

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Drumm ML, Konstan MW, Schluchter MD, et al. Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med* 2005;353:1443-53.

05-1469: Drumm et al., Gene modifiers of lung disease in cystic fibrosis

(Supplement)

Methods (Supplement)

Subjects: For the initial (first) study, the study population consisted of $\Delta F508$ homozygotes enrolled from 44 CF Centers, using entry criteria of FEV₁ (% predicted, by Knudsen equation) falling into the lowest (“severe”) and highest (“mild”) quartiles of lung function for age, as derived from the U.S. CFF Registry (1999) for $\Delta F508$ homozygotes. Subjects with “severe” disease had to be at least eight years, but could be no older than 25 years, because those subjects (by survival) would exceed the 25th percentile for their birth cohort. To be considered “severe”, subjects ages 8 to 25 years had current FEV₁ % predicted \leq 80, 79, 78, 77, 75, 72, 69, 67, 63, 60, 59, 57, 54, 50, 45, 39, 36, and 34 (respectively, age 8, 9...25). The lower limit of age for subjects with “mild” disease was 15 years, which was useful for a robust prediction of long-term outcome (survival); subjects age 15 to 33 years were enrolled in the mild group if their current FEV₁ % predicted was \geq 97, 92, 90, 87, 86, 84, 82, 77, 70, 68, 67, 62, 58, 54, 52, 50, 45, 32, and 32 (respectively, age 15, 16...33). Subjects age 34 and older were enrolled in the mild group regardless of final FEV₁, due to their survival status. Considerations in choice of entry criteria were also guided by recent modeling of the relationship between FEV₁ and survival in $\Delta F508$ subjects (1). Only one subject was enrolled from any extended cystic fibrosis family, based on the most extreme pulmonary phenotype. Exclusion criteria included: pulmonary function data after lung transplantation or lobectomy, and the presence of *Burkholderia cepacia* in sputum cultures; this criterion was not applied for “mild” subjects older than 34 years. Once

subjects were enrolled according to these initial eligibility criteria, all available pulmonary function data in the previous five years were collected, regardless of the subject's status with respect to acute illnesses. In analyses of pulmonary function, we used the pre-bronchodilator value, if available, if not we used the post-bronchodilator values (of ~18,000 PFTs obtained on the final 808 subjects included in the analyses, ~20% were post-bronchodilator).

For the replication (second) study, the sample consisted of 498 CF subjects whose sputum cultures were negative for *Burkholderia cepacia*, who had two pancreatic insufficient mutations in CFTR (n=467) or an abnormal sweat Cl⁻ and pancreatic exocrine insufficiency (n=31). These subjects were from the University of North Carolina (n=84), CWRU (n=120), Toronto (n=153), and those enrolled from 34 other sites into the ongoing Gene Modifier Study (n=141). The replication study population included a broad range of spirometry values, i.e., most of the subjects were not selected just from the extremes (“mild” and “severe”) of phenotype. As in the initial study, a mixed linear regression model of FEV₁ (% pred.) vs. age, with random subject-specific intercept and slope, was fit to data from each study site, and was used to obtain empirical Bayes estimates of FEV₁ (% pred.) at age 20 for each subject (see Supplement, Statistical Analysis). As in the initial study, replication subjects had to be at least eight years of age; replication subjects with FEV₁ (estimated at age 20) of >70% had to be at least 11 years of age (see Supplement, Statistical Analyses). Based on the genotype (*TGFβ1* codon 10) and phenotype analyses of the initial study (see below for details and methodology), the replication population excluded: 1) subjects with estimated FEV₁ (at age 20) of < 20% pred., a subset not associated with *TGFβ1* genotype; and 2) subjects

without spirometric data by age 35 yrs, which is necessary to get quantitative estimates of FEV₁ at age 20 yrs. Exclusion criteria included any “pancreatic sufficient” mutations, and allergic bronchopulmonary aspergillosis. These 498 subjects had many different genotypes in *CFTR*, including 353 ΔF508 homozygotes; 114 with 2 defined pancreatic insufficient mutations (including 16 subjects with ΔF508/N1303K, 13 subjects with ΔF508/G542X, 12 subjects with ΔF508/G551D, plus 70 subjects with 33 other combinations of two severe alleles); and 31 subjects with either one (n=26) or two (n=5) alleles without an identified *CFTR* mutation.

Data collection: For the initial study, only initials and DOB were provided on source documents (identifying information was blacked out). Clinical data that were provided on standard case report forms included: selection of race/ethnicity by enrolled subjects; presence of meconium ileus at birth; diagnosis of diabetes mellitus, and use of oral hypoglycemic agents or insulin; and diagnosis of asthma, using guidelines as modified from the ATS (airflow obstruction reversed by bronchodilators; clinical evidence of atopy; lab evidence of allergic disease, such as eosinophilia or elevated IgE). To reduce errors, all data, including >18,000 measures of FEV₁, were double-entered into an Oracle-based relational database. For the replication study, we used FEV₁ data available at UNC, CWRU, and Toronto, or as provided in source documents for the modifier study. There were 11,065 measures of FEV₁ in the replication study.

Validation of subjects for genetic studies: For the initial study, 32 of 840 subjects were excluded because they had inadequate spirograms (n=2), were not ΔF508 homozygotes on repeat genotyping (n=8), or because they did not achieve > 90% probability of being congruent with other subjects in the “severe” or “mild” category upon rigorous evaluation

(n=22). These 22 subjects were excluded using mixed model methodology. Specifically, for subjects in the younger age groups (age < 34), a mixed linear model was used to fit a linear regression of FEV₁ (% pred.) on age, using data from mild and severe subjects combined, with random subject-specific intercepts and slopes, and the model was used to obtain empirical Bayes estimates of the intercept and slope for each individual subject. These were used to obtain a predicted value for each subject's "final" FEV₁ (i.e., at the time of enrollment). Because several subjects appeared to be misclassified based on their final FEV₁ (% pred.), we screened outliers by fitting a logistic regression of the probability of being classified as severe, based on the subject's current age and final FEV₁ (% pred.). We excluded "mild" subjects, for whom the predicted probability of being severe was $\geq 10\%$, and excluded "severe" subjects, for whom the predicted probability of being severe was < 90%. This analysis was done twice – using all pulmonary function tests and using the best yearly FEV₁, and subjects who were nonconcordant (as defined above) by either analysis were excluded. This analysis was based on 643 subjects age < 34, and resulted in exclusion of 22 subjects. For the replication study, we used available CFTR genotypes, i.e., we did not repeat CFTR genotyping.

Genotyping: For the initial study, genomic DNA was extracted from peripheral blood leukocytes, using standard protocols (2), or from lymphoblastoid cell lines by using ‘Epicentre MasterPure Complete DNA and RNA purification kit’ (Epicentre, Madison, WI). After PCR, the $\Delta F508$ genotypes were verified by a two color, allelic discrimination assay on an MJ Research Opticon 2. Oligos for *CFTR*, *ADRB2*, and *TNF α* genotyping, using PCR, were purchased from IDT (Coralville, IA). Genotyping of genetic variants (Table 2) was performed by sequencing, or by Illumina BeadArrayTM technology (San Diego, CA), or by published methods. For the alleles that were sequenced (*α LAP-Z*; *MBL2* B, C, and D “null” alleles), PCR products were purified either using ‘QIAquick PCR purification kit’ (Qiagene, Valencia, CA) or ExoSAP-IT

Primers and sequences used for genotyping in the initial study			
Gene	Direction	Primer sequence 5'->3'	Used for
<i>αLAT</i> (Z allele)	Forward	CGATGCTCTCCCTGTTCTGA	PCR
	Reverse	GAGGGGAGACTTGGTATTTTGTTC	PCR and sequence
Control <i>αLAT</i>	Forward	CCCACCTTCCCCTCTCTCCAGGCAAATGGG	Multiplex PCR
	Reverse	GGGCCTCAGTCCCAACATGGCTAAGAGGTG	Multiplex PCR
<i>ACE</i>	Forward	CTGCAGACCACTCCCATCCTTTCT	PCR
	Reverse	GATGTGGCCATCACATTCGTCAGAT	PCR
<i>ADRB2</i>	Forward	GCGGCTTCTTCAGAGCAC	PCR for A46G
	Reverse	CCACCCACACCTCGTCCC	PCR for C79G
<i>ADRB2</i>		AAAAAAAAACCTTCTTGCTGGCACCCAAT	Sequence for A46G
		AAAAAAAAAAAAAAAAACGGACCACGACGTCACGCA	Sequence for C79G
<i>CFTR</i>	Forward	GGAGAGTACCTGAAAGAGGA	PCR
	Reverse	CATTGACAGTAGCTTACCCA	PCR
		Fam-AAGAAAATATCATCTTTGGTGTTTCCTAT- BHQ1	Genotype Wildtype
		Hex-ATCATAGGAAACACCATGATATTTTCTT-BHQ2	Genotype $\Delta F508$
<i>GSTM1</i>	Forward	CATGATCTGCTACAATCCAGAA	PCR
	Reverse	TTCTGAACACAACTTTACCATAC	PCR
<i>MBL2</i>	Forward	TTTCATTCCCTAAGCTAACAG	PCR
	Reverse	TCTGGAAGGTAAAGAATTGCAG	PCR
	Reverse	AAACATTCCTTGTGACACTGC	Sequence
	Forward	GCAGTGTACACAAGGAATGTTT	Sequence
<i>TNFα</i>	Forward	GCTTGTCCCTGCTACCCGC	PCR
	Reverse	GTCAGGGGATGTGGCGTCT	PCR
<i>TNFα</i>	Antisense	TGGAGGCTGAACCCCGTCC	Sequence

(USB, Cleveland, OH) as per manufacturer’s instructions. The purified PCR products

were sequenced using ‘Big Dye Terminator Cycle Sequencing Kit’ and run on ABI

PRISM 3100 or 310 according to the manufacturer's protocol (Applied Biosystem, Foster City, CA). *MBL2* structural ("null") variants (B, C, D) were combined to construct the O/O (null/null) genotype. The low-expression promoter variant (X) was combined with the normal structural (A) sequence to construct the XA allele (Table 2), since XA produces low levels of MBL2 protein (3). The assays performed by Illumina (*α1AP-S* and G1237A alleles; *GSTP1*; *IL-10*; *MBL2* promoter variant, Y and X, which are the normal and low expression variants, respectively, *NOS3*; *TGFβ1*) used an established BeadArray™ (Illumina) technology (4;5). Flanking SNPs around TGFβ1 were assayed using the ABI Prism 7900 HT System (Applied Biosystems, Foster City, CA). Other genetic variants (*ACE*; *ADRB2*; *GSTMI*; *TNFα*; and *CFTR* ΔF508) were tested using standard methods (6-8). For *ACE*, genotyping of the insertion (I) or deletion (D) polymorphism of the Alu repeat sequence in intron 16 of the *ACE* gene was performed by the method previously reported (6). Briefly, PCR products were run on 3% Agarose gel (NuSieve, Cambrex Bio Science Rockland Inc., Rockland, ME) and DNA fragments of 490 bp (insertion) and 190 bp (deletion) were visualized by staining with ethidium bromide. For *ADRB2*, the sequences containing the polymorphic bases were amplified by PCR (see "Primers and sequences used for genotyping"). The *ADRB2* genotype at codons 16 and 27 were determined from the PCR product by single base extension, using SNaPSHOT (Applied Biosystems, Inc., Foster City, CA). To detect the deletion polymorphism of *GSTMI*, PCR was performed with the modification from the published method (7), using exon 3 of *α1AT* gene as an internal control. Briefly, PCR products were also run on the 1 % Agarose gel (NuSieve, Cambrex Bio Science Rockland Inc., Rockland, ME) and the fragments visualized by staining with ethidium bromide. For

TNF α , the -308 allele was amplified using primers and PCR conditions as described (8) and genotyped using SNaPSHOT.

For *aIAP*, *ACE*, *ADRB2*, *GSTMI*, *MBL2*, and *TNF α* , PCR was performed in 25 μ l reaction volume containing 100 ng genomic DNA, 0.4 μ M each forward and reverse primer, 1X buffer (1mM (NH₄)₂SO₄, 67mM Tris-HCl (pH8.8), 0.01% Tween-20), 1-3 mM MgCl₂, 100-200 μ M dNTP mix and 0.75-1.25 units of AmpliTaq DNA polymerase (Perkin Elmer, Foster City, CA). PCR amplification was carried out using the GeneAmp 9700 thermocycler (Applied Biosystem, Foster City, CA). PCR cycling conditions comprised of initial denaturation at 94°C for 5 minutes followed by 30-35 amplification cycles (94°C for 30-60 sec, 55-60°C for 30 sec, 72°C for 30-45 sec) and concluded with 6-10 minutes extension at 72°C.

For the region flanking *TGF β 1*, we tested 31 SNPs in ~730 (median number of tested) subjects (~244 severe subjects and ~487 mild subjects). These SNPs were genotyped as described by the manufacturer (ABI Assays on Demand), and the corresponding reference sequence numbers (rs#) for the assays are given in Figure S1. Because one ABI genotyping kit could be used to genotype ~735 subjects, we elected to assay most of the severe subjects, while reducing the number of mild subjects. Of the 31 SNPs flanking *TGF β 1* (see Figure S1), 30 SNPs had a minor allele frequency of at least 11.5% and a majority (16) had a frequency of > 29%.

For the replication study, genotyping for the TGF β 1 codon 10 polymorphism was performed from DNA extracted from peripheral blood leukocytes or lymphoblastoid cell lines (UNC and Toronto) or buccal DNA (CWRU). Genotyping for codon 10 was performed by sequencing from PCR-amplified product (UNC and CWRU) or allele-

specific oligomerization (ASO; Toronto). For sequencing, PCR was performed in 25 μ L reaction volume containing 100 ng genomic DNA, 0.4 μ M each forward and reverse primer, 1X buffer (1mM (NH₄)₂SO₄, 67mM Tris-HCl (pH8.8), 0.01% Tween-20), 2.0mM MgCl₂, 100 μ M dNTP mix, 5% DMSO and 1.25U AmpliTaq DNA Polymerase (Perkin Elmer, Foster City, CA). PCR conditions comprised of initial denaturation at 94°C for 5 minutes followed by 35 amplification cycles (94°C for 30 sec, 58°C for 30 sec, 72°C for 45 sec) and concluded with 5 minutes extension at 72°C. Sequences were obtained using either the M13(-21) forward primer or F2 oligo when preparing the sequencing mixtures.

Primers and sequences used for codon 10 genotyping by sequencing in the replication study
Forward oligo: 5'-TGTAACGACGGCCAGTGGGATACTGAGACACCCCG-3' (with M13(-21)F tag)
Reverse oligo: 5'-CGGGTGACCTCCTTGGCGTAG-3'
Additional sequencing oligo (when M13(-21)F failed): Forward oligo: 5'-GGATACTGAGACACCCCG-3' (F2)

For genotyping Toronto samples by ASO, we used primers and ASO probes (see below), using a published protocol (9).

Primers and oligonucleotides used for ASO genotyping in the replication study
Forward oligo: 5'-ATTCAAGACCACCCACCTTCTG-3'
Reverse oligo: 5'-CACCAGCTCCATGTCGATAGTC-3'
Oligo for ASO, codon 10 T: 5'-CTGCTGCTGCTGCTGC-3'
Oligo for ASO, codon 10 C: 5'-CTGCTGCCGCTGCTGC-3'

Statistical Analysis: We explored data from the initial study to assist in the design and establish the analytic strategy for the replication (second) study. Specifically, a mixed model regression of FEV₁ was refit to include all subjects in the initial study, which determined empirical Bayes estimates of individual subject intercepts and slopes, as well as predicted FEV₁ (% pred.) at specific ages for each subject. Empirical analysis of the

initial dataset showed that, when considering individual subject predictions of FEV₁ (% pred.) at various ages, as well as their individual slopes, the FEV₁ (% pred.) at age 20 was the best age to distinguish between mild and severe groups, a result supported by previous modeling results (1). Therefore, estimated FEV₁ (% pred.) at age 20 was used as a phenotypic marker representing severity of lung function in the analyses. The estimate of FEV₁ (% pred.) at age 20 allowed us to rank-order subjects from the initial study on a common scale, and can be used regardless of the actual age of the subject. We derived three key insights from analysis of these data and the *TGFβ1* CC genotypes. First, subjects in the initial study with an FEV₁ < 20 % pred. (estimated for age 20) were quite young (usually < 10 yrs), and did not have an increased prevalence of the *TGFβ1* codon 10 CC genotype, which suggested the possibility of non-*TGFβ1* associated environmental (or other genetic) influences on this young group of subjects; therefore, we omitted such subjects from the replication study. Second, we noted a systematic reduction in FEV₁ (estimated for age 20) for those subjects without pulmonary function data before age 36, which likely reflects a “flattening” of FEV₁ decline in those survivors, and limits the ability to quantitate (estimate) FEV₁ at age 20 for these older subjects; therefore, we omitted such subjects from the replication study. Third, data from the initial study suggested the use of a dichotomized phenotype (see below) and a recessive risk model (Table 3) would provide the greatest power. Thus, we used the initial study to identify a dichotomized FEV₁ status (low vs. high) for the replication study. To identify the optimal FEV₁ for dichotomization, we eliminated those subjects in the initial study with FEV₁ < 20 % pred. (estimated at age 20 years), and those with no FEV₁ values before age 36 from the 808 subjects in the initial study, and calculated the optimal

(greatest odds ratio) breakpoint for dichotomized FEV₁ status in comparison with the prevalence of the *TGFβ1* codon 10 CC genotype in the remaining 662 subjects. For these 662 subjects in the initial study, the breakpoint was established to be 67.7% for the FEV₁ (odds ratio 3.12, p<0.0001); therefore, for the primary analysis of the replication study, we used the FEV₁ value (estimated for age 20) of 68% to divide the sample into low/high FEV₁ groups, and tested for association with the CC genotype. For the secondary analysis, we compared the quantitative measure of lung function (FEV₁ at a common age, 20 years) by codon 10 genotypes, and used the Wilcoxon test to analyze for differences in the CC genotype as compared to the other (TC/TT) genotypes. The p value based on the normal approximation is reported (Table 4), and was confirmed via 10,000 permutations of FEV₁ values to create an empirical null distribution for the Wilcoxon statistic.

Chi-square tests of departure from Hardy-Weinberg equilibrium (HWE) (10) were performed with the overall initial data on the 10 candidate genes, as well the SNPs in the *TGFβ1* region, and subjected to permutation multiple-comparison correction. In addition, subgroup HWE tests for the initial and replication studies were performed by severity status and (for the replication study) ΔF508 homozygote status and clinical recruitment site. Bonferroni corrections were used to create overall tests in these stratified HWE analyses. Risk models for the *TGFβ1* codon 10 genotypes can be expressed as $P(\text{severe}|\text{CT}) = \psi_1 P(\text{severe}|\text{TT})$, $P(\text{severe}|\text{CC}) = \psi_2 P(\text{severe}|\text{TT})$, and the recessive model specifies $\psi_1 = 1$, $\psi_2 > 1$ [notation from (11)]. Here, the term “severe” encompasses both the severe/mild phenotyping in the initial study and the dichotomization approach (FEV₁ < 68% pred.) used in the replication study. Standard conditional probability calculations (11) show that this model predicts a deficit of

heterozygotes among severe patients (compared to that predicted by HWE from the allele frequency in “severes”), and an excess of heterozygotes among mild patients. For testing HWE within subgroups, the component subgroup statistics are not independent, and an overall stratified p value was obtained using a Bonferroni correction for the number of subgroups examined.

Results (Supplement)

Initial study: Characteristics of subjects (Table 1).

The prevalence of diabetes mellitus increases with age in cystic fibrosis, and the prevalence in our subjects is similar to that reported in other populations (12). When we adjusted for age, we did not see any difference in the prevalence of diabetes mellitus in the severe group, as compared to the mild subjects.

The mild cohort was divided into two age groups, by the pre-determined upper age limit of 28 years for the “younger” mild group, in order to ensure relative comparability with the severe group with respect to age and (estimated) size of groups. In subsequent analyses (after enrollment was completed), mild subjects above and below 28 years were validated as being a congruent group by a “concordance” analysis to identify outliers (see details above, “Validation of subjects for genetic tests”).

Initial study: Genotype and allelic associations.

In the main analysis (Table 2), strong associations with phenotype were seen only for *TGFβ1* variants. Similar association evidence was seen when only the Caucasian subset was tested, and a similar distribution of the -509 and codon 10 genotypes was seen when only subjects with asthma were tested (data not shown). The prevalence of the genotypes and “minor” alleles for all genetic variants tested (Table 2) was similar to that previously reported in Caucasians. In subgroup analyses, the severe group was compared to either the younger or older mild groups (see Table S1). In these analyses, two p values for variants of *α1AP-Z* achieved nominal significance with $p < 0.05$, and two p values for *GSTM1* were suggestive with $p < 0.10$, but none was significant after multiple testing corrections. There was a tendency for the *α1AP-Z* allele (hetero- or homozygous) to be

more common in the severe cohort (prevalence 4.0%), when compared to older mild subjects (prevalence 0.4%), but less common compared to the younger mild subjects (prevalence 4.6%). There was a tendency for *GSTM1* “null” homozygotes to be more prevalent in the younger cohort of subjects with the mild phenotype (particularly males), which contrasts to a previous report in which “null” homozygosity was associated with clinical markers of more severe disease (13).

Because several recent abstracts have associated *MBL2* variants (O/O and/or XA/O) with worse lung function or survival in cystic fibrosis, we paid particular attention to our analyses of *MBL2* variants (14-17). In addition to the data shown in Table 2 for the O/O and XA/O genotypes, we saw no association of *MBL2* with severe disease by other analyses, including various combinations of several genotypes (O/O, A/O, XA/O, XA/XA), and analysis of the subsets of only those subjects who had *Ps. aeruginosa* or who were Caucasian. We also studied DNA from an additional 56 cystic fibrosis subjects with pancreatic insufficient mutations, who had died (n=9), or who had received a lung transplant (surrogate for death; n=47), and saw no association of the *MBL2* O/O genotype with survival. Specifically, only 3 of the 56 subjects were O/O (5.3%), and their mean (and median) age at death or transplant was 32 years, as compared to 28.7 years (mean age) for the other 53 subjects.

Initial study: Association of SNPs flanking *TGFβ1*.

Figure 1 displays the p values for association of SNP genotypes with the severity of disease (-log₁₀ scale) versus genomic position. The flanking SNPs were tested on ~730 subjects (~244 severe and ~487 mild subjects), as compared to 806 subjects for the 3 *TGFβ1* polymorphisms (-509 and codons 10 and 25). The p values of the 3 *TGFβ1*

polymorphisms were changed only slightly when plotted for the same 730 subjects who were tested for flanking SNPs.

Replication study: Characteristics of subjects

The slight increase in prevalence of the $\Delta F508$ homozygotes in the replication population reflects the contribution of the $\Delta F508$ homozygotes from the multicenter gene modifier study.

Hardy-Weinberg equilibrium testing: initial and replication studies

For the initial study (as reported in the manuscript; see Methods: genotyping), none of the SNPs showed significant overall departure from HWE, after correction for multiple testing; in addition, no SNP attained statistical significance at the 0.05 level in HWE testing when we tested the 662 subjects (defined by entry criteria of the replication study), and the minimum p value among the *TGF β 1* SNPs was p=0.08 for the -509 polymorphism. When patients in the initial study were tested by subgroup (“mild” and “severe”), an excess of *TGF β 1* codon 10 and -509 heterozygotes was observed among mild patients ($\chi_1^2 = 6.65$, p=0.009; 237 codon 10 TC heterozygotes vs. 211.8 expected, and $\chi_1^2 = 5.57$, p=0.02; 211 -509 CT heterozygotes vs. 190.3 expected). Non-significant deficits in heterozygotes were observed for each of these polymorphisms in severe patients. Similar and somewhat stronger results (data not shown) were observed when applying the FEV₁ < 68% pred. dichotomization to the initial study, and are consistent with the recessive model for genotype risk.

For the replication study, an overall deficit of heterozygotes was observed for the *TGF β 1* codon 10 polymorphism ($\chi_1^2 = 7.37$, p=0.007; 206 TC observed vs. 234.5 expected). The deficit could be attributed almost entirely to the more severe group of

patients. In the milder group, with $FEV_1 > 68\%$ pred., the numbers of observed vs. expected heterozygotes were nearly identical (110 vs. 111.3), while the more severe group, with $FEV_1 < 68$ pred., had a significant deficit of heterozygotes ($\chi_1^2 = 10.27$, $p=0.001$, Bonferroni-corrected $p=0.003$, 96 TC heterozygotes vs. 120.7 expected). The results were also broadly consistent with the recessive model for genotype risk, although not every subgroup analysis for initial and replication studies was significant. None of the subgroups in stratified analyses of HWE in the replication study by site (CWRU, Univ. North Carolina, Toronto) was significant at the 0.05 level. The subgroup analysis of the 353 $\Delta F508$ homozygotes in the replication study showed a deficit of codon 10 heterozygotes ($\chi_1^2 = 4.44$, $p=0.04$; 148 codon 10 TC heterozygotes vs. 166.6 expected) that reflected almost exactly the genotype distribution for all replication study patients, yielding no evidence that $\Delta F508$ status is a meaningful determinant of HWE.

Reference List (Supplement)

- (1) Schluchter MD, Konstan MW, Davis PB. Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients. *Stat Med* 2002; 21(9):1271-1287.
- (2) Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2 ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989.
- (3) Crosdale DJ, Ollier WE, Thomson W, Dyer PA, Jensenius J, Johnson RW et al. Mannose binding lectin (MBL) genotype distributions with relation to serum levels in UK Caucasoids. *Eur J Immunogenet* 2000; 27(3):111-117.
- (4) Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques* 2002; 32:S56-S61.
- (5) Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL et al. Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 48). Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2003: 69-78.
- (6) Ashavaid TF, Shalia KK, Nair KG, Dalal JJ. ACE and AT1R gene polymorphisms and hypertension in Indian population. *J Clin Lab Anal* 2000; 14(5):230-237.

- (7) Yim JJ, Park GY, Lee CT, Kim YW, Han SK, Shim YS et al. Genetic susceptibility to chronic obstructive pulmonary disease in Koreans: combined analysis of polymorphic genotypes for microsomal epoxide hydrolase and glutathione S-transferase M1 and T1. *Thorax* 2000; 55(2):121-125.
- (8) Wilson AG, di Giovine FS, Blakemore AI, Duff GW. Single base polymorphism in the human tumour necrosis factor alpha (TNF alpha) gene detectable by NcoI restriction of PCR product. *Hum Mol Genet* 1992; 1(5):353.
- (9) Zielenski J, Markiewicz D, Lin S-P, Huang F-Y, Yang-Feng TL, Tsui L-C. Skipping of exon 12 as a consequence of a point mutation (1898+5G to T) in the cystic fibrosis transmembrane conductance regulator gene found in a consanguineous Chinese family. *Clin Genet* 1995; 47:125-132.
- (10) Weir BS. *Genetic Data Analysis II: Methods for discrete population genetic data.* Sunderland, MA: Sinauer Associates, 1996.
- (11) Lee WC. Searching for disease-susceptibility loci by testing for Hardy-Weinberg disequilibrium in a gene bank of affected individuals. *Am J Epidemiol* 2003; 158(5):397-400.
- (12) Lanng S. Glucose intolerance in cystic fibrosis patients. *Paediatr Respir Rev* 2001; 2(3):253-259.
- (13) Baranov VS, Ivaschenko T, Bakay B, Aseev M, Belotserkovskaya R, Baranova H et al. Proportion of the GSTM1 0/0 genotype in some Slavic populations and its

- correlation with cystic fibrosis and some multifactorial diseases. *Hum Genet* 1996; 97(4):516-520.
- (14) Yarden J, Radojkovic D, Vlietinck R, Vankeerberghen A, Nuytten H, de Boeck C et al. Modulating genes in CF and COPD disease. *J Cystic Fibrosis* 2004; 3 Suppl. 1:S5. (Abstr.)
- (15) Buranawuti K, Cheng S, Merlo C, Zeitlin P, Rosenstein B, Mogayzel P et al. Variants in the mannose binding lectin gene modify survival of cystic fibrosis patients. *Pediatr Pulmonol* 2003; Suppl. 25:215. (Abstr.)
- (16) Davies JC, Turner MW, Klein N, the London MBL CF Study Group. Impaired pulmonary status in cystic fibrosis adults with two mutated *MBL-2* alleles. *Eur Respir J* 2004; 24:298-804.
- (17) Carlsson M, Sjoeholm AG, Eriksson L, Thiel S, Jensenius JC, Segelmark M et al. Deficiency of the mannan-binding lectin pathway of complement and poor outcome in cystic fibrosis: bacterial colonization may be decisive for a relationship. *Clin Exp Immunol* 2004; 139:306-313.

Figure Legends

Figure S1. Genomic position and identifying code number (rs# or Celera ID#, hCV) of the *TGF β 1* SNPs and the 31 flanking SNPs (UNC SNP# in parentheses; most of these SNPs had a minor allele frequency in Caucasians of > 29%). Pairwise linkage disequilibrium (LD) is shown as haplotype blocks (using default parameters of Haploview), as indicated by red bars below SNP positions, and as shown by GOLD plot of the initial study subjects, showing level of LD for all marker pairs. GOLD plot scale bar indicates level of LD (coefficient D').

Table S1. Initial study: Association of polymorphic genotypes in a subgroup analysis of subjects with severe lung phenotype (n=263), as compared with young subjects with mild lung phenotype (15-28 yrs; n=299), and as compared with older subjects with mild lung phenotype (≥ 29 yrs; n=246).

Gene	Genetic variant (change in base/amino acid)	SNP rs #	P value*					
			Young subjects (15-28 yrs) with mild lung phenotype (n=299)			Older subjects (≥ 29 yrs) with mild lung phenotype (n=246)		
			All	males (n=152)	females (n=147)	All	males (n=151)	females (n=95)
<i>$\alpha 1AP$</i>	S allele	17580	1.00	0.82	0.91	1.00	0.84	0.62
<i>$\alpha 1AP$</i>	Z allele	none	0.68	0.39	1.00	0.01	0.18	0.04
<i>$\alpha 1AP$</i>	G1237A	11568814	0.96	1.00	0.80	0.71	0.15	0.51
<i>ACE</i>	D or I	N.A.	0.17	0.39	0.43	0.53	0.54	0.68
<i>ADRB2</i>	A46G	1042713	0.62	0.97	0.25	0.20	0.15	0.60
<i>ADRB2</i>	C79G	1042714	0.26	0.30	0.48	0.52	0.61	0.82
<i>GSTM1</i>	14.3 kb del.	N.A.	0.07	0.09	0.47	0.53	0.72	0.14

<i>GSTP1</i>	A1375G	947894	0.89	0.28	0.61	0.40	0.29	0.79
<i>IL-10</i>	A-1082G	1800896	0.98	0.40	0.34	0.93	0.74	0.95
<i>MBL2</i>	O/O, “null”†	N.A.	0.68	1.00	0.68	0.68	0.71	1.00
<i>MBL2</i>	O/O or XA/O‡	N.A.	0.72	1.00	0.73	0.70	0.73	1.00
<i>NOS3</i>	T5220G	1799983	0.53	0.27	0.86	0.47	0.29	0.93
<i>TGFβ1</i>	C-509T	1800469	0.001	0.06	0.15	0.07	0.23	0.22
<i>TGFβ1</i>	codon 10	1982073	0.0004	0.05	0.008	0.05	0.48	0.009
<i>TGFβ1</i>	codon 25	1800471	0.06	0.29	0.08	0.20	0.62	0.04
<i>TNFα</i>	G-308A	1800629	0.60	0.36	0.78	0.51	0.89	0.74

*Fisher’s exact test.

†P values are based on comparisons of O/O (null) genotype vs. remaining genotypes.

‡P values are based on comparison of O/O and XA/O genotype vs. remaining genotypes.

Table S2. Initial study: Multivariate logistic regression genotype analysis of *TGFβ1* variants associated with severe lung disease in subjects with cystic fibrosis (severe, n=260 vs. mild, n=544), using covariates diabetes mellitus, meconium ileus, *Pseudomonas aeruginosa*, and asthma.

Genetic variants [rs#]	Recessive Model		Codominant Model		Dominant Model	
	P value	Odds Ratio (95% CI)*	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
All subjects (n=804)						
-509 (T) [1800469]	0.003	2.21 (1.31-3.74)†	0.01	1.84 (1.15-2.95)‡	0.12	1.28 (0.94-1.73)§
codon 10 (C) [1982073]	0.0002	2.24 (1.47-3.42)	0.0008	2.15 (1.37-3.36)¶	0.05	1.38 (1.01-1.89)**
Male (n=430)						
-509 (T) [1800469]	0.03	2.28 (1.11-4.69)	0.05	1.93 (0.99-3.74)	0.22	1.30 (0.85-1.99)
codon 10 (C) [1982073]	0.09	1.60 (0.92-2.99)	0.05	1.85 (0.99-3.48)	0.14	1.40 (0.90-2.19)
Female (n=374)						
-509 (T) [1800469]	0.07	2.06 (0.94-4.48)	0.10	1.78 (0.90-3.54)	0.27	1.28 (0.83-2.01)
codon 10 (C) [1982073]	0.0004	3.12 (1.65-5.87)	0.004	2.57 (1.34-4.91)	0.14	1.41 (0.90-2.23)

*CI denotes confidence interval

For each listed genetic variant, the higher-risk allele is indicated in parentheses, and odds ratios are presented for the highest-risk genotype vs. the lowest-risk genotype. Specific genotype comparisons are listed below.

†The value is for the comparison of the T/T genotype with the combined C/T and C/C genotypes.

‡The value is for the comparison of the T/T genotype with the C/C genotype.

§The value is for the comparison of the combined C/T and TT genotypes with the C/C genotype.

|| The value is for the comparison of the C/C genotype with the combined C/T and T/T genotypes.

¶The value is for the comparison of the C/C genotypes with the T/T genotype.

**The value is for the comparison of the combined C/C and C/T genotypes with the T/T genotype.

Figure S1

