Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis

Lei Sun1,2,18, Johanna M Rommens3,4,18, Harriet Corvol5,6, Weili Li1,7, Xin Li7, Theodore A Chiang7, Fan Lin3, Ruslan Dorfman3, Pierre-François Busson8, Rashmi V Parekh3, Diana Zelenika9, Scott M Blackman10, Mary Corey7, Vishal K Doshi11, Lindsay Henderson11, Kathleen M Naughton11, Wanda K O’Neal12, Rhonda G Pace1,2, Jaclyn R Stonebraker1,2, Sally D Wood12, Fred A Wright13, Julian Zielenski3, Annick Clement5,6, Mitchell L Drumm14,15, Pierre-Yves Boëlle6,8, Garry R Cutting10,11, Michael R Knowles12, Peter R Durie16,17, and Lisa J Strug1,7

Variants associated with meconium ileus in cystic fibrosis were identified in 3,763 affected individuals by genome-wide association study (GWAS). Five SNPs at two loci near SLC6A14 at Xq23.24 (minimum \( P = 1.28 \times 10^{-12} \) at rs3788766) and SLC26A9 at 1q32.1 (minimum \( P = 9.88 \times 10^{-9} \) at rs4077468) accounted for ~5% of phenotypic variability and were replicated in an independent sample of affected individuals (\( n = 2,372; P = 0.001 \) and 0.0001, respectively). By incorporating the knowledge that disease-causing mutations in CFTR alter electrolyte and fluid flux across surface epithelium into a hypothesis-driven GWAS (GWAS-HD), we identified associations with the same SNPs in SLC6A14 and SLC26A9 and established evidence for the involvement of SNPs in a third solute carrier gene, SLC9A3. In addition, GWAS-HD provided evidence of association between meconium ileus and multiple genes encoding constituents of the apical plasma membrane where CFTR resides (\( P = 0.0002 \); testing of 150 apical membrane genes jointly and in replication, \( P = 0.022 \)). These findings suggest that modulating activities of apical membrane constituents could complement current therapeutic paradigms for cystic fibrosis.

Cystic fibrosis is caused by mutations in the CFTR gene encoding the cystic fibrosis transmembrane conductance regulator1. The CFTR protein is a chloride channel located on the apical membrane of the surface epithelium, where ion conduction and solute trafficking contribute to the regulation of transepithelial fluid flow. Individuals with the same loss-of-function CFTR mutations have variable disease severity and different cystic fibrosis–associated effects on organs, including the lung, pancreas, liver, intestine and vas deferens; thus, additional features, including variation in other genes (referred to as modifier genes), may affect disease pathophysiology. Approximately 15% of individuals with cystic fibrosis have severe intestinal obstruction at birth, a complication known as meconium ileus2. Meconium ileus develops in utero and presents following birth with complete intestinal obstruction that requires either enema or surgical intervention. This neonatal complication is highly indicative of cystic fibrosis, occurs in either sex, shows notable heritability exceeding 88% (ref. 3) and is likely minimally affected by environmental influences.

The North American Cystic Fibrosis Gene Modifier Consortium has accumulated 3,763 participants with severe (pancreatic exocrine–insufficient) CFTR alleles and genome-wide genotype data at 543,927 SNPs4,5 (Table 1 and Online Methods). The definition of meconium ileus was consistent within the consortium, and disease status for each subject was recorded following rigorous chart review. A conventional GWAS for meconium ileus used a generalized estimating equations (GEE) model6 to include collected sibling pairs and led to the identification of five SNPs that associated with genome-wide significance (\( P < 5 \times 10^{-8} \)) from two regions that include SLC26A9 on chromosome 1 and SLC6A14 on the X chromosome (Fig. 1, Table 2 and Supplementary Fig. 1).

Received 25 July 2011; accepted 24 February 2012; published online 1 April 2012; doi:10.1038/ng.2221
sex-specific results are given in Supplementary Table 1). CFTR genotype (encoding p.Phe508del) was not a significant confounder or effect modifier when incorporated into the GWAS (Supplementary Fig. 2 and Supplementary Table 2). We then replicated the associations at SLC6A14 (minimum P = 0.001) and SLC26A9 (minimum P = 0.0001) with meconium ileus in an independent combined collection from North America and France (Table 2).

The signal intensity plots of the associated SNPs reflected autosomal and X-associated SNPs at SLC26A9 and SLC6A14, respectively. Imputation analysis using MACH and minimac8,9 identified the same regions of association for the genotyped SNPs (Online Methods and Supplementary Fig. 3). The five associated SNPs in SLC6A14 and two in SLC26A9 (Fig. 1b, c) are positioned just upstream of the respective transcription start sites, such that binding of activating or repressing transcription factors might be affected, as highlighted by ENCODE data10 (data not shown). Neither SLC6A14 nor SLC26A9 coding regions showed evidence of copy-number variants (CNVs); however, there was a gap in the reference sequence more than 10 kb upstream of the SLC26A9 locus.

The seven SNPs genotyped in the two genes (Table 2) accounted for <5% of the meconium ileus variation, as estimated by pseudo R² (ref. 11), likely reflecting the common problems in association studies of locus heterogeneity and low power, given the available sample size. Whereas a conventional GWAS is often undertaken for complex disease mapping, modifier gene studies could incorporate disease etiology and pathobiology information to increase power and

Table 1 Participants in the cystic fibrosis consortia of the meconium ileus study

| Consortium study                                      | Sample size | CFTR genotype | Status of cystic fibrosis subjects |
|------------------------------------------------------|-------------|---------------|-----------------------------------|
| CGS (Canadian Gene Modifier Study, population-based design) | 1,661       | 992 (59.7)    | MI n (M:F) = 252 (128:124)        |
| GMS–Lung Study (University of North Carolina/Case Western Reserve University Extremes of Lung Phenotype Study) | 1,120       | 1,116 (99.6)  | Non-MI n (M:F) = 943 (498:445)    |
| GMS–Liver Study (North American samples only)         | 80          | 58 (72.5)     | MI n (M:F) = 23 (14:9)            |
| TSS (Johns Hopkins University Family-Based Twin and Sibling Study) | 902          | 522 (57.9)    | Non-MI n (M:F) = 57 (35:22)       |
| Total                                                 | 3,763       | 2,688 (71.4)  | MI n (M:F) = 611 (319:292)        |

*Size of cystic fibrosis subject sample used for the meconium ileus GWAS and GWAS-HD discovery study is given by consortium site. The TSS contained sibling pairs. There were 3,199 unrelated individuals among the total of 3,763 studied. aMeconium ileus (MI) status is given separately for each center. bGender breakdown is indicated for each subject category.

Figure 1 Meconium ileus GWAS identifies SNPs associated with genome-wide significance. Association analysis was performed on all SNPs with MAF > 2% that passed quality control criteria. (a) Genome-wide Manhattan plot for meconium ileus. The solid black line corresponds to the genome-wide significance threshold (P < 5 × 10⁻⁸) and the dotted black line to the suggestive association threshold, expected once per genome scan (P < 1/543,927 = 1.84 × 10⁻⁸). A total of five SNPs in two regions (SLC6A14 on chromosome X and SLC26A9 on chromosome 1) had association evidence exceeding the genome-wide threshold. The SNPs rs4077468 and rs4077469 are in perfect LD and appear as one SNP, as they are separated by only 128 bp. (b) Regional plot for SLC26A9. LocusZoom viewer was used to show association evidence around SLC26A9 based on NCBI36/hg18. Symbol coloring reflects LD r² values of the HapMap Utah residents of Northern and Western European ancestry (CEU) sample with the most significant SNP. The significant SNP rs4077468 is 2.17 kb upstream of the transcription start site. A gap is present in the genomic sequence between the SLC26A9 and FAM72A genes in both the NCBI36.3 and GRCh37 primary reference assemblies. (c) Regional plot for SLC6A14. The association evidence around SLC6A14 is shown. rs3788766 is 0.95 kb upstream of the transcription start site and is within the binding site of the CEBPB transcription factor, as annotated by ENCODE (data not shown). The mRNA transcript corresponding to CXorf61 is downstream of SLC6A14.
account for heterogeneity. We therefore proposed a GWAS that took a hypothesis into consideration, which therefore is hypothesis driven (GWAS-HD), to systematically prioritize SNPs for genome-wide analysis. The highest priority markers were also evaluated as a set to test the statistical significance of the hypothesis used for prioritization (Supplementary Fig. 4).

GWAS-HD prioritization in the context of cystic fibrosis was based on the knowledge that a major source of pathophysiology in cystic fibrosis is impaired fluid and electrolyte flux in the epithelium of cystic fibrosis–affected organs. The polarized epithelial layer forms a highly selective barrier for organ surface and ductal linings. Transepithelial function is achieved by cell polarization, wherein many determinants and regulators of fluid, solute and ion transport reside at the apical membrane alongside CFTR, with other features contributing from the basolateral surface. We have shown in a mouse model that CFTR function in the gastrointestinal epithelium is critical for preventing intestinal obstructions

A list of 157 gene products (Fig. 2 and Supplementary Table 3) was annotated to be localized to the apical plasma membrane using AmiGO

To implement the GWAS-HD for meconium ileus using the apical membrane hypothesis, we first prioritized the genome-wide markers by assigning the 3,814 SNPs of the apical membrane genes to a high-priority group and all remaining genome-wide SNPs to a low-priority group. We then performed two statistical analyses (Online Methods and Supplementary Fig. 4). In the first, which was analogous to a conventional GWAS, we conducted single-SNP association analysis using all of the 543,927 GWAS SNPs at the genome-wide level, but we up-weighted the 3,814 apical membrane SNPs via the stratified false discovery rate (SFDR) control

In the second analysis, which focused on only the 3,814 high-priority SNPs, we used multi-SNP analysis to test the prioritization hypothesis itself to determine whether multiple proteins present on the apical plasma membrane contribute to meconium ileus susceptibility.

As in the conventional GWAS, SNPs from SLC6A14 showed the strongest evidence for association with meconium ileus in the single-SNP GWAS-HD analysis (Supplementary Fig. 5), despite not being in the high-priority group, reflecting the robustness of whether multiple proteins present on the apical plasma membrane contribute to meconium ileus susceptibility.

Table 2 Discovery and replication of SNPs in SLC26A9 and SLC6A14

| SNP      | Chr. | Position (bp) | Gene           | Risk allele frequency | Discovery sample (611:3152) | Replication sample
|----------|------|---------------|-----------------|----------------------|-----------------------------|----------------------|
|          |      |               |                 | MI | Non-MI | OR | P     | OR | P     | OR | P     |
| rs4077468 | 1    | 204181380     | SLC26A9         | T  | 0.66 | 0.57 | 1.45 | 9.88 × 10⁻⁹ | 1.45 | 0.0005 | 1.27 | 0.0575 | 1.37 | 0.0001 |
| rs7512462 | 1    | 204166312     | SLC26A9         | T  | 0.66 | 0.57 | 1.45 | 2.14 × 10⁻⁸ | 1.31 | 0.0134 | 1.20 | 0.2120 | 1.27 | 0.0063 |
| rs7419153 | 1    | 204183932     | SLC26A9         | T  | 0.44 | 0.36 | 1.42 | 1.01 × 10⁻⁷ | 1.42 | 0.0007 | 1.20 | 0.1290 | 1.33 | 0.0004 |
| rs12047830| 1    | 204183322     | SLC26A9         | C  | 0.56 | 0.49 | 1.34 | 3.72 × 10⁻⁶ | 1.33 | 0.0054 | 1.27 | 0.0510 | 1.31 | 0.0007 |
| rs3788766 | X    | 115480867     | SLC6A14         | T  | 0.72 | 0.59 | 1.50 | 1.28 × 10⁻¹² | 1.14 | 0.1328 | 1.47 | 0.0006 | 1.25 | 0.0011 |
| rs5905283 | X    | 115479909     | SLC6A14         | C  | 0.61 | 0.50 | 1.34 | 1.69 × 10⁻⁸ | 1.04 | 0.6030 | 1.47 | 0.0002 | 1.19 | 0.0074 |
| rs12839137| X    | 115479578     | SLC6A14         | C  | 0.82 | 0.75 | 1.39 | 1.20 × 10⁻⁶ | 1.20 | 0.0666 | 1.15 | 0.3275 | 1.18 | 0.0386 |

Chr., chromosome; MI, meconium ileus; OR, odds ratio.

*The North American replication collection consisted of 1,140 individuals that correspond to the continuing enrolment at the North American sites. The French replication cohort consisted of 1,232 subjects collected at 38 French cystic fibrosis centers (Online Methods). The French samples were genotyped genome wide with the Illumina 660W-Quad BeadChip platform, and the North American replication samples were genotyped at the six SNPs using TaqMan Assays-on-Demand.*

The combined replication P value was calculated using the inverse-variance method, most powerful when different studies have the same direction of effect. The corrected type 1 error for the replication study was 0.007, using the conservative Bonferroni correction for the seven SNPs tested. *Meconium ileus status (MI:non-MI) breakdown for each site.*

There was another significantly associated SNP on chromosome 1 (rs4077469, P = 9.88 × 10⁻⁹) that was not included, as it is separated by only 128 bp from rs4077468 and is in perfect LD with this SNP.

© 2012 Nature America, Inc. All rights reserved.
SFDR\textsuperscript{17}. In addition, SNPs from SLC26A9 and two other apical genes, ATP2B2 and SLC9A3, showed association evidence with q values of <0.05 (Table 3). A gene-based analysis (Online Methods) of ATP2B2 (P = 0.0006) and SLC9A3 (P = 0.0001) provided evidence for allelic heterogeneity, as seen in comparing gene-level results to those from single-SNP analysis (Table 3). The association with SLC9A3 at the gene level was replicated in the French cohort (P = 0.017), whereas the association with ATP2B2 at the gene level was not (P = 0.283) (Supplementary Table 3).

Next, restricting analysis to the 3,814 SNPs annotated to the 155 apical membrane genes (which did not include SLC6A14), we tested the apical membrane prioritization hypothesis as part of the GWAS-HD. In this analysis, we observed evidence of association at genome-wide significance between meconium ileus and multiple genes encoding constituents of the apical plasma membrane (permutation P = 0.0002, testing all 3,814 SNPs jointly and not subject to multiple hypothesis testing; Fig. 2a,b). Even with the exclusion of SLC26A9 (as well as SLC6A14), the apical membrane hypothesis remained significant (P = 0.0058). Thus, GWAS-HD further establishes the involvement of other genes coding for apical membrane constituents, despite insufficient power to detect individual SNPs, even within the context of our prioritized GWAS. For comparison, we also constructed a null hypothesis list of genes encoding nuclear membrane–localized proteins. As expected, the 224 GO-annotated nuclear envelope genes tagged by 3,537 GWAS SNPs showed no relationship with meconium ileus (permutation P = 0.4639; Fig. 3).

The French cohort with genome-wide data provided independent validation of the apical membrane hypothesis (permutation P = 0.022; Fig. 2c,d and Online Methods). The statistical significance of this gene set (which excluded SLC6A14) in the French cohort remained after also excluding SLC26A9 (P = 0.021) and then both SLC26A9 and SLC9A3 (P = 0.023). Although analysis of the apical membrane hypothesis in a larger independent cohort should be considered as part of future efforts, replication in the smaller French cohort supports a common mechanism of contributing genes across the two study populations.

To determine which apical membrane genes were driving the association, the degree of genetic heterogeneity in meconium ileus and the common contributors across the French and North American samples, we implemented Lasso\textsuperscript{18}. In the North American sample, we jointly analyzed all 3,740 SNPs in the apical membrane genes (including in SLC26A9 and SLC9A3), as well as SNPs in SLC6A14, that were not in perfect linkage disequilibrium (LD) (Online Methods). Forty-eight SNPs spanning 36 different genes were retained by Lasso in the multivariate regression model (Supplementary Table 3). These SNPs jointly explained ~17% of the meconium ileus variation in the North American sample. The percentage explained by the same 48 SNPs in the French sample was 8.1% (Online Methods), with the drop presumably reflecting genetic heterogeneity and the smaller sample size of this replicate test collection. We then tested the significance of a score for each individual in the French cohort that was constructed from a weighted sum of the number of risk alleles (defined in the North American sample) across the 48 SNPs\textsuperscript{19}. The significant association between meconium ileus and this score (P = 0.0137; Online Methods) provided additional complimentary evidence for the nuclear envelope hypothesis (permutation P = 0.4639; Fig. 3).

**Table 3** Ranked SNPs with q values < 0.05 from GWAS or GWAS-HD

| SNP            | Chr. | Position (bp) | P value | q value<sup>a</sup> | Rank<sup>b</sup> | GWAS-HD q value<sup>a</sup> | Rank<sup>b</sup> | Apical<sup>c</sup> | Gene          |
|----------------|------|---------------|---------|---------------------|-----------------|-----------------------------|-----------------|------------------|---------------|
| rs3788766      | X    | 115480867     | 1.28 × 10<sup>-12</sup> | 0        | 1                | 0                           | 1               | 0                | SLC6A14       |
| rs4077469      | 1    | 204181380     | 9.88 × 10<sup>-9</sup> | 0.0018   | 3                | 0                           | 2               | 1                | SLC6A14       |
| rs4077469      | 1    | 204181508     | 9.88 × 10<sup>-9</sup> | 0.0018   | 2                | 0                           | 3               | 1                | SLC6A14       |
| rs7512462      | 1    | 204166218     | 2.14 × 10<sup>-8</sup> | 0.0023   | 5                | 0                           | 4               | 1                | SLC6A14       |
| rs7419153      | 1    | 204183932     | 1.01 × 10<sup>-7</sup> | 0.0091   | 6                | 0.0001                      | 5               | 1                | SLC6A14       |
| rs7419521      | 1    | 204177506     | 4.49 × 10<sup>-7</sup> | 0.0346   | 7                | 0.0003                      | 6               | 1                | SLC6A14       |
| rs12047830     | 1    | 204183322     | 3.72 × 10<sup>-6</sup> | 0.167    | 12               | 0.0022                      | 7               | 1                | SLC6A14       |
| rs5905283      | X    | 115479909     | 1.69 × 10<sup>-8</sup> | 0.0023   | 4                | 0.0045                      | 8               | 0                | SLC6A14       |
| rs1318819<sup>d</sup> | 3   | 10413649      | 2.11 × 10<sup>-5</sup> | 0.6309   | 18               | 0.0107                      | 9               | 1                | ATP2B2       |
| rs4684689<sup>d</sup> | 3   | 10423426      | 2.76 × 10<sup>-5</sup> | 0.6828   | 20               | 0.0123                      | 10              | 1                | ATP2B2       |
| rs495435<sup>d</sup> | 3   | 10443723      | 4.80 × 10<sup>-5</sup> | 0.7766   | 30               | 0.0191                      | 11              | 1                | ATP2B2       |
| rs12741299     | 1    | 204181139     | 1.23 × 10<sup>-4</sup> | 0.7766   | 81               | 0.0426                      | 12              | 1                | SLC6A14       |
| rs1874361      | 1    | 204174809     | 1.31 × 10<sup>-4</sup> | 0.7865   | 90               | 0.0426                      | 13              | 1                | SLC6A14       |
| rs1756316<sup>d</sup> | 5   | 550624        | 1.47 × 10<sup>-4</sup> | 0.796    | 98               | 0.0437                      | 14              | 1                | SLC9A3       |

Chr., chromosome.

<sup>a</sup>The q value is a genome-wide-adjusted P value that controls the false discovery rate. <sup>b</sup>The rank indicates the genome-wide ordering of the SNPs based on the original association evidence (GWAS) or the SDFR q value after incorporating the apical plasma membrane hypothesis (GWAS-HD). The GWAS and GWAS-HD rank results of all 543,927 SNPs are provided in Supplementary Table 3. <sup>c</sup>The SNPs from ATPT2B2 and SLC9A3 identified by GWAS-HD. Results in the French replication sample are P = 0.1606, 0.1948, 0.0674 and 0.0665 for rs1318819, rs4684689, rs495435 and rs17563161, respectively, where rs17563161 was imputed.

---

**Figure 3** Assessment of the nuclear envelope null hypothesis. A list was generated of 231 genes associated with the nuclear envelope, as defined by GO annotation. In total, 3,537 GWAS SNPs were within 10 kb of the boundaries of 224 tagged genes (NCBI36/hg18). A priori, the nuclear envelope list should not contain genes associated with meconium ileus (permutation P = 0.4639; Fig. 3).
evidence of common contributors between the two cohorts, with SNPs in SLC9A3 and SLC6A14 being two specific examples.

In summary, a conventional GWAS identified SNPs in SLC6A14 and SLC26A9 to be significantly associated with meconium ileus. GWAS-HD single-SNP analysis identified the same SNPs in SLC6A14 and SLC26A9, as well as SNPs in SLC9A3, and GWAS-HD multi-SNP analysis spanning 155 genes provided evidence that multiple constituents of the apical plasma membrane are collectively associated with meconium ileus. GWAS-HD can also be applied to other Mendelian disorders or even to complex traits, provided that there is a hypothesis based on underlying biology and that a list of participating, relevant genes can be compiled.

Although gene prioritization has been used in other approaches, such as pathway or gene enrichment analyses\(^{20–22}\), its use in GWAS-HD involves key differences. First, in contrast to the previous inclusion-exclusion approaches where all genotyped SNPs are not analyzed simultaneously, GWAS-HD performs parallel single-SNP analysis of all GWAS SNPs and multi-SNP analysis focusing on the set of SNPs of interest, respectively. The prioritized single-SNP analysis interrogates all available SNPs via the SFDR control, yet enables increased statistical power for regions favored a priori. For example, SLC6A14 would be omitted by inclusion-exclusion approaches, although it was the highest ranked gene for association with meconium ileus in the GWAS-HD analysis. Second, methods such as interactive pathway analysis can be restrictive, because contributing genes or proteins must relate to each other via direct or indirect links, which may be disturbed when a cog component (such as CFTR) is dysfunctional in the disease state. Although gene products that participate in maturation or delivery of CFTR may be contributory, consideration of local components of processes that may compensate for the ion and fluid flow disturbance in cystic fibrosis is enabled in the apical membrane hypothesis. Third, distinct from an exhaustive search of all plausible interactive pathways, GWAS-HD focuses on a single biological hypothesis and provides statistical significance for all genes involved jointly, alleviating the multiple testing burden.

It should be noted that the specific statistics or models used in our GWAS-HD application, such as SFDR, Lasso and the sum and score statistics, may not be the most powerful ones in any specific setting. For example, the adaptive rank truncated product statistic\(^{23}\) could be used to identify the common subset of associated apical membrane genes across two samples, and there are alternative weighting and prioritization approaches\(^{24}\).

The SLC9A3 gene codes for a sodium-hydrogen exchanger that, when disrupted, has been shown to decrease intestinal obstruction in a cystic fibrosis mouse model\(^{25}\). SLC6A14 codes for a sodium- and chloride-dependant neutral and basic amino-acid transporter\(^{26,27}\). SLC26A9 encodes an anion transporter, likely a chloride channel, with multiple modes that include chloride-bicarbonate exchanger and sodium-anion cotransporter capabilities\(^{28}\). The SLC26A9 protein has also been reported to physically interact with CFTR\(^{29}\) and to be influenced by CFTR activity, at least in lung-related tissues\(^{30}\).

It is notable that SLC9A3 has previously been associated with infections and pulmonary function in cystic fibrosis\(^{31}\). In the consortium discovery sample, rs6864158 (minor allele frequency (MAF) = 0.43) in SLC9A3 was associated with both a quantitative lung phenotype based on multiple forced expired volume in 1 s (FEV1) measurements (\(P = 0.0003\), analyzed previously\(^{32}\)) and with meconium ileus (\(P = 0.0001\)). This provides evidence that some genes may have a role in multiple cystic fibrosis comorbidities. Both SLC6A14 and SLC26A9 are robustly expressed in human lung epithelium and sweat gland, as well as in intestinal epithelium, as measured by RT-PCR (data not shown). We anticipate that meconium ileus modifier genes may also influence early pathology in other cystic fibrosis–affected organs and could therefore provide important insights into the mechanisms of cystic fibrosis disease severity and comorbidity.

These findings collectively have important practical implications for cystic fibrosis, where therapeutic strategies should consider pharmacologic modulation of epithelial function together with complementary approaches aimed at directly improving the function or delivery of the mutated CFTR gene product to the apical membrane\(^{32}\).

**METHODS**

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

**ACKNOWLEDGMENTS**

The authors would like to thank the study participants, the Cystic Fibrosis Foundation (CFF) Patient Registry, the University of North Carolina DNA Laboratory and the Genome Quebec and McGill University Innovation Centre. We thank I. Wong for technical help with the replication genotyping, the computer specialists from the French study J.-F. Vibert and M. Mahloul and clinical research assistants Michon, A. Blondel and P. Touche involved in the study design and subject recruitment of the French sample, E. Hawbaker and A. Dang for data coordination in the Gene Modifier Study, Y. Berthiaume, A. Sandford and P. Pare for ascertainment of Canadian phenotype data and DNA, J. Breaton, M. Christofi, N. Anderson, K. Keenan, C. Taylor and J. Avolio for coordinating and verifying the Canadian data collection, E. Crowdy for database development and management in Canada, S. Norris, A. Kohl, P. Miller, L. Charnin, W. Hannah and S. Adams for recruitment, phenotyping and data entry at the University of North Carolina, W. Wolf for DNA analysis and H. Kelkar, T. Randall and A. Xu for bioinformatics at the University of North Carolina, and N. Wang, K. Kaniecki, J. Bonner and C. Watson for technical assistance at Johns Hopkins University. This work was supported in part by Genome Canada, through the Ontario Genomics Institute per research agreement 2004-OGI-3-05 (to P.R.D.), with the Ontario Research Fund, Research Excellence Program; the University of Toronto McLaughlin Centre (to L.S.); the Ontario Ministry of Research and Innovation Early Researcher award and the Natural Sciences and Engineering Research Council (to L.S.); Cystic Fibrosis Canada (CFC; to P.R.D., IZ. and L.S.); the Natural Sciences and Engineering Research Council (NSERC; F250053–2008 to L.S.); the Canadian Institutes of Health Research (CIHR; MOP 84287 to L.S. and MOP 258916 to L.J.S.); the Lloyd Carr–Harris Foundation; the US National Institutes of Health (R01 HL68927 (G.R.C.), K23 DK083551 (S.M.B.), R01 HL088890 (M.R.K.), R01 DK066368 (M.R.K.), R01 HL095396 (M.R.K.) and HG-0004314 (L.J.S.)); the US CFF (CU350906P0, R025–CR07, KNOWLE00/A, RDP)–R026–CR07 and DRUMM0A00); the Flight Attendant Medical Research Institute (FAMRI2006 to G.R.C.); the Lawson Wilkins Pediatric Endocrine Society Clinical Scholar Award (to S.M.B.); INSERM; AP–HP; Université Pierre et Marie Curie Paris; Agence Nationale de la Recherche (R09186DS to H.C.); Direction Générale de la Santé; Association Vaincre la Mucoviscidose; and Groupement d’Intérêt Scientifique (GIS)–Institut des Maladies Rares. Funds for genome-wide genotyping of North American participants were generously provided by the US CFF. R.D. received a Joint Fellowship from the CIHR and Ontario Women’s Health Council.

**Contributing cystic fibrosis centers and principal investigators for the North American study collection:** 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. Abrams, University of Iowa Hospitals & Clinics, Iowa City, Iowa, USA; G. Ayliffe, Lutheran General Children's Hospital, Park Ridge, Illinois, USA; C. Allard, Hôpital de Châteauvêtre, Châtenay, France; T. Arndt; R. Amaro, University of Texas at Tyler Health Center, Tyler, Texas, USA; R. Anbar, State University of New York (SUNY) Upstate Medical University, Syracuse,
12. Hodges, C.A., Grady, B.R., Mishra, K., Cotton, C.U. & Drumm, M.L. Cystic fibrosis growth retardation is not correlated with loss of Cftr in the intestinal epithelium. Am. J. Physiol. Gastrointest. Liver Physiol. 301, G528–G536 (2011).

13. Carbon, S. et al. AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288–289 (2009).

14. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29 (2000).

15. GO-Consortium. The Gene Ontology in 2010: extensions and refinements. Nucleic Acids Res. 38, D331–D335 (2010).

16. 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).

17. 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).

18. Tibshirani, R. Regression shrinkage and selection via the lasso. J. R. Stat. Soc., B 58, 267–288 (1996).

19. Purcell, S.M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752 (2009).

20. Pedroso, I. & Breen, G. Gene set analysis and network analysis for genome-wide association studies. in Genetics of Complex Human Diseases—A Laboratory Manual (eds. Al-Chalabi, A. & Almasy, L.) 194–200 (Cold Spring Harbor Laboratory Press, New York, 2009).

21. Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genome-wide association studies. Am. J. Hum. Genet. 81, 1278–1283 (2007).

22. Cantor, R.M., Lange, K. & Sinnesheimer, J.S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. Am. J. Hum. Genet. 86, 6–22 (2010).

23. Yu, K. et al. Pathway analysis by adaptive combination of P-values. Genet. Epidemiol. 33, 700–709 (2009).

24. Thomas, D.C. et al. Use of pathway information in molecular epidemiology. Hum. Genomics 4, 21–42 (2009).

25. Bradford, E.M., Sartor, M.A., Gawenis, L.R., Clarke, L.L. & Shull, G.E. Reduced NHE3-mediated Na+ absorption increases survival and decreases the incidence of intestinal obstructions in cystic fibrosis mice. Am. J. Physiol. Gastrointest. Liver Physiol. 296, G886–G898 (2009).

26. Fragas, S., Serrao, M.P. & Soares-da-Silva, P. L-type amino acid transporters in two intestinal epithelial cell lines function as exchangers with neutral amino acids. J. Nutr. 132, 733–738 (2002).

27. Höglund, P.J., Adzic, D., Scicluna, S.J., Lindblom, J. & Fredriksson, R. The repertoire of solute carriers of family 6: identification of new human and rodent genes. Biochem. Biophys. Res. Commun. 335, 175–189 (2005).

28. Chang, M.H. et al. SLC26a9—anion exchanger, channel and Na+ transporter. J. Membr. Biol. 228, 125–140 (2009).

29. Chang, M.H. et al. SLC26a9 is inhibited by the R-region of the cystic fibrosis transmembrane conductance regulator via the STAS domain. J. Biol. Chem. 284, 28306–28318 (2009).

30. Bertrand, C.A., Zheng, R., Pilewski, J.M. & Frizzelli, R.A. SLC26A9 is a constitutively active, CFTR-regulated anion conductance in human bronchial epithelia. J. Gen. Physiol. 133, 421–438 (2009).

31. Dorfman, R. et al. Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis. Pediatr. Pulmonol. 46, 385–392 (2011).

32. Pedemonte, N. et al. Small-molecule correctors of defective ΔF508-CFTR cellular processing identified by high-throughput screening. J. Clin. Invest. 115, 2564–2571 (2005).

33. The ENCODE Project Consortium. A user’s guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 9, e1001046 (2011).
ONLINE METHODS

Human subjects. Consent was obtained from all participants of the North American Cystic Fibrosis Gene Modifier Consortium (NACFGMC), with procedural approval from the Institutional Review Boards of Johns Hopkins University, the University of North Carolina at Chapel Hill and Case Western Reserve University and the Research Ethics Board of The Hospital for Sick Children. Consent was also obtained for participants from France with procedural approval (Comité de Protection des Personnes (CPP) 2004/15) and information collection approval by Commission Nationale de l’Informatique et des Libertés (CNIL) (04.404).

Recruitment and inclusion. Individuals with cystic fibrosis and cystic fibrosis–related phenotypes, including impaired lung function and meconium ileus, were collected by the NACFGMC (Table 1). The meconium ileus GWAS was restricted to 3,763 subjects with severe (pancreatic exocrine–insufficient) CFTR genotypes who were of European decent. The 1,140 North American replication participants corresponded to the continuing collection at all sites (351 from TSS, 448 from the University of North Carolina at Chapel Hill and Case Western Reserve University and 341 from CGS) with known meconium ileus status or evidence of an abdominal scar.

There are 49 cystic fibrosis centers in France, caring for ~5,000–6,000 individuals with cystic fibrosis. Prospective enrollment of individuals with cystic fibrosis was initiated in 38 centers in 2006. The French replication cohort included 1,362 phenotyped individuals who were enrolled before June 2010, all of whom were over 6 years of age with both parents born in a European country and had two severe CFTR mutations.

Genotyping. NACFGMC GWAS subjects were genotyped simultaneously using the Illumina 610-Quad BeadChip, as previously described. DNA of the North American replication sample was extracted from whole blood or transformed lymphocytes and quantified with fluorimetry. Genotyping was performed with allele-specific fluorescent probes in TaqMan SNP Genotyping Assays (Custom or On-Demand; Applied Biosystems) as recommended. DNA of the French replication sample was obtained from whole blood and hybridized to the Illumina CNV370-Duo BeadChip for the first 299 subjects (those included before June 2009) and the Illumina 660W-Quad BeadChip for the other subjects at the Centre National de Genotyage (CNG).

Quality control filtering. The GWAS discovery subjects were first cleaned together as part of an initial quality control effort by NACFGMC (see ref. 5 for details) and then underwent additional quality control analysis for the purpose of this study. Samples with a heterozygosity proportion of <28% or sex incongruity and those of non-European ancestry, as determined by principal-component analysis using EIGENSTRAT34, were excluded. Using identity–by-descent (IBD) estimates from PLINK35, 12 cryptic full siblings were identified and adjusted for relationship. Only one randomly selected individual from each of the ten cryptic monozygotic pairs was retained, and parents of two cryptic parent-offspring pairs were removed from further analysis. In total, 3,763 samples were analyzed, among which only 14 individuals had genotype missing rates of >1% (maximum = 2.8%), and 543,927 SNPs with MAF of >2% were analyzed, among which 2,916 had missing rates of >2% (maximum = 10%). The missing rate for all 3,814 apical membrane SNPs and 15 SNPs from SLC6A14 was <2%.

For the North American replication sample, end-point fluorescence was measured with the plate reader component of the 7900HT Real Time PCR System (Applied Biosystems) and analyzed with TaqMan Genotyper software for allele discrimination with call rates of >95%. To insure quality control, 2% of the samples were run in duplicate, and 1% of the samples corresponded to individuals used in the initial GWAS control, with cross-platform concordance rates of >99%.

The French replication samples with missing rates of >5%, sex incongruity or pairwise IBD estimates of >40% were excluded, yielding 1,232 genotyped and phenotyped affected individuals. SNPs present only on the CNV370-Duo chip, with missing rates of >10% or MAF of <6%, were excluded. Overall, 554,792 SNPs were analyzed, of which 256,756 were genotyped on both chips.

GO annotation. Gene lists were generated based on annotation of genes encoding apical membrane and nuclear envelope constituents. A list of 157 apical membrane genes was annotated using the AmiGO tool13 (version 1.7; 28 March 2010), based on Gene Ontology data14 (GO:00163245) and using the cell location search phrase ‘apical plasma membrane’ with restriction to Homo sapiens. A list of 231 nuclear envelope genes was generated on the basis of GO annotation (GO:0005635).

Imputation. Using MACH and minima53, genome-wide imputation was conducted for the 3,763 GWAS subjects. The reference sample was the 87 CEU subjects extracted from the European continental group (EUR) of the 100 Genomes Project November 2010 release64. Imputed genotype data for 7,245,292 SNPs with MAF of >2% and estimated imputation accuracy of >0.3 (MACH R2) were analyzed, and no new regions with evidence of association at genome-wide significance were identified. In SLC26A9 and SLC9A3, the best imputed SNPs were only marginally more significant than any genotyped SNP (5.94 × 10−7 versus 9.88 × 10−7 and 1.09 × 10−6 versus 6.22 × 10−5, respectively), whereas in SLC6A14 the minimum P was given by a genotyped SNP.

Statistical methods. Generalized estimating equations (GEE) were used for the GWAS, with an exchangeable correlation structure to account for full sibbing relationships in the data (Geeqlm function in R). Genotypes were coded additively for autosomal SNPs and SNPs on the X chromosome in females, and 0 and 2 were used for SNPs on the X chromosome in males. A site covari-ate with four levels (Table 1) was included. Logistic regression in a sample of 3,199 unrelated individuals with the site covariate and the first seven principal components was also conducted, and the results were consistent with the GEE analysis of the full 3,763 subjects. (Therefore, the principal components were not included in subsequent analysis and permutation tests.) The French GWAS was used logistic regression with additive genotype coding (PLINK56 v1.07 for autosomal SNPs and R for X-chromosome SNPs).

GWAS-HD was used to accomplish two goals in parallel (Supplementary Fig. 4): (i) single-SNP analysis to establish the significance of individual SNPs at the genome-wide level after weighting according to a prioritization hypothesis (for example, the apical membrane hypothesis), analyzing and re-ranking all GWAS SNPs (for example, 543,927 SNPs in the meconium ileus study), and (ii) multi-SNP analysis to test the significance of the prioritization hypothesis itself, by assessing whether the group of SNPs defined by the hypothesis (the 3,814 SNPs annotated to 157 apical membrane genes; Supplementary Table 3) collectively showed significantly smaller P values (or larger association statistics; Fig. 2a,b) than would be expected under the null hypothesis of no association.

To carry out the first task, GWAS SNPs were assigned to a high-priority group (the 3,814 apical membrane SNPs) or a low-priority group (all other 540,113 SNPs). An SFDR16–17 was then applied, and q values were calculated separately in each group. A given SNP was determined to have statistical significance if its q value was <0.05; each SNP was re-ranked genome wide according to its q value (the original GWAS P values were used to guide order if q values were identical). This is equivalent to a weighted P-value approach57 but with robust weights. In the meconium ileus application, the 3,814 apical membrane SNPs were given weight of −124, and the remaining 540,113 SNPs were given weight of −0.13.

The second task was to determine the statistical significance of the apical membrane hypothesis used to prioritize the GWAS, which involved testing 3,814 SNPs simultaneously (or 3,420 SNPs in the French replication cohort). The meconium ileus phenotype was permuted (to preserve the LD pattern between SNPs simultaneously (or 3,420 SNPs in the French replication cohort). The empirical null hypothesis of no association was tested by permutation sampling (or 1,000 times in the French cohort). For each permutation sample, corresponding sets of SNPs were permuted and the significance of the prioritization hypothesis was assessed by a weighted association test (for a given site and the first seven principal components was also conducted, and the results were consistent with the GEE analysis of the full 3,763 subjects. (Therefore, the principal components were not included in subsequent analysis and permutation tests.) The French GWAS was used logistic regression with additive genotype coding (PLINK56 v1.07 for autosomal SNPs and R for X-chromosome SNPs).

GWAS-HD was used to accomplish two goals in parallel (Supplementary Fig. 4): (i) single-SNP analysis to establish the significance of individual SNPs at the genome-wide level after weighting according to a prioritization hypothesis (for example, the apical membrane hypothesis), analyzing and re-ranking all GWAS SNPs (for example, 543,927 SNPs in the meconium ileus study), and (ii) multi-SNP analysis to test the significance of the prioritization hypothesis itself, by assessing whether the group of SNPs defined by the hypothesis (the 3,814 SNPs annotated to 157 apical membrane genes; Supplementary Table 3) collectively showed significantly smaller P values (or larger association statistics; Fig. 2a,b) than would be expected under the null hypothesis of no association.

To carry out the first task, GWAS SNPs were assigned to a high-priority group (the 3,814 apical membrane SNPs) or a low-priority group (all other 540,113 SNPs). An SFDR16–17 was then applied, and q values were calculated separately in each group. A given SNP was determined to have statistical significance if its q value was <0.05; each SNP was re-ranked genome wide according to its q value (the original GWAS P values were used to guide order if q values were identical). This is equivalent to a weighted P-value approach57 but with robust weights. In the meconium ileus application, the 3,814 apical membrane SNPs were given weight of −124, and the remaining 540,113 SNPs were given weight of −0.13.

The second task was to determine the statistical significance of the apical membrane hypothesis used to prioritize the GWAS, which involved testing 3,814 SNPs simultaneously (or 3,420 SNPs in the French replication cohort). The meconium ileus phenotype was permuted (to preserve the LD pattern between SNPs) within data from each consortium site and independently 10,000 times (or 1,000 times in the French cohort). For each permutation sample, corresponding association analysis was performed using GEE, and a sum of the Wald X2 (1 degree of freedom) association statistics of the 3,814 (or 3,420 in the French cohort) SNPs was obtained. The empirical P value was calculated as the proportion of the permutation samples whose sum statistic was larger than that in the observed sample.

The gene-based analysis was similar to the permutation test, with summation limited to the SNPs within 10 kb of the boundaries of each gene. In total, 156 gene-based permutation tests were performed (155 apical membrane genes
and SLC6A14), and the conservative Bonferroni-adjusted significance level was 0.05/156 = 0.0003 (Supplementary Table 3).

To determine the specific subset of the apical membrane SNPs contributing to meconium ileus susceptibility, a multivariate analysis using penalized logistic regression (Lasso) was performed on 3,199 unrelated individuals (574 affected individuals) extracted from the original 3,763 subjects in the North American GWAS sample. The 3,814 apical membrane SNPs and 15 SNPs within SLC6A14 were considered in the joint analysis. After removing 93 SNPs in perfect LD with one another ($r^2 = 1$), 3,740 SNPs (and the site covariate) were included in the Lasso analysis using the glmnet package in R. The default option to standardize all predictors was turned off; and the optimal value of the tuning parameter $\lambda$ was chosen based on the mode of 50 $\lambda$ values obtained from 50 repeated 10-fold cross-validations (to maximize the deviance). The final model included 48 SNPs spanning 36 different genes.

We used pseudo $R^2$ (ref. 11) as an estimate of the phenotypic variance explained by the SNPs of interest. Calculations used the lrm function in R by regressing meconium ileus on (i) the 7 SNPs in Table 2 and (ii) the 48 SNPs selected by Lasso in the North American sample. The calculation was done separately in the North American (~5% and ~17%, respectively) and French (~4% and ~8%, respectively) samples. In the French sample, some of the 48 SNPs were not genotyped or did not pass quality control and were replaced by imputed SNPs with the highest LD.

34. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
35. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81, 559–575 (2007).
36. The 1000 Genomes Project Consortium. et al. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073 (2010).
37. Roeder, K., Bacanu, S., Wasserman, L. & Devlin, B. Using linkage genome scans to improve power of association in genomescans. Am. J. Hum. Genet. 78, 243–252 (2006).