ORIGINAL ARTICLE

Genome-wide Interrogation of Longitudinal FEV_1 in Children with Asthma

Kehua Wu[1,2*], Eric R. Gamazon[1*], Hae Kyung Im[3], Paul Geeleher[1], Steven R. White[1], Julian Solway[1], George L. Clemmer[4], Scott T. Weiss[4], Kelan G. Tantisira[4], Nancy J. Cox[1], Mark J. Ratain[1], and R. Stephanie Huang[1]

[1]Department of Medicine and [3]Department of Health Studies, The University of Chicago, Chicago, Illinois; [2]State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing, China; and [4]Channing 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 FEV_1 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 FEV_1 level. Budesonide-treated children with asthma had a slight but significant effect on FEV_1 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 ($P < 10^{-4}$ 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 (rs6763931; $P = 4 \times 10^{-6}$; 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; FEV_1

(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 R21 CA139278, University of Chicago Cancer Center Support Grant (#P30 CA14599), and Breast Cancer SPORE Career Development Award (CA125183). E.R.G. and N.J.C. are supported by R01 MH101820 and R01 MH090937. S.T.W. and K.G.T. are supported by U01 HL065899 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, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center was funded through a contract (HHSN268201000029C) 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 DA006227. 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.

Author contributions: Conceived and designed the study, K.W., E.R.G., H.K.I., M.J.R., and R.S.H. Performed the study, K.W., E.R.G., P.G., and R.S.H. Contributed reagents/materials/analysis tools, S.R.W., J.S., S.T.W., G.L.C., K.G.T., N.J.C., and M.J.R. Wrote the manuscript, K.W., E.R.G., and R.S.H. Edited and approved the manuscript, K.W., E.R.G., H.K.I., P.G., S.R.W., J.S., G.L.C., S.T.W., K.G.T., N.J.C., M.J.R., and R.S.H. Supervised the project, N.J.C., M.J.R., and R.S.H.

Correspondence and requests for reprints should be addressed to R. Stephanie Huang, Ph.D., 900 East 57th Street, KCBD 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

Originally Published in Press as DOI: 10.1164/rccm.201403-0460OC on August 5, 2014
Internet address: www.atsjournals.org

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_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 may, in addition, lead to a greater understanding of asthma pathophysiology. However, in most such studies conducted to date, the phenotype is typically defined assuming a single time-point (e.g., the binary classification with/without disease at a given time or response/no-response to treatment after one cycle). This binary "snapshot" phenotype reflects a single depiction of the dynamic biology, but fails to capture the progression of a disease over time of a chronic condition, such as asthma.

Longitudinal modeling of disease trajectory is a useful tool to characterize the disease progression and to identify disease-modifying effects of drugs (9, 10). In asthma, the most commonly used lung function indicator is FEV_1. This parameter is highly dependent on physiologic factors (e.g., age, body weight, height [11–14]) and pathologic conditions (e.g., severity of asthma), and is in addition partially genetically determined (5). Longitudinal modeling of FEV_1 allows for estimation of 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_1. All were performed in healthy subjects (11–13), with the exception of one study that included patients with asthma (age range, 6–88 yr). None of these modeling studies performed a genetic analysis (14).

In this study, we develop a method that applies nonlinear mixed-effects modeling (NONMEM) to longitudinal FEV_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_1 level over time. In our study, the "longitudinal phenotypes" used for GWAS were the estimates of the FEV_1 baseline level and the rate of change in FEV_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_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.

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_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.

Interindividual variability on FEV_1 was evaluated using an additive error model:

$$P_{ij} = P_{TVj} + \eta_{ij} \qquad (1)$$

where P_{ij} is the true value of the jth parameter for the ith subject. P_{TVj} is the population typical value (TV) for the jth parameter, and η_{ij} is an interindividual random effect, which quantifies the deviation of P_{ij} from P_{TVj} and is assumed to follow a normal distribution with mean of 0 and variance of ω_j^2.

Intraindividual variability was evaluated using a combined proportional and additive error model as follows:

$$FEV_{1_{obs}} = FEV_{1_{pred}} \cdot (1 + \varepsilon_1) + \varepsilon_2 \qquad (2)$$

where $FEV_{1_{obs}}$ are the observations and $FEV_{1_{pred}}$ is the corresponding model prediction. ε_1 and ε_2 are independent and normally distributed random variables with zero mean and variance of σ_1^2 and σ_2^2, respectively, that account for the residual unexplained variability. The final model was evaluated by a nonparametric bootstrap and the visual predictive check to assess the predictive performance and robustness (17, 18).

GWAS

The FEV_1 for each patient was calculated with equation 1 using NONMEM. The individual-level parameters ($theta_1$ and $theta_3$; *see* equation 3 in Results section) derived from the longitudinal model were the primary phenotypes used, which quantified the rate of change in FEV_1 with age and the baseline FEV_1 level for each

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

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

child with asthma. Genotype-phenotype associations were calculated assuming an additive genetic model. Details regarding the association analyses can be found in the METHODS section of the online supplement. For comparison with the longitudinal phenotypes, we also conducted GWAS on the single-time-point phenotypes (FEV_1 at 48 mo).

Functional Evaluation of Top Single-Nucleotide Polymorphism Associations

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:

$$FEV_1 = \exp(\text{theta}_1 \times \text{age} + \text{theta}_2 \times \text{height} - \text{theta}_3) + \text{theta}_{\text{drugeffect}} \quad (3)$$

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

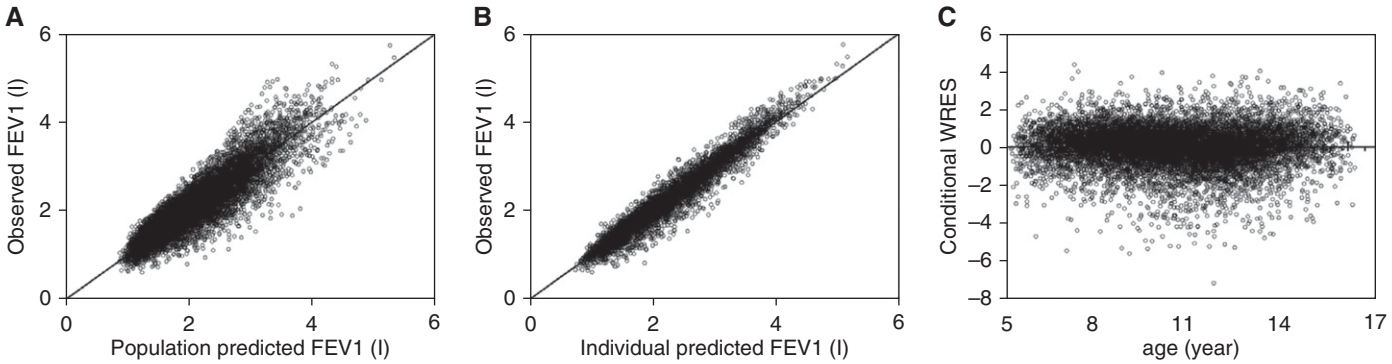

Figure 1. Goodness of fit of final model. (*A*) Relationship between observed FEV_1 and population typical FEV_1 predictions. (*B*) Relationship between observed FEV_1 and individual FEV_1 predictions. (*C*) Conditional weighted residuals (WRES) versus age. Most conditional WRES evenly distributed around 0. The *solid lines* are *diagonal lines* of identity in *A* and *B*.

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 FEV_1 (additive model $theta_{drugeffect}$, $P < 0.001$). The FEV_1 with long-term treatment of budesonide was predicted to be 0.103 ± 0.129 higher than FEV_1 in the placebo group. Nedocromil did not show any significant effect on FEV_1 when compared with the placebo group. We observed no significant treatment effect on the rate of change in FEV_1 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 FEV_1 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 FEV_1, and the relationship between observations and individual predicted FEV_1 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 FEV_1 observations reasonably well and was stable. Most of the observed FEV_1 fell within the 5th–95th percent prediction

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

Parameters	Description	NONMEM		Bootstrap	
		Estimate	RSE (%)*	Estimate	RSE (%)
$Theta_1$	Rate of change associated with age	0.0137	16.1	0.0138	16.8
$Theta_2$	Rate of change associated with height	0.0169	2.20	0.0169	2.30
$Theta_3$	Intercept among the ethnic groups				
White/Mexican American		1.89	1.78	1.89	1.84
African American		2.04	0.575	2.04	0.641
Others		1.95	0.742	1.94	0.787
$Theta_{drugeffect}$	Drug effect of budesonide	0.103	14.1	0.108	14.4
Interindividual variability (SD) (shrinkage %)†					
$Theta_1$		0.00722 (32.9)		0.00723	
$Theta_3$		0.0912 (26.4)		0.0915	
$Theta_{drugeffect}$		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.
‡The CV and SD for the proportional and additive residual unexplained variability, respectively.

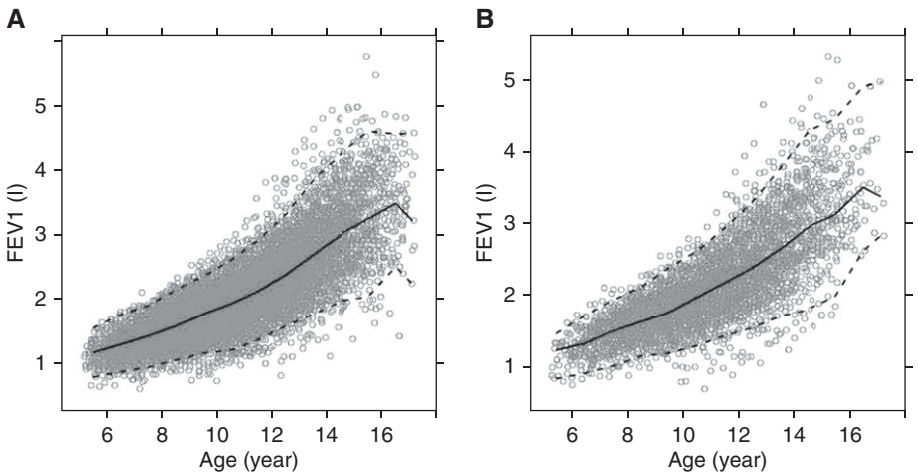

Figure 2. Visual predictive check of the final model. The median, the 5th, and 95th percent prediction intervals from visual predictive check simulation were superimposed with the observations. (*A*) Placebo and nedocromil treatment groups. (*B*) Budesonide treatment group. The medians of model simulations are shown by *solid lines* and 95% prediction intervals are encompassed by the *broken lines*. The *gray circles* refer to the observed FEV_1.

intervals (Figure 2), with less than 10% of the observations lying outside the intervals, suggesting that the final model adequately described most of the observed data. Therefore, we calculated, for each CAMP subject, the individual-level parameters ($theta_1$ and $theta_3$) generated by our final model as given by equation 1. These individual-level parameters were used in the subsequent GWAS analysis as phenotypes to explore the relationship between genetic variants and baseline FEV_1 level or the increase in FEV_1 in children with asthma.

GWAS

Because of the observed ancestry effect on FEV_1 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 FEV_1 between these two groups. Association analyses between 473,680 genotyped SNPs (which passed quality control) and the two modeled parameters ($theta_1$ and $theta_3$, equation 3) were performed. Forty-six and 53 SNPs were nominally associated with $theta_1$ ($P < 1 \times 10^{-4}$) and $theta_3$ ($P < 1 \times 10^{-4}$), 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 $theta_3$ (baseline FEV_1 level) and one SNP (rs17161791) for $theta_1$ (rate of change in FEV_1 with age) showed associations at P less than 0.05 (Table 3) in the nedocromil treatment group. Among them, two intronic SNPs (rs347412, $P = 1.39 \times 10^{-5}$; rs238349, $P = 1.42 \times 10^{-5}$), on chromosome 13 in the *DGKH* gene, showed the most significant associations with $theta_3$.

The variant allele for each of these four SNPs (rs347412, rs238349, rs6763931 [*see* Figure E1], and rs2304725) was associated with higher $theta_3$ and, therefore, lower FEV_1. For the other two SNPs (rs559389 and rs9366309), the variant allele was found to be associated with lower $theta_3$ leading to higher FEV_1. The variant allele of rs17161791 was associated with higher $theta_1$ resulting in higher FEV_1 increase rate. Not surprisingly, the association P values for the model parameters ($theta_1$ and $theta_3$) 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

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

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 $theta_1$, the other six SNPs are associated with $theta_3$.

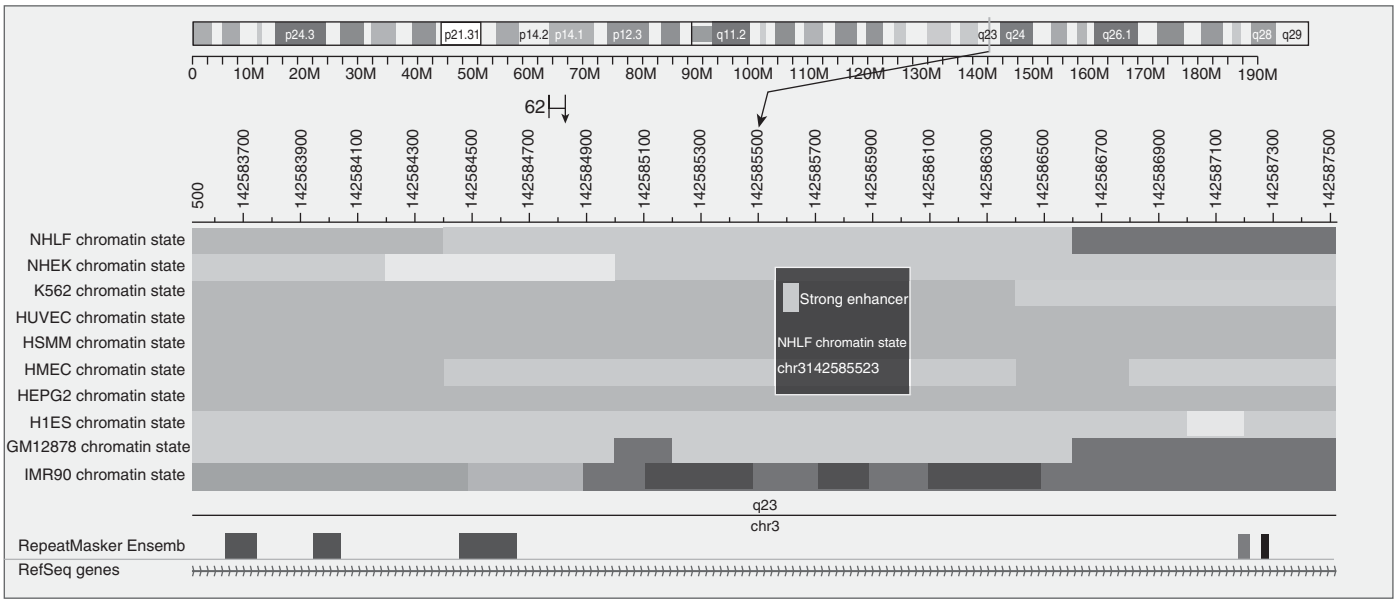

Figure 3. Regulatory function of rs6763931. The single-nucleotide polymorphism rs6763931 (located in an intron of *ZBTB38*, *black arrow*) overlaps a strong enhancer in normal human lung fibroblast (NHLF) and in a lymphoblastoid cell line (GM12878).

($P = 0.21$; *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 theta_3.

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.

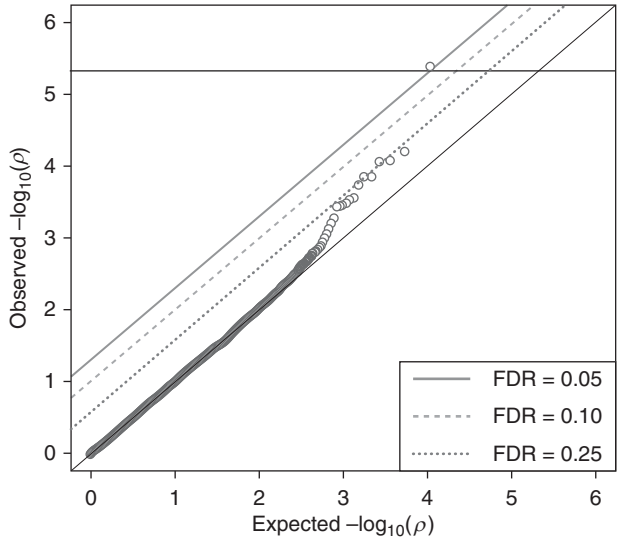

Figure 4. Quantile–quantile plot for the single-nucleotide polymorphisms (SNPs) overlapping strong enhancers. A highly significant association (rs6763931, FDR < 0.05) was identified using SNPs in regulatory regions in human lung fibroblast. The *P* values obtained from associations of the enhancer SNPs with theta_3 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

Analysis Using Enhancer SNPs in Lung Fibroblast	FDR < 0.05	FDR < 0.10	Bonferroni (0.05/N)	Analysis Using Known Lung Function-associated SNPs in the NHGRI Catalog	*P* < 0.05
Longitudinal	1	1	1	Longitudinal	5
FEV_1 at 48 mo	0	0	0	FEV_1 at 48 mo	0
Permuted longitudinal data (n = 100)	0	0	0		

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

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 FEV_1 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 FEV_1 parameters in both placebo and nedocromil groups. Of these, two intronic SNPs (rs347412 and rs238349 in the *DGKH* 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 ($P = 5.90 \times 10^{-5}$) (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 FEV_1 ($P < 0.05$) in our study: rs4762767 and rs58667 for $theta_1$ and rs1291183, rs12984174, and rs2571445 for $theta_3$. Specifically, rs1291183, located within the gene *YES1* on chromosome 18, was previously reported to be associated with percent predicted FEV_1 and percent predicted FVC ($P = 3.54\times10^{-6}$ and 5.47×10^{-5}, respectively) in populations of European descent with asthma (8). This SNP was associated with our longitudinal phenotype ($theta_3$ at $P < 0.05$) in placebo and nedocromil treatment groups. In addition, the SNPs rs2571445 and rs58667 were reported to be associated with FEV_1 or percent predicted FEV_1 with $P = 1.11\times10^{-12}$ 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 FEV_1/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 FEV_1 versus the baseline level even with the use of functional data. However, both $theta_1$ and $theta_3$ 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

1. Covar RA, Spahn JD, Murphy JR, Szefler SJ; Childhood Asthma Management Program Research Group. Progression of asthma measured by lung function in the childhood asthma management program. *Am J Respir Crit Care Med* 2004;170:234–241.
2. Hancock DB, Romieu I, Shi M, Sienra-Monge JJ, Wu H, Chiu GY, Li H, del Rio-Navarro BE, Willis-Owen SA, Weiss ST, *et al*. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in Mexican children. *PLoS Genet* 2009;5:e1000623.
3. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, Fujita K, Miyatake A, Enomoto T, Miyagawa T, *et al*. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* 2011;43:893–896.
4. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, Bleecker ER. Genome-wide association study of asthma identifies RAD50–IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol* 2010;125:328–335 e311.
5. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, *et al*.; Wellcome Trust Case Control Consortium; NSHD Respiratory Study Team. Genome-wide

association study identifies five loci associated with lung function. *Nat Genet* 2010;42:36–44.
6. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010; 42:45–52.
7. Li X, Howard TD, Moore WC, Ampleford EJ, Li H, Busse WW, Calhoun WJ, Castro M, Chung KF, Erzurum SC, *et al*. Importance of hedgehog interacting protein and other lung function genes in asthma. *J Allergy Clin Immunol* 2011;127:1457–1465.
8. Li X, Hawkins GA, Ampleford EJ, Moore WC, Li H, Hastie AT, Howard TD, Boushey HA, Busse WW, Calhoun WJ, *et al*. Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients. *J Allergy Clin Immunol* 2013;132: 313–320.e15.
9. Mould DR. Models for disease progression: new approaches and uses. *Clin Pharmacol Ther* 2012;92:125–131.
10. Mould DR, Denman NG, Duffull S. Using disease progression models as a tool to detect drug effect. *Clin Pharmacol Ther* 2007; 82:81–86.
11. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
12. Hibbert ME, Lannigan A, Landau LI, Phelan PD. Lung function values from a longitudinal study of healthy children and adolescents. *Pediatr Pulmonol* 1989;7:101–109.
13. Kristufek P, Brezina M, Ciutti P, Strmen J, Mayer M. Reference values and modelling of lung function development as a transcendent function of age, body height and mass. *Bull Eur Physiopathol Respir* 1987;23:139–147.
14. Ma G, Olsson B, Rosenborg J, Karlsson M. Quantifying lung function progression in asthma. Presented at the Annual Meeting of the Population Approach Group in Europe. June 9, 2009, St. Petersburg, Russia, p. 18.
15. The Childhood Asthma Management Program Research Group. Long-term effects of budesonide or nedocromil in children with asthma. *N Engl J Med* 2000;343:1054–1063.
16. Childhood Asthma Management Program Research Group. The Childhood Asthma Management Program (CAMP): design, rationale, and methods. *Control Clin Trials* 1999;20:91–120.
17. Yano Y, Beal SL, Sheiner LB. Evaluating pharmacokinetic/ pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodyn* 2001;28:171–192.
18. Karlsson M, Holford N. A tutorial on visual predictive checks. Presented at the Annual Meeting of the Population Approach Group in Europe. June 20, 2008, Marseille, France. p. 17.
19. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011; 473:43–49.
20. Kicic A, Hallstrand TS, Sutanto EN, Stevens PT, Kobor MS, Taplin C, Paré PD, Beyer RP, Stick SM, Knight DA. Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium. *Am J Respir Crit Care Med* 2010;181: 889–898.
21. Asmann YW, Necela BM, Kalari KR, Hossain A, Baker TR, Carr JM, Davis C, Getz JE, Hostetter G, Li X, *et al*. Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res* 2012;72:1921–1928.
22. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003;100:9440–9445.
23. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–216.
24. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
25. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–585.
26. Hindorff L, MacArthur J, Morales J, Junkins H, Hall P, Klemm A, Manolio T. A catalog of published genome-wide association studies [accessed 2014 Mar 12]. Available from: www.genome.gov/gwastudies
27. Wu R, Tong C, Wang Z, Mauger D, Tantisira K, Szefler SJ, Chinchilli VM, Israel E. A conceptual framework for pharmacodynamic genome-wide association studies in pharmacogenomics. *Drug Discov Today* 2011;16:884–890.
28. Sikorska K, Rivadeneira F, Groenen PJ, Hofman A, Uitterlinden AG, Eilers PH, Lesaffre E. Fast linear mixed model computations for genome-wide association studies with longitudinal data. *Stat Med* 2013;32:165–180.
29. Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. *Mol Med* 2010;16:247–253.
30. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, Daigo Y, Yamamoto K, Kubo M, Nakamura Y, *et al*. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum Mol Genet* 2010;19: 2303–2312.
31. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, *et al*. Many sequence variants affecting diversity of adult human height. *Nat Genet* 2008;40:609–615.
32. Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, Hammond N, Stolk L, Nica A, Inouye M, Hofman A, *et al*. Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet* 2009;5:e1000445.
33. Zhao J, Li M, Bradfield JP, Zhang H, Mentch FD, Wang K, Sleiman PM, Kim CE, Glessner JT, Hou C, *et al*. The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. *BMC Med Genet* 2010;11:96.
34. Lee JJ, Essers JB, Kugathasan S, Escher JC, Lettre G, Butler JL, Stephens MC, Ramoni MF, Grand RJ, Hirschhorn J. Association of linear growth impairment in pediatric Crohn's disease and a known height locus: a pilot study. *Ann Hum Genet* 2010;74: 489–497.
35. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, *et al*. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43:1082–1090.
36. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, *et al*. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 2014;507:371–375.
37. Sur IK, Hallikas O, Vähärautio A, Yan J, Turunen M, Enge M, Taipale M, Karhu A, Aaltonen LA, Taipale J. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 2012;338:1360–1363.