Margaritte-Jeannin, P., Budu-Aggrey, A., Ege, M., Madore, A-M., Linhard, C., Mohamdi, H., von Mutius, E., Granell, R., Demenais, F., Laprise, C., Bouzigon, E., & Dizier, M-H. (2021). Identification of OCA2 as a novel locus for the co-morbidity of asthma-plus-eczema. *Clinical and Experimental Allergy*, *52*(1), 70-81. https://doi.org/10.1111/cea.13972

Identification of *OCA2* as a novel locus for the co-morbidity of asthma-plus-eczema

Running title: *OCA2* a novel locus for asthma-plus-eczema

P Margaritte-Jeannin (1), A Budu-Aggrey (2), M Ege (3), AM Madore (4), C Linhard (1), H Mohamdi (1), E von Mutius (3), R Granell (2), F Demenais (1), C Laprise (4), E Bouzigon (1), MH Dizier (1)

(1) Université de Paris, UMRS 1124, INSERM, Paris, France

(2) Medical Research Council (MRC) Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

(3) Dr von Hauner Children's Hospital, Ludwig Maximilian University; Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich, Germany,

(4) Département des sciences fondamentales, Centre intersectoriel en santé durable (CISD), Université du Québec à Chicoutimi, Saguenay, QC, Canada

Corresponding author:

Marie-Hélène Dizier,

Université de Paris, UMRS 1124, INSERM,

Campus Saint-Germain-des-Prés

45 rue des Saints Pères - 75006 Paris, France.

Email: marie-helene.dizier@inserm.fr

Acknowledgements:

EGEA cooperative group. Coordination: V Siroux (epidemiology, PI since 2013); F Demenais (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 U 1018, Villejuif: O Dumas, F Kauffmann, N Le Moual, R Nadif, MP Oryszczyn (Egea1-2), R Varraso; Inserm U 1209 Grenoble: J Lepeule, V Siroux. Genetics: Inserm ex-U 393, Paris: J Feingold; Inserm UMR 1124, Paris: E Bouzigon, MH Dizier, F Demenais; 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 management and quality: Inserm ex-U155, Paris: J Hochez (Egea1); Inserm U 1018, Villejuif: N Le Moual, L Orsi; Inserm ex-U780, Villejuif: C Ravault (Egea1-2); Inserm ex-U794, Evry: N Chateigner (Egea1-2); Inserm UMR 1124, Paris: H Mohamdi; Inserm U1209, Grenoble: A Boudier, J Quentin (Egea1-2).

SLSJ: C Laprise is part of the Quebec Respiratory Health Network (RHN;), 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 Diseases and Allergies.

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.

GABRIELA: acknowledgement to Jon Genuneit and Roger Lauener for the field centres and to Michael Kabesch for DNA extraction etc.

ALSPAC: The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and AB-A will serve as guarantors for the contents of this paper.

"We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses"

Funding:

AB-A works in a research unit funded by the UK Medical Research Council (MC_UU_00011/1). ALSPAC GWAS data were generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Raquel Granell will serve as guarantors for the contents of this paper.

The funding source for GABRIELA was: "This work was supported by the European Commission as part of GABRIEL, contract number 018996 under the Integrated Program LSH-2004-1.2.5-1.).

The Saguenay−Lac-Saint-Jean asthma familial cohort is supported by the Canada Research Chair in the Environment and Genetics of Respiratory Diseases and Allergies.

Genotyping of GABRIELA study, Saguenay−Lac-Saint-Jean (SLSJ) asthma familial study and the Epidemiological study on the genetics and environment of asthma (EGEA) was supported by grants from the European Commission (No. LSHB-CT-2006-018996- GABRIEL) and the Wellcome Trust (WT084703MA).

Conflict Of Interest

All authors have no conflicts of interest.

Author contributions:

MHD conducted the design. PMJ performed the largest part of data analysis, ABA, CLinhard and HM a remaining part. MHD interpreted the findings and drafted the initial version of the manuscript. EB and PMJ contributed to drafting the manuscript. EvM, CLaprise, AMM, RG, ME, EB and MHD contributed to the data acquisition. All authors revised the manuscript and provided final approval of the version to be published.

Ethical statement:

Information may be found online in Supplementary Material

Data Availability Statement:

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

ABSTRACT

Background: Numerous genes have been associated with the three most common allergic diseases (asthma, allergic rhinitis or eczema) but these genes explain only a part of the heritability. In the vast majority of genetic studies, complex phenotypes such as co-morbidity of two of these diseases, have not been considered. This may partly explain missing heritability.

Objective: To identify genetic variants specifically associated with the co-morbidity of asthma-plus-eczema.

Methods: We first conducted a meta-analysis of four GWAS (genome-Wide Association Study) of the combined asthma-plus-eczema phenotype (total of 8,807 European-ancestry subjects of whom 1,208 subjects had both asthma and eczema). To assess whether the association with SNP(s) was specific to the co-morbidity, we also conducted a meta-analysis of homogeneity test of association according to disease status ("asthma-plus-eczema" *versus* the presence of only one disease "asthma only or eczema only"). We then used a joint test by combining the two test statistics from the co-morbidity-SNP association and the phenotypic heterogeneity of SNP effect meta-analyses.

Results: Seven SNPs were detected for specific association to the asthma-plus-eczema co-morbidity, two with significant and five with suggestive evidence using the joint test after correction for multiple testing. The two significant SNPs are located in the *OCA2* gene (Oculocutaneous Albinism II), a new locus never detected for significant evidence of association with any allergic disease. This gene is a promising candidate gene, because of its link to skin and lung diseases, and to epithelial barrier and immune mechanisms.

Conclusion: Our study underlines the importance of studying sub-phenotypes as co-morbidities to detect new susceptibility genes.

Keywords: GWAS, asthma, eczema, co-morbidity, phenotypic heterogeneity, EGEA, SLSJ, ALSPAC, GABRIELA

Key messages:

• A genome wide association study to identify new susceptibility genes for the phenotype of asthma-plus-eczema

• We identified OCA2, a new gene associated specifically with the phenotype of asthma-plus-eczema

• The candidate gene OCA2 is linked to skin and lung diseases, epithelial barrier and immunology

INTRODUCTION

The three most common allergic diseases, asthma, atopic dermatitis (or eczema) and allergic rhinitis (AR) may share genetic determinants as suggested by their strong associations at both the individual and family levels[1]. However, genes specifically involved in each disease also exist. Most genetic studies have focused specifically on either one of the allergic diseases and numerous susceptibility genes for asthma, AR and eczema have been found[2]. Many of these genes are involved in the immune response, and are not specifically associated to one of the diseases. Other genes related to epithelial barrier dysfunction as *C11orf30* a transcriptional regulator in keratinocyte, are also shared by the three diseases[3,4]. But some other genes were found to be more specifically associated to one disease, such as asthma genes *ORMDL3/GSDML*[5,6] involved in remodeling or eczema genes *SPRR3*[7], *SPINK5*[8] and *OVOL1*[9] involved in skin or epidermis development.

However, all the genes found associated with these diseases explain only a part of the heritability[10]. In most genetic studies, complex mechanisms such as Gene-Gene and Gene-Environment interactions or complex phenotypes, such as those taking into account simultaneously the three allergic diseases, were not considered. That may explain a part of this missing heritability. Few recent studies have considered the three allergic diseases simultaneously, rather in the sense of allergic disease defined by the presence of at least one of the diseases: asthma, eczema or AR. These studies indeed allowed detection of numerous new genes and loci especially involved or related to mechanisms affecting function of immune and epithelial cells[11-13].

Other studies have focused on the consideration of these diseases rather in the sense of co-morbidity. Studying more precise sub-phenotypes such as co-morbidities, could allow the detection of new genes. For example, some studies have focused on the co-morbidity of

asthma associated with allergic rhinitis and have detected new genes specifically associated to this co-morbidity, such as the *NFIA*[14], *ZBTB, CLEC16A*[15] and *MTNR1A*[16] genes.

It seemed of interest to study the genetics of the co-morbidity of asthma associated with eczema, which has less been investigated. From our knowledge, only one genetic study has studied this co-morbidity, an asthma-plus-eczema GWAS meta-analysis[17]. For the phenotype of this co-morbidity, the particular hypothesis of the atopic march was considered. It corresponds to a disease progression from infantile eczema to asthma in childhood. This study led to the detection of several genes already found to be associated with asthma and/or eczema and two other loci detected for allergic disease for the first time: a locus encompassing the *EFHC1* gene and another one located between *TMTC2* and *SLC6A15* genes.

To discover new genes in addition to those already found to be associated with either asthma or eczema, our goal here was to identify genetic variants specifically associated with the co-morbidity of asthma-plus-eczema. GWAS meta-analysis was conducted across four independent populations of European ancestry. Firstly, association with the co-morbidity of asthma-plus-eczema was investigated and then, to assess the specificity of the association with the co-morbidity, the homogeneity of the association was tested according to disease status defined by the presence of the two diseases “asthma-plus-eczema” *versus* the presence of only one disease “asthma only or eczema only”. A joint test was then applied by combining the two test statistics from the co-morbidity-SNP association and the phenotypic heterogeneity of SNP effect meta-analyses. The present study is thus the first GWAS meta-analysis of asthma-plus-eczema, testing the specificity of the associations with this co-morbidity.

MATERIAL AND METHOD

Populations

We studied 8,807 European-ancestry subjects from four independent studies, two population-based (GABRIELA and ALSPAC) and two family (EGEA and SLSJ) studies, which were part of the European consortium on the genetics of asthma[18]. A brief description of these studies with the definition of asthma and eczema phenotypes is provided in the Supplementary Material.

Genetic data

The EGEA, SLSJ and GABRIELA samples were genotyped using the Illumina 610-Quad, as part of the European Gabriel asthma GWAS consortium[18]. The ALSPAC samples were genotyped using the Illumina Human Hap 550-quad array (Illumina, Inc., San Diego, CA) by 23andMe. In all datasets, stringent quality criteria (QC) were used to select both individuals and SNPs as described previously[18]. To control for ethnicity/population stratification in the analysis, ancestry analysis was carried out in each dataset using the EIGENSTRAT2.0 software and HapMap data (CEU, YRI, JPT and CHB). Based on this analysis, putative non-European samples were flagged as outliers and eliminated from any subsequent genetic analyses.

Genetic association analyses were first conducted using genotyped data in EGEA, SLSJ and GABRIELA and imputed data for ALSPAC (imputation using the panel of the 1000 Genome Project, phase 1, version 3, release Dec 2013) for all SNPs included in both the 610-Quad chip and the 1000 genome panel. Then, to further investigate new loci showing evidence of association with the co-morbidity of asthma-plus-eczema, we used imputed SNPs in a region of 1Mb, 500kb on both sides of the top SNP(s) of each locus (imputed SNPs from HapMap2 (release 22) for EGEA, SLSJ and GABRIELA and 1000 genome for ALSPAC) .

Statistical analysis

For all association analyses in each of the four studies, we performed logistic regression using Stata® V14.1 or PLINK 1.9 assuming an additive model for SNP effect. Informative principal components for within-Europe diversity were included as covariates in all analyses. For the two family datasets (EGEA and SLSJ), logistic regression considered familial dependencies through the cluster within family and robust variance options of the logit function. To take into account the stratified random sampling in GABRIELA, inverse probability weights were introduced in the logistic regression analyses. Then, SNP effect estimates on disease status of the four studies were meta-analysed using a fixed-effects (inverse variance) model in order to increase power and to obtain more robust findings. As described below, different status for cases and/or controls were considered and subsequently meta-analysed. The whole strategy of analysis is described in Figure 1 and presented in the following paragraphs.

We firstly conducted a genome-wide meta-analysis of association between SNPs and the co-morbidity of asthma-plus-eczema, where cases were defined as having both asthma and eczema, and controls had neither asthma nor eczema; this test will be referred as the 'co-morbidity association test' hereafter. However, a SNP associated with asthma (including cases with asthma having or not eczema) or eczema (including cases with eczema having or not asthma) can be found associated with the co-morbidity asthma plus eczema. Consequently, an association shown between asthma-plus-eczema *vs.* no asthma no eczema is not sufficient to prove the specificity of the association to the co-morbidity. To verify this specificity of association, we then tested by meta-analysis, the homogeneity of SNP association between the two following phenotypes: "asthma-plus-eczema" *versus* "asthma alone or eczema alone" comparing by logistic regression SNP-association between these

two case groups. The latter test will be referred to hereafter as 'the phenotypic homogeneity test'. Such homogeneity tests of SNP effect between sub-phenotypes have been shown to have equivalent power to that of a multinomial regression-based test of heterogeneity.[19]

We then conducted on the whole genome, a joint test by combining the two test statistics obtained from the co-morbidity-SNP association and the phenotypic heterogeneity of SNP effect meta-analyses. These two tests follow a χ^2 distribution with 1 df, we summed their χ^2_{1df} results obtained at each SNP. However, we could not assume that this sum of χ^2_{1df} results follows a χ^2_{2df} distribution because the two tests were not independent due to sample overlapping, the two tests shared the individuals having asthma plus eczema. We used, as proposed by Ferrari [20], the approximation of the sum of correlated χ^2 variables by a parametrized gamma distribution with parameters depending on the Pearson correlation estimated between statistics obtained by the two tests.

For the joint test, the Bonferroni correction was applied to the Meff (effective number of independent SNPs calculated after discarding the dependence between tests due to Linkage Disequilibrium (LD) between SNPs from the total number of SNPs). In each of the four datasets, subjects' stratification according to "asthma plus eczema", "neither asthma nor eczema" or "asthma only or eczema only" status, led in each sub-phenotype group to small sample size per genotype. We thus selected only SNPs with MAF≥0.10 and/or SNPs having in each dataset sufficient expected sample size per genotypes (≥5 in either affected or unaffected subjects) depending on whether imputed or genotyped data were available. From the total number of tested SNPs (286,679), the Meff was estimated to 143,414 with the method of Li and Ji[21] and the threshold for significance to P=3.5x10^{-7} (0.05/143414).

We reported SNPs as showing significant (or suggestive) evidence of association specifically to the co-morbidity, those showing 1) a significant (or suggestive $P \leq 10^{-6}$) result with the joint test and 2) strong signals ($P \leq 10^{-4}$) for both the 'co-morbidity association test' and the 'phenotypic homogeneity test'.

Consistency of results across the four studies was assessed by use of the I^2 statistic, which describes the percentage of variation across studies that is due to heterogeneity rather than chance[22,23]. Thus to select only consistent results across studies, we applied $I^2 \leq 24\%$ as criteria to retain SNPs for both 'co-morbidity association test' and 'phenotypic homogeneity test', indicating no or little heterogeneity[24] of results across studies.

Next, to assess the consistency of our results according to the age at onset of asthma, all analyses were also conducted considering for asthma status, childhood-onset asthma before 16 years of age.

Finally, we repeated all analyses using imputed SNPs located in 1Mb region around the top genotyped SNPs of each locus showing significant or suggestive evidence of association.

Expression Quantitative Trait loci (eQTL) analysis and functional annotation

We investigated whether the SNPs (or their proxies, $r^2 \geq 0.8$) showing significant or suggestive evidence of association specifically to asthma-plus-eczema were cis-expression quantitative trait loci (cis-eQTLs). We queried existing eQTL databases in multiple tissues (GTEx[25], eQTLGen[26], BIOSQTL[27], Muther[28] and GHSExpress[29] browsers). Functional annotations of these SNPs (or proxies) were also done using ROADMAP and ENCODE (Encyclopedia of DNA Elements) data provided by the HaploReg tool[30].

RESULTS

Phenotypic description of the samples

The sample size of each of the four studies by affection status stratum is indicated in table I. Due to the mode of ascertainment, the proportion of subjects having both asthma and eczema was the strongest in EGEA (26%) and in SLSJ (23%) and the smallest in GABRIELA (16%) and in ALSPAC (10%). The proportion of men was similar in the four datasets, ranging from 46% to 57%. In contrast, the subjects were the oldest in SLSJ with a mean age equal to 23.7, then in EGEA and ALSPAC with respective mean ages equal to 16.5 and 13.9 and they were the youngest in GABRIELA with a mean age of 9.

Results of meta-analyses

The SNPs detected as specifically associated with the co-morbidity of asthma-plus-eczema are presented in Table II and in more details in Table S1. In addition, for SNPs presented in table II, the results obtained in each dataset are shown in Table S2. There was no inflation in the statistical tests with genomic inflation factor estimated to 1.02, 0.99 and 1.01 for the 'co-morbidity association test', 'phenotypic homogeneity test' and joint test respectively (see QQ plots in Figure S1). Results of the meta-analysis of the joint test are shown in the Manhattan plot in Figure 2.

When testing the association between SNPs and the co-morbidity of asthma-plus-eczema, 32 SNPs were found associated with "asthma-plus-eczema" at $P \leq 10^{-4}$, among which two were excluded because of $I^2 \geq 24\%$. When testing the specificity of this association, 25 SNPs reached the threshold of 10^{-4}, with none of them excluded because of $I^2 \geq 24\%$. Finally, two SNPs and five SNPs showed respectively significant ($P \leq 3.5 \times 10^{-7}$) and suggestive ($P \leq 10^{-6}$) results using the joint test, and had $P \leq 10^{-4}$ for both the 'co-morbidity association test' and the 'phenotypic homogeneity test'. None of these seven SNPs were excluded because of

I^2≥24%. The two SNPs (rs4778192 and rs2703978) showing significant evidence of association with the joint test were located in *OCA2* gene. The five SNPs with suggestive evidence, were located in *OCA2* (rs2311469, rs4778189 and rs2594897), *TBC1D14* (rs10937762) and *LRP1B* (rs1402470) genes. Forest plots are shown in Figure S2 for association and phenotypic homogeneity tests for each of the two significant SNPs (rs4778192 and rs2703978).

When the analyses were restricted to childhood-onset asthma, the five SNPs located in *OCA2* gene were detected in the same manner (two of them with significant evidence of association specifically to asthma-plus-eczema and the three others with suggestive evidence), while no SNP, including those in *LRP1* and *TBC1D14* was detected elsewhere.

In the regions of the three detected loci, we repeated association analyses using imputed SNPs located in a 1Mb region around the genotyped lead SNPs. These analyses supported all our initial findings, with similar or slightly improved significance of the results compared to those observed with genotyped SNPs. Moreover, additional signals with similar significance level were detected at imputed SNPs in *OCA2* gene (see Figure 3). The imputed SNP with highest significance level, rs924318, was in LD with the genotyped significant SNPs, strong LD with rs2703978 (r^2= 0.87) but not with rs4778192 (r^2=0.30). However, further conditional regression analyses of both co-morbidity association and phenotypic homogeneity tests showed that rs4778192 and rs2703978 were each with the joint test no longer significantly associated with asthma-plus-eczema (*P*≥0.05) when conditioning on rs924318. This indicates that the three SNPs represent the same association signal.

<u>eQTL analysis and functional annotation</u>

By interrogating databases of gene expression in target tissues, we identified a proxy rs10013696 of rs10937762 (r^2=0.82) located in *TBC1D14* that was associated with *TBC1D14*

expression in skin (P=5x10^{-9})[25]. We also identified two SNPs in *OCA2*, rs2594897 and rs2703978, associated in the whole blood with *HERC2* expression (P=5x10^{-12})[26]. Functional annotation of the three loci is presented in details in Table S3.

The associated SNPs located in *OCA2* and in *TBC1D14*, map to promotor and enhancer histone marks and DNase I hypersensitivity sites in numerous lung and skin cells, and included transcription factor (TF) binding sites that will be described in more detail in the discussion.

DISCUSSION

The aim of this study was to discover new genes, different from those already found associated with asthma or eczema, by identifying genetic variants specifically associated with the co-morbidity of asthma-plus-eczema. Our study identified a new locus significantly and specifically associated with the co-morbidity and located on chromosome 15q13 in the *OCA2* gene. Two other loci located on chromosomes 2q22 in the *LRP1B* gene and 4p16 in the *TBC1D14* gene were detected with a suggestive evidence of specific association to the co-morbidity. We did not conduct analysis in discovery sample(s) followed by replication analyses in independent sample(s). We directly conducted meta-analyses of four independent datasets and to ensure validity of our findings, we required strong consistency of results across studies assessed by I^2 (percentage of variation across studies that is due to heterogeneity rather than chance[22,23]) ≤24% indicating no or little heterogeneity[24]. We found strong consistency of results (I^2<10%) for the two *OCA2* and *TBC1D14* loci, but somehow less consistent results for the *LRP1B* locus (20%<I^2<24%).

For the three loci detected (*OCA2*, *TBC1D14 and LRP1B*), specificity of the association to the co-morbidity was also verified by an absence of association with the phenotype of "asthma only or eczema only" *vs* "no asthma and no eczema" and by a smaller evidence of association with each one of the diseases, asthma and eczema when tested separately.

All our findings were also well supported, by repeated association analyses using imputed SNPs to obtain a denser map of the regions of the three detected loci. These analyses strengthened the original findings, particularly for those in the *OCA2* gene.

To our knowledge, the single genetic study of asthma-plus-eczema reported to date in the literature is a GWAS meta-analysis of the co-morbidity defined as the atopic march[17]. This study was conducted to search for genetic associations with this co-morbidity but without

testing the specificity of the associations. The study led to the detection of five loci already found associated with one of the allergic diseases. For three of these loci (*IKZF3*[31,32,15], *AP5B1/OVOL1*[9] and *IL4/KIF3A*[9]), we found suggestive evidence of association with asthma-plus-eczema (10^{-5}<P<10^{-2}) in our present study. These results are not surprising since the "atopic march" meta-analysis enclosed the four datasets included in the current study. But for these three loci, there was not any indication of the specificity of this association with the asthma-plus-eczema co-morbidity. Furthermore, the two new loci identified in the "atopic march" meta-analysis (*EFHC1* and between *SLC6A15* and *TMTC2*) did not show indication of association in the current study (*P*>0.01). Regarding *CRNN/LCE5A* (*FLG*)[33], a known eczema locus that was also identified in the atopic march meta-analysis, we found evidence of association with asthma-plus-eczema (*P*=2x10^{-6}). However, we did not retain this locus due to the large measure of heterogeneity detected across studies (I^2>0.50).

None of our novel findings have been previously reported by published GWAS (GWAS-Catalog of Published Genome-Wide Association Studies, http://genome.gov/gwastudies) with significant evidence of association for asthma, eczema or more generally for allergic diseases. However, previous GWAS showed significant evidence of association between genetic variants belonging to *OCA2* loci/gene and diseases or phenotypes related to skin such as melanoma[34,35], cutaneous squamous cell carcinoma[36], and skin pigmentation[37]. Note that, underlying mechanisms shared by melanoma and asthma have been already suggested, due to genes found associated with both melanoma and asthma, as the *TYRP1* gene.[38,39] Suggestive evidence of associations of *OCA2* genetic variants with allergic sensitization, eczema and lung function phenotypes (FEV1/FVC ratio) have been also reported [40,41] Moreover, a polymorphism, located within an intron of the *HERC2* gene an

adjacent neighbor of *OCA2* which is involved in the regulation of OCA2 expression, was significantly associated with asthma response to diisiocyanate.[42]

The *OCA2* gene (Oculocutaneous Albinism II) codes for a melanosomal transmembrane protein which is involved in many biological processes as tyrosine transport. The tyrosine is a precursor to melanin synthesis. Interestingly, tyrosine is related to various pathological mechanisms of eczema.[43] *OCA2* is also involved in melanocyte differentiation, melanin biosynthesis and pigmentation. Melanocytes, along with keratinocytes and Langerhans cells, being positioned within the epidermis, form a physical skin barrier. A link between pigmentation, which depends on melanin synthesis, and the skin barrier function was shown, both in humans[44] and in mice[45] indicating that melanocytes producing melanin influence epidermal barrier function. In addition, accumulating evidence has shown that melanocytes are also active factors in the skin immune system and participate in immune responses.[46],[47]

The five SNPs detected in *OCA2* map to a large number of enhancer and promoter histone marks and to DNase I hypersensitivity sites in numerous lung and skin cells. Moreover, rs4778192 belongs to the binding sites of the transcription factor CEBPB, an important transcription factor involved in the regulation and expression of genes involved in immune and inflammatory responses.[48] The SNP rs2703978 belongs to the binding site of hoxa5 which probably activates antagonist genes against the release of keratinocyte growth factors and epidermal formation.[49] Lastly, rs8035720 a proxy of rs2594897, belongs to the binding site of the transcription factor CTCF, which controls MHC class II gene expression. It was recently shown that CTCF is a major driver of gene co-expression in the airways of asthmatic patients.[50] The two SNPs rs2594897 and rs2703978 were found associated in the

whole blood with expression of *HERC2*, this gene being like *OCA2*, also found associated to melanoma.[51]

Among the loci showing suggestive evidence of association to asthma-plus-eczema, the SNP rs10937762 is an intronic variant located in the *TBC1D14* gene (TBC1 Domain Family. Member 14). A variant of *TBC1D14* was found to be associated with eosinophil count, an important biomarker of allergy. The SNP rs10937762 in *TBC1D14* maps to a large number of enhancer and promoter histone marks and to DNase I hypersensitivity sites in numerous lung and skin cells. This SNP belongs to the binding site of the redox-sensitive nuclear factor (NF)-kappaB which is an important participant in a broad spectrum of inflammatory networks that regulate cytokine activity in airway pathology[52]. In addition, rs10013696 a proxy of rs10937762 was strongly associated with *TBC1D14* expression in skin[25].

Lastly, the SNP rs1402470 which showed suggestive evidence of association to asthma-plus-eczema, is an intronic variant located in the *LRP1B* gene (Low Density Lipoprotein Receptor-Related Protein 1B). Suggestive evidence of association was shown between *LRP1B* genetic variants and post bronchodilator FEV1[39], Diisocyanate-induced asthma[40] and childhood-onset asthma[53]. Besides that, a *LRP1B* mutation was shown to be a predictive marker for the presence of chronic obstruction pulmonary disease in patients with lung adenocarcinoma.[54]

In conclusion, the present study highlights that studying well-defined sub-phenotypic entities as the co-morbidity is a critical feature for the identification of new genes. Among the new three loci detected here for the co-morbidity of “asthma-plus-eczema” , *OCA2* is emerging as the most relevant candidate gene given the reached significance level of association, the high consistency across studies and its links to skin and lung diseases, and to epithelial barrier mechanisms and/or immune response, which have crucial roles in both asthma and eczema. Further confirmations of these findings as well as functional studies are

needed to bring greater insight into the role of these loci on the co-morbidity of asthma-plus-eczema. Deciphering molecular determinants of this co-morbidity could point to novel therapeutic approaches.

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; 67:1018-22.
2. Schoettler N, Rodríguez E ,Weidinger S, Ober C. Advances in asthma and allergic disease genetics: Is bigger always better? J Allergy Clin Immunol 2019; 144:1495-1506
3. Marenholz I, Bauerfeind A, Esparza-Gordillo J, Kerscher T, Granell R, Nickel R, et al. The eczema risk variant on chromosome 11q13 (rs7927894) in the population-based ALSPAC cohort: a novel susceptibility factor for asthma and hay fever. Hum Mol Genet 2011; 20:2443-9.
4. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, Imboden M, et al. A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization and their interaction with birth order. J Allergy Clin Immunol 2011; 128:996-1005.
5. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma, Nature 2007;448(7152):470-3.
6. Wu, H, Romieu I, Sienra-Monge JJ, del Rio-Navarro BE, London SJ. Genetic variation in ORM1-like 3 (ORMDL3) and gasdermin-like (GSDML) and childhood asthma. Allergy 2009; 64(4):629-35
7. Marenholz I, Rivera VA, Esparza-Gordillo J, Bauerfeind A, Lee-Kirsch MA, Ciechanowicz A, 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; 131:1644-9.
8. 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; 34:340-5.
9. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet 2011; 44:187-92.
10. 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.
11. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet 2017; 49:1752-7.
12. Zhu Z, Lee PH, Chaffin MD, Chung W, Loh PR, Lu Q, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat Genet 2018; 50:857-64.
13. Johansson A, Rask-Andersen M, Karlsson T, Ek WE. Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. Hum Mol Genet 2019.
14. Dizier MH, Margaritte-Jeannin P, Madore AM, Moffatt M, Brossard M, Lavielle N, et al. The nuclear factor I/A (NFIA) gene is associated with the asthma plus rhinitis phenotype. J Allergy Clin Immunol 2014; 134:576-82 e1.
15. Ferreira MA, Matheson MC, Tang CS, Granell R, Ang W, Hui J, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol 2014; 133:1564-71.
16. Sarnowski C, Laprise C, Malerba G, Moffatt MF, Dizier MH, Morin A, et al. DNA methylation within melatonin receptor 1A (MTNR1A) mediates paternally transmitted genetic variant effect on asthma plus rhinitis. J Allergy Clin Immunol 2016; 138:748-53.
17. Marenholz I, Esparza-Gordillo J, Ruschendorf F, Bauerfeind A, Strachan DP, Spycher BD, et al. Meta-analysis identifies seven susceptibility loci involved in the atopic march. Nat Commun 2015; 6:8804.

18. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010; 363:1211-21.
19. Morris AP, Lindgren CM, Zeggini E, Timpson NJ, Frayling TM, Hattersley AT, et al. A powerful approach to sub-phenotype analysis in population-based genetic association studies. Genet Epidemiol 2010; 34:335-43.
20. Ferrari A. A note on sum and difference of correlated chi-squared variables. arXiv preprint arXiv:1906.09982.
21. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 2005; 95:221-7.
22. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21:1539-58.
23. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557-60.
24. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010; 1:112-25.
25. Gibson G. Human genetics. GTEx detects genetic effects. Science 2015; 348:640-1.
26. Claringbould A, Vosa U, Esko T, Franke L, Consortium e. Trans-eQTL analysis in 25,000 individuals reveals clear differences between diseases in the types and number of causally involved biological pathways. European Journal of Human Genetics 2018; 26:108-9.
27. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet 2017; 49:139-45.
28. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 2011; 7:e1002003.
29. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One 2010; 5:e10693.
30. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012; 40:D930-4.
31. Hinds DA, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet 2013; 45:907-11.
32. Halapi E, Gudbjartsson DF, Jonsdottir GM, Bjornsdottir US, Thorleifsson G, Helgadottir H, et al. A sequence variant on 17q21 is associated with age at onset and severity of asthma. Eur J Hum Genet 2010; 18:902-8.
33. Weidinger S, O'Sullivan M, Illig T, Baurecht H, Depner M, Rodriguez E, et al. Filaggrin mutations, atopic eczema, hay fever, and asthma in children. J Allergy Clin Immunol 2008; 121:1203-9 e1.
34. Duffy DL, Zhao ZZ, Sturm RA, Hayward NK, Martin NG, Montgomery GW. Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. J Invest Dermatol 2010; 130:520-8.
35. Law MH, Bishop DT, Lee JE, Brossard M, Martin NG, Moses EK et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma, Nat Genet 2015;47(9):987-995.
36. Wei L, Allain DC, Bernhardt MN, Gillespie JL, Peters SB, Iwenofu OH, et al. Variants at the OCA2/HERC2 locus affect time to first cutaneous squamous cell carcinoma in solid organ transplant recipients collected using two different study designs. Br J Dermatol 2017; 177:1066-73.

37. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 2007; 39:1443-52.
38. Landi MT, Bishop DT, MacGregor S, Machiela MJ, Stratigos AJ, Ghiorzo P. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. Nat Genet 2020; 52(5):494-504.
39. Ding L, Abebe T, Beyene J, Wilke RA, Goldberg A, Woo JG Orcid profile, Martin LJ, Rothenberg ME, Rao M, Hershey GK, Chakraborty R, Mersha TB. Rank-based genome-wide analysis reveals the association of ryanodine receptor-2 gene variants with childhood asthma among human populations. Hum Genomics 2013; 7(1): 1-16.
40. Jang H, Kim M, Hong JY, Cho HJ, Kim CH, Kim YH, et al. Mitochondrial and Nuclear Mitochondrial Variants in Allergic Diseases. Allergy Asthma Immunol Res 2020; 12:877-84.
41. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald ML, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. BMC Genet 2015; 16:138.
42. Yucesoy B, Kaufman KM, Lummus ZL, Weirauch MT, Zhang G, Cartier A, et al. Genome-Wide Association Study Identifies Novel Loci Associated With Diisocyanate-Induced Occupational Asthma. Toxicol Sci 2015; 146:192-201.
43. Kurita M, Yoshihara Y, Ishiuji Y, Chihara M, Ishiji T, Asahina A, et al. Expression of T-cell immunoglobulin and immunoreceptor tyrosine-based inhibitory motif domain on CD4(+) T cells in patients with atopic dermatitis. J Dermatol 2019; 46:37-42.
44. Gunathilake R, Schurer NY, Shoo BA, Celli A, Hachem JP, Crumrine D, et al. pH-regulated mechanisms account for pigment-type differences in epidermal barrier function. J Invest Dermatol 2009; 129:1719-29.
45. Man MQ, Lin TK, Santiago JL, Celli A, Zhong L, Huang ZM, et al. Basis for enhanced barrier function of pigmented skin. J Invest Dermatol 2014; 134:2399-407.
46. Hong Y, Song B, Chen HD, Gao XH. Melanocytes and Skin Immunity. J Investig Dermatol Symp Proc 2015; 17:37-9.
47. Sil P, Wong SW, Martinez J. More Than Skin Deep: Autophagy Is Vital for Skin Barrier Function. Front Immunol 2018; 9:1376.
48. Kinoshita S, Akira S, Kishimoto T. A member of the C/EBP family, NF-IL6 beta, forms a heterodimer and transcriptionally synergizes with NF-IL6. Proc Natl Acad Sci U S A 1992; 89:1473-6.
49. Liang Y, Xia L, Du Z, Sheng L, Chen H, Chen G, et al. HOXA5 inhibits keratinocytes growth and epidermal formation in organotypic cultures in vitro and in vivo. J Dermatol Sci 2012; 66:197-206.
50. Pascoe CD, Obeidat M, Arsenault BA, Nie Y, Warner S, Stefanowicz D, et al. Gene expression analysis in asthma using a targeted multiplex array. BMC Pulm Med 2017; 17:189.
51. Amos C, Wang Li-E, Lee E, Gershenwald JE, Chen WV, Fang S, Kosoy R et al, Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. Hum Mol Genet . 2011; 20:5012-23
52. Schuliga M. NF-kappaB Signaling in Chronic Inflammatory Airway Disease. Biomolecules 2015; 5:1266-83.
53. Almoguera B, Vazquez L, Mentch F, Connolly J, Pacheco JA, Sundaresan AS, et al. Identification of Four Novel Loci in Asthma in European American and African American Populations. Am J Respir Crit Care Med 2017; 195:456-63.
54. Xiao D, Li F, Pan H, Liang H, Wu K, He J. Integrative analysis of genomic sequencing data reveals higher prevalence of LRP1B mutations in lung adenocarcinoma patients with COPD. Sci Rep 2017; 7:2121.

Table I: Phenotypic description of the EGEA, SLSJ, GABRIELA and ALSPAC samples

	EGEA	**SLSJ**	**GABRIELA**	**ALSPAC**
N	783	614	1649	5761
Age, yrs, mean (sd)	16.46 (0.29)	23.75 (0.62)	9.00 (0.04)	15.46 (0.33)
Gender-Men (%)	410(52.4)	282(45.9)	943(57.2)	2869 (49.8)
Asthma AND Eczema (%)	207 (26)	144 (23)	269 (16)	588 (10)
Neither Asthma NOR eczema (%)	260 (33)	163 (27)	713 (43)	3297 (57)
Asthma only OR eczema only (%)	316 (40)	307 (50)	667 (40)	1876 (33)

Table II: meta-analysis results of SNPs showing significant (or suggestive) evidence of association specifically to the co-morbidity "asthma-plus-eczema"

									Co-morbidity association test [c]			Phenotypic homogeneity test [d]			Joint test [e]
SNP	chr	Effect/baseline allele	Effect Allele Freq [a]	position (kb)[b]	SNP location	gene	Previous gene	Next gene	β_{fixed}	se(β_{fixed})	P	β_{fixed}	se(β_{fixed})	P	P
rs1402470	2	G/A	0.556	141 362	intron	*LRP1B*	*NXPH2*	*KYNU*	-0.250	0.055	4.90×10^{-6}	-0.230	0.056	4.41×10^{-5}	4.52×10^{-7}
rs10937762	4	G/A	0.495	6 917	intron	*TBC1D14*	*KIAA0232*	*CCDC96*	-0.244	0.054	6.10×10^{-6}	-0.222	0.053	3.29×10^{-5}	4.32×10^{-7}
rs2311469	15	T/G	0.444	27 824	intron	*OCA2*	*GABRG3*	*HERC2*	0.231	0.052	7.41×10^{-6}	0.221	0.054	3.84×10^{-5}	5.41×10^{-7}
rs4778189	15	G/A	0.556	27 827	intron	*OCA2*	*GABRG3*	*HERC2*	-0.232	0.052	6.63×10^{-6}	-0.224	0.054	3.24×10^{-5}	4.51×10^{-7}
rs4778192	**15**	**T/C**	**0.556**	**27 835**	**intron**	***OCA2***	***GABRG3***	***HERC2***	**-0.236**	**0.052**	$\mathbf{4.63 \times 10^{-6}}$	**-0.229**	**0.054**	$\mathbf{2.26 \times 10^{-5}}$	$\mathbf{2.84 \times 10^{-7}}$
rs2594897	15	T/C	0.278	27 900	intron	*OCA2*	*GABRG3*	*HERC2*	0.269	0.056	1.70×10^{-6}	0.225	0.058	9.58×10^{-5}	3.75×10^{-7}
rs2703978	**15**	**T/C**	**0.278**	**27 902**	**intron**	***OCA2***	***GABRG3***	***HERC2***	**0.270**	**0.056**	$\mathbf{1.60 \times 10^{-6}}$	**0.229**	**0.058**	$\mathbf{7.14 \times 10^{-5}}$	$\mathbf{2.98 \times 10^{-7}}$

a : estimated in CEU population from Phase 3 of the 1000 Genomes Project

b : SNP position in kilo base pairs (GRCh38.p12 :Genome Reference Consortium Human Build 38 patch release 12)

c: Meta-analysis of genome-wide association test of "asthma-plus-eczema" vs "no asthma, no eczema"

d: Meta-analysis of homogeneity test of association between "asthma-plus-eczema" vs "asthma only or eczema only"

e: Test combining statistics of the association and phenotypic homogeneity tests and following a gamma distribution with parameters depending on the correlation between statistics of the two tests

In bold: SNPs detected with significant evidence of association specifically to "asthma plus eczema", i.e detected at P≤10-4 for ' co-morbidity association test' AND for 'phenotypic homogeneity test' AND P≤ 3.5×10^{-7} for the joint test

The other SNPs were detected with suggestive evidence of association specifically to "asthma plus eczema", i.e detected at P≤10^{-4} for ' co-morbidity association test' AND for 'phenotypic homogeneity test' AND P≤ 1×10^{-6} for the joint test

Legends of figures:

Figure 1:

Whole analysis strategy for the detection of SNPs showing significant (and suggestive) evidence of association specifically with the co-morbidity asthma-plus-eczema. Note that * $Z_{fixed} = \beta_{fixed} / se(\beta_{fixed})$, ** Γ: gamma distribution with parameters depending on the correlation between statistics of test1 and test2.

Figure 2:

Manhattan plot of meta-analysis results for the joint test. The x axis represents chromosomal location and the y axis represents $-\log10P_{joint}$; the first horizontal line denotes $P=3.5x10^{-7}$, corresponding to the significance threshold, the second horizontal line denotes $P=10^{-6}$, corresponding to the suggestive threshold used for the joint test.

Figure 3:

Regional plots for the three regions of interest (imputed SNPs around 400kb of best genotyped SNPs). In regional plots (a-f), the x axis presents physical distance in megabases (build 37.3 coordinates) and the y axis presents $-\log10\ P_{joint}$ values for the joint test statistic.

SUPPLEMENTARY MATERIAL

Identification of OCA2 as a novel locus for the comorbidity of asthma-plus-eczema

Running title: OCA2 a novel locus for asthma-plus-eczema

P Margaritte-Jeannin (1), A Budu-Aggrey (2), M Ege (3), AM Madore (4), C Linhard (1), H Mohamdi (1), E von Mutius (3), R Granell (2), F Demenais (1), C Laprise (4), E Bouzigon (1), MH Dizier (1)

(1) Université de Paris, UMRS 1124, INSERM, Paris, France

(2) Medical Research Council (MRC) Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

(3) Dr von Hauner Children's Hospital, Ludwig Maximilian University; Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich, Germany,

(4) Département des sciences fondamentales, Centre intersectoriel en santé durable (CISD), Université du Québec à Chicoutimi, Saguenay, QC, Canada

Corresponding author :

Marie-Hélène Dizier, Université de Paris, UMRS 1124, INSERM, Campus Saint-Germain-des-Prés

45 rue des Saints Pères - 75006 Paris, France.

Email: marie-helene.dizier@inserm.fr

Populations

Participating studies

EGEA

The EGEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy) (https://egeanet.vjf.inserm.fr/index.php/en/) is a French longitudinal survey which has been described in detail previously[1]. A first survey took place between 1991 and 1995 and a 12-year follow-up was conducted between 2003 and 2007. The EGEA family sample consisted of 388 French nuclear families, 253 families ascertained through offspring with asthma and 135 families ascertained through one parent with asthma. Asthma status used for analysis was defined by the following criteria (present either at the time of data collection and/or at the 12-year follow-up): a positive answer to at least one of the two questions: 1/Have you ever had attacks of breathlessness at rest with wheezing?, 2/ Have you ever had an attack of asthma? associated with either the presence of bronchial-hyper responsiveness, hospitalization for asthma in life or asthma therapy. 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 12-year follow-up). Ethical approval was obtained from the relevant institutional review board committees (Cochin Port-Royal Hospital and Necker-Enfants Malades Hospital, Paris: n° 01-07-07, 04-05-03, 04-11-13 and 04-11-18). Written informed consent was signed by all participants. Written informed consent was signed by kin or guardians of the minors/children.

SLSJ

The Saguenay–Lac-Saint-Jean Familial Asthma Cohort(SLSJ) comprised 254 French-Canadian multigenerational families ascertained through one proband with asthma[2]. Definition of asthma used for analysis was similar to the one used in the EGEA study. Eczema was defined by a self-reported history of eczema on a questionnaire, based by a positive answer to at least one of the two following questions: 1) Do you have eczema? or 2) Have you ever had eczema? Positive answers were validated by physicians or through clinical records. The SLSJ study was approved by the ethic committees of the academic Integrated Health and Social

Services Centres of Saguenay-Lac-Saint-Jean (CIUSSSS) and of UQAC.

In the present study, we focused on offspring of the families in EGEA and SLSJ datasets, in order to limit the heterogeneity across studies according to the age.

ALSPAC

The ALSPAC (Avon Longitudinal Study of Parents and Children; http://www.bristol.ac.uk/alspac/) study is a population-based longitudinal prospective birth-cohort study which includes 14,062 children born in 1991 and 1992 in Avon, United Kingdom.Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study[3,4]. The initial number of pregnancies enrolled is 14,541 (for these at least one questionnaire has been returned or a "Children in Focus" clinic had been attended by 19/07/99). Of these initial pregnancies, there were a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age.

When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above. The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 24 is 913 (456, 262 and 195 recruited during Phases II, III and IV respectively), resulting in an additional 913 children being enrolled. . The total sample size for analyses using any data collected after the age of seven is therefore 15,454 pregnancies, resulting in 15,589 foetuses. Of these 14,901 were alive at 1 year of age.

A 10% sample of the ALSPAC cohort, known as the Children in Focus (CiF) group, attended clinics at the University of Bristol at various time intervals between 4 to 61 months of age. The CiF group were chosen at random from the last 6 months of ALSPAC births (1432 families attended at least one

clinic). Excluded were those mothers who had moved out of the area or were lost to follow-up, and those partaking in another study of infant development in Avon.

Eczema cases included those who responded yes to any of the 'Dr diagnosed eczema ever' variables at age 10 or 13 as well as those who had ever had eczema from ages 7 to 16 years. These were derived from questionnaires filled by mothers indicating if their child had had eczema in the past year at 81, 91, and 103 months and also at 10 and 13 years, as well as questionnaires filled by the young person indicating if they had had eczema in the past 12 months at 16 years. Asthma cases were defined by those responding yes to any of the 'Dr diagnosed asthma ever' variables at age 7.5 or 11 or 14 or 15 years. Controls were identified if they responded no to all the variables mentioned.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool" and reference the following webpage: http://www.bristol.ac.uk/alspac/researchers/our-data/.

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. A list of the ethics committee/institutional review board(s) that approved aspects of the study are available at http://www.bristol.ac.uk/alspac/researchers/researchethics/.

GABRIELA

The (GABRIEL Advanced Survey) are cross-sectional population-based surveys conducted in rural areas of Austria, Germany, and Switzerland. In total, 132,366 children aged 6 to 13 years were addressed through schools. In a first stage in fall/winter 2006, asthma, allergic disease, and contact to farming environments were assessed using a short parental questionnaire (n=79,888). In a second stage in spring/summer 2007, 9,668 children were selected among families consenting in writing to blood sampling, genetic testing and collection of environmental samples by stratified random sampling to ensure representation of children with high exposure to farming environments. Asthma was defined as a parental report of a doctor’s diagnosis ever of asthma or a reported diagnosis of obstructive bronchitis at least twice. Eczema was defined as a doctor’s diagnosis ever of eczema or

an itchy rash that was stronger or weaker for at least 6 months during the last 12 months as reported by the parents. Children without a doctor's diagnosis ever of asthma, obstructive bronchitis, eczema, and without itchy rash were selected as controls. The GABRIELA study was approved by the institutional review boards of the Bavarian Medical Association (for Bavaria), Ulm University (for Baden-Württemberg), the cantons Lucerne, Zurich and Thurgau (for Switzerland), Medical University of Innsbruck (for Austria), and Medical University of Wroclaw (for Poland) and informed consent was obtained from the parents

References

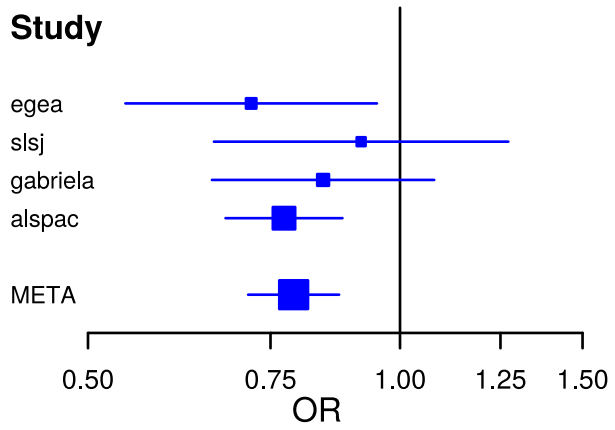
1. Kauffmann F, Dizier MH, Pin I, Paty E, Gormand F, Vervloet D, et al. Epidemiological study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy: phenotype issues. Am J Respir Crit Care Med 1997; 156:S123-9.
2. Laprise C. The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population. Genes Immun 2014; 15:247-55.
3 Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: The 'Children of the 90s'; the index offspring of The Avon Longitudinal Study of Parents and Children (ALSPAC). International Journal of Epidemiology 2013; 42: 111-127.
4 Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International Journal of Epidemiology 2013; 42:97-110.

Figure S2

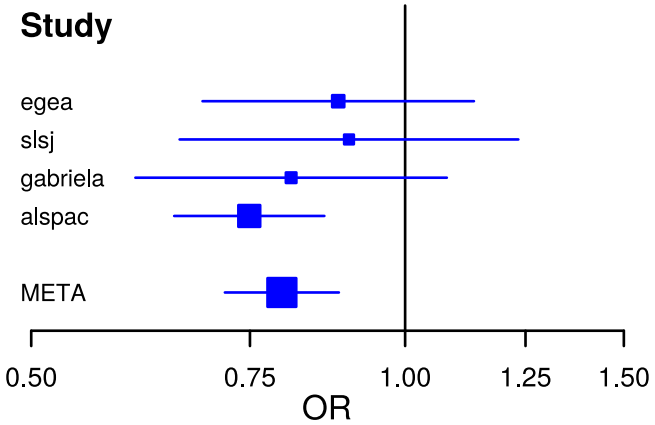
Forest Plots:

rs47778192 (OCA2)

Co-morbidity association Test

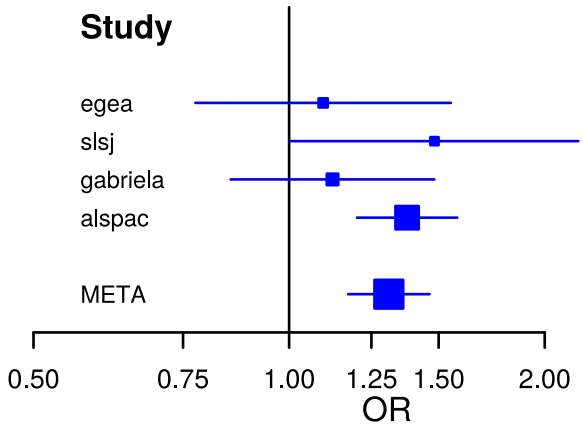


Phenotypic Homogeneity Test



rs2703978 (OCA2)

Co-morbidity association Test



Phenotypic Homogeneity Test

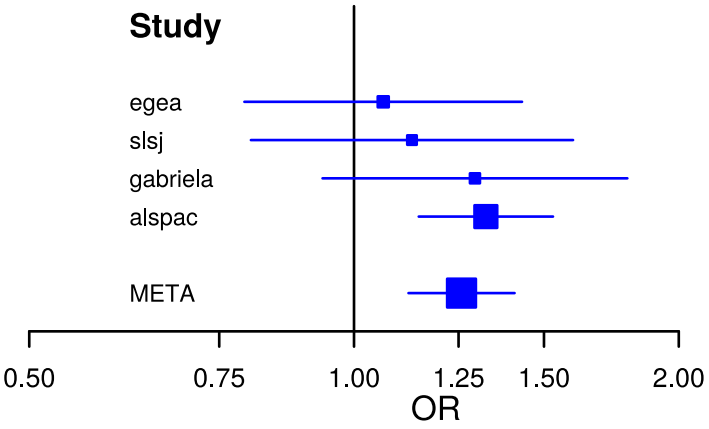


Figure 2:

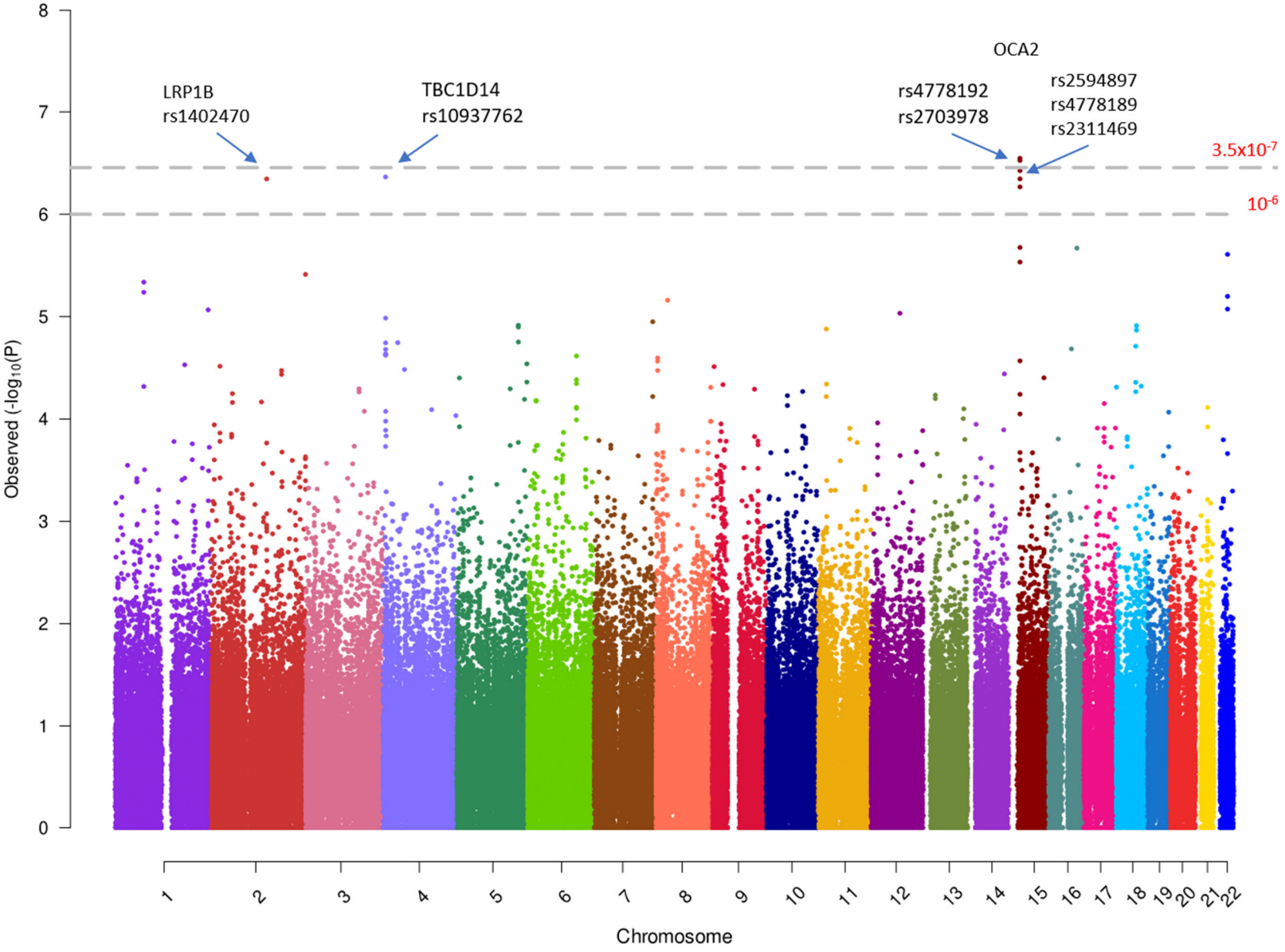
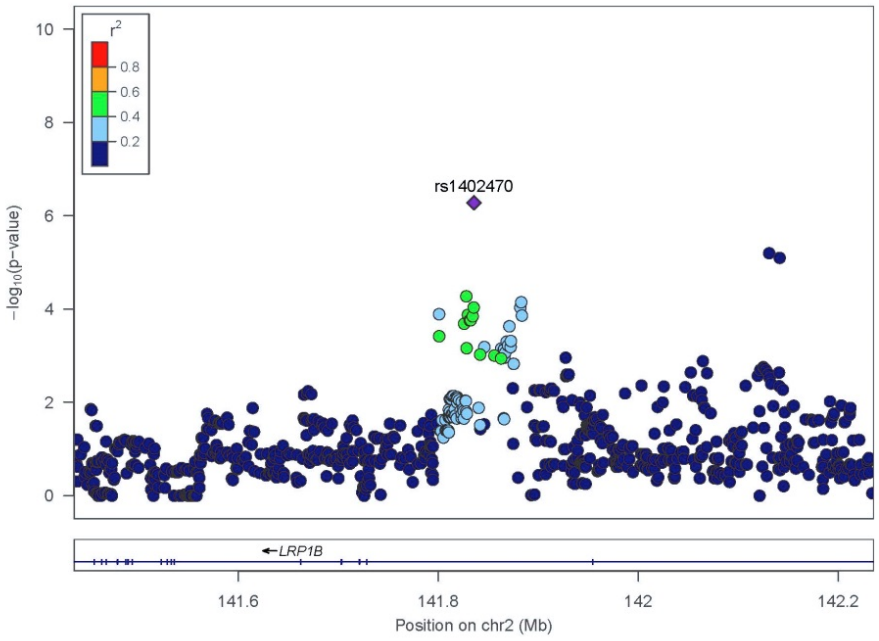
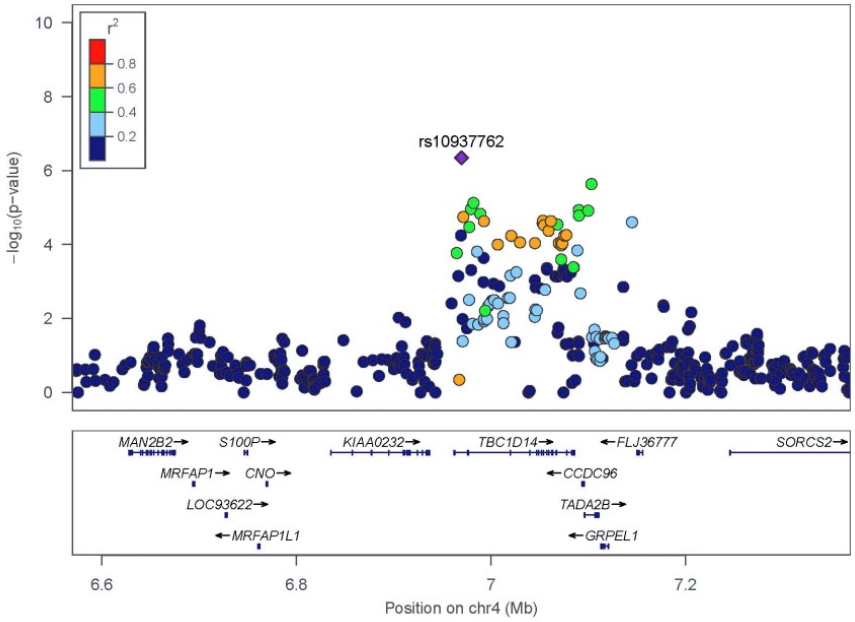

LRP1B
rs1402470
TBC1D14
rs10937762
OCA2
rs4778192
rs2703978
rs2594897
rs4778189
rs2311469
3.5x10-7
10-6
Observed (-log10(P))
Chromosome
0
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

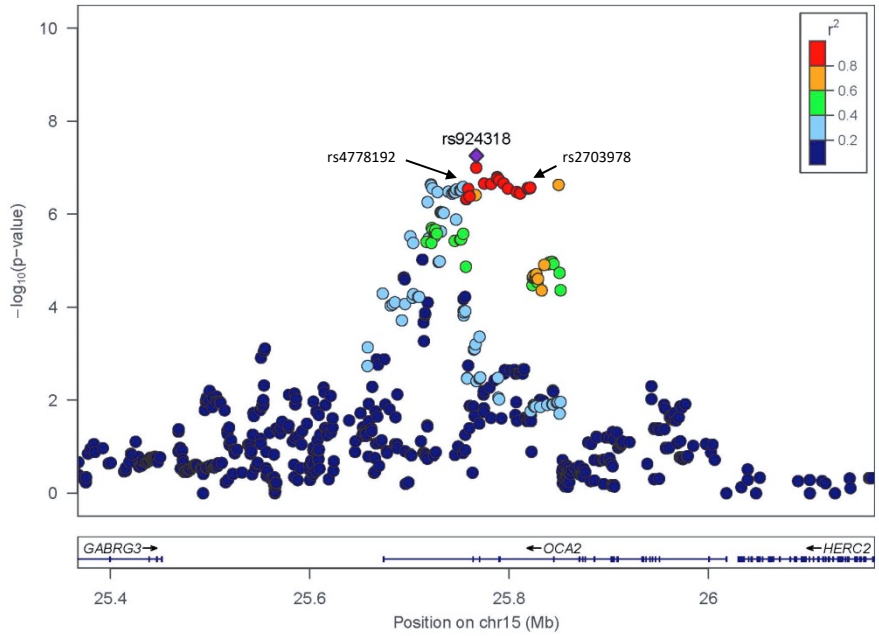
Figure 3



Region around the genotyped SNP rs1402470 located in *LRP1B*



Region around the genotyped SNP rs10937762 located in *TBC1D14*

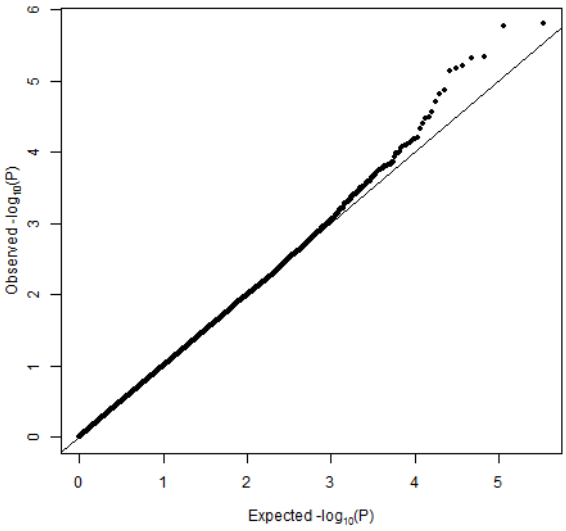


Region around the genotyped SNPs rs4778192 and rs2703978 located in *OCA2*

Figure S1:

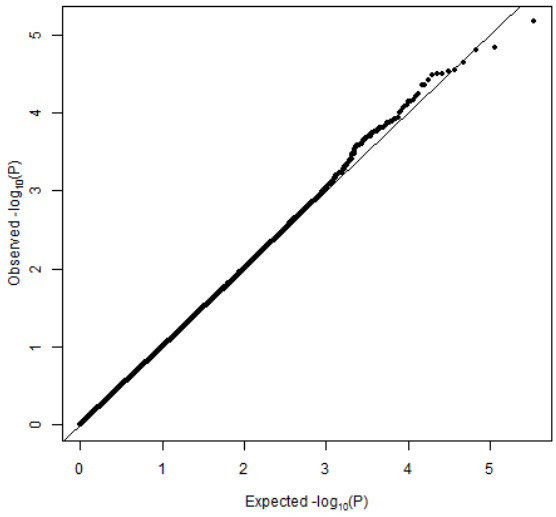
QQ plot of the test statistics of co-morbidity association test, phenotypic homogeneity test and joint test

Co-morbidity association test



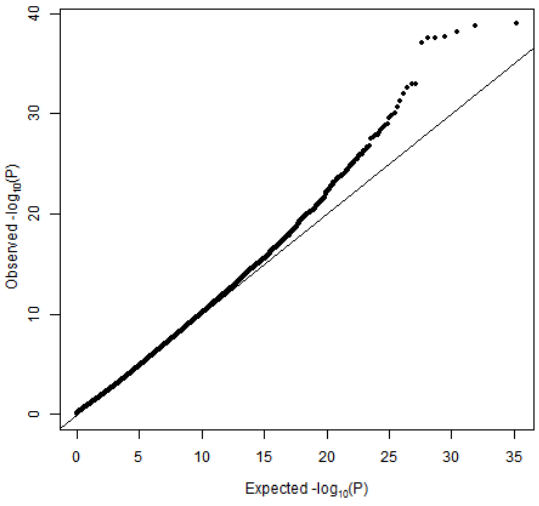
Lambda=1.02

Phenotypic homogeneity test



Lambda=0.99

Joint Test

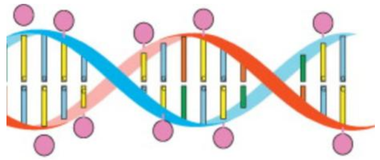


Lambda=1.01

Identification of *OCA2* as a novel locus for the co-morbidity of asthma-plus-eczema

1/ Genome-wide association study (GWAS)

Genetic variability (whole genome)



Phenotypic variability

No asthma, no eczema

Asthma-plus-eczema

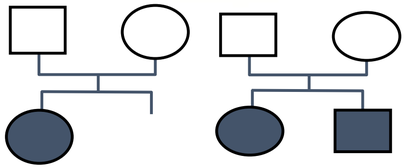
Asthma only

or eczema only

2/ Meta-analyses of GWAS among population and family based samples

Alspac

Gabriela



Egea

Slsj

OCA2

skin and lung diseases

epithelial barrier

immunology

CLINICAL & EXPERIMENTAL ALLERGY TRUSTED EVIDENCE IN ALLERGY

Table S1: Meta-analysis results for SNPS detected at P≤10^{-4} for "co-morbidity association test" AND/OR for "phenotypic homogeneity test" AND/OR P≤ 10^{-6} for the joint test

								Association test [c]				Phenotypic homogeneity test [d]				Joint test [e]
chr	snp	Effect/Baseline alleles	Effect Allele Freq [a]	position (pb) b[b]	gene	previous gene	next gene	I2 (%)	β_{fixed}	se (β_{fixed})	P	I2 (%)	β_{fixed}	se (β_{fixed})	P	P
1	rs7522428	T/C	0.763	74 712 676	CRYZ	ERICH3	TYW3	3.96	-0.243	0.059	3.36E-05	11.73	-0.218	0.061	3.43E-04	5.79E-06
1	rs10890142	G/A	0.237	74 713 445	CRYZ	ERICH3	TYW3	0.91	0.245	0.058	2.79E-05	0	0.221	0.061	2.90E-04	4.61E-06
1	rs12089980	T/C	0.086	152 459 455	.	CRNN	LCE5A	62.83	0.381	0.080	1.98E-06	73.76	0.310	0.083	1.88E-04	6.33E-07
1	rs2504053	G/A	0.657	179 759 991	FAM163A	TDRD5	TOR1AIP2	0	0.168	0.054	1.73E-03	0	0.212	0.054	8.69E-05	2.96E-05
1	rs12036109	G/A	0.556	239 902 578	CHRM3	LOC105373226	FMN2	0	-0.183	0.051	3.44E-04	0	-0.212	0.053	6.21E-05	8.60E-06
2	rs1402470	G/A	0.556	141 361 833	LRP1B	NXPH2	KYNU	23.49	-0.250	0.055	4.90E-06	19.68	-0.230	0.056	4.41E-05	4.52E-07
2	rs2196598	G/A	0.828	141 656 962	LRP1B	NXPH2	KYNU	0	-0.289	0.067	1.72E-05	29.37	-0.234	0.070	8.12E-04	6.48E-06
2	rs7355583	T/C	0.313	240 882 290	.	AGXT	C2orf54	0	-0.223	0.054	3.97E-05	0	-0.212	0.056	1.54E-04	3.87E-06
3	rs7651606	T/C	0.500	21 935 979	ZNF385D	SGOL1	UBE2E2-AS1	46.48	-0.134	0.053	1.10E-02	0	-0.209	0.054	9.87E-05	1.01E-04
3	rs1259380	C/A	0.692	105 271 040	.	ZPLD1	ALCAM	14.11	-0.228	0.057	5.66E-05	74.75	-0.172	0.058	3.00E-03	3.17E-05
3	rs333318	T/G	0.409	124 615 752	KALRN	ROPN1	UMPS	16.68	-0.209	0.053	7.88E-05	0	-0.112	0.054	3.82E-02	1.85E-04
3	rs1463230	T/C	0.470	150 316 769	.	LOC105374313	TSC22D2	0	-0.205	0.052	8.83E-05	0	-0.140	0.054	9.14E-03	8.43E-05
4	rs10937762	G/A	0.495	6 916 952	TBC1D14	KIAA0232	CCDC96	0	-0.244	0.054	6.10E-06	7.02	-0.222	0.053	3.29E-05	4.32E-07
4	rs870660	T/C	0.485	7 050 455	TADA2B	CCDC96	GRPEL1	40.74	0.214	0.053	5.69E-05	0	0.227	0.054	2.61E-05	1.56E-06
4	rs1381521	T/G	0.232	187 273 606	.	FAT1	ZFP42	1.78	0.123	0.058	3.44E-02	0	0.252	0.060	3.02E-05	9.29E-05
5	rs6872996	T/C	0.298	6 666 989	SRD5A1	NSUN2	PAPD7	0	-0.164	0.056	3.41E-03	0	-0.230	0.058	7.11E-05	3.98E-05
5	rs2160505	C/A	0.550	157 932 760	.	CLINT1	EBF1	0	-0.219	0.053	3.20E-05	0	-0.174	0.054	1.17E-03	1.21E-05
6	rs1832937	G/A	0.394	125 664 788	.	HDDC2	HEY2	0	0.217	0.055	7.24E-05	0	0.141	0.055	9.75E-03	7.71E-05
6	rs1777220	T/G	0.404	125 701 456	.	HDDC2	HEY2	0	0.209	0.053	8.49E-05	0	0.174	0.054	1.29E-03	2.42E-05
6	rs6941174	G/A	0.535	148 451 334	SASH1	SAMD5	UST	0	0.215	0.053	4.48E-05	56.49	0.109	0.054	4.38E-02	1.39E-04
7	rs10486733	T/G	0.450	43 233 525	HECW1	MRPL32	STK17A	0	0.080	0.053	1.34E-01	0.40	0.225	0.054	2.88E-05	1.94E-04
7	rs6977081	T/G	0.288	150 845 427	AOC1	TMEM176A	KCNH2	20.88	0.197	0.056	3.98E-04	0	0.220	0.056	8.13E-05	1.12E-05
8	rs10503644	G/A	0.636	17 839 867	.	MTUS1	FGL1	0	0.216	0.054	6.27E-05	0	0.100	0.056	7.56E-02	2.36E-04
8	rs3850744	G/A	0.621	17 926 694	PCM1	FGL1	ASAH1	0	0.211	0.053	8.12E-05	11.03	0.086	0.055	1.15E-01	3.53E-04

8	rs11775871	T/C	0.177	67 820 941	.	CPA6	PREX2	0	-0.266	0.070	1.59E-04	0	-0.348	0.071	1.07E-06	3.83E-07
8	rs909055	G/A	0.702	140 361 494	TRAPPC9	KCNK9	CHRAC1	0	0.128	0.056	2.35E-02	0	0.253	0.059	1.61E-05	4.91E-05
8	rs7834518	T/C	0.303	140 376 212	TRAPPC9	KCNK9	CHRAC1	0	-0.117	0.057	3.94E-02	0	-0.243	0.058	3.23E-05	1.05E-04
8	rs13265944	T/C	0.298	140 401 489	TRAPPC9	KCNK9	CHRAC1	0	-0.125	0.057	2.84E-02	0	-0.241	0.059	4.38E-05	1.06E-04
9	rs10811586	T/G	0.641	21 701 943	.	IFNE	MTAP	0	0.127	0.054	1.87E-02	0	0.216	0.054	7.00E-05	1.12E-04
9	rs1330148	G/A	0.535	107 786 143	.	KLF4	ACTL7B	14.26	-0.207	0.053	8.01E-05	0	-0.151	0.053	4.52E-03	5.12E-05
10	rs11004028	G/A	0.732	54 067 053	PCDH15	MBL2	MTRNR2L5	0	0.162	0.061	8.20E-03	0	0.253	0.063	5.67E-05	5.92E-05
10	rs17644321	T/C	0.732	54 089 217	PCDH15	MBL2	MTRNR2L5	0	0.161	0.061	7.84E-03	0	0.247	0.063	8.35E-05	7.40E-05
12	rs7316080	T/C	0.409	73 903 937	.	TRHDE	ATXN7L3B	0	0.213	0.052	4.66E-05	0	0.191	0.055	5.19E-04	9.28E-06
15	rs8036718	G/A	0.646	27 808 411	OCA2	GABRG3	HERC2	0	-0.228	0.053	2.01E-05	4.54	-0.205	0.055	1.99E-04	2.94E-06
15	rs6497238	T/C	0.646	27 808 632	OCA2	GABRG3	HERC2	0	-0.231	0.053	1.53E-05	0	-0.208	0.055	1.56E-04	2.11E-06
15	rs1498521	G/A	0.389	27 813 762	OCA2	GABRG3	HERC2	0	-0.244	0.053	3.77E-06	54.53	-0.200	0.055	2.47E-04	1.14E-06
15	rs8042881	G/A	0.389	27 815 483	OCA2	GABRG3	HERC2	0	0.244	0.053	3.96E-06	54.27	0.200	0.055	2.42E-04	1.17E-06
15	rs2311469	T/G	0.444	27 823 922	OCA2	GABRG3	HERC2	0	0.231	0.052	7.41E-06	0	0.221	0.054	3.84E-05	5.41E-07
15	rs4778189	G/A	0.556	27 827 082	OCA2	GABRG3	HERC2	0	-0.232	0.052	6.63E-06	0	-0.224	0.054	3.24E-05	4.51E-07
15	rs4778192	T/C	0.556	27 834 915	OCA2	GABRG3	HERC2	0	-0.236	0.052	4.63E-06	0	-0.229	0.054	2.26E-05	2.84E-07
15	rs2594897	T/C	0.278	27 899 904	OCA3	GABRG3	HERC2	0	0.269	0.056	1.70E-06	0	0.225	0.058	9.58E-05	3.75E-07
15	rs2703978	T/C	0.278	27 902 060	OCA4	GABRG3	HERC2	2.70	0.270	0.056	1.60E-06	0	0.229	0.058	7.14E-05	2.98E-07
15	rs2594909	T/C	0.621	27 905 916	OCA4	GABRG3	HERC2	0	-0.212	0.053	6.54E-05	0	-0.171	0.055	2.01E-03	2.71E-05
16	rs12929493	G/A	0.394	57 965 616	CNGB1	KIFC3	TEPP	0	-0.234	0.054	1.37E-05	0	-0.153	0.056	6.64E-03	2.07E-05
16	rs811047	G/A	0.429	72 226 794	.	PMFBP1	ZFHX3	0	-0.192	0.054	3.62E-04	0	-0.254	0.056	6.86E-06	2.14E-06
17	rs907092	G/A	0.520	39 766 006	IKZF3	GRB7	ZPBP2	0	0.212	0.053	6.63E-05	0	0.080	0.055	1.42E-01	3.46E-04
19	rs7246355	T/C	0.530	50 722 987	.	SHANK1	CLEC11A	31.13	-0.071	0.052	1.74E-01	0	-0.220	0.055	6.36E-05	3.74E-04
22	rs2009168	G/A	0.374	35 721 807	APOL5	APOL6	RBFOX2	0	-0.211	0.057	2.09E-04	0	-0.256	0.059	1.47E-05	2.47E-06
22	rs2016586	T/G	0.323	35 729 217	APOL5	APOL6	RBFOX2	0	-0.233	0.059	7.57E-05	0	-0.230	0.061	1.75E-04	6.35E-06

a : estimated in CEU population from Phase 3 of the 1000 Genomes Project

b: SNP position in base pairs (GRCh38.p12)

c : Meta-analysis of genome-wide association test of “asthma-plus-eczema” vs “no asthma, no eczema

d : Meta-analysis of homogeneity test of association between “asthma-plus-eczema” vs “asthma only or eczema only”

e : Test combining statistics of the co-morbidity association and phenotypic homogeneity tests and following a gamma distribution with parameters depending on the correlation between statistics of the two tests

in blue: SNPS detected with significant evidence of assocation specically to asthma plus eczema, i.e detected at $P \leq 10^{-4}$ for " co-morbidity association test" (with I2≤24%) AND for "phenotypic heterogenety test" (with I2≤24%) AND $P \leq 3.5 \times 10^{-7}$ for the joint test

in yellow : SNPS detected with suggestive evidence of assocation specically to asthma plus eczema, i.e detected at $P \leq 10^{-4}$ for " co-morbidity association test" (with I2≤24%) AND for "phenotypic heterogenety test" (with I2≤24%) AND $P \leq 10^{-6}$ for the joint test

Table S2: results per study of SNPs showing significant (or suggestive) evidence of association specifically to the co-morbidity "asthma-plus-eczema" in the meta-analysis of the four datasets

	CO-MORBIDITY ASSOCIATION TEST														
	EGEA			SLSJ			GABRIELA			ALSPAC			Meta-analysis		
SNP	β	se(β)	P	β	se(β)	P	β	se(β)	P	β	se(β)	P	β_{fixed}	se(β_{fixed})	P
rs1402470	-0.472	0.135	4.46E-04	-0.333	0.175	5.67E-02	-0.163	0.135	2.27E-01	-0.196	0.072	6.59E-03	-0.250	0.055	4.90E-06
rs10937762	-0.324	0.139	1.93E-02	-0.223	0.179	2.12E-01	-0.376	0.145	9.62E-03	-0.198	0.068	3.84E-03	-0.244	0.054	6.10E-06
rs2311469	0.314	0.143	2.81E-02	0.063	0.166	7.07E-01	0.171	0.126	1.75E-01	0.257	0.066	1.05E-04	0.231	0.052	7.41E-06
rs4778189	-0.314	0.143	2.81E-02	-0.075	0.166	6.49E-01	-0.171	0.126	1.75E-01	-0.257	0.066	1.05E-04	-0.232	0.052	6.63E-06
rs4778192	-0.331	0.143	2.04E-02	-0.086	0.167	6.05E-01	-0.171	0.126	1.75E-01	-0.258	0.066	1.01E-04	-0.236	0.052	4.63E-06
rs2594897	0.099	0.176	5.73E-01	0.378	0.195	5.25E-02	0.121	0.140	3.89E-01	0.318	0.069	4.68E-06	0.269	0.056	1.70E-06
rs2703978	0.092	0.177	6.02E-01	0.394	0.199	4.77E-02	0.118	0.141	4.04E-01	0.320	0.069	4.20E-06	0.270	0.056	1.60E-06

	PHENOTYPIC HOMOGENEITY TEST														
	EGEA			SLSJ			GABRIELA			ALSPAC			Meta-analysis		
SNP	β	se(β)	P	β	se(β)	P	β	se(β)	P	β	se(β)	P	β_{fixed}	se(β_{fixed})	P
rs1402470	-0.396	0.125	1.46E-03	-0.205	0.163	2.10E-01	-0.009	0.159	9.52E-01	-0.225	0.076	3.20E-03	-0.230	0.056	4.41E-05
rs10937762	-0.106	0.126	3.99E-01	-0.235	0.153	1.26E-01	-0.448	0.148	2.55E-03	-0.204	0.071	3.95E-03	-0.222	0.053	3.29E-05
rs2311469	0.104	0.128	4.18E-01	0.074	0.160	6.43E-01	0.213	0.147	1.48E-01	0.289	0.071	4.54E-05	0.221	0.054	3.84E-05
rs4778189	-0.104	0.128	4.18E-01	-0.082	0.159	6.04E-01	-0.213	0.147	1.48E-01	-0.291	0.071	4.17E-05	-0.224	0.054	3.24E-05
rs4778192	-0.124	0.128	3.34E-01	-0.104	0.160	5.16E-01	-0.211	0.147	1.51E-01	-0.289	0.071	4.72E-05	-0.229	0.054	2.26E-05
rs2594897	0.060	0.151	6.90E-01	0.096	0.176	5.86E-01	0.256	0.166	1.23E-01	0.281	0.073	1.25E-04	0.225	0.058	9.58E-05
rs2703978	0.062	0.151	6.80E-01	0.124	0.175	4.81E-01	0.258	0.166	1.20E-01	0.281	0.073	1.20E-04	0.229	0.058	7.14E-05

Table S3 : functional annotation of SNPs showing significant or suggestive evidence of association specifically with the "asthma-plus-eczema" comorbidity

Chr	SNP	Position in bp (GRCh38.p12)	Gene	Variant Location	Promoter histones marks	Enhancer histones marks	Dnase I hyper sensitive sites	Transcription factor binding sites
2	rs1402470	142 119 402	LRP1B	intronic variant				Dobox4, Mef2, Pax-4
4	rs10937762	6 916 952	TBC1D14	intronic variant	Lung, Lung Carcinoma Cell Line	Lung, Lung Carcinoma Cell Line, fetal lung fibroblasts Cell Line	Foreskin Melanocyte Primary Cells	NF-kappaB
					Foreskin Fibroblast and Melanocyte Primary Cells	Epidermal Keratinocyte, Foreskin Fibroblast, Keratinocyte and Melanocyte Primary Cells		
15	rs2311469		OCA2	intronic variant				cart1, Pou1f1, Pou2f2, TATA
15	rs4778192	28 080 061	OCA2	intronic variant	lung	Foreskin Melanocyte Primary Cells		AP1, CEBPB, lrf, NFY, Pbx3, RFX5, SP2
15	rs4778189	28 072 228	OCA2	intronic variant	Foreskin Fibroblast Primary Cells, Fetal lung, Adult Dermal Fibroblast Primary Cells	Foreskin Fibroblast Primary Cells, Fetal lung, Adult Dermal Fibroblast Primary Cells, Epidermal Keratinocyte Primary Cells, Lung Fibroblast Primary cells	Foreskin Fibroblast and Melanocyte Primary Cells, Fetal lung, Adult Dermal Fibroblast Primary Cells, Lung Fibroblast Primary cells	BATF,Irf,THAP1
15	rs2594897	28 145 050	OCA2	intronic variant		Foreskin Keratinocyte Primary Cells		VDR
15	rs2703978	28 147 206	OCA2	intronic variant		Foreskin Melanocyte Primary Cells		Arid3a, Cphx, Hoxa5, Hoxd8, Lhx3, pax4, pax6, Pou3f2, Pouf3f4, Pou6f1