Gene-environmental interaction study for discovery of new genetic factor for asthma

Junho Cha[1], Taehee Woo[2], Yongjun Choi[2], Sungkyoung Choi[2,3,*]

[1]Department of Molecular Life Science, Hanyang University (ERICA), Ansan, Korea
[2]Department of Applied Artificial intelligence, Hanyang University (ERICA), Ansan, Korea
[3]Department of Applied Mathematics, Hanyang University (ERICA), Ansan, Korea.

Abstract

- Asthma is a representative complex disease caused by the interaction of various environmental and immunological factors and genetic factors.
- To investigate the gene-environmental (smoke) interaction related in asthma, we conduct meta-analysis with Health examines study (HEXA) cohort data, Cardiovascular disease association study (CAVAS) cohort data and Korea association resource study (KARE) cohort data of Korean Genome and Epidemiology Study (KoGES), and genetic information was produced with Korean chips (KORV1.1).
- As a result of performing gene×smoke meta-analysis, 15 genes were identified and 2 genes were previously reported candidate gene with asthma.

Introduction

- Asthma, one of the major diseases occurring in the modern living environment, is a disease caused by an inflammatory reaction of the bronchial tubes when exposed to specific risk factors[1].
- Genome-wide association study finds associations for related traits or phenotypes at all genomic locations, discovered numerous asthma-related genetic variants[2]. However, the most complex traits of the discovered Single Nucleotide Polymorphism (SNP) explain only a small portion of the estimated heritability (missing heritability) [3].
- One solution suggested for this problem is a gene-environment interaction effect[4], and it is well known that asthma is caused by a combination of environmental and immunological factors[5].
- In this study, through previous studies, smoking status and allergic diseases were considered major environmental factors related to asthma[6,7]. In addition, it has been reported that the genetic risk of SNPs that can affect asthma and related traits is affected by age, sex, various other demographic and environmental factors[8].
- Therefore, an integrated analysis using mathematical and statistical models related to gene-environment interaction research will be conducted to discover candidate genomic markers related to asthma.

Materials & Methods

- **Quality Control (QC) of genetic data set**
  - ➢ Filter out
    - Insert-Deletion (InDel)
    - Multiallelic SNPs
    - Missing genotype call ratio > 0.05
    - Minor Allele Frequency (MAF) < 0.05
    - Hardy-Wein berg Equilibrium (HWE) p-value $\leq 10^{-5}$
  - ➢ Total 5,416,280 SNPs remained, after filtering 1,760,343 SNPs from the KORV1.1.

Table 1. General characteristics of subjects from HEXA, CAVAS, and KARE cohorts

	HEXA			CAVAS			KARE		
	Case (n = 975)	Control (n = 57,459)	p-value[a]	Case (n = 95)	Control (n = 2,908)	p-value	Case (n = 112)	Control (n = 5,308)	p-value
SEX									
Male	283(1.4%)	19,924(98.6%)	0.0002	37(3.1%)	1,164(96.9%)	0.8325	39(1.5%)	2,653(95.8%)	0.0048
Female	692(1.8%)	37,535(98.2%)		58(3.2%)	1,744(96.8%)		73(2.6%)	2,745(92.5%)	
AGE(years)[b]	55.4±8.4	53.8±8.0	< 0.0001	57.9±7.8	55.4±7.8	0.0025	53.3±7.9	51.5±8.5	0.0026
BMI(kg/m^2)[c]	24.3±3.2	23.9±2.9	0.0002	25.5±3.4	24.5±3.0	0.0002	25.0±3.5	24.6±3.0	0.1536
ALLER status[d]									
Non-ALLER	727(1.3%)	53,642(98.7%)	< 0.0001	74(2.7%)	2,695(97.3%)	< 0.0001	86(1.7%)	5,015(98.3%)	< 0.0001
ALLER	248(6.1%)	3,817(93.9%)		21(9.0%)	213(91.0%)		26(8.2%)	293(91.8%)	
SMOKE status[e]									
Non-Smokers	721(1.7%)	42,070(98.3%)	0.6348	72(3.0%)	2,123(97.0%)	0.5471	71(2.2%)	3,173(97.8%)	0.4397
Smokers	254(1.6%)	15,389(98.4%)		23(2.8%)	785(97.2%)		41(1.9%)	2,135(98.1%)	

[a]p-value from chi-square test or t-test
[b]Means ± standard deviation (SD)
[c]Body Mass Index (BMI)
[d]Allergy status (ALLER)
[e]Smoke status (non-Smoker: never smoker / Smoker: former smoker + current smoker)

- **Gene-environment (smoke) interaction study**
  - ➢ **(Step 1) : Logistic regression model by PLINK software**[9]

$$logit(ASTH = 2) = \beta_0 + \beta_1 \text{SNP} + \beta_2 \text{SEX} + \beta_3 \text{AGE} + \beta_4 \text{BMI} + \beta_5 \text{ALLER} + \beta_6 \text{SMOKE} + \boldsymbol{\beta_7} \text{SMOKE} \times \text{SNP} \quad \text{(Eq.1)}$$

(Genetic coding | SNP : AA=0, Aa=1, aa=2 / A : major allele, a : minor allele, Covariates | SEX, AGE, BMI, ALLER)

  - ➢ **(Step 2) : Meta-analysis model by METAL software**[10]

$$\text{Intermediate Statics} : Z_i = \Phi^{-1}\left(\frac{P_i}{2}\right) \times \text{sign}(\Delta_i)$$

$$W_i = \sqrt{N_i} \quad \text{(Eq.2)}$$

$$\text{Overall Z-Score} : \frac{\Sigma_i Z_i w_i}{\sqrt{\Sigma_i w_i^2}} \qquad \text{Overall Z-Score} : 2\Phi(|-Z|)$$

(N_i : Sample size for study i / P_i : P-value for study i / Δ_i : direction of effect for study i)

Results & Discussions

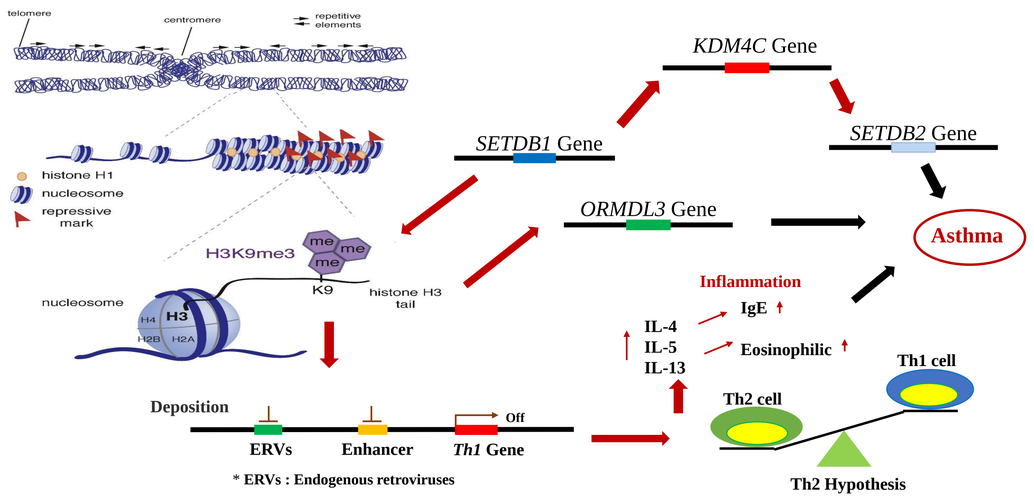
- **Results of gene×smoke meta-analysis**
  - ➢ Covariates: Age, Sex, BMI, and Allergy status
  - ➢ 15 Genes were identified as candidate genes associated with asthma, and 2 genes were previously reported candidate gene with asthma[10-11].

Table 2. Top list of SNP for gene-smoke interaction in asthma by meta analysis ($p \leq 3.0\times10^{-6}$).

CHR	rsID[a]	Allele1[b]	Allele2[c]	Weight	Z-score	p-value	Direction	Gene
10	rs17153428	a	t	66857	5.546	**Metal : 3.0×10^{-8}** HEXA : 3.0×10^{-7} CAVAS : 0.3212 KARE : 0.2789	+++	*RP11-383C5.4*
10	rs74743572	a	g	66857	-5.521	**Metal : 3.0×10^{-8}** HEXA : 4.0×10^{-7} CAVAS : 0.0349 KARE : 0.2353	---	*RP11-383C5.4*
10	rs57480332	t	c	66857	5.209	**Metal : 2.0×10^{-7}** HEXA : 2.0×10^{-6} CAVAS : 0.0418 KARE : 0.2186	+++	*RP11-383C5.3*
10	rs79219793	a	g	66857	-5.071	**Metal : 4.0×10^{-7}** HEXA : 1.0×10^{-6} CAVAS : 0.5597 KARE : 0.5857	---	*RP11-383C5.4*
9	rs4740872	t	g	66857	-4.82	**Metal : 1.0×10^{-6}** HEXA : 5.0×10^{-6} CAVAS : 0.2362 KARE : 0.3089	---	***KDM4C***
9	rs13292853	a	g	66857	-4.814	**Metal : 2.0×10^{-6}** HEXA : 2.0×10^{-6} CAVAS : 0.2557 KARE : 0.5588	---	***KDM4C***
1	rs71624514	t	c	66857	4.711	**Metal : 2.0×10^{-6}** HEXA : 1.0×10^{-6} CAVAS : 0.4197 KARE : 0.8869	+++	***SETDB1***
1	rs34207591	a	g	66857	-4.711	**Metal : 3.0×10^{-6}** HEXA : 1.0×10^{-6} CAVAS : 0.4197 KARE : 0.8869	---	***SETDB1***
2	rs7602146	c	g	66857	-4.65	**Metal : 3.0×10^{-6}** HEXA : 8.0×10^{-5} CAVAS : 0.0918 KARE : 0.0255	---	*AC113608.1*

[a]hg19, dbSNP150 version
[b]Minor allele
[c]Major allele

- **Novel gene (*SETDB1*) associated with asthma**[12].



- **Discussions**
  - ➢ This study is an association meta-analysis study of gene-environmental (smoke) interaction related to asthma, and it is significantly meaning that using Korean chips, several candidate genes related to asthma could be identified (containing previously reported candidate gene with asthma) and suggest a new candidate gene associated with asthma.
  - ➢ Further study, to more clearly verify the gene-environmental interaction associated with asthma, we will conduct gene-level and gene set-level meta-analysis.

References

1. Holgate, S.T., Asthma: a simple concept but in reality a complex disease. Eur J Clin Invest, 2011. 41(12): p. 1339-52.
2. Kim, K.W. and C. Ober, Lessons Learned From GWAS of Asthma. Allergy Asthma Immunol Res, 2019. 11(2): p. 170-187.
3. Manolio, T.A., et al., Finding the missing heritability of complex diseases. Nature, 2009. 461(7265): p. 747-53.
4. Pallares, L.F., Searching for solutions to the missing heritability problem. Elife, 2019. 8.
5. Rava, M., L.A. Smit, and R. Nadif, Gene-environment interactions in the study of asthma in the postgenomewide association studies era. Curr Opin Allergy Clin Immunol, 2015. 15(1): p. 70-8.
6. Kauffmann, F. and F. Demenais, Gene-environment interactions in asthma and allergic diseases: challenges and perspectives. J Allergy Clin Immunol, 2012. 130(6): p. 1229-40; quiz 1241-2.
7. Hancock, D.B., et al., Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. PLoS Genet, 2012. 8(12): p. e1003098.
8. Johansson, E. and T.B. Mersha, Genetics of Food Allergy. Immunol Allergy Clin North Am, 2021. 41(2): p. 301-319.
9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007, 81(3):559-575.
10. Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010, 26(17):2190-2191.
11. Heizmann, C.W., S100 proteins: Diagnostic and prognostic biomarkers in laboratory medicine. Biochim Biophys Acta Mol Cell Res, 2019. 1866(7): p. 1197-1206.
12. Torrano J, Al Emran A, Hammerlindl H, Schaider H: Emerging roles of H3K9me3, SETDB1 and SETDB2 in therapy-induced cellular reprogramming. Clin Epigenetics 2019, 11(1):43.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(No.2018R1C1B6008277). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-01343, Artificial Intelligence Center(Hanyang University ERICA)). This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (2019M3E5D3073365). This study was conducted with bioresources from National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2020-106).