Deep Multi-Omic Profiling Reveals Molecular Signatures that Underpin Preschool Wheeze and Asthma

Matthew Macowan, BSc (Hons), Céline Pattaroni, MSc, Katie Bonner, MD, Roxanne Chatzis, BSc, Carmel Daunt, BSc (Hons), Mindy Gore, PhD, Adnan Custovic, MD, PhD, Michael D. Shields, PhD, Ultan F. Power, PhD, Jonathan Grigg, MD, Graham Roberts, DM, Peter Ghazal, PhD, Jürgen Schwarze, MD, Steve Turner, MD, Andrew Bush, MD, Sejal Saglani, MD, Clare M. Lloyd, PhD, Benjamin J. Marsland, PhD

PII: S0091-6749(24)00869-8

DOI: https://doi.org/10.1016/j.jaci.2024.08.017

Reference: YMAI 16468

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 19 June 2024

Revised Date: 16 August 2024

Accepted Date: 23 August 2024

Please cite this article as: Macowan M, Pattaroni C, Bonner K, Chatzis R, Daunt C, Gore M, Custovic A, Shields MD, Power UF, Grigg J, Roberts G, Ghazal P, Schwarze J, Turner S, Bush A, Saglani S, Lloyd CM, Marsland BJ, Deep Multi-Omic Profiling Reveals Molecular Signatures that Underpin Preschool Wheeze and Asthma, *Journal of Allergy and Clinical Immunology* (2024), doi: https://doi.org/10.1016/j.jaci.2024.08.017.

Deep Multi-Omic Profiling Reveals Molecular Signatures that Underpin Preschool Wheeze and Asthma

Matthew Macowan, BSc (Hons)*[1]; Céline Pattaroni, MSc*[1]; Katie Bonner, MD[2]; Roxanne Chatzis, BSc[1]; Carmel Daunt, BSc (Hons)[1]; Mindy Gore, PhD[2]; Adnan Custovic, MD, PhD[2,3]; Michael D Shields, PhD[3]; Ultan F. Power, PhD[3]; Jonathan Grigg, MD[4]; Graham Roberts, DM[5,6,7]; Peter Ghazal, PhD[8]; Jürgen Schwarze, MD[9]; Steve Turner, MD[10,11]; Andrew Bush, MD[2]; Sejal Saglani, MD[12,13]; Clare M. Lloyd, PhD[13]; Benjamin J Marsland, PhD[1]

* These authors contributed equally.

[1]Department of Immunology, School of Translational Medicine, Monash University, Melbourne, Australia.

[2]Imperial Centre for Paediatrics and Child Health, and National Heart and Lung Institute, Imperial College London, United Kingdom.

[3]Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, United Kingdom.

[4]Centre for Child Health, Blizard Institute, Queen Mary University of London, United Kingdom.

[5]Human Development in Health School, University of Southampton Faculty of Medicine, Southampton, United Kingdom.

[6]NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom.

[7]David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Newport, Isle of Wight, United Kingdom.

[8]School of Medicine, Systems Immunity Research Institute, Cardiff University, Cardiff, United Kingdom.

[9]Centre for Inflammation Research, Child Life and Health, The University of Edinburgh, Edinburgh, United Kingdom.

[10]Child Health, University of Aberdeen, Aberdeen, United Kingdom.

[11]NHS Grampian, Aberdeen, United Kingdom.

[12]Royal Brompton Hospital, London, United Kingdom.

[13]National Heart & Lung Institute, Imperial College London, United Kingdom.

Corresponding author :

Céline Pattaroni
89 Commercial road
3004 Melbourne
Australia
Email : celine.pattaroni@monash.edu

Sources of funding :

The study is supported by a Wellcome Trust grant (108818), a L.E.W. Carty Charitable Fund awarded to C.P., funding from The Hospital Research Foundation Group (2022-SF-EOI-001) awarded to B.J.M and C.P., and a NHMRC Senior Research Fellowship (1154344) awarded to B.J.M.

Authors contributions :

M.M. and C.P. contributed equally to the study. C.P. led the research, while M.M. conducted data analysis, created figures, and co-wrote the manuscript. The conceptualization team included A.C., M.D.S., U.F.P., J.G., G.R., P.G., J.S., S.T., A.B., S.S., C.M.L., and B.J.M. Project administration was overseen by M.G. The investigation was carried out by M.M., C.P., R.C., and C.D., with samples provided by K.B. and S.S. The original draft was written by M.M. and C.P., with reviews and edits contributed by A.C., M.D.S., U.F.P., J.G., G.R., P.G., S.T., A.B., S.S., C.M.L., and B.J.M. All authors read and approved the manuscript.

Conflict of interest :

The authors declare no conflict of interest.

ABSTRACT

Background : Wheezing in childhood is prevalent, with over half of all children experiencing at least one episode by age six. The pathophysiology of wheeze, especially why some children develop asthma while others do not, remains unclear.

Objective : This study addresses the knowledge gap by investigating the transition from preschool wheeze to asthma using multi-omic profiling.

Methods : Unsupervised, group-agnostic integrative multi-omic factor analysis was performed using host/bacterial (meta-)transcriptomic and bacterial shotgun metagenomic datasets from bronchial brush samples paired with metabolomic/lipidomic data from bronchoalveolar lavage samples acquired from children 1-17 years old.

Results : Two multi-omic factors were identified: one characterising preschool-aged recurrent wheeze and another capturing an inferred trajectory from health to wheeze and school-aged asthma. Recurrent wheeze was driven by Type 1-immune signatures, coupled with upregulation of immune-related and neutrophil-associated lipids and metabolites. Comparatively, progression towards asthma from ages 1-18 was dominated by changes related to airway epithelial cell gene expression, Type 2-immune responses, and constituents of the airway microbiome, such as increased *Haemophilus influenzae*.

Conclusion : These factors highlighted distinctions between an inflammation-related phenotype in preschool wheeze, and the predominance of airway epithelial-related changes linked with the inferred trajectory toward asthma. These findings provide insights into the differential mechanisms driving the progression from wheeze to asthma and may inform targeted therapeutic strategies.

Clinical implications :

Multi-omic profiling identifies molecular signatures aiding early prediction and differentiation of preschool wheeze from progression to asthma.

Capsule summary :

This study identifies distinct multi-omic factors from lower airway samples, revealing Type 1-immune signatures for preschool wheeze and epithelial signatures with microbiome changes for asthma progression, offering insights into disease trajectories.

Key words : wheeze, asthma, multi-omics, gene expression, metagenomics, metabolomics, lipidomics, disease trajectory

Abbreviations : ALI (Acute Lung Injury), BAL (Bronchoalveolar Lavage), BH (Benjamini-Hochberg), CLR (centred log-ratio), COPD (Chronic Obstructive Pulmonary Disease), DCs (Dendritic Cells), DE / DA (Differential Expression / Differential Abundance), HLCA (Human Lung Cell Atlas), HMDB (Human Metabolome Database), KO (KEGG ortholog), LCMS (Liquid Chromatography-Mass Spectrometry), LF (Latent Factor), NA (Not Applicable), PCA (Principal Component Analysis), QC (Quality Control), ROS (Reactive Oxygen Species), scRNAseq (Single-cell RNA sequencing)

INTRODUCTION

Wheezing in childhood is remarkably prevalent, with over half of all children having at least one episode by 6 years of age (1). Due to disease heterogeneity and challenges associated with obtaining meaningful physiological parameters, such as lung function, in preschool-aged children, there exists a knowledge gap in the pathophysiology of wheeze. The reason some children with preschool wheeze continue to wheeze and develop asthma while others do not, even with comparable initial clinical presentation, remains an unanswered question. Indeed, more than 50% of children with recurrent preschool wheeze will have complete symptom resolution by school age and not develop asthma (2,3).

Multiple epidemiological studies have shown that atopy (high total IgE levels with the presence of elevated aero-allergen-specific IgE) in recurrent preschool wheezers is linked with asthma (3–5). In line with this, it is clear that Type 2 immune responses and airway remodelling are present in school-aged asthma (6). However, this is less clear in those with recurrent wheeze, and while some preschool children exhibit evidence of airway eosinophilia, a subgroup have a predominance of lower airway neutrophilia (7). In addition, it is unclear whether these children exhibit variations in their baseline immunological tone under steady-state conditions when acute symptoms are absent. Persistent underlying differences in cellular and molecular factors could represent important early therapeutic intervention targets and point to potential mechanisms driving disease.

There is evidence of lower airway microbial dysbiosis in both preschool children with recurrent wheeze and older children with asthma, and that dysbiosis may precede disease symptoms, sometimes by many years (8–10). Advances in multi-omics technologies have revolutionised respiratory research (11). In particular, the utilisation of shotgun metagenomics and metatranscriptomics facilitate precise characterisation of microbes, and allow interrogation of whether certain species are simply bystanders or whether they are transcriptionally active.

While some studies have performed multi-omic integration in the context of asthma (12), to our knowledge and as highlighted in a recent review (13), no study has simultaneously investigated the paired lower airway transcriptome, metabolome/lipidome, and microbiome

in the lower airways. Parallel investigation of the respiratory microbiome and circulating metabolome as performed in previous studies provides valuable insights, however it may overlook significant changes linked with disease. Exploration into the interplay between host and microbes in the lower airways has the potential to significantly advance our understanding of early-life asthma progression and pathophysiology.

To address these gaps in the field, we investigated the transition from preschool wheeze to asthma by using deep multi-omic profiling of lower airway samples from preschool wheezers, school-aged asthmatics, and preschool controls in the absence of acute symptoms. Data integration was conducted unsupervised, enabling impartial post-hoc correlation with clinical groups. Despite the inability to track individual patients over time due to the cross-sectional nature of our samples, we were able to investigate an inferred trajectory through a multi-omic bioinformatic approach. We define disease progression and trajectory throughout this manuscript as having a higher score for the multi-omic signature that captured the dynamic continuum from healthy preschool children, through to recurrent preschool wheeze and school-aged asthma. We identified two signatures: one related to an inferred trajectory toward childhood asthma characterised by Type 2 immune-related changes in the airway epithelium and presence of transcriptionally-active pathogens, and one specific to preschool wheeze distinguished by neutrophilic, Type 1 immune-associated signals and changes in local metabolite and lipid abundance. The modalities within each signature warrant further consideration as targets for future disease prevention and management strategies.

METHODS

Participants and sample collection

In order to investigate the progression from recurrent wheeze to asthma, three groups from the Breathing Together cohort were investigated (14). Our study examined male and female participants, and similar findings are reported for both sexes. Informed parental consent was obtained prior to inclusion in the study and ethical approval was obtained from the Regional Ethics Committee (REC reference: 16/LO/1518). **Preschool controls.** The first group consisted of preschool controls, aged 12 months to five years, with no history of wheezing, or other significant respiratory illness. Since lower airway sampling is not ethically justifiable in children without respiratory symptoms, lower airway research samples were obtained following informed parental consent, from children undergoing elective surgery that required clinically indicated intubation under general anaesthesia. Exclusion criteria for this group included recent respiratory tract illness (symptoms resolved for less than two weeks), other chronic respiratory diseases (e.g. cystic fibrosis), and a history of previous pituitary or ethmoid surgery. Four control patients were prescribed inhaled corticosteroids and six were prescribed bronchodilators in response to a prior isolated wheeze event, and were considered preschool controls for this study given the wheeze was not recurrent. A cytology brush was passed through the endotracheal tube and used to obtain lower airway epithelial cells as previously described (15). **Recurrent severe preschool wheeze**. The second group consisted of severe preschool wheezers, aged 12 months to five years, with a history of recurrent, severe wheezing and difficulty breathing, and scheduled for clinically indicated bronchoscopy. **Severe school-age asthma**. The third group included school-aged asthmatics, aged six to 18 years, with a confirmed diagnosis of severe asthma defined as evidence of airway hyper-responsiveness, bronchodilator reversibility (≥12%), or peak expiratory flow variability (≥10%). Both the preschool severe wheezers and school-age children with severe asthma were undergoing a clinically indicated bronchoscopy and additional lower airway samples were collected for research following informed parental consent and child assent, as previously described (9,16). For shotgun metagenomics, bronchial brushes were obtained using Copan eSwabs (COPAN Diagnostics), rotated five times, and subsequently stored at -80°C. Bronchial brushes, used for host transcriptomics and metatranscriptomics, were rotated three times to collect endobronchial cells, then immediately placed into an Eppendorf

with 700 µl of RLT lysis buffer (Qiagen) and 2-Mercaptoethanol (Sigma), followed by snap-freezing and storage at -80°C. Bronchoalveolar lavage (BAL) was performed from the right middle lobe with three aliquots of 1 mL/kg of 0.9% sodium chloride and BAL supernatants collected for metabolomics/lipidomics were stored at -80°C. Differential cell counts were performed on BAL for red blood cells and immune cell subsets. **Atopic status**. All children had assessments of total IgE and specific IgE to common aero-allergens. A total serum IgE level threshold of 30 IU/mL was used as previously described for associations with atopy in line with prescribing information for omalizumab, an anti-IgE monoclonal antibody indicated for asthma treatment (17). Atopy was therefore defined as a total serum IgE level >30 IU/mL, combined with the presence of at least 1 positive aero-allergen-specific IgE above the detection limit.

Host/bacterial RNA extraction, library preparation and sequencing

RNA was extracted using the Quick-RNA Microprep kit (Zymo Research) according to the manufacturer's protocol and eluted in 40µl of RNAse-free water. Following extraction, libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit (New England Biolabs) for Illumina following ribosomal RNA removal using the Ribo-Zero Magnetic Kit (Epicentre). Resulting libraries were sequenced on an Illumina NovaSeq 6000 platform using a S4 2x75 kit.

Host RNA sequencing data processing Raw RNA sequencing FASTQ files were processed using the NF-CORE rnaseq pipeline (version 3.10.1) (18). Briefly, reads underwent initial quality control (QC) with FastQC, were extracted with UMI-tools, adapters removed, and quality trimming performed with Trim Galore. Genomic contaminants and ribosomal components were removed with BBSplit and SortMeRNA respectively. Reads were aligned and quantified via a combination of STAR and Salmon tools, and sorted and indexed using SAMtools, and dereplicated with UMI-tools. An abundance matrix was produced, representing scaled and normalised transcripts per million reads, from which a re-estimated counts table was generated for downstream analysis. Gene data was filtered for a minimum of 100 reads per sampling group using the edgeR package (version 3.42.0) (19) filterByExpr function, followed by voom normalisation with limma (version 3.56.0) (20), and removal of non-protein-coding genes. Voom-normalised data matrix was saved for multi-omic integration.

Bacterial DNA extraction, library preparation and sequencing

Bronchial brush swabs were added to 1mL microbial-free DNA-free water (Qiagen) and centrifuged at 14,000xg for 10 min at 4°C. The pellet was resuspended and incubated with 300U of Lyticase (Sigma) for 30 min at 37°C with gentle shaking (500rpm) to enhance DNA recovery, as previously described (21). DNA was extracted from the resulting lysates using the DNeasy UltraClean Microbial Kit (Qiagen) according to the manufacturer's protocol and eluted in 40µL of microbial-free DNA-free water. All extraction steps were performed in microbial DNA-free conditions in a laminar flow hood, decontaminated and UV-treated prior to processing. Negative controls (ESwabs opened and closed at sampling site, extraction negative controls and PCR negative controls) were processed along samples. DNA was prepared for shotgun sequencing using the Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's protocol, with multiplexing via the IDT for Illumina Nextera DNA Unique Dual Indexes Set C (Illumina). Library pools were sequenced with paired-ends on a NovaSeq 6000 platform using a SP 2x250 kit.

Bacterial shotgun metagenomic sequencing data processing

Raw shotgun sequencing reads were processed as previously described using the Sunbeam pipeline for adaptor trimming, quality control, host genome decontamination, assembly of contiguous sequences, and co-assembly (22,23). The co-assembly was reformatted and filtered for sequences >500 base pairs using Anvi'o (version 7.1) (24). Contig taxonomy was estimated using Kraken 2 (25). Individual FASTQ files were aligned and mapped to the co-assembly using bowtie2 (26), per-sample contig counts generated using a custom bash script, then combined into a counts matrix in R. To decontaminate the contigs, any sequence found in the extraction negative controls was removed (Figure S1). Using the ratio of reads before and after decontamination, samples with >90% non-contaminant reads were retained. A decontaminated co-assembly FASTQ file was generated from the remaining contigs. Functional KEGG orthology (KO) was assigned using the GhostKOALA tool provided by the Kyoto Encyclopedia of Genes and Genomes (27), then agglomerated using taxonomy to produce taxa-gene pairs with their associated counts. Using the R microbiome package (version 1.22.0) (28), the dataset was filtered using a detection threshold of 1 in at least 10% of samples via the core function, and centred log-ratio (CLR)-normalised via the transform function. The final dataset was saved as a matrix for multi-omic data integration.

Bacterial metatranscriptomic sequencing data processing

Raw RNA sequencing reads were processed using the Sunbeam pipeline as above, which removed host reads from the dataset. The resulting FASTQ files were mapped to the decontaminated metagenomic co-assembly using bowtie2, and a corresponding counts matrix generated for multi-omics integration.

Metabolomic and lipidomic sample processing

100µL of BAL supernatant was used for concurrent metabolomics and lipidomics extraction. 400µL of extraction solvent (3:1 mixture of methanol and chloroform), supplemented with 0.5µM of generic internal standards (CHAPS, CAPS, and PIPES) and 5µM 2,6-di-tert-butyl-4-methylphenol (BHT), was added to the supernatant. Samples were shaken at 1,000rpm for one hour at 4°C, followed by centrifugation at 14,000xg for 10 min. Supernatants were evaporated using a Speedvac (Thermo Scientific) and re-solubilised in 100µL of a 3:1:1 solution of methanol, chloroform, and water for metabolomics or a 4.5:4.5:1 solution of butanol, methanol, and water for lipidomics. Samples were sonicated for 15 min on ice, centrifuged at 14,000xg for 10 min, and the supernatant collected. 10µL of each supernatant was mixed for quality control testing, and blanks were prepared as the solvent solutions alone.

Liquid chromatography-mass spectrometry (LCMS) data acquisition

LCMS were acquired on a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher) coupled with the Dionex Ultimate® 3000 RS (Thermo Fisher) high-performance liquid chromatography (HPLC) system. Chromatographic separation was performed on a ZIC-pHILIC column (5µm, polymeric, 150 x 4.6mm, SeQuant®, Merck). The mobile phase was 20mM ammonium carbonate and acetonitrile. The gradient program started at 80% and was reduced to 50% over 15min, then reduced further to 5% over 3min, followed by an 8min re-equilibration to 80%. Flow rate was 0.3mL/min and column compartment temperature was 40°C. Total run time was 32min with an injection sample volume of 10µL. The mass spectrometer operated with positive and negative polarity switching at 35,000 resolution at 200m/z, with a detection range of 85 – 1,275m/z in full scan mode. Electro-spray ionisation source (ESI) was set to 3.5kV for positive and 4.0kV for negative modes, sheath gas and auxiliary gas set to 50 and 20 arbitrary units respectively, capillary temperature set to 300°C, and probe heater

temperature set to 120°C. Samples were analysed in a single batch and randomised to account for LCMS system drift over time.

Metabolomics and Lipidomics LCMS data processing

Raw LCMS data were processed using the metabolome-lipidome-MSDIAL pipeline (see code and data availability section), incorporating MS-DIAL (version 5.1) (29), the human metabolome database (HMDB, version 4 - July 2021) (30), and the pmp (version 1.4.0) R package (31). MS-DIAL was used for deconvolution and peak detection with a minimum peak amplitude of 100,000. Peaks were identified using the MassBank database (version 2021.08) (32) with a retention time tolerance of 2min and accurate mass tolerance (AMT) of 0.002Da. Peaks were aligned using a retention time tolerance of 2min and AMT of 0.002Da, with gap-filling by compulsion. Peak intensity tables from MS-DIAL were imported into R, peaks filtered for intensities >5-fold higher than LCMS blanks, samples with >80% missing values and features with >20% missing values removed, and peaks filtered based on the percentage of variation in the QC samples with a maximum relative standard deviation of 25%. Data was normalised using probabilistic quotient normalisation (PQN), followed by random forest missing data imputation (using the pmp wrapper function of the missForest (version 1.4) (33) function), and subsequent generalised logarithmic (glog) transformation to stabilise variance across low and high intensity mass spectral features. MS1 data was further mapped to HMDB, with an AMT of 0.002Da, to increase the number of annotated features. Any unannotated features were filtered out, and the remaining dataset was subjected to two rounds of manual feature curation in MS-DIAL: firstly to identify and remove any poor-quality spectral features not filtered out during pre-processing with pmp, and secondly to select the best-quality spectrum-metabolite pairs using a combination of the MS-DIAL fill percentage and signal-to-noise ratio values, along with visual confirmation. The final dataset was saved in matrix format for multi-omic integration.

Statistical analysis and multi-omic data integration

Statistical analyses were performed in R (version 4.3.0) (34) and plots generated using ggplot2 (version 3.4.2) (35). A summary schematic for the multi-omic integration strategy is provided in Figure S2A. The three input matrices were assembled into a MultiAssayExperiment object (36). Data was integrated in an unsupervised, group-agnostic manner using Multi-Omics Factor Analysis v2 (MOFA+) (37), with slow convergence, a seed value of 2 for generation of

pseudo-random numbers, 5% minimum explained variance threshold for factor retention, and sampling age used as a covariate. An overview of the unsupervised multi-modal data integration is shown in Figure S2B, and we refer the reader to the original manuscript for technical details and mathematical derivations. As model inputs, we supplied voom-normalised host transcriptomics data with a read count threshold of 100, centred log-ratio-normalised metagenomic KEGG ortholog count data with a detection threshold of 1 in >10% of samples, and a combined small molecules dataset. Wilcoxon rank sum tests were used to assess categorical differences in the resulting latent factors, and Spearman correlation analyses were used to assess continuous variables. Statistical significance of clinical group Wheeze vs. Disease Score clustering was assessed by permutational multivariate analysis of variance (PERMANOVA) using the adonis2 function from the vegan R package (version 2.6–4) (38). Factors that showed significant differences between patient groups were used as predictors for age-corrected linear modelling on the individual MOFA-imputed datasets using a custom script built around limma (see code and data availability section). Benjamini-Hochberg multiple testing correction was used with a significance threshold of 0.05. Log fold-change threshold was 0.5 for host RNA and small molecules, and 0.25 for bacterial metagenomics. To summarise bacterial metagenomic changes following linear modelling, KO functional gene counts were stratified by taxa, filtered for a minimum count of 10, and the log-transformed value of the sum of increased (positive) and decreased (negative) gene counts calculated to determine the net change in gene count abundance per taxa along the MOFA factor. Spearman correlation analyses were performed on metagenomic and metatranscriptomic read counts as a proxy for inferring bacterial transcriptional activity.

Pseudo-bulk analysis

The full integrated human lung cell atlas (39) single-cell RNA sequencing dataset was subsetted to omit cells from individuals with cancer or any smoking history, nasal samples, and cells annotated as ‘native cell’. Gene expression data underwent pseudo-bulk processing by cell type for each study using decoupleR (40). Mode was set to ‘mean’, with a gene read count threshold of 100 and expression in at least 10 cells. The final dataset was batch-corrected with ComBat_seq (41), and contained 766 profiles, covering 50 cell types from 28 studies. For analysis, data was subsetted to include only multi-omic factor differential genes and used for principal component analysis and biplot generation. Labels around the periphery

were added to represent the biplot segment covered by a given cell type from the centre to the extents of 95% confidence ellipses.

Code and data availability

All microbiota and gene expression data have been deposited at the National Centre for Biotechnology Information database under the BioProject accession number PRJNA1080233 with minimal metadata. Comprehensive clinical metadata can be provided upon reasonable request. The data analysis pipeline for this study is described at https://github.com/mucosal-immunology-lab/BT_groups2-4_analysis. The shotgun sequencing pipeline and custom linear modelling script is described at https://github.com/mucosal-immunology-lab/microbiome-analysis/wiki.

RESULTS

Participants and samples

To investigate molecular signatures of wheeze and asthma, bronchial brushes were collected from preschool children with and without recurrent wheeze, and school-aged asthmatics during clinically stable disease; the present cohort did not extend to healthy school-aged children. Bronchoalveolar lavage (BAL) fluid was simultaneously collected from preschool wheezers and school-aged asthmatics. Demographic information for patients with samples that passed quality control criteria is provided in Table 1. Bronchial brush samples were processed for host and microbial transcriptomic sequencing and bacterial metagenomic characterisation. BAL samples were utilised for metabolomic and lipidomic profiling.

Multi-omics factor analysis reveals two sets of cohort-associated covariation patterns

Following quality control, the available datasets were used to build an integrative multi-omic model using multi-omics factor analysis (MOFA+), a statistical framework for integration of multi-omics data. This unsupervised approach integrates different types of biological data, identifies the main sources of variation, and generates a small set of key components, called latent factors. Through this process, each sample is assigned a single numeric score that indicates its position along each of the latent factors. MOFA+ identified four latent factors with a minimum explained variance of 5% (Figure 1A). Latent factor 2 (LF2) and latent factor 4 (LF4) showed significant differences between sample cohorts (Figures 1B-C) and were therefore the focus of further investigation. Although latent factors 1 and 3 contained unique covariation patterns, they were not explained by any available clinical metadata, and as such were not pursued further. LF2 explained 27.1% of overall sample variance, with significant differences between preschool wheezers and school-aged asthmatics ($p = 0.039$) but no difference between preschool controls and school age asthmatics. Importantly, LF2 was not associated with age differences (Spearman rho: -0.16; $p = 0.17$). As LF2 distinguished preschool wheezers from the other groups, it will be referred to as the “Wheeze Score”. Small molecules accounted for 70.4% of factor variation, followed by host transcriptomics (19.5%) and metagenomics (10.1%). LF4 explained 9.4% of sample variance, and was significantly different between all three groups (preschool controls vs. preschool wheezers: $p = 2.3e\text{-}04$; preschool wheezers vs. school-aged asthmatics: $p = 0.011$). As school-aged asthmatic samples

were simultaneously the oldest and had the highest LF4 scores, age was positively correlated with LF4 (Spearman rho: 0.46; p = 4.3e-05). However, variance explained by LF4 was also associated with inferred disease progression and will be referred to as the "Disease Score". Of note, some of the preschool controls were prescribed either inhaled corticosteroids (4/29) or bronchodilators (6/29) for a prior isolated wheezing episode (Figure S3A). In-depth characterisation of these samples revealed no significant impact of treatment on either the Wheeze or Disease Scores (Figure S3B-C). We define disease progression and trajectory here, by multi-omic inference, as an increased Disease Score: a scale representing a continuum from healthy preschool children, through to recurrent preschool wheeze and school-aged asthma. Host transcriptomics and metagenomics accounted for 68.1% and 31.9% of factor variation respectively.

Multi-omic Wheeze Score is characterised by an inflammatory signature

We next aimed to characterise changes explaining an increased Wheeze Score. Using the latent factor score as a continuous predictor for linear modelling of the individual datasets, corrected for age at sampling, significant changes were observed for host transcriptomics and small molecule composition, but not for metagenomic functional gene abundance. The top genes driving an increased Wheeze Score included both *S100A8* and *S100A9* (logFC: 1.983, p=1.45e-03), one of the most abundant cytosolic protein complexes found in neutrophils (42), chemotactic factors for neutrophils and other leukocytes (*CXCL8*, *CXCL2*, *CCL20*, *CSF3*), as well as others affecting neutrophil adhesion and motility (*SELL*, *PLAUR, AQP9*) (Figure 2A). Genes induced by pattern- and damage-associated molecular patterns (PAMPs and DAMPs) were also among the top upregulated (*IDO1*, *IL1B*, *CCL4*, *FPR1-3*, *SOCS3*) in addition to genes induced by type 1 inflammatory cytokines (*DUOX2, DUOXA2, BCL2A1*, *PLEK*). To gain a broader view of which cell types may be most affected by an increased Wheeze Score, we performed a pseudo-bulk analysis on a subset of the integrated human lung cell atlas (HLCA), the most comprehensive single-cell database currently available, capturing the majority of cell types in the airways with the exception of granulocytes. The full HLCA pseudo-bulk dataset was filtered for the 2,134 differentially expressed (DE) genes along the Wheeze Score and then used for principal component analysis (PCA) (Figures 2B and S5A). Using the location of each cell type cluster from the centre of the PCA biplot and the direction of key genes driving variation, it was evident that the majority of upregulated DE genes were more highly

expressed in immune cells than the structural cell compartment, suggesting an immune-driven signature in recurrent preschool wheeze. The Wheeze Score was associated with increased abundance of a large number of metabolites and lipids, except for decreased adenosine and monoethylglycinexylidide levels (Figure 2C). Among the top increased small molecules were Gram-negative bacterial-derived sulfonolipids (43), anti-inflammatory N-acylethanolamines (44), and triglycerides. Dihexosylceramides (Hex2Cer) and ceramides were increased, as were ether-linked phospholipids, arachidonic acid, and docosahexaenoic acid. Increased abundance was further observed for adenosine breakdown product hypoxanthine, and kynurenine metabolite quinolinate. Kynurenine pathway genes facilitating quinolinate production in addition to *IDO1* were also increased, including *KMO* (logFC: 1.435, p=3.71e-03) and *KYNU* (logFC: 1.698, p=8.93e-04) (Figure 2D). Using differential cell counts performed on fresh BAL samples, the Wheeze Score showed strong positive correlation with both neutrophils and leukocytes, and negative correlation with macrophages (Figures 2E and S4). Conversely, the Disease Score was not associated with changes in BAL immune cell composition. In summary, we identified a multi-omic factor specific to children with recurrent preschool wheeze at steady state characterised by enhanced neutrophilic and Type-1 high inflammation and increased abundance of both proinflammatory and immunomodulatory small molecules.

Multi-omic Disease Score is characterised by altered epithelium and bacterial gene counts

Given that the Wheeze Score distinguished preschool wheezers, but was similar between preschool controls and school-aged asthmatics following correction for patient age, we aimed to characterise the trajectory from health to asthma captured by the Disease Score. Linear modelling confirmed that Disease Score variation was explained by host gene expression and differentially-abundant (DA) bacterial functional gene counts. A notable type 2 inflammatory signature was observed among the top genes driving an increased Disease Score (Figure 3A). Clade B serine protease inhibitors (*SERPINB2*, *SERPINB10*, *SERPINB11*, *SERPINB3*), tetraspanins (*TM4SF1*, *UPK1B*), and type 2 inflammation-linked genes (*POSTN*, *LRRC31*) were all increased. Mucin gene expression patterns were also altered, with increased *MUC2* and *MUC5AC* paired with decreased *MUC5B* (logFC: -1.15, p=8e-03). This was associated with increases in mucociliary differentiation- and glycosaminoglycan-related genes (*B3GNT6*, *HS6ST2*, *KRT4*), and those relating to modulation of angiogenesis and smooth muscle function

(*SEMA3B*, *GSN*, *TMEM184A*, *EGFL6*). Lactotransferrin (*LTF*) was negatively associated with Disease Score, as were Type I, III, and VI collagens. Pseudo-bulk analysis of the 357 Disease Score DE genes highlighted that the majority of genes upregulated in disease were associated with airway epithelial and secretory cell populations (Figures 3B and S5B), suggesting that Th2-associated changes to the structural compartment were linked with disease trajectory. To assess microbial associations, differentially-abundant bacterial gene counts were stratified by species and the overall genetic shift determined by averaging gene counts increased and decreased with Disease Score (Figure 3C). Given the low microbial biomass and resulting functional gene sparsity, this method focused on broader metagenomic content rather than individual bacterial gene associations. Positive shifts in *Haemophilus* and *Neisseria* gene count averages were associated with disease progression, while *Streptococcus*, *Prevotella*, and *Veillonella* were negatively associated. To evaluate microbial transcriptional activity, Spearman correlations were performed between metagenomic and metatranscriptomic read counts. *Haemophilus influenzae* was the top Disease Score-associated species and showed the highest correlation (rho: 0.91, $p<0.0001$). Atopic individuals, identified as those with total systemic IgE levels over 30 IU/mL with at least one aero-allergen-specific IgE, had higher Disease Score values compared to non-atopic counterparts (Figure 3D). In summary, we identified a multi-omic signature of progression to asthma associated with type 2 inflammatory-driven changes to the airway epithelium and the presence of transcriptionally-active pathogens, such as *H. influenzae*.

DISCUSSION

In order to address the existing gaps in knowledge surrounding the transition from recurrent wheeze to asthma in childhood, we collected bronchial brushes and bronchoalveolar lavage (BAL) fluid for interrogation of host-microbe transcriptomics, microbiome, and small molecules. Challenges surrounding the invasive nature of this type of sampling has led to a predominance of upper respiratory (nasal and oropharyngeal) sample analysis, particularly in prior early life and childhood studies (45,46). Through deep multi-omic profiling of lower airway samples acquired in the absence of acute symptoms, we identified two multi-omic factors derived from host gene expression and bacterial metagenomics in the bronchi paired with small molecules from BAL fluid; one capturing a signature of preschool wheeze, and the

other characterising a trajectory of disease states from healthy preschool children through to wheeze and Type-2 high asthma.

The identified Wheeze Score exhibited a signature characteristic of neutrophilic and Type 1-driven immune responses. This was supported by the positive correlation between the Wheeze Score and proportions of neutrophils and leukocytes in BAL samples. Pseudo-bulk analysis revealed considerable transcriptional changes in immune cells, particularly among monocytes, macrophages, and dendritic cells (DCs); the absence of granulocytes in the HLCA database is a limitation to this approach. IL-8 (*CXCL8*), MIP-2α (*CXCL2*), and G-CSF (*CSF3*) are all potent chemokines driving neutrophil recruitment (47–51), while CCL20 stimulates rapid recruitment of DCs (52). The associated increase in ether-linked phospholipids supports elevated neutrophil infiltration, as ether-linked phosphatidylcholines constitute nearly half of neutrophil phosphatidylcholines, but are scarce in most tissues (53). Reportedly, the urokinase receptor (*PLAUR*) also plays a role in neutrophil chemotaxis unrelated to its protease activity (54). L-selectin (*SELL*) is expressed constitutively on most leukocytes, and participates in the "rolling stage" motility and adhesion of neutrophils and monocytes, allowing extravasation into inflamed tissue (55,56). Aquaporin 9 (*AQP9*) meanwhile contributes to regulation of lamellipodium formation and neutrophil motility by modulating cellular water flux (57). The upregulation of PAMP- and DAMP-induced genes can further contribute to cellular infiltration upon activation. Formyl peptide receptors (*FPR1-3*) recognise formylated peptides from bacteria and mitochondria, may have roles in viral defence, and have been shown to induce neutrophil and monocyte chemotaxis (58–60).

In addition to genes linked with enhancement of cell motility and recruitment, pro-inflammatory genes and small molecules were associated with an increased Wheeze Score. S100A8/S100A9 heterodimers, secreted primarily by neutrophils and macrophages, have critical roles in modulating leukocyte recruitment and stimulating phagocyte secretion of IL-1β, TNFα, and NF-κB target genes (61). The S100A8/S100A9 complex comprises ~45% of neutrophil cytosolic proteins in neutrophils, and ~1-5% in monocytes/macrophages (42,62). Sulfonolipids are produced by Gram-negative bacteria and elicit IL-1α/β, IL-6, and TNFα in macrophages (43,63). Increased ceramide production is associated with cellular stress (64), and a recent murine model demonstrated their contribution to neutrophil infiltration and reactive oxygen species (ROS) production (65). In relation to ROS production, dual oxidase 2

(*DUOX2*) and its maturation factor (*DUOXA2*) were increased. *DUOX2* is induced by Th1 cytokine IFNγ and viral infection, unlike Th2-induced *DUOX1*, and produces hydrogen peroxide (66).

Conversely, immunomodulatory signatures were also evident, likely to restrict excessive inflammation and tissue damage. Suppressor of cytokine signalling 3 (*SOCS3*) negatively regulates JAK/STAT3 signalling and limits severity of murine acute lung injury (ALI) (67,68). Indoleamine 2,3-dioxygenase (*IDO1*) is induced in antigen-presenting cells (APCs) and epithelial cells by type 1 cytokines (69), and converts tryptophan to downstream kynurenine metabolites. Monocytes and macrophages have the highest activities of kynurenine 3-monooxygenase (*KMO*) and kynureninase (*KYNU*), which drive conversion from kynurenine to quinolinate (70). The end product quinolinate can then be secreted to induce T cell apoptosis and suppress immune responses (71), and has an important role in replenishing cellular NAD+ levels to meet host energy demands in response to increased cellular stress (72). Overall, these data suggest that the lower airways in recurrent paediatric wheeze at baseline are characterised by increased immune cells (neutrophils, monocytes, and DCs), proinflammatory mediators, and pathways that limit excessive inflammatory damage. Importantly, respiratory viruses are the primary trigger of wheezing exacerbations in preschool children (73). As such, this multi-omic, immune-driven recurrent wheeze signature may reflect remnants of recurring infections, distinguishing it from the signature for progression to asthma.

While the Wheeze Score distinguished preschool-aged children with severe wheeze, the identified Disease Score captured an inferred trajectory from early childhood health to severe school-aged asthma in a stepwise manner, representing a set of potential host gene expression and bacterial alterations with utility as targets against disease progression. The Disease Score transcriptomic signature was suggestive of advancement toward Th2-high asthma, characterised by increased IL-13 and eosinophilia. While a transcriptional profile suggestive of eosinophilic recruitment was observed, no changes in BAL eosinophils proprtions were found. These changes were primarily associated with altered gene expression in epithelial and secretory cells rather than immune cells, highlighting a baseline epithelial dysfunction rather than an acute eosinophilic response. The heterogeneity of the airway epithelium confers diverse functionality and its contribution to various respiratory

diseases, including asthma, has been proposed and reviewed recently (74). Consistent biomarkers of Th2-high asthma that were recapitulated included upregulation of periostin (*POSTN*), *SERPINB2*, *SERPINB10*, and *MUC5AC*, and downregulation of lactotransferrin (*LTF*) and *MUC5B* (75–78). Airway mucus hypersecretion and secretory cell hyperplasia are features of asthma, alongside an increased *MUC5AC*/*MUC5B* ratio (79,80). *MUC2* is not typically expressed in the airways, but has been observed with airway irritation in asthma and chronic obstructive pulmonary disease (COPD) (80,81). Although collagen increases are associated with asthma, it has been reported that inhaled corticosteroids decrease collagen deposition (82). Their near ubiquitous use in the recurrent wheeze and asthma patients in this study likely explains the observed negative association between Disease Score and collagens I, III, and VI. Upregulation of other Th2-sensitive genes was also observed. *B3GNT6* encodes a protein with N-acetylglucosaminyltransferase activity that has an important role in biosynthesis of mucin-type glycoproteins and showed a trend to increase in IL-13-stimulated mucus secretory cells (83). Gelsolin (*GSN*) is secreted by human airway epithelial cells in response to IL-4 (84), and has actin depolymerisation activity. Inflammation-related cell death releases large amounts of filamentous actin into the airways, and therefore gelsolin upregulation may play an extracellular actin-scavenging role to prevent excessive airway surface liquid viscosity (84,85). Indeed, exogenous gelsolin administration has been shown to reduce sputum viscosity in cystic fibrosis (86). Uroplakin-1B (*UPK1B*) is a member of the transmembrane 4 superfamily shown to be increased in IL-13-treated primary esophageal epithelial cells, and was unresponsive to glucocorticoids in this context (87). Similarly, epithelial cell expression of *LRCC31* has been correlated with both esophageal eosinophilia and IL-13 expression, and has a role in increasing epithelial barrier function (88).

In addition to the shift in host gene expression, the Disease Score also captured changes in bacterial functional gene abundance of the lower airway microbiome. Due to low microbial biomass in the airways, utilisation of amplicon sequencing approaches is far more common, and most existing studies have focused on the upper airways. In a recent systematic review of microbiome studies in paediatric asthma (89), only 3/13 included studies investigated the lower airways by means of BAL or sputum samples, all of which were assessed by 16S amplicon sequencing. The combination of lower airway bronchial brush sampling and shotgun metagenomics used in this study is a novel resource for interpretation of host-microbe

interactions in recurrent wheeze and asthma. Colonisation with *Haemophilus influenzae* and *Neisseria* species have been previously linked to airway dysbiosis in children, and identified as risk factors for progression to asthma (8). Serological evidence of immune responses to *H. influenzae*, *Streptococcus pneumoniae*, and *Moraxella catarrhalis* have been reported in ~20% of wheezing children (90), and these species are the most commonly implicated in asthma development (91,92). The consistent detection of these bacteria across studies suggests they could be risk factors and influence asthma susceptibility and chronic inflammation in children with wheeze. While *M. catarrhalis* was detected in the current study, it was not associated with either the Wheeze or Disease scores. Importantly, we demonstrate by metatranscriptomics that these taxa are not simply bystanders, but are transcriptionally active in the lower airways. We found here that metagenomic composition was associated with the inferred disease trajectory but not with the type 1- and neutrophil-associated Wheeze Score. This suggests that the epithelial dysfunction creates a dysregulated microenvironment captured by the Disease Score, potentially shaping the dysbiotic microbiome independently of the inflammation seen in wheezers. Indeed, the signature captured by the Wheeze Score is probably linked to recent acute viral infections (93). Further, commensal taxa linked with respiratory homeostasis were transcriptionally-active and associated with decreased Disease Score values, including *Prevotella*, *Streptococcus*, and *Veillonella* species (94). This supports previous assertions that disease progression is linked with a shift from the Bacteroidetes phylum to Gammaproteobacteria, a phylum with many lung-associated Gram-negative pathogens.

While the unsupervised data integration approach and subsequent linear modelling used in this study accounted for age differences, the absence of bronchial brush samples from healthy school-aged children is a limitation. Similarly, the cross-sectional design of this study, while effective at identifying associations, precluded characterisation of longitudinal effects, and should be considered in future studies. Furthermore, while the use of untargeted metabolomics and lipidomics broadened the scope of potential findings, targeted validation would be required for investigation into potential biomarker utility.

There is certainly existing evidence of the importance of airway epithelial and smooth muscle changes in the pathogenesis of asthma (78,95–97); importantly, we now report key factors and modalities that distinguish children with recurrent wheeze from those who ultimately

develop asthma. While inflammatory measurements are common in cases of early life wheeze (4) and are of use in determining immediate treatment options (98), these are likely not impactful for prediction of future disease progression.

CONCLUSION

We report unsupervised, deep multi-omic profiling of lower airway samples in a paediatric cohort. We identified two distinct multi-omic signatures characterising preschool recurrent wheeze and the progression towards Type 2-high asthma, which were evident at steady-state in the absence of acute symptoms. The Wheeze Score revealed immune signatures involving neutrophil and Th1-associated pathways, while the Disease Score revealed an inferred trajectory toward childhood asthma marked by Type 2 immune-related gene expression changes in epithelial cells, altered bacterial abundance, and presence of transcriptionally-active pathogens such as *Haemophilus influenzae*. Our study emphasises the importance of the airway epithelial compartment in identification of predictive biomarkers that differentiate wheezers at risk of developing persistent asthma, rather than focusing on the inflammatory response at sampling. In a process made possible only by an integrated multi-omic approach, we were able to dissect progressive molecular signals driving disease and were able to distinguish these from preschool wheeze alone.

ACKNOWLEDGMENTS

We are grateful to the infants from the cohort and their families for their participation in this study, and also to the entire Breathing Together team. Support for High Performance Computing was provided by the MASSIVE High Performance Computing facility (www.massive.org.au).

REFERENCES

1. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. N Engl J Med. 1995 Jan 19;332(3):133–8.
2. Grad R, Morgan WJ. Long-term outcomes of early-onset wheeze and asthma. J Allergy Clin Immunol. 2012 Aug;130(2):299–307.
3. Bloom CI, Franklin C, Bush A, Saglani S, Quint JK. Burden of preschool wheeze and progression to asthma in the UK: Population-based cohort 2007 to 2017. J Allergy Clin Immunol. 2021 May;147(5):1949–58.
4. Laubhahn K, Schaub B. From preschool wheezing to asthma: Immunological determinants. Pediatr Allergy Immunol. 2023 Oct;34(10):e14038.
5. Wilson NM, Doré CJ, Silverman M. Factors relating to the severity of symptoms at 5 yrs in children with severe wheeze in the first 2 yrs of life. Eur Respir J. 1997 Feb;10(2):346–53.
6. Saglani S, Lloyd CM. Novel concepts in airway inflammation and remodelling in asthma. Eur Respir J. 2015 Dec;46(6):1796–804.
7. Guiddir T, Saint-Pierre P, Purenne-Denis E, Lambert N, Laoudi Y, Couderc R, et al. Neutrophilic Steroid-Refractory Recurrent Wheeze and Eosinophilic Steroid-Refractory Asthma in Children. J Allergy Clin Immunol Pract. 2017;5(5):1351-1361.e2.
8. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bønnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. N Engl J Med. 2007 Oct 11;357(15):1487–95.
9. Robinson PFM, Fontanella S, Ananth S, Martin Alonso A, Cook J, Kaya-de Vries D, et al. Recurrent Severe Preschool Wheeze: From Prespecified Diagnostic Labels to Underlying Endotypes. Am J Respir Crit Care Med. 2021 Sep 1;204(5):523–35.
10. Simpson JL, Daly J, Baines KJ, Yang IA, Upham JW, Reynolds PN, et al. Airway dysbiosis: Haemophilus influenzae and Tropheryma in poorly controlled asthma. Eur Respir J. 2016 Mar;47(3):792–800.
11. Singh S, Natalini JG, Segal LN. Lung microbial-host interface through the lens of multi-omics. Mucosal Immunol. 2022 May;15(5):837–45.
12. Chiu CY, Chou HC, Chang LC, Fan WL, Dinh MCV, Kuo YL, et al. Integration of metagenomics-metabolomics reveals specific signatures and functions of airway microbiota in mite-sensitized childhood asthma. Allergy. 2020 Nov;75(11):2846–57.
13. Cobos-Uribe C, Rebuli ME. Understanding the Functional Role of the Microbiome and Metabolome in Asthma. Curr Allergy Asthma Rep. 2023 Feb;23(2):67–76.
14. Turner S, Custovic A, Ghazal P, Grigg J, Gore M, Henderson J, et al. Pulmonary epithelial barrier and immunological functions at birth and in early life - key determinants of the development of asthma? A description of the protocol for the Breathing Together study. Wellcome Open Res. 2018;3:60.
15. Doherty GM, Christie SN, Skibinski G, Puddicombe SM, Warke TJ, de Courcey F, et al. Non-bronchoscopic sampling and culture of bronchial epithelial cells in children. Clin Exp Allergy. 2003 Sep;33(9):1221–5.
16. Bossley CJ, Fleming L, Gupta A, Regamey N, Frith J, Oates T, et al. Pediatric severe asthma is characterized by eosinophilia and remodeling without T(H)2 cytokines. J Allergy Clin Immunol. 2012 Apr;129(4):974-982.e13.

17. Hersh CP, Zacharia S, Prakash Arivu Chelvan R, Hayden LP, Mirtar A, Zarei S, et al. Immunoglobulin E as a Biomarker for the Overlap of Atopic Asthma and Chronic Obstructive Pulmonary Disease. Chronic Obstr Pulm Dis. 2020 Jan;7(1):1–12.
18. Patel H, Ewels P, Peltzer A, Botvinnik O, Sturm G, Moreno D, et al. nf-core/rnaseq: nf-core/rnaseq v3.10.1 - Plastered Rhodium Rudolph [Internet]. Zenodo; 2023. Available from: https://doi.org/10.5281/zenodo.7505987
19. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
20. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 2014 Feb 3;15(2):R29.
21. Pattaroni C, Macowan M, Chatzis R, Daunt C, Custovic A, Shields MD, et al. Early life inter-kingdom interactions shape the immunological environment of the airways. Microbiome. 2022 Feb 21;10(1):34.
22. Wypych TP, Pattaroni C, Perdijk O, Yap C, Trompette A, Anderson D, et al. Microbial metabolism of L-tyrosine protects against allergic airway inflammation. Nat Immunol. 2021 Mar;22(3):279–86.
23. Clarke EL, Taylor LJ, Zhao C, Connell A, Lee JJ, Fett B, et al. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. Microbiome. 2019 Mar 22;7(1):46.
24. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using illumina paired-end technology. PLoS One. 2013;8(6):e66643.
25. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019 Nov 28;20(1):257.
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.
27. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016 Feb 22;428(4):726–31.
28. Lahti L, Shetty S. microbiome R package. 2012.
29. Tsugawa H, Ikeda K, Takahashi M, Satoh A, Mori Y, Uchino H, et al. A lipidome atlas in MS-DIAL 4. Nat Biotechnol. 2020 Oct;38(10):1159–63.
30. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D608–17.
31. Jankevics A, Lloyd GR, Weber RJM. pmp: Peak Matrix Processing and signal batch correction for metabolomics datasets [Internet]. 2023. Available from: https://bioconductor.org/packages/pmp
32. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010 Jul;45(7):703–14.
33. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012 Jan 1;28(1):112–8.
34. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2023. Available from: https://www.R-project.org/
35. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org
36. Ramos M, Schiffer L, Re A, Azhar R, Basunia A, Cabrera CR, et al. Software For The

Integration Of Multi-Omics Experiments In Bioconductor. Cancer Research. 2017;77(21); e39-42.
37. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020 May 11;21(1):111.
38. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package [Internet]. 2022. Available from: https://CRAN.R-project.org/package=vegan
39. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madissoon E, et al. An integrated cell atlas of the lung in health and disease. Nat Med. 2023 Jun;29(6):1563–77.
40. Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinform Adv. 2022;2(1):vbac016.
41. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom Bioinform. 2020 Sep;2(3):lqaa078.
42. Edgeworth J, Gorman M, Bennett R, Freemont P, Hogg N. Identification of p8,14 as a highly abundant heterodimeric calcium binding protein complex of myeloid cells. J Biol Chem. 1991 Apr 25;266(12):7706–13.
43. Radka CD, Miller DJ, Frank MW, Rock CO. Biochemical characterization of the first step in sulfonolipid biosynthesis in Alistipes finegoldii. J Biol Chem. 2022 Aug;298(8):102195.
44. Mock ED, Gagestein B, van der Stelt M. Anandamide and other N-acylethanolamines: A class of signaling lipids with therapeutic opportunities. Prog Lipid Res. 2023 Jan;89:101194.
45. Faner R, Sibila O, Agustí A, Bernasconi E, Chalmers JD, Huffnagle GB, et al. The microbiome in respiratory medicine: current challenges and future perspectives. Eur Respir J. 2017 Apr;49(4):1602086.
46. Alamri A. Diversity of Microbial Signatures in Asthmatic Airways. Int J Gen Med. 2021;14:1367–78.
47. Schröder JM, Mrowietz U, Morita E, Christophers E. Purification and partial biochemical characterization of a human monocyte-derived, neutrophil-activating peptide that lacks interleukin 1 activity. J Immunol. 1987 Nov 15;139(10):3474–83.
48. Yoshimura T, Matsushima K, Tanaka S, Robinson EA, Appella E, Oppenheim JJ, et al. Purification of a human monocyte-derived neutrophil chemotactic factor that has peptide sequence similarity to other host defense cytokines. Proc Natl Acad Sci U S A. 1987 Dec;84(24):9233–7.
49. Kobayashi Y. The role of chemokines in neutrophil biology. Front Biosci. 2008 Jan 1;13:2400–7.
50. Rajarathnam K, Schnoor M, Richardson RM, Rajagopal S. How do chemokines navigate neutrophils to the target site: Dissecting the structural mechanisms and signaling pathways. Cell Signal. 2019 Feb;54:69–80.
51. Kim YM, Kim H, Lee S, Kim S, Lee JU, Choi Y, et al. Airway G-CSF identifies neutrophilic inflammation and contributes to asthma progression. Eur Respir J. 2020 Feb;55(2):1900827.
52. Thorley AJ, Goldstraw P, Young A, Tetley TD. Primary human alveolar type II epithelial cell CCL20 (macrophage inflammatory protein-3alpha)-induced dendritic cell migration. Am J Respir Cell Mol Biol. 2005 Apr;32(4):262–7.
53. Lodhi IJ, Wei X, Yin L, Feng C, Adak S, Abou-Ezzi G, et al. Peroxisomal lipid synthesis

regulates inflammation by sustaining neutrophil membrane phospholipid composition and viability. Cell Metab. 2015 Jan 6;21(1):51–64.
54. Waltz DA, Fujita RM, Yang X, Natkin L, Zhuo S, Gerard CJ, et al. Nonproteolytic role for the urokinase receptor in cellular migration in vivo. Am J Respir Cell Mol Biol. 2000 Mar;22(3):316–22.
55. Ivetic A, Hoskins Green HL, Hart SJ. L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling. Front Immunol. 2019;10:1068.
56. Ivetic A. A head-to-tail view of L-selectin and its impact on neutrophil behaviour. Cell Tissue Res. 2018 Mar;371(3):437–53.
57. Loitto VM, Forslund T, Sundqvist T, Magnusson KE, Gustafsson M. Neutrophil leukocyte motility requires directed water influx. J Leukoc Biol. 2002 Feb;71(2):212–22.
58. Grommes J, Drechsler M, Soehnlein O. CCR5 and FPR1 mediate neutrophil recruitment in endotoxin-induced lung injury. J Innate Immun. 2014;6(1):111–6.
59. Alessi MC, Cenac N, Si-Tahar M, Riteau B. FPR2: A Novel Promising Target for the Treatment of Influenza. Front Microbiol. 2017;8:1719.
60. Fiore S, Maddox JF, Perez HD, Serhan CN. Identification of a human cDNA encoding a functional high affinity lipoxin A4 receptor. J Exp Med. 1994 Jul 1;180(1):253–60.
61. Wang S, Song R, Wang Z, Jing Z, Wang S, Ma J. S100A8/A9 in Inflammation. Front Immunol. 2018;9:1298.
62. Hessian PA, Edgeworth J, Hogg N. MRP-8 and MRP-14, two abundant Ca(2+)-binding proteins of neutrophils and monocytes. J Leukoc Biol. 1993 Feb;53(2):197–204.
63. Hou L, Tian HY, Wang L, Ferris ZE, Wang J, Cai M, et al. Identification and Biosynthesis of Pro-Inflammatory Sulfonolipids from an Opportunistic Pathogen Chryseobacterium gleum. ACS Chem Biol. 2022 May 20;17(5):1197–206.
64. Uhlig S, Gulbins E. Sphingolipids in the lungs. Am J Respir Crit Care Med. 2008 Dec 1;178(11):1100–14.
65. James BN, Oyeniran C, Sturgill JL, Newton J, Martin RK, Bieberich E, et al. Ceramide in apoptosis and oxidative stress in allergic inflammation and asthma. J Allergy Clin Immunol. 2021 May;147(5):1936-1948.e9.
66. Harper RW, Xu C, Eiserich JP, Chen Y, Kao CY, Thai P, et al. Differential regulation of dual NADPH oxidases/peroxidases, Duox1 and Duox2, by Th1 and Th2 cytokines in respiratory tract epithelium. FEBS Lett. 2005 Aug 29;579(21):4911–7.
67. Jiang Z, Chen Z, Li L, Zhou W, Zhu L. Lack of SOCS3 increases LPS-induced murine acute lung injury through modulation of Ly6C(+) macrophages. Respir Res. 2017 Dec 29;18(1):217.
68. Zhao J, Yu H, Liu Y, Gibson SA, Yan Z, Xu X, et al. Protective effect of suppressing STAT3 activity in LPS-induced acute lung injury. Am J Physiol Lung Cell Mol Physiol. 2016 Nov 1;311(5):L868–80.
69. Benavente FM, Soto JA, Pizarro-Ortega MS, Bohmwald K, González PA, Bueno SM, et al. Contribution of IDO to human respiratory syncytial virus infection. J Leukoc Biol. 2019 Oct;106(4):933–42.
70. Heyes MP, Chen CY, Major EO, Saito K. Different kynurenine pathway enzymes limit quinolinic acid formation by various human cell types. Biochem J. 1997 Sep 1;326 (Pt 2)(Pt 2):351–6.
71. Murakami Y, Hoshi M, Imamura Y, Arioka Y, Yamamoto Y, Saito K. Remarkable role of indoleamine 2,3-dioxygenase and tryptophan metabolites in infectious diseases: potential role in macrophage-mediated inflammatory diseases. Mediators Inflamm.

2013;2013:391984.
72. Grant RS, Passey R, Matanovic G, Smythe G, Kapoor V. Evidence for increased de novo synthesis of NAD in immune-activated RAW264.7 macrophages: a self-protective mechanism? Arch Biochem Biophys. 1999 Dec 1;372(1):1–7.
73. Doss AMA, Stokes JR. Viral Infections and Wheezing in Preschool Children. Immunol Allergy Clin North Am. 2022 Nov;42(4):727–41.
74. Davis JD, Wypych TP. Cellular and functional heterogeneity of the airway epithelium. Mucosal Immunol. 2021 Sep;14(5):978–90.
75. Cao Y, Wu Y, Lin L, Yang L, Peng X, Chen L. Identifying key genes and functionally enriched pathways in Th2-high asthma by weighted gene co-expression network analysis. BMC Med Genomics. 2022 May 12;15(1):110.
76. Nie X, Wei J, Hao Y, Tao J, Li Y, Liu M, et al. Consistent Biomarkers and Related Pathogenesis Underlying Asthma Revealed by Systems Biology Approach. Int J Mol Sci. 2019 Aug 19;20(16):4037.
77. Southworth T, Van Geest M, Singh D. Type-2 airway inflammation in mild asthma patients with high blood eosinophils and high fractional exhaled nitric oxide. Clin Transl Sci. 2021 Jul;14(4):1259–64.
78. Zissler UM, Esser-von Bieren J, Jakwerth CA, Chaker AM, Schmidt-Weber CB. Current and future biomarkers in allergic asthma. Allergy. 2016 Apr;71(4):475–94.
79. Bonser LR, Erle DJ. Airway Mucus and Asthma: The Role of MUC5AC and MUC5B. J Clin Med. 2017 Nov 29;6(12):112.
80. Ordoñez CL, Khashayar R, Wong HH, Ferrando R, Wu R, Hyde DM, et al. Mild and moderate asthma is associated with airway goblet cell hyperplasia and abnormalities in mucin gene expression. Am J Respir Crit Care Med. 2001 Feb;163(2):517–23.
81. Rogers DF. Airway mucus hypersecretion in asthma: an undervalued pathology? Curr Opin Pharmacol. 2004 Jun;4(3):241–50.
82. Hoshino M, Takahashi M, Takai Y, Sim J. Inhaled corticosteroids decrease subepithelial collagen deposition by modulation of the balance between matrix metalloproteinase-9 and tissue inhibitor of metalloproteinase-1 expression in asthma. J Allergy Clin Immunol. 1999 Aug;104(2 Pt 1):356–63.
83. Jackson ND, Everman JL, Chioccioli M, Feriani L, Goldfarbmuren KC, Sajuthi SP, et al. Single-Cell and Population Transcriptomics Reveal Pan-epithelial Remodeling in Type 2-High Asthma. Cell Rep. 2020 Jul 7;32(1):107872.
84. Candiano G, Bruschi M, Pedemonte N, Caci E, Liberatori S, Bini L, et al. Gelsolin secretion in interleukin-4-treated bronchial epithelia and in asthmatic airways. Am J Respir Crit Care Med. 2005 Nov 1;172(9):1090–6.
85. Smith DB, Janmey PA, Lind SE. Circulating actin-gelsolin complexes following oleic acid-induced lung injury. Am J Pathol. 1988 Feb;130(2):261–7.
86. Vasconcellos CA, Allen PG, Wohl ME, Drazen JM, Janmey PA, Stossel TP. Reduction in viscosity of cystic fibrosis sputum in vitro by gelsolin. Science. 1994 Feb 18;263(5149):969–71.
87. Blanchard C, Mingler MK, Vicario M, Abonia JP, Wu YY, Lu TX, et al. IL-13 involvement in eosinophilic esophagitis: transcriptome analysis and reversibility with glucocorticoids. J Allergy Clin Immunol. 2007 Dec;120(6):1292–300.
88. D'Mello RJ, Caldwell JM, Azouz NP, Wen T, Sherrill JD, Hogan SP, et al. LRRC31 is induced by IL-13 and regulates kallikrein expression and barrier function in the esophageal epithelium. Mucosal Immunol. 2016 May;9(3):744–56.

89. Aldriwesh MG, Al-Mutairi AM, Alharbi AS, Aljohani HY, Alzahrani NA, Ajina R, et al. Paediatric Asthma and the Microbiome: A Systematic Review. Microorganisms. 2023 Apr 3;11(4):939.
90. Lehtinen P, Jartti T, Virkki R, Vuorinen T, Leinonen M, Peltola V, et al. Bacterial coinfections in children with viral wheezing. Eur J Clin Microbiol Infect Dis. 2006 Jul;25(7):463–9.
91. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. PLoS One. 2010 Jan 5;5(1):e8578.
92. Mthembu N, Ikwegbue P, Brombacher F, Hadebe S. Respiratory Viral and Bacterial Factors That Influence Early Childhood Asthma. Front Allergy. 2021;2:692841.
93. Invernizzi R, Lloyd CM, Molyneaux PL. Respiratory microbiome and epithelial interactions shape immunity in the lungs. Immunology. 2020 Jun;160(2):171–82.
94. Huffnagle GB, Dickson RP, Lukacs NW. The respiratory tract microbiome and lung inflammation: a two-way street. Mucosal Immunol. 2017 Mar;10(2):299–306.
95. Bush A. Cytokines and Chemokines as Biomarkers of Future Asthma. Front Pediatr. 2019;7:72.
96. Saglani S, Nicholson AG, Scallan M, Balfour-Lynn I, Rosenthal M, Payne DN, et al. Investigation of young children with severe recurrent wheeze: any clinical benefit? Eur Respir J. 2006 Jan;27(1):29–35.
97. Lezmi G, Gosset P, Deschildre A, Abou-Taam R, Mahut B, Beydon N, et al. Airway Remodeling in Preschool Children with Severe Recurrent Wheeze. Am J Respir Crit Care Med. 2015 Jul 15;192(2):164–71.
98. Bacharier LB, Guilbert TW, Jartti T, Saglani S. Which Wheezing Preschoolers Should be Treated for Asthma? J Allergy Clin Immunol Pract. 2021 Jul;9(7):2611–8.

FIGURE LEGENDS

Figure 1. MOFA+ multi-omic integration reveals wheeze-associated and disease-associated latent factors. A. Stacked barplot showing the contribution of each omic modality to the total variance explained by each factor. **B.** Density plot of the wheeze-associated latent factor (Latent Factor 2) coloured by group, with corresponding boxplot showing significant differences between preschool wheezers and school-aged asthmatics. **C.** Density plot of the disease-associated latent factor (Latent Factor 4) coloured by group, with corresponding boxplot showing significant differences between all groups. Sample sizes are n=29 for preschool controls, n=22 for preschool wheezers, and n=22 for school-aged asthmatics. Significance levels are indicated as p<0.05=*, p<0.01=**, p<0.001=***, p<0.0001=****.

Figure 2. Multi-omic-derived Wheeze Score (LF2) is characterised by changes in host gene expression and small molecules. A. Barplot representing the top limma DE genes along the Wheeze Score, with Benjamini-Hochberg (BH) multiple testing correction. **B.** Principal component analysis (PCA) biplot of principal components 1 and 2 of differential genes (with logFC>1) in panel A (subset of batch-corrected Integrated Human Lung Cell Atlas scRNAseq pseudo-bulk data; n=766 profiles of 50 pseudo-bulk cell types from 28 studies). PCA eigenvectors with eigenvalues >0.5 are shown, and coloured according to the Wheeze Score direction with higher expression. Eigenvectors touching the outer ring with open triangles extend beyond the plot bounds. Cell types occupying different sectors of the plot are indicated by lines at the plot edges. **C.** Barplot representing top limma DA small molecules along the Wheeze Score, with BH multiple testing correction. Lipid classes are indicated by coloured dots. **D.** Excerpt schematic of the kynurenine pathway, highlighting genes and small molecules upregulated along the Wheeze Score in purple. **E.** Dot plot summary of Spearman correlation analysis of log-transformed BAL immune cell and red blood cell percentages compared to either the Wheeze or Disease Score values. Dot size is proportional to -log10(p-value) and colour is scaled by Spearman rho. Significance values are indicated above their respective dots. Sample sizes are n=29 for preschool controls, n=22 for preschool wheezers, and n=22 for school-aged asthmatics. Significance levels are indicated as p<0.05=*, p<0.01=**, p<0.001=***.

Figure 3. Multi-omic-derived Disease Score (LF4) is characterised by changes in host gene expression and bacterial gene counts. A. Barplot representing the top limma DE genes along the Disease Score, with Benjamini-Hochberg (BH) multiple testing correction. **B.** Principal component analysis (PCA) biplot of principal components 1 and 2 of differential genes (with logFC>0.5) in panel A (subset of batch-corrected Integrated Human Lung Cell Atlas scRNAseq pseudo-bulk data; n=766 profiles of 50 pseudo-bulk cell types from 28 studies). PCA eigenvectors with eigenvalues >0.5 are shown, and coloured according to the Disease Score direction with higher expression. Eigenvectors touching the outer ring with open triangles extend beyond the plot bounds. Cell types occupying different sectors of the plot are indicated by lines at the plot edges. **C.** Barplots representing the number of differentially-abundant (DA) bacterial taxa-aggregated KEGG orthologs (KOs), i.e. functional gene counts, along the Disease Score, determined by limma with BH multiple-testing correction. Bars are coloured according to whether there are more DA gene counts in health or disease (left panel). Barplot representing Spearman correlation coefficients between taxa-aggregated metagenomic and metatranscriptomic sequencing reads as a measure of transcriptional activity. Black bars indicate significant correlation; grey bars are non-significant (right panel). **D.** Boxplot comparing Disease Score values between non-atopic and atopic patients, with dots coloured according to sample cohort. Sample sizes are n=29 for preschool controls, n=22 for preschool wheezers, and n=22 for school-aged asthmatics. Significance levels are indicated as p<0.1=., p<0.05=*, p<0.01=**, p<0.001=***.

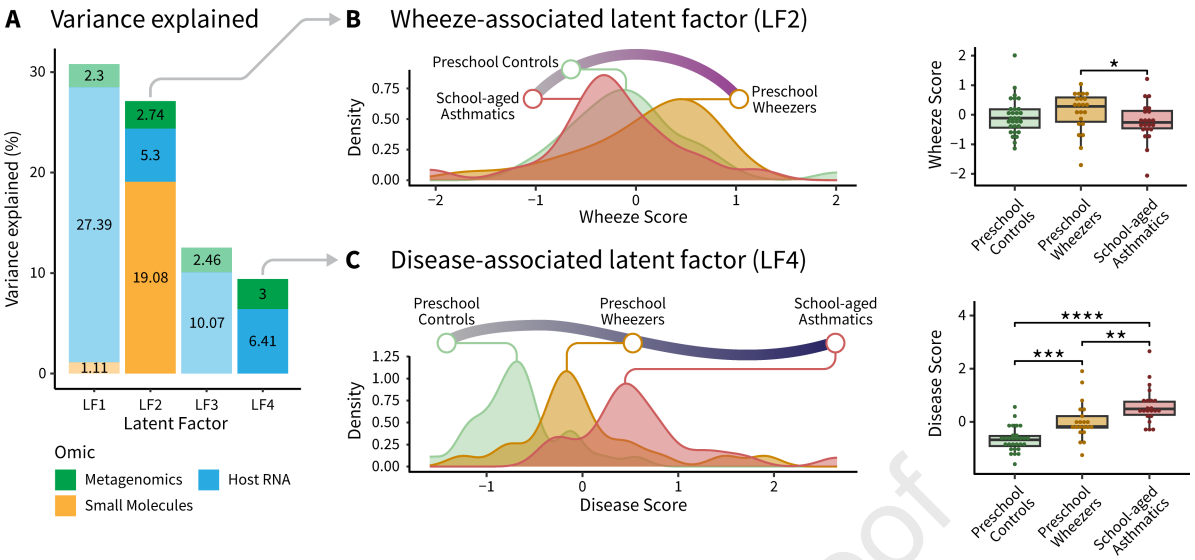
TABLES

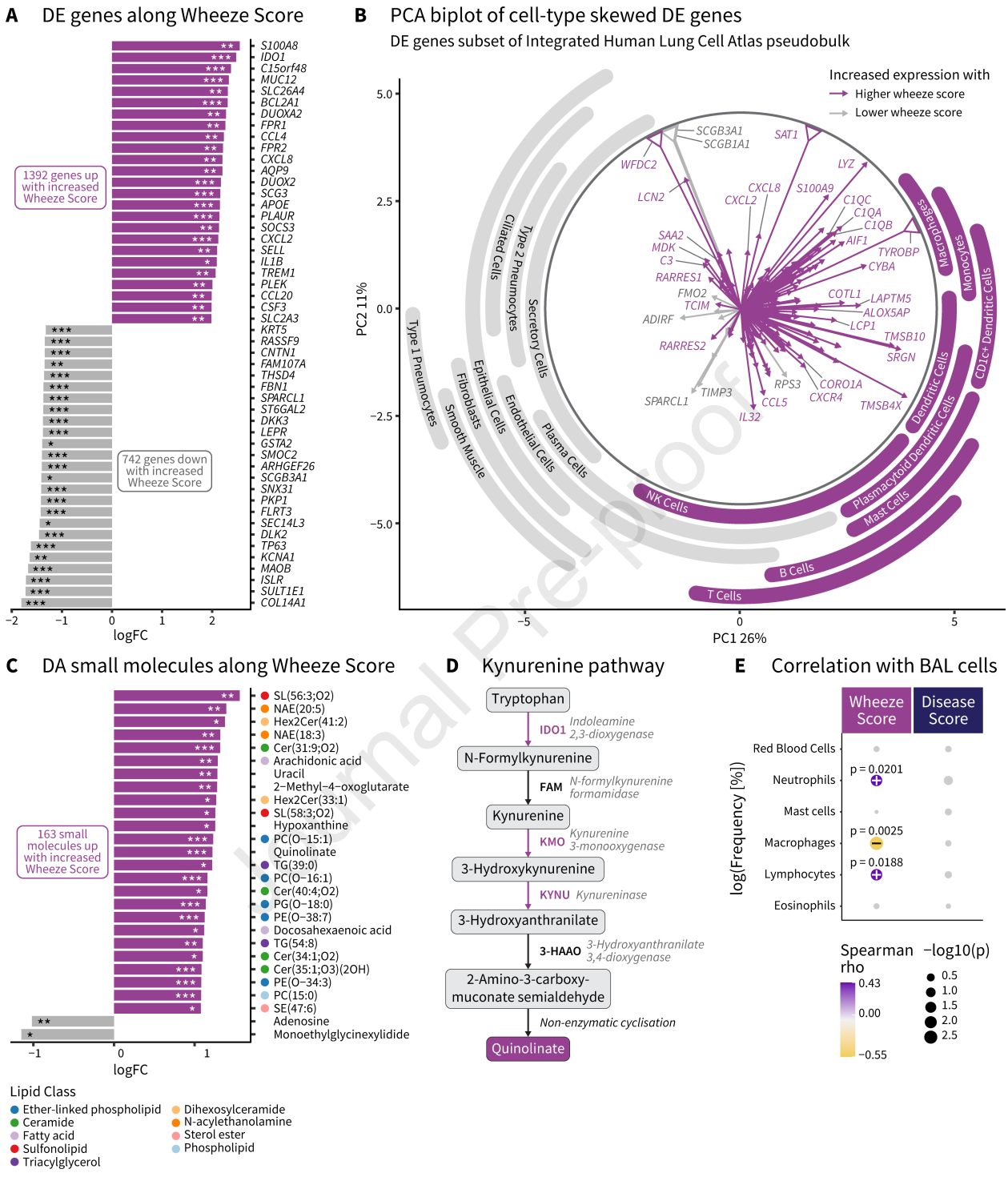
	Preschool controls	Preschool recurrent wheezers	School-aged asthmatics	Chi-squared test - Preschool controls vs. preschool recurrent wheezers	Chi-squared test - Preschool recurrent wheezers vs. school-aged asthmatics
Number of patients	29	22	22		
Age at sampling (years (range))	4 (1.5 - 5.6)	3.6 (2 - 5.8)	12.5 (6.7 - 17.1)		
Sex (Male)	58.6% (17/29)	63.6% (14/22)	36.4% (8/22)	ns (p = 1.00)	ns (p = 0.171)
Preterm delivery (<37 weeks)	0% (0/29)	0% (0/22)	13.6% (3/22)	ns (p = 1.00)	ns (p = 0.209)
Daycare*	93.1% (27/29)	50% (8/16, 6 NA)	NA	**** (p = 3.13e-03)**	NA
Inhaled corticosteroid usage#	13.8% (4/29)	86.4% (19/22)	90.1% (20/22)	**** (p = 3.73e-07)**	ns (p = 1.00)
Bronchodilators#	20.7% (6/29)	100% (22/22)	100% (22/22)	**** (p = 1.39e-07)**	ns (p = 1.00)
Normal vaginal delivery	62% (18/29)	72.7% (16/22)	54.5% (12/22)	ns (p = 0.699)	ns (p = 0.407)
Ethnicity (White)	65.5% (19/29)	59.1% (13/22)	59.1% (13/22)	ns (p = 1.00)	ns (p = 1.00)

Atopy (serum total IgE >30 IU/mL)*	58.3% (7/12, 17 NA)	54.5% (12/22)	81.8% (18/22)	ns (p = 1.00)	ns (p = 0.083)
Family history of asthma	37.9% (11/29)	77.3% (17/22)	81.8% (18/22)	*** (p = 1.68e-02)**	ns (p = 0.937)
Family history of eczema	34.5% (10/29)	31.8% (7/22)	50% (11/22)	ns (p = 0.893)	ns (p = 0.209)
Family history of allergic rhinitis	48.3% (14/29)	54.5% (12/22)	68.2% (15/22)	ns (p = 0.739)	ns (p = 0.665)

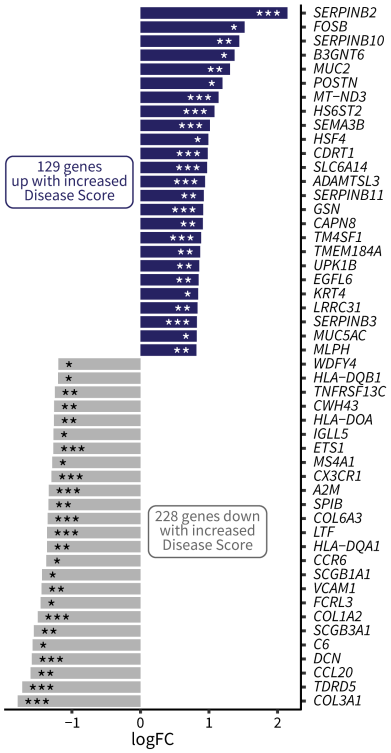
Table 1. Demographic information for patients included in multi-omic analysis.

Data is provided at either median (range) or percentage (n/n). Significance values represent results of Chi-squared tests for comparisons between two categorical variables. *Certain clinical variables are missing for some patients. #Inhaled corticosteroids or bronchodilators were prescribed to a small number of preschool controls for a single, isolated wheeze event. These children were considered controls for the purpose of clinical grouping as the wheeze was not recurrent.



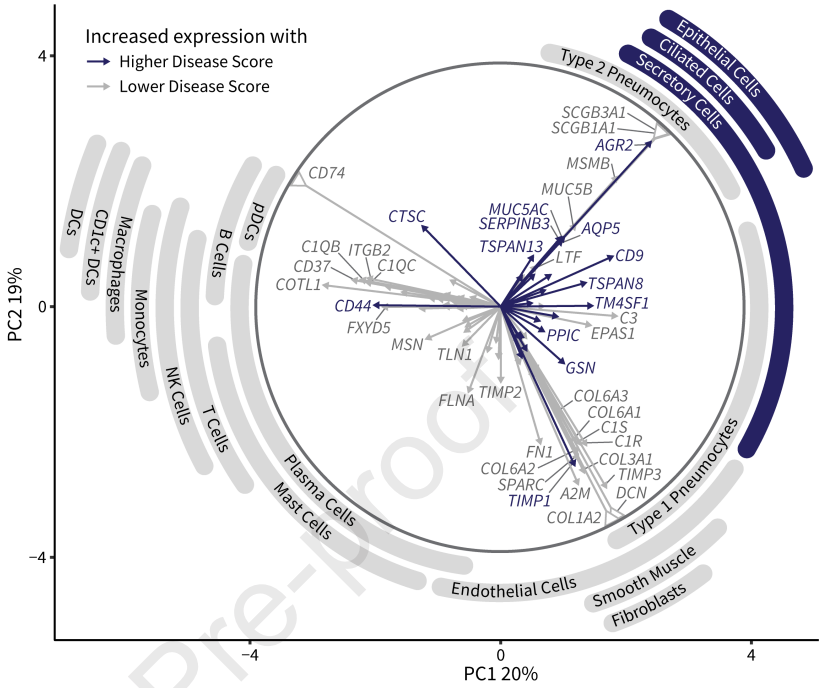


A DE genes along Disease Score

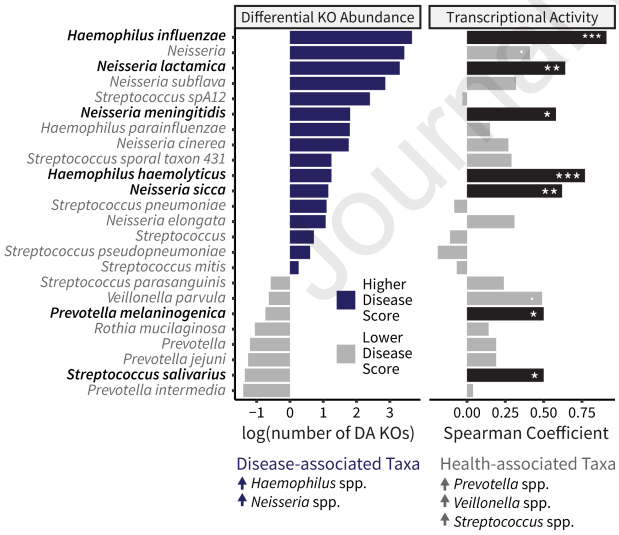


B PCA biplot of cell-type skewed DE genes

DE genes subset of Integrated Human Lung Cell Atlas pseudobulk



C Shift in taxa with DA KEGG orthologs



D Disease Score and Atopy