Genome-Wide Interrogation of Longitudinal FEV 1 in Children With Asthma

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi:10.1164/rccm.201403-0460OC

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:42668782

Terms of Use
This article was downloaded from Harvard University’s DASH repository, WARNING: This file should NOT have been available for downloading from Harvard University’s DASH repository.;This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Genome-wide Interrogation of Longitudinal FEV₁ in Children with Asthma

Kehua Wu¹²*, Eric R. Gamazon¹*, Hae Kyung Im³, Paul Geeleher¹, Steven R. White¹, Julian Solway¹, George L. Clemmer⁴, Scott T. Weiss⁴, Kelan G. Tantisira⁴, Nancy J. Cox¹, Mark J. Ratain¹, and R. Stephanie Huang¹

¹Department of Medicine and ³Department of Health Studies, The University of Chicago, Chicago, Illinois; ²State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing, China; and ⁴Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts

Abstract

Rationale: Most genomic studies of lung function have used phenotypic data derived from a single time-point (e.g., presence/absence of disease) without considering the dynamic progression of a chronic disease.

Objectives: To characterize lung function change over time in subjects with asthma and identify genetic contributors to a longitudinal phenotype.

Methods: We present a method that models longitudinal FEV₁ data, collected from 1,041 children with asthma who participated in the Childhood Asthma Management Program. This longitudinal progression model was built using population-based nonlinear mixed-effects modeling with an exponential structure and the determinants of age and height.

Measurements and Main Results: We found ethnicity was a key covariate for FEV₁ level. Budesonide-treated children with asthma had a slight but significant effect on FEV₁ when compared with those treated with placebo or nedocromil (P < 0.001). A genome-wide association study identified seven single-nucleotide polymorphisms nominally associated with longitudinal lung function phenotypes in 581 white Childhood Asthma Management Program subjects (P < 10⁻⁶) in the placebo ["discovery"] and P < 0.05 in the nedocromil treatment ["replication"] group. Using ChiP-seq and RNA-seq data, we found that some of the associated variants were in strong enhancer regions in human lung fibroblasts and may affect gene expression in human lung tissue. Genetic mapping restricted to genome-wide enhancer single-nucleotide polymorphisms in lung fibroblasts revealed a highly significant variant (rs676931; P = 4 × 10⁻⁶; false discovery rate < 0.05).

Conclusions: This study offers a strategy to explore the genetic determinants of longitudinal phenotypes, provide a comprehensive picture of disease pathophysiology, and suggest potential treatment targets.

Keywords: asthma; NONMEM; longitudinal model; FEV₁

(Received in original form March 11, 2014; accepted in final form August 3, 2014)

*These authors contributed equally to this work.

Supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH: UL1RR024999), NIH/NIGMS grant U01GM61393, and by NIH/NIGMS grant K08GM089941. R.S.H. also received support from NIH/NCI grant R21CA139278, University of Chicago Cancer Center Support Grant (#P30CA14599), and Breast Cancer SPORE Career Development Award (CA125183). E.R.G. and N.J.C. are supported by R01 MH1011820 and R01 MH009037. S.T.W. and K.G.T. are supported by U01 HL065999 and R01 NR013391. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the NIH (commonfund.nih.gov/GTEx). Additional funds were provided by the NCI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI/SAIC-Frederick, Inc. (SAIC-F) subcontracts to the Broad Institute, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center was funded through a contract (HHSN268201000025C) to the Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplement to University of Miami grant DAO06227. The datasets used for the analyses described in this manuscript were obtained from dbGaP through dbGaP accession number phs000424.v3.p1.c1 on 12/13/2013.


Correspondence and requests for reprints should be addressed to R. Stephanie Huang, Ph.D., 900 East 57th Street, KCBID Room 7148, The University of Chicago, Chicago, IL 60637. E-mail: rhuang@medicine.bsd.uchicago.edu

This article has an online supplement, which is accessible from this issue’s table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 190, Iss 6, pp 619–627, Sep 15, 2014
Copyright © 2014 by the American Thoracic Society
Originally Published in Press as DOI: 10.1164/rccm.201403-0460OC on August 5, 2014
Internet address: www.atsjournals.org

Wu, Gamazon, Im, et al.: Genomic Study of Longitudinal Lung Function 619
After one cycle). This binary time or response/no-response to treatment, assuming a single time-point (e.g., the binary date, the phenotype is typically defined as the presence or absence of disease at a given time point in an individual’s baseline lung function and change in lung function over time, which is independent of potentially confounding stature-related factors. To our knowledge, only a handful of studies have attempted longitudinal modeling of FEV\textsubscript{1} over time. In our study, the “longitudinal phenotype” refers to the fact that these phenotypes are derived from a longitudinal modeling approach. These phenotypes were independent of confounding anthropometric factors and characterize the development of lung function in these patients. We hypothesized that our method, which integrates GWAS and longitudinal modeling, may facilitate a better understanding of lung function in children with asthma over time than a standard analysis of a snapshot phenotype as currently used.

**Methods**

**Study Design**

This study was performed using the Childhood Asthma Management Program (CAMP) study dataset (15). Demographic information for the CAMP study participants is shown in Table 1. The CAMP study was approved by the Institutional Review Board in all eight study sites. Informed consent and assent were obtained from the participants and their guardians before enrolment (16).

**Lung Function over Time and Drug Effect Model Development**

Longitudinal FEV\textsubscript{1} (data collected every 2–4 mo in 4-yr period of time, before administering albuterol) was fitted to a nonlinear mixed-effects model with extended least squares regression using the NONMEM (Ellicott City, MD) program. Details regarding model development and covariate selection can be found in the METHODS section of the online supplement.

In this study, we develop a method that applies nonlinear mixed-effects modeling (NONMEM) to longitudinal FEV\textsubscript{1} observations with the goal of describing lung function in children with asthma in the presence and absence of treatment. Furthermore, GWAS was performed to identify genetic predictors of baseline lung function and the rate of change in FEV\textsubscript{1} level over time. In our study, the “longitudinal phenotypes” used for GWAS were the estimates of the FEV\textsubscript{1} baseline level and the rate of change in FEV\textsubscript{1} with age for each patient derived from the population-based model. (We use the term “longitudinal phenotypes” throughout to refer to the two phenotypes even though the FEV\textsubscript{1} baseline level is defined at a particular [initial] time-point; “longitudinal” refers to the fact that these phenotypes were derived from a longitudinal modeling approach.) These phenotypes were independent of confounding anthropometric factors and characterize the development of lung function in these patients. We hypothesized that our method, which integrates GWAS and longitudinal modeling, may facilitate a better understanding of lung function in children with asthma over time than a standard analysis of a snapshot phenotype as currently used.

**At a Glance Commentary**

**Scientific Knowledge on the Subject:** Genome-wide association studies of single-time-point phenotypes have been conducted for the discovery of genetic determinants of lung function and/or asthma risk in large human populations. Several single-nucleotide polymorphisms have been found to be reproducibly associated with pulmonary function.

**What This Study Adds to the Field:** We developed an integrative method that combines mixed-effects longitudinal modeling and a genome-wide association study. This approach may facilitate a better understanding of the progression of a complex phenotype over time and enable improved discovery of genetic variants associated with a dynamic phenotype.

Genome-wide association studies (GWASs) have greatly contributed to the identification of genes and genetic variants conferring susceptibility to complex diseases and other heritable traits, such as lung function (1–4). For example, a metaanalysis of GWASs has implicated several independent loci for association with forced expiratory volume in 1 second (FEV\textsubscript{1}) in the general population (5, 6). Some genomic loci were found to influence pulmonary function in several populations of patients with asthma (7). A subsequent study demonstrated that genes involved in airway remodeling were associated with lung function both in general populations and in patients with asthma (8). Using GWAS to identify the genetic factors associated with lung function, we hypothesized that our method, which integrates GWAS and longitudinal modeling, may facilitate a better understanding of lung function in children with asthma over time than a standard analysis of a snapshot phenotype as currently used.

**GWAS**

The FEV\textsubscript{1} for each patient was calculated with equation 1 using NONMEM. The individual-level parameters (theta\textsubscript{1} and theta\textsubscript{2}; see equation 3 in RESULTS section) derived from the longitudinal model were the primary phenotypes used, which quantified the rate of change in FEV\textsubscript{1} with age and the baseline FEV\textsubscript{1} level for each
from the Illumina Human BodyMap 2.0 of human tissues, we used RNA-Seq data expression of implicated genes in a variety of enhancers (20). To evaluate the relative candidate target genes for the implicated asthma and control subjects to determine of differential expression between (atopic) data GSE18965 from a microarray study of phenotypes, we also conducted GWAS on single-nucleotide polymorphisms (SNPs) performed simulations (n = 100) using strong enhancers in lung analysis for those variants mapping to \( P \) distribution of \( (Q-Q) \) plot, the expected and observed values from the pooled places in normal human lung section of the online supplement.

We annotated the GWAS-identified variants with chromatin status using ChromHMM (19). The WashU Epigenome Browser was used for visualization. Using the R package limma, we reanalyzed GEO data GSE18965 from a microarray study of differential expression between (atopic) airway epithelial cells from subjects with asthma and control subjects to determine candidate target genes for the implicated enhancers (20). To evaluate the relative expression of implicated genes in a variety of human tissues, we used RNA-Seq data from the Illumina Human BodyMap 2.0 project (21). For details, see the METHODS section of the online supplement.

We compared, in a quantile–quantile (Q–Q) plot, the expected and observed distribution of \( P \) values from the pooled analysis for those variants mapping to strong enhancers in lung fibroblast. We also performed simulations (\( n = 100 \)) using single-nucleotide polymorphisms (SNPs) that match the allele frequency and distance to nearest gene of the SNPs that overlap with the strong enhancers and generated the corresponding Q–Q plot for each simulation. We used a false discovery rate (FDR)-based multiple testing correction (22); FDR less than 0.05 was used to declare a significant association.

**Genetic Association with Longitudinal versus Single–Time–Point Phenotype Using Chromatin Profiling Data and the National Human Genome Research Institute Catalog**

Using the SNPs overlapping the chromatin states in normal human lung fibroblast (NHLF), we evaluated the gain in statistical power to detect a quantitative trait locus from the pooled analysis of the longitudinal phenotypes in relation to the single-time-point phenotypes.

We also compared the \( P \) values from each set of phenotypes for the SNPs that have been found to be reproducibly associated with lung function as curated in the National Human Genome Research Institute (NHGRI) catalog of published GWAS.

**Results**

**Longitudinal Model of FEV\(_1\) in Children with Asthma**

Both exponential and linear structural models were evaluated using the CAMP prebronchodilator FEV\(_1\) data. Before testing potential clinical covariates, we established a base model by evaluating age, body weight, height, sex, and body mass index as potential determinants of FEV\(_1\) change over time. Akaike information criterion (AIC) was used to compare the reference structural models (see Table E1). The best prediction for FEV\(_1\) was achieved by an exponential function of age and height:

\[
\text{FEV}_1 = \exp(\theta_1 \times \text{age} + \theta_2 \times \text{height} - \theta_3) + \text{thetaDrugEffect}
\]  

In equation 3, \( \theta_1 \) and \( \theta_2 \) are the rate of change in FEV\(_1\) associated with age and height, respectively. The \( \theta_3 \) refers to a baseline level for FEV\(_1\) (i.e., the FEV\(_1\) level at birth assuming the model is applicable to that age range). We evaluated the model that assumes a different definition of baseline level (defined at the mean age of 9, rather than age at birth) and found that the original \( \theta_3 \) and the new \( \theta_3 \) were significantly correlated (Spearman correlation of 0.97; \( P < 2.2 \times 10^{-16} \)). A comparison of the two models can be found in the METHODS section of the online supplement and Table E2.

The fit of the base model (including age and height) was not improved by the addition of sex, body weight, and body mass index in children with asthma (\( P > 0.05 \)). Using the base model, additional covariates (listed in Table 1) were evaluated. Among them, only race was found to be a key covariate. The FEV\(_1\) level was similar between whites and Mexican Americans, but it was significantly lower in African American children with asthma. Furthermore, we analyzed the model that

---

**Table 1. Demographic Information of the Patients at the Time of Enrollment**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Placebo</th>
<th>Nedocromil</th>
<th>Budesonide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>418</td>
<td>312</td>
<td>311</td>
</tr>
<tr>
<td>Age, yr</td>
<td>9.0 ± 2.2</td>
<td>8.8 ± 2.1</td>
<td>9.0 ± 2.1</td>
</tr>
<tr>
<td>Sex, male/female</td>
<td>234/184</td>
<td>206/106</td>
<td>181/130</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>292 (69.9)</td>
<td>218 (69.9)</td>
<td>201 (64.6)</td>
</tr>
<tr>
<td>African American</td>
<td>56 (13.4)</td>
<td>38 (12.2)</td>
<td>44 (14.1)</td>
</tr>
<tr>
<td>Mexican American</td>
<td>37 (8.9)</td>
<td>29 (9.3)</td>
<td>32 (10.3)</td>
</tr>
<tr>
<td>Others</td>
<td>33 (7.9)</td>
<td>27 (8.7)</td>
<td>34 (10.9)</td>
</tr>
<tr>
<td>Height, cm</td>
<td>55.3 ± 28.8</td>
<td>56.0 ± 28.7</td>
<td>56.8 ± 28.0</td>
</tr>
<tr>
<td>Body weight, kg</td>
<td>42.0 ± 16.0</td>
<td>42.2 ± 16.0</td>
<td>42.9 ± 16.8</td>
</tr>
<tr>
<td>Age of first symptoms, yr</td>
<td>3.0 ± 2.6</td>
<td>3.1 ± 2.4</td>
<td>3.1 ± 2.3</td>
</tr>
<tr>
<td>Year since diagnosis of asthma</td>
<td>4.9 ± 2.7</td>
<td>5.0 ± 2.7</td>
<td>5.2 ± 2.6</td>
</tr>
<tr>
<td>Maternal asthma, no/yes</td>
<td>301/102</td>
<td>224/81</td>
<td>225/79</td>
</tr>
<tr>
<td>Paternal asthma, no/yes</td>
<td>301/80</td>
<td>243/55</td>
<td>219/73</td>
</tr>
<tr>
<td>Vitamin D levels</td>
<td>36.7 ± 15.2</td>
<td>37.1 ± 16.8</td>
<td>39.9 ± 14.9</td>
</tr>
<tr>
<td>Mother smoked while pregnant, no/yes</td>
<td>349/65</td>
<td>266/46</td>
<td>271/39</td>
</tr>
<tr>
<td>Mother, dad, or other smoked after birth, no/yes</td>
<td>57/361</td>
<td>42/270</td>
<td>37/274</td>
</tr>
</tbody>
</table>

---

Wu, Gamazon, Im, et al.: Genomic Study of Longitudinal Lung Function 621
assumes an interaction between age and sex. Based on AIC, this interaction model did not perform as well as the base model (see Table E1).

Evaluating the different treatment arms, we observed that the additive drug effect model fit the data better than a proportional model, with lower values of AIC (see Table E1) and objective function value (OFV). We found that budesonide had a minor but statistically significant effect on the prebronchodilator FEV1 (additive model theta_drug effect, $P < 0.001$). The FEV1 with long-term treatment of budesonide was predicted to be $0.103 \pm 0.129$ higher than FEV1 in the placebo group. Nedocromil did not show any significant effect on FEV1 when compared with the placebo group. We observed no significant treatment effect on the rate of change in FEV1 with age. We tested the interaction between age and treatment (see Table E1) and found this model, based on AIC, to perform less optimally than our selected model.

The interindividual variability of theta_2 was very small ($<10^{-6}$), suggesting the rate of change in FEV1 with height was similar among the subjects. Therefore, in our final model, theta_2’s interindividual variability was fixed to zero. Figure 1 illustrates the relationship between the observed and population-predicted FEV1, and the relationship between observations and individual predicted FEV1 values using our final model. Most of the conditional weighted residuals were evenly distributed around 0 (Figure 1C).

To assess the stability of our model, bootstrapping validation analysis was performed. This analysis showed that the median parameter values and the corresponding relative standard error resulting from the bootstrapping agreed with the estimates from our final model (Table 2), suggesting that the final model fitted FEV1 observations reasonably well and was stable. Most of the observed FEV1 fell within the 5th–95th percent prediction

Table 2. The Estimates of the Parameters from the Final Model (Equation 3)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>NONMEM Estimate</th>
<th>RSE (%)</th>
<th>Bootstrap Estimate</th>
<th>RSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theta_1</td>
<td>Rate of change associated with age</td>
<td>0.0137</td>
<td>16.1</td>
<td>0.0138</td>
<td>16.8</td>
</tr>
<tr>
<td>Theta_2</td>
<td>Rate of change associated with height</td>
<td>0.0169</td>
<td>2.20</td>
<td>0.0169</td>
<td>2.30</td>
</tr>
<tr>
<td>White/Mexican American</td>
<td>Intercept among the ethnic groups</td>
<td>1.89</td>
<td>1.78</td>
<td>1.89</td>
<td>1.84</td>
</tr>
<tr>
<td>African American</td>
<td></td>
<td>2.04</td>
<td>0.575</td>
<td>2.04</td>
<td>0.641</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>1.95</td>
<td>0.742</td>
<td>1.94</td>
<td>0.787</td>
</tr>
<tr>
<td>Theta_drug effect</td>
<td>Drug effect of budesonide</td>
<td>0.103</td>
<td>14.1</td>
<td>0.108</td>
<td>14.4</td>
</tr>
</tbody>
</table>

Interindividual variability

(3) (shrinkage %)†

| Theta_1                     | 0.00722 (32.9)                                  | 0.00723          |
| Theta_2                     | 0.0912 (26.4)                                   | 0.0915           |
| Theta_drug effect           | 0.129 (67.3)                                    | 0.123            |

Intraindividual variability

(shrinkage %)‡

| Proportional (CV%)          | 5.91 (4.52)                                     | 5.91             |
| Additive (SD)               | 0.0863 (4.52)                                   | 0.0859           |

Definition of abbreviations: CV = coefficient of variation; NONMEM = nonlinear mixed-effects modeling; RSE = relative standard error.

†Percent RSE (100% × standard error/estimate).

‡The SD across the subjects for each parameter.

 warned that the CO2 level in the room was below the recommended level of 0.5%.

The CO2 levels in the room were monitored at regular intervals using a CO2 meter. The meter displayed an average CO2 level of 0.3% which was considered safe for the duration of the experiment. The CO2 meter was placed at a central position in the room to ensure accurate readings. The CO2 levels were also recorded at the beginning and end of the experiment to establish any potential changes. Throughout the experiment, the CO2 levels were consistently monitored and recorded to ensure a safe environment for the participants.

Figure 1. Goodness of fit of final model. (A) Relationship between observed FEV1 and population typical FEV1 predictions. (B) Relationship between observed FEV1 and individual FEV1 predictions. (C) Conditional weighted residuals (WRES) versus age. Most conditional WRES evenly distributed around 0. The solid lines are diagonal lines of identity in A and B.
GWAS

Because of the observed ancestry effect on FEV1 level and the potential for population stratification (and the small sample size of the African-American and Hispanic cohorts), we chose to focus on white subjects enrolled in the CAMP study for genome-wide association analysis.

A GWAS was performed on the pooled samples with treatment as a covariate and in the separate placebo (292 subjects) and nedocromil (218 subjects) groups, because there were no observed differences in longitudinal FEV1 between these two groups. Association analyses between 473,680 genotyped SNPs (which passed quality control) and the two modeled parameters (theta1 and theta3, equation 3) were performed. Forty-six and 53 SNPs were nominally associated with theta1 (P < 1 × 10^{-5}) and theta3 (P < 1 × 10^{-4}), respectively, in the placebo group. For replication, these SNPs were further evaluated (P < 0.05 and concordance of effect) in the nedocromil group. Six SNPs for theta3 (baseline FEV1 level) and one SNP (rs17161791 for theta1) were found to be associated with lower theta3 leading to higher FEV1. The variant allele of rs17161791 was associated with higher theta1 resulting in higher FEV1 increase rate. Not surprisingly, the association P values for the model parameters (theta1 and theta3) at all seven SNPs were improved in the pooled data analysis (placebo + nedocromil) relative to the placebo-alone analysis (Table 3).

Functional Evaluation of Top GWAS Associations

To functionally characterize our top SNP associations, we used ChromHMM (23) applied to ENCODE data (19, 24) from NHLF and a lymphoblastoid cell line (GM12878). We found that rs6763931 (intronic to ZBTB38 gene) overlaps a strong enhancer state in both NHLF and GM12878 (Figure 3). Consistent with this, we observed that the same SNP coincides with an active transcription start site in fetal lung fibroblast cells (IMR90; see Figure E2). Differential expression analysis of the genes at this locus between children with atopy and with asthma and nonatopic healthy individuals identified a nearby gene (103 kb away), RASA2, that was highly differentially expressed (P = 0.002; see Figure E3); in contrast, the host gene (ZBTB38) showed no evidence of differential expression.

Table 3. GWAS Results

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr (Location)</th>
<th>Allele* (Frequency)</th>
<th>Gene</th>
<th>Placebo</th>
<th>Nedocromil</th>
<th>Placebo + Nedocromil</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs347412</td>
<td>13 (intron)</td>
<td>A/G (0.5435)</td>
<td>DGKH</td>
<td>1.39 × 10^{-5} (0.03)</td>
<td>0.0249 (0.019)</td>
<td>1.42 × 10^{-6} (0.026)</td>
</tr>
<tr>
<td>rs238349</td>
<td>13 (intron)</td>
<td>C/A (0.5213)</td>
<td>DGKH</td>
<td>1.42 × 10^{-5} (0.034)</td>
<td>0.0497 (0.016)</td>
<td>3.54 × 10^{-6} (0.025)</td>
</tr>
<tr>
<td>rs559389</td>
<td>11 (intergenic)</td>
<td>T/C (0.5762)</td>
<td>—</td>
<td>5.28 × 10^{-5} (0.030)</td>
<td>0.0428 (0.016)</td>
<td>9.28 × 10^{-6} (0.023)</td>
</tr>
<tr>
<td>rs9366309</td>
<td>6 (intergenic)</td>
<td>C/T (0.6333)</td>
<td>—</td>
<td>3.32 × 10^{-5} (0.033)</td>
<td>0.0348 (0.017)</td>
<td>7.05 × 10^{-6} (0.024)</td>
</tr>
<tr>
<td>rs6763931</td>
<td>3 (intron)</td>
<td>G/A (0.5675)</td>
<td>ZBTB38</td>
<td>5.90 × 10^{-5} (0.031)</td>
<td>0.0107 (0.020)</td>
<td>4.05 × 10^{-6} (0.024)</td>
</tr>
<tr>
<td>rs2304725</td>
<td>3 (synonymous)</td>
<td>T/C (0.7091)</td>
<td>SLC6A11</td>
<td>3.87 × 10^{-5} (0.033)</td>
<td>0.0270 (0.019)</td>
<td>3.04 × 10^{-6} (0.026)</td>
</tr>
<tr>
<td>rs17161791</td>
<td>7 (intergenic)</td>
<td>T/C (0.7303)</td>
<td>—</td>
<td>3.01 × 10^{-5} (0.002)</td>
<td>0.0249 (0.001)</td>
<td>1.56 × 10^{-6} (0.002)</td>
</tr>
</tbody>
</table>

Definition of abbreviations: GWAS = genome-wide association studies; SNP = single-nucleotide polymorphism.
*The first allele is the common allele in whites followed by its frequency; bolded allele indicates effect allele.
†The last SNP is associated with theta1, the other six SNPs are associated with theta3.
(P = 0.21; see Figure E4) (20). (In all, four genes [RASA2, ZBTB38, RNF7, and SLC25A36] in a 1-Mb region centered at the SNP rs6763931 were tested; thus the differential expression of RASA2 meets Bonferroni significance.) The nongenic SNP rs559389, another top association, is in strong linkage disequilibrium ($r^2 = 0.80$ in 1,000 Genomes EUR) with variants (rs538322, rs3018303, and rs12366105) that also overlap regions of strong enhancer histone marks in NHLF.

Finally, we tested our top SNPs for association with expression in a variety of tissues. We found that rs238349 is a cis-acting expression quantitative trait locus in lung for diacylglycerol kinase, eta ($DGKH$; $P = 5.2 \times 10^{-5}$), using public GTEx RNA-seq data (Broad) (25). $DGKH$ is most highly expressed in prostate and lung in a comparison of 16 human tissues (see Table E3).

Taken together, these results provide strong evidence that our top SNPs are likely to mediate their phenotypic effect via transcriptional mechanisms in lung fibroblast and/or an immune-related tissue.

**Genetic Association with Longitudinal versus Single–Time-Point Phenotype**

*Comparison for genetic variants in enhancer regions in lung fibroblast cells.* Because of the small sample size, we did not expect any SNPs to reach genome-wide significance according to a conservative Bonferroni adjustment. Remarkably, the Q-Q plot in the pooled analysis restricted to regions enriched for functional SNPs (e.g., see Figure 4 for the Q-Q plot of SNPs overlapping strong enhancer regions in human lung fibroblast; n = 10,751 interrogated SNPs) showed a highly significant association (rs6763931; FDR < 0.05; $P = 4.05 \times 10^{-6}$) with theta3.

As expected, GWAS of single–time-point phenotypes in the pooled dataset yielded no genome-wide significant findings, nor was there a Bonferroni-adjusted significant association with any single–time-point phenotypes among the SNPs in strong enhancer regions in NHLF.

**Figure 3.** Regulatory function of rs6763931. The single-nucleotide polymorphism rs6763931 (located in an intron of ZBTB38, black arrow) overlaps a strong enhancer in normal human lung fibroblast (NHLF) and in a lymphoblastoid cell line (GM12878).

**Figure 4.** Quantile–quantile plot for the single-nucleotide polymorphisms (SNPs) overlapping strong enhancers. A highly significant association (rs6763931, FDR < 0.05) was identified using SNPs in regulatory regions in human lung fibroblast. The $P$ values obtained from associations of the enhancer SNPs with theta3 are shown as circles. The horizontal line at 0.05/N (significance level after Bonferroni correction) is also shown. FDR = false discovery rate.
Table 4, for example, shows a comparison of the association results between our longitudinal phenotypes and FEV$_1$ at 48 months (single time-point) for SNPs that intersect strong enhancer states in NHLF. Furthermore, none of the simulated datasets (n = 100) showed a significant association (FDR < 0.05).

Comparison for pulmonary function associated SNPs curated in the NHGRI catalog. The NHGRI GWAS catalog (26) lists more than 50 SNPs that have been found to be associated with at least one of the pulmonary functional terms (represented by FEV$_1$ or FEV$_1$/FVC or forced expiratory flow). Of these, 23 were genotyped in the CAMP dataset (see Table E4). Consistent with the observed improvement to detect a significant association with a longitudinal phenotype, but not with a single-time-point phenotype, using SNPs that overlap regulatory regions, no reproducible lung function–associated SNP as curated in NHGRI catalog (26) showed a nominally significant association (P < 0.05) (Table 4) with the single-time-point phenotypes; in contrast, we observed five SNP associations with our longitudinal phenotypes. These include rs4762767 and rs58667 for theta$_1$ and rs1291183, rs12984174, and rs2571445 for theta$_3$.

Discussion

In this study, we developed an integrative method that combines population-based mixed-effects modeling and GWAS to identify SNPs that may contribute to baseline FEV$_1$ or rate of change in FEV$_1$ with age in children with asthma. Current approaches to finding disease susceptibility loci are primarily based on single-time-point phenotypes yielding results that reflect only a snapshot of the dynamic biology of disease. These approaches, which focus on limited observations, are prone to bias. Population-based mixed-effects modeling, used in this study, considers all FEV$_1$ observations by concurrently fitting them. The random errors can be appropriately accounted for by repeated measures per subjects (9). Additionally, FEV$_1$ is highly dependent on the pathophysiologic condition in children (11–14). A significant association between the genotype and FEV$_1$ observed at a selected time-point may be related to a confounding factor (e.g., transient bronchospasm) rather than lung function. The individual-level parameters from our model (used as phenotypes here), which are defined independently of the clinical and pathophysiologic terms, reflect baseline lung function level and lung function progression in this disease setting resulting in a more appropriate and comprehensive understanding of lung function in children with asthma. The approach developed here can be extended to other diseases and/or drug effect (27) with longitudinal data. Sikorska and coworkers (28) proposed a two-step integrative method using the classical linear mixed-effect model as reference. Our method applies nonlinear mixed-effect modeling, and both linear and nonlinear models were tested.

Demonstrating the benefit of using longitudinal modeling of disease, we observed that budesonide had a small but significant effect on FEV$_1$ with ΔOFV of 19.34 when using mixed-effects modeling in FEV$_1$ over a period of up to 6 years. This treatment benefit was not reported in the previous CAMP study [2], when % FEV$_1$ change after bronchodilator at a selected time-point was compared with that of baseline. The initial CAMP study nevertheless concluded that inhaled corticosteroids, such as budesonide, are still useful, because they provide better control of asthma resulting in fewer hospitalizations and urgent care visits to a caregiver and reduced albuterol treatment for symptoms (15). Our study found the justification for the conclusion, because we observed a significant treatment effect of budesonide on FEV$_1$ level in children with asthma by using disease progression modeling. In our study, we did not detect any drug effect for nedocromil; indeed, the inclusion of drug effect did not give a significant change in OFV. This is consistent with the previous report based on the same dataset (15). The longitudinal modeling described here may provide a powerful tool to capture the long-term drug effect over time. When applied in clinical trials, our approach may detect additional drug effects that would be missed by the standard single–time-point strategy. This is consistent with some other previous reports in both healthy subjects and subjects with asthma (13, 14).

We have demonstrated that age and height are the essential physiologic determinants of lung function growth in children with asthma, and race is a key covariate for FEV$_1$ level. Hankinson and coworkers (11) reported lung function reference values in nonsmokers between the ages of 8 and 80 years. They also found age and height were predictive variables for FEV$_1$ function in children, consistent with the model developed in this study. Interestingly, in their study, whites and Mexican Americans had similar FVC and FEV$_1$. The values for FVC and FEV$_1$ were higher than in African Americans, which was corroborated by our results. Another study reported that age, height, and body weight were predictors of FEV$_1$ in patients with asthma ranging in age from 6 to 88 (14). However, in our study, body weight

Table 4. Comparison between Longitudinal Phenotype and Single–Time-Point Phenotype for SNPs Located in Strong Enhancer States in Human Lung Fibroblast and SNPs Known to Be Reproducibly Associated with Lung Function as Curated in the NHGRI Catalog

<table>
<thead>
<tr>
<th>Analysis Using Enhancer SNPs in Lung Fibroblast</th>
<th>Analysis Using Known Lung Function–associated SNPs in the NHGRI Catalog</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDR &lt; 0.05</td>
<td>FDR &lt; 0.10</td>
</tr>
<tr>
<td>Longitudinal</td>
<td>Bonferroni (0.05/N)</td>
</tr>
<tr>
<td>FEV$_1$ at 48 mo</td>
<td>P &lt; 0.05</td>
</tr>
<tr>
<td>Permutted longitudinal data (n = 100)</td>
<td>Longitudinal</td>
</tr>
<tr>
<td></td>
<td>FEV$_1$ at 48 mo</td>
</tr>
</tbody>
</table>

Definition of abbreviations: FDR = false discovery rate; NHGRI = National Human Genome Research Institute; SNP = single-nucleotide polymorphism.

Wu, Gamazon, Im, et al.: Genomic Study of Longitudinal Lung Function
did not improve the goodness of fit (ΔOFV <3.84; P > 0.05 after addition) likely because of the high correlation between height and body weight in children. Both exponential and linear models were used to describe the time course of FEV₁ in previous studies (11–14). Exponential function had the best fit for our model (see Table E1). Sex was also included in a previous report (12), although it did not show a significant difference in our study (ΔOFV <3.84, after addition).

We identified seven SNPs nominally associated with the modeled FEV₁ parameters in both placebo and nedocromil groups. Of these, two intronic SNPs (rs347412 and rs238349 in the DKGH gene) and another SNP (rs2304725 in SLC6A11 gene) have been reported to be associated with smoking cessation (29). Another two SNPs have been reported to be associated with height, namely rs347412 (30) and rs6763931 (P = 5.90 × 10⁻⁵) (31–33). The latter one has also been implicated in growth impairment (34). We interpret these findings to indicate that the SNPs we identified may also be implicated in physiologic growth on a macro scale in children, an observation that needs to be further validated.

GWAS of single–time–point phenotypes have been conducted for the discovery of genetic determinants of lung function and/or asthma risk in large human populations. Of the SNPs associated with pulmonary function in the NHGRI catalog (26), 23 were genotyped in the CAMP dataset. Of these, we found five SNPs were suggestively associated with baseline level and progression of FEV₁ (P < 0.05) in our study; rs4762767 and rs58667 for theta₁, rs1291183, rs12984174, and rs2571445 for theta₃. Specifically, rs1291183, located within the gene YEST1 on chromosome 18, was previously reported to be associated with percent predicted FEV₁ and percent predicted FVC (P = 3.54×10⁻⁸ and 5.47×10⁻⁵, respectively) in populations of European descent with asthma (8). This SNP was associated with our longitudinal phenotype (theta₃ at P < 0.05) in placebo and nedocromil treatment groups. In addition, the SNPs rs2571445 and rs58667 were reported to be associated with FEV₁ or percent predicted FEV₁ with P = 1.11×10⁻¹² and 3.95×10⁻⁷ in individuals of European ancestry (with rs58667, the association was seen in patients with asthma of European ancestry) (5, 8). rs12984174 was previously reported to be associated with percent predicted FVC (8.89×10⁻⁶) in subjects with asthma (8), as was rs4762767 with pulmonary function as measured by FEV₁/FVC (35).

Besides replicating these five previously reported SNPs, our study discovered seven additional loci that may be linked to lung function progression in asthma, suggesting improved biologic discovery from a longitudinal modeling approach. Indeed, trait mapping using a longitudinal phenotype, but not the single–time–point phenotypes, restricted to genetic variants that overlap enhancer regions in lung fibroblast identified a highly significant association. We should note that we were still underpowered to detect associations with the rate of change in FEV₁ versus the baseline level even with the use of functional data. However, both theta₁ and theta₃ showed improved replication relative to single–time–point phenotypes with respect to previously identified loci (as found in the NHGRI catalog). Longitudinal molecular–level and gene expression investigations in relevant cell types may further improve biologic discovery.

One limitation of our study, common in asthma genetic studies, is the small sample size, which results in only nominally significant findings from the GWAS. However, the use of functional and epigenomic datasets in relevant cell types to prioritize genetic variants allowed us to discover loci that pass Bonferroni significance (Figure 4), and our longitudinal approach (in contrast to the use of single–time–point phenotypes) enabled us to confirm lung function loci previously identified by other GWAS. One notable finding from the molecular and epigenomic datasets used here is the differential expression in children with atopy with asthma versus control subjects of an adjacent gene (RASA2), and not the host gene (ZBTB38), to the regulatory variant rs6763931 that overlaps an enhancer region in lung fibroblast. Although the exact mechanism for this connection remains to be fully elucidated, this finding is consistent with several recent studies (36, 37) showing the distal regulatory effects (and proposing potential mechanisms), at several hundred kilobases, of (noncoding) enhancer SNPs associated with complex human phenotypes. Future studies on the functional connection between rs6763931 and RASA2 are warranted.

In summary, our study developed a genetic locus mapping approach that combines nonlinear mixed-effects longitudinal modeling of phenotype and GWAS. This integrative approach allows us to identify new SNPs associated with longitudinal lung function in childhood asthma. These may offer insights into the mechanism underlying pulmonary function regulation in subjects with asthma and may further indicate potential treatment targets.

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank the National Institutes of Health GWAS Data Repository; the Contributing Investigators who contributed the phenotype data and DNA samples from their original study; and the NHLBI in collaboration with CARE, CAMP, and ACRN Clinical Network Investigators and the associated Data Coordinating Centers for their support to these studies.

References


