

Interaction between a novel *TGFB1* haplotype and *CFTR* genotype is associated with improved lung function in cystic fibrosis

Journal:	<i>Human Molecular Genetics</i>
Manuscript ID:	HMG-2007-W-01163.R2
Manuscript Type:	1 General Article - US Office
Date Submitted by the Author:	n/a
Complete List of Authors:	Bremer, Lindsay; Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine Blackman, Scott; Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine; Johns Hopkins University School of Medicine, Division of Pediatric Endocrinology Vanscoy, Lori; National Naval Medical Center, Department of Pediatrics McDougal, Kathryn; Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine Bowers, Amanda; Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine Naughton, Kathleen; Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine Cutler, David; Emory University School of Medicine, Department of Human Genetics Cutting, Garry; John Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine
Key Words:	transmission, tdt, qtdt, fbat

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Interaction between a novel *TGFB1* haplotype and *CFTR* genotype is associated with improved lung function in cystic fibrosis

Lindsay A. Bremer¹, Scott M. Blackman², Lori L. Vanscoy³, Kathryn E. McDougal¹,
Amanda Bowers¹, Kathleen M. Naughton¹, David J. Cutler⁴, Garry R. Cutting^{1*}

Author affiliations: ¹McKusick-Nathans Institute of Genetic Medicine and ²Division of Pediatric Endocrinology, Johns Hopkins University School of Medicine, Baltimore; ³Department of Pediatrics, National Naval Medical Center, Bethesda; and ⁴Department of Human Genetics, Emory University School of Medicine, Atlanta

***Corresponding author:** Dr. Garry R. Cutting, M.D.

Mailing address: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, Broadway Research Building 559, Baltimore, MD 21205 USA

Phone: (410) 955-1773

Fax: (410) 614-0213

E-mail: gcutting@jhmi.edu

ABSTRACT

Cystic fibrosis (CF), the most common lethal single gene disorder in Caucasians, is due to mutations in the *CFTR* gene. Twin and sibling analysis indicates that modifier genes, rather than allelic variation in *CFTR*, are responsible for most of the variability in severity of lung disease, the major cause of mortality in CF patients. We used a family-based approach to test for association between lung function and two functional SNPs (rs1800469, “-509” and rs1982073, “codon 10”) in the 5’ region of transforming growth factor-beta1 (*TGFBI*), a putative CF modifier gene. Quantitative transmission disequilibrium testing of 472 CF patient-parent-parent trios revealed that both *TGFBI* SNPs showed significant transmission distortion when patients were stratified by *CFTR* genotype. Although lung function and nutritional status are correlated in CF patients, there was no evidence of association between the *TGFBI* SNPs and variation in nutritional status. Additional tagging SNPs (rs8179181, rs2278422, rs8110090, rs4803455, and rs1982072) that capture most of the diversity in *TGFBI* were also typed but none showed association with variation in lung function. However, a haplotype composed of the -509 C and codon 10 T alleles along with the C allele of the 3’ SNP rs8179181 was highly associated with increased lung function in patients grouped by *CFTR* genotype. These results demonstrate that *TGFBI* is a modifier of CF lung disease and reveal a previously unrecognized beneficial effect of *TGFBI* variants upon the pulmonary phenotype.

INTRODUCTION

Cystic fibrosis (CF [MIM 602421]) is a common autosomal recessive genetic disorder caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene. There is a high degree of phenotypic variability in CF, especially in lung disease, the major cause of morbidity and mortality. Differences in lung disease severity are not entirely explained by *CFTR* genotype, as illustrated by the great extent of variability in patients homozygous for the most common mutation, $\Delta F508$ (1,2). Environmental factors such as infection with bacterial pathogens contribute to lung disease variation (3). However, affected twins and siblings demonstrate that variation in lung disease severity also has a strong genetic component, with heritability estimated at 0.6–0.8 (4). These observations indicate that modifier genes make a substantial contribution to CF lung disease, independent of *CFTR* genotype.

The moderate degree of similarity of CF lung disease severity between siblings suggests that a number of modifier genes are operating. These genes may individually be of small effect, requiring multiple replications to substantiate their relevance to CF lung disease. Indeed, numerous modifier genes have been examined, but few have withstood the test of replication (5). The most compelling biological candidate is transforming growth factor-beta1 (*TGFBI*). The *TGFBI* gene product is a secreted protein with numerous functions including involvement in cellular growth and differentiation, inflammation, and tissue fibrosis (6–8). Although several case-control studies have investigated *TGFBI* as a modifier of CF lung disease, the results have been conflicting (9–12). Discrepancies among these case-control studies may be due to population stratification, confounding of the studied trait by other phenotypes, interaction among

1
2
3 genes, undetected variants in the candidate gene that affect lung function, or differences
4 in power. The U.S. CF Twin and Sibling Study has recruited twins and siblings affected
5 with CF and their parents to evaluate association between candidate modifier genes and
6 variation in lung function. Use of these families permits the employment of
7 transmission-based methods that avoid spurious associations due to population
8 stratification. Patients were recruited only on the condition that they had a living sibling
9 with CF, thereby removing potential bias inherent in recruiting based on extreme
10 phenotypes and facilitating the analysis of the confounding effect of one trait upon
11 another. Additionally, patients bearing all *CFTR* genotypes were recruited, allowing the
12 investigation of gene-gene interaction between modifier genes and *CFTR*. Whenever
13 possible, the CF Twin and Sibling Study enrolled both parents of each patient to facilitate
14 the construction of haplotypes and the search for occult variation in candidate genes that
15 modify CF lung disease. Finally, the unfortunate relative commonness of CF and the
16 presence of a highly organized nationwide CF care system enabled recruitment of
17 hundreds of families, thus providing reasonable power for transmission-based association
18 studies.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

Association of *TGFBI* alleles with variation in lung function is dependent upon *CFTR* genotype

Patients in this study were similar for age, *CFTR* genotype distribution, body mass index and sex ratio compared to CF patients in the U.S. CF Patient Registry (Table 1) (13). Since lung disease is progressive in CF patients, pulmonary function measures of the forced expiratory volume in one second (FEV₁) were converted into disease-specific percentiles (FEV₁CF%) to facilitate comparisons among patients of different ages and sex (14). The mean percentiles of the CF-specific lung function measures were higher in our cohort than in unrelated patients (i.e. greater than 0.5); however, the entire spectrum of disease severity was represented. Quantitative transmission disequilibrium tests (QTDT) were performed on patient-parent-parent trios to evaluate association between polymorphisms in *TGFBI* and lung function in CF patients. QTDT tests for transmission distortion, or a skewing of the expected 50/50 ratio of alleles transmitted from heterozygous parents to offspring, across the spectrum of a quantitative trait. Though Z-scores of the cross-sectional (MaxFEV₁CF%) and longitudinal (AvgFEV₁CF%) lung function measures were employed, the results were essentially identical when non-transformed traits were used (data not shown). Two SNPs in *TGFBI* (rs1800469, “-509” and rs1982073, “codon 10”) that have shown variable association with lung function in case-control studies were typed in 472 trios (genotype frequency data provided in Supplementary Material, Table S1). The -509 SNP showed transmission distortion in patients stratified by the cross-sectional measure (p=0.011, achieving study-wide statistical significance) while transmission distortion of the codon 10 SNP approached

1
2
3 significance ($p=0.059$; Table 2). No difference in the parental origin of transmitted
4 codon 10 alleles was observed, in contrast to the report by Becker and colleagues (data
5 not shown) (15). These studies confirm that *TGFBI* is a modifier of lung function in
6 patients recruited by the CF Twin and Sibling Study.
7
8
9

10
11
12 To determine if *CFTR* genotype influenced *TGFBI* effect upon lung function,
13 patients were divided into two groups of nearly equal size: those who were homozygous
14 for the common $\Delta F508$ mutation (“ $\Delta F508$ homozygotes”) and all other genotypes (“non-
15 $\Delta F508$ homozygotes”). Eight patients with unknown *CFTR* genotype were excluded
16 from *CFTR*-genotype-specific analyses. The $\Delta F508$ homozygote and non- $\Delta F508$
17 homozygote groups had very similar means and standard deviations for cross-sectional
18 (0.68 ± 0.28 , 0.69 ± 0.26 , respectively) and longitudinal lung function (0.59 ± 0.24 , both
19 groups). When stratified by *CFTR* genotype it became apparent that non- $\Delta F508$
20 homozygotes were predominantly responsible for the observed transmission distortion of
21 both SNPs (Table 2). The -509 SNP demonstrated study-wide significant association
22 with the cross-sectional and the longitudinal measure of lung function in non- $\Delta F508$
23 homozygotes. Association between the codon 10 SNP and the cross-sectional measure of
24 lung function also achieved study-wide significance, and almost attained study-wide
25 significance for the longitudinal measure of lung function in non- $\Delta F508$ homozygotes
26 (Table 2).
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 Nutritional status, as measured by body mass index (BMI), is known to be
49 correlated with lung function measures in CF patients (16,17). As expected, the average
50 of Z-scores for BMI between ages 2–20 (AvgBMIZ) was correlated with both cross-
51 sectional and longitudinal lung function measures (Pearson’s $R=0.36$ and 0.44 ,
52 respectively; $p<0.0001$). QTDT performed in 457 trios in which both pulmonary and
53
54
55
56
57
58
59
60

1
2
3 BMI data were available revealed that neither *TGFBI* SNP showed association with
4 AvgBMIZ (data not shown). To further examine whether *TGFBI* alleles were associated
5 with variation in lung function or a composite of both lung function and BMI, we
6 adjusted for the confounding effect of BMI upon lung function using linear regression.
7
8 Association between the *TGFBI* SNPs and the cross-sectional and longitudinal
9 pulmonary traits adjusted for BMI (see Materials and Methods) was analyzed by QTDT.
10
11 The results were nearly identical to those for the unadjusted lung function measures,
12 demonstrating that variation in nutritional status did not account for the association
13 between *TGFBI* alleles and lung function (Table 2, “BMI-adjusted” columns). As noted
14 previously, transmission distortion of both the -509 and codon 10 SNPs was present only
15 in non- Δ F508 homozygotes.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 **Transmission analysis of additional SNPs identifies *TGFBI* haplotypes that are** 33 **associated with variation in lung function in CF patients**

34
35
36 To determine if certain combinations of variants in or near the *TGFBI* gene (i.e.
37 haplotypes) were associated with lung function measures, an additional eight SNPs
38 (rs8179181, rs28730295, rs2278422, rs8110090, rs11466334, rs1800472, rs4803455,
39 rs1982072) were typed in 445 trios with pulmonary data. The allele and genotype
40 frequencies observed in our population are similar to the frequencies reported in other
41 Caucasian subjects (data provided in Supplementary Material, Table S1). A
42 monomorphic SNP (rs28730295) and two SNPs with minor allele frequency (MAF)
43 below 0.05 (rs11466334, rs1800472) were dropped from further analysis. None of these
44 additional SNPs were individually associated with lung function measures in CF patients
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(data not shown), except for rs1982072, which was in almost complete linkage disequilibrium (LD) with the -509 SNP ($r^2=1.0$, $D'=1.0$).

Screening tools within PBAT (pedigree-based association testing) software (18) were used to select haplotypes of varying lengths and combinations of seven SNPs (-509, codon 10, rs8179181, rs2278422, rs8110090, rs4803455, and rs1982072) that were mostly likely to show association with the cross-sectional measure of lung function. Haplotypes comprised of -509 and codon 10 showed the highest power, especially in the presence of an additional SNP in intron 5 (rs8179181). Although the -509 and codon 10 SNPs were in strong LD with each other ($r^2=0.65$, $D'=0.99$), both SNPs were in low LD with the intron 5 SNP ($r^2<0.01$; $D'=0.10, 0.22$; respectively). The latter finding suggested that the improved power upon inclusion of the intron 5 SNP is not due to a high degree of ancestral correlation with the -509 and codon 10 SNPs, but that specific haplotypes composed of all 3 SNPs have greater association with variation in lung function than any SNP alone.

To test the above hypothesis, haplotypes composed of two SNPs (-509 and codon 10) and three SNPs (-509, codon 10 and intron 5; haplotype frequencies provided in Supplementary Material, Table S2) were tested for association with the cross-sectional lung function measure using QTDT and a second method, Family-Based Association Testing (FBAT). In the group of “All *CFTR* genotypes” the -509 T–codon 10 C (“T–C”) haplotype was significantly under-transmitted to patients with increasing lung function in both analyses (i.e. associated with more severe lung disease as indicated by the negative Z-statistic in FBAT; Table 3). The *TGFBI* haplotypes also demonstrated interaction with *CFTR* genotype, as was observed for individual SNPs. In non- $\Delta F508$ homozygotes, the T–C haplotype was under-transmitted with increasing lung function, as noted for the

entire group. Intriguingly, when the T–C haplotype was divided into two 3-SNP haplotypes based on the allele present at intron 5, the -509 T–codon 10 C–intron 5 C (“T–C–C”) haplotype was associated with decreased lung function, while the other haplotype, -509 T–codon 10 C–intron 5 T (“T–C–T”), showed no transmission distortion, even though the latter haplotype also contains a T at -509 and a C at codon 10. On the other hand, the -509 C–codon 10 T (“C–T”) haplotype was *over*-transmitted (i.e. associated with milder lung disease) in non- Δ F508 homozygotes, as shown by the positive Z-statistic. When the intron 5 SNP was included, the three-SNP haplotype -509 C–codon 10 T–intron 5 C (“C–T–C”) was *over*-transmitted to non- Δ F508 homozygous patients with improved lung function in both analyses while the -509 C–codon 10 T–intron 5 T (“C–T–T”) haplotype showed no transmission distortion (Table 3). Thus, haplotype transmission analysis provides further evidence that the association of genetic variants in *TGFBI* with CF lung function is better defined by haplotypes rather than individual alleles.

One *TGFBI* haplotype is correlated with improved lung function in non- Δ F508 homozygotes

To assess the relative magnitude of the effect of individual *TGFBI* haplotypes on the cross-sectional measure of lung function, we performed linear regression analysis under the assumption of additive, dominant, and recessive modes of inheritance (MOI). A measure of goodness of fit (Akaike’s information criterion) was calculated to compare models (19). For both univariate and multivariate regression, models which coded haplotypes according to additive or dominant MOIs had better fits than a model employing a recessive MOI. Since the resulting association estimates were qualitatively

1
2
3 similar in both additive and dominant scenarios, only the latter results are shown. Results
4 derived from the use of additive or recessive MOIs are available upon request. In the
5 univariate analysis, only the C–T–C haplotype had a significant relationship with lung
6 function in the complete group of patients (“All *CFTR* genotypes”; Table 4). The
7 positive value of the regression coefficient (β) associated with the C–T–C haplotype
8 indicates that this haplotype is associated with better lung function in CF patients. In
9 contrast, the C–T–T haplotype demonstrated no association with variation in lung
10 function, despite sharing with the C–T–C haplotype a C at -509 and a T at codon 10
11 (Table 4). Furthermore, the regression coefficient associated with the C–T–T haplotype
12 has a negative value, as opposed to the positive value for the C–T–C haplotype. In non-
13 $\Delta F508$ homozygous patients, the C–T–C haplotype was associated with an increase in
14 lung function of 15.9 percentile points, while the T–C–C haplotype was associated with a
15 decrease in lung function of 11.9 percentile points. Neither the C–T–T haplotype nor the
16 T–C–T haplotype demonstrated association with variation in lung function, despite
17 having the same alleles at -509 and codon 10 as the C–T–C and T–C–C haplotypes,
18 respectively (Table 4).
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 To verify that the magnitude of effect of the C–T–C haplotype (15.9 percentile
42 points) was greater than the effect of individual SNPs, we performed linear regression on
43 single alleles in 232 non- $\Delta F508$ homozygotes. As expected, the regression coefficients
44 were lower, at 12.0 ($p=0.045$) and 13.2 percentile points ($p=0.009$) for the -509 C and
45 codon 10 T alleles, respectively. Additionally, the -509 C–codon 10 T (C–T) haplotype
46 increased lung function by only 12.0 percentile points ($p=0.019$, $n=218$). These results
47 demonstrate that the “C” allele of the intron 5 SNP parses the ancestral haplotype block
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 of -509 and codon 10 SNPs into 3-SNP haplotypes that show enhanced effect upon lung
4
5 function.
6

7
8 The use of univariate regression described above allowed us to evaluate the
9
10 contribution of individual haplotypes to variation in lung disease; however, this modeling
11
12 could not account for the confounding effect of one haplotype upon another. For
13
14 example, association of the C–T–C haplotype with increased lung function might simply
15
16 be reflecting an absence of the T–C–C haplotype. Indeed, as these are the two most
17
18 common haplotypes in our population, patients who do not carry the former haplotype are
19
20 likely to carry the latter haplotype. To address this issue, we performed multivariate
21
22 linear regression to determine the effect of *TGFB1* C–T–C and T–C–C haplotypes (the
23
24 two haplotypes that showed association by univariate analysis) upon lung function while
25
26 simultaneously accounting for the effects of other haplotypes. When the relationship
27
28 between lung function and both the C–T–C and the T–C–C haplotypes was modeled in
29
30 non- $\Delta F508$ homozygotes, the C–T–C haplotype continued to have a positive influence on
31
32 CF lung disease while the T–C–C haplotype no longer showed a relationship with lung
33
34 function (Table 4). Patients who carried the former haplotype had an estimated increase
35
36 in lung function of 12.6 percentile points compared to patients who did not carry this
37
38 haplotype. This analysis implies that non- $\Delta F508$ homozygote CF patients with at least
39
40 one C–T–C haplotype will have better lung function than patients carrying *any* other
41
42 combination of the three SNPs that comprise this haplotype (i.e. T–C–C, T–C–T, C–T–T,
43
44 and C–C–C).
45
46
47
48
49
50
51

52
53 To confirm this conclusion, we plotted and compared the cross-sectional lung
54
55 function of patients bearing different combinations of *TGFB1* haplotypes (Figure 1).
56
57
58
59
60

1
2
3 Non- $\Delta F508$ homozygous patients carrying at least one C-T-C haplotype had
4
5 significantly higher median lung function (MaxFEV₁CF% = 0.84, n=106) than patients
6
7 carrying the T-C-C haplotype (MaxFEV₁CF%=0.58, n=51; p<0.0001, Mann-Whitney *U*
8
9 test). Of particular note, patients bearing both the C-T-C and T-C-C haplotypes (i.e. C-
10
11 T-C/T-C-C heterozygotes) had *higher* lung function (median MaxFEV₁CF%=0.79,
12
13 n=42) compared to those carrying the T-C-C haplotype in the absence of the C-T-C
14
15 haplotype (p=0.006, Mann-Whitney *U* test). This demonstrates the positive influence of
16
17 the C-T-C haplotype on lung function and argues against a negative influence of the T-
18
19 C-C haplotype. There was no difference in lung function between C-T-C/T-C-C
20
21 heterozygotes and patients bearing a C-T-C haplotype in *trans* with a non-T-C-C
22
23 haplotype. These results were unchanged when patients carrying no C-T-C or T-C-C
24
25 haplotypes (i.e. C-T-T, T-C-T, C-C-C; n=14) were also included in the group of T-C-
26
27 C-bearing patients (data not shown). Thus, the C-T-C haplotype manifests a dominant
28
29 positive effect on lung function in non- $\Delta F508$ homozygotes.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

Using family-based transmission analysis, we have shown that variation in the *TGFB1* gene is associated with variability in lung function in CF patients. More importantly, our study design facilitated the discovery of three new attributes of the modifier effect of *TGFB1*. First, our analyses revealed interaction between *TGFB1* and *CFTR*, as the effect of variation in *TGFB1* on lung function was primarily observed in patients who were *not* homozygous for $\Delta F508$. Second, variants in *TGFB1* were shown to modify lung function and not nutritional status, despite strong correlation between these two traits. Third, transmission and regression analyses revealed a *TGFB1* haplotype that is associated with better lung function, and hence mild lung disease, an observation of substantial therapeutic potential.

Several studies have demonstrated association between SNPs in *TGFB1* and variation in CF lung disease, while others have not. There are a number of key differences among these studies including design, power, definition of the pulmonary phenotype, and assignment of affection status. The best-powered case-control study to date defined severe and mild lung disease as having forced expiratory volume in 1 second (FEV₁) measurements in the lowest or highest quartiles for age, respectively. Of 808 $\Delta F508$ homozygous patients, those with severe lung disease were two times more likely to have the -509 TT and codon 10 CC genotypes than those with mildly impaired lung function. The authors replicated this finding in a second population of 498 CF patients, of whom 70% were $\Delta F508$ homozygotes and 30% had other “severe” *CFTR* genotypes associated with poorer clinical outcomes. Interestingly, this association was not observed

1
2
3 in the sub-group made up of $\Delta F508$ homozygotes only, suggesting that *TGFBI* variants
4 had greater modifying capacity in non- $\Delta F508$ homozygotes as seen in our population. In
5
6 contrast, Arkwright and colleagues showed in 171 $\Delta F508$ homozygotes that the codon 10
7
8 TT genotype was associated with an earlier decline in lung function . Brazova and
9
10 colleagues did not see association of *TGFBI* SNPs with variation in lung function in a
11
12 study of 118 Czech CF patients, of whom about half were $\Delta F508$ homozygotes and half
13
14 had other “severe” *CFTR* genotypes, and 268 control subjects . More recently, a study
15
16 examined the rate of decline in lung function of 511 Canadian CF patients stratified by
17
18 codon 10 genotype. Though a significant difference in the rate of decline was observed
19
20 between the three genotype groups (CC, CT, TT), the pattern was not entirely consistent
21
22 with previous reports as this study found that codon 10 heterozygotes (CT) had the
23
24 smallest annual decline in lung function (20). In the only other family-based study,
25
26 transmission distortion of -509 and codon 10 alleles at *TGFBI* was not observed in 34
27
28 pairs of extreme concordant and discordant $\Delta F508$ homozygote siblings using a
29
30 composite measure of lung function and body mass index . The current study of 472 trios
31
32 demonstrated that the -509 C and codon 10 T alleles had a beneficial modifier effect on
33
34 cross-sectional and longitudinal measures of lung function based upon FEV₁. From our
35
36 results we infer that the *absence* of the -509 C and codon 10 T alleles should be
37
38 associated with reduced lung function in CF patients. Indeed, the case-control study of
39
40 more than 1300 unrelated patients showed that the *presence* of the alternate alleles (-509
41
42 T and codon 10 C) were associated with lower FEV₁, fundamentally the same measure of
43
44 lung function used in this study. Thus, the results of our family-based study are
45
46 consistent with the association observed in the case-control study conducted by Drumm
47
48 and colleagues.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 The observations of the current study differ from all previous association studies
7
8 in that the modifier effect of *TGFBI* was dependent upon *CFTR* genotype. To reduce
9
10 genetic heterogeneity at the *CFTR* locus, prior studies primarily tested for association of
11
12 *TGFBI* with lung function in patients who were homozygous for the most common CF
13
14 mutation, $\Delta F508$. We detected association in patients who were *not* homozygous for
15
16 $\Delta F508$, but association was not observed in $\Delta F508$ homozygotes. This disparity does not
17
18 appear to be a function of power since the two groups had nearly identical numbers of
19
20 patients and had similar means and variances in lung function. We favor the concept that
21
22 the magnitude of the modifier effect conferred by *TGFBI* is smaller in $\Delta F508$
23
24 homozygotes than in CF patients carrying other *CFTR* mutations. By analyzing $\Delta F508$
25
26 homozygous patients with extreme phenotypes, the study by Drumm and colleagues had
27
28 80% power to detect modifiers that altered lung function by 0.7%. The 232 trios
29
30 homozygous for $\Delta F508$ in our study had only 11% power to detect the same change in
31
32 lung function. However, this number of trios did have reasonable power (80%) to detect
33
34 modifiers that account for about 11% of the total variance in lung function, an effect size
35
36 comparable to what was observed in the non- $\Delta F508$ homozygous trios. These findings
37
38 suggest that having two copies of $\Delta F508$ creates a unique lung disease environment that
39
40 is less responsive to variation in *TGFBI*. On the other hand, other *CFTR* mutations may
41
42 lead to lung pathology that is more amenable to alteration by modifiers such as *TGFBI*.
43
44 The latter concept is supported by the observation that the modifier effect of the
45
46 glutamate-cysteine ligase catalytic subunit (*GCLC*) gene upon severity of lung disease
47
48 could be observed only in CF patients with “mild” *CFTR* alleles (21).
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Patients were enrolled in the CF Twin and Sibling Study based on having a living
4 sibling with CF. Since patients were drawn from the entire spectrum of CF phenotypes,
5 we were able to test for association of *TGFBI* alleles with traits that are correlated with
6 lung function, such as nutritional status. Due to the ascertainment of only patients with
7 extreme phenotypes, prior studies could not distinguish with certainty whether *TGFBI*
8 variants were associated with lung disease severity, nutritional status or both . The
9 absence of association with a longitudinal measure of nutritional status and the presence
10 of association with lung function after adjustment for differences in nutritional status
11 indicates that variation in *TGFBI* primarily modifies CF lung disease. This observation
12 suggests that the lung should be the focus of studies examining the mechanism of the
13 *TGFBI* modifier effect in CF.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 Haplotype analysis is a powerful tool for discovering causal alleles that are highly
33 associated with typed markers and for determining whether combinations of alleles lead
34 to a greater effect upon phenotype than is caused by individual alleles. While single
35 alleles have been studied as modifiers in CF, *TGFBI* haplotypes have not previously been
36 explored. The *TGFBI* intron 5 SNP (rs8179181) by itself was not associated with
37 variation in lung function in this study. However, when the C allele of the intron 5 SNP
38 occurred on the same haplotype as -509 C and codon 10 T, this 3-SNP haplotype (C–T–
39 C) was shown to correlate with improved lung function in non- Δ F508 homozygous
40 patients. The intron 5 SNP has no functional role of which we are aware; thus we
41 propose that the -509 C–codon 10 T–intron 5 C haplotype contains a variant (or variants)
42 that modulate TGFB1 expression or function such that a protective outcome is conferred
43 on CF lungs. Another possible mechanism is that the codon 10 T and/or -509 C alleles in
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 combination with an additional variant on this haplotype are necessary to ameliorate CF
4 lung disease. The requirement of patients in this study to have a surviving affected
5
6 sibling probably accounts for the increased mean of cross-sectional and longitudinal lung
7
8 function measures compared to the CF population mean (Table1). The bias of the
9
10 patients in this study toward better lung function may have aided in the detection of the
11
12 “mild” *TGFB1* C–T–C haplotype.
13
14
15
16
17

18
19
20 Variation in *TGFB1* has been linked to several chronic pulmonary disorders in
21 addition to cystic fibrosis. For example, the -509 TT genotype was shown to associate
22 with the severity and diagnosis of asthma (22,23), similar to the observations in this study
23 of CF. In contrast, the -509 T and codon 10 C alleles were found to be protective against
24 chronic obstructive pulmonary disease (COPD) (24–26). These observations suggest that
25 the mechanism of *TGFB1* action upon lung function is context-specific. The *TGFB1*
26
27 -509 T and codon 10 C alleles have been linked to higher gene and protein expression
28 than the alternate alleles at these loci (27–32). *TGFB1* is known to promote fibrogenesis
29 by stimulating extracellular matrix production (33) and by inhibiting matrix degradation
30 (34). Alleles that decrease levels or activity of *TGFB1* may be predicted to stimulate an
31 appropriate balance between tissue repair and fibrosis that improves lung function in CF
32 patients. Our discovery that *TGFB1* variants associate with mild lung disease in a subset
33 of CF patients (non- $\Delta F508$ homozygotes) provides new opportunity to test the
34
35 aforementioned prediction and to possibly develop therapeutics that retard progression of
36
37 this life-limiting feature of CF.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

MATERIALS AND METHODS

Study subjects

CF twins and siblings (n=617) and their parents (n=606) from 303 families were recruited by the CF Twin and Sibling Study as previously described . Of these families, 88.4% had two children old enough to perform pulmonary function testing, 7.6% had three children, and 4.0% had one child. Sixteen dizygous and thirteen monozygous (MZ) twin pairs were included. Blood samples were obtained from patients and parents for standard DNA phenol/chloroform extraction. Raw pulmonary function test data, *CFTR* genotypes, and height and weight measurements were obtained from medical records. In some cases in which genotypes were unavailable, *CFTR* exons were sequenced to identify mutations. Written informed consent or assent was obtained from all subjects. Only families in which both parents of the patients were available (i.e. complete trios) were included in the present study. Analyses were conducted using data from all patients, as well as using data from subgroups of patients categorized by *CFTR* genotype: those homozygous for the $\Delta F508$ mutation ($\Delta F508$ homozygotes) and those bearing all other genotypes (non- $\Delta F508$ homozygotes).

Phenotype definition

The forced expiratory volume in one second (FEV_1), a lung function measure that is highly correlated with survival in cystic fibrosis patients (35,36), was used to derive cross-sectional ($MaxFEV_1CF\%$) and longitudinal ($AvgFEV_1CF\%$) measures, as previously described . All patients in this study with a longitudinal lung function measure also had a cross-sectional measure. However, 142 patients were too young to

1
2
3 have sufficient data for the longitudinal measure. Of the MZ twin pairs in which both
4 twins had pulmonary data, ten pairs had MaxFEV₁CF% and four pairs had
5 AvgFEV₁CF%. To include as many subjects as possible and to avoid randomly
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

have sufficient data for the longitudinal measure. Of the MZ twin pairs in which both twins had pulmonary data, ten pairs had MaxFEV₁CF% and four pairs had AvgFEV₁CF%. To include as many subjects as possible and to avoid randomly excluding one member of each pair, lung function measures were averaged for MZ twin pairs and included in analyses only if the twins' values were within ten percentiles of each other, as not to double-count genetically identical individuals. For MZ twin pairs in which only one of the twins had pulmonary data, that twin's data was included. The average of Z-scores for body mass index (BMI) between ages 2–20 years (AvgBMIZ) was used as a longitudinal marker of nutritional status. Pulmonary phenotypes were adjusted for nutritional status by regressing lung function measures on AvgBMIZ, and then for each individual the product of the regression coefficient and AvgBMIZ was subtracted from the lung function measure.

SNP selection and genotyping

Two SNPs in *TGFBI*, c.-1347C>T (rs1800469, -509) and c.29T>C (rs1982073, codon 10), were chosen based on the findings of a recent study on modifiers of CF lung disease . These SNPs were genotyped in all individuals using TaqMan Assays-on-Demand and Assays-by-Design, respectively (Applied Biosystems, Foster City, CA). Reactions were performed in 384-well plates in a total reaction volume of 10 µl with 10 ng of template DNA in a Bio-Rad iCycler thermal cycler. Quality control samples were included on each plate. Endpoint fluorescence readings were obtained using an ABI PRISM 7900HT sequence detection system (SDS; Applied Biosystems). Genotype-calling was conducted using SDS v.2.1 software and inheritance checking was performed by SIB-PAIR v.0.99.9 (<http://www2.qimr.edu.au/davidD>) (37). In addition, -509

1
2
3 genotypes were verified using allele-specific oligonucleotide linear arrays (Roche
4 Molecular Systems, Alameda, CA). The discrepancy rate between these two methods
5
6 was 0.91%. Unclear or erroneous genotypes were either repeated by TaqMan or
7
8 sequenced using the BigDye Terminator v.3.1 Cycle Sequencing Kit on an ABI 3100
9
10 sequencer (Applied Biosystems).
11
12
13
14

15 Tagger (<http://www.broad.mit.edu/mpg/tagger>) (38), the tag SNP selection
16
17 algorithm implemented in Haploview (<http://www.broad.mit.edu/mpg/haploview>) (39),
18
19 was used to select a minimal set of tag SNPs that adequately represented the genetic
20
21 diversity within the *TGFBI* gene region, based on patterns of linkage disequilibrium
22
23 found in the HapMap CEPH data (40) and Perlegen Caucasian data. The data source at
24
25 the time of SNP selection was HapMap data release 20/phaseII Jan06 on the NCBI build
26
27 35 assembly, dbSNP build 125. Criteria for selection of tag SNPs were an $r^2 > 0.8$ with
28
29 the untyped SNP, an Illumina design score > 0.6 , and an inter-SNP spacing of no less than
30
31 60 base-pairs. Eight tag SNPs (rs8179181, rs28730295, rs2278422, rs8110090,
32
33 rs11466334, rs1800472, rs4803455, rs1982072) within and 5 kb upstream of the *TGFBI*
34
35 gene were genotyped in 445 trios with pulmonary data using Illumina BeadArray
36
37 technology (Illumina, San Diego, USA). Allele and genotype frequencies are provided in
38
39 Supplementary Material, Table S1.
40
41
42
43
44
45
46
47

48 **Statistical analyses**

49
50 Genotype distributions were tested for Hardy-Weinberg equilibrium using the
51
52 ‘--unrelatedsOnly’ option in PEDSTATS v.0.6.6
53
54 (<http://www.sph.umich.edu/csg/abecasis/Pedstats>) (41), which performs an exact test in a
55
56 subset of unrelated individuals, as to avoid bias from correlated genotypes within
57
58
59
60

1
2
3 families. General statistics, linear regression, Mann-Whitney U tests, and t-tests were
4
5 performed in Intercooled Stata 8 (StataCorp, College Station, TX). Because correlation
6
7 among sibling marker genotypes may invalidate the results of family-based tests of
8
9 association in the presence of linkage, we tested for linkage of MaxFEV₁CF% to the
10
11 *TGFBI* gene region on chromosome 19. Single- and multipoint parametric linkage
12
13 analysis was performed using all SNPs with MAF>0.05 and two previously typed short
14
15 tandem repeat markers downstream of *TGFBI* (D19S400, D19S718) using Sequential
16
17 Oligogenic Linkage Analysis Routines (SOLAR v.4.0.7; <http://www.sfbr.org/solar>). No
18
19 linkage of the cross-sectional measure of lung function to this region was found in CF
20
21 patients.
22
23
24
25

26
27 Quantitative transmission disequilibrium testing (QTDT v.2.5.0;
28
29 <http://www.sph.umich.edu/csg/abecasis/QTDT>) (42) was used to perform family-based
30
31 tests of linkage disequilibrium. The orthogonal model implemented in QTDT was
32
33 adopted to test for association. To account for multiple testing in the presence of linked
34
35 polymorphisms, empirical p-values were calculated from 1,000 Monte-Carlo
36
37 permutations using the '-m' option. Though single-test p-values are reported, those
38
39 meeting the threshold for a global empirical significance level of 0.05 are denoted by
40
41 bold font.
42
43
44

45
46 Data analysis tools for continuous traits implemented within PBAT software
47
48 (<http://www.biostat.harvard.edu/~clange/default.htm>) were used to estimate the power of
49
50 the data set to detect an association between *TGFBI* haplotypes and MaxFEV₁CF% using
51
52 the conditional mean model (43) while simultaneously minimizing the number of tests for
53
54 which to correct. Haplotypes with greater than 80% power (at $\alpha=0.05$) were analyzed by
55
56 QTDT and by a second method, the Family-Based Association Testing program (FBAT;
57
58
59
60

1
2
3 <http://www.biostat.harvard.edu/~fbat/fbat.htm>) (44). Single-test p-values meeting the
4 threshold for a global empirical significance level of 0.05 are denoted by bold font and p-
5 values achieving significance after Bonferroni correction are marked with an asterisk.
6
7
8
9
10 Because QTDT does not internally generate haplotypes, PHASE v.2.1
11 (<http://stephenslab.uchicago.edu/software.html>) (45) was used to construct haplotypes
12 from the trio data via the '-P1' option. Patients with recombinant or ambiguous
13 haplotypes were excluded. The constructed haplotypes were treated as individual alleles
14 in QTDT. The 'hbat' command implemented within FBAT was employed to construct
15 and test haplotypes comprised of specified SNPs. The '-p' option in FBAT was
16 employed to compute empirical p-values from Monte-Carlo permutations and also to
17 perform the 'minimal p test', which calculates the significance of the smallest p-value.
18 Since the family-based association tests employed (QTDT, PBAT and FBAT) have
19 optimal power when traits are normally distributed, all quantitative phenotypes were also
20 ranked and converted to Z-scores using an inverse normal transformation.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 For univariate and multivariate linear regression of the cross-sectional lung
37 function measure (MaxFEV₁CF%) on *TGFBI* variants, alleles or haplotypes were coded
38 in an additive, dominant, or recessive fashion to determine the mostly likely mode of
39 inheritance. For the various models, alleles or haplotypes were coded as follows:
40 additive, 0 (zero copies of the allele/haplotype present), 1 (one copy, i.e. heterozygous),
41 or 2 (i.e. homozygous); dominant, 0 (zero copies) or 1 (at least one copy); recessive, 0
42 (zero or one copies) or 1 (two copies). Akaike's information criterion (AIC), a measure
43 of the goodness of fit of an estimated statistical model, was used to compare models .
44
45
46
47
48
49
50
51
52
53
54
55
56 Power analysis was performed using the online Genetic Power Calculator
57 (<http://pngu.mgh.harvard.edu/~purcell/gpc/>) (46).
58
59
60

ACKNOWLEDGMENTS

The authors sincerely thank the many CF patients and their families, research coordinators, nurses, and physicians involved with the CF Twin and Sibling Study, as this work would not have been possible without their participation. We also thank Nulang Wang for DNA extraction and *CFTR* genotyping, Michal Kulich, Ph.D. for providing conversion programs for CF-specific percentiles for FEV₁, and John McGready, Ph.D. for helpful statistical discussions.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Institutes of Health (HL68927) and the Cystic Fibrosis Foundation (CUTTIN06P0).

For Peer Review

REFERENCES

1. Kerem, E., Corey, M., Kerem, B.-S., Rommens, J., Markiewicz, D., Levison, H., Tsui, L.C. and Durie, P. (1990) The relation between genotype and phenotype in cystic fibrosis--analysis of the most common mutation (deltaF508). *N. Engl. J. Med.*, **323**, 1517–1522.
2. Burke, W., Aitken, M.L., Chen, S.H. and Scott, C.R. (1992) Variable severity of pulmonary disease in adults with identical cystic fibrosis mutations. *Chest*, **102**, 506–509.
3. Kerem, E., Corey, M., Gold, R. and Levison, H. (1990) Pulmonary function and clinical course in patients with CF after pulmonary colonization with *Pseudomonas aeruginosa*. *J. Pediatr.*, **116**, 714–719.
4. Vanscoy, L.L., Blackman, S.M., Collaco, J.M., Bowers, A., Lai, T., Naughton, K., Algire, M., McWilliams, R., Beck, S., Hoover-Fong, J. *et al.* (2007) Heritability of lung disease severity in cystic fibrosis. *Am. J. Respir. Crit. Care Med.*, **175**, 1036–1043.
5. Cutting, G.R. (2005) MODIFIER GENETICS: Cystic Fibrosis. *Annu. Rev. Genomics Hum. Genet.*, **6**, 237–260.
6. Blobe, G.C., Schiemann, W.P. and Lodish, H.F. (2000) Role of transforming growth factor beta in human disease. *N. Engl. J. Med.*, **342**, 1350–1358.
7. Bartram, U. and Speer, C.P. (2004) The role of transforming growth factor beta in lung development and disease. *Chest*, **125**, 754–765.
8. Sheppard, D. (2006) Transforming growth factor beta: a central modulator of pulmonary and airway inflammation and fibrosis. *Proc. Am. Thorac. Soc.*, **3**, 413–417.
9. Arkwright, P.D., Laurie, S., Super, M., Pravica, V., Schwarz, M.J., Webb, A.K. and Hutchinson, I.V. (2000) TGF-beta(1) genotype and accelerated decline in lung function of patients with cystic fibrosis. *Thorax*, **55**, 459–462.
10. Drumm, M.L., Konstan, M.W., Schluchter, M.D., Handler, A., Pace, R., Zou, F., Zariwala, M., Fargo, D., Xu, A., Dunn, J.M. *et al.* (2005) Genetic modifiers of lung disease in cystic fibrosis. *N. Engl. J. Med.*, **353**, 1443–1453.
11. Stanke, F., Tummler, B. and Becker, T. (2006) Genetic modifiers in cystic fibrosis. *N. Engl. J. Med.*, **354**, 88–90.
12. Brazova, J., Sismova, K., Vavrova, V., Bartosova, J., Macek, M., Jr., Lauschman, H. and Sediva, A. (2006) Polymorphisms of TGF-beta1 in cystic fibrosis patients. *Clin. Immunol.*, **121**, 350–357.

13. Cystic Fibrosis Foundation (2005) Cystic Fibrosis Foundation Patient Registry Annual Data Report 2005.
14. Kulich, M., Rosenfeld, M., Campbell, J., Kronmal, R., Gibson, R.L., Goss, C.H. and Ramsey, B. (2005) Disease-specific Reference Equations for Lung Function in Patients with Cystic Fibrosis. *Am. J. Respir. Crit. Care Med.*, **172**, 885–891.
15. Becker, T., Jansen, S., Tamm, S., Wienker, T.F., Tummler, B. and Stanke, F. (2007) Transmission ratio distortion and maternal effects confound the analysis of modulators of cystic fibrosis disease severity on 19q13. *Eur. J. Hum. Genet.*, **15**, 774–778.
16. Konstan, M.W., Butler, S.M., Wohl, M.E., Stoddard, M., Matousek, R., Wagener, J.S., Johnson, C.A. and Morgan, W.J. (2003) Growth and nutritional indexes in early life predict pulmonary function in cystic fibrosis. *J. Pediatr.*, **142**, 624–630.
17. Milla, C.E. (2004) Association of nutritional status and pulmonary function in children with cystic fibrosis. *Curr. Opin. Pulm. Med.*, **10**, 505–509.
18. Lange, C., DeMeo, D., Silverman, E.K., Weiss, S.T. and Laird, N.M. (2004) PBAT: tools for family-based association studies. *Am. J. Hum. Genet.*, **74**, 367–369.
19. Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Biomed. Eng.*, **19**, 716–719.
20. Dorfman, R., Sandford, A., Taylor, C., Huang, B., Frangolias, D., Wang, Y., Sang, R., Pereira, L., Sun, L., Berthiaume, Y. *et al.* (2008) Complex two-gene modulation of lung disease severity in children with cystic fibrosis. *J. Clin. Invest.*, **118**, 1040–1049.
21. McKone, E.F., Shao, J., Frangolias, D.D., Keener, C.L., Shephard, C.A., Farin, F.M., Tonelli, M.R., Pare, P.D., Sandford, A.J., Aitken, M.L. *et al.* (2006) Variants in the glutamate-cysteine-ligase gene are associated with cystic fibrosis lung disease. *Am. J. Respir. Crit. Care Med.*, **174**, 415–419.
22. Pulleyn, L.J., Newton, R., Adcock, I.M. and Barnes, P.J. (2001) TGFbeta1 allele association with asthma severity. *Hum. Genet.*, **109**, 623–627.
23. Silverman, E.S., Palmer, L.J., Subramaniam, V., Hallock, A., Mathew, S., Vallone, J., Faffe, D.S., Shikanai, T., Raby, B.A., Weiss, S.T. *et al.* (2004) Transforming growth factor-beta1 promoter polymorphism C-509T is associated with asthma. *Am. J. Respir. Crit. Care Med.*, **169**, 214–219.
24. Wu, L., Chau, J., Young, R.P., Pokorny, V., Mills, G.D., Hopkins, R., McLean, L. and Black, P.N. (2004) Transforming growth factor-beta1 genotype and susceptibility to chronic obstructive pulmonary disease. *Thorax*, **59**, 126–129.
25. Cledon, J.C., Lange, C., Raby, B.A., Litonjua, A.A., Palmer, L.J., Demeo, D.L., Reilly, J.J., Kwiatkowski, D.J., Chapman, H.A., Laird, N. *et al.* (2004) The

- transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Hum. Mol. Genet.*, **13**, 1649–1656.
26. van Diemen, C.C., Postma, D.S., Vonk, J.M., Bruinenberg, M., Nolte, I.M. and Boezen, H.M. (2006) Decorin and TGF-beta1 polymorphisms and development of COPD in a general population. *Respir. Res.*, **7**, 89.
 27. Yamada, Y., Miyauchi, A., Goto, J., Takagi, Y., Okuizumi, H., Kanematsu, M., Hase, M., Takai, H., Harada, A. and Ikeda, K. (1998) Association of a polymorphism of the transforming growth factor-beta1 gene with genetic susceptibility to osteoporosis in postmenopausal Japanese women. *J. Bone Miner. Res.*, **13**, 1569–1576.
 28. Grainger, D.J., Heathcote, K., Chiano, M., Snieder, H., Kemp, P.R., Metcalfe, J.C., Carter, N.D. and Spector, T.D. (1999) Genetic control of the circulating concentration of transforming growth factor type beta1. *Hum. Mol. Genet.*, **8**, 93–97.
 29. Suthanthiran, M., Li, B., Song, J.O., Ding, R., Sharma, V.K., Schwartz, J.E. and August, P. (2000) Transforming growth factor-beta 1 hyperexpression in African-American hypertensives: A novel mediator of hypertension and/or target organ damage. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 3479–3484.
 30. Luedeking, E.K., DeKosky, S.T., Mehdi, H., Ganguli, M. and Kamboh, M.I. (2000) Analysis of genetic polymorphisms in the transforming growth factor-beta1 gene and the risk of Alzheimer's disease. *Hum. Genet.*, **106**, 565–569.
 31. Dunning, A.M., Ellis, P.D., McBride, S., Kirschenlohr, H.L., Healey, C.S., Kemp, P.R., Luben, R.N., Chang-Claude, J., Mannermaa, A., Kataja, V. *et al.* (2003) A transforming growth factorbeta1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. *Cancer Res.*, **63**, 2610–2615.
 32. Shah, R., Hurley, C.K. and Posch, P.E. (2006) A molecular mechanism for the differential regulation of TGF-beta1 expression due to the common SNP -509C-T (c. -1347C > T). *Hum. Genet.*, **120**, 461–469.
 33. Imitola, J., Imitola, R.A. and Massague, J. (1986) Transforming growth factor-beta stimulates the expression of fibronectin and collagen and their incorporation into the extracellular matrix. *J. Biol. Chem.*, **261**, 4337–4345.
 34. Overall, C.M., Wrana, J.L. and Sodek, J. (1989) Independent regulation of collagenase, 72-kDa progelatinase, and metalloendoproteinase inhibitor expression in human fibroblasts by transforming growth factor-beta. *J. Biol. Chem.*, **264**, 1860–1869.
 35. Beaty, T.H., Cohen, B.H., Newill, C.A., Menkes, H.A., Diamond, E.L. and Chen, C.J. (1982) Impaired pulmonary function as a risk factor for mortality. *Am. J. Epidemiol.*, **116**, 102–113.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
36. Kerem, E., Reisman, J., Corey, M., Canny, G.J. and Levison, H. (1992) Prediction of mortality in patients with cystic fibrosis. *N. Engl. J. Med.*, **326**, 1187–1191.
37. Duffy, D. (1997) Sib-pair: a program for non-parametric linkage/association analysis. *Am. J. Hum. Genet.*, **61 Suppl**, A197.
38. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat. Genet.*, **37**, 1217–1223.
39. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
40. (2003) The International HapMap Project. *Nature*, **426**, 789–796.
41. Wigginton, J.E. and Abecasis, G.R. (2005) PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics*, **21**, 3445–3447.
42. Abecasis, G.R., Cardon, L.R. and Cookson, W.O. (2000) A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.*, **66**, 279–292.
43. Lange, C., DeMeo, D., Silverman, E.K., Weiss, S.T. and Laird, N.M. (2003) Using the noninformative families in family-based association tests: a powerful new testing strategy. *Am. J. Hum. Genet.*, **73**, 801–811.
44. Laird, N.M., Horvath, S. and Xu, X. (2000) Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.*, **19 Suppl 1**, S36–S42.
45. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
46. Purcell, S., Cherny, S.S. and Sham, P.C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19**, 149–150.

LEGENDS TO FIGURES

Figure 1. Lung function in non- $\Delta F508$ homozygous CF patients carrying different combinations of *TGFBI* -509-codon 10-intron 5 haplotypes. Circles represent individual patients' cross-sectional lung function measures, given as MaxFEV₁CF%. Horizontal lines indicate median lung function. "Non-T-C-C" refers to haplotypes other than T-C-C and "non-C-T-C" refers to haplotypes other than C-T-C. Lung function was higher in patients carrying the C-T-C haplotype (C-T-C/non-T-C-C, n=106) compared to those carrying the T-C-C haplotype (T-C-C/non-C-T-C, n=51; p<0.0001, Mann-Whitney *U* test). Patients bearing both the C-T-C and T-C-C haplotypes (C-T-C/T-C-C heterozygotes, n=42) also had higher lung function than patients who carried the T-C-C but not the C-T-C haplotype (p=0.006, Mann-Whitney *U* test). There was no difference in lung function between patients with C-T-C/non-T-C-C haplotypes and patients who were C-T-C/T-C-C heterozygotes.

Table 1. Characteristics of CF patients stratified by lung function

Variable	Trait	
	Cross-sectional ^a	Longitudinal ^b
Number of patients	472	330
Mean \pm standard deviation	.69 \pm .27	.59 \pm .24
Age Range (yr)	6–40	10–40
Mean age (yr)	14.9 \pm 6.5	17.4 \pm 6.0
Homozygous for Δ F508 (%)	50.0	50.8
AvgBMIZ ^c	-.19 \pm .83 (n=457)	-.31 \pm .84 (n=320)
Sex (% male)	48.9	48.8

^a MaxFEV₁CF%, the best CF-specific percentile for FEV₁ within the most recent year of data

^b AvgFEV₁CF%, the average of CF-specific percentiles for FEV₁ over a minimum of 4 years of data

^c The average of all BMI Z-scores calculated from available height and weight information for subjects between ages 2–20

Table 2. Transmission of *TGFBI* variants in CF patients stratified by quantitative lung function measures

Trait	-509						Codon 10					
	n ^d	F	p-val ^e	BMI-adjusted ^f			n	F	p-val	BMI-adjusted		
				n	F	p-val				n	F	p-val
<i>Cross-sectional^a</i>												
All <i>CFTR</i> Genotypes	296	6.48	.011	287	6.15	.014	337	3.58	.059	329	3.64	.057
ΔF508 Homozygotes	142	1.18	.279	141	1.15	.289	153	<.01	.982	152	<.01	.967
Non-ΔF508 Homozygotes ^b	150	7.21	.008	142	7.03	.009	179	8.76	.003	172	9.63	.002
<i>Longitudinal^c</i>												
All <i>CFTR</i> Genotypes	201	3.10	.079	196	2.61	.107	231	1.33	.250	227	1.28	.258
ΔF508 Homozygotes	99	.48	.488	99	.34	.560	108	.05	.818	108	<.01	.993
Non-ΔF508 Homozygotes	100	4.76	.031	95	4.52	.035	121	4.23	.041	117	3.59	.060

^a Z-score of MaxFEV₁CF% (see Materials and Methods)

^b Refers to patients who are not homozygous for the ΔF508 mutation (but may have 0 or 1 ΔF508 mutation)

^c Z-score of AvgFEV₁CF% (see Materials and Methods)

^d Number of informative trios (i.e. those in which at least one parent is heterozygous)

^e Single test p-values are shown; those with study-wide significance (p<.05) using Monte-Carlo permutation are shown in bold

^f Lung function measure adjusted for nutritional status (AvgBMIZ; see Materials and Methods)

Table 3. Transmission of *TGFBI* haplotypes in CF patients stratified by cross-sectional lung function

Group	-509 codon 10 intron 5 ^a	Freq ^b	QTDT ^c			FBAT		
			n ^d	F	p-val ^e	n ^f	Z	p-val ^e
All <i>CFTR</i> genotypes	C-C	.095	126	.40	.525	72.0	.63	.531
	C-C-C	.087	109	1.23	.268	61.7	.90	.371
	C-T	.602	308	3.48	.063	164.0	1.79	.074
	C-T-C	.446	314	2.42	.120	164.9	1.71	.088
	C-T-T	.155	199	.03	.861	105.4	.22	.823
	T-C	.301	267	6.33	.012*	141.0	-2.42	.015*
	T-C-C	.241	239	6.43	.012	124.6	-2.15	.032
	T-C-T	.061	79	.22	.638	43.2	-.69	.493
Δ F508 homozygotes	C-C		57	2.39	.124	35.0	1.38	.167
	C-C-C		48	5.82	.017	29.8	2.01	.044
	C-T		145	.01	.977	82.0	.03	.976
	C-T-C		156	.41	.523	86.7	-.48	.635
	C-T-T		109	.62	.431	58.8	.59	.552
	T-C		131	1.58	.210	73.0	-1.19	.236
	T-C-C		112	.92	.340	62.6	-.69	.489
	T-C-T		48	.42	.520	26.9	-.53	.598
Non- Δ F508 homozygotes	C-C		69	.21	.644	37.0	-.53	.595
	C-C-C		61	.29	.590	31.9	-.64	.526
	C-T		163	6.67	.011*	82.0	2.32	.020
	C-T-C		158	7.71	.006*	79.2	2.71	.007*
	C-T-T		90	.26	.608	47.6	-.22	.825
	T-C		136	6.00	.015*	68.0	-2.13	.033
	T-C-C		127	7.33	.007*	62.0	-2.14	.033
	T-C-T		31	.01	.940	16.4	-.45	.653

^a *TGFBI* SNPs in haplotype: -509 (rs1800469), codon 10 (rs1982073), intron 5 (rs8179181)

^b Frequencies of FBAT-generated haplotypes derived from 445 trios (frequencies of additional rare haplotypes provided in Supplementary Material, Table S2)

^c Haplotypes were constructed using PHASE and treated as individual alleles in QTDT (haplotype frequencies provided in Supplementary Material, Table S2)

^d Number of informative trios

^e Single test p-values are shown; those with study-wide significance ($p < .05$) using Monte-Carlo permutation are shown in bold. P-values with study-wide significance after Bonferroni correction are marked with an asterisk (*).

^f Number of informative families

Table 4. Effect of *TGFBI* haplotypes on lung function

	-509 codon 10 intron 5	All <i>CFTR</i> genotypes (n=437)			Δ F508 homozygotes (n=221)			Non- Δ F508 homozygotes (n=213)		
		β^b	95% CI ^c	p-val	β	95% CI	p-val	β	95% CI	p-val
Univariate ^a										
	C-T-C	.071	(.018 – .124)	.009	-.010	(-.082 – .063)	.788	.159	(.081 – .236)	.0001
	C-T-T	-.048	(-.105 – .008)	.090	-.021	(-.096 – .055)	.590	-.081	(-.166 – .004)	.062
	T-C-C	-.044	(-.094 – .005)	.081	.027	(-.041 – .095)	.440	-.119	(-.192 – -.046)	.001
	T-C-T	.037	(-.044 – .118)	.365	.031	(-.070 – .132)	.543	.044	(-.092 – .181)	.523
Multivariate ^d										
	C-T-C	.062	(.003 – .121)	.040	.002	(-.077 – .081)	.966	.126	(.039 – .213)	.005
	T-C-C	-.019	(-.075 – .036)	.489	.027	(-.047 – .102)	.469	-.065	(-.146 – .016)	.115

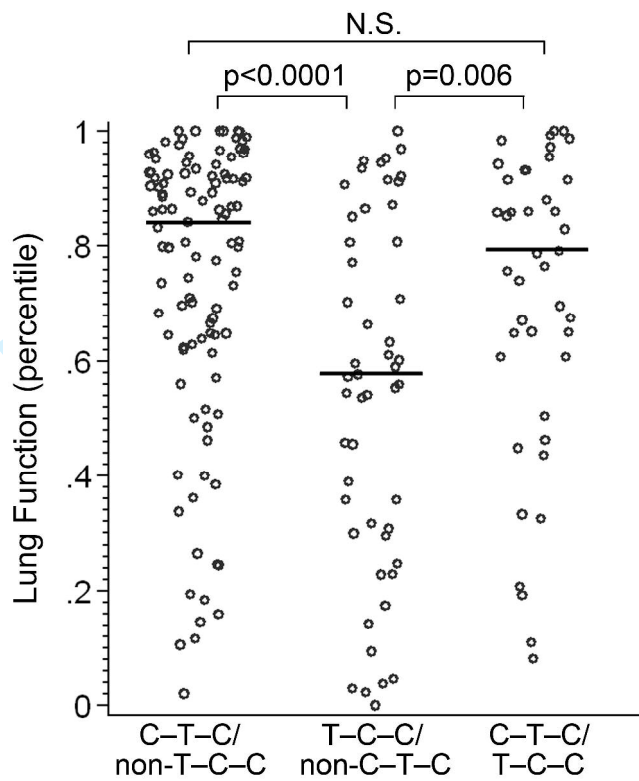
^a Univariate linear regression of MaxFEV₁CF%, assuming a dominant mode of inheritance

^b Regression coefficient

^c Confidence interval

^d Only the two haplotypes shown to be significantly associated (p<0.05) with lung function in univariate analysis (in non- Δ F508 homozygotes) were analyzed by multivariate linear regression

Figure 1



ABBREVIATIONS

CF, cystic fibrosis

QTDT, quantitative transmission disequilibrium test

MaxFEV₁CF%, cross-sectional measure of CF-specific percentiles for FEV₁

AvgFEV₁CF%, longitudinal measure of CF-specific percentiles for FEV₁

MZ, monozygous

BMI, body mass index

AvgBMIZ, average of Z-scores for body mass index between ages 2–20 years

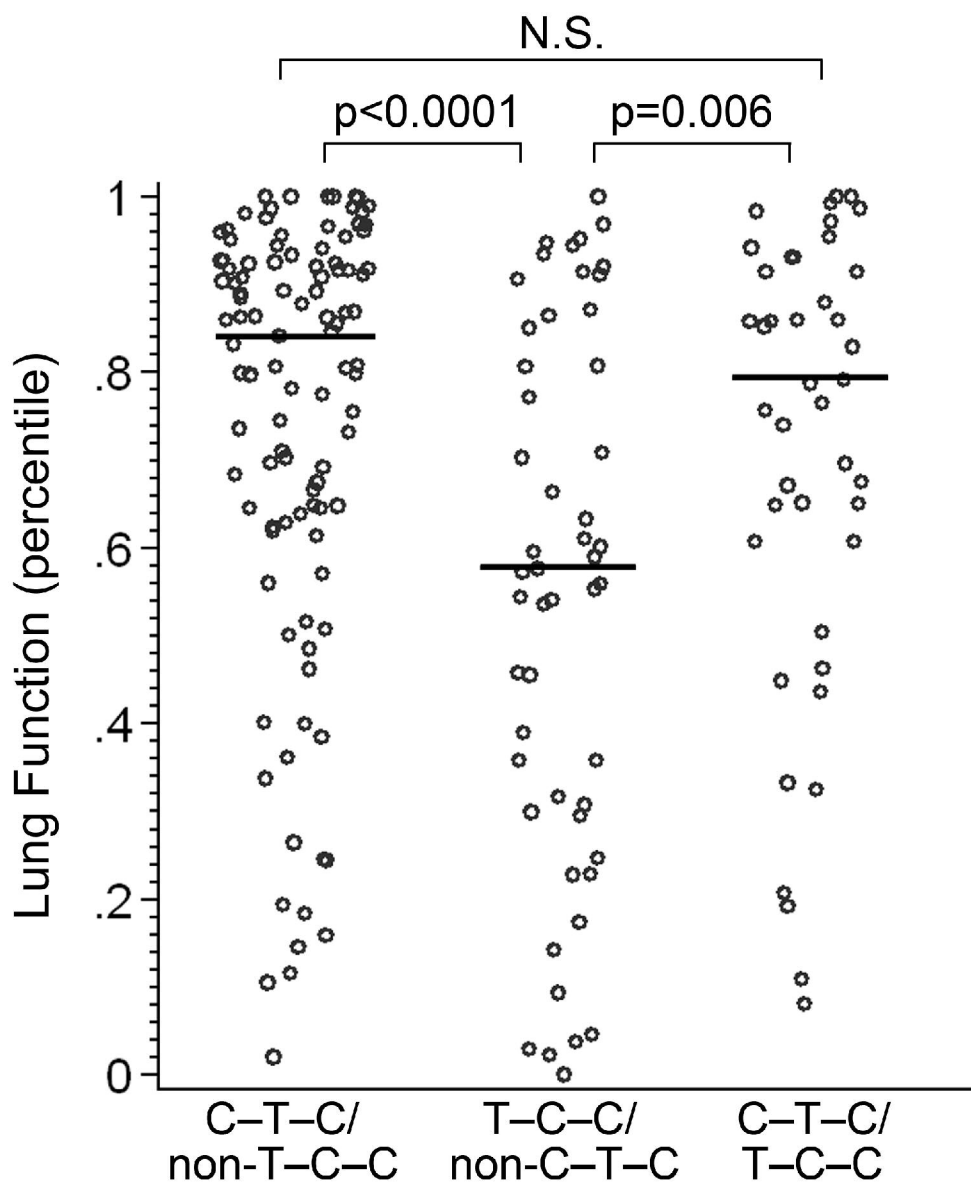
MAF, minor allele frequency

FBAT, family-based association test

MOI, mode of inheritance

FEV₁, forced expiratory volume in one second

LD, linkage disequilibrium



85x104mm (600 x 600 DPI)