BMJ Open Respiratory Research

The Wheezing, Asthma and Viral effects on the Epithelial Structure and function (WAVES) birth cohort study: rationale, design and methods to understand airway development and the role of early-life respiratory viral infections on childhood asthma inception

Tina Hartert ,[1,2] Christian Rosas-Salazar,[2] Dawn C Newcomb,[1] Brittney M Snyder,[3] Sergejs Berdnikovs,[4] Christopher McKennan,[5] Sarah Osmundson,[6] Zhouwen Liu,[7] Erica Tafoya,[8] Wais Folad,[8] Pat Simon,[8] Sara Reiss,[8] Kadijah Poleon,[8] Jacqueline-Yvonne Cephus,[8] Shelby N Kuehnle,[8] Kaitlin E McKernan,[8] Siyuan Ma,[7] Larry J Anderson,[9] Tebeb Gebretsadik[7]

To cite: Hartert T, Rosas-Salazar C, Newcomb DC, *et al*. The Wheezing, Asthma and Viral effects on the Epithelial Structure and function (WAVES) birth cohort study: rationale, design and methods to understand airway development and the role of early-life respiratory viral infections on childhood asthma inception. *BMJ Open Respir Res* 2026;13:e004049. doi:10.1136/bmjresp-2025-004049

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjresp-2025-004049).

Received 2 January 2026
Accepted 30 April 2026

For numbered affiliations see end of article.

Correspondence to
Professor Tina Hartert; tina.hartert@vumc.org

ABSTRACT

Introduction The airway epithelial barrier continues to develop during the first year of life, representing a critical time window of susceptibility to the effects of respiratory viruses that may contribute to the development of childhood asthma. This study specifically aimed to assess the impact of respiratory syncytial virus (RSV) infection on the longitudinal development of the airway epithelium, airway metabolism, DNA methylation and subsequent childhood asthma.

Methods and analysis The Wheezing, Asthma and Viral effects on the Epithelial Structure and function (WAVES) birth cohort study is a prospective, observational birth cohort that is enrolling mother–child dyads beginning in pregnancy and following children through age 5 years. It is designed as a quasi-natural randomisation of RSV infection in infancy, capturing the first RSV infection event during infancy and longitudinal nasal sampling from birth through age 5 years for multi-omic assays and assessment of clinical outcomes. The WAVES study protocol will enrol 250 pregnant women and their offspring (200 at high-risk for asthma and 50 at general-risk for asthma). The primary outcomes are pathways through which RSV infection intersects with airway epithelial development leading to asthma by age 5 years. We will employ a longitudinal systems biology approach to investigate the development of the airway epithelium following infant RSV infection in those who do and do not develop asthma. This approach will integrate single-cell and bulk RNA-sequencing gene expression, metabolomic and epigenetic data at the resolution of epithelial cell subsets.

Ethics and dissemination This study has been approved by the Vanderbilt University Medical Center IRB (#222277), and study results will be communicated in peer-reviewed publications and to the study participants and lay community. By analysing these longitudinal multi-omic datasets, we aim to uncover key mechanisms driving airway epithelial development and identify how these processes are altered by early-life RSV infection and contribute to the development of early-life respiratory morbidity and asthma.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ During the first year of life, the airway epithelial barrier is undergoing dynamic development, representing a critical time window of increased susceptibility to respiratory viruses. Infection with respiratory viruses, such as respiratory syncytial virus, during this window has been implicated in the development of childhood asthma and other airway morbidity.

WHAT THIS STUDY ADDS

⇒ By identifying airway epithelial growth patterns present at birth and how these patterns change throughout infancy in response to environmental factors, this study will provide new insights into how airway developmental processes may be shaped by external insults, such as respiratory syncytial virus infection.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Understanding the interplay between airway epithelial development, environmental exposures and susceptibility windows in early life lays the foundation for targeted interventions aimed at preventing acute and chronic respiratory diseases. Identifying key biologic pathways as potential therapeutic targets for asthma prevention may inform clinical and public health strategies to improve child airway health.

INTRODUCTION

The origin and clinical manifestations of asthma are tightly linked with dysfunctional physical, metabolic and functional barrier properties of airway epithelial cells (AECs). While immune cells are also important in asthma pathogenesis, the airway epithelium is the primary site of respiratory viral infection, and many established asthma risk factors are exposures that interact with the airway epithelium. The airway epithelial barrier continues to develop during the first year of life, representing a critical time window of susceptibility to the effects of respiratory viruses that may contribute to the development of childhood asthma.

Asthma, like most complex health conditions, is thought to develop from exposure to multiple exogenous and endogenous factors that interact to influence individual susceptibility. Although a proportion of childhood asthma is considered heritable, the increasing prevalence and marked differences in rates of asthma in different countries and in migrating populations over a relatively short period of time cannot be explained by genetic differences alone.[1] Along with results from multi-intervention trials establishing asthma prevention, these findings strongly support the critical role of environmental factors in asthma development and prevention.[2–5]

In our Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure (INSPIRE) cohort,[6] we have established that (1) early-life respiratory syncytial virus (RSV) infection (not just severe infection) is associated with the development of a non-allergic asthma phenotype that explains a substantial proportion of childhood asthma, (2) there is an underlying airway epithelial cell wheeze phenotype present in early childhood, characterised by profound developmental changes in basal and secretory subsets of nasal AECs (NAECs), (3) early-life RSV infection has a lasting impact on the DNA methylation in airway progenitor cells at functionally and clinically relevant asthma loci, and (4) NAECs in 2- to 3-year-old children who develop recurrent wheeze are characterised by altered glucose metabolism and metabolic reprogramming of epithelial cells associated with RSV infection in infancy.[7–10] Our published observations are critical, as they highlight two key factors in airway susceptibility to disease development: developmental reprogramming and early-life RSV infection. Furthermore, they reveal a mechanistic connection between respiratory viral infection-induced metabolic and epigenetic changes and their impact on epithelial development during early life. However, NAEC development in these studies was first assessed in samples collected at the age of 2 years, which is a limitation, as there are no samples prior to infection.[6] The Wheezing, Asthma and Viral effects on the Epithelial Structure and function (WAVES) study was established under the support of the National Institute of Allergy and Infectious Diseases (NIAID) Asthma and Allergic Diseases Cooperative Research Center in 2020 to extend the findings from the INSPIRE cohort to understand the role of early-life RSV infection on airway epithelial programming and development, assess developmental epithelial endotypes associated with asthma phenotypes longitudinally beginning at birth and include AECs before and following infant RSV infection.

Our motivation for this study stems from our findings and those of others, which led us to hypothesise (1) that distinct airway epithelial developmental endotypes are present at birth and play a critical role in shaping specific asthma phenotypes and (2) that these airway endotypes exhibit unique responses to environmental exposures such as RSV infection during infancy that contribute to asthma development. Developmental, metabolic and epigenetic programmes drive abnormal, pre-asthmatic airway barrier changes and development of the early-life epithelium that are key events in viral response and asthma inception.

To test our hypotheses, we conceptualised WAVES, a birth cohort study specifically designed to assess the impact of RSV infection on the longitudinal development of the airway epithelium, airway metabolism, gene expression, DNA methylation and subsequent childhood asthma. Using an integrated approach, we aim to identify clinically meaningful nasal mucosal endotypes, developmental pathways, metabolic changes and epigenetic alterations that will define normal and asthmatic airway epithelial cell biology and development and how respiratory viruses alter airway epithelial development to inform targeted prevention and treatments as outlined in figure 1.

The WAVES study was designed to address the following objectives: (1) characterise the longitudinal biologic differentiation pathways of the airway mucosa using single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing (RNA-seq) to define developmental differences present at birth associated with subsequent wheeze and asthma risk and whether development differs by resulting asthma phenotypes; (2) identify the longitudinal changes in airway epithelial cell metabolism in infants that precede the development of asthma phenotypes in early childhood and the specific metabolic pathways influenced by RSV infection that distinguish development of asthma phenotypes and endotypes; and (3) determine the effect of early-life RSV infection on airway epithelial DNA methylation and subsequent childhood asthma.

The guiding principle of WAVES is to focus on clinically relevant research, and this birth cohort was designed specifically to address key gaps in our understanding of childhood asthma development and the role of genetics, metabolism and early-life respiratory viral infection so that translation of the work from these projects may elucidate the developmental heterogeneity of asthma, which is critical to identifying biologic targets for prevention and treatment.

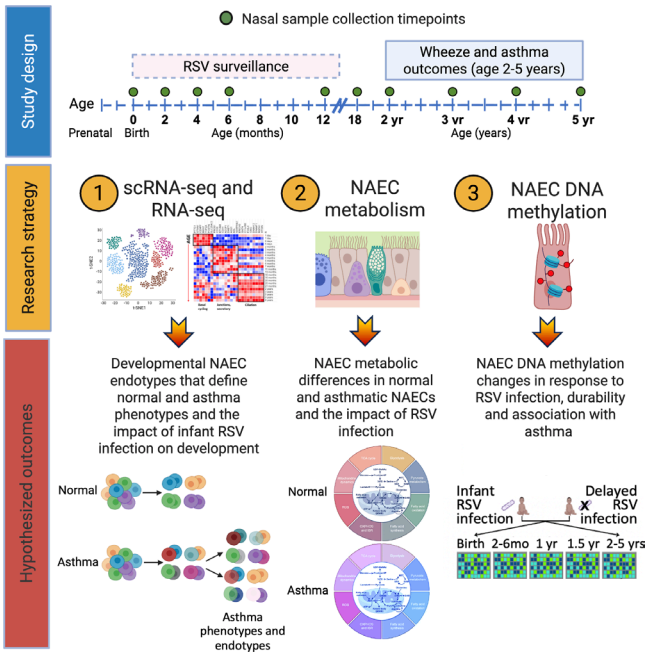

Figure 1 Schematic overview of study design, research strategy and hypothesised outcomes. AGE, advanced glycation end products; GDP, guanosine diphosphate; mo, month; NAEC, nasal airway epithelial cell; OXPHOS, oxidative phosphorylation; RNA-seq, RNA sequencing; ROS, reactive oxygen species; RSV, respiratory syncytial virus; scRNA-seq, single-cell RNA sequencing; TCA, tricarboxylic acid cycle; UDP-GABA, uridine diphosphate-γ-aminobutyric acid; yr, year; yrs, years.

METHODS AND ANALYSIS

Study design and target population: WAVES is a prospective observational pregnancy cohort that is enrolling mother–child dyads beginning in pregnancy and following children through age 5 years. It is designed as a quasi-natural randomisation of RSV infection in infancy, precisely and accurately defining the first RSV infection event through active and passive surveillance and performing annual follow-up for the assessment of clinical outcomes. The WAVES study protocol plans to enrol 250 pregnant women and their offspring (200 at high risk for asthma and 50 at general risk for asthma). High-risk participants include those with one or both parents or a sibling with asthma and/or allergic rhinitis. General risk, or low-risk, participants include those in whom neither of the biological parents have asthma or allergic rhinitis. Participants are pregnant women receiving their obstetrical care at Vanderbilt University Medical Centre (VUMC) or an affiliated site with intent to deliver at VUMC. The study population at VUMC is representative of the races and ethnicities of the middle Tennessee region, which includes white (62.5%), black (27.7%), other race (0.1%) and Hispanic of any race (22.2%) and spans urban, suburban and rural communities.[11] Patient and public involvement initiatives have been undertaken to inform elements of study design including a studio to provide feedback on individual and composite return of results, study newsletters and participant feedback from prior birth cohorts that informed the design of the WAVES study.

The WAVES study follows the Strengthening the Reporting of OBservational studies in Epidemiology (STROBE) checklist for cohort studies (online supplemental table 1).[12 13]

Ethics and safety: the Institutional Review Board (IRB) of VUMC (TN, USA) granted ethical approval (IRB #222277) for this study, and the biological mother of each child provided written informed consent for their and their child's participation. Although this is an observational study, we have a data safety and monitoring plan with an independent medical safety officer who is not involved in the research. A copy of the WAVES consent form is available at: https://medsites.vumc.org/centerforasthmaresearch/waves-study.

Eligibility criteria: eligible participants are healthy pregnant women receiving care in one of our academic medical centres affiliated practices and who intend to deliver at our medical centre. Inclusion and exclusion criteria are provided in online supplemental table 2.

Study overview: WAVES enrollment is during pregnancy, with a maternal pregnancy visit and birth delivery visit. Child visits are scheduled at birth and ages 2, 4, 6, 12, 18, 24, 36, 48 and 60 months. Children are followed annually for wheeze and asthma outcomes captured using the International Study of Asthma and Allergies in Childhood questionnaire, a standard validated asthma and allergic disease questionnaire.[14 15] Biweekly active and passive surveillance is conducted during the first year of life during RSV season. RSV is identified using polymerase chain reaction (PCR) for RSV detection, RSV serology at age 1 year and statistical modelling to identify age of infection not identified by PCR or RNA-seq are used to identify RSV infection and age of infection.[16] An objective, ordinal respiratory severity score is used to assess respiratory illness severity.[16 17] Nasal sampling at all time points includes collection for NAEC culture, nasal brush for bulk RNA-seq and nasal brush for NAEC preservation using 10× Chromium Single Cell Gene Expression Flex technology (10x Genomics, Pleasanton, CA) to fix and bank the biospecimens, which is crucial for obtaining high-quality data for single-cell analyses at birth and ages 4, 12 and 24 months while minimising batch effects. Testing includes specific IgE and skin testing at age 5 years. Figure 2 provides a longitudinal overview of study components.

Visits: each of the study visits is described in the online supplemental file 1. For exact details on specimen collection at each timepoint, refer to figure 2 and online supplemental table 3. To decrease participant burden, visits are scheduled in combination with regular obstetric and paediatric appointments. The designated visit windows for each scheduled visit in the first year of life are ±60 days, with a minimum of 28 days between visits. A visit may also be split into multiple visits to the research unit or home, or combination of phone and in-person contact.

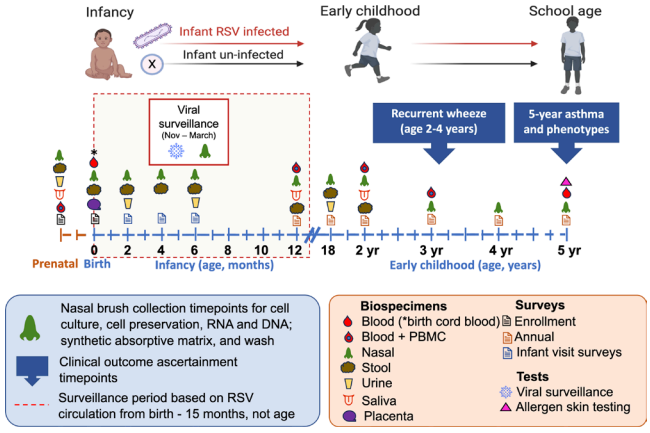

Figure 2 WAVES birth cohort timepoints, measurements and clinical outcomes. PMBC, peripheral blood mononuclear cells; RSV, respiratory syncytial virus.

In addition, RSV biweekly surveillance and illness visits are conducted from birth to 15 months beginning at the start of RSV season, as determined by RSV circulation in middle Tennessee using viral surveillance data, as further detailed in the online supplemental file 1.

Biospecimen collection: some specimens are collected in-person, and some are collected by the caregivers at home using remote collection kits and then returned to the study site. Specific details of nasal biospecimen collection are detailed below.

Surveys: while surveys may be completed during in-person study visits, most participants choose to respond to surveys at home via a secure Research Electronic Data Capture (REDCap) web portal. Additional paper copies of questionnaires or links to REDCap surveys are used as needed.

Scheduled nasal sample collections

Scheduled nasal sample collections from the biological mother and longitudinally from the infant beginning at birth are intended for NAEC culture, preservation for scRNA-seq, bulk RNA-seq, NAEC metabolomics, DNA methylation and microbiome. These are detailed in online supplemental file 1.

Blood samples

Cord blood is collected at birth, and child blood samples at 1, 2, 3 and 5 years of age. Planned sample analyses include RSV serology, absolute eosinophil count, allergen-specific IgE and genomic analyses on nasal, blood or saliva samples.

Normal airway epithelial cell study outcomes

We have a priori proposed three study outcomes focused on understanding airway epithelial development and the impact of infant RSV infection: (1) comprehensive longitudinal analysis of the developing airway epithelium that identifies developmental trajectories and endotypes associated with asthma and infant RSV infection status, (2) defining the longitudinal changes in the metabolism of AECs in infants that precede development of asthma in childhood including RSV infection status, (3) determining the effect of infant RSV infection on airway epithelial DNA methylation and subsequent asthma risk through identification of differentially methylated cytosine-phosphate-guanine sites in longitudinal nasal airway epithelial sample cell types.

Clinical study outcomes: clinical outcome definitions in the WAVES birth cohort are outlined in table 1.

Covariate assessment

Data on potential confounders to be included in statistical analyses as covariates are captured by detailed parental questionnaires or anthropometric measurements obtained longitudinally throughout the study. Child covariates include birth weight, gestational age, sex, race and ethnicity, body mass index percentile, other respiratory viruses in infancy, daycare, siblings, second-hand smoke exposure, breastfeeding, exposure to antibiotics including at nasal brush collection, health insurance and pets. Maternal covariates include pregnancy-related conditions (eg, gestational diabetes, pre-eclampsia and gestational weight gain), chronic conditions (eg, asthma, diabetes, obesity and allergic rhinitis) and demographic variables (eg, education, housing and employment) and socio-economic status latent variables via geocoding of linked addresses with neighbourhood variables. Maternal RSV vaccine and infant RSV immunoprophylaxis were introduced during the study period and are captured by report and electronic health record.

Power

Our preliminary results from scRNA-seq experiments have demonstrated its sensitivity in characterising cell differentiation and identifying related differential gene profiles.[8] In the WAVES high-risk cohort, we anticipate that 35% of children will develop asthma (n=60 with asthma allowing 15% attrition, 153 without asthma allowing 15% attrition); among those with asthma, half will have non-allergic and half will have allergic asthma. The estimated asthma prevalence in WAVES is thus consistent with ours and other established paediatric cohorts, which will provide more than enough children to power these studies as discussed below.[7 18 19] RSV infection is a natural quasi-randomisation exposure, as unlike other respiratory viruses, we and others have demonstrated that approximately half of infants have RSV infection and approximately half are uninfected in the first year of life.[9 20 21] In our bulk RNA-seq study, we will analyse transcriptomic data from 810 samples across longitudinal time points. These samples include those used for scRNA-seq analysis. Using the *RnaSeqSampleSize* tool and considering sample sizes ranging from 5 to 20, the power analysis shows that with five samples for scRNA-seq, we will have >90% power for detecting genes

Table 1 Definitions for childhood respiratory outcomes

Respiratory outcome	Outcome definition
Recurrent wheeze	Defined as two or more episodes of wheeze in the past 12 months assessed by parental or physician report between ages 2 and 5 years
Asthma	Physician-diagnosed asthma ever as confirmed in the EHR and either (a) wheeze in the last 12 months assessed by parental or physician report at age 5 years or (b) use of asthma medications in the last 12 months assessed by parental report or EHR pharmacy data at age 5 years
Allergic asthma	(1) **Childhood asthma**: as above **AND** (2) **Aeroallergen sensitisation**: defined as a positive (≥0.35 kU/L) specific IgE to at least one common aeroallergen in blood obtained at age 5 years*
Non-allergic asthma	(1) **Childhood asthma**: as above **AND** (2) **No aeroallergen sensitisation**: defined as a negative (<0.35 kU/L) specific IgE to common aeroallergens in blood obtained at age 5 years
No asthma (controls)	(1) Data available for all assessment criteria for childhood asthma and aeroallergen sensitisation **AND** (2) never meeting the definition of asthma, recurrent wheeze (defined as two or more episodes of wheeze in the past 12 months) during ages 2 to 5 years

*In sensitivity analyses, we will also use specific IgE positivity to two or more common aeroallergens in blood at age 5 years to define aeroallergen sensitszation and allergic asthma.
EHR, electronic health record.

across two contrasting conditions with at least a threefold change in gene expression differences, 5% false discovery rate (FDR), a maximum dispersion of 1e-4 inferred from our preliminary bulk RNA-seq data and an average read count of 10 per gene.[22] With our proposed 15 samples, we have >99% power to detect genes with a minimum of twofold difference, assuming a 5% FDR, a maximum dispersion of 1e-4 and an average of 10 read counts per gene. The participants in the general-risk cohort were not considered in the power calculations and are included to assess baseline and developmental characteristics in descriptive analyses compared with children born to parents with asthma in the high-risk group on which the study is powered. In all analyses in which they are included, we will adjust for parental asthma risk status.

Approaches for the assessment of statistical power in complex human single-cell genomic studies are still evolving and remain challenging, as cellular frequency, cell-to-cell variance in gene expression levels within cell population and read depth/cell influence power. To circumvent such known issues and ensure robust analysis of epithelial subset data with variable cell frequency, scRNA-seq data pseudobulking approach will be used, as is concordant with ours and others recently published best practices.[23–25] Our integrated dataset will contain transcriptomes of ~1.2 million cells from >120 individuals, indicating that for rare cell subsets with frequencies as low as 1%, an average of >100 cells/subject will be present. Our preliminary published data of basal and precursor cell subsets (at least 50–100 cells and 5000 reads/subset in a sample) in paediatric epithelial samples indicate that the use of a pseudobulk strategy (ie, DESeq2 on Regularised Log (Rlog) sum of counts pseudobulked data) results in similar power as would be estimated for bulk RNA-seq studies described above.[8] With 10 subjects per group/time point (ie, allergic asthma vs controls), we will have at least 80% power in detecting a cell type–specific differentially expressed gene with logFC >0.5 at an FDR<0.05.

Analysis

An overview of the nasal sampling types, biospecimen processing and analyses is summarised in figure 3. scRNA-seq and RNA-seq will be performed on the same nasal brush samples from the WAVES cohort at birth and ages 4, 12 and 24 months (scRNA-seq) and at birth and ages 2, 4, 6, 12, 18, 24, 36 and 48 months (bulk RNA-seq). These time points are supported by our preliminary data in which we found significant age-related changes

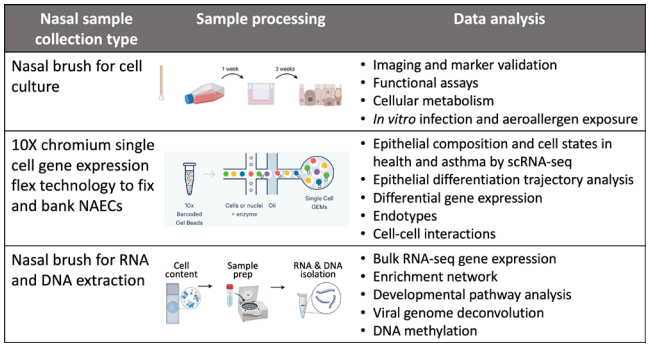

Figure 3 Plan for nasal sample collections, processing and data analysis. NAECs, nasal airway epithelial cells; scRNA-seq, single-cell RNA sequencing; RNA-seq, RNA sequencing.

in epithelial development and differences between those with recurrent wheeze and never wheeze (controls). scRNA-seq approaches will allow us to identify significant differences in epithelial subsets due to development and between asthma phenotypes. scRNA-seq brush data will be used as a reference dataset to deconvolve the bulk RNA-seq data.[26] Cell-type proportion estimates derived from this process, in combination with the CellDMC statistical framework,[27] will be employed to infer cell type–specific signals in AECs across the entire early childhood developmental trajectory (ages 0, 2, 4, 6 months and 1, 1.5, 2, 3 and 4 years).[28 29] Then, we plan to reduce dimensionality of genomic data to gene sets guided by enrichment analysis, followed by collapsing expression of each gene set of interest to a single variable, which measures standardised expression levels of the set of genes in a pathway. Following these variables longitudinally will allow for reconstructing a trajectory of a biologically meaningful event over time. For this purpose, we will use a longitudinal linear combination test for gene set analysis to perform (*i*) analysis of within-subject variation in identified trajectory and phenotype variables using a functional linear regression model, in which trajectory variables are treated as functional responses defined over a non-linear time domain, incorporating potential time-varying covariates; (*ii*) linear combination test to detect significant effect gene sets and covariates associated with the longitudinal events characterised by trajectory and phenotype variables.[30] We will also consider employing mixed-effects regression modelling (eg, lme4 and nlme) that handles flexible modelling of time and covariates, as well as repeated measures. This strategy provides rigorous statistical significance testing for trajectory differences between healthy and asthma phenotypes. To address multiple testing in genome-wide and other high-dimensional analyses (eg, RNA-seq, scRNA-seq and DNA methylation), we will control the FDR, for example, using the Benjamini–Hochberg procedure.

Sources of bias and mitigation

We designed a prospective cohort study in which exposure is collected before outcomes occur, providing a standard approach to reduce recall bias. Inclusion and exclusion criteria are clearly outlined, and both exposures and outcomes are objectively collected, defined and processed. To decrease selection bias, all eligible participants with data on the exposure and outcomes will be included; imputation will be performed if covariates are nonrandomly missing, and sensitivity analyses will be performed to evaluate the effects of missing data and potential misclassification of exposure. Attrition will be kept to a minimum through rigorous retention methods, and the characteristics of those lost to follow-up and those retained will be examined to assess potential impacts.

Confounding bias: our statistical analysis will control for potential covariates (ie, age and sex) using multivariable logistic regression models when assessing associations. Multicollinearity of variables will be assessed to support data dimension reduction and ensure model stability. Interaction terms will be considered as appropriate. Additional correction for systematic, unobserved batch or population variation confounding will be adjusted with our group's latent factor correction methodology,[31] which has been applied and validated in our previous work on molecular profiles of the airway epithelia.[10]

Internal validation: further in our statistical analyses, validation of findings will undergo cross-validation, generated from primary sample datasets, for example, to determine the association of unbiased developmental trajectory patterns with wheeze and RSV outcomes; Monte Carlo approaches will be used to simulate model outcomes and determine signal-to-noise ratios.[32 33]

External and cell composition validation: as part of the NIH Children's Allergy and Asthma Data Repository (CADRE) birth cohort data repository focused on asthma and allergic diseases, we have approval to use bulk RNA-seq data from other CADRE cohorts for validation (https://cadre.med.wisc.edu/). We will also include a subset of bulk RNA-seq samples to conduct orthogonal validation at a single critical time point in development identified through our analyses.

Functional validation: barrier function (permeability, junctions, cilia and mucus) will be measured in differentiated ALI as endpoint over time. Key gene markers and pathways will be validated in air–liquid interface (ALI) using qPCR and imaging approaches. Targeted isotope tracer (fluxomics) and Biolog metabolism assays will be used for functional validation of metabolomics analysis.

ETHICS AND DISSEMINATION

This observational longitudinal study has been approved by VUMC IRB. Study results will be communicated in peer-reviewed scientific publications and to study participants and the public.

Data sharing. Sharing of deidentified multi-omic data will be through depositing data in a public data repository per NIH requirements, such as NIAID ImmPort and dbGaP (Database of Genotypes and Phenotypes). ImmPort and dbGaP are data repositories funded by the NIH. We will deposit immunologic data into NIAID ImmPort, which is a portal designed for immunological research data, and genomic data into dbGaP (Database of Genotypes and Phenotypes).

SUMMARY

Our comprehensive and multi-disciplinary approach brings together a complementary group of investigators with a breadth of expertise, and this study will provide significant insights into the biology underlying the early-life origins of asthma, identifying airway endotypes and phenotypes with unprecedented precision and the role of early-life respiratory viral infection. This proposal has four fundamental innovative aspects that will advance the

field: (1) a unique systems biology approach and focus within the airway epithelial compartment and cell subtypes beginning at birth and longitudinally to define asthma heterogeneity, (2) the novelty and importance of the study objectives delineating both normal and asthmatic airway development and the impact of early-life respiratory viral infection, (3) creation of a unique airway epithelial cell shared resource and data for the scientific community and (4) the potential to lead to clinically important preventive and therapeutic interventions for asthma that target airway biology beginning at birth.

Key WAVES study outcomes include (1) identifying cell type-specific epithelial differentiation defects and their functional vulnerabilities to early-life RSV infection; (2) uncovering the timing and mistiming of critical developmental processes; (3) pinpointing cell types, developmental windows and biological pathways sensitive to epigenetic reprogramming; (4) determining metabolites that regulate epithelial development; (5) elucidating the impact of early-life respiratory viral infections on developmental trajectories; (6) defining precise postnatal windows that heighten epithelial barrier susceptibility; and (7) characterising RSV-induced epigenetic modifications, their persistence and link to childhood asthma phenotypes. These findings will transform our understanding of asthma pathogenesis and pave the way for biologically targeted prevention and treatment.

Our state-of-the-art longitudinal integration of single-cell, gene expression, metabolomic and epigenetic data at epithelial subset resolution is a necessary systems biology approach to understand the complex and heterogeneous development of the airway epithelium in health and specific asthma phenotypes and to delineate the impact on airway epithelial development of a common early-life asthma risk factor, RSV infection. Our WAVES birth cohort includes the following unique longitudinal and frequent assessments of the airway beginning at birth: (1) 10X Genomics Flex cell preservation for single-cell analyses, (2) cell culture for the functional study of metabolism and in vitro infection and (3) assessment of basal-to-specialised epithelial ALI differentiation in culture to understand developmental pathways involved. Our proposed metabolism studies integrated with information from other epithelial 'omics' in the postnatal developmental time frame will provide the necessary mechanistic systems biology framework for complete understanding of developmental reprogramming (figure 3).

This novel longitudinal study aims to address the development of the airway epithelial barrier beginning from birth using a multi-dimensional framework. By doing so, this study will address airway epithelial changes present at birth and the developmental susceptibility component of asthma at the epithelial subset-specific level by identifying normal and aberrant airway epithelial postnatal differentiation trajectories preceding the development of allergic and non-allergic asthma phenotypes and the impact of a common environmental asthma risk factor, RSV infection. This integrated approach is critical to unravelling the developmental origins of childhood asthma.

We will create and share novel multi-omics reference models for cost-effective reconstruction of NAEC postnatal developmental trajectories. This will be an invaluable data resource to the scientific community, particularly for studies of epithelial developmental reprogramming in large cohort studies of respiratory disease.

There are few effective primary preventive interventions for asthma, nor therapies for asthma based on the biology of asthmatic airway epithelial development for the vast majority of children with asthma. Identifying the timing and pathways driving allergic, non-allergic and normal airway epithelial development and the role of early-life infection may inform novel interventions for prevention and treatment(s) that target regulation of the normal development of the early-life airway epithelium.

Author affiliations
[1]Department of Medicine, Center for Asthma Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[2]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[3]Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[4]Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA
[5]Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[6]Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[7]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[8]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[9]Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, USA

Acknowledgements We would like to express our heartfelt gratitude to the families participating in the WAVES study for their time and commitment. We also extend our appreciation to the clinical staff in the Vanderbilt University Medical Center obstetrical clinics, paediatric clinics and labour and delivery for their invaluable support in facilitating this research. Additionally, we would like to thank the following WAVES team members for their dedication and expertise which are instrumental to the success of this work: Hiam Abdala-Valencia, Talissa Blackshere, Ferdinand Cacho, Lauren Chandler, Teresa Chipps, Kamry Coffey, Thi Duong, William Dupont, Haley Hickham, Loveis Jackson, Hannah Krause, Ally McBryar, Brittany O'Brian, Barron Patterson, Jeremy Richter, Christopher Sutton, Kisha Turner and Makenna Woods.

Contributors TH: study conceptualisation and design, funding acquisition, methodology, writing the original manuscript draft, review and editing of the manuscript, project administration, supervision, visualisation, interpretation of data and writing the original draft. CR-S: methodology, review and editing of the manuscript and interpretation of data. DN: biospecimen management, data curation, methodology, editing and review of the manuscript, laboratory management and administration. BS: methodology, interpretation of data, editing and review of the manuscript. SB: methodology, data analysis, editing and review of the manuscript and interpretation of data. CM: study power, methodology, interpretation of data and editing and review of the manuscript. SO: methodology and editing and review of the manuscript. ZL: methodology, data management, interpretation of data and editing and review of the manuscript. ET: data and biospecimen acquisition, supervision and editing and review of the manuscript. WF: project administration, data and biospecimen acquisition, supervision and editing

and review of the manuscript. PS: data and biospecimen acquisition, supervision and editing and review of the manuscript. SR: biospecimen management, project administration and editing and review of the manuscript. KP: biospecimen management and editing and review of the manuscript. JC: biospecimen management, methodology, project administration, data curation and editing and review of the manuscript. SK: methodology and data curation. KM: methodology and data curation. SM: methodology, interpretation of data and review and editing of the manuscript. LJA: conceptualisation, investigation, methodology and editing and review of the manuscript. TG: study design and power, methodology, review and editing of the manuscript, data management and curation and interpretation of data. All authors reviewed and approved the final manuscript. TH is the guarantor.

Funding This work was supported by NIH / NIAID U19 AI 095227 and the VUMC Center for Asthma Research.

Competing interests The authors have the following relevant competing interests: TH reports grants from the NIH, serving as co-chair of the American Thoracic Society Vaccines and Immunization advisory group, and as DSMB member for RSV vaccines for Pfizer. CRS reports grants from the Francis Family Foundation and the National Institutes of Health (NIH), personal fees from AstraZeneca and The KOL Connection, and presentation or writing fees from the American Academy of Pediatrics; the American Academy of Allergy, Asthma, and Immunology; and the American Association for Respiratory Care outside the submitted work. LJA reports paid consultancies on RSV vaccines for GSK, Janssen and AstraZeneca and on RSV antiviral drugs for Enanta and on influenza virus vaccines for Pfizer and our laboratory is currently receiving funding through Emory University from Pfizer for laboratory studies for RSV surveillance studies in adults and maternal antibody transfer to the infant and Sciogen for animal studies of RSV vaccines and previously received funding from Vernagen, LLC for animal studies of RSV and other vaccines. Our laboratory has received funding from an NIH SBIR (#R41AI167226) through Razi Pharmaceuticals, LLC. I am co-inventor on several CDC patents on the RSV G protein and its CX3C chemokine motif relative to immune therapy and vaccine development and on a device to capture coughed droplets from the lung and co-inventor on Emory patent filing for use of RSV platform VLPs with the F and G proteins for vaccines and G protein constructs for RSV vaccines. The other authors have no disclosures. These funders had no role nor influence on the study.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Vanderbilt University Medical Center Institutional Review Board, Ethics approval # 222277 Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iD

Tina Hartert https://orcid.org/0000-0001-7470-1166

REFERENCES

1 Pividori M, Schoettler N, Nicolae DL, *et al*. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med* 2019;7:509–22.
2 Melén E, Zar HJ, Siroux V, *et al*. Asthma Inception: Epidemiologic Risk Factors and Natural History Across the Life Course. *Am J Respir Crit Care Med* 2024;210:737–54.
3 Scott M, Roberts G, Kurukulaaratchy RJ, *et al*. Multifaceted allergen avoidance during infancy reduces asthma during childhood with the effect persisting until age 18 years. *Thorax* 2012;67:1046–51.
4 Chan-Yeung M, Ferguson A, Watson W, *et al*. The Canadian Childhood Asthma Primary Prevention Study: outcomes at 7 years of age. *J Allergy Clin Immunol* 2005;116:49–55.
5 Bisgaard H, Stokholm J, Chawes BL, *et al*. Fish Oil-Derived Fatty Acids in Pregnancy and Wheeze and Asthma in Offspring. *N Engl J Med* 2016;375:2530–9.
6 Larkin EK, Gebretsadik T, Moore ML, *et al*. Objectives, design and enrollment results from the Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure Study (INSPIRE). *BMC Pulm Med* 2015;15:45.
7 Rosas-Salazar C, Chirkova T, Gebretsadik T, *et al*. Respiratory syncytial virus infection during infancy and asthma during childhood in the USA (INSPIRE): a population-based, prospective birth cohort study. *Lancet* 2023;401:1669–80.
8 Berdnikovs S, Newcomb DC, Haruna N-F, *et al*. Single-cell profiling demonstrates the combined effect of wheeze phenotype and infant viral infection on airway epithelial development. *Sci Adv* 2025;11:eadr9995.
9 Lawless D, McKennan CG, Das SR, *et al*. Viral Genetic Determinants of Prolonged Respiratory Syncytial Virus Infection Among Infants in a Healthy Term Birth Cohort. *J Infect Dis* 2023;227:1194–202.
10 McKennan C, Newcomb DC, Berdnikovs S, *et al*. Impact of early-life respiratory syncytial virus infection on cell type-specific airway dna methylation. *Genomics* 2024.
11 Tennessee Department of Health. General health data: population: TN DOH. 2024. Available: https://www.tn.gov/health/health-program-areas/statistics/health-data/population.html [Accessed 14 Nov 2025].
12 Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;18:805–35.
13 von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
14 Asher MI, Keil U, Anderson HR, *et al*. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Respir J* 1995;8:483–91.
15 Ellwood P, Asher MI, Beasley R, *et al*. The international study of asthma and allergies in childhood (ISAAC): phase three rationale and methods. *Int J Tuberc Lung Dis* 2005;9:10–6.
16 McCallum GB, Morris PS, Wilson CC, *et al*. Severity scoring systems: are they internally valid, reliable and predictive of oxygen use in children with acute bronchiolitis? *Pediatr Pulmonol* 2013;48:797–803.
17 Golan-Tripto I, Goldbart A, Akel K, *et al*. Modified Tal Score: Validated score for prediction of bronchiolitis severity. *Pediatr Pulmonol* 2018;53:796–801.
18 Bacharier LB, Beigelman A, Calatroni A, *et al*. Longitudinal Phenotypes of Respiratory Health in a High-Risk Urban Birth Cohort. *Am J Respir Crit Care Med* 2019;199:71–82.
19 Stoltz DJ, Jackson DJ, Evans MD, *et al*. Specific patterns of allergic sensitization in early childhood and asthma & rhinitis risk. *Clin Exp Allergy* 2013;43:233–41.
20 Zylbersztejn A, Pembrey L, Goldstein H, *et al*. Respiratory syncytial virus in young children: community cohort study integrating serological surveys, questionnaire and electronic health records, Born in Bradford cohort, England, 2008 to 2013. *Euro Surveill* 2021;26.
21 Nyiro JU, Kombe IK, Sande CJ, *et al*. Defining the vaccination window for respiratory syncytial virus (RSV) using age-seroprevalence data for children in Kilifi, Kenya. *PLoS One* 2017;12:e0177803.
22 Zhao S, Li C-I, Guo Y, *et al*. Package 'RnaSeqSampleSize' [dataset]. 2023. Available: https://ftp.gwdg.de/pub/misc/bioconductor/packages/3.17/bioc/manuals/RnaSeqSampleSize/man/RnaSeqSampleSize.pdf
23 Murphy AE, Skene NG. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nat Commun* 2022;13:7851.
24 Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018;50:1–14.
25 Squair JW, Gautier M, Kathe C, *et al*. Confronting false discoveries in single-cell differential expression. *Nat Commun* 2021;12:5692.

26 Cobos FA, Panah MJN, Epps J, *et al*. Effective methods for bulk RNA-seq deconvolution using scnRNA-seq transcriptomes. *Genome Biol* 2023;24:177.
27 Zheng SC, Breeze CE, Beck S, *et al*. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* 2018;15:1059–66.
28 Stuart T, Butler A, Hoffman P, *et al*. Comprehensive Integration of Single-Cell Data. *Cell* 2019;177:1888–902.
29 Butler A, Hoffman P, Smibert P, *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20.
30 Khodayari Moez E, Hajihosseini M, Andrews JL, *et al*. Longitudinal linear combination test for gene set analysis. *BMC Bioinformatics* 2019;20:650.
31 McKennan C, Nicolae D. Accounting for unobserved covariates with varying degrees of estimability in high-dimensional biological data. *Biometrika* 2019;106:823–40.
32 Phelan KJ, Roskin KM, Burkle JW, *et al*. Early-life wheeze trajectories are associated with distinct asthma transcriptomes later in life. *J Allergy Clin Immunol* 2025;156:640–50.
33 Robert C. *Monte Carlo statistical methods*. 2nd edn. New York: Springer-Verlag, 2004:645.