

Genetic interactions model among *Eotaxin* gene polymorphisms in asthma

June-Hyuk Lee · Jason H. Moore · Sung-Woo Park ·
An-Soo Jang · Soo-Taek Uh · Yong Hoon Kim ·
Choon-Sik Park · Byung Lae Park · Hyoung Doo Shin

Received: 14 January 2008 / Accepted: 4 June 2008 / Published online: 20 August 2008
© The Japan Society of Human Genetics and Springer 2008

Abstract *Eotaxin* family (*Eotaxin* 1, 2 and 3) recruits and activates CCR3-bearing cells such as eosinophil, mast cells, and Th2 lymphocytes that play a major role in allergic disorders. We examined the polygenetic effects of the *Eotaxin* gene family in a Korean population. Gene–gene interactions were tested using a multistep approach with multifactor dimensionality reduction (MDR) method between asthmatics and normal controls. The overall best MDR model of the main effect single nucleotide polymorphisms (SNPs) included *EOT2* + 1272A > G and *EOT3* + 77C > T (model 1) [testing accuracy 0.597, cross-validation consistency (CVC) 10/10, $P < 0.001$]. The overall best MDR model of the SNPs with no main effects included *EOT2* + 304C > A, *EOT3* + 716A > G, and *EOT3* + 1579G > A (model 2) (testing accuracy 0.616, CVC 10/10, $P < 0.001$). Model 3 was obtained by including the MDR variables for models 1 and 2. This new composite model predicted asthma with better accuracy

than either model 1 or model 2 (testing accuracy 0.643, CVC 10/10, $P < 0.001$). The detection of statistical interaction models is one evidence of gene–gene interactions among *Eotaxin* genes, and this interaction is thought to influence the development of asthma. Although the models are limited to determining statistical interactions within a population, they may be useful for identifying groups at high risk of developing asthma.

Keywords Asthma · Epistasis · Polymorphism · *Eotaxin*

Abbreviations

MDR multifactor dimensionality reduction
CVC cross-validation consistency
OR odds ratio

Introduction

Asthma is a common respiratory disease characterized by intermittent airway obstruction, chronic airway inflammation, and airway remodeling (Davies et al. 2003). Asthma results from and progresses via a complex interaction of genetic and environmental factors (Steinke et al. 2008). *Eotaxin* is a member of a family of CC chemokines that coordinates the recruitment of inflammatory cells bearing the CCR3 receptor to sites of allergic inflammation (Rankin et al. 2000). *Eotaxin* 1, 2, and 3 messenger ribonucleic acid (mRNA) and proteins are expressed in the airways of asthmatics and normal controls. *Eotaxin* 1 may be important for eosinophilic inflammation in the early phase of the asthmatic response, whereas *eotaxin* 2 and *eotaxin* 3 may control eosinophil recruitment in the later stages of the allergic

J.-H. Lee · S.-W. Park · A.-S. Jang · S.-T. Uh ·
Y. H. Kim · C.-S. Park (✉)

Division of Allergy and Respiratory Medicine,
Department of Internal Medicine,
Genome Research Center for Allergy and Respiratory Diseases,
SoonChunHyang University Bucheon Hospital,
1174, Joong Dong, Wonmi Gu, Bucheon Si,
Gyeonggi Do 420-021, South Korea
e-mail: mdcspark@unitel.co.kr

J. H. Moore
Computational Genetics, Dartmouth Medical School,
706 Rubin Building HB 7937, One Medical Center Drive,
Lebanon, NH 03756, USA

B. L. Park · H. D. Shin
Department of Genetic Epidemiology, SNP Genetics, Inc.,
11th Floor, Mae Hun B/D, 13 Chongro 4Ga,
Chongro-gu, Seoul 110-834, South Korea

response (Berkman et al. 2001; Brown et al. 1998; Papadopoulos et al. 2001; Ying et al. 1999). The human *Eotaxin* gene families are located on chromosomes 17 and 7. *EOT2* + 1272A > G is associated with asthma development and *EOT1* + 123G > A with serum total immunoglobulin E (IgE) levels (Shin et al. 2003). Asthmatics exhibit a gene–dose effect between *EOT2* + 1272A > G and plasma eotaxin 2 levels (Min et al. 2005).

Given that asthma is a multifactorial disease, numerous genes may control its development, each playing only a small role in conferring a genetic predisposition to the disease phenotype. These genes may act independently or interact with other genes that exist in same biological pathway to produce a variable effect (Carlson et al. 2004).

Gene–gene and gene–environment interactions are difficult to detect and characterize using traditional parametric statistical methods such as logistic regression due to the sparseness of the data in high dimensions (Hahn et al. 2003). The multifactor dimensionality reduction (MDR) algorithm is a powerful tool for detecting gene–gene interactions. It uses an exhaustive search and a single classifier to identify the optimal combination of polymorphisms for predicting a discrete disease endpoint (Hahn et al. 2003; Ritchie et al. 2001). Recently, a flexible framework and MDR provided visualization of the information gained through incorporating the interactions of genetic factors (Moore et al. 2006). Furthermore, odds ratio (OR) based MDR provide more information regarding the effect of a certain genotype combination on the disease risk (Chung et al. 2007). A new generalized MDR, a framework based on the score of the generalized linear model, permits adjustment for covariates and handling both dichotomous and quantitative phenotypes (Lou et al. 2007).

Statistical epistasis is difficult to detect and distinguish in human studies due to its inherent nonlinearity. In its extreme form, epistasis can occur in the absence of detectable independent effects of any one polymorphism. This presents several computational and statistical challenges, especially in the context of genome-wide association studies (Moore and Ritchie 2004; Moore et al. 2006). We examined the polygenetic effects of the eotaxin gene family in a Korean population using a multistep approach and MDR analysis.

Materials and methods

Patients and controls

Three hundred asthmatics were enrolled from among patients at the Asthma Genome Research Center at Soonchunhyang University Bucheon Hospital and Seoul Hospital, Korea. Ethical approval was obtained from the

institutional review board of the hospital. All asthma patients had current symptoms, such as wheezing, dyspnea, or cough, and met the criteria for asthma as determined by the American Thoracic Society (Robert et al. 1987). Each patient showed airway reversibility as documented by an inhalant bronchodilator-induced improvement of forced expiratory in one second (FEV1) of more than 15% (Robert et al. 1999) and/or airway hyperresponsiveness as provocative concentration of methacholine required to cause a 20% decrease in FEV1 (PC20) less than 8 mg/ml (Robert et al. 1999). Three hundred normal subjects were recruited from the general population or among the spouses of asthmatic patients based on the following criteria: a negative screening questionnaire for respiratory symptoms (Ferris 1978), a predicted FEV1 > 75%, PC20 ≥ 8 mg/ml, total serum IgE < 300 IU/ml, and a normal chest X-ray. The clinical characteristics of all asthmatic and control subjects are presented in Table 1.

Genotyping by single-base extension and electrophoresis

To genotype polymorphic sites, amplifying and extension primers were designed for single-base extension (SBE). Primer extension reactions were performed with the SNaPshot dideoxynucleotide triphosphate (ddNTP) Primer Extension kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions.

Selection of polymorphisms

Single nucleotide polymorphisms (SNPs) with a rare allele frequency below 0.1 were excluded from the analysis. Between the two SNPs that linked completely ($|D'| = 1$) (Shin et al. 2003), the SNP that showed higher rare allele

Table 1 Clinical manifestations of asthma and control subjects

	Asthma	Control	P value
Number of subjects	300	300	
Age, year: median (range)	43 (8–70)	27 (8–71)	<0.0001
Gender (male/female)	124/176	136/164	0.16
Smoker (%)	26.4	41.4	0.0003
FVC%, pred	87.2 ± 1.1	88.8 ± 0.9	0.28
FEV1%, pred	79.7 ± 1.3	96.4 ± 0.9	<0.0001
% changes of FEV1 by bronchodilator	11.8 ± 0.9	2.8 ± 0.4	<0.0001
PC20 (mg/ml)	2.6 ± 0.3	24.8 ± 0.0	<0.0001
Total IgE (IU/ml)	436 ± 40.4	138 ± 10.5	<0.0001

Chi-square test was used for comparisons of categorical variables, Student's *t* test for continuous variables

FVC forced vital capacity, FEV1 forced expiratory volume in one second, IgE immunoglobulin E, pred predicted value

frequency was selected. Hardy–Weinberg equilibrium and calculation of D' for linkage disequilibrium (LD) were performed using PHASE v2.0.2 (Stephan et al. 2001) and Arelquin v2.0 (Hedric 1978). A total of 14 SNPs in *Eotaxin 1*, 2 and 3 genes were included in the analysis (six SNPs in *Eotaxin 1*, five in *Eotaxin 2*, and three in *Eotaxin 3*).

Multifactor dimensionality reduction analysis

Data were randomly divided as follows: 9/10 were used as a training set, and the remaining 1/10 was used for independent testing for cross-validation consistency (CVC). Cross-validation is a measure of the number of times a particular set of loci is identified in each possible 9/10 of the subjects. The threshold ratio is defined as the ratio of the number of affected individuals to that of unaffected individuals. The subjects are at high risk only when the threshold ratio exceeds 1.0.

A set of n genetic factors was selected, and all possible multifactor classes or cells were represented in n -dimensional space. Each multifactor class was labeled as either high or low risk, depending on the threshold ratio. This process was repeated for each possible cross-validation interval. When the final best model was selected, a model for high- and low-risk genotype combinations was formed using an adjusted threshold that was equal to the ratio of cases and controls in a model that maximizes the CVC and minimizes the prediction error (Hahn et al. 2003; Moore and Ritchie 2004; Ritchie et al. 2001).

Accuracy is defined as the proportion of subjects that are grouped correctly according to their status. When the CVC was maximal for one model and accuracy was maximal for another, statistical parsimony was used to choose the best model. Thus, when CVC and accuracy supported different models, the model with fewest loci/factors was selected (Hahn et al. 2003; Moore and Ritchie 2004; Ritchie et al. 2001).

Multistep approach using MDR, interaction information, and dendrogram

Gene–gene interactions were evaluated using the flexible four-step computational strategy (Moore et al. 2006). First, we applied a chi-square test of independence to obtain a list of SNPs with statistically significant main effects ($P \leq 0.05$) and a list of SNPs without significant main effects ($P > 0.05$). Second, we determined all possible combinations of SNPs from the main-effects list and no-main-effects list up to a maximum order of five using the MDR constructive induction algorithm (Hahn et al. 2003; Hahn and Moore 2004; Moore and Ritchie 2004; Moore et al. 2006, 2007; Ritchie et al. 2001, 2003). Third, we used a naïve Bayes classifier in the context of a tenfold cross

validation to estimate the testing accuracy of each best two-, three-, and four-factor model. A single best model from the main-effects analysis and the no-main-effects analysis that maximized the testing accuracy was selected. These models are the most likely to generalize to independent data sets. Statistical significance was evaluated using a sign test to compare the observed testing accuracies to those expected under the null hypothesis of no association. Models were considered significant at $P < 0.05$.

We then selected the best model derived from the list of SNPs with significant main effects, the best model from SNPs with no main effects, created new MDR attributes for each model, and placed these back into the data set in a process referred to as interleaving (Moore et al. 2006). We reran MDR with these new constructed attributes and reported the best final model. To estimate the contribution of associated genotype combination, we calculated the OR of each genotype combination in final model using OR-based MDR (OR-MDR) (Chung et al. 2007). Covariate analysis for the final model was performed using generalized MDR (GMDR) (Lou et al. 2007). Age and gender were used as covariates. Age was the continuous variable and gender was the discrete variable. Finally, we used the measure of interactions information to provide a statistical interpretation of the gene–gene interaction models (Andrew et al. 2006; Moore et al. 2006). Interaction information was measured among two given loci and case-control status using Shannon entropy (Jakulin et al. 2003). Let $H(X)$ be the Shannon entropy of X . The information gain (IG) was derived as follows:

$$IG(ABC) = I(A; B|C) - I(A; B)$$

$$I(A; B|C) = H(A|C) + H(B|C) - H(A; B|C)$$

$$I(A; B) = H(A) + H(B) - H(A; B)$$

where $I(A; B)$ denotes the dependency of correlation between A and B , and $I(A; B|C)$ denotes the interaction of A and B given C (Andrew et al. 2006). When the difference between these two parameters, $IG(ABC)$, is positive, it is defined as synergy or evidence of an attribute interaction, whereas when interaction information is negative, it is defined as redundant or evidence of an independency. An interaction dendrogram is presented in which interactions between two loci are indicated by different colors. These analyses were implemented using MDR (version 10.0; <http://www.epistasis.org>), OR-MDR (v1.2), and GMDR software (v0.7).

Results

Genotype distributions of the 14 SNPs were in Hardy–Weinberg equilibrium ($P > 0.05$, data not shown) in

asthmatics and control subjects. The two statistically significant SNPs selected by the chi-square test of independence included $EOT2 + 1272A > G$ ($P = 0.003$) and $EOT3 + 77C > T$ ($P = 0.018$, data not shown). These two SNPs were included in the first list of SNPs with significant main effects. The other 12 SNPs were not significant and were included in the second list. Table 2 summarizes the results of an exhaustive MDR analysis that evaluated the pairwise combinations for the two main-effects SNPs. Table 3 summarizes the results of an exhaustive MDR analysis that evaluated all possible two-, three-, and four-SNP models from the list of SNPs without main effects. In each table, the best model for each order is shown, along with its testing accuracy, CVC, and significance level as determined by the sign test. The overall best MDR model for the main-effects SNPs included $EOT2 + 1272A > G$ and $EOT3 + 77C > T$. This model (model 1) had a maximum testing accuracy of 0.597 and a maximum CVC of 10/10 (Fig. 1). This model was significant at the level of 0.001, indicating that all ten testing accuracies were greater than 0.5 during cross validation. Therefore, it is unlikely that these results fall under the null hypothesis of no association. The OR for model 1 was

Table 2 Summary of multifactor dimensionality reduction (MDR) analysis for the single nucleotide polymorphisms with significant main effects

Model	Testing accuracy	CVC	P value
$EOT2 + 1272A > G$	0.577	10	0.172
$EOT2 + 1272A > G, EOT3 + 77C > T$	0.597	10	0.001 ^a

P values were determined by sign test

CVC cross-validation consistency

^a Overall best MDR model

Table 3 Summary of multifactor dimensionality reduction (MDR) analysis for the single nucleotide polymorphisms with no significant main effects

Model	Testing accuracy	CVC	P value
$EOT2 + 1916A > C$	0.531	4	0.989
$EOT2 + 304C > A, EOT3 + 716A > G$	0.577	10	0.172
$EOT2 + 304C > A, EOT3 + 716A > G, EOT3 + 1579G > A$	0.616	10	0.001 ^a
$EOT2 + 304C > A, EOT2 + 447C > T, EOT3 + 716A > G, EOT3 + 1579G > A$	0.599	6	0.172

P values were determined by sign test

CVC cross-validation consistency

^a Overall best MDR model

2.439 [95% confidence interval (CI), 1.732–3.433]. The distribution of asthmatics and controls for model 1 is illustrated in Fig. 1 for all genotype combinations and for the new MDR-constructed variable. The overall best MDR model for SNPs with no main effects included $EOT2 + 304C > A$, $EOT3 + 716A > G$, and $EOT3 + 1579G > A$ (Table 3). This model (model 2) had a maximum testing accuracy of 0.616 and a maximum CVC of 10/10. Model 2 was significant at the level of 0.001, which indicates that all ten testing accuracies were greater than 0.5 during cross validation. Therefore, it is unlikely that this result falls under the null hypothesis of no association (Fig. 2). The OR for model 2 was 3.360 (95% CI, 2.318–4.871). The distribution of asthmatics and controls for model 2 is illustrated in Fig. 2 for all genotype combinations and for the new MDR-constructed variable. Model 2 was a better predictor of asthma than the significant main effects model. Model 3 was obtained by including the MDR variables for models 1 and 2 in the data set (Fig. 3). This new composite model had a testing accuracy of 0.643 ($P = 0.001$) and was a better predictor than either model 1 or 2. The global OR for model 3 was 3.287 (95% CI, 2.349–4.600). Among 243 possible genotype combinations of five SNPs in model 3, eight genotype combinations had relatively high OR (>1), and we could find out which genotype combinations of model 3 contributed more to case-control status (Table 5). In covariate analysis using GMDR, testing accuracies were similar between model 3 with and without age and gender adjustment (0.6075 vs. 0.6062, respectively) (Table 6).

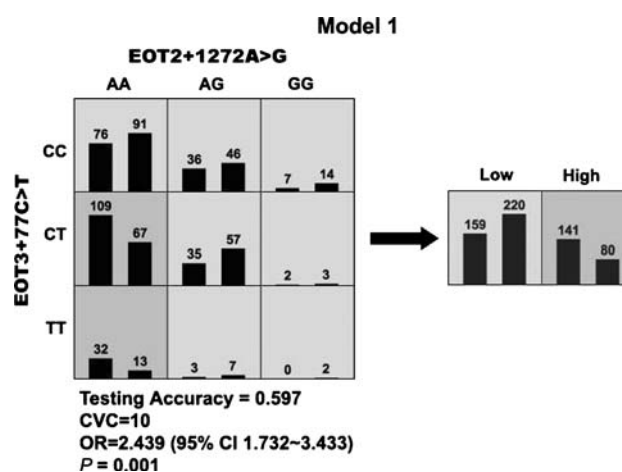


Fig. 1 Distribution of asthmatics (left bars) and controls (right bars) for each genotype combination from the two single nucleotide polymorphisms (SNPs) that had statistically significant main effects. High-risk genotype combinations are shaded dark grey and low-risk are shaded light grey. The new variable constructed by multifactor dimensionality reduction (MDR) is shown on the right. High high-risk group, Low low-risk group, CVC cross-validation consistency, OR odds ratio, CI confidence interval

Fig. 2 Distribution of asthmatics (left bars) and controls (right bars) for each genotype combination from the best combination of three single nucleotide polymorphisms (SNPs) that had no significant main effects. High-risk genotype combinations are shaded dark grey and low-risk are shaded light grey. White cells indicate no data was observed for that combination. The new variable constructed by multifactor dimensionality reduction (MDR) is shown below. High high-risk group, Low low-risk group, CVC cross-validation consistency, OR odds ratio, CI confidence interval

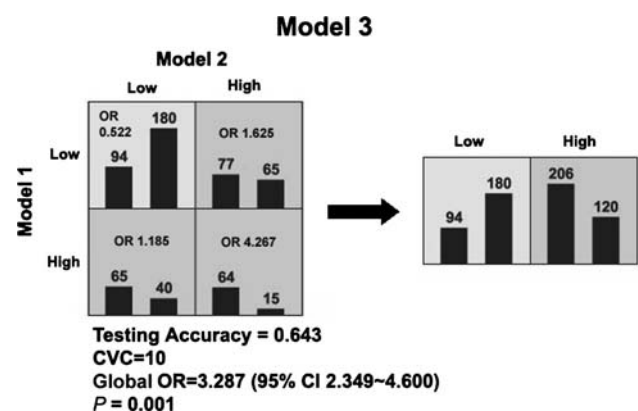
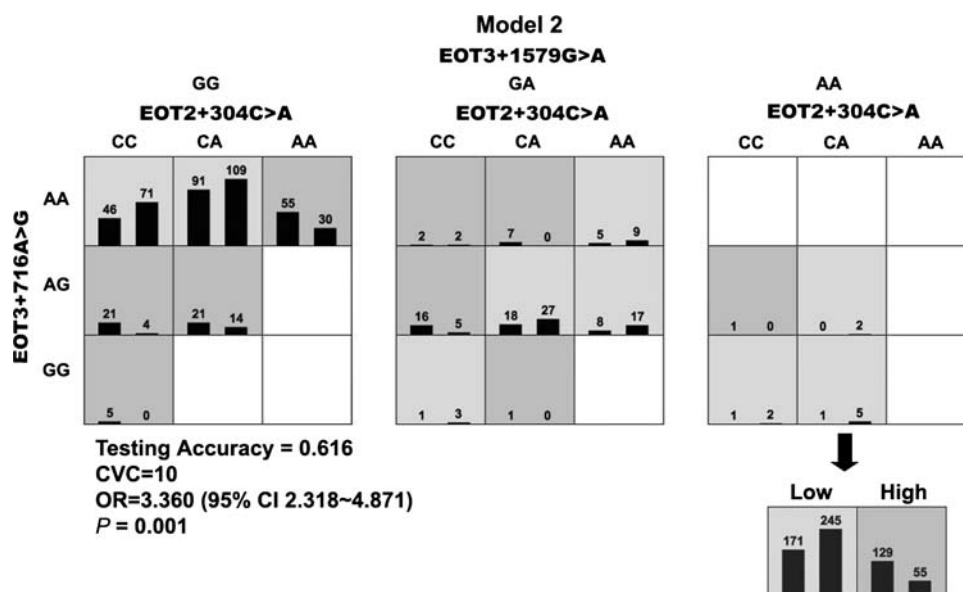


Fig. 3 Distribution of asthmatics (left bars) and controls (right bars) for the composite multifactor dimensionality reduction (MDR) model that combined the MDR variable for model 1 and the MDR variable for model 2. The number above the bar indicates the frequency of asthmatics or controls. High-risk genotype combinations are shaded dark grey and low-risk are shaded light grey. The final variable constructed by MDR is shown on the right. High high-risk group; Low low-risk group, CVC cross-validation consistency, OR odds ratio, CI confidence interval

The values of $|D'|$ for LD among SNPs in model 3 are presented in Table 4. $EOT3 + 1579G > A$ showed strong LD with $EOT2 + 1272A > G$ ($|D'| = 0.56$, $P < 0.001$) and with $EOT3 + 716A > G$ ($|D'| = 0.72$, $P < 0.001$) in asthmatics. $EOT + 304C > A$ also showed strong LD with $EOT3 + 716A > G$ ($|D'| = 0.51$, $P < 0.001$) and with $EOT3 + 1579G > A$ ($|D'| = 0.53$, $P < 0.001$) in controls.

Figure 4 summarizes the interaction information analysis (Moore et al. 2006). An interaction dendrogram is presented that highlights the amount of information gained about case-control status by putting two polymorphisms together using MDR. The interaction information analysis indicates that $EOT2 + 1272A > G$ and $EOT3 + 77C > T$

from model 1 have independent effects from one another and are independent from the SNPs in model 2.

Discussion

It is difficult to obtain clear statistical and biological evidence to determine the causes of complex polygenic diseases such as asthma. The MDR algorithm provides a nonparametric and genetic model-free alternative to logistic regression and is useful for detecting and characterizing nonlinear interactions among discrete genetic and environmental factors (Hahn et al. 2003; Moore and Ritchie 2004; Ritchie et al. 2001). This method uses data mining to identify new variables and so-called high-risk and low-risk groups from raw data. We applied MDR algorithms using a flexible multistep approach to identify genetic interactions that contribute to asthmatic phenotypes. We defined the main effect as an effect that any individual SNP associates with the subject's disease status. This main effect is difficult to replicate (Ioannidis 2007). When the single SNP effect is not present alone or is not strong enough, gene–gene interaction or gene–environmental interaction can be considered to characterize and identify the susceptibility of genes for disease risk. Epistasis is a description of the masking of the expression of one locus by alleles at another locus and quantitative differences among genotypes, calling any deviation from the additive combination of sing-locus genotypes (Wolf et al. 2000). Epistasis can occur in the absence of detectable independent effects of any one SNP.

An individual attribute was defined as a main effect contributing to disease when an SNP was determined to be significantly associated with the asthmatic phenotype by a

Table 4 Linkage disequilibrium among five single nucleotide polymorphisms (SNPs) of model 3

	<i>EOT2</i> + 304C > A	<i>EOT2</i> + 1272A > G	<i>EOT3</i> + 77C > T	<i>EOT3</i> + 716A > G	<i>EOT3</i> + 1579G > A
<i>EOT2</i> + 304C > A	~	0.07	0.24*	0.03	0.31*
<i>EOT2</i> + 1272A > G	0.12	~	0.12	0.13*	0.56*
<i>EOT3</i> + 77C > T	0.21*	0.27*	~	0.18*	0.05
<i>EOT3</i> + 716A > G	0.51*	0.21*	0.17	~	0.72*
<i>EOT3</i> + 1579G > A	0.53*	0.48*	0.22	0.82*	~

The values of Lewontin's $|D'|$ in asthma is in upright and those in control is in down-left area

* P value < 0.001 of D' from 0 by chi-square test

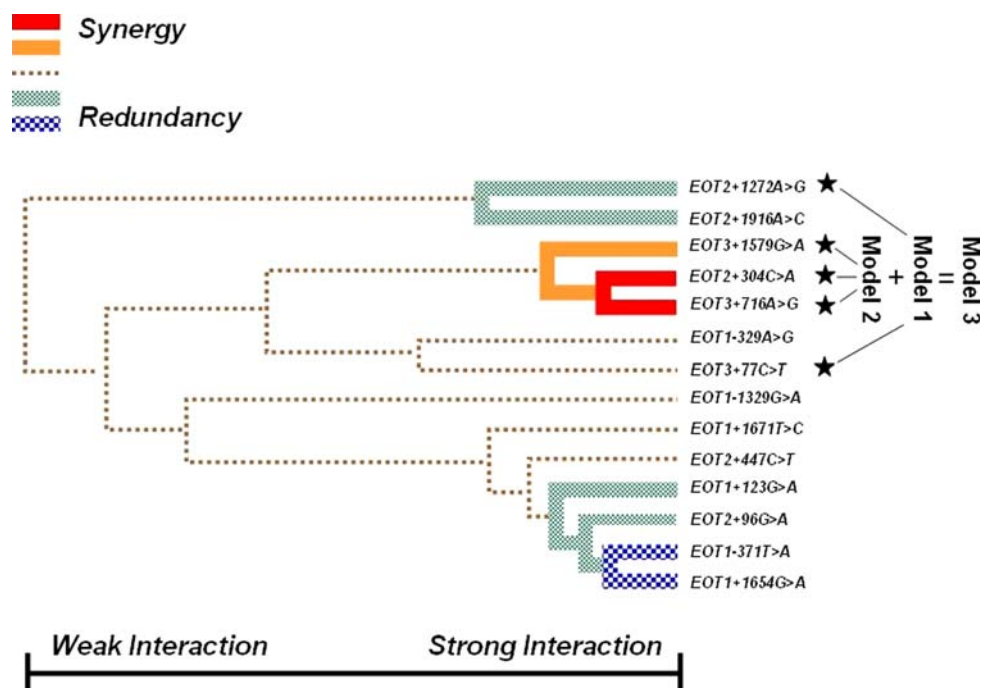


Fig. 4 Interactions dendrogram for the 14 polymorphisms modeled by MDR. A black or gray line (red or orange for online) suggests a positive information gain and can be interpreted as synergistic or non-additive relationship. Stripes lines suggest a loss of information and can be interpreted as redundancy or correlation. Dot lines indicate

independence or additivity. Independent effects of *EOT2* + 1272A > G and *EOT3* + 77C > T is comprised in model 1 and the strong synergistic effects of *EOT2* + 304C > A, *EOT3* + 716A > G and *EOT3* + 1579G > A are comprised in model 2. Five polymorphisms of model 1 and model 2 are comprised in hierarchical model 3

chi-square test ($P < 0.05$) (Moore et al. 2006). In main-effects analysis, of the 14 SNPs, *EOT2* + 1272A > G and *EOT3* + 77C > T were selected, and combination of these two made up model 1 (accuracy 0.597, CVC 10/10, OR 2.44, $P = 0.001$). Therefore, this combination accounted for one of the best models to predict asthma within the 14 SNPs of the *Eotaxin* genes examined. It was already reported that *EOT2* + 1272A > G was associated with asthma in three alternative models (dominant, recessive, and codominant) (Shin et al. 2003). We found that an *EOT2* + 1272A > G SNP contributes to one of the best two-loci models identified using MDR. The genotype combinations consisted of AA genotype of *EOT2* + 1272A > G and CT or TT genotypes of *EOT3* + 77C > T and were the high-risk group for asthma (Fig. 1).

Model 2 consisted of *EOT2* + 304C > A, *EOT3* + 716A > G, and *EOT3* + 1579G > A (accuracy 0.616, CV 10/10, OR 3.36, $P = 0.001$; Fig. 2). These were characterized as no-main-effect SNPs because they were not significantly associated with the disease phenotype in the first step of the analysis. Interaction dendrogram is used to visualize the nature of the dependencies. Interestingly, *EOT2* + 304C > A, *EOT3* + 716A > G, and *EOT3* + 1579G > A exhibited strong synergistic effects, suggesting nonadditive interactions (Fig. 4). We found that an *EOT2* + 1272A > G SNP contributes to one of the best two-loci models identified using MDR. These results indicate that our final model (model 3) comprised a pair of polymorphisms that had independent main effects and three polymorphisms with synergistic effects that were

independent of the two main-effects polymorphisms. Thus, model 3 was a hierarchical model consisting of a mix of main effects and interaction effects. We compared model 3 with the five-loci model gained from exhaustive MDR analysis, which was composed of $EOT1-329A > G$, $EOT2 + 304C > A$, $EOT2 + 447C > T$, $EOT3 + 77C > T$, and $EOT3 + 716A > G$. Accuracy was 0.672 and OR 3.173. However, CVC was only 5/10.

There are two primary reasons to perform separate MDR analyses on SNPs with and without significant independent main effects. First, an exhaustive analysis, by definition, looks at many more SNP combinations than a targeted approach, such as the one used here. An important concern in any combinatorial analysis is the risk of overfitting the data. MDR controls for overfitting in larger models through cross validation. Larger models that overfit the data are less likely to generalize to independent data and thus should have a lower testing accuracy. However, cross validation does not control over fitting within, for example, two-way or three-way models. The more two-way models that are exhaustively evaluated, the greater the chance of finding something interesting by chance. We reduced the total number of MDR evaluations by only exhaustively evaluating combinations among those SNPs with independent main effects and those without. This approach uses statistical knowledge about the nature of SNP univariate effects to reduce the total number of MDR evaluations. There are 91 possible pairwise combinations of SNPs that could be evaluated by an exhaustive search among the 14 eotaxin

SNPs. Reducing this to 12 based on their lack of a main effect reduced the total number of SNP pairs evaluated from 91 to 66. If we also consider all the three- and four-way combinations, the total number of SNP models evaluated by MDR drops nearly in half from 1,456 to 781. Second, organizing the analyses according to whether SNPs have a marginal effect significantly helps with the interpretation of the MDR models. The ability to disentangle the types of effects in an interaction model significantly helps with the understanding of that model.

To determine which genotype combinations contribute more to subject's status, every OR of each genotype combination of five SNPs in model 3 was calculated with OR-MDR (Table 3). The case that both model 1 and 2 were high-risk had a higher OR than only one was between two (4.267 vs. 1.625 or 1.18, respectively; Fig. 3). Of the 243 combinations, eight genotype combinations showed the OR > 1. The combination of C-A-CT-AG-G (in order of $EOT2 + 304C > A$, $EOT2 + 1272A > G$, $EOT3 + 77C > T$, $EOT3 + 716A > G$, and $EOT3 + 1579G > A$) was the best combination, and OR was 5.5 (Table 5). We performed the covariate analysis with GMDR. Age and gender were adjusted. The accuracy of model 3 without covariate adjustment was similar to that with adjustment (Table 6). But it was slightly different from the accuracy shown in Fig. 3, because GMDR uses score values instead of numbers of cases and controls to evaluate classification and prediction errors (Lou et al. 2007).

Table 5 The higher-risk genotype combinations and odds ratios (OR) in model 3

$EOT2 + 304C > A$	$EOT2 + 1272A > G$	$EOT3 + 77C > T$	$EOT3 + 716A > G$	$EOT3 + 1579G > A$	Frequency (case:control)	OR (95% CI)
A	A	C	A	G	28:21	1.24 (0.81–2.15)
CA	A	CT	A	G	37:33	1.12 (0.8–1.74)
A	A	CT	A	G	24:7	3.71 (1.77–8.43)
CA	A	T	A	G	22:5	3.4 (1.42–9.1)
A	A	T	A	G	4:2	1.5 (0.38–8.91)
C	A	C	AG	G	3:2	3.0 (0.75–14.75)
C	A	CT	AG	G	11:2	5.5 (1.38–24.6)
CA	A	CT	AG	G	15:4	2.5 (0.94–7.88)

The combinations that OR > 1 were presented among 243 combinations of five SNPs in model 3. All combinations are in the high-risk group. Confidence interval (CI) is usual asymptotic

Table 6 Comparison between model 3 with and without adjustment for age and gender as covariates

	Testing accuracy	Sign test (<i>P</i>)	CVC	OR (95% CI)
Without adjustment	0.6062	10 (0.001)	10/10	2.57 (0.57–11.65)
With adjustment	0.6075	10 (0.001)	10/10	2.59 (0.55–12.18)

Age is continuous variable and gender is discrete variable

CVC cross-validation consistency, OR odds ratio, CI confidence interval

Gene–gene interactions or epistasis can be defined biologically or statistically. Biological epistasis occurs when molecules such as deoxyribonucleic acid (DNA), RNA, proteins, and enzymes interact at the cellular level. Many of the important biological epistasis depends on specific locus to locus interactions at the individual level. In contrast, statistical epistasis can be defined as interindividual variation in DNA sequences detected at the population level. Statistical epistasis has been detected as some average estimates at the population level (Moore et al. 2006; Wolf et al. 2000). Recently, a concept of phenotypic landscape in hyperspace has emerged. This new theory is that the evolution of developmental interactions requires no simplifying assumptions about the number of underlying genetic and environmental factors. A landscape provides a concise summary of the patterns of genetic effects, gene–gene interactions, environmental effects, and gene–environmental interactions that produce the relationship between variation in underlying factors and phenotype (Rice 2002; Wolf 2002). Evolution occurs in a multidimensional genotypic space that cannot be justifiably reduced to a one- or two-dimensional representation (Wolf et al. 2000). Thus, analyses of gene–gene interactions are important for genetic and epidemiological studies of complex diseases such as asthma.

We previously reported that total IgE levels were associated with *EOT1* + *123G* > *A* in asthmatic subjects (Shin et al. 2003). Here, we included both asthmatic and normal subjects, but *EOT1* + *123G* > *A* was not found to be significantly associated with asthma in any epistatic model. We therefore hypothesized that the epistatic models reflected the effect of two or more SNPs that interacted with each other between asthmatics and controls. Therefore, the effect of *EOT1* + *123G* > *A* at the level of total IgE may be relatively weak in terms of its interactions with other SNPs compared with its effect alone.

Most biochemical analyses are unable to evaluate more than two factors at a time, and no biological method is universally accepted for performing experiments with multiple loci to reveal gene–gene interactions (Strohman 2002). Therefore, we are unable to provide biological evidence supporting the epistatic models predicted here. However, many studies have reported that eotaxin or *Eotaxin* genes are associated with other genes and cytokines. For example, eotaxin-2 is involved in airway inflammation and cooperates with interleukin-13 (IL-13) (Pompe et al. 2005). In addition, the IL-6-family cytokine oncostatin M (OSM) causes a dose-dependent increase in eotaxin release from murine fibroblasts (Langdon et al. 2003) and enhances IL-4 and IL-13-induced eotaxin-1 release from human airway smooth muscle (Faffe et al. 2005). Furthermore, interferon gamma (IFN- γ) enhances eotaxin expression in combination with tumor necrosis factor alpha (TNF- α)

mediated by a posttranscriptional mechanism (Matsukura et al. 2003).

Another limitation of this study is lack of replication. We did not provide the replication in an independent population to study subjects. Therefore, it might be required to test of usefulness of these models to other Korean populations or other races.

These statistical interaction models are considered useful for detecting groups that are genetically at high risk of developing asthma. However, although the models can be used to identify interactions between genetic variants that confer a risk for asthma, this approach cannot be translated into specific prevention and treatment strategies without interpreting the results under the biological context of asthma. Thus, further biological studies are needed before we can apply these models.

In conclusion, we developed three epistatic models for asthma by examining gene–gene interactions among polymorphisms within the *Eotaxin* gene family using a multistep approach with the MDR method. These models suggest that interactions within the *Eotaxin* gene family likely contribute to the development of asthma. Although the models are limited to determining statistical interactions within a population, they may be useful for identifying groups at high risk of developing asthma.

Acknowledgments This work was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (01-PJ3-PG6-01GN04-003) and Soonchunhyang University Research Fund (20050000). The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, see: <http://www.textcheck.com/cgi-bin/certificate.cgi?id=TVT7r8>

References

- Andrew AS, Nelson HH, Kelsey KT, Moore JH, Meng AC, Casella DP, Tosteson TD, Schned AR, Karagas MR (2006) Concordance of multiple analytical approaches demonstrates a complex relationship between DNA repair gene SNPs, smoking and bladder cancer susceptibility. *Carcinogenesis* 27:1030–1037
- Berkman N, Ohnonsa S, Chung FK, Breuer R (2001) Eotaxin-3 but not eotaxin gene expression is upregulated in asthmatics 24 hours after allergen challenge. *Am J Respir Cell Mol Biol* 24:682–687
- Brown JR, Kleimberg J, Marini M, Sun G, Bellini A, Mattoli S (1998) Kinetics of eotaxin expression and its relationship to eosinophil accumulation and activation in bronchial biopsies and bronchoalveolar lavage (BAL) of asthmatic patients after allergen inhalation. *Clin Exp Immunol* 114:137–146
- Carlson CS, Eberle MA, Kruglyak L, Nickerson DA (2004) Mapping complex disease loci in whole-genome association studies. *Nature* 429:446–452
- Chung Y, Lee SY, Elston RC, Park T (2007) Odds ratio based multifactor-dimensionality reduction method for detecting gene–gene interactions. *Bioinformatics* 23:71–76
- Davies DE, Wicks J, Powell RM, Puddicombe SM, Holgate ST (2003) Airway remodeling in asthma: new insights. *J Allergy Clin Immunol* 111:215–225

- Faffe DS, Flynt L, Mellema M, Moore PE, Silverman ES, Subramaniam V, Jones MR, Mizgerd JP, Whitherhead T, Imrich A, Panettieri RA, Shore SA (2005) Oncostatin M causes eotaxin-1 release from airway smooth muscle: synergy with IL-4 and IL-13. *J Allergy Clin Immunol* 115:514–520
- Ferris BG (1978) Epidemiology standardization project (American Thoracic Society). *Am Rev Respir Dis* 118:1–120
- Hahn LW, Moore JH (2004) Ideal discrimination of discrete clinical endpoints using multilocus genotypes. In *Silico Biol* 4:183–194
- Hahn LW, Ritchie MD, Moore JH (2003) Multifactor dimensionality reduction software for detecting gene–gene and gene–environmental interactions. *Bioinformatics* 19:372–382
- Hedric PW (1978) Gametic disequilibrium measures: proceed with caution. *Genetics* 117:331–341
- Ioannidis JP (2007) Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* 64:203–213
- Jakulin A, Bratko I, Smrke D, Demsar J, Zupan B (2003) Attribute interactions in medical data analysis. *Lect Notes Artif Intell* 2780:229–238
- Langdon C, Kerr C, Tong L, Richards CD (2003) Oncostatin m regulates eotaxin expression in fibroblasts and eosinophilic inflammation in C57BL/6 mice. *J Immunol* 170:548–555
- Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, Li MD (2007) A generalized combinatorial approach for detecting gene–by–gene and gene–by–environment interactions with application to nicotine dependence. *Am J Hum Genet* 80:1125–1137
- Matsukura S, Kokubu F, Kuga H, Kawaguchi M, Ieki K, Odaka M, Suzuki S, Watanabe S, Takeuchi H, Adachi M, Stellato C, Schleimer RP (2003) Differential regulation of eotaxin expression by IFN- γ in airway epithelial cells. *J Allergy Clin Immunol* 111:1337–1344
- Min JW, Lee JH, Park CS, Chang HS, Rhim TY, Park SW, Jang AS, Shin HD (2005) Association of eotaxin-2 gene polymorphisms with plasma eotaxin-2 concentration. *J Hum Genet* 50:118–123
- Moore JH, Ritchie MD (2004) The challenges of whole-genome approaches to common disease. *J Am Med Assoc* 291:1642–1643
- Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC (2006) A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol* 241:252–261
- Moore JH, Barney N, Tsai CT, Chiang FT, Gui J, White BC (2007) Symbolic modeling of epistasis. *Hum Hered* 63:120–133
- Papadopoulos NG, Papai A, Meyer J, Stanciu LA, Salvi S, Holgate ST, Johnston SL (2001) Rhinovirus infection up-regulates eotaxin and eotaxin-2 expression in bronchial epithelial cells. *Clin Exp Allergy* 31:1060–1066
- Pompe SM, Fulkerson PC, Blanchard C, Akei HS, Niklaidis NM, Zimmermann N, Molkentin J, Rothenberg ME (2005) Identification of a cooperative mechanism involving interleukin-13 and eotaxin-2 in experimental allergic lung inflammation. *J Biol Chem* 280:13952–13961
- Rankin SM, Conroy DM, Williams TJ (2000) Eotaxin and eosinophil recruitment: implications for human disease. *Mol Med Today* 6:20–27
- Rice SH (2002) A general population genetic theory for the evolution of developmental interactions. *Proc Natl Acad Sci USA* 99:15518–15523
- Ritchie MD, Hahn LW, Roodi N, Bailey R, Dupont WD, Parl FF, Moore JH (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 69:138–147
- Ritchie MD, Hahn LW, Moore JH (2003) Power of multifactor dimensionality reduction for detecting gene–gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol* 24:150–157
- Robert OC, John LH, Charles I, Neil RM, Karen ZV, Robert AW (1987) Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. *Am Rev Respir Dis* 136:225–244
- Robert OC, Richard C, Allan LC, Paul LE, John LH, Charles GI, Neil RM, Roy TM, Jask SW (1999) Guidelines for methacholine and exercise challenge testing-1999. *Am J Respir Crit Care Med* 161:309–329
- Shin HD, Kim LH, Park BL, Jung JH, Kim JY, Chung IY, Kim JS, Lee JH, Chung SH, Kim YH, Park HS, Choi JH, Lee YM, Park SW, Choi BW, Hong SJ, Park CS (2003) Association of Eotaxin gene family with asthma and serum total IgE. *Hum Mol Genet* 12:1279–1285
- Steinke JW, Rich SS, Borish L (2008) 5. Genetics of allergic disease. *J Allergy Clin Immunol* 121:S384–S387
- Stephan M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Strohma R (2002) Maneuvering in the complex path from genotype to phenotype. *Science* 296:701–703
- Wolf JB (2002) The geometry of phenotypic evolution in developmental hyperspace. *Proc Natl Acad Sci USA* 99:15849–15851
- Wolf JB, Brodie EDIII, Wade MJ (2000) Epistasis and the evolutionary process. In: Phillips PC, Otto SP, Whitlock MC (eds) *Beyond the average*. Oxford, New York, pp 21–22
- Ying S, Robinson DS, Meng Q, Barata LT, McEuen AR, Buckley MG, Walls AF, Askenase PW, Kay AB (1999) C–C chemokines in allergen-induced late-phase cutaneous response in atopic subjects: association of eotaxin with early 6-hour eosinophils and of eotaxin-2 and monocyte chemoattractant protein-4 with the later 24-hour tissue eosinophilia, and relationship to basophils and other C–C chemokines (monocyte chemoattractant protein-3 and RANTES). *J Immunol* 163:3976–3984