Published in final edited form as:

Nat Genet. 2011 June ; 43(6): 539–546. doi:10.1038/ng.838.

Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2

Fred A. Wright[1], **Lisa J. Strug**[2,3], **Vishal K. Doshi**[4], **Clayton W. Commander**[5], **Scott M. Blackman**[6], **Lei Sun**[3], **Yves Berthiaume**[7], **David Cutler**[8], **Andreea Cojocaru**[2], **J. Michael Collaco**[6], **Mary Corey**[2], **Ruslan Dorfman**[9], **Katrina Goddard**[10], **Deanna Green**[6], **Jack W. Kent Jr**[11], **Ethan M. Lange**[1,12], **Seunggeun Lee**[1], **Weili Li**[2], **Jingchun Luo**[13], **Gregory M. Mayhew**[1], **Kathleen M. Naughton**[4], **Rhonda G. Pace**[5], **Peter Paré**[14], **Johanna M. Rommens**[9,15], **Andrew Sandford**[14], **Jaclyn R. Stonebraker**[5], **Wei Sun**[1,12], **Chelsea Taylor**[2], **Lori L. Vanscoy**[16], **Fei Zou**[1], **John Blangero**[11], **Julian Zielenski**[9], **Wanda K. O'Neal**[5], **Mitchell L. Drumm**[17], **Peter R. Durie**[18,19], **Michael R. Knowles**[5,20], and **Garry R. Cutting**[4,6,20]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2]Child Health Evaluative Sciences, the Hospital for Sick Children, Toronto, Ontario, Canada

[3]Dalla Lana School of Public Health, the University of Toronto, Toronto, Ontario, Canada

[4]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[5]Cystic Fibrosis-Pulmonary Research and Treatment Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[6]Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[7]Department of Medicine, Centre de recherche Centre Hospitalier de l'Université de Montréal, Montréal, Quebec, Canada

[8]Department of Human Genetics, Emory University, Atlanta, GA, USA

[9]Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, Ontario, Canada

[20]Addresses for Correspondence: Michael Knowles, M.D. Cystic Fibrosis-Pulmonary Research and Treatment Center, University of North Carolina, Chapel Hill, North Carolina, USA, Knowles@med.unc.edu Telephone: (919) 966-6780 Fax: (919) 966-7524, Garry R. Cutting, MD, The Johns Hopkins Medical Institutions, 733 North Broadway, BRB 559, Baltimore, MD 21205. gcutting@jhmi.edu, Telephone: (410) 614-0211, Fax: (410) 614-0213.

Author Contributions F.A.W., L.J.S., L.S., D.C, M.C., R.D., L.L.V., W.K.O., M.L.D., P.R.D., A.H., RM., M.R.K., and G.R.C., worked on study design. Y.B., A.C., J.M.C., M.C., R.D., D.G., W.L., K.M.N., R.G.P., P.P., J.M.R., A.S., J.R.S., C.T., L.L.V., J.Z., M.L.D., P.R.D., M.R.K., and G.R.C. performed sample collection and phenotyping. F.A.W., C.W.C., S.M.B., D.C., K.G., E.M.L., J.L., R.G.P., J.R.S., F.Z., W.K.O., M.L.D., M.R.K., G.R.C. performed genotyping and data cleaning. F.A.W., L.J.S., V.D., C.W.C., S.M.B., L.S., A.C., J.M.C., M.C., R.D., K.G., J.K., E.M.L., S.L., W.L., G.M.M., J.M.R., W.S., C.T., F.Z., J.B., performed statistical analysis. F.A.W., L.J.S., V.D., L.S., R.D., J.M.R., W.O., M.L.D., P.D., M.R.K. and G.R.C. wrote the manuscript.

The authors have no competing financial interests.

[10]The Center for Health Research, Oregon Clinical Translational Research Institute, Portland, OR, USA

[11]Southwest Foundation for Biomedical Research, San Antonio, Texas, USA

[12]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[13]Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[14]James Hogg Research Centre, Providence Heart + Lung Institute, The University of British Columbia, Vancouver, British Columbia, Canada

[15]Department of Molecular Genetics, the University of Toronto, Toronto, Ontario, Canada

[16]Pediatric Pulmonology, CF Ctr., Naval Medical Center, Portsmouth, VA

[17]Case Western Reserve University Departments of Pediatrics and Genetics, Cleveland OH, USA

[18]Program in Physiology and Experimental Medicine, the Hospital for Sick Children, Toronto, Ontario, Canada

[19]Department of Pediatrics, University of Toronto, Ontario, Canada

Abstract

A combined genome-wide association and linkage study was used to identify loci causing variation in CF lung disease severity. A significant association (P=3. 34 × 10^{-8}) near *EHF* and *APIP* (chr11p13) was identified in *F508del* homozygotes (n=1,978). The association replicated in *F508del* homozygotes (P=0.006) from a separate family-based study (n=557), with P=1.49 × 10^{-9} for the three-study joint meta-analysis. Linkage analysis of 486 sibling pairs from the family-based study identified a significant QTL on chromosome 20q13.2 (LOD=5.03). Our findings provide insight into the causes of variation in lung disease severity in CF and suggest new therapeutic targets for this life-limiting disorder.

Lung disease is the major source of morbidity and mortality in cystic fibrosis (CF), a recessive disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Allelic variation in *CFTR* does not explain the wide variation in severity of lung disease[1] however studies of twins and siblings demonstrate substantial heritability underlying differences in lung function measures in CF patients ($h^2 > 0.5$)[2]. Candidate gene studies have produced conflicting results, with only a few large scale replications accounting for a small proportion of heritable variation in CF lung function[3,4]. Identification of other genetic modifiers could identify potential mechanisms for variation in lung function in CF, as well as for common diseases such as chronic obstructive pulmonary disease (COPD), and suggest new targets for intervention.

Whole-genome methods provide an attractive approach to identify modifier loci of Mendelian disorders. However CF presents numerous challenges, such as: (1) collecting multiple years of lung function measures to accurately classify lung disease severity; (2) selecting the appropriate study design to identify common and rare variants; (3) accruing

sufficient sample sizes, and (4) accounting for potential interaction between *CFTR* and modifier loci. To overcome these challenges, we formed a North American CF Gene Modifier Consortium to identify modifiers of lung disease severity and other phenotypes. For lung disease in CF, the forced expiratory volume in 1 second (FEV_1) is the most clinically useful measure of lung disease severity and is a well-established predictor of survival[5,6]. However, comparison of FEV_1 measures across a broad age range of CF patients is confounded by decline with age and mortality attrition. To account for these confounders, the Consortium developed a quantitative lung disease phenotype based on multiple measures of FEV_1 over 3 years[7] that displays robust genetic influence ($h^2 = 0.51$)[8].

The Consortium is composed of three samples of CF patients recruited using different study designs. The Genetic Modifier Study (GMS) consists of unrelated patients homozygous for the common CF allele *F508del* (HGVS nomenclature: *p.Phe508del*), recruited from extremes of lung function[9]. The Canadian Consortium for Genetic Studies (CGS) enrolled unrelated patients having pancreatic insufficiency from a population-based sample[10]. The CF Twin and Sibling Study (TSS) recruited families where two or more surviving children have CF[2]. The GMS and CGS were designed for association analysis, while the TSS was designed for both linkage and association, providing an opportunity to detect rarer variants or poorly tagged loci.

As many current Genome Wide Association Studies (GWASs) employ sample sizes that are several-fold larger than available for the CF population, we sought to maximize power by (1) testing association using combined data from GMS and CGS, followed by replication using the association evidence from TSS, and (2) testing linkage using the TSS, followed by SNP association testing in linked regions in the unrelated patients in GMS and CGS. We also restricted analysis to patients bearing two severe loss-of-function CFTR alleles and a subset of these patients that had identical *CFTR* genotypes (homozygosity for *F508del*).

RESULTS

Genome-wide association analysis of lung disease severity in CF

A total of 3,467 CF patients are represented in three study designs (Table 1, Supplementary Note). Patients in the GMS and 60% of the patients in the CGS and TSS are *F508del* homozygotes (*F508del/F508del*), while the remainder has other severe exocrine pancreatic *CFTR* genotypes[2,9,10]. The three samples showed consistent distributions of the lung disease phenotype, with the mid-range under-represented in GMS due to the extremes-of-phenotype design (Figure 1). Patients were contemporaneously genotyped using the Illumina 610-Quad array® in a single facility with stringent quality control (Online Methods). Association scans for the GMS and CGS used an additive model adjusted for sex and principal components as described[11]. Results were combined using a directional meta-analysis approach for (1) GMS and CGS, n=2,494 and (2) GMS and CGS *F508del/F508del*, n=1,978 (power analysis shown in Supplementary Figure 1).

The combined GMS and CGS analysis identified seven regions with suggestive association ($P \leq 1/570{,}725 = 1.75 \times 10^{-6}$) (Figure 2 and Table 2). Restricting analysis to *F508del/F508del* patients, the *EHF-APIP* region on 11p13 achieved genome-wide significance at

rs12793173 (P=3.34 × 10^{-8}, explaining 1.0% of the phenotype variation in GMS, 2.2% in CGS *F508del/F508del*). We verified the significance by permutation analysis and by developing an alternative conditional likelihood approach which acknowledged the GMS extremes of phenotype (Online Methods, Supplementary Figure 2). With the inclusion of CF-relevant covariates (sex, BMI and previously associated genes), association for rs12793173 was even stronger (P= 9.42 × 10^{-9} for GMS and CGS *F508del/F508del*; Supplementary Table 1). Two purported modifiers of CF lung disease, *TGFB1* and *IFRD1*, did not achieve genome-wide significance. *TGFB1* did, however, achieve *P*-values in the range of 10^{-3} to 10^{-4} in the GMS sample, depending on additional covariates (Supplementary Table 1).

The SNPs in the significant region and the six suggestive regions in GMS and CGS were evaluated for association in TSS using Merlin[12], while accounting for family structure. To be consistent with the GMS and CGS allelic effect, each replication test was one-sided, with the TSS sample (all or *F508del/F508del* patients) for each suggestive SNP chosen to be consistent with the GMS and CGS sample set providing maximum significance. Covariates for sex and four principal components[11] were included for TSS. The SNP attaining genome-wide significance in GMS and CGS (rs12793173, *F508del/F508del*) demonstrated significant association in the TSS *F508del/F508del* sample (P=0.006; Bonferroni corrected P = 0.041 for the seven replication tests; Table 2). Two of the suggestive SNPs provided modest evidence in TSS: rs9268905 near *HLA-DRA* (P=0.032) and rs1403543 near *AGTR2* (P=0.053), with neither significant after correcting for the seven replication tests.

We next performed a joint analysis, shown to be more powerful than testing followed by replication[13], using a weighted meta-analysis procedure (Online Methods). Using all patients, rs12793173 attained genome-wide significance (P=1.12 × 10^{-8}). For this patient set, rs568529, a SNP in high LD (r^2 > 0.9) with rs12793173, achieved slightly greater significance (P=9.75 × 10^{-9}). As in the earlier analysis, restricting to *F508del/F508del* patients increased the significance of *EHF-APIP* (P=1.49 × 10^{-9} for rs12793173 (Table 2), P=8.28 × 10^{-10} for rs568529). In the HLA class II region, a SNP (rs2395185, ~1kb from the suggestive SNP rs9268905 identified from GMS and CGS) approached genome-wide significance using all patients (P=9.02 × 10^{-8}; Supplementary Figure 3). SNPs in *AGTR2* remained suggestive for all patients (rs5952206, P=1.25 × 10^{-7}) and for *F508del/F508del* patients (rs7060450, P=3.67 × 10^{-7}).

Figure 3 shows the GMS and CGS results for an 800kb interval including *EHF-APIP*. The minimum *P*-value appears in an intergenic region 3' to both *EHF* and *APIP*. A second peak at rs286873 (P=5.62 × 10^{-7}) near *EHF* exhibited low linkage disequilibrium (r^2 < 0.2) with the primary SNP (Figure 3). After conditioning on the primary finding, rs286873 had regional statistical significance (rs12793173; corrected P=0.0029), suggesting additional regional genetic variants (Supplementary Figure 4). We repeated the testing after MACH imputation[14]. The imputed SNPs in the region identified the same *EHF/APIP* interval, with minimum P=1.45 × 10^{-8}, at rs535719, at a position 19kb closer to *APIP* than rs12793173. None of the imputed SNPs produced substantially improved association evidence (Supplementary Figure 5). Neither total copy number nor allele-specific copy number (Online Methods) models met genome-wide significance (illustrative Manhattan plot in

Supplementary Figure 6). Finally, after sequencing the exonic regions of *EHF* and *APIP* in 48 patients with mild pulmonary disease and 48 patients with severe pulmonary disease from the GMS, no additional genetic variation was found that offered insight into putative modifying roles (data not shown).

Linkage of lung disease severity in CF to chromosome 20q13.2

Linkage analysis revealed a genome-wide significant multipoint LOD score of 5.03 at rs4811626, located at 53.81 Mb (~85cM) on chromosome 20q13.2 (nominal $P=7.9 \times 10^{-7}$; genome-wide[15] $P=2.3 \times 10^{-3}$; Figure 4). Another, but more modest linkage signal was on chromosome 1p22.21, with multipoint LOD score of 2.48 for rs941031at 91.07 Mb (119 cM). Inclusion of BMI-Z, an important covariate of CF lung function (Supplementary Table 1), increased the LOD score for the linkage peak on 20q13.2 to 5.72 (genome-wide $P=5.05 \times 10^{-4}$ at rs4811645 which is 0.07cM (0.13Mb) from rs4811626; Figure 5) while linkage on chromosome 1p22.21 decreased to LOD 1.67. Thus, anthropometric measures are not major contributors to the linkage on 20q13.2 but may be playing a role on 1p22.21. We estimated that the QTL at 20q13.2 is approaching 50% of the variation in lung function in the CF sibling pairs (Supplementary Figure 7); however, this estimate is highly likely to be biased upward due to winner's curse[16].

A region of 1.31 Mb on 20q13.2, demarcated by 1 LOD unit below the maximum (when BMI-Z was used as a covariate), was analyzed for association in the combined GMS and CGS samples. A 16kb cluster of SNPs in high LD (rs6092179, rs6024437, rs8125625, rs6024454 and rs6024460; $r^2 > 0.8$) located ~200kb from *CBLN4* generated the lowest *P*-values in the combined GMS and CGS *F508del/F508del* samples (Figure 5). The SNP with the lowest *P*-value (rs6024460; $P=1.34 \times 10^{-4}$) reached regional significance (corrected $P = 0.041$). Association in the TSS identified a SNP (rs6069437) with marginal association (uncorrected $P = 0.014$) that displays weak LD with the GMS and CGS cluster of SNPs. Imputation did not identify any SNPs exhibiting a lower *P* value for association than rs6024460 (Supplementary Figure 8).

A combined false discovery rate approach corroborates genome-wide significance of loci on chromosomes 11 and 20

To evaluate association and linkage in a single framework, linkage information was used to reprioritize genome-wide association using extensions of the false discovery rate (FDR)[17] via the stratified FDR (SFDR)[18] and weighted FDR (WFDR)[19]. We (1) obtained linkage-weighted *q*-values representing the combined evidence at each SNP, and (2) re-ranked GWAS results by linkage-weighted *q*-values (see Online Methods). Results are presented from the WFDR; results were confirmed using the SFDR (data not shown). SNPs with *q*-values less than 0.05 were declared to be genome-wide significant (Table 3). SNPs in the *EHF-APIP* region on chromosome 11 are highly significant (low *q*-values), because of the strong association (Table 3). After accounting for linkage, the *q*-values for SNPs under the linkage peak on chromosome 20 are considerably decreased. The results presented in Table 3 illustrate that the linked SNPs on chromosome 20 are now top ranked genome-wide, while they were ranked 154th or lower, prior to incorporating the linkage information. The top-ranked SNP by the WFDR analysis was rs6092179 at 53.81 Mb on chr 20 (WFDR q-

value=0.015, Table 3). SNP rs6092179 is within an LD block containing 4 other SNPs (rs6024437, rs8125625, rs6024454 and rs6024460), all demonstrating association with CF lung function and *q*-value <0.05. A rank-based *q*-value Manhattan plot demonstrates that chromosome 11 and chromosome 20 both attain genome-wide significance (Supplementary Figure 9).

DISCUSSION

We identified two new loci containing genetic variants contributing to variation in lung function in CF patients. The success of this project reflected: 1) coordinated analysis of three independent samples of the CF population (representing ~15% of all patients in North America) where each study subject was characterized by the same quantitative measure of lung function; 2) simultaneous genotyping of samples using a single platform which allowed for data cleaning using relatedness assessments and removal of poor quality genotypes based on parent to child transmission predictions; 3) analyzing for loci with small effect sizes using association, and loci of major effect (even in the presence of substantial allelic heterogeneity) using linkage. Moreover, we garnered increased power from an extreme of phenotype sample, while a population-based sample allowed for the development of a phenotype with external validity.

The association at chr11p13 is in an intergenic region 3' to *APIP* and *EHF* with regulatory features including: i) significant conservation across species, ii) open chromatin (DNAase hypersensitivity and FAIRE-Seq), and iii) DNAase hypersensitive patterns suggesting cell-type-specificity (http://genome.ucsc.edu). The UCLA Gene Expression Tool (UGET, http://genome.ucla.edu/~jdong/GeneCorr.html)[20,21] indicates correlation of expression of nearby genes, including strong correlation of *EHF* to *ELF5*, both epithelial-specific transcription factors; *APIP* to *PDHX*, which have the same promoter region; and *EHF* to *APIP*. *APIP* (Apaf-1-interacting protein) is known to inhibit apoptosis by binding to APAF-1, an important activator of caspase-9[22, 23] and by APAF-1 independent activation of AKT and ERK1/2[24]. *EHF* is a member of epithelial-specific-Ets transcription factors that share a conserved Ets domain[25-27]. EHF can be induced in bronchial epithelial cells, smooth muscle cells and fibroblasts[28,29], leading to transcriptional repression of a subset of *ETS/AP-1*-responsive genes activated by MAP-kinase pathways[26,28], and in airway it may serve as an important regulator of differentiation under conditions of stress and inflammation[26,27]. Both genes show evidence of robust expression in lung and trachea, with *APIP* showing ubiquitous expression across tissues and *EHF* showing highest expression in trachea (http://www.ncbi.nlm.nih.gov/UniGene and http://www.ncbi.nlm.nig.gov/geo)[30]. Interestingly, cis-eQTL signatures for *APIP* are reported for lymphocytes and monocytes (eqtl.uchicago.edu). Comparing the eQTLs to the direction of phenotype-genotype association suggests that increased expression of *APIP* may be associated with decreased lung function, implying that inhibition of apoptosis worsens CF lung disease. This hypothesis is consistent with the emerging concepts that delayed neutrophil clearance, due to reduced apoptosis in neutrophils in the airways of CF patients, could lead to a hyperinflammatory state and more severe lung disease [31,32] and that inhibition of apoptosis contributes to goblet cell metaplasia, a central feature in CF airway pathophysiology[33].

All 5 genes within the 1 LOD support interval in the chromosome 20 linkage region (Figure 5) are expressed in either fetal or adult lung or in bronchial epithelial cells (http://genome.ucsc.edu/). The 16kb cluster of SNPs associated with lung function in the GMS and CGS samples is located ~200kb to 500 kb centromeric to the five genes. None of the SNPs lies within a segment of open chromatin identified in the 16kb region in Normal Human Bronchial Epithelia cells (http://genome.ucsc.edu). Neither eQTL in lymphocytes (eqtl.uchicago.edu), miRNA (http://www.mirbase.org) nor DNaseI hypersensitive sites in Small Airway Epithelial cells map to the 16kb region. However, this does not exclude the possibility that the associated region regulates expression of any of the five genes or more distant genes. Among the five genes, *MC3R* has been implicated in weight maintenance and regulation of energy balance in animals and humans[34-36]. Variation in resting energy expenditure has been correlated with lung function measurements, lung tissue damage and lung disease exacerbation in CF patients [37,38]. *MC3R* has also been implicated as a modulator of neutrophil accumulation in a murine model of lung inflammation[39], a key feature of CF lung disease, as noted above. Other genes of interest within the linkage peak encode Crk-associated substrate scaffolding (CASS) 4 (*CASS4/HEPL*), a relative of proteins implicated in cell attachment, migration establishing polarity, invasion and phagocytosis of bacterial pathogens[40] and Aurora kinase A (*AURKA*) which been shown to interact with Hef1/NEDD9, a member of the CASS family that mediates cytokinesis in late mitosis and facilitates disassembly of primary cilia[41].

Twin studies in adults demonstrate that FEV_1 is under strong genetic influence[42,43], and at least three loci (*GSTCD*, *TNS1* and *HTR4*) have been reproducibly associated with this measure[44-46]. Multiple replicated loci have also been associated with variation in the FEV_1/FVC ratio[45,46] and at least two of these loci (*HHIP* and *FAM13*) show reproducible association with COPD[44,47,48]. While the lung phenotype used here was based on FEV_1, none of the above loci coincides with the regions identified in this study and neither of the loci identified here occur within the top 2000 associations for FEV_1 or FEV_1/FVC[45,46].

Common variation in the *EHF/APIP* region is estimated to alter the lung function measure in the GMS and CGS *F508del/F508del* patients by ~0.2 units of the quantitative lung disease phenotype per allele (Table 2). Translated into more familiar clinical terms, the 0.2 unit difference is approximately equivalent to a mean difference in FEV1 percent predicted of 5.1 ± 1.9, corresponding to a mean difference in FEV_1 of 254 ± 86mL in patients over 18 years of age (Online Methods). The QTL on chromosome 20 may account for a sizeable fraction of lung function variation in CF. Using simulations described by Blangero and colleagues[16], we estimate that this locus accounts for a maximum of 46% and a minimum of 4% of the variance in the CF siblings (Online Methods).

In summary, our association and linkage approach provided complementary findings with the identification of two significant loci harboring genes of biologic relevance for CF. Of particular note for modifier searches in other monogenic diseases is the potential importance of minimizing variation in the causative gene. When we confined association analysis to patients with identical CFTR genotypes (i.e. *F508del/F508del*), one of the 7 suggestive loci achieved genome-wide significance, despite the reduction in sample size due to the exclusion of 38% of subjects in the CGS sample with other *CFTR* genotypes. The remaining

suggestive loci contain biologically intriguing candidate modifiers that will be evaluated in future studies. Finally, the identification of genetic loci that modify lung function in CF, should provide new insight leading to the development of novel therapies for this devastating condition.

ONLINE METHODS

Genotyping and quality control

DNA from whole blood or transformed lymphocytes was hybridized to the Illumina 610-Quad ® platform at Genome Quebec (McGill University and Genome Quebec Innovation Centre,) using the 96-well plates with CEPH and one replicate control per plate. Illumina BeadStudio® was used to call genotype, and identity confirmed by Sequenom® fingerprinting. SNPs were removed if they were monomorphic, missing > 10% calls or with >1% Mendelian error in TSS trios. Finally, 570,725 autosomal and X-chromosome SNPs were selected, as well as 158 chromosome Y SNPs and 138 mitochondrial SNPs. Duplicate discordance was 0.004% in GMS, and similar for the other studies.

Sample exclusions included: initial call rate below 98%, unexpected close relatives or duplicate enrollments, unresolved sex mismatches, aneuploidy or outlying heterozygosity (> 5 standard deviations from the mean of 31.6%). Overlapping from 542 Illumina GoldenGate ® SNPs in GMS revealed platform discordance of 0.07%. Families with >5% Mendelian errors were excluded. Twenty-eight patient samples were excluded (GMS6; CGS 17, TSS 5) due to genotyping failure or artifacts, two GMS samples excluded due to outlying ancestry (by PC analysis), and eight GMS samples excluded for > second degree relation with other samples. Reported findings were verified using Illumina GenomeStudio V1.0.2® module V1.0.10 and manually-assisted calling.

Association testing

Regressions for the lung phenotype were performed separately for GMS, all CGS, and CGS *F508del/F508del* using an additive model in PLINK v. 1.07 [49], adjusted for sex and genotype principal components (PCs)[11]. Using the PLINK z-statistics for GMS and CGS, the standard meta-analysis z-statistic[50] was $z = w_{GMS}z_{GMS} + w_{CGS}z_{CGS}$, with weights inversely proportional to standard errors, and common reference alleles for directional consistency. “Suggestive” association used the approximate threshold 1/(number of SNPs)=1/570,725=1.75×10^{-6}, and significant association the Bonferroni threshold $P < 0.05/570{,}725 = 8.76 \times 10^{-8}$. For males, X-chromosome genotypes followed PLINK defaults (0 or 1 minor alleles; alternative coding resulted in no qualitative changes).

Permutations of genotypes relative to phenotypes and covariates (1,000) were used to refine the thresholds. From this pool of permutations, 10,000 permuted meta-analyses were computed. The obtained significance thresholds for a genome-wide error 0.05 were $P = 1.07 \times 10^{-7}$ (GMS and CGS) and $P = 1.05 \times 10^{-7}$ (GMS and CGS *F508del/F508del*). Consequently, $P < 5 \times 10^{-8}$ achieves false positive error control at genome-wide $\alpha<0.05$, even correcting for two separate GWAS analyses. Regional multiple-comparisons correction (after highlighting a region) used the Bonferroni correction for the regional SNPs.

TSS association analysis was performed in 973 CF siblings and for the 557-patient *F508del/F508del* subset using the Merlin variance-components additive model framework[12], corrected for linkage, family structure, sex, and 4 PCs. Missing genotypes (0.125%) were inferred to increase power[51]. Joint analyses of GMS, CGS and TSS used the meta-analysis approach described above.

A combined conditional likelihood approach

We devised a novel approach using the assumption that CGS represents a random population sample, whereas GMS was conditional on the observed phenotypes. Letting g be the number of SNP minor alleles, the phenotypes y were pre-adjusted for sex and the study-specific PCs. We assumed an additive model $y = \beta_0 + \beta_1 g + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$. The full likelihood conditioned on GMS sampling was

$$\begin{aligned} L &= p(g_{CGS}, y_{CGS}; \beta_0, \beta_1, \sigma^2)\, p(g_{GMS} | y_{GMS}; \beta_0, \beta_1, \sigma^2) \\ &= p(g_{CGS})\, p(g_{GMS})\, p(y_{CGS} | g_{CGS}; \beta_0, \beta_1, \sigma^2)\, p(y_{GMS} | g_{CGS}; \beta_0, \beta_1, \sigma^2) / p(y_{GMS}; \beta_0, \beta_1, \sigma^2) \end{aligned}$$

where $p(y_{GMS}; \beta_0, \beta_1, \sigma^2) = \sum_{j=0}^{2} p(g_{GMS} = j)\, p(y_{GMS} | g_{GMS} = j; \beta_0, \beta_1, \sigma^2)$. Finally, we computed the SNP-specific statistic 2 × (log-likelihood ratio), with $\beta_1 = 0$ as the null and compared to 2 χ_1^2. The approach assumes the effect sizes are the same in GMS and CGS, which is true under the null.

Power Analyses

Power analyses for the combination of GMS and CGS assumed an additive genetic model, with effect β_1 on the average phenotype for each minor allele. The results for GMS and CGS *F508del/F508del* are in Supplementary Figure 1. For each simulation the weighted meta-analysis *P*-values were compared to 5×10^{-8}.

Genotype imputation

MACH (autosomes, http://www.sph.umich.edu/csg/abecasis/mach/) and IMPUTE (chromosome X, http://mathgen.stats.ox.ac.uk/impute/impute.html) imputation was conducted for 1162 GMS patients, 1,254 self-reported CGS “Caucasian” patients and 60 CEU reference samples from HapMap I/II. Some of these individuals were later used for TSS, and association analyses considered only unique subsets in GMS and CGS, respectively (Table 1). Imputation yielded data for ~2,544,000 autosomal and ~65,000 chromosome X SNPs.

Copy-number analysis

Copy number variants (CNVs) were detected using pennCNV (2008Nov19 version)[52] and genoCNV (version 1.08)[53] using default parameters in 1103 GMS and 1301 CGS samples. CNVs with fewer than 5 probes or showing <1% variation were used, resulting in 3,008/4,868 probes from genoCNV/pennCNV in GMS and 3015/4663 probes for genoCNV /pennCNV in CGS. Genotype PCs were used to control stratification.

Linkage Marker Selection

19,566 SNPs were selected from the Illumina platform with minor allele frequency >0.4 and r^2 <0.01 between adjacent SNPs, using Merlin[54]. HapMap II recombination data were used to integrate genetic and physical map positions. Average inter-marker distance was 0.18 cM, or 0.13 Mbp. Physical positions not appearing in HapMap were estimated assuming uniform recombination between known adjacent SNPs. The average marker information content was ~0.9 (multipoint) and ~0.31 (two-point).

Linkage Analysis

Variance components were estimated in SOLAR (Sequential Oligogenic Linkage Analysis Routines)[55], with similar results from Merlin[54], using multipoint IBD probabilities obtained from Merlin. LOD scores were computed with and without covariates (sex and average BMI Z-score). Multipoint LODs>2.0 was considered suggestive and LOD>3.7 was considered genome-wide significant[15].

WFDR and SFDR methods

Let P_i be the p-value of an association test for SNP i, i =1,…,m. Converting p-values to q-values[56] controls the FDR. SNPs with q-values less than the FDR threshold value (e.g. γ = 0.05) are declared significant. The expected proportion of false positives among all the positives is then controlled at level γ. Note that ranking SNPs by P-value or q-value are equivalent.

Let Z_i be the linkage score of SNP i obtained from a GWL study. For the SFDR method, m SNPs are divided into K disjoint strata based on the prior linkage information[57]. Cconsider K = 2 and assign each SNP i to stratum 1 (the high priority group) or stratum 2 (the low priority group) according to whether the linkage score Z_i exceeds a threshold C (we used C=3.3 corresponding to significant linkage[15]). Q-values are then calculated separately for each stratum of SNPs, achieving FDR control in each stratum (Sun et al., 2006). Ranks of the GWAS SNPs are determined by the q-values with the original association p-values used to break any q-value ties.

WFDR calculates a weighting factor W_i for each SNP i with weights subject to two constraints: $W_i \geq 0$_ and $W = \Sigma_i W_i /m = 1$. The weight W_i is proportional to the linkage signal Z_i for SNP i (e.g.W_i exp($B \cdot Z_i$) / ν, $\nu = \Sigma_i \exp(B \cdot Z_i)/m$,, and B=1) (Roeder et al., 2006), and the FDR procedure is applied to the set of weight-adjusted p-values, P_i/W_i, i =1, …,m. We use B=2 in the present analysis. The WFDR and SFDR were implemented in a perl program called *SFDR*, available at http://www.utstat.toronto.edu/sun/Software/SFDR/index.html.

Phenotype variation attributable to association and to linkage

The proportion of variation due to each SNP was measured as the change in regression sums of squares vs. the smaller model with the SNP removed[58]. Using the genome-scan threshold of P=5 ×10^{-8} and minimum P=3.34 × 10^{-8} in the chromosome 11p13 region for GMS and CGS *F508del/F508del* patients, we estimate a 57.4% reduction in effect size compared to the nominal result. Using the joint analysis based on GMS, CGS *F508del/F508del* and TSS

F508del/F508del patients, the observed minimum $P=8.28 \times 10^{-8}$ results in ~ 28.0% reduction of the effect size. Using the rough parallel to explained variation in the trait, the estimated explained variation for 11p13 remains 1%-2%. For a linkage study of comparable size (n=500 sibling pairs), with a phenotype heritability of 0.5, the bias attributed to the winner's curse varies from approximately 0.46 down to zero as the true (unmeasured) heritability attributable to the QTL increases[16]. While not possible to quantify the magnitude of this bias in this single study, these calculations provide an upper bound on the bias of 0.38 to 0.46 and a lower bound of 0.04 to 0.12.

Estimation of changes in the CF lung phenotype upon FEV1 %predicted and airway flow

Using 973 TSS individuals, a hypothetical quantity of 0.2 was added to each individual's lung phenotype, to correspond to the effect size observed for the significant association of SNPs near *EHF/APIP*. The average raw FEV1 (in liters) was then back-extrapolated[8] and FEV1 percent predicted values were generated using the predictive equations[59,60]. Height and age adjustments used to calculate the original quantitative lung phenotype were preserved. The average increase (mean ± SD) in FEV_1 percent predicted corresponding to a 0.2-unit increase of our lung phenotype was 5.09% ± 1.90% [n = 841; Range: 0.00 – 14.53%]. The corresponding average increase in raw FEV1 was 253.5 ± 85.9mL in adult subjects (>18 years) [n = 244; Range: 0.0 – 630.0mL].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the US National Institutes of Health grants: HL68927, K23DK083551, R01HL068890, R01DK066368, R01HL095396, P30DK027651, SOLAR-MH059490, HG-0004314; U.S. Cystic Fibrosis Foundation grants: CUTTIN06P0, COLLAC07A0, R025-CR07, KNOWLE00A0, RDP-R026-CR07, DRUMM0A00 and contract GENOMEQUEBEC07DDS0; Flight Attendant Medical Research Institute grant: FAMRI2006; Lawson Wilkins Pediatric Endocrine Society grant: LWPES Clinical Scholar Award; The Canadian Cystic Fibrosis Foundation; Genome Canada through the Ontario Genomics Institute as per research agreement 2004-OGI-3-05; Ontario Research Fund, Research Excellence Program; Lloyd Carr-Harris Foundation; Joint Fellowship of Canadian Institutes of Health Research and Ontario Women's Health Council. Funds for genome-wide genotyping were generously provided by the U.S. CFF.

The authors would like to thank the CFF Patient Registry, the University of North Carolina DNA Laboratory, Genome Quebec and McGill University Innovation Centre and the following for their contributions: **Manuscript preparation:** Patricia Cornwall **Study Design:** Ada Hamosh, Rita McWilliams **Recruitment:** Sonya Adams, Marilyn Algire, Nicole Anderson, Amanda Bowers, Jennifer Breaton, Colette Bucur, Leia Charnin, Mary Christofi, Barbara Coleman, John Dunn, Barry Elkind, Despina Fragolias, Julie Hoover-Fong, Katherine Keenan, Patricia Miller, Sarah Norris, Roxanne Rousseau, Sally Wood, Rossitta Yung, Catherine Yurk **Phenotyping:** Nicole Anderson, Jennifer Breaton, Mary Christofi, Despina Frangolias, Wan Ip, Katherine Keenan, Roxanne Rousseau, Rossitta Yung **Data Entry:** John Dunn, Sarah Norris, Teresa Lai **Genotyping:** Kaeleen Boden, Rebecca Darrah, Qiuju Huang, Kyle Kanieki, Fan Lin, Sarah Ritter, Nulang Wang, Yongqian Wang, Colleen Weiler, Whitney Wolf, Xiaowei Yuan **Analysis:** Tony Dang, Evan Hawbaker, Lindsay Henderson, Rita McWilliams, Yunfei Wang, Christopher Watson, Aaron Webel, **Bioinformatics:** Edgar Crowdy, Hong Dang, Hemant Kelkar, Tom Randall, Annie Xu

Appendix

Contributing North American CF Centers and Principal Investigators

Aaron,S., Ottawa General Hospital, Ottawa, Canada/Accurso,F., University of Colorado Health Sciences Center, CO/Acton,J., Cincinnati Children's Hospital and Medical Center, OH/Ahrens,R., University of Iowa Hospitals & Clinics, IA/Aljadeff,G., Lutheran General Children's Hospital, IL/Allard,C., Centre de Santé et de Services Sociaux de Chicoutimi, Chicoutimi, Canada/Amaro,R., University of Texas at Tyler Health Center, TX/Anbar,R., SUNY Upstate Medical University, NY/Anderson,P., University of Arkansas, AR/Atlas,A., Morristown Memorial Hospital, NJ/Bell,S., The Prince Charles Hospital, Australia/ Berdella,M., St. Vincents Hospital & Medical Center, NY /Biller,J., Children's Hospital of Wisconsin, WI/Black,H., Asthma & Allergy Specialists, Charlotte, NC/Black,P., Children's Mercy Hospital, MO/Boas,S., Children's Asthma Respiratory&Exercise Specialists, IL/ Boland,M., Children's Hospital of Eastern Ontario, Ottawa, Canada/Borowitz,D., Women & Children's Hospital of Buffalo, NY/Boswell,R., University of Tennessee, TN/Boucher,J., Centre Hospitalier Régional de Rimouski, Rimouski, Canada/Bowman,C.M., Medical University of South Carolina, SC/Boyle,M., Johns Hopkins Hospital, MD/Brown,C., California Pacific Medical Center, CA/Brown,D., Pediatric Pulmonary Associates., SC/ Brown,N., University of Alberta Hospitals, Edmonton, Canada/Caffey,L.F., University of New Mexico, NM/Chatfield,B., University of Utah Health Sciences Center, UT/ Chesrown,S., University of Florida, FL/Chipps,B., Sutter Medical Center, CA/Clancy,J.P., University of Alabama at Birmingham, AL/Cohen,R., Kaiser Permanente, OR/Colombo,J., University of Nebraska Medical Center, NE/Cronin,J., Women & Children's Hospital of Buffalo, NY/Cruz,M., St. Mary's Medical Center, FL/Cunningham,J., Cook Children's Medical Center, TX/Davidson,G., B.C. Children's Hospital, Vancouver, Canada/Davies,J, University of New Mexico, NM/Davies,L., University of New Mexico, SOM, NM/DeCelie-Germana,J., Schneider Children's Hospital, NY/Devenny,A., Royal Hospital for Sick Children, Scotland/DiMango,E., Columbia University Medical Center, NY/Doornbos,D., Via-Christi, St. Francis Campus, KS/Dozor,A., New York Medical College-Westchester Medical Center, NY/Dunitz,J., University of Minnesota, MN/Egan,M., Yale University SOM, CT/Eichner,J., Great Falls Clinic, MT/Ferkol,T., St. Louis Children's Hospital, MO/ Fiel,S., Morristown Memorial Hospital, NJ/Flume,P., Medical University of South Carolina, SC/ Freitag,A.,Hamilton Health Sciences Corporation, Hamilton, Canada/ Franco,M., Miami Children's Hospital, FL/Froh,D., University of Virginia Health System, VA/ Garey,N., Saint John Regional Hospital, Saint John, Canada/Geller,D., Nemours Children's Clinic Orlando, FL/Gershan,W., Children's Hospital of Wisconsin, WI/Gibson,R., Children's Hospital & Regional Medical Center, WA/Giusti,R., Long Island College Hospital, NY/Gjevre,J., Royal University Hospital, Saskatoon, Canada/Gondor,M., University of South Florida, FL/Gong,G., Phoenix Children's Hospital, AZ/Guill,M., Medical College of Georgia, GA/Gutierrez,H., University of Alabama at Birmingham, AL/ Hadeh,A., Drexel University College of Medicine, PA/Hardy,K., Children's Hospital - Oakland, CA/Hiatt,P., Texas Children's Hospital, TX/Hicks,D., Children's Hospital of Orange County, CA/Holmes,B., Regina General Hospital, Regina, Canada/Holsclaw,D., University of Pennsylvania, PA/Holzwarth,P., St. Vincent Hospital - Genetics, WI/

Honicky,R., Michigan State University, MI/Howenstine,M., Riley Hospital for Children, IN/ Hughes,D., IWK Health Centre, Halifax, Canada/Jackson,M., Grand River Hospital, Kitchener, Canada/James,P., Lutheran Hospital, IN/ Jenneret A., Hôtel Dieu de Montréal, Montréal, Canada/Joseph,P., University of Cincinnati, OH/Kanga,J., University of Kentucky, KY/Katz,M., Baylor College of Medicine, TX/Kent,S., Victoria General Hospital, Victoria, Canada/Kepron,W., Winnipeg Health Sciences Centre, Winnipeg, Canada/Knowles,M., University of North Carolina at Chapel Hill, NC/Konig,P., University of Missouri- Columbia, MO/Konstan,M., Case Western Reserve University, OH/ Kovesi,T., Children's Hospital of Eastern Ontario, Ottawa, Canada/Kramer,J., Oklahoma Cystic Fibrosis Center, OK/Kraynack,N., Children's Hospital Medical Center of Akron, OH/ Kumar,V., Hôpital Régional de Sudbury Regional Hospital, Sudbury, Canada/Lahiri,T., Fletcher Allen Health Care, VT/Landon,C., Pediatric Diagnostic Center, CA/Lands,L., Montréal Children's Hospital, Montréal, Canada/Lapin,C., Connecticut Children's Medical Center, CT/Larj,M., Wake Forest University Baptist Med. Ctr., NC/Ledbetter,J., TC Thompson Children's Hospital, TN/Lee,R., Naval Medical Center - Portsmouth, VA/ Leigh,M., University of North Carolina at Chapel Hill, NC/Lester,L., University of Chicago Children's Hospital, IL/Lever,T., Eastern Maine Medical Center, ME/Levy,H., Children's Hospital Boston, MA/Lieberthal,A., Kaiser Permanente Southern California, CA/Liou,T., University of Utah, UT/Lipton,A., National Naval Medical Center, MD/Lyttle,B., Children's Hospital of Western Ontario, London, Canada/Lothian,B., Royal University Hospital, Saskatoon, Canada/Lougheed,D., Hotel Dieu Hospital, Kingston, Canada/Malhotra,K., Grand River Hospital, Kitchener, Canada/Marcotte,J., Hôpital Sante-Justine, Montréal, Canada/Matouk,E., Montréal Chest Institute, Montréal, Canada/McCarthy,M., Providence Medical Center, WA/McColley,S., Children's Memorial Hospital & Northwestern University, IL/McCoy,K., Columbus Children's Hospital, OH/McNamara,J., Children's Hospitals and Clinics of Minneapolis, MN/Michael,R., Queen Elizabeth II Health Sciences Centre, Halifax, Canada/Miller,S., University of Mississippi Medical Center, MS/Milot,M., Centre de Santé et de Services Sociaux de Chicoutimi, Chicoutimi, Canada/Moffett,K., West Virginia University, WV/Montgomery,M., Alberta Children's Hospital, Calgary, Canada/ Moore,P., Vanderbilt University Medical Center, TN/Morgan,W., Tucson Cystic Fibrosis Center, AZ/Morris,R., Janeway Children's Health & Rehabilitation, St. John's, Canada/ Morse,M., Methodist Children's Hospital, TX/Moskowitz,S., Children's Hospital & Regional Medical Center, WA/Moss,R., Stanford University Medical Center, CA/ Murphy,P., University of Nebraska Medical Center, NE/Nakielna,E., St. Paul's Hospital, Vancouver, Canada/Nasr,S., University of Michigan Medical Center, MI/Nassri,L., Sparks Regional Medical Center, AR/Naureckas,E., University of Chicago Hospitals, IL/ Nielson,D., University of California at San Francisco, CA/Noseworthy,M., Janeway Children's Health & Rehabilitation, St. John's, Canada/Noyes,B., St. Louis University, MO/ Olivier,K., Wilford Hall USAF Med. Ctr. San Antonio, TX/Olson,E., University of Florida, FL/Omlor,G., Akron Children's Hospital, OH/Orenstein,D., Children's Hospital of Pittsburgh, PA/O'Sullivan,B., University of Massachusetts Memorial Health Care, MA/ Parker,H.W., Dartmouth-Hitchcock Medical Center, NH/Passero,M., Brown University Medical School Rhode Island Hospital, RI/Pasterkamp,H., Children's Hospital of Winnipeg, Winnipeg, Canada/Pedder,L., Hamilton Health Sciences Corporation, Hamilton, Canada/ Perkett,E., Vanderbilt University Medical Center, TN/Perry,G., University of Kansas

Medical Center, KS/Petit,N., Center Hospitalier Rouyn-Noranda, Rouyn-Noranda, Canada/Pian,M., University of California San Diego Children's Hospital, CA/Platzker,A., Children's Hospital of Los Angeles, CA/Prestidge,C., Children's Medical Center of Dallas, TX/Price,A., Children's Hospital of Western Ontario, London, Canada/Rabin,H., Foothills Medical Centre, Calgary, Canada/Radford,P., Phoenix Children's Hospital, AZ/Ratjen,F., The Hospital for Sick Children, Toronto, Canada/Regelmann,W., University of Minnesota, MN/Ren,C., University of Rochester Medical Center, Strong Memorial Hospital, NY/Retsch-Bogart,G., University of North Carolina at Chapel Hill, NC/Richards,W., Memphis Lung Physicians, MS/Riva,M., Via-Christi, St. Francis Campus, KS/Rivard,L., Centre Unversitaire de Santé de L'estrie, Sherbrook, Canada/Roberts,D., Providence Medical Center, AK/Rock,M., University of Wisconsin Hospital, WI/Rosen,J., Albany Medical College, NY/Royall,J., Childrens Hospital of Oklahoma, OK/Rubenstein,R., Children's Hospital of Philadelphia, PA/Ruiz,F., University of Mississippi Med. Ctr., MS/Scanlin,T., Children's Hospital of Philadelphia, PA/Schechter,M., Emory University School of Medicine, GA/Schmidt,H.J., Virginia Commonwealth University, VA/Schwartzman,M., Joe DiMaggio Children's Hospital, FL/Scott,P., Georgia Pediatric Pulmonology Assoc., PC, GA/Shay,G., Kaiser Permanente Medical Center, CA/Simon,R., University of Michigan Health System, MI/Smith,P., Long Island College Hospital, NY/Solomon,M., The Hospital for Sick Children, Toronto, Canada/Spencer,T., Children's Hospital of Boston, MA/Stecenko,A., Emory University, GA/Stokes,D., University of Tennessee, TN/Sullivan,B., Marshfield Clinic, WI/Taylor-Cousar,J., University of New Mexico, NM/Thomas,N., Pennsylvania State University College of Medicine, PA/Thompson,H., St. Luke's CF Clinic, ID/Toder,D., Children's Hospital of Michigan and Harper University Hospital, MI/Tullis,E., St.Michael's Hospital, Toronto, Canada/Turcios,N., University of Medicine & Dentistry of NJ, NJ/van Wylick,R., Hotel Dieu Hospital, Kingston, Canada/Varlotta,L., St. Christopher's Hospital for Children, PA/Vauthy,P., Toledo Children's Hospital, OH/Voynow,J., Duke University, NC/Wainwright,C., Royal Children's Hospital, Australia/Walker,P., St. Vincent's Hospital - Manhattan, NY/Warren,W.S., Hershey Medical Center, PA/Wilcox,P., Royal Jubilee Hospital, Victoria, Canada/Wilmott,R., St. Louis University, MO/ Wilcox,P., St. Paul's Hospital, Vancouver, Canada/Wojtczak,H., Naval Medical Center - San Diego, CA/Yee,W., New England Medical Center, MA/Zacher,C., St. Alexius Heart & Lung CF Clinic, ND/Zanni,R., Monmouth Medical Center, NJ/Zeitlin,P., Johns Hopkins Hospital, MD/Zuberbuhler,P., University of Alberta Hospitals, Edmonton, Canada.

References

1. Hamosh A, Corey M. Correlation between genotype and phenotype in patients with cystic fibrosis. The Cystic Fibrosis Genotype-Phenotype Consortium. N Engl J Med. 1993; 329:1308–1313. [PubMed: 8166795]
2. Vanscoy LL, et al. Heritability of lung disease severity in cystic fibrosis. Am J Respir Crit Care Med. 2007; 175:1036–1043. [PubMed: 17332481]
3. Cutting GR. Modifier genes in Mendelian disorders: the example of cystic fibrosis. Ann N Y Acad Sci. 2010; 1214:57–69. [PubMed: 21175684]
4. Chalmers JD, Fleming GB, Hill AT, Kilpatrick DC. Impact of mannose binding lectin (MBL) insufficiency on the course of cystic fibrosis: a review and meta-analysis. Glycobiology. 2010 [PubMed: 21045008]

5. Corey M, Edwards L, Levison H, Knowles M. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. J Pediatr. 1997; 131:809–814. [PubMed: 9427882]
6. Schluchter MD, Konstan MW, Davis PB. Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients. Stat Med. 2002; 21:1271–1287. [PubMed: 12111878]
7. Kulich M, et al. Disease-specific reference equations for lung function in patients with cystic fibrosis. Am J Respir Crit Care Med. 2005; 172:885–891. [PubMed: 15976373]
8. Taylor C, et al. A Novel Lung Disease Phenotype Adjusted for Mortality Attrition for Cystic Fibrosis Genetic Modifier Studies. Ped Pulm. 2011 Epub no. 10.1002/ppul.21456
9. Drumm ML, et al. Gene modifiers of lung disease in cystic fibrosis. N Engl J Med. 2005; 353:1443–1453. [PubMed: 16207846]
10. Dorfman R, et al. Complex two-gene modulation of lung disease severity in children with cystic fibrosis. J Clin Invest. 2008; 118:1040–1049. [PubMed: 18292811]
11. Li W, et al. Understanding the population structure of North American patients with cystic fibrosis. Clin Genet. 2010; 79:136–146. [PubMed: 20681990]
12. Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. Am J Hum Genet. 2007; 81:913–926. [PubMed: 17924335]
13. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–213. [PubMed: 16415888]
14. Li Y, Ding J, Abecasis GR. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. American Society of Human Genetics Annual Meeting - 56th Annual Meeting. 2006:416.
15. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995; 11:241–247. [PubMed: 7581446]
16. Goring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet. 2001; 69:1357–1369. [PubMed: 11593451]
17. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57(1):289–300.
18. Sun L, Craiu RV, Paterson AD, Bull SB. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet Epidemiol. 2006; 30:519–530. [PubMed: 16800000]
19. Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve power of association in genome scans. Am J Hum Genet. 2006; 78:243–252. [PubMed: 16400608]
20. Day A, Carlson MR, Dong J, O'Connor BD, Nelson SF. Celsius: a community resource for Affymetrix microarray data. Genome Biol. 2007; 8:R112. [PubMed: 17570842]
21. Day A, et al. Disease gene characterization through large-scale co-expression analysis. PLoS One. 2009; 4:e8491. [PubMed: 20046828]
22. Cao G, et al. Cloning of a novel Apaf-1-interacting protein: a potent suppressor of apoptosis and ischemic neuronal cell death. J Neurosci. 2004; 24:6189–6201. [PubMed: 15240811]
23. Cho DH, et al. Induced inhibition of ischemic/hypoxic injury by APIP, a novel Apaf-1-interacting protein. J Biol Chem. 2004; 279:39942–39950. [PubMed: 15262985]
24. Cho DH, et al. Suppression of hypoxic cell death by APIP-induced sustained activation of AKT and ERK1/2. Oncogene. 2007; 26:2809–2814. [PubMed: 17086211]
25. Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. Gene. 2008; 303:11–34. [PubMed: 12559563]
26. Tugores A, et al. The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor downstream of MAPK signaling cascades. J Biol Chem. 2001; 276:20397–20406. [PubMed: 11259407]
27. Kas K, et al. ESE-3, a novel member of an epithelium-specific ets transcription factor subfamily, demonstrates different target gene specificity from ESE-1. J Biol Chem. 2000; 275:2986–2998. [PubMed: 10644770]

28. Silverman ES, et al. Constitutive and cytokine-induced expression of the ETS transcription factor ESE-3 in the lung. Am J Respir Cell Mol Biol. 2002; 27:697–704. [PubMed: 12444029]
29. Wu J, et al. Regulation of epithelium-specific Ets-like factors ESE-1 and ESE-3 in airway epithelial cells: potential roles in airway inflammation. Cell Res. 2008; 18:649–643. [PubMed: 18475289]
30. Dezso Z, et al. A comprehensive functional analysis of tissue specificity of human gene expression. BMC Biol. 2008; 6:49. [PubMed: 19014478]
31. Dibbert B, et al. Cytokine-mediated Bax deficiency and consequent delayed neutrophil apoptosis: a general mechanism to accumulate effector cells in inflammation. Proc Natl Acad Sci U S A. 1999; 96:13330–13335. [PubMed: 10557320]
32. McKeon DJ, et al. Prolonged survival of neutrophils from patients with Delta F508 CFTR mutations. Thorax. 2008; 63:660–661. [PubMed: 18587042]
33. Harris JF, et al. Bcl-2 sustains increased mucous and epithelial cell numbers in metaplastic airway epithelium. Am J Respir Crit Care Med. 2005; 171:764–772. [PubMed: 15618464]
34. Butler AA. The melanocortin system and energy balance. Peptides. 2006; 27:281–290. [PubMed: 16434123]
35. Lee YS, Poh LK, Loke KY. A novel melanocortin 3 receptor gene (MC3R) mutation associated with severe obesity. J Clin Endocrinol Metab. 2002; 87:1423–1426. [PubMed: 11889220]
36. Savastano D, et al. Energy intake and energy expenditure among children with polymorphisms of the melanocortin-3 receptor 1-4. Am J Clin Nutr. 2009; 90:912–920. [PubMed: 19656839]
37. Dorlochter L, Roksund O, Helgheim V, Rosendahl K, Fluge G. Resting energy expenditure and lung disease in cystic fibrosis. J Cyst Fibros. 2002; 1:131–136. [PubMed: 15463819]
38. McCloskey M, et al. Energy balance in cystic fibrosis when stable and during a respiratory exacerbation. Clin Nutr. 2004; 23:1405–1412. [PubMed: 15556263]
39. Getting SJ, et al. A role for MC3R in modulating lung inflammation. Pulm Pharmacol Ther. 2008; 21:866–873. [PubMed: 18992358]
40. Tikhmyanova N, Little J. CAS proteins in normal and pathological cell growth control. Cell Mol Life Sci. 2010; 67:1025–1048. [PubMed: 19937461]
41. Pugacheva E, Jablonski S, Hartman T, Henske E, Golemis E. HEF1-dependent aurora A activation induces disassembly of the primary cilium. Cell. 2007; 129:1351–1363. [PubMed: 17604723]
42. Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and environmental influences on pulmonary function in adult twins. Am Rev Respir Dis. 1982; 125:409–415. [PubMed: 7200340]
43. McClearn GE, Svartengren M, Pedersen NL, Heller DA, Plomin R. Genetic and environmental influences on pulmonary function in aging Swedish twins. J Gerontol. 1994; 49:264–268. [PubMed: 7963289]
44. Wilk JB, et al. A genome-wide association study of pulmonary function measures in the Framingham heart study. PLoS Genet. 2009; 5:e1000429. [PubMed: 19300500]
45. Hancock D, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet. 2010; 42:45–52. [PubMed: 20010835]
46. Repapi E, et al. Genome-wide association study identifies five loci associated with lung function. Nat Genet. 2010; 42:36–44. [PubMed: 20010834]
47. Pillai SG, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet. 2009; 5:e1000421. [PubMed: 19300482]
48. Cho MH, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet. 2010; 42:200–202. [PubMed: 20173748]
49. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
50. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002; 21:589–624. [PubMed: 11836738]
51. Burdick JT, Chen WM, Abecasis GR, Cheung VG. In silico method for inferring genotypes in pedigrees. Nat Genet. 2006; 38:1002–1004. [PubMed: 16921375]

52. Wang K, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17:1665–1674. [PubMed: 17921354]
53. Sun W, et al. Integrated study of copy number states and genotype calls using high-density SNP arrays. Nucleic Acids Res. 2009; 37:5365–5377. [PubMed: 19581427]
54. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30:97–101. [PubMed: 11731797]
55. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62:1198–1211. [PubMed: 9545414]
56. Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2002; 64:479–498.
57. Yoo YJ, Bull SB, Paterson AD, Waggott D, Sun L. Were genome-wide linkage studies a waste of time? Exploiting candidate regions within genome-wide association studies. Genet Epidemiol. 2010; 34:107–118. [PubMed: 19626703]
58. Neter, J.; Wasserman, W.; Kutner, M. Applied Linear Regression Models. Irwin, IL: 1989. p. 282
59. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med. 1999; 159:179–187. [PubMed: 9872837]
60. Wang X, Dockery DW, Wypij D, Fay ME, Ferris BG Jr. Pulmonary function between 6 and 18 years of age. Pediatr Pulmonol. 1993; 15:75–88. [PubMed: 8474788]

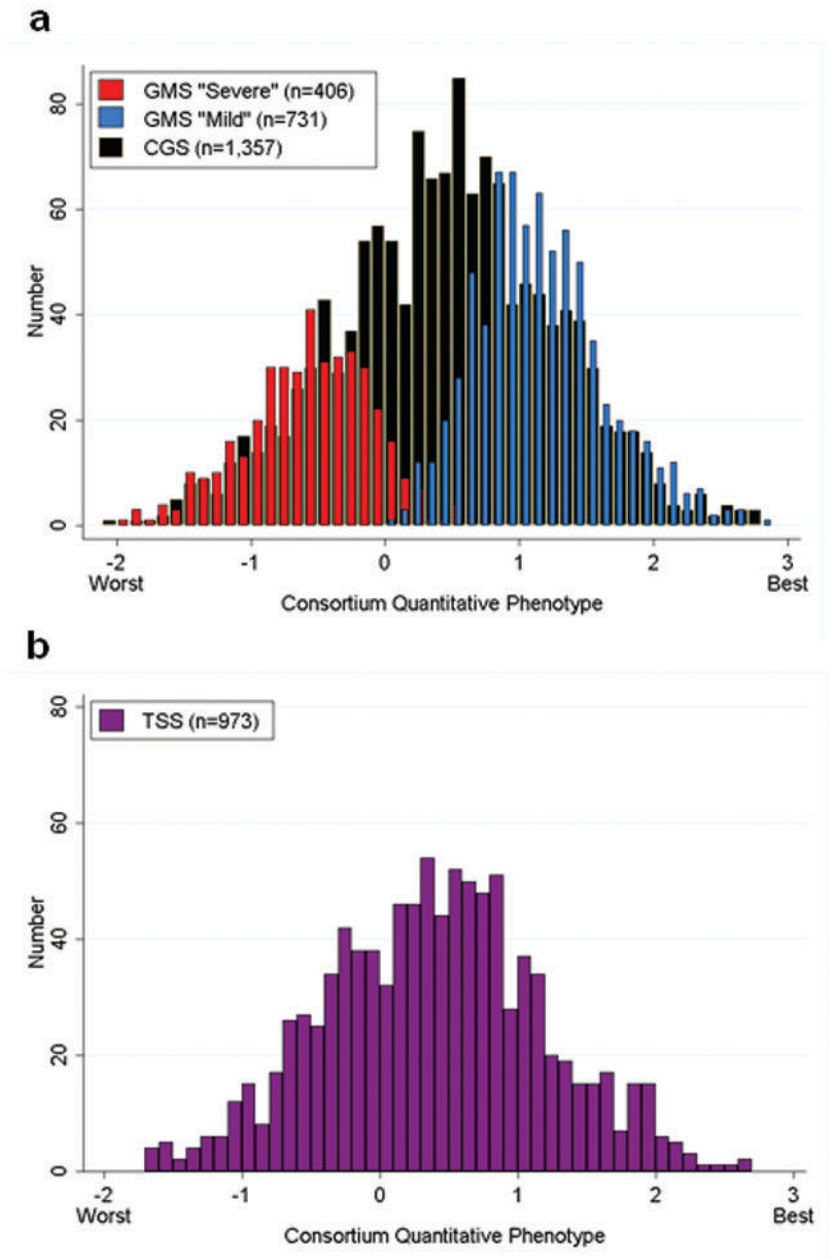

Figure 1.

Histograms of the Consortium lung phenotype for the three cystic fibrosis studies show similar average phenotypes. The phenotype mean is above zero due to a lower bound placed by the survival correction, as well as cohort effects of improving lung function. (a) The two designs using unrelated individuals. All of the patients in the Genetic Modifier Study (GMS) are *F508del/F508del* at CFTR. These patients were oversampled at extremes of an initial entry phenotype, in order to improve power, and the original severe/mild designations are colored separately. In contrast, the Canadian Consortium for Genetic Studies (CGS) is population based, representing a range of pancreatic insufficient CFTR genotypes. (b) Patients enrolled in the family-based Twin and Sibling Study (TSS) show a similar distribution of the Consortium lung phenotype as the population-based CGS.

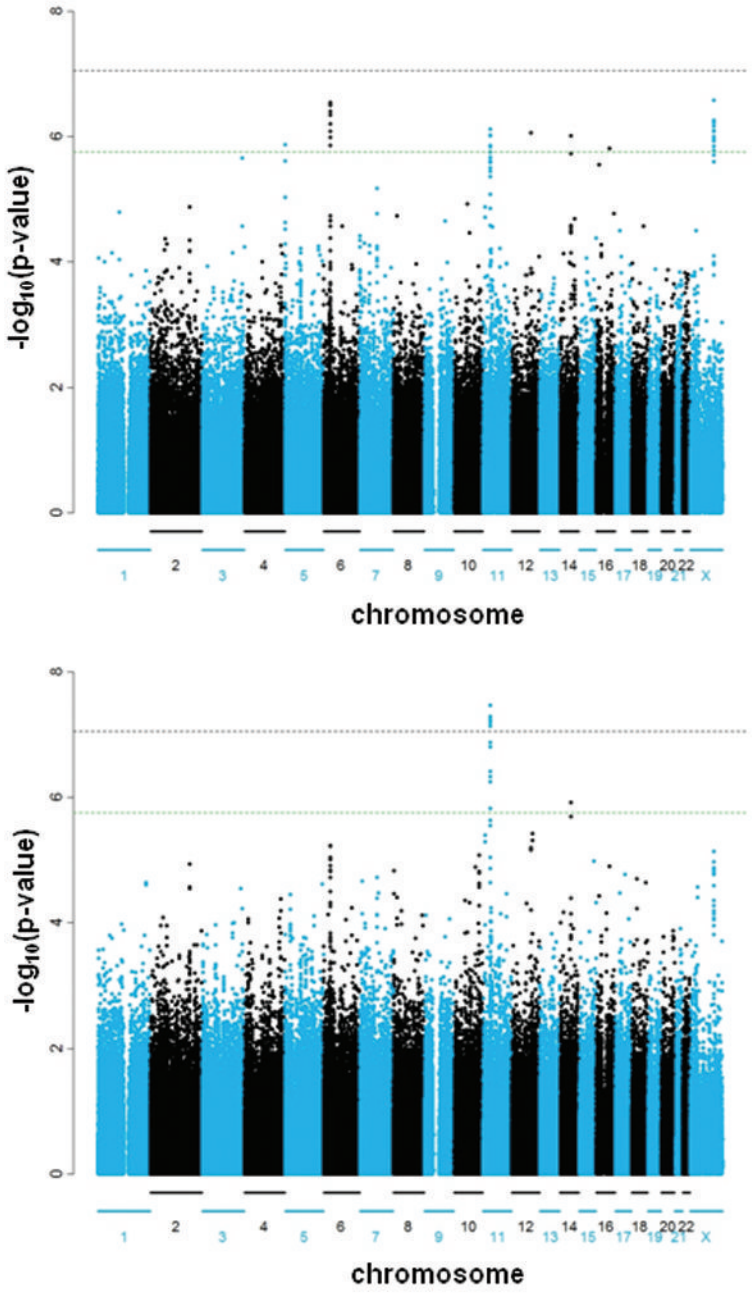

Figure 2.

Genome-wide Manhattan plots for the cystic fibrosis Consortium lung function phenotype, combining the association evidence from GMS and CGS samples across 570,725 SNPs. The black dashed line represents the Bonferroni threshold for genome-wide α=0.05, while the green dashed line is the suggestive association threshold, expected once per genome scan. SNPs are plotted in Mb relative to their position on each chromosome (alternating blue and black) (a) Results from GMS (n=1137, all of whom are *F508del/F508del*) combined with all of the CGS patients (n=1357). Seven regions reach suggestive significance. (b) Results from

the combined evidence of GMS (n=1137) and the CGS *F508del/F508del* (n=841). A region on chromosome 11p13 reaches genome-wide significance ($P=3.34 \times 10^{-8}$).

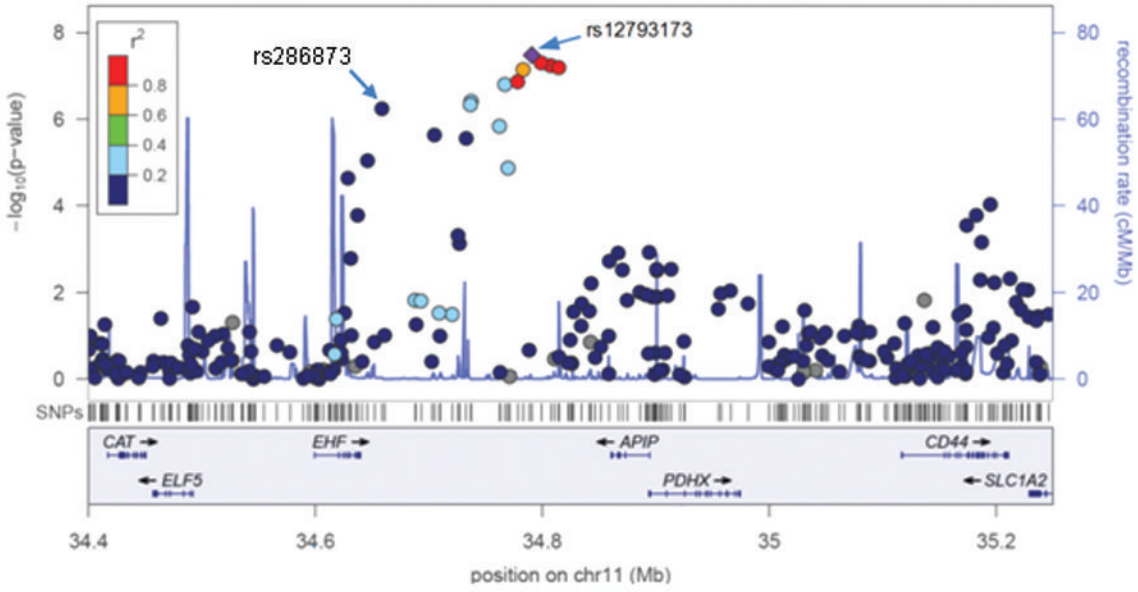

Figure 3.

A plot of the association evidence in GMS and CGS *F508del/F508del* in the chromosome 11p13 *EHF/APIP* region (NCBI build 36, LocusZoom viewer). Colors represent HapMap CEU linkage disequilibrium r^2 with the most significant SNP, rs12793173 (P=3.34 × 10^{-8}). The secondary peak at rs286873 has relatively low r^2 with the primary peak.

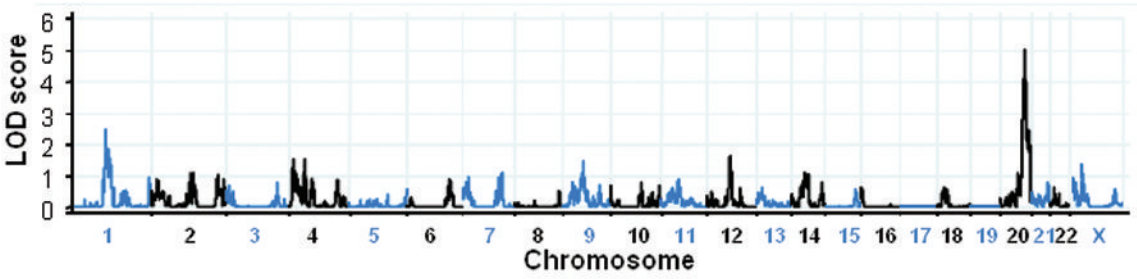

Figure 4.

Genome-wide linkage scan for the Consortium lung phenotype of 486 sibling pairs in the family-based TSS, adjusted for sex. A QTL with a genome-wide significant LOD=5.03 was found on 20q13.2. LOD scores with SNPs used in the linkage panel are plotted in cM relative to their position on each chromosome (alternating blue and black).

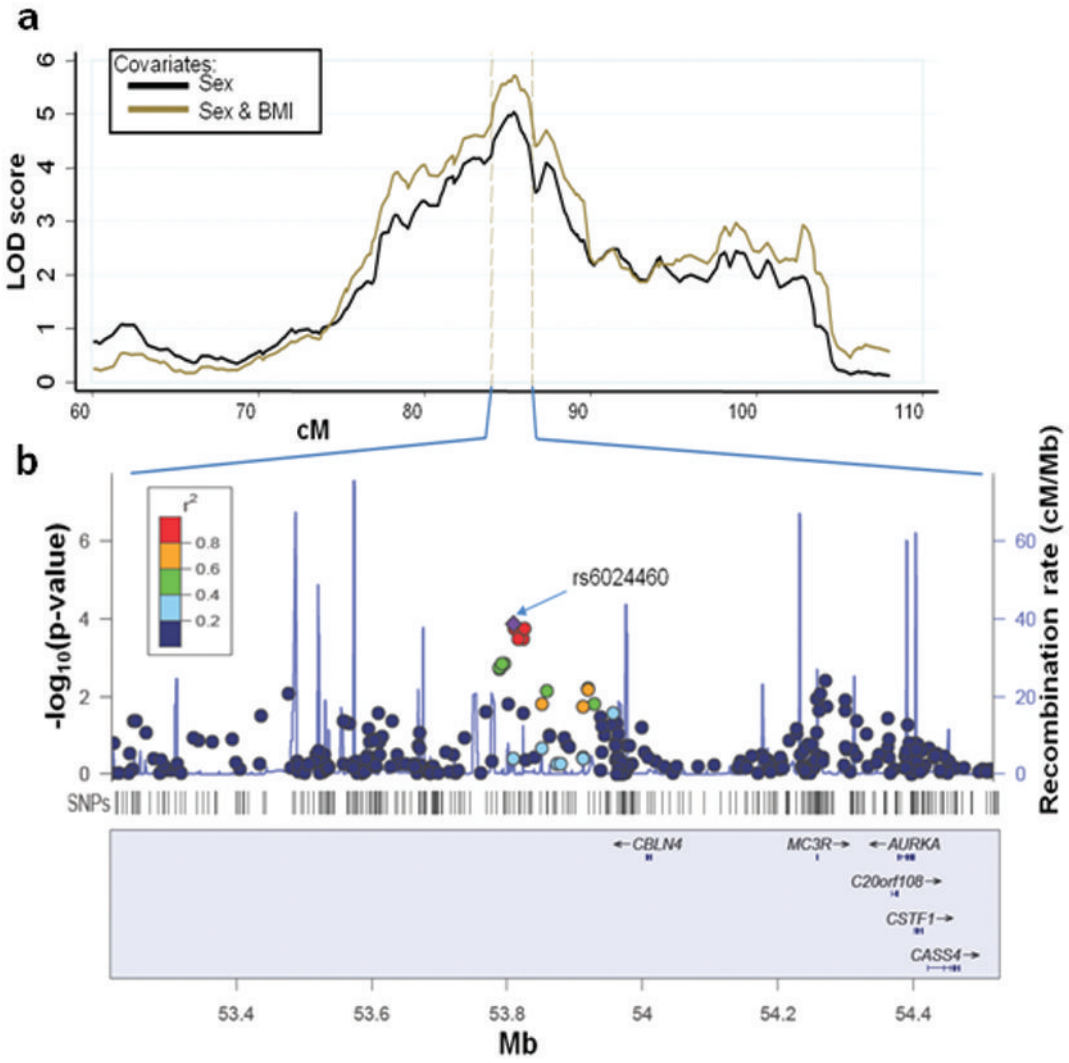

Figure 5.

Regional analysis of the QTL on chromosome 20q13.2 **(a)** A detailed chromosome 20 linkage plot for the Consortium lung phenotype in the TSS study, with covariates sex (essentially the same result as for no covariates) and with covariates sex and BMI. **(b)** Association evidence from the GMS and CGS *F508del/F508del* patients, in the 1-LOD support interval provided by TSS. A region centromeric to *CBLN4* and *MC3R* on 20q13.2 shows suggestive evidence of association, with the greatest evidence at rs6024460 (P=1.34 × 10^{-4}).

Table 1

Characteristics of patients enrolled by the three studies comprising the North American CF Gene Modifier Consortium

	Genetic Modifier Study (GMS)		Canadian Consortium for Genetic Studies (CGS)	Twins & Sibs Study (TSS)
Lead Institution(s)	Univ. of North Carolina/Case Western		Hosp. Sick Children	Johns Hopkins
Design	Extremes-of-Phenotype Unrelated		Population-Based Unrelated	Family-Based
Type of Evidence	Association		Association	Linkage and association
Number of patients	1,137		1,357	973 [a] (486 sibling pairs)
	Severe (*n* = 406)	Mild (*n* = 731)		
Age				
Mean ± SD (yrs)	15.2 ± 4.6	27.5 ± 9.8	18.5 ± 9.5	15.5 ± 7.8
Range (yrs)	8-25	15-56	6-49	6-55
Male *n* (%)	194 (47.8%)	405 (55.4%)	734 (54.1%)	521 (53.5)
Caucasian *n* (%)[b]	1,137 (100.0%)		1,180 (87.0%)	898 (92.3%)
***F508del/F508del n* (%)**	1,137 (100.0%)		841 (62.0%)	557 (57.2%)
Pancreatic Exocrine Insufficient *n* (%)	1,137 (100.0%)		1,357 (100.0%)	973 (100.0%)

[a] 420 two-sib families, 20 three-sib families, 1 four-sib family and 69 singletons.

[b] Based on self-identified ancestry and principal components analysis.

Table 2

Significant and suggestive association results for GMS and CGS, with replication values for TSS

SNP	Chr	Base pair[a]	Nearest Gene	Category[b]	Risk allele[c]	Non-risk allele[c]	(Minor allele) Freq.[d]	GMS coef.[e]	CGS *F508del/ F508del* coef.	CGS All coef	Analysis with max significance	*P*-value: GMS+CGS *F508del/ F508del*	*P*-value: GMS+ CGS All	*P*-value: TSS[f]	*P*-value: Joint[g]
rs12793173	11	34,790,780	*APIP/EHF*	Significant	C	T	(C) 0.24	0.16	0.20	0.12	GMS+CGS *F508del/F508 del*	3.34E-08	1.76E-06	0.006	1.49E-09
rs1403543	X	115,216,220	*AGTR2*	Suggestive	A	G	(G) 0.49	0.22	0.07	0.11	GMS+CGS All	1.61E-05	2.58E-07	0.053	1.71E-06
rs9268905	6	32,540,055	*HLA-DRA*	Suggestive	C	G	(C) 0.32	0.16	0.10	0.12	GMS+CGS All	1.42E-05	2.81E-07	0.032	1.21E-07
rs4760506	12	91,857,181	*EEA1*	Suggestive	G	A	(A) 0.45	0.16	0.10	0.10	GMS+CGS All	6.77E-06	8.56E-07	0.594	9.15E-05
rs12883884	14	69,586,936	*SLC8A3*	Suggestive	T	G	(G) 0.39	0.12	0.15	0.12	GMS+CGS All	1.20E-06	9.56E-07	0.223	7.81E-06
rs12188164	5	481,236	*AHRR*	Suggestive	A	C	(A) 0.38	0.08	0.12	0.15	GMS+CGS All	5.92E-04	1.34E-06	0.136	3.65E-06
rs11645366	16	60,934,654	*CDH8*	Suggestive	C	T	(T) 0.23	0.17	0.13	0.13	GMS+CGS All	1.23E-05	1.52E-06	0.182	7.03E-06

[a] NCBI build 36.

[b] Significant and suggestive imply $P \leq (0.05/570725) = 8.76 \times 10^{-8}$ or $P \leq (1/570725) = 1.75 \times 10^{-6}$, respectively, for at least one analysis (GMS+CGS *F508del/F508del* or GMS+CGS All).

[c] Alleles indexed to the forward strand of NCBI build 36; the risk allele is the allele associated with worse lung function.

[d] Minor allele frequencies are listed for all GMS +CGS All. Study-specific MAFs are provided in Supplementary Table 1.

[e] Coefficients refer to the average reduction in the Consortium lung phenotype for each copy of the risk allele.

[f] TSS direction-consistent association p-value, for TSS *F508del/F508del* only, or TSS All Patients, selected according to the GMS+CGS result with maximum significance.

[g] Joint meta-analysis P-value for GMS, CGS, and TSS, with selection of patients (*F508del/F508del* only, or All Patients) according to the GMS+CGS result with maximum significance

Nat Genet. Author manuscript; available in PMC 2012 March 07.

Table 3

Combined association and linkage-weighted FDR *q*-values and genome-wide ranks for SNPs with WFDR q-values genome-wide significant (< 0.05)

Chr	SNP	Base Pair	GMS+CGS *F508del/F508del* Association *P* value	FDR *q*-value[a]	FDR rank	WFDR *q*-value[b]	WFDR rank
11	rs93138	9716289	5.08E-06	0.0124	7	0.0383	16
11	rs11032829	34705078	2.29E-06	0.008	3	0.0277	8
11	**rs7924717**	**34732907**	**2.76E-06**	**0.008**	**1**	**0.0218**	**6**
11	rs10466455	34737512	3.86E-07	0.0124	6	0.0375	14
11	rs7929679	34762425	1.47E-06	0.008	5	0.0282	10
11	rs10836312	34767019	1.56E-07	0.008	2	0.0277	7
11	rs525202	34778524	1.34E-07	0.008	4	0.0277	9
20	rs7265042	53790816	1.14E-03	0.7852	1976	0.0459	18
20	rs6098782	53791974	1.84E-03	0.7349	865	0.0459	17
20	rs910668	53794753	1.09E-03	0.5029	175	0.015	2
20	rs6092176	53799109	1.51E-03	0.5581	255	0.015	5
20	**rs6092179**	**53812440**	**1.93E-04**	**0.484**	**154**	**0.015**	**1**
20	rs6024437	53813962	1.61E-04	0.7553	1116	0.0353	13
20	rs8125625	53820352	2.49E-04	0.516	207	0.015	3
20	rs6024454	53826840	2.56E-04	0.7615	1348	0.0381	15
20	rs6024460	53828948	1.34E-04	0.7349	854	0.0296	12
20	rs11907114	53862354	2.79E-03	0.554	250	0.015	4
20	rs1326022	54277432	1.16E-03	0.7344	824	0.0296	11

[a] Benjamini-Hochberg approach based on association *P*-value.

[b] Weighted FDR using combined linkage information and association *P*-values.

Rows in bold indicate the top ranked SNPs before incorporating linkage evidence (rs7924717 on chromosome 11) and after (rs6092179 on chromosome 20)