

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

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

**A combined genome-wide association and linkage study was used to identify loci causing variation in cystic fibrosis lung disease severity. We identified a significant association ( $P = 3.34 \times 10^{-8}$ ) near *EHF* and *APIP* (chr11p13) in p.Phe508del homozygotes ( $n = 1,978$ ). The association replicated in p.Phe508del homozygotes ( $P = 0.006$ ) from a separate family based study ( $n = 557$ ), with  $P = 1.49 \times 10^{-9}$  for the three-study joint meta-analysis. Linkage analysis of 486 sibling pairs from the family based study identified a significant quantitative trait locus on chromosome 20q13.2 ( $\log_{10}$  odds = 5.03). Our findings provide insight into the causes of variation in lung disease severity in cystic fibrosis and suggest new therapeutic targets for this life-limiting disorder.**

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

Whole-genome methods provide an attractive approach to identify modifier loci of Mendelian disorders. However, cystic fibrosis presents

numerous challenges, such as (i) collecting multiple years of lung function measures to accurately classify lung disease severity; (ii) selecting the appropriate study design to identify common and rare variants; (iii) accruing sufficient sample sizes; and (iv) accounting for potential interaction between *CFTR* and modifier loci. To overcome these challenges, we formed the North American Cystic Fibrosis Gene Modifier Consortium to identify modifiers of lung disease severity and other phenotypes. For lung disease in cystic fibrosis, the forced expiratory volume in 1 s (FEV<sub>1</sub>) is the most clinically useful measure of lung disease severity and is a well-established predictor of survival<sup>5,6</sup>. However, comparison of FEV<sub>1</sub> measures across a broad age range of individuals with cystic fibrosis 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<sub>1</sub> over 3 years<sup>7</sup> that displays robust genetic influence ( $h^2 = 0.51$ ) (ref. 8).

<sup>1</sup>Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. <sup>2</sup>Child Health Evaluative Sciences, the Hospital for Sick Children, Toronto, Ontario, Canada. <sup>3</sup>Dalla Lana School of Public Health, the University of Toronto, Toronto, Ontario, Canada. <sup>4</sup>McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. <sup>5</sup>Cystic Fibrosis-Pulmonary Research and Treatment Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. <sup>6</sup>Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. <sup>7</sup>Department of Medicine, Centre de recherche, Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada. <sup>8</sup>Department of Human Genetics, Emory University, Atlanta, Georgia, USA. <sup>9</sup>Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, Ontario, Canada. <sup>10</sup>The Center for Health Research, Oregon Clinical Translational Research Institute, Portland, Oregon, USA. <sup>11</sup>Southwest Foundation for Biomedical Research, San Antonio, Texas, USA. <sup>12</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. <sup>13</sup>Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. <sup>14</sup>James Hogg Research Centre, Providence Heart + Lung Institute, The University of British Columbia, Vancouver, British Columbia, Canada. <sup>15</sup>Department of Molecular Genetics, the University of Toronto, Toronto, Ontario, Canada. <sup>16</sup>Pediatric Pulmonology, Cystic Fibrosis Center, Naval Medical Center, Portsmouth, Virginia, USA. <sup>17</sup>Case Western Reserve University Departments of Pediatrics and Genetics, Cleveland, Ohio, USA. <sup>18</sup>Program in Physiology and Experimental Medicine, the Hospital for Sick Children, Toronto, Ontario, Canada. <sup>19</sup>Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada. Correspondence should be addressed to M.R.K. (knowles@med.unc.edu) or G.R.C. (gcutting@jhmi.edu).

Received 3 November 2010; accepted 22 April 2011; published online 22 May 2011; doi:10.1038/ng.838

**Table 1** Characteristics of patients enrolled by the three studies comprising the North American Cystic Fibrosis Gene Modifier Consortium

	Lead institution(s)	Design	Type of evidence	Number of subjects	Age		Male <i>n</i> (%)	Caucasian <i>n</i> (%) <sup>b</sup>	p.Phe508del/p.Phe508del <i>n</i> (%)	Pancreatic exocrine insufficient <i>n</i> (%)
					Mean ± SD <sup>a</sup>	Range <sup>a</sup>				
Genetic Modifier Study (GMS)	Univ. of North Carolina/Case Western	Extremes-of-phenotype unrelated	Association	Severe ( <i>n</i> = 406)	15.2 ± 4.6	8–25	194 (47.8)	1,137 (100.0)	1,137 (100.0)	1,137 (100.0)
				Mild ( <i>n</i> = 731)	27.5 ± 9.8	15–56	405 (55.4)			
Canadian Consortium for Genetic Studies (CGS)	Hospital for Sick Children	Population-based unrelated	Association	1,357	18.5 ± 9.5	6–49	734 (54.1)	1,180 (87.0)	841 (62.0)	1,357 (100.0)
Twins & Sibs Study (TSS)	Johns Hopkins	Family based	Linkage and association	973 <sup>c</sup>	15.5 ± 7.8	6–55	521 (53.5)	898 (92.3)	557 (57.2)	973 (100.0)

<sup>a</sup>Age and range in years. <sup>b</sup>Based on self-identified ancestry and principal components analysis. <sup>c</sup>Four hundred eighty-six sibling pairs from 420 two-sibling families, 20 three-sibling families, 1 four-sibling family and 69 singletons.

The Consortium is composed of three samples of individuals with cystic fibrosis recruited using different study designs. The Genetic Modifier Study (GMS) consists of unrelated individuals homozygous for the common cystic fibrosis allele p.Phe508del recruited from extremes of lung function<sup>9</sup>. The Canadian Consortium for Genetic Studies (CGS) enrolled unrelated individuals with pancreatic insufficiency from a population-based sample<sup>10</sup>. The cystic fibrosis Twin and Sibling Study (TSS) recruited families in which two or more surviving children have cystic fibrosis<sup>2</sup>. The GMS and CGS were designed for association analysis, whereas 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 (GWAS) use sample sizes that are several fold larger than those available for the population of individuals with cystic fibrosis, we sought to maximize power by (i) testing the association using combined data from GMS and CGS followed by replication using the association evidence from TSS and (ii) testing linkage using the TSS data followed by SNP association testing in linked regions in the unrelated individuals from GMS and CGS. We restricted our analysis to individuals with two severe loss-of-function *CFTR* alleles and a subset that had identical *CFTR* genotypes (homozygosity for p.Phe508del).

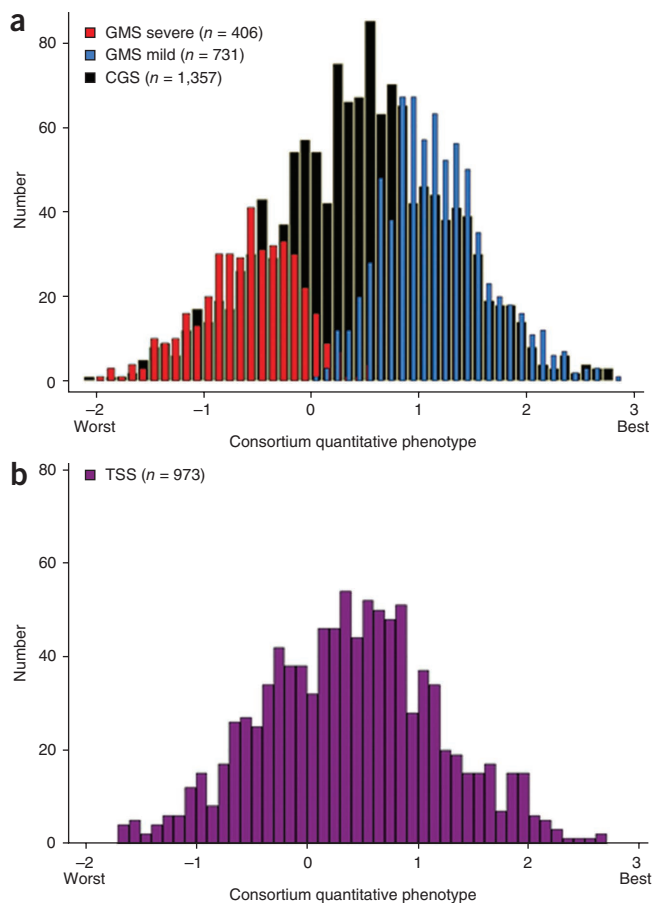
## RESULTS

### Association analysis of lung disease severity in cystic fibrosis

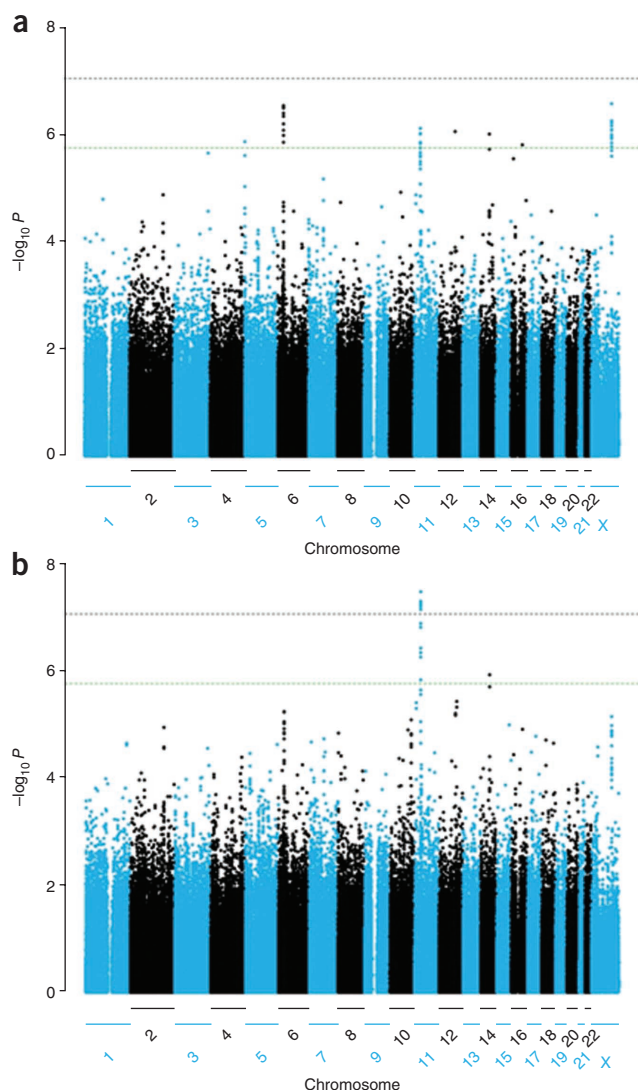
A total of 3,467 individuals with cystic fibrosis are represented in three study designs (Table 1 and Supplementary Note). Individuals in the GMS and 60% of those in the CGS and TSS are p.Phe508del homozygotes (p.Phe508del/p.Phe508del), whereas the remainder have other severe exocrine pancreatic *CFTR* genotypes<sup>2,9,10</sup>. The three samples showed consistent distributions of the lung disease phenotype, with the mid-range underrepresented in GMS because of the 'extremes of phenotype' design (Fig. 1). We contemporaneously genotyped affected individuals using the Illumina 610-Quad

array in a single facility with stringent quality control measures (Online Methods). Association scans for GMS and CGS used an additive model adjusted for sex and principal components as described<sup>11</sup>. We combined results using a directional meta-analysis approach for (i) GMS and CGS, *n* = 2,494, and (ii) GMS and CGS p.Phe508del/p.Phe508del, *n* = 1,978 (the power analysis is shown in Supplementary Fig. 1).

The combined GMS and CGS analysis identified seven regions with suggestive association ( $P \leq 1/570,725 = 1.75 \times 10^{-6}$ ) (Fig. 2 and Table 2). Restricting the analysis to individuals with p.Phe508del/p.Phe508del, the *EHF-APIP* region on 11p13 reached genome-wide significance at rs12793173 ( $P = 3.34 \times 10^{-8}$ , explaining 1.0% of the phenotype variation in GMS and 2.2% of the variation in CGS individuals with p.Phe508del/p.Phe508del). We verified this significance by permutation analysis and by developing an alternative conditional



**Figure 1** Histograms of the Consortium lung phenotype for the three cystic fibrosis studies showed similar average phenotypes. The phenotype mean is above zero because of a lower bound placed by the survival correction, as well as cohort effects of improving lung function. (a) The two designs using unrelated individuals are shown. All of the individuals in the Genetic Modifier Study (GMS) are p.Phe508del/p.Phe508del at *CFTR*. These individuals were oversampled at the extremes of an initial entry phenotype in order to improve power, and the original severe or 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) Individuals 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 (upper) dashed line represents the Bonferroni threshold for genome-wide  $\alpha = 0.05$ , and the green (lower) 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 ( $n = 1,137$ , all of whom are p.Phe508del/p.Phe508del) combined with all of the individuals from CGS ( $n = 1,357$ ). Seven regions reached suggestive significance. **(b)** Results from the combined evidence of GMS ( $n = 1,137$ ) and the CGS p.Phe508del/p.Phe508del individuals ( $n = 841$ ). A region on chromosome 11p13 reached genome-wide significance ( $P = 3.34 \times 10^{-8}$ ).

**Supplementary Table 1**). Two purported modifiers of cystic fibrosis lung disease, *TGFBI* and *IFRD1*, did not achieve genome-wide significance. *TGFBI* did, however, achieve  $P$  values in the range of  $P = 10^{-3}$  to  $P = 10^{-4}$  in the GMS sample, depending on additional covariates (**Supplementary Table 1**).

We evaluated the SNPs in the significant region and the six suggestive regions in GMS and CGS for association in TSS using MERLIN<sup>12</sup> while accounting for family structure. To be consistent with the GMS and CGS allelic effect, each replication test was one sided, and we chose the TSS sample (either all or individuals with p.Phe508del/p.Phe508del) for each suggestive SNP to be consistent with the GMS and CGS sample that provided maximum significance. We included covariates for sex and four principal components<sup>11</sup> for TSS. The SNP attaining genome-wide significance in GMS and CGS (rs12793173, p.Phe508del/p.Phe508del) showed significant association in the TSS p.Phe508del/p.Phe508del 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* ( $P = 0.053$ ), and neither was significant after correcting for the seven replication tests.

We next performed a joint analysis, which has been shown to be more powerful than testing followed by replication<sup>13</sup>, using a weighted meta-analysis procedure (Online Methods). Using all affected subjects, rs12793173 attained genome-wide significance ( $P = 1.12 \times 10^{-8}$ ). For this subject set, rs568529, a SNP in high linkage disequilibrium (LD) ( $r^2 > 0.9$ ) with rs12793173, achieved slightly greater significance ( $P = 9.75 \times 10^{-9}$ ). As in the earlier analysis, restricting the analysis to individuals with p.Phe508del/p.Phe508del increased the significance of *EHF-APIP* ( $P = 1.49 \times 10^{-9}$  for rs12793173 (**Table 2**) and  $P = 8.28 \times 10^{-10}$  for rs568529).

likelihood approach that acknowledged the GMS extremes of phenotype (Online Methods and **Supplementary Fig. 2**). With the inclusion of cystic fibrosis-relevant covariates (sex, body mass index (BMI) and previously associated genes), the association for rs12793173 was even stronger ( $P = 9.42 \times 10^{-9}$  for GMS and CGS p.Phe508del/p.Phe508del;

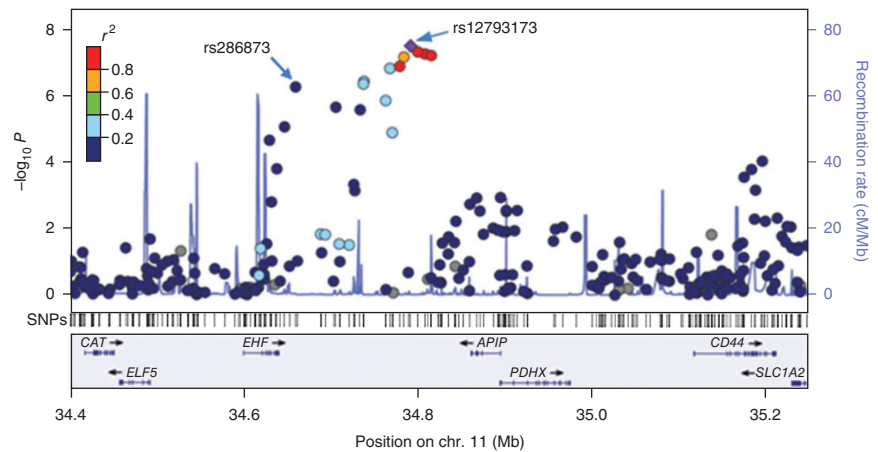
**Table 2** Significant and suggestive association results for GMS and CGS, with replication values for TSS

SNP	Chr.	Base pair <sup>a</sup>	Nearest gene	Category <sup>b</sup>	Risk allele <sup>c</sup>	Non-risk allele <sup>c</sup>	(Minor allele) frequency <sup>d</sup>	GMS coefficient <sup>e</sup>	CGS		Analysis with maximum significance	$P$ value, GMS + CGS p.Phe508del/p.Phe508del	$P$ value, GMS + CGS all	$P$ value, TSS <sup>f</sup>	$P$ value, joint <sup>g</sup>
									p.Phe508del/p.Phe508del	CGS all					
rs12793173	11	34,790,780	<i>APIP/EHF</i>	Significant	C	T	(C) 0.24	0.16	0.20	0.12	GMS + CGS p.Phe508del/p.Phe508del	$3.34 \times 10^{-8}$	$1.76 \times 10^{-6}$	0.006	$1.49 \times 10^{-9}$
rs1403543	X	115,216,220	<i>AGTR2</i>	Suggestive	A	G	(G) 0.49	0.22	0.07	0.11	GMS + CGS all	$1.61 \times 10^{-5}$	$2.58 \times 10^{-7}$	0.053	$1.71 \times 10^{-6}$
rs9268905	6	32,540,055	<i>HLA-DRA</i>	Suggestive	C	G	(C) 0.32	0.16	0.10	0.12	GMS + CGS all	$1.42 \times 10^{-5}$	$2.81 \times 10^{-7}$	0.032	$1.21 \times 10^{-7}$
rs4760506	12	91,857,181	<i>EEA1</i>	Suggestive	G	A	(A) 0.45	0.16	0.10	0.10	GMS + CGS all	$6.77 \times 10^{-6}$	$8.56 \times 10^{-7}$	0.594	$9.15 \times 10^{-5}$
rs12883884	14	69,586,936	<i>SLC8A3</i>	Suggestive	T	G	(G) 0.39	0.12	0.15	0.12	GMS + CGS all	$1.20 \times 10^{-6}$	$9.56 \times 10^{-7}$	0.223	$7.81 \times 10^{-6}$
rs12188164	5	481,236	<i>AHRR</i>	Suggestive	A	C	(A) 0.38	0.08	0.12	0.15	GMS + CGS all	$5.92 \times 10^{-4}$	$1.34 \times 10^{-6}$	0.136	$3.65 \times 10^{-6}$
rs11645366	16	60,934,654	<i>CDH8</i>	Suggestive	C	T	(T) 0.23	0.17	0.13	0.13	GMS + CGS all	$1.23 \times 10^{-5}$	$1.52 \times 10^{-6}$	0.182	$7.03 \times 10^{-6}$

Chr., chromosome.

<sup>a</sup>NCBI build 36. <sup>b</sup>Significant and suggestive imply  $P \leq (0.05/570,725) = 8.76 \times 10^{-8}$  or  $P \leq (1/570,725) = 1.75 \times 10^{-6}$ , respectively, for at least one analysis (GMS + CGS p.Phe508del/p.Phe508del or all GMS + CGS). <sup>c</sup>Alleles indexed to the forward strand of NCBI build 36; the risk allele is the allele associated with worse lung function. <sup>d</sup>Minor allele frequencies are listed for all GMS + CGS. Study-specific MAFs are provided in **Supplementary Table 1**. <sup>e</sup>Coefficients refer to the average reduction in the Consortium lung phenotype for each copy of the risk allele. <sup>f</sup>TSS direction-consistent association  $P$  value, for TSS p.Phe508del/p.Phe508del only or all TSS individuals, selected according to the GMS + CGS result with maximum significance. <sup>g</sup>Joint meta-analysis  $P$  value for GMS, CGS and TSS, with selection of subjects (p.Phe508del/p.Phe508del only or all subjects) according to the GMS + CGS result with maximum significance.

**Figure 3** A plot of the association evidence in GMS and CGS p.Phe508del/p.Phe508del individuals in the chromosome 11p13 *EHF-APIP* region (NCBI build 36, LocusZoom viewer). Colors represent HapMap CEU linkage disequilibrium  $r^2$  values with the most significant SNP, rs12793173 ( $P = 3.34 \times 10^{-8}$ ). The secondary peak at rs286873 has relatively low  $r^2$  with the primary peak.

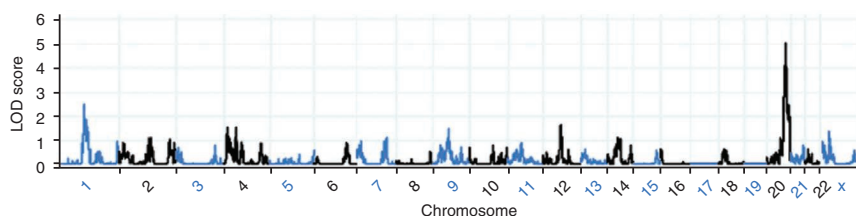


In the HLA class II region, a SNP (rs2395185, ~1 kb from the suggestive SNP rs9268905 identified from GMS and CGS) approached genome-wide significance using all affected subjects ( $P = 9.02 \times 10^{-8}$ ; **Supplementary Fig. 3**). SNPs in *AGTR2* remained suggestive for all affected subjects (rs5952206,  $P = 1.25 \times 10^{-7}$ ) and for individuals with p.Phe508del/p.Phe508del (rs7060450,  $P = 3.67 \times 10^{-7}$ ).

**Figure 3** shows the GMS and CGS results for an 800-kb interval including *EHF-APIP*. The minimum  $P$  value is in an intergenic region 3' to both *EHF* and *APIP*. A second peak at rs286873 ( $P = 5.62 \times 10^{-7}$ ) near *EHF* had low linkage disequilibrium ( $r^2 < 0.2$ ) with the primary SNP (**Fig. 3**). After conditioning on the primary finding, rs286873 had regional statistical significance (rs12793173; corrected  $P = 0.0029$ ), suggesting additional regional genetic variants (**Supplementary Fig. 4**). We repeated the testing after MACH imputation<sup>14</sup>. The imputed SNPs in the region identified the same *EHF-APIP* interval with a minimum  $P = 1.45 \times 10^{-8}$  at rs535719, which is at a position 19 kb closer to *APIP* than rs12793173. None of the imputed SNPs produced substantially improved association evidence (**Supplementary Fig. 5**). Neither the total copy number nor the allele-specific copy number (Online Methods) models met genome-wide significance (the illustrative Manhattan plot is shown in **Supplementary Fig. 6**). Finally, after sequencing the exonic regions of *EHF* and *APIP* in 48 individuals with mild pulmonary disease and 48 individuals with severe pulmonary disease from the GMS, we found no additional genetic variation that offered insight into putative modifying roles (data not shown).

#### Linkage of lung disease severity in cystic fibrosis analysis

Linkage analysis revealed a genome-wide significant multipoint log<sub>10</sub> odds (LOD) score of 5.03 at rs4811626 at 53.81 Mb (~85 cM) on chromosome 20q13.2 (nominal  $P = 7.9 \times 10^{-7}$ ; genome-wide<sup>15</sup>  $P = 2.3 \times 10^{-3}$ ; **Fig. 4**). Another, but more modest, linkage signal was on chromosome 1p22.21, with a multipoint LOD score of 2.48 for rs941031 at 91.07 Mb (119 cM). Inclusion of BMI  $z$  score (BMI-Z), an important covariate of cystic fibrosis 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 0.07 cM (0.13 Mb) from rs4811626; **Fig. 5**), whereas linkage on chromosome 1p22.21 decreased the LOD



**Figure 4** Genome-wide linkage scan for the Consortium lung phenotype of 486 sibling pairs in the family based TSS dataset adjusted for sex. We found a QTL with a genome-wide significant LOD of 5.03 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).

to 1.67. Thus, anthropometric measures are not major contributors to the linkage on 20q13.2 but may play a role on 1p22.21. We estimated that the quantitative trait locus (QTL) at 20q13.2 approaches 50% of the variation in lung function in the sibling pairs with cystic fibrosis (**Supplementary Fig. 7**); however, this estimate is highly likely to be biased upward because of winner's curse<sup>16</sup>.

We analyzed a 1.31-Mb region on 20q13.2, demarcated by 1 LOD unit below the maximum (when we used BMI-Z as a covariate), for association in the combined GMS and CGS samples. A 16-kb cluster of SNPs in high LD (rs6092179, rs6024437, rs8125625, rs6024454 and rs6024460;  $r^2 > 0.8$ ), located ~200 kb from *CBLN4*, generated the lowest  $P$  values in the combined GMS and CGS p.Phe508del/p.Phe508del samples (**Fig. 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 showed weak LD with the GMS and CGS cluster of SNPs. Imputation did not identify any SNPs with a lower  $P$  value for association than rs6024460 (**Supplementary Fig. 8**).

#### An FDR approach to combine association and linkage

To evaluate association and linkage in a single framework, we used linkage information to reprioritize genome-wide association using extensions of the false discovery rate (FDR)<sup>17</sup>, that is, the stratified FDR (SFDR)<sup>18</sup> and weighted FDR (WFDR)<sup>19</sup>. We (i) obtained linkage-weighted  $q$  values representing the combined evidence at each SNP and (ii) re-ranked GWAS results by linkage-weighted  $q$  values (Online Methods). Results are presented from the WFDR; we confirmed results using the SFDR (data not shown). We declared SNPs with  $q$  values less than 0.05 to be genome-wide significant (**Table 3**). SNPs in the *EHF-APIP* region on chromosome 11 were 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 were considerably decreased. The results presented in **Table 3** show that the linked SNPs on chromosome 20 are now top ranked genome wide, whereas they were ranked one hundred fifty-fourth or lower before incorporating the linkage information. The top-ranked SNP by the WFDR analysis was rs6092179 at 53.81 Mb on chromosome 20 (WFDR  $q$  value = 0.015; **Table 3**). rs6092179 is within an LD block containing four other SNPs (rs6024437, rs8125625, rs6024454 and rs6024460), all having association with cystic

**Table 3 Combined association and linkage weighted FDR  $q$  values and genome-wide ranks for SNPs with WFDR  $q$  values that are genome-wide significant ( $q < 0.05$ )**

Chr.	SNP	Base pair	GMS + CGS p.Phe508del/ p.Phe508del association $P$ value	FDR $q$ value <sup>a</sup>	FDR rank	WFDR $q$ value <sup>b</sup>	WFDR rank
11	rs10836312	34,767,019	$1.56 \times 10^{-6}$	0.0124	7	0.0383	16
11	rs12272777	34,807,478	$5.74 \times 10^{-8}$	0.008	3	0.0277	8
<b>11</b>	<b>rs12793173</b>	<b>34,790,780</b>	<b><math>3.34 \times 10^{-8}</math></b>	<b>0.008</b>	<b>1</b>	<b>0.0218</b>	<b>6</b>
11	rs525202	34,778,524	$1.34 \times 10^{-7}$	0.0124	6	0.0375	14
11	rs552627	34,783,256	$7.18 \times 10^{-8}$	0.008	5	0.0282	10
11	rs568529	34,799,082	$5.07 \times 10^{-8}$	0.008	2	0.0277	7
11	rs731727	34,814,410	$6.39 \times 10^{-8}$	0.008	4	0.0277	9
20	rs11907114	53,862,354	$2.79 \times 10^{-3}$	0.7852	1,976	0.0459	18
20	rs1326022	54,277,432	$1.16 \times 10^{-3}$	0.7349	865	0.0459	17
20	rs6024437	53,813,962	$1.61 \times 10^{-4}$	0.5029	175	0.015	2
20	rs6024454	53,826,840	$2.56 \times 10^{-4}$	0.5581	255	0.015	5
<b>20</b>	<b>rs6024460</b>	<b>53,828,948</b>	<b><math>1.34 \times 10^{-4}</math></b>	<b>0.484</b>	<b>154</b>	<b>0.015</b>	<b>1</b>
20	rs6092176	53,799,109	$1.51 \times 10^{-3}$	0.7553	1,116	0.0353	13
20	rs6092179	53,812,440	$1.93 \times 10^{-4}$	0.516	207	0.015	3
20	rs6098782	53,791,974	$1.84 \times 10^{-3}$	0.7615	1,348	0.0381	15
20	rs7265042	53,790,816	$1.14 \times 10^{-3}$	0.7349	854	0.0296	12
20	rs8125625	53,820,352	$2.49 \times 10^{-4}$	0.554	250	0.015	4
20	rs910668	53,794,753	$1.09 \times 10^{-3}$	0.7344	824	0.0296	11

Chr., chromosome.

<sup>a</sup>Benjamini-Hochberg approach based on association  $P$  values. <sup>b</sup>Weighted FDR using combined linkage information and association  $P$  values. Rows in bold indicate the top ranked SNPs before (rs12793173 on chromosome 11) and after incorporating linkage evidence (rs6024460 on chromosome 20).

fibrosis lung function and a  $q$  value  $< 0.05$ . A rank-based  $q$  value Manhattan plot shows that chromosome 11 and chromosome 20 both attained genome-wide significance (**Supplementary Fig. 9**).

## DISCUSSION

We identified two new loci containing genetic variants contributing to variation in lung function in individuals with cystic fibrosis. The success of this project reflected (i) coordinated analysis of three independent samples of the cystic fibrosis population (representing ~15% of all affected individuals in North America) where each study subject was characterized by the same quantitative measure of lung function; (ii) simultaneous genotyping of samples using a single platform that allowed for data cleaning using relatedness assessments and removal of poor quality genotypes based on parent-to-child transmission predictions; and (iii) analyzing for loci using both association and linkage, which can detect loci even in the presence of substantial allelic heterogeneity. Moreover, we garnered increased power from an extremes-of-phenotype sample, and a population-based sample allowed for the development of a phenotype with external validity.

The association at chromosome 11p13 is in an intergenic region 3' to *APIP* and *EHF* with regulatory features including (i) significant conservation across species, (ii) open chromatin (DNAse hypersensitivity and FAIRE-Seq) and (iii) DNAse hypersensitive patterns suggesting cell-type specificity (see URLs). The University of California at Los Angeles (UCLA) Gene Expression Tool (UGET; see URLs)<sup>20,21</sup> indicates correlation of expression of nearby genes, including strong correlation of *EHF* to *ELF5*, both of which are epithelial-specific

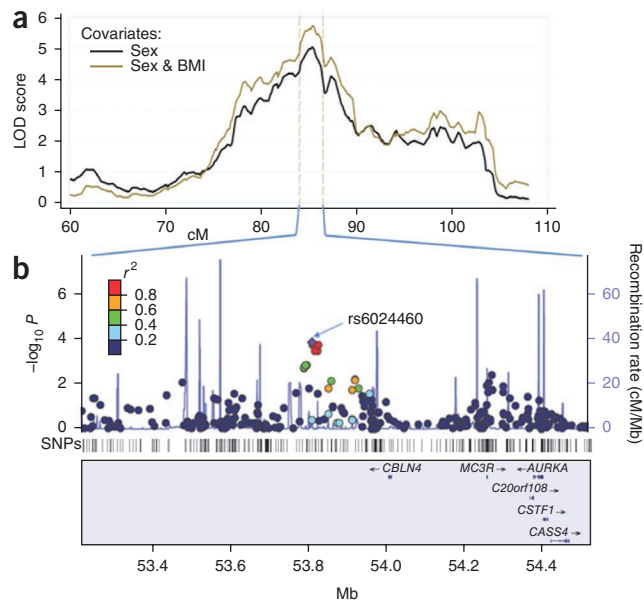
**Figure 5** Regional analysis of the QTL on chromosome 20q13.2.

(a) A detailed chromosome 20 linkage plot for the Consortium lung phenotype in the TSS study with the covariate sex (essentially the same result as for no covariates) and with the covariates sex and BMI. (b) Association evidence from the GMS and CGS p.Phe508del/p.Phe508del individuals in the 1-LOD support interval provided by TSS. A region centromeric to *CBLN4* and *MC3R* on 20q13.2 showed suggestive evidence of association, with the greatest evidence at rs6024460 ( $P = 1.34 \times 10^{-4}$ ).

transcription factors; *APIP* to *PDHX*, which have the same promoter region; and *EHF* to *APIP*. *APIP* (encoding the Apaf-1-interacting protein) is known to inhibit apoptosis by binding to APAF-1, an important activator of caspase-9 (refs. 22,23), and by APAF-1-independent activation of AKT and MAPK3 (ERK1) and MAPK1 (ERK2) (ref. 24). *EHF* is a member of epithelial-specific Ets transcription factors that share a conserved Ets domain<sup>25–27</sup>. *EHF* can be induced in bronchial epithelial cells, smooth muscle cells and fibroblasts<sup>28,29</sup>, leading to transcriptional repression of a subset of genes responsive to *ETS* and *AP-1* that are activated by MAP-kinase pathways<sup>26,28</sup>, and in the airway, it may serve as an important regulator of differentiation under conditions of stress and inflammation<sup>26,27</sup>. Both genes show evidence of robust expression in lung and trachea, with *APIP* showing ubiquitous expression across tissues and *EHF* showing the highest expression in trachea (see URLs)<sup>30</sup>. Notably, *cis* expression QTL (eQTL) signatures for *APIP* are reported for lymphocytes and monocytes

(see URLs). 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 cystic fibrosis lung disease. This hypothesis is consistent with the emerging concepts that delayed neutrophil clearance caused by reduced apoptosis in neutrophils in the airways of individuals with cystic fibrosis could lead to a hyper-inflammatory state and more severe lung disease<sup>31,32</sup> and that inhibition of apoptosis contributes to goblet cell metaplasia, a central feature in cystic fibrosis airway pathophysiology<sup>33</sup>.

All five genes within the one LOD support interval in the chromosome 20 linkage region (**Fig. 5**) are expressed in either fetal or adult lung or in bronchial epithelial cells (see URLs). The 16-kb cluster of SNPs associated with lung function in the GMS and CGS samples is



located ~200 kb to 500 kb centromeric to the five genes. None of the SNPs lies within a segment of open chromatin identified in the 16-kb region in normal human bronchial epithelia cells (see URLs). Neither eQTL in lymphocytes, miRNA (see URLs) nor DNaseI hypersensitive sites in small airway epithelial cells map to the 16-kb 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<sup>34–36</sup>. Variation in resting energy expenditure has been correlated with lung function measurements, lung tissue damage and lung disease exacerbation in individuals with cystic fibrosis<sup>37,38</sup>. *MC3R* has also been implicated as a modulator of neutrophil accumulation in a mouse model of lung inflammation<sup>39</sup>, a key feature of cystic fibrosis lung disease, as noted above. Other genes of interest within the linkage peak encode Crk-associated substrate scaffolding (*CASS*) 4 (encoded by *CASS4*, also known as *HEPL*), a relative of proteins implicated in cell attachment, migration establishing polarity, invasion and phagocytosis of bacterial pathogens<sup>40</sup>, and Aurora kinase A (encoded by *AURKA*), which has been shown to interact with Hef1, also known as NEDD9, a member of the *CASS* family that mediates cytokinesis in late mitosis and facilitates disassembly of primary cilia<sup>41</sup>.

Twin studies in adults show that FEV<sub>1</sub> is under strong genetic influence in the general population<sup>42,43</sup>, and at least three loci (*GSTCD*, *TNS1* and *HTR4*) have been reproducibly associated with this measure<sup>44–46</sup>. Multiple replicated loci have also been associated with variation in the ratio of FEV<sub>1</sub> to forced vital capacity (FVC)<sup>45,46</sup>, and at least two of these loci (*HHIP* and *FAM13*) showed reproducible association with COPD<sup>44,47,48</sup>. Although the lung phenotype used here was based on FEV<sub>1</sub>, none of the above loci coincides with the regions identified in this study, and neither of the loci identified here occur within the top 2,000 associations for FEV<sub>1</sub> or FEV<sub>1</sub>/FVC<sup>45,46</sup>.

Common variation in the *EHF-APIP* region is estimated to alter the lung function measure in the GMS and CGS individuals with p.Phe508del/p.Phe508del 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 predicted FEV<sub>1</sub> percent of  $5.1 \pm 1.9$ , corresponding to a mean difference in FEV<sub>1</sub> of  $254 \pm 86$  ml in affected subjects over 18 years of age (Online Methods). The QTL on chromosome 20 may account for a sizeable fraction of lung function variation in cystic fibrosis. Using simulations previously described<sup>16</sup>, we estimate that this locus accounts for a maximum of 46% and a minimum of 4% of the variance in the cystic fibrosis-affected 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 cystic fibrosis. 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 individuals with identical *CFTR* genotypes (that is, p.Phe508del/p.Phe508del), one of the seven suggestive loci achieved genome-wide significance despite the reduction in sample size because of 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 cystic fibrosis should provide new insight leading to the development of new therapies for this devastating condition.

**URLs.** UCSC Genome browser, <http://genome.ucsc.edu/>; UCLA gene expression tool, <http://genome.ucla.edu/~jdong/GeneCorr.html>; UniGene,

<http://www.ncbi.nlm.nih.gov/UniGene/>; NCBI, <http://www.ncbi.nlm.nih.gov/geo/>; eQTL signatures, [eqtl.uchicago.edu/](http://eqtl.uchicago.edu/); miRBase, <http://www.mirbase.org/>; MACH, <http://www.sph.umich.edu/csg/abecasis/mach/>; IMPUTE, <http://mathgen.stats.ox.ac.uk/impute/impute.html>; perl SFDR, <http://www.utstat.toronto.edu/sun/Software/SFDR/index.html>.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

*Note: Supplementary information is available on the Nature Genetics website.*

## ACKNOWLEDGMENTS

This work was supported in part by grants from the US National Heart, Lung and Blood Institute R01HL068927, R01HL068890, R01HL095396, the US National Institute of Diabetes and Digestive and Kidney Diseases R01DK066368, K23DK083551, P30DK027651, the US National Institute of Mental Health SOLAR-MH059490 and the US National Human Genome Research Institute HG-0004314; US Cystic Fibrosis Foundation grants CUTTIN00A0, CUTTIN06P0, COLLAC07A0, RDP-R025-CR07, KNOWLE00A0, RDP-R026-CR07, DRUMM0A00 and contract GENOMEQUEBEC07DDSO; Flight Attendant Medical Research Institute grant FAMRI2006; Lawson Wilkins Pediatric Endocrine Society grant LWPEP 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; and the Joint Fellowship of Canadian Institutes of Health Research and Ontario Women's Health Council. Funds for genome-wide genotyping were generously provided by the US Cystic Fibrosis Foundation (CFF).

The authors would like to thank the cases and families who participated in this study, 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:** P. Cornwall; **study design:** A. Hamosh, R. McWilliams; **recruitment:** S. Adams, M. Algire, N. Anderson, A. Bowers, J. Breaton, C. Bucur, L. Charmin, M. Christofi, B. Coleman, J. Dunn, B. Elkind, D. Frangolias, J. Hoover-Fong, K. Keenan, P. Miller, S. Norris, R. Rousseau, S. Wood, R. Yung and C. Yurk; **phenotyping:** N. Anderson, J. Breaton, M. Christofi, D. Frangolias, W. Ip, K. Keenan, R. Rousseau and R. Yung; **data entry:** J. Dunn, S. Norris and T. Lai; **genotyping:** K. Boden, R. Darrah, Q. Huang, K. Kanieki, F. Lin, S. Ritter, N. Wang, Y. Wang, C. Weiler, W. Wolf and X. Yuan; **analysis:** A. Dang, E. Hawbaker, L. Henderson, R. McWilliams, Y. Wang, C. Watson and A. Webel; **bioinformatics:** E. Crowdy, H. Dang, H. Kelkar, T. Randall and A. Xu.

**Contributing North American cystic fibrosis centers and principal investigators:** S. Aaron, Ottawa General Hospital, Ottawa, Ontario, Canada; F. Accurso, University of Colorado Health Sciences Center, Aurora, Colorado, USA; J. Acton, Cincinnati Children's Hospital and Medical Center, Cincinnati, Ohio, USA; R. Ahrens, University of Iowa Hospitals & Clinics, Iowa City, Iowa, USA; G. Aljaded, Lutheran General Children's Hospital, Park Ridge, Illinois, USA; C. Allard, Centre de Santé et de Services Sociaux de Chicoutimi, Chicoutimi, Quebec, Canada; R. Amaro, University of Texas at Tyler Health Center, Tyler, Texas, USA; R. Anbar, State University of New York (SUNY) Upstate Medical University, Syracuse, New York, USA; P. Anderson, University of Arkansas, Little Rock, Arkansas, USA; A. Atlas, Morristown Memorial Hospital, Morristown, New Jersey, USA; S. Bell, The Prince Charles Hospital, Brisbane, Queensland, Australia; M. Berdella, St. Vincent's Hospital & Medical Center, New York, New York, USA; J. Biller, Children's Hospital of Wisconsin, Milwaukee, Wisconsin, USA; H. Black, Asthma & Allergy Specialists, Charlotte, North Carolina, USA; P. Black, Children's Mercy Hospital, Kansas City, Missouri, USA; S. Boas, Children's Asthma Respiratory & Exercise Specialists, Glenview, Illinois, USA; M. Boland, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; D. Borowitz, Women's & Children's Hospital of Buffalo, Buffalo, New York, USA; R. Boswell, University of Tennessee, Memphis, Tennessee, USA; J. Boucher, Centre Hospitalier Régional de Rimouski, Rimouski, Quebec, Canada; C.M. Bowman, Medical University of South Carolina, Charleston, South Carolina, USA; M. Boyle, Johns Hopkins Hospital, Baltimore, Maryland, USA; C. Brown, California Pacific Medical Center, San Francisco, California, USA; D. Brown, Pediatric Pulmonary Associates, Columbia, South Carolina, USA; N. Brown, University of Alberta Hospitals, Edmonton, Alberta, Canada; L.F. Caffey, University of New Mexico, Albuquerque, New Mexico, USA; B. Chatfield, University of Utah, Salt Lake City, Utah, USA; S. Chesrown, University of Florida, Gainesville, Florida, USA; B. Chipps, Sutter Medical Center, Sacramento, California, USA; J.P. Clancy, University of Alabama at Birmingham, Birmingham,

Alabama, USA; R. Cohen, Kaiser Permanente, Portland, Oregon, USA; J. Colombo, University of Nebraska Medical Center, Omaha, Nebraska, USA; J. Cronin, Women & Children's Hospital of Buffalo, Buffalo, New York, USA; M. Cruz, St. Mary's Medical Center, West Palm Beach, Florida, USA; J. Cunningham, Cook Children's Medical Center, Fort Worth, Texas, USA; G. Davidson, BC Children's Hospital, Vancouver, British Columbia, Canada; L. Davies, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA; J. DeCelle-Germana, Schneider Children's Hospital, New Hyde Park, New York, USA; A. Devenny, Royal Hospital for Sick Children, Edinburgh, Scotland, UK; E. DiMango, Columbia University Medical Center, New York, New York, USA; D. Doornbos, Via-Christi St. Francis, Wichita, Kansas, USA; A. Dozor, New York Medical College-Westchester Medical Center, Valhalla, New York, USA; J. Dunitz, University of Minnesota, Minneapolis, Minnesota, USA; M. Egan, Yale University School of Medicine, New Haven, Connecticut, USA; J. Eichner, Great Falls Clinic, Great Falls, Montana, USA; T. Ferkol, St. Louis Children's Hospital, St. Louis, Missouri, USA; S. Fiel, Morristown Memorial Hospital, Morristown, New Jersey, USA; P. Flume, Medical University of South Carolina, Charleston, South Carolina, USA; A. Freitag, Hamilton Health Sciences Corporation, Hamilton, Ontario, Canada; M. Franco, Miami Children's Hospital, Miami, Florida, USA; D. Froh, University of Virginia Health System, Charlottesville, Virginia, USA; N. Garey, Saint John Regional Hospital, Saint John, New Brunswick, Canada; D. Geller, Nemours Children's Clinic, Orlando, Florida, USA; W. Gershan, Children's Hospital of Wisconsin, Milwaukee, Wisconsin, USA; R. Gibson, Children's Hospital & Regional Medical Center, Seattle, Washington, USA; R. Giusti, Long Island College Hospital, Brooklyn, New York, USA; J. Gjevre, Royal University Hospital, Saskatoon, Saskatchewan, Canada; M. Gondor, University of South Florida, St. Petersburg, Florida, USA; G. Gong, Phoenix Children's Hospital, Phoenix, Arizona, USA; M. Guill, Medical College of Georgia, Augusta, Georgia, USA; H. Gutierrez, University of Alabama at Birmingham, Birmingham, Alabama, USA; A. Hadeh, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA; K. Hardy, Children's Hospital, Oakland, California, USA; P. Hiatt, Texas Children's Hospital, Houston, Texas, USA; D. Hicks, Children's Hospital of Orange County, Orange, California, USA; B. Holmes, Regina General Hospital, Regina, Saskatchewan, Canada; D. Holsclaw, University of Pennsylvania, Philadelphia, Pennsylvania, USA; P. Holzwarth, St. Vincent Hospital, Green Bay, Wisconsin, USA; R. Honicky, Michigan State University, East Lansing, Michigan, USA; M. Howenstine, Riley Hospital for Children, Indianapolis, Indiana, USA; D. Hughes, IWK Health Centre, Halifax, Nova Scotia, Canada; M. Jackson, Grand River Hospital, Kitchener, Ontario, Canada; P. James, Lutheran Hospital, Fort Wayne, Indiana, USA; A. Jenneret, Hôtel Dieu de Montréal, Montréal, Quebec, Canada; P. Joseph, University of Cincinnati, Cincinnati, Ohio, USA; J. Kanga, University of Kentucky, Lexington, Kentucky, USA; M. Katz, Baylor College of Medicine, Houston, Texas, USA; S. Kent, Victoria General Hospital, Victoria, British Columbia, Canada; W. Kepron, Winnipeg Health Sciences Centre, Winnipeg, Manitoba, Canada; M. Knowles, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; P. Konig, University of Missouri, Columbia, Missouri, USA; M. Konstan, Case Western Reserve University, Cleveland, Ohio, USA; T. Kovesi, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; J. Kramer, Oklahoma Cystic Fibrosis Center, Tulsa, Oklahoma, USA; N. Kraynack, Children's Hospital Medical Center of Akron, Akron, Ohio, USA; V. Kumar, Hôpital Régional de Sudbury Regional Hospital, Sudbury, Ontario, Canada; T. Lahiri, Fletcher Allen Health Care, Burlington, Vermont, USA; C. Landon, Pediatric Diagnostic Center, Ventura, California, USA; L. Lands, Montréal Children's Hospital, Montréal, Quebec, Canada; C. Lapin, Connecticut Children's Medical Center, Hartford, Connecticut, USA; M. Larj, Wake Forest University Baptist Medical Center, Winston-Salem, North Carolina, USA; J. Ledbetter, TC Thompson Children's Hospital, Chattanooga, Tennessee, USA; R. Lee, Naval Medical Center, Portsmouth, Virginia, USA; M. Leigh, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; L. Lester, University of Chicago Children's Hospital, Chicago, Illinois, USA; T. Lever, Eastern Maine Medical Center, Bangor, Maine, USA; H. Levy, Children's Hospital Boston, Boston, Massachusetts, USA; A. Lieberthal, Kaiser Permanente Southern California, Panorama City, California, USA; T. Liou, University of Utah, Salt Lake City, Utah, USA; A. Lipton, National Naval Medical Center, Bethesda, Maryland, USA; B. Lyttle, Children's Hospital of Western Ontario, London, Ontario, Canada; B. Lothian, Royal University Hospital, Saskatoon, Saskatchewan, Canada; D. Lougheed, Hotel Dieu Hospital, Kingston, Ontario, Canada; K. Malhotra, Grand River Hospital, Kitchener, Ontario, Canada; J. Marcotte, Hôpital Sante-Justine, Montréal, Quebec, Canada; E. Matouk, Montréal Chest Institute, Montréal, Quebec, Canada; M. McCarthy, Providence Medical Center, Spokane, Washington, USA; S. McColley, Children's Memorial Hospital & Northwestern University, Chicago, Illinois, USA; K. McCoy, Nationwide Children's Hospital, Columbus, Ohio, USA; J. McNamara, Children's Hospitals and Clinics of Minnesota, Minneapolis, Minnesota, USA; R. Michael, Queen Elizabeth II Health Sciences Centre, Halifax, Nova Scotia, Canada; S. Miller, University of Mississippi Medical Center, Jackson, Mississippi, USA; M. Milot, Centre de Santé et de Services Sociaux de Chicoutimi,

Chicoutimi, Quebec, Canada; K. Moffett, West Virginia University, Morgantown, West Virginia, USA; M. Montgomery, Alberta Children's Hospital, Calgary, Alberta, Canada; P. Moore, Vanderbilt University Medical Center, Nashville, Tennessee, USA; W. Morgan, Tucson Cystic Fibrosis Center, Tucson, Arizona, USA; R. Morris, Janeway Children's Health & Rehabilitation, St. John's, Newfoundland and Labrador, Canada; M. Morse, Methodist Children's Hospital, San Antonio, Texas, USA; S. Moskowitz, Children's Hospital & Regional Medical Center, Seattle, Washington, USA; R. Moss, Stanford University Medical Center, Palo Alto, California, USA; P. Murphy, University of Nebraska Medical Center, Omaha, Nebraska, USA; E. Nakielna, St. Paul's Hospital, Vancouver, British Columbia, Canada; S. Nasr, University of Michigan Health System, Ann Arbor, Michigan, USA; L. Nassri, Sparks Regional Medical Center, Fort Smith, Arkansas, USA; E. Naureckas, University of Chicago Hospitals, Chicago, Illinois, USA; D. Nielson, University of California at San Francisco, San Francisco, California, USA; M. Noseworthy, Janeway Children's Health & Rehabilitation, St. John's, Newfoundland and Labrador, Canada; B. Noyes, St. Louis University, St. Louis, Missouri, USA; K. Olivier, Wilford Hall US Air Force Medical Center, San Antonio, Texas, USA; E. Olson, University of Florida, Gainesville, Florida, USA; G. Omlor, Akron Children's Hospital, Akron, Ohio, USA; D. Orenstein, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA; B. O'Sullivan, University of Massachusetts Memorial Health Care, Worcester, Massachusetts, USA; H.W. Parker, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA; M. Passero, Brown University Medical School Rhode Island Hospital, Providence, Rhode Island, USA; H. Pasterkamp, Children's Hospital of Winnipeg, Winnipeg, Manitoba, Canada; L. Pedder, Hamilton Health Sciences Corporation, Hamilton, Ontario, Canada; E. Perkett, Vanderbilt University Medical Center, Nashville, Tennessee, USA; G. Perry, University of Kansas Medical Center, Kansas City, Kansas, USA; N. Petit, Centre Hospitalier Rouyn-Noranda, Rouyn-Noranda, Quebec, Canada; M. Pian, University of California San Diego Children's Hospital, San Diego, California, USA; A. Platzker, Children's Hospital of Los Angeles, Los Angeles, California, USA; C. Prestidge, Children's Medical Center of Dallas, Dallas, Texas, USA; A. Price, Children's Hospital of Western Ontario, London, Ontario, Canada; H. Rabin, Foothills Medical Centre, Calgary, Alberta, Canada; P. Radford, Phoenix Children's Hospital, Phoenix, Arizona, USA; F. Ratjen, The Hospital for Sick Children, Toronto, Ontario, Canada; W. Regelmann, University of Minnesota, Minneapolis, Minnesota, USA; C. Ren, University of Rochester Medical Center, Strong Memorial Hospital, Rochester, New York, USA; G. Retsch-Bogart, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; W. Richards, Memphis Lung Physicians, Southaven, Mississippi, USA; M. Riva, Via-Christi, St. Francis Campus, Wichita, Kansas, USA; L. Rivard, Centre Universitaire de Santé de Lestrie, Sherbrooke, Quebec, Canada; D. Roberts, Providence Medical Center, Anchorage, Alaska, USA; M. Rock, University of Wisconsin Hospital, Madison, Wisconsin, USA; J. Rosen, Albany Medical College, Albany, New York, USA; J. Royall, Children's Hospital of Oklahoma, Oklahoma City, Oklahoma, USA; R. Rubenstein, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; F. Ruiz, University of Mississippi Medical Center, Jackson, Mississippi, USA; T. Scanlin, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; M. Schechter, Emory University, Atlanta, Georgia, USA; H.J. Schmidt, Virginia Commonwealth University, Richmond, Virginia, USA; M. Schwartzman, Joe DiMaggio Children's Hospital, Hollywood, Florida, USA; P. Scott, Georgia Pediatric Pulmonology Associates, PC, Atlanta, Georgia, USA; G. Shay, Kaiser Permanente Medical Center, Oakland, California, USA; R. Simon, University of Michigan Health System, East Lansing, Michigan, USA; P. Smith, Long Island College Hospital, Brooklyn, New York, USA; M. Solomon, The Hospital for Sick Children, Toronto, Ontario, Canada; T. Spencer, Children's Hospital of Boston, Boston, Massachusetts, USA; A. Stecenko, Emory University, Atlanta, Georgia, USA; D. Stokes, University of Tennessee, Memphis, Tennessee, USA; B. Sullivan, Marshfield Clinic, Marshfield, Wisconsin, USA; J. Taylor-Cousar, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA; N. Thomas, Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA; H. Thompson, St. Luke's Cystic Fibrosis Clinic, Boise, Idaho, USA; D. Toder, Children's Hospital of Michigan and Harper University Hospital, Detroit, Michigan, USA; E. Tullis, St. Michael's Hospital, Toronto, Ontario, Canada; N. Turcios, University of Medicine & Dentistry of New Jersey, New Brunswick, New Jersey, USA; R. van Wylick, Hotel Dieu Hospital, Kingston, Ontario, Canada; L. Varlotta, St. Christopher's Hospital for Children, Philadelphia, Pennsylvania, USA; P. Vauthy, Toledo Children's Hospital, Toledo, Ohio, USA; J. Voynow, Duke University, Durham, North Carolina, USA; C. Wainwright, Royal Children's Hospital, Melbourne, Victoria, Australia; P. Walker, St. Vincent's Hospital, New York, New York, USA; W.S. Warren, Hershey Medical Center, Hershey, Pennsylvania, USA; P. Wilcox, St. Paul's Hospital, British Columbia, Canada; R. Willmott, St. Louis University, St. Louis, Missouri, USA; H. Wojtczak, Naval Medical Center, Poway, California,

USA; W. Yee, New England Medical Center, Boston, Massachusetts, USA; C. Zacher, St. Alexius Heart & Lung Cystic Fibrosis Clinic, Bismarck, North Dakota, USA; R. Zanni, Monmouth Medical Center, Long Branch, New Jersey, USA; P. Zeitlin, Johns Hopkins Hospital, Baltimore, Maryland, USA; P. Zuberbuhler, University of Alberta Hospitals, Edmonton, Alberta, Canada.

#### AUTHOR CONTRIBUTIONS

FA.W., L.J.S., L.S., D.C., M.C., R.D., L.L.V., W.K.O., M.L.D., P.R.D., 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. FA.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. and G.R.C. performed genotyping and data cleaning. FA.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.W.K., E.M.L., S.L., W.L., G.M.M., J.M.R., W.S., C.T., F.Z. and J.B. performed statistical analysis. FA.W., L.J.S., V.K.D., L.S., R.D., J.M.R., W.K.O., M.L.D., P.R.D., M.R.K. and G.R.C. wrote the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- The Cystic Fibrosis Genotype-Phenotype Consortium. Correlation between genotype and phenotype in patients with cystic fibrosis. *N. Engl. J. Med.* **329**, 1308–1313 (1993).
- Vanscoy, L.L. *et al.* Heritability of lung disease severity in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **175**, 1036–1043 (2007).
- Cutting, G.R. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Ann. NY Acad. Sci.* **1214**, 57–69 (2010).
- Chalmers, J.D., Fleming, G.B., Hill, A.T. & Kilpatrick, D.C. Impact of mannose binding lectin (MBL) insufficiency on the course of cystic fibrosis: a review and meta-analysis. *Glycobiology* **21**, 271–282 (2010).
- Corey, M., Edwards, L., Levison, H. & Knowles, M. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J. Pediatr.* **131**, 809–814 (1997).
- Schluchter, M.D., Konstan, M.W. & Davis, P.B. Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients. *Stat. Med.* **21**, 1271–1287 (2002).
- Kulich, M. *et al.* Disease-specific reference equations for lung function in patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **172**, 885–891 (2005).
- Taylor, C. *et al.* A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. *Pediatr. Pulmonol.* published online, doi:10.1002/ppul.21456 (1 April 2011).
- Drumm, M.L. *et al.* Gene modifiers of lung disease in cystic fibrosis. *N. Engl. J. Med.* **353**, 1443–1453 (2005).
- Dorfman, R. *et al.* Complex two-gene modulation of lung disease severity in children with cystic fibrosis. *J. Clin. Invest.* **118**, 1040–1049 (2008).
- Li, W. *et al.* Understanding the population structure of North American patients with cystic fibrosis. *Clin. Genet.* **79**, 136–146 (2010).
- Chen, W.M. & Abecasis, G.R. Family-based association tests for genome-wide association scans. *Am. J. Hum. Genet.* **81**, 913–926 (2007).
- Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
- Li, Y., Ding, J. & Mach Abecasis, G.R. 1.0: rapid haplotype reconstruction and missing genotype inference. *American Society of Human Genetics Annual Meeting — 56th Annual Meeting* **416** (2006).
- Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**, 241–247 (1995).
- Görling, H.H., Terwilliger, J.D. & Blangero, J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.* **69**, 1357–1369 (2001).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
- Sun, L., Craiu, R.V., Paterson, A.D. & Bull, S.B. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genet. Epidemiol.* **30**, 519–530 (2006).
- Roeder, K., Bacanu, S.A., Wasserman, L. & Devlin, B. Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* **78**, 243–252 (2006).
- Day, A., Carlson, M.R., Dong, J., O'Connor, B.D. & Nelson, S.F. Celsius: a community resource for Affymetrix microarray data. *Genome Biol.* **8**, R112 (2007).
- Day, A. *et al.* Disease gene characterization through large-scale co-expression analysis. *PLoS ONE* **4**, e8491 (2009).
- Cao, G. *et al.* Cloning of a novel Apaf-1-interacting protein: a potent suppressor of apoptosis and ischemic neuronal cell death. *J. Neurosci.* **24**, 6189–6201 (2004).
- Cho, D.H. *et al.* Induced inhibition of ischemic/hypoxic injury by AP1, a novel Apaf-1-interacting protein. *J. Biol. Chem.* **279**, 39942–39950 (2004).
- Cho, D.H. *et al.* Suppression of hypoxic cell death by AP1-induced sustained activation of AKT and ERK1/2. *Oncogene* **26**, 2809–2814 (2007).
- Oikawa, T. & Yamada, T. Molecular biology of the Ets family of transcription factors. *Gene* **303**, 11–34 (2008).
- 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.* **276**, 20397–20406 (2001).
- 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.* **275**, 2986–2998 (2000).
- Silverman, E.S. *et al.* Constitutive and cytokine-induced expression of the ETS transcription factor ESE-3 in the lung. *Am. J. Respir. Cell Mol. Biol.* **27**, 697–704 (2002).
- 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.* **18**, 649–663 (2008).
- Dezso, Z. *et al.* A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol.* **6**, 49 (2008).
- 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. USA* **96**, 13330–13335 (1999).
- McKeon, D.J. *et al.* Prolonged survival of neutrophils from patients with Delta F508 CFTR mutations. *Thorax* **63**, 660–661 (2008).
- Harris, J.F. *et al.* Bcl-2 sustains increased mucous and epithelial cell numbers in metaplastic airway epithelium. *Am. J. Respir. Crit. Care Med.* **171**, 764–772 (2005).
- Butler, A.A. The melanocortin system and energy balance. *Peptides* **27**, 281–290 (2002).
- Lee, Y.S., Poh, L.K. & Loke, K.Y. A novel melanocortin 3 receptor gene (*MC3R*) mutation associated with severe obesity. *J. Clin. Endocrinol. Metab.* **87**, 1423–1426 (2002).
- Savastano, D.M. *et al.* Energy intake and energy expenditure among children with polymorphisms of the melanocortin-3 receptor 1–4. *Am. J. Clin. Nutr.* **90**, 912–920 (2009).
- Dorlöchter, L., Roksund, O., Helgheim, V., Rosendahl, K. & Fluge, G. Resting energy expenditure and lung disease in cystic fibrosis. *J. Cyst. Fibros.* **1**, 131–136 (2002).
- McCloskey, M. *et al.* Energy balance in cystic fibrosis when stable and during a respiratory exacerbation. *Clin. Nutr.* **23**, 1405–1412 (2004).
- Gettling, S.J. *et al.* A role for MC3R in modulating lung inflammation. *Pulm. Pharmacol. Ther.* **21**, 866–873 (2008).
- Tikhmyanova, N. & Little, J. CAS proteins in normal and pathological cell growth control. *Cell. Mol. Life Sci.* **67**, 1025–1048 (2010).
- Pugacheva, E.N., Jablonski, S.A., Hartman, T.R., Henske, E.P. & Golemis, E.A. HEF1-dependent aurora A activation induces disassembly of the primary cilium. *Cell* **129**, 1351–1363 (2007).
- Hubert, H.B., Fabsitz, R.R., Feinleib, M. & Gwinn, C. Genetic and environmental influences on pulmonary function in adult twins. *Am. Rev. Respir. Dis.* **125**, 409–415 (1982).
- McClearn, G.E., Svartengren, M., Pedersen, N.L., Heller, D.A. & Plomin, R. Genetic and environmental influences on pulmonary function in aging Swedish twins. *J. Gerontol.* **49**, 264–268 (1994).
- Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham heart study. *PLoS Genet.* **5**, e1000429 (2009).
- Hancock, D.B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* **42**, 45–52 (2010).
- Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* **42**, 36–44 (2010).
- Pillai, S.G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* **5**, e1000421 (2009).
- Cho, M.H. *et al.* Variants in *FAM13A* are associated with chronic obstructive pulmonary disease. *Nat. Genet.* **42**, 200–202 (2010).

## 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 96-well plates with Centre d'Etude du Polymorphisme Humain (CEPH) and one replicate control per plate. Illumina BeadStudio was used to call genotype, and identity was 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 Y chromosome SNPs and 138 mitochondrial SNPs. Duplicate discordance was 0.004% in GMS and was 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 s.d. 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 affected samples were excluded (in GMS, 6 were excluded, in CGS, 17 were excluded, and in TSS 5 were excluded) because of genotyping failure or artifacts, two GMS samples were excluded because of outlying ancestry (by principal component analysis), and eight GMS samples were excluded for greater than second degree relation with other samples. Reported findings were verified using the 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 p.Phe508del/p.Phe508del using an additive model in PLINK v. 1.07 (ref. 49) adjusted for sex and genotype principal components<sup>11</sup>. Using the PLINK  $z$  statistics for GMS and CGS, the standard meta-analysis  $z$  statistic<sup>50</sup> 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  $1/(\text{number of SNPs}) = 1/570,725 = 1.75 \times 10^{-6}$  and significant association used 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 of 0.05 were  $P = 1.07 \times 10^{-7}$  (GMS and CGS) and  $P = 1.05 \times 10^{-7}$  (GMS and CGS p.Phe508del/p.Phe508del). Consequently,  $P < 5 \times 10^{-8}$  achieved a 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.

The TSS association analysis was performed in 973 siblings with cystic fibrosis and for the 557-individual p.Phe508del/p.Phe508del subset using the MERLIN variance-components additive model framework<sup>12</sup> corrected for linkage, family structure, sex and four principal components. Missing genotypes (0.125%) were inferred to increase power<sup>51</sup>. Joint analyses of GMS, CGS and TSS used the meta-analysis approach as described above.

**A combined conditional likelihood approach.** We devised a new 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 principal components. 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

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

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  $2 \times (\log \text{likelihood ratio})$ , with  $\beta_1=0$  as the null and compared to  $\chi^2_1$ . 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 an effect  $\beta_1$  on the average phenotype for each minor allele. The results for GMS and CGS p.Phe508del/p.Phe508del are shown in **Supplementary Fig. 1**. For each simulation, the weighted meta-analysis  $P$  values were compared to  $P = 5 \times 10^{-8}$ .

**Genotype imputation.** MACH (autosomes) and IMPUTE (chromosome X; see URLs) imputation was conducted for 1,162 individuals with GMS, 1,254 self-reported Caucasian individuals with CGS and 60 CEU reference samples from HapMap I and 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 X chromosome SNPs.

**Copy-number analysis.** Copy number variants (CNVs) were detected using pennCNV (2008 November 19 version)<sup>52</sup> and genoCNV (version 1.08)<sup>53</sup> using default parameters in 1,103 GMS and 1,301 CGS samples. CNVs with fewer than five probes or showing <1% variation were used, resulting in 3,008 and 4,868 probes from genoCNV and pennCNV, respectively, in GMS and 3,015 and 4,663 probes for genoCNV and pennCNV, respectively, in CGS. Genotype principal components were used to control stratification.

**Linkage marker selection.** We selected 19,566 SNPs from the Illumina platform with minor allele frequency >0.4 and  $r^2 < 0.01$  between adjacent SNPs using MERLIN<sup>54</sup>. HapMap II recombination data were used to integrate genetic and physical map positions. The 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 contents were ~0.9 (multipoint) and ~0.31 (two-point).

**Linkage analysis.** Variance components were estimated in SOLAR (Sequential Oligogenic Linkage Analysis Routines)<sup>55</sup>, with similar results from MERLIN<sup>54</sup>, using multipoint identity-by-descent probabilities obtained from MERLIN. LOD scores were computed with and without covariates (sex and average BMI-Z). Multipoint LODs greater than 2.0 were considered suggestive, and LODs greater than 3.7 were considered genome-wide significant<sup>15</sup>.

**WFDR and SFDR methods.** Let  $P_i$  be the  $P$  value of an association test for SNP  $i$ ,  $i = 1, \dots, m$ . Converting  $P$  values to  $q$  values<sup>56</sup> controls the FDR. SNPs with  $q$  values less than the FDR threshold value ( $\gamma = 0.05$ ) were declared significant. The expected proportion of false positives among all the positives was then controlled at level  $\gamma$ . 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 were divided into  $K$  disjoint strata based on the prior linkage information<sup>57</sup>. Consider  $K = 2$  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<sup>15</sup>).  $Q$  values were then calculated separately for each stratum of SNPs, achieving FDR control in each stratum<sup>18</sup>. 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  $\bar{W} = \sum_i W_i / m = 1$ . The weight  $W_i$  is proportional to the linkage signal  $Z_i$  for SNP  $i$  (for example,  $W_i = \exp(B \cdot Z_i) / v$ ,  $v = \sum_i \exp(B \cdot Z_i) / m$ , and  $B = 1$ )<sup>19</sup>, and the FDR procedure was applied to the set of weight-adjusted  $P$  values,  $P_i/W_i$ ,  $i = 1, \dots, m$ . We use  $B = 2$  in the present analysis. The WFDR and SFDR were implemented in a perl program called SFDR (see URLs).

**Phenotype variation attributable to association and linkage.** The proportion of variation caused by each SNP was measured as the change in regression sums of squares versus the smaller model with the SNP removed<sup>58</sup>. Using the genome-scan threshold of  $P = 5 \times 10^{-8}$  and minimum  $P = 3.34 \times 10^{-8}$  in the chromosome 11p13 region for GMS and CGS individuals with p.Phe508del/p.Phe508del, we estimated a 57.4% reduction in effect size compared to the

nominal result. Using the joint analysis based on GMS, CGS p.Phe508del/p.Phe508del and TSS p.Phe508del/p.Phe508del subjects, the observed minimum  $P = 8.28 \times 10^{-8}$  resulted in ~28.0% reduction of the effect size. Using the rough parallel to explained variation in the trait, the estimated explained variation for 11p13 remained 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 0 as the true (unmeasured) heritability attributable to the QTL increases<sup>16</sup>. Although it is not possible to quantify the magnitude of this bias in this single study, these calculations provide an upper bound on the bias of 0.38–0.46 and a lower bound of 0.04–0.12.

**Estimation of changes in the cystic fibrosis lung phenotype upon predicted FEV<sub>1</sub> percent and airway flow.** Using 973 individuals from TSS, 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 FEV<sub>1</sub> (in liters) was then back extrapolated<sup>8</sup>, and FEV<sub>1</sub> percent predicted values were generated using the predictive equations<sup>59,60</sup>. Height and age adjustments used to calculate the original quantitative lung phenotype were preserved. The average increase (mean  $\pm$  s.d.) in FEV<sub>1</sub> percent predicted corresponding to a 0.2-unit increase of our lung phenotype was  $5.09\% \pm 1.90\%$  ( $n = 841$ , range 0.00–14.53%). The corresponding average increase in raw FEV<sub>1</sub> was  $253.5 \pm 85.9$  ml in adult subjects (defined as >18 years old) ( $n = 244$ , range: 0.0–630.0 ml).

49. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
50. van Houwelingen, H.C., Arends, L.R. & Stijnen, T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat. Med.* **21**, 589–624 (2002).
51. Burdick, J.T., Chen, W.M., Abecasis, G.R. & Cheung, V.G. *In silico* method for inferring genotypes in pedigrees. *Nat. Genet.* **38**, 1002–1004 (2006).
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.* **17**, 1665–1674 (2007).
53. Sun, W. *et al.* Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res.* **37**, 5365–5377 (2009).
54. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
55. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
56. Storey, J.D. A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002).
57. Yoo, Y.J., Bull, S.B., Paterson, A.D., Waggott, D. & Sun, L. Were genome-wide linkage studies a waste of time? Exploiting candidate regions within genome-wide association studies. *Genet. Epidemiol.* **34**, 107–118 (2010).
58. Neter, J., Wasserman, W., & Kutner, M. *Applied Linear Regression Models* 282 (Richard D. Irwin, Homewood, Illinois, USA, 1989).
59. Hankinson, J.L., Odencrantz, J.R. & Fedan, K.B. Spirometric reference values from a sample of the general U.S. population. *Am. J. Respir. Crit. Care Med.* **159**, 179–187 (1999).
60. Wang, X., Dockery, D.W., Wypij, D., Fay, M.E. & Ferris, B.G. Jr. Pulmonary function between 6 and 18 years of age. *Pediatr. Pulmonol.* **15**, 75–88 (1993).

