

FGF20 and PGM2 variants are associated with childhood asthma in family-based whole-genome sequencing studies

Julian Hecker^{1,2,*}, Sung Chun^{3,4}, Ahmad Samiei³, Cuining Liu⁵, Cecelia Laurie⁶, Priyadarshini Kachroo^{1,2}, Sharon M. Lutz^{2,7,8}, Sanghun Lee^{7,9}, Albert V. Smith¹⁰, Jessica Lasky-Su^{1,2}, Michael H. Cho^{1,2,11}, Sunita Sharma⁵, Manuel Enrique Soto Quirós¹², Lydiana Avila¹², Juan C. Celedón¹³, Benjamin Raby^{2,4}, Xiaobo Zhou^{1,2}, Edwin K. Silverman^{1,2,11}, Dawn L. DeMeo^{1,2}, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Christoph Lange^{1,2,7} and Scott T. Weiss^{1,2}

¹Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

²Harvard Medical School, Boston, MA 02115, USA

³Division of Pulmonary Medicine, Boston Children's Hospital, Boston, MA 02115, USA

⁴Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

⁵Division of Pulmonary Sciences and Critical Care Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

⁶Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

⁷Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

⁸Department of Population Medicine, PREcisiOn Medicine Translational Research (PROMoTeR) Center, Harvard Pilgrim Health Care, Boston, MA 02215, USA

⁹Department of Medical Consilience, Division of Medicine, Graduate School, Dankook University, Yongin-si, 16890, South Korea

¹⁰Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

¹¹Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

¹²Department of Pediatrics, Hospital Nacional de Niños, 10101 San José, Costa Rica

¹³Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA 15224, USA

*To whom correspondence should be addressed at: 181 Longwood Avenue, 4th Floor, Boston, MA 02115, USA; Tel: +1 6175252270; Fax: +1 6175250958; Email: rejhe@channing.harvard.edu

Abstract

Background: Asthma is a heterogeneous common respiratory disease that remains poorly understood. The established genetic associations fail to explain the high estimated heritability, and the prevalence of asthma differs between populations and geographic regions. Robust association analyses incorporating different genetic ancestries and whole-genome sequencing data may identify novel genetic associations. **Methods:** We performed family-based genome-wide association analyses of childhood-onset asthma based on whole-genome sequencing (WGS) data for the 'The Genetic Epidemiology of Asthma in Costa Rica' study (GACRS) and the Childhood Asthma Management Program (CAMP). Based on parent-child trios with children diagnosed with asthma, we performed a single variant analysis using an additive and a recessive genetic model and a region-based association analysis of low-frequency and rare variants. **Results:** Based on 1180 asthmatic trios (894 GACRS trios and 286 CAMP trios, a total of 3540 samples with WGS data), we identified three novel genetic loci associated with childhood-onset asthma: rs4832738 on 4p14 ($P = 1.72 \times 10^{-9}$, recessive model), rs1581479 on 8p22 ($P = 1.47 \times 10^{-8}$, additive model) and rs73367537 on 10q26 ($P = 1.21 \times 10^{-8}$, additive model in GACRS only). Integrative analyses suggested potential novel candidate genes underlying these associations: PGM2 on 4p14 and FGF20 on 8p22. **Conclusion:** Our family-based whole-genome sequencing analysis identified three novel genetic loci for childhood-onset asthma. Gene expression data and integrative analyses point to PGM2 on 4p14 and FGF20 on 8p22 as linked genes. Furthermore, region-based analyses suggest independent potential low-frequency/rare variant associations on 8p22. Follow-up analyses are needed to understand the functional mechanisms and generalizability of these associations.

Introduction

Asthma is a common chronic respiratory disease (1) characterized by substantial clinical heterogeneity. Patients differ across a range of clinical manifestations and, despite strong evidence for genetic predisposition, asthma etiology remains poorly understood (2). Recent large-scale genetic association analyses showed that the genetic architectures of childhood-onset and adult-onset asthma are partly distinct, likely due to age-of-onset-dependent disease mechanisms (3,4). Furthermore, the prevalence of childhood-onset asthma differs between populations and geographic regions

(2,5). Genetic association studies of asthmatic children that include multiple genetic ancestries could potentially facilitate the identification of the underlying biological mechanisms. Moreover, given the underlying differences between childhood-onset and adult-onset asthma as well as recall biases, targeted studies based on physician-diagnosed asthmatic children are required; biobank analyses are usually based on self-reported medical history (3,4).

In this work, we utilized whole-genome sequencing (WGS) data and performed genome-wide association analyses of childhood asthma based on two family-based studies: 'The Genetic Epi-

Table 1. Characteristics of the asthmatic children in GACRS and CAMP

	GACRS	CAMP ^a	Comparison test
N	894	286	
% Female	40.9	32.5	P = 0.013
Age (mean (SD) in years)	9.28 (1.86)	8.90 (2.16)	P = 0.008
Height (mean (SD) in meter)	1.33 (0.12)	1.34 (0.14)	P = 0.60
FEV ₁ (mean (SD) in liter)	1.80 (0.52)	1.65 (0.47)	P = 1.03e-05
FEV ₁ % predicted (mean (SD))	98.87 (16.84)	92.82 (13.40)	P = 1.17e-09
Maternal asthma history % yes	30.0	25.4	P = 0.16
Paternal asthma history % yes	25.4	22.6	P = 0.34
Race	Hispanic	75% White/10% African American/7% Hispanic/8% other	-

^aCAMP data at randomization. Comparison tests between GACRS and CAMP based on chi-squared or t-test, depending on the type of measurement.

demology of Asthma in Costa Rica' study (GACRS) (6) and the Childhood Asthma Management Program (CAMP) (7). For both studies, WGS data were generated as part of the NHLBI Trans-Omics for Precision Medicine (TOPMed) program (8). Given the family-based designs consisting of asthmatic trios (two parents and a child with asthma), we applied robust family-based association tests (FBATs) (9,10). We incorporated both the commonly used additive and a recessive model for the single variant analysis of common genetic variants. The latter was motivated by recent findings using recessive models for other complex diseases (11,12). Furthermore, WGS data allowed us to analyze region-based associations based on low-frequency and rare variants using a recently proposed FBAT framework (13,14). Our analysis incorporated a total of 3540 samples, representing 1180 asthmatic trios (894 GACRS trios and 286 CAMP trios). To the best of our knowledge, this is the largest WGS family-based association analysis of childhood-onset asthma to date.

Results

Baseline characteristics

After quality control, we kept WGS data for 894 GACRS and 286 CAMP trios. In Table 1, we report the baseline characteristics of the asthmatic offspring in both studies. For CAMP, these data reflect the randomization timepoint of the study. The two study populations possess overall comparable characteristics regarding the sex ratio, age distribution, parental asthma history and height. Differences are observed regarding lung function measurements and reported race. As the FBAT approach is robust against any confounding due to population substructure and admixture, the FBAT statistics can be computed for the combined sample of trios from GACRS and CAMP.

Genome-wide single variant association analysis

We performed power calculations for the single variant analysis based on an additive (multiplicative relative risk) and a recessive model (for minor and major allele). In Supplementary Material, Figure S1, we plotted the corresponding power curves based on a sample size of $n = 1180$ affected trios and established power formulas (15). These computations demonstrated that we have adequate power for common variants with substantial effect sizes.

In the combined GACRS + CAMP FBAT single variant analysis, we evaluated the additive association P -value for 9330567 variants. For the recessive model, we evaluated 7444202 association P -values (including a substantial number of variants with two recessive P -values, one for each allele, see section Materials and Methods). In Figure 1, we visualized the Miami

plot for the additive and recessive analysis. In Supplementary Material, Figures S2 and S3, we plotted quantile-quantile-plots for the additive and recessive analyses. These plots demonstrate the validity of the analyses and show that the exclusion of the lead SNP in the well-established 17q21 locus (16), and a 500 kb flanking region, removes a substantial part of the inflation of the test statistics.

Considering both the results for the additive and recessive models, we observed four genetic loci with variants that reached a genome-wide significant association P -value $P < 5 * 10^{-8}$. The well-established 17q21 locus was the most significant association in our additive GACRS + CAMP analysis (16). The additive GACRS + CAMP association P -value of the corresponding lead variant rs8076131 was $P = 5.35 * 10^{-11}$. The corresponding association P -values in GACRS and CAMP were $P = 3.89 * 10^{-10}$ and $P = 0.0239$, respectively (Supplementary Material, Table S4). Describing novel associations, variants in the 8p22 locus reached genome-wide significance in the additive GACRS + CAMP analysis. The lead variant rs1581479 obtained an association P -value of $P = 1.47 * 10^{-8}$ (Fig. 2, Table 2). The corresponding association P -values in GACRS and CAMP were $P = 1.83 * 10^{-4}$ and $P = 2.43 * 10^{-6}$, respectively.

Based on the recessive model, we observed genome-wide significant associations in the 4p14 locus. The lead variants rs4832738, rs7691795, and rs4832956 (in complete LD) reached an association P -value of $P = 1.72 * 10^{-9}$ (Table 2). All three P -values were based on the recessive model test for the major allele with an estimated allele frequency of 50.4% in GACRS + CAMP (Table 2 and Supplementary Material, Table S4). Supplementary Material, Figures S4–S6 visualize region association plots for 4p14. Supplementary Material, Figure S4 is based on standard LD correlation (LocusZoom, EUR LD), Supplementary Material, Figure S5 is based on a recessive genotype coding in 1000 Genomes phase 3 EUR, and Supplementary Material, Figure S6 is based on a recessive genotype coding in GACRS + CAMP (see section Materials and Methods). We added the two region plots in Supplementary Material, Figures S5 and S6 since LD does not correspond to the correlation between test statistics for the recessive model.

Considering GACRS separately, we observed highly significant low-frequency variants with minor allele frequencies (MAFs) of approximately 3% on 10q26 in the additive analysis. The lead variant rs73367537 had an additive association P -value of $P = 1.21 * 10^{-8}$ in GACRS and all significant variants were in strong LD (Supplementary Material, Fig. S7). However, the corresponding additive association P -value in CAMP was nearly significant ($P = 0.059$) but with an opposite direction of effect (Table 2, Supplementary Material, Table S4). Aside the established 17q21 associations, the identified SNPs on 4p14, 8p22, and 10q26 were not previously described in the childhood-onset asthma context

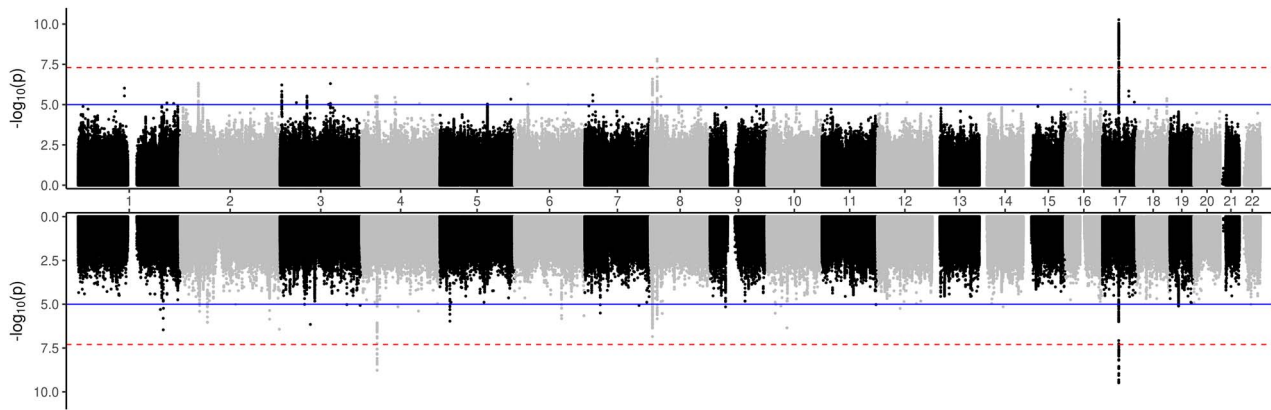
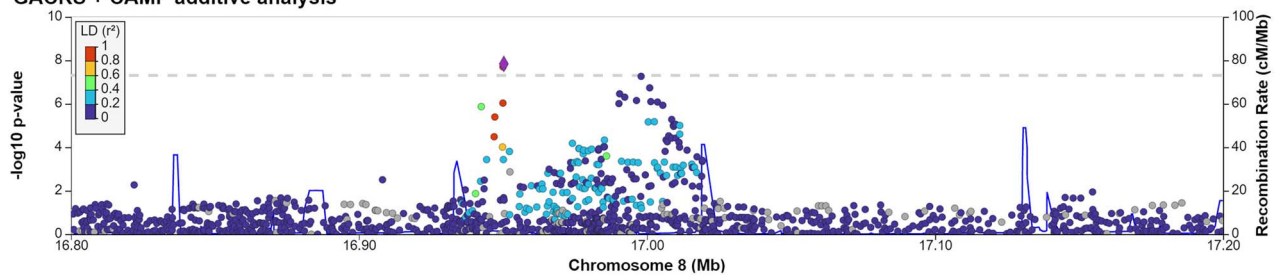


Figure 1. Miami plot for the GACRS + CAMP single variant analyses based on the additive (top) and recessive (bottom) model.

A

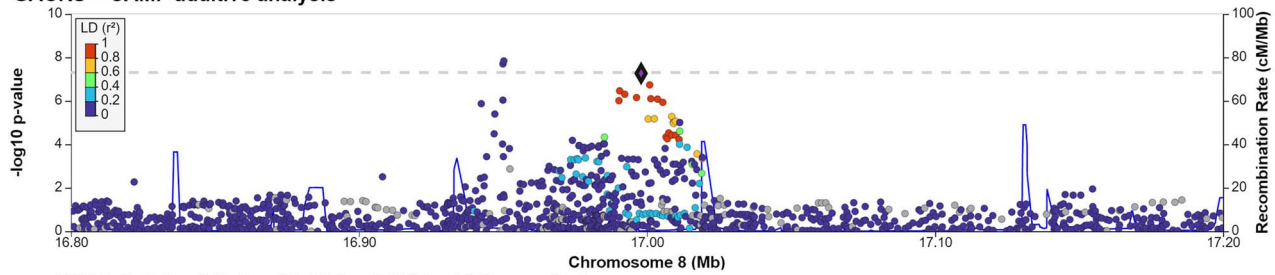
GACRS + CAMP additive analysis



GWAS Catalog hits for GACRS + CAMP additive analysis

B

GACRS + CAMP additive analysis



GWAS Catalog hits for GACRS + CAMP additive analysis

Figure 2. Region association plots for 8p22 based on the additive GACRS + CAMP analysis (LocusZoom, EUR LD). Reference variant rs1581479 (**A**) and reference variant rs17515041 (**B**). The LD between rs1581479 and rs17515041 is $r^2 = 0.1$ and $r^2 = 0.14$ in GACRS and CAMP, respectively (estimated based on parental genotypes). The LD pair tool estimated the LD between both variants to be $r^2 = 0.14$ and $r^2 = 0.11$ in EUR (European) and AMR (Ad Mixed American), respectively.

(GWAS catalog (17), June 2022) and therefore represent novel genetic associations for childhood asthma.

Single variant results replicate previously identified genetic associations for childhood asthma

Among the 87 genetic variants in our additive GACRS + CAMP analysis that were previously reported as childhood-onset asthma associations in the Ferreira et al. analysis (3) (see section

Materials and Methods), 13 (15%) had a nominal significant additive GACRS + CAMP association P -value ($P \leq 0.05$). A binomial test for enrichment (null hypothesis : success probability ≤ 0.05) resulted in a significant P -value of $P = 3.85 \times 10^{-4}$, assuming independent genetic variants and thus neglecting residual LD in this pruned set of variants. All of the 13 nominal significant associations had a consistent direction of effect. Considering directions of effects only, 70 out of 87 (80%) had a consistent direction in the additive GACRS + CAMP analysis. The

Table 2. Novel genetic associations for childhood asthma in either GACRS +CAMP or GACRS/CAMP only, based on an additive or recessive model

SNP	CHR	BP (hg38)	A1	A2	TEST ALLELE	MAF GACRS +CAMP	Z GACRS + CAMP	P GACRS + CAMP	Z GACRS	P GACRS	Z GACRS	P GACRS	Z CAMP	P CAMP	cluster	model	colocalization
rs4832738 ^a	4	37 813 371	G	A	A	0.4964	6.03	1.72e-09	5.65	1.62e-08	2.22	2.62e-02	4p14	recessive		eQTL: PGM2 in GTEx lung, fetal lung, immune cell types in eQTL Catalogue. Cluster contains meQTLs for cg20524452 and cg04038500 (blood).	
rs1581479	8	16 950 316	C	T	C	0.1801	5.67	1.47e-08	3.74	1.83e-04	4.71	2.43e-06	8p22-1	additive			
rs17515041	8	16 997 985	C	CTCTTG	C	0.4267	-5.43	5.50e-08	-4.12	3.70e-05	-3.71	2.11e-04	8p22-2	additive		eQTL: FGF20 in GTEx spleen. Cluster contains meQTLs for cg04451175 and cg06872257 (blood).	
rs73367537	10	120 390 728	T	C	T	0.0261	4.09	4.30e-05	5.70	1.21e-08	-1.89	5.88e-02	10q26	additive			

^aIn complete LD with rs7691795 and rs4832956. The column 'cluster' refers to the fine mapping results. The column 'colocalization' summarizes the results of the eQTL colocalization analysis and the overlap analysis between the fine mapped signal cluster and external meQTL results. eQTL: expression quantitative trait locus; meQTL: methylation quantitative trait locus.

corresponding binomial test for enrichment (null hypothesis: success probability ≤ 0.5) resulted in a p-value of $P = 4.18 \times 10^{-9}$. The list of the 13 nominal significant associations can be found in [Supplementary Material, Table S1](#).

To get a quantitative understanding of the replication results, we performed a replication power analysis for our analysis based on 1180 affected trios using established power formulas for the TDT/FBAT (15). We extracted the GARCS/CAMP minor allele frequencies (estimated based on parental genetic data) and the reported odds ratios from the Ferreira *et al.* analysis for all 87 available associations/variants. We applied the *empirical_bayes* function (18) in the *winnescurse* R package (see section code and data availability) to the log odds ratios to correct for the Winner's curse (19). Next, we converted the adjusted odds ratios to relative risks (assuming a prevalence of 0.27) and computed the power for each variant to reach nominal significance ($P \leq 0.05$) in our GACRS + CAMP analysis. The expected number of nominally significant associations based on these computations was approximately 19.78. The probability of observing 13 or fewer nominally significant associations based on these power parameters was approximately $P = 0.04$ (based on 100 000 simulations). Although the number of observed associations is less than expected, we conclude that the deviations from the theoretical expectations are modest. In general, replication analyses for asthma risk loci are challenging, given the heterogeneity of asthma and asthma diagnoses.

Statistical fine mapping of single variant associations

We performed statistical fine mapping for all identified novel loci using DAP-G (see section [Materials and Methods](#)) to pinpoint genetic variants that most likely explain the observed associations. [Supplementary Material, Table S4](#) contains all variants in each identified cluster, where variants within a cluster are in strong LD with each other. This table also contains the posterior inclusion probabilities for each genetic variant as computed by DAP-G. In line with the regional association plots for the 8p22 association ([Fig. 2](#)), DAP-G found evidence for an independent signal with lead variant [rs17515041](#) in the 8p22 locus. The corresponding association P-value was $P = 5.5 \times 10^{-8}$ in the GACRS + CAMP analysis, whereas the cohort-specific association P-values for GACRS and CAMP were $P = 3.7 \times 10^{-5}$ and $P = 2.11 \times 10^{-4}$, respectively ([Table 2](#) and [Supplementary Material, Table S4](#)). The LD between [rs1581479](#) and [rs17515041](#) is $r^2 = 0.1$ and $r^2 = 0.14$ in GACRS and CAMP, respectively, estimated based on parental genotypes. These two clusters on 8p22 will be denoted by 8p22-1 and 8p22-2 in the following. For the other loci, only one signal cluster was identified. [Supplementary Material, Table S4](#) contains additional information including annotations for all genetic variants located in the identified clusters ([Material and Methods](#)).

Expression quantitative trait loci colocalization analysis

We tested the single variant associations on 4p14, 8p22-1, 8p22-2 and 10q26 for colocalization with expression quantitative trait loci (eQTLs). Since there are two independent associations in the 8p22 locus, approximately 50 kb apart, we used a radius of 25 kb between the two associations when extracting the FBAT statistics for the colocalization analysis. For the other loci, we used a radius of 50 kb (see section [Materials and Methods](#)).

The lead variant [rs4832738](#) on 4p14 colocalized with an eQTL signal for PGM2 (Phosphoglucomutase 2) in fetal lung (false

discovery rate (FDR) $5 * 10^{-5}$), GTEx lung tissue (FDR $< 10^{-5}$) and several immune cell types in the eQTL Catalogue data. According to the GTEx eQTL results, the asthma risk increasing allele increases PGM2 expression in the lung.

For rs1989754 on 8p22-2, in strong LD with the lead variant rs17515041 on 8p22-2 ($r^2 > 0.98$ in AMR (Ad Mixed American) and EUR (European) populations in LDpair tool, see section code and data availability), we observed colocalization with an eQTL signal for FGF20 (Fibroblast Growth Factor 20) in spleen tissue in GTEx (FDR 0.0064). Here, the asthma risk decreasing allele increases expression in the spleen tissue. We tested rs1989754 since rs17515041 is an Indel and was removed in the eQTL studies. For the GWAS lead variants rs1581479 on 8p22-1 and rs73367537 on 10q26, no significant colocalizations with gene expression data were observed.

Overlap with meQTL results

Investigating the overlap between the identified signal clusters and recent methylation quantitative trait loci (meQTLs) results in blood (20) showed that rs3862103 and rs1721100 on 8p22-2 are clumped meQTLs for cg04451175 and cg06872257, respectively. On 4p14, we observed that rs34738612 and rs11721535 are clumped meQTLs for cg20524452 and cg04038500, respectively. A clumped meQTL represents an approximately fine mapped meQTL signal using LD clumping (see section [Materials and Methods](#)).

Replication of novel associations in UK Biobank and additional support for 10q26 by GACRS findings

Based on the publicly available summary statistics (see section code and data availability) for the analysis of childhood asthma in the UK Biobank (3), we investigated the association *P*-values of our identified variants on 4p14, 8p22 and 10q26. The identified variants failed to replicate (rs4832738 *P*=0.003829 with opposing direction of effect, rs1581479 *P*=0.6833 and rs73367537 *P*=0.1962). Also, rs1989754, in strong LD with the second association rs17515041 on 8p22-2, was not significant (*P*=0.3078). Association data for rs17515041 was not available. We additionally sought evidence for replication of our signal on 10q26. Since we observed heterogeneous effects for 10q26 in GACRS and CAMP, we performed a similar analysis as described by Ferreira *et al.* in the UK Biobank (see section [Materials and Methods](#)) but partitioned the samples according to the UK Biobank derived variable 'White British' ancestry (combination of self-report and genetic principal component analysis). Overall, after quality control, we obtained 369 010 samples, including 11 738 asthmatics with childhood-onset and 357 272 controls. To avoid extreme imbalances, we restricted the set of controls to randomly chosen 50 000 controls. Then, we performed the association analysis of rs73367537 in two distinct groups defined by the British European ancestry variable. In this stratified analysis, we observed a significant risk effect of rs73367537-T, consistent with the effect in GACRS, in samples with British European ancestry (9702 cases, 41 437 controls, *P*=0.0014) and a suggestive significant protective effect of rs73367537-T in the remaining participants (2036 cases, 8563 controls, *P*=0.064). Accordingly, a standard test for equality of effects in both subsamples had a significant two-sided *P*-value of *P*=0.0012.

Furthermore, we also found additional support for the 10q26 association in independent Costa Rica data. In 2007, Celedón *et al.* performed a linkage analysis of asthma and airway responsiveness in eight large pedigrees from Costa Rica (21). These pedigrees partly overlap with GACRS, but the 894 GACRS trios in

our analysis were not included in the linkage analysis. Celedón *et al.* found evidence for linkage signals for asthma and airway responsiveness. One shared signal was located on Chromosome 10 with genetic coordinates between 145 and 151 cM (21). The lead variant on 10q26 in GACRS, rs73367537, is located in this genetic region. We ran an FBAT analysis of rs73367537 in all members of the corresponding extended pedigrees with physician's diagnosed asthma and WGS data available. The resulting *P*-value was *P* = 0.045 (13 informative FBAT families) with consistent risk effect direction as in the GACRS main analysis.

In summary, the UK Biobank analysis replicated both the association and observed heterogeneity of the 10q26 signal. Furthermore, the association was replicated in an independent part of GACRS.

Region-based association analysis

The sliding window approach constructed a total of 1331 478 unique regions. As described in the Material and Methods section, we evaluated the regions using a two-stage procedure.

In the first stage, we considered all regions that are located within a radius of 20 kb around one of the associations identified in the single variant analysis. This included the lead variants in the signal clusters on 4p14, 8p22-1, 8p22-2 and 10q26, as well as the lead variant on 17q21. We obtained a total number of 98 regions, with an average number of 27 variants (minimum number 2 and maximum number 50). Among these regions, 16 showed a significant FBAT ACAT association *P*-value after Bonferroni correction for *n* = 98 tests and a significance level of $\alpha = 0.05$, i.e. $P < \frac{0.05}{98} = 0.00051$. The 16 regions were located in the 8p22 and 10q26 loci. Since the single variant associations on 10q26 are low-frequency variants with MAF of around 3% themselves, we focused on the six regions on 8p22 ([Supplementary Material, Table S2](#)). We further analyzed the conditional significance of these regions ([Table 3](#)), conditioning on the lead variants in the signal clusters 8p22-1 and 8p22-2 identified in the single variant analysis. Most of the regions remained nominally significant, even though, in general, the conditional approach reduces power substantially due to robustness of the model-free FBAT approach. [Table 3](#) also reports the FBAT ACAT *P*-values for the 6 regions, based on GACRS and CAMP separately. In [Supplementary Material, Table S6](#), we report the results of FBAT haplotype analysis based on the two common genetic variants and all rare variants that had a nominal significant single variant FBAT additive association *P*-value in GACRS + CAMP, GACRS or CAMP. Overall, these results suggest that both common and low-frequency/rare variants play a role in the architecture of the 8p22 association with childhood-onset asthma and that the observed signals appear to be highly consistent between GACRS and CAMP. Therefore, our first stage analysis identified potential additional, independent rare/low-frequency variant associations on 8p22.

In the second stage, we considered the ACAT *P*-values for all 1331 478 regions. The quantile-quantile plots for the ACAT and the four sub-statistics is displayed in [Supplementary Material, Figure S8](#). None of these regions were significant after correction for multiple testing. We considered all suggestively significant regions with an FBAT ACAT *P*-value $P < 10^{-5}$. The corresponding regions were located on Chromosomes 1, 3, 16 and 17 and are described in [Supplementary Material, Table S3](#). [Supplementary Material, Table S5](#) contains additional information including annotations for all low-frequency/rare genetic variants with a *P*-value below 0.05 in either GACRS + CAMP or GACRS and CAMP individually, that are located within one of the described regions.

Table 3. Genetic regions in the 8p22 locus with significant FBAT ACAT association *P*-value in the first stage of testing, *P*-values shown are uncorrected for multiple testing (see [Supplementary Material, Table S2](#))

CHR	Start BP (hg38)	End BP (hg38)	Number of variants	ACAT	ACAT conditional on rs1581479	ACAT conditional on rs17515041	GACRS ACAT	CAMP ACAT
8	16 948 949	16 951 656	30	2.77e−04	9.31e−03	4.68e−03	2.43e−03	1.08e−01
8	16 949 866	16 953 451	39	2.03e−05	7.15e−03	1.21e−03	3.86e−04	2.12e−02
8	16 995 849	16 999 748	35	2.89e−04	9.53e−02	3.46e−02	1.42e−04	2.85e−02
8	17 003 908	17 007 616	29	5.27e−05	4.14e−02	3.00e−02	2.76e−03	1.17e−02
8	17 005 897	17 009 766	28	7.33e−05	1.01e−01	1.98e−02	1.18e−02	1.48e−02
8	17 012 081	17 015 804	36	3.13e−04	3.91e−02	4.33e−04	8.53e−04	4.76e−01

The table also reports the (uncorrected) conditional FBAT ACAT association *P*-values when conditioning on one of the common variants identified in the single variant analysis (Table 2) (*P*-values based on 1 000 000 simulations). The last two columns contain the FBAT ACAT (uncorrected) *P*-values for the regions using only GACRS and CAMP data, respectively (*P*-values based on 10 000 000 simulations). FBAT: Family-Based Association Test; ACAT: Aggregated Cauchy Association Test.

Overall, the second stage analysis did not reveal genome-wide significant associations.

Discussion

We performed a WGS-based association analysis based on two family-based studies and identified three novel genetic associations for childhood asthma.

Our family-based analysis has several advantages compared to recent large-scale case–control studies. First, the trio design provides robustness against population stratification, allowing a joint analysis of the combined data from GACRS and CAMP. Second, asthma status in our analysis is based on comparable definitions of asthma in both study populations, including a physician's diagnosis. This more rigorous definition was used instead of self-report or a simple questionnaire-based diagnosis of asthma used in other asthma GWAS. Third, the incorporation of different ethnicities and genetic ancestries provides potential insights into the generalizability of the results. Fourth, our association analysis is based on whole-genome sequencing data that allowed us to robustly identify low-frequency variant associations in the GACRS cohort, as well as potential low-frequency/rare variant signals on 8p22. The main limitation of our analysis is the relatively small sample size compared to large-scale association analyses based on unrelated samples, which limits the power, especially for the genome-wide region-based analysis of low-frequency and rare variants. Another limitation of the analysis is that our eQTL colocalization testing using the JLIM approach does not take into account functional annotations or experimental validation. The resolution level of the colocalization approach depends on the underlying LD structure, and the relatively small overall sample size makes it difficult to distinguish causal variants from tagged variants. Furthermore, eQTL studies and GWAS tend to be biased towards different types of variants, emphasizing careful interpretation of colocalization results (22). Therefore, future studies are needed to confirm the role of the proposed genes.

Besides the three novel genetic loci associated with childhood asthma, we confirmed 13 (15%) of previously identified asthma GWAS associations. Moreover, 70 out of 87 previously identified loci (80%) had a consistent direction of effect.

The study design and analysis strategy played a crucial role in the identification of the three novel loci. For the signal on 8p22, both studies had similar association signals for two independent common variants and together these signals reached genome-wide significance in the combined analysis. Furthermore, our region-based analysis suggested that there may be additional, independent low-frequency/rare variant associations in this region. The 4p14 association reached genome-wide significance

in the recessive model. Finally, the association signal on 10q26 consists of low-frequency variants, enabled by the WGS data, and was observed in GACRS only and may have ancestry- or environment-specific effects.

The 4p14 locus was the most significant locus using the recessive association model. SNP rs13122762, part of the signal cluster, is located at the 5' upstream of *PGM2*. The signal cluster also contained clumped meQTL for cg20524452 and cg04038500, but no previous epigenome-wide association was reported for these CpG sites in the EWAS atlas (23). Furthermore, we found evidence for colocalization with eQTL for *PGM2* in several relevant tissues and datasets, including both lung tissues and immune cell types. *PGM2* encodes an enzyme implicated in glucose metabolism and catalyzes the conversion of the nucleoside breakdown products ribose-1-phosphate and deoxyribose-1-phosphate to the corresponding 5-phosphopentoses and the interconversion of glucose-1-phosphate to glucose-6-phosphate.

Recently, early embryonic lethality was reported in *Pgm2* knockout mice before E9.5 when lung development starts, suggesting important roles of *Pgm2* in embryonic development (24). Future studies in inducible and conditional knockout mice of *Pgm2* in specific lung cell types, especially in Ovalbumin (OVA)- or house dust mite-induced asthmatic models may pinpoint important cell types and possible molecular mechanisms of how *PGM2* may contribute to asthma susceptibility. Glucose metabolism was shown to be dysregulated in asthma development (25–27). It is conceivable that the expression levels of *PGM2*, influenced by genetic variants on 4p14, may impact asthma susceptibility possibly through fine-tuning the glucose metabolic status in immune cells, thereby priming epigenetic features of the genes and impacting the inflammatory response in airways in response to allergen or viral infection.

On 8p22, we found two common variant associations (i.e. two signal clusters), that were in weak LD. In addition, we observed potentially associated low-frequency/rare variants in close proximity to both common variants. However, the exact relationship of these rare variants to the two separate common variants is hard to explore because of the LD structure of the region and the limited sample size. The association was also supported by clumped meQTL associations for cg04451175 and cg06872257, which were reported to be associated with aging and gestational age (23). For the first signal cluster 8p22–1, we found no colocalization with eQTLs. For the second signal cluster, 8p22–2, gene expression colocalization analysis suggested *FGF20* as the gene of interest. *FGF20* belongs to the important fibroblast growth factor family of genes that have essential roles during embryonic lung development. We previously reported that the genetic variants nearby *FGF20* are linked with allergic asthma (28). *FGF20* binds to

its membrane receptor FGFR1, which activates the MAPK pathway (29) and the beta-catenin/Wnt pathway (30) to modulate cell proliferation and differentiation. Furthermore, previous findings (31) in *Fgf20* knockout mice demonstrated that the potential genetic interaction between *FGF20* and *SOX2* may determine the differentiation of progenitor cells in cortical hair cells during embryonic development. Given that *SOX2* is an important transcription factor that determines the differentiation of proximal airway epithelial progenitor cells during embryonic lung development, *FGF20* may also modulate the function of airway epithelial cells during fetal lung development or post-natally. Our study suggests that future characterization of the function of *FGF20* in both murine and cellular models relevant to asthma is clearly warranted.

The results for 10q26 are more difficult to interpret. This locus was statistically significant in GACRS, and was independently implicated in a linkage study based on independent samples from extended pedigrees of Costa Rican children with asthma. FBAT testing in this independent set also resulted in a significant association ($P = 0.045$). It was also replicated in the UK Biobank in British Europeans. The signal did not replicate in CAMP or the remaining UK Biobank participants but demonstrated heterogeneity of the effect. While there were no eQTL analyzed in this region, fine-mapped variants overlap with a distal, non-promoter DNase I hypersensitive site (DHS) that is correlated with a promoter DHS for *WDR11* (32,33). The low-frequency variants on 10q26 associated with childhood asthma in GACRS are also present with similar allele frequencies in CAMP and the UK Biobank, but the analyses indicate that the genetic effects on asthma differ. We hypothesize that the effects of the identified variants on 10q26 are modified by either a genetic or environmental factor.

The replication analysis provided mixed results. The associations on 4p14 and 8p22 did not replicate in the UK Biobank childhood asthma analysis by Ferreira *et al.* (3). We also investigated reported associations between the lead variants in both loci and other phenotypes, including asthma-related outcomes, in the GWAS catalog, FinnGen database, and publicly available UK Biobank summary statistics via open target genetics (34,35). No significant result was observed. Further analyses are needed to analyze the exact reasons for the lack of replication. Potential explanations include differences in phenotype definitions or modification of genetic risk by population-specific factors. Asthma status in GACRS and CAMP was directly assessed by a physician's diagnosis and not self-reported. We also note that the association on 10q26 did not replicate in the overall UK Biobank analysis but demonstrated significance and effect heterogeneity in the stratified analysis according to the 'White British' variable.

Our genome-wide region-based rare variant analysis did not identify any significant signals. Suggestive signals include a region on 3p26 that consists of intronic variants for the gene *SUMF1*, which has been hypothesized to be associated with COPD (36). A second region on 16q12 contains an intronic splice site variant for *ABCC11* with a Combined Annotation Dependent Depletion (CADD) score (37) above 24.

Our study design has many strengths, but the major weakness is that we lack a large enough sample size to identify rare variant associations for asthma in genome-wide analyses. Also, as mentioned above, the eQTL colocalization results are not functionally validated. This will be subject of future studies. However, our disease definition and trio design set a high standard for asthma genetic association studies, and we have adequate power to detect common variants.

Materials and Methods

Study populations

Our analysis is based on data from two studies with asthmatic children: 'The Genetic Epidemiology of Asthma in Costa Rica' study (GACRS) and the Childhood Asthma Management Program (CAMP). GACRS is a cohort of children between 6 and 14 years of age with physician-diagnosed asthma and at least two episodes of respiratory symptoms (wheezing, cough or dyspnea) or a history of asthma attacks in the previous year (6). Additional criteria were that the children have at least six great-grandparents born in the Central Valley of Costa Rica. The population of the Central Valley of Costa Rica is a genetic isolate with one of the highest prevalences of asthma worldwide (approximately 27% in children) (38). For further details about enrollment and phenotyping protocols, we refer to previous publications (6).

CAMP was a multicenter clinical trial designed to determine the long-term effects of three inhaled treatments for childhood asthma (7). This cohort consists mainly of trios and a similar age distribution as GACRS. CAMP only included children with mild to moderate persistent asthma, as defined by the presence of symptoms or by the use of an inhaled bronchodilator at least twice weekly or the use of daily medication for asthma (39). Children with too many symptoms during the run-in period, such that they could not be taken off medication, were excluded. Additional exclusion criteria included severe chronic sinusitis or nasal polyposis, use of more than four sprays of nasal steroids daily (only beclomethasone allowed), current use of metoclopramide, ranitidine or cimetidine treatment for gastroesophageal reflux or evidence of severe asthma. GACRS and CAMP used similar protocols for phenotyping subjects.

Ethics

Written parental consent and participating child's assent were obtained. For GACRS data, the study was approved by the Mass General Brigham Human Research Committee at Brigham and Women's Hospital (Boston, MA; protocol No. 2000P001130) and the Hospital Nacional de Niños (San José, Costa Rica).

For CAMP data, all study procedures were approved by the Mass General Brigham Human Research Committee at Brigham and Women's Hospital (Boston, MA; protocol No. 2011P000710).

Whole-genome sequencing and quality control

Whole-genome sequencing data for GACRS and CAMP were generated as part of the National Heart, Lung and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) program (8). Details on DNA sample handling, quality control, library construction, clustering and sequencing, read processing and sequence data quality control are described on the TOPMed website (<https://topmed.nhlbi.nih.gov/topmed-whole-genome-sequencing-methods-freeze-8>). Variant calls were obtained from TOPMed data freeze 8 with variant call format files aligned to the GRCh38 genome reference. We extracted biallelic SNPs and insertions/deletions that passed quality control, only considering individual variant information with at least 10x sequencing depth for the analyses. Based on genome-wide identity-by-descent estimates generated by kinship-based inference for GWAS (KING) (40) and additional information from the TOPMed Data Coordinating Center, we identified and removed genetic duplicates, pedigree discrepancies and sex mismatches. After this step, we kept complete trios for the family-based association analysis.

Single variant association analysis

Based on the trio data, we performed single variant family-based association tests (FBAT v204) using additive and recessive genetic models (9,10). Our primary analysis was based on the combined GACRS and CAMP dataset (GACRS + CAMP), but we also considered GACRS and CAMP separately. The advantage of the FBAT approach is that the association test is robust to population stratification and admixture (41), allowing the combination of both studies. For the single variant analysis, we excluded genetic variants with a minor allele frequency (MAF) below 1%, more than four Mendelian errors, genotype missing rate > 2%, TOPMed freeze 8 reported batch effects or deviations from Hardy–Weinberg proportions ($P < 10^{-8}$). Furthermore, we only considered association tests with at least 10 informative families to ensure valid asymptotic inference. Testing based on an additive genetic model provides one association P -value per variant, whereas testing based on the recessive genetic model can be performed based on the minor allele as well as based on the major allele. We performed recessive tests on the minor allele for all variants where this test had at least 10 informative families. For the recessive tests on the major alleles, however, we restricted testing to variants with a major allele frequency of at most 60%. The corresponding quality control computations were performed using PLINK (version v1.90b6.12 and v2.00a2.3LM). The MAFs and Hardy–Weinberg deviations were assessed based on parental genotype data. We considered only autosomal genetic data in our analyses. We declared association tests with P -values $P < 5 \times 10^{-8}$ as genome-wide significant. Since the FBAT statistic equals the TDT for asthmatic trios, we used existing power formulas for the TDT to investigate the expected power of our combined GACRS + CAMP analysis based on 1180 trios (15).

Replication of previously reported associations for childhood asthma

We downloaded all genetic associations from the Ferreira *et al.* analysis of childhood-onset asthma from the NHGRI-EBI GWAS catalog (3,17) (Study accession GCST007800, June 18, 2021). Among the 105 reported genetic variants, 103 were autosomal and therefore potentially overlapping with our analysis. Further, 87 out of the 103 variants (84%) were available in the GACRS + CAMP after the described quality control procedure. We considered the corresponding additive association P -values and direction of effect for comparison.

Statistical fine mapping

Due to the high variant density in the WGS data, signal clusters contain a large number of variants in strong linkage disequilibrium (LD). To identify potentially causal genetic variants, we performed statistical fine-mapping for all regions with significant single variant associations with an association P -value $P < 5 \times 10^{-8}$. For the fine mapping, we included all variants within a window of ± 100 kb around the lead association. Since the FBAT tests are score test statistics, we obtain association z -scores but no effect estimates with corresponding standard errors. This motivated using the DAP-G tool (42) for fine mapping since only z -scores and the corresponding LD/correlation matrices are required as input. The LD matrices were estimated based on parental WGS data. For associations discovered using the recessive model, we replaced the standard LD matrix with an empirical correlation matrix based on the parental genotype data and a recessive coding. The *ld_control* parameter was set to 0.25. We considered all signal clusters, groups of genetic variants in strong LD, with a

DAP-G posterior inclusion probability of at least 50% for further analyses.

eQTL colocalization analysis

We tested for colocalization between the identified genetic associations with asthma ('primary trait') and eQTLs ('secondary trait') using the joint likelihood mapping (JLIM) approach (43). We used the FBAT association P -values within a region of ± 50 kb around the lead variant for the primary trait (in the presence of two associations within the same loci, we used a smaller radius). For the secondary trait, we used the eQTL association P -values of the corresponding intervals from the following eQTL datasets: all immune cell types in the eQTL Catalogue (44), all tissues in the Genotype Tissue Expression project (GTEx; version 8) (45) and a study of *cis*-eQTL in developing human fetal lung (sample superset of dataset previously described (46)). For GTEx tissues, we used the eQTL association statistics calculated only with subjects of European ancestry, available for download from the GTEx consortium. In the case of fetal lung eQTLs, we used the full multi-ethnic eQTL data since the underlying individual-level genotype data were available. In addition to the association statistics, JLIM requires the LD matrices of primary and secondary traits to account for the correlation of association statistics between SNPs. For the additive model with the primary trait (asthma), LD matrices were derived directly from parental WGS genotype. Considering the secondary trait (eQTL), LD matrices were calculated from the genotypes of non-Finnish European ancestry ($n = 404$; 1000 Genomes Project phase 3 (47)) for the eQTL Catalogue and GTEx; and from the genotypes of all post-QC multi-ethnic subjects ($n = 386$) for fetal lung eQTLs. For the recessive model, we treated the alleles of each SNP as if they were distinct SNPs. Specifically, for the primary trait, the empirical correlation of parental genotype data based on a recessive coding was used. For the secondary trait, the LD between alleles of the same variants was set to 1, and the LD between alleles of different variants was set to the underlying LD between the SNPs.

For all colocalization analyses, we examined all genes of which transcription start sites are within 1 Mbps from the FBAT lead variant. Due to low imputation accuracies, especially in datasets with non-European genetic ancestry, Indels, genetic variants with $MAF < 5\%$, and tri-allelic SNPs were filtered out from the analyses. JLIM was applied in the default setting with a maximum adaptive permutation of 100 000 iterations and outputs a colocalization P -value for each gene. The FDR was calculated separately for each tissue/cell type and locus using the Benjamini–Hochberg procedure (48). All colocalizations with FDR below 5% were considered to be significant.

Overlap with methylation quantitative trait loci results in recent large-scale analysis

We used results from a recent large-scale DNA methylation quantitative trait loci (meQTLs) analysis to investigate the overlap between the identified signal clusters in the DAP-G fine mapping analysis and meQTL associations in blood (20). We extracted the clumped meQTL results from this analysis, using the reference SNP ID (rsID) as the identifier. Clumping corresponds to linkage disequilibrium (LD) clumping, aiming to extract the most significant meQTL while excluding other meQTL associations in LD.

Region-based association analysis

To test for an association between low-frequency/rare genetic variants and childhood asthma, we applied the recently published framework for region-based association analysis in family-based

designs (14). This framework evaluates different association test statistics (Burden, SKAT (49,50), Higher Criticism (51) and maximum single variant (51)) and computes an overall association P-value using the Aggregated Cauchy Association Test (ACAT) statistic (52). The evaluation of the individual sub-statistics is based on simulations and does not require asymptotic theory. For the region-based analysis, we considered genetic variants with MAF below 5% and at least five informative families (single variant FBAT, additive model). Also, we excluded variants with more than two Mendelian errors, variant missing rate > 2%, TOPMed freeze 8 reported batch effects or deviations from Hardy–Weinberg proportions ($P < 10^{-8}$). All region-based test statistics were computed based on an additive genetic model in GACRS+CAMP. The number of simulations for the P-value computation is chosen adaptively and the maximum number of simulations was set to 10 000 000.

We partitioned all genetic variants into regions by using a sliding window approach with a size of 4 kb (but at most 50 genetic variants) and with a shift of 2 kb. For the evaluation of significance, we considered a two-stage approach. First, we evaluated all regions in a ± 20 kb radius around a significant association identified in the single variant analysis. Here, we used a Bonferroni correction corresponding to the number of regions in this first stage. In the second stage, we considered all remaining regions. Additionally, for significant regions in the first stage, we also performed a conditional test, where we conditioned on the corresponding single variant association to confirm the independence of the potential observed low-frequency/rare variant signal. All region-based computations were performed using FBAT v204 and v208 (see section code and data availability).

Annotation of genetic variants, region association plots and Miami plot

We used two different resources to annotate potential causal variants to genes and functional categories: SNPnexus (53–57) and the combined SNP-to-gene (cS2G) strategy (58). SNPnexus combines different external resources to assist in selecting functionally relevant genetic variants and is accessible using a web interface (see section code and data availability). The cS2G strategy includes seven constituent SNP-to-gene strategies (Exon, Promoter, two fine-mapped cis-eQTL strategies, EpiMap enhancer-gene linking, Activity-By-Contact (ABC) and Cicero), and scores can be downloaded (see section code and data availability). Region association plots were created using the web application of LocusZoom (59) (see section code and data availability). The EUR LD information for the LocusZoom plots was used. Region association plots for the recessive association data were created using internal R code (see section code and data availability). The Miami plot for the single variant additive and recessive analysis was generated using the R package *miamiplot* (see section code and data availability).

Replication analysis: UK Biobank

We used the summary statistics from the UK Biobank analysis of childhood asthma (3) as a replication dataset (see section code and data availability). In addition, we used the individual genotype data in the UK Biobank (60) to perform more specific stratified analyses using a similar approach as conducted by Ferreira *et al.*

In particular, we kept only individuals with no deviations between self-reported and genetically inferred sex, no putative sex chromosome aneuploidies, no excess relatives, that are no outliers for heterozygosity or missing rate, and that have available genetic principal component data. Based on this subset of individuals, we identified individuals as asthmatics with childhood-onset using self-reported asthma onset before 16 years

of age (field 3786, ‘What was your age when the asthma was first diagnosed?’) and no difference of more than 10 years between the self-reported onset and the age where asthma was diagnosed by a physician (field 22 147, ‘age at which doctor diagnosed asthma’), if information was available. We defined individuals as controls that had no self-reported asthma, no physician-diagnosed asthma, and no asthma indicated by field 6152 or 20002. Asthma status was regressed on the imputed genotype data using logistic regressions adjusting for the first 10 principal components, genotyping array and sex. For the stratified analysis, we extracted the UK Biobank variable ‘in.white.British.ancestry.subset’ (data field 22 006), which represents a combination of self-reported information and genetic principal component analysis.

Code and data availability

Summary statistics for the single variant analyses: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001974. The FBAT software (v204 and v208) is available at <https://sites.google.com/view/fbatwebpage>. PLINK (version v1.90b6.12 and v2.00a2.3LM) was downloaded from <https://www.cog-genomics.org/plink/>. DAP-G (v1.0.0) was downloaded from <https://github.com/xqwen/dap>. JLIM (version 2.5) is available at <https://github.com/cotsapaslab/jlim>. C2G scores are available at <https://alkesgroup.broadinstitute.org/cS2G>. SNPnexus is available at <https://www.snp-nexus.org/v4/>. Region association plots were created using <https://my.locuszoom.org/> (v0.14.0). Region association plots for recessive association data was performed using an adaption of the R code available at <https://github.com/Geeketics/LocusZooms>. UK Biobank website (<https://www.ukbiobank.ac.uk/>). Data processing and plots were created using R (version 4.1.0) (<https://www.r-project.org/>). Summary statistics for the childhood-onset analysis by Ferreira *et al.* (3) are available at https://genepi.qimr.edu.au/staff/manuelF/gwas_results/main.html. The LDpair tool is available at <https://ldlink.nci.nih.gov/?tab=ldpair>. The meQTL analysis results were assessed using <http://api.godmc.org.uk/v0.1> (20). Miami plot was generated using the R package *miamiplot* (<https://github.com/juliedwhite/miamiplot>). The winnerscurse R package is available at <https://rdrr.io/github/amandaforde/winnerscurse/>.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

This research was conducted by using the UK Biobank resource under application number 20915. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. A full list of authors for the NHLBI TOPMed Consortium is provided at <https://www.nhlbiwgs.org/topmed-banner-authorship>.

Conflict of Interest statement: Dr Weiss reports personal fees from UpToDate, outside the submitted work, and Dr Weiss is on the Scientific Board of Histolix; Dr Silverman received grant support from GlaxoSmithKline and Bayer, outside the submitted work; Dr Cho has received grant funding from GSK and Bayer and speaking or consulting fees from AstraZeneca, Illumina and Genentech, outside the submitted work; Dr Celedón has received research materials from GSK and Merck (inhaled steroids) and Pharmavite (vitamin D and placebo capsules) in order to provide medications free of cost to participants in NIH-funded studies, unrelated to this

manuscript. Dr DeMeo has received honoraria from Novartis and grant support from Bayer, outside the submitted work. Dr Lasky-Su is a Scientific Advisor to Precion.

Funding

National Heart, Lung, and Blood Institute [U01HL089856, U01HL089897, P01HL120839, P01HL132825, K99HL159234, R01HL123915, R01HL141826, R01HL155742, R01HL125734], National Institute of Mental Health [R01MH129337] and the National Human Genome Research Institute [R01HG008976, 2U01HG008685]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute, the National Human Genome Research Institute, the National Institute of Mental Health, or the National Institutes of Health.

Molecular data for the Trans-Omics for Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Genome Sequencing for 'NHLBI TOPMed: The Genetic Epidemiology of Asthma in Costa Rica' (phs000988.v4.p1) was performed at Northwest Genomics Center (HHSN268201600032I, 3R37HL066289-13S1). Genome Sequencing for 'NHLBI TOPMed: Childhood Asthma Management Program (CAMP)' (phs001726.v1.p1) was performed at the Northwest Genomics Center (HHSN268201600032I). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I).

References

- Soriano, J.B., Abajobir, A.A., Abate, K.H., Abera, S.F., Agrawal, A., Ahmed, M.B., Aichour, A.N., Aichour, I., Aichour, M.T.E., Alam, K. et al. (2017) Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir. Med.*, **5**, 691–706.
- Carr, T.F. and Bleecker, E. (2016) Asthma heterogeneity and severity. *World Allergy Organ. J.*, **9**, 41.
- Ferreira, M.A.R., Mathur, R., Vonk, J.M., Szwajda, A., Brumpton, B., Granell, R., Brew, B.K., Ullemar, V., Lu, Y., Jiang, Y. et al. (2019) Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am. J. Hum. Genet.*, **104**, 665–684.
- Pividori, M., Schoettler, N., Nicolae, D.L., Ober, C. and Im, H.K. (2019) Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir. Med.*, **7**, 509–522.
- Lai, C.K.W., Beasley, R., Crane, J., Foliaki, S., Shah, J., Weiland, S. and Group, the I.P.T.S (2009) Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax*, **64**, 476–483.
- Hunninghake, G.M., Soto-Quiros, M.E., Avila, L., Ly, N.P., Liang, C., Sylvia, J.S., Klanderma, B.J., Silverman, E.K. and Celedón, J.C. (2007) Sensitization to *Ascaris lumbricoides* and severity of childhood asthma in Costa Rica. *J. Allergy Clin. Immunol.*, **119**, 654–661.
- (1999) The Childhood Asthma Management Program (CAMP): design, rationale, and methods. *Control. Clin. Trials*, **20**, 91–120.
- Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M. et al. (2021) Sequencing of 53 831 diverse genomes from the NHLBI TOPMed program. *Nature*, **590**, 290–299.
- Lake, S.L., Blacker, D. and Laird, N.M. (2000) Family-based tests of association in the presence of linkage. *Am. J. Hum. Genet.*, **67**, 1515–1525.
- Laird, N.M., Horvath, S. and Xu, X. (2000) Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.*, **19**, S36–S42.
- Guindo-Martínez, M., Amela, R., Bonàs-Guarch, S., Puiggròs, M., Salvaro, C., Miguel-Escalada, I., Carey, C.E., Cole, J.B., Rüeger, S., Atkinson, E. et al. (2021) The impact of non-additive genetic associations on age-related complex diseases. *Nat. Commun.*, **12**, 2436.
- O'Connor, M.J., Schroeder, P., Huerta-Chagoya, A., Cortés-Sánchez, P., Bonàs-Guarch, S., Guindo-Martínez, M., Cole, J.B., Kaur, V., Torrents, D., Veerapen, K. et al. (2022) Recessive genome-wide meta-analysis illuminates genetic architecture of type 2 diabetes. *Diabetes*, **71**, 554–565.
- Hecker, J., Xu, X., Townes, F.W., Loehlein Fier, H., Corcoran, C., Laird, N. and Lange, C. (2018) Family-based tests for associating haplotypes with general phenotype data: Improving the FBAT-haplotype algorithm. *Genet. Epidemiol.*, **42**, 123–126.
- Hecker, J., William Townes, F., Kachroo, P., Laurie, C., Lasky-Su, J., Ziniti, J., Cho, M.H., Weiss, S.T., Laird, N.M. and Lange, C. (2020) A unifying framework for rare variant association testing in family-based designs, including higher criticism approaches, SKATs, and burden tests. *Bioinformatics*, **36**, 5432–5438.
- Deng, H.W. and Chen, W.M. (2001) The power of the transmission disequilibrium test (TDT) with both case-parent and control-parent trios. *Genet. Res.*, **78**, 289–302.
- Stein, M.M., Thompson, E.E., Schoettler, N., Helling, B.A., Mag-naye, K.M., Stanhope, C., Igartua, C., Morin, A., Washington, C., Nicolae, D. et al. (2018) A decade of research on the 17q12-21 asthma locus: piecing together the puzzle. *J. Allergy Clin. Immunol.*, **142**, 749–764.e3.
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J. et al. (2017) The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res.*, **45**, D896–D901.
- Ferguson, J.P., Cho, J.H., Yang, C. and Zhao, H. (2013) Empirical bayes correction for the Winner's Curse in genetic association studies. *Genet. Epidemiol.*, **37**, 60–68.
- Palmer, C. and Pe'er, I. (2017) Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.*, **13**, e1006916.
- Mín, J.L., Hemani, G., Hannon, E., Dekkers, K.F., Castillo-Fernandez, J., Luijk, R., Carnero-Montoro, E., Lawson, D.J., Burrows, K., Suderman, M. et al. (2021) Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.*, **53**, 1311–1321.
- Celedón, J.C., Soto-Quiros, M.E., Avila, L., Lake, S.L., Liang, C., Fournier, E., Spesny, M., Hersh, C.P., Sylvia, J.S., Hudson, T.J. et al. (2007) Significant linkage to airway responsiveness on chromosome 12q24 in families of children with asthma in Costa Rica. *Hum. Genet.*, **120**, 691–699.
- Mostafavi, H., Spence, J.P., Naqvi, S. and Pritchard, J.K. (2022) Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv*, 2022.05.07.491045.
- Li, M., Zou, D., Li, Z., Gao, R., Sang, J., Zhang, Y., Li, R., Xia, L., Zhang, T., Niu, G. et al. (2019) EWAS Atlas: a curated knowledgebase of

- epigenome-wide association studies. *Nucleic Acids Res.*, **47**, D983–D988.
24. Balakrishnan, B., Verheijen, J., Lupo, A., Raymond, K., Turgeon, C., Yang, Y., Carter, K., Whitehead, K., Kozicz, T., Morava, E. et al. (2019) A novel phosphoglucomutase-deficient mouse model reveals aberrant glycosylation and early embryonic lethality. *J. Inherit. Metab. Dis.*, **42**, 998–1007.
 25. Park, S., Choi, N.-K., Kim, S. and Lee, C.-H. (2018) The relationship between metabolic syndrome and asthma in the elderly. *Sci. Rep.*, **8**, 9378.
 26. Perez, M.K. and Piedimonte, G. (2014) Metabolic asthma: is there a link between obesity, diabetes, and asthma? *Immunol. Allergy Clin. N. Am.*, **34**, 777–784.
 27. Peters, M. C., Schiebler, M., Bleecker, E. R., Castro, M., Erzurum, S. C., Hastie, A. T., Levy, B. D., Cardet, J. C., Israel, E., Phillips, B. et al. (2019) Impaired glucose metabolism is associated with asthma severity in the severe asthma research program-3 (SARP-3) cohort. A31. *Asthma: Clinical Studies I, American Thoracic Society International Conference Abstracts, American Thoracic Society*, pp. A1280–A1280.
 28. Bunyavanich, S., Schadt, E.E., Himes, B.E., Lasky-Su, J., Qiu, W., Lazarus, R., Ziniti, J.P., Cohain, A., Linderman, M., Torgerson, D.G. et al. (2014) Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC Med. Genet.*, **7**, 48.
 29. Su, Y., Yang, L.M. and Ornitz, D.M. (2021) FGF20-FGFR1 signaling through MAPK and PI3K controls sensory progenitor differentiation in the organ of Corti. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.*, **250**, 134–144.
 30. Cai, X., Tao, W. and Li, L. (2022) Glioma cell-derived FGF20 suppresses macrophage function by activating β -catenin. *Cell. Signal.*, **89**, 110181.
 31. Yang, L.M., Cheah, K.S.E., Huh, S.-H. and Ornitz, D.M. (2019) Sox2 and FGF20 interact to regulate organ of Corti hair cell and supporting cell development in a spatially-graded manner. *PLoS Genet.*, **15**, e1008254.
 32. Kim, Y.-J., Osborn, D.P., Lee, J.-Y., Araki, M., Araki, K., Mohun, T., Käsäkoski, J., Brandstack, N., Kim, H.-T., Miralles, F. et al. (2018) WDR11-mediated Hedgehog signalling defects underlie a new ciliopathy related to Kallmann syndrome. *EMBO Rep.*, **19**, 269–289.
 33. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. et al. (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.
 34. Mountjoy, E., Schmidt, E.M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., Fumis, L., Hayhurst, J., Buniello, A., Karim, M.A. et al. (2021) An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.*, **53**, 1527–1533.
 35. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A. et al. (2021) Open targets genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.*, **49**, D1311–D1320.
 36. Weidner, J., Jarenbäck, L., de Jong, K., Vonk, J.M., van den Berge, M., Brandsma, C.-A., Boezen, H.M., Sin, D., Bossé, Y., Nickle, D. et al. (2017) Sulfatase modifying factor 1 (SUMF1) is associated with chronic obstructive pulmonary disease. *Respir. Res.*, **18**, 77.
 37. Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M. and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
 38. Pearce, N., Ait-Khaled, N., Beasley, R., Mallol, J., Keil, U., Mitchell, E. and Robertson, C. (2007) Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax*, **62**, 758–766.
 39. The Childhood Asthma Management Program Research Group (2000) Long-term effects of budesonide or nedocromil in children with asthma. *N. Engl. J. Med.*, **343**, 1054–1063.
 40. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. and Chen, W.-M. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics*, **26**, 2867–2873.
 41. Laird, N.M. and Lange, C. (2006) Family-based designs in the age of large-scale gene-association studies. *Nat. Rev. Genet.*, **7**, 385–394.
 42. Wen, X., Lee, Y., Luca, F. and Pique-Regi, R. (2016) Efficient integrative multi-SNP association analysis via deterministic approximation of posteriors. *Am. J. Hum. Genet.*, **98**, 1114–1129.
 43. Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R. and Cotsapas, C. (2017) Limited statistical evidence for shared genetic effects of eQTLs and autoimmune disease-associated loci in three major immune cell types. *Nat. Genet.*, **49**, 600–605.
 44. Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J. et al. (2021) A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.*, **53**, 1290–1299.
 45. The GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
 46. Kho, A.T., Chhabra, D., Sharma, S., Qiu, W., Carey, V.J., Gaedigk, R., Vyhldal, C.A., Leeder, J.S., Tantisira, K.G. and Weiss, S.T. (2016) Age, sexual dimorphism, and disease associations in the developing human fetal lung transcriptome. *Am. J. Respir. Cell Mol. Biol.*, **54**, 814–821.
 47. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1092 human genomes. *Nature*, **491**, 56–65.
 48. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.
 49. Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J.D. and Lin, X. (2013) Family-based association tests for sequence data, and comparisons with population-based association tests. *Eur. J. Hum. Genet.*, **21**, 1158–1162.
 50. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, **89**, 82–93.
 51. Mukherjee, R., Pillai, N.S. and Lin, X. (2015) Hypothesis testing for high-dimensional sparse binary regression. *Ann. Stat.*, **43**, 352–381.
 52. Liu, Y. and Xie, J. (2019) Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.*, **115**, 393–402.
 53. Chelala, C., Khan, A. and Lemoine, N.R. (2009) SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*, **25**, 655–661.
 54. Dayem Ullah, A.Z., Lemoine, N.R. and Chelala, C. (2012) SNPnexus: a web server for functional annotation of novel and

- publicly known genetic variants (2012 update). *Nucleic Acids Res.*, **40**, W65–W70.
55. Dayem Ullah, A.Z., Lemoine, N.R. and Chelala, C. (2013) A practical guide for the functional annotation of genetic variations using SNPnexus. *Brief. Bioinform.*, **14**, 437–447.
56. Dayem Ullah, A.Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N.R. and Chelala, C. (2018) SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.*, **46**, W109–W113.
57. Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A.Z., Lemoine, N.R. and Chelala, C. (2020) SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res.*, **48**, W185–W192.
58. Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K.K., Nasser, J., Jagadeesh, K.A., Weiner, D.J., Shi, H., Fulco, C.P., O'Connor, L.J. et al. (2022) Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat. Genet.*, **54**, 827–836.
59. Boughton, A.P., Welch, R.P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G.R. and Boehnke, M. (2021) LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics*, **37**, 3017–3018.
60. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.