

GSDMA drives the most replicated association with asthma in naïve CD4⁺ T cells

Anne-Marie Madore¹, Lucile Pain¹, Anne-Marie Boucher-Lafleur¹, Jolyane Meloche¹, Andréanne Morin^{2,3}, Marie-Michelle Simon⁴, Bing Ge^{2,4}, Tony Kwan^{2,4}, Warren A. Cheung^{2,4}, Tomi Pastinen^{2,4,5}, Catherine Laprise^{*1,6}.

¹Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, Quebec, G7H 2B1, Canada; ²Department of Human Genetics, McGill University, Montreal, Quebec, H3A 0C7, Canada; ³Department of Human Genetics, University of Chicago, IL 60637, USA; ⁴McGill University and Génome Québec Innovation Centre, Montreal, Quebec, H3A 0G1, Canada; ⁵Center for Pediatric Genomic Medicine, Kansas City, MO 64108, USA; ⁶Centre intersectoriel en santé durable, Université du Québec à Chicoutimi, Saguenay, Quebec, G7H 2B1, Canada

***Corresponding author**

Catherine Laprise: catherine.laprise@uqac.ca

Abstract

Background The 17q12-21 locus is the most replicated association with asthma. However, no study had described the genetic mechanisms underlying this association considering all genes of the locus in immune cell samples isolated from asthmatic and non-asthmatic individuals. **Objective** This study takes benefit of samples from naïve CD4⁺ T cells and eosinophils isolated from the same 200 individuals to describe specific interactions between genetic variants, gene expression and DNA methylation levels for the 17q12-21 asthma locus. **Methods and Results** After isolation of naïve CD4⁺ T cells and eosinophils from blood samples, next generation sequencing was used to measure DNA methylation levels and gene expression counts. Genetic interactions were then evaluated considering genetic variants from imputed genotype data. In naïve CD4⁺ T cells but not eosinophils, 20 SNPs in the fourth and fifth haplotype blocks modulated both *GSDMA* expression and methylation levels, showing an opposite pattern of allele frequencies and expression counts in asthmatics compared to controls. Moreover, negative correlations have been measured between methylation levels of CpG sites located within the 1.5 kb region from the transcription start site of *GSDMA* and its expression counts. **Conclusion** Availability of sequencing data from two key cell types isolated from asthmatic and non-asthmatic individuals allowed identifying a new gene in naïve CD4⁺ T cells that drives the association with the 17q12-21 locus, leading to a better understanding of the genetic mechanisms taking place in it.

Key words

GSDMA, naïve CD4⁺ T cells, eosinophils, asthma, 17q12-21, gene expression, methylation, genetic variants, sequencing

Introduction

The 17q12-21 locus harboring *ORMDL3* sphingolipid biosynthesis regulator 3 (*ORMDL3*) and Gasdermin B (*GSDMB*) genes exhibits one of the most significant and replicated associations with asthma [MIM: 600807].¹⁻¹² Even though this locus was associated with this disease without considering any specific age of onset as it was done using the Saguenay–Lac-Saint-Jean (SLSJ) asthma familial cohort before,¹¹ it is principally with the condition of childhood.^{13, 14} Several studies explored the complex interactions between SNPs, gene expressions and epigenetic modifications at this locus, primarily focusing on the levels of expressed *ORMDL3* and *GSDMB*.^{12, 15-25} Given that DNA methylation and gene expression signals are cell-type specific,^{26, 27} isolated materials such as lymphocytes from blood¹⁷⁻²¹ and, to a lesser extent, bronchial epithelial cells from biopsies of bronchi or bronchoalveolar lavages^{15, 16} have been used in a number of studies to investigate such interactions.

The study published by Schmiedel and colleagues described *ORMDL3* and *GSDMB* expression in the largest diversity of isolates, including eight different lymphocyte types, monocytes and dendritic cells.²⁰ They described genetic variants and epigenetic marks (H3K27ac, CTCF binding and DNase I hypersensitivity sites [DHS]), which regulated *ORMDL3* expression in various immune cell types and pinpointed T lymphocytes to be key effectors in this locus. However, this and further studies did not look at the underlying genetic mechanisms and the interactions among all the other genes in the locus, and they did not include individuals with asthma to decipher the disease-specific processes. Moreover, there is a lack of data from isolated granulocytes, even though eosinophils are well known key cells in asthma physiopathology.²⁸

Furthermore, although DNA methylation profiles are key epigenetic mechanisms to understanding the association between 17q12-21 locus and asthma, they have been less studied than expression levels so far, and they were assessed with methods targeting specific methylation sites (pyrosequencing and microarrays).^{29, 30} Recently, a methylome bisulfite sequencing capture panel was designed to target

regulatory regions specific to immune cells with the consideration on DHS or active chromatin sites, hypomethylated footprints, and autoimmune SNPs from genome-wide association study (GWAS) catalog,³¹ giving a better coverage of the 17q12-21 locus compared to widely used DNA methylation arrays (e.g. Illumina methylation beadchips).

This publication addresses the aforementioned missing components of earlier studies by using an integrative multi-omic approach to characterize the functional genetic mechanisms at this locus in eosinophils and naïve CD4⁺ T cells. Both were selected because of their important role regarding the immune response in the asthma pathogenesis, and according to previous results of methylation analyses with the SLSJ cohort; a French-Canadian one located in the northeastern part of the Quebec province.^{20, 28, 32, 33} Naïve CD4⁺ T cells and eosinophils were obtained from asthmatic and non-asthmatic individuals who were participants in this cohort.³⁴ This study combines genomic (asthma-associated SNPs), epigenomic, and transcriptomic (DNA methylation levels and gene expression counts from next generation sequencing) data to better decipher the genetic mechanisms underlying the association between the 17q12-21 locus and this disease.

Subjects and Methods

Figure 1 shows the different types of data used in this study regarding the sampling and sizing (number of samples, SNPs, CpGs methylation levels or gene expression counts) accessible for each analysis performed considering quality control filtering steps and covariates availability.

SLSJ asthma cohort

The complete description of recruitment and clinical evaluation carried out with the SLSJ cohort can be found in Laprise et al.³⁴ Briefly, families were included in the study through probands ascertained and the general and respiratory health of all individuals were evaluated using a standardized questionnaire and pulmonary function tests according to the American Thoracic Society guidelines.³⁵

The SLSJ asthma cohort includes 1,394 individuals distributed in 271 families from which 1,214 individuals (representing 254 families) have genotypic information and 1,200 of them have imputed data available. See Table 1 for the phenotypic characteristics of the 1,200 individuals included in the association analysis with asthma. From these individuals, 215 were recruited for isolating naïve CD4⁺ T cells and eosinophils to measure methylation and expression levels by next generation sequencing. After being isolated and the quality control step (see below the isolation section), 173 naïve CD4⁺ T cell and 146 eosinophil samples were available for analyses. Shown in Table 1, their phenotypic characteristics. The *Centre intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean* ethics committee approved the study, and all subjects gave informed consent.

Genetic data

The genotypes for the SLSJ asthma cohort were obtained from Illumina Human610-Quad BeadChip. The pre-phasing step for the imputation process was performed with the Shapeit2 Software using the duoHMM method that combines estimated haplotypes with pedigree information to take advantage of known structure related to the latter.³⁶ Impute2 Software was used for imputation³⁷ with the 1,000

Genome Project database (phase 3)³⁸ and the UK10K one as reference samples. See Supplemental Data for details about quality control steps. Based on the review published by Stein et al.,¹² the 17q12-21 locus was defined as chr17:37,826,875-38,134,519 for our analyses. This region includes 11 sequences coding for microRNAs or genes (Post-GPI attachment to proteins 3 [*PGAP3*], Erb-B2 receptor tyrosine kinase 2 [*ERBB2*], microRNA [miR]-4728, Migration and invasion enhancer 1 [*MIEN1*], Growth factor receptor-bound protein 7 [*GRB7*], IKAROS Family Zinc Finger 3 [*IKZF3*], Zona Pellucida-Binding Protein 2 [*ZPBP2*], Gasdermin A [*GSDMA*] and B [*GSDMB*], *ORMDL3*, as well as Leucine-Rich Repeat-Containing 3C [*LRRC3C*]). *ORMDL3*, *GSDMB* and *ZPBP2* were the most replicated asthma associated genes. After quality filtering, association with asthma was evaluated for 762 SNPs for this specific region in 1,180 DNA samples.

Isolation of naïve CD4⁺ T cells and eosinophils

Naïve CD4⁺ T cells, isolated in this project, were defined as units with CD3⁺, CD4⁺, CD45RA⁺, CD45RO⁻ markers. These and eosinophils were from 200 ml blood samples from 215 individuals, a subset of the SLSJ asthma cohort, using the EasySep Human Naïve CD4⁺ T Cell Isolation Kit (STEMCELL Technologies, Vancouver, BC, Canada) and anti-CD16 MicroBeads respectively according to protocols described (Miltenyi Biotec, Auburn, CA, USA)^{39, 40}. A description of the different steps can be found in the Supplemental Data.

Targeted bisulfite sequencing (DNA methylation levels)

Bisulfite sequencing was done for a custom methyl capture panel of identified functional immune genetic regions as previously described.^{31, 41} These covered a total of 4,609,564 CpGs for a sum of 822,884 regions and 119,089,296 bp sequenced.³¹ The MCC-Seq methylation was developed and optimized in collaboration with R&D at Roche NimbleGen. Details of the protocols used can be found in the Supplemental Data. After quality control filtering, methylation data were available for samples from 192 naïve CD4⁺ T cells and 183 eosinophils and 1,320 CpGs for analyses of the former type as well as 1,055 CpGs of the latter one.

RNA sequencing (gene expression counts)

RNA sequencing was performed at the McGill University and Génome Québec Innovation Centre with the Illumina TruSeq Stranded Total RNA Sample Prep Kit (Illumina, CA, USA). A detailed description of the method used can be found in the Supplemental Data. After quality control filtering and given that the availability of all covariates included in statistical models, transcriptomic data were accessible for 159 naïve CD4⁺ T cell and 139 eosinophil samples. Using as a cut-off ten counts in totality while considering all samples together, a total of seven genes were expressed by naïve CD4⁺ T cells and six by eosinophils.

Statistical methods

Genetic association and definition of haplotype blocks

The significance threshold was calculated with Nyholt's method⁴² according to the number of independent SNPs in the 17q12-21 asthma locus and was set to 0.001. In this work, we carried out the MFQLS test implemented in the Workbench for Integrated Superfast Association study with Related Data (WISARD) toolkit, a quasi-likelihood method of estimation extended from the Cochran-Armitage test, suited for familial cohorts considering kinship coefficients, and can be used to associate multiple phenotypes with covariates.^{43, 44} The analyses were made with the asthma phenotype using sex, age and smoking status (a never or an ever smoker) as covariates.

The Haploview Software was used to delimit the haplotype blocks of the 17q12-21 asthma locus with an algorithm taken from Gabriel et al.^{45, 46} Methylation blocks for any cell type were also established using r^2 values between each adjacent CpGs⁴⁷ in a similar way haplotype blocks were defined with linkage disequilibrium ones (D' and r^2). For methylation, they were identified as two or more CpGs for which all pairwise r^2 value is equal or above 0.5 as described by Guo and colleagues.⁴⁷

Estimation of cell proportion and surrogate variables

The cell-type contribution approach was used to estimate the probable variation between the purity of samples from naïve CD4⁺ T cells and eosinophils and the values calculated, and the latter was employed as covariates in our analysis models. The method developed by Houseman⁴⁸ and implemented in the RnBeads R package was applied using the whole genome methylation sequencing data. The reference samples used for hematocytes included ten types: CD14⁺, CD19⁺, CD4⁺, CD8⁺, eosinophils, granulocytes, neutrophils, PBMCs and white blood cells.²⁷ Moreover, in order to take into account hidden confounders and relatedness between samples, surrogate variables were calculated using the R package sva.⁴⁹ They were by considering all covariates included in models analyzed.

Analyses of gene expression counts and DNA methylation levels

Gene expression counts were normalized on the library size and log₂ transformed while DNA methylation levels had undergone a logit transformation using the R packages DESeq2 and car respectively.⁵⁰ Correlations between gene expression counts from the two cell types, or from within a unique one, were performed using Pearson's r. Correlations between gene expression counts and DNA methylation levels were assessed with this and also with a negative binomial model (using normalized-only values for the former) considering age at sampling, sex, smoking status (a never or an ever smoker), proportion of naïve CD4⁺ T cells or eosinophils, according to the origin of data analyzed and surrogate variables as covariates. The differences, in gene expression counts (normalized-only data) and DNA methylation levels (non-transformed percentages) in accordance with the asthma status, were assessed using a negative binomial and a binomial model respectively, both considering the same covariates as for the previous analyses. False discovery rate (FDR) method was applied to take into account multiple testing.

QTL analyses were performed using the package MatrixEQTL in R⁵¹ for the gene expression counts (expression quantitative trait loci: eQTLs) and for the DNA methylation levels (methylation quantitative trait loci: mQTLs) with same covariates as previously. In mQTL analyses, CpGs were considered to be

linked specifically if they were located within the gene body or the region within 1.5 kb from the transcription start site (TSS). As information from these focused on genetic variation in haplotype blocks interacting with the expression or methylation, significance threshold was calculated using Bonferroni correction considering the number of these structures included and the one of genes expressed by each cell type. The significance threshold was set to 0.001 for naïve CD4⁺ T cells and eosinophils ($0.05/[7*7]$ and $0.05/[7*6]$ respectively).

Several analyses were performed on stratified samples according to asthma status or on affected and unaffected mother-child duos. In accordance with the small number of samples included, 1,000 analyses were performed after permutations between the asthmatic and non-asthmatic individuals to estimate the chance each result has to be a false positive.

Analyses using the mediation package in R were performed to describe more accurately the link between DNA methylation levels measured within the 1.5 kb from the *GSDMA* TSS region, asthma-associated SNPs in the fourth and fifth haplotype blocks identified in the SLSJ cohort and its gene expression counts. In the models used, methylation was considered as the mediator of the SNP-expression association of *GSDMA*. The same covariates as usual were applied to them, and 1,000 bootstraps were run to estimate the confidence intervals.

Results

This project used genomic, epigenomic and transcriptomic data to better understand the genetic interactions underlying the association between a specific region of the 17q12-21 locus (chr17:37,826,875-38,134,519) and asthma in two immune cell types (naïve CD4⁺ T cells and eosinophils). SNPs, gene expression and methylation data were first analyzed separately to establish and compare their profiles in naïve CD4⁺ T cells (CD3⁺, CD4⁺, CD45RA⁺, CD45RO⁻) and eosinophils. Then, genetic interactions were assessed using eQTL and mQTL analyses and correlations between DNA methylation levels and gene expression counts (Figure 1). This figure also indicates the number of samples as well as the genetic data used for each analysis after quality control filtering.

Haplotype blocks in the 17q12-21 locus are associated with asthma in the SLSJ cohort

The analysis of linkage disequilibrium between 762 SNPs in the 17q12-21 locus identified seven haplotype blocks in the SLSJ asthma cohort compared to the two from 17 SNPs in the European population¹² (Figures S1-S2). Of the 248 SNPs associated with asthma with $p < 0.05$, 20 remained of statistical significance after correction for multiple testing ($p < 0.001$) and all were located in the fourth as well as the fifth haplotype blocks (Table S1). SNP rs869402 in the fourth block and rs9303281 in the fifth one exhibited the most statistically significant associations ($p = 5.35e-04$ and $2.60e-04$, respectively). The analysis including only individuals with age of onset <17 years old as asthmatic individuals gave very similar results ($p = 1.73e-04$ and $p = 2.01e-04$ for the same two SNPs; see Table S1).

Gene expression and DNA methylation profiles harbored by naïve CD4⁺ T cells and eosinophils in the 17q12-21 asthma locus are different

Naïve CD4⁺ T cells and eosinophils express six genes in common at the 17q12-21 asthma locus

Gene expression analyses demonstrated that of the eleven ones located in the 17q12-21 locus, seven were detectable in naïve CD4⁺ T cells (*PGAP3*, *ERBB2*, *MIEN1*, *IKZF3*, *GSDMB*, *ORMDL3* and *GSDMA*) and six in eosinophils (*PGAP3*, *ERBB2*, *MIEN1*, *IKZF3*, *GSDMB* and *ORMDL3*). Detailed

results for correlations between gene expression counts between the two cell types or within each one were available in Tables S2-S4. In both cell types, the most significant pairwise correlation was between *GSDMB* and *ORMDL3* (naïve CD4⁺ T cells: $r = 0.668$ [FDR <1.0e-20]; eosinophils: $r = 0.888$ [FDR <1.0e-20]) (Figure S3). No genes were differentially expressed in analyses that compare cells isolated from individuals with and without asthma (Table S5).

Methylation profiles in the 17q12-21 asthma locus are different in naïve CD4⁺ T cells and eosinophils

DNA methylation levels were measured in naïve CD4⁺ T cells and eosinophils in order to compare their profiles and determine if levels of CpGs in this region are associated with asthma. Using a custom sequencing panel, a total of 1,320 and 1,055 CpGs have been measured after quality control filtering for naïve CD4⁺ T cells and eosinophils respectively compared to the 166 CpGs evaluated by the Infinium Human Methylation 450k BeadChip in this same locus. Of these 166 CpGs, 156 and 115 were among the 1,320 and 1,055 CpGs included in this study on naïve CD4⁺ T cells and eosinophils, respectively. In the same way that haplotype blocks were evaluated for SNPs of the 17q12-21 asthma locus, r^2 were computed for each neighboring pairs of CpGs in naïve CD4⁺ T cells and eosinophils separately to establish cell-type-specific methylation patterns in this region as previously described by Guo and colleagues⁴⁷ (Figure 2). Some differences were observed between the two cell types particularly within or near 1.5 kb regions from the gene TSS. The most obvious one was in *IKZF3* as correlation values were between 0.20 and 0.74 (mean $r^2 = 0.57$) in naïve CD4⁺ T cells compared to 0.02 and 0.55 (mean $r^2 = 0.19$) in eosinophils. Moreover, the *IKZF3* promoter region included three 68–90 bp long methylation blocks encompassing four, five and six CpGs respectively in naïve CD4⁺ T cells compared to only one block of two CpGs in eosinophils (Figure 2). The methylation levels were compared between asthmatics and non-asthmatics individuals in naïve CD4⁺ T cells and eosinophils (Table S6). Eighty-six CpGs were associated with asthma in naïve CD4⁺ T cells (FDR 5%) of which, six were located within the 1.5 kb region from four gene TSSs (*ERBB2*, *PGAP3*, *IKZF3* and *ORMDL3*). Sixty-three CpGs were differentially methylated in eosinophils of which, 11 CpGs were located within five TSS regions (*ERBB2*, *PGAP3*, *IKZF3*, *GSDMB* and *ORMDL3*).

Associations between genetic, epigenetic and transcriptomic data at the 17q12-21 locus are specific to naïve CD4⁺ T cells and eosinophils

Several genetic interactions were analyzed to better understand the mechanisms underlying the association between the 17q12-21 locus and asthma. CpGs located within the 1.5 kb region from TSS and gene body were investigated separately given that they usually have opposite impacts on expression.⁵²

SNPs of the 17q12-21 locus show different associations with gene expression counts (eQTLs) or DNA methylation levels (mQTLs) in naïve CD4⁺ T cells and eosinophils

In order to measure genetic interactions specific to asthma, disease-associated SNPs in the 17q12-21 locus were tested for association with gene expression and methylation levels at CpGs located within 1.5 kb region from TSSs. In naïve CD4⁺ T cell samples, 78 eQTLs in *IKZF3*, *GSDMB*, *ORMDL3* and *GSDMA* and 58 mQTLs in *MIEN1* and *GSDMA* were observed (Figure 3a). In eosinophil ones, 21 eQTLs in *GSDMB* and *ORMDL3* and no mQTL were detected (Figure 3b).

Genetic interactions observed for asthma-associated SNPs were compared to mQTLs for methylation sites located within gene bodies or intergenic regions and to results considering all SNPs in the 17q12-21 locus for comprehensive and complementary purposes. The genes and the numbers of QTLs varied depending on the location of the CpGs (1.5 kb region from TSS or gene body) and the type of SNPs (asthma-associated or not) investigated. See Tables S7 to S10 and Figures S4 and S5 for all the significant results.

Different DNA methylation-expression associations are observed in naïve CD4⁺ T cells and eosinophils at the 17q12-21 locus

Regression analyses detected correlations between DNA methylation levels and gene expression counts in naïve CD4⁺ T cells (Table 2). Gene expression counts of *GSDMA* was negatively correlated

with methylation levels of two CpGs located within its 1.5 kb region from TSS ($r = -0.238$ [FDR = 0.039] and -0.255 [FDR = 0.028]). Furthermore, those of *ORMDL3* had a positive correlation with one CpG within its 1.5 kb region from TSS ($r = 0.308$ [FDR = $7.55e-05$]). The expression of *ERBB2*, *IKZF3*, *ORMDL3* and *GSDMA* correlated with methylation levels of CpGs within their gene body. All significant correlations also including intergenic methylation sites can be seen in Table S11. In eosinophils, significant correlations were also uncovered for CpGs located in the gene body of *ERBB2* and *IKZF3* (Tables 3 and S12) but none with CpGs within the 1.5 kb from TSSs.

***GSDMA* expression is associated with genetic variations and methylation levels of the 17q12-21 locus in naïve CD4⁺ T cells**

Different analyses in this study showed that genetic, epigenetic and expression of *GSDMA* were interacting. First, all asthma-associated SNPs (two in the fourth haplotype block and 18 in the fifth one) modulated both the gene expression counts and the DNA methylation levels within the 1.5 kb region from TSS of *GSDMA* (Figures 3a and 4). All these SNPs formed mQTLs with the same two methylation sites located 18 bp (chr17:38,119,207) and 27 bp (chr17:38,119,198) upstream of *GSDMA* TSS (Figure 4d and 4e). The two CpGs were positively correlated ($r^2 = 0.710$ and FDR = $1.23e-47$), and each was negatively with *GSDMA* expression ($r = -0.238$ with FDR = 0.039 and $r = -0.255$ with FDR = 0.028; Figure 4f). mQTLs were also observed for CpGs located in *GSDMA* gene body and one of these three DNA methylation sites has its level correlated with its expression ($r = -0.192$ with FDR = 0.031).

Since the association between genetic variants and the disease occurring at the 17q12-21 locus is mainly driven by individuals with early-onset asthma,^{13, 14} correlation analyses were performed with DNA methylation levels from different mother-child duos for the two CpGs included in the significant mQTLs at *GSDMA* TSS. Analyses were between unaffected mother and affected child duos, all asthma status mother and affected child duos and affected mother and affected child duos to demonstrate the possible heritability of these DNA methylation levels. Also, they showed an increase

of the correlation values for the two CpGs with Pearson's r of 0.068 to 0.412 for the 17:38,119,198 CpG and of 0.112 to 0.238 for the 17:38,119,207 CpG in the case of the duos with unaffected mother and affected child and those with affected mother and affected child respectively (Figure S6).

Mediation analyses were performed to find out the proportion of the association between methylation levels near *GSDMA* TSS and its expression signal modulated by the 20 SNPs included in both eQTLs and mQTLs. These demonstrated no significant model (data not shown).

Finally, analyses of genetic interactions (correlation between methylation and expression levels of *GSDMA*, eQTLs and mQTLs) were conducted on individuals with and without asthma separately, in order to decipher if the latter is driven by subjects suffering from this disease. Neither of the correlations between methylation and expression, nor mQTL analysis provided significant results after permutations. However, eQTL analysis including SNPs located in the fourth and fifth haplotype blocks revealed 15 eQTLs to be significantly associated with *GSDMA* expression in individuals with asthma and 19 eQTLs in controls (See Table S13 for eQTL results regarding these subjects). Interestingly, minor allele frequencies of the asthma-associated SNPs were lower in asthmatic individuals for all SNPs compared to controls and expression level was higher in subjects with the disease in comparison with the others, showing an opposite pattern in the two groups (Figure 5).

Discussion

In this study, we described genetic, epigenetic and transcriptomic profiles as well as their interactions in a specific region (chr17:37,826,875-38,134,519) of the 17q12-21 asthma locus in naïve CD4⁺ T cells and eosinophils. The limit of this region was selected according to the review by Stein et al.¹² to include genes from *PGAP3* to *GSDMA*. The feasibility to measure and analyze various data types of cells obtained from the same individuals increase the chance of detecting valid genetic interactions. Furthermore, using data from imputed SNPs combined with next generation sequencing for individuals suffering from asthma or not allowed for a more complete description regarding the genetic interactions specific to this disease phenotype in these two cell types, an approach that was never used for this locus.

Genetic analysis performed in this study identified seven haplotype blocks in the 17q12-21 locus in the SLSJ cohort, in which 20 SNPs located in the fourth and fifth ones were significantly associated with asthma after correction for multiple testing. Among them, eight were already in the literature.^{6, 11, 53-55} The haplotype blocks observed in the SLSJ cohort were similar to those described previously for the European population¹² except that the second European one was represented by the third to the fifth in the SLSJ cohort. Given that imputed data were used in the SLSJ asthma cohort, the larger number of SNPs analyzed (762 compared to 17) improved the precision in linkage disequilibrium and led to the definition of more blocks and to a better accuracy in interpreting eQTL and mQTL results.

For cell-type-specific expression at this locus, seven genes (*PGAP3*, *ERBB2*, *MIEN1*, *IKZF3*, *GSDMB*, *ORMDL3* and *GSDMA*) were in naïve CD4⁺ T cells as previously also observed in tissue-resident memory ones or lymphoblastoid cell lines (LCLs).^{12, 19, 20, 29} In eosinophils, only one study has explored the gene expression.¹⁷ Thus, our findings show that six of the seven genes expressed by naïve CD4⁺ T cells were as well in eosinophils (excluding *GSDMA*), which significantly contributes to the research field. Our results also confirm the strong correlation observed between *GSDMB* and *ORMDL3* expression,²⁰ independent of the asthma status. Regarding the methylation profiles, the difference at

3.2 kb upstream of the *IKZF3* TSS with blocks specific to naïve CD4⁺ T cells may be of interest since this gene is known to regulate the lymphocyte development⁵⁶. Interestingly, none of the CpGs correlated in this region are included in the Infinium Human Methylation 450k BeadChip, and thus these results were obtained thanks to the custom capture panel used in this study.

To better understand the genetic interactions between SNPs, gene expression counts and DNA methylation levels, eQTL and mQTL analyses for each cell type were carried out. SNPs associated with the asthma phenotype in the SLSJ cohort were considered in order to identify QTLs, which are more likely to be involved in the disease. The two cell types exhibited different eQTL and mQTL patterns. Interestingly, both the expression and methylation levels of *GSDMA* in naïve CD4⁺ T cells were modulated by all asthma-associated SNPs (fourth and fifth haplotype blocks) and all measured mQTLs considering the 1.5 kb region from the TSS converged to two CpGs. The methylation profile at these two sites correlated with one another and both had a negative correlation with *GSDMA* expression counts; similar findings have been previously reported in LCLs.²⁴ This gene is involved in pyroptosis, a process that induces cell death and the subsequent massive release of cellular contents and triggers strong inflammation.^{57, 58} Thus, the approach used in this study that combined various omic data on asthma phenotype pinpointed *GSDMA* as a new gene of interest in this locus in naïve CD4⁺ T cells.

The association between asthma and the 17q12-21 locus was first established in childhood asthma⁶ and was then confirmed to be mainly driven by individuals with the early-onset form of the disease.^{13, 14} Using a familial cohort with a mean age of onset of seven and 22 years old for probands and other family members respectively can be considered a caveat for the study of this locus, principally since gene expression counts and DNA methylation levels are changing over time. However, the genetic association was replicated in a previous study performed with the SLSJ asthma cohort¹¹ and analyses using the imputed data in this study show similar results when taking account of all samples or only those with age of onset under 17 years old. Moreover, it was demonstrated that methylation marks

can be transmitted to other generations.⁵⁹ Analyses done on DNA methylation and mother-child duos in this study showed a marked increase of the correlation values between these levels in affected mother and affected child duos compared to the ones in the unaffected mother and affected child duos. Although the sample size was small in these stratified analyses, the above-mentioned increases strengthened the hypothesis that these two CpGs may be inherited and thus, the observation made in this study on *GSDMA* expression counts and methylation levels are of interest, even considering the age-specific association of this locus.

In this study, several observations previously described were reported; eQTLs between rs7216389, rs2290400 and rs4795405 as well as *ORMDL3* expression,^{18, 19, 21, 60, 61} mQTLs between rs7216389 and rs4065275 as well as methylation within the 1.5 kb region from *ORMDL3* TSS.²⁹ Most interestingly, using the sequencing data for methylation and expression measures combined with the imputed genetic one provided new insights for the interactions in the 17q12-21 locus. This study design enabled the investigation of almost all potential genetic interactions with DNA methylation levels or gene expression counts in this locus for the two cell types. Only few studies have looked at all the genes in this locus for eQTL and even fewer ones for mQTL interactions.^{12, 19} The analyses considering all SNPs in the region also highlighted the important difference in the number of eQTLs and mQTLs observed in naïve CD4⁺ T cells compared to eosinophils, suggesting more complex interactions in the former type.

The investigation of asthma-associated SNPs in the eQTL and mQTL analyses allowed the detection of genetic interactions specific to this disease. Logically, its next step was to look for correlations, eQTLs and mQTLs in asthmatic and non-asthmatic individuals separately as an attempt to assess for a specific contribution of subjects suffering from the disease to the genetic interactions observed, focusing on the *GSDMA* gene. Unfortunately, it was not possible to conclude from the mQTLs data that the asthma status contributed to these interactions as p values resulting from permutations were not significant. However, it is interesting to note that the data in this study showed that eQTLs between

SNPs located in the fourth and fifth haplotype blocks and *GSDMA* expression counts were linked to the asthma status of the individuals. In fact, 15 and 19 significant eQTLs were found in asthmatic individuals and controls respectively, showing a strong link between disease-associated SNPs and *GSDMA* expression counts in the two groups. More importantly, the pattern was opposite in both, the minor allele frequency is being lower in affected individuals for all SNPs and the expression being higher. This emphasizes the potential interest of this gene in the study of asthma.

To conclude, it will clearly be of interest to perform this approach on other immune cell types involved in asthma pathophysiology to obtain a more complete picture. Nevertheless, by using sequencing data to measure DNA methylation level and gene expression counts on naïve CD4⁺ T cells and eosinophils from the same individuals (asthmatic and non-asthmatic), this study pinpointed *GSDMA* as a key new actor in the 17q12-21 locus and further confirmed the role of naïve CD4⁺ T cells in the pathogenesis of this disease.

Supplemental Data

Supplemental Data include 6 figures and 13 tables.

Declaration of Interest

The authors declare no competing interests.

Acknowledgments

We wish to thank the participants recruited in the SLSJ asthma cohort for their valuable participation in this study. We also thank Dominique Fournier, scientific language expert, for the revision of this manuscript (<http://www.serviceslinguistiquesdf.com/home>). Anne-Marie Boucher-Lafleur, Andréanne Morin and Jolyane Meloche were supported by a *Fonds de recherche du Québec – Santé* (FRQS) Master, Doctoral or Postdoctoral training award respectively. Lucile Pain and Anne-Marie Boucher-Lafleur were supported by a Master or a Doctoral training award respectively from the Quebec Respiratory Health Network (RHN). Lucile Pain was also supported by the Natural Sciences and Engineering Graduate Scholarship from the Université du Québec à Chicoutimi Foundation (FUQAC). This project was supported by operating grants from the Canadian Institute of Health Research (CIHR; Laprise & Pastinen). Catherine Laprise is part of the Quebec Respiratory Health Network (RHN; <https://rsr-qc.ca/en/>), the investigator of CHILD Study, the director of the *Centre intersectoriel en santé durable de l'UQAC* and the chairholder of the Canada Research Chair in the Environment and Genetics of Respiratory Disorders and Allergies (<http://www.chairs.gc.ca>). The GWAS data were made available by the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1 Post genomic approaches to understand the molecular basis of asthma aiming at a preventive or therapeutic control. This study makes use of data generated by the UK10K Consortium. A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310.

Web resources

OMIM, <http://www.omim.org/>

PLINK, <https://www.cog-genomics.org/plink/1.9/>

WISARD, http://statgen.snu.ac.kr/wisard/?act=aa_fam

References

1. Andiappan, A.K., Sio, Y.Y., Lee, B., Suri, B.K., Matta, S.A., Lum, J., Foo, S., Koh, G., Liu, J., Zolezzi, F., et al. (2016). Functional variants of 17q12-21 are associated with allergic asthma but not allergic rhinitis. *J Allergy Clin Immunol.* 137(3), 758-66 e3.
2. Bonnelykke, K., Sleiman, P., Nielsen, K., Kreiner-Moller, E., Mercader, J.M., Belgrave, D., den Dekker, H.T., Husby, A., Sevelsted, A., Faura-Tellez, G., et al. (2014). A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet.* 46(1), 51-5.
3. Demenais, F., Margaritte-Jeannin, P., Barnes, K.C., Cookson, W.O.C., Altmuller, J., Ang, W., Barr, R.G., Beaty, T.H., Becker, A.B., Beilby, J., et al. (2018). Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet.* 50(1), 42-53.
4. Ferreira, M.A., McRae, A.F., Medland, S.E., Nyholt, D.R., Gordon, S.D., Wright, M.J., Henders, A.K., Madden, P.A., Visscher, P.M., Wray, N.R., et al. (2011). Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. *Eur J Hum Genet.* 19(4), 458-64.
5. Galanter, J.M., Gignoux, C.R., Torgerson, D.G., Roth, L.A., Eng, C., Oh, S.S., Nguyen, E.A., Drake, K.A., Huntsman, S., Hu, D., et al. (2014). 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.* 134(2), 295-305.
6. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., Cookson, W., et al. (2010). A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med.* 363(13), 1211-21.
7. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet.* 94(4), 559-73.
8. Torgerson, D.G., Ampleford, E.J., Chiu, G.Y., Gauderman, W.J., Gignoux, C.R., Graves, P.E., Himes, B.E., Levin, A.M., Mathias, R.A., Hancock, D.B., et al. (2011). Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet.* 43(9), 887-92.
9. Wan, Y.I., Shrine, N.R., Soler Artigas, M., Wain, L.V., Blakey, J.D., Moffatt, M.F., Bush, A., Chung, K.F., Cookson, W.O., Strachan, D.P., et al. (2012). Genome-wide association study to identify genetic determinants of severe asthma. *Thorax.* 67(9), 762-8.
10. Yan, Q., Brehm, J., Pino-Yanes, M., Forno, E., Lin, J., Oh, S.S., Acosta-Perez, E., Laurie, C.C., Cloutier, M.M., Raby, B.A., et al. (2017). A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur Respir J.* 49(5).
11. Madore, A.M., Tremblay, K., Hudson, T.J., Laprise, C. (2008). Replication of an association between 17q21 SNPs and asthma in a French-Canadian familial collection. *Hum Genet.* 123(1), 93-5.
12. Stein, M.M., Thompson, E.E., Schoettler, N., Helling, B.A., Magnaye, K.M., Stanhope, C., Igartua, C., Morin, A., Washington, C., 3rd, Nicolae, D., et al. (2018). A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. *J Allergy Clin Immunol.* 142(3), 749-64 e3.
13. Ferreira, M.A.R., Mathur, R., Vonk, J.M., Szwajda, A., Brumpton, B., Granell, R., Brew, B.K., Ulleamar, V., Lu, Y., Jiang, Y., et al. (2019). Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am J Hum Genet.* 104(4), 665-84.

14. Pividori, M., Schoettler, N., Nicolae, D.L., Ober, C., 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(6), 509-22.
15. Li, X., Hastie, A.T., Hawkins, G.A., Moore, W.C., Ampleford, E.J., Milosevic, J., Li, H., Busse, W.W., Erzurum, S.C., Kaminski, N., et al. (2015). eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy.* 70(10), 1309-18.
16. Nicodemus-Johnson, J., Myers, R.A., Sakabe, N.J., Sobreira, D.R., Hogarth, D.K., Naureckas, E.T., Sperling, A.I., Solway, J., White, S.R., Nobrega, M.A., et al. (2016). DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight.* 1(20), e90151.
17. Liu, Y.P., Rajamanikham, V., Baron, M., Patel, S., Mathur, S.K., Schwantes, E.A., Ober, C., Jackson, D.J., Gern, J.E., Lemanske, R.F., Jr., et al. (2017). Association of ORMDL3 with rhinovirus-induced endoplasmic reticulum stress and type I Interferon responses in human leucocytes. *Clin Exp Allergy.* 47(3), 371-82.
18. Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature.* 448(7152), 470-3.
19. Murphy, A., Chu, J.H., Xu, M., Carey, V.J., Lazarus, R., Liu, A., Szeffler, S.J., Strunk, R., Demuth, K., Castro, M., et al. (2010). Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Hum Mol Genet.* 19(23), 4745-57.
20. Schmiedel, B.J., Seumois, G., Samaniego-Castruita, D., Cayford, J., Schulten, V., Chavez, L., Ay, F., Sette, A., Peters, B., Vijayanand, P. (2016). 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. *Nat Commun.* 7, 13426.
21. Sharma, S., Zhou, X., Thibault, D.M., Himes, B.E., Liu, A., Szeffler, S.J., Strunk, R., Castro, M., Hansel, N.N., Diette, G.B., et al. (2014). A genome-wide survey of CD4(+) lymphocyte regulatory genetic variants identifies novel asthma genes. *J Allergy Clin Immunol.* 134(5), 1153-62.
22. Verlaan, D.J., Berlivet, S., Hunninghake, G.M., Madore, A.M., Lariviere, M., Moussette, S., Grundberg, E., Kwan, T., Ouimet, M., Ge, B., et al. (2009). Allele-specific chromatin remodeling in the ZBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet.* 85(3), 377-93.
23. Al Tuwaijri, A., Gagne-Ouellet, V., Madore, A.M., Laprise, C., Naumova, A.K. (2016). Local genotype influences DNA methylation at two asthma-associated regions, 5q31 and 17q21, in a founder effect population. *J Med Genet.* 53(4), 232-41.
24. Moussette, S., Al Tuwaijri, A., Kohan-Ghadr, H.R., Elzein, S., Farias, R., Berube, J., Ho, B., Laprise, C., Goodyer, C.G., Rousseau, S., et al. (2017). Role of DNA methylation in expression control of the IKZF3-GSDMA region in human epithelial cells. *PLoS One.* 12(2), e0172707.
25. Naumova, A.K., Al Tuwaijri, A., Morin, A., Vaillancourt, V.T., Madore, A.M., Berlivet, S., Kohan-Ghadr, H.R., Moussette, S., Laprise, C. (2013). Sex- and age-dependent DNA methylation at the 17q12-q21 locus associated with childhood asthma. *Hum Genet.* 132(7), 811-22.
26. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 167(5), 1415-29 e19.

27. Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., Greco, D., Soderhall, C., Scheynius, A., Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 7(7), e41361.
28. Sastre, B., Rodrigo-Munoz, J.M., Garcia-Sanchez, D.A., Canas, J.A., Del Pozo, V. (2018). Eosinophils: Old Players in a New Game. *J Investig Allergol Clin Immunol*. 28(5), 289-304.
29. Acevedo, N., Reinius, L.E., Greco, D., Gref, A., Orsmark-Pietras, C., Persson, H., Pershagen, G., Hedlin, G., Melen, E., Scheynius, A., et al. (2015). Risk of childhood asthma is associated with CpG-site polymorphisms, regional DNA methylation and mRNA levels at the GSDMB/ORMDL3 locus. *Hum Mol Genet*. 24(3), 875-90.
30. Berlivet, S., Moussette, S., Ouimet, M., Verlaan, D.J., Koka, V., Al Tuwaijri, A., Kwan, T., Sinnott, D., Pastinen, T., Naumova, A.K. (2012). Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. *Hum Genet*. 131(7), 1161-71.
31. Cheung, W.A., Shao, X., Morin, A., Siroux, V., Kwan, T., Ge, B., Aissi, D., Chen, L., Vasquez, L., Allum, F., et al. (2017). Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol*. 18(1), 50.
32. Barnes, P.J. Cellular and molecular mechanisms of asthma and COPD. England: 2017 The Author(s). published by Portland Press Limited on behalf of the Biochemical Society.; 2017 Jul 1. 1541-58 p.
33. Liang, L., Willis-Owen, S.A.G., Laprise, C., Wong, K.C.C., Davies, G.A., Hudson, T.J., Binia, A., Hopkin, J.M., Yang, I.V., Grundberg, E., et al. (2015). An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature*. 520(7549), 670-4.
34. Laprise, C. (2014). The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population. *Genes Immun*. 15(4), 247-55.
35. (1979). ATS statement--Snowbird workshop on standardization of spirometry. *Am Rev Respir Dis*. 119(5), 831-8.
36. O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., et al. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet*. 10(4), e1004234.
37. Howie, B.N., Donnelly, P., Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 5(6), e1000529.
38. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature*. 526(7571), 68-74.
39. Ferland, C., Guilbert, M., Davoine, F., Flamand, N., Chakir, J., Laviolette, M. (2001). Eotaxin promotes eosinophil transmigration via the activation of the plasminogen-plasmin system. *J Leukoc Biol*. 69(5), 772-8.
40. Guilbert, M., Ferland, C., Bosse, M., Flamand, N., Lavigne, S., Laviolette, M. (1999). 5-Oxo-6,8,11,14-eicosatetraenoic acid induces important eosinophil transmigration through basement

membrane components: comparison of normal and asthmatic eosinophils. *Am J Respir Cell Mol Biol.* 21(1), 97-104.

41. Allum, F., Shao, X., Guenard, F., Simon, M.M., Busche, S., Caron, M., Lambourne, J., Lessard, J., Tandre, K., Hedman, A.K., et al. (2015). Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat Commun.* 6, 7211.
42. Nyholt, D.R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet.* 74(4), 765-9.
43. Thornton, T., McPeck, M.S. (2007). Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet.* 81(2), 321-37.
44. Lee, S., Choi, S., Qiao, D., Cho, M., Silverman, E.K., Park, T., Won, S. (2018). WISARD: workbench for integrated superfast association studies for related datasets. *BMC Med Genomics.* 11(Suppl 2), 39.
45. Barrett, J.C., Fry, B., Maller, J., Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21(2), 263-5.
46. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. The structure of haplotype blocks in the human genome. *United States2002 Jun 21.* 2225-9 p.
47. Guo, S., Diep, D., Plongthongkum, N., Fung, H.L., Zhang, K., Zhang, K. (2017). Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet.* 49(4), 635-42.
48. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 13, 86.
49. Leek, J.T., Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *United States2007 Sep.* 1724-35 p.
50. Love, M.I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12), 550.
51. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 28(10), 1353-8.
52. Heyn, H. (2014). A symbiotic liaison between the genetic and epigenetic code. *Front Genet.* 5, 113.
53. Galanter, J., Choudhry, S., Eng, C., Nazario, S., Rodriguez-Santana, J.R., Casal, J., Torres-Palacios, A., Salas, J., Chapela, R., Watson, H.G., et al. ORMDL3 gene is associated with asthma in three ethnically diverse populations. *United States2008 Jun 1.* 1194-200 p.
54. Shahid, M., Sabar, M.F., Bano, I., Rahman, Z., Iqbal, Z., Fatim Ali, S.S., Ghani, M.U., Iqbal, M., Husnain, T. (2015). Sequence variants on 17q21 are associated with the susceptibility of asthma in the population of Lahore, Pakistan. *J Asthma.* 52(8), 777-84.

55. Tomita, K., Sakashita, M., Hirota, T., Tanaka, S., Masuyama, K., Yamada, T., Fujieda, S., Miyatake, A., Hizawa, N., Kubo, M., et al. (2013). Variants in the 17q21 asthma susceptibility locus are associated with allergic rhinitis in the Japanese population. *Allergy*. 68(1), 92-100.
56. Li, L., Ding, X., Wang, X., Yao, Q., Shao, X., An, X., Yan, N., Jiang, Y., Wang, W., Shi, L., et al. Polymorphisms of IKZF3 Gene and Autoimmune Thyroid Diseases: Associated with Graves' Disease but Not with Hashimoto's Thyroiditis. Switzerland: 2018 The Author(s). Published by S. Karger AG, Basel.; 2018. 1787-96 p.
57. Jorgensen, I., Miao, E.A. (2015). Pyroptotic cell death defends against intracellular pathogens. *Immunol Rev*. 265(1), 130-42.
58. Ding, J., Wang, K., Liu, W., She, Y., Sun, Q., Shi, J., Sun, H., Wang, D.C., Shao, F. (2016). Pore-forming activity and structural autoinhibition of the gasdermin family. *Nature*. 535(7610), 111-6.
59. Xavier, M.J., Roman, S.D., Aitken, R.J., Nixon, B. (2019). Transgenerational inheritance: how impacts to the epigenetic and genetic information of parents affect offspring health. *Hum Reprod Update*.
60. Caliskan, M., Bochkov, Y.A., Kreiner-Moller, E., Bonnelykke, K., Stein, M.M., Du, G., Bisgaard, H., Jackson, D.J., Gern, J.E., Lemanske, R.F., Jr., et al. (2013). Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med*. 368(15), 1398-407.
61. Lluís, A., Schedel, M., Liu, J., Illi, S., Depner, M., von Mutius, E., Kabesch, M., Schaub, B. (2011). Asthma-associated polymorphisms in 17q21 influence cord blood ORMDL3 and GSDMA gene expression and IL-17 secretion. *J Allergy Clin Immunol*. 127(6), 1587-94 e6.

Figures

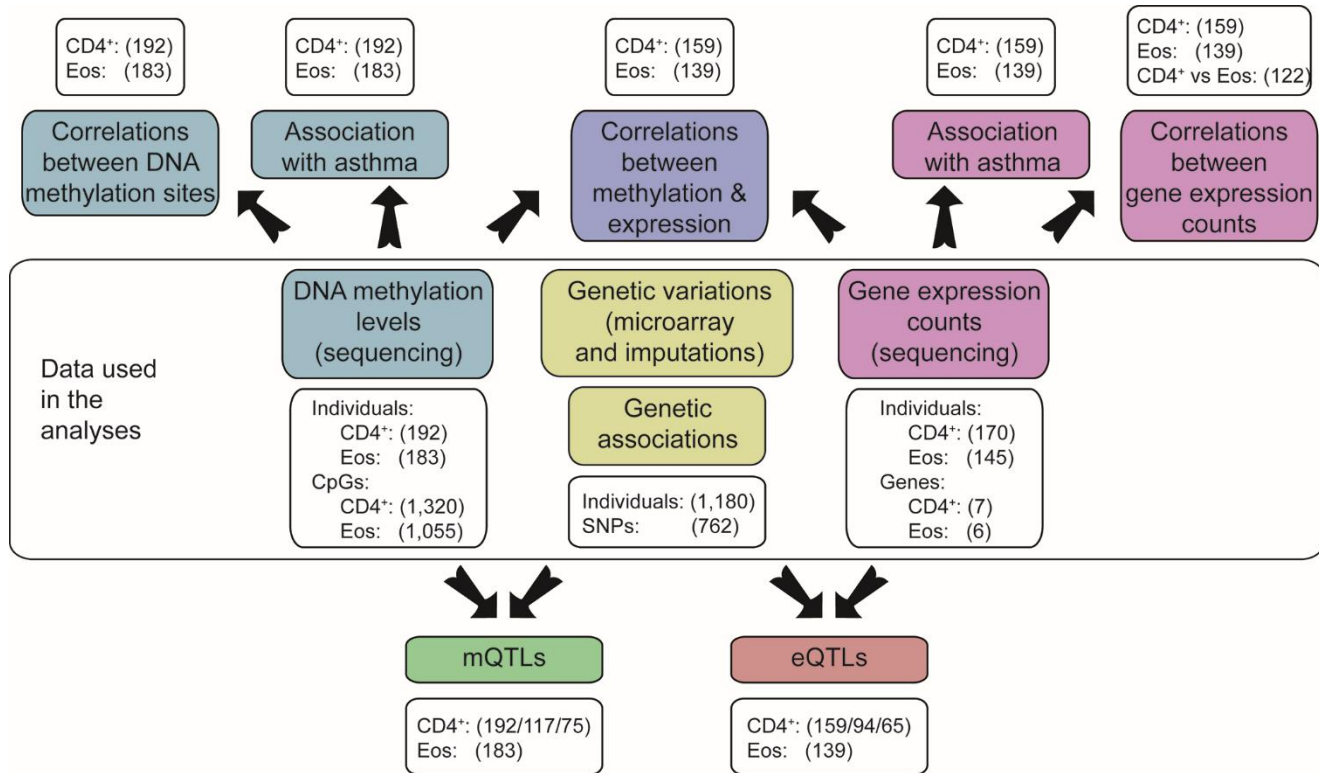


Figure 1. Schematic view of the data used and analyses performed in this study

Data used for the analyses included genetic variations from microarrays with imputations, as well as methylation levels and expression counts from next generation sequencing. The top part of the figure shows the analyses performed with methylation and expression data. The bottom part presents the integration of these data with genetic variations. Each white box indicates the number of samples and the one of genetic data (SNPs, CpGs or genes) included in the analyses, considering quality control filtering and the presence of all covariates in the models used. When three numbers are listed within parentheses, the first corresponds to the number of samples for the principal analysis (with all considered), the second where only individuals with asthma were analyzed and the third where controls were only. Abbreviations: CD4⁺ = naïve CD4⁺ T cells, Eos = eosinophils.

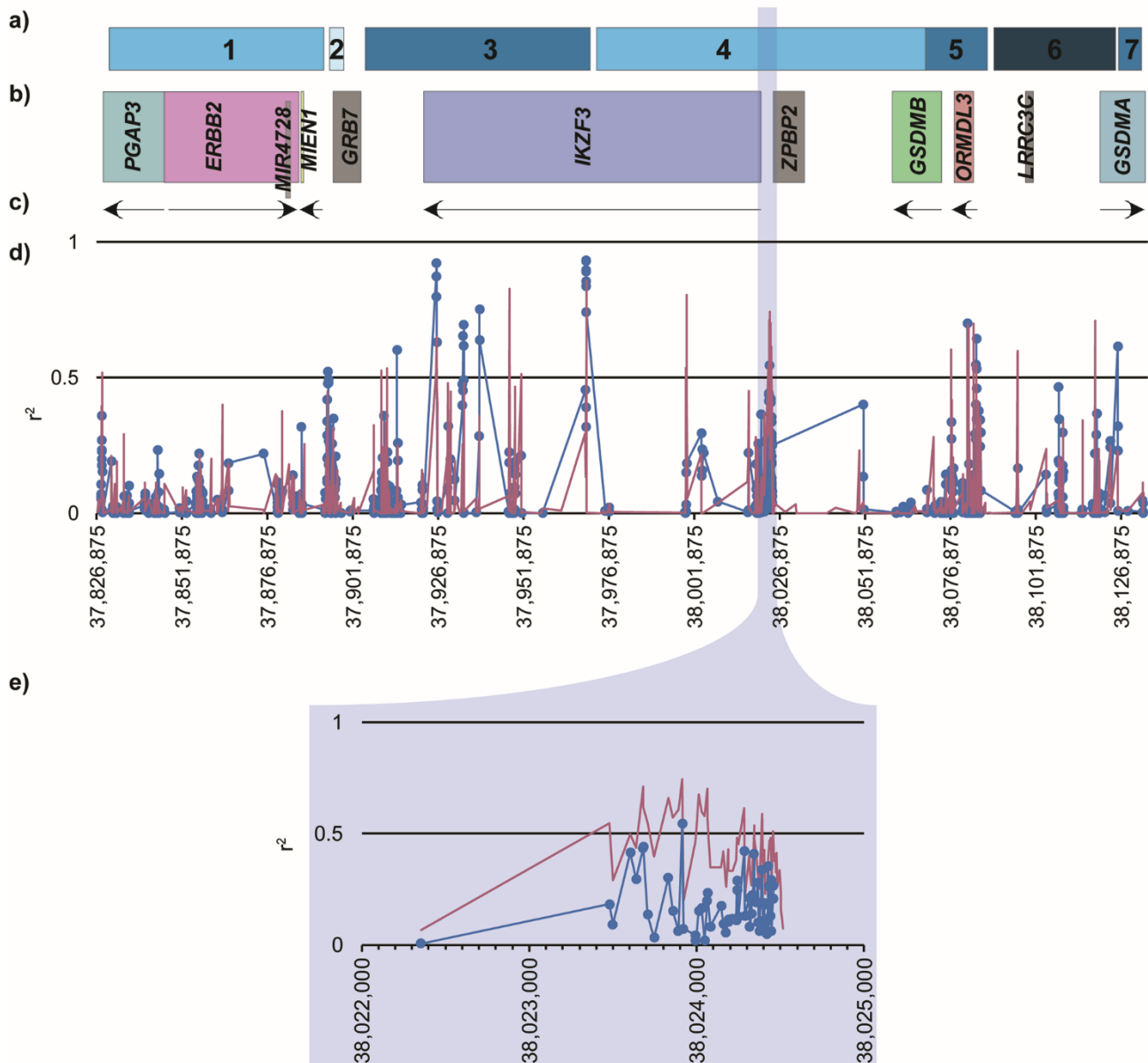


Figure 2. Correlation between methylation sites in naïve CD4⁺ T cells and eosinophils

Sections (a) and (b) depict the seven haplotype blocks identified in the Saguenay–Lac-Saint-Jean

asthma cohort and the eleven genes located in this locus. Genes colored in grey are those not

expressed in this study by naïve CD4⁺ T cells. The section (c) indicates the strand of each gene

expressed by one of the two cell types studied, and the section (d) depicts r^2 that correlates between

methylation sites in samples recovered from naïve CD4⁺ T cells (pink line) and eosinophils (blue line

with round markers) across the region of interest. Section (e) zooms in the promoter region of *IKZF3*

that includes several methylation blocks in naïve CD4⁺ T cells but not eosinophils (they are defined by two or more CpGs with r^2 values >0.5).

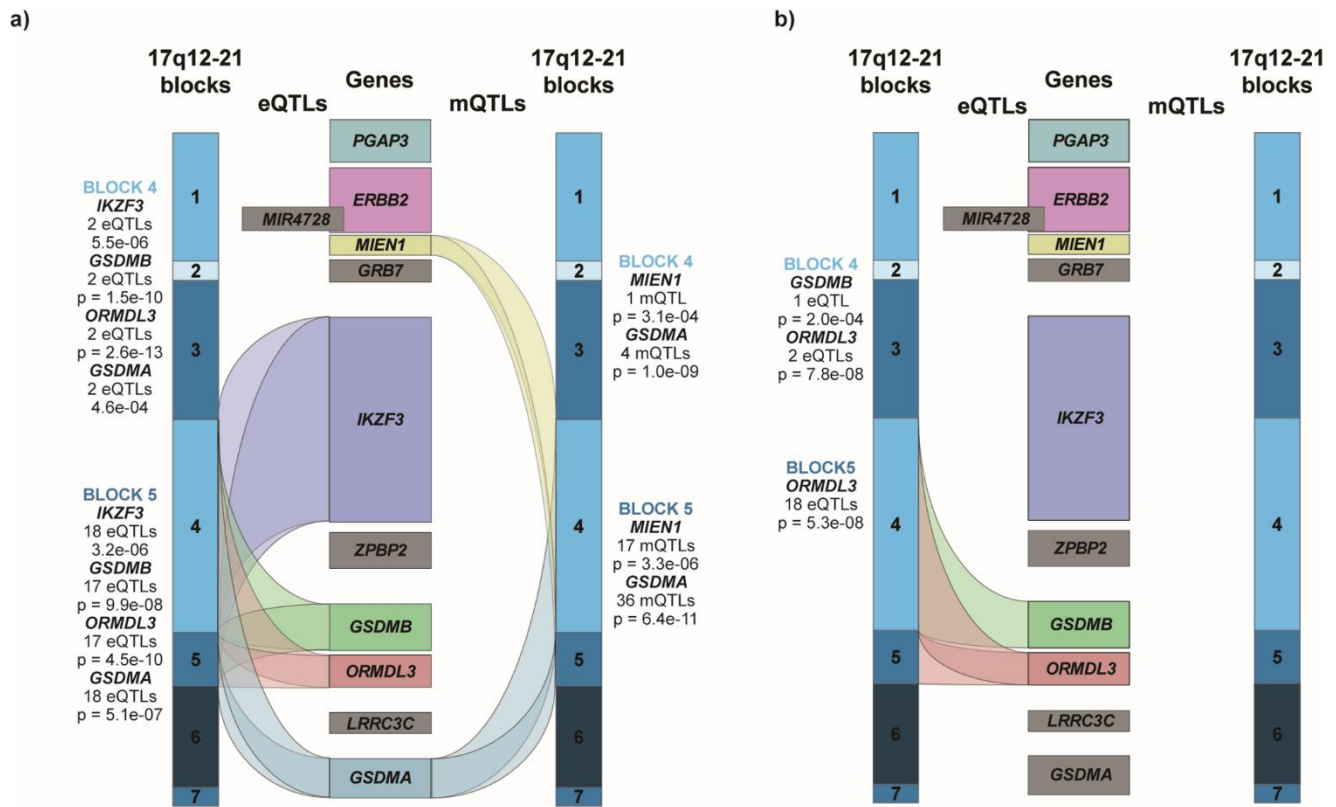


Figure 3. eQTLs and mQTLs of the 17q12-21 locus by cell types

This figure illustrates the eQTLs and mQTLs considering SNPs associated with asthma identified in naïve CD4⁺ T cells (a) and eosinophils (b). The seven haplotype blocks detected in the Saguenay–Lac-Saint-Jean asthma cohort for the chr17:37,826,875-38,134,519 region are at the outer edges in different shades of blue and the middle are genes located in the same area. Genes colored in grey are those not expressed by each cell type in this study. The number of eQTLs (left part of any figure) and mQTLs (right part) linked to each gene for each haplotype block are shown as indicated. P values for the most statistically significant QTLs are listed for each gene in each block.

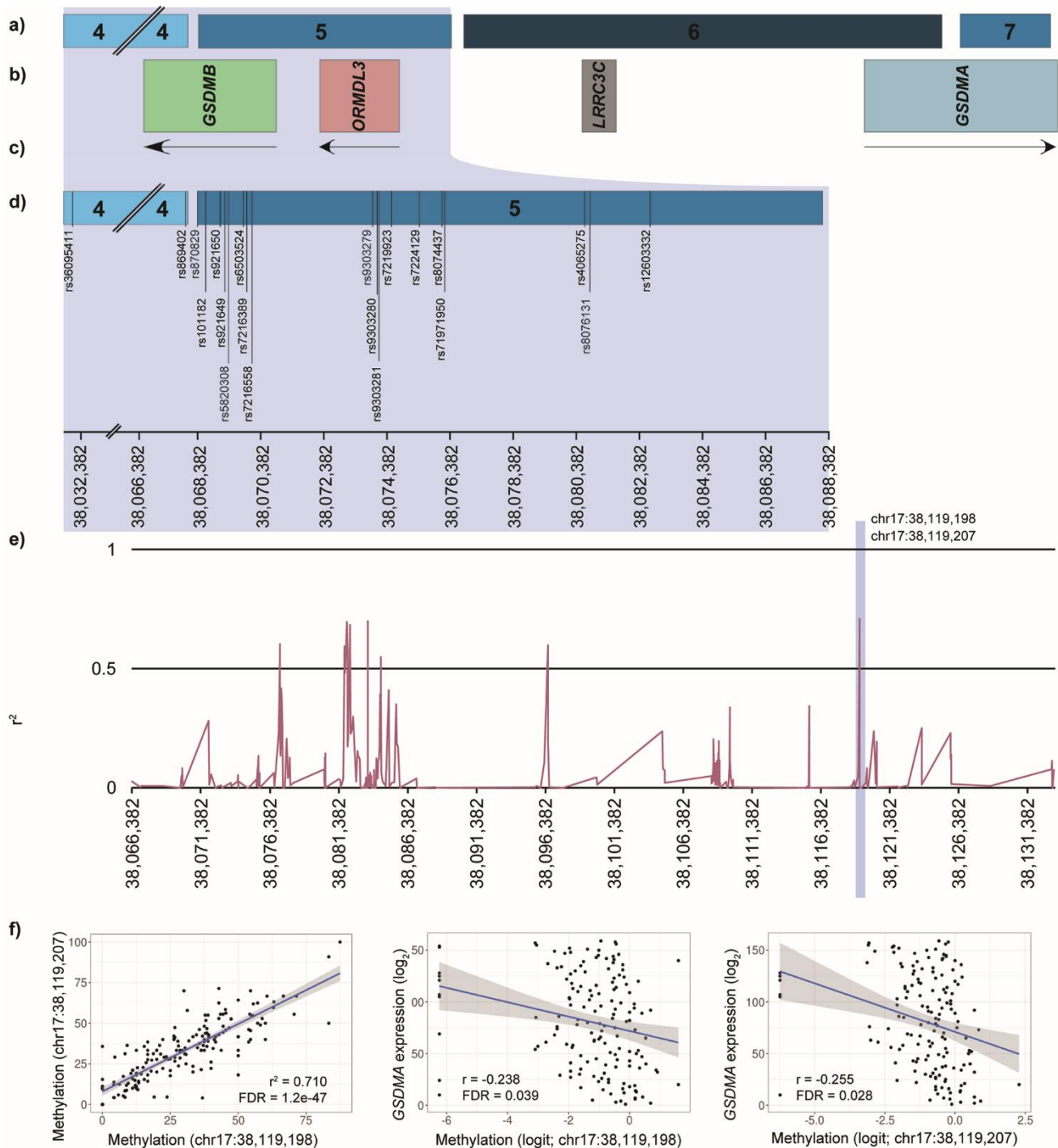


Figure 4. Methylation and expression profiles of *GSDMA* in samples from naïve CD4⁺ T cells

This figure shows: (a) the 17q12-21 region from the fourth to the seventh haplotype blocks; (b) the genes located in this region, and (c) the strand of these. Genes in grey are those not expressed by naïve CD4⁺ T cells. The section (d) zooms in the fourth and fifth haplotype blocks to show all SNPs in eQTLs and mQTLs that were identified in naïve CD4⁺ T cells for *GSDMA*. All SNPs included in eQTLs

and are in mQTLs with the two correlated CpGs indicated in (e), where r^2 is depicted for correlations between adjacent methylation sites located in the region including the end of the fourth block to the seventh one. The section (f) presents correlation between the two methylation sites included in mQTLs and the one between each of them and *GSDMA* expression.

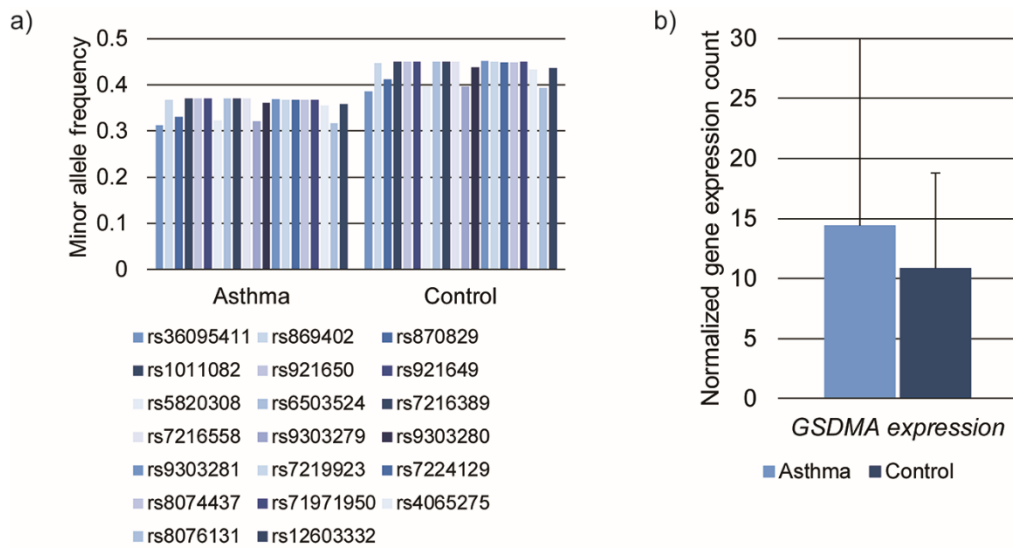


Figure 5. Minor allele frequencies of asthma-associated SNPs and *GSDMA* expression counts by the disease status

This figure shows: (a) the minor allele frequency of all 20 asthma-associated SNPs in the fourth and fifth haplotype blocks of individuals living with asthma and controls from the Saguenay–Lac-Saint-Jean cohort and (b) the *GSDMA* expression counts in the same groups. Standard deviations are presented for mean values.

Tables

Table 1: Phenotypic characteristics of the individuals genotyped and those included in the analyses performed on naïve CD4⁺ T cells and eosinophils

	GWAS		Naïve CD4 ⁺ T cells ^b (n=203)	Eosinophils ^b (n=183)
	Proband ^a (n=242)	Family members ^a (n=958)		
M:F ratio	01:01.1	01:01.2	1:1.07	1:1.01
Age, mean (range)	17 (3–62)	43 (0–93)	48 (18–81)	47 (18–81)
Age, median	15	43	52	49
Smoking status ^c				
Non-smokers, n (%)	203 (85)	435 (46)	118 (59)	111 (61)
Ex-smokers, n (%)	10 (4)	301 (32)	47 (23)	42 (23)
Smokers, n (%)	25 (11)	206 (22)	36 (18)	28 (15)
PC ₂₀ , mean (SD) ^d	2.3 (3.93)	9.8 (5.81)	6.69 (5.59)	5.87 (5.43)
FEV ₁ , mean % pred, (SD) ^e	91.78 (33.29)	95.03 (43.70)	95.47 (24.08)	94.73 (24.56)
FEV ₁ /FVC ratio, mean (SD) ^f	79.65 (30.93)	78.85 (35.71)	78.42 (24.42)	78.36 (24.93)
Asthma, n (%) ^g	242 (100)	344 (36)	122 (60)	117 (64)
Age of onset, mean (range)	7 (0–46)	22 (0–75)	12.53 (11.84)	12.09 (11.66)

^aProband^as are the first affected family members recruited and the ones referred to others who joined the study. ^bNumber of samples from naïve CD4⁺ T cells and eosinophils for which methylation and/or expression data were available. ^cEx-smokers are defined as individuals who stopped smoking for one year or more. Data are available for 242 probands, 957 family members, 203 and 183 individuals with samples from naïve CD4⁺ T cells and eosinophils respectively. ^dThe geometric mean of the provocative methacholine concentration inducing a 20% fall in FEV₁ (PC₂₀) calculated for 184 probands, 608 family members, 190 and 163 individuals with samples from naïve CD4⁺ T cells and eosinophils respectively. ^eThe mean forced expiratory volume in 1 s (FEV₁) as % of predicted value calculated for 213 probands, 737 family members, 196 and 176 individuals with samples from naïve CD4⁺ T cells and eosinophils respectively. ^fThe mean FEV₁ (L)/FVC (forced vital capacity) (L) ratio calculated as %. Data are available for 201 probands, 701 family members, 184 and 165 individuals with samples from naïve CD4⁺ T cells and eosinophils respectively. ^gPresent or past documented clinical history of asthma. Data are available for 242 probands, 951 family members, 203 and 183 individuals with samples from naïve CD4⁺ T cells and eosinophils respectively.

1 **Table 2.** Significant associations between gene expression counts and DNA methylation levels of CpGs located within 1.5 kb from a TSS or
 2 a gene body in naïve CD4⁺ T cells or eosinophils

Gene	Position TSS (hg19)	Position CpG (hg19)	Gene is methylated (1.5 kb from TSS) ^a	Gene is methylated (gene body) ^a	Pearson's r	Coeff	SE	P value	FDR
Naïve CD4 ⁺ T cells									
<i>ERBB2</i> / +	37,844,336	37,843,621	<i>ERBB2</i>	<i>PGAP3</i>	-0.235	-0.078	0.019	5.26e-05	0.012
		37,856,890		<i>ERBB2</i>	0.283	0.048	0.012	6.88e-05	0.015
		37,947,691		<i>IKZF3</i>	0.414	0.049	0.012	5.85e-05	0.013
		37,947,697		<i>IKZF3</i>	0.380	0.055	0.013	3.40e-05	0.009
		37,947,773		<i>IKZF3</i>	0.398	0.057	0.013	1.38e-05	0.004
		37,951,189		<i>IKZF3</i>	0.359	0.041	0.011	1.02e-04	0.020
		37,999,368		<i>IKZF3</i>	0.319	0.041	0.011	3.07e-04	0.045
		37,999,641		<i>IKZF3</i>	0.426	0.044	0.009	2.93e-06	0.001
		38,020,606	<i>IKZF3</i>		0.288	0.078	0.021	2.56e-04	0.040
		38,083,300		<i>ORMDL3</i>	0.262	0.079	0.022	3.64e-04	0.050
		38,120,370		<i>GSDMA</i>	0.409	0.136	0.021	7.48e-11	9.06e-08
<i>MIEN1</i> / -	37,886,816	37,937,926		<i>IKZF3</i>	-0.235	-0.041	0.011	1.57e-04	0.028
<i>IKZF3</i> / -	38,020,441	37,927,002		<i>IKZF3</i>	-0.145	-0.047	0.012	5.45e-05	0.012
		37,929,495		<i>IKZF3</i>	-0.269	-0.072	0.017	3.23e-05	0.009
		37,937,926		<i>IKZF3</i>	-0.233	-0.053	0.012	5.14e-06	0.002
		37,938,973		<i>IKZF3</i>	-0.096	-0.043	0.011	6.10e-05	0.013
		37,949,341		<i>IKZF3</i>	-0.380	-0.087	0.019	4.30e-06	0.002
		38,072,832		<i>GSDMB</i>	-0.185	-0.021	0.006	3.16e-04	0.046
<i>GSDMB</i> / -	38,074,903	37,926,502		<i>IKZF3</i>	-0.430	-0.045	0.013	3.48e-04	0.050
		37,947,691		<i>IKZF3</i>	-0.380	-0.034	0.009	2.69e-04	0.041
		38,077,703		<i>ORMDL3</i>	-0.167	-0.109	0.026	2.22e-05	0.006
		38,080,398		<i>ORMDL3</i>	-0.361	-0.091	0.019	1.54e-06	6.99e-04
		38,081,492		<i>ORMDL3</i>	-0.346	-0.028	0.007	1.57e-04	0.028
		38,081,812		<i>ORMDL3</i>	-0.470	-0.056	0.010	3.19e-08	2.32e-05
		38,081,829		<i>ORMDL3</i>	-0.510	-0.051	0.009	1.32e-08	1.13e-05

		38,081,995	ORMDL3	-0.548	-0.105	0.013	9.14e-15	2.21e-11
		38,082,018	ORMDL3	-0.478	-0.057	0.011	2.89e-07	1.56e-04
		38,082,022	ORMDL3	-0.549	-0.109	0.014	1.63e-15	5.92e-12
		38,082,035	ORMDL3	-0.442	-0.096	0.016	7.54e-10	8.42e-07
		38,082,073	ORMDL3	-0.413	-0.141	0.024	1.94e-09	2.01e-06
		38,082,206	ORMDL3	-0.432	-0.051	0.010	2.03e-07	1.19e-04
		38,082,235	ORMDL3	-0.365	-0.049	0.012	3.21e-05	0.009
		38,082,346	ORMDL3	-0.494	-0.106	0.014	3.35e-14	6.08e-11
		38,082,448	ORMDL3	-0.494	-0.084	0.012	1.08e-11	1.58e-08
		38,082,658	ORMDL3	-0.395	-0.040	0.010	1.12e-04	0.022
<i>ORMDL3 / -</i>	38,084,057	38,077,051		-0.367	-0.095	0.021	6.66e-06	0.002
		38,077,323	ORMDL3	-0.315	-0.024	0.007	2.58e-04	0.040
		38,077,621	ORMDL3	-0.338	-0.092	0.021	1.01e-05	0.003
		38,077,703	ORMDL3	-0.349	-0.113	0.022	2.06e-07	1.19e-04
		38,080,398	ORMDL3	-0.328	-0.063	0.017	1.71e-04	0.029
		38,081,492	ORMDL3	-0.340	-0.024	0.006	1.77e-04	0.030
		38,081,812	ORMDL3	-0.433	-0.052	0.009	2.28e-09	2.20e-06
		38,081,829	ORMDL3	-0.479	-0.050	0.007	2.09e-11	2.76e-08
		38,081,995	ORMDL3	-0.560	-0.094	0.012	3.17e-16	1.53e-12
		38,082,018	ORMDL3	-0.436	-0.053	0.009	2.30e-08	1.86e-05
		38,082,022	ORMDL3	-0.625	-0.113	0.011	7.42e-25	9.31e-21
		38,082,035	ORMDL3	-0.478	-0.099	0.013	4.69e-14	7.57e-11
		38,082,073	ORMDL3	-0.541	-0.151	0.019	5.00e-15	1.45e-11
		38,082,206	ORMDL3	-0.382	-0.042	0.009	9.95e-07	4.66e-04
		38,082,235	ORMDL3	-0.372	-0.046	0.010	3.57e-06	0.001
		38,082,346	ORMDL3	-0.551	-0.114	0.011	1.28e-24	9.31e-21
		38,082,448	ORMDL3	-0.516	-0.080	0.010	1.14e-14	2.37e-11
		38,082,658	ORMDL3	-0.387	-0.040	0.009	6.40e-06	0.002
		38,082,778	ORMDL3	-0.376	-0.053	0.012	1.19e-05	0.004
		38,082,921	ORMDL3	-0.403	-0.036	0.008	1.25e-05	0.004
		38,084,297	ORMDL3	0.308	0.085	0.022	7.55e-05	0.016
<i>GSDMA / +</i>	38,119,225	37,834,869	PGAP3	-0.304	-0.181	0.033	2.74e-08	2.10e-05
		38,119,198	GSDMA	-0.238	-0.129	0.035	2.50e-04	0.039
		38,119,207	GSDMA	-0.255	-0.154	0.041	1.61e-04	0.028

		38,119,649	<i>GSDMA</i>	-0.192	-0.152	0.041	1.84e-04	0.031
Eosinophils								
<i>PGAP3</i> / -	37,844,323	37,922,462	<i>IKZF3</i>	-0.183	0.036	0.009	1.00e-04	0.039
<i>ERBB2</i> / +	37,844,336	37,881,117	<i>ERBB2</i>	-0.030	0.093	0.024	1.25e-04	0.044
		38,004,064	<i>IKZF3</i>	-0.084	0.187	0.045	3.79e-05	0.020
		38,083,354	<i>ORMDL3</i>	-0.014	0.251	0.056	7.09e-06	0.005
<i>IKZF3</i> / -	38,020,441	37,828,163	<i>PGAP3</i>	0.005	0.107	0.027	8.80e-05	0.038
		37,938,833	<i>IKZF3</i>	0.106	0.058	0.015	6.91e-05	0.033
		37,938,939	<i>IKZF3</i>	0.034	0.070	0.015	2.93e-06	0.003
		38,020,407	<i>IKZF3</i>	0.069	-0.133	0.028	2.70e-06	0.003

- 3 ^aGenes are indicated in the fourth and/or fifth columns if the CpG position is within 1,500 bp from the transcription start site (Gene is
- 4 methylated [1.5 kb from TSS]) or in the gene body (Gene is methylated [gene body]). Abbreviations: TSS = transcription start site.