A novel role for cilia function in atopy: ADGRV1 and DNAH5 interactions

Pierre-Emmanuel Sugier, Myriam Brossard, Chloé Sarnowski, Amaury Vaysse, Andréanne Morin, Lucile Pain, Patricia Margaritte-Jeannin, Marie-Hélène Dizier, William O. C. M. Cookson, Mark Lathrop, et al.

HAL Id: inserm-01616494

https://inserm.hal.science/inserm-01616494

Submitted on 13 Oct 2017

1 **A novel role for cilia function in atopy: *ADGRV1* and *DNAH5* interactions**

2

3 Pierre-Emmanuel Sugier[a,b], MSc, Myriam Brossard[a]*, PhD, Chloé Sarnowski[a]*, PhD, Amaury

4 Vaysse[a], PhD, Andréanne Morin[c,d], MSc, Lucile Pain[d], MSc, Patricia Margaritte-Jeannin[a], PhD,

5 Marie-Hélène Dizier[a], PhD, William O.C.M. Cookson, MD, D.Phil[e]., Mark Lathrop[c], PhD, Miriam

6 F. Moffatt, D.Phil.[e], Catherine Laprise[d]‡, PhD, Florence Demenais[a]‡, MD, Emmanuelle

7 Bouzigon[a]‡, MD, PhD

8

9 *These authors shared second authorship.

10 ‡These authors shared senior authorship.

11 [a]Genetic Variation and Human Diseases Unit, INSERM, Université Paris Diderot, Université

12 Sorbonne Paris Cité, Paris, France

13 [b]Université Pierre et Marie Curie, Paris, France

14 [c]McGill University and Génome Québec Innovation Centre, Montréal, Canada

15 [d]Département des Sciences Fondamentales, Université du Québec à Chicoutimi, Chicoutimi, QC,

16 Canada

17 [e]Section of Genomic Medicine, National Heart and Lung Institute, London, UK

18 **Correspondance:**

19 Florence Demenais, MD

20 UMR-946, INSERM, Université Paris-Diderot

21 27 rue Juliette Dodu

22 75010 Paris

23 Tel: 33 (0) 1 53 72 50 02

24   Fax: 33 (0) 1 53 72 50 49

25   Email: florence.demenais@inserm.fr

26   **Disclosure of potential conflict of interest:** all authors declare no conflict of interest

27

36

37   Manuscript count: 3809

38   Abstract count: 248

39

ABSTRACT

Background: Atopy, an endotype underlying allergic diseases, has a substantial genetic component.

Objective: Our goal was to identify novel genes associated with atopy in asthma-ascertained families.

Methods: We implemented a three-step analysis strategy in three datasets: The Epidemiological study on the Genetics and Environment of Asthma (EGEA) dataset: 1,660 subjects; The Saguenay-Lac-Saint-Jean (SLSJ) dataset: 1,138 subjects; and The Medical Research Council (MRC) dataset: 446 subjects). This strategy included a single-SNP genome-wide association study (GWAS), the selection of related gene pairs based on statistical filtering of GWAS results and text-mining filtering using GRAIL and SNP-SNP interaction analysis of selected gene pairs.

Results: We identified the 5q14 locus, harboring the adhesion G protein-coupled receptor V1 (*ADGRV1*) gene, that showed genome-wide significant association with atopy (rs4916831; $P_{meta}=6.8 \times 10^{-9}$). Statistical filtering of GWAS results followed by text-mining filtering revealed relationships between *ADGRV1* and three genes showing suggestive association with atopy ($P \leq 10^{-4}$). SNP-SNP interaction analysis between *ADGRV1* and these three genes showed significant interaction between *ADGRV1* rs17554723 and two correlated SNPs (rs2134256 and rs1354187) within dynein axonemal heavy chain 5 (*DNAH5*) gene ($P_{meta-int}=3.6 \times 10^{-5}$ and 6.1×10^{-5}, that met the multiple-testing corrected threshold of 7.3×10^{-5}). Further conditional analysis indicated that rs2134256 alone accounted for the interaction signal with rs17554723.

Conclusion: As both DNAH5 and ADGRV1 contribute to function of cilia, this study suggests that cilia dysfunction may represent a novel mechanism underlying atopy. Combining GWAS and epistasis analysis driven by statistical and knowledge-based evidence represents a promising approach for identifying new genes involved in complex traits.

Key Messages:

- *ADGRV1* genetic variants are associated with atopy in asthma families

- Interaction between *ADGRV1* and *DNAH5* variants is associated with atopy; these two genes are involved in ciliary function

- Use of a strategy that combines genome-wide association analysis and epistasis analysis driven by statistical and knowledge-based evidence can successfully identify new genes underlying complex traits.

Capsule summary:

This study in three family-based studies identified association between *ADGRV1* and atopy and interaction between *ADGRV1* and *DNAH5*, two genes that contribute to ciliary functions.

Key words: atopy, asthma, genetics, genome-wide association study, gene-gene interaction, text-mining, *ADGRV1*, *DNAH*5, ciliary function

Abbreviations:

ADGRV1: adhesion G protein-coupled receptor V1

DNAH5: dynein, axonemal, heavy chain 5

SNP: single nucleotide polymorphism

GWAS: genome-wide association study

SPT: skin prick test

EGEA: Epidemiological study on the Genetics and Environment of Asthma

SLSJ: Saguenay-Lac-Saint-Jean study

MRCA: Medical Research Council funded collection of nuclear families with Asthma

MRCE: Medical Research Council funded collection of nuclear families with Eczema

86   GRAIL: Gene Relationships Across Implicated Loci

87   QC: quality control

88   MAF: minor allele frequency

89   PCs: principal components

90   LD: linkage disequilibrium

91   ORs: odds-ratios

92   GTEx: Genotype-Tissue Expression

93   ETS: environmental tobacco smoke

94   CI: confidence interval

95

INTRODUCTION

Allergies and asthma are among the most common diseases in industrialized countries. Although environmental factors play an important role in allergic diseases, estimates of heritability of allergy, which range between 25% and 80%, suggest significant genetic contribution.[1] Genome-wide association studies (GWAS) have identified a number of loci associated with allergic diseases (i.e., asthma, atopic dermatitis, rhinitis),[2,3] but these loci only explain a small part of the genetic risk. Part of the difficulty encountered in identifying the genetic factors involved in these allergic diseases is due to the heterogeneity of these diseases and the uncertainty of diagnosis. However, this problem can be alleviated by the study of an endotype underlying allergic diseases, such as allergic sensitization or atopy.

Atopy is characterized by the production of allergen-specific immunoglobulin E against environmental allergens. Estimates of heritability of atopy range from 40% to 85%.[4,5] Many candidate genetic studies of atopy have been conducted but have often led to inconsistent results.[6] While the first GWAS of allergic sensitization only reported a few loci,[7-10] two recent large-scale meta-analyses of allergic sensitization[11] and self-reported allergy[12] increased the number of associated loci to 10 and 16 loci, respectively. However, other loci may influence atopy as it is well known that GWAS alone cannot reveal the whole genetic landscape underlying complex phenotypes.

Heterogeneity across studies, which may be caused by variability in the genetic background of the populations, environmental exposures, or study design, may be a limitation of meta-analyses of GWAS for identifying new loci associated with a trait. Notably, the importance of data sampling was recently highlighted by a positional cloning study of eczema, where association with *ANO3/MUC15* genetic variants was only found in family samples ascertained through asthmatic

119   subjects but neither in families ascertained through eczema patients nor in a case/control study of

120   eczema.[13]

121   Another limitation of GWAS is that they typically focus on the analysis of individual single

122   nucleotide polymorphisms (SNPs) and are underpowered to detect genetic factors which have a

123   small marginal effect but rather interact with each other. Gene-gene interaction analysis (or

124   epistasis analysis) has the ability to reveal novel genes involved in complex traits but raises an

125   enormous multiple-testing problem when performed at the genome-wide level. Statistical and

126   biological filtering pipelines can be used to limit the search for SNP-SNP interactions.[14] Following

127   the "guilt-by-association" assumption which states connected genes are usually participating in the

128   same or related cellular functions,[15] search for interactions can be restricted to genes pointed out

129   by a preliminary GWAS (e.g., interactions of genes harboring significant association signals with

130   genes harboring suggestive associations) and showing relationships based on prior knowledge. One

131   knowledge-based approach that can be particularly useful to prioritize genes for epistasis analysis

132   is text-mining of the literature as it can highlight relationships between genes[16] according to their

133   co-occurrence with the same words in scientific articles.

134

135   The objective of this study was to identify novel genetic factors influencing atopy by combining a

136   GWAS and epistasis analysis driven by statistical and knowledge-based evidence in three family

137   samples ascertained through asthmatic subjects: the French Epidemiological study on the Genetics

138   and Environment of Asthma (EGEA; 1,660 subjects), the French-Canadian Saguenay-Lac-Saint-

139   Jean study (SLSJ; 1,138 subjects) and the Medical Research Council UK study (MRC; 446

140   subjects). Our overall analysis strategy included three main steps: (1) a genome-wide single-SNP

141   association analysis, (2) the selection of related gene pairs based on statistical filtering from GWAS

7

142 results and text-mining filtering using the Gene Relationships Across Implicated Loci (GRAIL)

143 approach,[17] and (3) a SNP-SNP interaction analysis for the selected gene pairs.

144

145 **MATERIALS AND METHODS**

146 *Study datasets and definition of atopy*

147 The EGEA study combines a case-control and a family-based study of asthma. The whole study

148 population includes 388 families ascertained through at least one asthmatic proband recruited in

149 chest clinics (1,705 probands and first-degree relatives) plus 415 population-based controls (total

150 of 2,120 subjects). All subjects were born in France and were of European ancestry. The protocol

151 of this study has been described elsewhere.[18-20] Atopy was assessed by skin prick tests (SPT)

152 performed in 1,978 subjects. A positive SPT response was defined as a wheal diameter $\geq$ 3mm to

153 at least one of 11 aeroallergens belonging to three groups (indoor allergens, outdoor allergens,

154 molds). After quality control (QC) of genotypic data, 925 atopic and 735 non-atopic subjects were

155 included in the analysis.

156 The Saguenay-Lac-Saint-Jean and Quebec City Familial Asthma Collection (SLSJ) consists of a

157 French-Canadian founder population panel of 253 multigenerational families from Saguenay-Lac-

158 Saint-Jean region, ascertained through two asthmatic probands.[21] This study has been described

159 elsewhere.[21] Skin tests were done in 1,195 SLSJ subjects and atopy was defined similarly as in

160 EGEA. After QC of genotypic data, the analysis dataset included 641 atopic and 497 non-atopic

161 subjects.

162 The Medical Research Council (MRC) UK study includes 207 nuclear families, recruited through

163 at least one proband with childhood-onset asthma (MRCA sample). The study protocol has been

164 described elsewhere.[22] Atopy was defined similarly as in EGEA. To increase the number of

165 unaffected subjects (controls), we included subjects from another MRC-UK dataset that were

8

166   recruited through probands with eczema (MRCE sample). Only subjects without asthma, without

167   eczema and with low IgE levels were used as controls in this study. We checked that the age and

168   gender distributions were similar in MRCA and MRCE samples. After QC of genotypic data, the

169   analysis sample included 106 atopic and 340 non-atopic subjects. The whole UK sample will be

170   subsequently designated as the MRC sample.

171   Protocols of EGEA, SLSJ and MRC studies have been approved by the local ethical committees.

172   All adult participants and child's legal guardians provided written informed consent.

173

174   *Genotyping*

175   Both EGEA and SLSJ datasets were genotyped using the Illumina 610-Quad array (Illumina, San

176   Diego, CA), as part as of the Gabriel asthma consortium GWAS.[23] Stringent quality criteria were

177   applied to select both individuals and SNPs and have been previously detailed.[23,24] After QC, there

178   was a final set of 501,167 autosomal SNPs for analysis. The offspring in MRCA families and

179   MRCE controls were genotyped using the Illumina Sentrix HumanHap300 BeadChip (307,981

180   autosomal SNPs), as part of the first asthma GWAS.[22,23] QC for MRC samples has been detailed

181   elsewhere.[22,25] In order to get a number of SNPs in MRC sample as large as in EGEA and SLSJ

182   samples, SNP imputation was performed using MACH v1.00 software[26] and HapMap2 release 21

183   CEU haplotypes as reference panel. Imputed SNPs were kept for analysis if their imputation quality

184   score (rsq)[27] was ≥ 0.5 and minor allele frequency was $\geq 5\%$.

185

186   *Descriptive statistics and strategy of analysis*

187   Descriptive statistics of atopy together with sex, age and asthma status were assessed in each

188   dataset using Stata® V14.1 (distributed by Stata Corporation, College Station, Texas, USA). The

189   workflow of our three-step analysis strategy is summarized in Figure 1 and presented in the

190   following paragraphs.

191

192   ***Genome-wide single-SNP analysis***

193   We performed a two-stage GWAS. In the first stage, association analysis between individual SNPs

194   and atopy was carried out in the EGEA dataset. This analysis was based on a logistic regression

195   model assuming an additive model for SNP effect, using Stata® V14.1. This model was adjusted

196   for significant effects of age and sex and two principal components (PCs) to account for population

197   structure. We took into account familial dependencies using the cluster and robust options of the

198   logit function in Stata®. Test of SNP effect was based on a Wald-test. In a second stage, the SNPs

199   reaching $P{\leq}10^{-4}$ in EGEA were followed–up in SLSJ and MRC. The association analysis in SLSJ

200   and in MRC used the same model as in EGEA. The results of stage 2 datasets and, then, of the

201   three datasets were combined using a fixed-effects meta-analysis. SNPs were declared significantly

202   associated with atopy if the three datasets meta-analysis P-value (P_{meta}) reached the genome-wide

203   significance level of $1.5{\times}10^{-7}$. This threshold was obtained by dividing the type I error of 5% by

204   the effective number of independent SNPs in the Illumina 610-Quad array.[28]

205

206   ***Selection of gene pairs using both statistical and text-mining filters***

207   The statistical filtering consisted of selecting two sets of genes using the GWAS results: genes

208   showing significant association with atopy (set-1) and genes showing suggestive association with

209   atopy (set-2). The set-1 included all genes lying at a distance of 50 kb or less from SNPs reaching

210   the genome-wide significance level in the GWAS meta-analysis. The set-2 included all genes that

211   were at most 50 kb apart from SNPs having $P{\leq}10^{-4}$ in the stage 1 EGEA dataset and were not part

212   of set-1. To assign SNPs to genes, we used NCBI dbSNP Build 137 and human Genome Build

213   37.3. We further filtered gene pairs (formed by crossing set-1 genes with set-2 genes) through

214   GRAIL[17] text-mining of PubMed abstracts (available in October 2014). For each gene, GRAIL

215   builds a vector of words where the elements of this vector are weights that take values between 0

216   and 1 depending on how often a word is found with a gene in an abstract. Then, GRAIL computes

217   pairwise similarity between genes from gene/word vectors and ranks the similarities between each

218   gene from set-1 and all genes of the genome. The P_{GRAIL} of a gene from set-2 with a gene from set-

219   1 is equal to the proportion of all genes that have similarity with the set-1 gene greater than the

220   similarity between set-2 and set-1 genes (i.e. rank divided by total number of genes across the

221   genome). We used the threshold of $P_{GRAIL} \leq 0.10$, as recommended,[17] to select related gene pairs

222   for further epistasis analysis.

223

224   *SNP-SNP interaction analysis for selected gene pairs*

225   As for the single-SNP association analysis, we performed a two-stage SNP-SNP interaction

226   analysis. At stage 1, we analyzed all SNP-SNP interactions for the GRAIL-selected gene pairs in

227   the EGEA dataset. For each gene, we considered all SNPs lying within gene boundaries. Pairwise

228   SNP-SNP interactions were evaluated by logistic regression assuming an additive model for SNP

229   main effects and interaction and adjusting for the same covariates (age, sex, PCs) as in the GWAS,

230   using Stata® V14.1. We used the same coding scheme as usually proposed for SNP-SNP

231   interaction modelling.[29] We modeled the additive effect of a SNP by coding the genotypes of

232   homozygotes for the minor allele, heterozygotes and homozygotes for the major allele as 1, 0, and

233   -1; the interaction term between two SNPs was obtained by multiplication of these genotypic values

234   for the two SNPs. Test of interaction was based on a likelihood-ratio test which follows a Chi-

235   square distribution with one degree of freedom. We discarded all SNP pairs for which one or more

236   of the nine genotypic combinations appeared in fewer than five subjects (cases or controls). In a

237   second stage, all SNP pairs showing suggestive evidence for interaction in EGEA ($P_{int} \leq 5 \times 10^{-3}$)

238   were followed-up in SLSJ and MRC. The results of the stage 2 datasets and, then, of the three

239   datasets were meta-analyzed using a fixed-effects model.

240   To correct for multiple testing, we computed, for each gene pair investigated, the effective number

241   of independent interaction tests from the eigenvalues of the correlation matrix of products of SNP

242   variables, using an extension of Li and Ji's method.[30] The corrected threshold to declare an

243   interaction statistically significant was equal to the 5% type I error divided by the sum of effective

244   number of independent interaction tests over all gene pairs tested.

245

246   **Stratified analyses according to asthma status**

247   Because family samples were ascertained through asthmatic probands, we investigated whether

248   SNP associations and SNP-SNP interactions detected with atopy might be related to the presence

249   of asthma. Single-SNP and SNP-SNP interaction analyses were repeated in the two groups of

250   asthmatic and non-asthmatic subjects separately. These analyses were performed for the SNPs that

251   showed significant results in the meta-analyses of the three datasets. Homogeneity of the odds-

252   ratios (ORs) between the two groups was tested using the Cochran's Q statistic.[31]

253

254

255   **RESULTS**

256   *Descriptive statistics*

257   A total of 1,660 EGEA, 1,138 SLSJ and 446 MRC subjects were included in this study. The

258   proportion of atopic subjects was similar in EGEA and SLSJ (55.7% and 56.3% respectively) but

259   was lower in MRC (23.8%; $P \leq 10^{-3}$). In each study, there was a higher proportion of males in atopic

12

260 than in non-atopic subjects and atopic subjects were younger than non-atopic subjects (see Table

261 E1 in the Online Repository). As expected, the proportion of asthmatic subjects was higher in

262 atopic than in non-atopic subjects in all datasets (Table E1). In EGEA (respectively in SLSJ and

263 MRC), 78.0% (75.0% and 78.3%) of atopic subjects had positive SPT to indoor allergens, 55.5%

264 (77.5% and 52.8%) to outdoor allergens, and 34.8% (14.8% and 12.3%) to molds.

265

266 *Genome-wide single-SNP analysis*

267 In the stage 1 EGEA dataset, no SNP reached the genome-wide significance level of 1.5×10^{-7} (see

268 quantile-quantile (QQ) plot and Manhattan plot in Figures E1 and E2). However, 73 SNPs lying in

269 47 loci showed associations with atopy exceeding the screening threshold of $P \leq 10^{-4}$. These SNPs

270 were followed-up in the stage 2 SLSJ and MRC datasets and meta-analyzed (Table E2). The SNP

271 rs4916831 within *ADGRV1* gene at 5q14 reached the genome-wide significance level

272 ($P_{meta}=6.8 \times 10^{-9}$) in the overall meta-analysis of the three datasets (Table I). Four other SNPs at that

273 locus, in moderate linkage disequilibrium (LD) with rs4916831 (r^2 ranging between 0.51 and 0.79),

274 showed suggestive association ($4.3 \times 10^{-7} \leq P_{meta} \leq 3.8 \times 10^{-6}$; Table I).

275

276 *Selection of gene pairs using both statistical and text-mining filters*

277 The gene set-1 included *ADGRV1*, the only gene significantly associated with atopy. There were

278 30 genes that lied fewer than 50 kb apart from the 65 SNPs at 46 loci having $P \leq 10^{-4}$ in EGEA (after

279 exclusion of *ADGRV1* SNPs) and formed gene set-2 (Table E3). When GRAIL was applied to 30

280 gene pairs (date accessed: 04/24/2015), formed by each of these 30 genes with *ADGRV1*, three

281 genes were related with *ADGRV1* at $P_{GRAIL}<0.10$: *DNAH5* on 5p15 ($P_{GRAIL}=0.084$), *CHD7* on 8q12

282 ($P_{GRAIL}=3.2 \times 10^{-3}$) and *ATP8B1* on 18q21 ($P_{GRAIL}=0.016$).

283

SNP-SNP interaction analysis for selected gene pairs

In the stage 1 EGEA dataset, the three GRAIL-selected gene-pairs (*ADGRV1/DNAH5, ADGRV1/CHD7, ADGRV1/ATP8B1)* were each examined for SNP-SNP interactions, making a total of 5,324 SNP pairs. There were 37 SNP pairs that reached $P_{int} \leq 5x10^{-3}$ in EGEA and were followed-up in SLSJ and MRC at stage 2. Two of these SNPs pairs, which are related to the *ADGRV1* and *DNAH5* gene pair, met the multiple-testing corrected threshold, estimated to be $7.3x10^{-5}$ (see Table E4), in the meta-analysis of the three datasets (Table II). The two significant interactions involved the same SNP rs17554723 within *ADGRV1* and two SNPs within *DNAH5*, rs2134256 ($P_{meta-int}=3.6x10^{-5}$) and rs1354187 ($P_{meta-int}=6.1x10^{-5}$), that are in moderate LD ($r^2=0.50$; D'=0.95). However, further conditional regression analysis in each of the strata defined by genotypes at *ADGRV1* rs17554723 showed that *DNAH5* rs1354187 was no longer significantly associated with atopy ($P \geq 0.15$) when conditioning on *DNAH5* rs2134256. The most significant SNP pair shows a pattern of interaction in which the ORs for atopy associated with TT (or CC) genotype at *DNAH5* rs2134256 are in opposite direction according to the genotype, AA (or GG), at *ADGRV1* rs17554723 (Figure 2). This pattern was consistent in all three datasets (Figure 2).

Stratified analyses according to asthma status

Association analyses of atopy with the genome-wide significant *ADGRV1* rs4916831 SNP in asthmatic and non-asthmatic subjects did not show any relationship with presence of asthma in the stage 1 and stage 2 datasets and meta-analysis of the three datasets ($P_{Cochran}$ for test of homogeneity between the two groups ≥ 0.82; Table E5A). In the meta-analysis, the evidence for association was even stronger although not significantly in non-asthmatics (P=7.8x10^{-6}) than in asthmatics (P=1.4x10^{-4}). Similarly, interaction analyses for *ADGRV1* and *DNAH5* SNPs did not show any relationship with asthma ($P_{Cochran} \geq 0.30$; Table E5B). The evidence for interaction was only

308 significant in non-asthmatic subjects (Table E6B); this can be at least partly explained by the larger

309 sample size of non-asthmatic (N=1,849) than asthmatic subjects (N=1,354).

310

311 **Functional annotations of SNPs showing significant results**

312 All SNPs that show significant association (or interaction) with atopy are intronic. The two

313 *ADGRV1* SNPs, rs4916831 and rs17554723 on 5q14, detected through GWAS and interaction

314 analysis, lie 120 kb apart in introns 83 (rs4916831) and 70 (rs17554723) and are in low LD

315 (r^2=0.20, D'=0.75). The two *DNAH5* SNPs (rs2134256 and rs1354187) at 5p15.2 are located in

316 introns 58 and 60 (8 kb apart) but only rs2134256 accounts for the interaction signal (see above).

317 By interrogating the Genotype-Tissue Expression (GTEx) database,[32] rs4916831 was found

318 associated with *ADGRV1* expression in esophagus mucosa (P=7.5x10^{-7}).[32] We also investigated

319 whether the *ADGRV1* and *DNAH5* SNPs (as well as their proxies, r^2≥0.80) map to functionally

320 important regulatory regions using HaploRegV4.[33] As shown in Table E6, these SNPs and/or

321 proxies map to binding sites of various transcription factors (TFs). In addition, four proxies of

322 *ADGRV1 rs4916831* map to enhancer histone marks in lung and skin while a proxy of *DNAH5*

323 rs2134256 maps to promoter and enhancer marks in hematopoietic stem cells.

324

325 **DISCUSSION**

326 By combining genome-wide single-SNP analysis and epistasis analysis driven by statistical and

327 knowledge-based evidence in three asthma-ascertained family datasets, we identified significant

328 association of atopy at a novel 5q14 locus harboring *ADGRV1* gene and significant interaction

329 between *ADGRV1* and *DNAH5* genetic variants.

330 The interaction between *ADGRV1* and *DNAH5* variants has biological relevance as these two genes

331 are both involved in ciliopathies and ciliary function. Ciliopathies comprise a group of disorders

15

associated with genetic mutations encoding defective proteins, which result in either abnormal formation or function of cilia.[34] Mutations in the adhesion G protein-coupled receptor V1 (*ADGRV1*) gene cause Usher syndrome type IIC, a ciliopathy characterized by hearing loss and visual impairment,[35,36] while mutations of dynein axonemal heavy chain 5 (*DNAH5*) gene cause primary ciliary dyskinesia type 3, a ciliopathy which combines upper and lower tract respiratory manifestations, male infertility, and situs inversus.[37] The ADGRV1 protein (also called GPR98) is a component of the Usher protein network that functions in stereocilia of inner ear hair cells and photoreceptor cilia. The heavy chain of axonemal dynein, encoded by *DNAH5*, is part of a microtubule-associated motor protein complex that is responsible for cilia mobility, especially in respiratory epithelial cells where cilia motility is essential for mucus transport and airway clearance.[38] Although the respective function of ADGRV1 and DNAH5 proteins was initially described in different organs, these proteins may also have related functions. Indeed, the cilium in photoreceptors is ultrastructurally very similar to the nasal ciliated epithelium and the nasal ciliated epithelium of Usher syndrome II patients was found to have a lower ciliary beat frequency than healthy controls.[39] Moreover, Usher syndrome has been reported to be associated with bronchiectasis, sinusitis and reduced nasal mucociliary clearance.[40]

Besides the involvement of both ADGRV1 and DNAH5 proteins in cilia functions, which supports the statistical interaction found between these two genes, both *DNAH5* and *ADGRV1* genes have been previously associated with respiratory diseases and related phenotypes. Recent GWAS reported significant association of *DNAH5* variants with total lung capacity in chronic obstructive pulmonary disease[41] and suggestive association with Immunoglobuline E grass sensitization.[9] However, the SNP reported by that latter study was not in LD with the *DNAH5* SNPs interacting with *ADGRV1* SNP in this study ($r^2 < 0.13$). Based on an approach similar to ours, which combined genome-wide expression data in nasal epithelial cells, allele frequency variation between

populations and literature search to select candidate genes, nominal association of asthma with *DNAH5* was reported and stronger association was found with *KIF3*, a gene involved in transport of protein complexes within cilia and potentially in allergen clearance as *DNAH5*.[42] In addition, *DNAH5* belongs to the same gene family as *DNAH9* which showed interaction with environmental tobacco smoke (ETS) exposure for bronchial hyperresponsiveness in EGEA and SLSJ families.[43] Moreover, suggestive association of *ADGRV1* with asthma has been recently reported by a meta-analysis of GWAS.[44] Though most previously reported associations concern asthma or respiratory phenotypes, the interaction between *ADGRV1* and *DNAH5* SNPs associated with atopy in the present study appears independent of asthma, as shown by the stratified analysis on asthma. Although the mechanism by which these two genes influence atopy is still unknown, we can hypothesize that they are involved in dysfunction of cilia that move secreted mucus containing trapped foreign particles up and out of the airways, which favors allergic sensitization. This is supported by recent observations of a differential mRNA expression of both *ADGRV1* and *DNAH5* genes in sputum from House Dust Mite (HDM)-sensitized wheezing subjects as compared to non-atopic controls.[45] Furthermore, *DNAH5* as well as other genes of the same family including *DNAH9* were among the highest-ranking co-expression hubs in one of the HDM-wheezing associated gene modules, which was strongly enriched with genes involved in the function of ciliated epithelial cells.[45] All these observations suggest cilia-related genes may constitute an important emerging pathway for atopy.

The strategy used in this study, that enabled identifying novel relevant candidates for atopy, combined genome-wide single-SNP analysis and gene-gene interaction analysis based on both statistical filtering of GWAS results and text-mining filtering. It is of note that our three-step strategy was designed *a priori* and SNP-SNP interaction tests were only performed for gene pairs selected through our two filtering processes. The genome-wide single-SNP analysis pointed

380   towards one gene (*ADGRV1*) which harbored the lead SNP rs4916831 reaching genome-wide

381   significance and four other SNPs showing suggestive association. By increasing the density of

382   SNPs through Hapmap2-based imputation at that locus, an additional SNP (r^2=0.80 with

383   rs4916831) reached genome-wide significance and six other SNPs had P-values within one order

384   of magnitude of the genome-wide threshold (results not shown), which strengthens our finding.

385   Further conditional analysis in that region showed that association with atopy was only accounted

386   for by the lead genotyped SNP. The subsequent statistical and text-mining filters, used prior to

387   epistasis analysis, made it possible to detect gene-gene interaction by lowering the multiple testing

388   burden. Indeed, use of both filters reduced the number of interaction tests by 9-fold as compared

389   to using the statistical filter only. The text-mining filter was based on GRAIL that was shown to be

390   successful in pointing out true disease regions that were validated.[17] Although many sources of

391   biological information can be used to connect genes, such as co-expression gene networks or

392   protein-protein interaction networks, the advantage of GRAIL is to provide a broader framework

393   for revealing gene-gene relationships of any origin through literature search. However, GWAS and

394   candidate gene studies, which are driven by researchers' expectations, can create a bias towards

395   genes that are frequently reported in the literature. An appropriate approach would be to utilize the

396   existing knowledge and to correct for potential bias but, to our knowledge, such method does not

397   exist yet.

398

399   In conclusion, this study shows that the proposed strategy that combines GWAS and epistasis

400   analysis driven by statistical and knowledge-based evidence can successfully identify strong

401   candidate genes for complex phenotypes as atopy. The interaction between *DNAH5* and *ADGRV1,*

402   two genes involved in cilia functions, is of biological relevance and provides a novel mechanism

403   underlying atopy. Further studies, including functional and experimental studies, are needed to

404   confirm the current findings and to identify the functional variants.

405

REFERENCES

1. Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev* 2011; 242:10-30.

2. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014; 42:D1001-6.

3. Bonnelykke K, Sparks R, Waage J, Milner JD. Genetics of allergy and allergic sensitization: common variants, rare mutations. *Curr Opin Immunol* 2015; 36:115-26.

4. Los H, Postmus PE, Boomsma DI. Asthma genetics and intermediate phenotypes: a review from twin studies. *Twin Res* 2001; 4:81-93.

5. Thomsen SF, Ulrik CS, Kyvik KO, Ferreira MA, Backer V. Multivariate genetic analysis of atopy phenotypes in a selected sample of twins. *Clin Exp Allergy* 2006; 36:1382-90.

6. Vercelli D. Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol* 2008; 8:169-82.

7. Andiappan AK, Wang de Y, Anantharaman R, Parate PN, Suri BK, Low HQ, *et al*. Genome-wide association study for atopy and allergic rhinitis in a Singapore Chinese population. *PLoS One* 2011; 6:e19719.

8. Castro-Giner F, Bustamante M, Ramon Gonzalez J, Kogevinas M, Jarvis D, Heinrich J, *et al.* A pooling-based genome-wide analysis identifies new potential candidate genes for atopy in the European Community Respiratory Health Survey (ECRHS). *BMC Med Genet* 2009; 10:128.

9. 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.

10. Wan YI, Strachan DP, Evans DM, Henderson J, McKeever T, Holloway JW, *et al.* A genome-wide association study to identify genetic determinants of atopy in subjects from the United Kingdom. *J Allergy Clin Immunol* 2011; 127:223-31, 31 e1-3.

468   11.   Bonnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, *et al.* Meta-analysis of

469   genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet* 2013;

470   45:902-6.

471   12.   Hinds DA, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM, *et al.* A genome-wide association

472   meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat*

473   *Genet* 2013; 45:907-11.

474   13.   Dizier MH, Margaritte-Jeannin P, Madore AM, Esparza-Gordillo J, Moffatt M, Corda E, *et al.* The

475   ANO3/MUC15 locus is associated with eczema in families ascertained through asthma. *J Allergy Clin*

476   *Immunol* 2012; 129:1547-53 e3.

477   14.   Sun X, Lu Q, Mukherjee S, Crane PK, Elston R, Ritchie MD. Analysis pipeline for the epistasis search

478   - statistical versus biological filtering. *Front Genet* 2014; 5:106.

479   15.   Li ZC, Huang MH, Zhong WQ, Liu ZQ, Xie Y, Dai Z, *et al.* Identification of drug-target interaction

480   from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics*

481   2016; 32:1057-64.

482   16.   Luo Y, Riedlinger G, Szolovits P. Text mining in cancer gene and pathway prioritization. *Cancer*

483   *Inform* 2014; 13:69-79.

484   17.   Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Sklar P, *et al.* Identifying relationships

485   among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions.

486   *PLoS Genet* 2009; 5:e1000534.

487   18.   Kauffmann F, Dizier MH, Annesi-Maesano I, Bousquet J, Charpin D, Demenais F, *et al.* EGEA

488   (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness

489   and atopy)-- descriptive characteristics. *Clin Exp Allergy* 1999; 29 Suppl 4:17-21.

490   19.   Kaufmann F, Dizier MH, Pin I, Paty E, Gormand F, Vervloet D, *et al.* Epidemiological Study of the

491   Genetics and Environment of Asthma, Bronchial Hyperresponsiveness, and Atopy. *Am J Respir Crit*

492   *Care Med* 1997; 156:123-9.

493 20. Bouzigon E, Nadif R, Le Moual N, Dizier MH, Aschard H, Boudier A, *et al.* [Genetic and

494   environmental factors of asthma and allergy: Results of the EGEA study]. *Rev Mal Respir* 2015;

495   32:822-40.

496 21. Laprise C. The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young

497   founder population. *Genes Immun* 2014; 15:247-55.

498 22. Moffatt MF, Kabesch M, Liang L*,* Dixon AL, Strachan D, Heath S, *et al.* Genetic variants regulating

499   ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; 448:470-3.

500 23. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, *et al.* A large-scale, consortium-

501   based genomewide association study of asthma. *N Engl J Med* 2010; 363:1211-21.

502 24. Sarnowski C, Sugier PE, Granell R, Jarvis D, Dizier MH, Ege M, *et al.* Identification of a new locus at

503   16q12 associated with time to asthma onset. *J Allergy Clin Immunol* 2016.

504 25. Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF*, et al.* A cross-platform analysis

505   of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res* 2013;

506   23:716-26.

507 26. Li Y, Abecasis GR. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am. J.*

508   *Hum. Genet.* 2006; S79:2290.

509 27. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;

510   10:387-406.

511 28. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and

512   significant p-value thresholds in commercial genotyping arrays and public imputation reference

513   datasets. *Hum Genet* 2012; 131:747-56.

514 29. Herold C, Steffens M, Brockschmidt FF, Baur MP, Becker T. INTERSNP: genome-wide interaction

515   analysis guided by a priori information. *Bioinformatics* 2009; 25:3275-81.

516 30. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation

517   matrix. *Heredity* 2005; 95:221-7.

518 31. Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950; 37:256-66.

519   32. The GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:

520   multitissue gene regulation in humans. *Science* 2015; 348:648-60.

521   33. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory

522   motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; 40:D930-4.

523   34. Fliegauf M, Benzing T, Omran H. When cilia go bad: cilia defects and ciliopathies. *Nat Rev Mol Cell*

524   *Biol* 2007; 8:880-93.

525   35. Weston MD, Luijendijk MW, Humphrey KD, Moller C, Kimberling WJ. Mutations in the VLGR1

526   gene implicate G-protein signaling in the pathogenesis of Usher syndrome type II. *Am J Hum Genet*

527   2004; 74:357-66.

528   36. Besnard T, Vache C, Baux D, Larrieu L, Abadie C, Blanchet C, *et al.* Non-USH2A mutations in USH2

529   patients. *Hum Mutat* 2012; 33:504-10.

530   37. Leigh MW, Zariwala MA, Knowles MR. Primary ciliary dyskinesia: improving the diagnostic

531   approach. *Curr Opin Pediatr* 2009; 21:320-5.

532   38. Olbrich H, Horvath J, Fekete A, Loges NT, Storm van's Gravesande K, Blum A*, et al.* Axonemal

533   localization of the dynein component DNAH5 is not altered in secondary ciliary dyskinesia. *Pediatr*

534   *Res* 2006; 59:418-22.

535   39. Armengot M, Salom D, Diaz-Llopis M, Millan JM, Milara J, Mata M, *et al.* Nasal ciliary beat frequency

536   and beat pattern in retinal ciliopathies. *Invest Ophthalmol Vis Sci* 2012; 53:2076-9.

537   40. Bonneau D, Raymond F, Kremer C, Klossek JM, Kaplan J, Patte F. Usher syndrome type I associated

538   with bronchiectasis and immotile nasal cilia in two brothers. *J Med Genet* 1993; 30:253-4.

539   41. Lee JH, McDonald ML, Cho MH, Wan ES, Castaldi PJ, Hunninghake GM, *et al.* DNAH5 is associated

540   with total lung capacity in chronic obstructive pulmonary disease. *Respir Res* 2014; 15:97.

541   42. Kovacic MB, Myers JM, Wang N, Martin LJ, Lindsey M, Ericksen MB, *et al.* Identification of KIF3A

542   as a novel candidate gene for childhood asthma using RNA expression and population allelic

543   frequencies differences. *PLoS One* 2011; 6:e23714.

544   43. Dizier MH, Nadif R, Margaritte-Jeannin P, Barton SJ, Sarnowski C, Gagne-Ouellet V, *et al.* Interaction

545   between the DNAH9 gene and early smoke exposure in bronchial hyperresponsiveness. *Eur Respir J*

546   2016; 47:1072-81.

547   44. Almoguera B, Vazquez L, Mentch F, Connolly J, Pacheco JA, Sundaresan AS, *et al.* Identification of

548   Four Novel Loci in Asthma in European and African American Populations. *Am J Respir Crit Care*

549   *Med* 2016 Sep 9; [Epub ahead of print].

550   45. Jones AC, Troy NM, White E, Hollams EM, Gout AM, Ling KM, *et al.* Persistent activation of

551   interlinked Th2-airway epithelial gene networks in sputum-derived cells from aeroallergen-sensitized

552   symptomatic atopic asthmatics. *bioRxiv* 2016; https://doi.org/10.1101/063602

553

SNP	Position (kb)[*]	Alleles[†]	MAF[‡]	EGEA (N=1,660) beta (se) [§]	$P^{\|}$	SLSJ (N=1,138) beta (se) [§]	$P^{\|}$	MRC (N=446) beta (se) [§]	$P^{\|}$	Meta-Analysis beta (se) [§]	P_{stage2}[**]	Overall Meta-Analysis beta (se) [§]	P_{meta}[††]	$P_{Cochran}$[‡‡]
rs4244205	90,188	A/G	0.41	-0.35 (0.08)	1.1×10^{-5}	-0.19 (0.10)	6.0×10^{-2}	-0.14 (0.18)	0.45	-0.18 (0.09)	4.4×10^{-2}	-0.27 (0.06)	3.8×10^{-6}	0.35
rs4916831	90,212	A/G	0.44	-0.40 (0.08)	1.0×10^{-6}	-0.32 (0.11)	2.3×10^{-3}	-0.21 (0.17)	0.23	-0.29 (0.09)	1.2×10^{-3}	-0.35 (0.06)	**6.8×10^{-9}**	0.59
rs10060641	90,213	T/C	0.38	-0.39 (0.08)	7.2×10^{-7}	-0.17 (0.11)	0.11	-0.24 (0.19)	0.22	-0.18 (0.09)	4.8×10^{-2}	-0.30 (0.06)	4.3×10^{-7}	0.23
rs12054681	90,217	C/A	0.37	-0.39 (0.08)	9.7×10^{-7}	-0.14 (0.11)	0.18	-0.29 (0.21)	0.16	-0.17 (0.09)	6.5×10^{-2}	-0.30 (0.06)	7.8×10^{-7}	0.18
rs949787	90,251	G/T	0.28	-0.33 (0.08)	5.5×10^{-5}	-0.23 (0.10)	2.0×10^{-2}	-0.19 (0.20)	0.36	-0.22 (0.09)	1.2×10^{-2}	-0.28 (0.06)	3.2×10^{-6}	0.65

[*]Position in kilobases (kb) according to NCBI dbSNP Build 137

[†]Major allele/Minor allele

[‡]Minor allele frequency

[§]beta is the regression coefficient for a one-unit increase of the effect allele in logistic regression assuming an additive model; se is the standard error associated with the regression coefficient.

[‖]*P* is the *P*-value associated with the Wald test of SNP effect.

[**]P_{stage2} is the *P*-values associated with the Wald test of meta-analyzed SNP effect in the stage 2 datasets (SLSJ and MRC).

[††]P_{meta} is the *P*-value associated with the Wald test of meta-analyzed SNP effect in the three datasets (EGEA, SLSJ, MRC); the *P*-value is shown in bold when it reached the multiple-testing corrected threshold of 1.5×10^{-7}.

[‡‡]$P_{Cochran}$ is the *P*-value associated with Cochran's Q test of homogeneity across the three datasets.

Table II. SNP pairs showing significant interaction for atopy

SNPs	Chr[*]	Genes[†]	Alleles[‡]	MAF[§]	Stage 1 EGEA (N=1,660) Main effect beta (se)[‖]	Interaction beta (se)[‖]	P_{int}[**]	Stage 2 SLSJ (N=1,138) Main effect beta (se)[‖]	Interaction beta (se)[‖]	P_{int}[**]	MRC (N=446) Main effect beta (se)[‖]	Interaction beta (se)[‖]	P_{int}[**]	Meta-Analysis Main effect beta (se)[‖]	Interaction beta (se)[‖]	$P_{stage2\text{-}int}$[††]	Overall Meta-analysis Main effect beta (se)[‖]	Interaction beta (se)[‖]	$P_{meta\text{-}int}$[‡‡]	$P_{Cochran}$[§§]
rs17554723	5	*ADGRV1*	A/G	0.33	-0.06 (0.10)	-0.38 (0.12)	3.0×10^{-3}	0.02 (0.13)	-0.42 (0.16)	1.1×10^{-2}	0.13 (0.23)	-0.28 (0.33)	0.40	0.04 (0.11)	-0.39 (0.14)	6.1×10^{-3}	-0.02 (0.08)	-0.38 (0.09)	**3.6×10^{-5}**	0.84
rs2134256[‖‖]	5	*DNAH5*	T/C	0.25	0.13 (0.09)			-0.06 (0.13)			0.03 (0.26)			-0.04 (0.12)			0.06 (0.07)			
rs17554723	5	*ADGRV1*	A/G	0.33	0.03 (0.09)	-0.34 (0.11)	3.3×10^{-3}	0.13 (0.11)	-0.35 (0.14)	1.4×10^{-2}	0.19 (0.21)	-0.16 (0.29)	0.59	0.14 (0.10)	-0.32 (0.12)	9.3×10^{-3}	0.08 (0.06)	-0.33 (0.08)	**6.1×10^{-5}**	0.92
rs1354187[‖‖]	5	*DNAH5*	T/C	0.36	0.08 (0.08)			-0.04 (0.11)			0.05 (0.22)			-0.02 (0.10)			0.04 (0.06)			

[*]Chr is the chromosome number where the SNP is located

[†]Gene symbol of gene where SNP lies

[‡]Major allele/Minor allele

[§]Minor allele frequency

[‖]beta for the main effect is the regression coefficient for a one-unit increase of the effect allele in logistic regression assuming an additive model; beta for interaction is the regression coefficient for homozygotes for the minor allele at the two loci or homozygotes for the major allele at the two loci with respect to heterozygotes at either one or the two loci using the coding scheme under an additive genetic model described in the methods section; se is the standard error associated with the regression coefficient.

[**]P_{int} is the P-value of the likelihood-ratio test for interaction between SNPs (which follows a chi-square distribution with one degree of freedom assuming an additive model).

[††]$P_{stage2\text{-}int}$ is the P-values associated with the Wald test of meta-analyzed interaction effect in the stage 2 datasets (SLSJ and MRC).

[‡‡]$P_{meta\text{-}int}$ is the P-values associated with the Wald test of meta-analyzed interaction effect in the three datasets (EGEA, SLSJ, MRC); $P_{meta\text{-}int}$ is shown in bold when it reached the multiple-testing corrected threshold of 7.3×10^{-5}.

[§§]$P_{Cochran}$ is the P-value associated with Cochran's Q test of homogeneity across the three datasets.

[‖‖]The two *DNAH5* SNPs, rs1354187 and rs2134256, showing significant interaction with *ADGRV1* SNP are in moderate linkage disequilibrium (r^2=0.50; D'=0.95).

FIGURE LEGENDS

Figure 1. Three-step analysis strategy

Figure 2. Odds-ratio (ORs) and 95% confidence intervals for atopy associated with each genotype at *DNAH5* rs2134256 (TT, CT, CC) in each of the strata defined by genotypes at *ADGRV1* rs17554723 (AA, AG, or GG). These ORs were calculated using the genotype coding scheme defined in the text and are shown for each of the three datasets (EGEA, SLSJ, MRC) and for the combined dataset.
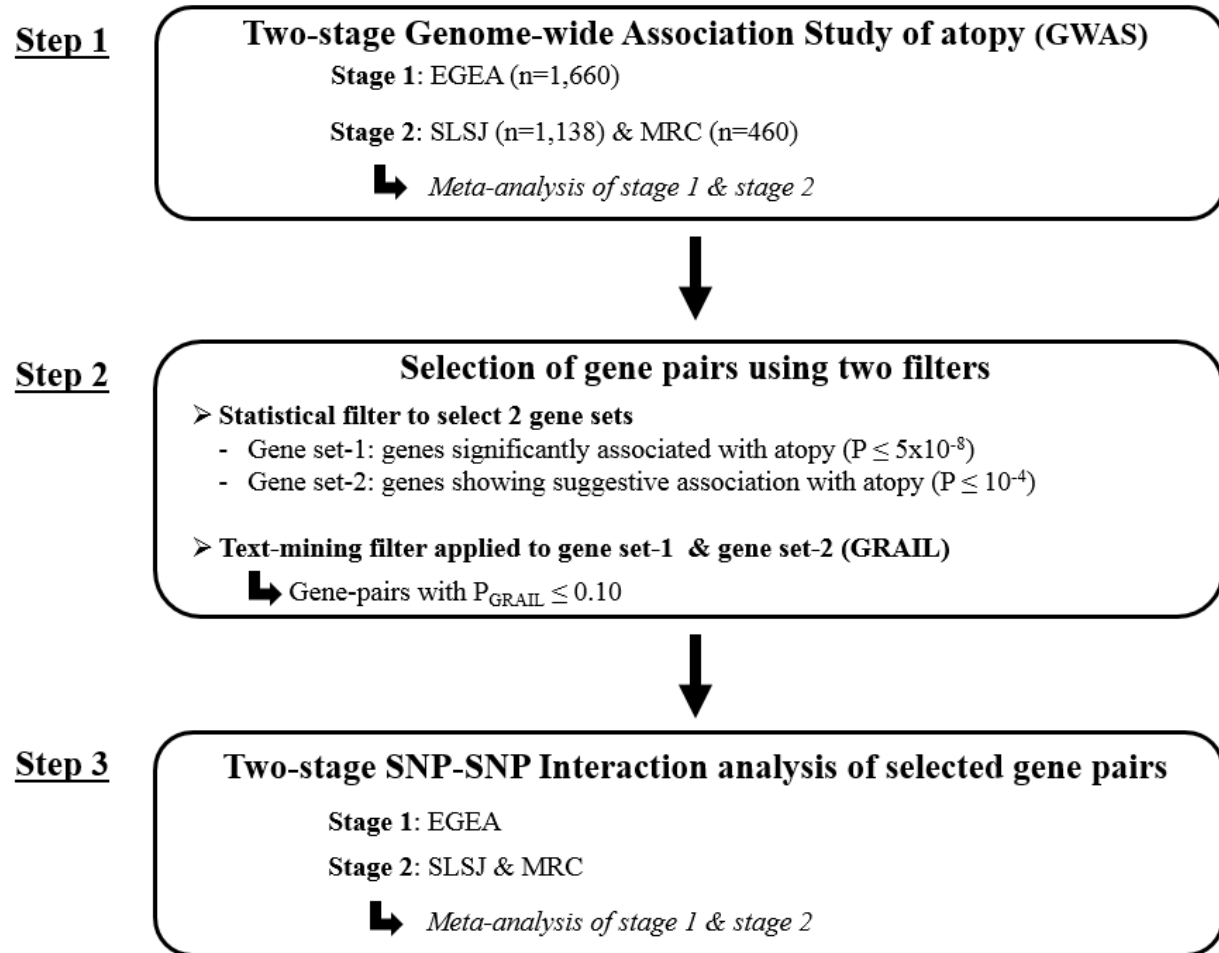
Figure 1.

Step 1

Two-stage Genome-wide Association Study of atopy (GWAS)
Stage 1: EGEA (n=1,660)

Stage 2: SLSJ (n=1,138) & MRC (n=460)

↳ *Meta-analysis of stage 1 & stage 2*

Step 2

Selection of gene pairs using two filters
➢ **Statistical filter to select 2 gene sets**
  - Gene set-1: genes significantly associated with atopy ($P \leq 5 \times 10^{-8}$)
  - Gene set-2: genes showing suggestive association with atopy ($P \leq 10^{-4}$)

➢ **Text-mining filter applied to gene set-1 & gene set-2 (GRAIL)**
  ↳ Gene-pairs with $P_{GRAIL} \leq 0.10$

Step 3

Two-stage SNP-SNP Interaction analysis of selected gene pairs
Stage 1: EGEA
Stage 2: SLSJ & MRC

↳ *Meta-analysis of stage 1 & stage 2*

Figure 2.