Original Article
Allergy Asthma Immunol Res. 2011 October;3(4):265-272.
http://dx.doi.org/10.4168/aair.2011.3.4.265
pISSN 2092-7355 • eISSN 2092-7363

AAIR
Allergy, Asthma & Immunology Research

Asthma-Predictive Genetic Markers in Gene Expression Profiling of Peripheral Blood Mononuclear Cells

Seung Woo Shin,[1] Tae Jeong Oh,[2] Se-Min Park,[1] Jong Sook Park,[1,3] An Soo Jang,[1,3] Sung Woo Park,[1,3] Soo Taek Uh,[4] Sungwhan An,[2] Choon-Sik Park[1,3]*

[1]*Genome Research Center for Allergy and Respiratory Disease, Soonchunhyang University Bucheon Hospital, Bucheon, Korea*
[2]*Genomictree Inc., Daejeon, Korea*
[3]*Division of Allergy and Respiratory Medicine, Department of Internal Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Korea*
[4]*Division of Allergy and Respiratory Medicine, Soonchunhyang University Seoul Hospital, Seoul, Korea*

Purpose: We sought to identify asthma-related genes and to examine the potential of these genes to predict asthma, based on expression levels. **Methods:** The subjects were 42 asthmatics and 10 normal healthy controls. PBMC RNA was subjected to microarray analysis using a 35K array; *t*-tests were used to identify genes that were expressed differentially between the two groups. A multiple logistic regression analysis was applied to the differentially expressed genes, and area under the curve (AUC) values from receiver operating characteristic (ROC) curves were obtained. **Results:** In total, 170 genes were selected using the following criteria: $P \leq 0.001$ and ≥ 2-fold change. Among these genes, 57 were up-regulated and 113 were down-regulated in asthmatics versus normal controls. A multiple logistic regression analysis was done using more stringent criteria ($P \leq 0.001$ and ≥ 5-fold change), and eight genes were selected as candidate asthma biomarkers. Using these genes, 255 models (2^8-1) were generated. Among them, only 85 showed $P \leq 0.05$ by multiple logistic regression analysis. Based on the AUCs from ROC curves for the 85 models, we found that the best model consisted of the genes *MEPE, MLSTD1*, and *TRIM37*. The model showed 0.9928 of the AUC with 98% sensitivity and 80% specificity. **Conclusions:** *MEPE, MLSTD1*, and *TRIM37* may be useful biomarkers for asthma.

Key Words: Asthma; gene expression profiling; PBMC; ROC

INTRODUCTION

Asthma is a common and heterogeneous respiratory disease characterized by intermittent airway obstruction and respiratory symptoms that are related to chronic airway inflammation and remodeling.[1] Pathological features of airway remodeling include goblet cell hyperplasia, subepithelial fibrosis, collagen deposition, mucosal gland hyperplasia, smooth muscle hypertrophy, and changes in the extracellular matrix.[1]

Inflammation and remodeling are the main causes of airway hyperresponsiveness and chronic airway obstruction, characteristic features of asthma. These changes are attributed to the altered expression of genes associated with transcriptional pathways, inflammatory processes, apoptosis, and cell proliferation.[2]

However, the complex nature of the asthma phenotype, together with genetic heterogeneity and environmental influences, has made it difficult to uncover genetic aspects of this common condition.[3] Thus, the search for asthma-specific genes is an important issue because such genes may promote the early detection, positive prognosis, and treatment of the disease. To this end, large-scale, high-throughput, whole-genome studies are needed to understand the genomic contribution to asthma.

Although many approaches have been used to investigate relationships between diseases and genes, the high-throughput microarray is one of the most important. Microarray technology allows the monitoring of gene expression on a genomic scale. It

Correspondence to: Choon-Sik Park, MD, PhD, Division of Allergy and Respiratory Medicine, Department of Internal Medicine, Soonchunhyang University Bucheon Hospital, 1174 Jung-dong, Wonmi-gu, Bucheon 420-021, Korea.
Tel: +82-32-621-5105; Fax: +82-32-621-5023; E-mail: mdcspark@unitel.co.kr

Co-correspondence to: Sungwhan An, PhD, Genomictree Inc., 829 Tamnip-dong, Yuseong-gu, Daejeon 305-510, Korea.
Tel: +82-42-861-4550; Fax: +82-42-861-4552; E-mail: genomictree1@korea.com
Received: October 11, 2010; Accepted: March 4, 2011

- Seung Woo Shin and Tae Jeong Oh contribute equally as the first author.
- There are no financial or other issues that might lead to conflict of interest.

also enables the classification of genes causing various diseases and for a diagnosis to be made at the genetic level.[4,5]

Several types of tissues have been used in human microarray studies of asthma, including airway epithelial cells and airway inflammatory cells from patients and healthy subjects.[6-10] Nasal mucosal cells are a good choice for genomic studies.[11]

Although an analysis of gene expression in the airways can provide important information, obtaining airway tissues, including airway epithelial cells, using bronchoalveolar lavage (BAL) or brushing is difficult, especially in severe asthma[12]; thus, peripheral blood mononuclear cells (PBMCs) have been used as an alternative.[13-16] PBMCs contain lymphocytes, monocytes, and dendritic cells, which are recognized sources of allergic response mediators.

Over the past decade, using gene expression profile analysis, the diagnosis of multifactorial complex diseases and their subphenotypes have been attempted. Although the diagnostic value of the expression of a single gene is regarded as minimal, multi-gene analyses can increase the predictability, and possibly lead to usable biomarkers.

In the present study, we identified candidate genes affecting asthma development using an mRNA expression chip. Additionally, we evaluated the diagnostic value of differential gene expression levels in the discrimination of asthmatics from normal controls through an additional analysis of the area under the curve (AUC) values from receiver operating characteristic (ROC) curves, sensitivity, and specificity.[17] Finally, the selected genes were analyzed according to asthma severity.

MATERIALS AND METHODS

Study subjects

The subjects were recruited from the Genome Research Center for Allergy and Respiratory Diseases at Soonchunhyang University Hospital. All subjects were Korean. The protocol was approved by the local ethics committee of Soonchunhyang University Hospital. Written informed consent was obtained from all subjects.

All patients met the definition of asthma given in the Global Initiative for Asthma (GINA) guidelines.[1] Each patient showed airway reversibility, as documented by a positive bronchodilator response of a greater than 15% increase in FEV1 and/or airway hyperreactivity to less than 10 mg/mL methacholine. Normal controls were recruited from hospital personnel; they gave negative answers to a screening questionnaire for respiratory symptoms and had an FEV1 $\geq$75% of the predicted value, a PC20 methacholine $\geq$10 mg/mL, and a normal chest X-ray.

Subjects were assessed using a skin-prick test for 24 common inhaled allergens, including house dust mites, *Alternaria*, *Aspergillus*, pollen, dogs, cats, and cockroaches. Atopy was defined as having a wheal reaction with the allergen extract equal to or greater than that with histamine (1 mg/mL) or 3 mm in diameter. Total IgE was measured using the UniCAP system (Pharmacia Diagnostics, Uppsala, Sweden). Peripheral venous blood was obtained when the subjects were in a stable state. No patient had an exacerbation or systemic steroid treatment within 6 weeks of the study. Clinical profiles of the asthma patients and control subjects are summarized in Table 1.

cDNA synthesis and microarray hybridization

A human oligonucleotide microarrays (Genomictree Inc., Daejeon, Korea) containing 34,560 oligonucleotide probes, representing human genes and gene transcripts, were used for gene expression analysis. The 34,560 oligoncleotide probes were purchased from (Qiagen, Germantown, MD, USA). The synthesis of target cDNA probes and hybridization were performed according to previously described.[8] Briefly, each 50 μg of total RNA was mixed with 2 μg of oligo-24N (dT) (GenoTech, Daejeon, Korea) in 15.4 μL of RNase free water and incubated at 65°C for 10 minutes. After incubation, the single stranded cDNA was

Table 1. Clinical profiles for asthma patients and normal controls, mild asthma and moderate to severe asthma subjects

	Normal controls	Asthmatics	*P* value	Mild asthma	Moderate to severe asthma	*P* value
Number	10	42		30	12	
Age (median [range], yr)	59 (48-78)	47 (21-70)	0.111	55 (32-78)	57 (28-77)	0.989
Male (%)	20.0	38.1	0.462	36.6	41.6	1.000
Smoking (NS/SM/ES)	8/1/1	29/7/6	1.000	23/4/2	6/3/3	0.332
Atopy (%)	30.0	54.7	0.291	40.0	96.6	**0.004**
FVC, % predicted	98.0±14.2	87.0±14.4	**0.030**	92.6±9.1	73.1±15.9	**0.001**
FEV1, % predicted	108.0±16.0	88.1±18.3	**0.003**	97.3±11.4	65.0±9.5	**<0.001**
PC20 methacholine (mg/mL)	25.0±0.0	9.0±10.4	**<0.001**	10.2±10.5	5.5±10.2	**0.016**
Body mass index (kg/m^2)	25.7±2.9	25.0±3.9	0.519	25.1±4.1	24.6±3.4	0.811
Log (total IgE concentration (IU/mL))	1.6±0.5	2.1±0.5	1.000	2.0±0.6	2.4±0.4	**0.018**

Mild asthma and moderate to severe asthma are separated by FEV1 value of 80% predicted and *P* values were acquired by *t*-test or Fisher's exact test.
NS, non-smoker; SM, smoker; ES, ex-smoker.

synthesized in the presence of Cy3-dUTP or Cy5-dUTP (1 mM each, NEN Life Science Products, Boston, MA, USA) at 42°C for 2 hours. Total RNA from test samples and common reference RNA pooled from 10 normal samples were labeled with Cy5 and Cy3, respectively. Both of Cy3 and Cy5-labeled cDNA were purified using PCR purification kit (Qiagen) as recommended by manufacturer. The purified cDNA was resuspended in 100 μL of hybridization solution containing 5X SSC, 0.1% SDS, 30% formamide, 20 μg of Human Cot-1 DNA, 20 μg of poly A RNA and 20 μg of Yeast tRNA (Invitrogen, Grand Island, NY, USA). The hybridization mixtures were heated at 100°C for 2-3 minutes and directly pipetted onto microarrays. The arrays hybridized at 42°C for 12-16 hours in the humidified hybridization chamber (Genomictree Inc.). The hybridized miroarrays were washed with 2×SSC/0.1% SDS for 5 minutes, 0.1×SSC/0.1% SDS for 10 minutes, and 0.1×SSC for 2 minutes two times. The washed microarrays were immediately dried using the microarray centrifuge (Genomictree Inc.).

Statistical methods

For microarray analysis, background-corrected values for each probe on the BeadChip array were extracted using BeadStudio (Illumina, San Diego, CA, USA); *P* values were computed from the background model characterized by the chance that the target sequence signal was distinguishable from the negative controls on the same chip. Normalization was performed for the extracted values. The statistical significance of the microarray data was calculated by *t*-tests using 'R' (ver. 2.8.1; http://www.r-project.org/). If the distribution of the mRNA expression levels showed normality and equality in the variances in each group (i.e., asthma and normal controls), a classic two-sample *t*-test was used. If the distribution of mRNA expression levels showed normality and inequality in the variances for each group, the Welch two-sample *t*-test was used. If the distribution of mRNA expression levels did not show normality in each group, then the Mann-Whitney U-test was applied. For the diagnostic gene expression values used to discriminate the asthmatics from the normal controls, we selected candidate genes with $P<0.001$ and a fold-change ≥ 5 between the two groups. Fold changes in gene expression were obtained by dividing the mean value for the asthmatics by that for the normal controls. After filtering by *P* value and fold-change, a power set was made using the candidate genes, and a multiple logistic regression analysis was performed with elements in the power set. ROC curves for all models were obtained and AUCs were calculated[19,20] to select disease marker genes.

RESULTS

Comparison of gene expression between the asthmatics and normal controls

To identify genes that may be related to asthma, we applied a high-throughput gene expression microarray consisting of 15,054 featured genes on RNA samples, which were obtained from normal controls (n=10) and asthma patients (n=42). All values in the microarray cell were normalized and used for feature selection. The overall approach is depicted in Fig. 1.

To evaluate overall differences in gene expression levels in PBMCs between asthmatics and normal controls, we calculated gene expression as shown on a volcano plot. To identify differentially expressed genes between the asthmatics and normal controls, we applied two types of *t*-tests and the Mann-Whitney U-test to the difference in mean expression level between the two groups. We selected genes having a *P* value<0.001 and a change of 2-fold or greater. Volcano plots of significance against the fold-change values for each gene in the PBMCs revealed that the expression levels were quite different between the asthmatics and normal controls (Fig. 2A). Using the criteria of $P\leq 0.001$ and a fold change ≥ 2, we identified 57 genes showing a significant increase in expression and 113 genes showing a significant decrease in expression in the asthmatics compared with the normal controls (see Table S1 in additional file 1). A heat map of these differentially expressed genes (n=170) revealed distinctive RNA expression profiles for the PBMCs (Fig. 2B). The up- and down-regulated genes were well-classified by this hierarchical clustering method.

Because 170 genes would be too many to handle as biomarkers, we used the more stringent criteria of $P\leq 0.001$ and a fold change ≥ 5 to the differences in gene expression between the asthmatics and normal controls. Consequently, we identified eight genes (Fig. 3).

Development of genetic biomarkers for the diagnosis of bronchial asthma

The genes that met these criteria are listed in Table 2. The expression of *ZFP161* and *NOX5* was increased, while that of *LMAN1*, *MEPE*, *MLSTD1*, *TRIM37*, *KNS2*, and *CCT5* was de-

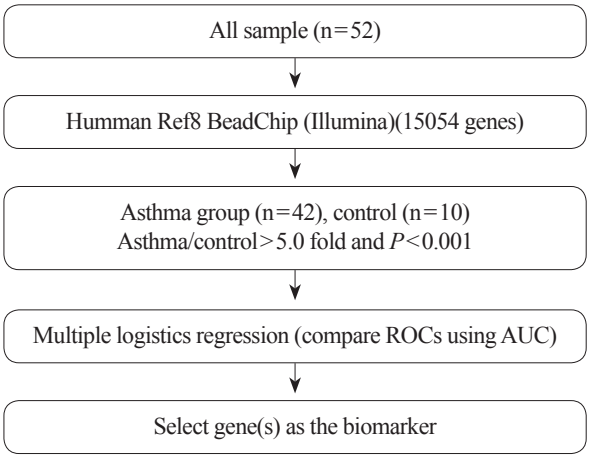

Fig. 1. Gene expression profiling strategy and general workflow.

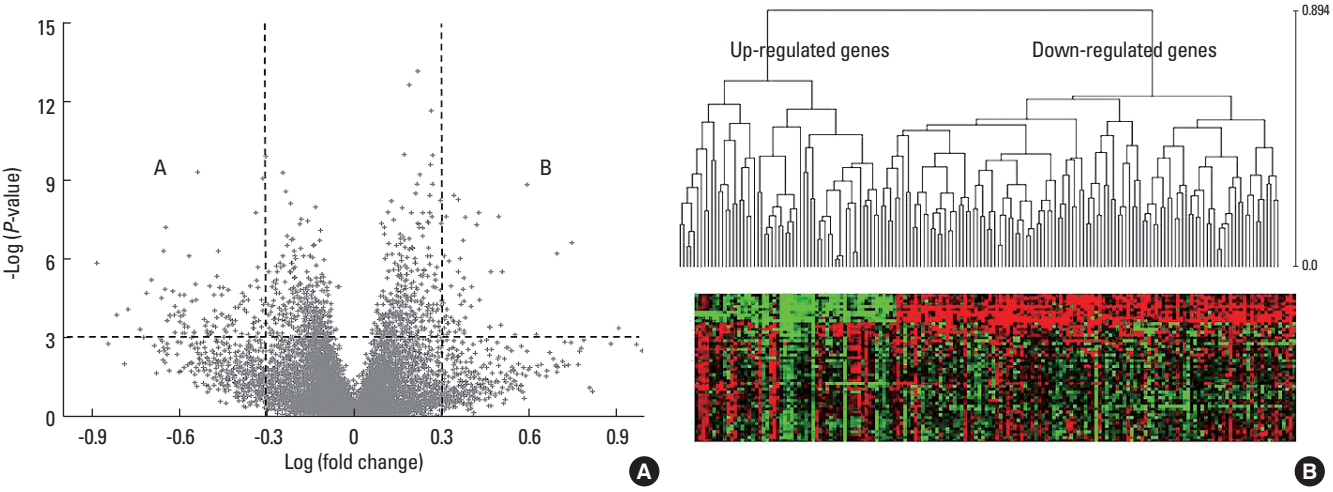

Fig. 2. (A) Volcano graph. The x-axis represents the logarithm of the fold-change value, while the y-axis represents the negative logarithm of the *P* value. A and B denote areas satisfying the following criteria: ≥2-fold change and $P \leq 0.001$. (B) Hierarchical clustering graph and heat map.

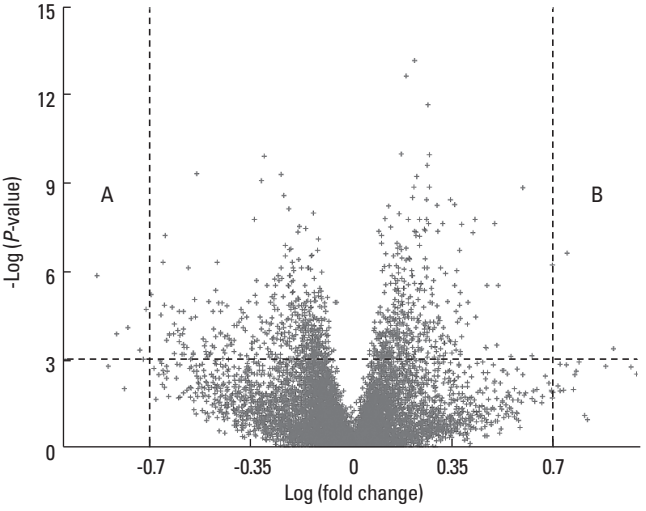

Fig. 3. Volcano graph. The x-axis represents the logarithm of the fold-change value, while the y-axis represents the negative logarithm of the *P* value. A and B denote areas satisfying the following criteria: ≥5-fold change and $P \leq 0.001$.

creased in the PBMCs of the asthmatics versus the normal controls.

Selection of the best discriminative model using multiple logistic regression analysis and ROC curve analysis

Using the eight genes as candidate biomarkers to discriminate between asthmatics and normal controls, we performed a multiple logistic regression analysis for all elements of the power set and measured the AUCs from ROC curves. First, we made 255 (2^8-1) models for all elements of the power set and measured the *P* values of the variable for each model. We separated the 255 models into eight groups (Groups 1-8). Group n indicates models made of n genes for multiple logistic regression analysis. Among the 255 models, only 85 showed $P \leq 0.05$. All data for the AUC and *P* values are presented in Supplementary Table S2 (see additional file 2). As the number of genes increased, the *P* value decreased (Fig. 4). Only models comprising fewer than three genes showed significant *P* values (i.e., <0.05).

Table 2. List of genes meeting the criteria of $P \leq 0.001$ and a fold change of ≥5-fold

Gene	*P* value	Fold-change*	Name
ZFP161	0.0004342765	8.09	Zinc finger protein 161 homolog
NOX5	0.0000002402	5.60	NADPH oxidase, EF-hand calcium binding domain
LMAN1	0.0000198657	0.19	Lectin, mannose-binding, 1
MEPE	0.0009997948	0.18	Matrix extracellular phosphoglycoprotein
MLSTD1	0.0004841020	0.18	Fatty acyl CoA reductase 2
TRIM37	0.0000814375	0.16	Tripartite motif-containing 37
KNS2	0.0001356000	0.15	Kinesin light chain 1
CCT5	0.0000014420	0.13	Chaperonin containing TCP1, subunit 5

*Fold change (asthma/normal control).

The AUC and ROC were calculated for Groups 1, 2, and 3. Among them, single genes or each combination of two or three genes having the five highest AUC values are presented in Fig. 5A. The best AUC value (0.9928) was observed for a combination of three genes, *MEPE*, *MLSTD1*, and *TRIM37* showing asymptotic *P* values: 0.000001, asymptotic 95% confidence interval (lower bound: 0.977, upper bound: 1). The ROC curves and AUC values for the other genes and combinations are presented in Supplementary Figure S1 (see additional file 4).

Discriminating power of the combination of MEPE, MLSTD1, and TRIM37 between asthmatics and normal controls

To evaluate the discriminating power of the combination of

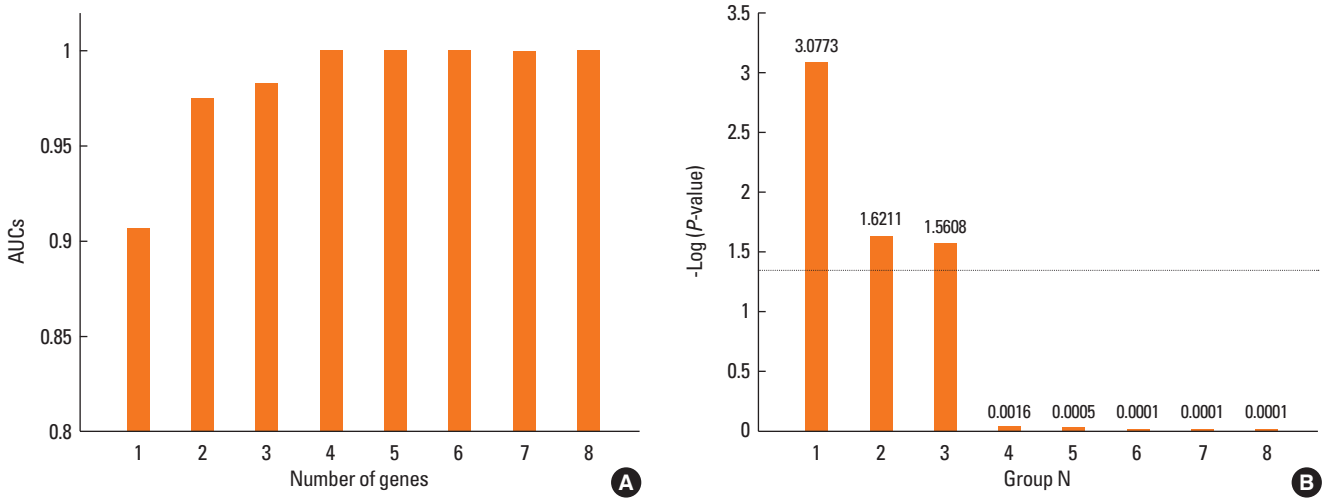

Fig. 4. (A) Distribution of the average AUCs for the top five in group n. (B) Distribution of the average log of the *P* values for the top five in group n. The dashed line indicates the cut-off value ($P<0.05$).

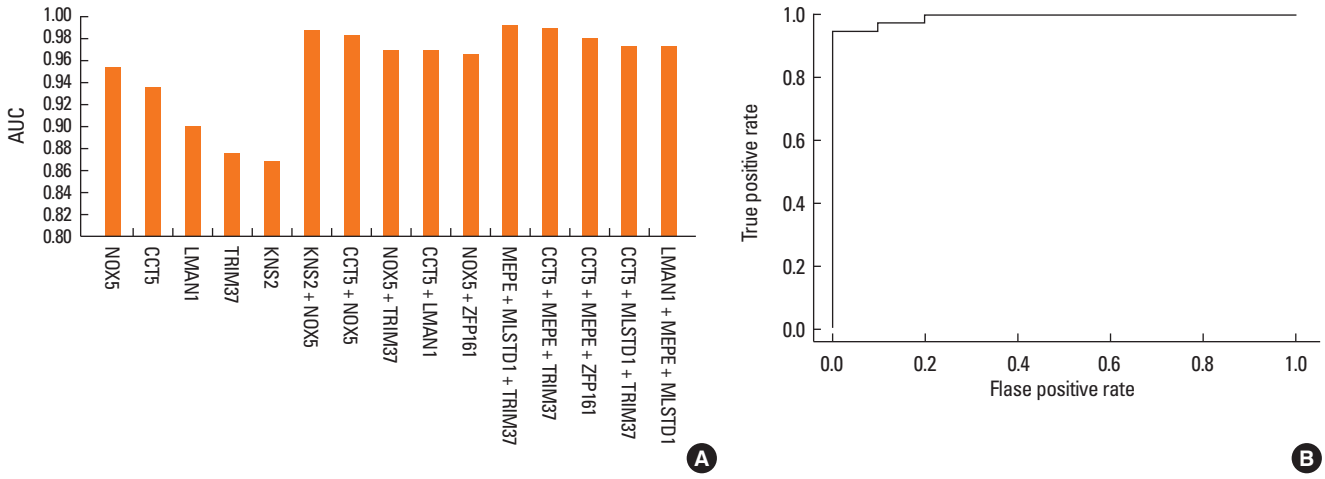

Fig. 5. (A) Values of the AUCs for each of the eight genes, two-gene combinations, and three-gene combinations. (B) ROC curve of the best model, consisting of *MEPE*, *MLSTD1*, and *TRIM37* (*P* value: 0.000001; asymptotic 95% confidence interval lower bound: 0.977, upper bound: 1; AUC: 0.9928).

MEPE, *MLSTD1*, and *TRIM37* between asthmatics and normal controls, we calculated the sensitivity and specificity using a contingency table of 42 asthmatics and 10 normal controls. As shown in Table 3, the sensitivity and selectivity were 0.98 and 0.80, respectively, while the accuracy was 0.942. To evaluate the diagnostic accuracy of the three-gene combination in an independent data set, we applied three-fold cross-validation (CV). The average results of three-fold CV for sensitivity, specificity, and accuracy were 1, 1, and 1, respectively.

Table 3. Contingency table for the best model

Prediction outcome	Actual value		Sensitivity	Specificity
	Asthma	Normal		
Asthma	41	2	0.98	0.80
Normal	1	8		

*Fold change (asthma/normal control).

Analysis of the selected genes according to asthma severity

We divided the asthma group into mild asthma and moderate-to-severe asthma according to their FEV1% (threshold: 80), and we analyzed the genes for each paired group as follows: normal controls versus mild asthma, normal controls versus moderate-to-severe asthma, and mild asthma versus moderate-to-severe asthma. By this analysis, we could predict genes involved in asthma development. Our complete results are presented in Supplementary Table S3 (see additional file 3). Table 4 shows the results for the eight genes.

Based on our results, *TRIM37* is predictive for the occurrence of asthma, while *NOX5*, *LMAN1*, *KNS2*, and *MLSTD1* are pre-

Table 4. The selected 8 genes analysis according to asthma severity

Group	Num.	Selected 8 genes
Normal controls vs. moderate to severe asthma	Up: 22	
	Down: 31	TRIM37, CCT5
Normal controls vs. mild asthma	Up: 58	NOX5
	Down: 106	TRIM37, LMAN1, KNS2, MLSTD1
Mild asthma vs. moderate to severe asthma	Up: 0	
	Down: 3	

dictive for mild asthma. *CCT5* is predictive for moderate-to-severe asthma.

DISCUSSION

We identified genes related to asthma using a microarray analysis of PBMCs. Using these data, we found that a combination of three genes, *MEPE*, *MLSTD1*, and *TRIM37*, showed the best discriminating power between asthmatics and normal controls. To our knowledge, this is the first reported study to select disease markers using ROC analyses of PBMCs. In our microarray analysis, when a *P* value threshold of <0.001 and a fold change of ≥2 were applied, a total 170 genes were selected. Of these genes, 57 were up-regulated and 113 were down-regulated in PBMCs. The genes included inflammatory and immune response genes, such as *NOX5*, *MALT1*, *TNFRSF10C*, *GRK5*, *CXCL3*, *RELA*, *CD40*, *ABR*, *RELB*, *REBB2*, *PGLYRP1*, *CD82*, *RPE*, *CFTR*, *KITLG*, and *IKBKG*, which would be expected to be expressed differentially in asthmatics versus normal controls. These findings suggested that our gene chip study was performed adequately.

There have been several previous reports of gene expression analyses in asthma. Previous gene array studies have shown that *CLCA1*[21,22], *SerpineB2*[7,21,22], *MUC5AC*[6,21], *AGR2*[21], *CPA3*[6,22], and *tryptase*[6,22] are over-expressed in bronchial epithelial cells or the airway mucosa of asthmatics. Such studies have used tissues obtained from bronchial epithelial cells or the mucosa of healthy controls and subjects with allergic asthma. Genes such as *NFκB* p65 subunit (*RELA*) and scinderin (*SCIN*) were expressed at similar levels in our and these other studies, suggesting that the expression of certain genes in the airways is reflected in the peripheral blood cells of similar subjects. However, many genes were expressed differently in the present study than in previous reports. This may be attributable to the different tissue sources used, such as airways and peripheral blood cells. For mechanistic or pathogenesis studies, airway tissue, such as bronchial mucosa, would be the best target. However, obtaining airway tissues or airway epithelial cells using BAL or brushing is not convenient for asthmatics, especially those with severe asthma.[12] Thus, PBMCs were used for the development of diagnostic gene markers in the present study.

First, we narrowed the number of candidate gene markers from the 170 genes that were initially selected, because 170 would be too many to handle. By applying more stringent criteria ($P<0.001$ and ≥5-fold change), eight candidate genes were selected for modeling. Using these genes, a multiple logistic regression analysis was performed to identify disease marker genes. Because we did not know the complex dependencies among the eight selected genes,[23] we made all 255 (2^8-1) possible models using the genes and measured the *P* values to examine their validity. ROC curves and the AUC were measured to assess the predictability of the gene markers for asthma.

Of the 255 models, only 85 showed a *P* value <0.05. Furthermore, the number of genes analyzed in combination was inversely correlated with the *P* values. Those models made of more than three genes showed insignificant statistics when a *P*-value <0.05 was applied, so we compared the AUC values only for those genes or combinations of genes having *P* values <0.05. After performing a multiple logistic regression analysis, the best model (*MEPE*, *MLSTD1*, *TRIM37*) and the second best model (*CCT5*, *MEPE*, *TRIM37*) were found to show almost perfect classification. The AUC values were 0.9928 and 0.9904, respectively. Considering that an AUC value of '1' indicates perfect classification, 0.9928 and 0.9904 indicate very high predictability for the markers. The sensitivity and selectivity for the best model were 0.98 and 0.80, respectively, while the accuracy was 0.942. Thus, these models are sufficient to discriminate between asthmatics and normal controls. For three-fold CV, the results showed perfect classification. This means that the genes are very effective for independent data. Thus, the genes may be used as disease markers for asthma in PBMCs. The approach of selecting gene markers using ROC curves has been applied to select biomarkers in other diseases, including gastric cancer, knee osteoarthritis, and lung cancer.[24-26]

In previous studies, the most up-regulated gene and/or down-regulated gene have been selected and subjected to logistic regression analyses and measurements of the AUC from ROC curves. However, such models might not have the best discriminative performance as disease markers. In contrast, in the current study, we selected several candidate genes that satisfied our criteria. We applied a multiple logistic regression analysis for all possible (2^8-1) models and selected the one with the highest AUC. In our experiment, a model consisting of three genes showed the best performance.

We also analyzed the eight genes selected according to asthma severity. Our results show that *TRIM37* was predictive of the occurrence of asthma, while *NOX5*, *LMAN1*, *KNS2*, and *MLSTD1* were predictive of mild asthma, and *CCT5* was predictive of moderate-to-severe asthma.

Additionally, we analyzed the effect of asthma mediation on the expression levels of the three candidate genes. We found no difference in terms of mRNA level between the treated and untreated subjects using *t*-tests and multiple logistic regression analyses, indicating that there was no effect of mediation on the

expression of the three genes.

To date, the functions of the genes used as diagnostic markers of asthma have not been characterized. *MEPE* is an inhibitor of mineralization *in situ* and in cell cultures where altered expression is associated with oncogenic osteomalacia and hypophosphatemic rickets. *MLSTD1* is related to fatty alcohol synthesis in mammals.[27] *CCT5* is associated with cell proliferation, the cell cycle, morphological changes, and apoptosis.[28] TRIM37 encodes a member of the tripartite motif (*TRIM*) family, which is involved in diverse cellular functions, such as developmental patterning and oncogenesis. We searched various databases to determine any relationship between asthma and the selected genes; however, we were unable to find any connection. Thus, the functional relation of these genes with asthma remains to be solved; however, these genes do highlight the complex mechanism(s) underlying the pathogenesis of asthma. Furthermore, they comprise a set of potential biomarkers for use in discriminating between asthmatics and normal individuals. The three genes presented here as biomarkers were selected by computational methods alone. Additional functional studies are needed to determine the exact mechanism or indirect relationship among these genes with asthma.

In summary, a 35K whole-genome mRNA expression study of PBMCs from asthmatics and normal controls resulted in the selection of 170 genes based on criteria of $P \leq 0.001$ and a fold change ≥ 2. By applying more stringent criteria ($P \leq 0.001$ and a fold change ≥ 5), we selected eight genes as candidate asthma biomarkers. Using the AUCs from ROC curves, we identified a genetic biomarker consisting of three genes (*MEPE*, *MLSTD1*, *TRIM37*) having a diagnostic accuracy of 0.9928 AUC, with 98% sensitivity and 80% specificity. This marker may prove to be useful diagnostically after validation using a larger number of samples.

ADDITIONAL FILES

Additional file 1 - Table S1, Excel file, 170 genes ($P \leq 0.001$ and fold change ≥ 2).

Additional file 2 - Table S2, Excel file, 255 models (AUC, *P*-value).

Additional file 3 - Table S3, Excel file, gene list according to asthma severity.

Additional file 4 - Figure S1, PowerPoint file, ROC curves and AUCs for all models.

ACKNOWLEDGMENTS

This work was supported by a grant from the Korea Health 21 R&D Project, Ministry of Health, Welfare, and Family Affairs, Republic of Korea (A010249). Blood samples were generously provided by the Collaborative Biobank of Korea at Soonchunhyang University Bucheon Hospital.

REFERENCES

1. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald M, Gibson P, Ohta K, O'Byrne P, Pedersen SE, Pizzichini E, Sullivan SD, Wenzel SE, Zar HJ. Global strategy for asthma management and prevention: GINA executive summary. Eur Respir J 2008; 31:143-78.
2. Cohn L, Elias JA, Chupp GL. Asthma: mechanisms of disease persistence and progression. Annu Rev Immunol 2004;22:789-815.
3. Sengler C, Lau S, Wahn U, Nickel R. Interactions between genes and environmental factors in asthma and atopy: new developments. Respir Res 2002;3:7.
4. De Vos J, Thykjaer T, Tarte K, Ensslen M, Raynaud P, Requirand G, Pellet F, Pantesco V, Rème T, Jourdan M, Rossi JF, Ørntoft T, Klein B. Comparison of gene expression profiling between malignant and normal plasma cells with oligonucleotide arrays. Oncogene 2002; 21:6848-57.
5. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A 1999;96:6745-50.
6. Laprise C, Sladek R, Ponton A, Bernier MC, Hudson TJ, Laviolette M. Functional classes of bronchial mucosa genes that are differentially expressed in asthma. BMC Genomics 2004;5:21.
7. Lilly CM, Tateno H, Oguma T, Israel E, Sonna LA. Effects of allergen challenge on airway epithelial cell gene expression. Am J Respir Crit Care Med 2005;171:579-86.
8. Yang SH, Kim JS, Oh TJ, Kim MS, Lee SW, Woo SK, Cho HS, Choi YH, Kim YH, Rha SY, Chung HC, An SW. Genome-scale analysis of resveratrol-induced gene expression profile in human ovarian cancer cells using a cDNA microarray. Int J Oncol 2003;22:741-50.
9. Chamberland A, Madore AM, Tremblay K, Laviolette M, Laprise C. A comparison of two sets of microarray experiments to define allergic asthma expression pattern. Exp Lung Res 2009;35:399-410.
10. Baines KJ, Wood LG, Gibson PG. The nutrigenomics of asthma: molecular mechanisms of airway neutrophilia following dietary antioxidant withdrawal. OMICS 2009;13:355-65.
11. Guajardo JR, Schleifer KW, Daines MO, Ruddy RM, Aronow BJ, Wills-Karp M, Hershey GK. Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. J Allergy Clin Immunol 2005;115:243-51.
12. Kavuru MS, Dweik RA, Thomassen MJ. Role of bronchoscopy in asthma research. Clin Chest Med 1999;20:153-89.
13. Hansel NN, Hilmer SC, Georas SN, Cope LM, Guo J, Irizarry RA, Diette GB. Oligonucleotide-microarray analysis of peripheral-blood lymphocytes in severe asthma. J Lab Clin Med 2005;145:263-74.
14. Katsunuma T, Kawahara H, Suda T, Ishii T, Ohya Y, Akasawa A, Saito H, Oshida T, Sugita Y. Analysis of gene expressions of T cells from children with acute exacerbations of asthma. Int Arch Allergy Immunol 2004;134:29-33.
15. Aoki T, Matsumoto Y, Hirata K, Ochiai K, Okada M, Ichikawa K, Shibasaki M, Arinami T, Sumazaki R, Noguchi E. Expression profiling of genes related to asthma exacerbations. Clin Exp Allergy 2009; 39:213-21.
16. Subrata LS, Bizzintino J, Mamessier E, Bosco A, McKenna KL, Wikström ME, Goldblatt J, Sly PD, Hales BJ, Thomas WR, Laing IA, LeSouëf PN, Holt PG. Interactions between innate antiviral and atopic immunoinflammatory pathways precipitate and sustain asthma exacerbations in children. J Immunol 2009;183:2793-800.

17. Kraft P, Wacholder S, Cornelis MC, Hu FB, Hayes RB, Thomas G, Hoover R, Hunter DJ, Chanock S. Beyond odds ratios--communicating disease risk based on genetic profiles. Nat Rev Genet 2009;10:264-9.
18. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 1987;162:156-9.
19. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997;30:1145-59.
20. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006;27:861-74.
21. Kuperman DA, Lewis CC, Woodruff PG, Rodriguez MW, Yang YH, Dolganov GM, Fahy JV, Erle DJ. Dissecting asthma using focused transgenic modeling and functional genomics. J Allergy Clin Immunol 2005;116:305-11.
22. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, Ellwanger A, Sidhu SS, Dao-Pick TP, Pantoja C, Erle DJ, Yamamoto KR, Fahy JV. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. Proc Natl Acad Sci U S A 2007;104:15858-63.
23. Kruskal WH, Tanur JM. International encyclopedia of statistics. Vol. 1. New York: Free Press; 1978. Linear hypotheses; 523–41.
24. Chong PK, Lee H, Zhou J, Liu SC, Loh MC, Wang TT, Chan SP, Smoot DT, Ashktorab H, So JB, Lim KH, Yeoh KG, Lim YP. ITIH3 is a potential biomarker for early detection of gastric cancer. J Proteome Res 2010;9:3671-9.
25. Peat G, Thomas E, Duncan R, Wood L. Is a "false-positive" clinical diagnosis of knee osteoarthritis just the early diagnosis of pre-radiographic disease? Arthritis Care Res (Hoboken) 2010;62:1502-6.
26. Rom WN, Goldberg JD, Addrizzo-Harris D, Watson HN, Khilkin M, Greenberg AK, Naidich DP, Crawford B, Eylers E, Liu D, Tan EM. Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. BMC Cancer 2010;10:234.
27. Cheng JB, Russell DW. Mammalian wax biosynthesis. I. Identification of two fatty acyl-Coenzyme A reductases with different substrate specificities and tissue distributions. J Biol Chem 2004;279:37789-97.
28. Grantham J, Brackley KI, Willison KR. Substantial CCT activity is required for cell cycle progression and cytoskeletal organization in mammalian cells. Exp Cell Res 2006;312:2309-24.