Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Interactions with early-life exposures modulate polygenic risk of wheeze and asthma in preschool-aged children

Jihoon Choi[1], Amirtha Ambalavanan[2], Yang Zhang[3], Ruixue Dai[4], Elinor Simons[5], Hind Sbihi[6], Sonia Anand[7], Guillaume Paré[7], Diana Lefebvre[8], Stuart Turvey[6], Piush Mandhane[9], Allen Becker[5], Meghan Azad[5], Theo Moraes[10], Malcolm Sears[7], Padmaja Subbarao[7], and Qingling Duan[1]

[1]Queen's University
[2]Queen's University Department of Biomedical and Molecular Sciences
[3]Queen's University School of Computing
[4]The Hospital for Sick Children
[5]University of Manitoba Children's Hospital Research Institute of Manitoba
[6]The University of British Columbia
[7]McMaster University
[8]McMaster University Faculty of Health Sciences
[9]University of Alberta
[10]University of Toronto

October 5, 2020

Abstract

Background: Asthma is a multifactorial disease with numerous associated genetic and environmental risk factors, however, gene-environment interactions are poorly understood in modulating disease risk. This study determines the polygenic effects of multiple genetic loci and interactions with environmental exposures during early infancy on risk of recurrent wheeze and asthma in pre-school aged children. Methods: We conducted genome-wide association studies (GWAS) and applied a thresholding method to calculate genetic risk scores (GRS) of recurrent wheeze and asthma in 2835 children of the CHILD Cohort Study. Recurrent wheeze was defined as two or more episodes in one year between ages 2-5 years and asthma was diagnosed at age 5 years. In addition, we tested for interaction effects between the GRS and environmental exposures on these respiratory outcomes. Results: GWAS identified associations with known asthma loci on chromosome 17q12 - 17q21 ($p < 5e\text{-}8$). GRS analysis determined that the weighted addition of alleles at four childhood-asthma loci correlated with more than 2-fold higher prevalence of recurrent wheeze ($p = 1.5e\text{-}08$) and asthma ($p = 9.4e\text{-}08$) between high vs. low GRS groups. In addition, the GRS interacts with breastfeeding ($p = 0.02$) and traffic air pollution (NO_2; $p < 0.01$) during the first year of life to modulate risk of recurrent wheeze and childhood-onset asthma. Conclusions: This study reports polygenic effects of multiple genetic loci, which interact with early-life exposures, to determine risk of respiratory outcomes during early childhood. Thus, asthma risk may be determined early in infancy when exposures may modulate genetic risk.

Keywords

Asthma

Wheeze

Breastfeeding

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Traffic Air Pollution

Gene-Environment Interactions (GxE)

Genetic Risk Score (GRS)

Genome-Wide Association Studies (GWAS)

Introduction

Asthma is a complex, multifactorial disease with numerous reported genetic and environmental risk factors.(1) It is estimated that genetic factors contribute to approximately 55-74% of asthma heritability, which is the proportion of the phenotypic variability attributed to genetic differences.(2,3) Genetic studies of asthma to date have reported hundreds of genes associated with asthma and related respiratory outcomes such as lung function, however, these genetic loci account for only a small fraction of the estimated heritability.(4) For example, Vicente *at al.* reported that 31 well-replicated asthma variants collectively explain for only 2.5% of asthma heritability. Moreover, this study estimated that all common genetic variants (i.e. single nucleotide polymorphisms or SNPs) account for only 14% of asthma heritability.(5) The difference between the estimated asthma heritability and that explained by known genetic variants is referred to as the *missing heritability* .

In this manuscript, we test the hypothesis that the missing heritability of asthma may be explained in part by polygenic effects of multiple genetic factors as well as interactions with environmental exposures. Genetics studies of asthma to date have focused on associations of single variants, with few evaluating the polygenic or additive effects of multiple genetic factors.(6) Such polygenic effects are assessed as genetic risk scores (GRS), which is calculated as the total number of risk alleles multiplied by their individual weights or allelic effect sizes.(7,8) Moreover, we investigate the potential interactions between genetic risk and environmental exposures during early infancy on risk of developing childhood-onset asthma.(9) Previous studies have reported interactions between a well-established locus on chromosome 17q21 and exposures to rhinovirus infection and cigarette smoke.(10–14)

This is the first genetic study of asthma using data from the Canadian CHILD Cohort Study (N=3,455) and the first study to determine GRS using results from the largest genome-wide association study (GWAS) of childhood-onset asthma to date.(15) Furthermore, we test for interactions between the GRS and modifiable exposures (e.g., breastfeeding, pet ownership, cigarette smoke, traffic air pollution) on respiratory outcomes including recurrent wheeze and asthma by age 5. Recurrent wheeze between ages 2-5 years is the primary outcome of this study given that clinical diagnosis of asthma during early childhood is known to be difficult due to heterogeneity of asthma phenotypes.(16) In addition, standard assessments of lung health for the diagnosis of asthma such as spirometry measures and methacholine challenge tests are not typically used for children before age 7.(17)

In addition to determining the polygenic effects of multiple genetic variants associated with childhood-onset asthma, this manuscript is one of few studies to assess genetic interactions with environmental exposures on respiratory outcomes during infancy. The GRS may be used to assess genetic risk of developing recurrent wheeze and asthma early in childhood, when disease risk may be modified by exposures.

Methods

Study population

The CHILD Cohort Study is a Canadian birth cohort consisting of 3455 children and their parents.(18,19) Mothers were initially recruited during the second or third trimesters of pregnancy from four provinces across Canada: British Columbia (Vancouver), Alberta (Edmonton), Manitoba (Winnipeg, Morden and Winkler), and Ontario (Toronto). Their children have been followed from birth onwards, with a home visit at 3-4 months of age, repeated questionnaire surveys examining environmental exposures, and clinic visits for detailed assessments and obtaining biospecimens (e.g., blood, stool, urine, nasal swabs) at ages 1, 3, and 5 years.(19–21) Written informed consent was obtained from the parents of children enrolled in this study.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Protocols were approved by the Human Research Ethics Boards at Queen's University, McMaster University, the Hospital for Sick Children, and the Universities of Manitoba, Alberta, and British Columbia.

Lung health and questionnaire data

Health questionnaires were used to determine presence and frequency of wheeze and its triggers, and physician diagnosed asthma among the children. These questionnaires were administered at 3 and 6 months postpartum, repeated every 6 months from ages 1 year to 3 years, then annually to 5 years. Recurrent wheeze was defined as two or more episodes of wheeze symptoms in one year between 2-5 years of age. Asthma diagnosis by age 5 was determined by a pediatric consultant with expertise in asthma at clinic visits. This diagnosis resulted in three classifications: definite, possible or no asthma. Children who showed consistent symptoms and objective evidence of bronchodilator response (greater than 12% enhancement in FEV_1) were classified as *definite* asthma. Those who had consistent symptoms but did not show response to bronchodilator or had no symptoms but showed response to bronchodilator were classified as *possible* asthma. In order to maximize statistical power, we combined definite and possible asthma cases into a single case group and those with no asthma symptoms as the control group.

In addition to recurrent wheeze and physician diagnosed asthma, questionnaires completed by parents at 3, 6, 12, 18, 24, 30 months of age documented exposures to pre- and postnatal cigarette smoking, traffic air pollution measured by nitrogen dioxide (NO_2), pet ownership, and breastfeeding duration. Exposure to NO_2 was measured using city-specific land use regression (LUR) models, while accounting for residential mobility and temporal variance in ambient concentrations.(22) Estimates of exposure were measured for each residence reported by participants from address at conception to the address where the participating child resided at age 1 year.(22) All measures were temporally adjusted based on local ambient monitoring data per biweekly basis.(22) NO2 exposure were grouped into three categories: high, moderate, and low exposure. High was defined as 1 standard deviation above the mean (z score $>$ 1), low as 1 standard deviation below the mean (z score $<$ -1) and moderate exposure fell between high and low groups (-1 $<$ z score $<$ 1).

Genotyping arrays

Cord blood was collected at birth for DNA extraction and genotyping (Genetic and Molecular Epidemiology Laboratory, McMaster University). DNA was extracted using QiaSymphony automated large sample nucleic acid purification system. Subjects were genotyped using the Illumina HumanCoreExome BeadChip comprising over 500,000 single nucleotide variants (SNVs).

Genotype data pre-processing and imputations

Quality control of genotyped SNV data was carried out at both subject- and SNV-level using Plink 1.9, an open-source whole-genome analysis toolset.(23) Details are included in **Supplemental Data 1** . In brief, subjects who were missing more than 10% of genotypes were omitted, as were those with sex discrepancies (i.e. reported sex did not match sex as determined by genetic markers on the sex chromosomes), and related subjects as determined using identity by descent measures. The processed genotype data was then imputed using the Michigan imputation server.(24) First, chromosomes were phased using ShapeIT (v2.r790), then SNVs were imputed using the Minimac algorithm and the Haplotype Reference Consortium reference panel (r1.1 2016).(25–27) SNVs that passed the quality score R^2>0.3 were retained for further analysis.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was applied to determine genetic differences such as allele frequencies due to different ethnicities in our admixed cohort (i.e. population substructure). We used SMARTPCA from the Eigensoft package to conduct PCA of the genomics data.(28) In addition, we used 1000 Genomes Project phase 3 database as a reference cohort to identify ethnicities within the CHILD Study based on genetics information.(29)

Genome-wide association study (GWAS)

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Genome-wide association analysis of recurrent wheeze and asthma diagnosis used a logistic regression model, adjusted for sex and three significant PCs using Plink 1.9.(23) All variants with minor allele frequencies (MAF) less than 1% in the population were filtered out due to the lack of statistical power for detecting associations with rare variants. For significantly associated SNVs, local plots of recombination rates were visualized using LDassoc function of the LDlink web application.(30) Manhattan plots and quantile-quantile (qq) plots of GWAS analysis were generated using the qqman package in R.(31)

Genetic risk score (GRS) analysis and classification

In order to compute the polygenic effects of multiple SNVs, genetic risk scores (GRS) were calculated using summary statistics from the largest GWAS of childhood asthma published to date.(15) Our GRS used the pruning and thresholding approach as previously described (32): 1) Identified shared SNVs between the discovery (published GWAS summary statistics) and target (CHILD study) cohorts; 2) Eliminated redundant signals by pruning SNVs in high linkage disequilibrium (LD) using a window size of 50kb, shift distance of 2, and LD threshold (R^2) of > 0.8; 3) Employed a stepwise, forward regression analysis to calculate GRS for each individual, starting with the most significant SNV from the published GWAS, by summing the risk alleles weighted by the effect size (β) obtained from the discovery cohort; 4) Identified a set of genetic variants that best predicted the target trait using regression analysis while accounting for sex and the first three PCs as covariates.

Subjects were classified into high (1), moderate (0) and low (-1) GRS groups: defined as high if their GRS was 1 standard deviation higher from the mean (z score > 1), and low if 1 standard deviation lower than the mean (z score $<$ -1). Individuals who fall between high and low risk group (-1 $<$ z score < 1) were defined as having a moderate risk.

SNV annotations

Associated variants from the GWAS and GRS analysis were annotated using the Ensembl Variant Effect Predictor to determine their potential impact on gene transcription and translation (e.g., transcript expression, splicing, protein sequence) as well as reported associations with traits and diseases.(33) Non-coding variants were assessed for potential regulatory effects (i.e. expression quantitative trait loci) using RegulomeDB and Genotype-Tissue Expression (GTEx) databases.(34,35)

Gene-environment interactions

Interaction effects between the GRS and environmental exposures were investigated using a multiple regression model with an interaction term. This analysis used GRS classified as high, moderate, and low in addition to pre- and post-natal exposure to cigarette smoke, pets (dog, cat, furry, and non-furry animals), nitrogen dioxide, and breastfeeding duration. Sex and the first three PCs were included as covariates in the model.

Heritability estimates

Heritability, which is the proportion of variance in a particular trait (i.e. recurrent wheeze, asthma) accounted for by genetic factors was estimated using the Dsquared function in the modEvA package, which calculates the amount of deviance accounted for by a generalized linear model.(36) As above, this analysis included sex and the 3 first PCs as covariates.

Results

A total of 557,006 SNVs were genotyped in 2967 infants from the CHILD Cohort Study using the Illumina HumanCoreExome BeadChip. Following quality control (QC), a total of 2835 children remained for analysis, of whom 381 (13.4%) had experienced recurrent wheeze from ages 2-5 years, and 372 (13.1%) were diagnosed with asthma (including both definite and possible cases). **Supplemental Data 1 and Figure S1** outlines the QC steps and the number of subjects omitted from analysis at each step. Furthermore, we determined the genetic variance (e.g. allele frequencies) explained by each of the principal components (PCs).**Supplemental Figure S2** depicts changes in the eigenvalue (i.e. amount of variance captured) per addition of each PC,

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

which demonstrates that PCs 1, 2 and 3 collectively capture the majority of genetics variance in the CHILD Cohort. Moreover, we overlaid PC plots of CHILD subjects with participants of the 1000 Genomes Project (see**Supplemental Figure S3)** , which identified that 2090 CHILD subjects originated from Central Europe (approximately 74% of the CHILD Study cohort).

GWAS of recurrent wheeze and asthma validated associations with known asthma loci

GWAS of recurrent wheeze identified 98 significantly associated SNVs ($p < 5e\text{-}8$) in the full (multi-ethnic) cohort (n=2835,**Figure 1a**). No significant findings were identified in the smaller, Caucasian sub-cohort (n=2090, **Supplemental Figure S4**). All of the associated variants are located on chromosome 17q12 - 17q21, are in strong LD (r2>0.8), and maps to known genes*GRB7* , *IKZF3* , *ZPBP2* , *GSDMB* , and *ORMDL3*(**Figure 1b**). Annotation analysis identified that 5% of these variants are coding or regulatory (3% missense, 1% splice acceptor, and 1% located within known regulatory regions) whereas the majority are non-coding (53% intronic, 19% downstream, 13% intergenic, 4% located within the 5' or 3'-untranslated regions (UTRs), 2% are synonymous; **Supplemental Figure S5)** . Moreover, 28 SNVs are cis-eQTLs identified from lymphoblastoid cell lines, which were correlated with the expression of nearby genes. Annotations of the identified variants using Ensemble Variant Effect Predictor database show association with traits such as allergy, atopic march, hay fever and eczema.(33) Additional annotation details of associated SNVs can be found in **Supplemental Data 2** .

GWAS of physician diagnosed asthma in both the multi-ethnic, full cohort (**Figure 2a**) and the Caucasian sub-cohort (**Figure 2b**) identified one significant association for variant rs145454327. This SNV is located within an intronic region of the Calmodulin-regulated spectrin-associated protein 2 (*CAMSAP2)* gene on chromosome 1. Chromatin state model analysis of fetal lung tissue using RegulomeDB identified that this variant is located within an enhancer region (chr1:200753800..200754400), and affects the motif for 4 transcription factors: GATA-4, TBP, FOXP1, and Sox8 (**Supplemental Data 3**).

Polygenic effects of four asthma loci on chromosome 17 were associated with prevalence of recurrent wheeze and asthma during childhood

GRS analysis determined that the combined, weighted effects of 4 reported asthma variants (rs3816470, rs3902920, rs8076131, rs12603332) on chromosome 17q12-21 had the strongest association with recurrent wheeze between ages 2-5 in the CHILD cohort (**SupplementalData 4**). This association using a regression model was significant for the full, multi-ethnic cohort ($p = 1.5e\text{-}08$) and the Caucasian sub-cohort ($p = 8.9e\text{-}05$; **Supplemental Figure S6**). The SNV rs3816470 was associated with reduced risk (i.e. protective effect) whereas variants rs3902920, rs8076131, rs12603332 were associated with increased risk of recurrent wheeze. Three of these variants (rs3816470, rs8076131, and rs12603332) are known eQTLs correlated with the expression of nearby genes *ORMDL3* and*GSDMB* within lung and immune cells, reported by the Genotype-Tissue Expression (GTEx) project.

In a separate GRS analysis targeting asthma diagnosis by age 5 in the CHILD Study, the most correlated GRS (p=9.4e-08) consisted of the same four variants as for recurrent wheeze described above.**Supplemental Figure S7** demonstrates that GRS including additional variants is less associated with asthma and recurrent wheeze (shown in p-value).

The prevalence of recurrent wheeze was 8.7% among children in the low GRS group, 14.7% in moderate GRS group and 19.7% in high GRS group (**Figure 3a**). This shows a 2.3-fold increase in wheeze prevalence between the low vs. high GRS groups. Similarly, the prevalence of asthma in the low GRS group was 9.3%, 15.5% in the moderate GRS group and 19.9% in the high GRS group (**Figure 3b**). This indicates a 2.2-fold increase in asthma prevalence between the low and high GRS groups.

Environmental exposures modulate polygenic risk of recurrent wheeze

Several environmental exposures were tested for interaction with GRS in a multiple regression model (see list in **Supplemental Data 5**) accounting for sex and the top three PCs. Breastfeeding status at 12 months of age showed a significant interaction (p=0.02) with GRS to modify prevalence of recurrent wheeze. As shown

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

in **Figure 3c** , breastfeeding at 12 months was associated with a 2.2- and 1.5-fold lower prevalence of recurrent wheeze among those with low and moderate GRS, respectively. Interestingly, the prevalence of recurrent wheeze was not modified by breastfeeding among those in the high GRS group. In addition, nitrogen dioxide (NO_2) exposure in the first six months of life significantly interacts with GRS to increase prevalence of recurrent wheeze for the first quarter or 0-3 months of age (p=0.005) and for the second quarter or months 4-6 (p=0.01;**Supplemental Data 4**). **Figures 3d** & **3e**demonstrate that the prevalence of recurrent wheeze increases with genetic risk, however, the prevalence is highest for those exposed to high levels of NO_2. Specifically, among those with high GRS, the prevalence was 2.7-fold higher among those exposed to high NO_2 compared with low exposure from birth to 3 months of age. This effect of NO_2 exposure is reduced but still significant in the second quarter (4-6 months) among those with high GRS, showing that those with high NO_2 exposure had a 1.5-fold higher prevalence of wheeze compared to those with low exposure.

Interaction analysis did not identify any modifying effects of environmental exposures on GRS for physician diagnosed asthma (**Supplemental Data 6**).

Heritability estimates

Each of the four variants identified in the GRS analysis individually accounts for up to 1.4% of the heritability of recurrent wheeze (rs3816470, 1.3%; rs3902920, 1.3%; rs8076131, 1.4%; rs12603332, 1.4%). In comparison, the GRS, which combined the weighted effects of all four variants, explained for 2.0% of the variance in recurrent wheeze (**Table 1**). When the breastfeeding interaction term was added, 2.7% of variance in recurrent wheeze was explained. When both breastfeeding and NO_2 exposure interaction terms are added, 3.6% of recurrent wheeze was explained.

The four variants identified in GRS each explained for up to 1.5% of the variability in asthma (rs3816470, 1.25%; rs3902920, 1.12%; rs8076131, 1.13%; rs12603332, 1.51%). When the weighted effect of the four variants were combined in the GRS, this explained for 1.8% of the variance in asthma. When breastfeeding interaction term is added, 2.0% of asthma can be explained. When both breastfeeding and all NO_2 exposure interaction terms are added, 2.1% of asthma could be explained (**Table 1**).

Discussion

This manuscript reports polygenic effects of multiple genetic variants as well as modulating effects of environmental exposures on risk of recurrent wheeze and asthma during early childhood. First, we determined GRS for recurrent wheeze and asthma based on four genetic variants located on chromosome 17q12-21. These loci were previously associated with childhood-onset asthma in the largest published GWAS to date.(15) In addition, we observed that breastfeeding until one year reduced prevalence of recurrent wheeze, especially among children with moderate or low genetic risk but not among those with high genetic risk. This suggests that breastfeeding at one year of life reduces risk of recurrent wheeze but does not offset high genetic risk. In contrast, NO_2 exposure appeared to increase the risk of respiratory outcomes, particularly among children with moderate or high genetic risk but did not affect those in the low GRS group. Overall, our findings support the notion that asthma is a polygenic disease that is modulated by interactions with environmental exposures, which could account in part for the missing heritability. Moreover, our results indicate that while asthma may be difficult to diagnose in preschool-aged children, risk of respiratory outcomes can be detected during early infancy using recurrent wheeze between ages 2-5 years.

Our GWAS of recurrent wheeze identified 98 associated SNVs, albeit all are in strong linkage disequilibrium (LD, $r^2 > 0.8$), suggesting that they represent a single associated locus. Of these, 19 have been previously correlated with asthma or a related trait (e.g., allergy, atopic march, hay fever and eczema) and only two are coding. For example, SNV rs2305479 is a missense variant located within the *GSDMB* gene, resulting in an amino acid substitution (Glycine to Arginine at position 304/416 of the Gasdermin-B protein), and has been previously implicated in inflammatory processes. An earlier study reported that rs2305479 leads to abnormal sulfatide transport, which might compromise the integrity of the epithelial cell barrier and promote inflammatory processes in inflammatory bowel disease (IBD).(37) In addition, rs11078928 codes for a splice acceptor, which could result in an alternative mRNA isoform. Further experiments are needed to determine

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

the biological effects of these genetic variants as well as others in the same associated LD block.

In addition to SNVs associated with recurrent wheeze on chromosome 17, which is a well-established locus for childhood asthma, we identified that SNV rs145454327 on chromosome 1 was also associated with asthma by age 5 years.(15) This intronic variant was previously associated with childhood-onset asthma in the largest GWAS to date.(15) Furthermore, chromatin state model analysis of fetal lung tissue identified that this variant could potentially alter the motif for 4 transcription factors, GATA-4, Tbp, FOXP1, and Sox8. GATA-4 have been previously implicated with pathogenesis in bronchial asthma (38) and FOXP1 is known to regulate epithelial cell fate and regeneration during lung development.(39)

While our GWAS of recurrent wheeze and asthma both identified associations with loci previously associated with asthma, there was no overlap of associated loci. These results may reflect the phenotypic heterogeneity of asthma as well as challenges in clinical diagnosis among young children before age 7 (when spirometry is difficult and the use of methacholine challenge tests are not typical).(16,17) In addition, we combined definite and possible asthma into one case group for GWAS, which may have further affected the accuracy of asthma diagnosis by age 5. However, the association of SNPs on chromosome 17q12-21, a well-replicated locus for childhood-onset asthma, with recurrent wheeze in this study indicates that asthma risk may be assessed among young children, even prior to the diagnosis of asthma.

In addition to individual SNV associations with recurrent wheeze and asthma, we identified the polygenic effects of four genetic variants correlated with both recurrent wheeze (p = 1.53e-08) and asthma diagnosis (p=9.39e-08). While the four variants identified in the GRS analysis are non-coding, located either within intronic and 3'UTRs, these have been previously identified to have potentially regulatory effects. For example, the three risk variants associated with increased prevalence of recurrent wheeze (rs3816470, rs8076131, and rs12603332) are known loci that alter the expression of genes (eQTLs), as reported by the Genotype-Tissue Expression (GTEx) project. Similarly, Tonchevaet *al.* reported that the protective SNV rs3902920 alters the expression of *ORMDL3* , a well-established asthma locus.(40) All three risk variants identified in the GRS are correlated with increased expression of *ORMDL3/GSDMB* in lung cells and immune cells (leukocytes) while the protective variant is correlated with decreased expression of the same genes. Studies have found that mice expressing increased levels of human *ORMDL3* or *GSDMB* have an asthma phenotype characterized by increased airway responsiveness and increased airway remodeling in the absence of airway inflammation.(41) This suggests that the four variants used in calculating GRS in this study may affect wheeze and asthma by regulating the expression of these well-established asthma genes.

It is note-worthy that while GWAS of asthma did not yield associations with loci on chromosome 17, our GRS analysis indicates that SNVs on this chromosome may be used to assess risk of asthma as well as recurrent wheeze. This may be explained by the modest effects of each individual SNV on chromosome 17, which could not be detected in an univariate analysis (i.e. GWAS) of single SNVs but the additive effects of these variants could be detected in our GRS analysis.

In addition to polygenic effects of multiple genetic variants, we determined that breastfeeding until 12 months and NO_2exposure during the first 6 months of life significantly interact with genetic factors to modulate risk of recurrent wheeze during early childhood. Specifically, we observed lower prevalence of recurrent wheeze among infants who were breastfed at 12 months compared to those who were not breastfed until 12 months. The protective effects of breastfeeding on wheeze in the CHILD study was previously reported by Azad *et al.* (42) In the current study, this protective effect was detected in children with modest or low genetic risk but not in the high GRS group, which suggests that high genetic risk offsets the protective effects of breastfeeding at one year of life. In contrast to breastfeeding, we detected higher prevalence of recurrent wheeze among infants exposed to higher NO_2 in the first 6 months of life. The detrimental effects on NO_2 on recurrent wheezing during the first two years of life in the CHILD study was previously reported.(43) However, our study shows that this effect is most evident among children with high genetic risk, whereby those exposed to high NO_2 have a 2.7-fold higher wheeze prevalence than those exposed to low NO_2.

This GRS study is the first to use results from the largest GWAS of childhood-onset asthma published

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

to date, which reported distinct association signals between childhood vs. adult-onset disease, suggesting different mechanisms of disease.(15) Prior GRS studies used summary statistics from GWAS of both adult- and childhood-onset asthma. Thus, our study uniquely focused on early-onset disease associations, resulting in an associated GRS based on four genetic variations on chromosome 17, which was the most significant peak from the childhood-onset asthma GWAS. We observed more than a 2-fold higher prevalence of recurrent wheeze and asthma between the high vs. low GRS groups. Moreover, compared to single associated SNVs, the GRS accounts for higher heritability estimates of recurrent wheeze and asthma (h2 = 2% vs. 1.4%). In fact, the GRS based on the additive effects of risk alleles at these four loci accounts for a similar heritability as 31 of the most replicated asthma variants (h2 = 2.5%). While a heritability estimate of 2% may seem low, it was estimated that all common variants account for up to 14% of asthma heritability.(5) Furthermore, we demonstrated that genetic risk interacts with environment exposures, which accounts for some of the missing heritability of asthma. In fact, this interaction accounts for up to 3.6% of the heritability of recurrent wheeze, which is an increase from considering the GRS alone.

While this manuscript reports a novel GRS of asthma and recurrent wheeze that is based on additive effects of childhood-onset asthma genes and identifies novel interactions with environmental exposures, our study has several limitations. First, while our GRS analysis utilized results from an independent GWAS of childhood asthma, the largest to date, replication of our GRS results in an independent asthma cohort is needed to validate our findings. A major challenge for replication is the identification of another cohort with both genetics and environmental exposures data in order to assess the effects of gene-environment interactions on childhood asthma. A second limitation is that our results are *in silico* -based and experimental studies are needed to test the functional impact, if any, of the identified variants on chromosomes 1 and 17. Given the large LD block on chromosome 17, which consists of numerous SNVs spanning multiple genes, identification of casual variants will be challenging. Finally, we have yet to assess additional environmental exposures as well as the long-term effects (i.e. beyond the first few year of life) on respiratory outcomes among the children as they grow. For example, methacholine challenge tests are ongoing in the CHILD Cohort Study at the age 8 clinical visits, which is ongoing and will result in more accurate diagnosis of asthma.

Conclusion

It is well established that both genetic and environmental factors contribute to risk of asthma and related traits (i.e. recurrent wheeze), however, few studies have evaluated polygenic effects as well as interaction effects of multiple genes and environmental factors. This lack of evidence is likely due to heterogeneity of asthma phenotypes as well as the lack of studies with both available exposure and genetic datasets. The CHILD Cohort Study successfully addressed these limitations with in-depth phenotyping, as well as the availability of environmental and genetics data to allow for assessment of interaction effects. In this study, we calculated the polygenic effects of multiple genetic risk factors, which account for greater heritability of asthma and recurrent wheeze compared to each individual variant. Moreover, we observed modulating effects of breastfeeding at one year of life and exposure to traffic air pollution (NO_2) on risk of recurrent wheeze during childhood. Our study identifies a period during early infancy when genetic risk of developing asthma and respiratory diseases may be assessed, and when early life interventions might modify risk of developing life-long respiratory diseases.

REFERENCES

1. Sham P, Bishop DT. *Analysis of multifactorial disease* . Bios 2000

2. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis*1990;**142** :1351–1358.

3. Thomsen SF, Ulrik CS, Kyvik KO, Ferreira MAR, Backer V. Multivariate genetic analysis of atopy phenotypes in a selected sample of twins.*Clin Exp Allergy* 2006;**36** :1382–1390.

4. Neville MDC, Choi J, Lieberman J, Duan QL. Identification of deleterious and regulatory genomic variations in known asthma loci.*Respir Res* Published Online First: 2018. doi:10.1186/s12931-018-0953-2

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

5. Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunol*2017;**6** :e165.

6. Belsky DW, Sears MR, Hancox RJ, Harrington HL, Houts R, Moffitt TE et al. Polygenic risk and the development and course of asthma: An analysis of data from a four-decade longitudinal study. *Lancet Respir Med*Published Online First: 2013. doi:10.1016/S2213-2600(13)70101-2

7. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. 2018. doi:10.1038/s41576-018-0018-x

8. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat. Rev. Genet. 2016. doi:10.1038/nrg.2016.27

9. Morales E, Duffy D. Genetics and Gene-Environment Interactions in Childhood and Adult Onset Asthma. Front. Pediatr. 2019. doi:10.3389/fped.2019.00499

10. 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* Published Online First: 2007. doi:10.1038/nature06014

11. Bisgaard H, Bønnelykke K, Sleiman PMA, Brasholt M, Chawes B, Kreiner-Møller E et al. Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. *Am J Respir Crit Care Med* Published Online First: 2009. doi:10.1164/rccm.200809-1436OC

12. Granell R, Henderson AJ, Timpson N, St. Pourcain B, Kemp JP, Ring SM et al. Examination of the relationship between variation at 17q21 and childhood wheeze phenotypes. *J Allergy Clin Immunol* Published Online First: 2013. doi:10.1016/j.jaci.2012.09.021

13. Bønnelykke K, Ober C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. J. Allergy Clin. Immunol. 2016. doi:10.1016/j.jaci.2016.01.006

14. Dizier MH, Bouzigon E, Guilloud-Bataille M, Siroux V, Lemainque A, Boland A et al. Evidence for gene × smoking exposure interactions in a genome-wide linkage screen of asthma and bronchial hyper-responsiveness in EGEA families. *Eur J Hum Genet* Published Online First: 2007. doi:10.1038/sj.ejhg.5201830

15. 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* Published Online First: 2019. doi:10.1016/S2213-2600(19)30055-4

16. Carr TF, Bleecker E. Asthma heterogeneity and severity. World Allergy Organ. J. 2016. doi:10.1186/s40413-016-0131-2

17. Ducharme FM, Dell SD, Radhakrishnan D, Grad RM, Watson WTA, Yang CL et al. Diagnosis and management of asthma in preschoolers: A Canadian Thoracic Society and Canadian Paediatric Society position paper.*Can Respir J* Published Online First: 2015. doi:10.1155/2015/101572

18. Sears MR, Greene JM, Willan AR, Wiecek EM, Taylor DR, Flannery EM et al. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. *N Engl J Med* 2003;**349** :1414–1422.

19. Subbarao P, Anand SS, Becker AB, Befus AD, Brauer M, Brook JR et al. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. *Thorax*2015;**70** :998–1000.

20. Takaro TK, Scott JA, Allen RW, Anand SS, Becker AB, Befus AD et al. The Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort study: assessment of environmental exposures. *J Expo Sci Environ Epidemiol* 2015;**25** :580–592.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

21. Moraes TJ, Lefebvre DL, Chooniedass R, Becker AB, Brook JR, Denburg J et al. The canadian healthy infant longitudinal development birth cohort study: Biological samples and biobanking. *Paediatr Perinat Epidemiol* Published Online First: 2015. doi:10.1111/ppe.12161

22. Sbihi H, Allen RW, Becker A, Brook JR, Mandhane P, Scott JA et al. Perinatal Exposure to Traffic-Related Air Pollution and Atopy at 1 Year of Age in a Multi-Center Canadian Birth Cohort Study. *Environ Health Perspect* Published Online First: 2015. doi:10.1289/ehp.1408700

23. 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* Published Online First: 2015. doi:10.1186/s13742-015-0047-8

24. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A et al. Next-generation genotype imputation service and methods. *Nat Genet* Published Online First: 2016. doi:10.1038/ng.3656

25. Delaneau O, Coulonges C, Zagury JF. Shape-IT: New rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics* Published Online First: 2008. doi:10.1186/1471-2105-9-540

26. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2014;**31** :782–784.

27. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A et al. A reference panel of 64,976 haplotypes for genotype imputation.*Nat Genet* Published Online First: 2016. doi:10.1038/ng.3643

28. 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* Published Online First: 2006. doi:10.1038/ng1847

29. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A et al. A global reference for human genetic variation. Nature. 2015. doi:10.1038/nature15393

30. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* Published Online First: 2015. doi:10.1093/bioinformatics/btv402

31. D. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw* Published Online First: 2018. doi:10.1101/005165

32. Choi SW, Mak TSH, O'Reilly PF. A guide to performing Polygenic Risk Score analyses. *bioRxiv* Published Online First: 2018. doi:10.1101/416545

33. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A et al. The Ensembl Variant Effect Predictor. *Genome Biol* Published Online First: 2016. doi:10.1186/s13059-016-0974-4

34. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;**22** :1790–1797.

35. The GteC. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;**45** :580–585.

36. Barbosa AM, Brown J., Jiménez-Valverde A, Real R. modEvA: Model Evaluation and Analysis. R Packag. version 0.5.2/r40. 2014.

37. Chao KL, Kulakova L, Herzberg O. Gene polymorphism linked to increased asthma and IBD risk alters gasdermin-B structure, a sulfatide and phosphoinositide binding protein. *Proc Natl Acad Sci U S A*Published Online First: 2017. doi:10.1073/pnas.1616783114

38. Caramori G, Lim S, Ito K, Tomita K, Oates T, Jazrawi E et al. Expression of GATA family of transcription factors in T-cells, monocytes and bronchial biopsies. *Eur Respir J* Published Online First: 2001. doi:10.1183/09031936.01.00040701

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

39. Li S, Wang Y, Zhang Y, Lu MM, DeMayo FJ, Dekker JD et al. Foxp1/4 control epithelial cell fate during lung development and regeneration through regulation of anterior gradient 2. *Development*2012;**139** :2500–2509.

40. Toncheva AA, Potaczek DP, Schedel M, Gersting SW, Michel S, Krajnov N et al. Childhood asthma is associated with mutations and gene expression differences of ORMDL genes that can interact. *Allergy Eur J Allergy Clin Immunol* Published Online First: 2015. doi:10.1111/all.12652

41. Das S, Miller M, Broide DH. Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases. In: *Advances in Immunology* . 2017 doi:10.1016/bs.ai.2017.06.001

42. Azad MB, Vehling L, Lu Z, Dai D, Subbarao P, Becker AB et al. Breastfeeding, maternal asthma and wheezing in the first year of life: A longitudinal birth cohort study. *Eur Respir J* Published Online First: 2017. doi:10.1183/13993003.02019-2016

43. Emenius G, Pershagen G, Berglind N, Kwon HJ, Lewné M, Nordvall SL et al. NO 2 , as a marker of air pollution, and recurrent wheezing in children: A nested case-control study within the BAMSE birth cohort.*Occup Environ Med* Published Online First: 2003. doi:10.1136/oem.60.11.876

Table 1 - Heritability estimation of GRS and GRS interacting with environmental factors

Trait / Variance explained (%)	GRS	GRS x BF	GRS x BF x NO2
Recurrent wheeze (2y-5y)	2.04	2.72	3.64
Asthma diagnosis at 5 years	1.77	1.98	2.1

Figure legends

Figure 1. Genome-wide association study of recurrent wheeze between age 2-5 years. a) Manhattan plot showing SNVs on chromosome 17q12 - 21 associated with recurrent wheeze in the CHILD Study Cohort ($p < 5e\text{-}8$). Y-axis depicts the -log10 transformed p-value and x-axis shows chromosomal positions. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively. **b)** Locus-zoom plot showing the specific region on chromosome 17q12 - 21 associated with recurrent wheeze. Red circles indicate that the SNVs are in strong LD, and numbers within each circle indicate that they are likely to be regulatory. These SNVs span 5 well-known asthma genes: *GRB7* ,*IKZF3* , *ZPBP2* , *GSDMB* , and *ORMLD3* .

Figure 2. Genome-wide association study of asthma at 5 years. a) Manhattan plot showing one SNV on chromosome 1q32.1 associated with asthma in the multi-ethnic CHILD Study Cohort ($p < 5e\text{-}8$). Y-axis depicts the -log10 transformed p-value and x-axis shows chromosomal positions. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively. **b)** Manhattan plot showing one SNV on chromosome 1q32.1 associated with asthma diagnosis in the Central European CHILD Study Cohort ($p < 5e\text{-}8$). Y-axis depicts the -log10 transformed p-value and x-axis shows chromosomal positions. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively.

Figure 3. GRS moderating the risk of recurrent wheeze and asthma. Y-axis indicates prevalence of recurrent wheeze (age 2-5 years) or asthma (diagnosed by age 5), and x-axis indicates low/moderate/high GRS. **a)** The prevalence of recurrent wheeze increases with GRS. **b)** The prevalence of asthma increases with GRS. **c)** The prevalence of recurrent wheeze decreases among children who were breastfed until 12 months, particularly among those with low or moderate genetic risk. **d)** The prevalence of recurrent wheeze increases among children who are exposed increasing amounts of NO_2 in the three months of life as well as**e)** during months 4-6, especially among those with moderate to high genetic risk.

Supplemental Figure S1. Flowchart of subject-level quality control process. From a total of 2967 children genotyped, subject level QC removed 56 who did not match the reported sex, 39 for genotype

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

heterozygosity and 37 who were likely related as indicated by identity by descent of alleles, resulting in 2835 subjects for further analysis.

Supplemental Figure S2. Principal component analysis. The y-axis shows the fraction of the variance explained by the individual PC (variance explained / total variance) and the x-axis displays the index of eigenvalues tested. The red vertical line at x=3 indicates the inclusion threshold we chose, which is where a steep decline in explained variance is observed.

Supplemental Figure S3. Overlay of the first two principal components (PC1 and PC2) from CHILD cohort and the 1000 Genomes Project (phase 3). The two dotted lines in the plot indicate the inclusion threshold of PC1 and PC2, chosen to identify subjects with Central European (CEU) ancestry in CHILD cohort.

Supplemental Figure S4. Manhattan plot showing GWAS results of recurrent wheeze in the Caucasian sub-cohort of the CHILD Study. The Y-axis depicts the -log10 transformed p-value and the x-axis shows chromosomal positions of SNVs. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively.

Supplemental Figure S5. Annotations of 98 SNVs associated with recurrent wheeze in the full, admixed CHILD Study cohort. The Pie chart on the left shows the breakdown of all predicted consequences and pie chart on the right shows the predicted consequences of protein-coding variants.

Supplemental Figure S6. Identification of the GRS associated with recurrent wheeze. This plot illustrates the process in determining the variants for inclusion in the GRS analysis. The y-axis is -log10 transformed p-value of the GRS and the x-axis shows the number of variants included in that GRS from the top associated 100 variants (results from previously published GWAS). The GRS with the lowest p-value occur with the inclusion of first (i.e. most strongly associated) 4 variants, which is indicated by the red colored bar (p=1e-7.8).

Supplemental Figure S7. GRS performance in predicting recurrent wheeze and asthma. Performance of GRS is compared for prediction of recurrent wheeze at 2-5 years and asthma diagnosed at 5 years. Y-axis of the plot shows -log10 transformed p-value and x-axis shows the number of variants included in the GRS. Both plots show that best performance is reached when first 4 variants are utilized in the GRS.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

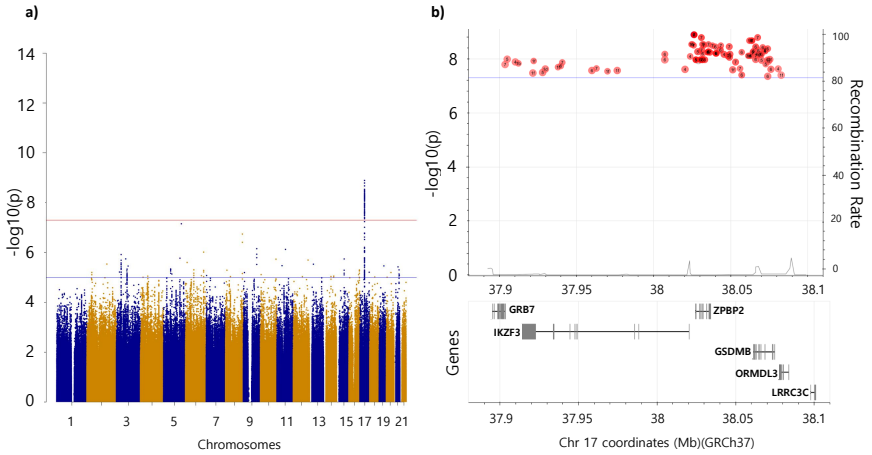

Figure 1: **Genome-wide association study of recurrent wheeze between age 2-5 years. a)** Manhattan plot showing SNVs on chromosome 17q12 - 21 associated with recurrent wheeze in the CHILD Study Cohort (p < 5e-8). Y-axis depicts the -log10 transformed p-value and x-axis shows chromosomal positions. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively. **b)** Locus-zoom plot showing the specific region on chromosome 17q12 - 21 associated with recurrent wheeze. Red circles indicate that the SNVs are in strong LD, and numbers within each circle indicate that they are likely to be regulatory. These SNVs span 5 well-known asthma genes: *GRB7* ,*IKZF3* , *ZPBP2* , *GSDMB* , and *ORMLD3* .

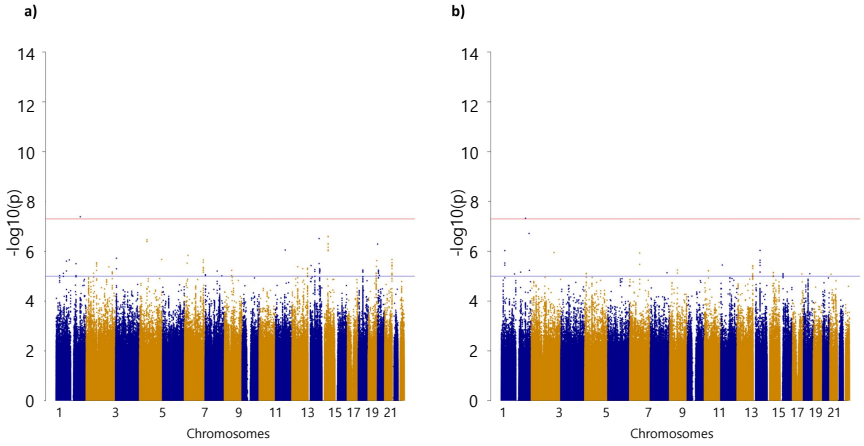

Figure 2: **Genome-wide association study of asthma at 5 years. a)** Manhattan plot showing one SNV on chromosome 1q32.1 associated with asthma in the multi-ethnic CHILD Study Cohort (p < 5e-8). Y-axis depicts the -log10 transformed p-value and x-axis shows chromosomal positions. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively. **b)** Manhattan plot showing one SNV on chromosome 1q32.1 associated with asthma diagnosis in the Central European CHILD Study Cohort (p < 5e-8). Y-axis depicts the -log10 transformed p-value and x-axis shows chromosomal positions. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

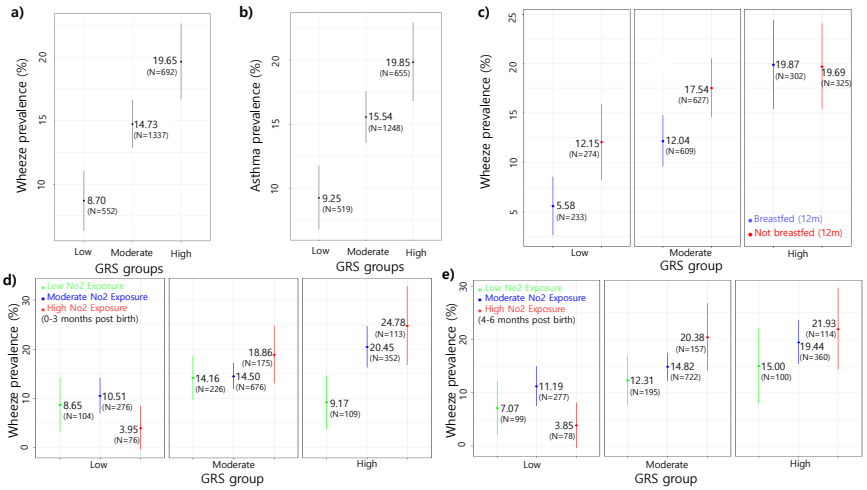

Figure 3: **GRS moderating the risk of recurrent wheeze and asthma.** Y-axis indicates prevalence of recurrent wheeze (age 2-5 years) or asthma (diagnosed by age 5), and x-axis indicates low/moderate/high GRS. **a)** The prevalence of recurrent wheeze increases with GRS. **b)** The prevalence of asthma increases with GRS. **c)** The prevalence of recurrent wheeze decreases among children who were breastfed until 12 months, particularly among those with low or moderate genetic risk. **d)** The prevalence of recurrent wheeze increases among children who are exposed increasing amounts of NO_2 in the three months of life as well as**e)** during months 4-6, especially among those with moderate to high genetic risk.

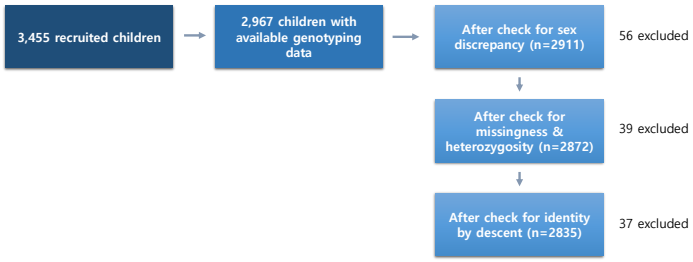

Figure 4: **Supplemental Figure S1. Flowchart of subject-level quality control process.** From a total of 2967 children genotyped, subject level QC removed 56 who did not match the reported sex, 39 for genotype heterozygosity and 37 who were likely related as indicated by identity by descent of alleles, resulting in 2835 subjects for further analysis.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

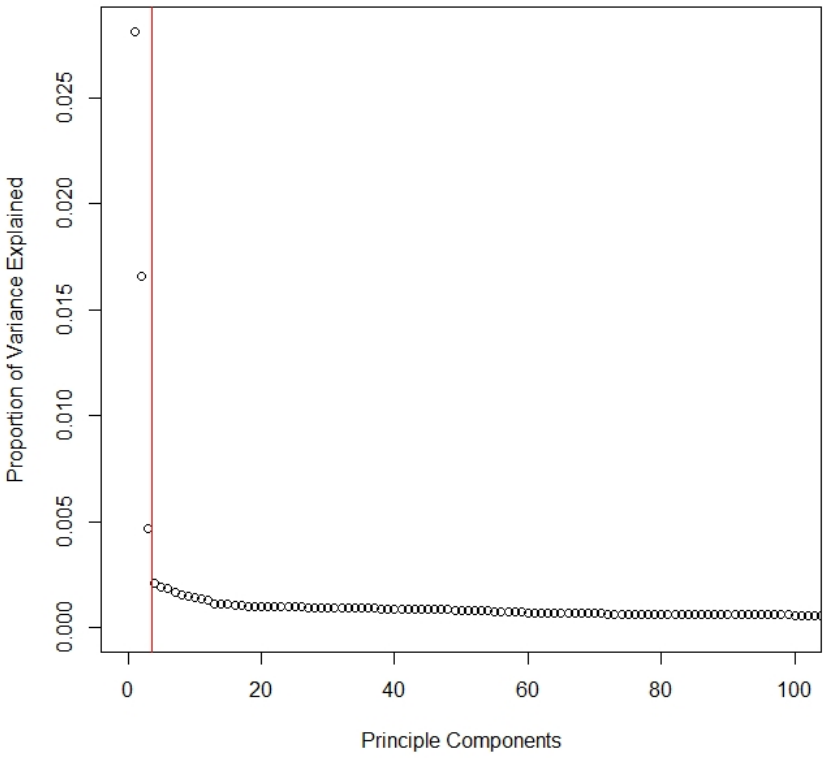

Figure 5: **Supplemental Figure S2. Principal component analysis.** The y-axis shows the fraction of the variance explained by the individual PC (variance explained / total variance) and the x-axis displays the index of eigenvalues tested. The red vertical line at x=3 indicates the inclusion threshold we chose, which is where a steep decline in explained variance is observed.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

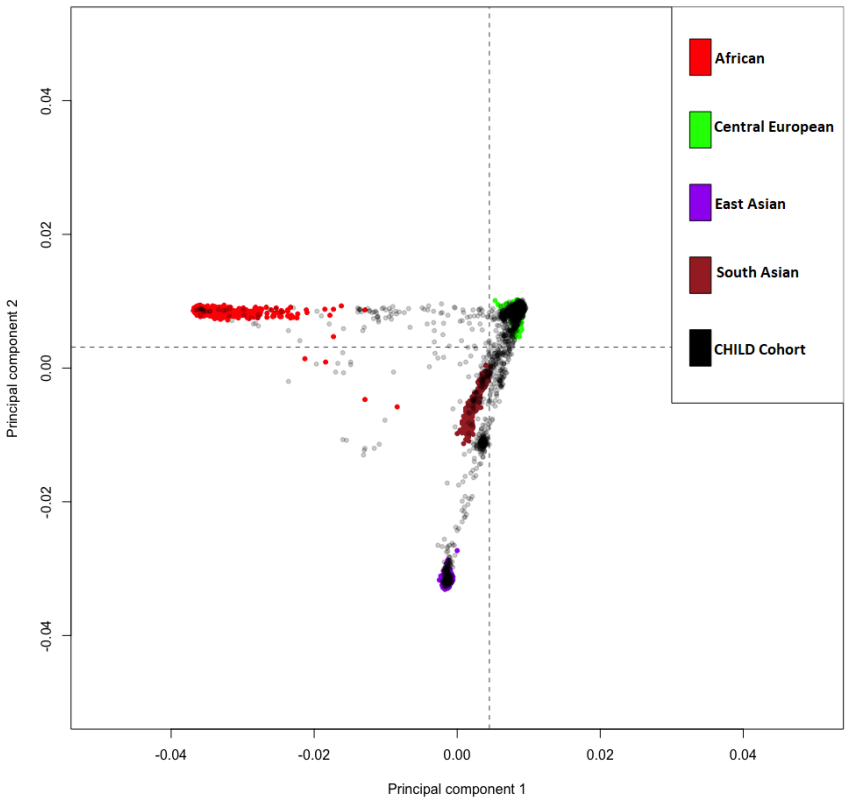

Figure 6: **Supplemental Figure S3. Overlay of the first two principal components (PC1 and PC2) from CHILD cohort and the 1000 Genomes Project (phase 3).** The two dotted lines in the plot indicate the inclusion threshold of PC1 and PC2, chosen to identify subjects with Central European (CEU) ancestry in CHILD cohort.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

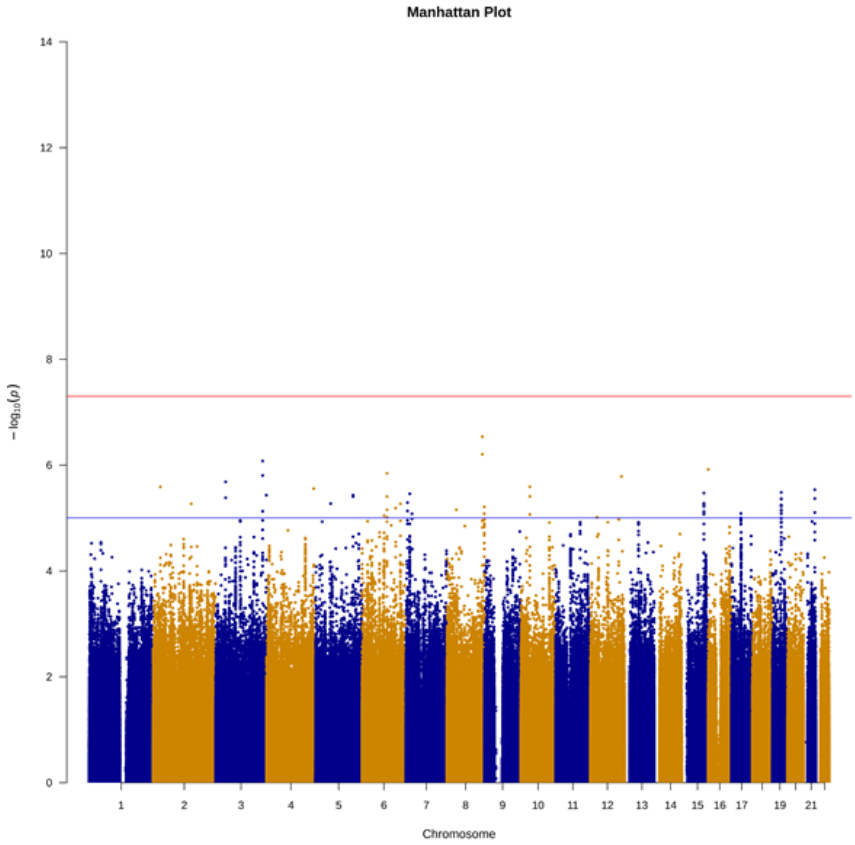

Figure 7: **Supplemental Figure S4. Manhattan plot showing GWAS results of recurrent wheeze in the Caucasian sub-cohort of the CHILD Study.** The Y-axis depicts the -log10 transformed p-value and the x-axis shows chromosomal positions of SNVs. Suggestive significance (p=5e-5) and genome-wide significance (p=5e-8) are respectively marked as blue and red horizontal lines, respectively.

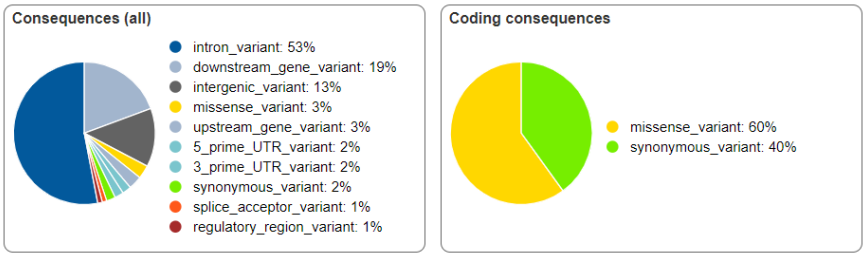

Figure 8: **Supplemental Figure S5. Annotations of 98 SNVs associated with recurrent wheeze in the full, admixed CHILD Study cohort.** The Pie chart on the left shows the breakdown of all predicted consequences and pie chart on the right shows the predicted consequences of protein-coding variants.

Posted on Authorea 5 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159986525.53189546/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

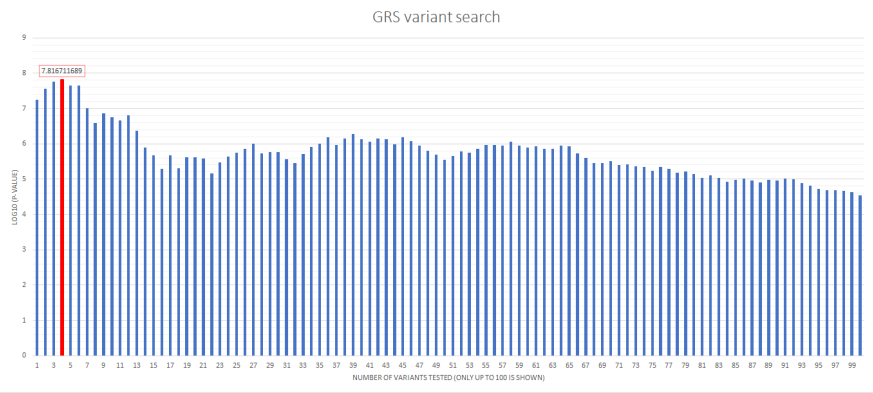

Figure 9: **Supplemental Figure S6. Identification of the GRS associated with recurrent wheeze.** This plot illustrates the process in determining the variants for inclusion in the GRS analysis. The y-axis is -log10 transformed p-value of the GRS and the x-axis shows the number of variants included in that GRS from the top associated 100 variants (results from previously published GWAS). The GRS with the lowest p-value occur with the inclusion of first (i.e. most strongly associated) 4 variants, which is indicated by the red colored bar (p=1e-7.8).

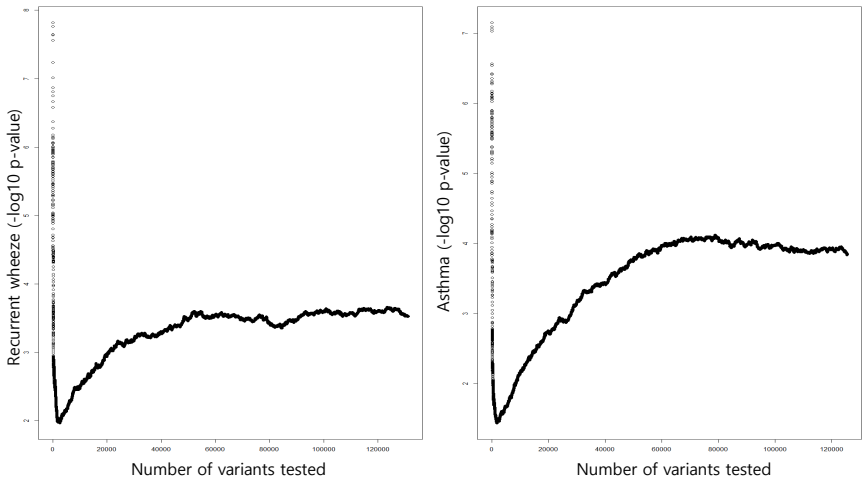

Figure 10: **Supplemental Figure S7. GRS performance in predicting recurrent wheeze and asthma.** Performance of GRS is compared for prediction of recurrent wheeze at 2-5 years and asthma diagnosed at 5 years. Y-axis of the plot shows -log10 transformed p-value and x-axis shows the number of variants included in the GRS. Both plots show that best performance is reached when first 4 variants are utilized in the GRS.