



PhD Thesis

Kasper Fischer-Rasmussen, MSc

Genetic Heterogeneity in Early Childhood Wheeze

Author: Kasper Fischer-Rasmussen

Title: Genetic Heterogeneity in Early Childhood Wheeze

Principal supervisor: Klaus Bønnelykke, MD, PhD¹

Principal co-supervisor: Anders Gorm Pedersen, PhD²

Co-supervisor: Anders Ulrik Eliassen, PhD^{1,3,4}

Affiliations: ¹COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark

²Department of Health Technology, Section for Bioinformatics, Technical University of Denmark, Kongens Lyngby, Denmark

³Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁴Department of Pediatrics, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA

Assessment committee: Allan René Linneberg, MD, PhD, Professor (Chair)
Institute of Clinical Medicine
University of Copenhagen, Denmark

Qingling Duan, PhD, Associate Professor at the Department of Biomedical and Molecular Sciences & School of Computing, Queen's University, Canada

Michael H. Cho, MD, Assistant Professor at Channing Division of Network Medicine, and Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Submitted: 15th of January 2025

List of Papers

My PhD thesis is based on three scientific papers that highlights investigations into wheeze genetic risk factors for childhood wheeze. The papers are listed below.

- I. **Kasper Fischer-Rasmussen**, Raquel Granell, Anders Ulrik Eliassen, Eskil Kreiner, Casper-Emil Tingskov Pedersen, Yang Luo, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, David Strachan, Ashish Kumar, Anne-Marie Nybo Andersen, Bjarke Feenstra, Frank Geller, Valérie Siroux, Florence Demenais, Emmanuelle Bouzigon, Vincent Jaddoe, Ralf van der Valk, Liesbeth Duijts, Jordi Sunyer, Mónica Guxens, Marcella Marinelli, Mariona Bustamante, Joachim Heinrich, Marie Standl, John Curtin, Chih-Mei Chen, Wenche Nystad, Bo Jacobsson, Ronny Myhre, Craig E Pennell, Denise Daley, Carole Ober, Dorret Boomsma, Abdel Abdellaoui, Angela Simpson, Clare Murray, John W Holloway, Sam Collins, Steve Turner, S Hasan Arshad, Anhar Ullah, Erik Melén, John Henderson[†], Hans Bisgaard[†], Anders Gorm Pedersen, Adnan Custovic, Judith M. Vonk, Gerard H. Koppelman, Michael Kabesch, Klaus Bønnelykke, GABRIEL consortium, the EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium
Genetic Characterization of Preschool Wheeze Phenotypes
Submitted
- II. **Kasper Fischer-Rasmussen**, Casper-Emil Tingskov Pedersen, Luo Yang, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, Per Magnus, Yunsung Lee, Marianna Niemi, Tarja Laitinen, Þórunn Olafsdóttir, Guðmar Þorleifsson, Ingileif Jonsdóttir, Christian Erikstrup, Lotte Aas Hindhede, Carole Ober, Zhaozhong Zhu, Carlos A. Camargo, Anders Børghlum, Anders Gorm Pedersen, Anders Ulrik Eliassen & Klaus Bønnelykke
Genome-Wide Association Study of Severe Preschool Wheeze Uncovers Novel Insights into the Pathobiology
Manuscript in preparation
- III. **Kasper Fischer-Rasmussen**, Casper-Emil Tingskov Pedersen, Astrid Sevelsted, Luo Yang, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, Anders Børghlum, Thomas Munk Laursen, Bjarni Vilhjálmson, Anders Gorm Pedersen, Anders Ulrik Eliassen & Klaus Bønnelykke
Early Childhood Trajectories of Inhaled Corticosteroid Prescriptions Implicate the Presence of Wheeze Endotypes
Manuscript in preparation

Additional research

In addition to my three main papers, I have contributed to additional research projects during my PhD. Below is a list of selected unpublished works nearing finalization and those published in peer-reviewed journals or preprints.

- I. Lee-Sarwar KA, **Fischer-Rasmussen K**, Bønnelykke K, Bisgaard H, Chawes B, Kelly RS, Lasky-Su J, Zeiger RS, O'Connor GT, Bacharier LB, Carey VJ.
Omega-3 Fatty Acids Interact with DPP10 Region Genotype in Association with Childhood Atopy.
Nutrients. 2023 May 22;15(10):2416.
- II. Jensen SK, Pedersen CE, **Fischer-Rasmussen K**, Melgaard ME, Brustad N, Kyvsgaard JN, Vahman N, Schoos AM, Stockholm J, Chawes B, Eliassen A., Bønnelykke K
Genetic predisposition to high BMI increases risk of early life respiratory infections and episodes of severe wheeze and asthma.
European Respiratory Journal. 2024 Jan 1.
- III. Saarentaus E, **Fischer-Rasmussen K**, Sliz E, Bizaki-Vallaskangas A, Laitinen T, Toppila-salmi S, Kankaanranta H, Kettunen J, Bønnelykke K, Palotie A, Mäkitie A.
131 genetic loci highlight immunological pathways and tissues in nasal polyposis and asthma.
Submitted.
- IV. Jensen SK, **Fischer-Rasmussen K**, Eliassen A, Hesselberg L, Brustad N, Melgaard M, Werge T, Schoos AM, Stockholm J, Chawes B, Pedersen CE, Bønnelykke K.
From genotype to endotype in early childhood asthma.
Manuscript in preparation
- V. Frederikke R Skov, Tamo Sultan, **Kasper Fischer-Rasmussen**, Bo L Chawes, Jakob Stockholm, Nilo Vahman, Klaus Bønnelykke, Ann-Marie M Schoos.
Type 2-high airway inflammation in childhood asthma distinguishes a more severe phenotype.
Submitted

Table of Contents

List of Papers	1
Additional research	2
Acknowledgements	6
Abbreviations	7
Additional learning & acknowledgements	9
Supervision	9
Research exchange	9
Use of Artificial Intelligence	10
Summary	11
Resumé	13
Background	15
Wheeze & progression to childhood asthma	15
Genome-wide association studies of wheeze	16
17q12-21	16
Cadherin-related family member 3 (CDHR3)	17
Fucosyltransferase 2 (FUT2)	17
Strategies for dissecting genetic heterogeneity	17
Cross-trait analysis	18
Trait-based SNP clustering	18
Electronic health records	18
Prescription trajectories	19
Genomic measurements	20
Genotyping	20
RNA sequencing	21
DNA methylation	22
Aims and objectives	23
Paper I	23
Paper II	23
Paper III	24
Materials and methods	25
Paper I	25
EAGLE & GABRIEL consortia	25
Wheeze phenotypes	25
GWAS meta-analysis	26

Post-GWAS analyses	26
Paper II	27
Registry based discovery cohorts	27
Wheeze phenotypes	27
COPSAC clinical birth cohorts	28
Wheeze phenotypes	29
Nasal transcriptome & methylome	29
MARC-35 study	30
Wheeze phenotypes	30
Nasal transcriptome	30
The MoBa cohort	31
Wheeze phenotype	31
GWAS meta-analysis	32
Post-GWAS analyses	33
Paper III	35
iPSYCH cohort	35
Wheeze phenotypes	36
Genotypes & quality control	36
Risk factors	37
Latent class analysis	38
Within-class likelihood ratio-test	39
Results	41
Paper I	41
Discovery analyses	41
Analysis of known asthma variants	42
Genetic Similarity Between Adult Traits and Preschool wheeze	43
Paper II	46
GWAS results	46
Heritability and genetic correlations	48
Prioritization of causal genes at genomic risk loci	49
Association-based groupings of severe wheeze risk loci	50
Transcriptome-wide association analysis	52
Clinical & transcriptomic fingerprint of genetic risk score	53
Paper III	56
Latent class analysis	56
Perinatal risk profiles	58
Genetic risk profiles	60
Discussion	63

Main findings	63
Strengths and limitations	64
Interpretation	68
Conclusion	74
Future research	74
References	77
Appendix	91
Paper I	92
Paper II	192
Paper III	288

Acknowledgements

A special thanks to Klaus Bønnelykke, my primary supervisor, for his constant support and mentorship, and for involving me in additional projects that enriched my experience. His trust and encouragement to explore different themes and approaches have been invaluable.

I would like to thank my master's supervisor, Anders Gorm Pedersen, for introducing me to statistical genetics and Bayesian statistics and expanding my perspectives on data analysis. I'm also grateful for his connection to COPSAC, which made my thesis possible.

Thank you to Anders Eliassen for his guidance, from genetic analyses to helping me prioritize tasks and stay focused on the most effective paths forward. I would also like to especially thank the remaining genetics group, Casper and Yang for support, but also the whole of COPSAC for providing serious and useful feedback as well as an abundance of fun and humor during social events and breaks. I would also like to extend my utmost gratitude to the physicians and nurses in the COPSAC clinic and to their unrelenting quest for collecting high quality data - and lots of it! I would also like to thank the families of COPSAC the cohort for their persistent efforts in providing data for invaluable research.

Most importantly, I would like to thank my partner, Melanie, for providing unconditional support, especially during the last and busiest year of my PhD and for traveling with me to Boston while taking care of our children. Thank you to my two children, Andrea and August for being an endless source of joyful mischief and love. Starting a family has been the most rewarding experience during this time, and I would like to dedicate this thesis to them.

Abbreviations

ATC = Anatomical Therapeutic Chemical Classification System

BIC = Bayesian Information Criterion

BLRT = Bootstrapped Likelihood Ratio Test

CA = Current Asthma

COLOC = Colocalization

COPSAC = Copenhagen Prospective Studies on Asthma in Childhood

DBDS = Danish Blood Donor Study

DEPICT = Data-driven Expression Prioritized Integration for Complex Traits

DGIdb = Drug-Gene Interaction Database

EAGLE = EARly Genetics and Lifecourse Epidemiology Consortium

EHR = Electronic Health Records

EW = Early-onset wheeze

ETR = Early-transient redeemers

EWAS = Epigenome-wide association study

FIZI = Functionally informed Z-score imputation

GSMR = Generalised Summary-data-based Mendelian Randomisation

GWAS = Genome-wide association study

ICD = International Classification of Disease

ICS = Inhaled Corticosteroid

LCA = Latent class analysis

LDSC = Linkage Disequilibrium Score Regression

LMRT = Lo-Mendell-Rubin's Test

LRT = Likelihood-ratio test

LRTI = Lower-respiratory tract infections

LOR = Late-onset redeemers

MARC-35 = 35th Multicenter Airway Research Collaboration

MAGMA = Multi-marker Analysis of GenoMic Annotation

MLE = Maximum Likelihood Estimation

MoBa = Norwegian Mother, Father and Child Cohort Study

NCRR = National Center for Register-based Research

NEC = Nasal Epithelial Cells

PC = Principal Component

PoPS = Polygenic priority score

PR = Persistent Redeemers

PRS = Polygenic Risk Score

PTR = Prolonged-transient Redeemers

PW = Persistent wheeze

TEW = Transient-early wheeze

TROLS = Trouble-some Lung Symptoms

TWAS = Transcriptome-wide association study

UKB = UK Biobank

Additional learning & acknowledgements

After completing my master's thesis in February 2021, I joined COPSAC in March of the same year. My experience at COPSAC has been highly rewarding, allowing me to acquire a range of valuable skills. Beyond gaining a deeper understanding of the complex processes involved in conducting and disseminating research, I have also developed practical skills such as effective time management, task prioritization, and communication. In this section, I aim to briefly outline several tasks, while not directly related to research, that have provided important learning opportunities and will here acknowledge the people who helped me grow during these experiences.

Supervision

I have also gained experience with supervision during my PhD. I have been a co-supervisor for two master students in bioinformatics, working partly on the TWAS and EWAS pipelines that I've included in paper II. I would like to thank Tobias Kragsholm and Weizhe Zhang for asking questions that forced me to fine-tune my knowledge in the underlying machinery of the FUSION software. I've gained invaluable experience in efficient communication, consistent follow-up and time-prioritization during these times.

Additionally, I was invited to become a statistical mentor for a clinical PhD student. My role has been to provide sparring on model selection, interpretation of results and potential pitfalls of statistical methods. I would like to thank Frederikke Rosenvinge Skov, for pushing me to clearly and concisely communicate statistical methodology and forcing me to brush up on basic methods and tests. This experience has been invaluable.

Research exchange

I've been fortunate to be invited to spend 3 months in Boston as a visiting PhD student at the research group EMNet at Harvard Medical School & Massachusetts General Hospital. EMNet coordinates the MARC-35 cohort that includes infants hospitalized with bronchiolitis and I was

given the opportunity to explore the various omics layers in their database. I would like to thank Kohei Hasegawa, Zhaozhong Zhu, Heidi Makrinioti, Ryohei Shibata, Hide Miyachi & Carlos Camargo for welcoming me as part of the team during my time in Boston. I am extremely grateful for being granted this experience and I hope to cultivate this collaboration throughout my future career.

Use of Artificial Intelligence

I have used artificial intelligence (AI) for writing assistance throughout my thesis. AI (ChatGPT 4.5) has only been used for suggesting rephrasing of sentences initially written by myself. At no point, during writing have AI been used to generate scientific content or used as an encyclopedia, nor has AI-generated text been directly copy-pasted into this thesis at any point. It has only been used for proofreading and for suggestions on how to rephrase sentences. All AI-generated text has been thoroughly reviewed by myself and corrected and rephrased to match my own writing style.

Additional to text, the front page image has also been generated by AI (Dall-E) using the prompt: *“Create a modernist artwork of how DNA affects wheezing of the lungs, in matte colours on a white background. Add colours and shapes to the DNA molecules so that they look different representing genetic heterogeneity”*.

Summary

Wheezing in preschool children presents a significant risk of progression to childhood asthma and understanding heterogeneity among these symptoms is crucial for optimal prevention and treatment strategies. Genetics is thought to play a detrimental role in asthma progression and this thesis presents three research projects that aim to elucidate the underlying mechanisms of wheeze subtypes in order to understand the genetic heterogeneity of preschool wheeze. The first study is a meta-analysis across 22 birth cohorts defining wheeze trajectories based on prior knowledge of remission patterns. Secondly, hospitalization records of asthma and bronchitis are used to define severe wheeze across 5 nordic biobanks comprising the largest GWAS of preschool wheeze to date. Lastly, prescription records are utilized to define trajectories of medication consumption using latent class analysis and subsequent genetic characterization to elucidate putative endotypes.

In **paper I**, we compiled GWAS results from the EAGLE and GABRIEL consortia, creating the largest GWAS of temporal wheezing phenotypes in preschool children, involving over 18,000 children from 22 studies. This study investigated phenotypes including early-onset wheeze (EOW), transient-early wheeze (TEW), persistent wheeze (PW), and current asthma (CA), with the aim of elucidating genetic differences and similarities between different symptom trajectories. We integrate publicly available GWAS of asthma and asthma-related comorbidities to characterize the genetic overlap preschool wheeze subtypes. We conclude that persistent wheeze is genetically similar with both childhood- and adult-onset asthma, atopic traits and low lung function. Transient-early wheeze displayed little to no association with asthma and atopy, but a significant overlap with low lung function.

In **paper II**, we disregard the timing of wheeze symptoms and include nearly 20,000 preschoolers with severe wheeze defined by at least one hospitalization with asthma, bronchitis or bronchiolitis before the age of seven. We apply a multitude of statistical analyses and GWAS-specific methods to unveil molecular mechanisms involved in risk of severe wheeze. We identified eight novel loci involved in airway inflammation, lung function and growth—five of which replicated in an independent cohort. We identify significant genetic correlations with BMI,

height and c-reactive protein, SNP-groups with distinct associations to atopic traits and lung function, clinical fingerprint and transcriptomic signatures of individuals with high genetic risk severe wheeze. This comprehensive approach aims to improve our detailed understanding of the molecular biology of preschool wheeze and to identify loci with distinct association patterns, representing genetic driving forces behind potential endotypes.

In **paper III**, we employed latent class analysis to identify trajectories in use of inhaled corticosteroids leveraging the Danish National Prescription Registry coupled with the iPSYCH cohort. We subsequently characterize the identified trajectories by their association patterns with perinatal and genetic risk factors. Lastly, we carry out statistical testing to evaluate if risk factors have distinct impacts on the trajectories and our results suggest the presence endotypes of wheeze. We identified an array of distinctly associated risk factors like c-section, smoking during pregnancy, premature birth, *IL-13*, *FUT2* and *HLA-DQA1* and genetic risk for asthma, allergy and eosinophils and a number of putatively distinct associations. These results implicated that the proposed endotypes differ in genetic influence, T2-high response and infection propensity.

Resumé

Hvæsen hos småbørn udgør en betydelig risiko for udvikling af børneastma, og det er afgørende at forstå variationen blandt disse symptomer for at optimere forebyggelses- og behandlingsstrategier. Genetik menes at spille en væsentlig rolle for udviklingen af børneastma, og denne afhandling præsenterer tre forskningsprojekter, der stiler mod at belyse de underliggende mekanismer bag forskellige subtyper af hvæsesymptomer for at forstå den genetiske heterogenitet bag symptomerne. Den første undersøgelse er en meta-analyse af 22 fødselskohorter, der definerer hvæse-forløb baseret på tidligere viden om remission. For det andet bruges hospitalsregistre over astma, bronkitis og bronkiolit til at definere svær hvæsen på tværs af fem nordiske biobanker, hvilket udgør det største genome-wide association study (GWAS) af hvæsen hos småbørn til dato. Endelig anvendes recepter til at definere forløb for medicinforbrug ved hjælp af latent class analysis og efterfølgende karakterisering af perinatale og genetiske risikofaktorer for at belyse evidens for mulige hvæse-endotyper.

I **artikel I** kombinerede vi GWAS-resultater fra EAGLE- og GABRIEL-konsortierne, hvilket resulterede i det største GWAS af tidsafhængige hvæsefænotyper hos småbørn og involverede over 18.000 børn fra 22 studier. Vi undersøgte fænotyper, herunder tidligt opstået hvæsen (EOW), forbigående tidlig hvæsen (TEW), vedvarende hvæsen (PW) og aktuel astma (CA) med det formål at klarlægge genetiske forskelle og ligheder mellem forskellige symptomforløb. Vi integrerer offentligt tilgængelige GWAS resultater af astma og astmarelaterede komorbiditeter for at karakterisere det genetiske overlap med hvæsen hos småbørn. Vi konkluderer, at vedvarende hvæsen genetisk ligner både børne- og voksenastma, atopiske træk og lav lungefunktion. Forbigående tidlig hvæsen viste kun lidt eller ingen sammenhæng med astma og atopiske tilstande, men vi observerede et betydeligt overlap med lav lungefunktion i voksenalderen.

I **artikel II** ser vi bort fra tidspunkterne for hvæsesymptomer og inkluderer næsten 20.000 førskolebørn med svær hvæsen, defineret ved mindst én hospitalsindlæggelse med astma, bronkitis eller bronkiolitis før syvårsalderen. Vi anvender en bred vifte af statistiske analyser og GWAS-specifikke metoder for at løfte sløret for molekulære mekanismer involveret i svær

hvæsen. Vi identificerede otte nye loci involveret i luftvejsinflammation, type-2 diabetes, lungefunktion og vækst, hvoraf fem blev replikeret i en uafhængig kohorte. Vi identificerede signifikante genetiske korrelationer med BMI, højde og CRP, grupper af loci med ensartede associationer til atopiske træk og lungefunktion samt kliniske og transkriptomiske signaturer hos individer med høj genetisk risiko for svær hvæsen. Denne omfattende tilgang sigter mod at forbedre vores detaljerede forståelse af de molekylære mekanismer bag hvæsen hos småbørn og at identificere genetisk loci med distinkte associationsmønstre, der potentielt repræsenterer genetiske drivkræfter bag hvæse-endotyper.

I **artikel III** anvendte vi latent class analysis (LCA) til at identificere receptforløb af inhalationssteroider ved hjælp af det Danske Receptregister koblet med iPSYCH-kohorten. Vi karakteriserede derefter de identificerede forløb ud fra deres associationsmønstre med perinatale og genetiske risikofaktorer. Til sidst udfører vi en statistisk test for at vurdere, om risikofaktorer har forskellige indvirkninger på forløbene, og vores resultater indikerer tilstedeværelsen af endotyper af hvæsen. Vi identificerede en række distinkt associerede risikofaktorer såsom kejsersnit, rygning under graviditet, for tidlig fødsel, polymorphismer i *IL-13*, *FUT2* og *HLA-DQA1* og genetisk risiko for astma, allergi og eosinofile samt en række formodede distinkte associationer. Disse resultater indikerer, at de foreslåede endotyper adskiller sig genetisk ved T2-immunrespons og tilbøjelighed til infektion.

Background

Wheeze & progression to childhood asthma

Asthma is a common condition in childhood (1) and wheeze is likely involved in progression of childhood asthma (2–4). The European Respiratory Society defines wheezing as “a soft polyphonic noise or whistling sound heard mainly during expiration, caused by turbulent airflow occurring simultaneously in many airways of different caliber” (5). Wheezing is often triggered by respiratory viral infections and subsequent development of childhood asthma is likely mediated by Human Rhinovirus (HRV) rather than Respiratory Syncytial Virus (RSV) (6). However, most children experience transient episodes of wheeze, and will not develop asthma (5,7,8) thus understanding the underlying causes of progression will likely aid in limiting over-diagnosis, wasted resources and risk of side-effects (5,9,10).

A family history of asthma is a strong predictor of disease-development and wheezing displays a moderate to high heritability (11). Additionally, a Danish twin-study included twins down until 3 years of age, and saw increasing impact of genetic liability with younger age (12). A small study comparing monozygotic twins reared together and apart, suggested that shared familial environment is a limited confounder of twin-mediated risk of asthma (13). As family studies indicate genetic influence of wheeze, there is still a lack of understanding the specific underlying mechanisms of and only few genetic loci have been investigated (14–18).

Wheezing has limited predictive value for childhood asthma (19), highlighting substantial heterogeneity of the condition. Endotypes based on multi-omics data in infants with bronchiolitis have sought to define more homogenous phenotypes (20–22). One bronchiolitis endotype was characterized by high genetic risk of asthma, high airway abundance of *H. influenzae*, and rhinovirus-A and -C infections and presented with a substantial risk of developing childhood asthma (22), highlighting genetics as a key contributor to wheeze heterogeneity. Furthermore, temporal variations of wheeze have suggested different effects of genetic variants in *GSDMB* and *CDHR3* (14) and resulted in the discovery of the novel locus *ANXA1* (15). Transient early

wheezing has been shown to be influenced by at least three COPD-genes indicating a role in poor adult lung function (23), providing novel support for a genetic endotype not influencing later asthma development (7). Thus genetically heterogeneous wheeze subtypes are already identified motivating further understanding of the molecular underpinnings of wheeze, taking steps towards better prevention and treatment of asthma in childhood.

Genome-wide association studies of wheeze

Genome-wide association studies (GWAS) provide valuable insights into genetics which enables characterization of disease risk mediated by common SNPs (24). GWAS have established that asthma with onset in childhood has substantially higher heritability than disease compared to onset in adulthood (25,26). Furthermore, a number of genetic variants associated with preschool wheeze have been identified, but mainly in association with severe asthma exacerbations (27,28) and only a single study has used wheeze as a phenotype and discovered novel loci (15).

17q12-21

The first GWAS of childhood asthma was published in 2007 and identified genetic variants in the 17q12-21 (17q) locus suggesting regulation of *ORMDL3* levels in lymphoblastoid cells (29). Subsequently the 17q locus has been the most replicated locus for childhood asthma spanning multiple genes including *IKZF3*, *ZPBP2*, *PGAP3*, *ERBB2*, *GSDMA*, *GSDMB* and *ORMDL3* (30). Genotypes in the locus have been shown to affect transcription levels of both *ORMDL3* and *GSDMB*, highlighting core genes in the region (31). *ORMDL3* is involved in regulation of sphingolipid biosynthesis (32) and *GSDMB* is involved in pyroptosis (32) and has been shown to regulate airway inflammation upon viral infection (33). Additionally, the 17q locus has been shown to increase risk of wheeze episodes after infection with HRV, but not upon infection with RSV with significant genotype-independent inductions of both *ORMDL3* and *GSDMB* upon infection (34).

Cadherin-related family member 3 (*CDHR3*)

A GWAS focusing on a severe subtype of childhood asthma, identified a coding variant in the *CDHR3* gene (rs6967330, C529Y) associated with asthma exacerbations and increased cell-surface expression of CDHR3 (27). The CDHR3 protein contains binding pockets for rhinovirus-C (RVC) and the coding variant mediates infection in host cells while increasing the risk of RVC-related illnesses in early childhood (35,36). Moreover, the impact of the *CDHR3* variant is dependent on the 17q locus, and this interaction is linked to the induction of IL-17A in response to viral infection (37).

Fucosyltransferase 2 (*FUT2*)

A GWAS meta-analysis of COPSAC_{severe} and iPSYCH₂₀₁₂ identified *FUT2* as a risk factor for asthma exacerbations and established that *FUT2* is involved in an epistatic interaction with *ABO*. *FUT2* is associated with early-onset asthma but not with asthma in later childhood. *FUT2* encodes the enzyme fucosyltransferase 2, which determines the ability to secrete the soluble ABO blood group antigens and express them on mucosal surfaces, including the bronchial epithelium. Furthermore, *FUT2-ABO* epistasis increases risk of respiratory infections with *S. pneumoniae* (28). *FUT2* is involved in the susceptibility to both respiratory and gastrointestinal infections in infants (38) and genetic variation in the *FUT2* gene influences infection by norovirus, rotavirus and *H. pylori* (39–41), highlighting an important role of *FUT2* in receptivity to infections in early childhood.

Strategies for dissecting genetic heterogeneity

GWAS is a powerful method to identify the underlying molecular mechanisms and unravel the heterogeneity of wheeze. Interpretation of GWAS findings are not straightforward (24) and subsequent analysis is crucial for determining likely effector genes, association to comorbidities or disease subtypes.

Cross-trait analysis

Asthma has various comorbidities including rhinitis, sinusitis, obesity, infections, respiratory disorders and mental disorders (42,43). To characterize the molecular causes for preschool wheeze it is advantageous to leverage GWAS results from asthma-related comorbidities to identify shared genetic influence (44).

If individual data is available, polygenic risk scores (PRS) may be utilized for a powerful integration of large-scale GWAS of specific subtypes of childhood asthma (45), allowing a flexible modeling approach by inclusion of covariates and interaction terms. In contrast, when only summary statistics are available, as in meta-analyses, investigating genetic correlations between traits can be an unbiased approach (46). Methods for estimating genetic correlation have different assumptions (47) and for smaller GWAS of sub-phenotypes, it might be necessary to rely on methods that do not assume significant heritability of traits.

Trait-based SNP clustering

Large GWAS of broadly defined phenotypes have higher statistical power to detect multiple genomic risk loci compared to smaller GWAS of more narrow phenotype definitions (24).

Broadly defined phenotypes consist of several distinct endotypes with different etiologies and the identified loci likely reflect unique molecular mechanisms (14,48). To gain more detailed mechanistic insights, clustering SNPs according to their phenotypic associations with related traits or comorbidities may provide valuable insights into genetic drivers of specific endotypes (49).

Electronic health records

Most studies of wheeze rely on parentally reported questionnaires (8,16,50–52), providing a straightforward and detailed description of wheeze. However, disagreement between parents and doctors on symptoms (53,54), ascertainment bias (47,7), difficulties in wheeze definition and heterogeneous mechanisms attenuates robust interpretation of genetic studies of wheeze (56). These challenges have prompted speculations that electronic health records (EHRs) may provide

more robust definitions of wheeze (56).

EHRs mitigate the dependence on questionnaires and limit the impact of ascertainment bias. Additionally, diagnoses made by healthcare professionals yield more accurate and standardized definitions of disease compared to parentally reported symptoms. Furthermore, data is ascertained upon the patient's encounter with the health care system (57). This provides a significant advantage, as asthma and wheeze symptoms frequently occur in irregular intervals and may not be evident during recruitment visits in a birth cohort (58). A challenge in studying wheeze through EHRs is that no official diagnosis for wheeze exists in the ICD classification system. Defining wheeze from hospital admissions, requires clinical expertise in order to determine which diagnoses accurately represents wheezing symptoms in childhood. Wheeze defined by hospitalization records represents a severe subtype and it has previously been demonstrated that defining asthma from hospitalization records provides a powerful approach for GWAS (27,28).

Prescription trajectories

Apart from severity, timing of symptoms is also an important driver of wheeze heterogeneity (7,59,60). Temporal subtypes have been identified based on parentally reported wheeze, showing differences in asthma progression, sensitization and lung function (7,8,14,55), while prescription patterns of asthma medication in EHRs, have been characterized by progression into adulthood, severity and adverse events (16, 17, 18,19, 26).

Leveraging prescription data allows reliable longitudinal assessment of symptom trajectories, representing milder forms of wheeze compared to hospital admissions. Temporal phenotypes can be inferred by applying clustering methods to longitudinal prescription data. Subsequent risk factor analyses, including genetics, can then provide insights into the presence of endotypes. Latent class analysis (LCA) is a finite mixture model used for categorical data and has previously proven successful in identifying temporal wheeze based on questionnaire data (66). LCA assumes that the observed data arise from a mixture of underlying distributions and Maximum Likelihood Estimation (MLE) is used to find the parameters for the underlying distributions that maximizes the likelihood of the data (67). Fitting LCA models with varying

numbers of classes enables the identification of the most likely number of temporal subtypes of wheeze.

Genomic measurements

To effectively characterize the genetic heterogeneity of wheezing disorders in early childhood we need robust measurements of genomic features to provide reliable results. GWAS is able to characterize how single nucleotide polymorphism (SNP) affects phenotypic traits in large scale studies, where functional genomics measurements such as gene expression and DNA methylation provide evidence for mechanistic roles of disease associated loci.

Genotyping

The Danish Neonatal Screening Biobank (DNSB) contains dried blood spots (DBS) from ~2 million danes born after 1981 where all Danish newborns are screened for a variety of congenital disorders by analyzing blood samples extracted shortly after birth and stored on special filter paper (68). To measure genome-wide genotypes we use the iPSYCH protocol as an example of how genotypes can be extracted from biological samples, quantified and called (69). Despite the small amount of DNA that can be quantified from DBS, it is possible to reliably obtain genome-wide SNP data through unselected DNA amplification by a process called whole-genome amplification (wga) of DNA extracted from the DBS (70). Whole-genome amplification is a PCR-based method using random primers to copy small amounts of DNA in a massive parallel manner in order to provide detectable quantities of DNA for subsequent genotype calling (71). The process of genotyping starts by addition of the fragmented PCR-product (wgaDNA) to a genotyping chip for hybridization with fluorescent probes on the array. In short, SNPs come in three variant genotypes (AA, AB and BB) and they hybridize to a fluorescently labeled sequence on the array. The brightness of the fluorescent signals is proportional to the strength of the binding and clustering algorithm is subsequently applied to partition the samples into three categories, namely AA, AB and BB based on the strength of the fluorescent signal. This corresponds to homozygotes and heterozygotes of each allele (72).

Genotype-calling algorithms are used to convert a fluorescent signal to a genotype (73). SNP arrays typically quantify 100-800 thousands SNPs, and samples are typically excluded if the call-rate is below 95 % and due to discrepancies between genetically inferred and actual sex (69). To improve coverage of the genotyped variants, genome-wide imputation is performed by leveraging reference panels with robust genotyping that leverage the LD structure between variants to infer the genotype of uncalled variants (74). Reference panels such as the HapMap project (75) and the 1000 genomes project (76) are often utilized in imputation and form the basis of most GWAS research to date.

RNA sequencing

An RNA sequencing experiment consists of isolating RNA from biological samples, converting it to complementary DNA (cDNA) using reverse transcriptase, library preparation by PCR-amplification and lastly, sequencing the sample on a next generation sequencing (NGS) platform (77).

RNA is an unstable molecule since RNA-degrading factors (RNAses among others) are prevalent resulting in careful handling of samples in the laboratory (78). For this reason, it is important to quantify the degradation of RNA before sampling, which is often performed using a BioAnalyzer to obtain an RNA integrity number (RIN) between 1 and 10, with the highest number signifying the least amount of degradation. RNA is captured by various methods depending on the research objectives. Capture using poly-T oligos select mRNA (coding RNA) since a key-characteristic of mRNA is that they have a '3-polyA tail and this allows for complementary binding between the molecule and oligo. Commercial kits such as RiboMinus allows for polyA-independent depletion of ribosomal RNA, allowing for capture of both coding and non-coding RNA. Ribosomal RNA is depleted from the sample since it is the most abundant form of RNA and provides little value of cellular mechanisms (79). To prevent further degradation of the sample, RNA must be quickly converted into cDNA, which can be amplified with PCR to construct DNA libraries allowing large quantities of DNA to be robustly sequenced (77).

After sequencing the resulting reads are quality controlled, often using methods such as FastQC (80) or MultiQC (81). Reads are then mapped to the genome and read mapping in

RNA-sequencing experiments is a considerable challenge since reads must be mapped across splice-junctions due to alternative splicing. Finally, gene expression must be estimated through various methods with different considerations that take into account systematic variation such as library size, sequence composition and read depth (how many reads were mapped to the same region). To account for systematic variation it is necessary to normalize read counts by gene length and the number of mapped reads in the sample and this is done by defining a normalizing factor called reads per kilobase of transcript per million mapped reads (RPKM) (77).

DNA methylation

DNA methylation is a well-known epigenetic modification consisting of attachment or removal of a methyl-group from a cytosine (81). In order to quantify DNA methylation at the genome-wide level, extracted DNA samples are treated with bisulfite. This process converts cytosine to uracil, except in the case where cytosine is methylated. Unconverted cytosines are thus a marker of where DNA methylation has occurred in the genome. The bisulfite treated DNA is subjected to whole-genome amplification before being transferred to a Methylation-chip containing CpG beads allowing methylation-dependent fluorescent hybridization (83–85). Quantification of brightness determines the methylation-ratio between 0 and 1 of a site. If a site contains no methylation the value is 0, if a single allele is methylated the value is 0.5 and the value is 1 upon homozygous methylation (83).

Aims and objectives

In this thesis we aim to characterize the genetic heterogeneity of wheeze in three scientific articles. In **paper I**, we leverage GWAS results from wheeze subtypes defined from expert clinical knowledge and use cross-trait analysis to understand genetic enrichment between traits informing about genetic overlap with asthma, atopic traits and lung function. In **paper II**, we define wheeze using early-childhood diagnoses of asthma, bronchitis and bronchiolitis in electronic health records to carry out the largest GWAS of preschool wheeze to date. We identify novel loci, prioritize genes and highlight SNP-clusters based on their associations to relevant asthma-comorbidities. Thirdly, we identify temporal subtypes of wheeze in a large cohort linked to electronic health records. The identified subtypes show distinct associations with perinatal risk factors, candidate SNPs and polygenic risk scores to argue for the presence of distinct endotypes.

Paper I

The aim of this study is to enhance the understanding of genetic susceptibility and the underlying mechanisms of clinically defined subtypes of preschool wheeze through large-scale genome-wide association studies (GWAS). This is done by compiling GWAS of smaller birth cohorts with wheeze based on questionnaires or doctor's diagnosis. The study focuses on estimating genetic correlations to related traits to suggest differences in etiology of wheeze subtypes.

Paper II

This study leverages a broader definition, agnostic of temporal patterns, pooling all individuals with asthma, bronchitis or bronchiolitis diagnosis before age six based on electronic health records. GWAS results are combined across multiple cohorts resulting in ~20.000 cases and ~280.000 controls. We aim to identify molecular mechanisms influencing wheeze, but also to characterize specific risk loci in terms of associations with asthma-related traits. We aim to characterize specific molecular mechanisms by prioritizing genes to genomic risk loci, identify

SNP-clusters with similar associations to asthma-related comorbidities and to integrate expression data from the nasal epithelium.

Paper III

This paper aims to identify distinct subtypes of wheeze through temporal patterns of inhaled corticosteroid prescriptions in electronic health records. We subsequently characterize identified subtypes using perinatal risk factors, known severe asthma loci and polygenic risk scores of asthma-related comorbidities and traits. We employ a statistical procedure to identify distinct associations between the subtypes to provide evidence of potential wheeze endotypes.

Materials and methods

Paper I

EAGLE & GABRIEL consortia

To facilitate the large-scale genetic analyses of various childhood traits and conditions, the EARly Genetics and Lifecourse Epidemiology (EAGLE) consortium was established. The consortium primarily consists of longitudinal, population-based cohorts, although additional cohorts with extensive data on early childhood phenotypes have been included. Many of these cohorts adopt a broad research scope, gathering data on a range of somatic and psychiatric traits in addition to environmental exposures (86).

The GABRIEL Consortium (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community) represents a collaborative initiative with the aim of including smaller studies in large scale genetic research. This coordinated approach has proven effective in assembling sufficient sample quantities to conduct robust genome-wide association studies (GWAS) on asthma (87).

Wheeze phenotypes

Preschool wheeze was characterized by three or more episodes of wheezing or a diagnosis of asthma before the age of 3. Persistent wheeze (PW) included children who experienced preschool wheeze and had at least one wheezing episode or an asthma diagnosis between ages 5 and 7. Contrarily, transient early wheeze (TEW) referred to children who wheezed during preschool but were symptom-free between ages 5 and 7. We included a broader phenotypes encompassing both PW and TEW named early-onset wheeze (EOW). Current asthma (CA) was defined as at least one wheezing episode or an asthma diagnosis between ages 5 and 7, regardless of the age of onset, thus including both PW and children with late-onset wheezing or asthma

(after age 3). Late-onset wheeze or asthma was not analyzed separately due to the low number of cases. All children who did not meet the criteria for any of the case groups were used as controls for the analysis of subsequent phenotypes.

GWAS meta-analysis

Over 13,000 individuals of European descent were recruited from 15 distinct research initiatives, all part of the GABRIEL and EAGLE consortia (86,87). Ethical approval was obtained from local research ethics committees for each study, and informed consent was provided by participants or, where applicable, by the child's legal guardians. Genotyping and subsequent imputation were conducted within each study and association analyses were carried out using logistic regression models, applying an additive model for single nucleotide polymorphisms (SNPs). Variants were excluded from the meta-analysis if they were present in fewer than 3 studies to ensure robustness across cohorts. We also excluded variants with minor allele frequency below 5%, or if they were not found in the 1000 Genomes reference panel (88). A fixed-effects inverse-variance weighted meta-analysis was performed using the METAL software (version released on 2020-05-05) (89).

Variants were selected for the replication phase based on two criteria: a p-value below $5e-5$, and the variant was located in a region not previously associated with asthma. Given the weakened signal at the standard genome-wide significance threshold ($5e-8$), we adopted a more lenient p-value threshold ($5e-5$) in the discovery phase, allowing for approximately 5% false positives among the chosen variants (90). Replication was attempted in over 5,000 children of European descent, recruited from seven additional studies. Informed consent was obtained from parents in each replication study. The replication studies conducted their own genotyping and quality control processes (see Supplementary Tables 2 and 5).

Post-GWAS analyses

Due to attenuated signals at the genome-wide cutoff ($P < 5e-8$), we were unable to characterize the genetic profile of preschool wheeze independently. Instead we sought to estimate the shared

influence for asthma-related comorbidities/traits, allowing interpretations of broad molecular mechanisms by leveraging traits such as asthma ever, childhood onset asthma, adult onset asthma, atopy (atopic dermatitis & allergy), eosinophil count and lung function (7, 8, 9,10, 2, 11, 7,12, 13).

In this study, estimating genetic correlation (shared heritability) using LD-score regression (46) was not feasible. Instead, we used Mendelian Randomization (MR) as an alternative approach to approximate genetic correlation, since the heritability of the wheezing phenotypes was relatively low (see Supplementary Table 15). Prior research has demonstrated that MR methods can be influenced by genetic correlation, as both approaches aim to answer similar questions (47). The related traits were treated as exposures and the wheeze phenotypes as outcomes, and the genetic overlap was interpreted as a log odds ratio for wheeze status for each unit increase of trait-risk (96). Multiple testing correction was performed using the Benjamini-Hochberg method.

Paper II

Registry based discovery cohorts

Severe wheeze was defined by at least one acute hospitalization within the first seven years of life for conditions such as bronchiolitis, bronchitis, asthmatic bronchitis, or asthma. The study utilizes cohorts with genetic data linked to public electronic health registries of inpatient hospitalizations. The discovery meta-analysis included cohorts from Denmark (iPSYCH & the DBDS), Iceland (deCODE (100–102)) and Finland (FinnGen (103)). Cases were included from the registries using ICD-8 and ICD-10 codes (see supplementary material), while controls were children with full follow-up by age six and no hospitalizations for these diagnoses.

Wheeze phenotypes

Wheezing disorders were defined in iPSYCH, FinnGen and DBDS as participants with at least one hospitalization for a primary diagnosis before age six—such as bronchiolitis (ICD-8 code:

913, 390), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), or asthma (ICD-8 code: 493; ICD-10 codes: J45-46). Those without such hospitalizations in that age gap were used as controls. deCODE was the only cohort that included cases based on asthma diagnoses only (ICD-8 code: 493; ICD-10 codes: J45-46). In this cohort one hospitalization for a primary diagnosis before age seven was used as a criteria, excluding those without such diagnoses in that time period. This resulted in the inclusion of 19,461 severe wheeze cases and 284,631 controls across the studies (see Table 1).

COPSAC clinical birth cohorts

The COPSAC₂₀₀₀ cohort is a single-center, prospective clinical study of 411 children born to asthmatic mothers, designed to investigate gene-environment interactions in the development of atopic diseases. Newborns were enrolled within the first month of life and followed up at 1 month, every 6 months until age 7, and again at ages 12 and 18, with additional visits for any respiratory or skin symptoms (104).

The COPSAC₂₀₁₀ cohort is a prospective study involving 700 children born to unselected mothers in Zealand, Denmark, between 2008 and 2011. Initial visits began at 1 week of age, followed by visits at 1, 3, 6, 12, 18, 24, 30, and 36 months, and annually thereafter until age 10. Extra visits were scheduled for respiratory or skin symptoms (105). The COPSAC cohorts were utilized in post-GWAS analysis to examine the clinical impact of identified genetic variants and polygenic risk scores on childhood asthma phenotypes.

Table 1: Cohort case-control numbers of each cohort included in the discovery. Cohorts are all coupled to EHRs based in Denmark, Iceland and Finland.

Study	Phenotype	Time points	Cases	Controls	Total	Case %
iPSYCH	Severe wheeze	0 - 6 yr	9,026	103,479	112,605	8,10 %
FinnGen	Severe wheeze	0 - 6 yr	6,137	110,625	116,762	5,26%
deCODE	Severe asthma	0 - 7 yr	2,671	29,505	32,176	8,30 %
DBDS	Severe wheeze	0 - 6 yr	1,627	41,022	42,649	3,81 %
Total	Severe wheeze	0 - 7 yr	19,461	284,631	304,092	6,40%

Wheeze phenotypes

Asthma was diagnosed longitudinally by research physicians in the COPSAC birth cohorts using a standardized treatment algorithm. The diagnosis required a specific burden of symptoms documented in diaries, a response to treatment with inhaled corticosteroids, and a relapse after withdrawing treatment. Symptom diaries were collected and evaluated by trained physicians during scheduled clinic visits.

Nasal transcriptome & methylome

Biological material was extracted from nasal scrapings from the COPSAC₂₀₁₀-participants during the 6-year visit.

In short, RNA was extracted and quantified using SMART-seq V4 Ultra Low input RNA kit and cDNA libraries were prepared with the Illumina Nextera XT kit, according to the manufacturer instructions. The quality of the RNA expression data was checked using FastQC (80) and MultiQC (81), and all samples passed these tests. FastQ files were mapped to the hg19 genome using STAR, excluding 60 samples without genotype information and 98 with less than 8 million exon-mapped reads. Subsequently 17 outliers were identified and removed since they displayed read counts above 3 standard deviations from the median. This resulted in 381 samples. Genes with less than one count per million were removed in more than 85% of samples including genes on sex chromosomes. The remaining 15,363 genes were normalized using logCPM (counts per million) (106,107).

DNA methylation was profiled using the Illumina 850k EPIC array. To ensure accuracy, the data was cleaned and processed using a software package called minfi in R (108). Probes were removed if they were located on the sex chromosomes, mapped to common SNPs or were assigned to multiple locations of the EPIC array. In total, 707,112 CpG's passed quality control and were normalized to make the measurements comparable across samples. M-values were used for all analyses.

The methods for sample collection, extraction, library preparation, sequencing and quality control have been described in further detail previously (109).

MARC-35 study

The 35th Multicenter Airway Research Collaboration (MARC-35) is a prospective multicenter cohort study on infants hospitalized for bronchiolitis (110). Coordinated by the Emergency Medicine Network (EMNet), an international research collaboration involving 247 hospitals, MARC-35 enrolled infants from 17 sites across 14 U.S. states. The study followed a standardized protocol during three consecutive bronchiolitis seasons (November 1 to April 30) between 2011 and 2014, with a median enrollment age of 3.2 months (111). Bronchiolitis diagnosis followed the American Academy of Pediatrics guidelines (112) and detailed inclusion and exclusion criteria have been published elsewhere (21).

Wheeze phenotypes

The definition of asthma in the MARC-35 study was defined by physician-diagnosis of asthma by age 5 years, including asthma medication use (e.g., albuterol inhaler, inhaled corticosteroids, montelukast) or asthma-related symptoms in the previous year (113). Breathing problems were defined as a child experiencing a cough that disrupts sleep or causes vomiting, or when the child has wheezing or shortness of breath without a cough (114,115).

Nasal transcriptome

We incorporated RNA sequencing data from the MARC-35 cohort, which includes infants hospitalized for bronchiolitis. Nasal epithelial samples for RNA sequencing were taken within 24 hours of hospitalization.

RNA was isolated from nasopharyngeal samples and quantity and quality control was carried out using Agilent Bioanalyzer 2100 (Agilent, Palo Alto, CA) with the RNA 6000 Nano kit. Samples

were depleted of ribosomal RNA with NEBNext rRNA Depletion Kits (New England Biolabs, Ipswich, MA). RNA was sequenced on an Illumina NovaSeq6000 using a S4 100PE Flowcell (Illumina, San Diego, CA). No samples were excluded due to inadequate sequencing depth. Transcript abundances from reads were estimated in Salmon (116) using the human transcriptome (hg38). For the differential gene expression analysis, Salmon's estimated transcript abundances were imported into DESeq2 and normalized to transcripts per million (TPM) (117). Data acquisition processes, including nasal sampling, handling RNA isolation, sequencing and quantification protocols, have been described in further detail (118–120).

The MoBa cohort

The Norwegian Mother, Father, and Child Cohort Study (MoBa) enrolled over 100,000 children at birth, along with 90,000 mothers and 70,000 fathers, over a 10-year period from 50 hospitals across Norway. Participating parents were asked to complete questionnaires, providing comprehensive information about themselves and their children from 6 months of age through to adulthood, beyond 19 years. The cohort and genotyping methods have been previously described in detail (121).

Wheeze phenotype

We utilized the detailed early-life questionnaire from the MoBa cohort coupled with genotype data to replicate findings of our GWAS study (121). We defined asthma and wheeze from 0-7 years by using questionnaires at 6-months, 18-months, 3-year, 5- year and 7-year questionnaires. Children were included in the case group if the parents had answered yes to the following questions in the 6-month questionnaire: “DD245: *Has had asthma?*” or “DD937-DD939: *Has you your child had asthma? Did you go to a doctor/clinic for this? Been admitted to hospital?*” of “DD931-DD933: *Has you your child had RSV/bronchitis? Did you go to a doctor/clinic for this? Been admitted to hospital?*”.

Children were included in the case group if the parents had answered yes to the following questions in the 18-month questionnaire: “EE824-EE825: *Yes, has asthma now or has had*

previously” or “EE197: has asthma now” or “EE299: Had symptoms of wheeze/whistling in the chest”.

Children were included in the case group if the parents had answered yes to the following questions in the 3-year questionnaire: *“GG70-GG71: Child has or has had asthma?”* or *“GG143: Child has had bronchitis between 8 and 18 months”*. The 3-year questionnaire contains no questions about wheeze specifically.

Children were included in the case group if the parents had answered yes to the following questions in the 5-year questionnaire: *“LL58-LL60: Child has had asthma AND it is confirmed by a doctor”* or *“LL64-LL66: Child has experienced wheeze symptoms AND they were confirmed by a doctor”*.

Children were included in the case group if the parents had answered yes to the following questions in the 7-year questionnaire: *“JJ623, JJ16, JJ12, JJ169: Child has had asthma AND it is confirmed by a doctor”* or *“JJ243: Child has had bronchitis AND it is confirmed by a doctor”* or *“JJ635: Child has had wheeze symptoms”*.

Whenever possible, confirmation by a physician was used as an inclusion criterion. We included 19,621 cases following the above criteria, and 29,134 controls. Controls were defined as childrens whose parents had answered no to all questions regarding asthma or wheezing.

GWAS meta-analysis

Plink2.0 was used by respective cohorts, to perform genotype-phenotype association analyses with logistic regression (122), adjusting for gender and the top 10 principal components (PCs) to account for population stratification (123). To combine results across cohorts, we applied inverse variance-weighted meta-analysis using METAL (89). Variants were included if they had a minor allele frequency (MAF) greater than 0.05% and were present in at least 3 cohorts to ensure robust support. A heterogeneity test was conducted to assess variation in effect sizes across the six cohorts.

We replicated findings in the independent Norwegian MoBa cohort, ensuring no overlap with our discovery cohort, which included individuals from Denmark, Iceland, and Finland. The replication phenotype was defined as a diagnosis of asthma, bronchitis, and/or wheeze between 6

months and 7 years of age (see Supplementary for phenotype details). For descriptions of all novel loci, regardless of replication outcomes, refer to the Supplementary material.

Post-GWAS analyses

We mapped SNPs to genes using a consensus score based on 12 different SNP-to-gene features based on distance, functional consequence, DEPICT, PoPS, MAGMA, colocalization with xQTL signals in nasal epithelial cells and OpenTargets V2G-predictor (51, 52, 53, 54–56, 57, 58). We interpreted genes with the majority of features as the most likely causal gene in the locus.

LD-score regression (46,132) was used to estimate SNP-based heritability and shared heritability (genetic correlation, rg) with asthma-related comorbidities and measurements utilizing publicly available GWAS summary statistics (45, 17, 19, 11, 46, 47, 48, 49, 50). We further analyzed each identified wheeze locus for associations with these traits with the aim of identifying SNP-clusters based on phenotypic similarity.

We employed TWAS to explore regulatory mechanisms and provide additional support for loci with increased chance of achieving genome-wide significance in a larger study (139). We also used FUSION for Epigenome-wide association study (EWAS) based on the methylome of nasal epithelium used solely for gene prioritization and not interpreted independently.

Polygenic risk scores (PRS) for severe wheeze were calculated using PRS-cs (45) in the COPSAC and MARC-35 cohorts using plink (140). A phenome-wide association study (PheWAS) was conducted in the COPSAC cohorts to identify clinical and para-clinical outcomes associated with the PRS, involving 944. In the MARC-35 cohort, we examined associations between the wheeze-PRS and asthma at age six, infant breathing problems, and rhinovirus infections in 326 European infants.

To estimate the transcriptomic profile of high-risk individuals, we regressed RNA levels on standardized PRS using limma, using raw CPM values since limma doesn't handle negative values (141). In the COPSAC2010 models, we adjusted for Bioanalyzer RNA concentration, RNA-handling site, RNA integrity number (RIN), gender, age, and the top 5 principal

components. For the MARC-35 models, adjustments included age, gender, and the top 2 principal components. Next, we conducted pathway and GO term enrichment analysis by including all genes with nominal p-values ($P < 0.05$), identifying enriched pathways and GO terms using enrichR (142–144). This analysis was applied to both the COPSAC and MARC-35 cohorts.

We further assessed the predictive ability of the severe wheeze PRS in the MoBa cohort, comparing it with PRSs from three large asthma GWAS studies (16,66, 67). AUC values were calculated for the four PRS models in relation to the wheeze phenotype defined in the replication analysis to determine if the severe wheeze GWAS captured additional genetic information. DeLong's test was used to evaluate significant differences between the PRS models (146).

Paper III

iPSYCH cohort

Data was collected from a subset of participants in the iPSYCH Consortium, which consists of two cohorts: iPSYCH2012, including individuals born between 1981 and 2005 (69), and iPSYCH2015, covering those born between 1981 and 2008 (97). The iPSYCH cohorts are linked to all Danish public registries via the Central Person Registry (CPR), with data access facilitated by Statistics Denmark. The National Center for Register-based Research (NCRR) at Aarhus University manages data transfer and infrastructure.

Prescription data were sourced from the National Prescription Registry (NPR), where each entry reflects redeemed medication, not necessarily whether it was consumed (147). NPR began in 1995, and before 1996, children's prescriptions were registered under their mother's social security number (147). To ensure that we did not accidentally include maternal prescriptions in our analysis, we included individuals born after 1998. For a balance between sample size and follow-up duration, we included iPSYCH participants with complete data for their first nine years of life, excluding those who migrated during this period. The Migration Registry runs until 2017, allowing us to include individuals born in 1998 or later who turned nine by 2017. We included prescriptions for inhaled corticosteroids (ICS) as the main outcome. We also included additional medications for sub-analyses, including beta-2 agonists (B2), nasal steroids, antihistamines, steroid creams, and antibiotics (see Table 2).

Table 2. Medication types included in this study, defined by drug class. All medications were defined from the following Anatomical Therapeutic Chemical (ATC) codes.

Drug class	ATC definition
Inhaled corticosteroid	R03AK06, R03AK07, R03AK08, R03AK10, R03AK11, R03BA01, R03BA02, R03BA05, R03BA07, R03BA08
Beta-2 agonist	R03AC02, R03AC03, R03AC04, R03AC12, R03AC13, R03AC18, R03AL01, R03BB01, R03CC02, R03CC03, R03CC12
Nasal steroid	R01AC01, R01AC02, R01AC03, R01AD01, R01AD04, R01AD05, R01AD08, R01AD09, R01AD11, R01AD12, R01AD58
Antihistamines	R06AE07, R06AA02, R06AX27, S01GX, R06A
Steroid cream	D07AA, D07AB, D07AC, D07AD, D11AH01, D11AH02
Antibiotics	J01CE02, J01CA04, J01CR02

Wheeze phenotypes

We examined trajectories based on sequentially redeemed ICS prescriptions, as a proxy for wheeze symptoms in early childhood. We marked the presence of ICS in each life-year without accounting for the number of prescriptions, as the registry lacks information on actual medication use. Time points (life-years) were organized into 365-day intervals starting from birth. Since multiple prescriptions may not reflect disease severity, we used binary indicators to denote whether any ICS prescription was redeemed within a life-year. We conducted latent class analysis on yearly ICS prescriptions to define wheeze phenotypes in the iPSYCH cohort that differ in timing of symptoms.

Genotypes & quality control

Our analyses were confined to iPSYCH participants with available genotype data (refer to

Supplementary Figure 1). Individuals exhibiting high levels of relatedness or substantial genotype missingness were omitted. Additionally, SNPs were excluded based on the following criteria: elevated missingness (>5%), substantial deviation from Hardy-Weinberg equilibrium (HWE) ($P < 1e-5$), or a minor allele frequency (MAF) below 1%. Comprehensive protocols for sample and genotype quality control have been extensively outlined (148).

Risk factors

For each phenotype, we calculated yearly prevalence and odds ratios (OR) compared to non-prescribers for medications related to asthma comorbidities, including beta-2 agonists, nasal steroids, antihistamines, steroid creams, and antibiotics (see Table 1). A similar analysis was conducted for hospitalizations due to asthma (ICD-8 code: 493; ICD-10 codes: J45-46), wheezing (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J21, J205, J210, J40, J41, J42, J44), and lower respiratory tract infections (ICD-8: 471, 480-486; ICD-10: J100, J110, J12-J18) during the first six years of life.

We examined associations between the trajectories and 22 risk factors related to birth circumstances, socioeconomic status, and familial medication history. We also analyzed parental prescription history for the 10 years prior to the child's birth, focusing on the redemption of prescriptions for asthma (ICS & beta-2 agonists), allergies (nasal steroids & antihistamines), steroid creams, and antibiotics (see Table 3). We assessed the association between known wheeze risk loci identified in previous genome-wide association studies (GWAS) and we explored relationships between polygenic risk scores (PRS) of asthma-related traits.

Table 3. Risk factors included in the study to characterize the identified ICS trajectories. Each category is shown in this table in the left column and the specific covariates are listed to the right.

Category	Covariates
Birth circumstances	Birth weight, gestational age, cesarean section, gender, maternal smoking during pregnancy and neonatal distress
Familia/socioeconomic indicators	Parental age, low parental education, number of siblings and household income
Parental prescription history	Asthma (ICS & b-2 agonists), allergies (nasal steroids & antihistamines), steroid creams & antibiotics
Preschool asthma/wheeze loci	<i>GSDMB (rs2305480-G), CDHR3 (rs6967330-A), FUT2 (rs601338-G), HLA-DQA1 (rs1071630-C), IL-33 (rs340933-G & rs1342326-C), IL1RL1 (rs10189629-C), IL-13 (rs20541-A), ABO (rs505922-C), TSLP (rs1043828-C) & ANXA1 (rs75260654-T)</i>
Polygenic risk scores	Asthma phenotypes (Childhood onset asthma, adult-onset asthma and severe asthma), BMI, FEV1/FVC, immune cells (Eosinophils & neutrophils), Psychiatric traits (Depression, bipolar, autism, anxiety & ADHD), atopic traits (Atopic dermatitis, allergic sensitization and allergic rhinitis) and C-reactive protein levels.
Psychiatric diagnoses	Schizophrenia (ICD-10 codes: F20), bipolar disorder (ICD-10 codes: F30-F31), affective disorder (ICD-10 codes: F30-F39), autism (ICD-10 codes: F84.0, F84.1, F84.5, F84.8 or F84.9) and attention-deficit/hyperactivity disorder (ADHD) (ICD-10 codes: F90.0)

Latent class analysis

We conducted latent class analysis (LCA) using the R package poLCA (149), to identify latent classes in longitudinal ICS prescriptions. Only individuals with at least one ICS prescription in the first 9 years of life were included to avoid class imbalance, as non-ICS users make up 79.2% of the sample. The optimal number of classes was determined using several criteria: Bayesian information criterion (BIC), Lo-Mendell-Rubin (LMR) test (67), bootstrapped likelihood ratio test (BLRT) with 1000 bootstraps (67), model entropy, average silhouette width (ASW), parsimony, and interpretability, ensuring the absence of the “Salsa effect” (150), where

trajectories only differ in prevalence rather than temporal patterns. We did not include covariates during clustering analysis.

To ensure cluster stability, we iteratively removed 20% of the data during model fitting for each number of clusters (k). Furthermore, we downsampled the data from 100% to 10%, applying the same fitting procedure and repeatedly removing 20% of the data at each subset to evaluate the robustness of clustering across varying sample sizes (66).

To characterize the identified clusters, we carried out risk factor analysis using multinomial regression, with clusters as endpoints with non-ICS redeemers as the references group. All analyses were adjusted for gender, cohort (iPSYCH2012 or iPSYCH2015), birth year, and psychiatric status. In the parental prescription history analysis, parental birth year was also included as a covariate, along with the child's birth year. Neonatal respiratory distress was adjusted for gestational age. For the genetic risk factor analysis, we further adjusted for the top 10 principal components (PCs) to account for population stratification. All continuous variables were standardized to a zero-mean and unit standard deviation.

Within-class likelihood ratio-test

To assess whether a given risk factor differs significantly among the four phenotypes, we excluded individuals who never redeemed an ICS prescription. We subsequently fitted two models: one including the risk factor and another excluding the risk factor, otherwise keeping the same covariates in both models. We then compared the models using the likelihood ratio test (LRT). If including the risk factor in the model leads to significant improvement in likelihood, this suggests that the risk factor explains substantial variation among the phenotypes.

We fit the following models to the data excluding the non-redeemer class, to assess if the risk factor explains substantial variation among the phenotypes.

$$M_r: p_i = \beta_0 + \beta_1 r_i + \beta_{cov}^T cov_i + \epsilon_i$$

$$M_{null}: p_i = \beta_0 + \beta_{cov}^T cov_i + \epsilon_i$$

M_r represents the model including the risk factor (perinatal, SNP or PRS) and M_{null} represents the model including only covariates. The risk factor is denoted r and cov , represents the included covariates gender, cohort-status (iPSYCH12 or iPSYCH15), birth year and psychiatric status in all models and gestational age, parental birth year and top 10 principal components in specified models. The residual error is represented by ϵ . The ratio of likelihood of the two models and quantifies how much better the risk-factor model fits to the data.

$$\lambda = \frac{L(M_r)}{L(M_{null})}$$

Where $L(M_r)$ is the likelihood of the risk factor model and $L(M_{null})$ the likelihood of the model without covariates (null model). The null-hypothesis is that the two models explain the data equally well, i.e. have identical likelihoods. A significance level for the null-hypothesis is calculated by assuming that the likelihood-ratio test statistic follows a chi-square distribution with k degrees of freedom.

$$-2 \log \lambda = -2(\log(L(M_r)) - \log(L(M_{null})))$$

$$-2 \log \lambda \sim \chi^2(k)$$

We interpret a rejection of the null hypothesis as the indication of significantly distinct associations between the phenotypes given in the risk factor analysis. All statistical analyses were corrected for multiple comparisons using the Benjamini-Hochberg procedure to control the false-discovery rate to 5 %.

Results

In this section, key findings for each study are presented. We highlight genetic differences between temporal wheeze subtypes in relation to shared genetics with asthma comorbidities, novel loci, genetic correlation, SNP-clusters and integration of nasal gene expression with GWAS results. Furthermore, we identify a clinical and transcriptomic fingerprint of individuals with high risk of severe wheeze and identify trajectories of redeemed inhaled corticosteroids in electronic health records that might represent wheeze endotypes.

Paper I

This study aims to improve understanding of genetic susceptibility and mechanisms behind wheeze symptoms in preschool children through large-scale GWAS. We combined GWAS from smaller birth cohorts with longitudinal records of wheeze symptoms, defining temporal subtypes based on persistence and remission patterns between three and six years (7). We carried out meta-analysis and characterized the influence of known asthma susceptibility loci for each phenotype and subsequently estimated genetic overlap with asthma-related comorbidities.

Discovery analyses

After correcting for multiple testing in the GWAS analyses ($P < 5e-8$), only the 17q12-21 locus (17q) was significantly associated with three out of the four phenotypes, namely EOW, PW and CA (see Table 1) and the lead variants in the 17q region for PW (rs11078926, OR = 1.43 (1.30 - 1.578), $P = 2.14E-11$), EOW (rs2305480, OR = 1.26 (1.17 - 1.33), $P = 2.30E-16$) and CA (rs1008723, OR = 1.38 (1.30 - 1.52), $P = 1.75E-17$), where all in high LD (minimum $r^2 = 0.82$) (see Supplementary Table 8). No genome-wide significant loci were found for TEW (see Supplementary Figure 1).

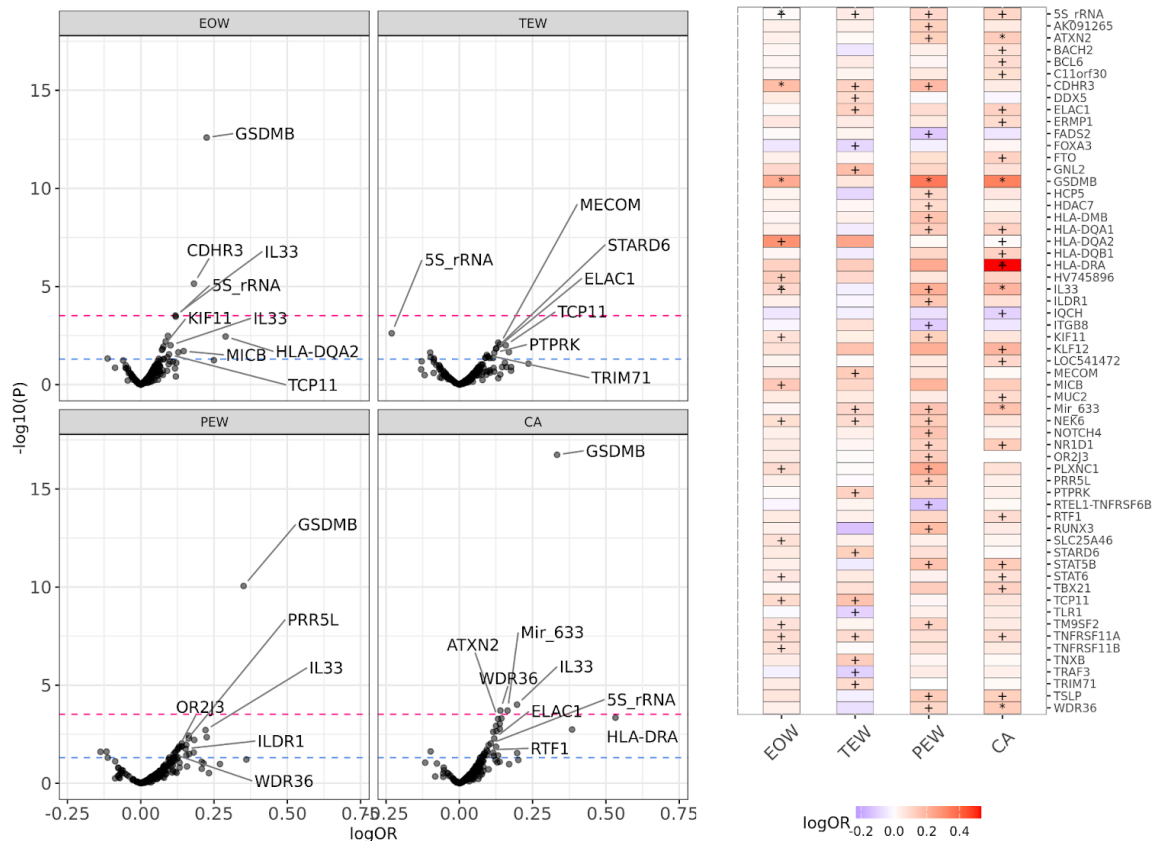
Analysis of known asthma variants

207 independent asthma-associated SNPs were used to characterize genetic similarity of the wheezing phenotypes to asthma. 58/207 SNPs were nominally associated with wheeze ($P < 0.05$) with consistent direction of effect. Most of the robust associations were observed for EOW, PW and CA (see Figure 1), but eight associated SNPs displayed opposite direction of effect (see Supplementary Table 12). Three SNPs were positively associated with EOW ($FDR < 0.05$), located in the loci *GSDMB*, *CDHR3* and *IL33*. Only *GSDMB* passed FDR in PW and SNPs (*ATXN2*, *GSDMB*, *IL33*, *LBH* & *WDR3*) passed FDR in CA. We observed limited directional replication in TEW and *CDHR3* was the only SNP with consistent asthma and wheeze effects (rs6967330-A, OR=1.12, $P=0.044$) (27).

Table 1. Association of the 17q12 locus represented by rs7219923 across the wheezing phenotypes. 17q12 is associated with all phenotypes except TEW characterized by remission before age six.

Phenotype	Nearest gene	SNP	EA/NEA	OR	95 % CI	P-value	N
CA	<i>GSDMB</i>	rs7219923	T/C	1.335	(1.241 - 1.438)	1.33E-14	15,074
EOW				1.24	(1.173 - 1.311)	3.05E-14	17,383
PW				1.369	(1.237 - 1.516)	1.28E-09	13,878
TEW				1.063	(0.979 - 1.153)	0.1423	12,604

Figure 1. Left panel: Volcano plots display the four wheezing phenotypes—EOW, TEW, PW, and CA, from top to bottom. The pink dashed line represents an FDR threshold below 0.05, while the blue dashed line indicates nominal significance ($P < 0.05$). These plots illustrate trends in the risk effects of known asthma-related variants and their impact on wheeze, with log-odds for wheeze status on the x-axis and $-\log_{10}(p\text{-value})$ on the y-axis. Right panel: Variants showing significant associations in at least one phenotype are displayed. Cells are color-coded according to log odds-ratio, with stars marking significance levels (+ for $P < 0.05$, * for $FDR < 0.05$).

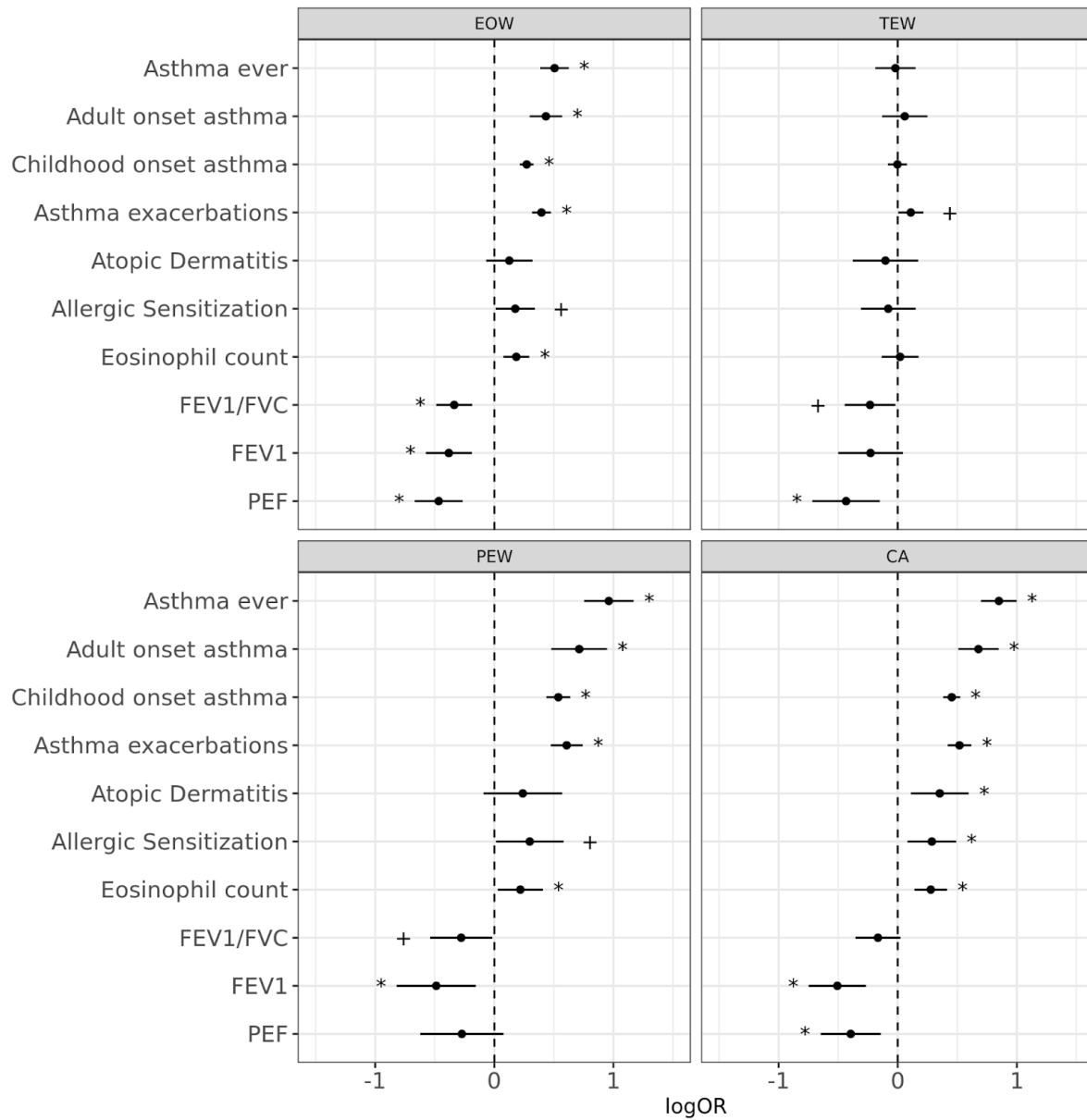


Genetic Similarity Between Adult Traits and Preschool wheeze

EOW overlapped with asthma related traits; both childhood- and adult-onset, allergy, eczema, eosinophil counts and lung function measurements. PW and CA displayed a similar trend in genetic overlap. In contrast, TEW did not overlap with asthma, eczema, allergy or eosinophils, whereas lung function measurements (especially PEF) displayed a negative relationship between lung function and risk of TEW. PEF was the only trait passing FDR (OR = 0.64, (CI =

0.48-0.86), $P = 2.73e-3$, whereas FEV1/FVC (OR = 0.79, (CI = 0.64-0.98), $P = 0.032$) and asthma exacerbations (OR = 1.12, (CI = 1.004-1.24), $P = 0.041$) displayed a suggestive overlap with TEW (see Figure 2). Results are shown in Supplementary Table 12 and scatter plots of SNP effect sizes between traits and preschool wheeze are shown in Supplementary Figures 5-14.

Figure 2. Log odds-ratios reflecting shared genetic links between adult traits and preschool wheeze. The association results show variant coefficients comparing adult traits (y-axis) with preschool wheeze (boxes), along with 95% confidence intervals. Stars denote the significance levels of the associations (+ for $P < 0.05$, * for $FDR < 0.05$).



Paper II

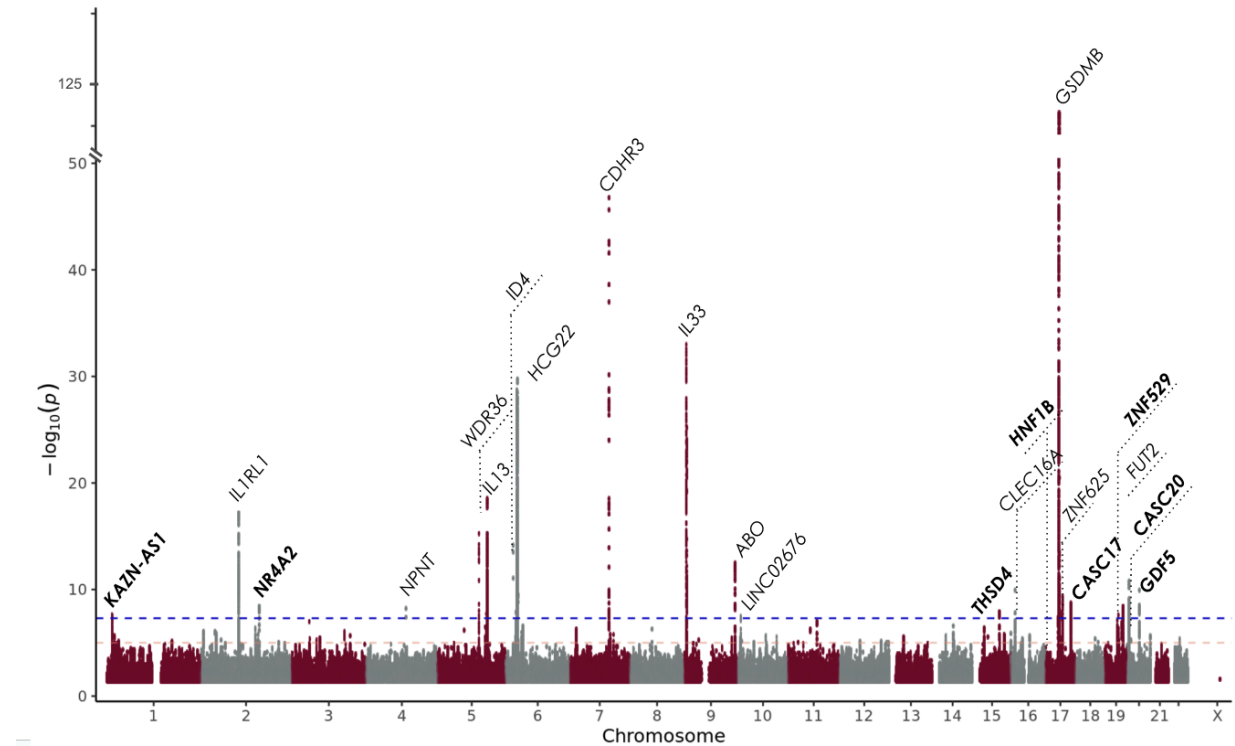
This study uses a broad definition, combining all individuals diagnosed with asthma, bronchitis, or bronchiolitis before age six using electronic health records. GWAS results from multiple cohorts, totaling ~20,000 cases and ~280,000 controls, aim to uncover molecular mechanisms influencing wheeze and identify specific risk loci linked to asthma-related traits.

GWAS results

The study included 19,461 children with severe wheeze and 284,631 controls. A total of 23 loci reached genome-wide significance in the meta-analysis ($P < 5 \times 10^{-8}$), with 8 of these being novel findings (see Figure 1 and Tabel 1).

The GWAS analysis identified novel loci, including *CASC20* (OR = 1.08, 95% CI = 1.06–1.11, $P = 1.85 \times 10^{-11}$), *GDF5* (OR = 1.08, 95% CI = 1.05–1.11, $P = 1.48 \times 10^{-10}$), *CASC17* (OR = 1.07, 95% CI = 1.05–1.1, $P = 2.38 \times 10^{-9}$), *NR4A2* (OR = 1.11, 95% CI = 1.07–1.15, $P = 4.45 \times 10^{-9}$), *THSD4* (OR = 1.09, 95% CI = 1.06–1.12, $P = 2.53 \times 10^{-8}$), *KAZN-AS1* (OR = 1.08, 95% CI = 1.05–1.11, $P = 3.18 \times 10^{-8}$), *ZNF529* (OR = 1.07, 95% CI = 1.05–1.1, $P = 3.20 \times 10^{-8}$), and *HNF1B* (OR = 1.06, 95% CI = 1.04–1.09, $P = 1.99 \times 10^{-7}$) (see Supplementary Figure 2). Five of these seven novel loci were replicated in association with preschool wheeze in the MoBa cohort, which included 19,621 cases and 29,134 controls (refer to Supplementary data). The replicated loci were *GDF5* (OR = 1.06, 95% CI = 1.03–1.09, $P = 2.25 \times 10^{-5}$), *NR4A2* (OR = 1.08, 95% CI = 1.04–1.13, $P = 1.35 \times 10^{-4}$), *CASC17* (OR = 1.05, 95% CI = 1.03–1.8, $P = 1.77 \times 10^{-4}$), *THSD4* (OR = 1.07, 95% CI = 1.03–1.1, $P = 4.27 \times 10^{-4}$), and *CASC20* (OR = 1.08, 95% CI = 1.02–1.08, $P = 1.13 \times 10^{-3}$) (see Table 1).

Figure 1: The Manhattan plot of the severe wheeze GWAS presents negative log-transformed p-values on the y-axis against chromosomal positions on the x-axis. Genomic risk loci are annotated with the prioritized consensus gene, with newly identified loci, not previously associated with asthma or wheezing disorders, highlighted in bold. The blue line indicates the genome-wide significance cutoff ($P < 5e-8$).



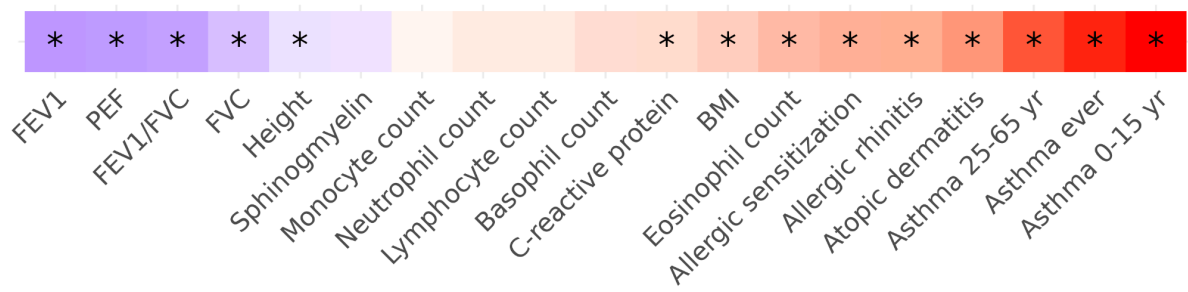
Heritability and genetic correlations

Heritability was estimated to be 12.66%. Within individual cohorts, heritability estimates were 17.85% for iPSYCH12, 13.15% for iPSYCH15, 18.22% for FinnGen, 18.19% for DBDS, and 5.19% for deCODE. Intercepts above 1 were observed for iPSYCH15, FinnGen, DBDS, and deCODE, while iPSYCH12 had an intercept of 0.99, and the meta-analysis intercept was 0.98 (see Supplementary Table 5). Significant positive genetic correlations were found between wheeze and asthma (0–15 years), asthma (25–65 years), lifetime asthma, atopic dermatitis, allergic rhinitis and sensitization, eosinophil count, BMI, and C-reactive protein levels in adults. Significant negative correlations were noted with height, FEV1, FVC, FEV1/FVC ratio, and PEF. Multiple testing adjustments were applied using the Benjamini-Hochberg procedure (FDR < 0.05) (see Figure 2).

Table 1: Fine-mapped variants are presented when a 95% credible set is available; otherwise, the lead SNP is reported. Replication results in the MoBa cohort are displayed on the right. Loci not previously linked to asthma or other wheezing disorders are shown in bold, and loci with significant replication ($P < 0.05$) have their p-values emphasized in bold.

Discovery: Severe wheeze (0-7 years)									Replication: MoBa (wheeze 0-7 years)		
Chr	Position	SNP	EAF	EA/NEA	OR	CI 95 %	P-value	Mapped gene	OR	CI 95 %	P-value
17	38062196	rs2305480	0.4797	G/A	1.32	(1.29 - 1.35)	2.70E-124	<i>GSDMB</i>	1.12	(1.09-1.15)	2.28E-15
7	105658451	rs6967330	0.2112	A/G	1.22	(1.19 - 1.26)	1.78E-47	<i>CDHR3</i>	1.13	(1.09-1.17)	3.87E-12
9	6213387	rs928413	0.7495	G/A	1.17	(1.14 - 1.2)	1.12E-33	<i>IL33-2</i>	1.05	(1.02-1.09)	1.24E-03
6	31009055	rs2517546	0.1259	T/C	1.25	(1.2 - 1.3)	9.81E-29	<i>HCG22</i>	1.10	(1.06-1.14)	4.72E-07
9	6114076	rs439190	0.7261	T/C	1.13	(1.1 - 1.16)	1.09E-19	<i>IL33-1</i>	1.03	(1-1.06)	5.54E-02
5	131996669	rs847	0.269	T/C	1.13	(1.1 - 1.16)	3.30E-19	<i>IL13</i>	1.04	(1.01-1.08)	9.14E-03
2	102955082	rs13408661	0.1398	G/A	1.16	(1.12 - 1.2)	7.54E-18	<i>ILIR1</i>	1.06	(1.01-1.1)	8.48E-03
5	110427795	rs79881201	0.3358	T/C	1.1	(1.07 - 1.13)	7.63E-15	<i>WDR36</i>	1.02	(1-1.05)	9.68E-02
6	19842661	rs9350191	0.8538	T/C	1.14	(1.1 - 1.18)	7.75E-15	<i>IDA</i>	1.05	(1.01-1.09)	1.52E-02
9	136151806	rs600038	0.7852	C/T	1.1	(1.07 - 1.13)	9.94E-12	<i>ABO</i>	1.02	(0.99-1.05)	2.40E-01
20	6470265	rs928202	0.6057	G/A	1.08	(1.06 - 1.11)	1.85E-11	<i>CASC20</i>	1.05	(1.02-1.08)	1.13E-03
16	11277358	rs11644510	0.3444	C/T	1.08	(1.06 - 1.11)	1.33E-10	<i>CLEC16A</i>	1.01	(0.98-1.03)	7.06E-01
20	34025756	rs143384	0.5648	A/G	1.08	(1.05 - 1.1)	1.48E-10	<i>GDF5</i>	1.06	(1.03-1.09)	2.25E-05
17	47461433	rs17637472	0.4148	A/G	1.08	(1.05 - 1.1)	5.17E-10	<i>PHB1</i>	1.02	(0.99-1.05)	1.35E-01
17	69191316	rs1008348	0.4256	A/G	1.07	(1.05 - 1.1)	2.38E-09	<i>CASC17</i>	1.05	(1.03-1.08)	1.77E-04
2	157083697	rs72907878	0.1271	G/A	1.11	(1.07 - 1.15)	4.45E-09	<i>NR4A2</i>	1.08	(1.04-1.13)	1.35E-04
4	106819053	rs34712979	0.2563	A/G	1.08	(1.05 - 1.11)	7.35E-09	<i>NPNT</i>	0.98	(0.92-1.05)	5.73E-01
15	71804587	rs62015884	0.177	T/C	1.09	(1.06 - 1.12)	2.53E-08	<i>THSD4</i>	1.07	(1.03-1.1)	4.27E-04
1	14440292	rs12026570	0.7644	A/G	1.08	(1.05 - 1.11)	3.18E-08	<i>KAZN-AS1</i>	1.01	(0.98-1.04)	4.84E-01
19	37048117	rs3096616	0.2866	C/T	1.07	(1.05 - 1.1)	3.20E-08	<i>ZNF529</i>	0.98	(0.95-1.02)	3.19E-01
10	8939198	rs78539108	0.8656	T/A	1.1	(1.06 - 1.13)	3.53E-08	<i>LINC02676</i>	1.03	(0.99-1.07)	1.67E-01
19	49218111	rs33988101	0.4824	G/T	1.07	(1.04 - 1.09)	5.65E-08	<i>FUT2</i>	1.00	(0.97-1.02)	8.14E-01
17	36102035	rs7405696	0.4858	G/C	1.06	(1.04 - 1.09)	1.99E-07	<i>HNF1B</i>	0.99	(0.96-1.01)	2.94E-01

Figure 2: The genetic correlations between severe wheeze and related traits are illustrated below. Squares are color-coded and scaled based on the genetic correlation estimate (r_g) with red indicating positive and blue negative correlation. Stars indicate correlations that remain significant after multiple testing corrections using the Benjamini-Hochberg procedure ($FDR < 0.05$).



Prioritization of causal genes at genomic risk loci

A total of 62 genes were prioritized across the 23 loci, with *GSDMB* and *CDHR3* receiving the highest scores, being prioritized by 7 out of 12 features. Genes scoring in the upper third quartile were prioritized by at least 4 features. These genes, in descending order, are *GSDMB*, *CDHR3*, *IL-13*, *HNFB*, *THSD4*, *NPNT*, *CLEC16A*, *FUT2* and *GDF5*. Among these top-scoring genes, three—*HNFB*, *THSD4* and *GDF5*—were associated with novel loci (see Table 2).

Association-based groupings of severe wheeze risk loci

The clustering analysis revealed seven distinct terminal branches in the hierarchical dendrogram for SNPs and five branches for traits (refer to Supplementary Table 11). Among these, three branches contained multiple SNPs and were defined as SNP clusters, while branches with a single SNP included *GSDMB*, *HCG22*, *FUT2* and *ABO*.

One of the SNP clusters (Cluster 2) was associated with low adult lung function and showed minimal connections to asthma-related traits. Another cluster (Cluster 7) was linked to a high risk of atopic conditions, such as asthma, allergy, and eczema, with strong associations to eosinophil levels and moderately elevated counts of other immune cells. The final SNP cluster (Cluster 4) displayed a mixed phenotypic profile; two variants within this cluster (*GDF5* and *CASC20*) were notably associated with height but did not exhibit other shared association patterns (see Figure 3).

Table 2: Gene prioritization. Overview of methods used for each locus and their prediction. Rows signify wheeze risk loci, while columns contain various information; genomic coordinates, prioritization methods and lastly the consensus gene.

CHR	SNP	Nearest gene	Intron	Exon	Splice variant	AlphaMissense	DEPICT	MAGMA	PoPS	TWAS + COLOC	eQTL	EWAS + COLOC	OpenTargets	Consensus gene
7	rs6967330	CDHR3	CDHR3	CDHR3	-	CDHR3	ATXN7L1	CDHR3	CDHR3	-	-	-	CDHR3	CDHR3
17	rs2305480	GSDMB	ORMDL3	-	GSDMB	ZPBP2	GJD3	ORMDL3	GSDMB	GSDMB	GSDMB	GSDMB	GSDMB	GSDMB
5	rs847	IL13	IL13	IL13	-	IL13	IRF1	IL13	IL13	-	-	-	IL13	IL13
17	rs7405696	HNF1B	HNF1B	-	-	-	-	HNF1B	HNF1B	-	-	HNF1B	HNF1B	HNF1B
15	rs62015884	THSD4	THSD4	-	-	-	THSD4	THSD4	THSD4	-	-	-	THSD4	THSD4
4	rs34712979	NPNT	NPNT	-	-	-	NPNT	INTS12	NPNT	-	-	-	NPNT	NPNT
16	rs11644510	SOC1	CLEC16A	-	-	-	SOC1	CLEC16A	-	-	-	CLEC16A	CLEC16A	CLEC16A
19	rs33988101	MAMSTR	RASIP1	FUT2	-	FUT2	NUCB1	RASIP1	MAMSTR	FUT2	FUT2	IZUMO1	MAMSTR	FUT2
20	rs143384	GDF5	UQC1	GDF5	-	GDF5	-	UQC1	GDF5	UQC1	FAM83C	UQC1	CPNE1	GDF5
9	rs600038	ABO	-	-	-	-	CELP	-	-	-	ABO	-	ABO	ABO
6	rs9350191	ID4	-	-	-	-	MBOAT1	ID4	-	-	-	-	ID4	ID4
2	rs13408661	IL1RL1	IL18RAP	-	-	-	IL1RL2	IL18R1	IL1RL1	-	IL18RAP	IL1RL1	IL18RAP	IL1RL1
5	rs79881201	WDR36	WDR36	-	-	-	-	TSLP	WDR36	-	TSLP	-	TSLP	WDR36
19	rs3096616	ZNF529	-	-	-	-	-	ZNF260	ZNF529	-	LINC00665	-	ZNF529	ZNF529
17	rs1008348	CASC17	CASC17	-	-	-	-	-	-	-	-	-	CASC17	CASC17
20	rs928202	CASC20	CASC20	-	-	-	-	-	-	-	-	-	BMP2	CASC20
9	rs928413	IL33-2	-	-	-	-	-	IL33-2	-	-	-	-	ERMP1	IL33-2
10	rs78539108	LINC02676	LINC02676	-	-	-	-	-	-	-	-	-	-	LINC02676
2	rs72907878	NR4A2	LINC01876	-	-	-	NR4A2	-	-	-	-	-	GPD2	NR4A2
17	rs17637472	PHB1	ZNF652	-	-	-	FAM117A	ZNF652	-	-	-	ABI3	GNMT2	ZNF652
6	rs2517546	HCG22	MICB-DT	HCP5	-	MICB	-	-	-	-	-	-	HLA-C	HCG22
9	rs439190	IL33-1	-	-	-	-	-	RP11-21817.2	-	-	-	-	RANBP6	IL33-1
1	rs12026570	KAZN-ASI	-	-	-	-	-	-	-	-	-	-	KAZN	KAZN-ASI

Figure 3: Trait-based clustering of SNPs was conducted by converting SNP association statistics to z-scores, followed by standardization to zero-mean and unit variance to adjust for sample size differences. The resulting SNP clusters are displayed with unscaled z-scores, representing the actual direction and strength of associations. Missing SNP values were imputed using the trait's median value, and only traits with at least one genome-wide significant wheeze-associated SNP ($P < 5e-8$) were included. Positive z-scores are shown in red, negative in blue, and cells marked with '*' indicate associations significant after Benjamini-Hochberg correction ($FDR < 0.05$).

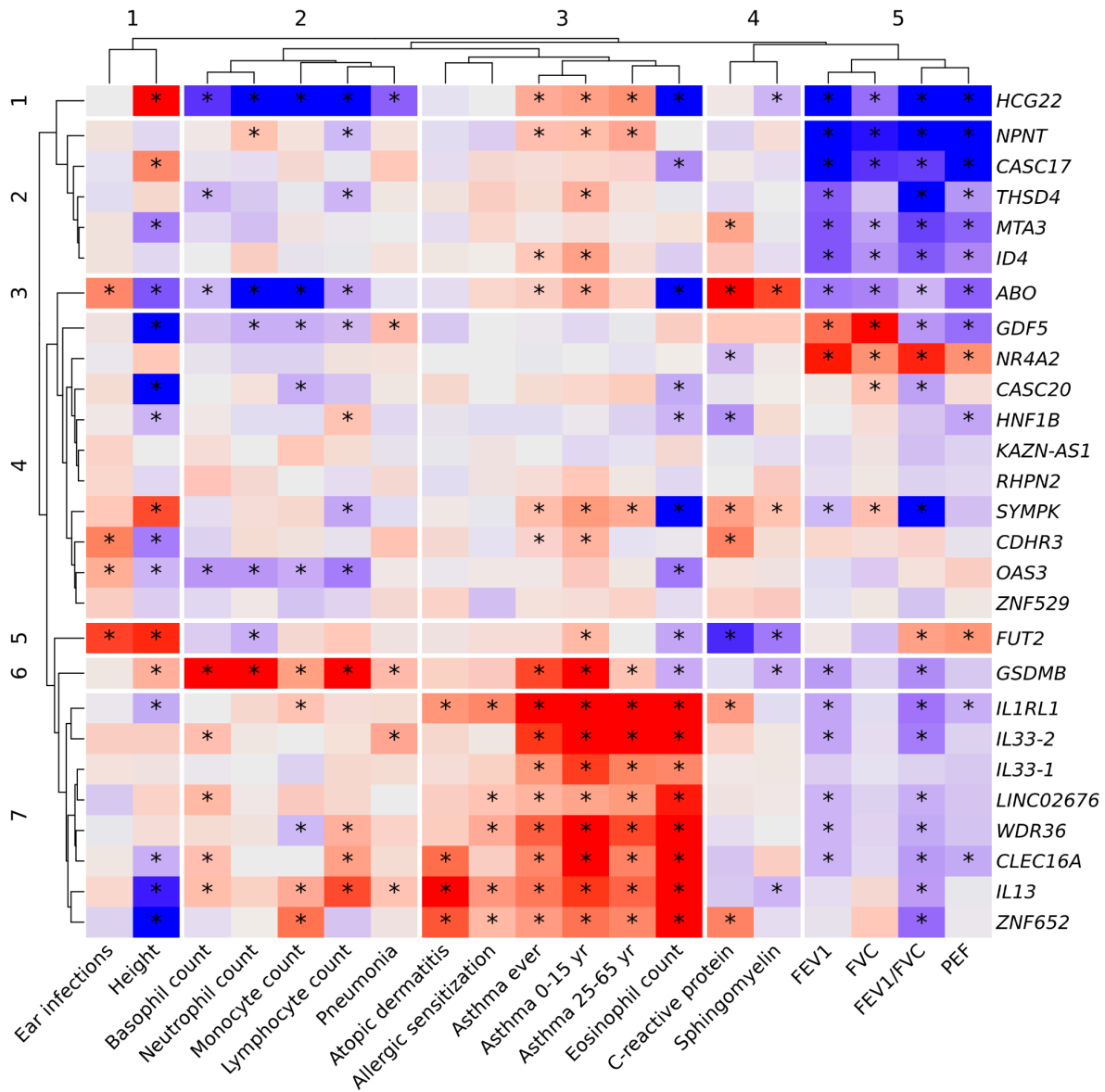
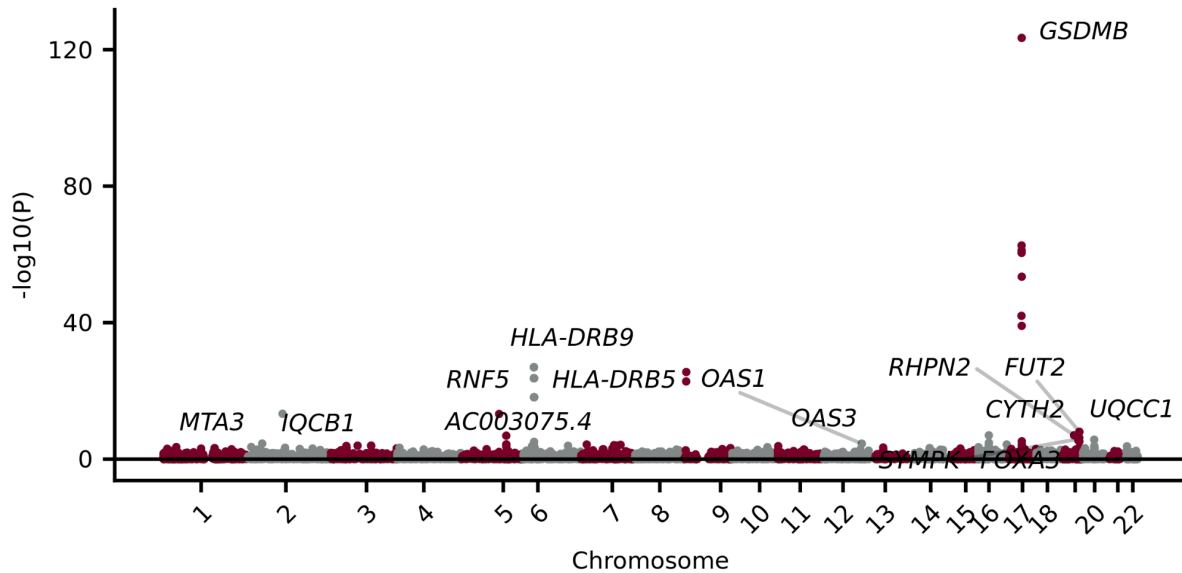


Figure 4: The Manhattan plot displays TWAS results using the COPSAC2010 NEC expression data. Genes are labeled if the TWAS P-value meets the Benjamini-Hochberg correction threshold and if the probability of colocalization (shared variant) exceeds 75%.



Transcriptome-wide association analysis

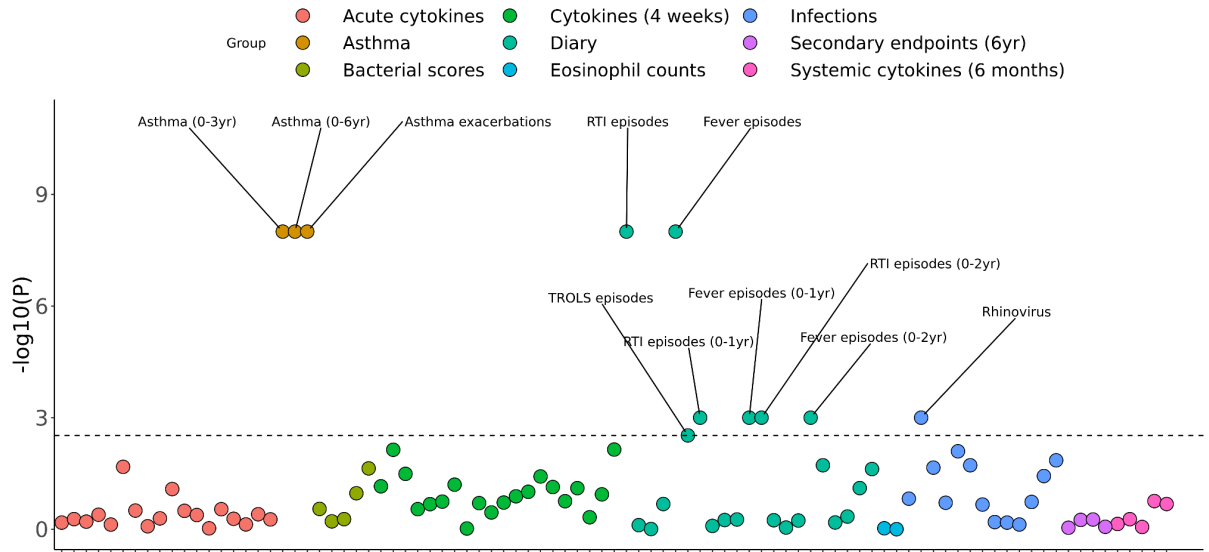
A total of 118 genes met the threshold for multiple testing correction in the TWAS. Among these, *MTA3*, *IQCB1*, *RNF5*, *HLA-DRB5*, *HLA-DRB9*, *AC003075.4*, *OAS1/OAS3*, *GSDMB*, *RHPN2*, *SYMPK*, *CYTH2*, *FUT2* and *UQCC1* had a probability of a shared variant exceeding 75% (see Figure 4). For three of these genes—*MTA3* (OR = 1.06, 95% CI = 1.04–1.09, P = 1.84E-06), *OAS1/OAS3* (OR = 1.06, 95% CI = 1.03–1.08, P = 8.02E-06), and *RHPN2* (OR = 1.1, 95% CI = 1.06–1.14, P = 1.06E-07)—the shared GWAS and eQTL variants had not previously been associated with childhood asthma or asthma in general. Two of these three novel TWAS loci showed borderline significance in replication with the MoBa cohort: *MTA3* (OR = 1.03, 95% CI = 1.0–1.06, P = 4.03E-02) and *OAS1/OAS3* (OR = 1.03, 95% CI = 1.0–1.06, P = 4.76E-02) (see Figure 4).

Clinical & transcriptomic fingerprint of genetic risk score

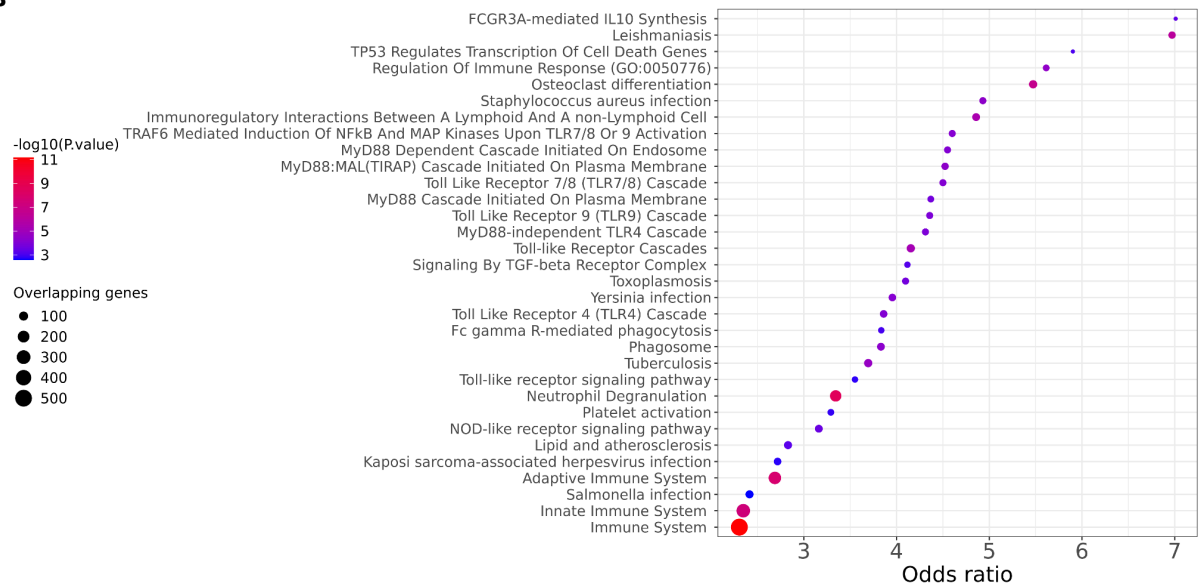
We observed a strong association between the severe wheeze polygenic risk score (PRS) and outcomes such as asthma, acute severe wheeze, infant respiratory infections, fever episodes, and rhinovirus respiratory illnesses within the COPSAC cohorts (see Figure 5A). Additionally, the PRS was linked to genes expressed in nasal epithelium, enriched in pathways related to Fc gamma receptor (Fc γ) signaling, Toll-like receptor cascades (TLR4, TLR7, TLR8, and TLR9), as well as *Staphylococcus aureus* and other bacterial and parasitic infections. The most significantly enriched functional terms included FCGR4A-mediated IL-10 synthesis (OR = 7.01, P = 4.3e-4), Leishmaniasis (OR = 6.97, P = 7.5e-7), and TP53-regulated transcription of cell death genes (OR = 5.9, P = 9.72e-4) (see Figure 5B).

Figure 5: **A:** Clinical profile of the PRS. A PheWAS of clinical and para-clinical endpoints in the COPSAC2000 and COPSAC2010 cohorts, displaying negative-log-p-values on the y-axis and endpoints along the x-axis. Colors indicate different categories of endpoints, arranged alphabetically. The horizontal dotted line marks the significance threshold after multiple testing corrections (FDR < 0.05). **B:** Functional enrichment of airway epithelium genes associated with the severe wheeze PRS, excluding the HLA region. Only functional terms meeting the Benjamini-Hochberg threshold (FDR < 0.05) are shown. The x-axis represents the odds ratio for functional enrichment; points are colored by enrichment p-values and sized according to the number of overlapping genes within each functional term.

A

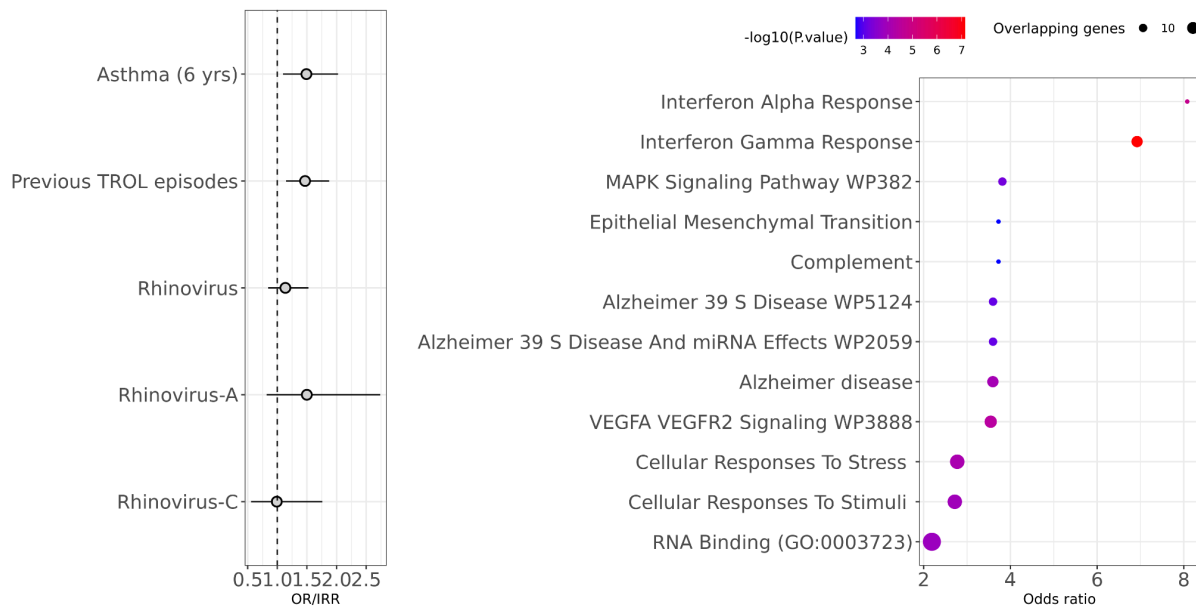


B



The PRS was also linked to asthma at age six among bronchiolitis patients (OR = 1.49, 95% CI = 1.09–2.02, $P = 0.011$) and to the number of respiratory issues in infancy before the initial hospitalization in the MARC-35 cohort (IRR = 1.46, 95% CI = 1.15–1.88, $P = 2.2e-3$). A positive association was observed between PRS and rhinovirus subtypes, with a trend toward significance. Functional enrichment analysis of PRS-associated nasal epithelium genes in the MARC-35 cohort indicated enrichment in pathways related to viral infections and inflammatory signaling cascades. The top enriched terms included interferon-alpha response (OR = 8.07, $P =$

Figure 6: The right panel shows the associations between the PRS (0-7) in the MARC-35 cohort with relevant traits based on the COPSAC PheWas, with the PRS as exposure. Binary outcomes are analyzed with logistic regression while count phenotypes (Previous TROL episodes) are analyzed using quasi-poisson regression. The right panel shows enriched pathways using enrichR, among genes associated with the PRS in the MARC-35 expression data in the nasal epithelium sampled within 24 hours of hospitalization. All analyses were performed in infants of European ancestry hospitalized with bronchiolitis and adjusted for gender, age and top 2 principal components.



1.5e-5), interferon-gamma response (OR = 6.9, $P = 7.2\text{e-}8$), and MAPK signaling pathway (OR = 3.81, $P = 5.27\text{e-}4$) (see Figure 6).

Severe wheeze, Pividori et al. (2019), Namjou et al. (2022), and Tsuo et al. (2022) reported AUC values for predicting wheeze in MoBa of 0.561, 0.537, 0.536, and 0.551, respectively. DeLong's test indicated that the severe wheeze PRS yielded significantly higher AUC values compared to the other scores (see Table 3).

Table 3: AUC values for predictive ability of three published PRSs in comparison with severe wheeze. We used wheeze as defined in MoBa as an outcome.

DeLong's test				
PRS	AUC	Z	dAUC (CI 95%)	P-value
Pividori et al. 2019	0.54	7.21	(0.017-0.03)	5.79E-13
Namjou et al. 2022	0.54	7.61	(0.018-0.031)	1.69E-14
Tsuo et al. 2022	0.55	3.10	(0.003-0.016)	1.90E-03

Paper III

This paper presents a method to identify temporal wheeze phenotypes using machine learning on longitudinal corticosteroid prescription records. The resulting trajectories are analyzed for associations with perinatal and genetic risk factors, and statistical tests assess whether these factors show distinct associations across phenotypes, suggesting potential endotypes.

Latent class analysis

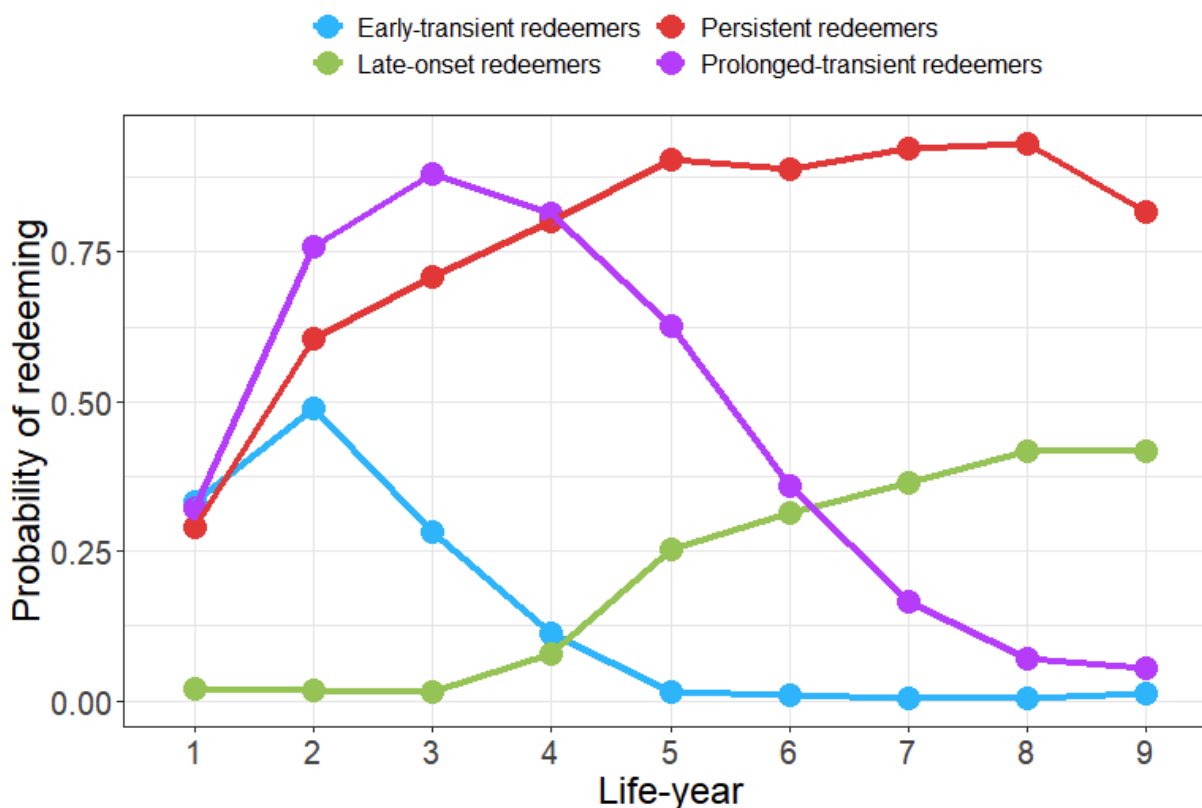
A total of 7,527 children with at least one ICS prescription before age 9 were included. Based on the elbow plot of BIC versus the number of classes (see Supplementary Figure 2), we identified two potential solutions, with inflection points at 4 and 6 classes. Sensitivity analysis on the number of clusters supported a 4-class model. Additionally, as the 4-class model provided the most parsimonious solution, all subsequent analyses were conducted using the results from this 4-class model (see Table 1).

Table 1: This table presents model selection results based on the number of classes, including BIC, entropy, average silhouette width (ASW), and p-values from the Lo-Mendell-Rubin's likelihood ratio test and the bootstrapped likelihood ratio test (BLRT). P-values for Lo-Mendell-Rubin and BLRT tests compare models with k classes versus $k-1$ classes to assess whether adding an additional class significantly improves the model. For example, the second row evaluates whether a 2-class model outperforms a 1-class model. ASW, Lo-Mendell-Rubin, and BLRT are undefined for the 1-class model.

K	BIC	Entropy	ASW	Lo-Mendell-Rubin	BLRT
1	77261	5.1	NA	NA	NA
2	69226	4.6	0.123	0.00e+00	0.00e+00
3	66657	4.5	0.138	0.00e+00	0.00e+00
4	65245	4.3	0.120	4.16e-305	0.00e+00
5	64823	4.3	0.108	2.13e-99	6.40e-02
6	64440	4.3	0.142	1.64e-91	8.10e-02
7	64355	4.2	0.148	1.13e-31	3.43e-01
8	64281	4.2	0.144	1.91e-28	4.24e-01
9	64209	4.2	0.129	1.05e-27	4.09e-01
10	64156	4.2	0.123	1.40e-24	4.52e-01

We classified the four identified classes clinically as follows: early-transient redeemers (ETR, $n = 3,919$, 52%), late-onset redeemers (LOR, $n = 1,302$, 17%), persistent redeemers (PR, $n = 781$, 10%), and prolonged-transient redeemers (PTR, $n = 1,525$, 21%). ETR was characterized by 50% redeeming inhaled corticosteroids (ICS) in the second year of life, with no further prescriptions from age 6 onward. PTR had high ICS redemption (~70–80%) from ages 2–4, followed by a decline starting at age 5, with less than 25% redeeming by age 7 and near-zero frequencies by age 9. PR displayed consistent ICS redemption across the study period, with approximately 25% in the first year, nearly 100% from ages 5–8, and a slight decrease by age 9.

Figure 1: Identified trajectories in ICS prescription patterns. The x-axis represents life years of ICS redemption, while the y-axis shows the frequency of ICS redemption within each year. The four phenotypes are categorized as early-transient redeemers (ETR), (PR), and prolonged-transient, persistent redeemers (PR) and late-onset redeemers (LOR).



LOR showed no ICS prescriptions in the first three years, with redemption peaking around ages 8–9, when approximately 50% were redeeming ICS (see Figure 1). Children who did not redeem any ICS prescriptions during the period ($n = 28,809$) were excluded from this analysis and used as controls in subsequent analyses (see Figure 1).

Perinatal risk profiles

After adjusting for multiple comparisons, ETR, LOR, PR, and PTR were associated with 19, 13, 11, and 19 out of 22 perinatal risk factors, respectively. Premature birth, cesarean section,

gestational age, gender, and parental history of asthma and allergies showed significant associations across all trajectories following Benjamini-Hochberg correction. ETR had the strongest associations with maternal smoking (OR = 1.48, CI = 1.4–1.6, P = 0.00), presence of older siblings (OR = 1.09, CI = 1.0–1.1, P = 1.87e-4), maternal age (OR = 0.89, CI = 0.88–0.92, P = 1.56e-7), and paternal age. LOR showed no perinatal risk factors as the strongest association. PR was most strongly associated with cesarean section (OR = 1.50, CI = 1.3–1.8, P = 3.47e-8) and female gender (OR = 0.77, CI = 0.7–0.8, P = 1.54e-10). PTR had the highest odds ratios for prematurity (OR = 2.02, CI = 1.7–2.4, P = 4.88e-15), gestational age (OR = 0.81, CI = 0.8–0.82, P = 0.00), and maternal antibiotic use (OR = 1.50, CI = 1.3–1.7, P = 5.05e-8). Additionally, neonatal respiratory distress showed moderate associations only with PR (OR = 1.49, CI = 1.1–2.1, P = 0.024) and PTR (OR = 1.38, CI = 1.1–1.8, P = 0.01), even after adjusting for gestational age (see Table 2).

Within-class LRT indicated that prematurity, cesarean section, neonatal respiratory distress, maternal smoking during pregnancy, birth weight, gestational age, gender, presence of older siblings, parental age, and maternal antibiotic use were distinctly associated across classes (see Table 2). Paternal asthma, allergy, and eczema did not differ significantly across ICS trajectories. Maternal antibiotic use showed a unique association with both ETR and PTR, but was strongest for PTR. Socioeconomic factors, such as household income and low parental education, either had no association or showed similar effects across trajectories.

Table 2: Multinomial regression results across the four phenotypes, with baseline controls incorporated. The LRT column provides the FDR-corrected *p*-value for within-class LRT tests.

Exposure	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			LRT
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
Cesarian section	1.14	1.04-1.2	3.90e-03	1.29	1.13-1.5	1.19e-04	1.50	1.26-1.8	3.47e-06	1.13	0.97-1.3	1.09e-01	3.025e-02*
Smoking during pregnancy	1.48	1.37-1.6	0.00e+00	1.23	1.1-1.4	4.85e-04	0.90	0.76-1.1	2.32e-01	1.19	1.05-1.4	7.81e-03	1.146e-06*
Premature	1.47	1.29-1.7	1.33e-08	2.02	1.7-2.4	4.88e-15	1.74	1.35-2.2	2.08e-05	1.28	1.02-1.6	3.53e-02	1.379e-02*
Gestational age	0.86	0.84-0.9	0.00e+00	0.81	0.78-0.8	0.00e+00	0.87	0.81-0.9	1.46e-05	0.91	0.86-1	5.76e-04	2.840e-02*
Birth weight	0.88	0.85-0.9	1.33e-13	0.87	0.82-0.9	1.32e-08	0.96	0.89-1	2.24e-01	0.96	0.91-1	1.43e-01	2.936e-02*
Female-gender	0.88	0.85-0.9	4.35e-13	0.84	0.8-0.9	1.66e-09	0.77	0.72-0.8	1.54e-10	0.87	0.82-0.9	3.44e-06	4.139e-02*
Neonate resp. distress	1.01	0.84-1.2	9.03e-01	1.38	1.08-1.8	1.00e-02	1.49	1.05-2.1	2.38e-02	0.79	0.55-1.1	1.91e-01	0.000e+00*
Older siblings	1.09	1.04-1.1	1.87e-04	0.98	0.91-1.1	5.82e-01	1.01	0.91-1.1	9.14e-01	0.91	0.83-1	2.46e-02	2.339e-03*
Distance to older sib.	0.98	0.92-1	4.47e-01	0.95	0.86-1	2.32e-01	1.12	1.01-1.3	3.91e-02	1.09	1-1.2	5.65e-02	4.139e-02*
Mothers age	0.89	0.85-0.9	1.59e-07	0.91	0.85-1	3.75e-03	1.07	0.98-1.2	1.47e-01	0.97	0.9-1	3.37e-01	4.539e-03*
Fathers age	0.89	0.85-0.9	8.17e-07	0.94	0.88-1	6.66e-02	1.09	1-1.2	4.83e-02	0.96	0.89-1	2.24e-01	3.529e-03*
Low education father	1.45	0.94-2.2	9.25e-02	1.63	0.89-3	1.13e-01	2.16	1.04-4.5	3.80e-02	1.44	0.73-2.9	2.98e-01	7.883e-01
Low education mother	1.29	0.96-1.7	9.42e-02	1.07	0.67-1.7	7.84e-01	1.46	0.83-2.6	1.93e-01	0.94	0.54-1.6	8.13e-01	7.367e-01
Household income	0.85	0.8-0.9	4.30e-06	0.87	0.78-1	5.65e-03	0.93	0.82-1.1	2.39e-01	0.90	0.81-1	4.10e-02	7.367e-01
Mothers Asthma	1.75	1.58-1.9	0.00e+00	1.85	1.58-2.2	4.66e-15	1.72	1.37-2.2	3.51e-06	1.52	1.27-1.8	7.63e-06	6.694e-01
Mothers Allergy	1.24	1.13-1.4	9.25e-06	1.52	1.32-1.8	3.56e-09	1.45	1.18-1.8	4.49e-04	1.39	1.18-1.6	6.69e-05	8.600e-02
Mothers Steroid_cream	1.18	1.07-1.3	4.76e-04	1.27	1.11-1.5	7.27e-04	1.17	0.95-1.4	1.33e-01	1.27	1.09-1.5	2.71e-03	7.367e-01
Mothers Antibiotics	1.19	1.08-1.3	3.24e-04	1.50	1.29-1.7	5.05e-08	1.02	0.83-1.3	8.63e-01	1.11	0.95-1.3	1.96e-01	2.652e-02*
Fathers Asthma	1.36	1.17-1.6	7.93e-05	1.72	1.39-2.1	7.42e-07	1.61	1.15-2.2	5.02e-03	1.63	1.27-2.1	1.11e-04	3.448e-01
Fathers Allergy	1.16	1.02-1.3	2.01e-02	1.32	1.1-1.6	2.96e-03	1.49	1.14-2	3.42e-03	1.49	1.21-1.8	1.35e-04	1.417e-01
Fathers Steroid_cream	1.21	1.07-1.4	2.47e-03	1.28	1.07-1.5	7.66e-03	1.32	1.01-1.7	4.58e-02	1.09	0.88-1.4	4.40e-01	7.367e-01
Fathers Antibiotics	1.14	1.03-1.3	1.20e-02	1.21	1.03-1.4	1.65e-02	1.18	0.94-1.5	1.56e-01	1.12	0.95-1.3	1.84e-01	8.879e-01

Genetic risk profiles

After adjusting for multiple comparisons, ETR, LOR, PR, and PTR were associated with 3, 3, 8, and 4 out of 11 SNPs, respectively. *GSDMB*, *CDHR3*, and *TSLP* were associated with ETR following correction, though no single SNP was most strongly linked to ETR. For LOR, *HLA-DQA1* showed the strongest association (OR = 1.16, CI = 1.1–1.3, P = 5.04e-4), but without a significant LRT. The strongest predictors for PR included *GSDMB* (OR = 1.52, CI = 1.4–1.7, P = 5.3e-15), *CDHR3* (OR = 1.45, CI = 1.3–1.6, P = 2.31e-9), *IL-33(1)* (OR = 1.37, CI = 1.17–1.6, P = 5.53e-5), *IL-33(2)* (OR = 1.25, CI = 1.098–1.4, P = 9.44e-4), *IL1RL1* (OR =

Table 3: Single-SNP associations and interactions with the phenotypes are displayed, with odds ratios on the y-axis and the nearest mapped gene on the x-axis. Stars denote significant ($FDR < 0.05$) within-class LRT results, indicating differences among phenotypes for significantly associated risk factors.

Locus	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	LRT
<i>GSDMB</i>	1.09	1.03-1.1	8.38e-04	1.36	1.26-1.5	4.44e-16	1.52	1.37-1.7	5.33e-15	1.20	1.1-1.3	1.06e-05	4.790e-10*
<i>CDHR3</i>	1.17	1.1-1.2	4.99e-07	1.16	1.06-1.3	1.90e-03	1.45	1.28-1.6	2.31e-09	1.07	0.96-1.2	2.09e-01	8.371e-03*
<i>FUT2</i>	1.00	0.95-1	9.78e-01	1.14	1.06-1.2	4.73e-04	1.05	0.95-1.2	3.37e-01	1.01	0.94-1.1	7.36e-01	3.517e-02*
<i>HLA-DQA1</i>	1.02	0.97-1.1	4.65e-01	1.06	0.98-1.1	1.58e-01	1.04	0.94-1.2	4.17e-01	1.16	1.07-1.3	5.04e-04	9.113e-02
<i>ANXA1</i>	0.93	0.79-1.1	3.38e-01	0.89	0.69-1.1	3.35e-01	0.62	0.42-0.9	1.74e-02	1.12	0.88-1.4	3.62e-01	9.113e-02
<i>IL33-1</i>	1.06	0.99-1.1	1.06e-01	1.05	0.95-1.2	3.14e-01	1.37	1.17-1.6	6.53e-05	1.06	0.95-1.2	3.12e-01	3.151e-02*
<i>IL33-2</i>	1.01	0.95-1.1	7.08e-01	1.12	1.01-1.2	2.86e-02	1.25	1.09-1.4	9.44e-04	1.13	1.02-1.3	2.05e-02	3.517e-02*
<i>IL1RL1</i>	0.95	0.89-1	2.11e-01	1.16	1.03-1.3	1.52e-02	1.24	1.05-1.5	1.22e-02	1.13	1-1.3	5.40e-02	8.223e-03*
<i>IL13</i>	1.03	0.97-1.1	3.19e-01	1.05	0.96-1.1	3.17e-01	1.27	1.13-1.4	7.76e-05	0.99	0.9-1.1	9.14e-01	3.151e-02*
<i>TSLP</i>	1.08	1.03-1.1	1.32e-03	1.13	1.05-1.2	1.45e-03	1.31	1.19-1.5	1.44e-07	1.13	1.04-1.2	2.55e-03	3.151e-02*
<i>ABO</i>	1.01	0.96-1.1	7.09e-01	1.04	0.96-1.1	3.15e-01	1.08	0.98-1.2	1.27e-01	0.99	0.91-1.1	7.98e-01	5.299e-01
<i>GSDMBxCDHR3</i>	1.09	1-1.2	6.19e-02	1.15	1-1.3	4.53e-02	1.22	1.02-1.5	3.32e-02	1.19	1.02-1.4	2.31e-02	5.110e-01
<i>FUT2xABO</i>	0.94	0.88-1	1.08e-01	1.05	0.94-1.2	3.54e-01	1.06	0.92-1.2	4.09e-01	1.00	0.89-1.1	9.38e-01	3.125e-01

1.24, CI = 1.05–1.5, $P = 1.22e-2$), *IL-13* (OR = 1.27, CI = 1.1–1.4, $P = 7.76e-5$), and *TSLP* (OR = 1.31, CI = 1.2–1.5, $P = 1.44e-7$). For PTR, *FUT2* had the strongest association (OR = 1.14, CI = 1.1–1.2, $P = 4.73e-4$). The T allele of the *ANXA1* SNP (rs75260654) showed a negative association with PR after multiple testing corrections (OR = 0.62, CI = 0.4–0.9, $P = 1.74e-2$). While epistatic effects of *GSDMBxCDHR3* and *FUT2xABO* did not pass correction, *GSDMBxCDHR3* showed a consistent positive trend across all phenotypes (see Table 3).

Within-class LRT revealed distinct associations across phenotypes for *GSDMB*, *CDHR3*, *FUT2*, *IL-33*, *IL-13*, and *TSLP*.

Polygenic risk scores (PRS) for childhood-onset asthma, severe asthma, adult-onset asthma,

Table 4: Polygenic risk score (PRS) analysis results from multinomial regression across the four phenotypes, with baseline controls included. The baseline group consists of individuals who did not redeem any ICS prescriptions within the first 9 years of life.

PRS	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			LRT
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
BMI	1.10	1.06-1.14	4.17e-08	1.08	1.03-1.14	2.46e-03	1.04	0.97-1.12	2.40e-01	1.05	0.99-1.11	7.57e-02	4.757e-01
Allergic rhinitis	1.03	0.99-1.06	1.37e-01	1.06	1.01-1.12	2.67e-02	1.19	1.11-1.28	2.98e-06	1.12	1.06-1.18	1.34e-04	3.029e-03*
Allergic sensitization	1.00	0.97-1.04	8.45e-01	1.04	0.99-1.1	1.28e-01	1.18	1.09-1.26	1.04e-05	1.10	1.04-1.17	6.61e-04	6.130e-04*
Atopic dermatitis	1.05	1.02-1.09	3.69e-03	1.03	0.97-1.08	3.36e-01	1.12	1.04-1.2	1.79e-03	1.08	1.02-1.14	6.31e-03	3.384e-01
Adult onset asthma	1.06	1.03-1.1	5.01e-04	1.15	1.09-1.22	1.80e-07	1.28	1.19-1.38	5.66e-11	1.22	1.15-1.3	1.48e-11	2.549e-06*
Childhood onset asthma	1.12	1.08-1.16	9.99e-11	1.27	1.2-1.34	0.00e+00	1.73	1.61-1.87	0.00e+00	1.38	1.3-1.46	0.00e+00	0.000e+00*
Severe asthma TAGC	1.13	1.09-1.18	3.24e-11	1.25	1.18-1.32	3.51e-14	1.52	1.4-1.64	0.00e+00	1.32	1.25-1.41	0.00e+00	2.605e-11*
FEV1/FVC	0.89	0.85-0.92	1.07e-10	0.82	0.77-0.87	3.01e-12	0.80	0.74-0.86	1.07e-08	0.86	0.81-0.91	1.20e-06	5.108e-02
Autism	0.92	0.89-0.96	1.30e-05	0.98	0.93-1.04	5.55e-01	1.01	0.94-1.09	7.68e-01	1.01	0.95-1.07	8.31e-01	3.392e-02*
Anxiety	1.00	0.96-1.03	8.82e-01	1.05	0.99-1.11	1.04e-01	1.00	0.92-1.08	9.44e-01	1.05	0.99-1.11	1.19e-01	3.795e-01
ADHD	1.13	1.08-1.17	2.46e-10	1.05	0.99-1.11	1.14e-01	1.02	0.94-1.1	6.83e-01	1.08	1.01-1.14	1.91e-02	5.108e-02
Bipolar	0.99	0.96-1.03	7.48e-01	0.96	0.91-1.02	2.18e-01	0.98	0.9-1.06	5.53e-01	0.97	0.91-1.03	3.05e-01	8.114e-01
Depression	1.05	1.02-1.09	3.55e-03	1.07	1.01-1.13	2.11e-02	1.03	0.96-1.11	4.38e-01	1.07	1.01-1.14	2.01e-02	8.114e-01
CRP	1.03	1-1.07	9.06e-02	1.03	0.98-1.08	2.71e-01	0.99	0.92-1.07	8.58e-01	1.03	0.97-1.09	3.15e-01	8.114e-01
Eosinophils	1.04	1-1.08	6.29e-02	1.04	0.98-1.1	1.92e-01	1.23	1.14-1.34	6.40e-07	1.17	1.1-1.25	1.40e-06	1.304e-04*
Neutrophils	1.01	0.97-1.05	5.01e-01	1.04	0.98-1.11	1.59e-01	1.06	0.98-1.15	1.54e-01	1.08	1.02-1.16	1.49e-02	3.735e-01

eosinophil counts, allergic rhinitis and sensitization, and autism showed notable differences in effect sizes. PRSs related to asthma subtypes had significantly higher effects for PR compared to other phenotypes, with PR showing the strongest genetic effects across all SNPs and PRSs. PR and LOR were strongly associated with allergy and eosinophil PRSs, while ETR and PTR showed no significant effects with these scores. The PRS for autism indicated a protective effect for ETR and was not significantly associated with other phenotypes after correction (see Table 4). Distinct associations were observed for allergic rhinitis, allergic sensitization, adult onset asthma, childhood onset asthma, severe asthma, autism and eosinophils.

Discussion

Main findings

In **paper I**, we highlight key genetic differences between wheeze phenotypes in preschool children. Specifically, we observed that children with recurrent wheezing or asthma symptoms until age 3 were genetically similar to adults with asthma, atopy, high eosinophils and low lung function. Genetic similarity to adult asthma and atopic traits were largely driven by children with persistent symptoms. Contrarily, genetic risk for low lung function was the only adult trait that showed significant overlap with transient wheezing in early childhood. Our results indicate that recurrent wheezing that remit at age 6, is genetically dissimilar to asthma and retention in lung function and environmental conditions are stronger influences for this phenotype.

In **paper II**, we discovered eight novel wheeze loci, near genes with roles in airway inflammation, infections, type-2 diabetes, growth and some loci with unknown molecular functions. Five out of eight novel wheeze loci replicate in an independent cohort supporting the identification of novel risk loci. Additionally, we employ a gene-prioritization framework that was able to reliably prioritize genes at 9 out of 23 loci. Furthermore, we identified trait-based SNP clusters with distinct associations to atopy, immune cells and lung function. Lastly, we show that genetic liability to severe wheeze can already be observed in infants that display increasing episodes of breathing problems with increasing wheeze liability. We also highlight molecular pathways involved in T2-response to infections in healthy children and pathways related to virus response and airway remodeling in wheezing infants.

In **paper III**, we identified four distinct trajectories of early-life ICS prescriptions, each shaped by unique risk factors. Early-transient redeemers exhibited low genetic influence but were highly impacted by *in utero* tobacco exposure, the presence of older siblings, and younger parents. This group was also associated with higher polygenic risk scores for ADHD. Prolonged-transient redeemers were characterized by maternal antibiotic use before birth, lower gestational age, a distinct association with the *FUT2* locus, no association with atopic risk scores, and strong

connections to asthma polygenic risk scores (PRS). Persistent redeemers showed a high prevalence of c-section births and male gender, with significant genetic effects across most examined SNPs and elevated atopic PRS, including asthma. Late-onset redeemers demonstrated minimal influence from perinatal risk factors but had high polygenic risk scores for asthma and atopic traits, with a strong yet nonspecific association to *HLA-DQA1*. Overall, our findings highlight that these phenotypes are distinct and shaped by various risk factors, indicating the presence of different endotypes.

Strengths and limitations

This thesis has several strengths that contribute to the robustness of the main findings. First, large sample size in all three studies ensures high statistical power to estimate precise statistical coefficients in the presence of a true effect. This is especially important in genome-wide association studies where effect sizes for SNPs are small and the large number of tests requires adequate sample size to acquire meaningful results (24). In **paper I**, we gathered longitudinal wheeze-status and genotypes on more than 18,000 children across 22 studies and in **paper II**, we obtained hospitalization status and genotypes of 304,092 children across four biobank-sized cohorts phenotyped. This leads to another strength of this thesis, which is the use of meta-analysis that minimizes bias of a single study design and resolves conflicting evidence by prioritizing estimates that are consistent across studies. Furthermore, ensuring that all cohorts in the meta-GWAS in **paper I** and **paper II** follow the same analysis plan, we minimize heterogeneity in phenotyping, caused by questionnaires, self-reported or different guidelines and practices in hospital regimens across countries.

Most GWAS of childhood asthma have focused on relatively broad phenotypes (26,92) and a few have leveraged phenotypes from 0-6 years and temporal definitions of wheeze (15,27,28). **Paper II**, leverages a unique phenotype of severe wheeze in preschoolers, defined by in-patient hospitalizations with asthma, bronchitis or bronchiolitis. Aggregating hospitalization records across countries with similar healthcare systems (Denmark, Iceland and Finland), further improves phenotype harmonization and studies with large sample size which allows the

discovery of loci with smaller effect sizes. Despite limited predictive utility of low-effect loci, inclusion of all genetic markers in polygenic prediction substantially improves performance compared to methods relying on fewer, but stronger loci (45), suggesting that robust GWAS findings, in aggregate, capture relevant biology despite small effect size.

Rigorous phenotypic definitions are expected to increase power of genetic studies of wheeze (56) and **paper II** have demonstrated that leveraging a severe wheeze phenotype coupled with large sample size allows for robust detection of novel loci. Despite pooling different diagnoses in a single severe phenotype, **paper II**, still manages to identify robust loci that replicate in an independent cohort with a milder phenotype defined on parentally reported wheeze. This highlights large sample size and use of a severe phenotype as a major strength of this thesis.

Paper I defines phenotypes based on complete remission or persistence by age six, resulting in more interpretable phenotypic distinctions compared to those derived from data-driven approaches, which may include a mix of symptomatic and asymptomatic individuals across time points. While the phenotyping approach in **paper I** enhances the rigor of phenotype interpretation, it may also exclude clinically relevant phenotypes. Data-driven approaches have identified several patterns of remission and persistence of wheeze symptoms, primarily in birth cohorts with longitudinal wheeze records (8,16,52,151). However, differences in sampling frequencies, timing and sample size have challenged replication of these phenotypes (51,66). Furthermore, birth cohorts often have missing data due to attrition (152) that complicates replication of cohorts across studies, partly due to reliance on imputation (51). To mitigate part of this problem, the use of electronic health records have been hypothesized to allow identification of more homogenous wheeze phenotypes (56).

Paper III utilizes electronic health records to identify temporal wheeze phenotypes through longitudinal patterns of inhaled corticosteroid (ICS) prescriptions, representing another significant strength of the thesis, since data collection occurs each time a patient interacts with the healthcare system (57). This approach facilitates real-world tracking of asthma-like symptoms that often present intermittently and may not be detectable during scheduled recruitment visits in a birth cohort study (58). Longitudinal records of ICS prescriptions coupled with perinatal risk factors derived from birth registries, genetics and polygenic risk scores

provides a strong foundation for **paper III** in proposing the existence of endotypes based on statistical distinctness of risk factor associations.

A solid data foundation is a major strength of **paper II**, where polygenic risk scores are calculated in the COPSAC cohorts. The COPSAC cohorts follow a prospective study design with detailed recordings of longitudinal measurements of asthma, atopy, exacerbations and infection from infancy and throughout childhood. This allows for explorative analysis of how genetic risk of wheeze affects the comprehensive clinical presentation of children with high genetic risk for severe wheeze. Furthermore, strict adherence to study protocols ensures a high degree of sampling consistency resulting in exceptional data quality and robust analyses, with limited influence of recall bias due to the reliance on both scheduled and acute visits.

Additional to the COPSAC cohorts, **paper II** also leverages data from the MARC-35 cohort enrolling infants after hospitalization with bronchiolitis across 14 US states in three consecutive bronchiolitis seasons. The multi-center design ensures a diverse representation of viral genotypes, allowing for broader conclusions about the consequences of bronchiolitis symptoms that might otherwise be biased by specific viral genotypes associated with particular hospitals or regions (18), possibly influencing severity (154,155). Furthermore, children were followed for 6 years after hospitalization through structured interviews to assess symptoms and development of asthma. This allows for detailed characterization of how genetic liability to wheeze influences progression to childhood asthma in a symptomatic cohort. Both COPSAC and MARC-35 provide RNA sequencing data sampled from the nasal epithelium of children allowing profiling of the transcriptomic signatures of high risk individuals to reveal potentially dysregulated pathways in nasal epithelial cells. Sampling in children represents a significant methodological strength, providing meaningful interpretation compared with the GWAS phenotype from 0-7 years in **paper II**.

The many strengths of this thesis are accompanied with several limitations. In **paper I**, we did not include late-onset wheeze due to low numbers of children with this phenotype, thus limiting our interpretations to children with wheeze at age three. Additionally, we had relatively sparse genotyping panels with on average ~2.4 million imputed SNPs. Therefore, summary statistics were imputed using z-score imputation using the Gaussian imputation algorithm, which managed to impute app. 500-600 thousand additional SNPs for each GWAS. This partially mitigated the

issue, but not entirely, limiting the ability to estimate overlap with related traits due to limited tagging of SNPs. Even though this study represents the largest GWAS on preschool wheeze, we only have a modest sample size compared to GWAS with strong genome-wide signals. This limited the use of robust methods for estimating genetic overlap, requiring alternative methods with fewer corrections for confounding. Another limitation is the inclusion of the remaining phenotypes in the control population for each GWAS. This was done to maximize the probability of detecting phenotype-specific signals, however a side-effect is a reduction in statistical power for general asthma genetics. This could potentially be an alternative explanation for the attenuated signal in the 17q locus for transient early wheeze.

Despite the retention of signals at the genome-wide significance level in **paper I**, it still represents the biggest GWAS study of temporal phenotypes of wheeze defined by remission or persistence between age three and six, but focusing on subtypes in a small age-gap will expectedly reduce sample size. In contrast, **paper II**, focused on increasing sample size to maximize power. Despite relying on a severe phenotype, severe asthma may be distinct from severe bronchitis and severe bronchiolitis, thus increasing heterogeneity of the phenotype. Evidence for phenotype heterogeneity is observed in the modest estimate of heritability and low-effect size of genetic variants. This represents a potential limitation that complicates the characterization of identified loci within the COPSAC cohorts, as the limited effect size in a relatively small population and challenges in phenotype harmonization of the GWAS, hinders decisions regarding the most appropriate COPSAC phenotype to designate as an endpoint, potentially increasing number of statistical tests. It is also a limitation of **paper II** that the replication was carried out in a phenotype based on parentally reported wheeze, since parents and medical doctors often disagree on definitions of wheeze (53,54), which is evident by the failure of replicating known asthma-loci, such as *FUT2*.

Lack of replication is a severe limitation of **paper III**, meaning that we cannot adequately ensure that our findings are robust enough to represent true endotypes. Replication is a possibility in the Danish Blood Donor Study (98,99), since the cohort contains both genotypes and EHR-based prescription records (156). Furthermore, the lack of observed measurements like serum IgE and eosinophils to robustly measure atopic-status and lung function measurements (FEV1/FVC), remain a challenge that is only partially mitigated by the inclusion of polygenic risk scores in **paper III**. Lastly, **paper III** is conducted in a cohort with overrepresentation of psychiatric

illnesses, and the adjusted associations between psychiatric traits and ICS-phenotypes suggest complex relationships between psychiatric traits, early wheeze and socioeconomic status. Shortly, asthma has previously been suggested to be a mediator between socioeconomic status and mental illnesses in children (157). However, it is unknown how these relationships may influence our results, but an in depth analysis is beyond the scope of this thesis and is considered as a theme for future research.

Additionally, all studies rely on individuals of European descent which severely limits the relevance of genetics studies, since interpretations cannot be extended to the general population. This caused all **paper II**'s analyses in the MARC-35 cohort to be restricted to children of European ancestry and thereby limiting statistical power and clinical relevance.

Interpretation

The literature is inconsistent regarding the association between transient wheeze and 17q locus. In contrast to our findings, multiple variants in the 17q-region have been associated with LCA-derived transient wheeze (0-11 years), in a multi-ethnic birth cohort (158), suggesting that 17q should be deemed a wheezing locus. In **paper I**, we defined transient wheeze based on children who wheezed more than 3 times or were diagnosed with asthma before age 3 and simultaneously free of wheeze and asthma at age 6. This is in contrast to the previously mentioned study, in which approx. 20 % of transient wheezers were symptomatic at age 6, creating a significant overlap with the persistent phenotype in **paper I**. In line with our findings, a study combining UK birth cohorts (STELAR) found no association between transient wheeze and 17q in 7719 children (151). It should be noted that some cohorts in the two studies described above occur in the discovery phase of this study, but our study also includes additional cohorts and the large sample size indicating that 17q is not a wheezing locus, but confined to persistent phenotypes. **Paper II** does not provide support for the specificity of the 17q locus, but **paper III** identifies a significantly lower effect of 17q in transient symptoms compared to persistent symptoms in an independent cohort. The likelihood ratio test employed in **paper III**, suggests that 17q might be associated with early-transient redeemers but the association is significantly weaker compared to the remaining phenotypes also displaying non-overlapping confidence

intervals. Association with 17q to the transient phenotype could also be a result of phenotypic heterogeneity in the early-transient phenotype, suggesting that 17q is in fact not a wheezing locus.

Paper II suggested the existence of 8 novel loci that had never been associated with asthma or a wheeze-related trait before, namely *CASC20*, *GDF5*, *CASC17*, *NR4A2*, *THSD4*, *KAZN-AS1*, *ZNF529* and *HNFB1B*. Additionally, by integrating the nasal transcriptome with the severe wheeze GWAS results, we identified two additional novel loci, namely *OAS1/OAS3* and *MTA3*. The loci *GDF5*, *NR4A2*, *CASC17*, *THSD4*, *CASC20*, *MTA3* and *OAS1/OAS3* all replicated in an independent cohort. Loci known to be associated with preschool asthma also replicated in the GWAS of **paper II**, but with lower effect size compared to GWASs of severe asthma with lower sample size (27,159). This suggests that many of the known loci are specific to asthma, i.e. *GSDMB*, *CDHR3*, *FUT2*, *HLA-DQA1*, *IL-33*, *IL1RL1*, *IL-13*, *ABO* and *TSLP*. These loci were utilized in **paper III** and distinct associations among the phenotypes were discovered, implying different genetic mechanisms. Most of the loci were strongly associated with persistent redeemers compared to other phenotypes and *CDHR3*, *IL-33*, *IL1RL1*, *IL-13* and *TSLP* were most strongly associated with persistent ICS use. The influence of *GSDMB*, *CDHR*, *IL33* and *IL1RL1* have all previously been implicated in temporal preschool wheeze phenotypes with evidence of distinct associations (151,158,160), however the influence of *IL-13*, *TSLP*, *FUT2* and *HLA-DQA1* on timing of wheeze symptoms have not been investigated.

IL-13 was only associated with persistent redeemers, and is involved in a number of pathways with T2-high response as the most established (161). *TSLP* is a cytokine with diverse roles in immunity (162), which is evident since the locus was associated with all the phenotypes in **paper III**. However, the *TSLP* was significantly stronger for persistent redeemers emphasizing a strong genetic component for this phenotype. Prolonged-transient redeemers were distinctly influenced by the *FUT2* locus, highlighting a role of infection susceptibility in this trajectory (14,15,16,17). Late-onset redeemers presented a significant, but not distinct, association with *HLA-DQA1* potentially excluding the HLA-region from conferring wheeze risk in early childhood.

The Tucson Children's Respiratory Study (N=125) observed that TEW had poor lung function shortly after birth, while PW had normal lung function capacity at birth coupled with a

subsequent decline until age six (7). In contrast, an Australian study (N=243) (163) reported that only PW had reduced lung function in infancy, compared to TEW which had normal lung function. COPSAC₂₀₀₀ with higher statistical power (N=403) showed PW correlated with reduced lung function shortly after birth, with further deterioration to age 7 years, and that TEW showed a weak signal of reduced lung function at birth (164). The ALSPAC study with a large sample size (N=6265) reported poor lung function at age 7 for both TEW and PW (8). Thus, the literature mainly agrees with TEW having reduced lung function compared to healthy controls and **paper I** suggests that TEW also carry genetic susceptibility to poor adult lung function driven by multiple lung function loci, most strongly seen for PEF. Contrarily only phenotypes characterized by recurrent wheeze were genetically similar to adult asthma, allergy and eosinophils levels. Consistently, **paper II** shows that severe wheeze has a substantial genetic overlap with asthma and lung function, atopic traits including eosinophil counts, implicating involvement of T2-high/atopic pathways in the etiology of severe wheeze, which likely represents a persistent phenotype. Severe wheeze also negatively correlates with height and growth patterns have previously been shown to be involved in wheezing (11–14) and childhood asthma has been suggested to attenuate adult height (15). Loci that drive this association may reveal mechanisms of impaired lung development leading to wheeze through attenuated fetal lung development (165). Severe wheeze was also positively correlated with adult BMI and genetic predisposition to high BMI has previously been linked to increased risk of severe wheeze in preschool children (166).

Paper II, tried to dissect the identified genetic correlations by partitioning genomic risk loci in groups with similar phenotypic associations. The association-based grouping analysis revealed suggestive clusters characterized by strong association with high eosinophil levels, risk of atopic disease, weak to moderate involvement with other immune cell counts and lung function. This cluster contains loci likely involved in T2-high pathways and progression to childhood asthma. Another cluster displayed weak associations with asthma, atopic traits and showed strong association with adult lung function coupled with weak signals for all immune cell counts. The genes in this cluster may prove to be interesting effectors of the progression from childhood wheeze and adult COPD (167–169). **Paper III**, examined genetic overlap through the use of polygenic risk scores and found that ETR was, contrary to **paper I**, associated with asthma scores, however the effects were significantly lower than for the remaining phenotypes.

Consistent with **paper I**, ETR, including PTR, were not influenced by genetic risk of eczema, allergy or eosinophil counts, confirming the presence of the non-atopic genetic signature in TEW suggested in **paper I**. PTR, which displayed later remission patterns compared to ETR, were more strongly influenced by genetic risk for asthma, indicating that this could represent non-atopic asthma. Similarly to **paper I**, **paper III** shows that PR is strongly associated with asthma and allergy genetics, whereas LOR displayed a similar association pattern to PR revealing some association patterns omitted in **paper I**. Additionally, lung function was investigated by using the PRS for FEV/FVC and it was equally associated with all phenotypes, including LOR. Only ETR and PTR were associated with genetic risk of high BMI consistent with the notion that overlap is thought to be higher among non-atopic compared to atopic asthmatics (77, 78). Our results indicate that the atopic phenotypes PR and LOR likely representing true asthma, are not influenced by genetic BMI.

We were not able to evaluate the impact of preschool wheeze phenotypes on clinical outcomes in **paper I**. In contrast, the polygenic risk score (PRS) for severe wheeze was associated with asthma and wheezing phenotypes, notably infant respiratory distress symptoms in both COPSAC cohorts and the MARC-35 cohort, in **paper II**. Specifically, respiratory tract infections, rhinovirus respiratory illnesses and fever episodes in the first year of life, were correlated with the PRS after correction for multiple testing in the COPSAC cohorts. Similarly, **paper III** showed that PR and PTR were moderately influenced by neonatal respiratory distress, maternal infection propensity of PTR and that all phenotypes were associated with hospitalizations for asthma, wheeze and pneumonia from 0-3 years, highlighting susceptibility to wheeze in infancy and early childhood. This is further underlined by **paper II**, by examining the wheeze-PRS in a cohort of infants with bronchiolitis. Among infants with severe bronchiolitis, those with high wheeze-PRS experienced more frequent respiratory symptoms in infancy and had an increased risk of developing asthma at age six. This highlights early infancy is a period for intervention to prevent wheeze developing to childhood asthma.

We were only able to perform functional enrichment in **paper II**, in order to implicate the involvement of biological pathways in preschool wheeze. We regressed genes expressed in nasal

epithelial cells on the severe wheeze PRS in both the COPAC2010 cohort and the MARC-35 study and carried out functional enrichment of suggestively associated genes.

Enrichment analysis in COPSAC₂₀₁₀ revealed pathways involved $\text{Fc}\gamma$ -receptor signaling, toll-like receptors cascades (TLR4, TLR7, TLR8 & TLR9), *S. aureus* and other bacterial and parasitic infections, highlighting potential molecular signatures for genetic risk of wheeze. In summary, binding of IgE to $\text{Fc}\gamma$ -receptor is a consequence of allergic inflammation in airway epithelium (172) and $\text{Fc}\gamma$'s increases susceptibility to bacterial infections by interacting with TLR4 (173). TLR4 is particularly responsive to bacterial lipopolysaccharide (LPS) and the capsid proteins of respiratory syncytial virus (RSV) (23), whereas TLR7 and TLR8 recognize single-stranded viral RNA (24). TLR9 is associated with eosinophil activation in response to viral ligands (25) and infections with *S. aureus* are closely linked to increased rates of wheezing and asthma exacerbations in children and young adults, as *S. aureus* endotoxins trigger IL-33 production in the airway mucosa (26). Although leishmaniasis and toxoplasmosis are of limited clinical relevance to asthma or wheeze (27, 28), their associated genes may be enriched due to eosinophilic pathway activation, commonly involved in parasitic infections (10), highlighting the importance type-2 proinflammatory genes in the nasal epithelium at six years of age.

Enrichment of PRS-associated genes in MARC-35 revealed interferon alpha ($\text{IFN-}\alpha$) and gamma ($\text{IFN-}\gamma$), MAPK signaling and epithelial mesenchymal transition (EMT). $\text{IFN-}\gamma$ in nasal epithelium of bronchiolitis infants is correlated with asthma progression at age 5 (21) and these results indicate that high genetic risk of severe wheeze may affect this process. Furthermore, MAPK signaling is a key process of transforming extracellular signals to intracellular responses and can be involved in a wide range of mechanisms, among apoptosis. Furthermore, EMT is the process by which epithelial cells are converted into mesenchymal cells (embryonic connective tissue), leading to fibrosis (174). EMT can be induced upon infection with rhinovirus (175), which has been shown to induce EMT factors such as IL-15RA in the presence of house dust mites (176). In comparison, RSV-infection does not seem to cause EMT in airway tissue (174). Interestingly, EMT is considered a core pathological mechanism for COPD (177), suggesting EMT as a possible avenue of research to delineate the relationship with childhood asthma.

Overall, the enriched pathways involve T2-high response to both bacterial and viral infections through recruitment of eosinophils and induction of IL-33 in children at age six and pathways related to virus response and airway remodeling in infants.

Conclusion

In summary, the studies presented in this thesis suggest the genetic heterogeneity of wheeze can be explained by the presence of genetic endotypes. This is highlighted by limited genetic overlap between atopic traits and transient wheeze, while persistent wheeze displayed stronger genetic similarity to atopic traits and asthma at older ages. Additionally, we identified loci in the GWAS of severe wheeze that had never before been associated with asthma suggesting wheeze-specific loci and we could further identify clusters of SNPs with distinct associations to atopy and lung function, suggesting the definition of severe wheeze from 0-7 contains several distinct subtypes. Lastly we identified potential endotypes based on timing of ICS prescriptions in early childhood. Identified trajectories displayed distinct associations to various risk factors, and results suggest that transient wheeze is less influenced by genetics and more influenced by environmental exposures. Contrarily, we confirmed that persistent wheezing is highly influenced by genetics, gender and mode of delivery. Conclusively, coherent results between the three papers suggests that wheeze endotypes exist and they are influenced by different mechanisms. Two groups are especially evident, namely a group influenced by atopic mechanisms associated with asthma, hayfever and eczema and another group without evidence of atopy and relatively strong risk for low adult lung function. This is evident from the genetic overlap analyses in **paper I**, association-based grouping analysis in **paper II** and the LRT analysis for PRS associations with ICS-phenotypes in **paper III**. A third, but less pronounced group is characterized by genetic risk for infections. No evidence for such a phenotype is provided in **paper I**, but SNP-clusters in **paper II** suggests that some genes like OAS3 and CDHR3 mediate wheeze risk through infectious susceptibility. Lastly, the increased antibiotic use of mothers in the prolonged-transient phenotype suggests an infection propensity that may or may not be inherited, coupled with associations to FUT2 and no association between the early-transient and prolonged phenotypes.

Future research

Given the knowledge I have acquired during my PhD, I would take a different approach if I were to initiate **paper I** today. I would have leveraged cross-trait meta-analysis with related traits using MTAG (178), in order to identify shared signals between traits. In order to prevent

drowning of the wheeze signals, I select subsampled individuals from UK Biobank to align sample size and perform GWAS on adult asthma, eosinophils, neutrophils, lung function in the subsampled UK Biobank for inclusion in cross-trait meta-analysis. This could possibly have highlighted key SNPs that drive shared signals pointing towards more specific mechanisms for temporal preschool wheeze.

A different direction for future research could involve systematically assessing whether GSMR yields unbiased or comparable estimates to genetic correlation through a proof-of-concept paper. This investigation could be conducted through simulations of GWAS results with different sample size and heritability. Such a study would be of considerable interest for publication, particularly because GWAS of disease subtypes or rare phenotypes frequently yield few genome-wide significant variants, thereby limiting the applicability of genetic correlation via LD-score regression. This could potentially highlight the utility of alternative methods to estimate genetic correlation and extract additional value from low-powered GWAS of rare phenotypes.

In future studies based on **paper II**, it could be of interest to extend the discovery population to individuals of all ancestries. In the iPSYCH population some individuals are of middle eastern ancestry, and coupled with the Qatari Birth Cohort (179), we could extend findings of this GWAS to middle eastern individuals who constitute a significant share of the population in the Nordic countries. Integration of the eMERGE network (180) and Mass General biobank (181) into the current meta-analysis provide interesting avenues to leverage the full potential of the MARC-35 cohort in order to decipher the role genetics in the progression from severe bronchiolitis to asthma.

Since we observed a subset of loci that was weakly associated with asthma and strongly associated with adult lung function, a cross-trait GWAS between preschool wheeze and adult lung function (FEV1/FVC) (182) could be of considerable interest as this could pinpoint specific loci acting on preschool wheeze and adult COPD. This could reveal interesting molecular functions involved in life-time pulmonary deterioration leading to improved treatment and prevention of adult COPD.

Additionally, we could improve the gene-prioritization framework by investigating how to

weight each method according to their ability to prioritize known asthma genes. This could be supported by literature review of the experimental literature with the final aim to compile a list of “ground truth” genes for childhood asthma loci.

Experimental studies with the aim of characterizing the function of *GDF5* locus, could add tremendous value to this study. Such studies could focus on bronchial fibroblasts, since *GDF5* seems to be expressed in both lung tissue and fibroblasts. Mouse models could investigate the role of *GDF5*, in growth (as the genetic variant is strongly associated with human height) and explore the role of this protein in airway inflammation.

As an extension of **paper III**, It could be of immediate interest to identify ICS trajectories in the general population, allowing for larger sample size, as this holds the potential to discover novel temporal phenotypes with low prevalence. Such a study would leverage the full potential of the electronic health records with minimal ascertainment bias, and an increased sample size would allow more robust study of rare events like neonatal respiratory distress, and medication during pregnancy. Phenotypes identified in a large cohort, could be genetically profiled in a smaller cohort using polygenic risk scores as outcomes for GWAS.

Additionally, It would be of substantial interest to the research community to elucidate how psychiatric overrepresentation in the iPSYCH cohort influences analyses of somatic illnesses. A proposed strategy for such a study could be to perform mediation analysis, with psychiatric traits as a mediator between all analysis in **paper III**. It could also be of interest to quantify how much gene-environment correlations contribute to polygenic risk scores of psychiatric traits (183).

References

1. Lai CKW, Beasley R, Crane J, Foliaki S, Shah J, Weiland S, et al. Global variation in the prevalence and severity of asthma symptoms: Phase Three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax*. 2009 Jun 1;64(6):476–83.
2. He P, Moraes TJ, Dai D, Reyna-Vargas ME, Dai R, Mandhane P, et al. Early prediction of pediatric asthma in the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort using machine learning. *Pediatr Res*. 2024 Jun;95(7):1818–25.
3. Neuman Å, Bergström A, Gustafsson P, Thunqvist P, Andersson N, Nordvall L, et al. Infant wheeze, comorbidities and school age asthma. *Pediatr Allergy Immunol*. 2014 Jun;25(4):380–6.
4. Amin P, Levin L, Epstein T, Ryan P, LeMasters G, Khurana Hershey G, et al. Optimum Predictors of Childhood Asthma: Persistent Wheeze or the Asthma Predictive Index? *J Allergy Clin Immunol Pract*. 2014 Nov;2(6):709-715.e2.
5. Gaillard EA, Kuehni CE, Turner S, Goutaki M, Holden KA, De Jong CCM, et al. European Respiratory Society clinical practice guidelines for the diagnosis of asthma in children aged 5–16 years. *Eur Respir J*. 2021 Nov;58(5):2004173.
6. Makrinioti H, Hasegawa K, Lakoumentas J, Xepapadaki P, Tsolia M, Castro-Rodriguez JA, et al. The role of respiratory syncytial virus- and rhinovirus-induced bronchiolitis in recurrent wheeze and asthma—A systematic review and meta-analysis. *Pediatr Allergy Immunol*. 2022 Mar;33(3):e13741.
7. 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.
8. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax*. 2008 Nov 1;63:974–80.
9. Lindensmith J, Morrison D, Deveau C, Hernandez P. Overdiagnosis of Asthma in the Community. *Can Respir J*. 2004 Jan;11(2):111–6.
10. Mukherjee M, Stoddart A, Gupta RP, Nwaru BI, Farr A, Heaven M, et al. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Med*. 2016 Dec;14(1):113.
11. Ullemar V, Magnusson PKE, Lundholm C, Zettergren A, Melén E, Lichtenstein P, et al. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy*. 2016 Feb;71(2):230–8.
12. Thomsen SF. Exploring the origins of asthma: Lessons from twin studies. *Eur Clin Respir J*. 2014 Jan;1(1):25535.

13. Hanson B, McGue M, Roitman-Johnson B, Segal NL, Bouchard TJ, Blumenthal MN. Atopic disease and immunoglobulin E in twins reared apart and together. *Am J Hum Genet.* 1991 May;48(5):873–9.
14. Haider S, Granell R, Curtin J, Fontanella S, Cucco A, Turner S, et al. Modeling Wheezing Spells Identifies Phenotypes with Different Outcomes and Genetic Associates. *Am J Respir Crit Care Med.* 2022 Apr 15;205(8):883–93.
15. Granell R, Curtin JA, Haider S, Kitaba NT, Mathie SA, Gregory LG, et al. A meta-analysis of genome-wide association studies of childhood wheezing phenotypes identifies ANXA1 as a susceptibility locus for persistent wheezing. *eLife.* 2023 May 25;12:e84315.
16. Dai R, Miliku K, Gaddipati S, Choi J, Ambalavanan A, Tran MM, et al. Wheeze trajectories: Determinants and outcomes in the CHILD Cohort Study. *J Allergy Clin Immunol.* 2022 Jun;149(6):2153–65.
17. Klaassen EMM, Penders J, Jöbsis Q, Van De Kant KDG, Thijs C, Mommers M, et al. An ADAM33 Polymorphism Associates with Progression of Preschool Wheeze into Childhood Asthma: A Prospective Case-Control Study with Replication in a Birth Cohort Study. He W, editor. *PLOS ONE.* 2015 Mar 13;10(3):e0119349.
18. Klaassen EMM, Van De Kant KDG, Jöbsis Q, Penders J, Van Schooten FJ, Quaak M, et al. Integrative genomic analysis identifies a role for intercellular adhesion molecule 1 in childhood asthma. *Pediatr Allergy Immunol.* 2014 Mar;25(2):166–72.
19. Skytt N, Bønnelykke K, Bisgaard H. “To wheeze or not to wheeze”: That is not the question. *J Allergy Clin Immunol.* 2012 Aug;130(2):403-407.e5.
20. Raita Y, Camargo CA, Bochkov YA, Celedón JC, Gern JE, Mansbach JM, et al. Integrated-omics endotyping of infants with rhinovirus bronchiolitis and risk of childhood asthma. *J Allergy Clin Immunol.* 2021 Jun;147(6):2108–17.
21. Raita Y, Pérez-Losada M, Freishtat RJ, Harmon B, Mansbach JM, Piedra PA, et al. Integrated omics endotyping of infants with respiratory syncytial virus bronchiolitis and risk of childhood asthma. *Nat Commun.* 2021 Jun 14;12(1):3601.
22. Zhu Z, Shibata R, Hoffman KL, Cormier J, Mansbach JM, Liang L, et al. Integrated nasopharyngeal airway metagenome and asthma genetic risk endotyping of severe bronchiolitis in infancy and risk of childhood asthma. *Eur Respir J.* 2024 Sep 26;2401130.
23. Kerkhof M, Boezen HM, Granell R, Wijga AH, Brunekreef B, Smit HA, et al. Transient early wheeze and lung function in early childhood associated with chronic obstructive pulmonary disease genes. *J Allergy Clin Immunol.* 2014 Jan;133(1):68-76.e1-4.
24. Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primer.* 2021 Aug 26;1(1):59.
25. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med.* 2019 Jun 1;7(6):509–22.
26. Ferreira MAR, Mathur R, Vonk JM, Sz wajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am J Hum*

- Genet. 2019 Apr 4;104(4):665–84.
27. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet.* 2014 Jan;46(1):51–5.
 28. Ahluwalia TS, Eliassen AU, Sevelsted A, Pedersen CET, Stokholm J, Chawes B, et al. FUT2–ABO epistasis increases the risk of early childhood asthma and Streptococcus pneumoniae respiratory illnesses. *Nat Commun.* 2020 Dec 16;11:6398.
 29. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature.* 2007 Jul 26;448(7152):470–3.
 30. Kim KW, Ober C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res.* 2019;11(2):170.
 31. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, et al. A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. *J Allergy Clin Immunol.* 2018 Sep;142(3):749-764.e3.
 32. Das S, Miller M, Broide DH. Chapter One - Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases. In: Alt FW, editor. *Advances in Immunology* [Internet]. Academic Press; 2017 [cited 2020 Aug 27]. p. 1–52. Available from: <http://www.sciencedirect.com/science/article/pii/S006527761730038X>
 33. Liu T, Liu S, Rui X, Cao Y, Hecker J, Guo F, et al. Gasdermin B, an asthma-susceptibility gene, promotes MAVS-TBK1 signalling and airway inflammation. *Eur Respir J.* 2024 May;63(5):2301232.
 34. Çalışkan M, Bochkov YA, Kreiner-Møller E, Bønnelykke K, Stein MM, Du G, et al. Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med.* 2013 Apr 11;368(15):1398–407.
 35. Bochkov YA, Watters K, Ashraf S, Griggs TF, Devries MK, Jackson DJ, et al. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc Natl Acad Sci U S A.* 2015 Apr 28;112(17):5485–90.
 36. Bønnelykke K, Coleman AT, Evans MD, Thorsen J, Waage J, Vissing NH, et al. Cadherin-related Family Member 3 Genetics and Rhinovirus C Respiratory Illnesses. *Am J Respir Crit Care Med.* 2018 Mar;197(5):589—594.
 37. Eliassen AU, Pedersen CET, Rasmussen MA, Wang N, Soverini M, Fritz A, et al. Genome-wide study of early and severe childhood asthma identifies interaction between CDHR3 and GSDMB. *J Allergy Clin Immunol.* 2022 Sep;150(3):622–30.
 38. Barton SJ, Murray R, Lillycrop KA, Inskip HM, Harvey NC, Cooper C, et al. *FUT2* Genetic Variants and Reported Respiratory and Gastrointestinal Illnesses During Infancy. *J Infect Dis.* 2019 Feb 15;219(5):836–43.
 39. Kindberg E, Akerlind B, Johnsen C, Knudsen JD, Heltberg O, Larson G, et al. Host genetic resistance to symptomatic norovirus (GGII.4) infections in Denmark. *J Clin Microbiol.*

2007 Aug;45(8):2720–2.

40. Imbert-Marcille BM, Barbé L, Dupé M, Le Moullac-Vaidye B, Besse B, Peltier C, et al. A FUT2 Gene Common Polymorphism Determines Resistance to Rotavirus A of the P[8] Genotype. *J Infect Dis.* 2014 Apr 15;209(8):1227–30.
41. Azevedo M, Eriksson S, Mendes N, Serpa J, Figueiredo C, Resende L, et al. Infection by *Helicobacter pylori* expressing the BabA adhesin is influenced by the secretor phenotype. *J Pathol.* 2008 Jul;215(3):308–16.
42. Rogliani P, Laitano R, Ora J, Beasley R, Calzetta L. Strength of association between comorbidities and asthma: a meta-analysis. *Eur Respir Rev.* 2023 Mar 31;32(167):220202.
43. De Groot EP, Duiverman EJ, Brand PLP. Comorbidities of asthma during childhood: possibly important, yet poorly studied. *Eur Respir J.* 2010 Sep 1;36(3):671–8.
44. Van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. *Nat Rev Genet.* 2019 Oct;20(10):567–81.
45. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun.* 2019 Apr 16;10(1):1776.
46. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015 Mar;47(3):291–5.
47. Kraft P, Chen H, Lindström S. The Use of Genetic Correlation and Mendelian Randomization Studies to Increase Our Understanding of Relationships between Complex Traits. *Curr Epidemiol Rep.* 2020 Jun;7(2):104–12.
48. Tsuo K, Zhou W, Wang Y, Kanai M, Namba S, Gupta R, et al. Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. *Cell Genomics.* 2022 Dec;2(12):100212.
49. Udler MS, Kim J, Von Grotthuss M, Bonàs-Guarch S, Cole JB, Chiou J, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *Langenberg C, editor. PLOS Med.* 2018 Sep 21;15(9):e1002654.
50. Granell R, Henderson AJ, Sterne JA. Associations of wheezing phenotypes with late asthma outcomes in the Avon Longitudinal Study of Parents and Children: A population-based birth cohort. *J Allergy Clin Immunol.* 2016 Oct;138(4):1060-1070.e11.
51. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol.* 2011 Jun;127(6):1505-1512.e14.
52. Oksel C, Granell R, Haider S, Fontanella S, Simpson A, Turner S, et al. Distinguishing Wheezing Phenotypes from Infancy to Adolescence. A Pooled Analysis of Five Birth Cohorts. *Ann Am Thorac Soc.* 2019 Jul;16(7):868–76.
53. Michel G, Silverman M, Strippoli MPF, Zwahlen M, Brooke AM, Grigg J, et al. Parental understanding of wheeze and its impact on asthma prevalence estimates. *Eur Respir J.* 2006

Dec 1;28(6):1124–30.

54. Levy ML, Godfrey S, Irving CS, Sheikh A, Hanekom W, Bush A, et al. Wheeze Detection: Recordings vs. Assessment of Physician and Parent. *J Asthma*. 2004 Jan;41(8):845–53.
55. Larsson H. The importance of selection bias in prospective birth cohort studies. *JCPP Adv*. 2021 Oct;1(3):e12043.
56. Wolters AAB, Kersten ETG, Koppelman GH. Genetics of preschool wheeze and its progression to childhood asthma. *Pediatr Allergy Immunol*. 2024 Jan;35(1):e14067.
57. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annu Rev Public Health*. 2016 Mar 18;37(1):61–81.
58. McCormack MC, Enright PL. Making the diagnosis of asthma. *Respir Care*. 2008 May;53(5):583–90; discussion 590-592.
59. Brand PLP, Caudri D, Eber E, Gaillard EA, Garcia-Marcos L, Hedlin G, et al. Classification and pharmacological treatment of preschool wheezing: changes since 2008. *Eur Respir J*. 2014 Apr;43(4):1172–7.
60. Pescatore AM, Dogaru CM, Duembgen L, Silverman M, Gaillard EA, Spycher BD, et al. A simple asthma prediction tool for preschool children with wheeze or cough. *J Allergy Clin Immunol*. 2014 Jan;133(1):111-118.e13.
61. Mommers I, Van Boven JF, Schuiling-Veninga CC, Bos JH, Koetsier M, Hak E, et al. Real-World Dispensing Patterns of Inhalation Medication in Young Adult Asthma: An Inception Cohort Study. *Clin Epidemiol*. 2023 Jun;Volume 15:721–32.
62. Von Bülow A, Hansen S, Sandin P, Ernstsson O, Janson C, Lehtimäki L, et al. Severe asthma trajectories in adults: findings from the NORDSTAR cohort. *Eur Respir J*. 2023 Sep;62(3):2202474.
63. Lee E, Rhee EH, Kim K, Kim HS, Kim WK, Song DJ, et al. Frequency of exacerbation and degree of required asthma medication can characterize childhood longitudinal asthma trajectories. *Ann Allergy Asthma Immunol*. 2023 Oct;131(4):444–50.
64. Markus AF, Rijnbeek PR, Kors JA, Burn E, Duarte-Salles T, Haug M, et al. Real-world treatment trajectories of adults with newly diagnosed asthma or COPD. *BMJ Open Respir Res*. 2024 Feb;11(1):e002127.
65. Pape K, Cowell W, Sejbaek CS, Andersson NW, Svanes C, Kolstad HA, et al. Adverse childhood experiences and asthma: trajectories in a national cohort. *Thorax*. 2021 Jun;76(6):547–53.
66. Oksel C, Granell R, Mahmoud O, Custovic A, Henderson AJ. Causes of variability in latent phenotypes of childhood wheeze. *J Allergy Clin Immunol*. 2019 May;143(5):1783-1790.e11.
67. Sinha P, Calfee CS, Delucchi KL. Practitioner’s Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. *Crit Care Med*. 2021 Jan;49(1):e63–79.
68. Mei JV, Alexander JR, Adam BW, Hannon WH. Use of filter paper for the collection and

- analysis of human whole blood specimens. *J Nutr.* 2001 May;131(5):1631S-6S.
69. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry.* 2018 Jan;23(1):6–14.
 70. Hollegaard MV, Grove J, Grauholm J, Kreiner-Møller E, Bønnelykke K, Nørgaard M, et al. Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet.* 12:58–58.
 71. Wang X, Liu Y, Liu H, Pan W, Ren J, Zheng X, et al. Recent advances and application of whole genome amplification in molecular diagnosis and medicine. *MedComm.* 2022 Mar;3(1):e116.
 72. Zhao S, Jing W, Samuels DC, Sheng Q, Shyr Y, Guo Y. Strategies for processing and quality control of Illumina genotyping arrays. *Brief Bioinform.* 2018 Sep 28;19(5):765–75.
 73. Illumina. Illumina GenCall Data Analysis Software. [Internet]. Illumina; Available from: https://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf.
 74. Shi M, Tanikawa C, Munter HM, Akiyama M, Koyama S, Tomizuka K, et al. Genotype imputation accuracy and the quality metrics of the minor ancestry in multi-ancestry reference panels. *Brief Bioinform.* 2023 Nov 22;25(1):bbad509.
 75. International HapMap Consortium. The International HapMap Project. *Nature.* 2003 Dec 18;426(6968):789–96.
 76. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015 Oct 1;526(7571):68–74.
 77. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc.* 2015 Nov;2015(11):pdb.top084970.
 78. Houseley J, Tollervey D. The Many Pathways of RNA Degradation. *Cell.* 2009 Feb;136(4):763–76.
 79. Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, et al. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA.* 2015 Oct;21(10):1731–45.
 80. Andrews S FastQC: a quality control tool for high throughput sequence data. [Internet]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
 81. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016 Oct 1;32(19):3047–8.
 82. Khodasevich D, Smith AR, Huen K, Eskenazi B, Cardenas A, Holland N. Comparison of DNA methylation measurements from EPIC BeadChip and SeqCap targeted bisulphite sequencing in PON1 and nine additional candidate genes. *Epigenetics.* 2022 Dec 9;17(13):1944–55.
 83. Contributed by Daniel J. Weisenberger, David Van Den Berg, Fei Pan, Benjamin P.

- Berman,, and Peter W. Laird. Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform [Internet]. Available from:
https://www.illumina.com/documents/products/appnotes/appnote_dna_methylation_analysis_infinium.pdf
84. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet.* 2005 May;37(5):549–54.
 85. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. Whole-genome genotyping with the single-base extension assay. *Nat Methods.* 2006 Jan;3(1):31–3.
 86. EARly Genetics Lifecourse Epidemiology (EAGLE) consortium, Early Growth Genetics (EGG) consortium, Middeldorp CM, Felix JF, Mahajan A, McCarthy MI. The Early Growth Genetics (EGG) and EARly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. *Eur J Epidemiol.* 2019 Mar;34(3):279–300.
 87. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med.* 2010 Sep 23;363(13):1211–21.
 88. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015 Oct 1;526(7571):68–74.
 89. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma Oxf Engl.* 2010 Sep 1;26(17):2190–1.
 90. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet.* 1995 Nov;11(3):241–7.
 91. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat Genet.* 2018 Aug;50(8):1072.
 92. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med.* 2019 Jun;7(6):509–22.
 93. Sliz E, Huilaja L, Pasanen A, Laisk T, Reimann E, Mägi R, et al. Uniting biobank resources reveals novel genetic pathways modulating susceptibility for atopic dermatitis. *J Allergy Clin Immunol.* 2022 Mar;149(3):1105-1112.e9.
 94. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell.* 2020 Sep 3;182(5):1214-1231.e11.
 95. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet.* 2019 Mar;51(3):481–93.
 96. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun.* 2018 Jan 15;9(1):224.

97. Bybjerg-Grauholm J, Bøcker Pedersen C, Bækvad-Hansen M, Giørtz Pedersen M, Adamsen D, Søholm Hansen C, et al. The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders [Internet]. *Epidemiology*; 2020 Dec [cited 2023 Mar 20]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.11.30.20237768>
98. Erikstrup C, Sørensen E, Nielsen KR, Bruun MT, Petersen MS, Rostgaard K, et al. Cohort Profile: The Danish Blood Donor Study. *Int J Epidemiol*. 2022 Oct 4;dyac194.
99. Hansen TF, Banasik K, Erikstrup C, Pedersen OB, Westergaard D, Chmura PJ, et al. DBDS Genomic Cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. *BMJ Open*. 2019 Jun;9(6):e028401.
100. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*. 2015 May;47(5):435–44.
101. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009 Dec 17;462(7275):868–74.
102. Eggertsson HP, Jonsson H, Kristmundsdottir S, Hjartarson E, Kehr B, Masson G, et al. GraphTyper enables population-scale genotyping using pangenome graphs. *Nat Genet*. 2017 Nov;49(11):1654–60.
103. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023 Jan 19;613(7944):508–18.
104. Bisgaard H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol*. 2004 Oct;93(4):381–9.
105. Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Følsgaard NV, Kreiner-Møller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clin Exp Allergy J Br Soc Allergy Clin Immunol*. 2013 Dec;43(12):1384–94.
106. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014 Feb 3;15(2):R29.
107. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010 Mar 2;11(3):R25.
108. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Hancock J, editor. *Bioinformatics*. 2017 Feb 15;33(4):558–60.
109. Morin A, McKennan CG, Pedersen CET, Stokholm J, Chawes BL, Malby Schoos AM, et al. Epigenetic landscape links upper airway microbiota in infancy with allergic rhinitis at 6 years of age. *J Allergy Clin Immunol* [Internet]. 2020 Jul 18 [cited 2020 Oct 21]; Available from: <http://www.sciencedirect.com/science/article/pii/S0091674920309672>

110. Hasegawa K, Mansbach JM, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalised for bronchiolitis. *Eur Respir J*. 2016 Nov;48(5):1329–39.
111. on behalf of the MARC-35 Investigators, Mansbach JM, Hasegawa K, Geller RJ, Espinola JA, Sullivan AF, et al. Bronchiolitis severity is related to recurrent wheezing by age 3 years in a prospective, multicenter cohort. *Pediatr Res*. 2020 Feb;87(3):428–30.
112. Ralston SL, Lieberthal AS, Meissner HC, Alverson BK, Baley JE, Gadomski AM, et al. Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatrics*. 2014 Nov 1;134(5):e1474–502.
113. Camargo CA, Ingham T, Wickens K, Thadhani R, Silvers KM, Epton MJ, et al. Cord-Blood 25-Hydroxyvitamin D Levels and Risk of Respiratory Infection, Wheezing, and Asthma. *Pediatrics*. 2011 Jan 1;127(1):e180–7.
114. Hasegawa K, Stewart CJ, Celedón JC, Mansbach JM, Tierney C, Camargo CA. Circulating 25-hydroxyvitamin D, nasopharyngeal airway metabolome, and bronchiolitis severity. *Allergy*. 2018 May;73(5):1135–40.
115. Dumas O, Hasegawa K, Mansbach JM, Sullivan AF, Piedra PA, Camargo CA. Severe bronchiolitis profiles and risk of recurrent wheeze by age 3 years. *J Allergy Clin Immunol*. 2019 Apr;143(4):1371-1379.e7.
116. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017 Apr;14(4):417–9.
117. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2016 Feb 29;4:1521.
118. Raita Y, Pérez-Losada M, Freishtat RJ, Hahn A, Castro-Nallar E, Ramos-Tapia I, et al. Nasopharyngeal metatranscriptome profiles of infants with bronchiolitis and risk of childhood asthma: a multicentre prospective study. *Eur Respir J*. 2022 Jul;60(1):2102293.
119. Zhu Z, Camargo CA, Raita Y, Freishtat RJ, Fujiogi M, Hahn A, et al. Nasopharyngeal airway dual-transcriptome of infants with severe bronchiolitis and risk of childhood asthma: A multicenter prospective study. *J Allergy Clin Immunol*. 2022 Apr 26;S0091-6749(22)00556-5.
120. Fujiogi M, Raita Y, Pérez-Losada M, Freishtat RJ, Celedón JC, Mansbach JM, et al. Integrated relationship of nasopharyngeal airway host response and microbiome associates with bronchiolitis severity. *Nat Commun*. 2022 Aug 30;13(1):4970.
121. Helgeland Ø, Vaudel M, Juliusson PB, Lingaas Holmen O, Juodakis J, Bacelis J, et al. Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. *Nat Commun*. 2019 Oct 1;10(1):4448.
122. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015 Dec;4(1):7.
123. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006 Aug;38(8):904–9.

124. Karolchik D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004 Jan 1;32(90001):493D – 496.
125. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010 Sep 1;38(16):e164–e164.
126. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science.* 2023 Sep 22;381(6664):eadg7492.
127. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun.* 2015;6:5890.
128. Weeks EM, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat Genet.* 2023 Aug;55(8):1267–76.
129. De Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. Tang H, editor. *PLOS Comput Biol.* 2015 Apr 17;11(4):e1004219.
130. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016 Mar;48(3):245–52.
131. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016 Dec;17(1):122.
132. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015 Nov;47(11):1236–41.
133. Budu-Aggrey A, Kilanowski A, Sobczyk MK, 23andMe Research Team, Shringarpure SS, Mitchell R, et al. European and multi-ancestry genome-wide association meta-analysis of atopic dermatitis highlights importance of systemic immune regulation. *Nat Commun.* 2023 Oct 4;14(1):6172.
134. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun.* 2017 Sep 19;8(1):599.
135. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet.* 2018 Oct 15;27(20):3641–9.
136. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. *Nature.* 2022 Oct 27;610(7933):704–12.
137. Said S, Pazoki R, Karhunen V, Võsa U, Ligthart S, Bodinier B, et al. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nat Commun.* 2022 Apr 22;13(1):2198.
138. Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, Davey Smith G, et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target

- mendelian randomisation. Locasale JW, editor. PLOS Biol. 2022 Feb 25;20(2):e3001547.
139. Hammond RK, Pahl MC, Su C, Cousminer DL, Leonard ME, Lu S, et al. Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *eLife*. 2021 Jan 18;10:e62206.
 140. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559–75.
 141. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015 Apr 20;43(7):e47–e47.
 142. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013 Dec;14(1):128.
 143. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W90–7.
 144. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc*. 2021 Mar;1(3):e90.
 145. Namjou B, Lape M, Malolepsza E, DeVore SB, Weirauch MT, Dikilitas O, et al. Multiancestral polygenic risk score for pediatric asthma. *J Allergy Clin Immunol*. 2022 Nov;150(5):1086–96.
 146. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988 Sep;44(3):837–45.
 147. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol*. 2016 Oct 27;dyw213.
 148. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet*. 2023 Feb;55(2):198–208.
 149. Linzer DA, Lewis JB. **poLCA** : An R Package for Polytomous Variable Latent Class Analysis. *J Stat Softw* [Internet]. 2011 [cited 2024 Sep 30];42(10). Available from: <http://www.jstatsoft.org/v42/i10/>
 150. Tein JY, Cox S, Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Struct Equ Model Multidiscip J*. 2013 Oct;20(4):640–57.
 151. Haider S, Granell R, Curtin J, Fontanella S, Cucco A, Turner S, et al. Modeling Wheezing Spells Identifies Phenotypes with Different Outcomes and Genetic Associates. *Am J Respir Crit Care Med*. 2022 Apr 15;205(8):883–93.
 152. Teixeira R, Doetsch J, Freitas AI, Lorthe E, Santos AC, Barros H, et al. Survey of data collection methods and retention strategies in European birth cohorts of children and adults

- born very preterm. *Paediatr Perinat Epidemiol*. 2022 Sep;36(5):706–14.
153. Hasegawa K, Mansbach JM, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalized for bronchiolitis. *Eur Respir J*. 2016 Nov;48(5):1329–39.
 154. Martinello RA, Chen MD, Weibel C, Kahn JS. Correlation between Respiratory Syncytial Virus Genotype and Severity of Illness. *J Infect Dis*. 2002 Sep 15;186(6):839–42.
 155. Papadopoulos NG, Gourgiotis D, Javadyan A, Bossios A, Kallergi K, Psarras S, et al. Does respiratory syncytial virus subtype influences the severity of acute bronchiolitis in hospitalized infants? *Respir Med*. 2004 Sep;98(9):879–82.
 156. Ghouse J, Ahlberg G, Andreassen L, Banasik K, Brunak S, Schwinn M, et al. Association of Variants Near the Bradykinin Receptor B2 Gene With Angioedema in Patients Taking ACE Inhibitors. *J Am Coll Cardiol*. 2021 Aug;78(7):696–709.
 157. Omura M, Cortese S, Bailhache M, Navarro MC, Melchior M, Van Der Waerden J, et al. Associations between symptoms of attention-deficit hyperactivity disorder, socioeconomic status and asthma in children. *Npj Ment Health Res*. 2024 Apr 16;3(1):22.
 158. Hallmark B, Wegienka G, Havstad S, Billheimer D, Ownby D, Mendonca EA, et al. Chromosome 17q12-21 Variants Are Associated with Multiple Wheezing Phenotypes in Childhood. *Am J Respir Crit Care Med*. 2021 Apr 1;203(7):864–70.
 159. Ahluwalia TS, Eliassen AU, Sevelsted A, Pedersen CET, Stokholm J, Chawes B, et al. FUT2-ABO epistasis increases the risk of early childhood asthma and *Streptococcus pneumoniae* respiratory illnesses. *Nat Commun*. 2020 Dec 16;11(1):6398.
 160. Savenije OE, Mahachie John JM, Granell R, Kerkhof M, Dijk FN, De Jongste JC, et al. Association of IL33–IL-1 receptor–like 1 (IL1RL1) pathway polymorphisms with wheezing phenotypes and asthma in childhood. *J Allergy Clin Immunol*. 2014 Jul;134(1):170–7.
 161. Wills-Karp M. Interleukin-13 in asthma pathogenesis. *Immunol Rev*. 2004 Dec;202(1):175–90.
 162. Ebina-Shibuya R, Leonard WJ. Role of thymic stromal lymphopoietin in allergy and beyond. *Nat Rev Immunol*. 2023 Jan;23(1):24–37.
 163. Turner SW, Palmer LJ, Rye PJ, Gibson NA, Judge PK, Cox M, et al. The Relationship between Infant Airway Function, Childhood Airway Responsiveness, and Asthma. *Am J Respir Crit Care Med*. 2004 Apr 15;169(8):921–7.
 164. Bisgaard H, Jensen SM, Bønnelykke K. Interaction between asthma and lung function growth in early life. *Am J Respir Crit Care Med*. 2012 Jun 1;185(11):1183–9.
 165. Maritz GS, Cock ML, Louey S, Joyce BJ, Albuquerque CA, Harding R. Effects of fetal growth restriction on lung development before and after birth: A morphometric analysis. *Pediatr Pulmonol*. 2001 Sep;32(3):201–10.
 166. Jensen SK, Pedersen CET, Fischer-Rasmussen K, Melgaard ME, Brustad N, Kyvsgaard JN, et al. Genetic predisposition to high BMI increases risk of early life respiratory infections and episodes of severe wheeze and asthma. *Eur Respir J*. 2024 Sep;64(3):2400169.

167. Bui DS, Lodge CJ, Burgess JA, Lowe AJ, Perret J, Bui MQ, et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med*. 2018 Jul;6(7):535–44.
168. Bisgaard H, Nørgaard S, Sevelsted A, Chawes BL, Stokholm J, Mortensen EL, et al. Asthma-like symptoms in young children increase the risk of COPD. *J Allergy Clin Immunol*. 2021 Feb;147(2):569-576.e9.
169. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N Engl J Med*. 2016 May 12;374(19):1842–52.
170. Appleton S, Adams R, Wilson D, Taylor A, Ruffin R. Central obesity is associated with nonatopic but not atopic asthma in a representative population sample. *J Allergy Clin Immunol*. 2006 Dec;118(6):1284–91.
171. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol*. 2020 Feb;145(2):537–49.
172. Wang Y, Bai C, Li K, Adler KB, Wang X. Role of airway epithelial cells in development of asthma and allergic rhinitis. *Respir Med*. 2008 Jul;102(7):949–55.
173. Golebski K, Hoepel W, Van Egmond D, De Groot EJ, Amatngalim GD, Beekman JM, et al. FcγRIII stimulation breaks the tolerance of human nasal epithelial cells to bacteria through cross-talk with TLR4. *Mucosal Immunol*. 2019 Mar;12(2):425–33.
174. Talukdar SN, McGregor B, Osan JK, Hur J, Mehedi M. Respiratory Syncytial Virus Infection Does Not Induce Epithelial-Mesenchymal Transition. *J Virol*. 2023 Jul 27;97(7):e0039423.
175. Minor DM, Proud D. Role of human rhinovirus in triggering human airway epithelial-mesenchymal transition. *Respir Res*. 2017 Dec;18(1):110.
176. Radzikowska U, Eljaszewicz A, Tan G, Stocker N, Heider A, Westermann P, et al. Rhinovirus-induced epithelial RIG-I inflammasome suppresses antiviral immunity and promotes inflammation in asthma and COVID-19. *Nat Commun*. 2023 Apr 22;14(1):2329.
177. Su X, Wu W, Zhu Z, Lin X, Zeng Y. The effects of epithelial-mesenchymal transitions in COPD induced by cigarette smoke: an update. *Respir Res*. 2022 Aug 31;23(1):225.
178. Han X, Gharahkhani P, Hamel AR, Ong JS, Rentería ME, Mehta P, et al. Large-scale multitrait genome-wide association analyses identify hundreds of glaucoma risk loci. *Nat Genet*. 2023 Jul;55(7):1116–25.
179. Sadoun E, Leventakou V, Casas M, Ahmed HF, Kogevinas M, Fthenou E. A birth cohort study in the Middle East: the Qatari birth cohort study (QBiC) phase I. *BMC Public Health*. 2017 Dec;17(1):836.
180. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med Off J Am Coll Med Genet*. 2013 Oct;15(10):761–71.
181. Boutin NT, Schecter SB, Perez EF, Tchamitchian NS, Cerretani XR, Gainer VS, et al. The

- Evolution of a Large Biobank at Mass General Brigham. *J Pers Med*. 2022 Aug 17;12(8):1323.
182. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019 Mar;51(3):481–93.
183. Abdellaoui A, Dolan CV, Verweij KJH, Nivard MG. Gene–environment correlations across geographic regions affect genome-wide association studies. *Nat Genet*. 2022 Sep;54(9):1345–54.

Appendix

- IV. **Kasper Fischer-Rasmussen**, Raquel Granell, Anders Ulrik Eliassen, Eskil Kreiner, Casper-Emil Tingskov Pedersen, Yang Luo, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, David Strachan, Ashish Kumar, Anne-Marie Nybo Andersen, Bjarke Feenstra, Frank Geller, Valérie Siroux, Florence Demenais, Emmanuelle Bouzigon, Vincent Jaddoe, Ralf van der Valk, Liesbeth Duijts, Jordi Sunyer, Mónica Guxens, Marcella Marinelli, Mariona Bustamante, Joachim Heinrich, Marie Standl, John Curtin, Chih-Mei Chen, Wenche Nystad, Bo Jacobsson, Ronny Myhre, Craig E Pennell, Denise Daley, Carole Ober, Dorret Boomsma, Abdel Abdellaoui, Angela Simpson, Clare Murray, John W Holloway, Sam Collins, Steve Turner, S Hasan Arshad, Anhar Ullah, Erik Melén, John Henderson[†], Hans Bisgaard[†], Anders Gorm Pedersen, Adnan Custovic, Judith M. Vonk, Gerard H. Koppelman, Michael Kabesch, Klaus Bønnelykke, GABRIEL consortium, the EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium
Genetic Characterization of Preschool Wheeze Phenotypes
Submitted
- V. **Kasper Fischer-Rasmussen**, Casper-Emil Tingskov Pedersen, Luo Yang, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, Per Magnus, Yunsung Lee, Marianna Niemi, Tarja Laitinen, Þórunn Olafsdóttir, Guðmar Þorleifsson, Ingileif Jonsdóttir, Christian Erikstrup, Lotte Aas Hindhede, Carole Ober, Zhaozhong Zhu, Carlos A. Camargo, Anders Børghlum, Anders Gorm Pedersen, Anders Ulrik Eliassen & Klaus Bønnelykke
Genome-Wide Association Study of Severe Preschool Wheeze Uncovers Novel Insights into the Pathobiology
Manuscript in preparation
- VI. **Kasper Fischer-Rasmussen**, Casper-Emil Tingskov Pedersen, Astrid Sevelsted, Luo Yang, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, Anders Børghlum, Thomas Munk Laursen, Bjarni Vilhjálmson, Anders Gorm Pedersen, Anders Ulrik Eliassen & Klaus Bønnelykke
Early Childhood Trajectories of Inhaled Corticosteroid Prescriptions Implicate the Presence of Wheeze Endotypes
Manuscript in preparation

Fischer-Rasmussen, K et al.

Genetic Characterization of Preschool Wheeze Phenotypes

Genetic characterization of preschool wheeze phenotypes

Kasper Fischer-Rasmussen¹, Raquel Granell³, Anders Ulrik Eliassen, PhD¹, Eskil Kreiner, MD, PhD¹, Casper-Emil Tingskov Pedersen, PhD¹, Yang Luo¹, Bo Chawes, MD, DMSc¹, Jakob Stokholm, MD, PhD¹, Ann-Marie Malby Schoos, MD, PhD¹, David Strachan⁴, Ashish Kumar^{5,6}, Anne-Marie Nybo Andersen⁷, Bjarke Feenstra⁷, Frank Geller⁷, Valérie Siroux⁸, Florence Demenais⁸, Emmanuelle Bouzigon⁸, Vincent Jaddoe^{9,10}, Ralf van der Valk^{9,10}, Liesbeth Duijts^{9,10,11}, Jordi Sunyer^{12,13,14}, Mónica Guxens^{12,13,14,15}, Marcella Marinelli^{13,18}, Mariona Bustamante^{12,14}, Joachim Heinrich^{19,20}, Marie Standl²¹, John Curtin^{23,24}, Chih-Mei Chen²⁵, Wenche Nystad²⁶, Bo Jacobsson^{27,28}, Ronny Myhre²⁷, Craig E Pennell^{32,33,34,35}, Denise Daley³⁶, Carole Ober³⁷, Dorret Boomsma³⁸, Abdel Abdellaoui³⁸, Angela Simpson, Clare Murray, John W Holloway^{39,40}, Sam Collins^{39,41}, Steve Turner^{42,43}, S Hasan Arshad^{39,41,44}, Anhar Ullah⁴⁵, Erik Melén^{5,6}, John Henderson^{3†}, Hans Bisgaard^{1†}, Anders Gorm Pedersen, MSc, PhD², Adnan Custovic²², Judith M. Vonk³⁰, Gerard H. Koppelman^{29,30}, Michael Kabesch²⁵, Klaus Bønnelykke, MD, PhD¹, GABRIEL consortium, the EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium

¹COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark,

²Department of Health Technology, Section for Bioinformatics, Technical University of Denmark, Kongens Lyngby, Denmark

³MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

⁴Population Health Research Institute, St George's, University of London, Cranmer Terrace, London, SW17 0RE, United Kingdom

⁵Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

⁶Sach's Children and Youth Hospital, Södersjukhuset, Stockholm, Sweden

⁷Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

⁸Université Paris Cité, INSERM, UMR 1124, Group of Genomic Epidemiology of Multifactorial Diseases, 75006 Paris, France

⁹The Generation R Study Group, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

¹⁰Department of Pediatrics, division of Respiratory Medicine and Allergology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

¹¹Department of Neonatal and Pediatric Intensive Care, division of Neonatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

¹²ISGlobal, 08036 Barcelona, Spain

¹³Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain

¹⁴Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain

¹⁵Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands

¹⁶CIBER Epidemiología Y Salud Pública, Madrid, Spain

¹⁷Center for Research in Environmental Epidemiology (CREAL), Barcelona, Catalonia, Spain

¹⁸Hospital del Mar Research Institute (IMIM), Barcelona, Catalonia, Spain

¹⁹LMU– Ludwig-Maximilians-Universität Munich, Div. Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, University of Munich Medical Center, Munich, Germany

²⁰Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Munich, Germany

²¹Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764, Neuherberg, Germany

²²Pediatric allergy, Imperial College London, London, UK

²³Division of Immunology, Immunity to Infection and Respiratory Medicine, Manchester Academic Health Science Centre, School of Biological Sciences, University of Manchester, UK

²⁴NIHR Manchester Biomedical Research Centre, Manchester University Hospital NHS Foundation Trust, Manchester, UK

²⁵Department of Pediatric Pneumology and Allergy, St. Hedwig's Hospital of the order of St. John, University Children's Hospital Regensburg (KUNO), Steinmetzstr. 1-3, 93049 Regensburg, Germany

- ²⁶Division of Epidemiology, The Norwegian Institute of Public Health, Oslo, Norway.
- ²⁷Department of Genes and Environment, Norwegian Institute of Public Health, N-0403 Oslo, Norway
- ²⁸Department of Obstetrics and Gynecology, Sahlgrenska Academy, Gothenburg, Sweden
- ²⁹Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- ³⁰Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- ³¹Observational and Pragmatic Research Institute, Singapore, Singapore
- ³²Hunter Medical Research Institute, New Lambton Heights, NSW 2305, Australia
- ³³School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, NSW 2308, Australia
- ³⁴Mothers and Babies Research Centre, New Lambton Heights, NSW 2305, Australia
- ³⁵John Hunter Hospital, New Lambton Heights, NSW 2305, Australia
- ³⁶Center for Heart and Lung Innovation, University of British Columbia, Vancouver, British Columbia, Canada
- ³⁷Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada
- ³⁸Department of Human Genetics, University of Chicago, Chicago, IL, USA
- ³⁹Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

⁴⁰NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK

⁴¹Human Development and Health, University of Southampton Faculty of Medicine, Southampton, UK

⁴²Clinical and Experimental Sciences, University of Southampton Faculty of Medicine, Southampton, UK

⁴³Child Health, University of Aberdeen, Aberdeen, AB25 2ZG, UK

⁴⁴Women and Children Division, NHS Grampian, Aberdeen, AB25 2ZG, UK

⁴⁵National Heart and Lung Institute, Imperial College London, UK

⁴⁶Division of Immunology, Immunity to Infection and Respiratory Medicine at the University of Manchester, School of Biological Sciences, The University of Manchester, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, UK.

Corresponding author

Correspondence and requests for reprints should be addressed to Klaus Bønnelykke, M.D., Ph.D., Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital and University of Copenhagen, Ledreborg Alle 34, 2820 Gentofte, Denmark. E-mail: kb@copsac.com.

Funding

COPSAC (Copenhagen Prospective Studies on Asthma in Childhood) is funded by private and public research funds, which are all listed on www.copsac.com. The Lundbeck Foundation, Danish State Budget, Danish Council for Strategic Research, Danish Council for Independent Research, and The Capital Region Research Foundation have provided core support for COPSAC. The study is further supported by the following NIH grants: R01 HL129735 and R01 HL141826. C.E.W. was supported by the Swedish Heart–Lung Foundation (Hjärt–Lungfonden 20170734; 20180290). This project has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 946228).

Author Contributions: The guarantor of the study is K.B., who is responsible for the integrity of the work as a whole, from conception and design of the study to conduct of the study and the acquisition of data, the analysis and interpretation of data, and the writing of the manuscript. K.B. had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. K.F.R., R.G and E.K. were responsible for data analysis and K.F.R wrote the first draft of the manuscript. K.F.R., R.G., A.U.E, E.K., C.E.T., Y.L., B.C., J.S. and A.M.S., contributed to the design of the study, interpretation of the data, and writing of the manuscript. All co-authors have contributed substantially to the analyses and/or interpretation of the data and have provided important intellectual input and approval of the final version to be published. The lead author (K.B.) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Data sharing statement: The COPSAC (Copenhagen Prospective Studies on Asthma in Childhood) biobank is publicly available at the Danish National Biobank (www.biobankdenmark.dk), and data will become available in the Danish Data Archive (www.sa.dk) upon request to the corresponding author.

Disclosure

During the preparation of this work the KFR used ChatGPT 4.5 in order to improve readability and language. After using this tool, KFR reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Abstract

Rationale

Preschool wheeze is a heterogeneous and poorly understood clinical syndrome. As a result, current treatments are insufficient and prevention is not possible.

Objective

To increase understanding of the genetic susceptibility and underlying disease mechanisms of wheeze phenotypes in early childhood through large-scale genome-wide association study (GWAS) analyses.

Methods

We performed meta-analyses of GWAS on preschool wheeze phenotypes, including early transient and persistent wheeze, defined by asthma/wheeze at age 3 and subsequent remission or persistent at age 6 respectively. The discovery analyses included data on more than 13,000 children from 15 cohorts. Suggestive findings were followed-up for replication through meta-analyses of data from 7 additional cohorts including up to 5,000 children. Additionally, we utilized genetic variants associated with asthma-related traits in adulthood (adult asthma, atopy, eosinophils and lung function) to quantify the degree to which genetic risk influencing asthma-related adult traits also influences genetic risk of preschool wheeze.

Measurements & Main Results

Variants near the *GSDMB* gene in the 17q-region showed genome-wide significant association with persistent (rs11078926, OR = 1.43 (1.30 - 1.578), $P = 2.14 \times 10^{-11}$), but not with early transient wheeze (rs1054609, OR = 1.08 (0.98 - 1.18), $P = 0.094$). Additionally, increased genetic risk to persistent wheeze was associated with genetic

risk for asthma at older ages, atopy, eosinophil count and lower adult lung function. In contrast, genetic risk to transient early wheeze was only associated with low lung function.

Conclusions

The genetic susceptibility to preschool wheeze exhibited significant enrichment within genetic risk loci associated with asthma, atopic traits, and diminished lung function in adulthood. Furthermore, the observed enrichment displayed distinct patterns for persistent and transient phenotypes of wheezing. These results indicate distinct etiologies of wheeze phenotypes, which could inform studies in optimization of prevention and treatment strategies.

Introduction

Asthma and asthma-like symptoms in preschool age, often termed preschool wheeze, is a heterogenous and poorly understood clinical syndrome with both genetic and environmental etiologies. It is common in childhood as the most frequent cause of acute hospitalizations in early life, and is associated with a substantial burden on quality of life for affected children and families and socioeconomic costs for society (1–4). The incomplete understanding of recurrent preschool wheeze is probably one main reason for the current insufficiencies in treatment and prevention of the disorder (3). Wheeze is often used to describe symptoms such as whistling in the chest while breathing and can arise from various conditions such as bronchitis or bronchiolitis, triggered by lower respiratory tract infections, especially rhinovirus (5). Furthermore, recurrent wheezing in preschool children is a strong risk factor for asthma at school age (6).

Evidence from twin studies estimate the heritability to be as high as 80% for childhood asthma and 60 % for ever having wheezed during childhood (7,8). While multiple susceptibility loci have been discovered in genome-wide association studies (GWAS) on asthma (9–14), very little is known regarding genetic determinants of preschool wheezing (38284918). GWAS using specific phenotype definitions may have improved power for discovery of susceptibility variants and thereby elucidate underlying mechanisms (15). This is exemplified by previous studies of severe exacerbations in preschool children identifying *CDHR3* as a novel susceptibility gene (16), and of longitudinal wheeze phenotypes from birth to 18 years identifying *ANXA1* as associated with persistent wheeze (17); both of which were not discovered in larger GWAS using less specific definitions of asthma.

In this study, we performed a GWAS using data from the EAGLE and GABRIEL consortia (18,19), in order to create the largest study of temporal wheezing phenotypes in early life comprising more than 18,000 children from 22 studies. We examined early-onset wheeze (EOW), defined as recurrent wheeze or asthma in the first 3 years of life, as well as subtypes of EOW in terms of transient-early wheeze (TEW), with symptoms resolving before age 6, and persistent wheeze (PW), with symptoms persisting to age 6. We also investigated current asthma by age 6 (CA), defined as cross-sectional asthma at age 6 irrespective of age at onset. Due to a small number of children with late-onset wheezing across the studies, a specific GWAS on this phenotype was not performed. Our objective was to characterize genetic differences and similarities among the four phenotypes of preschool wheeze to illuminate the underlying molecular mechanisms.

Methods

Preschool wheeze phenotypes

Preschool wheeze was defined as three or more wheeze episodes or an asthma diagnosis before the age of 3. Transient early wheeze (TEW) was defined as preschool wheezers that were symptom-free between age 5 and 7 years. Persistent wheeze included children with preschool wheeze and one wheezing episode or an asthma diagnosis between ages 5 and 7 years. Early-onset wheeze (EOW) consisted of both TEW and PW. Current asthma (CA) was defined with at least one wheezy episode or asthma diagnosis between ages 5 and 7 years, irrespective of age at onset, thus including both PW and children with late onset of wheezing or asthma (after 3 years). Late onset wheeze or asthma was not analyzed separately in this study due to insufficient number of cases (see Supplementary Table 1 & 3). All children not fulfilling case criteria were used as controls for any subsequent phenotype.

Discovery Analysis

A total of more than 13,000 individuals of European descent were recruited from 15 separate research initiatives all participating in the GABRIEL and EAGLE consortia (18,19). Ethical approval was granted by local research ethics committees for each individual study, and participants or child's legal guardians provided informed consent, where applicable. Genotyping and subsequent imputation was conducted within each study. Association analysis was executed using logistic regression models,

implementing an additive model for single nucleotide polymorphisms (SNPs). Variants were excluded in the meta-analysis if they occurred in less than 3 studies, had a minor allele frequency below 5 % or did not occur in the 1000 genomes reference panel (20) (see Supplementary Table 2 & Supplementary Table 4). Fixed-effects inverse-variance weighted meta-analysis was performed utilizing the software METAL (version released 2020-05-05) (21). Functionally-Informed Z-score Imputation (FIZI) was used to impute summary results to increase genotype density. Z-scores were converted to logOR and standard errors using equations from Zhu et al. 2016 (22) (see Supplementary Material & Supplementary Table 4).

Replication Analysis

Variants were carried forward to the replication phase based on two criteria. 1) the p-value should be less than $5e-5$ and 2) the variant is in a locus not previously known to be associated with asthma. Due to the attenuated signal at the standard genome-wide significance cutoff ($5e-8$), we utilized a more liberal p-value cutoff ($5e-5$) in the discovery analysis, allowing ca. 5 % false positives among the selected variants (23). Replication was sought in more than 5000 individuals of european descent recruited from 7 different studies. Informed consent was gathered from the parents in each individual replication study. The replication studies carried out their own genotyping and quality control (see Supplementary Table 2 & Supplementary Table 5).

Analysis of Known Asthma Variants

To investigate if known asthma variants were associated with the wheezing phenotypes, we gathered lead variants from previous GWAS studies on asthma, including phenotypes of asthma ever (24–26) adult asthma, childhood asthma (13,14) and early childhood asthma with severe exacerbations (9,12). This resulted in 138 independent variants ($r^2 < 0.05$) available for association with EOW, TEW, PW and CA respectively. Variants showing similar direction of effect and a false discovery rate below 5 %, were considered associated with preschool wheeze. Correction for multiple testing was performed using the Benjamini-hochberg method.

Genetic Similarity Between Adult Traits and Preschool wheeze

Obtaining reliable estimates of genetic correlation (shared heritability) using LD-score regression (27) was not feasible in this study. Instead, we leveraged MR, to approximate genetic correlation since the wheeze phenotypes displayed weak heritability (see Supplementary Table 15). Previous studies have shown that methods for Mendelian Randomization (MR) is confounded by genetic correlation and MR and genetic correlation answer similar questions (28). To characterize genetic similarity of the phenotypes with related traits, we instead utilized Generalised Summary-data-based Mendelian Randomisation (GSMR) (29) a tool originally developed for Mendelian Randomization analysis to compare association coefficients of alleles between traits from published GWAS and the GWAS of preschool wheeze. GSMR was selected for its ability to accurately account for sampling variance in both the exposure and outcome GWAS, which was particularly important given the limited power the wheeze GWAS. We omitted the HEIDI-outlier analysis, as our study did not aim to establish causal inference

(see Supplementary Code). The related traits were included based on published large-scale GWAS studies on childhood onset asthma, adult onset asthma (14), asthma ever (26) severe exacerbations in early childhood (9,12), allergic sensitization (30), atopic dermatitis (31), eosinophil count (32) and lung function measurements (33). For detailed description of phenotype definitions of included traits (see Supplementary Material). Related traits were included as exposures and the wheeze phenotypes as outcomes and the genetic overlap can be interpreted as a logOR for wheeze status (29). Correction for multiple testing was performed using the Benjamini-hochberg method.

Lung Development Genes

To investigate if the wheezing phenotypes were associated with genetic risk factors involved in lung development, we aimed to characterize variants located near genes previously known to be involved in lung development. The list of genes was compiled as described in Portas et al. 2020 (34) and we mapped SNPs to genes based on proximity using BEDTools (35).

Results

Discovery analyses

The only locus that reached genome-wide significance in the GWAS analyses was the widely known 17q12-21 locus (17q). Variants at this locus were associated with three out of the four phenotypes, namely EOW, PW and CA (see Table 1). We did not find any genome-wide significant loci for TEW (see Supplementary Figure 1). The lead variant in the 17q region for PW (rs11078926, OR = 1.43 (1.30 - 1.578), $P = 2.14E-11$), EOW (rs2305480, OR = 1.26 (1.17 - 1.33), $P = 2.30E-16$) and CA (rs1008723, OR = 1.38 (1.30 - 1.52), $P = 1.75E-17$), where all strongly correlated (minimum $r^2 = 0.82$) (see Supplementary Table 8). Association results outside 17q are shown in Supplementary Figures 2-3. All suggestive SNPs within a 1Mb window with $r^2 < 0.01$, are shown in Supplementary Table 11.

Replication analysis

Ten suggestive SNPs ($P < 5e-5$) in the discovery phase were unknown asthma loci, and were carried forward to the replication phase (see Supplementary Table 2). None of these SNPs showed evidence of replication in independent cohorts. One variant, rs2214823 near *HGF* with suggestive association to TEW, showed a trend towards significance in the replication phase (OR = 1.11, (0.99-1.26) $P = 0.071$) (see Supplementary Table 5).

Analysis of known asthma variants

In total, 138 independent SNPs were used to characterize genetic similarity of the wheezing phenotypes to the known genetics of asthma. In total 29 out of 138 SNPs were nominally associated ($P < 0.05$) with the same direction of effect, such that the risk allele for asthma was also associated with a higher risk of wheeze. Eight SNPs passing nominal significance showed opposite direction of effect (see Supplementary Table 12).

Most of the significant associations with the same direction of effect were seen for EOW, PW and CA (see Figure 1). Three loci passed correction for multiple testing ($FDR < 0.05$) with directional consistency in EOW, namely *GSDMB*, *CDHR3* and *IL33*. Only *GSDMB* passed FDR in PW. Lastly, 5 loci passed FDR in CA, specifically *ATXN2*, *GSDMB*, *IL33*, *LBH* & *WDR36*. For TEW, limited directional replication was present, and no loci passed FDR. *CDHR3* was the only locus with a positive association with TEW (rs6967330-A, OR=1.12, $P=0.044$), that had previously been associated with preschool asthma (9).

Genetic Similarity Between Adult Traits and Preschool wheeze

EOW significantly overlapped with all of the asthma related traits, including asthma at older ages, atopic traits, and lung function. A similar pattern of association was seen for PW with significant overlap with 7 out of 10 traits, and CA with overlap for 9 out of 10 traits. TEW showed a distinct association pattern compared to the other phenotypes and only showed significant overlap with peak expiratory flow (PEF) (OR = 0.64, (CI = 0.48-0.86), $P = 2.73e-3$). FEV1/FVC (OR = 0.79, (CI = 0.64-0.98), $P = 0.032$) and

asthma exacerbations (OR = 1.12, (CI = 1.004-1.24), P = 0.041) showed nominally significant overlap with TEW (see Figure 2). Results are shown in Supplementary Table 12 and scatter plots of SNP effect sizes between traits and preschool wheeze are shown in Supplementary Figures 5-14.

Lung Development genes

When only considering variants in genes associated with lung development, the number of statistical tests was reduced from ~3.1M to ~10K. The variant rs2214823 near the *HGF* gene, was significantly associated with TEW after correcting for multiple testing (see Supplementary Figure 3). No other signals were detected across the four phenotypes. The TEW-risk allele of rs2214823 was associated with a lower PEF in adults (P = 6.86e-03), but not FEV1 or FEV1/FVC [34] (see Supplementary Table 8).

Discussion

Wheezing in preschool age, and particularly in the first 3 years of life, is a common clinical syndrome that is still poorly understood. It clearly represents heterogeneous mechanisms, as also indicated by the different temporal subtypes with a large proportion of children experiencing remission before age 6 (6,36). Diagnosing and subtyping wheezing illnesses in the first 3 years of life is challenging both clinically and in a research context. The mechanistic understanding of this process remains a black box due to limited lung function assessment options and the absence of markers of airway inflammation in this age group. For this reason, many clinicians and researchers are hesitant to diagnose asthma before age 3.

Our GWAS on early onset wheeze or asthma (with onset before 3 years of age) showed genetic similarity with previous GWAS on childhood onset asthma (before 12 yrs), adult onset asthma (onset between 20 and 60 yrs), and severe exacerbations in age 2-6 years. It also showed genetic similarity with the atopic traits characterizing later asthma, including blood eosinophil count, allergic sensitization and atopic dermatitis, as well as low lung function was also associated with EOW. These findings are consistent with the shared genetic architecture and age at onset characteristics that have previously been described for these traits and asthma at older ages (37,38). This indicates that the mechanisms involved in later asthma are also involved in wheeze in the first 3 years of life. SNPs showing the strongest associations with early onset wheeze were generally SNPs that were also among the strongest loci in GWAS of asthma in later childhood (6-12 years), including *GSDMB*, *IL33*, *ATXN2*, *LBH* & *WDR36*. One exception is the

finding that *CDHR3* was the second strongest candidate SNP for early onset wheeze, while it has not been among top SNPs in GWAS studies of asthma in school age or adulthood. *CDHR3* was discovered in a GWAS of severe exacerbations at age 2-6 years and functions as a rhinovirus C receptor (39) and the known role of viral infections in preschool wheeze. This indicates that *CDHR3* plays a particular role for preschool asthma, which is also supported by the lack of association with current asthma by age 6 in the current study.

Our two subtypes of early onset wheeze, PEW and TEW, differed in their pattern of genetic associations. PEW showed similarity in genetic risk with asthma, atopic traits and low lung function, and also showed associations to previously reported asthma genes. This genetic association pattern also resembled the pattern seen for our phenotype of current asthma by age 6. In contrast, TEW was mainly influenced by risk of low PEF and possibly also FEV1/FVC and preschool asthma exacerbations. The associations with previously reported asthma genes were also markedly different. There were no FDR significant loci with consistent direction of effect, and *CDHR3* was the only locus with nominal significance shared with the other phenotypes. Also, based on the qq-plot, the overall genetic signal for transient early wheeze seemed weaker than for the other phenotypes.

The Tucson Children's Respiratory Study (N=125) found that TEW was associated with lower lung function shortly after birth, while PW had normal lung function at birth and lower lung function at age 6 (6). In contrast, the Perth birth cohort study (N=243) (40) reported reduced infant lung function in children with PW but not TEW, while

COPSAC2000 with higher statistical power (N=403) found that persistent wheeze was associated with reduced lung function already at birth, with further reduction to age 7 years, and TEW showed a trend of reduced lung function at birth (41). The large ALSPAC study (N=6265) reported reduced lung function at age 7 for both TEW and PW (36). Regarding atopic traits, most studies have shown that PW, but not TEW, is associated with atopic traits such as allergic sensitization (6,36). The literature thus suggests that wheeze restricted to the first 3 years of life is associated with reduced lung development and early lung function decline, while wheeze persisting after 3 years is more atopic in nature, but that both etiologies involve reduced lung function by the age of six, possibly through different mechanisms. Our analysis of shared genetic influence supports these findings by showing that genetic risk for asthma exacerbations, low PEF and FEV1/FVC are associated with TEW, while PW and CA were consistent with genetic risk for asthma at older ages, atopy and lung function. In line with our results, a study from the PIAMA cohort showed that TEW was associated with known COPD genes (42). Overall, this suggests that TEW is distinct from childhood asthma and more likely involves impaired lung development and as a result impaired respiratory function.

We did not identify novel susceptibility loci at the genome-wide association level in our study. A recent GWAS study of temporal childhood wheeze phenotypes was performed in 9658 subjects from the STELAR & UNICORN cohorts. This study utilized LCA-derived temporal phenotypes based on 5 time-points to define wheeze trajectories from birth to 18 years and identified one novel genome-wide significant variant near *ANXA1* associated with wheeze persisting from birth until adolescence (17). The

identified *ANXA1* variant was not present in our study and we had no strong tagging variant, potentially due to the low allele frequency (~2%). We were therefore unable to test if this variant replicated in our study.

In the replication phase, a single variant rs2214823 near *HGF*, showed a trend towards replication. The *HGF* variant also showed significant association to transient early wheeze after adjustment for multiple testing when considering only genes involved in fetal lung development. No associations were found between the *HGF* variant and EOW, PW or CA. Additionally, we found that the variant was associated with PEF in UK Biobank, indicating a possible role in adult lung function. Even though replication failed for rs2214823, these results suggest that TEW may be predisposed to wheezing in early childhood due to underdeveloped airways, a view shared by a previously mentioned study from Martinez and colleagues (6). HGF has been shown to play a role in models of pulmonary fibrosis in inducing tissue repair and preventing fibrotic remodeling (43). There is substantial evidence that pulmonary fibrosis involves suggesting an aberrant re-activation of embryonic and developmental pathways (44), and thus it is plausible that HGF may play a role in lung development. Furthermore, HGF is expressed in fetal lung fibroblasts, myofibroblasts and smooth muscle cells (45).

We found that the 17q12-21 locus was strongly associated with PW, EOW and CA but not with TEW, despite the relatively large sample size in the discovery phase. In contrast to our findings, a study combining multi-ethnic US birth cohorts (CREW) found associations between multiple 17q variants and LCA-derived transient wheeze from 0 to 11 years (46). In the present study, we used a different approach to define transient

wheeze based on children who wheezed more than 3 times or were diagnosed with asthma before age 3. Simultaneously, they were completely free of wheeze and asthma at age 6. This contrasts the definition in the CREW-study in which ~20 % wheezed at age 6, meaning that their transient phenotype could potentially overlap with our persistent phenotype. In line with our findings, a study combining UK birth cohorts (STELAR) found no association between transient wheeze and 17q in 7719 children (47). It should be noted that some cohorts in the two studies described above occur in the discovery phase of this study.

CDHR3 was the only asthma candidate locus in our study that was associated with both persistent (OR=1.17, $P=6.2e-3$ and transient wheeze (OR=1.12, $P=0.044$). The above-mentioned study from STELAR found that *CDHR3* was associated with persistent and intermittent wheeze, but not with transient and late-onset wheeze. This contradicts our finding, but our definition of TEW differed slightly from that of STELAR in that we require at least 3 episodes of wheeze before the age of three, in order to classify as TEW, contrary to a single episode in STELAR. The higher symptom burden of TEW in this study may explain the different results.

This study represents the largest GWAS of preschool wheeze phenotypes to date and thereby increased power for identification of susceptibility loci and understanding the genetic background of preschool wheeze. Another strength is the inclusion of a large number of well characterized birth cohorts with wheeze phenotypes characterized longitudinally, thereby minimizing the risk of recall bias.

It is a limitation of our study that we did not include a late onset wheeze phenotype. This decision was taken because of the expected low numbers of children with this phenotype and therefore limited statistical power. Also, in spite of being the largest GWAS on preschool wheeze to date, the sample sizes were still relatively small compared to previous GWAS using other definitions of asthma, which is also reflected in the paucity of genome-wide significant associations. Given limited statistical power in our discovery analyses, we were not able to estimate genetic correlation with other traits using the standard approach of LD-score regression. LD score regression is more robust than our approach since LD-structure is taken into account (27). In contrast to the previous GWAS study of longitudinal wheeze (17), we did not have longitudinal information on wheezing beyond the age of 6 years of age.

Conclusions

We found that recurrent wheeze or asthma in the first 3 years of life showed genetic similarity with asthma at older ages, atopy, eosinophils, and low lung function. This was driven by children with symptoms persisting to age 6. In contrast, children with symptoms that remitted before age 6 did not display genetic similarity to asthma at older ages or atopic traits, but only to genetics of reduced lung function in adulthood and risk of early-life asthma exacerbations. Also, heritability seemed weaker for this phenotype. These results suggest that transient early wheeze is influenced by other molecular mechanisms than persistent phenotypes and that environmental factors play a larger role for early transient wheeze. Future studies should address if identification of distinct genetic mechanisms in children with early onset preschool wheeze can help improve prediction of the disease course and treatment of the disease.

References

1. Lozano P, Sullivan SD, Smith DH, Weiss KB. The economic burden of asthma in US children: Estimates from the National Medical Expenditure Survey. *J Allergy Clin Immunol*. 1999 Nov;104(5):957–63.
2. Kocevar VS, Bisgaard H, Jönsson L, Valovirta E, Kristensen F, Yin DD, et al. Variations in pediatric asthma hospitalization rates and costs between and within Nordic countries. *Chest*. 2004 May;125(5):1680–4.
3. Bush A. Practice imperfect--treatment for wheezing in preschoolers. *N Engl J Med*. 2009 Jan 22;360(4):409–10.
4. Pardue Jones B, Fleming GM, Otililio JK, Asokan I, Arnold DH. Pediatric acute asthma exacerbations: Evaluation and management from emergency department to intensive care unit. *J Asthma*. 2016 Jul 2;53(6):607–17.
5. Lemanske Jr. RF, Jackson DJ, Gangnon RE, Evans MD, Li Z, Shult PA, et al. Rhinovirus illnesses during infancy predict subsequent childhood wheezing. *J Allergy Clin Immunol*. 2005 Sep;116(3):571–7.
6. 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.
7. Ullemar V, Magnusson PKE, Lundholm C, Zettergren A, Melén E, Lichtenstein P, et al. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy*. 2016 Feb;71(2):230–8.
8. van Beijsterveldt CEM, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J*. 2007 Mar 1;29(3):516–21.
9. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet*. 2014 Jan;46(1):51–5.
10. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet*. 2018 Jan;50(1):42.
11. Han Y, Jia Q, Jahani PS, Hurrell BP, Pan C, Huang P, et al. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat Commun*. 2020 Dec;11(1):1776.
12. Ahluwalia TS, Eliassen AU, Sevelsted A, Pedersen CET, Stokholm J, Chawes B, et al. FUT2–ABO epistasis increases the risk of early childhood asthma and Streptococcus pneumoniae respiratory illnesses. *Nat Commun*. 2020 Dec 16;11:6398.
13. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019 Jun 1;7(6):509–22.
14. Ferreira MAR, Mathur R, Vonk JM, Sz wajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am J Hum Genet*. 2019 Apr 4;104(4):665–84.
15. Wolters AAB, Kersten ETG, Koppelman GH. Genetics of preschool wheeze and its progression to childhood asthma. *Pediatr Allergy Immunol*. 2024 Jan;35(1):e14067.
16. Carlsson CJ, Vissing NH, Sevelsted A, Johnston SL, Bønnelykke K, Bisgaard H. Duration of wheezy episodes in early childhood is independent of the microbial trigger. *J Allergy Clin Immunol*. 2015;

17. Granell R, Curtin JA, Haider S, Kitaba NT, Mathie SA, Gregory LG, et al. A meta-analysis of genome-wide association studies of childhood wheezing phenotypes identifies ANXA1 as a susceptibility locus for persistent wheezing. *eLife*. 2023 May 25;12:e84315.
18. EARly Genetics Lifecourse Epidemiology (EAGLE) consortium, Early Growth Genetics (EGG) consortium, Middeldorp CM, Felix JF, Mahajan A, McCarthy MI. The Early Growth Genetics (EGG) and EARly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. *Eur J Epidemiol*. 2019 Mar;34(3):279–300.
19. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010 Sep 23;363(13):1211–21.
20. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015 Oct 1;526(7571):68–74.
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma Oxf Engl*. 2010 Sep 1;26(17):2190–1.
22. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016 May;48(5):481–7.
23. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 1995 Nov;11(3):241–7.
24. Han Y, Jia Q, Jahani PS, Hurrell BP, Pan C, Huang P, et al. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat Commun*. 2020 Apr 15;11(1):1776.
25. Olafsdottir TA, Theodors F, Bjarnadottir K, Bjornsdottir US, Agustsdottir AB, Stefansson OA, et al. Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis. *Nat Commun*. 2020 Dec;11(1):393.
26. Tsuo K, Zhou W, Wang Y, Kanai M, Namba S, Gupta R, et al. Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. *Cell Genomics*. 2022 Dec;2(12):100212.
27. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015 Mar;47(3):291–5.
28. Kraft P, Chen H, Lindström S. The Use of Genetic Correlation and Mendelian Randomization Studies to Increase Our Understanding of Relationships between Complex Traits. *Curr Epidemiol Rep*. 2020 Jun;7(2):104–12.
29. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun*. 2018 Jan 15;9(1):224.
30. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat Genet*. 2018 Aug;50(8):1072.
31. Sliz E, Huilaja L, Pasanen A, Laisk T, Reimann E, Mägi R, et al. Uniting biobank resources reveals novel genetic pathways modulating susceptibility for atopic dermatitis. *J Allergy Clin Immunol*. 2022 Mar;149(3):1105–1112.e9.
32. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020 Sep 3;182(5):1214–1231.e11.
33. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019 Mar;51(3):481–93.
34. Portas L, Pereira M, Shaheen SO, Wyss AB, London SJ, Burney PGJ, et al. Lung

- Development Genes and Adult Lung Function. *Am J Respir Crit Care Med*. 2020 Sep 15;202(6):853–65.
35. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010 Mar 15;26(6):841–2.
 36. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax*. 2008 Nov 1;63:974–80.
 37. Zhu Z, Lee PH, Chaffin MD, Chung W, Loh PR, Lu Q, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat Genet*. 2018 Jun;50(6):857–64.
 38. Ferreira MAR, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Age-of-onset information helps identify 76 genetic variants associated with allergic disease. Bouzigon E, editor. *PLOS Genet*. 2020 Jun 30;16(6):e1008725.
 39. Bochkov YA, Watters K, Ashraf S, Griggs TF, Devries MK, Jackson DJ, et al. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc Natl Acad Sci U S A*. 2015 Apr 28;112(17):5485–90.
 40. Turner SW, Palmer LJ, Rye PJ, Gibson NA, Judge PK, Cox M, et al. The Relationship between Infant Airway Function, Childhood Airway Responsiveness, and Asthma. *Am J Respir Crit Care Med*. 2004 Apr 15;169(8):921–7.
 41. Bisgaard H, Jensen SM, Bønnelykke K. Interaction between asthma and lung function growth in early life. *Am J Respir Crit Care Med*. 2012 Jun 1;185(11):1183–9.
 42. Kerkhof M, Boezen HM, Granell R, Wijga AH, Brunekreef B, Smit HA, et al. Transient early wheeze and lung function in early childhood associated with chronic obstructive pulmonary disease genes. *J Allergy Clin Immunol*. 2014 Jan;133(1):68-76.e1-4.
 43. Panganiban RAM, Day RM. Hepatocyte growth factor in lung repair and pulmonary fibrosis. *Acta Pharmacol Sin*. 2011 Jan;32(1):12–20.
 44. Chanda D, Otoupalova E, Smith SR, Volckaert T, De Langhe SP, Thannickal VJ. Developmental pathways in the pathogenesis of lung fibrosis. *Mol Aspects Med*. 2019 Feb;65:56–69.
 45. He P, Lim K, Sun D, Pett JP, Jeng Q, Polanski K, et al. A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell*. 2022 Dec;185(25):4841-4860.e25.
 46. Hallmark B, Wegienka G, Havstad S, Billheimer D, Ownby D, Mendonca EA, et al. Chromosome 17q12-21 Variants Are Associated with Multiple Wheezing Phenotypes in Childhood. *Am J Respir Crit Care Med*. 2021 Apr 1;203(7):864–70.
 47. Haider S, Granell R, Curtin J, Fontanella S, Cucco A, Turner S, et al. Modeling Wheezing Spells Identifies Phenotypes with Different Outcomes and Genetic Associates. *Am J Respir Crit Care Med*. 2022 Apr 15;205(8):883–93.

Acknowledgements

John Henderson (1958-2019)

This article is dedicated to the memory of our wonderful colleague and friend Professor John Henderson, whose contribution to our work cannot be overstated.

Hans Bisgaard (1955-2022)

We would like to acknowledge the huge work of late professor Hans Bisgaard, who was the founder of COPSAC and was head of the clinical research center for more than 25 years. Hans was a dedicated, innovative physician-scientist who pushed the asthma research field forward. He contributed immensely to pediatric research through the COPSAC birth cohorts, as well as a vast amount of other clinical studies on childhood asthma. Hans's impressive work and ideas live on in the studies conducted in the birth cohort. Thank you for being a great inspiration to us all.

Figure legends

Figure 1. Left panel: Volcano plots of the four wheezing phenotypes, EOW, TEW, PW & CA (Top to bottom). The pink dashed indicates an FDR below 0.05 and the blue dashed line represents the nominal significance level ($P < 0.05$). Plots show if there is a trend in the risk effects of known asthma variants and the effect in wheeze, with log-odds for wheeze status on the x-axis and $-\log_{10}(\text{p-value})$ on the y-axis. Right panel: Variants with a nominally significant association in at least one phenotype. Cells are coloured according to log odds-ratio and stars indicates level of significance (+: $P < 0.05$, * FDR < 0.05).

Figure 2. Log odds-ratios increase for wheeze risk of shared genetic influence between adult traits and preschool wheeze. Association results of variant coefficients between the adult traits (y-axis) and preschool wheeze (boxes), including 95 % confidence intervals for the association result. Stars mark the significance of the association, (+: $P < 0.05$, * FDR < 0.05).

Tables

Table 1. Association of the 17q12 locus represented by rs7219923 across the wheezing phenotypes.

Phenotype	Nearest gene	SNP	EA/NE A	OR (95 % CI)	P-value	N
CA	GSDMB	rs7219923	T/C	1.335 (1.241 - 1.438)	1.33E-14	15,074
EOW				1.24 (1.173 - 1.311)	3.05E-14	17,383
PW				1.369 (1.237 - 1.516)	1.28E-09	13,878
TEW				1.063 (0.979 - 1.153)	0.1423	12,604

Table 2. Association results containing novel variants carried forward to the replication stage. Results from both phases are shown for the allele conferring risk in the discovery phase.

Phenotype	Chr	SNP	EA/NEA	Nearby gene	Discovery				Replication			
					OR	P	N	Phet	OR	P	N	Phet

					(95% CI)				(95% CI)			
TEW	7	rs2214823	A/G	HGF/ CACNA2D1	1.25 (1.15, 1.35)	2.32E-07	12,604	0.116	1.11 (0.99, 1.26)	0.071	7,323	0.066
EOW	18	rs1187725 4	A/G	DCC	1.16 (1.10, 1.23)	2.33E-07	17,383	0.933	0.95 (0.85, 1.04)	0.37	7,355	0.307
PW	6	rs6941858	G/C	LOC10013185 3/ ARG1	1.75 (1.41, 2.17)	3.48E-07	13,878	0.069	0.72 (0.38, 1.38)	0.33	6,642	0.325
CA	9	rs1255425 3	T/C	PRUNE2/ LOC10012879 0	1.35 (1.20, 1.51)	3.97E-07	15,074	0.749	1.03 (0.84, 1.26)	0.76	4,653	0.761
EOW	12	rs1281796 7	A/C	RPL31P10/ SRP14P1	1.15 (1.09, 1.22)	8.75E-07	17,383	0.012	1.00 (0.91, 1.10)	0.93	7,355	0.934
EOW	15	rs1896796	A/G	SH3GL3/ ADAMTSL3	1.15 (1.09, 1.22)	1.22E-06	17,383	0.993	1.01 (0.92, 1.11)	0.79	7,355	0.793
CA	10	rs1276710 8	A/C	SLC29A3	1.29 (1.17, 1.44)	1.36E-06	15,074	0.954	1.06 (0.89, 1.27)	0.5	4,653	0.504
PW	1	rs3752543	C/T	CAMTA1	1.37 (1.21, 1.56)	1.71E-06	13,878	0.102	1.00 (0.81, 1.23)	0.99	6,642	0.998
EOW	1	rs1208233 0	G/C	SLC35D1	1.30 (1.17, 1.45)	1.74E-06	17,383	0.479	0.97 (0.82, 4.46)	0.8	7,355	0.796

TEW	6	rs505000	C/T	SLC22A2/ SLC22A3	1.27 (1.15, 1.40)	3.29E-06	12,604	0.711	1.10 (0.93, 1.29)	0.25	7,323	0.098
-----	---	----------	-----	---------------------	-------------------------	----------	--------	-------	----------------------	------	-------	-------

Table 3. *HGF* association results across four phenotypes.

Phenotype	Nearest gene	SNP	EA/N EA	OR (95 % CI)	P-value	N
TEW	HGF	rs2214823	A/G	1.251 (1.149 - 1.362)	2.32E-07	12,604
EOW				1.093 (1.032 - 1.159)	0.0026	17,383
PW				0.977 (0.881 - 1.085)	0.6661	13,878
CA				0.991 (0.919 - 1.069)	0.8223	15,074

Figures

Figure 1

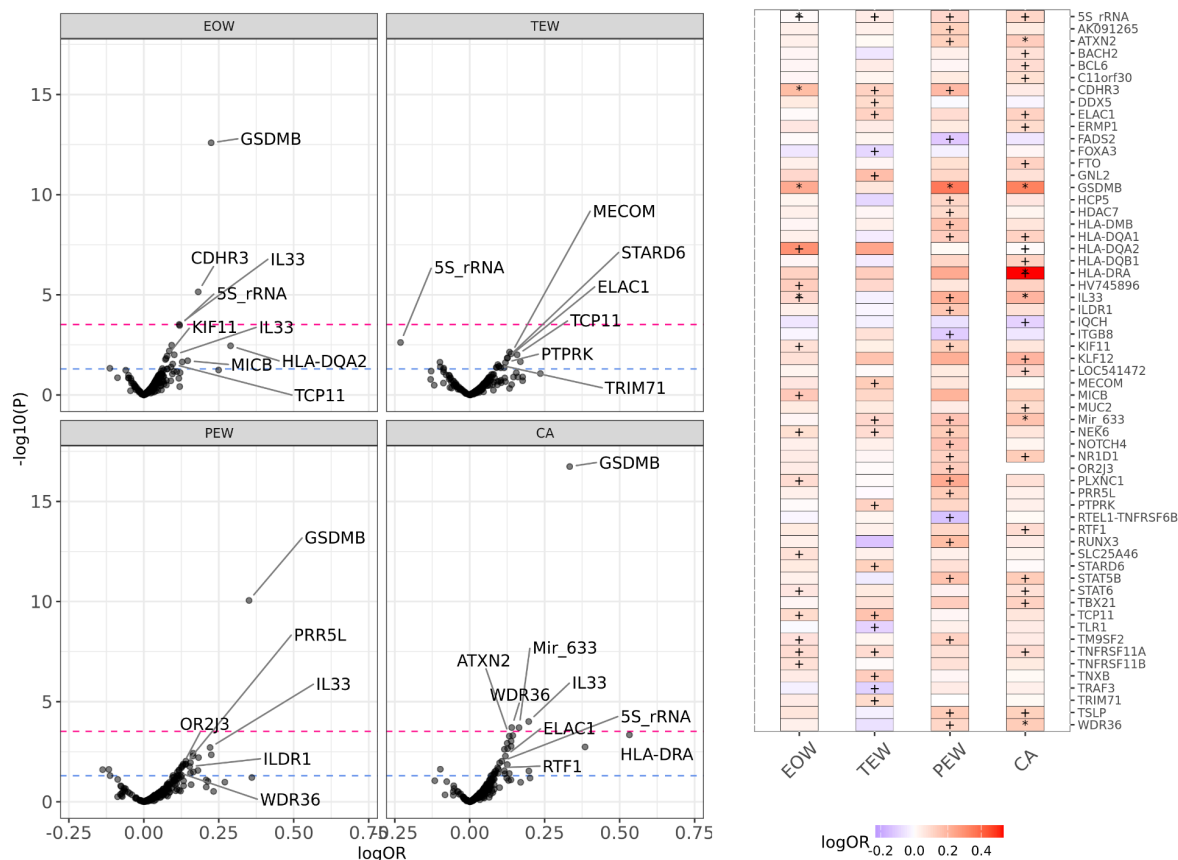
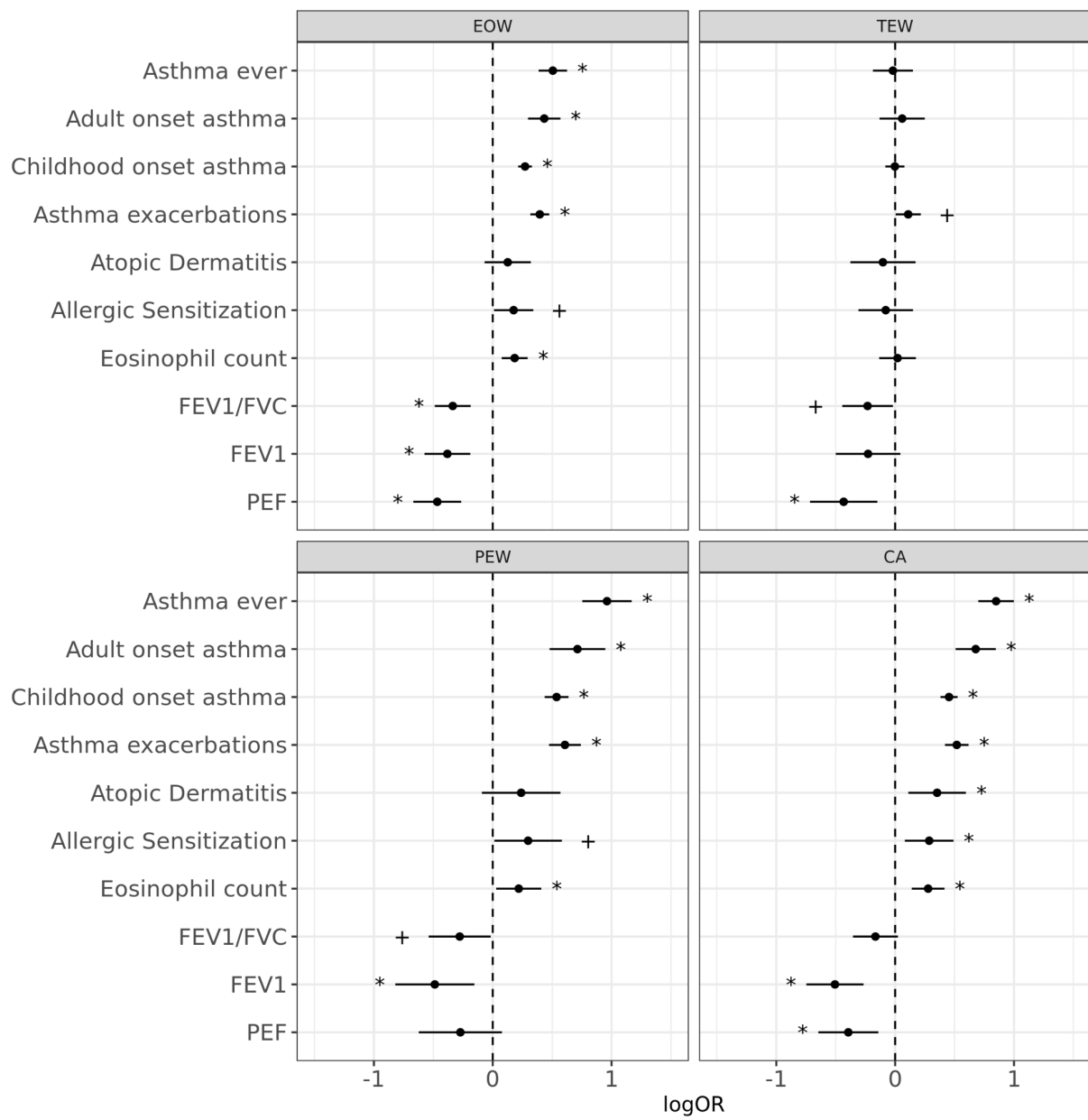


Figure 2



Supplementary material

Genetic characterization of preschool wheeze phenotypes

Kasper Fischer-Rasmussen¹, Raquel Granell³, Anders Ulrik Eliassen, PhD¹, Eskil Kreiner, MD, PhD¹, Casper-Emil Tingskov Pedersen, PhD¹, Yang Luo¹, Bo Chawes, MD, DMSc¹, Jakob Stokholm, MD, PhD¹, Ann-Marie Malby Schoos, MD, PhD¹, David Strachan⁴, Ashish Kumar^{5,6}, Anne-Marie Nybo Andersen⁷, Bjarke Feenstra⁷, Frank Geller⁷, Valérie Siroux⁸, Florence Demenais⁸, Emmanuelle Bouzigon⁸, Vincent Jaddoe^{9,10}, Ralf van der Valk^{9,10}, Liesbeth Duijts^{9,10,11}, Jordi Sunyer^{12,13,14}, Mónica Guxens^{12,13,14,15}, Marcella Marinelli^{13,18}, Mariona Bustamante^{12,14}, Joachim Heinrich^{19,20}, Marie Standl²¹, John Curtin^{23,24}, Chih-Mei Chen²⁵, Wenche Nystad²⁶, Bo Jacobsson^{27,28}, Ronny Myhre²⁷, Craig E Pennell^{32,33,34,35}, Denise Daley³⁶, Carole Ober³⁷, Dorret Boomsma³⁸, Abdel Abdellaoui³⁸, Angela Simpson, Clare Murray, John W Holloway^{39,40}, Sam Collins^{39,41}, Steve Turner^{42,43}, S Hasan Arshad^{39,41,44}, Anhar Ullah⁴⁵, Erik Melén^{5,6}, John Henderson^{3†}, Hans Bisgaard^{1†}, Anders Gorm Pedersen, MSc, PhD², Adnan Custovic²², Judith M. Vonk³⁰, Gerard H. Koppelman^{29,30}, Michael Kabesch²⁵, Klaus Bønnelykke, MD, PhD¹, GABRIEL consortium, the EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium

Cohort recruitment details, acknowledgements and funding statements	3
ALSPAC	3
B58C (British 1958 birth cohort)	5
BAMSE	6
CAPPS/SAGE	7
COAST	9
COPSAC2000	10
COPSAC2010	10
DNBC	11
EGEA	13
GENERATION R	14
GINI/LISA	16
IoW	18
INMA	19
MAAS	20
ISAAC II / MAGICS	22
NTR	23
RAINE	24
PIAMA	25
	129

MOBA	27
SWS	29
SEATON	30
Imputation of GWAS summary statistics	32
Data Visualization	32
Genetic Similarity Between Adult Traits and Preschool wheeze	33
Published GWAS on Asthma and Asthma-related Traits in Adulthood	33
Asthma ever in UK Biobank	34
Childhood vs. adult onset asthma in UK Biobank	34
Asthma exacerbations (0-6 year)	35
Atopic dermatitis	36
Allergic sensitization	36
Eosinophil count	36
Lung function measurements	37
Opposite Direction of Effects: Analysis of Known Asthma Variants	37
Supplementary figures	38
Supplementary tables	52
Supplementary code	59

Cohort recruitment details, acknowledgements and funding statements

ALSPAC

Recruitment

Pregnant women resident in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992 were invited to take part in the study. 20,248 pregnancies have been identified as being eligible and the initial number of pregnancies enrolled was 14,541. Of the initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above: The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented in the released data and reflecting enrolment status at the age of 24 is 906, resulting in an additional 913 children being enrolled (456, 262 and 195 recruited during Phases II, III and IV respectively). The phases of enrolment are described in more detail in the cohort profile paper and its update (see footnote 5 below). The total sample size for analyses using

any data collected after the age of seven is therefore 15,447 pregnancies, resulting in 15,658 fetuses. Of these 14,901 children were alive at 1 year of age (1,2).

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: [Explore data and samples | Avon Longitudinal Study of Parents and Children | University of Bristol](#)

Acknowledgements and Funding

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. FLG genotyping was funded by an MRC centenary award (awarded to L Paternoster).

B58C (British 1958 birth cohort)

Recruitment

The British 1958 birth cohort is an ongoing follow-up of all persons born in England, Scotland and Wales during one week in 1958. At ages 7, 11, 16, 23, 33 and 42 years, a history of hay fever or allergic rhinitis was obtained by interview with a parent (aged 7-16) or the cohort member (ages 23-42). At the age of 44-45 years, the cohort were followed up with a biomedical examination and blood sampling, from which a DNA collection was established as a nationally representative reference panel, and the presence of circulating IgE antibodies to house dust mite, mixed grass pollen and cat fur was ascertained by Hytec enzyme immunoassay, with a detection threshold of 0.35 kU/L.

Acknowledgements and Funding

We acknowledge use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC

GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research.

The funding agencies did not have any influence on design and conduct of the study; collection, management, and interpretation of the data; or preparation, review, or approval of the manuscript

BAMSE

Recruitment

BAMSE is a Swedish birth cohort study. A total number of 4,089 newborn infants were recruited between 1994 and 1996 in the Stockholm area¹⁹. The first questionnaire data, dealing with parental allergic diseases, socio-economic status and residential characteristics, was obtained when the children were about 2 months. Similar questionnaires, with a focus on the children's symptoms related to asthma and allergic diseases including eczema, were answered by the parents when the children were approximately 1, 2, 4, 8 and 12 years old. At 8 years of age, all children were invited to clinical testing, and blood samples were obtained from 2,480 children (~60%). DNA was extracted from 2,033 samples after exclusion of samples with too little blood, lack of

questionnaire data, or if parental consent to genetic analysis of the sample was not obtained.

Acknowledgements and Funding

BAMSE was supported by The Swedish Research Council, The Swedish Heart-Lung Foundation, Stockholm County Council (ALF) and the Strategic Research Programme (SFO) in Epidemiology at Karolinska Institutet. Genome-wide genotyping in BAMSE was funded by the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1.

CAPPS/SAGE

Recruitment

The Canadian Asthma Primary Prevention Study (CAPPS) was initiated in 1995 to assess the effectiveness of a multifaceted intervention program applied in the first twelve months of life on the primary prevention of asthma and other atopic disorders in high-risk infants (E6). Infants at high risk for developing asthma and other allergic disorders were identified and their mothers were recruited during their third trimester of pregnancy. Infants at high risk were defined as those who had a parent with asthma or two first-degree relatives with allergies or atopic disorders. The study had two recruitment centers in Canada (Vancouver and Winnipeg).

The Study of Asthma Genes and the Environment (SAGE), is derived from a populationbased cohort of 16,320 children born in province of Manitoba, Canada, between January 1, 1995 through December 31, 1995. A survey on child health and home environment exposures was mailed to the parents of this birth cohort in 2002. Following parent response to the mailed survey and approval to be contacted, a subset of children was invited to join the study at age 8-10 years. This included all children with parent-declared asthma and a comparable number of children without asthma, randomly selected following stratification of the received surveys by urban / rural location and family income. All recruited children underwent clinical assessment for asthma by a pediatric allergist (AB or JL) on the basis of the Canadian Asthma Consensus Guidelines.

Acknowledgements and Funding

The authors thank all families for their enthusiastic participation in this study. The authors also thank Diane Gagné, Muriel Grenon, and Dr. Paul Bégin for their invaluable participation in acquiring the subjects. Supported by the Canadian Institutes of Health Research and the Respiratory Health Network of the Fonds de la Recherche en Sante' du Que'bec, and by AllerGen NCE, Inc., which paid for the genotyping. P.B. and K.T. are supported by the Allergy, Genes and Environment Network (AllerGen). T.J.H. is recipient of an Investigator Award from the Canadian Institutes of Health Research and a Clinician–Scientist Award in Translational Research from the Burroughs Wellcome Fund. D.D. is supported by grants from the Canadian Institutes of Health Research, Institutes of Gender and Health, Genetics, Population and Public Health, and the IMPACT fellowship, and is also supported by the Lung

Association of British Columbia. C.L. is the chair-holder of the Canada Research Chair on genetic determinants in asthma and the director of the Genetics platform of the Respiratory Health Network of the Fonds de la Recherche en Sante' du Que'bec, which financially supports the Saguenay–Lac-St.-Jean sample.

COAST

Recruitment

The COAST longitudinal birth cohort study was designed to test for effects of immune development and early-life viral respiratory infections on the risk of allergic diseases and asthma (9). Expectant families were recruited during the prenatal period. Entry criteria required that at least one parent have self-reported allergies or asthma. All babies were delivered in Madison area hospitals. Babies were excluded from the study if they were 36 weeks gestation or less, or if they developed, during the neonatal period, significant respiratory distress or other health conditions affecting either the lung or the immune system. The COAST study was approved by the University of Wisconsin–Madison Human Subjects Committee (2013-1044), and informed consent was obtained before enrollment. The study population included 289 at birth and 275 at age 3 years.

Acknowledgements and Funding

The authors thank the children and families of the COAST (Childhood Origins of Asthma Birth Cohort Study) cohort for their support and commitment to their studies, and acknowledge and appreciate the unique efforts of the COPSAC and COAST research

teams. The Childhood Origins of Asthma Birth Cohort Study (COAST) is funded by NIH-NHLBI grant P01 HL070831.

COPSAC2000

Recruitment

The COPSAC2000 birth cohort study is a prospective clinical study of a birth cohort of 411 infants born to mothers with a history of asthma. The newborns were enrolled at the age of 1 month, the recruitment of which was previously described in detail.²⁵ The study was approved by the Ethics Committee for Copenhagen (KF 01- 289/96) and The Danish Data Protection Agency (2008-41-1754) and informed consent was obtained from both parents. For this study, only the parents were included.

Acknowledgements and Funding

COPSAC is funded by private and public research funds all listed on www.copsac.com. The Lundbeck Foundation (Grant no R16-A1694); The Danish Ministry of Health (Grant no 903516); Danish Council for Strategic Research (Grant no 0603-00280B); The Danish Council for Independent Research (Grant no 10-082884 and 271-08-0815); The Capital Region Research Foundation (No grant number) and NIH-NHLBI R01 HL129735 have provided core support for COPSAC.

COPSAC2010

Recruitment

Pregnant women from eastern Denmark were invited during 2008–2010 to a novel unselected ‘COPSAC₂₀₁₀’ cohort. The women visited the clinic during pregnancy weeks 24 and 36. Their children were followed at the clinic with deep phenotyping and collection of biological samples at nine regular visits until the age of 3 and at acute symptoms. Randomized controlled trials of high-dose vitamin D and fish oil supplements were conducted during pregnancy, and a trial of azithromycin for acute lung symptoms was conducted in the children with recurrent wheeze.

Acknowledgements and Funding

The authors gratefully express their gratitude to the children and families of the COPSAC cohort studies for all their support and commitment, and we acknowledge and appreciate the unique efforts of the COPSAC research team. COPSAC is funded by private and public research funds. The Lundbeck Foundation; the Pharmacy Foundation of 1991; Augustinus Foundation; the Danish Medical Research Council; and The Danish Pediatric Asthma Centre provided the core support for the study.

DNBC

Recruitment

The DNBC consists of approximately 100,000 pregnancies enrolled via general practitioners from across Denmark in 1996–2003 [Olsen J, Melbye M, Olsen SF, Sørensen TI a., Aaby P, Nybo Andersen AM, et al. The Danish National Birth Cohort--its background,

structure and aim.]. Mothers and their offspring have participated in multiple follow-ups, with the most recent 18 years after offspring birth. At the 11-year follow-up (2010–2014), approximately half of the caregivers (mostly mothers) responded to an online questionnaire regarding their child's health and a wide range of questions regarding housing and indoor home environment at the time their child turned 11 years. At the 18-year follow-up, DNBC children were invited to reply to an online questionnaire in which questions on asthma diagnosis, medication and symptoms were asked. Individuals who had relevant data available from the 11-year and 18-year follow-ups were included in this study (Fig. 1).

Acknowledgements and Funding

The Danish National Birth Cohort was established with a significant grant from the Danish National Research Foundation. Additional support was obtained from the Danish Regional Committees, the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Health Foundation and other minor grants. The DNBC Biobank has been supported by the Novo Nordisk Foundation and the Lundbeck Foundation. Follow-up of mothers and children have been supported by the Danish Medical Research Council (SSVF 0646, 271-08-0839/06-066023, O602-01042B, 0602-02738B), the Lundbeck Foundation (195/04, R100-A9193), The Innovation Fund Denmark 0603-00294B (09-067124), the Nordea Foundation (02-2013-2014), Aarhus Ideas (AU R9-A959-13-S804), University of Copenhagen Strategic Grant (IFSV 2012), and the Danish Council for Independent Research (DFF – 4183-00594 and DFF—4183-00152).

EGEA

Recruitment

The study population included 348 asthmatics and their family members and 416 control subjects. Probands (asthmatics and controls) aged 7 to 65 were ascertained in five French cities (Paris, Lyons, Marseilles, Montpellier, and Grenoble). All pro-bands and their two parents were born in France. In roughly 10,000 self-completed questionnaires (6,400 for potential cases and 3,600 for potential control subjects), 1,589 probands fulfilled inclusion criteria regarding place of birth and family structure (either proband with a spouse and at least one child, or proband with at least one sibling and two parents). A positive answer to four questions (Table 1) identified those with asthma. A total of 278 probands fulfilled this criterion. Seventy more who only answered positively to two or three questions were also included after examination of their medical record. Subjects with asthma comprised 213 adults and 135 children. Children control subjects were recruited in surgery hospitals. Two thirds of adult control subjects were population based (electoral rolls), and one-third were from surgery hospitals or check-up centers. Overall, half of the controls were population based.

Acknowledgements and Funding

The authors thank all the clinicians, interviewers, technicians, nurses, and secretaries who collected and coded the data and helped coordinate the study through the years. They also thank the subjects who participated in the study, without whose cooperation

the study would not have been possible. Supported by INSERM/MSD-Chibret Convention, INSERM Networks for Clinical Research Grant 489012 and Public Health Research Grant 4934009.

GENERATION R

Recruitment

The Generation R Study is a population-based prospective cohort study of pregnant women and their children from fetal life onwards in Rotterdam, The Netherlands. All children were born between April 2002 and January 2006, and currently followed until young adulthood. Of all eligible children in the study area, 61% were participating in the study at birth. Cord blood samples including DNA have been collected at birth. The study protocol was approved by the Medical Ethical Committee of the Erasmus Medical Centre, Rotterdam (MEC 217.595/2002/20). Written informed consent was obtained from parents of all participants.

Acknowledgements and Funding

The Generation R Study gratefully acknowledges the contributions of the children and their parents, the general practitioners, the hospitals and the midwives and pharmacies in Rotterdam. We would like to thank A. Abuseiris, K. Estrada, Dr. T. A. Knoch, and R. de Graaf as well as their institutions Biophysical Genomics, Erasmus MC Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and #

01 IG 07015 G, for access to their grid resources. We thank M. Jhamai, M. Ganesh, P. Arp, M. Verkerk, L. Herrera and M. Peters for their help in creating, managing and performing quality control for the genetic database. Also, we thank K. Estrada and C. Medina-Gomez for their support in the creation and analysis of imputed data. The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and the Faculty of Social Sciences of Erasmus University Rotterdam, the Municipal Health Service, Rotterdam area, the Rotterdam Homecare Foundation and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC; Rotterdam). The generation and management of genotype data for the Generation R Study were performed at the Genetic Laboratory of the Department of Internal Medicine at Erasmus Medical Center. The Generation R Study is made possible by financial support from the Erasmus Medical Center (Rotterdam), the Erasmus University Rotterdam and the Netherlands Organization for Health Research and Development (ZonMw; 21000074). The project received funding from the European Union's Horizon 2020 research and innovation programme (LIFECYCLE project, no 733206, 2016, EUCAN-Connect, project no. 824989, 2019; ATHLETE project, no. 874583, 2020). Dr Vincent Jaddoe received an additional grant from the Netherlands Organization for Health Research and Development (VIDI 016.136.361) and the European Research Council (Consolidator Grant, ERC-2014-CoG-648916). Dr Liesbeth Duijts received an additional grant from European Union's Horizon 2020 research and innovation programme (ENDOMIX project, no. 101136566, 2024). The researchers are independent from the funders. The study sponsors had no role in the study design, data analysis, interpretation of data, or writing of this report.

GINI/LISA

Recruitment

The influence of Life-style factors on the development of the Immune System and Allergies in East and West Germany PLUS the influence of traffic emissions and genetics (LISApplus) Study is a population based birth cohort study. A total of 3094 healthy, full-term neonates were recruited between 1997 and 1999 in Munich, Leipzig, Wesel and Bad Honnef. The participants were not preselected based on family history of allergic diseases.³⁹ A total of 5991 mothers and their newborns were recruited into the German Infant study on the influence of Nutrition Intervention PLUS environmental and genetic influences on allergy development (GINIplus) between September 1995 and June 1998 in Munich and Wesel. Infants with at least one allergic parent and/or sibling were allocated to the interventional study arm investigating the effect of different hydrolysed formulas for allergy prevention in the first year of life.⁴⁰ All children without a family history of allergic diseases and children whose parents did not give consent for the intervention were allocated to the non-interventional arm. Detailed descriptions of the LISApplus and GINIplus studies have been published elsewhere^{1,2}. DNA was collected at the age 6 and 10 years and 1511 children from the Munich study center from both studies were genotyped. For both studies, approval by the local Ethics Committees and written consent from participant's families were obtained.

Acknowledgements and Funding

The authors thank Martina Müller-Nurasyid for her help with imputation and analysis, all families for participation in the studies and the LISApplus and GINIplus study teams for their excellent work. The GINIplus study team wishes to acknowledge the following: Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg (Heinrich J, Brüske I, Schulz H, Flexeder C, Zeller C, Standl M, Schnappinger M, Ferland M, Thiering E, Tiesler C); Department of Pediatrics, Marien-Hospital, Wesel (Berdel D, von Berg A); Ludwig-Maximilians-University of Munich, Dr von Hauner Children's Hospital (Koletzko S); Child and Adolescent Medicine, University Hospital rechts der Isar of the Technical University Munich (Bauer CP, Hoffmann U); IUF- Environmental Health Research Institute, Düsseldorf (Schikowski T, Link E, Klümper C). The LISApplus study team wishes to acknowledge the following: Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology I, Munich (Heinrich J, Schnappinger M, Brüske I, Ferland M, Lohr W, Schulz H, Zeller C, Standl M); Department of Pediatrics, Municipal Hospital "St. Georg", Leipzig (Borte M, Gnodtke E); Marien Hospital Wesel, Department of Pediatrics, Wesel (von Berg A, Berdel D, Stiers G, Maas B); Pediatric Practice, Bad Honnef (Schaaf B); Helmholtz Centre of Environmental Research – UFZ, Department of Environmental Immunology/Core Facility Studies, Leipzig (Lehmann I, Bauer M, Röder S, Schilde M, Nowak M, Herberth G, Müller J, Hain A); Technical University Munich, Department of Pediatrics, Munich (Hoffmann U, Paschke M, Marra S); Clinical Research Group Molecular Dermatology,

Department of Dermatology and Allergy, Technische Universität München (TUM), Munich (Ollert M).

IoW

Recruitment

The IOWBC is a single-centre study designed to represent the community population. All children born on the Isle of Wight in a defined period were eligible for inclusion. Ethics approval was obtained from the local/national ethics committees at recruitment of the birth cohort between January 1989 and February 1990, and subsequently at each assessment. The cohort was recruited through the 1509 women who gave birth to 1536 children on the IOW during the recruitment period. All 1509 mothers were recruited and consented to complete questionnaires and provide samples soon after birth. Parental consent was obtained from 1456 of the 1536 children for inclusion into a longitudinal study of asthma and allergic disease.

Acknowledgements and Funding

Recruitment and initial assessment for the first 4 years of age was supported by the Isle of Wight Health Authority. The 10-year follow-up of this study was funded by the National Asthma Campaign, UK (Grant No 364) and the 18-year follow-up by NIH/NHLBI R01 HL082925-01. The 21 year assessment was funded by Medical

Research Council, UK (G23369). The 26 years assessment was supported by The David Hide Asthma & Allergy Research Trust.

INMA

Recruitment

Population-based birth cohorts were established as part of the INMA – INfancia y Medio Ambiente [Environment and Childhood] Project in several regions of Spain following a common protocol. This analysis uses the INMA subcohort of Menorca established between 1997 and 1998 in Menorca island. This project aims to study the associations between pre- and postnatal environmental exposures and growth, health, and development from early fetal life until adolescence and has been described previously in detail.⁴⁴ Pregnant women were enrolled during the 1st trimester of pregnancy at public primary health care centers or public hospitals and children were followed after birth until 18y of age. Informed consent was obtained from all participants and the study was approved by the Hospital Ethics Committee.

Acknowledgements and Funding

This study was funded by grants from Instituto de Salud Carlos III (97/0588, 00/0021-02, G03/176, 04/1705, 06/1756, 09/1958 and PS09/00432), Fundació La Marató de TV3, Obra Social la Caixa (IV Convocatoria de ayudas a la Investigación en Enfermedades Neurodegenerativas) and European Community (Contract No.

QLK4-CT-2000-00263). DNA extractions and genotyping was performed at the Spanish National Genotyping Centre (CEGEN-Barcelona). The authors would particularly like to thank all the participants for their generous collaboration. The authors are grateful to Mireia Garcia, Maria Victoria Estraña, Maria Victoria Iturriaga, Cristina Capo and Josep LLuch for their assistance in contacting the families and administering the questionnaires. A full roster of the INMA Project Investigators can be found at http://www.proyectoinma.org/presentacion-inma/listado-investigadores/en_listado-investigadores.html.

MAAS

Recruitment

The Manchester Asthma and Allergy Study is an unselected (i.e. population-based), birth cohort study. The setting is the maternity catchment area of Wythenshawe and Stepping Hill Hospitals, comprising of 50 square miles of South Manchester and Cheshire, UK, a stable mixed urban-rural population. Study was approved by the Local Research Ethics Committee. Informed consent was obtained from all parents.

Screening & Recruitment

All pregnant women were screened for eligibility at antenatal visits (8th-10th week of pregnancy). The study was explained to the parents, and informed consent for initial questionnaires and skin prick testing was obtained. Both parents completed a questionnaire about their and their partner's history of asthma and allergic diseases and

smoking habits. If the pregnant woman's partner was not present at the antenatal clinic visit, an invitation was sent for him to attend an open-access evening clinic for skin prick testing and questionnaire. Once both parents had completed questionnaires and skin prick testing, a full explanation of the proposed future follow-up for the child was given. Of the 1499 couples who met the inclusion criteria (<10 weeks of pregnancy, maternal age >18 years, questionnaire and skin test data available for both parents), 288 declined to take part in the study. A total of 1185 participants had at least some evaluable data. *Follow-up:* The children have been followed prospectively, and attended review clinics at ages 1, 3, 5, 8 and 11 years.

Acknowledgements and Funding

We would like to thank the children and their parents for their continued support and enthusiasm. We greatly appreciate the commitment they have given to the project. We would also like to acknowledge the hard work and dedication of the study team (post-doctoral scientists, research fellows, nurses, physiologists, technicians and clerical staff). This report includes independent research supported by National Institute for Health Research Respiratory Clinical Research Facility at Manchester University NHS Foundation Trust (Wythenshawe). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. MAAS was supported by the Asthma UK Grants No 301 (1995-1998), No 362 (1998-2001), No 01/012 (2001-2004), No 04/014 (2004-2007), the BMA James Trust and Medical Research Council, UK (G0601361) and The Moulton Charitable Foundation (2004-current); the Medical Research Council

(MRC) Grants G0601361, MR/K002449/1 and MR/L012693/1, and Angela Simpson is supported by the NIHR Manchester Biomedical Research Centre. The authors would like to acknowledge the North West Lung Centre Charity for supporting this project.

ISAAC II / MAGICS

Recruitment

This study was performed using asthma cases derived from the Multicenter Asthma Genetics in Childhood (MAGIC) study (E1) and reference children randomly selected from the cross-sectional International Study of Asthma and Allergies in Childhood Phase II (ISAAC II) in Munich and Dresden (E2). Written informed consent was obtained from the parents of all children from both populations and the study protocols were approved by the ethics committees of each participating center.

The ISAAC II study (E2) was conducted in Germany (Munich and Dresden) to assess the prevalence of asthma in school children at 9–11 years of age. A random subset of 800 German children was selected for a GWA study as previously described (E3). This random subset was investigated in this study for BHR, atopy, allergic rhinitis and atopic eczema and served as a reference to asthma cases derived from the MAGIC study in the GABRIEL consortium GWA study.

Acknowledgements & funding

Supported by the German ministry of education and research (BMBF) as part of the National Genome Research Network (NGFN), with grant NGFN 01GS0810, and by the European Commission as part of the GABRIEL (a multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract (no. 018996) under the Integrated Program LSH-2004-1.2.5-1. M.K. is a member of the German Center for Lung Research (DZL). R.C.P. is supported by a research fellowship from the GRK1441 (Allergy) program of Deutsche Forschungsgemeinschaft (DFG) and was supported by the PhD program “Molecular Medicine,” Hannover Biomedical Research School (HBRS).

NTR

Recruitment

In the Young Netherlands Twin Register (YNTR), twins are followed from birth onwards. Around every two to three years, a survey is sent out inquiring about physical and mental health. At ages 1, 2, 3, 5, 7, 10 and 12 years, the surveys are completed by the parents and/or teachers, from age 14 onward, twins and their non-twin siblings are asked to complete the surveys by themselves. At age 5 years of the twins, parents are asked a series of questions about allergies and rhinitis.⁴⁵ Adult and adolescent twins (ANTR) have been recruited through city councils and media, and complete a self-rating survey every few years. Surveys contain a series of questions about doctor diagnosed diseases and disorders. Subsamples of twin families have taken part in biobank and genotyping studies.

Acknowledgements and Funding

The Netherlands Organization for Scientific Research (NWO 480-15-001/674); Netherlands Twin Registry Repository: researching the interplay between genome and environment; Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007); Amsterdam Public Health (APH); KNAW Academy Professor Award (PAH/6635); the Avera Institute, Sioux Falls, South Dakota (USA) and Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995).

RAINE

Recruitment

The Western Australian Pregnancy Cohort (Raine) Study is a prospective pregnancy cohort where 2900 were recruited from King Edward Memorial Hospital between 1989 and 1991. Data were collected throughout pregnancy and the children have been followed-up at ages 1, 2, 3, 5, 8, 10, 14, 17, 18, 20, and 23. Ethics approval for this study was obtained from King Edward Memorial Hospital and Princess Margaret Hospital. Participants were consented to being involved in this study prior to each follow-up.

Acknowledgements and Funding

This study was supported by the National Health and Medical Research Council of Australia [grant numbers 572613, 403981 and 003209] and the Canadian Institutes of Health Research [grant number MOP-82893]. The authors are grateful to the Raine Study participants and their families, and to the Raine Study research staff for cohort

coordination and data collection. The authors gratefully acknowledge the NH&MRC for their long term contribution to funding the study over the last 25 years and also the following Institutions for providing funding for Core Management of the Raine Study: The University of Western Australia (UWA), Curtin University, Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, The Telethon Kids Institute, Women and Infants Research Foundation (King Edward Memorial Hospital), Murdoch University, The University of Notre Dame (Australia), and Edith Cowan University. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility).

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia

PIAMA

Recruitment

PIAMA is a birth cohort study consisting of two parts: a placebo controlled intervention study in which the effect of mite impermeable mattress covers was studied and a natural history study in which no intervention took place. Details of the study design have been published previously. Recruitment took place in 1996-1997. A screening questionnaire was distributed to pregnant women visiting one of 52 prenatal clinics at three regions in the Netherlands. A total of 10,232 pregnant women completed a

validated screening questionnaire. Mothers reporting a history of asthma, current hay fever or allergy to pets or house dust mite were defined as allergic. Based on this screening, 7862 women were invited to participate, of whom 4,146 women (1327 allergic and 2819 nonallergic) gave written informed consent. The response rates to the annual questionnaires ranged from 3030 (92%) at age 1 to 2732 (83%) at age 8 years. DNA was collected from 2162 children at age 4 and/or 8 years. Genome-wide genotyping was performed within the framework of the Gabriel Consortium. For this study, DNA samples from 213 children with parental reported doctor diagnosed asthma ever at age 8 years and from 213 controls without doctor diagnosed asthma or wheeze ever at age 8 years were provided. From these children, 50 cases (≥ 1 skin test positive) and 176 controls were selected. Specific allergens tested for were *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Alternaria alternata*, mixed grass pollen, mixed tree pollen, cat, and dog. Skin test positivity was defined as one or more positive skin test with a diameter of 3 mm or larger. For specific IgE to inhalant allergens, we included 161 cases (≥ 0.35 IU/ml) and 176 controls. We tested by Radio Allergo Sorbent Test for any of the following allergens: *Dermatophagoides pteronyssinus*, Fel d1, Can f1, *Dactylis Glomerata*, *Betula verrucosa*, *Alternaria Alternata*, milk and egg. Controls were defined as children without specific IgE (<0.35 IU/ml) to these allergens. Genotyping was performed with an Illumina Human610 quad array. SNPs were excluded that fulfilled one or more of the following criteria: p-value for test of Hardy-Weinberg equilibrium $\leq 1E-7$, genotyping call rate $<95\%$ or MAF $<1\%$. SNPs were imputed with IMPUTE version 2 software using HAPMAP CEU release #22

b36. Genome-wide association analyses were performed using SNPTTEST version 1.1.5. The Medical Ethical Committees of the participating institutes approved the study.

Acknowledgements & funding

The PIAMA study would like to thank all participants, and co-investigators of the study. The PIAMA study is supported by the Dutch Asthma Foundation (grant 3.4.01.26, 3.2.06.022, 3.4.09.081 and 3.2.10.085CO), the ZonMw (a Dutch organization for health research and development; grant 912-03- 031), and the ministry of the environment. Genome-wide genotyping was funded by the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1 Post genomic approaches to understand the molecular basis of asthma aiming at a preventive or therapeutic control.

MOBA

Recruitment

Norway has ~4.5 million inhabitants, and ~55 000 births a year. The target population of the study is all women who give birth in Norway. There are no exclusion criteria. All hospitals and maternity units with more than 100 births annually, altogether 52 units, are to be included, and by January 2006, 50 units participate in the study. For practical reasons, the sampling frame comprises pregnant women who attend routine ultrasound examination. Together with appointments for ultrasound scanning in week 17–18 of

pregnancy, the pregnant women receive a postal invitation that includes an informed consent form, the first questionnaire, an information brochure as well as consent form and questionnaire for the father.

The pregnancy is the unit of observation, and a woman can participate in the study with more than one pregnancy. Each pregnancy is given an identification number, and all other data, be it from the mother, father, or the child, are linked to this number. A participant is a pregnant woman who has sent in a written informed consent to participate. The consent states that biological material is only to be used for the study of causes of disease and that analyses can also be performed in other countries. No laboratory results will be provided to the participants. They agree that MoBa can access health registries as long as it is within the general aim of the study. The participants can choose to withdraw at any time (will not receive more questionnaires) or to be deleted from the study (so that all data are deleted). New sub-projects require a new written consent if they imply direct contact with the participants. A detailed protocol of the study including the consent can be found elsewhere (<http://www.fhi.no/morogbarn>).

Acknowledgements & funding

The donations of questionnaire data and biological material from MoBa participants are gratefully acknowledged. The study is supported by The Norwegian Ministry of Health, NIH/NIEHS (grant no. N01-ES-85433), NIH/NINDS (grant no. 1 U01 NS 047537) and Norwegian Research Council/FUGE (grant no. 151918/S10).

SWS

Recruitment

Participants were mother-child pairs from the Southampton Women's Survey; a cohort study with the objective, among others, of studying early life environmental influences on child growth and development, including respiratory disease and asthma in childhood[18]. During the period 1998-2002, 12,579 women aged 20-34 were recruited through their GP surgeries prior to conception. There were 1973 births before the end of 2003. Those born before 35 weeks and/or twins were excluded from the study and only one child per SWS mother was included. Parental consent was obtained and ethics approval was granted by the Southampton and South West Hampshire Local Research Ethics Committee (LREC Number 276/97, 307/97, 089/99, 125/98 and 06/Q1702/104).

Acknowledgements & funding

We thank the general practitioners and midwives in Southampton for their assistance in the SWS. Many staff make up the SWS Study Group and we are grateful for all their work in recruiting and interviewing the participants, and processing the data and the samples. Finally, we are extremely grateful to the participants in the SWS who have given so much time and information about themselves to the Survey. The SWS is grateful for financial support from the UK Medical Research Council, the University of Southampton, the Dunhill Medical Trust, the British Heart Foundation, and the Food Standards Agency. Additional support for ancillary studies has been received from the

BUPA Foundation, Hope, Cancer Research UK, the Biotechnology and Biological Sciences Research Council, the Arthritis Research Campaign, SPARKS (Sport Aiding medical Research for KidS), the British Lung Foundation, the National Osteoporosis Society of Great Britain, the Research Council of Norway, and the US National Institutes of Health.

SEATON

Recruitment

The database from the 15 year assessment will be used to contact participants by post. Non-respondents will be contacted a second time by post. We will ask staff in Data Management at the University of Aberdeen to identify and anonymously pass on letters of invitation to participants who do not respond to the second mailing. Participants will be asked to complete a respiratory questionnaire (online or by paper). We anticipate approximately 700 of the original cohort members to actively take part in this study and to get a more complete dataset, we will seek approval from the Grampian Safe Haven and Information Service Division to link the SEATON cohort to GP prescribing to identify those children who have developed asthma (identified by receipt of inhaled steroids).

Acknowledgements & funding

The authors thank the midwifery staff of the Aberdeen Maternity Hospital, Mr. George Henderson, Mrs. Alison Scaife, and Mrs. Gillian Moir for collecting and analyzing the

cord blood samples, and all the mothers and children who have taken part in the study. This study was funded by grants from Asthma UK and Tenovus Scotland neither of whom contributed to the study design, data analysis or preparation of this manuscript. The researchers are independent of Asthma UK and Tenovus Scotland.

Imputation of GWAS summary statistics

To increase the overlap between our summary results and the published GWAS we utilized Functionally-Informed Z-score Imputation (<https://pypi.org/project/pyfizi/>, <https://github.com/bogdanlab/fizi>) to impute up to 722,277 additional variants based on the 1000 genomes panel (phase 3), relying on the Gaussian Imputation Algorithm (3). FIZI-imputed variants with an imputation accuracy (r^2_{pred}) less than 0.9 were excluded from the analysis. Apart from increasing genotype density of the summary results, FIZI imputation also converts variant estimates to Z-scores. See **supplementary table 4** for more details on the number of imputed variants. We converted Z-scores back to effect sizes using the following equations from Zhu et al. 2016 (4).

$$\hat{b} = z / \sqrt{2p(1 - p)(n + z^2)}$$

$$SE = 1 / \sqrt{2p(1 - p)(n + z^2)}$$

Where p is the minor allele frequency of the SNP, z is the z-score and n is the sample size.

Data Visualization

Data analysis and visualization was conducted in R (version 4.2.2). Volcano plots, forest plots and heatmaps were created using 'ggplot2' (5), while Manhattan plots were created using the R-package 'ggfastman'.

Genetic Similarity Between Adult Traits and Preschool wheeze

Genetic correlation was approximated with GSMR by fitting a slope between coefficients of lead variants from the related traits and the coefficients of the same variants for the wheeze phenotypes. The intercept was fixed at zero to ensure the slope reflected a proportional relationship between variant coefficients. Each variant is weighted with its strength of association in both traits while accounting for residual linkage disequilibrium (LD) to decrease signal inflation. Variants were included if they were significantly associated with the related traits ($P < 5e-8$) and not in strong LD with each other ($r^2 < 0.05$). Coefficients estimated from GSMR can be estimated as odds-ratios, since the outcome GWAS is binary phenotype. No conditional analysis was performed to exclude variants in residual LD, since GSMR accounts for weak correlations between the included variants (6) (see Supplementary Code).

Published GWAS on Asthma and Asthma-related Traits in Adulthood

Shared genetic influence was characterized by utilizing existing knowledge of genetic risk factors for asthma related traits in adulthood, such as adult onset asthma, atopy (atopic dermatitis & allergic sensitization), eosinophil count and lung function.

Asthma ever in Global Biobank Meta-analysis initiative

See supplementary data under cohort description in Tsuo et al. 2022 for complete description of asthma ever for each cohort in the meta-analysis. We only included individuals of European descent (7).

Childhood vs. adult onset asthma in UK Biobank

We used the age at first diagnosis reported in fields 3786 and 22147 (the age was averaged when values in both fields were available. Those diagnosed before 19 were assigned to childhood onset asthma (COA, N = 14,873) and those diagnosed between 20 & 60 years of age were assigned to adult onset asthma (AOA, N = 26,582). A set of 300,671 controls who did not suffer from any allergic disease (asthma, hay fever, eczema, or other allergies) was used as controls (8).

Asthma exacerbations (0-6 year)

Children with repeated acute hospitalizations for asthma (ICD8 codes 493, ICD-10 codes J45-46) between ages 2 to 6 years were identified from the Danish National

Patient Register. Inclusion criteria required at least two hospitalizations lasting >1 day, with a minimum 6-month gap between episodes. Exclusions included comorbidities, chronic diagnoses affecting asthma risk, low birth weight (<2.5 kg), or gestational age <36 weeks. Cases were characterized based on asthma and acute bronchitis hospitalizations and concurrent atopy.

Control individuals (5328 non-asthmatic) were from the Inter99 cohort, a Danish study on ischemic heart disease prevention. Asthma cases in the iPSYCH study were defined using National Patient Registries data, identifying those with at least one hospitalization for asthma exacerbations in the first 6 years of life (ICD8-codes 493, ICD-10 codes J45-46). Individuals without such hospitalizations served as controls in both studies. In the combined dataset, 525,976 SNPs remained after quality control (9,10).

Atopic dermatitis

In FinnGen, cases needed ICD-10 code L20, ICD-9 code 6918, or ICD-8 code 691 in health registries, excluding ICD-9 code 6918X. Those lacking these codes were controls. In Estonian and UK Biobanks, L20 indicated AD cases, while no L20 record meant controls.

Due to heterogeneity in AD definitions, FinnGen conducted sensitivity analyses with stricter criteria: (1) ICD-10 code L20.0, ICD-9 code 6918B, or ICD-8 code 69100 ("AD strict"), and (2) KELA reimbursement codes 134, 395, or 317 ("AD reimb"). "AD strict"

had 6,412 cases and 253,666 controls, while "AD reimb" had 6,739 cases and 236,030 controls (11).

Allergic sensitization

See supplementary data under cohort description in Waage et al. 2018 for complete description of allergic sensitization for each cohort in the meta-analysis (7).

Eosinophil count

Eosinophil count was determined by flow cytometry gates. Samples were excluded based on certain criteria: pregnancy, acute medical/surgical illness, blood cancer, leukemia, lymphoma, chemotherapy, myelodysplastic syndrome, bone marrow transplant, congenital/hereditary anemia (e.g., sickle cell anemia or thalassemia), HIV, end-stage kidney disease, dialysis, EPO treatment, splenectomy, cirrhosis, and extreme measurements (e.g., WBC count $> 100 \times 10^9/L$ with $> 5\%$ immature cells or blasts).

WBC subtypes were represented using relative counts, and raw phenotypes were regressed on age, age-squared, sex, principal components, and cohort-specific covariates. WBC-related traits were log10 transformed before regression, and residuals were inverse normalized for cohort-level association analysis or GWAS (12).

Lung function measurements

Lung function measurements were derived from spirometry measurements in UK Biobank. We excluded FVC measurements from our analysis due to low association with airway obstruction and thus not believed to be asthma related (13).

Opposite Direction of Effects: Analysis of Known Asthma Variants

A small group of variants previously associated with asthma were also associated with wheeze phenotypes but with opposite direction of effect, although none of them were significant after adjusting for multiple testing. One SNP with opposite direction worth mentioning is rs174535 in the fatty acid desaturase (FADS) cluster. The T-allele of rs174535 was previously associated with a higher risk of adult-onset asthma (8) but was associated with a lower risk of persistent wheeze in our study. The T-allele of rs174535 is in high LD with the A-allele of rs1535 ($r^2 = 0.94$), and both alleles tag the major haplotype in individuals of European ancestry, which is associated with higher plasma levels of n-3 long-chain polyunsaturated fatty acids (LCPUFA), including DHA and EPA (14). In keeping with our results, a previous study found that children carrying the minor alleles at this locus (G allele of rs1535), and thereby lower levels of EPA and DHA, had increased risk of childhood asthma that was abrogated by higher dietary intake of EPA and DHA (15). Similarly, a randomized clinical trial with fish oil

supplements to pregnant mothers showed an increased risk of persistent wheeze in the children of mothers carrying the LCPUFA-reducing G-allele of rs1535, which was abrogated by the pregnancy supplementation (16). This suggests that genetic variation in the FADS region, and resultant changes in LCPUFA levels, has opposite effects on adult-onset asthma and preschool wheezing.

Supplementary tables

Supplementary table 1. The population used in this study comes from the EAGLE consortium which aims to understand how early childhood traits affect disease and well-being in later life [1]. The Gabriel Project also constitutes a substantial part of the included studies and this comprises one of the biggest genetic studies of asthma, genotyping 10,365 individuals with asthma diagnosis and 16,110 unaffected [2]. This table contains all basic characteristics from all included studies in the discovery phase.

See *STable1.DiscoveryCohortCharacteristics.xlsx*

Supplementary table 2. This table contains all basic characteristics from all included studies in the replication phase.

See *STable2.ReplicationCohortCharacteristics.xlsx*

Supplementary table 3. This table contains an overview of the phenotype definitions used in this study.

Phenotype name	Definition
Early onset wheeze (EOW)	Asthma diagnosis or 3 or more wheezy episodes before age 3.

Transient early wheeze (TEW)	Asthma diagnosis or 3 or more wheeze episodes before age 3, but free of symptoms and medicine in recent year by age 6 or closest school age status.
Persistent wheeze (PW)	Asthma diagnosis or 3 or more wheezy episodes before age 3 and current asthma or any wheezy episodes in recent year by age 6 or closest school age status.
Current asthma (CA)	Symptomatic or receiving medicine in recent year by age 6 or closest school age status.

Supplementary table 4. This table contains the number of cases and controls for each study included in the discovery phase.

Discovery	EOW		TEW		PW		CA	
Cohort	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
ALSPAC	1272	4667	496	5193	338	5873	682	5020
BAMSE	141	341	62	420	65	417	236	246
COPSAC 2000	61	255	15	303	48	288	45	262
DNBC	103	850	42	850	12	850	32	850
EGEA	203	1632	-	-	-	-	-	-
GENR	667	2294	383	1985	109	2259	286	3,089
INMA	163	644	91	648	50	692	139	636
LISA/GINI	74	963	46	1005	27	1100	118	1009
MAAS	294	568	151	545	143	719	166	696
MOBA	88	528	-	-	-	-	-	-
PIAMA	125	252	52	317	55	306	62	317
RAINE	82	1116	-	-	-	-	-	-
MAGICS	-	-	-	-	130	397	284	397
B58C TIDGC	-	-	-	-	-	-	40	2215
B58C WTCC	-	-	-	-	-	-	14	1319
TOTAL	3273	14110	1338	11266	977	12901	2104	12970
%	18.83%		10.62%		7.04%		13.96%	

Supplementary table 5. This table contains the number of cases and controls for each study included in the replication phase.

Replication	EOW		TEW		PW		CA	
Cohort	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
SWS	358	846	131	589	94	626	246	492
CAPPS/SAGE	81	182	50	149	21	149	49	201
NTR	150	1886	34	1992	112	1914	-	-
COAST	-	-	14	187	-	-	-	-
COPSAC 2010	124	555	88	559	30	617	146	501
SEATON	81	492	35	533	46	522	80	451
Isle of Wight	161	847	68	917	65	920	179	768
TOTAL	1235	6120	734	6589	471	6171	725	3928
%	20.2		11.1		7.6		18.5	

Supplementary table 6. This table contains the number of additional variants imputed from the summary stats. Each phenotype is represented by a row, containing information the number of SNPs included in the study (“N SNPs in study”), minimum number of individuals carrying this SNP before imputing (“Minimum N 90th percentile”), how many SNPs were included (“Included SNPs for imputation”), how many SNPs were imputed and how many total SNPs were included in the study.

Phenotype	N SNPs in study	Minimum N (90th percentile)	Included SNPs for imputation	Imputed SNPs	Total SNPs
EOW	2,467,019.00	10049	2,402,878.00	606,806.00	3,073,825.00
TEW	2,465,868.00	8436	2,360,928.00	599,260.00	3,065,128.00
PW	2,465,601.00	9167	2,288,960.00	539,994.00	3,005,595.00
CA	2,471,230.00	12025	2,242,028.00	515,792.00	2,987,022.00

Supplementary table 7. This table contains the results of the variants carried forward to the replication phase. The discovery effect size is also displayed in this table.

Phenotype	Chr	SNP	EA/NEA	Nearest gene	Discovery				Replication			
					OR (95% CI)	P	N	Phet	OR (95% CI)	P	N	Phet
TEW	7	rs2214823	A/G	<i>HOPV</i> <i>CACNA2D1</i>	1.23 (1.15, 1.35)	2.32E-40	12,604	0.116	1.11 (0.99, 1.26)	0.071	7,323	0.066
EOW	18	rs11872254	A/G	<i>DCC</i>	1.15 (1.10, 1.23)	2.30E-40	17,383	0.933	0.95 (0.85, 1.06)	0.37	7,355	0.307
PEW	6	rs6941858	G/C	<i>ZDC150N1A5B</i> <i>ABO2</i>	1.75 (1.41, 2.17)	3.48E-40	13,878	0.019	0.72 (0.38, 1.38)	0.33	6,642	0.323
CA	9	rs12554253	T/C	<i>PRKNG2</i> <i>ZDC150N1A5B</i>	1.35 (1.20, 1.51)	3.97E-40	15,874	0.749	1.05 (0.84, 1.28)	0.76	4,653	0.761
EOW	12	rs12867967	A/C	<i>APL11P1N</i> <i>SERPAP1</i>	1.13 (1.09, 1.22)	8.75E-40	17,383	0.012	1.08 (0.90, 1.10)	0.83	7,355	0.934
EOW	15	rs1896796	A/G	<i>SNPGL3</i> <i>ADAM10TSL3</i>	1.13 (1.09, 1.22)	1.22E-40	17,383	0.935	1.04 (0.92, 1.11)	0.79	7,355	0.793
CA	10	rs12767108	A/C	<i>SLC28A7</i>	1.24 (1.17, 1.44)	1.30E-40	15,874	0.954	1.06 (0.89, 1.27)	0.5	4,653	0.504
PEW	1	rs3752543	G/T	<i>CAMZB1</i>	1.57 (1.21, 1.96)	1.71E-40	13,878	0.102	1.08 (0.81, 1.23)	0.99	6,642	0.998
EOW	1	rs12882330	G/C	<i>SLC35D1</i>	1.50 (1.17, 1.85)	1.74E-40	17,383	0.479	0.97 (0.82, 1.46)	0.8	7,355	0.796
TEW	6	rs505060	G/T	<i>SLC22A7</i> <i>SLC22A7</i>	1.27 (1.15, 1.40)	3.28E-40	12,604	0.711	1.39 (0.93, 1.29)	0.25	7,323	0.098

Supplementary table 8. Top associated variants in 17q-21 locus across the four phenotypes.

r2	rs2305480	rs1008723	rs11078926	rs1054609
rs2305480	-	0.8477	0.8228	0.8308
rs1008723	-	-	0.8614	0.8308
rs11078926	-	-	-	0.9688
rs1054609	-	-	-	-

Supplementary table 9. For each phenotype, here is displayed the number of cases and controls and the number of suggestive ($P < 5e-5$) and genome-wide significant variants ($P < 5e-8$).

Phenotype	Cases	Controls	$P < 5e-8$	$5e-8 < P < 5e-5$
EOW	3273	14110	1	54
TEW	1338	11266	0	44
PW	977	12901	1	56
CA	2104	12970	1	55

Supplementary table 10. Results of associations between the *HGF*-variant and lung function measurements relevant for obstructive disease in UK Biobank.

SNP	EA/NEA	Lung function measurement	Log-odds	Std. Error	P-value	FDR corrected P-value	N
rs2214823	A/G	PEF	-0.0069	0.0025	0.00686	0.02058	341,818
		FEV1/FVC	-0.0041	0.0024	0.0917	0.13755	341,175
		FEV1	-0.0033	0.0024	0.158	0.158	341,175

Supplementary table 11. All variants in the wheeze GWAS, which showed at least suggestive association ($P < 5e-5$), with $r^2 < 0.01$ in a 1 Mb window.

See *STable11.TopLoci.xlsx*

Supplementary table 12. All candidate variants showing nominal significance ($P < 0.05$) in the wheeze phenotypes. They are all mapped to the closest gene.

See *STable12.NominalCandidateGenes.txt*

Supplementary table 13. Results of the genetic overlap analysis, for each trait and phenotype. Beta indicates the effect size of the overlap, and the standard error and p-value is associated with this. The column n.snps indicates the number of SNPs included in the analysis, which are all SNPs in a window of 1 Mb with an $r^2 < 0.05$. GSMR allows for residual correlation.

See *STable13.GeneticOverlapResults.xlsx*

Supplementary table 14. Overview of the SNPs including the in the genetic overlap analysis. The wheeze effect is shown in Z-scores since the FIZI-imputed summary results are used in this analysis to increase probability of overlap with the related traits.

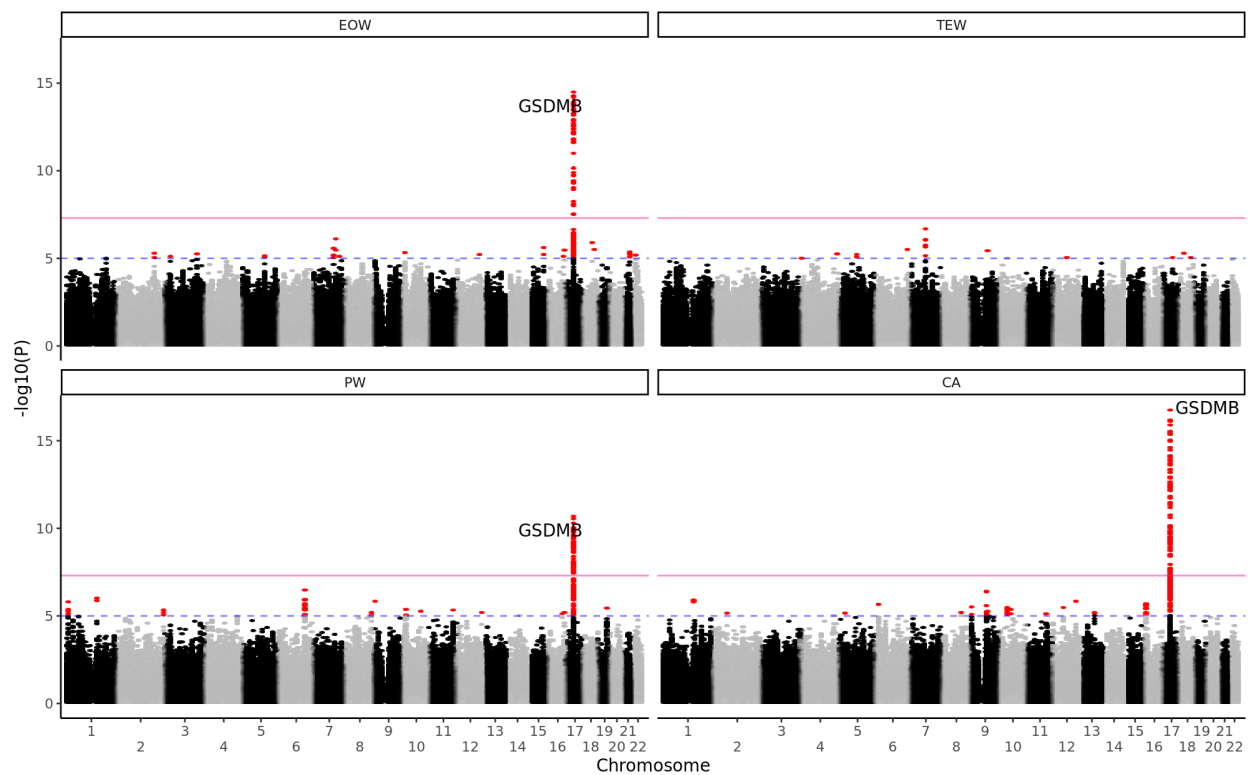
See *STable14.GeneticOverlapSNPs.xlsx*

Supplementary table 15. Heritability of the four wheeze phenotypes calculated with LD-score regression. Heritability is shown in the top row, followed by the lambda GC, mean chi square statistic and regression intercept.

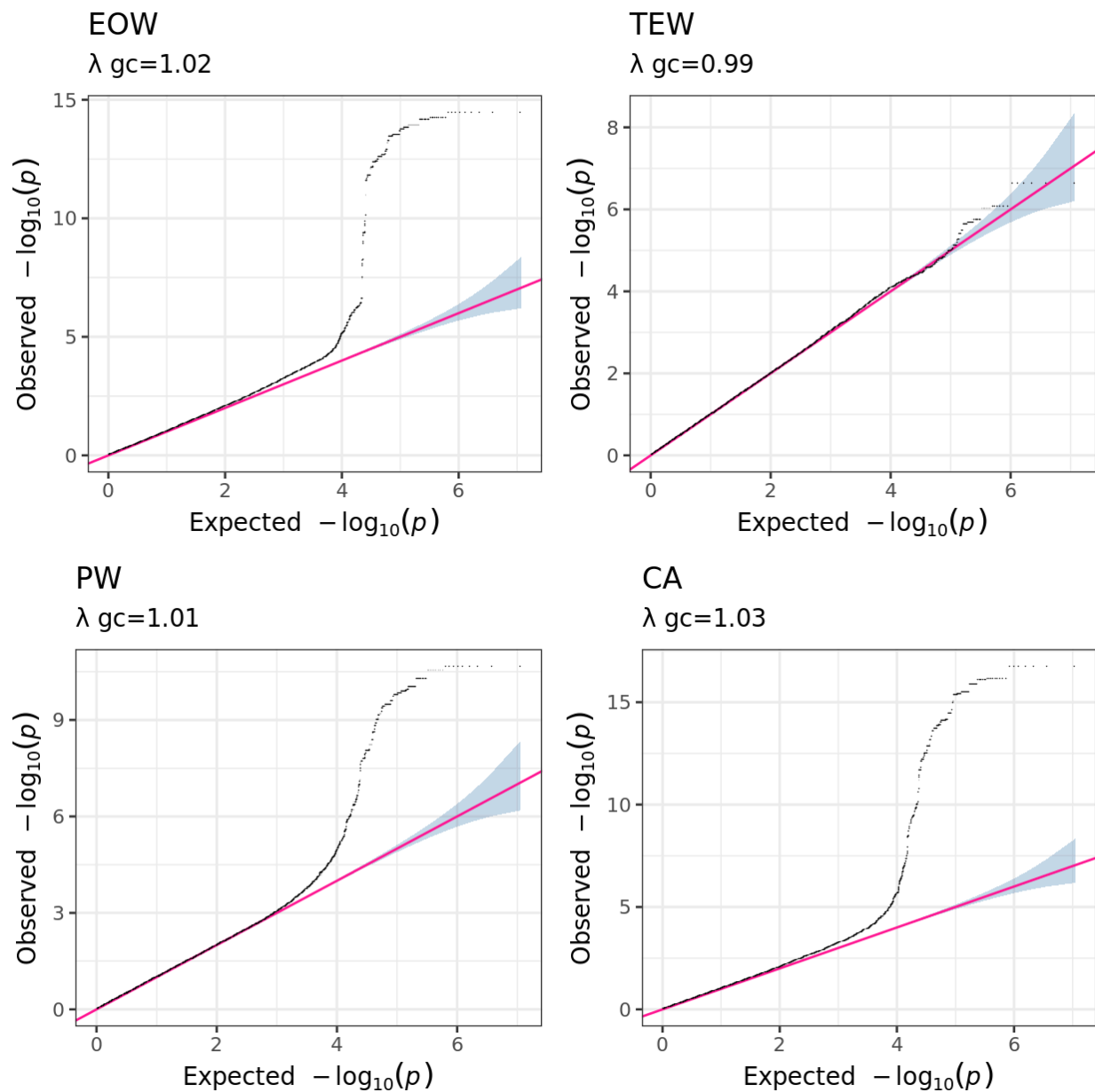
	EOW	TEW	PEW	CA
Heritability (SE)	0.347 (0.109)	-0.0032 (0.0969)	-0.0164 (0.1329)	0.0274 (0.0815)
λ GC	1.0195	0.9986	1.0016	1.0165
Mean chi2	1.0264	0.9983	1.0008	1.0215
Intercept (SE)	0.9923	0.9986	1.0021	1.0177

Supplementary figures

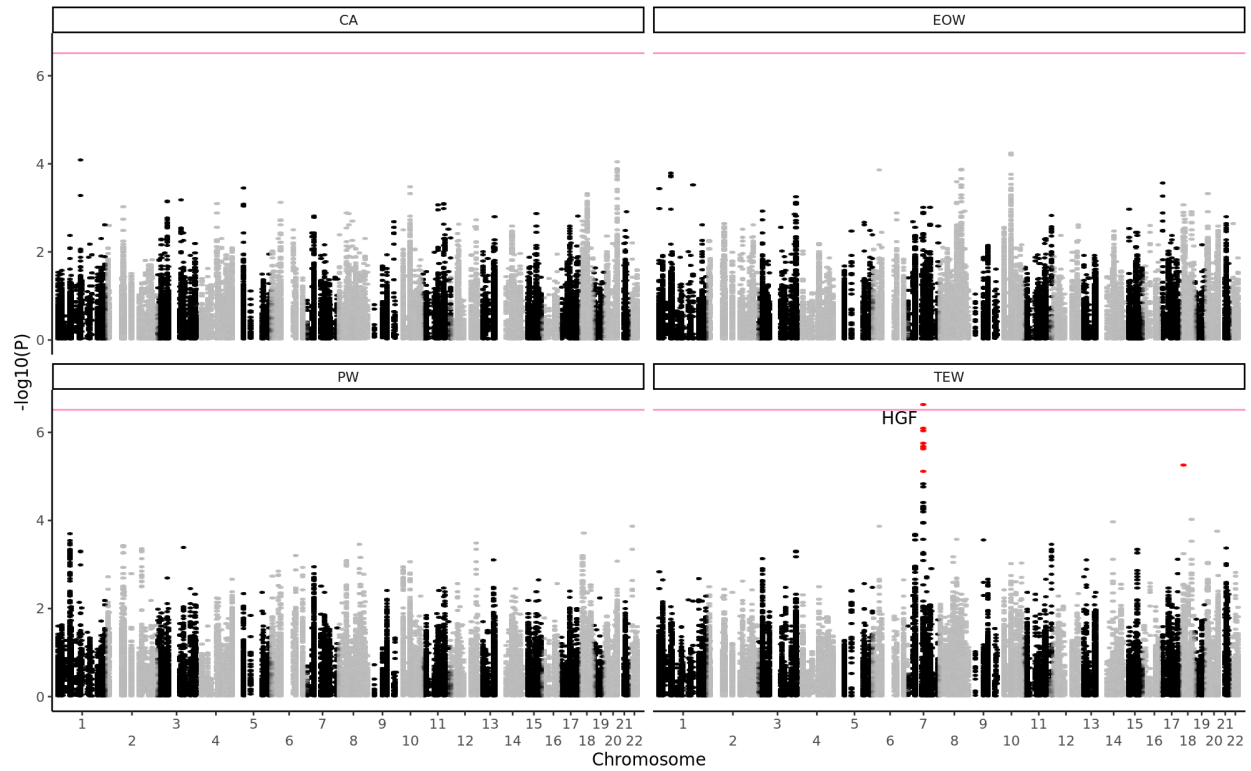
Supplementary figure 1. Association results of variants to the wheezing phenotypes. Pink solid line is the FDR line and blue-dashed is the nominal significance level. Top left going clockwise: EOW, TEW, PW & CA.



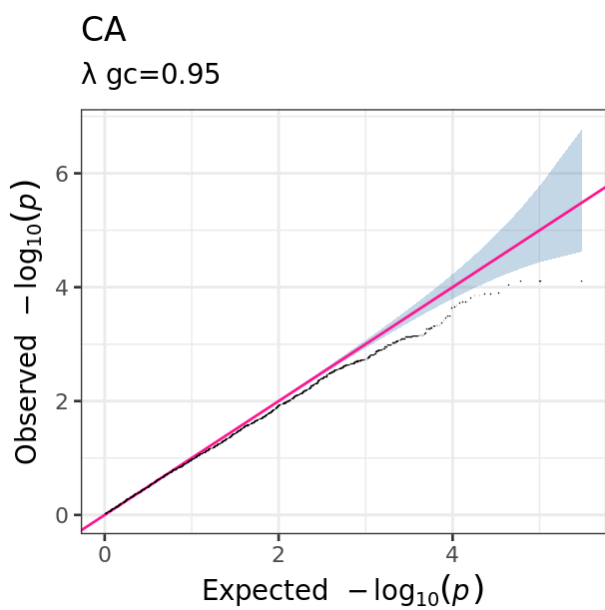
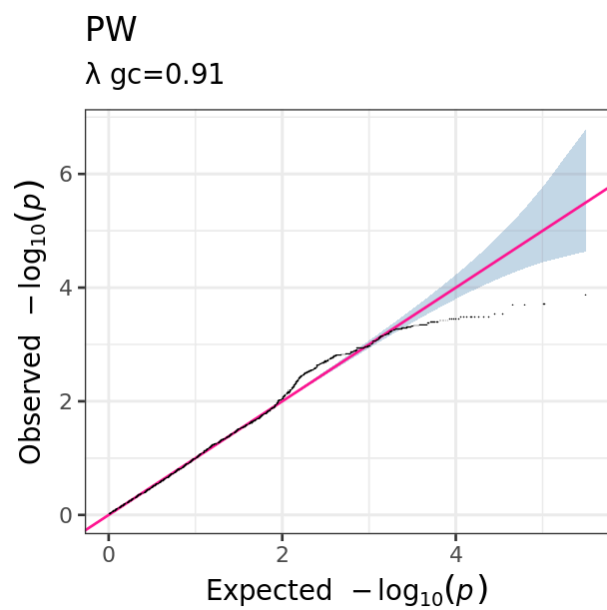
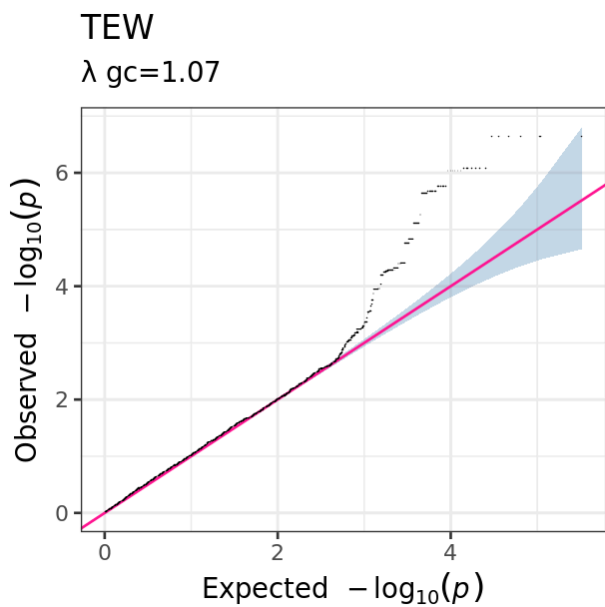
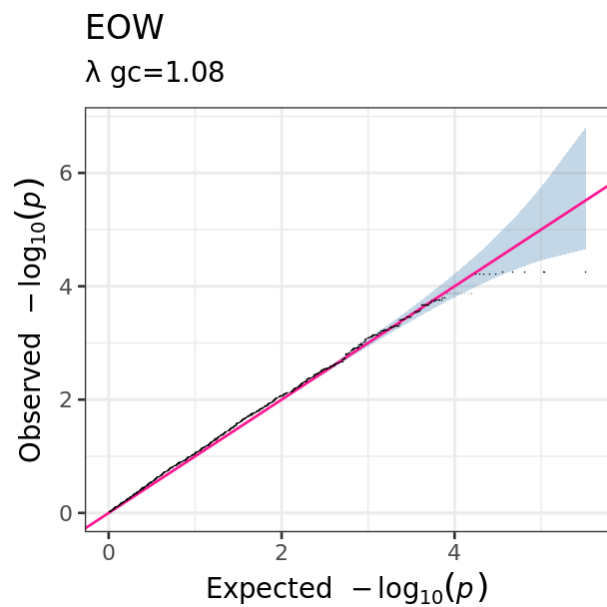
Supplementary figure 2. QQ-plots of the genome-wide association analysis, showing the expected vs. observed p-value distributions in the analysis. Top left going clockwise: EOW, TEW, PW & CA.



Supplementary figure 3. Association results of variants located within 10 kb from genes known to be differentially expressed during fetal lung development. Pink solid line is the FDR line and blue-dashed is the nominal significance level. Top left going clockwise: EOW, TEW, PW & CA.

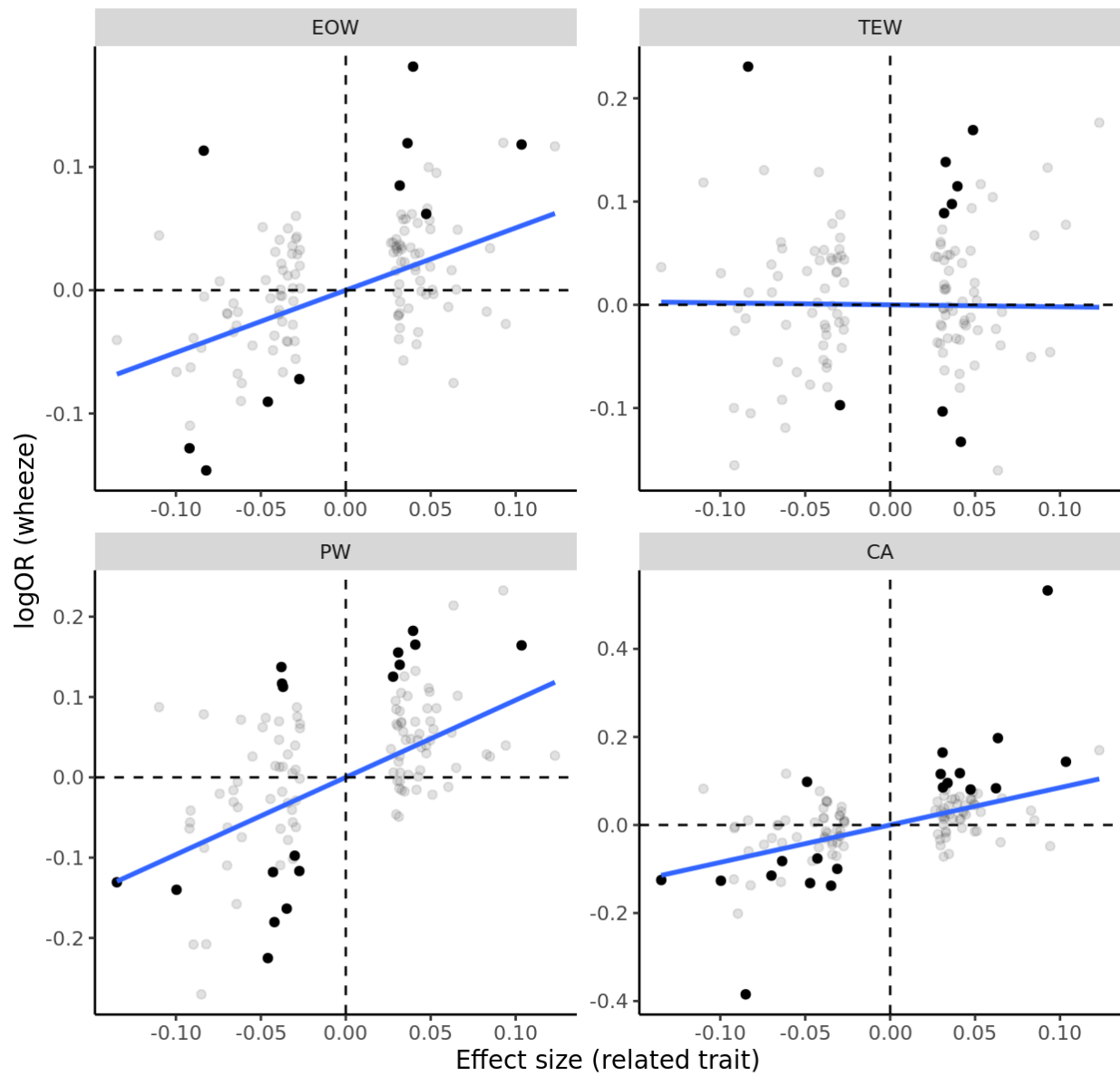


Supplementary figure 4. QQ-plots of the genome-wide association analysis of lung development genes, showing the expected vs. observed p-value distributions in the analysis. Top left going clockwise: EOW, TEW, PW & CA.

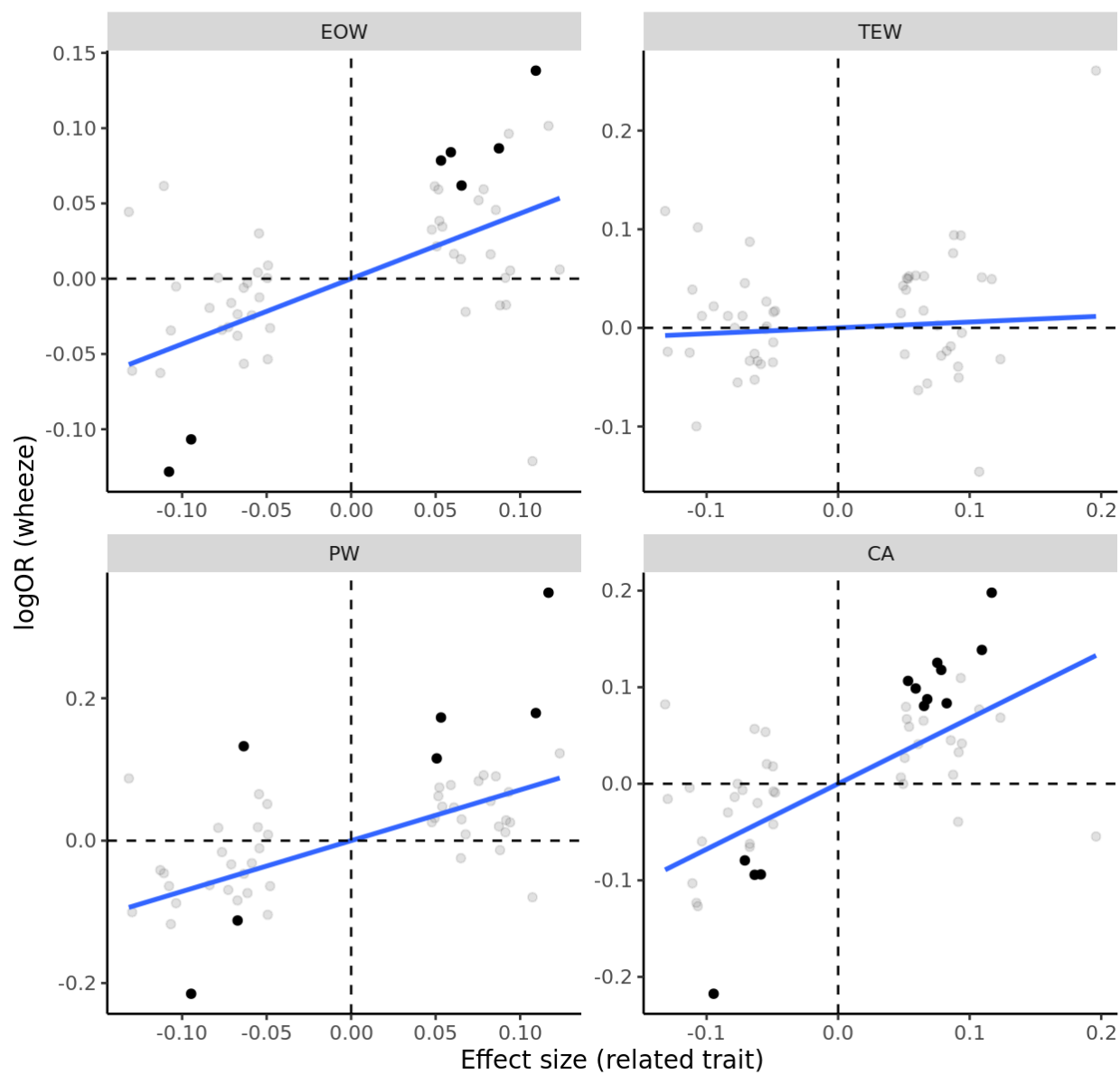


Supplementary figures 5-15. XY plots of SNP effects on related traits vs. SNP effects on wheeze phenotypes. Contains information on included SNPs, and their effects in related traits and investigated phenotypes as well as the closest gene. Black: $P < 0.05$ in wheeze GWAS. Grey: $P > 0.05$ in wheeze GWAS.

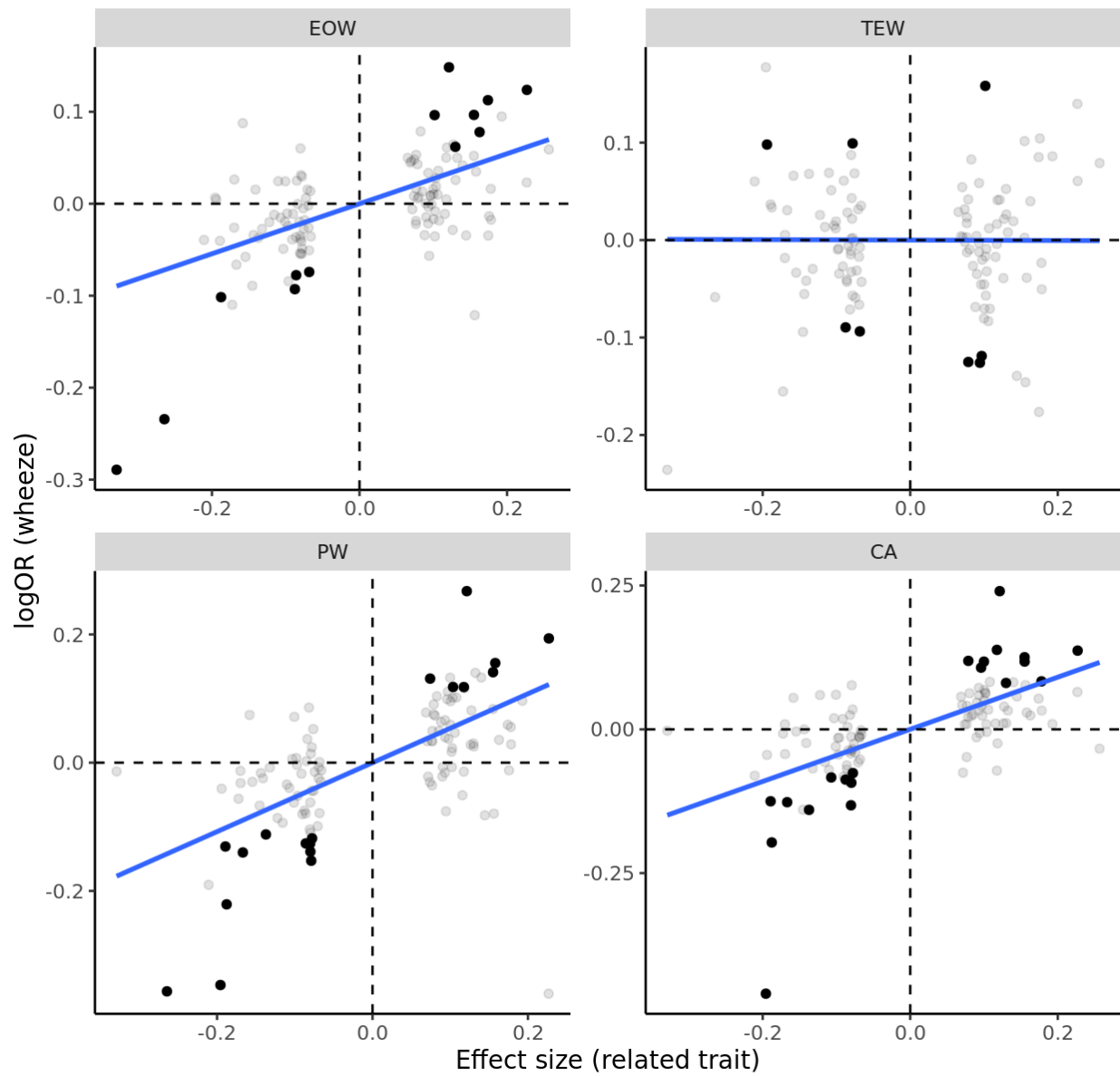
Asthma ever



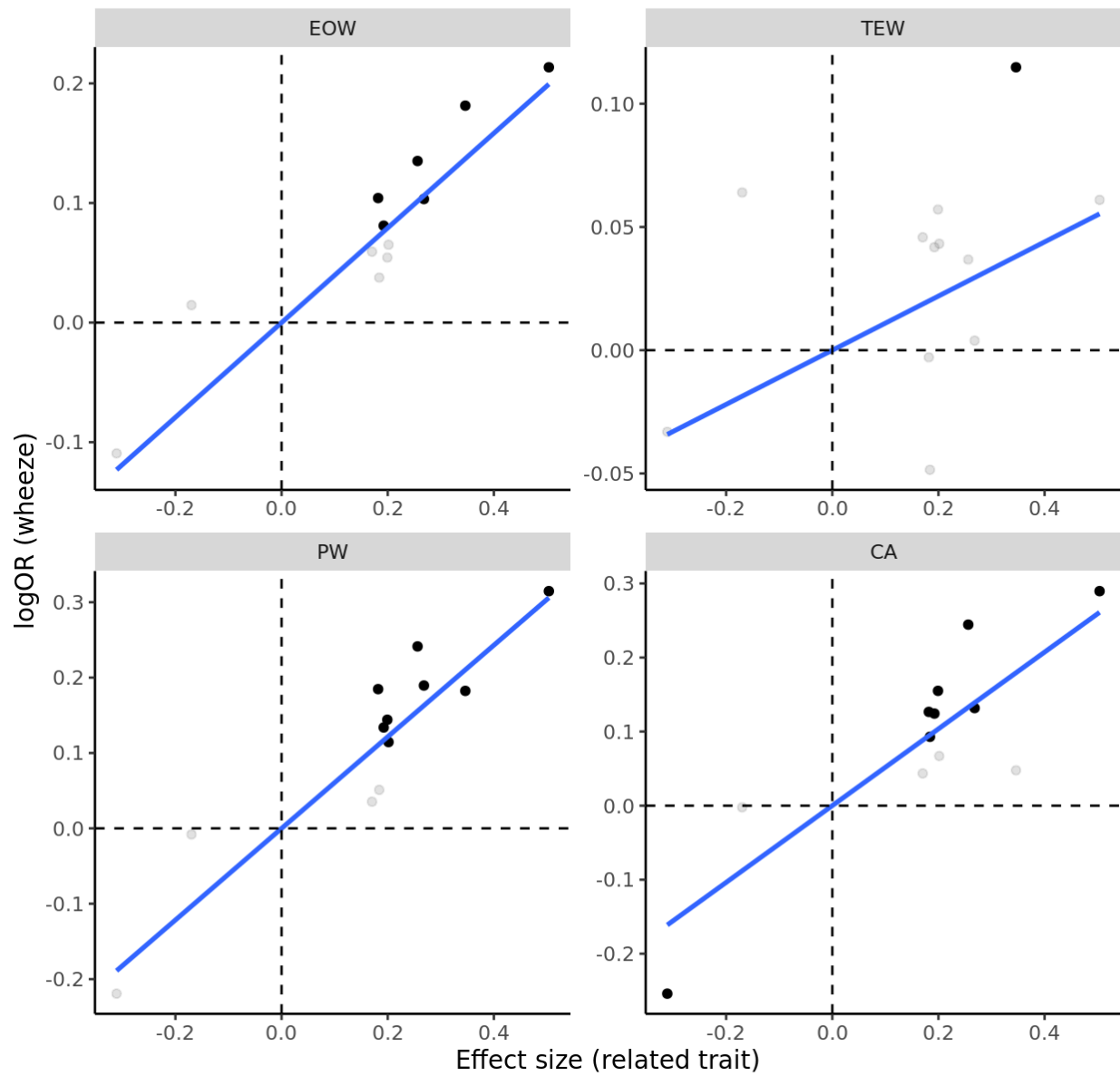
Adult onset asthma



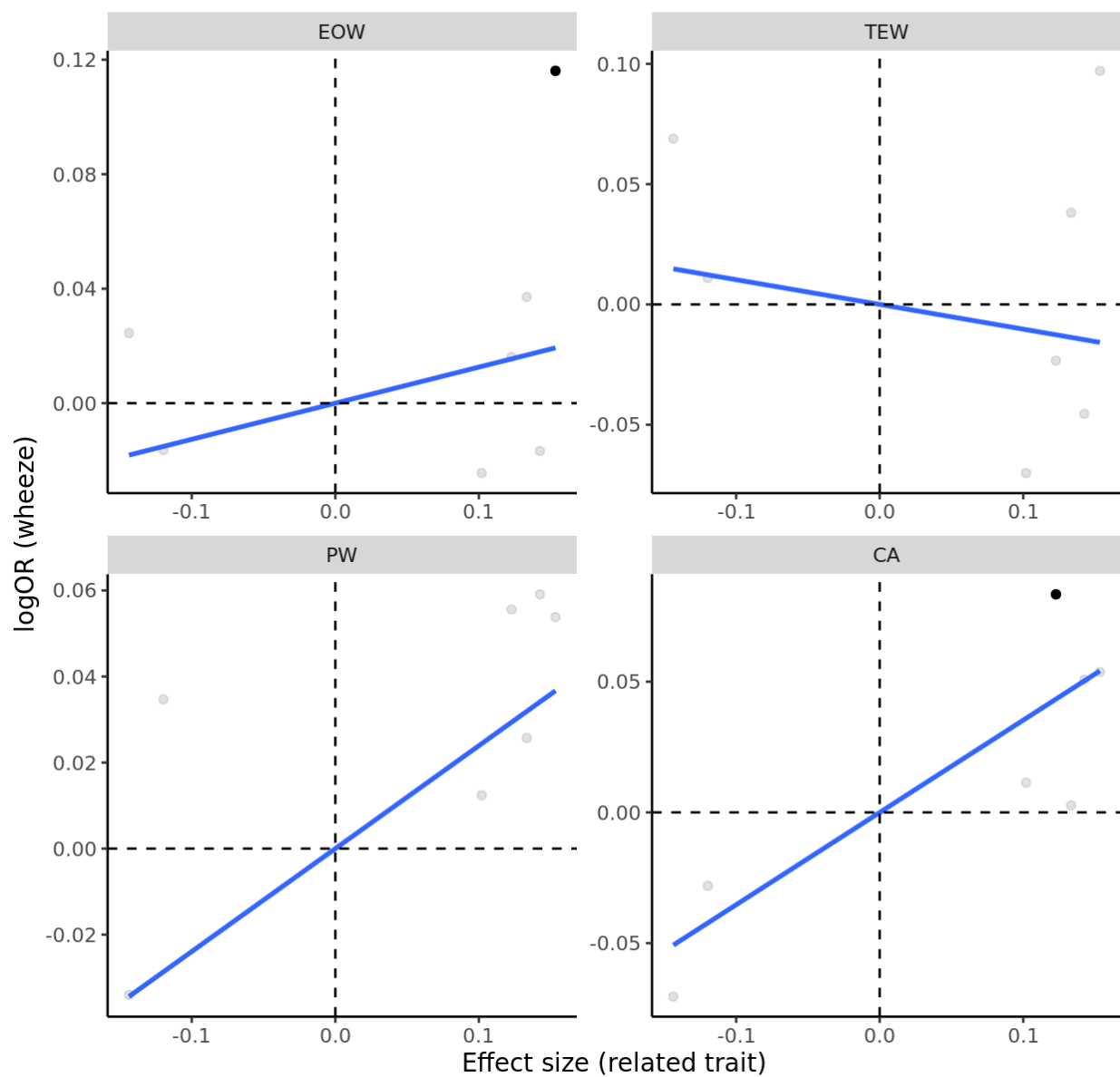
Childhood onset asthma



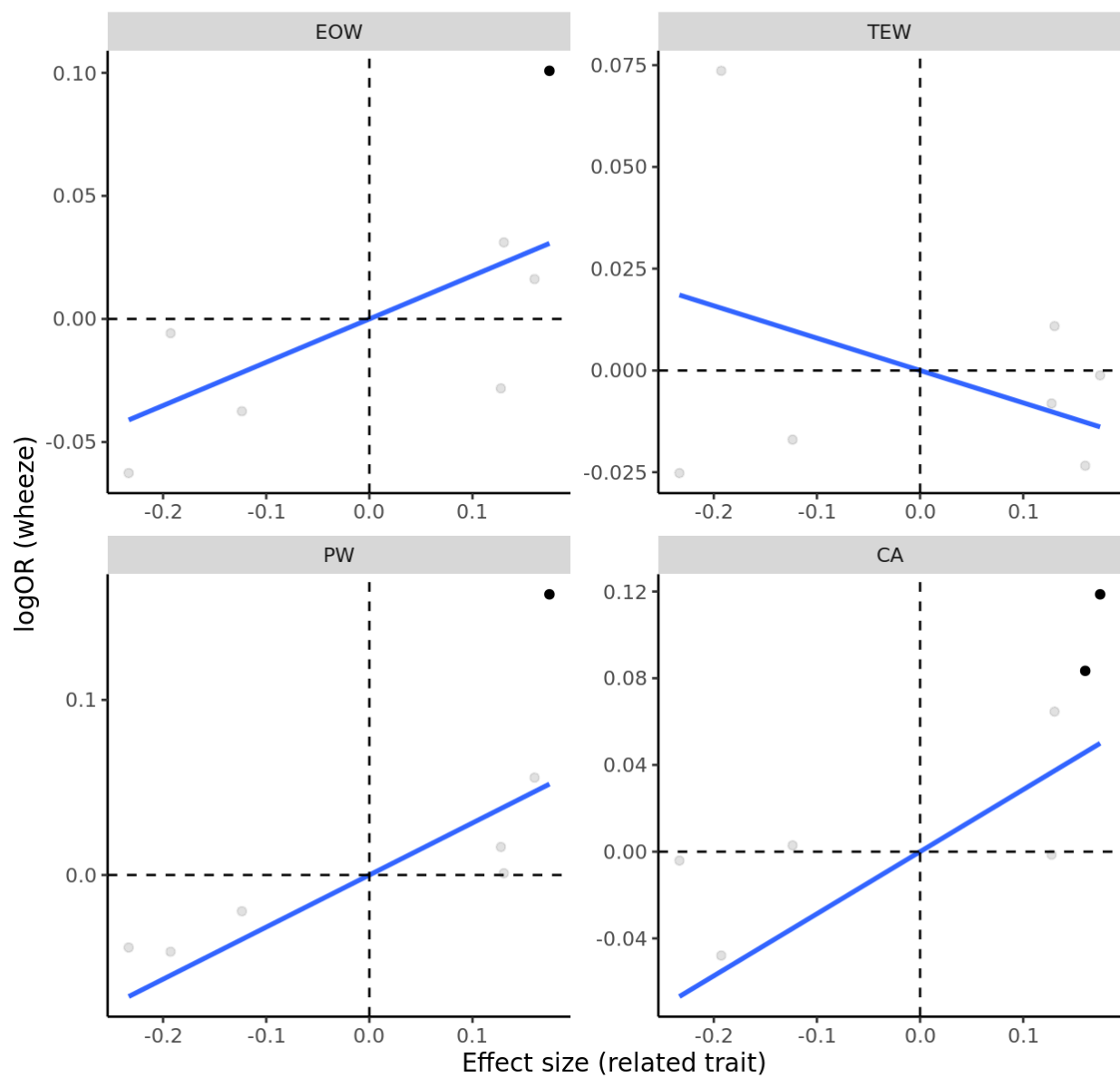
Asthma exacerbations



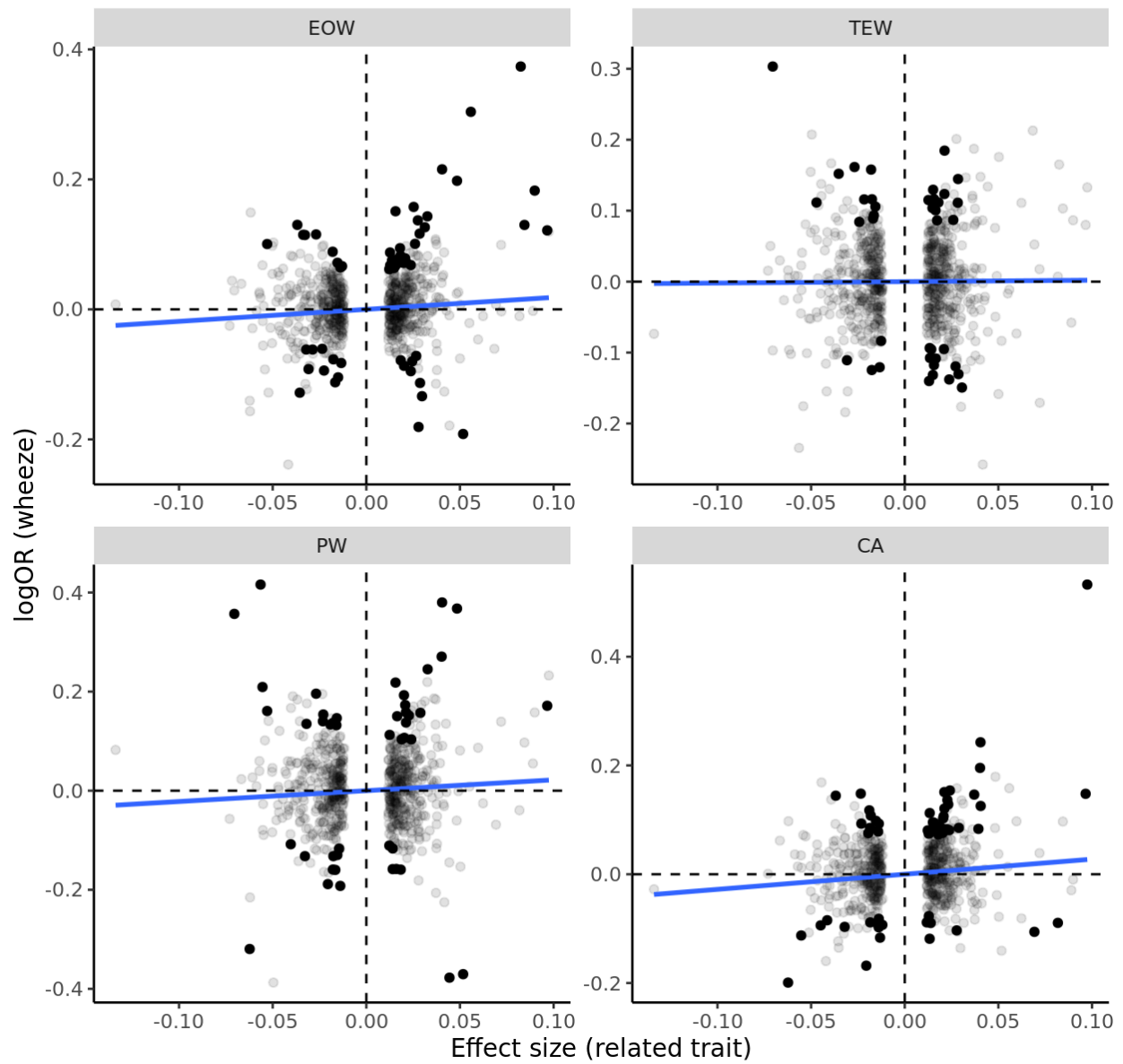
Atopic dermatitis

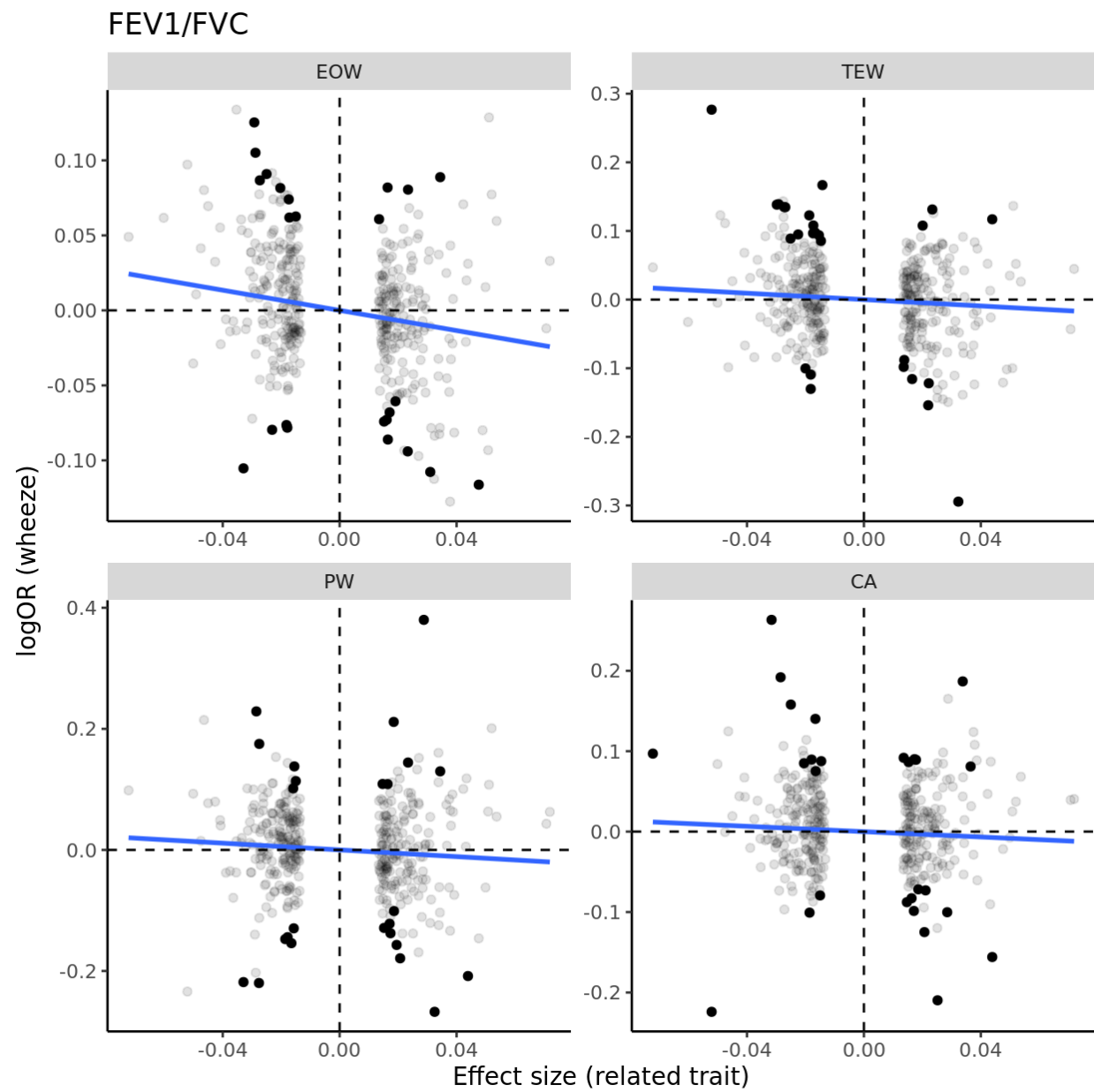


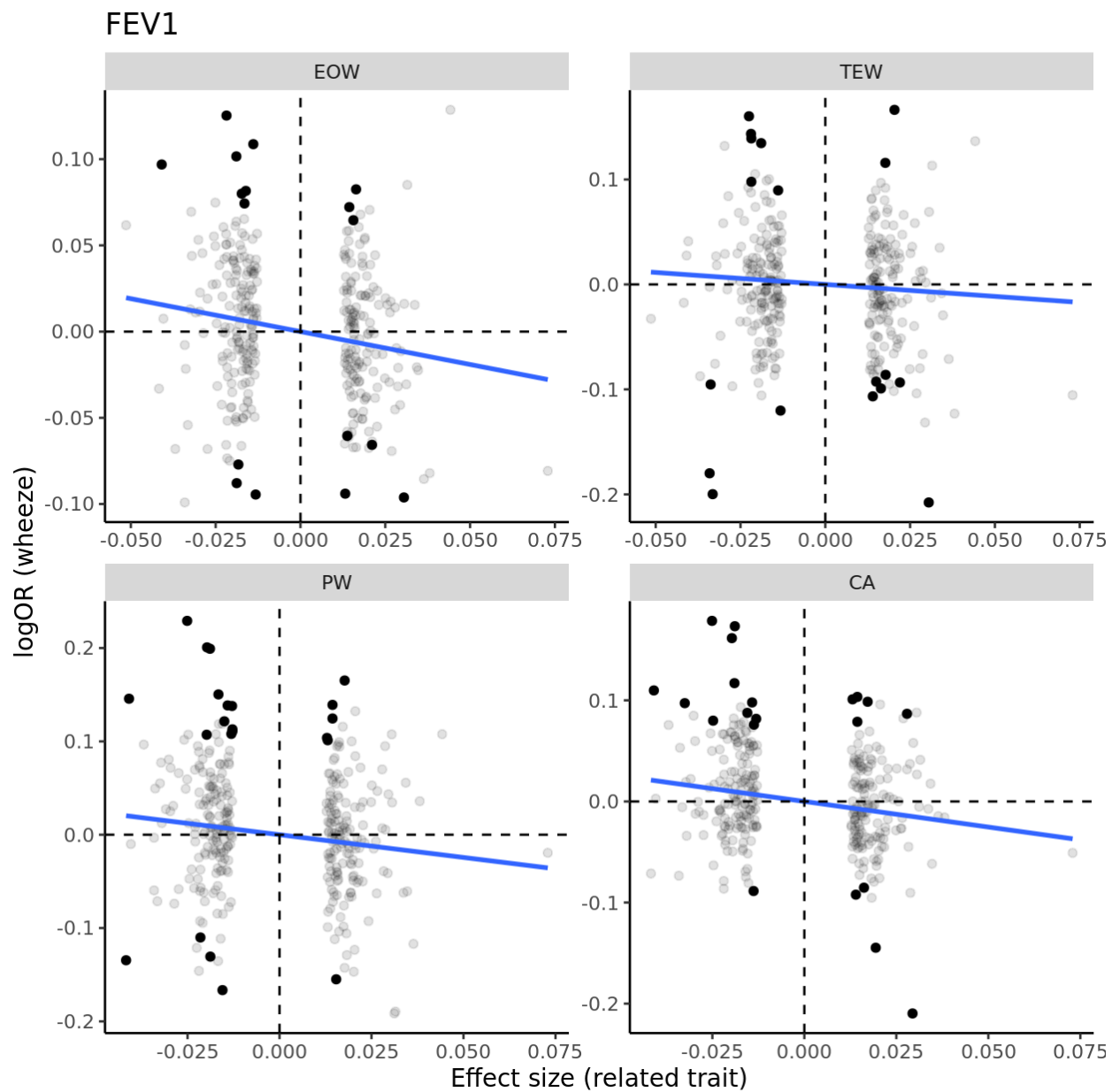
Allergic sensitization

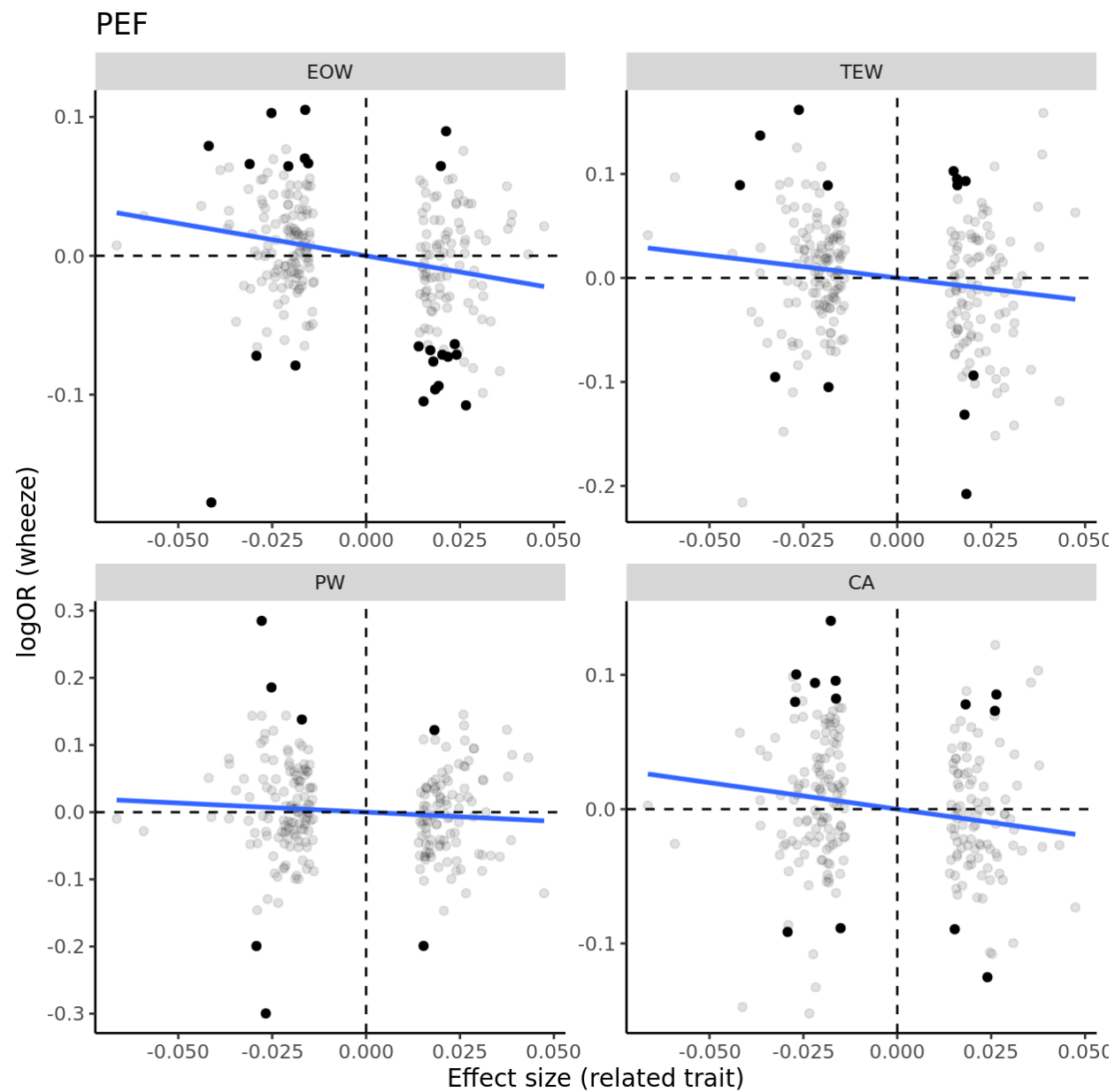


Eosinophil count









Supplementary code

Supplementary code 1. Command line parameters for the GSMR tool in the analysis of pleiotropic associations of related traits to wheezing phenotypes. The commands signify that SNPs with p-values less than $5e-8$ are taken forward in the analysis, but SNPs in LD with less than 0.05 are also included. The cutoff for the HEIDI test is set to zero, meaning that no variants will be filtered away. Chance correlations and mismatches in MAF between reference genome and summary stats are still considered using the `--diff-freq` and `--gsmr-ld-fdr` flags.

```
gcta64 \  
--bfile g1000_eur \  
--gsmr-file related_traits.txt wheeze_phenotypes.txt\  
--gsmr-direction 0 \ # Direction: related-trait -> wheeze  
--clump-r2 0.05 \ # Remove variants with  $r^2 > 0.05$   
--clump-kb 1000 \ # Remove variants within a 1MB window  
--gwas-thresh 5e-8 \ # Only include variants with  $P < 5e-8$  (related trait)  
--diff-freq 0.2 \ # Remove variants with large difference in MAF  
--heidi-thresh 0 \ # Turn off HEIDI pleiotropy-filter ( $P < 0$ )  
--gsmr-snp-min 3 \ # Only perform analysis if 10 variants pass criteria  
--gsmr-ld-fdr 0.05 \ # Account for chance correlations  
--out genetic_overlap \
```

References

1. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013 Feb 1;42(1):97–110.
2. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: The ‘Children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013 Feb;42(1):111–27.
3. Pasaniuc B, Zaitlen N, Shi H, Bhatia G, Gusev A, Pickrell J, et al. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*. 2014 Oct 15;30(20):2906–14.
4. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016 May;48(5):481–7.
5. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* [Internet]. Springer-Verlag New York; 2016 [cited 2022 Feb 10]. Available from: <https://ggplot2.tidyverse.org>
6. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun*. 2018 Jan 15;9(1):224.
7. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat Genet*. 2018 Aug;50(8):1072.
8. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019 Jun;7(6):509–22.
9. Ahluwalia TS, Eliassen AU, Sevelsted A, Pedersen CET, Stokholm J, Chawes B, et al. FUT2–ABO epistasis increases the risk of early childhood asthma and *Streptococcus pneumoniae* respiratory illnesses. *Nat Commun*. 2020 Dec 16;11:6398.
10. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet*. 2014 Jan;46(1):51–5.
11. Sliz E, Huilaja L, Pasanen A, Laisk T, Reimann E, Mägi R, et al. Uniting biobank resources reveals novel genetic pathways modulating susceptibility for atopic dermatitis. *J Allergy Clin Immunol*. 2022 Mar;149(3):1105–1112.e9.
12. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020 Sep 3;182(5):1214–1231.e11.
13. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019 Mar;51(3):481–93.
14. Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, et al. Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet*. 2011 Jul;7(7):e1002193.
15. Talaei M, Sdoná E, Calder PC, Jones LR, Emmett PM, Granell R, et al. Intake of n-3 polyunsaturated fatty acids in childhood, *FADS* genotype and incident asthma. *Eur Respir J*. 2021 Sep;58(3):2003633.
16. Bisgaard H, Stokholm J, Chawes BL, Vissing NH, Bjarnadóttir E, Schoos AMM, et al. Fish Oil–Derived Fatty Acids in Pregnancy and Wheeze and Asthma in Offspring. *N Engl J Med*.

2016 Dec 29;375(26):2530–9.

Paper II

Fischer-Rasmussen, K et al.

Genome-Wide Association Study of Severe Preschool Wheeze Uncovers Novel Insights into the Pathobiology

Genome-Wide Association Study of Severe Preschool Wheeze Uncovers Novel Insights into the Pathobiology

Kasper Fischer-Rasmussen¹, Casper-Emil Tingskov Pedersen¹, Luo Yang¹, Bo Chawes¹, Jakob Stokholm¹, Ann-Marie Malby Schoos¹, Per Magnus¹⁰, Yunsung Lee¹⁰, Marianna Niemi^{7,8}, Tarja Laitinen⁹, Þórunn Olafsdottir^{11,12}, Guðmar Þorleifsson¹², Ingileif Jonsdottir¹², Christian Erikstrup¹³, Lotte Aas Hindhede¹³, Carole Ober¹⁷, Zhaozhong Zhu¹⁴, Carlos A. Camargo¹⁴, Anders Børglum^{3,4,5}, Anders Gorm Pedersen², Anders Ulrik Eliassen^{1,15,16} & Klaus Bønnelykke¹

¹COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark

²Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

³The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, DK-8000 Aarhus, Denmark

⁴Center for Genomics and Personalized Medicine, Aarhus University, DK-8000 Aarhus, Denmark

⁵Department of Biomedicine, Aarhus University, DK-8000 Aarhus, Denmark

⁶Department for Congenital Disorders, Statens Serum Institut, DK-2300 Copenhagen, Denmark

⁷Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁸TAUCHI Research Center & Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁹Finnish Clinical Biobank Tampere, Tampere University and Tampere University Hospital, Tampere, Finland

¹⁰Norwegian Institute of Public Health, Oslo, Norway

¹¹Faculty of Medicine, Biomedical Center, University of Iceland, Reykjavik, Iceland

¹²deCODE Genetics/Amgen, Reykjavik, IC-101, Iceland

¹³Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark

¹⁴Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

¹⁵Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

¹⁶Department of Pediatrics, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA

¹⁷Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

Acknowledgements

We express our deepest gratitude to the children and families of the two COPSAC mother-child cohorts for all their support and commitment and acknowledge and appreciate the unique efforts of the COPSAC research team. We would like to acknowledge the huge work of late professor Hans Bisgaard, who was the founder of COPSAC and was head of the clinical research center for more than 25 years. Hans

was a dedicated, innovative physician-scientist who pushed the asthma research field forward. We also gratefully acknowledge the iPSYCH consortium for conducting the iPSYCH study and granting access to the iPSYCH data.

Abstract

Rationale

Severe wheeze is a predominant cause of hospitalization in early childhood and a risk factor for childhood asthma, yet the fundamental molecular mechanisms driving this condition remain largely elusive.

Objective

To illuminate the mechanisms underlying severe wheeze in preschool-aged children, uncovering novel disease pathways and identifying potential therapeutic targets.

Methods

We performed a meta-GWAS pooling data from 19,461 children who were hospitalized with severe wheeze before the age of seven, along with 284,631 controls, utilizing electronic health records across the Nordic countries.

Measurements & Main Results

Eight novel wheeze-loci were identified, including genes implicated in inflammation, diabetes, lung function, and growth, with five of these findings replicated in an independent cohort. Additionally, we propose the existence of SNP-clusters with unique influences on atopy, immune response, and lung function. Polygenic risk scores emphasize susceptibility to respiratory distress in infancy, highlighting a window for potential interventions for preventing asthma progression.

Conclusions

Severe wheeze is influenced by heterogeneous genetic mechanisms with pleiotropic effects to asthma-related comorbidities and symptoms of wheeze are already present in early infancy in genetically predisposed individuals.

Introduction

Preschool wheezing is a common clinical syndrome, characterized by acute episodes of obstructive airflow limitation, resulting in substantial morbidity, severe impact on quality of life and large health care expenditures (1). Severe disease requiring hospitalization is particularly common in the first years of life and is one of the most common causes of hospitalization in early childhood (2,3). In spite of the high prevalence and burden of the disease, the basic molecular causes of severe wheezing disorders in the first years of life are still largely undetermined. There is an overlap with early asthma, but only some of the children with severe wheeze later develop asthma, and the current inability to determine the underlying disease mechanisms challenges clinical decision-making and hampers the ability to optimally treat and prevent the disease (1–4). Viral respiratory infections are the most common trigger of acute wheezy episodes, but host factors making some children more prone to develop wheeze as a result of infections are hypothesized to play a detrimental role (5,6). Evidence for host susceptibility is underlined by studies showing that family history of asthma is a strong predictor of disease and twin studies estimate the heritability of wheezy symptoms and asthma to be approximately 60 % to 80 %, indicating an important role of genetic predisposition (7,8). Genome-wide association studies have in the past utilized definitions of asthma based on retrospective questionnaires with large sample sizes (9–12). We previously showed that focusing on a more severe and specific asthma phenotype requiring hospitalization in early life resulted in discovery of novel susceptibility loci and epistatic interactions that were not found in larger studies using more unspecific phenotypes (13–15).

Here, we wanted to shed light on the genetic mechanisms of severe wheezing disorders in preschool age to identify novel disease pathways and potential treatment targets. We conducted a meta-GWAS combining 19,461 children hospitalized with severe wheeze before age six, and 284,631 controls based on electronic health records from the Nordic countries. We applied a selection of computational methods to nominate putative causal genetic variants and prioritized genes and integrate our GWAS results with RNA expression and DNA methylation profiles to identify regulatory mechanisms associated with childhood asthma. Lastly, we calculated polygenic genetic risk scores (PRS) and evaluated their association with clinical endpoints and transcription levels in nasal tissues.

Methods

Registry based discovery cohorts

This genome-wide association study focuses on a severe phenotype defined by at least 1 acute hospitalization in the first 7 years of life due to a wheezing disorder, including bronchiolitis, bronchitis, asthmatic bronchitis, or asthma, by utilizing cohorts with genetic information coupled to public electronic health registries of inpatient hospitalizations.

The discovery meta-analysis included the following cohorts: iPSYCH (Denmark) (16,17), deCODE (Iceland) (18–20), FinnGen (Finland) (21) and the Danish Blood Donor Study (Denmark) (22,23).

Cases were defined based on at least one of the ICD-8 and ICD-10 codes listed in the supplementary material. Controls were defined as children with full follow-up in the first 6 years of life without any hospitalization for such diagnoses. An additional GWAS on the same diagnoses in the first 3 years of life was performed.

COPSAC clinical birth cohorts

The COPSAC₂₀₀₀ cohort is a single-center, prospective clinical study involving 411 children born to asthmatic mothers, aimed at investigating gene-environment interactions in the development of atopic diseases. Newborns were enrolled within the first month of life and attended follow-up visits at 1 month, every 6 months until age 7, then at ages 12 and 18, with additional visits scheduled for any respiratory or skin symptoms (24).

The COPSAC₂₀₁₀ cohort is a prospective study involving 700 children born to unselected mothers from Zealand, Denmark, between 2008 and 2011. Initial visits began at 1 week of age, followed by visits at 1, 3, 6, 12, 18, 24, 30, and 36 months, and yearly thereafter until age 10. Additional visits were arranged for any respiratory or skin symptoms (25). The COPSAC birth cohorts were used for post-GWAS analysis to characterize and understand the clinical impact on relevant childhood asthma phenotypes for the identified genetic variants and PRSs.

For definitions of asthma and wheeze in the COPSAC cohorts, see online supplement.

MARC-35 study

The 35th Multicenter Airway Research Collaboration (MARC-35) study, is a multicenter prospective cohort study of infants hospitalized for bronchiolitis (26). Emergency Medicine Network (EMNet) is an international research collaboration involving 247 hospitals and is the coordinator of the MARC-35 study. Investigators enrolled infants hospitalized with bronchiolitis at 17 sites across 14 U.S. states, following a standardized protocol during three consecutive bronchiolitis seasons (November 1 to April 30) from 2011 to 2014. Infants were enrolled at a median age of 3.2 months (27). Bronchiolitis diagnosis was based on the American Academy of Pediatrics guidelines (28) and inclusion and exclusion criteria have been described in detail elsewhere (29). For definitions of asthma and wheeze in the MARC35-study, see online supplement.

The MoBa cohort

The Norwegian Mother, Father, and Child Cohort Study (MoBa) enrolled more than 100,000 children at birth, 90,000 women and 70,000 fathers during a 10-year period from 50 hospitals across Norway. Participating parents were invited to answer highly detailed questionnaires acquiring detailed information of parents and children during ages 6 months until the child was above 19 years of age. The cohort, questionnaire and genotyping have previously been described in detail (30).

Association mapping

Plink2.0 was used to perform association analysis between genotype and phenotype (31) adjusting for gender and top 10 principal components (PCs) to account for population stratification (32). Inverse variance-weighted meta-analysis, implemented in METAL, was used to pool the summary results from each individual cohort (33).

Variants were included if they had a MAF > 0.05 % and were present in more than 3 cohorts to prioritize variants with robust support across studies. Heterogeneity test was performed to see if effect size varied more than expected across the six cohorts.

We subsequently used Susie to fine-map genome-wide significant loci to pinpoint putative causal variants (34) in the iPSYCH cohorts. We conducted in-sample fine-mapping in the iPSYCH cohort e.g. fine-mapping based on iPSYCH GWAS results with LD-reference using the same individuals, which has previously been shown to produce robust fine-mapping results (35). If a 95 % credible set could be computed, the variant with highest Posterior Inclusion Probability (PIP) was reported, otherwise we would report the lead variant. Summary statistics were mapped to assembly GRCh37/hg19 using the UCSC LiftOver (36).

A sensitivity analysis was performed to characterize the association of genomic risk loci ($P < 5e-8$), with each sub-diagnosis (see online supplement). Additionally, risk loci were investigated for associations with lower respiratory tract infections based on ICD codes for pneumonia, regardless of the suspected bacterial or viral cause (see online supplement). Children with any such diagnosis were considered cases, with the remaining individuals serving as controls.

Lastly, we investigated the epistatic role of *GSDMB* due to its well established strong association with early asthma and wheezing, robust replicability across heterogeneous studies and a previously demonstrated epistatic effect in relation to *CDHR3* (14). We regressed pairwise interactions between the *GSDMB* variant, and all meta-analysis genome-wide significant variants adjusted for marginal SNP-effects, gender, psychiatric status, cohort and top 10 principal components in the iPSYCH cohorts.

Heritability & genetic correlation

LD-score regression (37,38) was used to calculate the SNP-based heritability. Due to differences in ancestry among Danish, Icelandic and Finnish cohorts, we ran LD-score regression for each cohort separately to evaluate heritability contribution of each cohort to the meta-analysis. LD-score regression was subsequently used to estimate the shared heritability (genetic correlation - r_g) between related traits such as atopic dermatitis (39), allergic rhinitis and sensitization (40), immune cells (41), childhood and adult onset asthma (42), lung function measurements (43), childhood infections (44), BMI (45), height (46), C-reactive protein levels (47) and sphingomyelin measurements (48).

Gene prioritization

We mapped SNPs to genes using a consensus score based on 12 different SNP-to-gene features. We derived SNP-to-Gene distances from the Human genome of assembly GRCh37/hg19, using the NCBI RefSeq track (49). Intronic, exonic and splice variants in LD ($r^2 > 0.6$) with lead variants were used as additional three features and were mapped to GWAS summary statistics using ANNOVAR (50). The variant with the highest LD with the lead or fine-mapped variant was used to prioritize a gene. We utilized publicly available pathogenicity scores for missense variants, derived from AlphaMissense (51). Additionally, we used three gene prioritization tools; DEPICT, PoPS, and MAGMA as features for the consensus score (52–54). Additionally, we utilized TWAS and EWAS to prioritize genes that regulate expression and methylation levels in nasal epithelium using FUSION (55). We prioritized genes with TWAS and EWAS by identifying QTL variants that are in linkage disequilibrium ($r^2 > 0.5$) with a GWAS lead SNP and where the QTL and GWAS signals were colocalized. Lastly, we leveraged OpenTargets V2G (variant-to-gene) model to select the gene with the highest score based on RNA-, protein expression, enhancer-promoter interactions and variant effect predictor (VEP) (56) prioritizing the gene with the highest V2G-score. We assigned consensus scores to genes based on how many features prioritized each gene, selecting the gene with the highest score.

Drug target analysis

We investigated if potential drug targets were among the genes with high prioritization scores (upper 3rd quartile) by retrieving drug-target information from the Drug-Gene Interaction Database (57) (DGIdb, www.dgldb.org).

Association-based groupings of severe wheeze risk loci

We tested each of the risk variants identified in GWAS or TWAS for association with 27 binary and quantitative traits. We selected respiratory (43), atopic (39,40), asthma (9,42), infectious (44,58), immune traits (41), measurements (47,48) and anthropometric traits (45,46) suspected of involvement in preschool wheeze. To reduce noise, only traits with at least one wheeze SNP showing genome-wide significance for the trait ($P < 5e-8$) were included. All estimates in this analysis refer to the allele conferring risk of severe wheeze. All estimates were converted to z-scores and if a risk variant could not be determined, a proxy was used with $r^2 > 0.6$. Results were clustered using hierarchical clustering, with the number of splits determined by the k-means model that produced the highest silhouette width and most interpretable clusters. Results were visualized using the R-package *ComplexHeatmap* (59,60).

Replication

We replicated findings from the GWAS and TWAS in the independent MoBa cohort from Norway ensuring no sample overlap between our discovery cohort, which contains individuals from Denmark, Iceland and Finland. The replication phenotype was defined as a diagnosis of asthma, bronchitis and/or wheeze from 6-months to 7-years. See

supplementary for phenotype description. For description of all novel loci, regardless of replication results, see Supplementary material.

Expression and methylation data from nasal epithelial cells

RNA expression and DNA methylation from airway (nasal) epithelium were obtained from nasal scrapings collected at the age of 6 from participants in the COPSAC₂₀₁₀ cohort. The methods for data acquisition and handling, including sample collection, extraction, library preparation, and sequencing protocols, have been previously documented in detail (61).

Additionally, we included RNA sequencing data from airway (nasal) epithelium from the MARC-35 cohort, which consists of infants admitted for bronchiolitis. Nasal epithelial samples for RNA sequencing were collected within 24 hours of hospitalization due to bronchiolitis. Data acquisition process includes nasal sampling, RNA isolation, and sequencing methodologies in the MARC-35 cohort, as previously described in detail (62–64).

Transcriptome-wide association analysis

We use TWAS to characterize regulatory mechanisms of the genomic profile, and to provide further evidence for suggestive loci that might have been genome-wide significant with higher sample size (65). FUSION was used to compute predictive models based on cis-eQTL effect sizes (55) derived from COPSAC₂₀₁₀ individuals with both genotypes and expression data (n = 354). Genes that passed correction for multiple testing (Benjamini-Hochberg, FDR < 0.05) were validated by colocalization

analysis to compute the posterior probability of a shared GWAS and eQTL signal (66). Prior to running the TWAS, summary statistics were imputed with the ImpG algorithm to maximize overlap with QTL-models (67). We used the R-package *coloc* for colocalization analysis (66). A TWAS association was considered if the gene passed correction for multiple testing and colocalized with a probability of a shared variant above 75 %. We also used FUSION to perform Epigenome-wide association study (EWAS) utilizing the methylome of the airway (nasal) epithelium. In this analysis we utilized a colocalization cutoff of 95 % given the larger number of tests compared to the TWAS. EWAS was only included as a method in the gene prioritization scheme and not interpreted independently.

Differential expression analysis and functional enrichment

To estimate transcriptomic fingerprint of high-risk individuals we regressed RNA levels on the standardized PRS using limma (68). Models in COPSAC₂₀₁₀ were adjusted for Bioanalyzer RNA concentration, RNA-handling site, RNA-integrity number (RIN), Gender, age and top 5 PC's. MARC-35 models were adjusted for age, gender and top 2 PC's. We subsequently performed pathway/GO term enrichment by including all genes with nominal p-values ($P < 0.05$). Enriched pathways and GO terms were identified using enrichR (69–71). This analysis was performed both in the COPSAC cohorts and the MARC-35 cohort.

Clinical & transcriptomic fingerprint of genetic risk score

Polygenic risk scores (PRS) for severe wheeze were calculated using PRS-CS, inferring posterior effect sizes of all SNPs using a shrinkage prior based on the GWAS effect sizes (72). Posterior effect sizes were summarized to a polygenic risk score (PRS) in the COPSAC cohorts and the MARC-35 cohort using plink (73). A phenome-wide association study (PheWAS) was conducted in the COPSAC cohorts to identify clinical and para-clinical endpoints associated with the PRS; we included 944 children from both cohorts in this analysis and 354 had RNA sequencing data in COPSAC₂₀₁₀. In the MARC-35 cohort, we looked at associations between wheeze-PRS and asthma at age six, infant breathing problems and rhinovirus infections in 326 infants of European descent from the MARC-35 and 131 with RNA sequencing data (see Supplementary Material).

We explored the ability of the severe wheeze PRS to predict wheeze in the MoBa cohort in comparison to three large GWAS of asthma (74,75) or childhood asthma (12). We calculated AUC values between the four PRS's and the wheeze phenotype defined in the replication analysis to evaluate if the severe wheeze GWAS captures additional genetic information in relation to wheeze compared to large-scale GWAS. We used DeLong's test to evaluate significant differences between the PRS's (76).

Results

GWAS results

Main analysis (0-7 years)

We included 19,461 children with severe wheeze and 284,631 controls and a total of 23 loci reached genome-wide significance in the meta-analysis ($P < 5e-8$), 8 of which were novel (see Figure 1, Table 1 and Supplementary Figure 1). It was possible to calculate 95 % credible at 10 out of 23 loci by in-sample fine-mapping in the iPSYCH cohorts (see Supplementary Table 2).

The novel loci in the GWAS analysis were *CASC20* (OR = 1.08, (95% CI = 1.06 - 1.11), $P = 1.85E-11$), *GDF5* (OR = 1.08, (95% CI = 1.05 - 1.11), $P = 1.48E-10$), *CASC17* (OR = 1.07, (95% CI = 1.05 - 1.1), $P = 2.38E-09$), *NR4A2* (OR = 1.11, (95% CI = 1.07 - 1.15), $P = 4.45E-09$), *THSD4* (OR = 1.09, (95% CI = 1.06 - 1.12), $P = 2.53E-08$), *KAZN-AS1* (OR = 1.08, (95% CI = 1.05 - 1.11), $P = 3.18E-08$), *ZNF529* (OR = 1.07, (95% CI = 1.05 - 1.1), $P = 3.20E-08$) and *HNF1B* (OR = 1.06, (95% CI = 1.04 - 1.09), $P = 1.99E-07$) (see Supplementary Figure 2).

Five out of seven novel loci replicated in terms of association with preschool wheeze in the MoBa cohort including 19,621 cases and 29,134 controls (see Supplementary). The replicating loci were *GDF5* (OR = 1.06, (95% CI = 1.03 - 1.09), $P = 2.25E-05$), *NR4A2* (OR = 1.08, (95% CI = 1.04 - 1.13), $P = 1.35E-04$), *CASC17* (OR = 1.05, (95% CI = 1.03 - 1.8), $P = 1.77E-04$), *THSD4* (OR = 1.07, (95% CI = 1.03 - 1.1), $P = 4.27E-04$) and *CASC20* (OR = 1.08, (95% CI = 1.02 - 1.08), $P = 1.13E-03$) (see Table 1).

Secondary analysis (0-3 years)

We include 13,433 children hospitalized with wheeze before age 3 and 275,171 controls. A total of 12 loci reached genome-wide significance in the meta-GWAS of severe wheeze with onset before age three ($P < 5e-8$). All of these were also identified in the 0-6 years analysis, and scatter plots indicate that none of the loci were phenotype-specific (see Supplementary Figure 3). Additionally, most iPSYCH individuals had received their first wheeze diagnoses before the age of three (see Supplementary Figure 4).

A total of 5016 children had received an asthma diagnosis and 5309 had been diagnosed with bronchitis, whereas only 415 had a diagnosis with bronchiolitis (see Supplementary Figure 5) The variants *GSDMB* and *CDHR3* were strongly associated with both asthma, bronchitis and respiratory infections. *KAZN-AS1*, *MTA3*, *NPNT*, *HNF1B*, *CASC17* and *RHPN2* were only associated with bronchitis and not asthma (see Supplementary Figure 6).

Heritability and genetic correlations

The heritability for the 0–7-year analysis was estimated to be 12.66 % and in each cohort; iPSYCH12, iPSYCH15, FinnGen, DBDS and deCODE had heritability estimates of 17.85 %, 13.15 %, 18.22 %, 18.19 % and 5.19 % respectively. iPSYCH15, FinnGen, DBDS and deCODE all had intercepts above 1, whereas iPSYCH12 had an intercept of 0.99. The meta-analysis intercept was 0.98 (see Supplementary Table 5). We observed significant positive genetic correlations between wheeze and asthma 0-15 years, asthma 25-65 years, asthma ever, atopic dermatitis, allergic rhinitis & sensitization,

eosinophil count, BMI and C-reactive protein levels in adults. We observed significant negative correlations for height, FEV1, FVC, FEV1/FVC and PEF. Adjustments for multiple testing were made using the Benjamini-Hochberg procedure ($FDR < 0.05$) (see Figure 2 and Supplementary Table 6).

The heritability for the 0–3-year analysis was estimated to be 12.62 % and in each cohort; iPSYCH12, iPSYCH15, FinnGen and DBDS had heritability estimates of 20.6 %, 12.47 %, 15.53 % and 16.17 % respectively. iPSYCH15, FinnGen DBDS all had intercepts above 1, whereas iPSYCH12 had an intercept of 0.99. The meta-analysis intercept was 0.98. We observed similar genetic correlations as in the 0-7-year analysis (see Supplementary Table 7).

Epistatic effects with *GSDMB* were observed for *CDHR3* ($OR = 1.15$, (95% $CI = 1.09 - 1.22$), $P = 1.6e-6$), *ID4* ($OR = 0.91$, (95% $CI = 0.85 - 0.97$), $P = 9.9e-3$) and *NPNT* ($OR = 1.08$, (95% $CI = 1.02 - 1.13$), $P = 1.2e-2$) (see Supplementary Figure 5). *ID4* mediated the strongest wheeze risk in the absence of *GSDMB* risk alleles and *NPNT* was only associated with wheeze in the presence of *GSDMB* risk alleles (see Supplementary Table 8). There was some evidence of interaction with psychiatric case status and *MTA3* only conferred wheeze risk among individuals without a psychiatric diagnosis ($OR = 1.13$, ($CI = 1.06-1.21$), $P = 1.4e-4$), compared to those with any psychiatric diagnose ($OR = 1.01$, ($CI = 0.96-1.05$), $P = 0.73$). The effect of psychiatric status on severe wheeze was not influenced by *MTA3* genotype (see Supplementary Table 9).

Prioritization of causal genes at genomic risk loci

We prioritized 62 genes across the 23 loci. The highest scoring genes were *GSDMB* and *CDHR3* which were prioritized by 7 out of 12 features. Genes scoring above the upper 3rd quartile were prioritized by at least 4 features and these genes were *GSDMB*, *CDHR3*, *IL-13*, *HNFB1B*, *THSD4*, *NPNT*, *CLEC16A*, *FUT2* & *GDF5* in descending order. Among the 8 genes in the upper 3rd quartile, 3 genes were assigned to novel loci namely *HNFB1B*, *THSD4* and *GDF5* (see Table 2).

Drug target analysis

Among the confidently prioritized genes (*GSDMB*, *CDHR3*, *IL-13*, *HNFB1B*, *THSD4*, *NPNT*, *CLEC16A*, *FUT2* & *GDF5*) only *IL-13*, *HNFB1B*, *FUT2* and *CLEC16A* had any reported gene-drug interactions in DGIdb, displaying 12, 2, 1 and 1 drug-interaction respectively. The novel gene *HNFB1B* had two drugs with reported interactions namely platinum and metformin with interaction scores of 2.62 and 1.34 respectively. Platinum is currently under investigation for treatment of various cancer types and metformin is a known antidiabetic drug (see Supplementary Table 10).

Association-based groupings of severe wheeze risk loci

Clustering analysis identified seven distinct terminal branches of the hierarchical dendrogram representing SNPs and five representing traits (see Supplementary Table 11). Three branches contained more than one SNP and were designated SNP-clusters; branches containing a single SNP were *GSDMB*, *HCG22*, *FUT2* & *ABO*. One SNP-cluster (cluster 2) was characterized by low adult lung function and weak associations to the asthma-traits. Another cluster (cluster 7) was characterized by high

risk of atopic diseases (asthma, allergy and eczema), strong associations to eosinophils, and slightly elevated counts of other immune cells. The last SNP-cluster (cluster 4) was characterized by a heterogeneous phenotypic profile. Two variants in cluster 4 (*GDF5* and *CASC20*) were strongly associated with height, but did not share any other association patterns (see Figure 3, Supplementary Figure 7 & Supplementary Table 12).

Transcriptome-wide association analysis

A total of 118 genes passed correction for multiple testing in the TWAS, and among those *MTA3*, *IQCB1*, *RNF5*, *HLA-DRB5*, *HLA-DRB9*, *AC003075.4*, *OAS1/OAS3*, *GSDMB*, *RHPN2*, *SYMPK*, *CYTH2*, *FUT2* and *UQCC1*, had a probability of a shared variant of above 75 % (see Supplementary Figure 8). For three of these genes, the shared GWAS and eQTL variants in *MTA3* (OR = 1.06, (95% CI = 1.04 - 1.09), P = 1.84E-06), *OAS1/OAS3* (OR = 1.06, (95% CI = 1.03 - 1.08), P = 8.02E-06) and *RHPN2* (OR = 1.1, (95% CI = 1.06 - 1.14), P = 1.06E-07) had never been associated with childhood asthma or asthma before. Two out of three novel TWAS loci replicated with borderline significance in the MoBa cohort; *MTA3* (OR = 1.03, (95% CI = 1.0 - 1.06), P = 4.03E-02) & *OAS1/OAS3* (OR = 1.03, (95% CI = 1.0 - 1.06), P = 4.76E-02) (see Supplementary Table 13).

Clinical & transcriptomic fingerprint of genetic risk score

We found a strong association between the severe wheeze PRS and asthma, acute severe wheeze, infant respiratory infections, fever episodes and rhinovirus respiratory

illnesses in the COPSAC cohorts (See Figure 4A). Furthermore, the PRS was associated with genes expressed in the nasal (epithelium) enriched in pathways involved in fc gamma receptor (fcg) signaling, toll-like receptors cascades (TLR4, TLR7, TLR8 & TLR9), *S. aureus* and other bacterial and parasitic infections. The most strongly enriched functional terms were FCGR4A-mediated IL-10 synthesis (OR = 7.01, $4.3e-4$), Leishmaniasis (OR = 6.97, $P = 7.5e-7$) and TP53 Regulates Transcription Of Cell Death Genes (OR = 5.9, $P = 9.72e-4$) (see Figure 4B).

Only *HLA-DQB1* was significantly associated with the PRS in the COPSAC₂₀₁₀ RNA-expression data after correction for multiple testing (see Supplementary Figure 9).

The PRS was also associated with asthma at age six among bronchiolitis patients (OR = 1.49, (95% CI = 1.09 - 2.02), $P = 0.011$) and the number of breathing problems in infancy before index hospitalization in the MARC-35 cohort (IRR = 1.46, (95% CI = 1.15 - 1.88), $P = 2.2e-3$). The association between rhinovirus subtypes displayed a positive association with a trend towards significance. Functional enrichment of PRS associated nasal (epithelium) genes in the MARC-35 cohort revealed enrichment in terms related to viral infections and inflammatory signaling cascades. The most strongly enriched terms were interferon alpha response (OR = 8.07, $P = 1.5e-5$), interferon gamma response (OR = 6.9, $P = 7.2e-8$) and MAPK Signaling Pathway (OR = 3.81, $P = 5.27e-4$) (see Supplementary Figure 10).

Severe wheeze, Pividori et al. (2019), Namjou et al. (2022), and Tsuo et al. (2022) reported AUC values for predicting wheeze in MoBa of 0.561, 0.537, 0.536, and 0.551,

respectively. DeLong's test indicated that the severe wheeze PRS yielded significantly higher AUC values compared to the other scores (see Supplementary Table 14).

Discussion

This study represents the first GWAS of severe preschool wheeze requiring hospitalization and the largest GWAS on any preschool wheezing/asthma phenotype. The study reveals new genetic drivers of preschool wheeze, identifying risk loci, phenotype-specific genetic clusters, and the role of nasal epithelium in disease pathobiology.

We identified 8 novel risk loci that have not previously been associated with asthma or other childhood wheezing disorders, including *CASC20*, *GDF5*, *CASC17*, *NR4A*, *THSD4*, *KAZN-AS1*, *ZNF529* and *HNF1B*. Five of these, *GDF5*, *NR4A2*, *CASC17*, *THSD4* and *CASC20*, replicated using a related, but milder, phenotype in the MoBa cohort, indicating robust associations with preschool wheeze for these loci.

GDF5 (Growth Differentiation Factor 5) is the top gene at this locus, with *UQCC1* (ubiquinol-cytochrome c reductase complex assembly factor 1) as a strong alternative, supported by the TWAS results. The rs143384 variant likely impacts *UQCC1* levels in nasal epithelium, but *GDF5* RNA levels couldn't be measured in COPSAC₂₀₁₀. However, a missense variant in *GDF5*, in high LD ($r^2 > 0.6$), supports *GDF5* as the primary effector gene in this locus. *GDF5* is a member of the Bone Morphogenic Protein (BMP) family and BMPs play key roles in development, particularly in organogenesis and differentiation (77). *GDF5*, has been shown to promote generation of brown fat through the induction of *PGC1 α* expression in adipose tissue (78), suggesting a role in body development and fat metabolism. *UQCC1* involves generation of the majority of cellular ATP through mitochondrial oxidative phosphorylation. Specifically, *UQCC1* is involved in

assembly of mitochondrial complex III (79), indicating a housekeeping role of this gene.

The lead SNP in this locus rs143384 is highly associated with *UQCC1* levels in fibroblasts (NES = 0.51, $P = 1.6e-50$) and negatively associated with height ($b = -0.079$, $se = 0.001$, $P = 3.4e-1152$), further indicating a role in growth.

The risk allele in Nuclear receptor 4A2 (*NR4A2*), also known as *NURR1* (nuclear receptor related 1 protein), has previously been shown to increase adult lung function (FEV1/FVC, $b=0.033$, (CI = 0.027 - 0.038) $P=2e-21$) (43). *NR4A2* is an inhibitor of NF-kb-driven inflammation and is expressed in myeloid cells (80,81). Splicing and stability of *NRA4A2* has been shown to be inhibited by non-coding RNA *LUTAC1*, which is upregulated in COPD patients (82). The distance to the nearest gene was the feature determining this gene prioritization.

CASC17 is a long non-coding RNA (lncRNA) with little evidence from the literature linking this to any respiratory or immunological outcomes. An intronic variant in *CASC17* has previously been associated with low lung function (43) and prostate cancer (83).

CASC17 is prioritized by two features, namely physical distance and intron-mapping.

Previous studies have linked other variants in *THSD4* with ICS response in asthma and asthma severity. Response to inhaled corticosteroids have been shown to be influenced by genetic variants in *THSD4*, however the reported variant is not in LD with the finding in the present study (rs62015884, $r^2 = 0.01$) (84). Another study used whole-genome sequencing data to characterize moderate to severe asthma risk in a population with a mix of children and adults identified a genetic variant, rs11631778, near the *THSD4* gene, which was also not in LD with our lead variant ($r^2 = 0.004$) (85). *THSD4* is prioritized by six out of twelve features and is therefore a likely candidate for a causal

gene in the locus. *THSD4* encodes *ADAMTSL6*, a member of the ADAMTS/L superfamily, which aids formation of extracellular matrices by promoting fibrillin-1 assembly and genetic variants in *THSD4* has been connected to risk of aortic aneurysms (86). Additionally, abnormal fibrillin-1 levels have been proposed to play an important role in development of lung emphysema by decreasing elasticity of the connective tissue in the lung (87). This suggests that *THSD4* may confer asthma risk by compromising the structural integrity of the extracellular matrix through dysregulation of fibrillin-1.

CASC20 is an RNA-gene with unknown function, and the gene prioritization predicted *CASC20* using physical distance and intron-mapping. OpenTargets predict *BMP2*, but this is likely because OpenTargets do not include RNA-genes. *BMP2* is involved in bone development and adipose metabolism (88,89) and the lead SNP (rs928202) in the *CASC20*-locus has previously been associated with height ($b = -0.031$, $se = 0.001$, $P = 1.9e-183$) (46), indicating a role in growth.

Genetic correlation analyses showed expected strong association with asthma and lung function in UK Biobank. Genetic correlation with atopic traits were strongly positive, and we observed a strong correlation with eosinophils which provides further evidence that severe wheeze is influenced by T2-high/atopic pathways. Severe wheeze was also negatively correlated with adult height and positively correlated with adult BMI. The association-based grouping analysis suggests that the height-correlation may be driven by a few variants on chromosome 20 as it is relatively weak ($rg = -0.07$, $P = 1.9e-3$). Growth patterns have previously been shown to be involved in wheezing (11–14) and

childhood asthma has been suggested to attenuate adult height (15). BMI indicated a more polygenic mechanism behind the correlation, since no wheeze locus was genome-wide significant for BMI, but the genetic correlation was reasonably strong ($r_g = 0.16$, $P = 7.6e-6$). Genetic predisposition to high BMI has previously been linked to increased risk of severe wheeze in preschool children (90).

We selected genes based on the number of features, assuming all features equally contribute to gene prioritization. A previous study created a scoring scheme for gene prioritization for a BMI-GWAS, using a set of gold standard obesity genes located in genome-wide significant loci to rank methods according to how many gold standard genes they were able to prioritize (91,92). To the authors knowledge no gold standard asthma gene exists in the loci of the GWAS presented in this study thus making such scoring scheme is impossible. This highlights the need for future rare variant or experimental studies to establish causal genes based on GWAS findings from asthma. *HNF1B* is a novel asthma gene which is also a drug-target for metformin, which is an approved antidiabetic agent. The interaction score of 1.34 corresponds to the top 17 % of all drug-gene pairs in the DGIdb. The evidence of metformin use and asthma related outcomes and most of the literature suggests that metformin may reduce the number of and severity of asthma exacerbations and asthma-related emergency visits (93–98).

We identified an additional 3 novel risk loci in the TWAS analysis using the nasopharyngeal RNA sequencing data from the COPSAC₂₀₁₀ cohort, two of which

replicate in an independent cohort. One of these genes was *OAS1/OAS3* which are well-known members of the oligoadenylate synthetase family of genes and are involved in recognition of foreign nucleic acids in the cell and display antiviral activity (99). The lead SNP in *OAS1/OAS3* (rs1859331) has previously been associated with decreased eosinophils and white blood cells (41) and increased risk of severe COVID19 illness (100). The *MTA3* locus has been implicated in a gene-network target by BaP-induced H3K9-acetylation (101). BaP is used as a proxy to stimulate a response similar to airborne particulate matter. Thus, *MTA3* could increase susceptibility to wheeze involvement in environmental response to pollutants. The variant behind *MTA3* (rs10490558) has previously been associated with decreased lung function in adulthood (43).

The association-based grouping analysis revealed suggestive clusters among SNPs based on their associations with asthma related traits, suggesting that different risk loci increase risk for asthma and wheeze through different mechanisms and possibly different endotypes. We identified a cluster at terminal branch (cluster 7), characterized by strong association with high eosinophil levels and risk of atopic disease and weak to moderate involvement with the remaining immune cell counts and lung function. This cluster represents SNPs involved in T2-high pathways and consist of *ILRL1*, *IL-33*, *LINC0276*, *WDR36/TSLP*, *CLEC16A*, *IL-13* & *PHB1*. Cluster 7 is situated on a larger terminal branch and the closest neighbor is cluster 6 which contains only *GSDMB*. *GSDMB* is strongly associated with childhood asthma compared to the other asthma phenotypes, is only weakly associated to atopic traits and strongly associated with

neutrophils but not eosinophils. Neutrophilic childhood asthma might be driven by *GSDMB* related pathways while eosinophilic asthma with later-age allergies appears more polygenic, involving multiple effector genes like those in cluster 7 (102). This suggests that pathways involved in neutrophil response, formation or degradation may be mediated through the 17q risk variant.

Another cluster (cluster 2) displayed weak associations with asthma and the remaining atopic traits. This cluster is characterized by showing strong association with adult lung function and no/weak associations to immune cell counts or atopic phenotypes including asthma. SNPs in this cluster (*NPNT*, *CASC17*, *THSD4*, *MTA3* & *ID4*) may be effectors of the connection between childhood wheeze and adult COPD (103,104). Among the loci in cluster 2, *CASC17* and *MTA3* are most strongly associated with height and low lung function, indicating a role in low lung function due to impacts of growth rather than immunological drivers. This group likely contributes to T2-low asthma, characterized by a lack of eosinophilic inflammation and reduced responsiveness to inhaled corticosteroids (105). Another possibility is that cluster 2 represents loci involved in transient wheeze due to their weak association with asthma at older ages and associations to adult lung function (106).

The last cluster (cluster 4) is heterogeneous in nature and is characterized by a lack of an overall trend in phenotypic consequence. Additionally, strong associations were rare except for the associations between *UQCC1* and *CASC20* with adult height. The closest single SNP was *ABO*, which was characterized by showing stronger association with C-reactive protein, sphingomyelin, eosinophils, monocytes and neutrophils.

We found that *ID4* and *NPNT* potentially interact with *GSDMB* (see Supplementary Figure 5). From the stratified analysis *ID4* seemed to only confer wheeze risk in the presence of one or zero *GSDMB* risk alleles and had no effect among homozygous *GSDMB* risk-carriers. Contrarily, *NPNT* did only display an effect in the presence of two *GSDMB* risk-alleles. These results suggest complicated mechanisms of action for the 17q locus but should be replicated in an independent cohort to evaluate the robustness of the findings. We also observed a significant interaction between *MTA3* and psychiatric status, with *MTA3* only having an effect in children without psychiatric diagnoses, suggesting that we would have observed a stronger signal in a population without overrepresentation of psychiatric illnesses.

The PRS was associated with asthma and wheezing phenotypes, notably infant respiratory distress in both COPSAC cohorts and the MARC-35 cohort. Specifically, respiratory tract infections, rhinovirus respiratory illnesses and fever episodes in the first year of life, were associated with the PRS after correction for multiple testing in the COPSAC cohorts. Furthermore, among infants hospitalized with bronchiolitis, those with high wheeze-PRS experienced more frequent symptoms—such as night cough, vomiting-induced cough, and wheezing without cough— typically before 3 months of age, compared to those with low PRS. This shows an effect of genetic susceptibility already in early infancy, highlighting early infancy is a period for intervention to prevent future development of wheeze.

Functional enrichment of PRS-associated genes in nasal expression levels in COPSAC₂₀₁₀ revealed pathways involved fc gamma receptor (fcg) signaling, toll-like receptors cascades (TLR4, TLR7, TLR8 & TLR9), *S. aureus* and other bacterial and parasitic infections. Binding of IgE to fc gamma receptors is a consequence of allergic inflammation in airway epithelium (107) and fcg's increases susceptibility to bacterial infections by interacting with TLR4 in airway epithelium (108). TLR4 recognizes bacterial lipopolysaccharide (LPS) and capsid-proteins from RSV (109) while TLR7 & TLR8 recognizes single stranded viral RNA (110) and TLR9 have been linked to eosinophilic activation upon stimulation with viral ligands (111). Infections with *Staphylococcus aureus* correlate with wheezing and asthma exacerbations in children and young adults and *S. aureus* produce endotoxins that induce *IL-33* in airway mucosa (112). Leishmaniasis and toxoplasmosis are rare and thus of limited clinical significance (113,114), but are likely enriched among genes due to involvement of eosinophilic pathways in parasitic infections (10).

PRS-associated genes in MARC-35 were mainly enriched for pathways involved in viral infections with interferon response as the top functional terms. The relationship between interferon response and later asthma and wheeze have previously been mapped in a review (115). High IFN- γ levels in nasopharyngeal aspirates of infants with RSV-bronchiolitis associated with asthma diagnosis at 5 years old (29). Thus, we show that high genetic risk of severe wheeze in preschool is accompanied by dysregulated pathways tied to recognition of foreign antigens, cytokine signaling and type-2 immune response in the nasal epithelium.

Finally, we demonstrate that the severe wheeze PRS explains significantly more variation in non-severe wheezing from ages 0 to 7 years compared to prior large-scale GWAS of asthma. This suggests that the present GWAS captures additional heritability for wheeze compared to previous asthma GWAS. Notably, this holds true even though the out-of-sample phenotype represents a non-severe form of wheezing.

Conclusion

This GWAS contributes to the understanding of the genetic mechanisms of severe preschool wheeze. It identifies eight novel genetic loci for preschool wheeze, including genes involved with airway inflammation, diabetes, lung function and growth and some loci requiring further research, five of which replicate in an independent cohort. We confidently prioritized causal genes in 9 out of 23 loci, including key childhood asthma genes, as well as the novel loci *HNF1B*, *THSD4* and *GDF5*. Three SNP groups were identified, with distinct effects on atopy, immune response and adult lung function. Polygenic risk scores highlight genetic susceptibility to respiratory infections in early infancy, suggesting a critical window for interventions to prevent wheeze development.

References

1. Lozano P, Sullivan SD, Smith DH, Weiss KB. The economic burden of asthma in US children: Estimates from the National Medical Expenditure Survey. *J Allergy Clin Immunol*. 1999 Nov;104(5):957–63.
2. Kocevar VS, Bisgaard H, Jönsson L, Valovirta E, Kristensen F, Yin DD, et al. Variations in pediatric asthma hospitalization rates and costs between and within Nordic countries. *Chest*. 2004 May;125(5):1680–4.
3. Bush A. Practice imperfect--treatment for wheezing in preschoolers. *N Engl J Med*. 2009 Jan 22;360(4):409–10.
4. Pardue Jones B, Fleming GM, Otilio JK, Asokan I, Arnold DH. Pediatric acute asthma exacerbations: Evaluation and management from emergency department to intensive care unit. *J Asthma*. 2016 Jul 2;53(6):607–17.
5. Bønnelykke K, Vissing NH, Sevelsted A, Johnston SL, Bisgaard H. Association between respiratory infections in early life and later asthma is independent of virus type. *J Allergy Clin Immunol*. 2015 Jul 1;136(1):81-86.e4.
6. Carlsson CJ, Vissing NH, Sevelsted A, Johnston SL, Bønnelykke K, Bisgaard H. Duration of wheezy episodes in early childhood is independent of the microbial trigger. *J Allergy Clin Immunol*. 2015;
7. Ullemar V, Magnusson PKE, Lundholm C, Zettergren A, Melén E, Lichtenstein P, et al. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy*. 2016 Feb;71(2):230–8.
8. van Beijsterveldt CEM, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J*. 2007 Mar 1;29(3):516–21.
9. Han Y, Jia Q, Jahani PS, Hurrell BP, Pan C, Huang P, et al. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat Commun*. 2020 Dec;11(1):1776.
10. Olafsdottir TA, Theodors F, Bjarnadottir K, Bjornsdottir US, Agustsdottir AB, Stefansson OA, et al. Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis. *Nat Commun*. 2020 Dec;11(1):393.
11. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet*. 2018 Jan;50(1):42.
12. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019 Jun 1;7(6):509–22.
13. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet*. 2014 Jan;46(1):51–5.
14. Eliassen AU, Pedersen CET, Rasmussen MA, Wang N, Soverini M, Fritz A, et al. Genome-wide study of early and severe childhood asthma identifies interaction between CDHR3 and GSDMB. *J Allergy Clin Immunol*. 2022 Sep;150(3):622–630.
15. Ahluwalia TS, Eliassen AU, Sevelsted A, Pedersen CET, Stokholm J, Chawes B, et al. FUT2-ABO epistasis increases the risk of early childhood asthma and *Streptococcus pneumoniae* respiratory illnesses. *Nat Commun*. 2020 Dec 16;11(1):6398.
16. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry*. 2018 Jan;23(1):6–14.

17. Bybjerg-Grauholm J, Bøcker Pedersen C, Bækvad-Hansen M, Giørtz Pedersen M, Adamsen D, Søholm Hansen C, et al. The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders [Internet]. *Epidemiology*; 2020 Dec [cited 2023 Mar 20]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.11.30.20237768>
18. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*. 2015 May;47(5):435–44.
19. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009 Dec 17;462(7275):868–74.
20. Eggertsson HP, Jonsson H, Kristmundsdottir S, Hjartarson E, Kehr B, Masson G, et al. GraphTyper enables population-scale genotyping using pangenome graphs. *Nat Genet*. 2017 Nov;49(11):1654–60.
21. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023 Jan 19;613(7944):508–18.
22. Erikstrup C, Sørensen E, Nielsen KR, Bruun MT, Petersen MS, Rostgaard K, et al. Cohort Profile: The Danish Blood Donor Study. *Int J Epidemiol*. 2022 Oct 4;dyac194.
23. Hansen TF, Banasik K, Erikstrup C, Pedersen OB, Westergaard D, Chmura PJ, et al. DBDS Genomic Cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. *BMJ Open*. 2019 Jun;9(6):e028401.
24. Bisgaard H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol*. 2004 Oct;93(4):381–9.
25. Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Følsgaard NV, Kreiner-Møller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clin Exp Allergy J Br Soc Allergy Clin Immunol*. 2013 Dec;43(12):1384–94.
26. Hasegawa K, Mansbach JM, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalised for bronchiolitis. *Eur Respir J*. 2016 Nov;48(5):1329–39.
27. on behalf of the MARC-35 Investigators, Mansbach JM, Hasegawa K, Geller RJ, Espinola JA, Sullivan AF, et al. Bronchiolitis severity is related to recurrent wheezing by age 3 years in a prospective, multicenter cohort. *Pediatr Res*. 2020 Feb;87(3):428–30.
28. Ralston SL, Lieberthal AS, Meissner HC, Alverson BK, Baley JE, Gadomski AM, et al. Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatrics*. 2014 Nov 1;134(5):e1474–502.
29. Raita Y, Pérez-Losada M, Freishtat RJ, Harmon B, Mansbach JM, Piedra PA, et al. Integrated omics endotyping of infants with respiratory syncytial virus bronchiolitis and risk of childhood asthma. *Nat Commun*. 2021 Jun 14;12(1):3601.
30. Helgeland Ø, Vaudel M, Juliusson PB, Lingaas Holmen O, Juodakis J, Bacelis J, et al. Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. *Nat Commun*. 2019 Oct 1;10(1):4448.
31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015 Dec;4(1):7.
32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006 Aug;38(8):904–9.
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma Oxf Engl*. 2010 Sep 1;26(17):2190–1.

34. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the “Sum of Single Effects” model. Epstein MP, editor. *PLOS Genet.* 2022 Jul 19;18(7):e1010299.
35. Kanai M, Elzur R, Zhou W, Daly MJ, Finucane HK, Zhou W, et al. Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genomics.* 2022 Dec;2(12):100210.
36. Hinrichs AS. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* 2006 Jan 1;34(90001):D590–8.
37. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015 Mar;47(3):291–5.
38. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015 Nov;47(11):1236–41.
39. Budu-Aggrey A, Kilanowski A, Sobczyk MK, 23andMe Research Team, Shringarpure SS, Mitchell R, et al. European and multi-ancestry genome-wide association meta-analysis of atopic dermatitis highlights importance of systemic immune regulation. *Nat Commun.* 2023 Oct 4;14(1):6172.
40. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat Genet.* 2018 Aug;50(8):1072.
41. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell.* 2020 Sep 3;182(5):1214–1231.e11.
42. Ferreira MAR, Mathur R, Vonk JM, Sz wajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am J Hum Genet.* 2019 Apr 4;104(4):665–84.
43. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet.* 2019 Mar;51(3):481–93.
44. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun.* 2017 Sep 19;8(1):599.
45. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet.* 2018 Oct 15;27(20):3641–9.
46. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. *Nature.* 2022 Oct 27;610(7933):704–12.
47. Said S, Pazoki R, Karhunen V, Vösa U, Ligthart S, Bodinier B, et al. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nat Commun.* 2022 Apr 22;13(1):2198.
48. Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, Davey Smith G, et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. Locasale JW, editor. *PLOS Biol.* 2022 Feb 25;20(2):e3001547.
49. Karolchik D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004 Jan 1;32(90001):493D – 496.
50. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010 Sep 1;38(16):e164–e164.
51. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate

- proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023 Sep 22;381(6664):eadg7492.
52. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun*. 2015;6:5890.
 53. Weeks EM, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat Genet*. 2023 Aug;55(8):1267–76.
 54. De Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. Tang H, editor. *PLOS Comput Biol*. 2015 Apr 17;11(4):e1004219.
 55. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016 Mar;48(3):245–52.
 56. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016 Dec;17(1):122.
 57. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, et al. DGIdb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res*. 2024 Jan 5;52(D1):D1227–35.
 58. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016 Jul;48(7):709–17.
 59. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016 Sep 15;32(18):2847–9.
 60. Gu Z. Complex heatmap visualization. *iMeta*. 2022 Sep;1(3):e43.
 61. Morin A, McKennan CG, Pedersen CET, Stokholm J, Chawes BL, Malby Schoos AM, et al. Epigenetic landscape links upper airway microbiota in infancy with allergic rhinitis at 6 years of age. *J Allergy Clin Immunol* [Internet]. 2020 Jul 18 [cited 2020 Oct 21]; Available from: <http://www.sciencedirect.com/science/article/pii/S0091674920309672>
 62. Raita Y, Pérez-Losada M, Freishtat RJ, Hahn A, Castro-Nallar E, Ramos-Tapia I, et al. Nasopharyngeal metatranscriptome profiles of infants with bronchiolitis and risk of childhood asthma: a multicentre prospective study. *Eur Respir J*. 2022 Jul;60(1):2102293.
 63. Zhu Z, Camargo CA, Raita Y, Freishtat RJ, Fujiogi M, Hahn A, et al. Nasopharyngeal airway dual-transcriptome of infants with severe bronchiolitis and risk of childhood asthma: A multicenter prospective study. *J Allergy Clin Immunol*. 2022 Apr 26;S0091-6749(22)00556-5.
 64. Fujiogi M, Raita Y, Pérez-Losada M, Freishtat RJ, Celedón JC, Mansbach JM, et al. Integrated relationship of nasopharyngeal airway host response and microbiome associates with bronchiolitis severity. *Nat Commun*. 2022 Aug 30;13(1):4970.
 65. Hammond RK, Pahl MC, Su C, Cousminer DL, Leonard ME, Lu S, et al. Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *eLife*. 2021 Jan 18;10:e62206.
 66. Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. Epstein MP, editor. *PLOS Genet*. 2020 Apr 20;16(4):e1008720.
 67. Pasaniuc B, Zaitlen N, Shi H, Bhatia G, Gusev A, Pickrell J, et al. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*. 2014 Oct 15;30(20):2906–14.
 68. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015 Apr 20;43(7):e47–e47.
 69. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013 Dec;14(1):128.
 70. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a

- comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016 Jul 8;44(W1):W90–7.
71. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc.* 2021 Mar;1(3):e90.
 72. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun.* 2019 Apr 16;10(1):1776.
 73. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007 Sep;81(3):559–75.
 74. Namjou B, Lape M, Malolepsza E, DeVore SB, Weirauch MT, Dikilitas O, et al. Multiancestral polygenic risk score for pediatric asthma. *J Allergy Clin Immunol.* 2022 Nov;150(5):1086–96.
 75. Tsuo K, Zhou W, Wang Y, Kanai M, Namba S, Gupta R, et al. Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. *Cell Genomics.* 2022 Dec;2(12):100212.
 76. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988 Sep;44(3):837–45.
 77. Pregizer S, Mortlock DP. Control of BMP gene expression by long-range regulatory elements. *Cytokine Growth Factor Rev.* 2009 Oct;20(5–6):509–15.
 78. Hinoi E, Nakamura Y, Takada S, Fujita H, Iezaki T, Hashizume S, et al. Growth Differentiation Factor-5 Promotes Brown Adipogenesis in Systemic Energy Expenditure. *Diabetes.* 2014 Jan 1;63(1):162–75.
 79. Tucker EJ, Wanschers BFJ, Szklarczyk R, Mountford HS, Wijeyeratne XW, Van Den Brand MAM, et al. Mutations in the UQCC1-Interacting Protein, UQCC2, Cause Human Complex III Deficiency Associated with Perturbed Cytochrome b Protein Expression. Moraes CT, editor. *PLoS Genet.* 2013 Dec 26;9(12):e1004034.
 80. Mahajan S, Saini A, Chandra V, Nanduri R, Kalra R, Bhagyaraj E, et al. Nuclear Receptor Nr4a2 Promotes Alternative Polarization of Macrophages and Confers Protection in Sepsis. *J Biol Chem.* 2015 Jul;290(30):18304–14.
 81. McEvoy C, De Gaetano M, Giffney HE, Bahar B, Cummins EP, Brennan EP, et al. NR4A Receptors Differentially Regulate NF- κ B Signaling in Myeloid Cells. *Front Immunol* [Internet]. 2017 Jan 23 [cited 2024 Jul 30];8. Available from: <http://journal.frontiersin.org/article/10.3389/fimmu.2017.00007/full>
 82. Vierbuchen T, Agarwal S, Johnson JL, Galia L, Lei X, Stein K, et al. The lncRNA LUCAT1 is elevated in inflammatory disease and restrains inflammation by regulating the splicing and stability of NR4A2. *Proc Natl Acad Sci.* 2023 Jan 3;120(1):e2213715120.
 83. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* 2007 Aug;39(8):977–83.
 84. Dahlin A, Sordillo JE, McGeachie M, Kelly RS, Tantisira KG, Lutz SM, et al. Genome-wide interaction study reveals age-dependent determinants of responsiveness to inhaled corticosteroids in individuals with asthma. Padhukasahasram B, editor. *PLOS ONE.* 2020 Mar 2;15(3):e0229241.
 85. Chang D, Hunkapiller J, Bhangale T, Reeder J, Mukhyala K, Tom J, et al. A whole genome sequencing study of moderate to severe asthma identifies a lung function locus associated with asthma risk. *Sci Rep.* 2022 Apr 2;12(1):5574.
 86. Elbitar S, Renard M, Arnaud P, Hanna N, Jacob MP, Guo DC, et al. Pathogenic variants in THSD4, encoding the ADAMTS-like 6 protein, predispose to inherited thoracic aortic aneurysm. *Genet Med.* 2021 Jan;23(1):111–22.
 87. Robbesom AA, Koenders MMJF, Smits NC, Hafmans T, Versteeg EMM, Bulten J, et al.

- Aberrant fibrillin-1 expression in early emphysematous human lung: a proposed predisposition for emphysema. *Mod Pathol*. 2008 Mar;21(3):297–307.
88. Halloran D, Durbano HW, Nohe A. Bone Morphogenetic Protein-2 in Development and Bone Homeostasis. *J Dev Biol*. 2020 Sep 13;8(3):19.
 89. Blázquez-Medela AM, Jumabay M, Boström KI. Beyond the bone: Bone morphogenetic protein signaling in adipose tissue. *Obes Rev*. 2019 May;20(5):648–58.
 90. Jensen SK, Pedersen CET, Fischer-Rasmussen K, Melgaard ME, Brustad N, Kyvsgaard JN, et al. Genetic predisposition to high BMI increases risk of early life respiratory infections and episodes of severe wheeze and asthma. *Eur Respir J*. 2024 Sep;64(3):2400169.
 91. Hemerich D, Svenstrup V, Obrero VD, Preuss M, Moscati A, Hirschhorn JN, et al. An integrative framework to prioritize genes in more than 500 loci associated with body mass index. *Am J Hum Genet*. 2024 Jun;111(6):1035–46.
 92. Hendricks AE, Bochukova EG, Marenne G, Keogh JM, Atanassova N, Bounds R, et al. Rare Variant Analysis of Human and Rodent Obesity Genes in Individuals with Severe Childhood Obesity. *Sci Rep*. 2017 Jun 29;7(1):4394.
 93. Wen L, Zhong W, Chai Y, Zhong Q, Gao J, Guan L, et al. Association of Metformin Use with Asthma Exacerbation in Patients with Concurrent Asthma and Diabetes: A Systematic Review and Meta-Analysis of Observational Studies. *Can Respir J*. 2020 Aug 4;2020:1–8.
 94. Li C, Erickson SR, Wu C. Metformin use and asthma outcomes among patients with concurrent asthma and diabetes. *Respirology*. 2016 Oct;21(7):1210–8.
 95. Wu TD, Keet CA, Fawzy A, Segal JB, Brigham EP, McCormack MC. Association of Metformin Initiation and Risk of Asthma Exacerbation. A Claims-based Cohort Study. *Ann Am Thorac Soc*. 2019 Dec;16(12):1527–33.
 96. Ararat E, Landes RD, Forno E, Tas E, Perry TT. Metformin use is associated with decreased asthma exacerbations in adolescents and young adults. *Pediatr Pulmonol*. 2024 Jan;59(1):48–54.
 97. Wu TD, Fawzy A, Kinney GL, Bon J, Neupane M, Tejwani V, et al. Metformin use and respiratory outcomes in asthma-COPD overlap. *Respir Res*. 2021 Dec;22(1):70.
 98. Yen FS, Hsu CC, Shih YH, Pan WL, Wei JCC, Hwu CM. Metformin and the Development of Asthma in Patients with Type 2 Diabetes. *Int J Environ Res Public Health*. 2022 Jul 5;19(13):8211.
 99. Hornung V, Hartmann R, Ablasser A, Hopfner KP. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat Rev Immunol*. 2014 Aug;14(8):521–8.
 100. COVID-19 Host Genetics Initiative, COVID-19 Host Genetics InitiativeLeadership, Niemi MEK, Karjalainen J, Liao RG, Neale BM, et al. Mapping the human genetic architecture of COVID-19. *Nature*. 2021 Dec 16;600(7889):472–7.
 101. Ho SM. Environmental epigenetics of asthma: An update. *J Allergy Clin Immunol*. 2010 Sep;126(3):453–65.
 102. 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 Sep;5(5):1351-1361.e2.
 103. Bui DS, Lodge CJ, Burgess JA, Lowe AJ, Perret J, Bui MQ, et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med*. 2018 Jul;6(7):535–44.
 104. Bisgaard H, Nørgaard S, Sevelsted A, Chawes BL, Stokholm J, Mortensen EL, et al. Asthma-like symptoms in young children increase the risk of COPD. *J Allergy Clin Immunol*. 2021 Feb;147(2):569-576.e9.
 105. Kuruvilla ME, Lee FEH, Lee GB. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. *Clin Rev Allergy Immunol*. 2019 Apr;56(2):219–33.

106. Kerkhof M, Boezen HM, Granell R, Wijga AH, Brunekreef B, Smit HA, et al. Transient early wheeze and lung function in early childhood associated with chronic obstructive pulmonary disease genes. *J Allergy Clin Immunol*. 2014 Jan;133(1):68-76.e1-4.
107. Wang Y, Bai C, Li K, Adler KB, Wang X. Role of airway epithelial cells in development of asthma and allergic rhinitis. *Respir Med*. 2008 Jul;102(7):949–55.
108. Golebski K, Hoepel W, Van Egmond D, De Groot EJ, Amatngalim GD, Beekman JM, et al. FcγRIII stimulation breaks the tolerance of human nasal epithelial cells to bacteria through cross-talk with TLR4. *Mucosal Immunol*. 2019 Mar;12(2):425–33.
109. Rallabhandi P, Phillips RL, Boukhvalova MS, Pletneva LM, Shirey KA, Gioannini TL, et al. Respiratory Syncytial Virus Fusion Protein-Induced Toll-Like Receptor 4 (TLR4) Signaling Is Inhibited by the TLR4 Antagonists *Rhodobacter sphaeroides* Lipopolysaccharide and Eritoran (E5564) and Requires Direct Interaction with MD-2. Pier G, editor. *mBio*. 2012 Aug 31;3(4):e00218-12.
110. Diebold S. Recognition of viral single-stranded RNA by Toll-like receptors☆. *Adv Drug Deliv Rev*. 2008 Apr 29;60(7):813–23.
111. Månsson A, Cardell LO. Role of atopic status in Toll-like receptor (TLR)7- and TLR9-mediated activation of human eosinophils. *J Leukoc Biol*. 2009 Apr 1;85(4):719–27.
112. Bachert C, Humbert M, Hanania NA, Zhang N, Holgate S, Buhl R, et al. *Staphylococcus aureus* and its IgE-inducing enterotoxins in asthma: current knowledge. *Eur Respir J*. 2020 Apr;55(4):1901592.
113. Zangenberg M, Helleberg M, Stensvold CR, Nielsen HV. Imported leishmaniasis in Denmark. *Ugeskr Læg*. 2024 Apr 22;1–8.
114. Marstrand J, Kurtzhals JAL, Fuchs HJ, Nielsen HV, Jokelainen P. The disease burden of ocular toxoplasmosis in Denmark in 2019: Estimates based on laboratory testing of ocular samples and on publicly available register data. *Parasite Epidemiol Control*. 2021 Nov;15:e00229.
115. Makrinioti H, Custovic A, Hasegawa K, Camargo CA, Jartti T. The role of interferons in preschool wheeze. *Lancet Respir Med*. 2021 Jan;9(1):9–11.

Tables

Table 1: Fine-mapped variants are reported if a 95 % credible set could be computed, otherwise the lead SNP is reported. Replication in MoBa is shown on the right side. Loci not previously associated with asthma or other wheezing disorders are highlighted in bold, and loci with significant replication ($P < 0.05$) have p-values highlighted in bold.

Discovery: Severe wheeze (0-7 yerars)								Replication: MoBa (wheeze 0-7 years)			
Chr	Position	SNP	EAF	EA/NEA	OR	CI 95 %	P-value	Mapped gene	OR	CI 95 %	P-value
17	38062196	rs2305480	0.4797	G/A	1.32	(1.29 - 1.35)	2.70E-124	GSDMB	1.12	(1.09-1.15)	2.28E-15
7	105658451	rs6967330	0.2112	A/G	1.22	(1.19 - 1.26)	1.78E-47	CDHR3	1.13	(1.09-1.17)	3.87E-12
9	6213387	rs928413	0.7495	G/A	1.17	(1.14 - 1.2)	1.12E-33	IL33-2	1.05	(1.02-1.09)	1.24E-03
6	31009055	rs2517546	0.1259	T/C	1.25	(1.2 - 1.3)	9.81E-29	HCG22	1.10	(1.06-1.14)	4.72E-07
9	6114076	rs439190	0.7261	T/C	1.13	(1.1 - 1.16)	1.09E-19	IL33-1	1.03	(1-1.06)	5.54E-02
5	131996669	rs847	0.269	T/C	1.13	(1.1 - 1.16)	3.30E-19	IL13	1.04	(1.01-1.08)	9.14E-03
2	102955082	rs13408661	0.1398	G/A	1.16	(1.12 - 1.2)	7.54E-18	IL1RL1	1.06	(1.01-1.1)	8.48E-03
5	110427795	rs79881201	0.3358	T/C	1.1	(1.07 - 1.13)	7.63E-15	WDR36	1.02	(1-1.05)	9.68E-02
6	19842661	rs9350191	0.8538	T/C	1.14	(1.1 - 1.18)	7.75E-15	ID4	1.05	(1.01-1.09)	1.52E-02
9	136151806	rs600038	0.7852	C/T	1.1	(1.07 - 1.13)	9.94E-12	ABO	1.02	(0.99-1.05)	2.40E-01
20	6470265	rs928202	0.6057	G/A	1.08	(1.06 - 1.11)	1.85E-11	CASC20	1.05	(1.02-1.08)	1.13E-03
16	11277358	rs11644510	0.3444	C/T	1.08	(1.06 - 1.11)	1.33E-10	CLEC16A	1.01	(0.98-1.03)	7.06E-01
20	34025756	rs143384	0.5648	A/G	1.08	(1.05 - 1.1)	1.48E-10	GDF5	1.06	(1.03-1.09)	2.25E-05
17	47461433	rs17637472	0.4148	A/G	1.08	(1.05 - 1.1)	5.17E-10	PHB1	1.02	(0.99-1.05)	1.35E-01
17	69191316	rs1008348	0.4256	A/G	1.07	(1.05 - 1.1)	2.38E-09	CASC17	1.05	(1.03-1.08)	1.77E-04
2	157083697	rs72907878	0.1271	G/A	1.11	(1.07 - 1.15)	4.45E-09	NR4A2	1.08	(1.04-1.13)	1.35E-04
4	106819053	rs34712979	0.2563	A/G	1.08	(1.05 - 1.11)	7.35E-09	NPNT	0.98	(0.92-1.05)	5.73E-01
15	71804587	rs62015884	0.177	T/C	1.09	(1.06 - 1.12)	2.53E-08	THSD4	1.07	(1.03-1.1)	4.27E-04
1	14440292	rs12026570	0.7644	A/G	1.08	(1.05 - 1.11)	3.18E-08	KAZN-AS1	1.01	(0.98-1.04)	4.84E-01
19	37048117	rs3096616	0.2866	C/T	1.07	(1.05 - 1.1)	3.20E-08	ZNF529	0.98	(0.95-1.02)	3.19E-01
10	8939198	rs78539108	0.8656	T/A	1.1	(1.06 - 1.13)	3.53E-08	LINC02676	1.03	(0.99-1.07)	1.67E-01
19	49218111	rs33988101	0.4824	G/T	1.07	(1.04 - 1.09)	5.65E-08	FUT2	1.00	(0.97-1.02)	8.14E-01
17	36102035	rs7405696	0.4858	G/C	1.06	(1.04 - 1.09)	1.99E-07	HNF1B	0.99	(0.96-1.01)	2.94E-01

Table 2: Gene prioritization. Overview of methods used for each locus and their prediction.

CHR	POS	SNP	Nearest gene	Intron	Exon	Splice variant	AlphaMissense	DEPICT	MAGMA	PaPS	TWAS + COLOC	eQTL	EWAS + COLOC	OpenTargets	Consensus gene
1	14440292	rs12026570	KAZN-AS1	-	-	-	-	-	-	-	-	-	-	-	KAZN
2	102955082	rs13408661	IL1RL1	SLC9A4	-	-	IL1RL1	-	IL1RL1	IL1RL1	-	IL1RL1	IL1RL1	IL1RL1	IL1RL1
2	157083697	rs72907878	NR4A2	LINC01876	-	-	-	-	-	-	-	-	-	GP2	NR4A2
4	106819053	rs34712979	NPNT	NPNT	-	-	-	NPNT	INTS12	NPNT	-	-	-	NPNT	NPNT
5	110427795	rs79881201	WDR36	WDR36	-	-	-	-	TSLP	WDR36	-	TSLP	-	WDR36	WDR36
5	131996669	rs847	IL13	IL13	IL13	-	IL13	-	IL13	IL13	-	-	-	IL13	IL13
6	19842661	rs9350191	ID4	-	-	-	-	-	ID4	ID4	-	-	-	ID4	ID4
6	31009055	rs2517546	HCG22	HCG20	HCP5	HCG20	PPP1R18	-	-	MUC22	-	-	-	HLA-C	HCG20
7	105658451	rs6967330	CDHR3	CDHR3	CDHR3	-	CDHR3	-	CDHR3	CDHR3	-	-	-	CDHR3	CDHR3
9	6213387	rs928413	IL33-1	-	-	-	-	-	RP11-218T.2	-	-	-	-	RANBP6	IL33-1
9	6114076	rs439190	IL33-2	-	-	-	-	-	IL33-2	IL33-2	-	-	-	ERMP1	IL33-2
9	136151806	rs600038	ABO	-	-	-	-	-	-	-	-	ABO	-	ABO	ABO
10	8939198	rs78539108	LINC02676	LINC02676	-	-	-	-	-	-	-	-	-	-	LINC02676
15	71804587	rs62015884	THSD4	THSD4	-	-	THSD4	THSD4	THSD4	THSD4	-	-	-	THSD4	THSD4
16	11277358	rs11644510	SOC1	CLEC16A	-	-	-	-	CLEC16A	CLEC16A	-	-	CLEC16A	CLEC16A	CLEC16A
17	36102035	rs7405696	HNF1B	HNF1B	-	-	-	-	HNF1B	HNF1B	-	-	-	HNF1B	HNF1B
17	38062196	rs2305480	GSDMB	ORMDL3	-	GSDMB	ZFPBP2	-	ORMDL3	GSDMB	GSDMB	GSDMB	GSDMB	GSDMB	GSDMB
17	47461433	rs17637472	PHB1	ZNF852	-	-	-	-	ZNF852	-	-	-	-	AB1	ZNF852
17	69191316	rs1008348	CASC17	CASC17	-	-	-	-	-	-	-	-	-	CASC17	CASC17
19	37048117	rs3096616	ZNF529	-	-	-	-	-	ZNF529	ZNF529	-	-	-	ZNF529	ZNF529
19	49218111	rs33988101	MAMSTR	IZUMO1	FUT2	-	FUT2	-	RASIP1	MAMSTR	FUT2	FUT2	IZUMO1	MAMSTR	FUT2
20	6470265	rs928202	CASC20	CASC20	-	-	-	-	-	-	-	-	-	BMP2	CASC20
20	34025756	rs143384	GDF5	UQC1	GDF5	-	GDF5	-	UQC1	GDF5	UQC1	FAM83C	UQC1	CPNE1	GDF5

Table 4: Genetic correlation (rg) results. The trait column contains the name of the trait that is correlated with the severe wheeze GWAS. The genetic correlation with standard error, p-value and Benjamin-Hocberg correction is shown in the remaining columns.

Traits are sorted according to lowest p-value starting from the top.

Trait	Genetic correlation			
	rg	95 % CI	p-value	fdr-value
Asthma ever	0.69	0.63;0.75	5.53E-113	1.99E-111
Asthma 0-15 yr	0.73	0.62;0.84	9.61E-36	1.73E-34
PEF	-0.28	-0.34;-0.21	3.77E-16	3.25E-15
FEV1	-0.29	-0.36;-0.22	4.51E-16	3.25E-15
FEV1/FVC	-0.27	-0.34;-0.2	6.46E-15	3.88E-14
Pneumonia	0.44	0.31;0.57	3.04E-11	1.56E-10
Asthma 25-65 yr	0.47	0.31;0.62	4.40E-09	2.26E-08
Atopic dermatitis	0.34	0.23;0.46	1.22E-08	5.49E-08
FVC	-0.17	-0.24;-0.11	7.13E-08	2.85E-07
BMI	0.16	0.1;0.23	7.96E-07	2.87E-06
Eosinophil count	0.23	0.12;0.33	1.87E-05	6.12E-05
Allergic rhinitis	0.29	0.13;0.45	2.82E-04	8.45E-04
Allergic sensitization	0.27	0.12;0.42	4.13E-04	1.14E-03
Height	-0.08	-0.12;-0.03	2.10E-03	5.40E-03
C-reactive protein	0.12	0.04;0.2	2.94E-03	7.04E-03
Ear infections	0.15	0.05;0.25	3.28E-03	7.39E-03
Myringotomy	0.18	0;0.35	4.41E-02	9.33E-02
Tonsillectomy	0.10	0;0.2	5.84E-02	1.11E-01
Lymphocyte count	0.07	0;0.14	6.32E-02	1.14E-01
Strep Throat	0.14	-0.02;0.31	7.84E-02	1.34E-01
Chronic sinus infections	0.17	-0.04;0.39	1.17E-01	1.91E-01
No. of colds last year	-0.12	-0.28;0.04	1.52E-01	2.38E-01
Basophil count	0.10	-0.06;0.26	2.05E-01	2.98E-01
Sphingomyelin	-0.07	-0.17;0.04	2.07E-01	2.98E-01
Neutrophil count	0.06	-0.06;0.18	2.94E-01	4.07E-01
Monocyte count	0.03	-0.03;0.08	3.58E-01	4.71E-01

Figures

Figure 1: Manhattan plot of the severe wheeze GWAS showing negative log-transformed p-values on the y-axis and chromosomal positions on the x-axis. Genomic risk loci are marked by the prioritized consensus gene. Loci not previously associated with asthma or other wheezing disorders are in bold.

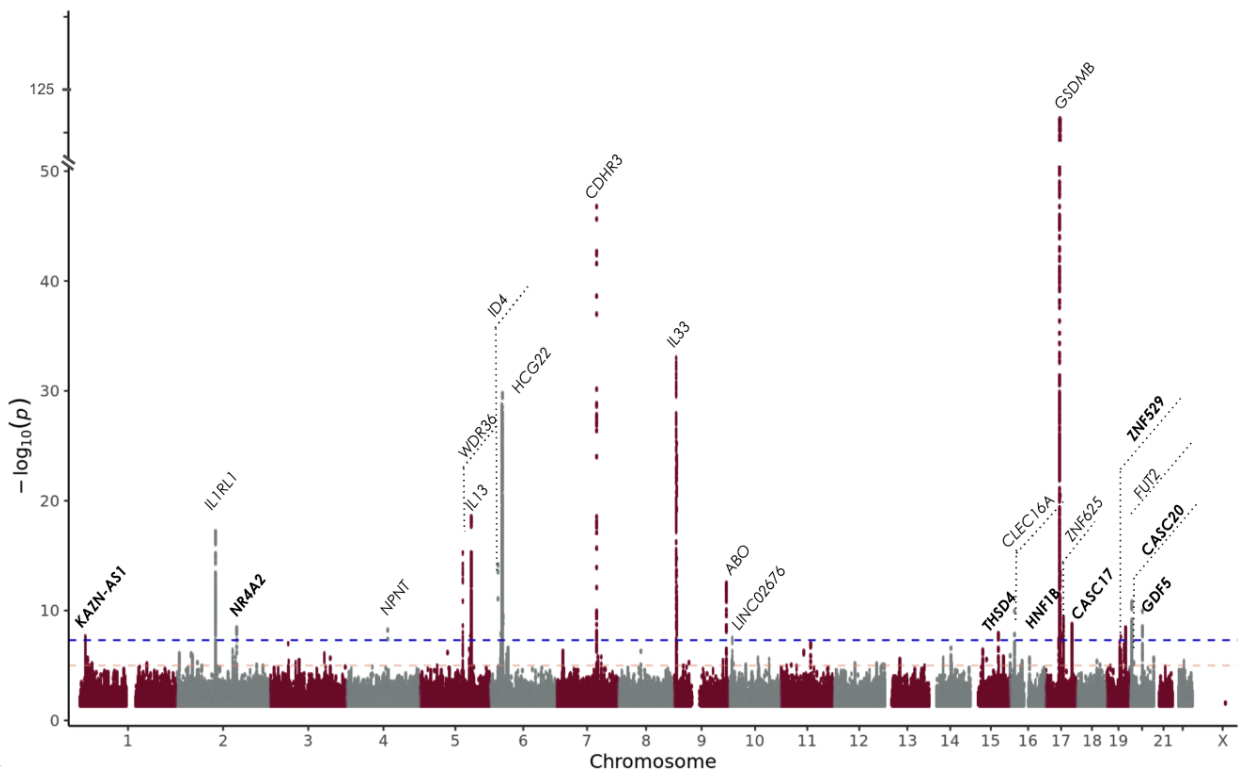


Figure 2: Genetic correlation between severe wheeze and related traits are shown below. The squares are colored and sized according to the genetic correlation estimate (rg), while stars represent genetic correlations that pass correction for multiple testing using the Benjamini-Hochberg procedure (FDR < 0.05).

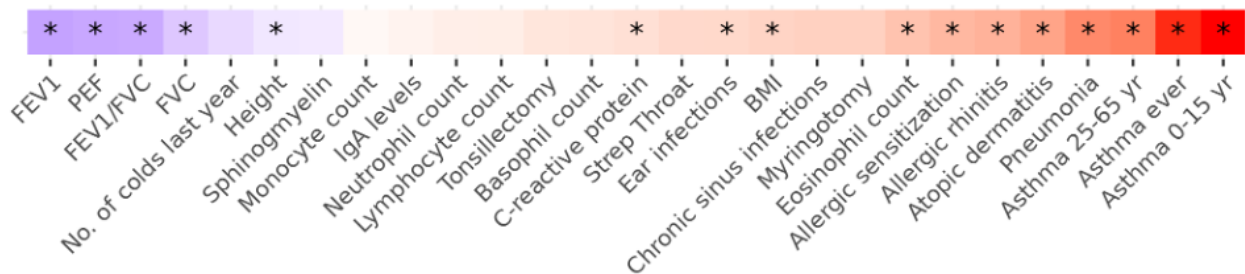


Figure 3: Trait-based clustering of SNPs. SNP association statistics were scaled to z-scores, and further scaled to zero-mean and unit variance to account for sample size imbalance. K-means clustering was used to determine the number of clusters among SNPs and traits, and hierarchical clustering was used to group SNPs with similar phenotypic consequences in each cluster. The SNP-clusters are visualized with unscaled z-scores to reflect the true direction and strengths of association. SNPs with missing values were imputed using the median value for a trait. Only traits with at least one genome-wide significant wheeze lead SNP ($P < 5e-8$) were included in the analysis. Red color indicates a positive z-score and blue indicates a negative. Cells are marked with ‘*’ if trait-association passes correction with Benjamini-Hochberg (FDR < 0.05).

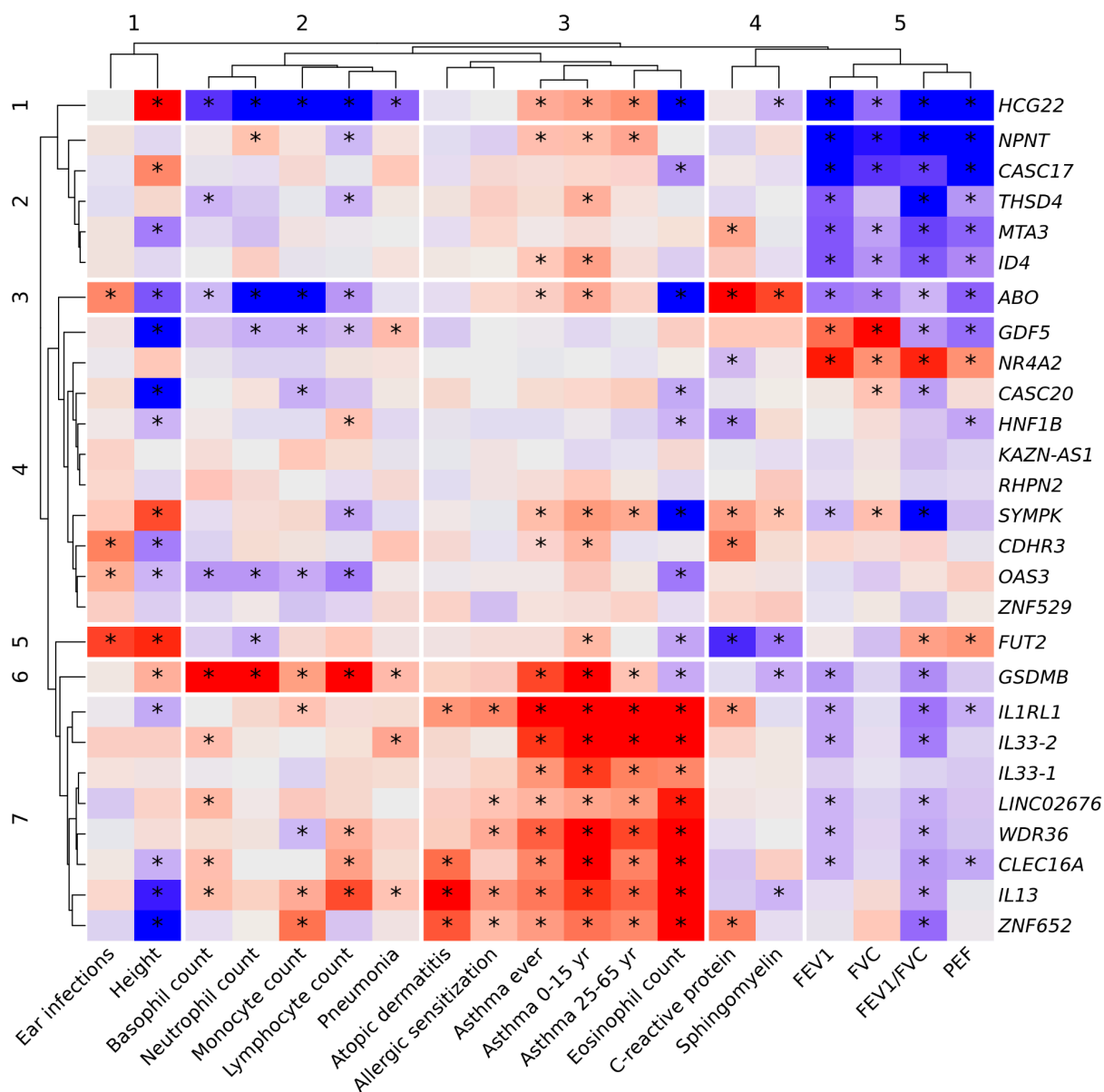
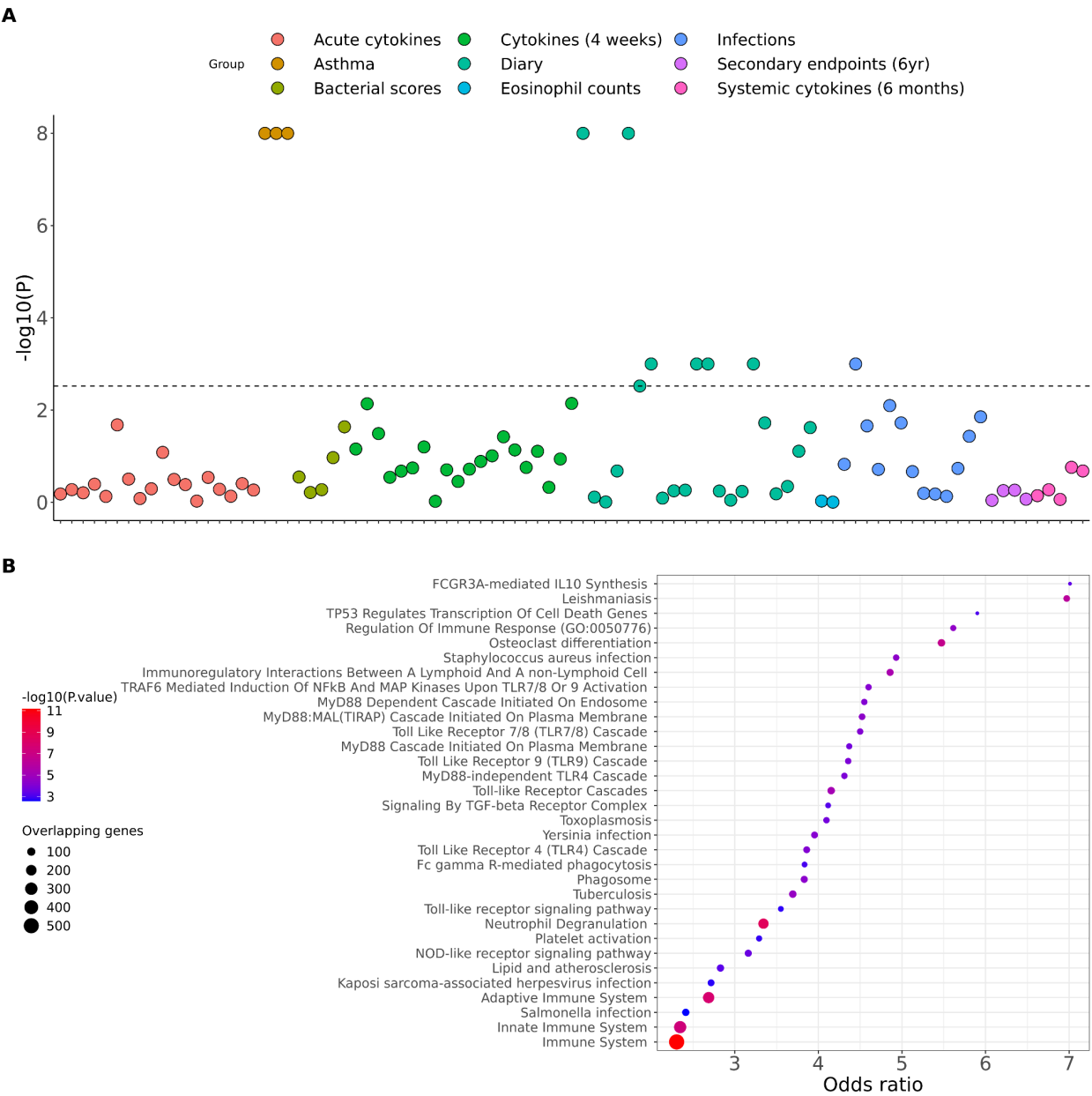


Figure 4: A: Clinical fingerprint of the PRS. PheWAS of clinical and para-clinical endpoints in COPSAC₂₀₀₀ + COPSAC₂₀₁₀, with negative-log-p-values on the y-axis and endpoints on the x-axis. Coloring represents categories of endpoints sorted after alphabetical order. The horizontal dotted line represents the significance cutoff after correcting for multiple testing (FDR < 0.05). **B:** Functional enrichment of airway

(epithelium) I genes associated with the severe wheeze PRS, excluding the HLA-region. Only functional terms passing Benjamini-Hochberg (FDR < 0.05) are included in the figure. X-axis show odds-ratio for functional enrichment, points are colored according to enrichment p-values and sized according to the number of overlapping genes in the functional term.



Supplementary material

Genome-Wide Association Study of Severe Preschool Wheeze Uncovers Novels Insights into the Pathobiology

Kasper Fischer-Rasmussen¹, Casper-Emil Tingskov Pedersen¹, Luo Yang¹, Bo Chawes¹, Jakob Stokholm¹, Ann-Marie Malby Schoos¹, Per Magnus⁹, Yunsung Lee⁹, Marianna Niemi^{10,11}, Tarja Laitinen¹², Þórunn Olafsdottir^{13,14}, Guðmar Þorleifsson¹⁴, Ingileif Jonsdottir¹⁴, Christian Erikstrup¹⁵, Lotte Aas Hindhede¹⁵, Carole Ober¹⁶, Zhaozhong Zhu¹⁷, Carlos A. Camargo, Jr.¹⁷, Anders Børghlum^{5,6,7}, Anders Gorm Pedersen⁴, Anders Ulrik Eliassen^{1,2,3} & Klaus Bønnelykke¹

Cohort recruitment details, acknowledgements and funding statements

Recruitment

Individuals diagnosed with schizophrenia, mood disorders, bipolar affective disorder, autism and attention-deficit/hyperactivity disorder were identified through linkage between Danish population-based registers along with a random sample of the same population that supplied the cases.[15](#) Dried blood spots for virtually all individuals were retrieved from the Danish Neonatal Screening Biobank and processed for genotyping. The design includes the ability to efficiently analyse prospectively collected cohort data within the iPSYCH case-cohort sample.[15](#) This particular design provides several advantages: As the cohort is randomly selected from the entire population, we are able to generate unbiased absolute risks and incidence rates and to estimate the effect sizes of genetic markers on risk of mental disorders, which is representative of the entire Danish population. To date, most genetic and epidemiological studies are based on convenient case-control samples, which are prone to biases.[15](#), [16](#) The iPSYCH₂₀₁₂ sample was preceded by four smaller Danish samples,[17](#), [18](#), [19](#), [20](#), [21](#), [22](#), [23](#), [24](#) all aiming to investigate the potential interplay between genes and the environment. Collectively, these forerunners informed on the best possible study design to use in the iPSYCH₂₀₁₂ sample ([Supplementary Text 1](#)). The following three paragraphs describe the resources and methods used to identify individuals included in the iPSYCH₂₀₁₂ sample.

Acknowledgements & funding

This study was supported by The Lundbeck Foundation (grant numbers R102-A9118 and R155-2014-1724), Denmark, the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project number 294838) and the Stanley Center for Psychiatric Research at Broad Institute and Centre for Integrated Register-based Research at Aarhus University. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation. Professor John J McGrath is supported by grant APP1056929 from the John Cade Fellowship from the National Health and Medical Research Council and the Danish National Research Foundation (Niels Bohr Professorship). We thank Betina Trabjerg, National Centre for Register-Based Research, Aarhus University, School of Business and Social Sciences, Aarhus, Denmark, for technical help in producing the principal component plot. We are indebted to the late Mads Vilhelm Hollegaard for his contribution to make sample material accessible for analysis from the Danish Neonatal Screening Biobank. Mads' pioneering work will be used in this and future studies.

The iPSYCH2012 had four precursors; all aiming at the potential interplay between genes and the environment and all of smaller scales. All studies included persons born in Denmark May 1981 or later. Collectively, these forerunners informed on the best possible study design to be used in the iPSYCH2012. In short, the very first selected 413 persons with schizophrenia like psychoses (N=186) and affective disorders (N=256) diagnosed up until October 1999. For each case, the samples in the Danish Neonatal

Screening Biobank physically located immediately before and after the sample were selected as controls(1). The second approach (SKIZO2005) utilized persons diagnosed with schizophrenia and bipolar affective disorder up until September 2005 first including 231 genetic markers and later whole genome genotyped(2-5). The third approach (SKIZO2006) utilized all persons diagnosed with schizophrenia up until December 31, 2006 reusing biological material already collected in the previous study(5-7). Finally, SKIZO2008 utilising new persons diagnosed with schizophrenia up until December 31, 2008 (8). Controls for SKIZO2005, SKIZO2006 and SKIZO2008 were individually time-matched using the following criteria, born in Denmark, had the same birthday as the corresponding case, and were alive and not diagnosed with schizophrenia when the case was first diagnosed. Genome wide genotyping was performed using the Illumina Human 610-quad beadchip. As demonstrated most persons diagnosed with schizophrenia up until 31 December 2008 had already been genotyped prior to the iPSYCH2012 sample, and was therefore not included here. The latter three designs include a total of 1887 cases of schizophrenia and 1863 individually time-matched controls. These individually matched designs are capable of estimating unbiased incidence rate ratios of the exposures/markers of interest(9), but challenging in situations where the matching is broken, e.g., due to missing information in the Danish Neonatal Screening Biobank or failure in genotyping/QC. The increased capability for the iPSYCH2012 sample including 30,000 randomly samples cohort members enabled us the unique possibility to introduce a large case cohort sample superior to previous studies while allowing inclusion of previously genotyped samples. Although, inclusion of

previously genotyped cases is methodologically straightforward, inclusion of previously genotyped controls is challenging as they are conditioned to be alive and without schizophrenia at different points in time (when then corresponding case was first diagnosed). The iPSYCH2012 case-cohort sample provides the rare opportunity to analyse genetic data and environmental data respecting the longitudinal nature of human life(10).

iPSYCH₂₀₁₅

Recruitment

The iPSYCH2015 Case-Cohort expands on the design used in the iPSYCH2012 version¹⁷. Recruitment is done through the use of unique Central Person Register (CPR) numbers. All residents in Denmark are assigned a CPR-number either upon birth or at the time of immigration to Denmark. CPR is linked to the national population-wide registries (e.g. Danish Psychiatric Central Research registry) to recruit participants¹⁷. We cross-referenced CPR with the availability of dried blood spot (DBS) in the Danish Neonatal Screening Biobank (DNSB). The DNSB has been storing residual DBS since May 1981 and encompasses close to all births in Denmark since then. While the DNSB is a screening Biobank with primary activities centred around new-born screening, ancillary and secondary purposes, including research, are allowed provided that relevant authorities have granted approval⁴⁴. Once the initial selection and data

generation is completed, the CPR-number allows continuous updates of information from population-based registers through pseudonymization.

Acknowledgements & funding

The Authors would like to thank Vivek Appadurai for the base calculations for the PCA plot. This study was supported by The Lundbeck Foundation, the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project number 294838) and Centre for Integrated Register based Research at Aarhus University. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation.

This study was supported by The Lundbeck Foundation, the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project number 294838) and Centre for Integrated Register based Research at Aarhus University. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation.

FinnGen

Recruitment

The FinnGen study is a nationwide cohort of Finnish individuals, combining genome information with digital health care and registry data. The FinnGen cohort consists of (a)

legacy samples recruited by specific research projects, collected based on study-specific consents (Supplementary Table 2) and later transferred to a biobank according to the Finnish Biobank Act 13 § (mainly from the Biobank of the Finnish Institute for Health and Welfare (THL)), and (b) prospective biobank samples recruited by Finnish biobanks, collected based on a wide biobank consent, according to the Finnish Biobank Act 11 §. Prospective samples in the FinnGen cohort originate from three complementary approaches: (i) samples from six regional hospital biobanks represent a vast variety of patients enrolled in specialized health care, (ii) samples from a private health care biobank allow for targeted enrichment of the FinnGen cohort with patients underrepresented in specialized health care, and (iii) participants recruited through Blood Service Biobank enrich the cohort with healthier individuals. In the present study, we included samples from altogether 224,737 participants (FinnGen release 5, August 2020) from the following biobanks: Auria Biobank, Helsinki Biobank, Biobank of Eastern Finland, Central Finland Biobank, Northern Finland Biobank Borealis, Finnish Clinical Biobank Tampere, THL Biobank, Blood Service Biobank, and Terveystalo Biobank.

Various means were utilized for informing individuals about ongoing biobank recruitment, including active informing by hospital personnel, locally distributed leaflets, national advertising campaigns on TV and in social media, and dedicated biobank research nurses operating in hospital districts. Information about the FinnGen research project was mostly distributed to the public in news articles, interviews, leaflets and on

the FinnGen website. Prospective samples collected by the hospital biobanks and the Blood Service Biobank were collected without any specific prioritization. All individuals volunteering to provide a biobank consent and to donate a biobank sample to a Finnish biobank were eligible for the FinnGen study Biobank recruitment and sampling protocols followed the protocols approved by the Finnish Medicines Agency (Fimea), the organization with the legal responsibility for supervising biobank operations in Finland. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol Nr HUS/990/2017.

DNA samples were obtained from participating biobanks. DNA was extracted either from whole blood or buffy coat according to the standard procedure of each biobank, and sent to a centralized logistics laboratory in the THL Biobank. THL Biobank laboratory aliquoted and normalized all samples in 96-well plates in 60 µl volume and 15 ng/µl concentration, and delivered the plated samples to Affymetrix Research Services Laboratory, Thermo Fisher Scientific (San Diego, USA).

Acknowledgements & funding

We thank all those who contributed samples and data for the FinnGen scientific project; and P. VandeHaar for technical consultation on PheWeb. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie, AstraZeneca UK, Biogen, Bristol Myers Squibb (and Celgene Corporation & Celgene International II), Genentech, Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., Rahway, NJ, USA, Pfizer,

GlaxoSmithKline Intellectual Property Development, Sanofi US Services, Maze Therapeutics, Janssen Biotech, Novartis, and Boehringer Ingelheim. The following biobanks are acknowledged for delivering samples to FinnGen: Auria Biobank (<https://www.auria.fi/biopankki/>), THL Biobank (<https://www.thl.fi/biobank>), Helsinki Biobank (<https://www.helsinginbiopankki.fi>), Biobank Borealis of Northern Finland (<https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere (https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (<https://www.ita-suomenbiopankki.fi/en>), Central Finland Biobank (<https://www.ksshP.fi/fi-FI/Potilaalle/Biopankki>), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (<https://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/>). All Finnish biobanks are members of the BBMRI.fi infrastructure (<https://www.bbMRI.fi>). The FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious services (<https://site.fingenious.fi/en/>) managed by FINBB. We acknowledge the BBMRI.fi for creating the original biobank map used in Fig. [Fig. 1a](#) and [1a](#) and providing us their permission to use it.

Open Access funding provided by University of Helsinki including Helsinki University Central Hospital.

deCODE

Recruitment

This study is based on whole-genome sequence data from the white blood cells of 2,636 Icelanders participating in various disease projects at deCODE Genetics ([Supplementary Tables 1](#) and [2](#)). In addition, a total of 104,220 Icelanders have been genotyped using Illumina SNP chips ([Supplementary Table 6](#)).

All participating individuals, or their guardians, gave their informed consent before blood samples were drawn. The family history of participants donating blood was incorporated into the study by including the phenotypes of first- and second-degree relatives and integrating over their possible genotypes. This integration is performed without the genotypes being stored.

All sample identifiers were encrypted in accordance with the regulations of the Icelandic Data Protection Authority. Approval for these studies was provided by the National Bioethics Committee and the Icelandic Data Protection Authority.

Acknowledgements

We thank all the participants in this study. This study was performed in collaboration with Illumina.

Funding

DBDS

Recruitment

The DBDS is an ongoing, national prospective cohort of blood donors. Before or during donation, blood donors aged 18 years or above who have previously donated blood at least once receive oral and written information about the study and are invited to join the cohort by signing a consent form. The participants can withdraw consent at any time; see [Table 1](#) for characteristics of participants (at baseline) and non-participants in the study. In the current dataset (28 February 2022), 138491 participants have completed a total of 371589 questionnaires ([Table 2](#)). To define the population of blood donors not included in DBDS, we identified all donations from non-participants (not yet invited blood donors) from 1 March 2010 to 28 February 2022 from SCANDAT. A single donation from each non-participating donor was sampled randomly as the index donation for that donor, and the age and measurements at that donation were used to characterize the donor. By this method, participants were slightly older, had more donations in the previous 5 years and had similar haemoglobin values compared with

non-participants. The number of blood donors who have declined to participate upon invitation is 4.0%.

Whenever a participant donates blood, the data from the donation becomes available for the study. The possibility for measuring biomarkers in consecutive samples from the same donor allows for detailed studies of biomarker dynamics. This is, to our knowledge, the most comprehensive platform for collecting sequential samples in any cohort study. By February 2022, a total of 2293916 unique plasma samples were available ([Table 2](#)).

After enrolment, the participants are followed at subsequent blood donations and through national health registers. The blood bank follow-up allows for additional data collection: the donor may be asked to fill out new questionnaires and biological sampling may take place. The participants are followed through the national registers also after ceasing their blood donor career. Outcomes, exposures and confounding factors can be assessed by register linkage via the unique personal identification number to a collection of nationwide registers regarding sociodemographics (education, employment, income, civil status, vital status, history of emigration) and health (filled prescriptions, general practitioner contacts, inpatient and outpatient hospital contacts, causes of death) as outlined elsewhere.¹⁶ Loss to follow-up in the traditional sense does not occur.

Acknowledgements

We thank the Danish blood donors for their valuable participation in the Danish Blood Donor Study, and the staff at the blood centres for making this study possible.

Funding

The initiation of the Danish Blood Donor Study was supported by the Danish Administrative Regions (02/2611) and the Danish Council for Independent Research (09–069412). The study is currently funded by the Danish Administrative Regions and Bio- and Genome Bank Denmark. Funding for specific projects has been received from the Danish Blood Donor Research Foundation, the A. P. Møller Foundation for the Advancement of Medical Science and CHALLENGE grants from the Novo Nordisk Foundation [NNF17OC0027864 (BERTHA) and NNF17OC0027594].

MoBa

Recruitment

Informed consent was obtained from all study participants. The administrative board of the Norwegian Mother, Father, and Child Cohort Study led by the Norwegian Institute of Public Health approved the study protocol. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committee for Medical Research Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The study was approved by

The Regional Committee for Medical Research Ethics (#2012/67).

The Norwegian Mother, Father, and Child Cohort Study is an open-ended cohort study that recruited pregnant women in Norway from 1999 to 2008. Approximately 114,000 children, 95,000 mothers, and 75,000 fathers of predominantly Norwegian ancestry were enrolled in the study from 50 hospitals all across Norway⁷. Anthropometric measurements of the children were carried out at hospitals (at birth) and during routine visits by trained nurses at 6 weeks; 3, 6, and 8 months; and 1, 1.5, 2, 3, 5, 7, and 8 years of age. Parents later transcribed these measurements to questionnaires. In 2012, the project Better Health By Harvesting Biobanks (HARVEST) randomly selected 11,490 umbilical cord blood DNA samples from the Norwegian Mother, Father, and Child Cohort Study's biobank for genotyping, excluding samples matching any of the following criteria: (1) stillborn, (2) deceased, (3) twins, (4) non-existing Medical Birth Registry data, (5) missing anthropometric measurements at birth in Medical Birth Registry, (6) pregnancies where the mother did not answer the first questionnaire (as a proxy for higher fallout rate), and (7) missing parental DNA samples. In 2016, HARVEST randomly selected a second set of samples, 5984, using the same criteria.

Acknowledgements

This work was supported by grants (to P.R.N.) from the European Research Council (AdG #293574), the Bergen Research Foundation ("Utilizing the Mother and Child Cohort and the Medical Birth Registry for Better Health"), Stiftelsen Kristian Gerhard Jebsen (Translational Medical Center), the University of Bergen, the Research Council

of Norway (FRIPRO grant #240413), the Western Norway Regional Health Authority (Strategic Fund “Personalized Medicine for Children and Adults”), the Novo Nordisk Foundation (grant #54741), and the Norwegian Diabetes Association; and (to S.J.) Helse Vest’s Open Research Grant (grant #912250). This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme (#262700), Better Health by Harvesting Biobanks (#229624) and The Swedish Research Council, Stockholm, Sweden (2015-02559), The Research Council of Norway, Oslo, Norway (FRIMEDBIO #547711, March of Dimes (#21-FY16-121). The Norwegian Mother, Father, and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research, NIH/NIEHS (contract no N01-ES-75558), NIH/NINDS (grant no.1 UO1 NS 047537-01 and grant no.2 UO1 NS 047537-06A1). We are grateful to all the families in Norway who are taking part in this ongoing cohort study.

Funding

Ø.H. and M.V. performed the analyses. O.L., J.J., J.B., B.J., H.L., K.H., R.T.L., G.P.K., C.S., and P.M. contributed to sample acquisition and genotyping. J.J. and J.B. assisted with genotype quality control. Ø.H., M.V., S.J., and P.R.N. wrote the manuscript with contributions from all authors. P.B.J., J.V.S., and A.M. critically revised the manuscript for important intellectual content. Ø.H., M.V., S.J., and P.R.N. designed the study. S.J. and P.R.N. directed the study. P.R.N. conceived the project, secured funding, and initiated the study.

COPSAC 2000

Recruitment

The COPSAC2000 birth cohort study is a prospective clinical study of a birth cohort of 411 infants born to mothers with a history of asthma. The newborns were enrolled at the age of 1 month, the recruitment of which was previously described in detail.²⁵ The study was approved by the Ethics Committee for Copenhagen (KF 01- 289/96) and The Danish Data Protection Agency (2008-41-1754) and informed consent was obtained from both parents. For this study, only the parents were included.

Acknowledgements and funding

COPSAC is funded by private and public research funds all listed on www.copsac.com. The Lundbeck Foundation (Grant no R16-A1694); The Danish Ministry of Health (Grant no 903516); Danish Council for Strategic Research (Grant no 0603-00280B); The Danish Council for Independent Research (Grant no 10-082884 and 271-08-0815); The Capital Region Research Foundation (No grant number) and NIH-NHLBI R01 HL129735 have provided core support for COPSAC.

COPSAC 2010

Recruitment

Pregnant women from eastern Denmark were invited during 2008–2010 to a novel unselected 'COPSAC₂₀₁₀' cohort. The women visited the clinic during pregnancy weeks 24 and 36. Their children were followed at the clinic with deep phenotyping and collection of biological samples at nine regular visits until the age of 3 and at acute symptoms. Randomized controlled trials of high-dose vitamin D and fish oil supplements were conducted during pregnancy, and a trial of azithromycin for acute lung symptoms was conducted in the children with recurrent wheeze.

Acknowledgements and funding

The authors gratefully express their gratitude to the children and families of the COPSAC cohort studies for all their support and commitment, and we acknowledge and appreciate the unique efforts of the COPSAC research team. COPSAC is funded by private and public research funds. The Lundbeck Foundation; the Pharmacy Foundation of 1991; Augustinus Foundation; the Danish Medical Research Council; and The Danish Pediatric Asthma Centre provided the core support for the study.

MARC-35

Recruitment

We conducted a multicentre prospective cohort study of infants (age < 1 year) hospitalised for bronchiolitis (severe bronchiolitis). This study, called the 35th Multicenter Airway Research Collaboration (MARC-35) [17], was coordinated by the

Emergency Medicine Network (EMNet, <http://www.emnet-usa.org/>), a collaboration of 235 participating hospitals. Using a standardised protocol, site investigators at 17 sites across 14 US states (supplementary table E1) enrolled infants hospitalised with an attending physician diagnosis of bronchiolitis during three consecutive bronchiolitis seasons from November 1, to April 30, 2011–2014. Bronchiolitis was defined by the American Academy of Pediatrics guidelines as acute respiratory illness with some combination of rhinitis, cough, tachypnoea, wheezing, crackles and retractions [18]. We excluded infants with previous enrolment; those who were transferred to a participating hospital >24 h after the original hospitalisation; those who were consented >24 h after hospitalisation; or those with known heart-lung disease, immunodeficiency, immunosuppression or gestational age <32 weeks. All patients were treated at the discretion of the treating physician. The institutional review board at each of the participating hospitals approved the study. Written informed consent was obtained from the parent or guardian.

The primary outcome was intensive care use, defined as admission to the intensive care unit and/or use of mechanical ventilation (continuous positive airway pressure and/or intubation during inpatient stay, regardless of location) at any time during the index hospitalisation [19, 28]. The secondary outcome was hospital length of stay of ≥ 3 days, defined using the median length of stay of 2 days and similar to the approach used in prior studies [4, 19].

With the exception of specimen collection (the NPA, blood, and nasal swab), all study participants were evaluated and treated as usual and without regard to this observational study. Parent/legal guardians were approached about participating after the medical team had finished their assessments and stabilized the study participant. If necessary, the recruiting took place the morning after admission, but no later than 24 h after admission to the ward or intensive care unit. Samples and information of the participants were used to study the possible genetic causes of severe bronchiolitis, recurrent wheezing, asthma, and related concepts. The institutional review board at each of the participating hospitals approved the study. Written informed consent was obtained from the parent or guardian.

Acknowledgements & funding

We thank Pamela Luna at Rice University (Houston, TX, USA) for her assistance with statistical analysis. The following individuals were MARC-35 site investigators and are collaborators on this paper. Anne K. Beasley, Phoenix Children's Hospital, Phoenix, AZ, USA; Juan C. Celedon, Children's Hospital of Pittsburgh, Pittsburgh, PA, USA; Ari R. Cohen, Massachusetts General Hospital, Boston, MA, USA; Michelle B. Dunn, Children's Hospital of Philadelphia, Philadelphia, PA, USA; Michael Gomez, Children's Hospital at St. Francis, Tulsa, OK, USA; Nancy Inhofe, Children's Hospital at St. Francis, Tulsa, OK, USA; Sujit Iyer, Dell Children's Medical Center of Central Texas, Austin, TX, USA; Federico R. Laham, Arnold Palmer Hospital for Children, Orlando, FL,

USA; Charles G. Macias, Texas Children's Hospital, Houston, TX, USA; Thida Ong, Seattle Children's Hospital, Seattle, WA, US; Brian M. Pate, The Children's Mercy Hospital & Clinics, Kansas City, MO, USA; Wayne G. Schreffler, Massachusetts General Hospital, Boston, MA, USA; Michelle D. Stevenson, Kosair Children's Hospital, Louisville, KY, USA; Richard T. Strait, Cincinnati Children's Hospital and Medical Center, Cincinnati, OH, USA; Stephen J. Teach, Children's National Medical Center, Washington, DC, USA; Henry T. Puls, The Children's Mercy Hospital & Clinics, Kansas City, MO, US; Amy D. Thompson, Alfred I. duPont Hospital for Children, Wilmington, DE, USA; Vincent J. Wang, Children's Hospital of Los Angeles, Los Angeles, CA, USA; and Ilana Waynik, MD, Connecticut Children's Medical Center, Hartford, CT, USA.

Wheeze, asthma and infection phenotypes

iPSYCH₂₀₁₂ + iPSYCH₂₀₁₅

In the iPSYCH study, severe episodes of wheeze/asthma requiring hospitalization were defined using longitudinal data from the Danish National Patient Registry (1). Individuals with at least one hospitalization for a primary diagnosis before age six of wheezing disorders—such as asthma (ICD-8 code: 493; ICD-10 codes: J45-46), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), bronchiolitis (ICD-10 code: J21) or RSV-bronchitis/bronchiolitis (ICD-10 codes: J205, J210)—were classified

as severe wheeze cases. Those without such hospitalizations in that age gap were used as controls.

Lower respiratory tract infections (LRTI) defined from pneumonia (ICD-8: 471, 480, 481, 482, 483, 484, 485, 486, ICD-10: J100, J110, J12, J13, J14, J15, J16, J17, J18) were analyzed to characterize identified wheeze loci. This analysis included 5672 cases and 108,611 controls.

FinnGen

Individuals with at least one hospitalization for a primary diagnosis before age six of wheezing disorders—such as bronchiolitis (ICD-8 code: 913, 390), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), or asthma (ICD-8 code: 493; ICD-10 codes: J45-46)—were classified as severe wheeze cases. Those without such hospitalizations in that age gap were used as controls.

deCODE

Individuals with at least one hospitalization for a primary diagnosis before age seven of asthma (ICD-8 code: 493; ICD-10 codes: J45-46)—were classified as cases. Those without such hospitalizations in that age gap were used as controls.

DBDS

Individuals with at least one hospitalization for a primary diagnosis before age six of wheezing disorders—such as bronchiolitis (ICD-8 code: 913, 390), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), or asthma (ICD-8 code: 493; ICD-10 codes: J45-46)—were classified as severe wheeze cases. Those without such hospitalizations in that age gap were used as controls.

COPSAC 2000 + 2010

Asthma was diagnosed longitudinally by research physicians in the COPSAC birth cohorts using a standardized treatment algorithm. The diagnosis required a specific burden of symptoms documented in diaries, a response to treatment with inhaled corticosteroids, and a relapse after withdrawing treatment. Symptom diaries were collected and evaluated by trained physicians during scheduled clinic visits.

MARC-35

The definition of asthma in the MARC-35 study was defined by physician-diagnosis of asthma by age 5 years, including asthma medication use (e.g., albuterol inhaler, inhaled corticosteroids, montelukast) or asthma-related symptoms in the previous year (2).

Breathing problems were defined as a child experiencing a cough that disrupts sleep or causes vomiting, or when the child has wheezing or shortness of breath without a cough (3,4).

MoBa

We utilized the detailed early-life questionnaire from the MoBa cohort coupled with genotype data to replicate findings of our GWAS study (5). We defined asthma and wheeze from 0-7 years by using questionnaires at 6-months, 18-months, 3-year, 5-year and 7-year questionnaires.

Children were included in the case group if the parents had answered yes to the following questions in the 6-month questionnaire: *“DD245: Has had asthma?”* or *“DD937-DD939: Has you your child had asthma? Did you go to a doctor/clinic for this? Been admitted to hospital?”* of *“DD931-DD933: Has you your child had RSV/bronchitis? Did you go to a doctor/clinic for this? Been admitted to hospital?”*.

Children were included in the case group if the parents had answered yes to the following questions in the 18-month questionnaire: *“EE824-EE825: Yes, has asthma now or has had previously”* or *“EE197: has asthma now”* or *“EE299: Had symptoms of wheeze/whistling in the chest”*.

Children were included in the case group if the parents had answered yes to the following questions in the 3-year questionnaire: *“GG70-GG71: Child has or has had*

asthma?” or “GG143: *Child has had bronchitis between 8 and 18 months*”. The 3-year questionnaire contains no questions about wheeze specifically.

Children were included in the case group if the parents had answered yes to the following questions in the 5-year questionnaire: “LL58-LL60: *Child has had asthma AND it is confirmed by a doctor*” or “LL64-LL66: *Child has experienced wheeze symptoms AND they were confirmed by a doctor*”.

Children were included in the case group if the parents had answered yes to the following questions in the 7-year questionnaire: “JJ623, JJ16, JJ12, JJ169: *Child has had asthma AND it is confirmed by a doctor*” or “JJ243: *Child has had bronchitis AND it is confirmed by a doctor*” or “JJ635: *Child has had wheeze symptoms*”.

Whenever possible, confirmation by a physician was used as an inclusion criterion. We included 19,621 cases following the above criteria, and 29,134 controls. Controls were defined as childrens whose parents had answered no to all questions regarding asthma or wheezing.

Descriptions of novel GWAS loci

KAZN-AS1

The lead SNP in this locus has never before been associated with any trait at the GSW ($P < 5e-8$), according to OpenTargets or GWAS Catalog. *KAZN-AS1* is a long non-coding RNA which is antisense-RNA of the *KAZN* gene, indicating a likely role of

translational regulation. The gene *KAZN* (Kazrin) has previously been shown to be involved in transcriptomic signatures of bronchial epithelial cells infected with respiratory syncytial virus (RSV) (80). However, very little evidence links this *KAZN* to the lead SNP, since the prioritization is only based on physical proximity. No other genes were prioritized by other features at this locus.

NR4A2

The risk allele in Nuclear receptor 4A2 (*NR4A2*), also known as *NURR1* (nuclear receptor related 1 protein), has previously been shown to increase adult lung function (FEV1/FVC, $b=0.033$, (CI = 0.027 - 0.038) $P=2e-21$) (43). *NR4A2* is an inhibitor of NF- κ B-driven inflammation and is expressed in myeloid cells (80,81). Splicing and stability of *NRA4A2* has been shown to be inhibited by non-coding RNA *LUTAC1*, which is upregulated in COPD patients (82). The distance to the nearest gene was the feature determining this gene prioritization.

THSD4

Previous studies have linked other variants in *THSD4* with ICS response in asthma and asthma severity. Response to inhaled corticosteroids have been shown to be influenced by genetic variants in *THSD4*, however the reported variant is not in LD with the finding in the present study (rs62015884, $r^2 = 0.01$) (84). Another study used whole-genome sequencing data to characterize moderate to severe asthma risk in a population with a

mix of children and adults identified a genetic variant, rs11631778, near the *THSD4* gene, which was also not in LD with our lead variant ($r^2 = 0.004$) (85). *THSD4* is prioritized by six out of twelve features and is therefore a likely candidate for a causal gene in the locus. *THSD4* encodes *ADAMTSL6*, a member of the ADAMTS/L superfamily, which aids formation of extracellular matrices by promoting fibrillin-1 assembly and genetic variants in *THSD4* has been connected to risk of aortic aneurysms (86). Additionally, abnormal fibrillin-1 levels have been proposed to play an important role in development of lung emphysema by decreasing elasticity of the connective tissue in the lung (87). This suggests that *THSD4* may confer asthma risk by compromising the structural integrity of the extracellular matrix.

HNF1B

HNF1B is prioritized by six out of twelve features and provides an avenue for understanding this locus. The *HNF1B* gene has been connected to monogenic subtypes of diabetes called maturity-onset diabetes of the young (MODY), also known as *HNF1B*-MODY or MODY-5 highlighting a the role of *HNF1B* in metabolic diseases. The risk allele of the *HNF1B* SNP (rs7405696-G) is correlated ($r^2 = 0.69$) with rs11651755-C that increases risk of type-2 diabetes (OR = 1.06, (CI = 0.06 - 0.07) $P = 1e-67$) derived from a GWAS of type-2 diabetes with 1.4 million participants (80). *HNF1B* has previously been linked to respiratory diseases in a genome-wide DNA methylation study in lung tissue, showing cg04136369 near *HFN1B* is differentially associated between

COPD and non-COPD patients (81). Concordantly, we find that the lead SNP in this locus likely drives methylation of *HNF1B* in nasal epithelial cells (see Table 2).

Furthermore, a study showed that asthmatics had lower levels of microRNA miR-199b-5p, in bronchial biopsies. The study also shows that miR-199b-5p is able to target *HNF1B* (82). Thus, *HNF1B* might present a novel avenue of research to investigate the link between childhood asthma and metabolic disorders.

CASC17

CASC17 is a long non-coding RNA (lncRNA) with little evidence from the literature linking this to any respiratory or immunological outcomes. An intronic variant in *CASC17* has previously been associated with low lung function (43) and prostate cancer (83). *CASC17* is prioritized by two features, namely physical distance and intron-mapping.

ZNF529

ZNF529 (Zinc Finger Protein 529) is a relatively understudied gene. *ZNF529* is a protein-coding gene involved in transcriptional regulation. It is predicted to function as a DNA-binding transcription repressor, specifically interacting with RNA polymerase II. *ZNF529* plays a role in the negative regulation of transcription by RNA polymerase II and is typically located in the nucleus.

CASC20

CASC20 is an RNA-gene with unknown function, and the gene prioritization predicted *CASC20* using physical distance and intron-mapping. OpenTargets predict *BMP2*, but this is likely because OpenTargets do not include RNA-genes. *BMP2* is involved in bone development and adipose metabolism (88,89) and the lead SNP (rs928202) in the *CASC20*-locus has previously been associated with height ($b = -0.031$, $se = 0.001$, $P = 1.9e-183$) (46), indicating a role in growth.

GDF5

GDF5 (Growth Differentiation Factor 5) is the top gene at this locus, with *UQCC1* (ubiquinol-cytochrome c reductase complex assembly factor 1) as a strong alternative, supported by the TWAS results. The rs143384 variant likely impacts *UQCC1* levels in nasal epithelium, but *GDF5* RNA levels couldn't be measured in COPSAC₂₀₁₀. However, a missense variant in *GDF5*, in high LD ($r^2 > 0.6$), supports *GDF5* as the primary effector gene in this locus. *GDF5* is a member of the Bone Morphogenic Protein (BMP) family. BMPs play key roles in development, particularly in organogenesis and differentiation (77). *GDF5*, has been shown to promote generation of brown fat through the induction of *PGC1 α* expression in adipose tissue (78), suggesting a role in body development and fat metabolism. *UQCC1* involves generation of the majority of cellular ATP through mitochondrial oxidative phosphorylation. Specifically, *UQCC1* is involved in assembly of mitochondrial complex III (79), indicating a housekeeping role of this gene. The lead SNP in this locus rs143384 is highly associated with *UQCC1* levels in

fibroblasts (NES = 0.51, $P = 1.6e-50$) and negatively associated with height ($b = -0.079$, $se = 0.001$, $P = 3.4e-1152$), further indicating a role in growth.

AlphaMissense

AlphaMissense utilizes an AlphaFold-derived model to compute pathogenicity scores for each variant in the human proteome. We excluded all synonymous missense variants in this analysis. The pathogenicity score of a variant reflects the structural, functional and evolutionary conservation of a protein domain, indicating that the amino acid substitution occurs in a site with important biological functions (6,7). If multiple genes in a locus were identified, the gene with the highest pathogenicity score and $R^2 > 0.6$ was prioritized.

DEPICT

DEPICT uses 14,461 reconstituted gene sets and genes are assigned to these sets by membership probabilities allowing clustering of genes based on how similar they are in terms of gene set membership. DEPICT defines a locus as all SNPs with $r^2 > 0.5$ with the lead SNP in a GWAS. Utilizing these definitions of gene sets and loci, DEPICT calculates enrichment for each of the reconstituted gene sets and selects genes that are similar in gene set membership across all the loci included in the analysis. Scores are

subsequently normalized for gene length and DEPICT empirically adjusts the false discovery rate (8).

MAGMA

MAGMA averages Chi-square statistics from GWAS summary results, in predefined boundaries to create a z-score, for the gene-phenotype association. MAGMA computes gene based p-values by approximating sampling distributions of the Chi-square statistic. LD is estimated by including a reference sample with similar ancestry as the GWAS (9).

PoPS

PoPS computes tau-score, by predicting MAGMA z-scores using 54,000 features through a penalized regressions approach in a leave-one-out-chromosome procedure, thereby prioritizing genes involved in pan-genome pathways (10).

Gene-prioritization scheme

Genes were scored according to how many features prioritized a given gene in a locus, i.e. allowing scores from 1 to 12. If it wasn't possible to select a consensus gene based on the majority-vote approach, the gene with the closest physical distance among the

equally prioritized genes was selected. UTR-variants and publicly available chromatin interaction data were unable to prioritize any genes in any loci, so they were not included. In the gene prioritization scheme, we chose a 10 kilobase window, since it has been shown that MAGMA retains similar power with a 10 kilobase window compared to no window (9). We extended this conclusion to define gene boundaries in DEPICT and PoPS for consistency and since PoPS's tau-score are based on predicted MAGMA's gene-based z-scores.

Drug-target interaction score

DGIdb stores an interaction score for each drug-gene pair that integrates scientific evidence with the observed frequency of a specific drug-gene pair. Information about clinical indication and approval status for each drug was retrieved from DrugBank (www.drugbank.ca).

In detail, the drug interaction score consists of three distinct components: the evidence score, the relative drug specificity, and the relative gene specificity. The evidence score is determined by aggregating the number of publications and sources that establish a connection between the gene and the drug. Relative drug specificity is calculated as the ratio of the average number of known gene partners for all drugs to the number of known partners for the specific drug in question. Similarly, gene specificity is assessed by dividing the average number of known drug partners for all genes by the number of

known drug partners for the specific gene. For additional details see <https://www.dgldb.org/about#interaction-scores>).

Association-based groupings of severe wheeze risk loci

If a proxy could not be identified we imputed the value using the median trait z-score. Since z-scores are highly correlated with sample size, hierarchical clustering was performed on z-scores normalized to zero-mean and unit-variance to account for differences in sample size among studies. We use k-means to identify the optimal number of splits by running k-means on the traits and variants and selecting the number of k with highest silhouette score. We included proxy variants ($r^2 > 0.6$) in the analysis and if a reliable proxy could not be found, we imputed using median-imputation. We subsequently compared clustering with imputed data containing no missing values. Clusters were visualized on unnormalized z-scores using the R-package *ComplexHeatmap* (11,12). For visualization purposes, z-scores were truncated between -10 to 10. The clustering was performed on scaled and un-truncated z-scores.

TWAS-pipeline (FUSION)

We used the FUSION software (13) to compute predictive models based on cis-QTL effect sizes. When predictive models have been computed, FUSION leverages the

ImpG algorithm (14), to impute genetically mediated expression and methylation levels on GWAS summary statistics. We carried out imputation by employing the single best eQTL through imputation using the *top1* model, as it represents the most straightforward and interpretable approach. The FUSION effect sizes should be interpreted as the covariance between GWAS and QTL effect sizes. The QTL-models were adjusted for the top 5 principal components for all genes. Genes that passed correction for multiple testing (Benjamini-Hochberg, FDR < 0.05) were validated by colocalization analysis to compute the posterior probability of a shared GWAS and eQTL signal (15). Prior to running the TWAS, summary statistics were also imputed with the ImpG algorithm to maximize overlap with QTL-models (14).

Supplementary tables

Supplementary Table 1: Cohort case-control numbers

Study	Phenotype	Time points	Cases	Controls	Total	Case %
iPSYCH	Severe wheeze	0 - 6 yr	9.026	103.479	112.605	8.10 %
FinnGen	Severe wheeze	0 - 6 yr	6.137	110.625	116.762	5.26%
deCODE	Severe asthma	0 - 7 yr	2.671	29.505	32.176	8.30 %
DBDS	Severe wheeze	0 - 6 yr	1.627	41.022	42.649	3.81 %
Total	Severe wheeze	0 - 7 yr	19.461	284.631	304.092	6.40%

Supplementary Table 2: Fine-mapped variants using SuSIE by in-sample fine-mapping in the iPSYCH cohorts

Chr	Position	SNP	EAF	EA/NEA	OR	CI 95 %	P-value	Consensus gene
17	38062196	rs2305480	0.4797	G/A	1.32	(1.29 - 1.35)	2.70E-124	<i>GSDMB</i>
7	105658451	rs6967330	0.2112	A/G	1.22	(1.19 - 1.26)	1.78E-47	<i>CDHR3</i>
9	6213387	rs928413	0.7495	G/A	1.17	(1.14 - 1.2)	1.12E-33	<i>IL33-2</i>
6	31009055	rs2517546	0.1259	T/C	1.25	(1.2 - 1.3)	9.81E-29	<i>HCG22</i>
9	6114076	rs439190	0.7261	T/C	1.13	(1.1 - 1.16)	1.09E-19	<i>IL33-1</i>
5	110427795	rs79881201	0.3358	T/C	1.1	(1.07 - 1.13)	7.63E-15	<i>WDR36</i>
6	19842661	rs9350191	0.8538	T/C	1.14	(1.1 - 1.18)	7.75E-15	<i>ID4</i>
20	34025756	rs143384	0.5648	A/G	1.08	(1.05 - 1.1)	1.48E-10	<i>GDF5</i>
15	71804587	rs62015884	0.177	T/C	1.09	(1.06 - 1.12)	2.53E-08	<i>THSD4</i>
19	49218111	rs33988101	0.4824	G/T	1.07	(1.04 - 1.09)	5.65E-08	<i>MAMSTR</i>
17	36102035	rs7405696	0.4858	G/C	1.06	(1.04 - 1.09)	1.99E-07	<i>HNF1B</i>

Supplementary Table 3: strongest eQTL tissue for GWAS findings in GTEx and trait from OpenTargets and GWAS catalog

Chr	Position	SNP	EAF	EA/NEA	OR	CI 95 %	P-value	Consensus gene	Tissue	eQTL gene (Tissue)	Trait
17	38062196	rs2305480	0.4797	G/A	1.32	(1.29 - 1.35)	2.70E-124	<i>GSDMB</i>	Whole Blood	<i>GSDMB</i>	White blood cell count
7	105658451	rs6967330	0.2112	A/G	1.22	(1.19 - 1.26)	1.78E-47	<i>CDHR3</i>	Nerve - tibial	RP4-568B10.1	Childhood asthma with severe exacerbations
9	6213387	rs928413	0.7495	G/A	1.17	(1.14 - 1.2)	1.12E-33	<i>IL33-2</i>	Artery - Aorta	<i>IL33</i>	Eosinophil counts
6	31009055	rs2517546	0.1259	T/C	1.25	(1.2 - 1.3)	9.81E-29	<i>HCG22</i>	Muscle - skeletal	<i>HLA-C</i>	Celiac disease
9	6114076	rs439190	0.7261	T/C	1.13	(1.1 - 1.16)	1.09E-19	<i>IL33-1</i>	Cells - cultured fibroblasts	<i>IL33</i>	Asthma
5	131996669	rs847	0.269	T/C	1.13	(1.1 - 1.16)	3.30E-19	<i>IL13</i>	Testis	<i>IL13</i>	Eosinophil counts
2	102955082	rs13408661	0.1398	G/A	1.16	(1.12 - 1.2)	7.54E-18	<i>IL18RAP</i>	Whole Blood	AC007278.3	Eosinophil counts
5	110427795	rs79881201	0.3358	T/C	1.1	(1.07 - 1.13)	7.63E-15	<i>WDR36</i>	Testis	<i>TSLP</i>	Eosinophil counts
6	19842661	rs9350191	0.8538	T/C	1.14	(1.1 - 1.18)	7.75E-15	<i>ID4</i>	-	-	FEV1
9	136151806	rs600038	0.7852	C/T	1.1	(1.07 - 1.13)	9.94E-12	<i>ABO</i>	Whole Blood	<i>ABO</i>	Haemoglobin concentration
20	6470265	rs928202	0.6057	G/A	1.08	(1.06 - 1.11)	1.85E-11	<i>CASC20</i>	Cells - Cultured fibroblasts	<i>BMP2</i>	Standing height
16	11277358	rs11644510	0.3444	C/T	1.08	(1.06 - 1.11)	1.33E-10	<i>CLEC16A</i>	Thyroid	RP11-396B14.2	Eosinophil counts
20	34025756	rs143384	0.5648	A/G	1.08	(1.05 - 1.1)	1.48E-10	<i>UQC1</i>	Cells - Cultured fibroblasts	<i>UQC1</i>	Standing height
17	47461433	rs17637472	0.4148	A/G	1.08	(1.05 - 1.1)	5.17E-10	<i>PHB1</i>	Cells - Cultured fibroblasts	ZNF652	Sex hormone-binding globulin levels adjusted for BMI
17	69191316	rs1008348	0.4256	A/G	1.07	(1.05 - 1.1)	2.38E-09	<i>CASC17</i>	-	-	PEF
2	157083697	rs72907878	0.1271	G/A	1.11	(1.07 - 1.15)	4.45E-09	<i>NR4A2</i>	Esophagus muscularis	LINC01876	FEV1
4	106819053	rs34712979	0.2563	A/G	1.08	(1.05 - 1.11)	7.35E-09	<i>NPNT</i>	Thyroid	RP11-710F7.2	FEV1/FVC
15	71804587	rs62015884	0.177	T/C	1.09	(1.06 - 1.12)	2.53E-08	<i>THSD4</i>	Adipose - Subcutaneous	RP11-100TQ24.2	FEV1/FVC
1	14440292	rs12026570	0.7644	A/G	1.08	(1.05 - 1.11)	3.18E-08	<i>KAZN-AS1</i>	-	-	Sodium in urine
19	37048117	rs3098616	0.2866	C/T	1.07	(1.05 - 1.1)	3.20E-08	<i>ZNF529</i>	Thyroid	CTD-2162K18.4	Impedance of whole body
10	8939198	rs78539108	0.8656	T/A	1.1	(1.06 - 1.13)	3.53E-08	<i>LINC02676</i>	-	-	Basal cell carcinoma
19	49218111	rs33988101	0.4824	G/T	1.07	(1.04 - 1.09)	5.65E-08	<i>MAMSTR</i>	Esophagus - Mucosa	<i>FUT2</i>	Serum alkaline phosphatase levels
17	36102035	rs7405696	0.4858	G/C	1.06	(1.04 - 1.09)	1.99E-07	<i>HNF1B</i>	-	-	Estimated glomerular filtration rate

Supplementary Table 5: SNP-based heritability estimates for each cohort separately

	Meta-analysis	iPSYCH12	iPSYCH15	FinnGen	DBDS	deCODE
Heritability (SE)	0.1266 (0.0211)	0.1785 (0.0315)	0.1315 (0.0384)	0.1822 (0.0468)	0.1819 (0.07)	0.0519 (0.0449)
λ GC	1,105	1,0649	1,0466	1,1113	1,0375	1,0046
Mean chi2	1,184	1,0791	1,0459	1,1429	1,0418	1,0117
Intercept (SE)	0.9836 (0.0074)	0.9985 (0.0062)	1.008 (0.0065)	1.0456 (0.009)	1.0167 (0.0068)	1.0005 (0.007)

Supplementary Table 6: Genetic correlation (rg) results for the 0-7-year phenotype.

The trait column contains the name of the trait that is correlated with the severe wheeze GWAS. The genetic correlation with standard error, p-value and Benjamin-Hochberg correction is shown in the remaining columns. Traits are sorted according to lowest p-value starting from the top.

Trait	Genetic correlation	Standard Error	P-value	Benjamini-Hochberg
Asthma ever	0.6314	0.03394	3.09E-77	8.65E-76
Asthma 0-15 yr	0.66	0.05338	4.01E-35	5.61E-34
Asthma 25-65 yr	0.5331	0.06446	1.34E-16	1.25E-15
PEF	-0.2755	0.03374	3.26E-16	2.28E-15
FEV1	-0.2894	0.03592	7.87E-16	4.41E-15
FEV1/FVC	-0.2673	0.03427	6.26E-15	2.92E-14
Atopic dermatitis	0.3486	0.06149	1.43E-08	5.73E-08
FVC	-0.1708	0.03231	1.25E-07	4.39E-07
BMI	0.1624	0.03288	7.90E-07	2.46E-06
Eosinophil count	0.2286	0.05469	2.92E-05	8.18E-05
Allergic rhinitis	0.2752	0.07596	0.0002915	0.000742
Allergic sensitization	0.2679	0.07515	0.0003642	0.0008498
Height	-0.07702	0.02487	0.001957	0.004215077
C-reactive protein	0.117	0.03961	0.003143	0.006286
Lymphocyte count	0.07064	0.03499	0.04352	0.081237333
Sphingomyelin	-0.0716	0.05232	0.1711	0.278444444
Basophil count	0.1096	0.08155	0.179	0.278444444
Neutrophil count	0.07034	0.0609	0.2481	0.365621053
Monocyte count	0.0291	0.02869	0.3104	0.43456

Supplementary table 7: Genetic correlation (rg) results of the 0-3-year phenotype.

Trait	Genetic correlation	Standard Error	P-value	Benjamini-Hochberg
Asthma ever	0.5785	0.04004	2.64E-47	7.39E-46
Asthma 0-15 yr	0.6157	0.06553	5.67E-21	7.93E-20
PEF	-0.2872	0.03535	4.48E-16	4.18E-15
FEV1	-0.3169	0.03967	1.36E-15	9.53E-15
Asthma 25-65 yr	0.4828	0.06073	1.87E-15	1.05E-14
FEV1/FVC	-0.2822	0.03743	4.76E-14	2.22E-13
BMI	0.1725	0.03183	6.00E-08	2.40E-07
FVC	-0.1951	0.03651	9.05E-08	3.17E-07
Atopic dermatitis	0.3018	0.06067	6.56E-07	2.04E-06
Eosinophil count	0.219	0.05518	7.21E-05	0.00020174
Allergic rhinitis	0.2559	0.08446	0.002444	0.006221090909
Allergic sensitization	0.238	0.0825	0.003915	0.009135
Height	-0.07202	0.02674	0.007068	0.01522338462
C-reactive protein	0.09838	0.03981	0.01347	0.02694
Lymphocyte count	0.08619	0.03701	0.01985	0.03705333333
Basophil count	0.1118	0.0787	0.1554	0.2559529412
Monocyte count	0.039	0.0301	0.1951	0.3034888889
Neutrophil count	0.06849	0.05984	0.2524	0.3719578947
Sphingomyelin	-0.05964	0.05668	0.2926	0.40964

Supplementary Table 8: Results of GSDMB-epistatic analyses in iPSYCH₂₀₁₂ +

iPSYCH₂₀₁₅. Stratum signifies the genotype stratum the analysis was performed in.

GSDMB (2), GSDMB (1) and GSDMB (0) denotes 2, 1 or 0 risk alleles. OR refers to the main-effect of the SNP within each stratum.

Stratum	OR	CI 95 %	P	Cases/Controls
CDHR3				
<i>GSDMB</i> (All)	1,23	(1,18-1,28)	5,51E-26	9026/103479
<i>GSDMB</i> (2)	1,39	(1,30-1,48)	7,17E-24	3198/28976
<i>GSDMB</i> (1)	1,19	(1,12-1,26)	1,35E-09	4268/51876
<i>GSDMB</i> (0)	1,06	(0,96-1,16)	0,24	1560/22624
ID4				
<i>GSDMB</i> (All)	1,15	(1,10-1,21)	1,77E-09	9026/103479
<i>GSDMB</i> (2)	1,07	(0,98-1,15)	0,09	3198/28976
<i>GSDMB</i> (1)	1,17	(1,09-1,25)	6,50E-06	4268/51876
<i>GSDMB</i> (0)	1,3	(1,16-1,46)	6,29E-06	1560/22624
NPNT				
<i>GSDMB</i> (All)	1,08	(1,04-1,11)	4,99E-05	9026/103479
<i>GSDMB</i> (2)	1,18	(1,11-1,25)	2,47E-08	3198/28976
<i>GSDMB</i> (1)	1,01	(0,95-1,06)	0,83	4268/51876
<i>GSDMB</i> (0)	1,06	(0,97-1,15)	0,96	1560/22624
GSDMB				
All	1,28	(1,24-1,32)	2,73E-52	9026/103479
<i>CDHR3</i> (2)	1,43	(1,21-1,69)	2,74E-05	332/3134
<i>CDHR3</i> (1)	1,44	(1,36-1,52)	2,29E-37	2980/28950
<i>CDHR3</i> (0)	1,2	(1,15-1,24)	1,84E-10	5714/71395
<i>ID4</i> (2)	1,24	(1,19-1,28)	2,24E-31	6867/75959
<i>ID4</i> (1)	1,42	(1,32-1,52)	9,76E-25	2031/25462
<i>ID4</i> (0)	1,2	(0,92-1,57)	0,17	128/2058
<i>NPNT</i> (2)	1,4	(1,25-1,57)	1,76E-08	663/6638
<i>NPNT</i> (1)	1,33	(1,26-1,39)	7,00E-28	3501/39457
<i>NPNT</i> (0)	1,22	(1,17-1,28)	9,90E-21	4862/57384

Supplementary Table 9: SNP interactions with psychiatric case status in iPSYCH₂₀₁₂ +

iPSYCH₂₀₁₅. Psychiatric (1), and Psychiatric (0) denotes children with psychiatric illnesses or not. OR refers to the main-effect of the SNP within each stratum.

Stratum	OR	CI 95 %	P	Cases/Controls
MTA3				
All	1,04	(1,01-1,07)	2,20E-02	9026/103479
Psychiatric status (1)	1,01	(0,96-1,05)	7,60E-01	6287/65496
Psychiatric status (0)	1,13	(1,06-1,21)	7,40E-04	2607/35991
Psychiatric status				
All	1,3	(1,24-1,36)	2,24E-27	9026/103479
MTA3 (2)	1,21	(1,13-1,29)	7,78E-09	4760/52913
MTA3 (1)	1,4	(1,3-1,51)	1,03E-17	3552/42278
MTA3 (0)	1,48	(1,25-1,76)	8,03E-06	714/8288

Supplementary Table 10: Drug-gene interactions among the most confidently prioritized genes across GWAS loci.

Gene	Drug	Regulatory approval	Indication	Interaction score
<i>CLEC16A</i>	GLATIRAMER	Approved		15,00
<i>IL13</i>	DECTREKUMAB	Not Approved	for treatment of Chron's disease,antiinflammatory agent	4,37
<i>IL13</i>	TNX-650	Not Approved	antineoplastic agent	8,75
<i>IL13</i>	QAX-576	Not Approved		8,75
<i>IL13</i>	CNTO-5825	Not Approved		8,75
<i>IL13</i>	ANRUKINZUMAB	Not Approved		8,75
<i>IL13</i>	ROMILKIMAB	Not Approved		1,46
<i>IL13</i>	ADALIMUMAB-ADBIM	Approved	DMARD,antiinflammatory agent	0,26
<i>IL13</i>	ROMILKIMAB	Not Approved		2,19
<i>IL13</i>	CENDAKIMAB	Not Approved		4,37
<i>IL13</i>	LEBRIKIZUMAB	Not Approved		8,75
<i>IL13</i>	TRALOKINUMAB	Approved	antiasthmatic agent	6,56
<i>IL13</i>	USTEKINUMAB	Approved	DMARD,antiinflammatory agent,Antipsoriatic Agents	0,43
<i>FUT2</i>	DEHYDRATED ALCOHOL	Approved		1,33
<i>HNF1B</i>	PLATINUM	Not Approved		2,62
<i>HNF1B</i>	METFORMIN	Approved	antidiabetic	1,34

Supplementary Table 11: We use the average silhouette width (ASW) of clusters to determine the number of SNP- and trait-clusters in the association-based grouping analysis. The highest ASW is highlighted in bold for both SNPs (k=7) and traits (k=5).

Average silhouette width		
K	SNP-cluster	Trait-clusters
1	0.000	0.000
2	0.227	0.184
3	0.221	0.269
4	0.225	0.296
5	0.156	0.333
6	0.214	0.301
7	0.231	0.292
8	0.190	0.231
9	0.153	0.248
10	0.141	0.201

Supplementary Table 12: Associations between meta-analysis SNPs and traits are shown in this table. Genes are sorted according to genomic positions, and we show the original and proxy SNP used in the analysis, combined with the r^2 . BETA signifies the estimate that refers to coefficients or log-odds ratios depending on the trait. We also include the Benjamini-Hochberg corrected p-value in the last column (FDR < 0.05).

Related traits associations (Google sheets).

Supplementary Table 13: SNP that were significant in the TWAS after correction for multiple testing and colocalization but identified in the GWAS. Novel SNPs are marked

with bold and novel SNPs that replicate in the MoBa cohort have bold p-values in the right side of the table.

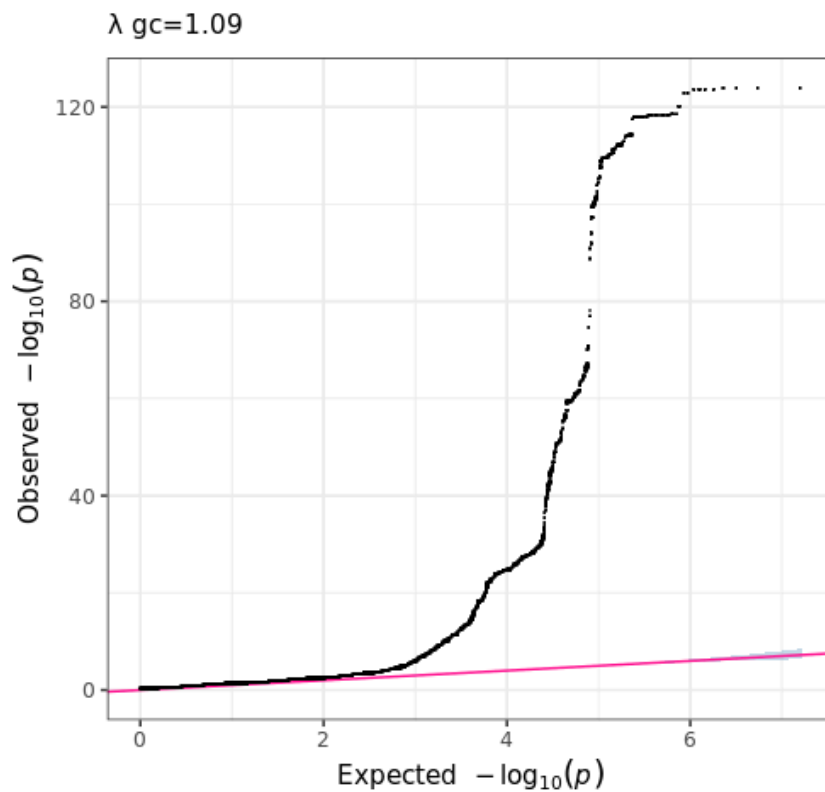
Discovery: Severe wheeze (0-7 years)								
Chr	Position	SNP	Mapped gene	EAF	EA/NEA	OR	CI 95 %	P-value
2	42502940	rs10490558	<i>MTA3</i>	0.71	A/G	1.06	(1.04-1.09)	1.84E-06
12	113376331	rs1859331	<i>OAS3</i>	0.69	A/C	1.06	(1.03-1.08)	8.02E-06
19	33532300	rs10411210	<i>RHPN2</i>	0.11	T/C	1.10	(1.06-1.14)	1.06E-07
19	46324880	rs11669073	<i>SYMPK</i>	0.30	C/T	1.07	(1.04-1.1)	1.78E-07
Replication: MoBa (wheeze 0-7 years)								
2	42502940	rs10490558	<i>MTA3</i>	0.71	A/G	1.03	(1-1.06)	4.03E-02
12	113376331	rs1859331	<i>OAS3</i>	0.69	A/C	1.03	(1-1.06)	4.76E-02
19	33532300	rs10411210	<i>RHPN2</i>	0.11	T/C	1.02	(0.97-1.07)	4.58E-01
19	46324880	rs11669073	<i>SYMPK</i>	0.30	C/T	1.02	(0.99-1.05)	2.05E-01

Supplementary Table 14: DeLong's test for generating confidence intervals for differences in AUC between the severe wheeze PRS and three previously published GWAS. The table contains the name of the study, PMID and sampled description. The AUC between the wheeze in MoBa and PRS's, dAUC contains the 95 % confidence interval of the differences between AUC values, DeLong's test generates a Z-statistics (Z-column) and respective p-value are all shown in the table. The AUC between the severe wheeze PRS and wheeze in MoBa was 0.561, which is used as a reference for all three tests.

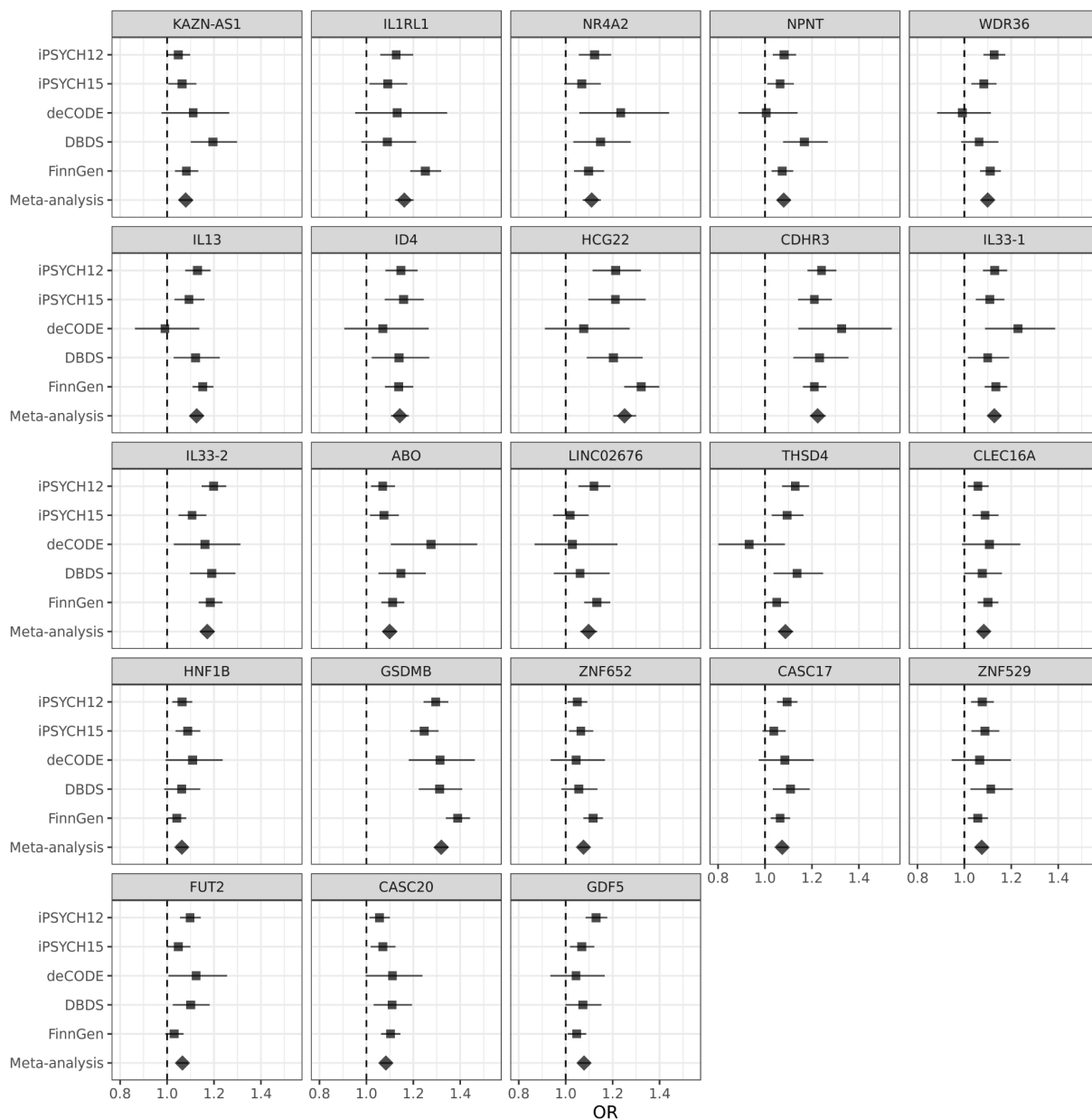
DeLong's test				
PRS	AUC	Z	dAUC (CI 95%)	P-value
Pivideri et al. 2019	0.54	7.21	(0.017-0.03)	5.79E-13
Namjou et al. 2022	0.54	7.61	(0.018-0.031)	1.69E-14
Tsuo et al. 2022	0.55	3.10	(0.003-0.016)	1.90E-03

Supplementary figures

Supplementary Figure 1: QQ-plot of meta-analysis

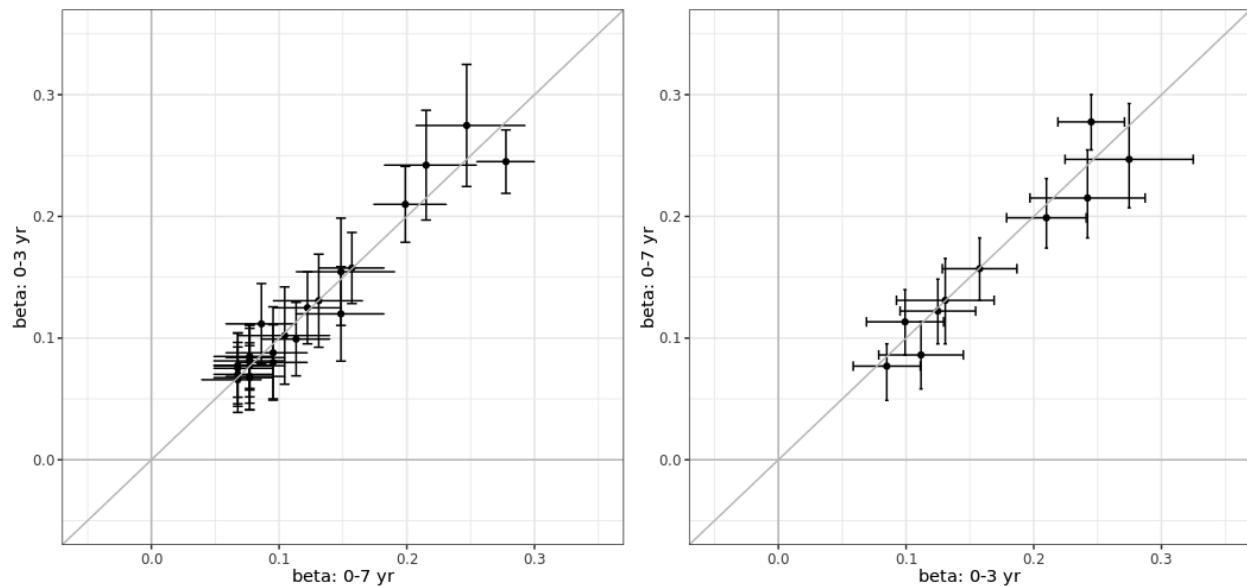


Supplementary Figure 2: Forest plot for genome-wide and TWAS findings, showing effect sizes for each individual study. The meta-analysis result is shown in the bottom of each panel.

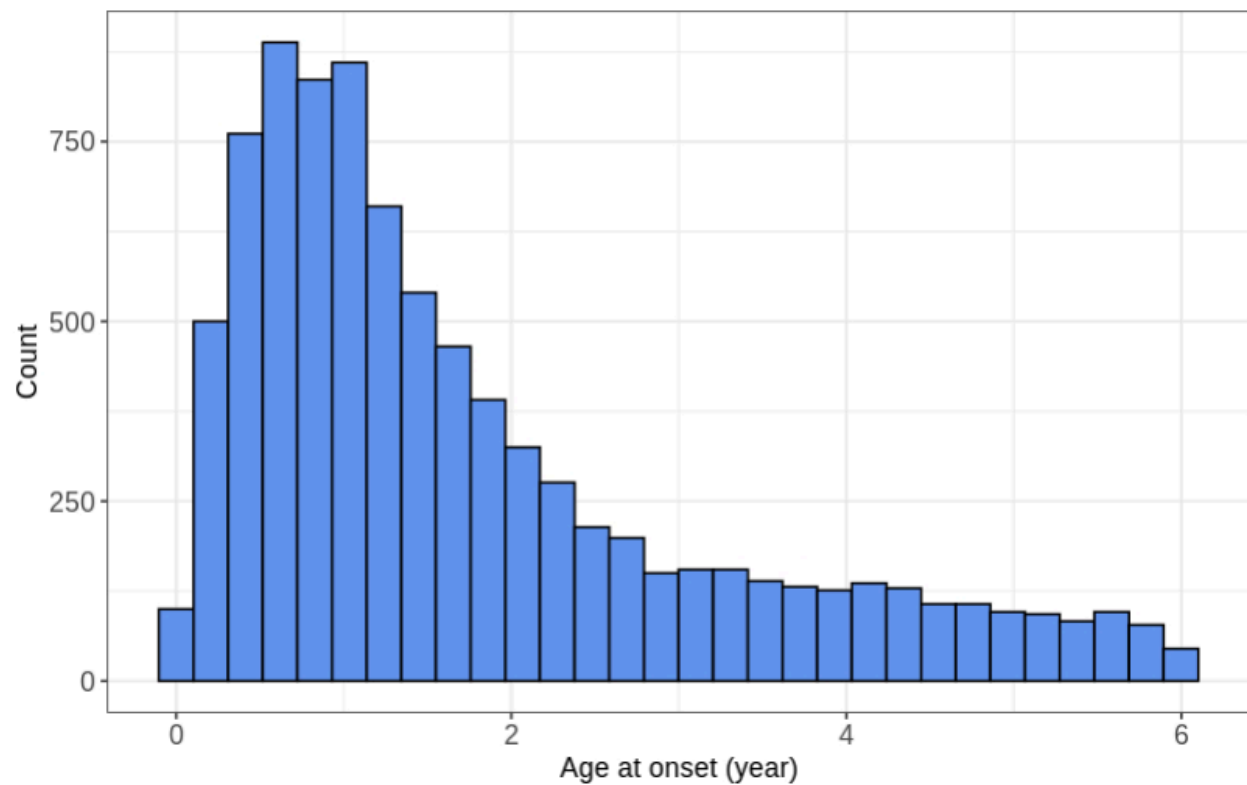


Supplementary Figure 3: The scatter plots show the effect sizes from the GWAS for severe wheeze ages 0-7 versus 0-3, using independent variants selected by clumping in Plink ($P < 5e-8$, $r^2 < 0.1$ within a 1 Mb window). The x-axis represents clumped

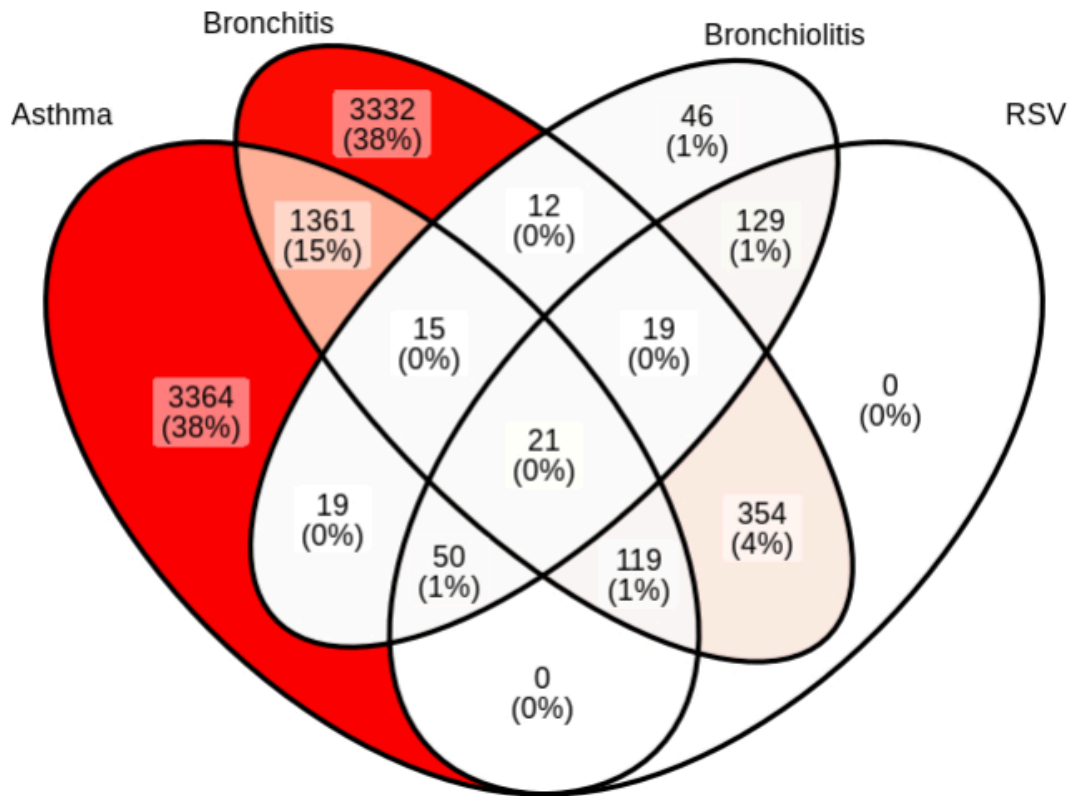
variants, and the y-axis shows their effect sizes in the other phenotype. Effect sizes are based on the full meta-analysis, excluding deCODE for the 0-3 phenotype.



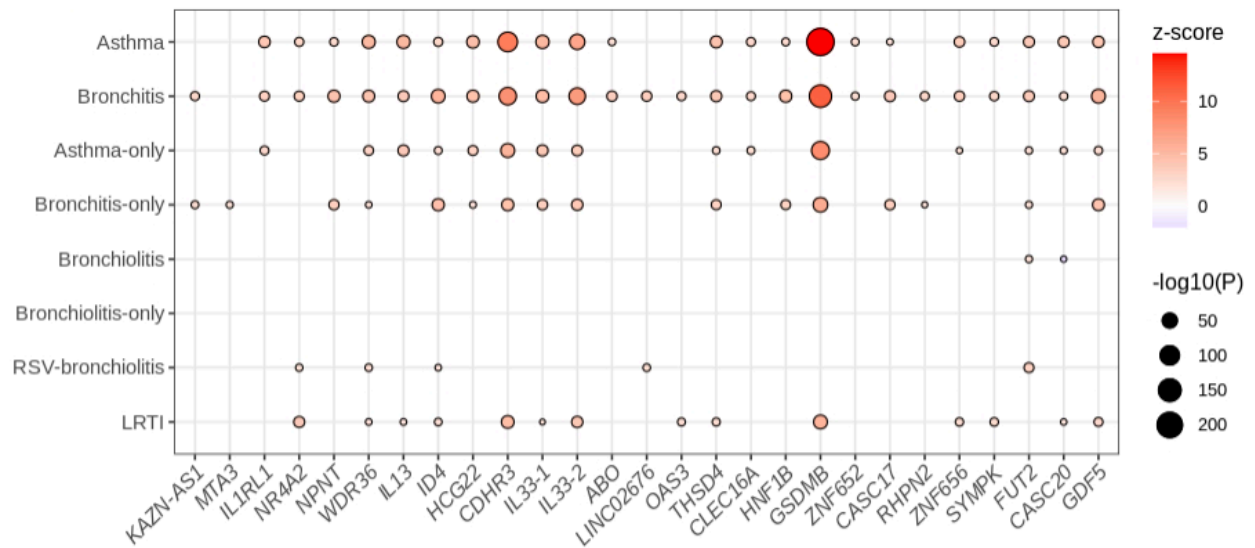
Supplementary Figure 4: The histogram displays the age distribution at the first wheezing diagnosis for the study's phenotype. Wheeze is defined as at least one hospitalization with bronchiolitis (ICD-10 code: 913, 390), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), or asthma (ICD-8 code: 493; ICD-10 codes: J45-46). The distribution is based on $iPSYCH_{2012} + iPSYCH_{2015}$ and does not represent the remaining cohorts.



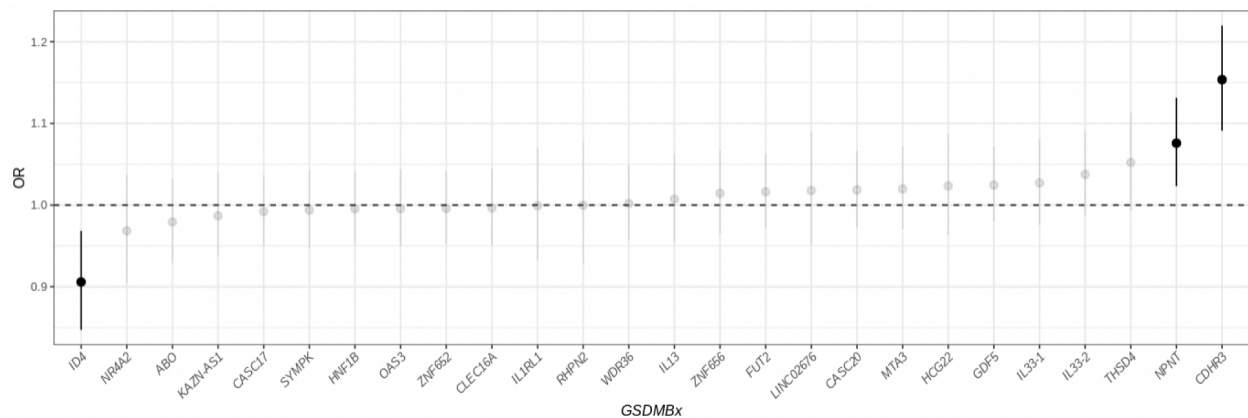
Supplementary Figure 5: Venn-diagram of diagnoses codes in the phenotype definition in iPSYCH. Asthma (ICD-8 code: 493; ICD-10 codes: J45-46), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), bronchiolitis (ICD-10 code: J21) or RSV-bronchitis/bronchiolitis (ICD-10 codes: J205, J210) named RSV for brevity.



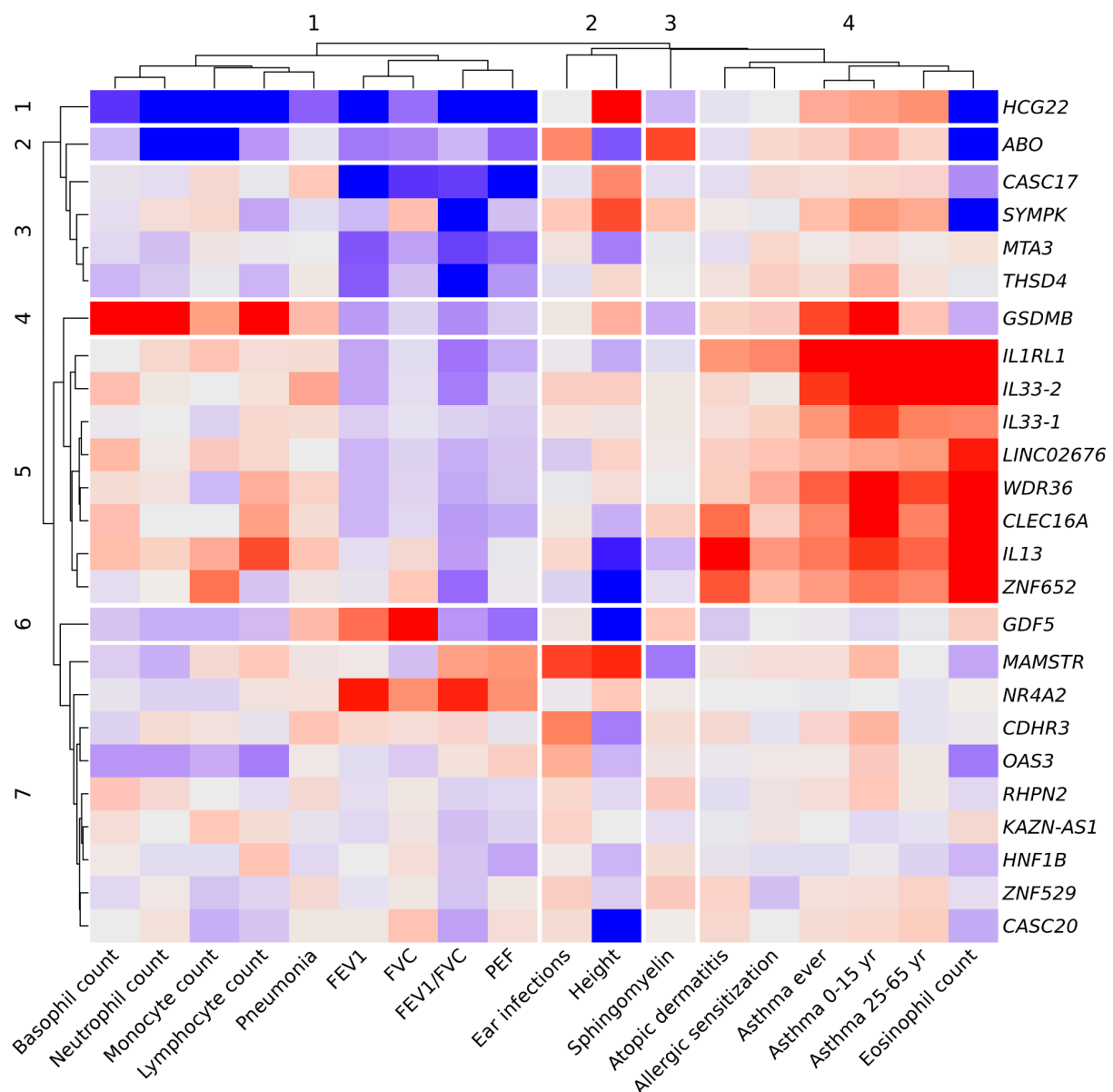
Supplementary Figure 6: Sensitivity analysis for findings across sub-diagnoses in $iPSYCH_{2012} + iPSYCH_{2015}$. Sub-diagnoses are shown on the y-axis with risk loci on the x-axis. The size of the point represents the p-value with bigger points signifying lower p-values and the redness representing the size of the z-score. Only SNP-trait associations with nominal p-values are shown ($P < 0.05$). See Supplementary Figure 5 for case numbers. LRTI, included 5672 cases and 108611 controls.



Supplementary Figure 5: *GSDMB*-interactions significant after multiple testing in $iPSYCH_{2012} + iPSYCH_{2015}$. Black solid dots represent significant interactions. The x-axis represents SNPs that interact with *GSDMB*, and the y-axis shows the odds-ratio of the interaction.

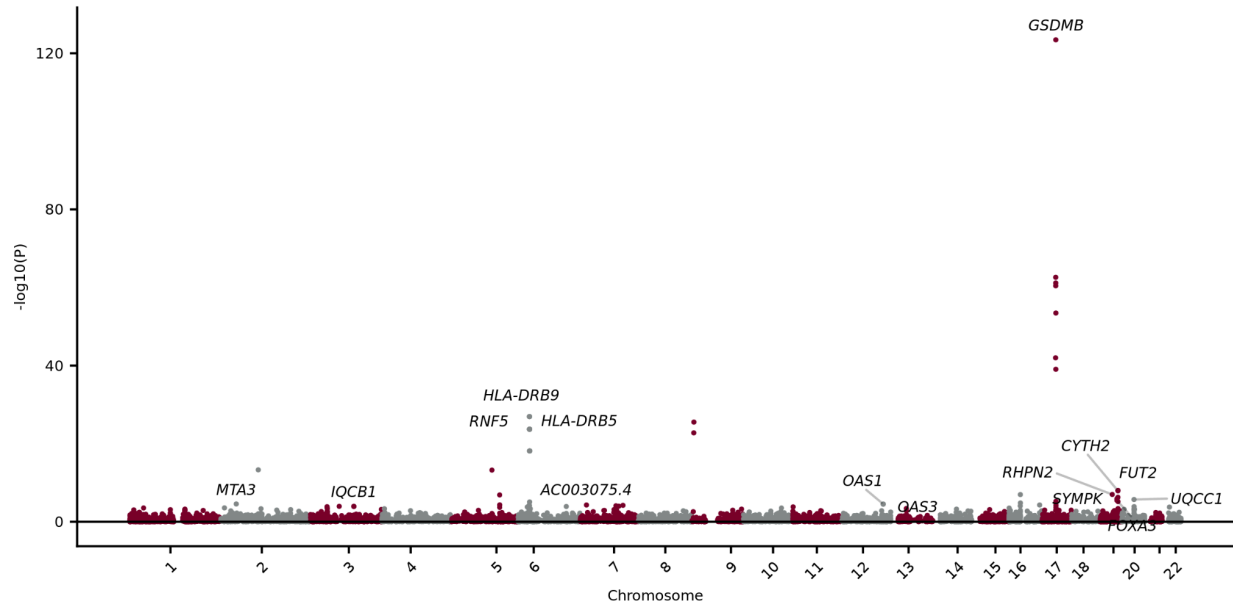


Supplementary Figure 7: Trait-based clustering of SNPs. SNP association statistics were scaled to Z-scores, and further scaled to zero-mean and unit variance to account for sample size imbalance. K-means clustering was used to determine the number of clusters among SNPs and traits, and hierarchical clustering was used to group SNPs with similar phenotypic consequences in each cluster. The SNP-clusters are visualized with unscaled Z-scores to reflect the true direction and strengths of association. SNPs with missing values were removed. Additionally, removing C-reactive protein from the HLA-region was excluded from this study. The HLA-region is represented by *HCG20* in our analysis. Only traits with at least one genome-wide significant wheeze lead SNP ($P < 5e-8$) were included in the analysis. Red color indicates a positive Z-score and blue indicates a negative.

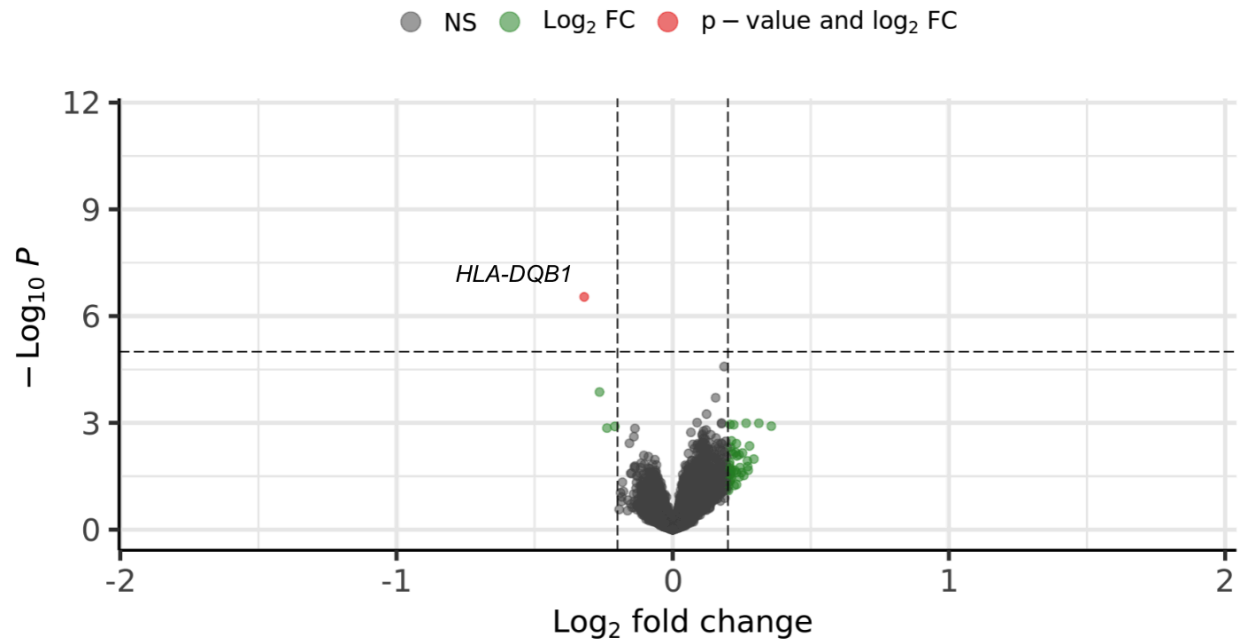


Supplementary Figure 8: Manhattan plot showing TWAS results using the COPSAC2010 NEC expression data. Gene are labeled if the TWAS P-value passes

Benjamini-Hochberg and if the probability of colocalization (Shared variant) is above 75 %.

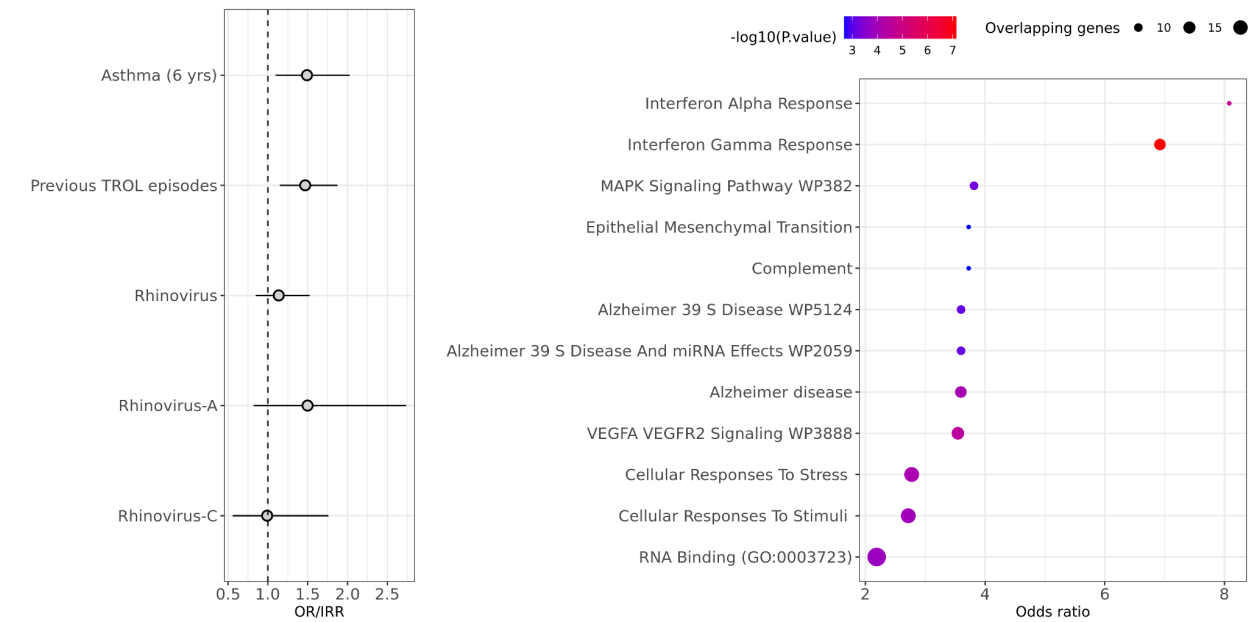


Supplementary Figure 9: Differential expression in COPSAC 2010 - PRS vs. RNA levels using Limma. Log-fold change pr. standard deviation of PRS is shown on the x-axis and p-value on the y-axis. Green color represents genes with large log-fold change, but not passing correction for multiple testing, while red dots represent significant findings (FDR < 0.05).



Supplementary Figure 10: The right panel shows the associations between the PRS (0-7) in the MARC-35 cohort with relevant traits based on the COPSAC PheWas, with the PRS as exposure. Binary outcomes are analyzed with logistic regression while count phenotypes (Previous TROL episodes) are analyzed using quasi-poisson regression. The right panel shows enriched pathways using enrichR, among genes associated with the PRS in the MARC-35 expression data in the nasal epithelium sampled within 24 hours of hospitalization. All analyses were performed in infants of

European ancestry hospitalized with bronchiolitis and adjusted for gender, age and top 2 principal components.



Paper III

Fischer-Rasmussen, K et al.

Early Childhood Trajectories of Inhaled Corticosteroid Prescriptions Implicate the Presence of Wheeze Endotypes

Early Childhood Trajectories of Inhaled Corticosteroid Prescriptions Implicates the Presence of Wheeze Endotypes

Kasper Fischer-Rasmussen¹, Casper-Emil Tingskov Pedersen¹, Astrid Sevelsted¹, Luo Yang¹, Bo Chawes¹, Jakob Stokholm¹, Ann-Marie Malby Schoos¹, Anders Børglum^{3,4,5}, Thomas Munk Laursen^{3,7}, Bjarni Vilhjalmsen^{4,6,7,8}, Anders Gorm Pedersen², Anders Ulrik Eliassen^{1,9,10} & Klaus Bønnelykke¹

¹COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark

²Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

³The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, DK-8000 Aarhus, Denmark

⁴Center for Genomics and Personalized Medicine, Aarhus University, DK-8000 Aarhus, Denmark

⁵Department of Biomedicine, Aarhus University, DK-8000 Aarhus, Denmark

⁶Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

⁷NCRR - National Centre for Register-based Research, Aarhus University, Aarhus, Denmark

⁸Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, Massachusetts

⁹Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

¹⁰Department of Pediatrics, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA

Acknowledgements

We would like to acknowledge the huge work of late professor Hans Bisgaard, who was the founder of COPSAC and was head of the clinical research center for more than 25 years. Hans was a dedicated, innovative physician-scientist who pushed the asthma research field forward. He contributed immensely to pediatric research through the COPSAC birth cohorts, as well as a vast amount of other clinical studies on childhood asthma. Hans's impressive work and ideas live on in the following studies conducted in the birth cohort. Thank you for being a great inspiration to us all.

Abstract

Rationale

Wheezing is a clinical symptom that covers a diverse array of etiologies, resulting in inadequate treatments and lack of preventive measures.

Objective

To enhance comprehension of the underlying disease mechanisms associated with wheeze phenotypes in early childhood by examining the trajectories of inhaled corticosteroid (ICS) use.

Methods

We leveraged the iPSYCH consortium to couple genetic information with electronic health records (EHR). Subsequently, ICS trajectories were identified with latent class analysis and characterized by perinatal, socioeconomic, familial, psychiatric and genetic risk factors. Lastly, we sought to test if risk factors were distinctly associated with ICS trajectories.

Measurements & Main Results

Early-transient redeemers are only weakly influenced by genetics and strongly influenced by lifestyle factors such as smoking during pregnancy, household income and parental age. Prolonged-transient redeemers displayed elevated maternal antibiotic consumption, distinct association with *FUT2* and weak associations with atopic traits. Persistent redeemers were characterized by birth by c-section, high genetic impact of known risk loci and strongly associated with both asthma and atopic risk scores. Late-onset redeemers showed weak influence of perinatal risk factors, genetic overlap with atopic traits and moderate influence of the HLA-locus.

Conclusions

Different associations with risk factors between the trajectories, including birth circumstances, parental medication use, genetic loci and trait susceptibility imply that the observed temporal prescription patterns represent endotypes.

Introduction

Asthma is defined as a chronic inflammatory disorder, but the disease is characterized by a heterogeneous disease course. This is particularly true in early childhood with a variety of different clinical courses, from transient wheezing that resolves naturally to persistent forms requiring ongoing treatment. Various trajectories of wheeze symptoms have been demonstrated to have distinct features in terms of atopy, bronchial responsiveness and lung function progression (2–11), indicating distinct underlying mechanisms. Early studies identified subtypes through expert clinical knowledge, but later studies have leveraged data-driven approaches to characterize phenotypes from longitudinal records of wheeze symptoms (2,4,12–19). Temporal phenotypes identified and characterized in birth cohorts have proven hard to replicate (20) and differences in sampling frequency, timing and sample size impact the result of clustering analyses (21), suggesting that substantial within-class heterogeneity may limit the identification of temporal patterns reflecting true endotypes.

Electronic Health Records (EHRs) may address these challenges since data-collection occurs through documentation of routine clinical practice, ensuring longitudinal records of virtually all health-related events. Furthermore, in countries with substantial medication reimbursement, such as the Scandinavian countries, practically all asthma medication redeemed at pharmacies is included in the national registries and linked to the personal identification number (21). A previous study of redeemed asthma medication prescriptions in Denmark for diagnosing childhood asthma has reported good specificity (86%) and reasonable sensitivity (63%) (22). Previous studies have

identified prescription patterns of asthma medication in EHRs, but their main focus has been progression to adult asthma, severity and adverse events, including parental mental illness (16, 17, 18,19, 26). Additionally, asthma and psychiatric disorders are highly intertwined with suggested shared etiology (28, 29), suggesting that psychiatric disorders may also influence temporal subtypes of wheeze.

This study seeks to delineate temporal trajectories of inhaled corticosteroid (ICS) prescriptions in the first 9 years of life, by leveraging Danish electronic health records and to assess whether perinatal and genetic risk factors offer evidence supporting the existence of distinct endotypes. Furthermore, we aim to characterize the overlap between the identified trajectories and mental illnesses.

Methods

Population

Data was obtained in a subset of the participants in the integrative psychiatric research (iPSYCH) consortium. iPSYCH is divided between two cohorts; iPSYCH₂₀₁₂ consisting of individuals born between 1981 and 2005 (30), and iPSYCH₂₀₁₅ with individuals born between 1981 and 2008 (31). The iPSYCH cohorts are linked to all Danish public registries using the central person registry (CPR), with access facilitated by Statistics Denmark. Data transfer and infrastructure are managed by the National Center for Register-based Research (NCRR) at Aarhus University.

We included prescriptions for inhaled corticosteroids (ICS), beta-2 agonists (B2), nasal steroids, antihistamines, steroid creams and antibiotics (see Supplementary). Data on prescription medication were available from the National Prescription Registry (NPR), where each registered prescription reflects a redeemed medication but not whether it was actually taken (21). NPR started in 1995 and prior to 1996, children's prescriptions were registered under mother's social security number (21) and therefore we included individuals born after 1998 to ensure adequate data quality. Due to a trade-off between sample-size and longest possible follow-up, we included iPSYCH participants with complete data for the first 9 years of life, excluding individuals who had migrated in the period. The Migration registry runs until 2017, meaning we can include individuals born earliest in 1998 and individuals who turn 9 in 2017 (see Supplementary Figure 1).

We restricted all analysis to iPSYCH individuals with genotype data (see Supplementary

Figure 1). Individuals with high relatedness or genotype missingness were excluded. Furthermore, SNPs were excluded if they exhibited a high rate of missingness ($> 5\%$), significant deviation from Hardy-Weinberg equilibrium (HWE) ($P < 1e-5$), or a low minor allele frequency (MAF $< 1\%$). Further details about sample and genotype QC have been described in rigorous detail (32).

Longitudinal prescription indications

We defined prescription trajectories based on redeemed ICS prescriptions, recording the presence of ICS each life-year without considering the number of prescriptions. This was done because the registry lacks data on actual medication use. Time-points (life-years) were defined as bins of 365 days starting from birth. Since multiple prescriptions don't necessarily indicate disease severity, we used binary indicators defined as any ICS-prescriptions redeemed in a life-year. We focused solely on ICS prescriptions for clustering, as beta-2 agonists are typically first-line treatments, indicating single episodes of wheeze, and often only used on a trial basis in cases of suspected asthma (33). Controls were individuals in the iPSYCH cohort born between 1998 and 2008 that had never redeemed an ICS prescription in the first 9 years of life.

Latent class analysis

We performed latent class analysis (LCA) using the R package poLCA (34), on yearly ICS prescriptions as described above. We only included individuals with at least a single ICS prescription in the first 9 years of life due to detrimental class imbalance; if non-ICS prescribers were included in the model they would constitute 79.2 % of the sample size.

We selected the number of classes based on the bayesian information criterion (BIC), Lo-Mendell-Rubins test (LMR-test) (35) the bootstrapped likelihood test (BLRT) with 1000 bootstraps (35), model entropy, average silhouette width (ASW), parsimony and interpretability, i.e. absence of the “Salsa effect” (36). The Salsa effect occurs when trajectories differ only in prevalence, indicating a single class split into multiple levels. Furthermore, we ensure that each class had a 70 % median posterior class membership probability (37) to ensure robustness of class assignments. For the LMR-test and BLRT a p-value below 0.005 (accounting for 9 tests), is considered a significantly improved model. We also compute the entropy and average silhouette width using binary distances for each model.

To account for potential cluster instability, we iteratively removed 20% of the data during the fitting process, for each value of k (number of clusters). Additionally, we downsampled the data from 100% to 10% and applied the same fitting procedure described above to each subset, iteratively removing 20 % of the data multiple times for each number of classes, now with varying sample size to assess robustness of the clustering.

Characterization by other asthma and allergy medications

For each phenotype we identified yearly prevalence and odds-ratios (OR) compared to non-prescribers, for prescription medications related to asthma-comorbidities such as beta-2 agonist, nasal steroid, antihistamines, steroid cream and antibiotics. The same analysis was performed with hospitalizations for asthma, wheezing and

lower-respiratory tract infections in the first six years of life.

Perinatal risk factors

We investigated associations between the trajectories and 22 risk factors describing birth-circumstances, socioeconomic status and familial medication history. From the Medical Birth Registry we retrieved variables related to the birth of the child, such as birth weight, gestational age, premature status, cesarean section, gender, maternal smoking during pregnancy and neonatal respiratory distress. We also leveraged registries related to socioeconomic status and family conditions, such as parental age, number of siblings and distance to older siblings. Lastly, we utilized parental prescription history 10 years prior to the birth of the child, including redemption of at least one asthma prescription (ICS & b2-agonist), allergy prescription (Nasal steroid & antihistamines), steroid cream and antibiotic (see Table 1 and Supplementary Figure 4).

Genetic risk factors

The association between known wheeze risk variants and class status was also assessed by analyzing genetic variants identified in previous genome-wide association studies (GWAS) of preschool wheeze and asthma. We read genotype dosages from iPSYCH files using the R package bigsnpr (38) to extract genetic variants.

We included variants, that were flipped to the asthma/wheeze risk allele, near the genomic loci *GSDMB* (rs2305480-G), *CDHR3* (rs6967330-A), *FUT2* (rs601338-G),

HLA-DQA1 (rs1071630-C), *IL-33* (rs340933-G & rs1342326-C), *IL1RL1* (rs10189629-C), *IL-13* (rs20541-A), *ABO* (rs505922-C), *TSLP* (rs1043828-C) & *ANXA1* (rs75260654-T) (39–41). We also characterized the effect of epistatic interactions known to influence preschool asthma (39,42).

We also investigated relationships between polygenic risk scores (PRS) and the phenotypes. We include PRS's of asthma phenotypes (43,44) (Childhood onset asthma, adult-onset asthma and severe asthma), BMI (45), FEV1/FVC (46), immune cells (47) (Eosinophils & neutrophils), Psychiatric traits (32,48–51)(Depression, bipolar, autism, anxiety & ADHD), atopic traits (52,53) (Atopic dermatitis, allergic sensitization and allergic rhinitis) and C-reactive protein levels (54).

Psychiatric status

We tested all associations for interaction with psychiatric status (see Supplementary Material), in order to determine if predisposition to psychiatric disorders influenced our results. Furthermore, we looked for associations between all psychiatric diagnoses and the prescription trajectories, to characterize the susceptibility to specific psychiatric illnesses.

Statistical analysis

Latent class analysis was performed to identify prescription trajectories, using binary indicators of yearly redeeming of inhaled corticosteroids (ICS). No covariates were

included in the model.

For perinatal, socioeconomic and genetic risk factor analysis, multinomial regression was used with LCA-derived classes as endpoints including non-prescribers of ICS as a reference. All analyses were adjusted for gender, cohort-status (iPSYCH₂₀₁₂ or iPSYCH₂₀₁₅), birth year and psychiatric status. Parental birth year was included in the model for the parental prescription history, additional to the birth year of the child. Neonatal respiratory distress was also adjusted for gestational age. For the genetic risk factor analysis, we also adjusted the top 10 principal components (PCs) to account for population stratification. All continuous variables were scaled to zero-mean and unit standard deviation. In a subsequent sensitivity analysis we included interaction terms with all included risk factors and psychiatric status.

We also sought to investigate the association between the identified ICS trajectories and specific psychiatric diagnoses. Since our temporal phenotypes are defined from 0-9 years and most psychiatric disorders debut both during and this period, we considered psychiatric diagnoses both an exposure and outcome in two separate analyses. When considering psychiatric as an outcome, we considered ICS trajectories as exposures in logistic regression analysis, contrary to multinomial when psychiatric diagnoses were included as an exposure. All analyses of psychiatric diagnoses were adjusted for gender, cohort-stats and birth year, maternal smoking during pregnancy, gestational age, number of older siblings, paternal low education and household income.

Results

Latent class analysis

We included 7527 children with at least one ICS prescription before age 9. We identified two potential solutions based on the elbow plot of BIC vs. number of classes (see Supplementary Figure 2). Two inflection points were observed, one at 4 classes and another at 6 classes. Sensitivity analysis on the number of clusters also indicated either 4 or 6 classes. Since the 4-class model constituted the most parsimonious solution, we performed all subsequent analyses on the results of the 4-class model (see Supplementary Table 1).

We clinically classified the 4 identified classes as early-transient redeemers (ETR, n = 3919 (52 %)), late-onset redeemers (LOR, n = 1302 (17 %)), persistent redeemers (PR, n = 781 (10 %)) and prolonged-transient redeemers (PTR, n = 1525 (21 %)). ETR was characterized by 50 % redeeming ICS in the 2nd life year, and subsequent zero prescriptions from the 6th life year and throughout the study period. LOR redeemed zero prescriptions in the first 3 years of life, peaking at age 8-9 with app. 50 % redeeming ICS. PR displayed persistent redeeming throughout the study period, with app. 25 % in the first life year with almost 100 % redeeming ICS from age 5-8 with a small decrease in the 9th year of life. PTR was characterized by a high proportion of redeemers (~ 70-80 %) from age 2-4, with decreasing prescription frequencies from age 5 and until age 7 where less than 25 % redeem ICS and until age 9 where prescription frequencies almost reached 0 % (see Figure 1). Individuals not included in the analysis were children who never redeemed ICS in the period (n = 28,809) and were used as

controls for subsequent analyses.

Characterization by other asthma and allergy medications

PR displayed highest prevalences of redeeming all categories (beta-2, nasal steroid, antihistamines, steroid creams and antibiotics), in the later years (3-9 years), whereas PTR had higher prevalences of beta-2 and antibiotics prescription in early childhood (0-3 years), specifically PTR displayed a slightly higher beta-2 prescription prevalences in the first two years of life (see Supplementary Tables 2 & 3 and Supplementary Figures 5 & 9).

In the first two life-years, PR, PTR and ETR had a higher prevalence of hospitalizations with both asthma, wheeze and lower-respiratory tract infections compared to non-redeemers, whereas LOR did not. At age 6, only PR and LOR had higher than baseline prevalence of hospitalizations with asthma and wheeze. Both PR, PTR and LOR had higher than baseline prevalence for hospitalizations for lower-respiratory tract infections at age 6, whereas ETR did not (see Supplementary Tables 4 & 5 and Supplementary Figures 10-12).

Perinatal risk profiles

ETR, LOR, PR and PRT were associated with 19, 13, 11 and 19 out of 22 perinatal risk factors respectively, after adjustment for multiple comparisons. Premature birth, cesarean section, gestational age, gender and parental asthma and allergy were significantly associated with all trajectories after Benjamini-Hochberg correction.

Maternal smoking (OR = 1.48, (CI = 1.4-1.6), $P = 0.00e+0$), older siblings (OR = 1.09, (CI = 1.0-1.1), $P = 1.87e-4$), maternal age (OR = 0.89, (CI = 0.88-0.92), $P = 1.56e-7$) and paternal age (OR = ?) were most strongly associated with ETR. No risk factors were most strongly associated with LOR. Cesarean section (OR = 1.50, (CI = 1.3-1.8), $P = 3.47e-8$) and female gender (OR = 0.77, (CI = 0.7-0.8), $P = 1.54e-10$) was the strongest risk factor for PR. Lastly, prematurity (OR = 2.02, (CI = 1.7-2.4), $P = 4.88e-15$), gestational age (OR = 0.81, (CI = 0.8-0.82), $P = 0.00e+0$) and maternal antibiotic use (OR = 1.50, (CI = 1.3-1.7), $P = 5.05e-8$) had highest odds-ratios for PTR. Additionally, neonatal respiratory distress was moderately associated with only PR (OR = 1.49, (CI = 1.1-2.1), $P = 0.024$) and PTR (OR = 1.38, (CI = 1.1-1.8), $P = 0.01$) after adjusting for gestational age (see Figure 2).

Within-class ANOVA suggested that prematurity, cesarean section, neonatal respiratory distress, smoking during pregnancy, birth weight, gestational age, gender, older siblings, parental age and mothers antibiotic use had distinct associations among classes (see Table 2).

Paternal asthma, allergy and eczema did not show distinct differences among the ICS trajectories. Only maternal antibiotic use was distinctly associated with ETR and PTR, but most strongly with PTR. Socioeconomic factors like household income, and low education level of parents were either not associated with or displayed equal effects across the trajectories.

Genetic risk profiles

ETR, LOR, PR and PRT were associated with 3, 3, 8 and 4 out of 11 SNPs after adjustment for multiple comparisons. *GSDMB*, *CDHR3* and *TSLP* were associated with ETR after correction for multiple testing. No SNPs were most strongly associated with ETR. *HLA-DQA1* (OR = 1.16, (CI = 1.1-1.3), $P = 5.04e-4$) was most strongly associated with LOR. *GSDMB* (OR = 1.52, (CI = 1.4-1.7), $P = 5.3e-15$), *CDHR3* (OR = 1.45, (CI = 1.3-1.6), $P = 2.31e-9$), *IL-33*, *IL1RL1*, *IL-13* (OR = 1.27, (CI = 1.1-1.4), $P = 7.76e-5$) and *TSLP* (OR = 1.31, (CI = 1.2-1.5), $P = 1.44e-7$) were the strongest predictors of PR. Whereas only *FUT2* (OR = 1.14, (CI = 1.1-1.2), $P = 4.73e-4$) was strongest for PTR. Within-class ANOVA indicated distinct associations between phenotypes for *GSDMB*, *CDHR3*, *FUT2*, *IL-33*, *IL-13*, and *TSLP*. The T allele of the *ANXA1* SNP (rs75260654) was negatively associated with PR after multiple testing corrections (OR = 0.62, (CI = 0.4-0.9), $P = 1.74e-2$). Epistatic effects of *GSDMB*×*CDHR3* and *FUT2*×*ABO* did not pass correction for multiple testing, however *GSDMB*×*CDHR3* displayed a positive trend for all phenotypes (see Figure 3 and Table 3). Polygenic risk scores for childhood onset asthma, severe asthma, adult onset asthma, eosinophil counts, allergic rhinitis and sensitization and autism showed indications of distinctness in effect sizes. PRs for asthma subtypes had considerably higher effects for PR, compared to the remaining phenotypes. The strongest genetic effects across all SNPs and PRs were observed for PR (see Supplementary Figure 5). PR and LOR were strongly associated with allergy and eosinophil PRs, whereas ETR and PTR displayed no effects with those scores. The PR for autism revealed a protective effect for ETR, and was not significantly associated with any other phenotypes after correction (see Table 4).

Psychiatric status

Only the PRS for childhood-onset asthma displayed interaction with psychiatric status for PR with a stronger effect in individuals without any psychiatric diagnoses (Psychiatric cases (OR = 1.59, (1.45-1.74), $P = 1.17\text{e-}29$) vs. Psychiatric controls (OR = 2.1 (1.84-2.39), $P = 6.43\text{e-}29$).

Schizophrenia, schizophrenia spectrum, bipolar, affective disorders, autism, ADHD and anorexia had median debut age of 21, 20, 23, 20, 10, 12 and 16 years of age respectively in the iPSYCH cohorts. We observed associations between schizophrenia spectrum with LOR, affective disorder with ETR & LOR and ADHD with ETR and PTR. ADHD was strongly associated with ETR, PTR and to a lesser extent with PR. All associations were corrected for multiple testing using the Benjamini-Hochberg procedure. The same results were observed regardless of psychiatric traits being considered exposures or outcomes (see Supplementary Tables 5 and 6).

Discussion

We identified four trajectories of ICS prescription patterns in the first 9 years of life, revealing distinct associations among perinatal and genetic risk factors.

To aid us in choosing the most parsimonious model we used the elbow method, in order to identify when model performance reached a saturation point and inclusion of additional parameters only lead to modest improvements, in terms of BIC. We leveraged additional methods to identify the correct inflection point. The Lo-Mendell-Rubin's test (LMR-test) also suggested a plateau in model improvement at 4 classes, when comparing a 3-class to a 4-class model ($LR = 1448$). In comparison, the LMR-tests showed smaller improvement between a 4-class and 5-class model ($LR = 492$). An inflection point was also observed at a 6-class model both in terms of BIC and LMR, however the BLRT-test did not find improvement in model-likelihood beyond a 4-class model. Considering the BIC, LMR-test and BLRT, we interpret the 4-class model as the most parsimonious model and therefore the least likely overfitted model.

Previous studies have included healthy individuals in the clustering analysis, naming this class "never/infrequent wheezes". However, since ~80 % of individuals in our study-population never redeemed an ICS in the first 9 years of life, this would have created too much class imbalance to identify robust classes (35). We included controls (defined as non-prescribers of ICS) after identifying latent classes, in subsequent risk factor analysis using the controls as a reference-class in multinomial regression. The advantage of this is that we are more confident in the identified classes considering the lower degree of class imbalance, however this does not allow adjustment of class-membership probabilities in subsequent analyses. Since we didn't include beta-2

agonists in our phenotype, our control population also contains children receiving beta-2 agonists, thus our controls can be characterized with a similar name as in other studies, namely “never/infrequent wheezers/redeemers”. Previous studies have reported 50-70% never/infrequent wheezers (13,15,16), whereas in our study this group constitutes 79.2 % of our sample. This could potentially be explained by a large proportion of our study population having psychiatric diagnoses and low drug-adherence among psychiatric patients (55) may result in written prescriptions that are never redeemed, potentially inflating the control group (56).

We found distinct associations between the ICS trajectories, despite similar direction of effects. As an example, birth by cesarean section was significantly associated with all phenotypes (see Table 2), but substantially higher for the PR (OR = 1.5, (CI = 1.3-1.8), $P = 3.47e-6$) suggesting that c-section increases risk for persistent redeeming throughout childhood, possibly leading to asthma. Previous studies have shown that c-section increases risk of asthma (57) and that this effect is mediated by changes in the infant gut microbiome (58), implicating the gut microbiome in mediating early persisting symptoms rather than remitting or late-onset wheeze. Additionally, premature birth and gestational age were most strongly associated with PTR and older siblings and maternal smoking were specific to ETR. These findings are similar to a study showing that the effect of cesarean section and prematurity on asthma-like symptoms increases with age, whereas the effect of older siblings decreases with age in the first three years of life (59). In conjunction with our results, this suggests that being born prematurely or via cesarean section leave lasting effects on asthma-like symptoms such

as wheeze, whereas the effect of older siblings and maternal smoking are more or less transient.

PTR displayed a distinct association to maternal antibiotic use (OR = 1.5, (CI = 1.4-1.7), $P = 5.5e-8$) and a previous study showed that antibiotic use 80 weeks before and after pregnancy increased risk of child's asthma and the authors attributed this to maternal infection susceptibility (60). Our observations that maternal antibiotic use 10 years prior to pregnancy is mostly associated with PTR, suggest that maternal infections influence the prolonged redeeming of ICS in this phenotype. ETR was distinctly associated with maternal smoking during pregnancy, birth weight (shared with PTR) and older siblings, suggesting substantial environmental predispositions. This suggests that ETR can arise from infants with low birth weight resulting in increased vulnerability to infections (61), combined with a high viral burden conferred by older siblings (62).

Neonatal respiratory distress was associated with PR and PTR suggesting congenital susceptibility of these phenotypes to respiratory adverse outcomes in infancy. Studies suggest that asthmatic mothers have a higher incidence of adverse neonatal respiratory outcomes even among term births (63) and infections in infancy are associated with asthma persisting until age 10 (64). Our observed association with PTR indicates that infant respiratory distress also influences symptoms remitting earlier than previously suggested. Respiratory distress was not associated with LOR or ETR and the within-case ANOVA analysis suggests that this difference is significant, indicating that respiratory distress in infancy is limited to phenotypes with early symptoms that persists. It could be speculated that this could be related to reduced lung function and/or airway

reactivity at birth, which has been shown to be a risk factor for early asthma and severe exacerbations (65).

Generally, PR displayed higher effect sizes for candidate SNPs suggesting a higher degree of heritability. This is highlighted by a statistically higher association of *GSDMB* with PR and PTR compared to ETR and LOR. Additionally, *CDHR3* was differentially associated with PR compared to the remaining phenotypes, suggesting innate susceptibility to viral infections (66–68), are strong genetic drivers for persistent ICS use. The loci associated most strongly with PR with distinct effects were *CDHR3*, *IL-33*, *IL1RL1*, *IL-13* and *TSLP*. SNPs in the pathway comprising *IL-33* and its cognate receptor *IL1RL1* were previously found to be associated with persistent wheeze and wheeze with later onset, but not with transient-early wheeze (69). This is consistent with our results showing that both loci in *IL-33* and *IL1RL1* being specifically associated with PR.

Associations between various temporal wheeze phenotypes and *IL-13*, *TSLP*, *FUT2* and *HLA-DQA1* have not been previously investigated. There was no observed association between *IL-13* and ETR, LOR or PTR, whereas PR displayed a strong association with the *IL-13* SNP (OR = 1.27, (CI = 1.1-1.4), P = 7.76e-5) and within-class ANOVA suggested this association to be distinct (Adjusted P-value = 0.03). *IL-13* is involved in a large array of mechanisms involved in asthma pathobiology including IgE production, eosinophil recruitment, goblet cell maturation, extracellular matrix protein production and myofibroblast differentiation (70), suggesting broad involvement in T2-high pathways.

Contrary to *IL-13*, *TSLP* was associated with all phenotypes, but significantly stronger for PR according to within-class ANOVA (Adjusted P-value = 0.03). *TSLP* is a cytokine that influences development, viability and recruitment of dendritic cells, T cells, B cells, neutrophils, mast cells, eosinophils, and innate lymphoid cells, indicating pleiotropic immune roles (71). Additionally, patients with asthma have increased airway expression of *TSLP* mRNA compared to healthy controls (72,73) and levels of *TSLP* have been shown to correlate with levels of airway obstruction (74). Thus, *TSLP* plays a role in various molecular mechanisms underlying the etiology of asthma, which accounts for its association with all trajectories.

PTR displayed a distinct association with *FUT2* (OR = 1.14, (CI = 1.1-1.2), P = 4.73e-4), which has previously been shown to increase risk of respiratory and gastrointestinal infections during infancy (75) and individuals with *FUT2*-polymorphisms are more susceptible to norovirus, rotavirus and *Helicobacter pylori* (76,77). Furthermore, *FUT2*-*ABO* epistasis increases risk of respiratory infections with *Streptococcus pneumoniae* in early childhood (39), further implicating the role of infection susceptibility in combination with the effect of maternal antibiotic use on PTR.

HLA-DQA1 was associated with LOR (OR = 1.16, (CI = 1.1-1.3), P = 5.04e-4), however the effect could not be determined to be distinct (Adjusted p-value = 0.09). The SNP in *HLA-DQA1* is located in the Human Leukocyte Antigen locus containing more than 140 genes involved in both innate and adaptive immunity (78) and confers risk of both autoimmune and infectious diseases (79) including hay fever (53). Despite the inability to determine if this association is specific to LOR, we thus provide preliminary evidence that HLA-variants influence asthma with later onset suggesting the involvement of

atopic mechanisms of the HLA-locus on asthma.

The previously identified epistatic interaction between *GSDMB* and *CDHR3* was only suggestively associated with the phenotypes, likely due to low power due to the low frequency of SNPxSNP interactions.

ETR was significantly associated with asthma scores, but displayed small effect sizes compared to the remaining phenotypes, whereas both ETR and PTR were weakly influenced by genetic risk of eczema, allergy and eosinophil levels. PTR displayed stronger associations with asthma genetic scores than ETR, suggesting higher genetic similarity to asthma but little influence by atopy, indicating that PTR represents non-atopic childhood wheeze. Contrarily, PR and LOR displayed strong association with both asthma and atopic scores. PR was strongly influenced by PRSs for asthma, allergy and eosinophils, and the within-class ANOVA indicated that the associations were stronger than for the other phenotypes (see Table 4). Congruently, LOR was strongly associated with increased eosinophils and hayfever, reinforcing the notion that wheeze symptoms in later childhood are associated with atopy (3,8,13).

The PRS for lung function (FEV1/FVC) was associated with all phenotypes and we found little evidence of distinct associations (Adjusted P-value = 0.051). Significant associations with the BMI-PRS were confined to ETR and PTR, with little evidence of distinctness between the phenotypes (Adjusted P-value = 0.41). However, high BMI is associated with asthma and the overlap is thought to be higher among non-atopic compared to atopic asthmatics (77, 78). Additionally, genetic risk of BMI is suggested to

increase risk of viral infection and wheeze in childhood, independently of observed BMI (82). Our results indicate that PR and LOR, i.e. atopic phenotypes that likely represent true asthma, are not influenced by genetic BMI. Genetic BMI seems to influence remitting phenotypes with little evidence of limited atopy and high burden of infections. ETR seems to be involved in non-genetic risk for infections, while PTR likely displays congenital predisposition to infectious diseases, perhaps through maternal susceptibility and *FUT2*.

Limitations to the study is that the study is performed in a cohort sampled from the general population, with increased probability of sampling individuals with psychiatric disorders (30,31). In our selected population 59 % of our study population had at least one psychiatric diagnosis and we addressed this issue by controlling for psychiatric status, in all association analysis in this study. In a sensitivity analysis, we observed limited interaction between psychiatric status and the risk factors, suggesting minimal influence of psychiatric illnesses on our results. One exception was the PRS for childhood-onset asthma that had a stronger effect on PR in individuals with no psychiatric disorders compared to psychiatric cases, implying that a stronger effect might have been observed in the absence of psychiatric cases.

The PRS for ADHD was strongly associated with ETR but only weakly associated with LOR and PTR and not with PR. An ADHD diagnosis was significantly associated with both ETR and PTR, indicating a non-genetic relationship between ADHD and PTR. While a positive association between ADHD and childhood asthma is well-documented (83), our findings indicate that neither the genetic predisposition to ADHD nor the

observed ADHD risk is linked to trajectories indicative of atopic asthma. Instead, these associations appear to relate to transient wheeze symptoms with minimal evidence of atopy. The association with genetic risk of ADHD and ETR and PTR could be confounded by non-adherence to medication, since individuals with predisposition to ADHD-symptoms may stop redeeming medication at specific time-points, not because they are symptom-free, but due of poor drug adherence (56).

We identified associations between polygenic risk scores for Autism with ETR but not with the remaining phenotypes. However, when investigating the associations between the ICS trajectories and observed psychiatric diagnoses, we adjusted for maternal smoking during pregnancy, gestational age, number of older siblings, paternal low education and household income, which caused the effect to attenuate, suggesting socioeconomic factors play a larger role than shared etiology. Consistently, the overlap between autism and childhood asthma is not well-established (83). The association with ADHD and ETR/PTR were largely unaffected by the before-mentioned covariates, indicating a more robust overlap between the two conditions. The PRSs for ADHD and autism were derived using participants from the iPSYCH cohort, but since the unadjusted association patterns between ETR and the observed ADHD and autism outcomes closely resembled those found in PRS-based analyses, we expect minimal inflation of the PRS analyses. Adjusting the PRS analyses for lifestyle covariates are beyond the scope of this paper due to the complex interactions between genetic risk of behavioral traits and environment (84). However, these results motivate further research into how gene-environment correlations affect the relationship between preschool

wheeze and mental disorders.

Another limitation of this study is the lack of observed measurement data. Previous studies examining wheeze trajectories frequently incorporate blood tests for sensitization and eosinophil counts (13,15,16), offering more precise indicators of atopic conditions. Additionally, lung function measurements were unavailable. To address these gaps, we used polygenic risk scores (PRS) to assess genetic predisposition for sensitization, rhinitis, eosinophils, and lung function traits as a proxy for observed measurements. It is also a limitation that trajectories are defined retrospectively thus limiting the diagnostic utility.

A strength of our study is the high longitudinal data-granularity compared to actively recruited cohorts, where data is often missing at random due to attrition (85). This allows the study of detailed longitudinal profiles of primary and secondary endpoints, exemplified by our investigation of hospitalization secondary-prescription trajectories. Another strength of our study is sample size. Coupling identified trajectories with 28 thousands controls, allows for accurate estimates in statistical analysis and the large case group ($n = 7527$) provides a substantial power for our within-class ANOVA test. Larger sample size also increases power for analysis of rare events, such as respiratory distress and prematurity (see Table 1). Additionally, fathers are more represented in EHR compared to the observational cohort, where fathers are historically underrepresented (86).

Conclusion

In summary, we identified four distinct trajectories of ICS prescriptions in early life, influenced by unique risk factors. Early-transient redeemers were characterized by low genetic effects and high influence of in utero tobacco smoke, presence of older siblings and young parents. ETR was also influenced by polygenic risk scores for ADHD. Prolonged-transient redeemers were characterized by maternal antibiotic use before birth, low gestational age, a distinct association with *FUT2*, no association with atopic risk scores and strong associations to asthma PRS. Persistent redeemers were characterized by high prevalence of birth by C-section and male-gender, with high genetic effects for almost all investigated SNPs and atopic polygenic risk scores, including asthma. Late-onset redeemers displayed weak impact of perinatal risk factors, high polygenic risk scores for asthma and atopic traits and showed a strong, but not distinct, association to *HLA-DQA1*. Overall, our findings show that the phenotypes are distinct and influenced by different risk factors, suggesting the presence of endotypes.

References

1. Granell R, Henderson AJ, Sterne JA. Associations of wheezing phenotypes with late asthma outcomes in the Avon Longitudinal Study of Parents and Children: A population-based birth cohort. *J Allergy Clin Immunol* [Internet]. [cited 2016 May 3]; Available from: <http://www.sciencedirect.com/science/article/pii/S0091674916003833>
2. Taussig LM, Wright AL, Holberg CJ, Halonen M, Morgan WJ, Martinez FD. Tucson children's respiratory study: 1980 to present. *J Allergy Clin Immunol*. 2003 Apr;111(4):661–75.
3. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax*. 2008 Nov 1;63:974–80.
4. Belgrave DC, Custovic A, Simpson A. Characterizing wheeze phenotypes to identify endotypes of childhood asthma, and the implications for future management. *Expert Rev Clin Immunol*. 2013 Oct;9(10):921–36.
5. Lowe LA, Simpson A, Woodcock A, Morris J, Murray CS, Custovic A, et al. Wheeze Phenotypes and Lung Function in Preschool Children. *Am J Respir Crit Care Med*. 2005 Feb 1;171(3):231–7.
6. Sonnappa S, Bastardo CM, Wade A, Saglani S, McKenzie SA, Bush A, et al. Symptom-pattern phenotype and pulmonary function in preschool wheezers. *J Allergy Clin Immunol*. 2010 Sep;126(3):519–526.e7.
7. Sherriff A, Peters TJ, Henderson J, Strachan D. Risk factor associations with wheezing patterns in children followed longitudinally from birth to 3 ½ years. *Int J Epidemiol*. 2001 Dec;30(6):1473–84.
8. 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.
9. Spycher BD, Cochrane C, Granell R, Sterne JAC, Silverman M, Pedersen E, et al. Temporal stability of multitrigger and episodic viral wheeze in early childhood. *Eur Respir J*. 2017 Nov;50(5):1700014.
10. Bisgaard H, Bønnelykke K. Long-term studies of the natural history of asthma in childhood. *J Allergy Clin Immunol*. 2010 Aug;126(2):187–97.
11. McCready C, Haider S, Little F, Nicol MP, Workman L, Gray DM, et al. Early childhood wheezing phenotypes and determinants in a South African birth cohort: longitudinal analysis of the Drakenstein Child Health Study. *Lancet Child Adolesc Health*. 2023 Feb;7(2):127–35.
12. Deliu M, Fontanella S, Haider S, Sperrin M, Geifman N, Murray C, et al. Longitudinal trajectories of severe wheeze exacerbations from infancy to school age and their association with early-life risk factors and late asthma outcomes. *Clin Exp Allergy*. 2020 Mar;50(3):315–24.
13. Haider S, Granell R, Curtin J, Fontanella S, Cucco A, Turner S, et al. Modeling Wheezing Spells Identifies Phenotypes with Different Outcomes and Genetic Associates. *Am J Respir Crit Care Med*. 2022 Apr 15;205(8):883–93.
14. Lau HX, Chen Z, Van Bever H, Tham EH, Chan YH, Yap QV, et al. Clinical predictors of wheeze trajectories and associations with allergy in Asian children. *Ann Allergy Asthma Immunol*. 2023 Oct;131(4):466–473.e6.
15. Oksel C, Granell R, Haider S, Fontanella S, Simpson A, Turner S, et al. Distinguishing Wheezing Phenotypes from Infancy to Adolescence. A Pooled Analysis of Five Birth Cohorts. *Ann Am Thorac Soc*. 2019 Jul;16(7):868–76.
16. Dai R, Miliku K, Gaddipati S, Choi J, Ambalavanan A, Tran MM, et al. Wheeze trajectories:

- Determinants and outcomes in the CHILD Cohort Study. *J Allergy Clin Immunol*. 2022 Jun;149(6):2153–65.
17. Sordillo JE, Coull BA, Rifas-Shiman SL, Wu AC, Lutz SM, Hivert MF, et al. Characterization of longitudinal wheeze phenotypes from infancy to adolescence in Project Viva, a prebirth cohort study. *J Allergy Clin Immunol*. 2020 Feb;145(2):716-719.e8.
 18. Tse SM, Rifas-Shiman SL, Coull BA, Litonjua AA, Oken E, Gold DR. Sex-specific risk factors for childhood wheeze and longitudinal phenotypes of wheeze. *J Allergy Clin Immunol*. 2016 Dec;138(6):1561-1568.e6.
 19. Oksel C, Haider S, Fontanella S, Frainay C, Custovic A. Classification of Pediatric Asthma: From Phenotype Discovery to Clinical Practice. *Front Pediatr*. 2018 Sep 20;6:258.
 20. Oksel C, Granell R, Mahmoud O, Custovic A, Henderson AJ. Causes of variability in latent phenotypes of childhood wheeze. *J Allergy Clin Immunol*. 2019 May;143(5):1783-1790.e11.
 21. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol*. 2016 Oct 27;dyw213.
 22. Moth G, Vedsted P, Schiøtz P. Identification of asthmatic children using prescription data and diagnosis. *Eur J Clin Pharmacol*. 2007 Apr 27;63(6):605–11.
 23. Mommers I, Van Boven JF, Schuiling-Veninga CC, Bos JH, Koetsier M, Hak E, et al. Real-World Dispensing Patterns of Inhalation Medication in Young Adult Asthma: An Inception Cohort Study. *Clin Epidemiol*. 2023 Jun;Volume 15:721–32.
 24. Von Bülow A, Hansen S, Sandin P, Ernstsson O, Janson C, Lehtimäki L, et al. Severe asthma trajectories in adults: findings from the NORDSTAR cohort. *Eur Respir J*. 2023 Sep;62(3):2202474.
 25. Lee E, Rhee EH, Kim K, Kim HS, Kim WK, Song DJ, et al. Frequency of exacerbation and degree of required asthma medication can characterize childhood longitudinal asthma trajectories. *Ann Allergy Asthma Immunol*. 2023 Oct;131(4):444–50.
 26. Markus AF, Rijnbeek PR, Kors JA, Burn E, Duarte-Salles T, Haug M, et al. Real-world treatment trajectories of adults with newly diagnosed asthma or COPD. *BMJ Open Respir Res*. 2024 Feb;11(1):e002127.
 27. Pape K, Cowell W, Sejbaek CS, Andersson NW, Svanes C, Kolstad HA, et al. Adverse childhood experiences and asthma: trajectories in a national cohort. *Thorax*. 2021 Jun;76(6):547–53.
 28. Liu X, Plana-Ripoll O, McGrath JJ, Petersen LV, Dharmage SC, Momen NC. Bidirectional Associations Between Asthma and Types of Mental Disorders. *J Allergy Clin Immunol Pract*. 2023 Mar;11(3):799-808.e14.
 29. Zhu Z, Zhu X, Liu CL, Shi H, Shen S, Yang Y, et al. Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. *Eur Respir J*. 2019 Dec;54(6):1901507.
 30. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry*. 2018 Jan;23(1):6–14.
 31. Bybjerg-Grauholm J, Bøcker Pedersen C, Bækvad-Hansen M, Giørtz Pedersen M, Adamsen D, Søholm Hansen C, et al. The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders [Internet]. *Epidemiology*; 2020 Dec [cited 2023 Mar 20]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.11.30.20237768>
 32. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet*. 2023 Feb;55(2):198–208.
 33. Global Strategy for Asthma Management and Prevention. [cited 2018 Aug 20]; Available

from: <http://ginasthma.org/>

34. Linzer DA, Lewis JB. **poLCA**: An *R* Package for Polytomous Variable Latent Class Analysis. *J Stat Softw* [Internet]. 2011 [cited 2024 Sep 30];42(10). Available from: <http://www.jstatsoft.org/v42/i10/>
35. Sinha P, Calfee CS, Delucchi KL. Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. *Crit Care Med*. 2021 Jan;49(1):e63–79.
36. Tein JY, Coxe S, Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Struct Equ Model Multidiscip J*. 2013 Oct;20(4):640–57.
37. Nagin D. Group-Based Modeling of Development [Internet]. Harvard University Press; 2005 [cited 2024 Aug 30]. Available from: <https://hup.degruyter.com/view/title/514871>
38. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. Stegle O, editor. *Bioinformatics*. 2018 Aug 15;34(16):2781–7.
39. Ahluwalia TS, Eliassen AU, Sevelsted A, Pedersen CET, Stokholm J, Chawes B, et al. FUT2–ABO epistasis increases the risk of early childhood asthma and *Streptococcus pneumoniae* respiratory illnesses. *Nat Commun*. 2020 Dec 16;11:6398.
40. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet*. 2014 Jan;46(1):51–5.
41. Granell R, Curtin JA, Haider S, Kitaba NT, Mathie SA, Gregory LG, et al. A meta-analysis of genome-wide association studies of childhood wheezing phenotypes identifies ANXA1 as a susceptibility locus for persistent wheezing. *eLife*. 2023 May 25;12:e84315.
42. Eliassen AU, Pedersen CET, Rasmussen MA, Wang N, Soverini M, Fritz A, et al. Genome-wide study of early and severe childhood asthma identifies interaction between CDHR3 and GSDMB. *J Allergy Clin Immunol*. 2022 Sep;150(3):622–30.
43. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019 Jun;7(6):509–22.
44. Namjou B, Lape M, Malolepsza E, DeVore SB, Weirauch MT, Dikilitas O, et al. Multiancestral polygenic risk score for pediatric asthma. *J Allergy Clin Immunol*. 2022 Nov;150(5):1086–96.
45. Privé F, Albiñana C, Arbel J, Pasaniuc B, Vilhjálmsson BJ. Inferring disease architecture and predictive ability with LDpred2-auto. *Am J Hum Genet*. 2023 Dec;110(12):2042–55.
46. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019 Mar;51(3):481–93.
47. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020 Sep 3;182(5):1214–1231.e11.
48. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shiri M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019 Mar;22(3):343–52.
49. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*. 2021 Jun;53(6):817–29.
50. Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team, Grove J, Ripke S, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019 Mar;51(3):431–44.
51. Purves KL, Coleman JRI, Meier SM, Rayner C, Davis KAS, Cheesman R, et al. A major role for common genetic variation in anxiety disorders. *Mol Psychiatry*. 2020

Dec;25(12):3292–303.

52. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshihara S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet.* 2021 Oct;53(10):1415–24.
53. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat Genet.* 2018 Aug;50(8):1072.
54. Said S, Pazoki R, Karhunen V, Vösa U, Ligthart S, Bodinier B, et al. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nat Commun.* 2022 Apr 22;13(1):2198.
55. Brite J, Friedman S, De La Hoz RE, Reibman J, Cone J. Mental health, long-term medication adherence, and the control of asthma symptoms among persons exposed to the WTC 9/11 disaster. *J Asthma.* 2020 Nov 1;57(11):1253–62.
56. Walid F. Gellad, Jerry Grenard, Elizabeth A. McGlynn. A Review of Barriers to Medication Adherence: A Framework for Driving Policy Options. [cited 2010 Jun 10]; Available from: http://www.rand.org/pubs/technical_reports/2009/RAND_TR765.pdf
57. Sevelsted A, Stokholm J, Bisgaard H. Risk of Asthma from Cesarean Delivery Depends on Membrane Rupture. *J Pediatr.* 2016 Apr;171:38–42.e1–4.
58. Stokholm J, Thorsen J, Blaser MJ, Rasmussen MA, Hjelmsø M, Shah S, et al. Delivery mode and gut microbial changes correlate with an increased risk of childhood asthma. *Sci Transl Med.* 2020 Nov 11;12(569).
59. Kyvsgaard JN, Chawes BL, Horner DLG, Hesselberg LM, Melgaard ME, Jensen SK, et al. Risk Factors and Age-Related Patterns of Asthma-Like Symptoms in Early Childhood. *J Allergy Clin Immunol Pract.* 2023 Jun;11(6):1773–1784.e10.
60. Blaser MJ, Bello MGD. Maternal antibiotic use and risk of asthma in offspring. *Lancet Respir Med.* 2014 Oct;2(10):e16.
61. Hviid A, Melbye M. The Impact of Birth Weight on Infectious Disease Hospitalization in Childhood. *Am J Epidemiol.* 2007 Feb 19;165(7):756–61.
62. Hardelid P, Verfuenden M, McMenamin J, Gilbert R. Risk factors for admission to hospital with laboratory-confirmed influenza in young children: birth cohort study. *Eur Respir J.* 2017 Sep;50(3):1700489.
63. Mendola P, Männistö TI, Leishear K, Reddy UM, Chen Z, Laughon SK. Neonatal health of infants born to mothers with asthma. *J Allergy Clin Immunol.* 2014 Jan;133(1):85–90.e4.
64. Montgomery S, Bahmanyar S, Brus O, Hussein O, Kosma P, Palme-Kilander C. Respiratory infections in preterm infants and subsequent asthma: a cohort study. *BMJ Open.* 2013 Oct;3(10):e004034.
65. Bisgaard H, Jensen SM, Bønnelykke K. Interaction between Asthma and Lung Function Growth in Early Life. *Am J Respir Crit Care Med.* 2012 Jun 1;185(11):1183–9.
66. Çalışkan M, Bochkov YA, Kreiner-Møller E, Bønnelykke K, Stein MM, Du G, et al. Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med.* 2013 Apr 11;368(15):1398–407.
67. Bochkov YA, Watters K, Ashraf S, Griggs TF, Devries MK, Jackson DJ, et al. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc Natl Acad Sci U S A.* 2015 Apr 28;112(17):5485–90.
68. Bønnelykke K, Coleman AT, Evans MD, Thorsen J, Waage J, Vissing NH, et al. Cadherin-related Family Member 3 Genetics and Rhinovirus C Respiratory Illnesses. *Am J Respir Crit Care Med.* 2018 Mar;197(5):589–594.
69. Savenije OE, Mahachie John JM, Granell R, Kerkhof M, Dijk FN, De Jongste JC, et al. Association of IL33–IL-1 receptor–like 1 (IL1RL1) pathway polymorphisms with wheezing phenotypes and asthma in childhood. *J Allergy Clin Immunol.* 2014 Jul;134(1):170–7.

70. Wills-Karp M. Interleukin-13 in asthma pathogenesis. *Immunol Rev.* 2004 Dec;202(1):175–90.
71. Ebina-Shibuya R, Leonard WJ. Role of thymic stromal lymphopoietin in allergy and beyond. *Nat Rev Immunol.* 2023 Jan;23(1):24–37.
72. Shikotra A, Choy DF, Ohri CM, Doran E, Butler C, Hargadon B, et al. Increased expression of immunoreactive thymic stromal lymphopoietin in patients with severe asthma. *J Allergy Clin Immunol.* 2012 Jan;129(1):104–111.e9.
73. Murrison LB, Ren X, Preusse K, He H, Kroner J, Chen X, et al. TSLP disease-associated genetic variants combined with airway TSLP expression influence asthma risk. *J Allergy Clin Immunol.* 2022 Jan;149(1):79–88.
74. Wang W, Li Y, Lv Z, Chen Y, Li Y, Huang K, et al. Bronchial Allergen Challenge of Patients with Atopic Asthma Triggers an Alarmin (IL-33, TSLP, and IL-25) Response in the Airways Epithelium and Submucosa. *J Immunol.* 2018 Oct 15;201(8):2221–31.
75. Barton SJ, Murray R, Lillycrop KA, Inskip HM, Harvey NC, Cooper C, et al. *FUT2* Genetic Variants and Reported Respiratory and Gastrointestinal Illnesses During Infancy. *J Infect Dis.* 2019 Feb 15;219(5):836–43.
76. Kindberg E, Akerlind B, Johnsen C, Knudsen JD, Heltberg O, Larson G, et al. Host genetic resistance to symptomatic norovirus (GGII.4) infections in Denmark. *J Clin Microbiol.* 2007 Aug;45(8):2720–2.
77. Imbert-Marcille BM, Barbé L, Dupé M, Le Moullac-Vaidye B, Besse B, Peltier C, et al. A *FUT2* Gene Common Polymorphism Determines Resistance to Rotavirus A of the P[8] Genotype. *J Infect Dis.* 2014 Apr 15;209(8):1227–30.
78. Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var.* 2022 Dec 21;9(1):49.
79. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* 2017 Dec;18(1):76.
80. Appleton S, Adams R, Wilson D, Taylor A, Ruffin R. Central obesity is associated with nonatopic but not atopic asthma in a representative population sample. *J Allergy Clin Immunol.* 2006 Dec;118(6):1284–91.
81. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol.* 2020 Feb;145(2):537–49.
82. Jensen SK, Pedersen CET, Fischer-Rasmussen K, Melgaard ME, Brustad N, Kyvsgaard JN, et al. Genetic predisposition to high BMI increases risk of early life respiratory infections and episodes of severe wheeze and asthma. *Eur Respir J.* 2024 Sep;64(3):2400169.
83. Kaas TH, Vinding RK, Stokholm J, Bønnelykke K, Bisgaard H, Chawes BL. Association between childhood asthma and attention deficit hyperactivity or autism spectrum disorders: A systematic review with meta-analysis. *Clin Exp Allergy J Br Soc Allergy Clin Immunol.* 2021 Feb;51(2):228–252.
84. Abdellaoui A, Dolan CV, Verweij KJH, Nivard MG. Gene–environment correlations across geographic regions affect genome-wide association studies. *Nat Genet.* 2022 Sep;54(9):1345–54.
85. Teixeira R, Doetsch J, Freitas AI, Lortie E, Santos AC, Barros H, et al. Survey of data collection methods and retention strategies in European birth cohorts of children and adults born very preterm. *Paediatr Perinat Epidemiol.* 2022 Sep;36(5):706–14.
86. Davison KK, Gicevic S, Aftosmes-Tobio A, Ganter C, Simon CL, Newlan S, et al. Fathers' Representation in Observational Studies on Parenting and Childhood Obesity: A Systematic Review and Content Analysis. *Am J Public Health.* 2016 Nov;106(11):e14–21.

Tables

Table 1: Characteristics of no redeemers compared to redeemers of any number of ICS. Odds-ratio (OR), confidence intervals (CI 95 %) and P-values refer to logistic regression with any ICS as outcome.

Characteristic	Non-redeemers	ICS redeemers	OR	CI 95%	P-value
Cesarian section	17.5%	19.3%	1.20	1.1-1.3	5.66e-08
Smoking during pregnancy	24.9%	31.7%	1.31	1.2-1.4	2.15e-19
Premature	5.3%	8%	1.57	1.4-1.7	8.58e-19
Gestational age	39.4	39.1	0.86	0.8-0.9	5.60e-34
Birth weight	3535.5	3482	0.90	0.9-0.9	2.62e-16
Older siblings	0.8	0.8	1.03	1-1.1	1.29e-01
Distance to older sib.	1545.1	1560.2	1.01	1-1.1	7.95e-01
Mothers age	29.6	29.1	0.92	0.9-1	2.99e-06
Fathers age	32.2	31.8	0.93	0.9-1	7.43e-05
Low education father	1.8%	2.3%	1.56	1.1-2.2	7.91e-03
Low education mother	0.8%	1.2%	1.20	0.9-1.5	1.32e-01
Household income	384453 DKK	363285 DKK	0.87	0.8-0.9	6.76e-08
Female-gender	40.64%	33.41%	0.86	0.8-0.9	2.31e-28
Neonate resp. distress	3.2%	4.3%	1.11	1-1.3	1.52e-01
Mothers Asthma	10.5%	15%	1.73	1.6-1.9	2.90e-40
Mothers Allergy	18.7%	20.4%	1.34	1.2-1.4	3.06e-15
Mothers Steroid cream	22.1%	22.7%	1.21	1.1-1.3	1.05e-07
Mothers Antibiotics	36.8%	36.5%	1.21	1.1-1.3	1.31e-07
Fathers Asthma	4.8%	6%	1.50	1.3-1.7	7.33e-12
Fathers Allergy	9%	9.6%	1.28	1.2-1.4	5.08e-07
Fathers Steroid cream	9.6%	9.7%	1.22	1.1-1.3	4.80e-05
Fathers Antibiotics	23%	22%	1.16	1.1-1.2	2.32e-04

Table 2: Results of multinomial regression across the 4 phenotypes and including baseline as controls. ANOVA designates the FDR-corrected p-value for the within-class ANOVA tests.

Exposure	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			ANOVA
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
Cesarian section	1.14	1.04-1.2	3.90e-03	1.29	1.13-1.5	1.19e-04	1.50	1.26-1.8	3.47e-06	1.13	0.97-1.3	1.09e-01	3.025e-02*
Smoking during pregnancy	1.48	1.37-1.6	0.00e+00	1.23	1.1-1.4	4.85e-04	0.90	0.76-1.1	2.32e-01	1.19	1.05-1.4	7.81e-03	1.146e-06*
Premature	1.47	1.29-1.7	1.33e-08	2.02	1.7-2.4	4.88e-15	1.74	1.35-2.2	2.08e-05	1.28	1.02-1.6	3.53e-02	1.379e-02*
Gestational age	0.86	0.84-0.9	0.00e+00	0.81	0.78-0.8	0.00e+00	0.87	0.81-0.9	1.46e-05	0.91	0.86-1	5.76e-04	2.840e-02*
Birth weight	0.88	0.85-0.9	1.33e-13	0.87	0.82-0.9	1.32e-08	0.96	0.89-1	2.24e-01	0.96	0.91-1	1.43e-01	2.936e-02*
Female-gender	0.88	0.85-0.9	4.35e-13	0.84	0.8-0.9	1.66e-09	0.77	0.72-0.8	1.54e-10	0.87	0.82-0.9	3.44e-06	4.139e-02*
Neonate resp. distress	1.01	0.84-1.2	9.03e-01	1.38	1.08-1.8	1.00e-02	1.49	1.05-2.1	2.38e-02	0.79	0.55-1.1	1.91e-01	0.000e+00*
Older siblings	1.09	1.04-1.1	1.87e-04	0.98	0.91-1.1	5.82e-01	1.01	0.91-1.1	9.14e-01	0.91	0.83-1	2.46e-02	2.339e-03*
Distance to older sib.	0.98	0.92-1	4.47e-01	0.95	0.86-1	2.32e-01	1.12	1.01-1.3	3.91e-02	1.09	1-1.2	5.65e-02	4.139e-02*
Mothers age	0.89	0.85-0.9	1.59e-07	0.91	0.85-1	3.75e-03	1.07	0.98-1.2	1.47e-01	0.97	0.9-1	3.37e-01	4.539e-03*
Fathers age	0.89	0.85-0.9	8.17e-07	0.94	0.88-1	6.66e-02	1.09	1-1.2	4.83e-02	0.96	0.89-1	2.24e-01	3.529e-03*
Low education father	1.45	0.94-2.2	9.25e-02	1.63	0.89-3	1.13e-01	2.16	1.04-4.5	3.80e-02	1.44	0.73-2.9	2.98e-01	7.883e-01
Low education mother	1.29	0.96-1.7	9.42e-02	1.07	0.67-1.7	7.84e-01	1.46	0.83-2.6	1.93e-01	0.94	0.54-1.6	8.13e-01	7.367e-01
Household income	0.85	0.8-0.9	4.30e-06	0.87	0.78-1	5.65e-03	0.93	0.82-1.1	2.39e-01	0.90	0.81-1	4.10e-02	7.367e-01
Mothers Asthma	1.75	1.58-1.9	0.00e+00	1.85	1.58-2.2	4.66e-15	1.72	1.37-2.2	3.51e-06	1.52	1.27-1.8	7.63e-06	6.694e-01
Mothers Allergy	1.24	1.13-1.4	9.25e-06	1.52	1.32-1.8	3.56e-09	1.45	1.18-1.8	4.49e-04	1.39	1.18-1.6	6.69e-05	8.600e-02
Mothers Steroid_cream	1.18	1.07-1.3	4.76e-04	1.27	1.11-1.5	7.27e-04	1.17	0.95-1.4	1.33e-01	1.27	1.09-1.5	2.71e-03	7.367e-01
Mothers Antibiotics	1.19	1.08-1.3	3.24e-04	1.50	1.29-1.7	5.05e-08	1.02	0.83-1.3	8.63e-01	1.11	0.95-1.3	1.96e-01	2.652e-02*
Fathers Asthma	1.36	1.17-1.6	7.93e-05	1.72	1.39-2.1	7.42e-07	1.61	1.15-2.2	5.02e-03	1.63	1.27-2.1	1.11e-04	3.448e-01
Fathers Allergy	1.16	1.02-1.3	2.01e-02	1.32	1.1-1.6	2.96e-03	1.49	1.14-2	3.42e-03	1.49	1.21-1.8	1.35e-04	1.417e-01
Fathers Steroid_cream	1.21	1.07-1.4	2.47e-03	1.28	1.07-1.5	7.66e-03	1.32	1.01-1.7	4.58e-02	1.09	0.88-1.4	4.40e-01	7.367e-01
Fathers Antibiotics	1.14	1.03-1.3	1.20e-02	1.21	1.03-1.4	1.65e-02	1.18	0.94-1.5	1.56e-01	1.12	0.95-1.3	1.84e-01	8.879e-01

Table 3: Results of genetic analysis using multinomial regression across the 4 phenotypes and including no redeemers as controls. ANOVA designates the FDR-corrected p-value for the within-class ANOVA tests.

Locus	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			ANOVA
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
<i>GSDMB</i>	1.09	1.03-1.1	8.38e-04	1.36	1.26-1.5	4.44e-16	1.52	1.37-1.7	5.33e-15	1.20	1.1-1.3	1.06e-05	4.790e-10*
<i>CDHR3</i>	1.17	1.1-1.2	4.99e-07	1.16	1.06-1.3	1.90e-03	1.45	1.28-1.6	2.31e-09	1.07	0.96-1.2	2.09e-01	8.371e-03*
<i>FUT2</i>	1.00	0.95-1	9.78e-01	1.14	1.06-1.2	4.73e-04	1.05	0.95-1.2	3.37e-01	1.01	0.94-1.1	7.36e-01	3.517e-02*
<i>HLA-DQA1</i>	1.02	0.97-1.1	4.65e-01	1.06	0.98-1.1	1.58e-01	1.04	0.94-1.2	4.17e-01	1.16	1.07-1.3	5.04e-04	9.113e-02
<i>ANXA1</i>	0.93	0.79-1.1	3.38e-01	0.89	0.69-1.1	3.35e-01	0.62	0.42-0.9	1.74e-02	1.12	0.88-1.4	3.62e-01	9.113e-02
<i>IL33-1</i>	1.06	0.99-1.1	1.06e-01	1.05	0.95-1.2	3.14e-01	1.37	1.17-1.6	6.53e-05	1.06	0.95-1.2	3.12e-01	3.151e-02*
<i>IL33-2</i>	1.01	0.95-1.1	7.08e-01	1.12	1.01-1.2	2.86e-02	1.25	1.09-1.4	9.44e-04	1.13	1.02-1.3	2.05e-02	3.517e-02*
<i>IL1RL1</i>	0.95	0.89-1	2.11e-01	1.16	1.03-1.3	1.52e-02	1.24	1.05-1.5	1.22e-02	1.13	1-1.3	5.40e-02	8.223e-03*
<i>IL13</i>	1.03	0.97-1.1	3.19e-01	1.05	0.96-1.1	3.17e-01	1.27	1.13-1.4	7.76e-05	0.99	0.9-1.1	9.14e-01	3.151e-02*
<i>TSLP</i>	1.08	1.03-1.1	1.32e-03	1.13	1.05-1.2	1.45e-03	1.31	1.19-1.5	1.44e-07	1.13	1.04-1.2	2.55e-03	3.151e-02*
<i>ABO</i>	1.01	0.96-1.1	7.09e-01	1.04	0.96-1.1	3.15e-01	1.08	0.98-1.2	1.27e-01	0.99	0.91-1.1	7.98e-01	5.299e-01
<i>GSDMBxCDHR3</i>	1.09	1-1.2	6.19e-02	1.15	1-1.3	4.53e-02	1.22	1.02-1.5	3.32e-02	1.19	1.02-1.4	2.31e-02	5.110e-01
<i>FUT2xABO</i>	0.94	0.88-1	1.08e-01	1.05	0.94-1.2	3.54e-01	1.06	0.92-1.2	4.09e-01	1.00	0.89-1.1	9.38e-01	3.125e-01

Table 4: Results of polygenic risk score (PRS) analysis using multinomial regression across the 4 phenotypes and including baseline as controls. Baseline here is individuals who never redeemed a single ICS in the first 9 years of life.

PRS	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			ANOVA
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
BMI	1.10	1.06-1.14	4.17e-08	1.08	1.03-1.14	2.46e-03	1.04	0.97-1.12	2.40e-01	1.05	0.99-1.11	7.57e-02	4.757e-01
Allergic rhinitis	1.03	0.99-1.06	1.37e-01	1.06	1.01-1.12	2.67e-02	1.19	1.11-1.28	2.98e-06	1.12	1.06-1.18	1.34e-04	3.029e-03*
Allergic sensitization	1.00	0.97-1.04	8.45e-01	1.04	0.99-1.1	1.28e-01	1.18	1.09-1.26	1.04e-05	1.10	1.04-1.17	6.61e-04	6.130e-04*
Atopic dermatitis	1.05	1.02-1.09	3.69e-03	1.03	0.97-1.08	3.36e-01	1.12	1.04-1.2	1.79e-03	1.08	1.02-1.14	6.31e-03	3.384e-01
Adult onset asthma	1.06	1.03-1.1	5.01e-04	1.15	1.09-1.22	1.80e-07	1.28	1.19-1.38	5.66e-11	1.22	1.15-1.3	1.48e-11	2.549e-06*
Childhood onset asthma	1.12	1.08-1.16	9.99e-11	1.27	1.2-1.34	0.00e+00	1.73	1.61-1.87	0.00e+00	1.38	1.3-1.46	0.00e+00	0.000e+00*
Severe asthma TAGC	1.13	1.09-1.18	3.24e-11	1.25	1.18-1.32	3.51e-14	1.52	1.4-1.64	0.00e+00	1.32	1.25-1.41	0.00e+00	2.605e-11*
FEV1/FVC	0.89	0.85-0.92	1.07e-10	0.82	0.77-0.87	3.01e-12	0.80	0.74-0.86	1.07e-08	0.86	0.81-0.91	1.20e-06	5.108e-02
Autism	0.92	0.89-0.96	1.30e-05	0.98	0.93-1.04	5.55e-01	1.01	0.94-1.09	7.68e-01	1.01	0.95-1.07	8.31e-01	3.392e-02*
Anxiety	1.00	0.96-1.03	8.82e-01	1.05	0.99-1.11	1.04e-01	1.00	0.92-1.08	9.44e-01	1.05	0.99-1.11	1.19e-01	3.795e-01
ADHD	1.13	1.08-1.17	2.46e-10	1.05	0.99-1.11	1.14e-01	1.02	0.94-1.1	6.83e-01	1.08	1.01-1.14	1.91e-02	5.108e-02
Bipolar	0.99	0.96-1.03	7.48e-01	0.96	0.91-1.02	2.18e-01	0.98	0.9-1.06	5.53e-01	0.97	0.91-1.03	3.05e-01	8.114e-01
Depression	1.05	1.02-1.09	3.55e-03	1.07	1.01-1.13	2.11e-02	1.03	0.96-1.11	4.38e-01	1.07	1.01-1.14	2.01e-02	8.114e-01
CRP	1.03	1-1.07	9.06e-02	1.03	0.98-1.08	2.71e-01	0.99	0.92-1.07	8.58e-01	1.03	0.97-1.09	3.15e-01	8.114e-01
Eosinophils	1.04	1-1.08	6.29e-02	1.04	0.98-1.1	1.92e-01	1.23	1.14-1.34	6.40e-07	1.17	1.1-1.25	1.40e-06	1.304e-04*
Neutrophils	1.01	0.97-1.05	5.01e-01	1.04	0.98-1.11	1.59e-01	1.06	0.98-1.15	1.54e-01	1.08	1.02-1.16	1.49e-02	3.735e-01

Figures

Figure 1: Trajectories identified in ICS pattern prescription use. Life years of ICS redeeming are shown on the x-axis and frequency of ICS redeeming within each life-year is shown on the y-axis. The four phenotypes are early-transient redeemers (ETR), late-onset redeemers (LOR), persistent redeemers (PR) and prolonged-transient redeemers (PTR).

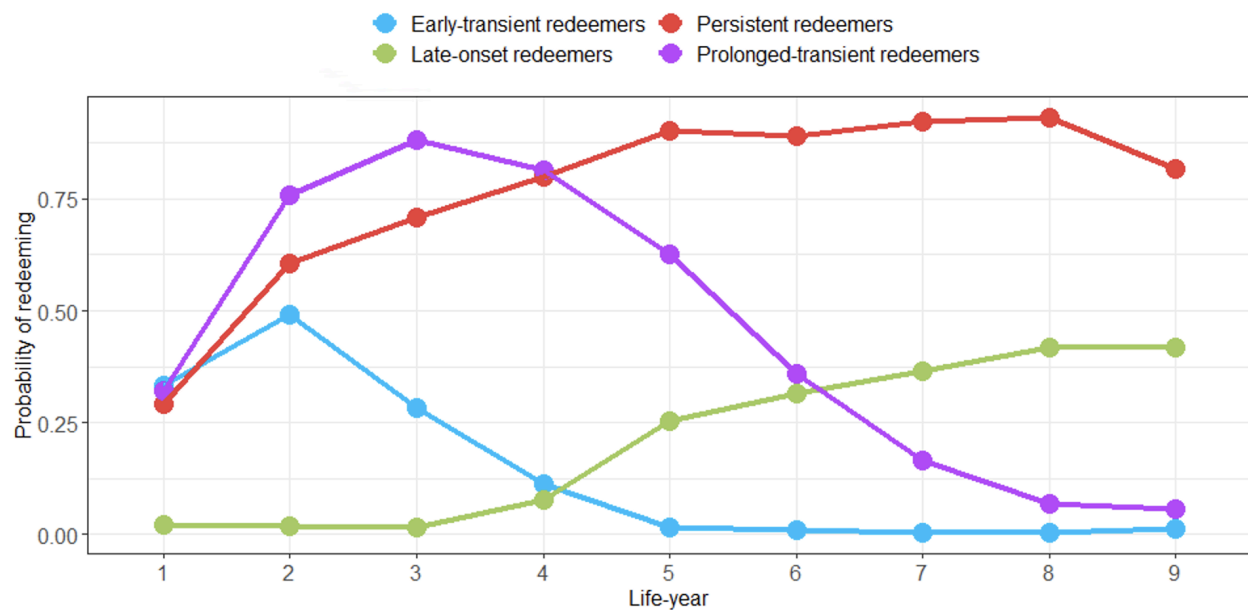


Figure 2: Heatmaps of log-odds ratios between perinatal risk factors and phenotypes. From the right: birth circumstances, socioeconomic, and family history. Cells are transparent if the association did not pass Benjamini-Hochberg correction ($FDR < 0.05$) and color scales are truncated to between 0.5 and -0.5. Stars indicate significant ($FDR < 0.05$) within-class ANOVA to indicate differences among phenotypes among

significantly associated risk factors.

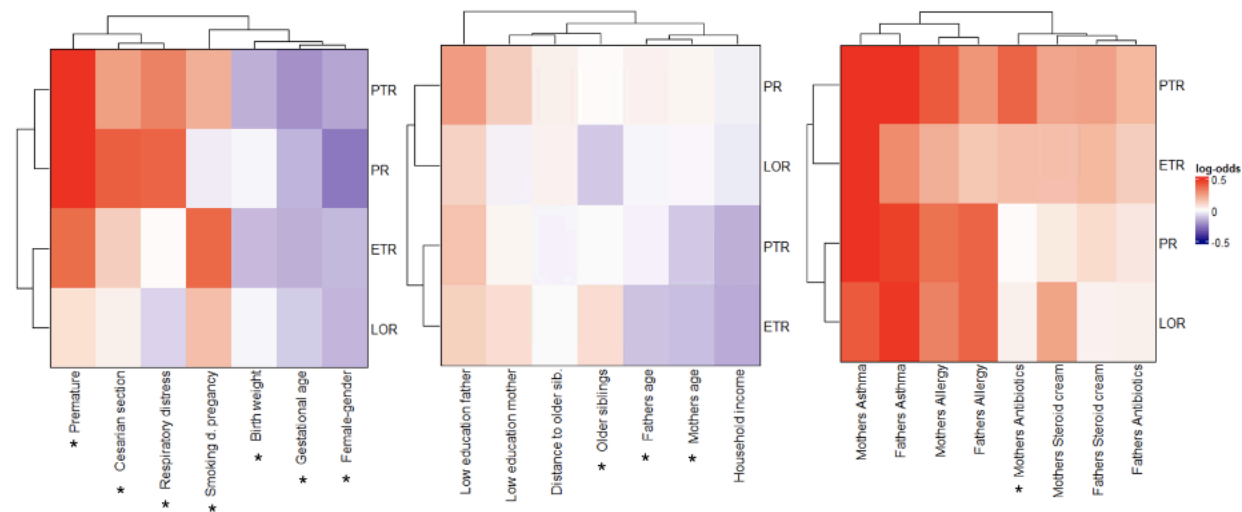
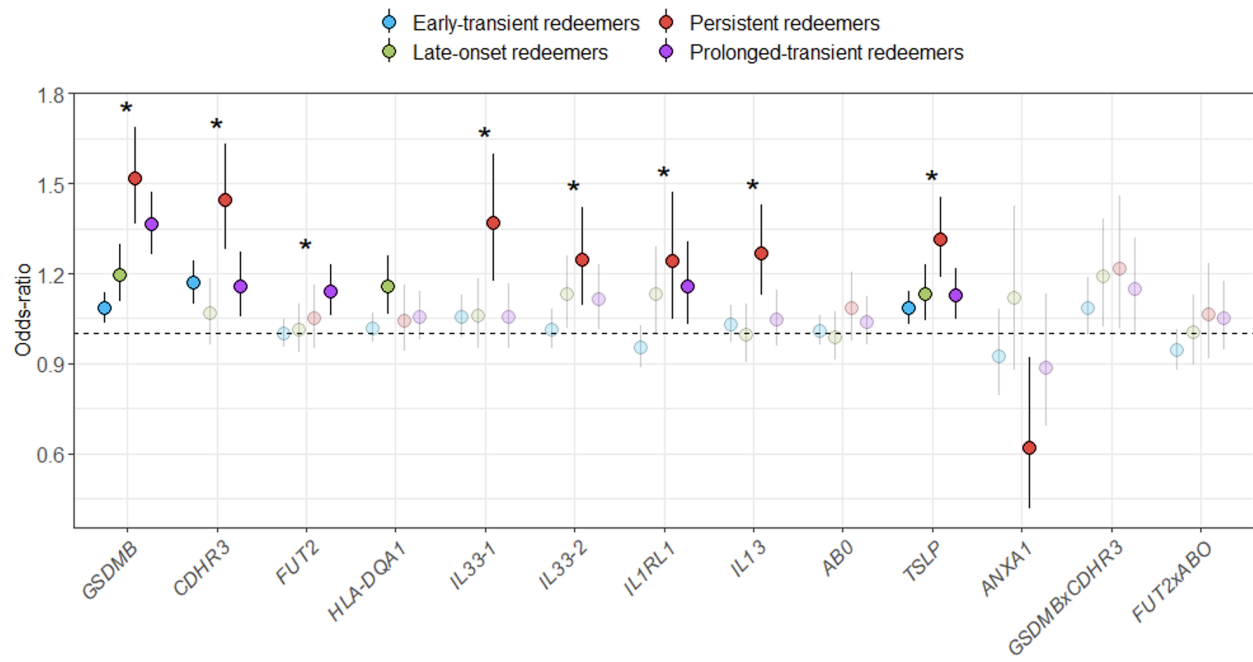


Figure 3: Single-SNP and interactions associated with the phenotypes. Odds-ratio on the y-axis and nearest mapped gene on the x-axis. SNPs are coloured according to the phenotype and SNPs that do not pass Benjamini-Hochberg correction ($FDR < 0.05$) are transparent highlighting more robustly associated signals. Stars indicate significant ($FDR < 0.05$) within-class ANOVA to indicate differences among phenotypes among significantly associated risk factors.



Early Childhood Trajectories of Inhaled Corticosteroid Prescriptions Implicates the Presence of Wheeze Endotypes

Kasper Fischer-Rasmussen, Casper-Emil Tingskov Pedersen, Astrid Sevelsted, Luo Yang, Bo Chawes, Jakob Stokholm, Ann-Marie Malby Schoos, Anders Børglum, Thomas Munk Laursen, Bjarni Vilhjalmsen, Anders Gorm Pedersen, Anders Ulrik Eliassen & Klaus Bønnelykke

Cohort recruitment details, acknowledgements and funding statements

iPSYCH 2012

Recruitment

Individuals diagnosed with schizophrenia, mood disorders, bipolar affective disorder, autism and attention-deficit/hyperactivity disorder were identified through linkage between Danish population-based registers along with a random sample of the same population that supplied the cases.¹⁵ Dried blood spots for virtually all individuals were retrieved from the Danish Neonatal Screening Biobank and processed for genotyping. The design includes the ability to efficiently analyse prospectively collected cohort data within the iPSYCH case-cohort sample.¹⁵ This particular design provides several advantages: As the cohort is randomly selected from the entire population, we are able to generate unbiased absolute risks and incidence rates and to estimate the effect sizes of genetic markers on risk of mental disorders, which is representative of the entire Danish population. To date, most genetic and epidemiological studies are based on convenient case-control samples, which are prone to biases.^{15, 16} The iPSYCH2012 sample was preceded by four smaller Danish samples,^{17, 18, 19, 20, 21, 22, 23, 24} all aiming to investigate the potential interplay between genes and the environment. Collectively, these forerunners informed on the best possible study design to use in the iPSYCH2012 sample ([Supplementary Text 1](#)). The following three paragraphs describe the resources and methods used to identify individuals included in the iPSYCH2012 sample.

Acknowledgements & funding

This study was supported by The Lundbeck Foundation (grant numbers R102-A9118 and R155-2014-1724), Denmark, the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project number 294838) and the Stanley

Center for Psychiatric Research at Broad Institute and Centre for Integrated Register-based Research at Aarhus University. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation. Professor John J McGrath is supported by grant APP1056929 from the John Cade Fellowship from the National Health and Medical Research Council and the Danish National Research Foundation (Niels Bohr Professorship). We thank Betina Trabjerg, National Centre for Register-Based Research, Aarhus University, School of Business and Social Sciences, Aarhus, Denmark, for technical help in producing the principal component plot. We are indebted to the late Mads Vilhelm Hollegaard for his contribution to make sample material accessible for analysis from the Danish Neonatal Screening Biobank. Mads' pioneering work will be used in this and future studies. The iPSYCH2012 had four precursors; all aiming at the potential interplay between genes and the environment and all of smaller scales. All studies included persons born in Denmark May 1981 or later. Collectively, these forerunners informed on the best possible study design to be used in the iPSYCH2012. In short, the very first selected 413 persons with schizophrenia-like psychoses (N=186) and affective disorders (N=256) diagnosed up until October 1999. For each case, the samples in the Danish Neonatal Screening Biobank physically located immediately before and after the sample were selected as controls(1). The second approach (SKIZO2005) utilized persons diagnosed with schizophrenia and bipolar affective disorder up until September 2005 first including 231 genetic markers and later whole genome genotyped(2-5). The third approach (SKIZO2006) utilized all persons diagnosed with schizophrenia up until December 31, 2006 reusing biological material already collected in the previous

study(5-7). Finally, SKIZO2008 utilising new persons diagnosed with schizophrenia up until December 31, 2008 (8). Controls for SKIZO2005, SKIZO2006 and SKIZO2008 were individually time-matched using the following criteria, born in Denmark, had the same birthday as the corresponding case, and were alive and not diagnosed with schizophrenia when the case was first diagnosed. Genome wide genotyping was performed using the Illumina Human 610-quad beadchip. As demonstrated most persons diagnosed with schizophrenia up until 31 December 2008 had already been genotyped prior to the iPSYCH2012 sample, and was therefore not included here. The latter three designs include a total of 1887 cases of schizophrenia and 1863 individually time-matched controls. These individually matched designs are capable of estimating unbiased incidence rate ratios of the exposures/markers of interest(9), but challenging in situations where the matching is broken, e.g., due to missing information in the Danish Neonatal Screening Biobank or failure in genotyping/QC. The increased capability for the iPSYCH2012 sample including 30,000 randomly samples cohort members enabled us the unique possibility to introduce a large case cohort sample superior to previous studies while allowing inclusion of previously genotyped samples. Although, inclusion of previously genotyped cases is methodologically straightforward, inclusion of previously genotyped controls is challenging as they are conditioned to be alive and without schizophrenia at different points in time (when then corresponding case was first diagnosed). The iPSYCH2012 case-cohort sample provides the rare opportunity to analyse genetic data and environmental data respecting the longitudinal nature of human life(10).

iPSYCH 2015

Recruitment

The iPSYCH2015 Case-Cohort expands on the design used in the iPSYCH2012 version¹⁷. Recruitment is done through the use of unique Central Person Register (CPR) numbers. All residents in Denmark are assigned a CPR-number either upon birth or at the time of immigration to Denmark. CPR is linked to the national population-wide registries (e.g. Danish Psychiatric Central Research registry) to recruit participants¹⁷. We cross-referenced CPR with the availability of dried blood spot (DBS) in the Danish Neonatal Screening Biobank (DNSB). The DNSB has been storing residual DBS since May 1981 and encompasses close to all births in Denmark since then. While the DNSB is a screening Biobank with primary activities centred around new-born screening, ancillary and secondary purposes, including research, are allowed provided that relevant authorities have granted approval⁴⁴. Once the initial selection and data generation is completed, the CPR-number allows continuous updates of information from population-based registers through pseudonymization.

Acknowledgements & funding

The Authors would like to thank Vivek Appadurai for the base calculations for the PCA plot. This study was supported by The Lundbeck Foundation, the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project number 294838) and Centre for Integrated Register based Research at Aarhus

University. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation.

This study was supported by The Lundbeck Foundation, the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project number 294838) and Centre for Integrated Register based Research at Aarhus University. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation.

Prescriptions

The iPSYCH cohort represents a comprehensive study that encompasses individuals with psychiatric disorders as well as a significant population-based control group born between 1981 and 2008. For this analysis, we included participants from the iPSYCH cohort who were born between 1998 and 2005 and had complete follow-up data available until the age of 15 in the National Prescription Registry (NPR). This selection aimed to ensure a thorough implementation phase for the NPR, which has been collecting data on redeemed medication prescriptions since its initiation in 1995.

To identify medications used for asthma (inhaled corticosteroids / beta-2 agonists), allergies (nasal steroids / antihistamines), eczema (steroid creams) and respiratory infections (antibiotics) the following Anatomical Therapeutic Chemical (ATC) codes were utilized:

Inhaled corticosteroid: R03AK06, R03AK07, R03AK08, R03AK10, R03AK11, R03BA01, R03BA02, R03BA05, R03BA07, R03BA08.

Beta-2 agonist: R03AC02, R03AC03, R03AC04, R03AC12, R03AC13, R03AC18, R03AL01, R03BB01, R03CC02, R03CC03, R03CC12.

Nasal steroid: R01AC01, R01AC02, R01AC03, R01AD01, R01AD04, R01AD05, R01AD08, R01AD09, R01AD11, R01AD12, R01AD58.

Antihistamines: R06AE07, R06AA02, R06AX27, S01GX, R06A.

Steroid cream: D07AA, D07AB, D07AC, D07AD, D11AH01, D11AH02

Antibiotics: J01CE02, J01CA04, J01CR02.

We also derived parental prescription history using the same ATC codes as above. However, we only consider prescriptions given 10 years prior to the birth of the child included in our analysis.

Wheeze, asthma and infection phenotypes

In the iPSYCH study, severe episodes of wheeze/asthma and lower respiratory tract infections (LRTI) requiring hospitalization were defined using longitudinal data from the Danish National Patient Registry (1). Individuals with at least one hospitalization for a primary diagnosis before age six (age 7 in deCODE) of wheezing disorders—such as

bronchiolitis (ICD-10 code: 913, 390), bronchitis (ICD-8 codes: 466, 490, 491; ICD-10 codes: J20, J40, J41, J42, J44), or asthma (ICD-8 code: 493; ICD-10 codes: J45-46)—were classified as severe wheeze cases. Those without such hospitalizations in that age gap were used as controls.

Psychiatric status

We defined psychiatric status based on data from psychiatric in-patient facilities, including diagnoses of schizophrenia (ICD-10 codes: F20), bipolar disorder (ICD-10 codes: F30-F31), affective disorder (ICD-10 codes: F30-F39), autism (ICD-10 codes: F84.0, F84.1, F84.5, F84.8 or F84.9) and attention-deficit/hyperactivity disorder (ADHD) (ICD-10 codes: F90.0). Psychiatric status was defined as having any of these diagnoses at any time during the lifetime of the participants. In the analysis of psychiatric status as a risk factor, controls were those without any psychiatric diagnoses in the study period.

Other registries

Familial factors such as parental age at birth (both mother and father), number of siblings was derived from Danish Civil Registration system (CPR registry).

Household income 5 years prior to birth was derived from the (IND) The Income Statistics Register (2) by summarizing stock, collective and interest income.

The maximum attained education in the study period (1998-2017) was derived from the (UDDA) Population Education Register (3).

Evidence for a 6-class model

LMR also suggested the presence of inflection points at 4 and 6 classes. The BLRT test suggested no further improvement in model likelihood after 4-classes. Neither entropy nor average silhouette width (ASW) did not display a strong trend, but ASW suggested a 7-class model to display the best separation among classes (see Supplementary Table 1). Since the 4-class model constituted the most parsimonious solution and the BLRT have shown to outperform the LMR-test (4) we performed all subsequent analyses on the results of the 4-class model.

Within Class ANOVA

To assess whether a given risk factor differs significantly among the four phenotypes, we excluded individuals who never redeemed an ICS prescription. We subsequently fitted two models: one including the risk factor and the other without it, keeping the same covariates in both models. We then compared the models using the likelihood ratio test (LRT). If including the risk factor in the model leads to significant improvement in likelihood, this suggests that the risk factor explains substantial variation among the phenotypes.

We fit the following models to the data excluding the baseline class, to assess if the risk factor explains substantial variation among the phenotypes.

$$M_r: p_i = \beta_0 + \beta_1 r_i + \beta_{cov}^T cov_i + \epsilon_i$$

$$M_{null}: p_i = \beta_0 + \beta_{cov}^T cov_i + \epsilon_i$$

Model M_r represents the model including the risk factor (perinatal, SNP or PRS) and M_{null} represents the model including only covariates. The risk factor is denoted r and cov , represents the included covariates gender, cohort-status (iPSYCH12 or iPSYCH15), birth year and psychiatric status in all models and gestational age, parental birth year and top 10 principal components in specified models. The residual error is represented by ϵ_i . The ratio of likelihood of the two models M_r and M_{null} quantifies how much better the risk-factor model fits to the data.

$$\lambda = \frac{L(M_r)}{L(M_{null})}$$

A significance level p-value is calculated by assuming that the likelihood-ratio test statistic, $-2 \log \lambda$, follows a chi-square distribution with k degrees of freedom.

$$-2 \log \lambda = -2(\log(L(M_r)) - \log(L(M_{null})))$$

$$-2 \log \lambda \sim \chi^2(k)$$

We interpret this as the indication of significantly distinct associations between the phenotypes given a significant result in the risk factor analysis. All statistical analyses were corrected for multiple comparisons using the Benjamini-Hochberg procedure to control the false-discovery rate to 5 %.

Supplementary tables

Supplementary Table 1: This table shows the results of model selection based on the number of classes. BIC, entropy, average silhouette width (ASW), p-value likelihood ratio from the Lo-Mendell-Rubins test and boot-strapped likelihood ratio (BLRT) are shown. P-values for Lo-Mendell-Rubin's test and BLRT are based on the k vs. $k-1$ classes, indicating if k number of classes results in substantial model improvement compared to $k-1$. For example, in the second line, we compare if a 2-class model is better than a 1-class model. ASW, L-Mendell-Rubin and BLRT are undefined for the 1-class model.

K	BIC	Entropy	ASW	Lo-Mendell-Rubin	BLRT
1	77261	5.1	NA	NA	NA
2	69226	4.6	0.123	0.00e+00	0.00e+00
3	66657	4.5	0.138	0.00e+00	0.00e+00
4	65245	4.3	0.120	4.16e-305	0.00e+00
5	64823	4.3	0.108	2.13e-99	6.40e-02
6	64440	4.3	0.142	1.64e-91	8.10e-02
7	64355	4.2	0.148	1.13e-31	3.43e-01
8	64281	4.2	0.144	1.91e-28	4.24e-01
9	64209	4.2	0.129	1.05e-27	4.09e-01
10	64156	4.2	0.123	1.40e-24	4.52e-01

Supplementary Table 2: Prevalences of comorbidities medications in 3-year bins from birth to 9 years for each phenotype

Medication	Non-redeemers	Early-transient redeemers	Prolonged-transient redeemers	Persistent redeemers	Late-onset redeemers
Beta-2 agonists (0-3)	34.6%	84.4%	95.7%	91.4%	61%
Beta-2 agonists (3-6)	10%	29.8%	81.9%	95.4%	55.5%
Beta-2 agonists (6-9)	3.5%	11.1%	34.7%	90.4%	62.7%
Nasal steroid (0-3)	0.6%	1.4%	2.4%	3.6%	1%
Nasal steroid (3-6)	1.5%	3.5%	4.7%	12.7%	6.5%
Nasal steroid (6-9)	3%	5%	5.9%	19.2%	17.3%
Antihistamines (0-3)	6.6%	11.8%	17%	20.9%	13.7%
Antihistamines (3-6)	7.8%	12.8%	18%	29.3%	24%
Antihistamines (6-9)	7.4%	12.7%	13.4%	32.9%	26.5%
Steroid cream (0-3)	19%	27.7%	34%	41.6%	37.1%
Steroid cream (3-6)	11.9%	17.3%	22.6%	29.3%	24.3%
Steroid cream (6-9)	10.4%	15.6%	18.4%	22.9%	21%
Antibiotics (0-3)	61.7%	89.3%	92.7%	91.9%	84%
Antibiotics (3-6)	48.1%	67.2%	75.7%	78.4%	72.4%
Antibiotics (6-9)	33.5%	47.9%	52.9%	58%	57.5%

Supplementary Table 3: Results of multinomial regressions of comorbidities

medications in 3-year bins from birth to 9 years for each phenotype

Medication	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers		
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value
Beta-2 agonists (0-3)	9.96	(9.07-10.93)	0.00e+00	45.12	(35.03-58.12)	0.00e+00	19.01	(14.69-24.6)	0.00e+00	2.44	(2.17-2.74)	0.00e+00
Beta-2 agonists (3-6)	3.46	(3.2-3.75)	0.00e+00	39.55	(34.45-45.41)	0.00e+00	176.57	(125.9-247.63)	0.00e+00	10.08	(8.97-11.33)	0.00e+00
Beta-2 agonists (6-9)	3.17	(2.82-3.57)	0.00e+00	13.71	(12.11-15.52)	0.00e+00	243.07	(189.86-311.21)	0.00e+00	42.61	(37.43-48.5)	0.00e+00
Nasal steroid (0-3)	2.22	(1.64-3.02)	2.91e-07	3.65	(2.53-5.26)	3.73e-12	5.55	(3.69-8.36)	2.22e-16	1.54	(0.87-2.72)	1.36e-01
Nasal steroid (3-6)	2.28	(1.87-2.77)	2.22e-16	3.04	(2.35-3.92)	0.00e+00	8.78	(6.95-11.09)	0.00e+00	4.25	(3.33-5.42)	0.00e+00
Nasal steroid (6-9)	1.58	(1.35-1.86)	1.53e-08	1.88	(1.5-2.35)	3.61e-08	6.96	(5.75-8.44)	0.00e+00	6.20	(5.28-7.28)	0.00e+00
Antihistamines (0-3)	1.73	(1.55-1.93)	0.00e+00	2.66	(2.3-3.06)	0.00e+00	3.36	(2.81-4.02)	0.00e+00	2.04	(1.73-2.41)	0.00e+00
Antihistamines (3-6)	1.57	(1.42-1.74)	0.00e+00	2.35	(2.05-2.7)	0.00e+00	4.39	(3.74-5.16)	0.00e+00	3.34	(2.92-3.83)	0.00e+00
Antihistamines (6-9)	1.61	(1.45-1.79)	0.00e+00	1.73	(1.48-2.02)	4.21e-12	5.37	(4.59-6.28)	0.00e+00	3.96	(3.47-4.51)	0.00e+00
Steroid cream (0-3)	1.48	(1.37-1.59)	0.00e+00	2.00	(1.79-2.23)	0.00e+00	2.72	(2.35-3.15)	0.00e+00	2.24	(1.99-2.52)	0.00e+00
Steroid cream (3-6)	1.41	(1.29-1.55)	1.21e-13	1.98	(1.74-2.25)	0.00e+00	2.77	(2.37-3.25)	0.00e+00	2.13	(1.86-2.43)	0.00e+00
Steroid cream (6-9)	1.46	(1.32-1.6)	9.55e-15	1.79	(1.56-2.05)	0.00e+00	2.32	(1.95-2.75)	0.00e+00	2.06	(1.79-2.37)	0.00e+00
Antibiotics (0-3)	4.61	(4.13-5.14)	0.00e+00	7.41	(6.06-9.08)	0.00e+00	5.93	(4.54-7.75)	0.00e+00	2.53	(2.16-2.95)	0.00e+00
Antibiotics (3-6)	1.90	(1.76-2.04)	0.00e+00	3.02	(2.67-3.42)	0.00e+00	3.33	(2.79-3.98)	0.00e+00	2.32	(2.04-2.63)	0.00e+00
Antibiotics (6-9)	1.61	(1.5-1.72)	0.00e+00	2.01	(1.81-2.24)	0.00e+00	2.40	(2.07-2.78)	0.00e+00	2.30	(2.05-2.58)	0.00e+00

Supplementary Table 4: Prevalences of related diagnoses in 3-year bins from birth to 6 years for each phenotype.

Medication	Non-redeemers	Early-transient redeemers	Prolonged-transient redeemers	Persistent redeemers	Late-onset redeemers
Asthma (0-3)	0.5%	7.6%	15.5%	17.5%	1.8%
Asthma (3-6)	0.1%	0.5%	5.4%	14.7%	4.2%
Wheeze (0-3)	1.3%	11.9%	22.3%	21.9%	3.7%
Wheeze (3-6)	0.1%	0.7%	6.3%	15.7%	5%
LRTI (0-3)	1.3%	3.8%	7.7%	7.8%	2.1%
LRTI (3-6)	0.3%	0.7%	2.1%	2.8%	1.4%

Supplementary Table 5: Results of multinomial regressions of comorbidities diagnoses in 3-year bins from birth to 9 years for each phenotype.

Diagnosis	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers		
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value
Asthma (0-3)	13.48	(11.01-16.5)	0.00e+00	29.97	(24.06-37.33)	0.00e+00	37.26	(28.75-48.29)	0.00e+00	2.97	(1.92-4.6)	1.14e-06
Asthma (3-6)	4.98	(2.79-8.87)	5.34e-08	50.72	(32.38-79.45)	0.00e+00	176.70	(113.46-275.17)	0.00e+00	43.24	(26.87-69.58)	0.00e+00
Wheeze (0-3)	9.29	(8.02-10.75)	0.00e+00	20.70	(17.44-24.56)	0.00e+00	21.04	(16.9-26.2)	0.00e+00	2.47	(1.81-3.37)	1.29e-08
Wheeze (3-6)	4.72	(2.88-7.74)	7.65e-10	42.84	(29.07-63.14)	0.00e+00	138.02	(93.7-203.29)	0.00e+00	36.95	(24.43-55.9)	0.00e+00
LRTI (0-3)	2.47	(2.03-3.01)	0.00e+00	5.04	(4.04-6.29)	0.00e+00	5.23	(3.91-7)	0.00e+00	1.33	(0.89-1.98)	1.67e-01
LRTI (3-6)	2.17	(1.38-3.4)	7.65e-04	6.63	(4.35-10.11)	0.00e+00	9.41	(5.77-15.35)	0.00e+00	4.71	(2.79-7.94)	6.79e-09

Supplementary Table 6: Psychiatric traits as exposures for ICS phenotypes. Analyses are adjusted for gender, cohort-status, birth year, maternal smoking during pregnancy, gestational age, number of older siblings, paternal low education and household income.

Diagnosis	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			Cases/controls
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
Schizophrenia	0.52	(0.16-1.69)	2.75e-01	0.44	(0.06-3.2)	4.14e-01	1.68	(0.4-7.08)	4.83e-01	2.54	(1.06-6.1)	3.67e-02	171/36165
Schizophrenia spectrum	1.32	(0.86-2.02)	2.11e-01	1.27	(0.66-2.45)	4.73e-01	1.24	(0.5-3.07)	6.46e-01	2.51	(1.52-4.12)	2.94e-04	854/35482
Bipolar disorder	0.64	(0.08-5.05)	6.75e-01	0.00	(0-0)	0.00e+00	0.00	(0-0)	0.00e+00	3.14	(0.68-14.5)	1.42e-01	97/36239
Affective disorders	1.53	(1.19-1.97)	9.48e-04	1.04	(0.66-1.62)	8.73e-01	1.47	(0.87-2.48)	1.48e-01	1.74	(1.21-2.5)	2.75e-03	2437/33899
Autism	0.93	(0.84-1.03)	1.86e-01	1.05	(0.9-1.22)	5.17e-01	1.15	(0.94-1.42)	1.81e-01	1.04	(0.88-1.23)	6.68e-01	10552/25784
ADHD	1.21	(1.1-1.34)	1.82e-04	1.48	(1.28-1.71)	1.89e-07	1.14	(0.92-1.41)	2.20e-01	1.18	(1-1.39)	5.39e-02	10486/25850
Anorexia	1.48	(0.96-2.29)	7.38e-02	1.00	(0.46-2.17)	9.97e-01	1.16	(0.42-3.2)	7.72e-01	0.86	(0.38-1.98)	7.24e-01	1056/35280
Psychiatric status	1.16	(1.04-1.28)	5.12e-03	1.44	(1.23-1.68)	3.84e-06	1.27	(1.03-1.58)	2.75e-02	1.19	(1-1.4)	4.44e-02	21287/15049

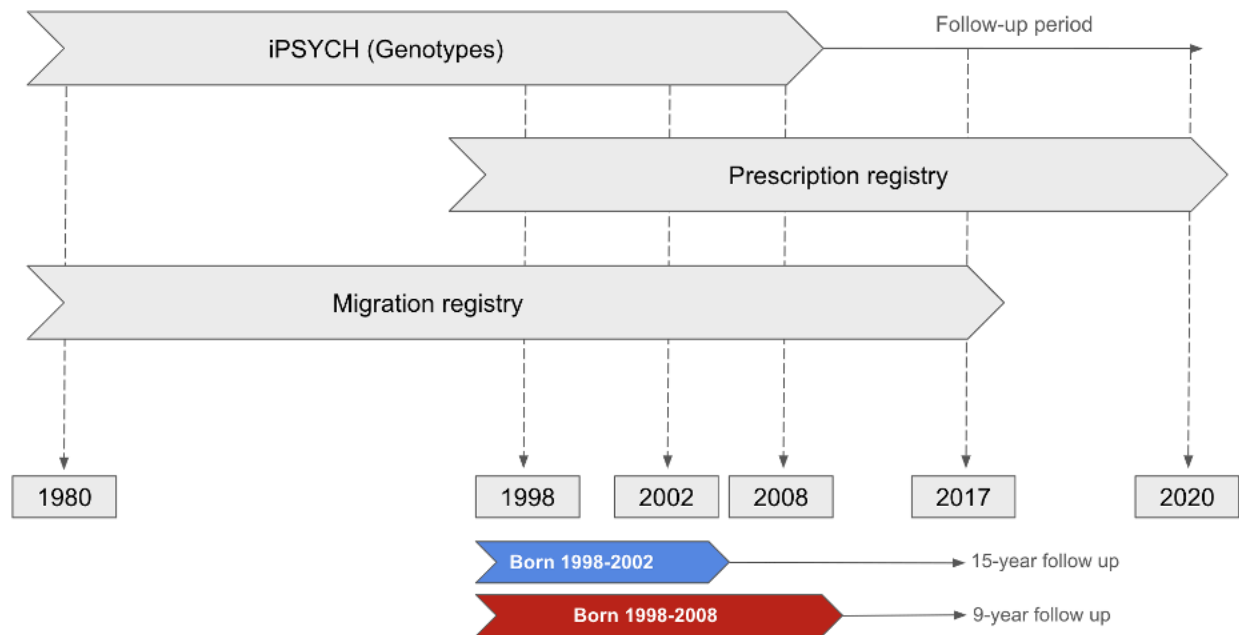
Supplementary Table 7: ICS phenotypes as exposures for psychiatric traits. Analyses are adjusted for gender, cohort-status, birth year, maternal smoking during pregnancy,

gestational age, number of older siblings, paternal low education and household income.

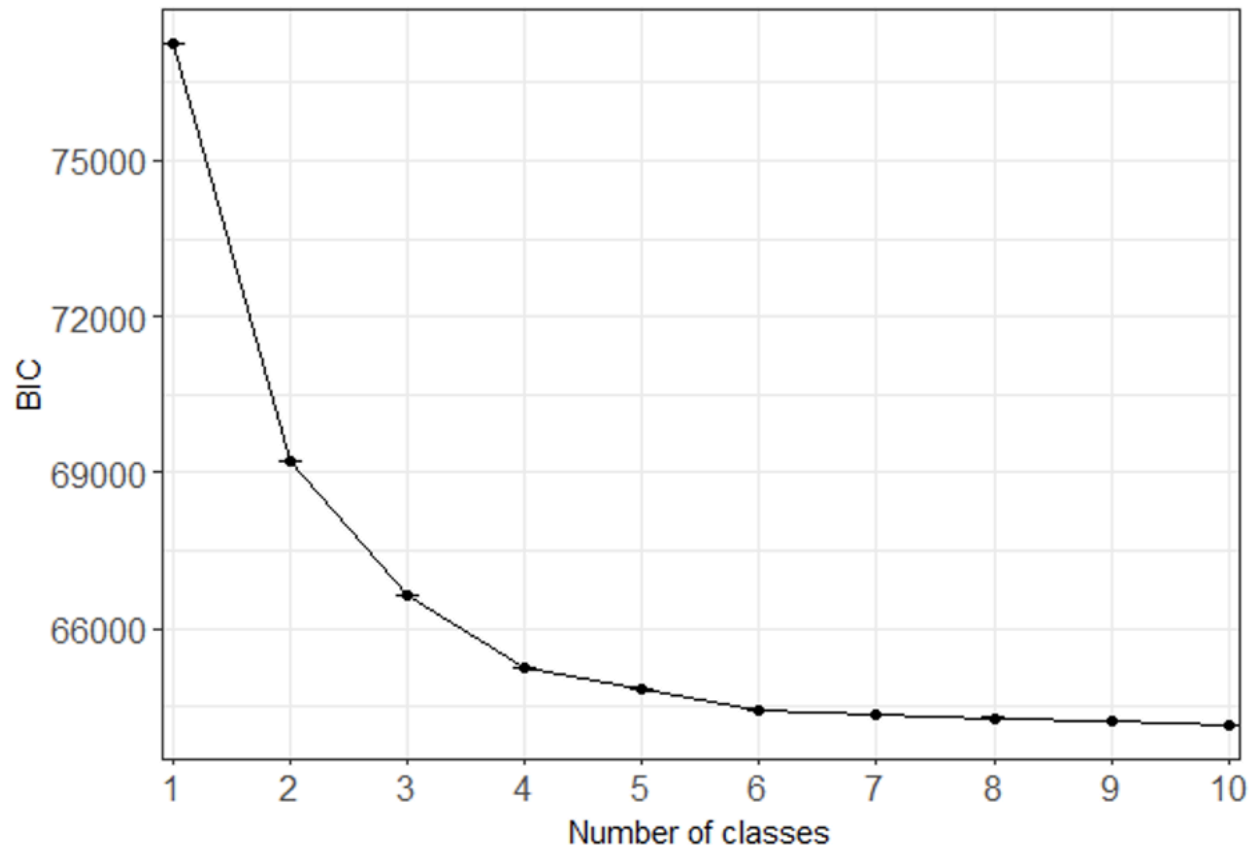
Diagnosis	Early-transient redeemers			Prolonged-transient redeemers			Persistent redeemers			Late-onset redeemers			Cases/controls
	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	OR	CI 95%	P-value	
Schizophrenia	0.46	(0.14-1.52)	2.06e-01	0.48	(0.06-3.5)	4.66e-01	1.85	(0.43-7.94)	4.05e-01	2.32	(0.96-5.64)	6.31e-02	171/36165
Schizophrenia spectrum	1.24	(0.81-1.92)	3.21e-01	1.30	(0.67-2.5)	4.36e-01	1.25	(0.51-3.11)	6.25e-01	2.39	(1.45-3.94)	6.39e-04	854/35482
Bipolar disorder	0.59	(0.07-4.67)	6.18e-01	0.00	(0-Inf)	9.92e-01	0.00	(0-Inf)	9.92e-01	2.90	(0.62-13.49)	1.75e-01	97/36239
Affective disorders	1.44	(1.11-1.86)	5.63e-03	1.05	(0.67-1.65)	8.31e-01	1.50	(0.89-2.54)	1.30e-01	1.65	(1.15-2.38)	7.23e-03	2437/33899
Autism	0.93	(0.84-1.03)	1.87e-01	1.05	(0.9-1.22)	5.31e-01	1.15	(0.94-1.43)	1.79e-01	1.04	(0.88-1.23)	6.72e-01	10552/25784
ADHD	1.22	(1.1-1.35)	1.58e-04	1.49	(1.29-1.73)	1.04e-07	1.16	(0.94-1.43)	1.80e-01	1.18	(1-1.4)	4.80e-02	10486/25850
Anorexia	1.37	(0.89-2.12)	1.54e-01	1.11	(0.51-2.42)	7.95e-01	1.10	(0.4-3.08)	8.51e-01	0.85	(0.37-1.97)	7.10e-01	1056/35280
Psychiatric status	1.17	(1.06-1.29)	2.66e-03	1.44	(1.24-1.68)	2.90e-06	1.27	(1.02-1.57)	3.15e-02	1.20	(1.01-1.41)	3.58e-02	21287/15049

Supplementary figures

Supplementary Figure 1: Overview of study population taking a starting point in the two iPSYCH cohorts. The prescription registry is used to define ICS trajectories, but we only include those where we have data on migration from the migration registry. The red and blue boxes represent two avenues, the maximum length of follow-up or 9-year follow up, but allowing for larger sample size.

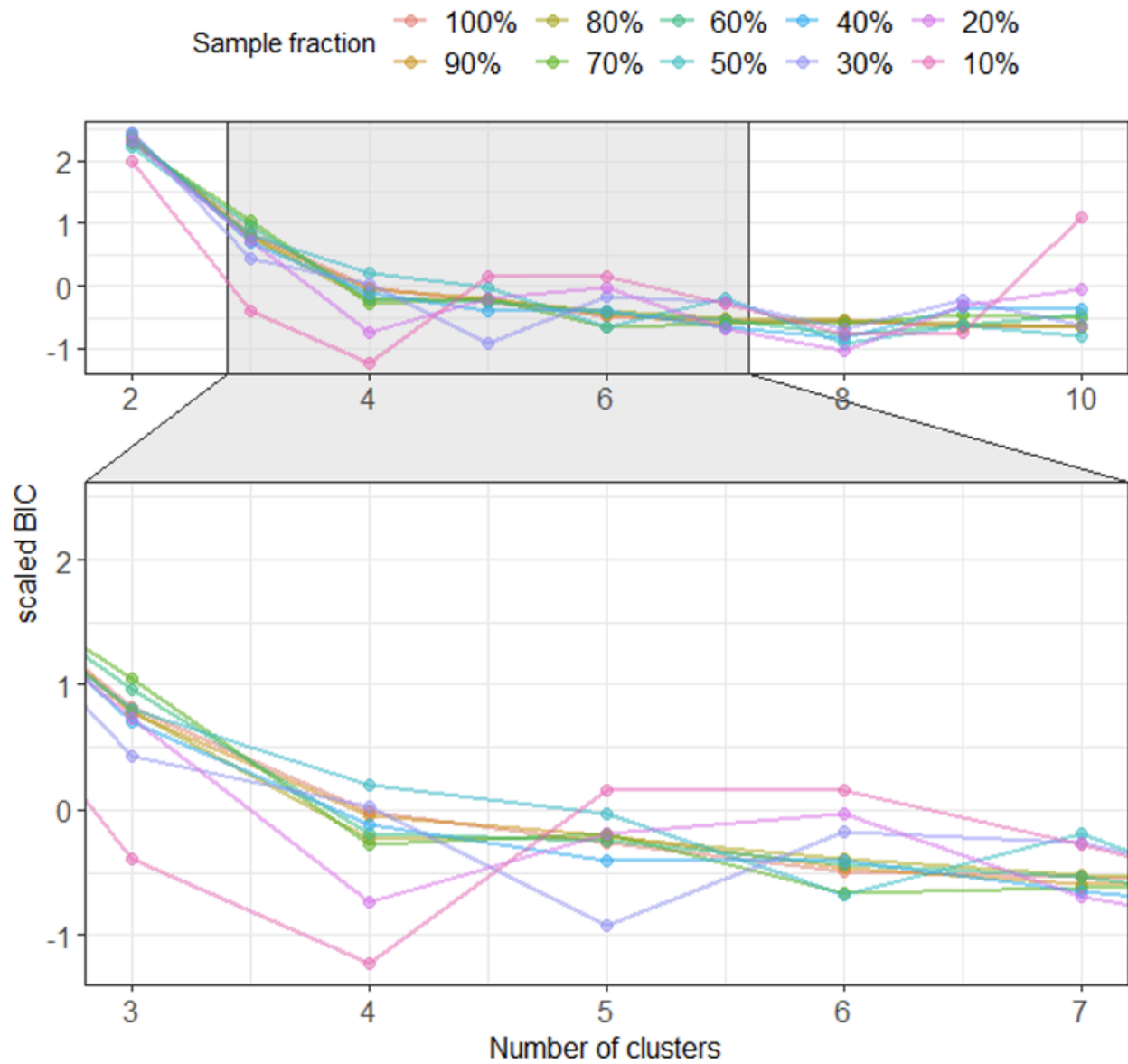


Supplementary Figure 2: Elbow-plot of BIC vs. number of classes. Since no global minimum was identified within 10-classes, we utilized an inflection point to inform about the best possible solution measured by model fit (BIC). Two potential inflection points were identified ($k=4,6$), however selecting the most parsimonious model resulted in a 4-class model being prioritized.

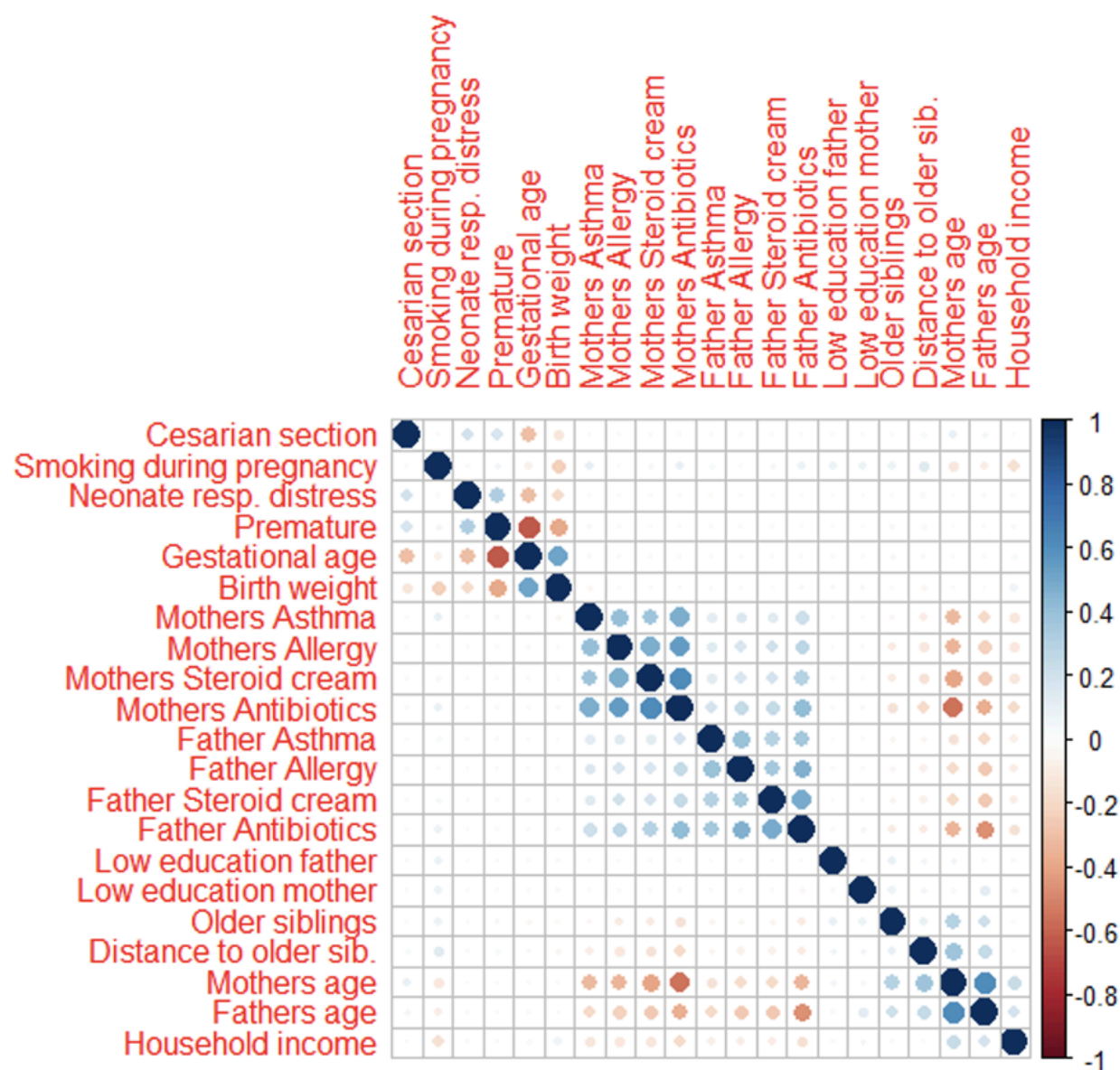


Supplementary Figure 3: Elbow-plot of BIC as a result of downsampling from 100 % sample size until 10 % sample size. The number of classes are shown on the x-axis and the scaled BIC (mean = 0, sd = 1) is shown on the y-axis to allow comparable BIC values among sample fractions. Each coloured curve represents an identical model selection procedure as the initial step. Curves are coloured according to the downsampling proportions as shown in the legend. Stability of the inflection point is

evaluated by zooming in on the 3-7 class models to allow closer visual inspection of the inflection point.

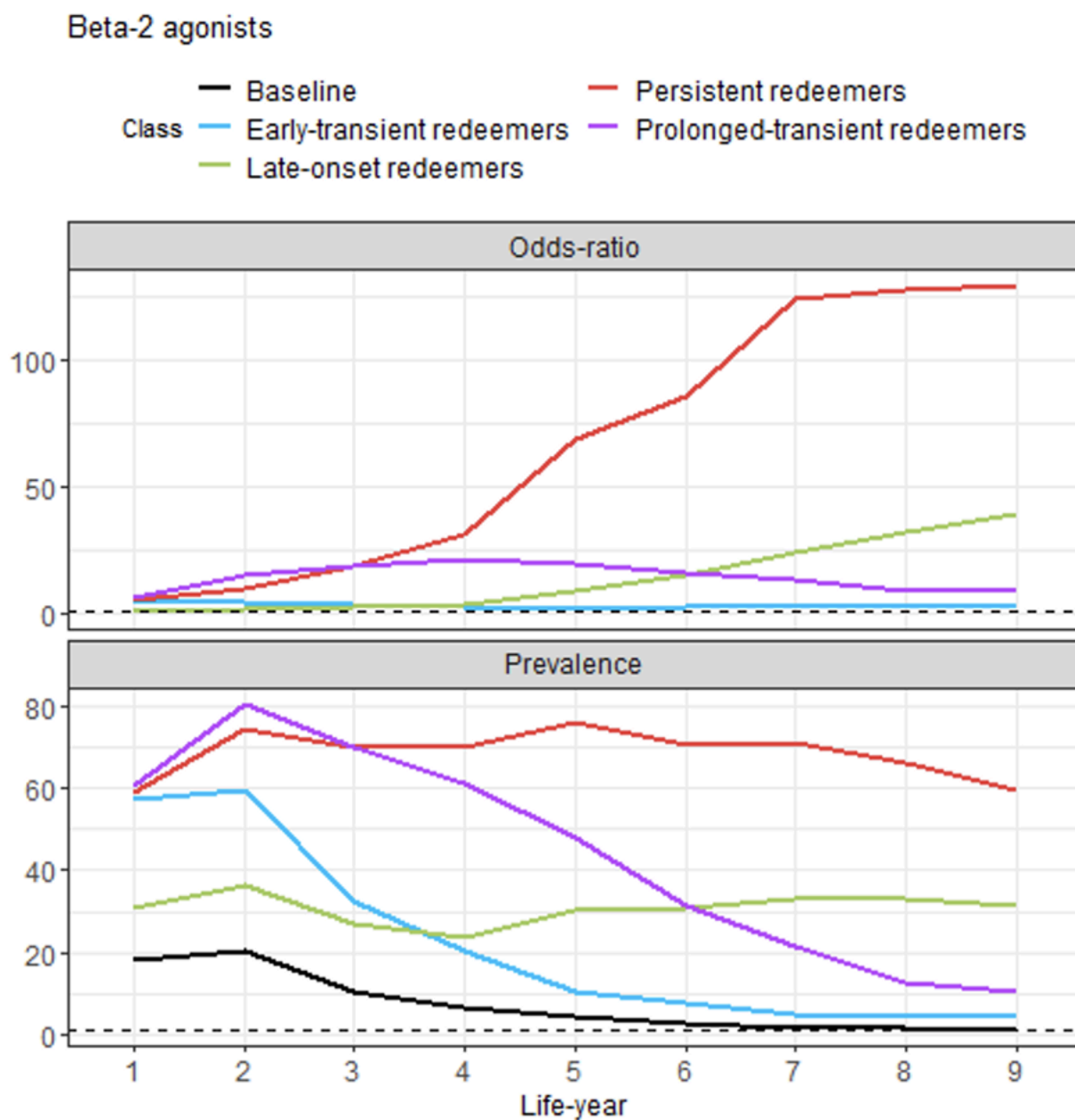


Supplementary Figure 4: Correlations between perinatal risk factors. All risk factors included in the analysis



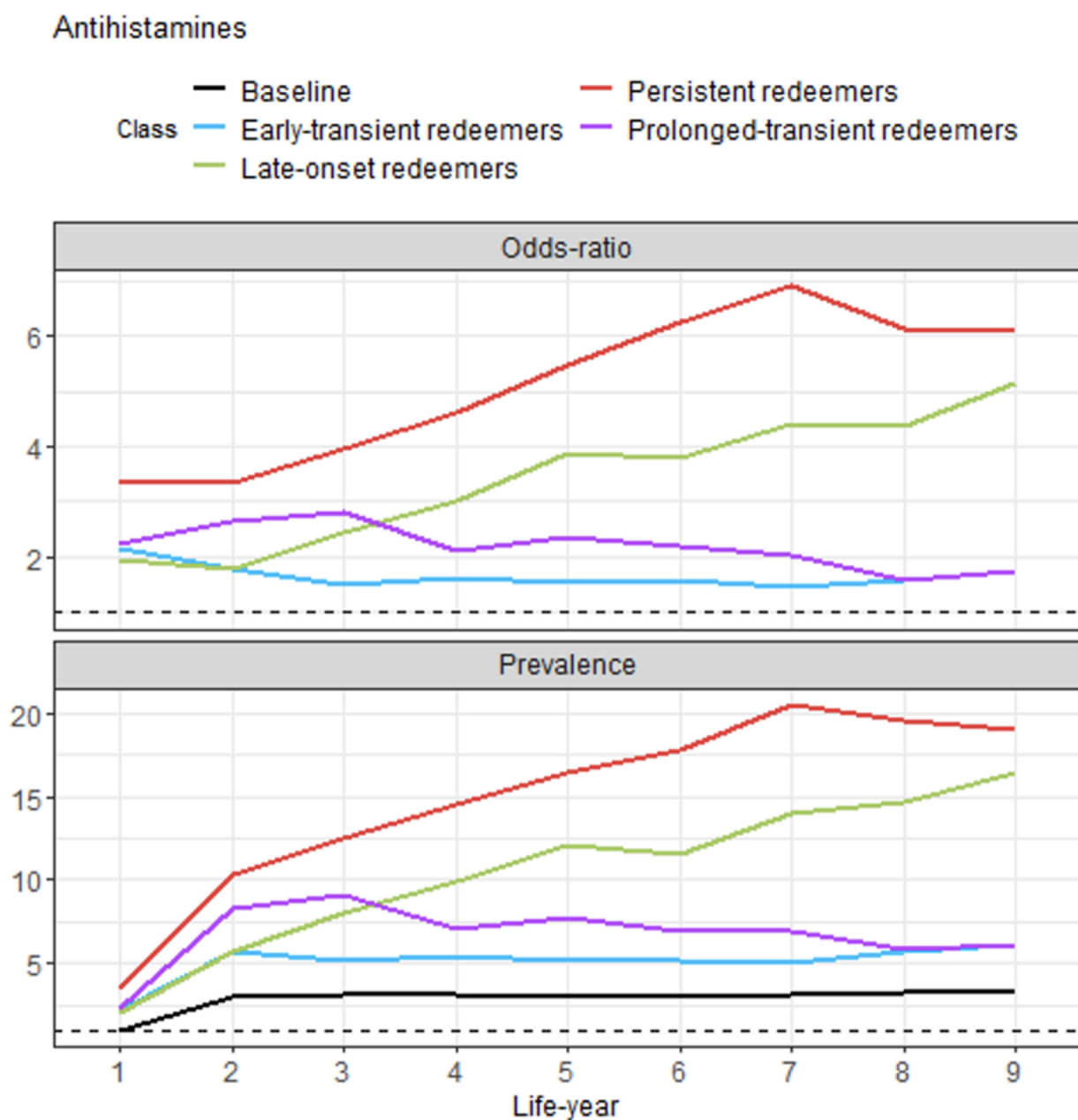
Supplementary Figure 5: Longitudinal prescription comorbidities for beta-2 agonists.

Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalence, including baseline.



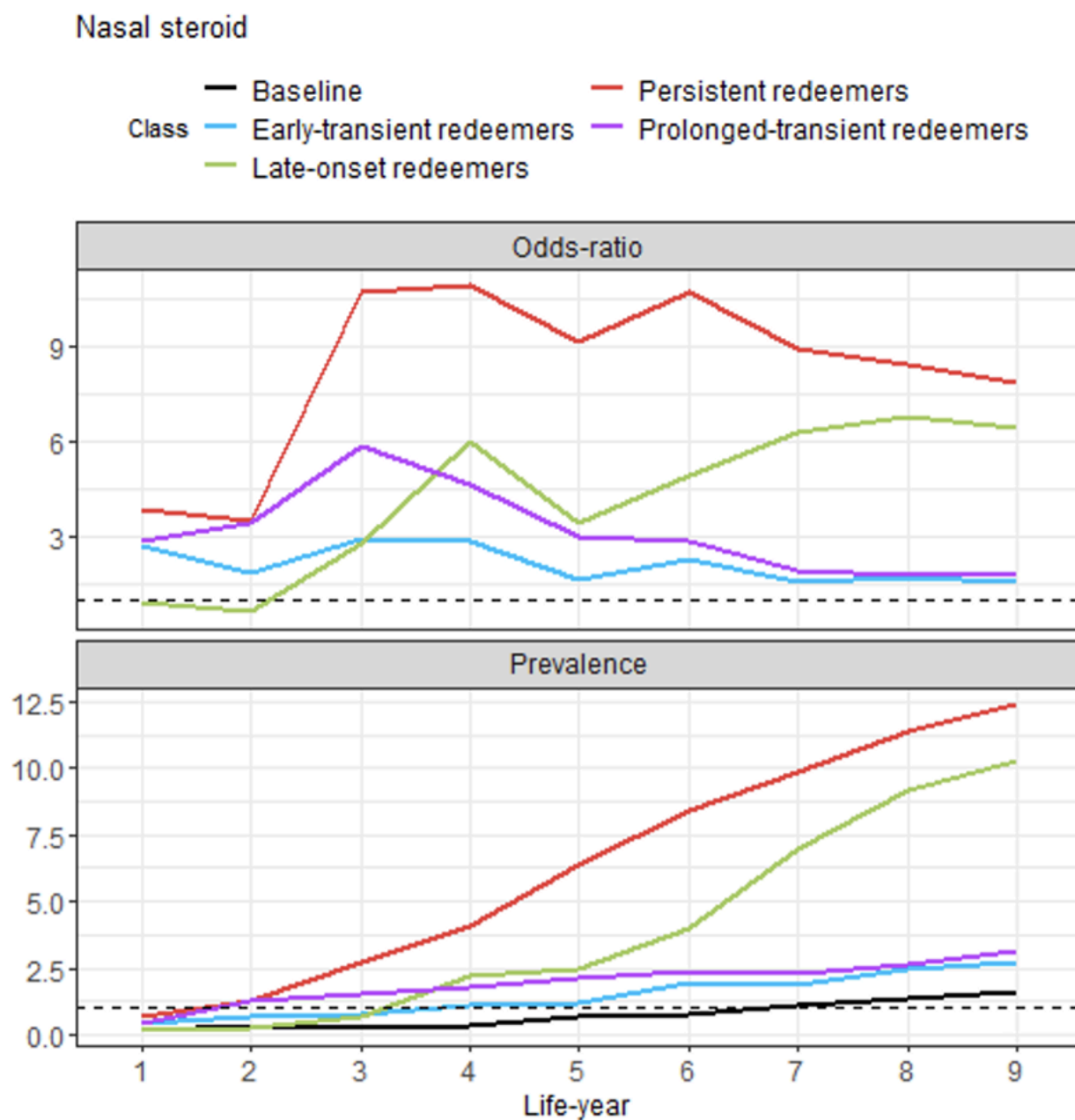
Supplementary Figure 6: Longitudinal prescription comorbidities for antihistamines.

Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalence, including baseline.

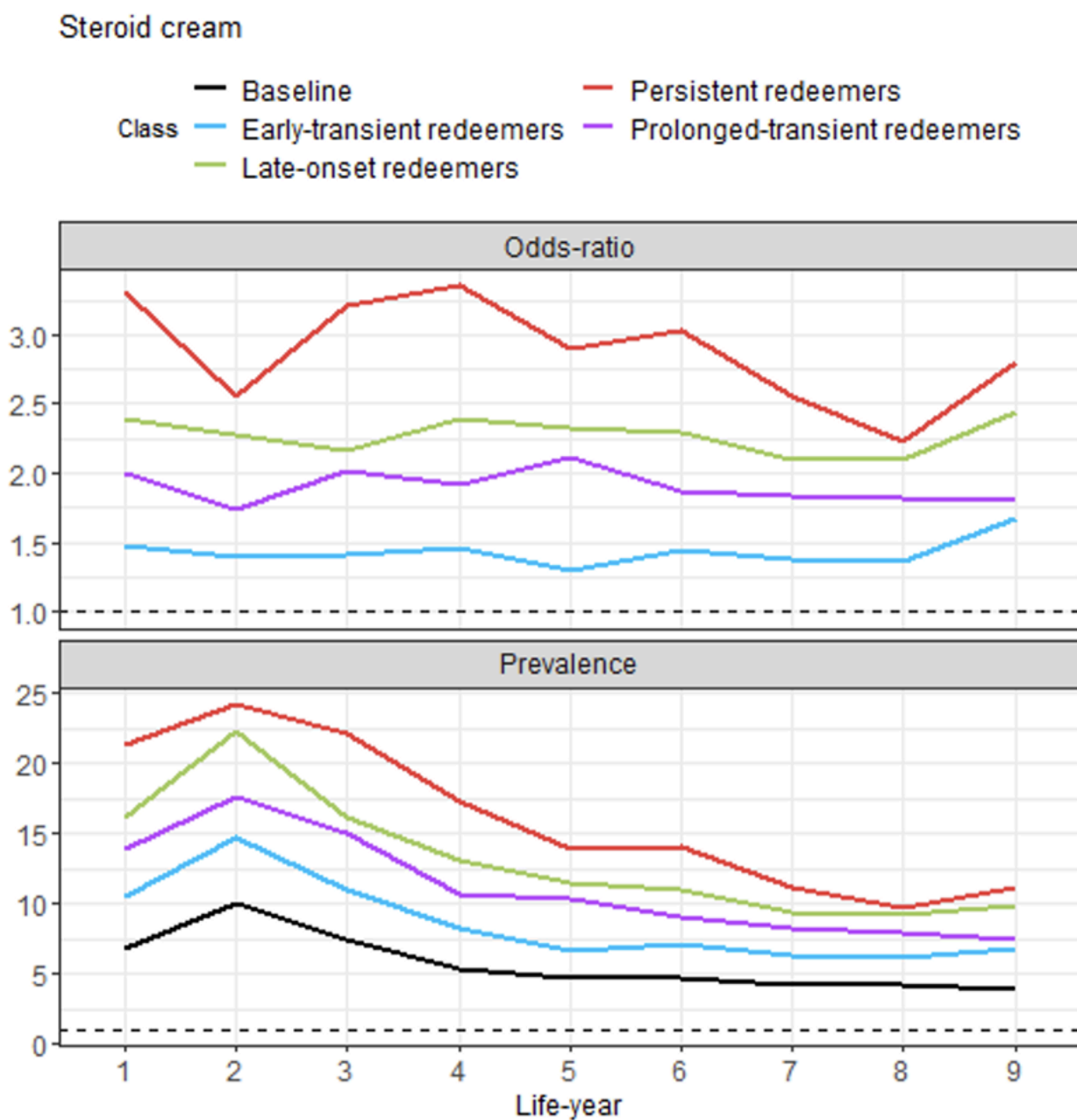


Supplementary Figure 7: Longitudinal prescription comorbidities for nasal steroids.

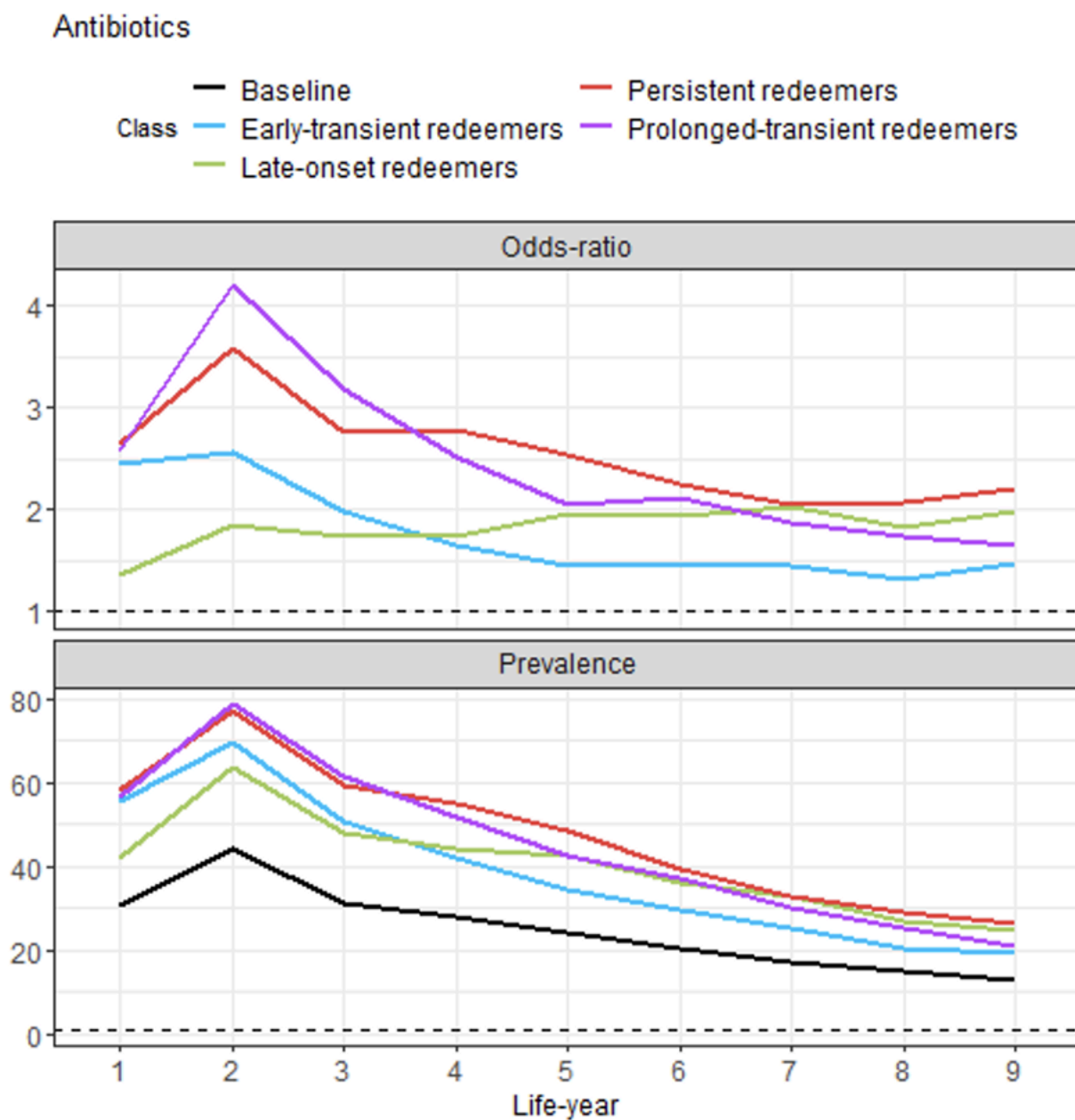
Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalence, including baseline.



Supplementary Figure 8: Longitudinal prescription comorbidities for steroid cream. Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalences, including baseline.

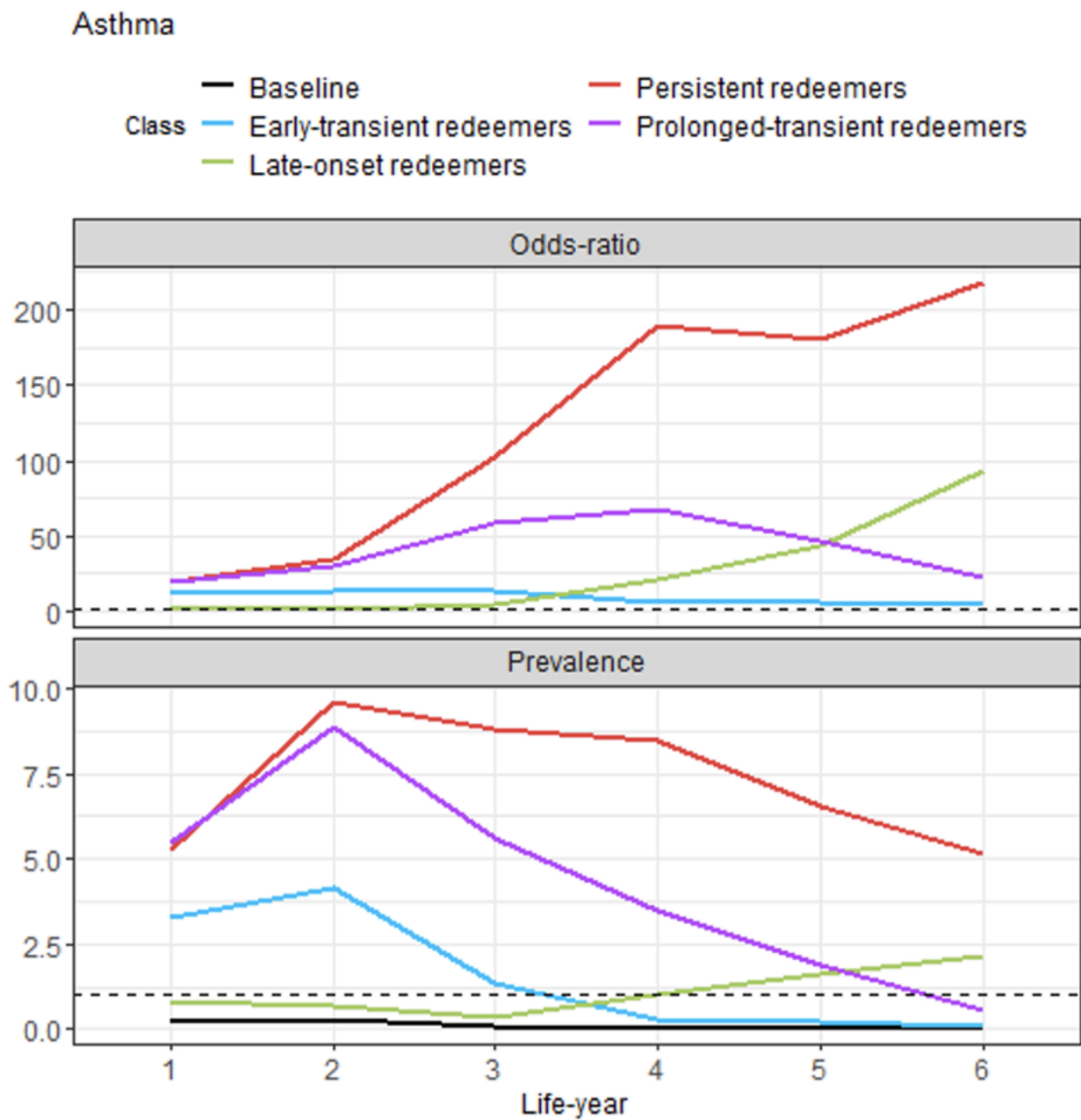


Supplementary Figure 9: Longitudinal prescription comorbidities for antibiotics. Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalences, including baseline.



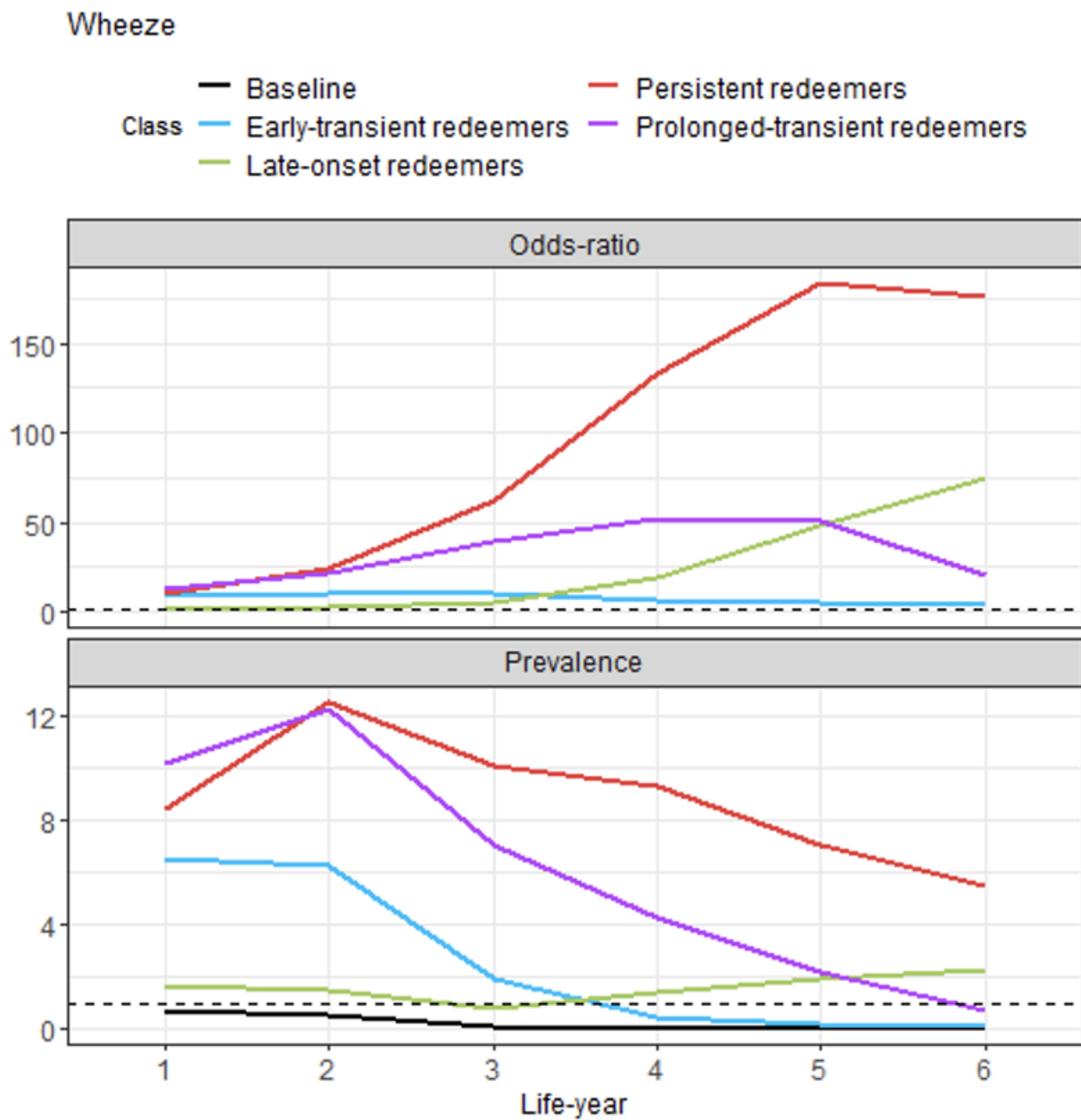
Supplementary Figure 10: Longitudinal asthma diagnoses. Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalences,

including baseline.



Supplementary Figure 11: Longitudinal wheeze diagnoses. Top plot shows odds ratios for each phenotype compared to baseline, while the bottom plot shows prevalences,

including baseline.

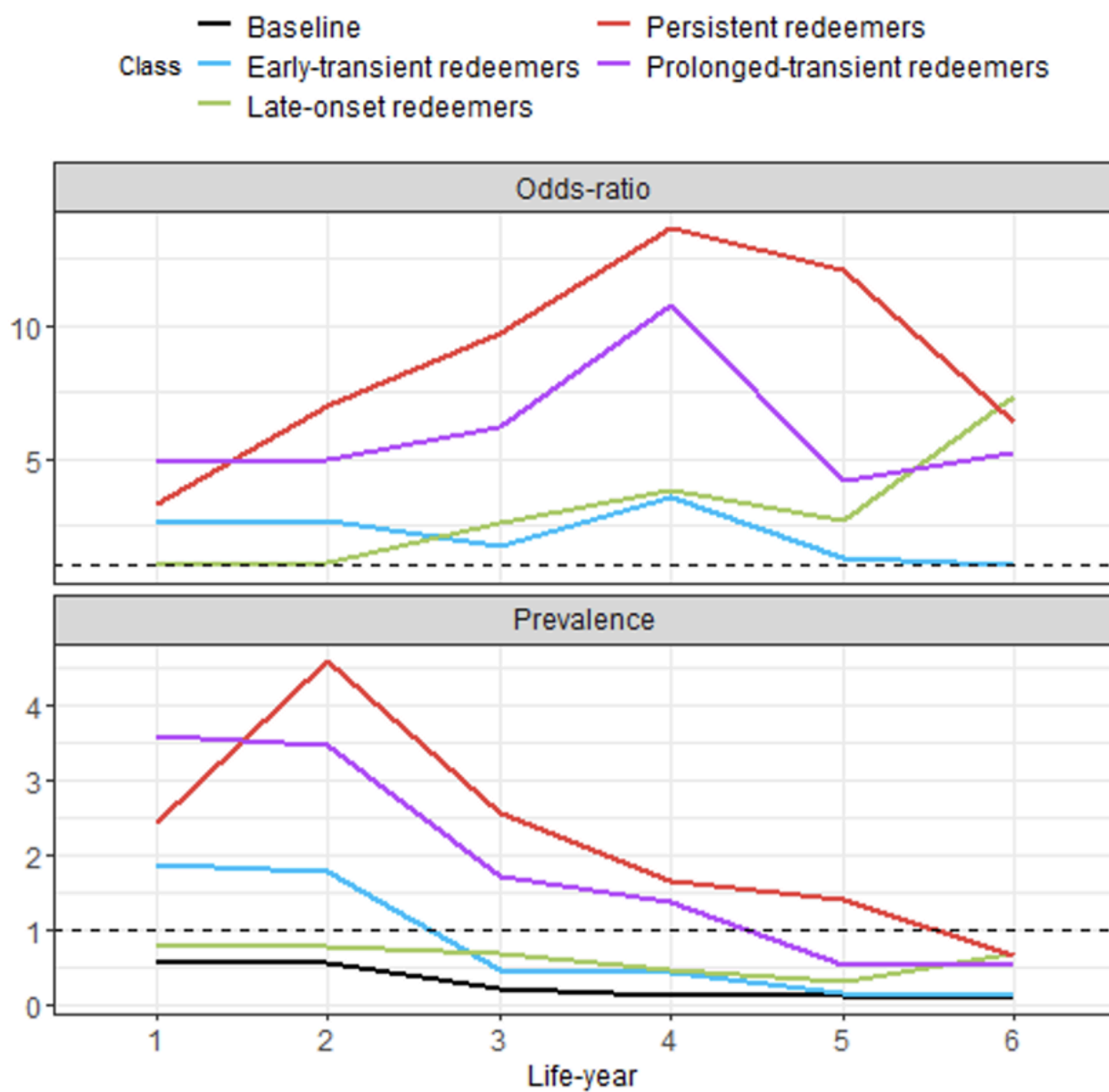


Supplementary Figure 12: Longitudinal lower-respiratory tract infections (LRTI)

diagnoses. Top plot shows odds ratios for each phenotype compared to baseline, while

the bottom plot shows prevalences, including baseline.

LRTI



References

1. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015 Nov 17;7:449–90.
2. Baadsgaard M, Quitzau J. Danish registers on personal income and transfer payments. *Scand J Public Health*. 2011 Jul;39(7_suppl):103–5.
3. Jensen VM, Rasmussen AW. Danish education registers. *Scand J Public Health*. 2011 Jul;39(7_suppl):91–4.
4. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Struct Equ Model Multidiscip J*. 2007 Oct 23;14(4):535–69.