OPEN ACCESS Freely available online

PLoS one

Pooled Genome-Wide Analysis to Identify Novel Risk Loci for Pediatric Allergic Asthma

Giampaolo Ricci[1]*, Annalisa Astolfi[2], Daniel Remondini[3,4], Francesca Cipriani[1], Serena Formica[1,2], Arianna Dondi[1], Andrea Pession[1,2]

1 Pediatric Unit, Department of Gynecologic, Obstetric and Pediatric Sciences, University of Bologna, Bologna, Italy, **2** Interdepartmental Centre for Cancer Research "G. Prodi," University of Bologna, Bologna, Italy, **3** Department of Physics, University of Bologna, Bologna, Italy, **4** Interdepartmental Centre "L. Galvani", University of Bologna, Bologna, Italy

Abstract

Background: Genome-wide association studies of pooled DNA samples were shown to be a valuable tool to identify candidate SNPs associated to a phenotype. No such study was up to now applied to childhood allergic asthma, even if the very high complexity of asthma genetics is an appropriate field to explore the potential of pooled GWAS approach.

Methodology/Principal Findings: We performed a pooled GWAS and individual genotyping in 269 children with allergic respiratory diseases comparing allergic children with and without asthma. We used a modular approach to identify the most significant loci associated with asthma by combining silhouette statistics and physical distance method with cluster-adapted thresholding. We found 97% concordance between pooled GWAS and individual genotyping, with 36 out of 37 top-scoring SNPs significant at individual genotyping level. The most significant SNP is located inside the coding sequence of C5, an already identified asthma susceptibility gene, while the other loci regulate functions that are relevant to bronchial physiopathology, as immune- or inflammation-mediated mechanisms and airway smooth muscle contraction. Integration with gene expression data showed that almost half of the putative susceptibility genes are differentially expressed in experimental asthma mouse models.

Conclusion/Significance: Combined silhouette statistics and cluster-adapted physical distance threshold analysis of pooled GWAS data is an efficient method to identify candidate SNP associated to asthma development in an allergic pediatric population.

Citation: Ricci G, Astolfi A, Remondini D, Cipriani F, Formica S, et al. (2011) Pooled Genome-Wide Analysis to Identify Novel Risk Loci for Pediatric Allergic Asthma. PLoS ONE 6(2): e16912. doi:10.1371/journal.pone.0016912

Editor: Jeffrey Whitsett, Cincinnati Children's Hospital Medical Center, United States of America

Received September 15, 2010; **Accepted** January 3, 2011; **Published** February 16, 2011

Funding: This study was supported by the "Fondazione del Monte di Bologna e Ravenna." S.F. is supported by the "Vanini-Cavagnino" grant, F.C. is supported by a grant from the University of Bologna, D.R. is supported by the UniBO Strategic grant "p53 and Non-Neoplastic Pathologies." The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: giampaolo.ricci@unibo.it

Introduction

Asthma is a chronic respiratory disease resulting from a complex interaction of multiple genetic and environmental factors. In the past decades, more than 200 asthma candidate genes have been identified using genetic association studies, positional cloning and knockout mouse approaches [1], but only in the recent years it has been possible to perform whole-genome investigations largely due to the genome-wide association studies (GWAS) [2–4], that have soon shown to be powerful tool to identify novel loci and susceptibility variants for common diseases.

Genome-wide association studies of pooled DNA samples were shown to be a valuable tool to identify in a fast, scalable and economical way candidate SNPs associated to a phenotype [5–7]. This method was applied to different SNP-array platforms and different diseases and QTL phenotypes, particularly those related to complex traits as intellectual and psychological abilities, multiple sclerosis or Alzheimer disease [8–13]. Many tools and analysis pipelines have also been developed to improve the ability to identify true associations among thousands of potential candidates [10,14–16]. However, very few studies evaluated the efficacy of different methods in selecting candidate SNPs for second-stage validation, invariably reporting a high percentage of false positive results.

No study was up to now applied to allergic asthma, even if the very high complexity of asthma genetics is an appropriate field to explore the potential of pooled GWAS approach. This high complexity ensues from the frequent causal association of asthma with several other allergic phenotypes that may constitute a complex confounding factor for the identification of asthma susceptibility genes in the classical case-control design of GWAS (asthmatic *vs* healthy subjects).

To address both these issues, in this study we performed a pooled GWAS on pediatric allergic asthma, comparing asthmatic children to allergic subjects, and introducing a new method of analysis that incorporate silhouette statistics with quality control, physical distance and cluster-adapted thresholding, that reached 97% efficiency in selecting significant susceptibility loci associated to allergic asthma onset.

Methods

Pediatric patients cohort

A total of 269 children of white European descent, aged 6–18 years, with a diagnosis of allergic asthma and/or allergic rhinoconjunctivitis (RC) visited at outpatient clinic of Pediatric Allergologic Unit were included in the study. Patients were divided into two groups: patients with *asthma* (also including patients with both allergic asthma and rhinoconjunctivitis) and patients with *rhinoconjunctivitis* (RC), affected by allergic rhinitis or rhinoconjunctivitis who had never shown asthmatic symptoms. The mean age at the blood sampling was 10.6±3.6 yrs for *asthma* group and 9.8±3.7 for RC group (**Table 1**). Among patients included into *asthma* group, 95.33% developed asthmatic symptoms before 10 years of age and this evidence is confirmed by the most recent Italian and European studies which show that a percentage around 90% of subjects with full blown asthma become symptomatic before 10 years of age [17,18]. Consequently we can assume that patients included in the RC group have a very low probability to develop asthma.

An independent validation set, including 35 patients with asthma and 44 with rhinoconjunctivitis (RC), was selected in a second stage of the study. Clinical, allergometric and spirometric characteristics of the patients were comparable to the ones included in the first stage of the study.

Ethics Statement

The study was conducted according to the principles expressed in the Declaration of Helsinki, and approved by Ethic Committee of University Hospital – S. Orsola-Malpighi of Bologna (*AllerGene* protocol no. 134/2008/U/Tess). Written informed consent was obtained from all the parents or guardians of the minors involved in the study.

Allergometric assays

Patients' allergometric assessment was evaluated through the determination of specific IgE levels (UNICAP1000, Phadia, Uppsala, Sweden) and by performing Skin Prick tests (Lofarma, Milan, Italy). Skin prick tests were performed for the main inhalant allergens: Grass pollen, *Parietaria mix*, *Compositae mix*, birch pollen, hazelnut pollen, *D. pteronyssinus*, *D. farinae* and *Alternaria a.*, *Cladosporium*, cat's epithelium and dog's dandruff. Referring to the ACAAI Practice Parameters [19] wheals with a diameter ≥ histamine (described as + + for a convention) was considered positive. For each patient total IgE and specific IgE for the main inhalant (*Cynodon d.*, *Phleum p.*, *Ambrosia e.*, *Artemisia v.*, *Parietaria J.*, *Betula v.*, *Corylus a.*, *Olea e.*, *D. pteronyssinus*, *D. farinae*, *Alternaria a.*, dog's dandruff, cat's epithelium) were determined; we considered positive specific IgE values higher than 1 kU/L to avoid immunological cross-reactivity. Only the sensitization for the most represented allergens has been reported (**Table 1**).

Assessment of pulmonary status of patients

Pulmonary status of patients was evaluated by performing spirometric tests (ZAN100, nSpire Health GmbH, Germany). In asthmatic children lung function tests were performed both for diagnostic purposes and to establish the efficacy of pharmacological therapy (inhalatory corticosteroids), while in patients with RC to exclude a silent pulmonary inflammatory condition. Measurements of lung function were performed both for *asthmatic* and *non asthmatic* children and **s**pirometric parameters were evaluated as variation between values at baseline and after administration of rapid-acting bronchodilator (e.g. after 200–400 μg salbutamol), both to demonstrate the reversibility of lung function abnormalities and to exclude an eventual airflow limitation also in patient with normal value at baseline. The measured variables were evaluated according to the GINA guidelines [20] and were

Table 1. Clinical, allergometric and spirometric characteristics of all the children included in the study.

	Asthma **135 pts**	***RC*** **134 pts**	***p value***	***OR***
Age at blood sampling – mean ± SD (yrs)	10.6±3.6	9.8±3.7	NS	-
Onset age – mean ± SD (yrs)	5.2±2.8	6.2±3.1	NS	-
Sex (male/female)	2.65	1.44	0.02	1.84
Atopic Dermatitis (%)	24.4	31.3	NS	-
Other allergic manifestations (%)	15.6	14.2	NS	-
Total IgE – g. mean (kU/L)	423.6	245.9	0.0003	-
Phleum p. – sIgE g. mean (kU/L)	18.4	18.7	NS	-
Phleum p. – sIgE + (%)	74.1	69.4	NS	-
Grass pollen – SPT + (%)	72.6	70.9	NS	-
D.pteronyssinus - sIgE g. mean (kU/L)	6.6	3.1	0.02	-
D. pteronyssinus – sIgE + (%)	39.3	21.6	0.002	2.34
D. pteronyssinus – SPT + (%)	39.3	21.6	0.002	2.34
Baseline FVC %	101.2±16.0	104.0±11.4	NS	-
Baseline FEV_1/FVC %	98.9±8.6	105.5±9.1	0.0005	-
Δ FEV_1 %	5.6±5.2	2.3±4.5	0.002	-

Patients were divided into A*sthma* group and RC group, detected at enrolment. Allergic sensitization was reported for the main inhalant allergens: grass pollens (*Phleum p.*) and house dust mites (*D. pteronyssinus*) both as specific IgE and Skin Prick test. Main spirometric parameters were expressed as rate of predicted values (%). Δ FEV_1 is the difference in FEV_1 values before and after salbutamol administration.

doi:10.1371/journal.pone.0016912.t001

reported the most representative indicators of asthmatic condition: FVC (forced vital capacity), FEV_1 (forced expiratory volume in one second), FEV1/FVC ratio (**Table 1**).

Clinical statistical analysis

Statistical significance for age at blood sampling and onset age of symptoms was evaluated by performing Mann Whitney test. The Chi squared test was performed to calculate the significance of gender differences, prevalence of atopic dermatitis and other allergic manifestations and the prevalence of sensitization to the major inhalant allergens, for which Odds Ratio (OR) was also estimated. Total and specific IgE values, expressed as geometric mean, were evaluated by Fisher exact test. Values of $P<0.05$ were considered as significant. Spirometric measurements were reported as average values and their significance was evaluated by Mann Whitney test (**Table 1**).

Power analysis

Sample dimension was calculated by modeling the LD coefficient (D') as a function of the genetic distance between the associated marker allele and the disease locus (assuming the worst-case scenario of the disease locus being located midway between two markers) and of the number of generations elapsed (G) since the mutation was introduced. G was chosen as representative of an outbred population ($G = 500$), and the recombination fraction between marker and disease locus was estimated from the number of markers on the array (roughly 500,000).

Sample dimension was then estimated, assuming a frequency of the predisposing allele of 0.15, a disease prevalence of 0.4 (prevalence of asthma in the allergic population is $\geq$40%) [21], a genotype relative risk of 1.7 (for both the homozygote and heterozygote), and a marker frequency of 0.2 (taken from Affymetrix product description). Under these conditions it is possible to identify such a susceptibility allele with a statistical power of 80%, a type I error rate of 0.05, and a study of 136 cases and 136 controls. The estimates were calculated using the Genetic Power Calculator web tool [22].

DNA extraction and sample pooling

DNA was extracted from all the blood samples with Nucleospin DNA kit (Macherey & Nagel, Duren, Germany) and individual DNA concentration was determined in triplicate with the Quant-iT Pico-Green dsDNA Assay Kit (Invitrogen, Milan, Italy). These triplicate values were used to calculate a mean concentration for each sample. Samples were excluded from the analysis if the coefficient of variation between the three replicates was <0.1.

DNA samples were assigned to the *Asthma* group if displaying symptoms of asthma, alone or associated to other allergic phenotypes, including RC, and assigned to the *RC* group if displaying rhinitis or rhinoconjunctivits alone or associated to other allergic phenotypes, excluding asthma. Each of the two groups was subdiveded into 4 independent groups of samples, each containing 31-36 individuals. Individual DNA samples were then added to their respective pools in equivalent molar amounts. The design of multiple replicate pools of distinct individuals was chosen as it decreases quantification error by reducing quantification error variance proportionally to the number of times a single pool is independently constituted (3 times), and measurement error variance by the number of independent analysis (24 times) [23].

Genome Wide SNP genotyping

Pooling was performed three times to account for pipetting variability. Each of these replicates was genotyped on Mapping 500K arrays (Affymetrix, Santa Clara; CA) following manufacturer's instructions. Detection rates were calculated with GTYPE 4.1 (Affymetrix) using the MPAM (*Modified Partitioning Around Medoids*) MDR algorithm, that represents the number of SNPs that passes the MPAM discrimination filter/total number of SNPs. MDR $\pm$ SD was 98.3 ± 1.8. Raw data were deposited in the GEO repository with the GSE24481 reference number.

Analysis of pooled SNP genotyping data

Probe intensity data was taken directly from the CEL file and used to calculate a Relative Allele Signal (RAS) score by the MPAM algorithm implemented in the SNP-MaP script (http://sgdp.iop.kcl.ac.uk/oleo/affy). This method calculates the average of RAS1 (sense) and RAS2 (antisense) values as a quantitative index of allele frequencies in pooled DNA [6].

Probes were filtered by overall average RAS ($0.1<RAS<0.9$) and AAD (*Average Absolute Difference*, $AAD<0.28$) to exclude data with low Minor Allele Frequency and high variability.

To control for possible confounding bias due to population stratification we applied Principal Component Analysis (PCA) on the RAS values from the Top 500 ancestry-informative SNPs identified by Drineas et al. [24], selected from a subset Population Reference Sample (POPRES) representative of 11 different populations and analyzed on Mapping 500K arrays (Affymetrix). PCA was done also in comparison to allele frequencies in CEU, YRI, CHB and JPT population taken from Hapmap data.

We used silhouette score [25,26] to assess SNP significance. Silhouette score is a measure to quantify the goodness of data clustering, defined as the average silhouette calculated over all cluster elements. For the *i*-th cluster element, silhouette is defined as

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

where a_i is the average (Manhattan) distance of the i-th element from each element of the cluster, and b_i is the average distance over the other clusters (in our case we have only two clusters, Asthma and RC). Silhouette values range from -1 (bad clustering) to 1 (optimal clustering).

We chose not to use an F test statistics because very probably the RAS values (ranging from 0 to1) and in particular the extreme values, do not fulfil the condition of gaussianity necessary for the application of the test. Also the absolute values of the differences were not considered since they do not take into account sample variability.

For the first step of our analysis we considered 1% top-scoring probes (about 3000 probes selected). We decided not to choose a too conservative threshold for single-probe significance (e.g. by comparison with silhouette values of label-reshuffled data) but we tried to evaluate a multi-probe significance by looking at the presence of significant SNP "blocks" (i.e. clustered together along the chromosome) that likely reflect the presence of LD blocks including a large number of (even weakly) significant probes. We evaluated the robustness of our results considering different starting datasets (5000 and 7000 top-scoring probes) and different parameters for clustering (chromosomal distance threshold from 20 Kb to 35 Kb, see below) obtaining a good agreement in the final SNP lists.

SNPs were mapped onto the Human Reference Sequence *hg19* GrCh37 assembly (UCSC Genome Browser). Then we clustered probes with genomic distance <30 Kb, and we considered only probes belonging to at least a 2-probe cluster. For cluster crossvalidation, we assigned randomly the calculated silhouettes

to all probes, and then repeated all the analysis steps up to calculating average cluster silhouette for each cluster size (100 times): average cluster silhouette obtained by reshuffling was compared to the real average cluster silhouette values, producing the final list of significant SNPs.

Validation of candidate SNPs

The LD structure of the significant clusters was visualized by GOLD heatmap and analyzed using Haploview software v. 4.1. on Caucasian phased genotypes from the International Hapmap Project (www.hapmap.org). SNAP tool (SNP annotation and proxy search) was used to find genes proxy to significant SNP clusters by Linkage Disequilibrium, by setting r^2 threshold to 0.8 and distance to 500 Kb, on Hapmap 22 release data (http://www.broadinstitute.org/mpg/snap/). For each cluster average silhouette score was reported.

Genotyping of the Top scoring SNPs for the validation study was performed using the MassArray system on a Sequenom MALDI-TOF device (Sequenom Inc., San Diego, CA) on the entire dataset. Primer sequences and PCR conditions are available upon request. Call rate was 100% for all SNPs and samples analyzed. SNPs were assessed for Hardy-Weinberg equilibrium and analyzed by a two-sided Fisher exact test to compare allele and genotype frequencies between cases and controls. Strength of association was estimated by conditional OR ±95%CI. Quality controls and allelic and genotypic association tests were performed using the SNPator package (http://www.snpator.org). Genotyping of the 24 top scoring SNPs associated with known genes was also performed on the Sequenom system for the independent validation set.

Gene expression

Gene expression analysis was performed on GSE6858 and GSE1301 Gene Expression Omnibus databases (http://www.ncbi.nlm.nih.gov/geo/), that report gene expression profiling data on two relevant mouse models of experimental asthma, analyzed on MG430 2.0 and MOE430A Affymetrix arrays, respectively. The first analyzes experimental asthma induced in BALB/c mice by sensitization and challenge with the allergen ovalbumin [27], while the second one explores the response of whole lungs of BALB/c mice to asthma induced by house dust mite sensitization and challenge, *vs* control mice. CEL files were downloaded, normalized with *rma* algorithm and filtered based on gene expression level (>5 in the log2 scale, at least 50% of samples) and InterQuartile Range (IQR >10% of average total IQR). Differential genes between allergen-challenged and control mice were selected by a modified *t*-test implemented in *limma* package [28], with $P<0.05$ cutoff. Probe sets corresponding to mouse orthologs of human genes were identified by the Affymetrix annotation in the NetAffx database.

Results

We recruited 269 children of self-reported white European ancestry, aged 6 to 18 years with a well documented diagnosis of allergic asthma and/or *rhinoconjunctivitis* (RC). Clinical, phenotypic and spirometric data of the subjects recruited in the *AllerGene* study are summarized in **Table 1**. Among the patients of the two groups no differences in the prevalence of atopic dermatitis and other allergic manifestations could be highlighted, while male gender was associated with asthmatic phenotype ($P=0.02$, OR 1.84). Asthmatic patients have higher total IgE values ($P=0.0003$), higher specific IgE values ($P=0.02$) and higher prevalence of sensitization to house dust mite ($P=0.002$) than patients with RC without asthmatic symptoms. According to pulmonary function tests at baseline, no differences in FVC were underlighted between *Asthma* and RC groups, while asthmatic patients showed a lower value of FEV_1/FVC ($P=0.0005$). Spirometric tests performed after administration of salbutamol revealed significant higher variation of FEV_1 in patients with asthma ($P=0.002$).

GWAS analysis by SNP arrays was undertaken on pooled DNA samples from 269 subjects of the study participants on the Mapping 500K Affymetrix platform. Patients were classified into *Asthma* (135) and *RC* (134) groups, and further subdivided into 4 subgroups considered biological replicates. Pooling was performed in triplicate, to account for technical variability (**Figure 1**). This multiple-pool strategy optimizes the power of the analysis by reducing the variance of each source of experimental error [23].

To control for possible stratification bias due to population structure we applied Principal Component Analysis to a list of 500 top scoring SNPs identified as ancestry-informative in a subset of the Population Reference Sample [24]. PCA showed that there is no recognizable substructure inside our data *(Asthma* and RC pools are mixed up), suggesting that no bias related to population structure was inflated into the data (**Figure S1 in File S1**). As expected the distribution of allele frequencies inside the pools is compatible with CEPH ancestry (Northern and Western european, **Figure S1 in File S1**).

Data reliability was assessed by checking Pearson correlation coefficient C among samples: technical replicates resulted more correlated among themselves (C>0.96). We also checked single probe average variance a) for triplicates, b) for classes (*Asthma*/RC) and c) for all arrays, confirming data homogeneity ($\mu_a=0.21$; $\mu_b=0.23$; $\mu_c=0.25$; $\sigma_a=0.54$; $\sigma_b=0.62$; $\sigma_c=0.58$) (**Figure S2 in File S1**).

Probes were then filtered by overall average RAS (*Relative Allele Signal*) and AAD (*Average Absolute Difference*) to exclude data with low Minor Allele Frequency and excessive variability (78.6% selection) (**Figure S3 in File S1**).

RAS values were analyzed by a Silhouette score statistic, that empirically showed to perform well in pooled-GWAS studies [14], and mapped to the human Reference Sequence *hg19* assembly. Significant SNPs were selected as clusters of top 1% significant silhouette scores within a genomic distance of 30 Kb, and crossvalidated to control for type I errors by random reshuffling average silhouette scores. Only clusters showing an average silhouette score above the cluster-adapted threshold were called significant (**Figure 2**).

Of the 113 significant clusters identified, SNPs representative of the first 37 clusters selected as the top-scoring (Silhouette $P<0.001$) were validated by individual genotyping by the MassArray genotyping system (Sequenom). Individual allelic and genotypic frequencies showed significant association with asthma for 36 SNPs (allelic $P<0.05$. **Table S1–S2 in File S1**), showing a 97% concordance between pooled GWAS and individual genotyping results. More than half of the identified candidate genetic loci were in significant LD with known genes from the Refseq database, or directly located inside introns or coding sequences (**Table S1 in File S1**). The most significant cluster as determined by average silhouette score statistic is located inside the coding sequence of *Complement component 5* gene (*C5*) that is already known as an asthma susceptibility gene reported in previous studies [29]. It is localized on chromosome 9q33.2, in LD with two other genes that regulate inflammation and significantly associated with allergic asthma, *GSN* and *RAB14*.

Biological relevance towards asthma physiopathology was supported by gene expression analysis of experimental asthma mouse models and literature-based functional studies of annotated

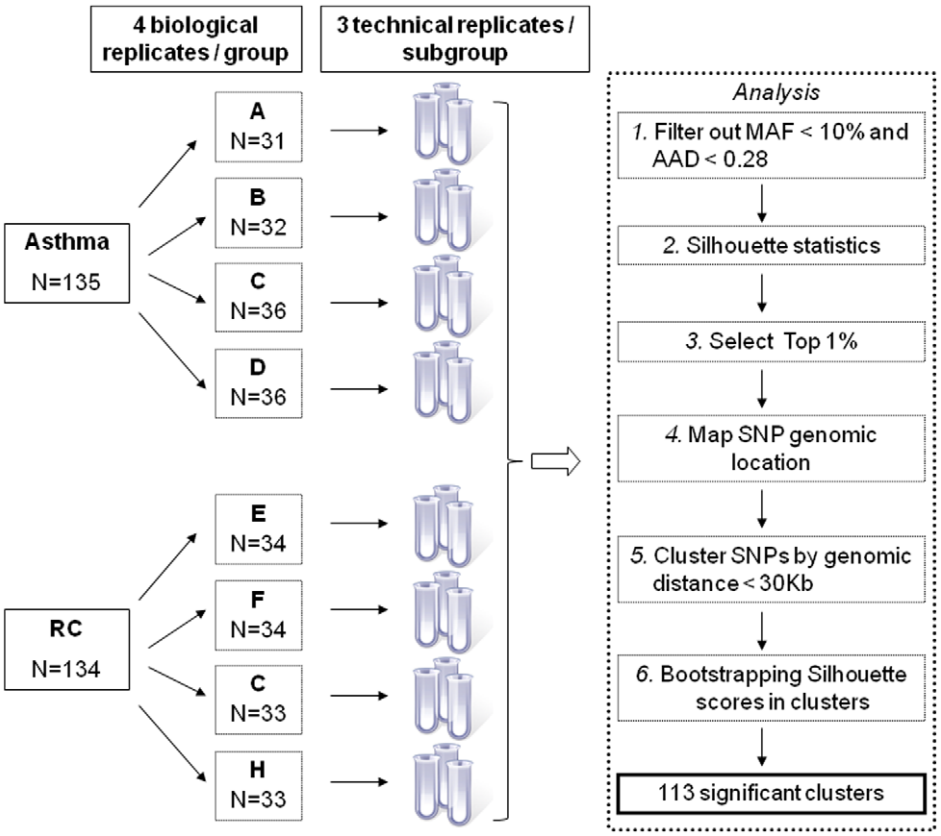

Figure 1. Data analysis pipeline. Asthma and RC samples were divided in 4 pools of 31–36 elements, labeled and hybridized to Mapping 500K SNP arrays. Different pools were replicated three times to account for tecnical and manual variability. Data were filtered, silhouette statistics was applied and top 1% SNP was mapped onto UCSC genome browser and only clusters of at least two SNPs with a genomic physical distance <30 Kb were retained. Cluster-adapted silhouette threshold was calculated by averaging reshuffled random cluster silhouette scores.
doi:10.1371/journal.pone.0016912.g001

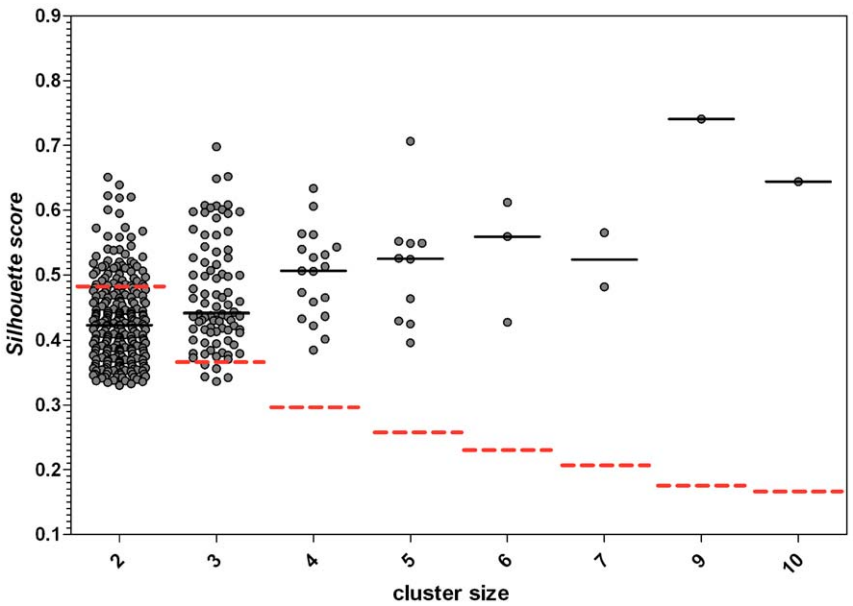

Figure 2. Details of cluster crossvalidation results. Silhouette average score for each cluster is shown as black circles and silhouette average score threshold calculated by random score reshuffling as red dashed lines.
doi:10.1371/journal.pone.0016912.g002

genes for many of the candidate genetic loci. Many of the genes harboring risk alleles are expressed differentially in asthmatic *vs* control mice in two different experimental asthma models (house dust mite or ovalbumin challenge, **Figure 3**) and have a role in the two key pathways involved in asthma physiopathology (**Table 2**): airway smooth muscle contraction or bronchoconstriction (*RYR2*, *TACR3*, *CHRM2*, *PDE5A*) [30–33] and regulation of the pleiotropic mechanisms of inflammation, as modulation of Th2 adaptive immunity and chemotaxis (*C5*, *GSN*, *IPCEF1*), antigen processing (*CPVL*, *PDIA6*, *PTPLB*), cytokine production (*LPAR1*, *FXR1*, *NFKBIZ*) [34–42].

To identify the candidate genetic loci more likely related to allergic asthma onset, we individually genotyped the 24 top scoring SNPs associated with known genes in an independent validation set. RYR2, CHRM2 and TNS3 polymorphysms were confirmed as associated with allergic asthma onset risk in an independent pediatric population (**Table S3 in File S1**).

Discussion

We report here the first pooled genome-wide association study of childhood allergic asthma along with the description of an analysis pipeline that efficiently identified candidate genetic loci linked to asthma development. Indeed the pooling approach poses limitations compared to individual genotyping, that range from the inflation of experimental error due to pooling construction, to the inability to adjust for population stratification and the lack of haplotype and subphenotype data. We addressed the first two points by performing multiple technical replicates of each subpool, and by checking the distribution of ancestry-informative SNPs through principal component analysis [24].

GWAS from pooled DNA samples follow generally a two-stage design in which candidates showing putative association are confirmed by individual genotyping [23]. However, while the statistics to identify the most significant candidates are well defined in GWAS from individual samples, there is no widespread agreement on which analysis pipeline should be used to prioritize SNPs for individual genotyping in pooled GWAS. Few studies explored the efficacy of different methods in selecting candidates; among them Pearson *et al.* compared many statistical methods and verified that the silhouette score was the most efficient method to rank SNPs [10]. However, Bossè *et al.* using silhouette statistics on T2DM dataset analyzed by pooled and individual genotyping documented just 30–40% concordance between the two methods [14]. However the authors also claim that a combination of silhouette scores with "absolute difference" testing is superior to silhouette analysis alone for validation studies including fewer than 1000 SNPs. Abraham *et al.* found 68% concordance, with the cluster method based on physical distance on the genome as the single best method [43]. These publications confirm that statistical tests are inadequate to reveal the most pertinent candidate SNPs in a pooled GWAS. While combination with absolute difference can help identifying those markers having the higher allele frequency variations between the two groups (and indeed absolute difference calculated on the 36 significant SNPs reported in our study is high, ranging from 0.086 to 0.24), combination with physical distance on the genome addresses the almost unique feature of SNP association studies, in which statistical association of one marker to a phenotype is expected to be shared with other markers inside a definite LD block. To address these points in the current study we used a modular approach to identify the most significant loci associated with asthma by combining silhouette statistics and physical distance method with cluster-adapted thresholding; in this way we found 97% concordance between pooled and individual genotyping, with 36 out of 37 top-scoring SNP significant at the individual genotyping level.

Moreover the most significant SNP is located inside the coding sequence of C5, that was already identified as an asthma

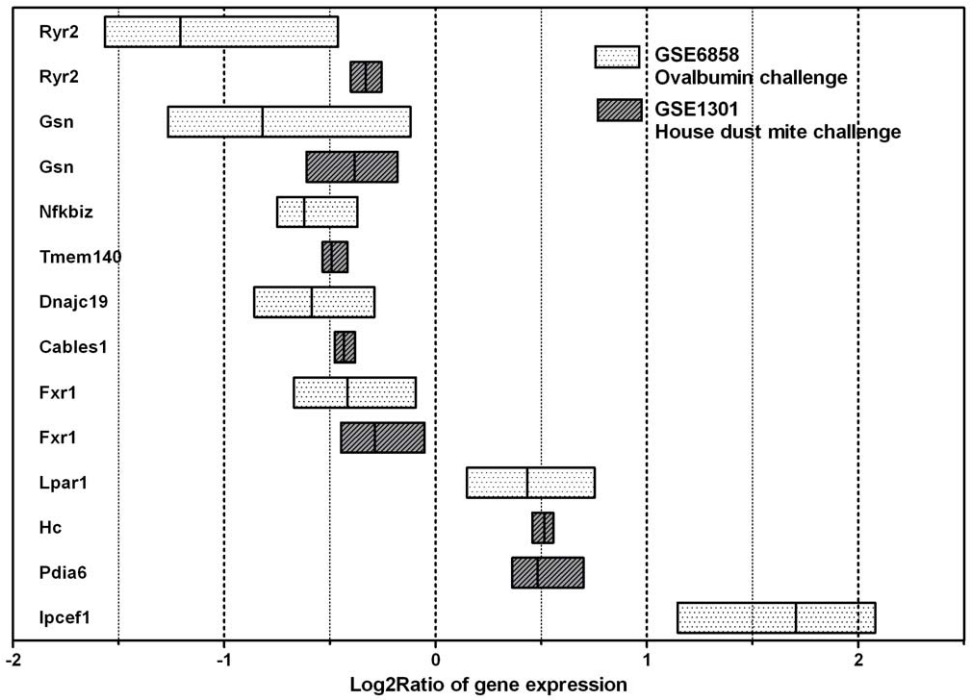

Figure 3. Gene expression profiling of experimental asthma. In two models of experimental asthma induced by Ovalbumin or House dust mite challenge in BALB/c mice there are significant differences in the expression of some genes identified in the GWAS study. Differential gene expression is shown as mean, maximum and minimum of $\log_2$ ratios of allergen-challenged *vs* control mice.
doi:10.1371/journal.pone.0016912.g003

Table 2. Candidate SNPs associated with allergic asthma onset in a pediatric population.

dbSNP ID	Chr	Minor Allele	Gene Simbol	Description	Risk Allele	Allelic p value	Average cluster Silhouette	Risk genotype	Genotypic p value	ASM contraction	Inflam-mation	Gene expression
rs25681	9q33.2	T	**C5**	complement component 5	**C**	**0.0011**	**0.577**	**CC**	**0.0045**	**X**	**X**	**X**
rs4679308	3q21.3	C	**CHCHD6**	coiled-coil-helix-coiled-coil-helix domain containing 6	**T**	**0.0249**	**0.536**	**TT**	**0.0564**			
rs334504	7p12.3	C	**TNS3**	tensin 3	**G**	**0.0118**	**0.520**	**GG**	**0.0173**			
rs17456162	2p25.1	G	**PDIA6**	protein disulfide isomerase isozyme A1	**G**	**0.0261**	**0.519**	**AG+GG**	**0.0261**		**X**	**X**
rs10760153	9q33.2	C	**RAB14**	member RAS oncogene family	**T**	**0.0004**	**0.502**	**TT**	**0.0009**			
rs4580655	4q24	G	**TACR3**	tachykinin receptor 3	**G**	**0.0001**	**0.497**	**GG**	**0.0008**	**X**		
rs790259	6q25.2	G	**OPRM1, IPCEF1**	opioid receptor, mu 1-interaction protein for cytohesin exchange factors 1	**A**	**0.0225**	**0.484**	**AA**	**0.0056**		**X**	**X**
rs3250	7q33	C	**TMEM14, C7orf49**	TMEM140/chromosome 7 open reading frame 49	**C**	**0.0033**	**0.477**	**CC+CT**	**0.0004**			**X**
rs1162394	3p25.3	C	**SRGAP3**	SLIT-ROBO Rho GTPase activating protein 3	**C**	**0.0077**	**0.476**	**CC+CG**	**0.0067**			
rs694936	3q12.3	T	**NFKBIZ**	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta	**A**	**0.0022**	**0.475**	**AA**	**0.0061**		**X**	**X**
rs7792231	7q33	T	**CHRM2**	cholinergic receptor, muscarinic 2	**C**	**0.0002**	**0.459**	**CC**	**0.0001**	**X**		
rs8093359	18q11.2	G	**CABLES1**	Cdk5 and Abl enzyme substrate 1	**G**	**0.0031**	**0.454**	**GG**	**0.01**			**X**
rs12820238	12q21.33	T	**C12orf12, EPYC**	chromosome 12 open reading frame 12/epiphycan	**T**	**0.0019**	**0.450**	**GT+TT**	**0.0018**			
rs2158623	7p15.1	G	**JAZF1**	JAZF zinc finger 1	**T**	**0.0256**	**0.445**	**TT**	**0.032**			
rs2460456	16q24.3	C	**SPG7**	spastic paraplegia 7	**T**	**0.007**	**0.442**	**TT**	**0.0026**			
rs531003	9q31.3	C	**LPAR1**	lysophosphatidic acid receptor 1	**G**	**0.0026**	**0.435**	**GG**	**0.0042**		**X**	**X**
rs506511	7p15.1	T	**CPVL**	carboxypeptidase, vitellogenic-like	**T**	**0.0044**	**0.434**	**TT**	**0.0116**		**X**	
rs6782299	3q26.33	G	**FXR1, DNAJC19**	fragile X mental retardation, autosomal homolog 1/DnaJ (Hsp40) homolog, subfamily C, member 19	**G**	**0.0003**	**0.433**	**GG+GT**	**0.0002**		**X**	**X**
rs10754593	1q43	G	**RYR2**	ryanodine receptor 2	**G**	**0.0003**	**0.431**	**CG+GG**	**0.0016**	**X**		**X**
rs10516997	4p14	C	**APBB2**	amyloid beta (A4) precursor protein-binding, family B, member 2	**G**	**0.0002**	**0.423**	**GG**	**0.0002**			
rs4837827	9q33.2	T	**GSN**	gelsolin	**C**	**0.0011**	**0.412**	**CC+CT**	**0.0056**		**X**	**X**
rs456290	5q33.3	C	**SGCD**	sarcoglycan, delta (35kDa dystrophin-associated glycoprotein)	**C**	**0.0002**	**0.410**	**CC+CT**	**0.0002**			
rs1456114	3q21.1	G	**PTPLB**	protein tyrosine phosphatase-like (proline instead of catalytic arginine), member b	**G**	**0.0032**	**0.403**	**AG+GG**	**0.0023**		**X**	
rs10518329	4q27	G	**PDE5A**	phosphodiesterase 5A, cGMP-specific	**G**	**0.0109**	**0.400**	**AG+GG**	**0.0088**	**X**		

Each cluster of significant SNP is represented here by the one showing the highest silhouette value. Allelic and genotypic *p* values are determined by individual genotyping carried out on the MassArray Sequenom platform. Biological relevance to asthma is shown by literature-based functional classification and gene expression meta-analysis of experimental asthma murine models (ovalbumin or house dust mite challenge). Results are shown by increasing values of allelic *P*. Complete results table, including SNPs not directly associated with Refseq genes (LD $r^2<0.8$ on CEPH Hapmap data) are shown in **Table S1–S2 in File S1**.

doi:10.1371/journal.pone.0016912.t002

susceptibility gene [29], and the vast majority of the other candidate genes regulate functions that are relevant to bronchial physiopathology, as immune- or inflammation-mediated mechanisms and airway smooth muscle contraction. Lastly, the high correlation with gene expression data in experimental asthma models (42%) further supports the results, since genetic variation is known to influence gene expression [44], and gene expression difference itself was proposed as an efficient method to refine the identification of candidate genes and functions in GWAS experiments [45]. Lastly, three out of 24 candidate SNPs located inside known genes were found significantly associated with asthma in an independent validation set.

In this study we identified candidate genetic variants that distinguish, within a population of allergic children, the subjects with and without asthma; these differences may be due to the differences between nasal and bronchial mucosa physiopathology, that may subtend different genetic pathways between asthma and rhinitis. In fact, even if upper and lower airways may be considered as a unique entity supporting the concept of a "united airways" [46], and are influenced by a common, evolving inflammatory process that may be sustained and amplified by interconnected mechanisms, there are many differences between nose and bronchi: smooth muscle is present from the trachea to the bronchioles explaining bronchoconstriction in asthma, cholinergic nerves are the predominant bronchoconstrictor pathway, α-adrenergic agonists are effective nasal vasoconstrictors in rhinitis whereas β2-adrenergic agonists are effective bronchodilators in asthma and many others.

Differences between children with and without asthma are evident from the analysis of the 269 patients recruited for the study from the clinical and allergometric evaluation; actually asthmatic children have higher total IgE values, higher specific IgE values and higher prevalence of sensitization to house dust mite than patients with RC without asthmatic symptoms. We suggest that these clinical and immunological differences reflect an intrinsic genetic diversity in the immunological mechanism of allergic response.

This is the first study done so far in which asthmatic children have been compared not to the healthy population but to allergic subjects with RC. By considering allergic children as the control group, we excluded any interferences of genes involved in the generic mechanisms of allergy, that mainly regulate the mechanisms of inflammation, alteration of epithelial and mucosal functions and modulation of the immune response to environmental factors, thus identifying only new potential genetic pathway explaining lower airway involvement. The limitation of using the RC population as control group is related to the possible later onset of asthma; however, since the mean age of the patients enrolled in the study is significantly higher than the onset age of asthma ($P<0.0001$), we can consider it a very unlikely event.

The vast majority of the identified candidate SNPs were not previously reported, apart from C5 (*Complement component 5*) whose polimorphisms were already associated with asthma development [29]. The anaphylatoxin C5a is found in high concentrations in the bronchoalveolar lavage fluid of asthmatics and mice with experimentally induced asthma; it can induce smooth muscle contraction, mucus secretion, increased microvascular permeability, leukocyte migration and activation, and degranulation of mast cells [47]. Moreover a very recent publication showed that signalling by complement factor C5a plays a key role in the development and severity of asthma, through the inhibition of IL-17-producing helper T cells and airway hyper-responsiveness [48]. Here we further supported the association on C5 with allergic asthma and furthermore suggested that the entire 9q33.2 region harbouring also Gelsolin and Rab14 is associated with asthma, thus identifying new candidate risk alleles in genes regulating inflammation and immunomodulation.

Indeed almost all of the candidate genes identified up to now belong to functional categories (innate immunity and immunoregulation, Th2 differentiation and effector function, mucosal immunity) implicated both in asthma physiopathology and allergic response in the upper airways and mucosa, very few of the identified genes regulate functions peculiar only of the asthmatic response (bronchial hyperresponsiveness and bronchoconstriction) [1]. Here we suggested for the first time that allergic children that develop or not asthma are genetically different, and that many of the predisposing genes regulate functions (airway smooth muscle contraction) that are uniquely relevant to asthma physiopathology and not to allergy. Indeed the fact that RYR2 and CHRM2 polymorphisms were already confirmed as significant in an independent validation set supports the view that genes regulating bronchoconstrictions should be given more attention than those regulating immune mechanisms that are relevant also to allergic manifestations in the upper airways. However a larger validation study is warranted to better define the most relevant target genes.

In summary, we described a new method of analysis of pooled GWAS data that proved to be efficient in identifying significant candidate SNPs associated to asthma onset within an allergic pediatric population and report a list of candidate susceptibility genes whose genetic variability appears to be associated to an increased risk of asthma development in children already carrying a genetic predisposition to allergic diseases.

Supporting Information

File S1 Supplementary tables and figures.
(DOC)

Acknowledgments

We thank all the affected children and their families for their participation in this study. We are grateful to the C.R.B.A. and Dr. Vilma Mantovani for providing access to the Sequenom MassArray platform. We thank Dr. Elena Baldi Cosseddu for useful suggestion and critical review of the statistical approaches and Dr. Letizia Spanti for helpful participation in the clinical data collection and analysis.

Author Contributions

Conceived and designed the experiments: GR AP. Performed the experiments: AA FC SF AD. Analyzed the data: AA DR GR. Contributed reagents/materials/analysis tools: AP. Wrote the paper: GR AA FC DR.

References

1. Vercelli D (2008) Discovering susceptibility genes for asthma and allergy. Nat Rev Immunol 8: 169–182.
2. Wu H, Romieu I, Shi M, Hancock DB, Li H, et al. (2010) Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. J Allergy Clin Immunol 125: 321–327.
3. Weiss ST, Raby BA, Rogers A (2009) Asthma genetics and genomics 2009. Curr Opin Genet Dev 19: 279–282.
4. Zhang J, Pare PD, Sandford AJ (2008) Recent advances in asthma genetics. Respir Res J 9: 4.
5. Butcher LM, Meaburn E, Liu L, Fernandes C, Hill L, et al. (2004) Genotyping pooled DNA on microarrays: A systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. Behavior Genetics 34: 549–555.
6. Meaburn E, Butcher LM, Schalkwyk LC, Plomin R (2006) Genotyping pooled DNA using 100K SNP microarrays: a step towards genome-wide association scans. Nucl Acids Res 34: e28.
7. Docherty SJ, Butcher LM, Schalkwyk LC, Plomin R (2007) Applicability of DNA pools on 500K SNP microarrays for cost-effective initial screens in genome-wide association studies. BMC Genomics 8: 214.

8. Butcher LM, Davis OS, Craig IW, Plomin R (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. Genes Brain Behav 7: 435–446.
9. Docherty SJ, Davis OS, Kovas Y, Meaburn EL, Dale PS, et al. (2010) A genome-wide association study identifies multiple loci associated with mathematics ability and disability. Genes Brain Behav 9: 234–247.
10. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, et al. (2007) Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. Am J Hum Genet 80: 126–139.
11. Comabella M, Craig DW, Camiña-Tato M, Morcillo C, Lopez C, et al. (2008) Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. PLoS One 3: e3490.
12. Bostrom MA, Lu L, Chou J, Hicks PJ, Xu J, et al. (2010) Candidate genes for non-diabetic ESRD in African Americans: a genome-wide association study using pooled DNA. Hum Genet 128: 195–204.
13. Viding E, Hanscombe KB, Curtis CJ, Davis OS, Meaburn EL, et al. (2010) In search of genes associated with risk for psychopathic tendencies in children: a two-stage genome-wide association study of pooled DNA. J Child Psychol Psychiatry 51: 780–788.
14. Bossé Y, Bacot F, Montpetit A, Rung J, Qu HQ, et al. (2009) Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. Hum Genet 125: 305–318.
15. Sebastiani P, Zhao Z, Abad-Grau MM, Riva A, Hartley SW, et al. (2008) A hierarchical and modular approach to the discovery of robust associations in genome-wide association studies from pooled DNA samples. BMC Genet 9: 6.
16. Medina I, Montaner D, Bonifaci N, Pujana MA, Carbonell J, et al. (2009) Gene set-based analysis of polymorphisms: finding pathways or biological processes associated to traits in genome-wide association studies. Nucleic Acids Res 37: W340–W344.
17. Punekar YS, Sheikh A (2009) Establishing the sequential progression of multiple allergic diagnoses in a UK birth cohort using the General Practice Research Database. Clin Exp Allergy 39: 1889–1895.
18. Sestini P, De Sario M, Bugiari M, Bisanti L, Giannella G, et al. (2005) Frequency of asthma and allergies in Italian children and adolescents: results from SIDRIA-2. Epidemiol Prev 29: 24–31.
19. Bernstein IL, Li JT, Bernstein DI, Hamilton R, Spector SL, et al. (2008) Allergy diagnostic testing: an updated practice parameter. Ann Allergy Asthma Immunol 100: S1–S148.
20. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, et al. (2008) Global Strategy for Asthma Management and Prevention: GINA executive summary. ERJ 1: 143–178.
21. Antonicelli L, Micucci C, Voltolini S, Feliziani V, Senna GE, et al. (2007) Allergic rhinitis and asthma comorbidity: ARIA classification of rhinitis does not correlate with the prevalence of asthma. Clin Exp Allergy 37: 954–960.
22. Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19: 149–150.
23. Sham P, Bader JS, Craig I, O'Donovan M, Owen M (2002) DNA Pooling: a tool for large-scale association studies. Nat Rev Genet 3: 862–871.
24. Drineas P, Lewis J, Paschou P (2010) Inferring geographic coordinates of origin for Europeans using small panels of ancestry informative markers. PLoS One 5: e11892.
25. Lovmar L, Ahlford A, Jonsson M, Syvänen AC (2005) Silhouette scores for assessment of SNP genotype clusters. BMC Genomics 6: 35.
26. Rousseeuw PJ (1987) Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics 20: 53–56.
27. Lu X, Jain VV, Finn PW, Perkins DL (2007) Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. Mol Syst Biol 3: 98.
28. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3, No. 1, Article 3.
29. Hasegawa K, Tamari M, Shao C, Shimizu M, Takahashi N, et al. (2004) Variations in the C3, C3a receptor, and C5 genes affect susceptibility to bronchial asthma. Hum Genet 115: 295–301.
30. Du W, Stiber JA, Rosenberg PB, Meissner G, Eu JP (2005) Ryanodine Receptors in Muscarinic Receptor-mediated Bronchoconstriction. J Biol Chem 280: 26287–26294.
31. Mizuta K, Xu D, Pan Y, Comas G, Sonett JR, et al. (2008) GABAA receptors are expressed and facilitate relaxation in airway smooth muscle. Am J Physiol Lung Cell Mol Physiol 294: L1206–L1216.
32. Zhou XB, Wulfsen I, Lutz S (2008) M2 muscarinic receptors induce airway smooth muscle activation via a dual, Gbetagamma-mediated inhibition of large conductance Ca2+-activated K+ channel activity. J Biol Chem 283: 21036–21044.
33. Mullershausen F, Lange A, Mergia E, Friebe A, Koesling D (2006) Desensitization of NO/cGMP signaling in smooth muscle: blood vessels versus airways. Mol Pharmacol 69: 1969–1974.
34. Köhl J, Baelder R, Lewkowich IP, Pandey MK, Hawlisch H, et al. (2006) A regulatory role for the C5a anaphylatoxin in type 2 immunity in asthma. J Clin Invest 116: 783–796.
35. Witke W, Sharpe AH, Hartwig JH, Azuma T, Stossel TP, et al. (1995) Hemostatic, inflammatory, and fibroblast responses are blunted in mice lacking gelsolin. Cell 81: 41–51.
36. Hosur V, Leppanen S, Abutaha A, Loring RH (2009) Gene regulation of alpha4beta2 nicotinic receptors: microarray analysis of nicotine-induced receptor up-regulation and anti-inflammatory effects. J Neurochem 111: 848–858.
37. Fietta A, Bardoni A, Salvini R, Passadore I, Morosini M, et al. (2006) Analysis of bronchoalveolar lavage fluid proteome from systemic sclerosis patients with or without functional, clinical and radiological signs of lung fibrosis. Arthritis Res Ther 8: R160.
38. Wang B, Pelletier J, Massaad MJ, Herscovics A, Shore GC (2004) The yeast split-ubiquitin membrane protein two-hybrid screen identifies BAP31 as a regulator of the turnover of endoplasmic reticulum-associated protein tyrosine phosphatase-like B. Mol Cell Biol 24: 2767–2778.
39. Harris J, Schwinn N, Mahoney JA, Lin HH, Shaw M, et al. (2006) A vitellogenic-like carboxypeptidase expressed by human macrophages is localized in endoplasmic reticulum and membrane ruffles. Int J Exp Pathol 87: 29–39.
40. Choi JW, Herr DR, Noguchi K, Yung YC, Lee CW, et al. (2010) LPA receptors: subtypes and biological actions. Annu Rev Pharmacol Toxicol 50: 157–186.
41. Lachance C, Thuraisingam T, Garnon J, Roter E, Radzioch D (2007) Posttranscriptional gene expression regulation in CpG-activated macrophages depends on FXR1P RNA-binding protein. FEMS Immunol Med Microbiol 51: 422–430.
42. Seshadri S, Kannan Y, Mitra S, Parker-Barnes J, Wewers MD (2009) MAIL regulates human monocyte IL-6 production. J Immunol 183: 5358–5368.
43. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, et al. (2008) A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics 1: 44.
44. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet. 10: 184–194.
45. Gorlov IP, Gallick GE, Gorlova OY, Amos C, Logothetis CJ (2009) GWAS meets microarray: are the results of genome-wide association studies and gene-expression profiling consistent? Prostate cancer as an example. PLoS One 4: e6511.
46. Bousquet J, van Cauwenberge P, Khaltaev N, Aït-Khaled N, Annesi-Maesano I, et al. (ARIA Workshop Group) (2001) Allergic rhinitis and its impact on asthma – ARIA Workshop Report. J Allergy Clin Immunol 108: S147–S336.
47. Lambrecht BN (2006) An unexpected role for the anaphylatoxin C5a receptor in allergic sensitization. J Clin Invest 116: 628–32.
48. Lajoie S, Lewkowich IP, Suzuki Y, Clark JR, Sproles AA, et al. (2010) Complement-mediated regulation of the IL-17A axis is a central genetic determinant of the severity of experimental allergic asthma. Nat Immunol 11: 928–935.