Variation in the *GST mu* Locus and Tobacco Smoke Exposure as Determinants of Childhood Lung Function

Carrie V. Breton[1], Hita Vora[1], Muhammad T. Salam[1], Talat Islam[1], Made Wenten[1], W. James Gauderman[1], David Van Den Berg[2], Kiros Berhane[1], John M. Peters[1], and Frank D. Gilliland[1]

[1]Department of Preventive Medicine and [2]Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, California

***Rationale*: The glutathione S-transferases (GSTs) are important detoxification enzymes.**

***Objectives*: To investigate effects of variants in *GST mu* genes on lung function and assess their interactions with tobacco smoke exposure.**

***Methods*: In this prospective study, 14,836 lung function measurements were collected from 2,108 children who participated in two Southern California cohorts. For each child, tagging single nucleotide polymorphisms in *GSTM2*, *GSTM3*, *GSTM4*, and *GSTM5* loci were genotyped. Using principal components and haplotype analyses, the significance of each locus in relation to level and growth of FEV_1, maximum midexpiratory flow rate (MMEF), and FVC was evaluated. Interactions between loci and tobacco smoke on lung function were also investigated.**

***Measurements and Main Results*: Variation in the *GST mu* family locus was associated with lower FEV_1 ($P = 0.01$) and MMEF (0.04). Two haplotypes of *GSTM2* were associated with FEV_1 and MMEF, with effect estimates in opposite directions. One haplotype in *GSTM3* showed a decrease in growth for MMEF (−164.9 ml/s) compared with individuals with other haplotypes. One haplotype in *GSTM4* showed significantly decreased growth in FEV_1 (−51.3 ml), MMEF (−69.1 ml/s), and FVC (−44.4 ml), compared with all other haplotypes. These results were consistent across two independent cohorts. Variation in *GSTM2* was particularly important for FVC and FEV_1 among children whose mothers smoked during pregnancy.**

***Conclusions*: Genetic variation across the *GST mu* locus is associated with 8-year lung function growth. Children of mothers who smoked during pregnancy and had variation in *GSTM2* had lower lung function growth.**

Keywords: FEV_1; *in utero*; glutathione S-transferase; tobacco smoke

Low lung function in adulthood is consistently associated with respiratory and cardiovascular disease morbidity and mortality, including chronic obstructive pulmonary disease, cystic fibrosis, asthma, and cardiovascular disease (1, 2). Lung function growth in childhood is one determinant of maximum attainable lung function in adulthood, and can have a lifelong effect on respiratory disease risk (3–5). Many factors can adversely affect lung development, including genetic variation (6, 7) and exposure to various environmental pollutants, such as tobacco smoke, ozone, and fine particulate matter (8, 9).

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

GSTM1 plays an important role in lung function development. The role of variants in other *GSTM* genes in lung development has yet to be determined.

What This Study Adds to the Field

Haplotypes of *GSTM2*, *GSTM3*, and *GSTM4* are associated with lung function growth during adolescence. The association between *GSTM2* haplotypes and lung function is modified by *in utero* exposure to tobacco smoke.

Genetic variation has the potential to affect childhood lung development, lung maturation, and adult lung injury and repair. Genes involved in oxidative stress may be particularly relevant, because oxidative stress may interfere with normal lung growth and contribute to the development of respiratory diseases (10). One family of genes essential for antioxidant defense is the cytosolic glutathione S-transferase (GST) genes. These genes code for important phase II detoxification enzymes, which are an integral part of the defense mechanism that protects against cellular damage from chemical agents. GSTs conjugate organic compounds, such as insecticides, herbicides, and carcinogens in tobacco smoke, and have antioxidant and antiinflammatory activities (11).

The *GST mu* class spans a 97-kb region on chromosome 1p13 (5′-*GSTM4-GSTM2-GSTM1-GSTM5-GSTM3*-3′, in that order). These genes are expressed in lung tissue and localize to the bronchial wall, suggesting *GST mu* sequence variation may contribute to susceptibility (12, 13). In particular, the frequently studied *GSTM1* deletion polymorphism is associated with reduction of enzymatic activity and individuals with this genotype have a higher level of carcinogen-DNA adducts and more cytogenetic damage (14). The *GSTM1* null genotype is also associated with increased susceptibility to respiratory disease and lung function deficits in children, people with asthma, and the general adult population (15–18). Moreover, negative effects of tobacco smoke on lung function and asthma are exacerbated in individuals with *GSTM1* null genotype (19, 20).

Of the five isoforms within the *GST mu* class (*GSTM1–5*), only the *GSTM1* deletion polymorphism has been well studied. Nevertheless, other *GST mu* genes have overlapping substrate specificities and should also be considered in assessing health outcomes. In this study, a comprehensive assessment of the other isoforms, *GSTM2*, *GSTM3*, *GSTM4*, and *GSTM5*, was conducted to evaluate their association with lung function growth in healthy children using data collected on 2,108 children over an 8-year time period as part of the Children's Health Study (21). We evaluated the global significance of the *GSTM2–5* region, followed by a more detailed exploration of the significance

(*Received in original form September 2, 2008; accepted in final form January 14, 2009*)

Supported by the National Institute of Environmental Health Sciences (Grants: 5P01 ES009581, 5P01 ES011627, and 5P30 ES007048), the United States Environmental Protection Agency (Grants: R826708–01 and RD831861–01), the National Heart, Lung and Blood Institute (Grants: 5R01HL61768 and 1R01HL76647), and the Hastings Foundation.

Correspondence and requests for reprints should be addressed to Frank Gilliland, M.D., Ph.D., Department of Preventive Medicine, USC Keck School of Medicine, 1540 Alcazar Street, CHP 236, Los Angeles, CA 90033. E-mail: gillilan@usc.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 179. pp 601–607, 2009
Originally Published in Press as DOI: 10.1164/rccm.200809-1384OC on January 16, 2009
Internet address: www.atsjournals.org

of specific genetic polymorphisms. Because the GSTs play a critical role in detoxification of tobacco smoke, we also evaluated whether variants in these genes might alter the effects of active smoking, current exposure to secondhand tobacco smoke (SHS), or maternal smoking during pregnancy on lung function growth. Some of the results of these studies have been previously reported in the form of an abstract (22).

METHODS

Study Subjects

Two cohorts of 4th grade children (n = 3,887), one in 1993 and the second in 1996, were enrolled and followed for 8 years, through 12th grade in public high school. Children were recruited from elementary schools in 12 southern California communities as part of an investigation of the long-term effects of air pollution on children's respiratory health (21, 23, 24). Pulmonary function and questionnaire data were obtained annually by trained field technicians at the schools. Details of the testing protocol have been previously published (24, 25). Questionnaire details are provided in the online supplement. The cohorts were restricted to non-Hispanic white (n = 2,108) and Hispanic white (n = 1,152) children due to relatively low sample sizes in other racial groups and concerns about possible population stratification bias. An additional 94 children with no pulmonary function data were excluded. Of the remaining 3,166 children, complete genetic information was available on 2,108 (67%). Although 46% of the original population was excluded, a comparison of these children to the included children across several baseline characteristics showed no substantial differences (*see* Table E1 in the online supplement). Excluded children had a slightly higher prevalence of SHS exposure and lower socioeconomic status.

The study protocol was approved by the institutional review board for human studies at the University of Southern California, and consent was provided by a parent or legal guardian for all study subjects.

Laboratory Methods

Buccal cells were collected from participants and used as the source of DNA for genotyping assays. DNA was extracted using a PUREGENE DNA isolation kit (Gentra, Minneapolis, MN). Twenty-two tagging single nucleotide polymorphisms (SNPs) in *GSTM2*, *GSTM3*, *GSTM4*, and *GSTM5* genes were genotyped using an Illumina BeadArray platform (Illumina Inc., San Diego, CA). In addition, one deletion polymorphism in *GSTM3* (RS1799735) was determined by real-time polymerase chain reaction using a TaqMan 7700 (Applied Biosystems, Foster City, CA) as described previously (15).

Single Nucleotide Polymorphism Selection and Genotyping for Illumina Assay

For each locus, we identified a set of SNPs from dbSNPs and other sequencing databases with an SNP density of 1 to 3 SNPs/kb over a region 20 kb upstream and 10 kb downstream of each gene. Details of the SNP selection process are provided in the online data supplement. SNPs were genotyped on the Illumina BeadArray platform, a highly accurate, high-throughput assay (Illumina, Inc.). Twenty-two SNPs were used in this analysis across the four genes, with an r^2_h of greater than 0.95 for each gene. Reference SNP ID (RS) numbers, gene location, minor allele frequencies, and Hardy-Weinberg Equilibrium values are shown in Table E2 and the GSTM locus is shown in Figure E1. Haplotype frequencies of unphased GSTM SNPs for Hispanic and non-Hispanic white children were estimated separately using TagSNPs (the program is available at http://www-hsc.usc.edu/~stram/tagSNPs.html) and are shown in Table E3. Some haplotypes were uncommon (<5%) and were collapsed into a combined "other" category. The estimated number of copies of each haplotype was used as a proxy for the true haplotype, a single imputation procedure that provides unbiased estimates and appropriate confidence intervals (26, 27). Three haplotype blocks were initially determined based on the Gabriel method (28): one for *GSTM3*, *GSTM5*, and a portion of *GSTM4*. We added three additional SNPs in *GSTM4* to that haplotype block because they were in high linkage disequilibrium (LD). We also created a custom haplotype block for *GSTM2* using the SNPs we had genotyped. Within each block, the haplotypes explained at least 95% of the variability of both ethnic groups. To understand the correlation between the *GSTM3* deletion polymorphism (RS1799735) and the *GSTM3* haplotype, RS1799735 was included in the *GSTM3* haplotype.

Statistical Methods

The data consisted of 14,836 pulmonary function tests recorded from 2,108 participants during 8 years of follow-up. Three measures of pulmonary function were evaluated: FVC, FEV_1, and maximal mid-expiratory flow rate (MMEF, also known as $FEF_{25–75}$). The primary exposures of interest were the *GSTM2–5* genotypes described above.

A hierarchical mixed effects model was used to relate 8-year growth in each lung function measure to principal components (PCs), haplotypes, and SNPs, with basic model structure that has been previously described (29). Growth patterns in lung function were modeled using linear splines with knots at ages 12, 14, and 16 (30), parameterized so that 8-year growth in lung function was estimated jointly with other model parameters. Random effects for the intercept and 8-year growth parameters were included at the subject level. We estimated and tested the effect of *GSTM2–5* SNPs on 8-year growth and on mean lung function level at age 18. We allowed for separate growth curves by sex, race, ethnicity, cohort, and asthma status at baseline across the age range of 10 to 18 years with knots at ages 12, 14, and 16. The model also included adjustments for height, height^2, body mass index (BMI), BMI^2, current asthma status, exercise or respiratory illness on the day of the test, any tobacco smoking by the child in the last year, *GSTM1* null genotype, and indicator variables for field technician. Handling of missing data for covariates is addressed in the online data supplement.

To address potential confounding by population stratification, four coefficients of ancestry variables were also included in the model (31, 32). These variables were constructed from 4 principal components derived from a set of 233 unlinked ancestry informative markers, which were selected to differentiate 4 parental populations (African, European, American Indian, and East Asian). Controlling for these ancestry variables provided adjustment for ancestral history beyond adjustment for typical self-reported racial and ethnic categories.

Principal components analysis was used to test the significance of the entire *GSTM2–5* locus in relation to 8-year lung function growth and level at age 18 (33). The set of PCs that explained greater than 80% of the SNP variance across the entire locus was included in the hierarchical model for lung function. Haplotype analyses were then used to investigate individual gene loci. Each evaluated locus contained only one haplotype block. A likelihood ratio test was used to evaluate the overall association between a given block and lung function growth and level. For single SNP analysis, an additive genetic model, in which genotype was coded as zero, one, or two for the number of minor alleles, was used for all SNPs. Single SNP analyses were adjusted for multiple testing within gene, but separately for each lung function outcome, using P_{ACT} (available at http://csg.sph.umich.edu/boehnke/p_act.php) (34).

Based on *a priori* hypotheses, we also evaluated whether genetic variation altered the effects of SHS, active smoking, and maternal smoking during pregnancy on lung function growth. Active smoking and SHS variables were incorporated as time-dependent covariates in the analysis. Interactions between these smoke variables and haplotypes were tested using likelihood ratio tests comparing models with and without the appropriate interaction terms for evidence of statistical significance.

Regression procedures in SAS were used to fit all models (35). Adjustment for multiple testing using P_{ACT} was done using R version 2.6.1 (R Foundation for Statistical Computing, Vienna, Austria) (34). Associations denoted as statistically significant were those that yielded a P value less than 0.05, assuming a two-sided alternative hypothesis.

RESULTS

Over 8 years of follow-up, an average of seven pulmonary function tests were performed per child. Eight-year growth in FVC, FEV_1, and MMEF averaged 1,569 ml, 1,367 ml, and 1,442 ml/s in girls and 2,848 ml, 2,441 ml, and 2,512 ml/s in boys. Approximately one-third of the sample was of Hispanic origin, 17% of children had mothers who smoked during their preg-

nancy, and 18% reported exposure to SHS in the household at baseline (Table 1).

Results from PC analysis suggested that variation in the GSTM family locus was significantly associated with lung function growth and level (global *P* values of 0.01, 0.04, and 0.13 for FEV_1, MMEF, and FVC, respectively). When haplotype analyses were conducted to test individual loci, *GSTM2* and *GSTM4* were associated with 8-year growth in several outcomes, whereas *GSTM3* and *GSTM5* were only associated with one outcome. *GSTM2* was associated with FEV_1 (Table 2) and MMEF, and *GSTM4* with FEV_1 (Table 3), FVC, and MMEF, whereas *GSTM3* and *GSTM5* were only associated with MMEF (Tables E4, E5).

GSTM2

For *GSTM2*, two haplotypes were observed: one that increased and one that decreased lung function growth. The most common haplotype (00101, with frequencies of 35% in non-Hispanic whites and 30% in Hispanic whites) was associated with lower FEV_1 (−38.6 ml, $P = 0.02$) and lower MMEF (−71.0 ml/s, $P = 0.02$) compared with other haplotypes (Table 4). Moreover, two copies of this risk haplotype had twice the deficit of lung function growth compared with one copy (Table E6). The protective haplotype (10000) occurred with a frequency of 28% in non-Hispanic whites and 44% in Hispanic whites. Eight-year growth in FEV_1 and MMEF was larger by 28.4 ml and 109.4 ml/s in individuals with this haplotype compared with individuals with other haplotypes, although no increase in strength of association was seen by number of haplotypes.

GSTM3

For *GSTM3*, a haplotype association was only observed with 8-year growth in MMEF ($P = 0.03$) (Table E4). The risk haplotype 001001 (frequency 8% in non-Hispanic whites, 6% in Hispanic whites) had a strong negative effect on lung function growth for MMEF (Table 4). Eight-year growth was 164.9 ml/s lower for MMEF in individuals with this risk haplotype compared with individuals with other haplotypes. The two SNPs underlying this haplotype were RS1537236 and the deletion polymorphism RS1799735. Single SNP analysis of RS1799735 also resulted in lower MMEF (−95.7 ml/s) compared with the common variant (Table E7).

It is possible that SNP RS1799735 is driving the haplotype effect; however, this SNP was split across two haplotypes, only one of which was associated with MMEF. To explore further, we conducted a likelihood ratio test to evaluate whether the haplotype had an independent effect on lung function beyond the effect of the SNP. We found the haplotype contributed a significantly independent association for MMEF (likelihood ratio test [LRT] $P = 0.01$). These results suggest that the haplotype captures additional variation in the *GSTM3* locus besides the deletion polymorphism that may be important for lung function.

GSTM4

For *GSTM4*, haplotype 1101000 occurred with a frequency of 22% in non-Hispanic whites and 16% in Hispanic whites, and showed significantly decreased rates of growth for FEV_1 (−51.3 ml), FVC (−44.4 ml), and MMEF (−69.1 ml/s) compared with all other haplotypes (Table 4). In addition, two copies of this haplotype showed twice the deficit in lung function growth compared with one copy (Table E6).

GSTM5

For *GSTM5*, a global haplotype association was only observed with 8-year growth in MMEF ($P = 0.01$) (Table E5). This association was driven by the "other" category, which constituted less than 5% of all haplotypes and no individual haplotype was associated with 8-year lung function growth. Therefore no further analyses of *GSTM5* are presented.

TABLE 1. BASELINE CHARACTERISTICS FOR THE 2,108 FOURTH GRADE CHILDREN'S HEALTH STUDY SUBJECTS INCLUDED IN LUNG FUNCTION ANALYSES

Variables	N	%
Sex		
Boys	1022	48.5
Ethnicity		
Non-Hispanic white	1398	66.3
Hispanic white	710	33.7
Ever diagnosed with asthma	310	15.0
Cohort		
Recruited in 1993	912	43.3
Recruited in 1996	1196	56.7
Exposure to secondhand smoke	361	17.6
Personal smoking in last year	36	1.7
In utero exposure to smoke	354	17.2
	Mean (SD)	
Height, cm	139.6 (6.6)	
BMI, kg/m^2	18.2 (3.5)	
Age, yr	10.0 (0.4)	

Definition of abbreviation: BMI = body mass index.

Consistency of Effect

The main haplotype associations for all genes were consistent across our two independent cohorts and were similar by ethnicity (Tables E8a–c). Haplotype associations were consistent with single SNP analyses (Table E7). Sensitivity analyses were conducted to evaluate whether presence of socioeconomic status (SES) and inclusion of children with asthma might have affected the observed associations. Results for *GSTM2*, *3*, and *4* haplotypes remained significant even after adjustment for SES and after restriction to only children without asthma (Table E9). Similarly, SES did not affect the associations in the joint effects model of *GSTM2* and *in utero* tobacco smoke exposure (Table E10).

Interaction of the GSTM Family Locus with Tobacco Smoke Exposure

We evaluated whether the genes might alter susceptibility to active smoking by the child (prevalence 1.7%), current exposure to secondhand tobacco smoke (prevalence 17.6%) or *in utero* tobacco smoke exposure (prevalence 17.2%). No interactions between GSTM loci and SHS or active smoking were observed, although the frequency of active smoking was particularly small, which limited the power to detect interactions for that exposure. *GSTM2* was the only gene to show significant interactions with *in utero* tobacco smoke exposure (Table 5). Children of mothers who smoked during pregnancy had lower FEV_1 (−127.6 vs. −23.2 ml) and FVC (−169.7 vs. −2.7 ml) if they had the *GSTM2* risk haplotype 00101. In a complementary fashion, the haplotype 10000 was protective in the exposed children. These associations were consistent with single SNP results for the SNPs underlying the haplotypes.

DISCUSSION

We found that genetic variation across the family of *GST mu* genes was associated with lung function development in children 10 to 18 years of age. In our population, we observed a global association between the entire region spanning *GSTM2*

TABLE 2. GLOBAL ASSOCIATION OF *GSTM2* WITH FEV_1 (ml) LEVEL AT AGE 18 AND 8-YEAR GROWTH USING HAPLOTYPE ANALYSIS: CHILDREN'S HEALTH STUDY, FOURTH GRADE COHORTS

Haplotype*	β†	*P* Value	Global *P* Value‡
Level at age 18			0.04
00101	Ref		
10000	32.9	0.12	
00000	−23.5	0.43	
10001	17.7	0.62	
01110	7.3	0.86	
Other	25.4	0.42	
Growth			
10000	53.8	0.005	
00000	15.6	0.56	
10001	34.9	0.27	
01110	28.7	0.44	
Other	49.1	0.08	

* Haplotypes were defined by the following SNPs, in order: RS2073483, RS574344, RS655315, RS619686, and RS12024479, where "1" is the variant and "0" the common variant allele.

† Difference in FEV_1 comparing each haplotype to reference haplotype. All models adjusted for covariates listed in the METHODS section.

‡ Global *P* value from likelihood ratio test of all haplotypes for growth and level at age 18 combined for the locus.

TABLE 3. GLOBAL ASSOCIATION OF *GSTM4* WITH FEV_1 (ml) LEVEL AT AGE 18 AND 8-YEAR GROWTH USING HAPLOTYPE ANALYSIS: CHILDREN'S HEALTH STUDY, FOURTH GRADE COHORTS

Haplotype*	β†	*P* Value	Global *P* Value‡
Level at age 18			0.003
0010000	Ref		
1101000	−26.3	0.24	
1101001	−46.9	0.10	
1000010	−25.2	0.47	
1000110	−1.9	0.96	
0000000	21.5	0.60	
1100000	30.5	0.41	
Other	78.5	0.17	
Growth			
1101000	−51.4	0.01	
1101001	−26.6	0.30	
1000010	−16.4	0.60	
1000110	29.6	0.38	
0000000	32.5	0.38	
1100000	4.7	0.89	
Other	79.9	0.12	

* Haplotypes were defined by the following SNPs, in order: RS12745189, RS668413, RS1010167, RS560018, RS650985, RS506008, and RS521999, where "1" is the variant and "0" the common variant allele.

† Difference in FEV_1 comparing each haplotype to reference haplotype. All models adjusted for covariates listed in the METHODS section.

‡ Global *P* value from likelihood ratio test of all haplotypes for growth and level at age 18 combined for the locus.

through *GSTM5* and lung function growth, as well as locus-specific haplotype associations for three of the four *GST mu* isoforms studied: *GSTM2*, *GSTM3*, and *GSTM4*. These associations remained after adjustment for presence or absence of *GSTM1* null genotype. Moreover, children with variants in *GSTM2* were more susceptible to effects of *in utero* tobacco smoke exposure on lung function than were children with the common variant. The joint effects of the variant and *in utero* exposure to maternal smoking were not altered when postnatal exposure to SHS was considered.

A substantial body of evidence supports a role for *GST mu* variants in lung function and growth (36, 37). *GSTM1*, the most–well-studied isoform, is associated with a reduction in enzymatic activity and has been implicated in respiratory diseases, including lung cancer, asthma, and deficits in lung function, although the data are not conclusive (15–18, 38, 39). *GST mu* genes are expressed to varying degrees in lung tissue, particularly in the bronchial wall, and therefore may have a direct effect on lung function development (12, 13). *GST mu* genes have a central role in phase II detoxification of reactive oxygen species. Failure to detoxify reactive oxygen species, which could be due to alterations in GST function caused by sequence variants, can enhance the inflammatory cascade, promote bronchoconstrictor mechanisms and airway hyperresponsiveness, contribute to asthma-like symptomatology, and impair lung development (36). *GST mu* genes are colocated across a 97-kb region on chromosome 1p13. Because the five isoforms have overlapping substrate specificities, variation in other isoforms has the potential to affect respiratory disease. However, no reports to date have been published demonstrating effects of variants in *GSTM2–5* on lung function in human populations.

For *GSTM2*, two haplotypes were associated with lung function growth. A risk haplotype conferred lower 8-year growth in FEV_1 and MMEF compared with other haplotypes, whereas a protective haplotype conferred higher 8-year growth in FEV_1 and MMEF. None of the SNPs constituting these haplotypes are known to be functional. A functional SNP in *GSTM2* does exist (rs592792); however, it did not track with the protective or risk haplotypes that we observed. Thus, a more extensive investigation is needed to determine which causal SNPs these haplotypes are marking.

The effects of *GSTM2* haplotypes showed significant variation for FEV_1 and FVC when prenatal exposure to tobacco smoke was considered. The magnitude of the genetic associations was substantially larger among the children of mothers who smoked during pregnancy. To our knowledge, this is the first study to demonstrate an interaction between *GSTM2* variants and prenatal tobacco smoke exposure on lung function. We previously found that *in utero* exposure to maternal smoking was associated with decreased lung function in children, especially for small airway flows (40), and an increased risk of asthma and wheeze among *GSTM*-null children (20). In this study, the *GSTM2* results remained significant after adjustment for *GSTM1*, indicating that they have independent effects.

GST mu genes are expressed in fetal tissue at low levels. However, the phase II enzymes are relatively uninducible, whereas activating enzymes are highly inducible. This suggests that fetuses are much more susceptible to environmental exposures than adults (41). Genetic polymorphisms in phase II enzymes that further inhibit enzyme activity may exacerbate this susceptibility. For example, subjects carrying the *GSTM1* null genotype had a higher susceptibility to DNA damage induced by tobacco smoke than *GSTM1*-positive ones (42). Our results for an interaction between *GSTM2* and *in utero* smoke exposure are consistent with the known role for GST genes in detoxification of carcinogens and with other studies showing that variation in the GST family of genes can alter the effects of cigarette smoke on respiratory outcomes.

The results for *GSTM3*, although not robust across all lung function outcomes in our study, are worth mentioning particularly because several studies have evaluated functional SNPs in this gene. The functional SNPs rs1799735, rs7483, and rs1332018 have been associated with risk of diseases such as Alzheimer, cancers, multiple sclerosis, and lung disease (43–48). In our analyses, we directly evaluated two of these functional SNPs

TABLE 4. THE ASSOCIATION OF *GSTM 2, 3,* AND *4* HAPLOTYPES COMPARED WITH ALL OTHERS WITH LUNG FUNCTION GROWTH: CHILDREN'S HEALTH STUDY, 4TH GRADE COHORTS

		8-yr Growth in Lung Function								
		FEV_1 (ml)			FVC (ml)			MMEF (ml/s)		
Gene	Haplotype	β*	95% CI	*P* Value	β*	95% CI	*P* Value	β*	95% CI	*P* Value
GSTM2†										
00101	None	Ref		0.02	Ref		0.15	Ref		0.02
	At least one	−38.6	(−71.4, −5.7)		−26.2	(−62.2, 9.8)		−71.0	(−130.9, −11.0)	
10000	None	Ref		0.10	Ref		0.15	Ref		0.001
	At least one	28.4	(−5.7, 62.5)		−0.1	(−37.5, 37.2)		109.4	(47.3, 171.6)	
GSTM3‡										
001001	None	Ref		0.14	Ref		0.99	Ref		0.002
	At least one	−44.5	(−103.2, 14.3)		2.1	(−62.3, 66.6)		−164.9	(−271.5, −58.2)	
GSTM4§										
1101000	None	Ref		0.01	Ref		0.03	Ref		0.05
	At least one	−51.3	(−88.8, −13.8)		−44.4	(−85.5, −3.4)		−69.1	(−137.2, −1.0)	

Definition of abbreviations: CI = confidence interval; MMEF = maximal midexpiratory flow.

* Difference in lung function measure comparing risk haplotype to all other haplotypes combined. All models include adjustment for covariates listed in the METHODS section.

† The risk haplotypes were defined by the following SNPs, in order: RS2073483, RS574344, RS655315, RS619686, and RS12024479, where "1" is the variant and "0" the common variant allele.

‡ The risk haplotype 001001 was defined by the following SNPs, in order: RS11101993, RS11101996, RS1537236, RS7483, RS10735234, and RS1799735.

§ The risk haplotype 1101000 was defined by the following SNPs, in order: RS12745189, RS668413, RS1010167, RS560018, RS650985, RS506008, and RS521999.

(rs1799735 and rs7483). We also evaluated a third SNP (rs10735234) that was in high LD (D′ = 0.96) with the functional SNP, rs1332018. A global test of the *GSTM3* locus suggested a negative association for haplotype h010001 with MMEF. We also observed an independent single SNP association for rs1799735, in which we observed a significant decrease in MMEF in subjects carrying the *GSTM3**B allele. This is in contrast to results observed in a small study of children with cystic fibrosis, in which the *GSTM3**B allele was found to increase lung function (FEV_1 and FVC) relative to the *GSTM3**A allele (16). However, this discrepancy may be explained by the presence of the disease itself. We observed no associations for the other two SNPs.

For *GSTM4*, consistent associations with all lung function outcomes suggest that haplotype 1101000 may be an important determinant for lung function growth. Of the SNPs included in our analyses, two are functional: RS506008 is a synonymous SNP and RS17024663 a missense mutation. These functional SNPs were not associated with any lung function outcomes in our single SNP analyses. RS506008 did not contribute to the observed risk haplotype and RS17024663 was excluded from haplotype creation due to a large number of missing data for that SNP. Variation in *GSTM4* is not well studied with respect to disease outcomes. One study did find that a C-T polymorphism in intron 6 of *GSTM4* conferred an increased risk for lung cancer (49). However, we observed no association for this SNP (rs650985) in our data nor did it drive the observed haplotype association.

A strength of this study was the long-term, prospective nature of the data with consistent follow-up and measurement of exposure and outcome data. However, certain limitations should also be considered. Haplotypes were used as part of the analytic strategy. The ability to detect a true disease variant using a haplotype approach can be attenuated if the variant is

TABLE 5. THE ASSOCIATION OF *GSTM2* HAPLOTYPES WITH 8-YEAR LUNG FUNCTION GROWTH IN FEV_1, FVC, AND MAXIMAL MIDEXPIRATORY FLOW STRATIFIED BY *IN UTERO* SMOKE EXPOSURE: CHILDREN'S HEALTH STUDY, FOURTH GRADE COHORTS

			In utero Smoke Exposed		*In utero* Smoke Unexposed		
Lung Function Growth	*GSTM2**	Haplotype	β†	95% CI	β†	95% CI	Interaction *P* Value‡
FEV_1 (ml)	00101	None	Ref		Ref		0.05
		At least one	−127.6	(−205.6, −49.5)	−23.2	(−60.1, 13.7)	
	10000	None	Ref		Ref		0.16
		At least one	105.2	(22.8, 187.5)	17.9	(−20.4, 56.3)	
FVC (ml)	00101	None	Ref		Ref		0.001
		At least one	−169.7	(−258.6, −80.9)	−2.7	(−42.6, 37.2)	
	10000	None	Ref		Ref		0.05
		At least one	100.6	(5.3, 195.9)	−13.2	(−54.6, 28.3)	
MMEF (ml/s)	00101	None	Ref		Ref		0.87
		At least one	−115.4	(−258.6, 27.8)	−65.8	(−133.0, 1.4)	
	10000	None	Ref		Ref		0.59
		At least one	116.2	(−33.8, 266.1)	116.3	(46.7, 186.0)	

Definition of abbreviations: CI = confidence interval; MMEF = maximal midexpiratory flow.

* The risk haplotypes were defined by the following SNPs, in order: RS2073483, RS574344, RS655315, RS619686, and RS12024479, where "1" is the variant and "0" the common variant allele.

† Difference in lung function measure comparing risk haplotype to all other haplotypes combined. All models include adjustment for covariates listed in the METHODS section.

‡ Wald *P* value for the interaction between haplotype and *in utero* smoke exposure.

distributed across several haplotypes in the population. This may be particularly relevant to the analysis of *GSTM3*, because the *GSTM3* deletion was split across two haplotypes. Moreover, we adopted the commonly used approach of clustering rare haplotypes into a composite category. This composite category likely included one or more haplotypes carrying the variant, thereby reducing the analytical power to contrast this variant against other haplotypes.

Confounding by population admixture is often a concern with genetic studies. We attempted to control for admixture by adjusting for ancestry variables in addition to typical adjustment for self-reported race and ethnicity. The ancestry variables provided better control for genetic descent of four distinct groups: African, European, American Indian, and East Asian. Adjusting for these variables did not appreciably change our results.

Misclassification of environmental or genetic risk factors may reduce the power to detect gene–environment interactions (50). The smoking exposures in this study may be of particular concern because they were assessed by self-report via questionnaire and smokers may be inclined to not report the fact that they smoke, particularly during the sensitive time of pregnancy, thereby underestimating true exposure. In cohort studies, such misclassification typically biases effect estimates toward the null. Despite the likelihood of some misclassification of smoking status, our study had a large sample size and we were able to observe significant differences in lung function growth of at least 100 ml or more.

To address concerns regarding multiple testing, we first assessed the entire family locus for association with lung function phenotypes, followed by test of each gene locus. For loci with significant associations, we followed with single SNP analyses that were corrected for multiple testing within gene using $P_{ACT,}$ a method for adjusting for multiple correlated tests particularly relevant for genetic studies and less conservative than a Bonferroni approach, which does not account for the correlations among SNPs.

In conclusion, we found that *GST mu* genes appear to play an important role in lung function development in children from 10 to 18 years of age. In general, the direction of the associations for haplotype and single SNP analyses were negative, implying that variants in these genes collectively tended to decrease lung capacity or small airway flow. *GSTM2*, in particular, may play a critical role in oxidative defense to environmental insults during fetal development. GSTs are a primary pathway for detoxifying lipid peroxidation products formed from tobacco smoke. Thus, children with *GST mu* variants may have decreased antioxidant defense compared with those with the common variant genotypes during prenatal and postnatal lung development. The *GST mu* genes constitute a larger family of GSTs with complex and overlapping substrates and functions. Future investigations into potential interactions between genes in this family may further clarify their individual roles in lung function development.

Conflict of Interest Statement: None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors thank all the participants and their parents of the Children's Health Study, as well as all the teachers and staff of the participating schools. They also thank the field team and laboratory personnel whose conscientious and courteous work ethic enabled them to gather and process the data efficiently.

References

1. Shi W, Bellusci S, Warburton D. Lung development and adult lung diseases. *Chest* 2007;132:651–656.
2. Chowdhuri S, Crook ED, Taylor HA Jr, Badr MS. Cardiovascular complications of respiratory diseases. *Am J Med Sci* 2007;334:361–380.
3. Schibler A. Physiological consequences of early-life insult. *Paediatr Respir Rev* 2006;7:103–109.
4. Bakke PS. Factors affecting growth of FEV1. *Monaldi Arch Chest Dis* 2003;59:103–107.
5. Dezateux C, Stocks J. Lung development and early origins of childhood respiratory illness. *Br Med Bull* 1997;53:40–57.
6. Kimura J, Deutsch GH. Key mechanisms of early lung development. *Pediatr Dev Pathol* 2007;10:335–347.
7. Warburton D, Gauldie J, Bellusci S, Shi W. Lung development and susceptibility to chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2006;3:668–672.
8. Milner AD, Rao H, Greenough A. The effects of antenatal smoking on lung function and respiratory symptoms in infants and children. *Early Hum Dev* 2007;83:707–711.
9. Kajekar R. Environmental factors and developmental outcomes in the lung. *Pharmacol Ther* 2007;114:129–145.
10. MacNee W. Pulmonary and systemic oxidant/antioxidant imbalance in chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2005;2: 50–60.
11. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. *Annu Rev Pharmacol Toxicol* 2005;45:51–88.
12. Anttila S, Hirvonen A, Vainio H, Husgafvel-Pursiainen K, Hayes JD, Ketterer B. Immunohistochemical localization of glutathione s-transferases in human lung. *Cancer Res* 1993;53:5643–5648.
13. Mace K, Bowman ED, Vautravers P, Shields PG, Harris CC, Pfeifer AM. Characterisation of xenobiotic-metabolising enzyme expression in human bronchial mucosa and peripheral lung tissues. *Eur J Cancer* 1998;34:914–920.
14. Reszka E, Wasowicz W. Significance of genetic polymorphisms in glutathione s-transferase multigene family and lung cancer risk. *Int J Occup Med Environ Health* 2001;14:99–113.
15. Gilliland FD, Gauderman WJ, Vora H, Rappaport E, Dubeau L. Effects of glutathione-s-transferase m1, t1, and p1 on childhood lung function growth. *Am J Respir Crit Care Med* 2002;166:710–716.
16. Flamant C, Henrion-Caude A, Boelle PY, Bremont F, Brouard J, Delaisi B, Duhamel JF, Marguet C, Roussey M, Miesch MC, *et al.* Glutathione-s-transferase m1, m3, p1 and t1 polymorphisms and severity of lung disease in children with cystic fibrosis. *Pharmacogenetics* 2004;14:295–301.
17. Romieu I, Sienra-Monge JJ, Ramirez-Aguilar M, Moreno-Macias H, Reyes-Ruiz NI, Estela del Rio-Navarro B, Hernandez-Avila M, London SJ. Genetic polymorphism of gstm1 and antioxidant supplementation influence lung function in relation to ozone exposure in asthmatic children in Mexico City. *Thorax* 2004;59:8–10.
18. Imboden M, Downs SH, Senn O, Matyas G, Brandli O, Russi EW, Schindler C, Ackermann-Liebrich U, Berger W, Probst-Hensch NM. Glutathione s-transferase genotypes modify lung function decline in the general population: SAPALDIA cohort study. *Respir Res* 2007;8:2.
19. He JQ, Ruan J, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Antioxidant gene polymorphisms and susceptibility to a rapid decline in lung function in smokers. *Am J Respir Crit Care Med* 2002;166:323–328.
20. Gilliland FD, Li YF, Dubeau L, Berhane K, Avol E, McConnell R, Gauderman WJ, Peters JM. Effects of glutathione s-transferase m1, maternal smoking during pregnancy, and environmental tobacco smoke on asthma and wheezing in children. *Am J Respir Crit Care Med* 2002;166:457–463.
21. Peters JM, Avol E, Navidi W, London SJ, Gauderman WJ, Lurmann F, Linn WS, Margolis H, Rappaport E, Gong H, *et al.* A study of twelve Southern California communities with differing levels and types of air pollution. I. Prevalence of respiratory morbidity. *Am J Respir Crit Care Med* 1999;159:760–767.
22. Breton C, Vora H, Wenten M, Gauderman J, Berg DVD, Gilliland F. GST mu genes are important predictors of lung function [abstract]. *Am J Respir Crit Care Med* 2008;177:A906.
23. Gauderman WJ, Gilliland GF, Vora H, Avol E, Stram D, McConnell R, Thomas D, Lurmann F, Margolis HG, Rappaport EB, *et al.* Association between air pollution and lung function growth in Southern California children: results from a second cohort. *Am J Respir Crit Care Med* 2002;166:76–84.
24. Peters JM, Avol E, Gauderman WJ, Linn WS, Navidi W, London SJ, Margolis H, Rappaport E, Vora H, Gong H Jr, *et al.* A study of twelve Southern California communities with differing levels and types of air pollution. Ii. Effects on pulmonary function. *Am J Respir Crit Care Med* 1999;159:768–775.
25. Gauderman WJ, Avol E, Gilliland F, Vora H, Thomas D, Berhane K, McConnell R, Kuenzli N, Lurmann F, Rappaport E, *et al.* The effect

of air pollution on lung development from 10 to 18 years of age. *N Engl J Med* 2004;351:1057–1067.
26. Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered* 2002;53:79–91.
27. Wallenstein S, Hodge SE, Weston A. Logistic regression model for analyzing extended haplotype data. *Genet Epidemiol* 1998;15:173–181.
28. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, *et al.* The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229.
29. Berhane K, Gauderman W, Stram D, Thomas D. Statistical issues in studies of the long term effects of air pollution: The Southern California children's health study (with discussion). *Stat Sci* 2005;19: 414–449.
30. Gauderman WJ, Vora H, McConnell R, Berhane K, Gilliland F, Thomas D, Lurmann F, Avol E, Kunzli N, Jerrett M, *et al.* Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. *Lancet* 2007;369:571–577.
31. Pritchard J, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–959.
32. Conti DV, Lee W, Li D, Liu J, Van Den Berg D, Thomas PD, Bergen AW, Swan GE, Tyndale RF, Benowitz NL, *et al.* Nicotinic acetylcholine receptor beta2 subunit gene implicated in a systems-based candidate gene study of smoking cessation. *Hum Mol Genet* 2008;17: 2834–2848.
33. Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. *Genet Epidemiol* 2007;31:383–395.
34. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of p values for multiple correlated tests. *Am J Hum Genet* 2007;81:1158–1168.
35. SAS/STAT. Version 9 [program]. 9.0 ed. Cary, NC: SAS Institute; 2002.
36. Sampsonas F, Archontidou MA, Salla E, Karkoulias K, Tsoukalas G, Spiropoulos K. Genetic alterations of glutathione S-transferases in asthma: do they modulate lung growth and response to environmental stimuli? *Allergy Asthma Proc* 2007;28:282–286.
37. Sandford AJ, Pare PD. Genetic risk factors for chronic obstructive pulmonary disease. *Clin Chest Med* 2000;21:633–643.
38. Aynacioglu AS, Nacak M, Filiz A, Ekinci E, Roots I. Protective role of glutathione S-transferase P1 (GSTP1) Val105Val genotype in patients with bronchial asthma. *Br J Clin Pharmacol* 2004;57:213–217.
39. Ye Z, Song H, Higgins JP, Pharoah P, Danesh J. Five glutathione S-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies. *PLoS Med* 2006;3:e91.
40. Gilliland FD, Berhane K, McConnell R, Gauderman WJ, Vora H, Rappaport EB, Avol E, Peters JM. Maternal smoking during pregnancy, environmental tobacco smoke exposure and childhood lung function. *Thorax* 2000;55:271–276.
41. Xu M, Moore JE, Leone-Kabler S, McCoy TP, Swank A, Nelson GB, Ross JA, Townsend AJ, Miller MS. Expression of glutathione S-transferases in fetal lung and liver tissue from parental strains and f1 crosses between c57bl/6 and balb/c f1 mice following in utero exposure to 3-methylcholanthrene. *Biochem Pharmacol* 2006;72:115–123.
42. Palma S, Cornetta T, Padua L, Cozzi R, Appolloni M, Ievoli E, Testa A. Influence of glutathione S-transferase polymorphisms on genotoxic effects induced by tobacco smoke. *Mutat Res* 2007;633:1–12.
43. Maes OC, Schipper HM, Chong G, Chertkow HM, Wang E. A GSTM3 polymorphism associated with an etiopathogenetic mechanism in Alzheimer disease. *Neurobiol Aging* 2008. (In press).
44. Kesarwani P, Singh R, Mittal RD. Association of GSTM3 intron 6 variant with cigarette smoking, tobacco chewing and alcohol as modifier factors for prostate cancer risk. *Arch Toxicol* 2008. Jul 31 (Epub ahead of print).
45. Liu X, Campbell MR, Pittman GS, Faulkner EC, Watson MA, Bell DA. Expression-based discovery of variation in the human glutathione S-transferase M3 promoter and functional analysis in a glioma cell line using allele-specific chromatin immunoprecipitation. *Cancer Res* 2005; 65:99–104.
46. Schnakenberg E, Breuer R, Werdin R, Dreikorn K, Schloot W. Susceptibility genes: GSTM1 and GSTM3 as genetic risk factors in bladder cancer. *Cytogenet Cell Genet* 2000;91:234–238.
47. Mann CL, Davies MB, Boggild MD, Alldersea J, Fryer AA, Jones PW, Ko Ko C, Young C, Strange RC, Hawkins CP. Glutathione S-transferase polymorphisms in MS: their relationship to disability. *Neurology* 2000;54:552–557.
48. Jourenkova-Mironova N, Voho A, Bouchardy C, Wikman H, Dayer P, Benhamou S, Hirvonen A. Glutathione S-transferase GSTM3 and GSTP1 genotypes and larynx cancer risk. *Cancer Epidemiol Biomarkers Prev* 1999;8:185–188.
49. Liloglou T, Walters M, Maloney P, Youngson J, Field JK. A t2517c polymorphism in the GSTM4 gene is associated with risk of developing lung cancer. *Lung Cancer* 2002;37:143–146.
50. Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: Assessment of bias and sample size. *Cancer Epidemiol Biomarkers Prev* 1999;8:1043–1050.