Sugier, P-E., Sarnowski, C., Granell, R., Laprise, C., Ege, M., Margaritte-Jeannin, P., Dizier, M-H., Minelli, C., Moffatt, M. F., Lathrop, M., Cookson, W. O. C. M., Henderson, J., von Mutius, E., Kogevinas, M., Demenais, F., & Bouzigon, E. (2019). Genome-wide interaction study of early-life smoking exposure on time-to-asthma onset in childhood. *Clinical and Experimental Allergy*, *49*(10), 1342-1351. https://doi.org/10.1111/cea.13476

Peer reviewed version

Link to published version (if available):
10.1111/cea.13476

Link to publication record in Explore Bristol Research
PDF-document

Genome-wide interaction study of early-life smoking exposure on time-to-asthma onset in childhood

Running title: gene-smoking exposure interaction in childhood asthma

Pierre-Emmanuel Sugier[1,2], Chloé Sarnowski[1], Raquel Granell[3], Catherine Laprise[4], Markus J Ege[5,6], Patricia Margaritte-Jeannin[1], Marie-Hélène Dizier[1], Cosetta Minelli[7], Miriam F. Moffatt[8], Mark Lathrop[9], William O.C.M. Cookson[8], A. John Henderson[3], Erika von Mutius[5,6,10], Manolis Kogevinas[11,12,13], Florence Demenais[1]*, Emmanuelle Bouzigon[1]*

1. Inserm, UMRS-1124, Genetic Epidemiology and Functional Genomics of Multifactorial Diseases Team, Université Paris Descartes, Paris, France

2. Université Pierre et Marie Curie, Paris, France

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

4. Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, QC, Canada

5 Dr von Hauner Children's Hospital, Ludwig Maximilian University, Munich, Germany

6. Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich, Germany

7. Population Health & Occupational Disease, National Heart and Lung Institute, Imperial College, London, UK

8. Section of Genomic Medicine, National Heart Lung Institute, Imperial College London, London, UK

9. McGill University and Génome Québec Innovation Centre, Montréal, Canada

10. Helmholtz Zentrum München - German Research Center for Environmental Health, Institute for Asthma and Allergy Prevention, Neuherberg, Germany

11. ISGlobal, Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

12. CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

13. Municipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain

* These authors contributed equally to this work.

Corresponding author:

Emmanuelle Bouzigon, MD, PhD

Inserm UMRS-1124, Genetic Epidemiology and Functional Genomics of Multifactorial Diseases Team, Université Paris Descartes

Institut de Génétique Moléculaire, 27 rue Juliette Dodu, 75010 Paris, France

Tel: 33 (0) 1 72 63 93 05 / Fax: 33 (0) 1 72 63 93 49

Email: emmanuelle.bouzigon@inserm.fr

ABSTRACT

Background: Asthma, a heterogeneous disease with variable age of onset, results from the interplay between genetic and environmental factors. Early-life tobacco smoke (ELTS) exposure is a major asthma risk factor. Only a few genetic loci have been reported to interact with ELTS exposure in asthma.

Objective: Our aim was to identify new loci interacting with ELTS exposure on time-to-asthma onset (TAO) in childhood.

Methods: We conducted genome-wide interaction analyses of ELTS exposure on time-to-asthma onset in childhood in five European-ancestry studies (totaling 8,273 subjects) using Cox proportional-hazard model. The results of all five genome-wide analyses were meta-analyzed.

Results: The 13q21 locus showed genome-wide significant interaction with ELTS exposure (P=4.3x10^{-8} for rs7334050 within *KLHL1* with consistent results across the five studies). Suggestive interactions (P<5x10^{-6}) were found at three other loci: 20p12 (rs13037508 within *MACROD2*; P=4.9x10^{-7}), 14q22 (rs7493885 near *NIN*; P=2.9x10^{-6}) and 2p22 (rs232542 near *CYP1B1*; P=4.1x10^{-6}). Functional annotations and the literature showed that the lead SNPs at these four loci influence DNA methylation in the blood and are located nearby CpG sites reported to be associated with exposure to tobacco smoke components, which strongly support our findings.

Conclusion and Clinical Relevance: We identified novel candidate genes interacting with ELTS exposure on time-to-asthma onset in childhood. These genes have plausible biological relevance related to tobacco smoke exposure. Further epigenetic and functional studies are needed to confirm these findings and to shed light on the underlying mechanisms.

Keywords: gene-environment interaction, environmental tobacco smoke exposure, childhood asthma, time-to-asthma onset

INTRODUCTION

Asthma, one of the most common chronic diseases, results from the interplay between genetic and environmental factors. Asthma has variable age of onset and variable expression over the life span. It is recognized that childhood-onset asthma may be distinct from later-onset asthma and may represent a more homogeneous subgroup often associated with allergy. The genetic component of asthma is substantial but the asthma loci, identified so far, explain only a part of the genetic risk [1]. One potential reason for this missing heritability is gene-environment (GxE) interaction because some genetic variants may confer risk only in the presence of environmental exposures.

Tobacco smoke exposure in early-life is a major risk factor for asthma. Interactions between genetic variants and early-life tobacco smoke (ELTS) exposure on asthma were first identified by genome-wide linkage scans and candidate gene studies [2]. ELTS exposure was then found to increase the risk of early-onset asthma associated with the 17q12-21 variants [3] identified by the first asthma GWAS [4]. Recently, a genome-wide interaction study (GWIS) of childhood-onset asthma reported interactions between *in utero* exposure to maternal smoking and the 18p11 locus and between exposure to parental smoking in childhood and the 6q26 locus but none of these interactions reached genome-wide significance [5]. A subsequent GWIS conducted for adult-onset asthma and active smoking revealed suggestive interactions on chromosomes 9p23 and 12p12 [6]. Therefore, similarly to the previously reported heterogeneity of SNP effect on asthma risk according to age of onset of asthma [3,7], the SNP-smoke exposure interactions may differ between childhood-onset and later-onset asthma and may vary according to time of exposure. Only two GWAS have considered either age of asthma onset in asthmatic children [8] or time-to-asthma onset (TAO) in asthmatic and non-asthmatic subjects [9] and have led to the discovery of new asthma loci, but no GWIS has yet taken into account the time-to-asthma onset.

In order to identify new asthma risk loci, we conducted a genome-wide interaction study of ELTS exposure on time-to-asthma onset in childhood in five European ancestry studies.

METHODS

Study populations

The total sample consisted of 8273 subjects of European-ancestry (2874 subjects with childhood-onset asthma and 5399 non-asthmatic subjects) from five studies, which were part of the Gabriel asthma consortium [7]. The five datasets were from three population-based studies (the ALSPAC [10] birth corhort from UK, the pan-European ECRHS cohort study [11] including sixteen centres from eight countries and the cross-sectional GABRIELA [12] survey conducted in rural areas of Austria, Germany and Switzerland) and two family-based studies (the French EGEA cohort study [13] and the French-Canadian SLSJ study [14]). A detailed description of these studies can be found in the Supplementary Information. These were the only Gabriel consortium studies of sufficient sample size that had data on age of asthma onset, ELTS exposure and imputed SNP data (Table S1-A). Written informed consent was signed by all participants or by kin or guardians for minors/children. Ethical approval was obtained for each study from the appropriate institutional ethics committees (ethical approval numbers are provided in the Supplementary Information).

Definition of time to asthma onset and ELTS exposure

The analysis of time-to-asthma onset phenotype requires to specify the event (here, asthma occurrence) and the time to event. Information on occurrence of asthma was based on report of doctor's diagnosis and/or on standardized questionnaires, as used in previous GWAS [7,9]. Childhood asthma was defined as having asthma at or before 16 years of age. The time to event was the time (in years) that has elapsed from birth to occurrence of childhood asthma during the

follow-up period. For those subjects who had childhood asthma before the end of the follow-up period, time to event was the age of onset of childhood asthma while for those subjects who were free of asthma upon last follow-up, time to event was their age at last follow-up. Individuals who developed asthma after 16 years of age were excluded. ELTS exposure was defined as being exposed to maternal smoking during pregnancy and/or parental smoking in early childhood. For more than 70% of the subjects included in the present study, the definition of the ELTS exposure was based on an exposure before 5 years of age.

Genotyping and imputation

SNP genotyping and imputation and quality control (QC) criteria are summarized in Table S1-B. Genotyping was carried out using the Illumina Human610-Quad array for all studies except for ALSPAC where the Illumina HumanHap550-Quad array was used. We used HapMap2 imputed data as available in Gabriel consortium studies. Imputation was performed as previously described [7]. We kept for analysis SNPs with imputation quality score (rsq) ≥ 0.5 and minor allele frequency $\geq 1\%$, making a total of 2.11 million SNPs for analysis.

Gene-environment-wide interaction analysis

Instead of artificially splitting the five datasets into discovery and replication sets, we conducted a meta-analysis of all five datasets to maximize statistical power. This approach has been previously used for the time-to-asthma onset GWAS [9]. As shown in Figure 1, each dataset was split in ELTS-exposed ($ELTS^+$) and ELTS-unexposed ($ELTS^-$) subjects. A genome-wide association analysis of TAO was conducted in each stratified dataset using the Cox proportional-hazard model. This model requires to specify an event indicator and time to event. For subjects who developed childhood asthma during the follow-up period, the event indicator was coded 1 and the time to

event was the age of onset of childhood asthma while for subjects who did not have asthma during the follow-up period, the event indicator was coded 0 and the time to event was the age at last follow-up. The hazard ratio of individual SNPs for TAO was estimated using the Cox model while adjusting for sex and the first four principal components to correct for population stratification and assuming an additive genetic model for SNP effect. A robust sandwich estimator of the variance with cluster on family was used to take into account familial dependencies in the family studies. The complex sampling design of the GABRIELA study was taken into account by using survey regression techniques to estimate robust standard errors ('svy' command in Stata). In each $ELTS^+/ELTS^-$ stratum, the hazard ratios (HRs) estimated from each of the five studies were combined using a fixed-effect meta-analysis with inverse variance weighting. A hazard ratio greater (lower) than one means an increased (decreased) risk of childhood asthma occurrence at a given time (given unaffected before) among all subjects at risk at that time that is associated with one unit increase in the number of effect alleles in the genotype. The SNPxELTS interaction was estimated as the difference (D_{HR}) between the $ELTS^+$ and $ELTS^-$ meta-analyzed HRs (on the log scale); the variance of D_{HR} was estimated, as previously explained [15]. The test statistic for SNPxELTS interaction (D_{HR} divided by the square root of its variance) was compared to a standard normal distribution. A SNPxELTS interaction was declared as genome-wide significant if it reached the threshold of 5×10^{-8}; an interaction reaching the threshold of $5x10^{-6}$ was considered as suggestive. Furthermore, the homogeneity of SNP HR estimates across all five studies in each $ELTS^+/ELTS^-$ stratum was assessed using the Cochran's Q homogeneity test, and the extent of heterogeneity was estimated using the I^2 statistic, which describes the percentage of variation across studies that is due to heterogeneity rather than to chance [16]. All analyses were conducted using Stata© V14.1.

Functional annotations of the loci interacting with ELTS exposure

To provide biological insight into our findings, we conducted a bioinformatic assessment of the loci detected by our genome-wide interaction meta-analysis. At each locus, we defined a list of SNPs to be interrogated that included the most significant SNP interacting with ELTS (designated as lead SNP) and all SNPs in LD with the lead SNP (r^2 comprised between 0.5 and 1). To pinpoint the most likely candidate genes at the identified loci, we searched for cis-expression quantitative trait loci (eQTLs) within at most 1 Mb of each investigated SNP by interrogating four eQTL studies in the blood (peripheral blood [17], lymphoblastoid cell lines [18,19], monocytes [20]) and the GTEx database that contains eQTL data from many tissues [21]. To complement the eQTL analysis, we searched for missense variants potentially tagged by the interaction signals using the HaploReg v4.1 tool [22]. To get greater insight into how the genetic variants interacting with ELTS may functionally influence TAO, we investigated whether the SNPs from the aforementioned SNP set were located in the vicinity of cis-regulatory DNA elements and transcription factor (TF) binding sites, using ROADMAP/ENCODE functional genomics data generated in a wide range of human cell types [23] and summarized in HaploReg v4.1 [22]. We also conducted a search in the Phenoscanner database [24] to assess whether the SNPs were previously reported in genetic association studies with diseases and traits as well as molecular phenotypes including DNA methylation.

RESULTS

The flow chart of the study as well as the sample size of each of the five datasets by $ELTS^+/ELTS^-$ stratum (total n=8273, of which 3187 were $ELTS^+$ and 5086 were $ELTS^-$) are shown in Figure 1.

Gene-environment-wide interaction analysis

The Manhattan plot of interaction *P*-values for the genome-wide interaction meta-analysis of ELTS exposure on TAO is shown in Figure 2. There was little inflation in the interaction test statistics (QQ plot in Figure S1, genomic inflation factor (lambda) = 1.003). A genome-wide significant interaction with ELTS exposure was found at the 13q21 locus ($P=4.3x10^{-8}$ for rs7334050). Besides the lead SNP at that locus, there were five additional variants that showed suggestive interactions ($1.3x10^{-7} < P < 3.9x10^{-6}$; Table S2). Suggestive interactions ($P \leq 5x10^{-6}$) were also observed at three other loci on chromosomes 20p12, 14q22 and 2p22. The results at all four loci are shown in Table 1 for the lead SNPs and in Table S2 for the additional SNPs. The G minor allele (MAF=0.14) of the significant SNP, rs7334050, at 13q21 conferred a hazard ratio higher than one in ELTS exposed subjects ($HR_{ELTS+} = 1.34$, 95% Confidence Interval (CI), 1.19-1.52) and a hazard ratio lower than one in unexposed subjects ($HR_{ELTS-} = 0.85$, 95% CI, 0.76-0.95). The SNPxELTS interaction hazard ratio was always in the same direction for all five studies (ranging from 1.16 to 3.39). There was homogeneity of rs7334050 HR estimates across the five studies in each $ELTS^+/ELTS^-$ stratum (*P* for Cochran's Q test >0.31; the I^2 estimates were equal to 0.0 in both $ELTS^+$ and $ELTS^-$ strata; Table S2 and Figure 3 for forest plot). All lead SNPs at the other three loci showed an opposite direction of effect according to exposure (Table 1) with consistent hazard ratios across studies in both $ELTS^+/ELTS^-$ strata (*P*-values for Cochran's Q test being greater than 0.15; Table S2 and Figure S2 for forest plots).

Functional annotations of the loci interacting with ELTS exposure

The search for cis-eQTLs at the four loci detected by this GWIS showed that the lead SNP rs232542 at 2p22 and three proxies (r^2 >0.99) were strong cis-eQTLs for the *CYP1B1* gene ($7.0x10^{-34} \leq P \leq 3.6x10^{-33}$) in the blood [17]. No cis-eQTL was found at the other three loci. However, the 13q21 lead SNP (rs7334050) is located in an intron of *KLHL1*, which is the only protein coding gene within 1 Mb on each side of that SNP. Similarly, the lead SNP (rs13037508) at 20p12 is an intronic variant in *MACROD2*, the sole protein coding gene within 500 kb of that SNP. The lead SNP (rs7493885) at 14q22 is closest to *NIN* and is in LD (r^2=0.65) with SNPs within *NIN* promoter. Moreover, the interrogated SNPs at all four loci did not tag any missense variant.

The colocalization of lead SNPs and proxies at the four loci with regulatory elements are shown in Table 2 and Table S3. The lead *KLHL1* SNP maps to binding sites of transcription factors (TFs) and a nearby SNP in strong LD (r^2=0.87) maps to histone marks in fetal lung and DNase I hypersensitive sites (DHSs) in hematopoietic stem cells. The lead SNPs (and proxies) at the other three loci colocalize with histone marks and/or DHSs in blood cells and the lungs and TF binding sites. Notably, these TFs include CTCF at 2p22, 14q22 and 20p12 loci and Ahr (Arhyl hydrocarbon receptor) and its partner Arnt (Arhyl hydrocarbon receptor nuclear translocator) at 14q22. CTCF functions as a transcriptional activator, repressor or insulator protein. It was recently shown that CTCF is a major driver of gene co-expression in the airways of asthmatic patients [25]. The Ahr and Arnt TFs play a major role in the regulation of biological responses to tobacco smoke components [26]. Finally, interrogation of the Phenoscanner database showed that the lead SNPs at all four loci were strongly associated with DNA methylation levels in the whole blood (Table 2). Associations of DNA methylation with 2p22 and 14q22 lead SNPs were also observed in neutrophils and immune cells. As shown in Table 2, the G allele of rs7334050 (13q21) and A allele

of rs13037508 (20p12) were associated with higher methylation while the G allele of rs7493885 (14q22) and C allele of rs232542 (2p22) were associated with lower methylation.

DISCUSSION

To our knowledge, this is the first study to examine SNP-by-ELTS exposure interactions on time-to-asthma onset using a genome-wide approach. We identified a significant interaction with ELTS exposure at the 13q21 locus and suggestive interactions at three other loci on chromosomes 2p22, 14q22 and 20p12. The evidence for these interactions rests on the results obtained in five large European-ancestry studies and the consistency of results across studies.

The SNP (rs7334050) showing significant interaction with ELTS exposure is located within the Kelch-like 1 (*KLHL1*) gene. *KLHL1* encodes a neuronal actin-binding protein that modulates voltage-gated calcium channels [27], known to play a role in airway smooth muscle contraction, cytokine production and airway inflammation [28]. Cigarette smoke was shown to enhance the expression of Ca2+ regulatory proteins leading to increased cell proliferation of airway smooth muscle and cytokine generation [29]. Interestingly, KLHL1 showed increased mutation frequency in lung tumors after exposure to benzopyrene [30] and is located nearby newborn blood DNA methylation modifications associated with prenatal exposure to arsenic [31]. Benzopyrene and arsenic are both known components of cigarette smoke.

Although the other three loci did not reach genome-wide significance, they harbor relevant candidates with biological function related to tobacco smoke exposure. The 20p12 lead SNP is within *MACROD2*, a gene encoding a deacetylase involved in removing ADP-ribose from mono-ADP-ribosylated proteins. This gene is part of the epigenetic signature of cigarette smoking and was found to colocalize with differentially methylated CpG sites in former smokers that never returned to never-smoker levels after 30 years of smoking cessation. The 14q22 lead SNP is located nearby *NIN*, which encodes a protein with a key role in ciliogenesis [32]. Genes involved in cilia function have been previously reported to interact with tobacco smoke exposure, either in early life for *DNAH9* on BHR [2] or in childhood for *PACRG* on childhood asthma [5]. A proxy of the 14q22

lead SNP colocalizes with AhR and Arnt TF binding sites, which play a crucial role in the biological response to polyaromatic hydrocarbons, and newborn blood DNA methylation changes near *NIN* were associated with maternal smoking during pregnancy [33]. Finally, at 2p22, the *CYP1B1* gene, whose expression is strongly associated with SNPs interacting with ELTS in this study, encodes a member of the cytochrome P450 superfamily of enzymes which metabolizes tobacco smoke components. The induction of *CYP1B1* expression in response to smoke exposure was recently confirmed by an epigenome-wide association study of cigarette smoking in lung cells [34]. Moreover, the *CYP1B1* locus, harboured DNA methylation changes associated with maternal smoking during pregnancy [33], prenatal exposure to drinking water arsenic [31] or active smoking [35]. Therefore, candidate genes at all four loci identified by this study show alterations (somatic mutations and/or DNA methylation changes) related to tobacco smoke exposure, which strongly support our statistical findings. Moreover, for three of these loci, the CpG sites associated with prenatal exposure to arsenic (*KLHL1*), smoking during pregnancy (*NIN*, *CYP1B1*) or active smoking (*CYP1B1*) were in close vicinity (from 0 to 23 kb) with the respective DNA methylation changes associated with the ELTS-interacting lead SNPs at each of these loci. We also observed that the effect alleles of 13q21 and 20p12 lead SNPs were associated with higher methylation while the effect alleles of 2p22 and 14q22 lead SNPs were associated with lower methylation. However, the relationships between genetic variation, cis-DNA methylation and cis-gene expression are quite complex [36,37] and require thorough investigations. Therefore, further asthma studies integrating genetic, epigenetic and gene expression data together with ELTS exposure in the same dataset are needed to confirm these results and uncover the underlying mechanisms.

To our knowledge, none of our findings have been previously reported by asthma GWAS (GWAS-Catalog of Published Genome-Wide Association Studies [38] and the Phenoscanner database [24]) or by asthma GWIS. One of the two suggestive interactions, previously found by

the sole GWIS of smoke exposure on childhood asthma [5], was replicated in our study (P=1.5x10^{-3} at 18p11 reported for *in utero* exposure) while the other one was not (P=0.86 at 6q26 reported for childhood exposure). None of the loci identified in our study was reported by that published study except for one SNP (rs4670230) on chromosome 2p22 that modestly interacted with *in utero* exposure (P=2.1x10^{-4}) but was not correlated with our lead signal (r^2=0). The published study discovery dataset, which underwent GWI analysis, and the current study had comparable sample sizes. However, the difference in the results might be partly due to differences in the definition of tobacco smoke exposure (the overall proportion of exposed subjects was 13% for *in utero* exposure and 51% for childhood exposure in the published study *versus* 36% for early-life exposure in this study), the outcome examined (asthma status *versus* time-to-asthma onset) and the model used for analysis (logistic regression *versus* Cox model).

Up to now, few GWIS have been conducted for asthma-related phenotypes and only one reported a genome-wide significant result for a rare variant (MAF=1.5%) interacting with dust mite exposure on lung function [39]. One of the difficulties in GWIS is the need of large scale studies to detect significant interaction, which in turn might be affected by heterogeneity in study designs, outcome and exposure definitions of the participating studies. To overcome these limitations, this study was restricted to childhood-onset asthma and we paid attention to use a definition of ELTS exposure so that exposure was likely to occur before the onset of asthma. We also checked that less than 5% of SNPs showed heterogeneity of HR estimates across studies in either ELTS$^+$ or ELTS$^-$ group (Cochran's test P-values ≤0.05), showing that potential phenotypic differences across studies induced a small amount of genetic heterogeneity and, therefore, had a minor impact on our results. We meta-analyzed all five studies to maximize statistical power. Power computation for a sample of the same size as the total sample of all five studies indicated that our GWIS had 80% power of detecting a SNPxELTS interaction hazard ratio of 1.8 or higher when the MAF is at least equal to

0.15. However, we acknowledge that this study did not have power to detect SNPs with smaller allele frequencies. Our findings were supported by the consistency of results across the five studies at all four loci. We also verified that all studies contributed to the interaction signals: for example, for the 13q21 genome-wide significant signal, the study-specific contribution (estimated by the ratio of the interaction test statistic of each study to the meta-analyzed interaction test statistic) ranged from 9% to 45%. Similar results were observed at the other three loci detected by this GWIS. Even though interactions of these three loci with ELTS exposure did not reach genome-wide significance, combining statistical results with biological and functional data greatly strengthened the evidence for the potential involvement of these loci. Moreover, our study which includes more than 8000 subjects, stands among the largest studies considered to date in GWIS of asthma phenotypes.

We detected SNPs that showed an opposite direction of effect according to exposure, as previously reported at several loci for asthma and other diseases [40]. It has been generally observed that the latter pattern of interaction is the one detected with highest power by the one degree-of-freedom SNP-by-ELTS exposure interaction test as performed in this study. We acknowledge that our study might not have enough power to detect other interaction patterns such as a genetic effect in one of the exposure groups and no effect in the other group or a SNP effect in the same direction but of different magnitude in the two groups. This is why we did not confirm the previously identified interaction of ELTS exposure with 17q12-21 variants [3] which had a higher effect in the ELTS exposed than unexposed group. However, we verified that this locus was detected by using a joint test of both SNP effect and SNP x ELTS interaction in TAO analysis (results not shown). All these results are in agreement with previous simulation studies which concluded that none of the many gene-environment-wide interaction models is universally the most powerful approach and the results of an analysis depend on the unknown underlying GxE model [41].

In conclusion, this study identified new loci interacting with ELTS exposure on childhood asthma. Candidate genes at these loci have biologically relevant functions related to tobacco smoke exposure. The colocalization of the ELTS-interacting variants with regulatory elements, their association with DNA methylation in the blood and the presence of nearby DNA methylation alterations associated with tobacco smoke exposure prompts for further epigenetic and functional studies to provide more insight into the underlying mechanisms.

ACKNOWLEDGEMENTS

We thank all participants who provided data for each study and to our valued colleagues who contributed to data collection, phenotypic characterization of the samples and genotyping.

This work was supported by the French National Agency for Research (ANR-11-BSV1-027-GWIS-AM, ANR-15-EPIG-0004-05 RESET-AID), Université Pierre et Marie Curie and Région Ile-de-France (DIM-SEnT) doctoral fellowships, the Fonds de Dotation Recherche en Santé Respiratoire. The UK Medical Research Council and Wellcome Trust (grant: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The Canada Research Chair held by C Laprise and the funding supports from Canadian Institutes of Health Research (CIHR) enabled the maintenance and continuation of the SLSJ asthma study. C. Laprise is the director of the Asthma Strategic Group of the Respiratory Health Network of the Fonds de la recherche en santé du Québec (FRSQ) and member of Allergen network. Genotyping was supported by grants from the European Commission (No. LSHB-CT-2006-018996-GABRIEL) and the Wellcome Trust (WT084703MA). GABRIELA was supported by the European Commission as part of GABRIEL (a multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community), European Commission Research Grant (LSHB-CT-2006-018996) and by the European Research Council (ERS-2009-AdG, project HERA 250268).

CONFLICT OF INTEREST

EvM received personal fees from Pharma Ventures, personal fees from Peptinnovate Ltd., OM Pharma SA, European Commission/European Research Council Executive Agency, Tampereen Yliopisto, University of Turku, HAL Allergie GmbH, Ökosoziales Forum Oberösterreich, Mundipharma Deutschland GmbH & Co. KG, outside the submitted work. All other authors have no relevant conflicts of interest

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Demenais F, Margaritte-Jeannin P, Barnes KC, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet.* 2018;50:42-53.
2. Dizier MH, Nadif R, Margaritte-Jeannin P, et al. Interaction between the DNAH9 gene and early smoke exposure in bronchial hyperresponsiveness. *Eur Respir J.* 2016;47:1072-1081.
3. 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.
4. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature.* 2007;448:470-473.
5. Scholtens S, Postma DS, Moffatt MF, et al. Novel childhood asthma genes interact with in utero and early-life tobacco smoke exposure. *J Allergy Clin Immunol.* 2014;133:885-888.
6. Vonk JM, Scholtens S, Postma DS, et al. Adult onset asthma and interaction between genes and active tobacco smoking: The GABRIEL consortium. *PLoS One.* 2017;12:e0172716.
7. 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.
8. Forno E, Lasky-Su J, Himes B, et al. Genome-wide association study of the age of onset of childhood asthma. *J Allergy Clin Immunol.* 2012;130:83-90 e84.
9. Sarnowski C, Sugier PE, Granell R, et al. Identification of a new locus at 16q12 associated with time to asthma onset. *J Allergy Clin Immunol.* 2016;138:1071-1080.
10. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol.* 2013;42:111-127.

11. Kogevinas M, Zock JP, Jarvis D, et al. Exposure to substances in the workplace and new-onset asthma: an international prospective population-based study (ECRHS-II). *Lancet.* 2007;370:336-341.
12. Ege MJ, Strachan DP, Cookson WO, et al. Gene-environment interaction for childhood asthma and exposure to farming in Central Europe. *J Allergy Clin Immunol.* 2011;127:138-144, 144 e131-134.
13. Kauffmann F, Dizier MH, Pin I, et al. Epidemiological study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy: phenotype issues. *Am J Respir Crit Care Med.* 1997;156:S123-129.
14. Laprise C. The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population. *Genes Immun.* 2014;15:247-255.
15. Winkler TW, Justice AE, Graff M, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet.* 2015;11:e1005378.
16. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539-1558.
17. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45:1238-1243.
18. Liang L, Morar N, Dixon AL, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* 2013;23:716-726.
19. Grundberg E, Small KS, Hedman AK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012;44:1084-1089.
20. Zeller T, Wild P, Szymczak S, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One.* 2010;5:e10693.

21. Consortium TG. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348:648-660.
22. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* 2016;44:D877-881.
23. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518:317-330.
24. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics.* 2016;32:3207-3209.
25. Pascoe CD, Obeidat M, Arsenault BA, et al. Gene expression analysis in asthma using a targeted multiplex array. *BMC Pulm Med.* 2017;17:189.
26. Baglole CJ, Maggirwar SB, Gasiewicz TA, Thatcher TH, Phipps RP, Sime PJ. The aryl hydrocarbon receptor attenuates tobacco smoke-induced cyclooxygenase-2 and prostaglandin production in lung fibroblasts through regulation of the NF-kappaB family member RelB. *J Biol Chem.* 2008;283:28944-28957.
27. Aromolaran KA, Benzow KA, Koob MD, Piedras-Renteria ES. The Kelch-like protein 1 modulates P/Q-type calcium current density. *Neuroscience.* 2007;145:841-850.
28. Valverde MA, Cantero-Recasens G, Garcia-Elias A, Jung C, Carreras-Sureda A, Vicente R. Ion channels in asthma. *J Biol Chem.* 2011;286:32877-32882.
29. Wylam ME, Sathish V, VanOosten SK, et al. Mechanisms of Cigarette Smoke Effects on Human Airway Smooth Muscle. *PLoS One.* 2015;10:e0128778.
30. Yu XJ, Yang MJ, Zhou B, et al. Characterization of Somatic Mutations in Air Pollution-Related Lung Cancer. *EBioMedicine.* 2015;2:583-590.

31. Rojas D, Rager JE, Smeester L, et al. Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. *Toxicol Sci.* 2015;143:97-106.
32. Gupta GD, Coyaud E, Goncalves J, et al. A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell.* 2015;163:1484-1499.
33. Joubert BR, Felix JF, Yousefi P, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet.* 2016;98:680-696.
34. Stueve TR, Li WQ, Shi J, et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Hum Mol Genet.* 2017;26:3014-3027.
35. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet.* 2016;9:436-447.
36. Chen L, Ge B, Casale FP, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell.* 2016;167:1398-1414 e1324.
37. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49:131-138.
38. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42:D1001-1006.
39. Forno E, Sordillo J, Brehm J, et al. Genome-wide interaction study of dust mite allergen on lung function in children with asthma. *J Allergy Clin Immunol.* 2017;140:996-1003 e1007.
40. Ober C, Vercelli D. Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet.* 2011;27:107-115.
41. Gauderman WJ, Mukherjee B, Aschard H, et al. Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. *Am J Epidemiol.* 2017;186:762-770.

42. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016;17:61.

Figure Legends

FIGURE 1. Flow chart of the genome-wide interaction study of early-life tobacco smoke (ELTS) exposure on time-to-asthma onset in childhood**.** This figure shows the number of subjects (N) by study in each ELTS exposure stratum.

FIGURE 2. Manhattan plot of the gene-environment-wide interaction meta-analysis for time-to-asthma onset in childhood**.** The x axis represents chromosomal location and the y axis represents $-\log_{10}(P)$ for tests of interaction between individual SNPs and ELTS exposure on time-to-asthma onset. The red horizontal line denotes the genome-wide significant threshold of $P=5\text{x}10^{-8}$ and the black dashed horizontal line a suggestive threshold of $P=5\text{x}10^{-6}$.

FIGURE 3. Forest Plot of the 13q21 lead SNP (rs7334050) according to ELTS exposure. Hazard ratios (HR) and 95% Confidence Intervals (CI) are plotted by study and by ELTS exposure stratum. The combined HR estimates over all five studies in each stratum and the combined SNP by ELTS exposure interaction HR are plotted as a diamond.

TABLE 1. Loci showing interaction ($P \leq 5x10^{-6}$) with early-life tobacco smoke (ELTS) exposure on time-to-asthma onset in childhood

CHR	SNP	Position (kb, build 37)	Closest Gene (distance of SNP to gene in kb)	E/R	EAF	SNP x ELTS interaction[a]		SNP effect in subjects exposed to ELTS		SNP effect in subjects non-exposed to ELTS	
						HR_{Int} [95% CI]	2-sided *P*	HR_{ELTS+} [95% CI]	2-sided *P*	HR_{ELTS-} [95% CI]	2-sided *P*
2p22	rs232542	38 328	*CYP1B1* (25)	C/T	0.69	1.33[1.18-1.50]	$4.1x10^{-6}$	1.18[1.07-1.30]	$7.0x10^{-4}$	0.89[0.82-0.96]	$1.7x10^{-3}$
13q21	**rs7334050**	**70 645**	***KLHL1* (0)**	**G/T**	**0.14**	**1.58[1.34-1.86]**	**$4.3x10^{-8}$**	1.34[1.19-1.52]	$2.6x10^{-6}$	0.85[0.76-0.95]	$3.2x10^{-3}$
14q22	rs7493885	51 317	*NIN* (20)	G/T	0.27	1.37[1.20-1.56]	$2.9x10^{-6}$	1.16[1.05-1.29]	$2.7x10^{-3}$	0.85[0.78-0.93]	$2.5x10^{-4}$
20p12	rs13037508	14 928	*MACROD2* (0)	A/T	0.63	1.48[1.27-1.72]	$4.9x10^{-7}$	1.20 [1.06-1.35]	$3.6x10^{-3}$	0.81[0.74-0.89]	$7.9x10^{-6}$

CHR, chromosome; SNP, single nucleotide polymorphism; E, effect allele / R, Reference allele; EAF, effect allele frequency; HR, hazard ratio; CI: confidence interval. In bold, SNP showing interaction with ELTS at the genome-wide significance level of $P \leq 5x10^{-8}$.

[a]The interaction effect size between SNP and ELTS exposure was estimated as the difference between the $ELTS^+$ and $ELTS^-$ combined SNP effect sizes obtained from the fixed-effects meta-analyses of the five studies in each $ELTS^{+/}ELTS^-$ stratum (as shown in Figure 1).

TABLE 2. Functional annotations of SNPs at significant (13q21) and suggestive (20p12, 14q22 and 2p22) loci interacting with ELTS exposure on time-to-asthma onset in childhood

Locus	SNP (r^2 with lead SNP)	Position (kb)	Regulatory elements			DNA methylation			
			Histone marks[a]	DNase I hyper sensitive sites[a]	Transcription factor binding sites	CpG site (position, kb)	Tissue	2-sided *P* [b] (Effect allele: Direction of association)	Study
13q21	**rs7334050**	70 645	No	No	FOXL1, PBX-1, POU1F1	cg14273027 (70 682)	Blood	$1.1x10^{-6}$ (G: positive)	BIOSQTL[37]
	rs73214641 (0.87)	70 653	Yes (fetal lung)	Yes (blood stem cells)					
20p12	**rs13037508**	14 928	Yes (lung, fetal lung)	Yes (stem cells)	BCL CEBPD, PAX-5 FOXD3, HDAC2, IRF, POU2F2, POU3F2, STAT, P30O, RXRA	cg04470754 (14 904)	Blood	$6.4x10^{-57}$ (A: positive)	BIOSQTL[37]
	rs2423868 (0.71)	14 929	Yes (lung)	Yes (lung, fetal lung)	CTCF, RAD21, RFX5				
14q22	**rs7493885**	51 317	Yes (blood stem cells)	No		cg25597366 (51 313)	Blood	$1.2x10^{-58}$ (G: negative)	BIOSQTL[37]
							Neutrophils	$7.1x10^{-15}$ (G: negative)	BLUEPRINT[36]
							Monocytes	$5.8x10^{-13}$ (G: negative)	BLUEPRINT[36]
							T cells	$8.3x10^{-9}$ (G: negative)	BLUEPRINT[36]
							Cord blood	$4.7x10^{-28}$ (G: negative)	ALSPAC[42]
							Blood	$3.6x10^{-46}$ (G: negative)	ALSPAC[42]

	rs8020067 (1.0)	51 318	Yes (T cells, blood stem cells)	No	AHR::ARNT, ARNT, HBP1, PAX-4				
	rs4901062 (0.99)	51 315	Yes (lung, fetal lung)	No	CTCF, PITX2, RAD21				
2p22	**rs232542**	38 328	Yes (fetal lung, blood cells)	No	YY1	cg02486145 (38 334)	Blood	$4.7x10^{-179}$ (C: negative)	BIOSQTL[37]
							Neutrophils	$1.9x10^{-11}$ (C: negative)	BLUEPRINT[36]
							Monocytes	$1.0x10^{-6}$ (C: negative)	BLUEPRINT[36]
							T cells	$2.9x10^{-11}$ (C: negative)	BLUEPRINT[36]
							Cord blood	$5.4x10^{-33}$ (C: negative)	ALSPAC[42]
							Blood	$4.4x10^{-34}$ (C: negative)	ALSPAC[42]
	rs232540 (0.99)	38 329	Yes (stem cells)	Yes (blood cells)	CTCF, ELTS, EVI-1, MYF, PEBP, RAD21, SMC3, TAL1				

r^2, linkage disequilibrium measure between a SNP and the lead SNP (in bold) at a locus; kb, kilobase (build 37)

[a] Histone marks represent promoters or enhancers. When regulatory elements colocalized with SNPs in tissues biologically relevant to asthma (blood cells and/or lung tissue), "Yes" is indicated in the corresponding column (data retrieved from Haploreg v4.1 [22])

[b] *P* is the *P*-value for association of SNP with DNA methylation levels at a CpG site (data retrieved from the Phenoscanner database [24]).