A meta-analysis of genome-wide association studies of asthma in Puerto Ricans

Qi Yan[1], John Brehm[1], Maria Pino-Yanes[2,3], Erick Forno[1], Jerome Lin[4], Sam S. Oh[5], Edna Acosta-Perez[6], Cathy C. Laurie[7], Michelle M. Cloutier[8], Benjamin A. Raby[9], Adrienne M. Stilp[7], Tamar Sofer[7], Donglei Hu[5], Scott Huntsman[5], Celeste S. Eng[5], Matthew P. Conomos[7], Deepa Rastogi[10], Kenneth Rice[7], Glorisa Canino[6], Wei Chen[1], R. Graham Barr[11,13], Esteban G. Burchard[5,12,13] and Juan C. Celedón[1,13]

Affiliations: [1]Division of Pediatric Pulmonary Medicine, Allergy, and Immunology, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA, USA. [2]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain. [3]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain. [4]Dept of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [5]Dept of Medicine, University of California San Francisco, San Francisco, CA, USA. [6]Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, Puerto Rico. [7]Dept of Biostatistics, University of Washington, Seattle, WA, USA. [8]Dept of Pediatrics, University of Connecticut, Farmington, CT, USA. [9]Channing Division of Network Medicine, Dept of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [10]Dept of Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA. [11]Dept of Epidemiology, Columbia University, New York, NY, USA. [12]Dept of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. [13]These authors contributed equally to this work.

Correspondence: Juan C. Celedón, Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of UPMC, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. E-mail: juan.celedon@chp.edu

@ERSpublications
Common allelic variants on chromosome 17q21 have the greatest effects on asthma in Puerto Ricans, a high-risk group http://ow.ly/OtZq3084XV1

Cite this article as: Yan Q, Brehm J, Pino-Yanes M, *et al.* A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur Respir J* 2017; 49: 1601505 [https://doi.org/10.1183/13993003.01505-2016].

ABSTRACT Puerto Ricans are disproportionately affected with asthma in the USA. In this study, we aim to identify genetic variants that confer susceptibility to asthma in Puerto Ricans.

We conducted a meta-analysis of genome-wide association studies (GWAS) of asthma in Puerto Ricans, including participants from: the Genetics of Asthma in Latino Americans (GALA) I-II, the Hartford–Puerto Rico Study and the Hispanic Community Health Study. Moreover, we examined whether susceptibility loci identified in previous meta-analyses of GWAS are associated with asthma in Puerto Ricans.

The only locus to achieve genome-wide significance was chromosome 17q21, as evidenced by our top single nucleotide polymorphism (SNP), rs907092 (OR 0.71, $p=1.2\times10^{-12}$) at *IKZF3*. Similar to results in non-Puerto Ricans, SNPs in genes in the same linkage disequilibrium block as *IKZF3* (*e.g. ZPBP2*, *ORMDL3* and *GSDMB*) were significantly associated with asthma in Puerto Ricans. With regard to results from a meta-analysis in Europeans, we replicated findings for rs2305480 at *GSDMB*, but not for SNPs in any other genes. On the other hand, we replicated results from a meta-analysis of North American populations for SNPs at *IL1RL1*, *TSLP* and *GSDMB* but not for *IL33*.

Our findings suggest that common variants on chromosome 17q21 have the greatest effects on asthma in Puerto Ricans.

This article has supplementary material available from erj.ersjournals.com

Received: July 27 2016 | Accepted after revision: Dec 18 2016

Introduction

Asthma is a disease with substantial heritability [1, 2]. Genome-wide association studies (GWAS) have identified several susceptibility loci for asthma, including the chromosome 17q21 region [3, 4]. This locus, which contains genes *IKZF3*, *ZPBP2*, *GSDMB* and *ORMDL3*, has been consistently replicated across diverse ethnic groups [3–13].

The burden of asthma varies among racial or ethnic groups. In the USA, current asthma is more common in Puerto Ricans (16.1%) than in non-Hispanic black people (11.2%), non-Hispanic white people (7.7%) or Mexicans (5.4%) [14]. Moreover, Puerto Ricans have greater morbidity from asthma than members of other racial or ethnic groups [15], and thus studying asthma in this ethnic group is relevant to public health and understanding disease pathogenesis. Although two previous GWAS of asthma included Puerto Ricans who participated in the Genetics of Asthma in Latino Americans Study (GALA I) [3] and in the Genes–Environments and Admixture in Latino Americans Study (GALA II) [13], no separate analysis was presented for Puerto Ricans.

We conducted a meta-analysis of GWAS of asthma, including only Puerto Rican participants from: GALA I, GALA II, the Hartford–Puerto Rico Study (Hartford–PR) [16] and the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) [17, 18]. Moreover, we examined whether susceptibility loci identified in two previous meta-analyses of GWAS of asthma (conducted by the GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community) and EVE (A Study to Identify Asthma-susceptibility Genes in Ethnically Diverse Populations) consortia [3, 4]) are associated with asthma in Puerto Ricans.

Methods

Study populations

Hartford–PR

Children were recruited in Hartford (CT, USA) (n=449) and San Juan (Puerto Rico) (n=678), as reported elsewhere [16]. At both study sites, the main recruitment tool was a screening questionnaire given to parents of children aged 6 to 14 years. All participants had to have four Puerto Rican grandparents. Asthma was defined as physician-diagnosed asthma and at least one episode of wheeze in the prior year. Control subjects had neither physician-diagnosed asthma nor wheeze in the prior year. Genome-wide genotyping was conducted using the HumanOmni2.5 BeadChip platform (Illumina Inc., San Diego, CA, USA), as previously described [16]. Imputation of non-genotyped single nucleotide polymorphisms (SNPs) was performed with IMPUTE2 [19], using data from the Phase I (November 2010 release) of the 1000 Human Genomes Project as the reference panel. After quality control measures, 948 children (523 with asthma) and ∼7 million genotyped and imputed SNPs were included in the GWAS. This analysis was conducted using logistic regression under an additive genetic model, adjusting for age, sex, study site and principal components calculated using smartPCA [20]. Written parental consent was obtained for participating children, from whom written assent was also obtained. The study was approved by the Institutional Review Boards of the University of Puerto Rico (San Juan, Puerto Rico), Brigham and Women's Hospital (Boston, MA, USA) and the University of Pittsburgh (Pittsburgh, PA, USA).

Support statement: This work was supported by grants HL079966 and HL117191 from the US National Institutes of Health (NIH), and by The Heinz Endowments. Q. Yan's contribution was supported by grant from Children's Hospital of Pittsburgh of UPMC. E. Forno's contribution was supported by grant HL125666 from the US NIH. D. Rastogi's contribution is supported by grant HL118733 from the NIH. The GALA II study was supported by US NIH grants to E. G. Burchard: HL088133, HL004464, HL117004, ES015794, ES24844, TRDRP 24RT 0025, MD006902, and GM007546. E.G. Burchard was also supported by the RWJF Amos Medical Faculty Development Award, the American Asthma Foundation and the Sandler Foundation. M. Pino-Yanes was supported by grant AC15/00015 from the Instituto de Salud Carlos III within the ERACoSysMed 1st Joint Transnational Call (SysPharmPedia 99) from the European Union, under the Horizon 2020. The HCHS/SOL was supported by NIH contracts N01-HC65233, N01-HC65234, N01-HC65235, N01-HC65236, N01-HC65237, and HHSN268201300005C AM03. The baseline examination of HCHS/SOL was carried out as a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contributed to the first phase of HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

GALA I

This study comprised subjects with and without asthma, recruited mainly from Puerto Rico, but also from New York City (NY, USA) [21, 22]. Subjects were included in the study if they were aged 8 to 40 years and self-identified all four grandparents as Puerto Rican. Cases had physician-diagnosed asthma and had experienced two or more symptoms in the previous two years (wheezing, coughing and/or shortness of breath). Control subjects had no symptoms of asthma or allergies, had no physician-diagnosis of asthma and no history of other chronic respiratory illness or allergies (eczema, hives or hay fever). All subjects were genotyped on the Affymetrix 6.0 GeneChip (Affymetrix, Santa Clara, CA, USA), and quality control was performed as described elsewhere [23]. Genotyped data was phased with SHAPE-IT [24] and imputation was performed with IMPUTE2 [19], using all populations from 1000 Genomes Project Phase I v3 [25] as reference. Association testing was conducted using logistic regression under an additive genetic model, adjusting for age, sex, and African and Native American ancestry estimates (calculated using ADMIXTURE [26]). A total of 437 unrelated subjects (251 with asthma) with complete data for all covariates were used in the current analysis. Written parental consent was obtained for participating children, from whom written assent was also obtained. The study was approved by the Institutional Review Boards of the University of California at San Francisco (UCSF; San Francisco, CA, USA) and at each participating centre.

GALA II

This is a case–control study of asthma in Latino children [27]. Cases and control subjects were recruited using a combination of community and clinic-based approaches from centers throughout the USA (Chicago (IL), Bronx (NY), Houston (TX), San Francisco Bay Area (CA) and Puerto Rico). Subjects were eligible if they were aged 8 to 21 years, had <10 pack–years of smoking history and were not current smokers, and self-reported having four grandparents of Latino ethnicity. Asthma was defined based on physician diagnosis and report of symptoms and medication use within the last two years. Control subjects had no history of asthma or allergies, and no wheeze or shortness of breath during their lifetime. The current analysis focused on 1786 participants (892 with asthma) with self-reported Puerto Rican ethnicity. All subjects were genotyped on the Axiom LAT1 array (World Array 4; Affymetrix), and quality control was performed as described elsewhere [28]. Imputation and association testing was performed as described above for GALA I. The study was approved by the Institutional Review Boards of UCSF and at each participating centre. All subjects and their parents provided written informed assent and written informed consent, respectively.

HCHS/SOL

This is a community-based cohort study of self-identified Hispanic/Latino individuals aged 18–74 years, who were recruited at four centres (Chicago, Miami, Bronx, San Diego). Participants were recruited using a two-stage sampling scheme, in which census block units were sampled first, and then household and individuals from these units [17, 18]. The HCHS/SOL study was approved by institutional review boards at participating institutions, and written informed consent was obtained from all participants.

The HCHS/SOL includes individuals who self-identified as Mexican, Central American, South American, Puerto Rican, Dominican or Cuban. However, genetic analysis groups were constructed based on a combination of these self-identified Hispanic/Latino background and genetic similarity. These genetic analysis groups largely overlap with the self-identified groups but using the genetic analysis groups in association testing and stratified analyses has advantages, as previously described [29]. Individuals from the Puerto Rican genetic analysis group were used in the current analysis. Cases (n=478) had self-reported current physician-diagnosed asthma. Control subjects (n=1388) had never been diagnosed with asthma. Individuals were genotyped at Illumina on the HCHS/SOL custom 15 041 502 B3 array, and imputed to 1000 Human Genomes Phase I data. Details about genotyping, imputation and quality control are provided elsewhere [29].

Association analysis was performed using GMMAT (Generalized linear Mixed Model Association Tests) [30], a mixed-model logistic regression that accounts for correlations due to relatedness, shared household and block group. The analysis was adjusted for five genetic principal components, study centre, sampling weights (to prevent potential selection bias resulting from the sampling scheme), age, sex, smoking status and pack–years of cigarette smoking.

In our primary meta-analysis, we included three cohorts of children (Hartford-PR, GALA I and GALA II) and one cohort of adults (HCHS/SOL) for maximum sample size. Since childhood asthma likely has genetic determinants that differ from those for adult-onset asthma, we repeated the meta-analysis after excluding HCHS/SOL.

Statistical methods

METAL (fast and efficient meta-analysis of genomewide association scans) [31] software was used to perform the meta-analysis of the four GWAS of asthma. METAL takes p-values across independent studies

as input, with sample size and effect direction taken into account. First, for each SNP, the coded and alternative alleles are determined and a Z-score is calculated based on the p-values and direction of effect in each study. Specifically, large positive Z-scores indicate small p-values where the coded allele is the risk allele, and large negative Z-scores indicate small p-values where the coded allele is protective. Formally the Z-score is:

$$Z_i = \Phi^{-1}(1 - P_i/2) \times \text{sign}(\Delta_i),$$

where Z_i is the Z-score for study i, P_i is the p-value for study i, Δ_i is the direction of effect for study i, and Φ^{-1} gives the percentile of a standard normal distribution. Then, the overall Z-score and p-value are calculated from a weighted sum of the individual Z-scores:

$$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}},$$

$$P = 2\Phi(|-Z|)$$

where Z is the overall Z-score, P is the overall p-value, and w_i is the weight for study i:

$$w_i = \frac{\text{MAF}_i(1 - \text{MAF}_i)N_i^{\text{cas}}N_i^{\text{con}}}{(N_i^{\text{cas}} + N_i^{\text{con}})}$$

where MAF is the minor allele frequency for study i, N_i^{cas} is the number of cases for study i and N_i^{con} is the number of controls for study i. This weighting is intended to closely approximate the results that would be obtained combining subject-level data across the studies, in an analysis that adjusts for study. Summary odds ratios were calculated by averaging the study-specific log-odds ratios, with weights reflecting the standard errors from the study-specific odds ratios.

We used the varLD [32, 33] software to assess whether there were statistically significant differences in regional linkage disequilibrium structures between our study populations. VarLD uses a Monte Carlo approach to calculate p-values, which were based on 10 000 permutations. For a comparison of two populations, a $p<0.05$ indicates a significant difference in linkage disequilibrium structures.

Results

The characteristics of participants in each of the four studies included in the meta-analysis are shown in table 1. The results reported here were calculated based on 7 485 508 imputed and genotyped SNPs in 2144 subjects with asthma and 2893 control subjects.

QQ plots for the meta-analysis (figures S1 and S3) show that neither the results from each of the four component studies nor the combined results were inflated in their test statistics. Moreover, these plots revealed an abundance of small p-values in the combined study. We expected to see SNPs with p-value$<5\times10^{-8}$, the standard significance threshold for GWAS. In the combined study, we found that 89 SNPs on chromosome 17q21 were significant at this level (figure 1 and table S1). Of these 89 SNPs, rs907092 on *IKZF3* showed the most significant association with asthma ($p=1.2\times10^{-12}$, figures 2 and 3).

TABLE 1 Summary of characteristics of participants included in the meta-analysis

	Hartford-PR	GALA I	GALA II	HCHS/SOL
Subjects n	948	437	1786	1866
Age years	10.0±2.7	18.2±9.5	12.9±3.4	48.2±14.0
Male sex	495 (52.2)	187 (42.8)	907 (50.8)	784 (42.0)
Asthma	523 (55.2)	251 (57.4)	892 (49.9)	478 (25.6)
Study sites	Hartford (CT) and San Juan (PR)	Puerto Rico and New York (NY)	Chicago (IL), Bronx (NY), Houston (TX), San Francisco (CA) and Puerto Rico	Bronx (NY), Chicago (IL), Miami (FL) and San Diego (CA)
Genotyping platform	Illumina 2.5M	Affymetrix 6.0 GeneChip	Axiom LAT1 array (Affymetrix)	Illumina HCHS/SOL custom 15 041 502 B3 array

Data are presented as mean±SD or n (%), unless otherwise stated. States within the USA are shown in brackets.

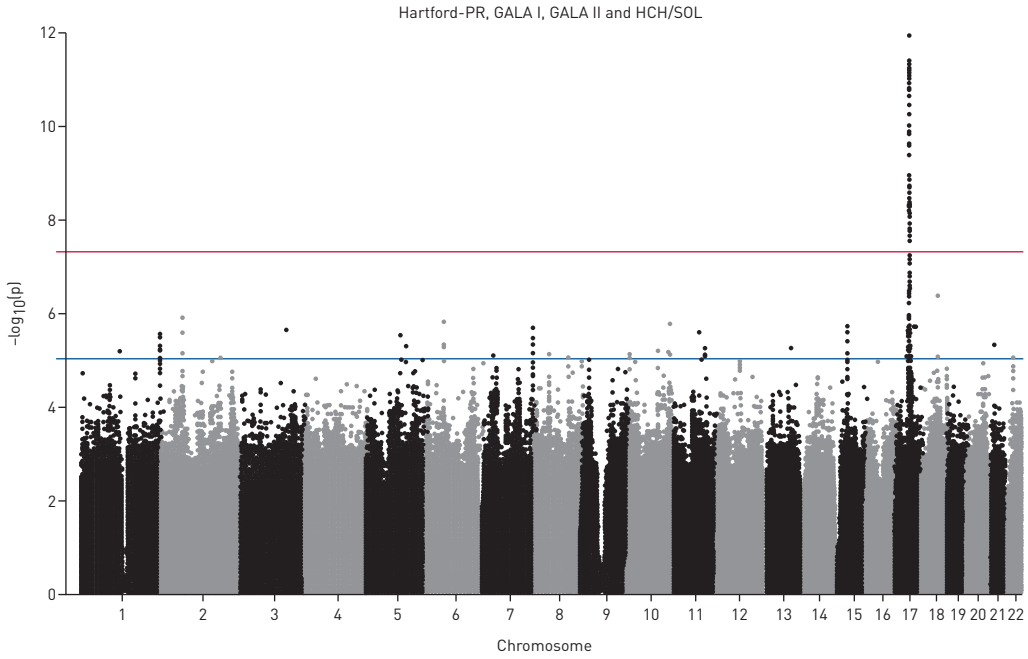

FIGURE 1 Manhattan plot showing the summary meta-analysis results of Hartford-PR, GALA I, GALA II, and HCHS/SOL. The chromosomal position of each SNP is displayed along the X-axis and the negative logarithm of the association p-value is displayed on the Y-axis. The blue line represents the suggestive significance line ($p<1\times10^{-5}$). The red line represents the genome-wide significance line ($p<5\times10^{-8}$).

This SNP was also associated with asthma at p<0.023 in all three studies of children and at p=0.092 in HCHS/SOL, a study of adults (figure 2 and figure S6).

Given our results for chromosome 17q21 and asthma, particularly in studies including children, we examined the linkage diseuilibrium pattern of this region among Puerto Ricans in Hartford–PR, GALA I and GALA II. Figures S8–S10 show that subjects in these studies had similar linkage disequilibrium patterns (varLD p-value=0.08 for Hartford-PR *versus* GALA I; p-value=0.46 for Hartford–PR *versus* GALA II (table S2)), and that nearly all significant SNPs gathered in a linkage disequilibrium bin spanning ~200 kb (see figures S8–S10). The *IKZF3-ZPBP2-GSDMB-ORMDL3* locus is included in this bin. We then compared the linkage disequilibrium pattern of chromosome 17q21 in Puerto Ricans to that of other ethnic groups (figure S11). Puerto Ricans had a linkage disequilibrium pattern that differed from that in Mexicans or Europeans (varLD p-value ⩽0.0007 for Hartford–PR *versus* Mexicans (in GALA I or GALA II) or 1000G Europeans (table S2)). For this analysis, Mexican genotypes were extracted from GALA I and GALA II, and European genotypes were extracted from the 1000 Genomes Project.

Next, we examined whether SNPs previously associated with asthma in either a meta-analysis of GWAS of Europeans [4] or a meta-analysis of ethnically diverse North American populations [3] were also

Study	Minor allele	MAF	p-value	OR
Hartford PR	A	0.31	5.0×10^{-6}	0.63
GALA I	A	0.31	0.023	0.63
GALA II	A	0.32	3.9×10^{-8}	0.66
HCHS/SOL	A	0.34	0.092	0.87
Summary			1.2×10^{-12}	0.71

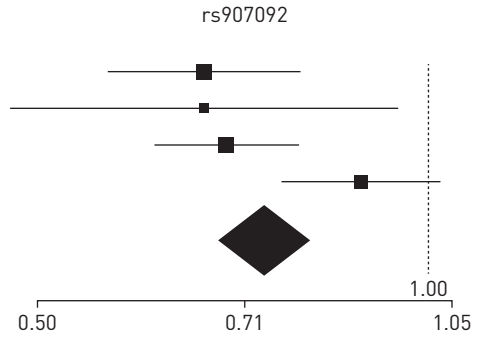

FIGURE 2 Forest plots of odds ratio and 95% confidence interval for the association with asthma. Forest plots for rs907092, the most significant SNP in the meta-analysis.

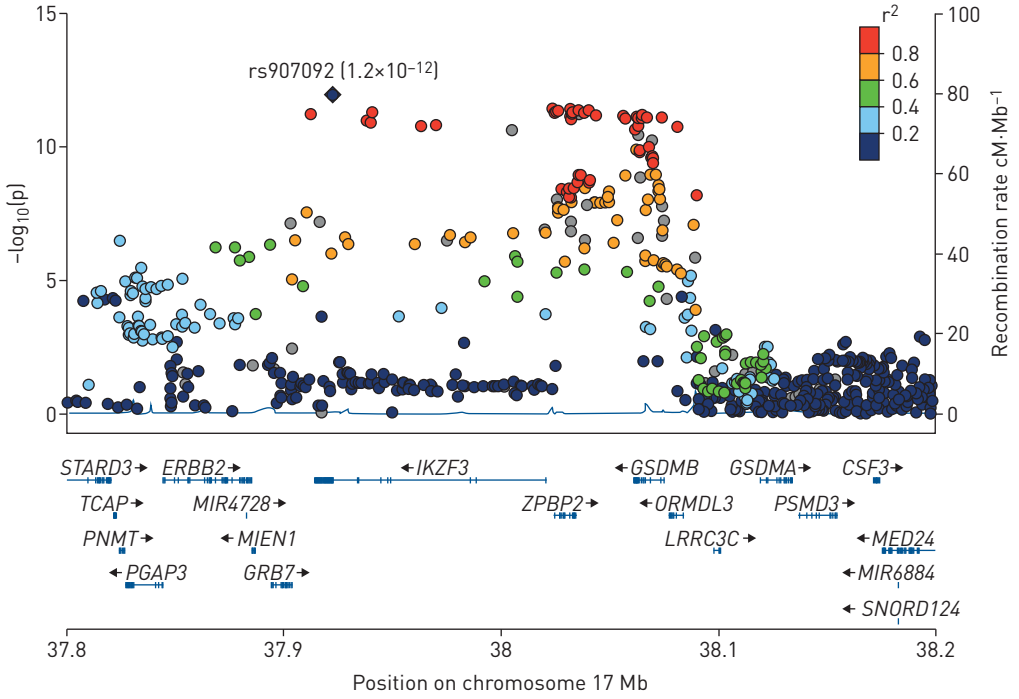

FIGURE 3 Results of the meta-analysis (Hartford-PR, GALA I, GALA II and HCHS/SOL) on the chromosome 17 region. The relative location of genes and the direction of transcription are shown in the lower portion of the figure, and the chromosomal position is shown on the *x* axis. The light blue line shows the recombination rate across the region (right *y* axis), and the left *y* axis shows the significance of the associations. The purple diamond shows the p-value for rs907092 that is the most significant SNP in the meta-analysis (Hartford-PR, GALA I, GALA II and HCHS/SOL). The circles show the p-values for all other SNPs and are colour coded according to the level of LD with rs907092 in the 1000 Genome Project Admixed American (AMR) population.

associated with asthma in Puerto Ricans (table 2). For the meta-analysis in Europeans, we were able to replicate associations for the SNP at *GSDMB* at p<0.005 (Bonferroni correction for 10 tests). For the meta-analysis of ethnically diverse populations, we were able to replicate the association with SNPs at *IL1RL1*, *TSLP* and *GSDMB* at p<0.01 (Bonferroni correction for 5 tests). The two SNPs (rs2305480 and rs11078927) at *GSDMB* identified in the two cited studies consistently had small p-values in Hartford–PR, GALA I, GALA II, as well as in the meta-analysis. Our top SNP (rs907092) at *IKZF3* was in high LD (r^2=0.84–0.85) with rs2305480 (the reported SNP at *GSDMB* from GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community)) and also in high LD (r^2=0.84–0.85) with rs11078927 (the reported SNP at *GSDMB* from EVE (A Study to Identify Asthma-susceptibility Genes in Ethnically Diverse Populations)). The r^2 between rs2305480 and rs11078927 was greater than 0.99. We did not replicate associations with SNPs in *IL33*, reported by both previous meta-analyses. Since this may be caused by differences in LD patterns in or near *IL33* between Puerto Ricans and ethnic groups included in the prior meta-analyses (*e.g.* Europeans and Mexicans; figure S12), we also show the top *IL33* SNPs from our meta-analysis in table S3. In this analysis, no *IL33* SNP was significantly associated with asthma (p>0.00035 in all instances, Bonferroni correction for 142 tests).

Since childhood asthma likely has genetic determinants that differ from those of adult-onset asthma, we repeated the meta-analysis of GWAS of asthma without HCHS/SOL, obtaining similar results (figures S4, S5 and S7) to those from the meta-analysis including HCHS/SOL: only SNPs on chromosome 17q21 reached a significance level of $p<5\times10^{-8}$. Moreover, we were able to replicate the same SNPs from the two previous meta-analyses of GWAS of asthma in the meta-analysis excluding HCHS/SOL (table S4). In addition, SNPs at *IL18R1* and *GSDMA* were replicated at p<0.005 (Bonferroni correction for 10 tests).

In order to check if there were independent SNPs associated with asthma, we performed conditional analyses in chromosome 17q21 by adjusting for the top SNP in our analysis (rs907092). Since adults did not show a strong genetic association in this region, we conducted the conditional analyses in a meta-analysis of the three cohorts of children. The results showed that no SNPs in chromosome 17q21 reached the Bonferroni- corrected significance level, $\alpha=4.1\times10^{-5}$ (1215 overlapped SNPs with minor allele frequencies (MAF) >0.05 across the three cohorts) and no SNPs in the *IKZF3-ZPBP2-GSDMB-ORMDL3* locus reached the Bonferroni-corrected significance level, $\alpha=2.5\times10^{-4}$ (197 overlapped SNPs with

TABLE 2 Results of the meta-analysis of genome-wide association studies of asthma in Puerto Ricans, for single nucleotide polymorphisms associated with asthma in two previous meta-analysis in Europeans (by the GABRIEL consortium) and ethnically diverse North Americans (by the EVE consortium)

SNP	Chr	Position	Ref	Alt	Gene	In cited paper	Hartford-PR			GALA I			GALA II			HCHS/SOL			Combined	
						OR	AAF	OR	p-value	AAF	OR	p-value	AAF	OR	p-value	AAF	OR	p-value	OR	p-value
Loci reported by the GABRIEL consortium [4]																				
rs3771166	2	102 986 222	G	A	*IL18R1*	0.85[#]	0.45	0.90	0.500	0.44	0.75	0.114	0.43	0.85	0.018	0.46	1.14	0.110	0.94	0.111
rs9273349	6	32 625 869	A	G	*HLA-DQ*	1.14[#]	0.61	0.98	0.860	NA	NA	NA	NA	NA	NA	0.62	1.22	0.016	1.13	0.079
rs1342326	9	6 190 076	A	C	*IL33*	1.27[#]	0.21	0.90	0.403	0.26	1.12	0.571	0.22	1.05	0.528	0.20	1.13	0.213	1.05	0.376
rs744910	15	67 446 785	G	A	*SMAD3*	0.89[#]	0.43	0.88	0.175	0.42	1.01	0.981	0.45	0.93	0.269	0.45	1.05	0.572	0.95	0.326
rs2305480	**17**	**38 062 196**	**G**	**A**	***GSDMB***	**0.76[#]**	**0.31**	**0.62**	**2.0×10^{-6}**	**0.31**	**0.54**	**0.003**	**0.32**	**0.71**	**6.7×10^{-6}**	**0.34**	**0.86**	**0.071**	**0.72**	**7.4×10^{-12}**
rs3894194	17	38 121 993	G	A	*GSDMA*	1.26[#]	0.37	1.23	0.034	0.39	0.92	0.648	0.40	1.23	0.002	0.39	0.94	0.406	1.11	0.023
rs2284033	22	37 534 034	G	A	*IL2RB*	0.92[#]	0.41	0.91	0.272	0.39	1.09	0.638	0.40	1.17	0.024	0.42	1.00	0.997	1.04	0.289
rs2073643	5	131 723 288	T	C	*SLC22A5*	0.89[#]	0.50	0.96	0.668	0.48	1.18	0.356	0.50	0.93	0.257	0.51	1.01	0.856	0.98	0.551
rs1295686	5	131 995 843	T	C	*IL13*	0.85[#]	0.59	1.01	0.943	0.62	0.75	0.136	0.62	0.90	0.159	0.61	1.07	0.398	0.96	0.408
rs11071559	15	61 069 988	C	T	*RORA*	0.88[#]	0.26	1.07	0.525	0.25	1.69	0.012	0.24	0.90	0.193	0.24	0.85	0.087	0.96	0.486
Replicated genes reported by the EVE consortium [3]																				
rs1102000	1	158 932 907	T	C	*PYHIN1*	0.76[¶]	0.08	1.09	0.651	0.06	0.50	0.061	NA	NA	NA	0.07	1.21	0.986	0.93	0.611
rs10173081	**2**	**102 957 348**	**C**	**T**	***IL1RL1***	**0.83[¶]**	**0.19**	**0.84**	**0.132**	**0.16**	**0.85**	**0.519**	**0.26**	**0.78**	**0.010**	**0.18**	**0.93**	**0.506**	**0.84**	**0.004**
rs1837253	**5**	**110 401 872**	**C**	**T**	***TSLP***	**0.84[¶]**	**0.25**	**0.73**	**0.004**	**0.23**	**0.65**	**0.041**	**0.27**	**0.83**	**0.014**	**0.26**	**0.96**	**0.701**	**0.82**	**2.5×10^{-4}**
rs2381416	9	6 193 455	A	C	*IL33*	1.18[¶]	0.38	0.98	0.847	0.39	1.00	0.980	0.37	1.09	0.249	0.35	1.13	0.157	1.07	0.155
rs11078927	**17**	**38 064 405**	**C**	**T**	***GSDMB***	**0.79[¶]**	**0.30**	**0.61**	**1.4×10^{-6}**	**0.31**	**0.55**	**0.003**	**0.31**	**0.71**	**7.9×10^{-6}**	**0.34**	**0.86**	**0.079**	**0.72**	**8.1×10^{-12}**

GABRIEL: A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community; EVE: A Study to Identify Asthma-susceptibility Genes in Ethnically Diverse Populations; Ref: reference allele; Alt: alternative allele; AAF: alternative allele frequency. Human Genome version: hg19. [#]: childhood onset odds ratio [4]; [¶]: overall odds ratio [3].

MAF>0.05 across the three cohorts). Thus, in addition to SNPs in linkage disequilibrium with rs907092, there seemed to be no other independent signals in chromosome 17q21.

Discussion

We report findings from the first GWAS of asthma conducted solely in Puerto Ricans. Although two of the cohorts included in the current meta-analysis (GALA I and II) were part of Hispanic/Latino cohorts included in previous studies, those studies did not separately analyse data from Puerto Ricans but instead combined data from Puerto Ricans and Mexicans. Such an approach can reduce statistical power, since Puerto Ricans and Mexicans differ markedly with regard to racial ancestry (*i.e.* Puerto Ricans have, on average, a greater proportion of African ancestry but a lower proportion of Native American ancestry than Mexicans), asthma burden (as Mexicans have a much lower asthma burden than Puerto Ricans) and allelic frequencies [15].

In our meta-analysis of four GWAS, the only locus to achieve a genome-wide significant association with asthma was chromosome 17q21, as evidenced by our top SNP, rs907092 ($p=1.2\times10^{-12}$), located on *IKZF3*. Similar to previous findings in non-Puerto Ricans, SNPs in genes located in the same linkage disequilibrium block as *IKZF3* on chromosome 17q21 (*e.g. ZPBP2*, *ORMDL3* and *GSDMB*) were also significantly associated with asthma in our meta-analysis. A previous candidate-gene study conducted in 399 Puerto Ricans and 301 Mexicans in GALA I, as well as 261 African Americans, found significant associations between two SNPs in *ORMDL3* (rs4378650 and rs12603332) and asthma in Mexicans and African Americans, but not in Puerto Ricans ($p=0.08$ for both SNPs) [6]. In our meta-analysis, these two SNPs ($p=4.0\times10^{-6}$ for rs4378650 and $p=5.5\times10^{-6}$ for rs12603332) were replicated for an association with asthma.

Consistent with prior findings suggesting stronger effects for SNPs in chromosome 17q21 on childhood asthma than in adult asthma among Europeans [4], we found genome-wide significant results for this locus in a meta-analysis restricted to the three cohorts of children, but not in the GWAS of asthma among adults in HCHS/SOL.

We also attempted to replicate previous findings for potential asthma-susceptibility SNPs from meta-analyses conducted in Europeans (by the GABRIEL consortium) [4] and ethnically diverse North American populations (by the EVE consortium) [3]. With regard to results from GABRIEL, we replicated findings for the SNP at *GSDMB* (on chromosome 17q21), but not for SNPs in any other gene. On the other hand, we report the first replication of results from the multi-ethnic EVE consortium for SNPs in *IL1RL1*, *TSLP* and *GSDMB* in a cohort composed exclusively of Puerto Ricans. Our lack of replication of most findings from GABRIEL in Puerto Ricans mimics negative findings from the EVE consortium, and may be due to ethnic differences in risk variants for asthma. In fact, EVE only replicated findings for chromosome 17q21 and *IL33* (albeit for different SNPs) from GABRIEL.

We did not replicate results for *PYHIN1* or *IL33* from EVE. *PYHIN1* was only associated with asthma among African-Americans and Afro-Caribbeans, and ancestry differences may account for our negative results. For SNPs with MAF $\geqslant$0.37 (the MAF of *IL33* SNP rs2381416 in our study cohorts), we had $\geqslant$70% statistical power to detect an odds ratio $\geqslant$1.10 at $\alpha\leqslant$0.01 in our meta-analysis [34]. Under the same assumptions, the statistical power to detect an odds ratio $\geqslant$1.20 at $\alpha\leqslant$0.01 was >99% in our meta-analysis. Of the 142 SNPs in *IL33* that were tested for association with asthma in the current study, only six were associated at $p<0.01$, and none remained significantly associated with asthma after a Bonferroni correction.

Expression of the 17q21 gene, *ORMDL3*, is regulated by asthma-associated SNPs [35], as replicated in our previous expression quantitative trait loci (eQTL) study (rs8067378, $p=2.6\times10^{-10}$) using the Hartford–PR data [36]. Lately, *ORMDL3* has been implicated in various cellular processes that could be relevant to asthma [35]. In addition to its effects on asthma susceptibility, the study from Tavendale *et al.* [11] also showed that *ORMDL3* was associated with asthma exacerbations and that the *ORMDL3* SNP-mediated expression is affected by rhinovirus infection, a common trigger of such exacerbations [37]. The SNP in the *IKZF3-ZPBP2-GSDMB-ORMDL3* region that was associated with asthma in the EVE meta-analysis was rs11078927 in *GSDMB*, which is replicated in our study ($p=8.1\times10^{-12}$ (table 2)). In EVE, the reported odds ratios for this SNP and asthma were 0.80 in European Americans and 0.78 in Latinos (non-significant in African Americans). In our study, we estimated odds ratios of 0.72 in all subjects and 0.66 in children only, suggesting a stronger effect of the *IKZF3-ZPBP2-GSDMB-ORMDL3* locus on childhood asthma in Puerto Ricans than in other ethnic groups. Thus, the effects of the major allele for SNP rs11078927 (*i.e.* OR 1.51 (1/0.66) in children) may contribute to asthma aetiology in Puerto Ricans.

The *ZPBP2-GSDMB-ORMDL3* locus was detected to have differences in expression due to allelic differences in lymphoblastoid cell lines (LCLs) and $CD4^+$ cells [38, 39]. Early studies of gene expression [40–42] revealed that the asthma-associated SNPs regulate the expression of *ORMDL3* and *GSDMB,* and that these two genes might be coregulated [38]. Findings from a prior study suggest that an asthma-associated 17q21 regulatory

haplotype affects transcriptional activity of *ZPBP2, GSDMB* and *ORMDL3*, with one non-coding variant in ZPBP2, rs12936231, differentially influencing the binding of the insulator protein CTCF in an allele-specific manner [38]; subsequent eQTL studies showed that SNP rs12936231 increases ORMDL3 and GSDMB expression in primary lymphocytes, whole blood and lung tissue [43, 44]. However, SNP rs12936231 (which has a minor allele frequency=0.44–0.46 in our cohorts) was non-significantly associated with asthma in our meta-analysis ($p=2.0\times10^{-6}$) and is only in moderate linkage disequilibrium ($r^2>0.51$) with our top SNP (rs907092) in *IKZF3*. This suggests that other functional polymorphisms on chromosome 17q21 may affect asthma risk in Puerto Ricans, likely through changes in expression of ORMDL3 (shown to cause experimental asthma in a transgenic murine model) [45], *GSDMB* or *ZPBP2*. Another SNP located within the promoter region of *ZPBP2* (rs4795397, $p=3.7\times10^{-12}$ in our meta-analysis (table S1)) is a putative functional polymorphism that shows asthma-associated allele-specific nucleosome occupancy [39]. However, the SNP's strong influence on *ZPBP2* promoter activity is masked by DNA methylation of exon 1 of this gene. In contrast, the *ORMDL3* promoter is fully unmethylated. The *ZPBP2* and *ORMDL3* genes show allelic differences in expression [38, 39]. It has been shown that the *IKZF3-ZPBP2-GSDMB-ORMDL3* haplotype acts differently in regulation of transcripts between Europeans and Africans [38], with a stronger association in Europeans than in Africans. In a study by Verlaan *et al.* [38], SNP rs8067378 was the most significant *cis*-eQTL in this four-gene region in both Europeans ($p=1.1\times10^{-18}$) and Africans ($p=4.3\times10^{-9}$). In an eQTL study in Puerto Ricans using our Hartford–PR dataset [36], the same SNP was also the most significant *cis*-eQTL ($p=2.6\times10^{-10}$) in the four-gene region. Although the two studies [36, 38] might not be comparable due to different statistical models and different tissues (*i.e.*, LCL used in the Verlaan *et al.* [38] study and globin-cleaned whole blood used in the Chen *et al.* [36] study), we speculate that the significance of associated SNPs with expression in Puerto Ricans is between Europeans and Africans, since these are two ancestral populations in Puerto Ricans. We further performed eQTL analyses between our top SNP (rs907092) and expression of the *IKZF3-ZPBP2-GSDMB-ORMDL3* locus using whole-blood RNA microarray data (n=121) from the Hartford–PR study. Findings from this secondary analysis, conducted using linear regression and adjusting for age, sex, asthma status, study site and the first two principal components, suggest that our top SNP (rs907092) may regulate expression of the *IKZF3-ZPBP2-GSDMB-ORMDL3* locus (p=0.07 for *IKZF3*, p=0.02 for *ZPBP2*, p=0.009 for *GSDMB* and p=0.0004 for *ORMDL3*).

We recognise several study limitations. First, we had insufficient statistical power to detect weak genetic effects of common SNPs (*e.g.* odds ratios between 0.9 and 1.10 for *IL33*) or rare susceptibility variants. Second, subjects in our study lived in different sites in mainland USA and Puerto Rico, and thus environmental differences may have confounded our results despite adjustment for study site in the four component studies. However, most minor allelic frequencies and effect estimates were similar across study cohorts. Third, misclassification of COPD as asthma is possible among adults in the HCHS/SOL cohort, given that 980 (52.5%) participants were former or current smokers (mean pack-years of 18.6). Such misdiagnosis could account for the less significant associations reported in HCHS/SOL, even after adjustment for cigarette smoking. In summary, our findings suggest that common allelic variants in the chromosome 17q21 locus have the greatest effects on, and are thus particularly important in, the pathogenesis of asthma in Puerto Ricans. We confirmed susceptibility variants in two previously reported genes (*IL1RL1* and *TLSP*). Future studies should aim to characterize functional variants that cause asthma in Puerto Ricans, either through primary genetic effects or through interactions with relevant exposures (*e.g.*, second-hand smoke and air pollution).

Acknowledgements

We thank the participants and staff of the Hartford-PR study, GALA I and GALA II, and the HCHS/SOL study for their contributions to this work.

Author contributions: conception and study design: R.G. Barr, E.G. Burchard and J.C. Celedón; data acquisition, analysis and interpretation: Q. Yan, M. Pino-Yanes, J. Brehm, E. Forno, J. Lin, S.S. Oh, E. Acosta-Perez, C.C. Laurie, M.M. Cloutier, B.A. Raby, A.M. Stilp, T. Sofer, D. Hu, S. Huntsman, C.S. Eng, M.P. Conomos, D. Rastogi, K. Rice, G. Canino and W. Chen; and drafting of the manuscript for intellectual content: Q. Yan, R.G. Barr, E.G. Burchard and J.C. Celedón. All authors approved the final version of the manuscript prior to submission.

References

1 Duffy DL, Martin NG, Battistutta D, *et al.* Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 1990; 142: 1351–1358.
2 Nieminen MM, Kaprio J, Koskenvuo M. A population-based study of bronchial asthma in adult twin pairs. *Chest* 1991; 100: 70–75.
3 Torgerson DG, Ampleford EJ, Chiu GY, *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse north American populations. *Nat Genet* 2011; 43: 887–892.
4 Moffatt MF, Gut IG, Demenais F, *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363: 1211–1221.
5 Bouzigon E, Corda E, Aschard H, *et al.* Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008; 359: 1985–1994.

6 Galanter J, Choudhry S, Eng C, *et al. ORMDL3* gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 2008; 177: 1194–1200.
7 Halapi E, Gudbjartsson DF, Jonsdottir GM, *et al.* A sequence variant on 17q21 is associated with age at onset and severity of asthma. *Eur J Hum Genet* 2010; 18: 902–908.
8 Leung TF, Sy HY, Ng MC, *et al.* Asthma and atopy are associated with chromosome 17q21 markers in Chinese children. *Allergy* 2009; 64: 621–628.
9 Madore AM, Tremblay K, Hudson TJ, *et al.* Replication of an association between 17q21 SNPs and asthma in a French-Canadian familial collection. *Hum Genet* 2008; 123: 93–95.
10 Sleiman PM, Annaiah K, Imielinski M, *et al. ORMDL3* variants associated with asthma susceptibility in north Americans of European ancestry. *J Allergy Clin Immunol* 2008; 122: 1225–1227.
11 Tavendale R, Macgregor DF, Mukhopadhyay S, *et al.* A polymorphism controlling *ORMDL3* expression is associated with asthma that is poorly controlled by current medications. *J Allergy Clin Immunol* 2008; 121: 860–863.
12 Bisgaard H, Bønnelykke K, Sleiman PM, *et al.* Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. *Am J Respir Crit Care Med* 2009; 179: 179–185.
13 Galanter JM, Gignoux CR, Torgerson DG, *et al.* Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The genes-environments & admixture in Latino Americans study. *J Allergy Clin Immunol* 2014; 134: 295–305.
14 Akinbami LJ, Moorman JE, Bailey C, *et al.* Trends in asthma prevalence, health care use, and mortality in the United States, 2001–2010. *NCHS Data Brief* 2012; 1–8.
15 Rosser FJ, Forno E, Cooper PJ, *et al.* Asthma in Hispanics. An 8-year update. *Am J Respir Crit Care Med* 2014; 189: 1316–1327.
16 Brehm JM, Acosta-Pérez E, Klei L, *et al.* African ancestry and lung function in Puerto Rican children. *J Allergy Clin Immunol* 2012; 129: 1484–1490.
17 Lavange LM, Kalsbeek WD, Sorlie PD, *et al.* Sample design and cohort selection in the Hispanic community health study/study of Latinos. *Ann Epidemiol* 2010; 20: 642–649.
18 Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, *et al.* Design and implementation of the Hispanic community health study/study of Latinos. *Ann Epidemiol* 2010; 20: 629–641.
19 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; 5: e1000529.
20 Price AL, Patterson NJ, Plenge RM, *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904–909.
21 Burchard EG, Avila PC, Nazario S, *et al.* Lower bronchodilator responsiveness in Puerto Rican than in Mexican subjects with asthma. *Am J Respir Crit Care Med* 2004; 169: 386–392.
22 Choudhry S, Burchard EG, Borrell LN, *et al.* Ancestry-environment interactions and asthma risk among Puerto Ricans. *Am J Respir Crit Care Med* 2006; 174: 1088–1093.
23 Torgerson DG, Gignoux CR, Galanter JM, *et al.* Case-control admixture mapping in Latino populations enriches for known asthma-associated genes. *J Allergy Clin Immunol* 2012; 130: 76–82 e12.
24 Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012; 9: 179–181.
25 Genomes Project C,Abecasis GR, Altshuler D, *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; 467: 1061–1073.
26 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009; 19: 1655–1664.
27 Nishimura KK, Galanter JM, Roth LA, *et al.* Early-life air pollution and asthma risk in minority children. The gala ii and sage ii studies. *Am J Respir Crit Care Med* 2013; 188: 309–318.
28 Pino-Yanes M, Thakur N, Gignoux CR, *et al.* Genetic ancestry influences asthma susceptibility and lung function among latinos. *J Allergy Clin Immunol* 2015; 135: 228–235.
29 Conomos MP, Laurie CA, Stilp AM, *et al.* Genetic diversity and association studies in us hispanic/latino populations: Applications in the hispanic community health study/study of latinos. *Am J Hum Genet* 2016; 98: 165–184.
30 Chen H, Wang C, Conomos MP, *et al.* Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet* 2016; 98: 653–666.
31 Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26: 2190–2191.
32 Teo YY, Fry AE, Bhattacharya K, *et al.* Genome-wide comparisons of variation in linkage disequilibrium. *Genome Res* 2009; 19: 1849–1860.
33 Ong RT, Teo YY. Varld: A program for quantifying variation in linkage disequilibrium patterns between populations. *Bioinformatics* 2010; 26: 1269–1270.
34 Lara M, Akinbami L, Flores G, *et al.* Heterogeneity of childhood asthma among Hispanic children: Puerto Rican children bear a disproportionate burden. *Pediatrics* 2006; 117: 43–53.
35 Ono JG, Worgall TS, Worgall S. 17q21 locus and *ORMDL3*: An increased risk for childhood asthma. *Pediatr Res* 2014; 75: 165–170.
36 Chen W, Brehm JM, Lin J, *et al.* Expression quantitative trait loci (EQTL) mapping in Puerto Rican children. *PLoS One* 2015; 10: e0122464.
37 Çalişkan M, Bochkov YA, Kreiner-Møller E, *et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* 2013; 368: 1398–1407.
38 Verlaan DJ, Berlivet S, Hunninghake GM, *et al.* Allele-specific chromatin remodeling in the *ZPBP2*/*GSDMB*/*ORMDL3* locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet* 2009; 85: 377–393.
39 Berlivet S, Moussette S, Ouimet M, *et al.* Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. *Hum Genet* 2012; 131: 1161–1171.
40 Verlaan DJ, Ge B, Grundberg E, *et al.* Targeted screening of cis-regulatory variation in human haplotypes. *Genome Res* 2009; 19: 118–127.
41 Stranger BE, Nica AC, Forrest MS, *et al.* Population genomics of human gene expression. *Nat Genet* 2007; 39: 1217–1224.

42 Dixon AL, Liang L, Moffatt MF, *et al.* A genome-wide association study of global gene expression. *Nat Genet* 2007; 39: 1202–1207.
43 Murphy A, Chu JH, Xu M, *et al.* Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood cd4+ lymphocytes. *Hum Mol Genet* 2010; 19: 4745–4757.
44 Hao K, Bossé Y, Nickle DC, *et al.* Lung EQTLS to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8: e1003029.
45 Miller M, Rosenthal P, Beppu A, *et al. ORMDL3* transgenic mice have increased airway remodeling and airway responsiveness characteristic of asthma. *J Immunol* 2014; 192: 3475–3487.