 15	-0.94	0.011
Q9NQ76	MEPE	Matrix extracellular phosphoglycoprotein	Post-translational protein phosphorylation	10 v 5	1.29	0.011
Q92820	GGH	Gamma-glutamyl hydrolase	Hydrolase	27 v 46	-0.50	0.011
Q4TZM4	HBB	Haemoglobin beta chain	Oxygen transport	5 v 7	1.47	0.013
P02750	LRG1	Leucine-rich alpha-2-glycoprotein	Extracellular protein	25 v 46	-0.51	0.014
Q8NEJ1	Unknown	Uncharacterized protein	Unknown	4 v 7	-0.87	0.016
P09668	CTSH	Pro-cathepsin H	Lysosomal protein degradation	11 v 15	-0.80	0.016
Q02818	NUCB1	Nucleobindin-1	Calcium/G-protein binding	16 v 39	-0.54	0.017
Q9NZP8	C1RL	Complement C1r subcomponent-like protein	Serine protease, complement pathway	25 v 45	-0.50	0.017

P51170	SCNN1G	Amiloride-sensitive sodium channel subunit gamma	Ion channel	2 v 9	0.91	0.018
Q9H497	TOR3A	Torsin-3A	Nucleotide binding	5 v 6	-2.00	0.018
A0A0X9UWL5	Unknown	GCT-A5 light chain variable region (Fragment)	Unknown	22 v 46	-0.41	0.019
P09603	CSF1	Macrophage colony-stimulating factor 1	Cytokine	26 v 39	0.45	0.019
Q9H665	IGFLR1	IGF-like family receptor 1	Cell membrane receptor IGF	24 v 34	-0.36	0.019
O00560	SDCBP	Syntenin-1	Multifunctional protein trafficker	2 v 24	0.49	0.020
Q01469	FABP5	Fatty acid-binding protein 5	Lipid metabolism	5 v 20	-0.75	0.024
P02766	TTR	Transthyretin	Thyroid hormone-binding protein	26 v 41	0.53	0.026
P69905	HBA1; HBA2	Haemoglobin subunit alpha	Oxygen transport	10 v 37	0.78	0.026
Q9UQV4	LAMP3	Lysosome-associated membrane glycoprotein 3	Adaptive immunity	16 v 34	-0.51	0.027
P08185	SERPINA6	Corticosteroid-binding globulin	Glucocorticoid/progestin transport	19 v 34	-0.55	0.028
P13598	ICAM2	Intercellular adhesion molecule 2	Leukocyte adhesion	8 v 18	0.75	0.028
P52848	NDST1	Bifunctional heparan sulphate N-deacetylase/N-sulfotransferase 1	Multifunctional enzyme/immune response	26 v 44	-0.76	0.029
Q99674	CGREF1	Cell growth regulator with EF hand domain protein 1	Cell adhesion, cell cycle	7 v 11	0.83	0.031
O15370	SOX12	Transcription factor SOX-12	Transcription factor	4 v 7	-1.72	0.034
Q01459	CTBS	Di-N-acetylchitobiase	Glycosidase	19 v 45	-0.34	0.035
Q9UL85	Unknown	Myosin-reactive immunoglobulin kappa chain variable region (Fragment)	Unknown	13 v 33	-1.01	0.035
P42785	PRCP	Lysosomal Pro-X carboxypeptidase	Carboxypeptidase	6 v 28	-0.86	0.036
Q9Y646	CPQ	Carboxypeptidase Q	Carboxypeptidase	5 v 7	-1.34	0.036
P60022	DEFB1	Beta-defensin 1	Bactericidal activity	11 v 20	-0.76	0.036
P14384	CPM	Carboxypeptidase M	Carboxypeptidase	20 v 30	-0.38	0.038
P07998	RNASE1	Ribonuclease pancreatic	Endonuclease	27 v 46	0.42	0.038
P31025	LCN1	Lipocalin-1	Transport protein	19 v 16	-0.61	0.039
Q96RR4	CAMKK2	Calcium/calmodulin-dependent protein kinase 2	Serine/threonine-protein kinase	10 v 15	-1.32	0.044
P24821	TNC	Tenascin	Cell adhesion	27 v 46	0.23	0.044
P12111	COL6A3	Collagen alpha-3(VI) chain	Cell adhesion	27 v 46	-0.41	0.045
Q7Z3B1	NEGR1	Neuronal growth regulator 1	Cell adhesion	13 v 25	-0.47	0.047
Q5T1S8	NCMAP	Noncompact myelin-associated protein	Myelin formation	27 v 46	-0.28	0.047
P04083	ANXA1	Annexin A1	Inflammatory/immune response	7 v 20	1.31	0.047
Q96PX8	SLITRK1	SLIT and NTRK-like protein 1	Synaptogenesis	3 v 13	-0.85	0.049
S6BGE0	Unknown	IgG H chain	Immune response	8 v 20	0.44	0.049

Table 4-4: Significantly altered urine protein abundances between pPRISm vs Term-born children

Gene Name	Protein Name	Protein Function	Replicates	Log$_2$FC	p-value	Inflammation of body cavity	Apoptosis myeloid cells	Quantity of leucocytes	Quantity of T-lymphocytes	Quantity of CD4+ T-lymphocytes	Quantity of CD8+ T-lymphocytes	Accumulation of neutrophils
— pPRISm vs PT$_c$ (n=27 vs 112)												
GLA	Alpha-galactosidase A	Lipid metabolism	2 v 6	-1.06	0.0001			●				
CLEC11A	C-type lectin domain family 11 member A	Osteogenesis	3 v 15	-1.39	0.001			●				
NAGLU	Alpha-N-acetylglucosaminidase	Glycosidase/ hydrolase	25 v 107	-0.34	0.011	●						
PGLYRP1	Peptidoglycan recognition protein 1	Innate immunity	27 v 111	-0.40	0.012	●		●	●	●		
DNASE1	Deoxyribonuclease-1	Serum endonuclease	27 v 111	-0.53	0.013	●		●	●	●	●	
MYH9	Myosin-9	Cell adhesion/shape	4 v 8	0.49	0.014	●						
SERPINA3	Alpha-1-antichymotrypsin	Serine protease inhibitor	25 v 100	-0.36	0.014	●	●					
CTSV	Cathepsin L2	Thiol protease	16 v 73	-0.33	0.015	●		●	●	●		
AGT	Angiotensinogen (Serpin A8)	Regulation of blood pressure	4 v 28	0.81	0.019	●		●	●		●	
ANXA1	Annexin A1	Inflammatory/immune response	7 v 29	1.46	0.026	●	●	●	●	●	●	
CLEC4G	C-type lectin domain family 4 member G	Substrate binder	27 v 106	-0.36	0.028	●		●	●	●	●	
SCGB1A1	Uteroglobin	Phospholipase A2 inhibitor	5 v 33	-1.04	0.032	●		●				
MGAM	Maltase-glucoamylase	Carbohydrate metabolism	27 v 111	-0.35	0.038	●						
B2M	Beta-2-microglobulin	Class I major histocompatibility complex	27 v 111	-0.60	0.040	●		●	●		●	
ANPEP	Aminopeptidase N	Aminopeptidase	23 v 103	-0.37	0.040		●	●	●	●		
CD14	Monocyte differentiation antigen CD14	Inflammatory/immune response	27 v 111	-0.39	0.045	●	●					

POLD vs PT$_c$ (n=21 vs 112)												
MMP9	Matrix metalloproteinase-9	Collagen degradation, leukocyte migration	21 v 105	0.30	0.018							●
AGT	Angiotensinogen (Serpin A8)	Regulation of blood pressure	5 v 28	0.55	0.030							●
S100A8	Protein S100-A8	Inflammatory/immune response	21 v 107	0.69	0.034							●
CTSC	Dipeptidyl peptidase 1 (Cathepsin C)	Thiol protease	19 v 77	0.31	0.038							●

Table 4-5: Significantly altered biological processes in pPRISM & POLD vs PT$_c$ identified by IPA software.
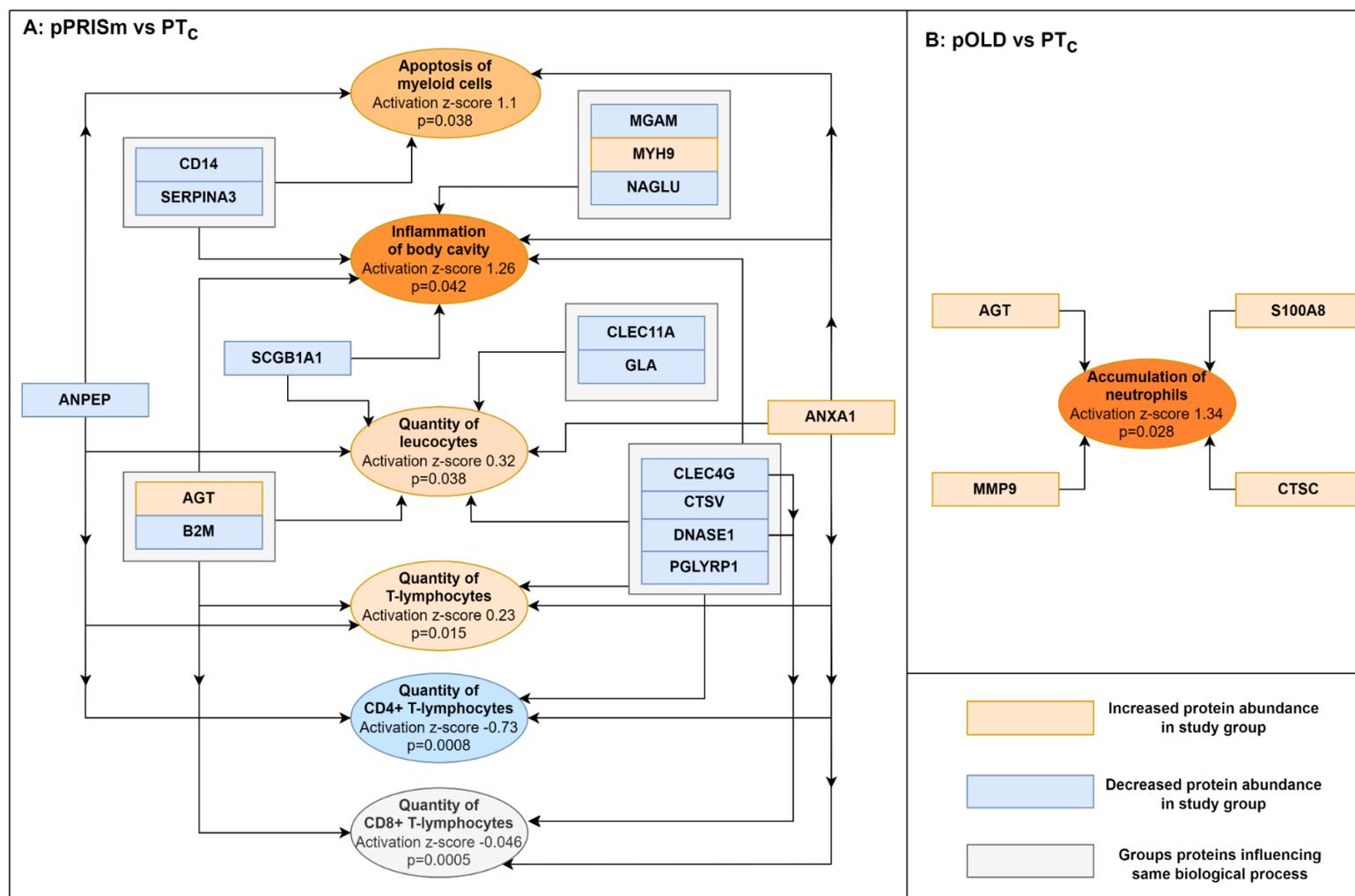
Figure 4-13: Urine proteins linked with significantly altered biological processes by IPA software within lung function phenotypes (pPRISm and POLD compared to PT$_c$).

POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PT$_c$: Preterm-born controls

Protein(s) included in model	Replicates	AUC (95% CI)	p-value	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
pPRISm vs PT$_c$ (n=27 v 112)							
DNASE1	27 v 112	0.66 (0.55, 0.78)	0.004	0.56 (0.37, 0.74)	0.72 (0.64, 0.80)	0.33 (0.19, 0.46)	0.87 (0.80, 0.94)
PGLYRP1	27 v 112	0.64 (0.53, 0.75)	0.012	0.93 (0.83, 1.00)	0.32 (0.23, 0.40)	0.25 (0.16, 0.33)	0.95 (0.87, 1.02)
B2M	27 v 112	0.63 (0.51, 0.75)	0.020	0.59 (0.41, 0.78)	0.66 (0.60, 0.75)	0.30 (0.18, 0.42)	0.87 (0.80, 0.94)
SERPINA3	25 v 100	0.66 (0.53, 0.79)	0.007	0.48 (0.28, 0.68)	0.82 (0.75, 0.90)	0.40 (0.23, 0.58)	0.86 (0.79, 0.93)
DNASE1 + PGLYRP1 + B2M + SERPINA3	25 v 100	0.73 (0.61, 0.84)	<0.001	0.80 (0.64, 0.96)	0.73 (0.64, 0.82)	0.43 (0.28, 0.57)	0.94 (0.88, 0.99)
POLD vs PT$_c$ (n=21 v 112)							
S100A8	21 v 107	0.64 (0.52, 0.76)	0.021	0.82 (0.66, 0.98)	0.50 (0.40, 0.59)	0.25 (0.15, 0.35)	0.93 (0.86, 1.00)
MMP9	21 v 105	0.64 (0.51, 0.77)	0.023	0.48 (0.26, 0.69)	0.76 (0.68, 0.84)	0.29 (0.14, 0.44)	0.88 (0.81, 0.95)
CTSC	19 v 77	0.66 (0.53, 0.79)	0.015	0.68 (0.48, 0.89)	0.65 (0.54, 0.76)	0.33 (0.18, 0.47)	0.89 (0.81, 0.97)
S100A8 + MMP9 + CTSC	19 v 77	0.76 (0.63, 0.90)	<0.001	0.84 (0.68, 1.00)	0.61 (0.50, 0.72)	0.35 (0.21, 0.49)	0.94 (0.87, 1.00)

Table 4-6: ROC Analysis of high replicate urine proteins implicated in related biological functions by IPA software.

POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PT$_c$: Preterm-born controls. AUC: Area Under the Curve. PPV: Positive Predictive Value. NPV: Negative Predictive Value.

Protein(s) included in model	Replicates	AUC (95% CI)	p-value	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
pPRISm vs PT$_c$ (n=27 v 112)							
DNASE1	27 v 112	0.62 (0.51, 0.74)	0.024	0.56 (0.37, 0.74)	0.69 (0.61, 0.78)	0.31 (0.18, 0.44)	0.87 (0.79, 0.94)
PGLYRP1	27 v 112	0.61 (0.53, 0.70)	0.005	0.54 (0.40, 0.67)	0.67 (0.61, 0.73)	0.28 (0.20, 0.37)	0.86 (0.80, 0.91)
B2M	27 v 112	0.58 (0.46, 0.71)	0.91	0.59 (0.41, 0.78)	0.64 (0.55, 0.73)	0.29 (0.17, 0.40)	0.87 (0.79, 0.94)
SERPINA3	25 v 100	0.63 (0.49, 0.76)	0.97	0.48 (0.28, 0.68)	0.80 (0.72, 0.88)	0.38 (0.21, 0.54)	0.86 (0.79, 0.93)
DNASE1 + PGLYRP1 + B2M + SERPINA3	25 v 100	0.65 (0.52, 0.78)	0.010	0.76 (0.59, 0.93)	0.66 (0.57, 0.75)	0.36 (0.23, 0.49)	0.92 (0.85, 0.98)
POLD vs PT$_c$ (n=21 v 112)							
S100A8	21 v 107	0.59 (0.47, 0.72)	0.92	0.82 (0.66, 0.98)	0.45 (0.35, 0.54)	0.23 (0.14, 0.33)	0.92 (0.85, 0.99)
MMP9	21 v 105	0.59 (0.45, 0.73)	0.90	0.38 (0.17, 0.59)	0.84 (0.77, 0.91)	0.32 (0.14, 0.50)	0.87 (0.81, 0.94)
CTSC	19 v 77	0.59 (0.46, 0.73)	0.90	0.58 (0.36, 0.80)	0.66 (0.56, 0.77)	0.30 (0.15, 0.45)	0.86 (0.78, 0.95)
S100A8 + MMP9 + CTSC	19 v 77	0.72 (0.57, 0.86)	0.002	0.47 (0.25, 0.70)	0.91 (0.85, 0.97)	0.56 (0.32, 0.81)	0.88 (0.80, 0.95)

Table 4-7: ROC Analysis (using Leave-One-Out Cross Validation model) of high replicate urine proteins implicated in related biological functions by IPA software.

POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PTc: Preterm-born controls. AUC: Area Under the Curve. PPV: Positive Predictive Value. NPV: Negative Predictive Value.
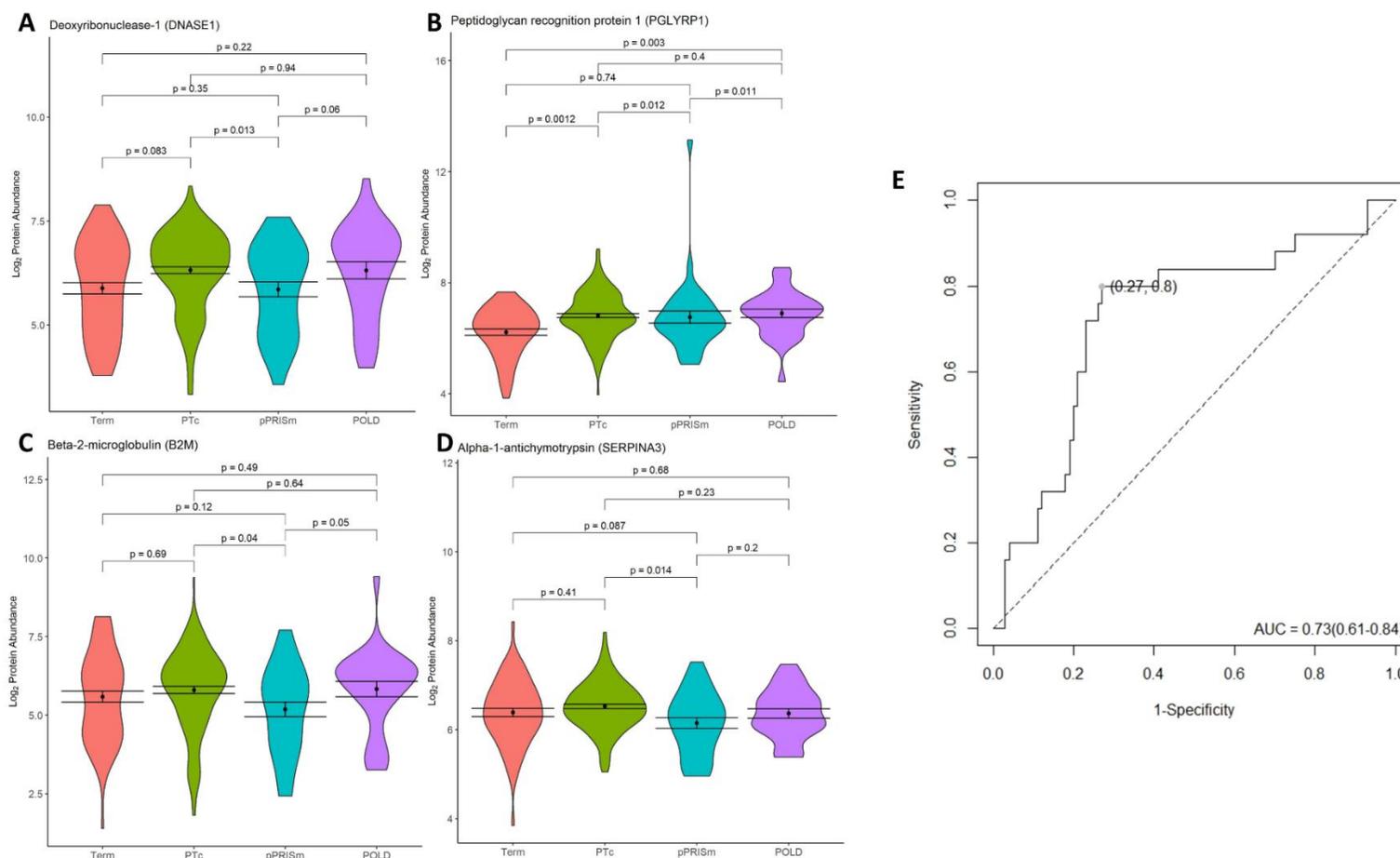
Figure 4-14: Significantly altered urine protein abundances in pPRISm vs PT$_c$ comparisons, showing violin plots for (A) DNASE1, (B) PGLYRP1, (C) B2M, and (D) SERPINA3, including comparisons with pPRISm and Term groups. (E) ROC Curve analysis for DNASE1, PGLYRP1, B2M and SERPINA3 in combination for pPRISm vs PT$_c$.

For violin plots, black dot represents mean, bars standard error of the mean. p-values given for between group comparisons. POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PT$_c$: Preterm-born controls. AUC: Area under the curve.
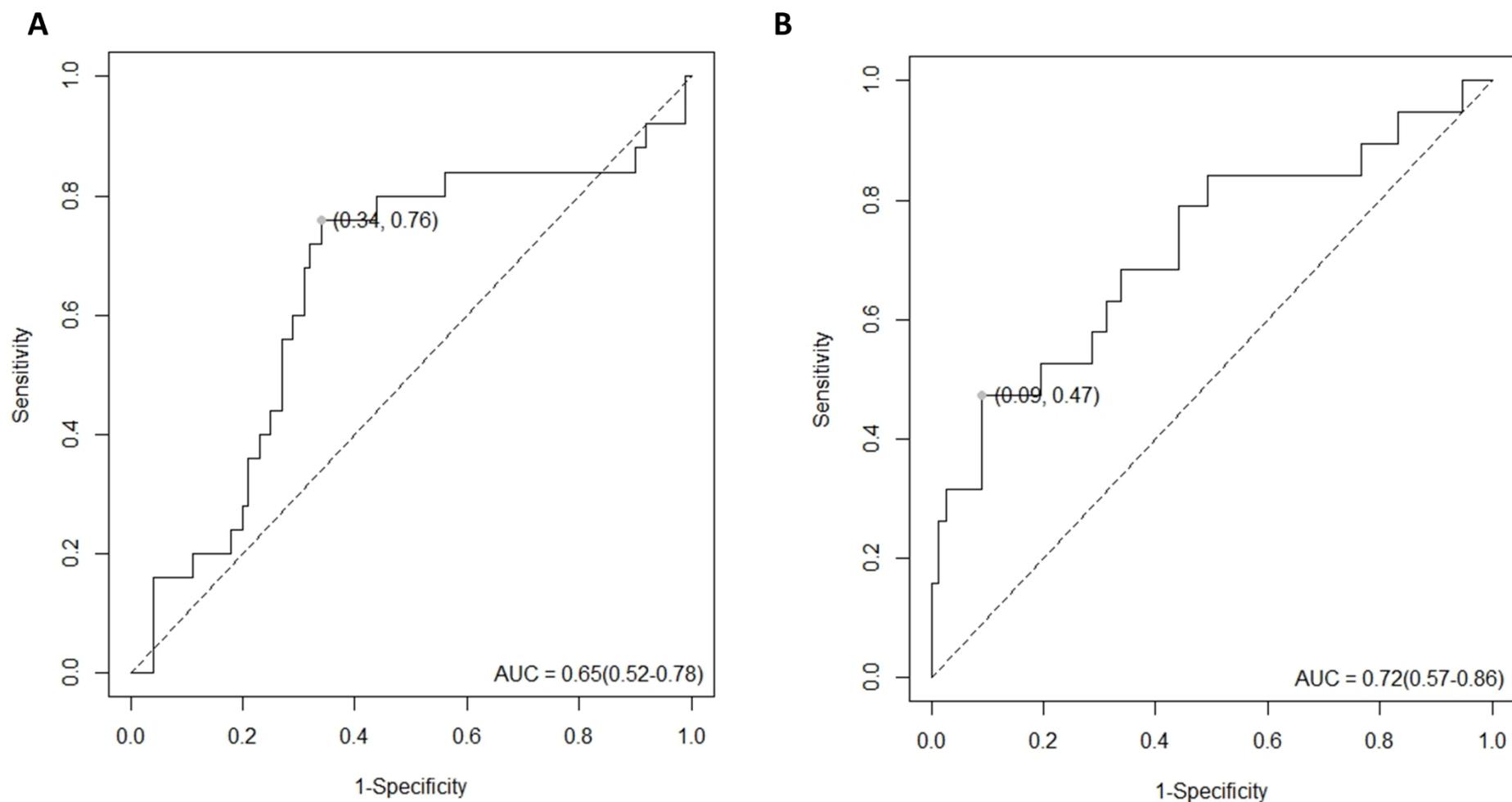
Figure 4-15: ROC Curves (using Leave-One-Out Cross Validation model) of protein groups with highest AUC values (A) pPRISm vs PT$_c$ including DNASE1 and PGLYRP1, (B) POLD vs PT$_c$ including S100A8, MMP9 and CTSC.

POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PT$_c$: Preterm-born controls

Univariable Models												
Variable	**DNASE1**			**PGLYRP1**			**B2M**			**SERPINA3**		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**
Sex, ref=Male	0.14	0.17	0.39	0.21	0.13	0.11	0.14	0.22	0.52	0.13	0.10	0.22
Age at testing, years	-0.07	0.06	0.23	0.05	0.05	0.32	-0.08	0.08	0.30	0.03	0.04	0.44
IUGR ref=No IUGR	-0.11	0.26	0.68	-0.34	0.21	*0.09*	-0.28	0.34	0.41	0.15	0.17	0.40
BPD ref=No BPD	-0.32	0.18	*0.08*	-0.24	0.14	0.10	-0.67	0.23	**0.005***	0.07	0.12	0.54
pPRISm ref=PT$_c$	-0.53	0.21	**0.011***	-0.40	0.16	**0.016***	-0.60	0.27	**0.027***	-0.37	0.13	**0.005***
Multivariable Models												
IUGR ref=No IUGR	-	-	-	-0.38	0.20	*0.06*	-	-	-	Not taken forward for multivariable model		
BPD ref=No BPD	-0.33	0.18	*0.06*	-	-	-	-0.69	0.23	**0.003***			
pPRISm ref=PT$_c$	-0.55	0.20	**0.008***	-0.42	0.16	**0.011***	-0.63	0.26	**0.018***			

Table 4-8: Univariable and multivariable linear regression analysis of early and current life factors and proteins of interest in pPRISm compared to PTc

SE: Standard error; IUGR: Intrauterine growth restriction; BPD: Bronchopulmonary dysplasia, pPRISm: prematurity-associated preserved ratio-impaired spirometry, PTc: preterm-born controls

Bold*: p value <0.05 Italic: p value <0.1. Dashes indicate variables that had p ≥0.1 on univariable analysis and therefore not included in multivariable model. Multivariable models only created where ≥2 univariable models had a p-value <0.1

4.3.4 Comparison between the POLD group with preterm- and term-born control groups

The POLD group had several significant differences on comparison to PT_c (Table 4-1) including; wheeze-ever (85.7% vs 57.1%, p=0.027), asthma (52.4% vs 18.8%, p=0.001), IUGR (38.1% vs 13.4%, p=0.006), positive BDR (57.1% vs 4.8%, p=<0.001) and highest FE_{NO} measurement (30.7 vs 17.2ppb, p=0.008), with a significantly higher proportion of POLD having a FE_{NO}>35ppb (33.3% vs 10.9%, p=0.012). When compared to pPRISm (Table 4-1), POLD had significantly higher rates of wheeze-ever (85.7% vs 55.6%, p=0.025) and IUGR (38.1% vs 7.4%, p=0.009).

44 (6.4%) proteins had a significantly different abundance when compared to PT_c (Figure 4-12; Table 4-9), and 70 (10.1%) had a significantly different abundance when compared to term-born subjects (Figure 4-12, Table 4-10). 18 proteins were common between the two comparisons. IPA linked four significantly altered proteins in POLD compared to PT_c to *Accumulation of neutrophils* (p=0.028); AGT, CTSC, MMP9, S100A8 (Table 5; Figure 13). IPA linked eight significantly altered proteins in POLD compared to Term-born with *Cellular infiltration by macrophages* (p=0.011); AGT, PLAU, C3, MMP9, CSF1, PROCR, IL6ST, PRCP.

ROC analysis (Table 4-6; Figure 4-16) demonstrated that S100A8, MMP9 and CTSC in combination had the highest predictive ability for identifying POLD from PT_c (AUC 0.76 (0.63 – 0.90), sensitivity 0.84 (0.68, 1.00), specificity 0.61 (0.50, 0.72), p=<0.001). Using the LOOCV model, S100A8, MMP9 and CTSC in combination performed similarly (AUC 0.72 (0.57 – 0.86), p=0.002) (Table 4-7; Figure 4-15). Results from univariable and multivariable linear regression modelling for these proteins are given in Table 4-11: Univariable and multivariable linear regression analysis of early and current life factors and urinary proteins of interest in POLD compared to PTc. No other early or current life factors were significantly associated with

S100A8 and CTSC abundance in univariable models. A history of BPD was significantly associated with MMP9 abundance in univariable modelling (p=0.017) and remained significant in the multivariable model BPD (p=0.017), along with POLD (p=0.024).

UniProt Accession	Gene Name	Protein Name	Protein Function	POLD vs PTc n = 23 v 111		
				Replicates	Log$_2$FC	p
B2R582	Unknown	cDNA, FLJ92374, highly similar to CLEC3B, mRNA	Unknown	2 v 12	-1.59	0.000003
Q9UHF0	TAC3	Tachykinin-3	Neuropeptide	3 v 22	0.64	0.0001
Q96IU4	ABHD14B	Protein ABHD14B	Acyltransferase	2 v 11	-1.01	0.0003
C9IYI1	DMKN	Dermokine	Cornified envelope assembly	10 v 39	1.06	0.002
P01709	IGLV2-8	Immunoglobulin lambda variable 2-8	Antigen recognition	23 v 105	0.35	0.002
P42785	PRCP	Lysosomal Pro-X carboxypeptidase	Carboxypeptidase	7 v 25	0.67	0.002
P31151	S100A7	Protein S100-A7 (Psoriasin)	Inflammatory/immune response	23 v 106	0.97	0.003
A0A494C0G5	AGRN	Agrin	Developmental protein	8 v 60	0.75	0.004
Q8WZ75	ROBO4	Roundabout homolog 4	Developmental protein, angiogenesis	23 v 111	0.43	0.004
P13473	LAMP2	Lysosome-associated membrane glycoprotein 2	Autophagy	20 v 98	0.51	0.006
Q58FF6	HSP90AB4P	Putative heat shock protein HSP 90-beta 4	Molecular chaperone, stress response	15 v 63	1.04	0.009
Q8NDA2	HMCN2	Hemicentin-2	Extracellular matrix	2 v 22	-0.68	0.009
Q92692	NECTIN2	Nectin-2	Modulator of T-cell signalling	15 v 64	0.38	0.012
Q99497	PARK7	Parkinson disease protein 7	Cell protection from oxidative stress	5 v 15	0.66	0.017
P08779	KRT16	Keratin, type I cytoskeletal 16	Cytokeratin	11 v 81	0.50	0.017
Q9UMS6	SYNPO2	Synaptopodin-2	Actin binding, cell migration	22 v 107	-0.49	0.017
P61970	NUTF2	Nuclear transport factor 2	mRNA transport	16 v 87	0.37	0.017
Q15274	QPRT	Nicotinate-nucleotide pyrophosphorylase	Pyridine nucleotide biosynthesis	3 v 11	1.40	0.018
P14780	MMP9	Matrix metalloproteinase-9	Collagen degradation, leukocyte migration	21 v 105	0.30	0.018
Q96RR4	CAMKK2	Calcium/calmodulin-dependent protein kinase 2	Serine/threonine-protein kinase	8 v 42	1.37	0.020
P41181	AQP2	Aquaporin-2	Fluid balance	14 v 70	0.63	0.021
P15289	ARSA	Arylsulfatase A	Lipid metabolism	19 v 76	0.60	0.021
P05543	SERPINA7	Thyroxine-binding globulin	Thyroid hormone transport protein	19 v 96	0.55	0.023
O95865	DDAH2	N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	Regulation of nitric oxide production	2 v 7	0.44	0.023
P31944	CASP14	Caspase-14	Epidermal differentiation	18 v 87	0.46	0.024
Q969E1	LEAP2	Liver-expressed antimicrobial peptide 2	Antimicrobial	18 v 99	1.14	0.024
P29966	MARCKS	Myristoylated alanine-rich C-kinase substrate	Actin/Calmodulin-binding	2 v 8	-0.45	0.025
P25815	S100P	Protein S100-P	Calcium signalling	15 v 85	0.42	0.027
P05091	ALDH2	Aldehyde dehydrogenase, mitochondrial	Oxidoreductase	4 v 26	0.43	0.029
P01019	AGT	Angiotensinogen	Regulation of blood pressure	5 v 28	0.55	0.030
P07998	RNASE1	Ribonuclease pancreatic	Endonuclease	23 v 111	0.32	0.030
Q96HD1	CRELD1	Protein disulfide isomerase	Isomerase	2 v 16	0.55	0.031

P05109	S100A8	Protein S100-A8	Inflammatory/immune response	22 v 107	0.69	0.034
P00738	HP	Haptoglobin	Haemoglobin recycling	23 v 111	0.55	0.035
P39059	COL15A1	Collagen alpha-1(XV) chain	Stabilizes microvessels/muscle cells	14 v 74	0.39	0.035
Q03154	ACY1	Aminoacylase-1	Hydrolase	4 v 18	0.65	0.035
P53634	CTSC	Dipeptidyl peptidase 1 (Cathepsin C)	Thiol protease	19 v 77	0.31	0.038
Q9Y4C0	NRXN3	Neurexin-3	Neuronal cell adhesion/recognition	16 v 50	0.46	0.038
Q96K68	Unknown	cDNA FLJ14473 fis, SNC73 mRNA	Unknown	23 v 111	0.28	0.042
O75882	ATRN	Attractin	Inflammatory response	23 v 97	0.30	0.043
Q9NZH0	GPRC5B	G-protein coupled receptor family C group 5 member B	Unknown	16 v 86	0.37	0.045
P43652	AFM	Afamin	Protein transport	23 v 111	-0.27	0.047
P68032	ACTC1	Actin, alpha cardiac muscle 1	Cytoskeleton	22 v 107	0.24	0.048
P51654	GPC3	Glypican-3	Protease inhibitor	9 v 49	-0.54	0.049

Table 4-9: Significantly altered urine protein abundances between POLD vs PT$_c$

UniProt Accession	Gene Name	Protein Name	Protein Function	POLD vs Term n = 23 v 46		
				Replicates	Log$_2$FC	p
Q9HCU0	CD248	Endosialin	Calcium binding/extra-cellular matrix	23 v 46	0.33	0.0003
O15370	SOX12	Transcription factor SOX-12	Transcription factor	5 v 7	-2.05	0.0003
Q16610	ECM1	Extracellular matrix protein 1	Multifunctional, protease binding	6 v 15	-1.49	0.0004
P69905	HBA1; HBA2	Haemoglobin subunit alpha	Oxygen transport	11 v 37	1.31	0.0004
C9IYI1	DMKN	Dermokine	Cornified envelope assembly	10 v 15	1.21	0.001
Q08380	LGALS3BP	Galectin-3-binding protein	Cell adhesion	23 v 46	-0.58	0.001
Q8WZ75	ROBO4	Roundabout homolog 4	Developmental protein, angiogenesis	23 v 46	0.60	0.001
Q14050	COL9A3	Collagen alpha-3(IX) chain	Extracellular matrix	8 v 7	0.70	0.001
A0A0S2Z4G4	TPM3	Tropomyosin 3 isoform 1 (Fragment)	Actin-binding	2 v 12	0.64	0.002
Q9UNN8	PROCR	Endothelial protein C receptor	Blood coagulation	23 v 46	0.49	0.003
O75594	PGLYRP1	Peptidoglycan recognition protein 1	Innate immunity	23 v 45	0.60	0.003
P98160	HSPG2	Basement membrane-specific heparan sulphate proteoglycan core protein	Basement membrane/angiogenesis	23 v 46	0.28	0.003
P15151	PVR	Poliovirus receptor	Natural killer cell adhesion/activation	14 v 22	0.93	0.004
P40189	IL6ST	Interleukin-6 receptor subunit beta	Cytokine binding	2 v 2	1.30	0.004
P61970	NUTF2	Nuclear transport factor 2	mRNA transport	16 v 41	0.64	0.004
P17936	IGFBP3	Insulin-like growth factor-binding protein 3	Growth regulation	16 v 31	1.39	0.005
U6FVB0	CD74-Ntrk1 fusion gene	Tyrosine-protein kinase receptor	Tyrosine-protein kinase	19 v 39	-0.51	0.005
Q16270	IGFBP7	Insulin-like growth factor-binding protein 7	Growth regulation	23 v 46	-0.80	0.006
P07998	RNASE1	Ribonuclease pancreatic	Endonuclease	23 v 46	0.45	0.008
D3DNU8	KNG1	Kininogen 1, isoform CRA_a	Vasodilation	23 v 46	-0.50	0.008
P09603	CSF1	Macrophage colony-stimulating factor 1	Cytokine	20 v 39	0.46	0.009
P19827	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	Serine protease inhibitor	7 v 24	0.78	0.009
P05154	SERPINA5	Plasma serine protease inhibitor	Serine protease inhibitor	19 v 46	-0.52	0.010
P61916	NPC2	NPC intracellular cholesterol transporter 2	Cholesterol metabolism	9 v 23	0.52	0.011
P00738	HP	Haptoglobin	Haemoglobin recycling	23 v 46	0.72	0.011
P07900	HSP90AA1	Heat shock protein HSP 90-alpha	Molecular chaperone	2 v 12	0.71	0.012
P01024	C3	Complement C3	Innate immunity	19 v 46	-0.35	0.012
Q9UQV4	LAMP3	Lysosome-associated membrane glycoprotein 3	Adaptive immunity	15 v 34	-0.49	0.013

Q4TZM4	HBB	Haemoglobin beta chain (Fragment)	Oxygen transport	4 v 7	2.21	0.013
P51654	GPC3	Glypican-3	Protease inhibitor	9 v 24	-0.77	0.013
Q8NDA2	HMCN2	Hemicentin-2	Extracellular matrix	2 v 9	-0.65	0.013
Q96S96	PEBP4	Phosphatidylethanolamine-binding protein 4	Extracellular exosome	12 v 18	0.74	0.013
A0A494C0G5	AGRN	Agrin	Developmental protein	8 v 5	0.62	0.013
Q14982	OPCML	Opioid-binding protein/cell adhesion molecule	Cell adhesion/opioid binding	15 v 45	-0.35	0.014
P01019	AGT	Angiotensinogen	Regulation of blood pressure	5 v 8	1.02	0.014
Q99985	SEMA3C	Semaphorin-3C	Developmental protein	8 v 14	2.13	0.015
P13473	LAMP2	Lysosome-associated membrane glycoprotein 2	Autophagy	20 v 46	0.50	0.015
P19652	ORM2	Alpha-1-acid glycoprotein 2	Transport protein	23 v 46	-0.46	0.016
P08779	KRT16	Keratin, type I cytoskeletal 16	Cytokeratin	11 v 38	0.52	0.017
P19835	CEL	Bile salt-activated lipase	Serine esterase, lipid degradation	15 v 22	0.57	0.018
O43866	CD5L	CD5 antigen-like (Apoptosis inhibitor expressed by macrophages)	Inflammatory response regulation	8 v 25	-0.79	0.018
P42785	PRCP	Lysosomal Pro-X carboxypeptidase	Carboxypeptidase	7 v 28	0.46	0.019
P41181	AQP2	Aquaporin-2	Fluid balance	14 v 46	0.69	0.020
P39059	COL15A1	Collagen alpha-1(XV) chain	Stabilizes microvessels/muscle cells	14 v 24	0.46	0.020
P0DOX7	Unknown	Immunoglobulin kappa light chain	Immune response	7 v 17	0.55	0.020
P52848	NDST1	Bifunctional heparan sulphate N-deacetylase/N-sulfotransferase 1	Multifunctional enzyme/immune response	23 v 44	-0.92	0.021
O15240	VGF	Neurosecretory protein VGF	Neurogenesis/neuroplasticity	22 v 46	-0.49	0.021
Q96K68	Unknown	cDNA FLJ14473 fis	Cell membrane	23 v 46	0.33	0.021
P00749	PLAU	Urokinase-type plasminogen activator	Plasminogen activator	13 v 31	-0.53	0.022
A0A193CHQ9	Unknown	10E8 heavy chain variable region (Fragment)	Unknown	19 v 43	0.53	0.023
Q9UMX5	NENF	Neudesin	Neurotrophic factor	14 v 21	0.48	0.025
A8TX70	COL6A5	Collagen alpha-5(VI) chain	Cell-binding protein	16 v 28	-0.52	0.025
P15144	ANPEP	Aminopeptidase N	Aminopeptidase	20 v 38	0.39	0.025
A0A140T9A1	COL11A2	Collagen alpha-2(XI) chain	Extracellular matrix	20 v 36	-0.55	0.029
Q6UVK1	CSPG4	Chondroitin sulphate proteoglycan 4	Cell proliferation/migration	23 v 46	0.58	0.031
P78380	OLR1	Oxidized low-density lipoprotein receptor 1	Inflammatory response	4 v 18	-0.44	0.033
P29508	SERPINB3	Serpin B3	Serine protease inhibitor	11 v 25	0.61	0.036
P98095	FBLN2	Fibulin-2	Binds to fibronectin	21 v 46	0.31	0.039
Q9NZH0	GPRC5B	G-protein coupled receptor family C group 5 member B	Unknown	16 v 34	0.43	0.039

Q9NRX4	PHPT1	14 kDa phosphohistidine phosphatase	Protein phosphatase	7 v 11	0.89	0.039
A0A2U8J953	IgH	Ig heavy chain variable region (Fragment)	Immune response	6 v 7	-1.25	0.041
P43121	MCAM	Cell surface glycoprotein MUC18	Cell adhesion	21 v 46	0.45	0.044
P02774	GC	Vitamin D-binding protein	Vitamin D transport/storage	8 v 14	-1.17	0.045
P01133	EGF	Pro-epidermal growth factor	Growth factor	23 v 46	-0.42	0.045
P01709	IGLV2-8	Immunoglobulin lambda variable 2-8	Antigen recognition	23 v 46	0.30	0.045
P33908	MAN1A1	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	Glycosidase/hydrolase	18 v 35	-0.39	0.045
P02766	TTR	Transthyretin	Thyroid hormone-binding protein	23 v 41	0.50	0.046
P14780	MMP9	Matrix metalloproteinase-9	Collagen degradation, leukocyte migration	21 v 46	0.26	0.046
Q12860	CNTN1	Contactin-1	Cell adhesion/signalling	17 v 45	-0.36	0.048
Q8NEJ1	Unknown	Immunoglobulin lambda-1 light chain-like	Cell membrane	3 v 7	-1.05	0.048

Table 4-10: Significantly altered protein abundances between POLD vs Term-born children.

Univariable Models									
Variable	**MMP9**			**S100A8**			**CTSC**		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**
Sex, ref=Male	0.15	0.10	0.15	-0.42	0.26	0.11	-0.02	0.13	0.88
Age at testing, years	-0.01	0.04	0.72	-0.04	0.09	0.65	0.01	0.05	0.78
IUGR ref=No IUGR	-0.02	0.13	0.90	0.32	0.34	0.36	-0.10	0.17	0.55
BPD ref=No BPD	-0.26	0.11	**0.017***	0.38	0.28	0.17	-0.03	0.15	0.85
POLD ref=PT$_c$	-0.30	0.13	**0.025***	-0.69	0.34	**0.043***	-0.31	0.16	*0.058*
Multivariable Models									
BPD ref=No BPD	-0.26	0.11	**0.017***	Not taken forward			Not taken forward		
POLD ref=PT$_c$	-0.30	0.13	**0.024***	for multivariable model			for multivariable model		

Table 4-11: Univariable and multivariable linear regression analysis of early and current life factors and urinary proteins of interest in POLD compared to PTc

SE: Standard error; IUGR: Intrauterine growth restriction; BPD: Bronchopulmonary dysplasia, POLD: prematurity-associated obstructive lung disease, PTc: preterm-born controls. Multivariable models only created where ≥2 univariable models had a p-vale <0.1
Bold*: p value <0.05 Italic: p value <0.1
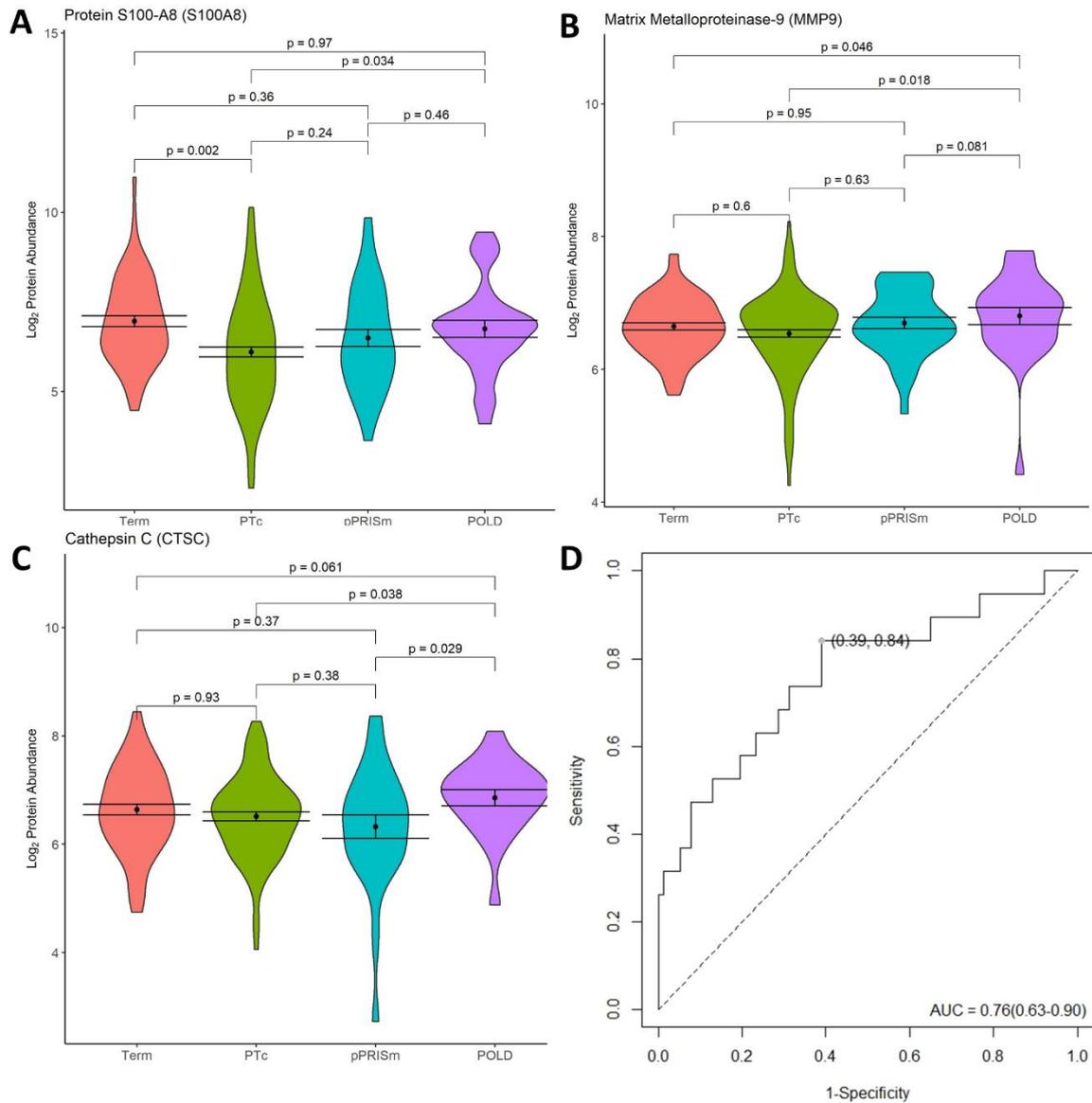
Figure 4-16: Significantly altered protein abundances in POLD vs PT$_c$ comparisons, showing violin plots for (A) S100A8, (B) MMP9 and (C) CTSC, including comparisons with pPRISm and Term groups. (D) ROC Curve analysis for S100A8, MMP9 and CTSC in combination for POLD vs PT$_c$.

For violin plots, black dot represents mean, bars standard error of the mean. p-values given for between group comparisons. POLD: Prematurity-related obstructive lung disease. pPRISm: Prematurity-related preserved ratio with impaired spirometry. PT$_c$: Preterm-born controls. AUC: Area under the curve

4.3.5 Comparison of urinary proteome before and after inhaled therapies.

Seven paired urine samples were available for those in the placebo treatment group, eleven from the ICS group and fourteen from the ICS/LABA group. Following the normalisation process which I had applied to the urinary proteome dataset, there were pre- and post RCT treatment samples available from five (71.4%) children in the placebo group, seven (63.6%) children in the ICS group and eight (57.1%) children in the ICS/LABA group. An exploratory analysis of the changes in the urinary proteome before after treatment revealed significant differences in eight proteins in the ICS group, ten in the ICS/LABA group and the majority of the altered proteins were seen in the placebo group, with nineteen significantly altered proteins (Table 4-12; Figure 4-17). There were no common significantly altered proteins between the ICS and ICS/LABA groups. Lipocalin-1 was detected in a minority of both the placebo and ICS/LABA groups samples; however, compared to post-treatment samples, lipocalin-1 was significantly elevated in pre-treatment samples in the placebo group, but significantly decreased in the pre-treatment samples in the ICS/LABA group. There were no other commonly detected significantly altered protein abundances between the placebo and ICS or ICS/LABA groups, nor were there any significantly altered proteins within the treatment groups that were common to the significantly altered proteins of interest in the POLD or pPRISm groups at baseline. Owing to the lack of a consistent pattern of significantly altered proteins between the treatment groups, with the majority of changes seen in the placebo group, and lack of correlation to proteins of interest identified in the POLD and pPRISm, I determined that enrichment analyses of these proteins would not be appropriate and may lead to unreliable results.

UniProt Accession	Gene Name	Protein Name	Replicates	Log$_2$FC	p-value
Placebo (n=5)					
Q96NY8	NECTIN4	Nectin-4	2 v 2	-0.51	<0.001
P05090	APOD	Apolipoprotein D	5 v 5	-0.67	0.001
Q16769	QPCT	Glutaminyl-peptide cyclotransferase	5 v 5	-0.60	0.002
Q92820	GGH	Gamma-glutamyl hydrolase	5 v 5	-0.61	0.008
P54802	NAGLU	Alpha-N-acetylglucosaminidase	5 v 5	0.48	0.009
Q16270	IGFBP7	Insulin-like growth factor-binding protein 7	5 v 5	-0.71	0.010
Q4G0X9	CCDC40	Coiled-coil domain-containing protein 40	3 v 3	0.83	0.012
P02656	APOC3	Apolipoprotein C-III	3 v 3	1.44	0.012
Q9BRK3	MXRA8	Matrix remodelling-associated protein 8	5 v 5	0.61	0.017
P05109	S100A8	Protein S100-A8	5 v 5	1.51	0.022
Q9UBC9	SPRR3	Small proline-rich protein 3	5 v 5	2.11	0.023
Q99715	COL12A1	Collagen alpha-1(XII) chain	5 v 5	-0.10	0.023
Q96DR8	MUCL1	Mucin-like protein 1	5 v 5	-0.70	0.031
P10451	SPP1	Osteopontin	5 v 5	-0.74	0.032
P05451	REG1A	Lithostathine-1-alpha	4 v 4	0.64	0.033
P01011	SERPINA3	Alpha-1-antichymotrypsin	4 v 4	-0.76	0.034
P31025	LCN1	Lipocalin-1	2 v 2	-0.16	0.035
Q9BRT3	MIEN1	Migration and invasion enhancer 1	4 v 4	1.24	0.035
Q9UMX5	NENF	Neudesin	2 v 2	0.55	0.035
P80188	LCN2	Neutrophil gelatinase-associated lipocalin	5 v 5	-0.44	0.037
P15586	GNS	N-acetylglucosamine-6-sulfatase	4 v 4	-0.48	0.049
ICS (n=7)					
Q8IUL8	CILP2	Cartilage intermediate layer protein 2	7 v 7	-0.39	0.007
P01034	CST3	Cystatin-C	7 v 7	-0.40	0.011
O60279	SUSD5	Sushi domain-containing protein 5	4 v 4	-1.43	0.015
P15328	FOLR1	Folate receptor alpha	7 v 7	-0.53	0.019
Q9NQ36	SCUBE2	Signal peptide, CUB and EGF-like domain-containing protein 2	2 v 2	-0.47	0.022
P0DJD8	PGA3	Pepsin A-3	3 v 3	0.92	0.028
Q8N114	SHISA5	Protein shisa-5	7 v 7	-0.43	0.044

P01344	IGF2	Insulin-like growth factor II	2 v 2	-0.76	0.047
ICS/LABA (n=8)					
P80188	LCN2	Neutrophil gelatinase-associated lipocalin	5 v 5	-0.73	0.004
P58400	NRXN1	Neurexin-1-beta	6 v 6	0.43	0.004
Q07654	TFF3	Trefoil factor 3	5 v 5	0.68	0.007
P78423	CX3CL1	Fractalkine	4 v 4	0.44	0.010
P00441	SOD1	Superoxide dismutase	8 v 8	0.47	0.012
P16284	PECAM1	Platelet endothelial cell adhesion molecule	2 v 2	1.59	0.025
P13521	SCG2	Secretogranin-2	7 v 7	-0.73	0.039
P19827	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	4 v 4	0.52	0.040
P17936	IGFBP3	Insulin-like growth factor-binding protein 3	4 v 4	-0.41	0.042
P31025	LCN1	Lipocalin-1	2 v 2	0.51	0.045

Table 4-12: Proteins significantly altered by treatment in the three RCT groups.
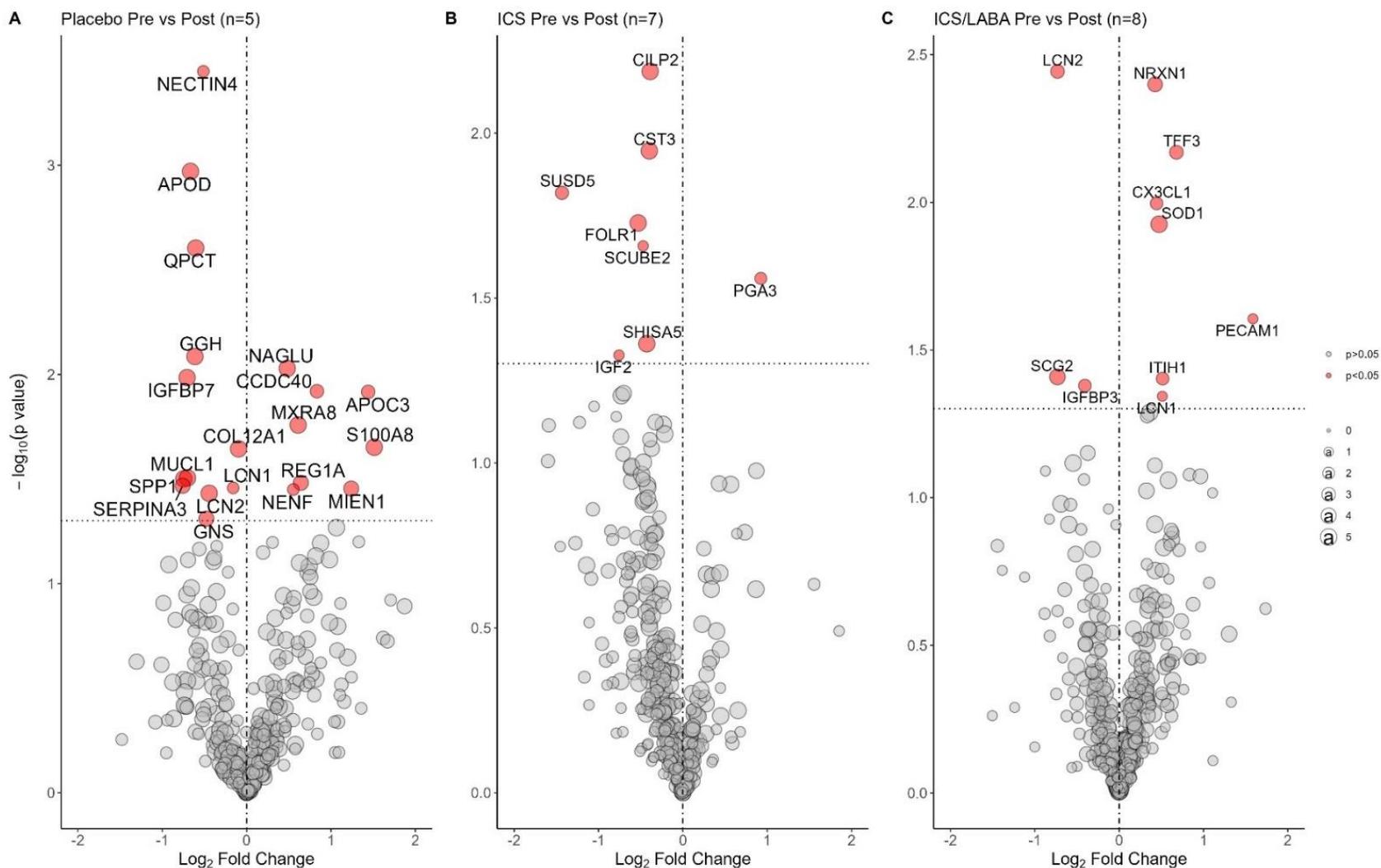
Figure 4-17: Volcano Plots demonstrating urine proteins which were significantly altered during the RCT treatments.

Vertical line represents Log2 fold change of 0. Horizontal line equivalent to a p-value of 0.05. Proteins with a significant difference between groups highlighted and labelled with respective gene name. Size of circle relative to replicate number. ICS: Inhaled corticosteroids. ICS/LABA: Inhaled corticosteroids in combination with long-acting β_2 agonists.

4.4 Discussion

In this novel exploratory study, I have characterised the urinary proteome of two major phenotypes of PLD, pPRISm and POLD, in one of the largest cohorts of preterm-born children studied to date. I have demonstrated persistent changes in the abundance of proteins related to inflammatory processes and immune-system function in preterm-born children with low lung function, several years after the initial pulmonary insult occurred in the neonatal period. In those with a pPRISm phenotype, there was evidence of multiple affected biological processes, with an ongoing, systemic inflammatory process, with a possible alteration in T-lymphocyte biology. Conversely, in those with a POLD phenotype, fewer related altered biological processes were identified, with a focus on myeloid cell lines including neutrophil and macrophage activity.

Historically, there has been a focus on highlighting the lung function deficits, and underlying mechanisms, in those preterm-born survivors with a history of BPD (Galderisi et al., 2019, Um-Bergström et al., 2022). As has been demonstrated previously in the RHiNO cohort (Cousins et al., 2023) and discussed in section 1.6.2, there is a high prevalence of lung function deficits within the preterm-born paediatric population (22.6% vs 9.2% of term-born children), a majority of whom do not have BPD. Whilst BPD in isolation is increasingly recognised as a poor predictor of future lung function (Hart et al., 2022, Corwin et al., 2018), it is clear that early lung injury does impact lung development (Um-Bergström et al., 2022, Simpson et al., 2018).

It is apparent there is greater complexity to PLD phenotypes, and a need to understand the biological mechanisms underlying current spirometry patterns in order to understand their pathogenesis, aid their identification and guide potential therapeutic interventions. For this reason, I have focussed these analyses by using current lung function phenotypes rather than

historical diagnoses. PRISm has recently been described in the adult population (Wan et al., 2021), and whilst there is a paucity of studies focussing on the paediatric age group, PRISm has been associated with increased respiratory symptoms, cardiovascular disease and increased all-cause mortality (Higbee et al., 2022, Wan et al., 2021). Within the RHiNO cohort, the pPRISm group have previously been demonstrated to have different association to bronchodilator response, fraction exhaled nitric oxide (FE$_{NO}$) and early/current life factors compared to POLD and PT$_c$ groups (Cousins et al., 2023). Yet there is little data available on underlying mechanisms.

This urinary proteomic data suggests multiple associations with systemic alterations in inflammatory and immune processes are present in the pPRISm group, with a likely increase in inflammation, quantities of leucocytes overall and T-lymphocytes in particular. Recent work has shown a relative decrease of CD4+ T-cells and increased CD8+ T-cells in brocheoalveolar lavage (BAL) fluid from young adults with former BPD, a similar finding to those with COPD. CD8+ T-cells were also negatively correlated with both FEV$_1$ and FEV$_1$/FVC (Um-Bergström et al., 2022). Adolescent survivors of severe BPD have also been noted to have an increase in bronchial wall CD8+ lymphocytes (Galderisi et al., 2019). A recent urine metabolomic study has linked early increases in proteins associated with leukocyte mediated immunity to the later development of BPD in infants born <29 weeks' gestation (Ahmed et al., 2022). There is a suggestion from my data that the quantity of CD4+ lymphocytes may be downregulated in the pPRISm group, of whom only 25.9% had a previous history of BPD. A relative increase in CD8+ T-cells number and function has also been associated with severity of COPD (Williams et al., 2021), and adult subjects with PRISm are known to be at heightened risk of developing COPD over time (Marott et al., 2021).

Four proteins (DNASE1, PGLYRP1, B2M and SERPINA3) showed good predictive ability for identifying pPRISm from PT_c in the ROC analysis. Deoxyribonuclease-1 (DNASE1) is a ubiquitous endonuclease which degrades the majority of circulating free DNA released from apoptosis and necrotic cell death, with DNASE1 deficiency being previously reported to be associated with autoimmune disease in animal models and humans (Keyel, 2017). Peptidoglycan recognition protein 1 (PGLYRP1), an innate proinflammatory and antibacterial protein, has been linked with asthma in animal models, with PGLYRP1-deficient mice exhibiting a decreased Th2/CD4+ response, with a less severe phenotype (Yao et al., 2013). Increased serum beta-2-microglobulin (B2M), the light chain of the class I major histocompatibility complex, has been linked with COPD disease progression, namely development of pulmonary fibrosis, alveolar wall thickening and decreased gas exchange capacity (Wu et al., 2020b). The anti-protease alpha-1-antichymotrypsin (SERPINA3) manipulates the immune and inflammatory response through inhibition of chymotrypsin and cathepsin G. Previous studies have identified increased SERPINA3 in serum from COPD subtypes associated with metabolic syndrome (Zhang et al., 2023), with genetic mutations resulting in SERPINA3 deficiency resulting in milder disease in patients with COPD (Sandford et al., 1998) and cystic fibrosis (CF) (Mahadeva et al., 2001). The reduced abundances I observed, of these four proteins in pPRISm were all significantly linked with a possible upregulation of inflammatory processes, with DNASE1, PGLYRP1 and B2M also being significantly linked with T-cell biology.

Significantly increased abundance of four proteins (MMP9, AGT, S100A8, CTSC) within the POLD group (when compared to PT_c) were linked with an increase in neutrophil accumulation, which is a reasonable hypothesis given the association of neutrophilic inflammation with wheezing in asthma (Ray and Kolls, 2017). Whether this is a specific phenotype of PLD or has similarities with neutrophilic asthma will need further investigation. Matrix

metalloproteinase-9 (MMP9) is a gelatinase protease, stored in neutrophils, involved in degradation of the extracellular matrix, enhancing inflammatory cell migration, activating proinflammatory cytokines and remodels the lung parenchyma (Davey et al., 2011). An increased abundance of MMP9 in respiratory samples has been linked with various lung diseases, including preterm neonates who develop BPD (Sweet et al., 2004), and in intubated paediatric patients with Acute Lung Injury (Kong et al., 2009). A recent urine metabolomic study demonstrated an early increase in MMP9 had a high predictive ability for development of BPD in extremely preterm infants (Ahmed et al., 2022). MMP9 had a significant association with BPD in the preterm-born cohort in univariable modelling, which remained in the multivariable regression model, along with a significant association with POLD. Studies from the RHiNO cohort have previously shown a significant association between BPD and the development of a POLD phenotype (Cousins et al., 2023). In older subjects, elevated serum MMP9 has been linked with COPD exacerbations (Wells et al., 2018) and FEV_1 decline in CF (Devereux et al., 2014).

Cathepsin C (CTSC) is a serine protease released by neutrophils that can result in increased tissue-degradation, being implicated in the pathophysiology of pneumonia and acute respiratory distress syndrome in mechanically ventilated adults (Seren et al., 2021). S100A8 is also associated with acute lung injury, being secreted by degranulating neutrophils and bronchial epithelium during infection/inflammation (Kotsiou et al., 2021). It has been shown to be increased in lung diseases resulting in tissue remodelling, including in bronchiolitis obliterans in children (Jerkic et al., 2020), and in adults with CF and COPD (Lorenz et al., 2008). These three proteins all have a role in tissue remodelling; recent publications from the RHiNO cohort using on hyperpolarised [129]Xe ventilation and diffusion MRI imaging have reported that the POLD group has significantly altered ventilation mechanics (Chan et al., 2023), which is likely to be related to tissue remodelling. MMP9, S100A8 and CTSC in combination had

good predictive ability for identifying the POLD group using ROC analysis. Whether these combinations of proteins have prospective predictive value for PLD phenotypes prior to the development of lung function deficits will require further work.

The analysis of the urinary proteome before and after inhaled therapies did not reveal a consistent pattern of proteome changes, with most altered protein abundances being observed in the placebo group. Overall, this part of the analysis appeared uninformative and unreliable. The reasons for this are unclear, however owing to the normalisation process I used to ensure robust comparisons between clinical and treatment groups, approximately one-third of all the RCT samples had to be excluded from the analysis. This may have limited the statistical power of the analysis to detect biologically meaningful changes between the groups. It is possible that differences were seen in the post-treatment EBC proteome analysis (as discussed in section 2.3.4) as this was sampling the lung more directly, and either inhaled therapies do not alter the systemic proteome (as reflected by the urine proteome), or twelve weeks of treatment is insufficient to detect significant alterations in systemic biology by this methodology. Future studies should aim to recruit larger number of participants to improve the discriminating ability of the analysis.

In this analysis, I have studied the urinary proteome. Whilst this is not a lung-specific sample type, it easily and non-invasively obtainable, and has been utilised in the study of respiratory diseases in neonates (Ahmed et al., 2022) and adults (Martelo-Vidal et al., 2022) previously. In addition, as urine lacks the same homeostatic controls as blood, proteome changes in urine may be detectable at an earlier stage of disease (Wu and Gao, 2015) which makes it an attractive sample type to study in preterm-born children, as they may be at a milder or pre-symptomatic stage of respiratory impairment, as their lung function continues to develop through adolescence into adulthood (Belgrave et al., 2018).

4.4.1 Strengths and limitations

The analysis presented in this chapter represents one of the largest proteomic analyses of urine in the paediatric population, and although lung dysfunction was present in approximately 30% of the preterm-born group, this is the first study to my knowledge that has examined the urinary proteome of this cohort. The regression modelling has demonstrated that many of the protein changes I have observed are primarily related to current lung function phenotype. I have used a robust TMT-methodology to quantify protein abundances and allow accurate comparisons between phenotypes, however there may have been proteins with low abundances/low TMT-tag counts that did not reach the limit of detection of the mass spectrometer.

To ensure accurate protein quantitation with Proteome Discoverer software I thoroughly examined and tested potential normalization strategies to overcome any dilutional effects on the urine samples analysed. At present there is no universally accepted proteomic analysis workflow, with differing analysis platforms (e.g. mass spectrometry, nuclear magnetic resonance, antibody-based platforms) and both labelled and unlabelled techniques, as different workflows lend themselves to analysing different biofluids and answering differing research questions (Cui et al., 2022, Pappireddi et al., 2019, Aslam et al., 2017). Owing to the variety of experimental designs, there is also no universally accepted normalisation method for proteomic workflows, and different analysis techniques will generate their own sets of limitations in this regard (Valikangas et al., 2018). This makes determining the appropriate normalisation method particularly important. This experiment utilised a TMT-based technology, which as I have previously discussed in section 1.5.1.1, allows for very accurate quantitation of individual proteins to allow for robust comparison of protein abundances between samples. Proteome Discoverer v2.1 is a commonly used software that was employed

to quantify and annotate the detected proteins, however my discussions with Thermo Fischer Scientific highlighted the limitations in the accuracy of the software when analysing data from samples with widely varying total protein load.

After thoroughly evaluating a range of commonly used extrinsic (including urinary creatinine and total protein content determined by two methods) normalisation methods, it became clear that within the technical limitations of the experimental design and Proteome Discoverer software, an intrinsic normalisation method using the median total protein abundance would generate the most reliable results for this analysis. This meant however that 64 (23.6%) of the samples that had been analysed by MS had to be excluded from the final statistical analysis. Whilst this ensured that a robust statistical comparison of the included samples was performed, a relatively high number of samples had to be excluded from the final analysis, and this may have resulted in important biological discoveries being undetectable. In addition, there were some significant differences between the included and excluded subjects in the POLD group. Of particularly note, there was a significantly lower rate of BPD in the POLD group whose samples were included in the analysis when compared to those excluded. This may have skewed the proteomic changes seen in this group, and given that previous work on the RHiNO cohort has shown a significant association between a history of BPD and a POLD phenotype in childhood (Cousins et al., 2023), as I have discussed in section 1.6.2.3, the proteome profile for POLD I have generated with this analysis may not accurately reflect the group as a whole.

Whilst this TMT-based methodology gives robust protein abundances for comparative purposes, it does not give absolute protein concentrations within a sample, which would need to be determined to directly apply this data clinically. Whilst this analysis lacks a

validation cohort, I am limited by the number of available large cohorts of preterm-born children who experienced a contemporary standard of neonatal care from which to sample.

4.5 Conclusions

In conclusion, in this chapter I have demonstrated distinct changes in the urinary proteome associated with the two recently described phenotypes of PLD; POLD and pPRISm. There was suggestion of proteins associated with the inflammatory and immune systems in the pPRISm group and of potential neutrophilic inflammation in the POLD group. I have also demonstrated potential predictive ability of combinations of proteins to identify the POLD and pPRISm phenotypes. Further work with specific targeting of these proteins is now required to confirm if these proteins can be used clinically to screen prospectively for preterm-born children at risk of future lung dysfunction, or whether they can be targeted therapeutically.

5   Metabolomic Analysis of Urine

5.1   Introduction:

As has previously been discussed in the introductory chapter (section 1.4), lung function impairments are known long-term consequences of preterm birth, including those with and without a neonatal diagnosis of bronchopulmonary dysplasia (BPD), also known as chronic lung disease of prematurity (CLD) (Kotecha et al., 2022b, Gibbons et al., 2023). Historically, studies have focused on pulmonary outcomes for those with BPD, however, as discussed in section 1.6.2, increasingly immature gestational age at birth and intra-uterine growth restriction (IUGR) appear to be more strongly associated with prematurity-associated lung disease (PLD) in childhood within the RHiNO cohort (Hart et al., 2022), who experienced a contemporary standard of neonatal care.

In Chapter 3, I explored the metabolome of EBC from subjects participating in the RHiNO cohort, finding metabolic changes associated with alterations of glutathione metabolism and the urine cycle, implying oxidative stress, in the children with a neonatal history of BPD (section 3.3.2). However, this analysis did not find associations between key metabolites of interest involved in these processes and spirometry parameters. Therefore, these analyses were extended to the study the urinary metabolome of samples taken during the RHiNO study, focusing on spirometry-based PLD phenotypes, including POLD and pPRISm (as described in section 1.6.2.3 and 4.2.2).

Whilst a proportion of individuals with PLD will respond to inhaled therapies (Goulden et al., 2021), the biological pathways implicated in the development of these PLD-associated phenotypes remain unclear, with most previous mechanistic work focusing on BPD diagnosed in the neonatal period (Um-Bergström et al., 2022) rather than current spirometry deficits. This limits the ability to accurately identify endotypes and target potential treatments. As discussed previously (section 4.1.1), urine lacks the same homeostatic mechanisms as blood, therefore systemic metabolite

changes accumulate and the urinary metabolome may show alterations prior to symptoms or histopathological changes, reflecting an earlier stage of pathogenesis (Wu and Gao, 2015). Therefore, whilst not a lung-specific sample, urine has been extensively used in metabolomic studies of respiratory disease, such as asthma and COPD, where it has been demonstrated to show metabolomic alterations before they occur in serum (Wang et al., 2013). Alterations in glutathione metabolism, lipid metabolism and lipid peroxidation have been implicated in severe asthma phenotypes and COPD (Moitra et al., 2023). The early urinary metabolome in preterm infants demonstrates specific changes, including increased myoinositol and taurine, that predict later development of BPD (Fanos et al., 2014), whilst the exhaled breath condensate from adolescent BPD survivors demonstrates distinct metabolite abnormalities possibly related to pulmonary surfactant composition and anti-inflammatory pathways (Carraro et al., 2015). However, how these patterns relate to current lung function of preterm-born children remains unclear.

5.1.1 Aims

The urinary metabolome of children with PLD has yet to be studied. Therefore, I performed an exploratory metabolomic analysis of urine taken from preterm-born, school-aged children with term-born matched controls, aiming to delineate the metabolic pathways underlying the PLD phenotypes of POLD and pPRISm.

5.2 Methods:

5.2.1 Participants:

This study was conducted on children recruited to the Respiratory Health Outcomes in Neonates study (RHiNO, EudraCT: 2015-003712-20) which has been extensively published (Goulden et al., 2021, Hart et al., 2022, Course et al., 2023b) and described previously in this thesis (sections 1.6, 2.2.1, 3.2.1 and 4.2.1).

5.2.2 Lung Function Assessment:

The same spirometry assessment as described in Chapter 4 (section 4.2.2) for the urinary proteome analysis was used to classify the PLD phenotypes for this urinary metabolome analysis. In brief, spirometry (Microloop, Care Fusion, UK) was performed by trained research nurses to ATS/ ERS guidelines (Miller et al., 2005) and normalised using GLI references (Quanjer et al., 2012). Respiratory medications were withheld prior to assessment (as described in 4.2.2) and children were free of respiratory infections for at least three weeks prior to testing. PLD phenotypes were categorised into POLD (FEV_1<LLN and FEV_1/FVC<LLN) and pPRISm (FEV_1<LLN and FEV_1/FVC≥LLN), as previously described (sections 1.6.2.3 and 4.2.2). Preterm-born control (PT_c) and Term-born children had FEV_1≥LLN. BPD was defined as oxygen-dependency at 28 days of age or greater for those born <32 weeks' gestation and at 56 days of age for those born ≥32 weeks' gestation) (Ehrenkranz et al., 2005). IUGR was defined as birthweight <10th percentile adjusted for sex and gestation (LMS Growth version 2.77, Medical Research Council, UK). Neonatal history was corroborated with medical records. Doctor-diagnosed asthma was self-reported by parents.

5.2.3 Urine Sampling:

Urine samples were obtained at the time of spirometry and immediately placed on ice. Samples were then aliquoted and stored at -80°C as soon as possible on the day of collection until further processing and analysis, as described below.

5.2.4 Metabolome analysis:

Urine samples were analysed using Gas Chromatography Time-of-Flight Mass Spectrometry (GCTOF-MS) at West Coast Metabolomics Centre (University of California, Davis) (Fiehn et al., 2008), who have previously published their analytical method (Fiehn et al., 2008), aimed at identifying constituents of primary metabolism (carbohydrates and sugar phosphates, amino acids, hydroxyl acids, free fatty acids, purines, pyrimidines, aromatics, and exposome-derived

chemicals). The analytical method and workflows used were similar to that of the EBC metabolome analysis, as described in section 3.2.3. A detailed description of the analytical method used for the urine samples is provided below.

To extract the metabolite content of the urine sample, 30μL of urine was mixed with 1ml of extraction solution (composed of acetonitrile, isopropanol and water in a 3:3:2 (v/v/v) ratio) on ice. Samples were vortexed for 10 minutes, agitated for 5 min at 4°C on an orbital mixing chilling/heating plate before centrifugation for 2 minutes at 14,000 rcf. The supernatant was divided into two 450μL aliquots, with one aliquot kept in reserve. 100μL was removed from the remaining aliquot for sample analysis with another 100μL removed for use in the pool sample. All processed samples were stored at -20°C pending analysis.

100μL of each sample was fractionated using an Agilent 6890 gas chromatograph (Agilent, Santa Clara, CA, USA), controlled using Leco ChromaTOF software v2.32 (LECO, St. Joseph, MI, USA), in a Rtx-5Sil MS column (Restek, Bellafonte, PA, USA) (30m length x 0.25mm internal diameter with 0.25μm film made of 95% dimethyl/5%diphenylpolysiloxane). Quality control (QC) samples comprised two method blanks (involving all the reagents and equipment used to control for laboratory contamination) and four calibration curve samples, which spanned one order of dynamic range and consisted of 31 pure reference compounds. Column temperature was maintained between 50-330°C, with a helium mobile phase. Injection volumes of 0.5μL were used, with injection temperatures starting at 50°C, ramped up to a maximum temperature of 250°C by 12°Cs^{-1}. Oven temperature program was set to 50°C for 1 min, then ramped at 20°C min^{-1} to 330°C, and held constant for 5 min. The analytical GC column was protected by a 10m long empty guard column which was cut by 20cm intervals whenever the reference mixture QC samples indicated problems caused by column contaminations. This sequence of column cuts has been validated, with no detrimental effects being detected with respect to peak shapes, absolute or relative

metabolite retention times or reproducibility of quantifications. This chromatography method yields retention and separation of primary metabolite classes (amino acids, hydroxyl acids, carbohydrates, sugar acids, sterols, aromatics, nucleosides, amines and miscellaneous compounds) with narrow peak widths of 2–3s and very good within-series retention time reproducibility of better than 0.2s absolute deviation of retention times. Automatic liner exchanges after each set of 10 injections were used, which reduces sample carryover for highly lipophilic compounds.

All spectra were acquired using a Leco Pegasus IV (LECO, St. Joseph, MI, USA) time-of-flight mass spectrometer, with unit mass resolution at 17 spectra s^{-1} from 80-500Da at -70eV ionization energy and 1800V detector voltage with a 230°C transfer line and a 250°C electron ion source. Raw data files were normalised to QC/pool samples using the systematic error removal by random forest (SERRF) method(Fan et al., 2019). Raw data files were processed and metabolites identified with the BinBase metabolomics database (Lai et al., 2018), using an algorithm based on the following: validity of chromatogram (<10 peaks with intensity>10^7 counts s^{-1}), unbiased retention index marker detection (MS similarity >800, validity of intensity range for high m/z marker ions), retention index calculation by 5th order polynomial regression. Spectra are cut to 5% base peak abundance and matched to database entries from most to least abundant spectra using the following matching filters: retention index window ±2,000 units (equivalent to about ±2s retention time), validation of unique ions and apex masses (unique ion must be included in apexing masses and present at >3% of base peak abundance), mass spectrum similarity must fit criteria dependent on peak purity and signal/noise ratios and a final isomer filter. Quantification of metabolites are reported as spectral peak height of the unique ion detected (m/z value) at the specific retention index. Peak heights are more precise for metabolites with low abundance than peak areas, due to the larger influence of baseline determinations on areas compared to peak heights. Raw data files

were processed, and metabolites annotated using the BinBase database (Lai et al., 2018) with a standardised algorithm.

5.2.5   Statistical analysis:

Demographics were compared using chi-squared or one-way ANOVA with Bonferroni correction as appropriate. Metabolite quantities were normalised using creatinine (as detected by MS), as recommended to account for dilutional effects (Li et al., 2022), $\log_{10}$ transformed and visually inspected for normality. Metabolites with mean and median peak intensities below the limit of detection were removed to ensure robust statistical comparisons. Fold changes between groups were calculated using mean metabolite quantity for each group. ANOVA with post-hoc Bonferroni correction was used to compare metabolite quantities between groups. Metabolite Set Enrichment Analysis (MSEA; identifying metabolic processes linked to significantly altered metabolites) was performed on all significantly altered metabolites between groups using the Small Molecule Pathways Database (SMPDB) (https://www.smpdb.ca), which is based upon the Human Metabolome Database (HMDB). Univariable and multivariable linear regression models identified significant associations between participant characteristics, spirometry values and metabolites of interest. $p < 0.05$ was considered statistically significant. I used R v4.0.4 (R Foundation for Statistical Computing, Austria), including R packages *"stringr", "dplyr", "ggplot2"* and *"ggpubr"*, as well as MetaboAnalyst v5.0 (www.metaboanalyst.ca) for all analyses in this chapter.

5.3 Results:

From 768 children (565 Preterm-born and 203 Term-born) recruited to RHiNO, urine was analysed from 292 participants; I excluded 1 sample from a PT_c subject as an outlier due to negligible overall metabolite detection. Demographics of the remaining 291 participants are given in Table 5-1.

Preterm-born children had higher rates of doctor-diagnosed asthma than the Term-born group (41 (20.8%) vs 6 (6.4%), p=<0.002). 51 (25.9%) of the Preterm-born subjects had BPD diagnosed in infancy (21 [41.2%] mild, 30 [58.8%] moderate/severe(Ehrenkranz et al., 2005)), and 48 (24.4%) had an FEV_1<LLN. Of those, 25 (52.1%) were classified as pPRISm and 23 (47.9%) as POLD. Paired urine samples were available from 28 RCT participants, 10 from the ICS group, 11 from the LICS/LABA group and 7 from the placebo group. 242 metabolites were detected and annotated in total, with 238 (98.4%) metabolites having mean and median abundances above the limit of detection across all samples (Table 8-3).

5.3.1 Comparisons between POLD and preterm- and term-control groups:

Comparison between the POLD group and PT_c group revealed several significant differences (Table 5-1) including increased wheeze-ever (82.6% vs 51.0%, p=0.009), asthma (39.1% vs 17.5%, p=0.017), IUGR (39.1% vs 12.1%, p=<0.001) and BPD (47.8% vs 22.2%, p=0.009). When compared to the pPRISm group (Table 1), POLD had higher wheeze-ever (82.6% vs 52.0%, p=0.025) and higher rates of IUGR (39.1% vs 8.0%, p=0.01).

Of 238 detected metabolites, 204 were detected in every sample from the POLD group. 49 (20.6%) metabolites were significantly altered when compared to PT_c (Figure 5-1;Table 5-2), and 69 (29.0%) when compared to Term-born children (Figure 5-1, Table 5-3), with 31 metabolites common between the two comparisons. Interestingly, all significantly altered metabolites were of lower quantity in the POLD group when compared with the PTc group, apart from beta-alanine, which was elevated (log2FC 0.55, p=0.047).

Variable	Term born (≥37/40) n = 94	Preterm born (≤34/40) n = 197	Preterm born Controls n = 149	POLD n = 23	pPRISm n = 25
Current Characteristics					
Sex (male), n(%)	50 (53.2)	108 (54.8)	83 (56.1)	9 (39.1)	16 (64.0)
Ethnicity (white), n(%)	91 (96.8)	190 (96.4)	143 (96.0)	22 (95.7)	25 (100)
Age at testing (years), mean (SD)	9.7 (1.2)	10.0 (1.2)	10.0 (1.2)	9.9 (1.4)	10.2 (1.2)
Weight (kg), mean (SD)	37.1 (10.8)	36.6 (10.3)	37.1 (10.1)	33.5 (10.8)	36.7 (10.7)
Body Mass Index (kg/m^2), mean (SD)	18.0 (3.4)	17.9 (3.3)	18.1 (3.2)	16.8 (3.1)	17.4 (3.5)
Wheeze-ever, n(%)	25 (26.6)	108 (54.8)***	76 (51.0)	19 (82.6)††[ə]	13 (52.0)
Doctor-diagnosed asthma, n(%)	6 (6.4)	41 (20.8)**	26 (17.5)	9 (39.1)†	6 (24.0)
Neonatal Characteristics					
Gestational age (weeks), mean (SD)	39.9 (1.2)	30.7 (2.8)***	30.9 (2.8)	29.5 (2.4)	30.7 (3.1)
Birthweight (g), mean (SD)	3522 (522)	1666 (607)***	1731 (590)	1313 (578)††	1605 (639)
Birthweight (z-score), mean (SD)	0.12 (1.0)	0.20 (1.35)	0.33 (1.31)	-0.35 (1.56)	-0.12 (1.22)
Intrauterine growth restriction, n(%)	4 (4.3)	29 (14.7)**	18 (12.1)	9 (39.1)†††[ə]	2 (8.0)
Antenatal smoking, n(%)	4 (4.3)	23 (11.7)‡*	19 (12.8)	2 (8.7)	2 (9.1)
Antenatal steroids, n(%)	2 (2.1)	167 (89.8)‡***	127 (85.2)	18 (78.3)	22 (88.0)
Mechanical ventilation, n(%)	1 (1.1)	82 (41.6)***	61 (40.9)	14 (60.9)	7 (28.0)
Bronchopulmonary Dysplasia (BPD), n(%)	0 (0)	51 (25.9)***	33 (22.2)	11 (47.8)††	7 (28.0)

Table 5-1: Participant demographics

Preterm vs Term: *p<0.05, **p<0.01, ***p<0.001. pPRISm/POLD vs Preterm born control: †p<0.05, ††p<0.01, †††p<0.001. pPRISm vs POLD: [ə]p<0.05, [əə]p<0.01, [əəə]p<0.001 by Chi-squared/independent samples t-test/ANOVA with post-hoc Bonferroni correction as appropriate.
pPRISm: Prematurity-related preserved ratio with impaired spirometry. POLD: Prematurity-related obstructive lung disease.
‡Antenatal steroid data missing for 11 preterm-born children. Antenatal smoking data missing for 4 preterm-born children.

MSEA linked 14 significantly altered metabolites between the POLD and PT_c groups to nine significantly altered metabolic processes (Table 5-4). Capric acid (log_2FC -0.28, p=0.003), caprylic acid (-0.18, 0.003), and ceratinic acid (-0.64, 0.014) were linked with β-oxidation of very-long chain fatty acids (p=0.004). Alanine (log_2FC -0.21, p=0.046), glutamic acid (-0.24, 0.023) and pyroglutamic acid (-0.17, 0.035) were linked with glutathione metabolism (p=0.008) (Figure 5-2). Comparisons of these six metabolites between all four study groups are shown in Figure 5-3. Significant reductions in capric acid, caprylic acid, ceratinic acid and glutamic acid were observed when compared to both the PT_c and Term-born groups (p<0.05). In multiple group comparisons, pyroglutamic acid was significantly lower in the POLD group when compared to the Term-born group (p=0.029), and near significantly lower when compared to PT_c (p=0.083). Univariable and multivariable linear regression analyses of these six metabolites with early- and current-life factors in the preterm-born group are given in Table 5-6. Alanine and glutamic acid had a significant association with only POLD in univariable analysis (p=0.046 and 0.022 respectively). In multivariable analysis, capric acid, caprylic acid, ceratinic acid and pyroglutamic acid all maintained a significant association with the POLD group (p=0.009, 0.025, 0.01 and 0.044 respectively). Figure 5-4 shows the relationship between these six metabolites and spirometry values (FEV_1, FVC, FEV_1/FVC and percent predicted forced expiratory flow between 25% and 75% of vital capacity ($FEF_{25\%-75\%}$)). Significant associations were seen between FEV_1 and capric acid, caprylic acid and ceratinic acid (p=0.013, 0.0034 and 0.005 respectively), FVC and capric and caprylic acid (p=0.043 and 0.028 respectively), FEV_1/FVC and ceratinic acid (p=0.024), and FEF_{25-75} and capric acid, caprylic and ceratinic acid (p=0.014, 0.0048 and 0.0018 respectively).

Metabolite	Retention Index	m/z	PubChem ID	% of samples	Log$_2$FC	p-value
POLD vs PT$_c$ n=23 v 149						
7-methylguanine	768706	294	11361	100	-0.27	<0.001
xylonic acid	589278	333	6602431	99.4	-0.66	<0.001
myristic acid	634414	285	11005	100	-0.32	0.001
citramalic acid	456203	247	1081	100	-0.44	0.002
capric acid	452386	229	2969	100	-0.28	0.003
caprylic acid	343457	201	379	100	-0.18	0.003
salicylic acid	480699	267	338	98.8	-0.56	0.003
1-monostearin	959214	203	24699	100	-0.32	0.004
tartaric acid	534291	292	444305	99.4	-3.74	0.004
ribose	553071	217	10975657	100	-0.22	0.005
2-hydroxypyrazinyl-2-propenoicacidethylester	493127	121	5371086	100	-0.31	0.008
ribitol	575497	217	827	100	-0.25	0.009
quinic acid	634900	345	6508	100	-1.70	0.011
erythritol	471922	217	222285	100	-0.22	0.012
2-ketoisocaproic acid	290473	89	70	100	-0.50	0.012
3,4-dihydroxycinnamic acid	748847	219	689043	100	-0.52	0.013
2,8-dihydroxyquinoline	626989	290	97250	100	-0.31	0.013
ceratinic acid	1033286	145	10469	86.0	-0.64	0.014
adenosine	918039	236	60961	100	-0.22	0.014
UDP-glucuronic acid	585473	217	17473	100	-0.52	0.015
benzoic acid	339067	179	243	100	-0.24	0.015
xylitol	567437	217	6912	100	-0.20	0.016
xylulose	553450	173	439205	100	-0.24	0.017
threitol	467595	217	169019	100	-0.32	0.017
biphenyl	426625	154	7095	97.1	-0.26	0.018
2-hydroxyhippuric acid	725465	206	10253	100	-0.88	0.019
pentose	540818	103	229	100	-0.96	0.020
isothreonic acid	489385	292	151152	100	-0.15	0.021
lactose	929908	204	11333	98.3	-0.46	0.021
N-acetylmannosamine	722897	319	439281	100	-0.35	0.022
glutamic acid	529100	246	33032	100	-0.24	0.023
glucuronic acid	665901	333	94715	100	-0.20	0.023
xanthine	701688	353	1188	100	-0.22	0.024

2-picolinic acid	383668	180	1018	100	-0.52	0.026
xylose	544100	103	135191	100	-0.80	0.026
galactinol	1015529	204	N/A	100	-0.37	0.028
sophorose	959716	319	N/A	98.3	-0.42	0.029
3,4-dihydroxybenzoic acid	620200	193	72	100	-0.62	0.029
2-monopalmitin	890356	129	123409	100	-1.12	0.030
butane-2,3-diol	205778	117	262	100	-0.97	0.030
pyroglutamic acid	485935	156	7405	100	-0.17	0.035
glutamine	600000	156	5961	100	-0.17	0.035
digitoxose	521798	117	94168	100	-0.06	0.040
kynurenic acid	726186	231	3845	100	-0.44	0.043
fumaric acid	390016	245	444972	100	-0.16	0.043
glycerol	344466	205	753	100	-0.43	0.045
alanine	244189	116	5950	100	-0.21	0.046
beta-alanine	435564	248	239	100	0.55	0.047
threonic acid	497572	292	5460407	100	-0.39	0.048
pPRISm vs PT$_c$ n = 25 v 149						
beta-mannosyl glycerate	774364	204	5460194	81.6	0.67	0.002
oleic acid	781527	339	445639	91.4	-0.52	0.021
pentitol	563801	307	827	100	-0.14	0.035

Table 5-2: Significantly altered urine metabolites in POLD and pPRISm groups when compared to preterm-born controls.

m/z: Mass-to-charge ratio. Log$_2$FC: Log$_2$ fold change between groups

Metabolite	Retention Index	m/z	PubChem ID	% of samples	Log$_2$FC	p-value
POLD vs Term n=23 v 94						
biphenyl	426625	154	7095	98.3	-0.42	0.0003
xylonic acid	589278	333	6602431	99.1	-0.94	0.0003
myristic acid	634414	285	11005	100	-0.48	0.0005
guanine	744307	352	764	99.1	-0.48	0.0007
threitol	467595	217	169019	100	-0.73	0.002
xylose	544100	103	135191	100	-1.14	0.002
enolpyruvate	234394	217	1005	100	-0.43	0.002
2-ketoisocaproic acid	290473	89	70	100	-0.50	0.002
glutamic acid	529100	246	33032	100	-0.41	0.002
pentose	540818	103	229	100	-1.51	0.002
butane-2,3-diol	205778	117	262	100	-0.81	0.003
2-hydroxyhippuric acid	725465	206	10253	100	-0.71	0.003
5-hydroxymethyl-2-furoic acid	497561	123	80642	100	-1.32	0.003
furoylglycine	553990	95	21863	100	-0.63	0.004
urea	323728	189	1176	100	-1.55	0.004
ribitol	575497	217	827	100	-0.34	0.005
allantoic acid	726050	259	203	100	-0.48	0.005
galactinol	1015529	204	NA	100	-0.48	0.005
citramalic acid	456203	247	1081	100	-0.63	0.005
6-deoxyglucitol	596111	319	151266	99.1	-2.15	0.005
1-monostearin	959214	203	24699	100	-0.43	0.005
quinic acid	634900	345	6508	100	-2.82	0.005
capric acid	452386	229	2969	100	-0.32	0.006
caprylic acid	343457	201	379	100	-0.19	0.006
4-pyridoxic acid	673225	309	6723	100	-0.29	0.008
7-methylguanine	768706	294	11361	100	-0.29	0.008
gluconic acid	693148	333	6857417	100	-0.33	0.009
hypoxanthine	619128	265	790	100	-0.55	0.010
tartaric acid	534291	292	444305	99.1	-4.82	0.010
2-hydroxypyrazinyl-2-propenoicacidethylester	493127	121	5371086	100	-0.42	0.011
erythritol	471922	217	222285	100	-0.34	0.011

mannose	645856	205	18950	100	-0.63	0.011
deoxypentitol	528774	231	270738	100	-0.44	0.012
ribose	553071	217	10975657	100	-0.29	0.013
uric acid	730691	441	1175	100	-0.29	0.013
N-acetylmannosamine	722897	319	439281	100	-0.47	0.014
threonic acid	497572	292	5460407	100	-0.44	0.014
azelaic acid	610551	317	19347555	99.1	0.10	0.015
hippuric acid	638579	206	NA	100	-0.39	0.015
gluconic acid lactone	645815	220	7027	100	-0.26	0.018
butyrolactam	277199	142	12025	100	-0.28	0.018
adenine	646534	264	190	100	-0.24	0.019
shikimic acid	611100	204	8742	100	-0.70	0.021
pyrophosphate	327517	110	1023	100	-0.51	0.021
pyroglutamic acid	485935	156	7405	100	-0.28	0.023
N-carbamoylaspartate	611345	257	93072	100	-0.40	0.023
1-methylinosine	1026110	259	65095	68.4	-1.17	0.023
parabanic acid	464991	100	67126	100	-0.29	0.024
UDP-glucuronic acid	585473	217	17473	100	-0.56	0.025
indole-3-lactate	764586	202	92904	100	-0.37	0.025
maleimide	245118	154	10935	100	-0.28	0.027
thymine	420133	255	1135	100	-0.33	0.030
aconitic acid	586815	229	643757	100	-0.24	0.031
kynurenic acid	726186	231	3845	100	-0.44	0.034
uracil	385735	241	1174	100	-0.43	0.034
pyrogallol	495011	239	1057	100	-0.54	0.034
xanthine	701688	353	1188	100	-0.28	0.035
raffinose	1120886	361	439242	100	-0.49	0.036
cholesterol	1078536	129	5997	100	-0.64	0.036
glycerol	344466	205	753	100	-0.43	0.038
3-hydroxyanthralinic acid	640146	354	NA	100	-0.34	0.040
2-picolinic acid	383668	180	1018	100	-0.40	0.041
lactic acid	217657	191	612	100	-0.82	0.041
methanolphosphate	289520	241	13130	61.5	-1.40	0.041
benzoic acid	339067	179	243	100	-0.33	0.042
N-acetylaspartic acid	548028	158	65065	100	-0.24	0.042

cellobiose	932179	204	6255	100	-0.55	0.046
4-hydroxyphenylacetic acid	542795	179	127	100	-0.32	0.046
asparagine	553743	231	6267	100	-0.22	0.049
pPRISm vs Term n = 25 v 94						
oleic acid	781527	339	445639	100	-0.79	0.002
beta-mannosyl glycerate	774364	204	5460194	100	0.39	0.009
lactic acid	217657	191	612	100	-1.05	0.012
furoylglycine	553990	95	21863	100	-0.98	0.013
methanolphosphate	289520	241	13130	99.2	-0.73	0.019
indole-3-lactate	764586	202	92904	98.3	-0.48	0.028
glycyl proline	691357	174	3013625	99.2	0.44	0.028
butane-2,3-diol	205778	117	262	100	-0.48	0.034
3-hydroxyanthralinic acid	640146	354	NA	100	-0.30	0.044
biphenyl	426625	154	7095	100	-0.16	0.044
anthranilic acid	530297	266	NA	100	-0.31	0.044
3-(3-hydroxyphenyl)propionic acid	583925	192	91	98.3	0.44	0.046
xylonic acid	589278	333	6602431	100	-0.47	0.047

Table 5-3: Significantly altered urine metabolites between POLD and pPRISm groups compared with term-born controls.
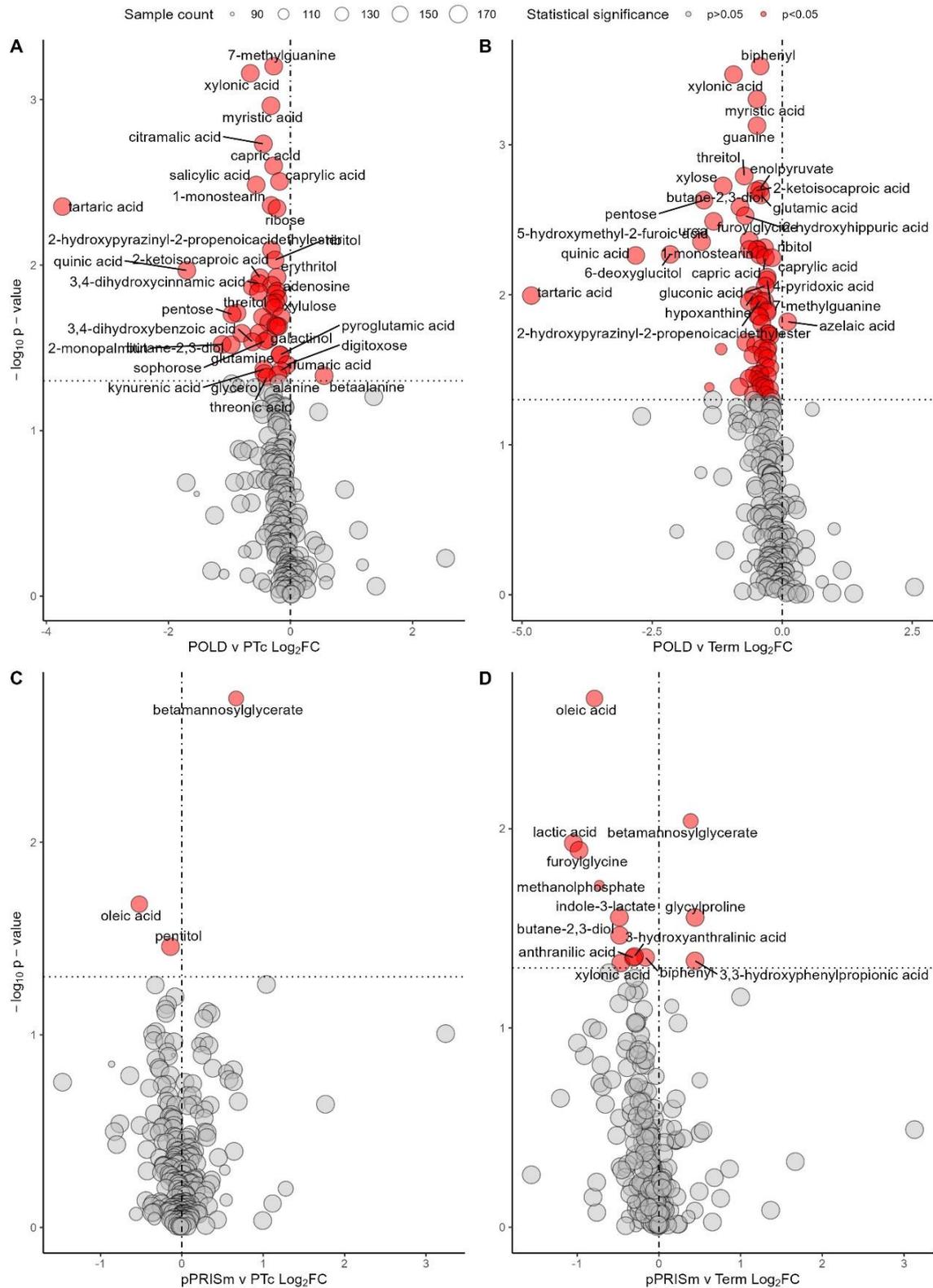
Figure 5-1: Volcano plots demonstrating significantly altered urine metabolites between groups (A) POLD vs PT$_c$ (B) POLD vs Term (C) pPRISm vs PT$_c$ (D) pPRISm vs Term.

Fold changes between groups log$_2$ transformed for visualization. Vertical line represents a Log$_2$FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to number of samples in which metabolite was detected. Metabolite name given if p<0.05. Log$_2$FC: Log$_2$ fold-change between groups.

Metabolic Process	Significantly altered metabolites	Enrichment Ratio	p-value
POLD vs PT$_c$			
Urea Cycle	Alanine, fumaric acid, glutamic acid, glutamine	6.7	0.002
β-oxidation of Very Long Chain Fatty Acids	Capric acid, caprylic acid, ceratinic acid	8.6	0.004
Aspartate Metabolism	Beta-alanine, fumaric acid, glutamic acid, glutamine	5.6	0.005
Glutathione Metabolism	Alanine, glutamic acid, pyroglutamic acid	7.0	0.008
Purine Metabolism	Adenosine, fumaric acid, glutamic acid, glutamine, xanthine	3.3	0.014
Glucose-Alanine Cycle	Alanine, glutamic acid	7.4	0.027
Amino Sugar Metabolism	Glutamic acid, glutamine, N-acetylmannosamine	4.4	0.027
Fatty Acid Biosynthesis	Capric acid, caprylic acid, myristic acid	4.2	0.032
Alanine Metabolism	Alanine, glutamic acid	6.0	0.045
POLD vs Term			
Aspartate Metabolism	Asparagine, glutamic acid, N-acetyl-L-aspartic acid, pyrophosphate, ureidosuccinic acid	4.4	0.004
Galactose Metabolism	D-mannose, galactinol, glycerol, pyrophosphate, raffinose,	4.1	0.006
Purine Metabolism	Adenine, guanine, glutamic acid, hypoxanthine, pyrophosphate, uric acid, xanthine	2.9	0.007
pPRISm vs PT$_c$			
No significant enrichment			
pPRISm vs Term			
Tryptophan Metabolism	3-hydroxyanthranilic acid, anthranilic acid	11.4	0.01

Table 5-4 Metabolite Set Enrichment Analysis demonstrating altered biological processes implicated by significantly altered urine metabolite quantities between POLD and pPRISm groups when compared to preterm- and term-born controls.

POLD: prematurity-associated obstructive lung disease. pPRISm: prematurity-associated preserved ratio impaired spirometry. PT$_c$: preterm-born controls.
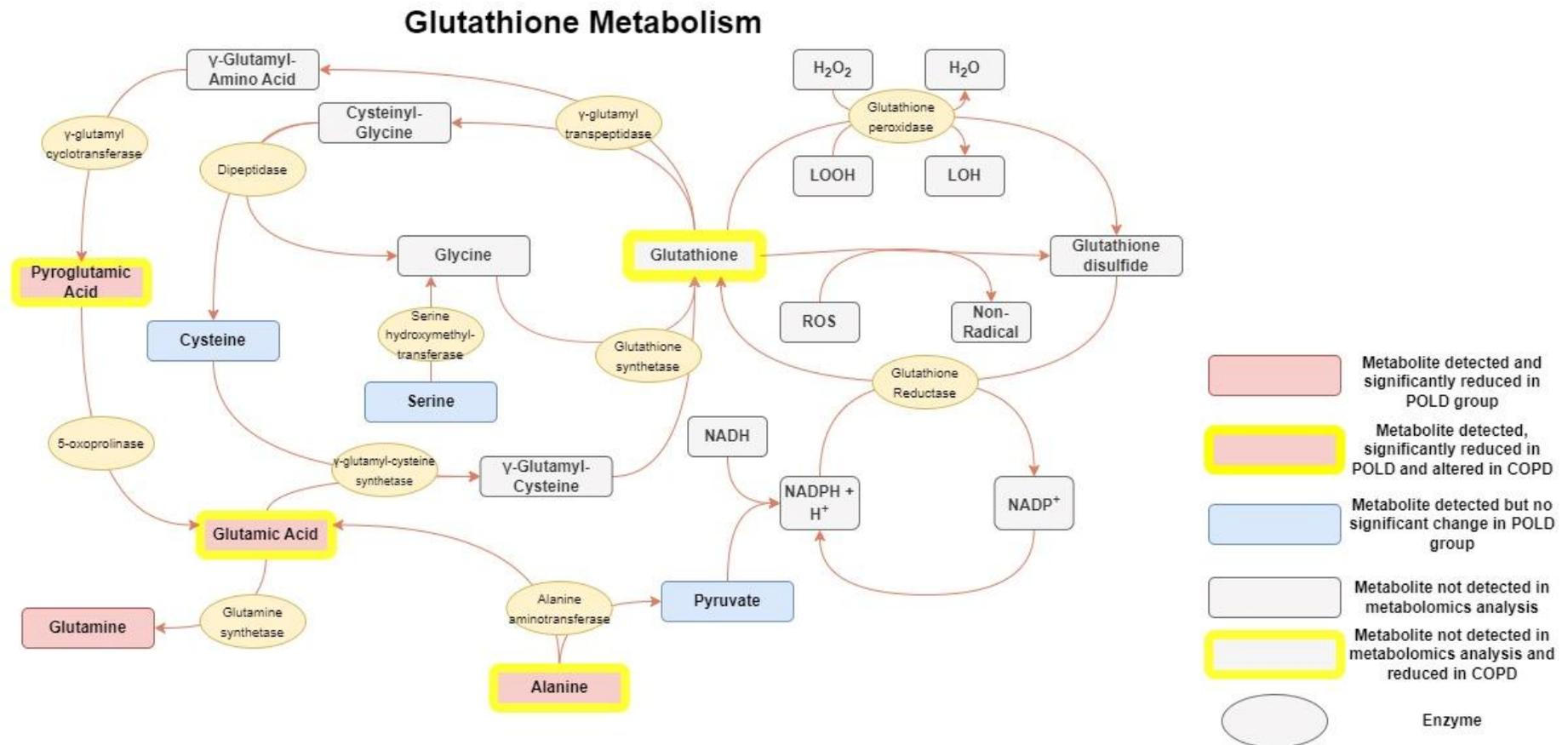
Figure 5-2: Diagram representing glutathione metabolism, highlighting altered urine metabolites within the POLD group and links with abnormalities in COPD described in the literature.
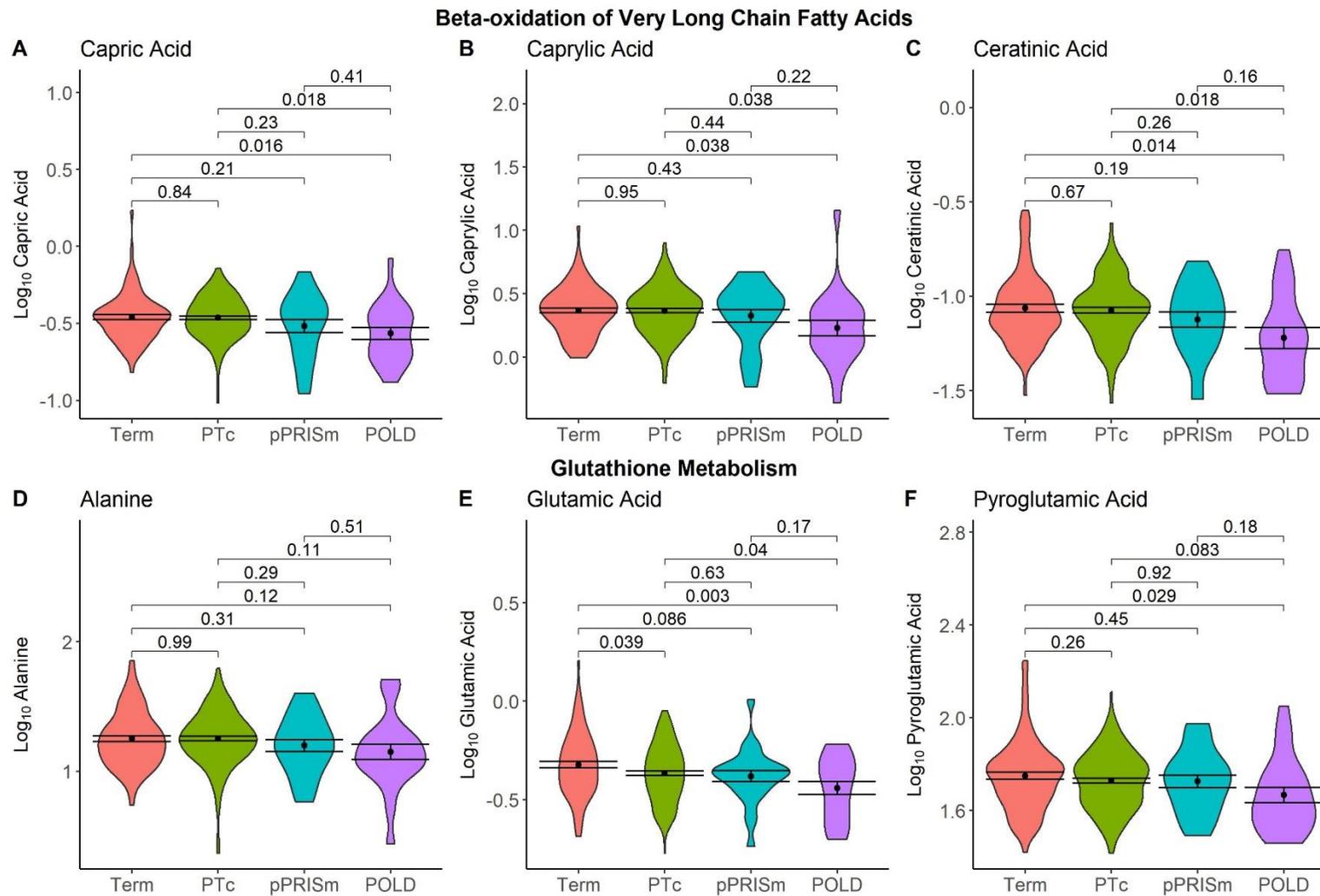
Figure 5-3: Violin Plots of Significantly Altered Urine Metabolites in POLD group, grouped by associated metabolic process.

Black dot and bars show mean and standard error of the mean (SEM). Bars give p-values from ANOVA with post-hoc Bonferroni correction for between group comparisons.

Variable	β-oxidation of Very Long Chain Fatty Acids									Glutathione Metabolism								
	Capric Acid			Caprylic Acid			Ceratinic Acid			Alanine			Glutamic Acid			Pyroglutamic acid		
	Beta	SE	p	Beta	SE	p	Beta	SE	p	Beta	SE	p	Beta	SE	p	Beta	SE	p
Univariable Models																		
Sex, ref=Male	-0.05	0.02	**0.019***	-0.02	0.03	0.51	-0.04	0.03	0.17	-0.04	0.03	0.19	0.03	0.02	0.15	-0.06	0.02	**<0.001***
Age at testing, years	0.01	0.01	0.22	0.01	0.01	0.62	0.01	0.01	0.27	0.02	0.01	0.21	0.01	0.01	0.25	-0.01	0.01	0.11
Weight, z-score	0.01	0.01	0.60	-0.001	0.01	0.94	0.003	0.01	0.82	-0.01	0.01	0.55	0.01	0.01	0.34	-0.02	0.01	**0.047***
BMI, z-score	0.002	0.01	0.81	-0.01	0.01	0.59	<0.001	0.01	0.93	-0.02	0.01	0.19	0.01	0.01	0.44	-0.02	0.01	**0.024***
Gestational age, weeks	0.003	0.004	0.41	0.01	0.01	**0.04***	0.005	0.005	0.33	0.01	0.01	0.35	-0.003	0.004	0.46	<0.001	0.004	0.99
Birthweight, z-score	-0.01	0.01	0.18	-0.01	0.01	0.36	-0.01	0.01	0.28	-0.002	0.01	0.83	-0.01	0.01	0.11	-0.01	0.01	*0.09*
IUGR, ref=No IUGR	-0.01	0.03	0.70	-0.04	0.04	0.35	-0.004	0.04	0.92	-0.05	0.05	0.31	0.01	0.03	0.71	0.01	0.03	0.84
BPD, ref=No BPD	0.01	0.03	0.66	0.01	0.04	0.89	-0.003	0.03	0.92	0.01	0.04	0.89	0.03	0.02	0.24	0.03	0.02	0.12
POLD, ref=PT$_c$	-0.10	0.04	**0.004***	-0.14	0.05	**0.004***	-0.15	0.05	**0.001***	-0.10	0.05	**0.046***	-0.08	0.03	**0.022***	-0.06	0.03	**0.036***
pPRISm, ref=PT$_c$	-0.05	0.03	0.12	-0.04	0.05	0.37	-0.05	0.04	0.22	-0.05	0.05	0.29	-0.02	0.03	0.64	-0.003	0.03	0.91
Asthma, ref=No	-0.04	0.03	0.20	-0.09	0.04	**0.021***	-0.10	0.03	**0.002***	-0.05	0.04	0.18	-0.03	0.03	0.25	-0.07	0.02	**0.003***
Multivariable Models																		
Sex, ref=Male	-0.05	0.02	**0.03***	-	-	-	-	-	-							-0.08	0.02	**<0.001***
BMI, z-score	-	-	-	-	-	-	-	-	-							-0.02	0.01	**0.019***
Gestational age, weeks	-	-	-	0.01	0.005	0.10	-	-	-	Not taken forward for multivariable analysis			Not taken forward for multivariable analysis			-	-	-
Birthweight, z-score	-	-	-	-	-	-	-	-	-							-0.01	0.01	**0.036***
POLD, ref=PT$_c$	-0.09	0.04	**0.009***	-0.06	0.05	**0.025***	-0.12	0.05	**0.010***							-0.06	0.03	**0.044***
Asthma, ref=No	-	-	-	-0.06	0.04	0.06	-0.08	0.03	**0.016***							-0.06	0.02	**0.009***

Table 5-5: Univariable and multivariable linear regression analyses of identified urinary metabolites of interest with early and current life factors in preterm-born children.

Variables with a p<0.1 in univariable analysis combined into multivariable model. ***** and **bold** indicates p<0.05. Dashes indicate a variable where p≥0.1 in univariable analysis and therefore not included in multivariable model. SE: Standard error, BMI: Body Mass Index, IUGR: Intrauterine growth restriction, POLD: prematurity-associated obstructive lung disease

In addition to alanine and glutamic acid, significant differences in fumaric acid ($\log_2$FC -0.16, p=0.043) and glutamine (-0.17, 0.035) were linked by MSEA to urea cycle metabolism (p=0.002) (Table 5-4). In addition to fumaric acid, glutamic acid, and glutamine, MSEA linked a significant increase of beta-alanine ($\log_2$FC 0.55, p=0.047) to aspartate metabolism (p=0.005). Results of linear regression analyses for these metabolites are shown in Table 5-6. Fumaric acid and glutamine remained significantly associated with the POLD group in multivariable models (p=0.021 and 0.012 respectively), but beta-alanine was no longer significant on multivariable analysis (p=0.16).

Figure 5-5 shows interactions between the significantly altered metabolic processes between the POLD and PT_c groups identified by MSEA. Direct relationships exist between alanine metabolism, aspartate metabolism and urea cycle. Purine and glutathione metabolism were also linked by glutamate metabolism, which showed a near-significant enrichment (p=0.08).

69 metabolites were significantly altered in the POLD group when compared to the Term-group (Table 5-3). MSEA linked 14 to three significantly altered metabolic processes (Table 5-4). As with the comparison with the PT_c group, aspartate metabolism (p=0.004) and purine metabolism (p=0.007) showed significant enrichment, however glutamic acid was the only common metabolite observed. Galactose metabolism showed a significant enrichment on comparison of the POLD with the term-born group (p=0.006) (Table 5-4).

5.3.2   Comparisons between pPRISm and preterm- and term-control groups:

Of 238 detected metabolites, 204 were also detected in every sample from the pPRISm group. 3 (1.3%) metabolites were significantly altered when compared to PT_c (Figure 5-1; Table 5-2) and 13 (5.5%) when compared to the Term-born subjects (Figure 5-1, Table 5-3), with two metabolites being common between the two comparisons (beta-mannosyl glycerate and oleic acid).
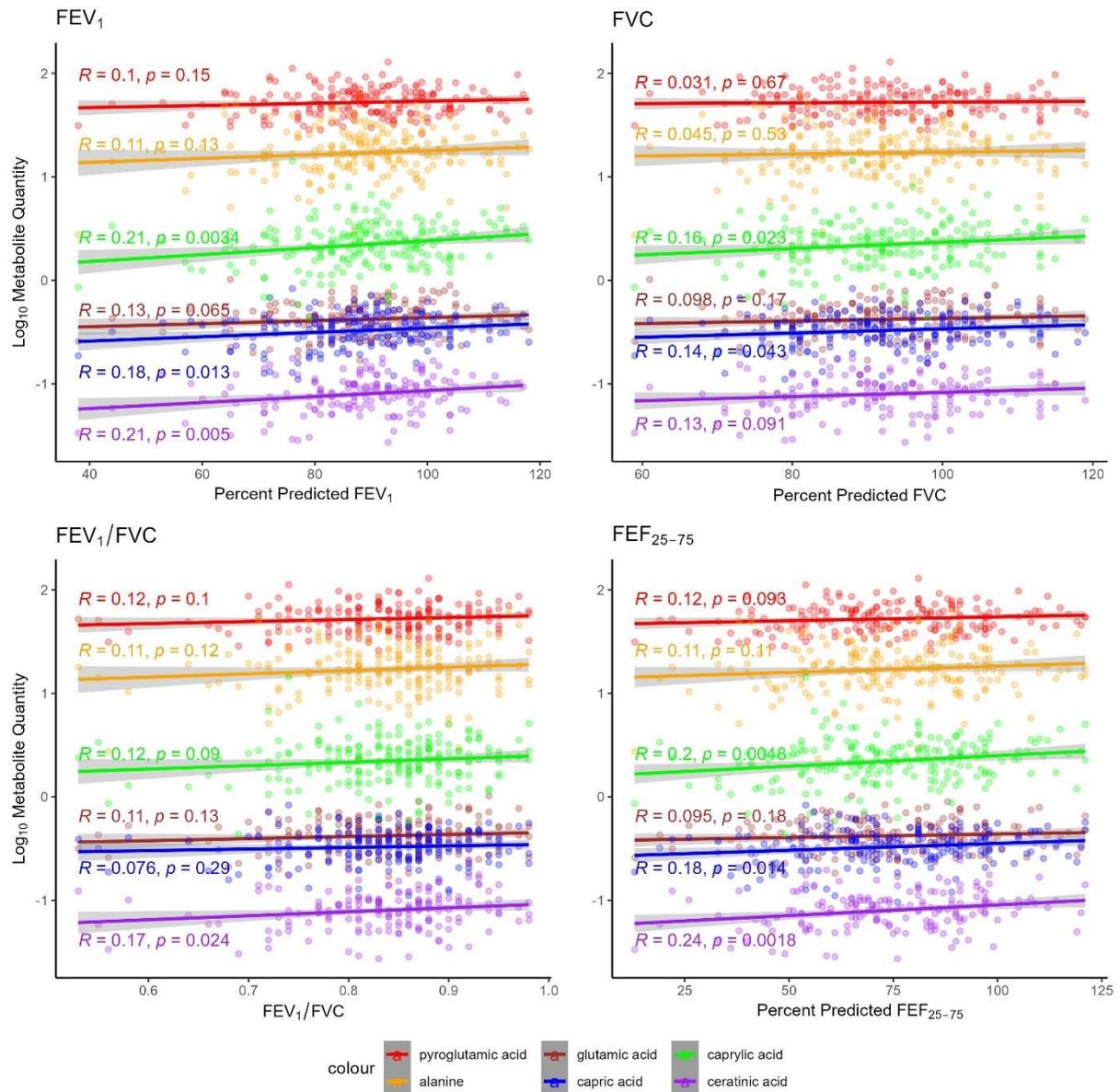
Figure 5-4 Scatter plots with linear regression lines demonstrating relationship between metabolites of interest and spirometry variables in preterm-born children.

Points represent individual metabolite measurements. Line represents linear regression model, with 95% confidence interval represented by grey shading. FEV1: Forced expiratory volume in one second. FVC: Forced vital capacity. FEF25-75: Forced expiratory flow between 25% and 75% of vital capacity.

Variable	Beta-Alanine			Fumaric Acid			Glutamine		
	Beta	SE	p-value	Beta	SE	p-value	Beta	SE	p-value
Univariable Models									
Sex, ref=Male	0.01	0.05	0.86	-0.02	0.02	0.30	-0.03	.02	0.16
Age at testing, years	-0.01	0.02	0.55	-0.004	0.01	0.64	0.002	0.01	0.79
Weight, z-score	-0.05	0.02	**0.021***	-0.02	0.01	*0.08*	-0.01	0.01	0.14
BMI, z-score	-0.05	0.02	**0.003***	-0.02	0.008	**0.049***	-0.01	0.01	*0.052*
Gestational age, weeks	<0.001	0.01	0.99	<0.001	0.004	0.95	0.003	0.004	0.47
Birthweight, z-score	-0.04	0.02	**0.032***	-0.01	0.01	0.19	-0.01	0.01	0.29
IUGR, ref=No IUGR	0.11	0.06	*0.098*	-0.005	0.03	0.88	-0.02	0.03	0.45
BPD, ref=No BPD	0.05	0.05	0.34	0.03	0.03	0.24	0.01	0.02	0.64
POLD, ref=PT$_c$	0.14	0.07	**0.043***	-0.07	0.03	*0.055*	-0.07	0.03	**0.034***
pPRISm, ref=PT$_c$	0.09	0.07	0.19	0.01	0.03	0.81	-0.02	0.03	0.61
Asthma, ref=No	0.01	0.06	0.90	-0.04	0.03	0.20	-0.04	0.03	0.12
Multivariable Models									
BMI, z-score	-0.04	0.02	**0.025***	-0.02	0.01	**0.021***	-0.02	0.01	**0.017***
Birthweight, z-score	-0.02	0.02	0.21	-	-	-	-	-	-
POLD, ref=PT$_c$	0.10	0.07	0.16	-0.08	0.03	**0.021***	-0.08	0.03	**0.012***

Table 5-6: Univariable linear regression analyses of identified urinary metabolites of interest identified as part of Aspartate Metabolism and Urea Cycle with early and current life factors in preterm-born children.

* and **bold** indicates p<0.05, *italics* indicate p<0.1. Dashes indicate a variable where p≥0.1 in univariable analysis and therefore not included in multivariable model. Variables with a p-value <0.1 in univariable model included in multivariable model. BMI: Body Mass Index, IUGR: Intrauterine growth restriction, BPD: Bronchopulmonary dysplasia, POLD: prematurity-associated obstructive lung disease, pPRISm: prematurity-associated preserved ratio impaired spirometry
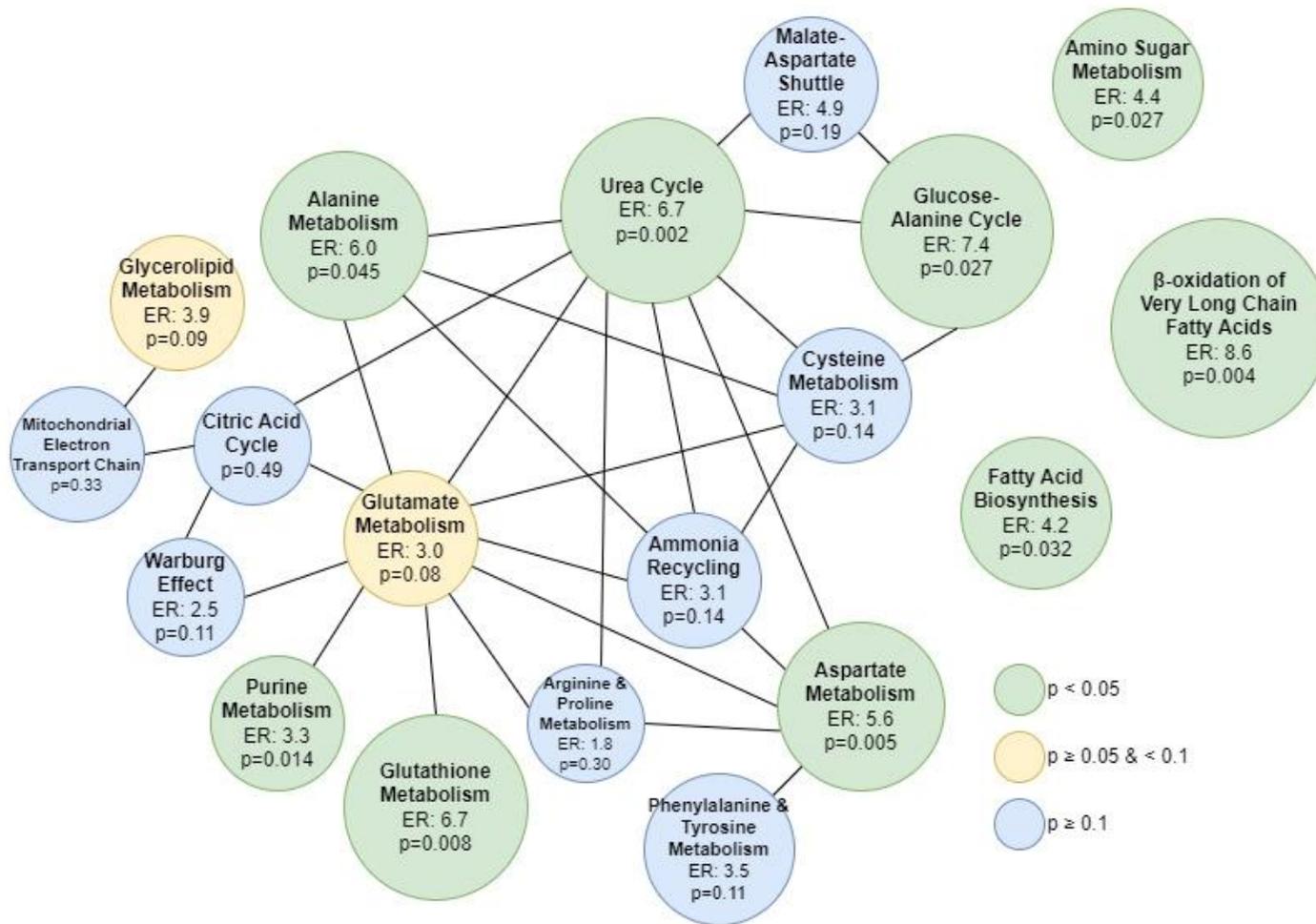
Figure 5-5: Interactions between metabolic processes identified by Metabolite Set Enrichment Analysis (MSEA) as significantly enriched in the urine metabolome of the POLD group compared to PT$_c$.

Processes colour coded according to their p-value. Size of circle relative to enrichment ratio of metabolic process. ER: Enrichment ratio

Beta-mannosyl glycerate (log$_2$FC 0.67, p=0.002), oleic acid (-0.52, 0.021) and pentitol (-0.14, 0.035) were significantly altered in the pPRISm group when compared to PT$_c$, however MSEA showed no significant links to any specific metabolic process (Table 5-4). Two altered metabolites in the pPRISm group when compared to the Term-born group (3-hydroxyanthranilic acid [log$_2$FC -0.30, 0.044] and anthranilic acid [-0.31, 0.044]) were significantly linked by MSEA to tryptophan metabolism (p=0.01).

5.3.3   Comparison of urinary metabolome before and after inhaled therapies:

Overall, paired urine metabolome results were available from 28 RCT participants, ten in the ICS group, eleven in the ICS/LABA group and seven in the Placebo group. From the 238 metabolites included in the analysis (as described above and detailed in Table 8-3), twelve were significantly altered in the ICS group, five in the ICS/LABA group and eight in the placebo group (Figure 5-6,Table 5-7). All significantly altered metabolites were detected in every set of paired samples analysed. Indol-3-acetate was significantly lower before treatment compared to post-treatment in both the ICS/LABA and Placebo groups (Log$_2$FC -0.63, p=0.045 and Log$_2$FC -0.78, p=0.005 respectively). No common significantly altered metabolites were seen between the ICS and ICS/LABA groups.

The results from MSEA for the three treatment groups are given in Table 5-8. Of the twelve significantly altered metabolites in the ICS group, five (Maltose-1, Pyrophosphate, N-carbamoylaspartate, Isoleucine and 2-ketoisocaproic acid) were differentially associated with five metabolic processes involved with carbohydrate, nucleotide, amino acid, and biotin (vitamin B7) metabolism. Of the five significantly altered metabolites in the ICS/LABA group, two (pyruvic acid and alpha-ketoglutarate) were linked to ten metabolic processes, predominantly covering carbohydrate and amino acid/protein metabolism and excretion. Seven metabolic processes were also identified as significantly altered in the placebo group, with Glycine and Serine Metabolism, Alanine Metabolism, Urea Cycle, Ammonia Recycling and Glutamate Metabolism being identified as altered in both the ICS/LABA and Placebo MSEA results, although with different metabolites implicated in both groups.
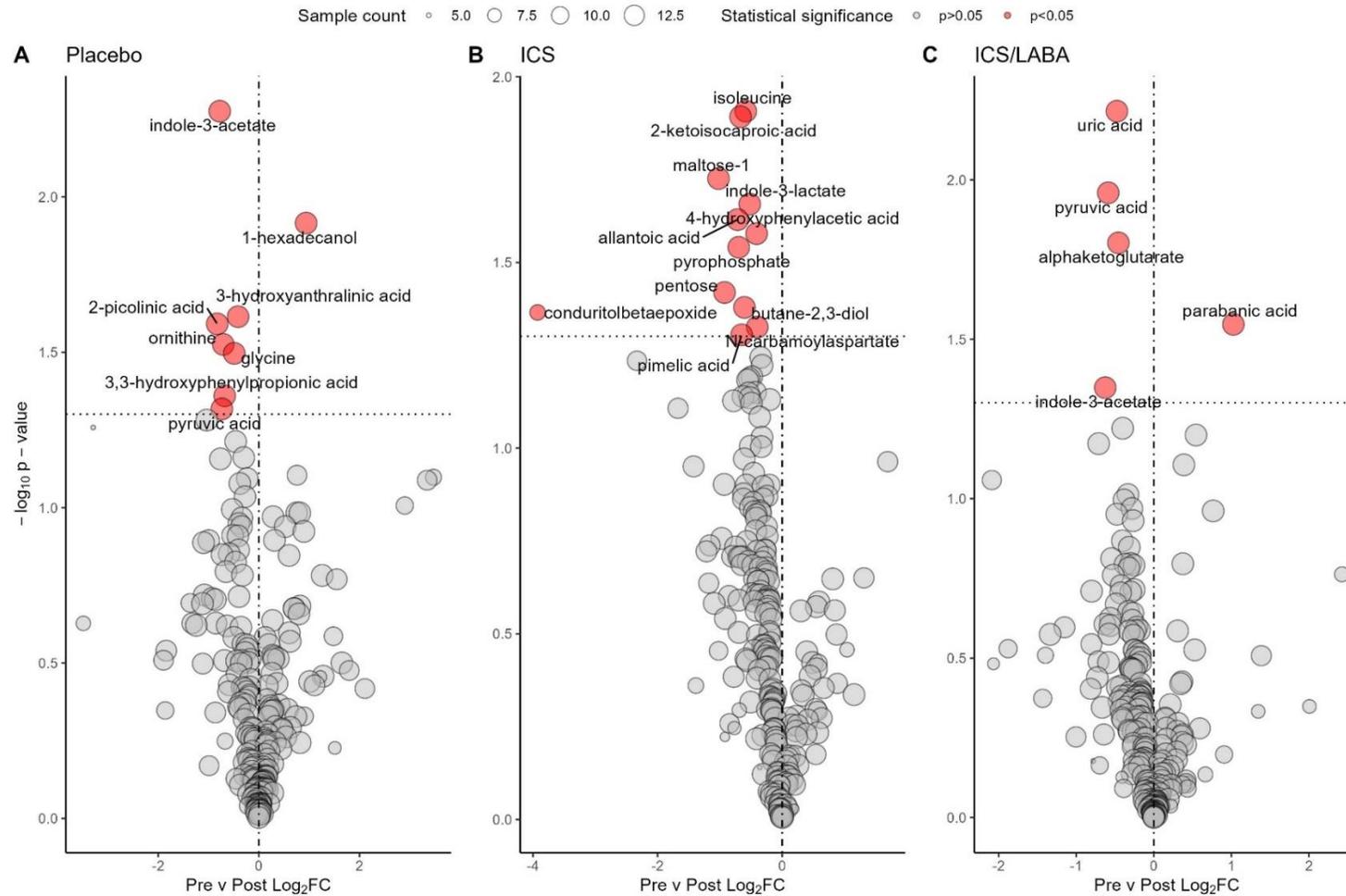
Figure 5-6: Volcano plots comparing urinary metabolome before to after treatment in the three RCT inhaler groups.

Vertical line represents a Log_2FC of 0. Horizontal line is equivalent to p-value 0.05. Size of point is relative to number of samples in which metabolite was detected. Metabolite name given if p<0.05. ICS: Inhaled corticosteroid. LABA: Long-acting β2 agonist; Log_2FC: Log_2 fold-change between groups

Metabolite	Retention Index	m/z	PubChem ID	% of participants	Fold Change	log$_2$FC	p value
Inhaled Corticosteroid (n=11)							
isoleucine	359251	158	6306	100	0.67	-0.59	0.012
2-ketoisocaproic acid	290473	89	70	100	0.63	-0.67	0.013
maltose-1	946601	204	439186	100	0.49	-1.02	0.019
indole-3-lactate	764586	202	92904	100	0.70	-0.52	0.022
allantoic acid	726050	259	203	100	0.61	-0.72	0.024
4-hydroxyphenylacetic acid	542795	179	127	100	0.75	-0.41	0.026
pyrophosphate	327517	110	1023	100	0.62	-0.70	0.029
pentose	540818	103	229	100	0.53	-0.93	0.038
butane-2,3-diol	205778	117	262	100	0.66	-0.61	0.042
conduritol beta-epoxide	675635	318	9989541	100	0.07	-3.93	0.043
N-carbamoylaspartate	611345	257	93072	100	0.76	-0.40	0.047
pimelic acid	523205	155	385	100	0.64	-0.65	0.050
Inhaled Corticosteroid/Long-Acting ß$_2$-agonist (n=10)							
uric acid	730691	441	1175	100	0.72	-0.48	0.006
pyruvic acid	213805	174	1060	100	0.67	-0.59	0.011
alpha-ketoglutarate	507392	198	51	100	0.73	-0.46	0.016
parabanic acid	464991	100	67126	100	2.04	1.03	0.028
indole-3-acetate	764586	202	92904	100	0.65	-0.63	0.045
Placebo (n=7)							
indole-3-acetate	764586	202	92904	100	0.58	-0.78	0.005
1-hexadecanol	679596	299	2682	100	1.92	0.94	0.012
3-hydroxyanthralinic acid	640146	354	N/A	100	0.75	-0.41	0.024
2-picolinic acid	383668	180	1018	100	0.56	-0.83	0.026
ornithine	619196	142	88747248	100	0.61	-0.71	0.030
glycine	368707	248	750	100	0.71	-0.49	0.032
3,(3-hydroxyphenyl)propionic acid	583925	192	91	100	0.62	-0.68	0.044
pyruvic acid	213805	174	1060	100	0.60	-0.73	0.048

Table 5-7: Significantly altered urine metabolites in the three RCT treatment groups.

m/z: mass-to-charge ratio. Log$_2$FC: Log$_2$ fold change. p values represent between group comparisons using a paired samples t-test

Metabolic Process	Significantly altered metabolites	Enrichment Ratio	p-value
ICS			
Starch and Sucrose Metabolism	Maltose-1, Pyrophosphate	10.8	0.013
Aspartate Metabolism	Pyrophosphate, N-carbamoylaspartate	9.5	0.016
Pyrimidine Metabolism	Pyrophosphate, N-carbamoylaspartate	5.9	0.041
Valine, Leucine and Isoleucine Degradation	Isoleucine, 2-ketoisocaproic acid	5.7	0.044
Biotin Metabolism	Pyrophosphate	20.9	0.048
ICS/LABA			
Glucose-Alanine Cycle	Pyruvic acid, Alpha-ketoglutarate	38.5	0.001
Alanine Metabolism	Pyruvic acid, Alpha-ketoglutarate	29.5	0.002
Cysteine Metabolism	Pyruvic acid, Alpha-ketoglutarate	19.2	0.004
Urea Cycle	Pyruvic acid, Alpha-ketoglutarate	17.9	0.004
Ammonia Recycling	Pyruvic acid, Alpha-ketoglutarate	16.1	0.005
Citric Acid Cycle	Pyruvic acid, Alpha-ketoglutarate	15.6	0.006
Gluconeogenesis	Pyruvic acid, Alpha-ketoglutarate	15.2	0.006
Glutamate Metabolism	Pyruvic acid, Alpha-ketoglutarate	10.4	0.013
Warburg Effect	Pyruvic acid, Alpha-ketoglutarate	8.8	0.018
Glycine and Serine Metabolism	Pyruvic acid, Alpha-ketoglutarate	8.5	0.019
Tryptophan Metabolism	Indole-3-acetate, Alpha-ketoglutarate	8.5	0.019
Malate-Aspartate Shuttle	Alpha-ketoglutarate	25.1	0.039
Pyruvaldehyde Degradation	Pyruvic acid	25.1	0.039
Placebo			
Glycine and Serine Metabolism	Glycine, Ornithine, Pyruvic acid	8.5	0.003
Alanine Metabolism	Glycine, Pyruvic acid	19.6	0.004
Urea Cycle	Ornithine, Pyruvic acid	11.9	0.011
Ammonia Recycling	Glycine, Pyruvic acid	10.8	0.013
Glutamate Metabolism	Glycine, Pyruvic acid	7.0	0.030
Arginine and Proline Metabolism	Glycine, Ornithine	6.4	0.035
Tryptophan Metabolism	Indole-3-acetate, 3-hydroxyanthralinic acid	5.7	0.044

Table 5-8: Metabolite Set Enrichment Analysis results for significantly altered metabolites in the RCT treatment groups

Treatment Group	Lung function group comparison			
	POLD vs PTc	POLD vs Term	pPRISm vs PTc	pPRISm vs Term
ICS	butane-2,3-diol 2-ketoisocaproic acid	indole-3-lactate 4-hydroxyphenylacetic acid	None	indole-3-lactate
ICS/LABA	None	None	None	None
Placebo	2-picolinic acid	3-hydroxyanthranillic acid 2-picolinic acid	None	3,(3-hydroxyphenyl) propionic acid

Table 5-9: Table showing common metabolites between baseline lung function group comparison and RCT treatment group comparison.

Metabolites in red were reduced in baseline comparison and in pre-treatment samples. Metabolites in blue were elevated in baseline samples and reduced in pre-treatment samples.

I next compared the metabolites altered by inhaled therapies to those metabolites identified as altered in the baseline samples when comparing the lung function groupings (Table 5-9). Six metabolites identified as significantly altered within the POLD group were also identified as altered within the treatment groups. All six were significantly reduced in the POLD group at baseline and were also significantly reduced in pre-treatment samples. Four (butane-2,3-diol, 2-ketoisocaproic acid, indole-3-lactate and 4-hydroxyphenylacetic acid) were significantly increased following ICS treatment and two (2-picolinic acid and 3-hydroyanthranilic acid) were significantly increased by placebo. The pPRISm group were also identified as having significantly reduced quantity of indole-3-lactate. 3,(3-hydroxyphenyl)propionic acid was significantly increased in the pPRISm group when compared to Term controls, and this was also identified as having a significantly reduced quantity in the pre-treatment samples in the placebo group. No common metabolites were seen in the ICS/LABA treatment group when compared to the baseline lung function group comparisons. In addition, none of these common metabolites were identified as significantly enriched in MSEA for either the baseline lung function group comparisons or the treatment group comparisons.

5.4 Discussion

In this novel, exploratory metabolomic analysis of urine from school-aged children with PLD, I have demonstrated significant differences in multiple metabolites linked with several metabolic processes

in the POLD group when compared to preterm- and term-born controls. Of particular interest were significant decreases in metabolites consumed and produced during fatty acid biosynthesis and metabolism, especially β-oxidation of very-long chain fatty acids, and glutathione metabolism, findings which are similar to those reported in studies of adults with COPD, as I will discuss below. It has previously been demonstrated that a neonatal history of BPD is significantly associated with development of an obstructive spirometry pattern in childhood (Cousins et al., 2023), and a recently published meta-analysis has demonstrated that this airway obstruction likely increases over the life course (Gibbons et al., 2023). There is increasing concern that PLD predisposes to early-onset COPD in adulthood (Pulakka et al., 2023), and my current exploratory urine metabolome analyses suggest that the altered metabolic activity present in childhood for those with a POLD phenotype is similar to adult studies of COPD. In contrast, minimal differences were noted for the urinary metabolome in the pPRISm phenotype when compared with the preterm- and term-born controls, implying less systemic active metabolic processes occurring within this group.

β-oxidation of very-long chain fatty acids occur in peroxisomes, where fatty acids are broken down before transportation to mitochondria, where further fatty acid degradation and energy release occurs (Schrader et al., 2020). The increased energy requirements secondary to airway inflammation and increased work of breathing in obstructive respiratory diseases such as COPD have been suggested to increase fatty acid consumption (Wada et al., 2005), with previous urine metabolomic studies supporting this finding with increased products of fatty acid catabolism (Wang et al., 2013). Previous metabolomic studies of airway samples in preterm infants who later developed BPD have also shown decreased quantities of metabolites involved in β-oxidation of fatty acids (Lal et al., 2018), as well as increases in acylcarnitines, which are released following β-oxidation of fatty acids during oxidative stress (Piersigilli et al., 2019). Similarly, altered β-oxidation of fatty acids (Callejon-Leblic et al., 2019) and increases in serum acylcarnitine have also been noted in COPD (Kim et al., 2021, Moitra et al., 2023). I observed significantly decreased capric and caprylic acids in the POLD group. Capric and

caprylic acids, both medium-chain fatty acids, have anti-inflammatory and antioxidant effects (Lee and Kang, 2017) in porcine models of intestinal disease. Whether these metabolites have similar roles in the lung is speculative. I also observed a reduction of the very-long chain fatty acid ceratinic (also known as hexacosanoic) acid in the POLD group, which is likely related to increased consumption for energy release owing to inflammatory processes and oxidative stress. β-oxidation of very-long chain fatty acids in peroxisomes leads to the production of hydrogen peroxide (H_2O_2) (Foerster et al., 1981), a reactive oxygen species (ROS) resulting in oxidative damage and altered intracellular signaling. Increase in peroxisome activity, due to increased fatty acid metabolism, leads to peroxisome-induced oxidative stress (Schrader and Fahimi, 2006), with peroxisomal enzymes responsible for fatty acid breakdown and H_2O_2 production disproportionately upregulated compared to H_2O_2-scaveging enzymes, such as catalase, in rodent models (Chu et al., 1995). Capric, caprylic and ceratinic acids had linear relationships with spirometry values across the preterm-born children, suggesting that β-oxidation of very-long chain fatty acids generally has an association with lung function.

Capric and caprylic acid, along with myristic acid were also implicated in fatty acid biosynthesis, another significantly altered process in the POLD group when compared with the Preterm and Term control groups. Myristic acid, a long-chain saturated fatty acid, was significantly decreased in the POLD group. Fatty acid metabolism impairments have been observed in airway secretions from patients with COPD both during the stable phase and during acute exacerbations (van der Does et al., 2017). Macrophage activity activates and regulates COPD-related pulmonary inflammation (Kotlyarov and Kotlyarova, 2021), and fatty acid metabolism is intrinsically linked with metabolic reprogramming of macrophages. Fatty acid biosynthesis has been shown to enhance pro-inflammatory activity and interleukin synthesis by macrophages, whereas fatty acid oxidation has a role in anti-inflammatory macrophage activity (Batista-Gonzalez et al., 2019).

I observed reduced levels of alanine, glutamic acid and pyroglutamic acid within the POLD group, which were significantly linked to glutathione metabolism. The lower levels of these metabolites are suggestive of lower glutathione concentration as all three are involved in glutathione synthesis (Figure 5-2). Glutathione provides potent defense against pulmonary oxidative injury, with studies of healthy adults demonstrating higher glutathione levels in the airways than in serum (Cantin et al., 1987). Animal models demonstrate pulmonary glutathione depletion enhances oxygen toxicity (Deneke et al., 1985). I noted decreased pyroglutamic acid and glutamic acid in the POLD group, both key intermediaries in glutathione synthesis and recycling. One pathway of glutathione consumption is in the removal of H_2O_2 by conversion of reduced glutathione to glutathione disulfide, catalysed by the peroxisomal enzyme glutathione peroxidase (Schrader and Fahimi, 2006). Reduced quantities of metabolites involved with glutathione metabolism, and thereby increased oxidative stress, have been observed in other respiratory pathologies, including COPD. Decreased alanine, pyroglutamic acid, glutamic acid and glutathione have been reported in a metabolomic study of murine lungs and in bronchoalveolar lavage fluid from adults with pulmonary inflammation and respiratory failure (Pacht et al., 1991, Ambruso et al., 2021). Pyroglutamic acid, glutamic acid, alanine and glutathione levels are decreased in targeted assay and/or metabolomic studies of serum from adults with COPD (Faucher et al., 2004, Callejon-Leblic et al., 2019), with pyroglutamic acid quantity being associated with a pulmonary emphysema phenotype (Callejon-Leblic et al., 2019). Whilst I did not detect glutathione in either its reduced or oxidized form, glutathione has a short half-life of approximately ten minutes (Hong et al., 2005), thus making its detection in urine challenging.

In contrast to the several altered metabolic pathways affected in the POLD group, suggestive of an ongoing active disease process, I observed a relatively stable metabolome within the pPRISm group when compared to the two Control groups. Only one metabolic process, namely tryptophan metabolism, was altered in the pPRISm group when compared to the Term control group. 3-hydroxyanthranillic acid was significantly lower in the pPRISm group and is a product of tryptophan

oxidation. Tryptophan is an essential amino acid, and deficiency limits protein synthesis, causing cellular dysfunction and decreased proliferation Reduced plasma tryptophan levels have been observed in COPD, particularly during acute exacerbations (Gulcev et al., 2016). Reduced tryptophan metabolism, as suggested by our results, can also lead to reduced production of kynurenine. Kynurenine promotes naïve CD4+ T-cells to become anti-inflammatory T-regulator lymphocytes, rather than highly-inflammatory Th17 lymphocytes (Stone et al., 2013).

Regarding the results for the changes in the urinary metabolome in the three RCT treatment groups, I have not identified a consistent pattern of changes comparing pre- to post-treatment samples in paired analysis. No common significantly altered metabolites or metabolic processes were identified between the ICS and ICS/LABA groups, and whilst although there was only one commonly altered metabolite in both the ICS/LABA and Placebo groups, five common significantly affected metabolic processes were seen in both groups. Comparing the metabolite changes seen in the treatment groups to those seen in the baseline lung function group comparisons, indole-3-lactate quantity was lower in both the POLD and pPRISm groups when compared to term-born controls, and its quantity was significantly increased by ICS treatment. Indole-3-lactate is also a product of tryptophan metabolism and has been identified as having anti-inflammatory properties in *in vitro* models of the preterm human intestine (Meng et al., 2020), where its production is related to intestinal microbiome composition. In contrast to my findings, a previous metabolomic study of human plasma from adults with asthma being treated with fluticasone propionate at similar doses to those used in the RHiNO study, demonstrated small but significant reductions in indolelactate following seven days of treatment compared to placebo (Daley-Yates et al., 2022). This suggests a different underlying mechanism to the reduced lung function seen in PLD (POLD and pPRISm) in this cohort compared to asthma. The fact that the changes in indole-3-lactate were only present in ICS monotherapy and not seen in combination therapy with LABA in my results suggest the dual effect of these drugs may influence different metabolic pathways.

5.4.1 Strengths and Limitations

This exploratory study represents the first time, to my knowledge, that the urinary metabolome of PLD has been studied in childhood. This analysis has been performed in one of the largest contemporary preterm-born paediatric populations available, who would have experienced modern standards of neonatal care. Using an untargeted approach, I have identified many individual metabolites and several significantly altered and biologically relevant metabolic pathways in the spirometry-defined clinical groups. Composition of the urinary metabolome can be affected by dietary intake (Stratakis et al., 2022), for which I had insufficient information to adjust for in the analyses. I adjusted the metabolite concentrations for dilutional effects using urinary creatinine, which is a widely accepted and recommended practice in urine metabolomic studies (Li et al., 2022). However, as these samples were collected at the time of spirometry, they were not necessarily early morning specimens, nor 24-hour urine collections, which may reveal greater metabolomic differences. These results require replication in a validation cohort, but this is currently limited by a lack of similar contemporaneous cohorts to study.

5.4.2 Conclusion

In conclusion, in this chapter I have demonstrated active metabolic processes with multiple significantly altered metabolites in the urinary metabolome of children with a POLD phenotype, including changes in β-oxidation of very-long chain fatty acids, fatty acid biosynthesis and glutathione metabolism. These changes imply increased cellular energy requirements and oxidative stress which have also been observed in COPD. In contrast, the metabolome appears more stable in pPRISm with a suggestion of altered tryptophan metabolism. Whether this phenotype is associated more with structural abnormalities rather than metabolic ones is speculative and will require further study.

6 Discussion: Integrating Proteomic and Metabolomic Findings

6.1 Overview

In this thesis, I have explored the mechanisms underlying different respiratory phenotypes in preterm-born school-aged children. To do this, I have utilized two different non-invasively collected samples, namely exhaled breath condensate, which is reflective of the biology of the respiratory tract (Davis et al., 2012), and urine which is reflective of systemic biology (Wu and Gao, 2015). These samples underwent untargeted proteome and metabolome analyses by mass spectrometry, using commercial analysis protocols, to provide a comprehensive representation of any altered biological mechanisms present in participants of the RHiNO study. I have used well-established risk factors to compare participant proteome/metabolome, such as history of BPD, upon which much of the existing published literature has focused, as well as newer PLD spirometry-based phenotypes, including PT_{low}, POLD and pPRISm, which have recently been defined within the RHiNO cohort (Cousins et al., 2023, Hart et al., 2022), as discussed in section 1.6.2.

This final chapter shall first give a recap of the key findings from the four results chapters and then discuss how these findings describe the biological mechanisms underlying differing phenotypes of PLD.

6.2 Key Findings

The key findings of this thesis are summarized below and in Table 6-1.

6.2.1 Analysis of the EBC proteome

- Despite a low overall protein content, mass spectrometry tandem-mass tag proteomic analysis of EBC was technically successful, and identified 210 separate proteins overall, with nineteen proteins common to all baseline samples.

- Cytokeratin abundance (KRT6A and KRT6B) was significantly increased in the BPD group at baseline when compared to preterm-born controls.

- Significantly reduced abundances of three proteins involved in desmosome structure (DSG1, DSC1, JUP) were significantly reduced within the BPD group compared to preterm-born controls in baseline samples.

- The reduced abundances of DSG1, DSC1 and JUP at baseline were predominantly observed in those participants with a history of BPD and current low lung function (PT$_{low}$; FEV$_1 \leq 85\%$ predicted).

- Abundance of DSG1, DSC1 and JUP increased significantly following 12-weeks combined inhaler therapy (ICS/LABA) to levels comparable with term-born controls at baseline, an effect seen predominantly in those preterm-born children in the RCT with a history of BPD. No significant changes were seen in these proteins with monotherapy (ICS) or placebo.

- 12-weeks of combined inhaler therapy also decreased the abundance of heat shock protein beta-1 (HSPB1), a protein associated with protein aggregation and folding, increased abundance of which has previously been associated with responses to oxidative stress and inflammation (Acunzo et al., 2012).

- PT$_{low}$ was associated with significantly reduced abundances of three antiproteases (ANXA1, SERPINB3 and SLPI) at baseline, however none of these proteins was altered by any of the inhaler therapies.

Sample Type	Analysis Methodology	Key Findings
Exhaled Breath Condensate	Proteomics	*Within the BPD group:* • Increased KRT6A and KRT6B • Reduced DSG1, DSC1 and JUP • Reduced abundances of DSG1, DSC1 and JUP most marked in those with low lung function (FEV$_1$≤85%)
	Metabolomics	*Within the BPD group:* • Reduced quantities of alanine, urea, pyroglutamic acid, and ornithine • Alterations in alanine and pyroglutamic acid linked to a reduction in glutathione metabolism • Alterations in alanine, urea and ornithine linked to alterations of urea cycle
Urine	Proteomics	*Within the pPRISm group:* • Altered abundances of 16 proteins linked to inflammation, and leucocyte and lymphocyte cell numbers. • Four proteins (DNASE1, PGLYRP1, B2M and SERPINA3) showed a reasonable ability to discriminate pPRISm from PT$_c$ *Within the POLD group:* • Increased abundances of four proteins (AGT, CTSC, MMP9 and S100A8) were linked with alterations of neutrophil biology • Three proteins associated with tissue remodeling (S100A8, MMP9 and CTSC) showed a strong ability to discriminate POLD from PT$_c$
	Metabolomics	*Within the POLD group:* • Reduced quantities of six metabolites differentially linked with β-oxidation of very-long chain fatty acids (capric acid, caprylic acid and ceratinic acid) and glutathione metabolism (alanine, glutamic acid and pyroglutamic acid).

Table 6-1: Summary of the key findings of baseline EBC and Urine sample data analyses

6.2.2 Analysis of the EBC metabolome

- Despite an overall low metabolite content, mass spectrometry-based metabolomic analysis of EBC was also technically successful, identifying 235 separate metabolites in total, with 38 being detected above the limit of detection in every baseline sample.

- Ten metabolites were significantly altered in the BPD group when compared to preterm-born controls, including reductions in alanine, urea, pyroglutamic acid, valine and ornithine.

- Fourteen metabolites were significantly altered in the BPD group when compared to term-born controls. As with the comparison to the preterm-born controls, significant reductions in alanine, urea, pyroglutamic acid and valine were also seen, as well as significant reductions in glycine, oleamide and glutamic acid.

- Metabolite set enrichment analysis (MSEA) significantly linked the reductions seen in alanine and pyroglutamic acid (as well as glycine in the comparison to term-born subjects) to alterations in glutathione metabolism. The changes observed in alanine, ornithine and urea were significantly linked to alterations in urea cycle metabolism.

- In univariable linear regression modelling, no significant associations were observed between these metabolites of interest and current lung function parameters.

- No metabolic processes were significantly enriched following inhaled therapies in those preterm-born children with low lung function ($FEV_1 \leq 85\%$ predicted).

6.2.3 Analysis of the urine proteome

- Following data normalisation, 785 proteins were detected across all samples included in the analysis, 129 of which were detected in every sample.

- 37 proteins were significantly altered in the pPRISm group when compared to PT_c, sixteen of which were significantly linked to six likely upregulated biological processes related to

inflammation, and leucocyte and lymphocyte cell numbers. In addition, there was a suggestion that CD4+ T-lymphocyte number may be downregulated.

- Four proteins (DNASE1, PGLYRP1, B2M and SERPINA3) showed a reasonable ability to discriminate pPRISm from the preterm-born cohort in ROC analysis.

- 44 proteins were significantly altered in the POLD group when compared to PT_c, of which only four (AGT, CTSC, MMP9 and S100A8) were significantly linked to one likely upregulated biological process related to neutrophil biology.

- Three proteins associated with tissue remodeling (S100A8, MMP9 and CTSC) showed a strong ability to discriminate POLD from the preterm-born cohort in ROC analysis.

- No significantly altered proteins in the POLD and pPRISm groups when compared to preterm- or term-born controls were significantly altered by inhaled therapies.

6.2.4 Analysis of the urine metabolome

- Overall, 242 different metabolites were detected across all samples, with 204 being detected in both every POLD and pPRISm sample analysed.

- Within the POLD group, 49 metabolites were significantly altered when compared to the preterm-born controls, of which fourteen were linked by MSEA to nine significantly altered biological processes, including β-oxidation of very-long chain fatty acids (capric acid, caprylic acid and ceratinic acid) and glutathione metabolism (alanine, glutamic acid and pyroglutamic acid).

- All six of these metabolites remained significantly associated with POLD in multivariable linear regression analyses when other relevant early and current life factors were considered.

- On comparing the pPRISm group to preterm-born controls, only three metabolites were significantly altered, which were not linked by MSEA to a biological process.

- On comparing the pPRISm group to term-born controls, thirteen significantly altered metabolites were observed, of which two (3-hydroxyanthranilic acid and anthranilic acid) were significantly linked to tryptophan metabolism by MSEA.

- Indole-3-lactate, a product of tryptophan metabolism, was significantly lower in both POLD and pPRISm groups at baseline and was significantly increased by ICS treatment.

6.3 Discussion

The results in this thesis represent the first time that proteomic and metabolomic analysis techniques have been applied to non-invasively collected samples in a large cohort of preterm-born school-aged children who would have experienced a contemporary standard of neonatal care. It is also the first time that the mechanisms underlying current lung function phenotypes of PLD in preterm-born school-aged children have been investigated using these techniques. In addition, these exploratory analyses represent some of the largest sample sizes of EBC and urine on which these proteomic and metabolomic techniques have been applied in the published literature.

6.3.1 Biological Mechanisms in Children with a History of Bronchopulmonary Dysplasia

BPD remains a key outcome metric for the neonatal community (Webbe et al., 2020), and as survival improves from increasingly immature gestations, and therefore increasingly immature stages of lung development, the incidence of BPD is rising (Stoll et al., 2015). In addition, much of the follow-up work with regard to lung function and respiratory morbidity in preterm-born survivors in later life has focused on those with a neonatal history of BPD, finding spirometry impairments that are generally more severe than their preterm-born counterparts without a BPD diagnosis (Kotecha et al., 2022b, Doyle et al., 2019b). Consequently, much of the mechanistic work studying these later lung function

impairments has also focused on those with a neonatal history of BPD (Um-Bergström et al., 2022, Filippone et al., 2012, Carraro et al., 2015).

However, as neonatal intensive care has advanced over the last thirty years, the clinical phenotype of BPD has changed significantly since its original description (Northway et al., 1967), and therefore it remains important to study cohorts of preterm-born survivors born at increasingly immature gestations and who would have experienced a contemporary standard of neonatal care. This includes cohorts with routine exposure to maternal antenatal corticosteroids, exogenous pulmonary surfactant replacement and modern neonatal ventilation strategies. Understanding the mechanisms underlying longer-term respiratory morbidity following BPD, including whether these are fixed or active processes, will aid in identification, prognostication, and therapeutics for this growing cohort of children.

The analysis of EBC revealed significant differences in both the proteome and metabolome of preterm-born children with a neonatal history of BPD when compared to preterm- and term-born controls (Figure 6-1). The EBC proteome in children with BPD demonstrated reduced abundance of proteins which are structural components of desmosomes (DSG1, DSC1 and JUP) as well as increased abundances of proteins known to form the cytoskeleton and intermediate filaments (KRT6A and KRT6B), which connect with desmosomes to give the cell structural strength (Herrmann et al., 2007). The changes seen in the desmosome proteins were most profound in those who had a history of BPD and current low lung function (PTl_{ow}); however, no difference was seen in any of these proteins within the PT_{low} group as a whole. DSG1, DSC1 and JUP all showed a significant increase in those children who entered the RCT and were treated with combination of ICS and LABA, and again this treatment effect was most noticeable in those in the PT_{low} group who had a history of BPD. Interestingly, no treatment effect was seen for these proteins in the monotherapy ICS group.

217

With regard to the EBC metabolome, I have demonstrated significant reductions in metabolites involved in the synthesis of glutathione and the urea cycle in those preterm-born children with a history of BPD. Taken together, the metabolite changes seen suggest a reduction in the production of glutathione, a potent airway antioxidant (Cantin et al., 1987, Grigg et al., 1993) as well as reduced urea cycle activity, leading to the accumulation of intracellular ammonia, potentially precipitating pulmonary oxidative stress (Li et al., 2023).

As has previously been discussed in this thesis (in section 2.4), there is evidence from *in vitro* studies of bronchial epithelium that increased inflammation leads to a reduction in desmosome size and number (Andersson et al., 2010, Carayol et al., 2002), as well as reduced desmosome size and number in bronchial wall biopsies taken from adults with both allergic and non-allergic asthma (Shahana et al., 2005). An increase in cytokeratins has also been identified as a potential marker of lung injury and inflammation in EBC from mechanically ventilated adults (Gessner et al., 2008) and serum studies of preterm infants who later develop BPD (Panahabadi et al., 2021). In conjunction with the reduced abundance of protease inhibitors (ANXA1, SERPINB3, CSTA and SLPI) seen in the BPD group, there is a suggestion of an active tissue remodeling process occurring in the airways of those subjects with a history of BPD.

Oxidative stress is the consequence of an imbalance between the production of reactive oxygen species (ROS) and their elimination by protective mechanisms, such as antioxidant enzymes (Hussain et al., 2016). Increased inflammation has been implicated in other pulmonary airway diseases, such as asthma, as a promoter of oxidative stress and reduced antioxidant defenses (Bezerra et al., 2023). My findings from the EBC metabolome analysis, mapping to a reduction in glutathione metabolism

218

and availability, and thereby impaired antioxidant defenses, has been noted in airway samples from other inflammatory paediatric lung diseases, including asthma (Fitzpatrick et al., 2009) and cystic fibrosis (Kettle et al., 2014). Reduced glutathione is also seen in broncheoalveolar lavage fluid taken from neonates on the first day of life who later develop BPD (Grigg et al., 1993, Collard et al., 2004). The fact that a reduction in glutathione metabolism has been implicated in early neonatal lung disease, as well as in the results from these analyses, suggests that this abnormality may persist many years after the initial pulmonary insult, although longitudinal studies from the neonatal period into later life would be required to confirm this speculation.

There is evidence from computed tomography (CT) radiological studies of preterm-born individuals in childhood, who would have experienced a comparable standard of neonatal care to those in the RHiNO cohort, that those with a neonatal history of BPD have significantly increased rates of pulmonary structural abnormalities, including bronchial wall thickening and decreased pulmonary attenuation, suggestive of reduced pulmonary vascularity, reduced alveolar complexity and potentially pre-emphysematous lung changes (Simpson et al., 2017). Linear and subpleural opacities have also been observed in CT scans in infancy (Mahut et al., 2007) and childhood/adolescence (Simpson et al., 2017, Aukland et al., 2009) in subjects with BPD, and likely represent early fibrotic lesions. These radiological appearances would be in keeping with the findings from my proteomic and metabolomic analyses of EBC from children with a history of BPD, with bronchial wall thickening, a consequence of airway inflammation (Tiddens et al., 2000), and altered lung tissue structure. In murine animal models, altered glutathione metabolism and consequent increased oxidative stress in the airways has been associated with abnormal lung tissue development (Robbins et al., 2021).

An interesting finding from my thesis is that proteomic and metabolomic changes were seen in the EBC of participants with a history of BPD, but this was not seen in the urine samples, where changes

were found to be associated with current lung function phenotypes. EBC primarily samples the epithelial lining fluid of the airway (Davis et al., 2012), and thereby the proteome and metabolome of a specific anatomical site, whereas urine samples the systemic proteome and metabolome (Martelo-Vidal et al., 2022) and reflects biological processes occurring within the body as a whole. As discussed extensively throughout this thesis, BPD occurs as a result of early pulmonary injury in preterm infants (Jobe and Bancalari, 2001), and the data from this thesis suggests an ongoing, active, inflammatory process in childhood, resulting in tissue remodeling and oxidative stress, which appears to be limited to the airways. In addition, combination inhaler therapy with ICS/LABA appears to reverse some of these structural changes in the airways for those with BPD and current low lung function (PT_{low}). As also discussed in my thesis, whilst it is recognized that those with BPD are at risk of later life lung function impairments (Kotecha et al., 2022b, Doyle et al., 2019b), within the RHiNO cohort, BPD does not appear to be the best predictor for later respiratory impairments (Hart et al., 2022), and this may be the reason that although I found proteomic and metabolomic changes in the EBC of the BPD group in RHiNO, these changes did not consistently associate with lung function impairments or spirometry values.

6.3.2 Biological Mechanisms in Lung Function Groups

Rather than solely focusing on outcomes based upon historic diagnoses of respiratory compromise, namely BPD, I have also used clinical phenotypes based upon current spirometry impairments to assess for mechanisms underlying PLD in this cohort of preterm-born children. In this thesis, I have used both an $FEV_1 \leq 85\%$ predicted cut-off for low lung function, which was used in the RHiNO study as entry criteria to an RCT of inhaled therapies (Goulden et al., 2021), as well as specific patterns of lung function impairment, namely POLD and pPRISm, defined by using GLI LLN thresholds for lung function impairment (Quanjer et al., 2012), and defined for the first time in the RHiNO cohort (Cousins et al., 2023).
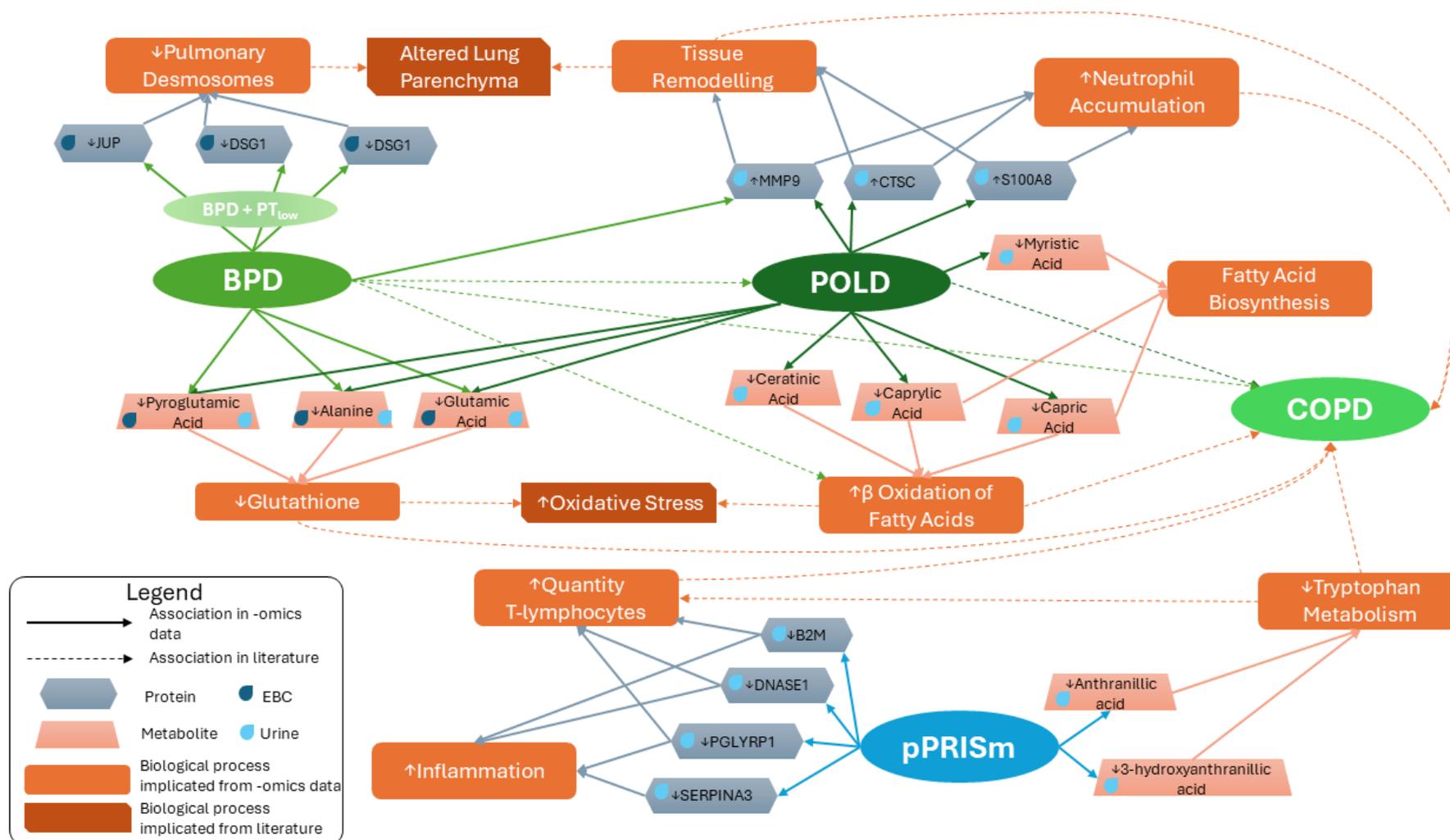
Figure 6-1: Summary diagram highlighting the key proteomic and metabolomic findings from EBC and urine samples and their relationships between different respiratory phenotypes of prematurity-associated lung disease.

Proteomic and metabolomic analyses of the urine samples revealed significant alterations associated with current lung function parameters using the 'LLN' (lower limit of normal) definitions described in the RHiNO study (Cousins et al., 2023), namely POLD and pPRISm (Figure 6-1). My analysis of the urinary proteome demonstrated multiple significantly altered protein abundances in the pPRISm group, which were linked to multiple biological processes concerned with inflammation and lymphocyte biology, with a suggestion of an alteration in the expression of CD4+ and CD8+ T-lymphocytes. A recent study of BAL fluid from preterm-born adults with a history of BPD has demonstrated a downregulation of CD4+ and upregulation of CD8+ cytotoxic T-cells compared to control subjects, as well as a positive correlation between CD4+/CD8+ ratio and FEV_1, and a negative correlation between numbers of CD8+ cells and FEV_1 and FEV_1/FVC (Um-Bergström et al., 2022). Similar alterations in CD4+ and CD8+ T-cells have also been observed in studies of adults with COPD (Eriksson Strom et al., 2020, Sales et al., 2017). I have also demonstrated changes to the urinary proteome in children with a POLD phenotype suggesting alterations of neutrophil biology, with increased abundance of three proteins (S100A8, MMP9 and CTSC) associated with tissue remodeling (Seren et al., 2021, Jerkic et al., 2020, Lorenz et al., 2008). The urinary metabolome analysis demonstrated multiple altered metabolic processes within the POLD group, including alterations in β-oxidation of very-long chain fatty acids and a reduction in glutathione metabolism. Conversely, few significantly altered metabolites were noted in the pPRISm group compared to the control groups.

As I have previously discussed in this thesis, there is increasing concern from the research evidence that individuals who are born preterm are at an increased risk of developing COPD in early adulthood (Bolton et al., 2015, Simpson et al., 2023, Doyle et al., 2019a). In a Scandinavian registry study of over two million individuals, those born extremely preterm had a 2-3 fold increased likelihood of having a obstructive respiratory pattern when aged 18-29 years compared to term-born controls, with a consequent increased risk of care episodes for COPD in young adulthood and middle age (Pulakka et

al., 2023). The increased abundances of S100A8, MMP9 and CTSC seen in the urinary proteome of the POLD group has also been observed in studies of adults with COPD (Lorenz et al., 2008, Wells et al., 2018, Merkel et al., 2005, Huang et al., 2020). The lungs of children with a POLD phenotype have also been demonstrated to have significant structural abnormalities on hyperpolarized [129]Xe ventilation and diffusion MRI imaging (Chan et al., 2023), further supporting the role of these tissue remodeling proteins in precipitating lung dysfunction for this group. The urine metabolome of the POLD group showed a reduction in metabolites which are indicators of glutathione metabolism, which I also observed in the EBC metabolome of those with BPD. Both analyses showed significant reductions of pyroglutamic acid, glutamic acid, and alanine in POLD and BPD compared to their respective control groups (Figure 6-1). Previously published data from the RHiNO study has demonstrated that a neonatal history of BPD is significantly associated with the development of a POLD phenotype in childhood (Cousins et al., 2023), and data from published meta-analysis (Gibbons et al., 2023) and longitudinal spirometry studies of preterm-born individuals (Doyle et al., 2019b) have also demonstrated that those with BPD demonstrate an increasingly obstructive respiratory phenotype as they age. Studies of adults with COPD have also demonstrated impairments in glutathione metabolism (Callejon-Leblic et al., 2019, Faucher et al., 2004), as have studies of neonates who later develop BPD, as discussed above (Grigg et al., 1993, Collard et al., 2004). The data from my analyses suggests a common mechanism for lung injury/impairment related to a lower quantity of glutathione, and, therefore, oxidative stress and impaired antioxidant mechanisms, in both BPD and POLD. This glutathione deficit may be present from birth and early neonatal lung injury; and may extend into development of COPD in later life. This hypothesis would need to be examined in longitudinal studies from the neonatal period and into adulthood.

PRISm has only recently been described in the adult literature (Higbee et al., 2022, Marott et al., 2021, Wan et al., 2021, Wijnant et al., 2020), and this phenotype in the preterm-born school-aged paediatric

population was first described by the RHiNO study (Cousins et al., 2023). In the adult population, PRISm has been associated with an increased risk of respiratory disease, cardiac disease, and all-cause mortality (Wan et al., 2021, Marott et al., 2021). Data from a follow-up study examining respiratory trajectories over a five-year period found that 50% of those with an initial PRISm phenotype developed COPD, however, interestingly 15% reverted to normal spirometry (Wijnant et al., 2020). Whilst the natural history and evolution of pPRISm requires further study, given the increased risk of significant respiratory-related morbidity and mortality in the adult population with this spirometry phenotype, it warrants further investigation in the preterm-born population.

The urine proteome analyses demonstrated associations between systemic alterations in inflammatory and immune processes in the pPRISm group, likely related to increases in inflammation and reductions in T-lymphocyte numbers, particularly CD4+ T-cells (Figure 6-1). As previously discussed in this thesis, alterations in T-lymphocyte numbers has been seen in bronchoalveolar lavage fluid from young adults with BPD, with a negative correlation between CD8+ cells and lung function (Um-Bergström et al., 2022), and CD8+ T-cell number is increased in histological studies of adolescents with a background of BPD (Galderisi et al., 2019). Whilst overall few urine metabolome differences were seen in the pPRISm group in my analyses, there was a suggestion of alterations in tryptophan metabolism on comparison with the term-born control subjects. Tryptophan is an essential amino acid required for protein synthesis and cellular proliferation, which has also been observed to be reduced in COPD exacerbations (Gulcev et al., 2016). Whilst the underlying biological mechanisms for development of PRISm in the adult population have yet to be fully investigated, many of the factors associated with development of PRISm, such as cigarette smoking, air pollution and obesity, are also associated with pulmonary and systemic inflammation (Huang et al., 2024). High-resolution computed tomography studies of adult PRISm patients have demonstrated minimal structural changes in this phenotype compared to controls and COPD patients, and overall, the underlying biological

mechanisms require further research (Lu et al., 2022, Huang et al., 2024). These analyses have identified that there are alterations in systemic biology in the pPRISm group, and further research is warranted to understand the evolution and longer-term implications of these mechanisms in the preterm-born population, and how they relate to the outcomes observed in the adult population.

6.3.3 Effect of Inhaled Therapies on Identified Biological Mechanisms

Despite the well-studied lung function impairments in later life following preterm birth (Kotecha et al., 2022b), there has been relatively little research in how to treat PLD, or how inhaled therapies may act on underlying pathological mechanisms. Two studies from the pre-/peri-surfactant era examined the use of inhaled corticosteroids in preterm-born children (Table 6-2). Chan et al. performed a double-blind, placebo-controlled cross-over study of fifteen low birth weight children, aged eight years who were born with a mean gestational age of 30.5 weeks. All were born before the routine use of exogenous pulmonary surfactant replacement. The children received twice daily 200µg of inhaled beclomethasone dipropionate for four weeks or placebo in a cross-over study. There was no significant effect on peak expiratory flow rate (PEFR), forced expiratory volume in one second (FEV_1) or airway hyperresponsiveness following treatment (Chan and Silverman, 1993). In the other study occurring during introduction of exogenous surfactant treatment, Pelkonen et al. studied eighteen children (median gestation at birth of 28 weeks and median of age 10.1 years) who had evidence of reversibility of airway obstruction as assessed by response to short-acting β2-agonists. These children received inhaled budesonide (0.8 mg/m^2/day for 1 month followed by 0.4 mg/m^2/day for 3 months) over a four-month period. No significant difference was noted for percent predicted FEV_1 (median 74% before and after treatment). However, PEFR diurnal variability improved suggesting decreased bronchial lability after treatment (Pelkonen et al., 2001). Neither of these studies examined underlying biological mechanisms. A more recently published study (Urs et al., 2023), examining a group of preterm-born children who would have experienced a contemporary standard of neonatal care, trialed the use of

either twice daily inhaled fluticasone propionate 125µg (another form of inhaled corticosteroid) or placebo for 12 weeks. However, reduced lung function was not a pre-requisite for entry into this trial.

Overall, a modest improvement was seen, with a 4% improvement in FEV_1, although a subset (23% of the cohort) had a more marked improvement of 6% (Table 6-2). Although it has been shown previously that fractional exhaled nitric oxide (FE_{NO}) is not increased in preterm-born individuals (Course et al., 2019), a decrease was seen following treatment, suggesting an element of airway inflammation in PLD. The RHiNO trial was the first to assess the use of a LABA in combination with ICS (as described in my introduction chapter, section 1.6.3.4) for the management of PLD, finding a clinically significant 14% increase in FEV_1 for combination inhaler therapy (Goulden et al., 2021), and the analyses in this thesis represent the first time the effect of inhaled treatments on underlying biological mechanisms associated with PLD phenotypes has been studied.

Study (author, year)	Country	Inhaled therapies investigated (Drug, dose, duration)	Active (n)	Placebo (n)	Results
Chan et al. 1993	United Kingdom	Beclomethasone dipropionate 200 µg twice daily	15	15	Mean FEV_1 1.3 litres post treatment in beclomethasone group, 1.25 litres in placebo group.
Pelkonen et al. 2001	Finland	Budesonide 0.8 mg/m^2/day for 1 month followed by 0.4 mg/m^2/day for 3 months	18	N/A	%pred FEV_1 74% pre and post treatment (p=0.50) Significant improvement in PEFR diurnal variation (p=0.02)
RHiNO (Goulden et al. 2022)	United Kingdom	Fluticasone propionate 100 µg twice daily for 12 weeks	20	14	%pred FEV_1 increased 7.7% (95% CI -0.3 to 15.7, p=0.16) Mean FE_{NO} reduced from 29.8 to 15.8 ppb
		Fluticasone propionate 100 µg and Salmeterol xinafoate 50 µg twice daily for 12 weeks	19	14	%pred FEV_1 increased 14.1% (95% CI 7.3 to 21.0, p=0.002) Mean FE_{NO} reduced from 25.2 to 15.9 ppb
PICSI (Urs et al. 2023)	Australia	Fluticasone propionate 125 µg twice daily for 12 weeks	87	83	0.30 (95% CI 0.15, 0.45) improvement in FEV_1 z-score Reduced FEV_1 bronchodilator response: -2.21 (-4.68, -0.26) z-score Mean FE_{NO} reduced from 15.2 to 10.5 ppb (p=<0.05)

Table 6-2 Summary of evidence from trials using inhaled corticosteroids in preterm-born children

FEV_1: forced expiratory volume in one second. FVC: forced vital capacity. FEF_{50}: forced expiratory flow over middle 50% of FVC. PEFR: peak expiratory flow rate. FE_{NO}: fractional exhaled nitric oxide. ppb: parts per billion. CI: Confidence interval. %pred: percentage predicted.

The most robust data from my analyses on the effect of inhaled therapies on biological mechanisms underlying PLD comes from the EBC proteomic analysis. This demonstrated that the reduced abundances of the desmosome-constituent proteins (DSG1, DSC1 and JUP) seen in those in the PT_{low} group with a history of BPD could be reversed to abundances comparable with the term-born controls using 12-weeks of combined (ICS/LABA) inhaler therapy. In this group, there was also reduced abundance of HSPB1 following combination inhaler treatment, with HSPB1 being important in the response to oxidative stress and inflammation (Acunzo et al., 2012). This combined treatment was associated with a significant 14% increase in FEV_1 when compared to placebo (Goulden et al., 2021), and the improvement in lung function is likely related to a reduction in airway inflammation and structural repair of the parenchyma, and, thereby, improved lung mechanics. There is a suggestion of a synergistic effect between ICS and LABA, as these changes were not seen in the ICS monotherapy group. In addition, no biologically significant changes were seen in the EBC metabolome within the treatment groups for those in the PT_{low} group. Regarding the urine samples, no significantly altered proteins in the POLD or pPRISm groups compared to controls were significantly altered by inhaled therapies. In the urine metabolome, ICS treatment significantly increased indole-3-lactate, which was seen as low in both the POLD and pPRISm groups at baseline, when compared to the preterm controls. However, this change was not seen in the combination inhaler therapy group. Indole-3-lactate has previously been demonstrated to play a role in the regulation of inflammation, with a multiomics-based study of human allergy demonstrating upregulation of interleukins-4 and -6 in macrophages (Zhen et al., 2022), and in murine models of necrotizing enterocolitis and intestinal disease, indole-3-lactate has been shown to directly inhibit genes that produce inflammatory cytokines (Huang et al., 2021), suppresses inflammatory T-cells and induces immunoregulatory T-cells (Cervantes-Barragan et al., 2017). The role of indole-3-lactate in lung pathology and the effect of corticosteroids on its quantity requires further study. Indole-3-lactate is a downstream product of tryptophan

metabolism, and MSEA had identified this as a significantly altered metabolic process in the pPRISm group compared to term-born controls, although indole-3-lactate was not included in that linkage.

6.3.4 Summary of findings:

Overall, the data presented in this thesis demonstrates that both risk factors for later PLD (including BPD) and current spirometry-based phenotypes (including POLD and pPRISm) are differentially associated with proteomic and metabolomic alterations, associated with differing biological processes. The EBC proteome in those with a history of BPD demonstrates changes in structural proteins, including cytokeratins and constituents of desmosomes, which are potentially amenable to combined inhaled therapies. The EBC metabolome for those with a history of BPD shows changes in metabolites related to processes including antioxidant mechanisms and oxidative stress, with common metabolite changes seen in the urine metabolome of those with a POLD phenotype. In addition, the urine metabolome in individuals with a POLD phenotype demonstrates changes associated with fatty acid metabolism and β-oxidation of fatty acids, with similar findings having been observed in adults with COPD. The urine proteome of those with a POLD phenotype demonstrates changes associated with and upregulation of neutrophil activation and tissue remodeling proteases. The urine proteome for those with a pPRISm phenotype shows multiple changes related to inflammation and T-lymphocyte biology, with a suggestion of altered CD4+ and CD8+ expression regulation, which has also been observed in studies of adults with a history of BPD and adults with COPD.

Overall, the data presented in this thesis suggests that despite the initial pulmonary insult, secondary to preterm birth, having occurred several years previously, there are ongoing, active biological processes related to inflammation, lymphocyte and neutrophil activation,

cellular metabolism and antioxidant mechanisms present in these children with PLD, which may be amenable to existing and novel treatments.

6.4 Strengths and Limitations

6.4.1 RHiNO study population

This thesis has used data from the RHiNO study, which represents the largest cohort of preterm-born children studied who would have experienced a modern standard of neonatal care, with high antenatal corticosteroid exposure and the routine use of exogenous pulmonary surfactant replacement. Consequently, my findings should be applicable to the current preterm neonatal population as they age. There are some limitations with the RHiNO cohort, namely that it predominantly includes an ethnically white population in which lower socio-economic groups are underrepresented (Hart et al., 2022), which is a common issue with cohort studies. Despite the large number of preterm-born children recruited, overall, the numbers of children with a PLD phenotype were relatively small. In addition, although the RCT stage of RHiNO included 53 children, and represents one of the few studies of the treatment of PLD (Goulden et al., 2021), the numbers per treatment group were relatively small. Overall, this may have limited the ability to detect significant proteomic and metabolomic changes in these groups. The findings from the EBC and urine proteomic and metabolomic analyses do require replication in a validation cohort, however there is a lack of comparable cohorts both within the UK and internationally with the same level of phenotyping and participant numbers. As the numbers with a PLD phenotype were relatively small overall within RHiNO this also limited the ability to perform internal validation studies. Also, these findings represent the first-time biological mechanisms of inhaled therapies have been examined in PLD and owing to the small numbers involved in the RCT, this has likely limited the discriminating power of my analyses to detect all the relevant proteomic and

metabolomic changes. Future studies should aim to target recruitment of larger numbers of subjects with a PLD phenotype.

6.4.2 EBC Samples

EBC is emerging as an attractive method of taking biological samples from the airways in a safe and non-invasive manner, as evidenced by a recent ERS technical guideline for its use in the study of respiratory diseases (Horvath et al., 2017). It is ideal for studying lung disease in a healthy paediatric population, where it can be sampled from large numbers of children in an acceptable and cost-effective manner. However, unlike invasively obtained samples from procedures such as bronchoscopy, it is not possible to know which lung compartment and section of the respiratory 'tree' (i.e. main bronchi, or lower branches of bronchii) is being sampled by EBC collection, as it gives more of a global impression of the biology of the airways. Also, as it is a passively collected sample and the epithelial lining fluid content is overall quite low, there may have been biologically relevant proteins or metabolites present in the samples that were below the limit of detection of the mass spectrometry technology used for analysis. ELF concentration in EBC is technically challenging to quantify and at present there is no accepted internal standard or reference compound present within EBC to adjust for dilutional effects (Horvath et al., 2017). Ideally, it would be preferable to confirm the findings of my analyses with more invasively collected samples such as broncheoalveolar lavage, however, there are significant ethical considerations of performing such invasive sampling in an otherwise healthy paediatric population.

6.4.3 Urine Samples

Urine is an easily available biofluid that provides the opportunity to study disease mechanisms, making it ideal for use in large cohort studies of clinically well paediatric

participants. The urine sampling performed in the RHiNO study was opportunistic at a time of assessment that was convenient for the family to attend. Therefore, the urine samples are from random timepoints throughout the day, and not necessarily early morning urine samples, which may have accumulated more proteomic and metabolomic changes from a prolonged period without regular micturition. In addition, 24-hour urine collections would have provided large sample volumes and may have revealed greater proteomic and metabolomic differences between subjects. Dietary intake can alter the urine metabolome (Stratakis et al., 2022), but unfortunately this data was not collected in the RHiNO study and therefore I have been unable to adjust for this in my analyses. Overall, the protein and metabolite content of the urine samples was far greater than EBC, as would be expected, but there may be some biologically relevant molecules present in the urine samples that were below the limit of detection for the mass spectrometry technology used for my analyses.

6.4.4 Data Normalisation

Normalisation of proteomic and metabolomic data is known to be challenging, particularly for excretory samples (Valikangas et al., 2018, Callister et al., 2006, Li et al., 2022) as have been used in this thesis. As discussed above (section 6.4.2), whilst EBC is increasingly used in the study of respiratory pathologies, at present there remains no widely accepted internal or external standard or method to normalize the resulting data (Horvath et al., 2017). For the proteomic analysis of EBC the same total protein load was used for each sample by the University of Bristol Proteomics Facility where possible, to attempt to ensure comparable results between samples and MS runs. This was not possible for the EBC metabolomic analysis, however, reassuringly, there appeared to be no significant relationship between EBC volume collected and both key protein and metabolite findings in the analysis.

Whilst urinary creatinine is a more widely accepted internal normalization factor for urine metabolomics experiments (Li et al., 2022), normalization of the urine proteome dataset was one of the more challenging aspects of this thesis, as described in section 4.2.4. The method employed, namely normalizing the data set to the median total protein content of each MS-run and excluding those samples with a +/-2-fold difference in total protein content, operated within the technical limitations of the Proteome Discoverer v2.1 software and ensured robust statistical comparisons between included samples. However, it did also result in a large proportion of the samples analysed being excluded from the final data analysis. Whilst there were minimal differences overall between the included and excluded samples, this did reduce the number of samples included in the analysis and may have reduced the statistical power of the experiment to detect significant differences in some proteins.

6.4.5 Proteome and Metabolome Analyses

The analyses presented in this thesis have used data from the analysis of excreted biological samples (i.e., EBC and urine), which are not as rich in protein and metabolite content as blood or whole cell/tissue lysates. As a result, there may be biologically important proteins and metabolites present in the samples which did not reach the limit of detection for the mass spectrometry analyses used. This is relevant not only for identifying individual compounds, but also the implicated biological processes, as this is based upon enrichment analysis methodology (IPA® and MSEA), which rely on over-representation of compounds to map to affected biological systems.

Both the proteomic and metabolomic analyses presented in this thesis were untargeted, meaning that the biological samples were analysed for all detectable compounds, and not aimed at identifying specific proteins or metabolites, so as to give as complete an analysis of the proteome and metabolome as possible. However, for the purposes of the data analyses,

this was restricted to those proteins and metabolites that could be identified and annotated with a high degree of confidence. This was important for the enrichment analyses and mapping to relevant biological processes, as these can only be performed with known compounds. However, there may have been significant differences in patterns in molecules that could not be accurately identified and annotated between the groups that may have had use in discriminating one phenotype from another.

Tandem Mass Tag (TMT) is a robust methodology in proteomics for comparing the same protein's abundance between samples, which has generated reliable results on changes in protein levels between clinical groups and before and after treatment. However, the relative abundances produced by TMT methodology does not allow for robust comparison of different protein abundances within the same sample (Pappireddi et al., 2019). This limits the ability to determine the effects of different proteins on one another within a single sample.

Whilst mass spectrometry is a highly sensitive analysis method, the reproducibility of findings between experiments has been challenging in both proteomic (Tabb et al., 2010) and metabolomic analyses (Ghosh et al., 2021), for example due to technical limitations of the equipment employed, different analysis platforms and workflows, 'noise' detected by the mass spectrometer and differences in characteristics between respective studies subjects. Unlike the genome, the proteome and metabolome are temporally dynamic and influenced by environmental factors, as well as host factors, meaning that finding equivalent cohorts that have been recruited and sampled in the same manner in order to replicate findings is challenging. These limitations also impact the application of these analysis methods to clinical practice.

6.4.6 Statistical Analysis Approaches

I have used robust statistical approaches for analysing my proteomic and metabolomic data sets that are well-established in the literature (Wang et al., 2022, Niziol et al., 2023, Juang et al., 2021, Bond et al., 2019), and for the metabolomics data, I have employed the use of MetaboAnalyst, which is frequently used by other groups analysing metabolome datasets (Pang et al., 2022, Chen et al., 2022). During the course of these analyses I have become proficient with using R (R Core Team, 2021) and learning to code in the R language, as well as employing a range of R packages. I have developed an understanding of handling the large datasets produced by MS-based analysis workflows, and the statistical techniques employed in proteomic and metabolomic analyses. Enrichment analyses and use of software that maps patterns of changes in the proteome and metabolome to biological processes have also been key skills I have learnt.

Owing to the exploratory nature of my analyses, I have used a p-value of <0.05 as a threshold for statistical significance in order to not exclude potentially important biological discoveries, which is a higher threshold than some other authors have used in published works. However, many of my findings have a much higher level of statistical significance.

Owing to the excretory nature of the samples, and, therefore, lower protein and metabolite content, I have been limited in some of the statistical approaches I could take to analyse my data. To ensure robust results, some analyses were restricted to include only proteins or metabolites that were either present in every sample analysed, or a high majority. This is particularly true for the EBC samples (see sections 2.3 and 3.3). This again may have excluded some biologically relevant compounds from enrichment analyses. Methods such as Principle Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) have been employed by other authors in the analysis on -omics data as multivariable dimension reduction techniques to identify patterns of changes able to discriminate phenotypes or

clinical groups (Alonso et al., 2015, Cui et al., 2022). However, this was not appropriate for my data owing to the low overall number of proteins or metabolites common to every sample analysed.

6.5 Implications for future research and clinical practice

There are a number of ways that the research presented in my thesis should be developed in future studies. Firstly, the results for this exploratory analysis of the EBC and urine proteome and metabolome analysis require replication in a validation cohort, using a population of different preterm- and term-born children of a similar age and who would have experienced a contemporary standard of neonatal care. If successful, then comparing to other age groups would be appropriate.

Direct measurement of compounds of interest identified in these analyses, using techniques such as Enzyme-Linked Immunosorbent Assay (ELISA), are required to accurately quantify their concentration, and would be necessary before any of the compounds of interest could potentially be employed clinically as a biomarker.

Longitudinal studies using these analytical techniques following a cohort of preterm infants from birth into childhood and adulthood would be useful to ascertain whether the implicated biological mechanisms are the same or evolve as subjects age and develop lung function impairments. In addition, this would ascertain whether the identified proteins and metabolites of interest could predict lung function impairments in later life, which would potentially allow for early risk stratification of preterm-born individuals clinically, and raises the possibility of targeted therapy to prevent future lung function impairments.

With regard to the potential for treatments to alter these identified metabolites of interest and biological processes, my analysis only detected significant relevant alterations in the EBC

proteome at the end of treatment. The proteome of the lungs (along with spirometry) also needs to be studied a period of time after the discontinuation of treatment to see whether the changes seen within the EBC proteome following combination inhaler therapy are maintained, along with the improvement seen in FEV_1 (Goulden et al., 2021). This will be important for clinical management to know whether a 12-week course of treatment will suffice, or whether treatment needs to be continued longer term in order for the lung function improvements to be maintained. In the future, studies of inhaled therapies for treatment of PLD should consider obtaining more invasively collected samples, such as broncheoalveolar lavage fluid or tissue biopsy, which may not only be richer in protein and metabolite content but also ensure sample acquisition from the distal airways and acinar units. This may overcome some of the issues identified with EBC in 6.4.2.

6.6 Thesis Conclusions

In this thesis, I have presented data that demonstrates that preterm-born school-aged children with differing phenotypes of PLD have ongoing active changes in biological processes associated with inflammation, immune system function and oxidative stress, as well as alterations in lung parenchyma structure.

In those with a history of BPD, there was evidence of persistent parenchymal structural changes primarily related to desmosomes, which for those with BPD and low lung function, appears to be modifiable with combination inhaler therapy (ICS/LABA). Glutathione metabolism, and thereby antioxidant defenses, appear to be impaired in both those with a history of BPD and a current POLD phenotype. In addition, POLD was associated with alterations in β-oxidation of fatty acids, fatty acid metabolism and neutrophil activity. These biological processes are also seen in adult subjects with COPD. In those with a pPRISm

phenotype, there appears to be altered T-lymphocyte biology and evidence of persistent inflammation.

Differing phenotypes of PLD are associated with different underlying biological processes and highlight the need for greater understanding of current lung function profiles in preterm-born subjects as they age to appropriately understand the underlying pathophysiology. Given that many of the biological processes I have identified have also been seen in studies of adults with COPD, there is a suggestion that some PLD phenotypes share a common pathophysiology.

7 References

ACUNZO, J., KATSOGIANNOU, M. & ROCCHI, P. 2012. Small heat shock proteins HSP27 (HspB1), alphaB-crystallin (HspB5) and HSP22 (HspB8) as regulators of cell death. *Int J Biochem Cell Biol,* 44**,** 1622-31.

AGUSTI, A. & FANER, R. 2019. Lung function trajectories in health and disease. *The Lancet Respiratory Medicine,* 7**,** 358-364.

AHMED, S., ODUMADE, O. A., VAN ZALM, P., SMOLEN, K. K., FUJIMURA, K., MUNTEL, J., ROTUNNO, M. S., WINSTON, A. B., STEEN, J. A., PARAD, R. B., VAN MARTER, L. J., KOUREMBANAS, S. & STEEN, H. 2022. Urine Proteomics for Noninvasive Monitoring of Biomarkers in Bronchopulmonary Dysplasia. *Neonatology,* 119**,** 193-203.

AKSENOV, A. A., ZAMURUYEV, K. O., PASAMONTES, A., BROWN, J. F., SCHIVO, M., FOUTOUHI, S., WEIMER, B. C., KENYON, N. J. & DAVIS, C. E. 2017. Analytical methodologies for broad metabolite coverage of exhaled breath condensate. *J Chromatogr B Analyt Technol Biomed Life Sci,* 1061-1062**,** 17-25.

ALBERTINE, K. H., JONES, G. P., STARCHER, B. C., BOHNSACK, J. F., DAVIS, P. L., CHO, S. C., CARLTON, D. P. & BLAND, R. D. 1999. Chronic lung injury in preterm lambs. Disordered respiratory tract development. *Am J Respir Crit Care Med,* 159**,** 945-58.

ALONSO, A., MARSAL, S. & JULIA, A. 2015. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol,* 3**,** 23.

AMBRUSO, S. L., GIL, H. W., FOX, B., PARK, B., ALTMANN, C., BAGCHI, R. A., BAKER, P. R., 2ND, REISZ, J. A. & FAUBEL, S. 2021. Lung metabolomics after ischemic acute kidney injury reveals increased oxidative stress, altered energy production, and ATP depletion. *Am J Physiol Lung Cell Mol Physiol,* 321**,** L50-L64.

ANDERSSON, K., SHEBANI, E. B., MAKEEVA, N., ROOMANS, G. M. & SERVETNYK, Z. 2010. Corticosteroids and montelukast: effects on airway epithelial and human umbilical vein endothelial cells. *Lung,* 188**,** 209-16.

ASKIE, L. M., DARLOW, B. A., FINER, N., SCHMIDT, B., STENSON, B., TARNOW-MORDI, W., DAVIS, P. G., CARLO, W. A., BROCKLEHURST, P., DAVIES, L. C., DAS, A., RICH, W., GANTZ, M. G., ROBERTS, R. S., WHYTE, R. K., COSTANTINI, L., POETS, C., ASZTALOS, E., BATTIN, M., HALLIDAY, H. L., MARLOW, N., TIN, W., KING, A., JUSZCZAK, E., MORLEY, C. J., DOYLE, L. W., GEBSKI, V., HUNTER, K. E., SIMES, R. J. & NEONATAL OXYGENATION PROSPECTIVE META-ANALYSIS, C. 2018. Association Between Oxygen Saturation Targeting and Death or Disability in Extremely Preterm Infants in the Neonatal Oxygenation Prospective Meta-analysis Collaboration. *JAMA,* 319**,** 2190-2201.

ASLAM, B., BASIT, M., NISAR, M. A., KHURSHID, M. & RASOOL, M. H. 2017. Proteomics: Technologies and Their Applications. *J Chromatogr Sci,* 55**,** 182-196.

AUKLAND, S. M., ROSENDAHL, K., OWENS, C. M., FOSSE, K. R., EIDE, G. E. & HALVORSEN, T. 2009. Neonatal bronchopulmonary dysplasia predicts abnormal pulmonary HRCT scans in long-term survivors of extreme preterm birth. *Thorax,* 64**,** 405-10.

BALFOUR-LYNN, I. M., FIELD, D. J., GRINGRAS, P., HICKS, B., JARDINE, E., JONES, R. C., MAGEE, A. G., PRIMHAK, R. A., SAMUELS, M. P., SHAW, N. J., STEVENS, S., SULLIVAN, C., TAYLOR, J. A., WALLIS, C. & PAEDIATRIC SECTION OF THE HOME OXYGEN GUIDELINE DEVELOPMENT GROUP OF THE, B. T. S. S. O. C. C. 2009. BTS guidelines for home oxygen in children. *Thorax,* 64 Suppl 2**,** ii1-26.

BANCALARI, E., ABDENOUR, G. E., FELLER, R. & GANNON, J. 1979. Bronchopulmonary dysplasia: clinical presentation. *J Pediatr,* 95**,** 819-23.

BAO, K., YUAN, W., ZHOU, Y., CHEN, Y., YU, X., WANG, X., JIA, Z., YU, X., WANG, X., YAO, L., WANG, S., XU, Y., ZHANG, Y., ZHENG, J. & HONG, M. 2019. A Chinese Prescription Yu-Ping-Feng-San Administered in Remission Restores Bronchial Epithelial Barrier to Inhibit House Dust Mite-Induced Asthma Recurrence. *Front Pharmacol,* 10**,** 1698.

BAPM. 2019. Perinatal Management of Extreme Preterm Birth before 27 weeks of gestation: A British Association of Perinatal Medicine (BAPM) Framework. Available: https://hubble-live-assets.s3.amazonaws.com/bapm/file_asset/file/30/Extreme_Preterm_28-11-19_FINAL.pdf.

BARALDI, E., GIORDANO, G., STOCCHERO, M., MOSCHINO, L., ZARAMELLA, P., TRAN, M. R., CARRARO, S., ROMERO, R. & GERVASI, M. T. 2016. Untargeted Metabolomic Analysis of Amniotic Fluid in the Prediction of Preterm Delivery and Bronchopulmonary Dysplasia. *PLoS One,* 11**,** e0164211.

BARDSEN, T., ROKSUND, O. D., BENESTAD, M. R., HUFTHAMMER, K. O., CLEMM, H. H., MIKALSEN, I. B., OYMAR, K., MARKESTAD, T., HALVORSEN, T. & VOLLSAETER, M. 2022. Tracking of lung function from 10 to 35 years after being born extremely preterm or with extremely low birth weight. *Thorax,* 77**,** 790-798.

BARROS, F. C., BHUTTA, Z. A., BATRA, M., HANSEN, T. N., VICTORA, C. G., RUBENS, C. E. & GROUP, G. R. 2010. Global report on preterm birth and stillbirth (3 of 7): evidence for effectiveness of interventions. *BMC Pregnancy Childbirth,* 10 Suppl 1**,** S3.

BATISTA-GONZALEZ, A., VIDAL, R., CRIOLLO, A. & CARRENO, L. J. 2019. New Insights on the Role of Lipid Metabolism in the Metabolic Reprogramming of Macrophages. *Front Immunol,* 10**,** 2993.

BEETON, M. L., MAXWELL, N. C., DAVIES, P. L., NUTTALL, D., MCGREAL, E., CHAKRABORTY, M., SPILLER, O. B. & KOTECHA, S. 2011. Role of pulmonary infection in the development of chronic lung disease of prematurity. *Eur Respir J,* 37**,** 1424-30.

BELGRAVE, D. C. M., GRANELL, R., TURNER, S. W., CURTIN, J. A., BUCHAN, I. E., LE SOUEF, P. N., SIMPSON, A., HENDERSON, A. J. & CUSTOVIC, A. 2018. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *Lancet Respir Med,* 6**,** 526-534.

BERRINGTON, J. E., HEARN, R. I., BYTHELL, M., WRIGHT, C. & EMBLETON, N. D. 2012. Deaths in preterm infants: changing pathology over 2 decades. *J Pediatr,* 160**,** 49-53 e1.

BEZERRA, F. S., LANZETTI, M., NESI, R. T., NAGATO, A. C., SILVA, C. P. E., KENNEDY-FEITOSA, E., MELO, A. C., CATTANI-CAVALIERI, I., PORTO, L. C. & VALENCA, S. S. 2023. Oxidative Stress and Inflammation in Acute and Chronic Lung Injuries. *Antioxidants (Basel),* 12.

BHANDARI, A. & PANITCH, H. B. 2006. Pulmonary outcomes in bronchopulmonary dysplasia. *Semin Perinatol,* 30**,** 219-26.

BJORKLUND, L. J., INGIMARSSON, J., CURSTEDT, T., JOHN, J., ROBERTSON, B., WERNER, O. & VILSTRUP, C. T. 1997. Manual ventilation with a few large breaths at birth compromises the therapeutic effect of subsequent surfactant replacement in immature lambs. *Pediatr Res,* 42**,** 348-55.

BLENCOWE, H., COUSENS, S., OESTERGAARD, M. Z., CHOU, D., MOLLER, A. B., NARWAL, R., ADLER, A., VERA GARCIA, C., ROHDE, S., SAY, L. & LAWN, J. E. 2012. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet,* 379**,** 2162-72.

BLOEMEN, K., VAN DEN HEUVEL, R., GOVARTS, E., HOOYBERGHS, J., NELEN, V., WITTERS, E., DESAGER, K. & SCHOETERS, G. 2011. A new approach to study exhaled proteins as potential biomarkers for asthma. *Clin Exp Allergy,* 41**,** 346-56.

BOLTON, C. E., BUSH, A., HURST, J. R., KOTECHA, S. & MCGARVEY, L. 2015. Lung consequences in adults born prematurely. *Thorax,* 70**,** 574-80.

BONADIES, L., CAVICCHIOLO, M. E., PRIANTE, E., MOSCHINO, L. & BARALDI, E. 2023. Prematurity and BPD: what general pediatricians should know. *Eur J Pediatr,* 182**,** 1505-1516.

BOND, A. R., IACOBAZZI, D., ABDUL-GHANI, S., GHORBEL, M. T., HEESOM, K. J., GEORGE, S. J., CAPUTO, M., SULEIMAN, M. S. & TULLOH, R. M. 2019. The cardiac proteome in patients with congenital ventricular septal defect: A comparative study between right atria and right ventricles. *J Proteomics,* 191**,** 107-113.

BORRILL, Z. L., ROY, K. & SINGH, D. 2008. Exhaled breath condensate biomarkers in COPD. *Eur Respir J,* 32**,** 472-86.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem,* 72**,** 248-54.

BREW, N., HOOPER, S. B., ALLISON, B. J., WALLACE, M. J. & HARDING, R. 2011. Injury and repair in the very immature lung following brief mechanical ventilation. *Am J Physiol Lung Cell Mol Physiol,* 301**,** L917-26.

BROSNAN, J. T. 2003. Interorgan amino acid transport and its regulation. *J Nutr,* 133**,** 2068S-2072S.

BROUGHTON, S., THOMAS, M. R., MARSTON, L., CALVERT, S. A., MARLOW, N., PEACOCK, J. L., RAFFERTY, G. F. & GREENOUGH, A. 2007. Very prematurely born infants wheezing at follow-up: lung function and risk factors. *Arch Dis Child,* 92**,** 776-80.

BUCZYNSKI, B. W., MADUEKWE, E. T. & O'REILLY, M. A. 2013. The role of hyperoxia in the pathogenesis of experimental BPD. *Semin Perinatol,* 37**,** 69-78.

BUSH, A., BUSST, C. M., KNIGHT, W. B., HISLOP, A. A., HAWORTH, S. G. & SHINEBOURNE, E. A. 1990. Changes in pulmonary circulation in severe bronchopulmonary dysplasia. *Arch Dis Child,* 65**,** 739-45.

CALLEJON-LEBLIC, B., PEREIRA-VEGA, A., VAZQUEZ-GANDULLO, E., SANCHEZ-RAMOS, J. L., GOMEZ-ARIZA, J. L. & GARCIA-BARRERA, T. 2019. Study of the metabolomic relationship between lung cancer and chronic obstructive pulmonary disease based on direct infusion mass spectrometry. *Biochimie,* 157**,** 111-122.

CALLISTER, S. J., BARRY, R. C., ADKINS, J. N., JOHNSON, E. T., QIAN, W. J., WEBB-ROBERTSON, B. J., SMITH, R. D. & LIPTON, M. S. 2006. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res,* 5**,** 277-86.

CAMPANELLA, A., DE SUMMA, S. & TOMMASI, S. 2019. Exhaled breath condensate biomarkers for lung cancer. *J Breath Res,* 13**,** 044002.

CANTIN, A. M., NORTH, S. L., HUBBARD, R. C. & CRYSTAL, R. G. 1987. Normal alveolar epithelial lining fluid contains high levels of glutathione. *J Appl Physiol (1985),* 63**,** 152-7.

CARAYOL, N., CAMPBELL, A., VACHIER, I., MAINPRICE, B., BOUSQUET, J., GODARD, P. & CHANEZ, P. 2002. Modulation of cadherin and catenins expression by tumor necrosis factor-alpha and dexamethasone in human bronchial epithelial cells. *Am J Respir Cell Mol Biol,* 26**,** 341-7.

CARRARO, S., GIORDANO, G., PIRILLO, P., MARETTI, M., RENIERO, F., COGO, P. E., PERILONGO, G., STOCCHERO, M. & BARALDI, E. 2015. Airway metabolic anomalies in adolescents with bronchopulmonary dysplasia: new insights from the metabolomic approach. *J Pediatr,* 166**,** 234-9 e1.

CAVALEIRO RUFO, J., PACIENCIA, I., MENDES, F. C., FARRAIA, M., RODOLFO, A., SILVA, D., DE OLIVEIRA FERNANDES, E., DELGADO, L. & MOREIRA, A. 2019. Exhaled breath condensate volatilome allows sensitive diagnosis of persistent asthma. *Allergy,* 74**,** 527-534.

CERVANTES-BARRAGAN, L., CHAI, J. N., TIANERO, M. D., DI LUCCIA, B., AHERN, P. P., MERRIMAN, J., CORTEZ, V. S., CAPARON, M. G., DONIA, M. S., GILFILLAN, S., CELLA, M., GORDON, J. I., HSIEH, C. S. & COLONNA, M. 2017. Lactobacillus reuteri induces gut intraepithelial CD4(+)CD8alphaalpha(+) T cells. *Science,* 357**,** 806-810.

CHAKRABORTY, M. & KOTECHA, S. 2013. Pulmonary surfactant in newborn infants and children. *Breathe,* 9**,** 476-488.

CHAKRABORTY, M., MCGREAL, E. P., DAVIES, P. L., NOWELL, M. A., JONES, S. & KOTECHA, S. 2013. Role of interleukin-6, its receptor and soluble gp130 in chronic lung disease of prematurity. *Neonatology,* 104**,** 161-7.

CHAKRABORTY, M., MCGREAL, E. P. & KOTECHA, S. 2010. Acute lung injury in preterm newborn infants: mechanisms and management. *Paediatr Respir Rev,* 11**,** 162-70; quiz 170.

CHAN, H. F., SMITH, L. J., BIANCARDI, A. M., BRAY, J., MARSHALL, H., HUGHES, P. J. C., COLLIER, G. J., RAO, M., NORQUAY, G., SWIFT, A. J., HART, K., COUSINS, M., WATKINS, W. J., WILD, J. M. & KOTECHA, S. 2023. Image Phenotyping of Preterm-Born Children Using Hyperpolarized (129)Xe Lung Magnetic Resonance Imaging and Multiple-Breath Washout. *Am J Respir Crit Care Med,* 207**,** 89-100.

CHAN, K. N. & SILVERMAN, M. 1993. Increased airway responsiveness in children of low birth weight at school age: effect of topical corticosteroids. *Arch Dis Child,* 69**,** 120-4.

CHEN, Y., LI, E. M. & XU, L. Y. 2022. Guide to Metabolomics Analysis: A Bioinformatics Workflow. *Metabolites,* 12.

CHU, S., HUANG, Q., ALVARES, K., YELDANDI, A. V., RAO, M. S. & REDDY, J. K. 1995. Transformation of mammalian cells by overexpressing H2O2-generating peroxisomal fatty acyl-CoA oxidase. *Proc Natl Acad Sci U S A,* 92**,** 7080-4.

CLYMAN, R. I., LIEBOWITZ, M., KAEMPF, J., ERDEVE, O., BULBUL, A., HAKANSSON, S., LINDQVIST, J., FAROOQI, A., KATHERIA, A., SAUBERAN, J., SINGH, J., NELSON, K., WICKREMASINGHE, A., DONG, L., HASSINGER, D. C., AUCOTT, S. W., HAYASHI, M., HEUCHAN, A. M., CAREY, W. A., DERRICK, M., FERNANDEZ, E., SANKAR, M., LEONE, T., PEREZ, J., SERIZE, A. & INVESTIGATORS, P.-T. T. 2019. PDA-TOLERATE Trial: An Exploratory Randomized Controlled Trial of Treatment of Moderate-to-Large Patent Ductus Arteriosus at 1 Week of Age. *J Pediatr,* 205**,** 41-48 e6.

COLLARD, K. J., GODECK, S., HOLLEY, J. E. & QUINN, M. W. 2004. Pulmonary antioxidant concentrations and oxidative damage in ventilated premature babies. *Arch Dis Child Fetal Neonatal Ed,* 89**,** F412-6.

CONG, Y. & ENDO, T. 2022. Multi-Omics and Artificial Intelligence-Guided Drug Repositioning: Prospects, Challenges, and Lessons Learned from COVID-19. *OMICS,* 26**,** 361-371.

CORWIN, B. K., TREMBATH, A. N. & HIBBS, A. M. 2018. Bronchopulmonary dysplasia appropriateness as a surrogate marker for long-term pulmonary outcomes: A Systematic review. *J Neonatal Perinatal Med,* 11**,** 121-130.

COSTELOE, K. L., HENNESSY, E. M., HAIDER, S., STACEY, F., MARLOW, N. & DRAPER, E. S. 2012. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ,* 345**,** e7976.

COURSE, C. & CHAKRABORTY, M. 2020. Management of Respiratory Distress Syndrome in Preterm Infants In Wales: A Full Audit Cycle of a Quality Improvement Project. *Sci Rep,* 10**,** 3536.

COURSE, C., WATKINS, W. J., MULLER, C. T., ODD, D., KOTECHA, S. & CHAKRABORTY, M. 2021. Volatile organic compounds as disease predictors in newborn infants: a systematic review. *J Breath Res,* 15.

COURSE, C. W., KOTECHA, S. & KOTECHA, S. J. 2019. Fractional exhaled nitric oxide in preterm-born subjects: A systematic review and meta-analysis. *Pediatr Pulmonol,* 54**,** 595-601.

COURSE, C. W., KOTECHA, S. J., COUSINS, M., HART, K., LOWE, J., WATKINS, W. J. & KOTECHA, S. 2023a. Association of Gestation and Fetal Growth Restriction on Cardiovascular Health in Preterm-Born Children. *J Pediatr,* 255**,** 42-49 e4.

COURSE, C. W., LEWIS, P. A., KOTECHA, S. J., COUSINS, M., HART, K., WATKINS, W. J., HEESOM, K. J. & KOTECHA, S. 2023b. Characterizing the urinary proteome of prematurity-associated lung disease in school-aged children. *Respir Res,* 24**,** 191.

COURSE, C. W., LEWIS, P. A., KOTECHA, S. J., COUSINS, M., HART, K., WATKINS, W. J., HEESOM, K. J. & KOTECHA, S. 2023c. Modulation of pulmonary desmosomes by inhaler therapy in preterm-born children with bronchopulmonary dysplasia. *Sci Rep,* 13**,** 7330.

COUSINS, M., HART, K., KOTECHA, S. J., HENDERSON, A. J., WATKINS, W. J., BUSH, A. & KOTECHA, S. 2023. Characterising airway obstructive, dysanaptic and PRISm phenotypes of prematurity-associated lung disease. *Thorax*.

COUSINS, M., HART, K., WILLIAMS, E. M. & KOTECHA, S. 2022. Impaired exercise outcomes with significant bronchodilator responsiveness in children with prematurity-associated obstructive lung disease. *Pediatr Pulmonol*.

CUI, M., CHENG, C. & ZHANG, L. 2022. High-throughput proteomics: a methodological mini-review. *Lab Invest,* 102**,** 1170-1181.

DALEY-YATES, P., KEPPLER, B., BREALEY, N., SHABBIR, S., SINGH, D. & BARNES, N. 2022. Inhaled glucocorticoid-induced metabolome changes in asthma. *Eur J Endocrinol,* 187**,** 413-427.

DAVEY, A., MCAULEY, D. F. & O'KANE, C. M. 2011. Matrix metalloproteinases in acute lung injury: mediators of injury and drivers of repair. *Eur Respir J,* 38**,** 959-70.

DAVIES, P. L., SPILLER, O. B., BEETON, M. L., MAXWELL, N. C., REMOLD-O'DONNELL, E. & KOTECHA, S. 2010. Relationship of proteinases and proteinase inhibitors with microbial presence in chronic lung disease of prematurity. *Thorax,* 65**,** 246-51.

DAVIS, M. D., MONTPETIT, A. & HUNT, J. 2012. Exhaled breath condensate: an overview. *Immunol Allergy Clin North Am,* 32**,** 363-75.

DEMIRKAN, F. G., YILMAZ, E., HANGUL, M., OZTURK, D., DEMIRKAN, H., SOYLAK, M. & KOSE, M. 2018. Exhaled breath condensate magnesium levels of infants with bronchiolitis. *Turk J Pediatr,* 60**,** 535-539.

DENEKE, S. M., LYNCH, B. A. & FANBURG, B. L. 1985. Transient depletion of lung glutathione by diethylmaleate enhances oxygen toxicity. *J Appl Physiol (1985),* 58**,** 571-4.

DESHPANDE, S. A. & NORTHERN, V. 2003. The clinical and health economic burden of respiratory syncytial virus disease among children under 2 years of age in a defined geographical area. *Arch Dis Child,* 88**,** 1065-9.

DEVEREUX, G., STEELE, S., JAGELMAN, T., FIELDING, S., MUIRHEAD, R., BRADY, J., GRIERSON, C., BROOKER, R., WINTER, J., FARDON, T., MCCORMICK, J., HUANG, J. T. & MILLER, D. 2014. An observational study of matrix metalloproteinase (MMP)-9 in cystic fibrosis. *J Cyst Fibros,* 13**,** 557-63.

DOMACHOWSKE, J., MADHI, S. A., SIMOES, E. A. F., ATANASOVA, V., CABANAS, F., FURUNO, K., GARCIA-GARCIA, M. L., GRANTINA, I., NGUYEN, K. A., BROOKS, D., CHANG, Y., LEACH, A., TAKAS, T., YUAN, Y., GRIFFIN, M. P., MANKAD, V. S., VILLAFANA, T. & GROUP, M. S. 2022. Safety of Nirsevimab for RSV in Infants with Heart or Lung Disease or Prematurity. *N Engl J Med,* 386**,** 892-894.

DOYLE, L. W., ANDERSSON, S., BUSH, A., CHEONG, J. L. Y., CLEMM, H., EVENSEN, K. A. I., GOUGH, A., HALVORSEN, T., HOVI, P., KAJANTIE, E., LEE, K. J., MCGARVEY, L., NARANG, I., NASANEN-GILMORE, P., STEINSHAMN, S., VOLLSAETER, M., VRIJLANDT, E. & ADULTS BORN PRETERM INTERNATIONAL, C. 2019a. Expiratory airflow in late adolescence and early adulthood in individuals born very preterm or with very low birthweight compared with controls born at term or with normal birthweight: a meta-analysis of individual participant data. *Lancet Respir Med,* 7**,** 677-686.

DOYLE, L. W., IRVING, L., HAIKERWAL, A., LEE, K., RANGANATHAN, S. & CHEONG, J. 2019b. Airway obstruction in young adults born extremely preterm or extremely low birth weight in the postsurfactant era. *Thorax,* 74**,** 1147-1153.

DU BERRY, C., NESCI, C., CHEONG, J. L. Y., FITZGERALD, T., MAINZER, R., RANGANATHAN, S., DOYLE, L. W., VRIJLANDT, E. & WELSH, L. 2022. Long-term expiratory airflow of infants born moderate-late preterm: A systematic review and meta-analysis. *EClinicalMedicine,* 52**,** 101597.

DUPREE, E. J., JAYATHIRTHA, M., YORKEY, H., MIHASAN, M., PETRE, B. A. & DARIE, C. C. 2020. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes,* 8.

EDWARDS, E. M., EHRET, D. E. Y., SOLL, R. F. & HORBAR, J. D. 2024. Survival of Infants Born at 22 to 25 Weeks' Gestation Receiving Care in the NICU: 2020-2022. *Pediatrics*.

EDWARDS, M. O., KOTECHA, S. J., LOWE, J., RICHARDS, L., WATKINS, W. J. & KOTECHA, S. 2015. Early-term birth is a risk factor for wheezing in childhood: A cross-sectional population study. *J Allergy Clin Immunol,* 136**,** 581-587 e2.

EDWARDS, M. O., KOTECHA, S. J., LOWE, J., RICHARDS, L., WATKINS, W. J. & KOTECHA, S. 2016. Management of Prematurity-Associated Wheeze and Its Association with Atopy. *PLoS One,* 11**,** e0155695.

EHRENKRANZ, R. A., WALSH, M. C., VOHR, B. R., JOBE, A. H., WRIGHT, L. L., FANAROFF, A. A., WRAGE, L. A., POOLE, K., NATIONAL INSTITUTES OF CHILD, H. & HUMAN DEVELOPMENT NEONATAL RESEARCH, N. 2005. Validation of the National Institutes of Health consensus definition of bronchopulmonary dysplasia. *Pediatrics,* 116**,** 1353-60.

EL-KHUFFASH, A., MULLALY, R. & MCNAMARA, P. J. 2023. Patent ductus arteriosus, bronchopulmonary dysplasia and pulmonary hypertension-a complex conundrum with many phenotypes? *Pediatr Res,* 94**,** 416-417.

ELLSBURY, D. L., ACARREGUI, M. J., MCGUINNESS, G. A. & KLEIN, J. M. 2002. Variability in the use of supplemental oxygen for bronchopulmonary dysplasia. *J Pediatr,* 140**,** 247-9.

ERIKSSON STROM, J., POURAZAR, J., LINDER, R., BLOMBERG, A., LINDBERG, A., BUCHT, A. & BEHNDIG, A. F. 2020. Airway regulatory T cells are decreased in COPD with a rapid decline in lung function. *Respir Res,* 21**,** 330.

FABIANO, A., GAZZOLO, D., ZIMMERMANN, L. J., GAVILANES, A. W., PAOLILLO, P., FANOS, V., CABONI, P., BARBERINI, L., NOTO, A. & ATZORI, L. 2011. Metabolomic analysis of bronchoalveolar lavage fluid in preterm infants complicated by respiratory distress syndrome: preliminary results. *J Matern Fetal Neonatal Med,* 24 Suppl 2**,** 55-8.

FAN, S., KIND, T., CAJKA, T., HAZEN, S. L., TANG, W. H. W., KADDURAH-DAOUK, R., IRVIN, M. R., ARNETT, D. K., BARUPAL, D. K. & FIEHN, O. 2019. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal Chem,* 91**,** 3590-3596.

FANOS, V., PINTUS, M. C., LUSSU, M., ATZORI, L., NOTO, A., STRONATI, M., GUIMARAES, H., MARCIALIS, M. A., ROCHA, G., MORETTI, C., PAPOFF, P., LACERENZA, S., PUDDU, S., GIUFFRE, M., SERRAINO, F., MUSSAP, M. & CORSELLO, G. 2014. Urinary metabolomics of bronchopulmonary dysplasia (BPD): preliminary data at birth suggest it is a congenital disease. *J Matern Fetal Neonatal Med,* 27 Suppl 2**,** 39-45.

FARROW, K. N. & STEINHORN, R. H. 2012. Sildenafil therapy for bronchopulmonary dysplasia: not quite yet. *J Perinatol,* 32**,** 1-3.

FAUCHER, M., STEINBERG, J. G., BARBIER, D., HUG, F. & JAMMES, Y. 2004. Influence of chronic hypoxemia on peripheral muscle function and oxidative stress in humans. *Clin Physiol Funct Imaging,* 24**,** 75-84.

FIEHN, O., WOHLGEMUTH, G., SCHOLZ, M., KIND, T., LEE, D. Y., LU, Y., MOON, S. & NIKOLAU, B. 2008. Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J,* 53**,** 691-704.

FILIPPONE, M., BONETTO, G., CORRADI, M., FRIGO, A. C. & BARALDI, E. 2012. Evidence of unexpected oxidative stress in airways of adolescents born very pre-term. *European Respiratory Journal,* 40**,** 1253-1259.

FITZPATRICK, A. M., TEAGUE, W. G., BURWELL, L., BROWN, M. S., BROWN, L. A. & PROGRAM, N. N. S. A. R. 2011. Glutathione oxidation is associated with airway macrophage functional impairment in children with severe asthma. *Pediatr Res,* 69**,** 154-9.

FITZPATRICK, A. M., TEAGUE, W. G., HOLGUIN, F., YEH, M., BROWN, L. A. & SEVERE ASTHMA RESEARCH, P. 2009. Airway glutathione homeostasis is altered in children with severe asthma: evidence for oxidant stress. *J Allergy Clin Immunol,* 123**,** 146-152 e8.

FOERSTER, E. C., FAHRENKEMPER, T., RABE, U., GRAF, P. & SIES, H. 1981. Peroxisomal fatty acid oxidation as detected by H2O2 production in intact perfused rat liver. *Biochem J,* 196**,** 705-12.

GALDERISI, A., CALABRESE, F., FORTAREZZA, F., ABMAN, S. & BARALDI, E. 2019. Airway Histopathology of Adolescent Survivors of Bronchopulmonary Dysplasia. *J Pediatr,* 211**,** 215-218.

GALLACHER, D., MITCHELL, E., ALBER, D., WACH, R., KLEIN, N., MARCHESI, J. R. & KOTECHA, S. 2020. Dissimilarity of the gut-lung axis and dysbiosis of the lower airways in ventilated preterm infants. *Eur Respir J,* 55.

GARCIA-CAMPOS, M. A., ESPINAL-ENRIQUEZ, J. & HERNANDEZ-LEMUS, E. 2015. Pathway Analysis: State of the Art. *Front Physiol,* 6**,** 383.

GESSNER, C., DIHAZI, H., BRETTSCHNEIDER, S., HAMMERSCHMIDT, S., KUHN, H., ESCHRICH, K., KELLER, T., ENGELMANN, L., SACK, U. & WIRTZ, H. 2008. Presence of cytokeratins in exhaled breath condensate of mechanical ventilated patients. *Respir Med,* 102**,** 299-306.

GHOSH, T., PHILTRON, D., ZHANG, W., KECHRIS, K. & GHOSH, D. 2021. Reproducibility of mass spectrometry based metabolomics data. *BMC Bioinformatics,* 22**,** 423.

GIANAZZA, E., ALLEGRA, L., BUCCHIONI, E., EBERINI, I., PUGLISI, L., BLASI, F., TERZANO, C., WAIT, R. & SIRTORI, C. R. 2004. Increased keratin content detected by proteomic analysis of exhaled breath condensate from healthy persons who smoke. *Am J Med,* 117**,** 51-4.

GIBBONS, J. T. D., COURSE, C. W., EVANS, E. E., KOTECHA, S., KOTECHA, S. J. & SIMPSON, S. J. 2023. Increasing airway obstruction through life following bronchopulmonary dysplasia: a meta-analysis. *ERJ Open Research***,** 00046-2023.

GOLDENBERG, R. L., CULHANE, J. F., IAMS, J. D. & ROMERO, R. 2008. Epidemiology and causes of preterm birth. *Lancet,* 371**,** 75-84.

GOULDEN, N., COUSINS, M., HART, K., JENKINS, A., WILLETTS, G., YENDLE, L., DOULL, I., WILLIAMS, E. M., HOARE, Z. & KOTECHA, S. 2021. Inhaled Corticosteroids Alone and in Combination With Long-Acting beta2 Receptor Agonists to Treat Reduced Lung Function in Preterm-Born Children: A Randomized Clinical Trial. *JAMA Pediatr*.

GREEN, K. J., GETSIOS, S., TROYANOVSKY, S. & GODSEL, L. M. 2010. Intercellular junction assembly, dynamics, and homeostasis. *Cold Spring Harb Perspect Biol,* 2**,** a000125.

GRIGG, J., BARBER, A. & SILVERMAN, M. 1993. Bronchoalveolar lavage fluid glutathione in intubated premature infants. *Arch Dis Child,* 69**,** 49-51.

GULCEV, M., REILLY, C., GRIFFIN, T. J., BROECKLING, C. D., SANDRI, B. J., WITTHUHN, B. A., HODGSON, S. W., WOODRUFF, P. G. & WENDT, C. H. 2016. Tryptophan catabolism in acute exacerbations of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis,* 11**,** 2435-2446.

GUPTA, S., SUBHEDAR, N. V., BELL, J. L., FIELD, D., BOWLER, U., HUTCHISON, E., JOHNSON, S., KELSALL, W., PEPPERELL, J., ROBERTS, T., SINHA, S., STANBURY, K., WYLLIE, J., HARDY, P., JUSZCZAK, E. & BABY, O. C. G. 2024. Trial of Selective Early Treatment of Patent Ductus Arteriosus with Ibuprofen. *N Engl J Med,* 390**,** 314-325.

HAMMITT, L. L., DAGAN, R., YUAN, Y., BACA COTS, M., BOSHEVA, M., MADHI, S. A., MULLER, W. J., ZAR, H. J., BROOKS, D., GRENHAM, A., WAHLBY HAMREN, U., MANKAD, V. S., REN, P., TAKAS, T., ABRAM, M. E., LEACH, A., GRIFFIN, M. P., VILLAFANA, T. & GROUP, M. S. 2022. Nirsevimab for Prevention of RSV in Healthy Late-Preterm and Term Infants. *N Engl J Med,* 386**,** 837-846.

HAMRICK, S. E. G., SALLMON, H., ROSE, A. T., PORRAS, D., SHELTON, E. L., REESE, J. & HANSMANN, G. 2020. Patent Ductus Arteriosus of the Preterm Infant. *Pediatrics,* 146.

HANSMANN, G., SALLMON, H., ROEHR, C. C., KOUREMBANAS, S., AUSTIN, E. D., KOESTENBERGER, M. & EUROPEAN PEDIATRIC PULMONARY VASCULAR DISEASE, N. 2021. Pulmonary hypertension in bronchopulmonary dysplasia. *Pediatr Res,* 89**,** 446-455.

HART, K., COUSINS, M., WATKINS, W. J., KOTECHA, S. J., HENDERSON, A. J. & KOTECHA, S. 2022. Association of early-life factors with prematurity-associated lung disease: prospective cohort study. *Eur Respir J,* 59.

HERRMANN, H., BAR, H., KREPLAK, L., STRELKOV, S. V. & AEBI, U. 2007. Intermediate filaments: from cell architecture to nanomechanics. *Nat Rev Mol Cell Biol,* 8**,** 562-73.

HIGBEE, D. H., GRANELL, R., DAVEY SMITH, G. & DODD, J. W. 2022. Prevalence, risk factors, and clinical implications of preserved ratio impaired spirometry: a UK Biobank cohort analysis. *Lancet Respir Med,* 10**,** 149-157.

HIGGINS, R. D., JOBE, A. H., KOSO-THOMAS, M., BANCALARI, E., VISCARDI, R. M., HARTERT, T. V., RYAN, R. M., KALLAPUR, S. G., STEINHORN, R. H., KONDURI, G. G., DAVIS, S. D., THEBAUD, B., CLYMAN, R. I., COLLACO, J. M., MARTIN, C. R., WOODS, J. C., FINER, N. N. & RAJU, T. N. K. 2018. Bronchopulmonary Dysplasia: Executive Summary of a Workshop. *J Pediatr,* 197**,** 300-308.

HILLMAN, N. H., KALLAPUR, S. G., PILLOW, J. J., MOSS, T. J., POLGLASE, G. R., NITSOS, I. & JOBE, A. H. 2010. Airway injury from initiating ventilation in preterm sheep. *Pediatr Res,* 67**,** 60-5.

HISLOP, A. A. 2002. Airway and blood vessel interaction during lung development. *J Anat,* 201**,** 325-34.

HOLGUIN, F. 2013. Arginine and nitric oxide pathways in obesity-associated asthma. *J Allergy (Cairo),* 2013**,** 714595.

HOLTHOFER, B., WINDOFFER, R., TROYANOVSKY, S. & LEUBE, R. E. 2007. Structure and function of desmosomes. *Int Rev Cytol,* 264**,** 65-163.

HONG, S. Y., GIL, H. W., YANG, J. O., LEE, E. Y., KIM, H. K., KIM, S. H., CHUNG, Y. H., HWANG, S. K. & LEE, Z. W. 2005. Pharmacokinetics of glutathione and its metabolites in normal subjects. *J Korean Med Sci,* 20**,** 721-6.

HORVATH, I., BARNES, P. J., LOUKIDES, S., STERK, P. J., HOGMAN, M., OLIN, A. C., AMANN, A., ANTUS, B., BARALDI, E., BIKOV, A., BOOTS, A. W., BOS, L. D., BRINKMAN, P., BUCCA, C., CARPAGNANO, G. E., CORRADI, M., CRISTESCU, S., DE JONGSTE, J. C., DINH-XUAN, A. T., DOMPELING, E., FENS, N., FOWLER, S., HOHLFELD, J. M., HOLZ, O., JOBSIS, Q., VAN DE KANT, K., KNOBEL, H. H., KOSTIKAS, K., LEHTIMAKI, L., LUNDBERG, J., MONTUSCHI, P., VAN MUYLEM, A., PENNAZZA, G., REINHOLD, P., RICCIARDOLO, F. L. M., ROSIAS, P., SANTONICO, M., VAN DER SCHEE, M. P., VAN SCHOOTEN, F. J., SPANEVELLO, A., TONIA, T. & VINK, T. J. 2017. A European Respiratory Society technical standard: exhaled biomarkers in lung disease. *Eur Respir J,* 49.

HORVATH, I., HUNT, J., BARNES, P. J., ALVING, K., ANTCZAK, A., BARALDI, E., BECHER, G., VAN BEURDEN, W. J., CORRADI, M., DEKHUIJZEN, R., DWEIK, R. A., DWYER, T., EFFROS, R., ERZURUM, S., GASTON, B., GESSNER, C., GREENING, A., HO, L. P., HOHLFELD, J., JOBSIS, Q., LASKOWSKI, D., LOUKIDES, S., MARLIN, D., MONTUSCHI, P., OLIN, A. C., REDINGTON, A. E., REINHOLD, P., VAN RENSEN, E. L., RUBINSTEIN, I., SILKOFF, P., TOREN, K., VASS, G., VOGELBERG, C., WIRTZ, H. & CONDENSATE, A. E. T. F. O. E. B. 2005. Exhaled breath condensate: methodological recommendations and unresolved questions. *Eur Respir J,* 26**,** 523-48.

HOWSON, C. P., KINNEY, M. V. & LAWN, J. E. 2012. March of Dimes, PMNCH, Save the Children, WHO. Born Too Soon: The Global Action Report on Preterm Birth.

HUANG, J. A.-O., LI, W., SUN, Y., HUANG, Z., CONG, R., YU, C. & TAO, H. 2024. Preserved Ratio Impaired Spirometry (PRISm): A Global Epidemiological Overview, Radiographic Characteristics, Comorbid Associations, and Differentiation from Chronic Obstructive

Pulmonary Disease. *International journal of chronic obstructive pulmonary disease,* 19**,** 753–764.

HUANG, S.-J., DING, Z.-N., XIANG, H.-X., FU, L. & FEI, J. 2020. Association Between Serum S100A8/S100A9 Heterodimer and Pulmonary Function in Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease. *Lung,* 198**,** 645-652.

HUANG, W., CHO, K. Y., MENG, D. & WALKER, W. A. 2021. The impact of indole-3-lactic acid on immature intestinal innate immunity and development: a transcriptomic analysis. *Sci Rep,* 11**,** 8088.

HUNDSCHEID, T., ONLAND, W., KOOI, E. M. W., VIJLBRIEF, D. C., DE VRIES, W. B., DIJKMAN, K. P., VAN KAAM, A. H., VILLAMOR, E., KROON, A. A., VISSER, R., MULDER-DE TOLLENAER, S. M., DE BISSCHOP, B., DIJK, P. H., AVINO, D., HOCQ, C., ZECIC, A., MEEUS, M., DE BAAT, T., DERRIKS, F., HENRIKSEN, T. B., KYNG, K. J., DONDERS, R., NUYTEMANS, D., VAN OVERMEIRE, B., MULDER, A. L., DE BOODE, W. P. & BENEDUCTUS TRIAL, I. 2023. Expectant Management or Early Ibuprofen for Patent Ductus Arteriosus. *N Engl J Med,* 388**,** 980-990.

HUSSAIN, T., TAN, B., YIN, Y., BLACHIER, F., TOSSOU, M. C. & RAHU, N. 2016. Oxidative Stress and Inflammation: What Polyphenols Can Do for Us? *Oxid Med Cell Longev,* 2016**,** 7432797.

HYSINGER, E. B., FRIEDMAN, N. L., PADULA, M. A., SHINOHARA, R. T., ZHANG, H., PANITCH, H. B., KAWUT, S. M. & CHILDREN'S HOSPITALS NEONATAL, C. 2017. Tracheobronchomalacia Is Associated with Increased Morbidity in Bronchopulmonary Dysplasia. *Ann Am Thorac Soc,* 14**,** 1428-35.

ILES, R. & EDMUNDS, A. T. 1997. Assessment of pulmonary function in resolving chronic lung disease of prematurity. *Arch Dis Child Fetal Neonatal Ed,* 76**,** F113-7.

ISAYAMA, T., LEE, S. K., YANG, J., LEE, D., DASPAL, S., DUNN, M., SHAH, P. S., CANADIAN NEONATAL, N. & CANADIAN NEONATAL FOLLOW-UP NETWORK, I. 2017. Revisiting the Definition of Bronchopulmonary Dysplasia: Effect of Changing Panoply of Respiratory Support for Preterm Neonates. *JAMA Pediatr,* 171**,** 271-279.

IWATA, J. & NISHIKAZE, O. 1979. New micro-turbidimetric method for determination of protein in cerebrospinal fluid and urine. *Clin Chem,* 25**,** 1317-9.

JAIN, V. G., PARIKH, N. A., RYSAVY, M. A., SHUKLA, V. V., TROTTA, M., JOBE, A., CARLO, W. A. & AMBALAVANAN, N. 2024. Histological Chorioamnionitis Increases the Risk of Bronchopulmonary Dysplasia. *Am J Respir Crit Care Med,* 209**,** 1272-1275.

JAKEWAYS, N., MCKEEVER, T., LEWIS, S. A., WEISS, S. T. & BRITTON, J. 2003. Relationship between FEV1 reduction and respiratory symptoms in the general population. *Eur Respir J,* 21**,** 658-63.

JENSEN, E. A., EDWARDS, E. M., GREENBERG, L. T., SOLL, R. F., EHRET, D. E. Y. & HORBAR, J. D. 2021. Severity of Bronchopulmonary Dysplasia Among Very Preterm Infants in the United States. *Pediatrics,* 148.

JERKIC, S. P., MICHEL, F., DONATH, H., HERRMANN, E., SCHUBERT, R., ROSEWICH, M. & ZIELEN, S. 2020. Calprotectin as a New Sensitive Marker of Neutrophilic Inflammation in Patients with Bronchiolitis Obliterans. *Mediators Inflamm,* 2020**,** 4641585.

JEWISON, T., SU, Y., DISFANY, F. M., LIANG, Y., KNOX, C., MACIEJEWSKI, A., POELZER, J., HUYNH, J., ZHOU, Y., ARNDT, D., DJOUMBOU, Y., LIU, Y., DENG, L., GUO, A. C., HAN, B., PON, A., WILSON, M., RAFATNIA, S., LIU, P. & WISHART, D. S. 2014. SMPDB 2.0: big improvements to the Small Molecule Pathway Database. *Nucleic Acids Res,* 42**,** D478-84.

JOBE, A. H. & BANCALARI, E. 2001. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med,* 163**,** 1723-9.

JOSHI, S. & KOTECHA, S. 2007. Lung growth and development. *Early Hum Dev,* 83**,** 789-94.

JUANG, H. H., CHEN, S. M., LIN, G., CHIANG, M. H., HOU, C. P., LIN, Y. H., YANG, P. S., CHANG, P. L., CHEN, C. L., LIN, K. Y. & TSUI, K. H. 2021. The Clinical Experiences of Urine

Metabolomics of Genitourinary Urothelial Cancer in a Tertiary Hospital in Taiwan. *Front Oncol,* 11**,** 680910.

KETTLE, A. J., TURNER, R., GANGELL, C. L., HARWOOD, D. T., KHALILOVA, I. S., CHAPMAN, A. L., WINTERBOURN, C. C., SLY, P. D. & AREST, C. F. 2014. Oxidation contributes to low glutathione in the airways of children with cystic fibrosis. *Eur Respir J,* 44**,** 122-9.

KEYEL, P. A. 2017. Dnases in health and disease. *Dev Biol,* 429**,** 1-11.

KIM, D. J., OH, J. Y., RHEE, C. K., PARK, S. J., SHIM, J. J. & CHO, J. Y. 2021. Metabolic Fingerprinting Uncovers the Distinction Between the Phenotypes of Tuberculosis Associated COPD and Smoking-Induced COPD. *Front Med (Lausanne),* 8**,** 619077.

KISTNER, A., CELSI, G., VANPEE, M. & JACOBSON, S. H. 2000. Increased blood pressure but normal renal function in adult women born preterm. *Pediatr Nephrol,* 15**,** 215-20.

KONG, M. Y., GAGGAR, A., LI, Y., WINKLER, M., BLALOCK, J. E. & CLANCY, J. P. 2009. Matrix metalloproteinase activity in pediatric acute lung injury. *Int J Med Sci,* 6**,** 9-17.

KONRAD MORGENROTH, M. E. 2008. Chapter 8 - Anatomy. *In:* PETER J. PAPADAKOS, B. L., LARAINE VISSER-ISLES (ed.) *Mechanical Ventilation Clinical Applications and Pathophysiology.* W.B. Saunders.

KOTECHA, S. 2000. Lung growth: implications for the newborn infant. *Arch Dis Child Fetal Neonatal Ed,* 82**,** F69-74.

KOTECHA, S. J., COURSE, C. W., JONES, K. E., WATKINS, W. J., BERRINGTON, J., GILLESPIE, D. & KOTECHA, S. 2022a. Follow-up study of infants recruited to the randomised, placebo-controlled trial of azithromycin for the prevention of chronic lung disease of prematurity in preterm infants-study protocol for the AZTEC-FU study. *Trials,* 23.

KOTECHA, S. J., GIBBONS, J. T. D., COURSE, C. W., EVANS, E. E., SIMPSON, S. J., WATKINS, W. J. & KOTECHA, S. 2022b. Geographical Differences and Temporal Improvements in Forced Expiratory Volume in 1 Second of Preterm-Born Children: A Systematic Review and Meta-analysis. *JAMA Pediatr*.

KOTECHA, S. J., WATKINS, W. J., PARANJOTHY, S., DUNSTAN, F. D., HENDERSON, A. J. & KOTECHA, S. 2012. Effect of late preterm birth on longitudinal lung spirometry in school age children and adolescents. *Thorax,* 67**,** 54-61.

KOTLYAROV, S. & KOTLYAROVA, A. 2021. Anti-Inflammatory Function of Fatty Acids and Involvement of Their Metabolites in the Resolution of Inflammation in Chronic Obstructive Pulmonary Disease. *Int J Mol Sci,* 22.

KOTSIOU, O. S., PAPAGIANNIS, D., PAPADOPOULOU, R. & GOURGOULIANIS, K. I. 2021. Calprotectin in Lung Diseases. *Int J Mol Sci,* 22.

KUMAR, A. & BACHHAWAT, A. K. 2012. Pyroglutamic acid: throwing light on a lightly studied metabolite. *Current Science,* 102**,** 288-297.

KUROVA, V. S., ANAEV, E. C., KONONIKHIN, A. S., FEDORCHENKO, K. Y., POPOV, I. A., KALUPOV, T. L., BRATANOV, D. O., NIKOLAEV, E. N. & VARFOLOMEEV, S. D. 2009. Proteomics of exhaled breath: methodological nuances and pitfalls. *Clin Chem Lab Med,* 47**,** 706-12.

LAHN-JOHANNESSEN LILLEBØE, H., SALVESON ENGESET, M., CLEMM, H. H., HALVORSEN, T., DRANGE RØKSUND, O., POTREBNY, T. & VOLLSÆTER, M. 2024. Expiratory airflow limitation in adults born extremely preterm: a systematic review and meta-analysis. *Paediatric Respiratory Reviews*.

LAHRA, M. M., BEEBY, P. J. & JEFFERY, H. E. 2009. Maternal versus fetal inflammation and respiratory distress syndrome: a 10-year hospital cohort study. *Arch Dis Child Fetal Neonatal Ed,* 94**,** F13-6.

LAI, Z., TSUGAWA, H., WOHLGEMUTH, G., MEHTA, S., MUELLER, M., ZHENG, Y., OGIWARA, A., MEISSEN, J., SHOWALTER, M., TAKEUCHI, K., KIND, T., BEAL, P., ARITA, M. & FIEHN, O. 2018. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat Methods,* 15**,** 53-56.

LAI, Z. J. & FIEHN, O. 2018. Mass spectral fragmentation of trimethylsilylated small molecules. *Mass Spectrometry Reviews,* 37**,** 245-257.

LAL, C. V., KANDASAMY, J., DOLMA, K., RAMANI, M., KUMAR, R., WILSON, L., AGHAI, Z., BARNES, S., BLALOCK, J. E., GAGGAR, A., BHANDARI, V. & AMBALAVANAN, N. 2018. Early airway microbial metagenomic and metabolomic signatures are associated with development of severe bronchopulmonary dysplasia. *Am J Physiol Lung Cell Mol Physiol,* 315**,** L810-L815.

LEE, S. I. & KANG, K. S. 2017. Function of capric acid in cyclophosphamide-induced intestinal inflammation, oxidative stress, and barrier function in pigs. *Sci Rep,* 7**,** 16530.

LI, D., SHEN, L., ZHANG, D., WANG, X., WANG, Q., QIN, W., GAO, Y. & LI, X. 2023. Ammonia-induced oxidative stress triggered proinflammatory response and apoptosis in pig lungs. *J Environ Sci (China),* 126**,** 683-696.

LI, T., TYNKKYNEN, T., IHANUS, A., ZHAO, S., MAKINEN, V. P. & ALA-KORPELA, M. 2022. Characteristics of Normalization Methods in Quantitative Urinary Metabolomics-Implications for Epidemiological Applications and Interpretations. *Biomolecules,* 12.

LI, Y., WANG, W. & ZHANG, D. 2019. Maternal diabetes mellitus and risk of neonatal respiratory distress syndrome: a meta-analysis. *Acta Diabetol,* 56**,** 729-740.

LIAO, Y., WANG, J., JAEHNIG, E. J., SHI, Z. & ZHANG, B. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res,* 47**,** W199-W205.

LIU, L., OZA, S., HOGAN, D., CHU, Y., PERIN, J., ZHU, J., LAWN, J. E., COUSENS, S., MATHERS, C. & BLACK, R. E. 2016. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet,* 388**,** 3027-3035.

LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F., YOUNG, N., FOSTER, B., MOSER, M., KARASIK, E., GILLARD, B., RAMSEY, K., SULLIVAN, S., BRIDGE, J., MAGAZINE, H., SYRON, J., FLEMING, J., SIMINOFF, L., TRAINO, H., MOSAVEL, M., BARKER, L., JEWELL, S., ROHRER, D., MAXIM, D., FILKINS, D., HARBACH, P., CORTADILLO, E., BERGHUIS, B., TURNER, L., HUDSON, E., FEENSTRA, K., SOBIN, L., ROBB, J., BRANTON, P., KORZENIEWSKI, G., SHIVE, C., TABOR, D., QI, L., GROCH, K., NAMPALLY, S., BUIA, S., ZIMMERMAN, A., SMITH, A., BURGES, R., ROBINSON, K., VALENTINO, K., BRADBURY, D., COSENTINO, M., DIAZ-MAYORAL, N., KENNEDY, M., ENGEL, T., WILLIAMS, P., ERICKSON, K., ARDLIE, K., WINCKLER, W., GETZ, G., DELUCA, D., MACARTHUR, D., KELLIS, M., THOMSON, A., YOUNG, T., GELFAND, E., DONOVAN, M., MENG, Y., GRANT, G., MASH, D., MARCUS, Y., BASILE, M., LIU, J., ZHU, J., TU, Z., COX, N. J., NICOLAE, D. L., GAMAZON, E. R., IM, H. K., KONKASHBAEV, A., PRITCHARD, J., STEVENS, M., FLUTRE, T., WEN, X., DERMITZAKIS, E. T., LAPPALAINEN, T., GUIGO, R., MONLONG, J., SAMMETH, M., KOLLER, D., BATTLE, A., MOSTAFAVI, S., MCCARTHY, M., RIVAS, M., MALLER, J., RUSYN, I., NOBEL, A., WRIGHT, F., SHABALIN, A., FEOLO, M., SHAROPOVA, N., et al. 2013. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics,* 45**,** 580-585.

LOPEZ-SANCHEZ, L. M., JURADO-GAMEZ, B., FEU-COLLADO, N., VALVERDE, A., CANAS, A., FERNANDEZ-RUEDA, J. L., ARANDA, E. & RODRIGUEZ-ARIZA, A. 2017. Exhaled breath condensate biomarkers for the early diagnosis of lung cancer using proteomics. *Am J Physiol Lung Cell Mol Physiol,* 313**,** L664-L676.

LORENZ, E., MUHLEBACH, M. S., TESSIER, P. A., ALEXIS, N. E., DUNCAN HITE, R., SEEDS, M. C., PEDEN, D. B. & MEREDITH, W. 2008. Different expression ratio of S100A8/A9 and S100A12 in acute and chronic lung diseases. *Respir Med,* 102**,** 567-73.

LOWE, J., GILLESPIE, D., HUBBARD, M., ZHANG, L., KIRBY, N., PICKLES, T., THOMAS-JONES, E., TURNER, M. A., KLEIN, N., MARCHESI, J. R., HOOD, K., BERRINGTON, J. & KOTECHA, S. 2020. Study protocol: azithromycin therapy for chronic lung disease of prematurity (AZTEC) - a randomised, placebo-controlled trial of azithromycin for the prevention of chronic lung disease of prematurity in preterm infants. *BMJ Open,* 10**,** e041528.

LOWE, J., WATKINS, W. J., EDWARDS, M. O., SPILLER, O. B., JACQZ-AIGRAIN, E., KOTECHA, S. J. & KOTECHA, S. 2014. Association between pulmonary ureaplasma colonization and

bronchopulmonary dysplasia in preterm infants: updated systematic review and meta-analysis. *Pediatr Infect Dis J,* 33**,** 697-702.

LU, J., GE, H., QI, L., ZHANG, S., YANG, Y., HUANG, X. & LI, M. 2022. Subtyping preserved ratio impaired spirometry (PRISm) by using quantitative HRCT imaging characteristics. *Respir Res,* 23**,** 309.

LUALDI, M. & FASANO, M. 2019. Statistical analysis of proteomics data: A review on feature selection. *J Proteomics,* 198**,** 18-26.

LUNDBERG, B., MERID, S. K., UM-BERGSTROM, P., WANG, G., BERGSTROM, A., EKSTROM, S., KULL, I., MELEN, E. & HALLBERG, J. 2024. Lung function in young adulthood in relation to moderate-to-late preterm birth. *ERJ Open Res,* 10.

MAGAGNOTTI, C., MATASSA, P. G., BACHI, A., VENDETTUOLI, V., FERMO, I., COLNAGHI, M. R., CARLETTI, R. M., MERCADANTE, D., FATTORE, E., MOSCA, F. & ANDOLFO, A. 2013. Calcium signaling-related proteins are associated with broncho-pulmonary dysplasia progression. *J Proteomics,* 94**,** 401-12.

MAHADEVA, R., SHARPLES, L., ROSS-RUSSELL, R. I., WEBB, A. K., BILTON, D. & LOMAS, D. A. 2001. Association of alpha(1)-antichymotrypsin deficiency with milder lung disease in patients with cystic fibrosis. *Thorax,* 56**,** 53-8.

MAHUT, B., DE BLIC, J., EMOND, S., BENOIST, M. R., JARREAU, P. H., LACAZE-MASMONTEIL, T., MAGNY, J. F. & DELACOURT, C. 2007. Chest computed tomography findings in bronchopulmonary dysplasia and correlation with lung function. *Arch Dis Child Fetal Neonatal Ed,* 92**,** F459-64.

MANISCALCO, M., FUSCHILLO, S., PARIS, D., CUTIGNANO, A., SANDUZZI, A. & MOTTA, A. 2019. Clinical metabolomics of exhaled breath condensate in chronic respiratory diseases. *Advances in Clinical Chemistry, Vol 88,* 88**,** 121-149.

MANSELL, T., VLAHOS, A., COLLIER, F., PONSONBY, A. L., VUILLERMIN, P., ELLUL, S., TANG, M. L. K., BURGNER, D., SAFFERY, R. & BARWON INFANT STUDY INVESTIGATOR, T. 2022. The newborn metabolome: associations with gestational diabetes, sex, gestation, birth mode, and birth weight. *Pediatr Res,* 91**,** 1864-1873.

MAROTT, J. L., INGEBRIGTSEN, T. S., COLAK, Y., VESTBO, J. & LANGE, P. 2021. Trajectory of Preserved Ratio Impaired Spirometry: Natural History and Long-Term Prognosis. *Am J Respir Crit Care Med,* 204**,** 910-920.

MARTELO-VIDAL, L., VÁZQUEZ-MERA, S., MIGUÉNS-SUÁREZ, P., SALGADO-CASTRO, F. J., BLANCO-APARICIO, M., MOSTEIRO-AÑÓN, M., MARTÍNEZ-CAPOCCIONI, G., MÉNDEZ-BREA, P., BRAVO, S. B., GONZÁLEZ-BARCALA, F. & NIETO-FONTARIGO, J. J. 2022. Application of urinary proteomics for biomarker discovery in respiratory diseases. *European Respiratory Journal,* 60**,** 1672.

MATTOLI, L., GIANNI, M. & BURICO, M. 2023. Mass spectrometry-based metabolomic analysis as a tool for quality control of natural complex products. *Mass Spectrom Rev,* 42**,** 1358-1396.

MCGOLDRICK, E., STEWART, F., PARKER, R. & DALZIEL, S. R. 2020. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev,* 12**,** CD004454.

MEDINA, T. M. & HILL, D. A. 2006. Preterm premature rupture of membranes: diagnosis and management. *Am Fam Physician,* 73**,** 659-64.

MEISTER, A. & ANDERSON, M. E. 1983. Glutathione. *Annu Rev Biochem,* 52**,** 711-60.

MENG, D., SOMMELLA, E., SALVIATI, E., CAMPIGLIA, P., GANGULI, K., DJEBALI, K., ZHU, W. & WALKER, W. A. 2020. Indole-3-lactic acid, a metabolite of tryptophan, secreted by Bifidobacterium longum subspecies infantis is anti-inflammatory in the immature intestine. *Pediatr Res,* 88**,** 209-217.

MENZIES-GOW, A., MANSUR, A. H. & BRIGHTLING, C. E. 2020. Clinical utility of fractional exhaled nitric oxide in severe asthma management. *Eur Respir J,* 55.

MERKEL, D., RIST, W., SEITHER, P., WEITH, A. & LENTER, M. C. 2005. Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. *PROTEOMICS,* 5**,** 2972-2980.

MILLER, M. R., HANKINSON, J., BRUSASCO, V., BURGOS, F., CASABURI, R., COATES, A., CRAPO, R., ENRIGHT, P., VAN DER GRINTEN, C. P., GUSTAFSSON, P., JENSEN, R., JOHNSON, D. C., MACINTYRE, N., MCKAY, R., NAVAJAS, D., PEDERSEN, O. F., PELLEGRINO, R., VIEGI, G., WANGER, J. & FORCE, A. E. T. 2005. Standardisation of spirometry. *Eur Respir J,* 26**,** 319-38.

MITRA, S., GARDNER, C. E., MACLELLAN, A., DISHER, T., STYRANKO, D. M., CAMPBELL-YEO, M., KUHLE, S., JOHNSTON, B. C. & DORLING, J. 2022. Prophylactic cyclo-oxygenase inhibitor drugs for the prevention of morbidity and mortality in preterm infants: a network meta-analysis. *Cochrane Database Syst Rev,* 4**,** CD013846.

MITRA, S., SCRIVENS, A., VON KURSELL, A. M. & DISHER, T. 2020. Early treatment versus expectant management of hemodynamically significant patent ductus arteriosus for preterm infants. *Cochrane Database Syst Rev,* 12**,** CD013278.

MOITRA, S., BANDYOPADHYAY, A. & LACY, P. 2023. Metabolomics of Respiratory Diseases. *In:* GHINI, V., STRINGER, K. A. & LUCHINAT, C. (eds.) *Metabolomics and Its Impact on Health and Diseases. Handbook of Experimental Pharmacology.*

MOKRES, L. M., PARAI, K., HILGENDORFF, A., ERTSEY, R., ALVIRA, C. M., RABINOVITCH, M. & BLAND, R. D. 2010. Prolonged mechanical ventilation with air induces apoptosis and causes failure of alveolar septation and angiogenesis in lungs of newborn mice. *Am J Physiol Lung Cell Mol Physiol,* 298**,** L23-35.

MONTI, C., ZILOCCHI, M., COLUGNAT, I. & ALBERIO, T. 2019. Proteomics turns functional. *J Proteomics,* 198**,** 36-44.

MOYER-MILEUR, L. J., NIELSON, D. W., PFEFFER, K. D., WITTE, M. K. & CHAPMAN, D. L. 1996. Eliminating sleep-associated hypoxemia improves growth in infants with bronchopulmonary dysplasia. *Pediatrics,* 98**,** 779-83.

NEYRINCK, A. M., RODRIGUEZ, J., ZHANG, Z., NAZARE, J. A., BINDELS, L. B., CANI, P. D., MAQUET, V., LAVILLE, M., BISCHOFF, S. C., WALTER, J. & DELZENNE, N. M. 2022. Breath volatile metabolome reveals the impact of dietary fibres on the gut microbiota: Proof of concept in healthy volunteers. *EBioMedicine,* 80**,** 104051.

NGUYEN, M., SHARMA, A., WU, W., GOMI, R., SUNG, B., HOSPODSKY, D., ANGENENT, L. T. & WORGALL, S. 2016. The fermentation product 2,3-butanediol alters P. aeruginosa clearance, cytokine response and the lung microbiome. *ISME J,* 10**,** 2978-2983.

NILSSON, T., MANN, M., AEBERSOLD, R., YATES, J. R., 3RD, BAIROCH, A. & BERGERON, J. J. 2010. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods,* 7**,** 681-5.

NIZIOL, J., OSSOLINSKI, K., PLAZA-ALTAMER, A., KOLODZIEJ, A., OSSOLINSKA, A., OSSOLINSKI, T., NIECZAJ, A. & RUMAN, T. 2023. Untargeted urinary metabolomics for bladder cancer biomarker screening with ultrahigh-resolution mass spectrometry. *Sci Rep,* 13**,** 9802.

NORMAN, K. C., MOORE, B. B., ARNOLD, K. B. & O'DWYER, D. N. 2018. Proteomics: Clinical and research applications in respiratory diseases. *Respirology,* 23**,** 993-1003.

NORTHWAY, W. H., JR., ROSAN, R. C. & PORTER, D. Y. 1967. Pulmonary disease following respirator therapy of hyaline-membrane disease. Bronchopulmonary dysplasia. *N Engl J Med,* 276**,** 357-68.

OHUMA, E. O., MOLLER, A. B., BRADLEY, E., CHAKWERA, S., HUSSAIN-ALKHATEEB, L., LEWIN, A., OKWARAJI, Y. B., MAHANANI, W. R., JOHANSSON, E. W., LAVIN, T., FERNANDEZ, D. E., DOMINGUEZ, G. G., DE COSTA, A., CRESSWELL, J. A., KRASEVEC, J., LAWN, J. E., BLENCOWE, H., REQUEJO, J. & MORAN, A. C. 2023. National, regional, and global

estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. *Lancet,* 402**,** 1261-1271.

ORSBURN, B. C. 2021. Proteome Discoverer-A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes,* 9.

PAANANEN, R., HUSA, A. K., VUOLTEENAHO, R., HERVA, R., KAUKOLA, T. & HALLMAN, M. 2009. Blood cytokines during the perinatal period in very preterm infants: relationship of inflammatory response and bronchopulmonary dysplasia. *J Pediatr,* 154**,** 39-43 e3.

PACHT, E. R., TIMERMAN, A. P., LYKENS, M. G. & MEROLA, A. J. 1991. Deficiency of alveolar fluid glutathione in patients with sepsis and the adult respiratory distress syndrome. *Chest,* 100**,** 1397-403.

PANAHABADI, S., HEINDEL, K., MUELLER, A., HOLDENRIEDER, S. & KIPFMUELLER, F. 2021. Increased circulating cytokeratin 19 fragment levels in preterm neonates receiving mechanical ventilation are associated with poor outcome. *Am J Physiol Lung Cell Mol Physiol,* 321**,** L1036-L1043.

PANG, Z., ZHOU, G., EWALD, J., CHANG, L., HACARIZ, O., BASU, N. & XIA, J. 2022. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc,* 17**,** 1735-1761.

PAPAGIANIS, P. C., PILLOW, J. J. & MOSS, T. J. 2019. Bronchopulmonary dysplasia: Pathophysiology and potential anti-inflammatory therapies. *Paediatr Respir Rev,* 30**,** 34-41.

PAPPIREDDI, N., MARTIN, L. & WUHR, M. 2019. A Review on Quantitative Multiplexed Proteomics. *Chembiochem,* 20**,** 1210-1224.

PARK, Y. H., FITZPATRICK, A. M., MEDRIANO, C. A. & JONES, D. P. 2017. High-resolution metabolomics to identify urine biomarkers in corticosteroid-resistant asthmatic children. *J Allergy Clin Immunol,* 139**,** 1518-1524 e4.

PATTI, G. J., YANES, O. & SIUZDAK, G. 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol,* 13**,** 263-9.

PELKONEN, A. S., HAKULINEN, A. L., HALLMAN, M. & TURPEINEN, M. 2001. Effect of inhaled budesonide therapy on lung function in schoolchildren born preterm. *Respir Med,* 95**,** 565-70.

PEREIRA-FANTINI, P. M. & TINGAY, D. G. 2016. The proteomics of lung injury in childhood: challenges and opportunities. *Clin Proteomics,* 13**,** 5.

PERNEMALM, M., DE PETRIS, L., BRANCA, R. M., FORSHED, J., KANTER, L., SORIA, J. C., GIRARD, P., VALIDIRE, P., PAWITAN, Y., VAN DEN OORD, J., LAZAR, V., PAHLMAN, S., LEWENSOHN, R. & LEHTIO, J. 2013. Quantitative proteomics profiling of primary lung adenocarcinoma tumors reveals functional perturbations in tumor metabolism. *J Proteome Res,* 12**,** 3934-43.

PIERSIGILLI, F., LAM, T. T., VERNOCCHI, P., QUAGLIARIELLO, A., PUTIGNANI, L., AGHAI, Z. H. & BHANDARI, V. 2019. Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. *Metabolomics,* 15.

PINTUS, M. C., LUSSU, M., DESSI, A., PINTUS, R., NOTO, A., MASILE, V., MARCIALIS, M. A., PUDDU, M., FANOS, V. & ATZORI, L. 2018. Urinary (1)H-NMR Metabolomics in the First Week of Life Can Anticipate BPD Diagnosis. *Oxid Med Cell Longev,* 2018**,** 7620671.

PULAKKA, A., RISNES, K., METSALA, J., ALENIUS, S., HEIKKILA, K., NILSEN, S. M., NASANEN-GILMORE, P., HAARAMO, P., GISSLER, M., OPDAHL, S. & KAJANTIE, E. 2023. Preterm birth and asthma and COPD in adulthood: a nationwide register study from two Nordic countries. *Eur Respir J,* 61.

QUAN-JUN, Y., JIAN-PING, Z., JIAN-HUA, Z., YONG-LONG, H., BO, X., JING-XIAN, Z., BONA, D., YUAN, Z. & CHENG, G. 2017. Distinct Metabolic Profile of Inhaled Budesonide and

Salbutamol in Asthmatic Children during Acute Exacerbation. *Basic Clin Pharmacol Toxicol,* 120**,** 303-311.

QUANJER, P. H., STANOJEVIC, S., COLE, T. J., BAUR, X., HALL, G. L., CULVER, B. H., ENRIGHT, P. L., HANKINSON, J. L., IP, M. S., ZHENG, J., STOCKS, J. & INITIATIVE, E. R. S. G. L. F. 2012. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J,* 40**,** 1324-43.

QUINN, L. A., HIRANI, S. H., WILLIAMS, T. C. & SINHA, I. P. 2021. Palivizumab immunoprophylaxis for infants with BPD has medium- and long-term benefits: myth or maxim? *Breathe (Sheff),* 17**,** 210110.

R CORE TEAM 2021. R: A language and environment for statistical computing. Version 4.0.4 ed.: R Foundation for Statistical Computing, Vienna, Austria.

RAFFAY, T. M., LOCY, M. L., HILL, C. L., JINDAL, N. S., ROGERS, L. K., WELTY, S. E. & TIPPLE, T. E. 2013. Neonatal hyperoxic exposure persistently alters lung secretoglobins and annexin A1. *Biomed Res Int,* 2013**,** 408485.

RAJU, T. N. K., PEMBERTON, V. L., SAIGAL, S., BLAISDELL, C. J., MOXEY-MIMS, M., BUIST, S., ADULTS BORN PRETERM CONFERENCE, S. & DISCUSSANTS 2017. Long-Term Healthcare Outcomes of Preterm Birth: An Executive Summary of a Conference Sponsored by the National Institutes of Health. *J Pediatr,* 181**,** 309-318 e1.

RAKOW, A., JOHANSSON, S., LEGNEVALL, L., SEVASTIK, R., CELSI, G., NORMAN, M. & VANPEE, M. 2008. Renal volume and function in school-age children born preterm or small for gestational age. *Pediatr Nephrol,* 23**,** 1309-15.

RAY, A. & KOLLS, J. K. 2017. Neutrophilic Inflammation in Asthma and Association with Disease Severity. *Trends Immunol,* 38**,** 942-954.

RESHETOVA, P., SMILDE, A. K., VAN KAMPEN, A. H. & WESTERHUIS, J. A. 2014. Use of prior knowledge for the analysis of high-throughput transcriptomics and metabolomics data. *BMC Syst Biol,* 8 Suppl 2**,** S2.

REUTER, S., MOSER, C. & BAACK, M. 2014. Respiratory distress in the newborn. *Pediatr Rev,* 35**,** 417-28; quiz 429.

RICE, S. J., LIU, X., MILLER, B., JOSHI, M., ZHU, J., CARUSO, C., GILBERT, C., TOTH, J., REED, M., RASSAEI, N., DAS, A., BAROCHIA, A., EL-BAYOUMY, K. & BELANI, C. P. 2015. Proteomic profiling of human plasma identifies apolipoprotein E as being associated with smoking and a marker for squamous metaplasia of the lung. *Proteomics,* 15**,** 3267-77.

ROBBINS, M. E., CHO, H. Y., HANSEN, J. M., LUCHSINGER, J. R., LOCY, M. L., VELTEN, M., KLEEBERGER, S. R., ROGERS, L. K. & TIPPLE, T. E. 2021. Glutathione reductase deficiency alters lung development and hyperoxic responses in neonatal mice. *Redox Biol,* 38**,** 101797.

ROMERO, R., ESPINOZA, J., KUSANOVIC, J. P., GOTSCH, F., HASSAN, S., EREZ, O., CHAIWORAPONGSA, T. & MAZOR, M. 2006. The preterm parturition syndrome. *BJOG,* 113 Suppl 3**,** 17-42.

ROMERO, R., QUINTERO, R., OYARZUN, E., WU, Y. K., SABO, V., MAZOR, M. & HOBBINS, J. C. 1988. Intraamniotic infection and the onset of labor in preterm premature rupture of the membranes. *Am J Obstet Gynecol,* 159**,** 661-6.

RYAN, R. M. 2006. A new look at bronchopulmonary dysplasia classification. *J Perinatol,* 26**,** 207-9.

SAIGAL, S. & DOYLE, L. W. 2008. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet,* 371**,** 261-9.

SALAM, M. T., ISLAM, T., GAUDERMAN, W. J. & GILLILAND, F. D. 2009. Roles of arginase variants, atopy, and ozone in childhood asthma. *J Allergy Clin Immunol,* 123**,** 596-602, 602 e1-8.

SALES, D. S., ITO, J. T., ZANCHETTA, I. A., ANNONI, R., AUN, M. V., FERRAZ, L. F. S., CERVILHA, D. A. B., NEGRI, E., MAUAD, T., MARTINS, M. A. & LOPES, F. 2017. Regulatory T-Cell Distribution within Lung Compartments in COPD. *COPD,* 14**,** 533-542.

SANDFORD, A. J., CHAGANI, T., WEIR, T. D. & PARE, P. D. 1998. Alpha 1-antichymotrypsin mutations in patients with chronic obstructive pulmonary disease. *Dis Markers,* 13**,** 257-60.

SANTHAKUMARAN, S., STATNIKOV, Y., GRAY, D., BATTERSBY, C., ASHBY, D., MODI, N. & MEDICINES FOR NEONATES INVESTIGATOR, G. 2018. Survival of very preterm infants admitted to neonatal care in England 2008-2014: time trends and regional variation. *Arch Dis Child Fetal Neonatal Ed,* 103**,** F208-F215.

SCHRADER, M. & FAHIMI, H. D. 2006. Peroxisomes and oxidative stress. *Biochim Biophys Acta,* 1763**,** 1755-66.

SCHRADER, M., KAMOSHITA, M. & ISLINGER, M. 2020. Organelle interplay-peroxisome interactions in health and disease. *J Inherit Metab Dis,* 43**,** 71-89.

SEREN, S., DERIAN, L., KELES, I., GUILLON, A., LESNER, A., GONZALEZ, L., BARANEK, T., SI-TAHAR, M., MARCHAND-ADAM, S., JENNE, D. E., PAGET, C., JOUAN, Y. & KORKMAZ, B. 2021. Proteinase release from activated neutrophils in mechanically ventilated patients with non-COVID-19 and COVID-19 pneumonia. *Eur Respir J,* 57.

SHAHANA, S., BJORNSSON, E., LUDVIKSDOTTIR, D., JANSON, C., NETTELBLADT, O., VENGE, P., ROOMANS, G. M. & GROUP, B. H. R. 2005. Ultrastructure of bronchial biopsies from patients with allergic and non-allergic asthma. *Respir Med,* 99**,** 429-43.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res,* 13**,** 2498-504.

SHAW, G. M. & O'BRODOVICH, H. M. 2013. Progress in understanding the genetics of bronchopulmonary dysplasia. *Semin Perinatol,* 37**,** 85-93.

SHENNAN, A. T., DUNN, M. S., OHLSSON, A., LENNOX, K. & HOSKINS, E. M. 1988. Abnormal pulmonary outcomes in premature infants: prediction from oxygen requirement in the neonatal period. *Pediatrics,* 82**,** 527-32.

SHEPHERD, E. G., CLOUSE, B. J., HASENSTAB, K. A., SITARAM, S., MALLESKE, D. T., NELIN, L. D. & JADCHERLA, S. R. 2018. Infant Pulmonary Function Testing and Phenotypes in Severe Bronchopulmonary Dysplasia. *Pediatrics,* 141.

SIMPSON, S. J., DU BERRY, C., EVANS, D. J., GIBBONS, J. T. D., VOLLSAETER, M., HALVORSEN, T., GRUBER, K., LOMBARDI, E., STANOJEVIC, S., HURST, J. R., UM-BERGSTROM, P., HALLBERG, J., DOYLE, L. W. & KOTECHA, S. 2023. Unravelling the respiratory health path across the lifespan for survivors of preterm birth. *Lancet Respiratory Medicine*.

SIMPSON, S. J., LOGIE, K. M., O'DEA, C. A., BANTON, G. L., MURRAY, C., WILSON, A. C., PILLOW, J. J. & HALL, G. L. 2017. Altered lung structure and function in mid-childhood survivors of very preterm birth. *Thorax,* 72**,** 702-711.

SIMPSON, S. J., TURKOVIC, L., WILSON, A. C., VERHEGGEN, M., LOGIE, K. M., PILLOW, J. J. & HALL, G. L. 2018. Lung function trajectories throughout childhood in survivors of very preterm birth: a longitudinal cohort study. *Lancet Child Adolesc Health,* 2**,** 350-359.

SMITH, C. A., WANT, E. J., O'MAILLE, G., ABAGYAN, R. & SIUZDAK, G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem,* 78**,** 779-87.

STARODUBTSEVA, N. L., KONONIKHIN, A. S., BUGROVA, A. E., CHAGOVETS, V., INDEYKINA, M., KROKHINA, K. N., NIKITINA, I. V., KOSTYUKEVICH, Y. I., POPOV, I. A., LARINA, I. M., TIMOFEEVA, L. A., FRANKEVICH, V. E., IONOV, O. V., DEGTYAREV, D. N., NIKOLAEV, E. N. & SUKHIKH, G. T. 2016. Investigation of urine proteome of preterm newborns with respiratory pathologies. *J Proteomics,* 149**,** 31-37.

STATSWALES. 2019. *Welsh Index of Multiple Deprivation* [Online]. Welsh Government. Available: https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation [Accessed 2022].

STOCKS, J., HISLOP, A. & SONNAPPA, S. 2013. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med,* 1**,** 728-42.

STOLL, B. J., HANSEN, N. I., BELL, E. F., SHANKARAN, S., LAPTOOK, A. R., WALSH, M. C., HALE, E. C., NEWMAN, N. S., SCHIBLER, K., CARLO, W. A., KENNEDY, K. A., POINDEXTER, B. B., FINER, N. N., EHRENKRANZ, R. A., DUARA, S., SANCHEZ, P. J., O'SHEA, T. M., GOLDBERG, R. N., VAN MEURS, K. P., FAIX, R. G., PHELPS, D. L., FRANTZ, I. D., 3RD, WATTERBERG, K. L., SAHA, S., DAS, A., HIGGINS, R. D., EUNICE KENNEDY SHRIVER NATIONAL INSTITUTE OF CHILD, H. & HUMAN DEVELOPMENT NEONATAL RESEARCH, N. 2010. Neonatal outcomes of extremely preterm infants from the NICHD Neonatal Research Network. *Pediatrics,* 126**,** 443-56.

STOLL, B. J., HANSEN, N. I., BELL, E. F., WALSH, M. C., CARLO, W. A., SHANKARAN, S., LAPTOOK, A. R., SANCHEZ, P. J., VAN MEURS, K. P., WYCKOFF, M., DAS, A., HALE, E. C., BALL, M. B., NEWMAN, N. S., SCHIBLER, K., POINDEXTER, B. B., KENNEDY, K. A., COTTEN, C. M., WATTERBERG, K. L., D'ANGIO, C. T., DEMAURO, S. B., TRUOG, W. E., DEVASKAR, U., HIGGINS, R. D., EUNICE KENNEDY SHRIVER NATIONAL INSTITUTE OF CHILD, H. & HUMAN DEVELOPMENT NEONATAL RESEARCH, N. 2015. Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012. *JAMA,* 314**,** 1039-51.

STONE, T. W., STOY, N. & DARLINGTON, L. G. 2013. An expanding range of targets for kynurenine metabolites of tryptophan. *Trends in Pharmacological Sciences,* 34**,** 136-143.

STRATAKIS, N., SISKOS, A. P., PAPADOPOULOU, E., NGUYEN, A. N., ZHAO, Y., MARGETAKI, K., LAU, C. E., COEN, M., MAITRE, L., FERNANDEZ-BARRES, S., AGIER, L., ANDRUSAITYTE, S., BASAGANA, X., BRANTSAETER, A. L., CASAS, M., FOSSATI, S., GRAZULEVICIENE, R., HEUDE, B., MCEACHAN, R. R., MELTZER, H. M., MILLETT, C., RAUBER, F., ROBINSON, O., ROUMELIOTAKI, T., BORRAS, E., SABIDO, E., URQUIZA, J., VAFEIADI, M., VINEIS, P., VOORTMAN, T., WRIGHT, J., CONTI, D. V., VRIJHEID, M., KEUN, H. C. & CHATZI, L. 2022. Urinary metabolic biomarkers of diet quality in European children are associated with metabolic health. *Elife,* 11.

SUGIMOTO, M., KAWAKAMI, M., ROBERT, M., SOGA, T. & TOMITA, M. 2012. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. *Curr Bioinform,* 7**,** 96-108.

SUN, Y. V. & HU, Y. J. 2016. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet,* 93**,** 147-90.

SURESH, G. K., DAVIS, J. M. & SOLL, R. F. 2001. Superoxide dismutase for preventing chronic lung disease in mechanically ventilated preterm infants. *Cochrane Database Syst Rev,* 2001**,** CD001968.

SVEGER, T., OHLSSON, K., POLBERGER, S., NOACK, G., MORSE, H. & LAURIN, S. 2002. Tracheobronchial aspirate fluid neutrophil lipocalin, elastase- and neutrophil protease-4-alpha1-antitrypsin complexes, protease inhibitors and free proteolytic activity in respiratory distress syndrome. *Acta Paediatr,* 91**,** 934-7.

SWEET, D. G., CURLEY, A. E., CHESSHYRE, E., PIZZOTTI, J., WILBOURN, M. S., HALLIDAY, H. L. & WARNER, J. A. 2004. The role of matrix metalloproteinases -9 and -2 in development of neonatal chronic lung disease. *Acta Paediatr,* 93**,** 791-6.

TABB, D. L., VEGA-MONTOTO, L., RUDNICK, P. A., VARIYATH, A. M., HAM, A. J., BUNK, D. M., KILPATRICK, L. E., BILLHEIMER, D. D., BLACKMAN, R. K., CARDASIS, H. L., CARR, S. A., CLAUSER, K. R., JAFFE, J. D., KOWALSKI, K. A., NEUBERT, T. A., REGNIER, F. E., SCHILLING, B., TEGELER, T. J., WANG, M., WANG, P., WHITEAKER, J. R., ZIMMERMAN, L. J., FISHER, S. J., GIBSON, B. W., KINSINGER, C. R., MESRI, M., RODRIGUEZ, H., STEIN,

S. E., TEMPST, P., PAULOVICH, A. G., LIEBLER, D. C. & SPIEGELMAN, C. 2010. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res,* 9**,** 761-76.

TAY-UYBOCO, J. S., KWIATKOWSKI, K., CATES, D. B., KAVANAGH, L. & RIGATTO, H. 1989. Hypoxic airway constriction in infants of very low birth weight recovering from moderate to severe bronchopulmonary dysplasia. *J Pediatr,* 115**,** 456-9.

TEIG, N., ALLALI, M., RIEGER, C. & HAMELMANN, E. 2012. Inflammatory markers in induced sputum of school children born before 32 completed weeks of gestation. *J Pediatr,* 161**,** 1085-90.

THEBAUD, B., GOSS, K. N., LAUGHON, M., WHITSETT, J. A., ABMAN, S. H., STEINHORN, R. H., ASCHNER, J. L., DAVIS, P. G., MCGRATH-MORROW, S. A., SOLL, R. F. & JOBE, A. H. 2019. Bronchopulmonary dysplasia. *Nat Rev Dis Primers,* 5**,** 78.

THEODORIDIS, G., GIKA, H. G. & WILSON, I. D. 2011. Mass Spectrometry-Based Holistic Analytical Approaches for Metabolite Profiling in Systems Biology Studies. *Mass Spectrometry Reviews,* 30**,** 884-906.

TIDDENS, H., SILVERMAN, M. & BUSH, A. 2000. The role of inflammation in airway disease: remodeling. *Am J Respir Crit Care Med,* 162**,** S7-S10.

UM-BERGSTRÖM, P., POURBAZARGAN, M., BRUNDIN, B., STRÖM, M., EZERSKYTE, M., GAO, J., BERGGREN BROSTRÖM, E., MELÉN, E., WHEELOCK, Å. M., LINDÉN, A. & SKÖLD, C. M. 2022. Increased cytotoxic T-cells in the airways of adults with former bronchopulmonary dysplasia. *European Respiratory Journal,* 60.

URS, R. C., EVANS, D. J., BRADSHAW, T. K., GIBBONS, J. T. D., SMITH, E. F., FOONG, R. E., WILSON, A. C. & SIMPSON, S. J. 2023. Inhaled corticosteroids to improve lung function in children (aged 6-12 years) who were born very preterm (PICSI): a randomised, double-blind, placebo-controlled trial. *Lancet Child Adolesc Health*.

VALIKANGAS, T., SUOMI, T. & ELO, L. L. 2018. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform,* 19**,** 1-11.

VAN DER DOES, A. M., HEIJINK, M., PERSSON, L. J., KLOOS, D., AANERUD, M., BAKKE, P., TAUBE, C., EAGAN, T., HIEMSTRA, P. S. & GIERA, M. 2017. Disturbed fatty acid metabolism in airway secretions of patients with Chronic Obstructive Pulmonary Disease. *European Respiratory Journal,* 50.

VISCARDI, R. M., TERRIN, M. L., MAGDER, L. S., DAVIS, N. L., DULKERIAN, S. J., WAITES, K. B., ALLEN, M., AJAYI-AKINTADE, A., AMBALAVANAN, N., KAUFMAN, D. A., DONOHUE, P., TUTTLE, D. J. & WEITKAMP, J. H. 2021. Randomized trial of azithromycin to eradicate Ureaplasma respiratory colonization in preterm infants: 2-year outcomes. *Pediatr Res*.

VISCARDI, R. M., TERRIN, M. L., MAGDER, L. S., DAVIS, N. L., DULKERIAN, S. J., WAITES, K. B., AMBALAVANAN, N., KAUFMAN, D. A., DONOHUE, P., TUTTLE, D. J., WEITKAMP, J. H., HASSAN, H. E. & EDDINGTON, N. D. 2020. Randomised trial of azithromycin to eradicate Ureaplasma in preterm infants. *Arch Dis Child Fetal Neonatal Ed,* 105**,** 615-622.

WADA, H., GOTO, H., SAITOH, E., IEKI, R., OKAMURA, T., OTA, T., HAGIWARA, S., KODAKA, T. & YAMAMOTO, Y. 2005. Reduction in plasma free fatty acid in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 171**,** 1465.

WADE, A. M. & TUCKER, H. N. 1998. Antioxidant characteristics of L-histidine 11The work described in this manuscript was partially sponsored and funded by Cytos Pharmaceuticals, LLC. *The Journal of Nutritional Biochemistry,* 9**,** 308-315.

WAGENAAR, G. T., TER HORST, S. A., VAN GASTELEN, M. A., LEIJSER, L. M., MAUAD, T., VAN DER VELDEN, P. A., DE HEER, E., HIEMSTRA, P. S., POORTHUIS, B. J. & WALTHER, F. J. 2004. Gene expression profile and histopathology of experimental bronchopulmonary dysplasia induced by prolonged oxidative stress. *Free Radic Biol Med,* 36**,** 782-801.

WALSH, M. C., YAO, Q., GETTNER, P., HALE, E., COLLINS, M., HENSMAN, A., EVERETTE, R., PETERS, N., MILLER, N., MURAN, G., AUTEN, K., NEWMAN, N., ROWAN, G., GRISBY, C., ARNELL, K., MILLER, L., BALL, B., MCDAVID, G., NATIONAL INSTITUTE OF CHILD, H. & HUMAN DEVELOPMENT NEONATAL RESEARCH, N. 2004. Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics,* 114**,** 1305-11.

WAN, E. S., BALTE, P., SCHWARTZ, J. E., BHATT, S. P., CASSANO, P. A., COUPER, D., DAVIGLUS, M. L., DRANSFIELD, M. T., GHARIB, S. A., JACOBS, D. R., JR., KALHAN, R., LONDON, S. J., NAVAS-ACIEN, A., O'CONNOR, G. T., SANDERS, J. L., SMITH, B. M., WHITE, W., YENDE, S. & OELSNER, E. C. 2021. Association Between Preserved Ratio Impaired Spirometry and Clinical Outcomes in US Adults. *JAMA,* 326**,** 2287-2298.

WANG, F., CHEN, S., JIANG, Y., ZHAO, Y., SUN, L., ZHENG, B., CHEN, L., LIU, Z., ZHENG, X., YI, K., LI, C. & ZHOU, X. 2018. Effects of ammonia on apoptosis and oxidative stress in bovine mammary epithelial cells. *Mutagenesis,* 33**,** 291-299.

WANG, L., TANG, Y., LIU, S., MAO, S., LING, Y., LIU, D., HE, X. & WANG, X. 2013. Metabonomic profiling of serum and urine by (1)H NMR-based spectroscopy discriminates patients with chronic obstructive pulmonary disease and healthy individuals. *PLoS One,* 8**,** e65675.

WANG, R., KANG, H., ZHANG, X., NIE, Q., WANG, H., WANG, C. & ZHOU, S. 2022. Urinary metabolomics for discovering metabolic biomarkers of bladder cancer by UPLC-MS. *BMC Cancer,* 22**,** 214.

WASSERZUG, O. & DEROWE, A. 2016. Subglottic Stenosis: Current Concepts and Recent Advances. *Int J Head Neck Surg,* 7**,** 97-103.

WATKINS, W. J., COURSE, C. W., COUSINS, M., HART, K., KOTECHA, S. J. & KOTECHA, S. 2024. Impact of ambient air pollution on lung function in preterm-born school-aged children. *Thorax.*

WATTERBERG, K. L., CARMICHAEL, D. F., GERDES, J. S., WERNER, S., BACKSTROM, C. & MURPHY, S. 1994. Secretory leukocyte protease inhibitor and lung inflammation in developing bronchopulmonary dysplasia. *J Pediatr,* 125**,** 264-9.

WEBBE, J. W. H., DUFFY, J. M. N., AFONSO, E., AL-MUZAFFAR, I., BRUNTON, G., GREENOUGH, A., HALL, N. J., KNIGHT, M., LATOUR, J. M., LEE-DAVEY, C., MARLOW, N., NOAKES, L., NYCYK, J., RICHARD-LONDT, A., WILLS-EVE, B., MODI, N. & GALE, C. 2020. Core outcomes in neonatology: development of a core outcome set for neonatal research. *Arch Dis Child Fetal Neonatal Ed,* 105**,** 425-431.

WELLS, J. M., PARKER, M. M., OSTER, R. A., BOWLER, R. P., DRANSFIELD, M. T., BHATT, S. P., CHO, M. H., KIM, V., CURTIS, J. L., MARTINEZ, F. J., PAINE, R., 3RD, O'NEAL, W., LABAKI, W. W., KANER, R. J., BARJAKTAREVIC, I., HAN, M. K., SILVERMAN, E. K., CRAPO, J. D., BARR, R. G., WOODRUFF, P., CASTALDI, P. J., GAGGAR, A., SPIROMICS & INVESTIGATORS, C. O. 2018. Elevated circulating MMP-9 is linked to increased COPD exacerbation risk in SPIROMICS and COPDGene. *JCI Insight,* 3.

WEN, Y. H., YANG, H. I., CHOU, H. C., CHEN, C. Y., HSIEH, W. S., TSOU, K. I., TSAO, P. N. & TAIWAN PREMATURE INFANT DEVELOPMENTAL COLLABORATIVE STUDY, G. 2019. Association of Maternal Preeclampsia with Neonatal Respiratory Distress Syndrome in Very-Low-Birth-Weight Infants. *Sci Rep,* 9**,** 13212.

WIJNANT, S. R. A., DE ROOS, E., KAVOUSI, M., STRICKER, B. H., TERZIKHAN, N., LAHOUSSE, L. & BRUSSELLE, G. G. 2020. Trajectory and mortality of preserved ratio impaired spirometry: the Rotterdam Study. *Eur Respir J,* 55.

WILKINS, M. R., PASQUALI, C., APPEL, R. D., OU, K., GOLAZ, O., SANCHEZ, J. C., YAN, J. X., GOOLEY, A. A., HUGHES, G., HUMPHERY-SMITH, I., WILLIAMS, K. L. & HOCHSTRASSER, D. F. 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y),* 14**,** 61-5.

WILLIAMS, M., TODD, I. & FAIRCLOUGH, L. C. 2021. The role of CD8 + T lymphocytes in chronic obstructive pulmonary disease: a systematic review. *Inflamm Res,* 70**,** 11-18.

WISHART, D. S., FEUNANG, Y. D., MARCU, A., GUO, A. C., LIANG, K., VAZQUEZ-FRESNO, R., SAJED, T., JOHNSON, D., LI, C., KARU, N., SAYEEDA, Z., LO, E., ASSEMPOUR, N., BERJANSKII, M., SINGHAL, S., ARNDT, D., LIANG, Y., BADRAN, H., GRANT, J., SERRA-CAYUELA, A., LIU, Y., MANDAL, R., NEVEU, V., PON, A., KNOX, C., WILSON, M., MANACH, C. & SCALBERT, A. 2018. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res,* 46**,** D608-D617.

WORLD HEALTH ORGANISATION 2019. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10).

WU, G., LUPTON, J. R., TURNER, N. D., FANG, Y.-Z. & YANG, S. 2004. Glutathione Metabolism and Its Implications for Health. *The Journal of Nutrition,* 134**,** 489-492.

WU, J. & GAO, Y. 2015. Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Rev Proteomics,* 12**,** 623-36.

WU, K. Y., JENSEN, E. A., WHITE, A. M., WANG, Y., BIKO, D. M., NILAN, K., FRAGA, M. V., MERCER-ROSA, L., ZHANG, H. & KIRPALANI, H. 2020a. Characterization of Disease Phenotype in Very Preterm Infants with Severe Bronchopulmonary Dysplasia. *Am J Respir Crit Care Med,* 201**,** 1398-1406.

WU, Z., YAN, M., ZHANG, M., WU, N., MA, G., WANG, B., FAN, Y., DU, X., DING, C. & LIU, Y. 2020b. beta2-microglobulin as a biomarker of pulmonary fibrosis development in COPD patients. *Aging (Albany NY),* 13**,** 1251-1263.

XIAO, J., LU, X., CHEN, X., ZOU, Y., LIU, A., LI, W., HE, B., HE, S. & CHEN, Q. 2017. Eight potential biomarkers for distinguishing between lung adenocarcinoma and squamous cell carcinoma. *Oncotarget,* 8**,** 71759-71771.

YAMAUCHI, K. & OGASAWARA, M. 2019. The Role of Histamine in the Pathophysiology of Asthma and the Clinical Efficacy of Antihistamines in Asthma Therapy. *Int J Mol Sci,* 20.

YAO, X., GAO, M., DAI, C., MEYER, K. S., CHEN, J., KEERAN, K. J., NUGENT, G. Z., QU, X., YU, Z. X., DAGUR, P. K., MCCOY, J. P. & LEVINE, S. J. 2013. Peptidoglycan recognition protein 1 promotes house dust mite-induced airway inflammation in mice. *Am J Respir Cell Mol Biol,* 49**,** 902-11.

YUM, S. K., KIM, M. S., KWUN, Y., MOON, C. J., YOUN, Y. A. & SUNG, I. K. 2018. Impact of histologic chorioamnionitis on pulmonary hypertension and respiratory outcomes in preterm infants. *Pulm Circ,* 8**,** 2045894018760166.

ZECHA, J., SATPATHY, S., KANASHOVA, T., AVANESSIAN, S. C., KANE, M. H., CLAUSER, K. R., MERTINS, P., CARR, S. A. & KUSTER, B. 2019. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics,* 18**,** 1468-1478.

ZHANG, Z., WANG, J., LI, Y., LIU, F., CHEN, L., HE, S., LIN, F., WEI, X., FANG, Y., LI, Q., ZHOU, J. & LU, W. 2023. Proteomics and metabolomics profiling reveal panels of circulating diagnostic biomarkers and molecular subtypes in stable COPD. *Respir Res,* 24**,** 73.

ZHEN, J., ZHAO, P., LI, Y., CAI, Y., YU, W., WANG, W., ZHAO, L., WANG, H., HUANG, G. & XU, A. 2022. The Multiomics Analyses of Gut Microbiota, Urine Metabolome and Plasma Proteome Revealed Significant Changes in Allergy Featured with Indole Derivatives of Tryptophan. *J Asthma Allergy,* 15**,** 117-131.

8 Appendices

Accession Number	Gene Name	Protein Name	Number of Samples	Accession Number	Gene Name	Protein Name	Number of Samples
H6VRG2	KRT1	Cytokeratin-1	218	P04040	CAT	Catalase	191
H6VRG3	KRT1	Cytokeratin-1	218	Q01469	FABP5	Fatty acid-binding protein 5	190
P01040	CSTA	Cystatin-A	218	P02768	ALB	Albumin	187
P02533	KRT14	Keratin, type I cytoskeletal 14	218	P04406	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	186
P02538	KRT6A	Keratin, type II cytoskeletal 6A	218	Q15517	CDSN	Corneodesmosin	185
P08779	KRT16	Keratin, type I cytoskeletal 16	218	Q9NZT1	CALML5	Calmodulin-like protein 5	184
P13645	KRT10	Keratin, type I cytoskeletal 10	218	P62736	ACTA2	Actin, aortic smooth muscle	165
P13647	KRT5	Keratin, type II cytoskeletal 5	218	Q7Z794	KRT77	Keratin, type II cytoskeletal 1b	165
P14923	JUP	Junction plakoglobin	218	P31151	S100A7	Protein S100-A7	164
P15924	DSP	Desmoplakin	218	P10599	TXN	Thioredoxin	156
P25311	AZGP1	Zinc-alpha-2-glycoprotein	218	P31025	LCN1	Lipocalin-1	154
P31944	CASP14	Caspase-14	218	Q13835	PKP1	Plakophilin-1	154
P35527	KRT9	Keratin, type I cytoskeletal 9	218	Q6UWP8	SBSN	Suprabasin	153
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal	218	P05089	ARG1	Arginase-1	152
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a	218	P12273	PIP	Prolactin-inducible protein	148
Q02413	DSG1	Desmoglein-1	218	Q06830	PRDX1	Peroxiredoxin-1	144
Q04695	KRT17	Keratin, type I cytoskeletal 17	218	P06702	S100A9	Protein S100-A9	142
Q08554	DSC1	Desmocollin-1	218	P60709	ACTB	Actin, cytoplasmic 1	137
Q8N1N4	KRT78	Keratin, type II cytoskeletal 78	218	Q96QE3	ATAD5	ATPase family AAA domain-containing protein 5	128
Q5D862	FLG2	Filaggrin-2	214	P32119	PRDX2	Peroxiredoxin-2	127
P81605	DCD	Dermcidin	210	A1A4E9	KRT13	Keratin 13	123
P07355	ANXA2	Annexin A2	198	Q6KB66	KRT80	Keratin, type II cytoskeletal 80	122
Q14CN4	KRT72	Keratin, type II cytoskeletal 72	198	P78386	KRT85	Keratin, type II cuticular Hb5	119

O75223	GGCT	Gamma-glutamylcyclotransferase	195	P04792	HSPB1	Heat shock protein beta-1	118
P20930	FLG	Filaggrin	194	P47929	LGALS7	Galectin-7	116
P04259	KRT6B	Keratin, type II cytoskeletal 6B	193	P05109	S100A8	Protein S100-A8	111
P01834	IGKC	Immunoglobulin kappa constant	104	Q5T749	KPRP	Keratinocyte proline-rich protein	50
Q08188	TGM3	Protein-glutamine gamma-glutamyltransferase E	102	Q5T749	KPRP	Keratinocyte proline-rich protein	50
Q13867	BLMH	Bleomycin hydrolase	101	Q9UI42	CPA4	Carboxypeptidase A4	48
Q96P63	SERPINB12	Serpin B12	99	P22531	SPRR2E	Small proline-rich protein 2E	43
B4DKJ0	Unknown	cDNA FLJ58539, highly similar to Keratin, type II cytoskeletal 4	95	Q16610	ECM1	Extracellular matrix protein 1	40
P69905	HBA1	Haemoglobin subunit alpha	93	P06733	ENO1	Alpha-enolase	39
Q86YZ3	HRNR	Hornerin	92	P08670	VIM	Vimentin	39
P0DUB6	AMY1A	Alpha-amylase 1A	91	Q9GZZ8	LACRT	Extracellular glycoprotein lacritin	35
P27482	CALML3	Calmodulin-like protein 3	88	W8QEY1	Unknown	Lactotransferrin	35
Q15323	KRT31	Keratin, type I cuticular Ha1	87	Q9UGM3	DMBT1	Deleted in malignant brain tumours 1 protein	34
Q7Z3Y8	KRT27	Keratin, type I cytoskeletal 27	84	Q9UJ41	RABGEF1	Rab5 GDP/GTP exchange factor	34
P61626	LYZ	Lysozyme C	83	O76011	KRT34	Keratin, type I cuticular Ha4	33
P02810	PRH1	Salivary acidic proline-rich phosphoprotein 1/2	81	Q9HCM4	EPB41L5	Band 4.1-like protein 5	33
P04083	ANXA1	Annexin A1	81	P62805	H4C1	Histone H4	32
Q9NSB2	KRT84	Keratin, type II cuticular Hb4	81	O76013	KRT36	Keratin, type I cuticular Ha6	31
P68871	HBB	Haemoglobin subunit beta	72	P01591	JCHAIN	Immunoglobulin J chain	30
Q14525	KRT33B	Keratin, type I cuticular Ha3-II	72	Q9C075	KRT23	Keratin, type I cytoskeletal 23	30
B2R853	Unknown	cDNA, FLJ93744, highly similar to Homo sapiens keratin 6E (KRT6E), mRNA	70	Q9Y224	RTRAF	RNA transcription, translation and transport factor protein	30
P01876	IGHA1	Immunoglobulin heavy constant alpha 1	66	P62328	TMSB4X	Thymosin beta-4	29
Q15828	CST6	Cystatin-M	66	P01857	IGHG1	Immunoglobulin heavy constant gamma 1	28
O60814	H2BC12	Histone H2B type 1-K	65	P14618	PKM	Pyruvate kinase PKM	28

P00441	SOD1	Superoxide dismutase	61	Q8TF66	LRRC15	Leucine-rich repeat-containing protein 15	28
P23490	LORICRIN	Loricrin	61	G3V1M9	PRB1	Basic salivary proline-rich protein 1	27
P29508	SERPINB3	Serpin B3	58	P12035	KRT3	Keratin, type II cytoskeletal 3	27
A0A0S2Z4G4	TPM3	Tropomyosin 3 isoform 1	57	P68431	H3C1	Histone H3.1	27
Q96DR8	MUCL1	Mucin-like protein 1	54	Q9BYR6	KRTAP3-3	Keratin-associated protein 3-3	25
O43790	KRT86	Keratin, type II cuticular Hb6	53	Q9HCY8	S100A14	Protein S100-A14	25
P22735	TGM1	Protein-glutamine gamma-glutamyltransferase K	24	Q92820	GGH	Gamma-glutamyl hydrolase	15
P82279	CRB1	Protein crumbs homolog 1	24	Q9UKX2	MYH2	Myosin-2	14
Q05639	EEF1A2	Elongation factor 1-alpha 2	23	P07339	CTSD	Cathepsin D	13
P04080	CSTB	Cystatin-B	20	P78385	KRT83	Keratin, type II cuticular Hb3	13
P15090	FABP4	Fatty acid-binding protein, adipocyte	20	P31947	SFN	14-3-3 protein sigma	12
P63104	YWHAZ	14-3-3 protein zeta/delta	20	P60174	TPI1	Triosephosphate isomerase	12
Q52LG2	KRTAP13-2	Keratin-associated protein 13-2	20	P42357	HAL	Histidine ammonia-lyase	11
Q5T750	XP32	Skin-specific protein 32	20	P02814	SMR3B	Submaxillary gland androgen-regulated protein 3B	10
Q6ZVX7	NCCRP1	F-box only protein 50	20	P0DPA2	VSIG8	V-set and immunoglobulin domain-containing protein 8	10
P35579	MYH9	Myosin-9	19	P16615	ATP2A2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	10
Q6E0U4	DMKN	Dermokine	19	Q6A163	KRT39	Keratin, type I cytoskeletal 39	10
Q8IUC0	KRTAP13-1	Keratin-associated protein 13-1	19	Q96DA0	ZG16B	Zymogen granule protein 16 homolog B	10
B4DJM5	Unknown	cDNA FLJ61294, highly similar to Keratin, type I cytoskeletal 17	18	Q9NQ38	SPINK5	Serine protease inhibitor Kazal-type 5	10
P01036	CST4	Cystatin-S	18	Q9UKK9	NUDT5	ADP-sugar pyrophosphatase	10
P02452	COL1A1	Collagen alpha-1(I) chain	18	A2IDD5	CCDC78	Coiled-coil domain-containing protein 78	9
P03973	SLPI	Secretory leukocyte peptidase inhibitor	18	P02812	PRB2	Basic salivary proline-rich protein 2	9
Q14533	KRT81	Keratin, type II cuticular Hb1	18	P06454	PTMA	Prothymosin alpha	9

Q3LI77	KRTAP13-4	Keratin-associated protein 13-4	18	P16104	H2AX	Histone H2AX	9
Q96QA5	GSDMA	Gasdermin-A	18	Q02383	SEMG2	Semenogelin-2	9
P13929	ENO3	Beta-enolase	17	Q3SY84	KRT71	Keratin, type II cytoskeletal 71	9
Q14574	DSC3	Desmocollin-3	17	Q6S8J3	POTEE	POTE ankyrin domain family member E	9
Q2PPJ7	RALGAPA2	RalGTPase-activating protein subunitalpha-2	17	Q6YFL4	KRTHB6	Type II keratin (Fragment)	9
Q6ZUA9	MROH5	Maestro heat-like repeat family member 5	17	Q92764	KRT35	Keratin, type I cuticular Ha5	9
P17066	HSPA6	Heat shock 70 kDa protein 6	15	Q9BZE2	PUS3	tRNA pseudouridine(38/39) synthase	9
P22528	SPRR1B	Cornifin-B	15	Q9HC84	MUC5B	Mucin-5B (MUC-5B)	9
Q14508	WFDC2	WAP four-disulfide core domain protein 2	15	P04279	SEMG1	Semenogelin-1	8
P05976	MYL1	Myosin light chain 1/3, skeletal muscle isoform	8	Q9BYR9	KRTAP2-4	Keratin-associated protein 2-4	7
P09228	CST2	Cystatin-SA	8	O43829	ZBTB14	Zinc finger and BTB domain-containing protein 14	6
P11055	MYH3	Myosin-3	8	Q5XKE5	KRT79	Keratin, type II cytoskeletal 79	6
P31949	S100A11	Protein S100-A11	8	Q9Y618	NCOR2	Nuclear receptor corepressor 2	6
P49454	CENPF	Centromere protein F	8	P13639	EEF2	Elongation factor 2	5
P52566	ARHGDIB	Rho GDP-dissociation inhibitor 2	8	P62937	PPIA	Peptidyl-prolyl cis-trans isomerase A	5
P00491	PNP	Purine nucleoside phosphorylase	7	A0A1B0GVI3	KRT10	Keratin, type I cytoskeletal 10 (Fragment)	4
P09211	GSTP1	Glutathione S-transferase P	7	D3DTX7	COL1A1	Collagen, type I, alpha 1, isoform CRA_a	4
P11021	HSPA5	Endoplasmic reticulum chaperone BiP	7	O95678	KRT75	Keratin, type II cytoskeletal 75	4
P47756	CAPZB	F-actin-capping protein subunit beta	7	P05455	SSB	Lupus La protein	4
P60900	PSMA6	Proteasome subunit alpha type-6	7	Q9BYR8	KRTAP3-1	Keratin-associated protein 3-1	4
Q8NEZ4	KMT2C	Histone-lysine N-methyltransferase 2C	7	Q9H1E1	RNASE7	Ribonuclease 7	4
Q8WVV4	POF1B	Protein POF1B	7	Q15149	PLEC	Plectin	2

Table 8-1: List of all proteins detected within EBC samples, in decreasing order of number of samples in which it was detected

Metabolite	% BPD Samples	BPD mean	BPD SD	% NoBPD Samples	NoBPD mean	NoBPD SD	% Term samples	Term mean	Term SD
1,2,4-benzenetriol	23.5	1.81	0.29	30.9	1.88	0.29	29.6	1.78	0.36
1,2-anhydromyoinositol	5.9	1.33	0.53	14.5	1.63	0.42	9.9	1.58	0.35
1-hexadecanol	100	2.50	0.11	99.1	2.48	0.14	100	2.48	0.12
1-kestose	0	1.48	0.20	4.5	1.51	0.30	1.4	1.50	0.29
1-methyladenosine	0	1.40	0.22	0.9	1.42	0.23	1.4	1.37	0.26
1-monostearin	100	2.73	07	99.1	2.75	07	100	2.72	0.10
2,3-bisphosphoglyceric acid	26.5	1.86	0.33	35.5	1.96	0.50	29.6	1.92	0.53
2,3-dihydroxybutanoic acid	0	1.45	0.23	0	1.44	0.29	2.8	1.44	0.33
2,8-dihydroxyquinoline	0	1.49	0.14	0	1.39	0.22	1.4	1.45	0.19
2-aminophenol	0	1.51	0.23	0	1.55	0.20	1.4	1.50	0.23
2-deoxytetronic acid	5.9	1.70	0.18	5.5	1.74	0.21	11.3	1.78	0.34
2-hydroxy-2-methylbutanoic acid	97.1	3.08	0.68	91.8	3.16	0.66	93.0	3.13	0.59
2-hydroxyglutaric acid	35.3	1.95	0.13	30.9	1.93	0.16	42.3	1.99	0.18
2-hydroxyhippuric acid	2.9	1.50	0.31	0	1.44	0.23	1.4	1.38	0.35
2-hydroxypyrazinyl-2-propenoic acidethylester	100	3.55	08	99.1	3.53	0.13	100	3.57	0.10
2-hydroxyvaleric acid	100	3.50	04	99.1	3.45	0.17	100	3.49	06
2-ketoisocaproic acid	100	3.04	0.31	98.2	3.07	0.33	98.6	3.11	0.31
2-methylglyceric acid	0	1.55	0.18	2.7	1.57	0.27	4.2	1.64	0.23
2-monopalmitin	100	2.98	0.55	99.1	2.90	0.57	100	2.81	0.47
2-picolinic acid	17.6	1.75	0.38	35.5	1.81	0.39	36.6	1.83	0.46
3,3-hydroxyphenyl-3-hydroxypropionic acid	0	1.50	0.23	1.8	1.49	0.22	4.2	1.53	0.38
3,3-hydroxyphenylpropionic acid	100	2.49	04	98.2	2.48	0.11	100	2.49	03
3,4-dihydroxybenzoic acid	0	1.71	0.18	3.6	1.70	0.22	5.6	1.69	0.36

3,4-dihydroxycinnamic acid	55.9	2.03	0.13	60.9	2.03	0.20	63.4	2.06	0.18
3,4-dihydroxyhydrocinnamic acid	8.8	1.62	0.27	16.4	1.72	0.43	12.7	1.70	0.29
3,4-dihydroxyphenylacetic acid	20.6	1.68	0.44	19.1	1.63	0.40	21.1	1.67	0.40
3,6-anhydro-D-galactose	2.9	1.47	0.28	0	1.44	0.26	1.4	1.51	0.30
3-aminoisobutyric acid	73.5	2.19	0.24	67.3	2.16	0.32	78.9	2.23	0.29
3-hydroxy-3,4-hydroxy-3-methoxyphenylpropionic acid	0	1.44	0.22	0	1.42	0.32	1.4	1.44	0.35
3-hydroxy-3-methylglutaric acid	0	1.46	0.18	0	1.41	0.27	1.4	1.49	0.27
3-hydroxyphenylacetic acid	0	1.48	0.20	0	1.50	0.18	2.8	1.54	0.27
3-hydroxypropionic acid	100	3.00	06	99.1	3.00	0.11	100	3.00	07
3-phosphoglycerate	0	1.48	0.17	1.8	1.44	0.25	1.4	1.50	0.24
4-hydroxybenzoate	44.1	2.00	0.21	47.3	1.99	0.20	52.1	2.00	0.27
4-hydroxyhippuric acid	14.7	1.65	0.40	20.9	1.69	0.37	28.2	1.81	0.38
4-hydroxyphenylacetic acid	2.9	1.45	0.37	5.5	1.48	0.36	5.6	1.53	0.41
4-methylcatechol	0	1.52	0.18	0	1.51	0.19	1.4	1.49	0.22
4-pyridoxic acid	0	1.37	0.24	0	1.40	0.28	1.4	1.43	0.28
5-aminovaleric acid	23.5	1.93	0.17	35.5	1.98	0.31	22.5	1.92	0.18
5-deoxy-5-methylthioadenosine	0	1.38	0.29	0	1.38	0.34	1.4	1.47	0.28
5-hydroxy-3-indoleacetic acid	0	1.39	0.32	0	1.42	0.20	1.4	1.42	0.29
5-hydroxymethyl-2-furoic acid	100	2.29	0.10	98.2	2.27	0.10	100	2.29	0.16
6-deoxyglucitol	0	1.45	0.25	0.9	1.45	0.26	1.4	1.44	0.29
7-methylguanine	0	1.41	0.32	0	1.44	0.27	1.4	1.45	0.28
aconitic acid	0	1.50	0.27	3.6	1.54	0.26	4.2	1.61	0.38
adenine	0	1.47	0.23	0	1.48	0.21	1.4	1.48	0.25
adenosine	2.9	1.53	0.25	1.8	1.55	0.39	7.0	1.58	0.40
adipic acid	100	3.28	0.23	99.1	3.27	0.21	100	3.26	0.18
alanine	100	3.25	0.27	99.1	3.42	0.43	100	3.42	0.36

allantoic acid	0	1.47	0.20	0	1.45	0.26	1.4	1.42	0.36
alloxanoic acid	8.8	1.84	0.15	15.5	1.83	0.21	11.3	1.88	0.15
alphaketoglutarate	0	1.47	0.26	4.5	1.51	0.40	7.0	1.53	0.43
aminomalonate	2.9	1.45	0.39	10	1.62	0.33	7.0	1.59	0.29
anthranilic acid	0	1.43	0.23	0	1.47	0.30	2.8	1.53	0.32
arabitol	100	2.77	07	99.1	2.80	0.11	100	2.83	0.27
arachidic acid	100	3.55	05	99.1	3.56	0.12	100	3.55	0.10
ascorbic acid	0	1.41	0.20	0	1.34	0.27	1.4	1.45	0.22
asparagine	100	2.70	08	99.1	2.73	0.18	100	2.73	0.17
aspartic acid	100	2.70	0.12	99.1	2.78	0.40	100	2.77	0.28
azelaic acid	97.1	2.32	0.14	96.4	2.33	0.17	93.0	2.30	0.22
benzoic acid	100	4.10	04	99.1	4.10	04	100	4.10	04
beta alanine	0	1.40	0.35	3.6	1.53	0.25	1.4	1.48	0.29
betagentiobiose	14.7	1.71	0.24	19.1	1.71	0.45	15.5	1.63	0.46
betamannosylglycerate	0	1.39	0.32	0	1.44	0.31	1.4	1.43	0.33
biphenyl	100	2.34	06	99.1	2.33	07	100	2.32	07
butane-2,3-diol	100	3.50	0.32	99.1	3.61	0.48	100	3.59	0.41
butyrolactam	88.2	2.19	0.47	89.1	2.28	0.35	90.1	2.24	0.37
capric acid	100	3.17	07	99.1	3.16	07	100	3.16	06
caprylic acid	100	3.91	04	99.1	3.89	05	100	3.91	05
catechol	0	1.53	0.22	0	1.51	0.26	2.8	1.43	0.36
cellobiose	79.4	2.25	0.32	81.8	2.31	0.48	77.5	2.28	0.40
ceratinic acid	91.2	2.20	0.12	91.8	2.20	0.13	90.1	2.18	0.17
cholesterol	61.8	2.11	0.23	68.2	2.16	0.35	57.7	2.09	0.38
citramalic acid	14.7	1.74	0.21	14.5	1.75	0.34	14.1	1.76	0.29
citric acid	91.2	2.38	0.29	96.4	2.49	0.44	97.2	2.54	0.58
citrulline	14.7	1.73	0.32	16.4	1.74	0.41	18.3	1.80	0.40

conduritolbetaepoxide	26.5	1.71	0.47	22.7	1.74	0.50	38.0	1.92	0.55
creatinine	29.4	1.77	0.73	30.9	1.89	0.76	25.4	1.83	0.95
cystathionine	0	1.51	0.27	3.6	1.51	0.37	2.8	1.56	0.30
cysteine	2.9	1.55	0.36	10.9	1.60	0.38	12.7	1.64	0.48
cysteineglycine	32.4	1.84	0.41	42.7	1.91	0.43	42.3	1.96	0.39
cystine	26.5	1.84	0.22	20.9	1.86	0.20	42.3	1.94	0.23
dehydroabietic acid	97.1	3.00	0.43	96.4	2.91	0.49	95.8	2.87	0.56
dehydroascorbic acid	58.8	2.00	08	49.1	1.98	0.15	50.7	2.02	0.20
deoxypentitol	0	1.48	0.16	5.5	1.54	0.30	4.2	1.55	0.35
digalacturonic acid	29.4	1.92	0.12	30	1.92	0.14	39.4	1.93	0.15
digitoxose	100	2.60	0.13	99.1	2.57	0.17	100	2.64	0.15
diglycerol	100	2.73	06	99.1	2.73	08	100	2.74	0.13
dihydro-3-coumaric acid	100	2.51	05	97.3	2.47	0.12	100	2.50	05
docosenoic acid	100	3.27	0.42	99.1	3.32	0.58	100	3.36	0.49
dodecanol	100	3.25	0.10	99.1	3.24	0.12	100	3.25	09
enolpyruvate	2.9	1.70	0.17	2.7	1.68	0.21	0	1.68	0.18
erythritol	23.5	1.70	0.34	26.4	1.67	0.46	19.7	1.71	0.60
erythronic acid lactone	0	1.44	0.31	0	1.47	0.21	1.4	1.45	0.25
erythrose major	2.9	1.47	0.31	10.9	1.62	0.31	7.0	1.55	0.30
ethanolamine	100	2.84	0.27	96.4	2.91	0.40	98.6	2.89	0.46
ferulic acid	0	1.44	0.22	0	1.41	0.28	1.4	1.43	0.27
fructose	76.5	2.27	0.37	72.7	2.37	0.62	80.3	2.40	0.54
fucose	100	2.85	06	99.1	2.84	08	100	2.85	0.15
fumaric acid	76.5	2.13	0.15	72.7	2.13	0.24	74.6	2.14	0.27
furoylglycine	55.9	1.95	0.53	43.6	1.98	0.45	42.3	1.84	0.56
galactinol	38.2	1.98	0.31	55.5	2.08	0.34	56.3	2.10	0.37
galactitol	38.2	1.77	0.50	48.2	2.08	0.68	42.3	1.99	0.57

galactonic acid	5.9	1.47	0.32	5.5	1.54	0.28	5.6	1.56	0.38
galactose	85.3	2.37	0.39	80.9	2.54	0.62	87.3	2.53	0.62
glucoheptulose	0	1.40	0.36	0	1.44	0.30	1.4	1.49	0.32
gluconic acid	2.9	1.53	0.21	8.2	1.60	0.33	14.1	1.66	0.40
gluconic acidlactone	0	1.46	0.25	0	1.41	0.30	4.2	1.48	0.31
glucose	47.1	2.12	0.46	54.5	2.25	0.53	66.2	2.36	0.57
glucuronic acid	0	1.57	0.30	4.5	1.62	0.25	5.6	1.62	0.37
glutamic acid	8.8	1.61	0.28	20.9	1.82	0.51	25.4	1.79	0.41
glutamine	100	3.98	0.38	99.1	4.09	0.29	100	4.01	0.38
glutaric acid	88.2	2.18	0.18	79.1	2.15	0.19	84.5	2.17	0.20
glycerol	100	3.92	0.17	99.1	3.96	0.21	100	3.87	0.15
glycerol-3-galactoside	14.7	1.50	0.48	15.5	1.64	0.32	28.2	1.80	0.40
glycerol alpha phosphate	0	1.47	0.20	0	1.46	0.18	1.4	1.44	0.29
glycine	100	3.42	0.33	99.1	3.53	0.35	100	3.55	0.27
glycolic acid	100	2.82	0.16	98.2	2.83	0.20	100	2.86	0.23
glycylproline	2.9	1.53	0.22	7.3	1.59	0.31	7.0	1.61	0.41
guanine	0	1.53	0.20	0	1.50	0.22	1.4	1.47	0.27
heptadecanoic acid	100	3.73	04	99.1	3.72	04	100	3.72	03
hippuric acid	5.9	1.60	0.23	10.9	1.67	0.39	8.5	1.71	0.53
histidine	32.4	1.97	0.14	53.6	2.07	0.30	56.3	2.05	0.24
homocystine	0	1.42	0.34	3.6	1.56	0.31	4.2	1.45	0.44
homovanillic acid	5.9	1.51	0.36	6.4	1.53	0.32	7.0	1.53	0.48
hydroxycarbamate	100	2.73	0.11	99.1	2.71	0.10	100	2.70	09
hydroxyproline dipeptide	2.9	1.68	0.16	4.5	1.65	0.23	2.8	1.72	0.37
hypoxanthine	0	1.38	0.32	2.7	1.46	0.33	2.8	1.43	0.45
indole-3-acetate	0	1.52	0.20	0.9	1.45	0.25	2.8	1.56	0.32
indoxylsulfate	0	1.37	0.35	4.5	1.41	0.32	4.2	1.45	0.45

isocitric acid	0	1.41	0.18	1.8	1.50	0.24	5.6	1.54	0.43
isoleucine	100	2.86	0.14	99.1	2.96	0.36	100	2.93	0.24
isopropyl benzene	100	3.75	0.51	99.1	3.82	0.47	100	3.66	0.48
isothreonic acid	5.9	1.75	0.14	10	1.78	0.19	8.5	1.83	0.37
itaconic acid	97.1	2.69	0.32	95.5	2.56	0.28	95.8	2.67	0.50
kynurenic acid	0	1.40	0.35	0	1.47	0.21	2.8	1.42	0.38
lactic acid	100	3.62	0.24	99.1	3.69	0.50	100	3.70	0.35
lactose	35.3	1.93	0.23	38.2	1.97	0.25	43.7	1.94	0.21
lactulose	55.9	2.09	0.33	50	2.14	0.54	46.5	2.11	0.46
lauric acid	100	3.98	05	99.1	3.97	05	100	3.97	0.10
leucine	100	3.48	0.83	99.1	3.39	0.73	100	3.46	0.71
levoglucosan	32.4	1.92	0.26	47.3	1.98	0.26	39.4	2.01	0.34
levoinositol	0	1.38	0.23	0	1.37	0.34	1.4	1.36	0.42
lysine	2.9	1.59	0.18	11.8	1.67	0.47	12.7	1.65	0.39
maleimide	100	2.43	0.28	85.5	2.33	0.30	93.0	2.38	0.29
malic acid	35.3	1.99	0.35	41.8	2.08	0.45	54.9	2.12	0.40
maltose-1	91.2	2.30	0.30	90.9	2.42	0.45	84.5	2.30	0.37
mannose	23.5	1.91	0.52	37.3	2.14	0.84	42.3	2.16	0.83
metanephrine	0	1.45	0.20	0.9	1.43	0.28	1.4	1.42	0.32
methanolphosphate	26.5	1.94	0.97	23.6	1.92	0.93	40.8	2.18	1.12
methionine	14.7	1.38	0.52	12.7	1.48	0.60	12.7	1.57	0.60
methylmaleic acid	0	1.62	0.17	0.9	1.62	0.16	1.4	1.62	0.16
montanic acid	100	2.46	0.10	97.3	2.42	0.13	100	2.46	09
myoinositol	26.5	1.92	0.33	37.3	1.98	0.43	50.7	2.05	0.47
myristic acid	100	3.41	04	99.1	3.40	05	100	3.40	04
N-acetylaspartic acid	2.9	1.59	0.24	3.6	1.62	0.32	8.5	1.63	0.40
N-acetylmannosamine	2.9	1.75	0.14	4.5	1.75	0.17	5.6	1.77	0.22

N-acetylputrescine	5.9	1.59	0.26	21.8	1.86	0.74	22.5	1.74	0.67
N-carbamoylaspartate	2.9	1.44	0.20	3.6	1.51	0.28	4.2	1.54	0.29
N-carbamylglutamate	0	1.47	0.31	0.9	1.51	0.24	4.2	1.52	0.34
nepsilontrimethyllysine	0	1.67	0.20	1.8	1.64	0.28	4.2	1.64	0.22
nicotinic acid	14.7	1.76	0.25	14.5	1.80	0.54	22.5	1.84	0.49
N-methylglutamic acid	88.2	2.30	0.26	89.1	2.29	0.24	91.5	2.28	0.28
nonadecanoic acid	100	2.74	06	99.1	2.71	07	100	2.71	07
noradrenaline	0	1.50	0.29	2.7	1.53	0.33	2.8	1.46	0.28
octadecanol	100	3.05	0.12	99.1	3.01	08	100	3.02	0.10
oleamide	82.4	2.48	0.39	83.6	2.38	0.34	97.2	2.54	0.32
oleic acid	97.1	2.56	0.18	99.1	2.56	0.14	100	2.57	0.11
ornithine	97.1	2.72	0.41	99.1	2.92	0.59	97.2	2.88	0.53
orotic acid	0	1.50	0.26	0	1.37	0.34	1.4	1.43	0.35
palatinitol	0	1.42	0.32	6.4	1.49	0.38	2.8	1.52	0.28
palmitoleic acid	5.9	1.67	0.19	0	1.59	0.18	1.4	1.56	0.26
parabanic acid	94.1	2.76	0.35	94.5	2.81	0.34	95.8	2.80	0.32
pcresol	100	2.53	0.15	97.3	2.55	0.19	98.6	2.59	0.31
pentitol	0	1.34	0.21	0	1.36	0.31	1.4	1.43	0.36
pentose	73.5	2.25	0.29	65.5	2.15	0.30	60.6	2.14	0.39
phenol	100	3.20	05	99.1	3.21	0.13	100	3.20	08
phenylalanine	100	2.70	09	99.1	2.75	0.22	100	2.73	0.17
phosphate	100	3.24	0.64	98.2	3.26	0.63	98.6	3.20	0.54
pimelic acid	100	2.48	0.15	98.2	2.48	0.15	98.6	2.47	0.20
pinitol	55.9	2.11	0.39	53.6	2.09	0.45	49.3	2.10	0.43
proline	52.9	1.99	0.49	55.5	2.18	0.58	64.8	2.22	0.49
pseudo uridine	0	1.48	0.16	0.9	1.49	0.30	4.2	1.56	0.46
psicose	26.5	1.72	0.55	25.5	1.76	0.74	22.5	1.73	0.68

ptolylglucuronide	41.2	1.98	0.42	41.8	1.98	0.39	38.0	1.96	0.52
putrescine	32.4	1.90	0.29	37.3	2.14	0.67	33.8	2.03	0.62
pyrogallol	11.8	1.62	0.30	11.8	1.71	0.31	12.7	1.67	0.29
pyroglutamic acid	97.1	3.55	0.37	99.1	3.75	0.45	100	3.71	0.39
pyrophosphate	100	3.35	0.27	99.1	3.34	0.30	100	3.36	0.28
pyruvic acid	52.9	2.04	0.44	51.8	2.19	0.61	54.9	2.15	0.50
quinic acid	32.4	1.84	0.55	28.2	1.87	0.57	38.0	1.98	0.63
quinolinic acid	0	1.54	0.20	0	1.56	0.20	1.4	1.61	0.26
raffinose	0	1.48	0.23	4.5	1.51	0.32	4.2	1.53	0.24
ribitol	50	2.17	0.66	51.8	2.20	0.67	66.2	2.41	0.65
ribonic acid	0	1.60	0.19	5.5	1.60	0.26	2.8	1.59	0.31
ribose	0	1.45	0.25	1.8	1.46	0.31	5.6	1.46	0.44
saccharic acid	20.6	1.54	0.48	16.4	1.58	0.39	19.7	1.60	0.45
salicylaldehyde	73.5	2.14	0.35	70.9	2.16	0.25	81.7	2.26	0.25
salicylic acid	61.8	2.09	0.23	71.8	2.16	0.37	71.8	2.16	0.37
serine	97.1	2.92	0.64	98.2	3.23	0.79	100	3.19	0.72
serotonin	2.9	1.47	0.24	4.5	1.59	0.47	5.6	1.62	0.47
shikimic acid	97.1	2.27	0.23	98.2	2.40	0.38	97.2	2.37	0.38
sinapinic acid	14.7	1.88	0.18	16.4	1.90	0.11	21.1	1.91	0.14
sophorose	0	1.54	0.26	0	1.47	0.33	1.4	1.38	0.44
sorbitol	100	2.53	0.17	96.4	2.67	0.45	98.6	2.65	0.36
succinic acid	100	2.58	0.16	98.2	2.61	0.27	100	2.64	0.25
sucrose	97.1	2.52	0.37	91.8	2.64	0.59	98.6	2.67	0.49
tagatose	67.6	2.15	0.32	70.9	2.31	0.54	73.2	2.31	0.51
tartaric acid	0	1.40	0.33	0.9	1.50	0.24	2.8	1.54	0.35
threitol	61.8	2.07	0.28	58.2	2.04	0.26	64.8	2.14	0.40
threonic acid	2.9	1.45	0.34	4.5	1.51	0.37	7.0	1.62	0.43

threonine	52.9	2.07	0.35	62.7	2.22	0.50	62.0	2.25	0.45
thymine	0	1.49	0.26	0.9	1.44	0.30	1.4	1.42	0.29
trehalose	97.1	2.42	0.25	98.2	2.46	0.24	94.4	2.41	0.25
triethanolamine	58.8	2.10	0.28	77.3	2.25	0.35	67.6	2.12	0.26
tryptophan	64.7	2.08	0.23	56.4	2.14	0.35	70.4	2.18	0.30
tyrosine	94.1	2.31	0.25	90	2.45	0.47	93.0	2.44	0.41
UDP-glucuronic acid	38.2	1.78	0.41	54.5	1.86	0.46	70.4	2.06	0.42
uracil	14.7	1.74	0.32	14.5	1.72	0.36	26.8	1.81	0.39
urea	97.1	3.12	0.44	99.1	3.35	0.55	100	3.32	0.57
uric acid	0	1.42	0.29	4.5	1.56	0.30	5.6	1.52	0.46
uridine	2.9	1.54	0.25	0	1.52	0.26	0	1.51	0.22
urocanic acid	2.9	1.55	0.28	9.1	1.65	0.41	7.0	1.68	0.25
valine	97.1	2.87	0.31	99.1	3.06	0.47	100	3.08	0.38
vanillic acid	8.8	1.54	0.30	10	1.55	0.39	21.1	1.65	0.34
xanthine	0	1.45	0.19	1.8	1.45	0.28	1.4	1.47	0.29
xanthosine	0	1.42	0.37	3.6	1.51	0.32	4.2	1.50	0.35
xylitol	58.8	2.01	0.12	52.7	2.06	0.26	50.7	2.05	0.27
xylonic acid	0	1.40	0.26	0.9	1.38	0.27	1.4	1.41	0.31
xylonic acid isomer	2.9	1.59	0.18	2.7	1.63	0.26	4.2	1.68	0.26
xylose	100	2.23	0.11	97.3	2.25	0.18	98.6	2.29	0.29
xylulose	0	1.55	0.20	0	1.49	0.19	2.8	1.55	0.32

Table 8-2: List of all detected metabolites in EBC, percentage of samples in which they were detected and mean and standard deviation of each metabolite by group (data log$_{10}$ transformed).

Metabolite	Retention Index	m/z	PubChem ID	No. of samples	Percentage of samples
1,2,4-benzenetriol	521803	239	10787	291	100
1,2-anhydromyoinositol	651472	318	119054	291	100
1-hexadecanol	679596	299	2682	287	98.6
1-kestose	1123027	361	440080	256	88.0
1-methyladenosine	829921	259	27476	287	98.6
1-methylinosine	829921	259	27476	201	69.1
1-monostearin	959214	203	24699	291	100
2,3-dihydroxybutanoic acid	384796	292	250402	291	100
2,8-dihydroxyquinoline	626989	290	97250	291	100
2-aminophenol	438445	150	NA	282	96.9
2-deoxytetronic acid	433456	189	150929	291	100
2-hydroxy-2-methylbutanoic acid	264833	145	95433	290	99.7
2-hydroxyglutaric acid	506306	247	43	291	100
2-hydroxyhippuric acid	725465	206	10253	291	100
2-hydroxypyrazinyl-2-propenoicacidethylester	493127	121	5371086	291	100
2-hydroxyvaleric acid	309587	131	98009	291	100
2-isopropylmalic acid	508690	275	5280523	203	69.8
2-ketoisocaproic acid	290473	89	70	291	100
2-methylglyceric acid	372491	219	560781	290	99.7
2-monopalmitin	890356	129	123409	291	100
2-picolinic acid	383668	180	1018	291	100
3,3-hydroxyphenyl-3-hydroxypropionic acid	632357	267	102959	291	100
3,3-hydroxyphenylpropionic acid	583925	192	91	290	99.7
3,4-dihydroxybenzoic acid	620200	193	72	291	100
3,4-dihydroxycinnamic acid	748847	219	689043	290	99.7

3,4-dihydroxyhydrocinnamic acid	673176	179	348154	291	100
3,4-dihydroxyphenylacetic acid	625046	179	547	291	100
3,6-anhydro-D-galactose	588886	231	16069996	291	100
3-aminoisobutyric acid	452655	248	64956	291	100
3-hydroxy-3,4-hydroxy-3-methoxyphenylpropionic acid	688753	297	NA	288	99.0
3-hydroxy-3-methylglutaric acid	521554	247	1662	291	100
3-hydroxyanthralinic acid	640146	354	NA	291	100
3-hydroxyphenylacetic acid	527648	164	12122	291	100
3-hydroxypropionic acid	269265	177	68152	291	100
3-phosphoglycerate	610734	227	724	282	96.9
4-hydroxybenzoate	537925	223	135	291	100
4-hydroxyhippuric acid	784581	294	151012	291	100
4-hydroxyphenylacetic acid	542795	179	127	291	100
4-methylcatechol	416586	268	9958	285	97.9
4-pyridoxic acid	673225	309	6723	290	99.7
5-aminovaleric acid	536657	174	138	291	100
5-deoxy-5-methylthioadenosine	967036	236	439176	291	100
5-hydroxy-3-indoleacetic acid	777606	290	1826	291	100
5-hydroxymethyl-2-furoic acid	497561	123	80642	290	99.7
6-deoxyglucitol	596111	319	151266	290	99.7
7-methylguanine	768706	294	11361	291	100
aconitic acid	586815	229	643757	291	100
adenine	646534	264	190	291	100
adenosine	918039	236	60961	291	100
adipic acid	474435	111	196	291	100
alanine	244189	116	5950	291	100
allantoic acid	726050	259	203	290	99.7

alloxanoic acid	785329	331	94146	189	64.9
alpha-ketoglutarate	507392	198	51	291	100
aminomalonate	455754	218	100714	291	100
anthranilic acid	530297	266	NA	287	98.6
arabitol	572730	103	94154	291	100
arachidic acid	856421	117	10467	291	100
ascorbic acid	672898	332	54670067	290	99.7
asparagine	553743	231	6267	291	100
aspartic acid	480387	232	5960	291	100
azelaic acid	610551	317	19347555	290	99.7
benzoic acid	339067	179	243	291	100
beta-alanine	435564	248	239	291	100
beta-gentiobiose	973116	204	441422	291	100
beta-mannosylglycerate	774364	204	5460194	231	79.4
biphenyl	426625	154	7095	285	97.9
butane-2,3-diol	205778	117	262	291	100
butyrolactam	277199	142	12025	291	100
capric acid	452386	229	2969	291	100
caprylic acid	343457	201	379	291	100
catechol	376695	254	289	291	100
cellobiose	932179	204	6255	291	100
ceratinic acid	1033286	145	10469	250	85.9
cholesterol	1078536	129	5997	290	99.7
citramalic acid	456203	247	1081	291	100
citric acid	617342	273	311	291	100
citrulline	621404	157	9750	291	100
conduritol-beta-epoxide	675635	318	9989541	220	75.6

creatinine	502599	115	588	291	100
cystathionine	772979	218	439258	239	82.1
cysteine	500158	220	5862	291	100
cysteine-glycine	715335	220	439498	289	99.3
cystine	804619	218	595	291	100
dehydroabietic acid	850374	239	94391	207	71.1
dehydroascorbic acid	633423	173	440667	291	100
deoxypentitol	528774	231	270738	291	100
digalacturonic acid	950338	233	439694	222	76.3
digitoxose	521798	117	94168	291	100
diglycerol	591074	103	42953	291	100
dihydro-3-coumaric acid	582960	192	91	291	100
docosenoic acid	911928	129	6433893	291	100
dodecanol	507619	243	8193	291	100
enolpyruvate	234394	217	1005	286	98.3
erythritol	471922	217	222285	291	100
erythronic acid lactone	407495	247	5325915	286	98.3
erythrose major	443306	205	439574	185	63.6
ethanolamine	342561	174	700	291	100
ferulic acid	732779	338	445858	273	93.8
fructose	639442	307	439709	291	100
fucose	578299	160	439650	291	100
fumaric acid	390016	245	444972	291	100
furoylglycine	553990	95	21863	291	100
galactinol	1015529	204	NA	291	100
galactitol	669079	319	5460044	291	100
galactonic acid	690882	292	128869	291	100

galactose	648756	319	439357	291	100
glucoheptulose	828606	217	5459879	290	99.7
gluconic acid	693148	333	6857417	291	100
gluconic acid lactone	645815	220	7027	291	100
glucose	659798	319	64689	291	100
glucuronic acid	665901	333	94715	291	100
glutamic acid	529100	246	33032	291	100
glutamine	600000	156	5961	291	100
glutaric acid	421596	261	743	291	100
glycerol	344466	205	753	291	100
glycerol-3-galactoside	805227	204	16048618	291	100
glycerol-alpha-phosphate	590747	357	754	291	100
glycine	368707	248	750	291	100
glycolic acid	227636	177	757	291	100
glycyl proline	691357	174	3013625	291	100
guanidinosuccinate	699521	444	NA	282	96.9
guanine	744307	352	764	287	98.6
heptadecanoic acid	751309	117	10465	291	100
hippuric acid	638579	206	NA	291	100
histidine	663790	154	6274	291	100
homocystine	874865	128	10010	289	99.3
homovanillic acid	601084	326	1738	291	100
hydroxycarbamate	325318	278	16639161	286	98.3
hydroxyproline dipeptide	879596	156	61159526	289	99.3
hypoxanthine	619128	265	790	291	100
indole-3-acetate	684929	202	802	291	100
indole-3-lactate	764586	202	92904	291	100

indoxyl sulfate	577333	277	10258	291	100
isocitric acid	617338	245	5318532	291	100
isoleucine	359251	158	6306	291	100
isomaltose	983199	160	439193	231	79.4
isopropylbenzene	240619	105	7406	291	100
isothreonic acid	489385	292	151152	291	100
itaconic acid	386511	147	811	291	100
kynurenic acid	726186	231	3845	291	100
kynurenine	769709	218	25245862	277	95.2
lactic acid	217657	191	612	291	100
lactose	935640	191	440995	283	97.3
lactulose	929908	204	11333	291	100
lauric acid	547906	117	3893	291	100
leucine	346357	158	6106	291	100
levoglucosan	569637	204	2724705	291	100
levoinositol	651238	432	NA	291	100
lysine	663483	317	5962	291	100
maleimide	245118	154	10935	291	100
malic acid	463180	233	525	290	99.7
maltose-1	946601	204	439186	291	100
mannose	645856	205	18950	291	100
metanephrine	621765	297	21100	291	100
methanolphosphate	289520	241	13130	162	55.7
methionine	483560	176	6137	290	99.7
methylmaleic acid	418804	259	643798	282	96.9
montanic acid	1087377	117	10470	250	85.9
myo-inositol	730022	305	892	291	100

myristic acid	634414	285	11005	291	100
N-acetylaspartic acid	548028	158	65065	291	100
N-acetylmannosamine	722897	319	439281	291	100
N-acetylputrescine	595523	174	122356	291	100
N-carbamoylaspartate	611345	257	93072	291	100
N-carbamylglutamate	651275	257	121396	291	100
n-epsilon-trimethyllysine	512366	118	440121	291	100
nicotinic acid	366992	180	938	290	99.7
N-methylglutamic acid	455629	98	439377	291	100
nonadecanoic acid	822782	117	12591	291	100
noradrenaline	754841	174	439260	291	100
octadecanol	755409	327	8221	291	100
oleamide	849710	144	5283387	211	72.5
oleic acid	781527	339	445639	271	93.1
ornithine	619196	142	88747248	291	100
orotic acid	586317	254	967	291	100
oxoproline	485935	156	7405	291	100
palatinitol	996670	204	88735	269	92.4
parabanic acid	464991	100	67126	288	99.0
p-cresol	280360	165	2879	291	100
pentitol	563801	307	827	291	100
pentose	540818	103	229	291	100
phenol	218927	151	996	291	100
phenylalanine	537804	192	6140	291	100
phosphate	361492	314	1004	289	99.3
pimelic acid	523205	155	385	291	100
pinitol	622466	260	164619	234	80.4

proline	364716	142	145742	291	100
pseudo uridine	813899	217	15047	291	100
psicose	635244	307	NA	291	100
p-tolyl glucuronide	847531	180	154035	291	100
putrescine	588119	174	1045	290	99.7
pyrogallol	495011	239	1057	291	100
pyrophosphate	327517	110	1023	291	100
pyruvic acid	213805	174	1060	291	100
quinic acid	634900	345	6508	291	100
quinolinic acid	581638	296	1066	290	99.7
raffinose	1120886	361	439242	230	79.0
ribitol	575497	217	827	291	100
ribonic acid	599680	292	5460677	291	100
ribose	553071	217	10975657	291	100
saccharic acid	699211	333	33037	291	100
salicylaldehyde	405583	193	6998	284	97.6
salicylic acid	480699	267	338	286	98.3
serine	395020	204	5951	291	100
serotonin	863824	174	5202	289	99.3
shikimic acid	611100	204	8742	291	100
sinapinic acid	788416	338	637775	150	51.5
sophorose	959716	319	NA	286	98.3
sorbitol	667922	217	5780	291	100
succinic acid	370608	247	1110	291	100
sucrose	915139	271	5988	291	100
tagatose	631835	307	439312	291	100
tartaric acid	534291	292	444305	289	99.3

threitol	467595	217	169019	291	100
threonic acid	497572	292	5460407	291	100
threonine	409568	218	6288	291	100
thymine	420133	255	1135	291	100
trehalose	948197	191	7427	290	99.7
triethanolamine	531892	262	7618	202	69.4
tryptophan	774603	130	6305	291	100
tyrosine	671252	218	6057	291	100
UDP-glucuronic acid	585473	217	17473	291	100
uracil	385735	241	1174	291	100
urea	323728	189	1176	291	100
uric acid	730691	441	1175	291	100
uridine	861508	217	6029	263	90.4
urocanic acid	699866	267	736715	291	100
valine	313502	144	6287	291	100
vanillic acid	597845	297	8468	291	100
xanthine	701688	353	1188	291	100
xanthosine	926133	325	64959	258	88.7
xanthurenic acid	795062	406	5699	290	99.7
xylitol	567437	217	6912	291	100
xylonic acid	589278	333	6602431	290	99.7
xylonic acid isomer	590775	189	10264	290	99.7
xylose	544100	103	135191	291	100
xylulose	553450	173	439205	291	100

Table 8-3: All identified urinary detected metabolites, and the number and percentage of samples in which they were detected.