EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

A Novel Locus for Exertional Dyspnea in Childhood Asthma

Sanghun Lee, Jessica Ann Lasky-Su, Christoph Lange, Wonji Kim, Preeti Lakshman Kumar, Merry-Lynn N. McDonald, Carlos A. Vaz Fragoso, Cecelia Laurie, Benjamin A. Raby, Juan C. Celedón, Michael H. Cho, Sungho Won, Scott T. Weiss, Julian Hecker

Please cite this article as: Lee S, Lasky-Su JA, Lange C, *et al*. A Novel Locus for Exertional Dyspnea in Childhood Asthma. *Eur Respir J* 2020; in press (https://doi.org/10.1183/13993003.01224-2020).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

A Novel Locus for Exertional Dyspnea in Childhood Asthma

Sanghun Lee[1,2] Jessica Ann Lasky-Su[3] Christoph Lange[2,3] Wonji Kim[3] Preeti Lakshman Kumar[4] Merry-Lynn N. McDonald[4] Carlos A. Vaz Fragoso[5] Cecelia Laurie[6] Benjamin A. Raby[3] Juan C. Celedón[7] Michael H. Cho[3] Sungho Won[8] Scott T. Weiss[3] Julian Hecker[3]

[1]Department of Medical consilience, Graduate school, Dankook university, South Korea
[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[3]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA
[4]Division of Pulmonary, Allergy and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, AL, USA
[5]Geriatric Medicine, Yale School of Medicine, New Haven, CT, USA
[6]Department of Biostatistics, University of Washington, Seattle, WA, USA
[7]Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA, USA
[8]Department of Public Health Science, Seoul National University, Seoul, South Korea.

Correspondence: Sanghun Lee, rehun@channing.harvard.edu

Running Head: A risk variant of exertional dyspnea in childhood asthma
Total word count: 2,550

Keywords: childhood asthma, whole-genome sequencing, family-based genome-wide association study, exertional dyspnea, *ACKR3*

Funding: This work was supported by the Industrial Core Technology Development Program (20000134) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea), the Bio-Synergy Research Project (NRF-2017M3A9C4065964) of the Ministry of Science, ICT and Future Planning through the National Research Foundation, and National Heart, Lung, and Blood Institute (P01 HL132825, Weiss PI).

Author Contributions

SL, CL, and JH conceptualized and designed the project, performed statistical analyses and interpretation, and drafted the manuscript. J.S.L and S.T.W assisted in data preparation and the analyses as well as manuscript preparation. PLK, MLM, and CAV contributed to the analysis on COPDGene. WK, CL, BAR, JCC, SW, and MC contributed to the critical revision of the manuscript. All authors contributed to the relevant sections and approved the final manuscript.

This article has an online supplement.

Abstract

Background: Most children diagnosed with asthma suffer from respiratory symptoms such as cough, dyspnea, and wheezing which are also important markers of overall respiratory function. A decade of genome-wide association studies (GWAS) have investigated the genetic susceptibility of asthma diagnosis itself, but few have focused on important respiratory symptoms that characterize childhood asthma.

Method: Using whole-genome sequencing (WGS) data for 894 asthmatic trios from a Costa Rican cohort, we performed family-based association tests (FBATs) to assess the association between genetic variants and multiple asthma-relevant respiratory phenotypes: cough, phlegm, wheezing, exertional dyspnea, and exertional chest tightness. We tested whether genome-wide significant associations replicated in two additional studies: 1) 286 WGS trios from the Childhood Asthma Management Program (CAMP), and 2) 2691 African American (AA) current or former smokers from the COPDGene study.

Results: In the 894 Costa Rican trios, we identified a genome-wide significant association between exertional dyspnea and single nucleotide polymorphism (SNP) rs10165869, located on chromosome 2q37.3 with a p value of 3.49×10^{-9} that was replicated in the CAMP cohort (p = 0.0222) with the same direction of association (combined p = $5.54\text{x}10^{-10}$), but was not associated in the AA subjects from COPDGene. We also found suggestive evidence of a link between SNP rs10165869 and the atypical chemokine receptor 3 (*ACKR3*) for the biological interpretation.

Conclusion: We identified and replicated a novel association between exertional dyspnea and SNP rs10165869 in childhood asthma which encourages to discover respiratory symptom associated variants in various airway diseases.

Tweetable Abstract

Our family-based association study using WGS data suggests that the replicated SNP variant, rs10165869 is linked with exertional dyspnea, likely through expression of ACKR3, among children with asthma.

INTRODUCTION

Asthma is the most common chronic respiratory disease of childhood and it is characterized by inflammation of the airways leading to airflow obstruction, and increased airway responsiveness that is generally caused by innate and adaptive immune responses to inhaled irritants and/or allergens [1]. The respiratory symptoms of an individual diagnosed with asthma consist of cough, dyspnea, phlegm, wheezing, and chest tightness, all of which are important markers of overall respiratory function and can signal asthma exacerbations [2].

Environmental exposures such as second-hand or personal exposure to tobacco smoke, viral respiratory infections, air pollutants, and allergens can contribute to, but do not account for all of these respiratory symptoms. Indeed, report or perception of respiratory symptoms differs among children with similar peak expiratory flow rate or FEV1 (forced expiratory volume in 1 second), and such perceptual accuracy of lung function has been shown to further vary by ethnicity [3]. Moreover, spirometry is usually measured at rest and therefore does not explain exertional symptoms (e.g. dyspnea) due to expiratory flow limitation and dynamic hyperinflation in response to exercise [4]. These data suggest that genetic variants may influence respiratory symptoms in children and that the magnitude of the effect of such variants may differ across ethnic groups.

To date, genome-wide association studies (GWASs) investigating the genetic susceptibility for asthma have been limited to asthma affection status, lung function measures, asthma severity, and response to asthma medications [5, 6]. Thus, GWAS of the respiratory symptom itself is lacking, particularly in children. Moreover, while there have been some studies of mucus

hypersecretion in adult smokers and in general population-based cohorts, no such studies have been conducted in individuals with asthma [7, 8].

In this study, we first performed a genome-wide association analysis using whole-genome sequencing (WGS) data for five major respiratory symptoms in a family-based study of childhood asthma in Costa Rica. Then, we attempted to replicate our findings in a WGS family-based multiethnic study of childhood asthma in North America. As childhood asthma has been associated with the development of chronic obstructive pulmonary disease (COPD), particularly in African Americans (AA) [9], we further attempted to replicate our findings in the Costa Rican study among African in the COPDGene study.

METHODS

GWAS Subjects

Subject recruitment and the study protocol for the Genetics of Asthma in Costa Rica Study (GACRS) have been previously described in detail [10, 11]. In brief, the GACRS is a family-based study of children with asthma and their parents (trios). All participants were aged 6 to 14 years and had asthma diagnosed by a physician and at least 2 respiratory symptoms (wheezing, cough, or dyspnea) or a history of asthma attacks in the previous year [12]. All participants also had at least 6 great-grandparents born in the Central Valley of Costa Rica, to ensure their descent from a founder population predominantly composed of Spaniards and Amerindians. The population was considered to be a semi-genetic isolate with Spanish and Amerindian ancestry due to topographical separation from the coasts and countries to the North and South by mountain ranges.

We first attempted to replicate our findings in the GACRS in a family-based study of North American children with asthma and their parents, including members of a diverse range of ethnicities from the Childhood Asthma Management Program (CAMP). CAMP was a multicenter clinical trial designed to determine the long-term effects of three inhaled treatments for childhood asthma [13]. All of the participants had to have increased airway responsiveness (a $PC_{20} \leq 12.5$ mg/ml of methacholine) and at least one of the following for at least six months in the previous year: respiratory symptoms at least twice per week, usage of an inhaled bronchodilator at least twice per week, or daily medication use for asthma. Children with severe asthma or other chronic medical conditions were excluded. Compared with GACRS, more participants in CAMP have moderate persistent asthma. We also attempted to replicate

our results in 2,691 current or former AA adult smokers in COPDGene. Subject recruitment and the study protocol for COPDGene have been previously described in detail [14].

Written parental consent and/or the subject's assent were obtained for each study protocol and ancillary genetic testing. Study protocols were approved by local Institutional Review Boards at each recruitment site for both studies, and by the Institutional Review Board of Brigham and Women's Hospital.

Whole-genome sequencing (WGS) data

WGS data were generated as part of the NHLBI Trans-Omics for Precision Medicine (TOPMed) WGS Program. Details regarding the laboratory methods, data processing, and quality control are described in the TOPMed website and in documents included in the TOPMed accession released on the database of Genotypes (TopMed Acc. number: phs000988 & phs001726) [15]. After evaluating genome-wide identity-by-descent (IBD) estimates generated by KING (Kinship-Based Inference for Genome-wide Association Studies), WGS data for 894 Costa Rican trios (a total of 2682 subjects), 286 CAMP trios (a total of 858 subjects, European American: African American: Latinos: others = 216: 28: 19: 23) and 2691 COPDGene AA subjects were obtained respectively [16]. For variant quality control, single nucleotide polymorphisms (SNPs) with Mendelian errors, genotyping rate (< 98%), allele frequency (< 1%) or deviations from Hardy-Weinberg proportions ($p < 10^{-8}$) were removed.

Respiratory Symptoms

In GACRS and CAMP study, questionnaires were used to obtain data on respiratory symptoms, including usual cough, usual phlegm, wheezing without cold, exertional dyspnea, and exertional chest tightness according to the standard practice in epidemiology studies of children between the ages of 6-14. They were administered at the same time as spirometry measurement and collection of blood samples. For each respiratory symptom, the child or her/his parent (usually the mother) had to answer "Yes" to the following questions: "Does the child usually have cough?", "Is the child usually congested?", "Is the child's lung sounded wheezy without cold or flu?", "Has there been an occasion when the child had an attack of shortness of breath when hurrying up on the level or walking up a slight hill?", and "Has there been an occasion when the child had an attack of chest tightness when hurrying up on the level or walking up a slight hill?" respectively.

Statistical analysis

Demographic characteristics and clinical features between groups depending on specific respiratory symptoms were compared using t-tests or Chi-square tests, as appropriate. Association analysis was conducted using the family-based association test (FBAT) software (version 2.04) under an additive genetic model adjusting for age, sex, body mass index (BMI), lung function (FEV/FVC), $\log_{10}$(total IgE), and usage on systemic oral corticosteroids in the previous year (as a proxy for asthma severity) [17]. For the top variants showing genome-wide significance, haplotype-based analysis after the same adjustment was also done using FBAT. As around 60-80% of the asthmatic subjects reported each respiratory symptom, the phenotype information for respiratory symptoms could be mixed with asthma affection status. To

distinguish our associations from asthma-related associations, we investigated the transmission-pattern in symptom-specific subgroups separately. Also, we analyzed the association p-value for asthma affection status of the top significant variant.

In the COPDGene cohort, single-variant association test method (https://genome.sph.umich.edu/wiki/EPACTS) was used after adjusting age, sex, Global Lung Function Initiative-calculated spirometric Z-scores [18], pack-years of smoking, and top 6 PC adjusting for population structure. We explored the association between the variant and gene by Open Targets Genetics (https://genetics.opentargets.org) and summary-Predixcan [19] was applied using cis-expression quantitative trait loci (eQTL) in lung tissue based on the GTEx release version 8 database (https://www.gtexportal.org/home). Topologically associated domains (TADs) in lung tissue were also checked in a three-dimensional genome browser (Hi-C Unifying Genomic Interrogator, https://yunliweb.its.unc.edu/hugin). The R statistical software (http://www.R-project.org) was used to evaluate these tests.

RESULTS

Descriptive characteristics of children with asthma

The characteristics of the 894 children with asthma who participated in the GACRS are summarized in Table 1. The median age of study participants was 9.1 years, with the sex ratio 1:0.7 (male:female), and a median age of asthma onset of 2.0 years (range=0–12 years). Most children (83.6%) had at least one positive skin test to allergens and their median FEV_1 (% predicted) was 97.5%. In CAMP, study participants had a median age of 8.8 years with a sex ratio of 1:0.5 (male:female) and a median age of asthma onset of 2.0 years (range= 0–11 years).

Table 2 shows the main characteristics of participants in the GACRS, according to the presence of each respiratory symptom. Subjects having cough or phlegm were slightly shorter than those without these symptoms, and children having exertional dyspnea or chest tightness had a higher body mass index than those without such symptoms. As expected, airway hyper-responsiveness and bronchodilator response were more common in children with wheezing, exertional dyspnea, or chest tightness than in those without these symptoms. Children with wheezing had more positive skin prick tests than those without wheezing, and children with phlegm showed slightly higher eosinophil counts than those without phlegm. Subjects with each respiratory symptom were likely to receive systemic corticosteroid medication compared to those without symptoms.

GWAS of respiratory symptoms

Genome-wide FBAT results are displayed in quantile–quantile and Manhattan plots (**Figure 1**). In this analysis, SNP rs10165869 (located on chromosome 2q37.3) was associated with exertional dyspnea (Z = 5.989, $p = 2.16\times10^{-9}$), and this association remained significant even

after a Bonferroni correction for the five phenotypes tested (i.e. *p*-value smaller than 1.0×10^{-8} or 5×10^{-8} divided by 5). SNPs nearby rs10165869, including rs6725280, rs1865671, rs7607911, and rs30102, and rs10168628had *p*-values that achieved genome-wide significance, ranging from 3.77×10^{-9} to 3.97×10^{-8} (figure 1D). A haplotype-based test using all 6 SNPs was significantly associated with exertional dyspnea (Z = 5.691, $p = 1.30\times10^{-8}$, supplementary Table 1, supplementary Figure 1) [20].

Using asthma affection status as the target phenotype, SNP rs10165869 is not significant with a *p*-value of 0.247. The subgroup analysis including the GACRS children with or without the exertional dyspnea symptom disclosed a different transmission behavior in both groups (Z = 3.744, $p = 1.81\times10^{-4}$, Number of informative family = 447 vs. Z = -4.704, $p = 2.55\times10^{-6}$, Number of informative family = 120, respectively) which clearly demonstrated the association between the minor allele (rs10165869) and exertional dyspnea.

The SNP, rs10165869 was replicated for exertional dyspnea in the ethnically diverse CAMP study (Z = 2.266, *p* = 0.023). A meta-analysis of data from the GACRS and CAMP yielded a combined *p*-value of 3.28×10^{-10}. The corresponding subgroup analysis in CAMP showed the same pattern as in GACRS (Z = 1.353, *p* = 0.176 vs. Z = -1.844, *p* = 0.065, respectively). Another subgroup analysis depending on ethnicities was also performed (supplementary Table 2). Besides, rs10165869 was also marginally associated in the AA subjects from COPDGene cohort but with the opposite effect direction (Z = -1.811, *p* = 0.070)

SNP rs10165869 and potential gene, atypical chemokine receptor 3 (*ACKR3*)

The set of potential genes functionally implicated by SNP rs10165869 are *ACKR3*, *COPS8*, *IQCA1*, and *COL6A3* from highest to lowest. Summary-Predixcan analysis in chromosome 2

suggested that only *ACKR3* expression in lung tissue remains after Bonferroni correction (Z = 4.78, $p = 1.72\times10^{-6}$, supplementary Table 3). The Hi-C Unifying Genomic Interrogator also captured a significant association between *ACKR3* and rs10165869 in lung tissue after Bonferroni correction ($p < 1.0\times10^{-11}$, supplementary Figure 2).

DISCUSSIONS

To our knowledge, this is the first family-based association study of the five major respiratory symptoms in childhood asthma, using WGS data. In this analysis, we identified SNP rs10165869 as a novel locus for exertional dyspnea among Costa Rican children with asthma, with replication in the independent and ethnically diverse CAMP study. Dose-response slope to methacholine across the cohort did not show any association with the locus after the same adjustment (p = 0.380), even though it was associated with both exertional phenotypes at baseline (Table2). Given that dyspnea in childhood asthma is influenced by sex [21], we performed a sex interaction analysis on SNP rs10165869 but it was not significant (p = 0.320).

ACKR3, also known as C-X-C chemokine receptor type 7 (CXCR7) is a G protein-coupled receptor (GPCR) for CXCL12, a chemokine that is involved in the inflammatory process regulating leukocyte extravasation into inflamed tissues [22, 23]. A CXCL12 has been found in high concentrations in bronchoalveolar lavage (BAL) from subjects with asthma, in whom BAL CXCL12 is correlated with circulating leukocytes, thus suggesting a role in airway inflammation and airway hyper-responsiveness [22]. Initially, CXCR4 had long been investigated as the only GPCR for CXCL12 until it was shown that ACKR3 has about 10-fold higher binding affinity for CXCL12 [23]. In the murine model having the characteristic features of asthma, knockdown of *ACKR3* by lentiviral delivery system in the lung reduces mucus secretion, allergic airway inflammation, serum allergen-specific IgE production, T-cell cytokine production, and airway hyper-responsiveness [24]. The predominant role of ACKR3 in CXCL12/CXCR4/ACKR3 axis

contributing to airway inflammation was also shown on the pulmonary epithelium, polymorphonuclear neutrophil (PMN), and transepithelial PMN migration [25].

Apart from inflammatory conditions, *CXCL12* and *ACKR3* are highly expressed under hypoxic conditions in proangiogenic environments such as various tumors, where CXCL12 enhances angiogenesis through ACKR3 activation [22, 26]. It was also shown that *ACKR3* expression responds more sensitively to hypoxia for up to 48 hours compared to *CXCR4* especially in lung endothelial cells shown in supplementary Figure 3 [27]. Endothelial cell dysfunction in the pathogenesis regeneration under alveolar hypoxia similar to that found in lung disease is chiefly mediated by ACKR3 [28]. In the GTEx database, the estimated effect of top SNPs including rs10165869 is mediated by increased expression of *ACKR3* in lung tissue (supplementary Table 3). Taken together, the role of ACKR3 on pulmonary epithelial and endothelial cells in the preclinical studies supports that the variant's effect (rs10165869) on the up-regulation of *ACKR3* in the lung could lead to more severe dyspnea symptom compared to asthma patients with wild type where exertional dyspnea means hypoxic condition due to chronic asthma. Previous GWAS also suggested an important role for rs7607316 and rs144060362 near the *ACKR3* region as genetic risk factors for airflow obstruction related to COPD ($p = 3\times10^{-6}$, 2×10^{-6} respectively) [29, 30]. Therefore, we also attempted to replicate the significant finding between the SNP and its phenotype to a broader range of older adult disease, COPD, but the reported association was not present in this population.

Our study has several limitations. First, answers to questions about respiratory symptoms are subjective and thus influenced by both the perception of symptoms and recall bias. However,

the misclassification seemingly non-differential, may lead to the null or weaker associations than what actually exist. Moreover, these concerns are ameliorated by replication of our results for exertional dyspnea in an independent multi-ethnic cohort of children living in a different geographic location. Second, this is a cross-sectional analysis. Although genotypes do not vary over time, we are unable to assess the temporal stability or variable severity of the reported symptoms in children with asthma. Third, we lack data on the use of controller medications for asthma in the GACRS, which would affect respiratory symptoms. Despite these challenges, our analysis highlights the significance of symptom-based GWAS to identify the genetic determinants of respiratory symptoms after adjusting the confounders such as lung function and asthma severity.

In conclusion, our study identified that SNP rs10165869 is linked with exertional dyspnea among children with asthma fostering a better understanding who could be free of exertional dyspnea. We hope that it would encourage to discover a wider range of phenotypes underlying respiratory symptom associated variants in various airway diseases.

Acknowledgements

WGS for the TOPMed program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for NHLBI TOPMed: The Genetic Epidemiology of Asthma in Costa Rica (phs000988.v1.p1) and Childhood Asthma Management Program (phs001726.v1.p1) were performed at the University of Washington Northwest Genomics Center (3R37HL066289-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

Conflict of Interest Statements

Dr. Celedón has received research materials from Pharmavite (vitamin D and placebo capsules) and Merck and GSK (inhaled steroids), in order to provide medications free of cost to participants in NIH-funded studies, unrelated to the current work. The other authors declare that there is no conflict of interest. Dr. Cho reports grants from NHLBI, grants from Bayer, grants from GSK, personal fees from AstraZeneca, personal fees from Illumina, outside the submitted work.

Table 1. Demographic characteristics and clinical features of Costa Rican subjects (N=894) and CAMP (N=286). Mean (SD) or N (%) is shown.

	Costa Rican Study subjects	CAMP study subjects	P value
Age	9.28 ± 1.86	8.90 ± 2.16	7.59×10^{-3}
Male gender	528 (59.06)	193 (67.48)	0.0134
Body mass index, kg/m^2	18.45 ± 3.92	18.24 ± 3.28	0.357
Spirometric measure			
FEV_1, L	1.80 ± 0.52	1.65 ± 0.47	1.03×10^{-5}
FEV_1 (% of predicted)	98.87 ± 16.84	92.82 ± 13.40	1.17×10^{-9}
FEV_1/FVC, %	84.36 ± 7.79	79.69 ± 8.07	2.20×10^{-16}
FEV_1/FVC (% of predicted)	94.87 ± 8.70	90.36 ± 8.94	5.36×10^{-13}
Bronchodilator Response as % of baseline FEV1	5.58 ± 10.20	10.86 ± 9.49	7.50×10^{-15}
Log_{10} Dose response slope from saline	1.15 ± 0.54	0.97 ± 0.47	1.92×10^{-7}
Blood tests			
Total serum IgE, IU/mL	725.26 ± 898.96	1084.29 ± 1945.05	2.87×10^{-3}
Eosinophil, $count/mm^2$	553.84 ± 403.93	511.96 ± 425.43	0.148
No. of positive skin tests to allergens	3.08 ± 1.82	5.56 ± 4.20	2.20×10^{-16}

Table 2. Demographic characteristics and clinical features between groups according to respiratory symptoms among children who participated in the Genetics of Asthma in Costa Rica Study. Mean (SD) or N (%) is shown.

	Cough			Phlegm			Wheezing			Exertional Dyspnea			Exertional Chest tightness		
	No	Yes	*p value*	No	Yes	P value	No	Yes	P value	No	Yes	P value	No	Yes	P value
N	190 (21.25)	704 (78.75)		414 (46.31)	479 (53.58)		105 (11.74)	789 (88.26)		193 (21.59)	701 (78.41)		266 (29.75)	624 (69.80)	
Age	9.49 ± 1.84	9.22 ± 1.86	0.0773	9.52 ± 1.87	9.07 ± (1.83)	3.07×10^{-4}	9.28 ± 1.84	9.28 ± 1.86	0.972	9.19 ± 1.94	9.30 ± 1.84	0.466	9.08 ± 1.77	9.35 ± 1.89	0.0428
Male gender	112 (58.95)	416 (59.09)	1	247 (59.66)	281 (58.66)	0.815	61 (58.10)	467 (59.19)	0.914	124 (64.25)	404 (57.63)	0.116	157 (59.02)	369 (59.13)	1
Height, m	1.35 ± 0.12	1.32 ± (0.12)	1.84×10^{-3}	1.35 ± 0.12	1.32 ± 0.12	5.13×10^{-4}	1.32 ± 0.12	1.33 ± 0.12	0.563	1.32 ± 0.13	1.33 ± 0.12	0.301	1.32 ± 0.12	1.33 ± 0.12	0.139
Body mass index, kg/m2	18.67 ± 4.28	18.39 ± 3.81	0.413	18.75 ± 4.08	18.20 ± 3.76	0.0347	18.34 ± 3.85	18.47 ± 3.93	0.756	17.86 ± 3.54	18.62 ± 4.00	0.0114	18.01 ± 3.98	18.63 ± 3.89	0.0335
No. of older siblings	1.33 ± 1.43	1.09 ± 1.27	0.0366	1.16 ± 1.28	1.12 ± 1.33	0.656	1.11 ± 1.35	1.14 ± 1.31	0.784	1.15 ± 1.27	1.14 ± 1.32	0.912	1.11 ± 1.30	1.15 ± 1.32	0.640
Spirometric measure															
FEV1/FVC, %	85.17 ± 7.53	84.14 ± 7.85	0.0967	84.56 ± 8.25	84.20 ± 7.37	0.495	84.80 ± 6.72	84.30 ± 7.92	0.492	84.91 ± 7.68	84.21 ± 7.82	0.266	84.89 ± 7.73	84.10 ± 7.81	0.161
FEV1/FVC (% of predicted)	95.84 ± 8.49	94.61 ± 8.74	0.0805	95.17 ± 9.29	94.64 ± 8.16	0.369	95.33 ± 7.72	94.81 ± 8.82	0.525	95.57 ± 8.51	94.68 ± 8.75	0.204	95.39 ± 8.60	94.60 ± 8.74	0.212
Absolute response to bronchodilator, mL	95.33 ± 124.61	82.61 ± 139.98	0.230	83.28 ± 143.08	86.84 ± 131.34	0.703	63.52 ± 101.68	88.12 ± 140.53	0.0325	74.24 ± 125.28	88.41 ± 139.85	0.180	73.39 ± 133.75	90.39 ± 138.37	0.0897
Dose-response slope to methacholine, umol	28.94 ± 36.59	27.21 ± 40.58	0.603	28.10 ± 39.11	27.17 ± 40.37	0.747	22.23 ± 31.92	28.35 ± 40.74	0.0901	22.17 ± 33.16	29.03 ± 41.27	0.0267	21.44 ± 28.64	30.30 ± 43.54	9.08×10^{-4}
the provocating dose (PD20) of 20% fall of FEV1	1.69 ± 2.21	1.84 ± 2.35	0.504	1.69 ± 2.36	1.91 ± 2.30	0.276	2.43 ± 2.86	1.71 ± 2.22	0.0379	2.20 ± 2.79	1.71 ± 2.18	0.0747	2.18 ± 2.63	1.67 ± 2.18	0.0298
Blood tests															
Total serum IgE, IU/mL	731.46 ± 982.71	723.59 ± 875.80	0.921	728.68 ± (927.43)	713.37 ± 853.38	0.799	583.52 ± 863.08	744.19 ± 902.50	0.0770	674.34 ± 793.46	739.24 ± 925.88	0.334	676.38 ± 797.89	749.62 ± 940.18	0.237
Positive IgE to dust mite	132 (69.47)	531 (75.43)	0.103	303 (73.19)	359 (74.95)	0.606	71 (67.62)	592 (75.03)	0.120	141 (73.06)	522 (74.47)	0.716	194 (72.93)	467 (74.84)	0.556
Positive IgE to cockroach	80 (42.11)	307 (43.61)	0.750	179 (43.24)	207 (43.21)	1	35 (33.33)	355 (44.61)	0.0350	81 (41.97)	306 (43.65)	0.714	104 (39.10)	282 (45.19)	0.100
Positive IgE to ascaris	67 (35.26)	267 (37.93)	0.571	155 (37.44)	178 (37.16)	0.962	29 (27.62)	305 (38.66)	0.0343	62 (32.12)	272 (38.80)	0.111	88 (33.08)	244 (39.10)	0.105
Eosinophil, count/mm^2	503.54 ± 382.75	567.39 ± 408.66	0.0481	517.06 ± 374.41	582.96 ± 424.11	0.0151	495.11 ± 436.92	561.62 ± 399.01	0.147	529.81 ± 406.98	560.45 ± 403.14	0.361	537.91 ± 407.68	563.57 ± 402.16	0.394
No. of positive skin tests to allergens	3.01 ± 1.86	3.10 ± 1.81	0.554	3.12 ± 1.81	3.06 ± 1.83	0.623	2.57 ± 1.96	3.15 ± 1.79	5.40×10^{-3}	3.15 ± 1.74	3.06 ± 1.84	0.531	3.18 ± 1.85	3.04 ± 1.80	0.335
No. of systemic corticosteroid medication last year	135 (71.05)	567 (80.54)	6.41×10^{-3}	308 (74.40)	393 (82.05)	7.08×10^{-3}	67 (63.81)	635 (80.48)	1.56×10^{-4}	137 (70.98)	565 (80.60)	5.42×10^{-3}	193 (72.56)	505 (80.93)	7.13×10^{-3}

Figure legends

Figure 1. GWAS in 894 Costa Rican trios for the respiratory symptoms such as cough (A), phlegm (B), wheezing without cold (C), exertional dyspnea (D), and exertional chest tightness (E), shown in Quantile-Quantile plot and Manhattan plot. The red line corresponds to a Bonferroni-corrected threshold for total 5 tests (p-value = 1.0×10^{-8}), and the blue line means genome-wide significant (p-value = 5×10^{-8}).

Supplementary Figure 1. Regional plot for an exertional dyspnea-associated locus in 400 kb upstream and downstream regions. Purple diamond represents the top-ranked SNP, rs10165869 and other SNPs are colored according to their r^2 value in relation to that SNP. The recombination rate correlations and LD map were estimated on the basis of the 1000 Genomes Project EUR datasets.

Supplementary Figure 2. Regions of chromatin contact for rs10165869 in lung tissue. Blue lines reflect -log(p-value) for the one-tailed test of whether observed Hi-C counts (black lines) were greater than the expected number of Hi-C counts (solid red lines). *ACKR3* is significantly associated with rs10165869 after tissue-specific False Discovery Rate (FDR, dashed red) and Bonferroni (dashed purple) corrections.

Supplementary Figure 3. CXCR4 (A) and ACKR3 (B) expression in pulmonary and cardiac microvascular endothelial cell (human) response to hypoxia in vitro: time course up to 48 hours

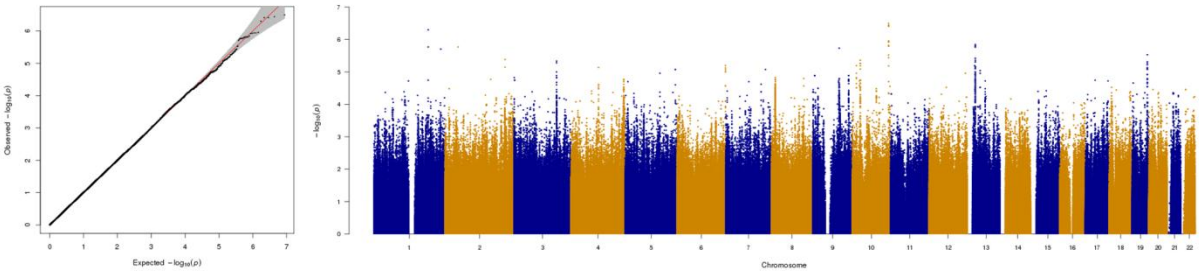
References

1. Martinez FD, Vercelli D. Asthma. *Lancet* 2013: 382(9901): 1360-1372.
2. Papadopoulos NG, Arakawa H, Carlsen KH, Custovic A, Gern J, Lemanske R, Le Souef P, Makela M, Roberts G, Wong G, Zar H, Akdis CA, Bacharier LB, Baraldi E, van Bever HP, de Blic J, Boner A, Burks W, Casale TB, Castro-Rodriguez JA, Chen YZ, El-Gamal YM, Everard ML, Frischer T, Geller M, Gereda J, Goh DY, Guilbert TW, Hedlin G, Heymann PW, Hong SJ, Hossny EM, Huang JL, Jackson DJ, de Jongste JC, Kalayci O, Ait-Khaled N, Kling S, Kuna P, Lau S, Ledford DK, Lee SI, Liu AH, Lockey RF, Lodrup-Carlsen K, Lotvall J, Morikawa A, Nieto A, Paramesh H, Pawankar R, Pohunek P, Pongracic J, Price D, Robertson C, Rosario N, Rossenwasser LJ, Sly PD, Stein R, Stick S, Szefler S, Taussig LM, Valovirta E, Vichyanond P, Wallace D, Weinberg E, Wennergren G, Wildhaber J, Zeiger RS. International consensus on (ICON) pediatric asthma. *Allergy* 2012: 67(8): 976-997.
3. Fritz GK, McQuaid EL, Kopel SJ, Seifer R, Klein RB, Mitchell DK, Esteban CA, Rodriguez-Santana J, Colon A, Alvarez M, Canino G. Ethnic differences in perception of lung function: a factor in pediatric asthma disparities? *Am J Respir Crit Care Med* 2010: 182(1): 12-18.
4. Marcus BS, McAvay G, Gill TM, Vaz Fragoso CA. Respiratory symptoms, spirometric respiratory impairment, and respiratory disease in middle-aged and older persons. *J Am Geriatr Soc* 2015: 63(2): 251-257.
5. Kim KW, Ober C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res* 2019: 11(2): 170-187.
6. Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology* 2017: 6(12): e165.
7. Zeng X, Vonk JM, de Jong K, Xu X, Huo X, Boezen HM. No convincing association between genetic markers and respiratory symptoms: results of a GWA study. *Respir Res* 2017: 18(1): 11.
8. Dijkstra AE, Boezen HM, van den Berge M, Vonk JM, Hiemstra PS, Barr RG, Burkart KM, Manichaikul A, Pottinger TD, Silverman EK, Cho MH, Crapo JD, Beaty TH, Bakke P, Gulsvik A, Lomas DA, Bosse Y, Nickle DC, Pare PD, de Koning HJ, Lammers JW, Zanen P, Smolonska J, Wijmenga C, Brandsma CA, Groen HJ, Postma DS, LifeLines Cohort Study g. Dissecting the genetics of chronic mucus hypersecretion in smokers with and without COPD. *Eur Respir J* 2015: 45(1): 60-75.
9. Hayden LP, Cho MH, Raby BA, Beaty TH, Silverman EK, Hersh CP, Investigators CO. Childhood asthma is associated with COPD and known asthma variants in COPDGene: a genome-wide association study. *Respir Res* 2018: 19(1): 209.
10. Hunninghake GM, Lasky-Su J, Soto-Quiros ME, Avila L, Liang C, Lake SL, Hudson TJ, Spesny M, Fournier E, Sylvia JS, Freimer NB, Klanderman BJ, Raby BA, Celedon JC. Sex-stratified linkage analysis identifies a female-specific locus for IgE to cockroach in Costa Ricans. *Am J Respir Crit Care Med* 2008: 177(8): 830-836.
11. Hunninghake GM, Soto-Quiros ME, Avila L, Ly NP, Liang C, Sylvia JS, Klanderman BJ, Silverman EK, Celedon JC. Sensitization to Ascaris lumbricoides and severity of childhood asthma in Costa Rica. *J Allergy Clin Immunol* 2007: 119(3): 654-661.

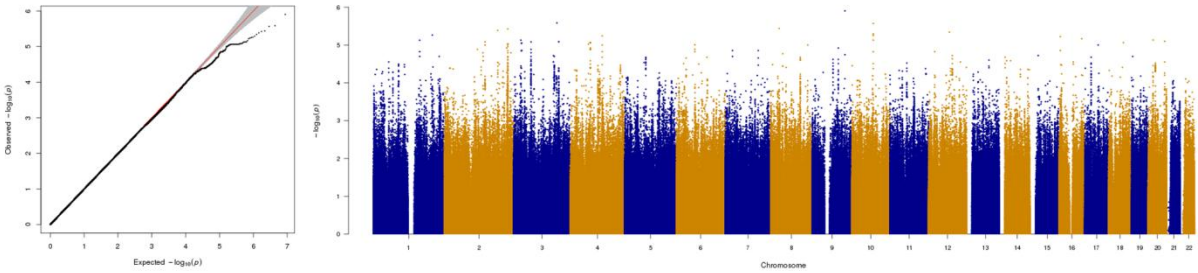
12. Weiss ST. NHLBI TOPMed: The Genetic Epidemiology of Asthma in Costa Rica.
13. Covar RA, Fuhlbrigge AL, Williams P, Kelly HW, the Childhood Asthma Management Program Research G. The Childhood Asthma Management Program (CAMP): Contributions to the Understanding of Therapy and the Natural History of Childhood Asthma. *Curr Respir Care Rep* 2012: 1(4): 243-250.
14. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010: 7(1): 32-43.
15. Regier AA, Farjoun Y, Larson DE, Krasheninina O, Kang HM, Howrigan DP, Chen BJ, Kher M, Banks E, Ames DC, English AC, Li H, Xing J, Zhang Y, Matise T, Abecasis GR, Salerno W, Zody MC, Neale BM, Hall IM. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* 2018: 9(1): 4038.
16. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010: 26(22): 2867-2873.
17. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 2000: 19 Suppl 1: S36-42.
18. Vaz Fragoso CA, McAvay G, Van Ness PH, Casaburi R, Jensen RL, MacIntyre N, Yaggi HK, Gill TM, Concato J. Phenotype of Spirometric Impairment in an Aging Population. *Am J Respir Crit Care Med* 2016: 193(7): 727-735.
19. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, Stahl EA, Huckins LM, Consortium GT, Nicolae DL, Cox NJ, Im HK. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018: 9(1): 1825.
20. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 2004: 26(1): 61-69.
21. Fu L, Freishtat RJ, Gordish-Dressman H, Teach SJ, Resca L, Hoffman EP, Wang Z. Natural progression of childhood asthma symptoms and strong influence of sex and puberty. *Ann Am Thorac Soc* 2014: 11(6): 939-944.
22. Janssens R, Struyf S, Proost P. Pathological roles of the homeostatic chemokine CXCL12. *Cytokine Growth Factor Rev* 2018: 44: 51-68.
23. Sanchez-Martin L, Sanchez-Mateos P, Cabanas C. CXCR7 impact on CXCL12 biology and disease. *Trends Mol Med* 2013: 19(1): 12-22.
24. Chang HC, Huang PH, Syu FS, Hsieh CH, Chang SL, Lu J, Chen HC. Critical involvement of atypical chemokine receptor CXCR7 in allergic airway inflammation. *Immunology* 2018: 154(2): 274-284.
25. Ngamsri KC, Muller A, Bosmuller H, Gamper-Tsigaras J, Reutershan J, Konrad FM. The Pivotal Role of CXCR7 in Stabilization of the Pulmonary Epithelial Barrier in Acute Pulmonary Inflammation. *J Immunol* 2017: 198(6): 2403-2413.
26. Zhang M, Qiu L, Zhang Y, Xu D, Zheng JC, Jiang L. CXCL12 enhances angiogenesis through CXCR7 activation in human umbilical vein endothelial cells. *Sci Rep* 2017: 7(1): 8289.
27. Costello CM, Howell K, Cahill E, McBryan J, Konigshoff M, Eickelberg O, Gaine S, Martin F, McLoughlin P. Lung-selective gene responses to alveolar hypoxia: potential role for the bone

morphogenetic antagonist gremlin in pulmonary hypertension. *Am J Physiol Lung Cell Mol Physiol* 2008: 295(2): L272-284.
28. Costello CM, McCullagh B, Howell K, Sands M, Belperio JA, Keane MP, Gaine S, McLoughlin P. A role for the CXCL12 receptor, CXCR7, in the pathogenesis of human pulmonary vascular disease. *Eur Respir J* 2012: 39(6): 1415-1424.
29. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald ML, Regan E, Mattheisen M, DeMeo DL, Parker M, Foreman M, Make BJ, Jensen RL, Casaburi R, Lomas DA, Bhatt SP, Bakke P, Gulsvik A, Crapo JD, Beaty TH, Laird NM, Lange C, Hokanson JE, Silverman EK, Investigators E, Investigators CO. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* 2015: 16: 138.
30. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, Smith AV, Heckbert SR, Smolonska J, Tang W, Loth DW, Curjuric I, Hui J, Cho MH, Latourelle JC, Henry AP, Aldrich M, Bakke P, Beaty TH, Bentley AR, Borecki IB, Brusselle GG, Burkart KM, Chen TH, Couper D, Crapo JD, Davies G, Dupuis J, Franceschini N, Gulsvik A, Hancock DB, Harris TB, Hofman A, Imboden M, James AL, Khaw KT, Lahousse L, Launer LJ, Litonjua A, Liu Y, Lohman KK, Lomas DA, Lumley T, Marciante KD, McArdle WL, Meibohm B, Morrison AC, Musk AW, Myers RH, North KE, Postma DS, Psaty BM, Rich SS, Rivadeneira F, Rochat T, Rotter JI, Soler Artigas M, Starr JM, Uitterlinden AG, Wareham NJ, Wijmenga C, Zanen P, Province MA, Silverman EK, Deary IJ, Palmer LJ, Cassano PA, Gudnason V, Barr RG, Loos RJ, Strachan DP, London SJ, Boezen HM, Probst-Hensch N, Gharib SA, Hall IP, O'Connor GT, Tobin MD, Stricker BH. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med* 2012: 186(7): 622-632.

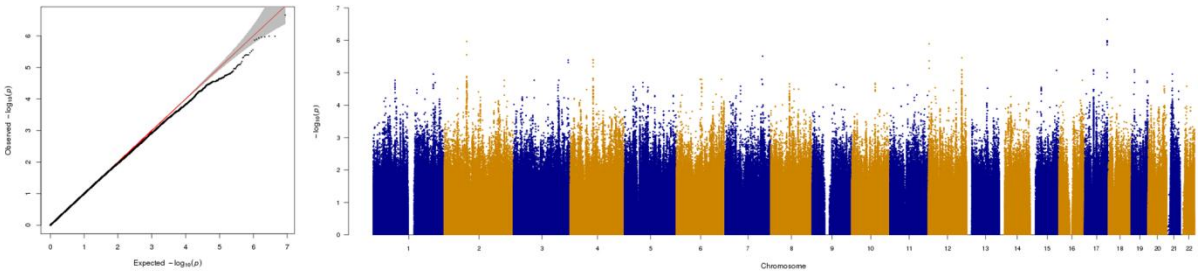
A. Cough



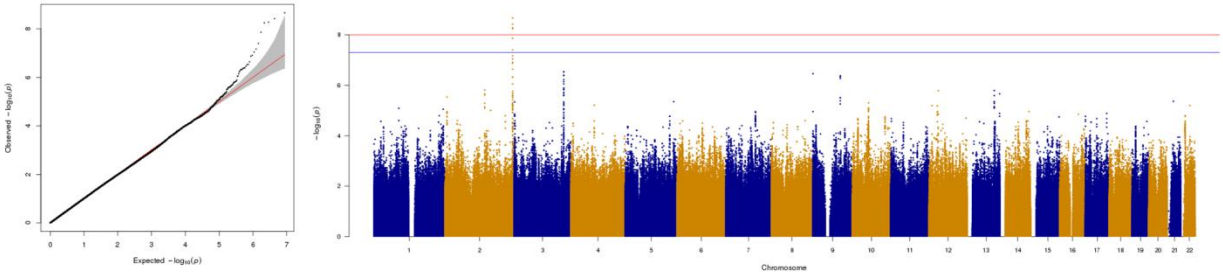
B. Phlegm



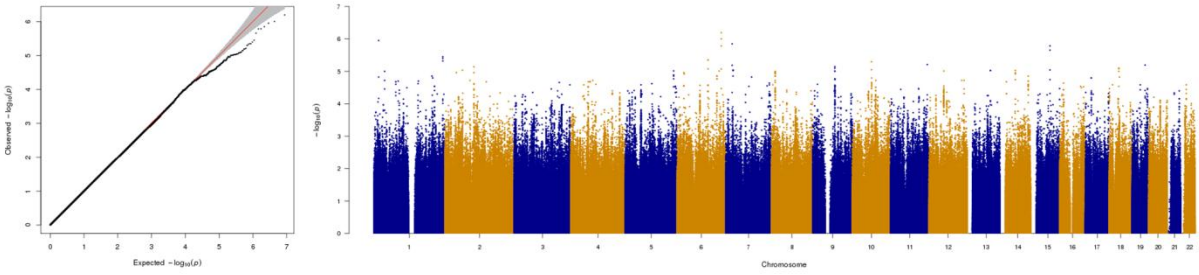
C. Wheezing without Cold



D. Exertional Dyspnea



E. Exertional Chest tightness



Supplementary Table 1. Haplotype-based test using the genome-wide significant SNPs such as rs10165869, rs6725280, rs7607911, rs1865671, rs30102, and rs10168628 where A1 and A2 haplotype mean all minor and major alleles of 6 SNPs respectively.

Haplotype	frequency	family numbers	S-E(S)	Var(S)	Z	P value
A1	0.278	482	77.501	185.445	5.691	1.30E-08
A2	0.651	520	-77.616	197.655	-5.521	3.38E-08

Supplementary Table 2. The genome-wide significant SNPs in Costa Rican WGS data and the replication of CAMP WGS data.

	Chr	Position	A1	A2	rsID	afreq	fam	S.E.S.	Var.S.	Z	p value
GACRS trios (N=894)	2	236932637	A	G	rs10165869	0.309	561	80.428	180.341	5.989	2.16E-09
	2	236940120	T	C	rs6725280	0.288	562	78.13	175.486	5.898	3.77E-09
	2	236935080	G	A	rs1865671	0.289	564	77.151	174.379	5.842	5.27E-09
	2	236929711	T	C	rs7607911	0.302	569	78.083	179.342	5.831	5.66E-09
	2	236908895	C	G	rs30102	0.304	584	76.225	180.017	5.681	1.37E-08
	2	236933062	A	G	rs10168628	0.336	606	76.898	196.043	5.492	3.97E-08
CAMP trios (N=286, dyspnea: 60.42%)	2	236932637	A	G	rs10165869	0.284	170	18.726	68.288	2.266	0.023445
	2	236940120	T	C	rs6725280	0.267	168	12.27	66.679	1.503	0.132928
	2	236935080	G	A	rs1865671	0.271	172	13.423	68.014	1.628	0.103599
	2	236929711	T	C	rs7607911	0.285	172	17.435	68.467	2.107	0.035107
	2	236908895	C	G	rs30102	0.275	168	16.441	67.900	1.995	0.046018
	2	236933062	A	G	rs10168628	0.338	197	16.772	79.905	1.876	0.060621
European American (N= 216, dyspnea: 61.21%)	2	236932637	A	G	rs10165869	0.308	135	13.77	56.454	1.833	0.066847
	2	236940120	T	C	rs6725280	0.287	136	8.27	55.512	1.11	0.266992
	2	236935080	G	A	rs1865671	0.288	137	8.958	55.985	1.197	0.231227
	2	236929711	T	C	rs7607911	0.306	136	12.479	56.197	1.665	0.095979
	2	236908895	C	G	rs30102	0.296	134	10.408	55.471	1.397	0.162271
	2	236933062	A	G	rs10168628	0.346	149	14.093	63.622	1.767	0.07726
African American (N= 28, dyspnea: 55.56%)	2	236932637	A	G	rs10165869	0.116	11	0.887	3.375	0.483	0.629201
	2	236940120	T	C	rs6725280	0.089	8	0.531	2.633	0.328	0.743281
	2	236935080	G	A	rs1865671	0.107	10	0.396	3.135	0.224	0.822818
	2	236929711	T	C	rs7607911	0.116	11	0.887	3.375	0.483	0.629201
	2	236908895	C	G	rs30102	0.098	9	1.022	2.873	0.603	0.546546
	2	236933062	A	G	rs10168628	0.196	14	-2.765	5.023	-1.234	0.217362

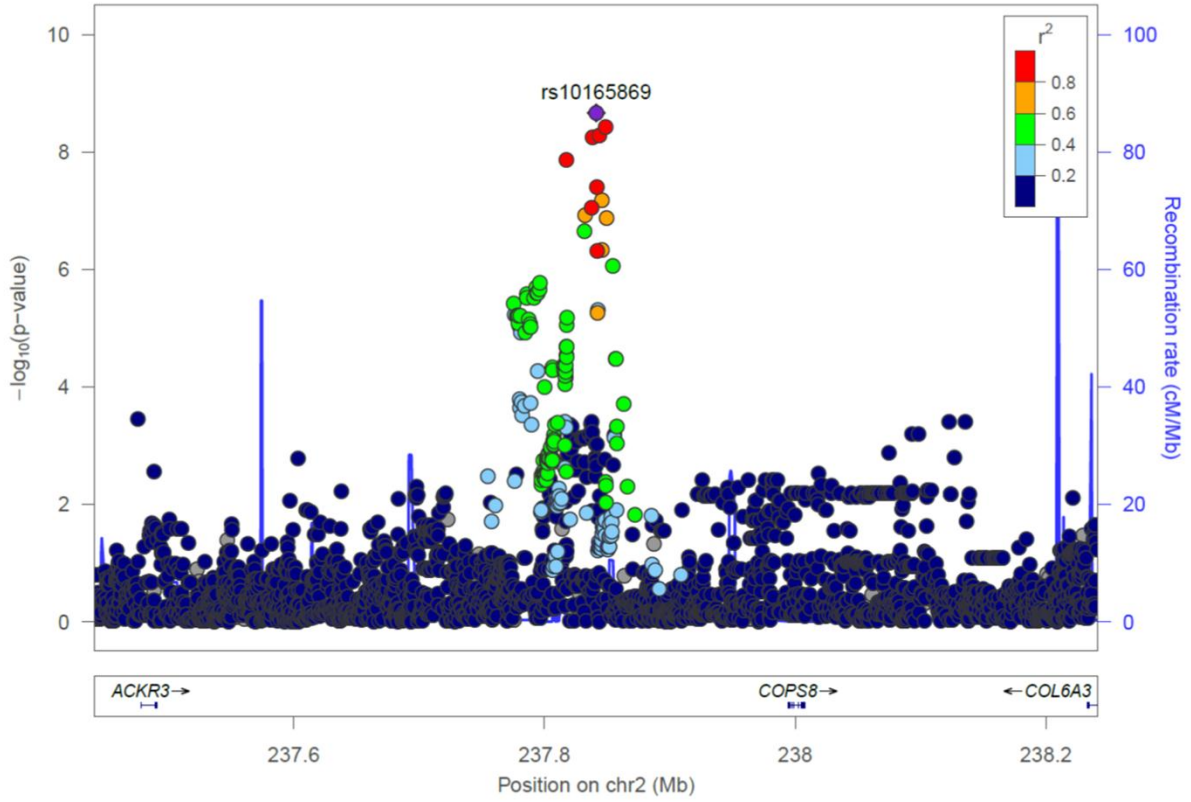
Latinos (N= 19, dyspnea: 63.16%)	2	236932637	A	G	rs10165869	0.365	10	0.719	3.004	0.415	0.678384
	2	236940120	T	C	rs6725280	0.355	10	0.118	3.079	0.067	0.946339
	2	236935080	G	A	rs1865671	0.368	11	0.719	3.44	0.387	0.698392
	2	236929711	T	C	rs7607911	0.368	11	0.719	3.44	0.387	0.698392
	2	236908895	C	G	rs30102	0.355	10	1.141	3.261	0.632	0.527503
	2	236933062	A	G	rs10168628	0.421	13	-1.496	4.047	-0.744	0.4571
Others (N= 23, dyspnea : 56.52%)	2	236932637	A	G	rs10165869	0.228	14	3.35	5.455	1.435	0.151425
	2	236940120	T	C	rs6725280	0.228	14	3.35	5.455	1.435	0.151425
	2	236935080	G	A	rs1865671	0.228	14	3.35	5.455	1.435	0.151425
	2	236929711	T	C	rs7607911	0.228	14	3.35	5.455	1.435	0.151425
	2	236908895	C	G	rs30102	0.228	15	3.87	6.294	1.542	0.12297
	2	236933062	A	G	rs10168628	0.370	21	6.94	7.213	2.584	0.009767

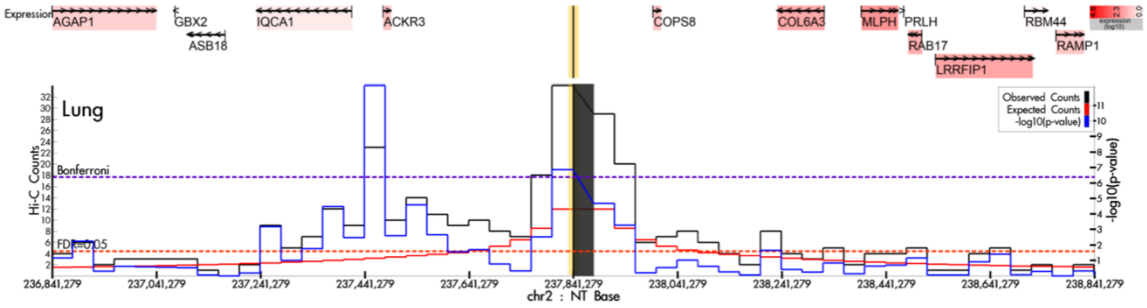
Supplementary Table 3. Top genes in chromosome 2 by summary-Predixcan analysis using lung tissues based on GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

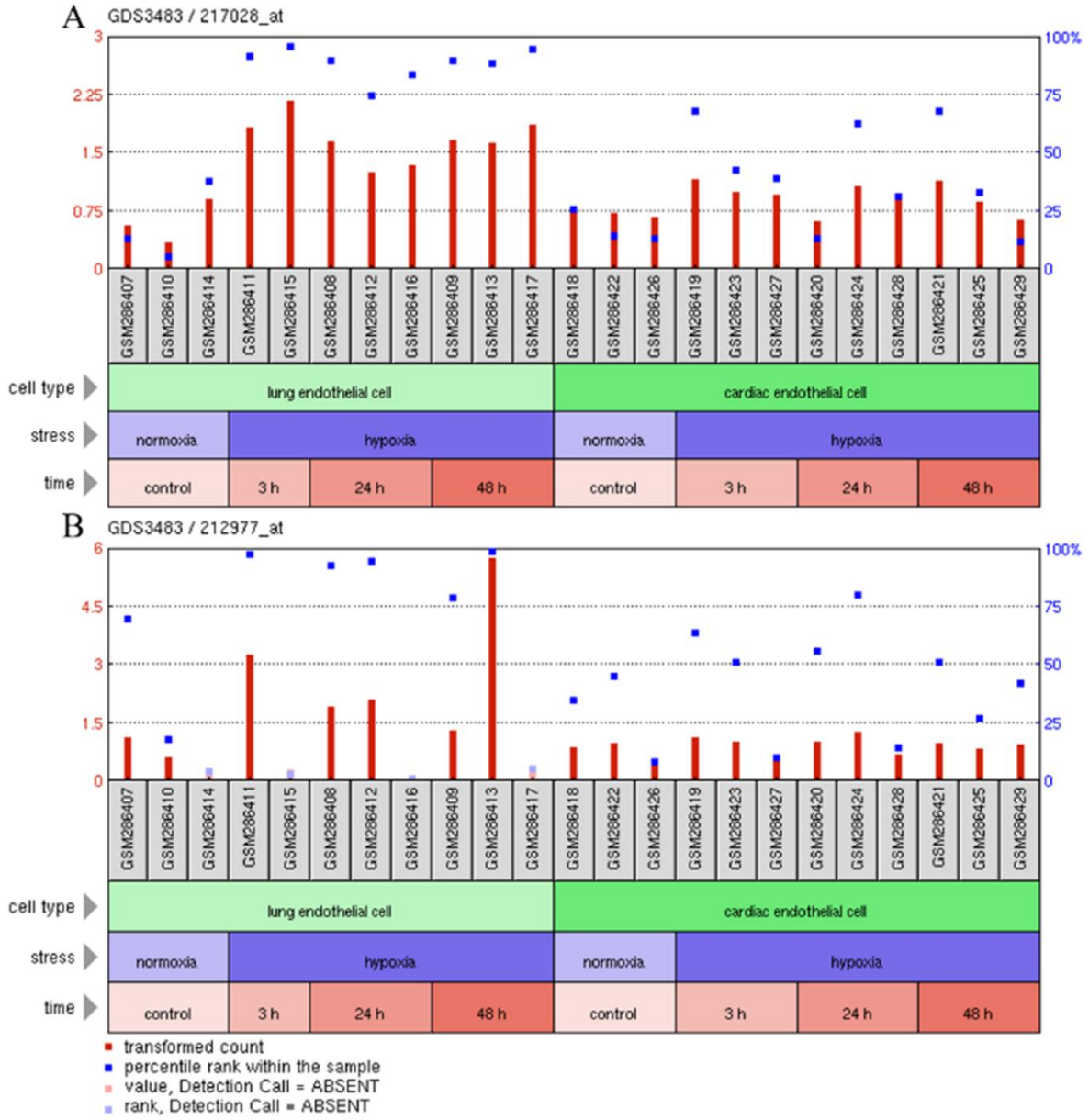
Gene	Gene name	Z	p value
ENSG00000144476.5	ACKR3	4.78	1.72E-06
ENSG00000115902.10	SLC1A4	3.38	7.27E-04
ENSG00000130294.15	KIF1A	2.89	3.83E-03
ENSG00000136689.18	IL1RN	2.87	4.08E-03
ENSG00000171150.7	SOCS5	-2.72	6.61E-03

Supplementary Table 4. Significant cis-eQTLs of top SNPs for *ACKR3* (ENSG00000144476.5) in lung tissues. Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

rsID	T-statistic	normalized effect size	p value
rs10165869_G	-2.2	-0.097	0.025
rs6725280_C	-2.6	-0.12	0.0091
rs1865671_A	-2.6	-0.11	0.0097
rs7607911_C	-2.5	-0.1	0.015
rs30102_C	1.9	0.084	0.052
rs10168628_G	-2	-0.082	0.046


rs10165869
r2
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
ACKR3
COPS8
COL6A3
237.6
237.8
238
238.2
Position on chr2 (Mb)


Expression
AGAP1
GBX2
ASB18
IQCA1
ACKR3
COPS8
COL6A3
MLPH
PRLH
RAB17
LRRFIP1
RBM44
RAMP1
Lung
Observed Counts
Expected Counts
-log10(p-value)
Hi-C Counts
-log10(p-value)
Bonferroni
FDR=0.05
236,841,279
237,041,279
237,241,279
237,441,279
237,641,279
237,841,279
238,041,279
238,241,279
238,441,279
238,641,279
238,841,279
chr2 : NT Base


A
GDS3483 / 217028_at
3
2.25
1.5
0.75
0
100%
75
50
25
0
GSM286407
GSM286410
GSM286414
GSM286411
GSM286415
GSM286408
GSM286412
GSM286416
GSM286409
GSM286413
GSM286417
GSM286418
GSM286422
GSM286426
GSM286419
GSM286423
GSM286427
GSM286420
GSM286424
GSM286428
GSM286421
GSM286425
GSM286429
cell type
lung endothelial cell
cardiac endothelial cell
stress
normoxia
hypoxia
normoxia
hypoxia
time
control
3 h
24 h
48 h
control
3 h
24 h
48 h
B
GDS3483 / 212977_at
6
4.5
3
1.5
0
100%
75
50
25
0
GSM286407
GSM286410
GSM286414
GSM286411
GSM286415
GSM286408
GSM286412
GSM286416
GSM286409
GSM286413
GSM286417
GSM286418
GSM286422
GSM286426
GSM286419
GSM286423
GSM286427
GSM286420
GSM286424
GSM286428
GSM286421
GSM286425
GSM286429
cell type
lung endothelial cell
cardiac endothelial cell
stress
normoxia
hypoxia
normoxia
hypoxia
time
control
3 h
24 h
48 h
control
3 h
24 h
48 h
transformed count
percentile rank within the sample
value, Detection Call = ABSENT
rank, Detection Call = ABSENT