

ORIGINAL ARTICLE

Epidemiology of Allergic Disease

The *COL5A3* and *MMP9* genes interact in eczema susceptibility

P. Margaritte-Jeannin^{1,2*} | M.-C. Babron^{1,2*} | C. Laprise³ | N. Lavielle^{1,2} |
 C. Sarnowski^{1,2} | M. Brossard^{1,2} | M. Moffatt⁴ | V. Gagné-Ouellet³ | A. Etcheto⁵ |
 M. Lathrop⁶ | J. Just⁷  | W. O. Cookson⁴ | E. Bouzigon^{1,2} | F. Demenais² |
 M.-H. Dizier^{1,2} 

¹Inserm, UMR-946, Genetic Variation and Human Diseases unit, Université Paris-Diderot, Paris, France

²Sorbonne Paris Cité, Institut Universitaire d'Hématologie, Univ Paris Diderot, Paris, France

³Université du Québec, Chicoutimi, Canada

⁴National Heart Lung Institute, Imperial College, London, UK

⁵Rheumatology Department, Cochin Hospital, AP-HP, INSERM U1153, Sorbonne Paris-Cité, Paris Descartes University, Paris, France

⁶Mc Gill University and Genome Quebec's Innovation Centre, Montréal, Canada

⁷Service d'Allergologie Pédiatrique, Centre de l'Asthme et des Allergies, Hôpital d'Enfants Armand-Trousseau-UPMC Paris 06, Paris, France

Correspondence

M.-H. Dizier, INSERM, UMR946, Université Paris-Diderot, Paris, France.
 Email: Marie-Helene.Dizier@inserm.fr

Funding information

This work was supported by the GWIS-AM grant ANR-11-BSV1-027ANR-USPC-2013-EDAGWAS The Canada Research Chair in Environment and Genetics of Respiratory Diseases and Allergy held by C. Laprise since 2005 allows the maintenance of the French Canadian study.

Summary

Background: Genetic studies of eczema have identified many genes, which explain only 14% of the heritability. Missing heritability may be partly due to ignored gene–gene (G-G) interactions.

Objective: Our aim was to detect new interacting genes involved in eczema.

Methods: The search for G-G interaction in eczema was conducted using a two-step approach, which included as a first step, a biological selection of genes, which are involved either in the skin or epidermis development or in the collagen metabolism, and as a second step, an interaction analysis of the selected genes. Analyses were carried out at both SNP and gene levels in three asthma-ascertained family samples: the discovery dataset of 388 EGEA (Epidemiological study on the Genetics and Environment of Asthma) families and the two replication datasets of 253 SLSJ (Sageunay–Lac-Saint-Jean) families and 207 MRCA (Medical Research Council) families.

Results: One pair of SNPs, rs2287807 in *COL5A3* and rs17576 in *MMP9*, that were detected in EGEA at $P \leq 10^{-5}$ showed significant interaction by meta-analysis of EGEA, SLSJ and MRCA samples ($P = 1.1 \times 10^{-8}$ under the significant threshold of 10^{-7}). Gene-based analysis confirmed strong interaction between *COL5A3* and *MMP9* ($P = 4 \times 10^{-8}$ under the significant threshold of 4×10^{-6}) by meta-analysis of the three datasets. When stratifying the data on asthma, this interaction remained in both groups of asthmatic and non-asthmatic subjects.

Conclusion: This study identified significant interaction between two new genes, *COL5A3* and *MMP9*, which may be accounted for by a degradation of *COL5A3* by *MMP9* influencing eczema susceptibility. Further confirmation of this interaction as well as functional studies is needed to better understand the role of these genes in eczema.

KEYWORDS

asthma, eczema, gene–gene interaction, logistic regression, meta-analysis

*These authors contributed equally to this study.

1 | INTRODUCTION

Atopic dermatitis (AD), or eczema, is the most common chronic inflammatory skin disorder. Eczema is also an allergic disease associated with asthma and allergic rhinitis. These co-morbidities may share genetic determinants as suggested by their strong associations at both the individual and family levels.¹ However, genes specific to eczema also exist, as first suggested by the observation that parental history of eczema represents a more potent risk factor for eczema than parental history of asthma or allergic rhinitis.^{1,2}

A number of susceptibility genes for eczema have been found through genetic studies³ based on candidate genes, positional cloning approaches and Genome-Wide Association Studies (GWAS). Many of these genes are involved in the immune response and are not specifically associated with eczema, but more generally to allergic diseases including asthma and allergic rhinitis. Among the genes more specifically associated with eczema are genes involved in the skin or epidermis development including *SPRR3*,⁴ *ACTL9*,⁵ *OVOL1*⁵ and also *FLG*,⁶ *SPINK5*,⁷ *C11orf30*⁸ which are related to epidermal barrier dysfunction and genes involved in collagen metabolism like *COL29A1*,⁹ *IL6*¹⁰ and *IL6R*.¹¹

However, all the genes associated with eczema explain only a small part of the heritability (14%-15%).^{12,13} Mechanisms such as gene-gene (G-G) or gene-environment interactions may account for this missing heritability. Indeed, most previous genetic studies on eczema focused on single-marker analysis. Five small-scale G-G interaction studies were conducted for eczema, which involved only on a very small number of genes.¹⁴⁻¹⁸

As it has been shown that ignoring interaction may hinder the detection of genetic factors,¹⁹ several methods have been developed to identify G-G interactions in association studies. They can be applied to the whole genome or to part of the genome, such as a linkage region. To alleviate the burden of multiple testing, which already exists in single-marker analysis but is even more crucial in two-marker analysis, two-step strategies have been proposed. They typically include a first step of marker selection, followed by a second step of two-marker tests applied to the selected markers.²⁰ The selection of SNPs at the first step can be based on single-marker association results or on biological information.²¹ We showed that, in case of strong interaction with a small marginal effect of each marker, a first-step selection based on association tests was not powerful, as the markers would not be selected at the first step.²² Moreover, we can assume that genetic variants that do not have a too small marginal effect would have already been detected by single-marker association tests, through the numerous GWAS conducted for eczema using large sample sizes^{5,8,11,23-26} which recently reached more than 115,000 subjects.¹³

The aim of the present study was to detect new interacting genes in eczema. To search for G-G interaction, we applied a two-step approach, with a first step of gene selection based on biological knowledge related to eczema. We selected genes which were, similarly to the genes previously found specifically associated with

eczema, involved either in the skin or epidermis development or in the collagen metabolism. In a second step, we performed cross-gene SNP-SNP interaction analyses followed by gene-based interaction analyses. These analyses were conducted in asthma-ascertained family samples from three studies: a discovery dataset, the Epidemiological study on the Genetics and Environment of Asthma (EGEA) and two independent replication datasets, the Saguenay-Lac-Saint-Jean (SLSJ) and the Medical Research Council (MRCA) studies.

2 | MATERIAL AND METHOD

2.1 | Discovery sample

The EGEA study has been described in detail previously.²⁷ The EGEA family sample consisted of 388 French nuclear families, 253 families ascertained through offspring with asthma (one proband in 90% of families and two probands in the remainder) and 135 families ascertained through one parent with asthma. Eczema status was based on the answer to the question: "Have you had eczema during childhood?" (either at the time of data collection and/or at the 10-year follow-up of the EGEA study). The same definition has already been used in previous linkage and association analyses of eczema conducted in EGEA.^{28,29} Ethical approval was obtained from the relevant institutional review board committees (Cochin Port-Royal Hospital and Necker-Enfants Malades Hospital, Paris). Written informed consent was signed by all participants or by kin or guardians of the minors/children.

2.2 | Replication samples

The SLSJ Familial Asthma Collection comprised 253 French Canadian multigenerational families ascertained through two probands with asthma.³⁰ Eczema was defined by a self-reported history of eczema on a questionnaire. Positive answers were validated by physicians or through clinical records. The SLSJ local ethics committee approved the study, and all subjects gave written informed consent.

The Medical Research Council (MRC) UK National family collection included 207 nuclear families, recruited through at least one proband with asthma (MRCA sample).³¹ Eczema definition was based on answers to questions on the presence of symptoms. The MRCA local ethics committee approved the study, and all subjects gave written informed consent.

2.3 | Genetic data

The EGEA and the SLSJ samples were genotyped using the Illumina 610-Quad, as part of the European Gabriel consortium asthma GWAS,³² while the MRCA sample was genotyped using the smaller Illumina Sentrix HumanHap300 BeadChip.³¹ To get a number of SNPs in the MRCA dataset as large as in the EGEA and SLSJ datasets, SNP imputation was performed using MACH v1.00 software (<http://www.sph.umich.edu/csg/abecasis/MACH/>) and HapMap2

CEU haplotypes as reference panel. Imputed SNPs with imputation quality score greater than or equal to 0.9 were retained. Stringent quality criteria (QC)³² were used to select both individuals and SNPs.

After genotyping, an ancestry analysis was carried out in the three European descent samples (EGEA, SLSJ and MRCA)³² and putative non-European samples were eliminated from subsequent analyses. Informative principal components for within-Europe diversity were included as covariates in analyses.

2.4 | Statistical analysis

2.4.1 | Two-step strategy in the Epidemiological study on the Genetics and Environment of Asthma discovery sample

The first step consisted in the selection of genes based on biological knowledge contained in the Gene Ontology (GO) database (<http://amigo.geneontology.org/>). From the GO biological process categories, we defined two groups. The first group included GO categories containing the term “collagen,” ie “collagen metabolic process” (GO:0032963), “collagen-activated signalling pathway” (GO:0038065), “regulation of collagen binding” (GO:0033341). The second group included GO categories containing the terms “epidermis” or “skin,” ie “skin development” (GO:0043588) and “epidermis development” (GO:0008544) after exclusion of sub-categories related to hair follicle, limb and nipple. After selection of the genes belonging to each group, all SNPs lying within the gene boundaries (Build 37.3) and present in our three datasets were retained for analysis. The first group, called the “collagen” group, comprised 121 genes (97 genes after excluding those having no SNP genotyped and 1942 SNPs after QC checking), while the second group, the “epidermis and skin” group, totalled 251 genes (176 genes after excluding those having no SNP genotyped and 2189 SNPs after QC checking) (Table S1). Only SNPs located within the genes were considered to limit the number of tests.

The second step consisted in conducting cross-gene SNP-SNP interaction analyses within each gene group, ie between SNPs located in two different genes belonging to the same group. We performed logistic regression using Stata® v14.1 (Stata Corporation, College Station, Texas) assuming an additive model for SNP main effect and interaction and taking familial dependencies into account through the cluster and robust options of the logit function. We used the same coding scheme as usually proposed for SNP-SNP interaction modelling.²¹ We modelled the additive effect of a SNP by coding the genotypes of homozygotes for the minor allele, heterozygotes and homozygotes for the major allele as 1, 0 and -1; the interaction term between two SNPs was obtained by multiplication of these genotypic values for the two SNPs. Test of interaction was performed by comparing the full model with SNP main effect and interaction to the restricted model without interaction, using a likelihood-ratio test which follows a Chi-square distribution with one degree of freedom. SNP pairs in which the expected number of joint genotypes in cases or in controls was

less than 5 were excluded. The total number of tested SNP pairs was 1.2×10^6 .

To take into account the dependence between tests due to the linkage disequilibrium (LD) between SNPs within the same gene, we calculated the effective number of independent SNP pairs. We first calculated the LD matrix of each gene, ie the correlation between the SNPs of that gene. Then, as proposed by Emily,³³ the correlation matrix of a gene pair is set to the Kronecker product of the LD matrices of the two genes. It is then possible to estimate the number of independent tests, K_{eff} , for a given gene pair, using the procedure of Moskvina and Schmidt.³⁴ Summing over all pairs of genes provides a total K_{eff} equal to 4.6×10^5 , thus giving a significance threshold of 10^{-7} after Bonferroni correction for multiple testing.

Pairwise interactions at the gene level within each group of biological categories were also tested using the AGGrEGATOR method.³³ Briefly, this method aggregates the SNP-SNP interaction P values to estimate a gene-based interaction P -value, based on the minimum- P -value, accounting for the number of SNP pairs and dependency between these SNP pairs as proposed by Connely and Boenhkne.³⁵ As 14,682 pairs of genes were tested, the Bonferroni significance threshold was equal to 3.4×10^{-6} .

2.4.2 | Replication in Saguenay–Lac-Saint-Jean and Medical Research Council

The SNP pairs exhibiting $P_{\text{SNP-SNP}} \leq 10^{-5}$ in EGEA were tested for replication in SLSJ and MRCA. For the MRCA imputed SNPs, logistic regression was carried out with allele dosage to take into account the uncertainty of imputation. The threshold of 10^{-5} was chosen to obtain both a strong indication of interaction and a reasonable number of SNP pairs (here 7, see Section 3) to be tested in the replication datasets. After Bonferroni correction for this number of SNP pairs, the threshold for significant replication was thus equal to 7×10^{-3} .

The gene-based interaction test was also applied to each replication sample for all pairs of genes with $P_{G-G} \leq 10^{-4}$ in EGEA. Three gene pairs were retained, leading to a Bonferroni threshold for significance of 0.017.

We then conducted a fixed-effects meta-analysis of SNP-SNP interactions in EGEA, SLSJ and MRCA for the SNP pairs detected in EGEA at $P_{\text{SNP-SNP}} \leq 10^{-5}$ and for the SNP pairs belonging to the gene pairs detected in EGEA at $P_{G-G} \leq 10^{-4}$. Gene $\times$ gene analyses were performed using AGGrEGATOR with P values of meta-analysis of SNP $\times$ SNP interaction test ($P_{\text{meta-SNP-SNP}}$) in EGEA, SLSJ and MRCA. The thresholds to declare significance were identical to those indicated above for analyses in the discovery sample (10^{-7} for SNP-SNP interactions and 3.4×10^{-6} for G-G interactions respectively).

Finally, to investigate whether the evidence for interaction was related to the presence of asthma in these asthma-ascertained samples, we performed a sensitivity analysis according to asthma status. In this sake, the SNP-SNP interaction tests were performed in subjects with and without asthma separately in the pooled EGEA, SLSJ and MRCA samples to have a sufficient sample size in each group of

subjects. Analyses were conducted in the pooled samples while adjusting for study and for principal components within study. Homogeneity of interaction between the two groups was tested. Gene-based interaction analyses were also repeated in subjects with and without asthma.

3 | RESULTS

3.1 | Descriptive statistics

Table 1 shows the number of genotyped subjects and their phenotypic distribution according to eczema status in the EGEA, SLSJ and MRCA samples. The total number of subjects was 1493 in EGEA, 1153 in SLSJ and 254 in MRCA. The proportion of subjects with eczema was similar in the three samples, ranging between 27% and 31%. The proportion of males was higher in subjects with eczema than without eczema in MRCA, while it was smaller in SLSJ and similar in EGEA. The reason may be that in MRCA, all subjects are children and most subjects with eczema are also asthmatic (89%) and it is known that the risk of asthma in children is higher among boys than among girls. The mean age was lower in subjects with eczema than in those without eczema in EGEA and SLSJ, while it was similar in these two groups in MRCA. The variation of age across studies, with younger subjects in MRCA and older subjects in SLSJ, is at least partly due to the ascertainment of the samples, only children in MRCA and the inclusion of fifty per cent of three-generation pedigrees in SLSJ. As expected, the proportion of subjects with asthma and atopy (defined by a positive skin prick test response to at least one aeroallergen) was much higher among individuals having eczema than among unaffected subjects. Note that in MRCA, whichever the eczema affection status, the proportion of asthmatics was higher than in the other two samples, while the proportion of atopics was smaller. Analysis of offspring only in MRCA and of subjects

belonging to parental and offspring generation in the other two samples might partly explain the higher proportion of asthmatics in MRCA families ascertained through asthmatic offspring. The smaller proportion of atopics in MRCA may be due to the smaller number of aeroallergens tested in MRCA (five allergens) than in the other two samples (ten or more allergens) (see Table 1).

3.2 | SNP-SNP and gene–gene interaction analyses

3.2.1 | Epidemiological study on the Genetics and Environment of Asthma discovery dataset

In the SNP-SNP interaction analysis, seven SNP pairs, three within genes of the “collagen” group and four within genes of the “epidermis and skin” group, had $P_{\text{SNP-SNP}} \leq 10^{-5}$ in EGEA (Table 2 and more details, Table S2). No pair reached the significance threshold of 10^{-7} . In the gene-based interaction analysis, three gene pairs (COL5A3 and MMP9, COL1A2 and COL6A1, CTSB and COL5A3) had $P_{\text{G-G}} \leq 10^{-4}$ (Table 3). None reached the significance threshold of 3.4×10^{-6} . As expected, these three pairs of genes included the SNP pairs with the smallest $P_{\text{SNP-SNP}}$ value.

3.2.2 | Saguenay–Lac-Saint-Jean and Medical Research Council replication datasets

Only one SNP pair (rs2287807 in COL5A3 and rs17576 in MMP9) among the pairs detected in EGEA was significantly replicated in SLSJ ($P_{\text{SNP-SNP}} = 1.24 \times 10^{-3}$, under the replication threshold of 7×10^{-3}) but not in MRCA ($P_{\text{SNP-SNP}} = 0.13$). However, meta-analysis of the EGEA, SLSJ and MRCA results for this SNP pair showed strong evidence for interaction ($P_{\text{meta-SNP-SNP}} = 1.11 \times 10^{-8}$, below the significance threshold of 10^{-7}) and strong improvement with respect to the interaction signal obtained in the EGEA discovery

Study	Phenotypes	No. (%)	Male sex (%)	Mean age (SD)	Asthma (%)	Atopy (%)
EGEA	All	1493 (100)	768 (51)	30.3 (0.44)	690 (46)	788 (53)
	ECZ–	1082 (72)	553 (51)	32.9 (0.51)	419 (39)	510 (47)
	ECZ+	411 (28)	215 (52)	23.3 (0.75)	271 (66)	278 (68)
SLSJ	All	1153 (100)	527 (46)	38.2 (0.64)	557 (48)	639 (56)
	ECZ–	844 (73)	421 (50)	39.9 (0.76)	366 (43)	424 (51)
	ECZ+	309 (27)	106 (34)	33.4 (1.17)	191 (62)	215 (70)
MRCA	All	254 (100)	144 (57)	11.2 (0.30)	176 (69)	104 (41)
	ECZ–	174 (69)	96 (55)	11.0 (0.37)	105 (60)	55 (32)
	ECZ+	80 (31)	48 (60)	11.5 (0.49)	71 (89)	49 (61)

TABLE 1 Phenotypic characteristics of genotyped individuals in the three asthma-ascertained samples

ECZ+, with eczema; ECZ–, without eczema; SD, standard deviation; EGEA, Epidemiological study on the Genetics and Environment of Asthma; MRCA, Medical Research Council; SLSJ, Saguenay–Lac-Saint-Jean.

Atopy: positive response of skin prick test to at least one allergen (house dust mite, cockroach, cat, timothy grass, olive, birch, ragweed, pellitory, Alternaria, Aspergillus and Cladosporium were tested in EGEA; house dust mite, dust, cat, herbs, trees, Alternaria, Aspergillus, Cladosporium, Hormodendrum and Penicillium in SLSJ; dust mite, cat, timothy grass, Aspergillus and Cladosporium in MRCA).

TABLE 2 SNP × SNP interaction test results in EGEA, SLSJ and MRCA for SNP pairs selected in EGEA ($P_{\text{SNP-SNP}} \leq 10^{-5}$)

SNP1	Gene	SNP2	Gene	EGEA $P_{\text{SNP-SNP}}$	SLSJ $P_{\text{SNP-SNP}}$	MRCA $P_{\text{SNP-SNP}}$	All samples ^a $P_{\text{meta-SNP-SNP}}$
Collagen							
rs1293288	CTSB	rs3745596	COL5A3	3.56×10^{-6}	8.65×10^{-1}	5.65×10^{-1}	1.20×10^{-3}
rs2521205	COL1A2	rs1053312	COL6A1	6.70×10^{-6}	4.36×10^{-1}	8.61×10^{-3}	2.88×10^{-2}
rs2287807	COL5A3	rs17576	MMP9	6.70×10^{-6}	1.24×10^{-3}	1.33×10^{-1}	1.11×10^{-8}
Epidermis or skin							
rs12854617	ATP8A2	rs6106524	TGM3	6.35×10^{-6}	4.18×10^{-1}	6.60×10^{-1}	4.69×10^{-4}
rs7586353	ABCA12	rs246406	ITGA2	6.73×10^{-6}	1.79×10^{-1}	3.37×10^{-1}	5.74×10^{-3}
rs6537155	GAB1	rs6026561	GNAS	8.60×10^{-6}	3.75×10^{-1}	8.73×10^{-1}	1.37×10^{-4}
rs3805251	GAB1	rs6026561	GNAS	9.25×10^{-6}	3.66×10^{-1}	8.78×10^{-1}	1.34×10^{-4}

EGEA, Epidemiological study on the Genetics and Environment of Asthma; MRCA, Medical Research Council; SLSJ, Saguenay–Lac-Saint-Jean.

^aMeta-analysis of interactive effects using fixed effect model.**TABLE 3** Gene × Gene interaction test results in EGEA, SLSJ and MRCA for SNP pairs selected in EGEA ($P_{\text{G-G}} \leq 10^{-4}$)

Gene1	Gene2	No of SNP pairs	EGEA $P_{\text{G-G}}^a$	SLSJ $P_{\text{G-G}}^a$	MRCA $P_{\text{G-G}}^a$	Meta-analysis $P_{\text{meta-G-G}}^b$
COL1A2	COL6A1	7	2.14×10^{-5}	2.87×10^{-1}	2.90×10^{-2}	7.18×10^{-2}
COL5A3	MMP9	8	4.49×10^{-5}	8.79×10^{-3}	5.23×10^{-1}	4.04×10^{-8}
COL5A3	CTSB	47	1.48×10^{-4}	2.61×10^{-2}	7.11×10^{-1}	9.86×10^{-3}

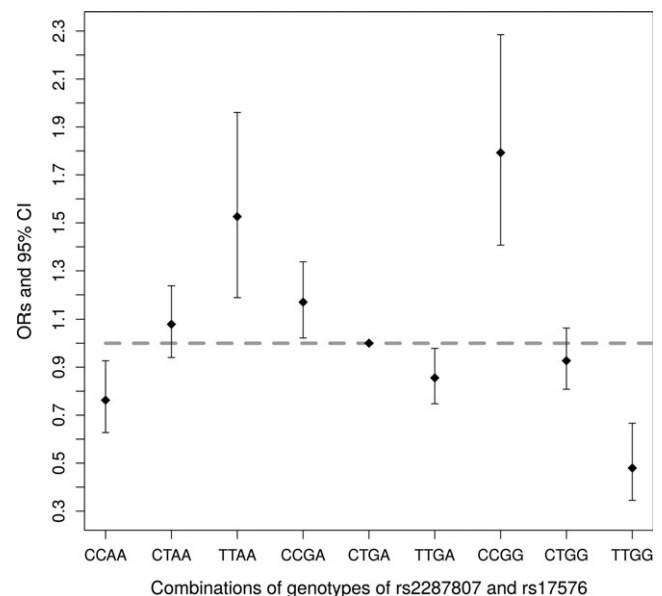
EGEA, Epidemiological study on the Genetics and Environment of Asthma; G-G, gene–gene; MRCA, Medical Research Council; SLSJ, Saguenay–Lac-Saint-Jean.

^aGene × gene analysis using AGGrEGATOR.^bGene × gene meta-analysis using AGGrEGATOR with P values of meta-analysis of SNP × SNP interaction test in EGEA, SLSJ and MRCA.

dataset ($P_{\text{SNP-SNP}} = 6.70 \times 10^{-6}$) (Table 2). This SNP pair was also the only one showing the same direction of interaction term in the three datasets. This SNP pair showed a pattern of interaction in which OR associated CC (or TT) genotype at rs2287807 (COL5A3) showed an opposite effect depending on the genotype AA (or GG) at rs17576 (MMP9) (Figure 1 and Table S3). More precisely, for subjects homozygous for allele A at rs17576, T allele at rs2287807 will increase the risk of eczema, while C allele will be protective. For the other subjects, heterozygous or homozygous for allele G, the effect of the alleles at rs2287807 will be inversed: The T allele will be protective and the C allele at risk.

In the gene-based interaction analysis, only one pair of genes (COL5A3 and MMP9) among the gene pairs detected in EGEA was significantly replicated in SLSJ ($P_{\text{G-G}} = 8.79 \times 10^{-3}$, below the replication threshold of 1.5×10^{-2}) but not in MRCA ($P_{\text{G-G}} = 0.52$). Moreover, this pair was also highly significant by the meta-analysis of the three datasets ($P_{\text{meta-G-G}} = 4.04 \times 10^{-8}$, well under the significance threshold of 3.4×10^{-6}). The other two gene pairs did not reach significance: $P_{\text{meta-G-G}} = 7.18 \times 10^{-2}$ for COL1A2 and COL6A1 pair, and $P_{\text{meta-G-G}} = 9.86 \times 10^{-3}$ for CTSB and COL5A3 pair (Table 3).

Finally, after stratification of the pooled sample according to asthma status, the SNP pair (rs2287807, rs17576) showed similar evidence for interaction in 1423 asthmatic ($P_{\text{SNP-SNP}} = 7.54 \times 10^{-5}$)

**FIGURE 1** Two-locus odds-ratio (ORs) and 95% confidence intervals for eczema estimated for the genotypic combinations of the two SNPs, rs2287807 (COL5A3) and rs17576 (MMP9). The ORs were computing using the estimates of SNP main effect and interaction obtained from the pooled analysis of Epidemiological study on the Genetics and Environment of Asthma (EGEA), Saguenay–Lac-Saint-Jean (SLSJ) and Medical Research Council (MRCA)

and in 1477 non-asthmatic subjects ($P_{\text{SNP-SNP}} = 1.55 \times 10^{-5}$). There was no heterogeneity between the two groups ($P_{\text{homogeneity}} = 0.16$), clearly indicating this interaction is not related to the presence of asthma. Likewise, in the gene-based interaction analyses, the *COL5A3-MMP9* gene pair showed similar results in asthmatic ($P_{\text{G-G}} = 3.09 \times 10^{-4}$) and non-asthmatic subjects ($P_{\text{G-G}} = 6.28 \times 10^{-5}$).

4 | DISCUSSION

This study identified the interactive effect of genetic variants in the *COL5A3* and *MMP9* genes on eczema risk in families ascertained through asthmatic subjects. Evidence for this interaction rests on the results obtained in the EGEA discovery sample and on the highly significant results obtained when the EGEA dataset was combined with the two SLSJ and MRCA replication datasets. This result is further strengthened by the fact that interaction was significant in both SNP-SNP and gene-based interaction analyses. Moreover, the interaction between these two genes in eczema susceptibility appears independently whichever the subject asthma status.

The EGEA results were significantly replicated in the French Canadian SLSJ dataset both at the SNP and gene level, but not in the MRCA sample. This might be partly explained by the shared French genetic background of EGEA and SLSJ families, but also by other phenotypic characteristics, shared by EGEA and SLSJ, but different in MRCA. In particular, age at onset is known to be a factor of genetic heterogeneity. Here, mean age was similar in EGA and SLSJ but much younger in MRCA, indirectly suggesting earlier onset of eczema and asthma in MRCA. The proportion of asthmatic and atopic subjects was similar in EGEA and SLSJ but different in MRCA. Other phenotypic differences not measured here such as severity of eczema or asthma could also have influenced the results.

Another explanation might be the reduced power in MRCA because of smaller sample size. Nevertheless, meta-analysis of EGEA, SLSJ and MRCA led to improvement of the SNP-SNP interaction signal ($P = 1.11 \times 10^{-8}$) compared to the one obtained in the meta-analysis of EGEA and SLSJ ($P = 2.55 \times 10^{-8}$).

Our study illustrates the advantage of SNP-SNP interaction analysis because the significant pair of SNPs was not detectable by single-SNP association test ($P = .81$ for rs2287807 and $.03$ for rs17576 in EGEA). Similarly, *COL5A3* and *MMP9* would not have been detected by gene-based association analysis ($P = .12$ for *COL5A3* and $.03$ for *MMP9* in EGEA).

Our findings also show the interest of a two-step strategy with selection of genes based on biological knowledge, prior to performing interaction analysis. Indeed, if the first step had been based on single-SNP association analysis, neither the genes nor the SNP pairs showing significant interaction would have been retained, because of the weak association signals observed in the discovery sample. Likewise, the SNP pair would not have been

detected by a one-step interaction strategy as the correction for multiple testing would have been too stringent (threshold for significance estimated to 10^{-13} for a genome-wide interaction study of a 500K chip³⁶).

We did not consider interaction across different GO classes because the hypothesis of a statistical interaction between genes belonging to the same biological process is more plausible. Similarly, we did not extend our search for interaction to other groups of genes, such as those involved in allergy and immunity, which have already been extensively investigated in atopy and related phenotypes such as asthma and eczema. Tested groups of genes included genes involved in regulatory T cell function,³⁷ toll-like receptor (TLR)-related pathway³⁸ and TH2 differentiation pathway (IL13, IL4, IL33, IL1RL1 and STAT6).^{15,39,40} Here, we focused on genes specifically related to eczema, a field poorly investigated until now, in order to reduce the multiple testing burden and to gain power. This strategy allowed detecting a significant interaction between the *COL5A3* and *MMP9* genes.

Note that the interaction detected here between two SNPs each with a very small or no marginal effect showed an inversion of the effect of genotypes at one variant depending on genotypes at the other variant. Such situations correspond to flip-flop models. The biological plausibility of flip-flop models may be discussed; however, several human diseases, including Alzheimer's disease,⁴¹ breast cancer,⁴² melanoma,⁴³ have been reported to be influenced by interaction of loci with no significant main effect at either locus and showing a flip-flop type of interaction.

The present study is the first one to investigate G-G interaction in the genetic susceptibility to eczema at a large-scale level. Neither *MMP9* nor *COL5A3* were detected in the three previous G-G interaction studies.¹⁴⁻¹⁸ We also verified in the GWAS-Catalog of Published Genome-Wide Association Studies (<http://www.ebi.ac.uk/gwas/>), that no significant association of SNPs in *COL5A3* nor in *MMP9* had been reported with eczema and asthma and any other allergy-related phenotypes, such as IgE sensitization, IgE levels. However, *MMP9* has been shown to be associated with either atopic⁴⁴ or non-atopic asthma⁴⁵ in candidate gene studies. To our knowledge, no association study of this gene with eczema has been conducted. However, it is of note that *MMP9* was shown to be over-expressed in the skin of patients with eczema⁴⁶ and differently expressed in lung tissue between pre- and post-corticosteroids.⁴⁷ It was also found that allergen exposure induces the release of *MMP9* into airway of allergic subjects.⁴⁸

SNP rs2287807 is located in intron 27 of the *COL5A3* gene. Examination of functional annotations using the HaploReg tool (<http://compbio.mit.edu/HaploReg>) shows that this SNP maps to promoter and enhancer histone marks in skin (fibroblast, melanocyte and keratinocyte cells) and is also DNase hypersensitive in skin cells. SNP rs17576 is located in exon 6 of the *MMP9* gene and corresponds to a missense substitution (Gln > Arg) at position 279 of the protein. This SNP also maps to promoter and enhancer histone marks in skin (fibroblast and keratinocyte cells).

The COL5A3 gene, located in 19p13.2 region, encodes an alpha chain for one of the low abundance fibrillar collagens. Type V collagen is found in tissues containing type I collagen, the collagen the most present in skin, and appears to regulate the assembly of heterotypic fibres composed of both type I and type V collagen. This gene belongs to both "collagen" and "epidermis and skin" groups defined here, and more precisely to the 3 GO biological process sub-categories, "collagen catabolic process," "collagen fibril organization" and "skin development."

Interestingly, the gene *MMP9* also belongs to the biological process sub-category "collagen catabolic process." *MMP9*, located on chromosome 20q13.12, is involved in the breakdown of extracellular matrix in normal physiological processes, such as tissue remodelling and collagen deposition. It also plays a significant role in inflammation by facilitating cellular traffic, including neutrophils and eosinophils. *MMP9* also codes for an enzyme which degrades type IV and V collagens. Studies showed experimental evidence of interaction (or binding) between *MMP9* and type IV and V collagens, eg COL4A1, COL4A2 and COL5A1, while COL5A3 and COL5A1⁴⁹ have also been experimentally shown to interact (to be in the same protein complex). Moreover, *MMP9* and COL5A3 were both showed experimentally to interact with the same protein COL1A2,^{49,50} which also belongs to the category of collagen catabolic process.

All these observations support our finding of a statistical interaction effect between COL5A3 and *MMP9*, which may be accounted for by a degradation of COL5A3 by *MMP9* influencing eczema susceptibility.

In conclusion, the present study detected two new genes having an interactive effect in eczema susceptibility. It highlights that G-G interaction analyses, beyond single-SNP association analysis, can greatly contribute to the identification of new genes. Further confirmation of the interaction of *MMP9* and COL5A3 as well as functional studies is needed to provide a full understanding of the role of these genes in eczema.

ACKNOWLEDGEMENTS

We thank the EGEA cooperative group:

EGEA cooperative group

Coordination: V Siroux (epidemiology, PI since 2013); F Deme-nais (genetics); I Pin (clinical aspects); R Nadif (biology); F Kauffmann (PI 1992-2012).

Respiratory epidemiology: Inserm ex-U 700, Paris: M Korobaeff (Egea1), F Neukirch (Egea1); Inserm ex-U 707, Paris: I Annesi-Maesano (Egea1-2); Inserm ex-U 1018, Villejuif: F Kauffmann, MP Oryszczyn (Egea1-2); Inserm U 1168, Villejuif: N Le Moual, R Nadif, R Varraso; Inserm U 1209 Grenoble: V Siroux.

Genetics: Inserm ex-U 393, Paris: J Feingold; Inserm U 946, Paris: E Bouzigon, F Deme-nais, MH Dizier; CNG, Evry: I Gut (now CNAG, Barcelona, Spain), M Lathrop (now Univ McGill, Montreal, Canada).

Clinical centers: Grenoble: I Pin, C Pison; Lyon: D Ecochard (Egea1), F Gormand, Y Pacheco; Marseille: D Charpin (Egea1), D Vervloet (Egea1-2); Montpellier: J Bousquet; Paris Cochin: A Lockhart (Egea1), R Matran (now in Lille); Paris Necker: E Paty (Egea1-2), P Scheinmann (Egea1-2); Paris-Trousseau: A Grimfeld (Egea1-2), J Just.

Data and quality management: Inserm ex-U155 (Egea1): J Hochez; Inserm U 1168, Villejuif: N Le Moual; Inserm ex-U780: C Ravault (Egea1-2); Inserm ex-U794: N Chateigner (Egea1-2); Grenoble: J Quentin (Egea1-2).

SLSJ: The Canada Research Chair in Environment and Genetics of Respiratory Diseases and Allergy 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 director of the SLSJ cohort. She is also member of Allergen network.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

J. Just  <http://orcid.org/0000-0002-5646-2429>

M.-H. Dizier  <http://orcid.org/0000-0001-8460-7667>

REFERENCES

1. Dold S, Wjst M, von Mutius E, Reitmeir P, Stiepel E. Genetic risk for asthma, allergic rhinitis, and atopic dermatitis. *Arch Dis Child*. 1992;67:1018-1022.
2. Wadonda-Kabondo N, Sterne JA, Golding J, Kennedy CT, Archer CB, Dunnill MG. Association of parental eczema, hayfever, and asthma with atopic dermatitis in infancy: birth cohort study. *Arch Dis Child*. 2004;89:917-921.
3. Bin L, Leung DY. Genetic and epigenetic studies of atopic dermatitis. *Allergy Asthma Clin Immunol*. 2016;12:52.
4. Marenholz I, Rivera VA, Esparza-Gordillo J, et al. Association screening in the Epidermal Differentiation Complex (EDC) identifies an SPRR3 repeat number variant as a risk factor for eczema. *J Invest Dermatol*. 2011;131:1644-1649.
5. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat Genet*. 2011;44:187-192.
6. Palmer CN, Irvine AD, Terron-Kwiatkowski A, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet*. 2006;38:441-446.
7. Kabesch M, Carr D, Weiland SK, von Mutius E. Association between polymorphisms in serine protease inhibitor, kazal type 5 and asthma phenotypes in a large German population sample. *Clin Exp Allergy*. 2004;34:340-345.
8. Esparza-Gordillo J, Weidinger S, Folster-Holst R, et al. A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat Genet*. 2009;41:596-601.

9. Soderhall C, Marenholz I, Kerscher T, et al. Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol.* 2007;5:e242.
10. Gharagozlou M, Farhadi E, Khaledi M, et al. Association between the interleukin 6 genotype at position -174 and atopic dermatitis. *J Invest Allergol Clin Immunol.* 2013;23:89-93.
11. Esparza-Gordillo J, Schaarschmidt H, Liang L, et al. A functional IL-6 receptor (IL6R) variant is a risk factor for persistent atopic dermatitis. *J Allergy Clin Immunol.* 2013;132:371-377.
12. Portelli MA, Hodge E, Sayers I. Genetic risk factors for the development of allergic disease identified by genome-wide association. *Clin Exp Allergy.* 2015;45:21-31.
13. Paternoster L, Standl M, Waage J, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet.* 2015;47:1449-1456.
14. Namkung JH, Lee JE, Kim E, et al. Association of single nucleotide polymorphisms in the IL-12 (IL-12A and B) and IL-12 receptor (IL-12Rbeta1 and beta2) genes and gene-gene interactions with atopic dermatitis in Koreans. *J Dermatol Sci.* 2010;57:199-206.
15. Ziyab AH, Davies GA, Ewart S, et al. Interactive effect of STAT6 and IL13 gene polymorphisms on eczema status: results from a longitudinal and a cross-sectional study. *BMC Med Genet.* 2013;14:67.
16. Shen C, Liu L, Jiang Z, et al. Four genetic variants interact to confer susceptibility to atopic dermatitis in Chinese Han population. *Mol Genet Genomics.* 2015;290:1493-1498.
17. Weidinger S, Baurecht H, Wagenpfeil S, et al. Analysis of the individual and aggregate genetic contributions of previously identified serine peptidase inhibitor Kazal type 5 (SPINK5), kallikrein-related peptidase 7 (KLK7), and filaggrin (FLG) polymorphisms to eczema risk. *J Allergy Clin Immunol.* 2008;122:560-568. e4
18. O'Regan GM, Campbell LE, Cordell HJ, Irvine AD, McLean WH, Brown SJ. Chromosome 11q13.5 variant associated with childhood eczema: an effect supplementary to filaggrin mutations. *J Allergy Clin Immunol.* 2010;125:170-174. e1-2
19. Culverhouse R, Suarez BK, Lin J, Reich T. A perspective on epistasis: limits of models displaying no main effect. *Am J Hum Genet.* 2002;70:461-471.
20. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet.* 2005;37:413-417.
21. Herold C, Steffens M, Brockschmidt FF, Baur MP, Becker T. INTERSNP: genome-wide interaction analysis guided by a priori information. *Bioinformatics.* 2009;25:3275-3281.
22. Babron MC, Guilleud-Bataille M, Sahbatou M, Demenais F, Genin E, Dizier MH. Comparative power of family-based association strategies to detect disease-causing variants under two-locus models. *Genet Epidemiol.* 2012;36:848-855.
23. Sun LD, Xiao FL, Li Y, et al. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nat Genet.* 2011;43:690-694.
24. Hirota T, Takahashi A, Kubo M, et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat Genet.* 2012;44:1222-1226.
25. Ellinghaus D, Baurecht H, Esparza-Gordillo J, et al. High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. *Nat Genet.* 2013;45:808-812.
26. Schaarschmidt H, Ellinghaus D, Rodriguez E, et al. A genome-wide association study reveals 2 new susceptibility loci for atopic dermatitis. *J Allergy Clin Immunol.* 2015;136:802-806.
27. 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-S129.
28. Guilleud-Bataille M, Bouzigon E, Annesi-Maesano I, et al. Evidence for linkage of a new region (11p14) to eczema and allergic diseases. *Hum Genet.* 2008;122:605-614.
29. Dizier MH, Margartitte-Jeannin P, Madore AM, et al. The ANO3/MUC15 locus is associated with eczema in families ascertained through asthma. *J Allergy Clin Immunol.* 2012;129:1547-1553. e3
30. 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.
31. 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.
32. 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.
33. Emily M. AGGREGATOR: a Gene-based GENE-Gene interACTiOn test for case-control association studies. *Stat Appl Genet Mol Biol.* 2016;15:151-171.
34. Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genet Epidemiol.* 2008;32:567-573.
35. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. *Am J Hum Genet.* 2007;81:1158-1168.
36. Becker T, Herold C, Meesters C, Mattheisen M, Baur MP. Significance levels in genome-wide interaction analysis (GWIA). *Ann Hum Genet.* 2011;75:29-35.
37. Bottema RW, Kerkhof M, Reijmerink NE, et al. Gene-gene interaction in regulatory T-cell function in atopy and asthma development in childhood. *J Allergy Clin Immunol.* 2010;126:338-346. 46.e1-10.
38. Reijmerink NE, Bottema RW, Kerkhof M, et al. TLR-related pathway analysis: novel gene-gene interactions in the development of asthma and atopy. *Allergy.* 2010;65:199-207.
39. Savenije OE, Mahachie John JM, Granell R, et al. Association of IL33-IL-1 receptor-like 1 (IL1RL1) pathway polymorphisms with wheezing phenotypes and asthma in childhood. *J Allergy Clin Immunol.* 2014;134:170-177.
40. Kabesch M, Schedel M, Carr D, et al. IL-4/IL-13 pathway genetics strongly influence serum IgE levels and childhood asthma. *J Allergy Clin Immunol.* 2006;117:269-274.
41. Martin BK, Frangakis CE, Rosenberg PB, et al. Design of Depression in Alzheimer's Disease Study-2. *Am J Geriatr Psychiatry.* 2006;14:920-930.
42. Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet.* 2001;69:138-147.
43. Vaysse A, Fang S, Brossard M, et al. A comprehensive genome-wide analysis of melanoma Breslow thickness identifies interaction between CDC42 and SCIN genetic variants. *Int J Cancer.* 2016;139:2012-2020.
44. Nakashima K, Hirota T, Obara K, et al. A functional polymorphism in MMP-9 is associated with childhood atopic asthma. *Biochem Biophys Res Comm.* 2006;344:300-307.
45. Pinto LA, Depner M, Klopp N, et al. MMP-9 gene variants increase the risk for non-atopic asthma in children. *Respir Res.* 2010;11:23.
46. Harper JI, Godwin H, Green A, et al. A study of matrix metalloproteinase expression and activity in atopic dermatitis using a novel skin wash sampling assay for functional biomarker analysis. *Br J Dermatol.* 2010;162:397-403.

47. Laprise C, Sladek R, Ponton A, Bernier MC, Hudson TJ, Laviolette M. Functional classes of bronchial mucosa genes that are differentially expressed in asthma. *BMC Genom.* 2004;5:21.
48. Kelly EA, Busse WW, Jarjour NN. Increased matrix metalloproteinase-9 in the airway after allergen challenge. *Am J Respir Crit Care Med.* 2000;162:1157-1161.
49. Niyibizi C, Fietzek PP, van der Rest M. Human placenta type V collagens. Evidence for the existence of an alpha 1(V) alpha 2(V) alpha 3 (V) collagen molecule. *J Biol Chem.* 1984;259:14170-14174.
50. Makela M, Salo T, Larjava H. MMP-9 from TNF alpha-stimulated keratinocytes binds to cell membranes and type I collagen: a cause for extended matrix degradation in inflammation? *Biochem Biophys Res Comm.* 1998;253:325-335.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Margaritte-Jeannin P, Babron M-C, Laprise C, et al. The COL5A3 and MMP9 genes interact in eczema susceptibility. *Clin Exp Allergy.* 2018;48:297–305.
<https://doi.org/10.1111/cea.13064>